Secuenciación masiva del genoma y transcriptomas de *Vitis vinifera* cv. Tannat: polifenoles durante el desarrollo del grano de uva

Mag. Cecilia Da Silva

Orientador: Dr. Francisco Carrau

Co-orientadores: Dra. Carina Gaggero

Dr. Massimo Delledonne

Tesis de Doctorado en Ciencias Biológicas

PEDECIBA

2016

IIBCE

UDELAR FQ FACULTAD DE QUIMICA

UNIVERSITA' DEGLI STUDI DI VERONA
Dipartimento di Scienze della Vita e della Riproduzione

ÍNDICE

Resumen…….....3

Introducción……..…….…….……4

Vitis vinífera……..5

Vitis vinífera cv Tannat..……..……..……..6

Desarrollo de la baya de uva…….8

Composición polifenólica de uvas y vinos Tannat….……..10

Características genéticas de Tannat.…….……..……..13

La tecnología de Secuenciación Masiva para el estudio de Tannat…..…….14

Objetivos……..……..16

Objetivo general……..……..……..16

Objetivos específicos……16

Bibliografía……..17

Resultados……...26

Capítulo I: El alto contenido de polifenoles en bayas del cultivar de vid Tannat es conferido principalmente por los genes que no se comparten con el genoma de referencia……..……27

Capítulo II: Los altos niveles de procianidinas galoileadas en las semillas de uva Tannat se correlacionan con la expresión de un nuevo gen epicatequina galoil transferasa……..……..43

Capítulo III: Desentrañando las bases genéticas del intenso color púrpura de las uvas Tannat…….……..……87

Conclusiones y perspectivas futuras…….……..……112

RESUMEN

Tannat (*Vitis vinifera*) es la variedad de vid vinífera más cultivada en Uruguay para la producción de vinos tintos de alta calidad. Sus bayas poseen niveles de compuestos polifenólicos inusualmente altos (antocianinas y taninos), produciendo vinos con intenso color púrpura y alto poder antioxidante. Los taninos dan estructura en boca a los vinos, se sintetizan en las semillas antes de envero y en Tannat más de un 40% de los taninos están galoileados, lo que determina un mayor poder antioxidante. Las antocianinas dan color a la piel de las uvas y al vino y su síntesis comienza durante el envero. Las nuevas tecnologías de secuenciación masiva permiten caracterizar variantes genómicas entre diferentes variedades y diferencias transcriptómicas entre diferentes tejidos y estadíos de desarrollo de la vid.

En esta tesis se secuenció el genoma del clon de Tannat uruguayo UY11 mediante la tecnología Illumina. Mediante RNA-Seq se secuenciaron los trancriptomas de cáscaras y semillas de uvas Tannat del clon francés 717 en diferentes momentos del desarrollo de la baya. Los transcriptomas se utilizaron para la anotación del genoma y para determinar patrones de expresión diferencial en los diferentes tejidos durante este proceso con la finalidad de dilucidar las bases genómicas y transcriptómicas de las características particulares de las uvas Tannat.

La particularidad de este trabajo radicó en el análisis por separado de cáscaras y semillas de uva, y también que en el muestreo realizado en el momento de 50% de envero se separaron las uvas totalmente verdes de las completamente coloreadas y se analizaron como muestras diferentes.

Al comparar el genoma de Tannat UY11 con el del clon de Pinot Noir PN40024 (genoma de referencia) se encontró una expansión de las familias génicas relacionadas con la biosíntesis de polifenoles. Los análisis de expresión diferencial focalizados en semillas permitieron determinar un gen que codifica para la enzima epicatequina galoil tranferasa (ECGT) involucrada en la síntesis de taninos galoileados. Se determinaron los patrones de expresión diferencial de las familias génicas involucradas en el proceso de cambio de color en envero y se determinó la participación de varios genes nuevos (no anotados previamente en el genoma de referencia) y varietales (ausentes en el genoma de referencia). También se determinaron las variantes genéticas en factores de transcripción reguladores de este proceso, encontrando que el retrotransposon Gret1 está presente en uno de los alelos de VvmybA1 en todos los 7 clones de Tannat analizados, y que existen SNPs no reportados previamente en VvMYBA2 y VvMYBA3 del clon UY11.

Este trabajo realiza aportes originales acerca de las bases moleculares de la biosíntesis de antocianinas y taninos durante el desarrollo de la uva vinífera insignia del Uruguay

INTRODUCCIÓN

Vitis vinifera

La familia Vitaceae posee 15 géneros, uno de ellos es el género *Vitis*, el cual se compone de alrededor de 60 especies (base de datos de GRIN, http://www.ars-grin.gov/cgi-bin/npgs/html/splist.pl?12735M). La Vid es uno de los cultivos perennes más antiguos e importantes en el mundo. Se encuentra en el puesto 14 de producción mundial según la FAO, con una producción de casi 68 millones de toneladas anuales, abarcando 7,2 millones de hectáreas de tierra cultivable (cifras del año 2010, FAOSTAT database: http://faostat.fao.org).

A nivel mundial casi el 98% de los viñedos están constituidos por variedades de *Vitis vinifera* subsp. sativa (*V. vinifera* de aquí en adelante). Esta se utiliza como una fuente de fruta fresca, pasas, jugo, y también se fermenta para producir vino, licores y bebidas destiladas. Cabe aclarar que todo el vino producido en el mundo proviene casi exclusivamente de esta especie. La mayoría de las uvas que se cosechan son fermentadas en el vino. Datos de la OIV (Organización Internacional del Vino) establecen que en el año 2008 se produjeron mundialmente 269.785.000 hectolitros de vino (StatOIV Extracts: http://www.oiv.int/oiv/info/esstatoivextracts2), por lo que el impacto para este producto es mucho mayor que el valor de las uvas mismas (Mullins *et al.* 1992).

V. vinifera es nativa de la zona al sur del Cáucaso y el Mar Caspio (Jay 1996). Las uvas cultivadas fueron domesticadas en el medio silvestre a partir de *V. vinifera subsp. sylvestris* (This 2006), que se dispersó ampliamente desde el Medio Oriente hacia Europa occidental (Zohary 2000). Más recientemente la vid ha sido ampliamente cultivada en el Nuevo Mundo y su cultivo se está expandiendo hacia Asia. Esta puede ser cultivada entre las latitudes 50°N al 40°S y hasta 3.000 metros sobre el nivel del mar. Su cultivo se concentra en las regiones con un clima de tipo mediterráneo, pero se cultiva en la mayoría de las regiones templadas.

El vino ha sido un componente importante de muchas culturas, desde el desarrollo de la vinificación en Irán entre 5440 y 5000aC (McGovern 2003). En un análisis de ánforas egipcias, objetos funerarios del rey egipcio Tutankamón (1332-1322 aC), se encontraron residuos de vino que sugieren que el vino se consideraba importante para la vida después de la muerte (Guasch-Jané *et al.* 2006). Hoy en día es un ícono tradicional de la dieta mediterránea (Panagiotakos *et al.* 2004).

Además de su importancia económica y cultural, el consumo de uva y vino tiene numerosos beneficios nutricionales y de salud para los seres humanos (German y Walzem 2000, Iriti y Faoro 2009, Wollin y Jones 2001). El vino se considera una fuente clave de

metabolitos secundarios promotores de la salud, especialmente los polifenoles. Por ejemplo, hay más de 200 compuestos polifenólicos de los vinos tintos que se cree que actúan como antioxidantes. En particular, un compuesto antioxidante, trans-resveratrol, se ha demostrado que desempeña un papel en la prevención de enfermedades del corazón (arteriosclerosis) (Bradamante et al. 2004) y el cáncer (Bruno et al. 2003). El resveratrol disminuye el proceso de envejecimiento en los animales (Baur y Sinclair 2006), actúa como una molécula de señalización en el cerebro (Dore 2005), y regula la expresión de genes que están involucrados en el ciclo celular y la proliferación celular en las células de la próstata humana (Jones *et al.* 2005). Hoy en día se considera que el efecto beneficioso para la salud del consumo de polifenoles se debe a sus efectos en procesos de transducción de señales, así como en su interacción con el microbioma humano (Kennedy, 2014; Dueñas et al., 2015).

Vitis vinifera cv. Tannat

Una visión general de la cantidad de *V. vinifera* plantadas en Uruguay muestra que las uvas viníferas representan el 76% de la viña, y de éstas, el 29% es *V. vinifera* cv. Tannat (cifras del año 2011 por INAVI: http://inavi.com.uy/sitio/home/home/index.php?menu=sub1b_2&menu2=sub2b_8&menu3=sub3b_20&t=secciones&secc=435&sub=778&sub1=778&sub2=779), conformando el viñedo de Tannat más extenso del mundo, representando el doble de hectáreas cultivadas en su región de origen en Francia. Esta variedad es considerada hoy en día la variedad insignia de Uruguay (Carrau 1997). Hasta hace algunos años Uruguay era el único país del continente americano donde se encontraba esta uva, por esta razón la industria vitivinícola uruguaya estableció una estrategia para producir vino Tannat utilizando tecnología avanzada en viticultura y vinificación (Boido *et al.* 2003).

El Tannat es un antiguo varietal (Durquety y Houbart 1982) originario de Madiran en el suroeste de Francia (Viala y Vermorel 1903), región donde su cultivo es dominante y casi exclusivo, aunque hoy en día es una variedad casi desconocida en Europa. Esta fue introducida a nuestro país por Don Pascual Harriague, un francés de la región vasca, que plantó unos viñedos de Tannat en 1870, 400 km al norte de Montevideo, en las afueras de la ciudad de Salto (Galanti 1919). Entre 1930 y 1970, los productores de vino pensaron en términos de cantidad cada vez mayor en lugar de aumentar la calidad, y por esta razón el vino Harriague (nombre por el cual era conocido al Tannat en aquellos tiempos) no tenía muy buena reputación (Carrau *et al.* 2011).

Luego, en la década del 70, los productores de uva uruguayos fueron alentados a sustituir las viejas plantas de Tannat por nuevos clones certificados de proveedores comerciales franceses y junto con mejores prácticas vitivinícolas lograron mejorar la calidad del vino. Desde entonces la estrategia de Uruguay ha sido la de promocionar a Tannat como uva emblemática del país; y así se están empezando a producir vinos tintos de alta calidad que están compitiendo con éxito en los mercados internacionales de vino (Carrau *et al.* 2011).

Tannat tiene una característica principal única: ser la variedad tinta con mayor contenido de polifenoles. Esto le da ciertas características particulares a sus vinos, como ser un intenso color púrpura, gran estructura en boca y alto poder antioxidante. Se ha demostrado que los vinos Tannat, en comparación con los vinos tintos elaborados a partir de otras variedades de uva, poseen los mayores contenidos de pigmentos (Alcalde-Eon *et al.* 2006). También estos vinos son ricos en contenido de taninos (2 o 3 veces el contenido de polifenoles totales de vinos como Cabernet Sauvignon), lo que le da estructura al vino aunque ésto puede resultar en vinos con bajas intensidades aromáticas (Carrau 1997).

Durante la última década, se ha demostrado que la dieta mediterránea parece tener efectos beneficiosos en enfermedades cardiovasculares, un hecho que ha sido conocido como la "paradoja francesa" (St. Legere *et al.* 1979). El consumo regular y moderado de vino, especialmente el vino tinto, parece ser una de las principales razones de estos efectos saludables. Recientemente se ha relacionado directamente estos efectos con el poder antioxidante de los vinos tintos, en particular con la presencia de compuestos fenólicos (e.g. flavonoides) y el resveratrol (Carrau 2011).

La cantidad de flavonoides en el vino tinto depende de la variedad de uva, área de cultivo, exposición al sol, la técnica de vinificación y la edad del vino (Auger *et al.* 2004, Burns et al. 2000, Dell'agli *et al.* 2004). En este contexto se estudió el vino Tannat y se encontró que la capacidad antioxidante aumenta con el envejecimiento del vino (Echeverry et al. 2005)

Otros compuestos como las procianidinas también se postulan como componentes responsables de la paradoja francesa, de los cuales Tannat presenta los más altos niveles entre las uvas tintas (Corder *et al.* 2006). Se ha demostrado que los efectos promotores de la salud de procianidinas galoileadas son más importantes que las de las procianidinas no galoileadas (Lizarraga *et al.* 2007; Kim *et al.* 2014). También se asocia al resveratrol con estos efectos saludables. La medición de los niveles de resveratrol en vinos Tannat resultó en un promedio de 2,7 mg L-1, un valor superior a la reportada para el Pinot Noir, Merlot y Cabernet (Gu *et al.* 1999).

Desarrollo de la baya de uva.

El desarrollo y la maduración de las bayas de uva se ha estudiado intensamente debido a la singularidad de este proceso en la biología de las plantas y al deseo de entender las características fisiológicas, bioquímicas y moleculares que determinan la calidad de frutas y vino. El desarrollo de la baya es un proceso dinámico y complejo que involucra una cascada de cambios bioquímicos (Coombe & McCarthy 2000). Durante este desarrollo hay dos períodos de crecimiento rápido separados por un período de crecimiento sustancialmente menor, resultando en un patrón de crecimiento doble sigmoidal (Figura 1). Aunque la transición de un período a otro es continua, el caracter distintivo de este modelo doble sigmoidal permite a los viticultores convenientemente describir el crecimiento de la baya en tres etapas: fases I y III demarcan los dos períodos de crecimiento rápido y la fase II el período lento (Coombe 1992, Coombe y McCarthy 2000).

La etapa inicial de crecimiento de la baya (Fase I) se caracteriza por un rápido crecimiento y rápida división mitótica celular, resultando en el desarrollo de los tres tejidos que componen la baya: piel, pulpa y semilla (Pratt 1971). Dos importantes ácidos orgánicos, ácido tartárico y málico, se sintetizan y acumulan en las vacuolas, alcanzando concentraciones máximas en el final de la Fase I. La biosíntesis de taninos, principalmente en piel y semillas, y de hidroxicinnamatos, que son precursores importantes de los compuestos fenólicos volátiles, también se produce principalmente durante esta fase. La fase II se caracteriza por ser una fase de retardo durante la cual no hay aumento en tamaño de las bayas debido a que cesa la división celular (Harris *et al.* 1968). En este período las bayas son duras, verdes y con una alta acidez; y las semillas han alcanzado su tamaño final (Nitsch *et al.* 1960). Los procesos biosintéticos de esta fase no están bien caracterizados, pero se sabe que durante esta fase, justo antes del envero (el inicio de la fase III) termina la síntesis de taninos en la piel y las semillas, junto con el inicio de la acumulación de azúcares en la pulpa (Coombe 1992). La fase III del crecimiento de la baya (maduración) se caracteriza por la aparición del color (la acumulación de antocianinas en la cáscara de las uvas tintas) y el ablandamiento de las bayas. En esta fase las bayas duplican su tamaño debido exclusivamente a la expansión celular y no a división mitótica (Coombe 1992). En el inicio de la fase III, los azúcares (principalmente glucosa y fructosa) siguen acumulándose, y disminuye la concentración de ácidos orgánicos (Coombe y McCarthy, 2000). La biosíntesis de muchos compuestos de aromas volátiles es inducida durante esta fase (Lücker *et al.* 2004, Martin y Bohlmann 2004, Mathieu *et al.* 2005, Wang y Luca 2005). Sin embargo, los taninos en semillas disminuyen a través de procesos oxidativos durante esta fase, resultando en que los taninos se unan a la cubierta de la semilla, reduciendo los

componentes astringentes dentro de la baya. Los taninos de la piel comienzan a interactuar y unirse con los antocianos y así aumenta el tamaño y la complejidad de los polímeros de taninos.

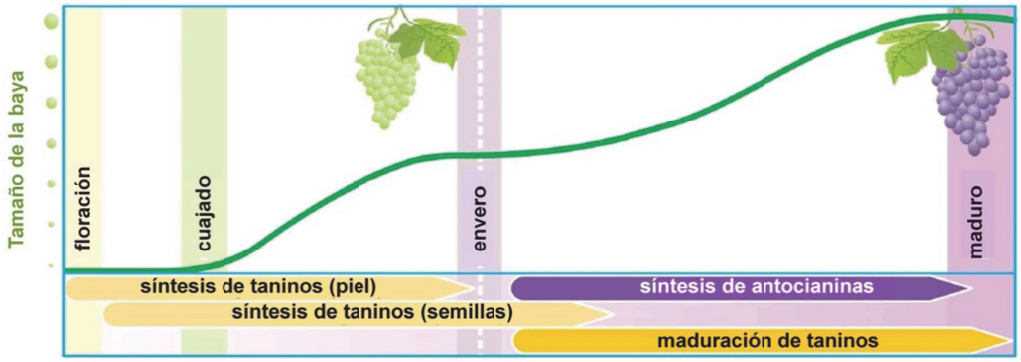


Figura 1: Esquema del desarrollo del grano de uva, indicando los periodos de síntesis de taninos y antocianos. Figura modificada de Robinson 2007.

Hay que tener siempre presente que la baya sintetiza los principales determinantes del aroma, color y sabor del vino. Los tres principales tejidos de la baya desempeñan diferentes funciones en la dispersión de semillas, cada uno contribuye a perpetuar las especies de uva. Esta especialización tiene un impacto directo en la calidad del vino, porque cada tejido contribuye con propiedades organolépticas específicas al vino (Grimplet ***et al.*** 2007).

El exocarpo del fruto (piel) contribuye a la integridad del conjunto de la baya con la protección de los tejidos internos contra daños mecánicos o de ataque de patógenos (Bargel y Neinhuis 2005). El color de la piel promueve la dispersión de semillas, proporcionando un alto contraste entre el follaje de fondo y las frutas (Burns y Dalen 2002), así como también proporciona protección contra la exposición a la luz UV (Solovchenko y Schmitz-Eiberger 2003). Para la elaboración del vino, la variabilidad de la composición de la piel juega un papel importante en la determinación del color, aroma y otras propiedades organolépticas del vino. La función principal del mesocarpo (la pulpa) es proporcionar un contenido alto de valor nutritivo para los agentes de dispersión, incluyendo una alta concentración de aminoácidos libres y azúcares de hexosa. Durante la producción de vino, el tejido de la pulpa aporta la mayoría de los azúcares, que se transforman en alcohol durante la fermentación. Los aromas surgen a partir de compuestos volátiles, tales como terpenos, norisoprenoides y tioles almacenados como compuestos conjugados a azúcares oaminoácidos en las vacuolas de las células del exocarpo (la piel) (Lund y Bohlmann 2006). Los aromas también están involucrados en la dispersión de semillas ya que sirven como

atrayentes importantes para los agentes de dispersión cuando los frutos están maduros y listos para consumir, pero hacen más susceptibles a la baya al ataque de patógenos (Slaughter 1999). La semilla, que contiene el embrión, es la principal fuente de monómeros de flavan-3-oles y procianidinas (taninos de semillas). Estos compuestos son responsables de importantes propiedades organolépticas del vino como amargor y astringencia (Robichaud y Noble 1990).

Composición polifenólica de uvas y vinos Tannat.

Las uvas y el vino contienen diferentes tipos de polifenoles. Las antocianinas son responsables del color del vino tinto, mientras que los flavan-3-oles (taninos) son los responsables de la astringencia del vino y la amargura. La polimerización de ambos grupos de polifenoles contribuye a la estabilización del color del vino añejado. Las antocianinas, así como otro grupo de polifenoles, los flavonoles, generalmente se encuentran como formas glucosídicas en casi todas las plantas superiores (Carrau *et al.* 2011). En la Figura 2 se muestra un esquema simplificado de la biosíntesis de polifenoles.

Los compuestos fenólicos se encuentran principalmente en las semillas y las capas internas de pieles de uva, mientras que la pulpa generalmente contiene cantidades muy bajas. Su concentración depende del cultivar y está influenciado por prácticas vitícolas y factores ambientales (Carrau *et al.* 2011).

El color del vino es una de las primeras características percibidas por los consumidores, influye en su aceptación e informa acerca de su edad, condiciones de almacenamiento, etc. Durante la vinificación las antocianinas se extraen desde la piel de la bayaal mosto, proporcionando la característica tonalidad de color rojo púrpura de los vinos tintos jóvenes (Carrau *et al.* 2011).

En el mercado internacional del vino hay una creciente demanda de vinos Tannat, ya que se considera un vino original, rico en estructura y con potencial de envejecimiento. De hecho, una de las características más notables de estos vinos es su color intenso, y se ha demostrado que los vinos Tannat, en comparación con los vinos tintos elaborados a partir de otras variedades de uva, poseen los mayores contenidos de pigmentos (Alcalde-Eon *et al.* 2006).

Por lo tanto, la fracción fenólica de esta variedad, hoy en día, está empezando a ser el objetivo de diferentes estudios. Las antocianinas principales y sus variaciones con los tratamientos clásicos en el viñedo también se han reportado para la uva Tannat (González-

Neves *et al.* 2004). La composición de antocianinas en los vinos Tannat jóvenes (Boido *et al.* 2006) y los cambios cualitativos y cuantitativos producidos durante el envejecimiento en las principales familias de pigmentos (antocianinas, piranoantocianos) han sido reportados en detalle (Boido *et al.* 2006). Las propiedades de color del vino se consideran con el fin de supervisar la evolución del color durante un período de 6 años de edad, y también para establecer la contribución del perfil de antocianinas a la calidad del vino (Boido *et al.* 2006).

Los flavonoles, que se encuentran en las pieles de la uva, se extraen al vino durante la vinificación. Su contenido en la uva Tannat fue estudiado y los principales compuestos cuantificados fueron los glucósidos y galactósidos de la quercetina y miricetina, y los glucósidos laricetin y syringetin (Boido *et al.* 2011).

Estos vinos se caracterizan también por su contenido alto en taninos, que le dan estructura al vino. Los taninos de la uva son componentes sensoriales importantes para la calidad del vino. En el vino, el término tanino describe una serie de compuestos polifenólicos, que incluye los monómeros de flavan-3-oles y las proantocianidinas.

Las proantocianidinas se encuentran en todos los tejidos de la uva, pero la piel contiene cantidades más bajas que las semillas, y sus características estructurales también son diferentes (Ricardo-da-Silva 1991). Los taninos de semilla de uva comprenden sólo las procianidinas (Labarbe *et al.* 1999, Prieur *et al.* 1994), mientras que los taninos de la piel abarcan tanto las procianidinas como las prodelfinidinas (Labarbe *et al.* 1999, Souquet *et al.* 1996). Las proantocianidinas de la piel tienen un mayor grado de polimerización medio (mDP) y un menor porcentaje de subunidades galoiliadas (unidas a ácido gálico) que los de las semillas (Cheynier *et al.* 1997, Moutounet *et al.* 1996).

Los flavan-3-oles de semillas mostraron valores mDP que van desde 2,5 a 3,0 con un aumento durante la maduración, mientras que el porcentaje de galoiliados (esterificación con ácido gálico) era alrededor del 40% con valores más bajos durante la madurez, siendo Tannat la variedad con la tasa de galoileados más alta (Boido *et al.* 2011). En los flavan-3-oles de la piel de Tannat, el porcentaje de prodelfinidinas varió entre 30% a 35%, pero las formas galoiliadas no se detectaron (Boido *et al.* 2011).

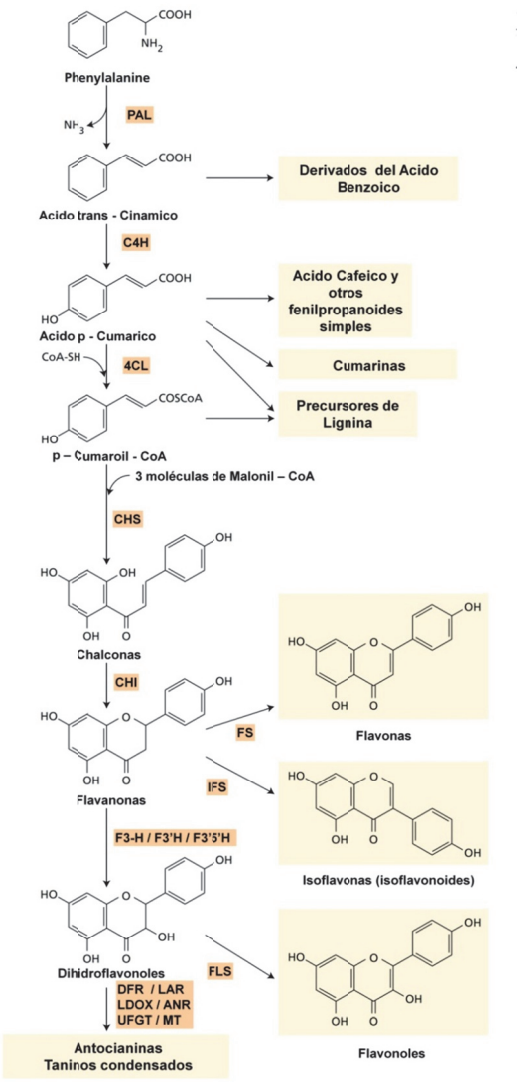


Figura 2: Esquema de la biosíntesis de fenoles a partir de fenilalanina. La formación de compuestos fenólicos, incluyendo fenilpropanoides simples, cumarinas, derivados del ácido benzoico, lignina, antocianinas, isoflavonas, taninos condensados y otros flavonoides, comienza con la fenilalanina. Figura modificada a partir de Taiz & Zeiger 2010.

Enzimas: PAL: fenilalanina amonio-liasa, C4H: cinamato 4-hidroxilasa, 4CL: 4-cumaril CoA ligasa, CHS: chalcona sintetasa, CHI: chalcona isomerasa, FS: flavona sintetasa, IFS: isoflavona sintetasa, F3-H = flavanona 3-hidroxilasa, F3'H: flavonol 3'-hidrolasa, F3'5'H: flavonol 3',5'-hidrolasa, FLS: flavonol sintetasa, DFR: dihidroflavonol 4-reductasa, LAR: leucoantocianidin reductasa, LDOX: leucoantocianidin dioxigenasa, ANR: antocianidina reductasa, UFGT: UDP-glucosa flavonoide 3-O-glucosil trasnferasa, MT: metiltransferasa.

Características genéticas de Tannat

El rango de tamaños de los genomas vegetales es muy amplio. Existen genomas vegetales pequeños como el de *Arabidopsis thaliana* (120 Mb) y otros varios órdenes de magnitud mayores como por ejemplo *Triticum aestivum* (16.000 Mb)

La vid tiene 38 cromosomas (n=19) y los híbridos interespecíficos son fértiles (Olmo 1979). Comparado con otras especies de plantas, su genoma es relativamente pequeño, 487 Mb (Jaillon *et al.* 2007), similar al Arroz (*Oryza sativa*, 430 Mb; Goff et al. 2002), el Carretón (*Medicago truncatula*, 500 Mb, http://medicago.org/) y el Álamo negro (*Populus trichocarpa*, 465 Mb; Tuskan *et al.* 2006).

Paradójicamente, muy pocos estudios han diferenciado genéticamente los clones dentro de una variedad de uva y, junto con Pinot Noir, la variedad Tannat es una de las excepciones. Muchos loci de microsatélites analizados amplificaron un solo alelo, lo que implica que el Tannat es una variedad muy homocigota. Teniendo en cuenta un conjunto de 15 microsatélites el nivel de homocigosis es del 53% en Tannat, en contraste con el 6% en Pinot Noir, 20% en Cabernet Franc y Chardonnay, y 33% en Cabernet Sauvignon (Gonzalez-Techera *et al.* 2004).

González Techera et al. (2004) demostraron que los clones de Tannat son genéticamente muy uniformes y que las diferencias ampelográficas asignadas a los diferentes clones (ENTAV 1995) son probablemente debido a diferencias epigenéticas. Sólo uno (VMCNg 1d12) de los 89 microsatélites utilizados en este estudio, pudo discriminar dos grupos de clones. En cada grupo se encontraron tanto clones antiguos uruguayos como clones comerciales franceses actuales, lo que sugiere que las fuentes de origen fueron las mismas para ambos grupos de clones y que podría haber habido dos diferentes fuentes de origen y propagación en la historia de estos clones de Tannat.

Históricamente el Tannat es la variedad dominante y casi exclusiva plantada en la región de Madiran de Francia (Viala & Vermorel 1903). El aislamiento geográfico puede haber promovido los eventos naturales de auto-fertilización que podrían explicar la alta frecuencia de loci homocigotas encontrados para esta variedad. Esto muestra que al contrario de otras variedades como Pinot Noir, donde la viabilidad de las progenies auto-fecundadas disminuye después de unas pocas generaciones (Bronner & Oliveira 1991), un proceso de endogamia no resulta perjudicial para Tannat.

Por otro lado, el análisis estadístico de las concentraciones de compuestos aromáticos de microvinificaciones también resultó en los mismos dos grupos de clones de

Tannat (González Techera *et al.* 2004). Curiosamente, este fue el primer informe de una correlación entre los compuestos relacionados con el aroma y marcadores moleculares dentro de clones de cultivares de *V. vinífera.*

La tecnología de secuenciación masiva para el estudio de Tannat.

La importancia económica de la vid ha alentado a muchos investigadores a estudiar su genoma y las bases fisiológicas y moleculares del desarrollo de las bayas, especialmente aquellas vías metabólicas relacionadas con procesos que afectan a la calidad del vino (Conde *et al.* 2007). La disponibilidad de métodos de análisis de alto rendimiento y la secuencia del genoma de *Vitis vinífera* cv. Pinot Noir (Jaillon *et al.* 2007) permiten caracterizar el genoma de otras variedades de vid y el transcriptoma durante el desarrollo de las bayas (Terrier *et al.* 2005, Waters *et al.* 2005, Deluc *et al.* 2007, Pilati *et al.* 2007, Zamboni *et al.* 2008, Rizzini *et al.* 2009, Zenoni *et al*. 2010, Sweetman *et al*. 2012, Venturini *et al.* 2013, Dal Santo et al. 2013). Pero pocos estudios han abordado los patrones de expresión génica de los diferentes tejidos de la baya (e.g. Grimplet *et al*. 2007, Ali *et al*. 2011, Gouthu *et al*. 2014).

Hay varios métodos para estudiar el transcriptoma de diferentes tejidos de la baya durante el desarrollo de ésta. Los análisis de microarrays son sensibles y precisos, pero limitados por la hibridación de fondo y el rendimiento de las diferentes sondas. El AffymetrixGeneChip de *Vitis* es muy eficaz, pero sólo contiene aproximadamente 14.500 genes únicos y la publicación del genoma de *V. vinífera* indica que hay al menos 30.434 genes que codifican para proteínas (Jaillon *et al*. 2007). El enfoque de RNA-Seq (Mortazavi *et al*. 2008) es el más acertado, y consiste en producir millones de lecturas cortas de cDNA que se asignan a un genoma de referencia para obtener un mapa del genoma de la transcripción, que consiste en la estructura de la transcripción y el nivel de expresión de cada gen. La visión holística del transcriptoma y su organización prevista por el método de RNA-Seq también revela muchas nuevas regiones transcritas, splicing alternativo, polimorfismos de nucleótido único (SNP), y la ubicación exacta de los límites de la transcripción (Cloonan y Grimmond 2008, Li *et al*. 2008, Lister *et al*. 2008, Mortazavi *et al*. 2008, Nagalakshmi *et al*. 2008, Sultan *et al*. 2008, Wilhelm *et al*. 2008). Por último, RNA-Seq genera mediciones absolutas, no relativas, de la expresión de genes, proporcionando un mayor conocimiento y precisión que los microarrays (Hoen *et al*. 2008, Marioni *et al*. 2008, Mortazavi *et al*. 2008).

Contrariamente a la tendencia general en los cultivos anuales, la mayoría de las mejoras en la vid y la producción de vino se basan en la evolución tecnológica en las prácticas agronómicas (viticultura) y enológica. Relativamente poco se ha hecho a nivel molecular para explorar la composición genética de la vid en comparación con otros cultivos de importancia económica (Peng *et al*. 2007). En la actualidad, el desarrollo de las tecnologías de secuenciación masiva están teniendo un gran impacto en la investigación genómica, proporcionando nuevas estrategias para analizar la complejidad funcional de los genomas y transcriptomas. Gracias a estas novedosas técnicas de alto rendimiento se dispone de un genoma de V. vinífera cv. Pinot Noir como genoma de referencia (Jaillon *et al*. 2007), lo cual es un importante punto de partida para el desarrollo de herramientas potentes para la manipulación genética de este cultivo. Los estudios con enfoques basados en genómica son una gran promesa para el mejoramiento molecular de variedades de uva de calidad con rasgos nuevos o mejorados.

La característica principal de Tannat es el alto nivel de polifenoles, lo cual le confiere a sus vinos un intenso color púrpura y mayores propiedades antioxidantes que otras uvas tintas. Estas características han promovido las plantaciones de Tannat en muchas regiones del Nuevo Mundo para la mezcla con otros vinos tintos como forma de añadir estructura y mejorar el potencial de envejecimiento de los vinos tintos de calidad. Pero muy poco se sabe de *V. vinífera* cv. Tannat, y menos aún de sus características genéticas. Con el fin de entender el porqué de estas características únicas se secuenció y analizó el genoma de Tannat, y se realizaron transcriptomas de cáscaras y semillas por separado durante el desarrollo de la baya. Estos estudios tuvieron énfasis en la biosíntesis de polifenoles, proporcionando así información útil y actualizada para la comprensión de esta variedad por parte de nuestros productores de uva y vino.

OBJETIVOS

OBJETIVO GENERAL

Aportar al conocimiento de *V. vinífera* cv. Tannat mediante el análisis del genoma y transcriptomas de cáscaras y semillas a lo largo del desarrollo de la baya.

OBJETIVOS ESPECÍFICOS

- Determinar variaciones genéticas involucradas en las características principales de las bayas de *Vitis vinifera* cv. Tannat.
- Determinar patrones de expresión génica durante el desarrollo de uvas relacionados con la biosíntesis de polifenoles en cáscaras y semillas.
- Proponer genes involucrados en la biosíntesis de polifenoles característicos de Tannat.
- Validar perfiles de expresión de genes claves en la biosíntesis de polifenoles en cáscaras y semillas de Tannat.

BIBLIOGRAFÍA

Alcalde-Eon C, Boido E, Carrau FM, Dellacassa E & Rivas-Gonzalo JC. 2006. Different formation of anthocyanin-derived pigments in Tannat, Caladoc, Marselan, Marzemino and Chaveñasco monovarietal wines produced in Uruguay. Am. J. Enol. Vitic. 57: 449-459.

Ali MB, Howard S, Chen S, Wang Y, Yu O, Kovacs LG & Qiu W. 2011. Berry skin development in Norton grape: distinct patterns of transcriptional regulation and flavonoid biosynthesis. BMC plant biology, 11(1):1.

Auer PL & Doerge RW. 2010. Statistical Design and Analysis of RNA Sequencing Data. Genetics 185: 405–416.

Auger C, Al-Awwadi N, Bornet A, Rouanet JM, Gasc F, Cros G & Teissedre PL. 2004. Catechins and procyanidins in Mediterranean diets. Food Res. Int. 37:233–245.

Bargel H & Neinhuis C. 2005. Tomato (Lycopersicon esculentum Mill.) fruit growth and ripening as related to the biomechanical properties of fruit skin and isolated cuticle. J Exp. Bot. 56:1049-1060.

Baur JA & Sinclair DA. 2006. Therapeutic potential of resveratrol: the in vivo evidence. Nat. Rev. Drug Discov. 5:493-506.

Boido E, Lloret A, Medina K, Fariña L, Carrau FM, Versini G, & Dellacassa E. 2003. Aroma composition of Vitis vinifera cv. Tannat: the typical red wine from Uruguay. J. Agric. Food Chem., 51:5408-5413.

Boido E, Alcalde-Eon C, Carrau FM, Dellacassa E & Rivas-Gonzalo JC. 2006. The aging effect on the pigment composition and color of Vitis vinifera l. cv Tannat wines. contribution of the main pigment families to wine color. J. Agric. Food Chemistry 54:6692-6704

Boido E, García-Marino M, Dellacassa E. Carrau F, Rivas-Gonzalo JC & MT Escribano-Bailón. 2011. Characterisation and evolution of grape polyphenol profiles of Vitis vinifera L. cv. Tannat during ripening and vinification. Aust. Journal Grape Wine Res. 17:383-393.

Bradamante S, Barenghi L & Villa A. 2004. Cardiovascular protective effects of resveratrol. Cardiovasc Drug Rev. 22:169-188.

Broner A & Olivera J. 1991. Création et étude de lignées chez le pinot noir (Vitis vinifera L.). J. Int. des Sci. Vigne Vin, 3, 133-148.

Bruno R, Ghisolfi L, Priulla M, Nicolin A & Bertelli A. 2003. Wine and tumors: study of resveratrol. Drugs Exp Clin Res, 29:257-261.

Burns KC & Dalen JL. 2002. Foliage color contrasts and adaptive fruit color variation in a bird-dispersed plant community. Oikos, 96:463-469.

Burns J, Gardner PT, O'Neil J, Crawford S & Morecroft I. 2000. Relationship among Antioxidant Activity, Vasodilation Capacity, and Phenolic Content of Red Wines. J Agric Food Chem, 48:220–230.

Câmara JS, Herbert P, Marques JC & Alves MA. 2004. Varietal flavor compounds of four grape varieties producing Madeira wines. Anal. Chim. Acta 2004, 513:203-207.

Carrau FM. 1997. The emergence of a new uruguayan wine industry. Journal of Wine Research, UK , 8:179-185.

Carrau FM, Medina K, Boido E, Fariña L, Gaggero C, Dellacassa E, Versini G, Henschke PA. 2005. de novo synthesis of monoterpenes by Saccharomyces cerevisiae wine yeasts. FEMS Microbiology Letters, 243:107-115.

Carrau FM, Dellacassa E & Boido E. 2008. Terpenoids in grapes and wines: origin and micrometabolism during the vinification process. Natural Product Communications, 3:577-592.

Carrau F, Boido E, Gaggero C, Medina K, Fariña L, Disegna E &Dellacassa E. 2011. Vitis vinifera Tannat, chemical characterization and functional properties. Ten years of research . In: Editor R. Filip (org.). Multidisciplinary approaches on food science and nutrition for the XXI century. Kerala, India, Research Signpost, Chapter. 4: pp 53-71.

Carro N, López E, Günata ZY, Baumes RL & Bayonove CL. 1996. Free and glycosidically bound aroma compounds in grape must of four non-floral Vitis vinifera varieties. Analusis, 24:254-258.

Cheynier V, Prieur C, Guyot S, Rigaud J & Moutounet M. 1997. The structures of tannins in grapes and wines and their interactions with proteins. In T. R. Watkins (Ed.), Proceedings of ACS symposium series 661, Wine: Nutritional and therapeutic benefits, 81.

Cloonan N & Grimmond SM. 2008. Transcriptome content and dynamics at single-nucleotide resolution. Genome Biol, 9:234.

Conde C, Silva P, Fontes N, Dias ACP, Tavares RM, Sousa MJ, Agasse A, Delrot S & Gerós H. 2007. Biochemical Changes throughout Grape Berry Development and Fruit and Wine Quality. Food, 1:1-22.

Coombe BG. 1992. Research on development and ripening of the grape berry. Am J Enol Vitic, 43:101-110.

Coombe BG & McCarthy MG. 2000. Dynamics of Grape Berry Growth and Physiology of Ripening. Aust J Grape Wine Res, 6:131-135.

Corder R, Mullen W, Khan NQ, Marks SC, Wood EG, Carrier MJ & Crozier A. 2006. Oenology: red wine procyanidins and vascular health. Nature, 444:566.

Dal Santo, S., Tornielli, G. B., Zenoni, S., Fasoli, M., Farina, L., Anesi, A., ... & Pezzotti, M. (2013). The plasticity of the grapevine berry transcriptome. Genome biology, 14(6), 1.

Dell'agli M, Buscialà A & Bosisio E. 2004. Vascular effects of wine polyphenols. Cardiovasc. Res. 63:593-602.

Deluc L, Bogs J, Walker AR, Ferrier T, Decendit A, Merillon JM, Robinson SP, Barrieu F. 2008. The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. Plant Physiol, 147:2041–2053.

Dore S. 2005. Unique properties of polyphenol stilbenes in the brain: more than direct antioxidant actions; gene/protein regulatory activity. Neurosignals, 14:61-70.

Dueñas M, Cueva C, Muñoz-González I, Jiménez-Girón A, Sánchez-Patán F, Santos-Buelga C, Moreno-Arribas MV, Bartolomé B. 2015. Studies on Modulation of Gut Microbiota by Wine Polyphenols: From Isolated Cultures to Omic Approaches. Antioxidants (Basel) 4:1-21. Review.

Durquety PM & Houbart JP. 1982. Two Tannat sports: "meunier" and "bulle". Progr. Agric. Vitic., 99:83-87.

Echeverry C, Ferreira M, Reyes-Parada M, Abin-Carriquiry JA, Blasina F, Gonzalez-Neves G & Dajas F. 2005. Changes in antioxidant capacity of Tannat red wines during early maturation. J. Food Eng., 69:147-154.

