Barcoding assessment of the Afghan *Citrus* population

Introduction. The identification and characterization of the local population of fruit trees could be of basic importance for developing countries. Usually, identification of the cultivated species relies on the use of phenotypic descriptors (UPOV). Phenotypic descriptors often are subjected to the observer's opinion and the environment can deeply affect the behavior of the plant determining errors in the species attribution. A more reliable system to identify species is based on the genetic analysis and particularly on the barcoding procedure which consists in the comparison of the sequences of highly conserved fragments located in the ribosomial or plastid DNA. Being highly conserved, such fragments accumulate mutations slowly. Furthermore many databases (PubMed, GenBank, OMIM) collect and store millions of sequences, which can be compared by means of specific software (BLAST) with unknown sequences. When properly queried a DNA database provides, along with the alignment, a similarity index which can be useful for identifying species or even sub-species, depending on the proximity of the taxonomic entities. Being highly conserved it has been possible to design universal primers which can be used with a great deal of species, allowing the comparison with the sequences stored in the databases.

Plant material. Afghan citrus seeds have been collected in three areas where citrus is intensively cultivated (Langman, Kunar e Nangarhar). We use seeds because they can be easily transported and many Citrus seeds originate clones due to the adventitious embryony. Furthermore most sequences for barcoding, being located on the plastid or ribosomial DNA, are inherited only from the maternal line, therefore germinating seeds produce plants whose organelle genomes are not affected by the pollen donor plant, but are expected to be the same of the seed donor plant. In each area some plants have been selected and batch of seeds have been collected from single fruit. Each batch has been labeled and delivered to the Plant Pathology Department of the University of Bologna. The seeds have been germinated in pots and for each batch a single seedling has been selected (tab.1). Total DNA has been extracted from the selected plants.

Table 1: list of the samples coming from Afghanistan. Origin and name is also reported.

Sample	Farmer Name	Province	District	Village	Species	Variety	Origin of the variety
1afganistan	Kabir-Rangbar	Laghman	Mehtarlam	Chardehi	Sour Orange	Local	Laghman
2afganistan	Kabir-Rangbar	Laghman	Mehtarlam	Chardehi	Sour Orange	Local	Laghman
5afganistan	Kabir-Rangbar	Laghman	Mehtarlam	Chardehi	Sour Orange	Local	Laghman
7afganistan	Haji-Amir-M	Kunar	Asadabad	Landi Tesha	Sour Orange	Local	Kunar
8afganistan	Haji-Amir-M	Kunar	Asadabad	Landi Tesha	Sour Orange	Local	Kunar
9afganistan	Haji-Amir-M	Kunar	Asadabad	Landi Tesha	Sour Orange	Local	Kunar
11afganistan	Haji-Amir-M	Kunar	Asadabad	Landi Tesha	Sour Orange	Local	Kunar
18afganistan	Haji-Hafez	Nangarhar	Surkhroad	Naghrak	Sour Orange	Local	Nangarhar
19afganistan	Haji-Hafez	Nangarhar	Surkhroad	Naghrak	Sour Orange	Local	Nangarhar
21afganistan	Haji-Ab-Malik	Nangarhar	Surkhroad	Sabzabad	Sour Orange	Local	Nangarhar
22afganistan	Haji-Ab-Malik	Nangarhar	Surkhroad	Sabzabad	Sour Orange	Local	Nangarhar
16afganistan	Mumtaz	Nangarhar	Bishud	PHDC	Lemon	Volkamar	South Africa

To verify that samples belong to a particular species, we have analysed also samples of Citrus aurantium, C. sinensis and many other Citrus whose origin is absolutely certain as control. Many samples of C. aurantium coming from different part of Tuscany have been analysed (tab 2).

