Caracterização de genes envolvidos no conteúdo de vitamina E em frutos de tomate

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Ciências, na Área de Botânica.

Orientadora: Maria Magdalena Rossi

São Paulo

Ficha Catalográfica

Almeida, Juliana

Caracterização de genes envolvidos no conteúdo de vitamina E em frutos de tomate.

115 p.

Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Botânica.

1. Tocoferol 2. tomate 3. Solanum pennellii 4. Metabolismo do fruto

I. Universidade de São Paulo. Instituto de Biociências. Departamento de Botânica.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Profa. Dra. Maria Magdalena Rossi

Orientadora

Os resultados obtidos neste trabalho fazem parte de um artigo que será publicado pelo *Journal of Experimental Botany* (ver Anexo II).

Almeida J, Quadrana L, Asis R, Setta N, de Godoy F, Bermudez L, Otaiza S, Correa da Silva JV; Fernie AR; Carrari F; Rossi M. **Genetic dissection of vitamin E biosynthesis in tomato.** *Journal of Experimental Botany (in press).*

À sociedade, pelos anos de formação,

ofereço

Aos meus pais com carinho,

dedico

"Para um espírito científico, todo o conhecimento é uma resposta a uma pergunta. Se não existe pergunta, não pode haver conhecimento científico. Nada vem sozinho, nada é dado. Tudo é construído."

Gaston Bachelard

Os pensadores mais admiráveis não separam seu trabalho de suas vidas. Encaram ambos demasiado a sério para permitir tal dissociação e desejam usar cada uma dessas coisas para o enriquecimento da outra."

Charles Wright Mills

Registro aqui meus sinceros agradecimentos:

Ao Departamento de Botânica do Instituto de Biociências da Universidade de São Paulo pelo suporte acadêmico e estrutural.

Ao Programa de Demanda Social da Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida. Ao Programa EU-SOL pelo auxílio financeiro para realização das colaborações internacionais.

À Profa. Maria Magdalena Rossi pela interlocução dedicada e sempre presente; por gerar desafios e problemas a serem enfrentados; pelo exemplo de competência e entusiasmo pela pesquisa.

Ao Prof. Fernando Carrari (INTA, AR) pelas valiosas contribuições e discussões fundamentais para realização desse trabalho; pela recepção amistosa. Ao Prof. Rámon Asis (UNC, AR) pela supervisão nas análises por HPLC; por me receber gentilmente em seu laboratório e casa.

A Leandro Quadrana, co-autor de parte deste trabalho, pela parceria construída.

Às colegas do GMP – Daniela, Fabiana, Junia, Luisa, Paula, Silvia e Thais – pelo aprendizado, companheirismo e vivência. À Dani, pelos ensinamentos sobre práticas de laboratório; pela sua generosidade inestimável. À Fabi, pela ajuda preciosa em todas as etapas; pela amizade e companhia nas investidas internacionais. À Luisa, por indicar caminhos; pelo acolhimento em terras argentinas. À Silvia, pela ajuda com as clonagens e sequenciamento. À Junia, pelo auxílio na manutenção das plantas.

À Profa. Marie Anne Van Sluys pelas portas sempre abertas.

A todos os colegas do GaTE – Andréia, Andrés, Bruno, Cushla, Danielle Quintanilha, Douglas, Érika, Guilherme, Hana, Jonas, Kleber, Leonor, Nathália, Mayra, Robson e Tatiana pelos momentos de trocas, intelectuais e "materiais". Em especial, à Mayra e Danielle pelas contribuições. À Nathalia, pela colaboração nas análises evolutivas do artigo. À Carolina Franco, do LAM, pelo suporte eficiente no sequenciamento.

Ao Danilo (Lafieco) pela ajuda de grande valor no cuidado com "os tomates".

À Profa. Déborah Santos pelas discussões sobre análises cromatográficas. À Profa. Sayuri Miyamoto (IQ-USP) pelos ensaios iniciais de HPLC.

Aos meus alunos que, ao longo desses anos, presentearam-me com as perguntas mais fundamentais. Aos colegas educadores, pelas conversas tão enriquecedoras. Aprendi muito com vocês; não há como sair incólume da experiência de ser professor.

Aos meus pais, Manoel e Cida, generosidade sem par, pacientes com minhas escolhas e ausências. Às minhas irmãs, Nadia e Andréia. Vocês são a base de tudo.

Ao Bruno, por todos os questionamentos que me fizeram repensar as escolhas; quem, seja pelo avesso ou pelo direito, ajudou definitivamente a construir esse "inédito viável".

A todos aqueles que contribuíram direta e indiretamente para realização desse trabalho, muito obrigada!

Abreviaturas e neologismos

AA – aminoácidos

APT – antranilato fosforribosiltransferase

BAC – do inglês "Bacterial Artificial Chromosomes", cromossomos artificiais bacterianos

BLAST – do inglês "Basic Local Alignment Search Tool", Ferramenta básica de busca de alinhamentos locais

- Brix índice da concentração de sólidos solúveis
- cDNA DNA complementar
- **cM –** centimorgan
- CHL clorofilase

CMS - 2-C-metil-D-eritritol 4-P citidililtransferase

CM - corismato mutase

CS - corismato sintase

- DEPC dietilpirocarbonato
- DNA ácido desoxirribonucléico

DAHPS - 3-deoxi-D-arabino-heptulosonato-7-P sintase

- DHQS 3-dehidroquinato sintase
- DHQ 3-dehidroquinato desidratase
- DNAse desoxirribonuclease
- dNTP desorribonucleotídeo trifosfatado

DXR – 2-C-metil-D-eritritol 4-P sintase

DXS - 1-deoxi-D-xilulose-5-P sintase

EDTA - ácido etilenodiamonotetracético

EPSPS - 5-enolpiruvilchiquimato-3-P sintase

EST – do inglês "Expressed Sequence Tag", fragmentos de sequências expressas

FPGS – folilpoliglutamato sintase

GC-MS – do inglês "Gas Chromatography/Mass Spectroscopy, cromatografia gasosa acoplada à espectrometria de massas"

- GFP do inglês "green fluorescent protein", proteína fluorescente verde
- GGPS geranilgeranil pirofosfato sintase

GGDR - geranilgeranil reductase

GPPS - geranil pirofosfato sintase

HGGT - homogentisato geranilgeranil transferase

HPLC - do inglês "High Perfomance Liquid Chromatography", cromatografia

líquida de alta eficiência

HPPD - 4-hidroxifenilpiruvato dioxigenase

HDR - 4-hidroxi-3-metilbut-2-enil-difosfato reductase

HDS - 4-hidroxi-3-metilbut-2-enil-difosfato sintase

HST - homogentisato solanesil transferase

HPT/ VTE2 - homogentisato fitil transferase

IL – do inglês "introgressed line", linhagens introgredidas

IPI – isopentenil difosfato δ-isomerase

ISPE – 4-(citidina 5'-P)-2-C-metil-D- eritritol quinase

ISPF - 2-C-metil-D-eritritol 2,4-ciclodifosfato sintase

Kb - kilo pares de base

LYCB – licopeno β-ciclase

Mb – mega pares de base

MEP – metileritritol fosfato

MPBQMT/ VTE3 - dimetil-fitilquinol metil transferase

nt – nucleotídeo

P – do inglês "phosphate", fosfato

PCR – do inglês "Polymerization Chain Reaction", reação de polimerização em cadeia

PAT – prefenato aminotransferase

PK/ VTE5 - fitol quinase

- PRAI fosforribosilantranilato isomerase
- PVPP polivinilpolipirrolidona
- RNA ácido ribonucléico
- **RNAi –** RNA de interferência
- ROS do inglês "Reactive oxygen species", espécies reativas de oxigênio
- SDH chiquimato desidrogenase
- **SEC14 –** transportador de fosfolipídeos

SK – do inglês "shikimate", chiquimato

SK – chiquimato quinase

TAT – tirosina aminotransferase

TBE - tris-borato-EDTA

TC/ VTE1 - tocoferol ciclase

TyrA – arogenato desidrogenase

Unigene – consenso de sequências de cDNA segundo a Solanaceae Genomics Network (http://solgenomics.net/).

VTE – vitamina E

γ-TMT/ VTE4 – γ-tocoferol C-metil transferase

Índice

Resumo	17
Abstract	19
I. Introdução	
1. Tomate	
1.1. Solanum lycopersicum: cultivar e espécie modelo	
1.2. Do fenótipo ao genótipo – o uso do germoplasma selvagem para de características complexas	estudo 23
2. A vitamina E	29
2.1. Aspectos gerais: estrutura, biossíntese e função nas plantas	
2.2. Controle da biossíntese de tococromanóis	
II. Objetivos	
III. Materiais e Métodos	39
1. Caracterização da rota de biossíntese da vitamina E em tomate	39
1.1. Busca das sequências de cDNA de S. lycopersicum envolvidas ne biossíntese de VTE	a 39
1.2. Predição da localização subcelular	40
1.3. Distribuição cromossômica dos genes identificados	40
 Mapeamento de QTL para conteúdo de tocoferol em fruto de tomat identificação de genes candidatos 	:e e 40
2.1. Material vegetal	41
2.2. Preparação dos extratos	41
2.3. Determinação por HPLC	41
2.4. Análise estatística	42
2.5. Mapeamento de QTL	42
3. Clonagem e caracterização dos alelos de <i>S. pennellii</i> para os gener candidatos identificados	s 43
3.1. Material vegetal	43
3.2. Obtenção de bactérias competentes - Escherichia coli	44
3.3. Extração de DNA genômico de tomate	44
3.4. Eletroforese em gel de agarose	45
3.5. Iniciadores	46

3.6. Extração de RNA e síntese de cDNA	. 47
3.7. Reação de Polimerização em Cadeia (PCR)	. 48
3.8. Clonagem dos produtos de PCR	. 48
3.9. Reação de sequenciamento	. 50
3.10. Análise das sequências e comparação dos alelos	. 50
IV. Resultados	. 51
1. Caracterização da rota de biossíntese da vitamina E em tomate	. 51
2. Mapeamento de QTL para conteúdo de tocoferol em fruto de tomate e identificação de genes candidatos	59
2.1. Mapeamento de QTL	. 59
2.2. Identificação de genes candidatos	. 64
 Clonagem e caracterização dos alelos de S. pennellii para os genes candidatos identificados 	67
V. Discussão	. 70
VI. Conclusões	. 79
VII. Referências Bibliográficas	. 80
Anexo I	. 89
Anexo II	97

Resumo

O tomate (Solanum lycopersicum) é uma das culturas de maior importância econômica atual, além de ser espécie modelo para estudos genéticos em plantas para a qual há diversos recursos genéticos e genômicos bem caracterizados. Tanto o fruto fresco quanto os produtos derivados do tomate constituem fonte de diversos antioxidantes, incluindo a vitamina E (VTE). Em plantas, a síntese dos compostos com atividade de VTE, tocoferóis e tocotrienóis, deriva dos precursores das vias do metileritritol fosfato e do chiquimato. O entendimento dos mecanismos responsáveis pela síntese, transporte e acúmulo da VTE em plantas cultivadas é de grande interesse devido sua implicação para saúde humana e na fisiologia vegetal. Loci para caracteres quantitativos (QTL) para o conteúdo de a-tocoferol em frutos maduros foram previamente mapeados nos cromossomos 6 e 9 em uma população de linhagens introgredidas (ILs) que combinam fragmentos cromossômicos da espécie selvagem Solanum pennellii no fundo genético de S. lycopersicum. Neste trabalho, após busca sistemática em banco de dados sequências expressas. foram identificados 41 loci envolvidos de estruturalmente na biossíntese de VTE em tomate. A análise da distribuição cromossômica desses loci revelou que a maioria dos genes que codificavam enzimas para a rota central do tocoferol estavam localizados nos cromossomos 7, 8 e 9. A variação do conteúdo dos isômeros de tocoferol (α , β , γ e δ) em fruto maduros das ILs foi acessada com intuito de se mapear QTL para cada uma das isoformas. Os perfis metabólicos obtidos permitiram descrever 12 QTL, que corroboram aqueles previamente descritos, além de identificar outros novos nos cromossomos 7 e 8. A análise integrada dos dados metabólicos, genômicos e genéticos levou à proposição de 16 loci candidatos que podem afetar o conteúdo de tocoferol em tomate. A comparação das seguências codificantes desses genes revelou polimorfismos de nucleotídeos e aminoácidos entre os alelos de S. lycopersicum e S. pennellii. Em conjunto, esses resultados representam um importante passo para o entendimento dos determinantes genéticos envolvidos na variação natural do conteúdo de VTE em frutos de tomate.

Abstract

Tomato (S. lycopersicum) is one of the most important worldwide crops, besides being a relevant model species for plant genetics studies in plants for which there are many genetic and genomic resources well characterized. Both, fresh fruits and tomato-derived products are source of several antioxidants, including vitamin E (VTE). In plants, the synthesis of VTE compounds, tocopherol and tocotrienol, derives from precursors of the shikimate and methylerythritol phosphate pathways. Understanding the mechanisms underlying synthesis, transport and accumulation of VTE in crops are of great interest because of its implications for human health and its role in plant physiology. Quantitative trait *loci* (QTL) for α-tocopherol content in ripe fruit for chromosome 6 and 9 have been mapped previously in a population of introgression lines (ILs) containing chromosome segments of the wild species Solanum pennellii in the genetic background of the S. lycopersicum. In this work, after a systematic survey in expressed sequences database, 41 loci involved structurally on VTE biosynthesis were identified. The analysis of chromosome distribution of these loci revealed that the majority of genes that encode enzymes for the tocopherol core pathway were located on chromosomes 7, 8 and 9. The variation of tocopherol isomers content (α , β , γ and δ) in ripe fruits of ILs was addressed in order to identify QTL for each isoform. The metabolic profiles obtained allowed to describe 12 QTL that corroborate those previously described while novel ones were identified in chromosome 7 and 8. The integrated analysis at the metabolic, genetic and genomic levels lead to propose 16 candidate *loci* putatively affecting tocopherol content in tomato. A comparative analysis revealed polymorphisms at nucleotide and amino acid levels between S. lycopersicum and S. pennellii candidate alleles. Together, these results represent an important step for understanding the genetic determinants of VTE natural variation in tomato fruit.

I. Introdução

1. Tomate

1.1. Solanum lycopersicum: cultivar e espécie modelo

O tomate (Solanum lycopersicum L., antigo Lycopersicon esculentum, Miller) atualmente desponta como uma das culturas de maior importância econômica mundial além de ser um modelo de estudo para estudos genéticos em plantas.

Sob a perspectiva agronômica, o tomate é a segunda hortaliça mais cultivada, depois da batata, com uma produção anual de cerca de 130 milhões de toneladas (FAO, Base de dados estatísticos, última atualização 2008; http://faostat.fao.org). O tomateiro é cultivado nos cinco continentes, que incluem regiões tropicais, subtropicais e temperadas. China e Estados Unidos destacam-se na produção mundial; o Brasil consolida-se na nona posição, cuja produção anual corresponde a 3,9 milhões de toneladas.

A produção de tomate é constituída de duas cadeias produtivas distintas, que determinam o destino da produção, consumo *in natura* ou abastecimento industrial. A produção para esses dois casos diverge desde as variedades escolhidas para o plantio até as formas de cultivo empregadas (Camargo *et al.*, 2006).

Atualmente, inúmeros são os desafios impostos às práticas de melhoramento moderno que buscam contemplar os diferentes setores envolvidos no mercado de consumo de tomate. Para a produção, o foco do melhoramento recai sobre o rendimento, resistência a patógenos e stress abióticos. Já para indústria, busca-se o aumento do conteúdo de sólidos solúveis (*brix*) nos produtos derivados do tomate. Para o consumo, o alvo é o incremento da qualidade do fruto, que envolve caracteres relacionados a teores nutricionais, aroma, sabor e textura. O fruto de tomate, para consumo fresco ou processado, possui alto valor nutricional em virtude da disponibilidade de diferentes tipos de micronutrientes como carotenóides (especialmente o licopeno), vitaminas C (VTC) e E (VTE), folato, compostos fenólicos e fibras (Abushita *et al.*, 1997; Beecher, 1998; Periago & Garcia, 2009).

S. lycopersicum é uma das espécies domesticadas da família Solanaceae, da qual também fazem parte batata (Solanum tuberosum), berinjela (Solanum melongena), pimenta (Capsicum sp) e tabaco (várias espécies). Essa família inclui mais de 3000 espécies, muitas das quais evoluíram em hábitats localizadas nas regiões andinas e amazônicas da América do Sul, os quais variam enormemente em relação à disponibilidade de água, altitude, temperatura e condições do solo. O centro da diversidade das solanáceas é estimado próximo ao Equador e, ao longo de sua história evolutiva, as espécies acumularam variações genéticas adaptativas ao mais diferentes nichos ecológicos. Em virtude dessa grande diversidade intra e interespecífica apresentada, além de sua importância econômica, Solanaceae tem se consolidado como modelo de estudo em genômica comparada e evolutiva (Knaap et al., 2004; Kahlau et al., 2006).

Dentro dessa família, o tomate é a espécie mais profundamente estudada. Tal fato ocorre, em parte, devido às características intrínsecas da espécie como genoma diplóide e de pequeno tamanho (950 Mb), curto tempo de geração e reprodução por autofecundação (Shibata, 2005). Além disso, importantes ferramentas para análise genética foram desenvolvidas e aplicadas com sucesso na espécie como, por exemplo, a transformação genética estável e o silenciamento gênico induzido por vírus (VIGS) (Barone *et al.*, 2008). Por fim, o tomate representa um modelo de estudo alternativo à espécie vegetal mais estudada, *Arabidopsis thaliana*, nos estudos de processos relacionados aos aspectos formação e desenvolvimento de frutos carnosos (Giovannoni, 2004).

A importância econômica aliada às características favoráveis da espécie contribuiu para que inúmeros recursos genéticos e genômicos fossem desenvolvidos em tomate, os quais incluem: (i) caracterização do germoplasma selvagem; (ii) populações de mapeamento; (iii) coleções de mutantes; (iv) mapas genéticos de alta densidade; (v) bibliotecas de BACs (*Bacterial Artificial Chromosomes*); (vi) mapas físicos; (vii) coleções de EST (*expressed-sequence tags*) e sequências de cDNA completas; e (viii) o genoma sequenciado em fase final de montagem (http://solgenomics.net).

O sequenciamento do genoma do tomate, que teve início em 2005, está sendo conduzido por um consórcio, formado por diferentes países, responsável pela obtenção das sequências, montagem e anotação do genoma de *S. lycopersicum* (*International Solanaceae Genomics Project* - SOL). Estudos anteriores gerados a partir da análise computacional de coleções de ESTs e de sequências genômicas, estimaram a existência de aproximadamente 35.000 genes na espécie, distribuídos preferencialmente nas regiões de eucromatina. Estas regiões representam menos de ¼ do DNA total presente no núcleo, o qual corresponde a 220 Mb dos 950 Mb do genoma haplóide do tomate (van der Hoeven *et al.*, 2002). O sequenciamento está em fase final de execução e um "esboço" do genoma já está disponível no banco de dados da *Solanaceae Genomics Network* (http://solgenomics.net/).

A disponibilidade da sequência do genoma amplia consideravelmente o poder de alcance em compreender os processos moleculares que determinam a morfologia e composição química da planta, mas ainda sim, constitui apenas a referência inicial que norteará os futuros estudos funcionais (Rothan & Causse, 2007; Alonso-Blanco *et al.*, 2009).

1.2. Do fenótipo ao genótipo – o uso do germoplasma selvagem para estudo de características complexas

Devido sua extraordinária variedade em caracteres fenotípicos, a diversidade natural é frequentemente utilizada como ponto de partida na investigação da base genética subjacente aos processos fisiológicos. Embora algumas das diferenças fenotípicas observadas sejam ocasionadas por variantes alélicas de um único gene (herança monogênica), a maioria dos caracteres de interesse agronômico é de herança quantitativa, sendo determinada por múltiplos *loci* espalhados pelo genoma, que são denominados *loci* para caracteres quantitativos (QTL, *quantitative trait loci*) (Fernie & Schauer, 2009; Stitt *et al.*, 2010). O fenótipo é, portanto, resultado da expressão de vários genes e suas possíveis interações que, por sua vez, estão sob influência ambiental (Mauricio, 2001; Zamir, 2001). No caso particular do tomate, existem 17 espécies selvagens dentro da seção *Lycopersicon* (Lippman *et al.*, 2007) as quais exibem grande variação no conteúdo de metabólitos em folhas e frutos (Schauer *et al.*, 2005). Esse fato torna o

germoplasma selvagem fonte valiosa para desvendar as bases genéticas do metabolismo, sendo que muitos esforços estão voltados para o melhoramento nutricional e industrial do cultivar de elite (Zamir, 2001; Fernie *et al.*, 2006; Tohge & Fernie, 2010).

Os caracteres de herança poligênica - o conteúdo de metabólitos, por exemplo - possuem distribuição fenotípica contínua, o que torna difícil a sua associação com as variantes genotípicas, em contraste aos caracteres qualitativos para os quais o fenótipo possui um número discreto de classes (Stitt *et al.*, 2010). Assim, abordagens experimentais e estatísticas sofisticadas foram desenvolvidas para analisar geneticamente as características de herança complexa, como o mapeamento de QTL. Tradicionalmente, nesse tipo de análise, uma população de plantas segregantes para o caráter de interesse é genotipada com marcadores moleculares. Busca-se, então, associar os valores fenotípicos observados com o genótipo, de forma a identificar *loci* cosegregantes ao fenótipo. Existem diversos tipos de populações experimentais utilizadas para mapeamento de QTL e a sua estrutura depende do esquema de cruzamento escolhido, que pode ser por autofecundação ou retrocruzamento (Stitt *et al.*, 2010; Rothan & Causse, 2007).

O advento dos marcadores moleculares que detectam polimorfismos ao nível do DNA, no início dos anos 90, alavancou o estudo e mapeamento genético de QTL. Por outro lado, possibilitou o desenvolvimento de populações de linhagens introgredidas (ILs) que, por sua vez, constituíram também uma poderosa ferramenta para o estudo de caracteres quantitativos determinados por um grande número de QTL de pequeno efeito em espécies autógamas. Isso porque essas linhagens, geradas por retrocruzamentos sucessivos da progênie F₁ com um dos genitores, contêm apenas um único fragmento introgredido do genótipo doador contrapondo-se ao fundo genético isogênico. Consequentemente, qualquer diferença fenotípica observada entre a IL e o parental recorrente é atribuída exclusivamente ao fragmento genômico introgredido. Outra vantagem dessa população experimental é o fato de constituírem um recurso homozigoto e permanente, uma vez que podem ser mantidas por autofecundação, permitindo seu emprego em diversas situações experimentais (Zamir, 2001; Lippman *et al.*, 2007; Rothan & Causse, 2007).

Em tomate, uma população de ILs pioneira que vem sendo utilizada nos últimos 15 anos para mapeamento de QTLs de diversas naturezas é a coleção desenvolvida por Eshed & Zamir (1995). Essa população combina fragmentos relativamente pequenos da espécie selvagem *Solanum pennellii* (LA716) no fundo genético de *S. lycopersicum* (M82) (Figura 1).

(a)



(b)



Figura 1. Esquema ilustrativo da obtenção das linhagens introgredidas (ILs) de *S. pennellii*.

(a) O esquema representa um exemplo de três linhagens para os cromossomos 1 e 6. As ILs foram obtidas através de sucessivos retrocruzamentos e seleção assistida por marcadores moleculares, duas seguido de gerações de autofecundação para obtenção dos fragmentos em homozigose. Adaptado de Zamir (2001). (b) Imagem de frutos maduros dos parentais S. lycopersicum (cv. M82); S. pennellii (LA716) e da geração F1 híbrida obtida (retirado de http://zamir.sgn.cornell.edu/).

S. pennellii, espécie originária de regiões áridas do oete do Peru, é um grupo relativamente distante de *S. lycopersicum*. A divergência entre ambas as espécies foi estimada em 2,7 milhões de anos (Kamenetzky *et al.*, 2010) e suas histórias evolutivas se diferenciam em termos de morfologia, composição bioquímica (especialmente de compostos secundários) e respostas a stress biótico e abiótico. Apesar das drásticas diferenças ecológicas entre as espécies, *S. pennellii* é sexualmente compatível com *S. lycopersicum* cujo cruzamento produz híbridos férteis (Figura 1). Dessa maneira, essa espécie

selvagem que possui um pequeno fruto verde foi escolhida como parental doador no desenvolvimento da primeira população de ILs utilizadas na identificação de QTL interespecíficos, clonagem e melhoramento genético (Lippman *et al.*, 2007).

A coleção consiste em 76 linhagens (genótipos) nas quais segmentos definidos do genoma de *S. pennellii* substituem regiões homólogas em um fundo genético de *S. lycopersicum*. Ao todo, as ILs cobrem por completo o genoma de tomate e os fragmentos relativamente pequenos estão bem delimitados através de marcadores moleculares (http://www.sgn.cornell.edu/maps/pe.pl) (Figura 2).



Figura 2: Representação esquemática da população de ILs de Solanum pennellii desenvolvidas por Eshed & Zamir (1995). (a) Visão geral das 76 ILs, apresentando os fragmentos de *S. pennelli* das diferentes linhagens introgredidas. Em cada fragmento se indica o nome da linhagem correspondente [retirado de Lippman *et al.* (2007)]. (b) Linhagens introgredidas para o cromossomo 1. As letras indicam os "BINs", nos quais é possível dividir o cromossomo 1 de acordo à sobreposição dos diferentes fragmentos introgredidos.

Uma vez estabelecida a coleção, a fenotipagem das ILs de *S. pennellii* para centenas de caracteres permitiu a identificação de mais de 2000 QTLs que afetam desde rendimento, morfologia da planta, compostos voláteis até o conteúdo nutricional, níveis de açúcares e vitaminas no fruto (Lippman *et al.*, 2007).

Apesar da complexa herdabilidade dos caracteres quantitativos para o conteúdo de metabólitos, os quais são dependentes do envolvimento de múltiplas vias biossintéticas, centenas de QTL afetando a composição metabólica do fruto de tomate foram identificados nos últimos anos devido ao desenvolvimento de ferramentas de alto rendimento e precisão para realização de perfis metabólicos (por exemplo, GC-MS) (Fernie & Schauer, 2009).

Um desses estudos, conduzido por Schauer et al. (2006), realizou um perfil metabólico detalhado nos frutos das 76 ILs. Os autores quantificaram 74 metabólitos de estrutura conhecida e descreveram 9 caracteres morfológicos relacionados a rendimento (brix, peso e tamanho do fruto, sementes por plantas, entre outros). Os resultados permitiram identificar 326 loci associados a rendimento (YALs, yield associated loci) e 889 loci metabólicos quantitativos (QMLs, quantitative metabolic loci) que explicam a variação relacionada ao conteúdo de aminoácidos, ácidos orgânicos e graxos, acúcares, álcoois, VTC e VTE. Dois desses QTL que explicam a variação para o conteúdo de α-tocoferol no fruto foram localizados nos cromossomos 6 e 9. Experimentos independentes disponíveis no Banco de Dados de Genômica Funcional de Tomate (Tomato Functional Genomics Database - TED, Fei et al., 2006; http://ted.bti.cornell.edu/) também revelaram diferenças no conteúdo de tocoferol associado com as mesmas regiões genômicas (Tabela 1). Entretanto, mecanismos que explicam essa variação não são atualmente os compreendidos, parcialmente devido à falta de conhecimento dos genes envolvidos na via biossintética da VTE em tomate.

Tabela 1: Mudanças no conteúdo de tocoferol observadas para as diferentes ILs. Valores representam porcentagem de aumento/decréscimo em relação ao controle (*S. lycopersicum*, M82).

IL	Schauer et al. 2006	Tomato Functional Genomics Database ¹			
		α	Y	δ	total
9.1	104,92	81,8	53,7	64,8	73,3
9.1.3	95,35	200	170,4	18,5	181,2
9.2	48,10	72,3	24,1	-53,7	52,5
6.1		80,3	33,3	27,8	65,3
6.2	122,71	****	****	***	***
6.2.2		258,4	251,9	225,9	254
6.3	44,52	172,3	7,4	-35,2	1 <mark>1</mark> 6,8

¹ Dados obtidos em Tomato Functional Genomics Database (http://ted.bti.cornell.edu). **** Valores não disponíveis para a IL.

---- ILs que não apresentaram mudanças significativas do conteúdo de tocoferol (p<0.05) em relação à *S. lycopersicum* (M82).

O conjunto de dados proveniente da caracterização fenotípica das ILs, gerados a partir de técnicas de metabolômica, constitui um marco para o estudo das bases genéticas responsáveis pela variação fenotípica. No entanto, o percurso para a resolução da base genética de um QTL tem se mostrado desafiador. Até agora, poucos QTL identificados em tomate foram caracterizados em seu nível genético/molecular.

Um desses exemplos está relacionado ao conteúdo de β -caroteno. Frutos maduros de *S. pennellii* são verdes, enquanto a maioria das variedades de tomate cultivada possui coloração vermelha. Curiosamente, algumas ILs apresentam frutos maduros com nova coloração, como os frutos de laranja escuro da IL 6-3. A clonagem posicional do gene dominante para cor laranja do fruto da IL 6-3, que codifica para enzima licopeno- β -ciclase envolvida na via biossintética dos carotenóides, revelou que a porção codificante do alelo de *S. pennellii* é idêntica comparada ao alelo de M82, porém carrega elementos adicionais em *cis* na região promotora. Consequentemente, o padrão de expressão dos alelos é diferencial; a licopeno β -ciclase na IL 6-3 é altamente expressa nos estágios finais de maturação do fruto, aumentando o conteúdo de β -caroteno que confere a cor laranja ao fruto (Ronen *et al.*, 2000).

Outro exemplo está relacionado ao conteúdo de açúcares no fruto. Fridman *et al.* (2000, 2004), identificaram uma região genômica no cromossomo 9 de *S. pennellii* responsável por um incremento de 25 % no *brix*. 28 Após o mapeamento genético e físico, a região foi sequenciada. A análise revelou a preseça de um gene codificante para uma invertase de parede celular que foi chamada de *lin5*. O estudo comparativo dos alelos de *S. pennellii* e *S. lycopersicum,* mostrou que a alteração de um único nucleotídeo no terceiro éxon do gene modificava as propriedades cinéticas da enzima. Esta invertase é determinante no carregamento de sacarose do floema para dentro da célula, sendo um fator importante para o incremento do conteúdo de açúcar no fruto.

Como explicitado nos exemplos anteriores, classicamente, na ausência de dados de genômica, a clonagem posicional era a única abordagem possível para a dissecção de um QTL. Porém, esta estratégia é extremamente laboriosa e demorada, sendo necessária a análise de grandes populações de indivíduos recombinantes. A disponibilidade recente de sequências de cDNA e do genoma nuclear criou a possibilidade de busca, ao longo da região genômica onde se localizam os QTL, de genes candidatos cuja variação poderia explicar a alteração fenotípica observada. Uma vez identificados, a validação dos genes candidatos pode ser feita por meio da manipulação do organismo vegetal, utilizando diversas ferramentas de genética reversa, muitas delas baseadas em mutagênese ou no silenciamento gênico específico através de tecnologias baseadas em RNAi (RNA de interferência) (Rothan & Causse, 2007). Essa abordagem se mostra pertinente para o estudo dos mecanismos que atuam na regulação de vias de biossíntese cujos componentes estruturais são conhecidos, como é o caso da vitamina E.

2. A vitamina E

2.1. Aspectos gerais: estrutura, biossíntese e função nas plantas

A evolução dos processos metabólicos aeróbios, tais como a respiração e a fotossíntese, levou inevitavelmente à produção de espécies reativas de oxigênio (ROS) em mitocôndrias, cloroplastos e peroxissomos. Ânions superóxidos (O2^{*-}), radicais lipídicos peroxil (ROO^{*}), peróxido de hidrogênio (H₂O₂), radical hidroxil (OH^{*}), oxigênio singlete (¹O₂) são continuadamente produzidos durante o metabolismo aeróbio. Apesar de atuar como sinalizadores em diversos processos celulares, os diferentes tipos de ROS gerados são capazes de causar danos oxidativos a proteínas, DNA e lipídios. Tais propriedades citotóxicas das ROS constituíram uma pressão seletiva no desenvolvimento de mecanismos dinâmicos e redundantes de detoxicação enzimática e não-enzimática em plantas. Dentre os principais mecanismos enzimáticos existentes nos organismos vegetais, encontram-se a superóxido dismutase, ascobarto peroxidase, glutationa peroxidase e a catalase. Já o sistema de antioxidantes não-enzimáticos é composto principalmente por ascorbato (VTC), glutationa, flavonóides, carotenóides, alcalóides e tococromanóis (VTE) (Apel & Hirt, 2004).

O termo "vitamina E" foi primeiramente introduzido por Evans e Bishop (1922) para descrever um importante fator da dieta que influenciava na reprodução animal. Mais de 40 anos se passaram até que a VTE fosse associada a propriedades antioxidantes (Epstein *et al.*, 1966 *apud* Munné-Bosch & Alegre, 2002). A partir de então, a função desses compostos lipofílicos e de outros antioxidantes tem sido extensivamente explorada em mamíferos. As diversas frentes de pesquisa incluíram tanto o estudo da sua localização e estrutura química que explica suas propriedades antioxidantes, quanto, especialmente, a atuação desses compostos para a saúde humana (Munné-Bosch & Alegre, 2002). Estudos recentes têm reforçado a hipótese dos efeitos benéficos da VTE na saúde humana com maior ênfase na prevenção de doenças cardiovasculares, câncer de mama, proteção contra o stress oxidativo induzido por nicotina além disfunções neurológicas (Ricciarelli *et al.*, 2002; Ros 2009; Zhang *et al.*, 2009; Das *et al.*, 2009).

Embora os óleos de sementes constituam a principal fonte de VTE da dieta humana, o tomate *in natura* e seus produtos derivados (pasta em conserva, por exemplo) juntamente com brócolis e espinafre apresentam os maiores teores de α -tocoferol dentre as hortaliças: 0,53; 4,66; 1,44; 1,96 mg/100 g de peso comestível, respectivamente (Chun *et al.*, 2006). O índice recomendado de ingestão diária para α -tocoferol é de 15 mg/dia (IOM, Institute of Medicine, 2000).

Como antioxidante, o principal papel de tocoferóis e tocotrienóis é eliminar radicais peroxil dos lipídeos, responsáveis pela propagação da peroxidação lipídica (Liebler, 1993). O controle da homeostase *redox* por

tocoferóis, ascorbato e glutationa aparenta estar intimamente interligado. Enquanto tocoferóis controlam a extensão da peroxidação lipídica, ascobarto e glutationa agem como tampões do potencial redox celular, controlando, entre outros, o nível de peróxido de hidrogênio (Foyer & Noctor, 2005). Dessa maneira, o ascorbato evita a formação de radicais hidroxil altamente reativos e, indiretamente, a peroxidação lipídica. Além disso, o ciclo do ascorbatoglutationa suporta a função antioxidante de tocoferóis através da reciclagem dos radicais tocoferoxil, resultado da reação de tocoferóis com os radicais peroxil (Figura 3). Além do seu papel na captura de radicais lipídicos peroxil, tococromanóis também atuam na eliminação (*quenching*) dos oxigênios singletes, isso porque uma molécula de α -tocoferol é capaz de desativar centenas de oxigênios singletes por meio de processos físicos (transferência de energia por ressonância). Adicionalmente, tocoferóis podem reagir quimicamente com oxigênios singletes e assim destruí-los (*scavenging*) (Falk & Munné-Bosch, 2010).