Escudero A, Campo E, Fariña L, Cacho J & Ferreira V. 2007. Analytical characterization of the aroma of five premium red wines. Insights into the role of odor families and the concept of fruitiness of wines. J. Agric. Food Chem., 55:4501-4510.

Ewing B. & Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res., 8:186–194.

Ewing B, Hillier L, Wendl MC & Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res., 8:175–185.

Fernandez-Gonzalez M & Di Stefano R. 2004. Fractionation of glycoside aroma precursors in neutral grapes hydrolysis and conversion by Saccharomyces cerevisiae. Lebensm-Wiss u Technol., 37:467-473.

Francis IL, Sefton MA & Williams PJ. 1992. Sensory descriptive analysis of hydrolysed precursor fractions from Semillion, Chardonnay, and Sauvignon Blanc grape juices. J. Sci. Food Agric., 59:511-520.

Galanti AN. 1919. La Industria Vitivinícola Uruguaya, Hemisferio Sur, Mendoza.

Gámbaro A, Boido E, Zlotejablko A, Medina K, Lloret A, Dellacassa E & Carrau FM. 2001. Effect of malolactic fermentation on the aroma properties of Tannat wine. Australian Journal of Grape and Wine Research, 7:27-32.

German JB & RL Walzem. 2000. The health benefits of wine. Annu Rev Nutr, 20:561-593.

Goff SA, Ricke D, Lan TH, Presting G, Wang R et al. 2002. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science, 296:92–100.

González Techera A, Jubany S, Ponce de León I, Boido E, Dellacassa E, Carrau FM, Hinrichsen P& Gaggero C. 2004 .Genetic diversity within clones of Vitis vinifera cv, Tannat. Vitis, 43:179-185.

Gouthu S, O'Neil ST, Di Y, Ansarolia M, Megraw M, Deluc LG. 2014. A comparative study of ripening among berries of the grape cluster reveals an altered transcriptional programme and enhanced ripening rate in delayed berries. J Exp Bot. 65:5889-5902.

Grimplet J, Deluc LG, Tillett RL, Wheatley MD, Schlauch KA, Cramer GR & Cushman jc. 2007. Tissue-specific mRNA expression profiling in grape berry tissues. BMC Genomics, 8:187.

Gu X, Creasy L, Kester A &Zeece m. 1999. Capillary electrophoretic determination of resveratrol in wines. J Agric Food Chem., 47:3223-3227.

Guasch-Jané MR, Andrés-Lacueva C, Jáuregui O & Lamuela-Raventós rm. 2006. First evidence of white wine in ancient Egypt from Tutankhamun's tomb. Journal of Archaeological Science 33:1075e1080

Gunata YZ, Bayonove CL, Tapiero C & Cordonnier RE. 1990. Hydrolysis of Grape Monoterpenyl-D-Glucosides by Various Glucosidases. J. Agric. Food Chem., 38:1232-1236.

Gunata Z, Bitteur S, Brillouet JM, Bayonove C & Cordonnier RE. 1988. Sequential enzymatic hydrolysis of potentially aromatic glysocides from grapes. Carbohydr. Res., 134:139-149.

Harris JM, Kriedmann PE & Possingham JV. 1968. Anatomical aspects of grape berry development. Vitis, 7:106–119

Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RH, de Menezes RX, Boer JM, van Ommen GJ, den Dunnen JT. 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res., 36: e141

Iriti M & Faoro F. 2009. Bioactivity of grape chemicals for human health. Natural product communications, 4:611-34.

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature, 449:463–467

Jay JM, editor. 1996. Modern Food Microbiology, Ed 5. Chapman & Hall, New York, pp 125–147

Jones SB, DePrimo SE, Whitfield ML & Brooks JD. 2005. Resveratrolinduced gene expression profiles in human prostate cancer cells. Cancer Epidemiol Biomarkers Prev, 14:596-604.

Kennedy DO. 2014. Polyphenols and the Human Brain: Plant "Secondary Metabolite" Ecologic Roles and Endogenous Signaling Functions Drive Benefits. Adv. Nutr. 5: 515–533.

Kim H-S, Quon MJ, Kim J-a. 2014. New insights into the mechanisms of polyphenols beyond antioxidant properties; lessons from the green tea polyphenol, epigallocatechin 3-gallate. Redox Biology 2:187–195.

Labarbe B, Cheynier V, Brossaud F, Souquet JM & Moutounet M. 1999. J. Agr. Food Chem., 47:2719-2723.

Li H, Ruan J & Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res, 18:1851–1858

Lizarraga D, Lozano C, Briedé JJ, van Delft JH, Touriño S, Centelles JJ, Torres JL, Cascante M. 2007. The importance of polymerization and galloylation for the antiproliferative properties of procyanidin-rich natural extracts. FEBS Journal 274, 4802–4811.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell, 133:523–536

Loscos N, Hernandez-Orte P, Cacho J & Ferreira V. 2007. J. Agric. Food Chem., 55:6674-6684.

Loulakakis KA, Roubelakis-Angelakis KA & Kanellis AK. 1996. Isolation of functional RNA from grapevine tissues poor in nucleic acid content. Am. J. Enol. Vitic., 47:181-185.

Luan F & Wüst M. 2002. Differential incorporation of 1-deoxy-D-xylulose into (3S)-linalool and geraniol in grape berry exocarp and mesocarp. Phytochemistry, 60:451-459.

Lücker J, Bowen P & Bohlmann J. 2004. Vitis vinifera terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (-)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries. Phytochemistry, 65:2649-2659.

Lucker J, Bouwmeester HJ, Schwab W, Blaas J, Van Der Plas LHW & Verhoeven HA. 2001. Expression of Clarkia S-linalool synthase in transgenic petunia plants results in the accumulation of S-linalyl-b-D-glycopyranoside.Plant J., 27:315-324.

Lund ST & Bohlmann J. 2006. The molecular basis for wine grape quality-a volatile subject. Science, 311:804-805.

Marioni JC, Mason CE, Mane SM, Stephens M & Gilad Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res, 18: 1509–1517

Martin DM & Bohlmann J. 2004. Identification of Vitis vinifera alphaterpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. Phytochemistry, 65:1223-1229.

Mateo JJ & Jiménez M. 2000. Monoterpenes in grape juice and wines. J. Chromatogr. A, 881:557-567.

Mateo JJ & Maicas SJ. 2005. Hydrolysis of terpenyl glycosides in grape juice and other fruit juices: a review. Appl. Microb. Biotech., 67:322-335.

Mathieu S, Terrier N, Procureur J, Bigey F & Günata Z. 2005. A carotenoid cleavage dioxygenase from Vitis vinifera L.: functional characterization and expression during grape berry development in relation to C13-norisoprenoid accumulation. J Exp Bot, 56:2721-2731.

McGovern PE. 2003. Ancient Wine: The Search for the Origins of Viniculture Princeton University Press, Princeton.

Mortazavi A, Williams BA, McCue K, Schaeffer L & Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods, 5:621–628

Moser C, Gatto P, Moser M, Pindo M & Velasco R. 2004. Isolation of Functional RNA From Small Amounts of Different Grape and Apple Tissues. Molecular Biotechnology, 26:95-99

Moutounet M, Rigaud J, Souquet JM & Cheynier V. 1996. Caracterisation structurale des tanins de las baie de raisin. Quelques exemples de l'incidence du cepage, du terroir et du monde de conduite de la vigne (1). Bull.OIV, 433-443.

Mullins MG, Bouquet A & Williams LE. 1992. Biology of the grapevine. Cambridge, Gran Bretaña, Cambridge University Press.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M & Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science, 320: 1344–1349.

Nitsch JP, Pratt C, Nitsch C & Shaulis NJ. 1960. Natural growth substances in Concord and Concord Seedless grapes in relation to berry development. Am J Bot, 47:566–576.

Olmo HP. 1979. Grapes. In: Simmonds NW, ed. Evolution of crop plants. London

Panagiotakos DB, Pitsavos C, Polychronopoulos E, Chrysohoou C, Zampelas A, et al. 2004. Can a Mediterranean diet moderate the development and clinical progression of coronary heart disease? A systematic review. Med Sci Monit, 10:193–198.

Peng FY, Reid KE, Liao N, Schlosser J, Lijavetzky D, Holt R, Martínez Zapater JM, Jones S, Marra M, Bohlmann J & Lund ST. 2007. Generation of ESTs in Vitis vinifera wine grape (Cabernet Sauvignon) and table grape (Muscat Hamburg) and discovery of new candidate genes with potential roles in berry development. Gene, 402:40–50

Petka J, Ferreira V, Gonzalez-Vinas MA & Cacho J. 2006. Sensory and chemical characterization of the aroma of a white wine made with Devin grapes. J. Agric. Food Chem., 54:909-915.

Pilati S, Perazzolli M, Malossini A, Cestaro A, Dematte L, Fontana P, Dal Ri A, Viola R, Velasco R, Moser C. 2007. Genome-wide transcriptional analysis of grapevine berry ripening reveals a set of genes similarly modulated during three seasons and the occurrence of an oxidative burst at veraison. BMC Genomics, 8:428-450.

Pratt C. 1971. Reproductive anatomy in cultivated grapes – a review. Am J Enol Vitic, 22:92–109

Prieur C, Rigaud J, Cheynier V & Moutounet M. 1994. Oligomeric and polimeric procyanidins from grape seeds. Phytochemistry, 36:781-784.

Puyares V, Ares G & Carrau FM. 2010. Studying the influence of bottle shape and colour on consumer expectations and willingness to purchase Tannat wine. Submitted to Food Quality and Preference.

Rabosto X, Carrau M, Paz A, Boido E, Dellacassa E & Carrau FM. 2006. Grapes and vineyard soils as sources of microorganisms for biological control of Botrytis cinerea. Am. J. Enol. Vitic., 53:332-338.

Rapp A. 1998. Volatile flavour of wine: Correlation between instrumental analysis and sensory perception. Nahrung, 42:351-363.

Reid KE, Olsson N, Schlosser J, Peng F & Lund ST. 2006. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol., 6:27-28.

Ricardo da Silva JM, Bourzeix M, Cheynier V & Moutounet M. 1991. Procyanidin composition of Chardonay, Mauzac and Grenache blanc grapes. Vitis, 30:245-252.

Rizzini FM, Bonghi C & Tonutti P. 2009. Postharvest water loss induced marked changes in transcript profiling in skin of wine grape berry. Postharvest Biol Technol, 52: 247–253.

Robichaud JL & Noble AC. 1990. Astringency and bitterness of selected phenolics In wine. Journal of The Science of Food And Agriculture, 53:343-353.

Robinson S. 2007. Flavonoid pathway genes in grapes. In: Final report to Grape and Wine Research & Development Corporatio. Project Number: 99/16A, CRVC Project Code 3.3.1. Cooperative Research Centre for Viticulture. CSIRO Plant Industry. Australia.

Rosillo L, Salinas MR, Garijo J & Alonso GL. 1999. A study of volatiles in grapes by dymanic headspace analysis application to the differentiation of some Vitis vinifera varieties. J. Chromatogr. A, 847:155-159.

Salzman RA, Fujita T, Zhu-Salzman K, Hasegawa PM & Bressan RA. 1999. An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. Plant Mol. Biol. Rep., 17:11-17.

Sefton MA. 1998. Hydrolytically-Released Volatile Secondary Metabolites from a Juice Sample of Vitis vinifera Grape cvs Merlot and Cabernet Sauvignon. Australian Journal of Grape and Wine Research. 4:30-38.

Sefton MA, Skouroumounis GK, Massy-Westropp RA & Williams PJ. 1989. Norisoprenoids in Vitis vinifera white wine grapes and the identification of a precursor of damascenone in these fruits. Aus. J. Chem., 42:2071–2084.

Sefton MA, Francis IL & Williams PJ. 1994. Free and bound volatile secondary metabolites of Vitis vinifera grapes cv Sauvignon Blanc. J Food Sci, 59:142-147.

Slaughter JC. 1999. The naturally occurring furanones: formation and function from pheromone to food. Biol Rev Camb Philos Soc, 74:259-276.

Solovchenko A & Schmitz-Eiberger M. 2003. Significance of skin flavonoids for UV-B-protection in apple fruits. J Exp Bot, 54:1977-1984.

Souquet J, Cheynier V, Broussaud F & Moutounet M. 1996. Polimeric proanthocyanidins from grape skins. Phytochem., 43:509-512.

St Leger AS, Cochrane AL & Moore F. 1979. Factors associated with cardiac mortality in developed countries with particular reference to the consumption of wine. Lancet, 1:1017-20.

Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A Parkhomchuk D, et al. 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science, 321:956–960

Sweetman C, Wong DC, Ford CM, & Drew DP. 2012. Transcriptome analysis at four developmental stages of grape berry (Vitis vinifera cv. Shiraz) provides insights into regulated and coordinated gene expression. BMC genomics, 13(1), 1.

Taiz L & Zeiger E. 2010. Plant Physiology, Fifth Edition. Sinauer Associates. Sunderland, MA. 623 paginas.

Terrier N, Glissant D, Grimplet J, Barrieu F, Abbal P, Couture C, Ageorges A, Atanassova R, Léon C, Renaudin JP, Dédaldéchamp F, Romieu C, Delrot S & Hamdi S. 2005. Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (Vitis vinifera L.) development. Planta, 222:832-847.

This P, Lacombe T & Thomas MR. 2006. Historical origins and genetic diversity of wine grapes. Trends in Genetics, 22:511-519.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, et al. 2006. The Genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science, 313:1596–1604.

Ugliano M, Bartowsky EJ, McCarthy J, Moio L & Henschke PA. 2006. Hydrolysis and transformation of grape glycosidically bound volatile compounds during fermentation with three Saccharomyces yeast strains. J. Agr. Food Chem., 54:6322-6331.

Varela P & Gámbaro A. 2006. Sensory descriptive analysis of Uruguayan Tannat wine: Correlation to quality assessment. J.Sens.Stud., 21:203–217.

Venturini, L., Ferrarini, A., Zenoni, S., Tornielli, G. B., Fasoli, M., Dal Santo, S., ... & Zamperin, G. (2013). De novo transcriptome characterization of Vitis vinifera cv. Corvina unveils varietal diversity. BMC genomics, 14(1), 1.

Versini G, Carlin S, Nicolini G, Dellacassa E & Carrau FM. 1999. Updating of varietal aroma components in wines. Vitiviniculture in the Southern hemisphere: proceedings of the 7th Latin-American Congress of viticulture and winemaking: Mendoza (Argentina), 325-349.

Viala P & Vermorel V. 1903. Traité général de viticulture. Tomo IV - Ampèlographie, 1a ed.; Masson et Cie. Editeurs, Paris.

Wang J & Luca VD. 2005. The biosynthesis and regulation of biosynthesis of Concord grape fruit esters, including 'foxy' methylanthranilate. Plant J, 44:606-619.

Waters DLE, Holton TA, Ablett EM, Lee LS & Henry RJ. 2005. cDNA microarray analysis of developing grape (Vitis vinifera cv. Shiraz) berry skin. Funct Integr Genomics, 5:40-58.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J & Bahler J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature, 453:1239–1243

Williams PJ, Strauss CR & Wilson B. 1982 b. Use of C18 reversed-phase liquid chromatography for the isolation of monoterpene glycosides and nor-isoprenoid precursors from grape juice and wines J. Chromatogr., 235:471-480.

Williams PJ, Strauss CR, Wilson B & Massy-Westropp RA. 1982 a. Studies on the hydrolysis of Vitis vinifera monoterpene precursor compounds and model monoterpene .beta.-D glucosides rationalizing the monoterpene composition of grapes. J. Agric. Food Chem., 30:1219-1223.

Wintheralter P, Sefton MA & Williams PJ. 1990. Volatile C13-norisoprenoid compounds in Riesling wine are generated from multiple precursors. Am. J . Enol .Vitic., 41: 277-283.

Wollin SD & Jones PJ. 2001. Alcohol, red wine and cardiovascular disease. J Nutr 131:1401-1404.

Zamboni A, Minoia L, Ferrarini A, Tornielli GB, Zago E, Delledonne M & Pezzotti M. 2008. Molecular analysis of post-harvest withering in grape by AFLP transcriptional profiling. J Exp Bot, 59:4145–4159

Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, Malerba G, Bellin D, Pezzotti M & Delledonne M. 2010. Characterization of Transcriptional Complexity during Berry Development in Vitis vinifera Using RNA-Seq. Plant Physiology, 152:1787–1795,

Zohary D & Hopf M. 2000. Domestication of plants in the Old World. 3rd edition. New York Oxford University Press;316.

RESULTADOS

CAPÍTULO I

VOLUME 25 NUMBER 12 DECEMBER 2013

THE PLANT CELL



VARIETAL GENES AND POLYPHENOLS IN TANNAT GRAPE BERRIES

www.plantcell.org

J. Troncy

Imp. F. CHAMPENOIS, Paris.

VARIETAL GENES AND POLYPHENOLS IN TANNAT GRAPE BERRIES

The Tannat grape berry is used to produce high quality wines with an intense purple color and remarkable antioxidant properties. Through reference-guided assembly of the genome combined with de novo assembly of the transcriptome, Da Silva et al. (pages 4777–4788) found variety-specific genes not present in the Pinot Noir reference genome that might contribute substantially to the unique characteristics of the Tannat berry. The cover image by Jules Troncy is from Ampélographie – Traité général de viticulture. Tome IV. 1903 by Pierre Viala and Victor Vermorell. Imp. F. Champenois, Paris.

This article is a *Plant Cell* Advance Online Publication. The date of its first appearance online is the official date of publication. The article has been edited and the authors have corrected proofs, but minor changes could be made before the final version is published. Posting this version online reduces the time to publication by several weeks.

LARGE-SCALE BIOLOGY ARTICLE

The High Polyphenol Content of Grapevine Cultivar Tannat Berries Is Conferred Primarily by Genes That Are Not Shared with the Reference Genome[W]

Cecilia Da Silva,[a,1] Gianpiero Zamperin,[b,1] Alberto Ferrarini,[b] Andrea Minio,[b] Alessandra Dal Molin,[b] Luca Venturini,[b] Genny Buson,[b] Paola Tononi,[b] Carla Avanzato,[b] Elisa Zago,[b,2] Eduardo Boido,[c] Eduardo Dellacassa,[c] Carina Gaggero,[a] Mario Pezzotti,[b] Francisco Carrau,[c] and Massimo Delledonne[b,3]

[a] Departamento de Biología Molecular, Instituto de Investigaciones Biológicas Clemente Estable, 11600 Montevideo, Uruguay
[b] Centro di Genomica Funzionale, Dipartimento di Biotecnologie, Universitá degli Studi di Verona, 37134 Verona, Italy
[c] Sección Enología, Facultad de Química, Universidad de la República,11800 Montevideo, Uruguay

ORCID IDs: 0000-0002-7100-4581 (M.D.); 0000-0002-8452-3380 (A.F.); 0000-0003-0600-5163 (G.Z.).

The grapevine (*Vitis vinifera*) cultivar Tannat is cultivated mainly in Uruguay for the production of high-quality red wines. Tannat berries have unusually high levels of polyphenolic compounds, producing wines with an intense purple color and remarkable antioxidant properties. We investigated the genetic basis of these important characteristics by sequencing the genome of the Uruguayan Tannat clone UY11 using Illumina technology, followed by a mixture of de novo assembly and iterative mapping onto the PN40024 reference genome. RNA sequencing data for genome reannotation were processed using a combination of reference-guided annotation and de novo transcript assembly, allowing 5901 previously unannotated or unassembled genes to be defined and resulting in the discovery of 1873 genes that were not shared with PN40024. Expression analysis showed that these cultivar-specific genes contributed substantially (up to 81.24%) to the overall expression of enzymes involved in the synthesis of phenolic and polyphenolic compounds that contribute to the unique characteristics of the Tannat berries. The characterization of the Tannat genome therefore indicated that the grapevine reference genome lacks many genes that appear to be relevant for the varietal phenotype.

INTRODUCTION

Grapevine (*Vitis vinifera*) cv Tannat is a grapevine cultivar originally from southwestern France that is now mainly cultivated in Uruguay. 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 (Carrau, 1997).

Tannat is the main grape berry produced for Gers wines and is the *V. vinifera* cultivar richest in tannins (Boido et al., 2011). Tannat seeds contain high levels of tannins, and Tannat skins at maturity contain high levels of anthocyanins. The total flavan-3-ol content of Tannat seeds (1946 mg/kg; Boido et al., 2011) is 6 times higher than the content reported for Pinot noir (317 mg/kg; Mattivi et al., 2009) under similar experimental conditions. Tannat grapes therefore produce a colorful wine, with high acidity and a high tannin content that is suitable for long-term aging (Alcalde-Eon et al., 2006; Boido et al., 2006, 2011). Tannat berries have other unusual characteristics relating to the high content of phenolic and polyphenolic compounds (e.g., the presence of compounds such as malvidin, delphinidin, and petunidin monoglucosides), which confer more intense pigments to wine (Alcalde-Eon et al., 2006), and much higher levels of resveratrol compared with cultivars such as Pinot Noir, Merlot, and Cabernet (Gu et al., 1999; Carrau et al., 2011).

Polyphenols have many useful biological functions in plants (Kutchan, 2005), including defense against biotic stress (Bhattacharya et al., 2010), reflecting their antimicrobial, antifungal, and antiherbivore properties (Dixon et al., 2002; Chong et al., 2009). Polyphenols acquired from the moderate consumption of red wine also provide pharmaceutical and nutritional benefits to humans (Lin and Weng, 2006), such as helping to prevent cancer and reducing the inflammation associated with coronary artery disease (Khan et al., 2010). Proanthocyanidins are the principal vasoactive polyphenols in red wine; they induce the endothelium-dependent dilatation of blood vessels and suppress the synthesis of the vasoconstrictive peptide endothelin-1 (Fitzpatrick et al., 1993; Hashimoto et al., 2001). The proanthocyanidin content of local red wines correlates strongly with increased longevity in areas such as Gers in France and Nuoro in Italy (Corder et al., 2006).

[1] These authors contributed equally to this work.
[2] Current address: Personal Genomics s.r.l, Strada le Grazie 15, 37134, Verona, Italy.
[3] Address correspondence to massimo.delledonne@univr.it.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Massimo Delledonne (massimo.delledonne@univr.it).
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.113.118810

 © 2013 American Society of Plant Biologists. All rights reserved.

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 (Myles et al., 2010; Zhang et al., 2012). However, recent studies have shown that reliance on a single reference genome may underestimate the variability among different genotypes (Springer et al., 2009; Swanson-Wagner et al., 2010; Gan et al., 2011). Plant genomes contain core sequences that are common to all individuals, as well as dispensable sequences comprising partially shared and nonshared genes that contribute to intraspecific variation (Morgante et al., 2007). The dispensable genome is more difficult to characterize because one of the main drivers is transposon activity, which influences genome size (Ammiraju et al., 2007; Ågren and Wright, 2011) and the expression and regulation of genes (Kobayashi et al., 2004). However, it is now suspected that such variability is not restricted to the number and location of transposable elements but may also involve functional sequences as previously shown in prokaryotes (Medini et al., 2005). Similarly in humans, Asian and African individuals possess ~5 Mb of population-specific DNA encoding 141 human RefSeq sequences that cannot be mapped against the current reference genome, suggesting that the human pangenome would include an additional 19 to 40 Mb of novel information (Li et al., 2010). In *Arabidopsis thaliana*, 221 genes that are not present in the reference Columbia-0 genome have been identified in other ecotypes (Gan et al., 2011; Schneeberger et al., 2011). More recently, the presence in specific cultivars of hundreds of genes not 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 (*Zea mays*; Hansey et al., 2012) and grapevine (Venturini et al., 2013).

We investigated the genetic basis of the unique phenotypic characteristics of Tannat berries, which are most notable for their high content of polyphenolic compounds, by sequencing the Uruguayan Tannat clone UY11 from the vines introduced into Uruguay in the 1870s (Gonzalez Techera et al., 2004). UY11 belongs to a predominantly homozygous group of clones including the French commercial clone 399, one of the most widely grown clones in Uruguay (Gonzalez Techera et al., 2004). We also produced an exhaustive annotation of Tannat genes by RNA-Seq analysis and identified many genes that are not shared with the reference genome and that are involved in the synthesis of phenolic and polyphenolic compounds.

RESULTS

Genome Sequencing, Assembly, and Analysis of Variants

Genomic DNA from the Uruguayan Tannat clone UY11 was used to generate 322,786,617 Illumina reads (2 × 100) representing 134-fold base pair coverage and ~260-fold physical coverage of the grapevine genome (see Supplemental Table 1 online). The genome was assembled using a hybrid approach based on iterative realignment to the reference genome and integration of de novo–assembled contigs with a reference genome (Gan et al., 2011) with the Pinot Noir PN40024 genome sequence as a reference (Jaillon et al., 2007). The reconstruction of 482 Mb of the Tannat genome (see Supplemental Table 1 online) required 11 cycles of iterative read mapping combined with de novo assembly (see Supplemental Figure 1 online). The reads were aligned with the final assemblies, identifying 8.6 Mb of polymorphic regions lacking read coverage that probably reflect complex polymorphisms as previously reported (Gan et al., 2011). The N50 length (the contig length at which 50% of the entire assembly is represented by contigs equal to or longer than this value) of contiguous read coverage between polymorphic regions was 97.2 kb.

To report complex alleles consistently, we defined all variants against the 12x grapevine reference genome assembly (Grimplet et al., 2012). There were 2,087,275 single-base differences between Tannat and PN40024, as well as 240,355 (11.5%) ambiguous calls. When the same procedure was applied to the Corvina cultivar, we found 426,415 ambiguous calls among 2,048,530 single-base differences (~20.8%), which is twice the number of heterozygous single-nucleotide polymorphisms (SNPs) found in the Tannat cultivar despite the similar total number of SNPs. These results concur with microsatellite data suggesting that Tannat has a lower degree of heterozygosity than other grapevine cultivars (Gonzalez Techera et al., 2004).

As expected, most of the identified SNPs were located outside known genes or within introns, with few located in coding sequences (CDSs) or untranslated regions (Figure 1A). The frequency

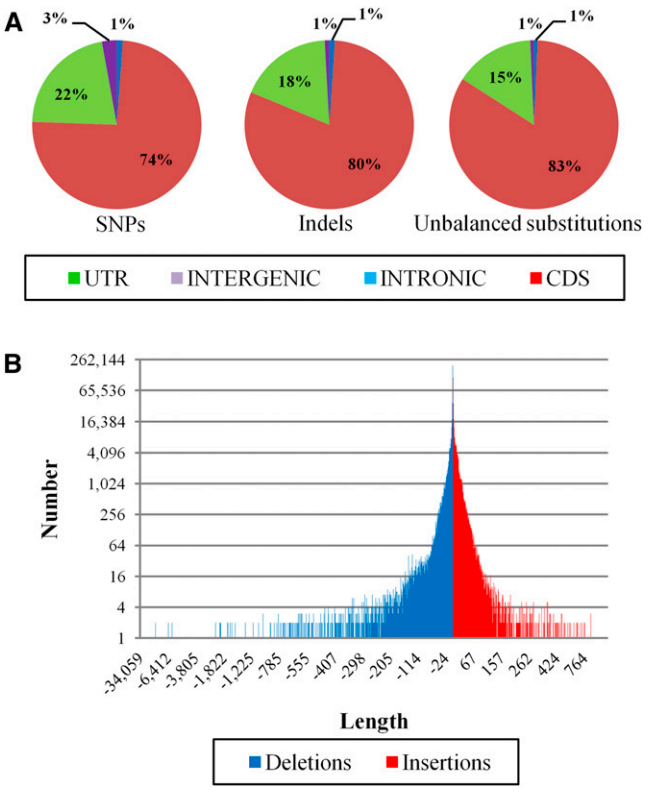


Figure 1. Variation in the Tannat Genome.

(A) Classification of SNPs, indels, and unbalanced substitutions based on the *V. vinifera* PN40024 genome and annotation. UTR, untranslated region.

(B) Distribution of lengths of deletions and insertion.

of SNPs within genes (0.2572%), CDSs (0.2408%), and untranslated regions (0.3336%) was lower than the whole genome frequency (0.4292%). The low frequency of SNPs in CDSs was attributed to the conservation of protein functions. We identified 20,843 genes containing SNPs, and these were enriched for major functional categories such as protein Tyr kinase activity, regulation of development, protein binding, cellular component organization or biogenesis, and transferase activity (see Supplemental Table 2 online). We also identified 644,772 insertion/deletion polymorphisms (indels) and 42,471 unbalanced substitutions representing the replacement of the PN40024 reference sequence with a different sequence. Most of these variants were located in intergenic regions, and the portion that mapped within known CDSs was lower than for SNPs (Figures 1A). This concurs with the greater potential impact of these types of variants on protein function. Although 57.5% of indels and unbalanced substitutions were shorter than 6 bp in length, 1.9% exceeded 100 bp (Figure 1B). 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 (Gan et al., 2011). 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.

Gene Annotation

A naive projection of the coordinates of the 29,971 nuclear protein-coding genes from PN40024 (V1 annotation; http://plants.ensembl.org/Vitis_vinifera/Info/Index) onto the Tannat genome predicted that 46.78% of the proteins contained sequence differences. Most of the changes were predicted not to affect protein functionality, and, 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 genome annotation. The remaining 6988 genes (23.3%) were predicted to encode proteins affected by deletions, truncations, or other disruptions (see Supplemental Figure 2 online). For these genes, it was not possible to transfer the reference annotation in a reliable manner.

To reannotate the Tannat genome, we performed an RNA-Seq analysis using a panel of four tissues/developmental stages (whole berry, skin at two developmental stages, and seeds), thus generating 395,863,776 2 $\times$ 100 reads (see Supplemental Table 3 online). We focused on berry tissues at early developmental stages because Tannat berries are rich in tannins, which are produced mostly prior to veraison (when berries begin to change color and enlarge) (Downey et al., 2003a; Conde et al., 2007). The expression data provided experimental support for 16,169 of the 22,983 genes reliably transferred from PN40024 to Tannat (see Supplemental Data Set 1 online). Reference-based assembly of the RNA-Seq data (Trapnell et al., 2012) allowed us to identify 81,759 putative transcripts and to reannotate 5796 genes corresponding to loci transferred from the V1 annotation that appeared to be disrupted in the Tannat genome, as well as 2866 nonannotated protein-coding genes with homology to sequences in the National Center for Biotechnology Information (NCBI) nonredundant (NR) and/or *Vitis vinifera* Gene Index (VvGI) databases. We ultimately characterized 31,645 loci including 13,932 genes (44%) encoding multiple protein isoforms (file available in General Transfer Format at http://ddlab.sci.univr.it/files/Tannat/annotation.gtf). The features of newly annotated (or reannotated) genes, such as the average mRNA length (1662 nucleotides) and the average number of exons per gene (6.77), were similar to those of PN40024 and other plant species (see Supplemental Table 4 online).

The gene annotations were validated further using the nonredundant core eukaryotic genes (CEGs) from the CEG mapping approach (CEGMA) pipeline (Parra et al., 2009). The presence of 232 of 248 CEGs (93.5%) in the Tannat gene set confirmed the quality of the annotation (see Supplemental Figure 3 online).

Transcriptome Assembly and Identification of Tannat-Specific Genes

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 (see Supplemental Table 5 online) with an average size of 1491 nucleotides, which is similar to the corresponding value for the PN40024 annotation (1331 nucleotides) and the Tannat annotation (1554 nucleotides).

We validated the assembled transcriptome using the NR core gene set (Parra et al., 2009). The high percentage of CEGs represented in the data set (95%) confirmed that the Tannat gene set based on the assembled transcriptome was almost complete. Furthermore, the highest fraction of CEGs detected compared with the Tannat annotation (93.5%) suggested that the assembled transcriptome contained more genes than the annotated Tannat genome. Among the 114,786 transcripts, a large fraction (88%) mapped with high confidence to the reconstructed Tannat genome. BLAST queries of the 13,707 transcripts that could not be mapped to the reconstructed genome against the NCBI NR protein database (BLAST Expect value [E-value] $\leq 1*10^{-5}$; http://www.ncbi.nlm.nih.gov/BLAST/) identified 731 sequences as contaminants (see Supplemental Table 6 online). We found that 9302 of the remaining 12,976 putative transcripts matched expressed grapevine sequences represented in the VvGI database v8.0 or other plant proteins, and these were considered as novel grapevine transcripts potentially restricted to *V. vinifera* cv Tannat. Out of the 9302 putative transcripts, 5052 had a high coding potential and a full open reading frame with defined start and stop codons (Kong et al., 2007).

Among the 5052 high-confidence putative protein-coding transcripts that could not be mapped against the Tannat genome, 4501 were validated by comparison with raw Tannat genomic reads. After clustering and manual inspection, these transcripts were grouped into 3035 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 1162 genes and that Tannat possesses a set of 1873 genes that are not shared with PN40024 (Table 1). Therefore, the Tannat genome appears to comprise 28,779 genes that are annotated on the

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

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

reference genome (referred to as known genes), 4028 genes previously unannotated or not assembled in the reference genome (referred to as novel genes), and 1873 genes that appear to be unique to Tannat (varietal genes).

All 34,680 genes mapping to the reconstructed genome or obtained from de novo transcriptome assembly were functionally annotated using the NCBI NR protein database, VvGI 8.0, Gene Ontology, and the Kyoto Encyclopedia of Genes and Genomes (see Supplemental Data Set 1 online). For genes transferred from the V1 annotation, we integrated the VitisNet annotations (Grimplet et al., 2012). Definitions could be assigned by homology searches to 20,170 of the genes (~58%). We also assigned Gene Ontology terms to 21,442 genes (61.8%), and 8543 genes encoding enzymes were classified by Enzyme Commission number. The gene categories are summarized in Supplemental Figure 4 online, revealing a higher proportion (>50%) of genes related to metabolic and cellular processes as well as binding and catalytic molecular functions.

Expansion of Gene Families Related to Polyphenol Biosynthesis

The Tannat cultivar is characterized by the accumulation of high levels of polyphenols in the berry skin and seed. We developed a hypothesis that the unusual metabolic profile of Tannat berries may reflect the amplification of genes encoding enzymes representing the phenylpropanoid pathway and its derivatives. Therefore, we checked the Tannat gene set for the presence of new family members in pathways related to phenol and polyphenol biosynthesis.

This approach identified 148 novel genes and 141 varietal genes corresponding to 23 different enzymes (see Supplemental Data Set 2 online). 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 (Figure 2; see Supplemental Data Set 2 online). 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 (Figure 2; see Supplemental Data Set 2 online). 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 (Figure 2; see Supplemental Data Set 2 online).

Expression of Genes Related to Polyphenol Biosynthesis

We investigated the expression of each of the 34,680 genes in the whole berry at 1 week post flowering (wpf) and in skins and seeds at 7 wpf (Figure 3A; see Supplemental Data Set 1 online). Most of the genes were expressed in at least one of the samples (29,982; fragments per kilobase of transcript per million fragments mapped [FPKM] > 0.01), whereas ~65% (22,764) were expressed in all samples. Approximately 35% of the genes (12,244) were differentially expressed (false discovery rate < 5%, absolute of the logarithmic fold change [|Log2FC|] > 1) between the two developmental stages (1 and 7 wpf) or between the two tissues (seeds and skin) at 7 wpf (Figures 3B to 3D). The large proportion of modulated genes matches the plethora of biochemical and physiological changes that occur in different tissues during fruit development (Giovannoni, 2001). A greater proportion of varietal genes were differentially expressed compared with the whole data set (54% versus 35%).

Most of the key genes representing the phenylpropanoid pathway and downstream pathways leading to flavonoids were strongly expressed, including CHS, F3H, and DFR (Figures 4A to 4F). The expression profiles were consistent with the biosynthesis of tannins in seeds during the first few weeks after fruit set (Bogs et al., 2005) and were highly correlated to the expression profile of *MYBPA1*, a gene encoding a transcription factor known to regulate berry tannin synthesis in grapevine (Bogs et al., 2007). A similar pattern was also observed for two annotated shikimate dehydrogenase genes and for two Ser carboxypeptidase-like genes, which are highly expressed at 1 wpf and strongly induced in seeds at 7 wpf compared with skin at the same time point. Genes encoding enzymes involved in the production of other flavonoids, such as FLS, which catalyzes the first committed step in flavonol biosynthesis, showed more diverse expression profiles. Some were expressed only in the berry skin and others only in the seeds (Figure 4G).

To evaluate the potential contribution of varietal genes to the extensive production and accumulation of tannins in Tannat, we determined the expression level of known, novel, and varietal genes relative to the overall expression level of each enzyme. Among the enzymes contributing to tannin biosynthesis, 14 were represented by the expression of varietal genes, with a contribution generally ranging from 11.16 to 81.24% (Figure 5A), but that in the case of 6'-deoxychalcone synthase reached 94% in the berry skin at 7 wpf (see Supplemental Data Set 2 online). The overall contribution to the expression of each enzyme was not strictly proportional to the number of gene copies among the known, novel, and varietal genes (Figure 5B), suggesting the significant differential regulation of genes encoding the same enzyme.

Regulation of Polyphenol Biosynthesis Pathway Genes

To investigate the regulation of polyphenol biosynthesis in Tannat berries, we analyzed four transcription factor superfamilies (basic

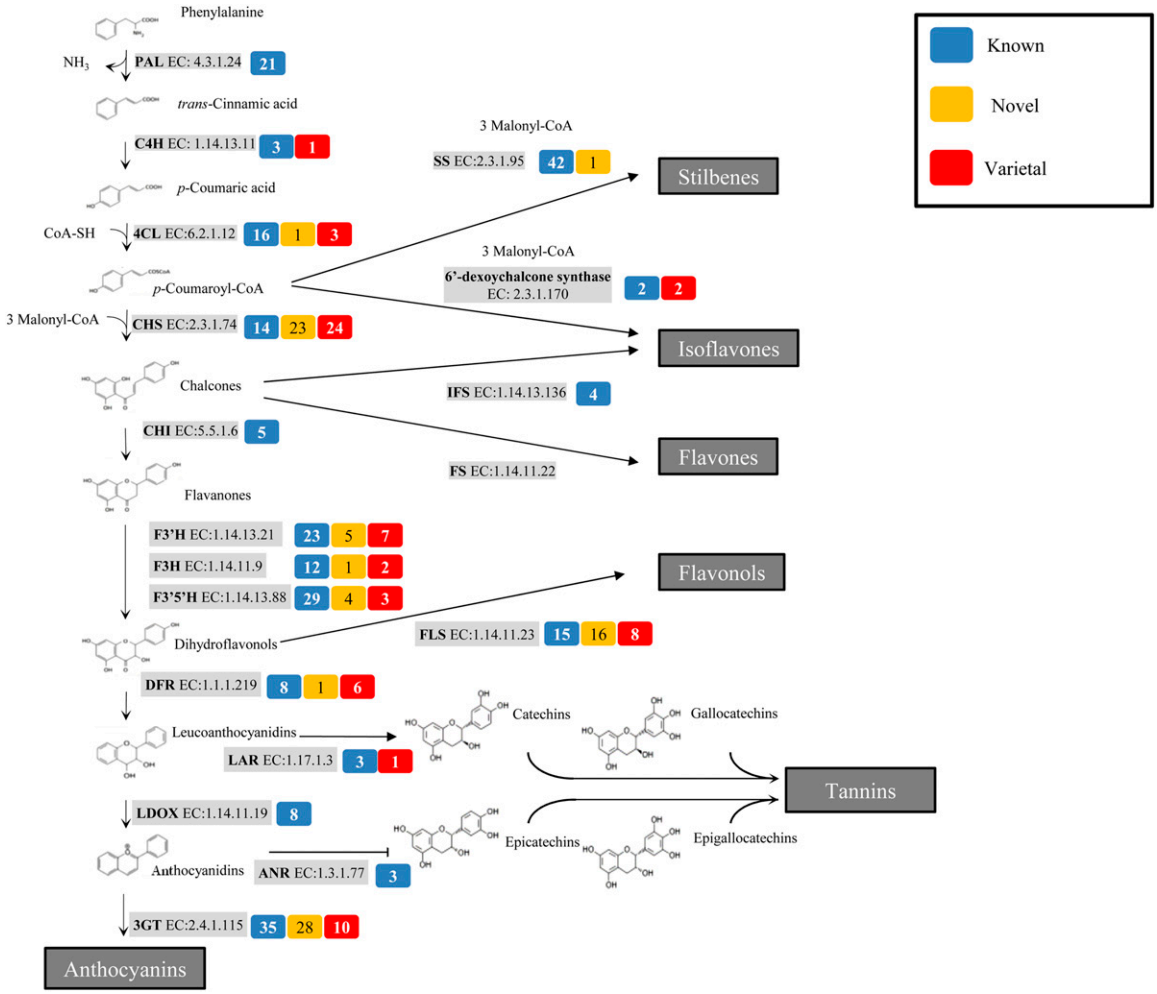


Figure 2. Schematic Diagram of Metabolic Pathways Involved in Polyphenol Biosynthesis.

