

Gramene, a Tool for Grass Genomics

Doreen H. Ware², Pankaj Jaiswal², Junjian Ni, Immanuel V. Yap, Xioakang Pan, Ken Y. Clark, Leonid Teytelman, Steven C. Schmidt, Wei Zhao, Kuan Chang, Sam Cartinhour, Lincoln D. Stein, and Susan R. McCouch*

Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724 (D.H.W., X.P., K.Y.C., L.T., S.C.S., W.Z., K.C., L.D.S.); Department of Plant Breeding, Cornell University, Ithaca, New York 14853–1901 (P.J., J.N., I.V.Y., S.R.M.); and United States Department of Agriculture–Agricultural Research Service Center for Agricultural Bioinformatics, 626 Rhodes Hall, Cornell Theory Center, Ithaca, New York 14853 (S.C.)

Gramene (<http://www.gramene.org>) is a comparative genome mapping database for grasses and a community resource for rice (*Oryza sativa*). It combines a semi-automatically generated database of cereal genomic and expressed sequence tag sequences, genetic maps, map relations, and publications, with a curated database of rice mutants (genes and alleles), molecular markers, and proteins. Gramene curators read and extract detailed information from published sources, summarize that information in a structured format, and establish links to related objects both inside and outside the database, providing seamless connections between independent sources of information. Genetic, physical, and sequence-based maps of rice serve as the fundamental organizing units and provide a common denominator for moving across species and genera within the grass family. Comparative maps of rice, maize (*Zea mays*), sorghum (*Sorghum bicolor*), barley (*Hordeum vulgare*), wheat (*Triticum aestivum*), and oat (*Avena sativa*) are anchored by a set of curated correspondences. In addition to sequence-based mappings found in comparative maps and rice genome displays, Gramene makes extensive use of controlled vocabularies to describe specific biological attributes in ways that permit users to query those domains and make comparisons across taxonomic groups. Proteins are annotated for functional significance using gene ontology terms that have been adopted by numerous model species databases. Genetic variants including phenotypes are annotated using plant ontology terms common to all plants and trait ontology terms that are specific to rice. In this paper, we present a brief overview of the search tools available to the plant research community in Gramene.

Gramene is a relational database and Web site that serves as a comparative mapping database for the grasses and a community resource for rice (*Oryza sativa*; Ware et al., 2002). The U.S. Department of Agriculture (USDA) funded the database in October 2000 to better leverage the newly available rice genomic sequence (Sasaki and Burr, 2000; Yu et al., 2002) as a foundation for the study of comparative genomics of agriculturally important members of the grass family. The first release of Gramene to the community was made in January 2002. Gramene provides an integrated resource of comparative genetic and physical maps between rice and other grasses based upon orthologous sequences and brings structure to the wealth of genetic and phenotypic information available among the grasses.

In this paper, we will walk through the Gramene Web site as a way to familiarize the reader with the main types of information available in Gramene. In addition to interactive Web-based access, all of the

raw data sets and annotations are downloadable in bulk form from the Gramene FTP site. Researchers wishing to create mirrors of Gramene or to reuse the Gramene database and visualizations can obtain the Gramene software under open source terms.

MAP SEARCH

The Gramene main navigation bar (Fig. 1A), found on all of the Gramene pages, provides explicit links to various database search tools, submission forms, documentation, and access to data sets found within the Web site. We begin by exploring the grass genetic and comparative maps available under the map search link.

Genetic maps are used as entry points to identify regions of colinearity among the cereals (Ahn and Tanksley, 1993; Gale and Devos, 1998; Feuillet and Keller, 1999). Researchers typically will begin by selecting a single genetic map to display, and will then add additional maps from the same or different species, building up a comparative display (Fig. 1B). One way to enter a map display is to select a map study and a specific linkage group from a set of pull down menus. Gramene also offers a “feature search” that allows researchers to identify maps that contain a particular feature of interest such as a genetic marker. Maps can be adjusted in a number of ways,

¹ This work was supported by the USDA Initiative for Future Agriculture and Food Systems (IFAfS) (grant no. 00–52100–9622) and by a USDA–Agricultural Research Service specific cooperative agreement (grant no. 58–1907–0–041).

² These authors contributed equally to the paper.

* Corresponding author; e-mail SRM4@cornell.edu; fax 607–255–6683.

www.plantphysiol.org/cgi/doi/10.1104/pp.015248.