Figura 3. Esquema de controle da homestase redox feito por tocoferóis. Tocoferóis contribuem para o controle da homeostase redox modulando a extensão da peroxidação lipídica em folhas. Tocoferol inibe a propagação da peroxidação lipídica por meio da captura de radicais peroxil e previne a peroxidação por meio da reação com outras ROS, como o oxigênio singlete, juntamente com os carotenóides. O ciclo ascobarto-glutationa atua na reciclagem dos radicais tocoferoxil a tocoferol. PUFAs (*Poly Insaturated Fatty Acid*, ácido graxo poliinsaturado) Adaptado de Falk & Munné-Bosch, 2010.

In planta, a função dos tocoferóis permanece ainda não completamente definida. Recentes estudos apontam que a interação entre α-tocoferol e carotenóides parece ser relevante na resposta ao stress foto-oxidativo,

protegendo pigmentos, proteínas e ácidos graxos poliinsaturados do aparato fotossintético contra espécies reativas de oxigênio (ROS) (Semchuk *et al.*, 2009). A oxidação desses compostos causa a perda de atividade do fotossintema II principalmente pela degradação da proteína D1 localizada no centro de reação (Krieger-Liszkay & Trebst 2006). Além disso, embora subestimada na literatura, tococromanóis não apenas desempenham função como antioxidantes lipofílicos, mas também operam como importantes agentes estabilizadores da estrutura da membrana (Falk & Munné-Bosch, 2010).

A síntese de VTE ocorre somente em organismos fotossintetizantes e seus produtos correspondem a moléculas anfipáticas, compostas por um anel hidroxicromanol polar, derivado homogentisato, e uma cadeia lateral poliprenil lipofílica, produtos da via do chiquimato (SK) e metil-eritritrol-fosfato (MEP), respectivamente (Munné-Bosch & Alegre, 2002). A excelente atividade antioxidante da VTE mencionada anteriormente dá-se pela capacidade do anel cromanol heterocíclico em doar hidrogênio aos radicais livres dos lipídeos.

Os compostos com atividade VTE, coletivamente chamados de tococromanóis, podem ser divididos em dois grupos de acordo com o grau de saturação da cadeia lipofílica. Tocoferóis, mais abundante em plantas, têm cadeia lateral saturada derivada do fitil 2P, enquanto tocotrienóis possuem cadeia insaturada derivado do geranilgeranil 2P. Existem quatro isômeros de tocoferóis/tocotrienóis - α , β , $\gamma \in \delta$ - que variam quanto ao número e posição dos grupos metila junto ao anel cromanol. Entretanto, α -tocoferol é o composto que apresenta maior atividade vitamínica (Munné-Bosch & Alegre, 2002; DellaPenna & Last, 2006) (Figura 4). Essa diferença decorre da acumulação e distribuição diferencial desse isômero no corpo dos animais. A retenção é mediada por uma proteína de transferência de α -tocoferol hepática (α -TTP), que se liga preferencialmente ao α -tocoferol e aumenta, portanto, sua transferência para membrana celular, determinando sua maior atividade vitamínica em relação às demais isoformas (Hosomi *et al.*, 1997).



Figura 4: Estrutura química do tocoferol e do tocotrienol (adaptado de Dellapenna & Last, 2006). A tabela indica o número/posição do radical metil no anel em α , β , γ e δ tocoferol e tocotrienol. A atividade vitamínica E de cada composto é relativa ao a-tocoferol (100%). nm, não mensurável. Em destaque, diferenças estruturais chaves.

Os tecidos vegetais variam enormemente no conteúdo e composição de tococromanóis. Nos tecidos fotossintéticos, a forma predominante é α -tocoferol, porém este aparece como componente secundário nas sementes, para as quais γ -tocoferol é o isômero predominante. Enquanto tocoferóis estão amplamente distribuídos em quase todos os tecidos vegetais de todas as plantas, os tocotrienóis estão presentes em diferentes clados não-relacionados (de modo marcante em monocotiledôneas), sendo quase exclusivamente encontrados em sementes e frutos (Dellapenna & Last, 2006; Falk & Munné-Bosch, 2010). No caso dos cereais, a existência de uma preniltransferase tocotrienol específica (HGGT) explicaria a capacidade dessas espécies de acumular tocotrienóis (Cahoon *et al.*, 2003). O acúmulo de tocotrienóis nas sementes sugere uma função específica para esses compostos que difere da função de tocoferóis nas folhas, mas que, até o momento, não foi compreendida (Falk & Munné-Bosch, 2010).

Em plantas, assume-se que a biossíntese de tococromanóis ocorre nos plastídeos e, até o momento, não foi demonstrado a possibilidade de transporte para nenhuma das isoformas de VTE dentro do corpo vegetal (Sun *et al.,* 2009).

Na via de biossíntese da VTE, a redução do hidroxifenilpiruvato a homogentisato inicia o conjunto de reações que compõem a "rota central da

VTE" na qual há sete enzimas envolvidas: p-hidroxifenilpiruvato desidrogenase (HPPD), homogentisato geranilgeranil transferase (HGGT), homogentisato fitil transferase (HPT/VTE2), dimetil-fitilquinolmetil transferase (MPBQMT/VTE3), γ-tocoferol metiltransferase (γ TMT/VTE4), tocoferol ciclase (TC/VTE1) e fitol quinase (PK/VTE5) (Li *et al.*, 2008). Após a ligação da cauda hidrofóbica ao anel cromanol pela VTE2 e posterior metilação, catalisada pela VTE3, o intermediário formado - 2,3-dimetil-5-fitil-1,4-hidroquinol (DMPQ) - é convertido em γ -tocoferol pela VTE1. O α -tocoferol é então produzido pela ação da VTE4. Tais enzimas descritas também são responsáveis pela síntese dos tocotrienóis (Li *et al.*, 2008). No caso dos tocoferóis, recentemente foi demonstrado que o precursor fitil 2P também é produzido a partir da conversão do fitol livre, produto da degradação da clorofila, pela VTE5 (Valentin *et al.*, 2005).

Ainda que a via biossintética de tocoferol tenha sido elucidada em 1979 (Soll & Schultz, 1979), a identificação dos genes envolvidos ocorreu apenas recentemente. Nas últimas décadas, por meio da combinação de abordagens genéticas e genômicas, os genes que codificam as enzimas para os passos da rota central da VTE foram identificados e clonados em várias espécies (Li *et al.,* 2008). Além disso, a caracterização de mutantes para VTE e plantas transgências tem contribuído de maneira significativa para o conhecimento da regulação da biossíntese de tococromanóis (Falk & Munné-Bosch, 2010; Mène-Saffrané & DellaPenna, 2009). Estes estudos sugerem que a VTE desempenha um papel para além de sua função antioxidante, que incluiria participação em diversos processos fisiológicos como germinação, senescência foliar, resposta ao stress abiótico, crescimento e particionamento de fotoassimilados (Falk & Munné-Bosch, 2010).

2.2. Controle da biossíntese de tococromanóis

A análise da via biossíntética dos tococromanóis revela que as enzimas podem influenciar no acúmulo e na composição da VTE (perfil dos isômeros) nos tecidos vegetais.

As primeiras observações relacionadas ao controle do fluxo da síntese de tocoferol envolveram os precursores homogentisato e fitil-2P. HPPD e GGDR foram selecionadas como enzimas candidatas para mediar a regulação,

baseadas em evidências in vivo de que a disponibilidade desses precursores limitava a biossíntese de tococromanóis. Células de cártamo (Carthamus tinctorius) em meio de cultura aumentavam em 18 vezes o conteúdo de tocoferol após o suprimento exógeno de homogentisato e fitil-2P (Furuya *et al.*, 1987). Entretanto, como mencionado anteriormente, o fitil 2P utilizado na biossíntese de tocoferol também resulta da reciclagem do fitol, produto da degradação da clorofila. Foi observado em tecidos foliares que, o acúmulo de fitil-2P produzido a partir da degradação da clorofila devido à senescência induzida pelo escuro, era concomitante ao aumento na biossíntese de tocoferol (Rise et al., 1989). A identificação do mutante para vte5 de arabidopsis, que contém apenas 20% do conteúdo de tocoferol da planta selvagem em folhas e semente, foi essencial para elucidar a dupla origem da cadeia fitil do tocoferol. A expressão da proteína VTE5 em Escherichia coli demonstrou que a enzima fosforilava eficientemente fitol em fitil P. Os resultados indicaram que aproximadamente 80% e 65% da cadeia fitil utilizada na síntese de tocoferóis em sementes e folhas, respectivamente, têm origem da reciclagem, dependente de VTE5, do fitol livre produzido a partir da degradação de clorofila em arabidopsis (Valentin et al., 2005).

Por outro lado, para testar a hipótese de que os níveis de homogentisato são limitantes para a biossíntese de tocoferol, o gene que codifica para HPPD foi superexpresso em folhas e sementes de arabidopsis e tabaco (Tsegaye *et al.*, 2002). Embora as linhagens transgênicas apresentassem aumento na atividade enzimática, apenas um acréscimo discreto de tocoferol foi observado, sugerindo que o conteúdo de homogentisato poderia ser um gargalo da atividade da HPPD. Essa hipótese foi confirmada pela co-expressão constitutiva, em tabaco, dos genes *hppd* de arabidopsis e *PDT* de levedura, que codifica para enzima bifuncional prefenato desidratase insensível ao feedback negativo da tirosina. Assim, essas plantas transgênicas mostraram um considerável acúmulo de tococromanóis, especialmente de α -tocotrienol (Rippert *et al.* 2004).

Mudanças no fluxo da rota do MEP também impactam o conteúdo de α tocoferol. A superexpressão do gene *dxs* em arabidopsis aumentou em 40% o conteúdo de α -tocoferol em folhas (Estévez *et al.*, 2001). Outro ponto importante na regulação da via é o da enzima VTE2. Folhas superexpressando *vte2* em arabidopsis aumentam em 4.4 vezes os níveis de tocoferol total nas linhagens transgênicas (Collakova & DellaPenna, 2003a). Por outro lado, arabidopsis superexpressando *vte4* possui alteração apenas na composição dos isômeros; α -tocoferol torna-se o composto predominante, apesar de não ser observada alteração no conteúdo total (Shintani & Dellapenna 1998). Desta forma, quando *vte2* e *vte4* são co-expressos constitutivamente em arabidopsis, a atividade de VTE é elevada pelo acúmulo de α -tocoferol em folhas e sementes; há diferenças tanto no conteúdo total quanto na composição dos isômeros (Collakova & DellaPenna, 2003b).

Finalmente, plantas transgênicas de arabidopsis expressando constitutivamente *vte1* apresentaram sete vezes mais tocoferóis que as plantas controles, evidenciando também a importância dessa enzima como passo limitante na síntese de VTE (Kanwischer *et al.*, 2005).

MEP e SK, vias componentes plastidiais da biossíntese de tococromanóis, produzem precursores para uma ampla gama de compostos do metabolismo da planta. Enquanto a rota do MEP é responsável pela formação geranilgeranil 2P, precursor também da síntese do de clorofilas, plastoquinonas, giberelina, carotenóides entre outros (DellaPenna & Pogson, 2006); a rota do SK determina a síntese de precursores para as vias do folato, fenilpropanóides, alcalóides e de aminoácidos aromáticos - fenilalanina, tirosina e triptofano (Tzin et al., 2010). Em concordância com a diversidade de rotas metabólicas nas quais os intermediários das vias do MEP e SK estão envolvidos, vários trabalhos evidenciam a comunicação (crosstalk) entre essas rotas associadas e o conteúdo de VTE.

Plantas transgênicas de tomate que superexpressam em fruto o gene fitoeno sintase (*psy*), que codifica para enzima chave da biossíntese de carotenóides, apresentam aumento do conteúdo de tocoferol. Os autores argumentam que esses resultados podem ser explicados por uma alteração no equilíbrio de síntese/degradação da clorofila nas linhagens transgênicas. Desse modo, um aumento da degradação da clorofila aumentaria o influxo do precursor fitil para síntese de tocoferol, estimulando a produção desses compostos (Fraser *et al.*, 2007). Além disso, o conteúdo de tococromanóis
também é alterado quando porção pós-corismato da via do chiquimato é manipulada. Plantas transgênicas de arabidopsis expressando constitutivamente a enzima bifuncional bacteriana corismato mutase (CM)/ prefenato desidratase (PDT) apresentam aumento significativos nos níveis de fenilalanina, assim como de γ-tocoferol, γ-tocotrienol e outros metabótitos secundários (Tzin *et al.*, 2009).

A síntese de tococromanóis também é alvo do controle de sinais ambientais e endógenos. O silenciamento do fator de transcrição responsivo à luz DE-ETIOLATED1 resulta em frutos com aumento nos teores de antioxidantes, incluindo carotenóides, flavonóides e tocoferóis (Davuluri *et al.*, 2005; Enfissi *et al.*, 2010).

Como visto, apesar do rápido progresso nos últimos anos, há muitas lacunas na compreensão dos mecanismos regulatórios que envolvem a síntese, transporte e acúmulo dos tococromanóis em plantas. Esses avanços tem sido significativos quando os genes da via biossíntética são conhecidos e bem caracterizados nas espécies estudadas.

Com o intuito de contribuir para o conhecimento da regulação da VTE em tomate, esse estudo tem como objetivo fornecer uma base para associação entre sequências gênicas e conteúdo de tocoferol no fruto por meio da análise integrada de dados metabólicos, genômicos e genéticos. A abordagem escolhida para iniciar a dissecção da base genética envolvida na determinação do conteúdo de VTE foi o mapeamento de QTL na população de ILs de *S. pennellii,* seguido da identificação de genes candidatos. Os resultados obtidos serão discutidos à luz dos mecanismos envolvidos no controle da biossíntese de VTE, abrindo estratégias para futuros estudos funcionais.

II. Objetivos

O objetivo geral do deste trabalho foi caracterizar as regiões genômicas de *S. pennellii* responsáveis pelo aumento do conteúdo de vitamina E no fruto. Este objetivo geral foi contemplado nos seguintes objetivos específicos:

1. Caracterização da rota de biossíntese da vitamina E em tomate.

2. Mapeamento de QTL para conteúdo de tocoferol em fruto de tomate e identificação de genes candidatos.

3. Clonagem e análise dos alelos de *S. pennellii* para os genes candidatos identificados.

III. Materiais e Métodos

1. Caracterização da rota de biossíntese da vitamina E em tomate

1.1. Busca das sequências de cDNA de S. lycopersicum envolvidas na biossíntese de VTE

Primeiramente, o mapa metabólico completo da biossíntese da VTE, que inclui precursores do metabolismo primário requeridos pelas vias componentes do SK e MEP até os compostos com atividade de VTE, foi reconstruído a partir das informações disponíveis nos mapas de referência do banco de dados do KEGG (*Kyoto Encycopledia of Genes and Genomes*; http://www.genome.jp/kegg/) e das publicações recentes disponíveis.

Para cada reação enzimática identificada, foram obtidas as sequências protéicas de A. thaliana que possuíam atividade enzimática ("enzyme comission number" - EC number) correspondente no KEGG. Essas sequências foram então utilizadas como referência no levantamento das sequências consenso de cDNA (unigenes) ortólogas de S. lycopersicum por meio do algoritmo TBLASTN (Altschul et al., 1990) na plataforma de dados da SGN (Solanaceae Genomics Network; http://solgenomics.net/; 247,317 sequências; 173,392,588 letras totais), banco que reúne dados genômicos e genéticos das espécies de solanáceas. Os critérios aqui adotados na determinação da ortologia entre arabidopsis e tomate foram: identidade ≥ 40% entre as sequências de aminoácidos e cobertura ≥ 65% em relação à proteína de arabidopsis. No caso de unigenes incompletos, a porcentagem de cobertura adotada foi > 30%. Em seguida, uma busca por BLASTN (Altschul et al., 1990) na base de dados do genoma de S. lycopersicum - versão 2.31 [Tomato WGS] Scaffolds (SL2.31) - 526 sequências; 374,251,055 letras totais - The International Tomato Genome Sequencing Project, disponível em http://solgenomics.net/] foi feita com o objetivo de identificar as sequências genômicas correspondentes aos unigenes.

1.2. Predição da localização subcelular

A partir das sequências de aminoácido deduzidas (Translate Tool, http://expasy.org/tools/dna.html), a predição *in silico* da localização subcelular dos unigenes de tomate foi feita com base nos programas computacionais TargetP (http://www.cbs.dtu.dk/services/TargetP) e ChloroP (http://www.cbs.dtu.dk/services/ChloroP). Ambos os programas sugerem o direcionamento celular das proteínas de acordo com sua estrutura primária. O primeiro prediz se há a presença de uma sequência sinalizadora na extremidade N-terminal que pode direcionar o produto protéico à mitocôndria, ao cloroplasto ou à via de secreção. A sequência analisada pode ainda ser classificada como "*any other location*", que neste estudo, foi considerada como citossólica. Já o ChloroP indica se há ou não presença de peptídeo sinal para o cloroplasto.

1.3. Distribuição cromossômica dos genes identificados

A distribuição cromossômica dos genes envolvidos na biossíntese de VTE em tomate foi determinada a partir da presença de homologia entre as sequências identificadas de *S. lycopersicum* (unigene ou sequência genômica correspondente) e marcadores moleculares com posição já definida no mapa genético desta espécie (Tomato-EXPEN-2000, http://solgenomics.net/index.pl) baseada em buscas utilizando o algoritmo BLASTN. O mapa genético foi representado graficamente utilizando-se o software Map Chart 2.2 (Voorrips *et al.,* 2000).

2. Mapeamento de QTL para conteúdo de tocoferol em fruto de tomate e identificação de genes candidatos

A determinação dos níveis de tocoferol em fruto de tomate foi realizada em colaboração com Leandro Quadrana do "Instituto de Biotecnologia do Instituto Nacional de Tecnología Agropecuária (IB-INTA)" sob supervisão do Prof. Ramón Asis da "Facultad de Ciencias Químicas de la Universidad Nacional de Córdoba" em ambas instituições argentinas.

2.1. Material vegetal

As sementes de tomate de *Solanum lycopersicum* L. (cv M82) e das linhagens introgredidas de *Solanum pennellii* foram gentilmente cedidas pelo Tomato Genetic Resource Center (http://tgrc.ucdavis.edu). As plantas de tomate foram crescidas em casa de vegetação no IB-INTA, em potes de 20 litros contendo terra vegetal e vermiculita (1:1). As condições da casa vegetação foram: fotoperíodo de 16/8 h, 24 ± 3 °C, 60% umidade e fotoirradiância de 140 ± 40 µmol m⁻² s⁻¹. Seis frutos maduros (45 dias depois da antese) foram coletados de seis plantas das ILs 6-1, 6-2, 7-4, 7-4-1, 7-5, 8-2, 8-2-1, 9-1, 9-2-6, cujos tecidos foram imediatamente congelados em nitrogênio líquido e armazenados em freezer -80 °C até o uso.

2.2. Preparação dos extratos

Para extração de tocoferóis, utilizou-se o protocolo descrito por Fraser *et al.* (2000) com modificações. Frutos de tomate foram triturados a pó fino em nitrogênio líquido e 500 mg do material foram extraídos com 1.5 mL de metanol. Após agitação em vortex por 1 min, 1 mL de clorofórmio foi adicionado e a mistura obtida sonicada em gelo por 5 min. Na sequência, foram adicionados 1 mL de solução Tris (Tris (pH 7.5 50 mM)/NaCl 1M) a cada amostra, realizada agitação da mistura em vortex 1 min, seguida de centrifugação a 3000 rpm por 5 min. A fase clorofórmica foi então recuperada e o pellet submetido a nova re-extração com 2 mL de clorofórmio. Os dois extratos obtidos foram reunidos em um tubo falcon (15 mL) e o volume final ajustado a 4 mL com clorofórmio. Uma alíquota de 2 mL do extrato foi seca em nitrogênio gasoso, ressuspendida em 0.2 mL de solução hexano/isopropanol (99.5:0.5) e, antes da injeção, filtradas em papel filtro.

2.3. Determinação por HPLC

O conteúdo de tocoferol foi determinado utilizando o cromatógrafo líquido "Hewlett-Packard series 1100 HPLC system" acoplado a um detector de fluorescência (Agilent Technologies, série 1200). A separação cromatográfica foi realizada em coluna de fase normal Metasil Si (250 mm × 4.6 mm, 5 µm, Varian, Metachem, Torrance, CA) operada em temperatura ambiente utilizando-se um sistema de eluição isocrático (fase móvel) de 99.5:0.5 hexano/isopropanol com fluxo de 1 mL.min⁻¹. Os compostos eluídos foram detectados por fluorescência com excitação a 296 nm e emissão a 340 nm. Os diferentes isômeros de tocoferol foram identificados e quantificados por meio da comparação do tempo de retenção e das áreas dos picos dos padrões adquiridos comercialmente (Merck, tocopherol set Calbiochem, Cat. No. 613424). Uma curva de calibração diária foi realizada utilizando uma solução de tocoferóis com concentração variando entre 0.31 e 5 µg/mL para cada isoforma. Desse modo, foi feita uma correlação linear entre as áreas dos picos e as concentrações injetadas de cada composto. A equação de regressão linear obtida para cada isômero foi utilizada para quantificar o conteúdo dos tocoferóis nas amostras estudadas.

2.4. Análise estatística

Os dados de conteúdos de α -, β -, γ -, δ -tocoferol e tocoferol total foram avaliados estatisticamente de acordo com Sokal e Rohlf (1981). Quando o conjunto de dados obedeceu ao critério de homoscedasticidade, com ou sem transformação logarítimica ou de raiz quadrada, os dados foram submetidos à análise por ANOVA seguido pelo teste de Dunnett (p<0.05) para comparação dos conteúdos de tocoferol entre as ILs e o controle. Na falta de homoscedasticidade, uma comparação não-paramétrica foi feita por meio do teste Krustal Wallis (p<0.05 e 0.1). Os testes estatísticos foram realizados utilizando o programa Bioestat 5.0 (Ayres *et al.,* 2007) e InfoStat v. 2009 (Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina).

2.5. Mapeamento de QTL

As regiões genômicas associadas aos QTL para VTE nos cromossomos 6, 7, 8 e 9 foram determinadas de acordo à posição dos marcadores que flanqueiam as regiões introgredidas de cada IL no mapa genético Tomato-EXPEN-2000 (http://solgenomics.net/index.pl). Os dados foram confrontados com as quantificações de VTE descritas previamente por Schauer *et al.* (2006) e as disponíveis no Tomato Functional Genomics Database (http://ted.bti.cornell.edu).

2.6. Identificação de genes candidatos relacionados à biossíntese de VTE

Ao longo das regiões cromossômicas onde se localizaram os QTL, foram triados *in silico* todos marcadores genéticos com posição definida. As sequências dos marcadores identificados foram usadas como referência na busca de unigenes no banco de dados da SGN (*All SGN Unigene sequences*; 247,317 sequências; 173,392,588 letras totais). A predição da função do produto gênico foi feita com base na homologia entre a sequência do unigene e proteínas homólogas previamente caracterizadas. Para isso, foi realizado uma análise por BLASTX (Altschul *et al.*, 1990) das sequências identificadas de tomate à luz do banco de dados de proteínas não-redundantes (nr) do NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/). Um gene foi considerado candidato se a comparação entre sua sequência de aminoácidos e de uma proteína de função conhecida apresentava: (i) identidade de aminoácidos \geq 40%; (ii) cobertura \geq 65% para unigenes completos ou \geq 30 % para unigenes incompletos; (iii) função cujo efeito pudesse direta ou indiretamente interferir no metabolismo da VTE.

3. Clonagem e caracterização dos alelos de *S. pennellii* para os genes candidatos identificados

3.1. Material vegetal

Os alelos selvagens dos genes candidatos foram clonados a partir das linhagens introgredidas de *S. pennellii*. Foram cultivadas plantas das seguintes linhagens: 6-2, 7-4-1, 8-2, 9-1, 9-2. As plantas de tomate foram cultivadas em casa de vegetação no Instituto de Biociências, Universidade de São Paulo, em vasos plásticos de 15 litros contendo terra vegetal e vermiculita (1:1) sob condições naturais de temperatura e iluminação. Para fins de clonagem, foram coletados tecidos de folha fonte (completamente expandida) os quais foram congelados imediatamente em nitrogênio líquido e armazenados a -80°C.

3.2. Obtenção de bactérias competentes - Escherichia coli

A obtenção de bactérias competentes foi realizada segundo descrito por Hanahan et al. (1983), com modificações. A partir de uma colônia isolada de E.coli, cepa DH10B, foi cultivado um pré-inóculo com 2 mL de meio SOB durante 18 horas a 37°C sob agitação de 200 rpm constante. O pré-inoculo foi transferido para 50 mL de meio LB e incubado a 37°C sob agitação constante até atingir entre 0,4 a 0,6 de densidade óptica (DO) a 600 nm. Em seguida, adicionou-se 0,5 mL de MgCl₂ (2 M) e a cultura foi resfriada em gelo por 15 minutos. Seguiu-se uma centrifugação suave a 3000 rpm (Centrífuga MPW-350R, rotor nº 11469) por 20 minutos a 4°C. O sobrenadante foi descartado, o precipitado ressuspendido em 15 mL de solução RFI e incubado em gelo por 10 minutos. Repetiu-se a centrifugação a 3000 rpm por 20 minutos a 4°C, sendo o novamente o sobrenadante desprezado. As bactérias foram ressuspendidas em 2 mL da solução RFII e a suspensão de células distribuída em microtubos contendo 50 µL de volume. Essas alíquotas foram imediatamente congeladas em nitrogênio líquido e armazenadas em freezer a -80°C.

LB (Luria Bertani): 10 g/L de bactotriptona, 5 g/L de extrato de levedura e 10 g/L de NaCl. Ajustar o pH 7,0 com NaOH 5N. Esterilizar na autoclave por 20 minutos a 120°C.

SOB: 20 g/L de bactotriptona, 5 g/L de extrato de levedura, 0,5 g/L NaCl, 2,5 mM KCl e 1 mM MgCl₂. Esterilizar na autoclave por 20 minutos a 120°C.

Solução RFI: Acetato de potássio pH 6,9 (30 mM), cloreto de manganês (50 mM), cloreto de potássio (100 mM), cloreto de cálcio (10 mM) e glicerol (15%). Esterilizar em autoclave por 15 minutos a 120°C.

Solução RFII: MOPS pH 7,0 (10 mM), cloreto de cálcio (75 mM), cloreto de potássio (10 mM) e glicerol (15%). Esterilizar em autoclave por 15 minutos a 120°C.

3.3. Extração de DNA genômico de tomate.

Para a extração de DNA de tomate foi usado o método de CTAB descrito por Bernatzky & Tanksley (1986) com modificações. Primeiramente, o material vegetal foi macerado em cadinho com nitrogênio líquido e PVPP

(polivinilpolipirrolidona, 1,5 g de PVPP para cada 3 g de material vegetal) e transferido para um tubo falcon contendo o tampão CTAB (ver preparação abaixo). Para cada 3 g de material vegetal foi usado 18 mL de tampão CTAB + 36 μL de β-mercaptoetanol, cuja preparação foi incubada em banho-maria a 65°C por 30 minutos, invertendo o tubo cada 5 minutos. Em seguida, foi adicionado 18 mL de clorofórmio: álcool isoamílico (24:1) e preparação levada à centrífuga a 5000 rpm (Centrífuga MPW-350R, rotor nº 11469) por 20 minutos em 22ºC. Após a transferência da fase aquosa para um tubo novo, foi adicionado o dobro do volume de fenol equilibrado:clorofórmio:álcool isoamílico (25:24:1). Em seguida, a fase aquosa foi transferida para um tubo falcon (15 mL) na qual foi adicionada o dobro de volume da solução de clorofórmio: álcool isoamílico (24:1). Após mais uma etapa de centrifugação (20 minutos, 5000 rpm a 22°C) a fase aquosa foi isolada. Para precipitação do DNA, acrescentouse 2/3 do volume da amostra de isopropanol seguido de mistura por inversão. A amostra foi então centrifugada a 3000 rpm por 5 a 10 minutos a 22ºC e o isopropanol descartado. O precipitado foi lavado com etanol 70% e centrifugado por 3 min a 3000 rpm. Finalmente, a amostra foi submetida à centrífuga a vácuo, para evaporação do etanol e, em seguida, ressuspendida em 100 µL de TE (10 mM de Tris-HCI (pH 8,0), 1mM de EDTA (pH 8,0). A integridade do DNA purificado foi verificada por eletroforese em gel de agarose.

Tampão CTAB: 4 g de CTAB (brometo de cetil-trimetilamônio), 56 mL de cloreto de sódio 5 M, 8 mL de EDTA 0,5 M, 20 ml de Tris pH 8 e 2 g de PVP (polivinilpirrolidona) em 200 mL de volume final. A solução foi esterilizada em autoclave. Posteriormente foi adicionado o β-mercaptoetanol em uma concentração final de 10 mM.

3.4. Eletroforese em gel de agarose

Para a visualização e análise de DNA genômico, plasmidial, produtos de PCR e RNA total foram realizadas eletroforeses em gel de agarose. Dependendo do tamanho dos fragmentos, a concentração de agarose utilizada foi 0,8 % ou 1,0 %, dissolvida em tampão TBE 1/2X (TBE 10x: 108 g/L de Trisbase, 55 g/l de ácido bórico e 40 ml/l de EDTA 0,5M pH 8,0).

As amostras foram aplicadas no gel após adição de solução tampão (azul de bromofenol 0,25%; xileno cianol 0,25%, glicerol 30%, água deionizada). Após a migração eletroforética, os géis foram corados em solução de brometo de etídio (Fluka, Cat. No. 46067) a 0.5 µg/mL, lavados em água destilada e visualizados em transiluminador UV (Vilber Lourmart). As imagens foram digitalizadas em fotodocumentador (Vilber Lourmart).

3.5. Iniciadores

O desenho dos iniciadores específicos para a amplificação dos genes de interesse foi realizado sobre as sequências dos respectivos unigenes utilizando-se o programa Oligo Analyzer 3.1 (http://www.idtdna.com). Ao todo, 16 pares de oligonucleotideos foram desenhados (Tabela 2) contendo, pelo menos, 40 % de GC, e evitando a formação de dímeros e grampos estáveis. Sempre que possível, os iniciadores foram ancorados de forma de incluir os códons de início (ATG) e terminação segundo a predição in silico (Translate Tools, http://www.expasy.ch/tools/dna.html). Verificou-se a funcionalidade dos pares de iniciadores a partir da amplificação utilizando DNA genômico de S. lycopersicum L (cv M82) como molde. Para cada reação de PCR de 50 µL, utilizou-se 50 ng de DNA, 2 µL da polimerase Elongase Enzyme Mix (Invitrogen, 10480-010), 0.2 mM de cada dNTP, 10 pmoles de cada primer, 5 μ L do Tampão A (300 mM Tris-SO₄, 90 mM (NH₄)₂SO₄ e 5 mM MgSO₄), 5 μ L do Tampão B (300 mM Tris-SO₄, 90 mM (NH₄)₂SO₄ e 10 mM MgSO₄) atingindo uma concentração final de 1,5 mM de MgCl₂. A reação foi processada em termociclador gradiente (Biocycler, modelo MJ96G) por 35 ciclos que correspondem: desnaturação a 94ºC por 30 s, temperatura de anelamento específica para os iniciadores utilizados por 30 s, extensão a 68ºC por 5 a 7 min., dependendo do fragmento. Esse procedimento permitiu corroborar o tamanho predito in silico das sequências genômicas por meio da análise do tamanho do produto de amplificação observado no gel.

Canad		Direto (5'→3')	Anelamento	Fragmento
Genes	Unigene	Reverso $(5' \rightarrow 3')$	(°C)	(pb)
ggps4	U575882	CGGAGATCAATTATGGTTTTCTC GCCAAGTCTAGATCATGTTTAGC	56	1029
hppd1	U580457	ATGGGCAAACAAGCCGCTG CCATCGAAGAATATGAAAAGACACTTG	55	1291
tyral	U570951	CCATGTTGTCTTTCACCCCAC GTTAAACCTGAGAACAACTGAGGG	56	1139
tyra2	U570951	CACATTCAACCATGTTTTCCC GGAACTAGAGGCACTTTCTTGA	56	1184
tat2	U563404	CATGGAGAACGGTACGACTAC GCTTATTGTTTCTTAGCATGCC	56	1272
vte2	U327540 (5′) U576207 (3′)	ATGGAATCTTTGCTTATTGGGTC GTGAGGTGAATGAATGTACAAGC	55	1223
vte3(1)	U578249	GCATTTGCTCAACTAATCATGGC GAACCTCTGTGAGAGACCAAG	54	1047
vtel	U570602	ATGGAGTGTGTTTCCGCCATTTC GCCTTTGAGCAAAATGCAAAAACTAG	57	1515
vte4	U584511	GTAATGGGCAGCCAATGCTATTC CATCACATGTCGAAAACCTGAATAA	52	1080
vte5	U583081	ATGGCTCTCTTTGCTTCTTCTG GTTGGCAGTCTTGTTTTATGATACC	55	885
sec14	U583419	ATGGATTCTTCATCCCCCATTAC CAGCCTAAAATGAAGCAAATCTATG	53	1275
chl	U574853	ATGGTGATAACTAGTACTTCCTCTG GGTTGTGGTTATATTGTACCCTTCG	50	969
prai	U564371	GGAAGGATGCTCTCAGTGTTC CGCAGTGAAATCTTTACACCTTTAG	58	828
fpgs	U581922(3')	CTGTAGTGTGAAGAACAAAAAATG GCTGACTCCTATTTGGTAAGAG	55	1457
apt	U566340	CATGGCTCTGTTTCAGTGTCG CGCGTTGATAATCTGGCTGA	54	1098
lycb	U570109	CTATAATGGAAACTCTTCTCAAGCC CTAGCAATAGAGAGCCTTTGAATG	57	1515

Tabela 2: Oligonucleotídeos específicos utilizados para amplificação dos genes candidatos.

3.6. Extração de RNA e síntese de cDNA.

Utilizou-se para a extração de RNA o reagente Trizol (Invitrogen, Cat. No.15596-026). As amostras de RNA foram ressuspendidas em 20 µL de água DEPC 0,1%. A concentração e a integridade do RNA foram avaliadas por espectrofotômetro Nanodrop ND-100 (Nanodrop Technologies) e por eletroforese em gel de agarose, respectivamente. Para a remoção de traços contaminantes de DNA das amostras, 1 µg de amostra foi submetida a

tratamento com DNAse I (Invitrogen, Cat. No.18068-015), segundo as especificações do fabricante. Na sequência, o cDNA foi sintetizado a partir do RNA tratado utilizando o kit SuperScript[®] III Reverse Transcriptase (Invitrogen, Cat. No. 18080-044), utilizando-se oligo DT como iniciador (Invitrogen) de acordo com o protocolo recomendado. A análise da efetividade da ação da DNAse nas amostras sintetizadas de cDNA foi feita por meio de PCR, utilizando iniciadores que se anelam em éxons diferentes do gene de actina TGGCATCATACCTTTTACAA 3' (senso 5' е reverso 5' TCCGGGCATCTGAACCCTCT 3'). As reações de PCR foram conduzidas empregando-se: 1X tampão, 0.2 mM de cada dNTPs, 0.2 mM de cada iniciador, 100 ng de cDNA e 1 unidade de Taq DNA polimerase (Invitrogen Cat. No.11615). O ciclo de amplificação utilizado foi de 94 °C por 10 min.; 35 ciclos de 94 °C por 30s, 50°C por 30s, 72 °C por 1 min; e uma extensão final a 72 °C por 10 min. Como controle positivo foi utilizado 50ng de DNA genômico. Dessa forma, o tamanho do fragmento amplificado a partir de DNA genômico corresponderia a 812 pb e de cDNA a 521 pb. A presença da banda com 812 pb na amostra de RNA indicava contaminação por DNA genômico e assim a necessidade de mais um tratamento com DNAse.