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.

helix-loop-helix [bHLH], R2R3 MYB, WD40, and WRKY) known to be involved in the regulation of this pathway (Lepiniec et al., 2006). The V1 annotation includes 106 bHLH, 131 R2R3 MYB, 56 WD40, and 61 WRKY genes. By contrast, with the observed expansion of enzyme gene families, we identified only two varietal genes representing the bHLH and R2R3 MYB superfamilies and four varietal genes and three novel genes representing the WD40 superfamily (see Supplemental Data Set 2 online). The contribution to overall expression levels by new genes (i.e., not present in the V1 annotation) in these four transcription factor superfamilies was low (1.4, 2.3, and 7% for the R2R3 MYB, bHLH, and WD40 superfamilies, respectively).

Origin of the Varietal Genes

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 (Velasco et al., 2007) and with internally produced Illumina genomic reads from the Corvina cultivar (Figure 6). 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 (Cipriani et al., 2010).

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

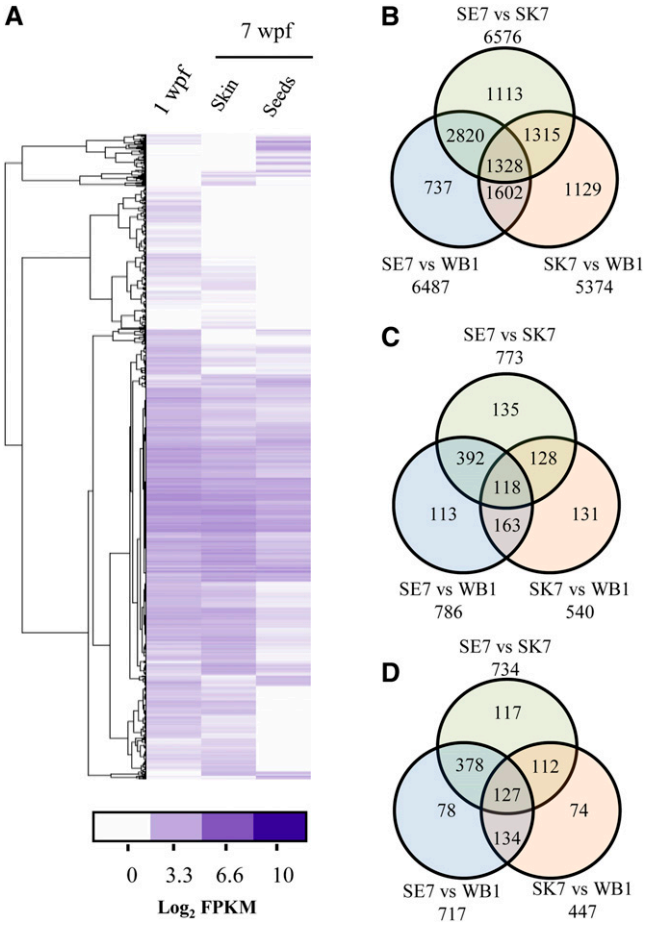


Figure 3. Analysis of Gene Expression in Tannat Berries.

(A) Hierarchical clustering of RNA-Seq read counts (Log2 FPKM) in three tissues of Tannat berries.

(B) Venn diagram showing numbers of commonly and uniquely differentially expressed known genes in pairwise comparisons of whole berry at 1 wpf (WB1), seeds at 7 wpf (SE7), and skin at 7 wpf (SK7).

(C) Venn diagram showing numbers of commonly and uniquely differentially expressed novel genes in pairwise comparisons among WB1, SE7, and SK7 samples.

(D) Venn diagram showing numbers of commonly and uniquely differentially expressed varietal genes in pairwise comparisons among WB1, SE7, and SK7 samples.

Tannat genome using a hybrid approach (Gan et al., 2011) that relied on both the de novo assembly of Illumina genomic reads and iterative mapping against the PN40024 reference genome (Jaillon et al., 2007). The highly contiguous genome assembly is comparable in length to the reference genome (98.9%), but divergence is demonstrated by the presence of more than two million SNPs and almost 700,000 indels or unbalanced substitutions. A striking feature of the Tannat genome is the low level of heterozygosity compared with common cultivated grapevine cultivars. The total number of SNPs is similar in Tannat and another representative cultivar (Corvina), but the Tannat genome contains only half as many ambiguous, potentially heterozygous bases, in agreement with previous reports (Gonzalez Techera et al., 2004). The relatively high level of homozygosity could reflect the geographic isolation of Tannat in southwestern France after its branching from Manseng Noir (Lacombe et al., 2013) and the resulting predominance of self-fertilization, but it may also indicate that Tannat is more tolerant to inbreeding (Lassalle, 1993) than Pinot varieties, where the viability of selfed progenies declines after a few generations (Bronner and Oliveira, 1991). Furthermore, the distribution of variants among Tannat genes indicates the fixation of mutations in genes related to specific biological processes such as development, the transfer of glycosyl groups, and protein Tyr kinase activity (see Supplemental Table 2 online). The enrichment of variation within particular functional classes of genes may reflect the environment in which Tannat was cultivated, and this is particularly interesting because genome-wide structural and gene content variations may drive the phenotypic variations that characterize different genotypes within a species (McHale et al., 2012).

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 (Chen et al., 2013); therefore, we named them varietal genes.

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 (Dixon et al., 2002) and whose expression is tightly regulated during berry development (Boss et al., 1996; Bogs et al., 2005). The production and accumulation of tannins in the seeds and skins occurs mainly during the early phases of berry development (Conde et al., 2007; Boido et al., 2011). We identified 141 varietal genes encoding 19 enzymes involved in the production of polyphenols (see Supplemental Data Set 2 online). 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 (Dixon et al., 2002). 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 (Pollack et al., 1999; Wang et al., 2012). Interestingly, varietal CHS genes account for more than 30% of total CHS gene expression, including VVV_00520, which is one of the two most strongly expressed CHS genes in our data set, on par with the known gene CHS2.

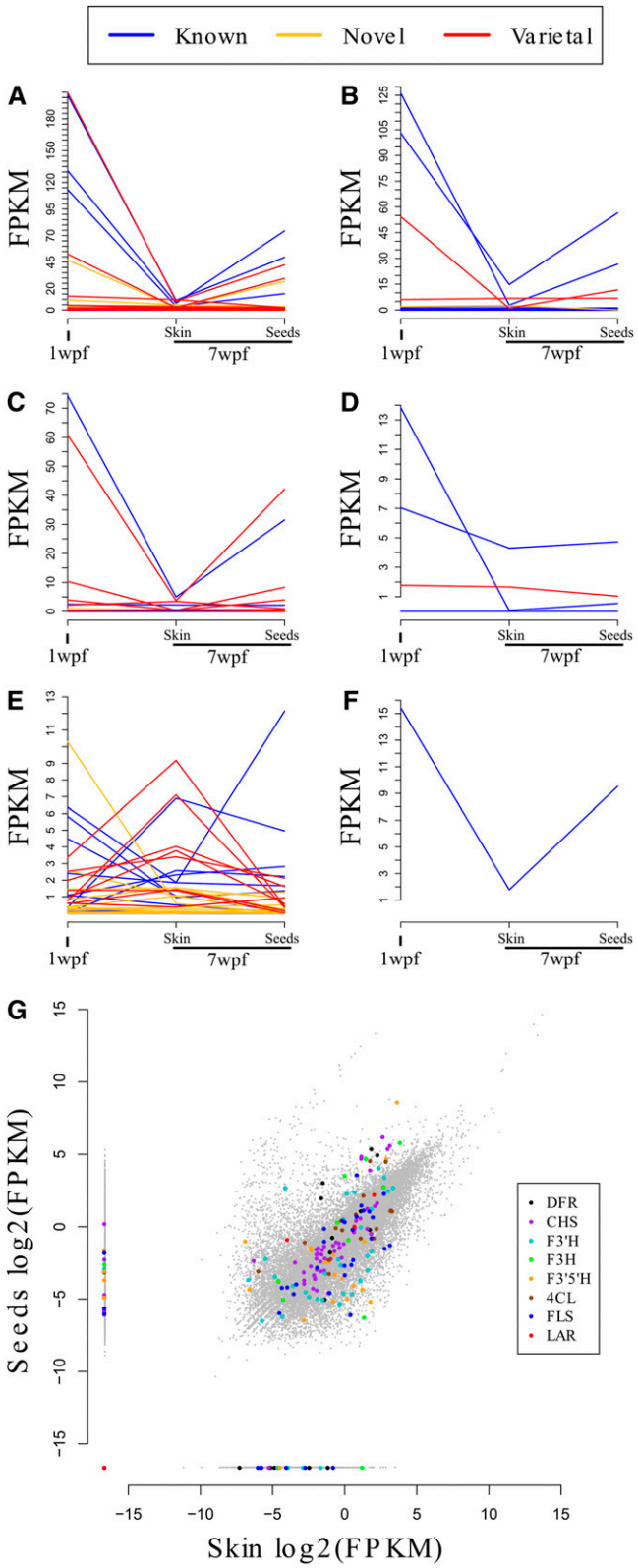


Figure 4. Comparison of Gene Expression in Different Tannat Berry Tissues.

(A) to **(F)** Expression profiles of CHS **(A)**, F3H **(B)**, DFR **(C)**, LAR **(D)**, FLS **(E)**, and MybPA1 **(F)** in whole berries at 1 wpf, skin at 7 wpf, and seeds at 7 wpf. **(G)** Scatterplot of RNA-Seq expression values in seeds and skin at 7 wpf. Selected genes involved in polyphenol biosynthesis are shown in color as indicated.

Most of the gene families representing key enzymes in the flavonoid pathway showed some degree of expansion, including FLS, F3H, F3′H, and flavonoid 3′5′-hydroxylase (F3′5′H). The biosynthesis of flavonols is induced in the grapevine flower but decreases postanthesis and is then restored in berry skins after veraison (Downey et al., 2003b). The most common flavonols in Tannat berry skins are different glucosylated forms of quercetin (Boido et al., 2011), which together with myricetin have cytoprotective activities (Echeverry et al., 2004).

FLS catalyzes the first committed step in flavonol biosynthesis. We inspected the expression of two previously characterized FLSs (FLS1 and FLS2), and we found that, in agreement with previous reports (Downey et al., 2003b), the first was expressed at minimal levels at 1 wpf, while the second was not expressed at all in the analyzed stages. However, we identified two varietal FLS genes expressed specifically in the berry skin at 7 wpf (VVV_00613 and VVV_00659), suggesting that they are involved in the biosynthesis of specific flavonols or similar compounds in the skin during the early stages of Tannat berry development. Three flavonoid hydroxylases (F3H, F3′H, and F3′5′H) catalyze common steps in the flavonoid and anthocyanin biosynthesis pathways (Jeong et al., 2006). F3′H is also responsible for the conversion of kaempferol to quercetin, a natural antioxidant present in particular in red grape berries and wines. RNA-Seq data revealed that F3′5′H genes were expressed at double the level of F3′H genes in the Tannat berries, agreeing with the reported sevenfold increase in the ratio of 3′,5′-dihydroxy/3′-hydroxy anthocyanin derivatives compared with Pinot Noir and Corvina, perhaps reflecting differences in the activities of these pathways among different cultivars (Mattivi et al., 2006). Interestingly, three of the five most strongly expressed F3′5′H genes were varietal, suggesting that expansion within this family might explain the high 3′,5′-dihydroxy/3′-hydroxy anthocyanin ratio observed in Tannat berries.

Whereas FLS catalyzes the first step in the isoflavonoid pathway, DFR catalyzes a key step in the flavonoid pathway common to anthocyanin and tannin biosynthesis and is completely devoted to tannin biosynthesis before veraison. DFR was represented by six varietal genes that, in addition to the eight genes present in the reference genome, accounted for more than 50% of overall DFR expression with a peak of 62% in the seeds at 7 wpf. A varietal DFR (VVV_00044) was among the two most strongly expressed DFR genes in our data set and showed an expression pattern consistent with tannin biosynthesis during the early phases of berry development. This varietal gene was slightly more specific for the seed at 7 wpf compared with the known DFR gene (Boss et al., 1996).

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

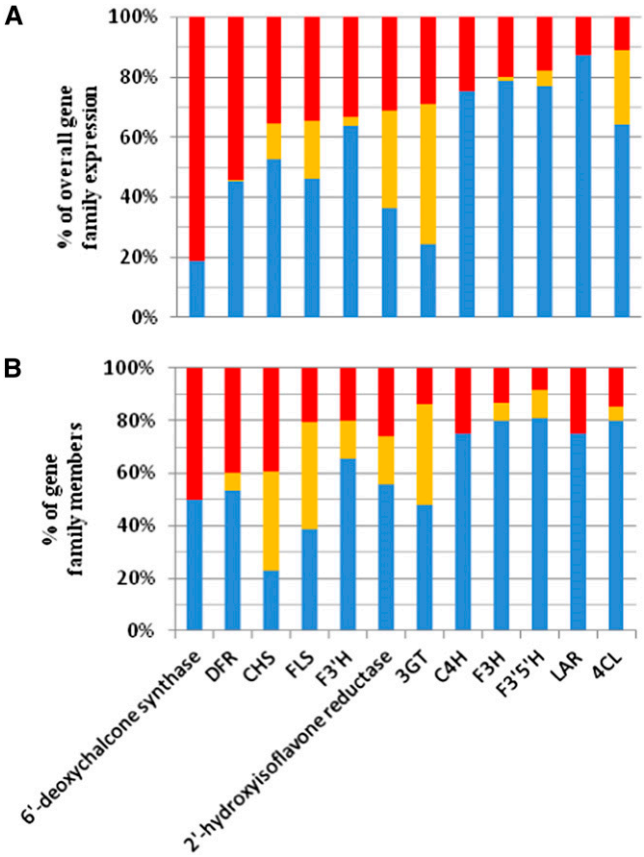


Figure 5. Family Expansion and Expression Contribution of Genes Related to Polyphenol Biosynthesis.

(A) Percentage contribution of known, novel, and varietal genes to the overall expression of gene families encoding enzymes in the phenylpropanoid and flavonoid biosynthesis pathways.
(B) Percentage of known, novel, and varietal genes in gene families encoding enzymes in the phenilpropanoid and flavonoid biosynthesis pathways.

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 (see Supplemental Data Set 2 online). It is worth noting that gene expression levels do not always correlate with the related protein biosynthesis and/or activity levels or with the abundance of their target metabolites. However, previous results strongly suggest a high correlation of expression levels of polyphenol-related genes with the concentration of important polyphenol metabolites, including tannins, in grapevine (Dal Santo et al., 2013).

In sharp 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 little or no expansion of gene members, and the observed contribution of varietal genes to the overall expression of those families was very limited. The expression patterns of regulators known to be involved in the regulation of polyphenol biosynthesis were nevertheless consistent with current knowledge concerning the regulation of the pathway (see Supplemental Data Set 2 online). For example, the WD40 transcription factor WDR1 and the bHLH transcription factor MycA1, which are known to coregulate the biosynthesis of polyphenols in the early and late stages of berry development, were expressed at high levels at 1 wpf and repressed at 7 wpf, in agreement with previous reports (Matus et al., 2010). We also detected the expression of Myc1, which can induce the expression of key structural genes representing the flavonoid pathway (Hichri et al., 2010), with a pattern similar to that of MycA1 but at markedly lower FPKM values. Two of the most important regulators for the biosynthesis of grapevine proanthocyanidins, the R2R3 MYB factors MybPA1 and MybPA2, also showed expression profiles that agree with previous reports and match the temporal and spatial regulation of tannin biosynthesis (Bogs et al., 2007; Terrier et al., 2009). MybPA1 was the most strongly expressed R2R3 MYB factor both at 1 wpf and in the seeds at 7 wpf, whereas MybPA2 was moderately expressed at 1 wpf and almost completely silent at 7 wpf in both tissues.

In conclusion, Tannat appears to rely on a standard palette of transcription factors to increase the level of polyphenols through the activation of an expanded set of genes encoding the enzymes involved in this pathway. 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.

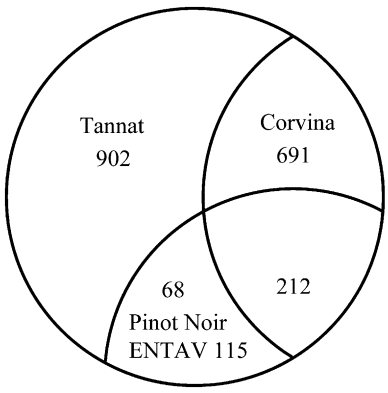


Figure 6. Categorization of the 1873 Genes Not Shared with PN40024.

Number of genes found in common among Tannat, Corvina, and Pinot Noir (ENTAV 115).

METHODS

Sample Collection

Unexpanded young leaves were collected from the old Uruguayan Tannat clone UY11 of grapevine (*Vitis vinifera*) (Gonzalez Techera et al., 2004) in a vineyard in Canelones, southern Uruguay. The samples were frozen in liquid nitrogen in the field and stored at –80°C.

Ten berry clusters from the French Tannat commercial clone 717 (ENTAV nomenclature: ENTAV 1995) were collected 1, 5, and 7 wpf during the 2011 to 2012 growing season, from a vineyard of 2000 plants in Melilla, Montevideo, southern Uruguay. The clusters were sampled randomly from 10 different plants, excluding the borders of the vineyard. Ten berries were randomly selected from each cluster and pooled with berries from the other plants on the same vineyard site, resulting in three independent pools of 30 berries for each developmental stage. Skin and seeds sampled at each time point were frozen separately in liquid nitrogen, except at 1 wpf, where the whole berries were used because the seeds were not fully developed.

Nucleic Acid Extraction

Frozen leaves were ground to powder under liquid nitrogen. Total DNA was extracted using the Qiagen Plant DNeasy kit (Qiagen) according to the manufacturer's instructions. The purity and quantity of the DNA were determined using a Nanodrop 1000 spectrophotometer (Thermo Scientific). Frozen berries were ground to powder using liquid nitrogen, and total RNA was extracted using the Spectrum Plant Total RNA kit (Sigma-Aldrich) according to the manufacturer's instructions. The purity and quantity of the RNA were determined using a Nanodrop 1000 spectrophotometer (Thermo Scientific). RNA integrity was determined using a Bioanalyzer 2100 (Agilent) with RNA 6000 Nano Kit I (Agilent).

Library Preparation and Sequencing

Genomic DNA (6 μg) was mixed with 750 μL of nebulization buffer (Illumina) and fragmented for 6 min in a nebulizer (Life Technologies) using compressed nitrogen at 35 p.s.i. to produce fragments with a typical size range of 150 to 700 bp and a peak at 500 bp. The sheared DNA was purified using the QIAquick PCR purification kit (Qiagen) and eluted in 30 μL of water. The quality of the fragmented DNA was determined using the Agilent DNA 1000 kit on an Agilent 2100 bioanalyzer, and DNA libraries were prepared using TruSeq DNA sample preparation kits (Illumina). Illumina RNA-Seq libraries were prepared from 2.5 μg of total RNA per sample according to the manufacturer's instructions.

The DNA-Seq and RNA-Seq libraries were size-selected at 350 to 550 bp using the Pippin Prep DNA size selection system (Sage Science). Library quality was determined using the Agilent High Sensitivity DNA kit on the Agilent 2100 bioanalyzer, and the quantity was determined by quantitative PCR using the KAPA Library Quantification kit (KapaBiosystems). Libraries were then pooled in equimolar concentrations and sequenced with the TruSeq Sequencing by Synthesis Kit v3-HS and TruSeq Paired End Cluster Kit v3-cBot-HS (Illumina) using an Illumina HiSequation 1000 sequencer according to the manufacturer's instructions to generate 100-bp paired-end reads.

Genome Assembly, Identification, and Analysis of Polymorphic Regions

The genomic sequences were assembled using the IMR/DENOM v0.3.3 pipeline (Gan et al., 2011) with default parameters and the 12x PN40024 genome as a reference (Jaillon et al., 2007). Chloroplast and mitochondrial sequences were excluded from the final reconstruction.

To identify polymorphic regions, we aligned genomic reads to the reconstructed Tannat genome using Burrows-Wheeler Aligner (BWA) with default parameters (Li and Durbin, 2010). Alignments were processed with SAMtools v0.1.18 (Li et al., 2009) to produce a histogram of the sequence coverage at each position in the genome. We identified polymorphic regions in the final genome as regions with a contiguous read coverage lower than four (Gan et al., 2011). The effects of polymorphisms detected in the IMR/DENOM pipeline on the functionality of proteins encoded by the V1 grapevine annotation were predicted using Variant Effect Predictor with EnsEMBL database v64 (McLaren et al., 2010).

The gene space of the assembled genome was assessed by aligning CEGs (Parra et al., 2009) with the Tannat genome using BLAST with a 65% identity threshold (Altschul et al., 1990).

Gene Annotation

V1 annotation gene models were translated to the reconstructed Tannat genome by taking structural variations in the reference genome into account and adjusting the coordinates accordingly using bespoke software (available on request). The reference-guided reconstruction of transcripts was performed using TopHat v2.0.6 (Trapnell et al., 2009) and Cufflinks v2.0.2 (Trapnell et al., 2010) as described (Trapnell et al., 2012) and was compared and merged with the translated V1 annotation using the Cuffcompare withol in the Cufflinks suite. We clustered the loci reconstructed by Cufflinks based on a comparison with sequences in the NCBI NR protein database using BLASTX (E-value $\leq 10^{-5}$) and VvGI database v8.0 (E-value $\leq 10^{-5}$) and manual inspection of alignments.

De Novo Assembly of Tannat Transcripts and the Identification of New Genes

De novo assembly of transcripts from RNA-Seq data was performed using the Velvet/Oases assembler with a k-mer value of 49, a minimum contig length of 200, an insert length of 310, and a standard deviation of 100 (Zerbino and Birney, 2008; Schulz et al., 2012). These parameters were optimized for scaffold N50 and total base coverage, after running the assembler across a range of parameters. Redundancy among assembled sequences was first removed using cd-hit-est with a threshold of 90% sequence identity (Li and Godzik, 2006).

We then applied five different filters to identify de novo–assembled transcripts missing from the genome assembly. Transcripts were aligned against the Tannat genome using GMAP v2012-11-07 with default parameters (Wu and Watanabe, 2005), and those with at least 80% of coverage and identity were discarded. Contaminant sequences were identified by comparison with NCBI NR database sequences using BLASTX (E-value $\leq 1*10^{-5}$) and were discarded. Putative grapevine transcripts were identified by sequence identity with VvGI using BLASTX (E-value $\leq 1*10^{-5}$). The coding potential of transcripts was computed with CPC v0.9 (Kong et al., 2007), and those with a complete open reading frame including start and stop codons and more than 50 codons overall were retained as high-confidence putative protein-coding transcripts. The Tannat genomic reads were aligned against transcripts using BWA with default parameters and the alignments were processed with BedTools suite v2.17.0 (Quinlan and Hall, 2010) to produce a histogram of the sequence coverage at each position in the transcripts. Those with <80% sequence coverage were discarded.

De novo–assembled transcripts that passed the above filters were classified as missing from the genome assembly and compared with different sets of genomic sequences to ascertain their presence in other genotypes. For PN40024, we obtained both the set of Sanger sequences used for the current version of the genome (downloaded from ftp://ftp-private.ncbi.nlm.nih.gov/pub/TraceDB/vitis_vinifera/) and two different

sets of Illumina genomic reads (downloaded from http://urgi.versailles.inra.fr/galaxy/u/nchoisne/h/pn40024-public-data). For Pinot Noir ENTAV115, we downloaded the final assembled sequences from GenBank (accession numbers AM423240 to AM489403; Velasco et al., 2007). PN40024 Sanger reads were aligned with BLAT v34x12 against the transcripts to find matches >100 bp with no gaps and sequence coverage ≥80% (Kent, 2002), whereas the other sets of sequences were aligned with BWA using default parameters. A histogram of the sequence coverage at each position was computed with genomeCoverageBed (Quinlan and Hall, 2010). Transcripts covered (≥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. A multiple alignment was performed for each cluster using mafft v6.864b (Katoh and Toh, 2008) and manually checked for consistency. Final clusters were considered to be genes. The gene space for genes missing from the assembled genome was assessed as described above.

Functional Annotation of Genes

Genes were functionally annotated by integrating the V1 manual annotation (Grimplet et al., 2012) and automatic annotations performed with Blast2GO (Conesa et al., 2005). Novel and varietal genes without assigned Enzyme Commission numbers were manually classified by BLAST using known polyphenol-related genes as reference sequences and discarding hits with an E-value >$1*10^{-10}$.

Expression of Polyphenol Pathway Genes

RNA-Seq reads were aligned against the Tannat genome and gene sequences with BWA using default parameters. We counted the mapping reads and calculated FPKM expression values for known, novel, and varietal genes using the R package DESeq with default parameters (Anders and Huber, 2010) to assess which genes were differentially expressed among the developmental stages and tissues under analysis (false discovery rate < 5%, |Log2FC| > 1).

Accession Numbers

Genomic sequence data from this article can be found in the NCBI Sequence Read Archive under accession numbers SRR863595 and SRR863618. Transcriptome sequence data from this article can be found in the NCBI Sequence Read Archive under accession numbers SRR866531, SRR866540, SRR866544, SRR866568, SRR866569, SRR866570, SRR866571, SRR866572, SRR866574, SRR866575, and SRR866576. Transcriptome shotgun assembly data from this article can be found in the NCBI Transcriptome Shotgun Assembly database under accession number GAKH00000000. Variation data are downloadable as IMR/DENOM sdi files from http://ddlab.sci.univr.it/files/Tannat/corvina.sdi and http://ddlab.sci.univr.it/files/Tannat/tannat.sdi. Annotations of known and novel genes can be downloaded from http://ddlab.sci.univr.it/files/Tannat/annotation.gtf. Mappings between varietal gene IDs and transcripts accession numbers can be downloaded from http://ddlab.sci.univr.it/files/Tannat/putative_transcripts2genes.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Number of Variants Identified at Each Iterative Mapping Step with IMR.

Supplemental Figure 2. Genes Affected by Potentially Disruptive Mutations.

Supplemental Figure 3. CEGMA Analysis of Tannat Genes.

Supplemental Figure 4. GO Gene Classification.

Supplemental Table 1. *V. vinifera* cv Tannat Genome Sequencing and Assembly Statistics.

Supplemental Table 2. Enrichment Analysis of Genes Containing SNPs.

Supplemental Table 3. *V. vinifera* cv Tannat Transcriptome Sequencing Statistics.

Supplemental Table 4. Comparison of Tannat Annotation with Other Plant Annotations.

Supplemental Table 5. *V. vinifera* cv Tannat Transcriptome Assembly Statistics.

Supplemental Table 6. Classification of Contaminants Present in de Novo–Assembled Transcripts Not Mapping on the Tannat Genome.

Supplemental Data Set 1. Definitions, GO and KEGG Annotations, Expression Values, and Differential Expression Tests for All 34,680 Tannat Genes.

Supplemental Data Set 2. Gene Number and Expression Levels for Gene Families Related to Polyphenol Biosynthesis; Expression Levels for Selected Genes Related to Polyphenol Biosynthesis.

ACKNOWLEDGMENTS

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 Consejo Superior de Investigaciones Cientificas Group 656 of UdelaR and ANII of Uruguay for supporting the travels and PhD fellowship of C.D.S.

AUTHOR CONTRIBUTIONS

C.D.S. performed research, prepared samples, and analyzed data. G.Z. performed research, analyzed data, contributed new computational tools, and wrote the article. A.F. and L.V. analyzed data and wrote the article. A.M. and A.D.M. analyzed data and contributed new computational tools. G.B., P.T., C.A., and E.Z. prepared samples. E.B. and E.D. contributed new analytic tools. C.G. prepared samples and wrote and reviewed the article. M.P. reviewed the article. F.C. designed the research. M.D. designed the research and wrote the article.

Received September 20, 2013; revised September 20, 2013; accepted November 14, 2013; published December 6, 2013.

REFERENCES

Ågren, J.A., and Wright, S.I. (2011). Co-evolution between transposable elements and their hosts: A major factor in genome size evolution? Chromosome Res. **19:** 777–786.

Alcalde-Eon, C., Boido, E., Carrau, F., Dellacassa, E., and Rivas-Gonzalo, J.C. (2006). Pigment profiles in monovarietal wines produced in Uruguay. Am. J. Enol. Vitic. **57:** 449–459.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Ammiraju, J.S.S., et al. (2007). Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus Oryza. Plant J. **52:** 342–351.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. **11:** R106.

Bhattacharya, A., Sood, P., and Citovsky, V. (2010). The roles of plant phenolics in defence and communication during Agrobacterium and Rhizobium infection. Mol. Plant Pathol. **11:** 705–719.

Bogs, J., Downey, M.O., Harvey, J.S., Ashton, A.R., Tanner, G.J., and Robinson, S.P. (2005). Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves. Plant Physiol. **139:** 652–663.

Bogs, J., Jaffé, F.W., Takos, A.M., Walker, A.R., and Robinson, S.P. (2007). The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. Plant Physiol. **143:** 1347–1361.

Boido, E., Alcalde-Eon, C., Carrau, F., Dellacassa, E., and Rivas-Gonzalo, J.C. (2006). Aging effect on the pigment composition and color of *Vitis vinifera* L. Cv. Tannat wines. Contribution of the main pigment families to wine color. J. Agric. Food Chem. **54:** 6692–6704.

Boido, E., García-Marino, M., Dellacassa, E., Carrau, F., Rivas-Gonzalo, J.C., and Escribano-Bailón, M.T. (2011). Characterisation and evolution of grape polyphenol profiles of *Vitis vinifera* L. cv. Tannat during ripening and vinification. Aust. J. Grape Wine Res. **17:** 383–393.

Boss, P.K., Davies, C., and Robinson, S.P. (1996). Expression of anthocyanin biosynthesis pathway genes in red and white grapes. Plant Mol. Biol. **32:** 565–569.

Bronner, A. andOliveira, A. (1991). Création et étude de lignées chez le pinot noir (*Vitis vinifera* L.). J. Int. Sci. Vigne Vin. **3:** 133–148.

Carrau, F., Boido, E., Gaggero, C., Medina, K., Disegna, E., and Dellacassa, E. (2011). *Vitis vinifera* Tannat, chemical characterization and functional properties. Ten years of research. In Multidisciplinary Approaches on Food Science and Nutrition for the XXI Century, Rosana Filip, ed (Kerala, India: Transworld Research Network), pp. 53–71.

Carrau, F.M. (1997). The emergence of a new Uruguayan wine industry. J. Wine Res. **8:** 179–185.

Chen, S., Krinsky, B.H., and Long, M. (2013). New genes as drivers of phenotypic evolution. Nat. Rev. Genet. **14:** 645–660.

Chong, J., Poutaraud, A., and Hugueney, P. (2009). Metabolism and roles of stilbenes in plants. Plant Sci. **177:** 143–155.

Cipriani, G., et al. (2010). The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. Theor. Appl. Genet. **121:** 1569–1585.

Conde, C., Silva, P., Fontes, N., Dias, A.C.P., Tavares, R.M., Sousa, M.J., Agasse, A., Delrot, S., and Gerós, H. (2007). Biochemical changes throughout grape berry development and fruit and wine quality. Food **1:** 1–22.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21:** 3674–3676.

Corder, R., Mullen, W., Khan, N.Q., Marks, S.C., Wood, E.G., Carrier, M.J., and Crozier, A. (2006). Oenology: Red wine procyanidins and vascular health. Nature **444:** 566.

Dal Santo, S., Tornielli, G.B., Zenoni, S., Fasoli, M., Farina, L., Anesi, A., Guzzo, F., Delledonne, M., and Pezzotti, M. (2013). The plasticity of the grapevine berry transcriptome. Genome Biol. **14:** r54.

Dixon, R.A, Achnine, L, Kota, P, Liu, C.J, Reddy, M.S.S, and Wang, L. (2002). The phenylpropanoid pathway and plant defence: A genomics perspective. Mol. Plant Pathol. **3:** 371–390.

Downey, M.O., Harvey, J.S., and Robinson, S.P. (2003a). Analysis of tannins in seeds and skins of Shiraz grapes throughout berry development. Aust. J. Grape Wine Res. **9:** 15–27.

Downey, M.O., Harvey, J.S., and Robinson, S.P. (2003b). Synthesis of flavonols and expression of flavonol synthase genes in the developing grape berries of Shiraz and Chardonnay (*Vitis vinifera* L.). Aust. J. Grape Wine Res. **9:** 110–121.

Echeverry, C., Blasina, F., Arredondo, F., Ferreira, M., Abin-Carriquiry, J.A., Vasquez, L., Aspillaga, A.A., Diez, M.S., Leighton, F., and Dajas, F. (2004). Cytoprotection by neutral fraction of tannat red wine against oxidative stress-induced cell death. J. Agric. Food Chem. **52:** 7395–7399.

Fitzpatrick, D.F., Hirschfield, S.L., and Coffey, R.G. (1993). Endothelium-dependent vasorelaxing activity of wine and other grape products. Am. J. Physiol. **265:** H774–H778.

Gan, X., et al. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. Nature **477:** 419–423.

Giovannoni, J. (2001). Molecular biology of fruit maturation and ripening. Annu. Rev. Plant Physiol. **52:** 725–749.

Gonzalez Techera, A., Jubany, S., Ponce de Leon, I., Boido, E., Dellacassa, E., Carrau, F.M., Hinrichsen, P., and Gaggero, C. (2004). Molecular diversity within clones of cv. Tannat (*Vitis vinifera*). Vitis **43:** 179–185.

Grimplet, J., Van Hemert, J., Carbonell-Bejerano, P., Díaz-Riquelme, J., Dickerson, J., Fennell, A., Pezzotti, M., and Martínez-Zapater, J.M. (2012). Comparative analysis of grapevine whole-genome gene predictions, functional annotation, categorization and integration of the predicted gene sequences. BMC Res. Notes **5:** 213.

Gu, X., Creasy, L., Kester, A., and Zeece, M. (1999). Capillary electrophoretic determination of resveratrol in wines. J. Agric. Food Chem. **47:** 3223–3227.

Hansey, C.N., Vaillancourt, B., Sekhon, R.S., de Leon, N., Kaeppler, S.M., and Buell, C.R. (2012). Maize (*Zea mays* L.) genome diversity as revealed by RNA-sequencing. PLoS ONE **7:** e33071.

Hashimoto, M., Kim, S., Eto, M., Iijima, K., Ako, J., Yoshizumi, M., Akishita, M., Kondo, K., Itakura, H., Hosoda, K., Toba, K., and Ouchi, Y. (2001). Effect of acute intake of red wine on flow-mediated vasodilatation of the brachial artery. Am. J. Cardiol. **88:** 1457–1460, A9.

Hichri, I., Heppel, S.C., Pillet, J., Léon, C., Czemmel, S., Delrot, S., Lauvergeat, V., and Bogs, J. (2010). The basic helix-loop-helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. Mol. Plant **3:** 509–523.

Jaillon, O., et al; French-Italian Public Consortium for Grapevine Genome Characterization (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449:** 463–467.

Jeong, S.T., Goto-Yamamoto, N., Hashizume, K., and Esaka, M. (2006). Expression of the flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase genes and flavonoid composition in grape (Vitis vinifera). Plant Sci. **170:** 61–69.

Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. **9:** 286–298.

Kent, W.J. (2002). BLAT—The BLAST-like alignment tool. Genome Res. **12:** 656–664.

Khan, N., Adhami, V.M., and Mukhtar, H. (2010). Apoptosis by dietary agents for prevention and treatment of prostate cancer. Endocr. Relat. Cancer **17:** R39–R52.

Kobayashi, S., Goto-Yamamoto, N., and Hirochika, H. (2004). Retrotransposon-induced mutations in grape skin color. Science **304:** 982.

Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L., and Gao, G. (2007). CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. **35** (Web Server issue)**:** W345–W349.

Kutchan, T.M. (2005). A role for intra- and intercellular translocation in natural product biosynthesis. Curr. Opin. Plant Biol. **8:** 292–300.

Lacombe, T., Boursiquot, J.M., Laucou, V., Di Vecchi-Staraz, M., Péros, J.P., and This, P. (2013). Large-scale parentage analysis in an extended set of grapevine cultivars (*Vitis vinifera* L.). Theor. Appl. Genet. **126:** 401–414.

Lassalle, D. (1993). Nouvéaux cépages INRA pour le sud de la France. Progrès Agricole et Viticole **110:** 511–517.

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N., and Caboche, M. (2006). Genetics and biochemistry of seed flavonoids. Annu. Rev. Plant Biol. **57:** 405–430.

Li, H, and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26:** 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:** 2078–2079.

Li, R., et al. (2010). Building the sequence map of the human pan-genome. Nat. Biotechnol. **28:** 57–63.

Li, W., and Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics **22:** 1658–1659.

Lin, J.K., and Weng, M.S. (2006). Flavonoids as nutraceuticals. In The Science of Flavonoids, E. Grotewold, ed (New York: Springer), pp. 213–238.

Mattivi, F., Guzzon, R., Vrhovsek, U., Stefanini, M., and Velasco, R. (2006). Metabolite profiling of grape: Flavonols and anthocyanins. J. Agric. Food Chem. **54:** 7692–7702.

Mattivi, F., Vrhovsek, U., Masuero, D., and Trainotti, D. (2009). Differences in the amount and structure of extractable skin and seed tannins amongst red grape varieties. Aust. J. Grape Wine Res. **15:** 27–35.

Matus, J.T., Poupin, M.J., Cañón, P., Bordeu, E., Alcalde, J.A., and Arce-Johnson, P. (2010). Isolation of WDR and bHLH genes related to flavonoid synthesis in grapevine (*Vitis vinifera* L.). Plant Mol. Biol. **72:** 607–620.

McHale, L.K., Haun, W.J., Xu, W.W., Bhaskar, P.B., Anderson, J.E., Hyten, D.L., Gerhardt, D.J., Jeddeloh, J.A., and Stupar, R.M. (2012). Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiol. **159:** 1295–1308.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., and Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics **26:** 2069–2070.

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005). The microbial pan-genome. Curr. Opin. Genet. Dev. **15:** 589–594.

Morgante, M., De Paoli, E., and Radovic, S. (2007). Transposable elements and the plant pan-genomes. Curr. Opin. Plant Biol. **10:** 149–155.

Myles, S., Chia, J.M., Hurwitz, B., Simon, C., Zhong, G.Y., Buckler, E., and Ware, D. (2010). Rapid genomic characterization of the genus vitis. PLoS ONE **5:** e8219.

Parra, G., Bradnam, K., Ning, Z., Keane, T., and Korf, I. (2009). Assessing the gene space in draft genomes. Nucleic Acids Res. **37:** 289–297.

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., and Brown, P.O. (1999). Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat. Genet. **23:** 41–46.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics **26:** 841–842.

Schatz, M.C., Witkowski, J., and McCombie, W.R. (2012). Current challenges in de novo plant genome sequencing and assembly. Genome Biol. **13:** 243.

Schneeberger, K., et al. (2011). Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. Proc. Natl. Acad. Sci. USA **108:** 10249–10254.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics **28:** 1086–1092.

Springer, N.M., et al. (2009). Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet. **5:** e1000734.

Swanson, Wagner, R.A, Eichten, S.R, Kumari, S, Tiffin, P, Stein,, J.C, Ware, D, and Springer, N.M. (2010). Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res. **20:** 1689–1699.

Terrier, N., Torregrosa, L., Ageorges, A., Vialet, S., Verriès, C., Cheynier, V., and Romieu, C. (2009). Ectopic expression of VvMybPA2 promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. Plant Physiol. **149:** 1028–1041.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics **25:** 1105–1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. **7:** 562–578.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. **28:** 511–515.

Velasco, R., et al. (2007). A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE **2:** e1326.

Venturini, L., et al. (2013). De novo transcriptome characterization of *Vitis vinifera* cv. Corvina unveils varietal diversity. BMC Genomics **14:** 41.

Wang, Y., Wang, X., and Paterson, A.H. (2012). Genome and gene duplications and gene expression divergence: A view from plants. Ann. N. Y. Acad. Sci. **1256:** 1–14.

Wu, T.D., and Watanabe, C.K. (2005). GMAP: A genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics **21:** 1859–1875.

Zerbino, D.R., and Birney, E. (2008). Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18:** 821–829.

Zhang, Y., Mao, L., Wang, H., Brocker, C., Yin, X., Vasiliou, V., Fei, Z., and Wang, X. (2012). Genome-wide identification and analysis of grape aldehyde dehydrogenase (ALDH) gene superfamily. PLoS ONE **7:** e32153.

