

(مقاله مروری)

تحقیقات غلات

دوره نهم / شماره اول / بهار ۱۳۹۸ (۱۰۱-۸۳)

پایگاه‌های داده‌های اصلی جهت مطالعه بیان ژن‌ها در گیاهان (مطالعه موردی: برنج)

مجتبی کردرستمی^۱ و بابک ربیعی^{۲*}

تاریخ پذیرش: ۱۳۹۸/۲/۱۵

تاریخ دریافت: ۱۳۹۷/۱۰/۲۶

چکیده

برنج به‌عنوان یکی از مهم‌ترین گیاهان زراعی، کوچک‌ترین ژنوم را در بین غلات دارد و به‌عنوان یک گیاه مدل برای مطالعات ژنتیکی مطرح است. کوچک بودن ژنوم این گیاه باعث شده است که مطالعات جامعی روی آن انجام شود و در نتیجه مقدار زیادی اطلاعات به‌دست آید. جمع‌آوری و نگهداری مناسب این اطلاعات (داده‌ها) و مدیریت هر چه بهتر داده‌های حاصل از آزمایش‌های مختلف در قالب یک پایگاه داده‌ها، جهت دسترسی محققین به آن‌ها برای جلوگیری از دوباره‌کاری و نیز مقایسه نتایج خود با نتایج سایر محققین، از اهمیت ویژه‌ای برخوردار است. برای رسیدن به این هدف، می‌توان از بیوانفورماتیک کمک شایان توجهی گرفت. بنابراین، ایجاد و توسعه پایگاه‌های داده‌های تخصصی و استفاده از ابزارهای بیوانفورماتیک برای پردازش داده‌ها، سازمان‌دهی کارآمد، تجزیه و تحلیل و تجسم آن‌ها امری ضروری است. در این مقاله، پایگاه‌های داده‌های عمده برای مطالعه بیان ژن‌ها در سه سطح RNA، پروتئین و متابولوم در گیاه برنج بررسی می‌شود و ویژگی‌های این پایگاه‌ها برای هر سطح مورد بحث و بررسی قرار می‌گیرد.

واژه‌های کلیدی: بیوانفورماتیک، پایگاه‌های داده مبتنی بر وب، مسیرهای متابولیک

۱- دانش‌آموخته دکتری، گروه بیوتکنولوژی کشاورزی، دانشکده علوم کشاورزی، دانشگاه گیلان، رشت، ایران

۲- استاد، گروه مهندسی تولید و ژنتیک گیاهی، دانشکده علوم کشاورزی، دانشگاه گیلان، رشت، ایران

* نویسنده مسئول: rabiei@guilan.ac.ir

مقدمه

روش‌های سنجش بیان ژن

تفاوت در بیان ژن‌ها در همه مراحل رشد موجودات زنده رخ می‌دهد. شناخت این تفاوت‌ها منجر به شناسایی نقش ژن و فرایندهای مولکولی که در شرایط گوناگون درون سیستم‌های زیستی رخ می‌دهد، می‌شود (Moody, 2001). روش‌های متفاوتی برای درک چگونگی تظاهر ژن‌ها و تفاوت آن در شرایط مختلف وجود دارد. در سطح اول بررسی بیان ژن می‌توان به دوره‌گیری کاهنده (Sargent and Dawid, 1983)، نوردن بلات (Callard et al., 1994)، کتابخانه EST (Adams et al., 1991)، تجزیه ترتیبی تظاهر ژن (Velculescu et al., 1995)، بیان افتراقی در PCR (Liang and Pardee, 1992) و تجزیه ریزآرایه اشاره کرد. در سطح دوم تجزیه بیان ژن می‌توان به بررسی پروتئین‌ها اشاره کرد. وسترن‌بلات، الکتروفورز ژل دوبعدی و ایمنی‌سنجی از راه‌های سنجش بیان ژن در سطح پروتئین هستند (Grigoriev, 2001). تجزیه و تحلیل تغییرات پساترجمه‌ای نیز سومین سطح بررسی بیان ژن می‌باشد. ایمنی‌سنجی و طیف‌سنجی جرمی نیز در این سطح قرار می‌گیرند (Zhang et al., 2018).

پایگاه‌های داده و اهمیت آن‌ها در زیست‌شناسی

پس از بررسی بیان یک ژن هدف در نمونه مورد مطالعه، قدم بعدی بررسی پایگاه‌های داده‌ها و مقایسه نتایج خود با نتایج سایر محققین است. برای رسیدن به این هدف، بیوانفورماتیک کمک شایان توجهی خواهد کرد. بر طبق نتایج مطالعات مختلف تا دو دهه گذشته، بیوانفورماتیک در هیچ یک از شاخه‌های علم قرار نداشت، اما اکنون به‌عنوان یک عنصر کلیدی در توسعه علوم پایه و کاربردی مطرح است (Chen et al., 2017). تعداد زیاد گزارش‌های مربوط به توالی اسیدهای نوکلئیک و پروتئین‌های تولید شده با استفاده از فناوری‌های قدیمی و جدید موجب شد که در سال ۱۹۷۹، ۳۵ دانشمند در دانشگاه راکفلر با یکدیگر ملاقات و فوریت ایجاد یک پایگاه داده بین‌المللی اسیدهای نوکلئیک را پیشنهاد دهند (Smith, 1990; Strasser, 2008). این پیشنهاد منجر به ایجاد پایگاه داده Genbank شد. پس از آن، شاهد فرآیندهای مشابه در زمینه ایجاد پایگاه داده‌ها بودیم. با این وجود، با افزایش توالی‌های نوکلئوتیدی، در سال ۱۹۸۷، یک پایگاه جامع به نام همکاری پایگاه داده بین‌المللی توالی نوکلئوتیدی (INSDC) ایجاد شد. در حال حاضر، این پایگاه از پایگاه

در گیاهان عالی، بسیاری از ژن‌ها برای ایجاد اندام‌های تخصصی و پاسخ به محیط‌های مختلف به‌طور هماهنگ عمل می‌کنند و بیان می‌شوند (Tardieu, 2013). ژن‌های تنظیم‌کننده‌ای که تمام این فعالیت‌ها را هماهنگ می‌کنند، شامل کسر بزرگی از ژنوم گیاهی می‌شوند (Riechmann and Ratcliffe, 2000). جهش در ژن‌های تنظیم‌کننده، یکی از دلایل اهلی‌سازی و به‌نژادی گیاهان زراعی مهم می‌باشد و تغییر در تنظیم بیان ژن نیز به‌عنوان یک اصل مرکزی برای تکامل گیاهان شناخته شده است (Clark et al., 2006). از نظر لغوی، تجزیه و تحلیل بیان ژن تعریف ساده‌ای دارد و عبارت از مطالعه نحوه تولید ژن‌های کارکردی، گونه‌های RNA کارکردی یا محصولات پروتئینی از طریق رونویسی ژن‌ها است. مطالعه تنظیم بیان ژن، بینش ما را در مورد فرآیندهای سلولی طبیعی (مانند تمایز) و یا غیرطبیعی (مانند پاسخ گیاه به انواع تنش‌ها) افزایش می‌دهد (Barrett et al., 2012). محققان ممکن است آنالیز بیان ژن را در هر یک از سطوح مختلف تنظیم بیان ژن (رونویسی، پساترجمه، ترجمه و پساترجمه) انجام دهند (Deans and Maggert, 2015). فرآیند رونویسی می‌تواند به‌روش‌های مختلف تنظیم شود. فرآیندهای تنظیم رونویسی اغلب در آزمایش‌های تجزیه بیان ژن مورد مطالعه و دستکاری قرار می‌گیرند. اتصال پروتئین‌های تنظیم‌کننده به مکان‌های اتصال DNA، مستقیم‌ترین روش برای تنظیم رونویسی است (Geertz and Maerkl, 2010). تأثیر تنظیمات اپی‌ژنتیک (مانند اثر متیلاسیون DNA بر بیان ژن)، نیز به‌عنوان یک ابزار قدرتمند برای بررسی میزان بیان ژن عنوان شده است (Tollefsbol, 2011). پس از رونویسی، در RNA یوکاریوتی به‌طور معمول شاهد حذف توالی اینترون و اضافه شدن دم پلی A هستیم. در این سطح از پساترجمه، ثبات RNA تأثیر قابل توجهی بر بیان ژن کارکردی (تولید پروتئین کارکردی) دارد (Gruber et al., 2014). RNAهای خاموشگر کوچک (siRNA) شامل مولکول‌های اسید نوکلئیک دو رشته‌ای هستند که در مسیر خاموش‌سازی RNA شرکت کرده و به‌طور معمول با کاهش فعالیت یک ژن، بیان ژن‌های خاص را تنظیم می‌کنند. تا کنون، نحوه تنظیم این RNAها به‌طور کامل درک نشده است. در حال حاضر نیز بیش‌تر مطالعات حول تأثیر miRNAs بر تنظیم بیان ژن یوکاریوت‌ها استوار است (Ghildiyal and Zamore, 2009).

