Journal of Geochemical Exploration 121 (2012) 62–68

Contents lists available at SciVerse ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/jgeoexp

Biological and geochemical markers of the geographical origin and genetic identity of potatoes

Paola Adamo [a,1], Mariavittoria Zampella [b,*,1], Christophe R. Quétel [b], Riccardo Aversano [a], Fabrizio Dal Piaz [c], Nunziatina De Tommasi [c], Luigi Frusciante [a], Massimo Iorizzo [a], Laura Lepore [c], Domenico Carputo [a]

[a] Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali, Università di Napoli Federico II, Naples, Italy
[b] European Commission, Joint Research Centre, Institute for Reference Materials and Measurements (IRMM), Geel, Belgium
[c] Dipartimento di Scienze Farmaceutiche e Biomediche, Università degli Studi di Salerno, Italy

ARTICLE INFO

Article history:
Received 6 December 2011
Accepted 1 July 2012
Available online 21 July 2012

Keywords:
Traceability
DNA analysis
Trace elements
Strontium isotopic signature
Secondary metabolites
Discriminant analysis

ABSTRACT

There is a growing interest in agriculture productions combining safety and quality attributes with clear regional identity. In the last few years several methods have been employed for food authentication and traceability. In this study we tested geochemical data for elemental concentrations of Mn, Cu, Zn, Rb, Sr and Cd and strontium isotope ratio in combination with biological data of 11 secondary metabolites and DNA as markers for the authentication of the origin of early potatoes at small geographical scale levels in Italy. DNA fingerprints through 12 SSR (simple sequence repeat) primer pairs allowed cultivar identification, confirming the discrimination power of molecular markers. Element concentrations, strontium isotope ratio and secondary metabolite data, through multivariate statistics (partial least squares discriminant analysis, PLS-DA), made it possible to clearly assign all the potato samples to the respective administrative regions of cultivation. The validation of the models was successful. It included external prediction tests on 20% of the data randomly selected from each administrative province and a study on the robustness of these multivariate data treatments to uncertainties on measurement results.

1. Introduction

Consumers in the European Union need guarantees on the geographical origin of food products, which they take as a pledge for safety and quality (regulations EEC 2081/92 and EC 1898/06). According to regulation EC 178/2002, "'traceability' means the ability to trace and follow a food, feed, food-producing animal or substance intended to be, or expected to be incorporated into a food or feed, through all stages of production, processing and distribution". Finding the appropriate tools to authenticate the origin of food products and establish their 'traceability' may be challenging.

Different combinations of chemical and/or genetic markers have been tested for the identification of crop varieties, animal species or genetically modified organism (GMO) products, for the establishment of frauds or counterfeiting, and for the determination of animal paternity (Kelly et al., 2005; Peres et al., 2007; Zhang et al., 2011). For instance, the combination of C, N, O, S and Sr isotopic signatures, alone or with element concentrations, was successful in the authentication of origin of wheat at a large geographical scale level (Europe), encompassing an extensive range of geographical and environmental characteristics (Goitom Asfaha et al., 2011). Similarly, DNA (deoxyribonucleic acid) analysis was employed in studies on the traceability and authenticity of different types of plant, including tomato (Caramante et al., 2010) and wheat (Pasqualone et al., 2010). An alternative approach may be the analysis of the metabolite profile whose synthesis is controlled by specific genes. In grape, for instance, Rocha et al. (2007) detected the monoterpenoid profile and proposed it as an efficient tool to trace the genetic footprint and potentially the varietal origin of wines.

Among horticultural crops cultivated in Italy, the potato ranks second in terms of economic importance, with about 2 million tons, obtained over an area of 90,000 ha. Under favorable climatic conditions potatoes can be grown in a cycle that is much earlier than the typical spring–summer one. In Italy this type of production mainly occurs in the southern regions of Apulia, Sicily and Campania, where tubers are planted starting from late November and are harvested from the end of March. Early potatoes are mainly exported, with a value of more than 50 million euros per year. They are considered typical regional products with specific quality attributes, such as low dry matter content and high reducing sugar content. They also possess historical and culinary characteristics that have induced the Italian Ministry of Agriculture to include them in the list of "traditional food products", and fund the TIPIPAPA project ("tipicizzazione e caratterizzazione di varietà precoci di patata con l'impiego di tecniche molecolari e spettroscopiche") to support fight against fraud on early

* Corresponding author. Tel.: +32 14 571 969; fax: +32 14 571 548.
E-mail address: Mariavittoria.ZAMPELLA@ec.europa.eu (M. Zampella).
[1] Co-first authors.

0375-6742/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.gexplo.2012.07.006

potatoes from the south of Italy. The main objective of this project was the identification of suitable markers of the geographical origin and genetic identity of potatoes. In the project, all production sites considered were less than 10 km distance from the sea and distributed over a rather small size area (about 58,600 km^2), and under relatively similar average meteorological conditions. If differences in climate characteristics are too small to allow the discrimination between production sites, variety discrimination may be rather difficult. As was shown recently (Zampella et al., 2011), geopedological factors governing the bioavailability of elements in soils and multivariate interactions between soil markers in plants offer a very good alternative. The strontium (Sr) isotopic signature and selected elemental concentrations ([Mn], [Cu], [Zn], [Rb], [Sr] and [Cd]) in early potatoes were successfully combined for authentication of the origin of the early potatoes studied. The prediction abilities for potatoes grown in soils developed from alluvial sediments and volcanic substrates were excellent. Potatoes grown in soils covering carbonate rocks could be grouped together but, in this case, more discriminative parameters are required to allow the differentiation between sites.

In the current paper, these geochemical data for trace elements and $n(^{87}Sr)/n(^{86}Sr)$ were re-examined, in combination with results from biological markers (secondary metabolites and DNA), to identify further ways of authenticating the origin of vegetables at such a small geographical scale level.

