DNA AND TRANSCRIPTOMIC MARKERS AS TOOLS TO AUTHENTICATE PRIMARY PRODUCTS IN FOOD CHAIN

Daria Scarano

Dottorato in Scienze Biotecnologiche – XXVI ciclo
Indirizzo Biotecnologie per le Produzioni Vegetali
Università di Napoli Federico II

Dottorato in Scienze Biotecnologiche – XXVI ciclo
Indirizzo Biotecnologie per le Produzioni Vegetali
Università di Napoli Federico II

DNA AND TRANSCRIPTOMIC MARKERS AS TOOLS TO AUTHENTICATE PRIMARY PRODUCTS IN FOOD CHAIN

Daria Scarano

Dottoranda:	Daria Scarano
Relatore:	Prof.ssa Rosa Rao
Coordinatore:	Prof. Giovanni Sannia

INDEX

RIASSUNTO

Il pomodoro appartiene alla famiglia delle *Solanaceae*, una delle più importanti famiglie tra le angiosperme per valore agronomico ed economico. La pianta originaria dell'America del Sud, si diffuse, poi, anche in America Centrale e a farla conoscere in Europa furono gli Spagnoli nel XVI secolo. La coltivazione della pianta del pomodoro era già diffusa in età precolombiana ma era utilizzata solo come pianta ornamentale poiché considerata velenosa a causa del suo alto contenuto di solanina. L'utilizzo del frutto come vegetale commestibile sembra risalire al periodo del 1600. Probabilmente in Italia furono avviati i primi tentativi di coltivazione che poi si è diffusa principalmente nei paesi del Mediterraneo (Soressi 1969; Esquinas-Alcazar e Nuez, 1995; Garcia-Martinez et al., 2006).
I sistemi di riproduzione vanno dall'allogamia, nelle specie selvatiche, all'autogamia, nella specie coltivata. Quest'ultima ha una limitata variabilità genetica, conseguenza soprattutto delle selezioni che si sono verificate durante la domesticazione e l'evoluzione delle moderne cultivar.
Il pomodoro riveste un importante ruolo economico in tutto il mondo grazie ai differenti impieghi sul mercato. È all'undicesimo posto per quantità nella lista dei prodotti più consumati al mondo con una produzione annua di 160 milioni di tonnellate nel 2012. L'Italia si classifica al settimo posto tra i paesi produttori, con circa 5 milioni di tonnellate (FAOSTAT 2012).
Inoltre, nel 2004 ha avuto inizio il sequenziamento del genoma del pomodoro, attraverso la fondazione del Tomato Genome Consortium, a cui hanno partecipato 10 diversi paesi da Europa, Asia e America. Al termine del progetto sono state definite le specifiche del genoma di pomodoro che presenta una dimensione di circa 930 Mbp e 34.727 geni codificanti per proteine (http://solgenomics.net/organism/Solanum_lycopersicum/genome).
Il pomodoro 'San Marzano' è un importante componente dell'agroalimentare "Made in Italy" noto in tutto il mondo per le caratteristiche proprietà organolettiche. È ampiamente impiegato per il consumo fresco ma il ruolo da protagonista viene assunto dai pelati inscatolati, e nel 1996 è stato insignito del marchio di denominazione di origine protetta (DOP) con la costituzione di un disciplinare di produzione *ad hoc* che ne regola la coltivazione.
Recenti indagini sul mercato agroalimentare italiano hanno evidenziato illeciti a carico della filiera del pomodoro e in particolare sulle produzioni 'San Marzano' DOP: falsi marchi DOP in etichetta, falsi disciplinari di qualità, assenza di tracciabilità e lavorazione in Italia di triplo concentrato cinese etichettato come *Made in Italy*.
La necessità di definire sistemi validi ed economici per tracciare i prodotti alimentari ha assunto un'importanza sempre maggiore da quando la globalizzazione del commercio e l'industrializzazione dei processi produttivi hanno reso impossibile il controllo diretto della produzione alimentare da parte dei consumatori.
Con il termine di frode alimentare si considera un atto o inganno che si traduce in una diminuzione del valore, economico o nutritivo, della merce perpetuata da produttori o rivenditori attraverso la modifica delle caratteristiche dell'alimento rendendolo diverso da quello di partenza.
Le frodi alimentari vengono messe in atto con diverse condotte illecite allo scopo di adulterare, contraffare, sofisticare, alterare i prodotti alimentari e dichiarare in maniera fraudolenta l'origine di provenienza geografica con il fine ultimo di trarre vantaggio economico.

Lo sviluppo di sistemi di tracciabilità per prodotti alimentari è diventato sempre più esigente poiché consumatori, e sempre più spesso anche produttori, sono alla ricerca di garanzia sulla qualità degli alimenti. Di conseguenza, la tracciabilità è diventata un elemento significativo nel settore della sicurezza alimentare.
Il Regolamento CEE178/2002 definisce la rintracciabilità come: "la possibilità di ricostruire e seguire il percorso di un alimento, di un mangime, di un animale destinato alla produzione alimentare o di una sostanza destinata o atta ad entrare a far parte di un alimento o di un mangime attraverso tutte le fasi della produzione, della trasformazione e della distribuzione". Nella pratica, tracciare una filiera agroalimentare, significa raccogliere i dati che si generano lungo il percorso dal campo alla tavola, ogni volta che si completa una fase produttiva in qualsiasi punto della filiera.
Negli ultimi anni, come conseguenza alle recenti emergenze riguardanti la BSE (Bovine Spongiform Encephalophaty) nei bovini, la contaminazione di diossina nei mangimi zootecnici e di mercurio nel pesce, è aumentata la richiesta di sistemi di identificazione, registrazione ed etichettatura di prodotti alimentari.
L'identificazione può essere effettuata in modo generico adottando un logo attraverso il quale è possibile tracciare le aziende che hanno partecipato alla produzione del prodotto alimentare, oppure indicando sulla confezione finale tutte le aziende della catena (nel caso il numero sia limitato), o, ancora, utilizzando codici a barre che permettono di inserire molte informazioni occupando uno spazio limitato.
Ma come si può dire se il prodotto corrisponde realmente a ciò l'etichetta dichiara? E, inoltre, che cosa rende un prodotto alimentare diverso da un altro?
La tracciabilità permette di garantire il consumatore da possibili frodi e tutelare scelte alimentari individuali. Le informazioni su un alimento sono essenziali per dare ai consumatori la libertà di scegliere un prodotto alimentare piuttosto che un altro. La scelta può riflettere lo stile di vita o di religione (il vegetarianismo, la preferenza per i prodotti biologici, assenza di maiale per ebrei e musulmani), o problemi di salute (ad esempio, assenza di arachidi, lattosio o glutine per le persone con particolari allergie). Inoltre, la possibilità di verificare con sistemi oggettivi e scientifici l'origine dei prodotti accresce il valore della certificazione di qualità (come ad esempio i marchi IGP, DOP) ed amplia le possibilità di scelta del consumatore. Ecco perché la descrizione erronea e l'etichettatura scorretta di un prodotto alimentare sono illegali, soprattutto se il cibo è stato processato eliminando la possibilità di distinguere i diversi componenti.
In questo scenario, marcatori del DNA si sono rivelati validi strumenti per l'autenticazione degli alimenti e per contrastare frodi alimentari: ad esempio, per determinare l'identità dei componenti degli alimenti sono stati utilizzati tratti del genoma mitocondriale in alimenti a base di carne e in prodotti ittici (Cai et al., 2011; Teletchea et al., 2008; Filonzi et al., 2010; Barbuto et al., 2010) e, allo stesso modo, per alimenti di origine vegetale è stato impiegato il genoma plastidiale (De Mattia et al., 2011). Inoltre, i marcatori molecolari sono stati ampiamente impiegati nella protezione di marchi di qualità, quali IGP per la carne (Arana et al., 2002) o DOP per alimenti di origine vegetale (Rao et al., 2009; Scarano et al., 2011). Infine, i marcatori del DNA possono essere impiegati nel controllo di filiera valutando la quantità di specie diverse presenti in un alimento (Sonnante et al., 2009; Pasqualone et al., 2007).
Nel presente lavoro di tesi è stata valutata la potenzialità di impiego di due tipologie di marcatori, marcatori del DNA e marcatori trascrittomici, ai fini della tracciabilità e autenticazione nella filiera agro-alimentare del pomodoro.

Il genoma del pomodoro, così come quello di altri organismi vegetali, include molte sequenze di DNA ripetute. Tra le diverse classi di sequenze altamente e mediamente ripetute in un genoma, d'interesse ai fini del *fingerprint* molecolare, vi è quella rappresentata dai microsatelliti. Conosciuti anche come sequenze semplici ripetute (SSR) o brevi ripetizioni in tandem (STR), essi rappresentano sequenze specifiche di DNA costituite da unità ripetute in tandem di lunghezza generalmente compresa tra 2 e 6 nucleotidi rappresentanti il *core*, ad alta variabilità e fiancheggiate da sequenze altamente conservate. Essi sono altamente polimorfici, codominanti, dispersi in tutto il genoma, disponibili in molte specie vegetali, specie specifici e richiedono un corto DNA-stampo (Zane et al. 2002). Tali marcatori sono stati utilizzati con successo per studi genetici in diverse specie che spaziano dall'uomo alle piante passando per il regno animale. Inoltre, grande utilità dei marcatori SSR è stata dimostrata in studi di tracciabilità di filiera sia per alimenti di origine animale che vegetale.

In questo lavoro di tesi, tredici marcatori molecolari SSR He et al., 2003; Smulders et al., 1997) sono stati impiegati per il *fingerprinting* di 145 genotipi di pomodoro ampiamente diffusi in Italia ed in particolar modo in Campania. I risultati hanno indicato che i marcatori SSR selezionati hanno un elevato polimorfismo nell'ambito della popolazione analizzata, riuscendo a discriminare il 72% dei genotipi all'analisi. Dall'analisi sono stati evidenziati un totale di 71 diversi alleli SSR con una media di 5.462 alleli per locus. La capacità di discriminare tra varietà, cultivar ed ecotipi analizzati è supportata dal potere discriminante calcolato per ciascuno dei tredici marcatori che ha mostrato una media di 0.484 con un valore massimo di 0.706 al locus LE20592. Gli indici genetici Ho (eterozigosità osservata) e He (eterozigosità attesa) hanno, invece, fornito informazioni sulla percentuale di genotipi eterozigoti ritrovati nella popolazione analizzata e indicazioni sulla possibilità che altri genotipi al di fuori di quelli analizzati siano eterozigoti ai loci in considerazione, rispettivamente. I valori riscontrati per Ho variano da 0 a 0.172 con una media di 0.086, indicando che in media solo l'8.6% dei 146 genotipi analizzati mostra un profilo allelico eterozigte, mentre per He il range è compreso tra 0.183 e 0.703, suggerendo quindi che ai loci considerati, dato il numero di alleli osservato, la possibilità di evidenziare eterozigosi è in media piuttosto elevata.

L'attribuzione di profili allelici SSR identificativi ha, inoltre, permesso di verificare che i genotipi all'analisi si relazionano tra loro in base a caratteristiche che li accomunano. Infatti, dall'analisi dell'intera popolazione, si evidenzia che le varietà industriali di pomodoro si raggruppano separatamente rispetto alle varietà locali coltivate in Campania. Inoltre, nell'ambito del gruppo di varietà Campane, è stato identificato un sottogruppo formato dai tipi 'San Marzano'.

Infatti, nella popolazione analizzata, 40 genotipi appartengono alla varietà 'San Marzano', tra cui 'San Marzano 2' e 'Kiros' (inclusi nel disciplinare di produzione del pomodoro 'San Marzano' DOP). I dati SSR prodotti dai 13 marcatori sono stati impiegati per studiare e approfondire le conoscenze relative alle relazioni genetiche esistenti tra i differenti tipi 'San Marzano'. In particolare, sei loci sono risultati monomorfici mostrando un unico allele, mentre per gli altri sette è stato evidenziato un numero di alleli compreso tra due e tre. L'analisi del parametro Ho ha rivelato che, per tutti a loci analizzati, i quaranta ecotipi mostrano uno stato allelico omozigote. Infine, il potere discriminante ha mostrato un valore medio di 0.146 ed un valore massimo di 0.461, confermando la potenzialità dei marcatori SSR selezionati di discriminare genotipi molto simili tra loro ed evidenziando una variabilità interna nell'ambito della popolazione 'San Marzano' in grado di discriminare 'San Marzano 2'

e ‘Kiros’ dagli altri tipi ‘San Marzano’ che possono essere ugualmente venduti come ‘San Marzano’ ma senza l’apposizione del marchio di qualità DOP.
Sei dei 13 loci SSR sono stati poi selezionati per l’analisi di 33 inscatolati di pomodoro pelato per studi di autenticazione nella filiera del pomodoro “San Marzano DOP dell’Agro-Sarnese Nocerino”. L’analisi molecolare ha avuto come finalità quella di verificare la corrispondenza tra quanto dichiarato sull’etichetta ed il contenuto dell’inscatolato. I profili allelici SSR ottenuti dall’analisi degli inscatolati sono stati confrontati con quelli prodotti in precedenza per i tipi ‘San Marzano’ inclusi nel disciplinare. Il confronto delle dimensioni alleliche e dello stato allelico a ciascun locus ha portato alla conclusione che gli inscatolati di pomodoro etichettati come “San Marzano DOP dell’Agro-Sarnese Nocerino” non contengono realmente bacche di pomodoro ‘San Marzano’ DOP.
Un ulteriore aspetto di interesse per la tracciabilità è rappresentato dalla possibilità di identificare l’origine geografica dei prodotti che vengono impiegati nelle filiere alimentari. In particolar modo, la componente ambientale di uno specifico luogo di produzione è strettamente legata alle caratteristiche organolettiche e nutritive di un alimento al punto da renderlo qualitativamente migliore rispetto ad altri simili.
Per il ‘San Marzano’ l’origine geografica è una componente fondamentale per l’attribuzione del marchio DOP. Infatti, la coltivazione del pomodoro ‘San Marzano’ deve essere effettuata in specifici areali della regione Campania e, quindi, è di grande interesse la selezione di marker ambiente-specifici idonei a tracciarne la provenienza geografica.
L’approccio trascrittomico prevede l’identificazione, mediante tecniche di deep sequencing, di profili di espressione genica caratteristici di specifici genotipi, da cui estrapolare uno o più geni la cui espressione può servire da bio-marcatore. Alla base c’è l‘assioma che cambiamenti fisiologici e biochimici di un organismo riflettono la modulazione di espressione di molti geni. Le nuove strategie di sequenziamento (NGS-New Generation Sequencing) hanno promosso la disponibilità di un nuovo approccio per lo studio delle variazioni di espressione genica e la quantificazione dei trascritti in campioni biologici, definito come RNA-seq. La strategia del sequenziamento dell’RNA offre numerosi vantaggi, oltre allo studio dell’ espressione genica, tra i quali l’analisi di organismi per i quali non è stata ancora determinata la sequenza genomica (organismi non modello), informazioni sulle giunzioni esone-introne e sui siti di splicing, scoperta di variazioni di sequenza (ad esempio SNP) e nuove isoforme dei trascritti.
L’approccio trascrittomico, basato sul sequenziamento dell’mRNA, ha fornito la possibilità di identificare markers ambiente-specifici. In particolare, è stato analizzato il trascrittoma proveniente da bacche di ‘Kiros’ raccolte da piante allevate in due diversi areali di produzione (Acerra e Brusciano) e, parallelamente, la stessa analisi è stata condotta su un genotipo di pomodoro ibrido da industria, ‘Docet’.
Il sequenziamento *paired-end*, da 36 milioni di *reads*, ha fornito sequenze la cui qualità è stata valutata mediante l’analisi del Phred-score associato a ciascuna base sequenziata. In media le sequenze hanno mostrato un grado di accuratezza nel *base-calling* pari al 99.99%.
Tramite l’impiego di opportuni software, le *reads* sono state mappate sul genoma di pomodoro di riferimento e, successivamente alla valutazione della qualità del mapping, i dati, ovvero il numero di reads assegnate a ciascun gene, sono stati normalizzati (CPM, Count Per Million, Robinson et al., 2010)
Utilizzando i dati normalizzati, è stata possibile una primaria discriminazione tra i campioni raccolti nei due ambienti di produzione, confermando quindi che la crescita

dello stesso genotipo in due ambienti diversi si traduce in una differenza di espressione genica tale da riuscire a separare nettamente i due ambienti.
Per le analisi successive, sono stati selezionati soltanto i geni che risultavano avere un'espressione superiore al valore soglia di 1 CPM. I geni così selezionati (circa 13700 per campione) sono stati sottoposti ad arricchimento funzionale al fine di individuare le categorie ontologiche (GO terms) più rappresentate per ogni campione. Dalle categorie ontologiche statisticamente più significative sono poi stati estrapolati geni ambiente-specifici. Tale selezione ha portato alla produzione di liste di geni che risultano, pertanto, espressi per un dato genotipo esclusivamente in un determinato ambiente ed assenti nello stesso genotipo allevato nell'altro ambiente. Infine, attraverso l'impiego di un software idoneo all'analisi di geni differenzialmente espressi (DEG), sono stati evidenziati i geni che mostrano la più alta probabilità di essere differenzialmente espressi per lo stesso genotipo nei due differenti areali di produzione (potenziali bio-marcatori ambiente-specifici).
Lo studio di tali geni selezionati ha evidenziato una forte presenza di prodotti genici coinvolti nella riparazione del DNA e nella risposta allo stress ossidativo (ad esempio MutS, MCM7, dUTP pyrophosphatase, RPA e ADNT1). Le piante impiegate nell'analisi sono state raccolte in campo, ad Agosto, a completa maturazione delle bacche e, pertanto è ipotizzabile che la presenza dei suddetti geni rispecchi un adattamento volto alla protezione dalle radiazioni UV.
Studi recenti hanno, inoltre, evidenziato come le radiazioni UV possono influenzare le qualità organolettiche e nutritive dei frutti di pomodoro (Jagadeesh et al., 2009; Liu et al., 2009). Alla luce di questo è ipotizzabile che l'adattamento delle piante alle forti radiazioni in campo, evidenziato al livello del trascrittoma, sia connesso anche alla modifica del bouquet dei composti che si accumulano nei frutti e che ne conferiscono aroma e proprietà nutritive.
L'analisi dell'espressione genica ha suggerito una chiara distinzione tra i due ambienti di coltivazione per i genotipi analizzati. In particolare, lo studio congiunto di due varietà di pomodoro diverse negli stessi due campi ha permesso di comprendere meglio quali geni, tra quelli differenzialmente espressi, sono fortemente influenzati dall'ambiente e quali sono maggiormente legati al genotipo, seppur, ovviamente, influenzati anch'essi dall'ambiente di coltivazione.
In conclusione, il presente lavoro di tesi ha dimostrato che l'impiego di marcatori del DNA e marcatori trascrittomici può essere d'aiuto nell'autenticazione e verifica dell'origine genetica e geografica delle materie prime impiegate nella filiera agro-alimentare del pomodoro ed essere strumento per combattere le frodi alimentari.
La tracciabilità, supportata da opportuni strumenti molecolari, può davvero rappresentare il ponte tra consumatori e produttori perché fornisce i mezzi per rendere i sistemi di produzione trasparenti e sicuri.

SUMMARY

Tomato (*Solanum lycopersicum*) has an important economic role in the world thanks to the different applications on the market. It is the eleventh among commodities produced worldwide with a yield average of 37 tons/ha and about 160 million tons of tomatoes were produced in the world in 2011. China is the largest producer followed by India and United States while Italy, with 6 million of tons, is the seventh producer country (FAOSTAT 2011; http://faostat.fao.org).
The tomato genome sequencing project started in 2004 by a consortium of 10 countries and the release of the complete tomato genome sequence was in 2012 (Tomato Genome Consortium).
Among different tomato varieties, ‘San Marzano’ is a premium variety covered by an EU Protected Designation of Origin (PDO) label, cultivated in different areas of Campania region.
Unfortunately, media constantly refer of unscrupulous producers that adulterate, alter or replace the premium products in tomato food chain with the goal to maximize illegally profits. There are rules aimed to protect against fraudulent substitution of quality product in food chain but this is not enough without tools able to verify the material incoming and outcoming.
Food traceability, that is the possibility to identify an agricultural product at every step of production, processing and commercialisation, from farm to table, is central for the identification of improper labelling of processed food and feed.
The present work evaluates the potential use of two types of markers, DNA and transcriptomic markers, in order to authenticate and trace ‘San Marzano’ tomato food chain.
A set of 13 SSR (He et al., 2003; Smulders et al., 1997) markers was used to characterize a population of 145 tomato genotypes. Selected SSRs revealed an high polymorphism able to discriminate 72% of the genotype under analysis. Furthermore, SSR allelic profiles allowed verifying that industrial varieties are grouped separately from the local varieties from Campania Region and suggest that several sub-grouping are present, such as ‘San Marzano’ types subgroup.
The same set of SSR was employed to study specifically the subgroup composed by 40 ‘San Marzano’ types, which contains 'San Marzano 2' and 'Kiros' (varieties included in the procedure guideline for tomato 'San Marzano' PDO production). The analysis revealed that all ‘San Marzano’ types were homozygotes and confirmed the power of selected SSR markers to discriminate 'San Marzano 2' and 'Kiros' from other 'San Marzano' types, which can also be sold as 'San Marzano' but without the PDO label.
Six out of the 13 SSR loci were then selected for analysis of 33 peeled tomatoes labeled as ‘San Marzano’ PDO in order to verify the correspondence between what is declared on the label and the contents. SSR allelic profiles were compared with those produced for 'San Marzano' PDO types showing no match. The analyzed peeled tomatoes do not contain tomato berries from 'San Marzano' PDO varieties.
On the other hand, the transcriptomic approach provides the identification of gene expression profiles characteristic of genotype-environment association. In particular, the transcriptome from berries of 'Kiros' plants collected from two different production areas (Acerra and Brusciano) was analyzed and the same analysis was conducted on ‘Docet' genotype.

Among up-regulated highlighted genes in both comparisons most are involved in UV and oxidative stress response and signalling transduction. The occurrence of genes whose products are involved in DNA damage repair should have a biological meaning of plant self-protection from UV-radiation and can reflect in fruits compounds bouquet.

Gene expression analysis suggested a clear distinction between locations for the analyzed genotypes. Furthermore, the combined study of two different tomato varieties in the same two fields has allowed a better understanding of which genes, among those differentially expressed, are strongly influenced by the environment and which are more related to the genotype, although, of course, influenced by the location.

In conclusion, the present work has shown that the use of DNA markers and transcriptomic markers can help in authentication of genetic and geographical origin of raw materials used in the tomato agro-food chain and can be a tool to contrast food fraud.

Traceability, supported by appropriate molecular tools, can really represent the bridge between consumers and producers making food production transparent and safety.

1. INTRODUCTION

1.1 Food Frauds and Traceability

Food fraud is a collective term used to define the deliberate and intentional substitution, addition, tampering, or misrepresentation of food, food ingredients, or food packaging or false statements made about a product for economic gain.
The desire to make a fraudulent profit by mislabeling of food products has always been a feature of societies. Food frauds have been conducted since antiquity: evidence has been found of counterfeit Roman seals on amphorae containing fraudulent olive oil and wine (Mello et al., 1982; Purcell, 1985; Armstrong, 2009). In 1861, Arthur Hill Hassall performed one of the earliest scientific studies about the authenticity of food. He employed microscopy to investigate the authenticity of coffee, an extremely expensive commodity at that time, finding that 31 out of 34 samples contained adulterants such as chicory, roasted wheat and burnt sugar.
In recent years, the negative impact of food-borne illness (Wilcock et al., 2004; Fischer et al., 2007; Knowles et al., 2007; Lobb and Mazzocchi, 2007) and the occurrence of food safety incidents (Wilcock et al., 2004; De Jonge et al., 2007; Lobb et al., 2007; Houghton et al., 2008) has been associated with reduced consumer confidence in food safety. Chemical and microbial food contaminants represent an important food safety topic (Tent, 1999), but an additional consumer interest is the knowledge of the genetic origin of the raw materials used for the production of a food product.
Food is a human necessity irreplaceable that moves huge economic interests inducing sometimes producers to obtain illegal profits through fraudulent actions, which have as main aim to make products easy to sell and gainfully.
There are different classifications for food fraud (Woolfe and Primrose, 2004):

- substitution of one ingredient by a similar but cheaper one;
- extending or adulterating food with a cheaper or base material;
- presence of undeclared ingredients;
- extending or adulterating food to increase value;
- non-declaration or false declaration of processes;
- over-declaration of a quantitative ingredient;
- false claims regarding geographical or production origin.

Consumers demand for a food product that must be safe – in terms of microbiological safety, contaminants, etc.– and correctly described in terms of its nature, composition, ingredients and origin.
Labelling legislation is there to ensure that food is properly described. It tries to protect consumers from being sold an inferior product with a false description and, in addition, to protect honest producers from unfair competition. Enforcement of this legislation ensures that correctly described products remain available to the consumer and that consumer confidence is maintained, which in turn ensures a market place for these foods.
A document of great importance in the agro-food sector is represented by the European regulation 178/2002, which defines the traceability as “the ability to trace and follow a food, feed, food producing animal or ingredients, through all stage of production and distribution”. The mentioned regulation is the fundamental law on food safety in Europe and, since its application from 1st January 2005, the definition

of a traceability system for the whole food sector has become mandatory in all member countries.
Practically, trace a food chain means collecting all data that are generated from field to table, every time a production stage is complete, at any point in the supply chain: at the seed level, in the farm, in the companies before, during and after the processing, until distribution to consumers.
Traceability helps to ensure consumer from possible fraud and protect individual food choices that can reflect lifestyle or religious concerns (e.g. vegetarianism, preference for organic products, absence of pork for Jews and Muslims), or health concerns (e.g. absence of peanuts, lactose or gluten for individuals with particular allergies).
In addition, the possibility to verify the origin of products increases the value of quality certification (such as PGI and PDO), favouring the development of local economy, through the promotion of products from specific places, and providing incentives for the conservation of local ecotypes preserving biodiversity.
Characteristics of a traceability system depend on objectives, costs and benefits. As proposed by Golan and the workgroup (2004), a traceability system might be characterized by: breadth, that is the amount of information the traceability system records; depth that is the capability of the system to capture relevant information along the different steps of the food chain (for instance from the filed to the processing industry); precision that reflects the degree of assurance with which the tracing system can identify a particular food product .
Traditional tracking systems are based on label and barcodes in which are included information about processes, companies, production areas and everything that happens to the product before, during and after the manufacturing and packaging (Regattieri et al., 2007).
Nowadays, genetic traceability, through the use of DNA-based markers, is extensively widespread because DNA analysis furnishes deeper levels of identification, such as individual, breed and species discrimination, resulting able to detect fraud and protect typical productions (Scarano et al., 2012). To determine the genetic identity of food components, portions of the mitochondrial genome were used in meat and fish products (Cai et al., 2011; Teletchea et al., 2008; Filonzi et al., 2010; Barbuto et al., 2010) and, similarly, in plant derived food the plastidial genome proved to be helpful (De Mattia et al., 2011); furthermore, molecular markers were widely employed in protection of quality label such the PGI label for meat (Arana et al., 2002) or the PDO label for fresh agro-food products (Rao et al., 2006) and plant derived foods (Scarano et al., 2011; Rao et al., 2009); finally, DNA markers can also help in the evaluation of the different species that are present in a food products or the quantitative declaration of a specific food component (Sonnante et al., 2009; Pasqualone et al., 2007).
So, the availability of analytical methods, which can ensure the authenticity of foods through traceability systems, plays a fundamental role in modern societies as a new tool to strike food frauds.

1.2 Identification of production areas in the agro-food sector

In the wide scenario of food authentication, great importance is attributed to the geographical origin of raw materials that constitute a food product.

Some foodstuffs are described as coming from a particular country or region, in order to protect and promote them, because peculiar characteristics of the final products depend primarily on the area in which they are produced.
European legislation (CEE No. 2081/92) has been developed for the protection of geographical indications and designations of origin for agricultural product and foodstuffs also by the assignment of quality labels, such as PDO and PGI.
The Protected Designation of Origin (PDO) is recognized and attributed to foodstuff whose peculiar characteristics depend primarily or exclusively from the geographic area in which they are produced. Production, processing and preparation must take place in a defined geographical area for a PDO product.
The title of Protected Geographical Indication (PGI) is given, instead, to those agricultural products and foodstuffs for which a given quality or other characteristic depends on the geographical origin and the production, processing and/or preparation takes place in a defined geographic area. To get an IGP then, at least one phase of the production process must take place in a specific area.
With these systems consumers receive valuable information on the quality and geographical origin of the products.
The UK Food Standards Agency (FSA) findings clearly demonstrated that “country of origin labelling” was “high on the consumers list of demands for change” (Food Standards Agency, 2001). The reasons vary from patriotism; specific culinary, organoleptic qualities or health benefits associated with regional products; a decreased confidence in the quality and safety of foods produced outside their local region or country; concern about animal welfare and “environmentally friendly” production methods (Kelly et al., 2005).
Geographical effects arise due to differences in the geological origin of the soils, soil pH, anthropogenic contaminants, atmospheric and climatic differences, and the interaction among certain trace elements.
Unfortunately, less expensive ingredients or components of dubious geographical origin may be fraudulently included for monetary gain in a food product.
A need exists to develop protocols enabling a foodstuffs geographic origin to be assessed. Techniques can be used to ‘‘fingerprint’’ the geographic origins of certain plant and animal materials; and these methodologies can form part of a suite of traceability tests (Polychroniadou et al., 1985).
The agro-food piracy can be viewed as a type of counterfeiting consisting in assign, to a food product, the name of another food product known for its organoleptic and/or safety and origin, despite being different.
Italy, thanks to its geographical position and climate, was able to obtain, over the centuries, an extraordinary amount of food with a high nutritional value and unique organoleptic characteristics, flavour, taste and colour. The brand “Made in Italy” is in fact very well known around the world. Many of these products have got the certification of quality by the European Union through the recognition of PDO, PGI and TSG (Traditional Speciality Guaranteed). And these products, that are the excellent expression of Italian food, are "hacked" in foreign markets, such as USA, Australia, New Zealand and China.
The agro-food piracy, in other words, is an illegal falsification of protected geographical indications and protected designations, exploiting quality, reputation and economic value of Italian food products, resulting in a fraud for consumers and in economic damage.
The agro-food piracy is a phenomenon which has causes and implications of different nature: Italian foods are highly appreciated by consumers around the world,

the food industry has a very important export, there are industries that are able to produce some "copy" of foods that, in some cases, it is not easy to distinguish a foreign product from the authentically Italian.
The ability to copy original Italian products is due to several reasons.
First of all, many Italians emigrated abroad bringing the "secrets" of traditional food, and in addition some Italian food companies have delocalised the production to countries where there is a wide availability of raw materials and the cost of work is significantly lower.
Sometimes these products are also imported in Italy with ambiguous labels. But the biggest damage is for exports to international markets because consumers can find food products very similar to the Italian but with different prices.
On the other hand, in the agro-food sector, the term "Italian sounding" is used to indicate those food items that "sound" from Italy, meaning they have a mix of Italian names, logos, images, slogans and packaging clearly attributable to Italy. In this case it is not like food piracy but rather imitations, low-cost copies of food products. Italian sounding is, therefore, synonym for products that "recall" Italian ones by the use of names and labels.
Beside frauds that impact on human health such as those that arise from chemical and/or microbiological contaminations, food frauds that do not reduce food security also represent a serious economic problem for companies and, more in general, for the Italian food industry, particularly for products with high production standards and quality certification.
Today, controls against food piracy are carried out by agencies, which are supported by Consortia of certified quality products. However, governamental actions are needed to put under control this phenomenon, which has enormous proportions, and never stops as it is alimented by profit.

1.3 Solanum lycopersicum

Tomato is the edible red fruit/berry of the *Solanum lycopersicum* (*Solanacee* family), plant.
The species originated in the South America and was spread throughout the world following the Spanish colonization of the Americas (Rick, 1976).
After the colonization, Spanish people distributed the tomato throughout their colonies in the Caribbean, Philippines, South Asia and in the entire Asian continent. The conquistador Hernán Cortés may have been the first to transfer the small yellow tomato from Mexico in 1521, although Christopher Columbus, may have taken them back in Europe as early as 1493.
Tomato grew easily in Mediterranean climates (Soressi 1969; Esquinas-Alcazar et al., 1995; García-Martínez et al. 2006), and cultivation began in the 1540s.
At the beginning, tomatoes were erroneously thought to be poisonous (although the leaves are poisonous) by Europeans who were suspicious of their bright, shiny fruit. Native versions were small, like cherry tomatoes, and most likely yellow rather than red. It was certainly being used as food by the early 17th century in Spain.
Tomato is now grown worldwide for its edible fruits, with thousands of cultivars having been selected with varying fruit types, and for optimum growth in differing growing conditions.
Tomato is commonly classified as determinate or indeterminate. Determinate, or bush, types bear a full crop all at once and top off at a specific height. Commercial growers who wish to harvest a whole field at one time prefer determinate types.

Indeterminate varieties develop into vines that never top off and continue producing until killed by frost. They are preferred by home growers and local-market farmers who want ripe fruit throughout the season.
As an intermediate form, there are plants sometimes known as vigorous determinate or semi-determinate; these top off like determinates, but produce a second crop after the initial crop.
Tomato plants are vines, initially decumbent, typically growing six feet or more above the ground if supported, although erect bush varieties have been bred, generally three feet tall or shorter. Indeterminate types are "tender" perennials, dying annually in temperate climates (they are originally native to tropical highlands), although they can live up to three years in a greenhouse in some cases. Determinate types are annual in all climates.
Tomato plants are dicots, and grow as a series of branching stems, with a terminal bud at the tip that does the actual growing.
The leaves are 10–25 cm long, odd pinnate, with five to 9 leaflets on petioles, each leaflet up to 8 cm long, with a serrated margin; both the stem and leaves are densely glandular-hairy.
The flowers, appearing on the apical meristem, have the anthers fused along the edges, forming a column surrounding the pistil's style. Flowers in domestic cultivars tend to be self-fertilizing. The flowers are 1–2 cm across, yellow, with five pointed lobes on the corolla; they are borne in a cyme of three to 12 together.
Tomato fruit is classified as a berry. As a true fruit, it develops from the ovary of the plant after fertilization, its flesh comprising the pericarp walls. The fruit contains hollow spaces full of seeds and moisture, called locular cavities. These vary, among cultivated species, according to type.
Cultivated tomatoes vary in size, from *tomberries*, about 5 mm in diameter, through *cherry* tomatoes, about the same 1–2 cm size as the wild tomato, up to *beefsteak* tomatoes 10 cm or more in diameter. The most widely grown commercial tomatoes tend to be in the 5–6 cm diameter range. Most cultivars produce red fruit, but a number of cultivars with yellow, orange, pink, purple, green, black, or white fruit are also available. Tomatoes grown for canning and sauces are often elongated, 7–9 cm long and 4–5 cm diameter; they are known as plum tomatoes, and have a lower water content.
Including *Solanum lycopersicum*, there are currently 13 species recognized in Solanum section Lycopersicon. Three of these species - *S. cheesmaniae*, *S. galapagense*, and *S. pimpinellifolium* - are fully cross compatible with domestic tomato. Four more species - *S. chmielewskii*, *S. habrochaites*, *S. neorickii*, and *S. pennelli* - can be readily crossed with domestic tomato, with some limitations. Five species - *S. arcanum*, *S. chilense*, *S. corneliomulleri*, *S. huaylasense*, and *S. peruvianum* - can be crossed with domestic tomato with difficulty (Ranc et al., 2008).
In the wild original state, tomatoes required cross-pollination (Rick et al., 1977); they were much more self-incompatible than domestic cultivars. As a floral device to reduce selfing, the pistil of wild tomatoes extends farther out of the flower than today's cultivars. The stamens were, and remain, entirely within the closed corolla (Chen et al., 2007).
When tomatoes were moved from their native areas, their traditional pollinators, (probably a species of bee) did not move with them. The trait of self-fertility became an advantage, and domestic cultivars of tomato have been selected to maximize this trait (Sharma, 2012).

Cultivated tomato has a limited genetic variability that is a direct consequence of the high level of self-pollination and different natural and artificial selection events that have occurred during the domestication and spread of modern cultivars (Rick, 1976).
First tomatoes introduced in Europe provided the entire genetic base of modern cultivars and consequently the current variety on the national and international market, are highly similar. It has been estimated that within the genus S. lycopersicum only the 5% of the total genetic variation from Solanum is present (Miller et al., 1990; Rick et al., 1975). The related wild species, however, are a rich source of genes and characteristics suitable for genetic improvement.
Breeding of new cultivars to get the best features, started more than 200 years ago in Europe, and especially in Italy (Stevens et al., 1986; Tigchelaar, 1986). By the beginning of the 20th century, tomato breeding yielded dozens of inbreds, known today as heirloom lines, displaying variation mostly in fruit shape, size, and color properties. Despite the differences in fruit traits of heirloom lines, they vary only in few genes, such as the fruit shape genes OVATE, FASCIATED, LOCULE NUMBER, and SUN (Rodriguez et al., 2011), and the fruit colour genes PSY1 (Fray, et al., 1993), beta-carotene (Ronen et al., 2000), Delta (Ronen et al., 1999), and HIGH-PIGMENT (Lieberman et al., 2004).
Tomato breeding has played a major role in developing varieties adapted to the new agricultural and processing technologies. An example is the development of cultivars suited for mechanical harvesting which are characterized by a determinate growth habit, concentrated fruit set and firm flesh (Gould, 1992). One of the major breeding objectives for tomatoes destined to industrial uses is to increase total soluble solids content (TSS or brix; mainly sugars and acids) of the fruits. Varieties with high brix values are more efficient for the production of the various concentrated products.
Other characters employed for the genetic improvement, were the growth habit and characteristics of the berry (size, shape, colour and morphology) (Bai et al., 2007). Another important objective has been the increase of productivity and the addiction of resistances to different diseases. In fact, one of the main limitations for the cultivation of tomato is the damage caused by pathogens, such as viruses, bacteria and fungi. Their control involves three main strategies that are cultural systems, the application of chemicals substances and the use of resistant varieties. In tomato, more than 20 major genes for disease resistance have been reported, which are used in tomato improvement and developed cultivars possessing multiple disease-resistance attributes (Kalloo, 1991).

1.3.1 The economic role

The great importance of the tomato cultivation is due to the use of berries for the greatly appreciated organoleptic characteristics, which are consumed both fresh and processed because of the different commercial ranges and its availability all year thanks to agricultural techniques and transformation methods of berries.
Tomato, for quantity, is the eleventh among commodities produced world-wide after sugar can, maize, rice, wheat, milk, potatoes, sugar beet, vegetables, soybeans, and cassava, with a yield average of 37 tons/ha (FAOSTAT 2012; http://faostat.fao.org).
About 160 million tons of tomatoes were produced in the world in 2012. China, the largest producer, accounted for more than one quarter of the global output, followed

by India and United States. Italy, with around 5 million of tons, is the seventh producer country (FAOSTAT 2012; http://faostat.fao.org).
In Italy, tomato is in the fourth place, after milk, wheat and grapes for production and after grapes, milk, olives and pig-meat for economic value. Northern Italy represent the 54% of all fields dedicated to the cultivation of tomatoes, the 41% is in South Italy, while the rest is in the Centre.
In 2013 Italy marks a record in the value of agro-food exports for 34 billion euro due to the increase of 7% of exports (data Coldiretti).
Furthermore, two-thirds of the revenues are obtained with the export of food products to the EU countries (+6%), but the *Made in Italy* is strong even in the Americas (+7%), and emerging markets from Asia, Africa and Oceania.
Wine is the first exported product, followed by fresh fruit and vegetables. In particular, canned and peeled tomatoes are among the most exported food products from Italy. According to the agreement between the tomato producers associations and the industries for canned food, in 2014 industrial tomatoes will be sold at about 92 euro per tons (http://agronotizie.imagelinenetwork.com) demonstrating a new economic growth in the tomato market that in recent years had suffered a significant reduction in prices (8.4 euro per kilos in 2012 and 8.6 euro per kilos in 2013; http://www.agricolturanews.it).

1.3.2 San Marzano Tomato

The ‘San Marzano’ tomato from Agro-Nocerino Sarnese (PDO) is a variety recognized as Protected Designation of Origin.
‘San Marzano’ tomato is long, nervous, and consistent. It is the only one that will not shatter during the processing; it remains entire and alive in the can.
According the oral tradition, it is said that the first seed of this cultivar arrived in Italy in the 1770s, as a gift from the Kingdom of Perù to the Kingdom of Naples, and that would have been planted in the area that corresponds to the locality of San Marzano (Salerno). Since then, it derives the origin of this famous tomato that over time, with various actions of selection, has acquired the characteristics of the ecotype known all over the world.
The tomato ‘San Marzano’ assumed great appreciation at the beginning of 1900, when started the first processing company by Francesco Cirio, which produced the famous "peeled" tomato for sauce.
In the recent past, the ‘San Marzano’ tomato was also called "red gold" for its the economic value in the Agro-Nocerino Sarnese area. In the 80s the crop suffered a drastic reduction, both in terms of area and production, but the action for the recovery of pure genetic lines and their improvement has allowed the preservation and revitalization on an international basis. For the tomato ‘San Marzano’ is now a new season of life and is required not only in Europe and America, but also in other continents.
The intrinsic parameters that have made the ‘San Marzano’ tomato famous and widespread are the typically sweet and sour taste, elongated berries with parallel longitudinal depressions, bright red colour, few seeds and fibers, and easy to peel. These, together with the chemical-physical characteristics and the nutritional value, make it unique, both fresh and transformed. For these reasons, the ‘San Marzano’ tomato is widely known and commercialized around the world (Porretta 1995).

