Location of chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant

Pietro Gramazio1, Jaime Prohens1*, Mariola Plazas1, Isabel Andjar1, Francisco Javier Herraiz1, Elena Castillo1, Sandra Knapp2, Rachel S Meyer3,4 and Santiago Vilanova1

Abstract

Background: Eggplant is a powerful source of polyphenols which seems to play a key role in the prevention of several human diseases, such as cancer and diabetes. Chlorogenic acid is the polyphenol most present in eggplant, comprising between the 70% and 90% of the total polyphenol content. Introduction of the high chlorogenic acid content of wild relatives, such as S. incanum, into eggplant varieties will be of great interest. A potential side effect of the increased level polyphenols could be a decrease on apparent quality due to browning caused by the polyphenol oxidase enzymes mediated oxidation of polyphenols. We report the development of a new interspecific S. melongena × S. incanum linkage map based on a first backcross generation (BC1) towards the cultivated S. melongena as a tool for introgressing S. incanum alleles involved in the biosynthesis of chlorogenic acid in the genetic background of S. melongena.

Results: The interspecific genetic linkage map of eggplant developed in this work anchor the most informative previously published genetic maps of eggplant using common markers. The 91 BC1 plants of the mapping population were genotyped with 42 COSII, 99 SSRs, 88 AFLPs, 9 CAPS, 4 SNPs and one morphological polymorphic markers. Segregation marker data resulted in a map encompassing 1085 cM distributed in 12 linkage groups. Based on the synteny with tomato, the candidate genes involved in the core chlorogenic acid synthesis pathway in eggplant (PAL, C4H, 4CL, HCT, C3′H, HQT) as well as five polyphenol oxidase (PPO1, PPO2, PPO3, PPO4, PPO5) were mapped. Except for 4CL and HCT chlorogenic acid genes were not linked. On the contrary, all PPO genes clustered together. Candidate genes important in domestication such as fruit shape (OVATE, SISUN1) and prickliness were also located.

Conclusions: The achievements in location of candidate genes will allow the search of favorable alleles employing marker-assisted selection in order to develop new varieties with higher chlorogenic content alongside a lower polyphenol oxidase activity. This will result into an enhanced product showing a lower fruit flesh browning with improved human health properties.

Keywords: Chlorogenic acid, Genetic map, Polyphenol oxidases, Solanum incanum, Solanum melongena, Synteny

Background

Eggplant (Solanum melongena L., Solanaceae; 2n = 2 = 24) ranks third in the genus Solanum, after potato and tomato, in total production and economic importance and is the most important Solanaceae crop native to the Old World [1]. The most nutritionally important bioactive constituents of the eggplant fruit are phenolics, which are responsible of the high antioxidant activity of eggplant [2-6]. The most abundant phenolics of eggplant are hydroxycinnamic acid (HCA) conjugates, which are synthesized by converting phenylalanine to cinnamic acid. Among HCA conjugates, chlorogenic acid (5-O-caffeoyl-quinic acid; CGA) constitutes between 70% to over 95% of the total phenolics content [7-10]. Growing interest in this compound is due to its many beneficial properties for the treatment for various metabolic and cardiovascular diseases and
Several in vitro and in vivo experiments have shown that CGA has anti-oxidant, anti-inflammatory, analgesic, antipyretic, neuroprotective, cardioprotective, anti-carcinogenic, anti-microbial, hypotensive, anti-obesity and anti-diabetic activity [5,11,12]. Moreover, CGA is highly stable at high temperatures, and its bioavailability in eggplant increases, as compared to the raw product, after cooking [3].

Great diversity in the content of total phenolics and CGA has been observed in eggplant, due both to genetic and environmental factors [2,6,7,9,10,13]. Some close wild relatives of cultivated eggplant, such as S. inca num [14,15], have high levels of CGA [7,9,16]. Solanum inca num is native to northern Africa and the Middle East to Pakistan [15], and is a cross-compatible with S. melongena [1,7]. Therefore, S. inca num shows promise for use in breeding programs for developing new eggplant varieties with increased phenolic content [5].

Raising the total phenolics content, however, may cause a negative effect on apparent quality of the fruit. When eggplant fruit flesh is cut, phenolics, mostly stored in vacuoles, become available to polyphenol oxidase enzymes (PPOs), which are present in chloroplasts. PPOs catalyse the oxidation of phenolics to quinones, which in turn, react non-enzymatically with oxygen in the air to give brown compounds, thus causing browning of fruit flesh [17]. Several authors have found differences in PPO activity between varieties of eggplant, which can lead to differences in the degree of browning in fruit flesh between varieties with similar content of total phenolics [18-20]. Molecular breeding for high CGA content and low PPO activity could contribute to developing improved cultivars with higher bioactive properties through a combination of high antioxidant activity and presenting a low degree of browning. For this purpose, a candidate gene approach shows promise, given that the genes involved in the CGA synthesis pathway, which include phenylalanine ammonia lyase, PAL; cinnamate 4-hydroxylase, C4H; 4-hydroxycinnamoyl-CoA ligase, 4CL; hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase, HCT; p-coumaroyl ester 3-hydroxylase, C3H; and, hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase, HQT (Figure 1), in addition to the PPO genes, are known [20-24].

