Partner 5

IRTA- Departament de Genètica Vegetal - Carretera de Cabrils s/n 08348 Cabrils(Barcelona), Spain and Departament de Genètica Molecular CID-CSIC, Barcelona
Pere ARÚS, Responsible Scientist
Scientific team: Enrique Graziano, Francesc Garriga-Calderé and Pere Puigdomènech

Subtitle: Comparative mapping in *Prunus* and in other dicotyledoneous species using *Arabidopsis* as a reference.

Task 1: Identification of genes from EST databases for use as anchor loci and determination of ESTs from mapped DNA

a) *Arabidopsis* EST mapping in *Prunus*

Materials and methods

The F2 population (N=80) obtained from the cross between 'Texas' almond and 'Earlygold' peach (abbreviated TxE) was used as mapping population. A map with 246 markers (235 RFLPs and 11 isozymes) had been previously constructed with this population.

Each *Arabidopsis* probe was first tested for polymorphism using DNA from 'Texas' and 'Earlygold' digested with five restriction enzymes: *BamH*I, *Dra*I, *EcoR*I. *Hind*III and *Mva*I. Probes that showed good resolution and polymorphism were studied for segregation in all individuals of the population. Methods for DNA digestion, Southern blot and radioactive probe hybridization were as described in Viruel et al. (1994).

Data were analyzed with two software programs, LINKEM (Vowden et al., 1995) for the analysis of goodness-of-fit to the one-locus expected segregation ratios and MAPMAKER for the map construction. Linkage groups were established with a LOD≥7.0 and we used the Kosambi mapping function for the conversion of recombination fractions into genetic distances.

Results and Discussion

A total of 198 *Arabidopsis* probes were tested for polymorphism in the parents of the TxE offspring. Only 77 of them were polymorphic in our population and gave a sufficiently good resolution. These probes produced 111 loci that were all of them placed in the map. A general characteristic of these probes was that they detected multiple loci more often than the rest of the probes used (most of them coming from Rosaceae genomic or cDNA libraries). Only 2 of them (3%) detected a single locus, in contrast with the 56% of probes detecting only one locus with Rosaceae probes. Seventeen additional markers were mapped using 4 pea probes obtained from Partner 3 (Dr. N. Ellis' group) (out of 32 assayed), 4 potato probes obtained from Partner 2 (Dr. C. Gebhardt's group) (out of 6 assayed) and 9 probes corresponding to known almond genes.

Locus distribution was similar to that found with other probes and spread throughout the eight linkage groups of the TxE map without any specific pattern of distribution. The complete map contained 374 markers for a total distance of 595 cM, which represents an average density of 1.59 cM per marker. This map can be found in Figure 1 of this report.

Using the map position of the RFLPs obtained with the 57 probes that detected two or more loci it was possible to have a first insight into the degree of duplication of the *Prunus* genome. The contingency table with the number of correspondences of duplicated loci produced by the same probe for each pair of linkage groups was significantly different from independence (χ^2 = 42.25; P<0.05). Two pairs of loci had clearly a higher number of common duplicate loci than expected. These were linkage group 2 and linkage group 6 (abbreviated PrG2 and PrG6), and PrG5-PrG7. The map comparison between the former two groups was consistent with the existence of a duplicated region between them (Fig. 2), but the maps of PrG5 and PrG7 gave a less clear pattern of similarity. These results would be in agreement with the previous data on Rosaceae and *Arabidopsis* probes detecting more than one locus in *Prunus*.

b) Sequencing of Rosaceae probes mapped in *Prunus* and 'in silico' mapping in *Arabidopsis*

Materials and methods

A total of 197 DNA probes corresponding to 218 RFLP loci from the T x E map (Joobeur et al, 1998), and 31 probes specific of the F x T map (Viruel et al, 1995) were sequenced using dideoxy terminators an run on a capillary automated DNA sequencer. We used the "ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit" with AmpliTaq DNA polymerase, FS (PE Applied Biosystems) and run the sequences in an ABI Prism 310 Genetic Analyzer (Perkin Elmer). The primers used were the universal M13 forward primer (5'-GTAAAACGACGGCCAGT-3') and the universal M13 reverse primer (5'-CAGGAAACAGCTATGAC-3') depending on the clones. After the sequencing, the electropherograms were studied and the resulting sequences edited where needed using the ABI Prism sequencing analysis program v. 3.0. All the sequences that were too short, had big compressions, or were considered of low quality were removed from this study, leaving a total of 161 independent sequences. The flanking plasmid stretches were removed manually using the same program.