2. Materials and methods

2.1. Description of the sampling areas

Potato tubers were collected on May 2009 from varieties (Spunta, Elvira, Sieglinde, Arinda, and Agria) grown off-season in typical potato producing areas of three administrative regions of southern Italy (Apulia, Sicily and Campania). Various sampling sites for each variety were selected in each region for a total of 12 sites. The sampling sites are all coastal sites (distance from the Mediterranean Sea less than 10 km) and, according to the geographical coordinates recorded by GPS (global positioning system), distributed within a range of 4°34′ latitude N and 3°46′ longitude E (Table 1). Details of the climatic, geological and soil characteristics of the sampling sites are reported elsewhere (Zampella et al., 2011). The weather conditions of all sampling sites during the potato growth season (01/01/09–15/06/09) are typical of a Mediterranean climate, and not very much differentiated, with average daily temperatures of 13–14 °C (11 °C only in Lecce) and total precipitations ranging between 262 mm (Foggia) and 781 mm (Messina). The main geological substrates underlying soils are alluvial sediments (Foggia and Messina), vulcanites (Naples and Catania) and carbonate rocks (all the others). Soils, according to requirements for potato cultivation, have loose, well drained, loam textures and low organic matter contents. In most cases pH is neutral–subalkaline, with the exception of NA-Afragola, LE-Alliste-Cisternella, ME-Monteforte S. Giorgio and CT-Acireale-Riposto where it is acid. Nitrogen content is very low in soils from FG-Manfredonia, due to the sandy texture. Exchangeable K, considered readily available for plant growth, is particularly high in Campania soils according to their K-rich volcanic parent material. All soils developed on volcanic substrates are enriched in amorphous Fe oxides, while high amounts of Fe in crystalline oxides occur in soils from Apulia and Sicily formed from calcareous rocks.

2.2. Potato sampling strategy

In each selected site, a sampling strategy with 3 replications was adopted. Potato tubers were harvested at physiological maturity from five plants for each replicate. About 30 tubers (3 samplings of 10 tubers each) of homogeneous size were collected from each locality. Already commercialized potatoes originating from the same fields in FG-Manfredonia and LE-Alliste-Cisternella were integrated into the study. Part of the tubers were stored at 5 °C for multi-element and Sr-isotope ratio analyses, the other part was washed with ultra-pure water without peeling and lyophilized for DNA and secondary metabolite studies. For lyophilization, each tuber was cut in eight identical longitudinal slices discarding the pieces from the top, immediately frozen in liquid N_2, and then homogenized.

2.3. Measurements in potato samples

2.3.1. Multi-element and Sr-isotope ratio

The analytical protocols developed and applied for multi-element and stable Sr-isotope ratio measurements (incl. the estimation of associated measurement uncertainties) in potatoes have been described elsewhere (Zampella et al., 2011). Briefly, for each potato sample a cubic section of the central pulp was produced under ultra-clean conditions, lyophilized and microwave-assisted acid digested (portions of ~0.5 g). The aliquot reserved for Sr isotopic measurements was further matrix separated by ion chromatography while the remaining fraction was evaporated to near-dryness and step-wise re-diluted in 2% HNO_3 for multi-element measurements with a quadrupole inductively coupled plasma mass spectrometer (ICPMS). A multi-collector ICPMS was used for the Sr isotope ratio measurements.

2.3.2. Secondary metabolites

Freeze-dried powder tubers (400 mg) were placed into a 2 mL screw cap tube along with 0.9 mL of extraction buffer (50% methanol, 2.5% metaphosphoric acid, 1 mM ethylenediaminetetraacetic acid, EDTA). Tubes were immersed in an ultrasonic bath cooled with ice (Grant, Chelmsford, Essex, UK), sonicated for 15 min at a maximum speed and centrifuged for 7 min at 4 °C (13,000 rpm). The supernatant

Table 1
Sample identity (ID), variety and geographical location of the studied potato samples.

Sample ID*	Variety	Region	Province	Site name	Latitude N	Longitude E
SP-P-A	Spunta	Apulia	Lecce (LE)	Alliste-Chianchi	39° 56.254	18° 06.344
SG-P-C	Sieglinde	Apulia	Lecce	Alliste-Cisternella	39° 54.655	18° 05.102
EL-P-E	Elvira	Apulia	Foggia (FG)	Manfredonia	41° 32.563	15° 53.514
SP-P-F	Spunta	Apulia	Bari (BA)	Bitonto	41° 06.641	16° 43.635
SP-P-I	Spunta	Apulia	Bari	Mola di Bari	41° 02.051	17° 07.388
AR-S-A	Arinda	Sicily	Siracusa (SR)	Contrada Milocca	37° 01.581	15° 15.862
SP-S-B	Spunta	Sicily	Siracusa	Agro di Cassibile	36° 58.593	15° 12.270
SP-S-C	Spunta	Sicily	Catania (CT)	Acireale	37° 39.885	15° 09.517
AR-S-E	Arinda	Sicily	Catania	Riposto	37° 41.306	15° 10.957
SG-S-F	Sieglinde	Sicily	Messina (ME)	Torregrotta	38° 12.503	15° 20.465
AR-S-G	Arinda	Sicily	Messina	Monteforte S. Giorgio	38° 12.536	15° 20.290
AGR-C-A	Agria	Campania	Napoli (NA)	Afragola	40° 55.230	14° 20.370

* ID code: XX-Y-Z, where XX = variety (SP = Spunta; SG = Sieglinde; EL = Elvira; AR = Arinda; AGR = Agria), Y = region (P = Apulia; S = Sicily; C = Campania), and Z = progressive letter for sample identification.

was transferred to a clean tube. The remaining pellet was re-extracted with 0.6 mL of extraction buffer and centrifuged (1 min, 13,000 rpm). The supernatants were combined, centrifuged, and concentrated in a Speed Vac (Eppendorf, Hamburg, Germany) prior to high-performance liquid chromatography (HPLC) analysis. Samples were kept chilled at all times and not exposed to bright light (Shakya and Navarre, 2006).

