Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: http://oatao.univ-toulouse.fr/
Eprints ID: 4528

To link to this article: DOI:10.1007/s11295-010-0272-3
http://dx.doi.org/10.1007/s11295-010-0272-3

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@inp-toulouse.fr

High resolution synteny maps allowing direct comparisons between the coffee and tomato genomes

Florent Lefebvre-Pautigny · Feinan Wu · Murielle Philippot · Michel Rigoreau ·
Priyono · Mohamed Zouine · Pierre Frasse · Mondher Bouzayen · Pierre Broun ·
Vincent Pétiard · Steven D. Tanksley · Dominique Crouzillat

Abstract Tomato (*Solanum lycopersicum*) and coffee (*Coffea canephora*) belong to the sister families Solanaceae and Rubiaceae, respectively. We report herein the mapping of a common set of 257 Conserved Ortholog Set II genes in the genomes of both species. The mapped markers are well distributed across both genomes allowing the first syntenic comparison between species from these two families. The majority (75%) of the synteny blocks are short (<4 cM); however, some extend up to 50 cM. In an effort to further characterize the synteny between these two genomes, we took advantage of the available sequence for the tomato genome to show that tomato chromosome 7 is syntenic to half of the two coffee linkage groups E and F with the putative break point in tomato localized to the boundary of the heterochromatin and euchromatin on the long arm. In addition to the new insight on genome conservation and evolution between the plant families Solanaceae and Rubiaceae, the comparative maps presented herein provide a translational tool by which coffee researchers may take benefit of DNA sequence and genetic information from tomato and vice versa. It is thus expected that these comparative genome information will help to facilitate and expedite genetic and genomic research in coffee.

F. Lefebvre-Pautigny · M. Rigoreau · P. Broun · V. Pétiard ·
D. Crouzillat (✉)
Nestle R&D Centre,
101 Avenue Gustave Eiffel, Notre Dame d'Oé, B.P. 49716, 37097
Tours Cedex 2, France
e-mail: dominique.crouzillat@rdto.nestle.com

F. Wu · S. D. Tanksley
Department of Plant Breeding and Genetics, Cornell University,
Ithaca, NY 14853, USA

M. Philippot · M. Zouine · P. Frasse · M. Bouzayen
Université de Toulouse, INP-ENSA Toulouse,
Génomique et Biotechnologie des Fruits,
Avenue de l'Agrobiopole, BP 32607, Castanet-Tolosan F-31326,
France

M. Philippot · M. Zouine · P. Frasse · M. Bouzayen
INRA, Génomique et Biotechnologie des Fruits,
Chemin de Borde Rouge,
Castanet-Tolosan F-31326, France

Priyono
Indonesian Coffee and Cocoa Research Institute,
Jl. PB. Sudirman 90,
Jember 68118, Indonesia

Introduction

The genus *Coffea* (family Rubiaceae) encompasses approximately 100 species all of which are native to the African continent and Madagascar and the Mascarene Islands (Davis et al. 2006). Two of these species *Coffea canephora* and *Coffea arabica*, are widely cultivated for the production of the coffee beverage. The former is diploid and allogamous, the latter, allotetraploid and preferentially autogamous. Approximately 70% of the world coffee

production is coming from *C. arabica* versus 30% for *C. canephora*. But the main source of raw material for soluble coffee is *C. canephora*.

While it is not widely known, coffee is one of the most valuable international exchanges for cultural-traded commodities. This is reflected in the fact that the stock exchange for raw coffee is the forth in value only to wheat, sugar, and soya in the international market (Pendergrast 2009). Additionally, more than 25 million people worldwide are linked to coffee culture and processing. Despite these economic aspects, coffee research suffers of a lack of investments both in terms of science and financial resources. Moreover, coffee is a perennial plant with a time from seed to seed of about 5 years, which makes genetic studies more difficult and time consuming. While some genomic information is publicly available for coffee (e.g., an expressed sequence tags (EST) database— Lin et al. 2005; Poncet et al. 2006), it lags far behind what is available for many other agricultural species. As a result, coffee researchers have only limited access to the plethora of genomic resources available for most major crop species.

Comparative genomics provides the opportunity to leverage genetic/genomic information from one species to another via comparative genetic maps (e.g., Tanksley et al. 1992; Livingstone et al. 1999; Mysore et al. 2001; Dirlewanger et al. 2004; Srinivasachary et al. 2007; Wu et al. 2009a, b). For example, this comparative approach has been used to leverage genomic information from the model species *Arabidopsis thaliana* to related crop species (Van Dodeweerd et al. 1999; Grant et al. 2000; Ku et al. 2000; Bevan and Walsh 2005). However, the value of comparative genomic information is inversely proportional to the evolutionary distance of the species being compared. Hence, genomic information from rice has been of great benefit to other crops species in the grass family, but has been used minimally for species from other families (Srinivasachary et al. 2007; Wilson et al. 1999). There have been attempts to use genomic information from *Arabidopsis* to coffee for synteny study (Mahé et al. 2007). However, coffee being only distantly related to *Arabidopsis*, the genome comparison approach is difficult to perform and is therefore of limited utility. Phylogenetically, the model species most closely related to coffee, for which significant genetics and genomic resources exist, is tomato. Both coffee and tomato belong to the Euasterid I clade of flowering plants and are likely derived from an ancestral species with a haploid number of ×=11 or 12 (Wu et al. 2006).

The tomato genome has already been linked to the genomes of other Solanaceous crops (e.g., potato, eggplant, pepper) via comparative genetic maps (Tanksley et al. 1992; Livingstone et al. 1999). Moreover, comparative

bacterial artificial chromosome (BAC) sequencing has shown a high degree of microsynteny conservation among these species (Wang et al. 2008). This high level of conservation can be attributed, at least in part, to the absence of polyploidization in the ancestral genome that gave rise to the family Solanaceae (Wang et al. 2008). Furthermore, recent studies suggest that the Solanaceae and Rubiaceae (which contains coffee) families descended from a diploid ancestor without any recent whole genome duplication (Wu et al. 2006). Hence, one might expect a reasonable level of both macro and micro (Guyot et al. 2009) synteny between the tomato and coffee genomes. If comparative genetic maps could be generated for coffee and tomato, it might allow coffee researchers to access to the large set of genetic and genomic data currently available for tomato—including the genome sequence of tomato which is currently being generated by a consortium of international scientists (http://www.sgn.cornell.edu/; Mueller et al. 2005a, b, 2009). The availability of a comparative map may open the way for coffee researchers to use the tomato genetic map and genome sequence for predicting the chromosomal positions of the large number of coffee ESTs currently available (Lin et al. 2005, http://sgn.cornell.edu/search/library_search.pl?w5c4_library_parameters=canephora).

Coffee and tomato diverged from their last common ancestor since approximately 86 MY, a long period of evolutionary time allowing for many chromosomal rearrangement to accumulate (Wikstrom et al. 2001). Therefore, a detailed understanding of the comparative relationships between the coffee and tomato genomes will likely have to await the sequencing of both genomes. However, the generation of comparative genetic maps for tomato and coffee, as described in this paper, may still allow a cross use of genetic information between the two species. The extent of the predictive value of these comparative maps will be determined in part by the extent to which the two genomes were rearranged subsequent to divergence from their last common ancestor.

Recently, a set of polymerase chain reaction (PCR)-based ortholog markers (Wu et al. 2006), have been used to compare genomes of solanaceous species (tomato, eggplant (Wu et al. 2009a), pepper (Wu et al. 2009b), and tobacco (Wu et al. 2009c)), and a related taxa such as Rubiaceae species: coffee. These markers are referred to Conserved Orthologous Set (COSII) markers. The primary objective of the current study aims at constructing a comparative map of the coffee and tomato genomes based on these COSII markers. This represents the first detailed comparative maps between species belonging to different families and might significantly enhance the potential for genetic/genomic research in coffee.

