# Research article

# **Open Access**

# A linkage map for the B-genome of Arachis (Fabaceae) and its synteny to the A-genome

Márcio C Moretzsohn<sup>\*1</sup>, Andrea VG Barbosa<sup>2</sup>, Dione MT Alves-Freitas<sup>3</sup>, Cristiane Teixeira<sup>1</sup>, Soraya CM Leal-Bertioli<sup>1</sup>, Patrícia M Guimarães<sup>1</sup>, Rinaldo W Pereira<sup>3</sup>, Catalina R Lopes<sup>2</sup>, Marcelo M Cavallari<sup>2</sup>, José FM Valls<sup>1</sup>, David J Bertioli<sup>3</sup> and Marcos A Gimenes<sup>1</sup>

Address: <sup>1</sup>Embrapa Recursos Genéticos e Biotecnologia, C.P. 02372, CEP 70.770-900, Brasília, DF, Brazil, <sup>2</sup>Departamento de Genética, IB-UNESP, Rubião Jr, CEP 18618-000, Botucatu, SP, Brazil and <sup>3</sup>Universidade Católica de Brasília, Campus II, SGAN 916, CEP 70.790-160, Brasília, DF, Brazil

Email: Márcio C Moretzsohn\* - marciocm@cenargen.embrapa.br; Andrea VG Barbosa - andreagobbi@gmail.com; Dione MT Alves-Freitas - dionebio@gmail.com; Cristiane Teixeira - cristi.teixeira@gmail.com; Soraya CM Leal-Bertioli - soraya@cenargen.embrapa.br; Patrícia M Guimarães - messenbe@cenargen.embrapa.br; Rinaldo W Pereira - rinaldo@pos.ucb.br; Catalina R Lopes - dtcatalina@terra.com.br; Marcelo M Cavallari - mmcavall@gmail.com; José FM Valls - valls@cenargen.embrapa.br; David J Bertioli - david@pos.ucb.br; Marcos A Gimenes - gimenes@cenargen.embrapa.br

\* Corresponding author

Published: 7 April 2009

BMC Plant Biology 2009, 9:40 doi:10.1186/1471-2229-9-40

This article is available from: http://www.biomedcentral.com/1471-2229/9/40

© 2009 Moretzsohn et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 5 December 2008 Accepted: 7 April 2009

### Abstract

**Background:** Arachis hypogaea (peanut) is an important crop worldwide, being mostly used for edible oil production, direct consumption and animal feed. Cultivated peanut is an allotetraploid species with two different genome components, A and B. Genetic linkage maps can greatly assist molecular breeding and genomic studies. However, the development of linkage maps for *A. hypogaea* is difficult because it has very low levels of polymorphism. This can be overcome by the utilization of wild species of *Arachis*, which present the A- and B-genomes in the diploid state, and show high levels of genetic variability.

**Results:** In this work, we constructed a B-genome linkage map, which will complement the previously published map for the A-genome of *Arachis*, and produced an entire framework for the tetraploid genome. This map is based on an  $F_2$  population of 93 individuals obtained from the cross between the diploid *A. ipaënsis* (K30076) and the closely related *A. magna* (K30097), the former species being the most probable B genome donor to cultivated peanut. In spite of being classified as different species, the parents showed high crossability and relatively low polymorphism (22.3%), compared to other interspecific crosses. The map has 10 linkage groups, with 149 loci spanning a total map distance of 1,294 cM. The microsatellite markers utilized, developed for other *Arachis* species, showed high transferability (81.7%). Segregation distortion was 21.5%. This B-genome map was compared to the A-genome map using 51 common markers, revealing a high degree of synteny between both genomes.

**Conclusion:** The development of genetic maps for *Arachis* diploid wild species with A- and B-genomes effectively provides a genetic map for the tetraploid cultivated peanut in two separate diploid components and is a significant advance towards the construction of a transferable reference map for *Arachis*. Additionally, we were able to identify affinities of some *Arachis* linkage groups with *Medicago truncatula*, which will allow the transfer of information from the nearly-complete genome sequences of this model legume to the peanut crop.

# Background

Peanut (*Arachis hypogaea* L.) is one of the most important crops in tropical and subtropical regions of the world. Peanut is used as both human and animal food, being a valuable source of protein and oil [1,2]. The genus *Arachis* (Leguminosae or Fabaceae) is native to South America and contains 80 described species assembled into nine taxonomical sections, according to their morphology, geographic distribution and sexual compatibility [3,4]. The *Arachis* section includes the species that can be crossed to *A. hypogaea* and encompasses 29 diploid species and the tetraploid species *A. hypogaea* and *A. monticola* [3,4].

Cultivated peanut is an allotetraploid  $(2n = 4 \times = 40 \text{ chromosomes})$  with two genome types, A and B, which are found separately in the wild species of the *Arachis* section. The A-genome species are diploids characterized by the presence of a so-called A chromosome pair [5], of reduced size and with a lower level of euchromatin condensation in comparison to the other chromosomes [6]. Diploid species of the section *Arachis* with 2n = 20 and lacking the A chromosome pair are usually considered to share the B-type genome, although they are much more heterogeneous and may present variant forms of this B-genome. One species, *A. glandulifera*, revealed very poor homologies with all A and B genome taxa, and is considered to have a D genome [7,8]. Three other species show 2n = 18 chromosomes [9-11] and their genomic affinities are not clear.

*Arachis hypogaea* was originated via hybridization of two diploid wild species, probably *A. duranensis* (A-genome) and *A. ipaënsis* (B-genome), followed by a rare spontaneous duplication of chromosomes [6,12-14]. The resulting tetraploid plant would have been reproductively isolated from its wild diploid relatives. This isolation, coupled with the origin through a probably single hybridization event [13,15-17], leads to a limited genetic diversity of peanut, as observed in different studies using molecular markers [13,15-17]. In contrast, wild diploid *Arachis* species are genetically more diverse [18-20], providing a rich source of variation for agronomical traits, and DNA polymorphisms for genetic and genomic studies [21-23].

