PCP
PLANT & CELL PHYSIOLOGY

A Maize Database Resource that Captures Tissue-Specific and Subcellular-Localized Gene Expression, via Fluorescent Tags and Confocal Imaging (Maize Cell Genomics Database)

Vivek Krishnakumar[1], Yongwook Choi[1], Erin Beck[1], Qingyu Wu[2], Anding Luo[3], Anne Sylvester[3], David Jackson[2] and Agnes P. Chan[1,*]

[1]The J. Craig Venter Institute, Rockville, MD, USA
[2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
[3]University of Wyoming, Laramie, WY, USA
*Corresponding author: E-mail, achan@jcvi.org; Fax, +1-301-795-7070.
(Received September 1, 2014; Accepted November 10, 2014)

Maize is a global crop and a powerful system among grain crops for genetic and genomic studies. However, the development of novel biological tools and resources to aid in the functional identification of gene sequences is greatly needed. Towards this goal, we have developed a collection of maize marker lines for studying native gene expression in specific cell types and subcellular compartments using fluorescent proteins (FPs). To catalog FP expression, we have developed a public repository, the Maize Cell Genomics (MCG) Database, (http://maize.jcvi.org/cellgenomics), to organize a large data set of confocal images generated from the maize marker lines. To date, the collection represents major subcellular structures and also developmentally important progenitor cell populations. The resource is available to the research community, for example to study protein localization or interactions under various experimental conditions or mutant backgrounds. A subset of the marker lines can also be used to induce misexpression of target genes through a transactivation system. For future directions, the image repository can be expanded to accept new image submissions from the research community, and to perform customized large-scale computational image analysis. This community resource will provide a suite of new tools for gaining biological insights by following the dynamics of protein expression at the subcellular, cellular and tissue levels.

Keywords: Fluorescent protein • Reporter gene expression.

Abbreviation: FP, fluorescent protein.

Introduction

Maize is an important crop plant globally and is also a powerful model system for genetic and functional studies among grain crops. Novel tools and resources are needed for deciphering biological functions of genes in space and time towards understanding plant developmental processes for crop improvement. A number of resource databases have been developed for maize, e.g. the maize genome database in Gramene (Monaco et al. 2014), the Panzea database which provides allelic variations across maize cultivars (Canaran et al. 2008), and the MaizeGDB database which is a repository of genetic, genomic and functional data for the research community (Lawrence et al. 2008).

Fluorescent proteins (FPs) have been widely used in plant model systems as direct markers to follow gene expression and protein localization in plant organelles and other subcellular compartments (Brown et al. 2005, Koroleva et al. 2005, Li et al. 2006, Xiao et al. 2010, Yang et al. 2013, Leonard et al. 2014, Mano et al. 2014). The optimization of FP variants for use in plant cells and recent improvement in the production of stable maize transformants have enabled the efficient use of FPs in maize to study directly the dynamics of protein localization, interactions and functions across subcellular and cellular levels in a crop plant.

In the present post-genomics era, researchers are increasingly faced with data-intensive research and challenging tasks of not only generating a large collection of biological images, but also managing, analyzing and sharing the resulting large data sets. However, these images are often stored in proprietary file formats post-acquisition, and require specific commercial software for viewing or manipulation. In addition, similar to other types of biological data, images require association with metadata, describing experimental conditions and acquisition settings, for proper interpretation. The non-open-access platform-dependent image formats and a lack of organized metadata have been major hurdles for sharing bioimages between research groups. Because of these challenges, there have been efforts to develop image management tools centering on a common data format, to support open access and analysis of bioimages (Moore et al. 2008, Kvilekval et al. 2010).

BISQUE (Bio-Image Semantic Query User Environment) is a web-based system developed with the goal of the exchange and exploration of biological images (Kvilekval et al. 2010). BISQUE is built upon a database of images and metadata, which allows users to store, visualize, organize, analyze and share biological images. Textual annotation of images is stored as tag-value pairs to enable efficient querying of images. Image analysis modules are available for automated high-throughput analysis in broad

Plant Cell Physiol. 56(1): e12(1–7) (2015) doi:10.1093/pcp/pcu178, Advance Access publication on 27 November 2014, available online at www.pcp.oxfordjournals.org

applications, including counting nuclei, tracking microtubules, measuring seed size or following root tip development (Goff et al. 2011, Fick et al. 2013).