Materials and methods

Tomato mapping population

The tomato map is based on 80 F2 individuals from the cross *Solanum lycopersicum* LA925 x *Solanum pennellii* LA716 (Wu et al. 2009a, b). The complete tomato map, including all COSII markers is available at http://sgn.cornell.edu/cview/map.pl?map_version_id=52

Coffee mapping population

Genetic diversity of *C. canephora* is subdivided in two main genetic groups, Guinean and Congolese groups coming from two different geographic origins (Dussert et al. 1999; Gomez et al. 2009). A consensus reference genetic map of *C. canephora* has been developed in Nestle Tours Centre in collaboration with the Indonesian ICCRI institute (Crouzillat et al. unpublished data). This map derives from a cross between two highly heterozygous genotypes, a Congolese group genotype (BP409) and a Congolese-Guinean hybrid parent (Q121). This same population was used to map 396 COSII loci. The segregating population is composed of 93 F1 individuals.

Identification of coffee COSII orthologs and primer design for mapping

It was first necessary to identify the coffee ortholog corresponding to each tomato COSII gene. This was accomplished by using the publicly available coffee EST database (Lin et al. 2005; Poncet et al. 2006) and employing the search/verification methods described in Wu et al. (2006). Coffee unigene sequences were downloaded for each COSII marker from the Coffee Genomic Network Website (http://coffee.pgn.cornell.edu/index.pl) or SOL website (http://www.sgn.cornell.edu/). The corresponding *Arabidopsis* cDNA and genomic sequence were downloaded for the same COSII from the TAIR website (http://www.arabidopsis.org/). Based on these three-way alignments (coffee unigene: *Arabidopsis* cDNA: *Arabidopsis* genomic sequence) using Clustalw: (http://bioinfo.hku.hk/services/analyseq/cgi-bin/clustalw_in.pl), it was possible to identify the predicted positions of introns for each of the coffee COSII genes. Coffee primers were then designed in exons flanking one or more introns. Amplified introns were expected to carry more polymorphic sites than exons and thus be more useful in the subsequent genetic mapping experiments. Two different software programs were used for primer design: Primer Express and Primer Select. Primer design conditions were as follows: predicted amplicon length ≤600 nt, primer length 17-24 nt, Tm≥55°C, hairpins (between and within primers) ≤4. PCR amplifications were performed at 55°C using total DNA from BP409 and Q121. Of the PCR products, 10 µl were loaded on 3% agarose gel in order to identify PCR product lengths for DNA sequencing. PCR products were sequenced both strands for the two parents of the mapping population. For identity control, sequences of the PCR products, derived from genomic DNA, were compared with the coffee EST-based unigene sequences. In order to identify polymorphisms, BP409 and Q121 sequences were compared using SeqScape software or SeqDoc website (http://research.imb.uq.edu.au/seqdoc/). With this approach, it was possible to identify simple sequence repeats (SSR), single-nucleotide polymorphism (SNP), and insertion/deletion polymorphisms in coffee COSII sequences useful for genetic mapping.

Coffee genetic mapping

COSII markers were mapped into the coffee mapping population (BP409 × Q121) using three types of polymorphism:

Restriction fragment length polymorphism (RFLP) Several COSII EST amplicons, derived from *C. canephora* (Wu et al. 2006), were used as source of DNA for labeled probes using ^{32}P. The screening and mapping of the polymorphic COSII probes was performed using eight restriction enzymes (*Dra*I, *Eco*RV, *Hin*dIII, *Hae*III, *Rsa*I, *Sca*I, *Hin*cII, and *Pvu*II).

SSR Primers pairs were designed in order to obtain a range of PCR amplicon length from 100 to 300 bp. Forward primers were tagged with a fluorescent dyes FAM^{TM}, VIC$^{®}$, NED^{TM} or PET^{TM}. Reagent volumes are 2.5 µl of genomic DNA, 10.5 µl of AmpliTaq Gold Master Mix (Applied Biosystems^{TM}), 1 µl of each primer (initial concentration of 20 µM), and 10.5 µl of water. PCR conditions are 94°C for 10 min, 30 cycles 94°C/55°C/72°C with respective durations of 30/30/60 s and finally 72°C during 7 min before cooling to 4°C. Mix for electrophoresis was 2 µl of PCR product, 20 µl of formamide and 0.5 µl of Genscan 500Liz size standard (Applied Biosystems^{TM}) and alleles detection was obtained using an ABiPrism 3100 analyzer (Applied Biosystems^{TM}). Alleles detection and identification were then computed with "GeneScan Analysis Software", and segregation determined using "Genotyper" software.

SNP SNPs indentified in sequenced amplicons from the two mapping parents were mapped using the MGB$^{®}$ (Minor Group Binder) TaqMan$^{®}$ technology with hot start DNA polymerase—one primer pair and two TaqMan$^{®}$ probes were designed by Applied Biosystems. The probes were tagged with FAM^{TM} or VIC$^{®}$ dyes fluorescence to

discriminate the two segregating alleles from the heterozygous parent(s). Reagent volumes for one genotyping assay were 2.5 μl of genomic DNA, 0.6 μl of 40X MGB® probes Master Mix, 9.4 μl of water, and 12.5 μl of TaqMan® 2X Universal PCR Master Mix (Applied Biosystems™). PCR reactions were performed with an ABiPrism 7500 thermocycler (Applied Biosystems™) under the following conditions: hot start DNA polymerase activation at 95°C during 10 min, followed by amplification step 40 cycles of 95°C for 15 s, 60°C for 1 min. End point reading for genotype assessment was made at the last step. Progenies genotyping was performed using allelic discrimination assay. Analysis of results has been done with "7500 System Software".

Additional SNPs were also detected using high resolution melting technology. After detection of SNP in one or both parents of the population, primers pairs were designed in order to have an amplified fragment length between 150 and 500 bp. Reagent components for one genotyping assay are, 2 μl of genomic DNA, 10 μl of Master Mix (2× containing fluorescent dye), 1 μl of primer mix (final concentration of 0.2 μM for each primers), 2.4 μl of MgCl2, 3.6 μl of water. Then PCR reactions were done with Lightcycler® 480 thermocycler (Roche Life Sciences). The PCR conditions were the following, activation of DNA polymerase and DNA denaturation by one cycle at 95°C during 10 min. Next step is the amplification of target DNA by 45 cycles of 95°C for 10 s, 60°C for 15 s, and 72°C for 25 s. The last step is the high resolution melting of PCR products, realized in three times, first is denaturation of PCR products at 95°C during 1 min, second is renaturation at 40°C during 1 min, and finally melting in an interval from 60°C to 95°C with a ramp rate of 0.02°C/s. During this last step, fluorescence is read 25 times per degree. Analysis of melting curves and segregation of alleles were performed using "Genescanning" software.

The linkage analysis and map calculations were performed using JoinMap® software version 4 (Stam 1993; Van Ooijen 2006). Population type was cross-pollinated type, as population under study resulted from a cross of two heterozygous diploid parents with no information available about linkage phases. The genetic maps of the two parents BP409 and Q121 were constructed separately since COSII markers could be mapped on one (backcross) or two (F_2) of these parental genetic maps due to allogamous status of *C. canephora*. Linkage groups were defined at logarithm of odds (LOD) of three or more. Kosambi's function was used to calculate genetic distances between two loci. The coffee COSII genetic map was then built using F_2 segregating loci as anchor markers in order to merge homologous parental linkage groups. Order of loci on consensus genetic map was compared to that on both parental genetic maps. If the order of markers on the consensus genetic map was clearly

Fig. 1 Coffee COSII synteny map. Shared syntenic blocs with tomato ▶ are deduced from the COSII loci shared with tomato. Each tomato chromosome is assigned a different color. Then the COSII loci on the coffee linkage groups and the syntenic blocs are colored according to the corresponding tomato chromosome (see Wu et al. 2009a for correspondence). Areas expected to be syntenic are extended 10 cM beyond last orthologous marker since syntenous tracts often do not extend beyond such distances (see text for discussion). Loci showing distorted segregation are indicated with *$P<0.01$ and **$P<0.001$

different from the order on the parental maps, the order of the parental linkage group was taken as a fixed reference order. Genetic maps were drawn using MapChart software version 2.1 (Voorrips 2002).

