

## REVIEW ARTICLE

## Molecular genetics and genomics of the Rosoideae: state of the art and future perspectives

Sara Longhi<sup>1</sup>, Lara Giongo<sup>1</sup>, Matteo Buti<sup>1</sup>, Nada Surbanovski<sup>1</sup>, Roberto Viola<sup>1</sup>, Riccardo Velasco<sup>1</sup>, Judson A Ward<sup>2</sup> and Daniel J Sargent<sup>1</sup>

The Rosoideae is a subfamily of the Rosaceae that contains a number of species of economic importance, including the soft fruit species strawberry (*Fragaria ×ananassa*), red (*Rubus idaeus*) and black (*Rubus occidentalis*) raspberries, blackberries (*Rubus* spp.) and one of the most economically important cut flower genera, the roses (*Rosa* spp.). Molecular genetics and genomics resources for the Rosoideae have developed rapidly over the past two decades, beginning with the development and application of a number of molecular marker types including restriction fragment length polymorphisms, amplified fragment length polymorphisms and microsatellites, and culminating in the recent publication of the genome sequence of the woodland strawberry, *Fragaria vesca*, and the development of high throughput single nucleotide polymorphism (SNP)-genotyping resources for *Fragaria*, *Rosa* and *Rubus*. These tools have been used to identify genes and other functional elements that control traits of economic importance, to study the evolution of plant genome structure within the subfamily, and are beginning to facilitate genomic-assisted breeding through the development and deployment of markers linked to traits such as aspects of fruit quality, disease resistance and the timing of flowering. In this review, we report on the developments that have been made over the last 20 years in the field of molecular genetics and structural genomics within the Rosoideae, comment on how the knowledge gained will improve the efficiency of cultivar development and discuss how these advances will enhance our understanding of the biological processes determining agronomically important traits in all Rosoideae species.

Horticulture Research (2014) 1, 1; doi:10.1038/hortres.2014.1; published online 22 January 2014

## THE ROSACEAE

The Rosaceae family comprises some 90 genera containing approximately 3000 species which include: economically important fruit trees such as cultivated apples (*Malus pumila* Mill.) and pears (*Pyrus* spp.); stone fruits including peaches (*Prunus persica*) and sweet cherries (*P. avium*); numerous ornamental species, including the roses (*Rosa* spp.), mountain ash (*Sorbus aucuparia*) and ninebark (*Physocarpus opulifolius*); and the soft-fruit species strawberry (*Fragaria ×ananassa*), raspberry (*Rubus idaeus*) and blackberry (*Rubus* spp.), among many others. Several classifications of the family based on morphology have been proposed (e.g., ref. 1), while Schulze-Menz<sup>2</sup> proposed the categorisation of the family into four subfamilies: Maloideae, Amygdaloideae, Rosoideae and Spiraeoideae based on chromosome number and fruit type. Molecular phylogenetic analyses strongly supported the monophyly of the Rosaceae and demonstrated that chromosome number is a better character for subclassification within the family than fruit morphology.<sup>3,4</sup>

## ROSOIDEAE PHYLOGENETICS

Morgan *et al.*<sup>3</sup> were the first to establish the monophyly of the Rosoideae within the Rosaceae, and the subsequent reclassification of the family by Potter *et al.*<sup>5</sup> retained the Rosoideae as a subfamily resulting from the first phylogenetic split within the family. Figure 1 shows a schematic depiction of the phylogeny of the Rosaceae adapted from Potter *et al.*<sup>5</sup> displaying the three main subfamilies Rosoideae, Spiraeoideae and Dryadoideae, and the tribes within each subfamily. According to this molecular classification,<sup>5</sup> the Rosoideae subfamily contains a single supertribe, the Rosodae, containing all genera except for *Filipendula*. Rosodae genera are

contained within three tribes, the Sanguisorbeae, the Potentilleae (containing the Fragariinae subtribe to which the genus *Fragaria* belongs) and the Colurieae. *Rosa* and *Rubus*, which are contained within the Rosodae, are not contained within any tribe. Plants of the Rosoideae are characterized by a base chromosome number of  $x=7$ . The majority of the plants within the subfamily are perennial herbs and shrubs; however, a small number of annual herbs and perennial tree species exist within the current classification.<sup>5</sup>

Rosoideae genera contain diverse numbers of species; the genus *Fragaria* for example contains relatively few species, with just 21 currently recognised and described,<sup>6</sup> while *Rosa* is thought to contain at least 120 species,<sup>7</sup> *Potentilla* up to 500 species<sup>8</sup> and *Rubus* some 750.<sup>9</sup> Species contained within the Rosoideae genera exhibit impressive morphological diversity and exist at numerous ploidy levels, and thus, the classification and phylogenetic determination of species within each genus have been the subject of various phylogenetic analyses, with molecular phylogenies presented for *Fragaria*,<sup>10</sup> *Potentilla*,<sup>8</sup> *Rosa*<sup>11</sup> and *Rubus*.<sup>9</sup> Within these studies, the evolutionary origins of some important polyploid species in relation to extant diploid progenitors have been determined.

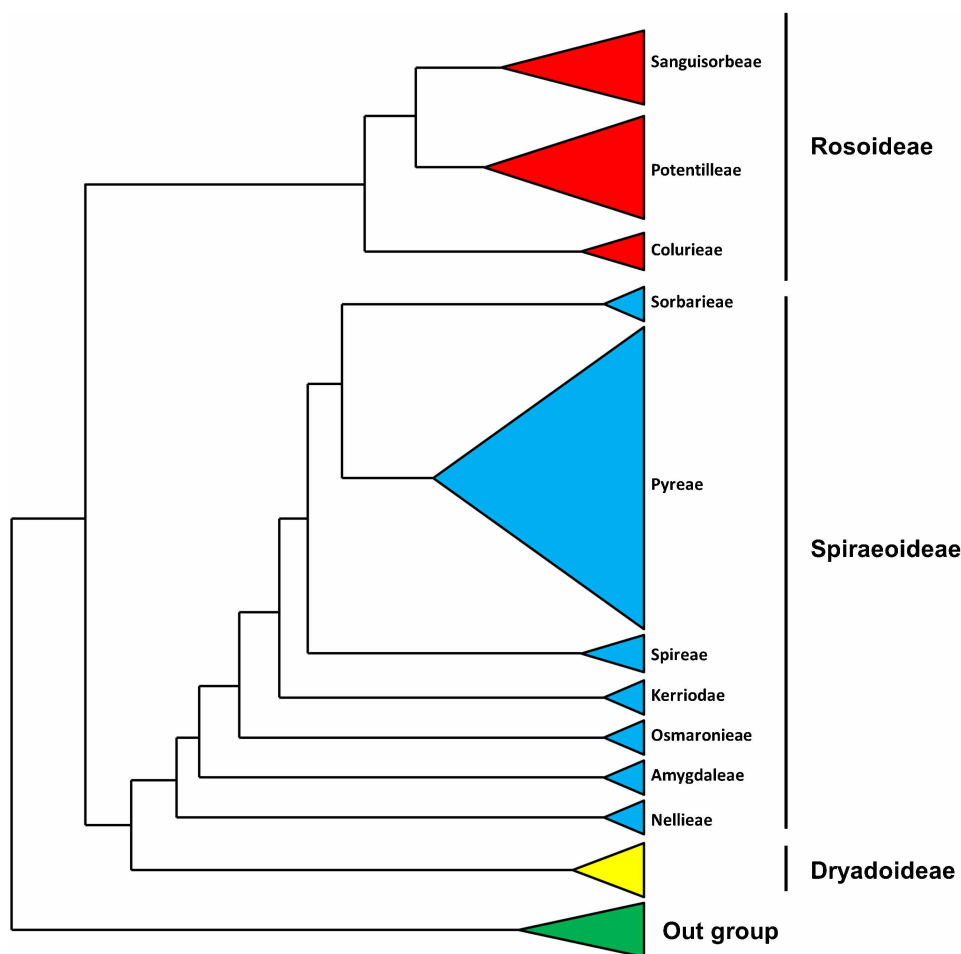
## THE ROSOIDEAE AND ITS ECONOMIC IMPORTANCE

The Rosoideae contains species prized for their ornamental value such as *Rosa* species, those that produce sweet edible aggregate and accessory 'fruits' including *Rubus idaeus* (red raspberry) and *Fragaria ×ananassa* (cultivated strawberry), highly dispersed wild herbaceous species such as *Potentilla reptans* and species that have become regarded as invasive pests such as *Rosa rubiginosa*.<sup>12</sup> The economic importance of the subfamily is exemplified by three main

<sup>1</sup>Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Italy and <sup>2</sup>Driscoll's, Watsonville, CA, USA.

Correspondence: DJ Sargent (dan.sargent@fmach.it)

Received: 4 November 2013; accepted: 24 November 2013



**Figure 1.** A schematic representation of the phylogeny of the Rosaceae adapted from Potter *et al.*,<sup>5</sup> showing three subfamilies: Rosoideae (red); Spiraeoideae (blue) and Dryadoideae (yellow) and the tribes within each family. The relative size of each triangle represents the relative number of genera contained/analysed in each tribe in the phylogenetic reconstruction by Potter *et al.*<sup>5</sup>

genera, *Fragaria*, *Rosa* and *Rubus*, although other genera of more minor economic value exist. *Fragaria* and *Rubus* species are valued for their 'fruits', which take the form of berries (swollen receptacles that bear the true fruits, the achenes on their surface) in *Fragaria*, and aggregates of drupelets in *Rubus*, while *Rosa* is prized for its large, often fragrant and colorful flowers and over 20 000 commercial rose cultivars are reported to exist worldwide.<sup>13</sup> Together, products of these three genera are of immense economic significance, which has prompted the need to continually produce improved varieties of both flowers and 'fruits' for a market that is constantly increasing its demands in terms of quality and quantity. As a consequence of the demand for new varieties, breeding programmes for the genetic improvement of key species have been established worldwide, and in line with almost all other crop plants, a huge investment in the development of molecular tools to assist these programmes, and to answer more fundamental biological questions, has been made over the past 20 years. This review will outline the major advances made in molecular genetics and genomics research in the Rosoideae over the past two decades and highlight the direction this research will take in the coming years in the light of technological advances made in the field of biological sciences.

#### ROSOIDEAE CYTOLOGY AND ESTIMATES OF GENOME SIZE

Cytological studies performed on members of the Rosoideae including *Fragaria*, *Potentilla*, *Rosa* and *Rubus* have demonstrated

the base chromosome number of the subfamily to be  $x=7$ , and further works have categorised and described the karyotypes of different species within the genera.<sup>14–17</sup> While the base chromosome number of members of the Rosoideae is the smallest for any of the plants within the Rosaceae,<sup>18</sup> polyploidisation events within each of the genera has led to species with diploid chromosome counts ( $2n=2x=14$ ) through to tetradecaploid ( $2n=14x=98$ ) being reported, as well as various aneuploids and interploidy hybrids.<sup>19</sup> Fertile interspecific hybrids have been reported within discrete ploidy levels in many of the genera, suggesting that within a genus, species are closely related<sup>20,21</sup> and while intragenic hybrids between *Fragaria* and *Potentilla*<sup>22</sup> have been produced, viable fertile hybrids were observed only at higher levels of ploidy.

The genomes of diploid Rosoideae species that have been investigated to date were shown to be relatively small, with estimates made using flow cytometry for diploid *Rosa* species ranging in size between subgenera from 0.78 pg/2C–1.29 pg/2C by Yokoya *et al.*<sup>23</sup> to 1.10 pg/2C–1.36 pg/2C by Rajapakse *et al.*,<sup>7</sup> who estimated a genome size in nucleotides of ~600 Mbp/C, while the estimate for *Rubus* species ranged from 0.58 pg/2C for *R. ideaus*<sup>24</sup> to 0.75 pg/2C for *R. sanctus*.<sup>25</sup> The genome size estimate of *F. vesca* was determined by flow cytometry through comparison with the 125 Mbp/C sequenced genome size of *Arabidopsis thaliana* (0.34 pg/2C) and was determined to be 164 Mbp/C (and thus presumably in the range of 0.45 pg/2C);<sup>26</sup> however, this was later

adjusted in the light of changes in the size estimates of the Arabidopsis genome<sup>27</sup> to be in the region of 206 Mbp/C. The genome size of the cultivated octoploid strawberry was also estimated in the same study and determined to be, following adjustment against new estimates of the genome size of *A. thaliana*, in the region of 703 Mbp, less than four times the estimated genome size of the diploid *Fragaria*, suggesting extensive loss of chromosomal content following polyploidisation.<sup>26</sup> Thus, Rosoideae genomes have been estimated to be among some of the smallest of all angiosperm genomes, with *Fragaria* having the smallest, followed by *Rubus* and then *Rosa*.

## MOLECULAR MARKER DEVELOPMENT AND APPLICATION

### Arbitrary marker systems

Numerous molecular marker systems have been employed in a large number of genetic analyses of Rosoideae species over the past 20 years. These include arbitrary Polymerase Chain Reaction (PCR)-based marker systems, such as amplified fragment length polymorphisms (AFLP), random amplified polymorphic DNA (RAPD) and inter-simple sequence repeats. These marker systems were used in *Fragaria*, *Rosa* and *Rubus* for a range of genetic analyses, including cultivar and sport identification and genetic fingerprinting of closely related germplasm,<sup>28–30</sup> parentage analyses and evaluations of genetic diversity among species,<sup>31–34</sup> linkage map construction,<sup>35–37</sup> and population genetics<sup>38,39</sup>. The inherent limitations of arbitrary marker systems in transferability and reproducibility between populations, species and laboratories, however, have meant that, as in the vast majority of plant species, more robust, sequence-characterized marker systems such as microsatellites have gained greater favour in genetic analyses in recent years.

### Microsatellite markers

**Microsatellite development.** Microsatellites, or simple sequence repeats (SSRs) are PCR-based, sequence-characterized markers that have become the marker of choice for genetic analysis. Prior to the advent of second-generation sequencing technologies, various protocols for the efficient characterisation of SSRs from enriched genomic libraries of Rosoideae species were reported,<sup>40–42</sup> along with the characterisation of SSR markers from expressed sequence tag (EST) libraries.<sup>43–45</sup> Subsequently, following the availability of high-throughput sequencing methodologies and the release of the *F. vesca* genome sequence,<sup>46</sup> SSR markers were developed in abundance for both *Fragaria*, where more than 4000 SSRs have been characterized from the genome sequence and mapped in both wild and cultivated species,<sup>47,48</sup> and *Potentilla* from which 74 SSR markers were characterised from 1476 sequences identified in 454 genomic sequence data derived from *P. pusilla*.<sup>49</sup>

**Cultivar identification.** In *Fragaria*, *Rosa* and *Rubus*, SSR markers have been successfully employed for cultivar identification. In *Fragaria*, a set of 10 SSRs in three multiplexes have been reported for fingerprinting cultivated strawberry germplasm,<sup>50,51</sup> and more recently, Chambers *et al.*<sup>52</sup> used 16 high repeat number SSR markers to discriminate between a collection of 219 cultivated strawberry varieties. In rose, 24 polymorphic SSR markers were used to characterise over 70 hybrid tea and rootstock varieties,<sup>53</sup> while just six SSR markers were needed to effectively discriminate between 65 accessions of old garden roses.<sup>54</sup> Likewise in *Rubus*, 21 SSR markers were used to discriminate between 148 wild and cultivated black raspberry accessions,<sup>55</sup> and two fingerprinting sets have been developed for red raspberry.<sup>56,57</sup>

**Transferability between species.** While SSR markers are robust and reliable for genetic analysis within a species, their transferability

between species and genera is more limited than for other marker types. Despite reports of transferability of SSR markers between Rosaceous genera being possible,<sup>58,59</sup> in practice, most of the markers reported display very low levels of polymorphism and thus there are only a few examples of SSR markers from other Rosaceous genera being employed successfully for genetic analysis in Rosoideae species.<sup>7</sup>

The rate of transferability and heterozygosity of SSR markers has been investigated within the Rosoideae. Within genera, the transferability of SSR markers is high and markers developed from one species have been shown to be directly applicable to other species. Davis *et al.*<sup>60</sup> and Zorilla-Fontanesi *et al.*<sup>61</sup> demonstrated up to 90% transferability of polymorphic SSR markers from diploid to octoploid strawberry and others<sup>47,48,62</sup> have employed large numbers of diploid *Fragaria* SSR markers in the construction of linkage maps of the cultivated strawberry. Likewise in *Rubus*, SSR markers developed for red raspberry were successfully used to genotype accessions of black raspberry and five other *Rubus* species by Fernandez Fernandez *et al.*,<sup>56</sup> while Marulanda *et al.*<sup>63</sup> demonstrated the transferability of SSR markers developed in a number of *Rubus* species to the Andean blackberry, *R. glaucus*. Similarly, high rates of SSR transferability have been observed in other Rosoideae genera, including *Potentilla*<sup>49</sup> where 86%–94% transferability between species was reported.

