

## Plant genomics and agriculture: From model organisms to crops, the role of data mining for gene discovery

Viswanathan Mahalakshmi

International Crops Research  
Institute for the Semi-Arid Tropics (ICRISAT)  
Patancheru, 502 324, Andhra Pradesh, India  
Tel: 91 40 3296161  
Fax: 91 40 3241329/3296182  
E-mail: v.mahalakshmi@cgiar.org

Rodomiro Ortiz\*

International Institute of Tropical Agriculture (IITA)  
Oyo Road, PMB 5320, Ibadan, Oyo State, Nigeria  
E-mail: r.ortiz@cgiar.org

**Keywords:** bio-informatics, DNA linkage maps, genetic markers, plant breeding.

**International mailing address:** IITA, c/o Lambourn & Co., Carolyn House, 26 Dingwall Road, Croydon, CR9 3EE, United Kingdom. Tel: 234 2 241 2626. Fax: 234 2 241 2221.

**Sequencing and analysing of the *Arabidopsis thaliana* genome, the first plant kingdom genome to be unraveled, will always remain a scientific landmark. International initiatives to sequence rice, the most important cereal in Asia, are underway. However as functional information piles up in *Arabidopsis* and rice, researchers working in other crops will benefit from this new knowledge and apply it to their studied plants or crop species. The increasing role of public databases of model organisms and bio-informatics in data mining, presents a new opportunity as well as a challenge to researchers to develop more focused molecular tools for gene discovery and deployment. The work presented in here describes how such an approach has benefited sorghum, a rainfed semi-arid tropical cereal.**

Over the last two decades, comparative genetics has shown that the organization of genes within plant genomes has remained more conserved over the evolutionary periods than previously thought (Gale and Devos, 1998). The complete sequencing of the *Arabidopsis thaliana* genome has been regarded as a landmark in plant sciences (Dennis and Surridge, 2000). The sequencing of the genome of a species is not the end by itself but is just the beginning of a new venture to unravel genetic information and to gain better insight into the genetics of other species under investigation. However, a question should be posed: What is the colinearity between model plant systems, which are receiving priority attention for DNA sequencing, and more important crop species belonging to the same taxon?

Insights from the *Arabidopsis* sequence, which has one copy of each gene and less than 10% repetitive DNA

sequences, are likely to boost developments in agricultural science, but the bigger challenge today is to use this information and to extend it to food, feed and fiber crops. Rice, one of the most important crops in the tropics, is the next model plant system receiving priority attention for genome sequencing. Its genome sequence was expected to be available by 2004, though today biotech multinationals (Syngenta and Monsanto) claim that they have completed the first draft of the rice genome (<http://www.nadii.com>; [http://www.syngenta.com/en/media/article.asp?article\\_id=126](http://www.syngenta.com/en/media/article.asp?article_id=126); Gerard, 2001). The controversy on whether other agricultural crops need complete genome sequencing or the information from *Arabidopsis* and rice will be sufficient for other crops, which have significantly more amounts of DNA but share a common gene set, will continue.

In the mid 1980s, when restriction fragment length polymorphism (RFLP) analysis was applied to plants, it was clear that the complementary DNA (cDNA) RFLP probes provided the tools for comparison of the arrangement of low-copy number genes among various plant species. This has as a result led to the elucidation of extensive genome comparison and conservation of gene order within homoeologous chromosomal segments. Studies showed that tomato and potato maps were very similar (Bonierbale et al. 1988), a close genomic relationship between cereals (Gale and Devos, 1998), and between *Brassica* crops and *Arabidopsis*. Wheat, rye, barley, maize, sorghum and millet seem to have genetic layout similar to rice; *i.e.*, the genes are in much the same order but the larger genomes have more junk DNA between genes (Gale and Devos, 1998). These findings suggest that improvement of other important food crops can benefit from information obtained in these model crop systems.

\*Corresponding author

Complete co-linearity between even closely related species is rare, where co-linearity is often interrupted by insertions, deletions and inversions. One of the few studies where contiguous DNA sequences have been compared shows complete co-linearity through three genes in the ~20kb Sh2-A1 regions in rice and sorghum (Chen et al. 1997). Good co-linearity between contiguous regions in rice was cross-mapped to barley and foxtail millet (Gale and Devos, 1998). However, interspecific analysis of R-like genes, specifically NBS-LRR (nucleotide binding site leucine-rich repeat) between dicot and monocot species revealed share motifs but provided no monocot specific signature, suggesting therefore, a re-arrangement of resistance gene (R-gene) loci between related monocot species (Leister et al. 1998).

As the functional information accumulates in *Arabidopsis* and rice, researchers working on other crops will discover how these results apply to their plant species. Many commercial crops (cotton, sugar beet) are relying on the co-linearity information on these model crops (Cook, 1998; Adam, 2000). Though eventually as the cost decline and the process of sequencing gets more automated, all food and commercially important crops may get fully sequenced. However, the agricultural research community need not wait to reap the benefit as they can make better-informed decisions based on current knowledge from the model crops. This article discusses how this can be achieved using our work on sorghum as an example, which benefited from the rice genome initiative. This article also highlights the benefits for sharing information in the public domain and the role of bioinformatics.

