the plant journal

SEB Society for Experimental Biology

The Plant Journal (2017) **90,** 330–343 doi: 10.1111/tpj.13496

Partial preferential chromosome pairing is genotype dependent in tetraploid rose

Peter M. Bourke[1], Paul Arens[1], Roeland E. Voorrips[1], G. Danny Esselink[1], Carole F. S. Koning-Boucoiran[1,†], Wendy P. C. van't Westende[1], Tiago Santos Leonardo[1,‡], Patrick Wissink[1], Chaozhi Zheng[2], Geert van Geest[1,3], Richard G. F. Visser[1], Frans. A. Krens[1], Marinus J. M. Smulders[1,*] and Chris Maliepaard[1,*]

[1]Plant Breeding, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands, [2]Biometris, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands and [3]Horticulture and Product Physiology, Department of Plant Sciences, Wageningen University & Research, Droevendaalsesteeg1, 6708 PB Wageningen, The Netherlands

Received 4 December 2016; revised 23 January 2017; accepted 24 January 2017; published online 31 January 2017.
*For correspondence (e-mails chris.maliepaard@wur.nl; rene.smulders@wur.nl).
†Present address: H. A. N. University of Applied Sciences, PO Box 6960, 6503 GL, Nijmegen, The Netherlands.
‡Present address: IESE Business School, University of Navarra, Spain.

SUMMARY

It has long been recognised that polyploid species do not always neatly fall into the categories of auto- or allopolyploid, leading to the term 'segmental allopolyploid' to describe everything in between. The meiotic behaviour of such intermediate species is not fully understood, nor is there consensus as to how to model their inheritance patterns. In this study we used a tetraploid cut rose (*Rosa hybrida*) population, genotyped using the 68K WagRhSNP array, to construct an ultra-high-density linkage map of all homologous chromosomes using methods previously developed for autotetraploids. Using the predicted bivalent configurations in this population we quantified differences in pairing behaviour among and along homologous chromosomes, leading us to correct our estimates of recombination frequency to account for this behaviour. This resulted in the re-mapping of 25 695 SNP markers across all homologues of the seven rose chromosomes, tailored to the pairing behaviour of each chromosome in each parent. We confirmed the inferred differences in pairing behaviour among chromosomes by examining repulsion-phase linkage estimates, which also carry information about preferential pairing and recombination. Currently, the closest sequenced relative to rose is *Fragaria vesca*. Aligning the integrated ultra-dense rose map with the strawberry genome sequence provided a detailed picture of the synteny, confirming overall co-linearity but also revealing new genomic rearrangements. Our results suggest that pairing affinities may vary along chromosome arms, which broadens our current understanding of segmental allopolyploidy.

Keywords: high-density integrated map, segmental allopolyploid, polyploid genetic linkage map, *Rosa hybrida*, meiotic chromosomal pairing behaviour.

INTRODUCTION

Polyploids are generally divided into two types – autopolyploids and allopolyploids. There continues to be debate about the definition of these categories, namely, whether they should be defined by their (presumed or known) mode of origin or their (observed) mode of inheritance (Ramsey and Schemske, 1998), otherwise described as the taxonomic or genetic definitions, respectively (Doyle and Egan, 2010). According to the taxonomic definition, polyploids are distinguished by their number of founder species (one in the case of autopolyploids, two or more in the case of allopolyploids), whereas the genetic definition distinguishes polysomic inheritance resulting from random pairing of chromosomes during meiosis (autopolyploid) from disomic inheritance resulting from non-random or preferential pairing (allopolyploid) (Doyle and Sherman-Broyles, 2016).

Theoretically, at least, it has long been recognised that there may also be intermediate forms of polyploidy, variously termed segmental allopolyploidy (Stebbins, 1947), partial preferential pairing (Wu *et al.*, 1992), incomplete

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

polysomy (Guimarães *et al.*, 1997), heterosomy (Roux and Pannell, 2015) or mixosomy (Soltis *et al.*, 2016). In these intermediate categories, the genetic definition (i.e. pairing behaviour during meiosis) takes precedence (Figure 1a). This pairing behaviour is primarily important as it determines the extent of recombination between homologues or homoeologues, providing a diagnostic of the type of polyploidy (Parisod *et al.*, 2010; Doyle and Sherman-Broyles, 2016). Furthermore, it may influence our ability to produce linkage maps and subsequently perform accurate quantitative trait locus (QTL) analysis, which is of importance for both fundamental and applied plant research. Understanding the pairing behaviour of polyploid species is also relevant for the study of genome evolution upon genome duplication.

Various ornamental crops, *Rosa* species in particular, have a long history of hybridisation and polyploid formation between and within species to generate the diversity of cultivars we have today (Smulders *et al.*, 2011). This breeding and selection is likely to have contributed to differences in homology between various chromosomes, thought to play a role in meiotic pairing behaviour (Bingham and Gillies, 1971; Lentz *et al.*, 1983). Pairing in other species has been shown to be under genetic control, for example the *Ph1* locus of hexaploid wheat (Okamoto, 1957; Riley and Chapman, 1958) or the *PrBn* locus in *Brassica napus* (Jenczewski *et al.*, 2003; Nicolas *et al.*, 2009), but may also be influenced by environmental factors (Bomblies *et al.*, 2015). Despite some advances in our understanding of polyploid meiotic regulation, there remain many unanswered questions, particularly in species with mixed inheritance types.

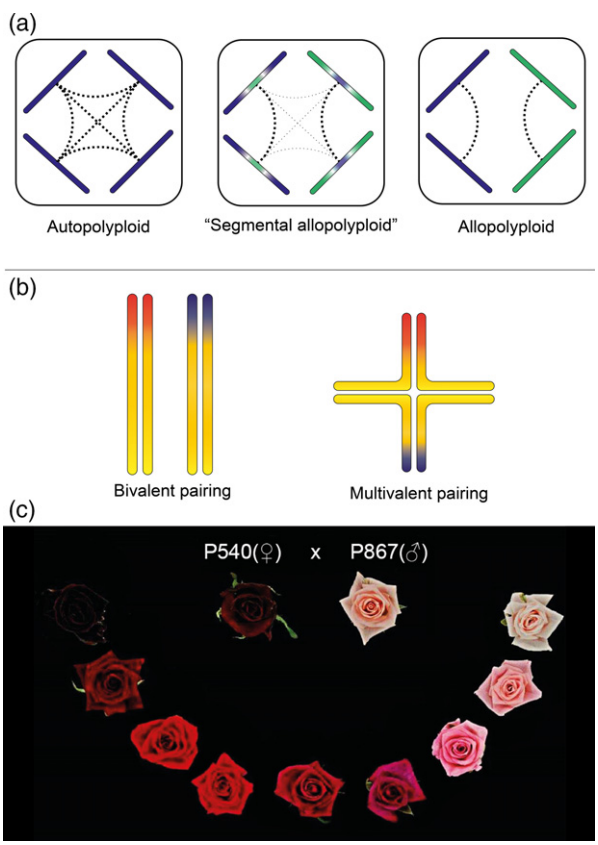


Figure 1. The meiotic pairing behaviour of tetraploid rose is not fully understood, yet exhibits many features of a segmental allopolyploid; mapping populations and high-resolution maps can shed light on such questions. (a) Schematic representation of auto- and allopolyploid pairing (in this case, for a tetraploid), with intermediate pairing behaviour of a segmental allopolyploid shown in between.
(b) Hypothesised model to account for variable rates of preferential pairing along a chromosome. If preferential pairing initiation (telomeric bouquet formation) is confined to one end of the chromosome, random pairing between the other telomeres can lead to both bivalent and multivalent pairing as shown. Some degree of preferential pairing is still expected given that pairing between one set of telomeres is highly preferential; this agrees with the observations of integrated consensus map 1 of P1.
(c) Parents P540 and P867 of the K5 rose population used in this study, with examples showing the range of segregation for flower colour in the F_1 progeny. [Colour figure can be viewed at wileyonlinelibrary.com].

Genetic mapping in rose

The genus *Rosa* consists of a highly complex and much-debated phylogeny, divided into four subgenera, of which the subgenus *Rosa* comprises over 95% of all species (Wissemann and Ritz, 2005). *Rosa* is generally divided into ten sections, four of which (Synstylae, Gallicanae, Indicae and Pimpinellifoliae) have contributed to the gene pool of the domesticated rose (Smulders *et al.*, 2011). *Rosa hybrida*, or the hybrid tea rose, is nowadays the most well-known and commercially important representative of the genus and has a complex mixture of hybrid perpetuals derived from China rose, Noisettes (*Rosa chinensis*), Bourbons as well as *Rosa gallica* and *Rosa alba*, and tea roses (*Rosa* × *odorata*) in its pedigree (Koning-Boucoiran *et al.*, 2012; Liorzou *et al.*, 2016). Given this complexity of origin, it is not surprising that it remains a poorly understood species genetically (Debener and Linde, 2009). Studies of the mode of inheritance of wild tetraploid rose populations have found evidence for both disomic and tetrasomic behaviour (Wissemann and Ritz, 2005; Joly *et al.*, 2006). Tetrasomic inheritance is generally assumed for *R. hybrida* (Gar *et al.*, 2011) with the possibility of some preferential pairing (Koning-Boucoiran *et al.*, 2012), although this has never been quantified. A large number of publications have studied linkage mapping in *Rosa*, including an integrated map at the diploid level (Spiller *et al.*, 2011) and maps at the tetraploid level (Gar *et al.*, 2011; Koning-Boucoiran *et al.*, 2012; Vukosavljev *et al.*, 2016), but none have taken full account of its pairing behaviour. An initiative is under way to sequence the genome of the diploid species *R. chinensis* (Bendahmane *et al.*, 2016), for which ultra-high-density genetic linkage maps will be likely to provide useful information for connecting and orientating scaffolds (Bartholomé *et al.*, 2015).