Understanding of eggplant genome organization, which is of great relevance for molecular breeding, has lagged behind that for other solanaceous crops such as potato, tomato and pepper. Several linkage maps for eggplant have been developed. Nunome et al. [25] developed a first intraspecific linkage map in eggplant using RAPD and AFLP markers. Two improved versions of the Nunome et al. [25] map were developed by adding SSR markers [26,27]. Doganlar et al. [28] developed the first interspecific map using RFLP markers resulting from crossing S. melongena and S. linnaeanum. The resolution of this map was further improved by adding COSII and AFLP markers [29,30]. Barchi et al. [31] also developed an intraspecific mostly AFLP and SSR marker. Finally, an intraspecific saturated integrated map of S. melongena was developed by Fukuoka et al. [32] from two F2 populations in which SSR and SNP markers were mapped. Of the markers used by Fukuoka et al. [32], many were obtained from Solanum orthologous (SOL) gene sets from a multiple alignment between the unigenes of eggplant, tomato and potato.

Here we report the development of a new interspecific S. melongena - S. inca num linkage map with the aim of locating, and in the future introgressing, S. inca num alleles involved in the biosynthesis of CGA in the genetic background of S. melongena. In order to devise molecular tools for minimizing browning associated with high CGA levels, PPO genes were also targeted. This new map is anchored to the tomato genetic map and previous eggplant maps, which will facilitate molecular breeding in eggplant for high CGA content and reduced browning as well as other morphological traits of importance in eggplant breeding.

Results
Genetic map construction
The mapping population was genotyped with 243 molecular markers comprising 42 COSII, 99 SSRs, 88 AFLPs, nine CAPS, four SNPs and the morphological marker PRICKLINESS. Genotypic data generated a genetic linkage map that spans 1085 cM distributed in 12 major and three minor linkage groups (Figure 2). Synteny with maps of Wu et al. [29], Fukuoka et al. [32], Barchi et al. [33] and Tomato-EXPEN 2000 [34] anchored the three minor linkage group to the corresponding major linkage groups (E05, E10, E11). The linkage groups ranged in length between 58.6 cM (E05) and 132.9 cM (E01) (Table 1). The average genome-wide density was 4.46 cM, with linkage group E01 having the lowest average density (5.77 cM inter-locus separation), and E08 showing the highest density (3.19 cM inter-locus separation). The number of loci per linkage group was highest in E06 (27) and lowest in E04 (16). Segregation distortion was observed for 22.6% of the markers (Table 1). The linkage groups with greater distortion were E02, E03 and E09 with around half of their markers skewed. A clear distortion in favour of S. inca num was found in E03 and E09 whereas in E02 and E06, alleles for S. melongena were more abundant. In order to develop a strong framework map, only markers joined at LOD > 3 were selected and those that had a lower LOD were discarded to avoid errors in positioning.

COSII analysis
A total of 35 (28.5%) out of the 123 COSII developed by Wu et al. [29] were polymorphic in our mapping
Figure 1 Biochemical pathway for chlorogenic acid (CGA) synthesis in eggplant. Enzymes involved in the CGA pathway are indicated: PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxilase; 4CL, 4-hydroxycinnamoyl-CoA ligase; HCT, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase; C3H, p-coumaroyl ester 3-hydroxilase; HQT, hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase [21-24].

Figure 2 The interspecific genetic linkage map SMIBC. Linkage groups were denominated E01 to E12 in agreement with the denomination of linkage groups in other eggplant genetic maps [28,32,33]. The map distances, given in cM, are shown on the left side of the linkage groups. Marker names are shown on the right side.
Seven other COSII markers were identified comparing the sequences of SSRs mapped with the tomato genome database (Sol Genomic Network). Six of them were EST-SSRs, obtained in silico, and one was a genomic SSR marker (Table 2). COSII markers allowed us to establish synteny with Wu et al. [29] eggplant map and with the Tomato EXPEN-2000 map [34] (Figures 3 and 4, Additional file 1: Figure S1) as well as with other members of the Asterid clade [35].

SSRs analysis
A total of 99 SSR markers of different sources were mapped in the BC1 population (Table 1). One hundred and twenty-eight of the 254 EST-SSRs, obtained in silico, using eggplant unigenes from the VegMarks database, showed homology with tomato unigenes (Sol Genomics Network). Of these, forty-seven were selected based on the theoretical position and screened for segregation in our parental (S. incanum and S. melongena) plants. Twenty of them (42.5%) showed polymorphism and were mapped. In addition, 71 genomic SSR markers from Nunome et al. [27], with an average of 4–5 per linkage group, were tested. Of these, 53 (74.6%) showed polymorphism and were positioned on the map. Most of them were also used by Fukuoka et al. [32] in the LWA2010 genetic integrated map, allowing the comparison between the two maps (Figures 3 and 4). Finally, 33 genomic SSRs (CSM markers) developed by Vilanova et al. [36] could be mapped. Seven of them were also used in Barchi et al. [33] in their genetic linkage map, enabling us to establish synteny (Figures 3 and 4).

AFLPs analysis
A total of 116 AFLP polymorphic bands were produced from 12 AFLP primer combination combined with three MseI primers. A total of 88 AFLP markers were identified and mapped (Table 1), with an average of 9.6 polymorphic bands per primer pair. Scoring only AFLP bands present in S. incanum and absent in S. melongena in a backcrossing population where S. incanum is a donor parent had the advantage that AFLPs could be scored in a codominant manner.