The High Polyphenol Content of Grapevine Cultivar Tannat Berries Is Conferred Primarily by Genes That Are Not Shared with the Reference Genome

Cecilia Da Silva, Gianpiero Zamperin, Alberto Ferrarini, Andrea Minio, Alessandra Dal Molin, Luca Venturini, Genny Buson, Paola Tononi, Carla Avanzato, Elisa Zago, Eduardo Boido, Eduardo Dellacassa, Carina Gaggero, Mario Pezzotti, Francisco Carrau and Massimo Delledonne

Plant Cell; originally published online December 6, 2013;
DOI 10.1105/tpc.113.118810

This information is current as of December 6, 2013

Supplemental Data	**http://www.plantcell.org/content/suppl/2013/11/20/tpc.113.118810.DC1.html**
Permissions	https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&issn=1532298X&WT.mc_id=pd_hw1532298X
eTOCs	Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain
CiteTrack Alerts	Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain
Subscription Information	Subscription Information for *The Plant Cell* and *Plant Physiology* is available at: http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY

CAPÍTULO II

Theoretical and Applied Genetics

High levels of galloylated procyanidins in Tannat grape seeds correlate with the expression of a new epicatechin galloyl transferase-like gene

--Manuscript Draft--

Manuscript Number:	
Full Title:	High levels of galloylated procyanidins in Tannat grape seeds correlate with the expression of a new epicatechin galloyl transferase-like gene
Article Type:	Original Article
Funding Information:	Agencia Nacional de Investigación e Innovación (ANII POSG 2012 1 8711) — MSc Cecilia Da Silva CSIC Universidad de la Republica UY (CSIC Group 656) — Dr. Francisco Carrau Fondazione Cariverona (Centro de Genomica Funzionale Vegetale) — Prof Massimo Delledonne
Abstract:	Vitis vinifera cv Tannat is the grape cultivar with the highest content of galloylated procyanidins in seeds. Although it has been shown that the health-promoting effects of galloylated procyanidins are stronger than those of non-galloylated procyanidins, the mechanism involved in galloylation of flavan-3-ols is still poorly understood. Based on the sequence of the epicatechin galloyl transferase (ECGT) recently reported for tea leaves, we found five putative ECGTs in Tannat with at least 50% aminoacid identity and 90% coverage. A phylogenetic analysis of ECGTs in plants show that there are two clearly distinct clades, both including grape sequences. By RNASeq analysis during Tannat berry development, separating grape skins and seeds, we show the expression pattern of these putative ECGTs, as well as other enzymes involved in the synthesis and regulation of the galloylated procyanidins precursors (gallic acid, β-glucogallin and epicatechin). The contents of galloylated procyanidins and their precursors during grape berry development, analyzing separately skins and seeds, was determined by HPLC-MS. Expression of two of the ECGT-like genes was confirmed by RT-qPCR in the same samples used for metabolome analysis. We find the highest concentration of galloylated procyanidins in Tannat grains (fruit set), one week after flowering. Our results show that the high content of galloylated procyanidins in Tannat grape seeds correlate with the expression of a new epicatechin galloyl transferase-like gene, different from the one reported in tea leaves.
Corresponding Author:	Francisco Carrau School of Chemistry, Universidad de la Republicar, Uruguay URUGUAY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	School of Chemistry, Universidad de la Republicar, Uruguay
Corresponding Author's Secondary Institution:	
First Author:	Cecilia Da Silva
First Author Secondary Information:	
Order of Authors:	Cecilia Da Silva Eduardo Boido Carina Gaggero Massimo Delledonne Francisco Carrau
Order of Authors Secondary Information:	

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Author Comments:	Montevideo, October 7th 2016 Dear Editor in Chief Theoretical Applied Genetics Dr. Albrecht E. Melchinger, We are sending you our manuscript "High levels of galloylated procyanidins in Tannat grape seeds correlate with the expression of a new epicatechin galloyl transferase-like gene" (Cecilia Da Silva , Eduardo Boido, Carina Gaggero, Massimo Delledonne, Francisco Carrau), which we would like to be considered for publication in Theoretical Applied Genetics. In our opinion this work makes a significant contribution to the understanding of the synthesis of galloylated procyanidins, some of the most bioactive compounds against cell oxidative reactions during nutrition or cell aging. We report here that the expression pattern of one putative ECGT, which is different from the tea ECGT, correlates with the high content of galloylated procyanidins in Tannat grape seeds. In our opinion the strength of this work is that it is not only based on genomics and transcriptomics but it also integrates a complete metabolomic analysis of the most complex plant metabolites. English was proof read by Proof Reading Services. net. Thank you for your attention and we look forward to your answer.
Suggested Reviewers:	Stella Grando stella.grando@fmach.it Vitis genetics and genomics
	Patricio Hinrichsen phinrichsen@inia.cl plant molecular biology and genomics
	Rafael Radi rafael.radi@gmail.com expert in antioxidant compounds and health
	Robin Buell buell@msu.edu plant metabolism and genomics

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Manuscript Click here to download Manuscript Da Silva et al 2016 TAG.doc

Click here to view linked References

High levels of galloylated procyanidins in Tannat grape seeds correlate with the expression of a new epicatechin galloyl transferase-like gene

Cecilia Da Silva [1,2¶] , Eduardo Boido [1], Carina Gaggero [2], Massimo Delledonne [3], Francisco Carrau [1*]

[1] Área Enología y Biotecnología de Fermentaciones, Facultad de Química, Universidad de la República, 11800 Montevideo, Uruguay.

[2] Departamento de Biología Molecular, Instituto de Investigaciones Biológicas Clemente Estable, 11600 Montevideo, Uruguay.

[3] Centro di Genomica Funzionale, Dipartimento di Biotecnologie, Universitá degli Studi di Verona, 37134 Verona, Italy.

* Corresponding author information: fcarrau@fq.edu.uy

¶ Present address: Cecilia Da Silva, Prof. Adj. Biología Molecular Vegetal, PDU Espacio de Biología Vegetal del Noreste, Universidad de la República sede Tacuarembó, Ruta 5 km 387, Tacuarembó, Uruguay

Running title: Galloylated procyanidins and a new ECGT-like gene in Tannat seeds

Abstract

Vitis vinifera cv Tannat is the grape cultivar with the highest content of galloylated procyanidins in seeds. Although it has been shown that the health-promoting effects of galloylated procyanidins are stronger than those of non-galloylated procyanidins, the mechanism involved in galloylation of flavan-3-ols is still poorly understood.

Based on the sequence of the epicatechin galloyl transferase (ECGT) recently reported for tea leaves, we found five putative ECGTs in Tannat with at least 50% aminoacid identity and 90% coverage. A phylogenetic analysis of ECGTs in plants show that there are two clearly distinct clades, both including grape sequences.

By RNASeq analysis during Tannat berry development, separating grape skins and seeds, we show the expression pattern of these putative ECGTs, as well as other enzymes involved in the synthesis and regulation of the galloylated procyanidins precursors (gallic acid, β-glucogallin and epicatechin).

The contents of galloylated procyanidins and their precursors during grape berry development, analyzing separately skins and seeds, was determined by HPLC-MS. Expression of two of the ECGT-like genes was confirmed by RT-qPCR in the same samples used for metabolome analysis.

We find the highest concentration of galloylated procyanidins in Tannat grains (fruit set), one week after flowering. Our results show that the high content of galloylated procyanidins in Tannat grape seeds correlate with the expression of a new epicatechin galloyl transferase-like gene, different from the one reported in tea leaves.

Keywords

galloylated procyanidins, ECGT, grape seed, Tannat, RNASeq, HPLC-MS

Introduction

Grapevine (*Vitis vinifera*) cv Tannat, a red wine grape historically grown in Southwestern France, was introduced into Uruguay in the 1870s. Nowadays, Tannat is the Uruguayan emblematic cultivar for the production and export of high-quality red wines (Carrau, 1997). Tannat is the first cultivated grapevine whose genome has been sequenced with high coverage (Da Silva et al. 2013). Tannat has unusually high levels of polyphenolic

compounds, producing wines with an intense purple colour, rich structure and remarkable antioxidant properties (Carrau et al. 2011; Echeverry et al. 2004; Khan et al. 2015). In spite of being considered as secondary metabolites, polyphenols perform many crucial biological functions in plants, including defense against biotic and abiotic stress. Plants synthesize specific polyphenols to attract pollinators and fruit-dispersing animals and also to deter herbivores. Polyphenols act as signaling molecules within the plant, in the rhizosphere and in the surrounding soil (Treutter 2006; Quideau et al. 2011; Weston and Mathesius 2013).

The polyphenolic compounds named as proanthocyanidins, or condensed tannins, are oligomers and polymers of flavan-3-ols. They are the major flavonoids found in grape berry, as dimers or oligomers of catechin and epicatechin and their gallic acid esters. Flavan-3-ols are the most structurally complex subclass of flavonoids, ranging from the simple monomers (+)-catechin and its isomer (-)-epicatechin, which can be hydroxylated to form gallocatechins and also undergo esterification with gallic acid, through to complex oligomeric and polymeric structures of up to 50 units. Proanthocyanidins consisting exclusively of epicatechin units are called procyanidins, and are the most abundant type in plants (Crozier et al. 2009). Proanthocyanidins are the main vasoactive polyphenols in red wine. They induce the endothelium-dependent dilatation of blood vessels and suppress the synthesis of the vasoconstriction peptide endothelin-1 (Fitzpatrick et al. 1993; Hashimoto et al. 2001). Their 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 (Corder et al. 2006). Polyphenols acquired from the moderate consumption of red wine also provide pharmaceutical and nutritional benefits in humans (Donovan et al, 1999; Lin and Weng 2006, Witte et al, 2014), e.g. helping to prevent cancer and reducing the inflammation associated with coronary artery disease (Khan et al. 2010; Khan et al. 2015). It has been recently postulated that the beneficial effects of polyphenols on cardiovascular and brain function may be explained by a signaling role of these phytochemicals within human cellular signal transduction pathways (Kennedy 2014). Polyphenols may also contribute to the maintenance of gut health by the modulation of the gut microbial balance through the stimulation of the growth of beneficial bacteria and the inhibition of pathogen bacteria, exerting prebiotic-like effects (Dueñas et al., 2015).

Procyanidins and oligomeric procyanidins (OPC) are a type of proanthocyanidins that yield cyanidin when depolymerized under oxidative conditions. Grape seed proanthocyanidins

include only procyanidins. Although grape tannins are present both in skins and seeds, there are higher levels in seeds (Hanlin et al. 2011). Esterification with 3,4,5-trihydroxybenzoic (gallic) acid is referred to as galloylation. Tannat is the grape cultivar with the highest galloylation rate in seeds (Boido et al. 2011). The mechanism involved in galloylation of flavan-3-ols is still poorly understood. The biosynthesis of gallic acid has been studied in oak (Muir et al, 2011) and also in grapevine (Bontpart et al. 2016). Galloylation of glucose with gallic acid yields β-glucogallin (Mittasch et al. 2014; Khater et al. 2012), the first specific metabolite which serves as the precursor of galloylated flavan-3-ols. Recently, an epicatechin galloyl transferase (ECGT) has been isolated from tea leaves (Liu et al. 2012). (see Fig 1). Several Myb transcription factors regulate proanthocyanidin biosynthesis in grapes (Bogs et al. 2007; Terrier et al. 2009; Koyama et al. 2014).

It has been shown that the health-promoting effects of galloylated procyanidins are stronger than those of non-galloylated procyanidins (Lizarraga et al. 2007; Kim et al. 2014). It was also demonstrated that galloylation of the procyanidin core structure was a prerequisite for its anti-viral activity against Influenza A (Dersken et al. 2014). Galloylation could represent a means for increasing the biological potency of phenolic compounds and improving their cancer-preventive activities (Lewandowska et al. 2016). Galloylation also improves the oenological and colloidal properties of proanthocyanidins (Vidal et al. 2003; Fabre et al. 2010).

We have investigated the genetic basis for the synthesis of galloylated proanthocyanidins during Tannat berry development, and we have determined the main polyphenolic compounds synthesized during this process. Based on the sequence of the *Camellia sinensis* ECGT, we performed a search for highly homologous sequences in plants. We report here that the expression pattern of one putative ECGT from Tannat, which is different from the tea ECGT, correlates with the high content of galloylated procyanidins in Tannat grape seeds.

Materials and Methods

Homologous search and phylogenetic analyses.

The search for proteins highly homologous to CsECGT (CN103740740) was performed using Blastp (Altschul et al. 1990) against the NCBI non-redundant protein database, considering ECGT-like proteins the ones with at least 50% identity and more than 90% coverage. Protein sequences were aligned using ClustalX v2.1 (Larkin et al. 2007) with

default parameters. Maximum Likelihood (ML) Phylogenetic tree was done through W-IQ-TREE (Trifinopoulos et al. 2016) using IQ-TREE V.1.4.3 (Nguyen et al. 2015), model selection was set to auto and bootstrapping (1,000 replicates) was done with the ultrafast bootstrapping algorithm (Minh et al. 2013). The best fitting model according to the Bayesian Information Criterion (Schwarz 1978) and the corrected Akaike Information Criterion (Burnham and Anderson 2004) was the WAG substitution model (Whelan and Goldman, 2001) with five FreeRate categories (Yang 1995, Soubrier et al. 2012).

Plant material

Grape berry samples from the French Tannat commercial clone 717 (ENTAV nomenclature:ENTAV1995) were all collected from a vineyard of 2000 plants in Melilla, Montevideo, Southern Uruguay. Sampling was done at 1, 5, 7 10 and 15 weeks after flowering (waf) for the 2011 to 2012 growing season; and at 1, 5, 10 and 15 waf for the 2014 to 2015 growing season. Ten clusters were sampled randomly from 10 different plants, excluding the borders of the vineyard. Ten berries were randomly selected from each cluster and pooled with berries from the other plants on the same vineyard site, resulting in three independent pools of 30 berries for each developmental stage. In the 10 waf samples (veraison) totally green and totally red berries were separated to form different samples. Skins and seeds were manually separated at each time point and frozen separately in liquid nitrogen, except at 1 wpf, where the whole berries were used because the seeds were not fully developed. Samples were stored at –80°C until use.

RNA extraction

Frozen berry tissues were ground to powder using a mortar and liquid nitrogen, and total RNA was extracted with the Spectrum Plant Total RNA kit (Sigma-Aldrich) according to the manufacturer's instructions. The RNA purity and quantity were determined using a Nanodrop 1000 spectrophotometer (Thermo Scientific). RNA integrity was determined using a Bioanalyzer 2100 (Agilent) with RNA 6000 Nano Kit I (Agilent).

Library preparation and sequencing.

Illumina RNA-Seq libraries for the 2011 to 2012 growing season samples were prepared from 2.5 µg of total RNA per sample according to the manufacturer's instructions. The RNA-Seq libraries were size-selected at 350 to 550 bp using the Pippin Prep DNA size

selection system (Sage Science). Library quality was determined using the Agilent High Sensitivity DNA kit on the Agilent 2100 bioanalyzer, and the quantity was determined by quantitative PCR using the KAPA Library Quantification kit (KapaBiosystems). Libraries were then pooled in equimolar concentrations and sequenced with the TruSeq Sequencing by Synthesis Kit v3-HS and TruSeq Paired End Cluster Kit v3-cBot-HS (Illumina) using an Illumina HiSequation 1000 sequencer according to the manufacturer's instructions to generate 100-bp paired-end reads.

Transcriptome sequence data of samples from 1, 5 and 7 waf can be found in the NCBI Sequence Read Archive under accession numbers SRR866531, SRR866540, SRR866544, SRR866568, SRR866569, SRR866570, SRR866571, SRR866572, SRR866574, SRR866575, and SRR866576 (Da Silva et al 2013); and of samples from 10 and 15 waf under accession numbers xxxx. Annotations can be downloaded from http://ddlab.sci.univr.it/ files/Tannat/annotation.gtf.

Differential expression analyses.

RNA-Seq reads from the 2011 to 2012 growing season were aligned against the Tannat genome with TopHat v2.0.4 (Trapnell et al. 2009). Mapping reads were counted with HTSeq 0.6.1 using the intersection-nonempty mode (Anders et al 2015), and expression values were calculated using the R package DESeq2 with default parameters (Love et al. 2014) to assess which genes were differentially expressed among the developmental stages and tissues under analysis (adjusted pvalue < 0.001, |Log2FC| >1). All the differential expression values are expressed against 1waf. In this study we only assessed the expression values of genes coding for enzyme and transcription factors directly involved in the synthesis of galloylated procyanidins: VvANR1 (VIT_00s0361g00040), VvANR2 (VIT_00s0361g00020), VvANR3 (VIT_00s0361g00010), VvECGT-like (VIT_03s0088g00050, VIT_03s0088g00260, VIT_03s0091g01240, VIT_03s0091g01290, VIT_11s0052g00790), VvgGT1 (VIT_03s0091g00040), VvgGT2 (VIT_03s0180g00200), VvgGT3 (VIT_03s0180g00320), VvMYB5a (VIT_08s0007g07230), VvMYB5b (VIT_06s0004g00570), VvMYBPA1 (VIT_15s0046g00170), VvMYBPA2 (VIT_11s0016g01320), VvMYBPAR (VIT_11s0016g01300), VvSDH1 (VIT_05s0020g02030), VvSDH2 (VIT_14s0030g00670), VvSDH3 (VIT_14s0030g00660), VvSDH4 (VIT_14s0030g00650).

RT-qPCR

All RNA samples were treated with RNase-free DNase Set (QIAGEN), and no DNA contamination was detected by PCR amplification. cDNA was synthesized using SuperScript® III First-Strand Synthesis SuperMix for qRT-PCR (INVITROGEN) according to the manufacturer's protocol. To validate RNA-Seq results, RT-qPCR amplification of four genes was performed on the 2011 to 2012 growing season; and to facilitate parallel comparisons to the metabolite data, amplification of the VvECGT-like genes VIT_03s0088g00260 and VIT_03s0091g01290 was performed on the 2014 to 2015 growing season. Primers used were designed by Primer3Plus (Untergasser et al. 2007) and are listed in Supplemental Table S1. Real Time PCRs were performed in a Corbett Rotor-gene RG-6000 (QIAGEN) using the EXPRESS SYBR® GreenER™ qPCR Supermix universal (INVITROGEN) according to the manufacturer's protocol. The thermocycler was programmed to run for 2 min at 50°C for UDG incubation, 2 min at 95°C, followed by 40 cycles of 15 s at 95°C, 60 s at 50°C, and the Melting Curve from 60°C to 95°C. Non-template controls were included for each primer pair. Specific annealing of the primer pairs used was controlled on the Melting Curve. Expression data were normalized against the grapevine ACT7 gene (VIT_04s0044g00580, Zenoni et al. 2010, Polesani et al. 2010, Perazzolli et al. 2012). The amplification efficiencies of the different primer combinations were analyzed and all were >90%. Relative expression was determined using the Pfaffl model (Pfaffl 2001), and the values were expressed relative to the 1 waf corresponding growing season. Each data point is the mean value of two biological replicates for the 2011 to 2012 growing season, and three biological replicates the 2014 to 2015 growing season. Students t-test was used to evaluate the statistical significance of the difference in values using the GraphPad Prism Software (La Jolla, CA, USA).

HPLC-MS Analyses of galloylated procyanidins and their precursors.

HPLC-MS analyses were performed using a HPLC-DAD 1200 (Agilent Technologies), equipped with auto sampler G1367B, binary pump model G1312A and triple Quad LC/MS 6410 LC detector. Gradients of solvent A (water-formic acid, 95:5, v/v) and solvent B (acetonitrile) were applied at flow rate 1.0 mL/min over a C18 reverse-phase 150 mm x 4.6 mm column (Phenomenex, Torrance, CA) as follows: 10 to 30% B from 0 to 40 min, then 30 to 35% B from 40 to 48 min, and 35 to 10% B from 48 to 49 min, then 10% from 49 to 55 min to reequilibrate the column to initial conditions. MS analyses were performed using electrospray ionization (ESI) and the experimental conditions were as follows: sheath gas

flow 10 L/min, gas temperature 300°C, nebulizer 40 psi, and capillary voltage 4000 V. The fragmentation voltage was 60 V and mass scan (MS) were recorded in negative ion mode between m/z 100 and 1600 a.m.u.

The quantification was carried out with the calibration curves obtained with (+)-catechin (Sigma) using the MS- peak (289 a.m.u.). The concentrations of the different studied compounds were calculated with their respective MS- peak area and expressed as (+)-catechin in μg/g of grape.

For each compound, variance analyses were performed with the model sample (nine samples, skin and seed in different recollection date) and repetition (duplicate) without interaction, using the software Statistica 7.1 (StatSoft, Tulsa, OK, 1984-2005).

Results

A phylogenetic analysis of plant epicatechin galloyl transferases-like (ECGT-like) proteins shows two distinct clades

The search for proteins highly homologous to the ECGT from tea (*Camellia sinenis*) (Patent CN103740740A) yielded 97 sequences (Supplementary Table 2S), almost all belonging to dicotyledons, except 3 from the monocotyledon *Phoenix dactylifera* (date palm fruit). This search retrieved 12 *Vitis vinifera* proteins that correspond to 7 genes according to the 12X version of the *Vitis vinifera* grape genome, classified as serine carboxypeptidase-like proteins (SCPL). When the search was performed in the *Vitis vinifera* cv. Tannat genome only 5, from these 7 genes encoding putative ECGTs, were found: VIT_03s0088g00050, VIT_03s0088g00260, VIT_03s0091g01240, VIT_03s0091g01290 and VIT_11s0052g00790; which correspond to the protein GenBank accessions XP_002266867, XP_002272116, XP_010647903, XP_002269439, XP_010656845, respectively. The 5 proteins present in Tannat are depicted in bold in Fig 2.

The phylogenetic analysis of the ECGT-like proteins in plants (Fig. 2, Supplementary Figure 1S) groups these sequences in 2 distinct main clades, independently of their species of origin. Green tea ECGT falls into the lower clade in Fig 2, closely related to sequences from *Diospyros kaki* (Kaki fruit) (BAH89272), *Eucalyptus grandis* (XP_010028337) and *Vitis vinifera* (XP_002272116, protein encoded by the gene

VIT_03s0088g00260), forming what we called subclade B, depicted in green colour. In the upper clade in Fig 2, there is a subclade formed as well by sequences from *Diospyros kaki*, *Eucalyptus grandis* and *Vitis vinifera* (BAF56655, XP_010028338, XP_002269439, CBI22419), together with sequences from *Fragaria vesca* (wild strawberry), *Jatropha curcas* (Barbados nut) and *Manihot esculenta* (cassava or juca) (XP_011457835, XP_012093157, OAY45348), which we called subclade A, depicted in red colour. The XP_002269439 and CBI22419 *Vitis vinifera* proteins correspond to the same gene VIT_03s0091g01290.

Expression pattern of Tannat enzymes and regulators related to the synthesis of galloylated procyanidins in berries

We have analyzed the expression pattern during Tannat berry development of the different enzymes and regulators related to the synthesis of galloylated procyanidins, i.e., SDHs that synthesize gallic acid from shikimate pathway metabolites; gGTs that form β-glucogallin; ANRs that synthesize epicatechin; putative ECGTs or epicatechin galloyl transferases, as well as several transcription factors that have been reported to regulate proanthocyanidin biosynthesis in grapevine.

In order to study the spatial and temporal expression pattern of the five ECGT-like genes found in Tannat, an RNA-Seq approach was performed throughout berry development, analyzing skins and seeds separately at six developmental stages, from fruit set to the ripe stage, during the 2011-2012 growing season. Each of these five genes has a unique expression pattern (Fig. 3A, Supplementary Table 3S). VIT_03s0091g01290 and VIT_11s0052g00790 expression is higher at 1waf and in seeds at 7waf; however, the highest level for VIT_03s0091g01290 is at 1 waf while for VIT_11s0052g00790 it is in seed. VIT_03s0088g00260 shows higher expression at 1waf and lower in skins of red berries at veraison (10 waf). VIT_03s0088g00050 shows low expression levels in seeds at 7waf and in skins at 15 waf while VIT_03s0091g01240 shows a relatively higher expression before veraison.

The differential expression analyses of each of the time points and tissues versus 1 waf (Supplementary Table 4S) show that only VIT_03s0088g00260, VIT_03s0091g01290 and VIT_11s0052g00790 are differentially expressed genes.

The expression patterns of the enzymes involved in the last stages of the biosynthesis of galloylated procyanidins (ANR, SDH, gGT, Fig. 3C-E) were also analyzed, as well as the

transcription factors known to regulate this metabolic pathway (MYBPA1, MYBPA2, MYBPAR, MYB5a, MYB5b, Fig. 3B). The enzymes are expressed mainly in green tissues, showing higher expression at 1waf and in seeds, while the transcription factors do not show such a clear pattern. The expression of MYBPAR is higher in all tissues before veraison; MYB5b expresses mainly in red skins; the highest expression of MYBPA2 and MYB5a is at 1waf and skins at 5waf, and later on in maturation the expression increases around veraison; while MYBPA1 expression is high at 1waf and in seeds at 7waf, and slightly lower in skins of red berries at veraison (10waf). The differential expression analyses for each of the time points and tissues versus 1 waf for these genes are shown in Supplementary Table 4S.

Relative expression levels of 5 genes were evaluated by RT-qPCR in the same samples used for RNA-Seq analysis. As expected, a high correlation between these two methods was found (a Pearson's correlation coefficient of 0.97, Supplementary Figure 2S). The relative expression of Tannat ECGT gene VIT_03s0088g00260 at different berry developmental stages and tissues was analyzed during two different growing seasons (2011-2012 and 2014-2015) aiming to determine if the expression pattern is maintained in different years. There is a high correlation between these two years as shown by a Pearson's correlation coefficient of 0.89 (Supplementary Figure 3S)

Galloylated procyanidins are produced mostly in Tannat grains and seeds during berry development.

The contents of galloylated procyanidins and their precursors were determined by HPLC-MS during Tannat berry development in the 2014-2015 growing season (Fig. 4, Suplementary Table 5S). A high proportion of galloylated phenols was confirmed for Tannat in seed (32%) and skin (26%). The concentrations of the precursors β-glucogallin and gallic acid were much lower (almost 30 times lower) than the other compounds. In fruit set (1 waf) and in seeds all the metabolites analyzed showed much higher concentration (at least 10 times higher) than in skins. Regarding the levels measured at different time points within each tissue, each metabolite shows a different profile: gallic acid and β-glucogallin in fruit set and seeds have a rather similar profile, their concentration is very high at 1waf, decrease at 5waf and then gradually increase until maturity (15 waf), while in skins the highest content of gallic acid is reached at 15 waf and that of β-glucogallin at 5waf. Epicatechin was not detected at 1waf. In seeds, the highest content of epicatechin is

reached in green grapes at veraison (10waf_sk_g), while in skins the highest concentration is at maturity (15waf_sk). Analyzing the sum of all the galloylated compounds there is a maximum peak at 1waf and also high levels in seeds with a slightly higher content in green grapes at veraison. In skins, the concentration of galloylated compounds starts from a very low level at 5 waf and continues to diminish to negligible levels.

Since these measurements of galloylated procyanidins were performed in samples from the 2014-2015 growing season, we used the same samples for confirmation of gene expression by RT-qPCR. Considering that the pattern of expression of total galloylated compounds showed a very high concentration at 1 waf, high levels in seeds and very low levels in skins, confirmation by RT-qPCR was only performed for VIT_03s0091g01290 and VIT_03s0088g00260.

Correlation between Tannat metabolome and transcriptome data identifies a new ECGT-like protein

The expression patterns obtained by RT-qPCR of the Tannat ECGT-like genes VIT_03s0091g01290 and VIT_03s0088g00260, were compared with the quantification pattern of total galloylated compounds in different berry tissues throughout the 2014-2015 growing season (Fig. 5). In seeds, the expression pattern of both genes correlates with the total galloylated compounds. However, in skins the expression pattern of VIT_03s0088g00260 does not correlate to the measured metabolites because in red grapes at veraison (10waf_sk_r) the expression of this gene rises, while galloylated compounds keep diminishing. In the case of VIT_03s0091g01290, the expression in skins was so low that it could not be detected by this methodology. Therefore, we conclude that the expression pattern of VIT_03s0091g01290 in skins correlates well with the concentration of galloylated compounds.

In summary, based on all the results, we propose that VIT_03s0091g01290 is the gene encoding an ECGT-like protein in Tannat. The genes and regulators with a similar expression pattern to VIT_03s0091g01290 are SDH3, gGT1, ANR1 and MYBPA1.

Discussion

Red wine and green tea share many secondary metabolites like galloylated procyanidins, that have been shown to be beneficial for human health (Namkung et al. 2010). Since Tannat is the grapevine cultivar showing the highest galloylation level in seeds (Boido et al. 2011) and galloylated compounds have proven to be more active than their non-galloylated counterparts (Kim et al. 2014), we decided to search for a Tannat putative galloyl transferase based on homology to the one reported for tea leaves.

Although large-scale sequencing projects rely on sequence similarity for protein annotation, sequence similarity is not always a guarantee of functional similarity. In fact, protein function has only been experimentally determined for a small fraction of all the annotated proteins (Kim and Buell 2015). Joshi and Xu (2007) show that high sequence identity is a better indicator of function similarity than e-value and that higher sequence identity has a higher probability of similar functionality. Hence, when we searched for proteins homologous to the galloyl transferase reported for tea leaves (Patent CN103740740A), we limited our search to proteins sharing at least 50% aminoacid identity and 90% sequence coverage. When analyzing the phylogenetic tree we noted that most of the accessions correspond to species from the Rosids group (*Citrus sinensis, Citrus clementina, Eucalyptus grandis, Vitis vinifera,*
Prunus mume, Prunus persica, Pyrus bretschneideri, Jatropha curcas, Ricinus communis, Manihot esculenta, Fragaria vesca, Populus trichocarpa). The evolutionary success of this monophyletic group with more than 70,000 species, including important crops such as fruit trees (orange, peach and pear) and legumes, is probably related to their ability to produce a great diversity of flavonoids (Zavala and Opazo, 2015). It must be noted that not all the species involved have been completely sequenced, we are limited to the sequences uploaded to GenBank up to now, and to the limits we have imposed. However, the phylogenetic analysis groups plants known to accumulate high amounts of galloylated flavan-3-ols, such as *Vitis vinifera* (Boido et al. 2011), *Camellia sinensis* (Jiang et al. 2013), *Diospyros kaki* (Li et al. 2010), and/or ellagitannins, that are polyphenols formed from the oxidative linkage of galloyl groups from pentagalloyl glucose in wood trees. Ellagitannins are not present in grapevine but they accumulate in species such as *Fragaria vesca* (Landete, 2011) and *Eucalyptus sp.* (Boulekbache-Makhlouf et al. 2010). We may have orthologous as well as paralogous proteins with high similarity to ECGT from *Camellia sinensis*. Different ECGTs might be expressed in different tissues, the ECGT reported for

tea is expressed in leaves and its probable ortholog in grape is VIT_03s0088g00260. In this work we show that the Tannat ECGT expressed in seeds that correlates with the content of galloylated metabolites is VIT_03s0091g01290.

In the literature, both VIT_03s0088g00260 and VIT_03s0091g01290 have been reported as related to proanthocyanidin pathways. In 2009, Terrier et al., working with Syrah grape cultivar, describe the VvMybPA2 transcription factor and show that ectopic expression of either VvMybPA1 or VvMybPA2 in grapevine hairy roots induce qualitative and quantitative changes of the proanthocyanidin profiles. Among the genes induced in common by VvMybPA1 and VvMybPA2 and linked to the proanthocyanidin pathway, they report one Glc acyltransferase (GSVIVT00038626001). In the 8X version of the Pinot reference genome, GSVIVT00038626001 corresponds to VvGAT1, while in the 12X version, VvGAT1 corresponds to GSVIVT01037059001. VvGAT1 is a close homolog to DkSCPL1, a gene of persimmon (*Diospyros kaki*), reported previously by Ikegami et al. (2007).

Carrier et al. (2013) performed an integrative strategy combining genetic mapping and transcriptomic approaches to select the most promising candidates involved in proanthocyanidin biosynthesis in grape berry. They analyzed the F1 progeny of Syrah x Grenache and focused on the following proanthocyanidin traits in berry skins: total quantities, percentage of galloylated units and mean degree of polymerization. In the list of selected candidate genes they include GSVIVG01037059001/VvGAT-like and GSVIVG01037791001. They also report one SNP in VvGAT-like that is significantly associated with the variation in galloylation rate.

GSVIVG01037059001/VvGAT-like corresponds to VIT_03s0088g00260, while GSVIVG01037791001 corresponds to VIT_03s0091g01290.

In 2014, Koyama et al. report a new transcription factor, VvMybPAR, in Cabernet Sauvignon. They show that in a transient reporter assay with a dual luciferase system, VvMYBPAR activates the promoters of several flavonoid pathway genes involved in proanthocyanidin biosynthesis, including one SCP (Vv3s0091g01290), closely related to DkSCPL1 (Ikegami et al. 2007). This SCP was identified as a putative proanthocyanidin biosynthetic gene because its gene expression was specifically observed at an early stage of berry development.

Persimmon (*Diospyrus kaki*) fruit is an interesting model because the high-molecular weight proanthocyanidins are accumulated in ‘tannin cells’ during their development. Enclosing young persimmon fruits on the tree in an ethanol-containing polyethylene bag causes insolubilization of these proanthocyanidins. Ikegami et al. (2007) examined the

effect of this treatment on *de novo* synthesis of proanthocyanidins. They performed suppression subtractive hybridization (SSH) to identify genes differentially expressed during this treatment and found one serine carboxypeptidase-like (SCPL) protein (DkSCPL1; BAF56655.1). In 2009, Akagi et al., report a second serine carboxypeptidase-like protein (DkSCPL2; BAH89272.1) involved in proanthocyanidin synthesis in persimmon fruit. In the phylogenetic analysis shown in our results (Fig 2, and in more detail in the supplementary Fig 1S), DkSCPL1 appears in the red subclade A as a close homolog to the protein encoded by VIT_03s0091g01290, while DkSCPL2 is the corresponding homolog in the green subclade B to the grape protein encoded by VIT_03s0088g00260. The spatio-temporal expression of these versions of SCPL proteins and its possible involvement in galloylation in persimmon remain to be elucidated.

Working with Syrah, Bontpart et al. (2016) found that two genes, VvSDH3 and VvSDH4, mainly expressed in immature berry tissues in which galloylated flavan-3-ols are accumulated, were able to produce gallic acid *in vitro*. The over-expression of VvSDH3 in hairy-roots increased the contents of gallic acid, β-glucogallin, and galloylated flavan-3-ols. In Syrah, the expression levels of VvSDH1, VvSDH2, and VvSDH4 were maximal at the green stage (18 daf), low at véraison, and intermediate at maturity. During the green stage, VvSDH1 was less expressed in the pulp than in skin and seeds, VvSDH2 was less expressed in skin than in pulp and seeds, whereas VvSDH4 expression was relatively high in the three tissues. VvSDH3 had a very low level of expression whatever the tissue or the developmental stage, except in the immature seeds.

In Tannat we found a similar level of expression for VvSDH1 and VvSDH2 through berry development. VvSDH3 and VvSDH4 were highly expressed in Tannat grain and seeds.

Khater et al., (2012) report the isolation and functional validation of three Syrah grapevine genes encoding glucosyltransferases that were called Vv galloyl Glucosyl Transferase (VvgGT) because of their preference to form glc-esters with gallic acid. The expression of the genes was studied during Syrah berry development showing that VvgGT1 is expressed in green tissues in both skin and seeds, VvgGT3 expression is almost seed specific, while VvgGT2 is predominantly expressed in the pulp.

In Tannat grapes we found a similar pattern of expression for VvgGTs, although VvGT1 is not expressed in skin at 7 waf and in skin of green grapes at veraison (10 waf_sk_g).

Regarding the synthesis of epicatechin, ANR1 was found to be the main gene induced in grain and seed, as has been reported in Syrah (Bogs et al. 2007; Terrier et al. 2009) and Norton grapes (Ali et al. 2011).

Grape berry development occurs in two successive growth phases. From flowering until veraison proanthocyanidin synthesis takes place, while anthocyanin biosynthesis coincides with the onset of berry ripening. Several MYB transcription factors (VvMYBPA1, VvMYBPA2, VvMYBPAR, VvMYB5A, VvMYB5B) have been shown to positively regulate the spatio-temporal expression of the proanthocyanidin-specific biosynthetic genes in different grape cultivars (Bogs et al. 2007; Terrier et al. 2009; Deluc et al. 2006; Ali et al. 2011; Koyama et al. 2014). VvMYBPA1 has been shown to play a significant role in the regulation of proanthocyanidins during early berry development and in seeds of Norton (Ali et al. 2011), Syrah (Bogs et al. 2007) and Tannat (this work). In Cabernet Sauvignon, VvMYBPA2 appears to be more important than VvMYBPA1 (Ali et al. 2011; Koyama et al. 2014). In Tannat, VvMYBPA2 expresses mostly in grain and at lower levels in skin at some stages, with no expression in seed. VvMYB5A and VvMYB5B are general regulators of the flavonoid pathway and their impact in Tannat berry development is not clear. VvMYBPAR, adjacent and similar to VvMYBPA2, has been recently shown to induce several genes of the proanthocyanidin pathway in Cabernet Sauvignon (Koyama et al. 2014), suggesting a polygenic regulation of proanthocyanidin biosynthesis by closely related MYB transcription factors. In Tannat, VvMYBPAR expresses in grain and early green stages (upto 7waf) both in seed and skin, suggesting a possible fine-tuning role in coordination with other transcription factors. The differential regulation and impact of these transcription factors in different grapevine cultivars might explain the differential contents of proanthocyanidins and merit further studies. In fact, it has already been shown that different accessions of the same species (*A. thaliana*, Bac-Molenaar et al. 2015; *O. sativa*, Dey et al. 2016) show polymorphisms and differential expression of key transcription factors.

HPLC-MS analysis of procyanidins in Tannat seeds shows the synthesis and presence of beta glucogallin that is probably immediately used as substrate by ECGT to form epicatechin gallate.

The highest concentration of galloylated flavan-3-ols (59%) was detected at fruit set, 1 waf while at veraison levels of galloylated compounds in seeds were at 20%, higher than the levels reported for many other grape varieties including the recent report for Syrah (Bontpart et al. 2016). During maturity galloylated phenols in skin decreased from 22% during veraison to 6% at harvest time. These data confirm that the galloylation process

occurs mainly previous to veraison, as also supported by the RNA transcription analysis. The enzyme isolated from tea leaves has been shown to gallolyate an epicatechin monomer (Liu et al, 2012). However, we cannot rule out the possibility that the galloylation process could function not only on epicatechin monomers but also on dimers, trimers or polymers, since all of these galloylated compounds were detected in our metabolome analysis.

Galloylated procyanidins have beneficial effects for human health and also play an important role in wine quality. Thus, both from a nutraceutical and food quality perspective, it is important to understand the mechanisms and regulation of procyanidins galloylation in a grape cultivar such as Tannat.

Acknowledgments

We thank the support of CSIC Group 656 of UdelaR, CSIC Iniciacion en la Investigacion and Pedeciba, Uruguay. We also thank Fondazione Cariverona (Completamento e attività del Centro di Genomica Funzionale Vegetale), Verona, Italy, for supporting this work. Cecilia Da Silva was awarded with a PhD and travel fellowships from ANII, Uruguay.

Author contribution statement

CDS performed research, prepared samples, analyzed data and wrote the article. EB performed HPLC MS analysis and reviewed the article. CG analyzed data, designed the research and wrote the article. MD reviewed the article. FC designed the research and reviewed the article.

References

Akagi T, Ikegami A, Suzuki Y, Yoshida J, Yamada M, Sato A, Yonemori K (2009) Expression balances of structural genes in shikimate and flavonoid biosynthesis cause a difference in proanthocyanidin accumulation in persimmon (*Diospyros kaki* Thunb.) fruit. Planta 230:899e915

Ali MB, Howard S, Chen S, Wang Y, Yu O, Kovacs LG, Qiu W. (2011) Berry skin development in Norton grape: distinct patterns of transcriptional regulation and flavonoid biosynthesis. BMC Plant Biol 11:7

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Anders S, Pyl PT, Huber W (2015) HTSeq- A Python framework to work with high-throughput sequencing data. Bioinformatics, 31:166–169.

Bac-Molenaar JA, Fradin EF, Rienstra JA, Vreugdenhil D, Keurentjes JJ (2015) GWA Mapping of Anthocyanin Accumulation Reveals Balancing Selection of MYB90 in *Arabidopsis thaliana*. PLoS One. 10:e0143212.

Bogs J, Jaffé FW, Takos AM, Walker AR, Robinson SP (2007) The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. Plant Physiol 143: 1347–1361

Boido E, García-Marino M, Dellacassa E, Carrau F, Rivas-Gonzalo JC, Escribano-Bailón MT (2011) Characterisation and evolution of grape polyphenol profiles of *Vitis vinifera* L. cv. Tannat during ripening and vinification. Aust J Grape Wine Res 17: 383–393

Bontpart T, Marlin T, Vialet S, Guiraud J-L, Pinasseau L, Meudec E et al (2016) Two shikimate dehydrogenases, VvSDH3 and VvSDH4, are involved in gallic acid biosynthesis in grapevine. J Exp Bot 67:3537–3550

Boulekbache-Makhlouf L, Meudec E, Chibane M, Mazauric JP, Slimani S, Henry M et al (2010) Analysis by high-performance liquid chromatography diode array detection mass spectrometry of phenolic compounds in fruit of *Eucalyptus globulus* cultivated in Algeria. J Agric Food Chem 58: 12615–12624

Burnham KP, Anderson DR (2004) Multimodel inference understanding AIC and BIC in model selection. Sociol Meth Res 33: 261-304

Carrau F, Boido E, Gaggero C, Medina K, Disegna E, Dellacassa E (2011) *Vitis vinifera* Tannat, chemical characterization and functional properties. Ten years of research. In: Research Signpost/Transworld Res. Network, Kerala, India. Editor. R. Filip.