که با این که غلات مختلف از یک جد مشترک منشأ گرفته‌اند، ولی اختلافات زیادی در اندازه ژنوم آن‌ها دیده می‌شود. دلایلی دیگر نیز وجود دارد تا برنج ماده آزمایشی خوبی برای کاربرد نشانگرهای مولکولی باشد. برنج دوّمین غله مهمّ دنیا می‌باشد و غذای بیش از ۵۰ درصد مردم جهان را تأمین می‌کند. به علت قابلیت رشد در شرایط محیطی متفاوت، امکان بررسی و مطالعه آن در محیط‌های مختلف وجود دارد. تنوع ارقام زراعی برنج بیش از سایر گیاهان زراعی است و مجموعه ژرم‌پلاسم برنج دارای بیش از ۱۲۰/۰۰۰ واریته می‌باشد. این نکات از دلایلی است که باعث شده محققین تحقیقات خود را روی این گیاه متمرکز سازند (Second, 1991).

پایگاه‌های داده اصلی بیان ژن در گیاهان

تا به حال پایگاه‌های داده مبتنی بر وب مختلفی جهت بررسی داده‌های بیان ژن و تحلیل داده‌های ریزآرایه در گیاهان ایجاد و معرفی شده‌اند (جدول ۱). سه پایگاه داده شامل GENEVESTIGATOR، DIURNAL و EMBL-EBI Gene Expression Atlas از پایگاه‌های معروف و پرکاربرد برای بیان ژن در گیاهان هستند که در ادامه به معرفی آن‌ها می‌پردازیم.

جدول ۱- پایگاه‌های داده مبتنی بر وب جهت بررسی داده‌های بیان ژن و تحلیل داده‌های ریزآرایه در گیاهان

Table 1. Web-based databases to study gene expression data and microarray data analysis in plants

Database	Website address	Reference
TAIR	http://www.arabidopsis.org	Garcia-Hernandez <i>et al.</i> , 2002; Rhee <i>et al.</i> , 2003
ArrayExpress	http://www.ebi.ac.uk/arrayexpress	Brazma <i>et al.</i> , 2003
GENEVESTIGATOR	https://www.genevestigator.ethz.ch	Zimmermann <i>et al.</i> , 2004
NASC Arrays	http://affymetrix.arabidopsis.info	Craigon <i>et al.</i> , 2004
AtGenExpress	http://www.weigelworld.org/resources/microarray/AtGenExpress	Schmid <i>et al.</i> , 2005
Botany array resource (BAR)	http://www.bar.utoronto.ca	Toufighi <i>et al.</i> , 2005
PlexDB/Barleybase	http://www.plexdb.org	Shen <i>et al.</i> , 2005
GEO	http://www.ncbi.nlm.nih.gov/geo	Barrett <i>et al.</i> , 2007
DIURNAL	http://diurnal.mocklerlab.org	Mockler <i>et al.</i> , 2007
EMBL-EBI Gene Expression Atlas	https://www.ebi.ac.uk/gxa	Petryszak <i>et al.</i> , 2015
VirtualPlant	http://virtualplant.org	Katari <i>et al.</i> , 2010

بیان ژن در انواع موجودات زنده فراهم کرده است. GENEVESTIGATOR دارای چهار بخش اصلی شامل: (۱) یک پایگاه داده، (۲) یک سرور برنامه، (۳) برنامه کاربردی Java و (۴) وبسایت برای پشتیبانی و مدیریت

پایگاه‌های داده‌های اصلی جهت مطالعه بیان ژن‌ها در گیاهان

داده DNA ژاپن (DDBJ) در موسسه ملی ژنتیک می‌شیمی ژاپن، پایگاه داده‌های نوکلئوتید اروپا - موسسه بیوانفورماتیک اروپا (EMBL) در کمبریج انگلستان و GenBank در مرکز ملی اطلاعات بیوتکنولوژی (NCBI) در بتسدا (Bethesda, MD) ایالات متحده آمریکا تشکیل شده است (Karsch-Mizrachi *et al.*, 2011; Cochrane *et al.*, 2015). از زمان ایجاد این پایگاه (INSDC)، توالی‌های نوکلئوتیدی از آزمایش‌های بیولوژیک در سرتاسر دنیا تهیه شده و در آن ذخیره می‌شود و اجازه دسترسی نامحدود به هر شخص یا موسسه‌ای را به این داده‌ها می‌دهند (Cochrane *et al.*, 2015). روز به روز به میزان داده‌های این پایگاه‌ها افزوده می‌شود. دسترسی راحت به این داده‌ها، شناخت هر چه بیش‌تر کارکرد ژن‌ها را باعث شده و در نهایت باعث برآورده شدن هر چه بیش‌تر منافع بشر شده است.

برنج: یک غله مدل

در بین غلات، برنج دارای کوچک‌ترین ژنوم از نظر اندازه (۳۴۰ میلیون جفت باز) می‌باشد و این امر سبب شده است تا برنج به عنوان گیاه زراعی ایده‌آل در تحقیقات تکاملی غلات و گونه‌های مهم زراعی مورد استفاده فراوان قرار گیرد. بنت و اسمیت (Bennet and Smith, 1991) اعلام کردند

پایگاه داده GENEVESTIGATOR

این پایگاه داده به آدرس <http://www.genevestigator.ethz.ch/> قابل دسترسی است و ابزار قدرتمندی برای زیست‌شناسان جهت مطالعه

شبانه‌روزی و تغذیه گیاهی. چهار روش تجزیه و تحلیل پایه در این پایگاه داده در دسترس هستند:

(۱) تجزیه و تحلیل یک آزمایش منفرد برای پیدا کردن ژن‌ها یا نمونه‌های خاص.

(۲) شناسایی شرایطی که بر بیان یک ژن خاص تاثیر می‌گذارند.

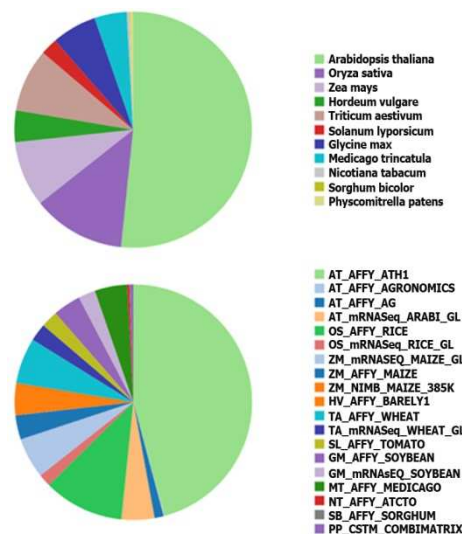
(۳) شناسایی ژن‌های خاصی که در شرایط ویژه بیان می‌شوند.