As a consequence, most of the linkage maps developed for *Arachis* included wild species as progenitors, the exception being the *A. hypogaea* map that has been recently published [24]. These maps are based on RFLP [25,26], RAPD [27], and more recently, microsatellite markers [24,28]. In this latter study [28] we used a diploid population from a cross between *A. duranensis* and the closely related *A. stenosperma*, both having A-type genomes, the former being the most probable A genome donor to cultivated peanut. This map, which essentially provides genetic information for half the genetic component of *A. hypogaea*, has more

recently been updated with new microsatellites, RGAs, AFLPs, and single-copy gene-based markers (anchor markers) (unpublished data).

Microsatellite markers are the ideal markers for the development of linkage maps, as they are multiallelic, highly polymorphic, typically co-dominant, and PCR-based markers. Additionally, they can often be transferred between different populations and even related species [28-31]. Therefore different maps constructed with common microsatellite markers can be aligned, allowing information from the different maps to be accumulated, helping to confirm linkage orders and providing information on the genome evolution of related species.

The aim of this study was to create a linkage map for the *Arachis* B-genome to complement the previously published A-genome map and effectively to provide a linkage map for tetraploid peanut in two separate diploid components. For that, we made an  $F_2$  population from a cross between the most probable B-genome donor of cultivated peanut, *A. ipaënsis* [13,14], and the very closely related *A. magna.* In order to facilitate map comparisons we used the same set of microsatellite markers used for the construction of the A-genome map, with the addition of some recently published markers, 75 newly developed microsatellite, 19 EST-STS markers and 11 strategically chosen anchor markers, which are single copy genic markers that are ideal for the alignment of genomes [32-34].

# Results

# Interspecific hybridization

Several crossings between *A. ipaënsis* and *A. magna* were made. Seven plants of *A. ipaënsis* (K30076) and six of *A. magna* (K30097) were used as female parents (see Additional file 1). A total of 993 flowers were cross-pollinated, of which 515 and 478 had *A. ipaënsis* and *A. magna* as female parents, respectively. A total of 556 viable seeds were obtained, being 313 (56%) from *A. ipaënsis* × *A. magna* crosses and 243 (44%) from *A. magna* × *A. ipaënsis* crosses. Hybrids were identified using the SSR marker Ah-282 visualized in 3% agarose gels. The number of seeds obtained from the 23 self-pollinated  $F_1$  individuals was high, ranging from 50 to 165, with an average of 92. The  $F_1$  plant obtained from cross 4 (see Additional file 1), which produced the highest number of seeds (165) was selected to generate the  $F_2$  mapping population.

# Marker development and analysis

### Genomic microsatellites

Forty primer pairs were developed using the three genomic libraries enriched for AC/TG and AG/TC repeats (see Additional file 2) and were screened against the progenitors of the mapping population. Repeats were, as expected, almost entirely composed of dinucleotides

(Table 1). Nine out of the 40 primer pairs (22.5%) were polymorphic, including one dominant marker (present in *A. ipaënsis* and absent in *A. magna*); seven (17.5%) were monomorphic; 13 (32.5%) did not amplify any fragment, and 11 (27.5%) did not allow precise analyses (Table 2).

A total of 556 genomic SSR markers (the 40 developed here plus 516 cited in literature) were tested against A. ipaënsis (K30076) and A. magna (K30097). Of these, 123 (22.1%) were polymorphic (including one dominant marker); 267 (48.0%) were monomorphic, and 166 (29.9%) did not amplify any interpretable fragment (Table 2).

# EST-SSR markers

Out of the 738 unique sequences obtained from the two *A. hypogaea* cDNA libraries enriched for expressed genes in response to *Cercosporidium personatum* [35], 61 (8.3%) presented SSRs with more than five repeats and 35 primer pairs could be designed (see Additional file 2). Frequencies of the SSR repeat types are shown in Table 1. Di- and trinucleotides were the most abundant repeats. Out of the 35 primer pairs screened against both progenitors, nine (25.7%) were polymorphic, 15 (42.9%) were monomorphic, six (17.1%) did not produced any amplification, and five (14.3%) resulted in low intensity or multiple-band patterns, and were excluded from the analyses (Table 2). The homologies between the sequences and genes are shown in Additional file 2.

Of the 189 EST-SSR markers screened against *A. ipaënsis* and *A. magna* (35 new plus 154 already published), only 17 (9.0%) did not amplify any product. A total of 43 EST-SSR markers (22.8%) were polymorphic, 106 (56.1%) were monomorphic, and 23 (12.1%) were excluded due to poor or confusing amplification patterns (Table 2).

# EST-STS markers

Nineteen primer pairs were designed from ESTs with homologies to plant genes involved in defense processes against biotic stress (see Additional file 2). Of these, two

Table 1: Characteristics of the n	newly developed markers
-----------------------------------	-------------------------

Repeat motif	Genomic SSR	EST-SSR
Dinucleotides	38 (95.0)	17 (48.6)
Trinucleotides	-	13 (37.1)
Tetranucleotides	-	I (2.9)
Di- and trinucleotides	l (2.5)	4 (11.4)
Di- and tetranucleotides	I (2.5)	-
Total	40	35

Number of the newly developed EST- and genomic SSR markers detected per repeat size class. Numbers in parentheses refer to the percentages of the total.

 Table 2: Polymorphism levels detected for the different markers.

	Genomic SSR	EST-SSR	EST-STS
New markers			
Polymorphic	9 (22.5%)	9 (25.7%)	2 (10.5%)
Monomorphic	7 (17.5%)	15 (42.9%)	10 (52.6%)
No amplification	13 (32.5%)	5 (14.3%)	I (5.3%)
Poor amplification	11 (27.5%)	6 (17.1%)	6 (31.6%)
Total	40	35	19
All markers			
Polymorphic	123 (22.1%)	43 (22.8%)	2 (10.5%)
Monomorphic	267 (48.0%)	106 (56.1%)	10 (52.6%)
No amplification	119 (21.4%)	17 (9.0%)	l (5.3%)
Poor amplification	47 (8.5%)	23 (12.1%)	6 (31.6%)
Total	556	189	19

Summary of the results obtained for the three types of markers detected after screening against the two BB genome species (A. *ipaënsis*, accession K30076 and A. *magna*, accession K30097) used as progenitors of the  $F_2$  mapping population.

detected polymorphism against both progenitors, ten were monomorphic, one did not amplify any product, and six resulted in low intensity or multiple band patterns, and were excluded from the analyses (Table 2).

