NUCLEAR AND MATRNAL PHYLOGENY WITHIN *CITRUS* AND FOUR RELATED GENERA BASED ON NUCLEAR GENES SEQUENCE SNPs AND MITOCHONDRIAL InDels

Garcia-Lor, Andres [1]; Curk, Franck [3,1]; Luro, François [3]; Navarro, Luis [1] and Ollitrault, Patrick [2,1]

1: Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada (Valencia), Spain
2: Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Montpellier, France
3: Unité de recherche GEQA, Institut National de la Recherche Agronomique (INRA), 20230 San Giuliano, France
Contact email: angarcia@ivia.es

Current Opinion Conferences — Plant Genome Evolution
Amsterdam, The Netherlands · 4-6 September, 2011
A *Current Opinion* Conference

INTRODUCTION

Despite considerable morphological differentiation *Citrus*, *Fortunella*, *Poncirus*, *Microcitrus* and *Eremocitrus* genera are sexually compatible. Species of these genera are mainly diploid (2n=18). The origin of cultivated *Citrus* from four basic taxa (*C. maxima, C. medica, C. reticulata* and wild citrus) is now well documented, but their phylogenetic relationships with *Citrus* wild species and related genera still unclear. In the present work we analyse their nuclear and maternal phylogeny by using respectively SNPs on gene sequences and mitochondrial Insertion-Delections (InDels).

MATERIAL AND METHODS

The varietal sample set was composed of 45 genotypes: 7 *C. reticulata*, 5 *C. maxima*, 5 *C. medica*, 5 *Fortunella*, 4 wild citrus "papeda" (*C. micrantha*, *C. hystrix*, *C. ichangensis*, *C. macroptera*), 3 *Poncirus*, 2 *Microcitrus*, 1 *Eremocitrus,* cultivated secondary species (2 *C. sinensis* -sweet orange-, 2 *C. aurantium* -sour orange-, 1 *C. paradisi* -grapefruit-, 1 *C. limon* -lemon-, 1 *C. aurantifolia* –mexican lime-, 2 *C. Clementina* –clementines-), recent hybrids from the 20th century breeding programs (1 *C.reticulata x C.sinensis* –tangor-, 1 *C.paradisi x C.tangerina* – tangelo-) and the haploid clementine. *Severinia buxifolia* was used as outgroup. PCR amplified fragments (190 to 941 bp) from 27 genes which are involved in primary and secondary metabolite biosynthesis pathways that determine the quality of citrus fruit (sugars, acids, flavonoids and carotenoids) and some of them involved in abiotic stresses response were directly sequenced by Sanger. The sequenced genes were blasted on the reference citrus genome sequence. They display a good dispertion on all chromosomes. Excepted PIP1 and NCED, which two locations were found for each gene.Therefore they will provide a good representation of the whole genome diversity and an important source of data for population genetics and phylogenomics. SNPs were mined using BioEdit and SeqMan softwares and phylogenetic analysis done in http://phylemon.bioinfo.cipf.es with different approaches (Phylip (v. 3.68), Phyml Best AIC Tree (v. 1.02b), PhyML (v. 3.00). For maternal phylogeny, 4 mitochondrial InDel markers have been used.

