Database

# PeanutMap: an online genome database for comparative molecular maps of peanut

Arun M Jesubatham[1] and Mark D Burow*[2,3]

Address: [1]Department of Computer Science, Texas Tech University, Lubbock, TX 79409, USA, [2]Texas Agricultural Experiment Station, Texas A&M University, 1102 East FM 1294, Lubbock, Texas 79403, USA and [3]Department of Plant and Soil Science, Texas Tech University, Lubbock, Texas 79409, USA

Email: Arun M Jesubatham - arunmoses99@yahoo.com; Mark D Burow* - mburow@tamu.edu

* Corresponding author

## Abstract

**Background:** Molecular maps have been developed for many species, and are of particular importance for varietal development and comparative genomics. However, despite the existence of multiple sets of linkage maps, databases of these data are lacking for many species, including peanut.

**Description:** PeanutMap http://peanutgenetics.tamu.edu/cmap provides a web-based interface for viewing specific linkage groups of a map set. PeanutMap can display and compare multiple maps of a set based upon marker or trait correspondences, which is particularly important as cultivated peanut is a disomic tetraploid. The database can also compare linkage groups among multiple map sets, allowing identification of corresponding linkage groups from results of different research projects. Data from the two published peanut genome map sets, and also from three maps sets of phenotypic traits are present in the database. Data from PeanutMap have been incorporated into the Legume Information System website http://www.comparative-legumes.org to allow peanut map data to be used for cross-species comparisons.

**Conclusion:** The utility of the database is expected to increase as several SSR-based maps are being developed currently, and expanded efforts for comparative mapping of legumes are underway. Optimal use of these data will benefit from the development of tools to facilitate comparative analysis.

## Background

Molecular maps are an important part of the genomics revolution. The use of DNA markers has expanded our knowledge of genetic linkage relationships considerably by removing the need for linkage to phenotypic markers. DNA-based maps were first developed in the 1980s [1], and have since expanded to encompass thousands of markers in some species [2-4]. Two RFLP-based maps of peanut have been published [5,6], and the recent develop-

ment of SSRs for peanut [7-10] is expected to result in rapid generation of additional maps.

Genetic markers have been developed for selection of qualitative traits and QTLs in multiple species. In peanut, DNA-based markers have been developed for nematode resistance, and were used for selection during of the final two generations of development of the variety 'NemaTAM' [11]. In addition to being useful in selection

programs, markers are useful for identification of biological relationships among accessions and species, and are an integral part of gene isolation by positional cloning [12] and ordered gene sequencing [13].

One of the most-important insights from genomics is the discovery of synteny among species of the same botanical family and, to a lesser extent, among different families. Comparative maps using genetic markers have demonstrated considerable similarity in chromosome structure and gene order among species in the same botanical family, with Poaceae being a striking example [14,15]. Marker analysis of several legume species has also provided evidence for the conservation of gene order in Fabaceae [16]. Conservation of gene order has been used in attempts to clone genes based on comparative information between species [17], and given insights into genome structure.

One of the limitations of genomics has been the lack of informatics resources for analysis of the large amounts of data produced, and for comparison of data among species. Databases exist for species of interest as model systems or having the greatest economic value, but are lacking in other species, including peanut. Among legumes, genome databases exist only for *Glycine* [18] and *Medicago* [19]. As awareness of the significance of comparative genomics has increased, there is a trend towards databases encompassing data from multiple species. Gramene has incorporated data from maize and rice [20], and a cross-legume database, called the Legume Information System, is currently being developed to combine data from species-specific databases and permit cross-legume comparisons [21].

A large amount of the older species-specific data were held in either the AceDB-type databases [22] or proprietary databases. AceDB had the advantage that it was a database developed specifically for genomic data; however, the lack of structured query language and tools for comparative maps have made this type of database problematic for comparative genomics. Proprietary databases make use of commercial software, such as Oracle [23]; such databases are very powerful, but software licencing is prohibitively expensive for all but the largest research projects, and the software lacks built-in tools for genomics. Recently, the USDA and NIH have co-sponsored development of the GMOD [24] suite of open-source programs for genomics. This software runs on open-source databases such as MySQL [25] or PostGresSQL [26], and on multiple operating systems, including the open-source Linux operating system. Parts of the Gramene [20] and LIS [21] databases are utilizing or migrating to the GMOD-based software. The component of this software for maps is called CMAP [27].