3.7. Reação de Polimerização em Cadeia (PCR)

A amplificação dos fragmentos completos de cDNA foram gerados por PCR utilizando a polimerase de DNA Taq Platinum *pfx* (Invitrogen, Cat. No.11708-013) que possui atividade *proofreading* 26 vezes maior que a Taq polimerase comum. Foram utilizados 150 ng de cDNA da IL correspondente, 0.2 mM de cada dNTP, 0.2 mM de cada iniciador, 1 mM de Mg₂SO₄, e 2,5 unidades de enzima. As condições de amplificação empregadas foram: 94 °C por 3 min; 35 ciclos de 94 °C por 15 s, temperatura de anelamento por 30 s, 68 °C por 4 min; e uma extensão final a 68 °C por 10 min.

3.8. Clonagem dos produtos de PCR

Purificação dos produtos de PCR e ligação no vetor

Os fragmentos amplificados com tamanhos esperados foram purificados da banda isolada do gel de agarose ou diretamente da reação de PCR com o kit "GFXTM PCR DNA and Gel Band Purification Kit" (Amersham Biosciences, Cat. No. 289034-70) seguindo as instruções do fabricante. Uma vez purificados, os produtos de PCR foram quantificados utilizando marcador de massa de concentração conhecida (High DNA Mass Ladder, Invitrogen, Cat. No. 10496016) por meio de eletroforese em gel de agarose 0,8%. Esses produtos foram submetidos à reação de ligação no vetor pCR-Blunt[®] II TOPO disponível no kit TOPO-Zero Blunt (Invitrogen, Cat. No. 45-0245) cujas condições foram: cerca de 50 ng (1 a 4 μ L) dos fragmentos de PCR purificados, 1 μ L de solução salina, 1 μ L de vetor e água milliQ até um volume final de 6 μ L. A reação foi incubada por 30 minutos à temperatura ambiente.

Transformação de E. coli por choque térmico

As transformações foram feitas adicionando-se 2 µL do produto de ligação a uma alíquota de 50 µL célula bacterianas E. coli competentes (ver item 3.2). Após incubação de 30 minutos no gelo, as células foram submetidas ao choque térmico (45 segundos a 42°C, seguidos de 2 minutos em gelo). Na sequência, foram adicionados 500 µL de meio LB (ver item 3.2) para a etapa de recuperação. As bactérias foram mantidas por 40 minutos sob agitação de 250 rpm a 37 °C. Posteriormente, 300 µL de cada cultivo foram plaqueados em meio LB sólido suplementado com 50 µg/mL de kanamicina, 40 µg/mL de X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) e 250 µg/mL de IPTG (isopropyl-beta-D-thiogalactopyranoside) selecão para das colônias transformantes. As placas de petri permaneceram em estufa a 37 °C por 18 horas.

Estoques de glicerol e minipreparações de DNA plasmidial

Os clones transformantes que apresentaram coloração branca foram cultivados em 2 mL de meio LB suplementado com 50 µg/mL kanamicina, crescido por 16 horas a 37º C sob agitação constante. Uma alíquota de 500 µL do cultivo bacteriano foi colocada em tubos criogênicos contendo 500µL de glicerol 50% para a obtenção de estoque mantido em freezer a -80 °C. O restante do cultivo foi submetido à extração do DNA plasmidial em pequena escala utilizando-se o kit Qiagen Miniprep (Qiagen, Cat. No. 27106). A integridade do DNA purificado foi verificada por eletroforese em gel de agarose.

3.9. Reação de sequenciamento

Para a reação de sequenciamento, foram utilizados 2 µL de DNA (250-500 ng); 10 pmoles de cada iniciador - SP6 (5' ATTTAGGTGACACTATAG 3') e T7 (5' TAATACGACTCACTATAGGG 3'); 2 μ L de BigDyeTM Terminator Sequencing Kit (Applied Biosystems), 2 µL de tampão de sequenciamento[™] e água milliQ até um volume final de 10 µL. A reação foi processada em termociclador (Biocycler, modelo MJ96G) sob as seguintes condições: 3 min a 96°C mais 40 ciclos que compreendem denaturação a 96°C por 10 s, anelamento a 50°C por 5 s e extensão a 60°C por 4 min. Após a amplificação, foi feita a purificação do produtos da reação. Para cada 10 µL de reação de sequenciamento, adicionou-se 1µL de EDTA 125mM; 1µL de acetato de sódio 3M; 25 µL de etanol 100%. Os tubos foram incubados por 15 min à temperatura ambiente e em seguida, levados para centrifugação a 14000 rpm por 25 min a 4°C. O sobrenadante foi descartado, acrescentou-se 35 µL de etanol 70% em cada um dos tubos e seguiu-se para nova centrifugação a 14000 rpm por 15 min a 4°C. Descartado o sobrenadante, a secagem completa das amostras foi feita em Speed Vac (Eppendorf, Concentrator 5301) durante 1 hora. Estas foram então ressuspendidas em 10 µL de High Dye (tampão de corrida) para a leitura no sequenciador ABI3130 (Applied Biosystems).

3.10. Análise das sequências e comparação dos alelos

Todos os fragmentos foram sequenciados em ambos os sentidos e as sequências obtidas foram analisadas e editadas manualmente, quando necessário, com o auxílio do programa CHROMAS LITE versão 2.0, sendo depositadas no GenBank sob os número de acesso: HQ014366 a HQ014383; HQ219713 a HQ219716. A detecção de polimorfismos foi feita a partir do alinhamento das sequências de nucleotídeos e aminoácidos dos alelos de *S. lycopersicum* e *S. pennellii* utilizando o programa MULTIALIN (http://www-archbac.u-psud.fr/genomics/multalin.html; Corpet, 1988).

IV. Resultados

1. Caracterização da rota de biossíntese da vitamina E em tomate

Os compostos com atividade de VTE, tocoferóis e tocotrienóis, são produtos de convergência das vias plastidiais do MEP e do SK. Com o intuito de identificar cada reação enzimática envolvida na formação de tococromanóis em plantas, o mapa metabólico completo da biossíntese, desde os precursores do metabolismo primário requeridos pelas vias do SK e do MEP, foi reconstruído baseado nos dados disponíveis no KEGG e na literatura científica recente. Após a reconstrução metabólica, foram identificadas 29 reações bioquímicas envolvidas na biossíntese de VTE, as quais são catalisadas por 28 enzimas. Somente um passo da via representada na Figura 5 permanece ainda obscuro no metabolismo das plantas, o qual envolve a fosforilação de fitil fosfato a fitil 2P. Esta rota alternativa de síntese do precursor fitil 2P, inicia-se a partir da remoção da cadeia lateral da clorofila, o fitol. Até o momento, só é reconhecida a atividade da enzima VTE5, responsável pela fosforilação do fitol a fitil P (Valentin et al., 2006), permanecendo, portanto, indeterminada qual enzima catalisa o passo subsequente dessa via, adição do grupo fosfato ao fitil fosfato. Os loci que codificam as enzimas de A. thaliana se encontram listados na tabela 3. Para cada um deles, foi feito um levantamento bibliográfico dos dados funcionais experimentais e, então, determinada a localização subcelular curada das enzimas nessa espécie.

Após uma busca extensiva no banco de dados de unigenes do Solanaceae Genomics Network (http://solgenomics.net) baseada em critérios de similaridade utilizando as sequências de arabidopsis como referência, o conjunto de genes que codificam as enzimas da rota de biossíntese da VTE em tomate foi identificado (Figura 6). As buscas permitiram a determinação de 41 diferentes *loci*. Para 11 enzimas, o número de genes identificados em tomate difere daqueles descritos em *A. thaliana*, de acordo com o critério adotado neste estudo (Tabela 3).



respectivamente As rotas relacionadas à VTE (carotenóides, clorofila, triptofano, fotato) onde se localizam os genes candidatos identificados não estão destacadas. As enzimas estão nomeadas de acordo com a abreviação da tabela 3. Genes cujos alelos da espécie selvagens foram clonados, sequenciados Figura 5. Mapa da biossíntese de VTE. As rotas de síntese de MEP, SK e central de tocoferol estão destacadas em vermelho, verde e azul, e analisados estão sublinhados.

```
Query= gi|15005009|dbj|BAB62076.1| APG1 [Arabidopsis thaliana]
      (338 letters)
Database: Lycopersicon Combined (Tomato) Unigenes
      42,257 sequences; 35,767,302 total letter
                                                                                                                 Score
                         Sequences producing significant alignments:
                                                                                                                 (bits) value
 SGN-U581492 Tomato 200607#2 [118 ESTs aligned] (*GB) gi|1419090|emb|C... 489
                                                                                                                        1e-138
 SGN-U578249 Tomato 200607#2 [27 ESTs aligned] (*GB) gi|157348021|emb|... 482 1e-136
 SGN-U581056 Tomato 200607#2 [15 ESTs aligned] (*GB) gi|1419090|emb|CA... 370
                                                                                                                        1e-103
 SGN-U578217 Tomato 200607#2 [9 ESTs aligned] (*GB) gi|157348021|emb|C... 243 1e-64
 > SGN-U581492 gi[1419090]emb[CAA64422.1 @ Tomato 200607#2 [118 ESTs aligned] (*GB) |
(e_value=7e-169) 37xDa chloroplast inner envelope membrane polypeptide precursor [Nicotiana
tabacum]; (*ATH) AT3G63410.1 (e_value=8e-142) Symbol: None | chloroplast inner envelope
membrane protein, putative (APG1), similar to SP:P23525 37 kDa inner envelope membrane
protein, chloroplast precursor (E37) {Spinacia oleracea}; contains Pfam profile PF01209;
methlytransferase, UbiE/COQ5 fa...; (*SWP) sp|P23525|IN37_SPIOL (e_value=4e-145) 37 kDa
inner envelope membrane protein, chloroplastic OS=Spinacia oleracea;
      Length = 1,535
Score = 489 bits (1258), Expect = 1e-138
Identities = 245/345 (71%), Positives = 275/345 (79%), Gaps = 7/345 (2%), Frame = +3
Query: 1 MASLMLNGAITF-----PKGLGSPGSNLHAKSIPRPTLLSVTRTSTPRLSVATKCXXX 53
MA3 +L+G+ F P G GS L+ K +P+ +++ T T R TKC
Sbjct: 240 MASSILSGSDNFKLLSGISPSGSSFLGSELNLKCLPQKGFVNLRATRTLR---PTKCSLS 410
Query: 54 XXXXXXXAQPRFIQHKKEAYWFYRFLSIVYDHVINPGHWTEDMRDDALEPADLSHPDMR 113
+QPRFIQHKKEA+WFYRFLSIVYDHVINPGHWTEDMRD+ALEPA+L +++
Sbjct: 411 ASRPA---SQPRFIQHKKEAFWFYRFLSIVYDHVINPGHWTEDMRDEALEPAELHARNLQ 581
Query: 114 VVDVXXXXXXXXXXXXVIVKTVKAKNVTILDQSPHQLAKAKQKEPLKECKIVEGDAEDLPFPT 173
VVDV IVK V AKNVTI+DQSPHQLAKA+QKEPLKECKI+EGDAEDLPFPT
Sbjct: 582 VVDVGGGTGFTILGIVKHVDAKNVTIIDQSPHQLAKARQKEPLKECKILEGDAEDLPFPT 761
Query: 174 DYADRYVSAGSIEYWPDPQRGIREAYRVLKIGGKACLIGPVYPTFWLSRFFSDVWMLFPK 233
D DRY+BAGSIEYWPDPQRGI+EAYRVL IGG ACLIGPVYPTFWLSRFF+DVWMLFPK
Sbjet: 762 DTFDRYIBAGSIEYWPDPQRGIKEAYRVLTIGGVACLIGPVYPTFWLSRFFADVWMLFPK 941
Query: 234 EEEYIEWFKNAGFKDVQLKRIGPKWYRGVRRHGLIMGCSVTGVKPASGDSPLQLGPKEED 293
EEEYIEWFK AGF V+LKRIGPKWYRGVRRHGLIMGCSVTGVKP+46+SPLQLGFK ED
Sbjet: 942 EEYIEWFKKAGFTQVKLRKIGFKWKGVRRHGLIMGCSVTGVKPYAGESPLQLGFKVED 1121
Query: 294 VEKPVNNPFSFLGRFLLGTLAAAWFVLIPIYMWIKDQIVPKDQPI 338
V KPV NPF FL RFLLG +AA+++VL+PIYMWIKDQIVPK QPI
Sbjct: 1122 VSKPV-NPFVFIMRFLLGIMAASYYVLVPIYMWIKDQIVPKGQPI 1253
```

Figura 1. Exemplo de resultado obtido para o gene vte3 nas buscas pelos loci de S. *lycopersicum* no banco de dados da SGN a partir das proteínas de A. thaliana (tBLASTN). Nesse caso, a identidade da sequência de aminoácidos de tomate considerada ortóloga à sequência de arabidopsis é de 260/338 (77%) enquanto a cobertura corresponde a 100% (338/338).

A predição da localização subcelular dos produtos protéicos dos *loci* de *S. lycopersicum* identificados foi feita utilizando-se os programas TargetP e ChloroP. As enzimas para as quais a predição não revelou nenhum peptídeo sinal foram consideradas, portanto, citossólicas. Em geral, a predição *in silico* está de acordo com as evidências experimentais reportadas para os ortólogos de arabidopsis (Tabela 3), exceto para duas proteínas. Essa discrepância pode ser observada entre os ortólogos da enzima geranil pirofosfato sintase (GPPS), cuja proteína de tomate apresenta direcionamento para a mitocôndria em contrapartida à localização plastidial descrita em arabidopsis, e para a enzima

bifuncional SDH/DHQ em cujos *loci* analisados de tomate não foi detectado peptídeo sinal.

Além disso, nenhuma das CM identificadas de tomate apresentou peptídeo sinal direcionado para o cloroplasto, sendo consideradas citossólicas da mesma maneira que uma das sequências protéicas ortólogas de arabidopsis. Finalmente, as últimas enzimas envolvidas na porção pós-corismato da via do chiquimato, TAT e HPPD estão, aparentemente, localizadas no citossol nas células de tomate.

O segundo passo na caracterização genética da biossíntese de tococromanóis em tomate incluiu a determinação da posição cromossômica dos 41 *loci* envolvidos nas vias do MEP, SK e rota central da VTE no mapa de ligação TOMATO EXPEN 2000 (Tabela 3 e Figura 9). O mapeamento foi baseado na ligação física entre marcadores de posição conhecida e as sequências de unigenes e/ou suas sequências genômicas correspondentes. Com exceção do cromossomo 12, todos os cromossomos de tomate possuem, pelo menos, um dos *loci* identificados. De modo particular, a maioria dos genes que codificam enzimas da rota central do tocoferol estão localizados em regiões genômicas dos cromossomos 7, 8 e 9, exceto por *hst e vte3*(2), que mapeiam no cromossomo 3, e *hppd*(2) localizado no cromossomo 5. Notavelmente, os quatro genes mapeados no cromossomo 9, *vte3*(1), *vte5*, *tyra*(2), e *ggps*(4) co-localizam com os QTL para α -tocoferol previamente descritos por Schauer *et al.* (2006).

Resultados

Tabela 3: Identificação dos genes de tomate envolvidos na biossíntese de VTE e genes candidatos associados a QTL. O padrão de cores responde ao apresentado na figura 5. Nd, não determinado.

ENZIMA ¹	ARABIDOPSIS THALIANA <i>LOCUS</i>	LOCALIZAÇÃO CURADA ²	UNIGENE TOMATE ³	PEPTÍDEO SINAL ⁴	ID GENÔMICA ⁵	MARCADOR LIGADO ⁶	POSIÇÃO CROMOSSÔMICA ⁷
1-deoxi-D-xilulose-5-P sintase EC 2.2.1.7 DXS	At4g15560 (717 aa)	cloroplasto	U567647 (1) U316204 (2)	cloroplasto nd	SL2.31sc05941 SL2.31sc03748	T1704 SSR67	1 (39 cM) 11 (24 cM)
2-C-metil-D-eritritol 4-P sintase EC 1.1.1.267 DXR	At5g62790 (477 aa)	cloroplasto	U585813	cloroplasto	SL2.31sc03701	C2_At5g23060	3 (102.5 cM)
2-C-metil-D-eritritol 4-P citidilitransferase EC 2.7.7.60 CMS	At2g02500 (302 aa)	cloroplasto	U566797	cloroplasto	SL2.31sc04323	TG528	1 (127.5 cM)
4-(citidina 5'-P)-2-C-metil-D- eritritol quinase EC 2.7.1.148 ISPE	At2g26930 (383 aa)	cloroplasto	U583224	cloroplasto	SL2.31sc04133	cTOC-4-C7	1 (19 cM)
2-C-metil-D-eritritol 2,4-ciclodifosfato sintase EC 4.6.1.12 ISPF	At1g63970 (231 aa)	cloroplasto	U568497	cloroplasto	SL2.31sc03923	C2_At3g27530	8 (78 cM)
4-hidroxi-3-metilbut-2-enil-difosfato sintase EC 1.17.7.1 HDS	At5g60600 (717 aa)	cloroplasto	U567167	cloroplasto	SL2.31sc03876	At5g60600	11 (79 cM)
4- hidroxi -3- metilbut -2-enil- difosfato reductase EC 1.17.1.2 HDR	At4g34350 (466 aa)	cloroplasto	U580658	cloroplasto	SL2.31sc04323	C2_At4g34350	1 (154 cM)
isopentenil difosfato ô-isomerase EC 5.3.3.2 IPI	At3g02780 (284 aa) At5g16440 (291 aa)	mitocôndria cloroplasto (Phillips <i>et al.</i> , 2008)	U577516 (1) U569721 (2)	cloroplasto cloroplasto	SL2.31sc06101 SL2.31sc03902	U49812 C2_At5g04270	4 (64 cM) 5 (112.7 cM)
geranil pirofosfato sintase EC 2.5.1.1 GPPS	At2g34630 (422 aa)	cloroplasto (Bouvier <i>et al.</i> , 2000)	U573523	mitocôndria	SL2.31sc03835	C2_At1g30360	8 (21.3 cM)
geranilgeranil pirofosfato sintase EC 2.5.1.29	At4g36810 (371 aa)	cloroplasto	U574849 (1) U571085 (2) U573348 (3)	cloroplasto cloroplasto nd	SL2:31sc03748 SL2:31sc04135 SL2:31sc03665	C2_At5g16710 C2_At1g19340 CT232	11 (31.4 cM) 4 (112 cM) 2 (90.1 cM) –
	At4g38460 (326 aa)	cloroplasto	U575882 (4)	cloroplasto	SL2.31sc03771	T0532	9 (30 cM)

3 (122 cM)	11 (26 cM) 4 (83.5 cM)	2 (83.4 cM)	1 (39 cM) 6 (101 cM)	4 (56 cM)	1 (57 cM)	4 (56.7 cM) 4 (21.5 cM)	2 (106 cM) 11 (45 cM)	4 (61 cM)	7 (0.4 cM) 9 (48 cM)
C2_At1G74470	T0408 T1560	C2_At3g01160	T1704 TG221	C2_At3g62940	C2_At2g45240	C2_At3g07950 TG370	T1480 TG147	C2_At2g22250	C2_At5g34850 T1212
SL2.31sc03701	SL2.31sc03748 SL2.31sc04135	SL2.31sc03665	SL2.31sc05941 SL2.31sc03622	SL2.31sc06101	SL2.31sc04323	SL2.31sc06101 SL2.31sc03604	SL2.31sc03665 SL2.31sc03748	SL2.31sc06101	SL2.31sc03731 SL2.31sc03771
cloroplasto	cloroplasto nd	cloroplasto	p p	cloroplasto	cloroplasto	cloroplasto cloroplasto	pe Pe	cloroplasto	cloroplasto cloroplasto
U564571	U581552 (1) U566921 (2)	U568781	U570855(1) U570070(2)	U582040	U577580	U563165(1) U563163(2)	U575627(1) U585231(2)	U567172	U567861 (1) U570951 (2)
cloroplasto	cloroplasto cloroplasto cloroplasto (Entus <i>et al.</i> , 2002)	cloroplasto	cloroplasto	nd cloroplasto	cloroplasto cloroplasto	cloroplasto	cloroplasto (Mobley <i>et al.</i> ,1999) cloroplasto (Eberhard <i>et al.</i> , 1996) citossol (Eberhard <i>et al.</i> , 1996)	cloroplasto	cloroplasto cloroplasto (Rippert <i>et al.</i> , 2009)
At1g74470 (467 aa)	At1g22410 (527aa) At4g33510 (432aa) At4g39980 (525 aa)	At5g66120 (442 aa)	At3g06350 (603 aa)	At2g21940 (276 aa) At4g39540 (300 aa)	At1g48860 (521 aa) At2g45300 (520 aa)	At1g48850 (380 aa)	At1g69370 (316 aa) At5g10870 (340 aa) At3g29200 (265 aa)	At2g2250 (475 aa)	At1g15710 (358 aa) At5g34930 (640 aa)
geranilgeranil reductase EC 1.3.1- GGDR	3-deoxi-D-arabino-heptulosonato-7-P sintase EC 2.5.1.54 DAHPS	3-dehidroquinato sintase EC 4.2.3.4 DHQS	chiquimato desidrogenase EC 1.1.1.25 SDH 3-dehidroquinato desidratase EC 4.2.1.10 DHQ	chiquimato quinase EC 2.7.1.71 SK	5-enolpiruvilchiquimato-3-P sintase EC 2.5.1.19 EPSPS	corismato sintase EC 4.2.3.5 CS	corismato mutase EC 5.4.99.5 CM	prefenato aminotransferase EC 2.6.1.57 PAT	arogenato desidrogenase EC 1.3.1.78 TyrA

os
g
Ē
su
ě
Γ.

10 (7.5 cM) 7 (43 cM)	7 (36.5 cM) 5 (70 cM)	3 (72.6cM)	7 (17 cM)-	9 (52 cM) 3 (4.6 cM)	8 (34 cM)	8 (41.8 cM)	9 (50.5 cM)	6 (59 cM)	6 (24 cM)	6 (26 cM)	9 (50.5 cM)
C2_At1g53000 C2_At1g03820	TG584 CLET-6-14	C2_At3g58490	сТОА-13-К15	T0565 TG324	C2_At4g32770	TG282	C2_At5g58240	C2_At5g17990	cLET-1-113	At5g41480	T1095
SL2.31sc05925 SL2.31sc03685	SL2.31sc03685 SL2.31sc03902	SL2.31sc06725	SL2.31sc03731	SL2.31sc04777 SL2.31sc04439	SL2.31sc04948	SL2.31sc03923	SL2.31sc03771	SL2.31sc05054	SL2.31sc05732	SL2.31sc05732	SL2.31sc03771
pu pu	p nd	cloroplasto	cloroplasto	cloroplasto cloroplasto	cloroplasto	cloroplasto	cloroplasto	cloroplasto	cloroplasto	mitocôndria	cloroplasto
U577103 (1) U563404 (2)	U580457(1) U578997(2)	U585005	U327540(5') U576207(3')	U578249(1) U581492(2)	U570602	U584511	U583081	U566340	U564371	U581922#	U583419
ри	citossol (Garcia <i>et al.</i> , 1999)	cloroplasto	cloroplasto	cloroplasto	cloroplasto	cloroplasto (Ferro <i>et al.</i> , 2009)	cloroplasto	cloroplasto	cloroplasto cloroplasto (Zhao e Last, 1995) cloroplasto	mitocôndria (Ravanel <i>et al.</i> , 2001)	membrana (Peterman <i>et al.</i> , 2004)
At5g53970 (414 aa)	At1g06570 (473 aa)	At3g11945.2 (393 aa)	At2g18950 (393 aa)	At3g63410 (338 aa)	At4g32770 (488 aa)	At1g64970 (348 aa)	At5g04490 (304 aa)	At5g17990 (444 aa)	At1 g07780 (275 aa) At1 g29410 (244aa) At5g05590 (275 aa)	At5g41480 (530 aa)	At3g51670 (409 aa)
tirosina aminotransferase EC 2.6.1.5 TAT	4-hidroxifenilpiruvato dioxigenase EC 1.13.11.27 HPPD	homogentisato geranilgeranil transferase/ homogentisato solanesil transferase EC 2.5.1 HGGT/HST	homogentisato fitil transferase EC 2.5.1 HPT (<i>vte2</i>)	dimetil-fittilquinol metil transferase EC 2.1.1 MPBQMT (<i>vt</i> e3)	tocoferol ciclase EC 5.3 TC (<i>vte1</i>)	γ-tocoferol C-metil transferase EC 2.1.1.95 γ-TMT (<i>vte4</i>)	fitol quinase EC 2.7 PK (<i>vt</i> e5)	antranilato fosforribosiltransferase EC 2.4.2.18 APT	fosforribosilantranilato isomerase EC 5.3.1.24 PRAI	folilpoliglutamato sintase EC 6.3.2.17 FPGS	transportador de fosfolipídeos SEC14

6 (32 cM)	6 (74 cM)	
T0834	cLET-19-J2	
SL2.31sc05732	SL2.31sc05054	
pu	cloroplasto	
U574853	U570109	
citossol (Schenk <i>et al.</i> , 2007)	cloroplasto (Ferro <i>et al.</i> , 2009)	oreviação.
At5g43860 (318 aa) At1g19670 (324 aa)	At3g10230 (369 aa)	/ww.genome.jp/kegg/) e al
clorofilase EC 3: 1.1.14 CHL	licopenoβ-ciclase LYCB	Nome da enzima e código (EC), de acordo com KEGG (w

² Localização da proteína de A. *thaliana* de acordo com The Plant Proteome Database (ppdb.tc.cornell.edu, Sun *et al.*, 2009) ou referência específica. ³ Número do unigene de tomate de acordo com SGN (solgenomics.net).

⁴ Predição da localização subcelular de acordo com os programas TargetP 1.1 (www.cbs.dtu.dk/services/TargetP/, Emanuelsson *et al.*, 2007) e ChloroP (www.cbs.dtu.dk/services/ChloroP). ⁵ Scaffold de S. *lycopersicum* que contém o gene correspondente (versão 2.31).

⁶ Marcador mais próximo ligado ao gene dentro do scaffold. ⁷ Posição cromossômica e genética no mapa Tomato-EXPEN 2000 v52. # 5' unigene incompleto.

2. Mapeamento de QTL para conteúdo de tocoferol em fruto de tomate e identificação de genes candidatos

2.1. Mapeamento de QTL

Em estudo anterior, Schauer et al. (2006), através da análise dos perfis metabólicos das ILs por GC-MS, descreveram dois QTL para o conteúdo de αtocoferol em fruto maduro localizados no cromossomo 6 e 9. Por outro lado, nossos resultados da localização cromossômica dos genes revelaram que, embora onze cromossomos possuam loci que codificam enzimas da via de síntese da VTE, a maioria dos genes da rota central estão reunidos nos cromossomos 7, 8 e 9. Diante do cenário revelado pelo mapeamento, levantouse a hipótese de que a presença desses genes nas regiões mencionadas poderia influenciar no acúmulo das diferentes isoformas de tocoferóis nos frutos, fato que poderia ser acessado pela avaliação detalhada da variabilidade fenotípica das linhagens com introgressões nos cromossomos anteriormente citados. Nesse sentido, a fim de investigar a presença de outros possíveis QTL para os diferentes isômeros de tocoferol, foi determinado o conteúdo desses compostos por HPLC de fase normal, método capaz de separar e quantificar os diferentes vitâmeros. As buscas por QTL na população das ILs concentrou-se, portanto, nas regiões identificadas que co-localizam com os genes da rota central da VTE e aquelas já associadas a QTL por Schauer et al. (2006). Dessa forma, foram obtidos perfis completos dos guatro isômeros (α -, β -, γ -, δ -) de tocoferol em fruto das ILs 6-1, 6-2, 7-4, 7-4-1, 7-5, 8-2, 8-2-1, 9-1, 9-2-6 e nos frutos controle correspondentes de S. lycopersicum (cv M82) além de tocoferol total.

A Figura 7 apresenta um perfil cromatográfico típico obtido para uma mistura das soluções padrão dos quatro isômeros de tocoferóis, uma amostra de M82 e da IL 9-2-6. Verifica-se que o método escolhido de extração/separação das amostras permitiu resolver e quantificar os compostos de interesse. Os perfis cromatográficos mostram que o α-tocoferol foi o composto principal, enquanto os demais isômeros ocorreram em menores quantidades. Para todas as ILs analisadas, os níveis de cada isoforma e suas razões são apresentados na Figura 8 e Tabela 4.



Figura 7. Cromatograma obtidos por HPLC para a quantificação dos isômeros de tocoferol. (a) mistura dos padrões de tocoferóis; (b) amostra controle (M82); (c) amostra IL 9-2-6.

A análise estatística dos dados revelou a existência de 12 QTL que influenciam os teores dos diferentes isômeros de tocoferol em todos os cromossomos analisados (Fig. 6). Considerando os marcadores que flanqueiam os fragmentos introgredidos de cada linhagem, os QTL foram mapeados e a sua localização no mapa genético está indicada na Figura 9.

Para o cromossomo 9, foram identificados QTL com efeito positivo, dois para α - e tocoferol total (ILs 9-1 e 9-2-6), um para β -tocoferol (IL 9-1), além de um QTL de efeito negativo para γ -tocoferol (IL 9-1).

Os níveis aferidos em frutos das ILs com fragmentos introgredidos de *S. pennellii* no cromossomo 6 mostraram diferenças significativas positivas comparados aos frutos de M82 nos conteúdos de β -tocoferol (IL 6-2), δ - e tocoferol total (IL 6-1).

No cromossomo 8, foi identificado um QTL positivo para tocoferol total (IL 8-2-1). 60 QTLs que influenciam negativamente o conteúdo de α - e β -tocoferol foram encontrados associados ao cromossomo 7, nas ILs 7-4 e 7-4-1, respectivamente).

Genótipo	Réplicas	Alfa (ug/gTF)	Beta (ug/gTF)	Gama (µg/gTF)	Delta (ug/gTF)	Total (ug/gTF)
	1	3 548	0.089	0.952	0 172	4 761
M82	2	2,773	0.077	0.353	0.192	3.395
M82	- 3	3.263	0.074	0.563	nd	3,900
M82	4	1.733	0.064	0.519	0.318	2.635
M82	5	1.902	0.057	0.290	0.019	2.268
M82	6	2.916	0.100	2.062	0.271	5,349
M82	7	2,417	0.091	0,188	nd	2,697
M82	8	3,004	0,112	0,110	nd	3,225
M82	9	3,984	0,131	0,588	0,069	4,772
M82	10	3,505	0,079	1,100	nd	4,684
M82	11	1,772	nd	nd	0,133	1,906
M82	12	3,601	0,063	1,610	nd	5,274
M82	13	2,698	0,059	0,127	nd	2,884
M82	14	4,144	0,098	1,180	0,094	5,516
M82	15	3,997	0,108	0,526	0,029	4,660
M82	16	3,405	0,092	1,546	nd	5,042
M82	17	4,323	0,137	1,386	nd	5,845
M82	18	2,164	0,061	0,180	nd	2,405
6-1	1	3,970	nd	1,334	0,132	5,436
6-1	2	2,603	0,081	1,012	0,621	4,317
6-1	3	3,661	0,102	1,183	0,439	5,385
6-1	4	2,786	0,061	0,984	0,604	4,435
6-1	5	6,489	0,143	1,240	0,124	7,996
6-1	6	5,098	0,101	0,962	0,104	6,266
6-2	1	3,263	0,105	0,108	0,067	3,543
6-2	2	3,798	0,108	1,614	0,991	6,511
6-2	3	4,802	0,147	0,181	0,111	5,241
6-2	4	2,561	0,164	0,091	0,036	2,852
6-2	5	4,013	0,083	2,162	0,255	6,514
6-2	6	3,069	0,129	0,112	0,015	3,325
7-4	1	1,096	0,049	0,095	nd	1,241
7-4	2	1,286	0,065	0,218	0,022	1,591
7-4	3	3,446	0,072	0,326	nd	3,845
7-4	4	2,659	0,108	0,624	nd	3,391
7-4	5	0,770	0,066	0,131	nd	0,967
7-4	6	2,821	0,067	0,972	nd	3,861

Tabela 4: Valores obtidos para os isômeros e tocoferol total expresso em µg de composto por g de tecido fresco (TF). nd, composto não-detectado na amostra.

7-4-1	1	2,947	0,055	0,423	0,025	3,451
7-4-1	2	2,947	0,045	0,817	0,501	4,310
7-4-1	3	1,650	0,033	0,518	0,173	2,374
7-4-1	4	3,343	0,057	1,666	nd	5,066
7-4-1	5	5,232	nd	1,566	nd	6,798
7-4-1	6	2,312	nd	1,003	nd	3,314
7-5	1	1,726	0,065	0,206	nd	1,996
7-5	2	1,532	0,080	0,315	0,050	1,977
7-5	3	3,974	nd	2,261	nd	6,235
7-5	4	3,195	0,201	0,466	nd	3,862
7-5	5	2,217	0,098	0,713	0,115	3,143
7-5	6	2,739	0,102	0,289	0,081	3,211
8-2	1	1,983	0,087	2,056	0,384	4,510
8-2	2	1,136	0,145	0,202	nd	1,482
8-2	3	1,473	0,063	1,881	0,356	3,773
8-2	4	4,826	0,088	2,711	0,263	7,889
8-2	5	1,756	0,084	1,050	nd	2,889
8-2	6	2,069	0,119	0,576	nd	2,764
9-1	1	5,696	0,125	0,320	0,196	6,337
9-1	2	5,356	0,190	nd	nd	5,545
9-1	3	7,162	0,246	0,214	0,132	7,754
9-1	4	5,707	0,206	0,137	0,050	6,099
9-1	5	5,012	0,102	0,265	0,044	5,423
9-1	6	4,895	nd	0,193	nd	5,088
8-2-1	1	3,079	0,077	0,106	nd	3,262
8-2-1	2	5,148	0,116	2,587	0,183	8,034
8-2-1	3	3,389	0,112	0,500	0,059	4,059
8-2-1	4	3,157	0,074	1,763	0,171	5,165
8-2-1	5	4,211	0,091	2,289	nd	6,591
8-2-1	6	4,211	0,106	1,204	nd	5,521
9-2-6	1	3,770	0,076	0,707	nd	4,553
9-2-6	2	6,034	0,091	1,673	nd	7,798
9-2-6	3	9,393	0,144	1,035	0,049	10,622
9-2-6	4	4,037	0,067	0,617	nd	4,722
9-2-6	5	3,806	0,073	1,264	0,039	5,183
9-2-6	6	6,617	0,101	0,696	nd	7,414







63

2.2. Identificação de genes candidatos

As regiões genômicas onde os QTL foram mapeados servem, portanto, como referência na busca de genes candidatos, isto é, genes que co-segregam com a região genômica identificada e que, pela função dos seus produtos, podem contribuir ou determinar as variações observadas (Tabor *et al.*, 2002). Obviamente, genes envolvidos estruturalmente na via de biossíntese da VTE poderiam explicar as diferenças fenotípicas detectadas entre as ILs e o controle M82. Dessa forma, os resultados da distribuição cromossômica dos genes da rota de biossíntese VTE foram associados ao mapeamento dos QTL para identificação de genes candidatos que determinam o conteúdo de VTE em fruto de tomate.

Os QTL identificados na região genômica do fragmento introgredido da IL 9-2-6 co-localizam com dois genes que codificam enzimas da rota central da VTE, *vte3*(1) e *vte5* (Figura 9), enquanto os QTL da IL 9-1 estão associados à região genômica que contém os genes *tyra*(2) e *ggps*(4).

Da mesma maneira, os demais genes codificantes enzimas da rota central da VTE também co-localizam com QTL: *vte2* e *hppd*(1), no cromossomo 7; enquanto no cromossomo 8, têm-se a presença dos genes *vte1* e *vte4* (Figura 9). Além dos genes mencionados, a presença dos *loci tyra*(1) em 0.4 cM e *tat*(2) em 43 cM no cromossomo 7 poderiam também explicar os QTL para α - e β -tocoferol mapeados nessas regiões.