## SNP markers

Ten anchor markers and one microsatellite distributed in six linkage groups of the AA map [28,36] were selected for mapping in the BB population. These selected markers were size monomorphic between the mapping parents as judged by electrophoresis in 4% polyacrylamide gel. The PCR products were sequenced and SNPs were identified for the 11 markers. In average, one SNP was identified per 200 bp, ranging from one SNP for every 42 bp to 627 bp. These markers were separated in two multiplex groups of five/six markers each and analyzed in the parents, the  $F_1$ hybrid and the  $F_2$  population.

### Genetic Mapping

A total of 745 SSR markers were evaluated, of which 166 (22.3%) were polymorphic between the parents. Using a minimum LOD score of 3.0 and a maximum recombination fraction of 0.35, 149 markers mapped into 10 linkage groups. These markers included 106 genomic SSRs, 32 EST-SSRs, two EST-STS, and nine anchor markers. The map covered a total distance of 1,294.4 cM (Figure 1). Groups ranged from 40.7 cM (5 markers) to 287.4 cM (31 markers), with an average distance of 8.7 cM between adjacent markers. Linkage groups were numbered according to the LG numbers of the AA genome map [28,36] by the identification of syntenic markers. Two SSR primer pairs amplified consistently two loci (RN9A05 and pPGSseq16C3) and these markers were identified by the numbers \_1 and \_2 after the marker names (Figure 1).



## Figure I

A linkage map for the B-genome of Arachis. Linkage map of Arachis based on an  $F_2$  population resultant from the cross A. *ipaënsis* × A. magna (B-genome). The map consists of 10 linkage groups and 149 codominant markers (genomic SSR, EST-SSR, STS, and SNPs). Distorted markers (P < 0.05) are identified by # after the loci names. Numbers on the left of each group are Kosambi map distances. Syntenic markers between the B- and A-genome maps [28,36] are indicated by colored blocks. Colors were assigned to the A-genome linkage groups so that syntenic LG are represented by corresponding colors.

Thirty-two markers (21.5%) out of the 149 mapped markers showed deviation from the expected 1:2:1 ratio, being 24 at P < 0.05 and eight at P < 0.01. Of these, 12 markers were skewed towards *A. magna*, three markers towards *A. ipaënsis*, and 17 towards the heterozygote. Linkage groups B2 and B10 had all distorted markers with an excess of *A. magna* alleles, while LGs B1, B4, and B7 had all distorted markers skewed towards the heterozygote. The three markers with an excess of *A. ipaënsis* alleles grouped on LGs B3, B5 and B8 that also had markers with an excess of *A. magna* alleles and towards the heterozygote. Distorted markers at P < 0.05 were identified by # (Figure 1). Groups B6 and B9 had no distorted markers.

### Synteny analysis

A total of 51 common markers mapped in the AA and BB genome diploid maps spanned the 10 linkage groups of both maps (Figure 1). Seven LGs of the BB map (B1, B2,

B3, B4, B5, B8, and B9) showed direct correspondences with seven groups of the AA map. Of these, five had all common markers mapped in the same order. From two (LG B8) to 11 (LG B3) collinear loci were identified per linkage group. The groups B2 and B10 showed common loci to group A2, and two segmental inversions were apparent (see Additional file 3). Group B2 was syntenic to the upper region of LG A2 with five collinear loci, and the group B10 in the lower region. Inversions were also detected in the LGs B1/A1 and B6/A6. Linkage groups B6 and B7 showed split syntenic relationships, with common markers mapping in two LG of the AA map, B6 with A6 and A10, and B7 with A7 and A8.

### Discussion

This linkage map was obtained using an  $F_2$  population derived from a cross between *A. ipaënsis* and *A. magna*. Several lines of evidence indicate that *A. ipaënsis* is the

most probable donor of the B-genome to A. hypogaea [6,13,14,37,38]. Arachis magna is also a B-genome species closely related to A. ipaënsis, as indicated by crossability data [3], high rates of pollen viability in hybrids [39], and molecular marker analyses [17,19,20,40]. The high fertility of the crosses and low polymorphism levels between the species (22.3% of SSR markers) observed here support this close relationship, and indeed even suggest that the two names could actually correspond to a single biological species. Further studies should be carried out to check this hypothesis, as it might have important implications for the incorporation of new wild alleles in cultivated peanut: so far there are many collected accessions of A. magna and only one available accession of A. ipaënsis. However, regardless the taxonomic status of the species, it is clear that both genomes used to construct the map are similar to the B-genome of A. hypogaea and that the linkage map is probably a good representation of it.

The DNA polymorphism within this population is lower than the populations used for the construction of previously published Arachis maps: 51% for RFLP probes in the A. stenosperma × A. cardenasii derived population [25]; 40% for RFLP probes in the Arachis hypogaea × synthetic amphidiploid {A. batizocoi × (A. cardenasii × A. diogoi)} $^{4\times}$ population [26]; and 47% for SSR markers in the A. duranensis × A. stenosperma derived population [28]. This low polymorphism has been compensated by the large number of SSR markers developed for Arachis over the past few years [19,20,28,40-45], which has enabled the development of this linkage map. On the other hand, the segregation distortion of 21.5% is in the same range as the distortion found in many intraspecific maps [46-48]. Linkage groups B2 and B10 had all distorted markers with an excess of A. magna alleles, while LG B1, B4, and B7 had all distorted markers skewed towards the heterozygote. These groupings of distorted markers suggest that some regions of the chromosome are more prone to segregation distortion, rather than the distortion being marker-specific.