Before starting the similarity searches, we analysed the sequences for mislabelling and other human errors with the GCG software package (REF GCG). We used the same package to assemble consensus sequences whenever we had overlapping runs. The similarity searches were performed using the BLAST family of algorithms, mainly through the WWW interface located at the School of the Biological Sciences of the University of Cambridge (http://www.bio.cam.ac.uk/cgi-bin/blast2/blastallsrs.pl). We used BLASTX to compare our sequences translated in the six possible reading frames against the protein databases (SwissProt plus translated GenBank, TrEMBL and TrEMBLNew) using the BLOSUM62 substitution matrix, and the SEG complexity filter. For direct DNA searches we used BLASTN against the whole set of nucleic acid databases (including ESTs, genome survey sequences, and recent additions), using the DUST complexity filter. We used TBLASTX to search our sequence translated in the

six reading frames against the whole nucleotide database dynamically translated to protein in the six reading frames. We performed all three BLAST searches with all the sequences. We assumed that a search was successful when the resulting E-value for a given alignment was lower than 10^{-8}, though we looked carefully at each alignment that had an E-value close to the threshold to try to avoid missing significant matches and wrong positives.

Only the sequences that were selected under the previous search were used to find similarities with *Arabidopsis* sequences using the database of the *Arabidopsis* Genome Initiative (www.arabidopsis.org). Criteria for accepting a match were: i) all *Arabidopsis* sequences with a TBLASTX E-value below 10^{-25} or, ii) for Rosaceae probes that did not have any match under these conditions, the best hit if its E-value was below 10^{-15}.

Results and Discussion

We sequenced 161 probes (68 cDNAs and 93 genomic DNAs) at least partially, yielding an average length of 404 base pairs. We found similarities at $E<10^{-8}$ in 98 sequences (61%). From these we found 77 that met the criteria established for considering them homologous to the *Arabidopsis* sequences and therefore be used as anchor points for map comparisons. To these sequences we added 18 more corresponding to *Prunus* probes already sequenced from a total of 25 included in the map. Homologies with *Arabidopsis* were also searched for 14 mapped isozyme genes obtained with nine enzyme systems (aspartate aminotransferase, phosphoglucomutase, phosphoglucoisomerase, shikimate dehydrogenese, leucine aminopeptidase, 6 phosphogluconate dehydrogenese, aconitase, malic enzyme and isocitrate dehydrogenase).

The total number of loci in the *Prunus* map corresponding to all these probes and isozymes was of 127, and the number of anchor loci found with them in *Arabidopsis* was of 347, that represents that we found 2.7 *Arabidopsis* putative anchor points for each *Prunus* locus. Considering these 127 *Prunus* anchor loci and the 111 obtained in the section a) of this work, we had a total of 238 markers in *Prunus* for the comparison with the *Arabidopsis* genome that would have 476 putative homologous loci in its map. In the case of *Prunus* that means that we have an average marker density of 2.5 cM/marker, that can be considered acceptable to detect any major DNA fragments conserved between both species.

c) Comparative mapping between *Prunus* and *Arabidopsis*

Materials and Methods

The 238 markers of the *Prunus* map and their corresponding 476 *Arabidopsis* putative homologous found by 'in silico' mapping as explained in sections a) and b) were aligned and the maps obtained were compared. We performed the map alignments using the FITMAPS program, developed in our laboratory (see Deliverables section).

The number of correspondences between each pair of linkage groups were assayed for independence using a χ^2 test with the contingency table.

A detailed analysis searching for possible small regions of synteny between *Prunus* and *Arabidopsis* was done for all the 40 possible pairwise linkage group combinations. A region of the map was accepted as possibly conserved if it met in both maps the following criteria: i) ≥ 4 markers; ii) ≤ 4.0 cM/marker average density; and iii) maximum distance between any two markers ≤ 8cM.

Results and Discussion

Three elements contribute to the complexity of the comparison between *Prunus* and *Arabidopsis*, and in general to the comparison among non-confamilial species. First, the fact recently discovered that the *Arabidopsis* genome is partly duplicated (Blanc et al, 2000; T.A.G.I, 2000). This means that two or more homologous sequences in *Arabidopsis* instead of one only may be found when comparing its genome with a strict diploid. The situation may become more difficult if the species compared is also a segmental or full polyploid, as it may occur in *Prunus* as suggested by the results described in section a) of this report. Second, we are comparing RFLP linkage data in *Prunus* with sequence data in *Arabidopsis*. The RFLP information is incomplete, as it corresponds to only a part of the loci detected by each probe (at least 63% of the *Arabidopsis* probes used for RFLP mapping in *Prunus* detected one or more loci that were not segregating). And third, in the sequences conserved there is, at least in *Arabidopsis*, only a partial conservation of the genes, some of them being selectively lost (Blanc et al., 2000; Ku et al., 2000). This may be problematic for the identification of small syntenic regions using non-high density linkage maps or DNA sequence data, as it occurs in this study.

