E

BMT/11/23
ORIGINAL: English
DATE: September 9, 2008

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES AND DNA PROFILING IN PARTICULAR

Eleventh Session
Madrid, September 16 to 18, 2008

COMPARATIVE GENOMIC HYBRIDIZATION FOR IDENTIFYING MUTATION VARIETIES

Document prepared by experts from Spain

COMPARATIVE GENOMIC HYBRIDIZATION FOR IDENTIFYING MUTATION VARIETIES

Gabino Ríos, Miguel A Naranjo, Domingo J Iglesias, Omar Ruiz-Rivero, Marion Geraud, Antonio Usach, Manuel Talón[§]

Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias, Carretera Moncada-Náquera km 4.5, 46113 Moncada (Valencia), Spain

[§]Corresponding author:

Email addresses:
GR: grios@ivia.es
MAN: mnaranjo@ivia.es
DJI: iglesias@ivia.es
ORR: oruiz@ibmcp.upv.es
MG: geraud@ensam.inra.fr
AU: anlous@ivia.es
MT: mtalon@ivia.es

ABSTRACT

Background

1. Many fruit-tree species, including relevant *Citrus* spp varieties exhibit a very peculiar reproductive biology that impairs breeding and strongly constrains genetic improvements. In citrus, juvenility increases the generation time while sexual sterility and self-incompatibility prevent the production of homozygous varieties. Genomic technology may provide citrus researchers with a new set of tools to address these various restrictions. In this work, we report a valuable genomics-based protocol for the structural analysis of deletion mutations on a heterozygous background.

Results

2. Two independent fast neutron mutants of self-incompatible clementine (*Citrus clementina* Hort. Ex Tan. cv. Clemenules) were the subject of the study. Both mutants, named 39B3 and 39E7, were expected to carry DNA deletion lesions in hemizygous dosage. Array-based Comparative Genomic Hybridization (array-CGH) with a 20K *Citrus* cDNA microarray allowed the identification of underrepresented genes in the two mutants. Comparison of putative deleted citrus gene hits against annotated plant genomes, especially poplar, predicted the presence of a large deletion in 39B3 mutant of about 700 kb and at least two deletions of approximately 100 and 500 kb in the other mutant. The deletion in 39B3 was further characterized through reciprocal PCR on available *Citrus* BACs that helped to draw a partial physical map of the deletion. Among the deleted genes, the function of *ClpC*-like, coding for a putative subunit of a multifunctional chloroplastic protease involved in chlorophyll *b* synthesis was directly related to the mutated phenotype since the mutant contained lower chlorophyll *a*/*b* ratio in green tissues. This observation confirmed a predictable consequence of the hemizygous deletion of the *ClpC*-like gene.

Conclusions

3. In this work, we report the use of array-CGH for the successful identification of genes included in a hemizygous deletion induced by fast neutron irradiation on *Citrus clementina*. The study of gene content and order into the 39B3 deletion also led to the unexpected conclusion that microsynteny and local gene colinearity in this species were higher with *Populus trichocarpa* than with the phylogenetically closer *Arabidopsis thaliana*. This work corroborates the potential of *Citrus* genomic resources to assist mutagenesis-based approaches for functional genetics, structural studies and comparative genomics, and hence to facilitate citrus variety improvement.

BACKGROUND

4. The rapid increase in the world's population, field degradation by soil salinization and erosion, and the likely fluctuations in climate caused by global warming will pose new and known challenges to agriculture during this century [1]. Crop improvements required to cope with these dares will likely be attained through agronomic advances, leading to a better use of fertilizers, protection agents or soil rescue, and exploitation of recent technologies for plant breeding. Despite the outstanding importance of genetics-based breeding applied to spontaneous mutations and conventional hybrids, molecular and genomic tools are still expected to develop their great potential for crop improvement through functional genetics analysis, involving gene and function discovery and genome modification. Genomic approaches allowing rapid identification of new plant genes, however, has not been followed by an evenly growing process of gene functional characterization, causing the so called "phenotypic gap" , a break between the achievement of genetic information and knowledge of gene function, and therefore gene exploitation [1,2]. In order to fill this gap, functional genetics has taken advantage of mutagenesis through two complementary ways. On one hand, traditional forward genetics approaches pursue the cloning of the gene causing a particular mutant phenotype related to an interesting pathway or process. Reverse genetics strategies, on the other hand, start from the known sequence of a specific gene to find a knock-out on it that may allow gene function discovery [3,4]. Isolation or detection of mutated genes become straightforward when the gene of interest is tagged by insertional mutagenesis techniques based on transposon or T-DNA technologies [5-7]. In recent years, the development of genetic maps with increasing complexity and TILLING (targeting induced local lesions in genomes) procedures for identification of single nucleotide changes [8,9] have also facilitated the analysis and discovery of non tagged mutants originated by classical chemical mutagenesis or irradiation [10]. However, only a few plant models with short life cycle, annotated genome and standard reproductive biology can fully benefit from these powerful approaches while in many species of agricultural interest, especially woody fruit trees, such techniques are now rather inaccessible.

5. Citrus, one the most important fruit crops worldwide, are woody, perennial trees requiring a juvenility period of several years and frequently are parthenocarpic and sexually self-incompatible [11,12]. These conditions considerably impair traditional breeding while current alternative methods for variety improvement such as marker-assisted selection, somatic hybridization or mutation breeding also are of limited use mostly because the *Citrus* genome is scarcely characterized. Genomic technology, including methods to rapidly identify and manipulate genes of agricultural interest, holds promise of improvements in several areas that may be difficult through traditional approaches. In recent years, *Citrus* has been the target of several genomic developments including large EST collections [13-16], cDNA and oligonucleotide-based microarrays [13,17,18], BAC libraries and BAC end sequencing (BES)

(to be published). However, functional studies, i.e. genetic transformation and the capability to perform reverse genetic analyses, are also considerably impaired. In citrus, high throughput transgenic programs such as the generation of RNA interference knockouts, activation tagging through enhancer elements, gene-trap T-DNA insertions, or transposon tagging systems have not yet been developed. Since no efficient tagging or insertional procedures are available in these species, other gene disruption methods including strategies based on genome-wide mutagenesis such as TILLING and fast neutron mutagenesis have been initiated. These approaches are non-transgenic and may have particular interest for the industry where the debate on genetically modified organisms has restricted crop improvement. Both approaches, however, are of limited usefulness as strategies for reverse genetics because of the lack of knowledge on *Citrus* genomic sequence and the large amount of space required for mutant populations of suitable size. ECOTILLING, however, on natural citrus variants and microarray-based detection of deletions on fast neutron citrus mutants in a more direct genetic strategy are apparently very straightforward approaches. In this work, we planned to explore the possibilities and potential of this last idea using two fast neutron *Citrus clementina* hemizygous mutants from the IVIA collection and a 20K cDNA citrus microarray.

6. Physical mutagenesis through fast neutron irradiation has been reported to cause variable genomic deletions ranging in size from several hundreds of base pairs to 12 kb in *Arabidopsis thaliana* [19,20]. Several approaches have been used to characterize deletions at the molecular level. These mostly include positional cloning [21], a method applicable to any kind of genetic lesion that, however, needs highly saturated genetic maps; PCR-based reverse genetics techniques [20,22], requiring a previous considerable knowledge of genomic sequence; and genomic subtraction procedures [23-25], which do not need sequence information but are strongly dependent on the gene dosage. The current ignorance on *Citrus* genome sequence and the hemizygous nature of the induced deletions preclude the use of these strategies although in principle enables array-based procedures, following the strategy used for the identification of homozygous gene deletions in *Arabidopsis* [26]. Although main applications of microarrays are transcriptome profiling analyses, microarrays can also be used to study DNA variation. Oligonucleotide arrays are particularly suited for the detection of single nucleotide mismatches during hybridization, and hence for the discovery of novel DNA variants or the determination of known variants. The origin of this technique relies on a cytogenetic method described 25 years ago named “Comparative Genomic Hybridization” (CGH) that used differential DNA hybridization on chromosome spreads for visualization of deleted or amplified genomic regions in tumour tissues [27]. Subsequently, different laboratories mostly working on cancer research applied independently the microarray technology to the genomic DNA hybridization procedure, a technique consequently named array-CGH [28-32]. Array-CGH was successfully utilized to detect gene duplications in *Arabidopsis* and rice [33], and to validate aneuploidy analysis performed by quantitative fluorescent PCR in *Arabidopsis* [34]. Therefore, this method has proven to be suitable to study chromosomal imbalances in plants.

7. For the characterization of the deleted regions we also leaned on comparative genomics with other dicots since physical citrus maps are not yet available. Comparative genomics takes advantage of available information on gene content and order in genomic DNA from different species to infer phylogenetic relationships and formulate hypotheses on DNA evolutionary dynamics. Whole genomes are preferentially compared when available, but more often relatively short stretches of DNA or polymorphic markers are used. A previous work, for instance, reported that microsynteny between the sequence of a 282 kb region from *Poncirus trifoliata* L. Raf., a *Citrus* close relative, and two regions from *Arabidopsis* chromosome 1 was too low to support the use of *Arabidopsis* sequence annotation for positional cloning strategies in *Citrus* or

Poncirus [35]. It has also been shown that papaya exhibited higher level of colinearity with the poplar than with the *Arabidopsis* genome despite that *Carica papaya* is a basal member of the Brassicales [36].