Here, we describe a public database of confocal images, which catalogs FP expression at the subcellular, cellular and tissue levels for a collection of maize marker lines that we have generated since 2005 as part of the National Science Foundation-funded Maize Fluorescent Protein localization project (Mohanty et al. 2009, Wu et al. 2013). The database currently contains information for over 150 marker constructs and more than 550 confocal images captured from specific subcellular compartments and tissue types to document expression of the FP marker. We expect that this community resource will provide a new set of valuable biological tools for studying the spatial and temporal expression of genes and molecular networks involved in maize development and deciphering their potential roles towards crop improvement.

Database Contents

The Maize Cell Genomics (MCG) Database is publicly accessible at http://maize.jcvi.org/cellgenomics. Two types of data are represented in the database: (i) information which describes fluorescent-tagged gene constructs used for maize marker line generation; and (ii) confocal images representing spatial and temporal expression of the fluorescent markers in the maize marker lines. A summary of reporter constructs and confocal images hosted in the MCG database is shown in **Table 1a**.

Reporter constructs carrying fluorescent markers

Reporter constructs were developed through translational and promoter fusion of fluorescent tags to selected genes. For translational fusion constructs, a fluorescent tag is inserted in-frame to the native genomic sequence of the gene of interest and expressed as a fusion to the translated protein. The native genomic sequence contains a 2–3 kb promoter region upstream of the translational start site, all the exons and introns, and a 1–2 kb terminator or 3′ region. To date, we have generated a collection of maize translational fusion constructs and reporter lines which cover all major subcellular compartments including the plasma membrane, endoplasmic reticulum, cytoplasm, nucleus, plasmodesmata and many others. A full list of subcellular structures showing fluorescent marker expression is shown in **Table 1b**. For promoter fusion constructs, the fluorescent tag is driven by a 2–3 kb promoter region upstream of the translational start site that is taken from a gene of interest to control cell/tissue-specific expression. In addition, the function of promoter fusion constructs has been expanded to serve as transactivators that can switch on target gene expression in a cell/tissue-specific manner, via the pOp–LhG4 transactivation system (Craft et al. 2005, Rutherford et al. 2005). We adapted the pOp–LhG4 transactivation system to be compatible with the MultiSite Gateway Pro kit (Invitrogen) to make constructs. Briefly, an AttR1–AttR2 Gateway cassette, and a cassette containing the bidirectional pOp promoter with the β-glucuronidase (GUS) and TagRFP-T gene on either side were sequentially cloned into the pTF101.1 binary vector, to make the destination vector, pAL10. Next, a maize native promoter and terminator from the gene of interest, and the LhG4 gene, were cloned into corresponding donor vectors, and then recombined into the destination vector. A list of tissue types showing fluorescent marker expression from each of the different construct types is shown in **Table 1c**. The overall collection of maize genes selected for fluorescent-tagging, with fluorescent tag used and Gene Ontology (GO) annotation, is shown as **Supplementary data (Table S1)**.

Confocal images for fluorescent marker expression

Expression of fluorescent tags in the maize marker lines was captured as images using confocal laser scanning microscopy (CLSM). For each marker line, plant tissues were collected from callus, seedling or mature plant, and screened for FP expression. The transgenic plants with the expected expression patterns were selected based on the promoter specificity reported in previous publications or patents.

Database Functions

Gene constructs for generating marker lines and confocal images are stored in the MCG database using the BISQUE image database management system as the database backend. Data objects for gene constructs consist of a description of the maize gene or promoter selected for FP fusion, and the type and location of the FP and PCR primer sequences used for building the gene–FP fusion construct. Each reporter construct is linked to one or more confocal image, when available, derived from the corresponding marker lines. Data objects for confocal images consist of a description of the maize specimen including Plant Ontology terms (Cooper et al. 2013) to describe tissue source, localization to subcellular compartments, marker construct used, experimenter, etc. Imaging parameters embedded in the original confocal image files (e.g. filters, channels, image size, etc.) are automatically captured and stored in the image data object.