All markers evaluated in this study were amplified using heterologous primers. Most of them were developed for A. hypogaea and A. stenosperma, and 74 markers were developed for species from other sections of the Arachis genus (50 primer pairs for A. pintoi of section Caulorrhizae and 24 for A. glabrata of section Rhizomatosae), confirming the high transferability of SSR markers within the Arachis genus. From 745 markers tested, 609 (81.7%) allowed the amplification of PCR products in A. ipaënsis and/or A. magna. As expected, the level of transferability varied among the different types of primers tested. Microsatellites based on expressed genic regions (EST-SSR and STSs) showed higher transferability levels (91.0% and 94.7%, respectively) than random genomic microsatellites (78.6%). This confirms previous findings that markers based on cDNA sequences are more transferable among species than random markers, such as genomic SSRs, since they are based on coding regions, which are generally more conserved that non coding regions [49-54].

The number of repeats found in the genomic microsatellite markers was, in general, higher (5 to 64 repeats) than the number in expressed genic microsatellites (5 to 16 repeats). This difference was not reflected in the polymorphism levels found for these two sources of primers: 22.8% of the EST-SSRs and 22.0% of the genomic SSRs. These findings are in agreement with our previous results for wild species and contrasts with cultivated peanut, where longer microsatellites have higher polymorphism [28].

The present map comprised 10 linkage groups, with 149 loci spanning a total map distance of 1,294.4 cM, which corresponds to the haploid chromosome number of the progenitor species n = 10 [3]. The total length obtained is similar to the sizes described for the other two co-dominant marker-based linkage maps published for diploid species of *Arachis*: 1,063 cM for an RFLP based map developed using an *A. stenosperma* × *A. cardenasii* cross [25] and 1,230.9 cM found for a microsatellite based map developed using an *A. duranensis* × *A. stenosperma* cross [28]. This size is also comparable to half of the 2,210.0 cM found for a published tetraploid map for *Arachis* spp. [26]. However, seventeen (10.2%) of the 166 segregating markers remained unlinked, suggesting that at least parts of the genome have not been covered by this map.

Twenty five percent of the mapped markers were developed from cDNA libraries (33 EST-SSR and two STS markers). Some of them had similarity to genes of known function, including genes involved in the photosynthesis process and in responses to biotic stresses. For instance, marker AHBGSD1002H08 (LG B8) showed similarity to a tissue specific gene coding for a prolin-rich protein of soybean (E-value =  $3.0 \times 10^{-27}$ ), that has the expression induced by salicylic acid, virus infection, circadian rhythm and salinic and drought stresses, indicating this gene may have an important role in the response to multiple internal and external factors [55]. Marker AHBGST1002B04 showed similarity to dihiydro-isoflavone redutase (Evalue =  $3.0 \times 10^{-57}$ ), that is an enzyme involved in the synthesis of different flavonoids, and some of them, such as flavones and the 3-deoxyanthocyanidina, are involved in the plant defense process [56]. Linkage maps that contain genic markers can facilitate the finding of genes of interest, as ESTs mapping in regions with QTLs are good candidates to be involved in the trait and being an alternative to positional cloning [47,57].

A total of 42 microsatellite markers in common with the A-genome map [28] were placed on this B-genome map. In order to increase the number of shared markers, nine anchor markers [32-34] selected from the A-map [36] were placed on the B-map using SNPs. The comparison of the 51 shared markers revealed associations between maps and apparently high levels of synteny, since all but one of the B linkage groups show single main correspondences to the A-map. This seems largely consistent with the observed for homeologous groups in the published tetraploid map of *Arachis* [26] with perhaps the main differences being: in the tetraploid study, one large B linkage group shows no marker correspondences to the A genome, whilst in this study no "orphan" linkage groups are present; and in this study two B linkage groups correspond to one A (B2 and B10 to A2), a situation not observed in the tetraploid map.

The integration of the A- and B-genome *Arachis* maps effectively increases the information content of both maps. The A-genome map contains candidate genes and QTLs for disease resistance, and has been aligned with the genomes of the model legumes *Lotus* and *Medicago* and

with the bean genetic map [36,58]. Much of this information is likely to be transferable to the B-map. As an example, Figure 2 shows an alignment of the B-map through the A-map with *Lotus*, whose genome sequence was recently published [59]. This type of alignment allows the inference of the position of candidate genes from a whole genome sequence on the B-genome map.

## Conclusion

Here we present a microsatellite-based map for the Bgenome of *Arachis* and its integration with an A-genome map. The development of these maps, based on markers that are highly transferable and simple to use will facilitate the identification and introgression of useful genes from both A-type and B-type wild genomes into cultivated peanut. These maps will also be used as reference for future cultivated peanut maps and for the development of introgression lines which are underway. Both the B-genome population described here and the A-genome population [28], have now been developed into  $F_5$  RIL (Recombinant



### Figure 2

**An example of synteny between A- and B- genomes of** *Arachis* and *Medicago*. Alignment of linkage group B3 of the developed map with the A-genome (LG A3) and *Medicago truncatula* (LG Mt4 and Mt7).

Inbred Lines) populations which will facilitate the even broader use of these map and marker resources.

# Methods

# Plant material

The  $F_2$  population composed of 93 plants was obtained by selfing a unique  $F_1$  plant derived from a cross between *A. ipaënsis* (accession K30076), used as the female parent, and *A. magna* (K30097), used as the male. Accession K30097 is the holotype of *A. magna*, while K30076 originate from the same collection site of the type specimen of *A. ipaënsis* [3,4]. Plants were obtained from the Brazilian *Arachis* germplasm collection, maintained at Embrapa Genetic Resources and Biotechnology – CENARGEN (Brasília-DF, Brazil).

# DNA extraction

Total genomic DNA was extracted from young leaflets essentially as described by Grattapaglia & Sederoff (1994) [60]. The quality and quantity of the DNA were evaluated in 1% agarose gel electrophoresis and spectrophotometer (Genesys 4 – Spectronic).