The number of markers common to each *Prunus-Arabidopsis* chromosome pair was not different from random (χ^2 of the contingency table of 29.64; n.s.). This indicates that it has not been possible to detect large DNA fragments (including entire chromosomes or arms of chromosomes) conserved between *Prunus* and *Arabidopsi*s.

The analysis of the correspondences between each pair of linkage groups of both species allowed us to identify thirteen short regions of putative synteny between both genomes. Their main characteristics are described in Table 1 and the maps showing the syntenic fragments and the markers involved are shown in Figure 3. These regions had average values of 6 markers per region, density of 2.1 cM/marker, longest gap of 5.4 cM, and mean distance covered of 12 cM. In total they represent a 19,8% of the *Prunus* and a 16,9% of the *Arabidopsis* map distance. This results suggest that detectable conserved regions still exist between these two species.

At least one syntenic region was identified in each of the linkage groups of *Prunus* and each chromosome of *Arabidopsis* (Fig 3). Some of them detected overlapping regions in the *Prunus* genome (PrG3, PrG5, and PrG8), or in the *Arabidopsis* genome (AtC2, AtC4 and AtC5), indicating possible duplications in *Arabidopsis* and *Prunus*, respectively. Four of the six conserved regions of AtC5 were overlapping in the same region of this chromosome. The corresponding regions in the *Prunus* genome of two of them were the fragments of PrG2 and PrG6 that we detected as possible duplications (see section a)).

d) Mapping and map comparisons with probes from other species

d.1) Comparison pea - *Prunus*

Fabaceae and Rosaceae are relatively close to each other when compared with the rest of the species included in the project. Information on the position of RFLPs obtained with common *Arabidopsis* probes and interchange of *Prunus* and pea probes was done between Partner 3 and our group with the objective of trying to compare our respective maps.

We received 36 pea probes from which 32 were assayed for polymorphism and only four of them found polymorphic and mapped. The *Arabidopsis* probes that produced RFLPs in both species were 23, detecting 29 loci. The 33 resulting loci were compared with the map of pea, but the pattern of similarity was unclear, due in part to the small number of markers used.

d.2) Comparison potato-*Prunus*

Potato probes (6) from a region of potato chromosome III that was conserved between potato and *Arabidopsis* detected by Partner 2 were also studied in *Prunus*. The hypothesis was that this region may have been conserved across different dicotyledoneous families. In this case we would expect to find linkage also between different RFLPs detected with these probes. Four of them were found polymorphic and allowed us to determine their position in the *Prunus* map. Each of the RFLPs detected were placed in a different linkage group: P3g6 in PrG2, P3c9 in PrG1, P1f3 in PrG6 and P1f11in PrG5, suggesting that this region is not conserved in *Prunus*.

References

Blanc et al. 2000. Plant Cell 12:1093-1101
Joobeur et al. 1998. Theor Appl Genet. 97:1034-1041
Ku et al., 2000. Proc. Natl. Acad. Sci. USA. 97(16):9121-9126
Paterson et al., 1996. Nat. Genet. 14: 380-382
The Arabidoppsis Genome Initiative. 2000. Nature
Viruel et al. 1995. Theor. Appl. Genet. 91:964-971

Dissemination

Graziano, E., F. Garriga-Calderé and P. Arús. 2001. Map comparison between Prunus and Arabidopsis. Plant and Animal Genome Conference. January 2001, San Diego. Poster and oral presentation to the workshop 'Fruit and Nut crops'.

Graziano, E. and P. Arús. Gene identification and sequence analysis of mapped RFLPs in Prunus. (in preparation).

Deliverables

FITMAPS: a program for alignment between maps. Briefly, FITMAPS generates a text file with a report of all the matches found between the two maps and a postscript graphic page, suitable for direct printing in postscript printers or retouching with standard image processing programs. Its input consists in a configuration file plus two plain text files containing the name and position of each marker, and a number of optional formatting parameters, including facing, underlining and color.

Figure 1. Linkage map of the *Prunus* TxE population. with 374 markers. RFLPs obtained using *Arabidopsis* EST probes are marked in blue ink. Isozyme genes are marked in red. RFLPs obtained with Rosaceae, pea or potato probes are marked with black ink.