Tomato chromosome 7 genomic sequence

The sequence of tomato chromosome 7 is currently being generated as part of the international initiative aimed at sequencing the entire tomato genome (M. Bouzayen personal communication, Mueller et al. 2009). In the current study, we utilized 15.25 Mb of non-redundant sequences generated for chromosome 7. Given that the sequencing effort is targeted to the gene-rich region of the genome, most of the sequence corresponds to the euchromatin part of tomato chromosome 7. Based on the chromosomal location of the sequenced BACs determined by BAC-fluorescence in situ hybridization (FISH) mapping, it has been estimated that 8.96 Mb of the sequences correspond to euchromatin, whereas 6.29 Mb to heterochromatin (M. Bouzayen personal communication). Among the 189 BACs and Fosmids sequenced on chromosome 7, 165 are distributed into 39 contigs; the remaining are singletons. Among the 189 chromosome 7 BACs and Fosmids used in this study, the sequences corresponding to 160 are available at Genbank.

In order to study the synteny between tomato chromosome 7 and coffee linkage groups E and F, for each COSII marker mapped on each of these two linkage groups, the corresponding tomato unigene (available at ftp://ftp.sgn. cornell.edu/COSII/tomato_mapping/) or the *Arabidopsis* ortholog (when no tomato unigene was identified for the marker) was blasted against the tomato chromosome 7 BAC sequences (Altschul et al. 1997). The marker was assigned to tomato chromosome 7 if the sequence identity exceeded 99% between the tomato unigene and the exonic part of the BAC sequences, or if the *Arabidopsis* ortholog had a single match on the BAC sequences. On the other hand, we also screened all available (up to 2,800) COSII sequences by a blast search on chromosome 7 to identify new COS markers on chromosome 7. The COSII sequences selected by this mean were then mapped on coffee genome.

Genetic linkage map (linkage groups A–J). Scale in cM from 0 to 225 (left axis).

A
- C2_At1g55480
- C2_At4g34350(A)
- C3_At1g06058
- C2_At4g21580
- C2_At5g04590
- C2_At2g32600(A)
- C2_At3g44600
- C2_At5g20890
- C2_At2g27450
- C2_At5g22750
- C2_At5g22580
- C2_At5g22620*
- C2_At3g44880
- C2_At2g28370
- C2_At5g60540
- C2_At5g60600
- C2_At2g28490
- C2_At4g29890
- C2_At2g28690
- C2_At5g59960
- C2_At4g37300
- C2_At5g58410
- C2_At4g10360
- C2_At5g58490
- C2_At5g58240
- C2_At3g25120
- C2_At1g03150
- C2_At3g12290
- C2_At4g02680
- C2_At5g43940
- C2_At4g02070
- C2_At1g02475
- C2_At4g01880
- C2_At4g01650
- end

B
- C2_At3g02910
- C2_At5g53000
- C2_At1g65000
- C2_At3g03070
- C2_At3g03100
- C2_At5g17560
- C2_At5g24170
- C2_At4g10810
- C2_At5g67490
- C2_At3g49260
- C2_At5g23880
- C2_At5g23940
- C2_At4g25720
- C2_At5g51700
- C2_At5g62390
- C2_At3g10670(A)
- C2_At3g22660
- C2_At3g56460
- C2_At2g40490
- C2_At3g10920
- C2_At3g55800
- C2_At1g54650
- C2_At3g55120
- C2_At2g36740
- C2_At5g04480
- C2_At5g04420
- C2_At5g04270
- C2_At1g61780
- C2_At2g28380
- C2_At4g21540
- C2_At4g21520
- C2_At2g21445
- C2_At1g29260
- C2_At4g18810
- C2_At2g21290
- C2_At1g49510
- C2_At1g49411
- C2_At5g25480
- C2_At3g19630
- C2_At2g15290
- C2_At2g14910
- C2_At4g33690
- C2_At2g14260
- C2_At3g19720
- C2_At1g10240
- C2_At5g49820
- C2_At2g02500
- C2_At4g14110
- C2_At1g64850
- C2_At3g20230
- C2_At5g26880
- C2_At3g63490
- C2_At5g52820
- C2_At2g38730
- C2_At3g04600
- C2_At2g46910
- C2_At1g54460
- C2_At4g21800
- C2_At1g54520
- C2_At1g56300
- C2_At1g56345
- C2_At1g09340
- C2_At3g05810
- C2_At1g48050
- C2_At3g61140
- C2_At2g38000(A)
- C2_At4g00740
- C2_At1g01730
- C2_At1g29990
- C2_At5g64130
- C2_At2g46370
- C2_At2g45620
- end

C
- top
- C2_At3g17300
- C2_At3g06440
- C2_At4g08300
- C2_At5g18580
- C2_At4g15930
- C2_At3g16810
- C2_At3g06790
- C2_At4g15520
- C2_At3g23590
- C2_At1g04190
- C2_At1g05385
- C2_At3g55800
- C2_At3g55120
- C2_At3g23490
- C2_At3g23400
- C2_At3g16840*
- C2_At5g20070*
- C2_At1g04690
- C2_At2g31970**
- C2_At2g42490
- C2_At4g24690
- C2_At3g06980
- C2_At4g02730**
- C2_At3g28040**
- C2_At3g12300
- C2_At5g64670*
- C2_At1g05350
- end

D
- top
- C2_At5g06360
- C2_At3g56160
- C2_At1g79975
- C2_At4g16710
- C2_At1g80360
- C2_At1g52590
- C2_At1g18430
- C2_At1g09760(A)
- C2_At1g16180
- C2_At5g48930
- C2_At3g07025
- C2_At3g17970
- C2_At3g17940
- C2_At5g23060
- C2_At1g68660
- C2_At5g41040
- C2_At5g63380
- C2_At5g63460
- C2_At5g63440
- C2_At3g48610
- C2_At2g42490
- C2_At4g24690
- C2_At3g06980
- C2_At1g03360
- C2_At2g31170

E
- C2_At1g06690
- C2_At2g44050
- C2_At3g09880*
- C2_At2g32950*
- C2_At4g37130
- C2_At4g15420
- C2_At4g03280
- C2_At4g03210
- C2_At1g03600
- C2_At2g20690
- C2_At1g03820
- C2_At4g28450
- C2_At2g42620
- C2_At2g42750
- C2_At4g11120
- C2_At5g46300
- C2_At1g06550
- C2_At2g38020
- C2_At1g17130
- C2_At1g78600
- C2_At1g53000
- C2_At5g53880
- C2_At3g15290
- C2_At1g52980
- C2_At3g15430
- C2_At3g20660
- C2_At1g55265
- C2_At3g20720
- C2_At5g54770
- C2_At3g21610
- C2_At4g26750
- C2_At1g55670
- C2_At1g55870
- C2_At5g54310
- C2_At5g08420
- end

