



OPEN

# Maximum-likelihood method identifies meiotic restitution mechanism from heterozygosity transmission of centromeric loci: application in citrus

SUBJECT AREAS:  
PLANT BREEDING  
GENETIC MARKERS  
POLYPOIDY IN PLANTS

Received  
30 November 2014

Accepted  
13 March 2015

Published  
20 April 2015

José Cuenca<sup>1\*</sup>, Pablo Aleza<sup>1\*</sup>, José Juárez<sup>1</sup>, Andrés García-Lor<sup>1</sup>, Yann Froelicher<sup>2</sup>, Luis Navarro<sup>1</sup> & Patrick Ollitrault<sup>1,2</sup>

<sup>1</sup>Crop Protection and Biotechnology Center. Instituto Valenciano de Investigaciones Agrarias (IVIA), <sup>2</sup>Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD).

Correspondence and requests for materials should be addressed to P.O. (ollitrault@cirad.fr) or L.N. (lnavarro@via.es)

\* These authors contributed equally to this work.

Polyplodisation is a key source of diversification and speciation in plants. Most researchers consider sexual polyplodisation leading to unreduced gamete as its main origin. Unreduced gametes are useful in several crop breeding schemes. Their formation mechanism, i.e., First-Division Restitution (FDR) or Second-Division Restitution (SDR), greatly impacts the gametic and population structures and, therefore, the breeding efficiency. Previous methods to identify the underlying mechanism required the analysis of a large set of markers over large progeny. This work develops a new maximum-likelihood method to identify the unreduced gamete formation mechanism both at the population and individual levels using independent centromeric markers. Knowledge of marker-centromere distances greatly improves the statistical power of the comparison between the SDR and FDR hypotheses. Simulating data demonstrated the importance of selecting markers very close to the centromere to obtain significant conclusions at individual level. This new method was used to identify the meiotic restitution mechanism in nineteen mandarin genotypes used as female parents in triploid citrus breeding. SDR was identified for 85.3% of 543 triploid hybrids and FDR for 0.6%. No significant conclusions were obtained for 14.1% of the hybrids. At population level SDR was the predominant mechanisms for the 19 parental mandarins.

Polyplodisation is a key source of species diversification and speciation in plants<sup>1–3</sup> and may occur by somatic chromosome doubling (somatic polyplodisation) or sexually through gametic nonreduction (sexual polyplodisation)<sup>4</sup>. Currently, most researchers consider sexual polyplodisation, leading to unreduced gamete, to be the main mechanism of polyplodisation in plants<sup>1,5,6</sup>.

Meiotic aberrations related to spindle formation, spindle function and cytokinesis can lead to unreduced gamete formation in plants. Up to seven major mechanisms of  $2n$  gamete formation have been cytogenetically characterised: premeiotic doubling, first-division restitution (FDR), chromosome replication during the meiotic interphase, second-division restitution (SDR), postmeiotic doubling, indeterminate meiotic restitution, and apospory<sup>7–9</sup>. However, FDR and SDR are the predominant mechanisms of  $2n$  gamete formation<sup>4</sup>. Failure of the first (FDR) or second (SDR) divisions leads to the formation of restitution nuclei with an unreduced chromosome number. A FDR  $2n$  gamete contains non-sister chromatids, while a SDR  $2n$  gamete contains two sister chromatids<sup>5,10,11</sup>.

The use of unreduced gametes in plant breeding<sup>9,12</sup>, resulting in the establishment of sexual polyplods, is useful for improvement of crops such as lily<sup>8,13,14</sup>, maize<sup>15</sup>, potato<sup>16–18</sup>, rose<sup>19</sup>, rye<sup>20</sup>, alfalfa<sup>21,22</sup>, banana<sup>23,24</sup> and citrus<sup>25–29</sup>.

Diploidy is the general rule in *Citrus* and its related genera, with a basic chromosome number  $x = 9$ <sup>30</sup>. However, triploid breeding has become an important strategic tool in the development of new seedless citrus commercial varieties<sup>25–29</sup>. Indeed, seedlessness is one of the most important economic traits related to fruit quality for fresh-fruit marketing of mandarins<sup>26,27,31</sup>. Very large triploid progenies have been obtained from  $2x \times 2x$  crosses<sup>32</sup> and several cultivars patented<sup>28,29</sup>.

Cytogenetic studies<sup>33</sup> showed that triploid embryos are associated with pentaploid endosperm, indicating that triploid hybrids result from the fertilisation of unreduced ovules by normal haploid pollen. According to the genotype, the frequency of duplication in the female gametes can range from below 1% to over 20%. Esen *et al.*<sup>34</sup>



proposed that, citrus,  $2n$  eggs result from the abortion of the second meiotic division in the megaspore. This hypothesis was corroborated by molecular marker analysis for clementine (*Citrus clementina* Hort. ex Tan.)<sup>35,36</sup>. The method proposed by Cuenca *et al.*<sup>37</sup> was successfully applied in populations of  $2n$  ovules of ‘Fortune’ mandarin and ‘Nules’ clementine, and it was concluded that SDR was the main restitution mechanism and that partial chromosome interference occurs<sup>36,37</sup>. By contrast, Chen *et al.*<sup>38</sup> proposed that  $2n$  eggs of sweet orange (*C. sinensis* (L.) Osb.) resulted from first meiotic division restitution.

The origin of  $2n$  gamete formation greatly impacts the gametic structures and, therefore, the polyploid populations and the efficiency of breeding strategies. Under FDR, non-sister chromatids retain parental heterozygosity from the centromere to the first crossover point. Under SDR, the two sister chromatids are homozygous between the centromere and the first crossover point (Figure 1<sup>5</sup>). As a consequence, several studies based on genetic markers indicate that FDR gametes transmit 70–80% of the parental heterozygosity, but SDR gametes transmit only 30–40%<sup>9,19,39–42</sup>. Thus, a tighter distribution is expected in FDR-derived populations than in SDR ones because a higher percentage of the parental genome is transferred intact, resulting in a more uniform gamete production<sup>43</sup>. Therefore, insights into the meiotic nuclear restitution mechanisms that produce unreduced gametes are crucial for the optimisation of breeding strategies based on sexual polyploidisation<sup>44</sup>.

The identification of the mechanisms driving the formation of  $2n$  gametes is complex. However, the use of cytological or marker analysis on polyploid progeny provide accurate or additional information on these mechanisms<sup>9,19,45</sup>. Molecular cytological approaches have been used successfully, including the unequivocal identification of genomes and recombinant segments in the sexual polyploid progenies<sup>11,14,45–47</sup>. Molecular marker analysis is also a valuable tool for the estimation of parental heterozygosity restitution (HR) through diploid gametes to polyploid progenies and, therefore, to identify the mechanisms underlying unreduced gamete formation<sup>22,35,38,39,41,48,49</sup>. Several previously developed methods are based on the analysis of HR rates for randomly chosen unmapped markers<sup>38</sup>. These methods require the analysis of a large set of molecular markers to encounter, by chance, the loci with HR lower than 50% that are only found under SDR<sup>50</sup>. However, when HR over 50% is observed for all loci, no

definitive conclusion can be reached without a prior knowledge of their location relative to a centromere. Significant FDR conclusions are therefore difficult to obtain with such non-mapped markers. Half-tetrad analysis (HTA,<sup>51</sup> based on multiple linked loci, is a powerful method for mapping centromeres or for determining the mode(s) of  $2n$  gamete formation. Tavoletti *et al.*<sup>10</sup> developed a multi-locus maximum-likelihood method of HTA that permits the estimation of both the relative frequencies of FDR and SDR  $2n$  gametes and the centromere location within a linkage group without relying on previously identified centromeric markers. The models described therein are all based on population analysis and suppose complete chiasma interference.

Cuenca *et al.*<sup>37</sup> proposed an approach that takes into account different models of chromosome interference (i.e., no interference, partial interference or complete chiasma interference) when testing for FDR and SDR, and for mapping centromeres to linkage groups. This approach is based on functions of heterozygosity restitution (HR) at the population level along a chromosome in relation to locus-centromere distance ( $d$ )<sup>52</sup>. Indeed, under FDR or SDR, HR is a direct function of the crossing over frequency between the considered locus and the centromere. It is, therefore, possible to implement the function ( $HR = f(d)$ ) according to the FDR and SDR hypotheses while also taking into account different models of chromosome interference (Figure 2).

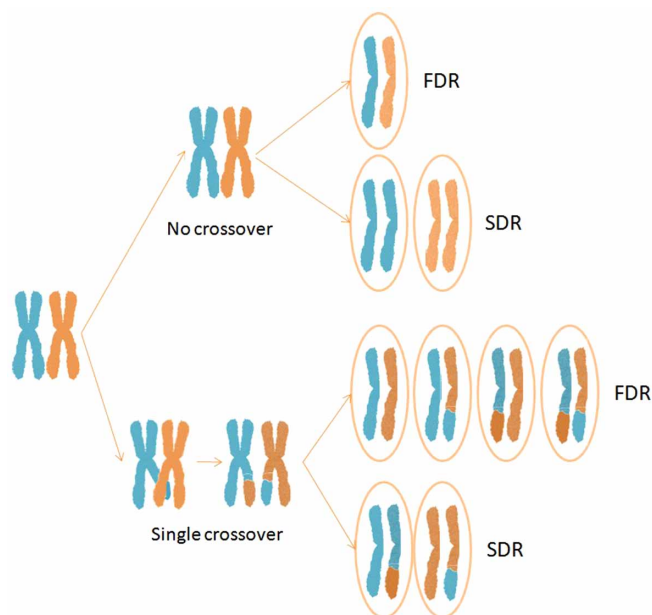
In the present work, we propose a maximum-likelihood approach to test the SDR/FDR mechanism based on the HR of unlinked markers located close to the centromere of different chromosomes. This approach can be applied at the individual or population level. We simulated  $2n$  gamete populations arising from FDR or SDR. This enabled us to identify the number of independent markers necessary to test in order to draw significant conclusions at the individual level in relation to marker/centromere distances, as well as the minimum population size necessary to be able to draw significant conclusions when analysing a defined number of unlinked markers.