Lewers *et al.*<sup>43</sup> reported that transferability of SSR markers between the genera was low (31%–19%), while Rousseau-Gueutin *et al.*<sup>64</sup> reported transferability of *Fragaria* EST-SSRs to *Potentilla* as 75%, while to *Rosa*, it was just 30%. Koning-Boucoiran *et al.*<sup>65</sup> found that just 17% of the *Fragaria* SSR markers tested transferred to *Rosa* and just 2% were heterozygous in a rose mapping population. Park *et al.*,<sup>44</sup> however, demonstrated higher rates of transferability from *Rosa* to *Fragaria*, reporting that 61% of the EST-SSR markers they developed for *Rosa* were transferred to *Fragaria*, 47% of which were reported to be polymorphic. These studies highlight the potential for cross-genera application of SSR markers; however, they also demonstrate that the more distantly related the two genera are, the lower the transferability and polymorphism of SSR loci will be. Thus in practice, as SSR development and characterisation has become easier, molecular genetics research has focused on applying SSR markers to species within the genera from which they were developed, and cross-genera studies have mainly utilized other marker systems for the evaluation of synteny between genera within the Rosaceae and the Rosoideae, such as restriction fragment length polymorphism (RFLP) and sequence-tagged site markers developed from ESTs.<sup>66,67</sup>

### RFLPs

RFLPs have been used to a limited extent for the molecular characterisation of Rosoideae species, including cultivar identification<sup>68</sup> and phylogenetic analyses.<sup>39</sup> Since RFLP probes are commonly developed from gene-coding sequence, they are often highly conserved between species. Such markers, termed conserved orthologous set markers, have found enormous utility in comparative mapping studies between plant genera and even families.<sup>69,70</sup> In the Rosoideae, they have been employed in the development of a linkage map of rose<sup>71</sup> and their conserved nature was also exploited, along with PCR-based markers designed from the conserved coding regions of ESTs, in a study of the conservation of synteny between *Fragaria* and *Prunus*.<sup>66</sup>

**Single nucleotide polymorphism (SNP) detection and analysis**  
Traditionally, SNPs in the genomes of Rosoideae species were identified through projects that employed Sanger sequencing to sequence EST collections<sup>72,73</sup> or through direct sequencing of PCR products amplified from genomic regions of interest.<sup>66,74,75</sup> SNP markers were then scored in progenies either by direct

sequencing of PCR products from specific genotypes or segregating progenies<sup>66,74,75</sup> or through the utilisation of the cleaved-amplified polymorphic site approach where suitable restriction enzymes were available, to differentially digest alternative alleles at a locus.<sup>76,77</sup> Such approaches, while effective and reliable, remain extremely low throughput since individual assays can rarely be multiplexed effectively, and the results are most usually visualised through agarose gel electrophoresis. Thus, for the development of large numbers of markers for mapping and large-scale surveys of populations or varieties, SNPs have not traditionally been used routinely.

Since the advent of second-generation sequencing technologies, however, high-throughput methods for SNP discovery and analysis have been developed, which have revolutionised the ability to discover, screen and associate SNPs in progenies and germplasm collections. High-throughput deep sequencing of both genomic DNA and mRNA (termed RNA-seq) has been used to identify SNPs in both *Fragaria*,<sup>78</sup> *Rubus*<sup>79</sup> and *Rosa*,<sup>80</sup> following which a number of downstream analysis methods are available for interrogation and analysis of thousands of SNP markers in a single genotyping assay. Novel techniques have been developed for assaying large numbers of SNPs using the Illumina sequencing platform; Celton *et al.*<sup>78</sup> exploited reduced representation restriction fragment libraries of the FV×FB selective mapping progeny to identify and map large numbers of SNP markers in diploid *Fragaria*, while Ward *et al.*<sup>79</sup>, using the 'genotyping by sequencing' (GBS) approach of Elshire *et al.*<sup>81</sup>, developed a densely saturated linkage map for a red raspberry mapping population. Likewise, using a novel sequence capture technique utilizing RNA baits termed targeted capture,<sup>82</sup> Tennessen *et al.*<sup>83</sup> developed a high density SNP-based linkage map of *F. vesca* subsp. *bracteata* composed entirely of segregating SNP markers.

The use of direct sequencing permits the identification and analysis of large numbers of SNPs without *a priori* knowledge of their genomic position or nature, but the data generated are technically demanding to analyse and interpret reliably and accurately. If SNP data are already available for an organism, however, a number of other genotyping assays have become available and have begun to be exploited recently in Rosoideae species. Kompetitive (sic) allele-specific PCR, which can be used to genotype individual SNPs in moderate numbers (96-1536) of individuals reliably and accurately, has been used to generate segregation data for mapping in tetraploid rose,<sup>65</sup> while high-throughput, massively multiplexed genotyping array technology, which permits the interrogation of tens of thousands of SNPs simultaneously using microarray technology, has been exploited recently for the development of genotyping tools for wild and cultivated strawberry species.<sup>84</sup> The development and exploitation of such technologies in the Rosoideae is still in its infancy and there are yet to be full reports of their application in the scientific literature; however, similar technology is currently being applied in related Rosaceous genera such as apple,<sup>85</sup> peach<sup>86</sup> and cherry,<sup>87</sup> and unpublished data suggest that the IStraw90 array for *F. ×ananassa* is effective at producing mapping data for markers spanning the cultivated strawberry genome.

## LINKAGE MAP DEVELOPMENT

Linkage mapping permits the location of genes controlling traits of importance to chromosomes through their association with molecular markers and thus, represents a powerful tool for the positional cloning of major genes, the characterisation of quantitative trait loci (QTL), and the anchoring and ordering of contigs and scaffolds of both physical maps and genome sequence assemblies. With the rapid developments made in molecular marker characterisation and the ease in which they can be used to genotype individuals, genetic linkage maps of Rosoideae species have continually evolved to become increasingly saturated, enabling the genetic

dissection of the genomes of members of the Rosoideae genera and the identification of loci controlling of traits of agronomic importance in *Fragaria*, *Rosa* and *Rubus*.

## FRAGARIA

### Diploid *Fragaria* linkage maps

*Fragaria* species are found in a range of ploidy levels from diploid to decaploid, while the most economically important species, *F. ×ananassa*, is a complex allo-octoploid. The phylogenetic origins of the octoploid species *Fragaria* have been investigated and evidence that a number of extant diploid species, including *F. vesca* and *F. iinumae*, have genomes similar to the diploid progenitors of the modern octoploid species has been presented.<sup>10</sup> Thus, the application of molecular markers to genetic mapping studies in *Fragaria* has permitted the investigation of the structure of both the diploid and octoploid genomes, comparative genomic analyses, the development of saturated maps of the cultivated species and the identification of markers useful for breeding.

Genetic investigations of the genome of the cultivated strawberry were initially confounded by its complex polyploid nature and thus, initial linkage studies focused on *F. vesca* a wild diploid closely related to the polyploid strawberry species.<sup>10</sup> The first linkage map for *F. vesca* spanning the expected seven chromosomes of the species was reported by Davis and Yu.<sup>35</sup> The map was developed from an F<sub>2</sub> population comprising 80 individuals derived from an intraspecific cross between *F. vesca* var. Baron Solemacher and the *F. vesca* accession WC6. The map was composed of a total of 80 markers resolved into seven linkage groups and covering a genetic distance of 445 cM. This map included 75 RAPD markers (64 dominant and 11 codominant), a gene-specific marker for the alcohol dehydrogenase locus, two isoenzyme markers, phosphoglucose isomerase (*Pgi-2*) and shikimate dehydrogenase (*Sdh*), and morphological markers for the major genes controlling the *runnerless* (*r*) and *yellow fruit color* (*c*) phenotypes.

Subsequently, a second diploid *Fragaria* linkage map was reported by Deng and Davis.<sup>88</sup> They used two F<sub>2</sub> populations derived from a cross between *F. vesca* subsp. *bracteata* DN1C and the *F. vesca* variety Yellow Wonder, and a cross between Yellow Wonder and the *F. nubicola* accession FRA520 containing 40 individuals each, to map six candidate genes: chalcone synthase, chalcone isomerase, flavanone 3-hydroxylase, dihydroflavonol 4-reductase, anthocyanidin synthase structural genes and a Del-like regulatory gene. The genes mapped to five of the seven *Fragaria* linkage groups, identified through comparative mapping of RAPD markers previously mapped by Davis and Yu.<sup>35</sup> In the study, they showed, through cosegregation, that a mutation in the flavanone 3-hydroxylase gene of the anthocyanin biosynthesis pathway was likely responsible for yellow fruit colour in the variety Yellow Wonder.

Following the development of these initial linkage maps using predominantly arbitrary PCR-based markers, a map characterizing an interspecific mapping progeny was reported by Sargent *et al.*<sup>20</sup>. A total of 73 molecular markers (66 SSR, 1 sequence-characterized amplified region (SCAR) and 6 gene-specific markers) and three genes controlling morphological traits were mapped in an F<sub>2</sub> population of 94 individuals (later reduced to 76) obtained from a cross between *F. vesca* f. *semperflorens* '815' and *F. bucharica* '601' (formerly *F. nubicola* '601') (FV×FB). The map covered 448 cM and was resolved into the seven linkage groups expected for the genus. A high degree of segregation distortion was observed along the linkage groups, which was suggested to have arisen from the interspecific nature of the cross.<sup>20</sup> This genetic differentiation, however, permitted greater numbers of markers to be mapped compared with previous intraspecific mapping progenies and thus, it was adopted as the international reference mapping progeny for *Fragaria*.

Successive studies have refined and improved the FV×FB reference map, continually adding markers, including SSRs, gene-specific markers, RFLPs, SNPs and ESTs, and increasing the degree of saturation along the linkage groups.<sup>46,61,66,76,78,89–93</sup> The most recent incarnation of the FV×FB map<sup>93</sup> comprises a total of 411 sequence characterised markers (SSR, RFLP, EST and gene-specific markers) mapped in the full progeny of 76 individuals, and a further 298 markers mapped using a selective mapping strategy using just six seedlings that divided the linkage groups into a total of 46 mapping bins covering the *Fragaria* genome.<sup>91</sup> The total length of the current FV×FB linkage map is 442.8 cM across seven linkage groups, comparable to that of the first diploid strawberry map which covered 445 cM<sup>35</sup> and is estimated to almost completely cover the diploid *Fragaria* genome.<sup>93</sup> Due to the exclusive use of sequence-characterized markers in its construction, the diploid *Fragaria* reference map was employed to anchor and orientate the genome sequence scaffolds that were derived from the sequencing of the *F. vesca* Hawaii 4 (FvH4) genome by Shulaev *et al.*<sup>46</sup>

The release of the FvH4 genome sequence<sup>46</sup> provided an essential resource for re-sequencing projects aiming to develop SNP-based linkage maps for *Fragaria*. In order to study male sterility in the gynodioecious diploid strawberry *F. vesca* subsp. *bracteata*, parental molecular maps of an intraspecific mapping population were developed.<sup>83</sup> The maps contained a total of 7802 SNP and in-del markers, and following linkage analysis, the seven expected linkage groups were resolved on both maps. The 4338 markers that mapped to the maternal map covered 410 cM, while the paternal map comprised 4305 markers spanning 406 cM. These lengths are comparable with the *F. vesca* diploid maps previously developed (445 cM in the map of Davis and Yu,<sup>35</sup> 442.8 cM in the map of Sargent *et al.*<sup>93</sup>).

#### Octoploid *Fragaria* linkage maps

The cultivated strawberry *F. ×ananassa* originated from the chance hybridization between the allo-octoploid species *F. virginiana* and *F. chiloensis*.<sup>94</sup> Several genomic formulas, based on cytological observations, have been proposed for the cultivated strawberry, the most commonly accepted of which is that of Bringhurst,<sup>95</sup> who proposed the genomic composition AAA'A'BBB'B' reflecting the contention that the allo-octoploid *Fragaria* genomes are completely diploidised and that segregation is completely disomic. Initial mapping efforts in the cultivated strawberry suggested that evidence for mixed disomic and polysomic inheritance was observed.<sup>96</sup> However, the growing availability of numerous codominant transferable markers previously mapped to the diploid reference map greatly facilitated the development of well-characterized linkage maps for both *F. ×ananassa* and *F. virginiana*<sup>47,62,97</sup> and novel techniques for studying segregation through microsatellite allele dose and configuration establishment<sup>98</sup> have demonstrated that segregation in the allo-octoploid *Fragaria* is exclusively disomic and that the genome is fully diploidised as suggested by the genome model proposed by Bringhurst.<sup>95</sup>

The first linkage maps of the cultivated strawberry were reported by Lerceteau-Köhler *et al.*<sup>96</sup>. The maps were constructed from a full-sib F<sub>1</sub> progeny comprising 133 individuals from a cross between the variety Capitola and the breeding line CF1116 ([Pajaro×Earlyglow]×Chandler) (CA×CF) and were generated using 789 AFLP markers and two putative genes, alcohol acetyl transferase and dihydro-flavonol 4-reductase. The female and male parental maps contained 235 and 280 markers and were resolved into 43 linkage groups covering 1604 cM and 1496 cM, respectively. Subsequently, the CA×CF map was extended using a larger population of 213 seedlings and new AFLP, SCAR and SSR markers were added to the existing AFLP framework.<sup>99</sup> The additional markers extended the maps to include 367 markers covering 2582 cM on

the maternal map and 440 markers covering 2165 cM on the paternal map. Integration of the two maps resulted in a final consensus map for the progeny spanning 2195 cM across 32 Linkage Group (LG)s. Through comparison of common markers mapped to the diploid *Fragaria* reference map, linkage groups representing four homoeologous groups were identified for six of the seven diploid LGs, with just three homoeologous groups recovered for LG2. The study highlighted extremely high levels of macrosynteny between the diploid and octoploid maps, suggesting the absence of major chromosomal rearrangements during the evolution of polyploid *Fragaria* species from their diploid progenitors and validating the use of the FV×FB linkage map as a reference for the genus. The addition of sequence-characterized markers also supported the conclusion that disomy was the predominant, if not the only, meiotic behaviour exhibited by the *F. ×ananassa* genome.

While the maps of the CA×CF progeny spanned the majority of the *F. ×ananassa* genome and were well saturated with markers, the predominant marker type used for the development of the maps was AFLPs. Subsequent maps of octoploid strawberry progenies employed larger numbers of transferable markers. In order to study sex determination, Spigler *et al.*<sup>97</sup> developed the first SSR-based linkage map of an octoploid strawberry species. A total of 210 SSR markers derived from various genomic sources were mapped in a *F. virginiana* mapping population raised from a cross between a female maternal line and a hermaphrodite male line. The progeny consisted of 184 individuals and the resultant consensus linkage map had a total length of 2373 cM across 42 linkage group fragments, 14 more than expected for a consensus map of the species. An updated map for the progeny was presented by Spigler *et al.*,<sup>100</sup> composed of linkage group fragments associated with 30 linkage groups per parent; however, in their second report, the authors chose to present the female and male linkage maps separately rather than as a consensus map.

Subsequently, two integrated linkage maps of the cultivated strawberry, *F. ×ananassa*, were reported. The first was developed to investigate the inheritance of resistance to *Verticillium dahliae* and was derived from a cross between the resistant cultivar Redgauntlet and the susceptible cultivar Hapil.<sup>47,101</sup> The integrated map had a total length of 2140.3 cM and was developed from a progeny of 174 seedlings. The map, which was almost the same length as the AFLP-based linkage map developed by Rousseau-Gueutin *et al.*,<sup>99</sup> contained 549 molecular markers (89% of which were SSR and gene specific markers) and was resolved into the expected 28 LGs for an integrated map. One of the groups, however, was composed of separate female and male maps since no biparental markers for that LG segregated in the progeny. While no data pertaining to wilt resistance were reported in either mapping study, due to the transferable nature of the markers, a comparison with maps developed in previous investigations and of loci on each of the four *F. ×ananassa* homeologous linkage groups corresponding to the seven linkage groups of the diploid reference map was possible. In line with other studies, the authors demonstrated a high degree of colinearity between the diploid and the octoploid *Fragaria* genomes. Moreover, this work highlighted regions of homozygosity within the *F. ×ananassa* genome also present in other *F. ×ananassa* linkage maps, including those of Rousseau-Gueutin *et al.*<sup>99</sup> and Zorilla-Fontanesi *et al.*,<sup>62</sup> which the authors suggested were evidence of selection in the genome of the cultivated strawberry, since such regions of homozygosity were not evident on the linkage map of *F. virginiana* presented by Spigler *et al.*<sup>100</sup> The second integrated SSR map reported for *F. ×ananassa* was composed of 338 markers, of which 250 were SSRs, and was developed from a mapping population obtained from a cross between two *F. ×ananassa* lines 232 (4-43×Vilanova) and 1393 (Gaviota×Camarosa) (232×1393), for the purposes of studying the genetics of fruit quality.<sup>62</sup> The total length of the map was 1259.8 cM spanning 37 linkage groups, exceeding the 28 expected

groups for an integrated map of the species; however, all the LGs included a marker present in the reference map (FV×FB) allowing the assignment of each linkage group fragment to one of the seven diploid LGs and permitting further map comparisons between the diploid and octoploid maps to be performed. Later, the resolution of the linkage map was increased through the mapping of additional markers.<sup>102</sup> In that study the O-methyltransferase locus, along with other candidate genes, was mapped and a QTL analysis to identify loci controlling the production of volatile compounds was performed. The total length of the integrated map was increased to 1400.1 cM and 363 markers (70% of which were SSRs) were mapped to 39 linkage groups fragments, all but one of which was assigned to one of the four homeology groups associated with each diploid linkage group.