F
- C2_At3g08710
- C2_At2g38270
- C2_At3g37510
- C2_At3g51840
- C2_At2g37240
- C2_At3g55140
- C2_At2g36930
- C2_At2g37025
- C2_At2g36720
- C2_At5g06130
- C2_At2g35920
- C2_At3g02420
- C2_At1g63780
- C2_At5g34850
- C2_At5g06410
- C2_At5g03790
- C2_At5g20350
- C2_At5g20180
- C2_At2g26830
- C2_At4g20080
- C2_At4g29120
- C2_At2g24530
- C2_At4g29490
- C2_At5g05920
- C2_At5g66940*
- C2_At5g02620*
- C2_At4g43230
- C2_At2g06050
- C2_At1g19130
- C2_At1g11440*
- C2_At5g57655
- C2_At2g24090
- C2_At2g24270
- C2_At5g57870
- C2_At1g55880
- C2_At5g58200
- C2_At3g54360*
- C2_At5g10200
- C2_At3g57810
- C2_At1g61620
- C2_At3g57280
- C2_At2g28250
- C2_At3g11710
- C2_At4g02720
- C2_At2g24390
- C2_At5g11490
- C2_At1g09760(B)
- C2_At5g60990
- C2_At3g11960

G
- top
- C2_At2g23750
- C2_At3g26900
- C2_At5g40950
- C2_At3g02040
- C2_At3g54470
- C2_At5g40500
- C2_At5g14520
- C2_At3g28050
- C2_At3g01480
- C2_At5g40200
- C2_At5g39410
- C2_At3g02220
- C2_At3g02420
- C2_At5g38840
- C2_At4g36970
- C2_At4g37280
- C2_At5g67530
- C2_At5g03390
- C2_At1g77470
- C2_At1g66520
- C2_At5g66530
- C2_At5g66290
- C2_At5g11130
- C2_At5g39790**
- C2_At2g18050
- C2_At4g36530
- C2_At4g35560
- C2_At4g00560
- C2_At5g66090
- C2_At1g65720
- C2_At1g65730
- C2_At3g51520
- C2_At1g78690
- C2_At5g37260
- C2_At1g65900
- C2_At5g37850
- C2_At1g77370
- C2_At1g55880
- C2_At5g10200
- C2_At4g08685
- C2_At5g15510
- C2_At3g57810
- C2_At5g64730
- C2_At1g61620
- C2_At3g57280
- C2_At2g28250
- C2_At3g11710
- C2_At4g02720
- C2_At3g17590
- C2_At5g23530
- C2_At2g18030
- end

H
- C2_At5g47040*
- C2_At4g32280
- C2_At5g12370**
- C2_At4g17790
- C2_At4g19003
- C2_At2g21100
- C2_At1g77470
- C2_At5g47010
- C2_At5g20910
- C2_At5g25940
- C2_At3g06730
- C2_At1g04940
- C2_At3g10670(B)
- C2_At1g62780**
- C2_At1g63770
- C2_At1g64150
- C2_At4g11570**
- C2_At1g63970*
- C2_At1g64355**
- C2_At2g45530*
- C2_At4g24880
- end

I
- top
- C2_At4g16440
- C2_At4g12590
- C2_At3g52155
- C2_At3g02870
- C2_At4g21770
- C2_At1g08550
- C2_At1g65270
- C2_At1g65230
- C2_At5g44650
- C2_At5g26360
- C2_At4g34350(B)
- C2_At4g21120
- C2_At2g32600(B)
- C2_At5g46250
- C2_At1g29950
- C2_At4g35070
- C2_At2g16920
- C2_At2g20940
- C2_At1g16210
- C2_At4g35250

J
- top
- C2_At1g53610
- C2_At1g07755
- C2_At5g45730*
- C2_At5g74670
- C2_At5g42660
- C2_At5g19680
- C2_At3g46940
- C2_At5g41270
- C2_At5g36210
- C2_At1g45976
- C2_At5g25960
- C2_At4g43310
- C2_At1g64480
- C2_At5g07960
- C2_At1g71010
- C2_At1g55730
- C2_At5g09010
- C2_At1g78230
- C2_At1g27600
- C2_At1g22850
- C2_At4g10050
- C2_At1g09920
- C2_At4g10030
- C2_At1g29900
- C2_At1g72030
- C2_At5g16620
- C2_At1g71950(A)
- C2_At1g21780
- C2_At2g28000(B)
- C2_At1g77290
- C2_At4g09830
- end

Table 1 Number and distribution of the COSII markers on the coffee genetic map. The COSII number and the average distance between two loci are indicated for each of the 11 coffee-linkage groups

	A	B	C	D	E	F	G	H	I	J	K	Total
Size (cM)	123	226	116	90	110	161	115	118	83	103	86	1,331
COSII	34	72	23	28	39	51	49	21	21	33	25	396
COSII distance	3.7	3.2	5.1	3.2	2.8	3.4	2.3	5.6	4.0	3.0	3.4	3.6

Results

Coffee COSII mapping

A total of 1,541 coffee COSII genes were identified in silico from a *C. canephora* EST public database (http://www.sgn.cornell.edu/). Among them, 990 were analyzed in order to detect DNA polymorphism on the two parents BP409 and Q121. In total, 396 loci (40%) were subsequently mapped. These were mapped as: SNPs 195 loci (49%), RFLPs 157 loci (40%), and SSRs 44 loci (11%).

Coffee COSII genetic map

The coffee COSII genetic map comprised 11 linkage groups covering 1,331 cM (Fig. 1 and Table 1). As the haploid number of *C. canephora* is $\times=n=11$, we assume that these 11 linkage groups correspond to the 11 coffee chromosomes. All the linkage groups were defined at LOD >3. The highest number of COSII mapped in a group is 72 for linkage group B (Table 1), and the lowest is 21 for the groups H and I. The average spacing between markers is 3.6 cM and the maximum is 32 cM. Only six COSII markers (2%) show a duplicated locus, demonstrating that the majority of the COSII selected in silico do indeed correspond to single copy genes. A set of 25 COSII genes showed significant distortions of segregation ($p \leq 0.01$), seven of them were mapped on linkage group H, six on linkage group C, four on group F, three on group K, two on linkage group E, one on groups A, G, and J. These biased loci are not randomly distributed among linkage groups; they are mainly located on centromeric areas or distal ends.

Comparison of the coffee and tomato COSII maps

From the 396 COSII loci mapped on coffee, 257 (65%) are in common with the tomato map (Table 2). On the basis of these common markers, it was possible to present a synteny map of the coffee genome (Fig. 1). An examination of this map reveals that the average length of syntenic tracts (comprised of at least two adjacent markers) is 12 cM (Fig. 1). In an effort to further examine the extent to which the coffee and tomato maps are conserved, we conducted the following experiment. The coffee genetic map was parsed into pairs of linked orthologous markers. The percentage of cases for which the two adjacent markers are syntenic (correspond to the same tomato chromosome) was recorded and plotted against the map distance (cM) between markers. As the maps distance between markers increases, one would expect a higher chance that chromosomal rearrangement(s) would have led to loss in synteny between the coffee and tomato genome. Further, if there is

Table 2 Distribution of the shared 257 COSII markers on tomato chromosomes and the 11 coffee linkage groups. COSII number shared between each tomato chromosome and coffee-linkage group is indicated. The *red* color shows the highest COSII number shared for each coffee-linkage group

		Coffee linkage groups											
		A	B	C	D	E	F	G	H	I	J	K	
Tomato chromosomes	1	3	17	6			3	1	2	2	1		35
	2	2	3			2	2	20		2		1	32
	3		8	1	11		1	1	1	1	1		25
	4	1				1		3			13	3	21
	5		3	2	1					2		6	14
	6	4	2		1			2		4			13
	7					19	13	2					34
	8		3		1	1	3		10		1		19
	9	5		4	2		6						17
	10	1	2			2	5			2	1	1	14
	11	9	3			1	1	4		1	1		20
	12			1	1	5	4		1		1		13
		25	41	14	17	31	38	33	14	10	23	11	257