# Marker development and analysis

The same set of microsatellite markers used in Moretzsohn et al., 2005 [28] was used for screening for polymorphism between the parents. In addition, some markers recently published [44,45] were used, as well as the newly developed one, as follows:

Development of genomic DNA libraries enriched for microsatellites Three libraries were developed using genomic DNA iso-

lated from leaves of A. hypogaea (section Arachis), A. glabrata (section Rhizomatosae) and A. pintoi (section Caulorrhizae). For each library, about nine micrograms of DNA were digested with Sau3AI (Amersham Biosciences, UK) and electrophoresed in 0.8% low melting agarose gels to select fragments ranging from 200 to 600 bp. The selected fragments were purified from the agarose gels using phenol/chloroform, and ligated into Sau3AI specific adaptors (5'-cagcctagagccgaattcacc-3' and 5'-gatcggtgaaatcggctcaggctg-3'). The ligated fragments were hybridized to biotinylated (AC)<sub>15</sub> and (AG)<sub>15</sub> oligonucleotides and isolated using streptavidin-coated magnetic beads (Dynabeads Streptavidin, Dynal Biotech, Norway). The eluted fragments were amplified using one adaptor-specific primer, cloned into the pGEM-T Easy vector (Promega, WI, USA) and transformed into DH5α E. coli cells with blue/white selection (Invitrogen, CA, USA). Plasmid DNAs of the positive clones were isolated using the 'CONCERT Rapid Plasmid Purification Miniprep System', as described by the manufacturer (Invitrogen, CA, USA) and sequenced with an ABI Prism 377 automated sequencer using the 'BigDye Terminator Cycle Sequencing Kit', version 3.1 (Applied Biosystems, CA, USA).

# EST-SSR and EST-STS marker development

EST-SSRs were developed from 883 EST sequences obtained from a recently constructed Suppression Subtractive Hybridization - SSH library of A. hypogaea enriched for expressed genes in response to Cercosporidium personatum [35] using the software described below. In addition, 14 A. hypogaea ESTs were selected due to their similarity to genes involved in defense mechanisms, identified using BlastX analyses [61]. From these, 12 sequences had no SSR repeats, but were used for primer design to develop STS (Sequence tagged sites) markers. Primers were also designed for an EST of unknown function (AHBGSI1002C10), for a sequence similar to a dienelactone hydrolase family protein of Arabidopsis thaliana (AHBGSI1006D06) and for three ESTs of putative intron adjacent sequences (AHBSI1001D05-I1, AHBSI1002C11-I1 and AHBSI1009D07-I2) that were selected using an unpublished software developed by Dr. Wellington Martins, Universidade Católica de Goiás, Brazil.

# Primer design

Sequences were processed and assembled by using the Staden package [62] with the repeat sequence finding module TROLL [63] and Primer3 [64]. Sequences with more than five motif repeats were chosen for primer design. The parameters for primer design were: (1) primer size ranging from 18 bp to 25 bp with an optimal length of 20 bp; (2) primer  $T_{\rm m}$  (melting temperature) ranging from 57 °C to 63 °C with an optimal temperature of 60 °C; and (3) GC content ranging from 40% to 60%. Default values were used for the other parameters.

# PCR amplifications

PCR reactions contained 5 ng of genomic DNA, 1 U of Taq DNA polymerase (Amersham Biosciences), 1× PCR buffer (200 mM Tris pH 8.4, 500 mM KCl), 1.5-2.0 mM MgCl<sub>2</sub>, 200  $\mu$ M of each dNTP, and 0.4  $\mu$ M of each primer, in a final reaction volume of 10 µl. Amplifications were carried out in a PTC100 thermocycler (MJ Research Inc., MA, USA). PCR conditions were: 96°C for 5 min, followed by 32 cycles of 96°C for 30 s, 48-62°C (annealing temperature depending on primer pair, see Additional file 2) for 45 s, 72°C for 1 min, with a final extension for 10 min at 72°C. PCR products were separated by electrophoresis on denaturing polyacrylamide gels (6% acrylamide:bisacrylamide 29:1, 5 M urea in TBE pH 8.3), stained with silver nitrate [65]. Some SSR markers highly contrasting between the progenitors of the mapping population were run on 3% agarose Metaphor (FMC Bioproducts, PA, USA) gels stained with ethidium bromide.

# SNPs identification and analysis

Ten anchor markers and one microsatellite distributed in six linkage groups of the AA map [28,36] were selected for mapping in the BB population. Markers from A-genome linkage groups that had few markers in common with an initial version of the B-map were preferentially chosen. The identification of SNPs and single base extension (SNaPshot) analysis was performed essentially as described by Alves et al. (2008) [66]. Primers were designed using the program Primo SNP 3.4, available at http://www.changbioscience.com/primo/primosnp.html (Chang Bioscience). The SNP in the consensus sequence of both progenitors was replaced by a degenerated IUPAC code for primer design. Non-homologous polynucleotides (dGACT)<sub>n</sub> were added to the 5'-end of each primer to enable the analysis in multiplexes (see Additional file 2), using the commercial system ABI PRISM<sup>®</sup> SNaPshot<sup>™</sup> Multiplex Kit (Applied Biosystems). Absence of hairpins and self-complementarity of all SNP primers were checked by the software Autodimer [67].

## Map construction

A total of 745 SSR, 19 STS and 11 SNP markers were screened against the two progenitors of the mapping population. These included the 105 newly developed markers (see Additional file 2) plus another 670 published microsatellite markers [19,20,28,40-45,68-70]. Polymorphic markers were analyzed on the mapping population consisting of 93 F<sub>2</sub> individuals. A  $\chi^2$  test was performed to test the null hypothesis of 1:2:1 segregation on all scored markers. The linkage analysis was done using Mapmaker Macintosh version 2.0 [71]. A minimum LOD score of 4.0 and maximum recombination fraction ( $\theta$ ) of 0.35 were set as thresholds for linkage groups determination with the "group" command. The most likely marker order within each LG was estimated by the matrix correlation method using the "first order" command. Marker orders were confirmed by comparing the log-likelihood of the possible orders using multipoint analysis ("compare" command) and by permuting all adjacent triple orders ("ripple" command). After establishment of the group orders, the LOD score was set to 3.0 in order to include additional markers in the groups. The "try" command was then used to determine the exact position of the new markers within each group. The new marker orders were again confirmed with the "first order", "compare", and/or "ripple" commands. Recombination fractions were converted into map distances in centimorgans (cM) using the Kosambi's mapping function.