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.,

Identifying and quantifying preferential pairing

Identifying pairing preferences to help formulate a model of meiosis has been a long-standing challenge for researchers. The earliest methods to determine pairing behaviour were cytological and relied on counting the frequency of bivalents and multivalents during diakinesis or metaphase I of meiosis (Lentz *et al.*, 1983; Sybenga, 1994), under the assumption that allopolyploids should exhibit more bivalent pairing than autopolyploids. However, certain autopolyploid species such as potato predominantly pair as bivalents (Swaminathan and Howard, 1953) and yet show no evidence for preferential pairing. Bivalent to multivalent ratios are no longer seen as an accurate method for determining pairing type.

More recent methods have used molecular markers, which carry a signature of the parental meioses. Comparing the observed segregation ratios of different marker types with those expected under an assumed pairing model is a simple test for pairing behaviour, but can be influenced by marker skewness, which may be caused by selection or other, unknown, factors. If preferential pairing is incomplete, marker segregation ratios cannot be tested against a fixed set of expected ratios (Allendorf and Danzmann, 1997). Repulsion-phase linkages (i.e. linkage between markers which target the same genomic region but which tag alternative homologues of the chromosome) are also sensitive to deviations from random pairing. Initially, the identification of significant repulsion linkages was taken as evidence of disomic behaviour (Da Silva *et al.*, 1993; Al-Janabi *et al.*, 1994). Later studies attempted to estimate the degree of preferential pairing from repulsion-phase recombination frequency estimates, although they relied on prior knowledge or assumptions about repulsion-phase inter-marker distances (Qu and Hancock, 2001; Cao *et al.*, 2004).

Recently, methods for creating integrated genetic maps in autotetraploid species using single nucleotide polymorphism (SNP) marker data have been developed (Hackett *et al.*, 2013; Bourke *et al.*, 2016). These methods offer the possibility of identifying repulsion-linkages that can help reveal pairing behaviour. Furthermore, methods to reconstruct offspring haplotypes in mapping populations of tetraploids have also become available (Hackett *et al.*, 2013; Zheng *et al.*, 2016). In particular, TetraOrigin (Zheng *et al.*, 2016) provides the most likely predicted pairing structures per offspring (either in bivalents or quadrivalents), enabling the estimation of population-wide meiotic pairing behaviour as well as the strength of preferential pairing (ρ) should it exist.

In this study, we used a genotyped tetraploid mapping population of rose to investigate its meiotic behaviour. The 68K WagRhSNP array (Koning-Boucoiran *et al.*, 2015; Schulz *et al.*, 2016; Vukosavljev *et al.*, 2016) was used to create an ultra-high-density tetraploid linkage map, enabling the quantification of preferential pairing, quadrivalent formation and double reduction during parental meiosis. In a previous study using simulated data it was hypothesised that preferential pairing could effectively be ignored in polyploid linkage mapping (up to a 70% deviation from random pairing) (Bourke *et al.*, 2016). In this study we had the opportunity to test this prediction by correcting each linkage map for the observed pairing behaviour and comparing them with those created under a random model.

RESULTS

Marker filtering and dosage assignment

All 68 893 SNPs on the WagRhSNP array were assayed by two probes, designed to anneal to the flanking sequence at both strands (Koning-Boucoiran *et al.*, 2015). Probes were independently scored in fitTetra (Voorrips *et al.*, 2011) and were filtered according to a number of quality criteria, such as containing at most 25% missing values (in a later filtering step after merging probes this was reduced to 10%), being non-skewed ($P < 0.001$) and containing fewer than 5% invalid or unexpected scores (Table S1 in the Supporting Information). Filtered probes were merged where possible, with conflicting probes being kept separate, using the mapping as a quality check. Although parents were genotyped in triplicate, two of the Parent 1 (P1) replicates showed high numbers of missing values and low concordance with the offspring (we suspect that some genotypes were incorrectly labelled during previous multiplication steps). Therefore, the replicate of P1 with the best concordance with the offspring scores was selected, whereas for P2 two of the three replicates were found to have few conflicts (0.4%) and could be merged. There were 28 109 segregating markers to begin mapping, with a good spread over the nine marker classes (Table S2). Nineteen pairs of duplicate individuals were identified (pairwise genotype correlation > 0.95). Two individuals with more than 10% missing values were also removed, resulting in a mapping population size of 151 individuals.

Linkage map construction under a tetrasomic model

We followed the mapping procedure as described for potato in Bourke *et al.* (2016), although it was not possible to associate homologue (simplex $\times$ nulliplex, or S $\times$ N) clusters into chromosomal groups on the basis of repulsion-phase linkages due to more noise in the data. Instead, coupling linkages with duplex $\times$ nulliplex (D $\times$ N) markers were used to provide these associations. We used new marker ordering software MDSmap (Preedy and Hackett, 2016), which greatly facilitated the creation of integrated chromosomal linkage maps in this study. Rose is a

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.,

predominantly tetrasomic species, and thus we can identify four homologous copies of each chromosome from each parent, assuming sufficient S × N and N × S markers exist. Over both parents and seven linkage groups there are therefore 4 × 7 × 2 = 56 homologue groups expected. Fifty-five of the expected 56 homologues were identified, following which we assigned every other marker type to both its most probable linkage group (chromosomes, termed integrated consensus map (ICM) groups 1–7, cf. Spiller *et al.* (2011)) and homologue(s) using a linkage logarithm of odds (LOD) threshold of 3.

Estimation of preferential pairing parameters

Given an integrated linkage map and a population with dosage scores, TetraOrigin (Zheng *et al.*, 2016) can be used to infer the identity-by-descent (IBD) probabilities for all offspring, useful for subsequent quantitative trait locus (QTL) analyses (Hackett *et al.*, 2014). Here, we used it to estimate preferential pairing between chromosomes. Using the assumption of bivalent pairing, we applied a χ^2 test to the predicted bivalent counts under the hypothesis of random pairing (Table 1). At a significance level of 0.001 three chromosomes exhibited unusual behaviour (ICM1 of P1 and ICM3 and ICM4 of P2), with an over-representation of one bivalent configuration and an under-representation of the other two, which was most extreme on chromosome ICM1 of P1 (96 out of 151 counts were of the same pairing configuration). For ICM5 in P1, there was a near-significant departure from random pairing ($P = 0.002$), but this arose from an under-representation of one configuration not two, which we have not attempted to model.

Table 1 Predicted pairing behaviour according to TetraOrigin, showing significance of the deviation from random pairing and the estimated preferential pairing parameter, ρ

Chm	12_34[a]	13_24	14_23	χ^2	*P*-value[b]	ρ
P1						
1	96	35	20	64.4	**1.00×10^{-14}**	0.302
2	55	50	46	0.8	0.7	–
3	56	48	47	1.0	0.6	–
4	60	56	35	7.2	0.03	–
5	68	50	33	12.2	0.002	–
6	61	56	34	8.2	0.02	–
7	66	47	38	8.1	0.02	–
Chm	56_78	57_68	58_67	χ^2	*P*-value[b]	ρ
P2						
1	63	51	37	6.7	0.04	–
2	58	55	38	4.6	0.1	–
3	77	43	31	22.6	**0.00001**	0.177
4	80	40	31	27.0	**1.3×10^{-6}**	0.196
5	67	48	36	9.7	0.01	–
6	52	52	47	0.3	0.9	–
7	67	42	42	8.3	0.02	–

[a]The predicted number of offspring with bivalent pairing between homologues 1 and 2, and the second bivalent pairing between 3 and 4.
[b]*P*-value of the χ^2 test of the hypothesis of random pairing, significant values ($P < 0.001$) are shown in bold.
Chm, Chromosome.

Preferential pairing estimated from repulsion-phase linkages

Preferential pairing has previously been deduced from repulsion-phase estimates between 'nearby' S × N marker pairs in other studies. We estimated the minimum resolution of recombination frequency (r_{min}) as approximately 0.0117 and found 73 386 repulsion S × N marker pairs in P1 and 103 496 pairs in P2 mapped to within this distance on the integrated map (Table S3). Of these, 2896 had significant evidence for non-tetrasomic behaviour (using a false discovery rate (FDR)-corrected threshold of 0.00081, where $\alpha = 0.05$). Strong preferential pairing was identified on ICM1 in P1, confirming our previous findings. The negative $\log_{10}(P)$ values of these tests are visualised in Figure 2. The strength of significance was not constant across the chromosome, partly due to uneven marker distribution. However, in the case of ICM1 of P1 this appears to also be due to differences in pairing behaviour between chromosome arms. A possible model explaining this behaviour is represented in Figure 1(b), in which variable pairing affinities are realised through a mixture of bivalent and non-random quadrivalent formation. Our estimated preferential pairing parameters corresponded well with those from TetraOrigin (Table 2), although ICM7 of P1 also shows a region of preferential pairing, which was not predicted by TetraOrigin. The strongest evidence for preferential pairing from repulsion-phase linkages remains that of ICM1 in P1 and ICM3 in P2.