8. The main objective of this work was to identify deleted genes on a heterozygous genetic *Citrus* background, provided by fast neutron generated mutants, through array-Comparative Genomic Hybridization. In addition, we also explored the possibility of using comparative genomics with annotated dicot genomes assisted by BAC end sequencing for the generation of realistic partial physical maps of the deleted *Citrus* regions.

RESULTS AND DISCUSSION

Procedure for the characterization of hemizygous deletions in *Citrus*

9. The proposed procedure to identify deleted genes is illustrated in Figure 1 and its potential to structurally characterize hemizygous deletions is exemplified below with *Citrus* mutants as starting plant material. This method uses microarrays to hybridize genomic DNA extracted from the deletion mutants to render a list of underrepresented genes. The putative deleted genes are then validated through gene dosage evaluation by real-time PCR with specific primers. Deleted genes may in this way contribute to the identification of the molecular mechanisms underlying the observed phenotypes by means of a candidate gene approach, validated by physiological analyses or genetic transformation [37]. An alternative for functional genetics may utilize a reverse genetics strategy to find additional phenotypes in a particular mutant based on reported bibliography in other species. In non-sequenced genomes or in plants with poor development of physical maps, further characterization of deletions at the structural level requires TBLASTX similarity searches against databases containing the sequence annotation of known eudicot genomes, such as *Arabidopsis thaliana*, *Populus trichocarpa* and *Vitis vinifera*. These searches produce a listing of homologous genes and syntenic genomic regions between these four species. Local physical maps of deletions are built allocating the deleted gene sequences and the syntenic genomic fragments from these other eudicots into a BES database of the species of interest. Lastly, specific PCR on the array of BACs contributes to confirm gene content and order on the lineal structure of the deletions. The results may also be used in comparative genomics analyses to study evolutionary dynamics and phylogenetics.

Identification of deleted alleles in 39B3 and 39E7 fast neutron mutants of *Citrus clementina*

10. Two mutants obtained by fast neutron mutagenesis of wild type *Citrus clementina*, were selected from the IVIA mutant collection for this study. The mutants, named 39B3 and 39E7, were expected to carry DNA deletion lesions in hemizygous dosage and showed an altered time delay of natural color break in fruit peel. The 39B3 mutant exhibited a delay in color change from green to orange while 39E7 was better characterized by an abnormal final yellowish color instead of the natural orange coloration. Putative deleted genes in the mutants were first identified through an approach based on genomic hybridization (array-CGH) that exploited a recently developed *Citrus* microarray containing 21240 cDNAs [13,14]. To this end, total genomic DNA from four independent samples of mutants 39B3 and 39E7 were Cy3 or Cy5-labelled and cohybridized with wild type DNA labeled with the complementary Cy5 or Cy3 probe on four independent microarray slides. Fluorescence intensity data were normalized and single EST showing a mutant/wild type signal ratio lower than 0.7 fold, with a P-value lower than 0.2 (39B3) or 0.1 (39E7), were selected as putative candidates.

11. The number of ESTs fulfilling these criteria was 24 and 78 for mutants 39B3 and 39E7, respectively. One of the 39B3 positives [GenBank: CX299090], composed of three unrelated sequences was discarded for subsequent analysis due to its chimerical nature. In order to validate the array-CGH results, gene dosage of several putative candidates was determined through real-time PCR quantification of mutant/wild type signals for candidate ESTs as related to a reference undeleted gene [GenBank: CX293764]. The results showed that gene dosage from 39B3 candidates ranged from 0.50 to 0.60 when genomic DNA from the 39B3 genotype was tested, while ranged from 0.96 to 1.15 when the assayed DNA proceeded from the 39E7 genotype (Table 1). Similar results, corroborating the presence of putative deleted genes at half dosage, were also obtained for the 39E7 mutant. Therefore, the developed array-CGH procedure proved to be an appropriate tool to identify genes in hemizygous content in the self-incompatible clementine.

Clustering of homologues of *Citrus* deleted genes in the poplar genome

12. To elucidate hypothetical clustering of *Citrus* deleted genes in the genome, microsynteny comparisons with homologous stretches from the sequenced genomes of *Arabidopsis thaliana*, *Populus trichocarpa* and *Vitis vinifera* [38-40] were performed. In order to compare protein sequences, TBLASTX, which searches for translations of a crude genome similar to a translated query, was utilized with an E-value cut-off of 10^{-5}. The homologous regions produced by the best TBLASTX hit of each of the *Citrus* candidate genes were located on the chromosome maps of *Arabidopsis*, poplar and grapevine. Homologues of *Citrus* genes were then grouped into clusters in each species when the distance between them was shorter than 250 kb. The second and third TBLASTX best alignments were similarly placed in the respective maps when they were included in an existing cluster. In this case, a binding line was drawn linking the second and third hits to the best hit of the same *Citrus* query. Thus, two chromosomal maps, one for each mutant, in the three species was obtained. Figure 2 represents in detail chromosome mappings of the 39B3 mutation, which was subjected to further analyses. The results indicate that the *Populus* mapping exhibited rather lower complexity than the *Arabidopsis* and grapevine ones since it included less chromosomes and only 3 clusters although the number of 39B3 candidate genes represented in the map was identical (21) for the three genomes. It is worth to note that the number of represented hits in these mappings is higher than 21 due to the inclusion of second or third homologues. In *Populus,* most of the candidate genes mapped to two different genome regions of approximately 700 kb long in chromosomes 12 and 15, two duplicated chromosomes probably originated during the recent genome duplication event occurred in this species [39]. These two clusters contained 17 and 15 hits respectively while the third one placed in chromosome 16 had only one hit. In contrast, the number of clusters in *Arabidopsis* and *Vitis* were 9 and 11, respectively, and none of them contained more than 11 hits. Furthermore, cluster number (and clustering density) of the homologues of 39E7 putative deleted genes was also lower (and higher) in *Populus* than in *Arabidopsis* or *Vitis*, although the differences were smaller (data not shown).

13. Overall, these observations suggest that *Populus* genomic regions homologous to the *Citrus* deletions were less fragmented than their *Arabidopsis* and *Vitis* counterparts, and consequently microsynteny on the considered segments was higher with the *Populus* genome. These results may be striking since *Citrus* and *Arabidopsis* belong to Sapindales and Brassicales orders (eurosids II clade) while *Populus* is included in the eurosids I clade, and *Vitis* is part of Vitaceae, a family outside of rosids [41].

Gene arrangement and partial physical map of the 39B3 deletion

14. The closer microsynteny observed between the 39B3 deletion and the two duplicated homologous regions in poplar enabled to predict gene structure by direct inference of gene position from the *Populus* sequences. This assumption leaded to the gene arrangement depicted in Figure 3. Twenty genes out of twenty-one having high similarity with *Populus* homologues were directly located on the *Citrus* deletion fragment by combining the two clusters found on *Populus* chromosomes 12 and 15, which shared 12 hits. Inclusion of the 21st gene, homologue of a *Populus* gene placed on chromosome 16, in the 39B3 deletion was based on his location on the right end of the *Citrus* BAC CCER1019D04 (named B12, see below), whose left end shared identity with other deleted gene [GenBank: CX295702]. The accession number and protein similarity of these 21 genes, numbered according to the ordered position of their homologues on the poplar genome (Figure 3), are depicted in Table 2 that also shows coding strand sense of poplar homologues. The coding strand was coincident for the *Populus* paralogous genes present in chromosomes 12 and 15, except for genes similar to *Citrus* CX308429, located in position 8 in Figure 3.

15. In order to get deeper insight into protein function and gene ontology (GO) of the putative deleted genes listed in Table 2, Blast2Go searches [42] were performed. The retrieved information revealed a great diversity of protein similarities and functions, most of them related to protein processing, redox balance, hydrolysis, RNA binding and regulation among others. Interestingly, GO analysis highlighted an elevated relative number of genes likely coding for plastid targeted proteins (6 out of 21) that may in principle be related to the 39B3 phenotype, i.e. delay in chloroplast degradation.

16. Furthermore, the recent sequencing of 46,000 *Citrus clementina* BAC ends (to be published) enabled the construction of a physical map of the 39B3 deleted region. To this end, two DNA sequences covering 700 kb along the *Populus* chromosomes 12 and 15, containing the genes homologous to the *Citrus* deleted candidates, were BLASTed against the *Citrus* BAC end sequences. The homology search rendered 33 BACs with a BLASTN E value lower than 10^{-5} for both paralogous regions. In subsequent analyses, redundant BACs were discarded, while additional candidate BACs were obtained by comparing these previous ones with the BES database to yield overlapping BACs. Moreover, BACs with both ends showing similarity to repetitive DNA that may cause ambiguous positioning and inaccurate gene dosage measurement were also discarded. Finally, a partial physical map containing 13 BACs systematically named B1 to B13 (Table 3) was provided by standard PCR of BAC end amplicons against BAC templates and *in silico* search of overlapping antiparallel ends (Figures 4A, B).