Table 1 Statistics of framework of SMIBC map

| Linkage groups | Statistics |
|---------------|------------|
|               | E01 | E02 | E03 | E04 | E05 | E06 | E07 | E08 | E09 | E10 | E11 | E12 | Total |
| Length (cM)   | 132.9 | 78.7 | 94.3 | 76 | 58.6 | 111.9 | 101.7 | 78.3 | 96.1 | 96.7 | 79.7 | 80.1 | 1085.0 |
| Number of markers | 23 | 18 | 21 | 16 | 18 | 27 | 21 | 20 | 22 | 19 | 19 | 19 | 243 |
| COSII         | 5 | 7 | 8 | 2 | 2 | 4 | 3 | 1 | 1 | 5 | 3 | 1 | 42 |
| SSRs          | 8 | 7 | 4 | 4 | 7 | 13 | 8 | 9 | 12 | 6 | 9 | 12 | 99 |
| AFLPs         | 9 | 3 | 7 | 10 | 9 | 8 | 9 | 5 | 8 | 7 | 7 | 6 | 88 |
| CAPS          | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 4 | 1 | 1 | 0 | 0 | 9 |
| SNPs          | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 4 |
| Morphological | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Average density (cM) | 5.77 | 4.37 | 4.49 | 4.75 | 3.25 | 4.14 | 4.84 | 3.19 | 4.36 | 5.09 | 4.19 | 4.21 | 4.46 |
| Gaps (>15 cM) | 2 | 0 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 3 | 16 |
| Skewed markers (P<0.05) | 0 | 9 | 11 | 0 | 3 | 5 | 2 | 1 | 11 | 7 | 2 | 4 | 55 |
| Percentage skewed markers | 0 | 50.0 | 52.3 | 0 | 16.6 | 18.5 | 9.5 | 5.0 | 50.0 | 36.8 | 10.5 | 21.0 | 22.6 |

The table shows the length in cM, the number of markers of each type, the average density in cM, the gaps larger than 15 cM and the skewed segregation for each linkage group. Linkage groups are designated as E01–E12.

Table 2 COSII markers identified in the SMIBC genetic map from CSM and EST-SSR markers based on sequence homology found after a BLASTN search on the SGN Cornell marker database [70]

| SMIBC LG | Current marker name | Previous marker name | Type of SSR |
|-----------|---------------------|-----------------------|-------------|
| E03       | C2_At3g17970        | CSM44                 | genomic SSR |
| E04       | C2_At1g09920        | SmFL32K22A            | EST-SSR     |
| E05       | C2_At3g24490        | LSS02D19A             | EST-SSR     |
| E06       | C2_At5g61910        | PLA03P07F             | EST-SSR     |
| E08       | C2_At5g46360        | ROT01104F             | EST-SSR     |
| E12       | C2_At4g32930        | LS004H02A             | EST-SSR     |
Figure 3 (See legend on next page.)
Mapping of CGA pathway genes

The six genes (Figure 1) involved in the core CGA synthesis pathway [21-24] were amplified and positioned in the SMIBC genetic map (Figure 2) based on the syntenic position with the Tomato EXPEN-2000 genetic linkage map [34] (Additional file 1: Figure S1, Additional file 2: Table S1).

**PAL (phenylalanine ammonia lyase)**

The search for phenylalanine ammonia lyase in SOL database give several orthologous gene sequence in tomato clustered in the same region of chromosome nine. A Blast of these orthologs in our local database allow to identify eggplant unigene OVS02A18A obtained by Fukuoka et al. [37]. A reciprocal Blast of OVS02A18A in SOL database shows the high homology with [SGN:Solyc09g007890.1.1] whose length is approximately 2.3 Kb and consist of two exons and one intron. In *S. melongena* and *S. incanum* parents a SNP (T/C) was found, after the amplification and sequencing of intron region, which was validated by restriction enzyme using a CAPS method (Additional file 3: Table S2). The gene could be mapped in the lower part of linkage group 6 that corresponded to unigene [SGN-U590064]. Although this tomato gene, positioned in the Tomato EXPEN-2000 map between markers C2_At3g51630 (92.5 cM) and T1789 (95.0 cM) (Figure 3), was not annotated as *C4H*, it is certainly orthologous to *C4H*.

**4CL (4-hydroxy-cinnamoyl-CoA ligase)**

Using a tomato 4CL sequence [SGN:Solyc03g117870.2], which was 3.6 Kb with five exons and four introns, it was possible to find the eggplant unigene SmFL38N19A. Analysis of the amplified sequences of introns three and four, allowed us to detect a polymorphism (A/G) that was transformed in a CAP marker (Additional file 3: Table S2). The gene could be mapped in the lower part of linkage group E03 at 90.9 cM from the linkage group end (Figure 2). The orthologous gene in tomato is positioned in linkage group 3 between COSII markers [SGN:C2_At1g09760] (133.3 cM) and [SGN:C2_At1g16180] (133.5 cM) (Figure 3).

