Natural Genetic Variation in Cassava (*Manihot esculenta* Crantz) Landraces: A Tool for Gene Discovery

L J C B Carvalho[1,*], C R B de Souza[2], J C de Mattos Cascardo[3], M A Valle Agostini[1], E Alano Vieira[4], J V Anderson[5] & J Lippolis[6]

Abstract

Cassava landraces are the earliest form of the modern cultivars and represent the first step in cassava domestication. Our forward genetic analysis uses this resource to discover spontaneous mutations in the sucrose/ starch and carotenoid synthesis/accumulation and to develop both an evolutionary and breeding perspective of gene function related to those traits. Biochemical phenotype variants for the synthesis and accumulation of carotenoid, free sugar and starch were identified. Six subtractive cDNA libraries were prepared to construct a high quality (phred > 20) EST database with 1,645 entries. Macroarray and microarray analysis was performed to identify differentially expressed genes aiming to identify candidate genes related to *sugary* phenotype and carotenoid diversity. cDNA sequence for gene coding for specific enzymes in the two pathways was obtained. Gene expression analysis for coding specific enzymes was performed by RNA blot and Real Time PCR analysis. Chromoplast-associated proteins of yellow storage root were fractionated and a peptide sequence database with 906 entries sequences (MASCOT validated) was constructed. For the sucrose/starch, metabolism a *sugary* class of cassava was identified, carrying a mutation in the BEI and GBSS genes. For the pigmented cassava, a pink color phenotype showed absence of expression of the gene CasLYB, while an intense yellow phenotype showed a down regulation of the gene CasHYb. Heat shock proteins were identified as the major proteins associated with carotenoid. Genetic diversity for the GBSS gene in the natural population identified 22 haplotypes and a large nucleotide diversity in four subsets of population. Single segregating population derived from F_2, half-sibling and S_1 population showed segregation for *sugary* phenotype (93% of individuals), waxy phenotype (38% of individuals) and glycogen like starch (2% of individuals). Here we summarize our current results for the genetic analysis of these variants and recent progress in mapping of loci and with large-effect genes.

Introduction

Biochemical phenotypes are essential for the definition of gene functions and to understanding gene regulation. Currently, such functional analysis uses either forward or backward genetic approaches with laboratory-induced mutants. Studies on genetics of cassava are rare, incomplete and most of the time difficult, because of long life cycle of the plant. Mutants have not been found in cassava. Some reports have attempted to identify phenotype variants in landraces for starch type without genetic analysis [1], laboratory-induced mutants in starch [2], and linamarin cleavage [3] in a restricted number of genetic backgrounds. In cassava, an alternative to laboratory-induced mutants is identifying biochemical phenotypes in landraces in their center of origin and domestication. Since cassava is not a selfing species, most of the collected plants are highly heterozygous, making it practically impossible to find in nature an inbred line that could carry a recessive character like a mutation in phenotype using conventional genetic analysis. Therefore, exploitation of such genetic variation for either quantitative or qualitative traits needs molecular species-specific tools for genetic analysis. In considering the constrains of producing offspring and the complexity of the cassava genome, our two complementary strategies are based on association genetics analysis using natural populations and single segregating population, the latter involving more time and field experiments.

For association studies, we accessed the center of origin and domestication of cassava [4] to explore the possibility of isolating useful natural mutants. Our search was focused on starch and carotene accumulation because relatively few major genes are involved. Because resources were often limited, our study was performed in a two-stage approach by using a subset of samples to identify biochemical phenotypes and SNPs. Instead of genotyping hundreds of controls for the characterization of haplotype tag SNPs (htSNPs), we genotyped sample cases and carried out preliminary tests of association to aid the selection of htSNPs. Once the subset has been genotyped, the whole set of loci will be tested for equilibrium to proceed. In addition, cross populations are being prepared, based on a modified backcross breeding design, to obtain single segregating populations with alternate new local adapted parental divergent from the antecessor parental identified by marker assisted recurrent selection (MARS). Here, we summarize our advances on this systematic exploitation of the naturally occurring variation as a complementary resource for the functional analysis of the cassava genome.

Biochemical phenotype characterization

The use of the candidate gene approach requires considerable knowledge of the physiology and biochemistry of the phenotype. This knowledge is available for starch and carotenoid accumulation in model plants as well as grain crops and has been applied successfully in carotenoid candidate gene analysis in *Solanaceae* [5]. However, the biology of biochemical phenotypes is usually species-specific and varies with the organ and storage tissue studied. Consequently, different mechanisms of regulating starch and carotenoid accumulation are involved, including genetic background of the cultivar, as well as the general environment.