One hindrance to the advancement of peanut genomics has been the lack of a genome database. Development of such a database would assist in the dissemination of genomic data, accelerate genomic research and varietal development, and foster comparative genomics with other legumes. In this paper, we present a new map database for peanut, called PeanutMap.

## Construction and content

The installation of PeanutMap was done on a PC with SCSI hard drives set up for Ultra 160 RAID 1 mirroring, and running the Redhat Linux Advanced Server v. 3.0 operating system [28]. The following software was installed before installing CMAP, with current versions listed after the software: libgd 2.0 [29], MySQL database v.4.0.20 [25], Perl v.5.8.0 [30], CPAN modules v. 1.390 [31], and Apache v. 2.0.46 [32]. Libgd is the C graphics library used by CMAP. The MySQL structured query language database was installed as binaries, and serves as the relational database system for data storage and retrieval. The Perl programming language is needed for execution of CMAP. The CPAN shell was installed and used to download all the required CPAN Perl modules. The CPAN shell was allowed to check automatically for dependencies among Perl modules; however, modules that the CPAN shell failed to install were installed manually. The Apache Web Server was installed as binaries, and scripts that came packaged with Apache are used to start and stop the web server.

CMAP version 0.10 was installed as the core of the PeanutMap system; CMAP is a cgi (computer gateway interface) application written entirely in Perl The CMAP software [27] is open source, originally written for the Gramene project and is now part of GMOD [24]. The locations for the cache, templates, and html documents were specified in the CMAP configuration file. Tables were then created for housing the CMAP data, using scripts that came with CMAP. A cronjob was written to remove the images from the cache folder on a daily basis.

The current data files were made in a multistage process. Linkage map data were entered in a Lotus 1-2-3 spreadsheet with columns in the order specified for CMAP. The data were exported in comma-separated variable format, and converted to tab-delimited ASCII files. The script supplied with the program was used to import the map data.