The most comprehensive linkage map developed for the cultivated strawberry to date was recently reported by Isobe et al.<sup>48</sup> In their study, the authors developed an integrated linkage map composed exclusively of transferable SSR markers from parental linkage maps of three mapping populations: the F<sub>1</sub> progeny of the crosses 02-19×Sachinoka containing 188 individuals and Kaorino×Akihime containing 140 individuals, and an F<sub>2</sub> progeny 0212921 containing 169 individuals derived from the selfing of an unnamed genotype. Constructed from five linkage maps, the integrated consensus map produced encompassed a larger amount of diversity than one map could individually and contained a total of 1856 loci distributed throughout 28 linkage groups spanning a total length of 2364.1 cM. This map is of a comparable length to the maps previously reported by Rousseau-Gueutin et al.<sup>99</sup> (2195 cM) and Sargent et al.<sup>47</sup> (2140.3 cM) and was nearly twice the length of the map developed by Zorilla-Fontanesi et al.<sup>102</sup> Given the uniformity in linkage group lengths of the majority of maps of the cultivated strawberry presented to date, and the high numbers of markers mapped in the study of Isobe et al.,<sup>48</sup> it is reasonable to assume that the current integrated linkage map covers the majority of the cultivated strawberry genome.

**Mapped traits and QTL analysis.** Molecular maps are essential for the investigation of the inheritance of important traits and the genes that control them, and thus, are an important precursor for the development of marker assisted selection (MAS) programs. Targeting the genetic elements controlling important agronomic traits like fruit quality, disease resistance or metabolic content, is a crucial step for crop improvement. These traits are, however, usually controlled by QTL, which makes the identification of the genetic determinants accounting for smaller percentages of the genetic variance challenging. Genetic analysis of QTL in polyploid species gives insight into the relationship of how gene copy number influences the trait of interest, and one of the major challenges to genetic research in the cultivated strawberry is understanding how homoeologous gene loci impact the genetic control of agronomically-important traits. The sections below describe some of the traits that have been characterized genetically in *Fragaria* species.

#### Disease resistance

There are a number of economically significant fungal pathogens which cause severe damage to strawberry crops worldwide. These diseases include *Botrytis cinerea* (grey mould), those of the genus *Phytophthora*, including *Phytophthora fragariae* (red stele) and *P. cactorum* (crown rot), *V. dahliae*, and members of the genus *Colletotrichum*, most notably *C. acutatum* (blackspot or anthracnose). As natural resistance to many of these diseases exists in *F. ×ananassa* germplasm, attention has been focused on the development of molecular markers to characterize and select for resistance through linkage mapping. An F<sub>1</sub> population containing 60 individuals obtained from a cross between the *F. ×ananassa* selection Md683 and the variety Senga Sengana was used in the

first study of marker-trait linkage in the cultivated strawberry.<sup>103,104</sup> The trait under investigation was resistance to the soilborne fungus *P. fragariae* var. *fragariae*, the cause of the red stele root rot disease.<sup>103</sup> Seven RAPD markers linked to the *Rpf1* gene were identified and one of them was cloned and later converted to a SCAR marker linked in coupling phase to the resistance gene.<sup>104</sup> Hitherto, however, there have been no reports of the use of this marker in breeding programmes. Indeed alone, it has been reported that the marker is not sufficient for the selection of comprehensive resistance to red stele, which to be effective involves the pyramiding of three resistance genes, *Rpf1*, *Rpf2* and *Rpf3*.<sup>105</sup> Using the map of Lerceteau-Köhler et al.,<sup>96</sup> five QTL for resistance to *P. cactorum* and five for resistance to *C. acutatum* were characterized and mapped by Denoyes-Rothan et al.<sup>106</sup> Later two SCAR markers linked to the *Rca2* anthracnose resistance gene conferring resistance to *C. acutatum* were developed,<sup>107</sup> and subsequently, the markers have been used in a programme of marker-assisted breeding, despite permitting selection for just *C. acutatum* pathogenicity group 2.<sup>105</sup> Recently, markers were identified linked to *V. dahliae* resistance in 'Redgauntlet' using the RG×H mapping population of Sargent et al.<sup>47</sup> A number of QTL were identified, with the most promising markers developed from a hydroxyproline rich glycoprotein on LG1, one allele of which was present in the majority of the resistant cultivars screened, and absent from susceptible genotypes (Šurbanovski et al., unpublished data), showing promise for the development of an MAS strategy for the selection of novel resistant material.

#### Flowering habit

Cultivated strawberries are divided, according to their flowering habits, into two primary types: short-day (SD) and day-neutral (DN). SD genotypes require a day length shorter than 14 h or temperatures below 15°C to initiate flowers, while DN plants require only moderate temperatures to begin flowering, being insensitive to photoperiod. DN plants are also defined as everbearing; however, not all everbearing plants have the same flowering behaviour as DN types. The first DN cultivars were released in 1979 and all are descended from an accession of *F. virginiana* subsp. *glauca* from the Utah mountains, with the DN flowering characteristic introduced by backcross breeding. In favourable conditions, DN cultivars can produce fruit continuously all summer and into the autumn. Many everbearing cultivars have existed for much longer than DN types, but typically, they have a different cropping pattern, with two or three peaks during each growing season.

In order to study the genetics of flowering habit in the cultivated strawberry, a map resolving a single linkage group was developed containing markers linked to the *Everbearing* (*Ev*) trait.<sup>108</sup> In the study, 199 F<sub>1</sub> seedlings from a cross between two Japanese octoploid varieties Ever Berry and Toyonoka were used to map 5 RAPD markers in the 39.7 cM surrounding the dominant gene (*Ev*) regulating day-neutrality.<sup>108</sup> This was the first report of DNA markers linked to *Ev* in strawberry, but as the markers were 11.8 cM and 15.8 cM on either side of the gene, they were not close enough to be useful in breeding selection. Later 127 seedlings obtained from a cross between the DN variety Tribute and the short day Honeoye (T×H) were scored with AFLP primer combinations to identify QTLs associated with the day neutrality habit.<sup>109</sup> The mapping approach allowed the identification of eight QTL linked to the trait, one of which explained the 36% of the variability, indicating that while day-neutrality is accepted as being controlled by a single dominant gene in cultivated strawberry progenies, the regulation of the trait under different environmental conditions may be polygenic, at least in the T×H cross. More recently, Gaston et al.<sup>110</sup> scored perpetual flowering as a dominant gene (PF) under Mendelian inheritance in the CA×CF population and mapped the trait to LG4b-f. When scored as a QTL, the locus was shown to colocalize with a

major dominant QTL that the authors reported also regulated the production of runners (RU), and thus, the locus was denoted *FaPFRU*. The results of this study have implications for the development of markers for the selection of these two traits, both of which are of immense importance to commercial strawberry production.

#### Fruit quality

Fruit quality traits are major targets in breeding programs and as such a better understanding of the genetic basis controlling these traits is a crucial step towards the implementation of MAS for fruit quality. Agronomic, physical and chemical traits have been evaluated in different studies to determine their genetic control.<sup>62,102,111</sup> In total, 87 unique QTLs were detected for 19 analysed traits in the study of Lerceteau-Köhler *et al.*,<sup>111</sup> 33 were detected for 14 of the 17 agronomic and quality traits studied in the 232×1393 progeny by Zorilla-Fontanesi *et al.*,<sup>62</sup> and 70 QTLs for 48 different volatile compounds were detected in the companion study performed by Zorilla-Fontanesi *et al.*<sup>102</sup> on the same population. Some of the considered traits were common between the studies and those such as fruit weight, firmness, soluble-solids content, titratable acidity, pH and anthocyanins, mapped to the same LG in the two mapping populations, and in all three works, a non-random distribution of QTLs was observed, with clustering of QTL observed. Another common feature of the three investigations was the presence of homoeo-QTLs, that is, QTL controlling a particular trait that mapped to orthologous positions on homoeologous linkage groups. In the two progenies studied in the three reports, plant width<sup>62</sup>, fruit shape, firmness, glucose and malate content, pH,<sup>111</sup> terpenes linalool and terpineol<sup>102</sup> were shown to be controlled by homoeo-QTLs, suggesting that more than one homoeologous gene copy regulates the expression of particular quality traits, and raising implications for marker development for MAS.

To date, linkage maps of the cultivated strawberry have been developed by scoring the presence or absence of alleles, i.e., using the alleles as single dose markers. This approach limits the efficiency of QTL mapping as information is discarded where two homeologous loci share a common (multidose) allele. If many of the most economically important traits in the cultivated strawberry are governed by homeo-QTL, a large amount of genetic information and potentially valuable markers would be lost through single-dose allele mapping, since homeo-QTL are likely to share alleles at loci linked to the genes regulating phenotypic expression. A possible solution to this limitation could be the scoring of multidose markers, as has been described by van Dijk *et al.*,<sup>98</sup> who were able to define map positions of multidose SSR markers through analysis following capillary electrophoresis by calculating allele dose based on their electropherogram peak intensities. The methodology proposed to score multidose markers is effective; however, it is labour intensive and is currently not automated, limiting its applicability. Furthermore, due to the nature of SSR variability and segregation in *F. ×ananassa*, reference markers and multiplexes would almost certainly have to be developed *de novo* for each progeny investigated.

#### RUBUS

The cultivated red raspberry, *R. idaeus* subsp. *idaeus*, has a complex pedigree and in many cases *R. idaeus* subsp. *idaeus* cultivars contain, to a greater or lesser degree, introgression from *R. idaeus* subsp. *strigosus* (the American red raspberry), a fact that has contributed to the high number of molecular markers heterozygous in many red raspberry mapping progenies. The distinction between the *idaeus* and *strigosus* subspecies is often dismissed by breeders due to the high levels of interfertility between the two species, but the geographic separation has certainly contributed to greater allelic diversity in modern cultivars. In contrast to the high diversity and thus, heterozygosity of *R. idaeus*, members of *R. occidentalis* (black raspberry) populations tend to be very homozygous

and very limited genetic diversity has been revealed in genetic studies of accessions from throughout the entire geographic range of the species.<sup>43,55</sup>

#### Linkage mapping in red raspberry

The first genetic linkage map for red raspberry (*R. idaeus* L.) was developed by Graham *et al.*<sup>37</sup> from a full-sib family from a cross between two cultivars Latham and Glen Moy (L×GM) belonging to different red raspberry subspecies, *R. idaeus* subsp. *strigosus* (American red raspberry) and *R. idaeus* subsp. *idaeus* (European red raspberry), respectively. The two cultivars used have very different morphological characteristics; while Glen Moy bears commercially acceptable fruits in terms of flavour, size and colour, on a spine-free plant, it is susceptible to diseases and low-temperature damage, while 'Latham' is in contrast extremely hardy, spiny and bears small fruits that are not of a commercially acceptable quality.

A total of 30 heterozygous SSR markers, 4 EST-SSR and 206 AFLP markers (240 markers in total) were mapped in the progeny to produce a linkage map covering a total of 789 cM across nine linkage groups. QTL analysis performed on the progeny identified two significant QTL on LG2 accounting for 48%–50% of the phenotypic variation related to spines on the canes, two QTL on LG8 that explained 33%–79% of the phenotypic variance associated with root sucker spread (diameter), and a single QTL also on linkage group 8 explaining 53% of the variation in root sucker density.

An additional 20 SSR markers were subsequently added to the same linkage map by Graham *et al.*<sup>112</sup>, along with additional markers that were scored, but not mapped in the study of Graham *et al.*<sup>37</sup> The updated map thus comprised 349 markers and covered a total distance of 669 cM. While some of the linkage groups from the initial study of Graham *et al.*<sup>37</sup> were joined following the addition of the new markers, there were still two groups, one from each parent, which could not be associated with each other and therefore, the updated map comprised eight linkage groups. The mapping population was scored for resistance to cane botrytis (*B. cinerea*), spur blight (*Didymella applanata*), cane spot (*Elsinoe veneta*) and yellow rust (*Phragmidium rubi-idaei*). The diseases were scored in two growing sites for presence or absence of the disease, as well as on a severity scale for cane botrytis and spur blight.

The gene controlling cane pubescence, gene *H*, was also mapped in the L×GM mapping population.<sup>112</sup> The homozygous form of the gene (*HH*) is rarely found because it is linked with a lethal recessive gene;<sup>113</sup> however, seedlings carrying a dominant allele for gene *H* in heterozygous form have been shown to be resistant to cane botrytis and spur blight,<sup>114,115</sup> but more susceptible to cane spot, powdery mildew (*Sphaerotheca macularis*) and yellow rust.<sup>116–118</sup> Gene *H* controlling cane pubescence was mapped to LG2 of 'Glen Moy' and its association with a QTL for resistance to cane botrytis and spur blight was confirmed.<sup>112</sup> An additional QTL for resistance to rust, spur blight and botrytis was identified on LG3 of the 'Latham' linkage map, along with QTL for resistance to cane spot on LG4 of the 'Glen Moy' map and rust on LG5 of the 'Latham' map.

A set of 23 EST-SSRs developed from cDNA libraries of roots and flower buds was subsequently mapped to the L×GM map by Woodhead *et al.*,<sup>45</sup> locating to six of the seven linkage groups of the map. Fourteen of the newly mapped markers were reported to be associated with QTL for disease resistance, fruit quality, fruit size and developmental stage, although the authors did not elaborate on how variation in these traits was quantified. Subsequently, additional phenotypic data relating to the stages of fruit development from bud break to over ripe fruit, fruit colour, the production of volatile compounds, cane height and cane splitting were scored in the L×GM progeny, and QTL and candidate genes associated variously with these traits were mapped to the seven *Rubus* linkage groups.<sup>119–121</sup> In its most recently updated version, 37 gene-specific markers were added to the L×GM linkage map, bringing the total

number of functional markers mapped in the progeny to 97.<sup>122</sup> The linkage map now contains a total of 223 molecular markers (57 AFLP and 69 SSR markers, along with the 97 functional markers), covering a total genetic distance of 840.3 cM across the seven expected groups for a consensus linkage map.<sup>122</sup>

#### Mapping aphid resistance

A second red raspberry linkage map was developed from a mapping progeny derived from a cross between the cultivars Malling Jewel and Malling Orion (MJ×MO) containing 94 seedlings. The map was composed of 95 AFLP and 22 SSR markers and covered a total distance of 505 cM over the expected seven linkage groups for a consensus map of the species.<sup>90</sup> The authors estimated that the MJ×MO genetic linkage map covered approximately 80% of the raspberry genome, with some regions in LG1, which was composed mainly of markers segregating in the maternal parent, and further regions of LG3 not completely covered. The map was used to determine the position of a major gene  $A_1$  involved in the resistance to biotype 1 of the aphid *Amphorophora idaei*, carried by the 'Malling Orion' parent, which was located on LG3, 5 cM from the codominant SSR marker Ru103a. Interestingly, the region on LG3 associated with the  $A_1$  resistance gene was also where QTLs for resistance to cane botrytis, spur blight and rust were located on the linkage map of Graham *et al.*<sup>112</sup> In addition to aphid resistance, dwarfism (*dw*) was also scored in the MJ×MO progeny. Different segregation models have been proposed for this trait, including a two gene model,<sup>123</sup> but the segregation data recorded for the MJ×MO progeny (3 : 1) suggested a single gene Mendelian model, in line with that proposed by Jennings.<sup>113</sup> However, the gene mapped to LG6 of the MJ×MO map, and was not linked to gene *H* as suggested by Jennings.<sup>113</sup>

#### Mapping *Phytophthora* root rot resistance

Linkage maps were also produced from a backcross progeny of the cross NY00-34 (Titan × Latham) × Titan comprising 159 individuals to investigate the inheritance of resistance to root rot (*Phytophthora fragariae* var. *rubi*). Linkage maps of the NY00-34 and Titan parents were generated comprising 138 AFLP, 68 RAPD and 20 resistance gene analogue polymorphism markers spanning 440 cM and seven linkage groups on the 'NY00-34' map, and 153 AFLPs, 47 RAPDs and 11 resistance gene analogue polymorphisms spanning 370 cM and seven linkage groups on the 'Titan' map.<sup>124</sup> QTL analyses performed on both parental maps for a number of disease criteria identified a region on LG1 of 'NY 00-34' and 'Titan' accounting for 30%, 61% and 25% and 26%, 33% and 29% of the variance for plant disease index, incidence of petiole lesions and root regeneration, respectively, while a second QTL region located on LG5 of 'NY 00-34' was responsible for 28%, 10%, 28% and 15% of the variance of plant disease index, stem lesion size, incidence of petiole lesions and root regeneration, respectively. A second QTL on LG7 of the Titan map was responsible for 15%, 18%, 14% and 16% of the variance for plant disease index, stem lesion size, incidence of petiole lesions and root regeneration score, respectively.