As a concrete application this new method has been used for investigating the unreduced gamete formation in citrus. Taking advantage of the centromere locations<sup>36</sup> within the nine linkage groups of the clementine reference genetic map<sup>53</sup>, we selected centromeric markers and used the proposed maximum-likelihood method to (i) check the potential variability of origin between individuals for two genotypes in which SDR was proposed to be the predominant polyploidisation mechanism as determined by population analysis (‘Fortune’ mandarin<sup>37</sup>, and clementine<sup>35,36</sup>, and (ii) shed light on the mechanism leading to unreduced gamete formation in a range of mandarin genotypes used as female parents in  $2x \times 2x$  triploid breeding programs.

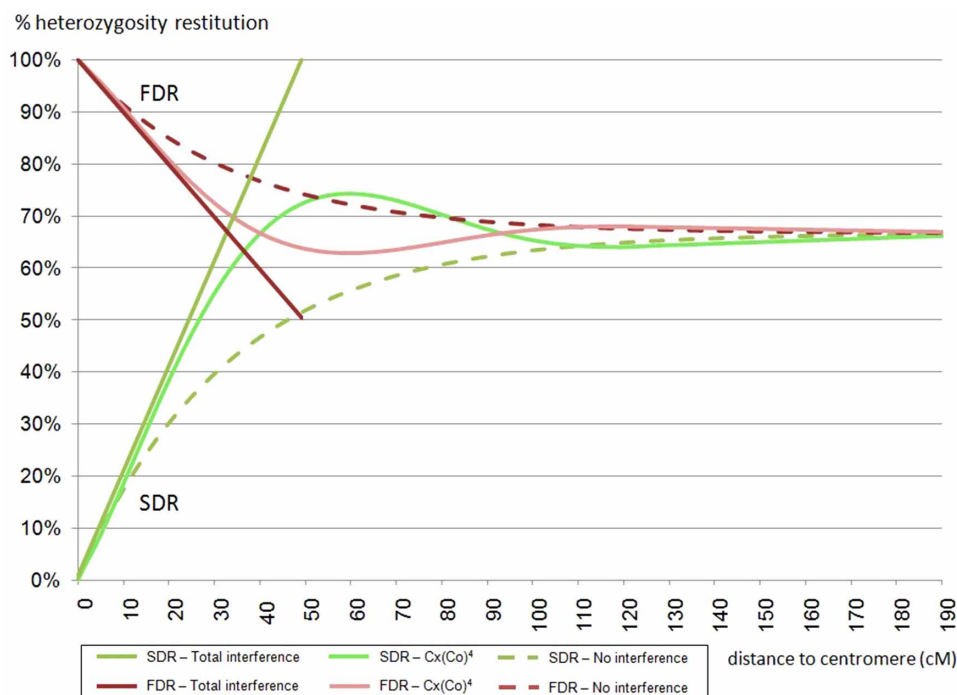
## Results

**Statistical method for the identification of meiotic restitution mechanism.** Identification of the restitution mechanism at an individual level. For loci heterozygous for the parent producing the  $2n$  gamete, the probabilities of a  $2n$  gamete being heterozygous or homozygous as a consequence of FDR or SDR mechanisms are direct functions of the marker-centromere distance.

To estimate such probabilities, the function relating HR rate and locus-centromere distance<sup>37</sup>, derived from the  $Cx(Co)^4$  partial chiasma interference model developed by Zhao and Speed<sup>52</sup> and Foss *et al.*<sup>54</sup>, could be used. Indeed, Cuenca *et al.*<sup>37</sup> showed that this model fit better to ‘Fortune’ mandarin data (SDR mechanism) than total or no interference models. However, since selected markers are located close to centromeres (as explained above), for our data, the  $Cx(CO)^4$  model and the total interference model are equivalent (Figure 2). To simplify mathematical calculations of probabilities, the total interference model was used. Marker-centromere distances ( $d$ ) in Morgan units were estimated from the centromere locations<sup>36</sup> in the clementine reference genetic map<sup>53</sup>.



**Figure 1** | Half tetrads resulting from no crossover and single crossover events under FDR and SDR mechanisms of unreduced gamete formation.



**Figure 2** | Rate (percentage) of heterozygosity restitution in the unreduced gametes under FDR and SDR mechanisms in function of the locus-centromere distance considering the total interference model, the no interference model and the  $Cx(Co)^4$  partial interference model (adapted from Cuenca *et al.*<sup>37</sup>).

The probabilities of a marker being inherited as heterozygous under the SDR [ $P_{SDR}(M_{He})$ ] or FDR [ $P_{FDR}(M_{He})$ ] mechanisms were directly estimated from the total interference model functions as  $P_{SDR}(M_{He}) = 2d$  and  $P_{FDR}(M_{He}) = (1 - d)$ . The probabilities of a marker being inherited as homozygous under SDR and FDR were estimated as  $P_{SDR}(M_{Ho}) = (1 - 2d)$  and  $P_{FDR}(M_{Ho}) = d$ , respectively.

Therefore, the LOD values used to compare the probabilities of a heterozygous or a homozygous diploid gamete occurring at a locus, under the two models (SDR/FDR), were calculated respectively as:

$$LOD(M_{He}) = \log[P_{SDR}(M_{He})/P_{FDR}(M_{He})] = \log(2d/(1 - d)) \quad (1)$$

and

$$LOD(M_{Ho}) = \log[P_{SDR}(M_{Ho})/P_{FDR}(M_{Ho})] = \log((1 - 2d)/d) \quad (2)$$

For each restitution model, the probability of a single unreduced gamete [ $P(G)$ ] presenting the observed allelic configuration for  $i$  unlinked markers ( $M_i$ ) is the product of the probabilities of the observed genotype at each locus,  $P(G) = \pi P_{M_i}$ , and therefore the LOD value to compare the SDR/FDR models is the sum of the LOD at each locus,

$$LOD(G) = \sum LOD_{M_i}, \quad (3)$$

where  $P_{M_i}$  and  $LOD_{M_i}$  are the probability and the LOD value of the observed genotype at the locus  $i$ , respectively.

As an example, if three unlinked loci ( $M_1$ ,  $M_2$  and  $M_3$ ) were heterozygous, homozygous and homozygous, respectively, the probabilities of observing such gametes [ $P(G)$ ; ( $M_{1He}-M_{2Ho}-M_{3Ho}$ )] are, respectively,

$$P_{SDR}(G) = 2d_1 \times (1 - 2d_2) \times (1 - 2d_3) \quad (4)$$

under SDR and

$$P_{FDR}(G) = (1 - d_1) \times d_2 \times d_3 \quad (5)$$

under FDR

The LOD value used to compare the probabilities of SDR/FDR models is

$$LOD(G) = \log(2d_1/(1 - d_1)) + \log((1 - 2d_2)/d_2) + \log((1 - 2d_3)/d_3) \quad (6)$$

where  $d_i$  is the distance from the locus  $i$  to its centromere.

LOD scores greater than 3 (the probability of the observed gamete is more than 1000-fold higher under the SDR model than the FDR one; LOD3) or greater than 2 (the probability of the observed gamete is more than 100-fold higher under the SDR model than the FDR one; LOD2) were considered as thresholds indicating that SDR was the mechanism involved in the single unreduced gamete formation, whereas LODs below  $-3$  (or  $-2$ ) indicate that FDR was the underlying mechanism; for LOD scores between  $-3$  and  $3$  (or between  $-2$  and  $2$ ), we considered that the mechanism could not be determined significantly.

*Identification of the restitution mechanism at population level.*

Considering an infinite population of  $2n$  gametes and a single locus, the probability of observing a sample of gametes [ $P(\text{Pop})$ ] with  $j$  heterozygous and  $k$  homozygous individuals under the SDR and FDR model are, respectively:

$$P_{SDR}(\text{Pop}) = C \times P_{SDR}(M_{He})^j \times P_{SDR}(M_{Ho})^k = C \times (2d)^j \times (1 - 2d)^k \quad (7)$$

$$P_{FDR}(\text{Pop}) = C \times P_{FDR}(M_{He})^j \times P_{FDR}(M_{Ho})^k = C \times (1 - d)^j \times (d)^k \quad (8)$$

where  $C$  is a combinatory coefficient constant for the observed sample. Therefore,

$$LOD(\text{Pop}) = \frac{(2d)^j \times (1 - 2d)^k}{(1 - d)^j \times (d)^k} \quad (9)$$

If  $i$  independent loci are analysed, the probabilities of the observed sample of gametes occurring under the SDR [ $P_{SDR}(\text{Pop})$ ] or FDR [ $P_{FDR}(\text{Pop})$ ] models are the products of the probabilities of the observed sample at each locus





$$P_{\text{SDR}}(\text{Pop}) = \pi C_i \times P_{\text{SDR}}(M_{\text{IHe}})^{ji} \times P_{\text{SDR}}(M_{\text{IH0}})^{ki} = \pi C_i \times (2d_i)^{ji} \times (1 - 2d_i)^{ki} \quad (10)$$

$$P_{\text{FDR}}(\text{Pop}) = \pi C_i \times P_{\text{FDR}}(M_{\text{IHe}})^{ji} \times P_{\text{FDR}}(M_{\text{IH0}})^{ki} = \pi C_i \times (1 - d_i)^{ji} \times (d_i)^{ki} \quad (11)$$

and therefore,

$$\text{LOD}(\text{Pop}) = \sum \frac{(2d_i)^{ji} \times (1 - 2d_i)^{ki}}{(1 - d_i)^{ji} \times (d_i)^{ki}} \quad (12)$$

where  $P(M_{\text{IHe}})$ ,  $P(M_{\text{IH0}})$ ,  $ji$ ,  $ki$ , and  $d_i$  are, respectively, the probability of heterozygous individuals, probability of homozygous individuals, number of heterozygous individuals, number of homozygous individuals and distance to centromere for the locus  $i$ .

At the population level, LOD scores greater than 3 were considered to indicate that SDR was the mechanism involved in unreduced gamete formation, whereas LODs below  $-3$  indicated that FDR was the underlying mechanism. When LOD scores between  $-3$  and  $3$  were obtained, we considered that the mechanism could not be significantly determined.