Sugary **cassava:** Normal cassava storage root accumulates a large amount of starch with distinct features such as clarity of the gel, excellent thickening (swelling capacity), neutral flavor and good-texture quality. All these properties are largely determined by the starch type and composition (amylose/amylopectin proportion). Some variation in amylose proportion has been observed, but the lack of variation in amylopectin structure is remarkable. A new class of cassava (named *sugary*) was reported [1], with high free sugar content, several variants in the starch type and composition, including amylose-free starch, glycogen-like starch, and a unique amylopectin structure when compared with normal cassava (**Fig. 1A**). Landrace CAS36.1 showed the highest glucose content (248.2 mg g^{-1} DWt) and a glycogen-like starch [1]. The

[1] Embrapa-Genetic Resources and Biotechnology, Brasilia-DF. Brazil.
[2] Universidade Federal do Para, Belém-PA. Brazil.
[3] Universidade Estadual de Santa Cruz, Ilhéus-BA. Brazil.
[4] Embrapa Cerrados Brasilia-DF, Brazil
[5] USDA-ARS, Plant Science Research Unit, Fargo-ND
[6] USDA-ARS National Animal Disease Center, Ames-AI

* Corresponding author. E-mail: carvalho@cenargen.embrapa.br

Q.Y. Shu (ed.), Induced Plant Mutations in the Genomics Era. Food and Agriculture Organization of the United Nations, Rome, 2009,**279-282**

sugar profile separated in a GC-MS analysis revealed free sugars such as arabinose, glucuronic acid, galacturonic acid, manose and xylose that were not observed in the profile for normal cassava (**Fig. 1B**). Enzyme activities and protein blot analyses revealed that the branching enzyme I (BEI), and its corresponding protein, is either low or not detected in landrace Cas36.1 [1]. Therefore, the *sugary* cassava phenotype revealed biochemical variants related to free sugar content and composition that could involve starch synthesis/degradation, as well as cell wall degradation and modification in amylopectin structure associated to the missing activity of BEI [1].

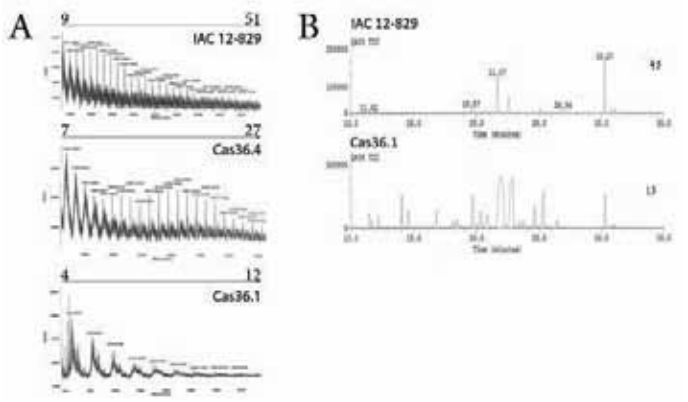

Figure 1 Sugar composition and amylopectin structure variation in cassava. **(A)** Amylopectin structure as revealed by digestion of amylopectin with β-glucosidase and polyglucan fragment identified by mass spectrometer analysis (MALDI-TOF apparatus). Polyglucan fragments varied from 9-51, 7-27 and 4-12 units of glucose for cv. IAC 12-829 and landraces Cas36.4 and Cas36.1, respectively. **(B)** Free sugar profile revealed by GC-MS analysis.