(۴) پیدا کردن هماهنگی و شباهت بین بیان ژن‌ها و یا شرایط تنظیم‌کننده بیان آن‌ها.

با استفاده از این پایگاه داده، به راحتی می‌توان شرایط بیولوژیک تنظیم‌کننده بیان یک ژن ویژه را پیدا کرد. برای مثال، می‌توان فهمید که میزان بیان آن ژن در اندام و بافت‌های مختلف چقدر است و عوامل داخلی یا خارجی چقدر بر بیان آن‌ها تاثیر می‌گذارند. همچنین می‌توان فهمید که نحوه بیان این ژن در ژنوتیپ‌های ویژه‌ای به چه صورت بوده است؟ از طرفی می‌توان نتایج خود را با هزاران آزمایش دیگر مقایسه و نتایج مشابه با نتایج مطالعه خود را پیدا کرد. همچنین می‌توان ژن‌های مختلف را در بافت‌ها، تیمارهای مختلف هورمونی و شیمیایی و ژنوتیپ‌های مختلف خوشه‌بندی و ارتباط آن‌ها با یکدیگر را تحت شرایط محیطی مختلف ارزیابی کرد. با بررسی داده‌های بیانی این پایگاه به راحتی می‌توان آزمایش مورد نظر را طراحی کرد که این امر منجر به طراحی ایده‌های جدید و پشتیبانی از تصمیمات آگاهانه می‌شود.

کاربر می‌باشد. به طور خاص، GENEVESTIGATOR یک سیستم چند سطحی Java / Java / Mysql است که در آن یک سطح کاربری جاوا با انواع مختلف سرور ارتباط برقرار می‌کند تا اطلاعات درخواست شده را بازیابی کند (Hruz et al., 2008). سیستم پایگاه داده پشت سرور برنامه با دقت خاصی طراحی شده تا اطلاعات چند موجود زنده را هم‌زمان در خود جای داده و امکان تغییر سریع بین موجودات زنده مختلف را فراهم کند. این پایگاه داده، هزاران آزمایش ریزآرایه و RNA-Seq را به خوبی با یکدیگر ادغام و به راحتی بیان ژن‌ها را در محیط‌های مختلف زیستی مانند انواع بیماری‌ها، داروها، بافت‌ها، سرطان‌ها، سلول‌ها یا ژنوتیپ‌ها بررسی می‌کند. تنوع بالای آزمایش‌های موجود در این پایگاه این امکان را برای شما فراهم می‌کند که ژن‌ها یا داده‌های خود را با طیف وسیعی از پرونده‌ها و مجموعه داده‌های موجود مقایسه کنید. این مجموعه دارای اطلاعات یک‌پارچه برای بیش از ۵۰۰ بیماری از ۱۸ حوزه درمانی اصلی است (<http://www.genevestigator.ethz.ch/>).

در مورد گیاهان، این پایگاه داده اطلاعات بیانی انواع مختلف گیاهان و تیمارهای مرتبط با گیاهان را طی ده سال به شرح زیر جمع‌آوری کرده است (شکل ۱). گیاهان مدل: آرابیدوپسیس و یونجه؛ گیاهان تک‌لپه: ذرت، برنج، گندم، جو، سورگوم؛ گیاهان دولپه: سویا، گوجه فرنگی، تنباکو؛ انواع اندام، بافت و سلول گیاهی؛ انواع ژنوتیپ‌های گیاهی؛ انواع تیمارهای هورمونی و شیمیایی؛ انواع تنش‌های زیستی و غیرزیستی؛ انواع مراحل رشد و نمو؛ انواع زمان‌بندی



شکل ۱- انواع گیاهان، ژن‌ها و تیمارهای مورد مطالعه در پایگاه GENEVESTIGATOR
Figure 1. Types of plants, genes and treatments studied at GENEVESTIGATOR

پایگاه داده DIURNAL

این پایگاه داده به آدرس <http://diurnal.mocklerlab.org> قابل دسترس است و شامل تعداد زیادی از آزمایش‌های بیان آرایه روی ژنوم‌های مختلف (آرابیدوپسیس، برنج، صنوبر و ...) تحت شرایط استاندارد نور و دما می‌باشد (Mockler *et al.*, 2007). همان‌طور که می‌دانیم، موجودات زنده تغییرات محیطی روزانه را در نور و دما تجربه می‌کنند که میزان این تغییرات با فصل و عرض جغرافیایی متفاوت هستند. همچنین موجودات زنده دارای یک ساعت زیستی درونی (Internal Biological Clock) هستند که به موجود زنده این اطمینان را می‌دهد که تمامی فرایندهای زیستی در داخل بدن آن‌ها به‌طور مناسب و مطابق با تغییرات روزانه محیطی انجام می‌گیرند (Woelfle *et al.*, 2004; Dodd *et al.*, 2005). تغییرات محیطی روزانه بر بیان ژن‌ها تاثیر می‌گذارند و این تاثیر، اساس فعالیت‌های زیستی مختص به ساعت و روز در گیاهان است. در پایگاه داده DIURNAL علاوه بر الگوهای بیانی معمول، نتایج آزمایش‌های مختلف نیز قرار گرفته‌اند که با مقایسه آن‌ها با نتایج خود می‌توان به الگوهای بیانی ژن‌ها تحت تیمارهای مختلف آزمایشی پی برد. در این پایگاه، انواع شرایط و تیمارهای آزمایشی ویژه روی گیاهان اعمال شده و به‌صورت مخفف بیان شده‌اند که عبارت‌اند از:

- ۱- LDHC: نور ۱۲ ساعت، تاریکی ۱۲ ساعت / گرما ۱۲ ساعت، سرما ۱۲ ساعت.
- ۲- LLHC: نور ۲۴ ساعت / گرما ۱۲ ساعت، سرما ۱۲ ساعت.
- ۳- LDHH: نور ۱۲ ساعت، تاریکی ۱۲ ساعت / گرما ۲۴ ساعت.
- ۴- LL_LDHC یا LLHH_LDHC: نمونه‌ها ابتدا تحت شرایط نور (۱۲ ساعت)، تاریکی (۱۲ ساعت) / گرما (۱۲ ساعت) و سرما (۱۲ ساعت) قرار گرفته و سپس در معرض نور (۲۴ ساعت) / گرما (۲۴ ساعت) قرار می‌گیرند.
- ۵- LL_LLHC یا LLHH_LLHC: نمونه‌ها ابتدا تحت شرایط نور (۲۴ ساعت) / گرما (۱۲ ساعت) و سرما (۱۲ ساعت) قرار گرفته و سپس در معرض نور (۲۴ ساعت) / گرما (۲۴ ساعت) قرار می‌گیرند.
- ۶- LL_LDHH یا LLHH_LDHH: نمونه‌ها ابتدا تحت شرایط نور (۱۲ ساعت)، تاریکی (۱۲ ساعت) / گرما

(۲۴ ساعت) قرار گرفته و سپس در معرض نور (۲۴ ساعت) / گرما (۲۴ ساعت) قرار می‌گیرند.