The protected designation of origin (PDO) "San Marzano Tomato Agro-Nocerino Sarnese" is reserved for tomatoes that satisfy conditions and requirements established by the rules of the procedure guideline (Disciplinare di produzione e trasformazione della Denominazione di Origine Protetta "Pomodoro San Marzano dell'agro sarnese-nocerino")
According to the procedure guideline, tomato lines obtained as a result of genetic improvement of 'San Marzano' ecotypes can be used to produce this tomato, provided that both the improvement and cultivation are carried out in the territory included in the procedure guideline and that plants and berries exhibit characteristics conform to the standards described in the same document.
The production area comprises the Agro-Sarnese Nocerino, with offshoots in some of the localities around Naples and Avellino.
The soils recognized by the procedure guideline appear very deep, soft, with a good amount of organic substances and a high amount of available phosphorus and exchangeable potassium. The hydrology of the area is very rich for the presence of numerous sources of water and abundant groundwater at different depths. The water for irrigation generally is derived from wells that feed directly from underground.
The climate of the area affected by the beneficent influence of the sea. The prevailing winds are hot, rains are abundant in the fall, winter and spring and although the rains are absent during the summer months, the relative humidity remains quite high.
Transplant normally occur in the first two weeks of April, however, can last until the first week of May.
According to the procedure guideline, the harvest of the fruits must be done exclusively by hand, so climb when they reach full maturity, and takes place in several stages.
After the assignment of the Protected Designation of Origin (PDO) in 1996 by the European Union, the San Marzano Consortium was established in June 1999 (http://www.consorziopomodorosanmarzanodop.it).
The San Marzano Consortium aims to achieve an action for economic development of the 'San Marzano' production areas, involving economic, social and institutional operators. Specifically, it wants: provide for the keeping of registers for producers and processors of the product that may use the PDO; develop and distribute improved cultivation techniques; improve the economic income of agricultural producers.
The verification of compliance with the procedure guideline along the pathway of the 'San Marzano' berry production is carry out by IS.ME.CERT (Istituto Mediterraneo di Certificazione Agroalimentare) that is involved in the control of the agro-quality
Although 'San Marzano' tomato has a gastronomically superior fruits very suitable for processing and organoleptic and chemical-physical parameters better than other tomato varieties, it also has undesirable agronomic traits such as the lack of genetic resistance against pathogens and a yield lower than modern hybrid cultivars (Monti et al. 2004).
Some of the phenotypic characteristics of 'San Marzano' plants such as the indeterminate growth habits and the susceptibility to major pathogens, increase its cost of production that is higher in respect to the costs for hybrid production. As a result 'San Marzano' market price is higher (40 euro-cent per kilos, http://www.consorziopomodorosanmarzanodop.it/) than hybrid berries with similar commercial destination (9.2 euro-cent per kilos, http://agronotizie.imagelinenetwork.com).

1.4 Whole Tomato Genome

The tomato sequencing project was initiated in 2004 by a consortium of 10 countries, with each of the following countries sequencing one chromosome: Korea (chromosome 2), China (chromosome 3), Great Britain (chromosome 4), India (chromosome 5), The Netherlands (chromosome 6), France (chromosome 7), Japan (chromosome 8), Spain (chromosome 9), and Italy (chromosome 12). The United States sequenced three chromosomes (chromosomes 1, 10, and 11).
The accession chosen for the genome sequencing was 'Heinz 1706', a processing inbred cultivar, which has a number of known introgressions from wild relatives, including resistance genes for Fusarium and Verticillium wilt (Ozminkowski, 2004).
The complete sequence of the plastidial genome (the plastome, corresponding now to chromosome 0) was available from 2006 (Kahlau et al., 2006).
Initially, the approach was sequencing a BAC tiling path of the euchromatin, which contains more than 90% of the genes, but spans less than 25% of the ≈900Mb genome (Mueller et al., 2009). Scaffolds were linked with two bacterial artificial chromosome (BAC)-based and anchored/oriented using a high-density genetic map, introgression line mapping and BAC fluorescence in situ hybridization (FISH).
In 2009, with more than 1,200 BACs sequenced, whole-genome 454 sequencing was used to the BAC sequences, and provides higher coverage for assembling the entire genome.
The release of the complete tomato genome sequence was in 2012 (Tomato Genome Consortium). The genome size is approximately 930 megabases (Mb), consistent with previous estimates (Michaelson et al., 1991) of which 760 Mb were assembled in 91 scaffolds aligned to the 12 tomato chromosomes.
For the tomato genome, 34.727 protein-coding gene were found, of which 30.855 are supported by RNA sequencing (RNA-Seq) data and 31.741 show high similarity to *Arabidopsis* genes. Furthermore, chromosomal organization of genes, transcripts, repeats and small RNAs (sRNAs) are very similar between tomato and potato (Tomato Genome Consortium, 2012).
The official functional annotation for the tomato genome is provided by the International Tomato Annotation Group (ITAG), a multinational consortium funded in part by the EU-SOL project, reporting 19.662 (56.6%) genes associated to GO (Gene Ontology) terms describing their functions and 2,108 unique GO terms (ftp://ftp.sgn.cornell.edu/genomes/Solanum_lycopersicum/annotation/).
The tomato sequencing consortium also released a draft genome for *S. pimpinellifolium*, and determined an estimated divergence of 0.6% compared to the 'Heinz 1706' reference genome.

1.5 DNA Markers

A molecular marker can highlight differences between two or more genomes, in terms of length, presence, absence or sequence of a known DNA trait. The recurrence of certain characteristics in different genomes can be used to define the degree of relatedness between groups of organisms.
Thanks to the recent advancements in molecular biology, DNA markers have become the most effective and rapid instrument in the analysis of the DNA of plant cultivars and animal breeds, and are also used to track the raw materials in food

industry processes (Kumar et al., 2009; Mafra et al., 2008; Woolfe et al., 2004; Caramante et al., 2010).
Usually, traditional DNA-based methods use specific DNA sequences as markers and can be divided in hybridization-based markers and Polymerase Chain Reaction (PCR)-based markers. In hybridization-based methods, species-specific DNA profiles are discovered by hybridizing DNA digested by restriction enzymes, and comparing it with labelled probes (DNA fragments of known origin or sequence), i.e. restriction fragment length polymorphisms (RFLPs). Instead, PCR-based methods involve the amplification of target loci by using specific or arbitrary primers, and a DNA polymerase enzyme. Fragments are then separated by electrophoresis and banding patterns are detected by different methods. PCR-based markers are: random amplification of polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), inter simple sequence repeats (ISSRs), microsatellites or simple sequence repeats (SSRs), and single nucleotide polymorphisms (SNPs). PCR-based methods are extremely sensitive, often faster than other technologies, and are widely used in agriculture, zootechny and food control (Grassi et al., 2006; Labra et al., 2004; Mane et al., 2006; Caramante et al., 2010; Pasqualone et al. 2007; Sonnante et al., 2009; Corrado et al., 2011).
On the other hand, a different approach utilize the combinations of variable non-coding and relatively conserved coding regions of the plastid genome and it have been proposed as new tool for species discrimination and discovery (Chase et al., 2007; Fazekas et al., 2009; Kress et al. 2009; Burgess et al., 2011). The DNA barcoding (www. barcoding.si.edu) is based on the designation of mtDNA fragment to act as a ‘‘barcode” to identify and delineate species. This approach was successfully employed in animals using a portion of the cytochrome oxidase 1 (cox1 or CO1) mitochondrial gene. Hebert et al. (2003) proposed that a single gene sequence would be sufficient to differentiate animal species, and proposed the use of the mitochondrial DNA gene cytochrome oxidase subunit I as a global bio-identification system for animals. Empirical support for the barcoding concept ranges from studies of invertebrates to birds (Hebert et al. 2004), but also of fish species, such as tuna (Terol et al., 2002), flatfish (Espineira et al., 2008), anchovy (Jérôme et al., 2008) and sharks (Barbuto et al., 2010). For plant barcode the majority preference is represented by a core-barcode, consisting of portions of two plastid-coding regions, rbcL+matK, to be supplemented with additional markers as required (CBOL Plant Working Group, 2009). The choice of rbcL+matK as a core barcode was based on the straightforward recovery of the rbcL region and the discriminatory power of the matK region (Hollingsworth et al., 2011).
From the first observations of Gregor Mendel, through the experiments of Thomas H. Morgan, up to the most modern technologies able to detect polymorphisms in the genome, DNA molecular markers have always accompanied the more ambitious studies of intra-species classification, individual characterization and genetic population structure.

1.5.1 SSR

Microsatellites are also known as Short Tandem Repeats (STRs, Edwards et al. 1991), Simple sequence repeats (SSRs, Jacob et al. 1991) or Simple Sequence Length Polymorphism (SSLP, Tautz 1989).

Hamada and colleagues (1982) demonstrated the existence of microsatellites in various eukaryotes, ranging from yeasts to vertebrates, and other studies confirmed the abundance of microsatellites in plants and in many other eukaryotes (Delseny et al., 1983; Tautz and Renz, 1984).
In particular, plants are rich in AT repeats, whereas in animals AC repeat is the most common and this seems to be the feature to discriminate plant and animal genomes (Powell et al., 1996).
SSRs are present in both coding and noncoding regions and are distributed throughout both the nuclear and plastidial genome (Provan et al., 2001; Chung et al., 2006; Soranzo et al., 1999; Rajendrakumar et al., 2007).
SSRs are composed by tandem repeated motifs of 2–6 bp, representing the *core*, and are characterized by a high degree of length polymorphism (Zane et al. 2002). The high length polymorphism is due to different number of repeats in the microsatellite regions, therefore they can be easily and reproducibly detected by polymerase chain reaction (PCR). Length polymorphism of the *core* is originated *in vivo* by the slippage of the DNA polymerase during replication and subsequent unequal crossing-over between homologous chromosomes during meiosis (Levinson and Gutman, 1987).
Differences in SSR alleles length are not always easily detected on agarose gel. An increase of the resolution in the electrophoretic separation is obtainable by polyacrylamide gel in combination with radioactive labeling or silver nitrate staining (Scrimshaw, 1992). But, the use of fluorescent primer in combination with the use of capillary electrophoresis has been shown to be a more efficient solution than traditional methods (Schwengel et al., 1994).
Microsatellites have been variously classified depending on their size, type of repeat unit and its location in the genome. Depending upon the number of nucleotides per repeat unit, SSRs have been classified as mono-, di-, tri-, tetra-, penta- or hexa-nucleotides. Several classifications have been attributed to the SSR because of the type of core: perfect, imperfect and compound microsatellites (Weber, 1990) or also simple imperfect, compound perfect or compound imperfect (Wang et al., 2009). Perfect repeats are tandem arrays of a single repeat motif, while in imperfect repeats perfect repeats are interrupted by non-repeat motifs. In compound microsatellites, two basic repeat motifs are present together in various configurations.
Based on their location in the genome, microsatellites can be classified as nuclear (nuSSR), mitochondrial (mtSSR) or chloroplastic SSRs (cpSSR) (Weising and Gardner, 1999; Powell et al., 1996; Soranzo et al., 1999; Provan et al., 2001).
SSR markers have gained considerable importance in plant genetics and breeding thanks to many desirable attributes including hypervariability, multiallelic nature, codominant, reproducibility, relative abundance, extensive genome coverage (including organellar genomes), chromosome specific location, possible automation and high throughput genotyping (Parida et al., 2009).
These markers have proven to be useful tool for paternity analysis, construction of high-density genome maps, mapping of genes, marker-assisted selection, and for establishing genetic and evolutionary relationships (Parida et al., 2009) and were also successfully employed in individual discrimination among species and identification of food components, both in animal and in plant kingdom (Orrù et al., 2006; Arana et al., 2002; Sardaro et al., 2013; Caramante et al., 2010; Turci et al., 2010; Corrado et al., 2011; Pereira et al., 2012).

1.6 Next Generation Sequencing

About 40 years ago, Maxam and Gilbert reported a novel approach for DNA sequencing in which terminally labeled DNA fragments were subjected to base-specific chemical cleavage and the reaction products were separated by gel electrophoresis (Maxam and Gilbert, 1977). In the same year, Frederick Sanger and colleagues described the use of chain-terminating dideoxynucleotide analogs that caused base-specific termination of primed DNA synthesis to reach the same goal (Sanger et al., 1977).
Over the next 30 years, the automated Sanger method had been the dominant approach and led to a number of important activities, including the completion of the only finished-grade human genome sequence (International Human Genome Consortium, 2004), a 13-year effort with an estimated cost of 2.7 billion of dollars.
The commercial launch of the first massively parallel pyrosequencing platform in 2005 conducted in the new era of high-throughput genomic analysis, the next-generation sequencing (NGS).
Already in 2008, the sequencing of a human genome needed 5-month and approximately $1.5 million by massive parallel sequencing (Wheeler et al., 2008), up to the Revolutionary Genome Sequencing Technologies program which had as goal the sequencing of a human genome for $1000 or less (http://www.genome.gov/27527585).
Starting ten years ago, a variety of massively parallel sequencing instruments, such as Genome Sequencer from Roche 454 Life Sciences (www.454.com), the Solexa Genome Analyzer from Illumina (www.illumina.com), the SOLiD System from Applied Biosystems (www.appliedbiosystems.com), the Heliscope from Helicos (www.helicos.com) and the commercialized Polonator (www.polonator.org), which were largely different from the Sanger-based capillary, were used to sequence the human and model organism genomes.
These second-generation machines are characterized by highly parallel operation, higher yield, simpler process, much lower cost per read, and production of shorter reads.
Although each instrument is different from the others, all massively parallel sequences machines share some common features. First, the initial preparatory steps are reduced and simplified. Second, amplification of the library fragments is needed for all platforms. Third, sequencing reactions are performed and detected automatically (Mardis, 2011).
The amplification of the library fragments by PCR, typically involve multiple primer pairs in a mixture that are combined with genomic DNA of interest. The use of multiplex primer pairs couples the high throughput of NGS platforms and the fact that each sequence read represents a single DNA product in the mixture due to the nature of the sequencing platforms (Mardis, 2013). Following the PCR, the resulting fragments have platform-specific adapters ligated to their ends to form a library that is suitable for sequencing.
Accuracy in NGS is achieved by sequencing a given region multiple times, enabled by the massively parallel process, with each sequence contributing to “coverage” depth.
The main processing feature of the data analysis is the computationally intensive conversion of image data into sequence reads, known as base calling. Image parameters such as intensity, background, and noise are used in a platform-

dependant algorithm to generate read sequences and error probability–related quality scores for each base.
The quality values calculated during NGS base calling provide important information for alignment, assembly, and variant analysis. Although the calculation of quality is different between platforms, the calculations are all related to the historically relevant Phred score, introduced in 1998 for Sanger sequence data (Ewing et al., 1998a; Ewing et al., 1998b). The Phred score quality value, *q*, uses a mathematical scale to convert the estimated probability of an incorrect call, *e*, to a log scale: $q=-10 * \log_{10}(e)$. Miscall probabilities of 0.1 (10%), 0.01 (1%), and 0.001 (0.1%) yield Phred scores of 10, 20, and 30, respectively.
Quality score resulting to be an important parameter for eliminating low-quality reads, trimming low-quality bases, improving alignment accuracy, and determining consensus-sequence and variant calls (Li et al., 2008).
However, alignment and assembly remain substantially difficult procedures because of the shorter reads lengths.
In general, obtained reads are used to create an assembly, a hierarchical data structure that maps the sequence data to a putative reconstruction of the target. It groups reads into contigs and contigs into scaffolds. Contigs provide a multiple sequence alignment of reads plus the consensus sequence. The scaffolds, sometimes called supercontigs or metacontigs, define the contig order and orientation and the sizes of the gaps between contigs.
NGS experiments generate volumes of data, which present challenges and opportunities for data management, storage, and, most importantly, analysis (Pop et al., 2008).
The high-throughput capacity of NGS has allowed to sequence entire genomes, from microbes to humans (Wheeler et al., 2008; Margulies et al., 2005; Pearsonet al., 20078; Smith et al., 2008; Quinn et al., 2008; Satkoski et al., 2008; Borneman et al., 200; Wang et al., 2008), has been used to map genomic structural variation, including deletions, insertions and rearrangements (Korbel et al., 2007; Campbell et al., 2008; Kim et al., 2008; Chen et al., 2008), is used to identify polymorphisms and mutations in genes implicated in cancer and in regions of the human genome implicated in disease (Yeager et al., 2008; Ding et al., 2008) and in metagenomic studies including analysis of microbial populations in ocean (Huber et al., 2007; Sogin et al., 2006) and soil (Urich et al., 2008).

1.6.1 RNA-seq

The physiological and biochemical changes of an organism reflect the transcriptional modulation of many genes.
The transcriptome is the complete set of transcripts, and their quantity, in a cell. Understanding the transcriptome is essential for interpreting the functional elements of the genome and revealing the molecular constituents of cells and tissues, and also for understanding development and disease.
Initially, Sanger sequencing of cDNA or EST libraries (Boguski et al., 1994; Gerhard et al., 2004) was used, but this approach is relatively low throughput, expensive and generally not quantitative.
The new NGS-strategy provided a powerful approach, termed RNA-Seq (RNA sequencing) for mapping and quantifying transcripts in biological samples, already applied to *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis*

thaliana, mouse and human (Nagalakshmi et al., 2008; Wilhelm et al., 2008; Mortazavi et al., 2008; Lister et al., 2008; Cloonan et al., 2008; Marioni et al., 2008; Morin et al., 2008).
In general, RNA-seq technology include the analysis of a population of RNA (total or fractionated, such as poly(A)+) converted to a library of cDNA fragments with adaptors attached to one or both ends. Each molecule, with or without amplification, is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing).
After sequencing, reads are aligned to a reference genome and compared with known transcript sequences, or assembled *de novo* to construct a genome-scale transcription map that consists of both the transcriptional structure and level of expression for each gene (Wang et al., 2009).
Expression levels are deduced from the total number of reads that map to the exons of a gene and normalized by the length of exons.
The most used normalization methods are RPKM/FPKM (reads or paired-end fragments per kilobase of exon model per million mapped reads) (Mortazavi et al., 2008), TMM (trimmed mean of M-values) (Robinson et al., 2010) and CPM (count per million) (Robinson et al., 2010).
The RNA-seq strategy offers several advantages in transcriptome analysis: is not limited in detection of transcripts for which the genomic sequence is know and this makes RNA-seq particularly attractive for non-model organisms that have no genomic sequences determined (Vera et al., 2008); short reads give information about how two exons are connected, while longer reads or pair-end short reads should reveal connectivity between multiple exons connection; of course, RNA-seq give an accurate view about gene expression levels (Marioni et al., 2008); finally, RNA-seq can also reveal sequence variations, for example, SNPs, isoform and novel splice junctions, and information of antisense regulation and intragenic expression (Cloonan et al., 2008; Morin et al., 2008; Carninci et al., 2005; Nagalakshmi et al., 2008; Graveley et al., 2010; Trapnell et al., 2010).
The unprecedented level of sensitivity and the high throughput of deep sequencing technologies suggest that RNA-seq is became the platform of choice for transcriptome for comprehensive studies of gene expression, differential splicing activity, and discovery of expressed SNPs.

1.7 Aim of the work

The aim of this research activity is to test the power of DNA and transcriptomic markers in the genetic authentication and geographic specificity of primary product in tomato food chain. This project is divided into several phases.
First of all, SSR molecular markers were employed for fingerprinting of about 150 tomato genotypes widespread in Italy, and in particular in Campania region. The aim has been to characterize and discriminate tomato varieties belonging to different types and destined to several uses, such as local ecotypes, hybrid and non-hybrid varieties, widely used in tomato chain for both fresh and processed products.
In the context of molecular fingerprinting, forty tomato ecotypes belonging to the ‘San Marzano’ variety have been characterized. This variety has a significant economic value in the local economy of the Campania region but also in the national and international tomato market. The ‘San Marzano’ tomato is a product with protected designation of origin and the PDO label implements the value of this product. Characterization through SSR markers allows to know the molecular profile of

ecotypes belonging to the ‘San Marzano’ variety useful in the assessment of the homogeneity of the ecotypes used for production of PDO ‘San Marzano’ products and allows to understand and study the genetic relationships within local ecotypes from the same production area and between genotypes with very similar or completely different features.
The molecular fingerprint becomes a tool for allocating the genetic origin of food products spread on the market walking in direction of traceability for tomato food chain. In particular, this work helps in protection of quality label attributed to foodstuffs, such as ‘San Marzano’ PDO.
In effect, in this study several tomato processed products, labeled as ‘San Marzano’ PDO, were analyzed in order to verify the correspondence between what is declared on the label and the content of the cans.
Another point of interest for traceability of primary products in food chains is the identification of production areas of food products.
Many food products owe their fame to their geographical location. The combination of soil, climate, flora and fauna with the cultivation techniques and processing from a specific place give to products organoleptic and nutritional properties often superior in compare with other products of the same category.
Once again the ‘San Marzano’ tomato has a central role: in fact, its cultivation is closely linked to a production area and the PDO logo, therefore, is not only synonymous of ‘San Marzano’ "genotype" but also of recognition of production locations and agricultural techniques.
Because this work is centred in the use of molecular techniques able to certify the raw materials in the tomato supply chain, a great interest in developing new tools for the identification of the geographical origin was born.
If at the genomic level markers are able to discriminate genotypes, at the transcriptomic level markers may prove to be a valuable aid in the discrimination of the cultivation environments of the same genotype. Therefore, in this work one of the goals is searching transcriptomic markers associated with specific production areas through the observation of gene expression of tomato genotypes grown in different environments.
The consolidation of the study of DNA markers for cultivar identification and the development of transcriptomic markers environment-specific can be helpful tools for traceability in tomato supply chain with the ultimate aim of countering and combating fraud regarding the genetic and/or geographic origin in order to protect and authenticate primary food products.

2. MATERIALS AND METHODS

2.1 Plant Material

2.1.1 Evaluation of Genetic Variability

Plant material used for the characterization by DNA microsatellites includes 146 tomato genotypes. This collection is composed by local ecotypes and varieties from Campania region (Italy) and hybrid cultivar for industrial use. Fifty-five genotypes, out of 145, are included in the project named “SALVE” (Salvaguardia della Biodiversità Vegetale della Campania) and represent part of the genetic variability available in Campania.
Forty ecotypes are related to the genotypes authorized by the “Disciplinare di produzione della Denominazione di Origine Protetta - Pomodoro San Marzano dell’Agro Sarnese-Nocerino" for the production of tomato products with PDO label. In particular ‘Kiros’ genotype and ‘San Marzano 2’ varieties are declared into the procedure guideline.
Industrial cultivars used in tomato food chain were mainly selected for their similarity to ‘San Marzano’ variety for shape and industrial destination.
For all samples leaves from three different plants were analyzed.
The list of tomato genotypes is reported in table 2.1.

Table 2.1: Name of the tomato genotypes, identification code, and hybrid material.

Genotype	Code	Hybrid	Germplasm Bank	Genotype	Code	Hybrid	Germplasm Bank
16 SMEC 2.1.1 (SM2*)	16 SM	no	CRA-A	Pisanello	PS1	no	Internal
17 SMEC 2.2.1.1 (SM4)	17 SM	no	CRA-A	Player	PLA	yes	Annalisa-Lodato
18 SMEC 1.1.1 (SM4)	18 SM A	no	CRA-A	Podium	PO	yes	Annalisa-Lodato
18 SMEC 1.2.1 (SM4)	18 SM B	no	CRA-A	Pomodorino giallo di Castel San Lorenzo	SLOR	no	CRA-ORT
19 SMEC 1.1.1.1 (SM4)	19 SM A	no	CRA-A	Pomodorino giallo di San Bartolomeo	S.BART	no	CRA-A
19 SMEC 1.2.1.1 (SM4)	19 SM B	no	CRA-A	Pomodorino giallo Beneventano	PDGB2	no	CRA-A
20 SMEC (SM2*)	20 SM	no	CRA-A	Pomodorino giallo di Camposano	GC1	no	CRA-A
20 SMEC 3 (SM2*)	20 SM 3	no	CRA-A	Pomodorino giallo di Camposano	GC3	no	CRA-A
21 SMEC 1.1.1 (SM4)	21 SM	no	CRA-A	Pomodorino giallo di Camposano	GC4	no	CRA-A
22 SMEC 2.2.1 (SM4)	22 SM	no	CRA-A	Pomodorino giallo di Cerreto	CER	no	CRA-A
24 SMEC 2.1.1.1 (SM4)	24 SM A	no	CRA-A	Pomodorino giallo di Vico Equense	EQ	no	CRA-A
24 SMEC 2.1.2.1 (SM4)	24 SM B	no	CRA-A	Pomodorino Agostino	AG	no	CRA-A
24 SMEC 2.2.1.1 (SM4)	24 SM C	no	CRA-A	Pomodorino di collina	PC	no	CRA-A
25 SMEC 1.2.1.1 (SM4)	25 SM	no	CRA-A	Pomodorino giallo del Beneventano	PDGB	no	CRA-A
26 SMEC 2.1.1.1 (SM4)	26 SM A	no	CRA-A	Pomodorino giallo di Montecalvo	MONT	no	CRA-A
26 SMEC 2.2.1.1 (SM4)	26 SM B	no	CRA-A	Pomodorino giallo di Visciano	PDG2	no	CRA-A
26 SMEC 2.2.2.1 (SM4)	26 SM C	no	CRA-A	Pomodorino Reginella	REG	no	CRA-A
27 SMEC 2.1.1.1 (SM4)	27 SM	no	CRA-A	Pomodorino Riccia San Vito	RSV6	no	CRA-A
29 SMEC 1.1.1.1 (SM4)	29 SM A	no	CRA-A	Pomodorino Riccia San Vito	RSV7	no	CRA-A
29 SMEC 1.1.2.1 (SM4)	29 SM B	no	CRA-A	Pomodorino rosa di Rofrano	PRF	no	CRA-A
30 SMEC 1.1.1.1 (SM4)	30 SM	no	CRA-A	Pomodorino rosso di Roccadaspide	RDA	no	CRA-ORT
31 SMEC 1.1.1.1 (SM4)	31 SM A	no	CRA-A	Pomodorino rosso selvatico	SEL	no	CRA-A
31 SMEC 1.1.2.1 (SM4)	31 SM B	no	CRA-A	Pomodoro San Marzano (SM)	SM	no	CRA-A
32 SMEC 1.1.1.1 (SM4)	32 SM A	no	CRA-A	Principe Borghese1	PB1	no	Semiorto
32 SMEC 1.1.2.1 (SM4)	32 SM B	no	CRA-A	Principe Borghese	PB	no	CRA-A
33 SMEC 1.1.1.1 (SM2*)	33 SM A	no	CRA-A	Principe Borghese Determinato	PBD1	no	Semiorto
33 SMEC 1.1.2.1 (SM2*)	33 SM B	no	CRA-A	PS 1296	PS1296	yes	Annalisa-Lodato
35 SMEC 1.1.1.1 (SM4)	35 SM	no	CRA-A	Ps1398	PS1398	yes	Seminis
37 SMEC 1 2.1.1 (SM4)	37 SM C	no	CRA-A	PS513	PS513	yes	Annalisa-Lodato
37 SMEC 1.1.1.1 (SM4)	37 SM A	no	CRA-A	Pummarola Riccia	PRI	no	CRA-ORT
37 SMEC 1.1.2.1 (SM4)	37 SM B	no	CRA-A	Pummarola scritta	PSC	no	CRA-ORT
A sole	SOL	no	CRA-ORT	Quadrato rosso	QR	no	CRA-ORT
Agro/Nocerino 5	AN5	no	Internal	Quarantino grande	QG	no	CRA-A
Agro/Nocerino 7	AN7	no	Internal	Quarantino piccolo	QP	no	CRA-A
Arsicolo di San Gregorio	ARSG	no	CRA-ORT	Red Setter	RS	no	Internal
Auricchio	AUR	no	CRA-ORT	Regent	RE	yes	ISI SEMENTI
Auspicio	AU	yes	Annalisa-Lodato	Roma V. F. Semiorto Sementi	RVF	no	Semiorto
Cannellino flegreo	CF	no	CRA-A	Romarzano	ROM	no	Semiorto
Cento scocche	CS	no	CRA-A	Rosso a punta	RPNT	no	CRA-ORT
Cilindrico	CLD	no	CRA-ORT	Sala	SLA	no	CRA-ORT
Cirio 3	C3	no	Annalisa-Lodato	San Marzano Cilindrico III (SM)	SMC1	no	Semiorto
Cohiba	CO	yes	Seminis	San Marzano Cilindrico IV (SM)	SMC2	no	Semiorto
Corbarino	CRB_ORT	no	CRA-ORT	San Marzano Morini (SM)	SMMO1	no	Semiorto
Crovarese	CRO	no	CRA-ORT	San Marzano Murano (SM)	SMMU1	no	Semiorto
Datterino	DAT	no	Semiorto	San Marzano Nano (SM)	SMN	no	Semiorto
Decio	DEC	yes	Annalisa-Lodato	Scipio	SCI	yes	Annalisa-Lodato
Defende	DEF	yes	Annalisa-Lodato	Seccagno	SECC	no	CRA-A
Diaz	DIAZ	yes	Annalisa-Lodato	Seccagno Pizzutiello	SPIZ	no	CRA-A
Discovery	DIS	yes	Annalisa-Lodato	SM 246 (SM)	SM246	no	CRA-A
Elba	EL	yes	ISI SEMENTI	Smart	SMA	yes	ISI SEMENTI
Ercole	E	yes	Annalisa-Lodato	Sorrento	SORR	no	CRA-A
Galeon	GA	yes	Seminis	Sorrento 61	S61	no	Semiorto
Genius	GE	yes	Annalisa-Lodato	Sorrento 62	S62	no	Semiorto
Giallo Auletta	GAU	no	CRA-ORT	Sorrento 65	S65	no	Semiorto
Giallo oblungo	OB	no	CRA-ORT	Sorrento Globoso Rosato Indeterminato	SGRI1	no	Semiorto
Guardiolo	GU_ORT	no	CRA-A	Sorrento Gragnano	SG	no	Semiorto
Gulliver	GU	yes	Annalisa-Lodato	Sorrento Semiorto	SS	no	Semiorto
Herdon	HE	yes	Seminis	Sorrento Tondo Liscio Rosato	STLR(A)	no	Semiorto
Insalataro Auletta	AUL	no	CRA-ORT	Sorrento_2	S-ORT	no	CRA-ORT
Intero liscio di San Gregorio Magno	SGM	no	CRA-ORT	Suerte	SU	yes	Annalisa-Lodato
Jet	JET	yes	Annalisa-Lodato	Talent	TA	yes	Annalisa-Lodato
Kiros 07 (SM*)	K07	no	ARCA	Tomito	TO	yes	Annalisa-Lodato
Kiros 09 (SM*)	K09	no	ARCA	Tondino Determinato	TD1	no	Semiorto
Kiros 10 (SM*)	K10	no	ARCA	Tondino Indeterminato	TI1	no	Semiorto
Lampadina Sala	LMPS	no	CRA-ORT	Tondo giallo di Roccadaspide	GRD	no	CRA-ORT
Leader	L	yes	Annalisa-Lodato	Tondo Liscio Indeterminato	TLI1	no	Semiorto
Logan	LO	yes	Seminis	Tondo Sala	TS	no	CRA-ORT
Lungo giallo di Capaccio	LGC	no	CRA-ORT	UC-82	UC-82	no	Internal
M82	M82	no	Internal	Ventura Determinato	VD1	no	Semiorto
Minidor	MI	yes	Annalisa-Lodato	Vesuviano	VES	no	CRA-A
P446	P446	yes	Annalisa-Lodato	Vesuviano pomodorino o Piennolo Rosso	PDS	no	CRA-A
Piennolo Pollena	POLL	no	CRA-A	Vesuviano pomodorino o Piennolo Rosso_2	PDS-SEME	no	CRA-A
Piennolo vesuviano	PDV	no	CRA-A	Vulcan	VU	yes	Annalisa-Lodato

(SM) = San marzano; (SM2)= San Marzano 2; (SM4)= San Marzano 4.
* included in the "Disciplinare di Produzione della Denominazione di Origine Protetta - Pomodoro San Marzano dell'Agro Sarnese-Nocerino

2.1.2 Evaluation of Traceability in Tomato Food Chain

For this purpose 33 peeled tomato, labelled as 'San Marzano' and exhibiting the PDO label, produced by seven different Companies were analysed. The molecular analysis was conduced on single berries extracted from each can (Table 2.2).

Table 2.2: Companies, identification code and label of 33 peeled tomatoes used for traceability studies. Solania samples are classified for their production number.

Company	Code	PDO Label
Solania	L219_V921	yes
Solania	L224_V926	yes
Solania	L226_V925	yes
Solania	L229_V883	yes
Solania	L229_V887	yes
Solania	L231_V884	yes
Solania	L233_V890	yes
Solania	L234_V892	yes
Solania	L236_V924	yes
Solania	L238_V885	yes
Solania	L239_V930	yes
Solania	L240_V928	yes
Solania	L242_V882	yes
Solania	L247_V891	yes
Solania	L216_V878	yes
Solania	L217_V927	yes
Solania	L218_V879	yes
Solania	L222_V888	yes
Solania	L223_V886	yes
Solania	L228_V882	yes
Solania	L232_V922	yes
Solania	L234_V892	yes
Solania	L235_V887	yes
Solania	L235_V923	yes
Solania	L236_V889	yes
Solania	L243_V893	yes
Solania	L245_V880	yes
Agrigenus	SM_Agrigenus	yes
Annalisa	SM_ANNALISA	yes
Auchan	SM_Auchan	yes
Biologico	SM_BIOLOGICO	no
Cento	SM_CENTO	yes
Danicoop	SM_Danicoop	yes

2.1.3 Evaluation of Environment-Dependent Transcripts

Transcriptomic analysis was carried out on berries of two tomato varieties: 'Kiros', 'San Marzano' genotype included in the "Disciplinare di produzione della Denominazione di Origine Protetta - Pomodoro San Marzano dell'Agro Sarnese-Nocerino", and 'Docet', hybrid variety for industrial uses with elongated berries. Genotypes were grown in two different environments near Naples, Acerra, area included in the procedure guideline for 'San Marzano' PDO production, and Brusciano, area not included in the same procedure guideline. Berries from three different plants (biological replicates) were collected in three different fructification steps in August and September 2012. Tomato fruits were immediately frozen in liquid nitrogen and stored at -80°C.
For each genotype a total of 12 samples were collected (Table 2.3).

Table 2.3: Description table of tomato fruits used for transcriptomic analysis. Genotype name, collection environment, biological replicates code, name of the group and name of the comparison.
PK: Kiros; D: Docet; 1, 2 or 3: number of the biological replicate; AC: Acerra; BR: Brusciano.

Genotype	Environment	Biological Replicates code	Sample name	Comparison
Kiros	Acerra	PK1AC	Kiros Acerra	Kiros Comparison
		PK2AC		
		PK3AC		
	Brusciano	PK1BR	Kiros Brusciano	
		PK2BR		
		PK3BR		
Docet	Acerra	D1AC	Docet Acerra	Docet Comparison
		D2AC		
		D3AC		
	Brusciano	D1BR	Docet Brusciano	
		D2BR		
		D3BR		

2.2 SSR Analysis for DNA Fingerprinting and Traceability

2.2.1 DNA Extraction and Quantification

DNA was extracted using 100 mg of frozen young leaves or 250 mg of lyophilized peeled tomato fruits. Leaves and berries were finely ground in a mortar in presence liquid nitrogen to obtain a powder.
Leaves genomic DNA was extracted with the "GenElute Plant Genomic DNA Kit" (G2N70 - SIGMA).
DNA extraction from peeled tomato fruits was performed including some modification to the protocol: the volumes of the Lysis Solution A and the Lysis Solution B were increased from 350 to 650 µl and from 50 to 95 µl, respectively; 230 µl of Precipitation Solution were added instead of 130 µl; finally, samples were centrifuged for 20' instead 5' to precipitate cell debris.

The extracted DNA was analyzed by agarose gel electrophoresis (1% w/v, containing ethidium bromide) in 1X TAE buffer and visualized by UV light (UV Gel Doc BIORAD), furthermore DNA was quantified at the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies).

2.2.2 DNA amplification for SSR analysis

Thirteen SSR loci, representative of various classes of repetition in the core, were selected (He et al., 2003; Smulders et al., 1997) (Table 2.4), and analyzed by PCR amplification followed by capillary electrophoresis.
The amplification reactions with DNA extracted from leaves were conducted in a final volume of 25 µl containing 20 ng of genomic DNA, 1X PCR buffer (Promega), 0.1 mM of dNTP, 0.2 mM of forward primer 5'-labeled with a fluorophore (Table 2.4), 0.2 mM of unlabeled reverse primer, 0.5 U of Taq polymerase (Promega). The amplification reactions were performed in a thermocycler with the following temperatures and reaction times: first step of DNA denaturation at 94°C for 3'; 35 cycles including denaturation at 94°C for 45'', annealing to the specific Ta for 45'' (Table 2.4) and elongation at 72° C for 1'30''; then, one final step of elongation at 72°C for 5'.
The DNA extracted from processed tomato products was amplified in a final volume of 25 µl, using 3-5 µl of the solution containing the extracted genomic DNA with the following procedure: DNA denaturation at 94°C for 5'; 39 cycles at 94°C for 50'', 50" at the specific Ta (Table 2.4) and elongation at 72°C for 5'; the final step of elongation was at 72°C for 15'.
The amplicones were separated by electrophoresis on 2% (w/v) agarose gel, prepared with ethidium bromide, and visualized through UV light (UV Gel Doc BIORAD). Amplicons sizes were calculated by comparison with the known size fragments of the "1 Kb Plus DNA Ladder" (Invitrogen).

Table 2.4: Features of SSR loci: name of the SSR locus, repeat type of core, expected fragment size in bp, polymorphic indicative content from literature (PIC), 5'-3' for forward and reverse primer, type of fluorescent label (only at the 5 'end of the forward primer) and annealing temperature (°C).

Marker	Repeat Motive	Fragment size (bp)	PIC	Primers sequences (5'-3') (Forward, Reverse)	Label (5'- forward)	Ta (°C)
LE20592	$(TAT)_{15-1}(TGT)_4$	158-167	0,58	CTGTTTACTTCAAGAAGGCTG	6-FAM	55
				ACTTTAACTTTATTATTGCCACG		
LE21085	$(TA)_2(TAT)_{9-1}$	98-113	0,36	CATTTTATCATTTATTTGTGTCTTG	6-FAM	50
				ACAAAAAAAGGTGACGATACA		
LEaat002	(AAT)12	106	0,55	GCGAAGAAGATGAGTCTAGAGCATAG	6-FAM	55
				CTCTCTCCCATGAGTTCTCCTCTTC		
LEat002	(AT)9	203	0,50	ACTGCATTTCAGGTACATACTCTC	6-FAM	57
				ATAAACTCGTAGACCATACCCTC		
LEcaa001	(CAA)7	105	0,33	AGAAGGCGTGAGAGGCAAC	6-FAM	52
				CTTAGCACTTGATGTTGATTGG		
LEct001	(CT)12	111	0,35	TCCAATTTCAGTAAGGACCCCTC	NED	52
				CCGAAAACCTTTGCTACAGAGTAGA		
LEctt001	(CTT)9	101	0,39	CCTCTCTTCACCTCTTTACAATTTCC	PET	57
				CACTGGTCATTAAGTCTACAGCC		
LEEF1Aa	$(TA)_8$ $(ATA)_9$	198-213	0,67	AAATAATTAGCTTGCCAATTG	6-FAM	50
				CTGAAAGCAGCAACAGTATTT		
LEga003	(GA)20	241	0,58	TTCGGTTTATTCTGCCAACC	VIC	52
				GCCTGTAGGATTTTCGCCTA		
LELE25	$(TA)_{11}$	211-217	0,36	TTCTTCCGTATGAGTGAGT	6-FAM	50
				CTCTATTACTTATTATTATCG		
LEta003	(TA)9	111	0,43	GCTCTGTCCTTACAAATGATACCTCC	VIC	52
				CAATGCTGGGACAGAAGATTTAATG		
LEta015	(TA)15	107	0,49	ATATGCATGGACAAATCTTGAGGG	PET	55
				CTCGCGCATCAAATTAATGTATCAG		
LEtat002	(TAT)12	196	0,58	ACGCTTGGCTGCCTCGGA	NED	55
				AACTTTATTATTGCCACGTAGTCATGA		

2.2.3 Capillary Electrophoresis for SSR Alleles Detection

The amplified fluorescent amplicons were analyzed by capillary electrophoresis on ABI Prism 3130-AVANT (Applied Biosystems). Information about the allele size is transferred to the computer and processed through the programs "ABI Prism GeneMapper" (v.4) (Applied Biosystems), obtaining electropherograms where each peak represents one SSR amplified allele.
Fluorescence used were: 6-FAM, VIC, NED and PET showing an emission spectrum in the blue, green, yellow and red, respectively (Figure 2.1). To perform a capillary electrophoresis run a mix composed by 1µl of PCR product (or diluition), 0.28 µl of GeneScan 500 Liz standard (Applied Biosystems) and 6.72 µl of formamide was used. The reaction was denatured for 5' at 95°C and placed in ice for 2'.

Electrophoretic run conditions include a voltage of 15kV, a temperature of 60°C and a time of 45'.

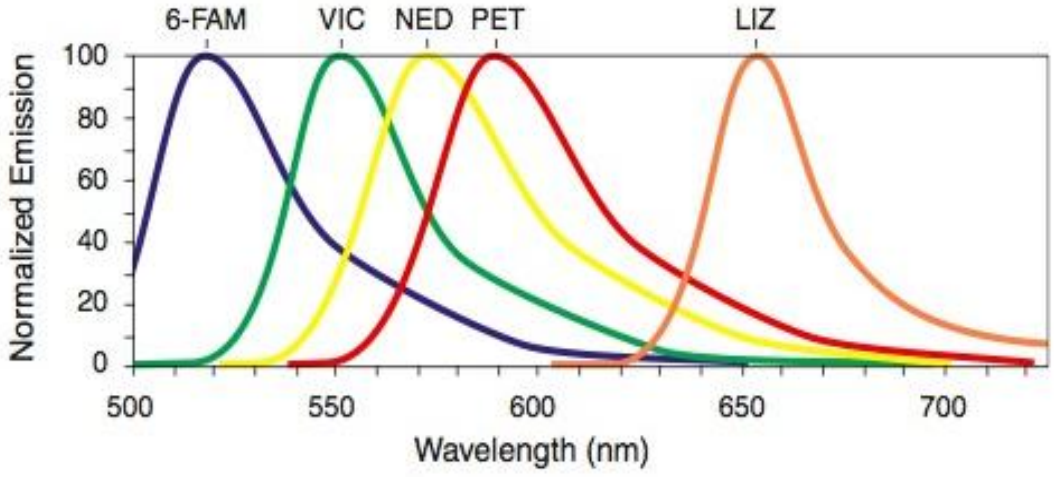


Figure 2.1: Emission spectrum for four fluorochromes used to analyze the amplified SSR allels by capillary electrophoresis (6-FAM, VIC, NED and PET) and the emission spectrum of the fluorescence associate to the ladder GeneScan 500 Liz (Applied Biosystems).