Buscas ao longo das regiões genômicas que contem os QTL identificados revelaram 6 genes candidatos que pertencem a vias metabólicas relacionadas à biossíntese de VTE associados aos QTL identificados. No cromossomo 6, a. 32 cM, foi identificado um gene que codifica para a enzima clorofilase (CHL, EC 3.1.1.14). Esta enzima catalisa o primeiro passo da degradação da clorofila durante a senescência e amadurecimento do fruto (Hörtensteiner, 2006), removendo a cadeia lateral fitol do anel pirrólico da clorofila. Como mencionado anteriormente, o fitol é utilizado como substrato da enzima VTE5 para produção de fitil 2P, precursor da rota de VTE. Além desse, outros três genes candidatos codificam enzimas das vias de síntese do folato e triptofano, as quais utilizam o intermediário corismato como precursor; fosforibosilantranilato isomerase (PRAI, EC 5.3.1.24), folilpoliglutamato sintase

(FPGS, EC 6.3.2.17) e antranilato fosforibosiltransferase (APT, EC 2.4.2.18), cujas posições no mapa genético são 24 cM, 26 cM e 59 cM, respectivamente. Finalmente, o gene codificante para licopeno β -ciclase (LYCB) localizado a 74 cM no cromossomo 6 também constitui um forte candidato para explicar as mudanças nos níveis de tocoferol da IL 6-2. Esta enzima, que foi extensivamente caracterizada após a clonagem posicional por Ronen *et al.* (2000) em tomate, catalisa o último passo da biossíntese de β -caroteno, via que compartilha o precursor geranilgeranil 2P com a rota da VTE. Colocalizado aos QTL identificados para a IL 9-2-6, foi identificado o gene que codifica para a proteína SEC14. Alguns estudos têm demonstrado o envolvimento dessa proteína no transporte de tocoferol nas células de mamíferos e no transporte de lipídeos nas plantas (Saito *et al.*, 2007; Bankaitis *et al.*, 2009).

Em conjunto, os resultados obtidos a partir da identificação e mapeamento dos genes da rota biossintética de VTE, quantificação dos níveis de tocoferóis nas ILs, e buscas por genes candidatos, permitiram propor a existência de 16 *loci* que afetam a produção e/ou acúmulo de tocoferol em tomate: *prai, fpgs, chl, apt* e *lycb* no cromossomo 6; *tyra*(1), *vte2, hppd*(1) e *tat*(2) localizados no cromossomo 7; *vte1* e *vte4* no cromossomo 8; e *ggps*(4), *tyra*(2), *vte5*, *sec14* e *vte3*(1) para o cromossomo 9 (Figuras 5, 8 e 9).



genético Tomato EXPEN 2000 disponível em Solanaceae Genomics Network (http://solgenomics.net/index.pl). Marcadores e genes estão indicados à esquerda dos cromossomos. Em vermelho, genes pertencentes à via do MEP; em verde, genes envolvidos na rota do SK; em azul, genes da rota central da VTE; sem destaque, genes que pertencem às vias associadas ao metabolismo da VTE. QTL para tocoferol estão indicados do lado direito Figura 9: Localização genômica dos genes envolvidos na biossíntese de VTE e candidatos. Todos os genes foram posicionados no mapa dos cromossomos. O padrão de cores dos genes está de acordo com a Figura 5.

3. Clonagem e caracterização dos alelos de *S. pennellii* para os genes candidatos identificados

Com o intuito de acrescentar evidências sobre o envolvimento dos genes identificados na determinação do conteúdo de tocoferol em tomate, assim como suas causas, foram analisadas as diferenças alélicas entre *S. lycopersicum* e *S. pennellii* dos16 genes candidatos associados aos QTL (genes sublinhados na Figura 5). Para tanto, as sequências codificantes dos alelos selvagens foram amplificadas a partir do cDNA das correspondentes ILs utilizando iniciadores ancorados nos códons de início e parada (Figura 10). Posteriormente, os fragmentos foram clonados e sequenciados em ambos os sentidos.



Figura 10: Amplificação dos alelos selvagens dos genes candidatos. Gel de agarose apresentando as bandas correspondentes às regiões codificantes dos alelos de *S. pennellii* amplificadas a partir de cDNA das ILs [8-2 (vte1; vte4), 7-4-1 (tyrA1; vte2; hppd1), 9-2 (vte5; sec14; vte3.1); 6-2 (chl; apt), pen (prai; tat2; tyrA2). O padrão de bandas do marcador de peso molecular 1 Kb Ladder (Invitrogen, Cat. No. 10787) está à esquerda de cada gel e o tamanho das bandas indicado em pares de bases (pb). O tamanho dos fragmentos são apresentados na tabela 2.

O par de alelos que codifica para a enzima PRAI mostrou-se o mais divergente uma vez que a proteína deduzida de *S. pennellii* possui 3 substituições e uma deleção de 26 aminoácidos em relação à proteína de *S. lycopersicum* (Figura 11).

prai pen	MLSVFNMGAQLQPRVIGLPRSQGVLDVEVYSTNRAQLRTSTSNRFMCTLNQSTGVSAQAAQTESRTLVKM
prai lyc	MLSVFNMGAOLOPRVIGLPRSOGVLDVEVYSINRAOLRTSTSNRFMCTLNOSTGVSAOAAOTESRTLVKM
Consensus	MLSVFNMGAOLOPRVIGLPRSOGVLDVEVYSINRAOLRTSTSNRFMCTLNOSTGVSAOAAOTESRTLVKM

prai pen	CGITSARDAALAAEAGANFIGMIIWPNSKRSVSLSTAKEISKVAREYGALPVGVFVDDNADTILRASDAA
prai lyc	$\tt CGITSARDAALAAEAGANFIGMIIWPNSKRSVSLSTAKEISKVAREYGALPVGVFVDDNANTILRASDAA$
Consensus	CGITSARDAALAAEAGANFIGMIIWPNSKRSVSLSTAKEISKVAREYGALPVGVFVDDNAdTILRASDAA

prai_pen	NLELVQ LHGNDSRDAFPVLVRERPLVYVLHADEEGGLLNSISNE
prai lyc	NLELVQVKTVIFHHLVSIVFNTQNPQHVQIEQLHGNDSRDAFPVLVRERPLVYVLHADEEGGLLNSISNE
Consensus	NLELVQLHGNDSRDAFPVLVRERPLVYVLHADEEGGLLNSISNE

prai_pen	ESSLVDWILVDSAKGGSGKGFNWAQFKLPSIRSKHGWLLAGGINPENVCEALSALKPNGVDVSSGICGQD
prai_lyc	ESSLVDWILVDSAKGGSGKGFNWAQFKLPSIRSKQGWLLAGGINPENVCEALSALKPNGVDVSSGICGQD
Consensus	${\tt ESSLVDWILVDSAKGGSGKGFNWAQFKLPSIRSK} {\tt h} {\tt GWLLAGGINPENVCEALSALKPNGVDVSSGICGQD}$

prai pen	GIKKDESRIQSFMNAVKSLHL
prai lyc	GIKKDESRIQSFMNAVKSLHL
Consensus	GIKKDESRIQSFMNAVKSLHL

(b)	
hppdl_pen	MGKQAAAAADDQQLPGDIEDKFKLVGFNNFVRTNPRSDFFTVKRFHHIEFWCGDATNTARRFSWGLGLPI
hppdl_lyc	MGKQAAAAADDQQLPGDIEDKFKLVGFNHFVRTNPRSDFFTVKRFHHIEFWCGDATNTARRFSWGLGLPI
Consensus	MGKQAAAAADDQQDPGDIEDKFKLVGFNnFVRTNPRSDFFTVKRFHHIEFWCGDATNTARRFSWGLGDPI ************************************
hppd1 pen	AAISDLSTGNSVHASYLLRSSSSSQLQFLFTAPYSPAISTPSSSSSIPTFSVSSHRSFTATHGLGVRAIAL
hppd1 lyc	AAISDLSTGNSVHASYLLRSSSSSQLQFLFTAPYSSAISTSSSASIPTFSVSSHRSFTATHGLGVRAIAL
Consensus	${\tt AAISDLSTGNSVHASYLLRSSSSSQLQFLFTAPYSpAISTpSSaSIPTFSVSSHRSFTATHGLGVRAIAL}$

hppd1 pen	EVENSRLAFSTCVSHGAKPVSEPVILNDEVVIAEVHLYGDVVLRFVSFLKDSNCFVFLPGFESVEGAHLD
hppd1_lyc	${\tt EVENSRLAFSTCVAHGAKPVSEPVILNDEVVIAEVHLYGDVVLRFVSFLKDSNCFVFLPGFESVEGAQLD}$
Consensus	${\tt EVENSRLAFSTCVaHGAKPVSEPVILNDEVVIAEVHLYGDVVLRFVSFLkDSNCFVFLPGFESVEGahLD}$
	*********** ***************************
hppd1_pen	YGIRRLDHAVGNVPELGPVVEYIKSFTGFHEFAEFTAEDVGTAESGLNSVVLANNDETVLFPLNEPVYGT
hppd1_lyc	${\tt YGIRRLDHAVGNVPELGPVVEYIKSFTGFHEFAEFTAEDVGTAESGLNSVVLANNDETVLFPLNEPVYGT}$
Consensus	${\tt YGIRRLDHAVGNVPELGPVVEYIKSFTGFHEFAEFTAEDVGTAESGLNSVVLANNDETVLFPLNEPVYGT}$

hppd1_pen	KRKSQIQTYLEHNEGAGVQHLALVTEDIFRTLREMRKRSGIGGFEFMPSPPPTYYKNLKTRAGDILSDEQ
hppd1_lyc	${\tt KRKSQIQTYLEHNEGAGVQHLalvTedIFRTLREMRKRSGIGGFEFMPSPPPTYYKNLKTRAGDILSDEQ}$
Consensus	KRKSQIQTYLEHNEGAGVQHLALVTEDIFRTLREMRKRSGIGGFEFMPSPPPTYYKNLKTRAGDILSDEQ
hppd1_pen	IQECEELGILVDRDDQGTLLQIFTKPVGDRPTIFLEIIQRIGCMLKNEEGELYQKGGCGGFGKGNFSELF
hppdl_lyc	IQECEELGILVDRDDQGTLLQIFTKPVGDRPTIFLEIIQRIGCMLKNEEGELYQKGGCGGFGKGNFSELF
Consensus	IQECEELGILVDRDDQGTLLQIFTKPVGDRPTIFLEIIQRIGCMLKNEEGELYQKGGCGGFGKGNFSELF ************************************
hondl non	K
hppd1_pen	*
Consensus	
23110011040	

Figura 11. Comparação da sequencias protéicas dos alelos de *S. pennellii* e *S. lycopersicum*. Alinhamento da sequência de aminoácidos das enzimas PRAI (a) e HPPD1 (b). Os sítios polimórficos estão indicados em verde.

(a)

A análise das sequências nucleotídicas revelou polimorfismos entre todos os pares de alelos (Anexo I). Diferenças nos tamanhos esperados dos amplicons foram observadas apenas entre os dois alelos dos genes *tyra(2), vte4, prai* e *chl*. Estes resultados indicam a ausência de regiões protéicas altamente variáveis e de grandes rearranjos nas regiões genômicas que englobam os genes analisados. As sequências foram traduzidas *in silico* e as proteínas de ambas as espécies comparadas. Todos os genes candidatos apresentam pelo menos um sítio polimórfico não-sinônimo (Tabela 5).

		Sequ codifica	ência ante (nt)	Proteína predita (aa)		nº nucleotídeos	nº de aa
GENE	UNIGENE	LYC	PEN	LYC	PEN	polimórficos ¹	polimórficos ¹
ggps(4)	U575882	1005	1005	334	334	11	3
tat(2)	U563404	1269	1269	422	422	10	6
tyra(1)	U567861	1134	1134	377	377	7	1
tyra(2)	U570951	1173	1182	390	393	24+9 inserções	9+3 inserções
vte4	U584511	1089	1086	362	361	9+3 deleções	2+1 deleções
vte1	U570602	1497	1497	498	498	12	5
vte3(1)	U578249	1020	1020	339	339	7	1
vte2	U327540 ² U576207 ³	1209	1209	402	402	8	3
vte5	U583081	882	882	293	293	10	4
hppd(1)	U580457	1263	1263 ⁴	420	421 ⁴	16	7
apt	U566340 ⁵	1097 ⁵	1097 ⁵	365	365	8	4
prai	U564371	906	828	301	275	12+78 deleções	3+26 deleções
fpgs	U581922 ⁵	1457 ⁵	1457 ⁵	485	485	13	4
chl	U574853	939	948	312	315	22+9 inserções	9+3 inserções
sec14	U583419	1275	1275	424	424	11	4
lycb	U570109	1497	1497	498	498	19	9

Tabela 1: Comparação dos alelos de S. lycopersicum e S. pennellii dos genes candidatos.

¹ Inserção/deleção nas sequências de S. pennellii.

² extremo 5'

³extremo 3'

⁴ A sequência analisada não apresentou polimorfismo no códon de parada.

⁵ A sequência analisada não inclui o extremo 3'.

LYC: S. lycopersicum.

PEN: S. pennellii.

V. Discussão

A disponibilidade atual de bancos de dados de sequências genômicas e expressas têm permitido a reconstrução de diversas vias metabólicas para diferentes espécies. Estes resultados fornecem um importante passo no estudo de aspectos específicos da síntese, acúmulo e transporte dos diferentes compostos dentro da célula. Desta forma, no caso do tomate, a publicação de versões preliminares do genoma de *S. lycopersicum* constitui uma plataforma fundamental para identificação de genes e de sua função proteica, ainda que imprecisões na montagem resultem em informações perdidas.

Neste trabalho, a fim de identificar passos regulatórios determinantes do conteúdo de VTE em fruto tomate, foi reconstruída a rota biossintética, desde as vias do MEP e do SK, que fornecem os precursores para a produção de tococromanóis, até as enzimas da sua rota central. Este levantamento resultou na identificação de 29 reações catalisadas por 28 enzimas. Posteriormente, buscou-se os genes que codificam estas enzimas no genoma de *S. lycopersicum*, identificando-se 41 *loci*, de acordo com os critérios metodológicos elegidos.

Durante o processo de reconstrução das vias bioquímicas, foram consideradas algumas premissas para o delineamento das reações envolvidas no metabolismo da VTE em plantas. A primeira delas está relacionada à rota do MEP e envolve a síntese dos precursores para os isoprenóides. Tipicamente, a enzima GPPS atua na biossíntese de monoterpenos, enquanto a GGPS é responsável pela produção da geranilgeranil 2P por meio da incorporação sequencial de três moléculas de isopentenil 2P (IPP) ao dimetilalil (DMAPP). O geranilgeranil 2P serve como um precursor para síntese de carotenóides, tococromanóis, giberelinas e clorofila (Aharoni *et al.*, 2005; Joyard *et al.*, 2009). Sendo assim, tradicionalmente assume-se que a GPPS não estaria envolvida na produção de tococromanóis. Dados recentes mostram que embora plantas de tomate silenciadas para GPPS exibissem fenótipo anão e níveis reduzidos de giberelina, estas não apresentavam alteração nos conteúdos de carotenóides e clorofila, indicando que a origem desses

70

pigmentos dá-se a partir de um *pool* de geranilgeranil 2P-independente do intermediário geranil 2P (Van Schie *et al.*, 2007). Estes resultados sugerem que GPPS pode não influenciar na síntese de VTE, embora experimentos adicionais sejam necessários para validação dessa hipótese.

Outro ponto importante diz respeito à reação de condensação de homogentisato e geranilgeranil 2P que leva a síntese de tocotrienóis. A enzima HGGT que catalisa exclusivamente essa reação foi previamente caracterizada apenas em espécies de gramíneas, cuja forma mais abundante de tococromanol é o tocotrienol (Cahoon et al., 2003). Até o momento, nenhum gene que codifica HGGT foi identificado em arabidopsis ou em tabaco e tomate, embora traços de tocotrienóis já tenham sido reportados nessas espécies de solanáceas (Chun et al., 2006). A presença desses compostos, entretanto, poderia ser explicada pela atividade promíscua de outras preniltransferases, como HST e/ou VTE2, na aceitação do substrato. Mudanças no fluxo através da SK, MEP e/ou rota central do tocoferol poderiam provocar alteração na preferência do substrato dessas enzimas (Herbers, 2003; Falk & Munné-Bosch, 2010). Esta hipótese é suportada pela observação de que plantas transgênicas de tabaco co-expressando o gene hppd de arabidopsis e prefenato desidrogenase (pdt) de levedura exibem altos níveis de tocotrienóis (Rippert et al., 2004). Além disso, o aumento no suprimento de homogentisato poderia afetar a especificidade pelo substrato da prenil transferase provocando a síntese de tocotrienol (Falk & Munné-Bosch, 2010). Adicionalmente, foi demonstrado que a HST de A. thaliana, enzima envolvida na produção de plastoquinona-9 que catalisa a condensação do solanesil 2P e homogentisato, também aceita farnesil 2P e geranilgeranil 2P como doadores de grupo prenil (Sadre et al., 2006; Tian et al., 2007), fornecendo mais uma evidência que suporta a produção de tocotrienol na ausência de uma HGGT específica.

Finalmente, apenas um ponto permanece ainda não resolvido dentro da via metabólica estudada, a ausência de uma fitol-P quinase que atuaria no passo subsequente à VTE5, produzindo fitil 2P a partir da degradação da clorofila (Ischebeck *et al.*, 2006; Valentin *et al.*, 2006).

No geral, todos os genes identificados que codificam enzimas da via MEP estão de acordo com as sequências previamente reportadas para tomate (Lois

et al., 2000; Paetzold *et al.*, 2009; Rodríguez-Concepción *et al.*, 2001; Rohdish *et al.*, 2000; Rodríguez-Concepción *et al.*, 2003; Botella-Pavía *et al.*, 2004; Ament *et al.*, 2006; GenBank *database direct submission* para IPI - GQ169536 e EU253957), com exceção dos genes codificantes para CMS e GGDR, os quais não haviam sido descritos anteriormente nesta espécie. Além disso, dois *loci* adicionais para GGPS foram identificados (GGPS3 e GGPS4). No caso da via do SK, os genes codificantes identificados para as enzimas que catalizam etapas anteriores do intermediário prefenato também já haviam sido descritos em *S. lycopersicum* (Görlach *et al.*, 1993; Bischoff *et al.*, 1996; Bischoff *et al.*, 2001; Schmid *et al.*, 1992; Görlach *et al.*, 1995; Eberhard *et al.*, 1996; Gasser *et al.*, 1988). Em particular, este trabalho identificou novos *loci* para SDH/DHQ e CM, além de descrever pela primeira vez os genes codificantes para PAT, TyrA e TAT em tomate.

Dados acerca da localização subcelular de proteínas constituem fonte de informação fundamental no processo de elucidação da função gênica. Embora seja aceito que as vias bioquímicas de MEP e SK em plantas ocorram no cloroplasto (Rippert et al., 2009; Friso et al., 2010), a existência de uma porção pós-corismato citossólica da via do chiquimato é objeto de controvérsia. A questão surgiu após a descrição de uma isoforma citossólica da CM em arabidopsis (Eberhard et al., 1996), sugerindo que as enzimas subsequentes da via (PAT, TyrA, TAT, HPPD) poderiam estar localizadas no citossol das células vegetais. Entretanto, a recém-identificada PAT apresenta peptídeo sinal predito direcionado ao cloroplasto em arabidopsis (Graindorge et al., 2010), da mesma maneira que as TyrAs descritas nessa espécie, cuja predição foi confirmada pela expressão das proteínas acopladas ao gene repórter GFP (Rippert et al., 2009). Por outro lado, análises de extratos protéicos de cloroplastos purificados por imunodetecção revelaram que a HPPD está confinada ao compartimento citossólico em células de arabidopsis em cultura (Garcia et al., 1999); enquanto que, para a TAT, a predição in silico não detecta peptídeo sinal. Essas evidências da presença de isoformas citossólicas geram questões acerca de uma via SK operante extra-plastidial (Ding et al., 2007), fato intrigante, principalmente em relação à síntese e ao transporte do precursor homogentisato ao longo da membrana do cloroplasto. Dessa forma,

72
a tirosina produzida no cloroplasto deveria ser transferida ao citossol, compartimento das enzimas TAT e HPPD, convertido no precursor homogentisato que, por sua vez, seria transportado ao cloroplasto para suprir a síntese de tococromanóis.

No caso das predições in silico dos loci de tomate, de modo geral, as sequências de aminoácidos deduzidas analisadas apresentaram sinalização de direcionamento para o cloroplasto. Para as enzimas TAT, CM e HPPDs, o direcionamento ao cloroplasto não pode ser confirmado pelas predições realizadas em concordância com as observações já reportadas em arabidopsis. Em contraste, para a enzima bifuncional SDH/DHQ, nenhum peptídeo sinal foi detectado em ambos os unigenes levantados. Isso se deve provavelmente a uma falha na identificação da sequência sinal dessas proteínas uma vez que análises de fracionamento subcelular sugerem que esta enzima atua no cloroplasto em tomate (Bischoff et al., 2001). Ding et al. (2007), entretanto, caracterizou funcionalmente uma isoforma citossólica da SDH/DHQ em tabaco além de identificar ortólogo correspondente 0 gene em tomate (BF096277/SGN-U578253). Este locus não foi incluído no presente estudo devido a sua identidade em relação à proteína de Arabidopsis ser inferior aos critérios adotados e à ausência de dados funcionais reportados até o momento.

O insucesso na predição da localização subcelular não é um cenário raro, uma vez que existe uma margem de erro intrínseca à identificação *in silico* do peptídeo sinal. A análise da sequência do gene *gpps* ilustra esse caso. Quando a sequência de *A. thaliana* de GPPS é submetida à análise da ferramenta TargetP, o sinal predito aponta para localização mitocondrial (dados não mostrados); em contrapartida, evidências experimentais revelam que essa enzima é direcionada ao plastídeo (Bouvier *et al.,* 2000). Existem descritas na literatura evidências ambíguas sobre a real localização da GPPS as quais sustentam tanto um direcionamento para o plastídeo quanto para o citossol dependendo da espécie estudada (Nagegowda, 2010).

Assim, os dados obtidos pela predição *in silico* constituem ponto de partida no estudo da localização celular de proteínas, mas a validação experimental é necessária para a definição do direcionamento celular das proteínas em estudo.

Seguindo na caracterização dos genes de tomate codificantes para as enzimas envolvidas na biossíntese de VTE, os 41 loci foram subsequentemente mapeados. A partir desses resultados, um perfil detalhado dos guatro isômeros de tocoferol em fruto foi realizado nas ILs que possuem as regiões que incluem os genes na rota central da VTE (IL 7-4, 7-4-1, 8-2, 8-2-1, 9-1 e 9-2-6) juntamente com aquelas do cromossomo 6 (IL 6-1 e 6-2) para as quais QTL para α-tocoferol foi previamente descrito (Schauer et al., 2006). Embora o conteúdo de VTE seja um caráter altamente afetado pelo ambiente (Schauer et al., 2008), nossos resultados indicam que, ao menos aqueles QTL mapeados no cromossomo 6 e 9 apresentam um nível relativamente alto de herdabilidade, pois estão de acordo com os experimentos anteriores reportados por (Schauer et al., 2006). Além disso, vale ressaltar que em oposição ao estudo desses autores, que foi conduzido com plantas crescidas em campo aberto, nossos dados foram obtidos a partir de material vegetal crescido em casa de vegetação.

A análise integrada dos dados metabólicos, genômicos e genéticos nos permite propor dezesseis *loci* candidatos que poderiam afetar o conteúdo de tocoferol em tomate: *prai, fpgs, chl, apt* e *lycb* no cromossomo 6; *tyra*(1), *vte2, hppd*(1) e *tat*(2) localizados nos cromossomo 7; *vte1* e *vte4* no cromossomo 8; e *ggps*(4), *tyra*(2), *vte5, sec14* e *vte3*(1) pertencentes ao cromossomo 9. Em plantas, diversos QTL que controlam o conteúdo de tocoferol forma identificados em soja, milho, colza (*Brassica napus*) e *Arabidopsis.* Assim como neste estudo, alguns desses QTL estão localizados em regiões do genoma que abrigam os genes da rota central do tocoferol (Li *et al.*, 2010; Chander *et al.*, 2008; Marwede *et al.*, 2005; Gilliland *et al.*, 2006).

Uma análise detalhada dos genes que co-localizam com os QTL identificados faz surgir hipóteses acerca da regulação do conteúdo de VTE em fruto. Nenhum dos QTL localizados no cromossomo 6 co-localiza com qualquer um dos genes das vias de MEP, SK, ou rota central da VTE, sugerindo que a variação no conteúdo observada nas IL 6-1 e 6-2 pode ser determinada, por exemplo, pelo efeito de genes pertencente a vias associadas à VTE. A IL 6-1 apresenta níveis elevados de δ - e tocoferol total em comparação com o controle *S. lycopersicum* (M82), enquanto a IL 6-2 apresenta aumento 74

significante dos níveis de β-tocoferol. A presença dos alelos de S.pennellii dos genes prai e fpgs na IL 6-1 poderia influenciar o influxo de hidroxifenilpiruvato para a rota de síntese de tocoferol por meio de um desvio do intermediário corismato das rotas de biossíntese do triptofano e folato, respectivamente. Além disso, a IL 6-1 compartilha com a IL 6-2 o alelo selvagem do gene chl. Como apresentado anteriormente, a hipótese que o fitol derivado da degradação da clorofila serve como um importante intermediário para a síntese de tocoferol foi demonstrada pela caracterização do mutante de arabidopsis vte5 (Valentin et al., 2006). Nesse sentido, a presença do alelo de chl de S. pennelllii nas linhagens 6-1 e 6-2 poderia resultar no aumento da disponibilidade do precursor fitil-2P para a síntese de tocoferol. No mesmo cromossomo, IL 6-2 também apresenta significante aumento dos níveis de βtocoferol em relação ao controle. Esta IL carrega também os alelos de apt e lycb. Ainda que esses genes estejam ligados a intermediários da via do tocoferol (corismato e geranilgeranil 2P), não há ligação clara que possa ser especificamente associada com a isoforma β-tocoferol.

A candidatura de genes que não estão diretamente envolvidos na rota estrutural da VTE ganha suporte pela complexa rede regulatória que age nos pontos de ramificação dessas vias. A enzima CM que atua na via do SK é inibida alostericamente em plantas por fenilalanina e tirosina e induzida por triptofano (Tzin & Galili, 2010). A variação alélica existente em *prai* e *apt* poderia modificar a síntese de triptofano, alterando os níveis de seus intermediários e que resultaria no aumento de precursor homogentisato para a síntese de tocoferol.

Para o cromossomo 7, dois QTL de efeito negativo foram detectados; enquanto da IL 7-4 apresenta baixos níveis de α -tocoferol, a IL 7-4-1 mostrou quantidades reduzidas da isoforma β . O fragmento genômico selvagem introgredido nestas linhagens possui os alelos de *tyra*(1), *vte2* e *hppd*(1). A IL 7-4 também contém o alelo de *S. pennellii* para o gene *tat*(2). Estes quatro candidatos, genes estruturalmente envolvidos na via, poderiam alterar a biossíntese de tocoferóis. Todavia, o modo pelo qual esses genes poderiam diferencialmente modificar os níveis das diferentes isoformas não está claro.

No cromossomo 8, a IL 8-2 apresentou níveis significativamente baixos da razão α/γ -tocoferol quando comparada ao controle. Interessantemente, esta IL carrega o alelo de *S. pennellii* dos genes *vte1* e *vte4* cujos produtos protéicos atuam na síntese de ambas isoformas de tocoferol. A baixa razão α/γ -tocoferol poderia ser causada pela diminuição e/ou aumento dos alelos selvagens da VTE1 e VTE4, respectivamente. Curiosamente, a IL 8-2-1 também possui os alelos de *S. pennellii* de *vte1* e *vte4*. Embora nenhuma alteração significativa tenha sido detectada na razão entre α/γ -tocoferol, um significante aumento no conteúdo total de tocoferol foi observado nos frutos dessa linhagem.

Finalmente, os níveis elevados observados para α -, β - e tocoferol total na IL 9-1 podem ser ocasionados pela presença dos alelos de S. pennelllii para os loci tyra(2) e ggps(4) que poderiam levar a um aumento na disponibilidade de precursores hidroxifenilpiruvato e fitil 2P para a via central do tocoferol. A elevada razão observada para $\alpha/\gamma \in \beta/\gamma$ suporta esta hipótese. É interessante ressaltar que tyra(2) apresenta maior número de polimorfismo entre S. pennellii e S. lycopersicum para as regiões codificantes analisadas, cujas alterações poderiam impactar a atividade da proteína. Nesse sentido, esses resultados reiteram a hipótese de Rippert et al. (2004), que o percursor hidroxifenilpiruvato é um intermediário chave no acúmulo de VTE in plantas. A IL 9-2-6 exibe aumento nos conteúdos de α- e tocoferol total. Essa linhagem possui os alelos selvagens dos genes vte5 e vte3. Enquanto o primeiro poderia aumentar a disponibilidade de fitil 2P via degradação do fitol (Valentin et al., 2006), o segundo poderia produzir uma proteína com maior atividade, por exemplo, resultando no acúmulo dos produtos finais da via, uma vez que um acúmulo de β-tocoferol não é observado juntamente com diferenças significativas na razão α/β.

Vale ressaltar que embora exista uma real ligação entre genes candidatos e o conteúdo de tocoferol em frutos, o efeito de *loci* não detectados dentro dos QTL não pode ser descartado, uma vez que as linhagens analisadas são constituídas por fragmentos introgredidos de tamanho considerável, cujas diferenças fenotípicas observadas no acúmulo das isoformas de tocoferol podem ser ocasionadas por outros mecanismos não acessados nesse estudo.

A clonagem e sequenciamento dos genes candidatos identificados de S. pennellii permitiu revelar um considerável grau de polimorfismo entre os alelos cultivados e selvagens. Foi possível confirmar a presença polimorfismos em todos os genes, que perfazem um total de 298 sítios polimórficos ao longo das sequência nucleotídicas analisadas para os 16 genes, 18712 bases de S. lycopersicum contra 18649 bases de S. pennellii. A maioria das alterações detectadas corresponde a polimorfismos de um único nucleotídeo, sendo que 107 polimorfismos resultam em mudanças na sequência protéica. Para quatro genes, tyra(2), vte4, prai e chl, foram encontrados inserções/deleções que alteram o tamanho da proteína, mas não o quadro de leitura. prai e tyra(2) destacam-se na análise comparativa por apresentarem as sequências que mais acumularam variação entre as duas espécies. Essas diferenças observadas sugerem o envolvimento das variantes alélicas na diversidade fenotípica, reforçando assim a candidatura dos genes. No entanto, a causalidade das alterações fenotípicas pode residir em regiões não analisadas neste trabalho, como por exemplo, sequências intrônicas, promotoras e ainda a presença ou ausência de outros elementos reguladores.

Este trabalho realizou um levantamento sistemático e detalhado dos genes que codificam as enzimas da rota biossintética da VTE em tomate, partindo da informação da rota de biossíntese da VTE previamente caracterizada na espécie modelo arabidopsis. O método de quantificação adotado permitiu descrever as diferenças de conteúdo dos diferentes isômeros de tocoferol em frutos das ILs de S. pennellii, corroborando os QTL previamente descritos, além de identificar outros novos. A associação entre a variação fenotípica observada e as sequências genômicas envolvidas na biossíntese de VTE identificadas em tomate S. lycopersicum, nos levou à proposição de 16 loci que apresentam variabilidade alélica entre a espécie cultivada e a selvagem e que, portanto, poderiam explicar a mudança no teor desses compostos em frutos de tomate. Os genes candidatos identificados suportam a interação entre as vias do MEP, SK, e rota central do tocoferol para controle do acúmulo de VTE na espécie. Além disso, genes de vias associadas à VTE podem contribuir na regulação do suprimento de intermediários da síntese de tocoferóis. Em conjunto, esses resultados representam um

importante passo para o entendimento dos componentes genéticos envolvidos na determinação e controle da variação natural do conteúdo de VTE em frutos de tomate, uma vez que fornece uma vasta plataforma de dados para estudos funcionais que irão contribuir para elucidação da regulação da biossíntese de tococromanóis na espécie.

VI. Conclusões

- A rota completa de biossíntese da VTE foi reconstruída e foram identificados 41 *loci* de *S. lycopersicum* envolvidos estruturalmente na via.
- A maioria dos genes da rota central da VTE estão localizados nos cromossomos 7, 8 e 9.
- A partir de um perfil completo do conteúdo de tocoferóis para os diferentes isômeros (α-, β-, γ-, δ-), 12 QTL para VTE foram identificados, alguns dos quais mostraram alta herdabilidade.
- A abordagem de genes candidatos implementada permitiu a identificação de 16 genes que co-localizam com os QTL descritos.
- A análise comparativa dos alelos de S. lycopersicum e S. pennellii revelou polimorfismos nas sequências de nucleotídeos e de aminoácidos que reforçam a candidatura dos genes na determinação das mudanças no conteúdo de tocoferóis observadas nos frutos das ILs.

VII. Referências Bibliográficas

- Abushita AA, Hebshi EA, Daood HG, Biacs, PA. 1997. Determination of antioxidant vitamins in tomatoes. *Food Chemistry* 60, 207–212.
- Aharoni A, Jongsma MA, Bouwmeester HJ. 2005. Volatile science? Metabolic engineering of terpenoids in plants. *Trends in Plant Science* 10, 594-602.
- Alonso-Blanco C, Aarts MG, Bentsink L, Keurentjes JJ, Reymond M, Vreugdenhil D, Koornneef M. 2009. What has natural variation taught us about plant development, physiology, and adaptation? *The Plant Cell* 21,1877-1896.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.
- Ament K, Van Schie CC, Bouwmeester HJ, Haring MA, Schuurink RC. 2006. Induction of a leaf specific geranylgeranyl pyrophosphate synthase and emission of (E,E)-4,8,12-trimethyltrideca-1,3,7,11-tetraene in tomato are dependent on both jasmonic acid and salicylic acid signaling pathways. *Planta* 224, 1197–1208.
- Apel K, Hirt, H. 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* 55, pp. 373–399.
- Ayres M, Ayres Júnior M, Ayres DL, Santos AA. 2007. Bioestat Aplicações estatísticas nas áreas das ciências bio-médicas. Sociedade Civil Mamirauá/MCT CNPq, Belém, PA, Brasil.
- Bankaitis VA, Mousley CJ, Schaaf G. 2009. The Sec14 superfamily and mechanisms for crosstalk between lipid metabolism and lipid signaling. *Trends in Biochemical Sciences* 35, 150–160.
- Barone A, Chiusano ML, Ercolano MR, Giuliano G, Grandillo S, Frusciante L. 2008. Structural and functional genomics of tomato. *International Journal* of *Plant Genomics*. 2008, doi:10.1155/2008/820274.
- Beecher G. R. 1998. Nutrient content of tomatoes and tomato products. *Proc. Soc. Exp. Biol. Med.* 218, 98-100.
- Bernatzky R, Tanksley SD. 1986. Genetics of actin-related sequences in tomato. *Theoretical and Applied Genetics* 72, 314–315.
- Bischoff M, Rösler J, Raesecke HR, Görlach J, Amrhein N, Schmid J. 1996. Cloning of a cDNA encoding 3-dehydroquinate synthase from a higher plant, and analysis of the organ-specific and elicitor-induced expression of the corresponding gene. *Plant Molecular Biology* 31, 69–76.
- Bischoff M, Schaller A, Bieri F, Kessler F, Amrhein N, Schmid J. 2001. Molecular characterization of tomato 3-dehydroquinate dehydrataseshikimate:NADP oxidoreductase. *Plant Physiology* 125, 1891–1900.
- Botella-Pavia P, Besumbes O, Phillips MA, Carretero-Paulet L, Boronat A, Rodriguez-Concepcion M. 2004. Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate

reductase in controlling the supply of plastidial isoprenoid precursors. *The Plant Journal* 40, 188–199.