## Materials and Methods

### Consensus map of sorghum

Construction of a linkage map is fundamental step for a detailed genetic study and marker-assisted breeding approach in any crop. Sorghum genome mapping based on DNA markers began in 1990s, and since then several maps have been published. Data from available relevant published maps of sorghum (Whitkus et al. 1992; Chittenden et al. 1994; Pereira et al. 1994; Ragab et al. 1994; Xu et al. 1994; Dufour et al. 1997; Ming et al. 1998; Nagamura et al. 1998; Tao et al. 1998; Boivin et al. 1999; Peng et al. 1999) were assembled and entered into a relational database (SQL7.0™). This curated data contain detailed information on the source including the material used, linkage group, distance, marker designation and alias marker designation in cases where the markers have been renamed by others. The relationship among the linkage group (LG) from various maps or all LG from a single map can be viewed graphically and the data in tabular format. The extent of relationship between maps from the various sources can be determined based on common markers provided as an option (search by markers). This will also depend on the degree to which the maps cover the genome, which is also displayed for visual comparison.

### Comparative mapping of sorghum with rice

After the publication of a fairly comprehensive map of rice (Harushima et al. 1998; <http://rgp.dna.affrc.go.jp/publicdata/geneticmap2000/index.html>), a consensus map between rice and sorghum was developed using these RFLP markers from rice (Nagamura et al. 1998). From this relationship between rice and sorghum genomes it was possible to generate a consensus linkage map for sorghum consisting of 381 DNA markers from rice. A high degree of co-linearity exists between the 10 sorghum linkage groups and the 12 rice linkage groups. For the rice markers, sequence information was obtained from the GenBank. The homology of these sequences was determined using Basic Local Alignment Tool program at National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>; <http://rgp.dna.affrc.go.jp/>). The entire data were collated and entered into a Research Database Management System (RDBMS) and is available at <http://www.icrisat.org/text/research/grep/homepage/syteny/intro.htm>. The relationship between rice and sorghum linkage groups or vice versa can be displayed both in tabular and graphical format. Markers, where information from Basic Local Alignment Search Tool (BLAST) was available as of December 2000, are linked to the detailed information on the protein, the organism, GenBank accession number of the sequence and the bit score.

## Results

### Consensus sorghum map

The assembling of the data into a RDBMS gives potential new avenues to explore this data. A linkage group from all the maps can be aligned together ([Figure 1](#)) or all the linkage groups from a given source can be aligned together ([Figure 2](#)). The information on the total distance and number of markers on a linkage group is readily displayed. The loci with more than a marker are marked in red. When the mouse is moved over the marker the details are displayed while click on gives complete information. A summary table listing the total number of markers for each map gives an indication the extent of the saturation of the map. The selected markers from the linkage groups are also given in a tabular form. Often the interest is the anchoring common probes across the maps, which can be used as landmark reference. Searching based on a marker will list the linkage groups in which the marker is found in various maps. Complete alignment of all sorghum maps onto a single map was not possible, *i.e.*, a bridging map for these maps, due to the lack of common loci across maps to compare. The other problem was inconsistency in probe location. As more and more data get accumulated it should be possible to construct a saturated map for sorghum.

### Syteny between rice and sorghum

Comparisons can be made between markers on the sorghum linkage groups and rice or *vice versa* ([Figure 3](#)). This gives the direct relationship between the linkage groups of

sorghum and rice. Most of the markers on rice linkage group 1 are found on sorghum linkage group G. Similarly most of the markers on sorghum linkage group A are found on rice linkage group 8. There is a fine and straight colinearity between the ten (A to J) sorghum linkage groups and the 12 (1 to 12) rice linkage groups (A=8; B=4; C=3, 10, 7; D=2; E=5; F= 7, 9; G= 1; H=6; I=12; J= 11). These results suggest there is an extensive conservation of gene order in rice and sorghum genomes. The practical aspects of this information would be in map-based cloning for specific genes. The information of the probe from BLAST search for homologs (Figure 4) would be an important resource for researchers to develop trait-based markers especially expressed sequence tags (EST).

## Discussion

*Arabidopsis thaliana* was described in 1777 by the British botanist William Curtis as a plant of no particular virtue or uses and remained so until recently when it was chosen to be first model flowering plant for genome sequencing. The reasons for choosing *Arabidopsis* as a model plant system are its small size, rapid life cycle, high seed production, and small genome size (120Mb). The elucidation of the full genomic sequence of *Arabidopsis* is only at the beginning of functional genomics research. We may be still regarded as functional genomics illiterates, despite the 26,000 genes that are known in the *Arabidopsis* sequence. About 70% of this genome appear to be duplicated, thereby about 15,000 different genes can be the repertoire for other crop plants to build on.

Among the cereal crops, rice has the smallest genome (430MB), which is four times larger than *Arabidopsis*. Insights into the genome sequence of *Arabidopsis* and rice will likely boost agricultural research, regardless of the fact that the other crop plant genomes are significantly larger. In crop plants, gene duplication or redundancy occurs to ensure adaptation of species in unpredictable environments and from mutations. Two factors contribute to gene redundancy: polyploidization (duplication of the entire set of chromosomes) and local gene duplications (copying of individual genes within chromosomes). Despite the apparent gene duplication, geneticists have identified mutations producing visible defects in plants and occurring in only one of the duplicated genes. If both genes had the same function, then one of these genes would compensate for the mutation in the other. Many researchers argue that full sequencing of other cereal crop would be a waste of resources because different cereals, although having significantly large amounts of DNA, they share common set of genes. The large genomes have more junk DNA between genes, and more emphasis should be laid on EST.

## Development of consensus map of sorghum

Restriction fragment length polymorphism (RFLP) markers proved to be valuable in establishing linkage maps of many crop species. In the last few years, several good high

density RFLP maps have been produced for sorghum (Chittenden et al. 1994; Xu et al. 1994; Dufour et al. 1996; Dufour et al. 1997; Tao et al. 1998). Generally, the alignment of the different sorghum maps will serve sorghum researchers to use all of the available mapping information effectively. Most of the available sorghum genome maps are far from saturated, which implies that there are uncertainties about the distance between loci and even linkage groups on the map. By bringing all these information together, it is hoped to help researchers to establish whether variation for a particular trait of interest in their study is due to the same major gene or allele of other population. Still these maps are far from complete and lack additional information on agronomic traits associated with these markers. We will continue with our efforts to search and update this information and place in public domain as and when they become available.