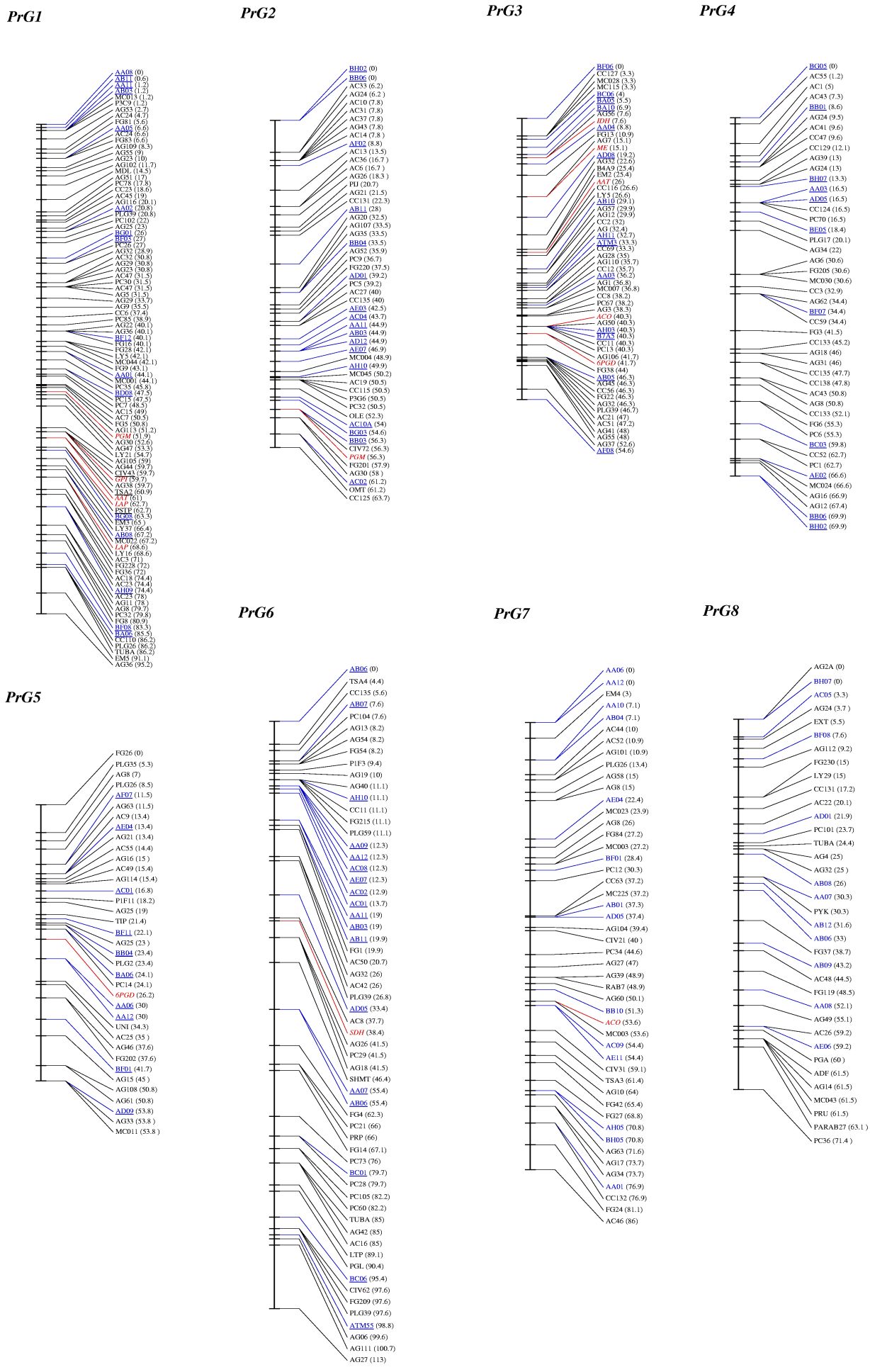


Figure 2. Correspondences between the map position of different loci obtained with the same probe in linkage groups 2 and 6 of *Prunus*