**Studies to check the power of the method.** We assessed the power of our method using simulated samples of diploid gametes arising from either the FDR or SDR mechanisms. From a theoretical infinite population with heterozygous and homozygous genotype frequencies directly defined by the considered locus-centromere distances [ $P_{\text{FDR}}(M_{\text{IHe}}) = (1 - d)$ ;  $P_{\text{FDR}}(M_{\text{IH0}}) = d$ ;  $P_{\text{SDR}}(M_{\text{IHe}}) = 2d$ ;  $P_{\text{SDR}}(M_{\text{IH0}}) = (1 - 2d)$ ] as explained above, individual gametes with information for nine markers (the haploid number of chromosome in *Citrus*) were randomly generated. Then, the LOD values of these gametes were calculated as described above. We estimated the proportion of gametes with significant solutions at LOD3 (LOD value  $> 3$  or  $< -3$ ) and LOD2 (LOD value  $> 2$  or  $< -2$ ) when analysing 1–9 markers mapped at the same centromere distance, but in different chromosomes, and for distances ranging from 0 to 20 cM.

Gamete populations were also generated in order to estimate the theoretical number of hybrids that would need to be analysed to obtain significant conclusions for a mechanism, depending on the number of markers used and the marker-centromere distances. From each theoretical population (FDR and SDR populations), 200 replicates of populations (with 1–100 gametes/population) were randomly generated. The generated population LODs were calculated as described above and, for each number of considered markers at a given centromere distance, we identified the minimum number of gametes needed in order to be able to reach a true significant conclusion for at least 99% of the generated populations (99% of replicates with  $\text{LOD} > 3$  for SDR or  $\text{LOD} < -3$  for FDR).

From 1000 randomly selected gametes with nine independent markers (at the same distance from their respective centromere) from a theoretical SDR and FDR infinite population, we analysed the percentage of replicates with significant LOD value (i.e., LOD3 and LOD2) at a given distance considering the data from 1–9 markers.

Curves corresponding to a significant true answer are shown in Figure 3. All curves display a vertical drop to 0, corresponding to the distance when the maximum theoretical LOD score (when all considered markers are in the most favourable combination for the model) is below the considered threshold. Compared with LOD3, the LOD2 threshold allows maintenance of the progressive decrease of the significant answer with increasing distance. As distance increases, more markers are needed to maintain a high level of significance.

At LOD3, the usefulness of only one marker is null for both the SDR (Figure 3a) and the FDR (Figure 3b) models at a very low marker distance from the centromere (0.1 cM). At 5 cM, at least five (for SDR) and six (for FDR) markers are necessary to maintain a 90% true significant identification of the mechanism. When all markers

were at least 10 cM from centromeres, nine markers were necessary to provide a 90% true significant answer for the SDR population, but only 78% significant true answers were obtained with nine markers for a FDR population. At 15 cM and nine markers, the true identification rates fall to 44% and 24% for SDR and FDR, and, at 20 cM, to 6.6% and 0%, respectively.

If the LOD2 threshold is considered, a single marker was informative in the first cM interval for the SDR model (Figure 3c) but significant replicate number decreases very quickly for FDR (Figure 3d). At 5 cM, at least four and five markers were necessary to provide 90% of true significant identification for SDR and FDR populations, respectively. With all markers at 10 cM from centromeres, at least eight markers were necessary to provide 90% true significant answers with an SDR or FDR population. For nine markers, the rate of true significant identification is improved for the SDR population at 15 cM and 20 cM (70% and 19%, respectively) as well as for the FDR population (59% and 14%, respectively) when compared with LOD3.

The rate of false identification (FDR significant conclusion [i.e.,  $\text{LOD} < -3$  or  $\text{LOD} < -2$ ] for a SDR population, or reciprocally) is very low for both models (SDR or FDR), whatever the centromere distance and the number of considered loci. At LOD3, it is under 0.1% for all conditions and it remains below 1% for the LOD2 threshold (Figure S1).

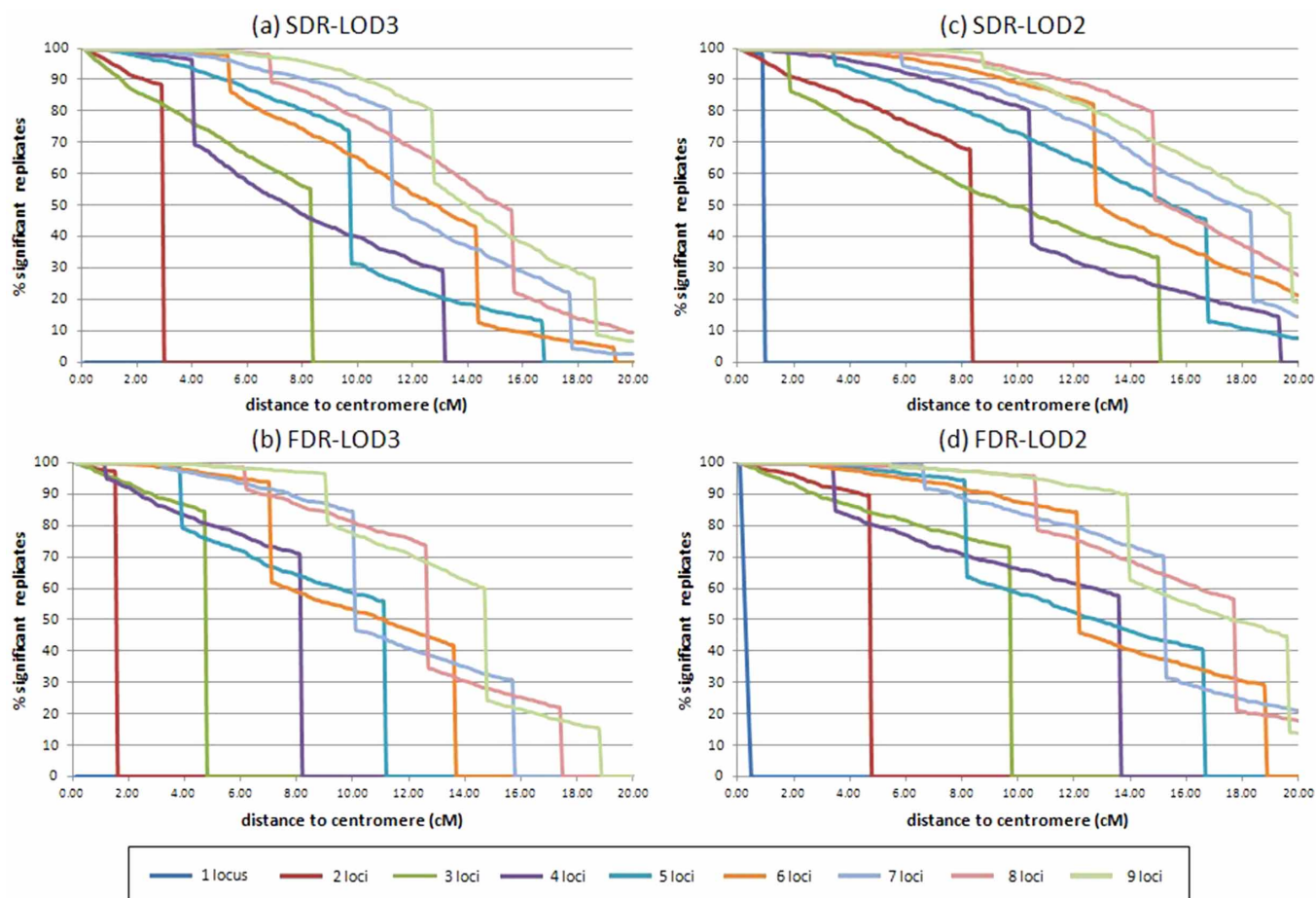
At the population level (Figure 4), due to the probabilities of the  $2n$  gamete genotypic structure under FDR and SDR models becoming similar as the distance to centromere rises, the number of hybrids needed to obtain significant conclusions for a mechanism increases as an exponential function and is more pronounced when analysing a single marker only.

For a concrete locus-centromere distance, the number of hybrids ( $h_m$ ) needed is related to the number of markers analysed as:  $h_m = h_1/m$ , being  $h_1$  the number of hybrids needed for one marker and  $m$ , the number of markers analysed. For example, for a SDR population model, at 20 cM, 58 hybrids are necessary if analysing only one marker, 29 are necessary for two markers, and 20 are necessary for three markers. The number of hybrids needed to provide the same level of conclusive answer is slightly lower for FDR (50 hybrids for one marker at 20 cM). With these population sizes, no false mechanism identification occurred for the generated populations.

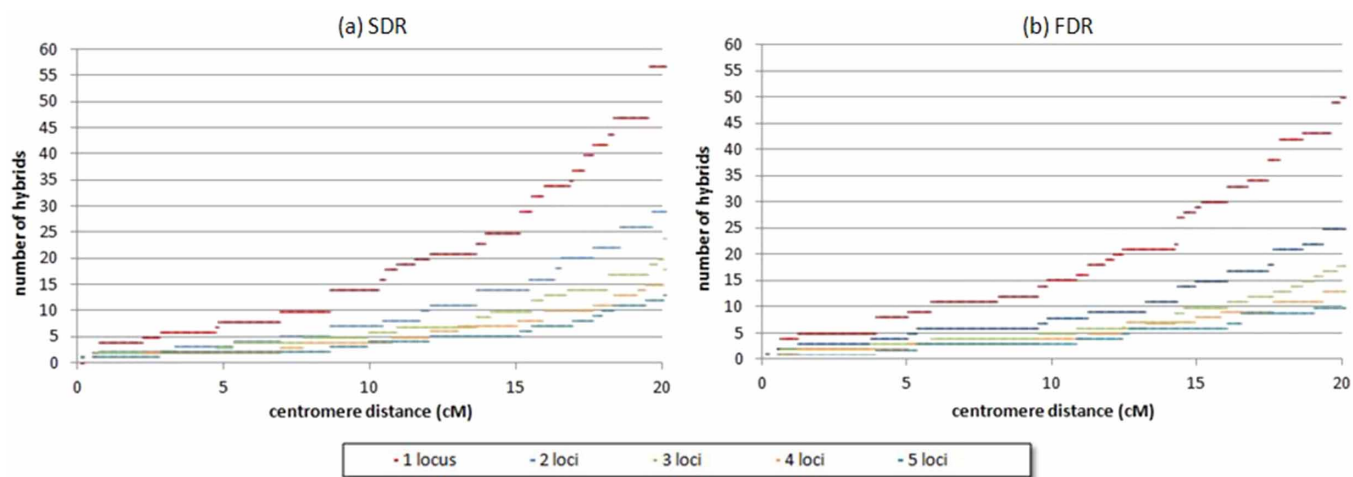
### Inference of allelic configuration of triploid hybrids and corresponding $2n$ gametes.