Pigmented cassava variants: Naturally occurring color variation associated with carotenoid accumulation was observed in cassava storage root. Carotenoid separation, identification and quantification by HPLC analyses indicate that total β-carotene is the major carotenoid form present and accounts for 54% to 77% of the total carotenoid in cassava storage roots. The carotenoid biosynthetic pathway is fully activated in cassava storage root, including the white phenotype. No detection of α-carotene in 24 landraces studied was observed, but variable amounts of lutein (an β-ring xanthophyll) were present. Yellow color intensity variation was associated with the accumulation of different carotenoids. Landraces with white storage root showed a profile with eight types, whereas intense yellow showed 17 types. Variation in total β-carotene content ranged from none in landrace Mirasol (pink) that accumulates only lycopene (99.81 μg g^{-1} DWt), to 49.91 (μg g^{-1} DWt) in landrace MC008 (deep yellow) that accumulates β-carotene (66%) its colorless precursor phytoene (31%) and its derived intermediate β-criptoxanthin (3%). Variations, in α and β-ring xanthophyll content, preferentially accumulate high amounts of lutein or violaxanthin, together with β-carotene, such as in landraces MC002 and MC016, respectively. The possible mechanisms responsible for this genetic variation include genetic regulation by genes coding for enzymes in the syntheses pathway as observed with the pink landraces, variants in the yellow that differentially synthesize β-carotene and β-ring xanthophylls (violaxanthin, neoxanthin, luteoxanthin and auroxanthin) but not α-ring xanthophylls (lutein) and carotenoid cleavage enzymes. In addition, the differential accumulation of β-carotene could possibly be associated with the sink capacity of chromoplasts related to carotenoid sequestering proteins as shown below.

Chromoplast-associated proteins analysis: Carotenoid accumulation is likely to be regulated by the carotenoid sequestering proteins in chromoplasts and may have a three-fold importance to cassava: firstly, to explain the range of β-carotene content in the yellow color gradient, secondly, to better understand the carotenoid protein sequestering mechanism in non-photosynthetic tissue and, thirdly, to account for protein content enhancement in cassava storage root due to the need for more functional proteins in the root tissue. In exploring this alternative, we first observed that protein content was 40-60% higher in intense yellow than in white cassava [6]. Correlation studies between total carotenoid and protein content indicated that the level of significant correlation is dependent on the method and procedure used to estimate protein and that protein is linearly correlated with β-carotene (**Fig. 2**). To dissect this correlation, a procedure combining chromoplast-enriched suspension, fractionation of carotenoid-protein complex in size exclusion chromatography (SEC) and protein blot analysis was developed. **Figure 3A and 3B** show two protein peaks separation in the SEC profile for white and yellow root. The carotenoid-protein complexes in peak 1 revealed β-carotene (**Fig. 3C**) as the major pigment present in yellow and traces of it in white landraces. 2DE gel [6] and SDS-PAGE analysis of fractions of peak 1 and peak 2 showed a large protein number difference between yellow and white color phenotypes (data not shown). To identify which proteins are specifically associated with β-carotene present in peak 1, the protein complex in each peak was sequenced.

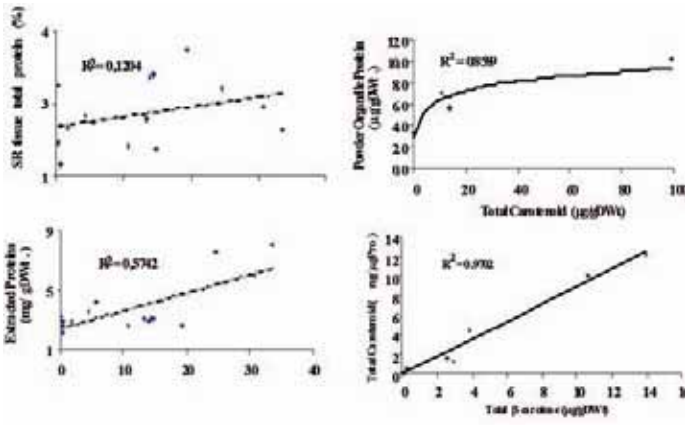

Figure 2 Correlation between total carotenoid and protein content. **(A)** crude protein (%, factor 3.24). **(B)** saline buffer soluble protein (mg/ gDWt, Bradford). **(C)** Enriched chromoplast suspension proteins quantified (Bradford). **(D)** Total carotenoid expressed on protein base. Total carotenoid was extracted and quantified by spectrometry. β-carotene was separated by HPLC and amount estimated by peak integration.

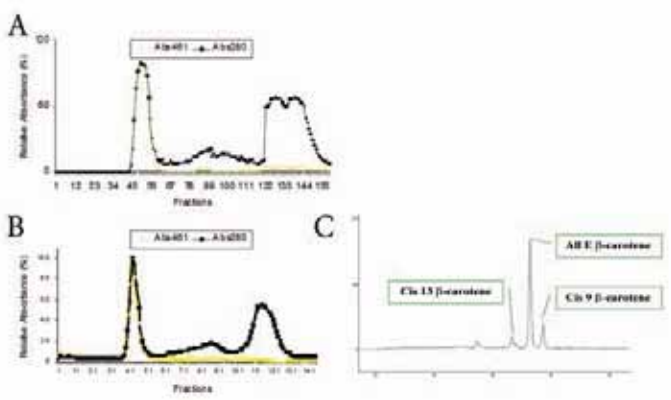

Figure 3 Protein-carotenoid complex characterization. Size exclusion chromatography separation of carotenoid-protein complex (Peak 1) and non-associated protein (Peak 2) in yellow **(A)** and white **(B)**. **(C)** HPLC carotenoid profile present in Peak 1 identifying β-carotene.