لازم به ذکر است که این تیمارها بیش‌تر با حروف مخفف نشان داده شده‌اند و بر اساس توضیحات این پایگاه داده، مخفف‌ها برای گونه‌های مختلف گیاهی مورد مطالعه در این پایگاه می‌تواند متفاوت باشد. یکی از ابزارهای مهم در این پایگاه داده، HAYSTACK نام دارد که به کاربر اجازه می‌دهد تا به داده‌های حجیم بیانی ریزآرایه توسط جستجوی الگوهای ویژه‌ای که توسط کاربر تعریف می‌شوند، دسترسی یابد. در حقیقت HAYSTACK (http://haystack.cgrb.oregonstate.edu) یک برنامه برای شناسایی ژن‌هایی است که سطوح بیان آن‌ها به‌طور مشابه در همه نمونه‌های یک مجموعه ریزآرایه تغییر می‌کنند. روش‌های مختلفی برای سازمان‌دهی و مطالعه الگوهای داده‌های ریزآرایه توسعه یافته است (Belacel *et al.*, 2006). یکی از این روش‌ها، روش تجزیه خوشه‌ای ژن‌ها و تیمارهای مورد مطالعه می‌باشد (روش تجزیه خوشه‌ای سلسله مراتبی، K-means، تجزیه به مولفه‌های اصلی و ...). در این برنامه از یک الگوریتم منطبق با الگو برای شناسایی ژن‌هایی که الگوهای بیانی آن‌ها متناسب با مدل تعریف شده توسط کاربر می‌باشد، استفاده می‌شود. یکی دیگر از کاربردهای HAYSTACK، کمک به محقق جهت شناسایی تمام ژن‌های موجود در یک مجموعه ریزآرایه که یک الگوی خاص بیان را در یک شرایط خاص نشان می‌دهند، می‌باشد. سومین کارایی برنامه HAYSTACK، شناسایی راه‌اندازها (Promoters) می‌باشد. این برنامه با اتصال به پایگاه‌های مهم از جمله Promoter (Toufighi *et al.*, 2005)، SIFT (Hudson and Quail, 2003)، یا MotifSampler (Thijs *et al.*, 2001, 2002) کمک شایان توجهی به محقق در شناسایی راه‌اندازها می‌کند. یک ابزار مهم دیگر در این پایگاه داده، ELEMENT است که جهت شناسایی عوامل سامان‌دهنده هم‌سو (cis-regulatory elements) به کار می‌رود. عوامل سامان‌دهنده هم‌سو به بخشی از DNA یا RNA گفته می‌شود که کار آن تنظیم بیان ژن است و روی همان مولکول DNA جای دارد. یک عامل سامان‌دهنده هم‌سو هم می‌تواند در بالادست توالی رشته رمزکننده باشد و سامان‌دهی آن را کنترل کند، هم می‌تواند در پایین‌دست آن و حتی در یکی از دو ناحیه رمزخوانی نشده قرار داشته باشد (Wittkopp and Kalay, 2012).

شبکه نشان داده می‌شوند که در این شبکه، هر ردیف نشان‌دهنده یک تیمار و هر ستون نشان‌دهنده یک ژن ویژه است. هر کدام از این ردیف و ستون‌ها با یک رنگ خاص (سبز و قرمز و سیاه) نشان داده می‌شوند. شدت این رنگ‌ها نشان‌دهنده شدت بیان یک ژن ویژه تحت یک تیمار ویژه است (Rajaram and Oono, 2010). رنگ سبز نشان‌دهنده افزایش بیان یک ژن، رنگ قرمز نشان‌دهنده کاهش بیان آن و رنگ سیاه نشان‌دهنده عدم تاثیر تیمار مربوطه در بیان ژن مزبور می‌باشد. با استفاده از ویجت جاوا اسکریپت موجود در پایگاه داده، محقق می‌تواند نقشه حرارتی ژن‌های مورد مطالعه خود را تهیه کند. این نقشه حرارتی می‌تواند یا از داده‌های موجود در پایگاه و یا از داده‌های خود محقق ساخته شود. دستورالعمل برای ساخت نقشه حرارتی از هر دو داده در پایگاه داده موجود است. برای کسب بهترین نتایج، پایگاه پیشنهاد می‌دهد که از تمامی نمونه‌های موجود در گالری پایگاه استفاده شود.

۲- Expression Atlas R Package on Bioconductor

نرم‌افزار R محیطی برای محاسبات آماری و تحلیل داده را فراهم می‌کند و به‌عنوان یک زبان استاندارد غیررسمی برای کارهای آماری و داده‌کاوی مطرح می‌شود. دلیل محبوبیت این نرم‌افزار در سال‌های اخیر در درجه اول رایگان بودن این نرم‌افزار است. همچنین این نرم‌افزار روی انواع سیستم عامل‌ها قابل نصب است. از دیگر مزایای این نرم‌افزار فراهم کردن زمینه‌ای بی‌نظیر برای نوشتن روش‌های آماری است و همچون دیگر نرم‌افزارهای باز، قابلیت توسعه دارد (Jombart, 2008). پایگاه داده Expression Atlas امکان را برای محقق فراهم می‌کند که داده‌ها و پکیج‌های آماده بررسی بیانی را برای نرم‌افزار R دانلود و به‌صورت آفلاین بیان ژن‌ها را بررسی کند. همچنین داده‌های خام RNA-seq و ریزآرایه برای دانلود و بررسی آفلاین در این پایگاه موجود است. پروتکل‌های توصیف نحوه تولید داده‌ها و تفسیر نمونه‌ها نیز در داخل پکیج دانلود شده R قرار گرفته است.

۳- iRAP: RNA-seq analysis tool

RNA-Seq یک روش نوین است که مطالعات ترانسکریپتوم را بر اساس آخرین نسل فناوری‌های توالی‌یابی انجام می‌دهد و بر بسیاری از مشکلات تکنیک ریزآرایه غلبه می‌کند. با استفاده از ابزار iRAP، توالی و کمیت رونوشت‌های RNA در یک ترانسکریپتوم تعیین شده،

این دو برنامه کاربردی در نهایت پایگاه داده DIURNAL را به‌عنوان یک مرجع مهم برای بررسی بیان ژن در گیاهان به‌ویژه برنج مطرح کرده است.

پایگاه EMBL-EBI Gene Expression Atlas

این پایگاه به آدرس <http://www.ebi.ac.uk/gxa> یک منبع علمی آزاد است که به کاربران این امکان را می‌دهد تا اطلاعات مختلفی در مورد ژن و پروتئین را در گونه‌ها و شرایط زیستی مختلف (بافت‌های مختلف، انواع سلول‌ها، مراحل رشد و ...) پیدا کنند. در سال ۲۰۰۸، موسسه بیوانفورماتیک اروپا (EMBL-EBI) یک پایگاه داده به نام Gene Expression Atlas را راه‌اندازی کرد (Petryszak et al., 2015) و چند سال بعد نیز این موسسه پایگاه داده Expression Atlas را راه‌اندازی کرد. Expression Atlas این امکان را به محقق می‌دهد که بررسی کند یک ژن در چه شرایطی بیان می‌شود و نحوه تغییر بیان آن تحت یک شرایط محیطی ویژه چگونه است؟ در این پایگاه داده، بیش از ۳۰۰۰ نتیجه آزمایش‌های بیانی مختلف (ریزآرایه و RNA-seq) از حدود ۴۰ موجود مختلف (از جمله حیوانات و گیاهان)، موجود است. اطلاعات بیانی برای هر دو مجموعه رونوشت‌های رمزکننده و غیررمزکننده ژنوم موجود هستند. همچنین اولین مجموعه داده پروتئومیکس در سال ۲۰۱۵ به این پایگاه داده اضافه شد. گیاهان مختلفی در این پایگاه داده وجود دارند که می‌توان به ذرت (۳۷ آزمایش)، برنج (۸۳ آزمایش)، آرابیدوپسیس (۵۳۳ آزمایش)، گندم (۱۵ آزمایش)، کلزا (۲۰ آزمایش) و انگور (۲۰ آزمایش) اشاره کرد. در این پایگاه داده یک نوار جستجو وجود دارد که با وارد کردن نام یک ژن می‌توان کلیه مشخصات آن را مشاهده کرد. همچنین این پایگاه از ابزارهای زیر تشکیل شده است:

۱- Expression Atlas Heatmap

در علم بیوانفورماتیک، روش‌های متعددی برای تفسیر و ترسیم اطلاعات بیان ژن وجود دارد که از آن‌ها می‌توان برای تفسیر میزان بیان ژن‌های ریزآرایه و RNA-seq استفاده کرد. یکی از این روش‌ها، نقشه حرارتی (Heat map) است. معمولاً این روش با ترکیب تجزیه خوشه‌ای، ژن‌های متعدد بیان‌شده تحت تیمارهای مختلف را گروه‌بندی می‌کند. این اطلاعات نشان می‌دهند که چه ژن‌هایی تحت تاثیر یک تیمار خاص بیان مشترک دارند (Zhao et al., 2014). در این نقشه، داده‌ها به‌صورت یک

iRAP به دقت انجام می‌شود.

میزان بیان یک ژن تحت یک تیمار ویژه مشخص و بیان افتراقی ژن‌ها تحت تیمارهای مختلف ارزیابی می‌شود. iRAP برای پردازش تمام داده‌های RNA-seq در Expression Atlas مورد استفاده قرار می‌گیرد.

پایگاه‌های داده اصلی بیان ژن در برنج

سه پایگاه ارایه شده در بالا از پایگاه‌های اصلی جهت مطالعه بیان ژن در گیاهان (به‌ویژه برنج) هستند، اما برخی از پایگاه‌ها به‌طور ویژه برای مطالعه بیان ژن در سطح RNA در گیاه برنج تهیه شده‌اند (جدول ۲). همچنین پایگاه‌هایی نیز به‌طور عمده جهت بررسی پروتئوم و متابولوم برنج وجود دارند که در جدول ۳ ارایه شده‌اند.

۴- The RNaseq-er REST API

این ابزار دسترسی آسان به نتایج تجزیه و تحلیل به‌روز و کامل داده‌های RNA-seq موجود در پایگانی نوکلئوتید اروپا (European Nucleotide Archive- ENA) را فراهم می‌کند. تجزیه و تحلیل هر توالی توسط یک تیم

جدول ۲- پایگاه‌های داده مبتنی بر وب جهت مطالعه بیان ژن در برنج

Table 2. Web-based databases to study gene expression in rice

Database	Website address	Reference
IC4R	http://ic4r.org	Zhang <i>et al.</i> , 2016
RiceFRIEND	http://ricefriend.dna.affrc.go.jp	Sato <i>et al.</i> , 2013a
RiceXPro	http://ricexpro.dna.affrc.go.jp	Sato <i>et al.</i> , 2013b
RGKbase	http://rgkbase.big.ac.cn/RGKbase	Wang <i>et al.</i> , 2013
Rice DB	http://ricedb.plantenergy.uwa.edu.au	Narsai <i>et al.</i> , 2013
TENOR	http://tenor.dna.affrc.go.jp	Kawahara <i>et al.</i> , 2016
OryzaExpress	http://plantomics.mind.meiji.ac.jp/OryzaExpress	Hamada <i>et al.</i> , 2011
Rice Expression Database	http://expression.ic4r.org	The IC4R Project Consortium, 2016
RICD	http://www.ncgr.ac.cn/ricd	Lu <i>et al.</i> , 2008
UniVIO	http://univio.psc.riken.jp	Kudo <i>et al.</i> , 2013
Rice Phylogenomic Database	http://ricephylogenomics.ucdavis.edu/index.shtml	Jung <i>et al.</i> , 2015
IntAct rice	http://www.ebi.ac.uk/ebisearch/search.ebi?db=int-act-experiments&t=Oryza+sativa	Orchard <i>et al.</i> , 2014
RiceNet	http://www.inetbio.org/ricenet	Lee <i>et al.</i> , 2015
ROAD (Rice Oligonucleotide Array Database)	http://www.ricearray.org	Cao <i>et al.</i> , 2012
RiceFOX	http://ricefox.psc.riken.jp	Sakurai <i>et al.</i> , 2011
RiceSRTFDB	http://www.nipgr.res.in/RiceSRTFDB.html	Priya and Jain, 2013

جدول ۳- پایگاه‌های داده اصلی مبتنی بر وب جهت مطالعه پروتئوم و متابولوم در برنج

Table 3. Major web-based databases to study proteome and metabolome in rice

Database	Website address	Reference
Proteome		
PRIN	http://bis.zju.edu.cn/prin	Gu <i>et al.</i> , 2011
DIPOS	http://www.riceresearch.info	Sapkota <i>et al.</i> , 2011
OryzaPG-DB	http://oryzapg.iab.keio.ac.jp	Helmy <i>et al.</i> , 2011
MCDRP	https://omictools.com/mcdrp-tool	Gour <i>et al.</i> , 2013
Metabolome		
EXPath	http://expa.th.itps.ncku.edu.tw	Chien <i>et al.</i> , 2015
RiceCyc	http://pathway.gamene.org/RICE/organism-summary?	Dharmawardhana <i>et al.</i> , 2013
OryzaCyc	http://www.plantcyc.org/databases/oryzacyc	Schläpfer <i>et al.</i> , 2017
KEGG rice	http://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_maps&org=dosa	Kanehisa <i>et al.</i> , 2013

پایگاه‌های داده اصلی ترانسکرپتوم برنج

پایگاه داده IC4R

وجود یک پایگاه داده یک پارچه که شامل انواع مختلف داده آمیکس باشد، برای محققین برنج اهمیت زیادی دارد. با توجه به وجود پایگاه‌های داده متعدد برای برنج، هیچ‌یک از آن‌ها انواع داده‌های آمیکس را به صورت یک جا با هم ادغام نکرده‌اند. با ظهور فناوری‌های توالی‌یابی، روزانه شاهد افزایش داده‌های بیانی در گیاهان مختلف (به ویژه برنج) هستیم و از این رو حضور یک پایگاه جامع که داده‌های مختلف بیانی را از پایگاه‌های داده مختلف جمع‌آوری و با هم ادغام کند، امری ضروری است. هر چند در برنج، پایگاه گرامینه (Gramene) به عنوان یک پایگاه ارزشمند برای بررسی و مطالعه ژنومیکس کارکردی مطرح است (Tabkhkar and Rabiei, 2014)، اما توانایی ادغام داده‌های آمیکس و ارایه پروفایل بیانی با وضوح بالا را ندارد (The IC4R Project Consortium, 2016). پایگاه داده IC4R به آدرس <http://ic4r.org> قابل دسترس است و به عنوان بهترین پایگاه برای ادغام انواع داده‌های آمیکس از پایگاه‌های داده مختلف مطرح می‌باشد. این پایگاه داده، بسیار پایدار بوده و داده‌های خود را از منابع مختلف از طریق انواع رابط برنامه‌نویسی وب (Web API) جمع‌آوری و در خود نگهداری می‌کند. این پایگاه به طور مداوم به روزرسانی می‌شود و در آخرین به روزرسانی خود (۲۰۱۸/۳/۲۷)، حاوی ۵۶,۲۲۱ ژن رمزکننده پروتئین، ۸۰,۰۳۸ رونوشت رمزکننده پروتئین، ۴,۳۷۳ RNA حلقوی، ۶,۲۵۹ رونوشت RNA بلند غیررمزکننده پروتئین و ۱,۵۰۳ مجموعه داده RNA-seq است. در صفحه اصلی سایت، یک جعبه جستجو وجود دارد که با وارد کردن نام ژن، پروتئین و یا رونوشت، می‌توان در مورد کلیه ویژگی‌های ژن مورد نظر اطلاعات کسب کرد. این پایگاه تعدادی جعبه ابزار اصلی دارد که در ادامه شرح داده می‌شوند:

۱- ابزار Genes

با کلیک روی این قسمت، صفحه جدیدی باز می‌شود و کلیه ژن‌های رمزکننده پروتئین برنج نشان داده می‌شود. این قسمت حدود ۵۶,۲۲۱ داده دارد که می‌توان بر اساس موارد مختلف در آن جستجو کرد. در این قسمت می‌توان یا بر اساس نام اختصاری ژن (Symbol) جستجو را انجام داد، یا بر اساس شناسه (ID) ژن در پایگاه‌های مختلف، ژن مورد نظر را پیدا کرد و یا جستجو را بر اساس کروموزومی که ژن

مورد نظر روی آن واقع شده است، انجام داد. برای هر ژن یک حرف اختصاری، کروموزوم، طول دقیق ژن بر حسب جفت باز و توضیح مختصری در مورد آن ژن داده شده است. با کلیک روی هر ژن، صفحه جدیدی باز می‌شود و در آن، مشخصات جامع‌تری از ژن، از جمله شرح مختصری از ژن، ویژگی‌های رونوشت‌های این ژن، داده‌های حاصل از آزمایش‌های RNA-seq در مورد این ژن، پروفایل بیانی این ژن در بافت‌های مختلف گیاه برنج، مرورگر ژنوم، تفسیر کارکرد این ژن در تیمارهای مختلف، هستی‌شناسی ژن (Gene ontology) و حتی نظرات محققین مختلف در مورد این ژن، ارایه شده است.

۲- ابزار LncRNAs

با کلیک روی این جعبه ابزار، صفحه جدید باز می‌شود که حدود ۶,۲۵۹ داده را در ارتباط با RNAهای بلند غیررمزکننده پروتئین در بر گرفته است. در این صفحه، شرح مختصری در مورد هر ژن از جمله مکان قرارگیری آن روی کروموزوم، طول دقیق آن، محتوای GC، محتوای پورین و تعداد اگزون ژن ارایه می‌شود. با کلیک روی نام هر ژن، یک صفحه جدید باز می‌شود که اطلاعات کامل‌تری از این نوع RNA از جمله توضیح ژن، مرورگر ژنوم، داده‌های حاصل از آزمایش‌های RNA-seq در مورد ژن، پروفایل بیانی ژن در بافت‌های مختلف گیاه برنج و قسمت نظرات محققین مختلف در مورد ژن در اختیار قرار می‌دهد.

۳- ابزار CircRNAs

با کلیک روی این جعبه ابزار، صفحه جدیدی باز می‌شود که حدود ۴,۳۷۳ داده را در ارتباط با RNA حلقوی برنج در بر گرفته است. در این صفحه، اطلاعات مختلفی در مورد ژن‌های مورد نظر، از جمله کروموزوم، طول دقیق ژن و نوع RNA حلقوی (اگزون، ناحیه بین‌ژنی و ...) در اختیار محقق قرار می‌گیرد. با کلیک روی نام هر ژن نیز صفحه جدیدی باز می‌شود که در آن، اطلاعات کامل‌تری از این نوع RNA از جمله اطلاعات پایه ژن، جزئیات دقیق‌تر ژن (شماره دسترسی، بافت و ...) و بخش نظرات محققین مختلف در مورد ژن ارایه می‌شود.

۴- ابزار HK-TS

با کلیک روی این جعبه ابزار، یک صفحه جدید باز می‌شود که حدود ۵۲,۷۲۱ داده را در ارتباط با ژن‌های خانه‌دار و ژن‌های مختص بافت (House-keeping &)

ریزآرایه حاصل از بافت‌ها و اندام‌های مختلف؛ مراحل مختلف رشدی و نموی، تیمارهای مختلف هورمونی و انواع شرایط محیطی در گیاه برنج ایجاد شده است. این پایگاه با هدف آرایه یک پلت‌فرم جامع برای شناسایی ژن‌های کارکردی مرتبط در مسیرهای مختلف بیولوژیک و یا فرآیندهای متابولیسم ایجاد شده است. این پایگاه هم قابلیت بررسی تک‌ژن‌ها را دارد و هم می‌تواند هم‌زمان چندین ژن را مورد بررسی قرار دهد. همان‌طور که در شکل ۲ نیز نشان داده شده است، در این پایگاه از دو طریق می‌توان یک یا چند ژن را جستجو و گراف شبکه بیان هم‌زمان دو یا چند ژن را رسم کرد. در حالت اول، بر اساس نام ژن و یا حروف اختصاری ژن می‌توان ژن را جستجو کرد، سپس جدول نتایج آن ژن آرایه می‌شود و در ادامه می‌توان تصویر شبکه ژنی مرتبط با آن ژن را مشاهده کرد. در حالت دوم، ژن‌های مورد نظر را می‌توان بر اساس کرموزوم‌ها مطالعه کرد.

ابزار HyperTree یک ابزار مهم جهت نمایش شبکه ژنی در این پایگاه داده است (شکل ۳). در حقیقت، HyperTree یک نمایشگر قوی مبتنی بر جاوا جهت نمایش درخت فیلوژنی است و با استفاده از این ابزار می‌توان رابطه خویشاوندی یک ژن را با سایر ژن‌ها در قالب یک درخت مشاهده کرد. این درخت قادر به نمایش درختان فیلوژنتیک بزرگ با صدها گره و یا بیش‌تر، چرخش و کشیدن صفحه نمایش در فضای دکارتی، جستجو و انتخاب گره‌ها، کپی خوشه‌ها برای به‌کارگیری در برنامه‌های دیگر، ایجاد رنگ و برچسب‌زنی شاخه‌ها در هر خوشه (برای مثال، اعضای یک خانواده) و بزرگ‌نمایی یک خوشه ویژه است.

پایگاه‌های داده‌های اصلی جهت مطالعه بیان ژن‌ها در گیاهان (Tissue-Specific Genes) در بر گرفته است. در این صفحه، میزان بیان ژن‌های مختلف در بافت‌های متفاوت برنج آرایه شده است و محقق می‌تواند این اطلاعات را در بافت‌های مختلف با یکدیگر مقایسه کند.

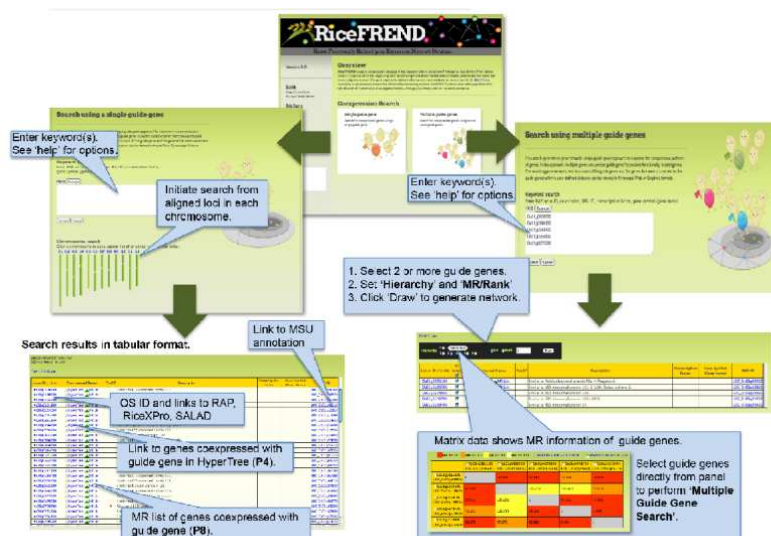
۵- ابزار Rice Expression Database

این ابزار به آدرس <http://expression.ic4r.org> قابل دسترس است و شامل اطلاعات جامع بیانی حاصل از آزمایش‌های RNA-seq در بافت‌ها و تیمارهای مختلف در گیاه برنج می‌باشد. در این ابزار، داده‌های موجود برای نه بافت، ۲۴ پروژه تحقیقاتی و ۲۸۴ آزمایش موجود است. همچنین، ابزار گرافیکی موجود در آن کمک قابل‌توجهی به محقق در فهم هر چه بیش‌تر داده‌های این پایگاه می‌کند.