Table 2 - Citrus coming from various public and private collections in Tuscany and used as control

n°	Species and variety	Place	Origin
1	Citrus Sinensis (Washington Navel)	Pescia (Italy)	Oscar Tintori
2	Citrus Sinensis (Ovale o Calabrese)	Pescia (Italy)	Oscar Tintori
3	Citrus Sinensis (Tarocco)	Pescia (Italy)	Oscar Tintori
4	Citrus Aurantium	Lucca (Italy)	Orto Botanico
5	Citrus Aurantium	Firenze (Italy)	Istituto Agronomico Oltremare
6	Citrus Aurantium (Foetifera)	Pescia (Italy)	Oscar Tintori
7	Citrus Aurantium (tipo)	Pescia (Italy)	Oscar Tintori
8	Citrus Aurantium (dolce del Gargano)	Pescia (Italy)	Oscar Tintori
9	Citrus Aurantifolia (Filippine Red lime)	Pescia (Italy)	Oscar Tintori
10	Citrus Aurantifolia (limetta)	Pescia (Italy)	Oscar Tintori
11	Citrus Aurantifolia (Messicani)	Pescia (Italy)	Oscar Tintori
12	Citrus Aurantium	Firenze (Italy)	Orto Botanico
13	Citrus Limon	Firenze (Italy)	Orto Botanico
14	Citrus Mitis	Firenze (Italy)	Orto Botanico
15	Citrus Histrix	Firenze (Italy)	Orto Botanico
16	Citrus Decumana	Firenze (Italy)	Orto Botanico
17	Citrus Grandis	Firenze (Italy)	Orto Botanico
18	Citrus Reticulata	Firenze (Italy)	Orto Botanico
19	Citrus Lumia	Firenze (Italy)	Orto Botanico
20	Citrus Medica	Firenze (Italy)	Orto Botanico

DNA extraction. DNA has been extracted from 50-70 mg fresh leaf material following the guidelines of the DNA invisorb DNA extraction kit producer (Stratec Italy). DNA has been electrophoresed on an agarose gel to validate quality and quantity (Fig.1).

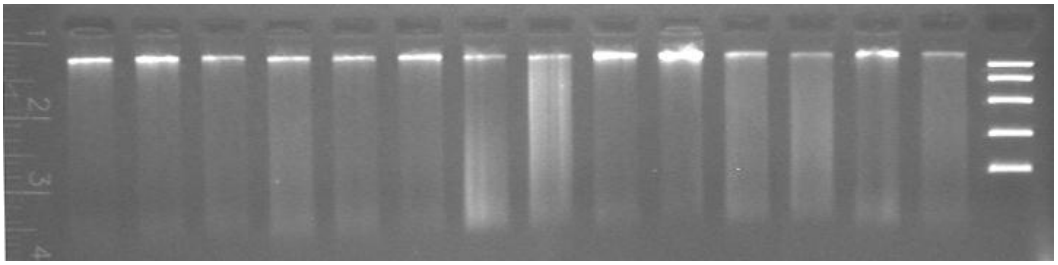

Figure 1. DNA extracted from Citrus samples after electrophoresing on an agarose gel.

PCR amplification and primers. Universal primers used in the present study have been retrieved in literature (Tab.3).

Even though the primers are “universal”and are expected to work properly with all higher plants, sometimes mutations arise in the sequence compromising the primers annealing during PCR and the efficiency of the reaction. The primers have therefore been tested by PCR amplifying and sequencing of the amplicons using some samples as template (Fig.2).

