Detection of natural and induced mutations from next generation sequencing data in sweet orange bud sports

M. Caruso[1], G. Las Casas[1], D. Scaglione[2], S. Gattolin[3], L. Rossini[4], G. Distefano[5], F. Cattonaro[2], A. Catara[6], G. Licciardello[6], M. Morgante[7], C. Licciardello[1a]

[1]Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria – Olivicoltura Frutticoltura Agrumicoltura (CREA-OFA), Corso Savoia 190, 95024 Acireale, Italy; [2]IGA Technology, lstituto di Genomica Applicata Technology Services (IGA-TS), via Linussio 51, 33100 Udine; [3]Parco Tecnologico Padano, via Einstein Albert, Lodi, Italy; [4]Università degli Studi di Milano (UniMi), via Celoria, Milano, Italy; [5]Università degli Studi di Catania (Unict) via Valdisavoia 5, Catania, Italy; [6]Laboratory for Phytosanitary Diagnoses and Biotechnology, Science and Technology Park of Sicily s.c.p.a., Blocco Palma I, via V. Lancia 57, Z.I. 95121 Catania, Italy; [7]Dipartimento di Scienze Agrarie ed Ambientali, Universita' di Udine, Via delle Scienze 208, 33100 Udine, Italy

Abstract

Somatic mutations are a cause of intraspecific diversity in many fruit crops, including citrus. In the case of sweet orange [*Citrus sinensis* (L.) Osbeck], intraspecific variability is determined only by somatic mutations. Tools for clonal fingerprinting are required by breeders and nurserymen and have important implications for traceability. With the aim of studying genomic variability and identifying mutational events responsible for varietal diversification, we deep-resequenced 22 accessions including navel, common and blood oranges using an Illumina platform. We also resequenced an induced mutant of Tarocco 'Scirè D2062' obtained using two cycles of gamma ray irradiation at 40 Gy. A robust and reliable set of single nucleotide polymorphisms (SNPs), structural variants (SVs) and indels, specific of each accession or common to varietal groups, was identified both in natural and induced mutants. A subset of SNPs, transposable element insertions and small indels (2 to 18 base pairs) was validated by Sanger sequencing, PCR amplification and high resolution melting analysis, confirming the results of the bioinformatics analysis. Moreover, to identify a reliable marker set for traceability of specific cultivars, we collected leaf and juice samples from many Italian citrus growing areas and used a KASP platform for their fingerprinting. These tools will be useful to prove true-to-typeness of specific sweet orange varieties.

Keywords: *Citrus sinensis*, resequencing, indel, single nucleotide polymorphisms, structural variants, traceability

INTRODUCTION

Sweet orange [*Citrus sinensis* (L.) Osbeck] is the most important citrus species for fresh market and processing, contributing with more than 50% to the global citrus production. Genome sequencing revealed that sweet orange originated from a complex hybridization of parents having mandarin and pummelo ancestry (Wu et al., 2014; Velasco and Licciardello, 2014). More recently, Wu and colleagues (2018) proposed that the species arose from pummelo introgression into a type-2 (early-admixture) mandarin (i.e; a mandarin having a small percentage of pummelo alleles).

All sweet orange varieties derive from somatic mutations and cannot be easily discriminated using traditional molecular markers. All clones differ for very few characters and share the majority of genome sequence with little variation. Mutations usually arise

[a] E-mail: concetta.licciardello@crea.gov.it

naturally, and they are selected by growers and maintained for several generations through clonal propagation. The long history of cultivation has led to the selections of hundreds of cultivars, divided into varietal groups based on fruit characteristics (common, navel, blood). Each varietal group is also divided into subgroups (i.e; Valencia types among the common ones; Tarocco, Moro and Sanguinello among the blood oranges). To induce additional variability, it is possible to use several kinds of mutagenic agents, such as gamma rays, trying to improve horticulturally important traits.

Generally, mutations are caused by nucleotide base substitutions, indels or structural variants (SVs), in addition to inversions, duplications and translocations of entire chromosomes. In citrus, most of the mutations responsible of specific phenotypic changes, such as in fruit color, ripening period, acidity, and fruit size remain unknown. One of very few examples shading light on the molecular basis of somatic mutations is related to anthocyanin pigmentation, a typical trait of blood oranges. It is known that all blood orange selections differ from common ones for the presence of a Copia-like retrotransposon located upstream a MYB gene regulating anthocyanin biosynthesis (Butelli et al 2012, Butelli et al., 2017).