# **Authors' contributions**

All authors read and approved the final manuscript. MCM carried out the analysis for genetic map construction, participated in the synteny analysis and drafted the manuscript. AVGB carried out the mapping population construction, participated in the development and analysis of SSR and STS markers and drafting the manuscript. DMTAF carried out the identification and analysis of SNP markers. CT and MMC participated in SSR and STS markers analysis. SCMLB and PMG participated in the SSR and synteny analyses. RWP coordinated the identification and analysis of SNP markers. CRL participated in conceiving the study. JV participated in the conception of the project and provided the germplasm. DJB participated in SSR, STS and SNP development and analysis, carried out the synteny analysis and participated in drafting the manuscript. MAG participated in conceiving the study, coordinated the SSR and STS markers development and analysis, and participated in drafting the manuscript.

# **Additional material**

## Additional File 1

Data of crossings between A. ipaënsis (accession K30076) and A. magna (K30097). The data provides the number of viable seeds obtained by crossing A. ipaënsis (accession K30076) and A. magna (K30097) and by selfing  $F_1$  hybrid individuals. Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2229-9-40-S1.doc]

# Additional File 2

**Features of the newly developed markers**. The data provides the details of the new set of markers, being 40 genomic SSR, 35 EST-SSR, 19 STS, *and 11 SNPs*. Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2229-9-40-S2.xls]

# Additional File 3

Relationships between the 10 linkage groups of the A- and B-genome maps. The data provides the affinities between the A- and B-genome linkage maps of Arachis. Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2229-9-40-S3.ppt]

# Acknowledgements

This work was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), PRODETAB Project number 004-01/01, and the Generation Challenge Program Projects G3005.05 and TLI.

### References

- I. Hammons RO: **The origin and history of the groundnut.** Chapman and Hall, London; 1994.
- 2. Holbrook CC, Stalker HT: Peanut Breeding and Genetic Resources. Plant Breeding Reviews 2003, 22:297-355.
- Krapovickas A, Gregory WC: Taxonomia del genero Arachis (Leguminosae). Bonplandia (Argentina) 1994, 8(1-4):1-186.
- Valls JFM, Simpson CE: New species of Arachis L. (Leguminosae) from Brazil, Paraguay and Bolivia. Bonplandia (Argentina) 2005, 14:35-64.
- 5. Husted L: Cytological studies on the peanut, *Arachis*. II. Chromosome number, morphology and behavior, and their application to the problem of the origin of the cultivated forms. *Cytologia* 1936, **7**:396-422.
- Seijo JG, Lavia GI, Fernandez A, Krapovickas A, Ducasse D, Moscone EA: Physical mapping of the 5S and 18S-25S rRNA genes by FISH as evidence that Arachis duranensis and A. ipaënsis are

the wild diploid progenitors of A. hypogaea (Leguminosae). American Journal of Botany 2004, 91(9):1294-1303.

- 7. Stalker HT: A New Species in Section Arachis of Peanuts with a D-Genome. American Journal of Botany 1991, 78(5):630-637.
- Robledo G, Seijo G: Characterization of the Arachis (Leguminosae) D genome using fluorescence in situ hybridization (FISH) chromosome markers and total genome DNA hybridization. Genetics and Molecular Biology 2008, 31(3):717-724.
- Lavia GI: Karyotypes of Arachis palustris and A. praecox (Section Arachis), two species with basic chromosome number x = 9. Cytologia 1998, 63(2):177-181.
- Lavia Gl, Fernández A: Genome size in wild and cultivated peanut germplasm. Plant Systematics and Evolution 2008, 272(1):1-10.
- Lavia G, Fernández A, Seijo J: Cytogenetic and molecular evidences on the evolutionary relationships among Arachis species. In Plant Genome: Biodiversity and Evolution, Phanerogams-Angiosperm Volume IE. Edited by: Sharma A. Enfield, NH: Science Publishers; 2008:101-134.
- 12. Halward TM, Stalker HT, Larue EA, Kochert G: Genetic variation detectable with molecular markers among unadapted germplasm resources of cultivated peanut and related wild species. *Genome* 1991, **34(6):**1013-1020.
- Kochert G, Stalker HT, Gimenes M, Galgaro L, Lopes CR, Moore K: RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, Arachis hypogaea (Leguminosae). American Journal of Botany 1996, 83(10):1282-1291.
- 14. Seijo G, Lavia GI, Fernandez A, Krapovickas A, Ducasse DA, Bertioli DJ, Moscone EA: Genomic relationships between the cultivated peanut (Arachis hypogaea, Leguminosae) and its close relatives revealed by double GISH. American Journal of Botany 2007, 94(12):1963-1971.
- Subramanian V, Gurtu S, Rao RCN, Nigam SN: Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. Genome 2000, 43(4):656-660.
- Herselman L: Genetic variation among Southern African cultivated peanut (Arachis hypogaea L.) genotypes as revealed by AFLP analysis. Euphytica 2003, 133(3):319-327.
- 17. Milla SR, Isleib TG, Stalker HT: Taxonomic relationshipsamong Arachis sect. Arachis species as revealed by AFLP markers. Genome 2005, 48(1):1-11.
- Hilu KW, Stalker HT: Genetic relationships between peanut and wild species of Arachis sect Arachis (Fabaceae): Evidence from RAPDs. Plant Systematics and Evolution 1995, 198(3– 4):167-178.
- Moretzsohn M, Hopkins M, Mitchell S, Kresovich S, Valls J, Ferreira M: Genetic diversity of peanut (Arachis hypogaea L.) and its wild relatives based on the analysis of hypervariable regions of the genome. BMC Plant Biology 2004, 4(1):11.
- Bravo JP, Hoshino AA, Angelici CMLCD, Lopes CR, Gimenes MA: Transferability and use of microsatellite markers for the genetic analysis of the germplasm of some Arachis section species of the genus Arachis. Genetics and Molecular Biology 2006, 29(3):516-524.
- 21. Kameswara Rao N, Reddy LJ, Bramel PJ: **Potential of wild species** for genetic enhancement of some semi-arid food crops. *Genetic Resources and Crop Evolution* 2003, **50(7)**:707-721.
- Dwivedi S, Bertioli DJ, Crouch JH, Valls JFM, Upadhyaya HD, Favero AP, Moretzsohn MC, Paterson AH: Peanut Genetics and Genomics: Toward Marker-assisted Genetic Enhancement in Peanut (Arachis hypogaea L). In Oilseeds Series: Genome Mapping and Molecular Breeding in Plants Volume 2. Berlim, Heidelberg: Springer; 2007:115-151.
- Stalker HT, Simpson CE: Germplasm resources in Arachis. In Advances in Peanut Science Edited by: Pattee HE, Stalker HT. Stillwater, OK: American Peanut Research and Education Society, Inc; 1995:14-53.
- Varshney RK, Bertioli DJ, Moretzsohn MC, Vadez V, Krishnamurthy L, Aruna R, Nigam SN, Moss BJ, Seetha K, Ravi K, et al.: The first SSR-based genetic linkage map for cultivated groundnut (Arachis hypogaea L.). Theoretical and Applied Genetics 2009, 118:729-739.
- Halward T, Stalker HT, Kochert G: Development of an RFLP Linkage Map in Diploid Peanut Species. Theoretical and Applied Genetics 1993, 87(3):379-384.