Table 3. Universal primer sequence for barcoding

Name	***5'□3' primer sequence***
matK f	CGTACAGTACTTTTGTGTTTACGAG
matK r	ACCCAGTCCATCTGGAAATCTTGGTTC
Rbcl_FW	ATGTCACCACAAACAGAGACTAAAGC
Rbcl_R634	GAAACGGTCCCTCCAACGCAT
Rbcl_R724	TCGCATGTCCCTGCAGTAGC
ITS1_5F_746	GGAAGTAAAAGTCGTAACAAGG
ITS1_4R_746	TCCTCCGCTTATTGATATGC
ITS2_S2_F497	ATGCGATACTTGGTGTGAAT
ITS2_S3R_497	GACGCTTCTCCAGACTACAAT
psbA_Fw	CGCGCATGGTGGATTCACAATCC
psbA_Rev	GTTATGCATGAACGTAATGCTC
*matK1F_plos**	ACCGTATCGCACTATGTATC
*matK1R_plos**	GAACTAGTCGGATGGAGTAG
*matK2F_plos**	ACGGTTCTTTCTCCACGAGT
*matK3F_plos**	GGTCCGATTTCTCTGATTCT
*matK2R_plos**	AGAATCAGAGAAATCGGACC
*matK3R_plos**	ACTCGTGGAGAAAGAACCGT

*Phylogenetic Relationships of Citrus and Its Relatives Based on matK Gene Sequences t.Penjor, Plos One, 2013

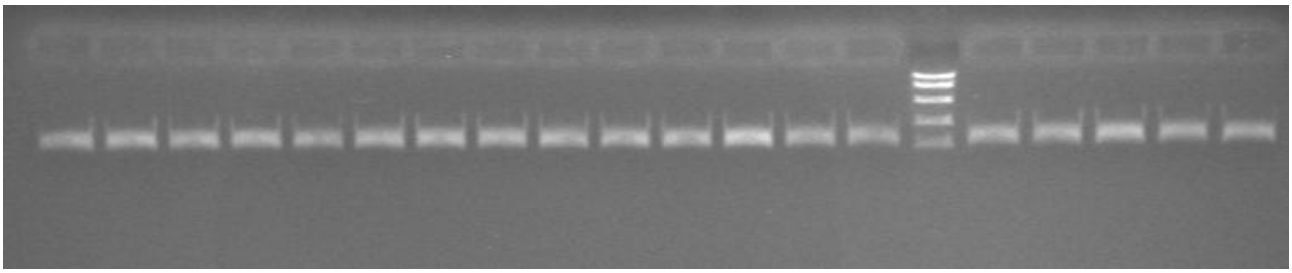

Figure 2. PCR amplification of some samples in the presence of the universal primers designed on the sequence of the photosystem II binding protein (psbA) found in literature.

The primers constructed on the sequence of the internal transcribed sequence 1 (ITS1) and 2 (ITS2) resulted in a faint amplification, while primers designed on the sequence of the maturase K (matK) and RuBisCo large chain unit (rbcl) genes have given very good amplification which have been sequenced. Unfortunately such sequences did not show any discriminatory ability being completely overlapping for most analysed samples. On the contrary, the primers of the photosystem II binding protein (psbA) gene showed either good amplification and sequencing results either a good discriminatory capacity. In figure 3 we show a robust single nucleotide polymorphism (sNP) discriminating C. aurantium from C. sinensis in all analysed samples.

After having amplified and sequenced PCR fragments of the psbA gene of all the collected samples we proceeded to perform a phylogenetic analysis. We have used the Mega6 software package which compares the sequence calculating a similarity matrix, transforms similarity coefficients into distances and makes a clustering using the Pair Group Method with Arithmetic mean (UPGMA) algorithm. The final output is represented by a dendrogram which clusters the samples (fig.4).