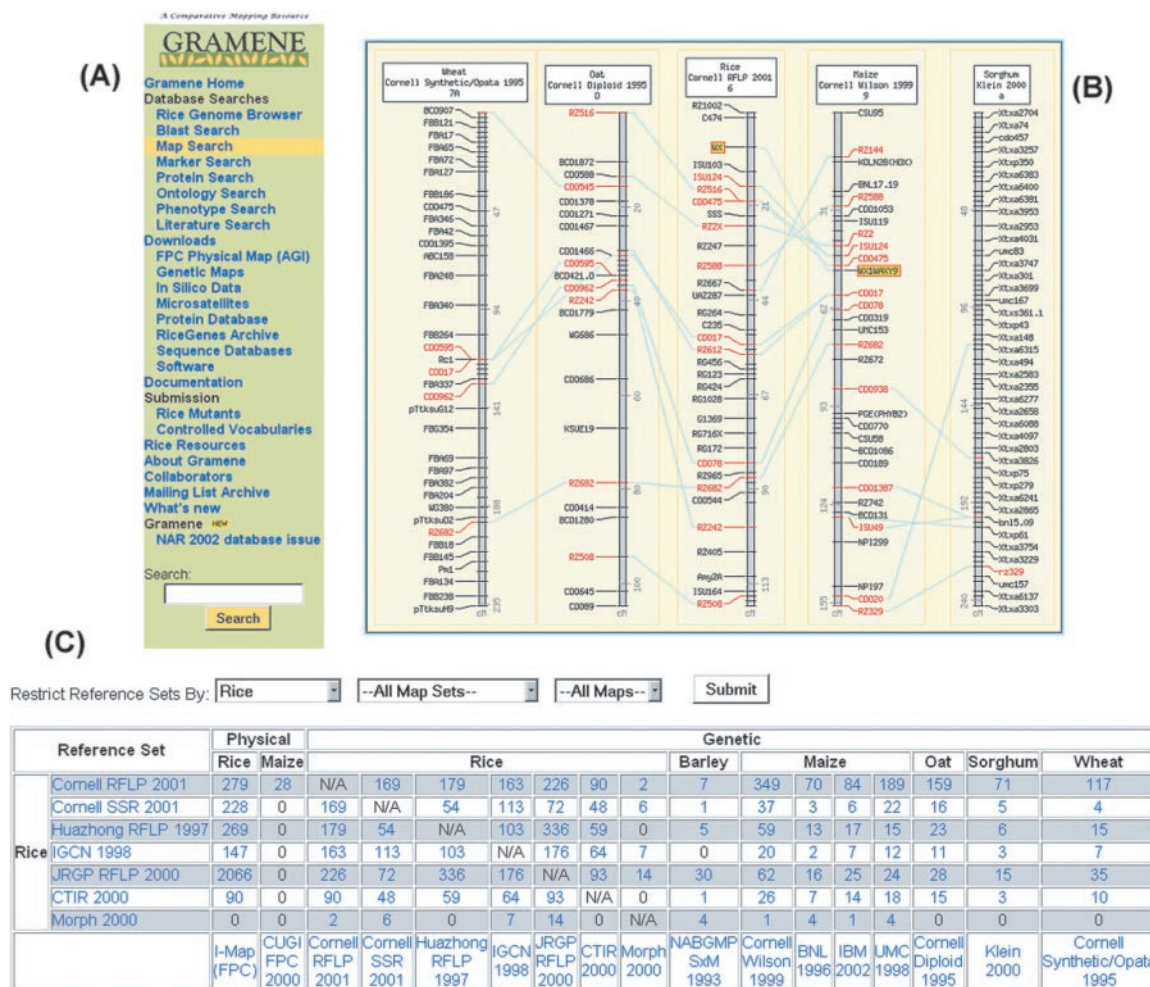


Figure 1. The Gramene comparative map viewer. A, The Gramene main navigation bar found throughout the Web site provides useful links to search tools, useful static pages such as documentation, collaborators, and mailing archives. B, Comparison between maps from different species: wheat 7A, oat (*Avena sativa*) D, rice 6, maize 9, and sorghum (*Sorghum bicolor*) a. Light blue lines connecting the adjacent maps indicate correspondences between the different maps. The rice waxy gene *wx* has been highlighted in a red box, along with the corresponding gene in maize. C, The matrix showing the number of correspondences for each pair of maps in Gramene. An on-line version of this figure can be viewed at www.plantphysiol.org.

including flipping linkage groups and expanding regions to see additional details.

Once a map view is generated, it provides the user with an interactive display containing links to external databases and internal displays within Gramene. For instance, when the researcher selects a genetic marker he/she is taken to a detail page that provides links to additional information related to that marker, such as marker assay conditions or polymorphism rate. If the genetic marker is found on several cereal maps in addition to rice, then the detail page will take the user directly to other species-specific databases, thereby facilitating the biologist's ability to traverse multiple databases and providing interoperability between plant community resources.