Molecular isolation and characterization of candidate genes responsible for naturally occurring variation

To find candidate genes of interest, global (macroarray and microarray) and specific gene expression analysis was performed. Six subtractive

cDNA were constructed and an EST database was assembled for *sugary* and pigmented cassava. Clones of cDNA fragments for genes coding for enzymes related to starch and carotenoid synthesis were sequenced. Chromoplast-associated proteins sequences and their corresponding genes were identified using three EST cassava databases (http://genoma.embrapa.br/genoma/, http://titan.biotec.uiuc.edu/cgi-bin/ESTWebsite/estima_start?seqSet=cassava, http://www.brc.riken.go.jp/lab/epd/Eng/catalog/pDNA.shtml) to identify differentially expressed proteins and genes.

Differentially expressed gene in *sugary* cassava: Global differentially expressed genes between normal and *sugary* cassava were accessed with a membrane base cDNA chip built up with 264 genes and hybridized with cDNA obtained by reverse transcription of total RNA population. Results indicated that the major genes differentially expressed are largely related to stress response such as up-regulated gene for ABA synthesis (two genes), transcription factor homolog related to hypoxia (2 genes), transport proteins for glucose/ABA (one gene)], nitrogen (one gene), and three unknown genes. Due to the low resolution of this technique, seven cDNA clones for gene coding for enzymes related to starch synthesis were cloned, sequenced and used for mRNA blot analysis. Results showed that the expression of the gene coding for BEI is missing in the *sugary* phenotype with glycogen-like starch [1]. For the waxy starch type, the analysis of cDNA and genomic DNA sequence of the GBSSI gene for normal and sugary cassava carrying waxy starch type showed a nonsense mutation on the N-terminal in the coordinates 337-372 of the mRNA (data not shown).

Differentially expressed genes in pigmented cassava: Global differentially expressed genes (microarray) between normal and pigmented cassava was accessed with a cDNA chip built up with 24,000 Euphorbiaceae genes and hybridized with cDNA obtained by reverse transcription of total RNA population. Preliminary results indicated that three carotenoid cleavage enzyme coding genes were upregulated in relation to pigmented cassava (not shown). Sequences of cloned cDNA fragments for the genes coding phytoene synthase (CasPSY), phytoene desaturase (CasPDS), carotenoid isomerase (CasCRTISO), lycopene β-cyclase (CasLCYb), β-ring hydroxilase (CasHYb), and neoxanthin synthase (CasNXS) were obtained. Transcript profiles for those genes across landraces contrasting carotenoid HPLC profiles consistently correlated with end products of carotenoid synthesis. Transcript levels for CasPSY were equivalent in all landraces evaluated, while trace values of CasLCYb and CasHYb were observed in landraces that accumulate only lycopene (Mirasol) or β-carotene (MC008), respectively. Taken together, these results indicated two kinds of color mutations related to carotenoid synthesis: the pink color genotype, which accumulates only lycopene with traces expression value detected for the gene CasLCYb, and the intense yellow genotype (MC008), which accumulates mainly β-carotene, indicating a down regulation of the gene coding for CasHYb.

Identification of carotenoid-associated proteins: From the SEC results above, proteins recovered from Peak 1 were sequenced by a classic PROTEOMIC procedure. The majority of the pigment-associated proteins in Peak 1 belong to the small heat shock proteins (sHSP) family class I and II. This provides evidence to support roles of sHSPs on the specificity of β-carotene accumulation in cassava, as observed for chaperonin21 for lycopene in tomato [7, 8]. The mechanism by which sHSPs promotes carotenoid accumulation is yet unknown. To gain knowledge on the sHSPs effects on β-carotene differential accumulation in yellow root phenotype, cDNA sequence for the gene coding for all the sHSPs identified in the carotenoid-protein complex were obtained and tested with Real time PCR. The results confirmed the protein blot experiment for fibrilin protein (not shown) and indicated traces of HSP18.1 protein in white cassava and a 7.2-fold higher expression value for the intense yellow phenotype. Together, these results indicate a possible gene mutation related to the carotenoid protein-sequestering mechanism in white cassava. Genes coding for sHSPs are not single copy but are distributed in six different classes, hence making it difficult to be used as a candidate gene in the genetic analysis.