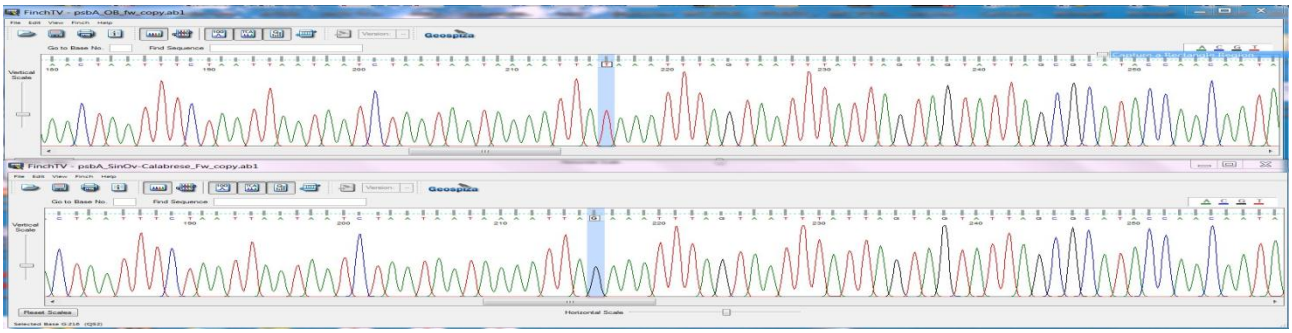

Figure 3. Electropherograms of a PsbA gene fragment showing a point mutation discriminating C. aurantium from C. sinensis samples.

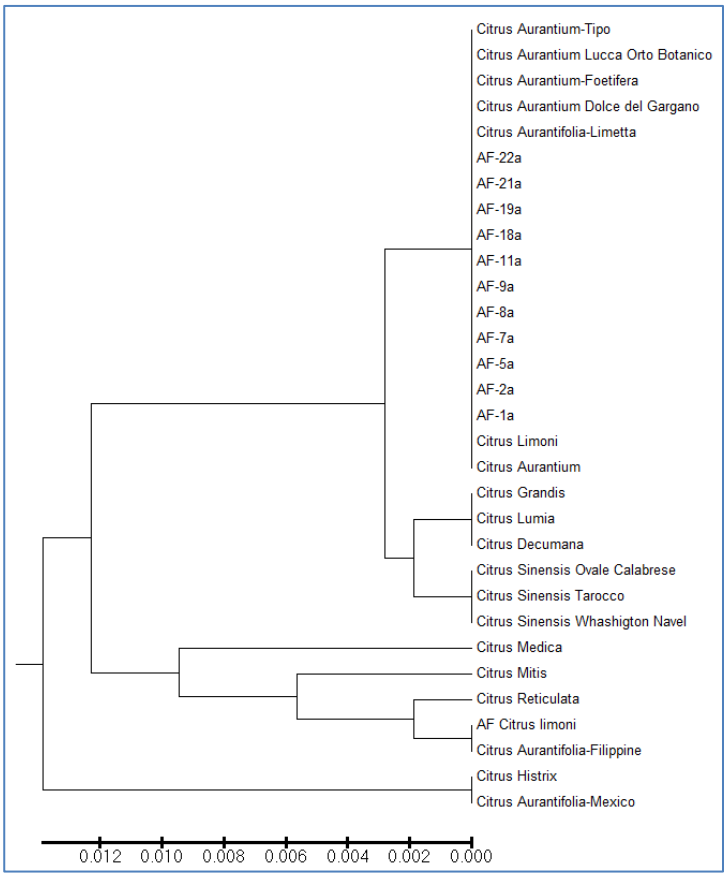

Figure 4. UPGMA dendrogram representing the genetic distances among the analysed samples

Results and conclusions. Before starting the present work we have interrogated the NCBI database in order to download the C. aurantium barcoding sequences. As the available sequences are very limited and showed discrepancies, casting doubts about the reliability of the data, we decided to provide references by analyzing accessions whose origin was certain. Therefore we have sampled C. aurantium along with other Citrus species from private and public collections (Vivai Oscar Tintori, Orto botanico of Florence, Istituto Agronomico per l'Oltremare) in order to compare samples coming from Afghanistan. Despite the limited region of the genome analysed containing, as expected, a low number of mutations, all the samples of C. aurantium, independently from their origin, have been clustered with a reasonable certainty. Also the C. sinensis samples cluster together while C. aurantifolia and C. limon need to be better characterized with additional sequence analysis.
According to the results of our analysis we can state with reasonable certainty that the Afghan citrus samples belong to the C. aurantium species. The discrepancies detected by the comparison with the data available in the database evidence the opportunity to introduce samples of certain origin in barcoding analysis in order to avoid biases due to errors in the sequences submitted to the database.