#### Development of a saturated sequence-characterized reference map for red raspberry

The most comprehensive linkage maps for red raspberry produced to date were reported recently by Ward *et al.*<sup>79</sup> Using GBS to generate SNP markers, supplemented with a genome-spanning SSR set, the authors produced highly saturated linkage maps of the parental lines of a mapping progeny produced from the cross Heritage × Tulameen (H×T) comprising 71 progeny. The study revealed almost twice the number of heterozygous markers in the Heritage genome than in the Tulameen genome. Linkage maps were composed only of markers segregating in the individual parental genotypes and spanned 462.7 cM across seven linkage

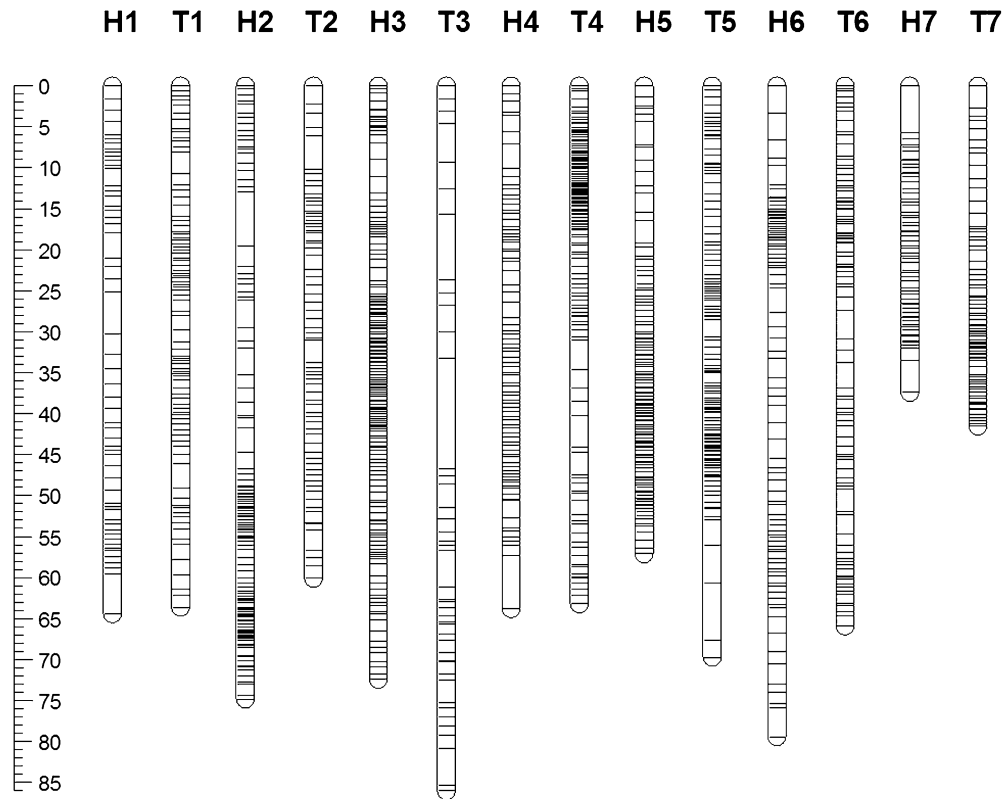
groups containing a total of 4521 SNP and 33 SSR markers in 487 genotyping bins on the Heritage linkage map, and 376.6 cM across seven linkage groups containing a total of 2391 SNP and 12 SSR markers in 274 genotyping bins on the Tulameen linkage map (Figure 2).

Due to the nature of the data generated by GBS, the authors proposed a novel imputation strategy that permitted linkage maps to be developed, despite ~30% missing data in each of the parental marker data sets. Almost complete colinearity was observed between the common SSR markers mapped between the H×T linkage map and those previously reported for the L×GM mapping progeny.<sup>37,112</sup> However, while the maps covered approximately the same physical distance in the two populations, significantly less genetic distance was covered by the H×T linkage maps,<sup>79</sup> suggesting that recombination rates in the parents of the L×GM map are much higher than those of other *Rubus* mapping populations. The density of molecular markers mapped in the study permitted the precise mapping of segregation distortion along the linkage groups of the two parental maps and the authors postulated the presence of lethal or detrimental sublethal effects of a number of genetic loci in the *R. idaeus* genome responsible for the high degree of segregation distortion observed on the H×T linkage maps. Since the SNP segregation data generated was derived from sequencing the progeny of the H×T mapping population, the dense saturation of these sequence-characterized markers across the seven *R. idaeus* linkage groups will be of immense utility for the anchoring and orientation of genome sequence scaffolds produced by the Heritage genome sequencing initiative.

#### Linkage map development in other *Rubus* species

*Linkage mapping in black raspberry.* Linkage mapping investigations were recently extended to other members of the *Rubus* genus. Bushakra *et al.*<sup>125</sup> developed a linkage map for black raspberry (*R. occidentalis*) for the purposes of comparison to maps of red raspberry and to other members of the Rosaceae. In their study, the authors developed an interspecific mapping population between an advanced selection of *R. occidentalis* (96395S1) displaying thornless canes and purple fruit, and the red raspberry cultivar Latham which is spiny and bears red fruits (S1×L). A linkage map of the S1×L progeny was constructed composed of 131 sequence-characterized markers; however, the majority of these were mapped in 'Latham' due to the inherently low genetic diversity in black raspberry.<sup>55</sup> The 96395S1 linkage map was composed of 29 markers over six linkage groups corresponding to linkage groups LG1–LG6 of the 'Latham' linkage map, and spanning a total genetic distance of 306 cM, while the 'Latham' map contained 114 markers spanning the expected seven linkage groups and covering a total genetic distance of 561 cM. Subsequently, Bushakra *et al.*<sup>125</sup> used the S1×L linkage map to identify compounds relating to the modification of cyanadin compounds in *Rubus*. The black raspberry fruits were characterized by two xylose-containing pigments, cyanidin 3-O-sambubioside and cyanidin 3-O-2G-xylosylrutinoside, while the red raspberries contained cyanidin 3-O-sophoroside and cyaniding 3-O-2G-glucosylrutinoside. In total, 27 QTL were identified in the progeny associated with the concentrations of five anthocyanins analysed with ultrahigh performance liquid chromatography that were detected in all 3 years of the study. On the 'Latham' parental map, two polyphenolic biosynthetic pathway genes and three transcription factor-derived markers were significantly associated with QTL for anthocyanin production. The QTL and their associated markers on LG2 and LG7, one of which encodes a putative component of the MYB/bHLH/WD protein complex involved in the regulation of anthocyanin biosynthetic pathway genes, were previously identified as influencing the concentrations of anthocyanin compounds in raspberry fruits by Kassim *et al.*<sup>126</sup> Thus, these loci were confirmed in the study of Bushakra *et al.*<sup>125</sup> and the authors concluded





**Figure 2.** The SSR- and SNP-based linkage map of the cultivar H×T mapping population developed by Ward *et al.*<sup>79</sup> comprising 33 SSR markers and 4521 SNP markers in 487 genotyping bins distributed throughout the seven linkage groups of the *Rubus* genome. Genetic distances are given in centiMorgan, cM.

that these loci were extremely important for the control of the accumulation of anthocyanins in raspberry fruits.

**Linkage mapping in blackberry.** As with other Rosoideae genera, *Rubus* species exist in a number of ploidy levels. Blackberry (*Rubus* subgenus *Rubus* Watson) is a species complex and thus, commercial cultivars have a number of species in their pedigree. Tupy is the most important commercial cultivar worldwide but has an uncertain pedigree (Comanche × wild Uruguayan blackberry) and an unknown ploidy level. The remaining important cultivars are contained in two groups: the western US developed cultivars (e.g., Marion, Black Diamond, Boysen) that have western and eastern US blackberry species as well as red raspberry (*R. idaeus* L.) in their pedigree and range in ploidy from 6× to 10× and the cultivars developed from eastern US blackberry species (i.e., Chester Thornless, Loch Ness, Ouachita, Prime-Ark 45) that are auto-tetraploids. The complex genetic background of the species complex has made linkage map construction for these species more challenging than for other polyploid Rosoideae species such as *F. ×ananassa* due to the complex polysomic inheritance patterns observed in autotetraploid segregating progenies. Using a full-sib BC<sub>1</sub> progeny of the cross PrimeJim (Arapaho × Ark.830) × Arapaho (PJ×AR) containing 188 individuals, a linkage map was developed for tetraploid blackberry to study the inheritance of two single gene traits, primocane fruiting and thornlessness.<sup>127</sup> Since PrimeJim is duplex for thorny habit (*ssSS*) and nulliplex for florican fruiting (*ffff*) and Arapaho is nulliplex for thorny habit (*ssss*) and duplex for florican fruiting (*ffff*), the progeny was observed to segregate 5 : 1 for both traits in accordance with tetrasomic inheritance in duplex × nulliplex crosses. A total of 120 molecular markers were scored in the progeny of which 40 uniparental markers segregated in simplex

(1:1), and six segregated in duplex (5:1). A further 74 markers segregated bi-parentally; 56 in double simplex (3:1), 14 in double duplex (35:1) and the remaining 15 in an 11:1 ratio indicating a simplex × duplex conformation (e.g., Aaaa × AAaa or the reciprocal). Highly distorted markers and those segregating 11:1 and 35:1 were excluded from the linkage analysis and a linkage map containing eight linkage groups for Arapaho and nine groups for PrimeJim was resolved; however, following comparison with common markers placed on other *Rubus* genetic maps, LG4 and LG5 in PrimeJim and LG5 in Arapaho were resolved into single linkage groups containing two fragments each.<sup>127</sup> Evidence for double reduction was observed at a number of loci, but no genotypes were observed that could only have originated from double reduction. The total length of the Arapaho and PrimeJim linkage maps was 788 cM and 768 cM, respectively. The thornless trait *S* was mapped to LG4 of the map, while primocane fruiting *F* was mapped to LG7.<sup>127</sup>

## ROSA

**Linkage mapping in diploid rose**

Debener and Mattiesch<sup>36</sup> were the first to present a genetic linkage map for *Rosa*, composed of AFLP and RAPD markers. The two parental linkage maps, produced from an F<sub>1</sub> population of 60 diploid *R. multiflora* accessions from the cross 93/1-117 × 93/1-119, were composed of a total of 305 molecular markers spanning the expected seven linkage groups for the species. In addition to the molecular markers, two morphological traits, petal number (*Blfo*) and petal colour (*Blfa*), were mapped to linkage groups 3 and 2 of the map, respectively. This linkage map was further extended to include an additional 469 markers comprising SSR and RFLP markers, along with gene-specific sequence characterised markers

derived from protein kinase and resistance gene analogue polymorphism genes.<sup>71</sup> The data set contained a significant number of biparental markers, which permitted the authors to construct an integrated map of both parents, spanning seven linkage groups and containing a total of 520 molecular markers covering 545 cM and estimated to cover more than 90% of the diploid rose genome.

Following the map of the 93/1-117 × 93/1-119 population, Crespel *et al.*<sup>128</sup> presented a linkage map of an interspecific cross between a dihaploid rose accession of the tetraploid *R. hybrida* and the diploid species *R. wichuriana* which contained 91 progeny. This linkage map, composed of AFLP markers, was used to map two major genes, recurrent blooming (*r4*) and double corolla (*db*), the latter being synonymous with the *Bflo* gene mapped earlier by Debener and Mattiesch,<sup>36</sup> along with two QTL for the number of thorns. The linkage map was later extended through the mapping of 64 EST- and genomic-SSR markers by Hibrand-Saint Oyant *et al.*,<sup>129</sup> which enabled the maps of the parental genotypes to be associated with one another. In that study, a QTL controlling flowering date was also identified. Linde *et al.*<sup>130</sup> developed a linkage map of a cross between parental lines differing in their susceptibility to powdery mildew (*Podosphaera pannosa*) and identified three QTL explaining over 80% of the variance for resistance to mildew race 9 and other QTL for resistance to natural populations of mildew. The progeny also segregated for four traits controlled by major genes, *Bflo* and *Bfla* mapped previously,<sup>36</sup> along with absence of prickles and a striped petal phenotype. The map was composed predominantly of AFLP and RGA markers, along with a modest number of SSR markers and a single SCAR, but the relatively high number (~30%) of biparental markers identified permitted an integrated linkage map of the progeny to be developed.

These three linkage maps, along with a fourth described by Shupert *et al.*<sup>131</sup> derived from the cross *R. chinensis* cultivar Old Blush × (*R. wichuriana* cultivar Basye's Thornless × Old Blush), were used by Spiller *et al.*<sup>132</sup> to develop an integrated consensus linkage map for rose. The consensus map contained a total of 597 markers, of which 206 were sequence characterised and 59 were common to all four linkage maps. The consensus map spanned a total map length of 530 cM and provided the means to propose a standardised linkage group nomenclature for *Rosa*. Perhaps most importantly, the consensus map permitted the location of ten phenotypic traits controlled by major genes, QTL for seven quantitative traits and 51 gene-based markers, providing a valuable resource for the development of tools for MAS in rose.

#### Linkage mapping in tetraploid rose

Despite the existence of many diploid rose species, modern rose cultivars are predominantly tetraploid, derived from around eight diploid and a smaller number of tetraploid progenitor species.<sup>13</sup> Diploid rose species are thus of immense value in understanding the genetics of specific traits, which would be desirable to introduce into commercial cultivars, and as a tool for unravelling the complex genetics of their tetraploid relations. However, such studies on the diploid are not directly applicable to rose breeding and are of limited value in understanding the inheritance of traits in commercial tetraploid rose genetic background.

Thus, following the development of diploid linkage maps for the genus, a linkage map developed from a tetraploid F<sub>2</sub> rose progeny was reported by Rajapakse *et al.*<sup>7</sup> The progeny studied was derived from a cross between an amphidiploid rose selection 86-7 and a tetraploid cultivar Basye's Blueberry (86-7 × BB). The meiotic behaviour of genotypes of this progeny had previously been investigated by Ma *et al.*,<sup>133</sup> who observed the multivalent formation of between 15% and 74%, and those authors suggested that the progeny exhibited partial tetrasomic inheritance. The linkage map of Rajapakse *et al.*<sup>7</sup> was produced from an F<sub>2</sub> population derived from the self-pollination of seedling 90-69 from the 86-7 × BB cross and

contained 52 plants. Approximately 70% of markers segregated in simplex in each parent, 20% in duplex, while the remaining 10% displayed other segregation types, including markers that displayed distorted segregation ratios. While the markers segregating in duplex appeared to display segregation patterns fitting to both 15:1 and 35:1 ratios, indicating that both disomic and tetrasomic inheritance occurred in the progeny, the population size was not large enough to discriminate between the two inheritance types. Thus, the resultant linkage map was composed exclusively of dominant markers segregating in simplex and therefore, separate parental maps were recovered, containing 15 linkage groups for the 86-7 map and 14 for the 82-1134 map.

Gar *et al.*<sup>134</sup> developed a linkage map of a cross between two tetraploid *R. hybrida* cultivars Golden Gate and Fragrant Cloud (GG × FC). Assuming tetrasomic inheritance due to the segregation ratios observed in the progeny, they mapped 403 markers, which were variously sequence-characterized RFLP, SSR and SNP (cleaved-amplified polymorphic site) markers (65%) and arbitrary AFLP markers (35%) to seven consensus linkage groups for each parent spanning 632 cM and 616 cM for the female and male parents, respectively. Each linkage group corresponded to a set of four coupling phase homeologous groups. Due to the high number of sequence-characterized markers mapped in the study, the authors were able to show the homology between the tetraploid map and the diploid consensus map of Spiller *et al.*<sup>132</sup> and to determine a high degree of conservation in macrosyntentic genome structure between the *Rosa* and *Fragaria* genomes.