**HCT (hydroxycinnamoyl-coA shikimate/quinate hydroxycinnamoyl transferase)**

The eggplant unigene ROT01O23W was identified using a tomato ortholog [SGN:Soly03g117600.2]. The gene in tomato (5.3 Kb) consists of three exons and two introns. The first intron was located in the 5’ UTR and there was insufficient data available for primers to be designed. The second intron was large (3.5 Kb) and we were unable to amplify it. No polymorphism between the two parental sequences was found after the amplification of first and the second exons. After a BLAST search with ROT01O23W unigene against database of contigs developed by Barchi et al. [38] we found a positive contig (22573:15433_PStI_67/3_NODE_1_L378; 15220_PStI_305E40_NODE_1_L282). On the basis of the contig sequence obtained, primers that partially amplify the second intron were developed. After the analysis of the sequenced amplicon, a SNP (T/A) was found and validated using HRM (Additional file 3: Table S2). The gene was mapped into the linkage group E03 at 89.6 cM very close to the 4CL gene (90.9 cM) (Figure 2). In the Tomato EXPEN-2000 map both genes also appear close together (Figure 3), separated only by 159.8 Kb in the tomato physical map.
Figure 4 (See legend on next page.)
C3 H (p-coumaroyl ester 3-hydroxilase)

An eggplant unigene YFR01120A was identified using a tomato orthologous sequence [SGN:Soly01g096670.2] which was approximately 3 Kb and contains three exons and two introns. Primers were developed to amplify the second intron of the gene where an Indel (TT) was found (Additional file 3: Table S2). Using this polymorphism as a CAP marker, the gene could be mapped in linkage group E01 at 85.9 cM from the linkage group end (Figure 2). Synteny study reveals that the linkage group 1 in eggplant and tomato are collinear except for minor position changes. The orthologous gene in tomato is located between the markers [SGN:T0852] (93.0 cM) and [SGN:cLPT-1-k7] (93.5 cM) (Figure 3).

HQT (hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase)

Using tomato ortholog sequence, the eggplant unigene YFR01H03A was identified. The sequence of the HQT gene in tomato [SGN:Soly07g005760.2] was 3.7 Kb and shows two exons and one large intron of 2.1 Kb. We tried to amplify the intron of the eggplant ortholog, but we were not able to obtain a clear band. The amplification of the two exons areas allowed us to detect a SNP (A/G) that was validated by HRM (Additional file 3: Table S2). The gene was located in the upper part of linkage group E07 at a distance of 3.0 cM from the first marker (Figure 2). The tomato orthologous gene is also positioned in the upper part of linkage group 7 between the markers [SGN:U176363] (0.2 cM) and [SGN:TG131] (2.0 cM) (Figure 4). Study of synteny based on 5 anchor points reveals certain collinearity between the two linkage groups although more anchor markers would be desirable.

Mapping of PPO genes

From the alignment of the PPO sequences published by Shetty et al. [20], primers were designed in order to amplify six PPO genes in S. melongena and S. incanum parental plants. In order to shorten the names, in this paper the SmePPO genes described by Shetty et al. [20] are here simply termed PPO. All of them were selectively amplified and sequenced except for PPO6, which was amplified in S. melongena but not in S. incanum. SNPs polymorphisms were found in the other five PPO genes. CAPS could be developed for PPO1 (C/A), PPO2 (C/G), PPO4 (G/A) and PPO5 (G/A and T/G) while the SNP of PPO3 (G/A) was validated by HRM (Additional file 3: Table S2). As we expected, based on synteny with Tomato EXPEN-2000 map [34] (Figure 4), all eggplant PPOs were mapped in SMIBC in the same genomic region in the linkage group E08, where PPO1 and PPO3 are situated at a distance of 42.2 cM, PPO2 and PPO4 at a distance of 42.4 cM, and PPO5 at 37.3 cM from the linkage group end (Figure 2).

Synteny reveals that PPO orthologous genes in tomato are located in an area of 95.5 Kb, comprising the markers [SGN:TG624] (36.70 cM) and [SGN:CIET-8-E2] (38.0 cM) (Figure 4). Several comparisons between eggplant and tomato PPO were made to correctly assign orthologous PPO genes between the two species, but a clear identification was not reached, probably due to high sequence similarity among the PPO genes of each species.

Mapping of other genes and traits of agronomic importance

The sequence of the OVADE gene [SGN:Soly02g085500.2], which determines the conversion from round to pear-shaped fruit in tomato, corresponded to eggplant unigene SmFL28E15A. The gene was mapped into the linkage E02 at 32.2 cM from the linkage group end (Figure 2). In Wu et al. [29] map, OVADE gene was mapped at 79.0 cM from the linkage group end (Figure 3), and in tomato the orthologous gene is positioned into linkage group 2 at 89.50 cM (Figure 3). The SISUN1 gene in tomato [SGN:SGN-U569959], which controls elongated and pointed fruit shape, is positioned in chromosome 10, near the marker [SGN:C2_AT5g10140] (52.80 cM) and shows a high identity with the eggplant contig (15541:35662_PstI_305E40_NODE_1_L230) [38]. The gene was mapped in SMIBC onto linkage group 10 at 79.0 cM from the linkage group end (Figure 2). The morphological marker PRICKLINESS could be mapped in linkage group E06 at 100.1 cM from the linkage group end (Figure 2).

Synteny and orthologous candidate genes with other maps

Synteny using common molecular markers was established in order to develop a genetic linkage map taking advantage of the information provided by previous eggplant genetic linkage maps (Figures 3 and 4, Additional file 2: Table S1). The reference maps were the F2 intraspecific maps developed by Nunome et al. [27], the integrated map LWA2010 derived from two F2 linkage maps.
published maps. The SMIBC was the first map that connects all the most informative previously published genetic maps for eggplant, allowing us to exploit the information generated by these maps (Figures 3 and 4). A total of 130 anchoring points connect SMIBC and Tomato EXPEN-2000, many of which are COSII, enabling us to establish synteny between the two genomes (Figures 3 and 4, Additional file 1: Figure S1, Additional file 2: Table S1). This allowed us to confirm the position of all candidate genes according to orthologous gene location in tomato. The syntenic relationships between these two maps are in agreement with those observed by Doganlar et al. [28], Wu et al. [29], and Fukuoka et al. [32] and provide information on the genome evolution of both crops.