17. This mapping contained three gaps, one at the 5' deletion junction and two internal ones (Figure 4B) limiting three main BAC clusters, composed of B1 to B4, B5 to B8, and B9 to B13. BACs B11 and B12 were connected by unigene aCL4690Contig1 coding for a putative subunit ClpD of an ATP-dependent Clp protease, whose sequence was shared by both BACs. Similarly B12 and B13 interaction is mediated by unigene aCL1915Contig2 (Table 2, 3). Real-time PCR quantification of gene dosage for some of the BAC ends (Figure 4A) confirmed the presence of these sequences at half dosage in the mutant genotype, indicating that the 39B3 mutation is a hemizygous deletion. Indeed, all analyzed BACs covered an internal segment of the deletion except B13 that exhibited haploid gene dosage on the left end and diploid dosage on the right one, suggesting that B13 contained the 3' border of the 39B3 deletion.

18. The above results indicated that the microsynteny between *Citrus* and *Populus* genomes was high enough to predict gene arrangement and to build a partial physical map of a *Citrus* genomic segment of about 700 kb, as inferred from the length of poplar homologous regions. Nevertheless, the observation that a 700 kb *Citrus* fragment only contains 21 genes may result striking considering an average distance of 10 Kb between adjacent genes, as deduced from the estimations of *Citrus* genome size (367 Mb) and gene number (35,000-40,000). It should be noted, however, that the microarray used in these analyses contains approximately between 2/3 and 1/2 of the estimated gene content of the *Citrus* genome, which may account for a major part of the hypothetical "loss" of deleted candidates. While this is a feebleness of the current available *Citrus* arrays, non-attributable to the array-CGH procedure, wider results are expected after the development of a more representative cDNA microarray. Other limitations of the method may also be related to the differential hybridization potential of different cDNAs including for instance cross-hybridizations. In this regard, oligonucleotide arrays are particularly suited for the detection of dissimilar DNA variants. Alternatively, synteny might be limited to several genes located on a bulk of non-conserved sequences inside this 700 Kb region, a possibility that may only be corroborated after genome sequencing.

19. Overall, the data indicated that the *Populus* genome is a useful model for comparative genomics and that can be used to characterize hemizygous deletions in *Citrus*.

The *Citrus* 39B3 deletion shows higher local gene colinearity with *Populus* than with *Arabidopsis*

20. Local gene colinearity between two genomic fragments is determined by the number of paralogous genes arranged in the same order. Therefore, not only permanence of genes in their original chromosomal location, but also conservation of gene order, affects local colinearity. In order to validate the gene arrangement postulated in Figure 3 and consequently to estimate gene colinearity of the 39B3 *Citrus* deletion with *Populus* homologous fragments, we mapped by PCR the 21 genes listed in Table 2 on the physical map of Figure 4B. All but three genes showed at least one PCR product on the array of 13 BACs, confirming that those genes were effectively included in the 39B3 deletion (Figure 5A). In addition, PCR reactions reproduced at the BAC size resolution the expected gene order outlined in Figure 3, corroborating the gene arrangement deduced by comparative genomics. In Figure 5B, the genes rendering a positive PCR signal were linked to the physical map position with an arrow. Moreover, genes 3, 4 and 9 corresponding to unigenes aCL3991Contig1, aC18005F10Rv_c and aC16014F08SK_c, respectively, did not show a detectable PCR signal on purified BACs, although their respective primers produced a band with the expected size when tested against genomic DNA from normal clementine variety (data not shown). These genes were most likely placed into the two reported internal gaps of the physical map, as suggested by the border situation of their neighbouring genes.

21. These results indeed confirm high local gene colinearity with poplar in the genomic region covered by 39B3 deletion. Taken together gene content and order conservation (Figures 2 and 5), it is inferred that in the studied DNA deleted segment there was higher gene colinearity with *Populus*, which diverged about 109 million years ago (Mya), than with *Arabidopsis*, splitting from the *Citrus* lineage about 87 Mya [41], despite gene colinearity is generally correlated with phylogenetic closeness. A similar conclusion has been reached in our group, after comparing the whole collection of *Citrus* BES with the poplar and *Arabidopsis* genomes (to be published), and also in previous works in papaya and melon. In papaya, BES alignment to the annotated genomes rendered higher gene colinearity with *Populus* than with *Arabidopsis*, although both *Arabidopsis* and papaya belong to the order

Brassicales [36]. In melon, microsynteny studies based on the sequence of two BACs also concluded that melon was closer to *Populus* than to *Arabidopsis* or *Medicago truncatula* [43]. These observations may be explained by a differential genome evolutionary dynamics in poplar and *Arabidopsis* lineages [44]. The more recent appraisals estimated that last whole genome duplications occurred not later than 60-65 Mya in *Populus* and around 24-40 Mya in *Arabidopsis* lineages [39,45-47]. Despite the older poplar event, genome rearrangements involving gene loss and translocation following these duplications were much more frequent in *Arabidopsis* ancestors [48]. Such a highly active genome dynamics probably caused the dispersion of genes and the subsequent reduction in synteny and gene colinearity with even related species. The different behaviour of *Populus* and *Arabidopsis* ancestral genomes still deserves further explanation. It has been suggested that woody long-lived species like poplar trees may undergo a slower genome dynamics due to their juvenile period that delays sexual fecundation for several years and to the recurrent contribution of gametes from aged individuals of previous generations [39]. In addition, species like *Arabidopsis thaliana* may have very active mechanisms for unequal or illegitimate recombination causing frequent chromosomal rearrangements such as translocations, insertions and deletions.

Chlorophyll *a*/*b* ratio is modified in 39B3 mutant

22. Structural studies describing gene arrangement on a particular deletion have outstanding importance for linking a specific mutant phenotype with an impaired gene. The 39B3 deletion removed at least a set of 21 genes including an elevated relative number of genes likely coding for plastid-targeted proteins (6 out of 21) that may in principle be related to the mutant phenotype, i.e. delay in chloroplast degradation. Among these genes, unigene aCL766Contig1, one of the 39B3 hits validated by real-time quantitative PCR (Table 1) coding for a ClpC-like protein, may have certain relevance in the 39B3 phenotype. Plant ClpCs are ATP-binding proteins located in the stroma of chloroplasts which have been found to be associated with the protein import machinery [49] and with the Clp protease complex [50]. In fact, ClpC has been related to protein translocation across the chloroplast inner envelope membrane and to multiple processes requiring proteolytic cleavage, as protein turnover and regulation [51,52]. In *Arabidopsis*, insertional mutagenesis in the *ClpC1* gene caused chlorosis, growth retard, photosynthetic damage and defects in chloroplast protein import [53-55] and no double knock-out of *ClpC1* and the less expressed *ClpC2* genes were obtained, suggesting that *ClpC* function is essential in plants [56]. In addition, a mutant impaired in *ClpC1* mRNA processing accumulated chlorophyllide *a* oxygenase protein (CAO), a key enzyme for the synthesis of chlorophyll *b* from chlorophyll *a*, leading to a reduced chlorophyll *a*/*b* ratio [57]. Interestingly, 39B3 mutants also contained a lower ratio of chlorophyll *a*/*b* in young and old leaves and fruit exocarp. Differences with normal values were low, ranging from 15% to 23% although they were statistically significant (Figure 6). This distinct chlorophyll composition was not accompanied by alterations in the total content of chlorophylls in the leaves although pigment levels in 39B3 fruit exocarp, as expected, were clearly higher (0.48 mg/g fresh weight) than in the peel of normal fruit (0.15 mg/g fresh weight) that has initiated chlorophyll degradation (Table 4). The chlorophyll accumulation observed in the 39B3 exocarp, however, is higher than the maximum reached in normal clementine fruits (0.35 mg/g f w) [58]**,** suggesting that the mutation also induced total chlorophyll build-up in the fruit peel. Indeed, fruit exocarps of a “wild type” clementine tree showing fruit color delay due to altered environmental conditions showed chlorophyll *a*/*b* ratios equivalent to those found in the standard cultivar (Figure 6) while total pigments had an intermediate value (0.29 mg/g f w) between those of normal and 39B3 genotypes (Table 4). Thus, other features found in *Arabidopsis ClpC1* knock-out lines, like reduction of total chlorophyll content or chloroplast defects were not produced by the hemizygous

39B3 deletion, suggesting that gene dosage may either compensate or affect differentially ClpC protein functions.

CONCLUSIONS

23. In this study, we propose a procedure for the genetic characterization of genomic hemizygous deletions in mutants from plant species with non-sequenced genomes. The procedure is illustrated with the study of the 39B3 *Citrus clementina* deletion, generated by fast neutron bombardment. The proposed strategy utilizes several genomic resources such as array-Comparative Genomic Hybridization (array-CGH) technology, EST and BAC end sequencing databases and poplar genome annotation.

24. The array-CGH results led to the conclusion that the 39B3 deletion removed at least 21 genes while a partial physical map of about 700 kb of the deleted region was inferred by comparison of two homologous genomic regions from poplar with a *Citrus* BES database.

25. Structural data including gene content and order in the deletion was utilized for microsynteny and local gene colinearity studies concluding that in the studied region *Citrus* is more similar to *Populus* than to *Arabidopsis*, a phylogenetically closer species. This observation supports previous works on other species and suggests that *Arabidopsis* lineage underwent a quicker genome evolutionary dynamics than the *Populus* one.