For the research community to identify maize marker lines of interest, the gene constructs and confocal images are searchable via the MCG web site, or the Maize BISQUE database. Reporter constructs are searchable on the MCG web site based on common gene name or public maize gene ID (e.g. GRMZM2G016439) (**Fig. 1a**). A BLAST server (Altschul et al. 1990) is available for querying the list of selected tagged genes using gene sequences. A sequence browser, JBrowse (Skinner et al. 2009), has been set up to provide a graphical display of the location of the fluorescent tag in the target gene, via mapped primer locations (**Fig. 1b**). To promote sharing of the MCG resources, the reporter constructs are also linked to plant community databases Gramene (Monaco et al. 2014) and MaizeGDB (Lawrence et al. 2008) via their Ensembl (Flicek et al. 2014) and GBrowse (Stein et al. 2002) browsers, respectively, for the Maize B73 genome (**Fig. 1b**). Confocal images can be queried per gene construct via the MCG web site, or the Maize BISQUE data organizer based on controlled

Table 1 A summary of maize marker constructs and their associated confocal images in the MCG database

(a) Four types of maize marker constructs have been created including gene–FP (translational fusion of FP with the native coding region and gene regulatory sequence), promoter–FP fusion, promoter–LhG4 fusion and pOp reporter. Confocal images were collected from the FP-tagged maize marker lines

	Gene–FP fusion	Promoter–FP fusion	pOp–LhG4 transactivation		Total
			Promoter–LhG4 fusion	pOp reporter	
No. of constructs	105	22	27	4	158
No. of images	469	60	23	Not applicable	552

(b) A list of subcellular structures showing fluorescent marker expression, and associated marker constructs and images available

Subcellular structures	No. of constructs	No. of images
Actin filament	1	4
Cell wall	4	23
Chloroplast or amyloplast	6	57
Cytoplasm	20	107
Golgi apparatus	1	1
Microtubules	4	22
Mitochondria	2	8
Nucleus	28	173
Perinuclear	2	9
Peroxisomes	1	4
Plasma membrane	10	45
Plasmodesmata	2	19
Protein bodies	1	2
Ribosomes	2	3
Rough endoplasmic reticulum	12	85
Smooth endoplasmic reticulum	12	83
Tonoplast	1	3
Unknown	3	9
Vacuoles	3	31
Vesicles between Golgi and cell wall	7	31
Vesicles between rough ER and Golgi	6	20

(c) A list of tissue types showing fluorescent marker expression, and associated marker constructs and images available

Tissues	No. of constructs	No. of images
Aleurone layer (PO:0005360)	1	1
Cultured callus (PO:0000009)	4	9
Ear (PO:0020136)	10	37
Ear meristem (PO:0009109)	1	1
Ear spikelet (PO:0006320)	1	2
Ear spikelet meristem (PO:0006378)	2	2
Embryo (PO:0009009)	6	15
Endosperm (PO:0009089)	1	2

(continued)

Table 1 Continued

(c) A list of tissue types showing fluorescent marker expression, and associated marker constructs and images available

Tissues	No. of constructs	No. of images
Inflorescence (PO:0009049)	5	13
Inflorescence bract of ear (PO:0006337)	4	4
Inflorescence meristem (PO:0000230)	4	4
Lateral root (PO:0020121)	8	19
Leaf (PO:0009025)	33	134
Leaf epidermis (PO:0006016)	22	97
Leaf mesophyll (PO:0005645)	7	11
Leaf vascular system (PO:0000036)	4	6
Prop root (PO:0000044)	1	2
Root (PO:0009005)	21	35
Root apical meristem (PO:0020147)	1	1
Scutellum (PO:0020110)	1	1
Seed (PO:0009010)	2	2
Shoot (PO:0009006)	2	2
Shoot apical meristem (PO:0020148)	7	11
Spikelet meristem (PO:0006327)	2	5
Tassel (PO:0020126)	11	34
Tassel meristem (PO:0009106)	5	20
Tassel spikelet meristem (PO:0006377)	4	8

vocabularies such as tissue source, subcellular localization, etc. (Fig. 2).

