

Genome sequence analysis and comparisons reveal ancestral hybridization and admixture events in the origins of some citrus cultivars

Gmitter Jr. F.G.¹, Ollitrault P.², Machado M.A.³, Reforgiato Recupero G.⁴, Talon M.⁵, Roose M.L.⁶, Navarro L.⁷, Wu G.⁸, Jaillon O.⁹, Morgante M.¹⁰, and Rokhsar D.S.⁹

¹University of Florida - Citrus Research and Education Center, USA; ²CIRAD, France; ³Centro APTA Sylvio Moreira, Instituto Agronômico, Brazil; ⁴C.R.A.-Istituto Superiore per l'Agricoltura, Italy; ⁵IVIA, Centro de Genómica, Spain; ⁶University of California at Riverside, Department of Botany and Plant Sciences, USA; ⁷IVIA, Centro de Protección Vegetal y Biotecnología, Spain; ⁸USDOE, Joint Genome Institute, USA; ⁹Genoscope - CNS, France; and ¹⁰University of Udine, Istituto di Genomica Applicata, Italy. fgmitter@ufl.edu

Since its inception, the International Citrus Genome Consortium (ICGC) has pursued development of freely available genomics resources and tools for the benefit of the citrus research community; these efforts were conducted by citrus scientists in concert with researchers from some of the world's major genome sequencing centers and corporations. An early ICGC goal was to produce a full-length genome assembly. Two reference genome sequences were made publicly available in January 2011, from a 'Clemenules' clementine-derived haploid and diploid 'Ridge Pineapple' sweet orange, produced using Sanger and 454 sequencing technologies respectively. Subsequent work has proceeded to improve the assemblies and their annotation; a brief summary of their characteristics and status of the assemblies will be presented. As sequencing technology has evolved and costs have dramatically decreased, other citrus genomes have been sequenced (the diploid 'Clemenules', 'Ponkan', 'Willowleaf', 'W. Murcott' mandarins; 'Siamese Sweet' and 'Chandler' pummelo; and sour orange, among others). A comparative analysis of these genomes has made it possible to describe the genomic contributions from ancestral outbred populations to major cultivar types, and to propose likely underlying models. The results confirmed previous reports that clementine arose from a hybridization event between 'Willowleaf' mandarin and sweet orange. Further, introgressions of *Citrus reticulata* and *Citrus maxima* germplasm, long implicated in its ancestry, clearly defined the segmented mosaic genome of sweet orange, through the ancient admixture of these species. Finally, sour orange was found to be a direct hybrid of a pummelo and mandarin. Interestingly, although the two pummelos sequenced appear to have their derivations from a single wild species (*C. maxima*), the mandarins sequenced reveal ancestral admixture of *C. maxima* with *C. reticulata*. Implications and future directions for collaborative research will be discussed.

S03O02

Sequencing of 150 citrus varieties: linking genotypes to phenotypes

Terol J.¹, Carbonell J.², Alonso R.², Tadeo F.R.¹, Herrero A.¹, Ibáñez V.¹, Muñoz J.V.¹, López-García A.¹, Hueso L.¹, Colmenero-Flores J.M.³, Conesa A.², Dopazo J.², and Talón M.¹

¹Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Genómica, Spain; ²Centro de Investigación Príncipe Felipe (CIPF), Bioinformatics and Genomics Department, Spain; and ³Instituto de Recursos Naturales y Agrobiología (IRNAS), Departamento de Biotecnología Vegetal, Spain. talon_man@gva.es

Elucidation of genomic sequences is an essential and necessary first step to identify genetic variants associated with agronomic traits of interest. The CITRUSEQ Consortium, a joint effort developed by public institutions and private companies has sequenced the genomes of 150 citrus varieties with the Illumina platform at an average coverage higher than 60x. Paired end genomic DNA libraries were constructed and average readings of (2x) 100 bp length with an average insert size of 500 bp were performed. For each variety, useful total average readings were higher than 220 million and total mapped readings approached 212 million. The haploid genome of *Citrus clementina* (www.phytozome.com) was used as a reference genome. Sequencing statistics indicate an average representation higher than 92% of the reference genome with coverage of 15x. More than 150 citrus genomes including the 3 ancestral taxa and many rootstocks and species of mandarins, oranges, lemons, grapefruits and limes were sequenced. The consortium has generated a web portal that allows searching and sequence comparison among the approximately 8 million genes found and the 45 billion base pairs read. Initial analyses suggest that the stored information can help unravel fundamental aspects of biological and evolutionary interest such as the origin of citrus, the mechanism of generation of spontaneous varieties, the identification of recombination hotspots, the effect of induced mutations, etc. Preliminary data suggest, for example, the presence of mitotic crossovers