The comparative map tool allows multiple maps to be viewed simultaneously providing utility for both between (Fig. 1B) and within (Fig. 2A) species map

comparisons. To add maps to the display, researchers simply select additional maps from the pull-down menus. This can be repeated ad infinitum. As an alternative, researchers who are primarily interested in comparative mapping may enter the map viewer via the matrix view (Fig. 1C), which provides a table with all pairwise combinations of maps and the number of correspondences (markers) between them. By referring to the matrix, a biologist may easily see that the Cornell RFLP map study (Causse et al., 1994; Van Deynze et al., 1998; Wilson et al., 1999) contains the highest number of correspondences between available cereal maps at this time (Fig. 1C) and will likely provide the most utility for cross-species comparisons in Gramene. However, the Japanese Rice Genome Project RFLP map study (Harushima et al., 1998) provides the highest number of correspondences to the rice physical map (Fig. 1C).

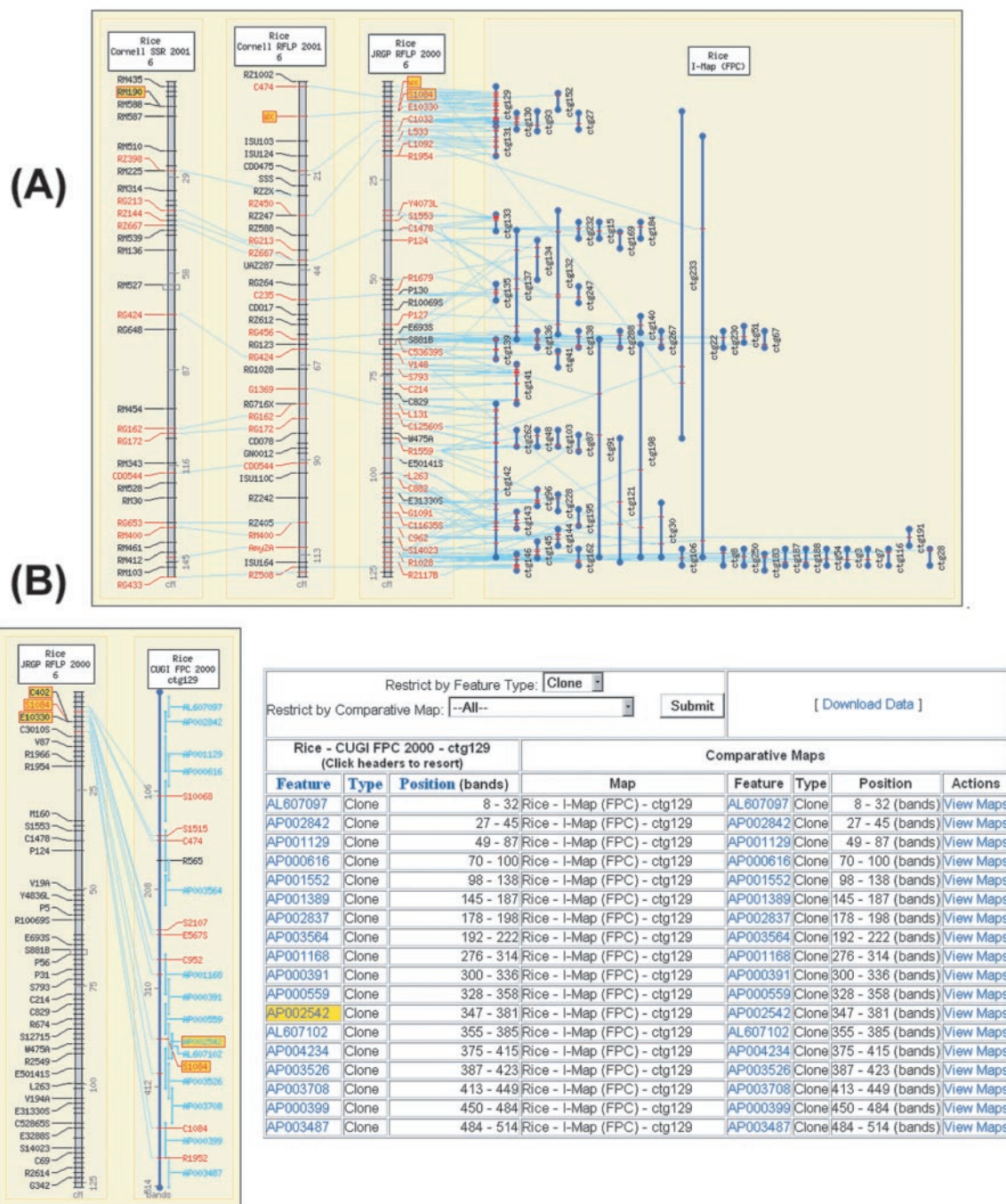


Figure 2. Genetic and physical comparative maps. A, Comparison of rice chromosome 6 mapped in three different genetic studies, and the rice I-Map. The waxy gene *wx* has been mapped on both the Japanese Rice Genome Project RFLP 2000 and Cornell RFLP 2001 maps. Highlighted SSR marker RM190, contained in the *wx* sequence (Bligh et al., 1995; Temnykh et al., 2001), suggests a correspondence with I-Map contig ctg129. B (map), Detailed view of ctg129 shows correspondences between sequenced BACs/PACs and the map for Japanese Rice Genome Project chromosome 6. The I-Map includes sequenced BACs/PACs that have been aligned to the contig by simulated digestion (Chen et al., 2002); these are shown as light blue vertical lines aligned to the appropriate positions along the contig. B (table), The table is customizable and in this example restricted to clones found on the contig. An on-line version of this figure can be viewed at www.plantphysiol.org.

Rather than limiting researchers to a single static view of map to map correspondences, our comparative map displays are dynamically generated based upon a set of automated and curated correspon-

dences at the individual marker level. The current correspondences are based on sequence similarity among markers, identity based on marker name, and curated manual correspondences. Each correspon-