Database Implementation

One of the goals of the Maize Fluorescent Protein project was to create a community resource of maize marker lines (Mohanty et al. 2009, Wu et al. 2013). During the initial phase of the project, confocal images were stored as raw image formats, such as JPEG or PNG. These file formats did not support the common z-stacks format of confocal images captured from multiple depths of a specimen. We have since adopted the BISQUE image management platform as our underlying framework for storing and managing confocal images and associated metadata (Kvilekval et al. 2010). BISQUE has been developed to support visualization, high-throughput analysis and sharing of bioimages. It uses an open source bioimage format and supports more than 100 file formats derived from different imaging platforms. In the MCG database, confocal images and associated metadata are uploaded to the Maize BISQUE database. PHP-based web pages at MCG are then used to retrieve and display data content from the Maize BISQUE database backend.

Conclusion and Future Direction

The MCG Database hosts a large collection of confocal images collected from maize marker lines that we have generated. MCG is a public web-based platform for direct retrieval and visualization of fluorescent marker expression and localization

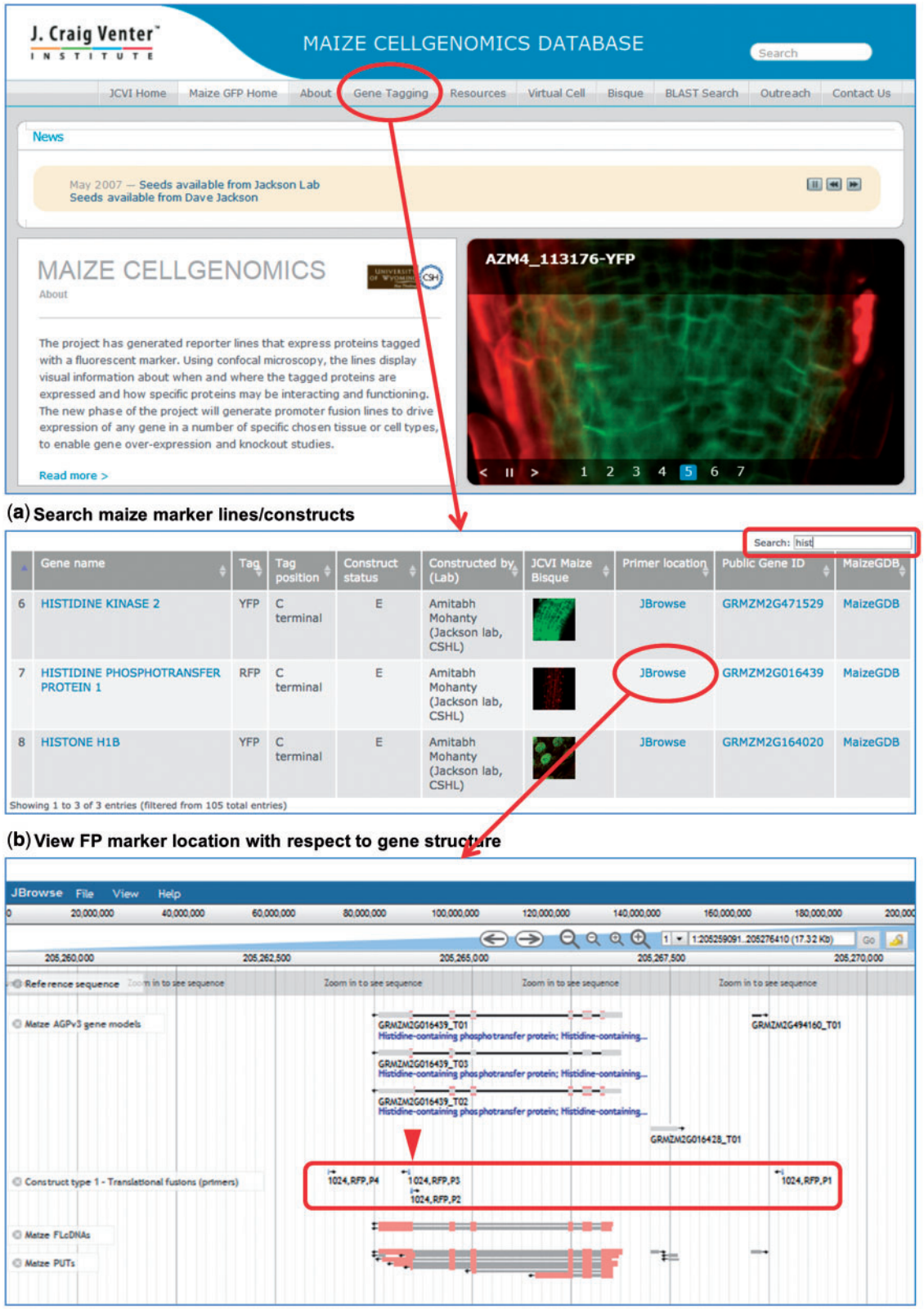

Fig. 1 The Maize Cell Genomics Database web site. (a) Searching of gene constructs based on user-provided keywords. For example, entering the keyword 'hist' in the search box will list three gene constructs: histidine kinase, histidine phosphotransfer protein 1 and histone H1B. The link out to JBrowse is highlighted with a red circle. The reporter constructs are also linked to plant community databases, Gramene and MaizeGDB, to promote resource sharing. (b) Viewing of PCR primer locations relative to gene structure, via JBrowse. The fluorescent tag is located between PCR primers P2 and P3 in gene–FP translational fusion constructs. In the example shown, the primer track is shown with a red rectangle. The gene is transcribed on the reverse strand, and the FP tag is indicated with a red arrowhead located at the beginning of the sixth exon.