Assignment of allelic configuration in heterozygous triploid hybrids was performed using the MAC-PR method for SSR markers<sup>55</sup> (Figure S2) adapted for *Citrus* by Cuenca *et al.*<sup>37</sup>. However, this method uses a 1:1 dosage correction from the relative allele signals for heterozygous diploid parents ( $A_1 : A_2$ ,  $A_1 : A_3$  or  $A_3 : A_4$ ). Therefore, for markers displaying  $A_1A_2 \times A_1A_3$  configuration in the parents, among the heterozygous triploid hybrids only the  $A_1A_2A_2/A_1A_1A_2$  or  $A_1A_3A_3/A_1A_1A_3$  configurations can be determined using these methods, while no direct allele dosage estimation can be obtained for a triploid with  $A_2/A_3$  heterozygosity without a reference for the relative  $A_2/A_3$  allele signal. Similarly, for markers displaying the  $A_1A_2 \times A_3A_4$  configuration, it is not possible to directly estimate allele dosage for the heterozygous triploid hybrids. In these situations, it is possible to use a 1:1 dosage correction between  $A_1$  and  $A_3$  (for example) from the peak ratios of  $A_1A_2A_3$  triallelic hybrids observed in the same family.

A concrete example can be the genotype assignment for the “Ellendale × Fortune” population (Additional file 1) and the mCrCIR07F11 marker. “Ellendale” shows 160/162 alleles, and “Fortune” 152/164 alleles. Hybrid#1 shows 152/160/162 allele configuration. This situation allows a 1:1 dosage correction for relative allele signals between 152/160 and 152/162. Similarly, hybrid #11



**Figure 3** | Percentage of replicates with significant LOD value considering a LOD3 for (a) theoretical SDR and (b) FDR populations, and considering a LOD2 for (c) SDR and (d) FDR populations.



**Figure 4** | Number of hybrids needed to obtain significant conclusions for (a) SDR and (b) FDR mechanisms.

shows 160/162/164 allele configuration for the same marker, and therefore, allows using a 1:1 dosage correction for relative allele signals between 160/164 and 162/164. All this 1:1 dosage corrections allow inferring the allele dosage for this marker in the remaining hybrids within this population.

**Identification of the unreduced gamete parental origin.** For each hybrid, determination of the  $2n$  gamete origin was carried out by

identifying the parent that passed double genetic information to the hybrid. For markers displaying  $A_1A_2 \times A_1A_1$  or  $A_1A_2 \times A_1A_3$  configurations, the identification of  $A_1A_2A_2$  or  $A_2A_2A_3$  (i.e., double dosage of  $A_2$ , the allele specific to the female parent) configurations in the hybrid would imply a female origin of the  $2n$  gamete. For the second combination, the observation of  $A_1A_3A_3$  or  $A_2A_3A_3$  (i.e., double dosage of  $A_3$ , the allele specific to the male parent) would indicate a male origin.



For markers displaying  $A_1A_2 \times A_3A_3$  configurations in the parents, the identification of  $A_1A_2A_3$ ,  $A_1A_1A_3$ , or  $A_2A_2A_3$  configurations in the hybrid resulted from a maternal origin of the unreduced gamete, while  $A_1A_3A_3$  or  $A_2A_3A_3$  resulted from a paternal origin.

For markers with  $A_1A_2 \times A_3A_4$  parental configuration, the identification of the following genotypes ( $A_1A_1A_3$ ,  $A_1A_1A_4$ ,  $A_1A_2A_3$ ,  $A_1A_2A_4$ ,  $A_2A_2A_3$ ,  $A_2A_2A_4$ ) and ( $A_1A_3A_3$ ,  $A_2A_3A_3$ ,  $A_1A_3A_4$ ,  $A_2A_3A_4$ ,  $A_1A_4A_4$ ,  $A_2A_4A_4$ ) implied, respectively, female and male origin of the  $2n$  gamete.

Once the parental origin of the  $2n$  gamete was identified, the inference of the allelic configurations of the unreduced gametes from triploid hybrid genotyping was carried out as previously described by Cuenca *et al.*<sup>37</sup>

A summary of triploid genotypes allowing inference of the  $2n$  gamete genotype and origin, either directly or by inferring allele doses from diploid parents or reference triploid hybrids, is given in additional table S1. Loci with complete differentiation between the parents ( $A_1A_2 \times A_3A_4$  or  $A_1A_2 \times A_3A_3$ ) are by far the best configurations as they allow unequivocal identification of the  $2n$  gamete parent and unambiguous determination of  $2n$  gamete structure. When the parental origin of a  $2n$  gamete has been determined by triploid patterns at other loci, the  $2n$  gamete structure can be inferred for all triploid hybrids for the loci sharing a single allele between the two parents.

Following the previous example for the “Ellendale  $\times$  Fortune” population (Additional file 1), hybrid #1 shows 152/160/162 allele configuration for the mCrCIR07F11 marker. This situation allows the unequivocal identification of the maternal parent as the  $2n$  gamete producer for this hybrid. Similarly, the observed configurations for the rest of the hybrids within this population (152/160, 152/162, 160/164 and 162/164) allow the identification of the maternal parent as the  $2n$  gamete producer for all hybrids with information for this marker. Once the female parent has been identified as the  $2n$  gamete producer for a hybrid, for example hybrid#1, we can infer the  $2n$  female gamete and male gamete configurations from the allelic and dosage observations for the other markers. In the situation that it is not possible to infer the  $2n$  gamete producer (hybrids #30, #36, #57 and #69), additional markers have been analysed.

In this work, 543 citrus triploid hybrids were analysed and allelic patterns of the markers (Additional file 1) allowed unequivocal identification of the origin of the double dosage for each analysed triploid hybrid. Female parents were the unreduced gamete producers leading to triploid hybrids for all studied parental combinations. No triploid hybrid arising from unreduced pollen was found. It was therefore possible to infer the maternal  $2n$  gamete genotypes for all hybrids and loci.

**Identification of the restitution mechanism at the individual level in citrus.** Between 4 and 7 SSR and InDel markers have been used to analyse all 543 triploid hybrids. Allelic segregation for homozygous diploid gametes has been analysed within each family by a chi-squared test. Some markers deviated from the 1:1 expected ratio in populations with a reduced number of hybrids. Considering population with more than 20 hybrids, only the mCrCIR06B05 marker in “Fortune”-derived populations ( $\chi^2 = 5,531$ ; p-value = 0,018) and for the CF-ACA01 and CI07C07 markers in “Hernandina  $\times$  Nadorcott” population ( $\chi^2 = 9,524$ ; p-value = 0,002 and  $\chi^2 = 6,737$ ; p-value = 0,009, respectively) showed significant segregation distortions.

Heterozygosity restitution ranged between 0% and 100% for the analysed  $2n$  gametes, with a mean value of 14,87%, whereas for markers, HR ranged between 0% and 54%, with a mean value of 15,49%. Distribution of HR for both hybrids and markers is clearly biased to values near 0% (Table S3).

LOD score testing the SDR/FDR hypothesis was estimated for each individual  $2n$  gamete from its inferred genotype, as described

in the statistical method section. Positive LODs were found for 523 hybrids of the 543 analysed (Figure 5), suggesting a large global predominance of the SDR mechanism. The LOD distribution for clementine  $2n$  gametes is displaced to higher values when compared with the distribution for ‘Fortune’ and other mandarin  $2n$  gametes. Fifty-seven diploid gametes occur with LOD between 9 and 10, and these correspond mostly to the ‘Fina’ clementine progeny (Figure 5).

When using LOD3 as the threshold, SDR was found to be the restitution mechanism underlying unreduced megagametophyte production for 424 (85.3%) of the analysed triploid hybrids (Table 1). For one triploid hybrid arising from ‘Ellendale’ and two arising from ‘Fortune’ (0.6%), the FDR mechanism was implicated. The other 70 (14.1%) triploid hybrids did not give significant conclusions for either the SDR or FDR mechanisms. All unreduced gametes arising from ‘Encore’, ‘Fallgo’, ‘Guillermina’, ‘Honey’, ‘Loretina’ and ‘Wilking’ were identified as having an SDR origin, whereas for 33 unreduced gametes arising from ‘Fortune’ (16.7%) no significant conclusions were obtained (Table 1).

When using LOD2 as the threshold, the percentage of gametes with unidentified origins decreased to 9%. Gametes attributed to SDR increased to 90.1%, with significance achieved for an additional three clementine gametes, another ten from ‘Fortune’ and an extra 11 from other mandarins. No additional  $2n$  gametes arising from FDR were identified.