Tuber extracts were dissolved in 100 μL of ultra-pure water, centrifuged (2 min, 13,000 rpm) and 20 μL of each solution was injected for analysis. Triplicate injections were made for each sample. HPLC with diode-array detection (DAD) analyses were carried out on an Agilent 1100 series system (Agilent Technologies, Palo Alto, CA, USA) using a monolithic Onyx column (50×2 mm, C18) at a flow rate of 0.6 mL/min, and injecting 20 μL of sample. Gradient elution of 0–1 min 100% buffer A (10 mM formic acid, pH 3.5, with ammonium hydroxide), 1–6.7 min 0–30% buffer B (100% methanol with 5 mM ammonium formiate), 6.7–11.7 min 40–70% buffer B, 11.7–14.56 min 70–100% buffer B, 14.56–20 min 100% buffer B. Ultraviolet (UV) detections were at 210, 280 and 254 nm. Structural identification of individual phenolic acids and amino acid in extracts was performed by associating the HPLC peaks of each compound with the corresponding UV and MS^n spectra comparison to those of standards when available. A method of external calibration (six points) was applied. LC–MS/MS analyses were performed on a quadrupole time-of-flight (Q-TOF) Premier instrument (Waters, Milford, MA) coupled to a 2690 Alliance HPLC. Instrument tuning and mass calibration of the spectrometer were performed using α-solanine ($[M+H]^+ = m/z$ 868.51) as standard. Chromatographic separations were conducted using the same conditions described above. Mass analyses were performed in positive mode, using a dependent MS/MS function. The following instrumental parameters were used: source temperature 80 °C, desolvation temperature 180 °C, cone gas flow rate 50 L/h, and desolvation 300 L/h. Capillary voltage was 3 kV, cone voltage 28, extraction cone 5 while fragmentation voltage was 40 V. The full mass scan ranged between m/z 250 and 1500 (Ruberto et al., 2007).

Typical uncertainties associated to the 11 secondary metabolites' data were estimated from RSDs (relative standard deviations) obtained from triplicates of six tubers. The average RSD was rounded to the immediate upper half unit and multiplied by two, obtaining the following relative expanded uncertainties (k=2): 2% (ascorbic acid), 3% (tyrosine), 8% (caffeoyl spermine), 7% (tryptophan), 4% (3-caffeoyl-5-O-feruloyl quinic acid, 3-C,5-FQA), 5% (bis-dihydro caffeoyl spermine), 2% (neochlorogenic acid), 2% (chlorogenic acid), 4% (tris(dihydro caffeoyl) spermidine), 10% (bis(dihydro caffeoyl) spermidine), and 2% (rutin).

2.3.3. DNA fingerprinting

DNA isolation was performed with the foodproof® GMO Sample Preparation Kit (BIOTECON Diagnostics GmbH, Potsdam, Germany) using 200 mg of each lyophilized tuber sample. Simple Sequence Repeat (SSR) analyses were carried out with 12 nuclear microsatellite (SSR) primer pairs. They were chosen from the potato genetic identification (PGI) kit (Ghislain et al., 2004) and selected based on quality criteria, genome coverage, and locus-specific information content. Polymerase chain reactions (PCR) were performed in a 10 μL volume containing 1× reaction buffer with 2.5 mM $MgCl_2$, 0.2 mM of each dNTP (deoxyribonucleotide triphosphate), 25 pM 6-carboxyfluorescein (FAM)-labeled forward SSR primer, 15 pM reverse SSR primer, 1 unit of Taq polymerase, and 40 ng of genomic DNA. PCR was carried out using the following cycling profiles: 4 min at 94 °C; 31 cycles of 45 s at 94 °C, 1 min at annealing temperature (Ta), and 1 min at 72 °C, with a final extension step of 5 min at 72 °C. PCR products were separated on an ABI PRISM® 3100 DNA Analyzer system. Size calibration was performed with the molecular weight ladder GenScan™ 500 ROX™ Size Standard. SSR alleles were detected and scored by using GeneScan® Analysis software as present (1) or absent (0). A similarity matrix was calculated by using the Dice coefficient. The dendrogram was built by applying the UPGMA (Unweighted Pair Group Method with Aritmetic Mean) method. Genetic similarity calculations and dendrogram construction were carried out with the NTSYS software.

2.4. Multivariate analysis

Partial least squares discriminant analysis (PLS-DA) was performed with the SIMCA-P Ver.12 software (Umetrics AB, Umeå, Sweden). Principal Component Analysis (PCA) was used only as an exploratory tool, at the earliest stage of our investigations. PLS-DA was then applied, to maximize the separation among classes, for the production of the projection models discussed in this paper. Prior to these calculations, the datasets were first unit variance scaled and mean-centered (Eriksson et al., 2001). In addition, Rb, Sr, Cd and chlorogenic acid concentration data required log-transformation to normalize the skewed distributions. Diagnostic tools applied for a numerical evaluation of the recognition and prediction abilities of a given model were the Variable Importance in the Projection (VIP), $R^2X_{(cum)}$, $R^2Y_{(cum)}$ and $Q^2_{(cum)}$ parameters. VIP identifies, at values greater than unity, the most influential variables for the model. R^2X and R^2Y parameters indicate the "goodness of fit" for "predictors" and "responses" respectively, and the Q^2 parameter indicates the "goodness of prediction". According to Eriksson et al. (2001), $Q^2 > 0.5$ is regarded as good, $Q^2 > 0.9$ as excellent, and differences between R^2 and Q^2 smaller than 0.2–0.3 are considered as good.

3. Results and discussion

3.1. Secondary metabolites

From a qualitative point of view all extracts showed similar chromatographic profile and were rich of different polyphenolic compounds. Two aromatic amino acids, tyrosine (TYR) and tryptophan (TRP), were identified in our tubers, with TYR being the most abundant in all cultivars (Table 2). Chlorogenic acid (5-*O*-caffeoyl quinic acid, 5-CQA) was the predominant phenolic acid in the analyzed potato cultivars, in accordance with former studies of potato polyphenols (Andre et al., 2007).