Fig. 2 Map distance between adjacent pairs of syntenic markers. Histograms showing the percentage of pair marker intervals (cM) from the coffee maps for which both markers correspond to same tomato chromosome. *Number above bars* indicates total number of paired marker intervals in each category
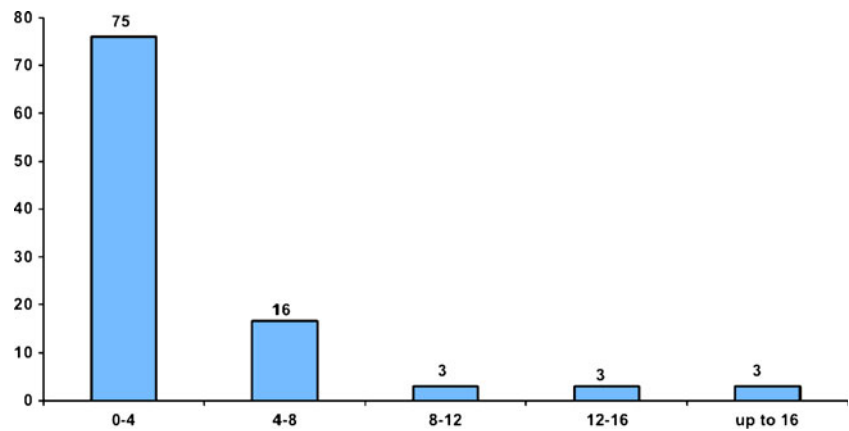
no predictive value, with regards to synteny, between the two markers, one would expect that the two markers would (by chance) still correspond to the same tomato chromosome approximately 8% of the time (since there are 12 chromosomes in the tomato genome, the probability that two markers correspond by chance to the same chromosome is, on average, $1/12 = 0.08$).

The results from this exercise are shown in Fig. 2. For intervals lower than 8 cM the predictive value ranged from 16-75%. However, for larger intervals (8-12 and 12-16 cM) the predictive value dropped to 3% lower than chance (8%). Thus, we propose that the predictive value of orthologous markers between the coffee and tomato genomes should not be generally extended beyond approximately 10 cM. We applied this guideline to the coffee comparative map in Fig. 1 for the putative syntenic correspondence (designated as colors for the tomato chromosomes) extends 10 cM on either side of each orthologous marker on the coffee map unless additional orthologous markers have been mapped in the region in question. The purpose of this exercise was not to deduce or characterize all genome rearrangements for which the coffee and tomato genomes (see subsequent sections for discussion of this topic), but rather to provide users a semi-quantitative tool for determining the probably counterparts in the tomato genome for as much of the coffee genome as possible (Table 2, electronic supplementary material).

Tomato chromosome 7 displays high syntenic relationship with coffee linkage groups E and F

Out of the 34 COSII markers mapped on tomato chromosome 7, up to 32 correspond to coffee linkage groups E and F (Table 2) stressing that the syntenic relationship between the two species is the highest in these particular regions of the tomato and coffee genomes. Therefore, these genomic regions offer a favorable material for more detailed analysis of the syntenic relationship between tomato and coffee genomes. Since tomato chromosome 7 is currently being

sequenced by a consortium of French scientists (M. Bouzayen et al. personal communication, Mueller et al. 2009), we thus took advantage of this sequence information to further investigate the fine scale synteny between tomato chromosome 7 and its counterparts in the coffee genome.

Screening for additional COSII markers by Blast searches against the genomic sequence of tomato chromosome 7

The tomato BAC sequences obtained on chromosome 7 were screened by blast for the presence of COSII. We screened the sequences for the presence of COSII mapped on chromosome 7 in the tomato map (67 COSII) and on coffee linkage groups E and F (90 COSII). The results presented in Table 3 shows that out of the 67 COSII mapped on tomato chromosome 7, only 54 were found in the BAC sequences. It is expected not to retrieve the whole set of COSII markers since the chromosome 7 sequence is currently incomplete. Among those 54 markers, 21 could be mapped and localized to linkage group E and F in coffee. On the other hand, five additional COSII markers mapped on coffee linkage group E and F but not present on the tomato genetic map were retrieved in the chromosome 7 BAC sequences which allowed to locate them on the tomato chromosome 7 genetic map. We also looked for other COSII that had been mapped neither on coffee nor on tomato. Blast search on tomato chromosome 7 BAC sequence and COSII set previously described by Wu et al. (2006) identified 18 additional COSII sequences, the coffee mapping of which was only successful for three due to a lack of polymorphism between the parents. Marker C2_At4g24820 maps on coffee linkage group F while C2_At2g42810 and C2_At3g20630 markers are located on linkage group E (Table 3, Fig. 3). These data illustrate the predictive value of coffee-tomato synteny. Overall, the present study allowed the selection of 35 COSII markers that are common between tomato chromosome 7 and coffee linkage groups E and F among which 20 are found on coffee linkage group E and 15 on linkage group F (Table 3,

Table 3 List of the COSII markers in tomato chromosome 7 and/or coffee linkage groups E and F

CosII name	Coffee unigene	Tomato unigene	Coffee group	Coffee cM	Tomato group	Tomato cM	Tomato BAC name	Genetic markers other than COSII assigned to tomato chromosome 7 BACs
C2_At4g15420	SGN-U309339	SGN-U323351	E	33.9	7	44.9	C07HBa0066L12	
C2_At2g20690	SGN-U301762	SGN-U323258	E	54.7	7	42.2	C07HBa0019I13	
C2_At2g42750	SGN-U308948	SGN-U321457	E	66.9	7	44.3	C07SLm0075L20	
C2_At2g42810	SGN-U304455	SGN-U317379	E	69	7	45	C07SLe0095F20/ C07HBa0023C09	
C2_At3g58790	SGN-U303091	SGN-U322143	E	72.3	7	44.6	C07HBa0018L21	
C2_At2g38020	SGN-U308527	SGN-U326005	E	74.3	7	49	C07HBa0116M01	
C2_At3g15290	SGN-U309818	SGN-U320561	E	82.9	7	63.5	C07HBa0046G04	
C2_At1g52980	SGN-U302053	SGN-U315903	E	84.7	7	68	C07HBa0227C07	
C2_At3g15430	SGN-U310705	SGN-U320751	E	85.4	7	68	C07HBa0227C07	
C2_At5g54310	SGN-U301021	SGN-U316297	E	92.8	7	75	C07HBa0140O20	
C2_At4g26750	SGN-U302171	SGN-U320345	E	94.7	7	80	C07HBa0061J13/ C07SLm0088O01	
C2_At1g55670	SGN-U300459	SGN-U314594	E	97.4	7	110	C07HBa0309F18/ C07SLm0032F07	
C2_At1g55870	SGN-U305006	SGN-U329762	E	97.5	7	90	C07HBa0309B15	
C2_At5g34850	SGN-U299906	SGN-U316123	F	33.4	7	0.4	C07SLm0046H06	
C2_At5g20350	SGN-U304405	SGN-U319705	F	36.5	7	0	C07HBa0127J08	
C2_At5g20180	SGN-U301390	SGN-U322000	F	39.5	7	6	C07HBa0037D07	
C2_At4g29490	SGN-U306923	SGN-U313555	F	50.8	7	22.3	C07HBa0108F03	
C2_At5g56940	SGN-U309065	SGN-U325196	F	54.6	7	11	C07SLe0077I03	
C2_At5g20520	SGN-U308094	SGN-U318539	F	56.0	7	22.3	C07HBa0041L08	
C2_At3g26060	SGN-U302141	SGN-U314061	F	60.9	7	28	C07HBa0287B22/ C07HBa0025K09	
C2_At5g57655	SGN-U300975	SGN-U315667	F	64.1	7	0.1	C07HBa0162M15	
C2_At2g24270	SGN-U299764	SGN-U314377	F	70.4	7	0.3	C07HBa0002D20/ C07HBa0166N19	
C2_At5g58200	SGN-U310609	SGN-U324859	F	77.2	7	0.4	C07SLe0111B06	
C2_At4g24820	SGN-U301422	SGN-U315297	F	130	7	13.5	C07HBa0030C22	
C2_At4g30580	SGN-U300974	SGN-U330308	-	-	7	2	C07HBa0111F22	
C2_At1g14850	SGN-U300685	SGN-U325116	-	-	7	3.5	C07HBa0078D19	
C2_At1g19140	SGN-U303267	SGN-U319846	-	-	7	23.5	C07HBa0002M15	
C2_At4g33250	SGN-U310651	SGN-U317255	-	-	7	29	C07SLm0121E18	
C2_At4g31040		SGN-U327356	-	-	7	29.2	C07SLm0141H03	
C2_At2g30520	SGN-U299771	SGN-U317317	-	-	7	31	C07HBa0049C13	
C2_At5g14520	SGN-U304670	SGN-U318628	G	19	7	38.3	C07HBa0212E04	
C2_At4g26680	SGN-U301367	SGN-U332887	-	-	7	38.4	C07HBa0212E04/ C07HBa0117J06	
C2_At2g04842		SGN-U331763[a]	-	-	7	40	C07SLm0093E04/ C07HBa0066L12	
C2_At2g32970	SGN-U312062	SGN-U315194	-	-	7	40.7	C07HBa0011E08	
C2_At2g04780		SGN-U316880	-	-	7	40.8	C07HBa0066L12	
C2_At3g21870	SGN-U304437	SGN-U326092	-	-	7	42.5	C07SLm0137M12	
C2_At2g20830	SGN-U307507	SGN-U317700	-	-	7	43	C07HBa0045O10	
C2_At3g21820		SGN-U327486	-	-	7	43	C07HBa0058M08	
C2_At2g20860	SGN-U307971	SGN-U318499	-	-	7	43.3	C07HBa0048I07	
C2_At1g17200		SGN-U320713	-	-	7	46	C07HBa0018L21	
C2_At1g53670		SGN-U317631	-	-	7	54	C07HBa0102J11	
C2_At3g14770	SGN-U301157	SGN-U331866	-	-	7	56.5	C07HBa0102J11	