- Burow MD, Simpson CE, Starr JL, Paterson AH: Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (Arachis hypogaea L.): Broadening the gene pool of a monophyletic polyploid species. Genetics 2001, 159(2):823-837.
- 27. Garcia GM, Stalker HT, Shroeder E, Lyerly JH, Kochert G: A RAPDbased linkage map of peanut based on a backcross population between the two diploid species Arachis stenosperma and A. cardenasii. Peanut Science 2005, 32:1-8.
- Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SCM, Gimenes MA, Martins WS, Valls JFM, Grattapaglia D, Bertioli DJ: A microsatellite-based, gene-rich linkage map for the AA genome of Arachis (Fabaceae). Theor Appl Genet 2005, 111(6):1060-1071.
- Choumane W, Winter P, Baum M, Kahl G: Conservation of microsatellite flanking sequences in different taxa of Leguminosae. Euphytica 2004, 138(3):239-245.
- Gutierrez MV, Patto MCV, Huguet T, Cubero JI, Moreno MT, Torres AM: Cross-species amplification of Medicago truncatula microsatellites across three major pulse crops. Theor Appl Genet 2005, 110(7):1210-1217.
- Katzir N, DaninPoleg Y, Tzuri G, Karchi Z, Lavi U, Cregan PB: Length polymorphism and homologies of microsatellites in several Cucurbitaceae species. Theoretical and Applied Genetics 1996, 93(8):1282-1290.
- Fredslund J, Schauser L, Madsen LH, Sandal N, Stougaard J: PriFi: using a multiple alignment of related sequences to find primers for amplification of homologs. Nucleic Acids Research 2005, 33:W516-W520.
- Fredslund J, Madsen LH, Hougaard BK, Sandal N, Stougaard J, Bertioli D, Schauser L: GeMprospector – online design of cross-species genetic marker candidates in legumes and grasses. Nucleic Acids Research 2006, 34:W670-W675.
- Fredslund J, Madsen LH, Hougaard BK, Nielsen AM, Bertioli D, Sandal N, Stougaard J, Schauser L: A general pipeline for the development of anchor markers for comparative genomics in plants. BMC Genomics 2006, 7:207.
- Nobile PM, Lopes CR, Barsalobres-Cavallari C, Quecim V, Coutinho LL, Hoshino AA, Gimenes MA: Peanut genes identified during initial phase of Cercosporidium personatum infection. Plant Science 2008, 174(1):78-87.
- 36. Bertioli D, Moretzsohn M, Madsen LH, Sandal N, Leal-Bertioli S, Guimarães P, Hougaard BK, Fredslund J, Schauser L, Nielsen AM, et al.: A comparison of genome synteny of Arachis with the model legumes Lotus japonicus and Medicago truncatula reveals a new feature of legume genomes. BMC Genomics 2009, 10:45.
- Raina SN, Rani V, Kojima T, Ogihara Y, Singh KP, Devarumath RM: RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (Arachis hypogaea) cultivars and wild species. Genome 2001, 44(5):763-772.
- Fávero AP, Simpson CE, Valls JFM, Vello NA: Study of the evolution of cultivated peanut through crossability studies among Arachis ipaënsis, A.duranensis, and A.hypogaea. Crop Science 2006, 46(4):1546-1552.
- Simpson ČÉ, Faries MJ: Advances in the characterization of diversity in section Arachis: archeological evidence, crossing results and their relationship in understanding the origins of Arachis hypogaea L. In Annals of the III SIRGEALC – Simpósio de Recursos Genéticos para a América Latina e Caribe Londrina, Paraná, Brazil; 2001:706.
- Gimenes MA, Hoshino AA, Barbosa AVG, Palmieri DA, Lopes CR: Characterization and transferability of microsatellite markers of the cultivated peanut (Arachis hypogaea). BMC Plant Biology 2007, 7:9.
- He G, Meng R, Newman M, Gao G, Pittman R, Prakash CS: Microsatellites as DNA markers in cultivated peanut (Arachis hypogaea L.). BMC Plant Biology 2003, 3(1):3.
- He GH, Meng RH, Gao H, Guo BZ, Gao GQ, Newman M, Pittman RN, Prakash CS: Simple sequence repeat markers for botanical varieties of cultivated peanut (Arachis hypogaea L.). Euphytica 2005, 142(1-2):131-136.
- Ferguson ME, Burow MD, Schulze SR, Bramel PJ, Paterson AH, Kresovich S, Mitchell S: Microsatellite identification and characterization in peanut (A. hypogaea L.). Theor Appl Genet 2004, 108(6):1064-1070.