An isomer of 5-CQA, neochlorogenic acid (3-*O*-caffeoyl quinic acid, 3-CQA) was identified on the basis of previously published UV and MS mass spectra (Shakya and Navarre, 2006); 3-CQA was present at 15–30% of 5-CQA. A flavonoid, rutin, could be also identified and quantified in our tuber extracts. Quantitative analysis pointed out that the most abundant metabolites in the analyzed tuber were 5-CQA and ascorbic acid. In particular, SP-P-A and EL-P-E showed the highest content of 5-CQA (1225 and 822 mg kg^{-1}, respectively) while all Arinda samples from Sicily had the lowest 5-CQA content. The sample SP-P-A emerged also for its high content in ascorbic acid, having 1660 mg kg^{-1} of this metabolite. The tubers with less ascorbic acid were identified as SG-P-C (378 mg kg^{-1}). Results obtained from the quantitative analysis let us hypothesize the presence of two groups among the cultivars which are distinguishable for the amount of ascorbic acid. Indeed, while Arinda, Spunta and Agria showed a very high relative amount of this metabolite (more than 35%), Sieglinde and Elvira had all a lower ascorbic acid quantity (about 25%). The results showed that all analyzed cultivars are appreciable source of ascorbic acid and polyphenols, which are important for their role in the plants' resistance to diseases, pests and other stresses. These compounds are also considered to have possible protective effects on human health (Friedman, 1997; Thomasset et al., 2007; Walters, 2003).

3.2. DNA fingerprinting

DNA markers are ubiquitous to most of the living organisms, unaffected by environment, and are free of epistatic interactions and

Table 2
Secondary metabolites content (mg kg^{-1}) of dried *Solanum tuberosum* tubers. All uncertainties indicated are expanded uncertainty $U = k \cdot u_c$ where u_c is the combined standard uncertainty. They are given in parentheses and include coverage factor $k = 2$.

Province	Sample ID	Ascorbic acid		TYR		Caffeoyl spermine		TRP		3-C,5-FQA		Bis-dihydro caffeoyl spermine		3-CQA		5-CQA		Tris(dihydro caffeoyl) spermidine		Bis(dihydro caffeoyl) spermidine		Rutin	
Bari	SP-P-I1	1081	(22)	219	(7)	7	(0.6)	18	(1)	38	(2)	13	(1)	62	(1)	445	(9)	7	(0.3)	5	(0.5)	13	(0.3)
	SP-P-I2	1202	(24)	241	(7)	8	(0.6)	28	(2)	71	(3)	17	(1)	87	(2)	590	(12)	8	(0.3)	5	(0.5)	14	(0.3)
	SP-P-I3	1151	(23)	220	(7)	8	(0.6)	23	(2)	53	(2)	10	(1)	113	(2)	577	(12)	8	(0.3)	5	(0.5)	14	(0.3)
Foggia	EL-P-E1	612	(12)	236	(7)	16	(1)	119	(8)	86	(3)	19	(1)	256	(5)	808	(16)	10	(0.4)	4	(0.4)	35	(1)
	EL-P-E2	631	(13)	222	(7)	13	(1)	117	(8)	42	(2)	8	(0.4)	240	(5)	955	(19)	9	(0.4)	7	(0.7)	33	(1)
	EL-P-E3	586	(12)	277	(8)	15	(1)	107	(7)	66	(3)	11	(1)	289	(6)	836	(17)	8	(0.3)	7	(0.7)	40	(1)
	EL-P-E-com	620	(12)	303	(9)	16	(1)	122	(9)	105	(4)	21	(1)	318	(6)	1512	(30)	8	(0.3)	4	(0.4)	53	(1)
Lecce	SP-P-A1	1632	(33)	309	(9)	10	(1)	29	(2)	61	(2)	21	(1)	143	(3)	1229	(25)	8	(0.3)	8	(0.8)	20	(0.4)
	SP-P-A2	1662	(33)	355	(11)	10	(1)	26	(2)	74	(3)	19	(1)	152	(3)	1079	(22)	7	(0.3)	9	(0.9)	19	(0.4)
	SP-P-A3	1689	(34)	416	(12)	10	(1)	30	(2)	72	(3)	26	(1)	189	(4)	1223	(24)	8	(0.3)	12	(1)	23	(0.5)
	SG-P-C1	373	(7)	548	(16)	13	(1)	45	(3)	191	(8)	44	(2)	337	(7)	808	(16)	9	(0.4)	5	(0.5)	18	(0.4)
	SG-P-C2	384	(8)	562	(17)	13	(1)	55	(4)	195	(8)	50	(3)	234	(5)	645	(13)	7	(0.3)	4	(0.4)	16	(0.3)
	SG-P-C3	489	(10)	452	(14)	14	(1)	52	(4)	114	(5)	39	(2)	199	(4)	613	(12)	6	(0.2)	5	(0.5)	15	(0.3)
	SG-P-C-com	367	(7)	408	(12)	7	(0.6)	37	(3)	105	(4)	44	(2)	207	(4)	950	(19)	6	(0.2)	9	(0.8)	23	(0.5)
Siracusa	SP-S-B1	1006	(20)	237	(7)	11	(1)	63	(4)	30	(1)	12	(1)	137	(3)	551	(11)	6	(0.2)	7	(0.7)	18	(0.4)
	SP-S-B2	916	(18)	229	(7)	11	(1)	74	(5)	25	(1)	8	(0.4)	189	(4)	656	(13)	7	(0.3)	6	(0.6)	17	(0.3)
	SP-S-B3	917	(18)	283	(8)	12	(1)	132	(9)	34	(1)	8	(0.4)	245	(5)	761	(15)	7	(0.3)	6	(0.6)	19	(0.4)
	AR-S-A1	846	(17)	214	(6)	18	(1)	36	(3)	55	(2)	11	(1)	146	(3)	464	(9)	7	(0.3)	5	(0.5)	20	(0.4)
	AR-S-A2	687	(14)	251	(8)	15	(1)	56	(4)	39	(2)	8	(0.4)	145	(3)	436	(9)	6	(0.2)	5	(0.5)	20	(0.4)
	AR-S-A3	829	(17)	201	(6)	15	(1)	64	(4)	42	(2)	8	(0.4)	119	(2)	437	(9)	6	(0.2)	4	(0.4)	18	(0.4)
Messina	SG-S-F2	522	(10)	386	(12)	13	(1)	58	(4)	62	(2)	17	(1)	225	(5)	617	(12)	7	(0.3)	4	(0.4)	14	(0.3)
	SG-S-F3	729	(15)	440	(13)	12	(1)	69	(5)	103	(4)	35	(2)	173	(3)	491	(10)	5	(0.2)	4	(0.4)	14	(0.3)
	AR-S-G1	667	(13)	477	(14)	16	(1)	60	(4)	65	(3)	22	(1)	95	(2)	279	(6)	5	(0.2)	7	(0.7)	13	(0.3)
	AR-S-G2	581	(12)	550	(17)	16	(1)	60	(4)	79	(3)	21	(1)	91	(2)	216	(4)	5	(0.2)	7	(0.71)	12	(0.2)
	AR-S-G3	504	(10)	275	(8)	10	(1)	38	(3)	32	(1)	14	(1)	37	(1)	123	(2)	4	(0.2)	5	(0.5)	11	(0.2)
Catania	SP-S-C1	893	(18)	419	(13)	12	(1)	37	(3)	64	(3)	15	(1)	86	(2)	294	(6)	5	(0.2)	6	(0.6)	17	(0.3)
	SP-S-C2	862	(17)	371	(11)	13	(1)	39	(3)	52	(2)	14	(1)	91	(2)	466	(9)	5	(0.2)	7	(0.7)	21	(0.4)
	SP-S-C3	901	(18)	373	(11)	13	(1)	45	(3)	40	(2)	15	(1)	95	(2)	343	(7)	5	(0.2)	8	(0.8)	18	(0.4)
	AR-S-E1	1068	(21)	250	(8)	11	(1)	52	(4)	26	(1)	16	(1)	48	(1)	209	(4)	5	(0.2)	8	(0.8)	13	(0.3)
	AR-S-E2	1041	(21)	232	(7)	11	(1)	49	(3)	24	(1)	18	(1)	36	(1)	183	(4)	4	(0.2)	8	(0.8)	12	(0.2)
	AR-S-E3	959	(19)	304	(9)	12	(1)	47	(3)	37	(1)	24	(1)	67	(1)	315	(6)	5	(0.2)	8	(0.8)	14	(0.3)
Napoli	AGR-C-A1	760	(15)	260	(8)	16	(1)	32	(2)	23	(1)	29	(1)	88	(2)	628	(13)	11	(0.4)	6	(0.6)	25	(1)
	AGR-C-A2	716	(14)	286	(9)	16	(1)	29	(2)	30	(1)	12	(1)	56	(1)	488	(10)	10	(0.4)	3	(0.3)	30	(1)
	AGR-C-A3	665	(13)	279	(8)	14	(1)	32	(2)	37	(1)	9	(0.5)	123	(2)	623	(12)	10	(0.4)	6	(0.6)	30	(1)