Ultra-high density *R. hybrida* linkage maps, tailored for segmental allopolyploidy

We modified our pairwise recombination frequency framework to include preferential pairing and recalculated the recombination frequency, LOD score and phase for all affected linkage groups and then remapped these chromosomes using the amended estimates. For the vast majority of markers in the dataset there was no change in the assignment to chromosomes or homologues (i.e. marker phasing was unaffected). We found almost no difference in map order (Figure 3) but there were some differences in map lengths (between −0.6 and 5.2 cM longer). We expected maps to be on average longer from previous work with simulated data, which showed that ignoring preferential pairing when it occurs will lead to an under-estimation of recombination frequency (Bourke *et al.*, 2016). In this population, the most extreme example of preferential

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.,

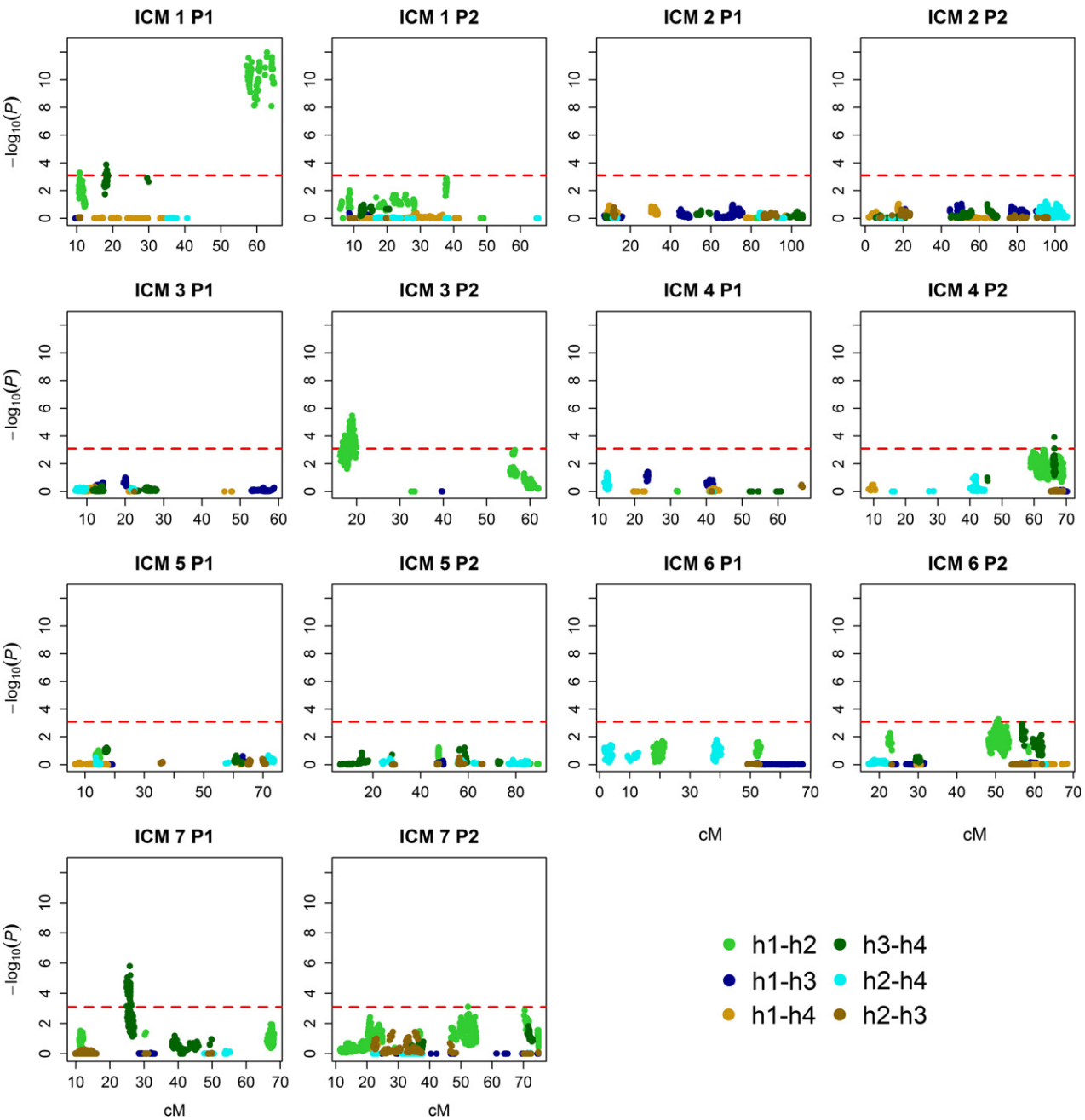


Figure 2. Evidence for both disomic and tetrasomic behaviour from pairs of closely mapped repulsion-phase S × N markers.
The $-\log_{10}(P)$ values of a binomial test against the hypothesis $r_{r_d} \geq \frac{1}{3}$ for S × N marker pairs in repulsion are plotted against their approximate genetic position. Regions of disomic behaviour [high values of $-\log_{10}(P)$] are not uniformly distributed, suggesting differences in pairing affinities along chromosomes. The significance threshold is shown as a dashed red line. h1–h2 refers to marker pairs from homologues 1 and 2. For simplicity, homologues 5–8 in parent 2 have been labelled 1–4. [Colour figure can be viewed at wileyonlinelibrary.com].

pairing was found on ICM 1 of P1 (ρ = 0.302, or about 45%), for which the corrected map was 5.21 cM longer.

All linkage groups were densely covered with markers, with the largest gap being 4.32 cM on ICM1 (Table 3). The final integrated map positions of 25 695 markers, using the corrections for ICM1, ICM3 and ICM4 are provided in Table S4. The closest-linked S × N marker positions on the SNP map to a set of previously mapped 1:1 segregating simple sequence repeat (SSR) and AFLP markers (Koning-Boucoiran *et al.*, 2012) are provided in Table S5, which facilitated numbering the linkage groups according to the ICM numbering (Spiller *et al.*, 2011).

© 2017 The Authors

Table 2 Significant preferential pairing identified using closely-mapping S × N marker pairs in repulsion

Parent	Chm	h_a	h_b	ρ	SD[a]	*N*[b]	N_s[c]
1	1	1	3	0.35	0.14	612	360
1	1	2	4	0.23	0.02	88	16
2	3	1	2	0.22	0.07	4676	2388
2	4	2	3	0.17	0.05	106	4
2	6	1	2	0.17	0.02	6760	4
1	7	2	3	0.16	0.07	2414	152
2	7	1	2	0.06	0.05	7538	2

[a]The standard deviation of the estimate for ρ, the preferential pairing parameter between homologues h_a and h_b.
[b]The number of non-negative recombination estimates on which the estimate of ρ was based.
[c]The number of repulsion-pairs that showed significant evidence for non-tetrasomic behaviour.
Chm, Chromosome.

Prediction of quadrivalents and double reduction

When TetraOrigin was rerun with the possibility of quadrivalent pairing, we found a relatively high proportion of quadrivalents predicted across all chromosomes in both parents, ranging from 38 to 64% in P1 and 44 to 67% in P2 (Table S6). At a significance threshold of 0.001, we still found significant deviations from random pairing on ICM1, ICM3 and ICM4 (in the same parents, between the same homologues), but also on ICM6 of P1 and ICM7 of P2. In ICM6 of P1, the significant score stemmed from an under-representation of one bivalent pairing configuration. There was also some evidence for preferential pairing on ICM7 of P2, similar to the prediction made using the repulsion-phase S × N markers (which was between the same two homologues). In other words, the results are consistent with those from the bivalent pairing model as well as from the analysis of repulsion-phase linkages, with a similar set of chromosomes and parents implicated each time. The double reduction plots (Figure S1) were more jagged than those from a previous analysis of potato (Bourke *et al.*, 2015), suggesting some difficulty in predicting double reduction in this dataset, although we did observe an increase in the rate of double reduction away from the centromeres, reaching a maximum of between 5% and 10% at the telomeres (Figure S1).

Haplotype diversity and pairing behaviour

Of the 25 695 mapped markers we could confidently assign full phase information to 20 090 of them (at least 10 linkages with LOD > 3 to a S × N homologue cluster), provided in Table S7. This set of markers was used to visualise the distribution of markers over the parental homologues (Figure S2). Using this marker subset, we compared the level of homology between homologues. There are 28 possible pairwise homologue combinations in a tetraploid cross, with 12 of these comparing within-parent haplotype diversity and 16 comparing between-parent diversity. There appeared to be no difference between the average within-parent homology versus between-parent homology (Figure S3) suggesting that both parents carry haplotypes from the same breeding pool. The dissimilarity between haplotypes per parent varied greatly per chromosome, but could not be used as a predictor of pairing behaviour (Figure S4). That said, for ICM1 of P1 we observed a high level of homology between h1 and h2 from 20 to 50 cM, and a high level of homology between h3 and h4 from 50 to 80 cM, consistent with the predicted preferential pairing conformation.

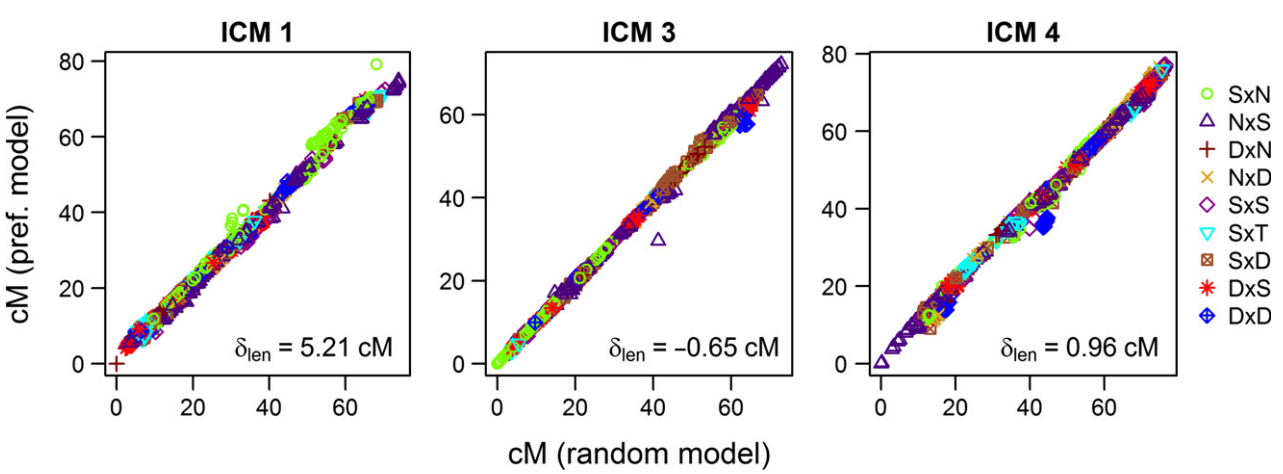


Figure 3. Comparison between the integrated genetic map (in centimorgans, cM) under a random bivalent model (*x*-axis) and a bivalent model that takes account of preferential pairing (*y*-axis) for rose chromosomes 1, 3 and 4.
δ_{len} refers to the difference between the tetrasomic and heterosomic map lengths. Symbols are given for each of the nine marker segregation types encountered in a tetraploid. Correcting for mild preferential pairing in an otherwise tetrasomic species has little impact on the map order or length. [Colour figure can be viewed at wileyonlinelibrary.com].