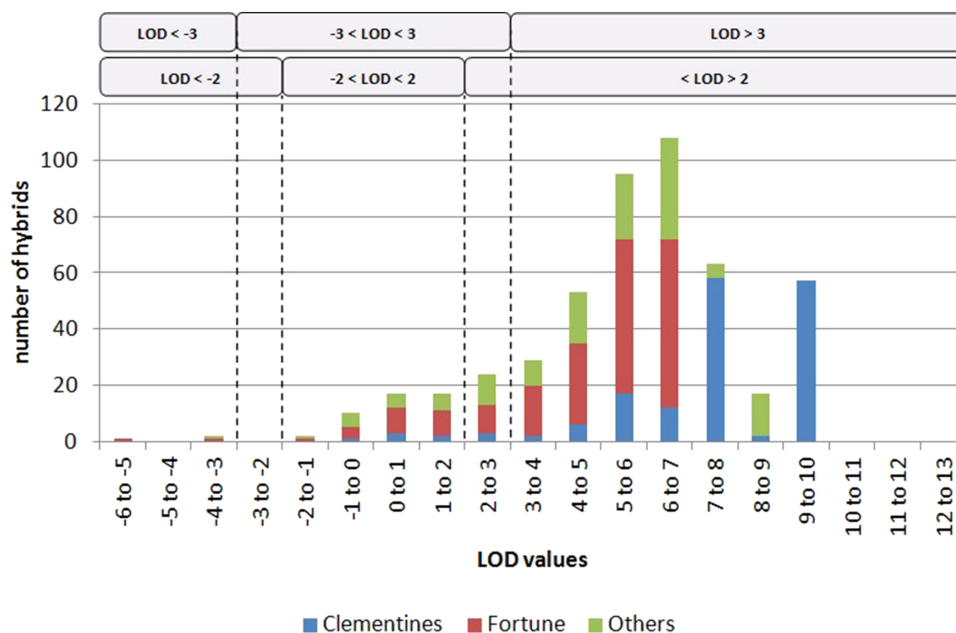
**Identification of the restitution mechanism at population level in citrus.** At the population level, all LOD scores were greater than 3, even for small populations with fewer than five hybrids. Therefore, SDR was identified as the preeminent restitution mechanism producing  $2n$  megagametophyte for all female parents analysed (Table 1).

## Discussion

**A powerful maximum-likelihood method to compare FDR and SDR hypothesis at the individual and population level has been developed.** In sexual polyploidisation, polyploids are generated by the formation of unreduced diploid gametes. From the cytogenetic point of view, two types of meiotic nuclear restitution leading to  $2n$  gamete formation are considered, FDR and SDR<sup>5,9,56,57</sup>.

The identification of the meiotic restitution mechanisms driving the formation of unreduced gametes is complex. However, molecular marker analysis is useful in such identification, and several methods, generally assuming complete chiasma interference, have been developed previously. The method proposed by Cuenca *et al.*<sup>37</sup>, based on the HR restitution curve along a linkage group, allows simultaneous identification of the restitution mechanism, raw centromere location, and comparison of several chromosome interference models. This approach is based on the analysis of genotype frequency in relatively large populations and provides global results of the preeminent mechanism; however, determination of the potential coexistence of the two mechanisms in the same progeny was not possible.

In this study, a maximum-likelihood approach based on marker HR with centromeric loci was developed and successfully applied both at the individual and population levels. Knowledge of marker-centromere distances greatly improves the statistical power of the comparison between the SDR and FDR hypotheses. For example, in this study, the restitution mechanism was identified in ‘Fortune’ as SDR at the population level with a LOD(SDR/FDR) of 933, whereas for the same population using 12 markers without information regarding marker-centromere distance, but with HR values under 50%<sup>37</sup>, the mechanism was identified as SDR with a LOD value of only 6.8. With the method proposed in the present paper, conclusions at the population level could therefore be obtained from smaller numbers of progeny and fewer markers than with non-located markers.



**Figure 5** | Frequency histogram of LOD values obtained for each individual *2n* gamete, indicating those arising from clementines, ‘Fortune’ mandarin and other mandarins analysed in this study.

The theoretical limits of our method were assessed by the simulation of populations arising from FDR or SDR mechanisms. At the population level, considering that the independent markers used are at the same distance from their respective centromeres, the power of the statistical test was directly linked to the product of the number of markers and the number of individuals. That means that the efficiency would be the same for *n* individuals with *m* markers as for *2n* individuals with *m/2* markers. Moreover, the necessary *n·m* geno-

typing points increase exponentially with increasing distance of the marker to the centromere. For example, to obtain a significant answer higher than 99%, it would be necessary a *n·m* higher than fifty-seven for markers at 20 cM, while a *n·m* value higher than eight and four would be sufficient for markers at 5 cM and 1 cM, respectively. The selection of markers as close as possible to their centromere is therefore a key element for successful analysis when low numbers of individuals and markers are used.

**Table 1** | LOD scores for progeny of 19 female parents analysed at population level and individuals within each population originated by SDR, FDR, or with unidentified origin

Group	Female parent	Nh	Nm	Cd average (cM)	LODs > +3		LODs -3 - +3		LODs < -3		Population LOD P <sub>SDR</sub> /P <sub>FDR</sub>
					number	(%)	number	(%)	number	(%)	
Clementine	‘Bruno’	17	6	7.8	15	(88.2)	2	(11.8)	0	(0.0)	98.9
	‘Clemenules’	23	5	4.4	22	(95.6)	1	(4.4)	0	(0.0)	143.3
	‘Fina’	87	6	3.7	83	(95.4)	4	(4.6)	0	(0.0)	699.3
	‘Guillermina’	14	6	8.1	14	(100.0)	0	(0.0)	0	(0.0)	91.1
	‘Hernandina’	22	5	4.4	20	(90.9)	2	(9.1)	0	(0.0)	139.0
	‘Loretina’	2	7	9.5	2	(100.0)	0	(0.0)	0	(0.0)	10.3
Mandarin	‘Imperial’	24	5	5.0	23	(95.8)	1	(4.2)	0	(0.0)	138.5
Hybrid mandarin	‘Ellendale’	69	5	9.1	50	(72.5)	18	(26.1)	1	(1.4)	282.7
	‘Encore’	3	5	4.9	3	(100.0)	0	(0.0)	0	(0.0)	17.9
	‘Fallglo’	3	5	3.7	3	(100.0)	0	(0.0)	0	(0.0)	21.6
	‘Fortune’	197	5	6.8	162	(82.2)	33	(16.7)	2	(1.1)	933.0
	‘Fortune’ × ‘Ellendale’	58	5	6.8	54	(93.1)	4	(6.9)	0	(0.0)	310.5
	‘Fortune’ × ‘Minneola’	35	4	5.2	28	(80.0)	6	(17.1)	1	(2.9)	145.7
	‘Fortune’ × ‘Murcott’	67	5	6.8	53	(79.1)	14	(20.9)	0	(0.0)	326.2
	‘Fortune’ × ‘Willowleaf’	37	4	6.6	28	(75.7)	8	(21.6)	1	(2.7)	150.6
	‘Honey’	1	4	6.1	1	(100.0)	0	(0.0)	0	(0.0)	5.1
	‘Kiyomi’	21	5	6.3	20	(95.2)	1	(4.8)	0	(0.0)	162.9
	‘Moncada’	8	4	10.3	4	(50.0)	4	(50.0)	0	(0.0)	22.1
	‘Nadorcott’	11	4	8.9	6	(54.5)	5	(45.5)	0	(0.0)	23.9
	‘Orri’	29	5	10.2	17	(58.6)	12	(41.4)	0	(0.0)	84.5
‘Orri’ × ‘Fortune’	17	5	10.2	14	(82.3)	3	(17.7)	0	(0.0)	67.5	
‘Orri’ × ‘Oronules’	12	5	10.2	3	(25.0)	9	(75.0)	0	(0.0)	17.0	
‘Ortanique’	6	5	6.6	6	(100.0)	0	(0.0)	0	(0.0)	28.1	
‘Umatilla’	5	4	11.3	1	(20.0)	4	(80.0)	0	(0.0)	9.6	
‘Wilking’	1	5	8.3	1	(100.0)	0	(0.0)	0	(0.0)	15.6	

Nh: number of hybrids within each population (pop). Nm: number of markers analyzed over each population. Cd: Centromere distance. *Italic format indicates different populations derived from ‘Fortune’ and ‘Orri’ Female parents*





In the study of citrus  $2n$  gamete progenies, significant results were obtained for all analysed populations, even for populations lower than five individuals.

One major improvement of our approach over existing methods is that it allows the identification of the restitution mechanism for each individual unreduced gamete. Simulation studies indicated that the proximity of markers to the centromeres is a key factor. With markers closer than 5 cM, five markers are sufficient to result in 95% significant answers, but significance diminishes to less than 78% and 0% for nine markers at 10 cM and 20 cM from their centromeres, respectively.

The importance of selecting markers very close to the centromere to obtain significant conclusions at the individual level is illustrated by the results of our citrus analysis. Indeed, a very high percentage of significant results at the individual level (95.4%) and with high LODs were obtained for the 'Fina' clementine progeny analysed with markers closer to centromeres than the other progenies.

Other mechanisms than meiotic restitution, also leading to unreduced gamete formation have been described, like pre-meiotic and post-meiotic genome doubling. However, both these mechanisms have only rarely been documented in plants<sup>4</sup>. Nevertheless, genetic configurations of the resultant unreduced gametes would be different than FDR or SDR-gametes.

In animals, pre-meiotic genome doubling leads to parthenogenesis<sup>58</sup>. Doubled chromosome number is reduced through meiosis and the resulting daughter chromosomes pair in the first meiotic prophase with their genetically identical counterpart. As a result, the genotype of the parent is passed on to the offspring unchanged. Analysing centromeric markers, this situation could be confused with FDR mechanism, if all markers resulted fully heterozygous in the offspring. However this situation was observed for only one of the 543 citrus diploid gametes analysed in the present work.

In case of post-meiotic doubling, meiotically formed haploid spores undergo an extra round of genome duplication, and consequently yield fully homozygous  $2n$  gametes. This situation could be also obtained in case of SDR, if all analysed centromeric markers resulted fully homozygous in the offspring. In the present work, 268 unreduced gametes resulted fully homozygous, but some heterozygous loci were observed in other unreduced gametes within the same populations, discarding a complete post restitution model at population level. At individual level, the analysis of telomeric markers allow analysing if homozygosity is maintained along the chromosome arm, and therefore concluding if the diploid gametes resulted from post-meiotic doubling or SDR. As an example, out of the 87 diploid gametes of "Fina" clementine analysed in the present study, 58 were totally homozygous for the 6 centromeric loci analysed. However, for the same population analysed with 104 markers including centromeric and telomeric loci, the HR at individual level ranged between 25% and 65%<sup>36</sup>. This broader marker study totally discard the pre- and post-meiotic doubling mechanisms at individual level. Similarly, additional marker information for the other families (data not shown) discarded the pre- and post-meiotic doubling hypothesis.