dence carries explicit supporting evidence, allowing researchers to understand the strength and significance of the map correspondences and to control the type of correspondences to display.

PHYSICAL MAPS

In addition to the side-by-side genetic map display shown in Figure 1B, the Gramene map viewer allows researchers to study the relationship between genetic maps and physical (clone) maps. The current physical maps contained in Gramene are fingerprint contig (FPC) maps generated from digested genomic clones in bacterial artificial chromosome (BAC) or P1 artificial chromosome (PAC) vectors (Chen et al., 2002). In Gramene, the rice FPC map has been enhanced to show the positions of both the original hybridized RFLP markers and sequence-based markers derived from genomic and cDNA clones. In addition, we have placed all available simple sequence repeat (SSR) markers onto the FPC map. By aligning sequence-based genetic markers to the emerging rice sequence from the International Rice Genome Sequencing Project (IRGSP) we are able to produce an integrated map that relates the genetic, physical and sequence maps of rice. This integrated map, called the "I-Map" allows researchers to move from the genetic maps of maize (*Zea mays*), wheat (*Triticum aestivum*), or barley (*Hordeum vulgare*) to the rice genome sequence.

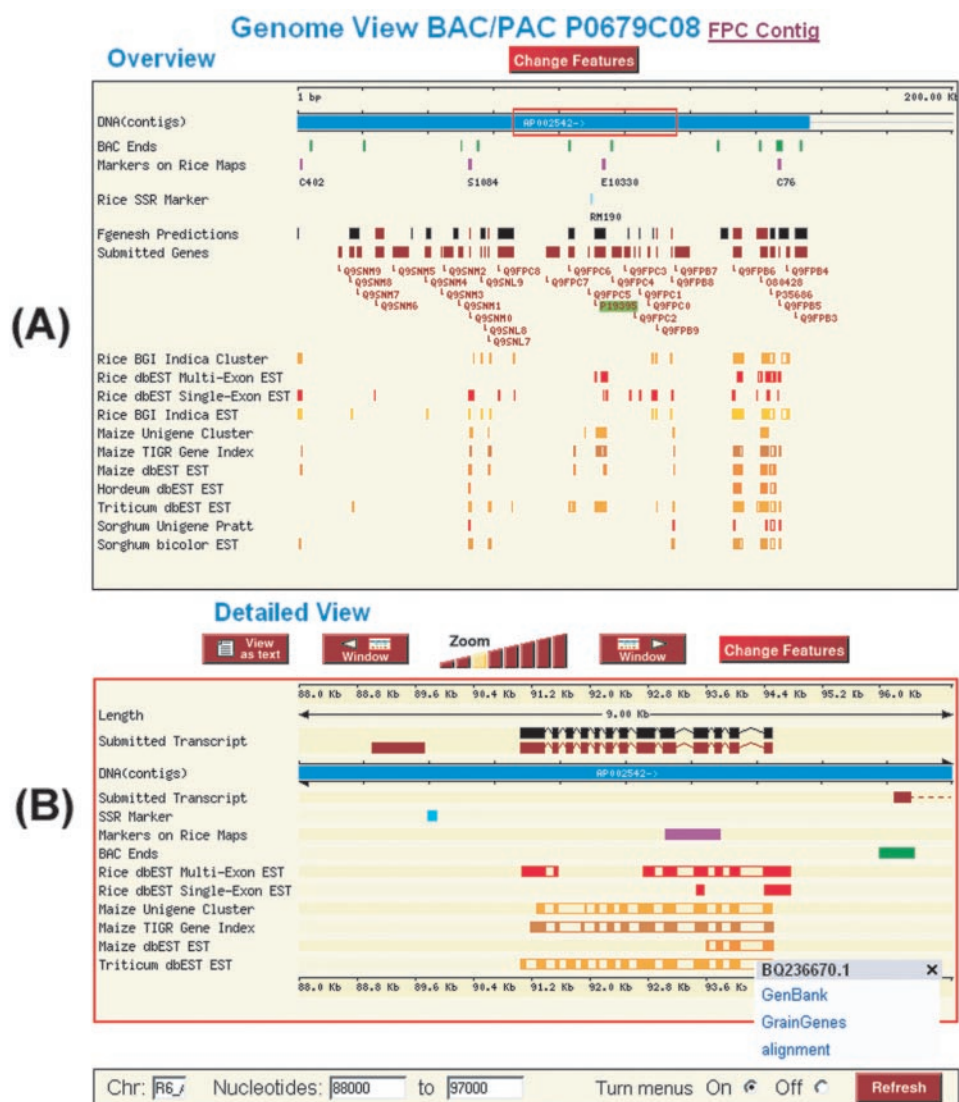
Figure 2 illustrates how researchers can use the I-Map to move between genetic, physical, and sequence maps within a species. In Figure 2A, the researcher searched for the *Waxy* locus, *wx*, and identified it in three genetic maps of rice, which are displayed side by side. The researcher then added the Clemson University Genomics Institute FPC map, thereby identifying a large contig, ctg129, that contains the region spanning *Waxy*. The researcher then clicked on ctg129, displaying the detailed view shown in Figure 2B (map), along with the tabular text summary shown in Figure 2B (table). The image and the table together provide an entry point to the sequenced clones that span the contig, and indicate that the *Waxy* locus is most likely to be located in clone AP002542 because this clone contains a genetic marker, S1084, that is close to *Waxy* on one of the genetic maps. By selecting a clone, the researcher is taken to the detailed display of the annotated sequence as described in the next section.