A general problem in the construction of genetic maps is that different types of molecular markers tend to map in a specific regions of the genome [27,32,39,40]. Our use of a wide variety of molecular markers including those used in previous eggplant genetic maps [27,29,32,33] allowed us to achieve better representation of the eggplant genome, which is important for effective molecular breeding and synteny studies in the future.

Distorted markers occurred in almost all linkage groups, except for E01 and E04. Segregation distortion in mapping populations is a well-documented phenomenon in different crops [41,42]. Similar distortion levels to those we observed have been reported by Doganlar et al. [28] in their interspecific map between S. melongena and S. linnaeanum, with 16% of markers being skewed towards the one or the other parent.

CGA is the major phenolic compound present in eggplant, which is one of the crops with highest content in hydroxycinnamic acids [5,8-10]. Until now, attempts to improve the CGA content in eggplant have been conducted in a classical way through hybridization and selection of materials with high CGA content [4,5,7]. The results of these programs have been positive, but due to moderate heritability ($H^2$) of the CGA content trait and total phenolics in general, genetic advances have been limited [7,43,44]. Although these studies indicate that phenotypic selection to obtain materials with higher contents of CGA is possible, it would be desirable to apply marker-assisted selection (MAS) to improve the efficiency of selection for this trait.

The candidate gene (CG) approach has been successful in both animal and plant genetics [45]. The involvement of CGs in CGA biosynthesis pathway has been studied in plant species. The PAL gene catalyses one of the first steps of the phenylpropanoid pathway, that produces hundreds of specialized metabolic products including lignins, flavonoids, alkaloids and many other important phenolics in plants [46]. In Populus trichocarpa, five PAL genes have been described that are involved in wood formation [47],

Table 3 Markers with conflicting position according to the synteny between SMIBC interspecific eggplant map and Fukuoka et al. [32] LWA2010 eggplant genetic map

| LG SMIBC map | Marker name | LG LWA2010 map |
|--------------|-------------|----------------|
| E01          | emf21A23    | E08            |
| E06          | emh01M15    | E03            |
| E09          | emh02A04    | E11            |
| E09          | emh21L21    | E07            |
| E09          | emh11O01    | E03            |
| E11          | Emf01E10    | E01            |

Linkage group (LG) in which the markers are positioned in SMIBC and LWA2010 maps are shown.
and in raspberry (*Rubus idaeus*) two PAL genes have been identified: *RiPAL1* associated with early fruit ripening events while *RiPAL2* involved with larger flower and fruit development [48]. Candidate genes involved in phenylpropanoid pathway have also been mapped in other plants like apple [49] and artichoke [50-52]. These are examples showing that homologs can have different functions, and paralogs can play differently roles in CGA or lignin synthesis. Even within the Solanaceae, different genes in the pathway are strongly associated with the abundance of CGA. *HQT* seems to be the most important contributor to CGA synthesis in tomato and *Nicotiana benthamiana* [24], while *HCT* and *C3H* are the most relevant in potato [53,54].

In this work we were able to locate all the genes involved in the core CGA pathway in eggplant. The positions of the genes are in agreement with those expected based on synteny with tomato. With the exception of *HCT* and *4CL*, which are co-situated at a genetic distance of 1.3 cm in the linkage group E03, all of the rest of the genes of the biosynthesis pathway of CGA are situated in separate linkage groups. This has important implications for breeding and for strategies based on pyramiding of favourable alleles introgressed from *S. incanum* in the genetic background of eggplant.

A potential problem for developing materials with higher CGA content is their oxidation mediated by polyphenol oxidase (PPO), which leads to the browning of the fruit flesh, reducing apparent fruit quality [5,18]. A positive phenotypic linear correlation was found in eggplant between the degree of fruit flesh browning and total content in phenolic and CGA content [4,43].

Several studies in other crops associated quantitative trait loci (QTL) for enzymatic browning with PPO alleles and these alleles were then used as markers for this trait [55,56]. On the other hand, silencing of PPO genes result in a reduced browning reaction as has been shown in potato and apple [57,58].

The studies of PPO genes suggest that in plants they form a gene family with high homology, usually have no introns and normally are present in several copies. For instance, tomato has seven PPOs, potato has six PPOs and soybean has eleven PPOs [56,59-61]. These features have complicated the search of polymorphisms between PPOs because amplification of a single isofrom is difficult. In this study we were able to detect polymorphism in all PPO genes with the exception of *PPO6* (amplified in *S. melongena* but not in *S. incanum*) leading us to speculate that in *S. incanum* this PPO gene is absent or presents substantial changes with respect to *S. melongena*. As in tomato [60], eggplant PPOs were mapped very closely to each other in linkage group E08; this confirms the existence of a cluster also in eggplant (Figure 2), although a correct assignment using synteny was not possible between eggplant and tomato orthologs, due to high sequence similarity among the PPO genes of each species. In eggplant PPO genes identity varying between 72% and 95% at nucleotide level and between 62% and 92% at amino acid level [20].