Whole genome resequencing of different sweet orange varieties might be helpful to identify the causative mutations leading to specific phenotypes, and to develop markers for clonal fingerprinting. In this paper we report the identification of different types of markers for sweet orange fingerprinting. After marker selection, we developed a system for clonal fingerprinting using different techniques. We tested the system analyzing several DNA samples of some of the most diffused blood orange clones collected from experimental and commercial farms and nurseries of different locations of Southern Italy. Moreover, we identified and validated mutations in an irradiated Tarocco clone.

MATERIALS AND METHODS

Resequencing of sweet orange varieties and data analysis

Twenty-two different sweet orange clones and varieties, including common, navel and pigmented ones were collected at Palazzelli experimental orchard of CREA (Siracusa, Italy) and resequenced using an Illumina HiSeq2000/2500. Among the blood oranges, three clones of Moro, three belonging to Sanguigno/Sanguinello group and eight Tarocco cultivars and mutants were resequenced. In addition, the genome of a Tarocco 'Scirè II gen', a 'Scirè D2062' mutant generated from two consecutive cycles of gamma ray irradiation (each cycle at a dosage of 40 Gy), was also resequenced.

Bioinformatics analysis for the discovery and filtering of single nucleotide polymorphisms (SNPs) and small indels (2 to 18 bp) was performed at Istituto di Genomica Applicata-Technology service (IGA-TS) in Udine (Italy) with custom pipelines (data not shown). SVs, specifically insertions of transposable elements, were detected using the methods described in Pinosio et al (2016).

Samples collection for marker validation

For marker validation, juice from 20 fruits and 20 leaves picked from different parts of the canopy were sampled from each resequenced clone. Juice was processed and filtered under sterile conditions. 500 µL were used to extract genomic DNA using TES buffer and the protocol described in Möller et al. (1992). Around 100 mg for each leaf sample was homogenized and used to extract DNA using TES buffer protocol.

To validate SNPs using KASP assay, samples of different blood orange cultivars were collected from plants coming from three different origins: Palazzelli experimental orchard; 19 commercial fields distributed in South East Sicily, Calabria, Basilicata and Sardinia; three nurseries, of which two in Sicily (respectively West and East coasts) and one in Apulia.

Marker Validation

A subset of filtered SNPs, indels and transposable element insertions was selected to identify reliable polymorphic markers for clonal fingerprinting. The analysis focused on seven resequenced clones, which represent some of the most diffused blood orange cultivars: Tarocco 'Ippolito M507', Tarocco 'Lempso C nuc', Tarocco 'TDV', Tarocco 'Meli C8158', Tarocco 'Gallo C898', Tarocco 'Scirè D2062', Tarocco 'Dal Muso 2b-14-10'. SNPs and small indel validation was also performed on the induced mutant. Sanger sequencing was used to validate SNPs. Bidirectional sequencing was performed using an ABI 3130 Genetic Analyser (Life Technologies). In addition, a KASP assay was performed using Master mix and KASP Assay mix according to manufacturer's instructions (www.lgcgroup.com). The KASP platform was based on 48 SNPs, including 3 SS SNPs from each resequenced cultivar, plus two to four GS SNPs. A total of 190 DNA samples were genotyped, of which 95 correspond to leaves and 95 to juice. Small indels were validated using High Resolution Melting (HRM), according to Caruso et al (2014). SVs were analyzed through classical PCR, and PCR products were visualized on 1% agarose gel. Two sets of primers were designed for each putative insertion: the first set amplified the left region, with the forward primer located in the reference genome and the reverse in the putative transposable element (sx); the second set amplified the right region using a forward primer designed in the putative transposable element and the reverse primer located in the genome (dx). Both indel and SV validations were only performed on samples collected from the same plants used for Illumina resequencing.

RESULTS and DISCUSSION

The use of custom bioinformatics pipelines developed at IGA-TS identified a set of SNPs, indels and SVs, specific of each cultivar/clone or shared among groups or subgroups, defined "sample-specific" (SS) and "group-specific" (GS) respectively. Different approaches were used to validate polymorphisms on leaf and juice samples collected in different growing areas of southern Italy, including the CREA experimental farm, and commercial orchards and nurseries.

SNP validation using Sanger sequencing

From a total of 212 SS and 39 GS SNPs, 52 SS and 15 GS SNPs were selected for validation using Sanger sequencing. 38 SS and 13 GS markers were validated by Sanger sequencing (Figure 1a), including some of the induced mutant. The missing validations could be due, among others, to the identification of false positives or to possible chimeric events of specific clones.

Validation of SS and GS SNPs using KASP

KASP assay was used for SNP validation on a larger number of samples of the seven pigmented varieties listed in M&M. This experiment was useful to identify a subset of reliable markers to be used for traceability of plant material collected in different growing areas and nurseries (Figure 1b).