- Bouvier F, Suire C, d'Harlingue A, Backhaus RA, Camara B. 2000. Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. *The Plant Journal* 24, 241–252.
- Cahoon EB, Hall SE, Ripp KG, Ganzke TS, Hitz WD, Coughlan SJ. 2003. Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production and increased antioxidant content. *Nature Biotechnology* 21, 1082–1087.
- Camargo, F. P.; Alves, H.S.; Camargo Filho, W.P.; Vilela N. J. 2006. Cadeia produtiva de tomate industrial no Brasil: resenha da década de 1990, produção regional e perspectivas. *Informações econômicas* 36, 11.
- Chander S, Guo YQ, Yang XH, Yan JB, Zhang YR, Song TM, Li JS. 2008. Genetic dissection of tocopherol content and composition in maize grain using quantitative trait *loci* analysis and the candidate gene approach. *Molecular Breeding* 22, 353–365.
- Chun J, Lee J, Ye L, Exler J, Eitenmiller RR. 2006. Tocopherol and tocotrienol contents of raw and processed fruits and vegetables in the United States diet. *Journal of Food Composition and Analysis* 19, 196–204.
- Collakova E, Dellapenna D. 2003a. Homogentisate phytyltransferase activity is limiting for tocopherol biosynthesis in Arabidopsis. *Plant Physiol*ogy 131, 632–642.
- Collakova E, Dellapenna D. 2003b. The role of homogentisate phytyltransferase and other tocopherol pathway enzymes in the regulation of tocopherol synthesis during abiotic stress. *Plant Physiol*ogy 133, 930-940.
- Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* 16, 10881–10890.
- Das S, Gautam N, Dey SK, Maiti T, Roy S. 2009. Oxidative stress in the brain of nicotine-induced toxicity: protective role of *Andrographis paniculata Nees* and vitamin E. *Applied Physiology, Nutrition, and Metabolism* 34,124–135.
- Davuluri GR, van Tuinen A, Fraser PD, Manfredonia A, Newman R, Burgess D, Brummell DA, King SR, Palys J, Uhlig J, Bramley PM, Pennings HM, Bowler C. 2005. Fruit-specific RNAi-mediated suppression of DET1 enhances tomato nutritional quality. *Nature Biotechnology* 23, 890–895.
- DellaPenna, D.; Last, R.L. 2006. Progress in the dissection and manipulation of plant vitamin E biosynthesis. *Physiologia Plantarum*. 126, 356–368.
- Ding L, Hofius D, Hajirezaei MR, Fernie AR, Börnke F, Sonnewald U. 2007. Functional analysis of the essential bifunctional tobacco enzyme 3dehydroquinate dehydratase/shikimate dehydrogenase in transgenic tobacco plants. *Journal of Experimental Botany* 58, 2053–2067.
- Eberhard J, Bischoff M, Raesecke HR, Amrhein N, Schmid J. 1996. Isolation of a cDNA from tomato coding for an unregulated, cytosolic chorismate mutase. *Plant Molecular Biology* 31, 917–922.

- Emanuelsson O, Brunak S, von Heijine G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* 2, 953– 971.
- Enfissi EM, Barneche F, Ahmed I, Lichtle C, Gerrish C, McQuinn RP, Giovannoni JJ, Lopez-Juez E, Bowler C, Bramley PM, Fraser PD. 2010. Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. *The Plant Cell* 22, 1190–1215.
- Entus R, Poling M, Herrmann KM. 2002. Redox Regulation of *Arabidopsis* 3-Deoxy-D-arabino-Heptulosonate 7-Phosphate Synthase. *Plant Physiology* 129, 1866–1871.
- Eshed Y, Zamir D. 1995. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141, 1147–1162.
- Estévez JM, Cantero A, Reindl A, Reichler S, León P. 2001. 1-Deoxy-dxylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *Journal of Biological Chemistry* 276, 22901-22909.
- Falk J, Munné-Bosch S. 2010. Tocochromanol functions in plants: antioxidation and beyond. *Journal of Experimental Botany* 61, 1549–1566.
- Fei Z, Tang X, Alba R, Giovannoni J. 2006. Tomato Expression Database (TED): a suite of data presentation and analysis tools. *Nucleic Acids Research* 34, D766–D770.
- Fernie AR, Yaakov T, Zamir D. 2006. Natural genetic variation for improving crop quality. *Current Opinion in Plant Biology* 9, 196–202.
- Fernie AR, Schauer N. 2009. Metabolomics-assisted breeding: A viable option for crop improvement? *Trends in Genetics* 25, 39–48.
- Ferro M, Brugiere S, Salvi D, Seigneurin-Berny D, Court M, Moyet L, Ramus C, Miras S, Mellal M, Le Gall S, Kieffer-Jaquinod S, Bruley C, Garin J, Joyard J, Masselon C, Rolland N. 2010. AT_CHLORO: a comprehensive chloroplast proteome database with sub-plastidial localization and information for functional genomics using quantitative label-free analyses. *Molecular & Cell Proteomics*, 10.1074/mcp.M900325-MCP200.
- Foyer CH, Noctor G. 2005. Redox homeostasis and antioxidant signaling: a metabolic interface between stress perception and physiological responses. *The Plant Cell* 17, 1866–1875.
- Fraser PD, Enfissi EM, Halket JH, Truesdale MR, Yu DM, Gerrish C, Bramley PM. 2007. Manipulation of phytoene levels in tomato fruit: Effects on isoprenoids, plastids and intermediary metabolism. *The Plant Cell* 19, 3194–3211.
- Fraser PD, Pinto ME, Holloway DE, Bramley PM. 2000. Application of highperformance liquid chromatography with photodiode array detection to the metabolite profiling of plant isoprenoids. *The Plant Journal* 24, 551–558.
- Fridman E, Pleban T, Zamir D. 2000. A recombination hotspot delimits a wildspecies quantitative trait *locus* for tomato sugar content to 484 bp within an invertase gene. *Proceedings of the National Academy of Sciences*, USA 97, 4718-4723.

- Fridman E, Carrari F, Liu Y, Fernie A, Zamir D. 2004. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* 305, 1786.
- Friso G, Majeran W, Huang M, Sun Q, van Wijk KJ. 2010. Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: Large-scale quantitative proteomics using the first maize genome assembly. *Plant Physiology* 152, 1219–1250.
- Furuya T, Yoshikawa T, Kimura T, Kaneko H. 1987. Production of tocopherols by cell-culture of safflower. *Phytochemistry* 26, 2741–2747.
- Garcia I, Rodgers M, Pepin R, Hssich T, Matringe M. 1999. Characterization and subcellular compartmentation of recombinant 4hydroxyphenylpyruvate dioxygenase from *Arabidopsis* in transgenic tobacco. *Plant Physiology* 119, 1507–1516.
- Gasser CS, Winter JA, Hironaka CM, Shah DM. 1988. Structure, expression, and evolution of the 5-enolpyruvylshikimate-3-phosphate synthase genes of petunia and tomato. *The Journal of Biological Chemistry* 263, 4280–4287.
- Gilliland LU, Magallanes-Lundback M, Hemming C, Supplee A, Koornneef M, Bentsink L, Dellapenna D. 2006. Genetic basis for natural variation in seed vitamin E levels in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* 103, 18834–18841.
- Giovannoni JJ. 2004. Genetic regulation of fruit development and ripening. Plant Cell 16,170–180.
- Görlach J, Beck A, Henstrand JM, Handa AK, Herrmann KM, Schmid J, Amrhein N. 1993. Differential expression of tomato (*Lycopersicon esculentum* L.) genes encoding shikimate pathway isoenzymes: I. 3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase. *Plant Molecular Biology* 23, 697–706.
- Görlach J, Raesecke HR, Abel G, Wehrli R, Amrhein N, Schmid J. 1995. Organspecific differences in the ratio of alternatively spliced chorismate synthase (LeCS2) transcripts in tomato. *The Plant Journal* 8, 451–456.
- Graindorge M, Giustini C, Jacomin AC, Kraut A, Curien G, Matringe M. 2010. Identification of a plant gene encoding glutamate/aspartate-prephenate aminotransferase: the last homeless enzyme of aromatic amino acids biosynthesis. *FEBS Letters* 584, 4357–4360.
- Hanahan, D. 1983. Studies on transformation of *E.coli* with plasmids. *Journal of Molecular Biology* 156, 557-580.
- Herbers K. 2003. Vitamin production in transgenic plants. *Journal of Plant Physiology* 160, 821–829.
- Hosomi A, Arita M, Sato Y, Kiyose C, Ueda T, *et al.* 1997. Affinity for alphatocopherol transfer protein as a determinant of the biological activities of vitamin E analogs. *FEBS Letters* 409, 105–108.
- Hörtensteiner S. 2006. Chlorophyll degradation during senescence. Annual Review of Plant Biology 57, 55-77.

- IOM (Institute of Medicine). 2000. Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids, National Academy Press, Washington, DC.
- Ischebeck T, Zbierzak AM, Kanwischer M, Dormann P. 2006. A salvage pathway for phytol metabolism in *Arabidopsis*. *The Journal of Biological Chemistry* 281, 2470–2477.
- Joyard J, Ferro M, Masselon C, Seigneurin-Berny D, Salvi D, Garin J, Rolland N. 2009. Chloroplast proteomics and the compartmentation of plastidial isoprenoid biosynthetic pathways. *Molecular Plant* 2, 1154–1180.
- Kamenetzky L, Asís R, Bassi S, *et al.* 2010. Genomic analysis of wild tomato (*Solanum pennellii*) introgressions determining metabolic and yield associated traits. *Plant Physiology*152, 1772-1786.
- Kahlau S, Aspinall S, Gray JC, Bock R. 2006. Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *Journal of Molecular Evolution* 63, 194–207.
- Kanwischer M, Porfirova S, Bergmüller E, Dörmann P. 2005. Alterations in tocopherol cyclase activity in transgenic and mutant plants of Arabidopsis affect tocopherol content, tocopherol composition, and oxidative stress. *Plant Physiology* 137, 713–723.
- Knaap E, Bohs L, Nee M: Spooner DM. 2004. Solanaceae a model for linking genomics with biodiversity. Comparative and Functional Genomics 5, 285-291.
- Krieger-Liszkay A, Trebst A. 2006. Tocopherol is the scavenger of singlet oxygen produced by the triplet states of chlorophyll in the PSII reaction centre. Journal of Experimental Botany 57, 1677–1684.
- Li Y, Wang Z, Sun X, Tang K. 2008. The current opinions on the functions of tocopherol based on the genetic manipulation of tocopherol biosynthesis in plants. *Journal of Integrative Plant Biology* 50, 1057–1069.
- Li H, Liu H, Han Y, Wu X, Teng W, Liu G, Li W. 2010. Identification of QTL underlying vitamin E contents in soybean seed among multiple environments. *TAG Theoretical and Applied Genetics* 120, 1405-1413.
- Liebler DC. 1993. The role of metabolism in the antioxidant function of vitamin E. *Critical Reviews in Toxicology* 23, 147–169
- Lippman ZB, Semel Y, Zamir D. 2007. An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Current Opinion in Genetics & Development* 17, 545–552.
- Lois LM, Rodríguez-Concepción M, Gallego F, Campos N, Boronat A. 2000. Carotenoid biosynthesis during tomato fruit development: regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase. *The Plant Journal* 22, 503– 513.
- Marwede V, Gul MK, Becker HC, Ecke W. 2005. Mapping of QTL controlling tocopherol content in winter oilseed rape. *Plant Breeding* 124, 20–26.
- Mauricio R. 2001. Mapping quantitative trait *loci* in plants: uses and caveats for evolutionary biology. *Nature Reviews Genetics* 2, 370-381.

- Mène-Saffrané L, DellaPenna D. 2009. Biosynthesis, regulation and functions of tocochromanols in plants. *Plant Physiology and Biochemistry* 48, 301– 309.
- Mobley EM, Kunkel BN, Keith B. 1999. Identification, characterization and comparative analysis of a novel chorismate mutase gene in *Arabidopsis thaliana*. *Gene* 240, 115–123.
- Munne-Bosch S, Alegre L. 2002. The function of tocopherols and tocotrienols in plants. *Critical Reviews in Plant Sciences* 21, 31–57.
- Nagegowda DA. 2010. Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Letters* 584, 2965–2973.
- Paetzold H, Garms S, Bartram S, Wieczorek J, Uros-Gracia EM, Rodriguez-Concepción M, Boland W, Strack D, Hause B, Walter MH. 2010. The isogene 1-deoxy-D-xylulose 5-phosphate synthase II controls isoprenoid profiles, precursor pathway allocation and density of tomato trichomes. *Molecular Plant* ssq032v1-ssq032.
- Peterman TK, Ohol YM, McReynolds LJ, Luna EJ. 2004. Patellin1, a novel Sec14-like protein, localizes to the cell plate and binds phosphoinositides. *Plant Physiology* 136, 3080–3094.
- Phillips MA, D'Auria JC, Gershenzon J, Pichersky E. 2008. The Arabidopsis thaliana type I Isopentenyl Diphosphate Isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis. *The Plant Cell* 20, 677–696.
- Ravanel S, Cherest H, Jabrin S, Grunwald D, Surdin- Kerjan Y, Douce R, Rebeille F. 2001. Tetrahydrofolate biosynthesis in plants: molecular and functional characterization of dihydrofolate synthetase and three isoforms of folylpolyglutamate synthetase in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* 98, 1523–1535.
- Ricciarelli R, Zingg JM, Azzi A. 2002. The 80th Anniversary of Vitamin E: Beyond Its Antioxidant Properties. *Biological Chemistry* 383, 457-465.
- Rippert P, Scimemi C, Dubald M, Matringe M. 2004. Engineering plant shikimate pathway for production of tocotrienol and improving herbicide resistance. *Plant Physiology* 134, 92–100.
- Rippert P, Puyaubert J, Grisollet D, Derrier L, Matringe M. 2009. Tyrosine and phenylalanine are synthesized within the plastids in *Arabidopsis*. *Plant Physiology* 149, 1251–1260.
- Rise M, Cojocaru M, Gottlieb HE, Goldschmidt E. 1989. Accumulation of αtocopherol in senescing organs as related to chlorophyll degradation. *Plant Physiol*ogy 89,1028–1030.
- Rodríguez-Concepción M, Ahumada I, Diez-Juez E, Sauret-Güeto S, Lois LM, Gallego F, Carretero-Paulet L, Campos N, Boronat A. 2001. 1-Deoxy-Dxylulose 5-phosphate reductoisomerase and plastid isoprenoid biosynthesis during tomato fruit ripening. *The Plant Journal* 27, 213–222.
- Rodríguez-Concepción M, Querol J, Lois LM, Imperial S, Boronat A. 2003. Bioinformatic and molecular analysis of hydroxymethylbutenyl diphosphate synthase (GCPE) gene expression during carotenoid accumulation in ripening tomato fruit. *Planta* 217, 476–482.

- Rohdich F, Wungsintaweekul J, Lüttgen H, Fischer M, Eisenreich W, Schuhr CA, Fellermeier M, Schramek N, Zenk MH, Bacher A. 2000. Biosynthesis of terpenoids: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase from tomato. *Proceedings of the National Academy of Sciences, USA* 97, 8251–8256.
- Ronen G, Carmel-Goren L, Zamir D, Hirschberg J. 2000. An alternative pathway to ß-carotene formation in plant chromoplasts discovered by map-based cloning of *Beta* and *old-gold* color mutations in tomato. *Proceedings of the National Academy of Sciences, USA* 97, 11102–11107.
- Ros E. 2009. Nuts and novel biomarkers of cardiovascular disease. *The American Journal of Clinical Nutrition* 89, 1649S–1656S.
- Rothan C, Causse M. 2007. Natural and artificially induced genetic variability in crop and model plant species for plant systems biology. *Experientia Supplementum* 97, 21-53.
- Sadre R, Gruber J, Frentzen M. 2006. Characterization of homogentisate prenyltransferases involved in plastoquinone-9 and tocochromanol biosynthesis. *FEBS Letters* 580, 5357–5362.
- Saito K, Tautz L, Mustelin T. 2007. The lipid-binding SEC14 domain. *Biochimica et Biophysica Acta* 1771, 719–726.
- Schauer N, Zamir D, Fernie AR. 2005. Metabolic profiling of leaves and fruit of wild species tomato: a survey of the Solanum lycopersicon complex. Journal of Experimental Botany 56, 297–307.
- Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR. 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature Biotechnology* 24, 447–454.
- Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR. 2008. Mode of inheritance of primary metabolic traits in tomato. *The Plant Cell* 20, 509–523.
- Schenk, N., Schelbert, S., Kanwischer, M., Goldschmidt, E.E., Dörmann, P., and Hörtensteiner, S. 2007. The chlorophyllases AtCLH1 and AtCLH2 are not essential for senescence-related chlorophyll breakdown in *Arabidopsis thaliana*. *FEBS Letters* 581, 5517–5525.
- Schmid J, Schaller A, Leibinger U, Boll W, Amrhein N. 1992. The *in-vitro* synthesized tomato shikimate kinase precursor is enzymatically active and is imported and processed to the mature enzyme by chloroplasts. *The Plant Journal* 2, 375–383.
- Semchuk NM, Lushchak OV, Falk J, Krupinska K, Lushchak VI. 2009. Inactivation of genes, encoding tocopherol biosynthetic pathway enzymes, results in oxidative stress in outdoor grown *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* 47, 384–390.
- Shibata D. 2005. Genome sequencing and functional genomics approaches in tomato. *Journal of General Plant Pathology* 71, 1–7.

Shintani D, DellaPenna D. 1998. Elevating the vitamin E content of plants through metabolic engineering. *Science* 282, 2098–2100

Sokal RR, Rohlf FJ. 1981. Biometry. Freeman Secon Edition, New York.

- Soll J, Schultz G. 1979. Comparison of geranylgeranyl and phytyl substituted methylquinols in the tocopherol synthesis of spinach chloroplasts. *Biochemical and Biophysical Research Communications* 91, 715–720.
- Stitt M, Sulpice R, Keurentjes J. 2010. Metabolic networks: how to identify key components in the regulation of metabolism and growth? Plant Physiology. 152, 428–444.
- Sun Q, Zybailov B, Majeran W, Friso G, Olinares, PDB, van Wijk KJ. 2009. PPDB, the Plant Proteomics Database at Cornell. *Nucleic Acids Research* 37, 969–974.
- Tabor HK, Risch NJ, Myers RM. 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature reviews genetics* 3, 1-7.
- Tian L, DellaPenna D, Dixon RA. 2007. The pds2 mutation is a lesion in the *Arabidopsis* homogentisate solanesyltransferase gene involved in plastoquinone biosynthesis. *Planta* 226, 1067–1073.
- Tohge T, Fernie AR. 2010. Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nature Protocols* 5, 1210-1227.
- Tsegaye Y, Shintani DK, DellaPenna D. 2002. Over-expression of the enzyme p-hydroxyphenylpyruvate dioxygenase in Arabidopsis and its relationship to tocopherol biosynthesis. *Plant Physiol Biochem* 40, 913–920.
- Tzin V, Malitsky S, Aharoni A, Galili G. 2009. Expression of a bacterial bifunctional chorismate mutase/prephenate dehydratase modulates primary and secondary metabolism associated with aromatic amino acids in *Arabidopsis*. *The Plant Journal* 60, 156–167.
- Tzin V, Galili G. 2010. New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Molecular plant* 10.1093/mp/ssq048.
- Valentin HE, Lincoln K, Moshiri F, Jensen PK, Qi Q, Venkatesh TV, Karunanandaa B, Baszis SR, Norris SR, Savidge B, Gruys KJ, Last RL. 2006. The *Arabidopsis vitamin E pathway gene5-1* mutant reveals a critical role for phytol kinase in seed tocopherol biosynthesis. *The Plant Cell* 18, 212–224.
- Van Schie CCN, Ament K, Schmidt A, Lange T, Haring MA, Schuurink RC. 2007. Geranyl diphosphate synthase is required for biosynthesis of gibberellins. *The Plant Journal* 52, 752–762.
- van der Hoeven R, Ronning C, Giovannoni J, Martin G, Tanksley SD. 2002. Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *The Plant Cell* 14, 1441– 1456.
- Voorrips RE. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* 93, 77–78.

- Zamir D. 2001. Improving plant breeding with exotic genetic libraries. *Nature Reviews Genetics* 2, 983–989.
- Zhang CX, Ho SC, Chen YM, Fu JH, Cheng SZ, Lin FY. 2009. Greater vegetable and fruit intake is associated with a lower risk of breast cancer among Chinese women. *International Journal of Cancer* 125,181–188.
- Zhao J, Last RL. 1995. Immunological characterization and chloroplast localization of the tryptophan biosynthetic enzymes of the flowering plant *Arabidopsis thaliana. The Journal of Biological Chemistry* 270, 6081–6087

Anexo I

Anexo I – Alinhamentos das sequências de aminoácidos deduzidas dos alelos de *S. lycopersicum* e *S. pennellii*.

ggps4_pen ggps4_lyc	MVFSMVMSFSPSLCLPRSRMVMQKAIQCSSSVSTASESVKFDLKTYWTTLIVDINQKLDE MVFSMVMSFSPSLCLPRSRMVMQKAIQCSSSVSTASESVKFDLKTYWTTLISDINQKLDE ************************************	60 60
ggps4_pen ggps4_lyc	AVPVKYPNQIYEAMRYSVLAKGAKRSPPIMCVAACELFGGNRLAAFPTACALEMVHAASL AVPVKYPNQIYEAMRYSVLAKGAKRSPPIMCVAACELFGGNRLAAFPTACALEMVHAASL ***********************************	120 120
ggps4_pen ggps4_lyc	IHDDLPCMDDDTTRRGLPANHTVFGVDMAILAGDALFPLGFQHIVSHTPSDLVPEDRILR IHDDLPCMDDDTTRRGLPANHTVFGVDMAILAGDALFPLGFQHIVSHTPSDLVPEDRVLR ***********************************	180 180
ggps4_pen ggps4_lyc	VITEIARAVGSTGMAAGQFLDLEGGPNAVDFVQEKKYGEMGECSAVCGALLAGASDEEIQ VITEIARAVGSTGMAAGQFLDLEGGPNAVDFVQEKKYGEMGECSAVCGALLAGASDEEIQ ***********************************	240 240
ggps4_pen ggps4_lyc	HMRKYGRAVGVLYRVVDDILEAKKTEDKTEGKKKKGKSYVSVYGIEKAVKVAEDLRAQAK HMRKYGRAVGVLYRVVDDILEAKKTENKTEGKKKKGKSYVSVYGIEKAVKVAEDLRAQAK ************************	300 300
ggps4_pen ggps4_lyc	RELVGLEKYGDKVMPLYSFLDYAADRGFSIDGQV 334 RELDGLEKYGDKVMPLYSFLDYAADRGFSIDGQV 334 *** ************************	
tat2_pen tat2_lyc	MENGTTTGRRRWNFKENEKLVSVSDLTVRSVLNKLMCCVDPADTRPTIPLGHGDPSAFPC MENGTTTGMRRWNFKENEKLVSVSDLTVRSVLNKLRCCVDPADTRPTIPLGHGDPSAFPC ******* *****************************	60 60
tat2_pen tat2_lyc	FRTTPIAEDAISDAVRSAMFNGYSSTVGILPARRAVAEYLSQDLPYKLSPDDIYLTSGCG FLTTTIAEDAISDAVRSAKFNGYSSTVGILPARRAVAEYLSQDLPYKLSPDDIYLTSGCG * **.	120 120
tat2_pen tat2_lyc	QAIEILLNALARPNANILLPTPGFPYYEAWGGFTQMEMRHFNLLPEKEWEVDLNAVESLA QAIEILLNALARPNANILLPSPGFPYYEAWGGFTQMEMRHFNLLPEKEWEVDLNAVESLA ********************	180 180
tat2_pen tat2_lyc	DENTVAMVIINPGNPCGNVYSEEHLKKVAETARKLGILVISDEVYAHLAFGSKPFVPMGI DENTVAMVIINPGNPCGNVYSEEHLKKVAETARKLGILVISDEVYAHLAFGSKPFVPMGI ************************************	240 240
tat2_pen tat2_lyc	FGSIAPVVTLGSISKRWIVPGWRLGWLVTNDPNGILKEHGDIDSIMGYLNISTDPATFIQ FGSIAPVVTLGSISKRWIVPGWRLGWLVTNDPNGILKEHGDIDSIMGYLNISTDPATFIQ ************************************	300 300
tat2_pen tat2_lyc	GAIPQILHETKDDFFSKIVNMLREDADICYERIKDIPCITCPSKPQGSMFLMVQLHLNLL GAIPQILHETKDDFFSKIVNMLREDADICYERIKDIPCITCPSKPQGSMFLMVQLHLNLL *********************************	360 360
tat2_pen tat2_lyc	EDIEDDLDFCAKLAKEESLIILPGVAVGLKNWLRITFACEPSYLEDGFQRLNAFYKRHAK EDIEDDLDFCAKLAKEESLIILPGVAVGLKNWLRITFACEPSYLEDGFQRLNAFYKRHAK ************************************	420 420
tat2_pen tat2_lyc	KQ 422 KQ 422 **	

tyrA1_pen	$\tt MLSFTPLQSKPTPTSNSNRFSNLTNPTSSTSRRHFSVSSPSQVSHHHGCRRLSIKAIDA$	60
tyra1_lyc	MLSFTPLQSKPTPTSNSNRFSNLTNPTSSTSRRHFSVSSPSQVSHHHGRRRLSIKAIDA ***********************************	60
tyrA1_pen tyra1_lyc	AQPYDYEALVSNQYAQSGRLKIAIVGFGNFGQFLAKSFVSKGHFVLAHSRTDYSQIANSL AQPYDYEALVSNQYAQSGRLKIAIVGFGNFGQFLAKSFVSKGHFVLAHSRTDYSQIANSL ************************************	120 120
tyrA1_pen tyra1_lyc	GVSFFQDPHDLCEQHPDVIVLCTSIISTETVLRSLPIQRLKRNTLFVDVLSVKEFPKNIF GVSFFQDPHDLCEQHPDVIVLCTSIISTETVLRSLPIQRLKRNTLFVDVLSVKEFPKNIF ************************************	180 180
tyrA1_pen tyra1_lyc	LQVLPTHFDILCTHPMFGPESGKDSWKDLIFMFDKVRIGEGRSRTARVDKFLDIFEKEGC LQVLPTHFDILCTHPMFGPESGKDSWKDLIFMFDKVRIGEGRSRTARVDKFLDIFEKEGC ************************************	240 240
tyrA1_pen tyra1_lyc	RMVPMTCAEHDKHAAGSQFITHTMGRVLEKLGLESTPINTKGYETLLNLVDNTASDSFDL RMVPMTCAEHDKHAAGSQFITHTMGRVLEKLGLESTPINTKGYETLLNLVDNTASDSFDL ************************************	300 300
tyrA1_pen tyra1_lyc	YYGLFMYNKNAMEELERLDLAFEALKKELFGHLHDLLRKQLFGKAEEAGQRRVLSKLPRN YYGLFMYNKNAMEELERLDLAFEALKKELFGHLHDLLRKQLFGKAEEAGQRRVLSKLPRN ************************************	360 360
tyrA1_pen tyra1_lyc	GYALPAPSSDAVKPENN 377 GYALPAPSSDAVKPENN 377 ******	

tyra2_pen tyra2_1yc	MFSLSSIQSNNIQSQSCSSLLFNHHHQHHHHSIISTRFHHHRLLFPLRAQNSDLTTSTNN MFSLSSIQSNNIQSQSSSSLLFNHHHQHSTISTRFHHHRLLFPLRAQNSDLTTATTN *********************************	60 57
tyra2_pen tyra2_lyc	NNYVDLDDNLTRLDKFSKSLSLSNIEENTSLNPLLCSNNKLKIAIIGFGNFGQFIAKSFI NNYVDLDDNLTRLDKFSKSLSISNIEENTSLNPLLCSNNKLKIAIIGFGNFGQFIAKSFI ************************************	120 117
tyra2_pen tyra2_1yc	KQGHIVLAHSRSDYSLIAQSLDVHFFQDPNDLCEQHPDIILLCTSINSLENVIRSLPIQK KQGHVVLAHSRSDYSLIAQSLNVHFFQDPNDLCEQHPDVILLCTSINSLENVIRSLPIQK ****:********************************	180 177
tyra2_pen tyra2_lyc	LKRNTLFVDVLSVKEFPKNIFLQSLPKEFDILCTHPMFGPTSGKDNWKGLPFMYDKVRIG LKRNTLFVDVLSVKEFPKNIFLQSLPKEFDILCTHPMFGPTSGKDNWKGLPFMYDKVRIG ************************************	240 237
tyra2_pen tyra2_lyc	QEESRIKRVNNFINIFVREGCRMVEMSCSEHDKYAAGSQFITHTIGRMLQKLGTQTTPIN QEESRIKRVNNFINIFVKEGCRMVEMSCSEHDKYAAGSQFITHTIGRMLQRLGTQTTPIN ************************************	300 297
tyra2_pen tyra2_lyc	TKGYESLLNLMENTTSDSFDLYCGLFMYNNNSMEVLEKLDAALDSLKRELFGQVLQKLEK TKGYESLLNLMENTTSDSFDLYCGLFMYNNNSMEVLEKLDAALDSLKRELFGQVLQKLEK ***********************************	360 357
tyra2_pen tyra2_lyc	RVEKGSRLVLPTPNFSKKIEKLKVERKELEALS 393 RVEKGSKLALPTPDFSKKIEKLKVERKELEALS 390 ******:*.****	

vte4_pen vte4_lyc	MGSQCYSGYSIQSLNPTCPSSSSSSVIFTLLKPQIHRRRIITCCNS-RRRRMASVAAMN MGSQCYSAYSIQSLNPTCPSSSSSSVIFTLLKPQIHRRRIITCCNSSRRRRMASVAAMN *******	59 60
vte4_pen vte4_lyc	AVSSSSVEVGIQNQQELKKGIADLYDESSGIWEDIWGDHMHHGYYEPKSSVELSDHRAAQ AVSSSSVEVGIQNQQELKKGIADLYDESSGIWEDIWGDHMHHGYYEPKSSVELSDHRAAQ ***********************************	119 120
vte4_pen vte4_lyc	IRMIEQALSFAAISEDPAKKPTSIVDVGCGIGGSSRYLAKKYGATAKGITLSPVQAERAQ IRMIEQALSFAAISEDPAKKPTSIVDVGCGIGGSSRYLAKKYGATAKGITLSPVQAERAQ *****	179 180
vte4_pen vte4_lyc	ALADAQGLGDKVSFQVADALNQPFPDGQFDLVWSMESGEHMPNKEKFVGELARVAAPGGT ALADAQGLGDKVSFQVADALNQPFPDGQFDLVWSMESGEHMPNKEKFVGELARVAAPGGT ***********************************	239 240
vte4_pen vte4_lyc	IILVTWCHRDLSPSEESLTPEEKELLNKICKAFYLPAWCSTADYVNLLQSNSLQDIKAED IILVTWCHRDLSPSEESLTPEEKELLNKICKAFYLPAWCSTADYVKLLQSNSLQDIKAED ************************************	299 300
vte4_pen vte4_lyc	WSENVAPFWPAVIKSALTWKGFTSVLRSGWKTIKAALAMPLMIEGYKKGLIKFAIITCRK WSENVAPFWPAVIKSALTWKGFTSVLRSGWKTIKAALAMPLMIEGYKKGLIKFAIITCRK ************************************	359 360
vte4_pen vte4_lyc	PE 361 PE 362 **	
vtel_pen vtel_lyc	MECVSAISPISKNGGFCRIRTHFATSIANGEVLLNNYSSTVLKLQSQKSRHAFVVKADSA MECVSAISPISKNGGFCRIRTHFATSIANGEVLLNNYSSAVLKLQSQKSRHAFVVKADSA ***********************************	60 60
vtel_pen vtel_lyc	VDTTKKENREPVKPVYSSTPSNRPLRTPHSGYHFDGSTRKFFEGWFFKVSIPECRQSFCF VDTTKKENREPVKPVYSSTPSNRPLRTPHSGYHFDGSTRKFFEGWFFKVSIPECRQSFCF ***********************************	120 120
vtel_pen vtel_lyc	MYSVESPSFTKKLSSLEELQYGPRFTGVGAQILGADDKYICQYSEESSNFWGSRHELMLG MYSVESPSFTKKLSSLEELQYGPRFTGVGAQILGADDKYICQYSEESSNFWGNRHELMLG ***********************************	180 180
vtel_pen vtel_lyc	NTFVAQNSAKPPNKEVRPQEFNRCVTEGFQVTPLWHQGSIRDDGRTDYTEIVKTASWEYS NTFVAQNSAKPPNKEVSPQEFNRCVTEGFQVTPLWHQGSIRDDGRTDYTEIVKTASWEYS ************************************	240 240
vtel_pen vtel_lyc	TRPIYGWGDVNSKQKSTAGWPAAFPVFEPHWQVCMAAGLSTGWIEWDGQRFEFQNAPSYS TRPIYGWGDVNSKQKSTAGWPAAFPVFEPHWQVCMAAGLSTGWIEWDGQRFEFQNAPSYS ***********************************	300 300
vtel_pen vtel_lyc	EKNWGGSFPRKWFWVQCSVFEGAIGDVALTAGGGLRLLPGLNETFESVALIGIHYRGIFY EKNWGGSFPRKWFWVQCSVFEGAIGDVALTAGGGLRLLPGLNETFESVALIGIHYRGIFY ************************************	360 360
vtel_pen vtel_lyc	EFVPWNASVSWEITPWGKWHISAENETHMVLLEATTEDPGTTLRAPTEETGLAPACRDTC EFVPWNASVSWEITPWGKWHISAENETHMVLLEATAEDPGTTLRAPTEEMGLAPACRDTC ************************************	420 420
vtel_pen vtel_lyc	FGELRLQLWERKSNGSKGKVILDVTSNMAGLEVGGGPWFNTWKGNAEMPEIVTRAINVPV FGELRLQLWERKSNGSKGKVILDVTSNMAGLEVGGGPWFNTWKGNAEMPEIVTRAINVPV ***********************************	480 480
vtel_pen vtel_lyc	DLDGIFSCVPSLLKPPGL 498 DLDGIFSCVPSLLKPPGL 498	