2.2.4 SSR Data Analysis

The analysis of the main genetic parameters for SSR markers was performed using the program GeneAlex 6 (Genetic Analysis in Excel) (Peakall et al., 2006). It was estimated the allele number (Na) per locus that is the number of SSR alleles found to each SSR locus. The observed heterozigosity (Ho) was calculated as number of Hets/N where Hets represents the heterozygotes and N is the total number of analyzed samples. The expected heterozigosity (He), was calculated as $1 - \Sigma pi^2$ where pi is the frequency of the i^{th} allele at each locus. Finally, the discrimination power (D) was computed as 1-ΣC where C is pi*[(N-pi)-1]/(N-1).
The calculated genetic distances (GD) were used to produce graphical representations (PCA-Principal Component Analysis) of the relationships among samples.

2.3 Transcriptomic analysis for Traceability

2.3.1 RNA Extraction, Quantity and Quality Evaluation

RNA extraction from 300 mg of frozen tomato fruits (Table 2.3) was performed by a phenol/chloroform procedure and lithium chloride precipitation. Extraction buffer, (750 µl) containing 100 mM Tris-HCl pH 8.5, 100 mM NaCl, 20 mM EDTA pH 8.0 and 1% SDS, and 750 µl phenol/chloroform 1:1, were added to the powder and the mixture vortexed and centrifuged at 13000 rpm at 4°C for 5'. Isopropanol (750 µl) were added to the supernatant and samples were incubated in ice for 5' and then centrifuged at 13000 rpm at 4°C for 10 min to perform DNA and RNA precipitation.
The upper phase was removed and the pellet was firstly dried and then suspended in 400 µl of DEPC-treated water (DEPC- Diethylpyrocarbonate; Sigma).

In order to have a RNA selective precipitation 1 volume of 4M Lithium Chloride (Sigma) was added to the samples then left on ice over-night.
The following day samples were centrifuged at 13000 rpm at RT for 20 min, supernatants were discarded and pellets suspended in 400 µl of DEPC-treated water. RNA was precipitated through the addition of 0.1 volume of 3M Sodium Acetate pH 7.2 and 1 volume of 96% ethanol and the incubation at -80°C for 10'. Pellets were collected by centrifugation at 13000 rpm at 4°C for 10' and suspended in 42 µl of DEPC-treated water. Two µl of isolated RNA were quantified by NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies).
RNA integrity was checked by horizontal electrophoresis on a 1.2% (w/v) agarose gel prepared without addition of ethidium bromide. Two µg of each sample were prepared with 20 µl of 10 X RNA Loading Buffer composed by 400 µl Formamide, 120 µl 37% formaldeyde, 5 µl loading buffer 10X (50% glycerol, 0.25% w/v bromophenol blue, 0.25% w/v xylene cyanol; Sigma), and 1.2 µl of 10 mg/µl ethidium bromide and than denatured at 65 °C for 5'. Run was performed at 50V for 20'. Gel visualization was performed using UV light (UV Gel Doc BIORAD).
In order to obtain high quantity of RNA for each samples more than one extraction was performed and total RNA from several extraction was pooled for each of 12 samples.
A treatment with DNAse I to remove DNA contaminations was done on total RNA. Ten µg of RNA were added with 1X DNAse I Reaction Buffer (Invitrogen), 6 U DNAse I Amplification Grade (Invitrogen) and sterile water until a final volume of 100 µl. After the incubation at RT for 15 min, reaction was stopped incubating samples on ice.
DNase I was removed by phenol/chloroform precipitation of RNA samples.
Finally RNA samples were analyzed quantitatively and qualitatively by NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies) and by Bioanalyzer (Agilent Technologies).

2.3.2 RNA-Sequencing and Data Analysis

Total purified RNA was converted to cDNA and sequenced on Illumina HiSeq1500 platform at the LabMedMolGe (Laboratory of Moleculare Medicine and Genomics Department of Medicine and Surgery, University of Salerno).
The process generates millions (36 M) of short (100 bp) reads sequenced from both ends of each cDNA fragment (paired-end sequencing).
The raw data for each sample consist of a long list of short sequences with associated quality scores (*fastq* format) (Ewing et al., 1998a; Ewing et al., 1998b).
All the steps described below refer to all the biological replicates analyzed per sample.
To evaluate the quality of the sequences the FastQC software was used (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). FastQC provide a simple way to do quality control checks on raw sequence data coming from high throughput sequencing pipelines. It give graphical outputs with information about the sequence such as per base sequence quality, per sequence quality score, per base sequence content, per sequence GC content, per base N content, sequence length distribution, sequence duplication level and so on.
The following step was to map the obtained reads to the tomato reference genome.
Tomato genome is available from 2012 (http://solgenomics.net/organism/Solanum_lycopersicum/genome), it is composed by

12 nuclear plus 1 plastidial chromosomes for a total of 950 Mbp with 34727 genes of which 56.6% are annotated with GO terms (19662). The total number of unique GO for the tomato genome is 2108. Tomato genome was downloaded from the Ensembl Plants database (http://plants.ensembl.org/info/data/ftp/index.html).
The script TopHat v.2.0.10 (http://ccb.jhu.edu/software/tophat) was used to align RNA-Seq reads to the tomato genome. The output files useful for the following analysis is the "accepted_hits.bam". This file contains a list of read alignments in SAM format, a compact short read alignment format (http://samtools.sourceforge.net).
The Qualimap (Garcia-Alcalde et al., 2012) software was used to check the mapping quality. Qualimap examines sequencing alignment data according to the features of the mapped reads and their genomic properties and provides an overview of the data that helps to detect biases in the sequencing and/or mapping of the data.
To continue with the data analysis, a table composed by i^{th} rows and the j^{th} columns, that tells how many reads (counts) have been mapped to gene i in sample j, was needed. The HTSeq Python package (http://www-huber.embl.de/users/anders/HTSeq/) was used to obtain the coverage from the mapping files produced by TopHat.
Than, the coverage was normalized using the CPM (Count Per Million) method proposed in the edgeR analysis pipeline (Robinson et al., 2010), calculated as the raw counts divided by the library sizes and multiplied by one million, in Rstudio v.3.0.2 (R Core Team, 2013) and a box plot representation of the normalized coverage was produced for each biological replicates for each sample.
The normalized coverage was first used to understand the relationship among genotypes grown in two different environments and their biological replicates. The table with the coverage data was elaborated using RStudio v.3.0.2 (R Core Team, 2013) by the "pcaMethod" package (Stacklies et al., 2007) to compute the PCA graph.
Later, to select the expressed genes for genotype in each environment, a threshold of 1CPM has been placed. Genes that showed coverage higher than 1 CPM in all three biological replicates per sample were selected.
Using the package "limma" (Smyth 2005) Venn diagrams were produced to start performing comparisons using lists of expressed genes.
Than, functional enrichment of the expressed genes for genotype in each environment was performed using the Blast2GO software (Conesa et al., 2005) at the default parameters, by introducing the functional annotation already available for the tomato genome (ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG2.3_release/).
The Blast2GO Fisher's exact test outputs, consisting in lists of GO terms per sample, were than used to further compare samples, highlighting GO terms specific for genotype in each environment.
Therefore, from the specific GO terms list per sample, expressed genes present exclusively in each sample were extrapolated by comparison with the list of expressed genes of the other sample from the same comparison.
Finally, the coverage of the specific expressed genes per sample was used to determine statistical differences in gene expression between samples from the same comparison by the "NOISeq" package (Tarazona et al., 2011) in order to find differentially expressed genes (DEG).

3. RESULTS

3.1 Characterization and Discrimination of Tomato Genotypes

A collection of 146 tomato genotypes (Tables 2.1) was analyzed at 13 SSR loci (Table 2.4) selected for their high polymorphism and high potential to discriminate the different varieties within the cultivated species of *S. lycopersicum*.

3.1.1 SSR amplification and Capillary Electrophoresis

The amplification products were separated on an agarose gel to verify the presence and the molecular size of the PCR fragments. For all tomato varieties, analyzed loci generated amplicones of specific molecular size corresponding to the range expected. Figure 3.1 is an example of the electrophoretic separation of the amplification products at the LEta015 locus where it is possible to verify that the observed size of the fragments obtained, approximately 100 bp, reflect the expected size from the literature (Table 2.4).

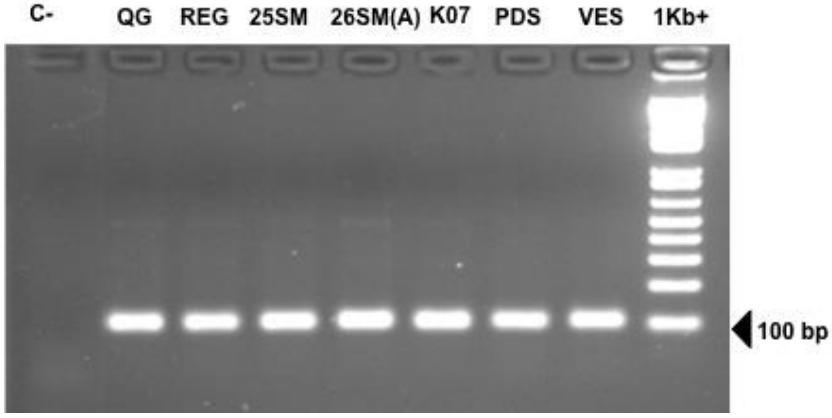


Figure 3.1: example of electrophoretic separation on agarose gel of genomic DNA amplification products obtained from seven of 146 tomato varieties at the SSR locus LEta015. QG: Quarantino Grande; REG: Reginella; 25SM: 25 SMEC 2.1.1.1; 26 SM(A): 26SMEC 2.2.1.1; K07: Kiros 07; PDS: Vesuviano pomodorino o Piennolo Rosso; VES: Vesuviano pomodorino o Piennolo Rosso 2; C-: negative control, mix without template; 1Kb+: 1Kb plus ladder (Invitrogen).

The amplified fragments were then analyzed by capillary electrophoresis. Figure 3.2 shows some examples of electropherograms of SSRs amplified at the LEta015 locus. The core of this SSR is characterized by a dinucleotide repeat. It is possible to note that the electrophoretic pattern consists of multiple peaks of different height and length. The highest peak is the allele while peaks that flank the highest are the stutter bands. These are due to the slippage of Taq polymerase during SSR amplification (Delmotte, 2001).

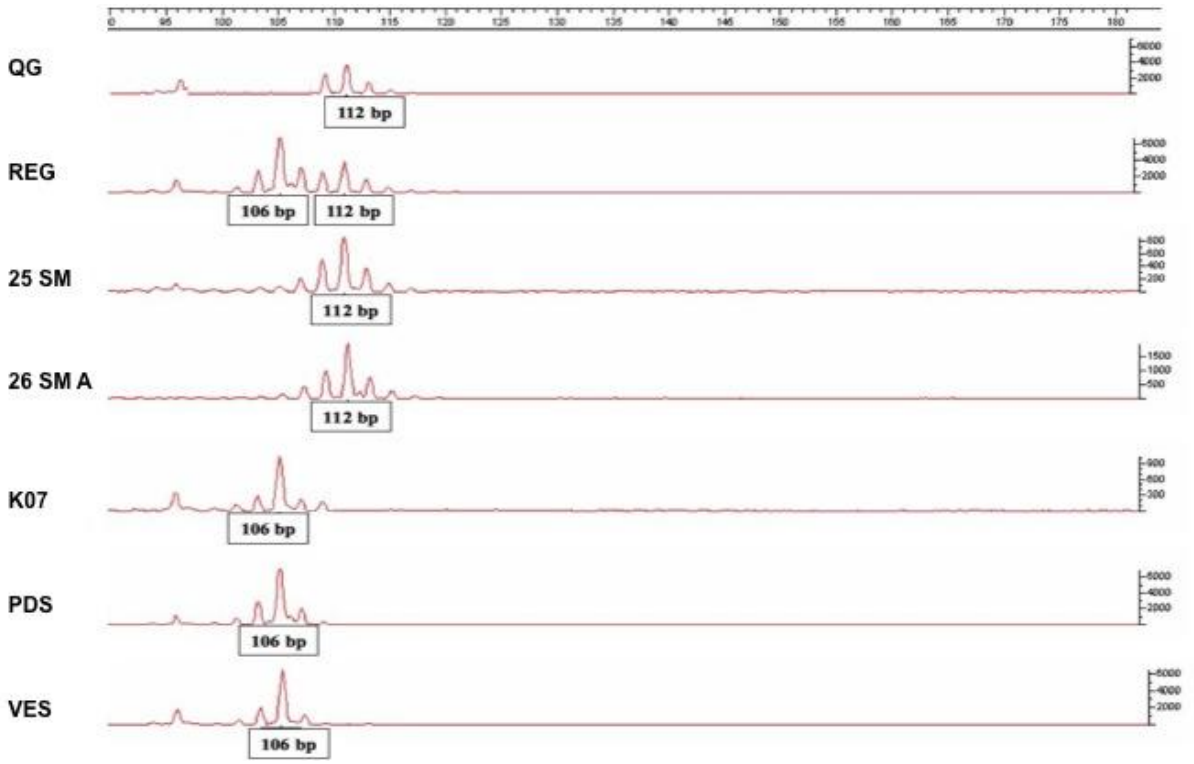


Figure 3.2: Examples of electropherograms of PCR products at the SSR locus LEta015. For each electropherogram the top bar indicates the allele size in bp while the vertical bar indicates the intensity of the peaks (RFU). QG: Quarantino Grande; REG: Reginella; 25SM: 25 SMEC 2.1.1.1; 26 SM(A): 26SMEC 2.2.1.1; K07: Kiros 07; PDS: Vesuviano pomodorino o Piennolo Rosso; VES: Vesuviano pomodorino o Piennolo Rosso 2.

3.1.2 SSR Data Analysis

The SSR data for the 13 SSR loci analyzed in 146 tomato varieties were elaborated by the use of the software Genalex 6 (Genetic Analysis in Excel) (Peakall et al., 2006) to calculate the main genetic indices needed to estimate the genetic variability (Table 3.1).

The 13 SSR loci were all polymorphic. The total number of observed alleles (Na) was 71, with the maximum value, equal to 18, for the locus LEEF1Aa (the most represented allele was of 203 bp present in 53% of tomato genotypes) and the lowest number of 2 at the locus LEcaa001. Data analyses for each SSR locus are summarized as follow:

-SSR locus LE20592: four alleles, of 161, 164, 167 and 176 bp, were present; the most represented allele was 167 bp present in 48% of the samples. The remaining tomato genotypes (52%) presented alleles of 164, 161 or 176 bp with frequencies of 35, 16 and 1%, respectively;

-SSR locus LE21085: this locus presented three alleles with size of 100, 102 and 116 bp; the highest allele frequency was for the allele of 102 bp with a percentage of 82%; for the other two alleles, 116 and 100 bp, the frequencies was equal to 10 and 5%, respectively.

-SSR locus LEaat002: four different alleles of 101, 104, 95, and 98 bp were recorded with frequency of 68, 24, 6 and 2%.

-SSR locus LEat002: a total of 4 different alleles were found; most of the tomato genotypes showed alleles of 201 bp and 197 bp (47% and 50% of the samples

respectively), while the alleles of 199 and 203 bp were present in 2% and 1% of samples.
-SSR locus LEcaa001: two alleles, 100 and 97 bp, are present showing frequencies of 90% and 10%.
-SSR locus LEct001: five different alleles of 97, 99, 101, 105 and 107 bp with variable frequency among samples (39, 2, 2, 12 and 45 %, respectively) are present.
-SSR locus LEctt001: the Na parameter was of five alleles for this locus (91, 94, 97, 100 and 103 bp) and the most represented allele (allele frequency of 84%) was of 94 bp; for the other alleles the frequency was between 1 and 8.7%.
-SSR locus LEga003: 7 alleles of 217, 227, 229, 231, 233, 236 and 239 bp existed and the highest allele frequency of 82% was reported for the allele of 233 bp.
-SSR locus LELE25: the locus presented three alleles of 218, 220 and 222 bp with a frequency of 5%, 85% and 10% respectively.
-SSR locus LEta003: six alleles are present; the allele size of 104 bp was the most represented with a frequency of 45%, followed by 106 bp (27%), 108 bp (18%), 102 (9%), 100 bp (0,7%) and 94 bp (0,3%).
-SSR locus LEta015: half of the analyzed samples showed the allele of 112 bp (allele frequency of 50%), other six alleles of 104, 106, 108, 110, 114 and 116 bp were present with frequencies of 2, 32, 13, 1, 1, and 1%.
-SSR locus LEtat002: alleles of 192, 195 and 198 bp were revealed for this locus, showing frequencies of 16, 34 and 50%, respectively.
The average value of Na was 5.462 alleles for the analyzed SSR loci.
(Ho) is the observed heterozygosity that is the real percentage of heterozygous samples found at the considered locus. This values ranged from 0, for LEcaa001 and LEct001 loci for which all samples were homozygote , to 0.172, for the LEga003 locus. The expected heterozigosity (He) that is the probability that other genotypes out of the population under analysis are heterozygotes at the considered locus, ranged from 0.183 to 0.703 for the set of SSR markers analyzed. The discrimination power (D) indicates the ability to discriminate among samples estimated for each locus after the analysis. This parameter has values between 0, for monomorphic markers, and 1, for polymorphic loci. The higher (near to 1) is the D value the greater is the power of the marker to discriminate genotypes. The maximum value for the D (0.706) was found at the locus LE20592 that showed the greatest discriminating power. The average value of the D for all loci under consideration was of 0.484.

Table 3.1: Main genetic indices and their average of the 13 SSR loci obtained from the analysis of the 146 tomato genotypes. Na: number of alleles; Ho: observed heterozygosity; He: expected heterozigosity; D: discrimination power; Allele size: size in bp of the alleles found to each SSR locus.

Locus	Na	Ho	He	D	Alleles size
LE20592	4	0,130	0,703	0,706	161, 164, 167, 176
LE21085	3	0,113	0,258	0,259	100, 102, 116
LEaat002	4	0,070	0,472	0,473	95, 98, 101, 104
LEat002	4	0,065	0,528	0,529	197, 199, 201, 203
LEcaa001	2	0,000	0,183	0,184	97, 100
LEct001	5	0,000	0,629	0,632	97, 99, 101, 105, 107
LEctt001	5	0,126	0,293	0,294	91, 94, 97, 100, 103
LEEF1Aa	18	0,154	0,687	0,690	124, 125, 126, 189, 195, 196, 200, 201, 203, 204, 205, 206, 207, 208, 209, 212, 213
LEga003	7	0,172	0,323	0,324	217, 227, 229, 231, 233,236, 339
LELE25	3	0,008	0,275	0,276	218, 220, 222
LEta003	6	0,077	0,684	0,686	94, 100, 102, 104, 106, 108
LEta015	7	0,079	0,635	0,637	104, 106, 108, 110, 112, 114, 116
LEtat002	3	0,124	0,606	0,608	192, 195, 198
Mean	5,462	0,086	0,483	0,485	-

3.1.3 Genetic Relationships between tomato varieties

Plant material analyzed was composed by local varieties cultivated in Campania Region and varieties for industrial uses. The genetic distance calculated on the SSR data revealed the discrimination of these two main groups, as reported in the PCA graph (Principal Component Analysis) (Figure 3.3), showing that varieties for industrial uses grouped together separating from local varieties cultivated in Campania Region. Furthermore, in the latter, it was possible to identify the subgroup of ‘San Marzano’ types.

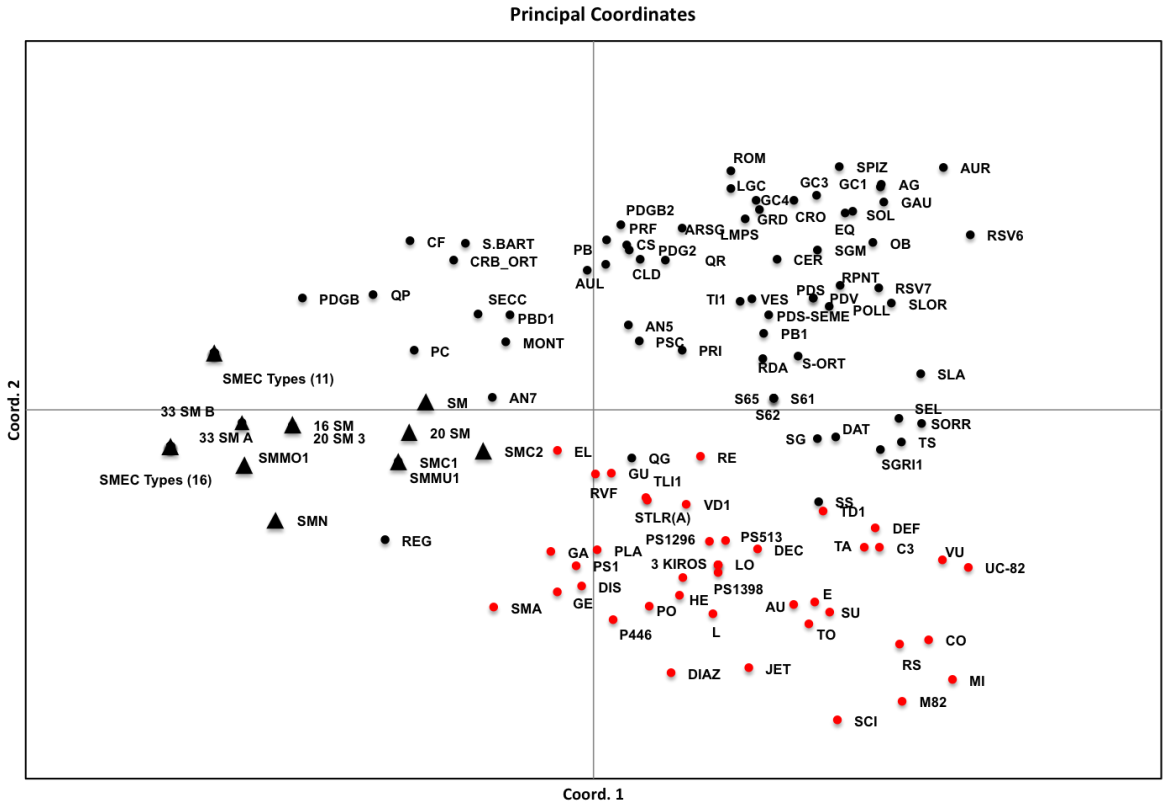


Figure 3.3: PCA graph based on genetic distance calculated on SSR data showing grouping of local varieties cultivated in Campania Region (black points and triangles) and industrial varieties (red points) of the146 tomato genotypes. Black triangles indicate the 'San Marzano' subgroup. Identification codes are reported in Table 2.1.

The dendrogram reported in figure 3.4, was constructed using the genetic distance among varieties calculated based on SSR analysis. It shows that 72% of samples were well discriminated. The remaining 28% were distributed among the three groups, and includes ecotypes and varieties with high similarity level or, eventually, a complete genetic uniformity, as expected in some cases for which are proposed cases of homonymy. It is composed as follows: three 'Sorrento' genotypes, 'Sorrento 61', 'Sorrento 62' and 'Sorrento 65', are identical and, furthermore, in the same cluster there are also 'Sorrento_2', 'Sorrento Globoso Rosato Indeterminato', 'Sorrento Semiorto', 'Sorrento Tondo Liscio Rosato (A)', 'Tondo Liscio Indeterminato', 'Tondo Determinato' and 'Sorrento Gragnano'; two 'Vesuviano' genotypes, 'Piennolo vesuviano' and 'Vesuviano pomodorino/Piennolo Rosso', showed the same SSR allelic profile and high similarity with 'Vesuviano pomodorino/Piennolo Rosso_2' and 'Vesuviano'.

Moreover, it was possible to identify a cluster composed by two accessions of 'Pomodorino giallo di Camposano', 'Tondo giallo di Roccadaspide', 'Lungo giallo di Capaccio' and 'Pomodorino giallo del Beneventano', where the common parameter is the fruit colour, and a group of cherry tomatoes, including 'Cento scocche', 'Corbarino', 'Cannellino flegreo', 'Pomodorino di collina', 'Pomodorino Giallo di Montecalvo', 'Seccagno', 'Pomodorino giallo di San Bartolomeo', 'Quarantino piccolo' and 'Pomodorino giallo del Beneventano_2'.

Finally, industrial varieties formed a cluster in which varieties with elongated fruits are grouped together, such as 'Player', 'Discovery', 'Genius', 'PS513', 'P446', 'PS1398',

‘Herdon’, ‘Diaz’, ‘Auspicio’ and ‘Talent’, and varieties with cherry fruit shape (‘Tomito’ and ‘Minidor’, ‘Vulcan’ and ‘Defender’) showed two grouping.

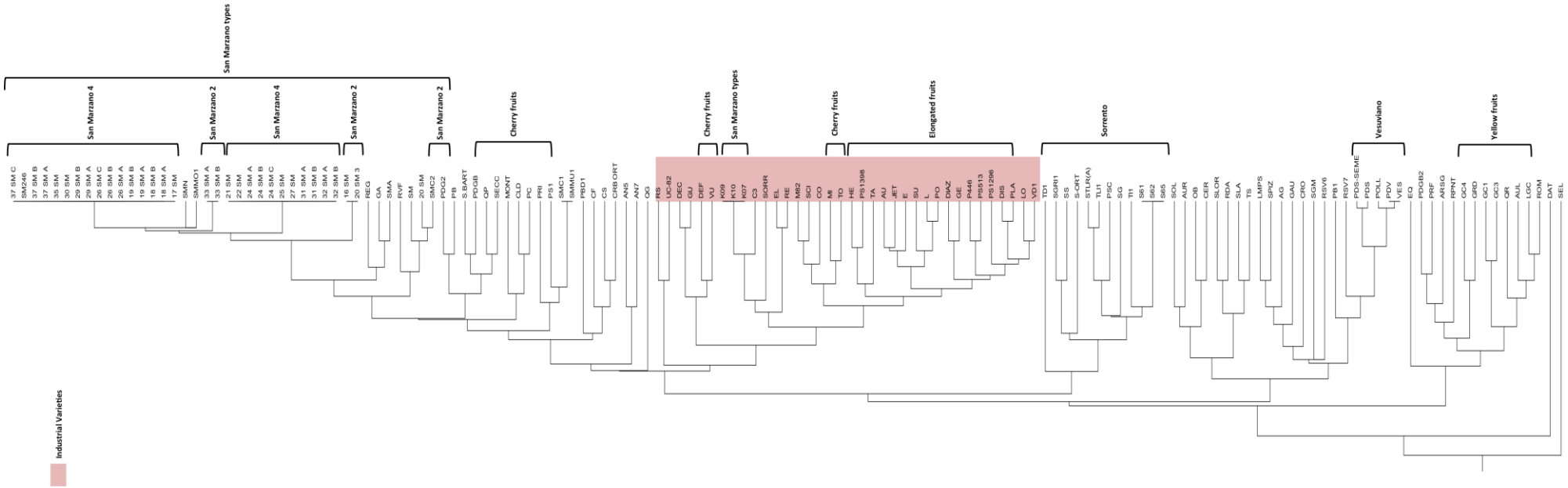


Figure 3.4: Dendrogram for graphical representation of the genetic relationships among 146 tomato genotypes. Identification codes are reported in Table 2.1.

3.2 Molecular Characterization of San Marzano types

In order to verify the possible use of the selected SSR molecular markers for authentication of the 'San Marzano dell'Agro Sarnese Nocerino' included in the PDO, along the peeled tomato chain, a group of 40 of 'San Marzano' types was analyzed (Table 2.1).
'San Marzano' plant material included types suitable for PDO production, that are five genotypes SMEC corresponding to 'San Marzano 2' and three 'Kiros' plants, both provided by the same germplasm bank CRA-A. The other 32 genotypes were 'San Marzano' types not used for PDO products but widely spread in Campania Region, as 'San Marzano 4' (27 genotypes) and other local types (5 genotypes) provided by CRA-A and Semiorto.
First of all the selected SSR were used for the molecular characterization of the 'San Marzano' types and the evaluation of the relative genetic relationships. All the analyzed genotypes resulted to be homozygous at the considered SSR loci.
The percentage of monomorphic and polymorphic loci resulted of 46.7% and 53.3%, respectively. Monomorphic loci LE21085, LEEF1Aa, LELE25, LEcaa001, LEctt001 and LEga003 presented a single allele in all samples that is 102 bp, 203 bp, 220bp, 100 bp, 94 bp and 233 bp, respectively. Among the seven polymorphic SSR loci, 2 of them showed two alleles and 5 had three alleles (Figure 3.5).

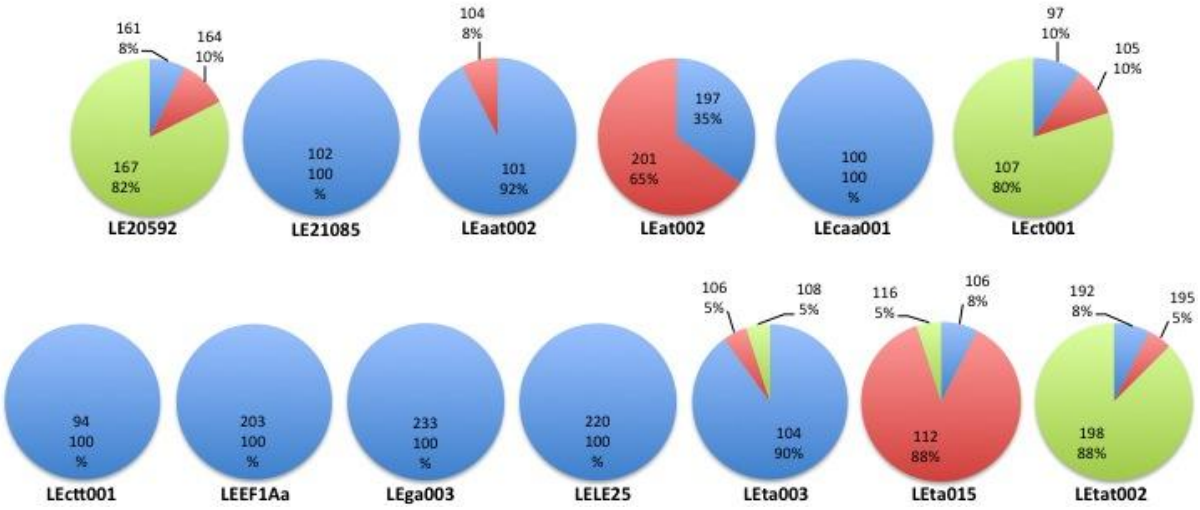


Figure 3.5: Allele frequencies at 13 SSR loci after the analysis of 40 'San Marzano' types.

Table 3.2: Genetic indices and their values observed by the analysis of 40 ‘San Marzano’ types at the 13 SSR loci. Na: number of alleles; Ho: observed heterozygosity; He: expected heterozigosity; D: discrimination power; Allele size: size in bp of the alleles found to each SSR locus.

Locus	Na	Ho	He	D	Allele size
LE20592	3	0,000	0,304	0,308	161, 164, 167
LE21085	1	0,000	0,000	0,000	102
LEaat002	2	0,000	0,139	0,141	101, 104
LEat002	2	0,000	0,455	0,461	197, 201
LEcaa001	1	0,000	0,000	0,000	100
LEct001	3	0,000	0,340	0,344	97, 105, 107
LEctt001	1	0,000	0,000	0,000	94
LEEF1Aa	1	0,000	0,000	0,000	203
LEga003	1	0,000	0,000	0,000	233
LELE25	1	0,000	0,000	0,000	220
LEta003	3	0,000	0,185	0,187	104, 106, 108
LEta015	3	0,000	0,226	0,229	106, 112, 116
LEtat002	3	0,000	0,226	0,229	192, 195, 198
Mean	1,923	0,000	0,144	0,146	-

Table 3.2 shows the values obtained for the different genetic indices characterizing the population taken into account.
Na parameter (number of alleles) is equal to 1 for six loci, while for the remaining seven, 6 loci had 3 alleles and one locus presented 2 alleles. A total of 25 different alleles were identified, with an average of 1.9 alleles per locus. The number of alleles per locus was lower comparing with the average obtained from the analysis of the total tomato collection (5.4). The Na average for ‘San Marzano’ types was also significantly lower compared with those observed by analysis conduced on different pool of cultivated tomato germplasm, with values of 2.5 (He et al., 2003), 3 (Smulders et al., 1997) and 4.8, which is reduced to 4.1 by limiting the analysis only to the Italian landraces (Mazzucato et al., 2008) and further reduced analyzing types belonging to one variety (Mazzucato et al., 2010).
The (Ho) parameter refers to the observed heterozygosity and resulted equal to 0 for all selected SSR loci because of the strictly self-pollinating nature of the ‘San Marzano’ variety. The D value of each SSR locus ranged from 0 to 0.461 (LEat002).
Genetic distance calculated on the SSR data was elaborated to obtain the dendrogram, showed in figure 3.6, showing the presence of two groups of ‘San Marzano Marzano 4’, composed by 16 and 11 genotypes, and two groups of ‘San Marzano 2’, composed by 2 SMEC samples respectively. Furthermore, a group composed by three ‘Kiros’ genotypes exists: ‘Kiros’ plants analyzed in this work represents the same genotypes collected in different years, hence the complete genetic identity between plants confirm the stability of the ‘Kiros’ genotype over the years

The use of SSR markers revealed to be useful for protection of the ‘San Marzano’ PDO chain allowing to distinguish ‘San Marzano 2’ and ‘Kiros’, included in the procedure guideline for PDO production, form the others ‘San Marzano’ types that can still be sold as ‘San Marzano’ but without the quality label.

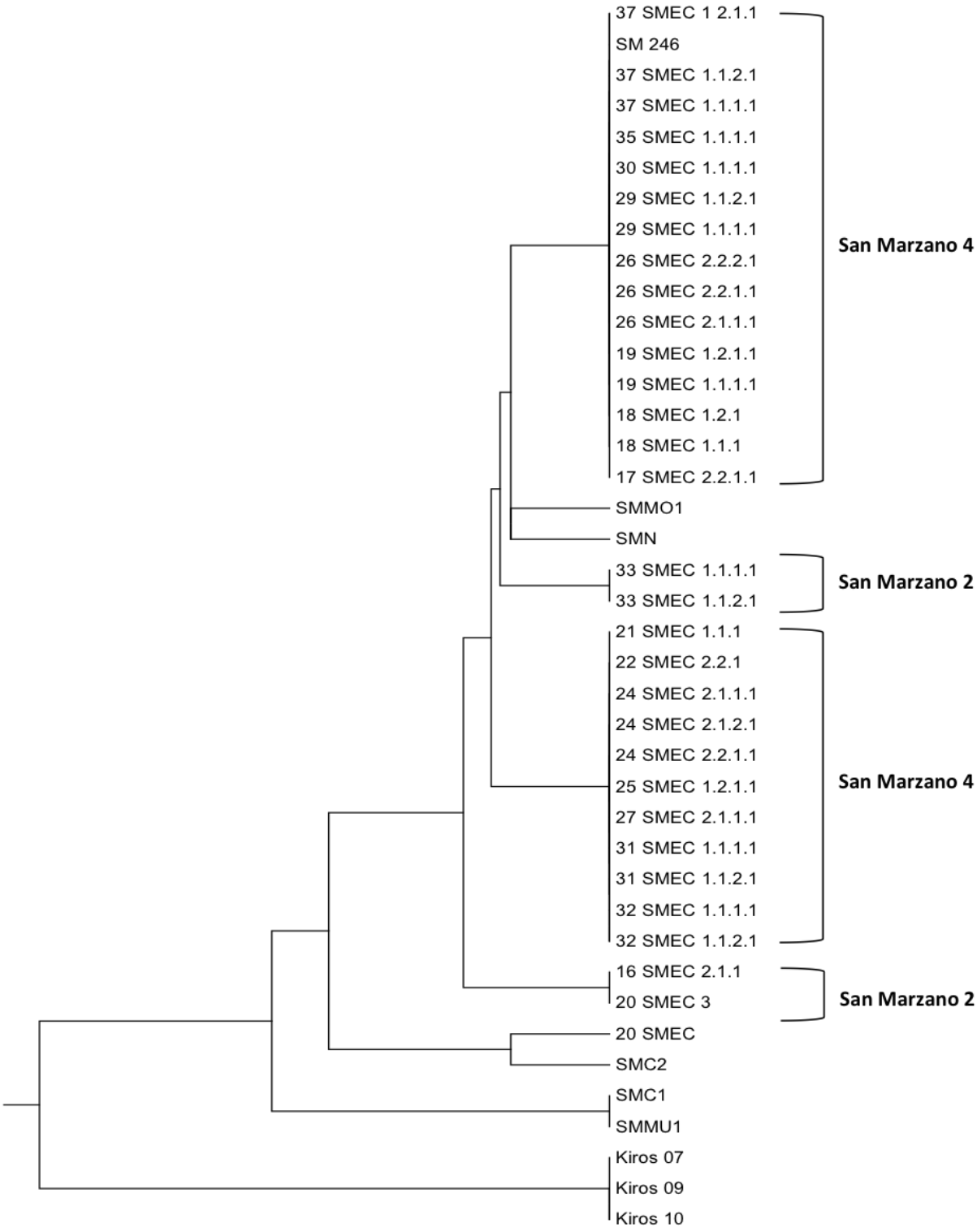


Figure 3.6: Graphical representation of the relationships among 40 ‘San Marzano’ types.

3.3 Authentication of San Marzano in Canned Tomatoes Through SSR Markers

in order to extend the authentication of the 'San Marzano' PDO to canned berries, thirty-three tomato cans all labelled as 'San Marzano' PDO (Table 2.2) were analyzed with selected SSR molecular markers.

3.3.1 DNA Extraction and SSR Analysis

the quality and the quantity of the DNA extraxted from canned berries could not be verified soon after the extraction. In fact, agarose gel electrophoresis of the extraction solution did not show any DNA bands as a possible consequence of the fact that the extracted DNA is highly degraded and present in very low amount. Furthermore, as DNA is degraded there is a limit in the size of the amplicones that can be obtained by PCR of DNA template extracted from processed tomato fruits (Caramante et al. (2010). For this reason a reduced set of SSR markers was used in the authentication analysis. Six SSR markers were selected, LE21085, LEaat002, LEct001, LEctt001, LEta003 and LEta015, as their amplification generate fragments in a range of 94-116 bp and also because of their high discrimination power (Table 3.1).
PCR amplification of DNA extracted from 33 peeled tomatoes generated fragments of the expected size for all the selected loci. Figure 3.7 shows an example of electrophoretic separation on agarose gel of the amplified fragments for the LEta015 locus.

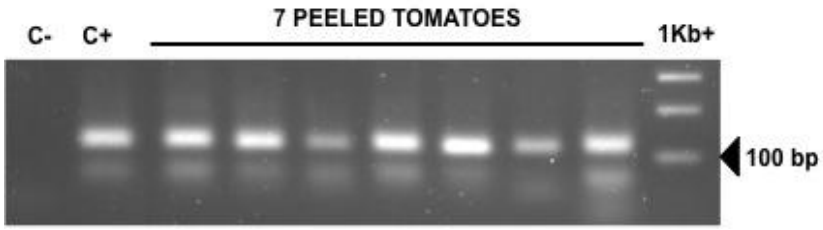


Figure 3.7: example of electrophoretic separation on agarose gel of the amplification products obtained from DNA extracted from seven peeled tomatoes at the SSR locus LEta015. C-: negative control, mix without template; C+: positive control, genomic DNA from tomato leaf; 1Kb+: 1Kb plus ladder (Invitrogen).

The amplification products were then separated by capillary electrophoresis (an example is reported in figure 3.8) in order to verify the correspondence in terms of allele size between the amplified fragment from processed tomatoes and the amplified alleles from the collection of 'San Marzano' types with special attention to 'San Marzano 2' and 'Kiros'. Results showed that the amplified alleles (LE21085: 100, 102, and 116 bp; LEta003: 104, 106 and 108 bp; LEaat002: 95, 98 and 101 bp; LEct001: 94, 97, 99 and 103 bp; LEctt001: 94, 103, 106 and 112 bp; LEta015: 102, 104, 106 and 110 bp) have the same size of the alleles found during the characterization of tomato genotype collection (Table 3.1).

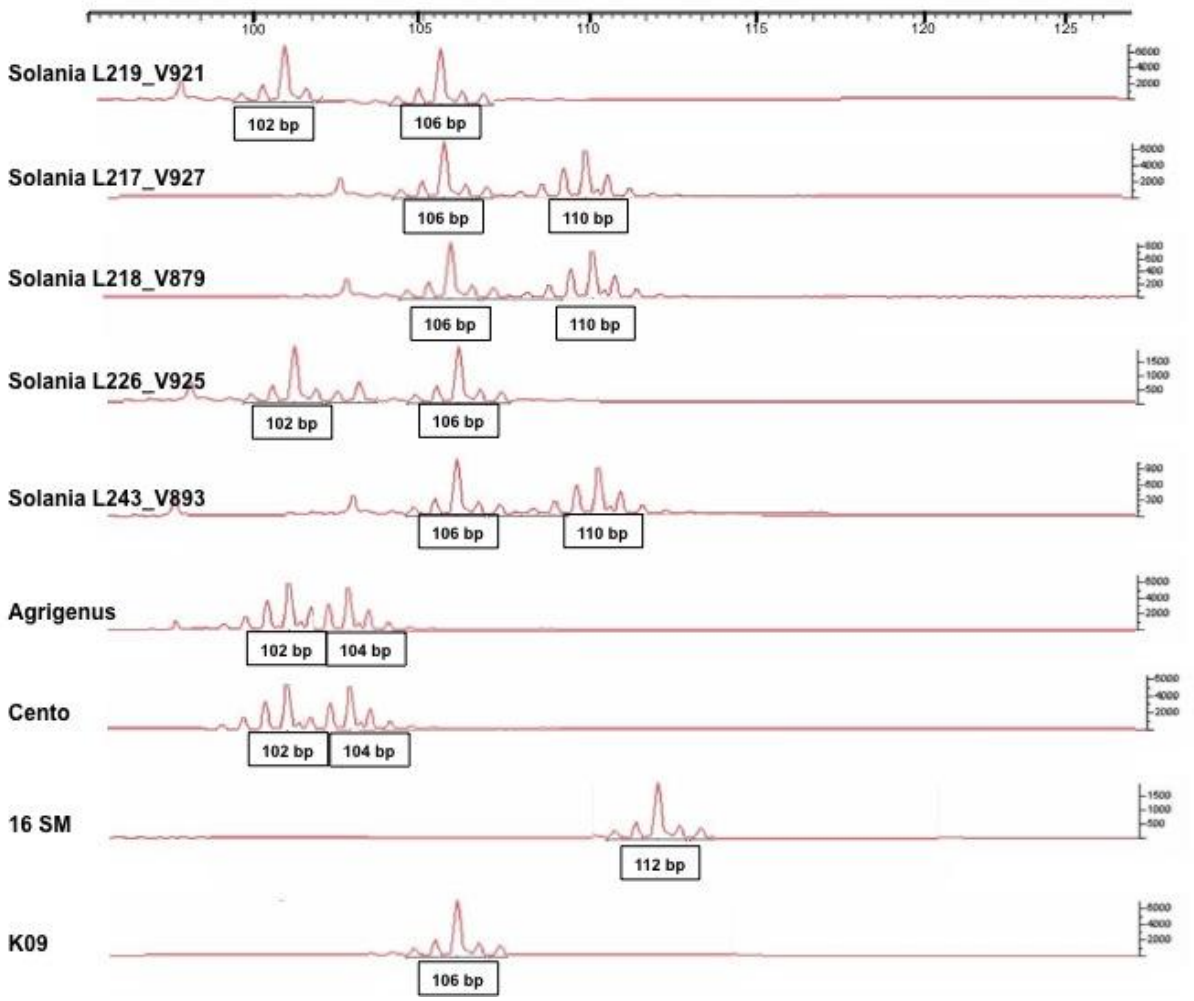


Figure 3.8**:** Examples of electropherograms at the SSR locus LEta015 for the PCR fragments produced from DNA extracted from processed tomato cans. Solania, Agrigenus and Cento are the manufacturing companies of tomato cans; Solania samples are classified for their production number; 16 SM: 16 SMEC 2.1.1, 'San Marzano2'; K09: Kiros 09.