**$2n$  megagametophytes arising from SDR are the preeminent source of triploid occurrence in  $2x \times 2x$  hybrid populations using mandarin-like parents.** Spontaneous occurrences of citrus triploid hybrids arising from the union of  $2n$  megagametophytes with haploid pollen have been noted since the seventies<sup>34,32,59</sup>. However, the frequency of such events is generally low<sup>32,60</sup> and extensive breeding programs based on this type of hybridisation require very effective methodologies for embryo rescue and ploidy evaluation of large progenies mandarins<sup>32</sup>. To date, very few cases of citrus triploid hybrid occurrence in  $2x \times 2x$  crosses from unreduced pollen have been reported<sup>35,38</sup>; our unpublished results].

In this study, the mechanism leading to triploid formation in  $2x \times 2x$  crosses was elucidated, both at individual and population level, for nineteen varieties used as female parents.

All the 543 triploid hybrids analysed originated from  $2n$  megagametophytes and, therefore, no  $2n$  pollen contributed to the production of triploids in our parental combinations. These results expand to a large range of genotypes the prior conclusion obtained from cytological studies<sup>34,57</sup> for 'Sukega' (*C. paradisi*  $\times$  *C. sinensis*), 'Temple' (*C. reticulata*  $\times$  *C. sinensis*) and clementine (*C. clementina*), indicating that in such  $2x \times 2x$  crosses, triploid embryos were associated with pentaploid endosperm. However, the occurrence of triploids arising from  $2n$  pollen at very low rates has been previously reported in studies using molecular markers for three selections of clementine ('Caffin', 'Commun' or 'SRA85' and 'Muskat'), 'King' mandarin pollinated with *C. deliciosa* ('Tardivo di Ciaculi', 'Willow Leaf'), *C. reticulata* ('Hansen', 'Ananas'), *C. paradisi* ('Star Ruby') and *C. sinensis* ('Tarroco Rosso', 'Sanguinelli')<sup>35</sup> and for *C. sinensis*  $\times$  *Poncirus trifoliata* hybridisations<sup>38</sup>.

When using the LOD3 threshold, SDR was identified as the restitution mechanism for 85.3% of the analysed triploid hybrids, no significant conclusions were obtained for 14.1% of the hybrids, and 0.6% of the analysed triploids were derived from FDR (one triploid hybrid arising from 'Ellendale' and two arising from 'Fortune'). When the LOD2 threshold was considered, the percentage of individuals with unidentified origin decreased to 9% and SDR levels increased to 90.1%. Moreover, we conducted individual level analysis of previously studied 'Fortune' mandarin progeny<sup>37</sup> and the progeny arising from 'Fina'<sup>36</sup>, and we confirmed SDR at the individual level for most hybrids, which concurs with the global-level conclusions proposed in these two studies. In the current study, six clementine genotypes were also analysed to discover their unreduced gamete formation mechanism. Results indicate that SDR is the most probable mechanism in the clementine group, in agreement with previous conclusions of Luro *et al.*<sup>35</sup>. For the other mandarin varieties, SDR was also the most probable mechanism at the individual level and, therefore, also at the population level. Taken together, our data and those of others suggest that SDR is the major mechanism underlying unreduced megagametophyte formation in most mandarin genotypes.

The mechanism leading to unreduced eggs or pollen was previously elucidated for several plant species<sup>4,12</sup>. Bretagnolle and Thompson<sup>5</sup> identified that both FDR and SDR are responsible for  $2n$  pollen formation, while SDR is more frequent in the formation of  $2n$  eggs. In potato,  $2n$  pollen arises predominantly by FDR<sup>16</sup>, while  $2n$  megagametophytes arise most frequently by SDR<sup>61</sup>, although SDR-FDR mixture in the formation of  $2n$  eggs has been also found<sup>62</sup>. Bilateral sexual polyploidisation can arise either from FDR and SDR in *Lilium*<sup>8,47,63</sup> and alfalfa<sup>22</sup>. Moreover, other examples of plant species where FDR and SDR may occur simultaneously has been described<sup>5</sup>, underlining the influence of genotype and environment on the expression of meiotic abnormality factors<sup>64,65</sup>.

**Implications for citrus triploid breeding.** The genetic and phenotypic consequences of FDR and SDR gametes are highly divergent, and are of potential importance for breeding applications, due to the different parental heterozygosity rate that each mechanism transmits to the polyploid progeny<sup>4</sup>.

Under FDR, the resulting  $2n$  gametes are heterozygous from the centromere to the first crossover point, and hence the gametes retain most parental heterozygosity and epistatic interactions. With the SDR mechanism, the resulting  $2n$  gametes are homozygous from the centromere to the first crossover point, but retain parental heterozygosity on the telomeric regions<sup>12</sup>. As a result, SDR- $2n$  gametes confer a lower level of heterozygosity than FDR- $2n$  and show a corresponding greater loss of parental epistasis<sup>5,66</sup>.





Table 2 | Number of hybrids within each population analysed in this study

# population	Population	Number of hybrids	# population	Population	Number of hybrids
1	Bruno × Chandler	17	10	Honey × N'15	1
2	Clemenules × Nadorcott	23	11	Imperial × Moncada	24
3	Ellendale × Fortune	69	12	Kiyomi × Nadorcott	21
4	Encore × Ellendale	3	13	Loretina × Chandler	2
5	Fallglo × N'15	3	14	Moncada × Ellendale	8
6	Fina × Nadorcott	87	15	Nadorcott × Ellendale	11
7	Fortune × 4 male parents	197	16	Orri × 2 male parents	29
7a	Fortune × Ellendale	58	16a	Orri × Fortune	17
7b	Fortune × Minneola	35	16b	Orri × Oronules	12
7c	Fortune × Murcott	67	17	Ortanique × Wilking	6
7d	Fortune × Willowleaf	37	18	Umatilla × Simeto	5
8	Guillermina × Chandler	14	19	Wilking × Fina	1
9	Hernandina × Nadorcott	22			

If an SDR origin of  $2n$  gametes is assumed for most mandarins, sexual polyploidisation may lead to a reduced average of HR and, therefore, loss of epistatic interactions. Therefore, when compared with interploid crosses using doubled diploids<sup>67,68</sup>, the sexual polyploidisation strategy should produce more polymorphic progeny by creating a larger number of new multilocus allelic combinations<sup>4</sup>. This provides the opportunity to select innovative products within the perspective of market segmentation as a commercial strategy.

Consequences of the SDR restitution mechanism would be clearly apparent for a character controlled by a single gene. If the gene is heterozygous in the female parent, most unreduced gametes will be homozygous for that gene if it is located near the centromere, but gametes will be mostly be heterozygous for the gene if it is telomere-proximal (partial interference model;<sup>37</sup>). Recently, Cuenca *et al.*<sup>69</sup> analysed the inheritance of resistance to *Alternaria* brown-spot fungal disease in citrus triploid progenies arising from crosses between diploid parents. They demonstrated that the resistance was controlled as a recessive trait by a single locus located near a centromere (10.5 cM from the centromere of chromosome 3). If a susceptible female parent is heterozygous, the SDR mechanism leads to approximately 80% homozygous unreduced gametes, half of having two resistant alleles. As *Alternaria* resistance is a major selective trait

when maternal heterozygous parents are used, sexual polyploidisation is a more effective strategy than the use of interploid crosses, which will result in only 16.7–22.5% of progeny being resistant.

For dominant traits controlled by a single centromeric locus, interploid crosses should be more interesting than  $2x \times 2x$  crosses. For characters controlled by major loci more distant than 30 cM from the centromere, the efficiency of the two triploid breeding strategies would be relatively similar. This information is now being used routinely in the mandarin triploid breeding program carried out in Spain<sup>70</sup>.

## Methods

**Plant materials.** Analyses were performed using 543 triploid hybrids derived from 19 different mandarin genotypes as female parents in  $2x \times 2x$  cross populations (Table 2). The mandarin genotypes include six clementine and 13 hybrid mandarins. Triploid hybrids were grown at the 'Instituto Valenciano de Investigaciones Agrarias' orchards in Moncada, Valencia, Spain. Practical details for the establishment of triploid populations from  $2x \times 2x$  crosses by embryo rescue and triploid selection by flow cytometry can be found in Aleza *et al.*<sup>32</sup>. All triploid genotypes in the present study were selected after ascertaining their hybrid nature by molecular marker analysis (data not shown). Taxonomic information about both female and male parental accessions is given in additional table S2 according to the standard classification system for the *Citrus* genus<sup>71,72</sup>.

Table 3 | Centromeric markers used for genotyping each triploid population

LG	Centromere Position (cM)	Marker id	Marker type	Reference	Marker Position (cM)	Centromere-distance (cM)	Populations analyzed (#)
1	60.66	mCrCIR06B05	SSR	73	50.27	10.39	7
		CID0806	InDel	74	55.17	5.49	8
		CIBE5720	SSR	75	58.45	2.21	4,5,11,12,16,17
		MEST539	SSR	In preparation	61.82	1.16	6
		MEST001	SSR	38	70.60	9.94	10,18,19
		mCrCIR07D05	SSR	37	75.60	14.94	1,13,14,17,18
2	56.87	CX2004	SSR	38	46.67	10.20	18
		CX6F23	SSR	38	49.53	7.34	1,2,4,5,6,7,9,10,11,12,13,15,16,17,19
		CIBE4225	SSR	75	86.33	4.26	4,12
3	90.59	CID5376	InDel	74	88.24	2.35	17
		MEST470	SSR	In preparation	88.76	1.83	6
		CX0124	SSR	In preparation	110.28	19.69	13,14,16
		mCrCIR07D06	SSR	37	16.33	0.19	1,7,8,13
4	16.14	CF-ACA01	SSR	In preparation	24.41	8.27	2,4,6,9,11
		CID0245	InDel	In preparation	20.94	2.18	2,5,6,9
5	23.12	MEST104	SSR	76	40.46	17.34	1,3,8,12,13,14,15,16,18,19
		MEST191	SSR	In preparation	10.86	4.46	1,5,8,10,11,12,13,15,16,19
		mCrCIR03B07	SSR	37	83.39	13.04	7
6	6.4	CX0114	SSR	In preparation	94.97	1.46	3
		CI07C07	SSR	73	98.02	1.59	2,3,6,9,10
		mCrCIR07B05	SSR	73	31.70	22.51	3
7	96.43	mCrCIR07F11	SSR	77	49.57	2.59	1,2,3,4,5,8,9,10,11,13,14,15,17,18,19
		CI08C05	SSR	73	55.14	2.98	7



**Selection of centromeric markers for the analysis of 2n gamete origin and formation mechanisms.** Triploid citrus hybrids obtained in  $2x \times 2x$  hybridisations arise from unreduced megagametophytes<sup>32–35,59,60</sup>. Therefore, markers heterozygous for the female parent and displaying polymorphism between the two parents were primarily selected for the molecular characterisation of triploid hybrids and analysis of 2n gamete origin.