## Comparative genome mapping of sorghum and rice

Geneticists and evolutionary botanists have long held interest in the phylogeny of crops and in chromosomal evolution. Until recently, the only means of addressing this problem was via cytological analysis of inter-specific hybrids and through surveys of naturally occurring chromosomal diversity within populations. Comparative genome mapping adds a very powerful technique for the mode and tempo of chromosomal evolution. The utility of comparative mapping outside this aspect of evolutionary biology is in the development of molecular markers for specific traits. The genome of sorghum is larger than rice, but considerably smaller than maize and sugarcane. The unified cereal genome concept of Bennetzen and Freeling (1993) received strong support from recent comparative genome mapping studies of Gramineae species (Bennetzen et al. 1998; Kellogg, 1998; McCouch, 1998; McCouch, 2001). These investigations have shown that the genetic content and linear order in large chromosomal segments have been conserved over millions of years, thereby leading to the construction of the first grass genome map that aligns several species (Moore et al. 1995; Gale and Devos, 1998).

There have been many efforts to make comparative maps between grass species such as maize and sorghum (Hulbert et al. 1990; Binelli et al. 1992; Whitkus et al. 1992), sugarcane, maize and sorghum (Dufour et al. 1997), and rice and sorghum (Paterson et al. 1995; Nagamura et al. 1998). The exact extent of co-linearity is often under estimated in crops whose maps are not saturated. Sorghum, maize and sugarcane belong the Andropogoneae tribe and the genome of sorghum is the smallest of the three. Comparison of maize, sugarcane and sorghum, revealed that maize is an ancient tetraploid while sugarcane is a polyploid (Guimaraes et al. 1997). Furthermore a simple organization of the comparative mapping suggested a high degree of conservation of genic regions and duplicated structure in these polyploid species, while sorghum standing

between these two food crops (Draye, 2001). Although rice is phylogenetically a distant crop from sorghum, it is considered the model crop for cereals due to its small genome size. The reports on comparative mapping of rice probes on sorghum indicates very high degree of colinearity and high degree of conservation of gene order (Paterson et al. 1995; Nagamura et al. 1998). The practical aspect of comparative mapping between two genomes would be in map-based gene cloning. As more and more information on functional regions of the genome becomes available for rice genome, sorghum and other related cereals would benefit immediately. The information on the rice–sorghum syntenic regions (Nagamura et al. 1998) was compiled with other information on sequence and homology was determined by BLAST from the GenBank data and was made available through <http://www.icrisat.org/text/research/grep/homepage/syteny/intro.htm> on the world wide web. The immediate use of these data would be in the development of sequence tagged marked site (STMS) based simple sequence repeats (SSR) markers for sorghum.

At the time ICRISAT started this project (2000), there was very little information available for sorghum in the public databases. At the time of the writing of this article (May 2001) there are over 150,000 nucleotide sequences of sorghum in the public domain database while for rice the number is just below 200,000. The completion first draft rice genome was announced by Syngenta and Monsanto in January 2001. In May 2001, the number of protein sequence available in public database for sorghum was just 463, while for rice 8400 protein sequences were available. These advances provide a new opportunity to develop markers based on known genes from rice and test them in sorghum and other related grass species. The debate concerning whether markers can be developed and deployed for other cereal crop improvement using this approach and without a molecular linkage map will continue. The scientific community should, however, use this information as appropriate. The findings that regulatory gene orders are more conserved across species than the resistance gene (disease) homologs, which are less conserved and are evolving faster (Leister et al. 1998), requires certain degree of caution while developing and deploying markers in marker-assisted breeding program for resistance traits using genome information from other species. It could be argued that in many instances disease resistance is defined as reaction to the disease and often the underlying physiological and biochemical mechanisms conferring resistance are not well elucidated. It is likely that the biochemical mechanisms, which confer resistance, are conserved across species and follow common pathways for a number of diseases.

We need to elucidate the biochemical functions of pathogen-avirulence (Avr) proteins and host-plant-resistance (R) proteins. The major break through came after the cloning of maize *Hm1* disease resistance gene against *Cochliobolus carbonum*, which encodes the NADPH-

dependent reductase that inactivates the toxin produced by the invading fungus (Johal and Briggs, 1992). The pace of cloning R genes increased during the 1990's with the cloning and characterization of R genes for resistance to several classes of pathogens, including viral, bacterial and fungal pathogens. Today we understand the mechanism of host-plant resistance and host-pathogen interaction. R genes are presumed to (a) enable plants to detect Avr-gene-specified pathogen molecules, (b) initiate signal transduction to activate defenses, and (c) have the capacity to evolve new R gene specificities rapidly. Comparison of R-proteins of several plant species revealed significant stretches of homologous regions of amino-acids. Today, 481 sequences, which contain putative R genes, have been identified (Staskawicz, 2001). Plant breeders have used disease resistance genes (R genes) to control plant disease since the turn of the century. Molecular cloning of R genes that enable plants to resist a diverse range of pathogens has revealed that the proteins encoded by these genes have several features in common. These findings suggest that plants may have evolved common signal transduction mechanisms for the expression of resistance to a wide range of unrelated pathogens. Characterization of the molecular signals involved in pathogen recognition and of the molecular events that specify the expression of resistance may lead to novel strategies for plant disease control.