SSR data were than analyzed using Genalex 6.4 (Genetic Analysis Exel) (Peakall et al., 2006) to estimate the main genetic parameters. Table 3.3 reports the number of alleles found and the calculated heterozygosity (Ho) for six SSR loci. The alleles number ranges between 3 and 5 and the allele size correspond to those obtained from the tomato genotype collection for the same six loci under analysis.
The study of the observed heterozigosity has given the minimum Ho value of 0.065 and the highest of 0.5 with an average of 0.341.

Table 3.3: Number of alleles (Na), observed heterozigosity (Ho) and allele size from the analysis of 33 peeled tomato at six SSR loci.

Locus	Na	Ho	Allele Size
LE21085	3	0,296	100, 102, 116
LEaat002	5	0,364	95, 98, 101, 107
LEct001	4	0,065	94, 97, 99, 103
LEctt001	4	0,424	94, 103, 106, 112
LEta003	3	0,400	104, 106, 108
LEta015	5	0,500	102, 104, 106, 110
Mean	4	0,34	-

These values were then compared with those obtained from the analysis of 'Kiros' and 'San Marzano 2' types (Table 3.2) finding no match in the percentage of heterozygotes, finding no match in the percentage of heterozygotes. As reported previously, 'San Marzano' types and 'San Marzano' PDO ('San Marzano 2' and Kiros') resulted homozygotes for all 13 SSR while Ho value of tomato cans was always higher than 0.
These results suggest that the 33 peeled tomatoes analyzed did not contain tomato fruits from 'San Marzano' PDO types.

3.4 Gene Expression Analyses

Gene expression analyses were carried out with the aim to study the transcriptomic profiles of the berries of two varieties, 'Kiros' and Docet' grown in two close geographic locations, Acerra and Brusciano (Figure 3.9).

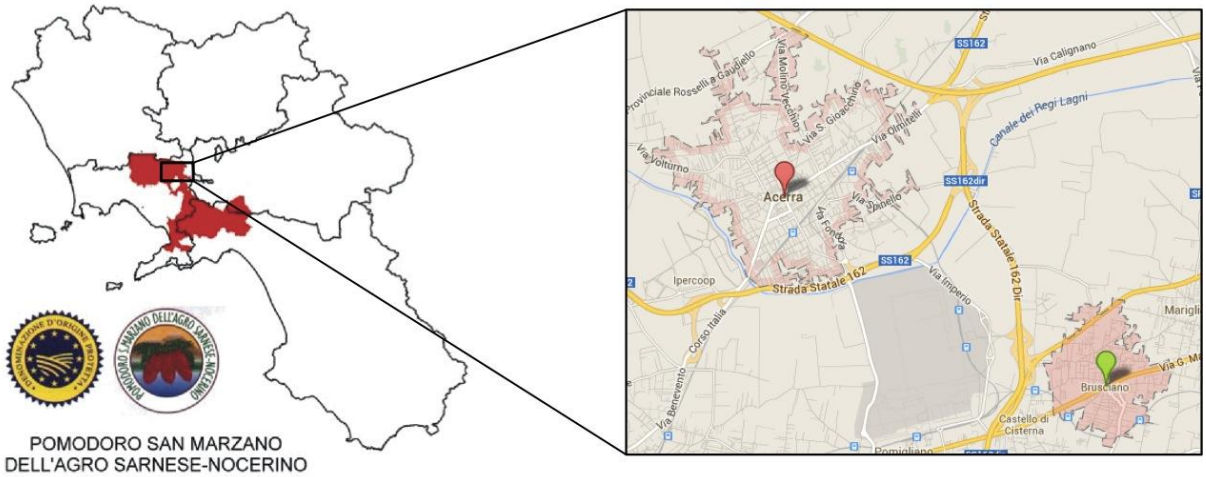


Figure 3.9: on the left: Campania Region area suitable for production of 'San Marzano' tomato PDO, on the right: zoomed localities, Acerra and Brusciano.

'Kiros' is included in the procedure guideline for the production of 'San Marzano' PDO, while 'Docet' is an hybrid variety with elongated berries used, similarly to 'Kiros' for the production of peeled tomatoes.
Acerra is one out 41 localities of Campania Region included in the list of production areas for the 'San Marzano' PDO; the soil has volcanic origin and is defined as typic haplustoll (Agriculture Dept, 2010) (Figure 3.10 A).
On the other hand, only part of Brusciano geographic area is included in the DOP procedure guideline but not the area selected for the experiment; the soil has alluvial origin and is defined as typic calciustepts (Agriculture Dept, 2010) (Figure 3.10 B).

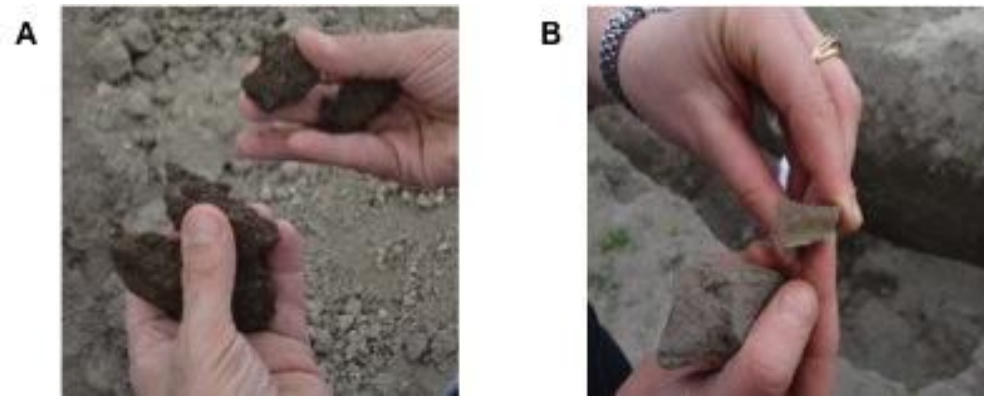


Figure 3.10: Soil sections from fields of Acerra (A) and Brusciano (B).

The study considered three biological replicates per genotype from each field (Table 2.3). The transcriptomic data obtained from the two varieties were then compared both as far as the locations and the genotypes.

3.4.1 RNA preparation for RNA-seq experiment

Total RNA was isolated from tomato fruits from 12 samples and visualized on agarose gel. Typically, the 28S (large subunit) and 18S (small subunit) cytosolic ribosomal RNA (rRNA) could be distinguished (Figure 3.11 A). RNA was analyzed quantitatively by NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies). The quality evaluation was performed by Bioanalyzer (Agilent Technologies) an automated bio-analytical device using microfluidics technology that provides electrophoretic separations in automated and reproducible manner. The result is visualized as an electropherogram where the amount of measured fluorescence correlates with the amount of RNA of a given size. In particular, first peak on the electropherogram represented the 18S rRNA while the following highest peaks is the 28S rRNA (Figure 3.10 B). RNA is considered of high quality when the large subunit of cytosolic rRNA is higher than 18S, as reported in figure 3.11 B, with a ratio value higher than 1.5.

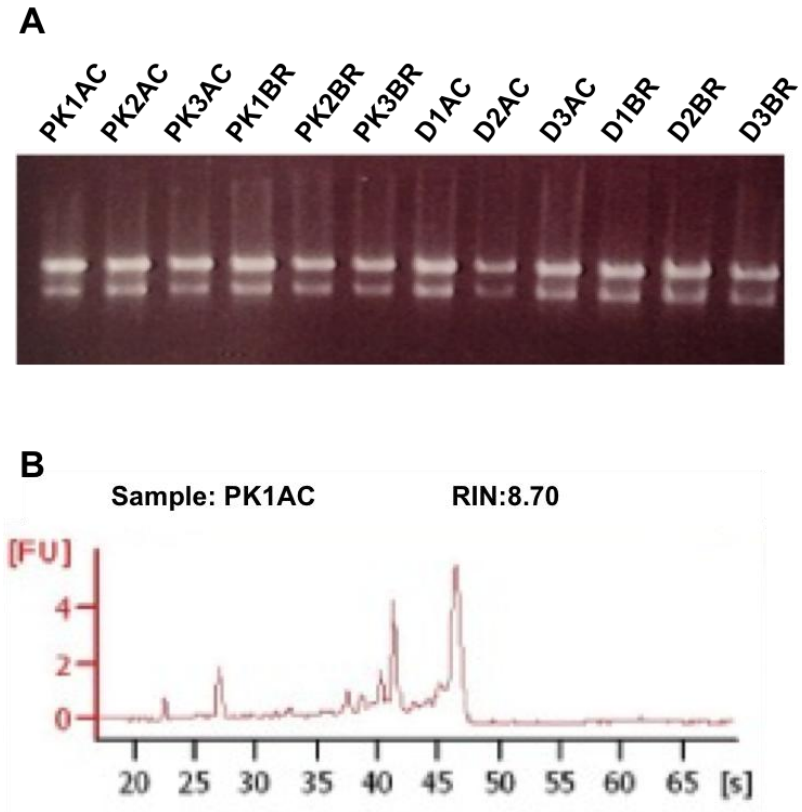


Figure 3.11: A: Electrophoresis on a 1.2% w/v agarose gel of 2 µg of isolated total RNA from tomato fruits. B: electropherogram of RNA sample PK1AC obtained by Bioanalyzer (Agilent Technologies).

To have a good quality RNA for RNA-seq experiments, there are crucial thresholds for parameters related to the absorbance (A260/A280>1.8 and A260/A230>1.8), concentration (more than 200 ng/µl), RIN (RNA Integrity Number) (>6.50) and 28s/18s ratio (>1.5).
The RNA parameters obtained are summarized in table 3.4. As reported, all the requirements have been accomplished. The lowest value of RNA concentration was 207.1 ng/µl for the sample D2BR and the highest 462.2 ng/µl for PK1AC sample; the ratio of the absorbance at 260 nm and 280 nm gave the minimum value of 1.99 for the sample PK1BR and higher than 2.00 for the other samples; the absorbance ratio A260/A230 was in a range from 2.02 and 2.23; the RNA integrity number (RIN) parameter was between 8.50 and 9.20; finally, the 28s/18s ratio, required higher than 1.5, was found from a minimum value of 1.5 (D1BR) to a maximum value of 2.9 (PK1AC).

Table 3.4: Quality parameters of RNA extracted from 12 tomato fruit samples.

Sample	Nucleic Acid Concentration	260/280	260/230	RIN	28s/18s
PK1AC	464,2 ng/µl	2,04	2,13	8,70	2,9
PK2AC	271,8 ng/µl	2,06	2,15	8,70	2,1
PK3AC	306,6 ng/µl	2,00	2,12	8,90	1,9
PK1BR	268,6 ng/µl	1,99	2,10	9,00	1,9
PK2BR	361 ng/µl	2,05	2,14	8,90	1,7
PK3BR	379,1 ng/µl	2,02	2,12	8,90	2
D1AC	242,2 ng/µl	2,16	2,12	8,90	1,9
D2AC	277,4 ng/µl	2,04	2,09	9,10	1,7
D3AC	340,8 ng/µl	2,03	2,07	9,20	1,8
D1BR	228,5 ng/µl	2,04	2,02	8,50	1,5
D2BR	207,1 ng/µl	2,16	2,08	8,50	1,8
D3BR	288,1 ng/µl	2,17	2,23	8,90	1,7

3.4.2 RNA-Seq sequences quality

RNA-seq generated around 36 millions reads of 100 bp length through a paired-end sequencing.
The raw data for each sample consist of a long list of short sequences with associated quality scores in *fastq* format. It is a text-based format for storing both a nucleotide sequence and its corresponding quality scores (http://www.asciitable.com/). An example of this type of file is reported in figure 3.12. Line 1 begins with a '@' character and is followed by a sequence identifier and an optional description; line 2 is the raw sequence nucleotides; line 3 begins with a '+' character and is optionally followed by the same sequence identifier (and any description); line 4 encodes the quality values for the sequence in line 2, and must contain the same number of symbols as nucleotides in the sequence.

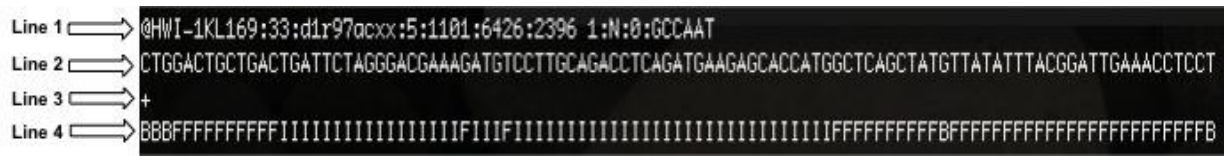


Figure 3.12: Example of *fastq* file format.

A graphical representation of the quality of the sequences is given by the FastQC software.
Figure 3.13 reports one of the outputs from FastQC. Symbols contained in the *fastq* file are translated in a graph where to each sequenced nucleotide (nucleotide number on x-axis) is assigned a position on the y-axes by the Phred quality score (Ewing et al., 1998a; Ewing et al., 1998b). The y-axis is divided in three coloured strips, red, orange and green (from the level to the top, bad to good quality, respectively) in order to make the identification of the quality easily intuitive. If most of the yellow boxes (representing groups of sequenced nucleotides) are in the green strip the sequence has a very good quality. Another way to read the sequence quality is given by the graph of the quality distribution over all sequence where it is possible to have an overview of the average of the quality among the analyzed read (Figure 3.14).
All these graphical representations are related to the base call accuracy from the Phred (Ewing et al., 1998a; Ewing et al., 1998b) score and for all the sequences obtained from the RNA sequencing of the 12 RNA samples the resulting observed Phred score was around 40 meaning a 99.99% of accuracy in base call.

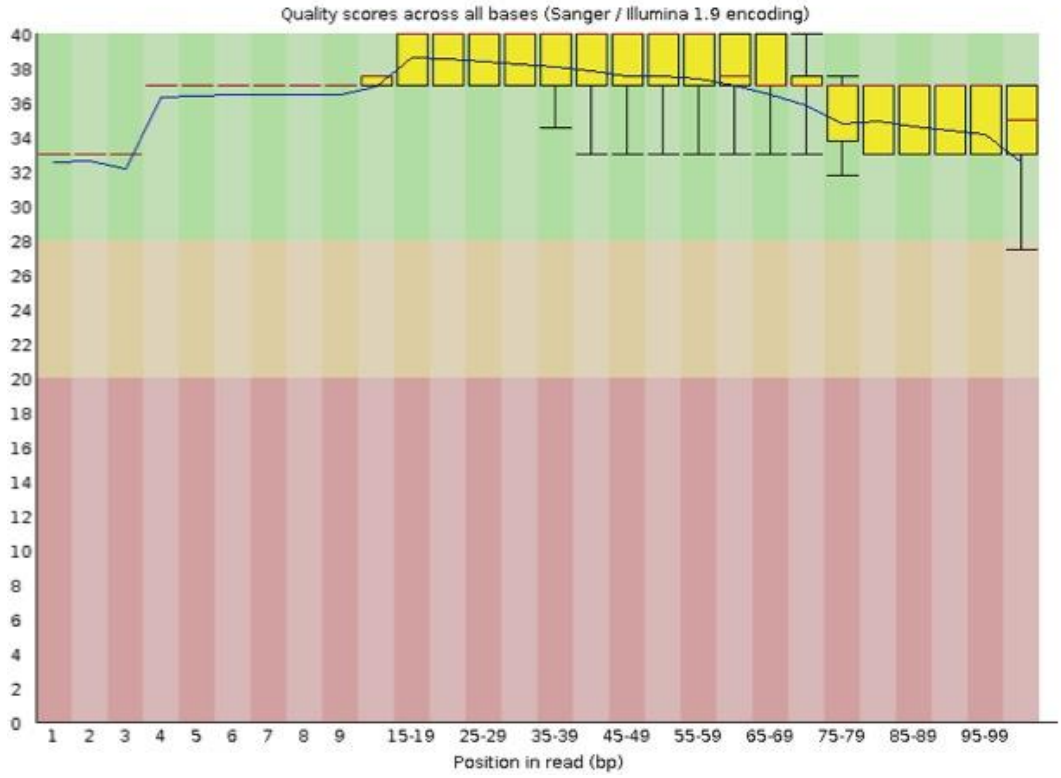


Figure 3.13: FastQC per base sequence quality graph.

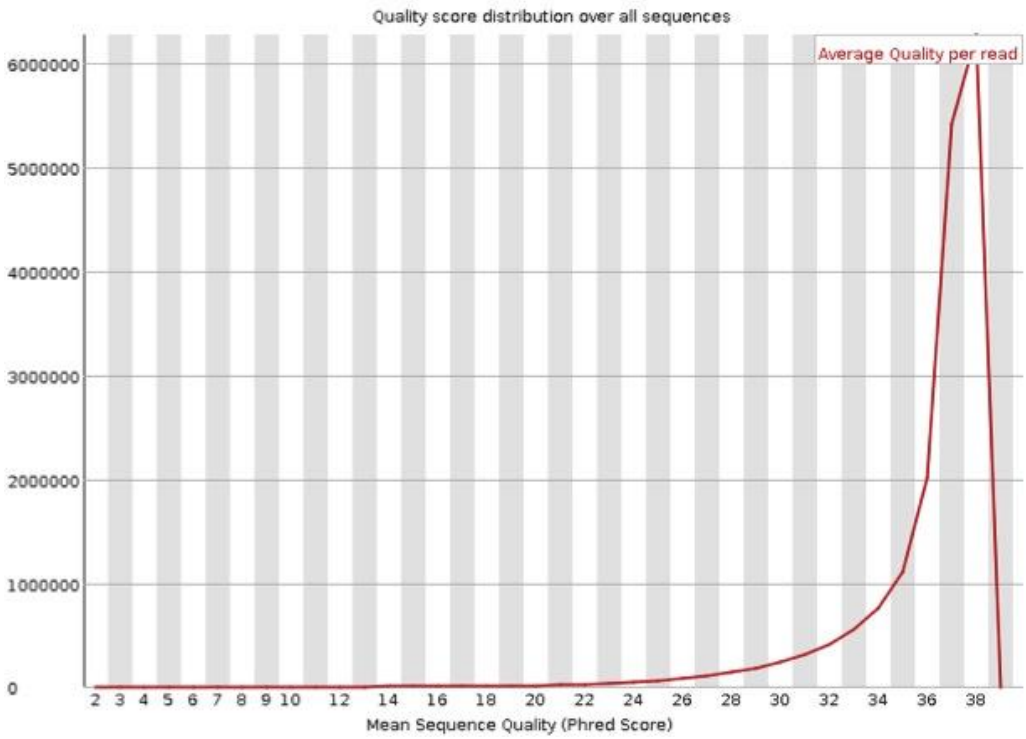


Figure 3.14: FastQC quality score distribution over all sequences graph.

3.4.3 Reads Mapping, Alignment Evaluation and Data Coverage

To use RNA-seq data to compare gene expression levels between samples, it is necessary to turn millions of short reads into a quantification of expression.
The first step in this procedure is the reads mapping. The goal of mapping is to find the unique location where each short read best matches the genome reference.
The tomato genome is available from 2012 (The Tomato Consortium, 2012; http://solgenomics.net/organism/Solanum_lycopersicum/genome). It consists of 13 sequenced chromosome (12 nuclear and 1 plastidial) for 930 Mbp with 34727 genes of which 56.6% are annotated with GO terms (19662). The total number of unique GO terms for the tomato genome is 2108.
The TopHat script (http://ccb.jhu.edu/software/tophat) was used to align RNA-Seq reads from the 12 different tomato samples to the tomato genome, in order to assign each transcript to tomato genes. To evaluate the quality of the alignment graphical outputs and tables were produced using the Qualimap software (Garcia-Alcalde et al., 2012) on the output files from TopHat. For all samples, 100% of the sequenced reads were correctly mapped. The average of the total numbers of mapped reads for the three biological replicates of ‘Kiros’ variety grown in Acerra environment (PKAC) was of 36,810132.75 and 37,312740.37 from the three biological replicates of the same varieties grown in Brusciano environment (PKBR). ‘Docet’ samples showed a number of mapped reads greater than ‘Kiros’ samples: 41,710174.47 reads for the plants grown in Acerra (DAC) and 40,486100.24 reads for those grown in Brusciano (DBR).
Count data from mapping were than normalized. The method used was the CPM (Counts Per Million), proposed in the edgeR analysis pipeline (Robinson et al., 2010)

that is the number of reads mapped to the feature (gene) per one million of aligned reads. Normalization facilitates accurate comparisons of expression levels between and within samples. Thus, the normalized coverage was used to produce box-plot graphs, graphical representation of the coverage through which is possible to verify that all samples have a comparable mean of the coverage before and after the normalization, showed in figure 3.15 A and B.

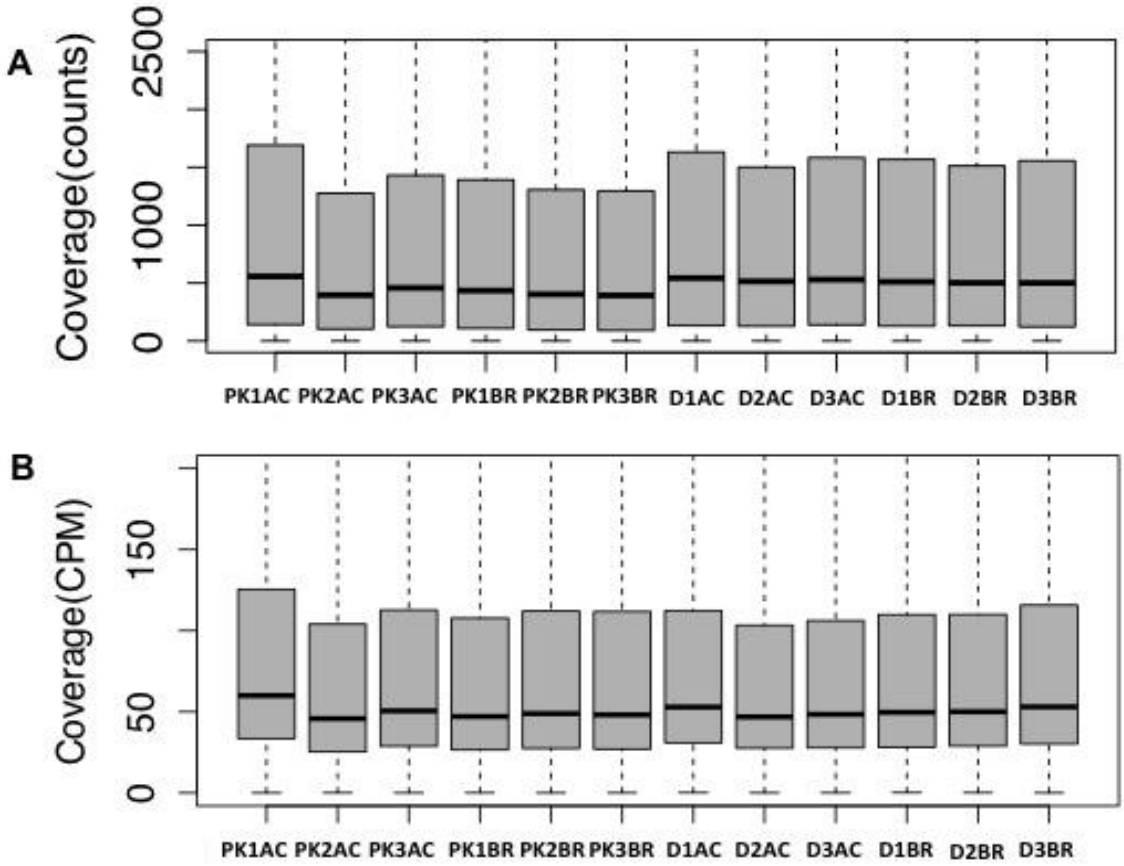


Figure 3.15: Box plots of coverage data before (A) and after (B) the normalization. For sample names see table 2.3.

3.4.4 Relationship among samples

Normalized data coverage was used for the construction of the PCA graph showed in figure 3.16. Graphical representation of the relationships among 12 tomato plants showed that samples are well grouped by the environment: plants from Acerra has been located on the left side of the graph while plants from Brusciano were all in the right side of the x-axis. This result suggests that the environmental component may have a greater influence than genotypic component.

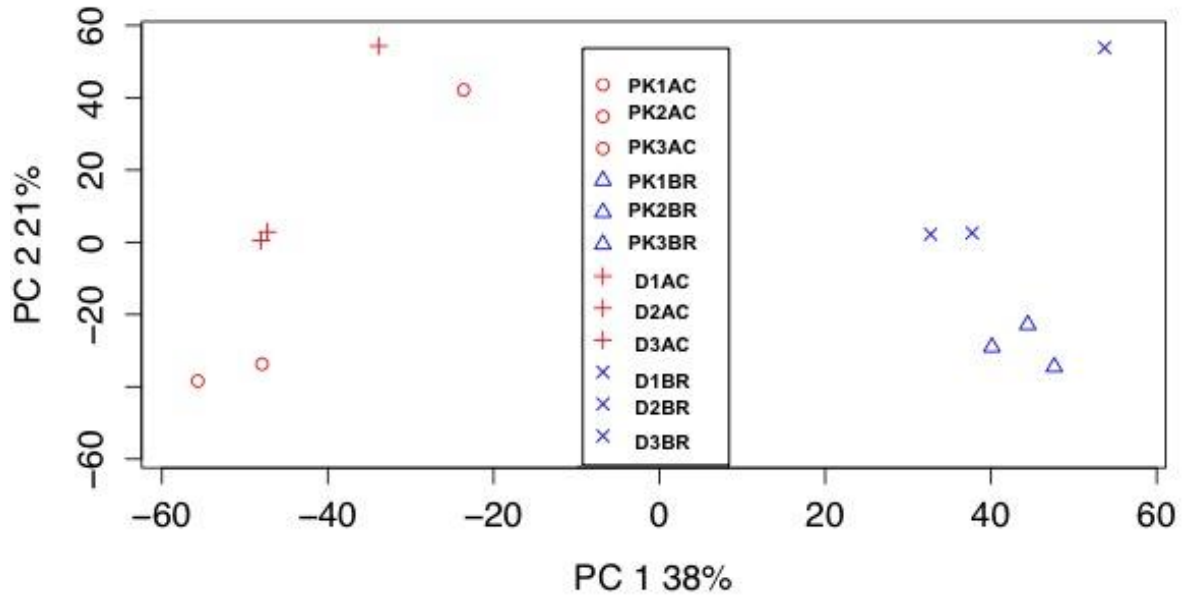


Figure 3.16**:** PCA graph for a graphic visualization of the relationship among 12 tomato samples based on normalized data coverage. For sample names see table 2.3.

3.4.5 Gene Expression

The objective of the transcriptomic analysis was to highlight transcripts that changed significantly in abundance across the selected experimental conditions.
Using the CPM normalized coverage it was possible to select the expressed genes, using a threshold of 1CPM, for both genotype in both environment.
A total number of 13,734 expressed genes were found for PKAC while PKBR showed 1,3690 expressed genes. 'Docet' genotype presented 13,709 and 13,722 expressed genes from Acerra and Brusciano, respectively.
Comparing the lists of the expressed genes, information about the number of common and specific genes for each comparison was obtained (Figure 3.17 A, B).
It is interesting to note, in both comparisons, that there is an high number of expressed common genes and, in contrast, there is a reduced number of specific genes for each sample that may be related to the environment.

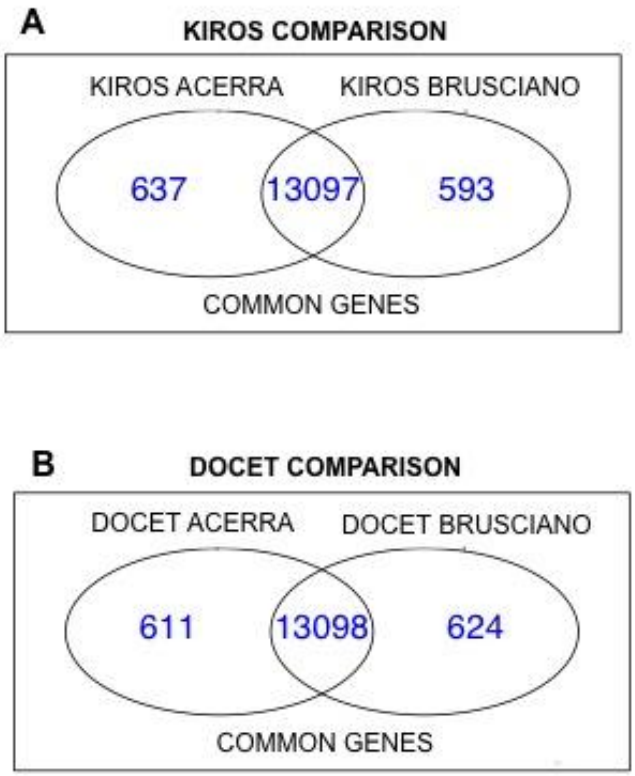


Figure 3.17: Venn diagram showing the number of overlapping and non-overlapping expressed genes in the comparisons between the two locations Acerra and Brusciano.

3.4.6 Functional Annotation of Expressed Genes

Functional enrichment of the expressed genes was performed using the Blast2GO tool (Conesa et al., 2005) at the default parameters. Outputs of functional enrichment are summarized in table 3.5. For all the analyzed samples, the most represented ontologic domain was molecular function (F), with 150, 150, 156 and 151 GO (Gene Ontology) terms for PKAC, PKBR, DAC and DBR, respectively.

Table 3.5: Summary table of functional enrichment outputs. The table reports the total number of GO terms, the number of GO terms for molecular function (F), cellular component (C) and biological process (P) domains per sample. PKAC: Kiros Acerra; PKBR: Kiros Brusciano; DAC: Docet Acerra; DBR: Docet Brusciano.

	PKAC	**PKBR**	**DAC**	**DBR**
Total Number of GO	346	355	369	350
Molecular Function (F)	150	150	156	151
Biological Process (P)	127	133	138	131
Cellular Component (C)	69	72	75	68

Comparing lists of GOs, information about the number of common and specific GO terms for each comparison was obtained.
A large number of common GOs was highlighted for both comparisons: PKAC/ PKBR had 87% GO on common while DAC/DBR 86% (Figure 3.18 A and B).

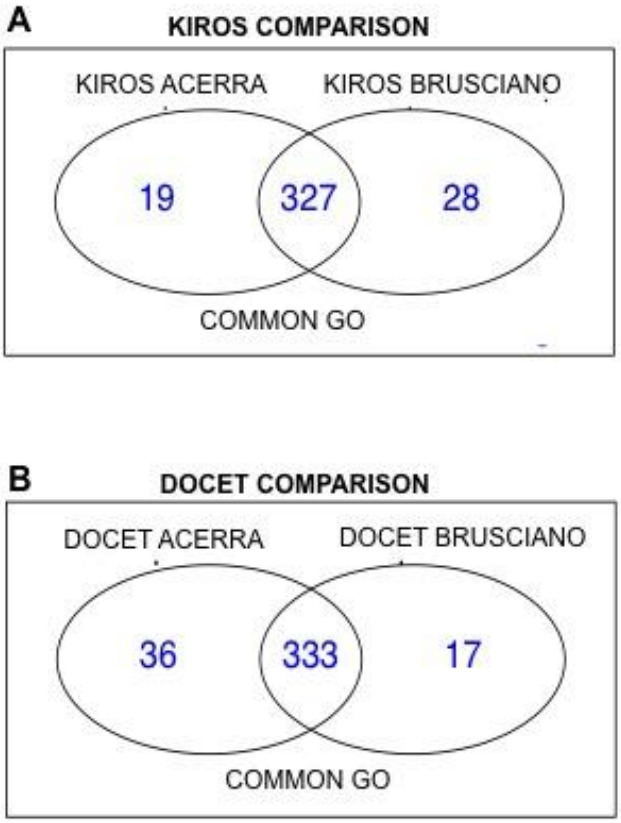


Figure 3.18: A: comparison between enriched GO terms from Kiros Acerra and Kiros Brusciano; B: comparison between enriched GO terms from Docet Acerra and Docet Brusciano.

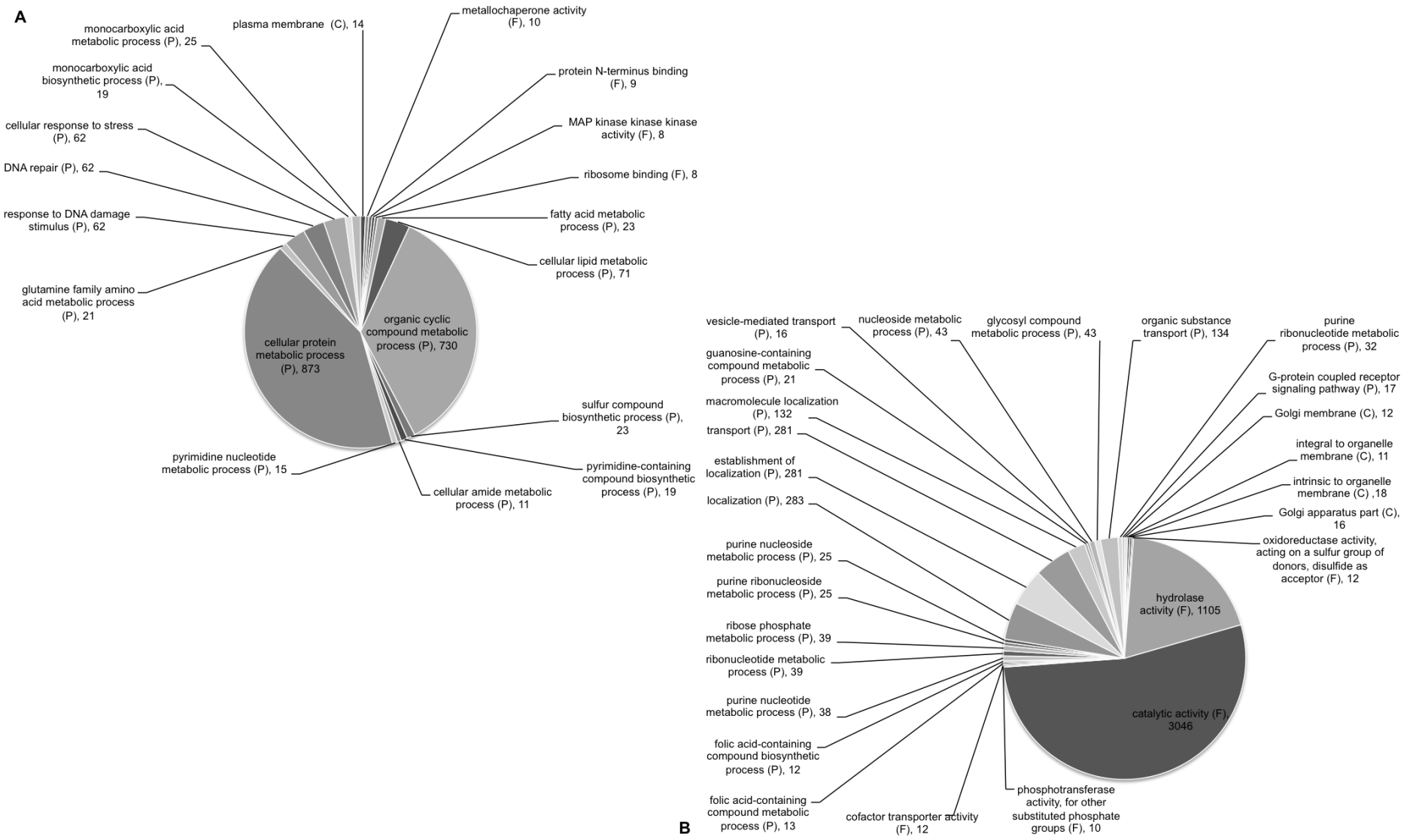
Lists of specific GOs per sample are summarized in tables 3.6 and 3.7.
Among the 19 GO terms specific for PKAC, terms “cellular protein metabolic process” and “organic cyclic compound metabolic process” included the great number of annotated genes (873 and 730 genes, respectively) (Figure 3.19 A). For PKBR the two biggest functional categories are represented by the “hydrolase activity” and “catalytic activity” with 1105 and 3046 annotated genes, respectively, both of them from the molecular function domain (Figure 3.19 B).
Similarly, for DAC “Organic cyclic compound metabolic process”, “cellular aromatic compound metabolic process”, “heterocycle metabolic process” and “cellular nitrogen compound metabolic process” showed the highest number of annotated genes (728, 702, 702 and 705 annotated genes, respectively) (Figure 3.19 C).
Finally, DBR sample showed as biggest category, among the 17 specific GO terms, the “isomerase activity” term, with 101 annotated genes (Figure 3.19 D).

Table 3.6: List of specific 19 GO terms for Kiros Acerra (PKAC) and 28 GO terms for Kiros Brusciano (PKBR) and their description (Term) and ontologic domain (Category). GOs are ordered by increasing value of FDR (False Discovery Rate) and P-value.

Sample	GO-ID	Term	Category	FDR	P-Value
	GO:0006631	fatty acid metabolic process	P	0.0089285617521523	6,46E+11
	GO:0016530	metallochaperone activity	F	0.009656079824241121	7,04E+10
	GO:0044255	cellular lipid metabolic process	P	0.017214463840834707	0.0013803579372662345
	GO:1901360	organic cyclic compound metabolic process	P	0.01730637253000509	0.0014028117187308018
	GO:0047485	protein N-terminus binding	F	0.017764997125264945	0.0014554704210705156
	GO:0044272	sulfur compound biosynthetic process	P	0.01823022112565378	0.0014988822134108135
	GO:0072528	pyrimidine-containing compound biosynthetic process	P	0.024473856728124226	0.0020406731205612006
	GO:0043603	cellular amide metabolic process	P	0.026910374323493507	0.0022751071842349244
	GO:0006220	pyrimidine nucleotide metabolic process	P	0.028947178334045837	0.0025052016926085086
PKAC	GO:0004709	MAP kinase kinase kinase activity	F	0.03266412974496479	0.0030082885325839162
	GO:0043022	ribosome binding	F	0.03266412974496479	0.0030082885325839162
	GO:0044267	cellular protein metabolic process	P	0.03504184480626451	0.003237451089015722
	GO:0009064	glutamine family amino acid metabolic process	P	0.041026807726611	0.0038380685903453638
	GO:0006974	response to DNA damage stimulus	P	0.044583516667149936	0.004218066026227345
	GO:0006281	DNA repair	P	0.044583516667149936	0.004218066026227345
	GO:0033554	cellular response to stress	P	0.044583516667149936	0.004218066026227345
	GO:0005886	plasma membrane	C	0.044614220161284586	0.004368097674129258
	GO:0072330	monocarboxylic acid biosynthetic process	P	0.04660267048039896	0.0046440197486277875
	GO:0032787	monocarboxylic acid metabolic process	P	0.048853407755010396	0.004882502111482736
	GO:0006760	folic acid-containing compound metabolic process	P	0.0012519762859766546	7,31E+10
	GO:0009396	folic acid-containing compound biosynthetic process	P	0.002470329348223506	1,52E+12
	GO:0006163	purine nucleotide metabolic process	P	0.00983107503729181	7,48E+11
	GO:0016671	oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor	F	0.013530217437022478	0.0011006568513556925
	GO:0000139	Golgi membrane	C	0.013530217437022478	0.0011006568513556925
	GO:0009259	ribonucleotide metabolic process	P	0.01810068714291899	0.0015092670569487943
	GO:0019693	ribose phosphate metabolic process	P	0.01810068714291899	0.0015092670569487943
	GO:0016787	hydrolase activity	F	0.018230891538636104	0.0015254203263007546
	GO:0031301	integral to organelle membrane	C	0.024720782084106507	0.0021259010740544817
	GO:0046128	purine ribonucleoside metabolic process	P	0.026308587034075452	0.0022930203690362102
	GO:0042278	purine nucleoside metabolic process	P	0.026308587034075452	0.0022930203690362102
	GO:0051179	localization	P	0.02892835017799913	0.0025801869566082895
	GO:0031300	intrinsic to organelle membrane	C	0.033317919271279	0.0030878606181109825
PKBR	GO:0003824	catalytic activity	F	0.03491461634350114	0.00324598408771655
	GO:0051234	establishment of localization	P	0.03599510936033935	0.003367351892512862
	GO:0006810	transport	P	0.03599510936033935	0.003367351892512862
	GO:0033036	macromolecule localization	P	0.036173243309611545	0.003394525737653843
	GO:1901068	guanosine-containing compound metabolic process	P	0.0367492145660247	0.00346992874316038
	GO:0016192	vesicle-mediated transport	P	0.03768839755563284	0.003591456826916784
	GO:0044431	Golgi apparatus part	C	0.03768839755563284	0.003591456826916784
	GO:0009116	nucleoside metabolic process	P	0.03982982290833142	0.003818664020845255
	GO:1901657	glycosyl compound metabolic process	P	0.03982982290833142	0.003818664020845255
	GO:0016780	phosphotransferase activity, for other substituted phosphate groups	F	0.04095623664721643	0.004081344907029412
	GO:0051184	cofactor transporter activity	F	0.04268588209574347	0.004303312343760309
	GO:0071702	organic substance transport	P	0.04281619737318955	0.004328889217277735
	GO:0009150	purine ribonucleotide metabolic process	P	0.04359468086108003	0.004420262527750416
	GO:0007186	G-protein coupled receptor signaling pathway	P	0.04950387070756757	0.005105715892268009
	GO:0007166	cell surface receptor signaling pathway	P	0.04950387070756757	0.005105715892268009

Table 3.7: List of specific 36 GO terms for Docet Acerra (DAC) and 17 GO terms for Docet Brusciano (DBR) and their description (Term) and ontologic domain (Category). GOs are ordered by increasing value of FDR (false discovery rate) and P-value.