Chapter 4. (Org.). Multidisciplinary approaches on Food Science and Nutrition for the XXI Century. pp. 53-71.

Carrau FM (1997) The emergence of a new Uruguayan wine industry. J Wine Res 8: 179–185

Carrier G1, Huang YF, Le Cunff L, Fournier-Level A, Vialet S, Souquet JM et al (2013) Selection of candidate genes for grape proanthocyanidin pathway by an integrative approach. Plant Physiol Biochem 72:87-95.

Corder R, Mullen W, Khan NQ, Marks SC, Wood EG, Carrier MJ et al (2006) Oenology: red wine procyanidins and vascular health. Nature 444: 566.

Crozier A, Jaganath IB, Clifford MN (2009) Dietary phenolics: chemistry, bioavailability and effects on health. Nat Prod Rep 26:1001–1043

Da Silva C, Zamperin G, Ferrarini A, Minio A, Dal Molin A, Venturini L et al (2013) The high polyphenol content of *Vitis vinifera* cv. Tannat berries is conferred mostly by genes that are not shared with the reference genome. Plant Cell 25:4777-4788

Deluc L, Barrieu F, Marchive C, Lauvergeat V, Decendit A, Richard T et al (2006) Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. Plant Physiol 140:499-511

Dey A, Samanta MK, Gayen S, Sen SK, Maiti MK (2016) Enhanced Gene Expression Rather than Natural Polymorphism in Coding Sequence of the OsbZIP23 Determines Drought Tolerance and Yield Improvement in Rice Genotypes. PLoS One 11:e0150763

Derksen A, Hensel A, Hafezi W, Herrmann F, Schmidt TJ, Ehrhardt C et al (2014) 3-O-galloylated procyanidins from *Rumex acetosa* L. inhibit the attachment of influenza A virus. PLoS One. 9:e110089.

Dueñas M, Cueva C, Muñoz-González I, Jiménez-Girón A, Sánchez-Patán F, Santos-Buelga C, Moreno-Arribas MV, et al (2015) Studies on Modulation of Gut Microbiota by Wine Polyphenols: From Isolated Cultures to Omic Approaches. Antioxidants (Basel) 4:1-21

Donovan JL, Bell JR, Kasim-Karakas S, German JB, Walzem RL, Hansen RJ et al. (1999) Catechin is present as metabolites in human plasma after consumption of red wine. The J Nutrit 129: 1662-1668

Echeverry, C., Blasina, F., Arredondo, F., Ferreira, M., Abin-Carriquiry, J.A., Vasquez, L et al (2004) Cytoprotection by neutral fraction of tannat red wine against oxidative stress-induced cell death. J Agric Food Chem 52: 7395–7399

Fabre S, Pinaud N, Fouquet E, Pianet I (2010) Colloidal behavior of wine galloylated tannins. Compt Rend Chimie 13: 561–565.

Fitzpatrick DF, Hirschfield SL, Coffey RG. (1993) Endothelium-dependent vasorelaxing activity of wine and other grape products. Am J Physiol 265: H774–8

Hanlin RL, Kelm MA, Wilkinson KL, Downey MO (2011) Detailed characterization of proanthocyanidins in skin, seeds, and wine of Shiraz and Cabernet Sauvignon wine grapes (*Vitis vinifera*). J Agric Food Chem 59:13265-13276

Hashimoto M, Kim S, Eto M, Iijima K, Ako J, Yoshizumi M et al (2001) Effect of acute intake of red wine on flow-mediated vasodilatation of the brachial artery. Am J Cardiol 88: 1457–1460

Ikegami A, Eguchi S, Kitajima A, Inoue K, Yonemori K (2007) Identification of genes involved in proanthocyanidin biosynthesis of persimmon (Diospyros kaki) fruit. Plant Sci 72:1037-1047

Jiang X, Liu Y, Li W et al (2013) Tissue-specific, development-dependent phenolic compounds accumulation profile and gene expression pattern in tea plant [*Camellia sinensis*]. PLoS One 8, e62315.

Joshi T, Xu D (2007) Quantitative assessment of relationship between sequence similarity and function similarity BMC Genom 8:222

Kennedy DO (2014) Polyphenols and the Human Brain: Plant “Secondary Metabolite” Ecologic Roles and Endogenous Signaling Functions Drive Benefits Adv Nutr 5: 515–533

Khan N, Adhami VM, Mukhtar H (2010) Apoptosis by dietary agents for prevention and treatment of prostate cancer. Endocr-Relat Cancer 17: R39–R52.

Khan, N.Q., Patel, B., Kang, S.S., Dhariwal, S.K., Husain, F., Wood, E.G et al (2015) Regulation of vascular endothelial function by red wine procyanidins: Implications for cardiovascular health. Tetrahedron 71: 3059-3065

Khater F, Fournand D, Vialet S, Meudec E, Cheynier V, Terrier N (2012) Identification and functional characterization of cDNAs coding for hydroxybenzoate/hydroxycinnamate glucosyltransferases co-expressed with genes related to proanthocyanidin biosynthesis. J Exp Bot 63:1201–1214

Kim H-S, Quon MJ, Kim J-a (2014) New insights into the mechanisms of polyphenols beyond antioxidant properties; lessons from the green tea polyphenol, epigallocatechin 3-gallate. Redox Biol 2:187–195

Kim J, Buell CR (2015) A revolution in plant metabolism: Genome-enabled pathway discovery. Plant Physiol 169:1532-1539

Koyama K, Numata M, Nakajima I, Goto-Yamamoto N, Matsumura H, Tanaka N (2014) Functional characterization of a new grapevine MYB transcription factor and regulation of proanthocyanidin biosynthesis in grapes. J Exp Bot 65:4433-4449

Landete JM. 2011. Ellagitannins, ellagic acid and their derived metabolites: a review about source, metabolism, functions and health. Food Res Internat 44: 1150–1160

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H et al (2007) Clustal W and Clustal X version 2.0. Bioinform 23:2947-2948.

Lewandowska U, Fichna J, Gorlach S (2016) Enhancement of anticancer potential of polyphenols by covalent modifications. Biochem Pharmacol. 109:1-13.

Li C, Leverence R, Trombley JD, Xu S, Yang J, Tian Y et al (2010) High molecular weight persimmon (*Diospyros kaki* L.) proanthocyanidin: a highly galloylated, A-linked tannin with an unusual flavonol terminal unit, myricetin. J Agric Food Chem 58:9033-9042

Lin J-K, Weng M-S. (2006) Flavonoids as nutraceuticals. In *The science of flavonoids* (ed. E. Grotewold), Springer, Berlin.

Liu Y, Gao L, Liu L, Yang Q, Lu Z, Nie Z, Wang Y, Xia T (2012) Purification and Characterization of a Novel Galloyltransferase Involved in Catechin Galloylation in the Tea Plant (*Camellia sinensis*). J Biol Chem 287: 44406–44417

Lizarraga D, Lozano C, Briedé JJ, van Delft JH, Touriño S, Centelles JJ et al (2007) The importance of polymerization and galloylation for the antiproliferative properties of procyanidin-rich natural extracts. FEBS J 274: 4802–4811

Love MI, Huber W and Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genom Biol 15: 550

Minh BQ, Nguyen MA, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol 30: 1188–1195

Mittasch J, Böttcher C, Frolova N, Bönn M, Milkowski C (2014) Identification of UGT84A13 as a candidate enzyme for the first committed step of gallotannin biosynthesis in pedunculate oak (*Quercus robur*). Phytochem 99:44-51.

Muir RM, Ibáñez AM, Uratsu SL, Ingham ES, Leslie CA, McGranahan GH et al (2011) Mechanism of gallic acid biosynthesis in bacteria (*Escherichia coli)* and walnut (*Juglans regia*). Plant Mol Biol 75: 555–565

Namkung W, Thiagarajah JR, Phuan P-W and Verkman AS (2010) Inhibition of Ca^{2+}-activated Cl^- channels by gallotannins as a possible molecular basis for health benefits of red wine and green tea. FASEB J 24:4178-4186

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) Iq-tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32: 268-274.

Perazzolli M, Moretto M, Fontana P, Ferrarini A, Velasco R, Moser C et al (2012) Downy mildew resistance induced by *Trichoderma harzianum* T39 in susceptible grapevines partially mimics transcriptional changes of resistant genotype. BMC Genom 13:660.

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT–PCR. Nuc Acid Res 29:e45

Polesani M, Bortesi L, Ferrarini A, Zamboni A, Fasoli M, Zadra C et al (2010) General and species-specific transcriptional responses to downy mildew infection in a susceptible (Vitis vinifera) and a resistant (V. riparia) grapevine species. BMC Genom 11:117.

Quideau S, Deffieux D, Douat-Casassus C, Pouységu L (2011) Plant polyphenols: chemical properties, biological activities, and synthesis. Angew Chemie 50: 586–621.

Schwarz G. (1978) Estimating the dimension of a model. The Ann Statist 6: 461-464.

Soubrier J, Steel M, Lee MS, Der Sarkissian C, Guindon S, Ho SY et al (2012) The influence of rate heterogeneity among sites on the time dependence of molecular rates. Mol Biol Evol 29: 3345-3358.

Terrier N, Torregrosa L, Ageorges A, Vialet S, Verriès C, Cheynier V et al (2009) Ectopic Expression of VvMybPA2 Promotes Proanthocyanidin Biosynthesis in Grapevine and Suggests Additional Targets in the Pathway. Plant Physiol 149:1028–1041.

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25:1105–11.

Treutter D (2006) Significance of flavonoids in plant resistance: a review. Environ Chem Lett 4:147–57

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucl Acids Res 44 (W1): W232-W235.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM (2007) Primer3 Plus, an enhanced web interface to Primer3 Nuc Acid Res 35: W71-W74.

Vidal S, Francis L, Guyot S, Marnet N, Kwiatkowski M, Gawel R, et al (2003) The mouth-feel properties of grape and apple proanthocyanidins in a wine-like medium. J Sci Food Agric 83: 564–573

Weston LA, Mathesius U (2013) Flavonoids: their structure, biosynthesis and role in the rhizosphere, including allelopathy. J Chem Ecol 39:283–97

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18: 691-699

Yang Z (1995) A space-time process model for the evolution of DNA sequences. Genetics 139: 993-1005.

Witte, AV, Kerti L, Margulies DS, Flöel A (2014) Effects of resveratrol on memory performance, hippocampal functional connectivity, and glucose metabolism in healthy older adults. The J Neurosci 34: 7862-7870.

Zavala K, Opazo JC (2015) Lineage-Specific Expansion of the Chalcone Synthase Gene Family in Rosids. PLoS One. 10:e0133400.

Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, Malerba G, et al (2010) Characterization of Transcriptional Complexity during Berry Development in *Vitis vinifera* Using RNA-Seq. Plant Physiol 152:1787–1795

Legends to Figures

Figure 1 – Simplified pathway for the biosynthesis of galloylated procyanidins. ANR, anthocyanidin reductase; SDH, shikimate dehydrogenase; gGT, galloyl glucosyl transferase; ECGT, epicatechin galloyl transferase. Dotted lines indicate pathways with several steps.

Figure 2 – Phylogenetic analysis of putative epicatechin galloyl transferases in plants. Simplified maximum likelihood tree of the 97 proteins homologous to CsECGT found in plants using W-IQ-TREE. Bootstrap support values are next to the nodes. The 5 proteins present in Tannat are depicted in bold. Numbers in parentheses represent GenBank accessions as follows: (1) XP_011100017, XP_011100015, XP_011100013 , XP_011088733; (2) XP_012836842; (3) XP_010028347; (4) EEF40337, XP_015576609; (5) XP_010936155; (6) XP_017698400, XP_017698405, XP_017698402; (7) KDO81474, XP_006472316, XP_006433653; (8) XP_002510077; (9) XP_012068198; (10) XP_015573545; (11) XP_016557336, XP_016548234, XP_015578548; (12) XP_015578548; (13) XP_010647903, CBI22421, XP_002266354; (14) XP_010028345, XP_010028346, XP_010028341, XP_010028339; (15) XP_009343914, XP_009375674; (16) XP_009343181, XP_009343180; (17) XP_008338878, (18) XP_008236292, XP_007201001, XP_008236287, XP_008236286, XP_008236285, XP_016650585; (19) XP_011457834, (20) XP_008236294, XP_007201135, XP_007200784, XP_008236346; (21) XP_002298786; (22) XP_006486896, XP_006422777, XP_006422770, XP_006486895, XP_006422780; (23) XP_012093156, XP_012093225, XP_012093227, KDP44531; (24)XP_010028338; (25) XP_011457835; (26) XP_012093157; (27) OAY45348; (28) CBI22419, XP_002269439; (29) BAF56655; (30) XP_010656845,

CBI17847; (31) XP_004296446; (32) XP_012075109, XP_012075152; (33) XP_011003871, XP_011003870, XP_011045684, XP_011045683; (34) XP_006369794, XP_002300184; (35) XP_002266867, CAN67390, XP_002266615, CAN78674; (36) XP_010065819, XP_010065814; (37) XP_004289769, (38) XP_012093160; (39) XP_015576597, (40) XP_010065863; (41) XP_004289771, XP_004289770, XP_011457832; (42) EEF40332, XP_015576596, (43) XP_012093158; (44) XP_010691546; (45) KNA17255, (46) XP_012093161; (47) OAY39327, OAY39328; (48) XP_010028337; (49) XP_002272116; (50) BAH89272; (51) AIW39897. The CsECGT (described in patent CN103740740) falls in clade 51. Complete ML phylogeny is shown in Supplementary Figure 1S

Figure 3 – Expression of putative ECGTs, SDHs, gGTs, ANRs and selected MYBs in seeds and skins during Tannat berry development.
Expression levels are presented as relative rlog-transformed counts through berry development. colours show the amount by which each gene deviates in each timepoint-tissue from the gene average across all timepoints-tissues. Genes were clustered by Euclidean distance using rlog-transformed expression counts. Samples identification correspond to the weeks after flowering (waf) in which the sample was taken, the tissue analized (SD, seeds; sk, skins), and whether the grapes were green (g) or red (r) at veraison. A) VvECGT- like: VIT_03s0088g00050, VIT_03s0088g00260, VIT_03s0091g01240, VIT_03s0091g01290, VIT_11s0052g00790; B) VvMYB5a, VIT_08s0007g07230; VvMYB5b, VIT_06s0004g00570; VvMYBPA1, VIT_15s0046g00170; VvMYBPA2, VIT_11s0016g01320; VvMYBPAR, VIT_11s0016g01300; C) VvANR1, VIT_00s0361g00040; VvANR2, VIT_00s0361g00020; VvANR3, VIT_00s0361g00010; D) VvSDH1, VIT_05s0020g02030; VvSDH2, VIT_14s0030g00670; VvSDH3, VIT_14s0030g00660; VvSDH4, VIT_14s0030g00650; E) VvgGT1, VIT_03s0091g00040; VvgGT2, VIT_03s0180g00200; VvgGT3, VIT_03s0180g00320.

Figure 4 – Contents of galloylated procyanidins and their precursors in seeds and skins during Tannat berry development. Content of Galic acid, Glucogallin, Epicatechin and the sum of total galloylated compunds assesed by HPLC-MS. Data are expressed per g of fresh berry weight, and represent the mean of 2 replicates ± SD. Mean values with the same letter are not significantly different according to ANOVA test ($P < 0.05$). Samples

identification correspond to the weeks after flowering (waf) in which the sample was taken, the tissue analized (SD, seeds; sk, skins), and whether the grapes were green (g) or red (r) at veraison.

Figure 5 – Correlation of putative ECGTs expression with galloylated procyanidins contents.

Relative expression levels of VIT_03s0088g00260 (A, B) and VIT_03s0091g01290 (C, D) assesed by RT-qPCR and shown as Log 2 (Fold of Change) against 1 waf during Tannat berry development in seeds and skins. ND stands for non-detected. Sum of total galloylated compunds assesed by HPLC-MS are also expressed as Log 2 (Fold of Change) against 1 waf during Tannat berry development in seeds and skins (E, F). Values shown represent the mean of 3 (for RT-qPCR) or 2 (for HPLC-MS) replicates ± SD. Significance of differential expression against 1waf for each timepoint and tissue was examined by Student's t-test, and represented as one asterisk ($P<0.05$) or two asterisks ($P<0.01$).

Samples identification correspond to the weeks after flowering (waf) in which the sample was taken, the tissue analized (SD, seeds; sk, skins), and whether the grapes were green (g) or red (r) at veraison.

Supplementary material

Figure 1S – Detailed phylogenetic analysis of putative epicatechin galloyl transferases in plants.

Maximum likelihood tree of the 97 plant proteins homologous to CsECGT using W-IQ-TREE. GenBank accessions are shown at each branch tip. Support values written on the branches are SH-aLRT support (%) and ultrafast bootstrap support (% respectively). The 5 proteins present in Tannat are depicted in bold.

Figure 2S – Scatterplot comparing RNAseq and RT-qPCR data for the samples taken in 2012 (2011-2012 growing season). The data compared are the Log2 (Fold of Change) of each timepoint-tissue against 1waf. The Pearson´s correlation index (r) between methods is also shown.

Figure 3S - Comparison of RTqPCR data for VIT_03s0088g00260 in samples taken in 2012 (2011-2012 growing season) and 2015 (2014-2015 growing season).
The data are expressed as the Log2 (Fold of Change) of each timepoint-tissue agains 1waf. The Pearson´s correlation index (r) between years is also shown. Samples identification correspond to the weeks after flowering (waf) in which the sample was taken, the tissue analized (SD, seeds; sk, skins), and whether the grapes were green (g) or red (r) at veraison.

Table 1S – Primers used for RT-qPCR

Table 2S – Details of the Blastp result of CsECGT against NCBI NR (Non Redundant) database.
Description, Maximum score, Total score, Query Coverage, e-value Identiy and NCBI accessions of the 97 protein sequences used for the phylogenetic analysis.

Table 3S – Rlog-transformed counts of RNA-Seq data for putative ECGTs, SDHs, gGTs, ANRs and selected MYB genes in Tannat.
Samples identification correspond to the weeks after flowering (waf) in which the sample was taken, the tissue analized (SD, seeds; sk, skins), and whether the grapes were green (g) or red (r) at veraison.

Table 4S- Differential expression analysis for putative ECGTs, SDHs, gGTs, ANRs and selected MYB genes in Tannat.
L2dc, Log2 (Fold of Change); lfcSE; standard error of Log2 (Fold of Change); padj, adjusted pvalue; DE, Differentially expressed. Samples identification correspond to the weeks after flowering (waf) in which the sample was taken, the tissue analized (SD, seeds; sk, skins), and whether the grapes were green (g) or red (r) at veraison.

Table 5S - Complete metabolome analysis of Tannat seeds and skins during berry development
Samples identification correspond to the weeks after flowering (waf) in which the sample was taken, the tissue analized (SD, seeds; sk, skins), and whether the grapes were green (g) or red (r) at veraison.

Fig 1

Click here to download Figure renamed_984af.jpg

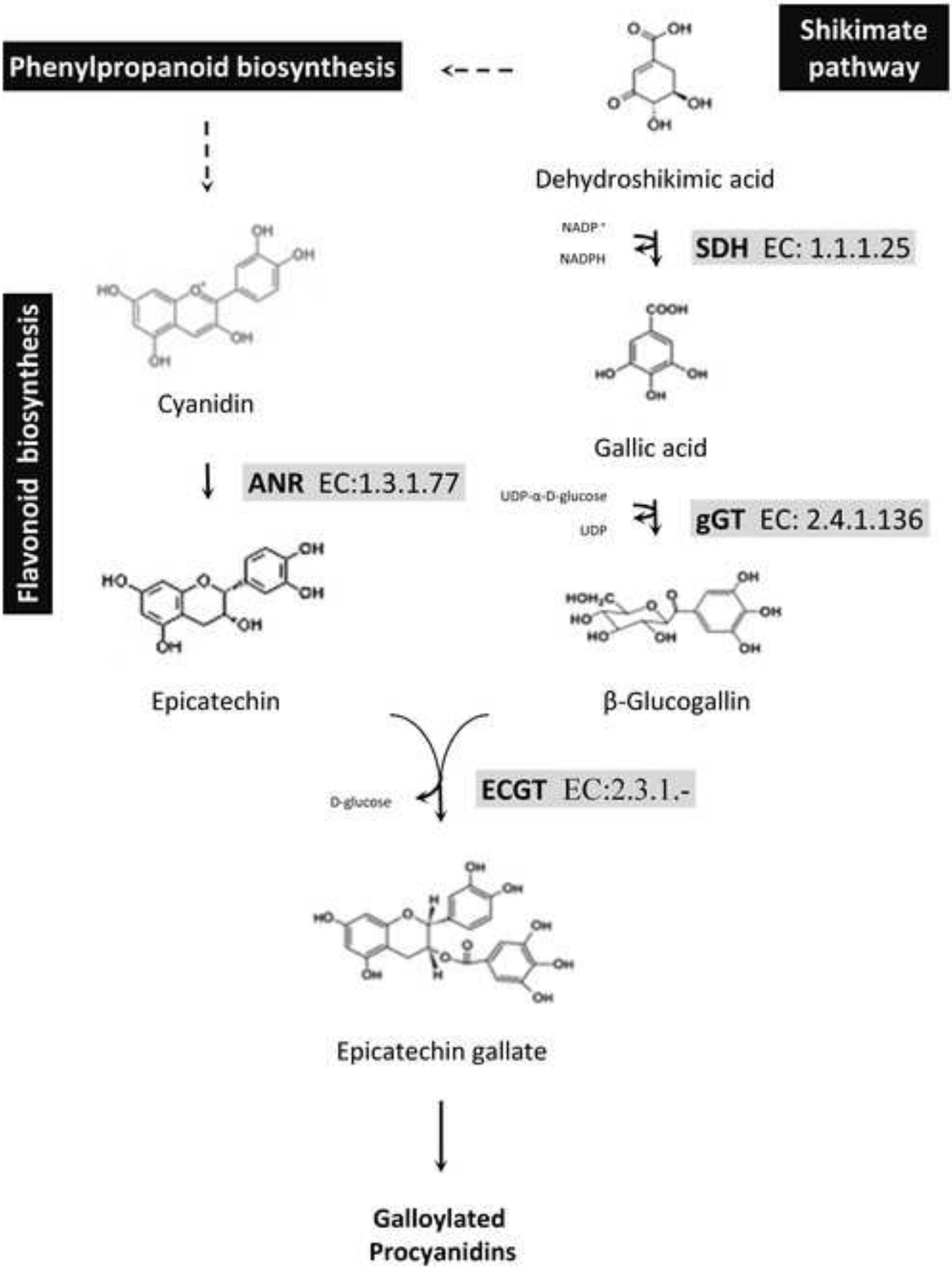


Fig 2

Click here to download Figure renamed_48603.jpg

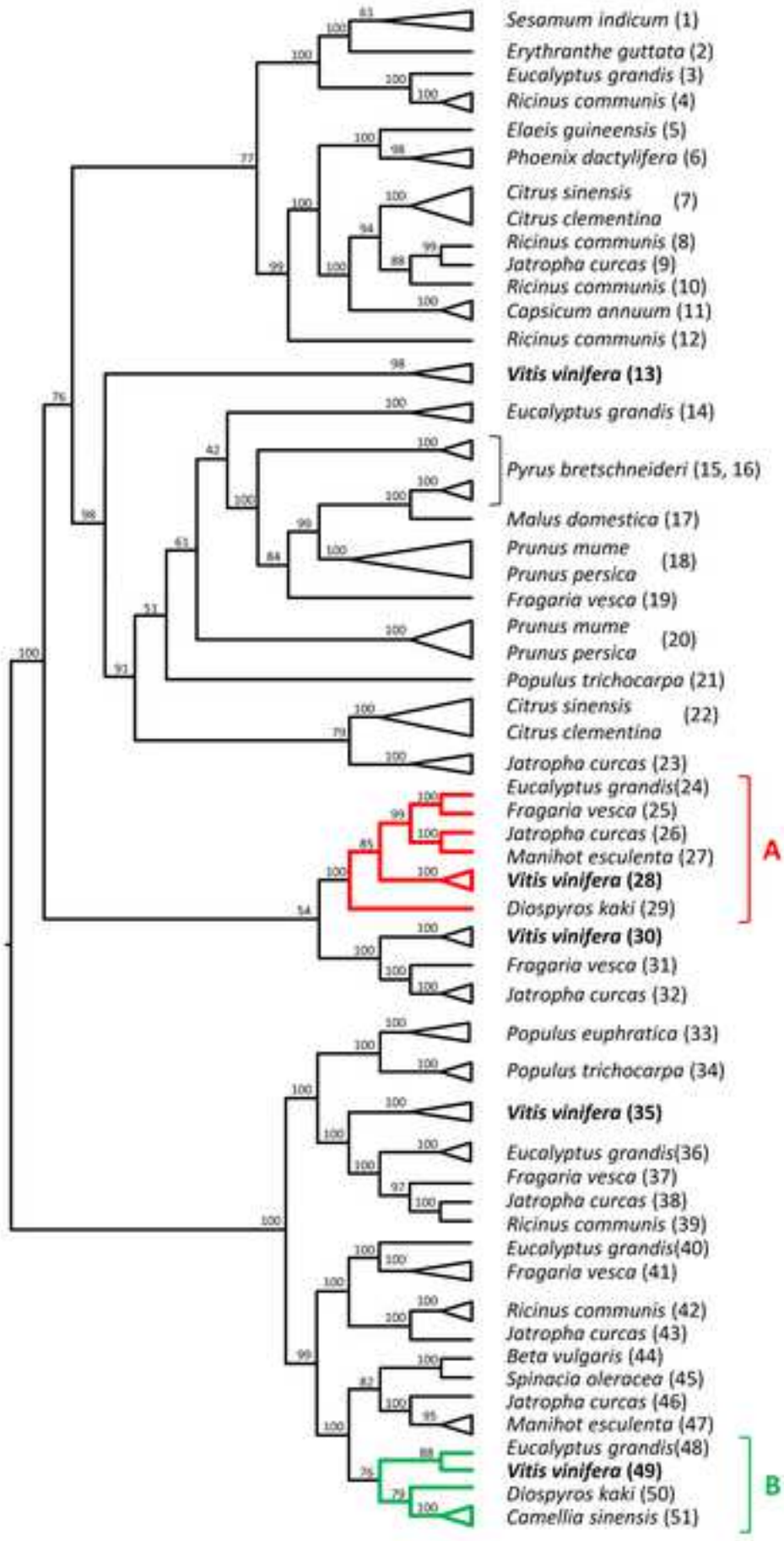


Fig 3

Click here to download Figure renamed_27cb5.jpg

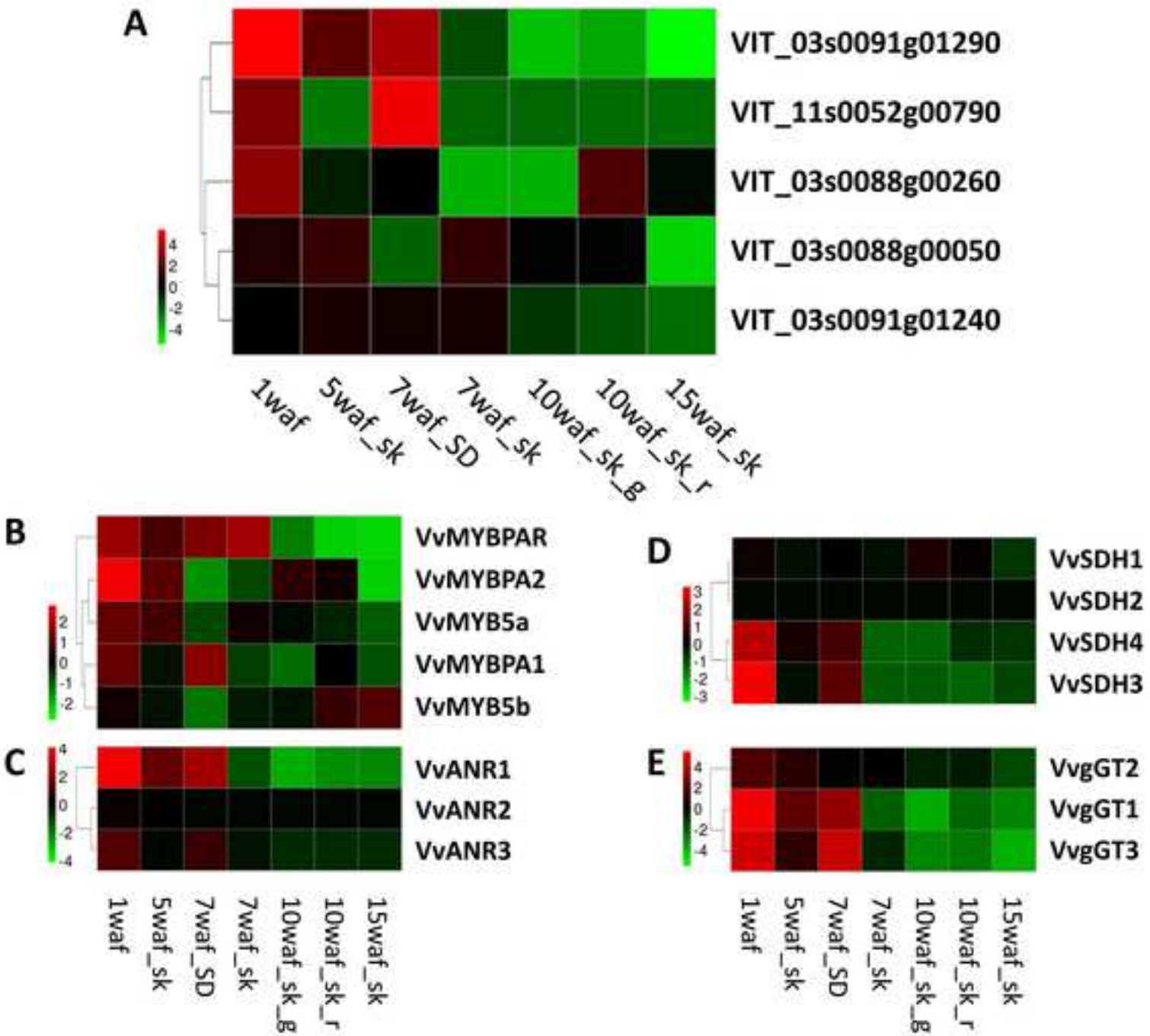


Fig 4

Click here to download Figure renamed_39e9b.jpg

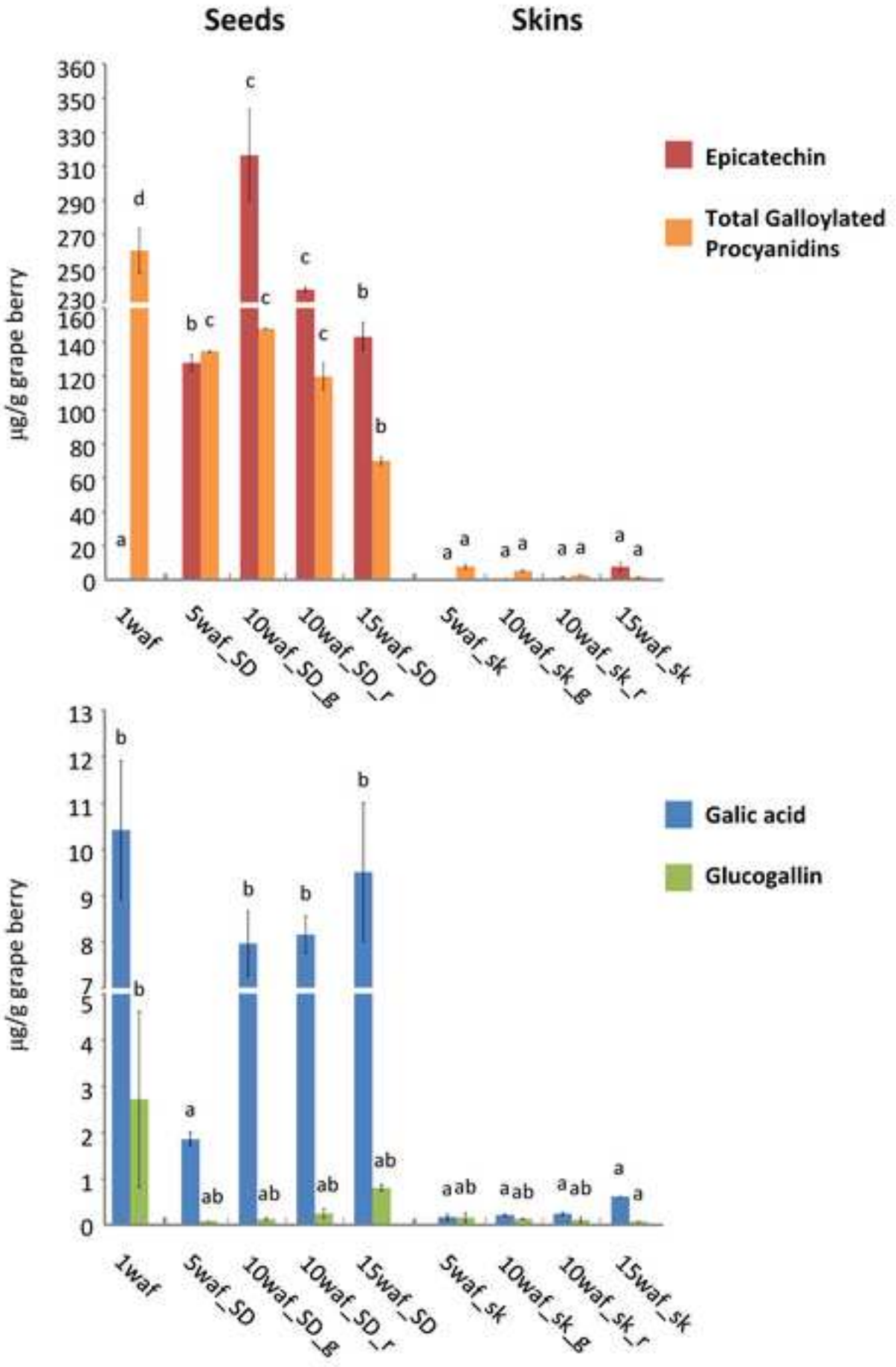


Fig 5

Click here to download Figure renamed_10b52.jpg

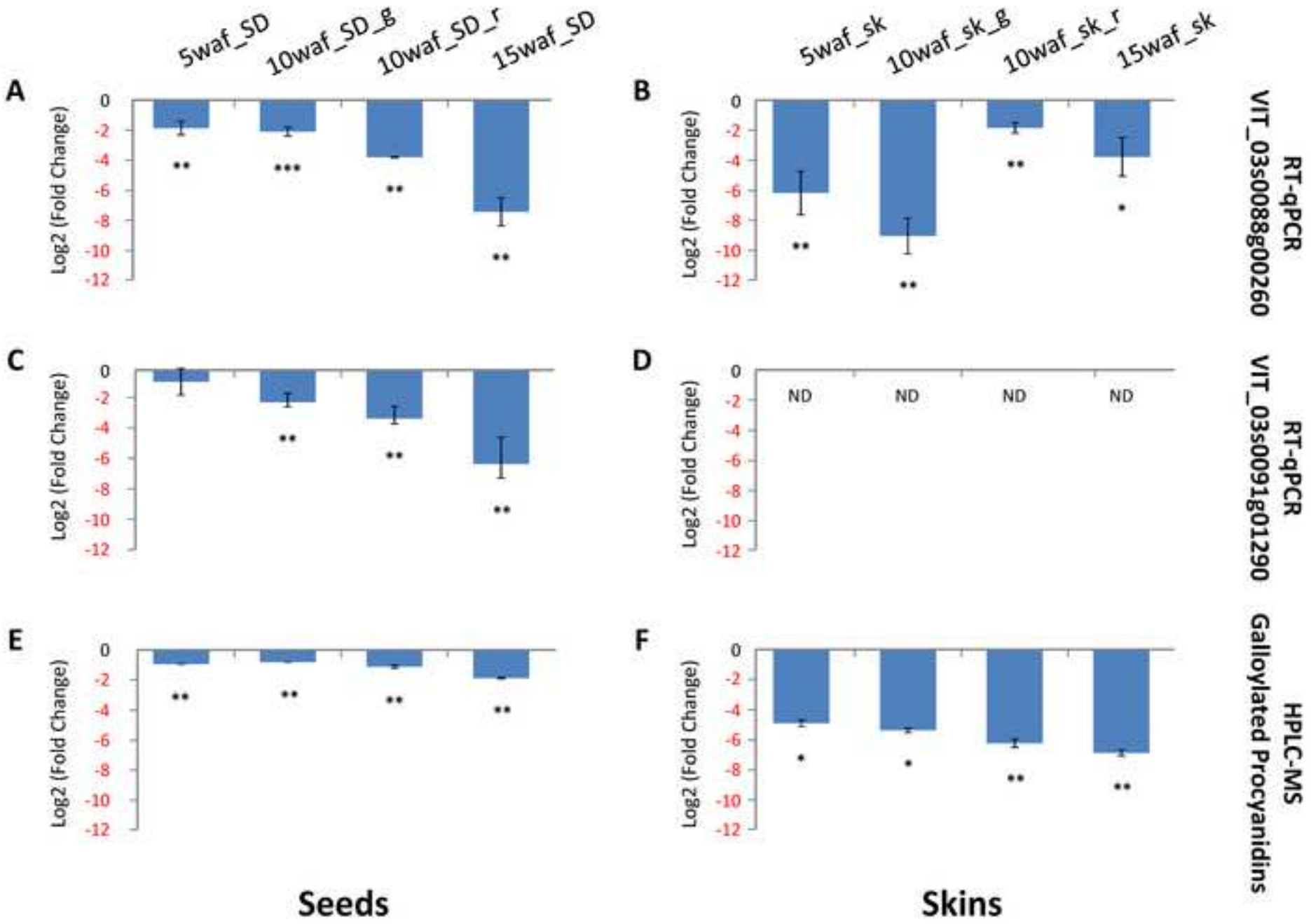


Fig 1S

Click here to download Supplementary Material (builds into PDF)
renamed_87495.jpg

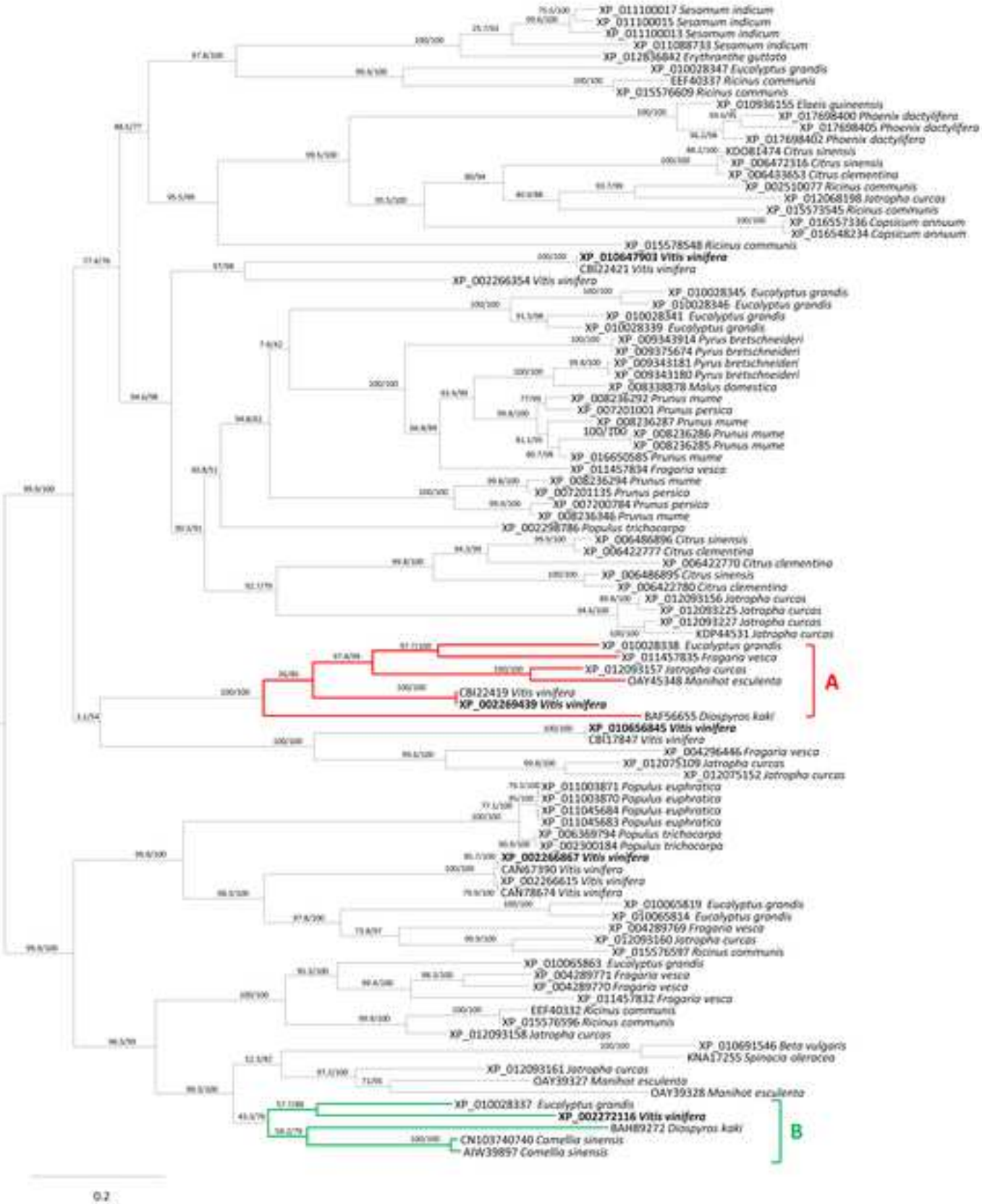


Fig 2S

Click here to download Supplementary Material (builds into PDF)
renamed_d7381.jpg