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.,

Table 3 Numbers of markers mapped per linkage group on the ultra-high density *Rosa hybrida* map. These maps represent the highest-density linkage maps in the genus Rosa published to date, helping to provide linkage information for current genome assembly efforts

ICM	Length (cM)	No. of markers	Max. gap (cM)
1	79.19	1865	4.32
2	108.67	6154	1.00
3	72.16	2912	1.18
4	77.30	2866	3.48
5	89.76	3799	0.69
6	71.82	4193	2.02
7	74.76	3906	0.53

ICM, integrated consensus map.

Synteny analysis with *Fragaria vesca*

Synteny has already been reported between *Rosa* and woodland strawberry (Gar *et al.*, 2011; Kirov *et al.*, 2014; Vukosavljev *et al.*, 2016). In this study, we provide the most detailed picture of this synteny to date (Figure 4a), which helps confirm the marker order of our maps. *Fragaria* chromosomes Fv1 and Fv6 are closely related to rose ICM2 and ICM3, with what appears to have been a reciprocal translocation occurring between them. We identified a major telomeric inversion between the gene orders on strawberry chromosome 5 and rose ICM7, as well as a second possible inversion on rose ICM6 or strawberry Fv2 (compare the outer track of Figure 4a). There is also a small fragment of the strawberry genome assembly that is currently unassigned (Fv0) with synteny to rose ICM5 from 34.8 to 48.1 cM. This suggests that this unassembled fragment should form part of chromosome Fv3, between 10.5 and 14.9 Mb. We were unable to locate the pericentromeric regions of rose from this comparison; this may be due to the method of marker development for the WagRhSNP array, which used transcriptome data (targeting gene-rich euchromatin rather than heterochromatin) to identify marker sequences.

A set of 176 non-segregating D $\times$ N markers and 7 non-segregating N $\times$ D markers, which were originally filtered from the dataset, were double-checked against the BLAST results. Although these markers cannot be genetically mapped, they provide an indication of disomic behaviour (since they are present on exclusively pairing homologues). Eighty-one of these markers produced a BLAST hit (72 of which had a BLAST E-value $<1 \times 10^{-20}$), and all 81 mapped to *Fragaria* chromosome 7, for which we have unambiguous evidence for synteny with rose ICM1 (Figure 4a). Their putative positions ranged from 12.14 to 19.87 Mb on the strawberry physical map (Table S8), which corresponds to the genetic region 47.7–65.9 cM on rose ICM1. Interestingly, this is precisely the region for which the most highly disomic repulsion-phase estimates were previously identified (Figure 4b) and they therefore provide further evidence for highly preferential pairing on this chromosome in this parent, which may be confined to a specific chromosomal region. Regarding the non-segregating N $\times$ D markers, only two produced BLAST hits, one to Fv4 and the other to Fv5 (Table S8), which show synteny to rose ICM4 and ICM7 respectively, consistent with our previous findings.

DISCUSSION

Non-uniform distribution of preferential pairing

The observations in this study suggest that preferential pairing is not uniformly distributed in the rose genome. We detected differences in pairing behaviour between parents, among chromosomes and even along a single chromosome where pairing behaviour can be preferential at one chromosome arm but tetrasomic at the other. The original description of segmental allopolyploidy does not preclude this sort of behaviour (Stebbins, 1947). Indeed, in keeping with Stebbins' description, *Rosa* displays many characteristics of an autopolyploid, with evidence for quadrivalent pairing and double reduction as well as the majority of our marker data fitting a tetrasomic segregation model. We have established with this work that rose may be categorised as a segmental allotetraploid, although it is impossible to predict whether this is ultimately a stable conformation (Sybenga, 1996) and whether it is genus-wide rather than population-specific. We have uncovered evidence that the strength of pairing can vary along a chromosome, potentially complicating genetic studies and methodologies, which invariably start with some assumptions about the uniformity of pairing behaviour. This is not the first reported instance of intra-chromosomal 'mixosomy'; in rainbow trout it has been proposed that there may be 'residual tetrasomy' (Allendorf and Danzmann, 1997) leading to variable pairing behaviour along a chromosome, albeit with disomic regions confined to the more central regions. A similar phenomenon has also recently been reported in peanut (Leal-Bertioli *et al.*, 2015; Nguepjop *et al.*, 2016). It is possible that the pairing behaviour in cultivated tetraploid *Rosa* is a consequence of its rather exotic pedigree, similar to an observation on the increased occurrence of mixed segregation patterns in trout populations derived from interspecific hybridisation (Allendorf *et al.*, 2015). On the other hand, if rose is the result of hybridisation among 7–14 species, it is surprising that most chromosomes of this cross exhibit random pairing. Species hybridisation may exhibit tetrasomic pairing if the underlying species are young, as suggested by a previous study of the low levels of variation in ribosomal and chloroplast markers in rose (Wissemann and Ritz, 2005). Cultivated roses have an unusual recent pedigree (which ornamental species in particular, but also many other

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.,

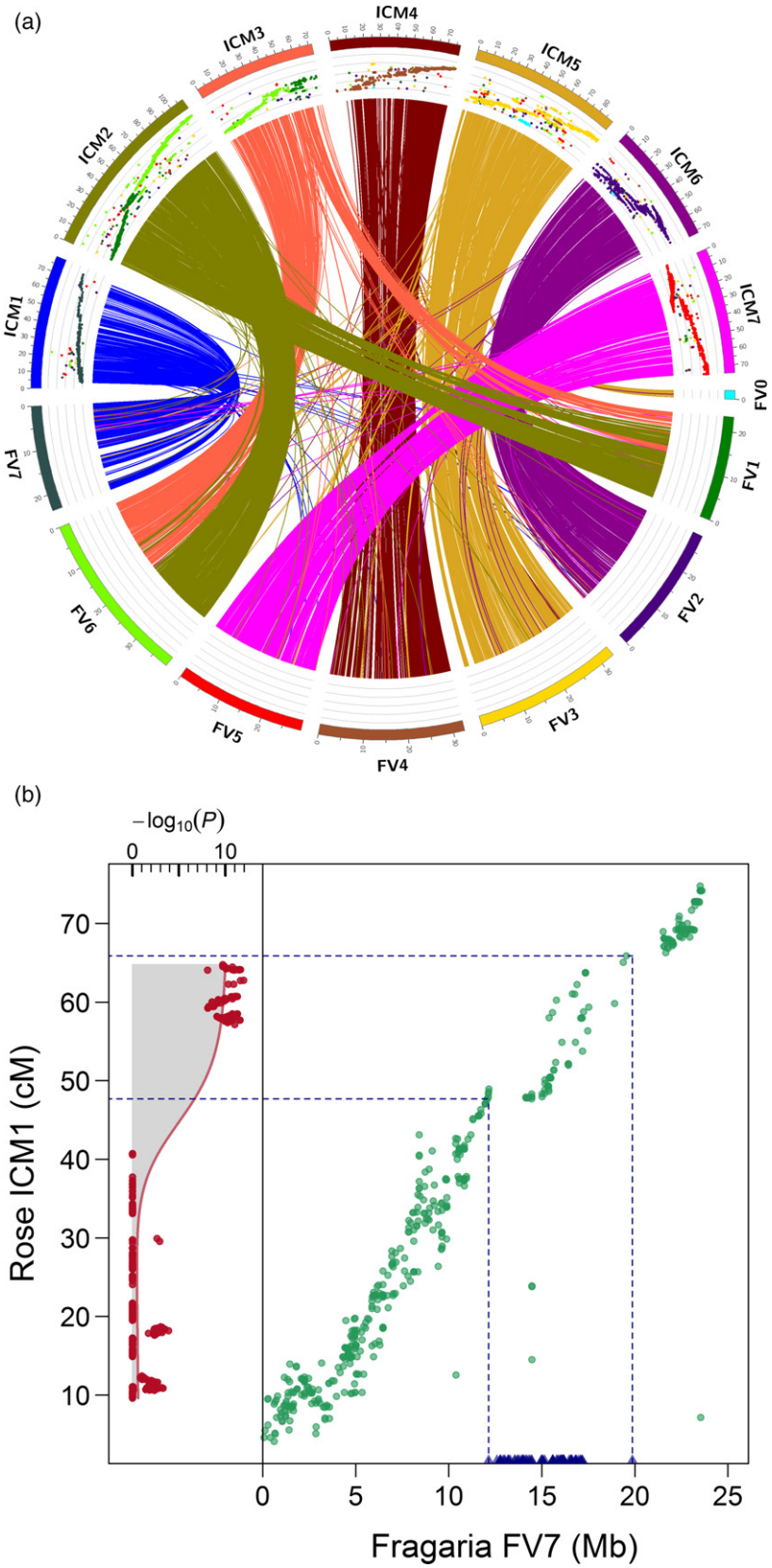


Figure 4. Overview of the synteny between *Rosa hybrida* and *Fragaria vesca*. (a) Synteny between integrated genetic linkage maps of cut rose and *F. vesca* pseudomolecules v.2.0.a1, with precise correspondence of position displayed in the outer track. FV0 indicates unassigned assembly of *F. vesca* (available from http://www.rosaceae.org). Rose genetic linkage group numbering follows the integrated consensus map (ICM) numbering of rose. Gene order is highly conserved between these species, and a number of interested features such as inversions (ICM6 and ICM7) and reciprocal translocations (ICM2 and ICM3) are clearly visible. (b) Detailed comparison of the syntenic relationship between *Fragaria* chromosome 7 and rose ICM1, with non-segregating D × N markers shown as triangles on the *x*-axis. Non-segregating D × N markers suggest complete disomic behaviour. These sets of markers would have mapped to the same region of rose ICM1 that contains strong evidence for disomic behaviour from repulsion-phase S × N marker pairs, confirming that this area differs in its pairing affinity from the rest of the chromosome. [Colour figure can be viewed at wileyonlinelibrary.com].