### Use of public domain databases

Today, the search for a gene of interest starts with sequence information, including EST, genome sequence and protein sequence. This information is publicly available for large-scale analysis from GenBank at NCBI and European Molecular Biology Laboratory (EMBL). Genome related public databases have already become an invaluable part of the scientific community and are the public window for the high-throughput genomic projects. The thrust of a high-throughput facility is the creation of large, well-organized, rigorous datasets. Public domain databases, most notably the model organism databases, have two major consumers: the focused scientific community actively studying that system, and the large scientific community interested in relating this specialized information to and from other systems. As mentioned in the earlier section the model system databases would be increasingly mined by other but related crop specialists to design markers for marker-assisted selection and –aided introgression methods.

Finding new genes that add value to agricultural crops and their products has immense value in agri-business. For rice genome sequence data, scientists are leveraging on data from other model crops for gene annotation (Yuan et al. 2001; <http://www.tigr.org/tdb/edb2/osa1/htmls/osa1.html>). Plant breeding has changed with the introduction of molecular techniques. Molecular markers allow for the extension of traditional breeding methods with one important difference – to transfer greater variety of genetic information in a more precise and controlled manner. Marker-assisted selection for important but complex traits,

which are often difficult to select in the routine breeding programs, will enhance the breeding programs in terms of better-focused products and save time and resources.

The availability of massive amounts of nucleotide sequence data has led to the development of innovative ways to examine these data as reflected in their functions. Data mining encompasses the use of pattern recognition technologies and statistical techniques to examine large data amounts. *De novo* generation of microsatellite markers through laboratory-based screening of SSR-enriched genomic libraries is highly time consuming and expensive. An alternative is to screen the public databases of related model species where abundant sequence data is already available. Beyond the cost savings, this approach also offers the possibility of identifying rare microsatellite motifs which would be uneconomical to identify through laboratory protocols. Unique databases are being assembled and made available to public, *i.e.*, Monsanto has made available to the public the data of 6655 SSRs for rice (Gerard, 2001).

The molecular maps for legume crops of the semi-arid tropics (*i.e.* chickpea, cowpea, groundnut and pigeonpea) are not very dense compared to the other related legume crops (*i.e.* soybean) but data from related species could be mined to develop maps and markers. For example, gene synteny research in *Vigna* species provide a means to detect orthologous genes by using common sets of chromosomes to probe their genomes (Fatokun et al. 1992). Cowpea (*V. unguiculata*) and mung bean (*V. radiata*) have QTL for seed weight in the same conserved region, which are arranged in identical order (Menancio-Hautea et al. 1992). This result was not surprising because it reflects the selection of chromosome regions controlling the most important character selected by farmers in these crop species: seed size. In cereal crops (*e.i.* maize and sorghum) similar findings for plant height were reported (Lin et al. 1995; Pereira and Lee, 1995), which reveals same domestication evolution pattern among crop species (Paterson et al. 1995). Hence, we envisage the successful integration of data mining, gene discovery, marker development and tactful gene deployment to become routine practice in crop improvement programs in the next decades.

## References

Adam, D. (2000). Now for the hard ones. *Nature* 408:792-793.

Bennetzen, J.L.; San Miguel, P.; Chen, M.; Tikhonov, A.; Francki, M. and Avramova, Z. (1998). Grass genomes. *Proceedings of the National Academy of Sciences, USA* 95:1975-1978.

Bennetzen, J.L. and Freeling, M. (1993). A unified grass genome: Synergy in synteny. *Genome Research* 7:301-306.

Binelli, G.; Giangranceschi, L.; Pe, M.E.; Taramino, G.; Busso, C.; Stenhouse, J. and Ottaviano, E. (1992). Similarity between maize and sorghum genomes as revealed by maize RFLP probes. *Theoretical and Applied Genetics* 84:10-16.

Boivin, K.; Deu, M.; Rami, J.F.; Trouche, G. and Hamon, P. (1999). Towards a saturated sorghum map using RFLP and AFLP markers. *Theoretical and Applied Genetics* 98:320-328.

Bonierbale, M.W.; Plaisted, R.L. and Tanksley, S.D. (1988). RFLP maps based on a common set of clones reveal modes of chromosomal evolution of potato and tomato. *Genetics* 120:1095-1103.

Chen, M.; Sanmiguel, P.; De Oliveira, A.C.; Woo, S.S.; Zang, H.; Wing, R.A. and Bennetzen, J.L. (1997). Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. *Proceedings of the National Academy of Sciences, USA* 94:3431-3435.

Chittenden, L.M.; Schertz, K.F.; Lin, Y.R.; Win, R.A. and Paterson, A.H. (1994). A detailed RFLP map of *Sorghum bicolor* × *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of sorghum chromosomes or chromosomal segments. *Theoretical and Applied Genetics* 87:925-933.

Cook, R.J. (1998). Towards a successful multinational crop plant genome initiative. *Proceedings of the National Academy of Sciences, USA* 95:1993-1995.

Dennis, C. and Suggidge, C. (2000). *A. thaliana* genome. *Nature* 408:491.

Draye, X.; Lin, Y.R.; Qian, X.Y.; Bowers, J.E.; Burow, G.B.; Morrell, L.P.; Peterson, D.G.; Presting, G.G.; Ren, S.-X.; Wing, R.A. and Paterson, A.H. (2001). Toward integration of comparative genetic, physical, diversity and cytomolecular maps for grasses and grains, using sorghum genome as a foundation. *Plant Physiology* 125:1325-1341.

Dufour, P.; Deu, M.; Grivet, L.; D'Hont, A.; Paulet, F.; Bouet, A.; Lanaud, D.; Glaszmann, J.C. and Hamon, P. (1997). Construction of a composite sorghum genome map for comparison with sugarcane, a related complex polyploid. *Theoretical and Applied Genetics* 94:409-418.