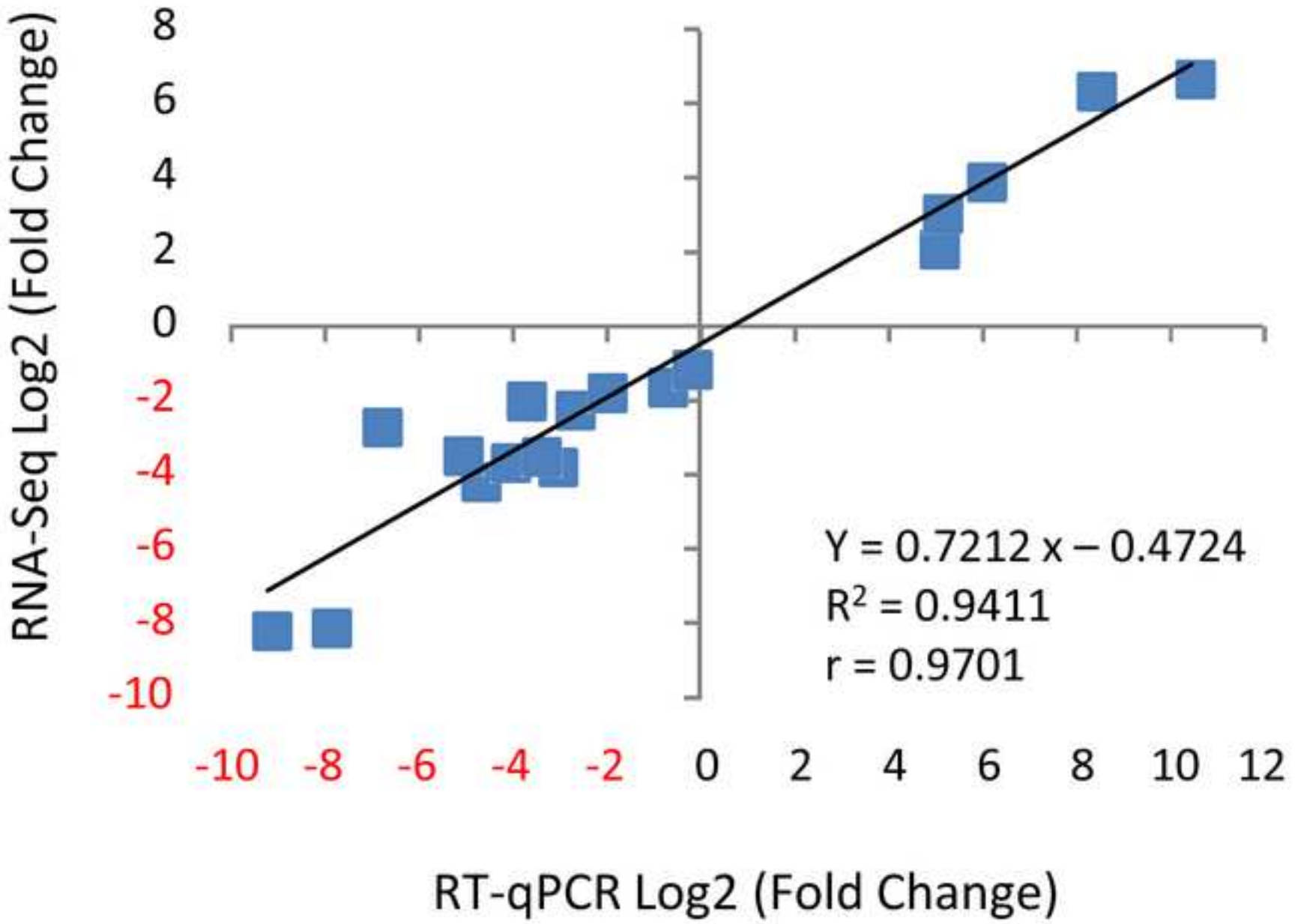
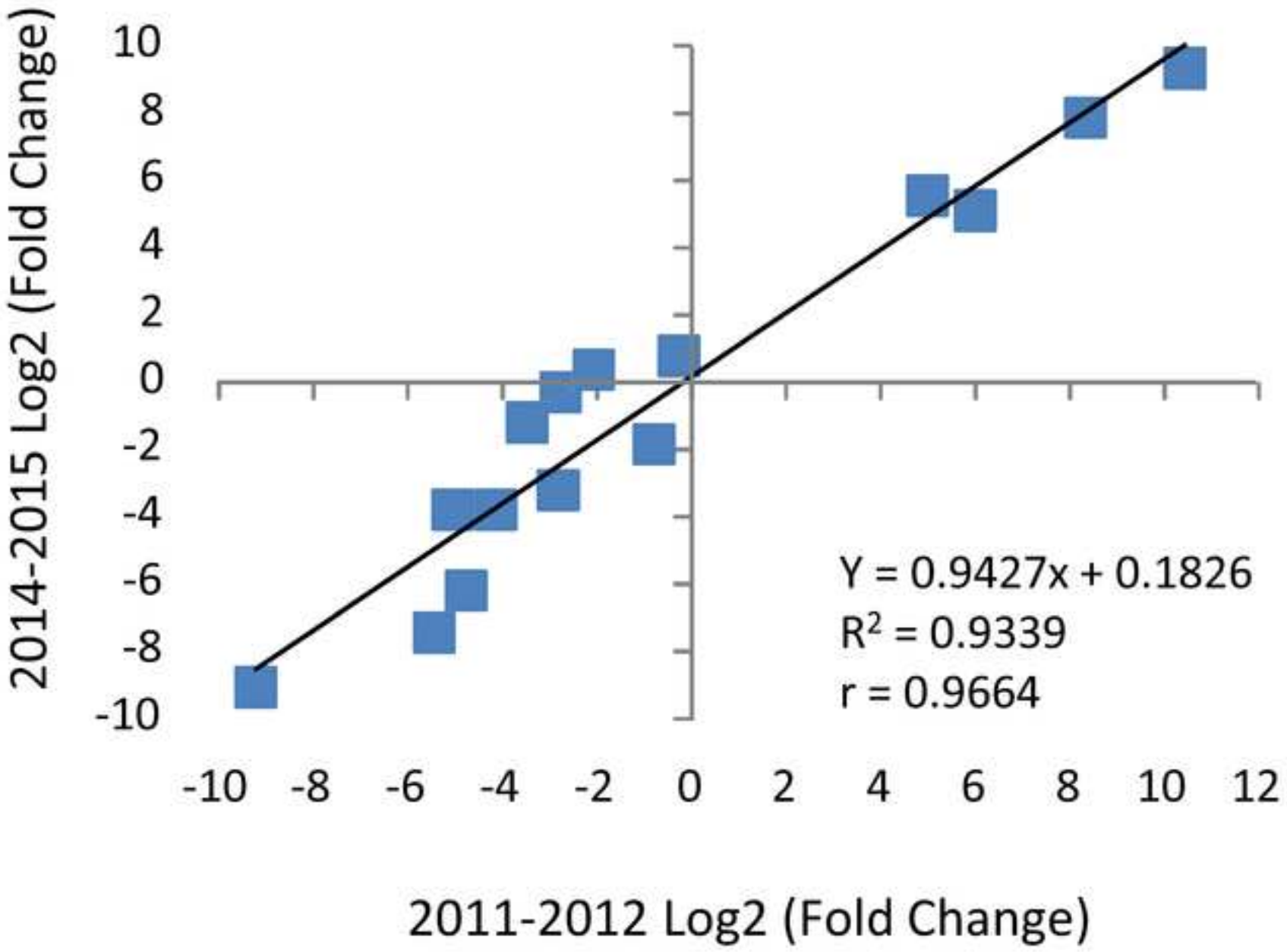


Fig 3S

Click here to download Supplementary Material (builds into PDF) Fig3S-Scatterplot comparing RT-qPCR data for samples taken in 2012 and 2015.jpg



Click here to download Supplementary Material (builds into PDF)
SuplementaryTables copy.xlsx

Table 1S – Primers used for the RT-qPCR

Gene name	Gene Annotation†	Sequence for the forward and reverse primers (5'->3')	Tm	Expected fragment size (bp)
VvACT7	VIT_04s0044g00580	F CTCCTTCGCCTCCGATTCTC	59.97	181
		R CTTGGCCGACCCACAATACT	60.04	
VvECGT-like	VIT_03s0088g00260	F TGGTCGAGGTGCAACAAGAG	60.25	137
		R CCCGGGTATGGAACAACCAA	59.96	
	VIT_03s0091g01290	F TTGCCCCAATCCTCATTAAG	55.36	92
		R TGGATTCCCGAGCAAGTATC	57.09	
VvF3'5'H	VVN_03608	F AGCTCTGCTCTACACTCTGTTTT	59.68	167
		R GCCGTGTGCTCCTCTATCATC	60.60	
VvF3'H	VVV_00903	F CCCAAGTTTTCGGGAAGTAA	55.57	168
		R CCCCTTGAGAATCATCGTACAG	58.28	

† According to Tannat annotation (Da Silva et al. 2013)

Table 2S – Details of the Blastp result of CsECGT agiains NCBI NR (Non Redundant) database.

Description	Max score	Total score	Query cover	E value	Ident	Accession
serine carboxypeptidase-like protein [Camellia sinensis]	974	974	100%	0.00E+00	98%	AIW39897
PREDICTED: serine carboxypeptidase-like 18 [Eucalyptus grandis]	706	706	92%	0.00E+00	75%	XP_010028337
PREDICTED: serine carboxypeptidase-like 18 [Jatropha curcas]	671	671	93%	0.00E+00	70%	XP_012093161
putative serine carboxypeptidase-like acyltransferase [Diospyros kaki]	652	652	93%	0.00E+00	68%	BAH89272
PREDICTED: serine carboxypeptidase-like 16 [Vitis vinifera]	639	639	93%	0.00E+00	67%	XP_002272116
hypothetical protein MANES_10G085800 [Manihot esculenta]	667	667	95%	0.00E+00	66%	OAY39327
hypothetical protein MANES_10G085900 [Manihot esculenta]	610	610	94%	0.00E+00	62%	OAY39328
PREDICTED: serine carboxypeptidase-like 18 [Jatropha curcas]	602	602	95%	0.00E+00	61%	XP_012093158
PREDICTED: serine carboxypeptidase-like 18 [Ricinus communis]	585	585	95%	0.00E+00	60%	XP_015576596
PREDICTED: serine carboxypeptidase-like 18 [Eucalyptus grandis]	587	587	96%	0.00E+00	59%	XP_010065863
hypothetical protein SOVF_081650 [Spinacia oleracea]	571	571	95%	0.00E+00	59%	KNA17255
PREDICTED: serine carboxypeptidase-like 18 [Beta vulgaris subsp, vulgaris]	569	569	96%	0.00E+00	58%	XP_010691546
PREDICTED: serine carboxypeptidase-like 18 [Fragaria vesca subsp, vesca]	545	545	93%	0.00E+00	57%	XP_004289770
PREDICTED: serine carboxypeptidase-like 17 [Fragaria vesca subsp, vesca]	553	553	93%	0.00E+00	57%	XP_011457832
PREDICTED: serine carboxypeptidase-like 17 [Fragaria vesca subsp, vesca]	548	548	98%	0.00E+00	56%	XP_004289771
PREDICTED: serine carboxypeptidase-like 17 [Vitis vinifera]	509	509	93%	2.00E-175	55%	XP_002266615
PREDICTED: serine carboxypeptidase-like 17 [Vitis vinifera]	506	506	93%	1.00E-174	55%	XP_002266867
serine carboxypeptidase, putative [Ricinus communis]	545	545	95%	0.00E+00	54%	EEF40332
hypothetical protein VITISV_009248 [Vitis vinifera]	505	505	93%	5.00E-174	54%	CAN78674
PREDICTED: serine carboxypeptidase-like 7 [Jatropha curcas]	514	514	93%	1.00E-177	53%	XP_012093160
hypothetical protein VITISV_034715 [Vitis vinifera]	498	498	93%	7.00E-171	53%	CAN67390
hypothetical protein MANES_07G053200 [Manihot esculenta]	482	482	91%	9.00E-165	53%	OAY45348
PREDICTED: serine carboxypeptidase-like 18 [Fragaria vesca subsp, vesca]	455	455	91%	3.00E-154	53%	XP_004289769
hypothetical protein CICLE_v10028360mg [Citrus clementina]	469	469	95%	7.00E-160	52%	XP_006422777
PREDICTED: serine carboxypeptidase-like 18 [Citrus sinensis]	465	465	95%	5.00E-158	52%	XP_006486896
PREDICTED: serine carboxypeptidase-like 7 isoform X1 [Sesamum indicum]	487	487	91%	1.00E-166	52%	XP_011100015
PREDICTED: serine carboxypeptidase-like 18 [Jatropha curcas]	478	478	91%	4.00E-163	52%	XP_012075109
serine carboxypeptidase, putative [Ricinus communis]	477	477	91%	1.00E-162	52%	EEF40337
PREDICTED: serine carboxypeptidase-like 18 [Eucalyptus grandis]	466	466	91%	3.00E-158	52%	XP_010028347
PREDICTED: serine carboxypeptidase-like 18 [Ricinus communis]	493	493	96%	6.00E-169	51%	XP_015576609
PREDICTED: serine carboxypeptidase-like 7 [Vitis vinifera]	483	483	95%	9.00E-165	51%	XP_002269439
unnamed protein product [Vitis vinifera]	482	482	95%	1.00E-164	51%	CBI22419
PREDICTED: serine carboxypeptidase-like 2 isoform X2 [Prunus mume]	481	481	95%	2.00E-164	51%	XP_008236286
PREDICTED: serine carboxypeptidase-like 2 [Jatropha curcas]	480	480	95%	5.00E-164	51%	XP_012093157
PREDICTED: serine carboxypeptidase-like 1 isoform X1 [Pyrus x bretschneideri]	478	478	95%	3.00E-163	51%	XP_009343181
PREDICTED: serine carboxypeptidase-like 2 isoform X1 [Prunus mume]	477	477	95%	1.00E-162	51%	XP_008236285
PREDICTED: serine carboxypeptidase-like 18 [Jatropha curcas]	472	531	95%	2.00E-160	51%	XP_012093227
PREDICTED: serine carboxypeptidase-like 7 [Ricinus communis]	493	493	93%	3.00E-169	51%	XP_015576597
PREDICTED: serine carboxypeptidase-like 18 [Jatropha curcas]	478	478	93%	2.00E-163	51%	XP_012093156
PREDICTED: serine carboxypeptidase-like 18 [Jatropha curcas]	472	472	93%	4.00E-161	51%	XP_012093225
hypothetical protein POPTR_0001s31980g [Populus trichocarpa]	492	492	92%	1.00E-168	51%	XP_006369794
hypothetical protein PRUPE_ppa016992mg [Prunus persica]	473	473	92%	1.00E-161	51%	XP_007201135
PREDICTED: serine carboxypeptidase-like 18 [Eucalyptus grandis]	484	484	91%	1.00E-165	51%	XP_010028339
PREDICTED: serine carboxypeptidase-like 13 isoform X1 [Sesamum indicum]	479	479	91%	2.00E-163	51%	XP_011100013
PREDICTED: serine carboxypeptidase-like 18 [Fragaria vesca subsp, vesca]	474	474	91%	2.00E-161	51%	XP_004296446
hypothetical protein CISIN_1g012236mg [Citrus sinensis]	472	472	91%	7.00E-161	51%	KDO81474
PREDICTED: serine carboxypeptidase-like 18 isoform X1 [Eucalyptus grandis]	472	472	91%	8.00E-161	51%	XP_010028341
PREDICTED: serine carboxypeptidase-like 17 isoform X2 [Citrus sinensis]	471	471	91%	1.00E-160	51%	XP_006472316
hypothetical protein CICLE_v10001059mg [Citrus clementina]	469	469	91%	8.00E-160	51%	XP_006433653
PREDICTED: serine carboxypeptidase-like 13 [Jatropha curcas]	469	469	91%	2.00E-159	51%	XP_012075152
PREDICTED: serine carboxypeptidase-like 1 [Pyrus x bretschneideri]	467	467	91%	8.00E-159	51%	XP_009343180
PREDICTED: serine carboxypeptidase-like 18 [Citrus sinensis]	461	461	91%	2.00E-156	51%	XP_006486895
PREDICTED: serine carboxypeptidase-like 18 isoform X2 [Phoenix dactylifera]	454	454	91%	1.00E-153	51%	XP_017698402
PREDICTED: serine carboxypeptidase-like 1 [Phoenix dactylifera]	450	450	91%	4.00E-152	51%	XP_017698400
hypothetical protein CICLE_v10028349mg [Citrus clementina]	450	450	91%	5.00E-152	51%	XP_006422780
hypothetical protein JCGZ_16364 [Jatropha curcas]	461	461	90%	1.00E-155	51%	KDP44531
PREDICTED: serine carboxypeptidase-like 17 isoform X1 [Ricinus communis]	474	474	96%	3.00E-161	50%	XP_015578548
PREDICTED: serine carboxypeptidase-like 18 isoform X1 [Phoenix dactylifera]	464	464	96%	2.00E-157	50%	XP_017698405
PREDICTED: serine carboxypeptidase-like 17 isoform X1 [Ricinus communis]	480	480	95%	7.00E-164	50%	XP_002510077
PREDICTED: serine carboxypeptidase-like 13 isoform X2 [Prunus mume]	477	477	95%	1.00E-162	50%	XP_008236292
PREDICTED: serine carboxypeptidase-like 2 [Prunus mume]	471	471	95%	2.00E-160	50%	XP_016650585
PREDICTED: serine carboxypeptidase-like 17 [Jatropha curcas]	468	468	95%	2.00E-159	50%	XP_012068198
PREDICTED: serine carboxypeptidase-like 18 [Elaeis guineensis]	465	465	95%	4.00E-158	50%	XP_010936155
hypothetical protein PRUPE_ppa005249mg [Prunus persica]	465	465	95%	5.00E-158	50%	XP_007201001
PREDICTED: serine carboxypeptidase-like 1 [Malus domestica]	452	452	95%	5.00E-153	50%	XP_008338878
hypothetical protein CICLE_v10028371mg [Citrus clementina]	444	444	95%	4.00E-150	50%	XP_006422770
PREDICTED: serine carboxypeptidase-like 17 isoform X2 [Capsicum annuum]	443	443	95%	1.00E-149	50%	XP_016548234
PREDICTED: serine carboxypeptidase-like 17 isoform X1 [Capsicum annuum]	442	442	95%	3.00E-149	50%	XP_016557336
PREDICTED: serine carboxypeptidase-like 18 [Vitis vinifera]	478	478	94%	3.00E-163	50%	XP_002266354
PREDICTED: serine carboxypeptidase-like 18 [Eucalyptus grandis]	479	479	94%	3.00E-163	50%	XP_010065814
PREDICTED: serine carboxypeptidase-like 18 [Eucalyptus grandis]	475	475	94%	7.00E-162	50%	XP_010065819
PREDICTED: serine carboxypeptidase-like 18 [Fragaria vesca subsp, vesca]	473	473	94%	2.00E-161	50%	XP_011457834
PREDICTED: serine carboxypeptidase-like 13 [Vitis vinifera]	466	466	94%	1.00E-158	50%	XP_010647903
unnamed protein product [Vitis vinifera]	468	468	94%	2.00E-157	50%	CBI22421
hypothetical protein POPTR_0001s29870g [Populus trichocarpa]	472	472	93%	1.00E-160	50%	XP_002298786
PREDICTED: serine carboxypeptidase-like 13 [Eucalyptus grandis]	459	459	93%	2.00E-155	50%	XP_010028338
PREDICTED: serine carboxypeptidase-like 18 isoform X2 [Populus euphratica]	488	488	92%	5.00E-167	50%	XP_011003871
PREDICTED: serine carboxypeptidase-like 18 isoform X2 [Populus euphratica]	487	487	92%	1.00E-166	50%	XP_011045684
PREDICTED: serine carboxypeptidase-like 18 isoform X1 [Populus euphratica]	483	483	92%	3.00E-165	50%	XP_011003870
PREDICTED: serine carboxypeptidase-like 18 isoform X1 [Populus euphratica]	483	483	92%	4.00E-165	50%	XP_011045683
hypothetical protein POPTR_0001s31980g [Populus trichocarpa]	476	476	92%	1.00E-162	50%	XP_002300184
PREDICTED: serine carboxypeptidase-like 18 [Erythranthe guttata]	475	475	92%	5.00E-162	50%	XP_012836842
PREDICTED: serine carboxypeptidase-like 2 [Prunus mume]	473	473	92%	2.00E-161	50%	XP_008236294
PREDICTED: serine carboxypeptidase-like 18 [Prunus mume]	468	468	92%	2.00E-159	50%	XP_008236346
hypothetical protein PRUPE_ppa024053mg [Prunus persica]	460	460	92%	2.00E-156	50%	XP_007200784
PREDICTED: serine carboxypeptidase-like 13 [Prunus mume]	459	459	92%	1.00E-155	50%	XP_008236287
PREDICTED: serine carboxypeptidase-like 13 [Sesamum indicum]	478	478	91%	2.00E-163	50%	XP_011100017
PREDICTED: serine carboxypeptidase-like 7 [Fragaria vesca subsp, vesca]	471	471	91%	2.00E-160	50%	XP_011457835
PREDICTED: serine carboxypeptidase-like 18 [Eucalyptus grandis]	467	467	91%	6.00E-159	50%	XP_010028346
PREDICTED: serine carboxypeptidase-like 7 [Sesamum indicum]	468	468	91%	1.00E-158	50%	XP_011088733
PREDICTED: serine carboxypeptidase-like 18 [Eucalyptus grandis]	463	463	91%	2.00E-157	50%	XP_010028345

PREDICTED: serine carboxypeptidase-like 7 [Pyrus x bretschneideri]	458	458	91%	1.00E-155	50%	XP_009375674
PREDICTED: serine carboxypeptidase-like 7 [Pyrus x bretschneideri]	456	456	91%	1.00E-154	50%	XP_009343914
PREDICTED: serine carboxypeptidase-like 13 [Vitis vinifera]	454	454	91%	1.00E-153	50%	XP_010656845
unnamed protein product [Vitis vinifera]	454	454	91%	1.00E-152	50%	CBI17847
PREDICTED: serine carboxypeptidase-like 17 [Ricinus communis]	449	449	91%	4.00E-152	50%	XP_015573545
serine carboxypeptitase-like protein 1 [Diospyros kaki]	437	437	91%	3.00E-147	50%	BAF56655

Table 3S – Rlog-transformed counts of RNA-Seq data for putative ECGTs, SDHs, gGTs, ANRs and selected MYB genes in Tannat.

Name	Gene	1waf	5waf_sk	7waf_SD	7waf_sk	10waf_sk_g	10waf_sk_r	15waf_sk
VvANR1	VIT_00s0361g00040	14.3393	11.9596	12.9456	9.0600	7.5293	7.9966	8.1560
VvANR2	VIT_00s0361g00020	-1.5837	-1.7661	-1.8597	-1.8623	-1.8593	-1.8603	-1.8606
VvANR3	VIT_00s0361g00010	2.6737	1.2194	2.2216	1.0688	0.6454	0.6265	0.6200
VvECGT-like	VIT_03s0088g00050	8.6101	9.0279	6.0804	8.9319	7.9041	7.8983	3.8796
VvECGT-like	VIT_03s0088g00260	14.5017	11.1358	11.6992	8.3164	8.2857	13.1917	11.5394
VvECGT-like	VIT_03s0091g01240	6.0320	6.6102	6.3681	6.5314	5.0148	4.5027	3.9699
VvECGT-like	VIT_03s0091g01290	12.8143	9.6378	11.1481	6.4749	4.1997	4.6464	3.0259
VvECGT-like	VIT_11s0052g00790	4.7210	0.0044	6.9763	0.3929	0.3526	0.2742	0.2371
VvgGT1	VIT_03s0091g00040	11.9505	8.7460	9.8553	4.5940	2.9797	4.4260	3.7152
VvgGT2	VIT_03s0180g00200	12.7964	12.0240	10.9490	11.1527	10.3894	10.4307	9.4710
VvgGT3	VIT_03s0180g00320	12.4157	8.9199	12.2357	7.1633	4.9908	5.2985	4.1451
VvMYB5a	VIT_08s0007g07230	10.9223	10.5523	9.0593	10.0548	9.7090	9.4032	8.8252
VvMYB5b	VIT_06s0004g00570	10.3257	9.8986	8.8774	9.8507	9.9121	10.7082	11.0600
VvMYBPA1	VIT_15s0046g00170	12.2260	10.9367	12.6026	10.4354	9.8631	11.1355	10.2193
VvMYBPA2	VIT_11s0016g01320	6.9595	5.2460	2.6704	3.4560	4.7762	4.5034	1.9497
VvMYBPAR	VIT_11s0016g01300	6.9438	6.0255	6.6119	6.9780	3.8085	2.9431	2.7739
VvSDH1	VIT_05s0020g02030	11.3944	11.0561	11.2329	10.9866	11.6620	11.3408	10.4735
VvSDH2	VIT_14s0030g00670	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
VvSDH3	VIT_14s0030g00660	11.4630	8.0995	9.5081	7.0446	7.1079	6.9603	7.3575
VvSDH4	VIT_14s0030g00650	12.5720	10.4176	10.9956	8.8085	8.8082	9.5059	9.4827

Table 4S- Differential expression analysis for putative ECGTs, SDHs, gGTs, ANRs and selected MYB genes in Tannat.

		5waf_sk vs 1waf				7waf_SD vs 1waf				7waf_sk vs 1waf			
Name	Gene	l2FC	lfcSE	padj	DE	l2FC	lfcSE	padj	DE	l2FC	lfcSE	padj	DE
VvANR1	VIT_00s0361g00040	-2.7904	0.1237	1.56E-110	YES	-1.6222	0.1382	9.11E-31	YES	-6.4330	0.1447	0.00E+00	YES
VvANR2	VIT_00s0361g00020	-0.9862	1.3044	NA	NO	-1.7370	1.1815	NA	NO	-1.8001	1.2155	NA	NO
VvANR3	VIT_00s0361g00010	-2.8762	0.7796	6.57E-04	YES	-0.7902	0.8061	4.05E-01	NO	-3.3268	0.9199	8.25E-04	YES
VvECGT-like	VIT_03s0088g00050	0.5052	0.1245	1.61E-04	NO	-3.3049	0.2172	5.40E-51	YES	0.3910	0.1400	1.17E-02	NO
VvECGT-like	VIT_03s0088g00260	-4.0832	0.1193	1.44E-253	YES	-3.3717	0.1339	6.14E-138	YES	-8.0312	0.1529	0.00E+00	YES
VvECGT-like	VIT_03s0091g01240	0.7378	0.2437	6.00E-03	NO	0.4349	0.2858	1.79E-01	NO	0.6301	0.2737	4.19E-02	NO
VvECGT-like	VIT_03s0091g01290	-3.6708	0.0977	0.00E+00	YES	-1.9056	0.1085	1.41E-67	YES	-7.5908	0.1492	0.00E+00	YES
VvECGT-like	VIT_11s0052g00790	-7.4784	1.0454	5.88E-12	YES	1.5006	0.4842	3.82E-03	YES	-6.0929	0.9744	2.09E-09	YES
VvgGT1	VIT_03s0091g00040	-3.7132	0.2069	3.42E-70	YES	-2.3837	0.2310	4.76E-24	YES	-9.0056	0.3222	1.95E-169	YES
VvgGT2	VIT_03s0180g00200	-0.9471	0.2025	1.10E-05	NO	-2.1934	0.2273	3.55E-21	YES	-2.0100	0.2265	6.49E-18	YES
VvgGT3	VIT_03s0180g00320	-4.1046	0.2086	3.30E-84	YES	-0.1855	0.2312	5.04E-01	NO	-6.2937	0.2438	1.45E-144	YES
VvMYB5a	VIT_08s0007g07230	-0.4582	0.1203	4.20E-04	NO	-2.3565	0.1443	1.36E-58	YES	-1.0684	0.1357	2.51E-14	YES
VvMYB5b	VIT_06s0004g00570	-0.5429	0.1338	1.61E-04	NO	-1.8858	0.1602	5.71E-31	YES	-0.6024	0.1503	1.86E-04	NO
VvMYBPA1	VIT_15s0046g00170	-1.6070	0.1826	1.39E-17	YES	0.4464	0.2036	4.60E-02	NO	-2.2320	0.2048	1.60E-26	YES
VvMYBPA2	VIT_11s0016g01320	-2.1563	0.3838	9.35E-08	YES	-5.9381	0.7001	1.30E-16	YES	-4.6599	0.5073	4.00E-19	YES
VvMYBPAR	VIT_11s0016g01300	-1.1473	0.2663	5.69E-05	YES	-0.3350	0.3007	3.39E-01	NO	0.0327	0.2916	9.49E-01	NO
VvSDH1	VIT_05s0020g02030	-0.4284	0.0937	1.80E-05	NO	-0.2030	0.1061	8.50E-02	NO	-0.5170	0.1053	3.46E-06	NO
VvSDH2	VIT_14s0030g00670	NA	NA	NA	NO	NA	NA	NA	NO	NA	NA	NA	NO
VvSDH3	VIT_14s0030g00660	-4.0703	0.1820	1.07E-108	YES	-2.3206	0.2018	1.34E-29	YES	-5.4382	0.2166	6.72E-137	YES
VvSDH4	VIT_14s0030g00650	-2.6242	0.1739	6.86E-50	YES	-1.9070	0.1948	9.22E-22	YES	-4.7174	0.1993	1.09E-121	YES

		10waf_sk_g vs 1waf				10waf_sk_r vs 1waf				15waf_sk vs 1waf			
Name	Gene	l2FC	lfcSE	padj	DE	l2FC	lfcSE	padj	DE	l2FC	lfcSE	padj	DE
VvANR1	VIT_00s0361g00040	-8.6145	0.1619	0.00E+00	YES	-7.9048	0.1445	0.00E+00	YES	-7.6865	0.1404	0.00E+00	YES
VvANR2	VIT_00s0361g00020	-1.8459	1.2410	NA	NO	-1.8817	1.2497	NA	NO	-1.8946	1.2549	NA	NO
VvANR3	VIT_00s0361g00010	-5.3376	1.1451	7.86E-06	YES	-5.4783	1.1367	3.86E-06	YES	-5.5754	1.1307	1.77E-06	YES
VvECGT-like	VIT_03s0088g00050	-0.8741	0.1369	6.01E-10	NO	-0.8916	0.1347	1.37E-10	NO	-6.7738	0.4197	1.43E-57	YES
VvECGT-like	VIT_03s0088g00260	-8.0884	0.1472	0.00E+00	YES	-1.5363	0.1187	2.87E-37	YES	-3.5715	0.1195	2.79E-194	YES
VvECGT-like	VIT_03s0091g01240	-1.4169	0.3012	6.42E-06	YES	-1.9970	0.3154	8.58E-10	YES	-3.0704	0.3701	3.82E-16	YES
VvECGT-like	VIT_03s0091g01290	-10.9823	0.3783	4.43E-183	YES	-10.1729	0.2749	3.27E-297	YES	-13.8999	0.7676	3.78E-72	YES
VvECGT-like	VIT_11s0052g00790	-6.7671	1.0955	2.19E-09	YES	-6.9137	1.0854	6.78E-10	YES	-7.0193	1.0780	2.03E-10	YES
VvgGT1	VIT_03s0091g00040	-11.9453	0.6715	2.38E-69	YES	-9.2386	0.3247	4.88E-176	YES	-10.1475	0.3828	3.29E-153	YES
VvgGT2	VIT_03s0180g00200	-2.9993	0.2043	1.49E-47	YES	-2.9435	0.2039	4.40E-46	YES	-4.2179	0.2061	8.29E-92	YES
VvgGT3	VIT_03s0180g00320	-9.2420	0.3183	3.34E-183	YES	-8.8770	0.2852	1.52E-210	YES	-10.7428	0.4045	8.90E-154	YES
VvMYB5a	VIT_08s0007g07230	-1.5173	0.1243	3.00E-33	YES	-1.9044	0.1248	2.07E-51	YES	-2.6716	0.1280	1.75E-95	YES
VvMYB5b	VIT_06s0004g00570	-0.5205	0.1361	2.76E-04	NO	0.4965	0.1336	4.39E-04	NO	0.9159	0.1331	1.72E-11	NO
VvMYBPA1	VIT_15s0046g00170	-2.9967	0.1857	3.19E-57	YES	-1.3560	0.1830	5.49E-13	YES	-2.4917	0.1841	7.19E-41	YES
VvMYBPA2	VIT_11s0016g01320	-2.6473	0.4081	3.15E-10	YES	-2.9735	0.4099	1.69E-12	YES	-7.7724	0.8267	2.22E-20	YES
VvMYBPAR	VIT_11s0016g01300	-4.0771	0.3933	2.64E-24	YES	-5.7023	0.5315	5.78E-26	YES	-5.8867	0.5339	1.48E-27	YES
VvSDH1	VIT_05s0020g02030	0.3362	0.0941	7.12E-04	NO	-0.0581	0.0943	6.16E-01	NO	-1.1777	0.0960	8.62E-34	YES
VvSDH2	VIT_14s0030g00670	NA	NA	NA	NO	NA	NA	NA	NO	NA	NA	NA	NO
VvSDH3	VIT_14s0030g00660	-5.3625	0.2016	7.43E-154	YES	-5.4772	0.1995	5.42E-164	YES	-5.0329	0.1928	1.31E-148	YES
VvSDH4	VIT_14s0030g00650	-4.7143	0.1815	9.62E-147	YES	-3.7837	0.1768	5.02E-100	YES	-3.7539	0.1765	4.99E-99	YES

Table 5S - Complete metabolome analysis of Tannat seeds and skins during berry development Samples identification corresponds to the weeks after flowering the sample was taken (waf), and the tissue analized (SD, seeds; sk, skins)

µg/g grape berry	1waf		5waf_SD		10waf_sk_SD		10waf_sk_SD		15waf_SD	
(+)-catechin	104.5 ± 20.0	b	135.4 ± 6.7	bc	269.3 ± 22.8	d	215.5 ± 22.3	cd	109.9 ± 0.9	b
(-)-epicatechin	0.00 ± 0.00	a	127.9 ± 4.9	b	316.3 ± 27.5	c	237.4 ± 1.4	c	143.0 ± 8.3	b
(+)-gallocatechin	23.6 ± 1.9	c	0.7 ± 0.1	b	0.3 ± 0.1	b	0.3 ± 0.1	b	0.243 ± 0.002	b
(-)-epigallocatechin	0.00 ± 0.00	a	1.2 ± 0.1	b	3.6 ± 0.1	e	2.51 ± 0.03	d	1.92 ± 0.02	c
(-)-epicatechin gallate	78.7 ± 28.0	b	29.4 ± 2.5	ab	35.5 ± 7.1	ab	35.6 ± 3.4	ab	13.0 ± 3.0	a
∑ dimer galloylated	172.0 ± 12.4	d	94.4 ± 1.8	c	100.4 ± 8.3	c	75.6 ± 4.0	bc	51.4 ± 0.2	b
∑ trimer galloylated	8.9 ± 2.5	bc	9.1 ± 1.1	bc	10.7 ± 0.5	c	7.8 ± 0.5	bc	5.3 ± 0.4	b
∑ tetramer galloylated	0.60 ± 0.04	c	0.27 ± 0.04	bc	0.29 ± 0.01	bc	0.13 ± 0.04	ab	0.08 ± 0.02	a
∑ trimer digalloylated	0.00 ± 0.00	a	1.31 ± 0.02	c	0.92 ± 0.38	bc	0.60 ± 0.08	ab	0.22 ± 0.04	a
∑ galloylated	260.3 ± 13.1	d	134.5 ± 0.5	c	147.8 ± 0.3	c	119.7 ± 8.0	c	70.0 ± 2.4	b
galic acid	10.4 ± 1.5	b	1.86 ± 0.15	a	7.9 ± 0.7	b	8.1 ± 0.4	b	9.5 ± 1.5	b
glucogallin	2.7 ± 1.9	b	0.08 ± 0.00	ab	0.1 ± 0.0	ab	0.2 ± 0.1	ab	0.8 ± 0.1	ab
% galloylated	59% 7%		32% 1%		19% 0.1%		20% 0.2%		21% 1%	

µg/g grape berry	5waf_sk		10waf_sk_g		10waf_sk_r		15waf_sk	
(+)-catechin	7.2 ± 2.0	a	7.6 ± 1.8	a	4.7 ± 1.0	a	3.3 ± 0.4	a
(-)-epicatechin	0.00 ± 0.00	a	0.4 ± 0.6	a	1.0 ± 0.9	a	7.6 ± 2.6	a
(+)-gallocatechin	8.8 ± 3.3	b	6.6 ± 0.2	ab	4.9 ± 0.2	ab	3.9 ± 0.9	ab
(-)-epigallocatechin	0.00 ± 0.00	a	0.00 ± 0.00	a	0.00 ± 0.00	a	0.00 ± 0.00	a
(-)-epicatechin gallate	3.7 ± 1.1	a	1.9 ± 0.1	a	0.7 ± 0.4	a	0.5 ± 0.1	a
∑ dimer galloylated	4.0 ± 2.4	a	3.2 ± 0.5	a	1.8 ± 0.2	a	0.7 ± 0.2	a
∑ trimer galloylated	0.00 ± 0.00	a	0.10 ± 0.02	a	0.00 ± 0.00	a	0.00 ± 0.00	a
∑ tetramer galloylated	0.00 ± 0.00	a	0.00 ± 0.00	a	0.00 ± 0.00	a	0.00 ± 0.00	a
∑ trimer digalloylated	0.00 ± 0.00	a	0.00 ± 0.00	a	0.00 ± 0.00	a	0.00 ± 0.00	a
∑ galloylated	7.7 ± 1.3	a	5.3 ± 0.7	a	2.5 ± 0.7	a	1.2 ± 0.3	a
galic acid	0.17 ± 0.06	a	0.21 ± 0.02	a	0.24 ± 0.03	a	0.614 ± 0.002	a
glucogallin	0.16 ± 0.14	ab	0.13 ± 0.01	ab	0.11 ± 0.05	ab	0.075 ± 0.003	a
% galloylated	26% 1%		22% 4%		16% 2%		6% 0.4%	

Different letters represent significantly different samples (p<0,01)

CAPÍTULO III

Unraveling the genetic basis of the deep purple colour of Tannat berries

Cecilia Da Silva [1,2¶] , Eduardo Boido [1], Carina Gaggero [2], Massimo Delledonne [3], Francisco Carrau [1*]

[1] Area Enología y Biotecnologia de Fermentaciones, Facultad de Química, Universidad de la República, 11800 Montevideo, Uruguay.
[2] Departamento de Biología Molecular, Instituto de Investigaciones Biológicas Clemente Estable, 11600 Montevideo, Uruguay.
[3] Centro di Genomica Funzionale, Dipartimento di Biotecnologie, Universitá degli Studi di Verona, 37134 Verona, Italy.

* Corresponding author information: fcarrau@fq.edu.uy

¶ Present address: Cecilia Da Silva, Prof. Adj. Biología Molecular Vegetal, PDU Espacio de Biología Vegetal del Noreste, Universidad de la República sede Tacuarembó, Ruta 5 km 387, Tacuarembó, Uruguay

Running title: Genetic basis of Tannat purple hue

ABSTRACT

Vitis vinifera cv. Tannat produces wines with higher pigment contents and darker purple hue, when compared to most red grape varieties. Anthocyanins are polyphenol pigments which confer red, purple and blue hues, depending on the relative proportion of the anthocyanins produced by the action of flavonoid 3'5' hydroxylase (delphinidin-3-glucoside, petunidin-3-glucoside, malvidin-3-glucoside) or flavonoid 3' hydroxylase (cyanidin-3-glucoside, peonidin-3-glucoside). Anthocyanin biosynthesis in grape is mainly regulated by the transcription factors MYBA1, MYBA2 and MYBA3.
In this work, we have performed RNA-Seq analysis of skins during berry development, taking green and red berries at 50% veraison as separate samples, as an experimental strategy to focus on the differential expression of genes involved in anthocyanin biosynthesis.
Several novel (not previously annotated in Pinot noir reference genome) and varietal (absent in reference genome) genes showed differential expression associated to colour change during veraison. Expression of a novel F3'5'H gene and a varietal F3'H gene was confirmed by RT-qPCR in the same samples used for RNA-Seq and also in another growing season. Genetic variants in the MYB transcriptional regulators of Tannat were analyzed, finding that retrotransposon Gret1 is present in one of the alleles of VvMYBA1 in all the Tannat clones analyzed, and that not previously reported SNPs are present in VvMYBA2 and VvMYBA3.
This work establishes the specific fingerprint of expression for gene families (CHS, CHI, F3H, F3'H, F3'5'H, DFR, LDOX, 3GT, OMT, MYBA1, MYBA2, MYBA3) directly involved in anthocyanin biosynthesis and regulation during Tannat veraison.