Further work investigating the modes of inheritance in tetraploid rose progenies was performed by Koning-Boucoiran *et al.*<sup>65</sup> Using a progeny from the cross P540 × P867 consisting of 184 genotypes, segregation of AFLP, nucleotide-binding site and SSR markers segregating in both simplex and duplex with both uniparental and biparental inheritance was scored. While the hypothesis of complete disomic inheritance in the progeny was rejected, marker segregation data displayed evidence for both disomic and tetrasomic inheritance, and so the authors proposed the hypothesis that the progeny conferred tetrasomic inheritance with some preferential pairing occurring on certain chromosomes. Evidence for double reduction at some codominant SSR loci was also observed; however, it was based on the assumption that all alleles scored mapped to the same locus, and as with the data of Castro *et al.*<sup>127</sup> for blackberry, the mapping data they presented did not conclusively support this assumption. In total, the expected 28 linkage groups (seven sets of four coupling phase linkage groups) were recovered for the female map and 30 groups were recovered for the male map, but not all of the groups contained SSR markers and so four female and five male groups remained unassigned. Using the map, the authors identified three QTL for prickles on the stem explaining 44.1% of the trait variance, and single QTL for both petal number and powdery mildew resistance explaining 12.7% and 8.5% of the variance, respectively.

#### GENOME SEQUENCING INITIATIVES

High molecular weight DNA library development and sequencing High molecular weight bacterial artificial chromosome (BAC) and fosmid libraries have been developed for *Fragaria*<sup>135,136</sup> and *Rosa*,<sup>137</sup> while in *Rubus*, a protocol for high molecular weight DNA extraction suitable for BAC library construction was published<sup>138</sup> and the use of the subsequently-constructed BAC libraries was reported in the literature,<sup>139</sup> but the characterisation of the BAC library was not. The 18 432-clone *Fragaria* BAC library was characterized through the development of pools for PCR-based screening and clones containing markers from the diploid *Fragaria* reference map were identified.<sup>135</sup> However, the library awaits full exploitation. Additionally, a 33 000-clone fosmid library has been developed from the *Fragaria* species *F. vesca* subsp. *americana*, which was

characterized through sequencing of a randomly selected set of 30 clones containing ~1 Mbp of genomic DNA sequence and gave the first insight into gene content, composition and distribution in the *Fragaria* genome.<sup>136</sup> A BAC library developed for *Rosa* was used to assemble a BAC contig containing a minimal tiling path of six clones spanning a ~400 kbp region of the telomeric section of LG1 of the *R. rugosa* genome containing a gene conferring resistance to the blackspot fungus (*Diplocarpon rosae*).<sup>137</sup> Subsequent PCR analysis and sequencing of the clones contained in the 340.1 kbp contig revealed the presence of sequences encoding nucleotide-binding site domains typical of resistance gene analogues and from this region, molecular markers tightly linked to the resistance locus were subsequently developed.<sup>140,141</sup> In *Rubus*, two BAC clones associated with QTL for *Phytophthora* root rot resistance located on LG3 and LG6 of the *Rubus* genome were sequenced and genes putatively associated with root vigour were identified, which the authors speculated may possibly have a role in *Phytophthora* root rot resistance.<sup>139</sup>

#### Plastid sequencing and applications

The chloroplast genome of *F. vesca* was recently described.<sup>46</sup> The genome is 155 691 bp long and encodes a total of 78 proteins, 30 tRNAs and 4 rRNAs, and its sequence has been utilized as a powerful tool for phylogenetic and population genetics studies within the genus. While previous studies in *Fragaria* have used chloroplast DNA sequences for phylogenetic analysis,<sup>64,142</sup> the publication of the complete chloroplast DNA sequence for *F. vesca* facilitated a phylogenetic treatment of the genus *Fragaria* by Njuguna *et al.*<sup>10</sup> based on whole chloroplast DNA re-sequencing. The study revealed compelling evidence for the maternal ancestry of the extant octoploid *Fragaria* species, indicating that the origins of the genus rest in East Asia and that the maternal origins of the polyploid *Fragaria* are rooted in the Americas. The study also permitted hypotheses to be formulated about the evolution of the mating and sexual systems in the genus. More recently, parental species-specific SNP markers identified through whole chloroplast re-sequencing were used in conjunction with nuclear SSR markers to study hybrid zones of *F. virginiana* and *F. chiloensis*, which showed that there is a high degree of natural interfertility, significant admixture and relatively little maternal bias among parental and hybrid species, and revealed the extent of the genetic complexity within octoploid *Fragaria* hybrid zones.<sup>143</sup>

Recently, the whole chloroplast genome sequence of *Potentilla micrantha* was sequenced using the Pacific Biosciences PacBio RS sequencing platform.<sup>144</sup> The genome spanned a total of 154 959 bp and was structurally similar to the *F. vesca* chloroplast genome, containing the two expected inverted repeats spanning 25 530 bp in length each, a large single copy and small single copy region covering 85 137 bp and 18 762 bp, respectively, and 120 genes, 21 of which were duplicated in the inverted repeats, giving a total of 141 genes of known function (Figure 3). Through comparison with the *F. vesca* chloroplast genome sequence, the authors determined that gene order within the chloroplasts of the two species was identical. The chloroplast genome of *P. micrantha* was the first to be sequenced using the PacBio RS platform, which, following *de novo* assembly, yielded a single, highly accurate contiguous sequence spanning the entire chloroplast, and suggesting that PacBio RS data may be of immense utility in whole-genome shotgun sequencing and BAC-sequencing projects.

#### Whole-genome shotgun sequencing projects

The advent of second generation sequencing platforms made whole-genome shotgun sequencing of moderately-sized genomes and *de novo* assembly of draft genome sequences a realistic prospect for non-model organisms. Indeed, the continuing development of second generation sequencing technology, with greater

throughput and longer sequencing reads, is driving down the cost of sequencing, and this has permitted the initiation of sequencing efforts for genomes of Rosoideae species by interested consortia, without the requirement for central funding for the projects. Thus, the sequencing efforts are community-driven, community-funded endeavours, the results of which are proving to be of immense benefit for the communities that developed them as well as many other researchers in the plant sciences field. The various sequencing projects in *Fragaria*, *Potentilla*, *Rosa* and *Rubus* that are currently underway are at various stages of completion. While the genome of *F. vesca* has been published and is publically available,<sup>46</sup> for others, such as those of *P. micrantha*<sup>144</sup> and *R. chinensis*, very little public information was available at the time of writing.

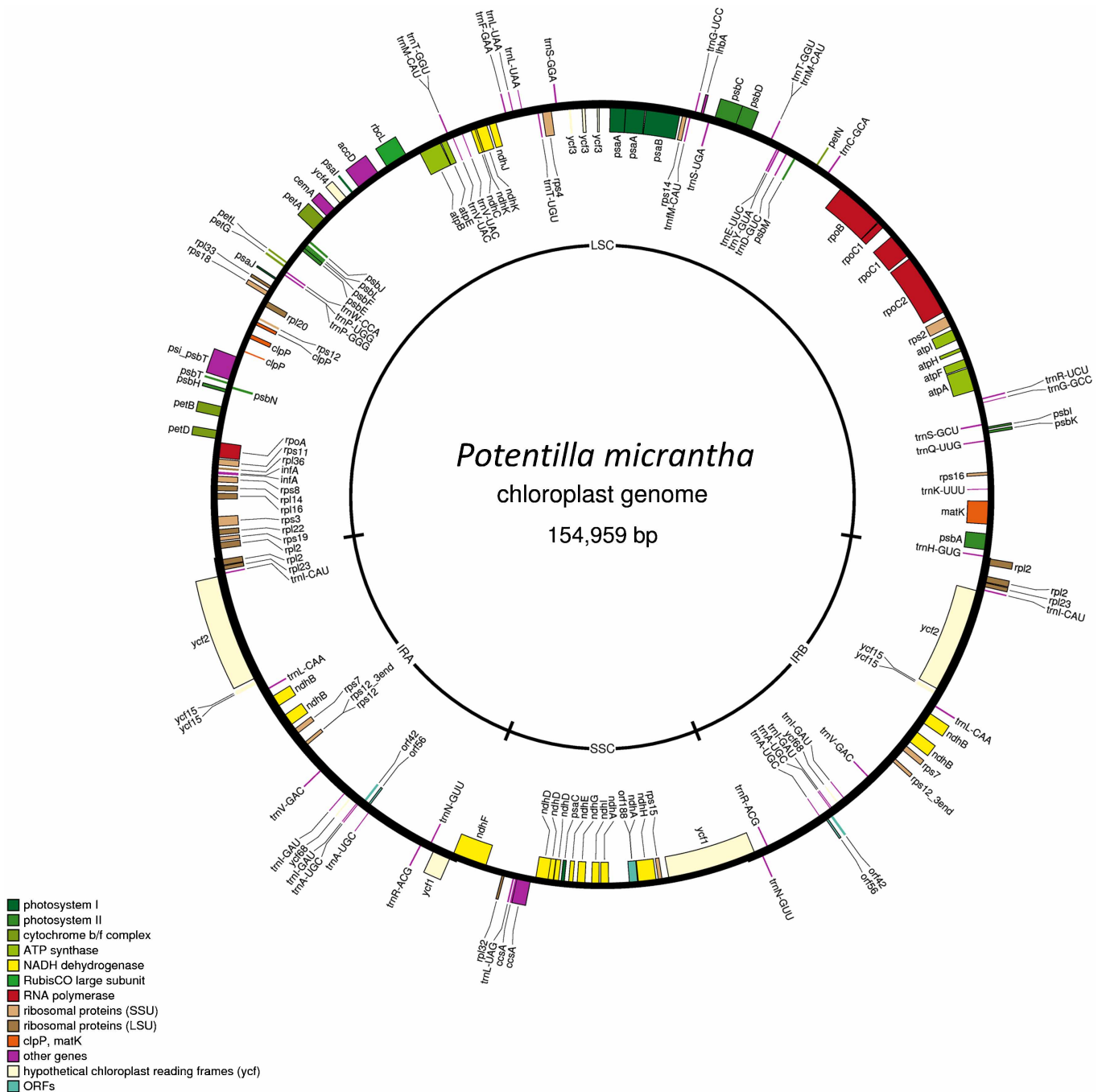
***Fragaria vesca* genome sequence.** The first whole-genome sequencing initiative in the Rosoideae was the genome of the woodland strawberry, *F. vesca*. The genome of *F. vesca* cultivar Hawaii 4 (FvH4) was sequenced by an international consortium, and since it was the first genome to be characterised using exclusively short-read technology, the project represented a significant milestone in plant genomics.<sup>145</sup> The 39× average depth of sequence coverage of the genome was achieved using predominantly data generated with the Roche 454 platform, supplemented with Illumina GAI and ABI/SOLID data, and was assembled *de novo* into ~3200 sequencing scaffolds with an N50 of 1.3 Mb. The majority of the sequenced genome (209.8 Mb, representing in excess of 95% of the total assembled genome length) was contained in 272 scaffolds of which 222, covering 197.7 Mb (94%), were anchored to the diploid *Fragaria* reference map to create seven pseudo-chromosomes (Figure 4).<sup>93</sup>

Multiplexed cDNA pools representing a comprehensive set of diverse tissues were sequenced and a set of 34 809 hybrid gene models, of which over 90% were supported by transcript-based evidence, were presented. A total of 957 genes identified in the sequence were unique to strawberry, of which 416 were assigned gene ontology categories, while the remaining 541 were previously unidentified predicted proteins. The *F. vesca* sequence, the first genome of the Rosoideae and second only to *Malus* to be sequenced for the Rosaceae, is one of the smallest plant genomes so far characterised, and is just 25% larger than the genome of *Arabidopsis thaliana*.<sup>26</sup> Its sequencing represents a gateway to functional and structural genomics investigations in the genus, and it has already found many applications in genetics and genomics research, enabling detailed comparisons with the genomes of other related genera,<sup>146</sup> permitting the fine mapping<sup>83</sup> and positional cloning<sup>147</sup> of genes controlling specific phenotypes, and enabling unprecedented numbers of markers to be identified and mapped in both diploid<sup>83</sup> and octoploid *Fragaria* species.<sup>48</sup>

#### *Rubus* genome sequencing initiatives

Two reference genome sequences are currently under construction for *Rubus*. One from the commercially important red raspberry, *R. idaeus*,<sup>148</sup> and the other from the black raspberry, *R. occidentalis*,<sup>149</sup> which is only grown commercially in specific regions of the United States. Both species are diploid, and the genome sizes of the two species are similar at approximately 350 Mb and 300 Mb for *R. idaeus* and *R. occidentalis*, respectively. Despite their close relationship and small genomes, the assembly strategy for the two genomes differs based on their relative heterozygosity.

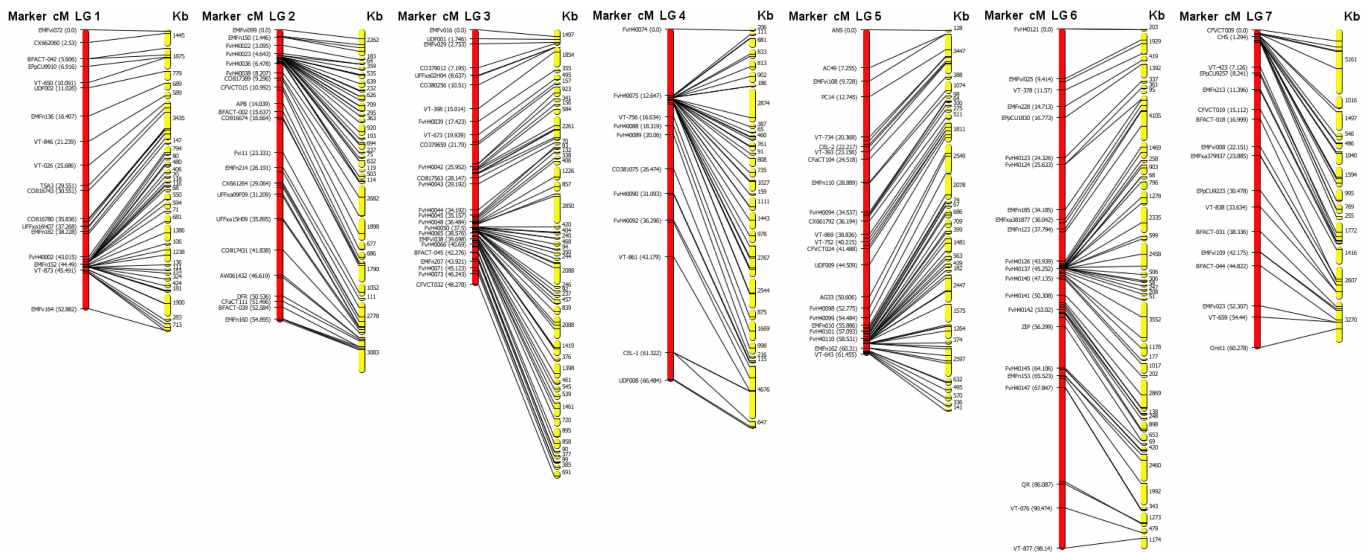
The cultivar Heritage was selected for the red raspberry sequencing project, because of its importance as one of the first commercially important primocane fruiting raspberry varieties. Heritage was initially sequenced to over 100× coverage using a combination of Roche 454 and Illumina technologies. Standard assembly algorithms, such as Newbler (Roche, Basel, Switzerland) and ABYSS resulted in fragmented assemblies. An assembly covering about



**Figure 3.** The chloroplast genome of *Potentilla micrantha* assembled from PacBio RS and Illumina HiSeq2000 data<sup>144</sup> detailing genes transcribed clockwise (inside circle) and counterclockwise (outside circle). Genes are coloured according to functional categorisation, IRA and IRB denote inverted repeats and LSC and SSC denote long and short single copy regions, respectively.