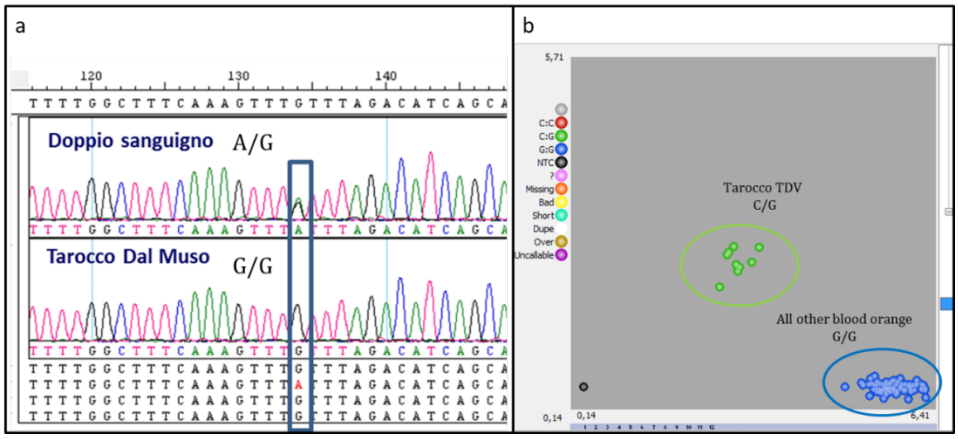

Figure 1. Example of SNP validation using Sanger sequencing (a); KASP assay used for genotyping 8 Tarocco 'TDV' samples, which were clearly discriminated form the rest of the analysed samples (b).

The effectiveness of clonal fingerprinting varied with the analyzed cultivars. For example, in the cases of Tarocco 'Lempso C nuc.' we validated all SNPs in most samples coming from experimental and commercial orchards and nurseries, both from leaves and juice (Table 1). Only a few samples were not correctly genotyped, probably due to an imprecise sample assignment (i.e; there are different 'Lempso' clones cultivated in Sicily), or to mislabelling. In the case of Tarocco 'Meli C8150' we validated SNPs on 100% samples coming from the fields. On the other hand, the success rate for SNP validation drastically decreased for other cultivars. Specifically, SS SNPs from 'Scirè D2062', Gallo 'C898' or 'Dal muso 2b-14-10' were not validated in many samples coming from commercial orchards and generically labelled as Tarocco 'Scirè', Tarocco 'Gallo' and Tarocco 'Dal muso'. This result is probably linked to the fact that these orchards are composed of heterogeneous populations of clones rather than cultivars.

Table 1. Examples of SNP validation using KASP assays in two Tarocco cultivars. Sample origin and tissue used for DNA isolation are indicated. "X" indicates the validation, dashes indicate missing samples, 'NV' indicates that the SNPs were not validated.

Variety	Origin	Tissue	juice			leaves		
			SNP1	SNP2	SNP3	SNP1	SNP2	SNP3
Tarocco 'Lempso C nuc'	experimental orchard	leaves/juice	X	X	X	X	X	X
	commercial orchard	leaves/juice	X	X	X	X	X	X
	commercial orchard	leaves/juice	X	X	X	X	X	X
	commercial orchard	leaves/juice	X	X	X	X	X	X
	commercial orchard	juice	NV	X	NV	-	-	-
	nursery	leaves	-	-	-	X	X	X
	nursery	leaves	-	-	-	X	X	X
	commercial orchard	leaves	NV	NV	NV	NV	NV	NV
	producers	leaves	NV	NV	NV	NV	NV	NV
	producers	leaves	NV	NV	NV	NV	NV	NV

	experimental orchard	leaves/juice	X	X	X	X	X	X
	commercial orchard	leaves/juice	X	X	X	X	X	X
	commercial orchard	leaves/juice	X	X	X	X	X	X
	commercial orchard	leaves/juice	X	X	X	X	X	X
	commercial orchard	leaves/juice	X	X	X	X	X	X
Tarocco 'Meli C8158'	producers	leaves	-	-	-	X	X	X
	producers	leaves	-	-	-	X	X	X
	producers	leaves	-	-	-	X	X	X
	producers	leaves	-	-	-	X	X	X
	producers	leaves	-	-	-	X	X	X
	producers	leaves	-	-	-	X	X	X
	producers	leaves	-	-	-	X	X	X

SS and GS indels

Custom pipelines identified SS and GS indels between 2 bp and 18 bp in the 23 resequenced genomes. From the 80 SS and 34 GS indels identified, we selected 26 SS and 13 GS indels for validation. 14 SS and 10 GS indels were correctly validated using HRM (Figure 2a). SS indels were also confirmed in the gamma ray irradiated clone (Tarocco 'Scirè II gen') as reported in Figure 2b.