Table 3 (continued)

CosII name	Coffee unigene	Tomato unigene	Coffee group	Coffee cM	Tomato group	Tomato cM	Tomato BAC name	Genetic markers other than COSII assigned to tomato chromosome 7 BACs
C2_At1g78620	SGN-U309305	SGN-U325052	-	-	7	57	C07HBa0102J11	
C2_At3g14910	SGN-U301769	SGN-U327185	-	-	7	61	C07HBa0079F09	
C2_At1g16900		SGN-U306380	-	-	7	62.5	C07HBa0079F09	
C2_At3g15380		SGN-U304912	-	-	7	69	C07HBa0137J06	
C2_At1g02180	SGN-U300475	SGN-U326199	-	-	7	73	C07SLe0121N11	
C2_At1g04970	SGN-U306315	SGN-U319356	-	-	7	84	C07HBa0061J13/ C07SLm0088O01	
C2_At1g50575	SGN-U311567	SGN-U322999	-	-	7	85.5	C07HBa0184E04	
C2_At4g16450		SGN-U320871	-	-	7	88.5	C07HBa0184E04/ C07HBa0033E17	
C2_At1g55880	SGN-U312097	SGN-U332466	G	72	7	89	C07HBa0033E17	
C2_At1g55840	SGN-U302634	SGN-U318286	-	-	7	92.5	C07HBa0309B15	
C2_At1g56050	SGN-U308896	SGN-U316710	-	-	7	100	C07HBa0215P04/ C07SLm0077G20	
C2_At5g56130	SGN-U307866	SGN-U316647	-	-	7	108	C07HBa0309F18/ C07SLm0032F07	
C2_At1g03820	SGN-U309861	SGN-U328972	E	43	7	49	-	
C2_At1g03600	SGN-U302366	SGN-U319106	E	44.5	7	55	-	
C2_At4g03210	SGN-U300349	SGN-U318702	E	44.7	7	57	-	
C2_At2g24090	SGN-U308290	SGN-U320263	F	0	7	1	-	
C2_At2g18710	SGN-U303383	SGN-U317205	F	0.1	7	3	-	
C2_At3g43230	SGN-U304184	SGN-U325966	F	16	7	22	-	
C2_At2g24120		-	-	7	0.4	-		
C2_At2g25620		SGN-U325025	-	-	7	0.4	-	
C2_At5g22010		SGN-U308003	-	-	7	2	-	
C2_At2g16060	SGN-U302107	SGN-U317560	-	-	7	3.5	-	
C2_At2g26590	SGN-U300895	SGN-U320442	-	-	7	15	-	
C2_At2g06925		SGN-U326975	-	-	7	19	-	
C2_At3g13050	SGN-U300474	SGN-U326276	-	-	7	27.5	-	
C2_At1g06690	SGN-U307430	SGN-U313507	E	2.6	-	-	C07HBa0165I16	TM10 cM 33 and T1110 cM 35
C2_At3g20630	SGN-U303035	SGN-U325263	E	82	-	-	C07HBa0140O20	T1366 cM 71 and T1257 cM 72
C2_At3g20720	SGN-U302701	SGN-U327660	E	88.5	-	-	C07HBa0130B18	T1738 cM 72.5 and TG438 cM 73
C2_At5g54770	SGN-U300441	SGN-U312660	E	90.9	-	-	C07HBa0226J04	between TG438 cM 73 and T0966 cM 78
C2_At2g06050	SGN-U300552	SGN-U315636	F	58.8	-	-	C07HBa0037F23/ C07HBa0143I18	between CT52 cM 2 and TG576 cM 2.9

COSII are assigned to tomato chromosome 7 either after genetic mapping or blast of the chromosome 7 BAC sequences. In all cases, sequence identity with tomato unigene is higher than 99% and sequence identity with Arabidopsis COSII or coffee unigene sequence higher than 79%

SGN-xx[a] this unigene was not in the SGN COSII reference table (ftp://ftp.sgn.cornell.edu/COSII/tomato_mapping/). It was found by BLAST

Fig. 3). Coffee group F matches the majority of the short arm of tomato 7 and coffee group E matches the majority of the long arm of tomato 7 (Fig. 3).

It is also important to mention that our blast search revealed for COSII markers C2_At5g48300 and C2_At1g78600, previously mapped on chromosome 12, a perfect match

(100% identity) with two sequences located on chromosome 7 but no homologous sequence in chromosome 12 nor in any other chromosome. Moreover, COSII maker C2_At5g48300 has been mapped on chromosome 7 in two other Solanaceae species closely related to tomato, pepper, and egg plant (Wu et al. 2009a, b). These two markers were therefore added to the

set of markers identified as common between tomato chromosome 7 and coffee linkage group E and F. Moreover, these two COSII genes are positioned in a similar order in tomato chromosome 7 and coffee linkage group E further supporting the high syntenic relationship between the two species chromosomal regions.