Since both consumers and industry prefer eggplant varieties with white flesh and a low degree of browning, future varieties with high phenolic content should also be bred to have limited browning [43]. Accordingly, the search for allelic variants for increasing CGA content should be conducted in concert with a search for allelic variants decreasing PPOs. In eggplant, several authors have found differences in PPO activity between varieties [18-20] which suggest mining the biodiversity of this crop and its relatives can lead to discovering desirable alleles useful to produce varieties with low browning in fruit flesh regardless of the total phenolic abundance. In addition, the study of selection sweeps at these loci due to the culturally distinctive organoleptic preferences, would be of interest to understand the history of evolution of eggplant. A reduction in browning is quite probably a significant domestication trait in eggplant and many other species.

In the present study an interspecific genetic linkage map was developed to better understanding the genetics of CGA content and fruit flesh browning in eggplant through the mapping of candidate genes involved in these processes. This was assisted by the use of synteny.
of the orthologous genes in tomato. The use of *S. incanum* in this work and in breeding programs is not accidental, as it is a close relative of eggplant [14,15] and produces fully fertile hybrids with regular meiosis [7,14]. *Solanum incanum* has characteristics of great interest for plant breeders, such as bacterial and fungal resistance, including to *Fusarium oxysporum*, and tolerance to various abiotic stresses, especially drought [7,62]. Recent discoveries have shown that *S. incanum* is a powerful source of variation for phenolic content, as its total phenolic content is several times higher than that of *S. melongena* [7,9]. *Solanum incanum* also exhibits absence of anthocyanins, presence of prickles and subovoid green fruits [7].

Building an interspecific genetic map from the backcross generation of two species that differ in many traits has two important advantages. First of all, in this manner the problem of low intraspecific polymorphism in eggplant [28,29,32,63] is partially overcome. Secondly, this BC1 population can be used as the starting point to develop a set of introgression lines of *S. melongena*, with *S. incanum* as the donor parent. Once obtained, these lines will be useful to understand and dissect candidate gene roles in CGA accumulation and browning, to detect QTLs and genes involved in the synthesis of other polyphenols present in *S. incanum* (mainly N-(E)-caffeoylputrescine, 3-O-Malonyl-5-O-(E)-caffeoylquinic acid and 5-O-Malonyl-4-O-(E)-caffeoylquinic acid) [7], as well as many other morphological and physiological traits of interest for breeding or for crop evolution and domestication studies. Candidate genes for traits thought to be important in domestication such as fruit shape and PRICKLINESS are mapped in SMIBC in agreement with their positions based on previous QTL studies in eggplant and in syntenic regions of tomato [64,65], also highlighting the common occurrence of parallelism in loci causing domestication phenotypes between species.

**Conclusion**

Despite its economic importance few genomic and biotechnological resources exist for eggplant. Eggplant contains high quantities of phenolics, in particular of CGA, a secondary metabolite with many properties beneficial for human-health and which in the last years is attracting a lot of attention from the scientific community. We report the development of a new interspecific map in eggplant using different types of markers. We have mapped the six genes involved in CGA synthesis pathway on five linkage groups, and determined that the PPO genes cluster together on LG8. This new map is anchored to the tomato genetic map as well as to four previous genetic maps of eggplant, and therefore constitutes an important tool for exploiting synteny and for molecular breeding in eggplant. The molecular tools we have developed will be of interest for eggplant breeding programmes, in particular those aimed at using marker assisted selection for improved CGA content combined with low browning effects. These tools and results, together with the recently published eggplant genome sequence [66] will greatly facilitate the development of a new generation of eggplant varieties with dramatically enhanced bioactive properties that can significantly contribute to better human nutrition.

**Methods**

**Plant materials**

Parentals for developing the interspecific genetic linkage map (denominated SMIBC, making reference to *S. melongena* and *S. incanum* [SMI] and backcross [BC] mapping population) were *S. melongena* accession AN-S-26 and *S. incanum* accession MM577 (Figure 5). The F1 hybrid was obtained using AN-S-26 as a female parent and MM577 as male parent. After that, the F1 hybrid was backcrossed (as a female) to the *S. melongena* AN-S-26 parental to obtain a BC1 population. BC1 seeds were germinated and 91 BC1 plants were randomly selected for the development of the genetic linkage map.

**Development of the SMIBC genetic map**

**DNA extraction**

Total genomic DNA was isolated from leaves of the parental, F1 and BC1 plants according to the CTAB procedure [67]. DNA quality was checked in 1% agarose gels. DNA concentrations were measured with a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, USA).

**COSII markers**

A total of 123 Universal Primers of COSII markers [68], previously mapped in an interspecific cross between *S. linnaeum* and *S. melongena* [29], were screened in parental and F1 plants, and in the BC1 mapping population. Amplicons were analyzed on an ABI PRISM 3100-Avant DNA genetic analyzer (Applied Biosystems, Foster City, California). Amplicon sequences were examined with Sequence Scanner v1.0 (Applied Biosystems) software and aligned using the CAPS Designer software.
three additional selective nucleotides at the 3' end of each primer. Each 5' end of the EcoRI primers was labelled with different fluorescent dyes (PET, FAM, NED, and VIC). AFLP fragment analysis was performed on an automatic capillary electrophoresis sequencer ABI PRISM 3100-Avant (Applied Biosystems, Foster City, CA). The data were analysed using GeneScan (Applied Biosystems) and Genotyper DNA Fragment Analysis (Applied Biosystems) software as above. Only AFLP bands that were present in in S. incanum and absent in S. melongena parental lines were scored.