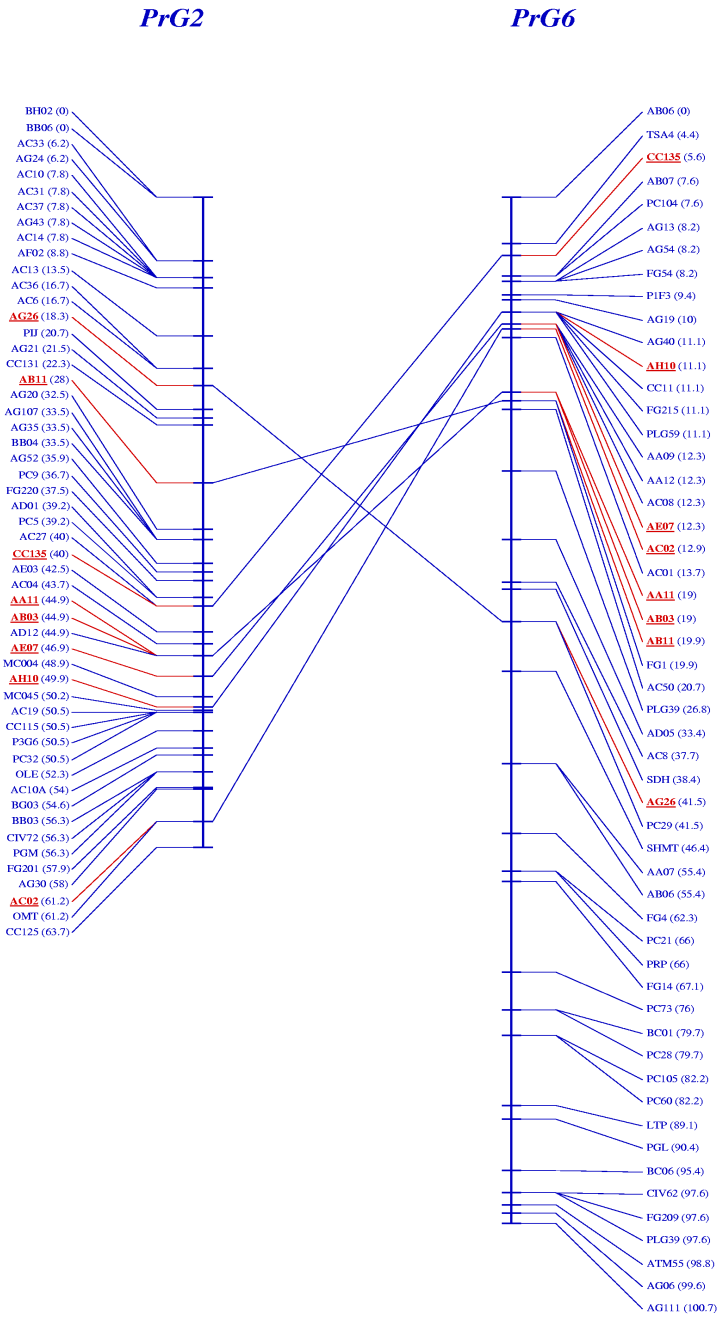


Figure 3. Syntenic regions between *Prunus* and Arabidopsis. PrG- are *Prunus* linkage groups and AtC- are *Arabidopsis* chromosomes. Anchor points of both maps ar marked with red ink and underlined.