GENE POLYMORPHISM

Code	Gene	Seq	SC	SNC	SNPc	Freq	SNPnc	Freq	InDelc	Freq	InDelnc	Freq
CHI	Chalcone isomerase	652	206	446	11	53.40	76	170.40	0	0	8	17.94
CHS	Chalcone synthase	574	574	0	20	35.40	-	-	0	0	-	-
FLS	Flavonol synthase	473	419	54	38	90.69	6	111.11	0	0	3	55.56
F3'H	Flavonoid 3'-hydroxylase	783	569	214	40	70.30	20	93.46	0	0	3	14.02
DFR	Dihydroflavonol 4-reductase	421	171	250	9	52.63	31	124	0	0	3	12.00
EMA	Enzyme malique	428	131	297	7	53.44	26	87.54	1	7.63	4	13.47
MDH	Malate dehydrogenase	712	712	0	28	39.33	-	-	0	0	-	-
ACO	Aconitase	695	250	445	5	20.00	40	89.89	0	0	2	4.49
TRPA	Vacuolar citrate/H+ symporter	795	657	138	40	60.88	16	115.94	0	0	1	7.25
INVA	Acid invertase	908	515	393	41	79.61	41	104.33	0	0	1	2.54
PEPC	Phosphoenolpyruvate carboxylase	694	61	633	2	32.79	50	78.99	0	0	4	6.32
PKF	Phosphofructokinase	775	406	369	16	39.41	32	86.72	0	0	3	8.13
DXS	1-deoxyxylulose 5-phosphate synthase	722	327	395	13	39.76	37	93.67	0	0	3	7.59
PSY	Phytoene synthase	606	97	509	5	51.55	40	78.59	0	0	2	3.93
HYB	β-Carotene hydroxylase	680	379	301	19	50.13	28	93.02	1	2.64	2	6.64
LCY2	Lycopene β-cyclase 2	738	738	0	65	88.08	-	-	5	6.77	-	-
LCYB	Lycopene β-cyclase b	941	941	0	37	39.32	-	-	0	0	-	-
NCED3	9-cis-epoxy hydroxy carotenoid dyoxygenase 3	560	560	0	23	41.07	-	-	0	0	-	-
AOC	Ascorbate oxydase	675	675	0	38	56.30	-	-	0	0	-	-
ATMR	ABCC-type ABC transporter	774	363	411	17	46.83	27	65.69	0	0	1	2.43
CCC1	Cation chloride cotransporter	762	762	0	34	44.62	-	-	0	0	-	-
HKT1	High-affinity K+ Transporter 1	238	116	122	9	77.59	9	73.77	0	0	1	8.20
LAPX	Ascorbate peroxidase	282	282	0	20	70.92	-	-	0	0	-	-
NADK2	NADH kinase	339	65	274	3	46.15	25	91.24	0	0	1	3.65
PIP1	Aquaporin PIP1A	190	103	87	5	48.54	21	241.38	0	0	0	0.00
SOS1	Salt Overly Sensitive 1	495	358	137	23	64.25	12	87.59	0	0	1	7.30
TSC	Tréhalose-6-Phosphate synthase	335	136	199	7	51.47	18	90.45	0	0	0	0.00
Total		16238	10564	5674	575		555		7		43	
Mean		562	391.26	210.15	21.30	53.50	29.21	104.09	0.26	5.68	2.17	10.67

(Seq) Sequence size; SC (Sequence Coding region); SNC (Sequence Non-Coding region); SNPc (SNPs Coding region); Freq (Frequency, X/Kb); SNPnc (SNPs Non-Coding region); InDelc (InDel Coding region); InDelnc (InDel Non-Coding region).

A total of 16.238 kb were amplified by PCR from 27 genes. The genes with the lowest frequency of SNPc and SNPnc were Aconitase (20 SNPs/Kb) and ATMR (65.69 SNPs/Kb) respectively and the highest FLS (90.69 SNPs/Kb) and PIP1 (241.38 SNPs/Kb). At InDelc and InDelnc the lowest values were 0 for some genes and the highest were EMA (7.63 InDels/Kb) and FLS (55.56 InDels/Kb) respectively. The average rate of SNPs/Kb and was 53.50 in coding region and 104.09 in non-coding region and the rate of InDel/Kb was 5.68 in coding region and 10.67 in non-coding region.