KEYWORDS
Anthocyanins, Tannat, RNA-Seq, MYB variants, flavonoid-3'5'-hydroxylase, flavonoid-3'-hydroxylase

INTRODUCTION

Vitis vinifera cv Tannat is the Uruguayan flagship grape variety for the production and export of premium red wines. The international wine market is increasingly demanding for Tannat wines due to their intense colour, mouthfeel structure and aging potential. These full bodied, deeply coloured and firm-tannin wines are very expressive in the nose with red fruit notes and a long lasting final bouquet (Carrau, 1997; Carrau et al., 2011). Anthocyanins are flavonoids responsible for grape and wine colour. Compared to other grape varieties, Tannat wines possess high pigment contents, with a large proportion of anthocyanins (86%). Monoglucosides of anthocyanins are the most abundant compounds, while acetyl derivatives of anthocyanins are more abundant than p-coumaroyl derivatives (Alcalde-Eon et al., 2006; González-Neves et al., 2004). The high polyphenolic content of Tannat wines confers antioxidant activity that increases with aging (Echeverry et al., 2004; Echeverry et al., 2005). The presence of high levels of flavan-3-ols (condensed tannins) in Tannat contribute to the structure of wines and the colour stabilization during aging (Boido et al., 2006). The genome of Tannat has been sequenced with high coverage and it was demonstrated that many genes not previously annotated or absent in the reference genome are related to polyphenol biosynthesis (Da Silva et al., 2013).

Anthocyanin pigments are the glycosylated form of anthocyanidin precursors or aglycones. Depending on pH, co-pigmentation, metal cations, modifications of the backbone and cultivar, anthocyanins exhibit colours ranging from orange-red to purple. During the ripening process of grapes, these pigments accumulate in the skin (and more rarely, also in the flesh of a few varieties). Anthocyanidins are synthesized in the cytosol, glycosylated and transported to the vacuole for storage. In the skins of vinifera grapes there are five types of anthocyanins. The final colour will depend on the relative proportion of the anthocyanins produced by the action of flavonoid 3'5' hydroxylase (delphinidin-3-glucoside, petunidin-3-glucoside, malvidin-3-glucoside) or flavonoid 3' hydroxylase (cyanidin-3-glucoside, peonidin-3-glucoside), resulting in more purple or red hues, respectively (see Fig 1). In Tannat mature grapes, anthocyanin monoglucosides constitute 70% of total anthocyanins, while 92% of these monoglucosides correspond to delphinidin derivatives (Boido et al., 2011).

Flavonoids perform a plethora of biological functions in plants: pigments absorb visible light and attract pollinators for seed dispersal; flavones, flavonols and anthocyanins act as UV-B filters and can also complex with DNA and protect it from oxidative damage; flavonoids control physiological traits of plant development, play a protective role against abiotic stresses and also contribute to defence against biotic stresses. Flavonoids are synthesized from the condensation of *p*-coumaroyl-CoA with three malonyl-CoA molecules by a chalcone synthase (CHS), which in turn produces a flavanone by a chalcone isomerase (CHI) and dihydrokaempferol by a flavanone-3-hydroxylase (F3H). Different enzymes will compete for the same substrates to give rise to flavanols, isoflavonoids, flavonols, flavones and anthocyanidins (see schematic pathway in Fig 1). Flavonoids show different substitutions catalyzed by isomerases, reductases, hydroxylases, methylases, and glycosyltransferases leading to high structural diversity and different spatio-temporal compartmentalization. Cross-talk between the tightly interconnected anthocyanin and proanthocyanidin pathways or between flavonoid and lignin pathways has been shown in many plant species (Mouradov & Spangenberg, 2014).

Many secondary metabolites are induced in grape leaves and berries in response to ultraviolet radiation or excessive light. The protective mechanisms include attenuation of radiation by filtering and antioxidant activity due to the scavenging of free radicals. Regulators belonging to the basic Leucine Zipper Domain (bZIP) and R2R3-MYB transcription factor families activate different metabolic pathways in response to increased UV levels, temperature or soil water availability (Matus, 2016). Recent models indicate that R2R3-MYB, beta helix-loop-helix (bHLH) and tryptophan and aspartic acid rich repeat (WDR) proteins form a MYB–bHLH–WDR (MBW) transcription complex that activates the genes encoding enzymes specifically involved in the late steps of the biosynthesis of anthocyanins and condensed tannins (Hichri et al., 2011). In grape, VvMYBA1 and VvMYBA2 specifically regulate anthocyanin synthesis during ripening by the upregulation of VvUFGT (also referred to as 3GT, see Fig 1), which encodes the enzyme that catalyzes conversion of anthocyanidin in anthocyanin. In white grape varieties, the absence of anthocyanin production is caused by a

mutation in the coding sequence of the VvMYBA2 gene together with a transposon insertion in the promoter of the VvMYBA1 gene (Kobayashi et al., 2004; Walker et al., 2007). Pink varieties have an insertion in an intron of VvMYBA1 (Shimazaki et al., 2011). A quantitative genetic analysis of anthocyanin variation in differently coloured grapes, ranging from white to the darkest purple, concludes that the continuous variation in anthocyanin content can be explained mainly by the actions of a single gene cluster of three regulatory genes: VvMYBA1, VvMYBA2 and VvMYBA3 (Fournier-Level et al., 2009).

Anthocyanin consumption is associated with inhibition of various forms of cancer, cardiovascular disease, neurodegenerative disease, and diabetes/metabolic syndrome. The health-promoting effects of anthocyanins are not only due to their antioxidant and anti-inflammatory properties but rather to their influence in modulating signal transduction pathways, inducing cell cycle arrest and stimulating apoptosis or autophagy of cancer cells (Lila et al., 2016; Lin et al., 2016). In particular, delphinidin-3-glucoside has been shown to suppress breast carcinogenesis (Yang et al., 2016), and prevent prostate cancer (Jeong et al., 2016). Anthocyanins have been proposed as promising nutraceutical compounds to improve gastrointestinal diseases, in combination with conventional treatments (Yang et al., 2016). Consumption of anthocyanins that are present in many fruits can modulate the inflammatory process associated with the interaction between the host and the gut microbiota, increasing the levels of the beneficial *Bifidobacterium* strains (Morais et al., 2016).

Therefore, a thorough understanding of anthocyanin synthesis and regulation in grapes can impact grape breeding, wine production, grape consumption and grape-derived nutraceutical and dye preparations.

In this work, we have performed an RNA-Seq analysis of skins during berry development, from fruit set to maturity. When grape bunches were at 50% veraison, green and red berries were sampled and analyzed separately, as an experimental strategy to focus on the differential expression of genes involved in anthocyanin biosynthesis. Several novel (not previously annonated in Pinot noir reference genome) and varietal (absent in reference genome) showed differential expression associated to colour change during veraison. Genetic variants in the MYB transcriptional regulators of Tannat were also determined.

MATERIALS AND METHODS

Plant Material

For the expression analyses, grape berry samples from the French Tannat commercial clone 717 (ENTAV nomenclature:ENTAV1995) were all collected from a vineyard of 2000 plants in Melilla, Montevideo, Southern Uruguay. Sampling was done at 1, 5, 10 and 15 weeks after flowering (waf) for the 2011 to 2012 and the 2014 to 2015 growing seasons. Ten clusters were sampled randomly from 10 different plants, excluding the borders of the vineyard. Ten berries were randomly selected from each cluster and pooled with berries from the other plants on the same vineyard site, resulting in three independent pools of 30 berries for each developmental stage. In the 10 waf samples (50% veraison) totally green and totally red berries were separated and analyzed as different samples. Skins were separated from the berry and frozen in liquid nitrogen, except at 1 waf, where the whole berries were used because the different berry tissues were not fully developed. Samples were stored at –80°C until use.

The presence of transposon Gret1 in the promoter region of MYBA1 transcription factor was evaluated in the following Tannat clones: Uruguayan clones UY7, UY9, UY11, UY15, and French commercial clones 398, 475, 717 (ENTAV nomenclature, ENTAV 1995) (Gonzalez-Techera et al. 2004), as well as in the sequenced Pinot Noir clone PN40024 (Jaillon et al. 2007).

Nucleic Acid Extraction

Frozen berry tissues were ground to powder using a mortar under liquid nitrogen, and total RNA was extracted with the Spectrum Plant Total RNA kit (Sigma-Aldrich) according to the manufacturer's instructions, previously adding Plant RNA Isolation Aid (AMBION) for the removal of common plant

contaminants such as polyphenolics and polysaccharides. The RNA purity and quantity were determined using a Nanodrop 1000 spectrophotometer (Thermo Scientific). RNA integrity was determined using a Bioanalyzer 2100 (Agilent) with RNA 6000 Nano Kit I (Agilent).
DNAs from Tannat clones used in this study were extracted and used in a previous study. Details on sampling and protocol DNA extraction can be found in Gonzalez-Techera et al. (2004). DNA of PN40024 was provided by Didier Merdinoglu.

Library preparation and sequencing
Illumina RNA-Seq libraries for the 2011 to 2012 growing season samples were prepared from 2.5 µg of total RNA per sample according to the manufacturer's instructions. The RNA-Seq libraries were size-selected at 350 to 550 bp using the Pippin Prep DNA size selection system (Sage Science). Library quality was determined using the Agilent High Sensitivity DNA kit on the Agilent 2100 bioanalyzer, and the quantity was determined by quantitative PCR using the KAPA Library Quantification kit (KapaBiosystems). Libraries were then pooled in equimolar concentrations and sequenced with the TruSeq Sequencing by Synthesis Kit v3-HS and TruSeq Paired End Cluster Kit v3-cBot-HS (Illumina) using an Illumina HiSequation 1000 sequencer according to the manufacturer's instructions to generate 100-bp paired-end reads.
Transcriptome sequence data of samples from 1waf and 5waf can be found in the NCBI Sequence Read Archive under accession numbers SRR866531, SRR866540, SRR866544, SRR866568, SRR866569, SRR866570 (Da Silva et al. 2013); and of samples from 10waf and 15waf under accession numbers XXX (in process). Annotations can be downloaded from http://ddlab.sci.univr.it/files/Tannat/annotation.gtf.

Differential expression analyses
RNA-Seq reads from the 2011 to 2012 growing season were aligned against the Tannat genome with TopHat v2.0.4 (Trapnell et al. 2009). Mapping reads were counted with HTSeq 0.6.1 using the intersection-nonempty mode (Anders et al. 2015), and expression values were calculated using the R package DESeq2 with default parameters (Love et al. 2014) to assess which genes were differentially expressed among the developmental stages of skins (adjusted pvalue < 0.001, |Log2FC| ≥2). The differential expression values are expressed against its previous time point. A principal component analysis was conducted on distances between samples using the rlog-transformed values with the R package ggplot2 (Wickham 2009).
Based on the differential gene expression analysis among the four time points, the software STEM (Short Time series Expression Miner, Ernst and Bar-Joseph, 2006) was used in order to cluster gene expression patterns with default settings and the "Normalize data" option. The input data file was the mean rlog-transformed values of the three replicates of each sample obtained with DESeq2.
WEGO (Web Gene Ontology Annotation Plot, Ye et al. 2006) (http://wego.genomics.org.cn/cgi-bin/wego/index.pl) was used to perform the GO functional classification for each STEM cluster or profile. Area-proportional Venn diagrams were generated using BioVenn (Hulsen et al. 2008, http://www.cmbi.ru.nl/cdd/biovenn/index.php). Heat maps were produced with R bioconductor (http://cran.r-project.org/).

Anthocyanin biosynthesis in Tannat
In order to evaluate the global balance of the biosynthesis of anthocyanins, the ratio between the expression of genes coding for F3'5'H enzyme and F3'H enzyme was calculated at the time points where skins were colored (10waf_sk_R and 15waf_sk). For this calculation only those genes with expression patterns concordant with the biosynthesis of anthocyanins were used.
The presence of Gret1 was assessed by PCR amplification on the 7 different Tannat clones and the Pinot Noir clone PN40024. Kobayashi et al., 2004 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., 2004. 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.).
The polymorphisms in the transcription factors related to anthocyanin biosynthesis MYBA1, MYBA2 and MYBA3 between Tannat clone UY11 and Pinot Noir clone PN40024 were obtained from analysis done in Da Silva et al., 2013.

RT-qPCR
All RNA samples were treated with RNase-free DNase Set (QIAGEN), and no DNA contamination was detected by PCR amplification. cDNA was synthesized using SuperScript® III First-Strand Synthesis SuperMix for qRT-PCR (INVITROGEN) according to the manufacturer's protocol. To validate RNA-Seq results, RT-qPCR amplification of four genes was performed on the 2011 to 2012 growing season (Da Silva et al., 2016). In order to confirm the expression of two key enzymes in Tannat anthocyanin biosynthetic pathway in different growing seasons, amplification of the genes VVN_03608 (F3'5'H) and VVV_00903 (F3'H) was performed with samples from the 2011 to 2012 and 2014 to 2015 growing seasons. Primers used were designed by Primer3Plus (Untergasser et al. 2007) and are listed in Supplemental Material 1. Real Time PCRs were performed in a Corbett Rotor-gene RG-6000 (QIAGEN) using the EXPRESS SYBR® GreenER™ qPCR Supermix universal (INVITROGEN) according to the manufacturer's protocol. The thermocycler was programmed to run for 2 min at 50°C for UDG incubation, 2 min at 95°C, followed by 40 cycles of 15 s at 95°C, 60 s at 50°C or 55°C, depending on the primer pair, and the Melting Curve from 60°C to 95°C. Non-template controls were included for each primer pair. Specific annealing of the primer pairs used was controlled on the Melting Curve. Expression data were normalized against the grapevine ACT7 gene (VIT_04s0044g00580, Zenoni et al. 2010, Polesani et al. 2010, Perazzolli et al. 2012). The amplification efficiencies of the different primer combinations were analyzed and all were >90%. Relative expression was determined using the Pfaffl model (Pfaffl 2001), and the values were expressed relative to its previous time point in each growing season. Each data point is the mean value of two biological replicates for the 2011 to 2012 growing season, and three biological replicates for the 2014 to 2015 growing season. Students t-test was used to evaluate the statistical significance of the differences in values using the GraphPad Prism Software (La Jolla, CA, USA).

RESULTS

Differential expression analyses in skins during Tannat berry development

The transcriptome variation during Tannat berry development was analyzed with RNA-Seq data from skins at the following developmental stages: skins from 5 waf (5waf_sk); skins from green (G) berries at 10 waf when veraison was at 50% (10waf_sk_G); skins from red (R) berries at 10 waf (50% veraison) (10waf_sk_R) and skins from mature berries at 15 waf (15waf_sk). First, a principal component analysis (PCA) was performed on distances between samples in order to visualize the overall similarity between samples and to ensure that it was correct to treat the green and red skin samples at veraison as separated samples (see Fig. 2). Differential expression analyses comparing each developmental stage with its previous stage yielded a total of 6873 differentially expressed genes (DEG) (adjusted pvalue < 0.001, |Log2FC| ≥2) (see Fig. 3), out of which 12% (817) are novel genes and 5% (336) are varietal genes compared to the Pinot Noir clone PN40024 (Da Silva et al. 2013). We found the greatest number of differentially expressed genes in the comparison between red skins (15waf_sk vs 10waf_sk_R); and the fewest number of DEG in the veraison comparison (10waf_sk_R vs 10waf_sk_G), as expected, due to the fact that samples were taken the same day from the same plants. Summarized results of DEGs in skins for the three comparisons are shown in Fig. 3 and detailed results in Supplemental Material 1. Relative expression levels of 4 genes were confirmed by RT-qPCR in the same samples used for RNA-Seq analysis and as

expected, a high correlation (Pearson's correlation coefficient of 0.97) was found between these two methods (Da Silva et al., 2016).
A time series expression data analysis was carried out with genes differentially expressed in at least one of the comparisons aiming to group genes based on their expression pattern through berry development. This analysis resulted in twelve significant profiles, with 10 profiles grouped in 4 clusters (see Fig. 4). Genes in Cluster 1 (Profiles 1, 9, 12) are progressively down-regulated during skin development. Genes in Cluster 2 (Profiles 11, 26, 34) are down-regulated at and after veraison. In Cluster 3 (Profiles 18, 40) genes are mainly up-regulated, while in Cluster 4 (Profiles 2, 13) genes are down-regulated during veraison, with a moderate rise in expression after veraison. In Profile 14 genes are down-regulated during veraison, with an abrupt rise in expression after veraison, and in Profile 38 genes are up-regulated during veraison, with an abrupt decrease in expression after veraison. The most interesting expression patterns are the ones showing up-regulation or down-regulation only during veraison, being possible candidates for the onset of this process.
In order to globally describe functions of genes included in each cluster/profile we classified genes based on their Gene Ontology (GO) class (see Fig. 4), and found some GO classes more represented in certain clusters/profiles. In Profile 38, for example, we found higher percentage of genes involved in the process of reproduction (GO:0000003, GO:0022414; 11%, while the mean in the other clusters/profiles was 6%), genes involved in the growth of an anatomical structure or organism (GO:0010926, GO:0032502, GO:0040007; 11%, while the mean in the other clusters/profiles was 8,5%), genes related to accumulation of pigments (GO:0043473; 15,4%, while the mean in the other cluster/profiles was 11%), genes involved in transport of substances into, out of or within a cell, or between cells (GO:0005215; 15,38%, versus 3,94%, the mean in the other clusters/profiles), and genes coding enzymes (GO:0003824, 42,56% versus 36,43%, the mean in the other clusters/profiles). In Profile 14 we found higher percentage of genes involved in the biosynthesis of constituent macromolecules, assembly, and arrangement of constituent parts of a cellular component (GO:0044085, GO:0016043; 9,35%, while the mean in the other clusters/profiles was 7,36%). In Cluster 4 we found higher percentage of genes involved in the process that results in a change in state or activity (in terms of movement, secretion, enzyme production, gene expression, etc.) as a response to a stimulus (GO:0050896; 22,31%, the mean in the other clusters/profiles was 13,96%). Details on genes assigned to each cluster/profile and GO categories can be found in Supplemental Material 2.

Anthocyanin biosynthesis in Tannat berry skins

One of the factors involved in the Tannat characteristic purple hue is the relative proportion of anthocyanins resulting from the activity of the enzymes flavonoid-3'-hydroxylase (F3'H) and flavonoid-3'5'-hydroxylase (F3'5'H). Anthocyanins resulting from F3'H impart a red hue (cyanidin-3-glucoside, peonidin-3-glucoside) while those formed from the action of F3'5'H contribute a purple hue (delphinidin-3-glucoside, petunidin-3-glucoside, malvidin-3-glucoside) (see Fig 1). In order to determine which of the genes coding for these enzymes are specifically involved in the synthesis of anthocyanins, we clustered by Euclidean distance their rlog-transformed expression values in each time point analyzed, and added the samples from 1waf to rule out genes that could be active in early stages of berry development (see Fig. 5). Here we determined that ten genes encoding F3'5'H (VIT_06s0009g02810, VIT_06s0009g02830, VIT_06s0009g02840, VIT_06s0009g02860, VIT_06s0009g02880, VIT_06s0009g02920, VIT_06s0009g02970, VIT_06s0009g03010, VVN_03332, VVN_03608) and four genes encoding F3'H (VIT_17s0000g07210, VIT_17s0000g07200,

VIT_02s0109g00310, VVV_00903) are involved in the biosynthesis of anthocyanins in Tannat berry skins. The ratio of these F3’5’H/F3’H genes expression in coloured skins at 10waf and 15 waf (10waf_sk_R and 15waf_sk) is 2.70 in each time point, showing that the balance is biased almost three times toward pigments giving purple/blue hues (delphinidin and its derivatives).

The relative expression of a novel and a varietal Tannat gene coding for the enzymes F3’5’H (VVN_03608) and F3’H (VVV_00903) was analyzed in skin development during different growing seasons (2011-2012 and 2014-2015) aiming to determine if the expression pattern is maintained in different years. We found highly concordant expression of these two genes between the RNA-Seq and RT-qPCR results in the growing season 2011-2012, as well as between RT-qPCR results from the two different growing seasons (see Fig. 6). The only discrepancy was the relative expression of VVN_03602 in the comparison between green skins (10waf_sk_G vs 5waf_sk) during 2014-2015 growing season. However, this value should not be taken into account because it was not statistically significant. These RT-qPCR analyses not only validate the expression of these two genes along different years, but also validate the genes themselves due to the fact that these genes were found through *de novo* assembly of transcriptomic data (Da Silva et al. 2013).

Among all the DEGs, we found 101 gene coding for enzymes directly involved in the anthocyanin biosynthesis (CHS, CHI, F3H, F3’H, F3’5’H, DFR, LDOX, 3GT, OMT, MYBA1, MYBA2, MYBA3, see Fig. 1). We clustered by Euclidean distance their rlog-transformed expression values in each time point analyzed (see Fig. 7) in order to analyze their expression patterns during berry skin development. We found a group of genes with higher expression only in coloured skins. Within these genes we found the ten F3’5’H specifically involved in anthocyanin biosynthesis, the three MYBs (MYBA1:VIT_02s0033g00410, MYBA2:VIT_02s0033g00390, MYBA3:VIT_02s0033g00450), four OMT (VIT_01s0010g03490, VIT_01s0010g03510, VVN_03939, VVV_01014), two CHS (VIT_16s0022g01020, VIT_03s0038g01460), two 3GT (VIT_04s0023g01290, VIT_11s0052g01630), and one DFR (VVV_01192). There was also a group of genes with moderately high expression in coloured skins where we found the four F3’H specifically involved in anthocyanin biosynthesis, five CHS(VIT_05s0136g00260, VIT_14s0068g00920, VVN_03266, VVV_00520, VVV_01133), three DFR (VIT_18s0001g12800, VVV_00044, VVV_01528), two CHI (VIT_13s0067g02870, VIT_13s0067g03820), two CHI (VIT_13s0067g02870, VIT_13s0067g03820), one 3GT (VVN_03231), and one LDOX (VIT_02s0025g04720).

Focusing on red vs green skins at veraison, we found 1286 DEGs, of which 51 code for enzymes directly related to the biosynthesis of anthocyanins (enzymes shown in Fig.1) and also the transcription factors MYBA1, MYBA2 and MYBA3, known to regulate anthocyanin biosynthesis. Within the DEGs with greater increase in coloured skins there are 6 genes encoding for flavonoid-3'5'-hydroxylases (F3’5’H), 4 for R-linalool synthases (monoterpene synthases), and genes for chalcone synthase, epicatechin galloyl transferase (ECGT-like), galactinol synthase (oligosaccharides biosynthesis), glutathione s-transferase (GST, transport of anthocyanins into the vacuole), serine carboxypeptidase-like (SCPL, candidate genes that could code for enzymes responsible for anthocyanin decoration with acyl groups), and phenol beta-glucosyltransferase (involved in starch and sucrose metabolism). Details of DEG results are in Supplemental Material 1.

Genetic variants in the MYB transcriptional regulators of anthocyanin biosynthesis

In order to understand anthocyanin synthesis in Tannat we searched for genetic variants in MYB transcription factors. For the MYBA1 gene we assessed the presence or absence of the retrotransposon Gret1 in the promoter region and the presence of SNPs in promoter region, introns and exons. 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. 8). We did not find any SNP for MYBA1 gene between Tannat clone UY11 and Pinot Noir clone PN40024. For MYBA2 and MYBA3 we also assessed SNPs between UY11 and PN40024. In MYBA2 we found one SNP in the promoter region and another in the third exon which results in a synonymous change. In MYBA3 we found four SNPs in the promoter region (see SNPs in Table 1).

DISCUSSION

Vitis vinifera cv Tannat is a red variety with high content of total pigments, resulting in wines with an intense purple colour, mouthfeel structure and aging potential. Anthocyanin monoglucosides in Tannat wines constitute 45% of total pigment content (Alcalde-Eon et al., 2006). However, the pigment contents of Tannat grapes at harvest are much higher and the sum of anthocyanin monoglucosides constitute 70% of total pigment content. Within this fraction of anthocyanin monoglucosides, the three anthocyanins that confer a blue/purple hue (delphinidin-3-glucoside, petunidin-3-glucoside, malvidin-3-glucoside) constitute 92% of the total and delphinidin-3-glucoside the 21% (Boido et al., 2011).
In Tannat grapes at harvest there are 10 times more anthocyanins produced by F3'5'H than those produced by F3'H (Mattivi et al., 2006; Boido et al., 2011). The ratio of the expression levels of F3'5'H divided by F3'H genes expressed at veraison is around 3, implying either a differential enzyme activity toward the same dihydrokaempferol substrate and/or the possibility that specific vacuolar transporters might be shifting the reaction equilibrium towards the synthesis of delphinidin derivatives. In this work, we have determined which specific genes coding for F3'5'H and F3'H enzymes are responsible for anthocyanin biosynthesis during Tannat veraison. The experimental strategy of analyzing skins from green and red berries -taken from the same bunches on the same day- as different samples allowed us to focus on the colour change and minimize the effect of the environment. Gouthu et al, 2014, also performed a comparative microarray study of ripening berries separating green and red berries at veraison. Comparing the list of genes reported as differentially expressed between red soft and green soft berries by Gouthu et al, 2014, we found 23 downregulated and 32 upregulated genes in common between Tannat and Pinot noir, while 3 genes reported by Gouthu et al, 2014, as upregulated in Pinot noir, are downregulated in Tannat. Costantini et al., 2015, also performed a microarray analysis with RNA isolated from grape skins of the F1 progeny from the cross 'Syrah'×'Pinot Noir' and identified a list of candidate genes involved in the fine regulation of grape colour. Comparing our results with those of Costantini et al., 2015, we found 48 genes in common. The genes commonly expressed at veraison by Tannat, Pinot noir and Syrah include the MYB transcription factors, anthocyanin acyltransferases, serine carboxypeptidase-like proteins, coumaroyl-shikimate-3-hydroxylase, anthocyanin methyl transferases, F3H, F3'H, F3'5'H (see details in Supplemental Material 3). When analyzing in detail the F3'H and F3'5'H genes, we noticed that in the 12x version of the reference genome PN0024, the F3'5'H genes definition is wrongly annotated although the EC number is correct. Hopefully, the detailed analysis of different gene families will allow a future improvement in the annotation of a reliable grape pan-genome.
As it happens with many other genomes, in the Tannat genome today around 40% of the genes lack a definition, GO or EC number. Therefore, it is reasonable to find 35% genes without definition, GO or EC number within the differentially expressed genes between red and green berries at 50% veraison. The Tannat genome contains 11.6% of novel genes (not previously annonated in PN0024 reference genome) and 5.4% of varietal genes (absent from PN0024). When analyzing the differentially expressed genes at veraison, 13.5% are novel and 4.7% are varietal genes. It is noteworthy that 17% of the down-regulated genes at veraison are novel and several bHLH transcription factors are down-regulated.

Since the pathways for the biosynthesis of anthocyanins and proanthocyanidins are highly interconnected in grape, we focused on the specific transcription factors known to regulate anthocyanin biosynthesis: MYBA1, MYBA2 and MYBA3. All three transcription factors were induced 2-3 fold during Tannat veraison. We found that the retrotransposon Gret 1 is present in one of the alleles of VvMYBA1 but not the other in all the Tannat clones analyzed, 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.
The SNPs detected in VvMYBA2 and VvMYBA3 comparing Tannat vs PN0024 are different from the ones reported in the ensembl plant database (http://plants.ensembl.org/). All of them are located in the promoter region except one located within the coding region. Although the SNP in the exon of MYBA2 results in a synonymous aminoacid change, this nucleotide difference could have an impact in the protein folding and conformation, affecting the protein activity.
In this work we have presented the fingerprint of Tannat genes upregulated and downregulated at veraison. The transcription factors are the master regulators of the colour change that takes place at veraison. In fact, purple tomatoes with high levels of anthocyanins were successfully engineered by overexpressing two different snapdragon transcription factors, a bHLH-type and an R2R3 MYB-type, under the control of a fruit-specific E8 promoter (Butelli et al., 2008). Nowadays, with the advent of the powerful CRISPR/Cas methodology to edit genomes (Belhai et al., 2013) and the availability of methods to achieve grapevine somatic embryogenesis (Tapia et al., 2009), the Tannat genetic variants in MYB transcription factors could be evaluated to determine their influence on pigment content in berries and introduced into other grape varieties and/or other crops.

ACKNOWLEDGMENTS
We thank the support of UdelaR (projects: CSIC Group 656 to FC and CSIC INI2015 to CDS) and Pedeciba, Uruguay. We also thank Fondazione Cariverona (Completamento e attività del Centro di Genomica Funzionale Vegetale), Verona, Italy, for supporting this work. Cecilia Da Silva was awarded a PhD and travel fellowships from ANII, Uruguay.We thank Didier Merdinoglu for providing us with DNA of PN40024.

AUTHOR CONTRIBUTIONS
CDS performed research, prepared samples, analyzed data and wrote the article. EB reviewed the article. CG analyzed data, designed the research and wrote the article. MD reviewed the article. FC designed the research and reviewed the article.

REFERENCES

Alcalde-Eon C, Boido E, Carrau F, Dellacassa E, Rivas-Gonzalo JC. 2006. Pigment Profiles in Monovarietal Wines Produced in Uruguay. Am. J. Enol. Viticult. 57:449-459.

Anders S, Pyl PT, Huber W. 2015. HTSeq- A Python framework to work with high-throughput sequencing data. Bioinformatics, 31:166–9.

Belhai K, Chaparro-Garcia A, Kamoun S, Nekrasov V. 2013. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant methods 9: 39.

Boido E, Alcalde-Eon C, Carrau F, Dellacassa E, Rivas-Gonzalo JC. 2006. Aging effect of the pigment composition and color of Vitis vinifera L. Cv. Tannat wines. Contribution of the main pigment families to wine color. J Agric Food Chem. 54:6692-6704.

Boido E, García-Marino M, Dellacassa E, Carrau F, Rivas-Gonzalo JC, Escribano-Bailón MT. 2011. Characterisation and evolution of grape polyphenol profiles of *Vitis vinifera* L. cv. Tannat during ripening and vinification. Aust J Grape Wine R 17: 383–393.

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EG, Hall RD, Bovy AG, Luo J, Martin C. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat Biotechnol. 26:1301-1308.

Carrau F, Boido E, Gaggero C, Medina K, Disegna E, Dellacassa E. 2011. Vitis vinifera Tannat, chemical characterization and functional properties. Ten years of research. In: Research Signpost/Transworld Res. Network, Kerala, India. Editor. R. Filip. Chapter 4. (Org.). Multidisciplinary approaches on Food Science and Nutrition for the XXI Century. pp. 53-71.

Carrau FM. 1997. The emergence of a new Uruguayan wine industry. *J Wine Res* **8**: 179–185.

Costantini L, Malacarne G, Lorenzi S, Troggio M, Mattivi F, Moser C, Grando MS. 2015. New candidate genes for the fine regulation of the colour of grapes. J Exp Bot. 66:4427-4440.

Da Silva C, Zamperin G, Ferrarini A, Minio A, Dal Molin A, Venturini L, Buson G, Tononi P, Avanzato C, Zago E, Boido E, Dellacassa E, Gaggero C, Pezzotti M, Carrau F, Delledonne M. 2013. The high polyphenol content of *Vitis vinifera* cv. Tannat berries is conferred mostly by genes that are not shared with the reference genome. Plant Cell 25:4777-4788.

Da Silva C, Boido E, Gaggero C, Delledonne M, Carrau F. 2016. High levels of galloylated procyanidins in Tannat grape seeds correlate with the expression of a new epicatechin galloyl transferase-like gene. Appl Gen (submitted).

Echeverry C, Ferreira M, Reyes-Parada M, Abin-Carriquiry JA, Blasina F, González-Neves G, Dajas F. 2005. Changes in antioxidant capacity of Tannat red wines during early maturation. J. Food Eng. 69:147-154.

Echeverry C, Blasina F, Arredondo F, Ferreira M, Abin-Carriquiry JA, Vasquez L, Aspillaga AA, Diez MS, Leighton F, Dajas F. 2004. Cytoprotection by neutral fraction of tannat red wine against oxidative stress-induced cell death. J Agric Food Chem. 52:7395-9.

Ernst J, Bar-Joseph Z. 2006. STEM: a tool for the analysis of short time series gene expression data. BMC bioinformatics, 7(1), 1.

Fournier-Level A, Le Cunff L, Gomez C, Doligez A, Ageorges A, Roux C, Bertrand Y, Souquet J-M, Cheynier V, This P. 2009. Quantitative genetic bases of anthocyanin variation in grape (Vitis vinifera L. ssp. sativa) berry: a quantitative trait locus to quantitative trait nucleotide integrated study. Genetics 183:1127-1139.

González-Neves G, Charamelo D, Balado J, Barreiro L, Bochicchio R, Gatto G, Gil G, Tessore A, Carbonneau A, Moutounet M. 2004. Phenolic potential of Tannat, Cabernet-Sauvignon and Merlot grapes and their correspondence with wine composition. Anal. Chim. Acta 513:191-196.

González Techera A, Jubany S, Ponce de León I, Boido E, Dellacassa E, Carrau FM, Hinrichsen P, Gaggero C. 2004. Molecular diversity within clones of *Vitis vinifera* cv. Tannat. Vitis 43:179-185.

Gouthu S, O'Neil ST, Di Y, Ansarolia M, Megraw M, Deluc LG. 2014. A comparative study of ripening among berries of the grape cluster reveals an altered transcriptional programme and enhanced ripening rate in delayed berries. J Exp Bot. 65:5889-5902.

Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V. 2011. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. J. Exp. Bot. 62:2465–2483.

Hulsen T, de Vlieg J, Alkema W. 2008. BioVenn–a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC genomics, 9:488.

Jaillon O et al; French-Italian Public Consortium for Grapevine Genome Characterization. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463–467.

Jeong MH, Ko H, Jeon H, Sung GJ, Park SY, Jun WJ, Lee YH, Lee J, Lee SW, Yoon HG, Choi KC. 2016. Delphinidin induces apoptosis via cleaved HDAC3-mediated p53 acetylation and oligomerization in prostate cancer cells. Oncotarget. doi: 10.18632/oncotarget.10790. [Epub ahead of print]

Kobayashi S, Goto-Yamamoto N, Hirochika H. 2004. Retrotransposon-induced mutations in grape skin color. Science 304:982.

Lila MA, Burton-Freeman B, Grace M, Kalt W. 2016. Unraveling Anthocyanin Bioavailability for Human Health. Annu Rev Food Sci Technol. 7:375-93.

Lin BW, Gong CC, Song HF, Cui YY. 2016. Effects of anthocyanins on the prevention and treatment of cancer. Br J Pharmacol. 2016 Sep 20. doi: 10.1111/bph.13627. [Epub ahead of print]

Love MI, Huber W and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15: 550.

Mattivi F, Guzzon R, Vrhovsek U, Stefanini M, Velasco R. 2006. Metabolite profiling of grape: Flavonols and anthocyanins. J Agric Food Chem. 54:7692-7702.

Matus JT. 2016. Transcriptomic and Metabolomic Networks in the Grape Berry Illustrate That it Takes More Than Flavonoids to Fight Against Ultraviolet Radiation. Front Plant Sci. 7:1337.

Morais CA, de Rosso VV, Estadella D, Pisani LP. 2016. Anthocyanins as inflammatory modulators and the role of the gut microbiota. J Nutr Biochem. 33:1-7.

Mouradov A, Spangenberg G. 2014. Flavonoids: a metabolic network mediating plants adaptation to their real estate. Front Plant Sci. 5:620.

Perazzolli M, Moretto M, Fontana P, Ferrarini A, Velasco R, Moser C, Delledonne M, I Pertot. 2012. Downy mildew resistance induced by Trichoderma harzianum T39 in susceptible grapevines partially mimics transcriptional changes of resistant genotype. BMC Genomics. 13:660.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res 29(9):e45

Polesani M, Bortesi L, Ferrarini A, Zamboni A, Fasoli M, Zadra C, Lovato A, Pezzotti M, Delledonne M, Polverari A. 2010. General and species-specific transcriptional responses to downy mildew infection in a susceptible (Vitis vinifera) and a resistant(V. riparia) grapevine species. BMC Genomics. 11:117.

Shimazaki M, Fujita K, Kobayashi H, Suzuki S. 2011. Pink-Colored Grape Berry Is the Result of Short Insertion in Intron of Color Regulatory Gene. PloS ONE 6: e21308.

Tapia E, Sequeida A, Castro A, Montes C, Zamora P, López R, Acevedo F, Prieto H. 2009. Development of grapevine somatic embryogenesis using an air-lift bioreactor as an efficient tool in the generation of transgenic plants. J Biotechnol. 139:95-101.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25:1105–11.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. 2007. Primer3Plus, an enhanced web interface to Primer3 Nucleic Acids Research, 35: W71-W74.

Walker AR, Lee E, Bogs J, McDavid DA, Thomas MR, Robinson SP. 2007. White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant J. 49:772-85.

Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer Science & Business Media.

Yang N, Sampathkumar K, Loo SC. 2016. Recent advances in complementary and replacement therapy with nutraceuticals in combating gastrointestinal illnesses. Clin. Nutr. pii: S0261-5614(16)30219-9. doi: 10.1016/j.clnu.2016.08.020. [Epub ahead of print]

Yang X, Luo E, Liu X, Han B, Yu X, Peng X. 2016. Delphinidin-3-glucoside suppresses breast carcinogenesis by inactivating the Akt/HOTAIR signaling pathway. BMC Cancer. 16:423.

Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J. 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 34:W293-7.

Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, Malerba G, Bellin D, Pezzotti M, Delledonne M. 2010. Characterization of Transcriptional Complexity during Berry Development in Vitis vinifera Using RNA-Seq. Plant Physiology. 152:1787–1795.

Legends to Figures and Tables

Figure 1 – Simplified pathway for the biosynthesis of Anthocyanins in Grapevine. CHS, Chalcone synthase; CHI, Chalcone isomerase; F3H, Flavanone 3-hydroxylase; F3'H Flavonoid 3'-hydroxylase; F3'5'H, Flavonoid 3'5'-hydroxylase; DFR, Dihydroflavonol 4-reductase; LDOX, Leucoanthocyanidin dioxygenase; 3GT, Anthocyanidin 3-O-glucosyltransferase; OMT, Flavonoid 3',5'-methyltransferase. EC numbers for each enzyme are shown. Numbers on blue, yellow and red boxes represent number of known, novel and varietal genes in Tannat, respectively. Figure modified from Da Silva et al. 2013.

Figure 2 - Principal component analysis plot of samples illustrating variation within and between samples replicates.

Figure 3 – Representation of differentially expressed genes within comparisons. Venn diagrams of up-regulated and down-regulated genes are shown. Differentially expressed genes included are the ones with a |Log2FC| ≥2 and adjusted pvalue < 0.001.

Figure 4 - Temporal expression profiles of differentially expressed genes. The twelve significant profiles grouped in six clusters are shown. Numbers on the top left corner indicate the ID of the profile. Numbers under each panel are the p-value. Numbers at the right of each cluster represent total genes assigned to the cluster (top) and number of genes not categorized by GO (bottom). The black lines show model expression profiles. The grey lines represent all individual gene expression profiles. The vertical axis is relative transcript abundance. The horizontal axis is developmental time of the four skins samples. The time series was normalized to start at 0. Histograms left to each cluster represent the functional categories distribution expressed as percentage of the amount of genes belonging to the cluster. Genes with no GO assigned were not considered in the analysis. Other Biological Process: GO:0022610; GO:0001906, GO:0016265, GO:0002376, GO:0040011, GO:0051704, GO:0032501, GO:0048511, GO:0016032; growth/developmental process: GO:0010926, GO:0032502, GO:0040007; cellular component biogenesis/organization: GO:0044085, GO:0016043; reproduction/reproductive process: GO:0000003, GO:0022414; biological regulation: GO:0065007; cellular process: GO:0009987; establishment of localization: GO:0051234; localization: GO:0051179: metabolic process: GO:0008152; pigmentation: GO:0043473; response to stimulus: GO:0050896; Other Cellular Component: GO:0031975, GO:0032991, GO:0031974, GO:0055044, GO:0005576, GO:0044421; cell: GO:0005623; cell part: GO:0044464; organelle/part: GO:0043226, GO:0044422; Other Molecular Function: GO:0016209, GO:0009055, GO:0030234, GO:0045735, GO:0005198; transcription & translation regulator activity/molecular transducer activity: GO:0060089, GO:0030528, GO:0045182; binding: GO:0005488; catalytic activity: GO:0003824; transporter activity: GO:0005215. Details of the genes assigned to each temporal profile and GOs are found in Supplemental Material 2.