TYR = tyrosine; TRP = tryptophan; 3-C,5-FQA = 3-caffeoyl-5-O-feruloyl quinic acid; 3-CQA = neochlorogenic acid; and 5-CQA = chlorogenic acid.

pleiotroic effects. For these reasons they can be analyzed objectively, providing an excellent tool for cultivar identification (Kalia et al., 2011). In this study we were able to detect a total of 36 alleles analyzing 8 microsatellite loci. Only 4 alleles were present in all varieties, while the other 32 showed a varying degree of polymorphism. The presence of private alleles was highlighted in each genotype and allowed unequivocally variety identification (Table 3). In particular, four SSR alleles (132 bp, 137 bp, 141 bp, and 253 bp) appeared only in Sieglinde, whereas two alleles were specific for Arinda (125 bp and 174 bp) and Elvira (250 bp and 292 bp). The 190 bp (STI0012), and 116 bp (STI0032) alleles were detected only in Spunta and Agria, respectively. Overall, the number of detected alleles per genotype varied from 16 (Elvira) to 22 (Agria) (Table 3) revealing a different degree of heterozygosity for each genotype. To obtain an estimate of the degree of differentiation among the groups, the genetic similarities among the five genotypes analyzed were calculated (data not shown). Pairwise similarity showed Sieglinde and Arinda to be the least similar varieties (0.44) and Spunta and Agria the most similar (0.76). To better appreciate the relationships among varieties, a UPGMA dendrogram was built using the similarity coefficients (Fig. 1). Two main groups were distinguished. The first main group (Cluster I) was comprised of Elvira, Arinda, Spunta and Agria, furtherly divided in two subgroups; Sieglinde formed the second main group (Cluster II). This molecular analysis confirmed that SSR fingerprinting is one of the most efficient methods that can clearly distinguish individuals based on differences in DNA electropherograms. One additional advantage of SSR markers is that the small dimension of their target sequences may allow the amplification also when DNA is degraded, e.g. DNA extracted from food (Pasqualone et al., 2010).

3.3. Discrimination of potato samples according to region

The elemental concentrations selected during our investigations on the early potatoes were [Mn], [Cu], [Zn], [Rb], [Sr] and [Cd]. In a previous work, we showed how the relationships that exist between the elemental concentration and Sr isotopic signature markers in potatoes and the geopedological properties of the cultivation areas could be used for authentication purposes (Zampella et al., 2011). It was found possible to "assign 26 potatoes (including 2 already commercialized samples) to their respective 8 sites of productions". The other 12 samples from 4 sites of production with similar characteristics in terms of geological substrate and soil properties were grouped together, underlying the need of more discriminative parameters to allow the differentiation between sites.

In this study, secondary metabolite data, alone and in combination with elemental concentration and Sr isotopic signature results described previously (Zampella et al., 2011), were tested as descriptor variables for the discrimination of early potatoes by PLS-DA according to the region of provenance.