RICE GENOME BROWSER

The rice genome browser in Gramene is a modified version of the Ensembl (Hubbard et al., 2002) genome browser, a tool developed collaboratively by the European Bioinformatics Institute and Sanger Center for the human genome project. In Gramene, the rice genome browser contains two views: an overview

(Fig. 3A) and detailed view (Fig. 3B). This browser integrates many types of information relating to the genome, including raw features such as cross-species sequence alignments, and derived annotations, such as predicted gene structures. A partial list of the information currently displayed on the browser includes GenBank annotations from IRGSP, gene predictions based on FGENESH (Salamov and Solovyev, 2000), genetic markers from cereals, rice coding sequences, sequence tagged DNAs ranging from rice BAC ends to expressed sequence tags (ESTs) to SSRs, and clusters derived from ESTs from rice and other cereals. Sequence tagged clones and genetic markers are downloaded from GenBank whereas the clusters (theoretical contigs) were obtained from multiple sources such as The Institute for Genomics Research (Quackenbush et al., 2001) and Beijing Genomics Institute (Yu et al., 2002). All data sources are acknowledged via the "Collaborators" link on the main navigation bar, whereas annotations generated within Gramene are detailed in a materials and methods section available from the "Documentation" link on the navigation bar.

The detailed view (Fig. 3B) allows users to zoom in and out or to move along the genomic clone. When users click on a feature or annotation, they are presented with a drop-down menu (Fig. 2B) which provides links to the detail pages in Gramene as well as to external databases such as GrainGenes or MaizeDB (Polacco and Coe, 1999; Matthews et al., 2003). Users can further customize the display by showing or hiding feature tracks, changing the color or type of feature, or compacting or expanding a track. A biologist can enter the genome browser by several methods; via a text search for a genomic clone or GenBank accession number, via a comparative map view by clicking on a genomic clone or genetic marker, or via the results of a BLAST search of the rice IRGSP clones. A biologist can traverse back to the physical and genetic map views by clicking on genetic markers that are displayed in the overview and detail panels or by following links from sequenced BAC ends that have been aligned to the genome. In the case of the rice PAC clone P0679C08 (Fig. 3A), RFLP markers C76 and S1084 and SSR marker RM190 all provide links back to the genetic maps found in Figure 2A. This image also shows the important role played by rice and cereal ESTs in the genome browser. In this case, the multi-exon gene prediction from IRGSP that is shown at the top of the detailed view is supported by multiple rice and cereal ESTs. These simultaneously validate the prediction and provide a basis for establishing future cross-species relationships. A biologist can obtain detailed information about each gene prediction by clicking on the transcript and selecting "protein information" from the drop down menu that appears. This will bring the researcher to a protein detail page from the



AP002542 Version 2 BAC/PAC P0679C08 Sequencing Phase 3 Sequenced by Rice Genome Research Program FPC Contig