Figure 5 – Expression levels for all genes coding for the enzymes F3'H and F3'5'H in Tannat. Expression levels are presented as relative rlog-transformed values through berry development. Red, black and green colors show the amount by which each gene deviates in each time point from the gene average across all time points. Genes were clustered by Euclidean distance using rlog-transformed expression values. Samples identification correspond to the weeks after flowering (waf) in which the sample was taken, and whether the skins were green (G) or red (R) at veraison. Expression patterns framed in purple and red were considered to correspond to genes encoding F3'5'H and F3'H directly related to anthocyanin biosynthesis, respectively. Hashtag symbols (#) indicate genes with significantly higher expression in red skins at veraison. Expression of genes in bold was confirmed by RT-qPCR.

Figure 6 – Correlation of the expression profiles of a varietal F3'H and a novel F3'5'H Tannat gene using the RNA-Seq or RT-qPCR methodology in two different growing seasons. Relative expression levels of VVV_00903 (F3'H) and VVN_0308 (F3'5'H) assessed by RNA-Seq in 2011 to 2012 growing season (identified as '2012 RNA-Seq') and by RT-qPCR in the 2011 to 2012 growing season

(identified as '2012 RT-qPCR') and 2014 to 2015 growing season (identified as '2015 RT-qPCR') are shown as Log_2(Fold Change) against its previous time point. Values shown represent the mean of 3 (for RNA-Seq and 2015 RT-qPCR) or 2 (for 2012 RT-qPCR) replicates ± SD. Significance of differential expression against its previous time point was examined by Student's t-test, and represented as one asterisk ($P<0.05$), two asterisks ($P<0.01$), or three asterisks ($P<0.001$). Samples identification correspond to the weeks after flowering (waf) in which the samples were taken, and whether the skins were green (G) or red (R) at veraison.

Figure 7 – Heat map of expression levels for all genes coding for the enzymes involved in the biosynthesis of anthocyanins. Gene expression shown correspond to genes coding for the enzymes detailed in Figure 1. Expression levels are presented as relative rlog-transformed values through berry development. Red, black and green colors show the amount by which each gene deviates in each time point from the gene average across all time points. Genes were clustered by Euclidean distance using rlog-transformed expression values. Samples identification correspond to the weeks after flowering (waf) in which the samples were taken, and whether the skins were green (G) or red (R) at veraison. Purple and red arrows identify genes encoding F3'5'H and F3'H directly related to anthocyanin biosynthesis, respectively. Expression of genes in bold was confirmed by RT-qPCR.

Figure 8 – 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.). Amplification products were separated on a 1% agarose gel. Image corresponds to the ethidium bromide-stained agarose gel.

Table 1 – Polymorphisms found in the transcription factors MYBA1, MYBA2, and MYBA3. The promoter region considered was within the 1000pb upstream of each gene. The exact SNP position is shown between brackets. The changes shown result from comparing Pinot Noir clone PN40024 (reference grape genome) to Tannat clone UY11.

Supplemental Material 1 – Tables with the primers used for RT-qPCR amplification; summary of the differentially expressed genes (DEG) in each comparison; list of all DEG in each comparison showing expression data; list of anthocyanin related DEG showing expression data; table with the mean rlog-transformed values for the anthocyanin related DEG in skin samples; table with the mean rlog-transformed values for all the genes coding F3'H and F3'5'H genes in grain (1waf) and skin samples.

Supplemental Material 2 – Tables with the list of genes assigned to each significant temporal profile calculated with STEM software, and Gene ontology (GO) categories distribution in each temporal profile.

Supplemental Material 3 – Comparison of genes reported in the literature for anthocyanin biosynthesis in other grape varieties. Table with the list of genes differentially expressed between green and red berry skins at veraison, in common in Tannat and Pinot noir (Gouthu et al., 2014). List of genes related to anthocyanin biosynthesis found in common by Costantini et al., 2015, and this work.

Supplementary tables can be found at:

https://drive.google.com/drive/folders/0BwbzC7laBG7dRVdrcl9Yd2FnVzA?usp=sharing

Figure 1

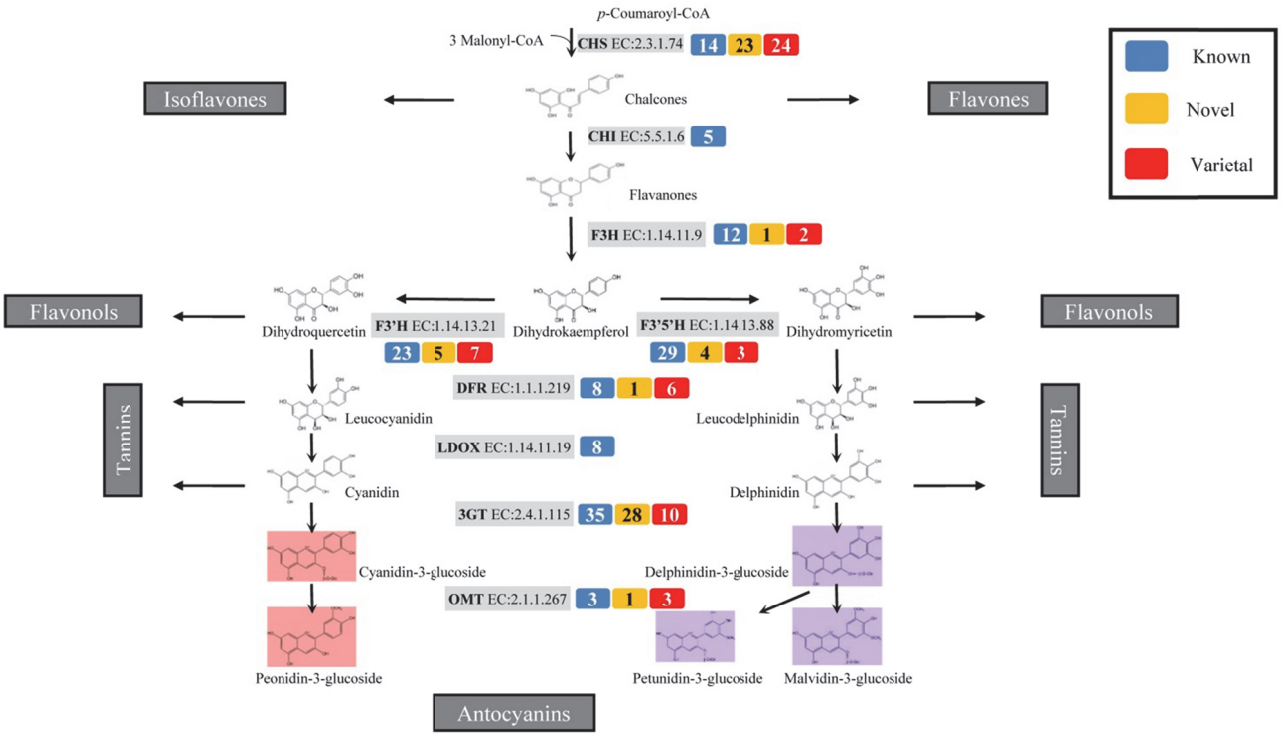

p-Coumaroyl-CoA
3 Malonyl-CoA
CHS EC:2.3.1.74 14 23 24
Isoflavones
Chalcones
Flavones
Known
Novel
Varietal
CHI EC:5.5.1.6 5
Flavanones
F3H EC:1.14.11.9 12 1 2
Flavonols
Dihydroquercetin
F3'H EC:1.14.13.21 23 5 7
Dihydrokaempferol
F3'5'H EC:1.14.13.88 29 4 3
Dihydromyricetin
Flavonols
DFR EC:1.1.1.219 8 1 6
Tannins
Leucocyanidin
Leucodelphinidin
Tannins
LDOX EC:1.14.11.19 8
Cyanidin
Delphinidin
3GT EC:2.4.1.115 35 28 10
Cyanidin-3-glucoside
Delphinidin-3-glucoside
OMT EC:2.1.1.267 3 1 3
Peonidin-3-glucoside
Petunidin-3-glucoside
Malvidin-3-glucoside
Antocyanins

Figure 2

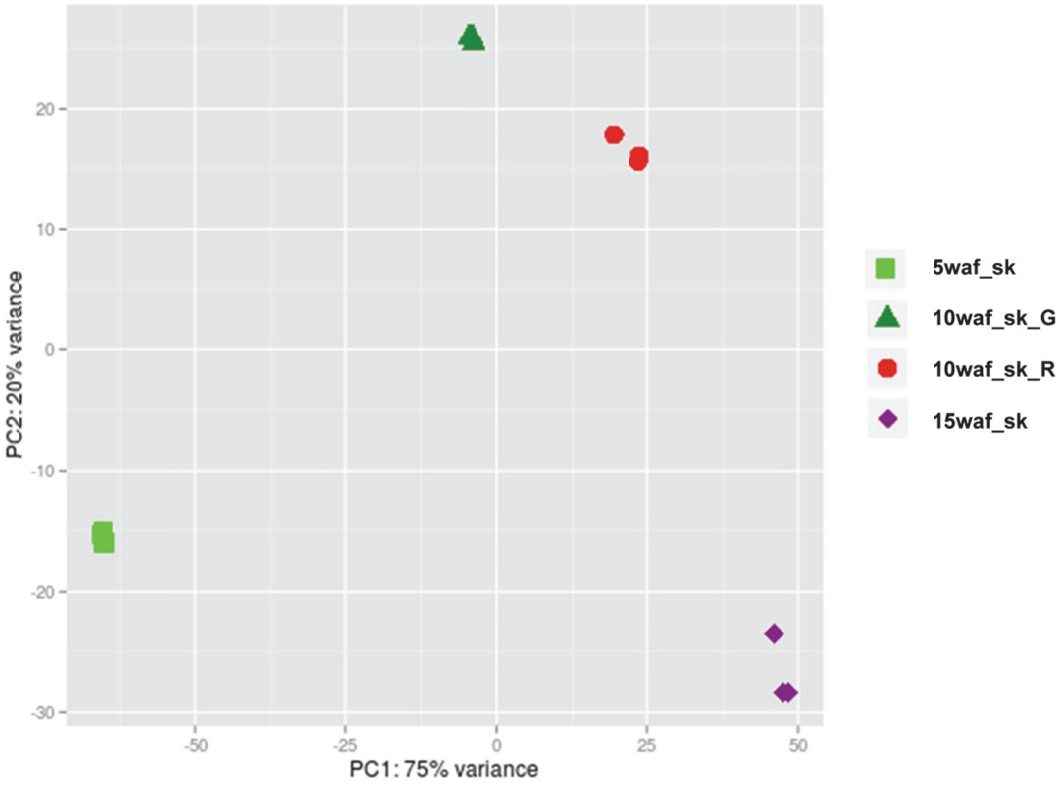

20
10
0
-10
-20
-30
PC2: 20% variance
-50
-25
0
25
50
PC1: 75% variance
5waf_sk
10waf_sk_G
10waf_sk_R
15waf_sk

Figure 3

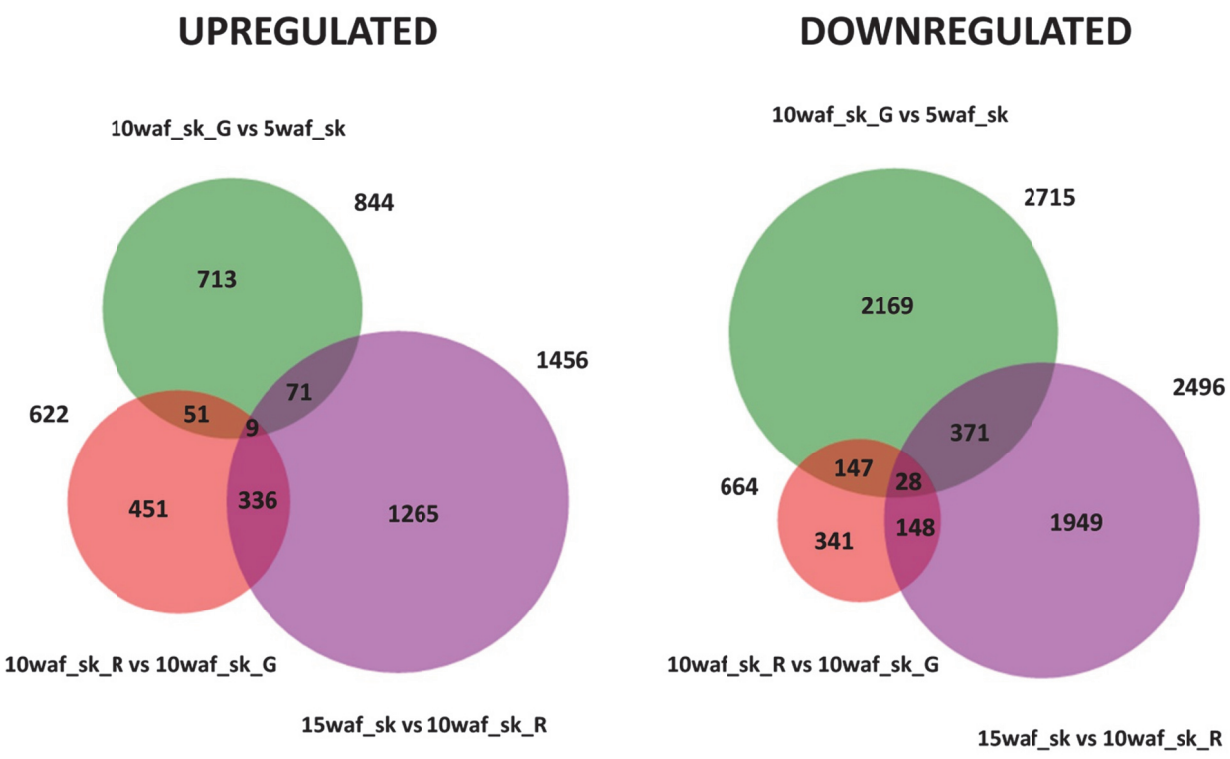

UPREGULATED
10waf_sk_G vs 5waf_sk
844
713
71
1456
622
51
9
451
336
1265
10waf_sk_R vs 10waf_sk_G
15waf_sk vs 10waf_sk_R
DOWNREGULATED
10waf_sk_G vs 5waf_sk
2715
2169
2496
371
664
147
28
341
148
1949
10waf_sk_R vs 10waf_sk_G
15waf_sk vs 10waf_sk_R

Figure 4

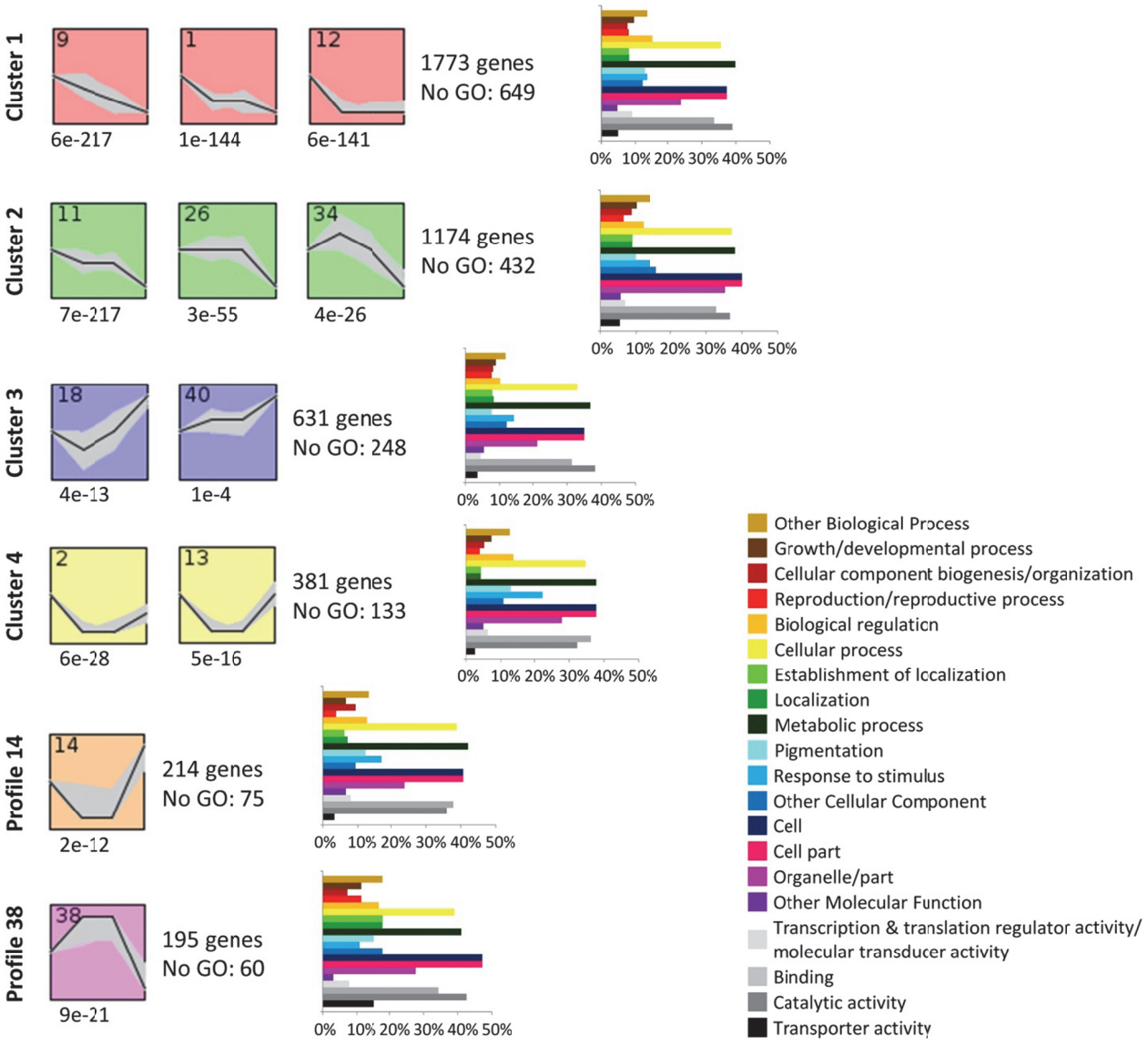

Cluster 1
9
1
12
6e-217
1e-144
6e-141
1773 genes
No GO: 649
Cluster 2
11
26
34
7e-217
3e-55
4e-26
1174 genes
No GO: 432
Cluster 3
18
40
4e-13
1e-4
631 genes
No GO: 248
Cluster 4
2
13
6e-28
5e-16
381 genes
No GO: 133
Profile 14
14
2e-12
214 genes
No GO: 75
Profile 38
38
9e-21
195 genes
No GO: 60
0% 10% 20% 30% 40% 50%
Other Biological Process
Growth/developmental process
Cellular component biogenesis/organization
Reproduction/reproductive process
Biological regulation
Cellular process
Establishment of localization
Localization
Metabolic process
Pigmentation
Response to stimulus
Other Cellular Component
Cell
Cell part
Organelle/part
Other Molecular Function
Transcription & translation regulator activity/ molecular transducer activity
Binding
Catalytic activity
Transporter activity

Figure 5

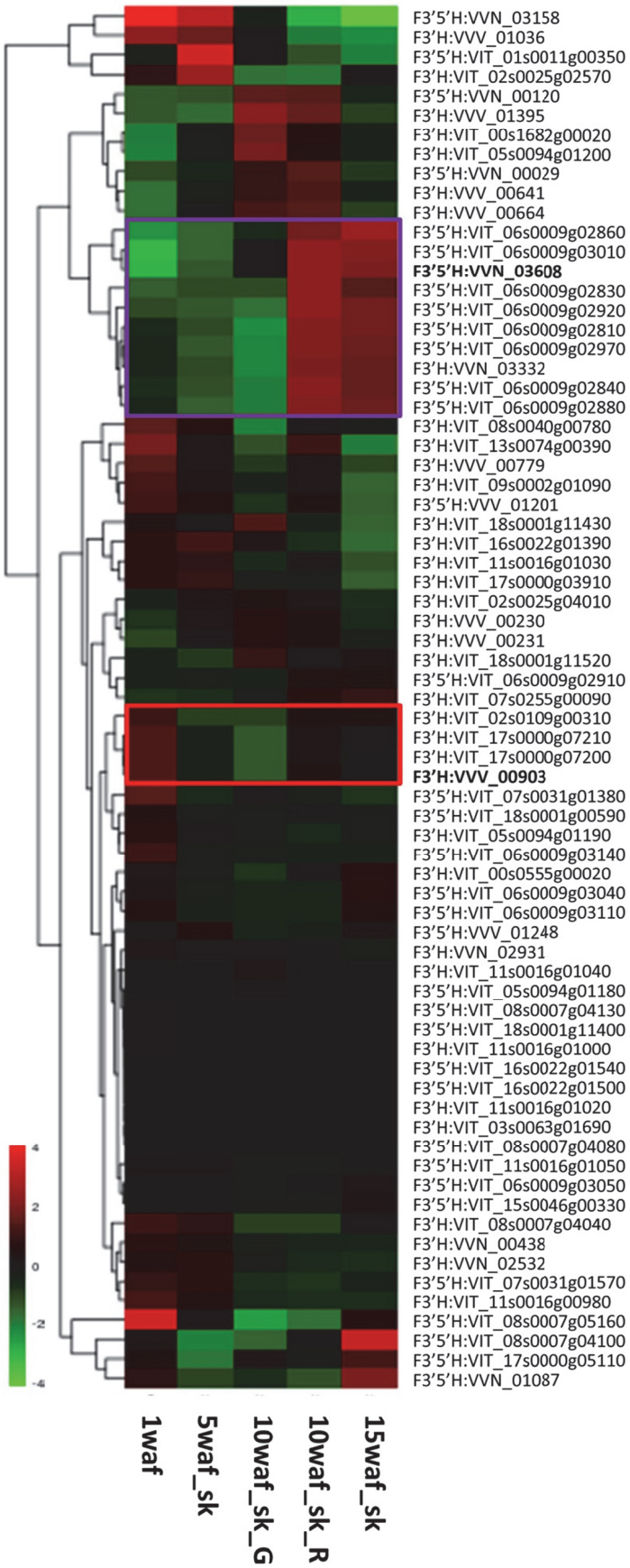

F3'5'H:VVN_03158
F3'H:VVV_01036
F3'5'H:VIT_01s0011g00350
F3'H:VIT_02s0025g02570
F3'5'H:VVN_00120
F3'H:VVV_01395
F3'H:VIT_00s1682g00020
F3'H:VIT_05s0094g01200
F3'5'H:VVN_00029
F3'H:VVV_00641
F3'H:VVV_00664
F3'5'H:VIT_06s0009g02860
F3'5'H:VIT_06s0009g03010
F3'5'H:VVN_03608
F3'5'H:VIT_06s0009g02830
F3'5'H:VIT_06s0009g02920
F3'5'H:VIT_06s0009g02810
F3'5'H:VIT_06s0009g02970
F3'H:VVN_03332
F3'5'H:VIT_06s0009g02840
F3'5'H:VIT_06s0009g02880
F3'H:VIT_08s0040g00780
F3'H:VIT_13s0074g00390
F3'H:VVV_00779
F3'H:VIT_09s0002g01090
F3'5'H:VVV_01201
F3'H:VIT_18s0001g11430
F3'H:VIT_16s0022g01390
F3'H:VIT_11s0016g01030
F3'H:VIT_17s0000g03910
F3'H:VIT_02s0025g04010
F3'H:VVV_00230
F3'H:VVV_00231
F3'H:VIT_18s0001g11520
F3'5'H:VIT_06s0009g02910
F3'H:VIT_07s0255g00090
F3'H:VIT_02s0109g00310
F3'H:VIT_17s0000g07210
F3'H:VIT_17s0000g07200
F3'H:VVV_00903
F3'5'H:VIT_07s0031g01380
F3'5'H:VIT_18s0001g00590
F3'H:VIT_05s0094g01190
F3'5'H:VIT_06s0009g03140
F3'H:VIT_00s0555g00020
F3'5'H:VIT_06s0009g03040
F3'5'H:VIT_06s0009g03110
F3'5'H:VVV_01248
F3'H:VVN_02931
F3'H:VIT_11s0016g01040
F3'5'H:VIT_05s0094g01180
F3'5'H:VIT_08s0007g04130
F3'5'H:VIT_18s0001g11400
F3'H:VIT_11s0016g01000
F3'5'H:VIT_16s0022g01540
F3'5'H:VIT_16s0022g01500
F3'H:VIT_11s0016g01020
F3'H:VIT_03s0063g01690
F3'5'H:VIT_08s0007g04080
F3'5'H:VIT_11s0016g01050
F3'5'H:VIT_06s0009g03050
F3'5'H:VIT_15s0046g00330
F3'H:VIT_08s0007g04040
F3'H:VVN_00438
F3'H:VVN_02532
F3'5'H:VIT_07s0031g01570
F3'H:VIT_11s0016g00980
F3'5'H:VIT_08s0007g05160
F3'5'H:VIT_08s0007g04100
F3'5'H:VIT_17s0000g05110
F3'5'H:VVN_01087
4
2
0
-2
-4
1waf
5waf_sk
10waf_sk_G
10waf_sk_R
15waf_sk

Figure 6

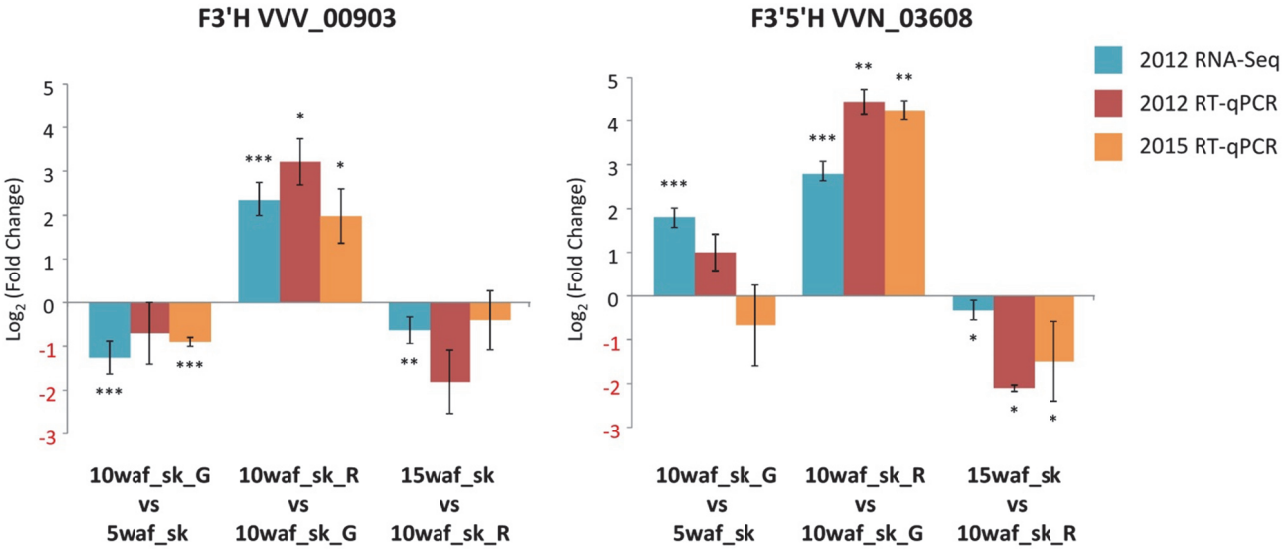

F3'H VVV_00903
F3'5'H VVN_03608
2012 RNA-Seq
2012 RT-qPCR
2015 RT-qPCR
Log2 (Fold Change)
10waf_sk_G vs 5waf_sk
10waf_sk_R vs 10waf_sk_G
15waf_sk vs 10waf_sk_R

Figure 7

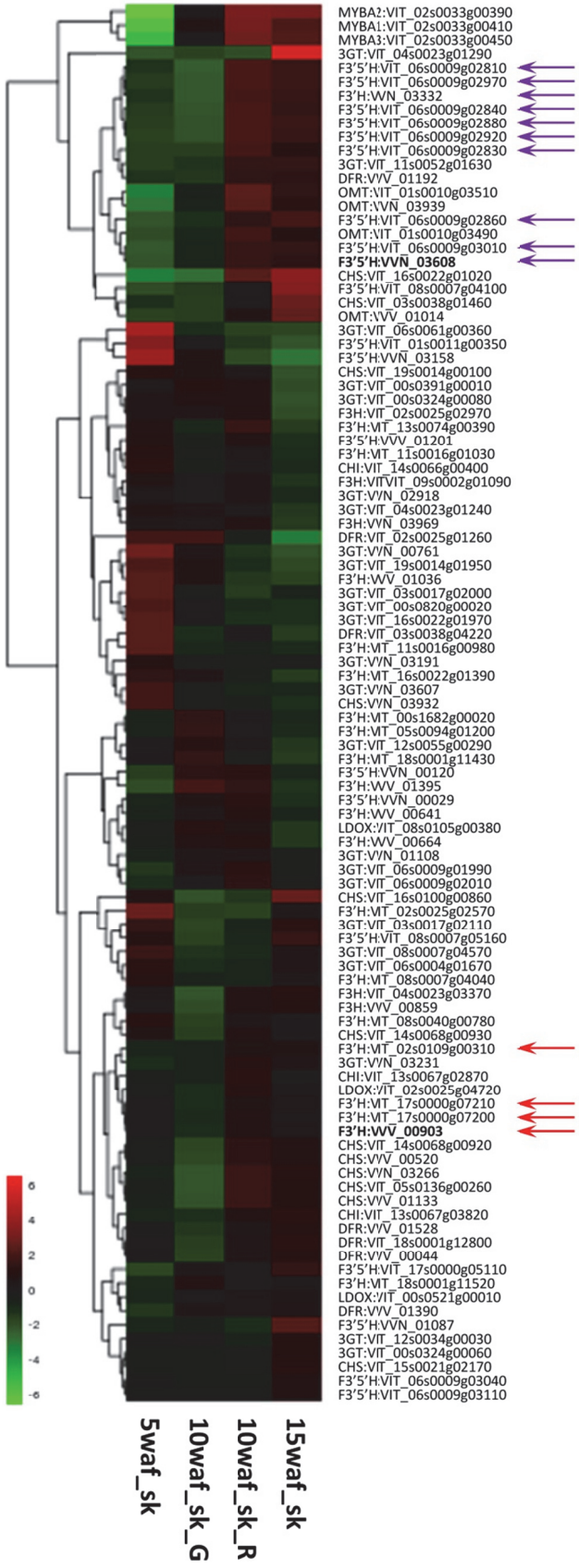

MYBA2:VIT_02s0033g00390
MYBA1:VIT_02s0033g00410
MYBA3:VIT_02s0033g00450
3GT:VIT_04s0023g01290
F3'5'H:VIT_06s0009g02810
F3'5'H:VIT_06s0009g02970
F3'H:VVN_03332
F3'5'H:VIT_06s0009g02840
F3'5'H:VIT_06s0009g02880
F3'5'H:VIT_06s0009g02920
F3'5'H:VIT_06s0009g02830
3GT:VIT_11s0052g01630
DFR:VVV_01192
OMT:VIT_01s0010g03510
OMT:VVN_03939
F3'5'H:VIT_06s0009g02860
OMT:VIT_01s0010g03490
F3'5'H:VIT_06s0009g03010
F3'5'H:VVN_03608
CHS:VIT_16s0022g01020
F3'5'H:VIT_08s0007g04100
CHS:VIT_03s0038g01460
OMT:VVV_01014
3GT:VIT_06s0061g00360
F3'5'H:VIT_01s0011g00350
F3'5'H:VVN_03158
CHS:VIT_19s0014g00100
3GT:VIT_00s0391g00010
3GT:VIT_00s0324g00080
F3H:VIT_02s0025g02970
F3'H:VIT_13s0074g00390
F3'5'H:VVV_01201
F3'H:VIT_11s0016g01030
CHI:VIT_14s0066g00400
F3H:VITVIT_09s0002g01090
3GT:VVN_02918
3GT:VIT_04s0023g01240
F3H:VVN_03969
DFR:VIT_02s0025g01260
3GT:VVN_00761
3GT:VIT_19s0014g01950
F3'H:VVV_01036
3GT:VIT_03s0017g02000
3GT:VIT_00s0820g00020
3GT:VIT_16s0022g01970
DFR:VIT_03s0038g04220
F3'H:VIT_11s0016g00980
3GT:VVN_03191
F3'H:VIT_16s0022g01390
3GT:VVN_03607
CHS:VVN_03932
F3'H:VIT_00s1682g00020
F3'H:VIT_05s0094g01200
3GT:VIT_12s0055g00290
F3'H:VIT_18s0001g11430
F3'5'H:VVN_00120
F3'H:VVV_01395
F3'5'H:VVN_00029
F3'H:VVV_00641
LDOX:VIT_08s0105g00380
F3'H:VVV_00664
3GT:VVN_01108
3GT:VIT_06s0009g01990
3GT:VIT_06s0009g02010
CHS:VIT_16s0100g00860
F3'H:VIT_02s0025g02570
3GT:VIT_03s0017g02110
F3'5'H:VIT_08s0007g05160
3GT:VIT_08s0007g04570
3GT:VIT_06s0004g01670
F3'H:VIT_08s0007g04040
F3H:VIT_04s0023g03370
F3H:VVV_00859
F3'H:VIT_08s0040g00780
CHS:VIT_14s0068g00930
F3'H:VIT_02s0109g00310
3GT:VVN_03231
CHI:VIT_13s0067g02870
LDOX:VIT_02s0025g04720
F3'H:VIT_17s0000g07210
F3'H:VIT_17s0000g07200
F3'H:VVV_00903
CHS:VIT_14s0068g00920
CHS:VVV_00520
CHS:VVN_03266
CHS:VIT_05s0136g00260
CHS:VVV_01133
CHI:VIT_13s0067g03820
DFR:VVV_01528
DFR:VIT_18s0001g12800
DFR:VVV_00044
F3'5'H:VIT_17s0000g05110
F3'H:VIT_18s0001g11520
LDOX:VIT_00s0521g00010
DFR:VVV_01390
F3'5'H:VVN_01087
3GT:VIT_12s0034g00030
3GT:VIT_00s0324g00060
CHS:VIT_15s0021g02170
F3'5'H:VIT_06s0009g03040
F3'5'H:VIT_06s0009g03110
6
4
2
0
-2
-4
-6
5waf_sk
10waf_sk_G
10waf_sk_R
15waf_sk

Figure 8

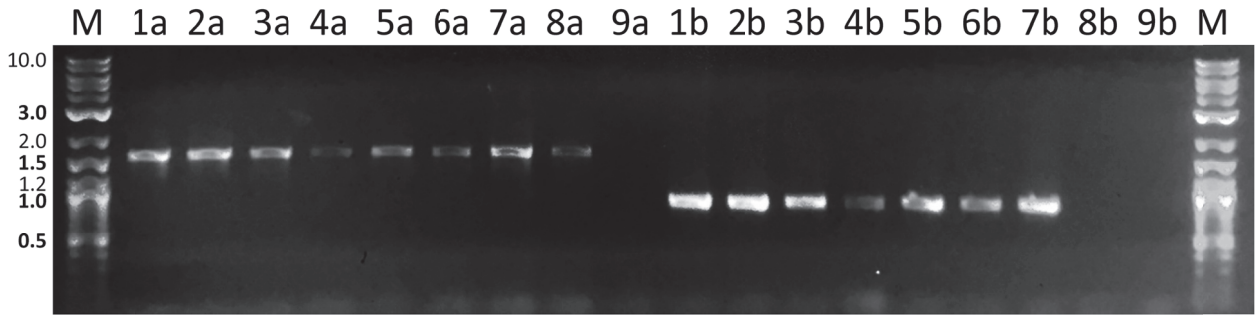

M 1a 2a 3a 4a 5a 6a 7a 8a 9a 1b 2b 3b 4b 5b 6b 7b 8b 9b M
10.0
3.0
2.0
1.5
1.2
1.0
0.5

Table 1

Gene	Type	Region	Change	Effect
MybA1	Gret1	Promoter		
MybA2	SNP	Promoter (ch2:14181334)	C → T/C	
	SNP	3° exon (ch2:14180207)	A → A/G	synonymous change
MybA3	SNP	Promoter (ch2:14309724)	A → A/G	
	SNP	Promoter (ch2:14309444)	A → A/T	
	SNP	Promoter (ch2:14309423)	T → G	
	SNP	Promoter (ch2:14309393)	G → G/T	

CONCLUSIONES Y PERSPECTIVAS FUTURAS

CONCLUSIONES

Este trabajo realiza aportes originales acerca de las bases moleculares de la biosíntesis de antocianinas y taninos durante el desarrollo de Tannat, la uva vinífera insignia del Uruguay.

Estos aportes fueron abordados utilizando la metodología de secuenciación masiva (NGS, por sus siglas en inglés, Next Generation Secuencing), y analizando los datos de secuenciación de ADN y de ARN (RNA-Seq) de diferentes maneras (ensamblado por genoma de referencia y *de novo*), y con diferentes enfoques (búsqueda de nuevos genes, caracterización de patrones de expresión, determinación de posibles funciones previamente desconocidas).

Así pudimos encontrar que el genoma del clon de Tannat UY11 posee genes que no se encuentran en el clon de Pinot Noir PN40024, los cuales llamamos genes varietales, y otros genes que aunque sí se encuentran en PN40024, no habían sido reportados previamente. Mediante la secuenciación de ARNm pudimos determinar regiones del genoma de UY11 transcriptas, codificantes para proteínas, y transcriptos ensamblados *de novo* que por corroboración con reads de ADN encontramos que provienen de genes también presentes en el genoma de PN40024, pero que no fueron incluidos en el genoma reconstruido de este clon. Los resultados determinaron que la mayor parte de las expansiones de familias génicas de vid codifican para enzimas relacionadas con la biosíntesis de polifenoles. De esta manera quedó en evidencia la variabilidad genética que existe entre las diferentes variedades de vid, y por lo tanto la importancia de realizar estudios de este tipo para poder entender el porqué de la tipicidad de cada variedad.

Enfocándonos en los genes que se expresan particularmente en semillas, y sus patrones de expresión en diferentes tejidos y momentos del desarrollo de la uva pudimos determinar la función de un gen que previamente estaba anotado con otra función. Este es un gran aporte, ya que este gen codifica para la enzima clave en la síntesis de taninos galoileados (ECGT). Estos compuestos están presentes en todas las variedades de uva vinífera en diferentes proporciones, pero hasta ahora no se había identificado el gen que codifica para la enzima responsable de esta síntesis.

Al enfocarnos en los genes con expresión diferencial justo en el momento en el que las cáscaras de uva se colorean (envero) pudimos determinar genes específicos dentro de familias génicas muy grandes que están involucrados con la aparición del color en cáscaras. Esto muestra que aunque diferentes genes codifiquen para una misma enzima, éstos no están involucrados en los mismos procesos fisiológicos. También así pudimos determinar un grupo de 195 genes que tienen un pico de expresión solo durante envero, de los cuales

para el 25% no se ha identificado su función. Comprobamos así que la utilización de esta tecnología es un gran avance en el descubrimiento de funciones génicas.

También realizamos un abordaje más clásico mediante la metodología de RT-qPCR, la cual utilizamos para validar niveles de expresión de genes que nos resultaron de interés en diferentes temporadas de crecimiento de la uva. Con esto validamos los datos obtenidos con NGS en dos sentidos: validamos la expresión de estos genes y validamos la existencia de aquellos genes que definimos mediante el ensamblado *de novo* de RNA-Seq.

Con esta tesis, estamos proporcionando las bases genéticas para la comprensión de las particularidades de Tannat, así como también de las variedades viníferas en general.

PERSPECTIVAS FUTURAS

Esta tesis ha sido el primer paso para el desarrollo de estudios con enfoque en las características que identifican a los vinos Tannat. Continuando en esta línea de trabajo, este año presentamos un proyecto CSIC I+D, en conjunto con la Dra. Laura Fariña de la Sección Enología de Facultad de Química, titulado “Estudios metabolómicos y transcriptómicos en *Vitis vinifera* cv. Tannat para la selección de un clon en base a potencial aromático”.

En esta tesis se presentan estudios realizados en un solo clon, y en un único viñedo. En el futuro de analizarán las variaciones génicas y transcriptómicas entre diferentes clones de Tannat que producen vinos con diferentes características organolépticas. También se determinará cómo influye la ubicación del viñedo (“terroir”) en la expresión de genes relacionados con características de interés en vitivinicultura.

Por otra parte, también está planteado clonar los genes de Tannat que codifican para putativas epicatequina galoil transferasas (ECGT) y analizar su expresión *in vitro*. El contar con una ECGT con alta actividad permitiría galoilear diferentes compuestos, analizar sus efectos beneficiosos para la salud e incluso producir preparaciones nutracéuticas.