Reordering the tomato chromosome 7 COSII markers by integrating the physical and genetic maps

Establishing contigs of BAC sequences in the tomato chromosome 7 revealed several inversions between the genetic map position of the marker and its physical position. Considering that the synteny study between tomato and coffee is based on marker location and order, the COSII map of tomato chromosome 7 was updated on the light of the integration of the data related to the physical and genetic positions of the markers used in the present study (Fig. 3). However, as the tomato genome sequence is still incomplete, we cannot exclude the possibility that other inversions might be revealed as the sequencing efforts progress further. Tomato chromosome 7 markers that show order discrepancy of particular importance for the present study are located in the extremity of the short and long arm, close to the telomeres. Comparison of the data presented in Fig. 3 and Table 3 reveals that four COSII markers are mis-ordered in the genetic map as revealed by their physical position determined by sequence analysis. Markers C2_At5g58200 and C2_At2g24270 were genetically mapped at position cM 0.4 and cM 0.3, respectively, while the analysis of the contig sequence shows that they are in the reverse order. Likewise, markers C2_At5g57655 maps at position cM 0.1 whereas it is found in a BAC contig containing marker T0256 located at cM 1.2. The position of the contig containing this marker was confirmed by FISH analysis to be lower than that of C2_At5g58200 and C2_At2g24270 (data not shown). Marker C2_At5g34850 located at cM 0.4 according to the tomato genetic map is contained in a BAC harboring also marker T1940 mapped at position cM 2.3. The order retained for these four markers on the basis of the position of the BAC contigs from top to bottom of the short arm of chromosome 7 is the following: C2_At5g58200, C2_At2g24270, C2_At5g57655, and C2_At5g34850. These loci order discrepancy is due to the low number of progenies used for tomato and coffee genetic mapping.

Relationship of chromosome structure and genome evolution

Location of heterochromatin and euchromatin areas on tomato chromosome 7 (Fig. 3) was deduced from both sequence data analysis and FISH experiments. Interestingly, syntenic blocks shared with E and F are localized mainly in the euchromatin

Fig. 3 Detailed synteny between coffee linkage groups E &F and tomato chromosome 7. COSII markers and unigenes localized on tomato 7 chromosome and the coffee linkage groups E & F. Common markers are connected by lines. Loci order on the tomato chromosome 7 has been modified according to the physical mapping information (see text for discussion). Disruption between syntenic areas between linkage groups E and F located in or near the border between heterochromatin and euchromatin of the long arm ▶

regions of tomato chromosome 7. Moreover, the boundary of synteny between coffee linkage groups E and F on tomato chromosome 7 is located in or near the border between heterochromatin and euchromatin of the long arm.

Discussion

While the genomes of coffee and tomato have undergone extensive rearrangements (e.g., inversions and transloca-tions) since divergence from their last common ancestor, the data presented in this study indicate that it is still possible to decipher synteny across most of their genomes. The majority (75%) of syntenic blocks are short (<4 cM); however, some conserved tracts extend >20 cM (Figs. 1 and 2). For example, the lower 50 cM of coffee-linkage group E matches, in near-perfect order, the loci on the long arm of tomato chromosome 7 (Fig. 3). Coffee-linkage group F matches the majority of the short arm of tomato 7 while linkage group E matches the majority of the long arm of tomato 7. However, there is no direct relationship between genetic and physical distances as illustrated by Xu et al. (2008) in cotton for example.

Our data show that the boundary of synteny between coffee linkage groups E and F and tomato chromosome 7 lies near the heterochromatin/euchromatin border of the long arm of this tomato chromosome strongly supporting the hypothesis of the occurrence of a major breakpoint rearrangement event in this region.

Studies, such as the one described here, would not be possible without some foundation of genomic information in the species being studied. In this case, we relied heavily on the many genetic and genomics resources now available for tomato (Mueller et al. 2009). But, we also took advantage of a publicly available EST database for coffee (Lin et al. 2005). The information on tomato chromosome 7 genomic sequence and physical mapping of the BACs were essential to perform the screening of a large number of COSII markers to select those useful for the present study. Also essential for this study was the availability of a set of single copy, orthologous genes that could be used for mapping in both tomato and coffee. We therefore utilized the set of previously published COSII genes that meet all of these criteria (Wu et al. 2006). These COSII genes have already been used for comparative mapping studies in
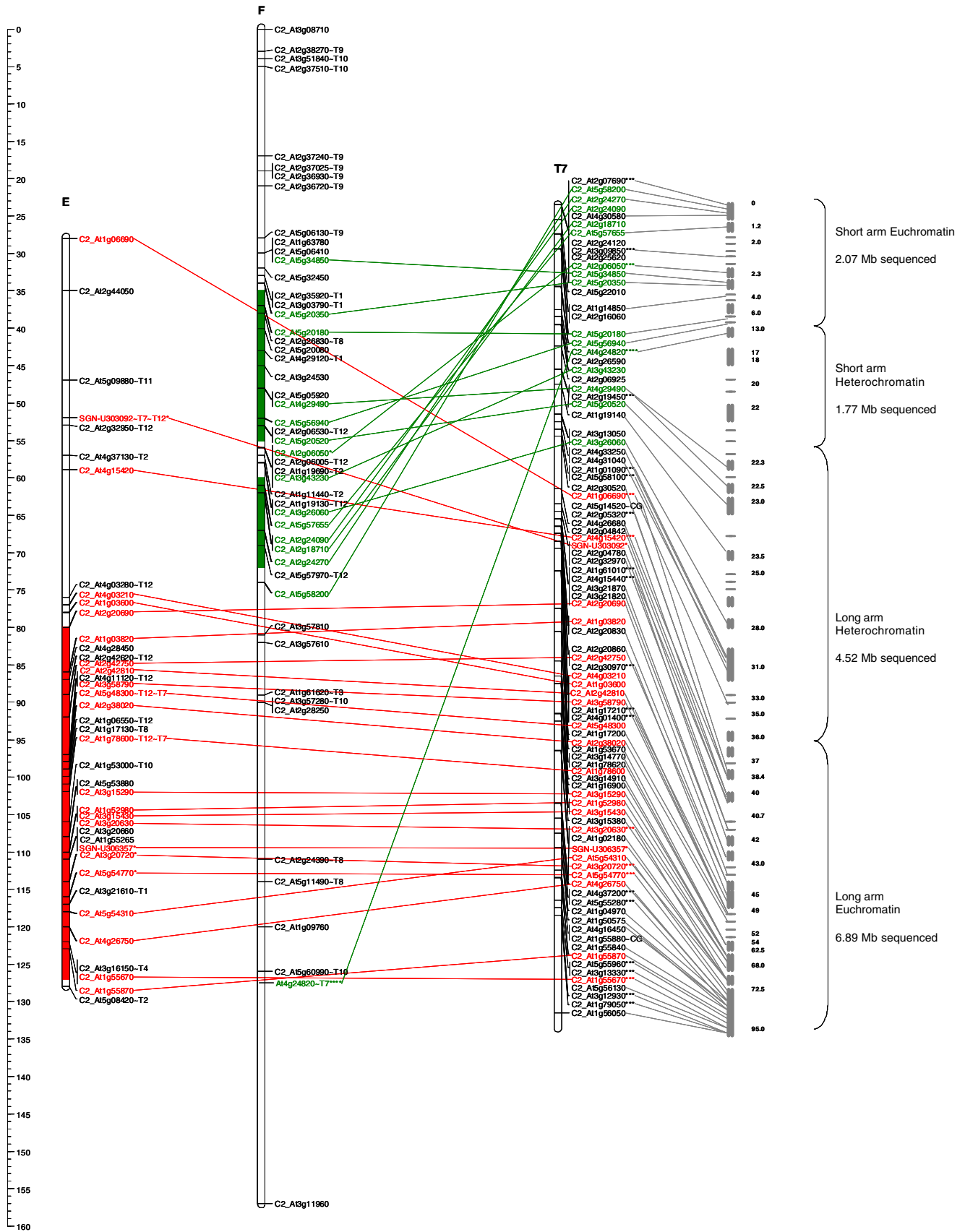
Solanaceous species (Wu et al. 2009a, b, c). The current study demonstrates that these COSII genes can also be used for comparative mapping across plant families—at least for the closely related plant families Solanaceae (tomato) and Rubiaceae (coffee). Understanding the modes and consequences of genome evolution across higher level taxa (such as families) may provide new insights into angiosperm diversification (Timms et al. 2006). The current study, comparing the genomes of coffee and tomato, represents one of the few cases in which synteny could be well-deciphered across plant families.