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.,

domesticated crop species, have experienced), but the same may be true, for example, for naturally occurring polyploid roses (including the unbalanced meiosis in pentaploid dog roses; Herklotz and Ritz, 2016). The genus *Rosa* is more than 30 million years old, but hybridsation between species has been extensive (Koopman *et al.*, 2008; Fougère-Danezan *et al.*, 2015), which may have led to widespread mixing of germplasm within hybrid rose varieties. Indeed, Zhang *et al.* (2013) found the same microsatellite haplotypes with flanking SNP markers in diploid and polyploid rose species, while Bruneau *et al.* (2007) found the same chloroplast gene haplotypes in multiple species. A more precise answer may therefore come from genome sequences of various diploid and polyploid accessions.

It is also worth noting here that preferential pairing and marker skewness are different phenomena that may be confused. Marker skewness can be caused by selection and can cause problems in linkage map construction (Van Ooijen and Jansen, 2013). For example, lethality or reduced fitness is unlikely to be caused by a combination of alleles present in one of the parents (since that parent is alive); therefore cross-parental lethal combinations are more probable, which will not result in a preferential pairing signal in one of the parents. Similarly, if selection is positive (e.g. at a resistance locus) then the skewness is not in favour of a particular *combination* of parental homologues but rather of specific allele(s); in cases where there are two functional copies in a single parent the effect would appear as selection *against* a particular homologue combination (i.e. 'unpreferential' pairing). In any case, in this study we only mapped non-skewed markers, so that all our evidence for preferential pairing comes from markers with no significant skewness. This is also visualised in Figure S5, showing that skewness levels among mapped markers were essentially uniform across the rose genome.

Predicting pairing behaviour through homology

We were interested to see whether pairing behaviour might be predicted from the homology as defined by our haplotype-specific linkage maps. We chose a simple dissimilarity measure to analyse the diversity between homologues in this study, similar to the most recent common ancestor (MRCA) measure used in population genetics studies (Joly *et al.*, 2015). We did not observe any clear relationship between homology as estimated from the mapped markers and pairing behaviour. However, as in all mapping studies, our data were biased towards higher diversities since only segregating markers can be included on a genetic map. We should therefore treat these results with caution as they do not give a complete picture of homology, which only sequencing can fully reveal. Telomeric homology, where pairing initiation is thought to occur (Sved, 1966; Sybenga, 1975; Cifuentes *et al.*, 2010), might have more influence than chromosome-length homology and could be a target of future sequencing efforts to clarify this.

Tailored genetic linkage maps

In this study we generated genetic linkage maps tailored to the particular pairing behaviour of each chromosome. We could have further refined our maps by allowing the estimate for the rate of preferential pairing to vary along the chromosome arms if this was deemed necessary. However, in this population the rate of preferential pairing on the affected chromosomes was sufficiently low that we could have ignored it in map construction. We can confirm our prediction that linkage mapping can be safely performed under the assumption of random pairing up to a level of preferential pairing as high as 70% ($\rho = 0.467$), after which estimates become seriously biased (Bourke *et al.*, 2016). This finding is likely to be welcomed by researchers in *Rosa*, for which the assumption of tetrasomic inheritance has generally been used. In this study, we used D $\times$ N markers to provide links between putative homologue clusters of S $\times$ N markers (helping to identify chromosomal linkage groups). However, these may not always be informative in cases of extreme preferential pairing, where simplex $\times$ simplex (S $\times$ S) markers may need to be used to provide cross-homologue linkages instead.

Synteny between *Rosa* and *Fragaria*

In this study we used the close relationship between *Rosa* and *Fragaria* to confirm the order and organisation of the linkage maps produced, as well as allowing us to position a set of non-segregating D $\times$ N markers which were indicative of extreme preferential pairing in a distal region of maternal chromosome ICM1. It also yielded a detailed picture of the high level of genomic conservation between these species (Figure 4a,b). Both *Rosa* and *Fragaria* belong to the Rosoideae clade (in the Potentilleae and Roseae tribes, respectively), whose divergence is estimated to have occurred approximately 60 million years ago (Mya) (Xiang *et al.*, 2017). High levels of gene conservation have also been reported in the grass family, for example (Bennetzen, 2007). On the other hand, despite an estimated divergence time of 27 Mya, there is a much more striking difference in the genomic configuration of *Brassica rapa* and *Arabidopsis thaliana* (Murat *et al.*, 2015), although this could be due to the fact that Brassicaceae possess an exceptionally high rate of biological radiation and diversification (Franzke *et al.*, 2011). It is tempting to speculate on why rates of genomic divergence differ between plant families (e.g. the level of activity of transposable elements within the genome; Bennetzen and Wang, 2014), although such a meta-analysis using data across multiple plant families has yet to appear in the literature. What our study does make clear is that the published reference sequence of

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.,

F. vesca continues to be a very useful genomic resource for the rose research community, at least until such time as a rose sequence becomes publically available.

Concluding remarks

In this study we have applied a novel approach for detecting and quantifying the level of preferential pairing in a tetraploid biparental population of cut rose, and used this information to refine our estimates of recombination to produce a tailored genetic map of this economically important species. We found conclusive evidence for parent-specific preferential pairing behaviour on three chromosomes whereas the other chromosomes showed tetrasomic behaviour. The picture that emerges is that the meiotic behaviour in polyploids may be difficult to predict without detailed analysis. We found that meiotic pairing behaviour can vary from chromosome to chromosome and from genotype to genotype, and perhaps even along chromosome arms. Tools which utilise chromosome-wide inheritance information across all homologues are very powerful for revealing underlying meiotic mechanisms and confirm our findings using repulsion-phase linkage information as well as information from syntenic mapping of non-segregating duplex markers to infer inheritance type. We uncovered a detailed picture of conserved synteny with the closely related woodland strawberry, highlighting both telomeric inversions and reciprocal translocations which occurred somewhere in the lineages of one or both of these species. Our ultra-high-density linkage map will facilitate downstream applications such as current efforts at genome assembly as well as providing a basis for detailed QTL analysis in the future.

EXPERIMENTAL PROCEDURES

Plant material, DNA isolation and genotyping

The tetraploid 'K5' cut rose mapping population, consisting of 172 individuals of the cross between 'P540' (mother) and 'P867' (father), was used in this study (Figure 1c). This population has previously been used in studies on powdery mildew (*Podosphaera pannosa*) resistance (Yan *et al.*, 2006), a range of morphological traits (Gitonga *et al.*, 2014, 2016), stomatal functioning (Carvalho *et al.*, 2015) and linkage map construction using AFLP and SSR markers (Koning-Boucoiran *et al.*, 2012). Genomic DNA was extracted from freeze-dried young leaves, using the DNeasy Plant Mini Kit (Qiagen, http://www.qiagen.com/) following the protocol of Esselink *et al.* (2003). Samples were sent to Affymetrix (http://www.affymetrix.com/) for genotyping using the WagRhSNP 68K Axiom SNP array (Koning-Boucoiran *et al.*, 2015). This array targets 68 893 SNPs with every SNP targeted with two probes (which we refer to as the P and Q probes). Both parents were genotyped in triplicate (two biological replicates and one technical replicate).

Genotype calling and data preparation

The SNP array data were converted into dosage scores using the fitTetra package (Voorrips *et al.*, 2011). Quality checks were subsequently performed to ensure consistency between the parental scores and those of the offspring, and to check for the possibility of 'shifts' in dosage assignments which can occur when fewer than the five possible dosage classes occur. As each SNP was targeted with two probes (P and Q), the genotype calls for these probe pairs had to be compared and merged if they were found to be consistent (<10% conflicts, where missing values were not considered conflicting), with conflicting scores made missing. Probes with more than 10% conflicting scores (including parental scores) were kept as separate markers, by appending the letters 'P' and 'Q' to the marker names. Markers with more than 10% missing values were removed from the dataset, as well as those markers showing highly skewed segregation patterns under either a tetrasomic or disomic model ($P < 0.001$).

In a tetraploid genotyped using bi-allelic SNP markers, the possible dosage classes range from nulliplex (zero copies of the alternative allele, coded 'N') to quadruplex (four copies, coded 'Q'), with marker segregation types defined by their parental scores. Marker dosage scores were recoded as described in Bourke *et al.* (2016), resulting in nine marker classes (S × N, N × S, D × N, N × D, S × S, S × T, S × D, D × S, D × D). Duplicate individuals were identified by genotype pairs with an unusually high correlation (>95% similar scores); these individuals were merged (if both scores were present and conflicting, the merged score was made missing). Individuals with more than 10% missing values were also removed.