## Utility

The PeanutMap homepage presents 10 choices, but map display and comparison is the most-frequently used option for users. Selecting "Maps" brings up a menu that lists all the available map sets. (In the terminology of the CMAP software, a "map" is a single linkage group, and a "map set" is a collection of linkage groups, such as is
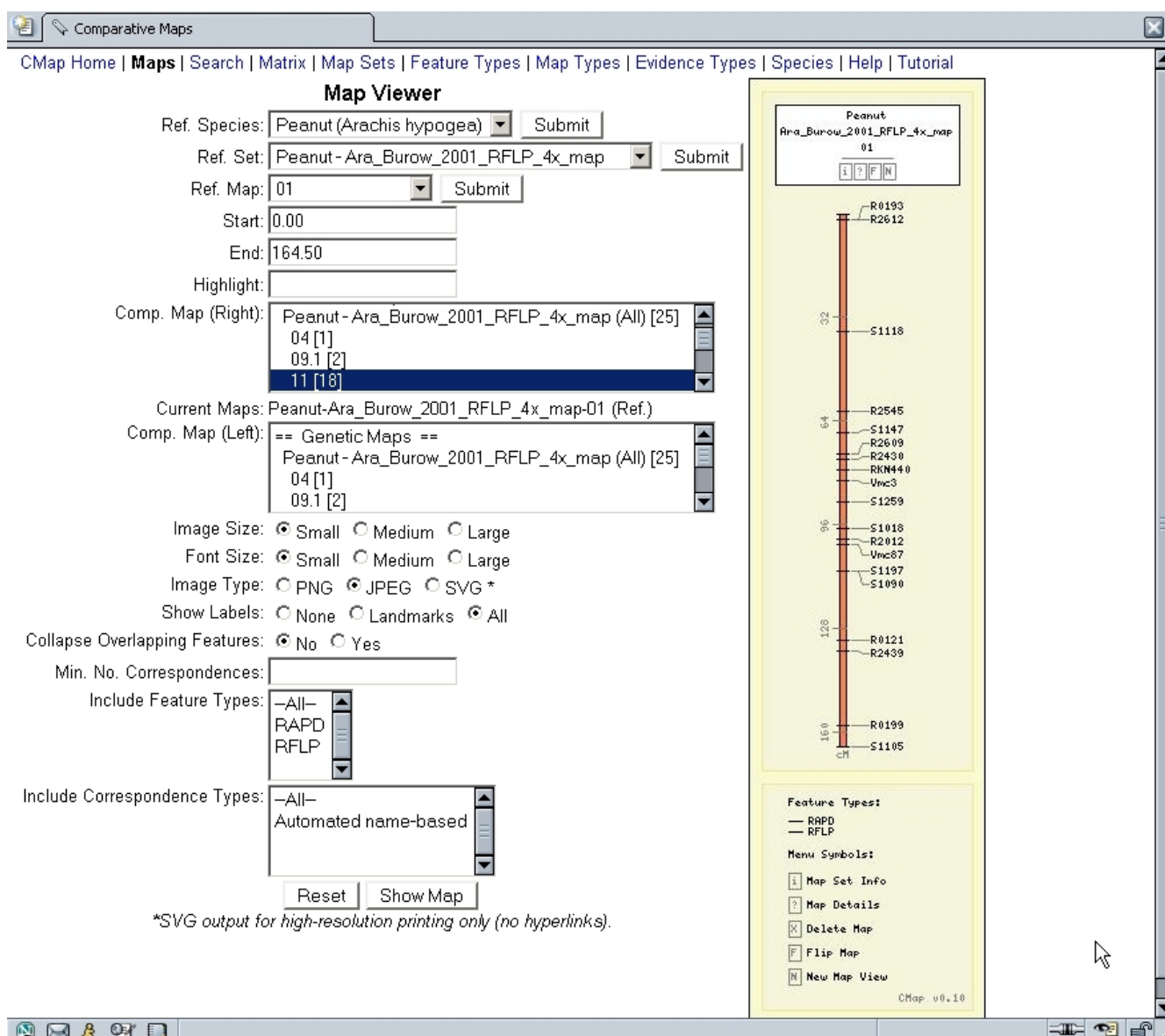
**Figure 1**
**Sample linkage group display with associated menus**. Linkage group 1 from the tetraploid map is displayed at right. At left can be seen menus for selection of map sets and maps within a set, and choices to limit the types of markers that are displayed.

found in a manuscript detailing mapping of the genome of an organism.) In PeanutMap, five map sets are available currently, two of which are large sets containing RFLP maps of the peanut genome. The first map set is from the cross *A. cardenasii* × *A. stenosperma* [5], and the other map set is from the cross between the synthetic amphidiploid TxAG-6 [*A. batizocoi* × (*A. cardenasii* × *A. diogoi*)][4x] and *A. hypogaea* cv. 'Florunner' [6]. The three small map sets are of markers associated with root-knot nematode resistance [33-35]. Once a particular map set is selected, a drop-down list of all linkage groups present in the set appears.

When a linkage group is selected for display, that linkage map and its associated markers and map distances are drawn (Fig. 1).