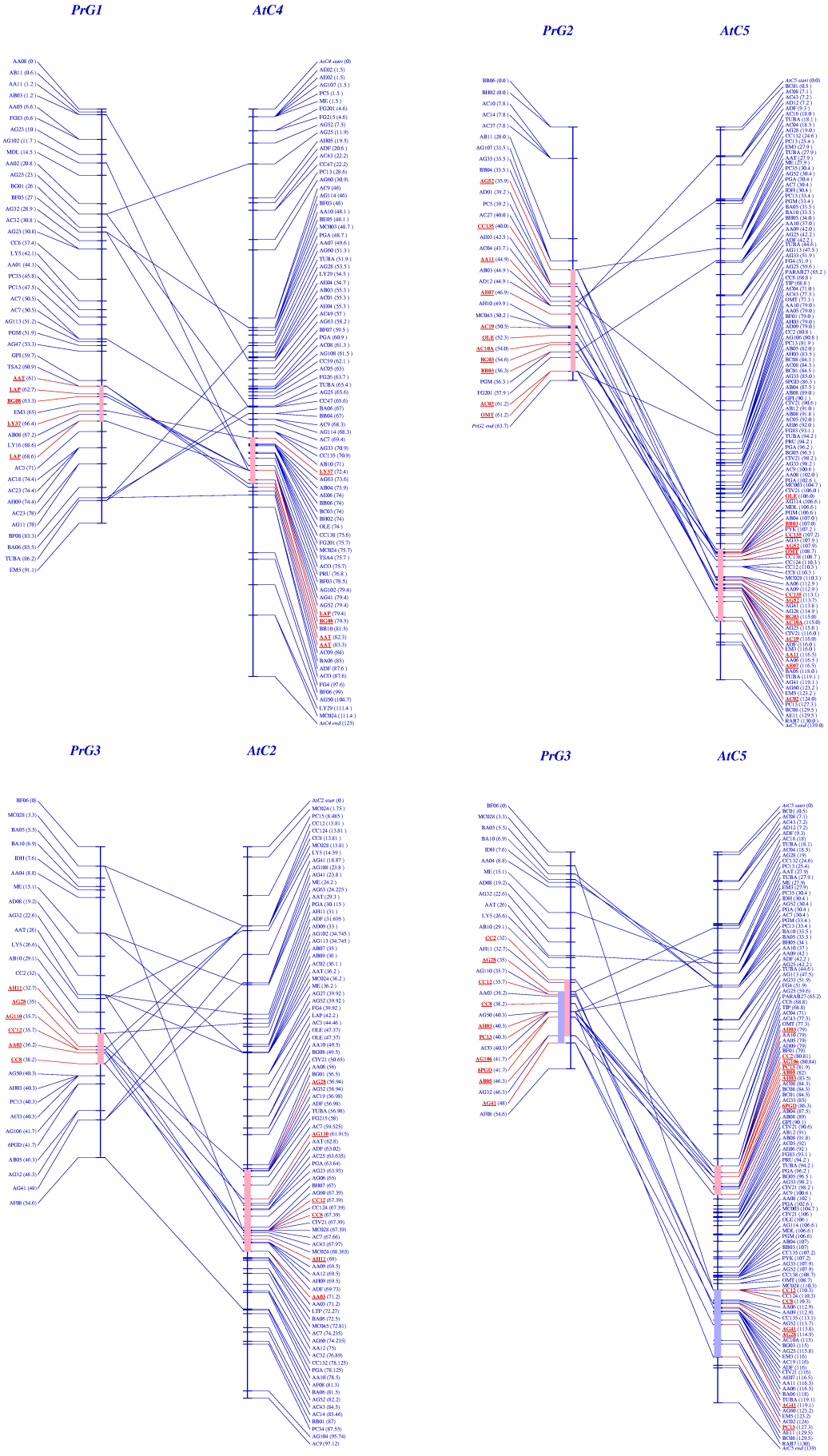


Figure 3. (continued)