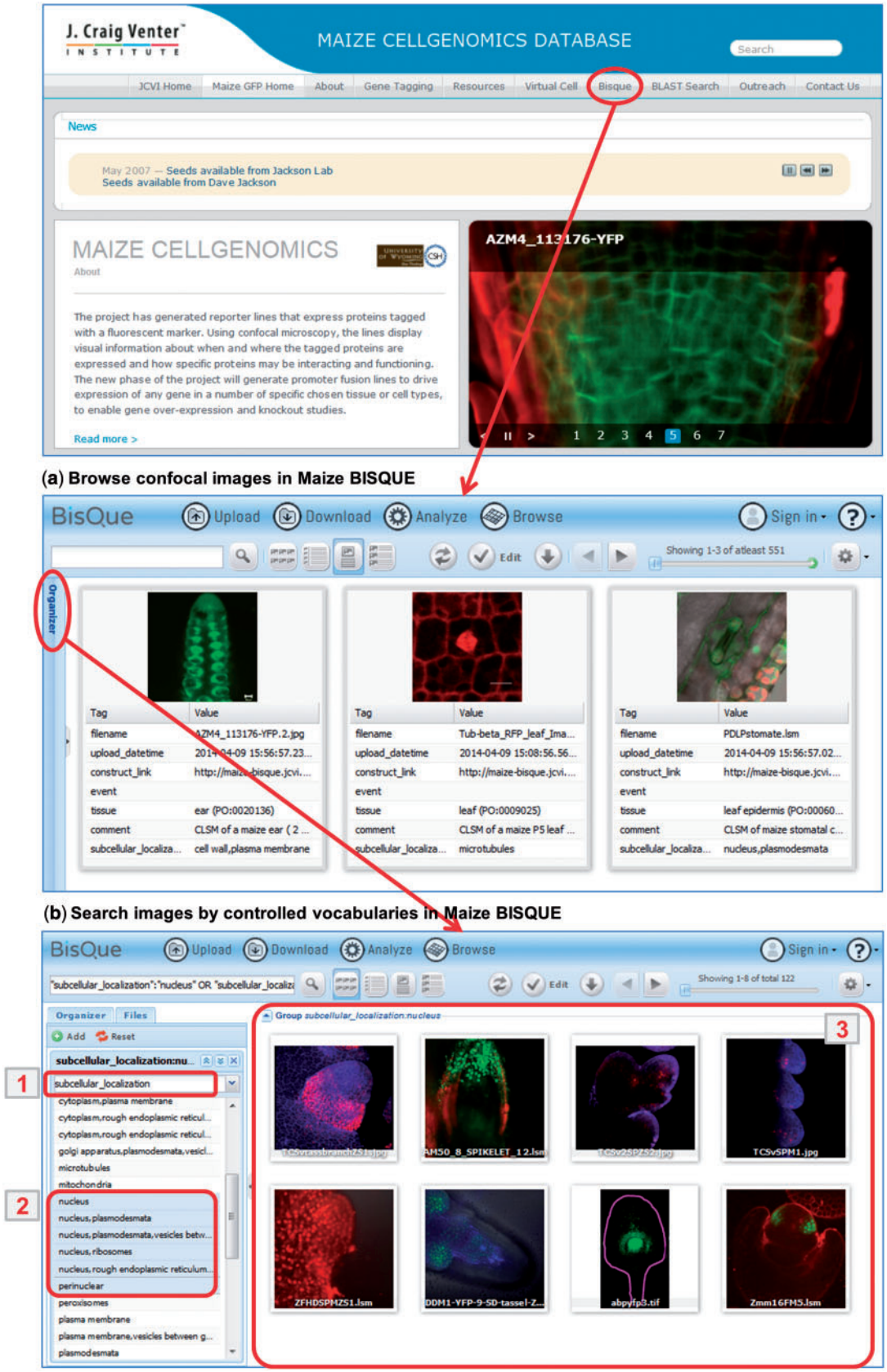

Fig. 2 Confocal images in the Maize BISQUE database. (a) Browsing confocal images in Maize BISQUE, with metadata that describe tissue types and subcellular locations (in Maize BISQUE, select from the top menu 'Browse' > 'Images'). (b) Using the data organizer (left-hand panel) in Maize BISQUE to search for confocal images based on controlled vocabularies. For example, to search for all images with nuclear or perinuclear expression: (Step 1) select 'subcellular localization' from the drop down box in the data organizer, (Step 2) multi-select subcellular structures of interest, i.e. 'nucleus' and 'perinuclear', and (Step 3) view images with fluorescence expression in the selected data set. Click on individual images for a full image view and annotations for tissue types, experimenter, etc.

at the subcellular and cellular levels. The biological materials including DNA constructs and seeds are available to the research community via online request from the project web site. To date, more than 400 community requests for constructs or seeds have been processed and with materials delivered from this project.

We expect that the maize marker lines and MCG will serve as a new resource for gaining biological insights into genes of interest via the dynamics of protein expression and localization. First, the current collection of maize marker lines covers all major subcellular compartments in the plant cell (Wu et al. 2013). The marker lines can be used to follow protein expression in situ and subcellular localization under various experimental conditions or mutant backgrounds, for example in time-series analysis via live cell imaging techniques. Secondly, the marker lines can be used to study cell-type-specific gene expression using fluorescence-activated cell sorting (FACS), or to induce misexpression of target genes in the context of the pOp–LhG4 transactivation system. Thirdly, the marker lines can serve as a biological resource for identifying in vivo interacting partners by co-immunoprecipitation of chromatin or protein complexes. For example, a translational fusion of yellow fluorescent protein (YFP) and RAMOSA1, a transcriptional regulator of inflorescence development in maize, was used to perform Chip-Seq experiments, which resulted in the identification of gene regulatory modules involved in inflorescence meristem determinacy specific to grasses (Eveland et al. 2014). Similarly, a translational fusion of YFP and COMPACT PLANT2 (CT2), a predicted GTP-binding protein, was used to demonstrate its subcellular localization to the plasma membrane, and also its interaction with a cell surface receptor protein, FEA2, through co-immunoprecipitation. Together, these analyses identified a new component of a signaling pathway involved in stem cell determinacy and shoot growth (Bommert et al. 2013).