Marker clustering and linkage group assignment

Initially, the simplex × nulliplex (S × N) and nulliplex × simplex (N × S) markers were clustered using the LOD for linkage. A LOD value of 4 was found to split the data evenly, with clusters of 15 or more markers selected as candidate homologue groups. In P1 this resulted in 29 clusters, one more than the 28 expected. Each cluster was then tested over a range of thresholds (from LOD 4 to LOD 10) to ensure the markers remained clustered together. In P2 at a LOD threshold of 4, 28 clusters were identified. Two of these clusters split at LOD 5, resulting in 30 P2 clusters. These clusters were subsequently assigned to linkage groups using their linkage to D × N (or N × D) markers, which provide cross-homologue linkage. Where more than four putative homologues were present, the predicted phasing across clusters was used to merge these into single homologues. In total, 28 P1 homologues across seven chromosomes were identified, and 27 P2 homologues across seven chromosomes in P2 (i.e. we missed one P2 homologue). Chromosomes were renumbered according to the ICM numbering previously introduced (Spiller *et al.*, 2011), through linkage with SSR markers from a previous study (Koning-Boucoiran *et al.*, 2012). All other marker segregation types were subsequently assigned to both chromosomes and homologues based on their linkage to S × N markers (LOD > 3).

Linkage analysis and map construction under a tetrasomic model

Pairwise linkage analysis between all marker types was performed per chromosome and per parent. The maximum likelihood framework used to estimate the (phased) recombination frequency and LOD score under the assumption of random bivalent pairing (i.e. no double reduction and no preferential pairing) has already been described elsewhere (Hackett *et al.*, 2013; Bourke *et al.*, 2016). Pairwise marker phasing was primarily based on likelihood maximisation (Hackett *et al.*, 2013), although minimum recombination

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.,

frequency was used to phase S × S marker pairs (as described in Bourke *et al.* (2016)).

A prototype version of the MDSMAP software was kindly made available by its authors for use in map ordering (Preedy and Hackett, 2016). Maps were produced using unconstrained weighted metric multidimensional scaling (with LOD^2 as weights and using Haldane's mapping function) followed by principal curve fitting in two dimensions. Poorly mapping markers were identified either as outliers in the PCO plots (judged by eye) or as those with a nearest-neighbour fit that exceeded 5. Such markers were removed, and up to three rounds of MDSMAP were performed until stable maps free of outlying markers were produced. The final map positions of the 705 pairs of (unmerged) P and Q probes were compared. In 235 cases, one of the probes was lost (either during the clustering stage or later when outliers were removed from the linkage maps) and the remaining probe was taken as the consensus marker. In cases where both probes were mapped, only 148 mapped less than 1 cM apart and had the same parental dosages; these probes could be merged while the rest were discarded. The mapping procedure was repeated using the amended dataset.

Estimation of a preferential pairing parameter

We modelled preferential pairing in the context of bivalent pairing by considering deviations ρ from the expected probabilities under a random model, where the probability of pairing between homologues 1 and 2, and between 3 and 4 (denoted 12/34) is $\frac{1}{3} + \rho$, and that between 13/24 or 14/23 is $\frac{1}{3} - (\rho/2)$. We avoided simultaneously estimating a preferential pairing factor and a recombination frequency (which was found to lead to an overestimation of the level of preferential pairing; Wu *et al.*, 2002), in favour of a two-step procedure that first estimates a map using two-point tetrasomic linkage analysis and then revises those estimates when a preferential pairing factor has been determined using a multipoint hidden Markov model (HMM) approach (Zheng *et al.*, 2016).

The marker data were simplified by rounding marker positions to the nearest cM, after which one of each marker segregation type was selected at each cM position. Where more than one marker of a particular type was present at a locus, the marker with the fewest missing values was chosen. TETRAORIGIN provides the most likely bivalent pairing in each individual (classes 12/34, 13/24 or 14/23). We renumbered homologues such that the pairing configuration with the highest predicted count in TETRAORIGIN was 12/34 etc. We then applied a χ^2 test on the counts of each class to test for deviations from $\frac{1}{3}$ ($P < 0.001$). If the number of structures predicted in each of the three pairing classes are n_1, n_2 and n_3 and assuming $n_1 \geq n_2$ and $n_1 \geq n_3$, the likelihood function given the observed counts is:

$$\mathcal{L}(\rho) \propto \left(\frac{1}{3}+\rho\right)^{n_1}\left(\frac{1}{3}-\frac{\rho}{2}\right)^{n_2+n_3}$$

Solving the likelihood equation leads to the maximum likelihood estimate for ρ:

$$\hat{\rho} = \frac{\frac{2}{3}n_1 - \frac{1}{3}(n_2+n_3)}{n_1+n_2+n_3}$$

TETRAORIGIN was re-run, allowing for the possibility of quadrivalent pairing in the offspring, to investigate the level of double reduction and to see whether preferential pairing was still predicted under a model that included quadrivalents. The rate of double reduction was determined per locus as the average of the haplotype probabilities that exceeded 1 across the population.

Preferential pairing estimated from repulsion-phase linkages

We estimated the error rate in the genotype data from pairs of duplicated individuals, of which 19 pairs were identified. The approximate error rate was taken as the average conflict rate over all duplicate pairs. We followed Brzustowicz *et al.* (1993) in their definition of the correspondence between the true (θ_0) and apparent (φ) rates of recombination given an error rate s (Brzustowicz *et al.*, 1993; Hackett and Broadfoot, 2003)

$$\phi = \theta_0(1-s) + (1-\theta_0)s$$

The minimum resolution of recombination frequency in a mapping population of size N and missing value rate of μ is $r_{\min} \approx [2(1-\mu)N]^{-1}$ (Bourke *et al.*, 2016), assuming error-free data. In the presence of errors, the above equation then implies:

$$r_{\min} \approx \frac{1-2\mu Ns}{2(1-\mu)N}$$

We identified all pairs of S × N or N × S markers that mapped within this distance ($r = 0.0117$) on different homologues (repulsion pairs). Under a purely disomic model, the repulsion-phase maximum likelihood estimate for the recombination frequency is given by $r_{\text{disom}} = (n_{00} + n_{11})/(n_{00} + n_{01} + n_{10} + n_{11})$, where n_{01} is the number of offspring with a dosage of 0 at marker A and a dosage of 1 at marker B etc. If inheritance is tetrasomic, this estimate never falls below $\frac{1}{3}$ (Qu and Hancock, 2001; Bourke *et al.*, 2015). This forms the basis of a binomial test (H_0:$r_{\text{disom}} \geq 1/3$), corrected for multiple testing using the FDR with $\alpha = 0.05$ (Benjamini and Hochberg, 1995). This identified cross-homologue pairs that showed significant deviations from a tetrasomic model. We then estimated a preferential pairing parameter for the homologue combinations that showed significant evidence of preferential pairing. For two repulsion-phase S × N markers A and B located at the same genetic position, the recombination frequency estimate between them is a function of independent assortment with no contribution from cross-overs (Qu and Hancock, 2001). Given a preferential pairing parameter ρ, the probabilities of observing each of the classes n_{00}, n_{01}, n_{10}, n_{11} in the offspring are

$$\frac{1}{6}-\frac{\rho}{4}, \frac{1}{3}+\frac{\rho}{4}, \frac{1}{3}+\frac{\rho}{4}, \frac{1}{6}-\frac{\rho}{4}$$

respectively. This leads to the following equation:

$$\frac{\partial}{\partial\rho}(\ln\mathcal{L}(\rho)) \propto (n_{00}+n_{11})\left(\frac{-1}{4\left(\frac{1}{6}-\frac{\rho}{4}\right)}\right) + (n_{01}+n_{01})\left(\frac{1}{4\left(\frac{1}{3}+\frac{\rho}{4}\right)}\right) = 0$$

Solving for ρ yields the maximum likelihood estimate for the preferential pairing parameter:

$$\hat{\rho} = \frac{2(n_{00}+n_{11}) - 4(n_{01}+n_{01})}{3(n_{00}+n_{01}+n_{10}+n_{11})}$$

We recorded the mean (and standard deviation) of the non-negative estimates to generate pairing-specific estimates of ρ.

Linkage analysis of a segmental allopolyploid

Having estimated the strength of preferential pairing for all affected linkage groups, we re-estimated the pairwise recombination frequencies for these cases. The maximum likelihood estimates for recombination frequency were derived in Mathematica (Wolfram Research Inc., 2014) and exported for use in R (R Core Team, 2016). A description of the likelihood model including preferential pairing is provided in Methods S1. Markers were

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.,

subsequently reassigned to chromosomes and homologues given the updated information, and the maps were re-calculated using MDSMAP (Preedy and Hackett, 2016).

Visualisation and quantification of haplotype diversity

A subset of 20 431 markers for which there were at least 10 significant linkages (LOD > 3) to each of the expected number of homologue clusters was used to define haplotypes. These allowed us to examine pairwise diversity between homologues to investigate the relationship between homologue similarity and pairing behaviour. We defined a simple dissimilarity measure between homologues A and B within a window centred at marker position *j* as:

$$\mathcal{D}_{\mathrm{A-B},j} = \frac{1}{N_j}\sum_{i=1}^{N_j} |d_{\mathrm{A},i} - d_{\mathrm{B},i}|$$

where $d_{A,i}$ is the SNP allele score of homologue A at marker position *i* (either '0' or '1') and N_j is the number of markers within a 5-cM sliding window around each marker position *j*.

Integration with SSR and AFLP map

A genetic linkage map of the same population using SSR and AFLP markers has already been published (Koning-Boucoiran *et al.*, 2012). We initially attempted to map all markers together, but found that the SSR/AFLP markers tended to cluster together at the ends of a linkage group, suggesting high tension between the two marker datasets. Linkage between SSR/AFLP markers and the set of S × N (or N × S) markers was evaluated, with the three most significant linkages recorded, giving these markers approximate positions on the SNP map.

