BIOINFORMATICS APPLICATIONS NOTE

Vol. 27 no. 3 2011, pages 435–436
doi:10.1093/bioinformatics/btq654

Databases and ontologies

Advance Access publication December 5, 2010

MASISH: a database for gene expression in maize seeds

M. Miquel[1], I. López-Ribera[1], M. Ràmia[2], S. Casillas[2], A. Barbadilla[2] and C. M. Vicient[1,*]

[1]Department of Molecular Genetics, Centre for Research in Agricultural Genomics, CSIC (IRTA-UAB), Jordi Girona, 18, 08034 Barcelona and [2]Plataforma Bioinformàtica de la UAB-Hospitals, Institut de Biotecnologia i de Biomedicina i Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

Associate Editor: Alfonso Valencia

ABSTRACT

Grass seeds are complex organs composed by multiple tissues and cell types that develop coordinately to produce a viable embryo. The identification of genes involved in seed development is of great interest, but systematic spatial analyses of gene expression on maize seeds at the cell level have not yet been performed. MASISH is an online database holding information for gene expression spatial patterns in maize seeds based on *in situ* hybridization experiments. The web-based query interface allows the execution of gene queries and provides hybridization images, published references and information of the analyzed genes.

Availability: http://masish.uab.cat/.

Contact: cvsgmp@cid.csic.es

Received on April 26, 2010; revised on November 19, 2010; accepted on November 23, 2010

The maize kernel is classified botanically as a caryopsis. In consequence, it is a fruit composed by one seed and the remnants of the seed coats and nucellus, and is permanently enclosed in the pericarp. The endosperm occupies most of the seed and is basically a storage organ that accumulates starch and proteins. The aleurone layer is part of the endosperm and consists in a continuous layer of large cubical cells, which accumulate protein and lipid granules and surrounds most of the endosperm. In the area of the pedicel, which connects the seed to the mother plant, the cells adopt a special morphology, typical of transfer cells and form the basal transfer cell layer. The embryo consists of an embryonic axis and a single cotyledon, which is called the scutellum. The embryo axis is formed by the plumule, covered by the coleoptile and the radicle, covered by coleorhiza. All these organs are almost completely surrounded by the scutellum, an organ whose major function is to accumulate nutrient reserves, mainly lipids and proteins. A single layer of cells directly in contact with the endosperm, which is called the scutellar epithelium, is important in the digestion and transport of the nutrients from the endosperm to the embryo axis during germination. Both endosperm and embryo derive from the fusion of gametes, but while the embryo is derived from the fertilized egg, triploid endosperm is derived from fertilized polar nuclei. Surrounding the endosperm and embryo lays the pericarp, a protective organ derived from the maternal tissues (more information at http://masish.uab.cat/masish/images/maizeseedanatomy.pdf).

*To whom correspondence should be addressed.

Full genome sequencing allows the identification of the complete catalog of genes in a species. However, the roles of a high proportion of these genes remain unknown. The description of temporal and spatial gene expression patterns is a first step in the determination of the functional roles of the genes. Microarrays are a useful tool to study gene expression in a tissue but, because it integrates data from all cell types used for RNA extraction, most spatial information is lost. On the other hand, it requires the existence of a microarray chip containing all, or at least most of the genes in the genome, which is not yet available for many plant species. Laser microdissection allows obtaining RNA samples from more homogeneous cell types and has been successfully used, for example, in the study of the maize shoot apical meristem or pericycle cell transcriptomes (Brooks *et al*., 2009; Dembinsky *et al*., 2007). The combination of laser microdissection with microarray hybridization is a very promising technique that has been allowed to produce, for example, a cell type transcriptome atlas in rice (Jiao *et al*., 2009). A complementary strategy is to define the specific cellular expression patterns by *in situ* hybridization (ISH). ISH is a high-resolution technique for the analysis of gene expression that allows determining the steady state concentration of a specific mRNA at the cellular level. ISH has been successfully applied in maize (Fontanet and Vicient, 2008). Large-scale surveys of gene expression patterns based on ISH analyses have been performed for animals as, for example, *Drosophila melanogaster* (Zhao *et al*., 2010), mouse (Richardson *et al*., 2010) and other mammals (Olsen *et al*., 2004). In plants, it has only been performed for wheat (Drea *et al*., 2005).

MASISH (Maize Seed In Situ Hybridization; http://masish.uab.cat/) consists of a database of patterns of gene expression in maize seeds based on ISH and a web-based interface that enables users to search and display images and related gene annotations (Fig. 1). The database contains two types of entries. Approximately half of the entries arise from the systematic search of PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) for publications on maize seed ISH. ISH images of the other genes were generated by the authors themselves as previously described (Fontanet and Vicient, 2008) using digoxygenin-labeled RNA probes and paraffin-embedded samples (and labeled as 'unpublished'). Some minor differences were introduced in the protocol in order to facilitate the systematic analyses. For example, hybridizations with two different probes were performed in a single slide using silicone insulators (Grace Bio-Labs JTR20-1.0). The probes used were obtained from clones of a cDNA library derived from scutellum dissected 30 days after pollination (Genebank entries AM937797 to AM938286). Control images (hybridization with sense probe) were performed for some