Forward genetic studies of naturally occurring variation

Our forward genetic analysis is focused on two alternative experimental designs. In the first one, a population genetic analysis is carried out with candidate genes derived from the biochemical phenotypes as described above and tested under an evolutionary perspective in a subset sample of a population including cassava ancestors (33 individuals), and landraces (121 individuals). The analytical procedure follows the rational and computer soft as (9,10). In the second, crossing populations based on a modified backcross breeding design to obtain single segregating populations for mapping, field evaluation and new cultivars are being prepared.

Genetic analysis of natural populations: Two candidate genes coding for starch synthesis (CasBEI and CasGBSSI) and three for carotenoid synthesis (CasPSY, CasLCYb and CasHYb) were selected and are sequenced across a population of 154 individuals. Here we show our preliminary analyses in a subset sample for the N-terminal region of the CasGBSSI. Table 1 summarizes the statistics describing the sequence diversity in the combined and each subset of the four classes of population. Larger nucleotide diversity is observed in the cassava ancestor followed by the pigmented cassava. Tagima's D value was highly significant for the combined subset sample indicating genetic neutrality. Haplotype number and diversity was also high for combined and ancestor subset.

Table 1. Genetic diversity estimates and neutrality test for GBSS (N-Terminal).

Penotype	Sample	Nucleotide Diversity (π)	Theta (θ)	Tagima's D	Haplotype number	Haplotype diversity
Ancestor	4	0,02044 (0,00689)	0,02187 (0,00157)	-0,66711 (NS)	22	0,966 (0,00031)
Normal	14	0,00978 (0,00021)	0,01659 (0,00688)	-1,73316 (NS)	12	0,978 (0,00119)
Sugary	8	0,00749 (0,00165)	0,00889 (0,00458)	-0,77880 (NS)	6	0,929 (0,084)
Pigmented	6	0,01992 (0,00619)	0,02416 (0,01212)	-1,10378 (NS)	5	0,933 (0,01481)
Combined	32	0,01232 (0,00245)	0,03068 (0,01006)	-2,29800 **	22	0,966 (0,018)

Values in parenthesis are standard deviation.
(NS) is not statistically significant.
** is statistically significant at P<0.0001.

The likely genealogical history of the genetic diversity showed in **Table 1** was inferred from RM networks constructed for haplotypes in the combined (**Fig. 4A**) and separate population samples (**Fig. 4B-E**).

Genetic analysis in single segregating populations: Field-tests for agronomic performance and morphological descriptors have been performed in the Cerrado region of Brazil since 2002 in a provenance test. Results showed identical *sugary* and pigmented phenotypes as observed early in the center of origin and domestication. In 2006, a breeding programme was initiated to transfer these new identified traits to local varieties. In addition, RAPD and SSR markers were also applied and showed a large genetic diversity among the *sugary* accessions. Agronomic performance indicated that the landrace Cas36.17 (waxy starch type) yielded about 9 ton/ha (12 month growth season base) while the best-adapted local variety (cv. Japonesinha) yielded 28 ton/ha. All the other *sugary* accessions showed mainly lower yield due to a severe attack of Xanthomonas axonopodis pv. manihotis. This information was used to better orientate the selection of landraces within individuals obtained in the crossed populations (F_1 segregating and self pollination population). Preliminary results for a half sibling population (55 individuals) tested in the 2006/2007 growth season revealed segregation of *sugary* phenotype

(93% of the individuals), waxy phenotype (38% of the individuals) and glycogen-like starch (2% of the individuals). Individuals (185) of the S_1 population for the best-adapted *sugary* landrace were planted in the 2007/2008 growth season in the Cerrado region and are being evaluated as above.