vte31_pen vte31_lyc	MANSIFISGTQSLSVCNRVSPNGLGFVGSDFNGNQFPKLGLSRTSYKIKTLAPICSISTS MANSIFISGTQSLSVCNRVSPNGLGFVGSDFNGNQFSKLGLSRTSYKIKTLAPICSISTS **********************************	60 60
vte31_pen vte31_lyc	RPASQPRFIQHKQEAFWFYRFLSIVYDHVINPGHWTEDMRDDALEPADLNDRNLTVVDVG RPASQPRFIQHKQEAFWFYRFLSIVYDHVINPGHWTEDMRDDALEPADLNDRNLTVVDVG **********************************	120 120
vte31_pen vte31_lyc	GGTGFTTLGIVEHVDAKNVTILDQSPHQLAKAKEKEPLKECKIIEGDAEDLPFSTDYADR GGTGFTTLGIVEHVDAKNVTILDQSPHQLAKAKEKEPLKECKIIEGDAEDLPFSTDYADR ************************************	180 180
vte31_pen vte31_lyc	YVSAGSIEYWPDPQRGIREAYRVLKPGGKACLIGPVHPTFWLSRFFADVWMLFPKEEEYI YVSAGSIEYWPDPQRGIREAYRVLKPGGKACLIGPVHPTFWLSRFFADVWMLFPKEEEYI **********************************	240 240
vte31_pen vte31_lyc	EWFEKAGFTDVQLKKIGPKWYRGVRRHGLIMGCSVTGVKSTPGDSPLQLGPKAEDVTKPV EWFEKAGFTDVQLKKIGPKWYRGVRRHGLIMGCSVTGVKSTPGDSPLQLGPKAEDVTKPV ************************************	300 300
vte31_pen vte31_lyc	NPFVFMLRFLLGATAATYYVLVPIYMWVKDQVVPEGEPL 339 NPFVFMLRFLLGATAATYYVLVPIYMWVKDQVVPEGEPL 339 ***********************************	
vte2_pen vte2_lyc	MESLLIGSFHRPSIPVPAEFSPLKTSSHAIVRVLKCKAWKRPKKHYSSSMKLQRQYITQE MESLLIGSFHRPSIPVPAEFSPLKTSSHAIVRVLKCKAWKRPKKHYSSSMKLQRQYITQE ************************************	60 60
vte2_pen vte2_lyc	HVGGSDLSTIAADKKLKRRFLVHASSEHPLESQPSKSPWDSVNDAVDAFYRFSRPHTIIG HVGGSDLSTIAADKKLKGRFLVHASSEHPLESQPSKSPWDSVNDAVDAFYRFSRPHTIIG ***********************************	120 120
vte2_pen vte2_lyc	TALSIISVSLLAVEKFSDFSPLFFTGVLEAIVAALFMNIYIVGLNQLSDIEIDKVNKPYL TALSIISVSLLAVEKFSDFSPLFFTGVLEAIVAALFMNIYIVGLNQLSDIEIDKVNKPYL ************************************	180 180
vte2_pen vte2_lyc	PLASGEYSVQTGVIVVSSFAILSFWLGWIVGSWPLFWALFISFVLGTAYSINLPLLRWKR PLASGEYSVQTGVIVVSSFAILSFWLGWIVGSWPLFWALFISFVLGTAYSINLPLLRWKR ***********************************	240 240
vte2_pen vte2_lyc	FAVVAAMCIFAVRAVIVQIAFYLHIQTYVYRRTAVLSRPLIFATAFMSFFSVVIALFKDI FAVVAAMCIFAVRAVIVQIAFYLHIQTYVYKRAAVLSRPLIFATAFMSFFSVVIALFKDI ************************************	300 300
vte2_pen vte2_lyc	PDIVGDKIFGIQSFTVRLGQERVFWICIGLLEMAYLVAIVVGAASSNTWSKYFTILGHSA PDIVGDKIFGIQSFTVRLGQERVFWICIGLLEMAYLVAIVVGAASSNTWSKYFTILGHSA ************************************	360 360
vte2_pen vte2_1yc	LALLLWTRAKSVDFSSKAAITSFYMFIWKLFYAEYLLIPLVR 402 LALLLWTRAKSVDFSSKAAITSFYMFIWKLFYAEYLLIPLVR 402 ********	

vte5_penMALFASSVLPLSLFKSNSDPYRLFSTKKGFNLGFQKLKTRRLHKTKVVVHFAMFSENPLI60vte5_lycMALFASSVLPLSLLKSNSDPYRLFSTKKGFNLGFQKLKTRRLHRTKVVVHFAMFSENLLI60vte5_penGDLIATALSGGIALSILRFWEETAKRGVFDQKTNRKLVHISIGLVFMLCWPMFSSGQQGA120vte5_lycGDLIATALSGGIALSILRFWEETAKRGVFDQKTNRKLVHISIGLVFMLCWPMFSSGQQGA120

vte5_pen vte5_lyc	ILAAFIPGLNIIKMFLLGLGIWKDDATVKSMSRFGDHRELLKGPLYYALSITCACAIYWR ILAAFIPGLNIIKMFLLGLGIWKDDATVKSMSRFGDHRELLKGPLYYALSITCACAIYWR ************************************	180 180
vte5_pen vte5_lyc	YSPISIGLICNLCAGDGIADIVGRRFGKQKLPYNKNKSFAGSIAMAAAGLLASIGFLHYF YSPISIGLICNLCAGDGIADIVGRRFGKQKLPYNKNKSFAGSIAMAAAGLLASIGFLHYF **********	240 240
vte5_pen vte5_lyc	SLFGYIQVNSKTVLGFLFMSLAAALVESHPLSSELDDNLTVPLTSVLVGSLVL 293 SLFGYIQVNSKTVLGFLFISLAAALVESHPLSSELDDNLTVPLTSVLVGSLVL 293 ************************************	
apt_pen apt_lyc	MALFQCRIPFSAVSPSVSPRIEFTKLFVGNSSNPFVKTTQKKFSSLCYALNDSNKHSSIS MALFQCRIPFSAVSPSVSPRIEFTKLFLGNSSNPFVKTTQKKFSSLCYALKNSNKHSSIS **********************************	60 60
apt_pen apt_lyc	TFSELIEALISGKDLTESESEDSLDFLLDGADEALISAFLVLLRAKGETFEEVVGLARAM TFSELIEALISGKDLTESESEDSLDFLLDGADEALISAFLVLLRAKGETFEEVVGLARAM ***********************************	120 120
apt_pen apt_lyc	IKHCRQVEGLGDAVDIVGTGGDGANTVNISTGASILAAACGAKVAKQGNRSSSSACGSAD IKHCRQVEGLGDAVDIVGTGGDGANTVNISTGASILAAACGAKVAKQGNRSSSSACGSAD ***********	180 180
apt_pen apt_lyc	VLEELGVAIELDPEGVKECVNKAGIGFMMSPIYHPAMKIVRPIRKKLKVKTVFNILGPML VLEELGVAIELDPEGVKECVNQAGIGFMMSPIYHPAMKIVRPIRKKLKVKTVFNILGPML ************************************	240 240
apt_pen apt_lyc	NPARVPFAVVGVYKEDLVHKMAKALQRFGMKRALVVHSEGLDEMSPLGPGLVLDVTPDNI NPARVPFAVVGVYKEDLVHKMAKALQRFGMKRALVVHSEGLDEMSPLGPGLVLDVTPDNI ************************************	300 300
apt_pen apt_lyc	KKFSFDPLDFGIPRCTLESLRGGGPEYNAEMLRRVLSGEKGSIADAFALNASAALLVSGR EKFSFDPLDFGIPRCTLESLRGGGPEYNAEMLRRVLSGEKGSIADAFALNASAALLVSGR :************************************	360 360
apt_pen apt_lyc	VDNLA- 365 VDNLAE 366 *****	
fpas pen	MKNISILCRFTRSIORNMLCSVAGLVHNTRHAFSTLREDPELDEMMEFLDSLKNFEKAGV	60
fpgs_lyc	MKNISILCRFTRSIQRNMLCSVAGLVHNTRHAFSTLREDPELDEMMEFLDSLKNFEKAGV	60
fpgs_pen fpgs_lyc	PKGAGTDSEDGFDLGRMRRLMELLGNPQSRFKTVHIAGTKGKGSTAAFLSSILRAEGYSV PKGAGTDSEDGFDLGRMRRLMELLGNPQSRFKTVHIAGTKGKGSTAAFLSSILRAEGYSV ***********	120 120
fpgs_pen fpgs_lyc	GCYTSPHIQTIRERITLGKWGEPVSAKVLNHHFNTRREVIERAVKLENGYLSHFEVFTAV GCYTSPHIQTIRERITLGKWGEPVSAKVLNHHFNTRREVIERAVKLENGYLSHFEVFTAV ***********	180 180
fpgs_pen fpgs_lyc	AFSLFAEENTEIAVVEAGLGGARDATNVISGSDLAISIITTVGEEHLDALGGSLESIAVA AFSLFAEENTEIAVVEAGLGGARDATNVISGSDLAISIITTVGEEHLDALGGSLESIAVA ***********************************	240 240
fpgs_pen fpgs_lyc	KSGIVKNGRPLVIGGPFVPSIECILRRKASSMSSPVVSASDPGNRSALRGFCEVSGIPRQ KSGIVKNGRPLVIGGPFVPSIECILRRKASSMSSPVVSASDPGNRSALRGFCEVSGIPRQ ************************************	300 300

fpgs_pen fpgs_lyc	LCDIVLQIEKDYDLSIELLGVKLRMLGAHQLQNAVTATCAALCLNKQGWSLSSGAIRAGL LCDIVLQIEKDYDLSIELLGVKLRMLGAHQLQNAVTATCAALCLNKQGWSLSSGAIRAGL ************************************	360 360
fpgs_pen fpgs_lyc	ESAFLQGRTQILSSEEANLLGVSGATVLLDGAHTKESARALANTLQMTFPKAKMVLVVAM ESAFLQGRTQILSSEEAKLLGVSGATVLLDGAHTKESARALAKTLQMTFPKAKMVLVVAM *****************	420 420
fpgs_pen fpgs_lyc	ANDKDHLGFASELLSVGDLDAVFSTEVDIAGAKSRMASASLLKCAWVNALRETNMKVLDL ANDKDHLGFASELLSVGDLDAVFSTEVDIAGAKSRMASASLLKCAWVNASREMNMKVLDL ***********************************	480 480
fpgs_pen fpgs_lyc	KVAES 485 KVAES 485 ****	
clh_pen chl_lyc	MVITSTSSVTCSTIFDIGNYKTKLLKIEPQTCTKHNSPPKALLIGTPSEAGNFPVLIFLH MVITSTSSVTCSTVFDIGNYKTKLLKIEPQTCSKHNSPPKSLLIGTPSEAGNFPVLIFLH ************	60 60
clh_pen chl_lyc	GYLLYNSFYSQLIQHLSSHGFIVVAPQLYLVEGADATEDIKSTAEVTNWLSEGLQHHLPP GYLLYNSFYSQLIQHLSSHGFIVVAPQLYLVEGADATEDIKSTAEVTHWLSEGLQHHLPP **********************************	120 120
clh_pen chl_lyc	DVEPNLKKLGLAGHSRGGKAAFSLALGRLATISTDLKFSALIGVDPVDGMEKGKQTPPSV DVEPNLTKLGLAGHSRGGKVAFSLALGRLASDLKFSALIGVDPVDGMEKGKQTPPSV ******.******************************	180 177
clh_pen chl_lyc	LTYVPRSFINLDMPVMVIGSGLGEVKKNPLFPACAPKGVNHRDFYNECCKPACYFVAKDY LTYVPRSFINLDMPVMVIGSGLGEVKKNPLFPACAPKGVNHRDFYNECCKPACYFVAKDY ************************************	240 237
clh_pen chl_lyc	GHNDMLDDETEGIRGKATYCLCKKGKSREPMRRFVGGVLVAFLEAYLEGNSSHLIAIRYG GHNDMLDEETKGIRGKATYCLCKKGKSREPMRRFVGGVLVAFLEAYLEGNSSHLIAIRDG ******:**:**:************************	300 297
clh_pen chl_lyc	HVALPVELQDIDFLV 315 HVALPVELQDIDFLV 312 *****	
lycb_pen lycb_lyc	METLLKPFPSLLLSSPTPHRSIFQQNPSFLSPTTQKKSRKCLLRNKSSKLFCSFLDLAPT METLLKPFPSLLLSSPTPYRSIVQQNPSFLSPTTKKKSRKCLLRNKSSKLFCSFLDLAPT ********************	60 60
lycb_pen lycb_lyc	SKPESLDVNISWVDPNSGRAQFDVIIIGAGPAGLRLAEQVSKYGIKVCCVDPSPLSMWPN SKPESLDVNISWVDPNSNRAQFDVIIIGAGPAGLRLAEQVSKYGIKVCCVDPSPLSMWPN ********************	120 120
lycb_pen lycb_lyc	NYGVWVDEFENLGLEDCLDHKWPMTCVHINDNKTKYLGRPYGRVSRKKLKLKLLNSCVEN NYGVWVDEFENLGLEDCLDHKWPMTCVHINDNKTKYLGRPYGRVSRKKLKLKLLNSCVEN ************************************	180 180
lycb_pen lycb_lyc	RVKFYKAKVWKVEHEEFESSIVCDDGKKIRGSLVVDASGFASDFIEYDKPRNHGYQIAHG RVKFYKAKVWKVEHEEFESSIVCDDGKKIRGSLVVDASGFASDFIEYDRPRNHGYQIAHG ************************************	240 240

lycb_pen lycb_lyc	SRPVLSYMEVKRRMVARLRHLGIKVRSVIEEEKCVIPMGGPLPRIPQNVMAIGGNSGIVH SRPVLSYMEVKRRMVARLRHLGIKVKSVIEEEKCVIPMGGPLPRIPQNVMAIGGNSGIVH ************************************	360 360
lycb_pen lycb_lyc	PSTGYMVARSMALAPVLAEAIVEGLGSTRMIRGSQLYHRVWNGLWPLDRRCVRECYSFGM PSTGYMVARSMALAPVLAEAIVEGLGSTRMIRGSQLYHRVWNGLWPLDRRCVRECYSFGM ************************************	420 420
lycb_pen lycb_lyc	ETLLKLDLKGTRRLFDAFFDLDPKYWQGFLSSRLSVKELGLLSLCLFGHGSNLTRLDIVT ETLLKLDLKGTRRLFDAFFDLDPKYWQGFLSSRLSVKELGLLSLCLFGHGSNMTRLDIVT ************************************	480 480
lycb_pen lycb_lyc	KCPVPLVRLIGNLAIESL 498 KCPLPLVRLIGNLAIESL 498 ***:******	
SEC14_PEN SEC14_LYC	MDSSSPITVQKALKSQESAPQSSPKPQSLKKSFVTSLMEATTLRTPSFKEDSYFTSHLKP MDSSSPITVQKALKSQESAPQSSPKPQSLKKSFVTSLMEATTLRTPSFKEDSYFTSHIKP ************************************	60 60
SEC14_PEN SEC14_LYC	SEKKALQELKDKLQSSSSSSSSMWGIPLLGGDEKADVILLKFLRARDFKVSDSLHMLEKCL SEKKALQELKDKLQSSSSSSSMWGIPLLGGDEKADVILLKFLRARDFKVSDSLHMLEKCV ************************************	120 120
SEC14_PEN SEC14_LYC	SWRKEFGADTILEEDFSGFKELEGVVAYMNGYDRDGHPVCYNAYGVFKDKEMYEKIFGDE SWRKEFGADTILEEDFSGFKELEGVVAYMNGYDRDGHPVCYNAYGVFKDKEMYEKIFGDE ************************************	180 180
SEC14_PEN SEC14_LYC	EKLKKFLRWRVQVLERGIEQLHFKPGGINSIIQVTDLKDMPKRELRVASNQILSLFQDNY EKLKKFLRWRVQVLERGIEQLHFKPGGINSIIQVTDLKDMPKRELRVASNKILSLFQDNY ************************************	240 240
SEC14_PEN SEC14_LYC	PEMVARKIFINVPWYFSVLYSMFSPFLTQRTKSKFVISKEGNVAETLYKFIRPEDIPVQY PEMVARKIFINVPWYFSVLYSMFSPFLTQRTKSKFVISKEGNVAETLYKFIRPEDIPVQY ************************************	300 300
SEC14_PEN SEC14_LYC	GGLSRPTDLQNGPPKPASEFTVKGGEKVNIQIEGIEGGATITWDIVVGGWDLEYSAEFVP GGLSRPTDLQNGPPKPASEFTVKGGEKVNIQIEGIEGGATITWDIVVGGWDLEYSAEFVP ************************************	360 360
SEC14_PEN SEC14_LYC	NGEGSYTIAVEKPRKIAASEEAIHNSFTSKEAGKMVLSVDNTASRKRKVAAYRYIVRKSA NGEGSYTIAVEKPRKIAANEEAIHNSFTSKEAGKMVLSVDNTASRKRKVAAYRYIVRKSA ************************************	420 420
SEC14_PEN SEC14_LYC	TISA 424 TISA 424 ****	

Anexo II

RESEARCH PAPER



Genetic dissection of vitamin E biosynthesis in tomato

Juliana Almeida¹, Leandro Quadrana², Ramón Asís³, Nathalia Setta¹, Fabiana de Godoy¹, Luisa Bermúdez¹, Santiago N. Otaiza³, Junia V. Corrêa da Silva¹, Alisdair R. Fernie⁴, Fernando Carrari^{2,5} and Magdalena Rossi^{1,5},*

¹ Departamento de Botânica-IB-USP, 277, 05508-900, São Paulo, SP, Brazil

² Instituto de Biotecnología, Instituto Nacional de Tecnología Agropecuaría (IB-INTA), and Consejo Nacional de Investigaciones

Científicas y Técnicas (CONICET), PO Box 25, B1712WAA Castelar, Argentina (partner group of the Max Planck Institute for Molecular Plant Physiology, Potsdam-Golm, Germany)

³ CIBICI, Facultad de Ciencias Químicas Universidad Nacional de Córdoba, CC 5000, Córdoba, Argentina

⁴ Max Planck Institute for Molecular Plant Physiology, Wissenschaftspark Golm, Am Mühlenberg 1, Potsdam-Golm, D-14476, Germany

¹⁵ ⁵ These authors contributed equally to this work

* To whom correspondence should be addressed. E-mail: fcarrari@cnia.inta.gov.ar; E-mail: mmrossi@usp.br

Received 25 October 2010; Revised 7 February 2011; Accepted 8 February 2011

Abstract

10

Vegetables are critical for human health as they are a source of multiple vitamins including vitamin E (VTE). In plants, the synthesis of VTE compounds, tocopherol and tocotrienol, derives from precursors of the shikimate and methylerythritol phosphate pathways. Quantitative trait loci (QTL) for α -tocopherol content in ripe fruit have 20 previously been determined in an Solanum pennellii tomato introgression line population. In this work, variations of tocopherol isoforms (α , β , γ , and δ) in ripe fruits of these lines were studied. In parallel all tomato genes structurally associated with VTE biosynthesis were identified and mapped. Previously identified VTE QTL on chromosomes 6 and 9 were confirmed whilst novel ones were identified on chromosomes 7 and 8. Integrated analysis at the metabolic, genetic and genomic levels allowed us to propose 16 candidate loci putatively affecting tocopherol 25 content in tomato. A comparative analysis revealed polymorphisms at nucleotide and amino acid levels between Solanum lycopersicum and S. pennellii candidate alleles. Moreover, evolutionary analyses showed the presence of codons evolving under both neutral and positive selection, which may explain the phenotypic differences between species. These data represent an important step in understanding the genetic determinants of VTE natural variation in tomato fruit and as such in the ability to improve the content of this important nutriceutical. 30

Key words: Fruit metabolism, Solanum pennellii, tocopherol, tomato, vitamin E.

Introduction

Vegetables are critical for human health as they are a source of multiple vitamins and other essential compounds. In particular, tomato fruits are an important dietary source of antioxidants for humans due both to the fact that have a high intrinsic content of these compounds and the elevated consumption of this crop by the western population. The main non-enzymatic antioxidants found

population. The main non-enzymatic antioxidants found in tomato fruits are ascorbic acid (VTC), lycopene and carotenoids, phenolics, and vitamin E (VTE) (Abushita *et al.*, 1997; Frusciante *et al.*, 2007). Recent studies have reinforced the hypothesis of beneficial effects of VTE on human health, mainly in the prevention of coronary heart disease, breast cancer, and protection against nicotineinduced oxidative stress in the brain (Das *et al.*, 2009; Ros 2009; Zhang *et al.*, 2009). Although its function in plants remains somewhat undefined, several reports link VTE to the protection of pigments, proteins, and polyunsaturated fatty acids of the photosynthetic apparatus against reactive

Abbreviations: IL, introgression line; LRT, likelihood ratio test; MEP, methylerythritol phosphate; QTL, quantitative trait loci; ROS, reactive oxygen species; SK, shikimate; VTC, vitamin C; VTE, vitamin E.

[©] The Author [2011]. Published by Oxford University Press [on behalf of the Society for Experimental Biology]. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

oxygen species (ROS) generated during photosynthesis (Semchuk *et al.*, 2009). It has additionally been proposed that VTE interacts with other antioxidant mechanisms in order to maintain cellular redox homeostasis (Foyer and Noctor, 2005).

The synthesis of VTE occurs in photosynthetic organisms and its major constituents are a group of amphipathic molecules containing a polar chromanol head group derived from homogentisate and a polyprenyl lipophilic side chain, products of the shikimate (SK) and methylerythritol phosphate (MEP) pathways, respectively. VTE compounds, collectively termed tocochromanols, can be classified into two groups on the basis of the degree of saturation of their hydrophilic tails. Tocopherols, which are the most abundant in plants, have saturated tails derived from phytyl 2P, whereas tocotrienols have an unsaturated tail derived from geranylgeranyl 2P. The VTE biosynthesis pathway proceeding from the reduction of hydroxyphenylpyruvate to homogentisate is considered the 'VTE core pathway' 70 and comprises seven enzymes: 4-hydroxyphenylpyruvate dioxygenase (HPPD, EC 1.13.11.27), homogentisic acid geranylgeranyl transferase (HGGT/HST, EC 2.5.1.-), homogentisate phytyl transferase (VTE2, EC 2.5.1.-), dimethyl-phytylquinol methyl transferase (VTE3, EC 2.1.1.-), 75 tocopherol cyclase (VTE1, EC 5.3.-.-), γ-tocopherol Cmethyl transferase (VTE4, EC 2.1.1.95), and phytol kinase (VTE5, EC 2.7.-.). There are four naturally occurring forms of tocopherols and tocotrienols (α , β , γ , and δ), 80 which differ in the position and number of methyl groups on the chromanol ring (Munné-Bosch and Alegre, 2002). Although all VTE isoforms are potent antioxidants in vitro, α -tocopherol is the most active in terms of vitamin activity, partly because it is retained in the human body in preference to other tocopherols and tocotrienols (Traber and Sies, 1996). In plants, tocochromanol has been found exclusively in plastids. Since it has not been proved thus far that any isoform can be transported within the plant, and the enzymes of the core pathway have been found in plastids (Sun et al., 2009), it is assumed that the biosynthesis also occurs in this compartment.

Although the VTE biosynthetic pathway was elucidated in 1979 (Soll and Schultz, 1979), the identification of the genes involved is much more recent. In the last decade, via the use of genetic and genomics-based methods, the genes 95 encoding the enzymes for most of the steps of the VTE core biosynthesis pathway have been identified and cloned. However, as yet, this is confined to the model organisms Arabidopsis thaliana and Synechocystis sp. PCC6803 (Li et al., 2008). Indeed, the characterization of VTE mutants 100 and transgenic lines has provided considerable insight into the regulatory network of tocochromanol biosynthesis (for review see Mène-Saffrané and DellaPenna, 2009; Falk and Munné-Bosch, 2010). These combined studies have additionally suggested roles for VTE compounds beyond their antioxidant function including their participation in diverse physiological processes including germination, photoassimilate partitioning, growth, leaf senescence, and plant responses to abiotic stress (Falk and Munné-Bosch, 2010). Moreover, several studies have demonstrated a close interaction between VTE and other metabolic pathways. Tomato fruits overexpressing phytoene synthase (PSY), a key enzyme in carotenoid biosynthesis, displayed increased levels of tocopherol (Fraser et al., 2007). Moreover, tocochromanol content is additionally affected when the post-chorismate pathway is manipulated. Arabidopsis transgenic plants overexpressing the bacterial bi-functional chorismate mutase (CM)/prephenate dehydratase (PDT), displayed significantly higher levels of phenylalanine, as well as γ -tocopherol and γ -tocotrienol, besides other secondary metabolites (Tzin et al., 2009). Plant tocochromanol biosynthesis is furthermore subjected to control by both environmental and endogenous signals. In agreement with this statement, the silencing of the light response factor DE-ETIOLATED1 resulted in tomato fruits with enhanced levels of antioxidants, including carotenoids, flavonoids, and tocopherol (Davuluri et al., 2005; Enfissi et al., 2010).

Cultivated tomato (Solanum lycopersicum) is the most consumed vegetable globally. The fact that its wild relatives display tremendous variation in metabolite content in both 130 leaves and fruits (Schauer et al., 2005), renders wild germplasm an important source for metabolic gene discovery focused on aiding efforts to improve the nutritional and industrial quality of crop species (Zamir, 2001; Fernie et al., 2006; Tohge and Fernie, 2010). Utilizing this approach, Schauer et al. (2006) reported a detailed metabolite profile of 76 tomato introgression lines (ILs) containing chromosome segments of the wild species Solanum pennellii in the genetic background of the cultivated S. lycopersicum (cv M82; Eshed and Zamir, 1995). Following the quantification 140 of 74 metabolites of known chemical structure, they were able to identify 889 quantitative fruit metabolic loci for variations in the content of amino and organic acids, sugars, alcohols, fatty acids, VTC, and VTE. Two of these quantitative trait loci (QTL), explaining variation in the 145 α -tocopherol fruit content, were located on chromosomes 6 and 9. Independent experiments available at the Tomato Functional Genomics Database (Fei et al., 2006; http:// ted.bti.cornell.edu/) have also revealed differences in tocopherol content associated with the exact same genomic regions. However, the mechanisms explaining these variations are currently not understood, partially due to a lack of knowledge of the complete VTE biosynthetic pathway in tomato.

The aim of the current report is to provide a framework 155 for associating gene sequence with fruit tocopherol content phenotypes by (i) characterizing and mapping all genes involved in the VTE biosynthesis pathway in tomato, (ii) identifying QTL for the content of the vitamers of VTE and their candidate genes, (iii) cloning and sequencing these genes from *S. pennellii*, and (iv) examining evolutionary patterns of candidates genes by comparing orthologues from *S. pennelli, S. lycopersicum*, and *A. thaliana*. The combined results of this study will be discussed in the context of the fundamental understanding of the accumulation of VTE opening further stages for functional analyses.

Materials and methods

Plant material

- 170 Tomato seeds from S. lycopersicum L. (cv M82) and S. pennellii introgressed lines were obtained from Tomato Genetic Resource Center (http://tgrc.ucdavis.edu). Tomato plants were grown in 20-1 pots under greenhouse conditions: 16/8 h photoperiod, 24 \pm 3 °C, 60% humidity, and 140 \pm 40 µmol m⁻² s⁻¹ incident photoirradiance. Cloning was carried out from fully expanded source leaves and mature fruits (60 d after flowering). For tocopherol quantification six ripe fruits were taken from six individual plants of ILs 6-1, 6-2, 7-4, 7-4-1, 7-5, 8-2, 8-2-1, 9-1, 9-2-6. Tissue was collected, immediately frozen into liquid nitrogen, and stored at -80 °C until use. 180

Survey of tocopherol biosynthesis enzymes, genome mapping, and expression analyses

The VTE pathway presented in Fig. 1 was outlined by combining data reported for enzymatic steps involved in SK, MEP, and VTE 185 biosynthesis available on KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/) and related scientific literature. Arabidopsis loci were obtained from the KEGG database and these sequences were used to perform TBLASTN

- (Altschul et al., 1990) searches against tomato expressed sequences 190 from the Lycopersicon Combined Build # 3 unigene database housed by the Solanaceae Genomics Network (http://solgenomic.net). The criteria used to determine orthology were $\geq 40\%$ identity at the amino acid level and $\geq 65\%$ coverage of the Arabidopsis protein. For uncompleted unigenes, the coverage cut-off was set at
- ≥30%. Based on BLASTN results, all unigene sequences were used 195 as queries to identify the corresponding genomic sequences from the S. lycopersicum genome assembly-version 2.31 (3,230 sequences; 781,381,961 total letters)—The International Tomato Genome Sequencing Project at the Solanaceae Genomics Network (http://
- 200 solgenomic.net). In silico prediction of the subcellular localization of tomato unigenes was performed using TargetP (http:// www.cbs.dtu.dk/services/TargetP) and ChloroP program (http:// www.cbs.dtu.dk/services/ChloroP) based on their deduced amino acid sequences.
- 205 The identified tomato genes were mapped onto the Tomato-EXPEN 2000 genetic map available at the Solanaceae Genomics Network. The genetic positions were obtained by BLASTN (Altschul et al., 1990) searches of the identified unigenes and/or their corresponding genomic sequences against the entire
- Tomato-EXPEN 2000 map marker sequences database (http:// solgenomics.net/index.pl). Map Chart software 2.2 (Voorrips, 2000) was used to construct the graphical representation of the genetic map.
- Expression data for the set of genes analysed here were extracted from a TOM1 microarray experiment previously published (Carrari et al., 2006). This experiment comprises transcript analyses from tomato fruits harvested along development and ripening stages (10, 15, 20, 21, 35, 49, 56, and 70 d after anthesis).

Tocopherol quantification by HPLC and QTL mapping

Tocopherol extraction was performed as described by Fraser 220 et al. (2000) with the following modifications: tomato fruit were ground to a fine powder in liquid nitrogen and 500 mg of material was extracted with 1.5 ml of methanol and, after vortexmixing, 1 ml of chloroform was added. Following 5 min of sonication, 1 ml of Tris buffer (50 mM Tris pH 7.5/1 M NaCl) 225 was added. The chloroform phase was recovered and the methanol phase (remaining pellet) was re-extracted with chloroform (2 ml). Chloroform extracts were pooled and adjusted to a final volume of 4 ml. Two millilitres were dried under nitrogen

gas and re-suspended in 0.2 ml of 99.5:0.5 hexane/isopropanol.

The tocopherol content was determined using a Hewlett-Packard

Vitamin E biosynthesis in tomato | 3 of 18

series 1100 HPLC system coupled with a fluorescence detector (Agilent Technologies series 1200). Separation was carried out on a normal-phase column Metasil Si (250 mm×4.6 mm, 5 µm, Varian; Metachem, Torrance, CA, USA) maintained at room temperature using an isocratic solvent system (mobile phase) consisting of 99.5:0.5 hexane/isopropanol with a flow rate of 1 ml \min^{-1} . Eluting compounds were detected and quantified by fluorescence with excitation at 296 nm and emission at 340 nm. Identification and quantification of tocopherol compounds was 240 achieved by comparison with the retention times and peak areas of standards purchased from Merck (tocopherol set; Calbiochem #613424). A daily calibration curve was carried out using a tocopherol solution with a concentration range between 0.31 μ g ml⁻¹ and 5 μ g ml⁻¹ for each isoform. Data of tocopherol 245 isoforms or total tocopherol content were statistically analysed according to Sokal and Rohlf (1981). When the data pull presented homoscedasticity, with or without data transformation using ln or square root, an ANOVA followed by a Dunnett test (P < 0.05) was used to compare tocopherol content between ILs 2.50 and M82 control. Due to lack of homoscedasticity, a nonparametric comparison was also performed by Kruskal-Wallis test (P<0.05 and P<0.1). Statistical tests were performed with Bioestat 5.0 (Ayres et al., 2007) and InfoStat v. 2009 (Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina).

The position of tocopherol QTL on the Tomato-EXPEN 2000 map was determined according to the flanking markers of the S. pennellii introgression fragments in the analysed ILs (Eshed and Zamir, 1995) and mapping of unigenes performed as described in Kamenetzky et al. (2010).

Identification of VTE-related pathway candidate genes

Candidate genes were surveyed along the genomic regions spanned by the identified VTE QTL as described in Bermúdez et al. (2008). All molecular markers mapped onto the selected genomic regions were identified in the comparison merging the Tomato-EXPEN 265 2000, the Tomato-EXPEN 1992, and the Tomato IL maps by using the comparative map web interface of the Solanaceae Genomics Network (Mueller et al., 2008). All marker sequences were used as query to identify the corresponding unigenes in the Solanaceae Genomics Network database. Gene product functions were determined according to homology to a previously characterized protein, whose function had been experimentally demonstrated in other related plant species, by using the BLASTX algorithm (Altschul et al., 1990) against the NCBI non-redundant (nr) protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The cut-off criteria were: $\geq 40\%$ identity at amino acid level and $\geq 65\%$ coverage of the orthologous Arabidopsis protein. For uncompleted unigenes the coverage cut-off was set at $\geq 30\%$.

Amplification, cloning, and sequencing

Total RNA from 100 mg of source leaf or fruit tissue was isolated 280 using Trizol reagent (Invitrogen, #15596-026) and 1 µg from each sample was treated with amplification grade DNase I (Invitrogen, #18068-015) to remove potential contamination of genomic DNA. The treated RNA was reverse-transcribed to first-strand cDNA and primed with oligo(dT) using Superscript First-Strand (Invi-285 trogen, #18080-044) following the manufacturer's protocol.

Primers were designed using the software Oligo Analyzer 3.1 (http://www.idtdna.com) on the basis of unigene sequences including the sequence surrounding the putative ATG start and stop codons. The primer sequences are available in Table S1 in Supplementary data available at JXB online.