In addition to providing insights about genome evolution across plant families, the current study also has applied implications. Sufficiently detailed syntenic maps provide a means by which to access genomic information across species. For example, relative to coffee, many more genetic and genomic resources exist for tomato—the genome of which is currently being sequenced by an international consortium of scientists (Mueller et al. 2009). The comparative maps provided here may allow coffee researchers to identify the orthologous counterparts in the tomato genome sequence. This could potentially expedite the identification of QTLs and cloning of genes in coffee. The value of this has already been demonstrated in other taxa (Kilian et al. 1997). Even through a large public EST database for coffee is now available, (Lin et al. 2005; Poncet et al. 2006), very few of these ESTs have been genetically mapped in coffee—limiting their use as candidate genes for cloning studies. The tomato-coffee syntenic map reported herein, combined with the tomato genome sequence, may allow the coffee map position of many of these ESTs to be predicted by in silico mapping. This could greatly increase both the number of mapped genetic markers in coffee and provide a rich source of candidate genes for QTL cloning experiments.

Conclusions

All COSII genes mapped on coffee genome were the result of publicly available information and databases: coffee EST database (http://sgn.cornell.edu/search/library_search.pl?w5c4_library_parameters=coffea), COSII sequences (Wu et al. 2006), tomato COSII map (Wu et al. 2006, http://sgn.cornell.edu/cview/), and Arabidopsis genome database (http://www.arabidopsis.org/). This highlights the importance and benefit of scientist working on orphan plants to have access to such data. Based on these data, we were able herein to produce the first comparative map for coffee and to link it to the intensive genetic and genome resources available from research in tomato and other Solanaceae species. This also represents the first detailed synteny study dedicated to a woody plant species. Because the Rubiaceae (coffee) and Solanaceae families are relatively closely related and share a

common, diploid ancestor (with most probably x=n=11 or 12), we are able to report significant genetic conservation between these two species. While the level of conservation between coffee and tomato is not as high as that observed among Solanaceae species, it is nevertheless sufficiently high to allow the prediction of orthologous positions between coffee and tomato across significant part of the genome.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bevan M, Walsh S (2005) The *Arabidopsis* genome: a foundation for plant research. Genome Res 15(12):1632–1642

Davis AP, Govaerts R, Bridson DM, Stoffelen P (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). Bot J Linn Soc 152:465–512

Dirlewanger E, Graziano E, Joobeur T, Garriga-Calderé F, Cosson P, Howad W, Arus P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci U S A 101(26):9891–9896

Dussert S, Lashermes P, Anthony F, Montagnon C, Trouslot P, Combes MC, Berthaud J, Noirot M, Hamon S (1999) In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Diversité génétique des plantes tropicales cultivées, 175–794.

Gomez C, Dussert S, Hamon P, de Kochko A, Poncet V (2009) Current genetic differenciation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. BMC Evol Biol 9:167

Grant D, Cregan P, Shoemaker RC (2000) Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proc Natl Acad Sci USA 97(8):4168–4173

Guyot R, de la Mare M, Viader V, Hamon P, Coriton O, Bustamante-Porras J, Poncet V, Campa C, Hamon S, de Kochko A (2009) Microcollinearity in na ethylene receptor coding gene region of the *Coffea canephora* genome is extensively conserved with *Vitis vinifera* and other distantdicotyledonous sequenced genomes. BMC Plant Biol 9:22

Kilian A, Chen J, Han F, Steffenson B, Kleinhofs A (1997) Towards map-based cloning of the barley stem rust resistance genes Rpg1 and rpg4 using rice as an intergenomic cloning vehicle. Plant Mol Biol 35(1-2):187–195

Ku HM, Vision T, Liu J, Tanksley SD (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. Proc Natl Acad Sci USA 97(16):9121–9126

Lin C, Mueller LA, McCarthy J, Crouzillat D, Pétiard V, Tanksley SD (2005) Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. Theor Appl Genet 112(1):114–130

Livingstone KD, Lackney VK, Blauth JR, Van Wijk R, Jahn MK (1999) Genome mapping in capsicum and the evolution of genome structure in the solanaceae. Genetics 152(3):1183–1202

Mahé L, Combes MC, Lashermes P (2007) Comparison between a coffee single copy chromosomal region and *Arabidopsis* duplicated counterparts evidenced high level synteny between the coffee genome and the ancestral *Arabidopsis* genome. Plant Mol Biol 64(6):699–711

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD (2005a) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. Plant Physiol 138(3):1310–1317

Mueller LA, Tanksley SD, Giovannoni JJ, Van Eck J, Stack S, Choi D, Kim BD, Chen M, Cheng Z, Li C, Ling H, Xue Y, Seymour G, Bishop G, Bryan G, Sharma R, Khurana J, Tyagi A, Chattopadhyay D, Singh NK, Stiekema W, Lindhout P, Jesse T, Lankhorst RK, Bouzayen M, Shibata D, Tabata S, Granell A, Botella MA, Giuliano G, Frusciante L, Causse M, Zamir D (2005b) The tomato sequencing project, the first cornerstone of the international Solanaceae project (SOL). Comp Funct Genomics 6(3):153–158

Mueller LA et al (2009) A snapshot of the emerging tomato genome sequence. The Plant Genome 2:78–92

Mysore KS, Tuori RP, Martin GB (2001) Arabidopsis genome sequence as a tool for functional genomics in tomato. Genome Biol 2(1):REVIEWS1003

Pendergrast M (2009) Coffee second only to oil? Tea Coffee Trade J 181(4):38–41

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the Coffea genus. Mol Genet Genomics 276:436–449

Srinivasachary, Dida MM, Gale MD, Devos KM (2007) Comparative analyses reveal high levels of conserved colinearity between the finger millet and rice genomes. Theor Appl Genet 115(4):489–499

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: Join Map. Plant J 3:739–744

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132(4):1141–1160

Timms L, Jimenez R, Chase M, Lavelle D, McHale L, Kozik A, Lai Z, Heesacker A, Knapp S, Rieseberg L, Michelmore R, Kesseli R (2006) Analyses of synteny between *Arabidopsis thaliana* and species in the Asteraceae reveal a complex network of small syntenic segments and major chromosomal rearrangements. Genetics 173(4):2227–2235

Van Dodeweerd AM, Hall CR, Bent EG, Johnson SJ, Bevan MW, Bancroft I (1999) Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. Genome 42(5):887–892

Van Ooijen JW (2006) JoinMap®4.0 software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93(1):77–78

Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J, Siepel A, Tanksley SD (2008) Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae. Genetics 180(1):391–408

Wikstrom N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. Proc Biol Sci 268(1482):2211–2220

Wilson WA, Harrington SE, Woodman WL, Lee M, Sorrells ME, McCouch SR (1999) Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. Genetics 153(1):453–473

Wu F, Mueller LA, Crouzillat D, Pétiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics 174(3):1407–1420

Wu F, Eannetta NT, Xu Y, Tanksley SD (2009a) A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. Theor Appl Genet 118(5):927–935

Wu F, Eannetta NT, Xu Y, Durrett R, Mazourek M, Jahn MM, Tanksley SD (2009b) A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. Theor Appl Genet 118(7):1279–1293

Wu F, Eannetta NT, Xu Y, Plieske J, Ganal M, Pozzi C, Bakaher N, Tanksley SD (2009c) COSII genetic maps of two diploid *Nicotiana* species provide a detailed picture of synteny with tomato and insights into chromosome evolution in tetraploid *N. tabacum*. Theor Appl Genet. doi:10.1007/s00122-009-1206-z

Xu Z, Kohel RJ, Song G, Cho J, Yu J, Yu S, Tomkins J, Yu JZ (2008) An integrated genetic and physical map of homeologous chromosomes 12 and 26 in Upland cotton (*G. hirsutum* L.). BMC Genomics 9:108