26. Among the deleted alleles, the function of *ClpC*-like, coding for a putative subunit of a protease involved in chlorophyll *b* synthesis was directly related to the mutant phenotype since green mutant tissues contained lower chlorophyll *a*/*b* ratio, a predictable consequence of the hemizygous deletion of the *ClpC*-like gene.

METHODS

Plant material

27. Plant material from approximately 6 years-old standard and 39B3 and 39E7 mutant lines of the clementine (*Citrus clementina* Hort. Ex Tan. cv. Clemenules) cultivar grown under standard agricultural practices at the Instituto Valenciano de Investigaciones Agrarias (IVIA) was used in this study. Commercial highly heterozygous clementine cultivars are considered “wild type” material while the 39B3 and 39E7 genotypes that belong to the IVIA mutant collection were obtained through bud irradiation with fast neutrons (5-6 Gy) at the Instituto Tecnologico e Nuclear (Sacavem, Portugal) in the frame of a much wider breeding program. Both mutants are expected to carry DNA deletion lesions in hemizygous dosage and showed altered patterns of colour change of fruit peel.

Array-CGH

28. The protocol was adapted from several published array-Comparative Genomic Hybridization (array-CGH) methods pursuing mainly the measurement of copy-number changes in human genomic DNA [59-61], and the study of large-scale genetic variation of the symbiotic bacteria *Sinorhizobium meliloti* [62]. Genomic DNA was isolated from leaves of “wild type” (control) and mutant plants, using DNeasy plant mini kit (QIAGEN, cat. nº 69104). Four Cy3 or Cy5-labelled independent biological samples from each mutant plant were co-profiled on four 20K *Citrus* cDNA microarrays containing 21240 EST, using Cy5 or Cy3-labelled control genomic DNA, respectively. Label probes were prepared as

follow: Cy3- or Cy5-dCTP fluorescent nucleotides (Amersham Biosciences) were incorporated directly in control and mutant genomic DNA (2µg) using BioPrime Array CGH Genomic Labelling System (Invitrogene, cat. nº 18095-011). Purified Cy5 and Cy3 labelled probes (about 50 µl each) were combined and mixed with 30 µg Cot-1 DNA (Invitrogene, cat. nº 15279-011), 100 µg yeast tRNA (Invitrogene, cat. nº 15401-011), and 346 µl TE buffer pH 7.4. Cot-1 DNA and yeast tRNA were used to block non-specific hybridization. Samples were laid on a microcon YM-30 filter (Millipore, cat. nº 42422), and subsequently centrifuged until sample volume was reduced to approximately 48 µl. Finally, 10.2 µl 20X SSC (Ambion, cat. nº 9763G) and 1.8 µl 10% SDS (Ambion, cat. nº 9823G) were added to the probe mixture to reach a final volume of 60 µl containing 3.4X SSC and 0.3% SDS. For microarray hybridization, probe mixture was denaturised by heating at 97 ºC during 5 minutes, and immediately incubated at 37 ºC during 30 minutes to block repetitive DNA sequences. Hybridization mixture was applied to a 37 ºC pre-warmed hybrid-slip (Sigma, cat. nº Z370274-100EA), and a pre-warmed array slide was lowered onto the mix. Microarrays were hybridized in darkness at 65 ºC overnight (16-20 hours) using a glass array cassette following manufacture instructions (Ambion, cat. nº AM10040). To prevent evaporation of hybridization solution during incubation, 5 µl of 3X SCC were poured into the reservoir inside the cassette chamber. Following hybridization, microarray slides were placed in a rack and cover slip removed by 10 minutes immersion in a washing chamber containing 2X SSC and 0.03% SDS at room temperature (RT). Microarray slides were passed through a series of washes on a shaking platform. Wash series were as follow: 2X SSC, 0.03% SDS for 5 min at 65 ºC, followed by 1X SSC for 5 min at RT, and 3X 15 min washes in 0.2X SSC at RT. After first wash slides were transferred to new racks to minimize transference of SDS to the next washing solution. Microarray slides were dried by centrifugation for 5 min at 300 rpm by using an Eppendorf 5804-R tabletop centrifuge. Arrays were immediately scanned at 5 µm. Cy3 and Cy5 fluorescence intensity was collected by using a ScanArray Gx (Perkin Elmer). The resulting images were overlaid and spots identified by the ScanArray Express program (Perkin Elmer). Spot quality was first measured by the signal-to-background method with parameters lower limit (200) and multiplier (2), and subsequently confirmed by visual test. Data analysis was performed using the Limma package from the R statistical computing software [63-65]. A mutant/wild type signal lower than 0.7, with a P-value not higher than 0.1 (39E7) or 0.2 (39B3) were the cut-off values for positive EST identification. The experimental design of microarray experiments has been loaded into the ArrayExpress database [66] under accessions E-MEXP-1432 and E-MEXP-1433.

Gene dosage measurements

29. Quantitative real-time PCR was performed on a LightCycler 2.0 instrument (Roche), using the LightCycler FastStart DNA MasterPLUS SYBR Green I kit (Roche, cat. nº 3515885001). Reaction composition and conditions followed manufacturer's instructions. Each individual PCR reaction contained 2 ng of genomic DNA from wild type or mutant, obtained with the DNeasy plant mini kit (QIAGEN, cat. nº 69104). Cycling protocol consisted of 10 min at 95 °C for pre-incubation, then 40 cycles of 10 sec at 95 °C for denaturation, 10 sec at 60 °C for annealing and 10-25 sec at 72 °C for extension. Fluorescent intensity data were acquired during the extension time. Specificity of the PCR reaction was assessed by the presence of a single peak in the dissociation curve after the amplification and through size estimation of the amplified product. For gene dosage measurements, we used the relative quantification-monocolor analysis from the LightCycler Software 4.0 package (Roche). This program compares the ratio of a target sequence to a reference DNA sequence in the mutant sample with the ratio of these sequences in a wild type sample. PCR and

normalized calculations were repeated in at least three independent samples from each mutant and wild type, rendering an estimation of target gene dosage in the mutant genotype. Primers for the reference sequence were obtained from CX293764.

Similarity searches

30. DNA sequences of *Citrus* unigenes containing positive array-CGH ESTs were used in online TBLASTX searches against genomic databases from the annotated genomes of *Arabidopsis thaliana* [67], *Populus trichocarpa* [68] and *Vitis vinifera* [69] at an E-value cut-off of 10^{-5}. For each gene, the best hit was placed on a chromosomal map while the second and third hits were only positioned in the map if they were located closer than 250 kb to any other hit. Two 700 kb regions from chromosomes 12 and 15 from the *Populus* genome including homologous genes to 39B3 array-CGH positive unigenes, were used as queries in a BLASTN local search on a *Citrus* BAC end sequence database. Only hits corresponding to those BAC ends showing an E-value lower than 10^{-5} in both chromosome searches were considered for the building of a local physical map of the 39B3 deletion.

BAC isolation and analysis

31. DNA from *Citrus* BACs was isolated with the Rapid Plasmid Miniprep System (Marligen Biosciences, cat. nº 11453-016). Purified BACs were used as templates in PCR reactions in a total volume of 15 µl, including 0.2 mM dNTP, 2 mM $MgCl_2$, 0.5 µM of each primer, 0.38 units of Netzyme DNA polymerase (Molecular Netline Bioproducts) and 0.1 ng of BAC DNA. After an initial denaturing step for 5 min at 95 °C, amplification was performed for 35 cycles of 30 sec at 95°C, 30 sec at 60 °C and 30 sec at 72 °C, followed by 5 min incubation at 72 °C. The PCR product was subjected to 1.5% agarose DNA electrophoresis.

Chlorophyll measurements

32. At least, three developing and mature leaves and fruit exocarp sectors from standard and 39B3 mutant lines of clementine were randomly collected per sample. Fruit exocarp tissues from a “wild type” clementine tree showing fruit colour delay due to altered environmental conditions were also sampled for chlorophyll analyses. Chlorophylls *a* and *b* were extracted with N,N-dimethylformamide for 72 h in the dark at 4 ºC and quantified through the absorbance at 647 and 664 nm following a reported procedure [70]. Absorbance was measured using a Varian Cary 50 UV-visible spectrophotometer (Varian).

AUTHORS' CONTRIBUTIONS

33. GR carried out the microarray hybridizations, standard PCR reactions, similarity searches and data analysis, and drafted the manuscript. MAN isolated DNA from *Citrus* BACs and carried out quantitative real-time PCR. DJI carried out mutant collection screenings, selected mutants and performed chlorophyll measurements. OR-R designed the array-CGH protocol. MG assisted in microarray hybridizations. AU provided plant material, identified mutant genotypes and carried out relevant work on the field. MT conceived and coordinated the project and elaborated the final manuscript.

ACKNOWLEDGEMENTS

Work at Centro de Genómica was supported by INIA grant RTA04-013, INCO contract 015453 and Ministerio de Educación y Ciencia grant AGL2007-65437-C04-01/AGR. We also thank Dr. José Marqués at the Instituto Tecnologico e Nuclear (Sacavem, Portugal) for irradiation with fast neutrons. Help and expertise of A. Almenar, E Blázquez, I. López, I. Sanchís and M. Sancho are gratefully acknowledged.