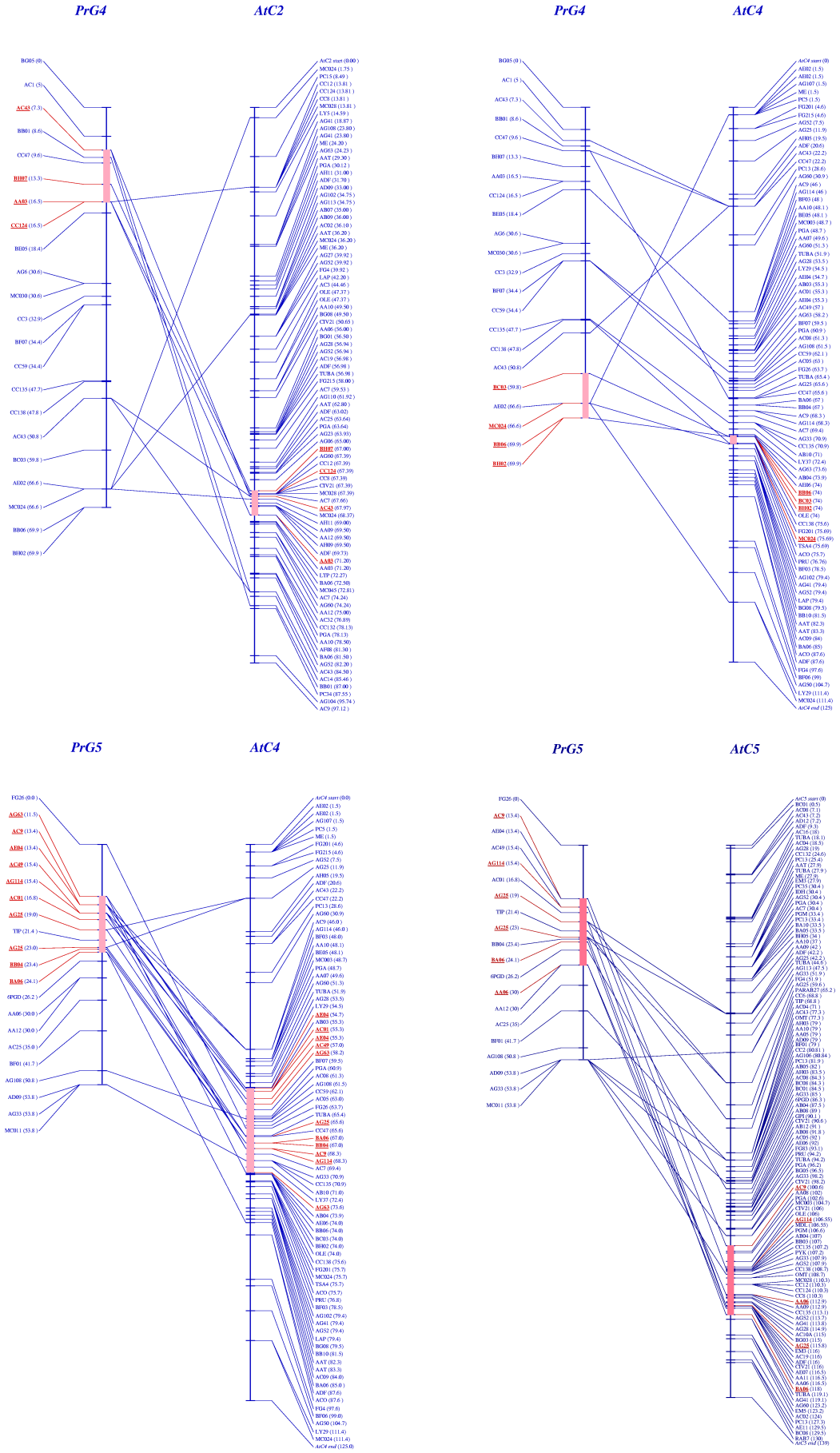
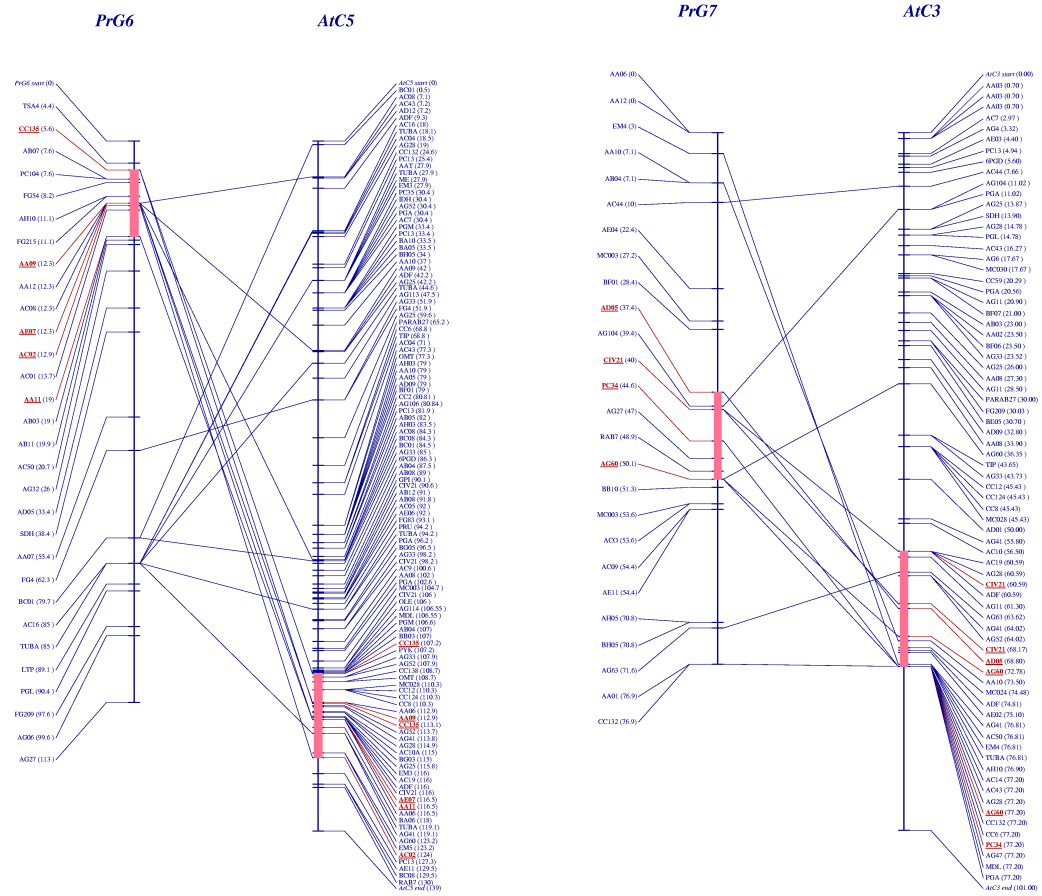


Figure 3. (continued).