The database software allows comparison among linkage groups in a map set. For example, when LG1 of the tetraploid map set [6] is displayed, the map viewer also presents a list of other linkage groups and the number of markers in common with LG1 (Fig. 1). LG11 has 18 markers in common with LG1, and selecting LG11 results in co-display of both linkage groups (Fig. 2). Markers common
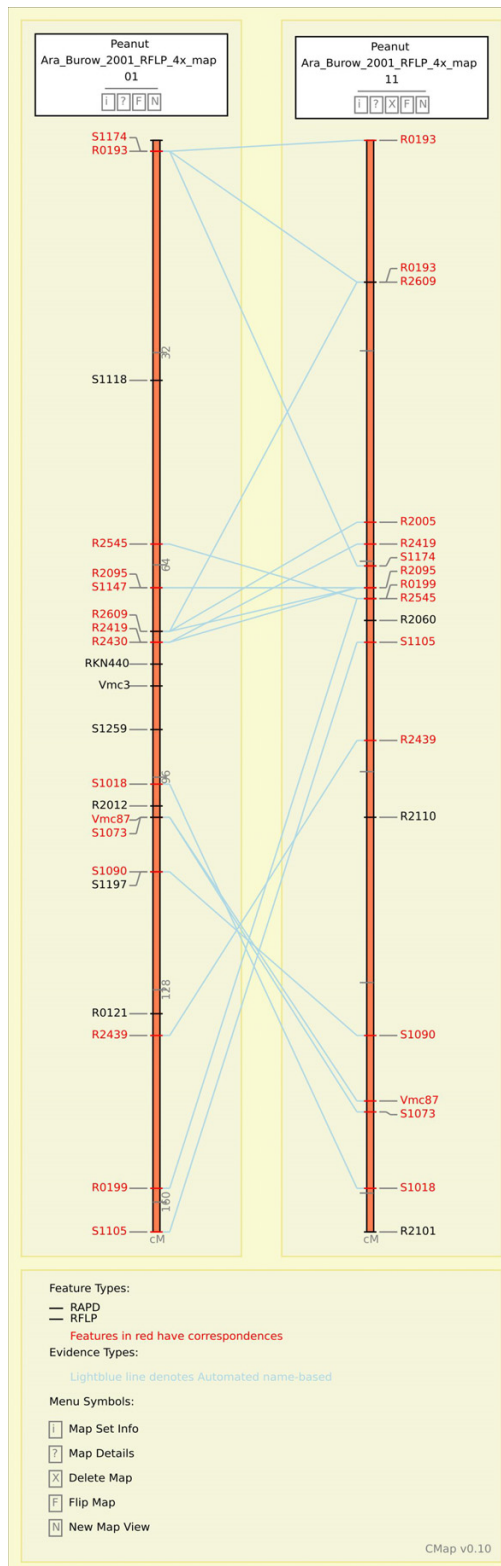