Sample	GO-ID	Term	Category	FDR	P-Value
DAC	GO:0016530	metallochaperone activity	F	0.00873141544475598	6,52E+11
	GO:0008308	voltage-gated anion channel activity	F	0.008901880989262164	6,85E+11
	GO:1901360	organic cyclic compound metabolic process	P	0.009871099522814028	7,74E+11
	GO:0006974	response to DNA damage stimulus	P	0.011465433987642056	9,19E+11
	GO:0006281	DNA repair	P	0.011465433987642056	9,19E+11
	GO:0033554	cellular response to stress	P	0.011465433987642056	9,19E+11
	GO:0033554	cellular response to stress	F	0.013136079779367659	0.0010876765651132432
	GO:0000139	Golgi membrane	C	0.013136079779367659	0.0010876765651132432
	GO:0006767	water-soluble vitamin metabolic process	P	0.0143729681342675	0.001202619065272818
	GO:0006766	vitamin metabolic process	P	0.0143729681342675	0.001202619065272818
	GO:0047485	protein N-terminus binding	F	0.015952806787256992	0.0013579815190779485
	GO:0070013	intracellular organelle lumen	C	0.021761776982244404	0.0019156939063393534
	GO:0043233	organelle lumen	C	0.021761776982244404	0.0019156939063393534
	GO:0043603	cellular amide metabolic process	P	0.02350209821842239	0.002103034936453834
	GO:0015108	chloride transmembrane transporter activity	F	0.02350209821842239	0.002103034936453834
	GO:0031301	integral to organelle membrane	C	0.02350209821842239	0.002103034936453834
	GO:0072527	pyrimidine-containing compound metabolic process	P	0.023676164361693217	0.0021277581557854475
	GO:0022839	ion gated channel activity	F	0.02379713156991211	0.0021570903689170765
	GO:0022836	gated channel activity	F	0.02379713156991211	0.0021570903689170765
	GO:0042579	microbody	C	0.025027956128543728	0.0023341004989141595
	GO:0005777	peroxisome	C	0.025027956128543728	0.0023341004989141595
	GO:0015631	tubulin binding	F	0.025104491900022584	0.0023485317814663778
	GO:0031300	intrinsic to organelle membrane	C	0.031077295484393187	0.0030427218414411693
	GO:0035091	phosphatidylinositol binding	F	0.035476072483854794	0.0035043186067724085
	GO:0044431	Golgi apparatus part	C	0.035765184680558386	0.00354326786056665
	GO:0006725	cellular aromatic compound metabolic process	P	0.03635444221859329	0.003612207797431408
	GO:0005083	small GTPase regulator activity	F	0.03898817203361654	0.0038852245809211142
	GO:0046483	heterocycle metabolic process	P	0.039077106292050384	0.003925391514508793
	GO:0072330	monocarboxylic acid biosynthetic process	P	0.03984665485236732	0.004167575754460266
	GO:0045454	cell redox homeostasis	P	0.040680828098271865	0.004278460131195356
	GO:0032787	monocarboxylic acid metabolic process	P	0.0407360349117735	0.0042961012995275365
	GO:0044272	sulfur compound biosynthetic process	P	0.04128933563672357	0.004366449207369953
	GO:0034641	cellular nitrogen compound metabolic process	P	0.04332436563307686	0.004606832603633391
	GO:0005253	anion channel activity	F	0.0471053399150186	0.005036247846811983
	GO:0006633	fatty acid biosynthetic process	P	0.0471053399150186	0.005036247846811983
	GO:0005216	ion channel activity	F	0.04858444692761096	0.005208501137794434
DBR	GO:0004222	metalloendopeptidase activity	F	0.00457593394838668	3,02E+11
	GO:0016853	isomerase activity	F	0.007164709457916596	5,14E+11
	GO:0003727	single-stranded RNA binding	F	0.007164709457916596	5,16E+11
	GO:0009064	glutamine family amino acid metabolic process	P	0.01276376970128936	0.0010049336400492204
	GO:0051087	chaperone binding	F	0.016149061756519454	0.0013324618067552 37
	GO:0009150	purine ribonucleotide metabolic process	P	0.022105965938606073	0.001862501488145195
	GO:0046128	purine ribonucleoside metabolic process	P	0.02727382154268933	0.0024100081971120757
	GO:0042278	purine nucleoside metabolic process	P	0.02727382154268933	0.0024100081971120757
	GO:0030118	clathrin coat	C	0.03123330951334713	0.002921884271730905
	GO:1901068	guanosine-containing compound metabolic process	P	0.038495912048446976	0.003623670977250674
	GO:0043648	dicarboxylic acid metabolic process	P	0.039203852975503606	0.0037244799311416847
	GO:0016192	vesicle-mediated transport	P	0.039203852975503606	0.0037244799311416847
	GO:0009116	nucleoside metabolic process	P	0.04192593617173519	0.004080257570939345
	GO:1901657	glycosyl compound metabolic process	P	0.04192593617173519	0.004080257570939345
	GO:0016780	phosphotransferase activity, for other substituted phosphate groups	F	0.04192593617173519	0.004190157479104273
	GO:0051184	cofactor transporter activity	F	0.043721422591471854	0.004433113446956327
	GO:0016160	amylase activity	F	0.043721422591471854	0.004433113446956327


A
plasma membrane (C), 14
metallochaperone activity (F), 10
monocarboxylic acid metabolic process (P), 25
protein N-terminus binding (F), 9
monocarboxylic acid biosynthetic process (P), 19
MAP kinase kinase kinase activity (F), 8
cellular response to stress (P), 62
ribosome binding (F), 8
DNA repair (P), 62
fatty acid metabolic process (P), 23
response to DNA damage stimulus (P), 62
cellular lipid metabolic process (P), 71
glutamine family amino acid metabolic process (P), 21
organic cyclic compound metabolic process (P), 730
cellular protein metabolic process (P), 873
sulfur compound biosynthetic process (P), 23
pyrimidine nucleotide metabolic process (P), 15
pyrimidine-containing compound biosynthetic process (P), 19
cellular amide metabolic process (P), 11
B
vesicle-mediated transport (P), 16
nucleoside metabolic process (P), 43
glycosyl compound metabolic process (P), 43
organic substance transport (P), 134
purine ribonucleotide metabolic process (P), 32
guanosine-containing compound metabolic process (P), 21
G-protein coupled receptor signaling pathway (P), 17
macromolecule localization (P), 132
Golgi membrane (C), 12
transport (P), 281
integral to organelle membrane (C), 11
establishment of localization (P), 281
intrinsic to organelle membrane (C) ,18
localization (P), 283
Golgi apparatus part (C), 16
oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor (F), 12
purine nucleoside metabolic process (P), 25
purine ribonucleoside metabolic process (P), 25
hydrolase activity (F), 1105
ribose phosphate metabolic process (P), 39
ribonucleotide metabolic process (P), 39
catalytic activity (F), 3046
purine nucleotide metabolic process (P), 38
folic acid-containing compound biosynthetic process (P), 12
folic acid-containing compound metabolic process (P), 13
cofactor transporter activity (F), 12
phosphotransferase activity, for other substituted phosphate groups (F), 10

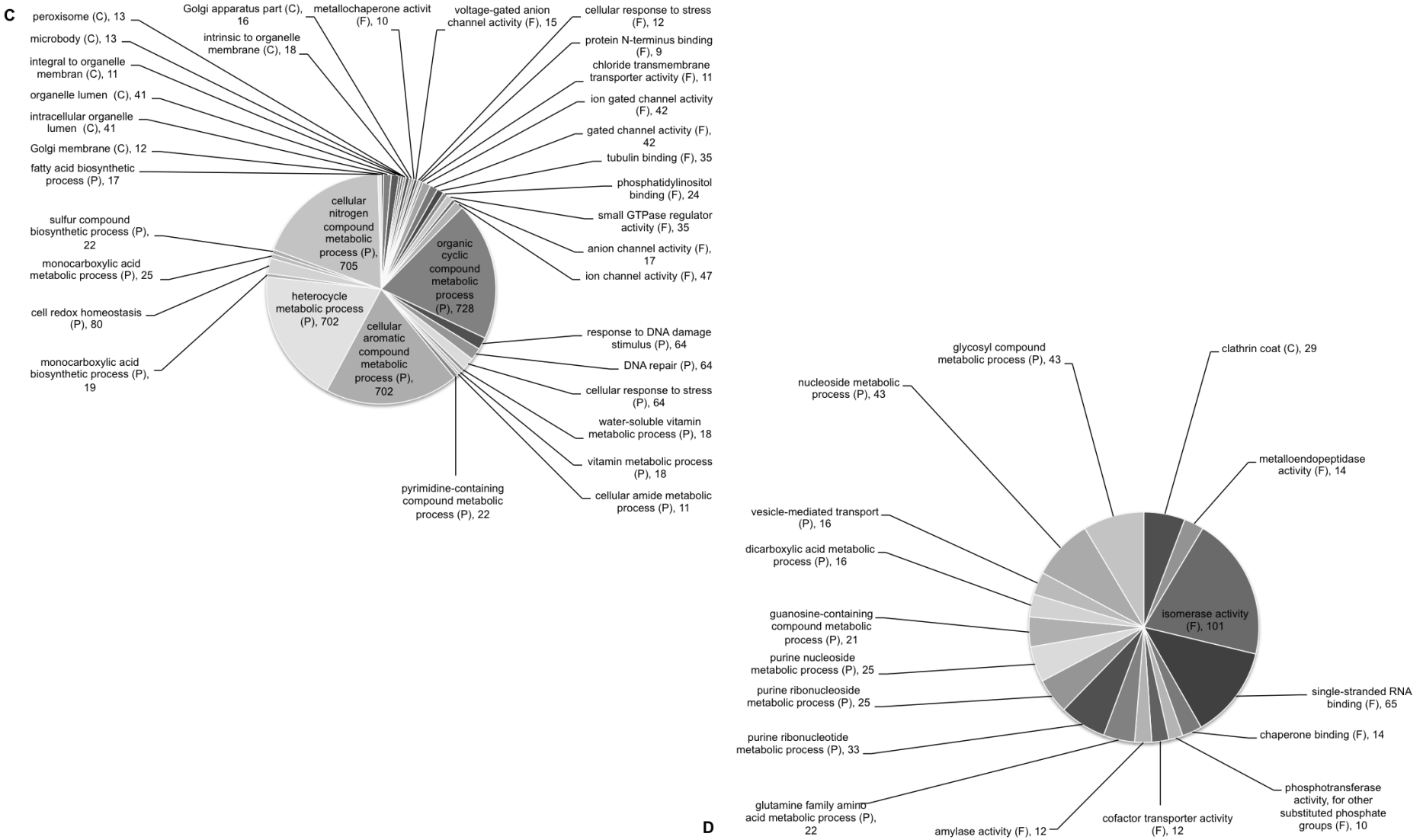


Figure 3.19: Graphical representation of the distribution of GO terms among the three ontologic domains (C) cellular component, (F) molecular function and (P) biological process for the 19 specific GO terms of PKAC (A), 28 specific GO terms of PKBR (B), 36 specific GO terms of DBR (C) and 17 specific GO terms of DBR (D). Numbers after the description represent genes included in GOs.

3.4.7 Differentially Expressed Gene Analysis

To determine the differentially expressed genes (DEG) among samples from the same genotype grown in different environments the NOISeq package (Tarazona et al., 2011) was used. NOISeq is a nonparametric approach for the identification of differentially expressed genes from count data that aims to be robust against the number of available reads. NOISeq creates a null or noise distribution of count changes by contrasting fold-change differences (M) and absolute expression differences (D) for all the genes in samples within the same condition. This reference distribution is then used to assess whether the (M, D) values computed between two conditions for a given gene are likely to be part of the noise or represent a true differential expression (Tarazona et al., 2011).
From lists of specific GO terms were extrapolated genes present exclusively in each samples for specific environment (i.e. expressed genes – higher than 1CPM – in one sample that are not present in the list of expressed genes of the same genotype harvested in the other environment). Then, tables with counts for each selected gene were created and analyzed by NOISeq. The DEG analysis was done by comparing counts of selected expressed genes from one genotype of one environment against counts (of the same genes) from the same genotype in the different experimental condition.
As a consequence, the outputs of the analysis were lists of exclusively up-regulated genes per sample.
Summary about number of total expressed genes and specific expressed genes per sample, number of total GO terms and specific GO terms, number genes environment specific from specific GO and number of up-regulated genes environment specific are summarized in table 3.8.

Table 3.8: Summary table of the expressed genes analysis.

	Kiros comparison		Docet comparison	
Sample	**PKAC**	**PKBR**	**DAC**	**DBR**
Total Expressed Genes	13743	13690	13709	13722
Specific Expressed Genes	637	593	611	624
Total GO terms	346	355	369	350
Specific GO terms	19	28	36	17
Genes environment specific from specific GOs	108	95	62	23
Up-regulated genes environment specific	97	30	43	9

From 19 specific GO terms of PKAC, a total of 108 genes environment specific were selected. When the NOISeq analysis was performed, the list of 108 genes was reduced to 97 up-regulated genes. These genes belong to two ontologic domains, molecular function (4,3%) and biological process (95,7%). Genes from biological process functional domain represent 10 classes of GO terms distributed as reported in figure 3.20 A.
Similarly, 95 unique genes from 28 specific GO terms were selected for PKBR and, after the NOISeq use, reduced to 30 up-regulated genes. The most represented ontologic domain is the molecular function (F) with 56.15% of terms, divided in “catalytic activity” (50%), “hydrolase activity” (46,9%) and “phosphotransferase activity” (3.1%) (Figure 3.20 B).

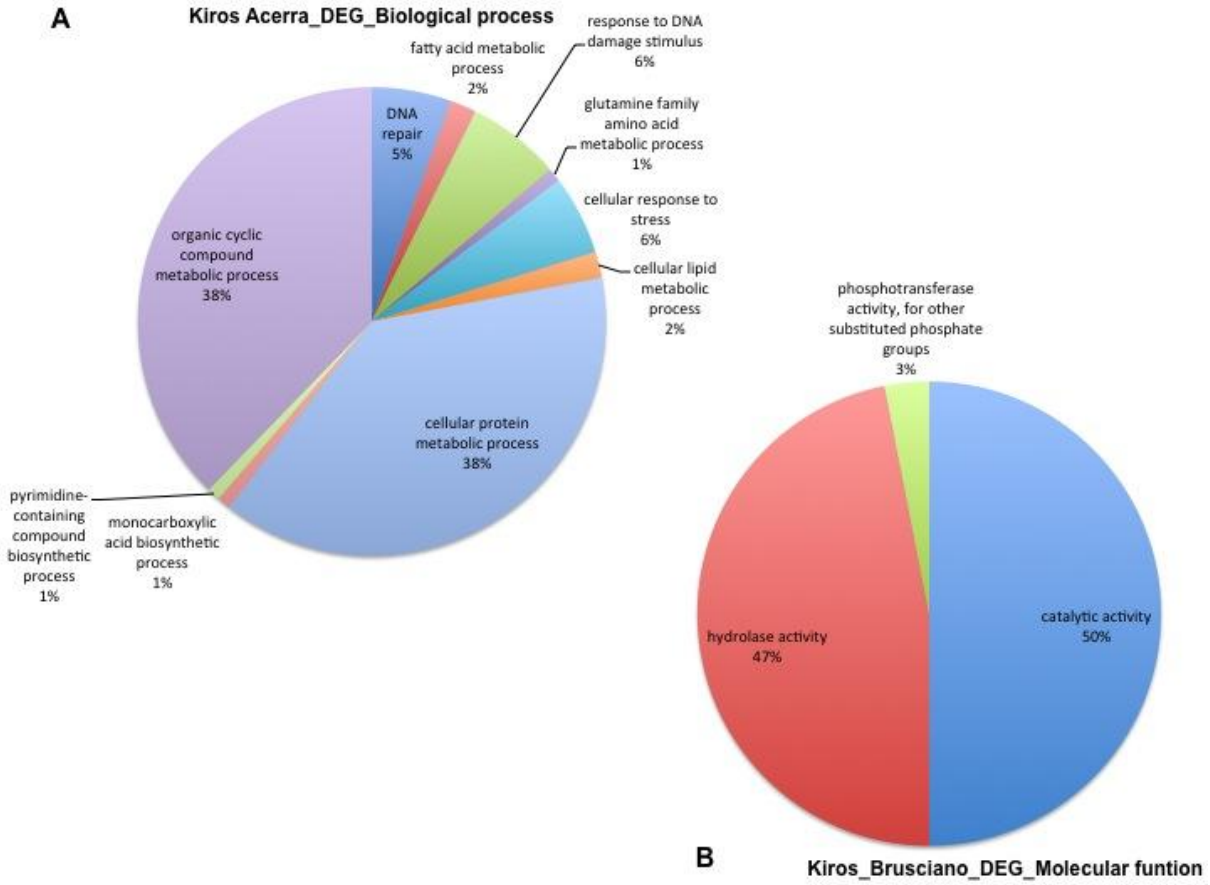


Figure 3.20: GO terms distribution for the most represented functional domain of differentially expressed genes from Kiros Acerra (A) and Kiros Brusciano (B) from.

From 36 specific GO terms DAC, 62 genes environment specific were isolated. The NOISeq analysis considered 43 differentially expressed genes up-regulated. The 81% of the GO terms associated to these DEG are related to the biological process domain (Figure 3.21 A).
Seventeen specific GO terms for DBR revealed 23 genes environment specific and, after the NOISeq procedure, 9 gene were considered up-regulated. These differentially expressed genes belong to two ontologic domains, molecular function and biological process. The biological process domain is represented by 7 classes of GO terms distributed as reported in figure 3.21 B.

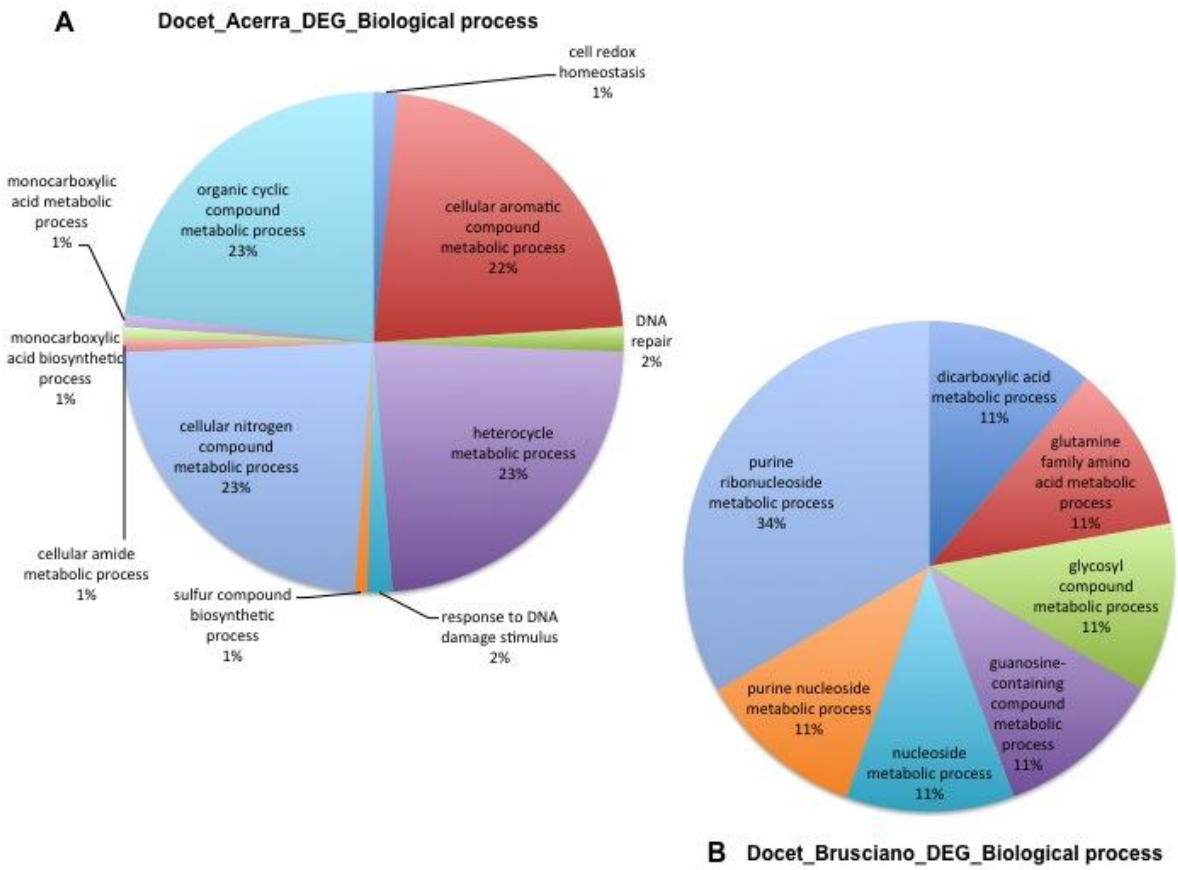


Figure 3.21: GO terms distribution for the most represented functional domain of differentially expressed genes from Docet Acerra (A) and Docet Brusciano (B).

According to the significant test conduced by NOISeq, genes with the highest difference in expression values and with a “prob-value” equal to 1 (the probability of differential expression is equivalent to 1-FDR, False Discoverate Rate) were selected and reported in table 3.9.
PKAC showed the highest number (16) of genes, followed by DAC, with three genes. Both samples from Brusciano, PKBR and DBR, showed only one gene with prob-value equal to 1.
Interestingly, three genes, Solyc06g04930.2.1, Solyc10g005040.2.1 and Solyc12g010800.1.1, were highlighted in common between PKAC and DAC samples. Finally, among the 16 genes of PKAC, two (Solyc02g071860.2.1 and Solyc07g066230.2.1) presented the same function of “Receptor like kinase (RLK).

Table 3.9: Table of up-regulated genes with prob-value equal to 1. Sample. Average of expression value: mean of the counts registered for each gene in both environments (Acerra and Brusciano); theta: statistical parameter from NOISeq analysis, it is a positive number when refers to first column of the average of expression value (Acerra) and negative for the second (Brusciano); prob-value: from NOISEq, probability of differential expression that is equivalent to 1-FDR; ITAG description: description of gene function from International Tomato Annotation Group.

Sample	Gene ID	Average of expression value		theta	prob-value	ITAG description
		Acerra	Brusciano			
PKAC	Solyc01g095510.2.1	10,00421739	0,951966782	3,05910436	1	Adenine nucleotide translocator 1
	Solyc11g073160.1.1	5,226612102	0,489307208	2,52329988	1	GTP cyclohydrolase/3 4-dihydroxy-2-butanone 4 phosphate synthase-Riboflavin biosynthesis protein (RibA)
	Solyc12g010800.1.1	3,065058773	0,099684574	2,49630506	1	BZIP transcription factor family protein-Basic leucine zipper
	Solyc10g005040.2.1	6,988922926	0,781684358	1,95075577	1	Harpin-induced protein-like
	Solyc06g048930.2.1	12,94094855	0,237866361	2,70579926	1	Two-component response regulator ARR3-Signal transduction
	Solyc06g069230.2.1	2,609037744	0,39422975	2,0110067	1	DNA mismatch repair protein muts (MSH2)
	Solyc04g007000.1.1	17,78832711	1,003488741	2,1829424	1	Ethylene-responsive transcription factor 4-Transcriptional factor B3
	Solyc08g074270.2.1	6,472892389	0,479346099	2,37286961	1	Cryptochrome 3
	Solyc01g079500.2.1	9,923620423	1,235006488	2,20459957	1	DNA replication licensing factor (MCM protein 7)
	Solyc07g018300.2.1	9,43147622	0,898301438	2,17376453	1	Single-stranded DNA binding protein p30 subunit- Replication protein subunit RPA32
	Solyc06g050270.1.1	7,754493749	0,365988912	2,53256724	1	CBL-interacting protein kinase 13-Serine/threonine protein kinase
	Solyc02g071860.2.1	5,041733719	0,96412056	1,82614989	1	Receptor like kinase (RLK)
	Solyc08g068390.2.1	4,580728376	0,792146363	1,80604979	1	Fatty acid oxidation complex subunit alpha-3-hydroxyacyl-CoA dehydrogenase NAD binding
	Solyc07g066230.2.1	6,765895677	1,208359736	1,79118067	1	Receptor like kinase (RLK)
	Solyc01g097970.2.1	4,538126011	0,777126526	1,78036008	1	DeoxyUTP pyrophosphatase subfamily 1
PKBR	Solyc06g083480.2.1	1,019448321	17,52321543	-2,0965339	1	Glucose/ribitol dehydrogenase
DAC	Solyc06g048930.2.1	3,15939533	0,081599231	3,83710013	1	Two-component response regulator ARR3-Signal transduction
	Solyc12g010800.1.1	2,947939065	0,074368487	2,62195704	1	BZIP transcription factor family protein-Basic leucine zipper
	Solyc10g005040.2.1	7,522101591	0,832112354	2,50364099	1	Harpin-induced protein-like
DBR	Solyc10g084690.1.1	0,597683816	4,511678282	-1,6648717	1	ADP-ribosylation factor GTPase activating protein 1-Arf GTPase activating protein

4. DISCUSSIONS

Technologies and biotechnologies able to depict the genetic identity of primary products in food chain are necessary to ensure the content of genetic material along the chain included those present in the processed foods.
Information on a food product is essential for consumers to let them choosing one food product over another. The choice can reflect lifestyle or religious (e.g. vegetarianism, preference for organic products, absence of pork for Jews and Muslims), or health concerns (e.g. absence of peanuts, lactose or gluten for individuals with particular allergies) but also the consumer's interest for premium quality products (PDO, PGI). Erroneous description and mislabelling of a food product are illegal, particularly if the food has been processed removing the ability to distinguish the components.
The occurrence of food safety incidents in recent years has been associated with reduced consumer confidence in food safety (Wilcock et al., 2004; Houghton et al., 2008). Chemical and microbial contaminants always represented an important food safety issue (Tent, 1999; Wilcock et al., 2004), but nowadays the knowledge of the genetic origin of the raw materials used for the production of a food product has become really important for consumers.
For these reasons, requests for food traceability and the food genomics, i.e. set of technologies and methods of control that allows following an agricultural product, from farm to table at every stage of preparation and processing, have been increased. The identification can be done in a generic way by adopting a logo evocative of the chain, with the possibility of going back to the same supply chain companies, or by indicating on the final pack all the companies of the chain (in the case the number is limited), or using barcodes (to enter a lot of information occupying a limited space).
Thanks to the recent advancements in molecular biology and genomic studies, genetic molecular markers have become the most effective and rapid instrument in the analysis of the DNA of plant cultivars and animal breeds, and are also used to track the raw materials in food industry processes (Kumar et al., 2009; Mafra et al., 2008; Woolfe & Primrose, 2004; Caramante et al., 2010).
The combinations of variable non-coding and relatively conserved coding regions of the plastidial genome have been proposed as new tool for species discrimination and discovery (Chase et al., 2007; Fazekas et al., 2009; Kress et al. 2009; Burgess et al., 2011). DNA barcoding (www.barcoding.si.edu) is based on the designation of mtDNA fragment to act as a ''barcode" to identify and delineate species. This approach is successfully employed in animals using a portion of the cytochrome oxidase 1 (cox1 or CO1) (Hebert et al., 2003; Hebert et al., 2004; Terol et al., 2002; Espineira et al., 2008; Jérôme et al., 2008; Barbuto et al., 2010). For plant barcode the majority preference is represented by a core-barcode, consisting of portions of two plastid-coding regions, rbcL+matK, to be supplemented with additional markers as required (CBOL Plant Working Group, 2009; Hollingsworth et al., 2011). The main limitation of DNA-barcoding use, in particular for plants, is represented by the low power in the intra-species discrimination (Nicolè et al., 2012).
On the other hand, DNA molecular markers (PCR-based) are extremely sensitive, often faster than other technologies, and are widely used in agriculture, zootechny and food control (Grassi et al., 2006; Labra et al., 2004; Mane et al. 2006; Teletchea et al., 2005; Caramante et al., 2010; Pasqualone et al. 2007; Sonnante et al., 2009; Corrado et al., 2011).

Nowadays, genetic traceability based on DNA markers is extensively widespread because DNA analysis furnishes different level of identification, such as individual, breed and species discrimination, all of them able to detect fraud and protect typical productions. Moreover, the newest frontier in the molecular analysis is represented by the new sequencing technologies that allow to obtain, in a relatively short time, information about the entire sequence of a genome or transcriptomic information that can provide a picture of the differences between two or more individuals, helping in species identification, individual discrimination and, why not, in track row materials in food chain.

4.1 Tomato Diversity

Tomato has been subject to intensive selection through domestication and breeding and nowadays cultivated varieties have a limited genetic base (Miller and Tanksley, 1990; Williams and St. Clair, 1993; Park et al., 2004). In contrast, for morphological traits like fruit size and shape cultivated genotypes show greater variation than wild species. Since 1930, introduction of genes stress resistance from wild species has been practised and has expanded the genetic diversity in modern varieties (Williams and St. Clair, 1993; Park et al., 2004; Sim et al., 2009). This process has led to genetic differentiation in contemporary tomato varieties (Sim et al., 2011; Sim et al., 2012).

Traditional agronomic productions, as those from Campania Region, are characterized by a rich genetic diversity. Their correct identification and protection are important tools to preserve local economies, production and processing. In addition, the existing genetic diversity in plants adapted to the soil and climatic condition of the region is an important source for genetic improvement and possible expansion of the range of varieties in response to new consumer demands.

But, markets globalization and industrialization of production processes resulted not only in the disappearance of some species and varieties, with a consequent lost of genetic variability, but also in more difficulties for consumers to check local supply chains. Therefore, the genetic traceability is important for the protection and conservation of plant genetic resources used in agriculture (genotypes, accessions and varieties with premium quality).

The possibility of verifying the genetic origin of food products increases the value of quality certification, such as PGI and PDO (Rao et al., 2009, Arana et al., 2002; Scarano et al., 2011; Pasqualone et al., 2007, Sonnante et al., 2009), encouraging the development of local economies, through the characterization and the promotion of local products, and provides incentives for the conservation of local ecotypes preserving biodiversity.

A collection of 145 tomato genotypes, including local varieties from Campania Region and industrial varieties, has been analyzed through SSR.

An interesting thing in this study of tomato diversity is that the selected set of SSR markers has been able not only to discriminate tomato genotypes but also to separate genotypes because of their major differences, highlighting how genotypes belonging to the same variety, with the same genetic features, cluster together. That is that the allelic composition at the SSR loci revealed to be so different between groups of genotypes to verify that genotypes with similar characteristics are grouped.

The identification of subpopulations with distinct features was supported by previous studies (within commercial European greenhouse cultivars, van Berloo et al., 2008;

within Italian tomato landraces, Mazzucato et al., 2008). Recently, by using a custom-made Illumina SNP-panel more than 200 tomato genotypes were analyzed revealing that the use of SNP marker can be useful to perform a population structure analysis through which separate landraces and varieties and identify subgroups within them (Corrado et al., 2013). Similarly, a collection of 426 tomato genotypes was examined by SNP array highlighting the possibility to identify subgroups (Sim et al., 2012).
Varieties and ecotypes evaluated in this work grouped distinguishing between industrial varieties from local.
Within groups, genotypes are related each other for some peculiarities. For example, 'Sorrento' genotypes are very close together, indicating that the name of each of them is related to a common genetic origin and the resulting phenotype of the plants. Similarly, for 'Vesuviano' variety a group in which, in some cases, the genotypes have extreme genetic similarity or complete identity for all SSR analyzed loci, has been identified. Interestingly, among local cultivated varieties from Campania Region it has been possible to identify a subgroup of 'San Marzano' types. Furthermore, industrial varieties could be divided in subgroups reflecting their different market classes and fruit shape.
Therefore, understanding how genetic variation is distributed within and among populations is important to germplasm management and crop breeding and the use of DNA-based markers offers an approach for population studies.

4.2 SSR Markers Efficiency for Tomato Varietal Discrimination

Traditionally, morphological markers based on the measurement of phenotypic data (i.e. colour, shape and size of fruits and leaves) were used to perform varietal identification. The International Plant Genetic Resources Institute (IPGRI, www.bioversityinternational.org) has provided a set of morphological and physiological descriptors used as a tool for varietal identification and protection. Their assessment is often complicate because of epistatic interactions, pleiotropic effects and interactions with the environment. In fact, morphological markers evaluate aspects related to the vegetative cycle and the physiological behaviour of individuals, such as the vegetative period of growth, flowering stages, fruit ripening, leaf fall, resistance to pests, tolerance to adverse environmental and the production that are parameters strictly related to the environmental conditions (Kalloo, 1991).
Another type of markers is represented by biochemical markers that investigate the polymorphism of gene products of primary or secondary metabolism, such as isoenzymes, terpenes, anthocyanin, flavonoids, storage proteins and enzymatic proteins (Tanksley, 1987). Despite many benefits relative to their use, biochemical parameters are limited in number and often do not appear to be polymorphic between relatively close genotypes (Foolad et al., 1993; Trakesley and Orton, 1983).
In contrast to the morphological, physiological and biochemical systems (influenced by abiotic/biotic factors), the identification of a variety based on DNA analysis offers a powerful tool to control and authenticate cultivar because DNA, that characterizes each variety, is a molecule more resistant to degradation than others (Martinez et al., 2003).
Among the different classes of DNA molecular markers, sequences with high repetition in a genome, such as microsatellites, result to be interesting for the purposes of the molecular fingerprint.

In recent years it has been shown how the analysis of these regions can be a tool to increase the value of quality labels (Rao et al., 2009) and for traceability in the food chain (Caramante et al., 2010, Alba et al., 2009). Microsatellites, also known as Simple Sequence Repeats (SSR) or short tandem repeats (STR), represent specific sequences of DNA consisting of repeated units representing the variable core (Hancock 1999; Jeffreys et al., 1985). The variability is detected on the length of the amplified fragments that are different in size because the core unit motif could be repeated differently among the different alleles.
In order to characterize and evaluate the relationships among local tomato varieties cultivated in Campania Region and between these and tomato varieties widely used in the tomato market, a collection of 145 tomato genotypes was analyzed by the use of 13 SSR molecular markers (He et al., 2003; Smulders et al., 1997).
Data obtained from the SSR analysis on 145 tomato genotypes revealed 71 alleles with an average of 5.462 alleles per locus and the maximum number of alleles (18) was found at the locus LEEF1Aa. This locus has a complex composed core so the number of repetitions for each part of the core could be different among the alleles.
The average of allele number was similar to those reported in previous studies (Mazzucato et al., 2008; Bredemeijer et al., 2002). Furthermore, the possibility to distinguish genotypes is not related exclusively to high number of alleles (He et al., 2003).
Despite the large alleles number and its average, for the analyzed tomato collection a low percentage of heterozigosity was found with an average value of 8.6% in contrast with the expected heterozigosity that had a mean of 48.3%. The whole studied tomato population is composed by a large group of local ecotypes from Campania region (more than 60% of samples).
Tomato is a self-pollinating species and local ecotypes are reproduced through self-fertilization, increasing the homozygosity level at all loci. That is why, as SSR data showed, loci with a consistent number of alleles, such as the locus LEaat002, LEat002, LEta003 and LEta015 with 4, 4, 6 and 7 alleles respectively, reported a heterozigosity value from 6.5 to 7.9%, meaning that a few percentage of the analyzed tomato genotypes are heterozygote at the considered loci. There were, moreover, SSR loci with homozigosity percentage equal to 100%, such as LEcaa001 and LEct001. For these loci, that showed 2 and 5 alleles among samples, is possible to say that analyzed genotypes are homozygotes for different alleles at the same locus and share a low level of genetic diversity.
However, the discrimination power values was between 0.2 and 0.4 for six loci and greater than 0.5 for 7 SSR loci, demonstrating that the set of molecular markers selected has a good ability to discriminate among the considered tomato population. In other words, the 72% of the analyzed genotypes showed a unique allelic profile able to identify and distinguish varieties from each other.
Closing, despite the discrimination power of SSR markers in tomato is lower than in other self-pollinating species (He et al., 2003; Smulders et al., 1997), the present study shows that selected SSR markers can be used not only for intra-species but also for intra-varietal discrimination, efficiently.

4.3 The ‘San Marzano’ case

To the ‘San Marzano' tomato variety was attributed the PDO label (Protected Designation of Origin, 1996) by the European Union. This label is a legal protection

for food and food products whose exhibit peculiar characteristics strictly related to the production area (Reg. CE 1263/96).
'San Marzano' tomatoes receive the PDO label exclusively when conditions and requirements of the PDO procedure guideline for production and processing are fulfilled. Different ‘San Marzano’ types, traditional and new, these last from the genetic improvement of the traditional ‘San Marzano’ ecotypes, exhibiting characteristics corresponding to the standards described are included in the ‘San Marzano’ procedure guideline (Reg. CE 1263/96). Specifically, the procedure guideline included as main genotypes ‘Kiros’ and ‘San Marzano 2’.
The main peculiarity of this variety are from berries: the typical sweet-and-sour flavour, the elongated shape with parallel longitudinal depressions, the bright red colour, the low number of seeds and placental fibres, the easily removable peel (Article 5 - Reg. CE 1263/96). These parameters, together with the chemical and organoleptic characteristics, make ‘San Marzano’ tomato unique both for fresh and transformed use.
Unfortunately, frauds of ‘San Marzano’ tomato can be generated by the replacement of the berries with berries from cultivars easier to cultivate and more economic. This substitution can be easily brought to light by the use of molecular markers SSR.
In the present study, forty 'San Marzano' types were analyzed to study the genetic relationships and features of this premium variety in order to protect the PDO label, both for fresh and processed products.
Despite the large number of ecotypes, the mean number of alleles per locus was 1.9, which is lower than the mean obtained from the analysis of the entire tomato population in this study. These data, together with the values of the discrimination power, which are also lower, suggest that the ‘San Marzano’ analyzed samples show a low level of genetic variability. Results indicate that ‘San Marzano’ types belong to a homogeneous population constituted, in some cases, from plants belonging to the same genotype whose seed was selected and produced in different years.
Furthermore, an interesting point is represented by the observed heterozigosity. This parameter refers to the percentage of heterozygote individuals revealed from each SSR locus analyzed. All SSR loci showed 0% of heterozygosis implying that all ‘San Marzano’ types were homozygotes to the selected loci.
This information is not surprising at all when a pool of tomato ecotypes from the most preserved and jealously guarded local varieties is analyzed.
As previously mentioned, tomato is a strongly self-pollinating species and this characteristic in the case of the ‘San Marzano’ tomato is a key point for preservation of the organoleptic and phenotypic characteristics that have made it one of the best known products in the world.
Despite the low observed variability and the presence of one allele per sample at most of the loci, not all loci were monomorphic in the population. This shows that ecotypes and accessions among the population are very similar but still different. Data obtained from SSR allowed to find SSR loci able to discriminate ecotypes within ‘San Marzano’ variety and to observe different ‘San Marzano’ clusters.
Results show that most of the analyzed samples (30) belong to three main groups, one ‘Kiros’ and two SMEC groups, and the remaining part is closely related to the SMEC groups indicating that the pool of analyzed samples represent a uniform population.
In particular, SSR analysis revealed that there are two SMEC groups, of 11 and 16 genotypes, composed only by genotypes from ‘San Marzano 4’ variety and other two

small SMEC groups, each of them constituted by 2 genotypes, composed exclusively by ‘San Marzano 2’.
In this study, it was also evaluated the homogeneity of the ‘Kiros’ genotype over the years observing the absence of molecular variability in time.
Furthermore, as described above, more than 100 tomato genotypes were characterized by the use of the same set of SSR. Through this, it was possible to verify that the population of ‘San Marzano’ is different from industrial varieties and also from other ecotypes and landraces from Campania Region.
The analysis demonstrates that simple sequences repeats are useful to confirm the homogeneity of a population that presents individuals morphologically similar each other and, furthermore, due to the polymorphism that characterizes these markers, it was also possible to highlight minimal genetic differences between ecotypes useful to separate ‘San Marzano 2’ and ‘Kiros’, included in the procedure guideline for ‘San Marzano’ PDO, from others ‘San Marzano’ types.

4.4 SSR as Tool to Trace ‘San Marzano’ Tomato in Food Chain

Traceability is the ability to trace and follow a food, feed, animal or substance intended to become part of a food or feed, through all stages of production, processing and distribution (EU regulation No. 178/2002). This definition is necessarily broad because food and feed are complex matrixes and traceability is a tool that should allow to track them, from row materials (e.g., plant cultivars) to final processed products and beyond, up to market distribution and consumers (Scarano et al., 2012).
The main difficulty to be faced in the molecular traceability of food supply chains is the analysis of DNA from complex food matrices. During the phases of processing of raw materials, in fact, the primary structure of the DNA molecule is modified due to chemical reactions of hydrolysis and oxidation that can lead to damage of the filament or to alterations in the sequence of nucleotides. The temperature and pH are the main factors that influence the stability of DNA.
The use of SSR markers for traceability in tomato food chain is a valuable tool to authenticate products with quality labels from improper replacement (Rao et al., 2009; Caramante et al., 2010; Alba et al., 2009, Sardaro et al., 2013).
'San Marzano' tomato is a classic example of a local variety with high value; from 1996 the Protected Designation of Origin (PDO) label protects the legitimate production of 'San Marzano' products but, unfortunately, it is frequently substituted with different cultivars with similar fruit shape and size (Scarano et al., 2011).
From the set of SSR markers used for varietal discrimination, six have been selected and tested (LE21085, LEaat002, LEct001, LEctt001, LEta015 and LEta003) (He et al., 2003; Smulders et al., 1997) in order to track the chain of peeled tomatoes labelled as 'San Marzano'.
In this work, thirty-three canned tomato, of which 27 belong to the same company and the rest are from six different processing companies, were tested. The analysis was performed on DNA extracted from single peeled berries from each can; the allelic profiles obtained from the peeled tomatoes labelled as ‘San Marzano’ did not match the profiles of the ‘San Marzano’ reference varieties (‘San Marzano 2’ and ‘Kiros’) and of other genetically similar accessions cultivated, indicating that it is possible to exclude the presence of fruits of these cultivars in the analyzed products.