REFERENCES

1. Miflin B: crop improvement in the 21st century. *J exp bot* 2000, 51:1-8.
2. Brown Sdm, Peters J: combining mutagenesis and genomics in the mouse - closing the phenotype gap. *Trends genet* 1996, 12:433-435.
3. Ostergaard L, yanofsky mf: establishing gene function by mutagenesis in arabidopsis thaliana. *Plant J* 2004, 39:682-696.
4. Bouchez D, Hofte H: functional genomics in plants. *Plant physiol* 1998, 118:725-732.
5. Feldmann Ka: T-DNA insertion mutagenesis in arabidopsis - mutational spectrum. *Plant J* 1991, 1:71-82.
6. Martienssen Ra: functional genomics: probing plant gene function and expression with transposons. *p natl acad sci usa* 1998, 95:2021-2026.
7. Koncz C, Nemeth K, Redei Gp, Schell J: T-DNA insertional mutagenesis in arabidopsis. *plant mol biol* 1992, 20:963-976.
8. Mccallum Cm, Comai L, Greene Ea, Henikoff S: targeting induced local lesions in genomes (tilling) for plant functional genomics. *Plant physiol* 2000, 123:439-442.
9. Mccallum Cm, Comai L, Greene Ea, Henikoff S: targeted screening for induced mutations. *nat biotechnol* 2000, 18:455-457.
10. Redei Gp, Koncz C: classical mutagenesis. In: *methods in arabidopsis research*. Edited by Koncz C, Chua N-H, Schell J. singapore: world scientific; 1992:16-82.
11. Iglesias Dj, Cercos M, Colmenero-Flores Jm, Naranjo Ma, Rios G, Carrera E, Ruiz-Rivero O, Lliso I, Morillon R, Tadeo Fr et al: physiology of citrus fruiting. *Braz J plant physiol,* in press.
12. Tadeo Fr, Cercos M, Colmenero-Flores Jm, Iglesias Dj, Naranjo Ma, Morillon R, Carrera E, Lliso I, Rios G, Ruiz-Rivero O et al: molecular physiology of development and quality of citrus. *Adv bot res,* in press.
13. Forment J, Gadea J, Huerta L, Abizanda L, Agusti J, Alamar S, Alos E, Andres F, Arribas R, Beltran Jp et al.: development of a citrus genome-wide est collection and cDNA microarray as resources for genomic studies. *Plant mol biol* 2005, 57:375-391.
14. Terol J, Conesa A, Colmenero Jm, Cercos M, Tadeo F, Agusti J, Alos E, Andres F, Soler G, Brumos J et al.: analysis of 13000 unique citrus clusters associated with fruit quality, production and salinity tolerance. *Bmc genomics* 2007, 8:31.
15. Bausher M, Shatters R, Chaparro J, Dang P, Hunter W, Niedz R: an expressed sequence tag (est) set from citrus sinensis l. osbeck whole seedlings and the implications of further perennial source investigations. *Plant sci* 2003, 165:415-422.
16. Shimada T, Kita M, Endo T, Fujii H, Ueda T, Moriguchi T, Omura M: expressed sequence tags of ovary tissue cDNA library in citrus unshiu marc. *Plant sci* 2003, 165:167-168.
17. Fujii H, Shimada T, Sugiyama A, Nishikawa F, Endo T, Nakano M, Ikoma Y, Shimizu T, Omura M: profiling ethylene-responsive genes in mature mandarin fruit using a citrus 22k oligoarray. *Plant science* 2007, 173:340-348.
18. Shimada T, Fuiii H, Endo T, Yazaki J, Kishimoto N, Shimbo K, Kikuchi S, Omura M: toward comprehensive expression profiling by microarray analysis in citrus: monitoring the expression profiles of 2213 genes during fruit development. *Plant sci* 2005, 168:1383-1385.
19. Bruggemann E, Handwerger K, Essex C, Storz G: analysis of fast neutron-generated mutants at the arabidopsis thaliana hy4 locus. *Plant J* 1996, 10:755-760.
20. Li X, Song Yj, Century K, Straight S, Ronald P, Dong Xn, Lassner M, Zhang Yl: a fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant J* 2001, 27:235-242.

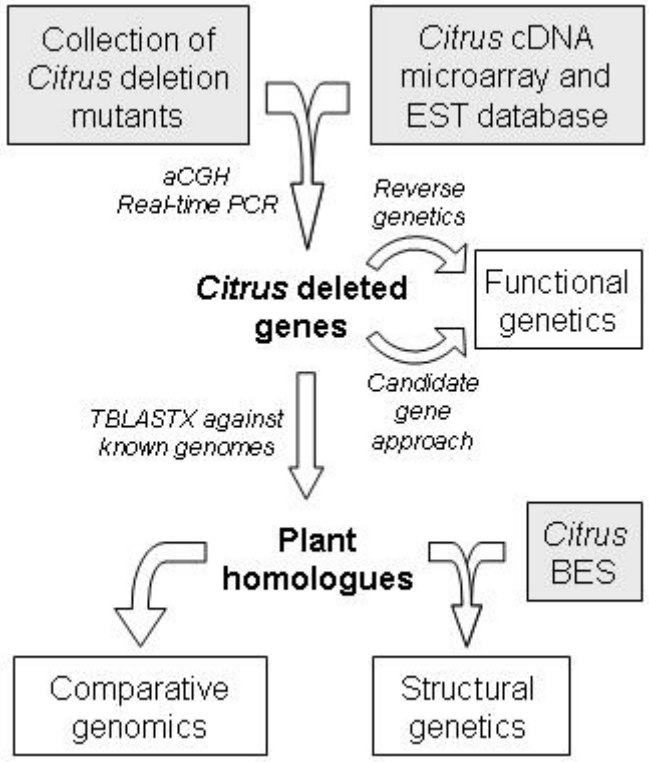
21. Gonzalez-Carranza Zh, Rompa U, Peters Jl, Bhatt Am, Wagstaff C, Stead Ad, Roberts Ja: hawaiian skirt: an f-box gene that regulates organ fusion and growth in arabidopsis. *Plant physiol* 2007, 144:1370-1382.
22. Li X, Lassner M, Zhang Yl: Deleteagene: a fast neutron deletion mutagenesis-based gene knockout system for plants. *compar funct genom* 2002, 3:158-160.
23. Straus D, Ausubel Fm: genomic subtraction for cloning DNA corresponding to deletion mutations. *P natl acad sci usa* 1990, 87:1889-1893.
24. Shirley Bw, Hanley S, Goodman Hm: effects of ionizing-radiation on a plant genome - analysis of 2 arabidopsis transparent-testa mutations. *Plant cell* 1992, 4:333-347.
25. Silverstone Al, Ciampaglio Cn, Sun Tp: the arabidopsis rga gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant cell* 1998, 10:155-169.
26. Gong Jm, Waner Da, Horie T, Li Sl, Horie R, Abid Kb, Schroeder Ji: microarray-based rapid cloning of an ion accumulation deletion mutant in arabidopsis thaliana. *P natl acad sci usa* 2004, 101:15404-15409.
27. Kallioniemi A, Kallioniemi Op, Sudar D, Rutovitz D, Gray Jw, Waldman F, Pinkel D: comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 1992, 258:818-821.
28. Geschwind Dh, Gregg J, Boone K, Karrim J, Pawlikowska-Haddal A, Rao E, Ellison J, Ciccodicola A, D'urso M, Woods R et al.: klinefelter's syndrome as a model of anomalous cerebral laterality: testing gene dosage in the x chromosome pseudoautosomal region using a DNA microarray. *Dev genet* 1998, 23:215-229.
29. Heiskanen Ma, Bittner Ml, Chen Y, Khan J, Adler Ke, Trent Jm, Meltzer Ps: detection of gene amplification by genomic hybridization to cDNA microarrays. *Cancer res* 2000, 60:799-802.
30. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo Wl, Chen C, Zhai Y et al.: high resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *nat genet* 1998, 20:207-211.
31. Pollack Jr, Perou Cm, Alizadeh Aa, Eisen Mb, Pergamenschikov A, Williams Cf, Jeffrey Ss, Botstein D, Brown Po: genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat genet* 1999, 23:41-46.
32. Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P: matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Gene chromosome canc* 1997, 20:399-407.
33. Fan Cz, Vibranovski Md, Chen Y, Long My: a microarray based genomic hybridization method for identification of new genes in plants: case analyses of arabidopsis and oryza. *J integr plant biol* 2007, 49:915-926.
34. Henry Im, Dilkes Bp, Comai L: molecular karyotyping and aneuploidy detection in arabidopsis thaliana using quantitative fluorescent polymerase chain reaction. *Plant J* 2006, 48:307-319.
35. Yang Zn, Ye Xr, Molina J, Roose Ml, Mirkov Te: sequence analysis of a 282-kilobase region surrounding the citrus tristeza virus resistance gene (ctv) locus in poncirus trifoliata l. raf. *Plant physiol* 2003, 131:482-492.
36. lai CwJ, Yu Qy, Hou Sb, Skelton Rl, Jones Mr, Lewis Klt, Murray J, Eustice M, Guan Pz, Agbayani R et al.: analysis of papaya bac end sequences reveals first insights into the organization of a fruit tree genome. *Mol genet genomics* 2006, 276:1-12.
37. Pflieger S, Lefebvre V, Causse M: the candidate gene approach in plant genetics: a review. *Mol breeding* 2001, 7:275-291.
38. Arabidopsis G, I: analysis of the genome sequence of the flowering plant arabidopsis thaliana. *Nature* 2000, 408:796-815.