**Figure 2**
**Comparison of homoeologous linkage groups in the same map set**. Linkage groups 1 and 11 from the tetraploid map are shown; associated menus are not shown. Lines connect corresponding markers in the two maps.
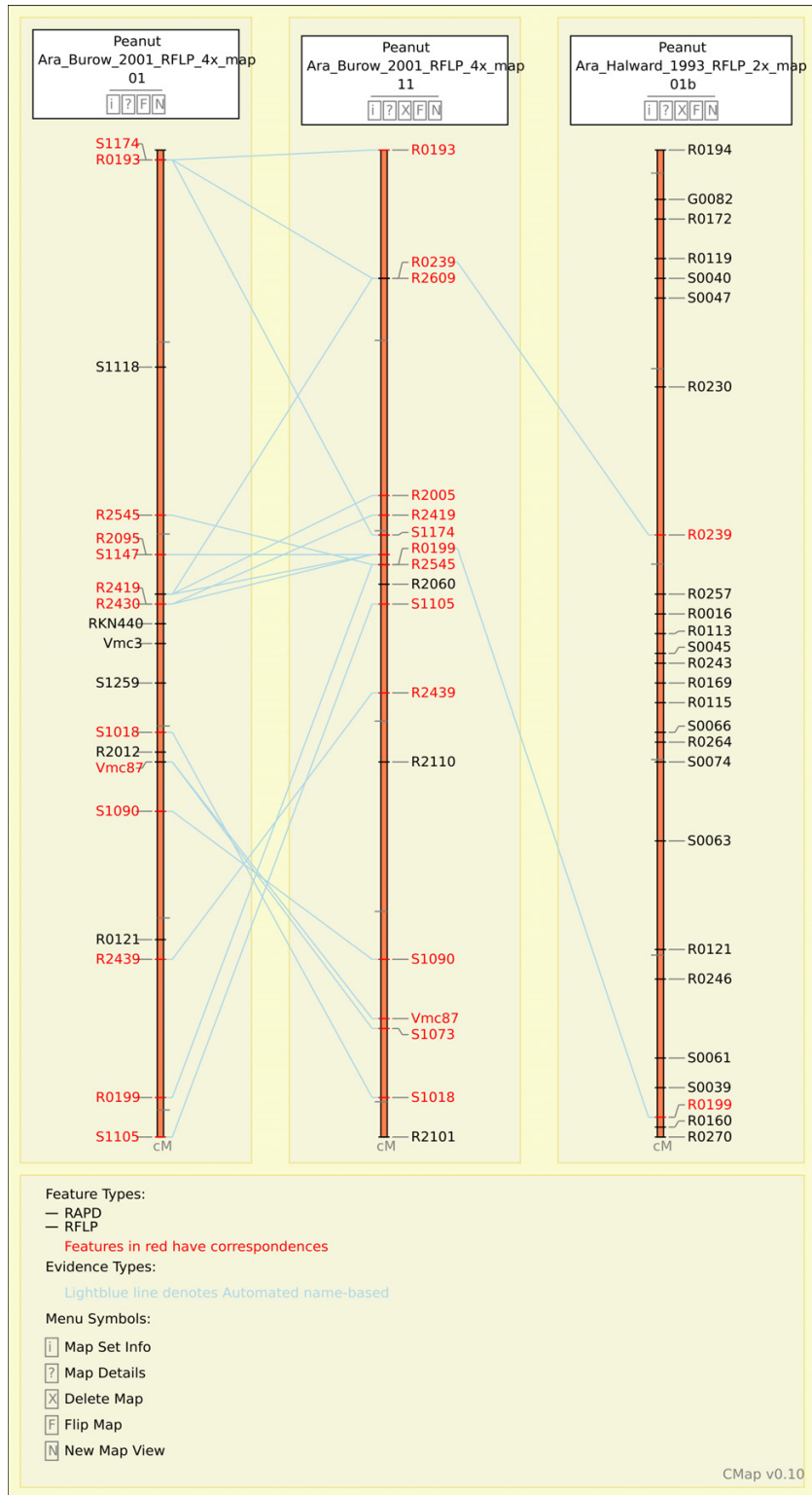
**Figure 3**
**Comparison of linkage groups within and between different map sets**. Three linkage groups are shown: linkage groups 1 and 11 from the tetraploid map, and 1b of the diploid map. Markers common to adjacent linkage groups are connected by lines.