252 Mb of the genome with an N50 scaffold size of about 218 000 bp was produced,<sup>148</sup> with an estimated SNP rate of approximately 1.5%. To improve the assembly, novel algorithms aimed at augmenting the problem of the heterozygous assembly were specifically developed (J Price, personal communication), and thus, currently, using a two-pronged approach which takes advantage of recently-developed Molecule technology (Illumina) to produce single-molecule derived 'long-read' contigs together with improved mate pair library construction methodologies, the contiguity of the assembly is being significantly improved.<sup>150</sup>

Parallel to the sequencing initiative in red raspberry, a preliminary draft genome of the black raspberry has been produced from the sequencing and assembly of data generated from clone 'ORUS 4115-3'.<sup>149</sup> The particular clone was chosen because of its highly homozygous genome, and thus extremely low SNP rate of just 0.06%, which is 25 times less than that of the red raspberry cultivar Heritage. A combination of overlapping fragment and mate pair libraries was sequenced to a combined coverage of over 370×, resulting in an assembly that produced 2226 scaffolds with an average size of approximately 107 000 bp, an N50 of approximately



**Figure 4.** The seven pseudochromosomes (yellow) of the *Fragaria vesca* ‘Hawaii 4’ (FvH4) genome sequence as anchored to the FV×FB reference map (red).<sup>93</sup> The pseudochromosomes are composed of a total of 222 sequencing scaffolds covering 197.7 Mb of sequence, representing 94% of the *F. vesca* genome. cM, centiMorgan.

350 000 bp and genome reconstruction of approximately 240 Mbp. The *Rubus* genome sequencing initiatives are currently ongoing, but the expectation is that both sequences will be published and made publically available in 2014.

Evidence that whole-genome sequencing is becoming a commonplace tool for genetic analysis in plant research increased recently when it was reported that the genome of a relatively unknown Rosoideae species *P. micrantha* was being sequenced using both Illumina and PacBio sequencing platforms.<sup>144</sup> The species, while possessing little perceived economic value, is being developed to study a number of evolutionary processes that have led to the expression of economically important traits in strawberry. The fact that whole-genome shotgun sequencing is being used as an economically viable tool to answer such questions in a non-model, non-crop plant, along with the increasing use of genotyping by sequencing for linkage map development and genomic analysis, suggests that *x*-generation sequencing is coming of age and will soon be as commonplace as the use of PCR in the plant molecular genetics and genomics toolkit.

### COMPARATIVE ANALYSES WITH OTHER SPECIES

Since, at the time of writing, only the genome of a single Rosoideae species, *F. vesca*, is publically available, interspecific synteny studies within the subfamily have relied on comparisons between the genetic positions of markers in *Rubus* and *Rosa* linkage maps with the *F. vesca* genome sequence to infer synteny between the genomes of the subfamily. Gar *et al.*<sup>134</sup> were the first researchers to evaluate the conservation of macrosynteny between genera within the Rosoideae. In their study, they exploited the conserved sequence of EST markers mapped in *Rosa* to infer evolutionary relationships between chromosome structure of rose and diploid strawberry. BLAST analysis of the sequence of 70 EST markers mapped to the GG×FC linkage map that had a single clear unambiguous position on the FvH4 genome sequence revealed similar distributions of markers in the genomes of the two species, with conserved chromosomal locations of 77% (54) of the markers. Indeed synteny between the genomes of the two species was high, with five *Rosa* linkage (RG3-7) groups corresponding to a single *Fragaria* chromosome each and the remaining two groups showing evidence for a single fusion–fission event in *Fragaria* leading to the evolution of *Fragaria* chromosome 6. In fact, just four translocations and six

inversions were required to place all orthologous markers in the same order between the two genomes. Later, using the L×S1 linkage map, Bushakra *et al.*<sup>67</sup> compared the genetic positions of 77 *Rubus* markers with conserved sequences to their positions on the FvH4 genome sequence and evaluated the conservation of synteny between the two genera. Each of the *Rubus* linkage groups in that study was shown to correspond with a single *Fragaria* chromosome, suggesting a high degree of conservation of synteny between *Rubus* and *Fragaria* as had been demonstrated previously between *Rosa* and *Fragaria*.<sup>134</sup> Following the conclusion that each of the seven haploid chromosomes of *Rubus* had a specific identifiable orthologue in *Fragaria*, the authors proposed that the *Rubus* linkage groups be renamed to correspond with those of *Fragaria*, and this revised nomenclature was adopted by Ward *et al.*<sup>79</sup> for their SNP-based map of *R. ideaus* and Castro *et al.*<sup>127</sup> on their blackberry linkage map.

Recently, large numbers of markers mapped in the cultivated strawberry<sup>48</sup> and *F. vesca* subsp. *bracteata*<sup>83</sup> have enabled a detailed comparison of the genomes within the genus *Fragaria*. While the studies revealed a high degree of colinearity between the FvH4 genome sequence and the genomes of both the cultivated strawberry and *F. vesca* subsp. *bracteata*, both investigations demonstrated that the mapped positions of a proportion of markers disagreed with their expected chromosomal locations on the FvH4 genome sequence. In the case of the comparison of the *F. vesca* subspecies,<sup>83</sup> 22 inversions, 11 intrachromosomal translocations and 40 interchromosomal translocations were observed between the *F. vesca* subsp. *bracteata* map and the ‘Hawaii 4’ pseudochromosomes. The reported differences were all supported with more than one mapped genetic marker and thus, the authors postulated that these differences could have been the result of real evolutionary differences between the *F. vesca* subspecies, errors in the genome assembly of the FvH4 sequence or rearrangements resulting from the hybrid nature of the linkage map (FV×FB) used to anchor the FvH4 sequence. The cost of map development using large numbers of SNP markers is becoming ever more affordable as sequencing technologies and GBS protocols develop, and thus, the development of a *F. vesca* linkage map saturated with sequence characterized SNP markers would enable a rigorous testing of the anchoring of the FvH4 sequence scaffolds, as well as the quality of the assembly of the sequencing scaffolds themselves.

### Comparative analyses of floral regulation in diploid *Rosa* and *Fragaria*

Comparative genomics has recently been used within the Rosoideae to demonstrate the role of orthologous genes in the regulation and control of a trait of economic importance in both rose and strawberry. Timing of flowering is an extremely important aspect of a plant's reproductive biology, and the ability to flower several times during a year is of significant economic importance in ornamental plants, such as roses for which the value is in the flowers, as well as fruit crops like strawberry. In diploid *Rosa* and *Fragaria*, two flowering types have been observed, a single period of flower initiation, described as 'once-flowering' in the context of roses by Iwata et al.<sup>151</sup> and more commonly referred to as 'short-day' flowering in the context of strawberry,<sup>152</sup> and continuous flowering, also referred to as day neutral, or *semper-florens* type in *Fragaria*.<sup>153</sup> Continuous flowering is controlled in strawberry by a recessive mutation in a single gene at the locus first described as *s* by Brown and Wareing<sup>154</sup> and later as the *seasonal flowering locus* (*sfl*) to denote the wild-type phenotype by Albani et al.<sup>152</sup> In rose, the CF phenotype was described first as *r4* by Crespel et al.<sup>128</sup> and later as *recurrent blooming* (*rb*) by Spiller et al.<sup>132</sup> The *sfl* and *rb* genes were mapped to linkage groups FG6 and RG3 of the diploid *Fragaria* and *Rosa* linkage maps respectively by Sargent et al.<sup>20</sup> and Spiller et al.,<sup>132</sup> which following the synteny comparisons of Gar et al.<sup>134</sup> were shown to be homologous chromosomes, suggesting that the genes controlling *sfl* and *rb* were also homologous.

The *rb* locus was shown to co-segregate initially with the RoSPINDLY gene by Remay et al.<sup>155</sup> and subsequently also with the RoKSN gene by Kawamura et al.<sup>156</sup>, which were confusingly mapped to LG4 of the diploid rose progeny from the cross The Fairy × *R. wichurana* (TF×RW) used in those studies. This linkage group, however, is the homologue of RG3 on the map of Spiller et al.<sup>132</sup> and the colocation of the BFACT-047 SSR previously shown to map closely to the *sfl* locus in the diploid *Fragaria* FV×FB progeny supported the homology of the *rb* and *sfl* loci in *Rosa* and *Fragaria*. RoKSN is a homologue of the floral repressor TERMINAL FLOWER 1 (TFL1), which acts as photoperiodically regulated repressor of flowering. In the TF×RW population, plants conferring the *rb* phenotype were shown to be homozygous for a mutant allele of the RoKSN gene containing a 9-kbp *copia*-like retrotransposon in its second intron. These same plants were shown not to accumulate RoKSN mRNA and it was hypothesized by the authors that the retrotransposon blocked the synthesis or accumulation of the mRNA and they proposed a model for the regulation of seasonal flowering by RoKSN.<sup>151</sup> In the same study, the authors showed a 2 bp deletion in the coding region of the *Fragaria* homologue of TFL-1 (termed *FvKSN* in that study) caused a truncated TFL-1 mRNA, which coincided with the *sfl* phenotype.<sup>151</sup> A concurrent study by Koskela et al.<sup>147</sup> provided functional evidence that TFL-1 is the gene responsible for floral repression in diploid *Fragaria*. In their study, the authors showed that TFL-1 cosegregated with the *sfl* locus and that it confers a short-day requirement for flowering through repression of FvAP1/FUL. Additionally, they demonstrated that the 2 bp deletion in the mutant TFL-1 gene leads to FT1 dependent long day flowering in *F. vesca*.

Such studies demonstrate the utility of both whole-genome sequences, and comparative genomics for the investigation of key traits of agronomic importance in closely related species. It is likely that other traits controlled by both single genes and QTL in the Rosoideae are governed by homologous genes, and thus, as sequencing data from a growing number of species within the subfamily accumulate, the elucidation of the genetic basis of traits in one species will help inform studies in other species and genera and enable the development of markers for MAS in numerous Rosoideae species.

### Development of genome-wide functional genomics resources for Rosoideae species

Numerous studies have been performed investigating the structure, function and expression patterns of specific genes in Rosoideae species. A comprehensive review of this body of excellent work is beyond the scope of this article; however, recently, a number of studies have exploited second-generation sequencing technology to characterize the global transcriptome of Rosoideae species challenged with a variety of different biotic and abiotic stresses.

Prior to the sequencing of the genome of FvH4, Folta et al.<sup>157</sup> published a data set derived from mRNA from *F. ×ananassa* cultivars that had been subjected to a range of treatments including harvesting at different stages of diurnal rhythms, plants subjected to different light conditions, the application of growth regulators along with biotic stresses and treatment with a variety of chemicals. mRNA was sequenced using the Roche 454 platform and assembled into over 32 000 contigs, representing a substantial proportion of the expressed portion of the cultivated strawberry genome, as well as characterizing allelic variation in the genes identified that the authors suggested could be applied to the development of genomics tools for breeding.

Likewise, in the other major Rosoideae species, for which genome sequences are currently not available, a number of approaches have been used to characterize the global transcriptome in specific tissues or during specific stresses. In *Rubus*, Ward et al.<sup>158</sup> sequenced mRNA pools derived from resistant and susceptible *R. idaeus* plants challenged with the *Phytophthora rubi* and their corresponding uninoculated controls. Exploiting the genetic similarity between *Rubus* and *Fragaria* transcripts, they aligned *Rubus* transcripts against the *F. vesca* genome sequence<sup>46</sup> and identified over 16 000 putatively expressed genes. Additionally, *de novo* assembly of transcripts enabled the identification of sequences unique to *Rubus* and also to the pathogen the plants were challenged with. The authors reported that this approach constituted a powerful analysis method in non-model organisms and could be used to investigate specific, testable hypotheses.

While no genome sequence currently exists for rose, much work has been performed characterizing the expressed portion of the *Rosa* genome. Using data generated with both the Roche 454 and the Illumina platforms, a transcriptome database resource was created from mRNA isolated from various stages of floral development, leaves and stems, as well as from drought stressed plants.<sup>159</sup> In total, 80 714 ESTs based on transcript clusters longer than 100 nucleotides and containing more than two sequences were identified from which 76 611 peptides were predicted, 20 997 of which were clustered into 13 900 protein families. Comparison with the *F. vesca* transcript set<sup>46</sup> revealed that 44 656 clusters had significant homology to 14 252 *Fragaria* gene predictions, with a mean identity of over 90%. Thus, 53% of the predicted transcripts were represented in the *Rosa* database. Similarly, Kim et al.<sup>160</sup> employed deep transcriptome sequencing from rose flowers of four cultivated varieties of *R. hybrida* to generate a database containing an average of 35 385 transcripts per cultivar.

Recently, Kang et al.,<sup>161</sup> exploiting the structural genomics resources developed for *F. vesca*,<sup>46</sup> performed genome-scale transcriptomic analyses of floral tissue of Hawaii 4 plants at various stages of fruit development from floral anthesis to enlarged fruits. By studying the developmental changes at the whole transcriptome level, they were able to determine that the endosperm and seed coat of strawberry achenes may play a more vital role in auxin and gibberellin biosynthesis than the embryo itself. Their study provided a global and comprehensive overview of the changes in gene expression during this economically important process, providing a platform for future studies of individual genes and pathways involved in fruit set and subsequent development both in *Fragaria* and other species within the Rosoideae and the Rosaceae as a whole.

These resources, along with the annotated gene predictions derived from the *F. vesca* 'Hawaii 4' genome sequence by Shulaev *et al.*<sup>46</sup>, provide a wealth of global genomic information from the expressed portion of the genomes of the three major Rosoideae genera. Such resources, in combination with the structural genomics resources developed for these species, afford researchers working with Rosoideae species the opportunity to link structural and functional information relating to traits of interest and to develop both a detailed understanding of the genetic mechanisms controlling important agronomic traits, and the development of functional markers for use in breeding and selection.

### CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The past two decades have seen prolific developments in the molecular genetics and genomics of Rosoideae species. Like in other plant species, a variety of techniques have been used to construct genetic linkage maps from segregating progenies, including the use of molecular markers ranging from isoenzymes and RFLPs,<sup>66</sup> arbitrary PCR-based markers, sequence-characterized PCR-based markers such as SSRs<sup>48</sup> and the development of array-based and second-generation sequencing-based SNP marker characterization and their application.<sup>79,83</sup> Developments in second- and third-generation sequencing technologies led to the first genome sequence of a Rosoideae plant, *F. vesca*,<sup>46</sup> and the promise of sequences of an additional four Rosoideae genomes in the near future. These sequence resources enable an improved understanding of the conservation of synteny between Rosoideae species and genera<sup>67,134</sup> and will assist in rapidly expanding our basic understanding of key agronomic traits.

Molecular genetics and genomics studies of Rosoideae species have benefited the whole production system, from breeding to nurseries and growers, by providing tools to effectively evaluate genetic diversity and to fingerprint for variety identification. The tools have also led to a greater understanding of the genetics of numerous traits of agronomic importance including pest and disease resistance,<sup>112</sup> anthocyanin and flavonol production affecting fruit quality<sup>102,125</sup>, the timing of flowering,<sup>147</sup> and have aided in the development of molecular markers linked to a range of traits desirable in modern soft-fruit and cut flower cultivars.<sup>141</sup>

To fully realize the potential of the molecular genetics and genomics data that have been amassed for Rosoideae species, MAS will need to be fully integrated into breeding programmes. To meet this goal, there must be an open dialogue between breeders and genomicists leading towards a more focused research strategy dependent on accurate phenotyping. This will lead to high-throughput and user-friendly markers for deployment in breeding and selection of strawberries, raspberries, roses and other Rosoideae crops of the future. With the continued development of ever more cost-effective genomics solutions, we predict that MAS will be quickly incorporated into breeding programs worldwide, permitting the development of varieties with improved resistance to pests and disease, quality and yield.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGMENTS

Rosoideae genome research and comparative genomics at FEM are supported by the research office of the Provincia autonoma di Trento.

### REFERENCES

- Hutchinson J. *The genera of flowering plants, vol. 1, Dicotyledons*. Oxford: Clarendon Press, 1964.
- Schulze-Menz GK. *Rosaceae*. Berlin: Gebrüder Borntraeger, 1964.