39. Tuskan Ga, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, rombauts S, Salamov A et al.: the genome of black cottonwood, populus trichocarpa (torr. & gray). *Science* 2006, 313:1596-1604.
40. Jaillon O, Aury Jm, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C et al.: the grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 2007, 449:463-4u5.
41. Wikstrom N, Savolainen V, Chase Mw: evolution of the angiosperms: calibrating the family tree. *P roy soc lond b bio* 2001, 268:2211-2220.
42. Conesa A, Gotz S, Garcia-Gomez Jm, Terol J, Talon M, Robles M: blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005, 21:3674-3676.
43. Deleu W, Gonzalez V, Monfort A, Bendahmane A, Puigdomenech P, Arus P, Garcia-Mas J: structure of two melon regions reveals high microsynteny with sequenced plant species. *Mol genet genomics* 2007, 278:611-622.
44. Kellogg Ea, Bennetzen Jl: the evolution of nuclear genome structure in seed plants. *Am J bot* 2004, 91:1709-1725.
45. Blanc G, Hokamp K, Wolfe Kh: a recent polyploidy superimposed on older large-scale duplications in the arabidopsis genome. *Genome res* 2003, 13:137-144.
46. Ermolaeva Md, Wu M, Eisen Ja, Salzberg Sl: the age of the arabidopsis thaliana genome duplication. *Plant mol biol* 2003, 51:859-866.
47. Bowers Je, Chapman Ba, Rong Jk, Paterson Ah: unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 2003, 422:433-438.
48. Vision Tj, Brown Dg, Tanksley Sd: the origins of genomic duplications in arabidopsis. *Science* 2000, 290:2114-2117.
49. Nielsen E, Akita M, Davilaaponte J, Keegstra K: stable association of chloroplastic precursors with protein translocation complexes that contain proteins from both envelope membranes and a stromal hsp 100 molecular chaperone. *Embo J* 1997, 16:935-946.
50. Desimone M, Weisswichert C, Wagner E, Altenfeld U, Johanningmeier U: immunochemical studies on the clp-protease in chloroplasts: evidence for the formation of a clpc/p complex. *Bot acta* 1997, 110:234-239.
51. Adam Z, Rudella A, Van Wijk Kj: recent advances in the study of clp, ftsh and other proteases located in chloroplasts. *Curr opin plant biol* 2006, 9:234-240.
52. Jackson-Constan D, Akita M, Keegstra K: molecular chaperones involved in chloroplast protein import. *biochim biophys acta* 2001, 1541:102-113.
53. Constan D, Froehlich Je, Rangarajan S, Keegstra K: a stromal hsp100 protein is required for normal chloroplast development and function in arabidopsis. *Plant physiol* 2004, 136:3605-3615.
54. Kovacheva S, Bedard J, Patel R, Dudley P, Twell D, Rios G, Koncz C, Jarvis P: in vivo studies on the roles of tic110, tic40 and hsp93 during chloroplast protein import. *Plant J* 2005, 41:412-428.
55. Sjogren Lle, Macdonald Tm, Sutinen S, Clarke Ak: inactivation of the clpc1 gene encoding a chloroplast hsp100 molecular chaperone causes growth retardation, leaf chlorosis, lower photosynthetic activity, and a specific reduction in photosystem content. *Plant physiol* 2004, 136:4114-4126.
56. Kovacheva S, Bedard J, Wardle A, Patel R, Jarvis P: further in vivo studies on the role of the molecular chaperone, hsp93, in plastid protein import. *Plant J* 2007, 50:364-379.
57. Nakagawara E, Sakuraba Y, Yamasato A, Tanaka R, Tanaka A: clp protease controls chlorophyll b synthesis by regulating the level of chlorophyllide a oxygenase. *Plant J* 2007, 49:800-809.

58. Alos E, Cercos M, Rodrigo Mj, Zacarias L, Talon M: regulation of color break in citrus fruits. Changes in pigment profiling and gene expression induced by gibberellins and nitrate, two ripening retardants. *J agric food chem* 2006, 54:4888-4895.
59. Lieu Pt, Jozsi P, Gilles P, Peterson T: development of a DNA-labeling system for array-based comparative genomic hybridization. *J biomol tech* 2005, 16:104-111.
60. Price Ts, Regan R, Mott R, Hedman A, Honey B, Daniels Rj, Smith L, Greenfield A, Tiganescu A, Buckle V et al.: sw-array: a dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data. *Nucleic acids res* 2005, 33:3455-3464.
61. Brown Po. Preparation of fluorescent DNA probe from human mRNA or total RNA using direct incorporation [*http://cmgm.stanford.edu/pbrown/protocols/5_hyb_human.html*].
62. Giuntini E, Mengoni A, De Fc, Cavalieri D, Ubin-Horth N, Landry Cr, Becker A, Bazzicalupo M: large-scale genetic variation of the symbiosis-required megaplasmid psyma revealed by comparative genomic analysis of sinorhizobium meliloti natural strains. *Bmc genomics* 2005, 6:158.
63. R statistical computing software [*http://www.r-project.org*].
64. Smyth Gk: linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat appl genet mol biol* 2004, 3:article3.
65. Gentleman Rc, Carey Vj, Bates Dm, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J et al: bioconductor: open software development for computational biology and bioinformatics. *Genome biol* 2004, 5:r80.
66. Arrayexpress database [*http://www.ebi.ac.uk/microarray-as/aer/?#ae-main[0*]].
67. Ncbi blast home [*http://www.ncbi.nlm.nih.gov/blast/blast.cgi*].
68. *Populus trichocarpa* genome v1.1 blast page [*http://genome.jgi-psf.org/cgi-bin/runalignment?db=poptr1_1&advanced=1*].
69. Grape genome blast server [*http://www.genoscope.cns.fr/cgi-bin/blast_server/projet_ml/blast.pl*].
70. Moran R: formulae for determination of chlorophyllous pigments extracted with n,n-dimethylformamide. *Plant physiol* 1982, 69:1376-1381.

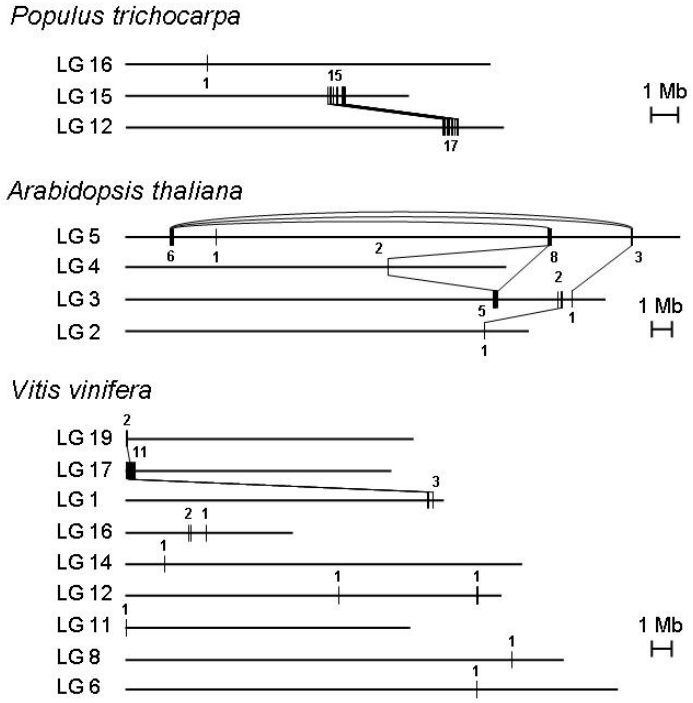
FIGURES

Figure 1 - Schematic guidelines for the characterization of hemizygous deletion *Citrus* mutants



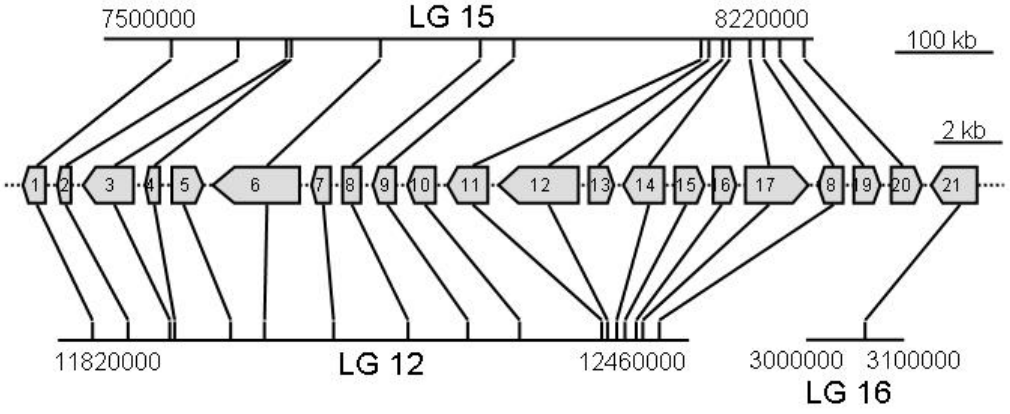
Arrows indicate successive steps of mutant characterization. Plant material and genomic resources are highlighted on grey boxes while gained knowledge on genetics and genomics are highlighted on white rectangles. Genes are shown in bold and methods and approaches in italics.