Figure 3. Rice genome browser. A, Overview section of rice clone PO679C08. The genomic clone is first feature in the view and is described as a DNA contig in blue and labeled with a GenBank accession number. The annotations provided in the overview represent datasets obtained from reference databases or individual researchers. All large datasets contain links back to the original database in the detail view and are available from the Gramene ftp site along with documentation of the mapping algorithm used to align the data sets. The overview and detail views are customizable by the researcher by selecting the change features button. A researcher may traverse to a physical map in the comparative map tool by selecting the FPC contig link. B, The detail view is customizable as described above and allows the user to zoom in or out, slide along the contig by selecting the window buttons, or select a view based upon nucleotide position. In the detail view, the researcher is provided with drop down menus for each feature type. In the example above, by selecting a wheat dbest EST, a user can link out directly to GenBank or to the species database GrainGenes, or obtain a Blast alignment of the EST to the genomic sequence. The detail display contains information on the mapped chromosome position, in the case of rice genomic clone, PO679C08 (GenBank accession no. AP002542) is located on rice chromosome 6 (R6) and was sequenced by the Rice Genome Research Program. An on-line version of this figure can be viewed at www.plantphysiol.org.

Gramene protein database, another useful resource for the cereal biologist.

THE USE OF ONTOLOGIES

We now discuss the curated portions of Gramene. An important organizing principle of the curated

portion of Gramene is its use of ontologies to describe information on proteins, genes, alleles, and phenotypes. Ontologies are controlled vocabularies that are shared by database communities working on different taxa. We use the gene ontology (GO; Ashburner et al., 2000) to describe the attributes of gene products in multiple species, the plant ontology (PO) to

describe anatomical features and developmental stages in diverse plant species, and the trait ontology (TO) to describe how, when and where a rice-specific character (or trait) is evaluated or quantified (Jaiswal et al., 2002). The use of controlled vocabularies allows Gramene to support complex queries and allows cross-referencing of information within and between databases (interoperability). For example, the GO is shared among multiple metazoan databases, whereas the PO is in use for Arabidopsis, maize, and rice. The use of shared ontologies allows researchers to relate genes that affect a particular life stage or organ in rice to ones that affect the corresponding life stage or organ in maize or Arabidopsis. Combined with map-based cross-species correspondence, this provides a powerful tool for researchers seeking to find candidate genes for quantitative trait loci and other traits.

PROTEIN DATABASE

The Gramene protein database is a collection of curated protein entries from all the species, subspe-

cies, and cultivars of *Oryza* sp. presently available from SPTreMBL and GenBank. The most recent updates of protein entries were imported en masse from the August 2002 release of SPTreMBL. The September 2002 release of Gramene has 12,310 entries representing 12,939 GenBank entries. Of these, 5,022 protein entries, or about 40% of the total, have annotations examined and validated by the Gramene curatorial staff.

A biologist will typically enter the protein database by searching for a protein name (e.g. granule-bound glycogen [starch] synthase), a gene name (e.g. *waxy*), a subspecies name (e.g. *indica* or *japonica*), a cultivar/strain name (e.g. Nipponbare or IR36), or a database accession number (e.g. P19395), or by typing the SPTreMBL database ID (e.g. UGST_ORYSA or Q9ZSU7). Other ways to enter the protein database are by searching for a particular protein domain (e.g. "zinc finger") or a GO term.

The protein detail page (Fig. 4A) displays the name of the protein, the gene name, the EC number (if the