Dufour, P.; Grivet, L.; D'Hont, A.; Deu, M.; Trouche, G.; Glaszmann, J.C. and Hamon, P. (1996). Comparative genetic mapping between duplicated segments of maize chromosomes 3 and 8 and homoeologous regions in sorghum and sugarcane. *Theoretical and Applied Genetics* 92:1024-1030.

Fatokun, C.A.; Menancio-Hautea, D.I.; Danesh, D. and Young, N.D. (1992). Evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics* 132:841-846.

- Gale, M.D. and Devos, K.M. (1998). Comparative genetics in the grasses. Proceedings of the National Academy of Sciences, USA 95:1971-1974.
- Gerard, F.B. (2001). The use of the Monsanto draft rice genome sequence research. Plant Physiology 125:1164-1165.
- Guimaraes, C.; Sills, G.R. and Sobral, W.S. (1997). Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. Proceedings of the National Academy of Sciences, USA 94:14261-14266.
- Harushima, Y.; Yano, M.; Shomura, A.; Sata, M.; Shimano, T.; Kuboki, Y.; Tamamoto, T.; Lin, S.Y.; Antonio, B.A.; Parco, A.; Kajiya, H.; Huang, N.; Yamamoto, K.; Nagamura, Y.; Kurata, N.; Khush, G.S. and Sasaki, T. (1998). A high density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148:479-494.
- Hulbert, S.H.; Richter, T.E.; Axtell, J.D. and Bennetzen, J.L. (1990). Genetic mapping and characterization of sorghum and related crops by means of maize probes. Proceedings of the National Academy of Sciences, USA 87:4251-4255.
- Johal, G.S. and Briggs, S.P. (1992). Reductase activity encoded by the HM1 disease resistance gene in maize. Reductase activity encoded by the HM1 disease resistance gene in maize. Science 258:985-987.
- Kellogg, E.A. (1998). Relationships of cereal crops and other grasses. Proceedings of the National Academy of Sciences, USA 95:2005-2010.
- Leister, D.; Kurth, J.; Laurie, D.A.; Yano, M.; Sasaki, T.; Devos, K.; Garner, A. and Schulze-Lefert, P. (1998). Rapid reorganization of resistance gene homologues in cereal genomes. Proceedings of the National Academy of Sciences, USA 95:370-375.
- Lin, Y.R.; Schertz, K.F. and Paterson, A.H. (1995). Comparative analysis of QTLs affecting plant height and maturity across the poaceae, in reference to an interspecific sorghum population. Genetics 141:391-411.
- McCouch, S.R. (2001). Genomics and synteny. Plant Physiology 125:152-155.
- McCouch, S.R. (1998). Toward a plant genomics initiative: Thoughts on the value of cross-species and cross-genera comparisons in the grasses. Proceedings of the National Academy of Sciences, USA 85:1983-1985.
- Menancio-Hautea, D.I.; Fatokun, C.A.; Kumar, L.; Danesh D. and Young, N.D. (1992). Comparative genome analysis of mung bean (*Vigna radiata*) and cowpea (*V. unguiculata*) using RFLP mapping data. Theoretical and Applied Genetics 86:797-810.
- Ming, R.; Liu, S.C.; Lin, Y.R.; da Silva, J.; Wilson, W.; Braga, D.; van Deynze, A.; Wenslaff, T.F.; Wu, K.K.; Moore, P.H.; Burnquist, W.; Sorrells, M.E.; Irvine, J.E. and Paterson, A.H. (1998). Detailed alignment of *Saccharum* and *Sorghum* chromosomes: Comparative organization of closely related diploid and polyploid genomes. Genetics 150:1663-1682.
- Moore, G.; Foote, T.; Helentjaris, T.; Devos, K.; Kurata, N. and Gale, M. (1995). Was there a single ancestral cereal chromosome? Trends in Genetics 11:81-82.
- Nagamura, Y.; Tanaka, T.; Nozawa, H.; Kaidai, H.; Kasuga, S. and Sasaki, T. (1998). Syntenic regions between rice and sorghum genomes. National Grassland Research Institute Report (Japan) 9:97-103.
- Paterson, A.H.; Lin, Y.R.; Li, Z.; Schertz, K.F.; Doebley, J.F.; Pinson, S.R.M.; Liu, S.C.; Stansel, J.W. and Irvine, J.E. (1995). Convergent domestication of cereal crops by independent mutations at corresponding loci. Science 269:1714-1718.
- Peng, Y.; Schertz, K.F.; Cartinhour, S. and Hart, G.E. (1999). Comparative genome mapping of *Sorghum bicolor* (L.) Moench using an RFLP map constructed in a population of recombinant inbred lines. Plant Breeding 118:225-235.
- Pereira M.G. and Lee M. (1995). Identification of genomic regions affecting plant height in sorghum and maize. Theoretical and Applied Genetics 90:380-388.
- Pereira, M.G.; Lee, M.; Bramel-Cox, P.; Woodman, W.; Doebley, J.; Whitkus, R. and Lee, M. (1994). Construction of an RFLP map in sorghum and comparative mapping in maize. Genome 37:236-243.
- Ragab, R.A.; Dronavalli, S.; Saghai-Marooof, S.A. and Yu, Y.G. (1994). Construction of a sorghum RFLP linkage map using sorghum and maize DNA probes. Genome 37:590-594.
- Staskawicz, B.J. (2001). Genetics of plant-pathogen interactions specifying plant disease resistance. Plant Physiology 125:73-76.
- Tao, Y.Z.; Jordan, D.R.; Henzell, R.G. and McIntyre, C.L. (1998). Construction of a genetic map in a sorghum recombinant inbred line using probes from different sources and its comparison with other sorghum maps. Australian Journal of Agricultural Research 49:729-736.
- Whitkus, R.; Doebley, J. and Lee, M. (1992). Comparative genome mapping of sorghum and maize. Genetics 132:1119-1130.