- Morgan DR, Soltis DE, Robertson KR. Systematic and evolutionary implications of rbcL sequence variation in Rosaceae. *Am J Bot* 1994; **81**: 890–903.
- Potter D, Gao F, Bortiri PE, Oh SH, Baggett S. Phylogenetic relationships in Rosaceae inferred from chloroplast *matK* and *trnL-trnF* nucleotide sequence data. *Plant Syst Evol* 2002; **231**: 77–89.
- Potter D, Eriksson T, Evans RC *et al*. Phylogeny and classification of Rosaceae. *Plant Syst Evol* 2007; **266**: 5–43.
- Staudt G. The species *Fragaria*, their taxonomy and geographical distribution. *Acta Hort* 1989; **265**: 23–33.
- Rajapakse S, Byrne DH, Zhang L, Anderson N, Arumuganathan K, Ballard RE. Two genetic linkage maps of tetraploid roses. *Theor Appl Genet* 2001; **103**: 575–583.
- Eriksson T, Donoghue MJ, Hibbs MS. Phylogenetic analysis of *Potentilla* using DNA sequences of nuclear ribosomal internal transcribed spacers (ITS), and implications for the classification of Rosoideae (Rosaceae). *Plant Syst Evol* 1998; **211**: 155–179.
- Alice LA, Campbell CS. Phylogeny of *Rubus* (Rosaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *Am J Bot* 1999; **86**: 81–97.
- Njuguna W, Liston A, Cronn R, Ashman TL, Bassil N. Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. *Mol Phylogenet Evol* 2013; **66**: 17–29.
- Matsumoto S, Kouchi M, Yabuki J, Kusunoki M, Ueda Y, Fukui H. Phylogenetic analyses of the genus *Rosa* using the *matK* sequence: molecular evidence for the narrow genetic background of modern roses. *Sci Hort* 1998; **77**: 73–82.
- Zimmermann H, Ritz CM, Hirsch H, Renison D, Wesche K, Hensen I. Highly reduced genetic diversity of *Rosa rubiginosa* L. populations in the invasive range. *Int J Plant Sci* 2010; **171**: 435–446.
- Rajapakse S, Zhang L, Ballard RE, Byrne DH. AFLP marker development in rose for genetic mapping: comparison of three restriction enzyme pairs. *Acta Hort* 2001; **546**: 619–627.
- Iwatsubo Y, Naruhashi N. Karyotypes of three species of *Fragaria* (Rosaceae). *Cytologia* 1989; **54**: 493–497.
- Iwatsubo Y, Naruhashi N. Cytotaxonomical studies of *Rubus* (Rosaceae) I. Chromosome numbers of 20 species and 2 natural hybrids. *J Jpn Bot* 1992; **67**: 270–275.
- Akasaka M, Ueda Y, Koba T. Karyotype analysis of wild rose species belonging to septets B, C, and D by molecular cytogenetic method. *Breeding Sci* 2003; **53**: 177–182.
- Nathewet P, Yanagi T, Hummer KE, Iwatsubo Y, Sone K. Karyotype analysis in wild diploid, tetraploid and hexaploid strawberries, *Fragaria* (Rosaceae). *Cytologia* 2009; **74**: 355–364.
- Potter D, Still SM, Grebenc T *et al*. Phylogenetic relationships in tribe Spiraeae (Rosaceae) inferred from nucleotide sequence data. *Plant Syst Evol* 2007; **266**: 105–118.
- Thompson MM. Survey of chromosome numbers in *Rubus* (Rosaceae:Rosoideae). *Ann Missouri Bot Garden* 1997; **84**: 128–164.
- Sargent DJ, Davis TM, Tobutt KR, Wilkinson MJ, Battey NH, Simpson DW. A genetic linkage map of microsatellite, gene-specific and morphological markers in diploid *Fragaria*. *Theor Appl Genet* 2004; **109**: 1385–1391.
- Randell RA, Howarth DG, Morden CW. Genetic analysis of natural hybrids between endemic and alien *Rubus* (Rosaceae) species in Hawaii. *Conserv Genet* 2004; **5**: 217–230.
- Bringhurst RS, Voth V. Breeding octoploid strawberries. *Iowa State J Res* 1984; **58**: 371–381.
- Yokoya K, Roberts AV, Mottley J, Lewis R, Brandham PE. Nuclear DNA amounts in roses. *Ann Bot* 2000; **85**: 557–561.
- Dickson EE, Arumuganathan K, Kresovich S, Doyle JJ. Nuclear-DNA content variation within the Rosaceae. *Am J Bot* 1992; **79**: 1081–1086.
- Meng RG, Finn C. Determining ploidy level and nuclear DNA content in *Rubus* by flow cytometry. *J Am Soc Hort Sci* 2002; **127**: 767–775.
- Akiyama Y, Yamamoto Y, Ohmido N, Ohshima M, Fukui K. Estimation of the nuclear DNA content of strawberries (*Fragaria* spp.) compared with *Arabidopsis thaliana* by using dual-step flow cytometry. *Cytologia* 2001; **66**: 431–436.
- Bennett LE, Burkhead JL, Hale KL, Terry N, Pilon M, Pilon-Smits EA. Analysis of transgenic Indian mustard plants for phytoremediation of metal-contaminated mine tailings. *J Environ Qual* 2003; **32**: 432–440.
- Debener T, Janakiram T, Mattiesch L. Sports and seedlings of rose varieties analysed with molecular markers. *Plant Breeding* 2000; **119**: 71–74.
- Congiu L, Chicca M, Cella R, Rossi R, Bernacchia G. The use of random amplified polymorphic DNA (RAPD) markers to identify strawberry varieties: a forensic application. *Mol Ecol* 2000; **9**: 229–232.
- Parent JG, Fortin MG, Page D. Identification of raspberry cultivars by random amplified polymorphic DNA (RAPD) analysis. *Can J Plant Sci* 1993; **73**: 1115–1122.

- 31 Koopman WJ, Wissemann V, de Cock K et al. AFLP markers as a tool to reconstruct complex relationships: A case study in *Rosa* (Rosaceae). *Am J Bot* 2008; **95**: 353–366.
- 32 Degani C, Rowland LJ, Saunders JA et al. A comparison of genetic relationship measures in strawberry (*Fragaria × ananassa* Duch.) based on AFLPs, RAPDs, and pedigree data. *Euphytica* 2001; **117**: 1–12.
- 33 Graham J, McNicol RJ. An examination of the ability of RAPD markers to determine the relationships within and between *Rubus* species. *Theor Appl Genet* 1995; **90**: 1128–1132.
- 34 Debnath SC. Inter-simple sequence repeat (ISSR)-PCR analysis to assess genetic diversity in a collection of wild cloudberry (*Rubus chamaemorus* L.) clones. *J Hort Sci Biotech* 2007; **82**: 727–732.
- 35 Davis TM, Yu H. A linkage map of the diploid strawberry, *Fragaria vesca*. *J Hered* 1997; **88**: 215–221.
- 36 Debener T, Mattiesch L. Construction of a genetic linkage map for roses using RAPD and AFLP markers. *Theor Appl Genet* 1999; **99**: 891–899.
- 37 Graham J, Smith K, MacKenzie K, Jorgenson L, Hackett C, Powell W. The construction of a genetic linkage map of red raspberry (*Rubus idaeus* subsp *idaeus*) based on AFLPs, genomic-SSR and EST-SSR markers. *Theor Appl Genet* 2004; **109**: 740–749.
- 38 Eriksson O, Bremer B. Genet dynamics of the clonal plant *Rubus-saxatilis*. *J Ecol* 1993; **81**: 533–542.
- 39 Harrison RE, Luby JJ, Furnier GR. Chloroplast DNA restriction fragment variation among strawberry (*Fragaria* spp) taxa. *J Am Soc Hort Sci* 1997; **122**: 63–68.
- 40 Graham J, Smith K, Woodhead M, Russell J. Development and use of simple sequence repeat SSR markers in *Rubus* species. *Mol Ecol Notes* 2002; **2**: 250–252.
- 41 Sargent DJ, Hadonou AM, Simpson DW. Development and characterization of polymorphic microsatellite markers from *Fragaria viridis*, a wild diploid strawberry. *Mol Ecol Notes* 2003; **3**: 550–552.
- 42 Kimura T, Nishitani C, Iketani H, Ban Y, Yamamoto T. Development of microsatellite markers in rose. *Mol Ecol Notes* 2006; **6**: 810–812.
- 43 Lewers KS, Styan SM, Hokanson SC, Bassil NV. Strawberry GenBank-derived and genomic simple sequence repeat (SSR) markers and their utility with strawberry, blackberry, and red and black raspberry. *J Am Soc Hort Sci* 2005; **130**: 102–115.
- 44 Park YH, Ahn SG, Choi YM et al. Rose (*Rosa hybrida* L.) EST-derived microsatellite markers and their transferability to strawberry (*Fragaria* spp.). *Sci Hort* 2010; **125**: 733–739.
- 45 Woodhead M, McCallum S, Smith K, Cardle L, Mazzitelli L, Graham J. Identification, characterisation and mapping of simple sequence repeat (SSR) markers from raspberry root and bud ESTs. *Mol Breed* 2008; **22**: 555–563.
- 46 Shulaev V, Sargent DJ, Crowhurst RN et al. The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet* 2011; **43**: 109–116.
- 47 Sargent DJ, Passey T, Surbanovski N et al. A microsatellite linkage map for the cultivated strawberry (*Fragaria × ananassa*) suggests extensive regions of homozygosity in the genome that may have resulted from breeding and selection. *Theor Appl Genet* 2012; **124**: 1229–1240.
- 48 Isobe SN, Hirakawa H, Sato S et al. Construction of an integrated high density simple sequence repeat linkage map in cultivated strawberry (*Fragaria × ananassa*) and its applicability. *DNA Res* 2013; **20**: 79–92.
- 49 Dobes C, Scheffknecht S. Isolation and characterization of microsatellite loci for the *Potentilla* core group (Rosaceae) using 454 sequencing. *Mol Ecol Resour* 2012; **12**: 726–739.
- 50 Govan CL, Simpson DW, Johnson AW, Tobutt KR, Sargent DJ. A reliable multiplexed microsatellite set for genotyping *Fragaria* and its use in a survey of 60 *F. × ananassa* cultivars. *Mol Breeding* 2008; **22**: 649–661.
- 51 Brunings AM, Moyer C, Peres N, Folta KM. Implementation of simple sequence repeat markers to genotype Florida strawberry varieties. *Euphytica* 2010; **173**: 63–75.
- 52 Chambers A, Carle S, Njuguna W et al. A genome-enabled, high-throughput, and multiplexed fingerprinting platform for strawberry (*Fragaria* L.). *Mol Breeding* 2013; **31**: 615–629.
- 53 Esselink GD, Smulders MJ, Vosman B. Identification of cut rose (*Rosa hybrida*) and rootstock varieties using robust sequence tagged microsatellite site markers. *Theor Appl Genet* 2003; **106**: 277–286.
- 54 Scariot V, Akkac A, Botta R. Characterization and genetic relationships of wild species and old garden roses based on microsatellite analysis. *J Am Soc Hort Sci* 2006; **131**: 66–73.
- 55 Dossett M, Bassil NV, Lewers KS, Finn CE. Genetic diversity in wild and cultivated black raspberry (*Rubus occidentalis* L.) evaluated by simple sequence repeat markers. *Genet Resour Crop Ev* 2012; **59**: 1849–1865.
- 56 Fernandez-Fernandez F, Antanaviciute L, Govan CL, Sargent DJ. Development of a multiplexed microsatellite set for fingerprinting red raspberry (*Rubus idaeus*) germplasm and its transferability to other *Rubus* species. *J Berry Res* 2011; **1**: 177–187.
- 57 Bassil NV, Nyberg A, Hummer KE, Graham J, Dossett M, Finn CE. A universal fingerprinting set for red raspberry. *Acta Hort* 2012; **946**: 83–87.
- 58 Dirlwanger E, Cosson P, Tavaud M et al. Development of microsatellite markers in peach *Prunus persica* (L.) Batsch and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theor Appl Genet* 2002; **105**: 127–138.
- 59 Gasic K, Han Y, Kertbundit S et al. Characteristics and transferability of new apple EST-derived SSRs to other Rosaceae species. *Mol Breed* 2009; **23**: 397–411.
- 60 Davis TM, DiMeglio LM, Yang R, Styan SM, Lewers KS. Assessment of SSR marker transfer from the cultivated strawberry to diploid strawberry species: functionality, linkage group assignment, and use in diversity analysis. *J Am Soc Hort Sci* 2006; **131**: 506–512.
- 61 Zorrilla-Fontanesi Y, Cabeza A, Torres AM et al. Development and bin mapping of strawberry genic-SSRs in diploid *Fragaria* and their transferability across the Rosoideae subfamily. *Mol Breed* 2011; **27**: 137–156.
- 62 Zorrilla-Fontanesi Y, Cabeza A, Dominguez P et al. Quantitative trait loci and underlying candidate genes controlling agronomical and fruit quality traits in octoploid strawberry (*Fragaria × ananassa*). *Theor Appl Genet* 2011; **123**: 755–778.
- 63 Marulanda M, Lopez AM, Uribe M. Molecular characterization of the Andean blackberry, *Rubus glaucus*, using SSR markers. *Genet Mol Res* 2012; **11**: 322–331.
- 64 Rousseau-Gueutin M, Richard L, Le Dantec L, Caron H, Denoyes-Rothan B. Development, mapping and transferability of *Fragaria* EST-SSRs within the Rosodae supertribe. *Plant Breed* 2011; **130**: 248–255.
- 65 Koning-Boucoiran CF, Gitonga VW, Yan Z et al. The mode of inheritance in tetraploid cut roses. *Theor Appl Genet* 2012; **125**: 591–607.
- 66 Vilanova S, Sargent DJ, Arus P, Monfort A. Synteny conservation between two distantly-related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry). *BMC Plant Biol* 2008; **8**: 67.
- 67 Bushakra JM, Stephens MJ, Atmadjaja AN et al. Construction of black (*Rubus occidentalis*) and red (*R. idaeus*) raspberry linkage maps and their comparison to the genomes of strawberry, apple, and peach. *Theor Appl Genet* 2012; **125**: 311–327.
- 68 Hubbard M, Kelly J, Rajapakse S, Abbott A, Ballard R. Restriction-fragment-length-polymorphisms in rose and their use for cultivar identification. *Hortscience* 1992; **27**: 172–173.
- 69 Fulton TM, van der Hoeven R, Eannetta NT, Tanksley SD. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell* 2002; **14**: 1457–1467.
- 70 Cabrera A, Kozik A, Howad W, Arus P, Iezzoni AF, van der Knaap E. Development and bin mapping of a Rosaceae Conserved Ortholog Set (COS) of markers. *BMC Genomics* 2009; **10**: 562.
- 71 Yan Z, Denneboom C, Hattendorf A et al. Construction of an integrated map of rose with AFLP, SSR, PK, RGA, RFLP, SCAR and morphological markers. *Theor Appl Genet* 2005; **110**: 766–777.
- 72 Folta KM, Staton M, Stewart PJ et al. Expressed sequence tags (ESTs) and simple sequence repeat (SSR) markers from octoploid strawberry (*Fragaria × ananassa*). *BMC Plant Biol* 2005; **5**: 12.
- 73 Ge AJ, Han J, Li XD et al. Characterization of SNPs in strawberry cultivars in China. *Genet Mol Res* 2013; **12**: 639–645.
- 74 Zhang J, Esselink GD, Che D, Fougere-Danezan M, Arens P, Smulders MJ. The diploid origins of allopolyploid rose species studied using single nucleotide polymorphism haplotypes flanking a microsatellite repeat. *J Hort Sci Biotechnol* 2013; **88**: 85–92.
- 75 Ila E, Sargent DJ, Girona EL et al. Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. *BMC Evol Biol* 2011; **11**: 9.
- 76 Ruiz-Rojas JJ, Sargent DJ, Shulaev V et al. SNP discovery and genetic mapping of T-DNA insertional mutants in *Fragaria vesca* L. *Theor Appl Genet* 2010; **121**: 449–463.
- 77 Kunihsa M, Fukino N, Matsumoto S. Development of cleavage amplified polymorphic sequence (CAPS) markers for identification of strawberry cultivars. *Euphytica* 2003; **134**: 209–215.
- 78 Celton JM, Christoffels A, Sargent DJ, Xu X, Rees DJ, Xu XM. Genome-wide SNP identification by high-throughput sequencing and selective mapping allows sequence assembly positioning using a framework genetic linkage map. *BMC Biol* 2010; **8**: 155.
- 79 Ward JA, Bhangoo J, Fernandez-Fernandez F et al. Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. *BMC Genomics* 2013; **14**: 2.
- 80 Koning-Boucoiran CF, Smulders MJ, Krens FA, Esselink GD, Maliepaard C. SNP genotyping in tetraploid cut roses. *Acta Hort* 2012; **953**: 351–356.
- 81 Elshire RJ, Glaubitz JC, Sun Q et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos ONE* 2011; **6**: e19379.
- 82 Cronn R, Knaus BJ, Liston A et al. Targeted enrichment strategies for next-generation plant biology. *Am J Bot* 2012; **99**: 291–311.