WITHIN AND BETWEEN SPECIES NUCLEAR DIVERSITY

Average SNP rate within and between *Citrus* species

SNP/Kb	*C. reticulata*	*C. maxima*	*C. medica*	Wild citrus
C. reticulata	1.54			
C. maxima	10.16	0.65		
C. medica	13.92	11.13	1.50	
Wild citrus	8.56	9.66	14.43	3.37

Within species SNP rates varied from 0.65 for *C. maxima* to 3.37 for wild citrus. Interspecific rates were much higher and varied from 8.56 between *C. reticulata* and wild citrus to 14.3 between *C. medica* and wild citrus.

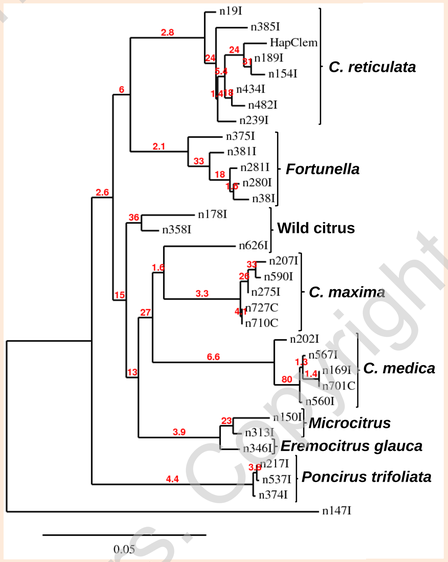
Implication of SNP polymorphism at proteic level

Population	S	π_T	π_{sil}	π_{syn}	π_{nonsyn}	π_{nonsyn}/π_{syn}
C. reticulata	9.15	0.0052	0.0084	0.0108	0.0027	**0.4111**
C. maxima	3.73	0.0035	0.0050	0.0041	0.0014	0.2051
C. medica	2.92	0.0029	**0.0043**	0.0048	0.0009	0.2559
Fortunella	8.54	0.0056	0.0086	0.0085	0.0035	0.2849
Wild citrus	**9.77**	**0.0090**	**0.0158**	**0.0152**	0.0035	0.2916
Microcitrus	5.58	0.0055	0.0083	0.0090	0.0033	0.1946
Eremocitrus	3.31	0.0060	0.0093	0.0128	**0.0037**	0.1675
Poncirus	**2.35**	**0.0029**	0.0048	**0.0032**	**0.0004**	**0.0880**
Media	6.17	0.0029	0.0081	0.0086	0.0024	0.2373
Whole pop	**40.77**	**0.0123**	**0.0209**	**0.0202**	**0.0057**	**0.5744**

(S) Segregating sites, (π_T) Nucleotide diversity total, (π_{sil}) Nucleotide diversity silent sites, (π_{syn}) Nucleotide diversity synonymous sites, ($\pi_{nonsyn/syn}$) Ratio Nucleotide diversity nonsynonymous / synonymous sites, (-) parameter not possible to be calculated, only one genotype, Red color (maximum value), Blue color (minimum value).

Non-synonymous subtitutions were lower than synonymous substitutions in most of the cases, $\pi_{nonsyn}/\pi_{syn} < 1$. Only in HYB in *C. reticulata* (1.42056), F3'H in *C. reticulata* (1.76712) and *Fortunella* (1.04845), EMA (2.27313), MDH (1.06473) and ATMR (1.876) in the whole population showed higher π_{nonsyn}/π_{syn} rate.

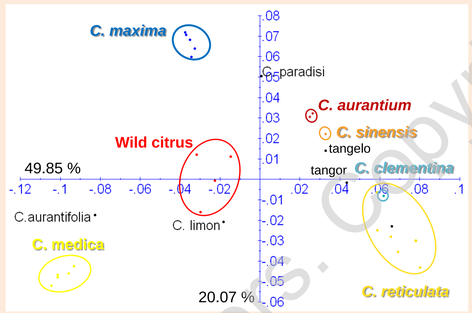
NUCLEAR PHYLOGENY BETWEEN RELATED GENERA AND *CITRUS* BASIC TAXA



Phylogenic tree of *Citrus* basic taxa and related genera
Branch support in red color. Best akaike (AIC) tree.