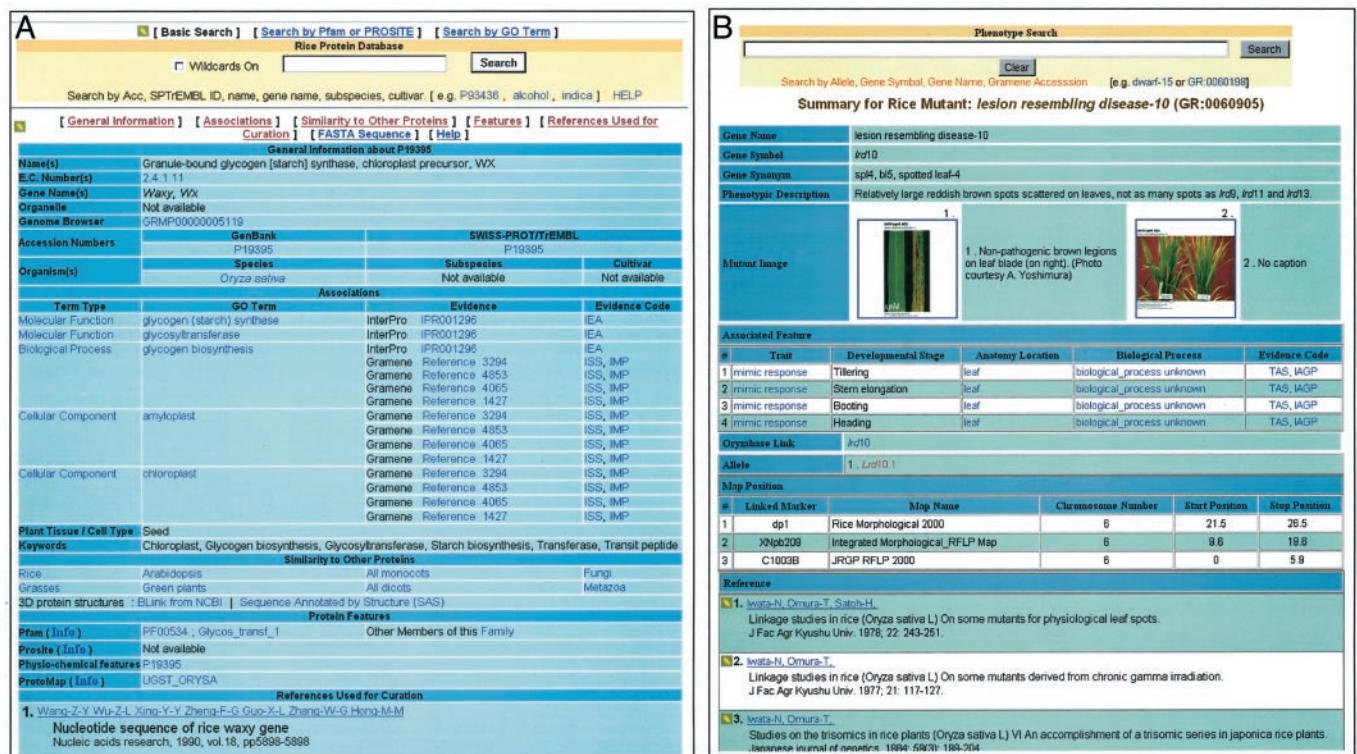


Figure 4. Detail pages from the protein and the genes and alleles databases. A, The protein database example is given for the waxy gene (accession no. P19395). It provides general information about the protein, its location on the rice genome via the genome browser, and database cross-references to GenBank and SPTreMBL. Associations are provided to GO terms for molecular function, biological process, and cellular component. The BLink to GenBank helps in finding protein homologs and orthologs. The protein feature section provides information on the physiochemical features, protein domains, protein family to which it belongs, and other features such as the transmembrane domains, targeting signal cleavage site, post translational modifications, etc. The reference section gives a list of all the citations used in curation of the respective protein entry, with an option to search for references citing this protein. The image has been modified to show all sections of the protein page. B, The main page of the rice gene and allele database is presented using *lrd10* as an example. The main fields include general information, association feature, OryzaBase link, allele, map position, references, etc. Phenotypic searching tool at the top of the page is searchable by using gene symbol, gene name, or Gramene accession number. An on-line version of this figure can be viewed at www.plantphysiol.org.

protein happens to be an enzyme and has been classified by the International Union for Biochemistry and Molecular Biology), the database cross references to GenBank and SPTreMBL, and its location on the rice genome, which links to the genome browser. The association section provides curated information on a gene product's molecular function, its role in a biological process, and its possible localization in a cellular component using the GO. The protein entry page also provides detail about the protein family to which the gene product belongs, based on the pairwise similarity to consensus domains identified by the Pfam and PROSITE databases. In the example of P19395 (Fig. 4A), the protein was assigned to a family of protein kinases (Pfam: PF00534; Glycos_transf_1). Gramene also allows the user to look for other curated entries from rice that share these features by clicking on the URL next to "Other Members of this Family," thus presenting a global overview of entries from all species and germplasm accessions in rice. In this example, there are 21 different protein entries in the database that are members of the protein family of glycosyl transferases (Pfam, PF00534; Glycos_transf_1). To further facilitate cross-species analysis, Gramene links to the NCBI BLINK service to facilitate a search for homologs and orthologs from rice, Arabidopsis, the grasses, all monocots, dicots, fungi, and/or metazoa.

The protein entry page may also provide information on in silico or experimentally analyzed features such as transmembrane domains, cleavage sites, and so forth. We currently annotate proteins using only GO terms, however in future, users will be able to search for information on the functional consequences of alterations in the protein using PO terms to describe affected plant anatomy and growth stages. As shown in Figure 4A, the supporting evidence for each assertion is documented, usually by referring to a literature reference. The same standard for documenting supporting evidence is upheld throughout the curated division of Gramene.

PHENOTYPE SEARCH

In September 2002, Gramene released a new database of rice genes and alleles, composed of classically identified phenotypic variants (marker genes) of rice. The gene and allele database is a curated resource providing collective information about publicly available mutant stocks of rice (*Oryza* sp.). It includes descriptions of phenotypic variants and alleles associated with morphological, developmental, and agronomically important phenotypes, variants of physiological or morphological characters, biochemical functions, and isozymes. The current version of the rice genes and alleles database houses information on over 400 phenotypic variants curated from the published literature.