- 83 Tennessen JA, Govindarajulu R, Liston A, Ashman TL. Targeted sequence capture provides insight into genome structure and genetics of male sterility in a gynodioecious diploid strawberry, *Fragaria vesca* subsp. *bracteata* (Rosaceae). *G3* 2013; **3**: 1341–1351.
- 84 Bassil NH. Development and preliminary evaluation of the IStraw90 Axiom® array in cultivated strawberry (*Fragaria × ananassa*). In: Proceedings of American Society for Horticultural Science Annual Conference; 22–25 July; Palm Desert, CA, USA. ASHS: Alexandria, VA, USA, 2013; 17.
- 85 Antanaviciute L, Fernandez-Fernandez F, Jansen J et al. Development of a dense SNP-based linkage map of an apple rootstock progeny using the *Malus* Infinium whole genome genotyping array. *BMC Genom* 2012; **13**: 203.
- 86 Verde I, Bassil N, Scalabrin S et al. Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *Plos ONE* 2012; **7**: e35668.
- 87 Peace C, Bassil N, Main D et al. Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. *Plos ONE* 2012; **7**: e48305.
- 88 Deng C, Davis TM. Molecular identification of the yellow fruit color (c) locus in diploid strawberry: a candidate gene approach. *Theor Appl Genet* 2001; **103**: 316–322.
- 89 Sargent DJ, Clarke J, Simpson DW et al. An enhanced microsatellite map of diploid *Fragaria*. *Theor Appl Genet* 2006; **112**: 1349–1359.
- 90 Sargent DJ, Rys A, Nier S, Simpson DW, Tobutt KR. The development and mapping of functional markers in *Fragaria* and their transferability and potential for mapping in other genera. *Theor Appl Genet* 2007; **114**: 373–384.
- 91 Sargent DJ, Cipriani G, Vilanova S et al. The development of a bin mapping population and the selective mapping of 103 markers in the diploid *Fragaria* reference map. *Genome* 2008; **51**: 120–127.
- 92 Illa E, Sargent DJ, Girona EL et al. Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. *BMC Evol Biol* 2011; **11**: 9.
- 93 Sargent DJ, Kuchta P, Girona EL et al. Simple sequence repeat marker development and mapping targeted to previously unmapped regions of the strawberry genome sequence. *Plant Genome* 2011; **4**: 165–177.
- 94 Darrow GM. *The strawberry: history, breeding and physiology*. New York: Holt, Rinehart and Winston, 1966.
- 95 Bringham RS. Cytogenetics and evolution in american *Fragaria*. *Hortscience* 1990; **25**: 879–881.
- 96 Lerceteau-Köhler E, Guerin G, Laigret F, Denoyes-Rothan B. Characterization of mixed disomic and polysomic inheritance in the octoploid strawberry (*Fragaria × ananassa*) using AFLP mapping. *Theor Appl Genet* 2003; **107**: 619–628.
- 97 Spigler RB, Lewers KS, Main DS, Ashman TL. Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. *Heredity* 2008; **101**: 507–517.
- 98 van Dijk T, Noordijk Y, Dubos T et al. Microsatellite allele dose and configuration establishment (MADCE): an integrated approach for genetic studies in allopolyploids. *BMC Plant Biol* 2012; **12**: 25.
- 99 Rousseau-Gueutin M, Lerceteau-Köhler E, Barrot L et al. Comparative genetic mapping between octoploid and diploid *Fragaria* species reveals a high level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry. *Genetics* 2008; **179**: 2045–2060.
- 100 Spigler RB, Lewers KS, Johnson AL, Ashman TL. Comparative mapping reveals autosomal origin of sex chromosome in octoploid *Fragaria virginiana*. *J Hered* 2010; **101**: S107–S117.
- 101 Sargent DJ, Fernandez-Fernandez F, Ruiz-Roja JJ et al. A genetic linkage map of the cultivated strawberry (*Fragaria × ananassa*) and its comparison to the diploid *Fragaria* reference map. *Mol Breed* 2009; **24**: 293–303.
- 102 Zorrilla-Fontanesi Y, Rambla JL, Cabeza A et al. Genetic analysis of strawberry fruit aroma and identification of O-methyltransferase FaOMT as the locus controlling natural variation in mesiferane content. *Plant Physiol* 2012; **159**: 851–870.
- 103 Haymes KM, Henken B, Davis TM, van de Weg WE. Identification of RAPD markers linked to a *Phytophthora fragariae* resistance gene (*Rpf1*) in the cultivated strawberry. *Theor Appl Genet* 1997; **94**: 1097–1101.
- 104 Haymes KM, van de Weg WE, Arens P, Maas JL, Vosman B, den Nijs AP. Development of SCAR markers linked to a *Phytophthora fragariae* resistance gene and their assessment in European and North American strawberry genotypes. *J Am Soc Hort Sci* 2000; **125**: 330–339.
- 105 Whitaker VM. Applications of molecular markers in strawberry. *J Berry Res* 2011; **1**: 115–127.
- 106 Lerceteau-Köhler E, Moing A, Guerin G et al. QTL analysis for fruit quality traits and resistance to *Colletotrichum acutatum* and *Phytophthora cactorum* in octoploid strawberry (*Fragaria × ananassa*). *Acta Hort* 2004; **649**: 93–97.
- 107 Lerceteau-Köhler E, Guerin G, Denoyes-Rothan B. Identification of SCAR markers linked to *Rca2* anthracnose resistance gene and their assessment in strawberry germplasm. *Theor Appl Genet* 2005; **111**: 862–870.
- 108 Sugimoto T, Tamaki K, Matsumoto J, Yamamoto Y, Shiwaku K, Watanabe K. Detection of RAPD markers linked to the everbearing gene in Japanese cultivated strawberry. *Plant Breed* 2005; **124**: 498–501.
- 109 Weebadde CK, Wang D, Finn CE et al. Using a linkage mapping approach to identify QTL for day-neutrality in the octoploid strawberry. *Plant Breed* 2008; **127**: 94–101.
- 110 Gaston A, Perrotte J, Lerceteau-Köhler E et al. PFRU, a single dominant locus regulates the balance between sexual and asexual plant reproduction in cultivated strawberry. *J Exp Bot* 2013; **64**: 1837–1848.
- 111 Lerceteau-Köhler E, Moing A, Guerin G et al. Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-QTL in their control. *Theor Appl Genet* 2012; **124**: 1059–1077.
- 112 Graham J, Smith K, Tierney I, MacKenzie K, Hackett CA. Mapping gene *H* controlling cane pubescence in raspberry and its association with resistance to cane botrytis and spur blight, rust and cane spot. *Theor Appl Genet* 2006; **112**: 818–831.
- 113 Jennings DL. Balanced lethals and polymorphism in *Rubus idaeus*. *Heredity* 1967; **22**: 465–479.
- 114 Jennings DL. Resistance to *Didymella-applanata* in red raspberry and some related species. *Ann Appl Biol* 1982; **101**: 331–333.
- 115 Jennings DL, Brydon E. Further-studies on resistance to *Leptosphaeria-coniothyrium* in the red raspberry and related species. *Ann Appl Biol* 1989; **115**: 499–506.
- 116 Jennings DL. Some evidence on the influence of the morphology of raspberry canes upon their ability to be attacked by certain fungi. *Hort Res* 1962; **1**: 100–111.
- 117 Jennings DL, McGregor GR. Resistance to cane spot (*Elsinoe-veneta*) in the red raspberry and its relationship to resistance to yellow rust (*Phragmidium-rubridaei*). *Euphytica* 1988; **37**: 173–180.
- 118 Anthony VM, Williamson B, Jennings DL, Shattock RC. Inheritance of resistance to yellow rust (*Phragmidium-rubridaei*) in red raspberry. *Ann Appl Biol* 1986; **109**: 365–374.
- 119 Graham J, Hackett CA, Smith K, Woodhead M, Hein I, McCallum S. Mapping QTLs for developmental traits in raspberry from bud break to ripe fruit. *Theor Appl Genet* 2009; **118**: 1143–1155.
- 120 McCallum S, Woodhead M, Hackett CA, Kassim A, Paterson A, Graham J. Genetic and environmental effects influencing fruit colour and QTL analysis in raspberry. *Theor Appl Genet* 2010; **121**: 611–627.
- 121 Woodhead M, Williamson S, Smith K et al. Identification of quantitative trait loci for cane splitting in red raspberry (*Rubus idaeus*). *Mol Breed* 2013; **31**: 111–122.
- 122 Woodhead M, Weir A, Smith K, McCallum S, MacKenzie K, Graham J. Functional markers for red raspberry. *J Am Soc Hort Sci* 2010; **135**: 418–427.
- 123 Keep E. Dwarfing in the raspberry, *Rubus idaeus* L. *Euphytica* 1968; **18**: 256–276.
- 124 Pattison JA, Samuelian SK, Weber CA. Inheritance of *Phytophthora* root rot resistance in red raspberry determined by generation means and molecular linkage analysis. *Theor Appl Genet* 2007; **115**: 225–236.
- 125 Bushakra JM, Krieger C, Deng D et al. QTL involved in the modification of cyanidin compounds in black and red raspberry fruit. *Theor Appl Genet* 2013; **126**: 847–865.
- 126 Kassim A, Poette J, Paterson A et al. Environmental and seasonal influences on red raspberry anthocyanin antioxidant contents and identification of quantitative traits loci (QTL). *Mol Nutr Food Res* 2009; **53**: 625–634.
- 127 Castro P, Stafne ET, Clark JR, Lewers KS. Genetic map of the primocane-fruiting and thornless traits of tetraploid blackberry. *Theor Appl Genet* 2013; **126**: 2521–2532.
- 128 Crespel L, Zhang D, Meynet J, Gudin S. Estimation of heterozygosity in two botanic rose species using AFLP markers. *Acta Hort* 2001; **546**: 187–191.
- 129 Hibrand-Saint Oyant L, Crespel L, Rajapakse S, Zhang L, Foucher F. Genetic linkage maps of rose constructed with new microsatellite markers and locating QTL controlling flowering traits. *Tree Genet Genomes* 2008; **4**: 11–23.
- 130 Linde M, Hattendorf A, Kaufmann H, Debener T. Powdery mildew resistance in roses: QTL mapping in different environments using selective genotyping. *Theor Appl Genet* 2006; **113**: 1081–1092.
- 131 Shupert DA, Byrne DH, Pemberton HB. Inheritance of flower traits, leaflet number and prickles in roses. *Acta Hort* 2007; **751**: 331–335.
- 132 Spiller M, Linde M, Hibrand-Saint Oyant L et al. Towards a unified genetic map for diploid roses. *Theor Appl Genet* 2011; **122**: 489–500.
- 133 Ma Y, Crane CF, Byrne DH. Meiotic behavior in a tetraploid rose and its hybrid progeny. *Hortscience* 2000; **35**: 1127–1131.
- 134 Gar O, Sargent DJ, Tsai CJ et al. An autotetraploid linkage map of rose (*Rosa hybrida*) validated using the strawberry (*Fragaria vesca*) genome sequence. *Plos ONE* 2011; **6**: e20463.

- 135 Bonet R, Girona EL, Sargent DJ et al. The development and characterisation of a bacterial artificial chromosome library for *Fragaria vesca*. *BMC Res Notes* 2009; **2**: 188.
- 136 Pontaroli AC, Rogers RL, Zhang Q et al. Gene content and distribution in the nuclear genome of *Fragaria vesca*. *Plant Genome* 2009; **2**: 93–101.
- 137 Kaufmann H, Mattiesch L, Lorz H, Debener T. Construction of a BAC library of *Rosa rugosa* Thunb. and assembly of a contig spanning *Rdr1*, a gene that confers resistance to blackspot. *Mol Genet Genom* 2003; **268**: 666–674.
- 138 Hein I, Williamson S, Russell J, Powell W. Isolation of high molecular weight DNA suitable for BAC library construction from woody perennial soft-fruit species. *Biotechniques* 2005; **38**: 69–71.
- 139 Graham J, Hackett CA, Smith K et al. Towards an understanding of the nature of resistance to *Phytophthora* root rot in red raspberry. *Theor Appl Genet* 2011; **123**: 585–601.
- 140 Biber A, Kaufmann H, Linde M, Spiller M, Terefe D, Debener T. Molecular markers from a BAC contig spanning the *Rdr1* locus: a tool for marker-assisted selection in roses. *Theor Appl Genet* 2010; **120**: 765–773.
- 141 Terefe-Ayana D, Kaufmann H, Linde M, Debener T. Evolution of the *Rdr1* TNL-cluster in roses and other Rosaceous species. *BMC Genomics* 2012; **13**: 409.
- 142 Potter D, Luby JJ, Harrison RE. Phylogenetic relationships among species of *Fragaria* (Rosaceae) inferred from non-coding nuclear and chloroplast DNA sequences. *Syst Bot* 2000; **25**: 337–348.
- 143 Salamone I, Govindarajulu R, Falk S, Parks M, Liston A, Ashman TL. Bioclimatic, ecological, and phenotypic intermediacy and high genetic admixture in a natural hybrid of octoploid strawberries. *Am J Bot* 2013; **100**: 939–950.
- 144 Ferrarini M, Moretto M, Ward JA et al. An evaluation of the PacBio RS platform for sequencing and *de novo* assembly of a chloroplast genome. *BMC Genom* 2013; **14**: 670.
- 145 Hamilton JP, Buell CR. Advances in plant genome sequencing. *Plant J* 2012; **70**: 177–190.
- 146 Jung S, Cestaro A, Troggio M et al. Whole genome comparisons of *Fragaria*, *Prunus* and *Malus* reveal different modes of evolution between Rosaceous subfamilies. *BMC Genom* 2012; **13**: 129.
- 147 Koskela EA, Mouhu K, Albani MC et al. Mutation in TERMINAL FLOWER1 reverses the photoperiodic requirement for flowering in the wild strawberry *Fragaria vesca*. *Plant Physiol* 2012; **159**: 1043–1054.
- 148 Ward JA, Price J, Clement MJ et al. Draft assembly and analysis of the highly heterozygous diploid red raspberry genome. In: Proceedings of Plant and Animal Genome XX Conference; 14–18 January 2012; San Diego, CA, USA. Sherago International, NJ, USA, 2012; 315.
- 149 Bassil N, Gilmore B, Dossett M et al. Developing genomics resources in black raspberry. In: Rosaceous Genomics Conference 6; 30 September–4 October 2012; Mezzocorona (Trento), Italy. Fondazione Edmund Mach: San Michele all'Adige, Italy, 2012; 48.
- 150 Geoff W, Michael K, Dmitry P, Tim B, John L. Production of long (1.5kb–15.0kb), accurate, DNA sequencing reads using an Illumina HiSeq2000 to support *de novo* assembly of the blue catfish genome. In: Proceedings of Plant and Animal Genome XXI Conference; 12–16 January 2013; San Diego, CA, USA. Sherago International, NJ, USA, 2013; 487.
- 151 Iwata H, Gaston A, Remay A et al. The TFL1 homologue KSN is a regulator of continuous flowering in rose and strawberry. *Plant J* 2012; **69**: 116–125.
- 152 Albani MC, Battey NH, Wilkinson MJ. The development of ISSR-derived SCAR markers around the SEASONAL FLOWERING LOCUS (SFL) in *Fragaria vesca*. *Theor Appl Genet* 2004; **109**: 571–579.
- 153 Nier S, Simpson DW, Tobutt KR, Sargent DJ. A genetic linkage map of an inter-specific diploid *Fragaria* BC1 mapping population and its comparison with the *Fragaria* reference map (FV×FN). *J Hort Sci Biotechnol* 2006; **81**: 645–650.
- 154 Brown T, Wareing PF. The genetical control of the everbearing habit and three other characters in varieties of *Fragaria vesca*. *Euphytica* 1965; **14**: 97–112.
- 155 Remay A, Lalanne D, Thouroude T, Le Couviour F, Hibrand-Saint Oyant L, Foucher F. A survey of flowering genes reveals the role of gibberellins in floral control in rose. *Theor Appl Genet* 2009; **119**: 767–781.
- 156 Kawamura K, Oyant LH, Crespel L, Thouroude T, Lalanne D, Foucher F. Quantitative trait loci for flowering time and inflorescence architecture in rose. *Theor Appl Genet* 2011; **122**: 661–675.
- 157 Folta KM, Clancy MA, Chamala S et al. A transcript accounting from diverse tissues of a cultivated strawberry. *Plant Genome* 2010; **3**: 90–105.
- 158 Ward JA, Ponnala L, Weber CA. Strategies for transcriptome analysis in nonmodel plants. *Am J Bot* 2012; **99**: 267–276.
- 159 Dubois A, Carrere S, Raymond O et al. Transcriptome database resource and gene expression atlas for the rose. *BMC Genom* 2012; **13**: 11.
- 160 Kim J, Park JH, Lim CJ et al. Small RNA and transcriptome deep sequencing proffers insight into floral gene regulation in *Rosa* cultivars. *BMC Genom* 2012; **13**: 18.
- 161 Kang CY, Darwish O, Geretz A, Shahan R, Alkharouf N, Liu ZC. Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry *Fragaria vesca*. *Plant Cell* 2013; **25**: 1960–1978.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>