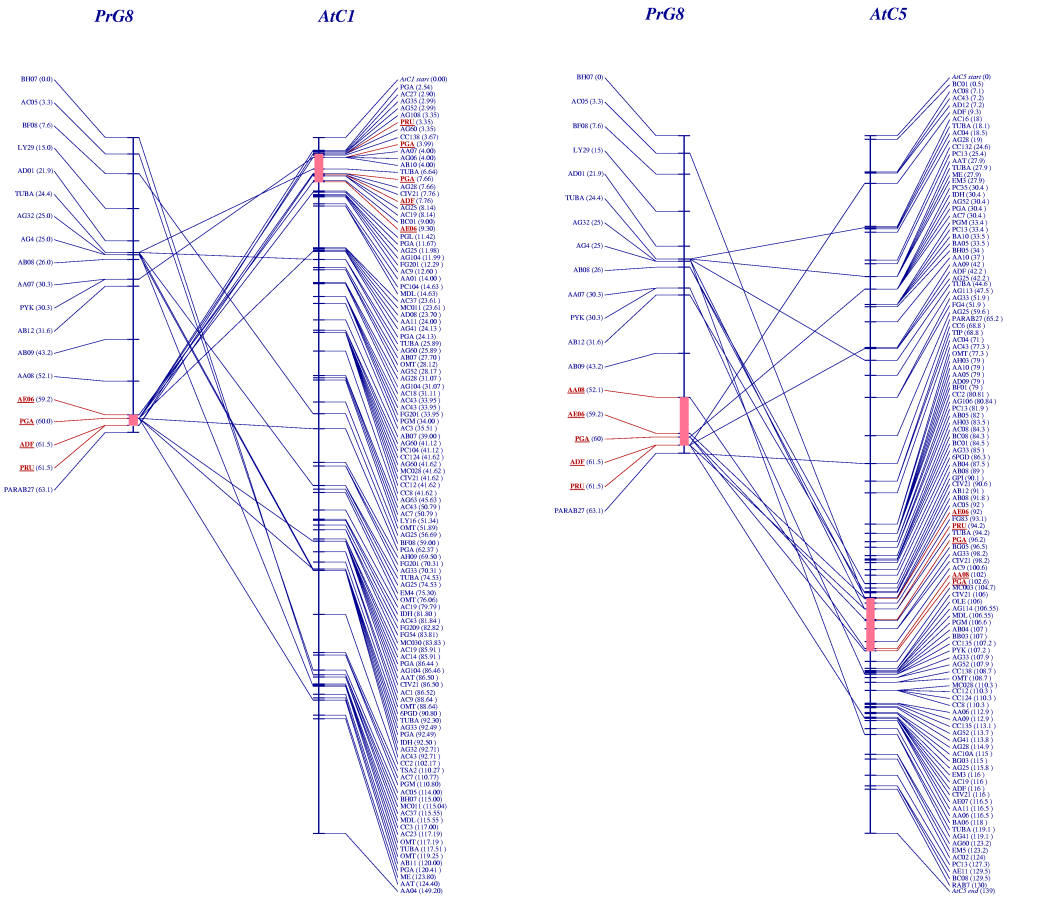


Table 1. Characteristics of the 13 conserved regions detected between *Prunus* and *Arabidopsis*

	Region1		**Region2**		**Region3**		**Region4**		**Region 5**		**Region6**	
Linkage groups	**PrG1**	**AtC4**	**PrG2**	**AtC5**	**PrG3**	**AtG2**	**PrG3 (a)**	**AtC5(a)**	**PrG3 (b)**	**AtC5(b)**	**PrG4**	**AtC2**
# loci	5	5	11	11	6	6	5	6	5	6	4	4
Distance (cM)	9	11	26	18	5	15	14	5	13	17	10	4
avg. Density	1,8	2,2	2,36	1,63	0,8	2,5	2,8	0,83	2,6	2,8	2,5	1
largest gap (cM)	3	7	5	7	2	5	8	1	8	8	7	3

	Region7		**Region8**		**Region9**		**Region10**		**Region11**		**Region12**		**Region13**	
Linkage groups	**PrG4**	**AtC4**	**PrG5**	**AtC4**	**PrG5**	**AtC5**	**PrG6**	**AtC5**	**PrG7**	**AtC3**	**PrG8**	**AtC1**	**PrG8**	**AtC5**
# loci	4	4	10	12	6	5	5	6	4	6	4	5	4	5
Distance (cM)	10	2	12	19	17	17	13	17	14	16	4	6	10	11
avg. Density	2,5	0,5	1,2	1,58	2,83	3,4	2,6	2,83	3,5	2,7	1	1,2	2,5	2,2
largest gap (cM)	7	2	4	5	6	6	6	7	8	7	1	4	7	6