**SSR markers**

Three sources of SSR markers were used: a) EST-SSRs developed in silico specifically for this study, b) genomic SSRs developed by Nunome et al. [27], and c) genomic SSRs developed using a genomic library [36]. Development of new SSR markers in silico was performed using 16,000 eggplant unigene sequences obtained from the VegMarks database (http://vegmarks.nivot.affrc.go.jp/VegMarks.jsp/index.jsp). The identification of perfect di- and tri-nucleotide motifs with a repeat of ≥6 times were identified using the software SciRoKo [69]. This allowed us the identification of 254 ESTs-SSRs, which were subjected to a BLASTN search of the SGN Cornell unigene database [70] (http://solgenomics.net/) against the tomato genome database. SSRs were validated in in silico analysis using the GeneScan software to obtain the electropherograms and polymorphisms were analyzed with Genotyper DNA Fragment Analysis software.

**AFLP markers**

Genomic DNA was digested with a two enzymes combination, EcoRI and MseI [71]. Then EcoRI and MseI adapters were ligated using an AFLP Core Reagent (Invitrogen, Carlsbad, CA) following the manufacturer’s instructions. A total of 12 AFLP primer combinations, generated by four EcoRI primers (EcoAGC, EcoACT, EcoACA, EcoACG) combined with three MseI primers (MseCA, MseC, MseCAC) were used. A pre-amplification PCR reaction was executed with primers based on the adapter sequences with one additional selective nucleotide at the 3' end of each primer (EcoA + MseC). Using an aliquot of the PCR product, a selective amplification reaction was performed with Eco and Mse primers combination with three additional selective nucleotides at
marker **PRICKLINESS** was developed by phenotyping the plants and taking data on the presence or absence of stem prickles.

**Linkage analysis and map construction**
The mapping population was genotyped and $\chi^2$ test ($\chi^2$ value $\leq \chi^2 = 0.05$) were performed to check if individual markers segregated following Mendelian ratios. Linkage analysis was carried out using Joinmap v4.0 software [75] to construct the map with settings LOD threshold of 3.0 and maximum recombination fraction $\theta = 0.4$. Kosambi mapping function [76] was used to convert recombination units into genetic distance (cM).

**Macro-synteny between SMIBC and other genetic linkage maps**
The synteny of SMIBC map with other eggplant genetic linkage maps was established using shared markers [27,29,32,33]. The macro-synteny of SMIBC with the Tomato EXPEN-2000 [34] was carried out mainly using COSII markers developed by Wu et al. [68]. Some SMIBC SSR markers were positioned based on reciprocal best-hit relationships using BLASTN search against the Tomato gene models CDS (ITAG release 2.30) sequence database. Synteny was visualized using the Circos software [77].

**Availability of supporting data**
All the data supporting our result are included in the article and in the Additional files.

**Additional files**

Additional file 1: Figure S1. Image of macro-synteny established between SMIBC interspecific eggplant genetic map and Tomato-EXPEN 2000 genetic map.

Additional file 2: Table S1. Data of the macro-synteny established between SMIBC interspecific eggplant genetic map, tomato and other eggplant genetic maps.

Additional file 3: Table S2. Data of molecular markers mapped in SMIBC interspecific eggplant genetic map.

**Competing interests**
The authors declare that they have no competing interests.

**Authors contributions**
PG, MP, IA, FJH, EC, and SV carried out the molecular markers analysis. PG and SV made the bioinformatics and syntenies studies. JP and SV conceived the study. JP, SK, RSM and SV designed the experiments. PG, JP, SK, RSM and SV wrote the paper. All authors read and approved the final manuscript.

**Acknowledgements**
This research has been partially funded by Ministerio de Economía y Competitividad and FEDER (grant AGL2012-34213). Pietro Gramazio is grateful to Universitat Politècnica de València for a pre-doctoral grant.

**Author details**
1Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain. 2Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK. 3Center for Genomics and Systems Biology, New York University, 12 Waverly Place, New York, NY 10003, USA. 4Center for Genomics and Systems Biology, New York University Abu Dhabi Research Institute, Abu Dhabi, United Arab Emirates.