Centromere positions in all nine clementine chromosomes are known<sup>36</sup>. Molecular markers within 20 cM of the centromere were used in this study because centromere-proximal markers are more informative with regard to the mechanisms of 2n gamete formation than centromere-distal markers<sup>53</sup>. Within this range, the lowest expected HR rate is greater than 80% for FDR, while the highest HR for SDR is 40% (Figure 2). Twenty-five markers were selected for genotyping the triploid progeny. Between four and seven of these centromeric markers were used for genotyping each population (Table 3).

**Genotyping of triploid hybrids.** *DNA extraction.* Leaf DNA of triploid hybrids and their parents was isolated using the Plant DNAeasy kit from Qiagen Inc. (Valencia, CA, USA), following the manufacturer's protocol.

*SSR and InDel analyses.* Polymerase chain reactions (PCRs) were performed with wellRED oligonucleotides (Sigma-Aldrich®, St Louis, MO, USA) in a Mastercycler epgradient S (Eppendorf Scientific Inc., Westbury, NY, USA). The reaction (volume, 15 µl) contained 0.8 U Taq polymerase (Fermentas®, Burlington, VT, USA), 0.1 mM of each dNTP, 5 mM MgCl<sub>2</sub>, 3 mM of each primer, and 30 ng of DNA in buffer containing 750 mM Tris-HCl (pH 9), 50 mM KCl, 200 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 0.001% bovine serum albumin. The PCR program was 94°C for 5 min; 40 cycles of 30 s at 94°C, 1 min at 55°C and 30 s at 72°C, and a final elongation of 10 min at 72°C. Separation was carried out by capillary gel electrophoresis (CEQ 8000 Genetic Analysis System; Beckman Coulter Inc., Fullerton, CA, USA). Data collection and analysis were carried out with GenomeLab GeXP (Beckman Coulter Inc.) version 10.0 software. Identification of allele doses in heterozygous triploid hybrids was carried out using the MAC-PR method<sup>55</sup> adapted for *Citrus* by Cuenca *et al.*<sup>37</sup>.

1. Otto, S. P. & Whitton, J. Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437 (2000).
2. Soltis, P. S. & Soltis, D. E. The Role of Hybridization in Plant Speciation. *Annu. Rev. Plant Biol.* **60**, 561–588 (2009).
3. Harlan, J. R. & DeWet, J. On O. Winge and a prayer: the origins of polyploidy. *The Botanical Review* **41**, 361–390 (1975).
4. De Storme, N. & Geelen, D. Sexual polyploidization in plants—cytological mechanisms and molecular regulation. *New Phytol.* **198**, 670–684 (2013).
5. Bretagnolle, F. & Thompson, J. D. Tansley Review no-78—Gametes with the somatic chromosome-number - mechanisms of their formation and role in the evolution of autopolyploid plants. *New Phytol.* **129**, 1–22 (1995).
6. Ramsey, J. & Schemske, D. Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Syst.* **33**, 589–639 (2002).
7. Peloquin, S. J., Yerik, G. L., Werner, J. E. & Darmono, E. Potato breeding with haploids and 2n gametes. *Genome* **31**, 1000–1004 (1989).
8. Lim, K., Shen, T., Barba-González, R., Ramanna, M. S. & Van Tuyl, J. Occurrence of SDR 2n gametes in *Lilium* hybrids. *Breeding Science* **54**, 13–18 (2004).
9. Dewitte, A., Van Laere, K. & Van Huylenbroeck, J. Use of 2n gametes in plant breeding. In *Plant Breeding* (ed Abdurakhmonov, D. I.) (InTech, 2012).
10. Tavoletti, S., Bingham, E. T., Yandell, B. S., Veronesi, F. & Osborn, T. C. Half tetrad analysis in alfalfa using multiple restriction fragment length polymorphism markers. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10918–10922 (1996).
11. Cai, X. & Xu, S. S. Meiosis-driven genome variation in plants. *Current Genomics* **8**, 151–161 (2007).
12. Ramanna, M. S. & Jacobsen, E. Relevance of sexual polyploidization for crop improvement - A review. *Euphytica* **133**, 3–8 (2003).
13. Barba-Gonzalez, R., Lokker, A. C., Lim, K., Ramanna, M. S. & Van Tuyl, J. Use of 2n gametes for the production of sexual polyploids from sterile Oriental— Asiatic hybrids of lilies (*Lilium*). *Theor. Appl. Genet.* **109**, 1125–1132 (2004).
14. Barba-Gonzalez, R., Lim, K., Ramanna, M. S., Visser, R. & Van Tuyl, J. Occurrence of 2n gametes in the F1 hybrids of Oriental— Asiatic lilies (*Lilium*): Relevance to intergenomic recombination and backcrossing. *Euphytica* **143**, 67–73 (2005).
15. Rhoades, M. M. & Dempsey, E. Induction of chromosome doubling at meiosis by the elongate gene in maize. *Genetics* **54**, 505–522 (1966).
16. Mok, D. W. S. & Peloquin, S. J. Three mechanisms of 2n pollen formation in diploid potatoes. *Can. J. Genet. Cytol.* **17**, 217–225 (1975).
17. Mendiburu, A. O. & Peloquin, S. J. The significance of 2n gametes in potato breeding. *Theoretical and Applied Genetics* **49**, 53–61 (1977).
18. Jongedijk, E., Ramanna, M. S., Savor, Z. & Hermesen, J. G. T. Formation of 1<sup>st</sup> division restitution (FDR) 2n-megaspores through pseudohomotypic division in ds-1 (desynapsis) mutants of diploid potato - routine production of tetraploid progeny from 2xFDR  $\times$  2xFDR crosses. *Theor. Appl. Genet.* **82**, 645–656 (1991).
19. Crespel, L. & Gudin, S. Evidence for the production of unreduced gametes by tetraploid *Rosa hybrida* L. *Euphytica* **133**, 65–69 (2003).
20. Lelley, T., Mahmoud, A. A. & Lein, V. Genetics and cytology of unreduced gametes in cultivated rye (*Secale cereale* L.). *Genome* **29**, 635–638 (1987).
21. Ortiz, R. & Peloquin, S. J. Breeding for 2n-egg production in haploid-X-species 2x-potato hybrids. *Am. Potato J.* **68**, 691–703 (1991).