Full-length cDNA fragments were generated by PCR using Taq Platinum pfx DNA polymerase (Invitrogen, #11708-013). The PCR reactions were conducted in a total volume of 50 µl containing 0.3 mM each dNTPs, 0.4 μ M each primer, 1.5× reaction buffer, 1 mM MgSO₄, ~150 ng of cDNA, and 2.5 IU of enzyme. The amplification conditions were: 94 °C for 3 min; 35

200

ъ.	
ð	
Ē.	
Ľ.	
8	
Ĕ	
Б	
d)	
Ĩ	
0	
ŝ	
Š	
\geq	
g	
E	
2	
g	
5	
-8	
5	
0	
2	
Ψ	
Ъ	
Ę	
σ	
ő	
5	
Q	
0	
З	
ž	
ä	
3	
ç	
t_{l}	
Ą.	
Ś	
é	
5	
Ř	
0,	
Ψ	
<u>a</u>	
<u>.0</u>	
p	
E	
ö	
~	
8	
Ť,	
0	
<u>ö</u>	
soci	
ssoci	
associ	
- associ	
-L- associ	
2TL- associ	
' QTL- associ	
nd QTL- associ	
and QTL- associ	
and QTL- associ	
is and QTL- associ	
ssis and QTL- associ	
resis and QTL- associ	
ithesis and QTL- associ	
vnthesis and QTL- associ	
synthesis and QTL- associ	
iosynthesis and QTL- associ	
biosynthesis and QTL- associ	
E biosynthesis and QTL- associ	
TE biosynthesis and QTL- associ	
VTE biosynthesis and QTL- associ	
n VTE biosynthesis and QTL- associ	
I in VTE biosynthesis and QTL- associ	
d in VTE biosynthesis and QTL- associ	
/ed in VTE biosynthesis and QTL- associ	
Jived in VTE biosynthesis and QTL- associ	
volved in VTE biosynthesis and QTL- associ	
involved in VTE biosynthesis and QTL- associ	
s involved in VTE biosynthesis and QTL- associ	
es involved in VTE biosynthesis and QTL- associ	
nes involved in VTE biosynthesis and QTL- associ	
jenes involved in VTE biosynthesis and QTL- associ	
genes involved in VTE biosynthesis and QTL- associ	
to genes involved in VTE biosynthesis and QTL- associ	
ato genes involved in VTE biosynthesis and QTL- associ	
mato genes involved in VTE biosynthesis and QTL- associ	
omato genes involved in VTE biosynthesis and QTL- associ	
tomato genes involved in VTE biosynthesis and QTL- associ	
of tomato genes involved in VTE biosynthesis and QTL- associ	
n of tomato genes involved in VTE biosynthesis and QTL- associ	
on of tomato genes involved in VTE biosynthesis and QTL- associ	
tion of tomato genes involved in VTE biosynthesis and QTL- associ	
ation of tomato genes involved in VTE biosynthesis and QTL- associ	
ication of tomato genes involved in VTE biosynthesis and QTL- associ	
tification of tomato genes involved in VTE biosynthesis and QTL- associ	
intification of tomato genes involved in VTE biosynthesis and QTL- associ	
lentification of tomato genes involved in VTE biosynthesis and QTL- associ	
Identification of tomato genes involved in VTE biosynthesis and QTL- associ	
. Identification of tomato genes involved in VTE biosynthesis and QTL- associ	
1. Identification of tomato genes involved in VTE biosynthesis and QTL- associ	
e 1. Identification of tomato genes involved in VTE biosynthesis and QTL- associ	
ye 1. Identification of tomato genes involved in VTE biosynthesis and QTL- associ	
able 1. Identification of tomato genes involved in VTE biosynthesis and QTL- associ	
Table 1. Identification of tomato genes involved in VTE biosynthesis and QTL- associ	

Enzyme ^a	A. <i>thaliana</i> locus (no. amino acids)	Curated localization ^b	Tomato unigene ^c	Signal peptide ^d	Genomic id ^e	Linked marker ^f	Chromosome position ^g (cM)
1-Deoxy-D-xylulose-5-P synthase (DXS, EC 2.2.1.7)	At4g15560 (717)	Chloroplast	U567647 (1) U316204 (2)	Chloroplast nd	SL2.31sc05941 SL2.31sc03748	T1704 TG523	1 (39 cM) 11 (29 cM)
2-C-methyl-D-erythritol 4-phosphate synthase (DXR, EC 1.1.1.267)	At5g62790 (477)	Chloroplast	U585813	Chloroplast	SL2.31sc03701	C2_At5g23060	3 (102.5 cM)
2-C-methyl-D-erythritol 4-phosphate cytidylyttransferase (CMS, EC 2.7.7.60)	At2g02500 (302)	Chloroplast	U566797	Chloroplast	SL2.31sc04323	TG528	1 (127.5 cM)
4-(Cyrtidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (ISPE, EC 2.7.1.148)	At2g26930 (383)	Chloroplast	U583224	Chloroplast	SL2.31sc04133	cTOC-4-C7	1 (19 cM)
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ISPF, EC 4.6.1.12)	At1g63970 (231)	Chloroplast	U568497	Chloroplast	SL2.31sc03923	C2_At3g27530	8 (78 cM)
4-Hydroxy-3-methylbut-2-enyl- diphosphate synthase (HDS, EC 1.17.7.1)	At5g60600 (717)	Chloroplast	U567167	Chloroplast	SL2.31sc03876	C2_At5g60600	11 (79 cM)
4-Hydroxy-3-methylbut-2-enyl- diphosphate reductase (HDR, EC 1.17.1.2)	At4g34350 (466)	Chloroplast	U580658	Chloroplast	SL2.31sc04323	C2_At4g34350	1 (154 cM)
lsopentenyl diphosphate ð-isomerase (IPI, EC 5.3.3.2)	At3g02780 (284) At5g16440 (291)	Mitochondria Chloroplast (Phillips <i>et al.</i> , 2008)	U577516 (1) U569721 (2)	Chloroplast Chloroplast	SL2.31sc06101 SL2.31sc03902	U49812 C2_At5g04270	4 (64 cM) 5 (112.7 cM)
Geranyl pyrophosphate synthase (GPPS, EC 2.5.1.1)	At2g34630 (422)	Chloroplast (Bouvier <i>et al.</i> ,2000)	U573523	Mitochondria	SL2.31sc03835	C2_At1g30360	8 (21.3 cM)
Geranylgeranyl pyrophosphate synthase (GGPS, EC 2.5.1.29)	<u>At4g36810</u> (371) <u>At4g38460</u> (326)	Chloroplast Chloroplast	U574849 (1) U571085 (2) U573348 (3) U575882 (4)	Chloroplast Chloroplast nd Chloroplast	SL2.31sc03748 SL2.31sc04135 SL2.31sc03665 SL2.31sc03665	C2_At5g16710 C2_At1g19340 CT232 T0532	11 (31.4 cM) 4 (112 cM) 2 (90.1 cM) 9 (30 cM)
Geranylgeranyl reductase (GGDR, EC 1.3.1-)	At1g74470 (467)	Chloroplast	U564571	Chloroplast	SL2.31sc03701	C2_At1G74470	3 (122 cM)
3-Deoxy-D-arabino-heptulosonate-7-P synthase (DAHPS, EC 2.5.1.54)	At1g22410 (527) At4g33510 (432) At4g39980 (525)	Chloroplast Chloroplast Chloroplast (Entus <i>et al.</i> ,2002)	U581552 (1) U566921 (2)	Chloroplast nd	SL2.31sc03748 SL2.31sc04135	T0408 T1560	11 (26 cM) 4 (83.5 cM)
3-Dehydroquinate synthase (DHQS, EC 4.2.3.4)	At5g66120 (442)	Chloroplast	U568781	Chloroplast	SL2.31sc03665	C2_At3g01160	2 (83.4 cM)
Shikimate dehydrogenase (SDH, EC 1.1.1.25) / 3-Dehydroquinate dehydratase (DHQ, EC 4.2.1.10)	At3g06350 (603)	Chloroplast	U570855(1) U570070(2)	nd	SL2.31sc05941 SL2.31sc03622	Т1704 ТG221	1 (39 cM) 6 (101 cM)
Shikimate kinase (SK, EC 2.7.1.71)	At2g21940 (276) At4g39540 (300)	nd Chloroplast	U582040	Chloroplast	SL2.31sc06101	C2_At3g62940	4 (56 cM)

4 of 18 | Almeida et al.

Enzyme ^a	A. thaliana locus Ino. amino acids)	Curated localization ^b	Tomato unigene ^c	Signal peptide ^d	Genomic id ^e	Linked marker ^f	Chromosome position ^g (cM)
5-Enolpyruvylshikimate-3-P synthase (EPSPS, EC 2.5.1.19)	At1g48860 (521) At2g45300 (520)	Chloroplast Chloroplast	U577580	Chloroplast	SL2.31sc04323	C2_At2g45240	1 (57 cM)
Chorismate synthase (CS, EC 4.2.3.5)	At1g48850 (380)	Chloroplast	U563165(1) U563163(2)	Chloroplast Chloroplast	SL2.31sc06101 SL2.31sc03604	C2_At3g07950 TG370	4 (56.7 cM) 4 (21.5 cM)
Chorismate mutase (CM, EC 5.4.99.5)	At1g69370 (316) At5g10870 (340) At3g29200 (265)	Chloroplast (Mobley <i>et al.</i> , 1999) Chloroplast (Eberhard <i>et al.</i> , 1996) Cytosol (Eberhard <i>et al.</i> , 1996)	U575627(1) U585231(2)	פ פ	SL2.31sc03665 SL2.31sc03748	T1480 TG147	2 (106 cM) 11 (45 cM)
Prephenate aminotransferase (PAT, EC 2.6.1.57)	At2g22250 (475)	Chloroplast	U567172	Chloroplast	SL2.31sc06101	C2_At4g39830	4 (61 cM)
Arogenate dehydrogenase (TyrA, EC 1.3.1.78)	<u>At1g15710 (</u> 358) At5g34930 (640)	Chloroplast Chloroplast (Rippert <i>et al.</i> , 2009)	U567861 (1) U570951 (2)	Chloroplast Chloroplast	SL2.31sc03731 SL2.31sc03771	C2_At5g34850 T1212	7 (0.4 cM) 9 (48 cM)
Tyrosine aminotransferase (TAT, EC 2.6.1.5)	<u>At5g53970</u> (414)	pu	U577103(1) U563404 (2)	pu	SL2.31sc05925 SL2.31sc03685	C2_At1g53000 C2_At1g03820	10 (7.5 cM) 7 (43 cM)
4-Hydroxyphenylpyruvate dioxygenase (HPPD, EC 1.13.11.27)	<u>At1g06570</u> (473)	Cytosol (Garcia <i>et al.</i> , 1999)	U580457(1) U578997(2)	nd Chloroplast	SL2.31sc03685 SL2.31sc03902	TG584 CLET-6-14	7 (36.5 cM) 5 (70 cM)
Homogentisate geranylgeranyl transferase/ homogentisate solanesyl transferase (HGGT/HST, EC 2.5.1)	At3g11945.2 (393)	Chloroplast	U585005	Chloroplast	SL2.31sc06725	C2_At3g58490	3 (72.6cM)
Homogentisate phytyl transferase [HPT (VTE2), EC 2.5.1]	<u>At2g18950(393)</u>	Chloroplast	U327540(5') U576207(3')	Chloroplast	SL2.31sc03731	cTOA-13-K15	7 (17 cM)
Dimethyl-phytyquinol methyl transferase [MPBQMT (VTE3), EC 2.1.1]	<u>At3g63410</u> (338)	Chloroplast	U578249(1) U581492(2)	Chloroplast Chloroplast	SL2.31sc04777 SL2.31sc04439	T0565 TG324	9 (52 cM) 3 (4.6 cM)
Tocopherol cyclase [TC (VTE1), EC 5.3]	At4g32770(488)	Chloroplast	U570602	Chloroplast	SL2.31sc04948	C2_At4g32770	8 (34 cM)
γ-Tocopherol C-methyl transferase [γ-ΤΜΤ (VTE4), EC 2.1.1.95]	At1g64970 (348)	Chloroplast (Ferro et al., 2010)	U584511	Chloroplast	SL2.31sc03923	TG282	8 (41.8 cM)
Phytol kinase [PK (VTE5), EC 2.7]	At5g04490(304)	Chloroplast	U583081	Chloroplast	SL2.31sc03771	C2_At5g58240	9 (50.5 cM)
Anthranilate phosphoribosyltransferase (APT, EC 2.4.2.18)	At5g17990(444)	Chloroplast	U566340	Chloroplast	SL2.31sc05054	C2_At5g17990	6 (59 cM)
Phosphoribosylanthranilate isomerase (PRAI, EC 5.3.1.24)	At1g07780 (275) At1g29410 (244) At5g05590 (275)	Chloroplast Chloroplast (Zhao and Last, 1995) Chloroplast	U564371	Chloroplast	SL2.31sc05732	cLET-1-113	6 (24 cM)

Vitamin E biosynthesis in tomato | 5 of 18

Table 1. Continued

Continued	
÷	
ble	
Tal	

Enzyme ^a	<i>A. thaliana</i> locus (no. amino acids)	Curated localization ^b	Tomato unigene ^c	Signal peptide ^d	Genomic id ^e	Linked marker ^f	Chromosome position ^g (cM)
Folylpolyglutamate synthase (FPGS, EC 6.3.2.17)	<u>At5g41480</u> (530)	Mitochondria (Ravanel <i>et al.</i> , 2001)	U581922 ^h	Mitochondria	SL2.31sc05732	At5g41480	6 (26 cM)
Phospholipid transporter SEC14	At3g51670(409)	Membrane (Peterman et al., 2004)	U583419	Chloroplast	SL2.31sc03771	T1095	9 (50.5 cM)
Chlorophyllase (CHL, EC 3.1.1.14)	<u>At59</u> 43860 (318) At1919670 (324)	Cytosol (Schenk <i>et al.</i> , 2007)	U574853	nd	SL2.31sc05732	T0834	6 (32 cM)
Lycopeno β-cyclase LYCB	At3g10230(369)	Chloroplast(Ferro <i>et al.</i> , 2010)	U570109	Chloroplast	SL2.31sc05054	cLET-19-J2	6 (74 cM)
 ^a Enzyme name, number according to KE ^b A. thaliana enzyme localization according ^c Tornau unigene number according to SI ^d Sub-cellular localization prediction accor ^e S. <i>lycopersicum</i> scaffold spanning the α ^f Closest markers within scaffold. ^g Chromosome and genetic position in Toi ^h 5' incomplete unigene Nd, not determined 	GG (www.genome.jp 10 The Plant Prote GN (solgenomics.net ding to TargetP 1.1.1 orresponding gene (/ mato-EXPEN 2000 v	/kegg/) and abbreviation. ome Database (ppdb.tc.comell.ed t) software (www.cbs.dtu.dk/service version 2.31).	lu, Sun <i>et al., 2</i> 009) or s ss∕TargetP/) (Emanuelss	pecific reference. on <i>et al.</i> , 2007) anc	ł ChloroP (www.	obs.dtu.dk/serv	ces/ChloroP).

cycles of 94 °C for 15 s; primer-specific annealing temperature for 30 s; and then 68 °C for 2 min. Amplification products were purified with GFX purification Kit (Amersham Biosciences, 300 #289034-70) and cloned into a pCR-Blunt II TOPO vector using a TOPO-Zero Blunt cloning kit (Invitrogen, #45-0245). Plasmid DNA was isolated using a Qiagen Miniprep Kit (#27106) and inserts were sequenced with BigDye Terminator (Applied Biosystems, #4336919) on an ABI3700 automated sequencer (Applied 305 Biosystems). Sequence data from this article have been deposited in the GenBank Data Libraries under accession number HQ014366-HQ014383 and HQ219713-HQ219716. Polymorphisms were detected at nucleotide and amino acid levels by aligning S. pennellii and S. lycopersicum sequenced alleles (excluding primer regions) using the MULTALIN program (http://wwwarchbac.u-psud.fr/genomics/multalin.html; Corpet, 1988).

Evolutionary analyses

The coding region alignments of A. thaliana, S. lycopersicum, and S. pennellii were performed with BioEdit Sequence Alignment Editor (Hall, 1999) using the ClustalW package (v1.81 Thompson et al., 1994) and were manually curated according to amino acid alignment. Non-synonymous (d_N) and synonymous (d_S) distances and their SE values were estimated with MEGA 4.1 software (Tamura et al., 2007) using the Nei-Gojobori method (p-distance). 320 In order to preserve the reading frames, the alignment gaps were deleted prior to estimation of $d_{\rm S}$ and $d_{\rm N}$. Codon bias was determined by the effective number of codons (N_c) value computed in the CodonW program (mobyle.pasteur.fr/cgi-bin/portal.py?form=codonw). N_c varies between 21 for maximum codon bias, when only one codon is used per amino acid, and 61 for minimum codon bias, when synonymous codons for each amino acid are used at similar frequencies. One-way ANOVA with Tukey's posthoc test in the InfoStat software was performed to evaluate significant differences in codon usage. 330

In order to compare codon evolution models to determine selective constraint, three models were fitted using the CODEML program of the PAML suite (Yang, 2007). The first model, M0, assumes that all codons across the sequences have the same level of $d_{\rm N}$ and $d_{\rm S}$ and estimates these values and the $d_{\rm N}/d_{\rm S}$ ratio (ω). ω is a signal of the selection at protein level thus, $0 \le \omega \le 1$ indicates purifying selection, $\omega=1$ neutral selection, and $\omega>1$ points to the presence of positive selection. The model M1a proposes the existence of two classes of codon, a proportion with $0 < \omega < 1$ and the remainder of codons with $\omega{=}1.$ Finally, model M2a divides 340 codons into three classes: those with $0 \le \omega \le 1$, $\omega = 1$, and $\omega \ge 1$. The fit of model M0 versus M1a or M1a versus M2a is evaluated by a likelihood ratio test comparing twice the difference in log likelihoods with a χ^2 distribution (Yang, 2007).

Results

Identification, mapping, and expression analyses of tomato genes involved in VTE biosynthesis

VTE compounds, tocopherols, and tocotrienols, are products of the convergence of the plastidial MEP and SK pathways. With the aim of identifying every metabolic step in tocochro-350 manol biosynthesis in tomato, the two routes from their primary metabolism precursors were linked to the VTE core pathway (Fig. 1) based on data available in the KEGG database. After an in-depth search of the unigene database deposited in the Solanaceae Genomics Network (http://solgenomics.net), using Arabidopsis loci as reference sequences, all the putative tomato enzyme encoding genes were identified (Table 1). Twenty-nine biochemical reactions constitute the



Fig. 1. VTE biosysthesis pathway. The MEP, SK, and tocopherol core pathways are highlighted in red, green, and blue, respectively. Candidate genes of the VTE-related pathways (carotenoid, chlorophyll, tryptophan and folate metabolism, and the SEC14 protein) are not highlighted. Enzymes are named according to their abbreviations in Table 1. Genes for which wild alleles were cloned, sequenced, and analysed are underlined.

360

365

VTE biosynthesis pathway and these are catalyzed by 28 enzymes, for which a total of 41 different tomato encoding loci were described. For eleven of the enzymes, the number of surveyed tomato genes differs from those described for *Arabidopsis*, according to the criteria adopted here. Only one point in the pathway shown in Fig. 1 remains obscure in plant metabolism, which is the phosphorylation of phytyl P to provide phytyl 2P from phytol via VTE5, as an alternative to the MEP pathway (Valentin *et al.*, 2006).

The subcellular localization of the protein products of previously identified *S. lycopersicum* loci were predicted by the TargetP and ChloroP softwares. Enzymes that did not contain a predicted targeting peptide were considered cytosolic. This *in silico* prediction was in general agreement with the experimental evidence reported for the *Arabidopsis* orthologues (Table 1). For two proteins, however, predictions revealed unexpected results. The tomato geranyl pyrophosphate synthase (GPPS) enzyme was predicted to be targeted to the mitochondria, while for the bi-functional shikimate dehydrogenase/3-dehydroquinate dehydratase (SDH/DHQ) no signal peptides were detected for any of the identified loci. Moreover, none of the tomato CMs presented a predicted plastid signal peptide. The fact that the last enzyme of the post-chorismate portion of the SK pathway, tyrosine aminotransferase (TAT), and HPPD appeared to be localized in the cytosol in *Arabidopsis* cells provides intrigue regarding the transport of homogentisate across the chloroplast envelope (Joyard *et al.*, 2009). These predictions, with regard to the tomato proteins, also failed to propose a subcellular localization for TAT, even though one of the HPPD unigenes exhibited a chloroplast signal peptide prediction.

As a second step in the characterization of the genetic basis of tocochromanol biosynthesis, the 41 tomato loci involved in MEP, SK, and VTE core pathways were localized onto the tomato genetic map (Tomato-EXPEN 2000) (Figs 1, 2, Table 1). Mapping was based on physical linkage between mapped markers and the unigenes and/or

385

380



Fig. 2. Genomic localization of tocopherol biosynthesis and candidate genes. All genes were localized in the Tomato-EXPEN 2000 genetic map available at the Solanaceae Genomics Network (http://solgenomics.net/index.pl). Markers and genes are indicated on the left side of the chromosomes. Tocopherol QTL are indicated on the right side of the chromosomes. Gene colour code is in accordance with Fig. 1.

their corresponding genomic sequences. With the exception of chromosome 12, all other tomato chromosomes harbour at least one of the identified loci. Interestingly, most of the tocochromanol core pathway enzyme-400 encoding loci were localized on chromosomes 7, 8, and 9, except for *hst* and vte3(2), mapping to chromosome 3, and hppd(2), mapping to chromosome 5. Remarkably, the four genes mapped on chromosome 9, vte3(1), vte5, 405 arogenate dehydrogenase [tyra(2)], and geranylgeranyl pyrophosphate synthase [ggps(4)] all co-localize with a previously reported QTL for α -tocopherol content described by Schauer et al. (2006).

420

Of the 41 identified genes, the expression patterns of 27 of them could be evaluated across fruit development and 410 ripening, by retrieving data from a previously published microarray experiment (Carrari et al., 2006). This analysis revealed that all 27 genes were expressed in tomato fruits in at least one time point. The *omeSOM model of neural clustering recently developed (Milone et al., 2010) revealed 415 six groups. Whilst these clusters grouped genes from all three evaluated pathways, no specific pathway patterns were identified (data not shown).

QTL for fruit tocopherol content and identification of candidate genes

As mentioned above, these mapping results revealed that the majority of the VTE core pathway enzyme-encoding genes are grouped within chromosomes 7, 8, and 9. Previously, using GC-MS analysis, Schauer et al. (2006) reported two QTL for fruit a-tocopherol content localized to chromosomes 6 and 9. In order to investigate further the presence of other QTL associated with the genomic regions that harbour the tocopherol core biosynthesis genes, and as such to obtain a precise and detailed quantification, an HPLC protocol for measuring all four tocopherol isoforms was applied.

425

430

The content of α -, β -, γ -, δ -, and total tocopherol was determined from ripe tomato fruits from the ILs 6-1, 6-2, 7-4, 7-4-1, 7-5, 8-2, 8-2-1, 9-1, 9-2-6 as well as fruits from the corresponding S. lycopersicum control (cv M82). The amount of each isoform and their ratios are presented in 435 Fig. 3, whereas the identified QTL are positioned on the genetic map presented in Fig. 2. On chromosome 9, two QTL were identified for α - and total tocopherol (ILs 9-1 and 9-2-6), one for β -tocopherol (IL 9-1), and one for γ -tocopherol (IL 9-1). Both QTL for total tocopherol are in 440 agreement with the results reported in Schauer et al. (2006). The QTL on IL 9-2-6 co-localize with two VTE core pathway encoding genes: vte3(1) and vte5 (Fig. 2), whilst the QTL on IL 9-1 spans the genomic region containing tyra(2) and ggps(4). Measurements performed in fruits 445 from ILs with S. pennellii introgressions on chromosome 6 showed significant differences from M82 fruits in the levels of β -tocopherol (IL 6-2), δ -, and total tocopherol (IL 6-1). Moreover, QTL were also identified on chromosome 7 (ILs 7-4 and 7-4-1 for α - and β -tocopherol, respectively) and 8 450 (IL 8-2-1 for total tocopherol) also co-localizing with the genes encoding the VTE core pathway enzymes: vte2 and

hppd(1) on chromosome 7 and vte1 and vte4 on chromosome 8 (Fig. 2). Besides the mentioned genes, the presence of tyra(1) at 0.4 cM and tat(2) at 43 cM on chromosome 7 could also be responsible for the α - and β -tocopherol QTL mapped onto these regions.

By surveying the genomic regions spanning the identified OTL, six novel candidate genes belonging to VTE-related pathways were found. On chromosome 6, at 32 cM, 460 a chlorophyllase encoding gene (CHL, EC 3.1.1.14) was identified. It has been proposed that the first step in the degradation of chlorophyll during senescence and fruit ripening is catalysed by this enzyme (Hörtensteiner, 2006) through the synthesis of phytol, providing phytyl 465 2P to the VTE pathway via VTE5. Moreover, three enzyme-encoding genes for the two branching pathways from chorismate to tryptophan and folate were also identified: a phosphoribosylanthranilate isomerase (PRAI, EC 5.3.1.24), a folylpolyglutamate synthase (FPGS, EC 470 6.3.2.17), and an anthranilate phosphoribosyltransferase (APT, EC 2.4.2.18) mapping at 24, 26, and 59 cM, respectively. Finally, the lycopene β -cyclase (LYCB) encoding gene was found at 74 cM of chromosome 6. This enzyme, which catalyses the last step in β-carotene bio-475 synthesis, has been deeply characterized after a positional cloning strategy by Ronen et al. (2000) and constitutes a strong candidate for tocopherol content due to the common precursor geranylgeranyl 2P. Another putative candidate, the SEC14 protein-encoding gene, was found on 480

- chromosome 9. Several studies have demonstrated the involvement of this protein in tocopherol transport in mammalian cells and lipid traffic in plants (Saito et al., 2007; Bankaitis et al., 2009). With the exception of the chl
- these novel candidates are expressed in tomato fruits at least 485 one time point of the developmental analysis performed by Carrari et al. (2006).

Taken together the results obtained from tomato gene identification, mapping, tocopherol quantification, and QTL localization, 16 candidate loci putatively affecting 490 tocopherol content in tomato can be proposed: *prai*, *fpgs*, *chl*, *apt*, and *lycb* on chromosome 6; tyra(1), vte2, hppd(1), and tat(2) located on chromosome 7; *vte1* and *vte4* on chromosome 8; and ggps(4), tyra(2), vte5, sec14, and vte3(1) on chromosome 9 (Figs 1–3).

495

500

505

455

Allele characterization of QTL-associated candidates

The identification of 16 QTL-associated candidate genes prompted us to unearth the allelic differences between S. lycopersicum and S. pennellii (underlined genes in Fig. 1). The coding regions of wild alleles were thus subsequently cloned from the corresponding ILs using primers annealing to the initial and stop codons. Although only minor size differences were observed between the two alleles for most of the genes, TyrA(2)-, VTE4-, PRAI-, and CHL-encoding genes exhibited different amplicon length. These results indicated that neither vast allelic polymorphisms, nor large genomic rearrangements, span the chromosomal region encompassing the analysed genes. This comparison revealed



Fig. 3. Tocopherol content. Tocopherol content was determined by HPLC. Grey bars indicate means of six biological replicates. Significant differences compared with the M82 control cultivar (black bars) according to Dunnett test (P<0.05) and/or Kruskal-Wallis test (P<0.05 ** and P<0.1*) are indicated.

that all analysed candidate genes present at least one nonsynonymous polymorphism (Table 2). The most divergent 510 alleles are those encoding the PRAI enzyme for which the S. pennelli allele encodes a protein 26 amino acids shorter in comparison with the S. lycopersicum allele.

VTE biosynthesis genes: evolutionary analyses of cultivated and wild tomato alleles

The fate of cellular metabolic networks generally depends on the products of many loci. The inter-relationships between loci at the phenotypic level raise the question of whether they evolved independently. In this work QTL for tocopherol content using S. pennelli ILs were mapped, and 520 16 candidate loci linked to those QTL were identified and the wild species alleles cloned. Furthermore, in order to study how the structure of the VTE metabolic pathway could have influenced protein evolution rates, the evolutionary pattern among the candidate genes and their 525

515

10 of 18 | Almeida et al.

Table 2. Comparison of S. lycopersicum and S. pennellii alleles of cDNA encoding VTE candidate genes. LYC, S. lycopersicum; PEN, S. pennellii.

Gene	Unigene	Coding sequence (nt)		Predicted protein (no. amino acids)		No. of polymorphic nucleotides ^a	No. of polymorphic amino acids ^a
		LYC	PEN	LYC	PEN		
ggps(4)	U575882	1005	1005	334	334	11	3
tat(2)	U563404	1269	1269	422	422	10	6
tyra(1)	U567861	1134	1134	377	377	7	1
tyra(2)	U570951	1173	1182	390	393	24+9insertions	9+3insertions
vte4	U584511	1089	1086	362	361	9+3deletions	2+1deletion
vte1	U570602	1497	1497	498	498	12	5
vte3(1)	U578249	1020	1020	339	339	7	1
vte2	U327540 (covers 5' end)	1209	1209	402	402	8	3
vte5	U583081	882	882	293	293	10	4
hppd(1)	U580457	1263	1263 ^b	420	420 ^b	16	7
apt	U566340 ^c	1097 ^c	1097 ^c	365	365	8	4
prai	U564371	906	828	301	275	12+78deletions	3+26deletions
fpgs	U581922 °	1457 ^c	1457 ^c	485	485	13	4
chl	U574853	939	948	312	315	22+9insertions	9+3insertions
sec14	U583419	1275	1275	424	424	11	4
lycb	U570109	1497	1497	498	498	19	9

^a Nucleotide or amino acid insertions or deletions in S. pennellii sequences.

^b S. pennellii does not present a stop codon along the analysed region.

^c Probably lacking 3' end.s

paralogues was also investigated, estimating the pairwise synonymous (d_S) , non-synonymous (d_N) , and d_N/d_S divergence between S. lycopersicum and S. pennellii (Fig. 4). The $d_{\rm S}$ and $d_{\rm N}$ values varied greatly between genes, ranging from 4.8 to 16 times for $d_{\rm S}$ and $d_{\rm N}$, respectively. Four 530 genes, ggps(2), tyra(2), chl, and lycb, displayed particularly high values of d_N , above the mean. The d_N/d_S also varied remarkably among the 22 loci analysed with the highest value being 17.4 times the lowest. The genes of 535 MEP, post-chorismate SK, and tocopherol biosynthesis core pathways displayed values similar to or lower than the average with the exception of those presenting more than one locus. Among candidate genes of related pathways apt, chl, and lycb displayed higher d_N/d_S values than the mean. Interestingly, the five genes for which more than 540 one locus was identified displayed variations of d_N/d_S values between paralogues. The d_N/d_S values for the different ggps, tat, tyra, vte3 and hppd varied by 3.5, 2.1, 7.6, 3.4, and 1.9 times, respectively. Differences in d_N/d_S 545 values between the paralogues might be due to high $d_{\rm N}$ caused by a weak selection at non-synonymous sites, or related to the intensity of natural selection on synonymous codon usage. To investigate the existence of codon usage bias, the effective number of codons (N_c) was calculated for each gene and species (S. pennellii and S. lycopersicum). 550 No statistically significant differences (ANOVA, P>0.01) in N_c were observed between ggps, tat, tyra, and vte3 paralogues, suggesting that the d_N/d_S differences are most probably due to a constraint relaxation for one of the gene copies rather than codon bias. In contrast, for the hppd 555 pair a significant N_c bias was observed (ANOVA, P<0.01). This can be also visualized in Fig. 4 comparing d_N and d_S values, suggesting that the d_N/d_S rate differences between hppd paralogues might not be explained by constraint relaxation.

560

Although d_N/d_S is a useful indicator of selective pressure, there is some oversimplification in its application since it may be that only certain codons in a gene can change in a way that enhances fitness whereas all others cannot accept substitutions without cost to fitness. Thus, to better explore 565 patterns of sequence variation, the orthologous sequences from A. thaliana, S. lycopersicum, and S. pennellii for all 22 genes under study were aligned and a Likelihood Ratio Test (LRT) with three models of molecular evolution was applied. The first, M0, assumes that all positions across the 570 sequences have the same level of d_N and d_S . The second, M1a, proposes that a proportion of codons is under purifying selection while the remainder have neutral evolution. Finally, M2a divides codons into three classes, those with purifying selection, those with neutral evolution 575 pattern, and the remainder with positive selection. In order to avoid data misinterpretation, the $N_{\rm c}$ was estimated and the comparison between species did not show statistically significant differences (P>0.01), indicating that there is no codon bias usage. Comparison of the three models showed 580 that for 19 genes M1a displayed the best fit, indicating that,


Fig. 4. Evolutionary rates at synonymous (dS) and non-synonymous (dN) sites of candidate genes linked to tocopherol QTL. Genes belonging to MEP, SK, and tocopherol core pathways are highlighted in red, green, and blue, respectively. VTE-related pathway candidates are not highlighted. Lines indicated estimated means.

even when a proportion of the codons for every gene are evolving neutrally, there is no support for positive selection. Nevertheless, for ggps(2), ggps(4), and vte1, M2a presented a better fit than M1a, with different levels of significance, showing that these genes exhibited signs of positive selection. Interestingly, while ggps(2) and vte1showed a higher proportion of codons evolving under positive selection (~25%) than ggps(4) (0.8%), they displayed lower significance levels. This can be explained by the high ω value of ggps(4) indicating that there is no variation in the synonymous sites along the codons under positive natural selection (Table 3).

Discussion

585

590

The two pathways feeding the main precursors of VTE biosynthesis, MEP and SK, as well as the tocopherol core route, were surveyed across the tomato genome with the aim of identifying regulatory steps of the VTE fruit biosynthesis. The metabolic reconstruction of VTE metab-600 olism resulted in the identification of 29 reactions catalysed by the protein products of 28 genes, for which 41 different S. lycopersicum loci were identified (Fig. 1, Table 1). In the course of reconstructing these routes, certain assumptions concerning the presence or absence of reactions involved in VTE metabolism had to be made. First, in the MEP 605 pathway, GPPS is the enzyme leading to monoterpene biosynthesis, while GGPS is responsible for the production of geranylgeranyl 2P by the sequential coupling of three isopentenyl 2P (IPP) molecules to dimethylallyl 2P (DMAPP). Geranylgeranyl 2P serves as a precursor for 610 carotenoid, tocochromanol, gibberelin, and chlorophyll biosynthesis (Aharoni et al., 2005; Joyard et al., 2009). Hence, the classical trend is to assume that GPPS would not be involved in tocochromanol production. Tomato plants silenced for GPPS even displayed a dwarfed phenotype and 615 reduced gibberellin levels, and did not alter carotenoid and chlorophyll content (Van Schie et al., 2007). These results thus indicated that pigments are originated from a geranyl 2P-independent geranylgeranyl 2P pool and suggested that GPPS might not influence VTE biosynthesis. However, 620 further functional experiments that validate this hypothesis are clearly needed. Secondly, HGGT, which condenses homogentisate with geranylgeranyl 2P during tocotrienol synthesis, was identified in grass species in which tocotrienols are the most abundant tocochromanol forms (Cahoon 625 et al., 2003). Neither in Arabidopsis nor in tomato was HGGT identified. However, tocotrienol traces have been detected in both tomato and tobacco (Chun et al., 2006). This might be explained by the promiscuous activity in substrate acceptance of other prenyl transferases such as 630 homogentisate solanesyl transferase (HST) and/or the homogentisate phytyl transferase (VTE2). Flux changes through the SK, MEP, and/or tocopherol pathway could conceivably shift the substrate preference of both these enzymes (Herbers, 2003; Falk and Munné-Bosch, 635 2010). This hypothesis is supported by the fact that under certain conditions tobacco plants, which do not have a canonical HGGT, and co-express the Arabidopsis HPPD and the yeast prephenate dehydrogenase, exhibit higher levels of tocotrienols (Rippert et al., 2004). Therefore, the 640

12 of 18 | Almeida *et al.*

Table 3. Parameter estimates and tests of selection	on
---	----

Gene	MO		M1a			M2a				
	InL ^a	ω ^b	InL ^a	ω0 ^c	p0d, ^e	InL ^a	ω0 ^f	թ0 ^ց	ω 2 ^h	ρ2 ^{ij}
ggps(1)	-2640.1	0.041	-2598.9	0.015	0.734**	-2598.9	0.012	0.734	2.165	0.000
ggps(2)	-2594.4	0.096	-2551.5	0.028	0.730**	-2549.9	0.032	0.746	-	0.254*
ggps(3)	-2444.5	0.014	-2411.5	0.018	0.812**	-2411.5	0.018	0.813	-	0.187
ggps(4)	-2233.7	0.061	-2200.8	0.019	0.811**	-2196.7	0.020	0.811	-	0.008**
tat(1)	-2718.5	0.101	-2690.3	0.045	0.792**	-2690.3	0.045	0.792	-	0.000
tat(2)	-2857.8	0.077	-2841.7	0.046	0.817**	-2841.7	0.046	0.817	_	0.000
tyra(1)	-2487.3	0.002	-2460.9	0.024	0.741**	-2460.9	0.024	0.741	2.629	0.000
tyra(2)	-2657.4	0.108	-2615.1	0.024	0.612**	-2615.1	0.024	0.612	1.124	0.000
vte4	-2425.8	0.063	-2391.8	0.028	0.765**	-2391.8	0.028	0.765	-	0.000
vte1	-3430.5	0.075	-3373.5	0.029	0.730**	-3371.9	0.039	0.743	_	0.257*
vte3(1)	-2138.5	0.051	-2124.3	0.031	0.900**	-2124.3	0.031	0.900	_	0.000
vte3(2)	-2256.5	0.057	-2231.6	0.022	0.842**	-2231.6	0.022	0.842	-	0.000
vte2	-2693.8	0.096	-2653.3	0.030	0.739**	-2653.3	0.030	0.739	_	0.000
vte5	-2302.1	0.079	-2287.6	0.023	0.541**	-2287.6	0.023	0.541	2.258	0.000
hppd(1)	-2586.5	0.064	-2552.8	0.024	0.818**	-2552.8	0.024	0.820	_	0.016
hppd(2)	-2617.9	0.037	-2579.5	0.020	0.805**	-2579.5	0.020	0.805	_	0.000
apt	-2447.7	0.099	-2400.6	0.024	0.771**	-2399.3	0.033	0.799	-	0.201

^a Log likehood of model.

^b Parameter estimate assuming a single d_N/d_S ratio per gene.

 c Estimated d_{N}/d_{S} for proportion of codons (ρ_{0}) under purifying selection; the rest of codons assumed to be evolving neutrally.