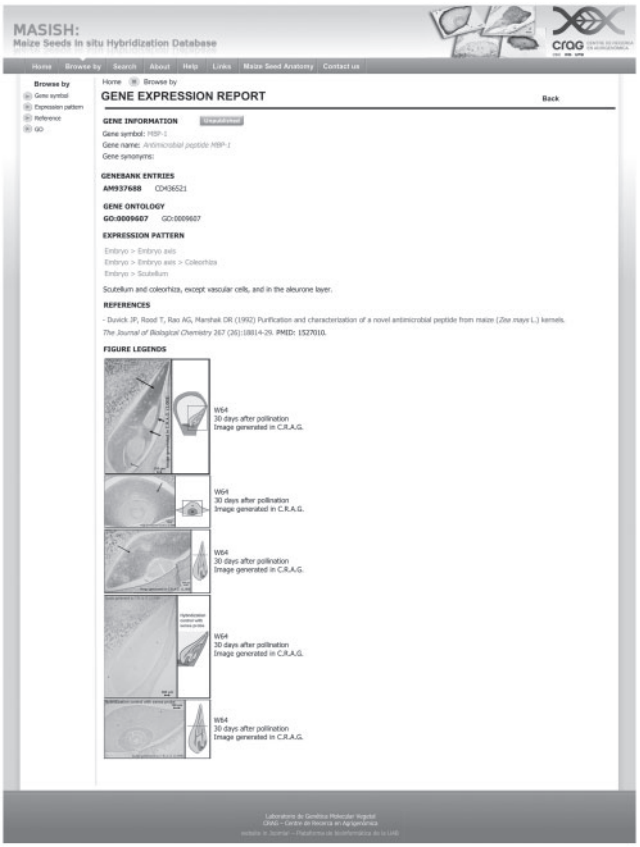

Fig. 1. Screenshot of a representative MASISH gene expression report showing the gene information on the top and the ISH images on the bottom. Images can be expanded by clicking on them.

of the genes, showing no hybridization. These control images are also available in the report of the corresponding genes, while genes still lacking controls are clearly labeled as 'unverified'. The database includes information concerning the possible functions of genes, GenBank entries and related publications. This database is expandable, so that the final aim is to obtain spatial information of the expression of all those genes that are transcribed in the maize seed at any point of development.

MASISH database is a relational MySQL database installed in a Linux Ubuntu Server. The searching interface is written in PHP and is wrapped in a Joomla web site that further includes information on the project and tools for the authors to easily maintain and update the database. MASISH database interfaces with GenBank database (http://www.ncbi.nlm.nih.gov/) and the AMIGO Gene Ontology web site (http://amigo.geneontology.org/cgi-bin/amigo/go.cgi) to retrieve information from these databases and link out gene reports. Information is accessed through browsing the MASISH database by gene name, expression pattern, reference or Gene Ontology ID and through text-based queries by gene symbol/name/synonym, PubMed ID, authors or GenBank accession, either including or excluding unpublished and/or unverified genes. Images for each gene are displayed initially as a set of thumbnails. Each thumbnail image is linked to the original full-sized image that can be downloaded.

Great efforts are being made in the recent times in the identification of all genes and their roles in maize. A second draft of the maize genome has been recently released, which contains most of the maize B73 genome sequence (http://www.maizesequence.org/; B73 RefGen_v2). Bioinformatics tools have been recently developed in order to identify and annotate the maize genome (Andorf *et al*., 2010; Montalent and Joets, 2010). High-throughput pyrosequencing of Expressed Sequence Tags (ESTs) has been applied in order to determine a maize transcriptome atlas at the organ level (Vega-Arreguín *et al*., 2009) and will also be of great help in the maize gene annotation. In this context, the establishment of an expandable expression pattern database such as MASISH provides a body of knowledge to suggest hypotheses that can facilitate the identification of the functions of genes that encode proteins of currently unknown function. In addition, it may also be helpful for comparing gene expression patterns of homologous genes in different species.

ACKNOWLEDGEMENTS

The authors would like to thank Emilio Centeno for help designing and implementing the MASISH database.

Funding: Spanish Ministry of Science and Innovation (BIO2007-64791 and 200720I037); Centre CONSOLIDER on Agrigenomics; Xarxa de Referencia en Biotecnologia of the Generalitat de Catalunya; Plataforma Bioinformàtica de la UAB-Hospitals (to A.B.).

Conflict of interest: none declared.

REFERENCES

Andorf,C.M. *et al*. (2010) The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps. *Bioinformatics,* **26**, 434–436.

Brooks,L. 3rd *et al.* (2009) Microdissection of shoot meristem functional domains. *PLoS Genet*., **5,** e1000476.

Dembinsky,D. *et al.* (2007) Transcriptomic and proteomic analyses of pericycle cells of the maize primary root. *Plant Physiol*., **145**, 575–588.

Drea,S. *et al.* (2005) Systematic spatial analysis of gene expression during wheat caryopsis development. *Plant Cell*, **17**, 2172–2185.

Fontanet,P. and Vicient,C.M. (2008) Maize embryogenesis. *Methods Mol. Biol.*, **427**, 17–29.

Jiao,Y. *et al*. (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat. Genet.*, **41**, 258–263.

Montalent,P. and Joets,J. (2010) EuGène-maize: a web site for maize gene prediction. *Bioinformatics*, **26**, 1254–1255.

Olsen,L. *et al*. (2004) CVD: the intestinal crypt/villus in situ hybridization database. *Bioinformatics*, **20**, 1327–1328.

Richardson,L. *et al*. (2010) EMAGE mouse embryo spatial gene expression database: 2010 update. *Nucleic Acids Res*., **38**, D703–D709.

Vega-Arreguín,J.C. *et al*. (2009) Deep sampling of the Palomero maize transcriptome by a high throughput strategy of pyrosequencing. *BMC Genomics*, **10**, 299.

Zhao,J. *et al.* (2010) FlyTED: the drosophila testis gene expression database. *Nucleic Acids Res.* **38**, D710–D715.