Synteny analysis with *F. vesca*

The expressed sequence tags (ESTs) from which the WagRhSNP 68K Axiom SNP array markers were derived were used in a BLASTn search against the woodland strawberry genome assembly (*F. vesca* v.2.0.a1 pseudomolecules, downloaded on 19 September 2016 from http://www.rosaceae.org) with 'N' used at the SNP position, as provided in Supplementary Table 2 of Koning-Boucoiran *et al.* (2015). An E-value threshold of 1×10^{-20} was used to retain only highly homologous hits, with multiple hits filtered out to avoid targeting multigene families. The rose linkage maps were orientated according to the order on the *Fragaria* genome sequence (as in Vukosavljev *et al.*, 2016), and the resulting synteny was visualised using CIRCOS software (Krzywinski *et al.*, 2009). The most likely map positions of a set of non-segregating D × N markers were also determined based on the position of their hit in the strawberry genome, providing additional indications for disomic segregation.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Christine Hackett and Dr Katherine Preedy (BioSS) for sharing a developmental version of MDSMAP, which was used in this study. The authors also wish to acknowledge the two anonymous referees whose comments helped improve the manuscript. Funding for this research was provided through the TKI polyploids project 'A genetic analysis pipeline for polyploid crops', project number BO-26.03-002-001. The authors declare they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

PA, CKB, WPCvtW and MJMS generated the SNP data; PMB, REV, and CM developed the methodology; PMB, CZ and GvG contributed to software development, PMB, REV, GDE, TSL, PW and CM analysed the data; PMB wrote the manuscript; PA, REV, RGFV, FAK, MJMS and CM edited the manuscript. All authors read and approved the final manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Average estimated probability of double reduction per parental meiosis over the seven rose chromosomes.

Figure S2. Visualisation of meiotically tailored high-density linkage maps, showing the distribution of segregating marker alleles across all eight parental homologues for each of the seven rose linkage groups (ICM1–ICM7), also listed in Table S7 using original marker dosage coding.

Figure S3. Average within-parent and between-parent dissimilarity (*D*) for each of the seven rose chromosomes. Multiple points correspond to each set of homologue pairs within these categories.

Figure S4. Average pairwise dissimilarity between parental haplotypes, plotted against genetic position.

Figure S5. Proportion of offspring with simplex scores (dosage 1) at 10 651 S × N and N × S marker positions across the rose linkage maps.

Table S1. Breakdown of probes and single nucleotide polymorphism marker numbers during genotyping and filtering steps.

Table S2. Number of single nucleotide polymorphism markers in each of the nine marker classes after filtering and dosage conversion.

Table S3. Comparison of all pairs of closely-mapping S × N markers in repulsion, giving the estimated strength of preferential pairing per pair.

Table S4. Genetic map positions of 25 695 single nucleotide polymorphism markers on the integrated ultra-high-density linkage map of rose.

Table S5. Top three significant linkages between a set of previously-mapped simple sequence repeat and AFLP markers and the 1:1 segregating single nucleotide polymorphism markers mapped in this study, with estimated recombination frequency, LOD and phase.

Table S6. Output of TETRAORIGIN when quadrivalent pairing is allowed showing large numbers of predicted quadrivalents, but still showing non-random pairing of the predicted bivalents.

Table S7. Phased marker positions of 20 090 single nucleotide polymorphism markers for which highly supported phase information (>10 significant linkages to a homologue) was available.

Table S8. BLASTn results of 81 non-segregating D × N and 2 non-segregating NxD markers, showing their most likely position on the *Fragaria vesca* physical map.

Methods S1. Derivation of pairwise maximum likelihood estimates for recombination frequency, LOD score and phase given prior estimates for chromosome and parent-specific preferential pairing factor(s) ρ.

REFERENCES

Al-Janabi, S.M., Honeycutt, R.J. and Sobral, B.W.S. (1994) Chromosome assortment in Saccharum. *Theor. Appl. Genet.* **89**, 959–963.

Allendorf, F.W. and Danzmann, R.G. (1997) Secondary tetrasomic segregation of *MDH-B* and preferential pairing of homeologues in rainbow trout. *Genetics*, **145**, 1083–1092.

© 2017 The Authors

Allendorf, F.W., Bassham, S., Cresko, W.A., Limborg, M.T., Seeb, L.W. and Seeb, J.E. (2015) Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. *J. Hered.* **106**, 217–227.

Bartholomé, J., Mandrou, E., Mabiala, A., Jenkins, J., Nabihoudine, I., Klopp, C., Schmutz, J., Plomion, C. and Gion, J.-M. (2015) High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytol.* **206**, 1283–1296.

Bendahmane, M., Just, J., Vergne, P. *et al.* (2016). The Rose Genome Sequencing Initiative, Prospects and Perspectives. Abstract W659. In Plant and Animal Genome XXIV Conference (https://pag.confex.com/pag/xxiv/webprogram/Paper19873.html).

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B (Methodological)*, **57**, 289–300.

Bennetzen, J.L. (2007) Patterns in grass genome evolution. *Curr. Opin. Plant Biol.* **10**, 176–181.

Bennetzen, J.L. and Wang, H. (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* **65**, 505–530.

Bingham, E.T. and Gillies, C.B. (1971) Chromosome pairing, fertility, and crossing behavior of haploids of tetraploid alfalfa, *Medicago sativa* L. *Can. J. Genet. Cytol.* **13**, 195–202.

Bomblies, K., Higgins, J.D. and Yant, L. (2015) Meiosis evolves: adaptation to external and internal environments. *New Phytol.* **208**, 306–323.

Bourke, P.M., Voorrips, R.E., Visser, R.G.F. and Maliepaard, C. (2015) The double reduction landscape in tetraploid potato as revealed by a high-density linkage map. *Genetics*, **201**, 853–863.

Bourke, P.M., Voorrips, R.E., Kranenburg, T., Jansen, J., Visser, R.G. and Maliepaard, C. (2016) Integrating haplotype-specific linkage maps in tetraploid species using SNP markers. *Theor. Appl. Genet.* **129**, 2211–2226.

Bruneau, A., Starr, J.R. and Joly, S. (2007) Phylogenetic relationships in the genus *Rosa*: new evidence from chloroplast DNA sequences and an appraisal of current knowledge. *Syst. Bot.* **32**, 366–378.

Brzustowicz, L., Merette, C., Xie, X., Townsend, L., Gilliam, T. and Ott, J. (1993) Molecular and statistical approaches to the detection and correction of errors in genotype databases. *Am. J. Hum. Genet.* **53**, 1137–1145.

Cao, D., Osborn, T.C. and Doerge, R.W. (2004) Correct estimation of preferential chromosome pairing in autotetraploids. *Genome Res.* **14**, 459–462.

Carvalho, D.R., Koning-Boucoiran, C.F., Fanourakis, D., Vasconcelos, M.W., Carvalho, S.M., Heuvelink, E., Krens, F.A. and Maliepaard, C. (2015) QTL analysis for stomatal functioning in tetraploid *Rosa*× *hybrida* grown at high relative air humidity and its implications on postharvest longevity. *Mol. Breeding*, **35**, 1–11.

Cifuentes, M., Grandont, L., Moore, G., Chèvre, A.M. and Jenczewski, E. (2010) Genetic regulation of meiosis in polyploid species: new insights into an old question. *New Phytol.* **186**, 29–36.

Da Silva, J.A.G., Sorrells, M.E., Burnquist, W.L. and Tanksley, S.D. (1993) RFLP linkage map and genome analysis of *Saccharum spontaneum*. *Genome*, **36**, 782–791.

Debener, T. and Linde, M. (2009) Exploring complex ornamental genomes: the rose as a model plant. *Crit. Rev. Plant Sci.* **28**, 267–280.

Doyle, J.J. and Egan, A.N. (2010) Dating the origins of polyploidy events. *New Phytol.* **186**, 73–85.

Doyle, J.J. and Sherman-Broyles, S. (2017) Double trouble: taxonomy and definitions of polyploidy. *New Phytol.* **213**, 487–493.

Esselink, G., Smulders, M. and Vosman, B. (2003) Identification of cut rose (*Rosa hybrida*) and rootstock varieties using robust sequence tagged microsatellite site markers. *Theor. Appl. Genet.* **106**, 277–286.

Fougère-Danezan, M., Joly, S., Bruneau, A., Gao, X.-F. and Zhang, L.-B. (2015) Phylogeny and biogeography of wild roses with specific attention to polyploids. *Ann. Bot.* **115**, 275–291.

Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A. and Mummenhoff, K. (2011) Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci.* **16**, 108–116.

Gar, O., Sargent, D.J., Tsai, C.-J., Pleban, T., Shalev, G., Byrne, D.H. and Zamir, D. (2011) An autotetraploid linkage map of rose (*Rosa hybrida*) validated using the strawberry (*Fragaria vesca*) genome sequence. *PLoS ONE*, **6**, e20463.

Gitonga, V.W., Koning-Boucoiran, C.F., Verlinden, K., Dolstra, O., Visser, R.G., Maliepaard, C. and Krens, F.A. (2014) Genetic variation, heritability and genotype by environment interaction of morphological traits in a tetraploid rose population. *BMC Genet.* **15**, 1.

Gitonga, V.W., Stolker, R., Koning-Boucoiran, C.F.S., Aelaei, M., Visser, R.G.F., Maliepaard, C. and Krens, F.A. (2016) Inheritance and QTL analysis of the determinants of flower color in tetraploid cut roses. *Mol. Breeding*, **36**, 143.

Guimarães, C.T., Sills, G.R. and Sobral, B.W.S. (1997) Comparative mapping of Andropogoneae: Saccharum L. (sugarcane) and its relation to sorghum and maize. *Proc. Natl Acad. Sci. USA*, **94**, 14261–14266.

Hackett, C.A. and Broadfoot, L.B. (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity*, **90**, 33–38.

Hackett, C.A., McLean, K. and Bryan, G.J. (2013) Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. *PLoS ONE*, **8**, e63939.

Hackett, C.A., Bradshaw, J.E. and Bryan, G.J. (2014) QTL mapping in autotetraploids using SNP dosage information. *Theor. Appl. Genet.* **127**, 1885–1904.