PLS-DA was first applied to all 11 secondary metabolites as descriptor variables, and the model was then progressively refined to obtain the best recognition and prediction abilities according to the region where the potatoes were cultivated. Eventually, the optimum model for the discrimination of potatoes grown in the three regions was based on seven metabolites (caffeoyl spermina, 3-C,5-FQA, bis-dihydrocaffeoyil spermina, 3-CQA, 5-CQA, tris(dihydro caffeoyl) spermidine, rutin) and had five PCs (Fig. 2a, Table 4). Apulia and Sicily were rather well sorted according to PC 1, whereas PC 2 essentially separated Campania from the other regions (Fig. 2a). The most important role for the discrimination between potatoes from Apulia, Sicily and Campania was played by tris(dihydro caffeoyl) spermidine, 3-CQA and 5-CQA. Potatoes grown in Campania were characterized by higher amounts of tris(dihydro caffeoyl) spermidine and potatoes cultivated in Apulia by rather higher amounts of 3-CQA and 5-CQA than potatoes cultivated in the other studied regions. The combination of secondary metabolites with elemental concentrations and Sr isotope ratio data further improved the discrimination of potatoes according to the region of cultivation (Fig. 2b, Table 4). The PLS-DA model obtained from all data (18 descriptor variables: 11 secondary metabolites, 6 element concentrations and Sr isotope ratio) had six principal components. The PLS-DA score-plot (PC 1 vs. PC 2) showed clear and well separated clusters according to the region of provenance of potatoes (Fig. 2b). As indicated by the Q^2 parameter (Table 4), the prediction ability of this model was good and higher than the one obtained with only secondary metabolites. The most relevant variables (indicated by the distribution of VIP parameters) for discrimination according to the region of origin were, in decreasing order, tris(dihydro caffeoyl) spermidine, 5-CQA, [Sr], 3-C,5-FQA, rutin, [Mn] and [Rb]. Potatoes from Campania were characterized by higher amounts of tris(dihydro caffeoyl) spermidine, rather higher amounts of Rb and quite lower amounts of Mn than potatoes from the other two regions. Potatoes from Apulia showed rather higher amounts of 5-CQA, 3-C,5-FQA and 3-CQA than potatoes from the other two regions. Potatoes from Sicily had rather higher amounts of Sr and quite lower amounts of rutin than potatoes from the other two regions. It is well accepted that the elemental composition of plants is the reflection of the bio-available nutrients present in the underlying soils from which they were cultivated and, thus, that it may provide unique markers of the geographical origin (Kelly et al., 2005). Concerning the secondary metabolites, although the role of polyamines in tubers is still unclear, some authors reported that they may be implicated in environmental stress resistance (Walters, 2003). Tuber metabolite content may be influenced by agronomic practices. 5-CQA and ascorbic acid synthesis, for example, has been reported to increase with increasing levels of nitrogen fertilizer (Rosenthal and Jansky, 2008). In addition, an inverse relationship has been reported between soil potassium levels and chlorogenic acid levels in potato tubers (Mondy et al., 1967). Other factors may be important in determining metabolite content. Hajslova et al. (2005) demonstrated that geographical variations as well as variety are equally or more important variants determining the secondary

Table 3
Detected SSR alleles (bp) obtained using eight microsatellite markers on five potato varieties. In bold the private alleles scored for each locus and cultivar.

Variety	Alleles (bp) per SSR locus								
	STI0032	STM1053	STM5127	STM5114	STG0001	STI0012	STM1052	STM1106	Total alleles per variety
Elvira	107, 110,	172	241, **250**	**292**, 298	120, 135, 139	168, 171, 184	210, 228	157	16
Sieglinde	107	172	241, **253**, 271, 274	286, 289, 298	127, **132**, 143	168, 171, 184	210, 228	**137**, **141**	19
Spunta	110, 119, 122	122, 169	241, 244, 274	289, 298	127, 135, 143	165, 184, **190**	210, 219, 228	157	20
Arinda	107, 119, 122	172	241, 244	286, 289, 298	120, **125**, 135	168, **174**	210, 219	157	17
Agria	107, **116**, 119, 122	169, 172	241, 244	286, 289, 298	127, 139, 143	165, 168, 171, 184	210, 219, 228	157	22
Total alleles per locus	5	3	6	4	7	6	3	3	

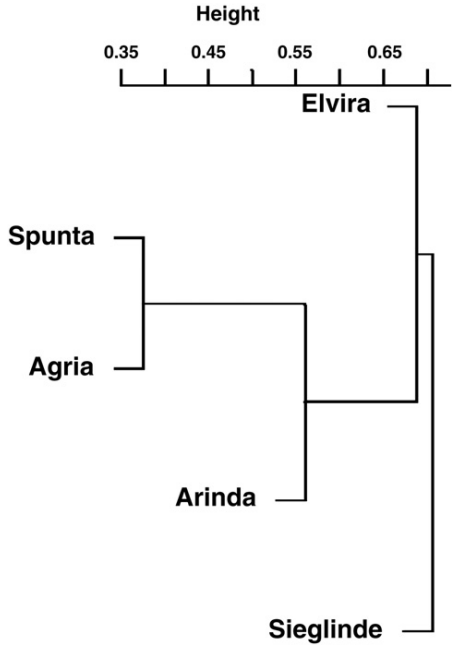

Fig. 1. UPGMA phenogram showing the relationship among varieties estimated through SSR markers.

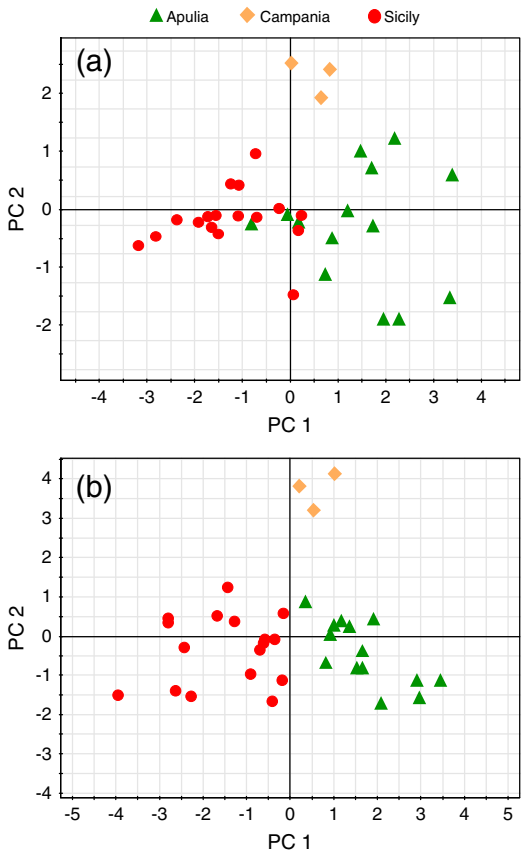

Fig. 2. PLS-DA (partial least squares discriminant analysis) score plots for (a) 7 secondary metabolites as variables: (caffeoyl spermina, 3-C,5-FQA, bisdihydrocaffeoyil spermina, 3-CQA, 5-CQA, tris(dihydro caffeoyl) spermidine, rutin); (b) 18 variables (11 secondary metabolites + 6 element concentrations + Sr isotope ratio) of 34 potato samples. Samples were sorted according to region of provenance.