Xu, G.W.; Magill, C.W.; Schertz, K.F. and Hart, G.E. (1994). A RFLP linkage map of *Sorghum bicolor* (L.) Moench. Theoretical and Applied Genetics 89:139-145.

Yuan, Q.; Quackenbush, J.; Sultana, R.; Perlea, M.; Salzberg, S.L. and Buell, C.R. (2001). Rice bioinformatics. Analysis of rice sequence data and leveraging the data to other plant species. Plant Physiology 125:1166-1174.

## Appendix I

### Figures

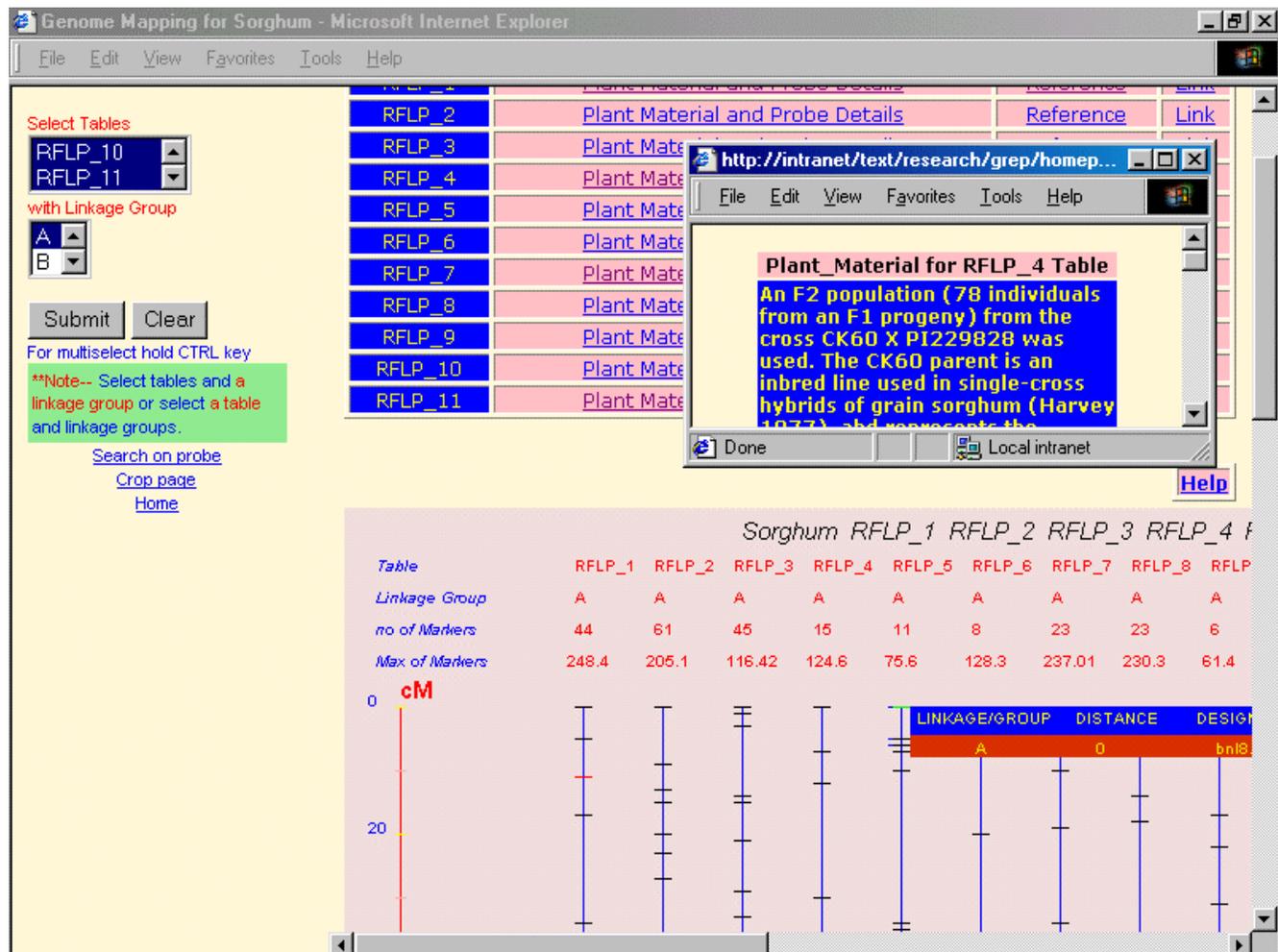


Figure 1. Sorghum linkage group A from all the maps with the plant material information inserted for RFLP\_4 map.