Herklotz, V. and Ritz, C.M. (2016) Multiple and asymmetrical origin of polyploid dog rose hybrids (*Rosa* L. sect. *Caninae* (DC.) Ser.) involving unreduced gametes. *Ann. Bot.* mcw217. doi: 10.1093/aob/mcw217.

Jenczewski, E., Eber, F., Grimaud, A., Huet, S., Lucas, M.O., Monod, H. and Chèvre, A.M. (2003) *PrBn*, a major gene controlling homeologous pairing in oilseed rape (*Brassica napus*) haploids. *Genetics*, **164**, 645–653.

Joly, S., Starr, J.R., Lewis, W.H. and Bruneau, A. (2006) Polyploid and hybrid evolution in roses east of the Rocky Mountains. *Am. J. Bot.* **93**, 412–425.

Joly, S., Bryant, D. and Lockhart, P.J. (2015) Flexible methods for estimating genetic distances from single nucleotide polymorphisms. *Methods Ecol. Evol.* **6**, 938–948.

Kirov, I., Van Laere, K., De Riek, J., De Keyser, E., Van Roy, N. and Khrustaleva, L. (2014) Anchoring linkage groups of the *Rosa* genetic map to physical chromosomes with Tyramide-FISH and EST-SNP markers. *PLoS ONE*, **9**, e95793.

Koning-Boucoiran, C.F.S., Gitonga, V.W., Yan, Z., Dolstra, O., Van der Linden, C.G., Van der Schoot, J., Uenk, G.E., Verlinden, K., Smulders, M.J.M. and Krens, F.A. (2012) The mode of inheritance in tetraploid cut roses. *Theor. Appl. Genet.* **125**, 591–607.

Koning-Boucoiran, C.F.S., Esselink, G.D., Vukosavljev, M. *et al.* (2015) Using RNA-Seq to assemble a rose transcriptome with more than 13,000 full-length expressed genes and to develop the WagRhSNP 68k Axiom SNP array for rose (Rosa L.). *Front. Plant Sci.* **6**, 249.

Koopman, W.J., Vosman, B., Sabatino, G.J. *et al.* (2008) AFLP markers as a tool to reconstruct complex relationships in the genus *Rosa* (*Rosaceae*). *Am. J. Bot.* **95**, 353–366.

Krzywinski, M.I., Schein, J.E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1653.

Leal-Bertioli, S., Shirasawa, K., Abernathy, B., Moretzsohn, M., Chavarro, C., Clevenger, J., Ozias-Akins, P., Jackson, S. and Bertioli, D. (2015) Tetrasomic recombination is surprisingly frequent in allotetraploid *Arachis*. *Genetics*, **199**, 1093–1105.

Lentz, E., Crane, C., Sleper, D. and Loegering, W. (1983) An assessment of preferential chromosome pairing at meiosis in *Dactylis glomerata*. *Can. J. Genet. Cytol.* **25**, 222–232.

Liorzou, M., Pernet, A., Li, S., Chastellier, A., Thouroude, T., Michel, G., Malécot, V., Gaillard, S., Briée, C. and Foucher, F. (2016) Nineteenth century French rose (*Rosa* sp.) germplasm shows a shift over time from a European to an Asian genetic background. *J. Exp. Bot.* **67**, 4711–4725.

Murat, F., Louis, A., Maumus, F., Armero, A., Cooke, R., Quesneville, H., Crollius, H.R. and Salse, J. (2015) Understanding Brassicaceae evolution through ancestral genome reconstruction. *Genome Biol.* **16**, 262.

Nguepjop, J.R., Tossim, H.-A., Bell, J.M., Rami, J.-F., Sharma, S., Courtois, B., Mallikarjuna, N., Sane, D. and Fonceka, D. (2016) Evidence of genomic exchanges between homeologous chromosomes in a cross of peanut with newly synthetized allotetraploid hybrids. *Front. Plant Sci.* **7**, 123–345.

Nicolas, S.D., Leflon, M., Monod, H., Eber, F., Coriton, O., Huteau, V., Chèvre, A.-M. and Jenczewski, E. (2009) Genetic regulation of meiotic cross-overs between related genomes in *Brassica napus* haploids and hybrids. *Plant Cell*, **21**, 373–385.

© 2017 The Authors

Okamoto, M. (1957) Asynaptic effect of chromosome V. *Wheat Inf. Serv.* **5**, 6.

Parisod, C., Holderegger, R. and Brochmann, C. (2010) Evolutionary consequences of autopolyploidy. *New Phytol.* **186**, 5–17.

Preedy, K.F. and Hackett, C.A. (2016) A rapid marker ordering approach for high-density genetic linkage maps in experimental autotetraploid populations using multidimensional scaling. *Theor. Appl. Genet.* **129**, 2117–2132.

Qu, L. and Hancock, J. (2001) Detecting and mapping repulsion-phase linkage in polyploids with polysomic inheritance. *Theor. Appl. Genet.* **103**, 136–143.

R Core Team. (2016) *R: A Language and Environment For Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramsey, J. and Schemske, D.W. (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* **29**, 467–501.

Riley, R. and Chapman, V. (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat. *Nature*, **182**, 713–715.

Roux, C. and Pannell, J.R. (2015) Inferring the mode of origin of polyploid species from next-generation sequence data. *Mol. Ecol.* **24**, 1047–1059.

Schulz, D.F., Schott, R.T., Voorrips, R.E., Smulders, M., Linde, M. and Debener, T. (2016) Genome-wide association analysis of the anthocyanin and carotenoid contents of rose petals. *Front. Plant Sci.* **7**, 1798. doi: 10.3389/fpls.2016.01798

Smulders, M.J.M., Arens, P., Koning-Boucoiran, C.F.S. *et al.* (2011). *Rosa*. In *Wild Crop Relatives: Genomic and Breeding Resources Plantation and Ornamental Crops* (Kole, C., ed). Berlin: Springer-Verlag, pp. 243–275.

Soltis, D.E., Visger, C.J., Marchant, D.B. and Soltis, P.S. (2016) Polyploidy: pitfalls and paths to a paradigm. *Am. J. Bot.* **103**, 1146–1166.

Spiller, M., Linde, M., Hibrand-Saint Oyant, L., Tsai, C.-J., Byrne, D.H., Smulders, M.J.M., Foucher, F. and Debener, T. (2011) Towards a unified genetic map for diploid roses. *Theor. Appl. Genet.* **122**, 489–500.

Stebbins, G.L. (1947) Types of polyploids: their classification and significance. *Adv. Genet.* **1**, 403–429.

Sved, J. (1966) Telomere attachment of chromosomes. Some genetical and cytological consequences. *Genetics*, **53**, 747.

Swaminathan, M.S. and Howard, H. (1953) Cytology and genetics of the potato (*Solanum tuberosum*) and related species. *Bibliogr. Genet.* **16**, 1–192.

Sybenga, J. (1975) The quantitative analysis of chromosome pairing and chiasma formation based on the relative frequencies of MI configurations. *Chromosoma*, **50**, 211–222.

Sybenga, J. (1994) Preferential pairing estimates from multivalent frequencies in tetraploids. *Genome*, **37**, 1045–1055.

Sybenga, J. (1996) Chromosome pairing affinity and quadrivalent formation in polyploids: do segmental allopolyploids exist? *Genome*, **39**, 1176–1184.

Van Ooijen, J.W. and Jansen, J. (2013). *Genetic Mapping in Experimental Populations*. Cambridge: Cambridge University Press.

Voorrips, R.E., Gort, G. and Vosman, B. (2011) Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics*, **12**, 172.

Vukosavljev, M., Arens, P., Voorrips, R. *et al.* (2016) High-density SNP-based genetic maps for the parents of an outcrossed and a selfed tetraploid garden rose cross, inferred from admixed progeny using the 68k rose SNP array. *Hortic. Res.* **3**, 16052.

Wissemann, V. and Ritz, C.M. (2005) The genus Rosa (Rosoideae, Rosaceae) revisited: molecular analysis of nrITS-1 and atpB-rbcL intergenic spacer (IGS) versus conventional taxonomy. *Bot. J. Linn. Soc.* **147**, 275–290.

Wolfram Research Inc. (2014). *Mathematica Version 10.0*. Champaign, Illinois.

Wu, K., Burnquist, W., Sorrells, M., Tew, T., Moore, P. and Tanksley, S. (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theor. Appl. Genet.* **83**, 294–300.

Wu, R., Ma, C.-X. and Casella, G. (2002) A bivalent polyploid model for linkage analysis in outcrossing tetraploids. *Theor. Popul. Biol.* **62**, 129–151.

Xiang, Y., Huang, C.-H., Hu, Y., Wen, J., Li, S., Yi, T., Chen, H., Xiang, J. and Ma, H. (2017) Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol. Biol. Evol.* **34**, 262–281.

Yan, Z., Dolstra, O., Prins, T.W., Stam, P. and Visser, P.B. (2006) Assessment of partial resistance to powdery mildew (*Podosphaera pannosa*) in a tetraploid rose population using a spore-suspension inoculation method. *Eur. J. Plant Pathol.* **114**, 301–308.

Zhang, J., Esselink, G., Che, D., Fougère-Danezan, M., Arens, P. and Smulders, M. (2013) The diploid origins of allopolyploid rose species studied using single nucleotide polymorphism haplotypes flanking a microsatellite repeat. *J. Hortic. Sci. Biotechnol.* **88**, 85–92.

Zheng, C., Voorrips, R.E., Jansen, J., Hackett, C.A., Ho, J. and Bink, M.C. (2016) Probabilistic Multilocus Haplotype reconstruction in outcrossing Tetraploids. *Genetics*, **203**, 119–131.

© 2017 The Authors
The *Plant Journal* published by John Wiley & Sons Ltd and Society for Experimental Biology.,