The principal parameter used to investigate about the genetic origin of the fruits from tomato cans was the observed heterozigosity, revealing that the analyzed samples showed a percentage of heterozygosity between 6.5 and 50% at the considered SSR loci. This data are completely in contrast with those obtained by the analysis of the 'San Marzano' types, including 'San Marzano2' and 'Kiros', that showed a percentage of homozigosity of 100% for all analyzed loci.
Results indicate that the canned tomatoes probably do not contain what is declared on their label.
The analysis were conduced on individual berries from cans and this leaves no doubt to define that in the 33 analyzed canned tomato the 'San Marzano' variety is not present.
Furthermore, alleles from peeled tomatoes were compared with the allelic profiles obtained for the varieties 'San Marzano' used for PDO products. This comparison showed that profiles obtained for commercial canned tomatoes do not correspond with profiles observed for the 'San Marzano' ecotypes authorized by the PDO procedure guideline. On the other hand the lab that hosted my PhD programme previously demonstrated that SSR alleles are unaltered during tomato processing (Caramante et al., 2010), therefore supporting that what here observed is not the resulta al whatever alteration occurred during processing.
Comparison between SSR data obtained by the analysis of 33 peeled tomatoes and SSR allelic profile of 'San Marzano' types demonstrates that PDO 'San Marzano' varieties or other 'San Marzano' types were not used for production of analyzed cans, that therefore represented a fraud case.

4.5 Study of Gene Expression Patterns Associated to Specific Production Areas

Plants can alter their growth, physiology and characteristics in relation to environmental conditions. The prevailing model in crop science partitions the phenotypic variance into genetic (G), environmental (E) and interaction (G × E) components. In an agronomic context, the interaction is often extended to include management practices as a component of the environment, or explicitly as a third factor in the interaction (Messina et al., 2009).
A genotype can produce different phenotypes in different environments. This property is known as phenotypic plasticity (Sultan, 2000; Nicotra et al., 2010; Agrawal, 2001).
The current vision implies that the phenotypic plasticity has a genetic substrate and its heritability, and the potential importance for the evolution of species is well known (Bradshaw, 2006; Lande, 2009).
The knowledge about phenotypic plasticity comes from studies conducted on plants that document the wide range of phenotypes produced by individual genotypes in response to conflicting conditions. The most studied parameters are simple descriptors such as size of the plant, the number of branches and the length of the internodes. Van Kleunen and Fischer (2001) showed that two genotypes of *Ranunculus reptans* respond differently under the same growth conditions and concluded, therefore, that there is a phenotypic adaptation caused by a genetic variation of the genotype. Therefore, the phenotypic variability is a direct consequence not only of the differences linked to the genotype but also of the variability attributable to the gene expression.

The objective of this part of the project was to analyze transcriptomic variation gene expression variation associated with growth environment.
West-Eberhard (2003) defined the genome as the full complement of DNA in a cell, the phenotype as including all traits of an organism other than its genome and the genotype as the genetic makeup by which an individual or one of its traits can be characterised in genetic comparisons with other individuals or their phenotypic traits. The term phenome seeks symmetry with genome, and further emphasises that the phenotype is the individual outside the genome (Mahner and Kary, 1997). The profile of mRNA, an operational definition of gene expression, is also part of the phenotype (Nachtomy et al., 2007; Sadras et al., 2013).
Genotypes selected for this study represent two varieties, Kiros and Docet, used for identical industrial destination: peeled canned tomato. The former, included in the guideline for the production of 'San Marzano' PDO, is characterized by indeterminate growth and consequent berries scalar harvest occurring according to their ripening stage; it also undesirable agronomic traits such as the lack of genetic resistance against pathogens (Cucumber mosaic virus - CMV, tomato spotted wilt virus - TSWV, susceptibility to Pyrenochaeta lycopersici and Meloidogyne hapla) and a yield lower than modern hybrid cultivars (Monti et al. 2004). The other genotype, 'Docet', is an hybrid cultivar, with determined vegetative habitus and contemporary fruit ripening; berries, around 80/85 gr, have elongated shape, joint-less character, excellent consistency and resistance to over-ripening; furthermore, this variety is resistant to TSWV (Tomato spotted wilt virus), with intermediate resistance to *Pseudomonas syringae*.
Although molecular fingerprinting is able to trace tomato products along the food-chain, it is unable to reveal information on the geographical localization of the plants from which berries were harvested. Our study investigated the possibility to gain information on geographical origin of berry's production by comparing transcriptomic chances driven by growth location of tomato berries.
To study changes that occur at the transcriptomic level, two experimental fields near Naples were selected, Acerra and Brusciano (Acerra is a locality included in the procedure guideline for 'San Marzano' growth), where the two tomato varieties were cultivated, '.
Unlike genome, uniform across organisms from the same genotype, transcriptome is organism-specific and inevitably environmentally sensitive. As mentioned by Cobb et al. (2013) "the phenome of an organism is dynamic and conditional, representing a complex set of responses to a multi-dimensional set of endogenous and exogenous signals that are integrated over the evolutionary and developmental life history of an individual. Phenotypic information can be envisioned as a continuous stream of data that changes over the course of development of species or individuals, in response to different environmental conditions".
Sequencing data were first used to have an overview of the relationship among samples under analysis. It has been possible to verify that data obtained from the sequencing can separate samples for their harvest environment. Only by the use of mapped reads it was immediately possible to verify that berries from Acerra perfectly separate from those harvest in Brusciano, for both cultivars. These first data provided from the beginning a clear indication that the two cultivation environments differentially influence the gene expression profiles.
The study demonstrated that 39% of tomato genes are switched on. An high number of common genes were found after the comparison of the expressed genes: both comparison, the Kiros comparison and the Docet comparison, revealed 91.4% of

common genes. The most interesting thing is that the remaining 8.6% of genes, a very small amount if compared to the 91.4% (few hundreds against many thousands), was suitable to detect differences sufficient to provide a strong separation of the samples collected in the two environments, a clearly consequence of the different genotypic performance related to environments.
These data suggest that there is a connection between transcriptomic modulations and harvest environments that could be used for identification of the geographical origin.
The objective of this study, however, is not only to verify that the same genotype cultivated in two different environmental conditions produces different gene expression patterns, but also to identify new (transcriptomic) markers in order to make the genotype-environment association easily detectable. These reasons led to further analyze lists of expressed genes.
Thus, annotated expressed genes per sample were than converted in enriched lists of Gene Ontology (GO) terms. The Gene Ontology project provides an ontology of defined terms representing gene product properties. The ontology covers three domains: cellular component, the parts of a cell or its extracellular environment; molecular function, the elemental activities of a gene product at the molecular level, such as binding or catalysis; and biological process, operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms (http://www.geneontology.org).
The 87.5% and the 86.3% of GO terms, respectively for both environmental comparisons, were common between compared samples reflecting the high percentage of common genes highlighted among samples from the same genotype.
Both varieties showed biological process (P) as the most represented functional domain.
After the GO comparison, genes sample-specific, from GO terms not in common, were selected and analyzed trough their counts to detect differential expression within samples of the two comparisons. The analysis gave always up-regulated genes as outputs because the input tables were created in order to compare genes that are expressed in one condition (environment) and not in the other for the same genotype.
The number of expressed genes that showed a differential expression statistically significant was in a range of 9 to 97 genes.
For each sample, differentially expressed genes are grouped according the main three functional domains, cellular component, molecular function and biological process. In particular, for Kiros Brusciano the most represented domain among differentially expressed genes is molecular function, while for the other three samples is the biological process domain.
The analysis of differentially expressed genes highlights genes showing a greater difference in the expression level through statistical parameters. In fact, for each gene up-regulated in each condition it was assigned a statistical probability value (from 0 to 1) that indicates which is the possibility that the given gene is actually differentially expressed. In this way, it is possible to further reduce the number of genes on which investigate in order to select the transcriptomic markers genotype-environment specific using, for example, the top genes from the DEG lists. Through this selection, first genes to study as transcriptomic markers are those with a probability equal to 1, i.e. the maximum probability value indicating that the gene is actually differentially expressed in the two environments.

PKAC, that showed the highest number of differentially expressed genes (97), presented 16 genes certainly up-regulated in function of PKBR, while PKBR showed one specific up-regulated gene. In the Docet comparison there are three genes up-regulated with the maximum probability to be differentially expressed for DAC and one for DBR.

First gene (Solyc01g095510.2.1) is the adenine nucleotide translocator 1 belonging to the mitochondrial carrier family. In Arabidopsis this family consist of two isoforms of the ADP/ATP carrier (At3g08580 and At5g13490), the dicarboxylate/tricarboxylate carrier (At5g19760), the phosphate carrier (At5g14040), the uncoupling protein (At3g54110), and a carrier protein (At4g01100) named ADNT1 with a primary function to catalyze the exchange between cytosolic AMP and intra-mitochondrial ATP and, in theory, with an additional function of ADP/ATP carrier (Palmieri et al., 2008). An ADNT1-mediated AMP/ATP exchange is likely to occur across the inner mitochondrial membrane when AMP is the predominant adenine nucleotide present in the cytosol. It is known that cytosolic AMP increases markedly in plant tissues during emergence from dormancy and during stresses such as anoxia and is primarily converted to ATP during recovery from these stresses (Saglio et al., 1980; Standard et al., 1983; Raymond et al., 1985). Moreover, the ADNT1, expressed in particular in non-photosynthetic tissues, have a role not only in the subcellular regulation of adenylate metabolism but also in the oxidative phosphorylation, as demonstrated through the knockout of the ADNT1 gene in tomato where a reduced rate of root growth was observed (Carrari et al., 2003; Nunes-Nesi et al., 2007).

Gene Solyc11g073160.1.1 corresponds to the GTP cyclohydrolase/3 4-dihydroxy-2-butanone 4-phosphate synthase (RibA), involved in the riboflavin biosynthesis. Riboflavin (vitamin B2) is the universal precursor of the flavocoenzymes riboflavin phosphate (FMN) and flavin adenine dinucleotide (FAD), which act as intermediates in the transfer of electrons in biological oxidation-reduction reaction. More than 100 enzymes of animal and microbial systems are known to bind FAD or FMN. These enzymes, called flavoproteins or flavoenzymes, include many oxidases and dehydrogenase. Plant flavoproteins also include enzymes functioning in the biosynthesis of aromatic amino acid, quinones, lignin, flavonoids and alkaloids (Bartz and Brecht, 2002). The availability of the A. thaliana genome allowed the identification of the ribAB gene of Arabidopsis based on its similarity with homologous genes from *B. subtilis* and *E. coli* (Herz et al., 2000). The ribA and ribB genes of *E. coli* specify GTP cyclo-hydrolase II (GCHYII) and 3,4-dihydroxy-2-butanone-4-phosphate synthase (DHBPS); the ribA gene of *B. subtilis* specifies a bifunctional enzyme with both activities. Plant GTP cyclohydrolase/3,4-dihydroxy-2-butanone-4-phosphate synthase catalyze two reaction in the riboflavin biosynthetic pathway that are the GTP conversion in 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone-5'-phosphate and the condensation of pyrimidine and 3,4-dihydroxy-2-butanone-4-phosphate to obtain 6,7-dimethyl-8-ribityllumazine (Herz et al., 2000; Fisher et al., 2006).

The Solyc12g010800.1 gene encodes for a bZIP transcription factor from the family protein of basic leucine zipper. Transcription factors are key regulatory proteins of the last steps in signal transduction cascades activated in response to stress (for example salt stress, Yanez et al., 2009). Another study demonstrates that sucrose regulated transcription factor bZIP was shown to affect the amino acid metabolism through regulation of the expression of Asn synthetase1 and Pro dehydrogenase2 in Arabidopsis (Hanson et al., 2008).

Harpin-induced protein-like is coded by Solyc10g005040.2.1. It is involved in multiple signaling pathways including salicylic acid, jasmonic acid and ethylene (Miao et al., 2010)

The gene product of Solyc06g048930.2.1 is named two-component response regulator ARR3 and is part of the signal transduction in plant. Signalling pathways are structured around two conserved proteins: a histidine protein kinase (HK) and a response regulator protein (RR) that are phosphorylated at His and Asp residues, respectively. Phospho-transfer from the HK to the RR results in activation of the RR and generation of the output response of the signaling pathway. In Arabidopsis, it has been proposed that response regulators could be classified into two groups, type-A and type-B, based on their architecture (Imamura, et al., 1999). The type-A response regulators are mainly composed of a receiver domain and short N- and C-terminal extensions, whereas the type-B response regulators have a receiver domain and a largely extended C-terminal region that is supposed to be an output domain. By using in vitro phosphotransfer analysis, ARR3 has been shown to accept a phosphoryl group from a phospho-HPt domain of an *E. coli* hybrid histidine kinase (Urao et al., 2000; West and Stock, 2001).

DNA mismatch repair protein MutS (MSH2) is a component of the mismatch repair (MMR) system, a DNA repair pathway essential for the correct maintenance of genetic information across many generations and is coded by the Solyc06g069230.2.1 tomato gene. The system is best known for its role in the correction of base substitutions and small insertion/deletion loops (IDL) generated during every round of replication. In A. thaliana, MSH2 works as heterodimer with MSH6, MSH3 and MSH7 to produce heterodimers AtMutSα, AtMutSβ and AtMutSγ. AtMutSα recognize base–base mispairs and small IDL, AtMutSβ participate in the repair of IDL heterologies, AtMutSγ preferentially binds some base–base mispairs (Culligan et al., 2000; Wu et al., 2003). Tomato MSH2 is related to *P. hybrida* MSH2 and to the MSH2 orthologues of *V. vinifera* and *A. thaliana* (Tam et al., 2009).

The ethylene-responsive transcription factor 4 (Solyc04g007000.1.1) belongs to the molecular function domain of DNA binding. Ethylene-responsive factor (ERF) proteins activate PR genes by binding to the GCC box (GCCGCC) of their promoters, thereby regulating the plant defence response to biotic and to abiotic stress (Zhang et al., 2004, Park et al., 2001; Zhang et al., 2005). These proteins contain unique transcription factors to the plant lineage and are classified into four subfamilies: AP2 (for apetala2 from Arabidopsis), DREB (for dehydration response element-binding protein), ERF, and RAV (for related to ABI3/VP1, where ABI3 is the abscisic acid-insensitive3 transcription factor of Arabidopsis and VP1 is the viviparous1 transcription factor from maize) (Suzuki, et al., 1997; Li et al., 2011). The tomato ethylene-responsive transcription factor-4 is part of the RAV proteins family involved in defence pathway (Li et al., 2011).

The tomato gene Solyc08g074270.2.1 codes for the Cryptochrome 3 (Cry3). Cryptochromes are a ubiquitous class of blue light photoreceptors, found in higher and lower plants (Kanegae and Wada, 1998), insects (Emery et al., 1998) and mammals (van der Horst et al., 1999). In all organisms studied, cryptochromes are involved in the control of circadian timing (Somers et al., 1998; Ceriani et al., 1999) and in Arabidopsis they are also involved in the control of flowering time by photoperiod (Guo et al., 1998; Mockler et al., 1999). Additionally, plant cryptochromes are involved in phototropic response (Ahmad et al., 1998) and in photomorphogenetic events like the inhibition of seedling and internode elongation and the biosynthesis of anthocyanins (Lin et al., 1998; Ninu et al., 1999).

Cryptochromes have most likely evolved from photolyases, another class of blue-UV-light-absorbing flavoproteins with a distinct function: they are involved in the photoreactivation of UV-damaged DNA. A study on Cryptochrome 3 of Arabidopsis suggests that its role is carried out in the mitochondria due to the N-terminal localization domain. It works in association with a folate chromophore playing a role as a general antenna pigment for efficient energy transfer (Klar et al., 2007).
The gene Solyc01g079500.2.1, encoding for the MCM7 protein, was found among the up-regulated genes. MCM proteins are part of the pre-RC (replication complex). In Arabidopsis thaliana, the pre-RC consists of six origin recognition complex (ORC) proteins and six canonical MCM proteins (MCM2 to MCM7) (Masuda et al., 2004). In particular, loss-of-function approaches revealed that MCM7 is required for gametophyte development and is maternally required for embryo development in Arabidopsis (Springer et al., 2000).
Single-stranded DNA binding protein p30 subunit (replication protein subunit RPA32-Solyc07g018300.2.1) is involved in the DNA replication. Replication protein A (RPA) is a single-stranded DNA (ssDNA)-binding protein complex comprising a heterotrimeric combination of a large (70 kDa), middle (32 kDa) and small (14 kDa) subunit (Wold 1997). RPA is phosphorylated in a cell cycle-dependent manner (Din et al., 1990; Dutta and Stillman 1992) and in response to DNA damage (Lee and Kim 1995; Zernik-Kobak et al., 1997). Hyperphosphorylation is involved in response to DNA-damaging agents, such as UV or ionizing radiation (Carty et al., 1994; Liu and Weaver, 1993) and cellular apoptosis (Treuner et al., 1999) and the phosphorylation occurs primarily on the N-terminal of RPA32 (Lee and Kim, 1995; Henricksen et al., 1996).
Solyc06g050270.1.1 tomato gene produces a CBL-interacting protein kinase (CIPK). Recently, the group of plant calcium sensor proteins has been extended by the identification of calcineurin B-like (CBL) proteins from Arabidopsis. These proteins are most similar to both the regulatory B subunit of calcineurin (CNB) and neuronal calcium sensors (NCS) of animals (Kudla et al., 1999). CBL proteins contain EF hand motifs as structural basis for calcium binding and interact specifically with a group of Ser/Thr protein kinases designated as CBL-interacting protein kinases (CIPKs; Shi et al., 1999; Kim et al., 2000). CIPKs most likely represent targets of calcium signals sensed and transduced by CBL proteins.
Solyc02g071860.2.1 and Solyc07g066230.2.1 genes produce kinases of the family of receptor-like kinases (RLK). In particular, RLK-genes found as up-regulated produce RLK-LRR (leucine-rich repeats) similar to Arabidopsis BRI1 gene, which is involved in the control of plant growth and development (Shiu and Bleecker, 2001).
The 3-hydroxyacyl-CoA dehydrogenase (HCDH), corresponding to the Solyc08g068390.2.1 tomato gene, is an enzyme involved in fatty acid metabolism responsible for the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA. In peroxisomes 3-hydroxyacyl-CoA dehydrogenase forms, with enoyl-CoA hydratase (ECH) and 3,2-trans-enoyl-CoA isomerase (ECI) a multifunctional enzyme where the N-terminal domain bears the hydratase/isomerase activities and the C-terminal domain the dehydrogenase activity. There are also two mitochondrial enzymes: one that is monofunctional and the other, which is, like its peroxisomal counterpart, multifunctional (Birktoft et al., 1987).
The essential enzyme dUTP pyrophosphatase (Solyc01g097970.2.1 gene) is specific for dUTP and is critical for the fidelity of DNA replication and repair. dUTPase hydrolyzes dUTP to dUMP and pyrophosphate, simultaneously reducing dUTP levels and providing the dUMP for dTTP biosynthesis. dUTPase decreases the intracellular

concentration of dUPT so that uracil cannot be incorporated into DNA during the replication (Mol et al., 1996).
Glucose dehydrogenase (Solyc06g083480.2.1) catalyses the oxidation of D-glucose without prior phosphorylation to D-beta-gluconolactone using NAD or NADP as a coenzyme. The enzyme is a tetrameric protein, each of the 4 identical subunits containing 262 amino acid residues. This family is a subset of a more general family of short-chain dehydrogenases and reductases (SDR) whose involvement has been demonstrated in a variety of primary (lipid synthesis, chlorophyll biosynthesis or degradation) and secondary metabolisms (terpenoids, steroids, phenolics and alkaloids) (Tonfack et al., 2011).
Solyc10g084690.1.1 correspond to ADP-ribosylation factor GTPase activating protein 1 (Arf). This name describes a family of small GTPase activating proteins, which are important for the regulation of the ADP ribosylation factor ARF, a member of the Ras superfamily of GTP-binding proteins (Vitale et al., 1998). ARFs regulate membrane trafficking by cycling between the GTP-bound active and GDP-bound inactive forms. ARFs are activated by GTP-exchange factors (GEFs) and inactivated by GTPase-activating proteins (GAPs) (Kahn 2009).
Among up-regulated highlighted genes in both comparisons most are involved in UV-response.
DNA in plants is continuously damaged by UV irradiation from sunlight. UV is known to induce DNA damage, although plants generally have a higher tolerance for UV than animals. Field-grown crops are also known to suffer continuous UV-induced DNA damage. Furthermore, the formation of reactive oxygen species in cells due to UV irradiation, biotic stresses and secondary metabolism, causes cellular components, including DNA, to be oxidized and therefore susceptible to oxidative modification. In addition, the fidelity and integrity of DNA are constantly challenged by chemical substances in the environment, ionizing radiation and errors that occur during DNA replication or proofreading. This accumulated damage blocks a number of critical processes, such as transcription and replication, and can eventually cause cell death. Thus, UV damage can reduce the growth and yield of plant crops, but plants have evolved several DNA-repair pathways (Kimura et al., 2006) and many genes can be involved in the UV-response (Maffei, 1998).
The analyzed plants of this study were collected in open-field when berries showed the perfect ripening status, in august. So, the presence of genes whose products are involved in DNA damage repair, such as MutS, MCM7, dUTP pyrophosphatase and RPA, certainly has a biological meaning of plant self-protection from UV-radiation. Moreover, the presence of genes involved in the light signal regulation, the criptochrome 3, and in response to oxidative stress, ADNT1, is an additional information about plant response to light stress.
In particular, transcripts of these genes are strongly marked in the Kiros Acerra sample indicating a good response of the ‘Kiros’ genotype to the Acerra environment.
In recent years, UV-C irradiation has also been tested as a postharvest treatment to increase ascorbic acid and total phenolic contents (Jagadeesh et al., 2009), and improve nutritional qualities by increasing lycopene content (Liu et al., 2009) in tomato fruit. In light of these, it is conceivable that genotype adaptation to the environment, and than to UV-irradiation, can also affect the organoleptic characteristics of the edible product. In effect, genes like bZIP, RibA, Cry3 and HCDH are also involved in several biosynthetic pathway such as amino-acid, carotenoids, fatty acid, lignin, flavonoids, alkaloids, vitamins and anthocyanin.

Furthermore, other genes and their products from Kiros Acerra, such as ARR3, ethylene-responsive transcription factor 4, riboflavin, CIPK, harpin-induced protein-like, RLKs and bZIP, are implicate in the plant signalling transduction in response to biotic and abiotic stresses, indicating the presence of networks of signal transduction to cope with the environment, to control metabolism, and to realize developmental programs.
In contrast, the same genotype showed different up-regulated genes in Brusciano environment. In particular, one gene, encoding for the glucose dehydrogenase, involved in primary and secondary metabolism, resulted to have the highest probability to be really differentially expressed in function of Kiros Acerra.
The analysis of the major differentially expressed genes in Docet comparison revealed that the hybrid cultivar showed in Brusciano environment one gene certainly differentially up-regulated, that is the ADP-ribosylation factor GTPase activating protein 1 (Arf), while in Acerra there are three up-regulated genes, which are included in the gene list of Kiros Acerra. These are ARR3, bZIP and harpin-induced protein-like.
These data suggest that the Acerra environment, the combination of its soil composition and climate, drive plants to switch on genes related to the signalling transduction probably because of some biotic and/or abiotic stress that are predominant in the environment.
All genes described above can be involved in the control of developmental and metabolic processes and may also affect fruit quality.
However, first objective of the differentially expressed genes analysis was to select transcriptomic markers able to distinguish the geographical origin of a given tomato genotype. The main purpose lies in the implementation of instruments for ‘San Marzano’ tomato traceability in order to have new tools to discover false claims regarding geographical origin of production.
The obtained results suggest that a significant distinction between the two environments is reflected in the transcriptome of ‘Kiros’ genotype. Therefore, data analysis has made possible the selection of transcriptomic markers capable to discriminate ‘Kiros’ genotype collected in Acerra or Brusciano.
Surely, the complete list of up-regulated genes per sample can be considerate as transcriptomic markers but those genes that showed the greatest difference in expression between two environments represent, undoubtedly, the most interesting transcriptomic markers.
On the other hand, the analysis of ‘Docet’ genotype in the same breeding conditions of ‘Kiros’ supported the comprehension of which genes are strongly influenced by the environment and which are, instead, due to the close relationship genotype-environment allowing, at the end, to select those genes that, despite are obviously influenced by the surrounding conditions, remain closely linked to the expression of the given genome in a given environment.
Concluding, with recent developments in high-throughput gene expression screening, it is possible to obtain gene expression profiles in tomato that correlate with a specific grown environment. The identification of a set of genes related to the environment enables to perform rapid and easy-to-use tests to certificate and protect, by the authentication of the geographical origin of production, agro-products, such as ‘San Marzano’ tomato for which the association with a given environment is crucial to getting the organoleptic and physiological characteristics that make it different from others tomato varieties.

5. CONCLUSIONS

Plant genetics and biotechnology gave an essential contribution to the development of agriculture and society through the production of new varieties and novel high quality products of plant origin. Recently, genetic techniques have also made contributions to the food industry, with new tools of investigation and analysis. Among these, the possibility to know the identity of the genetic components of a food have assumed increasing notoriety even among non-specialists, thanks to the action of mass communication in the dissemination of information on the techniques of DNA analysis. The application of these techniques to genetic traceability in the agro-food sector, represents a powerful tool to protect both producers and consumers, to ensure freedom of choice and ensure the accuracy of labeling. In addition, genetic traceability plays a key role to protect the production characteristics of Italian agro-food industry, and so to ensure our nation's privileged position as a supplier of premium products in world.

This study demonstrates that DNA markers such as SSR provide very effective for the discrimination of tomato genotypes. Local varieties from Campania region, including ‘San Marzano’ types, and hybrid varieties, were successfully discriminated confirming that markers appropriately selected can provide valuable support to germplasm identification even when analysing closely related types.
The same kind of markers was used hereinafter also in the authentication of commercial tomato products in order to perform traceability in ‘San Marzano’ tomato chain. Simple sequence repeats have proven to be a valuable tool in the evaluation of the content of tomato canned products labeled as “San Marzano-PDO”. Through the high polymorphism of these markers has been possible to characterize a large population of 'San Marzano' ecotypes in order to obtain information about the allelic composition. These information were essential for traceability in the ‘San Marzano’ tomato supply chain because the knowledge of which and how many alleles are present in the different ‘San Marzano’ types, allowed us to perform a very accurate analysis of the final product.
Canned tomatoes labelled as ‘San Marzano’ are the most defrauded tomato food products and this study has shown that through the use of SSR is possible to detect frauds, at the expense of the genotype, employing a tool easy to use, fast and inexpensive.
Beside the genetic characterization of tomato genotypes both for the variability study and for the San Marzano food chain authentication, this study describe a transcriptomic markers as a possible tool for the identification of the geographic origin of the berries production. The study of the genes expressed in different locations allowed to associate gene expression patterns to production areas. Genes showing an expression related to the genotype-environment interaction can be transformed into transcriptomic markers to be used to contribute to the identification of the geographical origin of the production.
The sum of the effects that the local environment has on the production of the product can be translated as *terroir*, the set of special characteristics that the geography, geology and climate of a certain place, interacting with plant genetics, express in agricultural products. The concept may be extended to any agricultural product and food but, originally, was coded in relation to wine and unique features of the French vineyard.

The use of transcriptomic markers, therefore, may help in highlighting fraud at the expense of the geographical area of production, that in the case of the ‘San Marzano’ tomato is a key factor to obtain the product that reflects the characteristics that made it famous and protected by the EU label PDO.
In conclusion, the present study showed that the use of DNA and transcriptomic markers and their combination can authenticate primary products in food chain and contrast food fraud meeting and supporting the ideal characteristics of a traceability system based on breadth, depth and precision.
Therefore, in light of this, traceability can really be the bridge between consumers and producers, restoring consumer confidence and contribution to the transparency of the production system.

6. REFERENCES

Agrawal A. A. (2001) Phenotypic Plasticity in the Interactions and Evolution of Species. Science's Compass, Vol. 294, pp. 321-326.

Agriculture Dept (US) (Ed.) (2010). Keys to Soil Taxonomy, 2010. Government Printing Office. ISO 690

Ahmad M., Jarillo J.A., Smirnova O., Cashmore A.R. (1998). Cryptochrome blue-light photoreceptors of Arabidopsis implicated in phototropism. Nature, Vol. 392, pp. 720-723.

Alba V., Sabetta W., Blanco A., Pasqualone A., Montemurro C. (2009) Microsatellite marker to identify specific alleles in DNA extracted from monovarietal virgin olive oils. European Food Research and Technology, Vol. 229, pp. 375-382.

Arana A., Soret B., Lasa I., Alfonso L. (2002) Meat traceability using DNA markers: application to the beef industry. Meat Science, Vol. 61, pp. 367-373.

Armstrong DJ. (2009). Food chemistry and US food regulations. Journal of Agricultural Food Chemistry, Vol. 57, pp. 8180-8186.

Ewing B., Green P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities, Genome Research, Vol. 8, pp. 186-194.

Bai Y., Feng X., van der Hulst R., Lindhout P. (2004) A set of simple PCR markers converted from sequence specific RFLP markers on tomato chromosomes 9 to 12. Molecular Breeding, Vol. 13, pp. 281-287.

Barbuto M., Galimberti A., Ferri E., Labra M., Malandra R., Galli P., Casiraghi M. (2010) DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of ‘‘palombo” (Mustelus spp.). Food Research International, Vol. 43, pp. 376–381.

Bartz J. A., Brecht J. K. (2002) - Postharvest Physiology and Pathology of Vegetables, CRC Press.

Birktoft J. J., Holden H. M., Hamlin R., Xuong N. H., Banaszak L. J. (1987) Structure of L-3-hydroxyacyl-coenzyme A dehydrogenase: preliminary chain tracing at 2.8-A resolution. Proc Natl Acad Sci U S A, Vol. 84, pp. 8262-8266.

Boguski M. S., Tolstoshev C. M, Bassett D. E. Jr. (1994) Gene discovery in dbEST. Science, Vol. 265, pp.1993-1994.

Borneman A. R., Forgan A. H., Pretorius I. S., Chambers P. J.(2008) Comparative genome analysis of a Saccharomyces cerevisiae wine strain. FEMS Yeast Research, Vol. 8, pp. 1185-1195.

Bradshaw A.D. (2006) Unraveling phenotypic plasticity-why should we bother? New Phytologist Vol. 170, pp. 644-648.

Bredemeijer G., Cooke R., Ganal M., Peeters R., Isaac P., Noordijk Y., Rendell S., Jackson J., Roder M. S., Wendehake K., Dijcks M., Amelaine M., Wickaert V., Bertrand L., Vosman B. (2002) Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theoretical and Applied Genetics, Vol. 105, pp. 1019-1026.

Burgess K. S., Fazekas A. J., Kesanakurti P. R., Graham S. W., Husband B. C., Newmaster S. G., Percy D. M., Hajibabaei M., Barrett S. C. H. (2011) Discriminating plant species in a local temperate flora using the rbcL+matK DNA barcode. Methods in Ecology and Evolution, Vol. 2, pp. 333–340.

Cai Y. S., Zhang L. A., Shen F. J., Zhang W. P., Hou R., Yue B. S., Li J., Zhang Z. H. (2011). DNA barcoding of 18 species of Bovidae. Chinese Science Bulletin, Vol. 56, pp.164–168.

Campbell P. J., Stephens P. J., Pleasance E. D., O'Meara S., Li H., Santarius T., Stebbings L. A., Leroy C., Edkins S., Hardy C., Teague J. W., Menzies A., Goodhead I., Turner D. J., Clee C. M., Quail M. A., Cox A., Brown C., Durbin R., Hurles M. E., Edwards P. A. W., Bignell G. R., Stratton M. R., Futreal P. A. (2008) Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nature Genetics, Vol. 40, pp. 722-729.

Caramante M., Corrado G., Monti L. M., Rao R. (2010) Simple Sequence Repeats are able to trace tomato cultivars in tomato food chains. Food Control, Vol. 22, pp. 549-554.

Carninci P., Kasukawa T., Katayama S., Gough J., Frith M., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., et al. (2005) The transcriptional landscape of the mammalian genome. Science, Vol. 309, pp. 1559-1563.

Carrari F., Coll-Garcia D., Schauer N., Lytovchenko A., Palacios-Rojas N., Balbo I., Rosso M., Fernie A. R. (2005) Deficiency of a plastidial adenylate kinase in Arabidopsis results in elevated photosynthetic amino acid biosynthesis and enhanced growth. Plant Physiology, Vol. 137, pp. 70-82.

Carty M. P., Zernik-Kobak M., McGrath S., Dixon K. (1994) UV light-induced DNA synthesis arrest in HeLa cells is associated with changes in phosphorylation of human single-stranded DNA-binding protein. EMBO Journal, Vol. 13, pp. 2114-2123.

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences 106: 12794–12797.

Ceriani M. F., Darlington T. K., Staknis D., Mas P., Petti A. A., Weitz C.J., Kay, S.A. (1999) Light-dependent sequestration of TIMELESS by CRYPTOCHROME. Science, Vol. 285, pp. 553-556.

Chase M. W., Cowan R. S., Hollingsworth P. M., van den Berg C., Madrinan S., Petersen P., Seberg O., Jørgsensen T., Cameron K. M., Carine M., Pedersen N., Hedderson T. A. J., Conrad F., Salazar G. A., Richardson J. E., Hollingsworth M. L., Barraclough T. G., Kelly L., Wilkinson M. (2007) A proposal for a standardised protocol to barcode all land plants. Taxon, Vol. 56, pp. 295–299.

Chen J., Kim Y. C., Jung Y. C., Xuan Z., Dworkin G., Zhang Y., et al. (2008) Scanning the human genome at kilobase resolution. Genome Research, Vol. 18, pp. 751-762.

Chen K. Y., Cong B., Wing R., Vrebalov J., Tanksley S. D. (2007) Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. Science, Vol. 318, pp 643-645.

Chung A. M., Staub J. E., Chen J. F. (2006) Molecular phylogeny of Cucumis species as revealed by consensus chloroplast SSR marker length and sequence variation. Genome Vol. 49, pp. 219-229.

Cloonan N., Forrest A. R. R., Kolle G., Gardiner B. B. A., Faulkner G. J., Brown M. K., Taylor D. F., Steptoe A. L., Wani S., Bethel G., Robertson A. J., Perkins A. C., Bruce S. J., Lee C. C., Ranade S. S., Peckham H. E., Manning J. M., McKernan K. J., Grimmond S. M. (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nature Methods, Vol. 5, pp. 613-619).

Cobb J. N., DeClerck G., Greenberg A., Clark R., McCouch S. (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics, Vol. 126, pp. 867-887.

Conesa A., Götz S., Garcia-Gomez J. M., Terol J., Talon M., Robles M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics, Vol. 21, pp. 3674-3676.

Corrado, G., Imperato, A., La Mura M., Perri E., Rao R. (2011) Genetic diversity among olive varieties of southern italy and the traceability of olive oil using SSR markers. Journal of Horticultural Science and Biotechnology, Vol. 86, pp. 461-466.

Corrado G., Piffanelli P., Caramante M., Coppola M., Rao R. (2013) SNP genotyping reveals genetic diversity between cultivated landraces and contemporary varieties of tomato. BMC genomics, Vol. 14, pp. 835-849.

Culligan K., Hays J. (2000). Arabidopsis MutS homologs-AtMSH2, AtMSH3, AtMSH6, and a novel AtMSH7-form three distinct protein heterodimers with different specificities for mismatched DNA. Plant Cell, Vol. 12, pp. 991-1002.

De Mattia F., Bruni I., Galimberti A., Cattaneo F., Casiraghi M., Labra M. (2011). A comparative study of different DNA barcoding markers for the identifcation of some members of Lamiacaea. Food Research International, Vol. 44, pp. 693–702.

De Jonge, J., Van Trijp, J.C.M., Renes, R.J. & Frewer, L.J. (2007) Understanding consumer confidence in the safety of food: its twodimensional structure and determinants. Risk Analysis, Vol. 27, pp. 729–740.

Delmotte F., Leterme N., Simon J. C. (2001). Microsatellite allele sizing: difference between automated capillary electrophoresis and manual technique. BioTechniques, Vol. 31, pp. 810-818.

Delseny M., Laroche M., Penon P. (1983) Detection of sequences with Z-DNA forming potential in higher plants. Biochemical and Biophysical Research Communications, Vol. 116, pp.113-120.

Din S., Brill S. J., Fairman M. P., Stillman B. (1990) Cell-cycle-regulated phosphorylation of DNA replication factor A from human and yeast cells. Genes and Development, Vol. 4,, pp. 968-977.

Ding L., Getz G., Wheeler D. A., Mardis E. R., McLellan M. D., Cibulskis K., Sougnez C., Greulich H., Muzny D. M., Morgan M. B., Fulton L., Fulton R. S., Zhang Q., Wendl M. C., Lawrence M. S., Larson D. E., Chen K., Dooling D. J., Sabo A., Hawes A. C., Shen H., Jhangiani S. N., Lewis L. R., Hall O., Zhu Y., Mathew T., Ren Y., Yao J., Scherer S. E., Clerc K., Metcalf G. A., Ng B., Milosavljevic A., Gonzalez-Garay M. L., Osborne J. R., Meyer R., Shi X., Tang Y., Koboldt D. C., Lin L., Abbott R., Miner T. L., Pohl C., Fewell G., Haipek C., Schmidt H., Dunford-Shore B. H., Kraja A., Crosby S. D., Sawyer C. S., Vickery T. V., Sander S., Robinson J., Winckler W., Baldwin J., Chirieac L. R., Dutt A., Fennell T., Hanna M., Johnson B. E., Onofrio R. C., Thomas R. K., Tonon G., Weir B. A., Zhao X., Ziaugra L., Zody M. C., Giordano T., Orringer M. B., Roth J. A., Spitz M. R., Wistuba I. I., Ozenberger B., Good P. J., Chang A. C., Beer D. G., Watson M. A., Ladanyi M., Broderick S., Yoshizawa A., Travis W. D., Pao W., Province M. A., Weinstock G. M., Varmus H. E., Gabriel S. B., Lander E. S., Gibbs R. A., Meyerson M., Wilson R. K. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. Nature, Vol. 455, pp. 1069 -1075.

Dutta A., Stillman B. (1992) cdc2 family kinases phosphorylate a human cell DNA replication factor, RPA, and activate DNA replication. The EMBO Journal, Vol. 11, pp. 2189-2199.

Edwards A., Civitello A., Hammond H. A., Caskey C. T. (1991) DNA typing and genetic mapping with trimeric and tet- rameric tandem repeats. American Journal of Human Genetics, Vol. 49, pp. 746-756.

Emery P., So W.V., Kaneko M., Hall J.C., Rosbash M. (1998) CRY, a Drosophila clock- and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. Cell, Vol. 95, pp. 669-679.

Espineira M., Nerea González-Lavín N., Vieites J. M., Santaclara, F. (2008) Development of a method for the genetic identification of flatfish species on the basis of mitochondrial DNA sequences. Journal of Agriculture and Food Chemistry, Vol. 56, pp. 8954–8961.

Esquinas-Alcazar J., Nuez F. (1995) Situacion taxonomica, domesticacion y divusion del tomate. In: Nuez F. (eds) El cultivo del tomate. Ed Mundi-Prensa, Madrid, Barcelona, Mexico, pp 13-42.

Ewing B., Hillier L., Wendl M. C., Green P. (1998a) Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research, Vol. 8, pp. 175-185

Ewing B., Green P. (1998b) Base-calling of automated sequencer traces using phred. II. Accuracy assessment. Genome Research, Vol. 8, pp. 185-194.

Fazekas A. J., Kesanakurti P.R., Burgess K. S., Percy D.M., Graham S.W., Barrett S.C.H., Newmaster S. G., Hajibabaei M., Husband B.C. (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Molecular Ecology Resources, Vol. 9, pp. 130–139.

Filonzi L., Chiesa S., Vaghi M., Nonnis Marzano F. (2010). Molecular barcoding reveals mislabelling of commercial fish products in Italy. Food Research International, Vol. 43, pp. 1383–1388.

Fischer M., Bacher A. (2006) Biosynthesis of vitamin B2 in plants. Physiologia Plantarum Vol. 126, pp. 304-318.

Food Standards Agency Draft Consultation Paper. Food labelling: Country of origin. http://www.food.gov.uk/multimedia/pdfs/ori- gin-labelling.pdf.

Foolad M. R., Jones R. A., Rodriguez R. L. (1993) RAPD markers for constructing intraspecific tomato genetic maps. Plant Cell Reports, Vol. 12, pp. 293-297.

Fray R., Grierson D. (1993) Identification and genetic-analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. Plant Molecular Biology, Vol. 22, pp. 589-602.

Garcia-Martinez S., Andreani L., Garcia-Gusano M., Geuna F., Ruiz J. J. (2006) Evaluation of amplified fragment length polymorphism and simple sequence repeats for tomato germplasm fingerprinting: utility for grouping closely related traditional cultivars. Genome, Vol. 49, pp. 648-656.

García-Alcalde F., Okonechnikov K., Carbonell J., Cruz L. M., Götz S., Tarazona S., Dopazo J., Meyer T. F., Conesa A. (2012) Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics Applications Note, Vol. 28, pp. 2678-2679.

Gerhard D. S. et al. (2004) The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). Genome Research, Vol. 14, pp. 2121-2127.

Golan E., Krissof B., Kuchler F., Nelson K., Price G. (2004). Traceability in the US foos supply: economic theory and industry studies. Agricultural Economic report no. AER830. Eshington, DC: Economic research service of the United States department of Agriculture.

Gould W.A. (1992) Tomato production. Processing and technology CTI publications INC.

Graveley B., Brooks A., Carlson J., Duff M., Landolin J., Yang L., Artieri C., van Baren M., Boley N., Booth B., et al. (2010) The developmental transcriptome of Drosophila melanogaster. Nature, Vol. 471, pp. 473-479.

Grassi F., Labra M., Minuto L. (2006) Molecular diversity in Ligurian local races of common beans (Phaseolus vulgaris L.). Plant Biosystems, Vol. 140, pp. 17–20.

Guo H.W., Yang W.Y., Mockler T.C., Lin C.T. (1998). Regulation of flowering time by Arabidopsis photoreceptors. Science, Vol. 279, pp. 1360-1363.

Hamada H. M., Petrino M. G., Kakunaga T. (1982) A novel repeated element with Z-DNAforming potential is widely found in evolutionarily diverse eukaryotic genomes. Proceedings of the National Academy of Sciences USA, Vol. 79, pp. 6465-6469.

Hancock J. M. (1999) Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: Goldstein D., Schlotterer C., editors. Microsatellites - Evolution and Applications. New York: Oxford University Press, pp. 1-9.