To our knowledge, while image data for whole plants are available in various plant databases such as the MaizeGDB (Lawrence et al. 2008), the OryzaBase (Kurata and Yamazaki 2006) and the RIKEN Arabidopsis Genome Encyclopedia II (RARGE II) (Akiyama et al. 2014), only a limited number of databases exist for plant images at the cellular level (Brown et al. 2005, Koroleva et al. 2005, Li et al. 2006, Xiao et al. 2010, Mano et al. 2014). The use of BISQUE as an image management tool has enabled the systematic and scalable organization of the large volume of confocal images generated for the maize marker lines, and direct online public access of the images by the research community. By serving as a database for both images and metadata, BISQUE has provided a unified framework and opened up future opportunities for performing high-throughput computational analysis of maize confocal images to study the dynamics of labeled organelles and more specifically individual FP molecules under different experimental conditions.

Biological images represent a rich source of biological information that can be mined and analyzed computationally. One of the goals of MCG is to provide a public database of confocal images that describe FP expression patterns in maize marker lines, to foster research collaboration and sharing of biological resources. A potential future direction is to enrich the data content of MCG by encouraging submission of image data from the research community that describe existing or new maize marker lines, and linking to other image resources, such as iPlant BISQUE, to broaden dissemination of plant image data and enable scientific discoveries (Goff et al. 2011).

Supplementary data

Supplementary data are available at PCP online.

Funding

This work was supported by the National Science Foundation grants DBI-0501862 and IOS-1027445 to D.J., A.S. and A.P.C.

Acknowledgments

The authors would like to thank Kristian Kvilekval and Dmitry Fedorov of the BISQUE development team at the University of California Santa Barbara for generous assistance and discussion for setting up the JCVI Maize BISQUE, the JCVI Information Technology department for infrastructure support, and Jose Gutierrez-Marcos and Paula McSteen for advice on the project.

Disclosures

The authors have no conflicts of interest to declare.

References

Akiyama, K., Kurotani, A., Iida, K., Kuromori, T., Shinozaki, K. and Sakurai, T. (2014) RARGE II: an integrated phenotype database of Arabidopsis mutant traits using a controlled vocabulary. *Plant Cell Physiol.* 55: e4.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.

Bommert, P., Je, B.I., Goldshmidt, A. and Jackson, D. (2013) The maize Galpha gene COMPACT PLANT2 functions in CLAVATA signalling to control shoot meristem size. *Nature* 502: 555–558.

Brown, J.W., Shaw, P.J., Shaw, P. and Marshall, D.F. (2005) Arabidopsis nucleolar protein database (AtNoPDB). *Nucleic Acids Res.* 33: D633–D636.

Canaran, P., Buckler, E.S., Glaubitz, J.C., Stein, L., Sun, Q., Zhao, W. et al. (2008) Panzea: an update on new content and features. *Nucleic Acids Res.* 36: D1041–D1043.

Cooper, L., Walls, R.L., Elser, J., Gandolfo, M.A., Stevenson, D.W., Smith, B. et al. (2013) The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol.* 54: e1.

Craft, J., Samalova, M., Baroux, C., Townley, H., Martinez, A., Jepson, I. et al. (2005) New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in Arabidopsis. *Plant J.* 41: 899–918.

Eveland, A.L., Goldshmidt, A., Pautler, M., Morohashi, K., Liseron-Monfils, C., Lewis, M.W. et al. (2014) Regulatory modules controlling maize inflorescence architecture. *Genome Res.* 24: 431–443.

Fick, R.H.J., Fedorov, D., Roeder, A.H.K. and Manjunath, B.S. (2013) Simultaneous cell tracking and image alignment in 3D CLSM imagery

of growing Arabidopsis thaliana sepals. *In* IEEE 10th International Symposium on Biomedical Imaging (ISBI). pp. 914–917.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S. et al. (2014) Ensembl 2014. *Nucleic Acids Res.* 42: D749–D755.