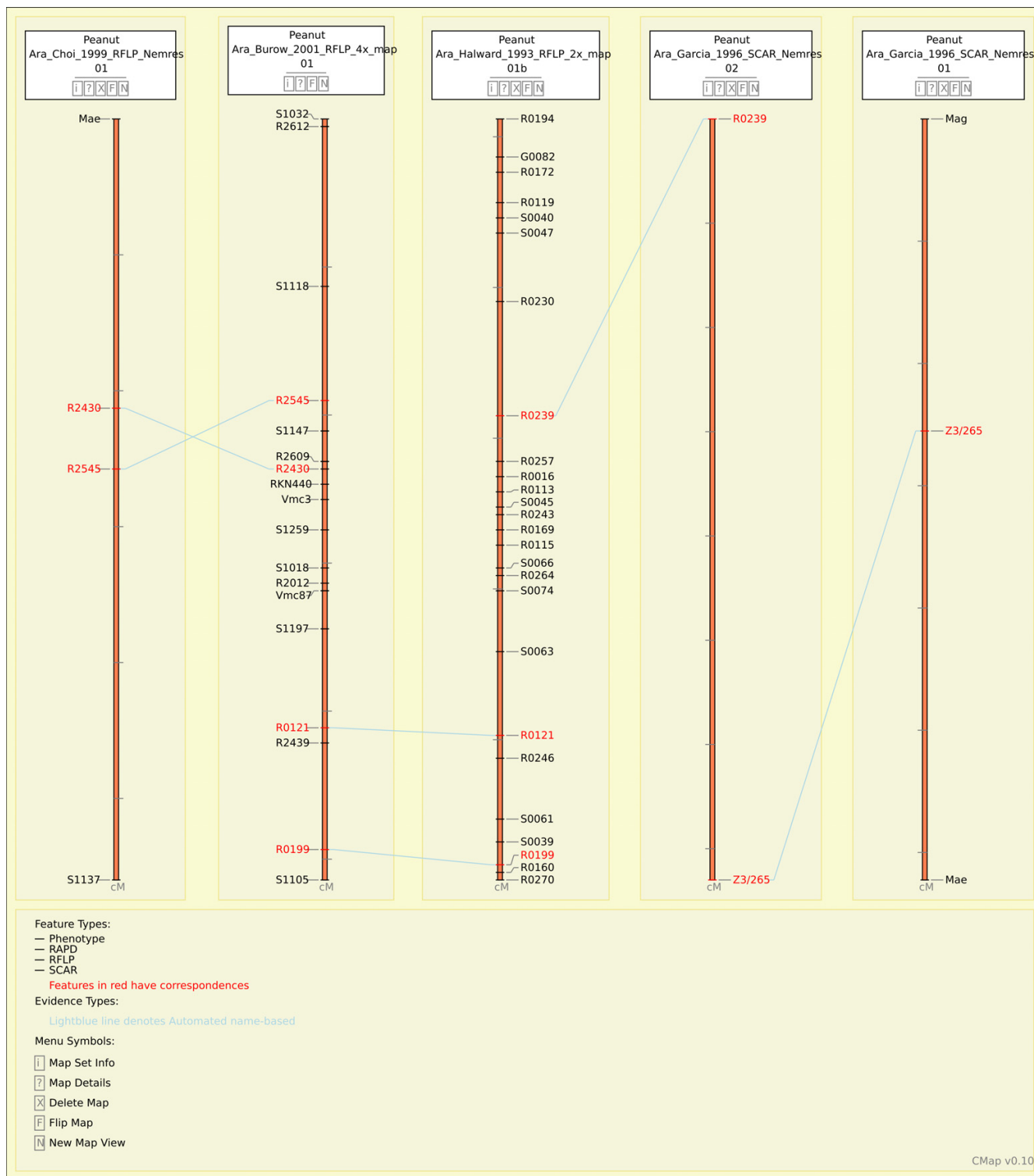
**Figure 4**
**Co-display of phenotypic and marker data**. The root-knot nematode resistance trait is denoted by 'Mae' and 'Mag', *Meloidogyne arenaria* resistance genes denoting reduced egg masses and reduced galling, respectively. Two linkage maps from genomic map sets are shown also, one of the tetraploid linkage group 1 and one of diploid linkage group 1b.