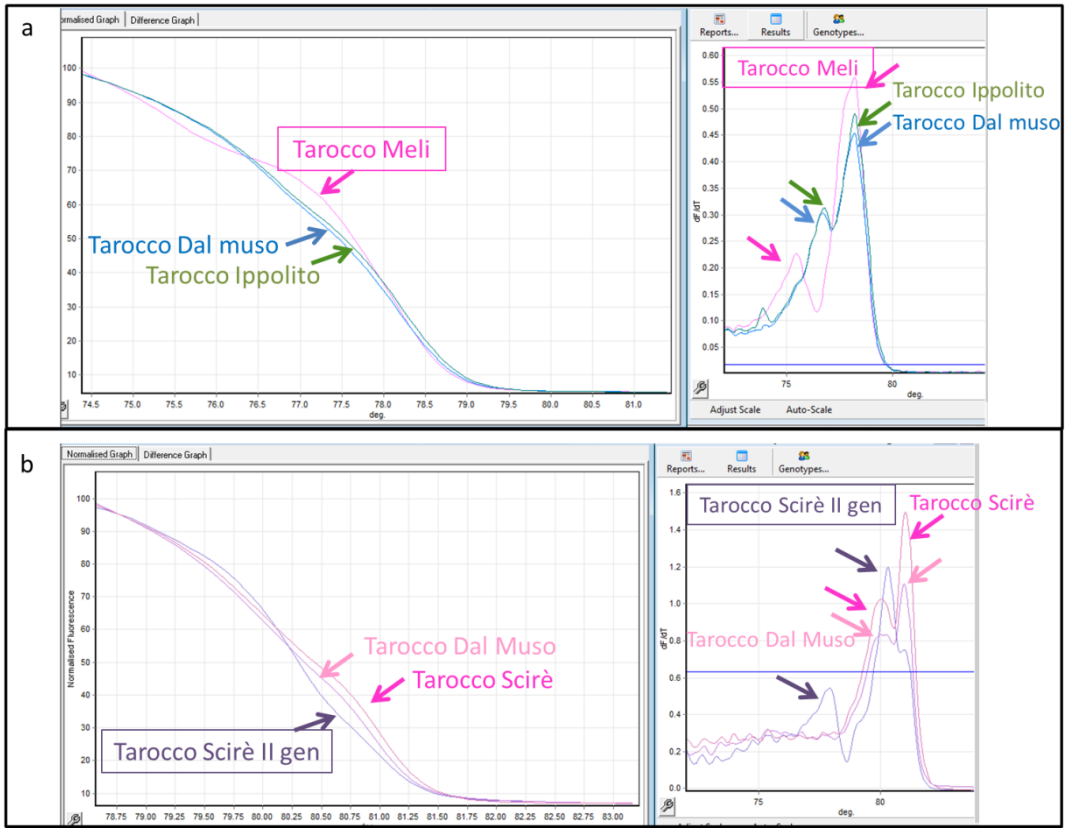

Figure 2. High Resolution Melting plots showing indel validations. Tarocco 'Meli C8158' shows a different melting curve compared to other two Tarocco cultivars (a); validation of a deletion in the irradiated Tarocco 'Scirè II gen' (b).

Validation of structural variants

A total of 40 SS and 75 GS SVs were identified among the 22 sweet orange resequenced clones (the analysis was not performed on the irradiated clone). 13 SS and 9 GS SVs were selected for validation through PCR amplification.

With the strategy described in M&M, 8 SS and 7 GS SVs were correctly amplified and validated (Figure 3), confirming the reliability of the bioinformatics pipeline.