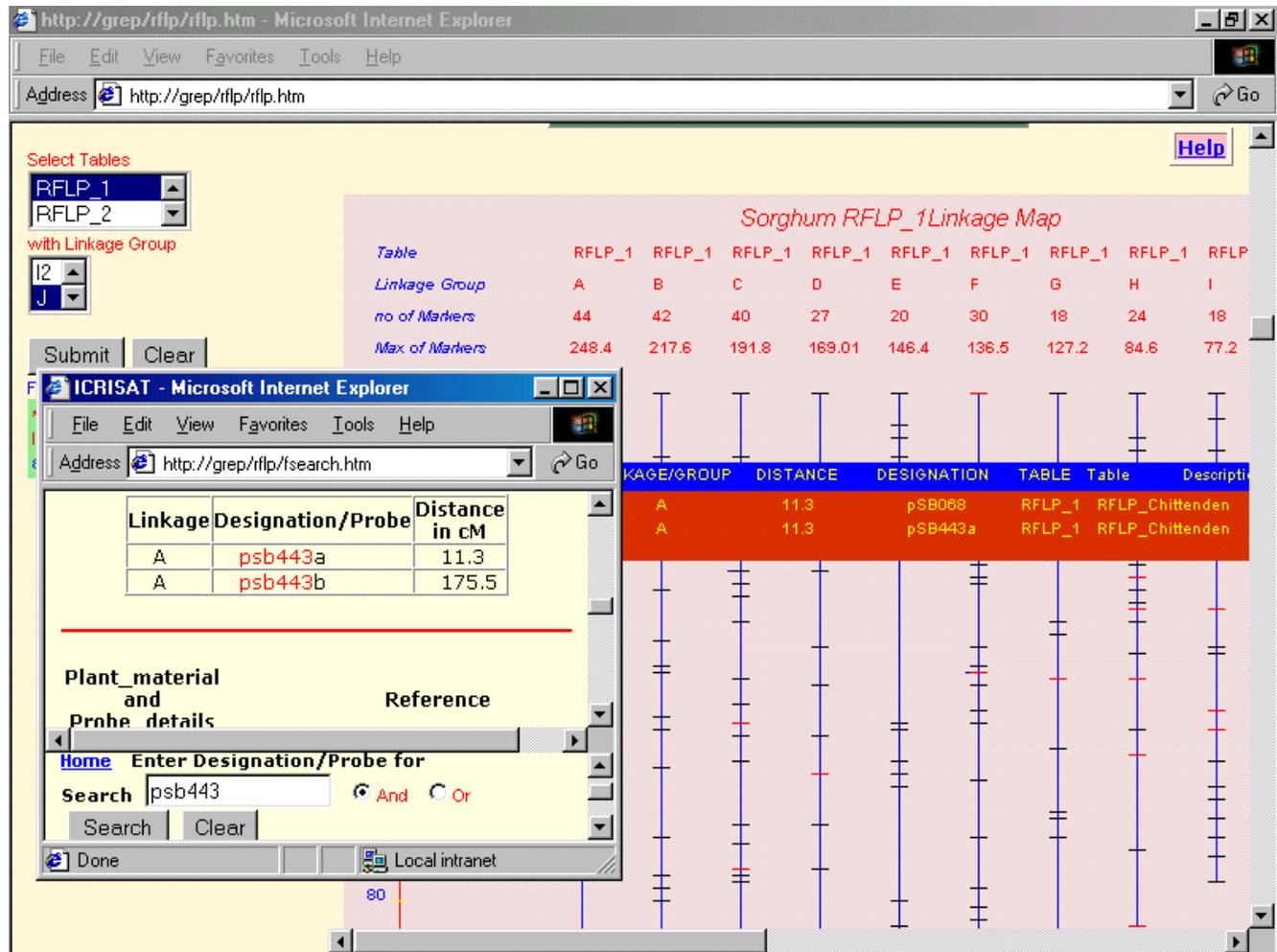


Figure 2. All linkage groups from the map of RFLP\_1 (Chittenden et al. 1994) with the results of a search for a marker (psb443) inserted.