to both linkage groups are highlighted, and the association illustrated by lines connecting corresponding markers in the two maps.

PeanutMap can also be used to find associations among linkage groups in different map sets. For example, in the LG1–LG11 comparison above, LG1a and LG1b of the diploid map set [5] are indicated to have 2 corresponding markers (not shown); selecting LG1b brings up a figure highlighting markers common to two or more of the three linkage groups displayed (Fig. 3).

PeanutMap can also be used for display of phenotypic data. To date, only root-knot nematode resistance has been mapped, and these data are present in three small map sets. Potential relationships among these genes can be seen by displaying these map sets plus the corresponding full-length linkage groups from the large map sets (Fig. 4). In this comparison, it appears that the markers for nematode resistance [33-35] are probably located on the same linkage group, and at least some of the mapped markers could correspond to the same gene. The database is also capable of displaying mapped QTLs, and these will be added once such data become available.

The datasets in PeanutMap have been made available to the Legume Information System, which is compiling map data from different legume species [21]. This will permit data held in different legume databases to be used for comparison of synteny in gene order among different species.

## Discussion
PeanutMap is a useful addition to the tools for genetic mapping of peanut. It is the only peanut-specific genome database known to the authors. Several map sets are already present and available for use for mapping and for phenotypic analysis. It is expected that additional maps will be forthcoming, especially as SSR and SNP-based markers are mapped. We plan to update the database with additional genomic information as it becomes available.

Use of the SQL-formatted CMAP software will allow interoperability and data exchange with other genome databases, facilitating comparative mapping of peanut with other legumes and perhaps species outside the family. Incorporation of data from PeanutMap into the Legume Information System are an example of this.

## Conclusion
PeanutMap is a graphics-oriented database that makes the current peanut map data available in a web-accessible format, and allows comparative mapping of linkage data. This will undoubtedly accelerate the pace and usefulness of mapping the peanut genome, and will further allow integration of different peanut maps and facilitate comparison of peanut and other legumes.

## Availability and requirements
The PeanutMap database is web-accessible at the URL http://peanutgenetics.tamu.edu/cmap, and has been tested to work with the Netscape 7.1 [36], Mozilla Firefox 1.0 [37], and Internet Explorer 6.0 [38] web browsers.

## Abbreviations
BAC – Bacterial Artificial Chromosome

GMOD – Generic Model Organism Database

LIS – Legume Information Systems

NIH – National Institute for Health

QTL – Quantitative Trait Locus

SNP – Single Nucleotide Polymorphism

SSR- Single Sequence Repeats

RFLP – Restriction Fragment Length Polymorphism

USDA – United States Department of Agriculture

v – Version

YAC – Yeast Artificial Chromosome

## Authors' contributions
AMJ installed the software and initial version of the data, and wrote the initial draft of the manuscript. MDB conceived the idea of the database, provided direction for its installation, updated the database and web page design, and is responsible for its maintenance. He also revised and wrote the subsequent drafts of this manuscript.

## Acknowledgements

## References
1. Botstein D, White RL, Skolnick M, Davis RW: **Construction of a genetic linkage map in man using restriction fragment length polymorphisms.** *Am J Hum Genet* 1980, **32:**314-331.
2. Hudson TJ, Stein LD, Gerety SG, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH, Hu X, Colbert AME, Rosenberg C, Reeve-Daly MP, Rozen S, Hui L, Wu X, Vestergaard C, Wilson KM, Bae JS, Maitra S, Ganiatsas S, Evans CA, DeAndelis MM, Ingalls KA, Nahf RW, Horton LT Jr, Anderson MO, Collymore AJ, Ye W, Kouyoumjian V, Zemsteva IM, Tam J, Devine R, Courtney DF, Reynaud MT, Nguyen H, O'Connor TJ, Fizames C, Faur S, Gyapay G, Dib C, Morissette J, Orlin JB, Birren BW, Goodman N, Weissenbach J, Hawkins TL, Foote S, Page DC, Lander ES: **An STS-based map of the human genome.** *Science* 1995, **270:**1945-1954.