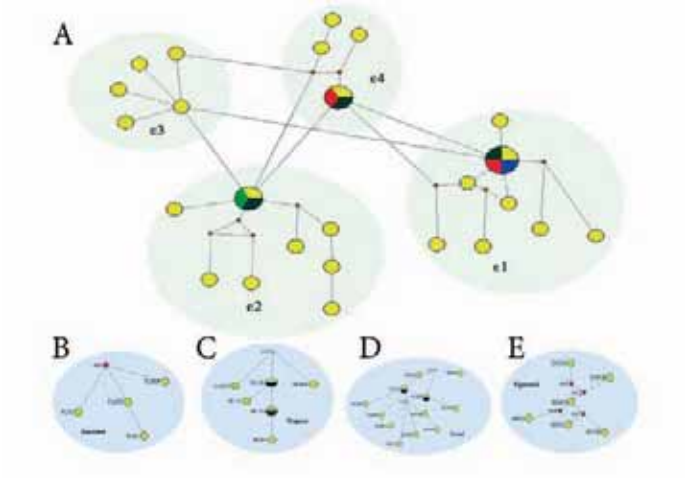

Figure 4 RM network of GBSS sequence haplotype. **(A)** Combined samples and sample subset for Cassava Ancestor **(B)**, *sugary* cassava **(C)**, normal cassava **(D)** and pigmented cassava **(E)**.

Concluding remarks and prospects

Cumulatively, our results confirm the importance of landraces, the efficiency of developing species-specific molecular tools directly from the naturally occurring variations and point out new research for cassava, related to protein enhancement in the storage root.

The exploitation of these variants will increase and become more systematic and efficient with the development of more permanent mapping populations, high-throughput genotyping technologies, improved QTL mapping statistical methods and more precise assays for phenotype analysis.

ACKNOWLEDGEMENTS

For financial support, we would like to thank Conselho Nacional de Desenvolvimento Cientifica e Tecnológico – CNPq (Grant # 480410.2001-1); Programmea. Nacional de Pesquisa em Biotecnologia – CENARGEN (Project Nº 060302058) and IAEA (contract # 13188.). The germplasm derived from this study was donated to Harvest-Plus Initiative and deposited in CIAT germplasm collection.

BIBLIOGRAPHY

1. Carvalho, L.J.C.B., Souza, C.R.B., Cascardo, J.M.C., Junior, C.B., Campos, L. Identification and characterization of a novel cassava (*Manihot esculenta* Crantz) clone with high free sugar content and novel starch. *Plant Molecular Biology* **56**, 634-659 (2004).

2. Raemkers, K., Schreuder, M., Suurs, L., Furrer-Verhorst, H., Vincken, J.P., Vetten, N., Jacosen, E., Visser, R. Improved Cassava Starch by Antisense Inhibition of Granule-bound Starch Synthase I. *Molecular Breeding* **16**, 163-172 (2005).

3. Siritunga, D., Sayre, R.T. Engineering cyanogen synthesis and turnover in cassava (*Manihot esculenta*). *Plant Molecular Biology* **56**, 661-669 (2004).

4. Schaal, B.A., Olsen, K.M., Carvalho, L.C.B.J. In: Darwin's Harvest: New Approaches to the Origins, Evolution, and Conservation of Crops edited by Timothy J. Motley, Nyree Zerega, and Hugh Cross. Columbia University Press, New York, NY, p.390 (2005).

5. Thorup, T.A., Tanyolac, B., Livingstone, K.D., Popovsky, S., Paran, I., Jahn, M. Candidate gene analysis of organ pigmentation loci in the Solanaceae. *PNAS* **97**, 11192-11197 (2000).

6. Carvalho, L.J.C.B., Agostini, M.A.V., Junior, C.B. Sixth International Scientific Meeting of Cassava Biotechnology Network. Cali, Colombia (2004).

7. Neta-Sharir, I., Isaacson, T., Lurie, S., Weiss, D. Dual Role for Tomato Heat Shock Protein 21: Protecting Photosystem II from Oxidative Stress and Promoting Color Changes during Fruit Maturation. *Plant Cell* **17**, 1829–1838 (2005).

8. Sabehat, A., Weiss, D., Lurie, S. The Correlation between Heat-Shock Protein Accumulation and Persistence and Chilling Tolerance in Tomato Fruit. *Plant Physiol.* **110**, 531–537 (1996).

9. Templeton, A.R., Maxwell, T., Posada, D., Stengard, J.H., Boerinkle, E., Sing, C.F. Tree scanning: a method for using haplotype trees in phenotype/genotype association studies. *Genetics* **169**, 441-453 (2005).

10. Posada, D., Maxwell, T.J., Templeton, A.R. TreeScan: a bioinformatic application to search for genotype/phenotype associations using haplotype trees. *Bioinformatics Applications Note* **21**, 2130-2132 (2005).