metabolite content of potato tubers than the farming system. Our results suggested that 5-CQA content was lower in soils with acid pH value. In addition, the higher quantity of 3-C,5-FQA, bis(dihydro caffeoyl) spermidine, and TYR content in Sieglinde tubers compared to the other cultivars confirmed the genetic differences resulting from molecular analysis with SSR markers (Fig. 1).

The combination of data from markers strictly related to the geopedological properties of the cultivation sites (elemental concentration and Sr isotopic signature) and from markers more generally related to environmental conditions, including climate and agricultural practices (secondary metabolites), allowed us to well discriminate all the potatoes according to the respective regions considered (Apulia, Campania, Sicily). The discrimination and prediction abilities are satisfactory although reasons explaining how these markers relate to each other remain to be clarified. This approach is different than the previous one (Zampella et al., 2011), where discrimination was based on relationships between geopedological markers in the plant and some selected properties of the associated soils. It had allowed a good recognition rate down to production site levels, irrespective of the regions concerned. With these two studies we thus have established complementary ways of working toward the authentication of origin of vegetables grown in a relatively small size region with little differences in terms of climatic and meteorological characteristics.

3.4. Validation of the statistical models

The appropriate number of principal components in PLS-DA models was determined by cross-validation (Eriksson et al., 2001). Response permutation tests (Eriksson et al., 2001; Van Der Voet, 1994) (20 permutations) indicated that the estimated predictive power of all PLS-DA models was significant. External validation of both PLS-DA models discussed previously (on data from 7 metabolites and from all 18 variables) was successful. It was run by computing predictions for an independent set of data (from 7 samples taken at random in each of the 7 provinces) not considered for the model calibration. Two new models sorted by region were generated with the dataset from the remaining 27 samples (values in italics in Table 4). There was a slight degradation of $R^2X_{(cum)}$, $R^2Y_{(cum)}$ and $Q^2_{(cum)}$ values (particularly for the model based on data from 18 variables) although they remained higher than 0.5, indicating goodness of fit and prediction. Besides, both new models were able to assign all samples of the independent set of observations to the right region. The test we run on the robustness of our PLS-DA models to the uncertainty of measurement results was successful too. PLS-DA-based models assume exactness of the input data, while uncertainties associated to these data are not necessarily negligible. Following the approach described recently elsewhere (Goitom Asfaha et al., 2011), we added to all original data the product of their expanded uncertainty (k = 2) by a randomly generated number ranging from −1 to +1. A new series of PLS-DA models was produced from this simulated dataset and the results (values between brackets in Table 4) were compared with those obtained from the original dataset and discussed above. We

Table 4
Main parameters estimated for the PLS-DA models used for the discrimination of potato samples according to the region of provenance ($n_{(training\ set)}$ is the number of samples in the training set; PCs = number of principal components; SMs = secondary metabolites; E = elemental concentrations; and SrIR = Sr isotope ratio). The parameters of models testing the robustness to measurement uncertainties are in brackets; the parameters of models used for the external prediction test are in italics.

	7 Description variables (7 SMs)			18 description variables (11 SMs + 6 E + SrIR)		
$n_{(training\ set)}$	34	(34)	*27*	34	(34)	*27*
PCs	5	(5)	*4*	6	(6)	*3*
$R^2X_{(cum)}$	0.95	(0.94)	*0.94*	0.80	(0.81)	*0.52*
$R^2Y_{(cum)}$	0.79	(0.78)	*0.74*	0.91	(0.90)	*0.74*
$Q^2_{(cum)}$	0.66	(0.59)	*0.51*	0.73	(0.74)	*0.56*
$R^2Y_{(cum)} - Q^2_{(cum)}$	0.13	(0.19)	*0.23*	0.18	(0.16)	*0.18*

observed that the models were globally robust to the fluctuations imposed to data. The decreases of $R^2X_{(cum)}$, $R^2Y_{(cum)}$ and $Q^2_{(cum)}$ values were at maximum of 0.01, 0.01 and 0.07 respectively. All $R^2X_{(cum)}$, $R^2Y_{(cum)}$ and $Q^2_{(cum)}$ values remained higher than 0.5, indicating goodness of fit and prediction. An external prediction test was also applied on this simulated dataset, after taking out 7 samples as described previously. For both models all samples of the independent set of observations could be assigned to the right region (in the case of the model based on 7 metabolites the sample from Afragola was assigned to both Campania–correct- and Apulia–wrong-regions). These simulations are a useful indicator of the potential sensitivity of these multivariate models to the quality of the experimental input data and may provide an indication of the maximum uncertainty tolerable for the models proposed.

4. Conclusions

The combination of secondary metabolites, element concentrations and Sr isotope ratio allowed us to discriminate 34 early potatoes (including 2 already commercialized samples) according to their respective regions of origin, while DNA fingerprints allowed us to identify the cultivars. The prediction abilities of the PLS-DA models for the region of origin of potatoes were excellent, despite the relatively small size and the little differences in terms of climatic and meteorological characteristics. These results highlighted the importance of secondary metabolites not only from a health standpoint, but also in terms of possible indicators of geographical provenance of potato tubers. Their use improved the discrimination power of trace elements and Sr isotopic signature reported by other Authors to characterize potatoes from Italy (Di Giacomo et al., 2007; Zampella et al., 2011) and USA (Anderson et al., 1999). The validity of the multidisciplinary approach used in this research deserves future research involving a larger sample size and different years of production.