Hanson J., Hanssen M., Wiese A., Hendriks M., Smeekens S. (2008) The sucrose regulated transcription factor bZIP11 affects amino acid metabolism by regulating the expression of asparagine synthetase 1 and proline dehydrogenase 2. Plant Journal, Vol. 53, pp. 935-949.

Hassall H. A. (1861) Adulterations Detected or Plain Instructions for the Discovery of Frauds in Food and Medicine. Longman, Green, Longman and Roberts.

He C., Poysa V., Yu K. (2003) Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among Lycopersicon esculentum cultivars. Theoretical and Applied Genetics, Vol. 106, pp. 363-373.

Henricksen L. A., Carter T., Dutta A., Wold M. S. (1996) Phosphorylation of human replication protein A by the DNA-dependent protein kinase is involved in the modulation of DNA replication. Nucleic Acids Research, Vol. 24, pp. 3107-3112.

Hebert P. D. N., Cywinska A., Ball S. L., de Waard J. R. (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society, Vol. 270, pp. 313–322.

Hebert P. D. N., Stoeckle M. Y., Zemlak T. S., Francis C. M. (2004) Identification of birds through DNA barcodes. PLoS Biology, Vol. 2, pp. 1657–1663.

Herz S., Eberhardt S., Bacher A. (2000) Biosynthesis of riboflavin in plants. The ribA gene of Arabidopsis thaliana species a bifunctional GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase. Phytochemistry, Vol. 53, pp. 723-731.

Hollingsworth P. M., Graham S. W., Little D. P. (2011) Choosing and Using a Plant DNA Barcode. PLoS ONE Vol. 6, pp. 19254-19267.

Houghton J.R., Rowe G., Frewer L.J., Van Kleef E., Chryssochoidis G., Kehagia, O., Korzen-Bohr S., Lassen J., Pfenning U., Strada A. (2008) The quality of food risk management in Europe: perspectives and priorities. Food Policy, Vol. 33, pp. 13–26.

Huber J. A., Mark Welch D. B., Morrison H. G., Huse S. M., Neal P. R., Butterfield D. A., Sogin M. L. (2007) Microbial population structures in the deep marine biosphere. Science, Vol. 318, pp. 97-100.

Imamura A., Hanaki N., Nakamura A., Suzuki T., Taniguchi M., Kiba T., Ueguchi C., Sugiyama T., Mizuno T. (1999) Compilation and characterization of Arabidopsis response regulators implicated in His-Asp phosphorelay signal transduction. Plant Cell Physiology, Vol. 40, pp. 733-742.

International Human Genome Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature, Vol. 431, pp. 931-945.

Jacob H. J., Lindpaintnesr K., Kusumir E. L., Bunkery K., Mao I. P., Gantenv D., Dzau J., Lander E. S. (1991) Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. Cell, Vol. 67, pp. 213-224.

Jagadeesh S.L., Charles M.T., Gariepy Y., Goyette B., Raghavan G.S.V., Vigneault C., (2011) Influence of postharvest UV-C hormesis on the bioactive components of tomato during post-treatment handling. Food Bioprocess Technology, Vol. 4, pp. 1463-1472.

Jeffreys A. J., Wilson V., Thein S. L. (1985) Hypervariable ‘minisatellite’ regions in human DNA. Nature, Vol. 314, pp. 67-73.

Jérôme M., Martinsohn J. T., Ortega D., Carreau P., Verrez-Bagnis V., Mouchel O. (2008) Toward fish and seafood traceability: Anchovy species determination in fish products by molecular markers and support through a public domain database. Journal of Agriculture and Food Chemistry, Vol. 56, pp. 3460–3469.

Kahlau S., Aspinall S., Gray J. C., Bock R. (2006) Sequence of the Tomato Chloroplast DNA and Evolutionary Comparison of Solanaceous Plastid Genomes. Journal of Molecular Evolution, Vol. 63, pp.194-207

Kahn R. A. (2009) Toward a model for Arf GTPases as regulators of traffic at the Golgi. FEBS letters, Vol. 583, pp. 3872-3879.

Kalloo, G., (1991) Genetic improvement of tomato. Monogr on Theoretical and Applied Genetics 14.

Kanegae T., Wada M. (1998) Isolation and characterization of homologues of plant blue-light photoreceptor (cryptochrome) genes from the fern Adiantum capillus-veneris. Molecular and General Genetics, Vol. 259, pp. 345-253.

Kelly S., Heaton K., Hoogewerff J. (2005) Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. Trends in Food Science & Technology, Vol. 16, pp. 555-567.

Kim K. N., Cheong Y. H., Gupta R., Luan S. (2000) Interaction specificity of Arabidopsis calcineurin B-like calcium sensors and their target kinases. Plant Physiology, Vol. 124, pp. 1844-1853.

Kim P. M., Lam H. Y., Urban A. E., Korbel J. O., Affourtit J., Grubert F., et al. (2008) Analysis of copy number variants and segmental duplications in the human genome: evidence for a change in the pro- cess of formation in recent evolutionary history. Genome Research, Vol.18, pp. 1865-74.

Kimura S., Sakaguchi K. (2006) DNA repair in plants. Chemical Reviews, Vol. 106, pp. 753-766.

Klar T., Pokorny R., Moldt J., Batschauer A., Essen L. O. (2007) Cryptochrome 3 from Arabidopsis thaliana: Structural and Functional Analysis of its Complex with a Folate Light Antenna. Journal of Molecular Biology, Vol. 366, pp. 954-964.

Korbel J. O., Urban A. E., Affourtit J. P., Godwin B., Grubert F., Simons J. F., et al. Paired-end mapping reveals extensive structural variation in the human genome. Science, Vol. 318, pp. 420-426.

Kress W.J., Erickson D. L., Jones F.A., Swenson N.G., Perez R., Sanjur O., Bermingham E. (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proceedings of the National Academy of Sciences of the USA, Vol. 106, pp. 18621–18626.

Kudla J., Xu Q., Harter K., Gruissem W., Luan S. (1999) Genes for calcineurin B-like proteins in Arabidopsis are differentially regulated by stress signals. Proceedings of the National Academy of Sciences USA, Vol. 96, pp. 4718-4723.

Kumar P., Gupta V. K., Misra A. K., Modi D. R., Pandey B. K. (2009). Potential of molecular markers in plant biotechnology. Plant Omics, Vol. 2, pp. 141–162.

Labra M., Miele M., Ledda B., Grassi F., Mazzei M., Sala F. (2004) Morphological characterization, essential oil composition and DNA genotyping of Ocimumbasilicum L. cultivars. Plant Science, Vol. 167, pp. 725–731.

Lande R. (2009) Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. Journal of Evolutionary Biology, Vol. 22, pp 1435-1446.

Lee S.H., Kim D. K. (1995) The Role of the 34-kDa Subunit of Human Replication Protein A in Simian Virus 40 DNA Replication in Vitro. Journal of Biological Chemistry, Vol. 270, pp. 12801-12807.

Levinson G., Gutman G. A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Molecular Biology and Evolution, Vol. 4, pp. 203-221.

Li C. W., Su R.C., Cheng C. P., Sanjaya, You S. J., Hsieh T. H., Chao T. C., Chan M. S. (2011) Tomato RAV Transcription Factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway. Plant Physiology, Vol. 156, pp. 213-227.

Li H., Ruan J., Durbin R. (2008) Mapping short DNA sequencing reads and calling variants using map- ping quality scores. Genome Research, Vol.18, pp. 1851-1858.

Lieberman M., Segev O., Gilboa N., Lalazar A., Levin I. (2004) The tomato homolog of the gene encoding UV-damaged DNA binding protein 1 (DDB1) underlined as the gene that causes the high pigment-1 mutant phenotype. Theoretical and Applied Genetics, Vol. 108, pp. 1574-1581.

Lin C., Yang H., Guo H., Mockler T., Chen J., Cashmore A.R. (1998) Enhancement of blue-light sensitivity of Arabidopsis seedlings by a blue light receptor cryptochrome 2. Proceedings of the National Academy of Sciences USA, Vol. 95, pp. 2686-2690.

Lister R., O'Malley R. C., Tonti-Filippini J., Gregory B. D., Berry C. C., Millar A. H., Ecker J. R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell, Vol. 133, pp. 523-536.

Liu J., Stevens C., Khan V.A., Lu J.Y., Wilson C.L., Adeyeye O., Kabwe M.K., Pausey, P.L., Chalutz E., Sultana T., Droby S. (1993) Application of ultraviolet-C light on storage rots and ripening of tomatoes. Journal of Food Protection, Vol. 56, pp. 868-872.

Liu L.H., Zabaras D., Bennett L.E., Aguas P., Woonton B.W. (2009) Effects of UV-C, red light and sun light on the carotenoid content and physical qualities of tomatoes during post-harvest storage. Food Chemistry, Vol. 115, pp. 495-500.

Liu V. F., Weaver, D. T. (1993) The ionizing radiation-induced replication protein A phosphorylation response differs between ataxia telangiectasia and normal human cells. Molecular and Cellular Biology, Vol. 13, pp. 7222-7231.

Lobb A., Mazzocchi M. (2007) Domestically produced food: consumer perceptions of origin, safety and the issue of trust. Acta Agriculturae Scandinavica, Vol. 4, pp. 3–12.

Maffei M. (1998) Biochimica vegetale II. Edizioni Piccin.

Mafra I., Ferreira I. M. P. V. O., Oliveira M. B. P. P. O. (2008) Food authentication by PCR-based methods. European Food Research and Technology, Vol. 227, pp. 649–665.

Mahner M., Kary M. (1997) What exactly are genomes, genotypes and phenotypes? And what about phenomes? Journal of Theoretical Biology, Vol. 186, pp. 55-63.

Mane B. G., Tanwar V. K., Girish P. S., Dixit V. P. (2006) Identifcation of species origin of meat by RAPD–PCR technique. Journal of Veterinary Public Health, Vol. 4, pp. 87–90.

Mardis E. R. (2011) A dacade's perspective on DNA sequencing technology. Nature, Vol. 470, pp. 198-203.

Mardis E.R. (2013). Next-generation sequencing platforms. Annual review of analytical chemistry (Palo Alto, Calif.), Vol. 6, pp. 287-303.

Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., Berka J., Braverman M. S., Chen Y. J., Chen Z., Dewell S. B., Du L., Fierro J. M., Gomes X. V., Godwin B. C., He W., Helgesen S., Ho C. H., Irzyk J. P., Jando S. C., Alenquer M. L. I., Jarvie T. P., Jirage K. B., Kim J. B., Knight J. R., LanzaJ. R., Leamon J. H, Lefkowitz S. M., Lei M., Li J., Lohman K. L., Lu H., Makhijani V. B., McDade K. E., McKenna M. P., Myers E. W., Nickerson E., Nobile J. R., Plant R., Puc B. P., Ronan M. T., Roth G. T., Sarkis G. J., Simons J. F., Simpson J. W., Srinivasan M., Tartaro K. R., Tomasz A., Vogt K. A., Volkmer G. A., Wang S. H., Wang Y., Weiner M. P., Yu P., Begley R. F., Rothberg J. M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature, Vol. 437, pp. 376 -380.

Marioni J., Mason C., Mane S., Stephens M., Gilad Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Research, Vol. 18, pp. 1509-1517.

Martinez I., Aursand M., Erikson U., Singstad T. E., Veliyulin E., van der Zwaag C. (2003) Destructive and non destructive analytical techniques for authentication and composition analyses of foodstuffs. Trends in Food Science & Technology, vol. 14, pp. 489-498.

Masuda H. P., Ramos, G. B., de Almeida-Engler J., Cabral L. M., Coqueiro V. M., Macrini C. M., Ferreira P. C., Hemerly A. S. (2004). Genome based identification and analysis of the pre-replicative complex of Arabidopsis thaliana. FEBS Letters, Vol. 574, pp.192-202.

Maxam A. M., Gilbert W. (1977) A new method for sequencing DNA. Proceedings of the National Academy of Sciences USA, Vol. 74, pp. 560-564.

Mazzucato A., Papa R., Bitocchi E., Mosconi P., Nanni L., Negri V., Picarella M. E., Siligato F., Soressi G. P., Tiranti B., Veronesi F. (2008) Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (Solanum lycopersicum L.) landraces. Theoretical and Applied Genetics, Vol.116, pp. 657-669.

Mazzucato A., Ficcadenti N., Caioni M., Mosconi P., Piccinini E., Reddy Sanampudi V. R., Sestili S., Ferrari V. (2010). Genetic diversity and distinctiveness in tomato (Solanum lycopersicum L.) landraces: The Italian case study of ‘A pera Abruzzese’. Scientia horticulturae, Vol. 125, pp. 55-62.

Mello E., Lusuardi S., Meloni S., Oddone M. (1982) Provenance studies of Roman amphorae by instrumental neutron activation analysis. Journal of Radioanalytical and Nuclear Chemistry, Vol. 69, pp. 259-270.

Messina C., Hammer G., Dong Z., Podlich D., Cooper M. (2009) Modelling crop improvement in a G*E*M framework via gene-trait-phenotype relationship. In: Sadras

V.O., Calderini, D.F. (Eds.), Crop Physiology: Applications for Genetic Improvement and Agronomy. Academic Press, San Diego, pp. 235-265.

Miao W., Wang X., Song C., Wang Y., Ren Y., Wang J. (2010) Transcriptome analysis of Hpa1xoo transformed cotton revealed constitutive expression of genes in multiple signaling pathways related to disease resistance. Journal of Experimental Botany, Vol. 61, pp. 4263-4275.

Michaelson M. J., Price H. J., Ellison J. R., Johnston J. S. (1991) Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry. American Journal of Botany, Vol. 78, pp. 183-188.

Miller J. C., Tanksley S. D. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. Theoretical and Applied Genetics, Vol. 80, pp. 437-448.

Mockler T. C., Guo H., Yang H., Duong H., Lin C. (1999) Antagonistic actions of Arabidopsis cryptochromes and phytochrome B in the regulation of floral induction. Development, Vol. 126, pp. 2073-2082.

Mol C. D., Harris J. M., McIntosh E. M., Tainer J. A. (1996) Human dUTP pyrophosphatase: uracil recognition by a β-hairpin and active sites formed by three separate subunits. Structure, Vol. 4, pp. 1077-1092.

Monti L. M., Santangelo E., Corrado G., Rao R., Soressi G. P., Scarascia Mugnozza G. T. (2004) Il “San Marzano”: Problematiche e prospettive in relazione alla sua salvaguardia e alla necessità di interventi genetici. Agroindustria, Vol. 3, pp. 161-170.

Morin R. Bainbridge M., Fejes A., Hirst M., Krzywinski M., Pugh T. J., McDonald H., Varhol R., Jones S. J. M., Marra M. A. (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques, Vol. 45, pp. 81-94.

Mortazavi A., Williams B. A., McCue K., Schaeffer L., Wold B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods, Vol. 5, pp. 621-628.

Mueller L. A., Lankhorst R. K., Tanksley S. D., Giovannoni J. J., White R., Vrebalov J., Fei Z., van Eck J., Buels R., Mills A. A., Menda N., Tecle I. Y., Bombarely A., Stack S., Royer S. M., Chang S. B., Shearer L. A., Kim B. D., Jo S. H., Hur C. G., Choi D., Li C. B., Zhao J., Jiang H., Geng Y., Dai Y., Fan H., Chen J., Lu F., Shi J., Sun S., Chen J., Yang X., Lu C., Chen M., Cheng Z., Li C., Ling H., Xue Y., Wang Y., Seymour G. B., Bishop G. J., Bryan G., Rogers J., Sims S., Butcher S., Buchan D., Abbott J., Beasley H., Nicholson C., Riddle C., Humphray S., McLaren K., Mathur S., Vyas S., Solanke A. U., Kumar R., Gupta V., Sharma A. K., Khurana P., Khurana J. P., Tyagi A., Sarita, Chowdhury P., Shridhar S., Chattopadhyay D., Pandit A., Singh P., Kumar A., Dixit R., Singh A., Praveen S., Dalal V., Yadav M., Ghazi I. A., Gaikwad K., Sharma T. R., Mohapatra T., Kumar Singh N., Szinay D., de Jong H., Peters S., van Staveren M., Datema E.,. Fiers M. W. E. J., van Ham R. C. H. J., Lindhout P., Philippot M., Frasse P., Regad F., Zouine M., Bouzayen M., Asamizu E.,

Sato S., Fukuoka H., Tabata S., Shibata D., Botella M. A., Perez-Alonso M., Fernandez-Pedrosa V., Osorio S., Mico A., Granell A., Zhang Z., He J., Huang S., Du Y., Qu D., Liu L., Liu. D, Wang J., Ye Z., Yang W., Wang G., Vezzi A., Todesco S., Valle G., Falcone G., Pietrella M., Giuliano G., Grandillo S., Traini A., D'Agostino N., Chiusano M. L., Ercolano M., Barone A., Frusciante L., Schoof H., Jöcker A., Bruggmann R., Spannagl M., Mayer K. X. F., Guigó R., Camara F., Rombauts S., Fawcett J. A., Van de Peer Y., Knapp S., Zamir D., Stiekema W. (2009) A snapshot of the emerging tomato genome sequence. The Plant Genome, Vol. 2, pp. 78–92.

Nachtomy O., Shavit A., Yakhini Z. (2007) Gene expression and the concept of the phenotype. Studies in History and Philosophy of Biological and Biomedical Sciences, Vol. 38, pp. 238-254.

Nagalakshmi U., Wang Z., Waern K., Shou C., Raha D., Gerstein M., Snyder M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science, Vol. 320, pp. 1344-1349.

Nicolè S., Erickson D. L., Ambrosi D., Bellucci E., Lucchin M., Papa R., Kress W. J., Barcaccia G. (2011) Biodiversity studies in Phaseolus spp. by DNA barcoding. Genome, Vol. 54, pp 529-545.

Nicotra A.B., Atkin O.K., Bonser S.P., Davidson A.M., Finnegan E.J., Mathesius U., Poot P., Purugganan M.D., Richards C.L., Valladares F., van Kleunen M. (2010) Plant phenotypic plasticity in a changing climate. Trends in Plant Science, Vol.15, pp. 684-692.

Ninu L., Ahmad M., Miarelli C., Cashmore A.R., Giuliano G. (1999) Cryptochrome 1 controls tomato development in response to blue light. The Plant Journal, Vol. 18, pp. 551-556.

Nunes-Nesi A., Carrari F., Gibon Y., Sulpice R., Lytovchenko A., Fisahn J., Graham J., Ratcliffe R. G., Sweetlove L. J., Fernie A. R. (2007) Deficiency of mitochondrial fumarase activity in tomato plants impairs photosynthesis via an effect on stomatal function. The Plant Journal, Vol. 50, pp. 1093-1106.

Orrù L., Napolitano F., Catillo G., Moioli B. (2006) Meat molecular traceability: How to choose the best set of microsatellites? Meat Science, Vol. 72, pp. 312-317.

Ozminkowski R. (2004) Pedigree of variety Heinz 1706. Report of the Tomato Genetics Cooperative 54, 26.

Palmieri L., Santoro A., Carrari F., Blanco E., Nunes-Nesi A., Arrigoni R., Genchi F., Fernie A. R., Palmieri F. (2008) Identification and Characterization of ADNT1, a Novel Mitochondrial Adenine Nucleotide Transporter from Arabidopsis. Plant Physiology, Vol. 148, pp. 1797-1808.

Parida S. K., Kalia S. K., Sunita K., Dalal V., Hemaprabha G., Selvi A., Pandit A., Singh A., Gaikwad K., Sharma T. R., Srivastava P. S., Singh N. K., Mohapatra T. (2009) Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. Theoretical and Applied Genetics, Vol. 118, pp. 327-338.

Park J. M., Park C. J., Lee S. B., Ham B. K., Shin R., Paek K. H. (2001) Over-expression of the tobacco Tsi1 gene encoding an EREBP/AP2-type transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. Plant Cell, Vol.13, pp. 1035-1046.

Park Y. H., West M. A. L., St Clair D. A. (2004) Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (Lycopersicon esculentum L.). Genome, Vol. 47, pp. 510–518.

Pasqualone A., Montemurro C., Summo C., Sabetta W., Caponio F., Blanco A. (2007) Effectiveness of microsatellite DNA markers in checking the identity of protected designation of origin extra virgin olive oil. Journal of Agricultural and Food Chemistry, Vol. 55, pp. 3857-3862.

Peakall R., Smouse P., (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes (electronic), Vol. 6, pp. 288-295.

Pearson B. M., Gaskin D. J., Segers R. P., Wells J. M., Nuijten P. J., van Vliet A. H. (2007) The complete genome sequence of Campylobacter jejuni strain 81116 (NCTC11828). Journal of Bacteriology, Vol. 189, pp. 8402-8403.

Pereira L., Martins-Lopes P., Batista C., Zanol G. C., Clímaco P., Brazão J., Eiras-Dias J. E., Guedes-Pinto H. (2012) Molecular Markers for Assessing Must Varietal Origin. Food Analytical Methods, Vol. 5, pp. 1252-1259.

Polychroniadou A., Vafopoulou A. (1985) Variations of major mineral constituents of ewe milk during lactation. Journal of dairy science, Vol. 68, pp. 147-150.

Pop M., Salzberg S. L. (2008) Bioinformatics challenges of new sequencing technology. Trends in Genetics, Vol. 24, pp. 142-149.

Porretta S. (1995) Quality evaluation of canned whole tomatoes. Food Science and Technology International, Vol. 1, pp. 97-104.

Powell W., Machray G. C., Provan J. (1996) Polymorphism revealed by simple sequence repeats. Trends in Plant Science, Vol. 1, pp. 215-222.

Provan J., Powell W., Hollingsworth P. M. (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends in Ecology & Evolution, Vol. 16, pp. 142-147.

Purcell N. (1985). Wine and wealth in ancient Italy. Journal of Roman Studies, Vol. 75, pp. 1-19.

Quinn N. L., Levenkova N., Chow W., Bouffard P., Boroevich K. A., Knight J. R., Jarvie T. J., Lubieniecki K. P., Desany B. A., Koop B. F., Harkins T. T., Davidson W. S. (2008) Assessing the feasibility of GS FLX Pyrosequencing for sequencing the Atlantic salmon genome. BMC Genomics, Vol. 9, pp. 404.

R Core Team, 2013

Rajendrakumar P., Biswal A. K., Balachandran S. M., Srinivasarao K., Sundaram R. M. (2007) Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. Bioinformatics, Vol. 23, pp. 1-4.

Ranc N., Muños S., Santoni S., Causse M. (2008) A clarified position for solanum lycopersicum var. cerasiforme in the evolutionary history of tomatoes (solanaceae). BMC Plant Biology, Vol. 8, pp. 130-146.

Rao R., Corrado G., Bianchi M, Di Mauro A. (2006) (GATA)4 DNA fingerprinting identifies morphologically characterized 'San Marzano' tomato plants. Plant Breeding, Vol. 125, pp. 173-176.

Rao R., Caramante M., Blanco A., Lanteri S., Lucchin M., Mazzucato A. (2009). Innovazioni genetiche per l'identificazione e la protezione di prodotti tipici italiani. Italian Journal of Agronomy, Vol. 4, pp. 95-101.

Raymond P., Alani A., Pradet A. (1985) ATP production by respiration and fermentation, and energy charge during aerobiosis and anaerobiosis in 12 fatty and starchy germinating seeds. Plant Physiology, Vol. 79, pp. 879-884.

Regattieri A., Gamberi M., Manzini R. (2007) Traceability of food products: General framework and experimental evidence. Journal of Food Engineering, Vol. 81, pp. 347-356.

Rick C. M., Fobes J. F. (1975). Allozyme variation in the cultivated tomato and closely related species. Bulletin of the Torrey Botanical Club, Vol. 102, pp. 376-384.

Rick C. M. (1976) Tomato, Lycopersicon esculentum (Solanaceae). In: Simmonds NW (eds) Evolution of crop plants. Longman Group, London, pp. 268-273.

Rick C. M., Fobes J. F., Holle M. (1977) Genetic variation in Lycopersicon pimpinellifolium: Evidence of evolutionary change in mating systems. Plant Systematics and Evolution, Vol. 127, pp. 139-170.

Robinson M.D., Oshlack A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology, Vol. 11, R25.

Rodriguez G. R., Munos S., Anderson C., Sim S., Michel A., Causse M., Gardener B. B. M., Francis D., van der Knaap E. (2011) Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. Plant Physiology, Vol. 156, pp. 275-285.

Ronen G., Carmel-Goren L., Zamir D., Hirschberg J. (2000) An alternative pathway to beta-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, Vol. 97, pp.11102-11107.

Ronen G., Cohen M., Zamir D., Hirschberg J. (1999) Regulation of carotenoid biosynthesis during tomato fruit development: Expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. The Plant Journal, Vol. 17, pp. 341-351.

Sadras V. O., Rebetzkeb G. J., Edmeadesc G.O. (2013) The phenotype and the components of phenotypic variance of crop traits. Field Crops Research, Vol. 154, pp. 255-259.

Saglio P. H., Raymond P., Pradet A. (1980) Metabolic activity and energy charge of excised maize root tips under anoxia control by soluble sugars. Plant Physiology, Vol. 66, pp. 1053-1057.

Sanger F., Nicklen S., Coulson A. R. (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences USA, Vol., 74, pp. 5463-5467.

Sardaro M. L. S., Marmiroli M., Maestri E., Marmiroli N. (2013) Genetic characterization of Italian tomato varieties and their traceability in tomato food products. Food Science & Nutrition, Vol. 1, pp. 54-62.

Satkoski J. A., Malhi R., Kanthaswamy S., Tito R., Malladi V., Smith D. (2008) Pyrosequencing as a method for SNP identification in the rhesus macaque (Macaca mulatta). BMC Genomics, Vol.9, pp. 256.

Scarano D., Corrado G., Caramante M., Rao R. (2011) SSR fingerprinting reveals mislabelling of commercial ‘San Marzano’ tomato products. Minerva Biotecnologica, Vol. 23, pp. 42-44.

Scarano D., Montemurro C., Corrado G., Blanco A., Rao R. (2012) DNA markers as a tool for genetic traceability of primary product in agri-food chains. Italian Journal of Agronomy, Vol. 7, pp. 346-350.

Schwengel D. A., Jedlicka A. E., Nanthakumar E. J., Weber J. L., Levitt R. C. (1994). Comparision of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. Genomis, Vol. 22, pp. 46-54.

Scrimshaw B. J. (1992). A simple nonradioactive procedure for visualization of (dC-dA)n dinucleotide repeat length polymorphisms. Bio/Techniques, Vol. 13, pp. 189.

Sharma V. P. (2012) Nature at Work - the Ongoing Saga of Evolution. Springer

Shi J., Kim K. N., Ritz O., Albrecht V., Gupta R., Harter K., Luan S., Kudla J. (1999) Novel protein kinases associated with calcineurin B-like calcium sensors in Arabidopsis. Plant Cell, Vol.11, pp. 2393-2405.

Shiu S. H., Bleecker A. B. (2001). Plant receptor-like kinase gene family: diversity, function, and signaling. Science Signaling, Vol. 113, pp. 22.

Sim S., Robbins M. D,. Chilcott C., Zhu T., Francis D. M. (2009) Oligonucleotide array discovery of polymorphisms in cultivated tomato (Solanum lycopersicum L) reveals patterns of SNP variation associated with breeding. BMC Genomics, Vol. 10, pp. 10.

Sim S. C., Robbins M. D., Van Deynze A., Michel A. P., Francis D. M. (2011) Population structure and genetic differentiation associated with breeding history and selection in tomato (Solanum lycopersicum L.). Heredity, Vol. 106, pp. 927-935.

Sim S. C., Van Deynze A., Stoffel K., Douches D. S., Zarka D., Ganal M. W., Chetelat R. T., Hutton S. F., Scott J. W., Gardner R. G., Panthee D. R., Mutschler M., Myers J. R., Francis, D. M. (2012) High-density SNP genotyping of tomato (Solanum lycopersicum L.) reveals patterns of genetic variation due to breeding. PLoS One, Vol. 7, e45520.

Smith D. R., Quinlan A. R., Peckham H. E., Makowsky K., Tao W., Woolf B., Shen L., Donahue W. F., Tusneem N., Stromberg M. P., Stewart D. A., Zhang L., Ranade S. S., Warner J. B., Lee C. C., Coleman B. E., Zhang Z., McLaughlin S. F., Malek J.A., Sorenson J. M., Blanchard A. P., Chapman J., Hillman D., Chen F., Rokhsar D. S., McKernan K. J., Jeffries T. W, Marth G. T., Richardson P. M. (2008) Rapid whole-genome mutational profiling using next-generation sequencing technologies. Genome Research, Vol.18, pp. 1638-1642.

Smulders M. J. M., Bredemeijer G., Rus-Kortekaas W., Arens P., Vosman B. (1997). Use of short microsatellites from database sequences to generate polymorphisms among Lycopersicon esculentum cultivars and accessions of other Lycopersicon species. Theoretical and Applied Genetics, Vol. 97, pp. 264-272.

Smyth G.K. (2005). Limma: linear models for microarray data. In: 'Bioinformatics and Computational Biology Solutions using R and Bioconductor'. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber, Springer, New York, pp. 397-420.

Sogin M. L., Morrison H. G., Huber J. A., Mark Welch D., Huse S. M., Neal P. R., Arrieta J. M., Herndl G. J. (2006) Microbial diversity in the deep sea and the underexplored “rare bio- sphere.” PNAS, Vol. 103, pp. 12115–12120.

Somers D.E., Devlin P.F., Kay S.A. (1998) Phytochromes and cryptochromes in the entrainment of the Arabidopsis circadian clock. Science, Vol. 282, pp. 1488-1490.

Sonnante G., Montemurro C., Morgese A., Sabetta W., Blanco A., Pasqualone A. (2009) DNA microsatellite region for a reliable quantification of soft wheat adulteration in durum wheat-based foodstuffs by Real-Time PCR. Journal of Agricultural and Food Chemistry, Vol. 57, pp. 10199-10204.

Soranzo N., Provan J., Powell W. (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. Genome, Vol. 42, pp. 158-161.

Soressi GP (1969) Il pomodoro. Ed Agricole, Bologna

Springer P.S., McCombie W.R., Sundaresan V., Martienssen R.A. (1995). Gene trap tagging of PROLIFERA, an essential MCM2- 3-5-like gene in Arabidopsis. Science, Vol. 268, pp. 877-880.

Stacklies W., Redestig H., Scholz M., Walther D., Selbig J. (2007) pcaMethods -- a Bioconductor package providing PCA methods for incomplete data. Bioinformatics, Vol. 23, pp.1164-1167.

Standard S. A., Perret D., Bray C. M. (1983) Nucleotide levels and loss of vigor and viability in germinating wheat embryos. Journal of Experimental Botany, Vol. 34, pp. 1047-1054.

Stevens M. A., Rick C. M. (1986). Genetics and breeding. The Tomato Crop: A Scientific Basis for Improvement, Atherton J. G., Rudich J., Eds., pp. 35-109, Chapman and Hall, New York (NY) USA.

Sultan S. E. (2000) Phenotypic plasticity for plant development, function and life history. Trends in plant science, Vol. 5, pp 537-542.

Suzuki M., Kao C. Y., McCartyl D. R. (1997) The Conserved B3 Domain of VIVIPAROUSI Has a Cooperative DNA Binding Activity. The Plant Cell, Vol. 9, 799-807.

Tam S. M., Samipak S., Britt A., Chetelat R. T. (2009) Characterization and comparative sequence analysis of the DNA mismatch repair MSH2 and MSH7 genes from tomato. Genetica, Vol. 137, pp. 341-354.

Tanksley S. D., Bernatzky R. (1987). Molecular markers for the nuclear genome of tomato. Plant Biology, vol. 4, Tomato Biotechnology, Nevins D. J., Jones R. A. Eds., pp. 37–44, Alan R. Liss, New York (NY) USA.

Tanksley S. D., Orton T. J. (1983). Isozymes. Plant Genetics and Breeding, Elsevier, Amsterdam, The Netherlands.

Tarazona S., Garcia-Alcalde F., Dopazo J., Ferrer A., Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. Genome research, Vol. 21, pp. 2213-222.

Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research, Vol. 17, pp. 6463-6471.

Tautz D., Renz M. (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research, Vol. 12, pp. 4127-4138.

Teletchea F., Bernillon J., Duffraisse M., Laudet V., Hänni C. (2008) Molecular identfication of vertebrate species by oligonucleotide microarray in food and forensic samples. Journal of Applied Ecology, Vol. 45, pp. 967–975.

Tent H. (1999) Research on food safety in 21st century. Food Control, Vol. 10, pp. 239–247.

Terol J., Mascarell R., Fernandez-Pedrosa V., Perez-Alonso M. (2002) Statistical validation of the identification of tuna species: Bootstrap analysis of mitochondrial DNA sequences. Journal of Agriculture and Food Chemistry, Vol. 50, pp. 963–969.

Tigchelaar E. C. (1986). Tomato breeding. Breeding for Vegetable Crops. Bassett M. J., Ed., pp. 135-171, AVI, Westport, Conn, USA.

Tonfack L. B., Moummou H., Latché A., Youmbi E., Benichou M., Pech J. C., Rest B. (2011). The plant SDR superfamily: involvement in primary and secondary metabolism. Current Topics in Plant Biology, Vol. 12, 41-53.

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature, Vol. 485, pp. 635-641.

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

Treuner K., Okuyama A., Knippers R., Fackelmayer, F. O. (1999) Hyperphosphorylation of replication protein A middle subunit (RPA32) in apoptosis Nucleic Acids Research, Vol. 27, pp. 1499-1504

Turci M., Sardaro M. L. S., Visioli G., Maestri G. (2010) Evaluation of DNA extraction procedures for traceability of various tomato products. Food Control, Vol. 21, pp.143-149.

Urao T., Yamaguchi-Shinozaki K., Shinozaki K. (2000) Two-component systems in plant signal transduction. Trends in Plant Science, Vol. 5, pp. 67-74.

Urich T., Lanzen A., Qi J., Huson D. H., Schleper C., Schuster S. C. (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLoS ONE, 3:e2527.

van Berloo R., Zhu A. G., Ursem R., Verbakel H., Gort G., van Eeuwijk F. A. (2008) Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. Theoretical and Applied Genetics, Vol. 117, pp. 89–101.

van der Horst G.T., Muijtjens M., Kobayashi K., Takano R., Kanno S., Takao M., de Wit J., Verkerk A., Eker A.P., van Leenen D., Buijs R., Bootsma D., Hoeijmakers J.H. Yasui A. (1999) Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. Nature, Vol. 398, pp. 627-630.

van Kleunen M., Fischer M. (2001) Adaptive evolution of plastic foraging responses in a clonal plant. Ecology, Vol. 82, pp. 3309-3319.

Vera J. C., Wheat C. W., Fescemyer H. W., Frilander M. J., Crawford D. L., Hanski I., Marden J. H. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology, Vol.17, pp. 1636-1647.

Vitale N., Moss J., Vaughan M. (1998) Molecular characterization of the GTPase-activating domain of ADP-ribosylation factor domain protein 1 (ARD1). The Journal of Biological Chemistry, Vol. 273, pp. 2553-2560.

Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W., Zhang J., Li J., Zhang J., Guo Y., Feng B., Li H., Lu Y., Fang X., Liang H., Du Z., LiD., Zhao Y., Hu Y., Yang Z., Zheng H., Hellmann I., Inouye M., Pool J., Yi X., Zhao J., Duan J., Zhou Y., Qin J., Ma L., Li G., Yang Z., Zhang G., Yang B., Yu C., Liang F., Li W., Li S., Li D., Ni P., Ruan J., Li Q., Zhu H., Liu D., Lu Z., Li N., Guo G., Zhang J., Ye J., Fang L., Hao Q., Chen Q., Liang Y., Su Y., A. san, Ping C., Yang S., Chen F., Li L., Zhou K., Zheng H., Ren Y., Yang L., Gao Y., Yang G., Li Z., Feng X., Kristiansen K., Ka-Shu Wong G., Nielsen R., Durbin R., Bolund L., Zhang X., Li S., Yang H., Wang J. (2008) The diploid genome sequence of an Asian individual. Nature, Vol. 456, pp.60-65.

Wang M. L., Barkley N. A., Jenkins T. M. (2009) Microsatellite markers in plants and insects. Part I. Applications of biotechnology. Genes Genomes Genomics, Vol. 3, pp. 54-67.

Wang Z., Gerstein M., Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews, Vol. 10, pp. 57-63.

Weber J. L. (1990) Informativeness of human (dC-dA)n(dG- dT)n polymorphisms. Genomics, Vol. 7, pp. 524-530.

Weising K., Gardner R. C. (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome, Vol. 42, pp. 9-19.

West A. H., Stock A. M. (2001) Histidine kinases and response regulator proteins in two-component signaling systems. Trends in Biochemical Sciences, Vol.26, pp. 369-376.

West-Eberhard M.J., (2003) Developmental Plasticity and Evolution. Oxford University Press, New York.

Wheeler D. A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A,, He W., Chen Y. J., Makhijani V., Roth G. T., Gomes X., Tartaro K., Niazi F., Turcotte C. L., Irzyk G. P., Lupski J. R., Chinault C., Song X., Liu Y., Yuan Y., Nazareth L., Qin X., Muzny D. M., Margulies M., Weinstock G. M., Gibbs R. A., Rothberg J. M. (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature, Vol. 452, pp. 872-876.

Williams C. E., St. Clair D. A. (1993). Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of Lycopersicon esculentum. Genome, Vol. 36, pp. 619–630.

Wilhelm B. T., Marguerat S., Watt S., Schubert F., Wood V., Goodhead I., Penkett C. J., Rogers J., Bähler J. (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature, Vol. 453, pp. 1239-1243.

Wilcock A., Pun M., Khanona J., Aung M. (2004) Consumer attitudes, knowledge and behaviour: a review of food safety issues. Trends in Food Science & Technology, Vol. 1, pp. 56–66.

Wold M. S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. Annual Review of Biochemistry, Vol. 66, pp. 61-92.

Woolfe M., Primrose, S. (2004) Food forensics: using DNA technology to combat misdescription and fraud. Trends in Biotechnology, Vol. 22, pp. 222-226.

Wu S. Y., Culligan K., Lamers M., Hays J. (2003). Dissimilar mispair-recognition spectra of Arabidopsis DNA-mismatch-repair proteins MSH2*MSH6 (MutSa) and MSH2*MSH7 (MutSc). Nucleic Acids Research, Vol. 31, pp. 6027-6034.

Yanez M., Caceres S., Orellana S., Bastıas A., Verdugo I., Ruiz-Lara S., Casaretto J.A. (2009) An abiotic stress-responsive bZIP transcription factor from wild and cultivated tomatoes regulates stress-related genes. Plant Cell Reports, Vol. 28, pp. 1497-1507.

Yeager M., Xiao N., Hayes R. B., Bouffard P., Desany B., Burdett L., Orr N., Matthews C., Qi L., Crenshaw A., Markovic Z, Fredrikson K. M., Jacobs K. B., Amundadottir L., Jarvie T. P., Hunter D. J., Hoover R., Thomas G., Harkins T. T., Chanock S. J. (2008) Comprehensive resequence analysis of a 136 kb region of human chromosome 8q24 associated with prostate and colon cancers. Human Genetics, Vol. 124, pp. 161-170.

Zane L., Bargelloni L., Patarnello T. (2002) Strategies for microsatellite isolation: a review. Molecular Ecology, Vol. 11, pp. 1-16.

Zernik-Kobak M., Vasunia K., Connelly M., Anderson C. W., Dixon K. (1997) Sites of UV-induced Phosphorylation of the p34 Subunit of Replication Protein A from HeLa Cells The Journal of Biological Chemistry, Vol. 272, pp. 23896-23904.

Zhang H., Zhang D., Chen J., Yang Y., Huang Z., Huang D., Wang X. C., Huang R. (2004) Tomato stress-responsive factor TSRF1 interacts with ethylene responsive element GCC box and regulates pathogen resistance to Ralstonia solanacearum. Plant Molecular Biology, Vol. 55, pp. 825-834.

Zhang X., Zhang Z., Chen J., Chen Q., Wang X. C., Huang R. (2005) Expressing TERF1 in tobacco enhances drought tolerance and abscisic acid sensitivity during seedling development. Planta, Vol. 222, pp. 494-501.

Stage

Four months stage (from September 2013 to December 2013) at the Dr. Ana Conesa's laboratory, Centro de Investigación Príncipe Felipe - Laboratory of Genomica de la Expresion Genica, Valencia, Spain. During this period it was made the elaboration of the data produced by the sequencing of the tomato transcriptome, with particular attention to the analysis of differentially expressed genes.

Publications

Scarano D., Corrado G., Rao. R. Study of molecular diversity of tomato to protect a PDO label. Proceedings of the 57th Italian Society of Agricultural Genetics Annual Congress Foggia, Italy – 16/19 September, 2013

Scarano D., Montemurro C., Corrado G., Blanco A., Rao R. (2012) DNA markers as a tool for genetic traceability of primary product in agri-food chains. Italian Journal of Agronomy, Vol. 7, pp. 346-350.

Scarano D., Corrado G., Caramante M., Rao R. (2011) SSR fingerprinting reveals mislabelling of commercial 'San Marzano' tomato products. Minerva Biotecnologica, Vol. 23, Suppl. 1 to No. 2, pp. 42-4

Scarano D., Corrado G., Rao. R. DNA testing as a means to protect 'San Marzano' PDP products. Proceedings of the Joint Meeting AGI-SIBV-SIGA Assisi, Italy – 19/22 September, 2011

Proceedings of the 57th Italian Society of Agricultural Genetics Annual Congress
Foggia, Italy – 16/19 September, 2013
ISBN **978-88-904570-3-6**

Poster Communication Abstract – 5.27

STUDY OF MOLECOLAR DIVERSITY OF TOMATO TO PROTECT A PDO LABEL

SCARANO D., CORRADO G., RAO R.

Department of Agriculture, University of Naples "Federico II", Via Università 100, 80055 Portici (Italy)

SSR, biodiversity, San Marzano

'San Marzano' is a tomato premium variety covered by an EU Protected Designation of Origin (PDO) label, cultivated in different areas of Campania region. In the recent past the 'San Marzano' was also called "red gold" for its economic value but in the '80s this crop suffered a drastic reduction both in terms of cultivated area and production. Nowadays, the 'San Marzano' PDO tomato is appreciated worldwide, and it plays a central role for the economy of the production areas. Unfortunately, media constantly refer of unscrupulous producers that adulterate, alter or replace the premium products with the goal to maximize illegally profits. There are rules aimed to protect against fraudulent substitution of quality product in food-chain but this is not enough without tools able to verify the material incoming and outcoming.