Received: 18 June 2014 Accepted: 25 November 2014

**References**
1. Dauray MC, Lester RN: *Eggplant*. In *Topical Plant Breeding*. Edited by Charrier A, Jacquot M, Harjon S, Nicolas D, Plymouth, UK: CIRAD, SciencePublishers Inc; 2001:199 221.
2. Hansson PM, Yang R, Tsou D, Ledesma D, Enge L, Lee T: Diversity in eggplant (*Solanum melongena*) for superoxide scavenging activity, total phenolics, and ascorbic acid. *J Food Compos Anal* 2006, 19:594 600.
3. Lo Scalo R, Fibiani M, Meninella G, Rotino GL, Dal Sasso M, Culfici M, Spallino A, Braga PC: Thermal treatment of eggplant (*Solanum melongena L*) increases the antioxidant content and the inhibitory effect on human neutrophil burst. *J Agric Food Chem* 2010, 58:3371 3379.
4. Plazas M, Lpez-Gresa MP, Vilanova S, Torres C, Hurtado M, Gramazio P, Andjar I, Herriz FJ, Bells JM, Prohens J: Diversity and relationships in key traits for functional and apparent quality in a collection of eggplant: fruit phenolics content, antioxidant activity, polyphenol oxidase activity, and browning. *J Agric Food Chem* 2013, 61:8871 8879.
5. Plazas M, Andjar I, Vilanova S, Hurtado M, Gramazio P, Herraz FJ, Prohens J: Breeding for chlorogenic acid content in eggplant: interest and prospects. *Not Bot Horti Agrobot Cluj-Napoca* 2013, 41:126 35.
6. Menella G, Lo Scalo R, Fibiani M, D Alessandro A, Francesc G, Toppino L, Acciarri N, de Almeida AE, Rotino GL: Chemical and bioactive quality traits during fruit ripening in eggplant (*S. melongena L*) and allied species. *J Agric Food Chem* 2012, 60:1821 11831.
7. Prohens J, Whitaker BD, Plazas M, Vilanova S, Hurtado M, Blasco M, Gramazio P, Stommel JR: Genetic diversity in morphological characters and phenolic acids content resulting from an interspecific cross between eggplant, *Solanum melongena*, and its wild ancestor (*S. incanum*). *Ann Appl Biol* 2013, 162:242 257.
8. Singh AP, Luthria D, Wilson T, Vorsa N, Singh V, Banuelos GS, Pasakdee S: Polyphenols content and antioxidant capacity of eggplant pulp. *Food Chem* 2009, 114:955 961.
9. Stommel JR, Whitaker BD: Phenolic acid content and composition of eggplant fruit in a germplasm core subset. *J Am Soc Hort Sci* 2003, 128:704 710.
10. Whitaker BD, Stommel JR: Distribution of hydroxycinnamid acid conjugates in fruit of commercial eggplant (*Solanum melongena L*) cultivars. *J Agric Food Chem* 2003, 51:3448 3454.
11. Cho A, Jeon S, Kim M, Yeo J, Seo K, Choi M, Lee M: Chlorogenic acid exhibits anti-obesity property and improves lipid metabolism in high-fat diet-induced-obese mice. *Food Chem Toxicol* 2010, 48:957 963.
12. Dos Santos MD, Almeida MC, Lopes NP, De Souza GEP: Evaluation of the anti-inflammatory, analgesic and antipyretic activities of the natural polyphenol chlorogenic acid. *Biol Pharm Bull* 2006, 29:2236 2240.
13. Luthria D, Singh AP, Wilson T, Vorsa N, Banuelos GS, Vinayak BT: Influence of conventional and organic agricultural practices on the phenolic content in eggplant pulp: plant-to-plant variation. *Food Chem* 2010, 121:406 411.
14. Lester RN, Hasan SMZ: Origin and domestication of the brinjal eggplant, *Solanum melongena*, from *S. incanum*, in Africa and Asia. In *Solanaceae III: Taxonomy, Chemistry, Evolution*. Edited by Hawkes JG, Lester RN, Nee MH, Estrada N, Kew, London, UK: Royal Botanic Gardens; 1991:369 388.
15. Knapp S, Varontoncova MS, Prohens J: Wild relatives of the eggplant (*Solanum melongena L*: Solanaceae): new understanding of species names in a complex group. *PloS One* 2013, 8:67039.
16. Ma C, Dastralmchi K, Whitaker BD, Kennelly EJ: Two new antioxidant malonated caffeoylquinic acids isomers in fruits of wild eggplant relatives. *J Agric Food Chem* 2011, 59:9645 9651.
17. Ramirez EC, Whitaker JR, Virador VH: Polyphenol Oxidase. In *Handbook of Food Enzymology*. Edited by Whitaker JR, Voragen AGJ, Wong DWS, New York: Marcel Dekker, Inc; 2003:509 524.
18. Concellin A, Zaro MI, Chaves AR, Vicente AR: Changes in quality and phenolic antioxidants in dark purple American eggplant (*Solanum melongena L* cv. *Luca*) as affected by storage at OC and 10C. *Postharvest Biol Technol* 2012, 66:35 41.
61. Tran L, Taylor J, Constabel C. The polyphenol oxidase gene family in land plants: lineage-specific duplication and expansion. BMC Genomics 2012, 13:395.
62. Yamakawa K, Mochizuki H. Nature and inheritance of Fusarium-wilt resistance in eggplant cultivars and related wild Solanum species. Bull Veg Crops Res Soc 1979, 6:19-27.
63. Stgel A, Portis E, Toppino L, Rotino GL, Lanteri S. Gene-based microsatellite development for mapping and phylogeny studies in eggplant. BMC Genomics 2008, 9:357.
64. Frary A, Frary A, Daunay M, Huvenaars K, Mank R, Doğanlar S. QTL hotspots in eggplant (Solanum melongena) detected with a high resolution map and CIM analysis. Euphytica 2014, 197:211-228.
65. Portis E, Barchi L, Toppino L, Lanteri S, Acciarri N, Felicioni N, Fusari F, Barbierato V, Cericola F, Val G, Rotino GL. QTL Mapping in eggplant reveals clusters of yield-related loci and orthology with the tomato genome. PLoS One 2014, 9:e89499.
66. Hirakawa H, Shirasawa K, Miyatake K, Nunome T, Negoro S, Ohyama A, Yamauchi H, Sato S, Isobe T, Tabata S, Fukuda H. Draft genome sequence of eggplant (Solanum melongena L.): the representative solanum species indigenous to the old world. DNA Res 2014, 1:12. doi:10.1093/dnares/dsu027
67. Doyle J, Doyle J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 1987, 19:11-15.
68. Wu F, Mueller LA, Crouzillat D, Piard V, Tanksley SD. 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 2006, 174:1407-1420.
69. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 1999, 41:95-98.
70. Rozen S, Skaltsky H. Primer3 on the WWW for general users and for biologist programmers. Bioinformatics 1999, 132:365-386.
71. van Ooijen J. JoinMap 4. In Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands; 2006.