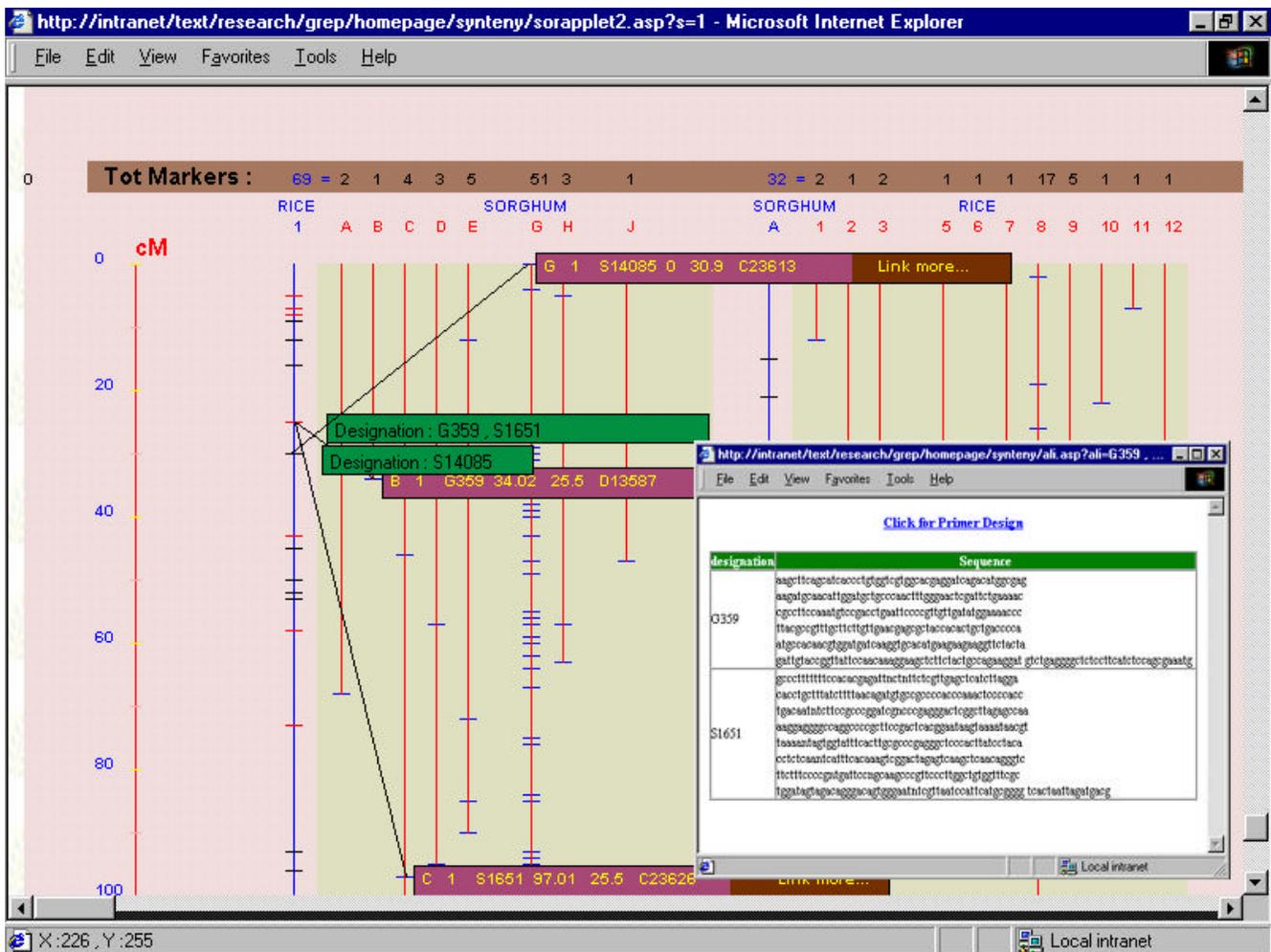


Figure 3. Syntenic regions of rice with sorghum for rice linkage group 11 and syntenic regions of sorghum with rice for sorghum linkage group A. The insert of the details of the markers on rice linkage group 1.

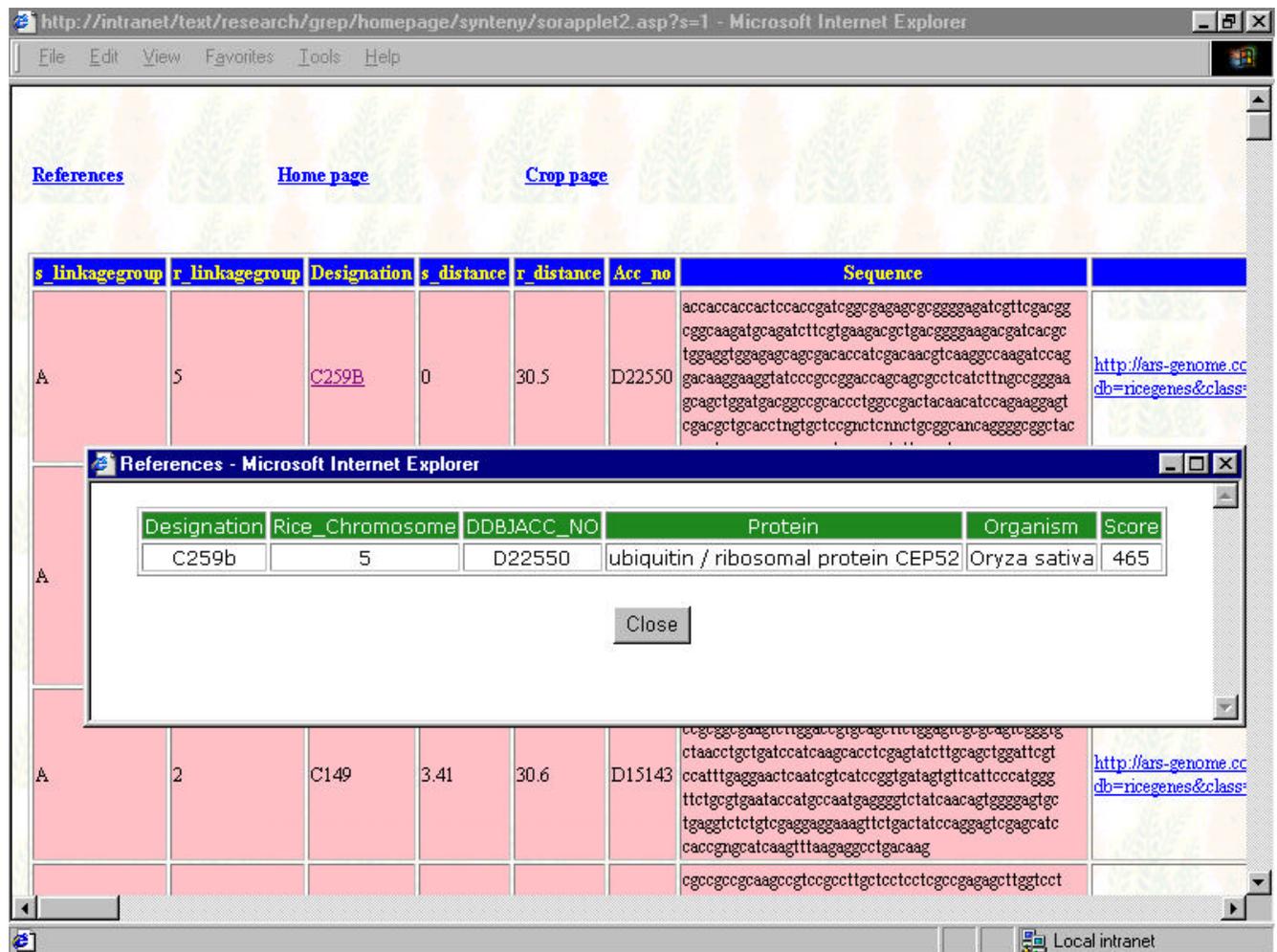


Figure 4. Detailed information of sorghum linkage group A and the BLAST information for a maker (e.i. C259b).