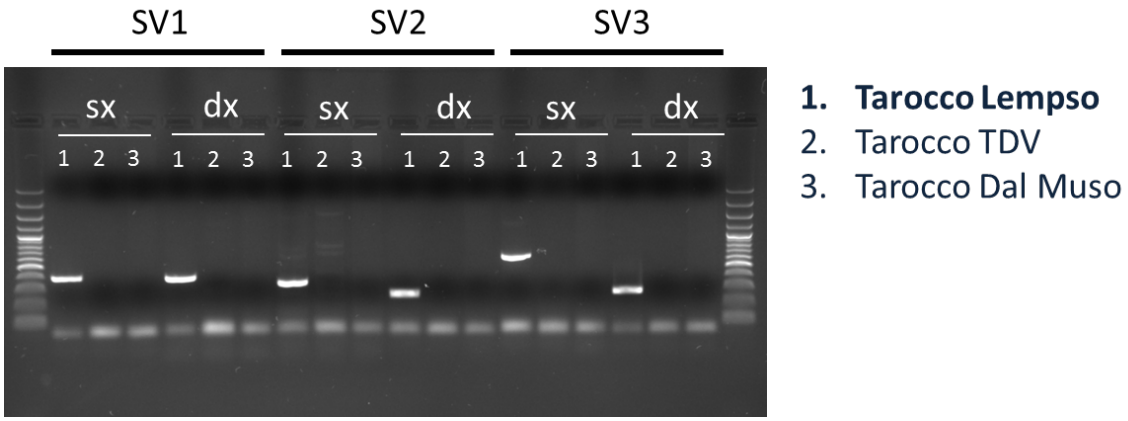

Figure 3. Agarose gel electrophoresis confirming three transposable element insertions in Tarocco 'Lempso C nuc', in comparison with two negative controls ('TDV' and 'Dal Muso 2b-14-0').

CONCLUSIONS

Resequencing of 15 pigmented sweet orange cultivars and mutants identified a reliable set of SNPs, indels and SVs. In particular:

- A subset of SS SNPs correctly discriminated specific Tarocco commercial clones (i.e. 'Ippolito M507', 'Meli C8158', 'TDV', 'Lempso C nuc') with traceable origin from the mother blocks to nurseries to the field.
- For other cultivars, such as 'Scirè', 'Gallo', or 'Dal Muso', SS SNPs were not always validated, probably due to (1) uncertain origin or (2) presence of clonal populations rather than a single clone. Also, we cannot exclude mislabeling during the sample collection.
- The correspondence between leaf and juice samples was confirmed in the 92.3% of the samples. This allow the use of most of the identified SNPs for juice traceability. The few differences that occurred between leaf and juice samples might be due to chimeric events occurring in L1 or L2 cell layers, which are not equally represented in the two tissues.
- Several SS and GS SVs and indels were validated. They could be used as additional tools for clonal traceability.
- Resequencing of a gamma ray induced mutant identified a number of SNPs and indels comparable to those identified among natural mutants.

ACKNOWLEDGEMENTS

This work was funded in part by MIUR and EU through the National Program PON R&C 2007-2013, project 'IT-Citrus genomics' (PON 01_01623), leaded by Science and Technology Park of Sicily, Catania (Italy) and was partially supported by the "QUALITRACE - Messa a punto e validazione di tool genetici e chimici per la tracciabilità integrata e la valorizzazione della qualità dell'arancia rossa di sicilia IGP" project (D.M. 19527/7303/2016), founding granted by Italian Ministry of Agriculture Food and Forestry.

Literature cited

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato Recupero, G., Martin, C. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell *24*, 1242–1255.

Butelli, E., Garcia-Lor, A., Licciardello, C., Las Casas, G., Hill, L., Reforgiato Recupero, G., Keremane, M.L., Ramadugu, C, Krueger, R., Xu, Q., et al., (2017). Changes in Anthocyanin Production during Domestication of Citrus. Plant Physiol. *173*, 2225–2242

Caruso, M., Distefano, G., Pietro Paolo, D., La Malfa, S., Russo, G., Gentile, A., Reforgiato Recupero, G. (2014). High resolution melting analysis for early identification of citrus hybrids: A reliable tool to overcome the limitations of morphological markers and assist rootstock breeding. Sci Hortic *180*, 199–206.Moller, E.M., Bahnweg, G., Sandermann H., Geiger, H.H. (1992). A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. Nuc Ac Res. *20*, 6115-6116

Pinosio, S., Giacomello, S., Faivre-Rampant, P., Taylor, G., Jorge, V., Le Paslier, M.C., Zaina, G., Bastien, C., Cattonaro, F., Marroni, F., Morgante, M. (2016). Characterization of the poplar pan-genome by genome-wide identification of structural variation. Mol. Biol. Evol. *33*, 2706–2719.

Velasco, R., Licciardello, C. (2014). A genealogy of the citrus family. Nat. Biotechnol. *32*, 640–642.

Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Takita, M.A., et al., (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat. Biotechnol. *32*, 656–662.

Wu, G.A., Terol, J., Ibanez, V., López-García, A., Pérez-Román, E., Borredá, C., Domingo, C., Tadeo, F.R., Carbonell-Caballero, J., Alonso, R., et al., (2018). Genomics of the origin and evolution of Citrus. Nature *554*, 311–316.