^e Estimated q_N/q_S for proportion of codons under purifying selection. ^e Test of M1a versus M0. ** χ^2 test using P < 0.01. ^f Estimated d_N/d_S for proportion of codons (ρ_0) under purifying selection. ^g Estimated proportion of codons under purifying selection.

^h Estimated d_N/d_S for proportion of codons (ρ_2) under positive selection.

Estimated proportion of codons under positive selection.

⁷ Test of M2a versus M1a. * χ^2 test using P<0.10; ** χ^2 test using P<0.01.

, not available.

enhanced supply of homogentisate may affect the substrate specificity of prenyl transferases leaking through tocotrienol synthesis (Falk and Munné-Bosch, 2010). In addition, A. thaliana HST, which is essential for plastoquinone-9 biosynthesis-catalysing the condensation of solanesyl 2P (SDP) and HGA-also accepts farnesyl 2P and geranylgeranyl 2P as prenyl donors (Sadre et al., 2006; Tian et al., 2007), providing further evidence in support of tocotrienol production in the absence of a specific HGGT.

Prephenate aminotransferase (PAT, EC 2.6.1.57), which converts prephenate intermediate into arogenate, has been characterized in bacteria. Although its activity had also been detected in plants, no associated loci were identified with this enzymatic function (Tzin et al., 2009). Recently, two independent reports performed biochemical and functional characterization of this plant enzyme thus, completing the identification of the genes involved in phenylalanine and tyrosine biosynthesis (Graindorge et al., 2010; Maeda et al., 2010).

One missing links still remains within the metabolic network under study, which is the absence of a phytol-P kinase that could provide phytyl 2P as an alternative to the MEP pathway (Ischebeck et al., 2006; Valentin et al., 2006).

All the MEP enzyme-encoding genes surveyed here are in accordance with sequences previously reported for tomato (Lois et al., 2000; Rohdich et al., 2000; Rodríguez-Concepción et al., 2001, 2003; Botella-Pavía et al., 2004; Ament et al., 2006; Paetzold et al., 2010; and GenBank database direct submission for IPI -GQ169536 and EU253957-), with the exception of the 2-C-methyl-D-670 erythritol 4-phosphate cytidylyltransferase (CMS)- and geranylgeranyl reductase (GGDR)-encoding genes, which have not been reported before in this species. Moreover, two additional loci for GGPS were identified here. For the SK pathway, enzyme-encoding genes involved in 675 reactions upstream of the prephenate intermediate have already been reported in tomato (Gasser et al., 1988; Schmid et al., 1992; Görlach et al., 1993, 1995; Eberhard et al., 1996; Bischoff et al., 1996, 2001). Moreover, novel loci for SDH/DHQ and CM were identified, whilst the loci encoding PAT, TyrA, and TAT were first described in tomato.

Protein localization data are highly valuable information in order to elucidate gene function. The MEP and SK pathways are well accepted to operate in the chloroplast (Rippert et al., 2009). Even so, the recent identification of some cytosolic isoforms points to the existence of an extraplastidial SK pathway (Ding et al., 2007). The results presented here for the in silico prediction of the subcellular localization of the tomato deduced protein sequences in-690 dicated that the reactions of VTE metabolism occur mostly in chloroplasts. In contrast, for the bi-functional SDH/ DHQ no signal peptide was detected, most probably due to the failure to detect true targeting, since reported analysis of subcellular fractions indicated that tomato SDH/DHQ is localized in the chloroplast (Bischoff et al., 2001). However,

685

665

Ding et al. (2007) functionally characterized a cytosolic SDH/DHQ isoform in tobacco and identified the tomato orthologue (BF096277/SGN-U578253). This locus was not included in this study due to its low identity to Arabidopsis protein according to the Pipeline criteria adopted and the absence of functional support so far. Unsuccessful subcellular prediction is not an unusual scenario, since in silico, signal peptides are occasionally misidentified. However, these predictions indicated mitochondrial targeting for the A. thaliana GPPS (data not shown), whilst experimental data revealed that this enzyme is directed to the plastid (Bouvier et al., 2000). It is important to point out the existence of some ambiguous reports on actual GPPS localization where experimental evidence supports its localization either in plastid or in the cytosol depending on the species (Nagegowda, 2010). Moreover, for the TAT and CM enzymes, the plastidial localization could not be confirmed by in silico analysis in agreement with the 715 undetermined subcellular occurrence of the post-chorismate portion of the SK pathway.

Following the identification of the tomato genes encoding all enzymes for the tocochromanol synthesis pathway, the 41 loci were further mapped. Furthermore, a detailed profile for all tocopherol isoforms was performed in fruits from ILs harbouring introgressed regions spanning the VTE biosynthesis core pathway genes (ILs 7-4, 7-4-1, 8-2, 8-2-1,

9-1, and 9-2-6) along with those in chromosome 6 (ILs 6-1 and 6-2) for which α -tocopherol QTL have been previously reported (Schauer et al., 2006). Although VTE content has been demonstrated to be a highly environmentally affected trait (Schauer et al., 2008), the results reported here indicate that at least those QTL mapped on chromosomes 6 and 9 show a relatively high heritability level as they are in

accordance with previous experiments reported by Schauer 730 et al. (2006). Moreover it is worth noting that, as opposed to the previous study, which was carried out on field-grown plants, the data reported here were obtained from greenhouse plants.

Integrated analysis at metabolic, genomic, and genetic levels allowed us to propose 16 candidate loci putatively affecting tocopherol content in tomato: prai, fpgs, chl, apt, and lycb on chromosome 6; tyra(1), vte2, hppd(1), and tat(2) located on chromosome 7; vte1 and vte4 on chromosome 8; and ggps(4), tyra(2), vte5, sec14, and 740 vte3(1) on chromosome 9. In plants, several QTL controlling tocopherol content have been identified in soybean, maize, oilseed rape, and Arabidopsis. As revealed in this report, only some of those QTL localized to areas of the genome where tocopherol core pathway genes occur 745 (Marwede et al., 2005; Gilliland et al., 2006; Chander et al., 2008; Li et al., 2010).

Detailed analysis of the identified QTL together with the co-localizing genes raised interesting features concerning VTE content regulation. None of chromosome 6 QTL colocalize with any of the MEP, SK, or tocopherol core pathway genes, thus suggesting that tocopherol content variation observed in ILs 6-1 and 6-2 might be determined by the effect of genes belonging to VTE-related pathways.

IL 6-1 showed elevated δ - and total tocopherol content in 755 comparison with S. lycopersicum (M82) control. This line harbours the S. pennellii alleles of the prai, fpgs, and chl genes. The first two could be regulating hydroxyphenylpyruvate fluxes into the tocopherol pathway by the deviation of chorismate from tryptophan and folate biosynthetic routes. On the other point of the pathway drawn in Fig. 1, the hypothesis that chlorophyll degradation-derived phytol serves as an important intermediate for tocopherol synthesis has been demonstrated by characterization of the Arabidopsis vte5 mutant (Valentin et al., 2006). In this sense, the 765 presence of S. pennellii chl allele in IL 6-1 might be raising the phytyl 2P input into tocopherol synthesis. On the same chromosome, IL 6-2 displayed significantly higher levels of β -tocopherol than the control. This IL carries S. pennellii alleles of chl, apt, and lycb. Even when these genes are linked to intermediate metabolites of the tocopherol pathway (chorismate and geranylgeranyl 2P) no evident links can be specifically associated with the β -isoform. The candidature of genes not directly involved in the VTE structural pathway is supported by the regulatory network 775 acting on branching points. The CM acting on the SK pathway is allosterically feedback-inhibited in plants by phenylalanine and tyrosine and induced by tryptophan (Tzin and Galili, 2010). Allelic variation in prai and apt could be modifying tryptophan synthesis, altering influx 780 through the SK pathway and then, increasing homogentisate precursor, finally resulting in the tocopherol content variation observed in ILs 6-1 and 6-2.

Two QTL have been detected on chromosome 7; while IL 7-4 displayed lower levels of α-tocopherol, IL 7-4-1 showed 785 reduced amounts of the β-isoform. The introgressed wild genome fragments in these ILs harbour S. pennellii alleles of tyra(1), vte2, and hppd(1). IL 7-4 also spans the wild allele of tat(2). These four candidates can alter total precursor influx to tocopherol biosynthesis. Nevertheless, the way that 790 these genes could differentially modify the amounts of tocopherol isoforms is currently unclear.

On chromosome 8, the IL 8-2 displayed a significantly lower α/γ -tocopherol ratio when compared with control (Fig. 3). Interestingly, this IL bears S. pennellii alleles of 795 vtel and vte4 genes whose protein product activities synthesize these two tocopherol isoforms. Therefore, the low α/γ -tocopherol ratio could be caused by lower VTE1 and/or higher VTE4 activity of wild alleles. Intriguingly, IL 8-2-1 also harbours the wild alleles of vte1 and vte4. Even 800 though no significant alteration in α/γ -tocopherol ratio was observed, a significant increase in total tocopherol was detected in the fruits.

IL 9-1 exhibits increased levels of α -, β -, and total tocopherol most probably due to differential activity levels 805 of the enzymes encoded by the S. pennellii allele of the tyra(2) and ggps(4) loci that could lead to higher input of hydroxyphenylpyruvate and phytyl 2P to the tocopherol core pathway. Elevated α/γ and β/γ ratios support this hypothesis (Fig. 3). Interesting to note is the fact that 810 tyra(2) shows one of the highest d_N values, indicating protein divergences between S. pennellii and S. lycopersicum

700

(Fig. 4). In this sense, the results presented here reinforce the hypothesis of Rippert *et al.* (2004) that hydroxyphenyl-pyruvate is a key step in the accumulation of VTE in plants. IL 9-2-6 exhibits increased levels of α- and total tocopherol and harbours the *S. pennellii* allele of the *vte5* gene, which could improve the input of phytyl 2P via the phytol alternative pathway (Valentin *et al.*, 2006). However, levels of β-tocopherol are not increased in IL 9-2-6, which could be explained by the more efficient activity of the *S. pennellii vte3* allele. This is also reflected in the significant differences found in the α/β ratio (Fig. 3).

Although there is a reliable link between the identified candidate genes and fruit tocopherol content, the effect of unidentified loci within the QTL cannot be discarded and, due to the considerable size of the *S. pennellii* fragments in the ILs, the differences observed in tocopherol accumulation could also be caused by undetected genes located within them.

Candidate gene approach has proved to be extremely powerful for studying the genetic architecture of complex traits (Zhu and Zao, 2007). By revealing the pattern of molecular genetic variation, the evolutionary analyses offer complementary data for strengthening gene candidature 835 (Moyle and Muir, 2010). In this sense, the pattern of selection of different genes within a metabolic pathway allows the determination of whether they are subject to equivalent evolutionary forces underlying trait phenotypic variation. Comparing S. lycopersicum and S. pennellii 840 alleles, out of the 22 genes studied, those encoding MEP, post-chorismate SK, and VTE core pathway enzymes presented d_N/d_S ratios below the mean, excluding those with more than one loci. In contrast, out of the six analysed candidate genes of related pathways, apt, chl, and lycb 845 displayed d_N/d_S ratios values above the mean (Fig. 4). This constraint relaxation cannot be related to a higher regionspecific mutation rate on chromosome 6 because two other genes mapped also on chromosome 6, prai and fpgs, presented low d_N/d_S ratios. Therefore, these results suggest 850 a strong purifying selection for tocopherol central biosynthesis genes while there is a more relaxed constraint for those genes of related pathways. When the selection constraint is evaluated codon by codon applying a likelihood ratio test, new insights about the evolutionary history 855 of the genes are revealed. Nineteen genes exhibit purifying selection associated with neutral evolving codons (Table 3) in agreement with the d_N/d_S analysis and previous reports that concluded that significant heterogeneity of evolutionary rates in metabolic pathway genes is mainly ascribed to 860 differential constraint relaxation rather than to positive selection (Livingston and Anderson, 2009; Yang et al., 2009). Even so, three loci exhibited patterns consistent with positive selection evolving codons. In tomato, loci showing positive selection have been identified associated with biotic 865 and abiotic stresses (Jiménez-Gomez and Maloof, 2009). It would be unexpected to envisage signs of positive selection in loci of major biosynthetic pathways that feed multiple metabolic routes such as MEP and SK. However, the signs of diversifying selection found for ggps(2) and ggps(4)870

could be explained by the existence of two other paralogues evolving under a more conservative evolutionary pattern. In the case of *vte1*, the reduction in protein negative selective pressure might indicate that this is not a committed step in tocopherol production.

Studies of evolutionary rates of genes in the plant anthocyanin (Lu and Rausher, 2003) and carotenoid (Livingstone and Anderson, 2009) pathways have demonstrated that upstream genes in the pathway evolved more slowly than downstream genes. However, this seems not to 880 be a constant trend. Downstream genes in the gibberellin pathway did not exhibit elevated substitution rates and instead, genes encoding either the branch point enzyme or those catalysing multiple steps in the pathway showed the lowest evolutionary rates due to strong purifying selection 885 (Yang et al., 2009). This observation is in close agreement with the theory of pathway fluxes, which indicates that natural selection would target enzymes controlling metabolic fluxes between converging pathways. Consequently, these branch points are usually targets of selection, experi-890 encing higher evolutionary constraints (Flowers et al., 2007). In this sense, the lowest value of d_N/d_S was observed for the tyra(1) gene whose protein product shares its substrate with phenylalanine/tyrosine biosynthesis resulting in branching points, whilst the tyra(2) paralogue displays 895 a relaxed evolutionary constraint, indicative of a functional divergence.

Genes responsible for adaptive morphological and physiological differences between species carry signatures of positive selection (Aguileta et al., 2010). In this sense, 900 regarding the variation in VTE content observed in the S. pennellii introgressed lines in comparison with that of S. lycopersicum, the presence of neutral and/or positive evolving codons could result in novel protein features being a source of new functional profiles. Even when coding 905 sequences are relevant to phenotype, they might not be the location at which key evolutionary changes occur. Analyses across coding sequences do not reveal allelic differences in regulatory sequences that could also be determining the observed phenotypic variations. In fact, a co-response 910 analysis of 32 of the genes identified here revealed an intricate network suggesting these pathways to be finely regulated at the level of gene expression (data not shown).

This report describes a comprehensive survey of the genes encoding VTE biosynthesis pathway enzymes in tomato, 915 and the methods adopted allowed the identification of novel tocopherol QTL. By an integrated analysis of the genome sequence data together with a well-characterized biosynthetic pathway, like that for VTE in Arabidopsis model species, this genetic/genomic approach described loci and 920 allelic variations that probably impact antioxidant content in tomato fruit. The identified candidate genes support cross-talk between the MEP, SK, and tocopherol core pathways through the control of VTE accumulation in tomato fruit. In addition, the VTE-related pathway genes 925 might contribute to regulation of the supply of intermediates for plastid tocopherol biosynthesis. The data produced provide a platform for functional studies that will

980

990

1000

contribute to elucidation of the biosynthesis and catabolism of tocochromanols, and their role in plant physiology. 930

Supplementary data

Supplementary data are available at JXB online. Supplementary Table S1 lists the gene primers used.

Acknowledgements

The authors thank Gregorio Ceccantini for assistance with 935 statistical analyses. This work was partially supported by grants from FAPESP, CNPq, and USP (Brazil); Max Planck Society (Germany); INTA, CONICET, and ANP-CyT (Argentina); and under the auspices of the EU SOL

Integrated Project FOOD-CT-2006-016214. J.A. was a re-940 cipient of a CAPES (Brazil). L.B., F.G., and N.S. were recipients of a FAPESP (Brazil) fellowships. L.Q. was a recipient of an ANPCyT (Argentina) fellowship. R.A. and F.C. are members of CONICET (Argentina). This work

was carried out in compliance with current laws governing 945 genetic experimentation in Brazil and in Argentina.

References

Abushita AA, Hebshi EA, Daood HG, Biacs PA. 1997.

Determination of antioxidant vitamins in tomatoes. Food Chemistry 60, 207-212.

Aguileta G, Lengelle J, Marthey S, et al. 2010. Finding candidate genes under positive selection in non-model species: examples of genes involved in host specialization in pathogens. Molecular Ecology 19, 292-306.

Aharoni A, Jongsma MA, Bouwmeester HJ. 2005. Volatile 955 science? Metabolic engineering of terpenoids in plants. Trends in Plant Science 10, 594-602.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403-410.

Ament K, Van Schie CC, Bouwmeester HJ, Haring MA, Schuurink RC. 2006. Induction of a leaf specific geranylgeranyl pyrophosphate synthase and emission of (E, E)-4,8,12-trimethyltrideca-1,3,7,11-tetraene in tomato are dependent on both jasmonic acid and salicylic acid signaling pathways. Planta 224, 1197-1208.

Ayres M, Ayres Júnior M, Ayres DL, Santos AA. 2007. Bioestat aplicações estatísticas nas áreas das ciências bio-médicas. Sociedade Civil Mamirauá/MCT CNPq, Belém, PA, Brasil.

Bankaitis VA, Mousley CJ, Schaaf G. 2009. The Sec14 superfamily and mechanisms for crosstalk between lipid metabolism and lipid signaling. Trends in Biochemical Sciences 35, 150–160.

Bermúdez L, Urias U, Milstein D, Kamenetzky L, Asis R, Fernie AR, Van Sluys MA, Carrari F, Rossi M. 2008. A candidate gene survey of quantitative trait loci affecting chemical composition in 975 tomato fruit. Journal of Experimental Botany 59, 2875-2890.

Bischoff M, Rösler J, Raesecke HR, Görlach J, Amrhein N, Schmid J. 1996. Cloning of a cDNA encoding 3-dehydroquinate synthase from a higher plant, and analysis of the organ-specific and elicitor-induced expression of the corresponding gene. Plant Molecular Biology 31, 69-76.

Bischoff M, Schaller A, Bieri F, Kessler F, Amrhein N, Schmid J. 2001. Molecular characterization of tomato 3-dehydroguinate

dehydratase-shikimate:NADP oxidoreductase. Plant Physiology 125, 1891 - 1900

Botella-Pavía P, Besumbes O, Phillips MA, Carretero-Paulet L, 985 Boronat A, Rodriguez-Concepcion M. 2004. Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate reductase in controlling the supply of plastidial isoprenoid precursors. The Plant Journal 40, 188-199.

Bouvier F, Suire C, d'Harlingue A, Backhaus RA, Camara B. 2000. Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. The Plant Journal 24, 241-252.

Cahoon EB, Hall SE, Ripp KG, Ganzke TS, Hitz WD,

Coughlan SJ. 2003. Metabolic redesign of vitamin E biosynthesis in 995 plants for tocotrienol production and increased antioxidant content. Nature Biotechnology 21, 1082-1087.

Carrari F, Baxter C, Usadel B, et al. 2006. Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behaviour. Plant Physiology 142, 1380-1396.

Chander S, Guo YQ, Yang XH, Yan JB, Zhang YR, Song TM,

Li JS. 2008. Genetic dissection of tocopherol content and composition in maize grain using quantitative trait loci analysis and the 1005 candidate gene approach. Molecular Breeding 22, 353-365.

Chun J, Lee J, Ye L, Exler J, Eitenmiller RR. 2006. Tocopherol and tocotrienol contents of raw and processed fruits and vegetables in the United States diet. Journal of Food Composition and Analysis 19, 196 - 204.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Research 16, 10881-10890.

Das S, Gautam N, Dey SK, Maiti T, Roy S. 2009. Oxidative stress in the brain of nicotine-induced toxicity: protective role of Andrographis paniculata Nees and vitamin E. Applied Physiology, Nutrition, and Metabolism 34. 124-135.

Davuluri GR, van Tuinen A, Fraser PD, et al. 2005. Fruit-specific RNAi-mediated suppression of DET1 enhances tomato nutritional quality. Nature Biotechnology 23, 890-895.

Ding L, Hofius D, Hajirezaei MR, Fernie AR, Börnke F, Sonnewald U. 2007. Functional analysis of the essential bifunctional tobacco enzyme 3-dehydroquinate dehydratase/shikimate dehydrogenase in transgenic tobacco plants. Journal of Experimental Botany 58, 2053-2067.

Eberhard J, Bischoff M, Raesecke HR, Amrhein N, Schmid J. 1025 1996. Isolation of a cDNA from tomato coding for an unregulated, cytosolic chorismate mutase. Plant Molecular Biology 31, 917-922.

Emanuelsson O, Brunak S, von Heijine G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols 2, 953-971.

960

16 of 18 | Almeida *et al.*

1040

1065

Enfissi EM, Barneche F, Ahmed I, et al. 2010. Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. The Plant Cell 22, 1190-1215.

Entus R, Poling M, Herrmann KM. 2002. Redox regulation of 1035 Arabidopsis 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase. Plant Physiology 129, 1866-1871.

Eshed Y, Zamir D. 1995. An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141, 1147-1162.

Falk J, Munné-Bosch S. 2010. Tocochromanol functions in plants: antioxidation and beyond. Journal of Experimental Botany 61, 1549-1566.

Fei Z, Tang X, Alba R, Giovannoni J. 2006. Tomato Expression 1045 Database (TED): a suite of data presentation and analysis tools. Nucleic Acids Research 34, D766–D770.

Fernie AR, Yaakov T, Zamir D. 2006. Natural genetic variation for improving crop quality. Current Opinion in Plant Biology 9, 196-202.

Ferro M, Brugiere S, Salvi D, et al. 2010. AT_CHLORO: 1050 a comprehensive chloroplast proteome database with sub-plastidial localization and information for functional genomics using quantitative label-free analyses. Molecular and Celllular Proteomics 9, 1063-1084.

Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt, PS, Eanes WF. 2007. Adaptive evolution of metabolic pathways in. Drosophila. Molecular Biology and Evolution 24, 1347-1354.

Foyer CH, Noctor G. 2005. Redox homeostasis and antioxidant 1060 signaling: a metabolic interface between stress perception and physiological responses. The Plant Cell 17, 1866-1875.

Fraser PD, Enfissi EM, Halket JH, Truesdale MR, Yu DM, Gerrish C, Bramley PM. 2007. Manipulation of phytoene levels in tomato fruit: effects on isoprenoids, plastids and intermediary metabolism. The Plant Cell 19, 3194-3211.

Fraser PD, Pinto ME, Holloway DE, Bramley PM. 2000. Application of high-performance liquid chromatography with photodiode array detection to the metabolite profiling of plant isoprenoids. The Plant Journal 24, 551-558.

Frusciante L, Carli P, Ercolano MR, Pernice R, Di Matteo A, Fogliano V, Pellegrini N. 2007. Antioxidant nutritional quality of tomato. Molecular Nutrition and Food Research 51, 609-617.

Garcia I, Rodgers M, Pepin R, Hssich T, Matringe M. 1999. Characterization and subcellular compartmentation of recombinant 4hydroxyphenylpyruvate dioxygenase from Arabidopsis in transgenic 1075 tobacco. Plant Physiology 119, 1507-1516.

Gasser CS, Winter JA, Hironaka CM, Shah DM. 1988. Structure, expression, and evolution of the 5-enolpyruvylshikimate-3-phosphate synthase genes of petunia and tomato. Journal of Biological Chemistry **263.** 4280-4287. 1080

Gilliland LU, Magallanes-Lundback M, Hemming C, Supplee A, Koornneef M, Bentsink L, Dellapenna D. 2006. Genetic basis for natural variation in seed vitamin E levels in. Arabidopsis thaliana. Proceedings of the National Academy of Sciences, USA 103, 18834-18841.

Görlach J, Beck A, Henstrand JM, Handa AK, Herrmann KM, Schmid J, Amrhein N. 1993. Differential expression of tomato

(Lycopersicon esculentum L.) genes encoding shikimate pathway isoenzymes: I. 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase. Plant Molecular Biology 23, 697-706.

Görlach J, Raesecke HR, Abel G, Wehrli R, Amrhein N,

Schmid J. 1995. Organ-specific differences in the ratio of alternatively spliced chorismate synthase (LeCS2) transcripts in tomato. The Plant Journal 8, 451-456.

Graindorge M, Giustini C, Jacomin AC, Kraut A, Curien G, Matringe M. 2010. Identification of a plant gene encoding glutamate/ aspartate-prephenate aminotransferase: the last homeless enzyme of aromatic amino acids biosynthesis. FEBS Letters 584, 4357-4360.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment 1100 editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95-98.

Herbers K. 2003. Vitamin production in transgenic plants. Journal of Plant Physiology 160, 821-829.

Hörtensteiner S. 2006. Chlorophyll degradation during senescence. Annual Review of Plant Biology 57, 55-77.

Ischebeck T, Zbierzak AM, Kanwischer M, Dormann P. 2006. A salvage pathway for phytol metabolism in Arabidopsis. Journal of Biological Chemistry 281, 2470-2477.

Joyard J, Ferro M, Masselon C, Seigneurin-Berny D, Salvi D, Garin J, Rolland N. 2009. Chloroplast proteomics and the compartmentation of plastidial isoprenoid biosynthetic pathways. Molecular Plant 2, 1154–1180.

Jiménez-Gomez JM, Maloof JN. 2009. Sequence diversity in three tomato species: SNPs, markers, and molecular evolution. BMC Plant 1115 Biology 9, 85.

Kamenetzky L, Asís R, Bassi S, et al. 2010. Genomic analysis of wild tomato introgressions determining metabolic- and yieldassociated traits. Plant Physiology 152, 1772-1786.

Li Y, Wang Z, Sun X, Tang K. 2008. The current opinions on the 1120 functions of tocopherol based on the genetic manipulation of tocopherol biosynthesis in plants. Journal of Integrative Plant Biology 50, 1057-1069.

Li H, Liu H, Han Y, Wu X, Teng W, Liu G, Li W. 2010. Identification of QTL underlying vitamin E contents in soybean seed among multiple environments. TAG Theoretical and Applied Genetics 120, 1405-1413.

Livingstone K, Anderson S. 2009. Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. Journal of Heredity 100, 754-761.

Lois LM, Rodríguez-Concepción M, Gallego F, Campos N, Boronat A. 2000. Carotenoid biosynthesis during tomato fruit development: regulatory role of 1-deoxy-d-xylulose 5-phosphate synthase. The Plant Journal 22, 503-513.

Lu Y, Rausher MD. 2003. Evolutionary rate variation in anthocyanin 1135 pathway genes. Molecular Biology and Evolution 20, 1844-1853.

Maeda H, Yoo H, Dudareva N. 2010. Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate. Nature Chemical Biology 7, 19–21.

1085

1095

1125

1130

Vitamin E biosynthesis in tomato | 17 of 18

)	controlling tocopherol content in winter oilseed rape. <i>Plant Breeding</i> 124, 20–26.	reductoisom fruit ripening
i	Mène-Saffrané L, DellaPenna D. 2009. Biosynthesis, regulation and functions of tocochromanols in plants. <i>Plant Physiology and Biochemistry</i> 48 , 301–309.	Rodríguez- Boronat A. hydroxymetł
	Milone D, Stegmayer G, Kamenetzky L, Lopez M, Lee J-M, Giovannoni JJ, Carrari F. 2010. *omeSOM: a software for	during carot 476–482.
)	clustering and visualization of transcriptional and metabolite data mined from interspecific crosses of crop plants. <i>BMC Bioinformatics</i> 11 , 438–447.	Rohdich F, Eisenreich Bacher A. 2
	Mobley EM, Kunkel BN, Keith B. 1999. Identification,	methyl-d-ery

characterization and comparative analysis of a novel chorismate mutase gene in. *Arabidopsis thaliana. Gene* **240,** 115–123.

114(

1144

1155

1165

Moyle L, Muir CD. 2010. Reciprocal insights into adaptation from agricultural and evolutionary studies in tomato. *Evolutionary Applications* **3**, 409–421.

Mueller LA, Mills A, Skwarecki B, Buels R, Menda N, Tanksley SD. 2008. The SGN Comparative Map Viewer. *Bioinformatics* **24**, 422–423.

Munné-Bosch S, Alegre L. 2002. The function of tocopherols and tocotrienols in plants. *Critical Reviews in Plant Sciences* 21, 31–57.

Nagegowda DA. 2010. Plant volatile terpenoid metabolism: biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Letters* **584**, 2965–2973.

Paetzold H, Garms S, Bartram S, Wieczorek J, Uros-Gracia EM, Rodriguez-Concepción M, Boland W, Strack D, Hause B, Walter MH. 2010. The isogene 1-deoxy-D-xylulose 5-phosphate synthase II controls isoprenoid profiles, precursor
pathway allocation and density of tomato trichomes. *Molecular Plant* 3, 904–916.

Peterman TK, Ohol YM, McReynolds LJ, Luna EJ. 2004. Patellin1, a novel Sec14-like protein, localizes to the cell plate and binds phosphoinositides. *Plant Physiology* **136**, 3080–3094.

- Phillips MA, D'Auria JC, Gershenzon J, Pichersky E. 2008. The Arabidopsis thaliana type I isopentenyl diphosphate isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis. *The Plant Cell* 20, 677–696.
- Ravanel S, Cherest H, Jabrin S, Grunwald D, Surdin-Kerjan Y, Douce R, Rebeille F. 2001. Tetrahydrofolate biosynthesis in plants: molecular and functional characterization of dihydrofolate synthetase and three isoforms of folylpolyglutamate synthetase in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences, USA* 98, 1523–1535.

Rippert P, Scimemi C, Dubald M, Matringe M. 2004. Engineering plant shikimate pathway for production of tocotrienol and improving herbicide resistance. *Plant Physiology* **134**, 92–100.

Rippert P, Puyaubert J, Grisollet D, Derrier L, Matringe M. 2009.

Tyrosine and phenylalanine are synthesized within the plastids in. *Arabidopsis. Plant Physiology* **149**, 1251–1260.

Rodríguez-Concepción M, Ahumada I, Diez-Juez E, Sauret-Güeto S, Lois LM, Gallego F, Carretero-Paulet L, Campos N, Boronat A. 2001. 1-Deoxy-D-xylulose 5-phosphate
reductoisomerase and plastid isoprenoid biosynthesis during tomato
fruit ripening. *The Plant Journal* 27, 213–222.
Rodríguez-Concepción M, Querol J, Lois LM, Imperial S,
Boronat A. 2003. Bioinformatic and molecular analysis of

hydroxymethylbutenyl diphosphate synthase (GCPE) gene expression during carotenoid accumulation in ripening tomato fruit. *Planta* **217,** 1200 476–482.

Rohdich F, Wungsintaweekul J, Lüttgen H, Fischer M, Eisenreich W, Schuhr CA, Fellermeier M, Schramek N, Zenk MH, Bacher A. 2000. Biosynthesis of terpenoids: 4-diphosphocytidyl-2-Cmethyl-d-erythritol kinase from tomato. *Proceedings of the National Academy of Sciences, USA* 97, 8251–8256.

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J. 2000. An alternative pathway to β -carotene formation in plant chromoplasts discovered by map-based cloning of *Beta* and *old-gold* color mutations in tomato. *Proceedings of the National Academy of Sciences, USA* **97,** 11102–11107.

Ros E. 2009. Nuts and novel biomarkers of cardiovascular disease. *American Journal of Clinical Nutrition* **89,** 1649S–1656S.

Sadre R, Gruber J, Frentzen M. 2006. Characterization ofhomogentisate prenyltransferases involved in plastoquinone-9 andtocochromanol biosynthesis. FEBS Letters 580, 5357–5362.

Saito K, Tautz L, Mustelin T. 2007. The lipid-binding SEC14 domain. *Biochimica et Biophysica Acta* **1771**, 719–726.

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR. 2008. Mode of inheritance of primary metabolic traits in tomato. *The Plant Cell* **20**, 509–523.

Schauer N, Semel Y, Roessner U, et al. 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature Biotechnology* **24**, 447–454.

Schauer N, Zamir D, Fernie AR. 2005. Metabolic profiling of leaves 122: and fruit of wild species tomato: a survey of the *Solanum lycopersicon* complex. *Journal of Experimental Botany* **56**, 297–307.

Schmid J, Schaller A, Leibinger U, Boll W, Amrhein N. 1992. The *in-vitro* synthesized tomato shikimate kinase precursor is enzymatically active and is imported and processed to the mature enzyme by chloroplasts. *The Plant Journal* **2**, 375–383.

Schenk N, Schelbert S, Kanwischer M, Goldschmidt EE, Dörmann P, Hörtensteiner S. 2007. The chlorophyllases At CLH1 and AtCLH2 are not essential for senescence-related chlorophyll breakdown in. *Arabidopsis thaliana. FEBS Letters* **581**, 12: 5517–5525.

Semchuk NM, Lushchak OV, Falk J, Krupinska K, Lushchak VI. 2009. Inactivation of genes, encoding tocopherol biosynthetic pathway enzymes, results in oxidative stress in outdoor grown *Arabidopsis thaliana. Plant Physiology and Biochemistry* **47**, 384–390.

Sokal RR, Rohlf FJ. 1981. *Biometry*. Second Edition. Freeman, New York.

Soll J, Schultz G. 1979. Comparison of geranylgeranyl and phytyl substituted methylquinols in the tocopherol synthesis of spinach chloroplasts. *Biochemical and Biophysical Research Communications*91, 715–720.

1245

1240

1230

18 of 18 | Almeida *et al.*

1250

1260

1270

Sun Q, Zybailov B, Majeran W, Friso G, Olinares PDB, van Wijk KJ. 2009. PPDB, the Plant Proteomics Database at Cornell. *Nucleic Acids Research* **37**, 969–974.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596–1599.

Thompson J, Higgins D, Gibson T. 1994. CLUSTAL W: improving 1255 the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Tian L, DellaPenna D, Dixon RA. 2007. The pds2 mutation is a lesion in the *Arabidopsis* homogentisate solanesyltransferase gene involved in plastoquinone biosynthesis. *Planta* **226**, 1067–1073.

Tohge T, Fernie AR. 2010. Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nature Protocols* **5**, 1210–1227.

Traber MG, Sies H. 1996. Vitamin E in humans: demand and delivery. *Annual Review of Nutrition* **16,** 321–347.

Tzin V, Malitsky S, Aharoni A, Galili G. 2009. Expression of a bacterial bi-functional chorismate mutase/prephenate dehydratase modulates primary and secondary metabolism associated with aromatic amino acids in. *Arabidopsis. The Plant Journal* **60**, 156–167.

Tzin V, Galili G. 2010. New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Molecular Plant* **3**, 956–972.

Valentin HE, Lincoln K, Moshiri F, et al. 2006. The *Arabidopsis vitamin E pathway gene5-1* mutant reveals a critical role for phytol kinase in seed tocopherol biosynthesis. *The Plant Cell* **18**, 212–224.

Van Schie CCN, Ament K, Schmidt A, Lange T, Haring MA, Schuurink RC. 2007. Geranyl diphosphate synthase is required for biosynthesis of gibberellins. *The Plant Journal* **52**, 752–762.

Voorrips RE. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* **93**, 77–78.

Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24, 1586–1591.

Yang Y, Zhang F, Ge S. 2009. Evolutionary rate patterns of the gibberellin pathway genes. *BMC Evolutionary Biology* **9**, 206–217.

Zamir D. 2001. Improving plant breeding with exotic genetic libraries. *Nature Reviews Genetics* **2**, 983–989.

Zhang CX, Ho SC, Chen YM, Fu JH, Cheng SZ, Lin FY. 2009. Greater vegetable and fruit intake is associated with a lower risk of breast cancer among Chinese women. *International Journal of Cancer* **125,** 181–188.

Zhao J, Last RL. 1995. Immunological characterization and chloroplast localization of the tryptophan biosynthetic enzymes of the flowering plant Arabidopsis thaliana. *Journal of Biological Chemistry* 270, 6081–6087.

Zhu M, Zhao S. 2007. Candidate gene identification approach: progress and challenges. *International Journal of Biological Sciences* **3**, 420–427. 1275

1280

1285

1290