The main entry point to the genes and alleles database is the "Phenotype search" link on the

Gramene navigation bar. A researcher can search the database using a gene symbol, gene name, or a Gramene accession number. The database also provides a full text search of the phenotype descriptions. The search result presents a comprehensive summary of all the data associated with the particular phenotypic variant in the database and provides links to associated objects. Figure 4B shows the detail page associated with the lesion mimic disease-10 locus *lrd10*. The page displays *lrd10*'s name, symbol, synonym, phenotypic description, literature references, and two images that demonstrate the phenotype. Other curated information includes information about the gene's allelic variants, the genetic background in which the alleles are observed, and the environmental conditions in which the alleles are assayed. Links to Gramene's sequence, map, and protein divisions are provided when feasible.

The gene and allele database is in the early stages of development, and more phenotypic variants will be curated and useful features will be added in the future. Users are encouraged to send suggestions and comments to help improve the utility of the phenotype search by contacting the Gramene curators by e-mail at gramene@gramene.org or by using the feedback button that appears at the bottom of every page.

THIRD-PARTY SUBMISSION OF DATA

The inclusion of curated literature is often limited by time and language constraints. To facilitate data retrieval and integration and to encourage community involvement, a Web-accessible submission form is provided to allow community submission of newly identified mutants or mapped genes that are accepted for publication or already available in scientific journals. The submission forms are available from the Gramene navigation bar under the submission link.

FUTURE ENHANCEMENTS

Over the past two years, Gramene has built an information resource for the grass community to use the rapidly emerging rice genome sequence and the reservoir of genetic maps available from the cereal community. In the next year, we aim to stay current with the emerging rice genome sequence, to continue to enhance existing tools within Gramene, and to continue the development of controlled vocabularies in collaboration with colleagues in the greater biological community. The major new effort will be to curate and incorporate phenotypes associated with quantitative trait loci studies published in the literature. This will substantially enhance the existing genes and alleles database and provide a valuable resource for researchers seeking to identify the genes associated with agronomically significant traits among the cereals.

Received September 24, 2002; returned for revision September 29, 2002; accepted October 7, 2002.

LITERATURE CITED

- Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. *Proc Natl Acad Sci USA* **90**: 7980–7984
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29
- Bligh H, Till R, Jones C (1995) A microsatellite sequence closely linked to the *Waxy* gene of *Oryza sativa*. *Euphytica* **86**: 83–85
- Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE et al. (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* **138**: 1251–1274
- Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S et al. (2002) An integrated physical and genetic map of the rice genome. *Plant Cell* **14**: 537–545
- Feuillet C, Keller B (1999) High gene density is conserved at syntenic loci of small and large grass genomes. *Proc Natl Acad Sci USA* **96**: 8265–8270
- Gale MD, Devos KM (1998) Comparative genetics in the grasses. *Proc Natl Acad Sci USA* **95**: 1971–1974
- Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A et al. (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* **148**: 479–494
- Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T et al. (2002) The Ensembl genome database project. *Nucleic Acids Res* **30**: 38–41
- Jaiswal P, Ware D, Ni J, Chang K, Zhao W, Schmidt S, Pan X, Clark K, Teytelman L, Cartinhour S et al. (2002) Gramene: development and integration of trait and gene ontologies for rice. *Comp Funct Genom* **3**: 132–136
- Matthews DE, Carollo VL, Lazo GR, Anderson OD (2003) GrainGenes, the genome database for small-grain crops. *Nucleic Acids Res* **31**: (in press)
- Polacco M, Coe E Jr (1999) MaizeDB: the maize genome database. In S Letovsky, ed, *Bioinformatics: Databases and Systems*. Kluwer Academic Publishers, Boston, pp 151–162
- Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J (2001) The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Res* **29**: 159–164
- Salamov AA, Solovyev VV (2000) Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res* **10**: 516–522
- Sasaki T, Burr B (2000) International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. *Curr Opin Plant Biol* **3**: 138–141
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* **11**: 1441–1452
- Van Deynze AE, Sorrells ME, Park WD, Ayres Park, Fu HY, Cartinhour SW, Paul E, McCouch SR (1998) Anchor probes for comparative mapping of grass genera. *Theor Appl Genet* **97**: 356–369
- Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S et al. (2002) Gramene: a resource for comparative grass genomics. *Nucleic Acids Res* **30**: 103–105
- Wilson WA, Harrington SE, Woodman WL, Lee M, Sorrells ME, McCouch SR (1999) Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoideae. *Genetics* **153**: 453–473
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* **296**: 79–92