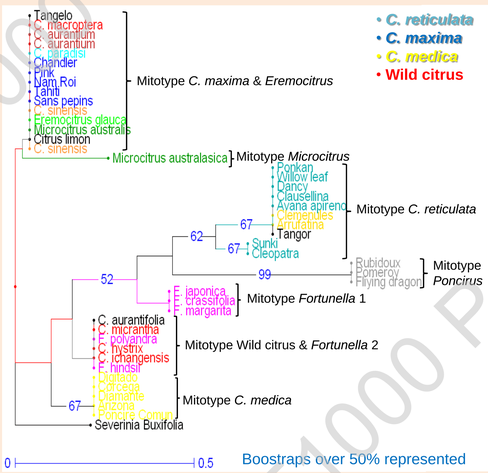
The phylogeny analysis was made joining together the 27 sequences for each genotype. The best model that fits with our data was TVM+I+G+F (with SH-like branch supports alone). It takes into account the nucleotide substitution model TVM (5 substitution classes), the proportion of invariable sites (I), the nucleotide frequency (F) and the Gamma distribution (G). The first specie that was separated from the others was *Poncirus trifoliata*. After that two clearly groups were formed, one which includes *C. reticulata* and *Fortunella* and the other group formed by the wild citrus, *C maxima*, *C. medica*, *Microcitrus* and *Eremocitrus*. These two last genera of Australian origin are strongly associated. The only ancestor group that did not formed a well defined clade was the wild citrus.

RELATION BETWEEN SECONDARY SPECIES AND BASIC *CITRUS* TAXA



Genetic relationship between secondary *Citrus* species and basic taxa (factorial analysis; axes 1/2)

Almost 70% of the diversity is explained by the first two axes. The basic citrus taxa are well distinguished. Only *C. macroptera* is clearly differentiated from the other wild Citrus (axes 3 not show) and should be of inter-specific origin. Secondary species are positioned in between their supposed parents:
C. sinensis between *C. maxima* and *C. reticulata* gene pools.
C. paradisi between *C. sinensis* and *C. maxima.*
C. lemon between *C. aurantium* and *C. medica.*
C. aurantifolia between *C. medica* and *C. micrantha.*
The complete similarity of this figure with the genetic organization displayed by previous SSRs studies suggests a predominantly neutral behavior of the present SNPs markers.



Mitochondrial diversity and cytoplasm genome origin of secondary *Citrus* species (NJ tree)

Nine mitotypes were encountered. Some of them associated *Citrus* taxa with genotypes of related genera. Two *Fortunella* were associated with wild *Citrus* mitotype while *Eremocitrus* and *microcitrus australis* were associated with *C. maxima* mytotype. *C. macroptera* belongs to the *C. maxima* mitotype, it comfort the hypothesis of an inter-specific origin. The low number of markers do not allow establishing an inter-generic maternal phylogeny. However, as basic taxa of cultivated citrus are clearly differentiated, it gives information on the cytoplasmic genomes origin of the secondary species. Interestingly *C. sinensis*, *C. aurantium*, *C. paradisi* and *C. limon* share the *C. maxima* mitotype confirming its contribution in the genesis of these secondary species. *C. aurantifolia* share the *C. micrantha* mitotype.

CONCLUSIONS:

1130 SNPs *loci* were revealed in this work. The average SNP rate is 6.4 time higher between than within *Citrus* basic taxa. Nuclear phylogenetic analysis suggests that *C. reticulata* and *Fortunella* form a consistent clade differentiated from the clade including the other basic taxa of cultivated citrus (*C. maxima, C. medica* and wild citrus). Within *Citrus* diversity study is in agreement with previous molecular markers study and provides a powerful tool for *Citrus* genotyping and management of the *Citrus* Germplasm collection. A new mitotype was observed for *Microcitrus australasica* while two different mitotypes were identified for *Fortunella*.