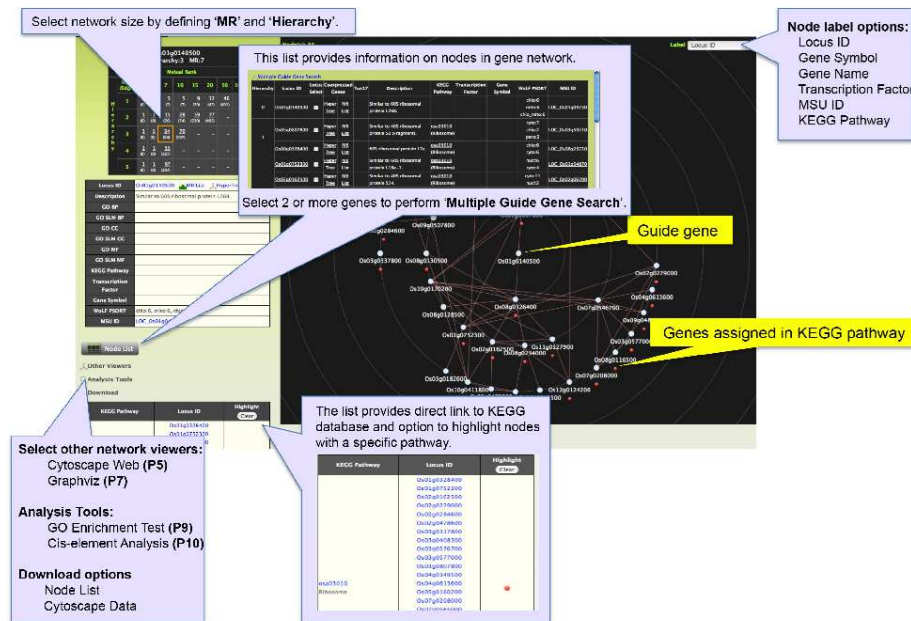
پایگاه‌های داده RiceXPro و RiceFRIEND

برخی از پایگاه‌های داده برای نمایش شبکه بیان هم‌زمان دو یا چند ژن (GCN: Gene Co-expression Network) در برنج به‌کار می‌روند. در حقیقت، شبکه بیان هم‌زمان دو یا چند ژن یک گراف بدون جهت (Undirected) است که در آن هر گره به یک ژن مربوط است و اگر دو ژن با هم بیان مشترک داشته باشند، هر جفت گره به یک لبه مرتبط می‌شوند.

یکی از پایگاه‌های داده برای بررسی و رسم شبکه بیان هم‌زمان دو یا چند ژن، RiceFRIEND می‌باشد. این پایگاه داده به آدرس <http://ricefrend.dna.affrc.go.jp> قابل دسترس است و بر اساس مجموعه وسیعی از داده‌های



شکل ۲- نحوه جستجوی یک ژن در پایگاه داده RiceFRIEND
Figure 2. How to search a gene in the RiceFRIEND database



شکل ۳- ابزار HyperTree به عنوان یک ابزار مهم جهت نمایش شبکه ژنی
Figure 3. The HyperTree tool as an important tool for displaying the genetic network

خطی، پیکتوگراف و ...) را برای ژن مورد نظر دریافت کرد. راه دیگر جستجو، کلیک روی گزینه Rice Global Gene Expression Profile است که از این طریق می توان با انتخاب دسته مورد نظر، انواع گراف را در مورد ژن مورد نظر پیدا کرد (شکل ۴).

پایگاه های داده پروتئین و پروتئوم در برنج

قبل از اینکه به بررسی پایگاه های داده پروتئین و پروتئوم در برنج بپردازیم، ابتدا بهتر است در باره پروتئوم بیش تر بدانیم. پروتئوم به مجموعه کل پروتئین های یک موجود زنده اطلاق می شود. در حقیقت، مجموعه ای از پروتئین ها که در زمانی ویژه، تحت یک شرایط ویژه و در یک مکان زیست شناختی اعم از سلول، بافت یا ارگانیسم بیان می شوند را پروتئوم می نامند. پایگاه های عمده برای بررسی پروتئوم برنج شامل PRIN (Gu *et al.*, 2011; <http://bis.zju.edu.cn/prin> Sapkota *et al.* DIPOS, (<http://www.riceresearch.info> Helmy *et al.*, 2011; <http://www.riceresearch.info> Gour) MCDRP (<http://oryzapg.iab.keio.ac.jp> (*et al.*, 2014; <https://omictools.com/mcdrp-tool> هستند (جدول ۳).

پایگاه داده RiceXPro نیز به آدرس <http://ricexpro.dna.affrc.go.jp> قابل دسترسی است و یک پایگاه داده جامع برگرفته از داده های هزاران آزمایش ریزآرایه می باشد که به راحتی بیان ژن ها را در محیط های مختلف بیولوژیک مانند انواع بافت ها، سلول ها یا ژنوتیپ ها تحت شرایط مزرعه ای در برنج بررسی می کند. در این پایگاه می توان نتایج تاثیر تیمارهای مختلف هورمونی را بر بیان ژن های مختلف در کل دوره رشد گیاه برنج نیز مشاهده کرد. این پایگاه داده بخشی از یک پروژه تجزیه و تحلیل ترانسکریپتوم برنج با استفاده از فناوری ریزآرایه است که هدف آن مشخص کردن مشخصات بیان تمام ژن های پیش بینی شده در برنج و ارائه اطلاعات مرجع جهت استفاده در ژنومیکس کارکردی این غله مهم می باشد. در این پایگاه دو نوع جستجو وجود دارد. در حالت اول، می توان در مجموعه داده های موجود در پایگاه جستجو و نحوه و میزان بیان ژن مورد نظر را پیدا کرد (شکل ۴). پس از انتخاب مجموعه داده، می توان بر اساس کلمات کلیدی و کروموزوم ها جستجو را انجام داد یا می توان با کلیک روی قسمت Analysis tools آزمون t-student، همبستگی و غیره را برای ژن مورد نظر انجام داد. پس از انتخاب کلمه کلیدی یا کروموزوم می توان انواع گراف (بارگراف، گراف

The image shows the RiceXPro database interface with several callouts explaining search steps:

- Global Gene Expression Profile:** Select a category (Field/Development or Plant Hormone) for overall view of transcriptome.
- Data Sets:** Select data set. Detailed information on samples in each data set.
- Search options:** Enter keyword(s). See 'help' for options.
- Chromosome / Region view:** Initiate search from aligned loci in each chromosome.
- Search genes located in between two loci.**
- Plant Hormone:** Enter keyword(s). See 'help' for options.
- Chromosome / Region view (Plant Hormone):** Initiate search from aligned loci in each chromosome.
- Search genes located in between two loci.**
- Analysis Tools (See P10):**

شکل ۴- نحوه جستجوی یک ژن در پایگاه داده RiceXPro
Figure 4. How to search a gene in the RiceXPro database