22. Barcaccia, G., Tavoletti, S., Mariani, A. & Veronesi, F. Occurrence, inheritance and use of reproductive mutants in alfalfa improvement. *Euphytica* **133**, 37–56 (2003).
23. Ortiz, R. Occurrence and Inheritance of 2n pollen in *Musa*. *Annals of Botany* **79**, 449–453 (1997).
24. Ssebuliba, R., Tenkouano, A. & Pillay, M. Male fertility and occurrence of 2n gametes in East African Highland bananas (*Musa* spp.). *Euphytica* **164**, 53–62 (2008).
25. Starrantino, A. Use of triploids for production of seedless cultivars in citrus improvement program. *Proceedings of the International Society of Citriculture* **1**, 117–121 (1992).
26. Ollitrault, P., Dambier, D., Luro, F. & Froelicher, Y. Ploidy manipulation for breeding seedless triploid citrus. *Plant Breeding Reviews* **30**, 323–352 (2008).
27. Reforgiato Recupero, G., Russo, G. & Recupero, S. New promising citrus triploid hybrids selected from crosses between monoembryonic diploid female ant tetraploid male parents. *HortScience* **40**, 516–520 (2005).
28. Aleza, P., Cuenca, J., Juárez, J., Pina, J. A. & Navarro, L. 'Garbi' mandarin: a new late-maturing triploid hybrid. *HortScience* **45**, 139–141 (2010).
29. Cuenca, J., Aleza, P., Juárez, J., Pina, J. A. & Navarro, L. 'Safor' mandarin: a new citrus mid-late triploid hybrid. *HortScience* **45**, 977–980 (2010).
30. Krug, C. Chromosome number in the subfamily *Aurantioidae* with special reference to the genus *Citrus*. *Botanical Gazette* **104**, 602–611 (1943).
31. Navarro, L. *et al.* Programa de obtención de híbridos triploides de mandarina en España. *Phytoma* **170**, 36–41 (2005).
32. Aleza, P., Juárez, J., Cuenca, J., Ollitrault, P. & Navarro, L. Recovery of citrus triploid hybrids by embryo rescue and flow cytometry from  $2x \times 2x$  sexual hybridisation and its application to extensive breeding programs. *Plant Cell Rep* **29**, 1023–1034 (2010).
33. Esen, A. & Soost, R. K. Unexpected triploids in *Citrus*: their origin, identification, and possible use. *J. Hered.* **62**, 329–333 (1971).
34. Esen, A., Soost, R. K. & Geraci, G. Genetic evidence for the origin of diploid megagametophytes in *Citrus*. *J. Hered.* **70**, 5–8 (1979).
35. Luro, F. *et al.* Identification and evaluation of diplogyny in clementine (*Citrus clementina*) for use in breeding *Acta Hort.* (ISHS) **663**, 841–848 (2004).
36. Aleza, P. *et al.* Genetic mapping of centromeres of the nine *Citrus clementina* chromosomes using half-tetrad analysis and recombination patterns in unreduced and haploid gametes. *BMC Plant Biology* **15**, 80 (2015).
37. Cuenca, J. *et al.* Multilocus half-tetrad analysis and centromere mapping in citrus: evidence of SDR mechanism for 2n megagametophyte production and partial chiasma interference in mandarin cv 'Fortune'. *Heredity* **107**, 462–470 (2011).
38. Chen, C. *et al.* Origin and frequency of 2n gametes in *Citrus sinensis*  $\times$  *Poncirus trifoliata* and their reciprocal crosses. *Plant Science* **174**, 1–8 (2008).
39. Barone, A., Gebhardt, C. & Frusciant, L. Heterozygosity in 2n gametes of potato evaluated by RFLP markers. *Theor. Appl. Genet.* **91**, 98–104 (1995).
40. Douches, D. S. & Quiros, C. F. Genetic strategies to determine the mode of 2n egg formation in diploid potatoes. *Euphytica* **38**, 247–260 (1988).
41. Vorsa, N. & Rowland, L. J. Estimation of 2n megagametophyte heterozygosity in a diploid blueberry (*Vaccinium darrowii* Camp) clone using RAPDs. *J. Hered.* **88**, 423–426 (1997).
42. Buso, J. A., Reifschneider, F. J. B., Boiteux, L. S. & Peloquin, S. J. Effects of 2n-pollen formation by first meiotic division restitution with and without crossover on eight quantitative traits in 4x-2x potato progenies. *Theor. Appl. Genet.* **98**, 1311–1319 (1999).
43. Douches, D. S. & Maas, D. L. Comparison of FDR and SDR derived tetraploid progeny from  $2x \times 4x$  crosses using haploids of *Solanum tuberosum* L that produce mixed modes of 2n eggs. *Theor. Appl. Genet.* **97**, 1307–1313 (1998).
44. Errico, A., Cammareri, M. & Conicella, C. Meiotic nuclear restitution mechanisms in a triploid lily. *Caryologia* **58**, 367–373 (2005).
45. Lim, K. B., Ramanna, M. S., de Jong, J. H., Jacobsen, E. & van Tuyl, J. M. Indeterminate meiotic restitution (IMR): a novel type of meiotic nuclear restitution mechanism detected in interspecific lily hybrids by GISH. *Theor. Appl. Genet.* **103**, 219–230 (2001).
46. Takahashi, C., Leitch, I. J., Ryan, A., Bennett, M. D. & Brandham, P. E. The use of genomic in situ hybridization (GISH) to show transmission of recombinant chromosomes by a partially fertile bigeneric hybrid, *Gasteria lutzii*  $\times$  *Aloe aristata* (*Aloaceae*), to its progeny. *Chromosoma* **105**, 342–348 (1997).
47. Karlov, G. I., Khrustaleva, L. I., Lim, K. B. & Van Tuyl, J. M. Homoeologous recombination in 2 n-gametes producing interspecific hybrids of *Lilium* (*Liliaceae*) studied by genomic in situ hybridization (GISH). *Genome* **42**, 681–686 (1999).
48. Bastiaanssen, H. J. M., Ramanna, M. S., Huigen, D. J. & Jacobsen, E. Selection of diploid tuberous *Solanum* hybrids for 2n-egg formation using  $2x \cdot 4x$ -crosses. *Euphytica* **101**, 325–339 (1998).
49. Hayashi, M., Kato, J., Ohashi, H. & Mii, M. Unreduced 3x gamete formation of allotriploid hybrid derived from the cross of *Primula denticulata* (4x)  $\times$  *P. rosea* (2x) as a causal factor for producing pentaploid hybrids in the backcross with pollen of tetraploid *P. denticulata*. *Euphytica* **169**, 123–131 (2009).
50. Park, T. *et al.* Genetic positioning of centromeres using half-tetrad analysis in a 4x-2x cross population of potato. *Genetics* **176**, 85–94 (2007).
51. Mendiburu, A. O. & Peloquin, S. J. Gene-centromere mapping by 4X-2X matings in potatoes. *Theor Appl Genet* **54**, 177–180 (1979).
52. Zhao, H. & Speed, T. P. Statistical analysis of half-tetrads. *Genetics* **150**, 473–485 (1998).



53. Ollitrault, P. *et al.* A reference genetic map of *C. clementina* Hort. ex Tan.; citrus evolution inferences from comparative mapping. *BMC Genomics* **13**, 593 (2012).
54. Foss, E., Lande, R., Stahl, F. & Steinberg, C. Chiasma interference as a function of genetic distance. *Genetics* **133**, 681–691 (1993).
55. Esselink, G. D., Nybom, H. & Vosman, B. Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method. *Theor. Appl. Genet.* **109**, 402–408 (2004).
56. Ramsey, J. & Schemske, D. W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* **29**, 467–501 (1998).
57. Brownfield, L. & Kohler, C. Unreduced gamete formation in plants: mechanisms and prospects. *J. Exp. Bot.* **62**, 1659–1668 (2011).
58. Stenberg, P. & Saura, A. in *Lost Sex. The Evolutionary Biology of Parthenogenesis* (eds Schön, I., Martens, K., & van Dijk, P.) Ch. 4, 63–74 (Springer, 2009).
59. Esen, A. & Soost, R. K. Precocious development and germination of spontaneous triploid seeds in *Citrus*. *J. Hered.* **64**, 147–154 (1973).
60. Geraci, G., Esen, A. & Soost, R. K. Triploid progenies from  $2x \times 2x$  crosses of *Citrus* cultivars. *J. Hered.* **66**, 177–178 (1975).
61. Stelly, D. M. & Peloquin, S. J. Diploid female gametophyte formation in 24-chromosome potatoes - genetic-evidence for the prevalence of the 2<sup>nd</sup> meiotic division restitution mode. *Canadian Journal of Genetics and Cytology* **28**, 101–108 (1986).
62. Conicella, C., Barone, A., Delgiudice, A., Frusciant, L. & Monti, L. M. Cytological evidences of SDR-FDR mixture in the formation of 2n eggs in a potato diploid clone. *Theor. Appl. Genet.* **81**, 59–63 (1991).
63. Khan, N. *et al.* Relevance of unilateral and bilateral sexual polyploidization in relation to intergenomic recombination and introgression in *Lilium* species hybrids. *Euphytica* **171**, 157–173 (2010).
64. Pécrix, Y. *et al.* Polyploidization mechanisms: temperature environment can induce diploid gamete formation in *Rosa* sp. *J. Exp. Bot.* doi:10.1093/jxb/err052 (2011).
65. Mason, A., Nelson, M., Yan, G. & Cowling, W. Production of viable male unreduced gametes in Brassica interspecific hybrids is genotype specific and stimulated by cold temperatures. *BMC Plant Biology* **11**, 103 (2011).
66. Ortiz, R., Simon, P., Jansky, S. & Stelly, D. Ploidy manipulation of the gametophyte, endosperm and sporophyte in nature and for crop improvement: a tribute to Professor Stanley J. Peloquin (1921–2008). *Annals of Botany* **104**, 795–807 (2009).
67. Aleza, P., Juárez, J., Hernández, M., Ollitrault, P. & Navarro, L. Implementation of extensive citrus triploid breeding programs based on  $4x \times 2x$  sexual hybridizations. *Tree Genetics and Genomes* **8**, 1293–1306 (2012).
68. Aleza, P., Juárez, J., Cuenca, J., Ollitrault, P. & Navarro, L. Extensive citrus triploid hybrid production by  $2x \times 4x$  sexual hybridizations and parent-effect on the length of the juvenile phase. *Plant Cell Rep.* **31**, 1723–1735 (2012).
69. Cuenca, J. *et al.* Genetically based location from triploid populations and gene ontology of a 3.3-mb genome region linked to alternaria brown spot resistance in citrus reveal clusters of resistance genes. *Plos One* **8**, e76755 (2013).
70. Navarro, L. *et al.* The triploid mandarin breeding program in Spain. *XII International Citrus Congress*. Valence (Spain): 18–23 November 2012. International Society of Citriculture (2012).
71. Swingle, W. T. & Reece, P. C. in *The Citrus Industry, Vol 1* (eds Reuther, W., Webber, H. & Batchelor, L.) 190–430 (University of California, Berkeley, USA, 1967).
72. Tanaka, T. Fundamental discussion of *Citrus* classification. *Studia Citrologica* **14**, 1–6 (1977).
73. Froelicher, Y. *et al.* Characterization of microsatellite markers in mandarin orange (*Citrus reticulata* Blanco). *Molecular Ecology Resources* **8**, 119–122 (2008).
74. Ollitrault, F. *et al.* Development of indel markers from *Citrus clementina* (Rutaceae) BAC-end sequences and interspecific transferability in *Citrus*. *Am J Bot* **99**, e268–73 (2012).
75. Ollitrault, F. *et al.* Development of SSR markers from *Citrus clementina* (Rutaceae) BAC end sequences and interspecific transferability in *Citrus*. *Am J Bot* **97**, e124–9 (2010).
76. Garcia-Lor, A. *et al.* A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the ‘true citrus fruit trees’ group (Citrinae, Rutaceae) and the origin of cultivated species. *Ann Bot*: doi: 10.1093/aob/mcs227 (2012)
77. Kamiri, M. *et al.* Evidence for non-disomic inheritance in a *Citrus* interspecific tetraploid somatic between *C. reticulata* and *C. lemon* hybrid using SSR markers and cytogenetic analysis. *Plant Cell Rep*: doi:10.1007/s00299-011-1050-x (2011).

## Acknowledgments

This work was supported by a grant (PrometeoII/2013/008) from the Generalitat Valenciana, Spain. Authors thanks Frédérique Ollitrault, María Hernández and Ana M. Moreno their collaboration to this paper.

## Author contributions

J.C. and P.A.: Contributed to plant recovery, analysed the data and wrote the manuscript; J.J.: Contributed to plant recovery; A.G.L.: Contributed to the data analysis; Y.F.: Contributed to plant recovery; L.N. and P.O.: managed the work and contributed to write and revise the manuscript; J. C. and P. O.: implemented the statistical method.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Cuenca, J. *et al.* Maximum-likelihood method identifies meiotic restitution mechanism from heterozygosity transmission of centromeric loci: application in citrus. *Sci. Rep.* **5**, 9897; DOI:10.1038/srep09897 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>