Figure 2 - Chromosome mapping of poplar, *Arabidopsis* and grapevine homologues of the 39B3 *Citrus* deleted genes



The first TBLASTX hit for each *Citrus* deleted gene with an E value cut-off < 10^{-5} is represented on linkage groups (LG) from *Populus trichocarpa*, *Arabidopsis thaliana* and *Vitis vinifera*. Homologues of *Citrus* genes were grouped into clusters in each species when the distance between them was shorter than 250 kb. Second and third hits are only represented when they are located in a previously identified cluster, and in this case are linked to the first hit by a line. The value on each cluster indicates the hit number of the cluster.

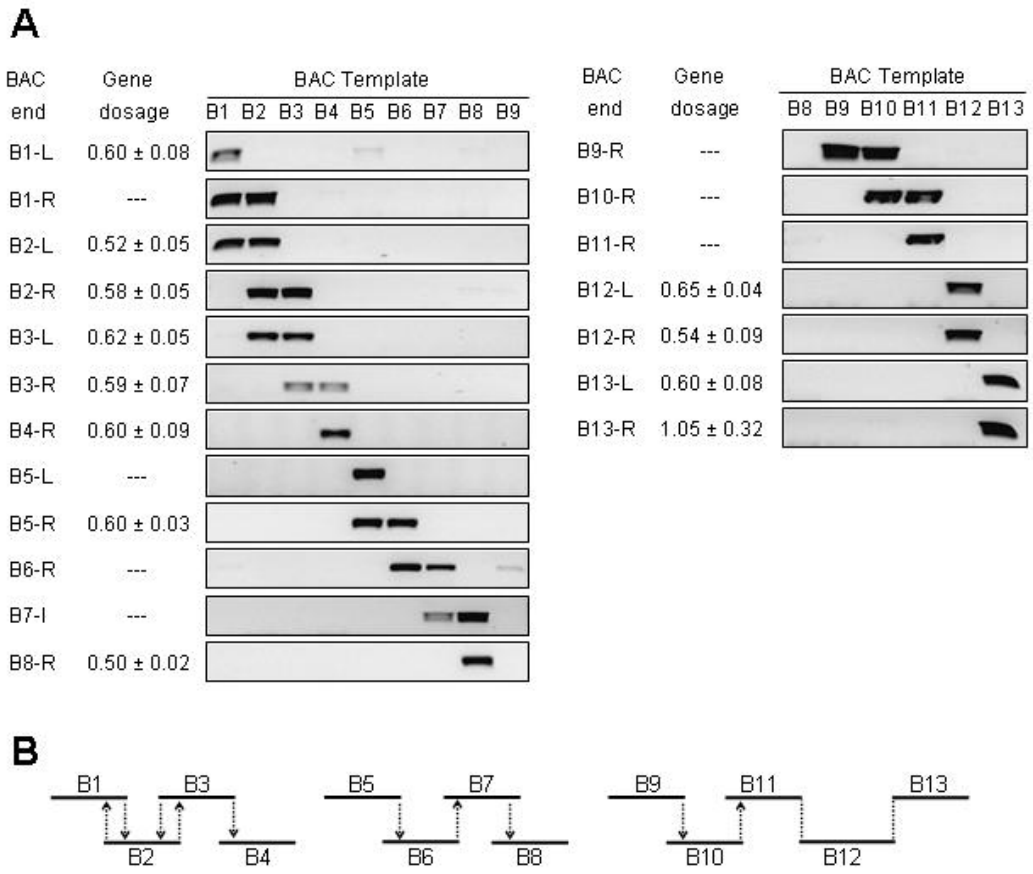
Figure 3 - Gene composition of the *Citrus* 39B3 deletion inferred from poplar homologous regions



The 39B3 deleted *Citrus* genes are arranged in the centre of the figure in the order inferred from the position of their *Populus* homologues found in linkage groups 12, 15 and 16. Genes

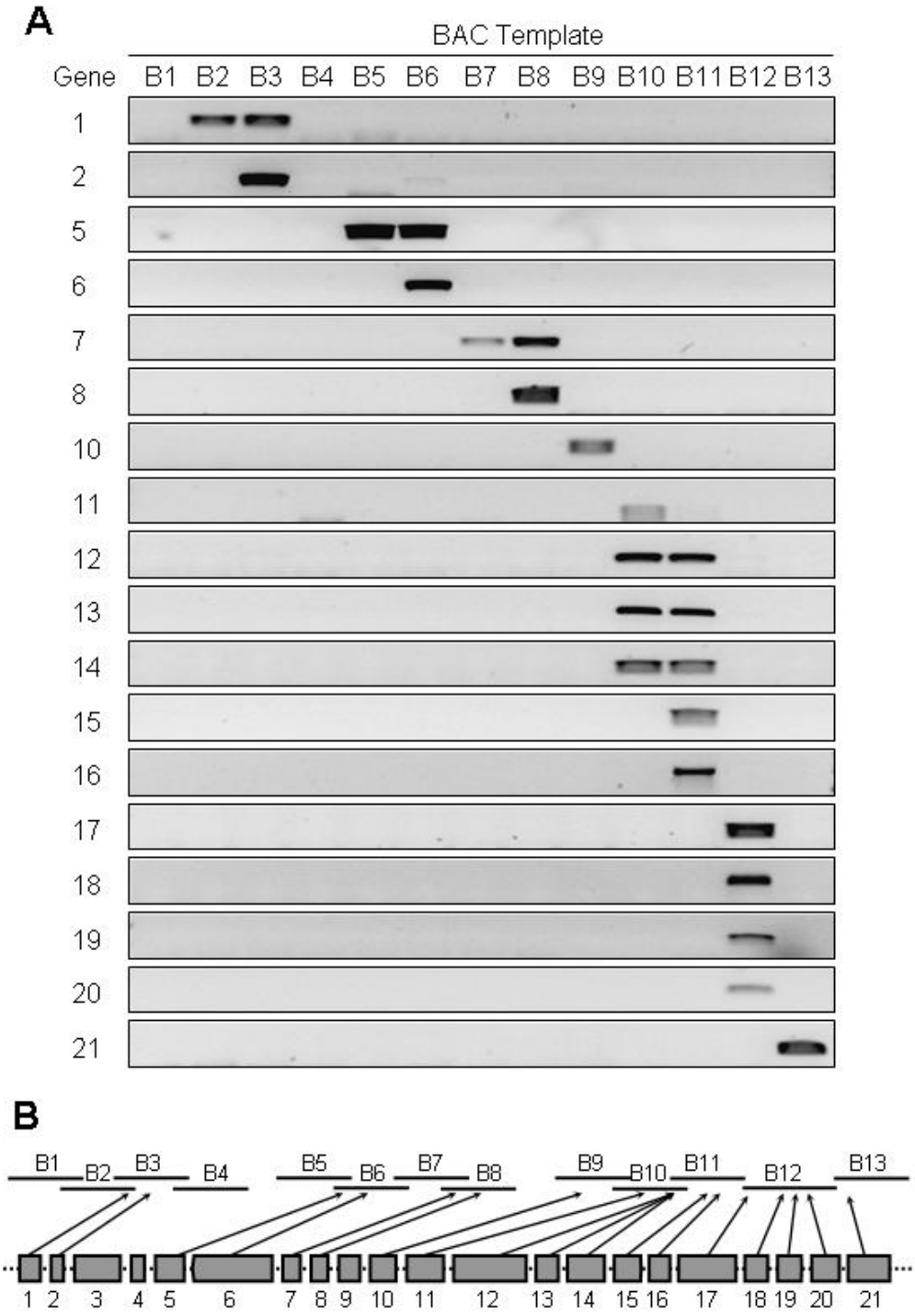
are numbered following this order. Strand sense deduced from poplar counterparts is indicated by an arrow.

Figure 4 - Local physical mapping of the *Citrus* 39B3 deletion



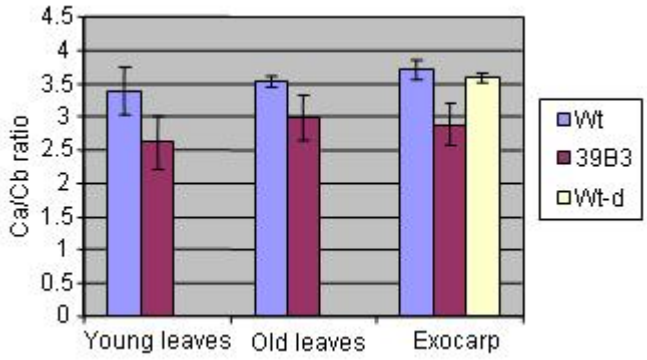
(A) Electrophoretic analysis of PCR products showing overlapping BACs. Purified BAC templates are distributed horizontally and divided in two panels. Primer pairs were designed from BAC end sequences containing non-repetitive DNA and named with the number of the BAC plus "-L" for left end and "-R" for right end according to the drawing orientation. B7-I primers amplify an internal sequence from B7 instead of an end. Gene dosage measurements for some of the primer pairs in the 39B3 genotype are shown on the left side of the electrophoretic images. (B) Physical map of the *Citrus* 39B3 deletion. Horizontal lines represent BACs, which are numbered from left to right. Vertical arrows show overlapping as inferred from PCR reactions and the head of the arrow indicates the BAC template. The vertical lines without arrow show connection of B11 with B12 by sequence of unigene aCL4690Contig1 and B12 with B13 by aCL1915Contig2 instead of a PCR reaction.