Acknowledgments

Contribution from the Dipartimento di Scienze del Suolo, della Pianta, dell'Ambiente e delle Produzioni Animali (no. 271) is gratefully acknowledged. Authors thank Ministero delle Politiche Agricole e Forestali (MiPAAF) for funding the TIPIPAPA project ("tipicizzazione e caratterizzazione di varietà precoci di patata con l'impiego di tecniche molecolari e spettroscopiche"), Mr. R. Garramone for technical help and all the potato farmers for collaboration.

References

Anderson, K.A., Magnuson, B.A., Tschirgi, M.L., Smith, B., 1999. Determining the geographic origin of potatoes with trace metal analysis using statistical and neural network classifiers. Journal of Agricultural and Food Chemistry 47, 568–575.

Andre, C.M., Oufir, M., Guignard, C., Hoffmann, L., Hausman, J., Evers, V.D., Larondelle, Y., 2007. Antioxidant profiling of native Andean potato tubers (*Solanum tuberosum* L) reveals cultivars with high levels of b-carotene, a-tocoferol, chlorogenic acid, and petanin. Journal of Agricultural and Food Chemistry 55, 10839–10849.

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 22, 549–554.

Di Giacomo, F., Del Signore, A., Giaccio, M., 2007. Determining the geographic origin of potatoes using mineral and trace element content. Journal of Agricultural and Food Chemistry 55, 860–866.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 2001. Multi-and Megavariate Data Analysis: Principles and Applications. Umetrics AB, Umeå.

Friedman, M., 1997. Chemistry, biochemistry, and dietary role of potato polyphenols. A review. Journal of Agricultural and Food Chemistry 45, 1523–1540.

Ghislain, M., Spooner, D.M., Rodríguez, F., Villamón, F., Núñez, J., Vásquez, C., Waugh, R., Bonierbale, M., 2004. Selection of highly informative and user-friendly microsatellites (SSRs) for genotyping of cultivated potato. Theoretical and Applied Genetics 108, 881–890.

Goitom Asfaha, D., Quétel, C.R., Thomas, F., Horacek, M., Wimmer, B., Heiss, G., Dekant, C., Deters-Itzelsberger, P., Hoelzl, S., Rummel, S., Brach-Papa, C., Van Bocxstaele, M., Jamin, E., Baxter, M., Heinrich, K., Kelly, S., Bertoldi, D., Bontempo, L., Camin, F., Larcher, R., Perini, M., Rossmann, A., Schellenberg, A., Schlicht, C., Froeschl, H., Hoogewerff, J., Ueckermann, H., 2011. Combining isotopic signatures of n(^{87}Sr)/n(^{86}Sr) and light stable elements (C, N, O, S) with multielemental profiling for the authentication of provenance of European cereal samples. Journal of Cereal Science 53, 170–177.

Hajslova, J., Schulzova, V., Slanina, P., Janne, K., Hellenas, K.E., Andersson, C.H., 2005. Quality of organically and conventionally grown potatoes: four-year study of micronutrients, metals, secondary metabolites, enzyme browning, and organoleptic properties. Food Additives and Contaminants 22, 514–534.

Kalia, R.K., Rai, M.K., Kalia, S., Singh, R., Dhawan, A.K., 2011. Microsatellite markers: an overview of the recent progress in plants. Euphytica 177, 309–334.

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 and Technology 16, 555–567.

Mondy, N.I., Owens Mobley, E., Bond Gedde-Dahl, S., 1967. Influence of potassium fertilization on enzymatic activity, phenolic content and discoloration of potatoes. Journal of Food Science 32, 378–381.

Pasqualone, A., Alba, A., Mangini, G., Blanco, A., Montemurro, C., 2010. Durum wheat cultivar traceability in PDO Altamura bread by analysis of DNA microsatellites. European Food Research and Technology 230, 723–729.

Peres, B., Barlet, N., Loiseau, G., Montet, D., 2007. Review of the current methods of analytical traceability allowing determination of the origin of foodstuffs. Food Control 18, 228–235.

Rocha, S.M., Coelho, E., Zrostlıkova, J., Delgadillo, I., Coimbra, M.A., 2007. Comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry of monoterpenoids as a powerful tool for grape origin traceability. Journal of Chromatography A 1161, 292–299.

Rosenthal, S., Jansky, S., 2008. Effect of production site and storage on antioxidant levels in specialty potato (*Solanum tuberosum* L.) tubers. Journal of the Science of Food and Agriculture 88, 2087–2092.

Ruberto, G., Renda, A., Daquino, C., Amico, V., Spatafora, C., Tringali, C., De Tommasi, N., 2007. Polyphenol constituents and antioxidant activity of grape pomace extracts from five Sicilian red grape cultivars. Food Chemistry 100, 203–210.

Shakya, R., Navarre, D.A., 2006. Rapid screening of ascorbic acid, glycoalkaloids, and phenolics in potato using high-performance liquid chromatography. Journal of Agricultural and Food Chemistry 54, 5253–5260.

Thomasset, S.C., Berry, D.P., Garcea, G., Marczylo, T., Steward, W.P., Gescher, A.J., 2007. Dietary polyphenolic phytochemicals promising cancer chemopreventive agents in humans? A review of their clinical properties. International Journal of Cancer 120, 451–588.

Van Der Voet, H., 1994. Comparing the predictive accuracy of models using a simple randomization test. Chemometrics and Intelligent Laboratory Systems 25, 313.

Walters, D.R., 2003. Polyamines and plant disease. Phytochemistry 64, 97–107.

Zampella, M., Quétel, C., Paredes, E., Goitom Asfaha, D., Vingiani, S., Adamo, P., 2011. Soil properties, Sr isotopic signature and multi-element profiles to authenticate the origin of vegetables from small scale regions: illustration with early potatoes from south Italy. Rapid Communications in Mass Spectrometry 25, 2721–2731.

Zhang, Jia, Zhang, Xiaoshuan, Dediu, Lorena, Victor, Cristea, 2011. Review of the current application of fingerprinting allowing detection of food adulteration and fraud in China. Food Control 22, 1126–1135.