Goff, S.A., Vaughn, M., McKay, S., Lyons, E., Stapleton, A.E., Gessler, D. et al. (2011) The iPlant Collaborative: cyberinfrastructure for plant biology. *Front. Plant Sci.* 2: 34.

Koroleva, O.A., Tomlinson, M.L., Leader, D., Shaw, P. and Doonan, J.H. (2005) High-throughput protein localization in Arabidopsis using Agrobacterium-mediated transient expression of GFP–ORF fusions. *Plant J.* 41: 162–174.

Kurata, N. and Yamazaki, Y. (2006) Oryzabase. An integrated biological and genome information database for rice. *Plant Physiol.* 140: 12–17.

Kvilekval, K., Fedorov, D., Obara, B., Singh, A. and Manjunath, B.S. (2010) Bisque: a platform for bioimage analysis and management. *Bioinformatics* 26: 544–552.

Lawrence, C.J., Harper, L.C., Schaeffer, M.L., Sen, T.Z., Seigfried, T.E. and Campbell, D.A. (2008) MaizeGDB: the maize model organism database for basic, translational, and applied research. *Int. J. Plant Genomics* 2008: 496957.

Leonard, A., Holloway, B., Guo, M., Rupe, M., Yu, G., Beatty, M. et al. (2014) tassel-less1 encodes a boron channel protein required for inflorescence development in maize. *Plant Cell Physiol.* 55: 1044–1054.

Li, S., Ehrhardt, D.W. and Rhee, S.Y. (2006) Systematic analysis of Arabidopsis organelles and a protein localization database for facilitating fluorescent tagging of full-length Arabidopsis proteins. *Plant Physiol.* 141: 527–539.

Mano, S., Nakamura, T., Kondo, M., Miwa, T., Nishikawa, S., Mimura, T. et al. (2014) The Plant Organelles Database 3 (PODB3) update 2014: integrating electron micrographs and new options for plant organelle research. *Plant Cell Physiol.* 55: e1.

Mohanty, A., Luo, A., DeBlasio, S., Ling, X., Yang, Y., Tuthill, D.E. et al. (2009) Advancing cell biology and functional genomics in maize using fluorescent protein-tagged lines. *Plant Physiol.* 149: 601–605.

Monaco, M.K., Stein, J., Naithani, S., Wei, S., Dharmawardhana, P., Kumari, S. et al. (2014) Gramene 2013: comparative plant genomics resources. *Nucleic Acids Res.* 42: D1193–D1199.

Moore, J., Allan, C., Burel, J.M., Loranger, B., MacDonald, D., Monk, J. et al. (2008) Open tools for storage and management of quantitative image data. *Methods Cell Biol.* 85: 555–570.

Rutherford, S., Brandizzi, F., Townley, H., Craft, J., Wang, Y., Jepson, I. et al. (2005) Improved transcriptional activators and their use in mis-expression traps in Arabidopsis. *Plant J.* 43: 769–788.

Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J. and Holmes, I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.* 19: 1630–1638.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A. et al. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.* 12: 1599–1610.

Wu, Q., Luo, A., Zadrozny, T., Sylvester, A. and Jackson, D. (2013) Fluorescent protein marker lines in maize: generation and applications. *Int. J. Dev. Biol.* 57: 535–543.

Xiao, Y.L., Redman, J.C., Monaghan, E.L., Zhuang, J., Underwood, B.A., Moskal, W.A. et al. (2010) High throughput generation of promoter reporter (GFP) transgenic lines of low expressing genes in Arabidopsis and analysis of their expression patterns. *Plant Methods* 6: 18.

Yang, Y.G., Lv, W.T., Li, M.J., Wang, B., Sun, D.M. and Deng, X. (2013) Maize membrane-bound transcription factor Zmbzip17 is a key regulator in the cross-talk of ER quality control and ABA signaling. *Plant Cell Physiol.* 54: 2020–2033.