- Proite K, Leal-Bertioli SCM, Bertioli DJ, Moretzsohn MC, da Silva FR, Martins NF, Guimaraes PM: ESTs from a wild Arachis species for gene discovery and marker development. BMC Plant Biology 2007, 7:7.
- Cuc LM, Mace ES, Crouch JH, Quang VD, Long TD, Varshney RK: Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (Arachis hypogaea). BMC Plant Biology 2008, 8:55.
- Sibov ST, De Souza CL, Garcia AAF, Garcia AF, Silva AR, Mangolin CA, Benchimol LL, De Souza AP: Molecular mapping in tropical maize (Zea mays L.) using microsatellite markers. I. Map construction and localization of loci showing distorted segregation. Hereditas 2003, 139(2):96-106.
- Flandez-Galvez H, Ford R, Pang EC, Taylor PW: An intraspecific linkage map of the chickpea (*Cicer arietinum* L.) genome based on sequence tagged microsatellite site and resistance gene analog markers. *Theor Appl Genet* 2003, 106(8):1447-1456.
- gene analog markers. Theor Appl Genet 2003, 106(8):1447-1456.
  48. Truco MJ, Antonise R, Lavelle D, Ochoa O, Kozik A, Witsenboer H, Fort SB, Jeuken MJW, Kesseli RV, Lindhout P, et al.: A high-density, integrated genetic linkage map of lettuce (Lactuca spp.). Theor Appl Genet 2007, 115(6):735-746.
- Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JMM, Echt C, et al.: Crossspecies transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 2004, 109(6):1204-1214.
- Gupta PK, Varshney RK: The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 2000, 113(3):163-185.
   Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS: Trans-
- Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS: Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Genet Genomics* 2003, 270(4):315-323.
- Thiel T, Michalek W, Varshney RK, Graner A: Exploiting EST databases for the development and characterization of genederived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet 2003, 106(3):411-422.
- 53. Varshney RK, Sigmund R, Borner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A: Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Science 2005, 168(1):195-202.
- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A, Hamon P: SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the Coffea genus. Mol Genet Genomics 2006, 276(5):436-449.
- 55. He CY, Zhang JS, Chen SY: A soybean gene encoding a prolinerich protein is regulated by salicylic acid, an endogenous circadian rhythm and by various stresses. Theor Appl Genet 2002, 104(6-7):1125-1131.
- Winkel-Shirley B: Biosynthesis of flavonoids and effects of stress. Current Opinion in Plant Biology 2002, 5(3):218-223.
- Zheng BS, Yang L, Zhang WP, Mao CZ, Wu YR, Yi KK, Liu FY, Wu P: Mapping QTLs and candidate genes for rice root traits under different water-supply conditions and comparative analysis across three populations. Theor Appl Genet 2003, 107(8):1505-1515.
- Hougaard BK, Madsen LH, Sandal N, Moretzsohn MD, Fredslund J, Schauser L, Nielsen AM, Rohde T, Sato S, Tabata S, et al.: Legume anchor markers link syntenic regions between Phaseolus vulgaris, Lotus japonicus, Medicago truncatula and Arachis. Genetics 2008, 179(4):2299-2312.
- Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, et al.: Genome Structure of the Legume, Lotus japonicus. DNA Research 2008, 15:227-239.
- Grattapaglia D, Sederoff R: Genetic Linkage Maps of Eucalyptus grandis and Eucalyptus urophylla Using a Pseudo-Testcross: Mapping Strategy and RAPD Markers. Genetics 1994, 137(4):1121-1137.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic Local Alignment Search Tool. Journal of Molecular Biology 1990, 215(3):403-410.
- Staden R: The Staden sequence analysis package. Molecular Biotechnology 1996, 5(3):233-241.

- Martins W, de Sousa D, Proite K, Guimaraes P, Moretzsohn M, Bertioli D: New softwares for automated microsatellite marker development. Nucleic Acids Research 2006, 34(4):e31.
- 64. Castelo AT, Martins W, Gao GR: TROLL-Ťandem Repeat Occurrence Locator. Bioinformatics 2002, 18(4):634-636.
- Creste S, Neto AT, Figueira A: Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Molecular Biology Reporter 2001, 19(4):299-306.
- Alves DM, Pereira RW, Leal-Bertioli SC, Moretzsohn MC, Guimaraes PM, Bertioli DJ: Development and use of single nucleotide polymorphism markers for candidate resistance genes in wild peanuts (Arachis spp). Genet Mol Res 2008, 7(3):631-642.
   Vallone PM, Butler JM: AutoDimer: a screening tool for primer-
- Vallone PM, Butler JM: AutoDimer: a screening tool for primerdimer and hairpin structures. Biotechniques 2004, 37(2):226-231.
- Hopkins MS, Casa AM, Wang T, Mitchell SE, Dean RE, Kochert GD, Kresovich S: Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. Crop Science 1999, 39(4):1243-1247.
- Palmeri DA, Hoshino AA, Bravo JP, Lopes CR, Gimenes MA: Isolation and characterization of microsatellite loci from the forage species Arachis pintoi (Genus Arachis). Molecular Ecology Notes 2002, 2(4):551-553.
- Palmieri DA, Bechara MD, Curi RA, Gimenes MA, Lopes CR: Novel polymorphic microsatellite markers in section Caulorrhizae (Arachis, Fabaceae). Molecular Ecology Notes 2005, 5(1):77-79.
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L: MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1987, 1(2):174-181.