In the present study we use the SSR molecular markers to study the diversity in a wide population of 'San Marzano' with the purpose to obtain a molecular fingerprinting to use for the protection of the PDO label especially in tomato food chain.

A set of 13 SSR markers were used to characterize a population of 45 of ecotypes and accessions of 'San Marzano', all related to the genotypes included in the 'San Marzano' guideline for production ("Disciplinare di Produzione"). Moreover, among the 'San Marzano' genotypes of the "Disciplinare di Produzione" we characterized the 'Kiros' cultivar collected in different years. As expected for self-pollinating species and, in particular, for strictly self-pollinating cultivars like 'San Marzano', most of the loci revealed the presence of one allele per sample but not all loci were monomorphic in the population. This shows that ecotypes and accessions among the population are very similar but still different and it is possible to find SSR loci able to discriminate among 'San Marzano' varieties and to observe different 'San Marzano' clusters. We also evaluated the homogeneity of the 'Kiros' genotype over the years observing the absence of molecular variability in time. Furthermore, we characterized with the same set of SSR markers more than 100 tomato genotypes, both for industrial processing and fresh market and also local ecotypes and landraces. This database helps us to verify the genetic identity of the new plants. We compare the SSR allelic profile of a new genotype with the allelic profile present in our database in order to assess or not a correspondence between varieties. Through this, it was possible to verify that our population of 'San Marzano' is different from all hybrid varieties for industrial uses and also from ecotypes and landraces of Campania region.

This work explains that through SSR we can distinguish 'San Marzano' tomato from the others varieties. Furthermore, SSR alleles are preserved in food-chain, as demonstrated previously, so properly selected SSR are a very useful tool for the protection of 'San Marzano' product defended by PDO label.

Italian Journal of Agronomy 2012; volume 7:e45

DNA markers as a tool for genetic traceability of primary product in agri-food chains

Daria Scarano,[1] Cinzia Montemurro,[2] Giandomenico Corrado,[1] Antonio Blanco,[2] Rosa Rao[1]

[1]Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali, Università di Napoli Federico II, Portici (NA);
[2]Dipartimento di Biologia e Chimica Agro-forestale ed Ambientale, sezione di Genetica e Miglioramento genetico, Università di Bari Aldo Moro, Bari, Italy

Abstract

The agri-food components of the *Made in Italy* are well known all over the world, therefore they may significantly contribute to the Italian economy. However, also owing to a large number of cases of improper labelling, the Italian agro-food industry faces an ever-increasing competition. For this reason, there is a decline of consumers' confidence towards food production systems and safety controls. To prevent erroneous classification of products and to protect consumers from false in-store information, it is important to develop and validate techniques that are able to detect mislabelling at any stage of the food-chain. This paper describes some examples of genetic traceability of primary products in some important plant food chains such as durum wheat, olive and tomato, based on DNA analysis both of raw material and of processed food (pasta, olive oil, and peeled tomato).

Introduction

In the agro-food world market the label *Made in Italy* is certainly very popular, but recent data show a decrease of the Italian performance on international markets (http://www.oecd.org) probably due to the high competition of global market systems. Italian agro-food provides an important contribution to the National economy: a variety of typical products (*i.e.* wines, pasta, sauces, oils and gastronomic preparations), that have quality labels [*i.e.* Protected Designation of Origin (PDO) or Protected Geographical Indication (PGI)] and belong to consolidated cultural and culinary traditions (tomato *San Marzano*, apple *Annurca*, apricot of the Vesuvio, olives *Terra di Bari*, *Altamura bread* and many others) are popular all over and give important contributions to the export of Italian agro-products. Currently, the number of food frauds is increasing and recent data suggest that the Italian products are the most affected because of their good quality and well known fame (http://www.[illegible]/3715.html).

Markets globalization largely affects supply chain networks that, in consequence, became extremely complex to control. For example, although Italian agriculture has a long standing tradition in tomato production, presently the processed tomatoes are largely imported from China (42% of total imports in 2010; www.coldiretti.it). At the same time, there is a growing decline of consumers' confidence for food production systems and controls as well as for the credence attributes such as *healthy and safe* food chains. This is due to the progressive loss of contact between consumers and the food production systems and it is boosted by recent food crises: the cases of the Bovine Spongiform Encephalopathy (BSE) and of the contamination with dioxin or with mercury of animal feed and fish respectively. For these reasons there is an increasing demand for identification and labelling of food products (Ammendrup and Fussel, 2001; Caporale *et al.*, 2001; Carcea *et al.*, 2009; Wen *et al.*, 2010).

In this context, technologies able to trace primary products processed in food chains represent a key issue that is receiving growing attention both from producers and consumers for the relevant contributions they may offer in respect to fraud and mislabelling reports and their connections with the possibility to track the food geographic origin.

Traceability

The traceability is the ability to trace and follow a food, feed, an animal or substance intended to become part of a food or feed, through all stages of production, processing and distribution (EU regulation No. 178/2002). This definition is necessarily broad because food and feed are complex matrixes and traceability is a tool that should allow to track them, from row materials (*e.g.*, plant cultivars) to final processed products and beyond, up to market distribution and consumers. As a consequence, no traceability system is complete. Therefore different systems should be integrated to satisfy different requirements.

The important parameters that characterize a traceability system are breadth, depth, precision and running costs. Breadth relates to the

Correspondence: Dr. Rosa Rao, Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali, Università di Napoli Federico II, via Università 100, 80055 Portici (NA), Italy.
Tel. +39.081.2539204 - Fax: +39.081.2539481. E-mail: rosa.rao@unina.it

Key words: molecular markers, SSR, AFLP, tomato, olive oil, durum wheat pasta.

Received for publication: 2 May 2012.
Accepted for publication: 24 August 2012.

Italian Journal of Agronomy 2012; 7:e45
doi:10.4081/ija.2012.e45

Italian Journal of Agronomy 2012; volume 7:e45

DNA markers as a tool for genetic traceability of primary product in agri-food chains

Daria Scarano,[1] Cinzia Montemurro,[2] Giandomenico Corrado,[1] Antonio Blanco,[2] Rosa Rao[1]

[1]Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali, Università di Napoli Federico II, Portici (NA);
[2]Dipartimento di Biologia e Chimica Agro-forestale ed Ambientale, sezione di Genetica e Miglioramento genetico, Università di Bari Aldo Moro, Bari, Italy

Abstract

The agri-food components of the *Made in Italy* are well known all over the world, therefore they may significantly contribute to the Italian economy. However, also owing to a large number of cases of improper labelling, the Italian agro-food industry faces an ever-increasing competition. For this reason, there is a decline of consumers' confidence towards food production systems and safety controls. To prevent erroneous classification of products and to protect consumers from false in-store information, it is important to develop and validate techniques that are able to detect mislabelling at any stage of the food-chain. This paper describes some examples of genetic traceability of primary products in some important plant food chains such as durum wheat, olive and tomato, based on DNA analysis both of raw material and of processed food (pasta, olive oil, and peeled tomato).

Introduction

In the agro-food world market the label *Made in Italy* is certainly very popular, but recent data show a decrease of the Italian performance on international markets (http://www.oecd.org) probably due to the high competition of global market systems. Italian agro-food provides an important contribution to the National economy: a variety of typical products (*i.e.* wines, pasta, sauces, oils and gastronomic preparations), that have quality labels [*i.e.* Protected Designation of Origin (PDO) or Protected Geographical Indication (PGI)] and belong to consolidated cultural and culinary traditions (tomato *San Marzano*, apple *Annurca*, apricot of the Vesuvio, olives *Terra di Bari*, *Altamura bread* and many others) are popular all over and give important contributions to the export of Italian agro-products. Currently, the number of food frauds is increasing and recent data suggest that the Italian products are the most affected because of their good quality and well known fame (http://www.[illegible]/3715.html).

Markets globalization largely affects supply chain networks that, in consequence, became extremely complex to control. For example, although Italian agriculture has a long standing tradition in tomato production, presently the processed tomatoes are largely imported from China (42% of total imports in 2010; www.coldiretti.it). At the same time, there is a growing decline of consumers' confidence for food production systems and controls as well as for the credence attributes such as *healthy and safe* food chains. This is due to the progressive loss of contact between consumers and the food production systems and it is boosted by recent food crises: the cases of the Bovine Spongiform Encephalopathy (BSE) and of the contamination with dioxin or with mercury of animal feed and fish respectively. For these reasons there is an increasing demand for identification and labelling of food products (Ammendrup and Fussel, 2001; Caporale *et al.*, 2001; Carcea *et al.*, 2009; Wen *et al.*, 2010).

In this context, technologies able to trace primary products processed in food chains represent a key issue that is receiving growing attention both from producers and consumers for the relevant contributions they may offer in respect to fraud and mislabelling reports and their connections with the possibility to track the food geographic origin.

Traceability

The traceability is the ability to trace and follow a food, feed, an animal or substance intended to become part of a food or feed, through all stages of production, processing and distribution (EU regulation No. 178/2002). This definition is necessarily broad because food and feed are complex matrixes and traceability is a tool that should allow to track them, from row materials (*e.g.*, plant cultivars) to final processed products and beyond, up to market distribution and consumers. As a consequence, no traceability system is complete. Therefore different systems should be integrated to satisfy different requirements.

The important parameters that characterize a traceability system are breadth, depth, precision and running costs. Breadth relates to the

Correspondence: Dr. Rosa Rao, Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali, Università di Napoli Federico II, via Università 100, 80055 Portici (NA), Italy.
Tel. +39.081.2539204 - Fax: +39.081.2539481. E-mail: rosa.rao@unina.it

Key words: molecular markers, SSR, AFLP, tomato, olive oil, durum wheat pasta.

Received for publication: 2 May 2012.
Accepted for publication: 24 August 2012.

Italian Journal of Agronomy 2012; 7:e45
doi:10.4081/ija.2012.e45

pagepress

Review

amount of available information which should include key attributes (for example product or process characteristics). Depth describes how far, back or forward inside the food *netchain,* the system is able to capture the relevant information (for example, a traceability system for cultivar authentication should be able to authenticate the genetic material both back to cultivation and forward to processed food). Precision reflects the capability to provide information in a particular step of a food chain (*e.g.,* in cultivar authentication precision is the ability to identify contaminating genetic material). Running costs, that include the resources to set-up and run the system, should be in a range that would incentivize industries to activate traceability systems from which they could receive consistent benefits as they can improve the appropriate use and reliability of information, effectiveness and productivity of the organization.

The tracking method involves monitoring of stuff flow from the raw material to final sale. This can be essentially done by the manufacturer through a process of self-certification to be released, for example, after the measurement of specific physicochemical parameters that are supposed to be almost invariable during transition from raw material to final product. A company that is able to follow the entity flow and to monitor at any time raw material, semi-processed or processed products that are within the limits of its own responsibility, is a company able to streamline identification processes, to individuate causes of errors, to monitor the efficiency of each stage and of the overall process. In addition, the possibility to verify the geographic origin of the primary product increases the value of quality certification (such as PGI, PDO) (Rao *et al*., 2009a), supporting the development of local economies through the commercialization of typical food products.

For all the above mentioned considerations, we believe that the development of a highly comprehensive and efficient traceability system requires an integrated interdisciplinary approach of the different expertises active in the agro-food sciences, that allow to study the system from different points of views.

Genetic traceability

Genetic traceability, based on DNA analysis, refers to the ability of a system to identify the species or genotypes of food stuff components. This system is depth and precise as may allow to identify and trace food components along the *netchain*. For example it has been demonstrated that the DNA sequence of a single mitochondrial gene (in animals) or chloroplast gene (in plants) is fairly along species but is very much alike in individuals of the same species (Waugh, 2007; Lahaye *et al*., 2008). As a consequence, the nucleotide sequence polymorphism of these genes could be used as a *barcode*, theoretically able to identify every species (http://barcoding.si.edu/).

DNA barcoding in animals is typically based on the mitochondrial gene *cox1* sequence variations available in comprehensive databases. The system is presently routinely used for animal species identification and has successfully contributed towards food authentication. Species discrimination has been very successful in seafood industry (Nicolè *et al*., 2012) allowing to highlight frauds caused by substitution of expensive fish species by cheap one (Eugene *et al*., 2008; Filonzi *et al.,* 2010) as well as in meat industry (Chen *et al*., 2010). In plants a number of different chloroplast genes have been proposed, but there is not yet any universally accepted barcode (Lahaye *et al*., 2008). Nevertheless DNA barcoding proved effective in tracing out olive oil adulteration by canola and sunflower oil, not always easy to identify by using fatty acid analysis (Kumar *et al*., 2011).

For many plant species, the market price of an edible product is largely dependent on the cultivated varieties. Compared to animal samples, the correct identification of the cultivated plant variety requires a deeper level of genetic information as it often relies on intra-species genetic variability. The analysis of such variability at molecular level represents a reliable tool to identify DNA fingerprints of cultivated varieties. Application of DNA fingerprint in plant food traceability allows the authentication of cultivars in commercial edible products. DNA fingerprint may be determined through different types of molecular markers; the most commonly used markers are:

RFLP, Restriction Fragment Length Polymorphism (Van Ooijen *et al*., 1994; Sandbrink *et al*., 1995; Smulders *et al*., 1997);
RAPD, Randomly Amplified Polymorphic DNA (Stevens *et al*., 1995; Grandillo and Tanksley, 1996);
AFLP, Amplified Fragment Length polymorphism (Rao *et al*., 2009b; Rony *et al*., 2009);
VNTR, Variable Number of Tandem Repeat, or minisatellites (Andreakis *et al*., 2004);
SSR, Simple Sequence Repeat, or microsatellites (Smulders *et al*., 1997; He *et al*., 2003, Corrado *et al*., 2009);
CAPS, Cleaved Amplified Polymorphic Sequence (Yang *et al*., 2004; Caramante *et al*., 2009);
COS, Conserved Ortholog Set (Fulton *et al*., 2002; Van Deynze *et al*., 2007; Labate *et al*., 2009);
SNP, Single Nucleotide Polymorphism (Labate and Baldo, 2005);
In/Del, Insertion Deletion (Ching *et al*., 2004).

They are all characterized by significant discrimination power although technical feasibility and running costs may differ consistently. The direct identification of polymorphism at the DNA level provides a powerful approach for the authentication of row and processed food components as a set of DNA sequences may univocally identify a genotype.

The application of DNA fingerprint to the identification, characterization and traceability of plant species and cultivars in food chains has been demonstrated in different studies. For example, Terzi and co-workers (2004) highlighted the possibility to identify wheat species used for pasta production through AFLP markers. Similarly, SSR polymorphism was able to discriminate apple varieties and to authenticate the cultivar *Annurca* in processed food such as apple purée and nectar (Melchiade *et al*., 2007).

Tomato traceability through Simple Sequence Repeat allelic profiles

As mentioned for animal food chains, also for plant food chains it happens that, due to economic interests, premium varieties are replaced with varieties of lower quality. For example, *San Marzano*, a traditional tomato local variety with DPO, is frequently substituted with different cultivars with similar fruit shape and size (Scarano *et al*., 2011). Recently, it has been reported that SSR alleles are stable in the tomato food chain (Caramante *et al*., 2010; Turci *et al*., 2010) and that SSR allelic profiles successfully trace tomato cultivars in peeled, diced and cherry canned tomatoes (Caramante *et al*., 2010). Moreover, it was shown that SSR fingerprinting is useful to evidence erroneous labelling of processed tomato, possible consequence of the failure of the internal traceability system in establishing correct associations between registration numbers and genetic identity of samples entering and exiting the industrial process (Caramante *et al*., 2010).

Olive oils traceability through DNA marker profiles

Olive (*Olea europea* L.) is one of the oldest and most important crops

in the Mediterranean area. The database of the olive germplasm (http://www.oleadb.it) contains information on 5435 accessions, stored in more than 100 different collections. About two-thirds of the varieties are present in the Southern European countries. Unfortunately, the characterization of olive germplasm resources is complicated by the very large number of available accessions, not always properly classified, that originate several cases of synonymy or homonymy (Bartolini *et al*., 2005). In this scenario the evaluation of olive molecular diversity is particularly important both for discrimination of olive varieties and clones and for the valorisation and protection of high quality extra-virgin olive oils (Doveri *et al*., 2008; Baldoni *et al*., 2009; Rao *et al*., 2009b; Rony *et al*., 2009).

Cultivar traceability in extra-virgin olive oils has been demonstrated by a number of papers (Busconi *et al*., 2003; Pasqualone *et al*., 2004; Pafundo *et al*., 2005; Muzzalupo *et al*., 2007; Pasqualone *et al*., 2007a; Consolandi *et al*., 2008; Montemurro *et al*., 2008; Alba *et al*., 2009) and recently the possibility to identify the varieties included in simple olive oil blends was also reported (Pasqualone *et al*., 2007b; Corrado *et al*., 2011). In addition, Montemurro and colleagues (2008) demonstrated the high discrimination power of AFLP markers in olive oils identification. These authors show that one AFLP primer combination, revealing 29 polymorphic bands, identifies ten extra virgin olive oils prepared in the laboratory from ten different Italian cultivars. The extension of this analysis to commercial monovarietal olive oils would validate the powerfulness of AFLP markers application in olive food industry.

The improvement of DNA-based methodology to authenticate varieties used for olive oil production represent an important requisite to certify and protect quality labels, against fraud and mislabelling.

Quantification of soft wheat adulteration in durum wheat-based foodstuffs by real-time PCR

Pasta is a traditional Italian product, made of durum wheat (*Triticum turgidum* L. Thell. subsp. *turgidum* convar. *durum* (Desf.) MK.). Currently, Italian legislation interdicts the production of pasta containing soft wheat (*Triticum aestivum* L. Thell. subsp. *vulgare* Vill. MK.). Only a maximum of 3% of *T. aestivum* can be tolerated to account for cross contamination during the agricultural process (DPR Reg. 187/2001). However, Italy allows importation of pasta totally or partially prepared using *T. aestivum*. It is obvious that the composition of such product should be clearly indicated (Sonnante *et al*., 2009). Consequently, there is a strong interest in the development of molecular methods able to detect soft wheat in pasta.

Monitoring the presence of soft wheat in durum wheat semolina and pasta preparations has always been of interest for Italian food industry and analytical methods used to discriminate between the two species were previously based on the analyses of protein fractions (Cantagalli, 1969; Garcia-Faure *et al*., 1969; Stevenson *et al*., 1994). However, proteins have a reduced stability in comparison with nucleic acids especially in processed food as bakery products that are exposed to high temperature. Recently, a new method based on DNA screening for sequences localized in the D-genome, has been developed (Bryan *et al*., 1998; Alary *et al*., 2002; Arlorio *et al*., 2003; Terzi *et al*., 2003; Pasqualone *et al*., 2007c; Prins *et al*., 2010). A microsatellite region mapping exclusively on the wheat D-genome proved to be able to detect and quantify soft wheat in durum wheat semolina and in some typical breads made in Southern Italy like *Pane di Altamura* and *Pane di Matera*, awarded with PDO and PGI marks, respectively that have to be prepared exclusively employing durum wheat (Pasqualone *et al*. 2007c; Sonnante *et al*., 2009) .

Conclusions

Validation of product quality, and safety in agri-food sectors has become a priority. To meet this requirement the development of a highly comprehensive and efficient traceability system, that integrates multidisciplinary approaches, is highly desirable.

Genetic traceability based on DNA markers offers a valuable contribution for the identification of genetic material along the production chains, also because DNA is a molecular label difficult to remove or alter. It may restore consumers' confidence in respect of possible frauds and protect individual food choices as it can partially verify the information upon food labels.

References

Alary R, Serin A, Duviau MP, Joudrier P, Gautier MF, 2002. Quantification of common wheat adulteration of durum wheat pasta using real-time quantitative polymerase chain-reaction (PCR). Cereal Chem. 79:553-557.

Alba V, Sabetta W, Blanco A, Montemurro C, 2009. Microsatellite markers to identify specific alleles in DNA extracted from monovarietal virgin olive oils. Eur. Food Res. Technol. 229:375-82.

Ammendrup S, Fussing L, 2001. Legislative requirements for the identification and traceability of farm animals within the European Union. Rev. Sci. Tech. OIE 20:437-444.

Andreakis N, Giordano I, Pentangelo A, Fogliano V, Graziani G, Monti LM, Rao R, 2004. DNA fingerprinting and quality traits of Corbarino cherry-like tomato landraces. J. Agr. Food. Chem. 52:3366-3377.

Arlorio M, Coisson J, Cereti E, Travaglia F, Papasso M, Martelli A, 2003. Polymerase chain reaction (PCR) of puroindoline b and ribosomal/puroindoline b multiplex PCR for the detection of common wheat (Triticum aestivum) in Italian pasta. Eur. Food Res. Technol. 216:253-258.

Baldoni L, Cultrera NG, Mariotti R, Ricciolini C, Arcioni S, Vendramin GG, Buonamici A, Porceddu A, Sarri V, Ojeda MA, Trujillo I, Rallo L, Belaj A, Perri E, Salimonti A, Muzzalupo I, Casagrande A, Lain O, Messina R, Testolin R, 2009. A consensus list of microsatellite markers for olive genotyping. Mol. Breeding 24:213-231.

Bartolini G, Prevost G, Messeri C, Carignani G, 2005. Olive germplasm: cultivars and world-wide collections. Available from: http://apps3.fao.org/wiews/olive/oliv.jsp

Busconi M, Foroni C, Corradi M, Bongiorni C, Cattapan F, Foghera C, 2003. DNA extraction from olive oil and its use in the identification of the production cultivar. Food Chem. 83:127-134.

Bryan GJ, Dixon A, Gale MD, Wiseman G, 1998. A PCR-based method for the detection of hexaploid bread wheat adulteration of durum wheat and pasta. J. Cereal Sci. 28:135-145.

Cantagalli P, 1969. Control of the adulteration of durum wheat macaroni by the immunodiffusion method. Ann. Ist. Super. Sanità 5:414-416.

Caporale V, Giovannini A, Francesco C, Calistri P, 2001. The importance of traceability of animals and animal products in epidemiology. Rev. Sci. Tech. OIE 20:372-378.

Caramante M, Corrado G, Monti LM, Rao R, 2010. Simple Sequence Repeats are able to trace tomato cultivars in tomato food chains. Food Control 22:549-554.

Caramante M, Rao R, Monti LM, Corrado G, 2009. Discrimination of 'San Marzano' accessions: A comparison of minisatellite, CAPS and SSR markers in relation to morphological traits. Sci. Hortic. 120:560-564.

Carcea M, Brereton P, Hsu R, Kelly S, Marmiroli N, Melini F, Soukoulis

C, Wenping D, 2009. Food authenticity assessment: ensuring compliance with food legislation and traceability requirements. In: R. Poms and S. Cauvein (eds.) Quality Assurance and Safety of Crops & Foods. ICC and Blackwell Publ. Ltd., pp 93-100.

Chen SY, Liu YP, Yao YG, 2010. Species authentication of commercial beef jerky based on PCR-RFLP analysis of the mitochondrial 12S rRNA gene. J Genet Genomics 37:763-769.

Consolandi C, Palmieri L, Doveri S, Maestri E, Marmiroli N, Reale S, Lee D, Baldoni L, Tosti N, Severgnini M, De Bellis G, Castiglioni B, 2007. Olive variety identification by ligation detection reaction in a universal array format. J. Biotech. 129:565-574.

Corrado G, Imperato A, La Mura M, Perri E, Rao R, 2011. Genetic diversity among olive varieties of southern italy and the traceability of olive oil using SSR markers. J. Hortic. Sci. Biotech. 86:461-466.

Corrado G, La Mura M, Ambrosino O, Pugliano G, Varricchio P, Rao R, 2009. Relationships of Campanian olive cultivars: comparative analysis of molecular and phenotypic data. Genome 52:692-700.

Doveri S, Sabino Gil F, Díaz A, Reale S, Busconi M, da Câmara Machado A, Martín A, Fogher C, Donini P, Lee D, 2008. Standardization of a set of microsatellite markers for use in cultivar identification studies in olive (Olea europaea L.). Sci. Hortic. 116:367-373.

Eugene H, Wong K, Hanner RH, 2008. DNA barcoding detects market substitution in North American seafood. Food Res. Int. 41:828-837.

Filonzi L, Chiesa S, Vaghi M, Nonnis Marzano F, 2010. Molecular barcoding reveals mislabelling of commercial fish products in Italy. Food Res. Int. 43:1383-1388.

Fulton TM, van der Hoeven R, Eannetta NT, Tanksley SD, 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457-1467.

Garcia-Faure R, Merck-Luengo JG, Garcia-Olmedo F, 1969. Detection of flour or farina from Triticum aestivum in macaroni by starch gel electrophoresis of water-soluble proteins. Cereal Chem. 46:623-625.

Kumar S, Kahlon TS, Chaudhary S, 2011. A rapid screening for adulterants in olive oil using DNA barcodes. Food Chem. 127:1938-1941.

Grandillo S, Tanksley SD, 1996. Genetic analysis of RFLPs, GATA microsatellites and RAPDs in a cross between L. esculentum and L. pimpinellifolium. Theor. Appl. Genet. 92:957-965.

He C, Poysa V, Yu K, 2003. Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among Lycopersicon esculentum cultivars. Theor. Appl. Genet 106:363-373.

Labate JA, Baldo AM, 2005. Tomato SNP discovery by EST mining and resequencing. Mol. Breeding 16:343-349.

Labate JA, Robertson LD, Wu F, Tanksley SD, Baldo AM, 2009. EST, COSII, and arbitrary gene markers give similar estimates of nucleotide diversity in cultivated tomato (Solanum lycopersicum L.). Theor. Appl. Genet. 118:1005-1014.

Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough T.G, Savolainen V. 2008. DNA barcoding the floras of biodiversity hotspots. P. Natl. Acad. Sci. USA 105:2923-2928.

Melchiade D, Foroni I, Corrado G, Santangelo I, Rao R, 2007. Authentication of the 'Annurca' apple in agro-food chain by amplification of microsatellites loci. Food Biotech. 21:33-43.

Montemurro C, Pasqualone A, Simeone R, Sabetta W, Blanco A, 2008. AFLP molecular markers to identify virgin olive oils from single Italian cultivars. Eur. Food Res. Technol. 226:1439-1444.

Muzzalupo I, Pellegrino M, Perri E, 2007. Detection of DNA in virgin olive oils extracted from destoned fruits. Eur. Food Res. Technol. 224:469-475.

Nicolè S, Negrisolo E, Ecchер G, Mantovani R, Patarnello T, Erickson DL, Kress WJ, Barcaccia G, 2012. DNA barcoding as a reliable method for the authentication of commercial Seafood Products. Food Technol. Biotech. (In press).

Pafundo S, Agrimonti C, Marmiroli N, 2005. Traceability of plant contribution in olive oil by Amplified Fragment Length Polymorphisms. J. Agric. Food Chem. 53:6995-7002.

Pasqualone A, Montemurro C, Caponio F, Blanco A, 2004. Identification of Virgin Olive Oil from Different Cultivars by Analysis of DNA Microsatellites. J. Agric. Food Chem. 52:1068-1071.

Pasqualone A, Montemurro C, Grinn-Gofron A, Sonnante G, Blanco A, 2007a. Detection of soft wheat in semolina and durum wheat bread by analysis of DNA microsatellites. J. Agric. Food Chem. 55:3312-3318.

Pasqualone A, Montemurro C, Summo C, Sabetta W, Caponio F, Blanco A, 2007b. Effectiveness of microsatellite DNA markers in checking the identity of Collina di Brindisi PDO Extra Virgin olive oil. J. Agric. Food Chem. 55:3857-3862.

Pasqualone A, Montemurro C, Summo C, Sabetta W, Caponio F, Blanco A, 2007c. Effectiveness of microsatellite DNA markers in checking the identity of protected designation of origin extra virgin olive oil. J. Agric. Food Chem. 55:3857-3862.

Prins TW, van Dijk JP, Van Hoef AMA, Voorhuijzen MM, Broeders S, Trapmann S, Seyfarth R, Hougs L, Schoen CD, Aarts HJM, Kok EJ. 2010 Towards a multiplex cereal traceability tool using padlock probe ligation on genomic DNA. Food Chem. 118:966-973.

Rao R, Caramante M, Blanco A, Lanteri S, Lucchin M, Mazzucato A, 2009a. Innovazioni genetiche per l'identificazione e la protezione di prodotti agricoli italiani. Ital. J. Agron. 3:95-101.

Rao R, La Mura M, Corrado G, Ambrosino O, Foroni I, Perri E, Pugliano G, 2009b. Genetic diversity of olive cultivars using AFLP and morphological traits. J. Hortic. Sci. Biotech. 84:261-266.

Rao R, Baalbaki R, Kalaitzis P, Talhouk SN, 2009. Molecular characterization of Lebanese olive germplasm. Tree Genet. Genomes 5:109-115.

Sandbrink JM, Van Coijen JW, Purimahua CC, Vrielink M, Verkerk R, Zabel P, Lindhout P, 1995. Localization of genes for bacterial canker resistance in Lycopersicon peruvianum using RFLPs. Theor. Appl. Genet. 95:444-450.

Scarano D, Corrado G, Caramante M, Rao R, 2011. SSR fingerprinting reveals mislabelling of commercial 'San Marzano' tomato products. Minerva Biotecnol. 23:42-44.

Smulders MJM, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B, 1997. Use of short microsatellites from database sequences to generate polymorphisms among Lycopersicon esculentum cultivars and accessions of other Lycopersicon species. Theor. Appl. Genet. 97:264-272.

Sonnante G, Montemurro C, Morgese A, Sabetta W, Blanco A, Pasqualone A, 2009. DNA microsatellite region for a reliable quantification of soft wheat adulteration in durum wheat-based foodstuffs by Real-Time PCR. J. Agric. Food Chem. 57:10199-10204.

Stevens MR, Lamb EM, Rhoads DD, 1995. Mapping the Sw-5 locus for tomato spotted wilt virus resistance in tomatoes using RAPD and RFLP analyses. Theor. Appl. Genet. 90:451-456.

Stevenson A, McCarthy PK, Griffin M, 1994. Polyclonal antisera against unheated and heated common wheat specific gamma and omega gliadins for the detection and adulteration of durum wheat and durum wheat products with common wheats. Food Agric. Immunol. 6:435-442.

Terzi V, Malnati M, Barbanera M, Stanca AM, Faccioli P, 2003 Development of analytical systems based on real time PCR for Triticum species-specific detection and quantitation of bread wheat contamination in semolina and pasta. J. Cereal Sci. 38:87-94.

Terzi V, Morcia C, Giovanardi D, D'Egidio MG, Stanca AM, Faccioli P, 2004. DNA-based analysis for authenticity assessment of monova-

rietal pasta. Eur. Food Res. Technol. 219:428-431.
Turci M, Sardaro MLS, Visioli G, Maestri G, 2010. Evaluation of DNA extraction procedures for traceability of various tomato products. Food Control 21:143-149.
Van Deynze AE, Wilkins TA, Stoffel K, Lee M, Stelly D, Kozik A, 2007. A set of informative markers designed specifically for breeding cotton. XV PAG, San Diego, California, 657.
Van Ooijen JW, Sandbrink JM, Vrielink M, Verkerk R, Zabel P, Lindhout P, 1994. An RFLP linkage map of Lycopersicon peruvianum. Theor. Appl. Genet. 89:1007-1013.
Wen J, Hu C, Zhang L, Luo P, Zhao Z, Fan S, Su T, 2010. The application of PCR-RFLP and FINS for species identification used in sea cucumbers (Aspidochirotida: Stichopodidae) products from the market. Food Control 21:403-407.
Waugh J, 2007. DNA barcoding in animal species: progress, potential and pitfalls. BioEssays 29:188-197.
Yang W, Bai X, Kabelka E, Eaton C, Kamoun S, van der Knaap E, Francis DM, 2004. Discovery of single nucleotide polymorphisms in Lycopersicon esculentum by computer aided analysis of expressed sequence tags. Mol. Breeding 14:21-34.

MINERVA BIOTEC 2011;23 Suppl. 1 to No. 2-62-4

SSR fingerprinting reveals mislabelling of commercial 'San Marzano' tomato products

D. SCARANO, G. CORRADO, M. CARAMANTE, R. RAO

Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali, Portici, Naples, Italy

Introduction

The tomato is one of the most frequently consumed vegetable world-wide (http://faostat.fao.org/default.aspx). In many countries, production is largely based on hybrid varieties that are incessantly introduced by breeding companies. However, some traditional cultivar is still playing a significant role in the world market. One notable example is the 'San Marzano', a classic example of local variety with premium value[1]. Since 1999, a EU Protected Designation of Origin (PDO) label protects the legitimate production of 'San Marzano'.

Unfortunately, there have been cases in the food sector that involved the substitution of a premium product with a less expensive or less desirable item, suggesting the possibility of deliberate mislabelling for economic gain[2]. Food genetic traceability (*i.e.* a set of DNA-based technologies and methods that allow to identify an agricultural product at every step of production, processing and commercialisation, from farm to table and *viceversa*) is central for the identification of improper labelling of processed food and feed. In food traceability by DNA-testing, the molecular analysis aims at the identification of DNA polymorphisms that are indicative of a certain genotype.

The tomato genome, as well as those of other plants, includes many repeated DNA sequences. Among the different classes of genomic repetitive sequences of interest for the molecular fingerprint, there are microsatellites, also known as Simple Sequence Repeats (SSRs) or Short Tandem Repeats (STRs)[3].

The combination of the PCR technique, like SSRs, and very sensitive detection systems, such as capillary electrophoresis, has recently become the most common approach for the traceability of foods deriving from animal or plant sources[2].

In a recent work, we showed that SSR allelic profiles are conserved in the tomato food-chain[4]. The present study shows that SSR markers can identify products of the agro-food chain but also reveal mislabelling of commercial products. Our data provided information that can be used for setting up extensive controls and certification systems of tomato products.

Materials and Methods

The plant material includes seven lots of tomato (*Solanum lycopersicum* L.) berries, which were intended for industrial processing in a local factory. These lots were identified by a number according to a blind coding system. We also used three peeled tomato cans labelled as 'San Marzano', of brands distributed at national and international level.

Genomic DNA isolation and PCR amplification were performed as previously reported[4]. The SSR loci employed were LEaat002, LEcaa001, LEct001, LEctt001, LEta003, LEtat002, LEta015, LEEF1Aa, LE20592 and LE21085[5]. Fluorescent amplicons were analyzed by capillary electrophoresis (ABI Prism

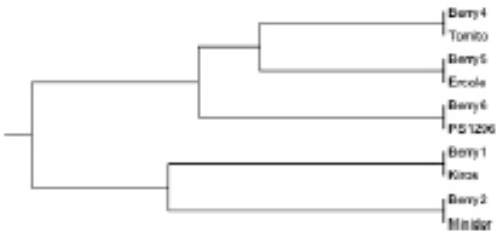


Figure 1.—Graphical representation of the relationship among the analyzed samples.

3100 Avant, Applied Biosystems) using the ABI Prism Genotyper 2.5 (v. 3.7) software (Applied Biosystems). Data were compared to a database of SSR profile of 42 tomatoes cultivar, already available in the laboratory.

For cluster analysis, monomorphic SSR loci were excluded. The estimation of the genetic distance and the corresponding tree were obtained as previously described[1,4].

Results

The set of 10 SSR markers used was useful for the discrimination of the seven lots of tomato berries, hereinafter indicated as "Berry 1", "Berry 2" and so forth. The allelic profiles obtained from berries were then compared with a SSR database produced in our lab and it was possible to reveal the genetic identity of five of the seven lots considered (Figure 1).

The same set of 10 SSR were used on three commercial tomato products. Seven primer pairs (LEcaa001, LEct001, LEctt001, LEta003, LEta015, LE20592 and LE21085) successfully amplified fragments from 94 to 167 bp from DNA isolated by three commercial products, while the others, expected

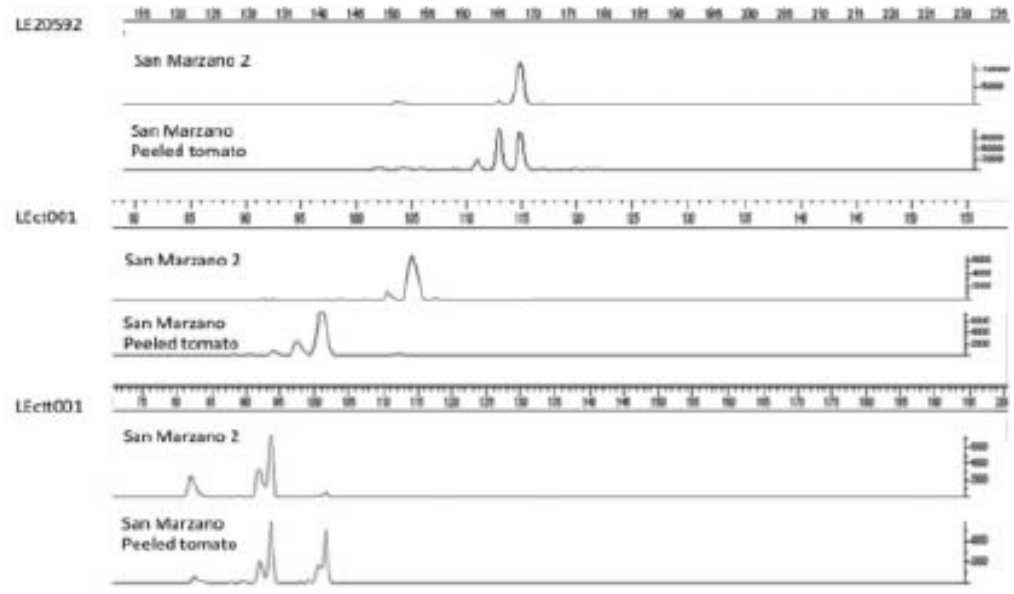


Figure 2.—Electropherograms of three SSR loci. 'San Marzano 2': DNA from fruits of true 'San Marzano' variety; 'San Marzano' peeled tomato: DNA peeled tomato cans labelled as 'San Marzano'.

to yield larger fragments, failed to give consistent amplifications. The allelic profiles of the processed tomato was heterozygous at four loci while at the same loci the 'San Marzano' and 'Kiros' tomatoes were, as expected, homozygous (Figure 2). Furthermore, at some loci we did not find the same alleles of the 'San Marzano' control in the peeled tomato. The allelic profiles obtained from the peeled tomatoes labelled as 'San Marzano' did not match the profiles of the 'Kiros', 'San Marzano 2' varieties and of other genetically similar accessions cultivated in the PDO area[1], indicating that it is possible to exclude the presence of fruits of these cultivars in the analysed products.

Discussion

The food industry is the second largest manufacturing industry in Italy and food products are a significant part of the "Made in Italy" export. Canned and peeled tomatoes are indeed the most exported food product from Italy. According to the agreement between the tomato producers associations and the industries for canned food, the former is sold at about 40 euro cents per kilo (http://consorziopomodorosanmarzanodop.it), the latter at about 9 euro cents per kilo (http://agronotizie.imagelinenetwork.com). Hence, the possibility of verifying the genetic identity of food products provides a means to protect the value of quality labels such as PGI and PDO.

DNA molecular markers such as SSR, are suitable for the genetic identification of raw and finished material along the production chain, since DNA is a molecular label very difficult to remove or alterate[5].

The results of the blind analysis allowed the identification of five out of seven fruit lots, implying the need to extend the molecular analysis to other tomato varieties. We have also investigated three peeled tomato cans labelled as 'San Marzano' and SSR fingerprint revealed a possible case of mislabelling, as it was possible to exclude the presence of the 'San Marzano 2' and the 'Kiros' cultivars. These two cultivars, along with varieties that shall be deemed as to be essentially derived from 'San Marzano' are the only cultivars allowed to be used for PDO productions.

Conclusions

In conclusion, we demonstrated that the use of SSR markers is a useful approach to identify products exiting the tomato food chain and to determine mislabelling in commercial tomato products.

References

1. Caramante M, Rao R, Monti LM, Corrado G. Discrimination of 'San Marzano' accessions: A comparison of minisatellite, CAPS and SSR markers in relation to morphological traits. Sci Hortic. 2009; 120:560-564.
2. Tsolchea F, Maudet C, Hanni C. Food and forensic molecular identification: update and challenges. Trends Biotechnol. 2005;23:359-366.
3. Hancock JM. Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: Goldstein D, Schlotterer C, editors. Microsatellites-Evolution and Applications. New York: Oxford University Press. 1999;1-9.
4. Caramante M, Corrado G, Monti LM, Rao R. Simple Sequence Repeats are able to trace tomato cultivars in tomato food chains. Food Control. 2010; 22:549-554.
5. Martinez I, Aursand M, Erikson U, Singstad T, Veliyulin E, van der Zwaag C. Destructive and non destructive analytical techniques for authentication and composition analyses of foodstuffs. Trends in Food Science & Technology. 2003;14:489-498.

Proceedings of the Joint Meeting AGI-SIBV-SIGA
Assisi, Italy – 19/22 September, 2011
ISBN **978-88-904570-2-9**

Poster Communication Abstract – 8.29

DNA TESTING AS A MEANS TO PROTECT 'SAN MARZANO' PDO PRODUCTS

SCARANO D., CORRADO G., RAO R.

Department of Soil, Plant, Environmental and Animal Production Sciences, School of Biotechnology, University of Naples "Federico II, Via Università 100, 80055 Portici (Italy)

SSR, blind analysis, 'San Marzano', peeled tomato, traceability

The tomato is one of the most frequently consumed vegetable world-wide and its production is largely based on hybrid varieties. However, some traditional cultivars play a significant role in the world market. One example is the 'San Marzano', a well-known local variety whose production is defended by an EU Protected Designation of Origin (PDO) label.

In the food sector, there have been cases that involved the substitution of a premium product with a less expensive or less desirable item, suggesting the possibility of deliberate mislabelling for economic gain. The present study shows that SSR markers can identify products of the tomato-food chain but also reveal mislabelling of commercial products.

We firstly used ten SSR markers to discriminate and seven blind coded lots of tomato berries, allowing the identification of five samples. Furthermore we also analysed commercial peeled tomatoes that were labelled as 'San Marzano'. Out of the ten SSR employed, seven successfully amplified fragments smaller than 200 bp from DNA isolated from tomato products. The allelic profiles obtained from the peeled tomatoes labelled as 'San Marzano' did not match the profiles of the 'Kiros', 'San Marzano 2' or of other genetically close accessions that can be used for the PDO production. Thus, molecular fingerprinting indicated that it is possible to exclude the presence of 'San Marzano' fruits in the analysed commercial products.

We demonstrated that selected SSR markers are a useful tool to protect the value of products entering and exiting the tomato food chain, as they are able to reveal mislabelling in commercial products.