3.   Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE: **A high-density genetic map of** *Sorghum bicolor* **(L.) Moench based on 2926 AFLP, RFLP, and SSR markers.** *Plant Mol Biol* 2002, **48**:483-499.
4.   Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Del-Monte TA, Ding X, Garza JJ, Marler BS, Park CH, Pierce GJ, Rainey KM, Rastogi VK, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Paterson AH: **A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission, and evolution of cotton (***Gossypium***).** *Genetics* 2004, **166**:389-417.
5.   Halward T, Stalker HT, Kochert G: **Development of an RFLP linkage map in peanut species.** *Theor Appl Genet* 1993, **87**:379-394.
6.   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**:823-837.
7.   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 Sci* 1999, **39**:1243-1247.
8.   He G, Meng R, Newman M, Gao G, Pittman RN, Prakash CS: **Microsatellites as DNA markers in cultivated peanut (***Arachis hypogaea* **L.).** *BMC Plant Biol* 2003, **3**:3-8.
9.   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**:1064-1070.
10.  Moretzsohn M, de Carvalho , Hopkins MS, Mitchell S, Kresovich S, Valls JFM, Ferreira ME: **Genetic diversity of peanut (***Arachis hypogaea L***) and its wild relatives based on the analysis of hypervariable regions of the genome.** *BMC Plant Biology* 2005, **4**:11-20.
11.  Simpson CE, Starr JL, Church GL, Burow MD, Paterson AH: **Registration of 'NemaTAM' peanut.** *Crop Sci* 2003, **43**:1561.
12.  Martin GB, de Vicente MC, Tanksley SD: **High-resolution linkage analysis and physical characterization of the** *Pto* **bacterial resistance locus in tomato.** *Mol Plant Microb Interact* 1993, **6**:26-34.
13.  The Arabidopsis Initiative: **Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.** *Nature* 2002, **408**:796-815.
14.  Moore G, Devos KM, Wang Z, Gale MD: **Cereal genome evolution. Grasses, line up and form a circle.** *Curr Biol* 1995, **5**:737-9.
15.  Bennetzen JL, Freeling M: **The unified grass genome: synergy in synteny.** *Genome Res* 1997, **7**:301-306.
16.  Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR: **Estimating genome conservation between crop and model legume species.** *Proc Natl Acad Sci USA* 2004, **101**:15289-15294.
17.  Miftahudin , Chikmawati T, Ross K, Scoles GJ, Gustafson JP: **Targeting the aluminum tolerance gene Alt3 region in rye, using rice/rye micro-colinearity.** *Theor Appl Genet* 2005:906-913.
18.  Grant D, Imsande MI, Shoemaker RC: **SoyBase, the USDA-ARS soybean genome database.** 2003 [http://soybase.agron.iastate.edu]. verified Sept. 13, 2005
19.  Lamblin A-FJ, Crow JA, Johnson JE, Silverstein KAT, Kunau TM, Kilian A, Benz D, Stromvik M, Endré G, VandenBosch KA, Cook DR, Young ND, Retzel EF: **MtDB: database for personalized data mining of the model legume** *Medicago truncatula* **transcriptome.** *Nucl Acids Res* 2003, **31**:196-201.
20.  Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, McCouch S, Stein L: **Gramene: a resource for comparative grass genomics.** *Nucl Acids Res* 2002, **30**:103-105.
21.  Gonzales MD, Archuleta E, Farmer A, Kajendran K, Grant D, Shoemaker R, Beavis WD, Wright ME: **The Legume Information System (LIS) : an integrated information resource for comparative legume biology.** *Nucl Acids Res* 2005, **33**:D660-D665.
22.  Durbin R, Mieg JT: **A** *C. elegans* **database.** 1991. Documentation, code and data available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov

23.  [http://www.oracle.com].
24.  [http://www.gmod.org].
25.  [http://www.mysql.com].
26.  [http://www.postgresql.org].
27.  [http://www.gmod.org/cmap/index.shtml].
28.  [http://www.redhat.com].
29.  [http://www.boutell.com/gd].
30.  [http://www.perl.org].
31.  [http://cpan.perl.org].
32.  [http://httpd.apache.org].
33.  Garcia GM, Stalker HT, Shroeder E, Kochert G: **Identification of RAPD, SCAR, and RFLP markers tightly linked to nematode resistance genes introgressed from** *Arachis cardenasii* **into** *Arachis hypogaea.* *Genome* 1996, **39**:836-845.
34.  Burow MD, Starr JL, Simpson CE, Paterson AH: **Identification of RAPD markers in peanut (***Arachis hypogaea***) associated with root-knot nematode resistance derived from** *A. cardenasii.* *Mol Breeding* 1996, **2**:307-319.
35.  Choi K, Burow MD, Church G, Burow G, Paterson AH, Simpson CE, Starr J: **Genetics and mechanism of resistance to** *Meloidogyne arenaria* **in peanut germplasm.** *J Nematol* 1999, **31**:283-290.
36.  [http://www.netscape.com].
37.  [http://www.mozilla.org].
38.  [http://www.microsoft.com].