Figure 5 - Gene arrangement on the physical map of the *Citrus* 39B3 deletion



(A) Primer pairs designed for the putative deleted genes included in the 39B3 deletion (Table 2) were utilized in PCR reactions on the BAC templates shown in Table 3. Genes are numbered and arranged vertically, on the left side of the electrophoretic image, and BAC templates are listed horizontally. (B) *Citrus* genes included in the 39B3 deletion and arranged as drawn in Figure 3 but without indication of strand sense, are connected with arrows to the deletion physical map according to PCR results.

Figure 6 - Chlorophyll *a*/*b* ratio in green tissues from 39B3 mutant and "wild type" cultivar of *Citrus clementina*



Chlorophyll *a* and *b* content was measured in young and old leaves and fruit peel exocarp from "wild type" clementine cultivar (Wt) and 39B3 mutant (39B3). Measurements were also taken from exocarps of a normal "wild type" clementine tree showing fruit colour delay (Wt-d) due to altered environmental conditions. The relative content of chlorophyll *a* to chlorophyll *b* is represented as the Ca/Cb ratio. Data are average of 3 (exocarp) or 5 (leaves) independent determinations and error bars show standard deviation.

TABLES

Table 1 - Gene dosage measurement of deleted genes in 39B3 and 39E7 *Citrus* mutants

Accession numbers and the corresponding assembled unigenes for several candidate ESTs are shown. Array-CGH data are signal ratios of mutant with respect to wild type samples for each single EST. Gene dosage was calculated through real-time PCR on 39B3 and 39E7 genomic templates. A horizontal line separates array-CGH hits from 39B3 mutant (up) from 39E7 mutant (down).

Unigene	EST accession number (GenBank)	Array-CGH		Real-time PCR gene dosage	
		Mutant/wt ratio	P value	Template 39B3	Template 39E7
aCL4690Contig1	CX295702	0.59	0.10	0.56 ± 0.02	0.99 ± 0.09
aCL1915Contig2	DY300024	0.62	0.12	0.60 ± 0.08	0.96 ± 0.04
aCL3317Contig1	DY265056	0.62	0.10	0.60 ± 0.03	0.98 ± 0.03
aC20009H03SK_c	CX308429	0.63	0.10	0.50 ± 0.02	0.98 ± 0.11
aCL766Contig1	CX288964	0.65	0.13	0.56 ± 0.06	0.96 ± 0.11
aCL7097Contig1	FC868864	0.65	0.12	0.59 ± 0.04	1.15 ± 0.07
aCL2087Contig2	DY300006	0.57	0.05	1.05 ± 0.13	0.59 ± 0.11
aCL6684Contig1	DY265447	0.60	0.05	1.04 ± 0.12	0.59 ± 0.12
aCL6641Contig1	FC930062	0.62	0.05	1.14 ± 0.19	0.58 ± 0.08
aCL3902Contig1	DY267778	0.62	0.05	1.12 ± 0.10	0.64 ± 0.04
aC05139C12SK_c	CX296347	0.66	0.05	1.24 ± 0.22	0.61 ± 0.05
aC01019E12SK_c	CX288357	0.66	0.05	1.04 ± 0.07	0.63 ± 0.11

Table 2 - Gene components of the *Citrus* 39B3 deletion

Unigenes are numbered according to the ordered position of their homologues on the poplar genome. The accession numbers correspond to 39B3 candidate ESTs. The coding frame of *Populus* homologues follow the proposed strand (+) or the complementary reverse one (-).

Nº	Citrus unigene	EST accession number (GenBank)	Strand	Similarity
1	aC01006D04SK_c	CX287243	-	Hypothetical protein
2	aC20006C06SK_c	CX308114	-	Ubiquitin conjugating enzyme
3	aCL3991Contig1	DY278065	-	Sterile alpha motif (SAM) domain-containing protein
4	aC18005F10Rv_c	CX305429	-	Sialyltransferase-like protein
5	aCL3317Contig1	DY265056	+	Hypothetical protein
6	aCL766Contig1	CX288964	-	ATP-dependent Clp protease, clpC homolog
7	aC01012C02SK_c	CX287682	-	Alpha-mannosidase
8	aC20009H03SK_c	CX308429	?	Mei2-like protein
9	aC16014F08SK_c	CX304691	-	Putative pol polyprotein
10	aCL6210Contig1	DY282423	-	Hypothetical protein
11	aCL8592Contig1	DY267639	-	Tudor domain-containing protein
12	aCL1065Contig1	DY282340	-	Putative amidase
13	aC32108G01EF_c	FC921733	+	Hypothetical protein
14	aCL503Contig1	CX292510	-	Respiratory burst oxidase homolog
15	aCL6269Contig1	DY263746	+	FHA domain-containing protein
16	aCL7097Contig1	FC868864	+	Putative pentatricopeptide (PPR) repeat protein
17	aCL4690Contig1	CX295702	+	ERD1 protein, chloroplast precursor
18	aCL8011Contig1	FC923875	-	Fe-superoxide dismutase
19	aIC0AAA60DF12RM1_c	DY284274	+	Poly(A)-binding protein II-like
20	aCL1848Contig1	FC875470	+	Hypothetical protein
21	aCL1915Contig2	DY300024	-	Tubulin-specific chaperone C-related

Table 3 - Listing of BACs included in the *Citrus* 39B3 deletion

BACs are numbered according to the ordered position in the deletion from B1 to B13. BES are named with the number of the BAC plus "-L" for left end and "-R" for right end according to the drawing orientation in Figure 4B. Nhf: no hits found.

Nº	BAC	Ends	BES ID (GenBank)	BLASTX against plant proteins	E value
B1	CCL021E18	B1-L	ET070583	Nhf	--
		B1-R	ET070584	gi|91805627| hypothetical protein	9e-24
B2	CCL011O24	B2-L	ET086992	Nhf	--
		B2-R	ET086991	gi|7576215| hypothetical protein	7e-47
B3	CCER1037B12	B3-L	ET077105	gi|7576215| hypothetical protein	3e-90
		B3-R	ET077106	Nhf	--
B4	CCER1032N17	B4-L	ET101817	gi|25411577| probable retroelement pol polyprotein	2e-06
		B4-R	ET101816	Nhf	--
B5	CCL011N15	B5-L	ET087145	gi|6469119| mitochondrial phosphate transporter	5e-56
		B5-R	ET087144	Nhf	--
B6	CCER1045A09	B6-L	ET077286	gi|92895029| Polynucleotidyl transferase (retrotransposon protein)	8e-63
		B6-R	ET077285	gi|30027167| auxin response factor-like protein	6e-85
B7	CCH3037D01	B7-L	ET112059	gi|87240692| Helix-loop-helix DNA-binding	1e-21
B8	CCER1005N09	B8-L	ET079746	Nhf	--
		B8-R	ET079745	gi|79331867| AML1; RNA binding / nucleic acid binding	2e-09
B9	CCH3005L04	B9-L	ET081228	gi|33113977| putative copia-type pol polyprotein	2e-85
		B9-R	ET081227	gi|51968598| peroxisomal Ca-dependent solute carrier-like protein	2e-21
B10	CCER1033B14	B10-L	ET102435	gi|51968598| peroxisomal Ca-dependent solute carrier-like protein	2e-37
		B10-R	ET102434	Nhf	--
B11	CCL011K21	B11-L	ET086761	gi|25402907| protein F5M15.26 (retrotransposon protein)	4e-78
		B11-R	ET086760	gi|14334878| putative ATP-dependent Clp protease ClpD	4e-57
B12	CCER1019D04	B12-L	ET098996	gi|14334878| putative ATP-dependent Clp protease ClpD	4e-35
		B12-R	ET098995	gi|6729532| putative protein	3e-28
B13	CCL032E17	B13-L	ET094320	gi|6729532| putative protein	9e-34
		B13-R	ET094321	Nhf	--

Table 4 - Total chlorophyll content in green tissues from 39B3 mutant and "wild type" cultivar of *Citrus clementina*

Total chlorophyll content (mg/g fresh weight) was measured in young and old leaves and fruit peel exocarp from samples shown in Figure 6. Data are average of 3 (exocarp) or 5 (leaves) independent determinations. Standard deviation is shown.

	Young leaves	Old leaves	Exocarp
Wt	0.76 ± 0.25	2.23 ± 0.19	0.15 ± 0.05
39B3	0.71 ± 0.24	2.49 ± 0.17	0.48 ± 0.01
Wt-d	--	--	0.29 ± 0.01

ADDITIONAL FILES

Additional file 1

Title: Primers employed in real-time and standard PCR experiments.

Format: XLS (Microsoft Excel)

[End of document]