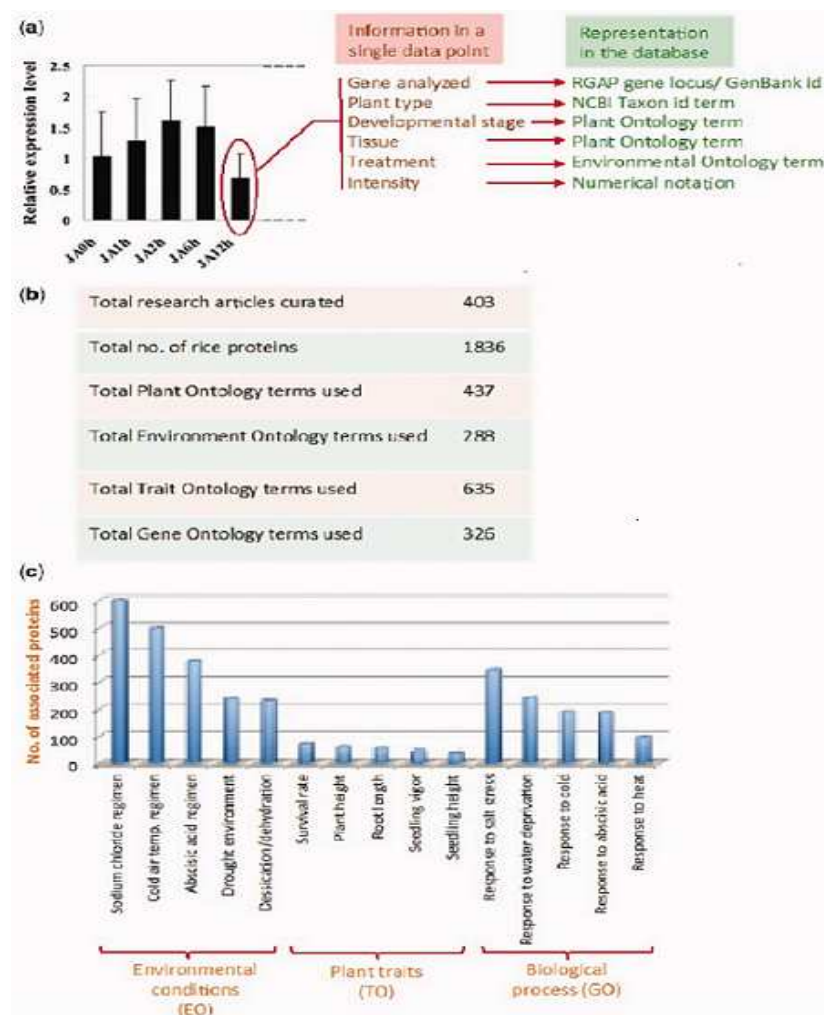
پایگاه داده (Rice Genome Annotation Project) استفاده شده و در مورد ژن‌ها/ پروتئین‌هایی که امکان استفاده از شناسه RGAP نبوده است، از شناسه پایگاه GenBank استفاده شده است. علاوه بر این، در برخی مطالعات، ژن برنج در سیستم گیاهی هترولوگوس مورد مطالعه قرار گرفته است. در چنین مواردی، اطلاعات تجربی ارایه شده بیانگر ژن هترولوگوس می‌باشد. بنابراین، اگر ژن از آرآبیدوپسیس گرفته شده باشد، شناسه ژنی پایگاه TAIR مورد استفاده قرار می‌گیرد، در غیر این صورت شناسه ژنی پایگاه GenBank برای ژن هترولوگ مربوطه استفاده شده است. بافت گیاهی و مرحله رشد و نمو مورد نظر بیش‌تر با اصطلاحات «هستی‌شناختی گیاهی» (Plant Ontology) PO:0025034- مرسوم مشخص می‌شوند (مانند PO:0005352-PO:0000074 به ترتیب برای برگ- آوند چوبی-پارنشیم). همچنین در این پایگاه شرایط رشد مانند دما، وضعیت آب، نور یا حضور هورمون‌ها، فلزات و غیره با اصطلاحات رایج موجود در بیش‌تر پایگاه‌های داده مشخص شده است. به‌منظور بیان هر چه دقیق‌تر شرایط محیطی، همراه اصطلاحات رایج در این پایگاه سعی شده که از سایر اطلاعات از جمله زمان تیمار، غلظت و غیره استفاده شود. شاید اینجا سوال پیش آید که آیا این علائم باعث سردرگمی

پایگاه داده MCDRP (Manually curated) (database of rice proteins)

این پایگاه داده به آدرس اینترنتی <https://omictools.com/mcdrp-tool> در دسترس است و داده‌های موجود در این پایگاه بر اساس داده‌های تجربی و مروری منتشر شده می‌باشند. در حالت عادی، داده‌های تجربی اغلب به‌صورت تصاویر یا نمودارها ارایه شده‌اند و بدیهی است که برای هر جستجوی رایانه‌ای، مناسب نیستند. اما در این پایگاه با دیجیتاله کردن تمام داده‌های موجود، دسترسی به آن‌ها بسیار راحت‌تر شده است. توضیحات جامعی در مورد این پایگاه و نحوه جمع‌آوری داده‌ها در آن در شکل ۵ ارایه شده است. همان‌طور که در شکل ۵-a نشان داده شده است، هر نقطه داده در یک گراف یا تصویر، نشان‌دهنده داده‌های مختلف مانند نام ژن، نوع گیاه، بافت و شرایط رشد می‌باشد. در طول سال‌ها، چندین الگوریتم و علائم استاندارد برای نشان دادن این نوع داده‌ها توسعه داده شده است. بنابراین، مراجعه‌کننده با حداقل پیچیدگی ممکن در طول جستجوی یک ژن/ پروتئین خاص مواجه خواهد شد. در این پایگاه، برای نشان دادن ژن/ پروتئین برنج، از شناسه‌های موجود در پایگاه RGAP

رونوشت‌های ژن LOC_Os05g46480 موجود در ریشه گیاه برنج است که با تکنیک RT-PCR مورد سنجش قرار گرفته است. نسخه فعلی پایگاه داده، اطلاعات بیش از ۱۸۰۰ ژن برنج از تقریباً ۴۰۰ مقاله پژوهشی را شامل می‌شود (شکل ۵-ب). این داده‌ها نتیجه بیش از ۴۰۰۰ آزمایش (با بیش از ۴۰۰ تکنیک آزمایشگاهی از جمله تجزیه و تحلیل بیان ژن، تجزیه فعالیت‌های بیوشیمیایی، تعامل پروتئین-پروتئین و تعامل DNA-پروتئین) هستند. در نهایت در شکل ۵-ج نیز می‌توان انواع بافت‌ها، تیمارها، ژن‌ها و صفات مورد مطالعه در این پایگاه داده را مشاهده کرد.

خواننده نمی‌شود؟ جواب بله است. برای حل این مشکل، در مواقع خاص، پایگاه شناسه‌های جدیدی را تعریف می‌کند که مختص به خود این پایگاه است. در این موارد، تمامی شناسه‌ها با یک عدد خاص شروع می‌شوند (مانند: TO:1000413-Ascorbic acid). همچنین در این پایگاه سعی شده برای فهم بهتر مراجعه‌کننده از نوع تکنیک آزمایشگاهی مورد استفاده، از شناسه‌های ویژه برای این تکنیک‌ها استفاده شود. به‌عنوان مثال، اگر کنار داده‌ای شناسه [LOC_Os05g46480/ PO:0009005 (roots)] قرار داشت، نشان‌دهنده این است که این داده‌ها میزان



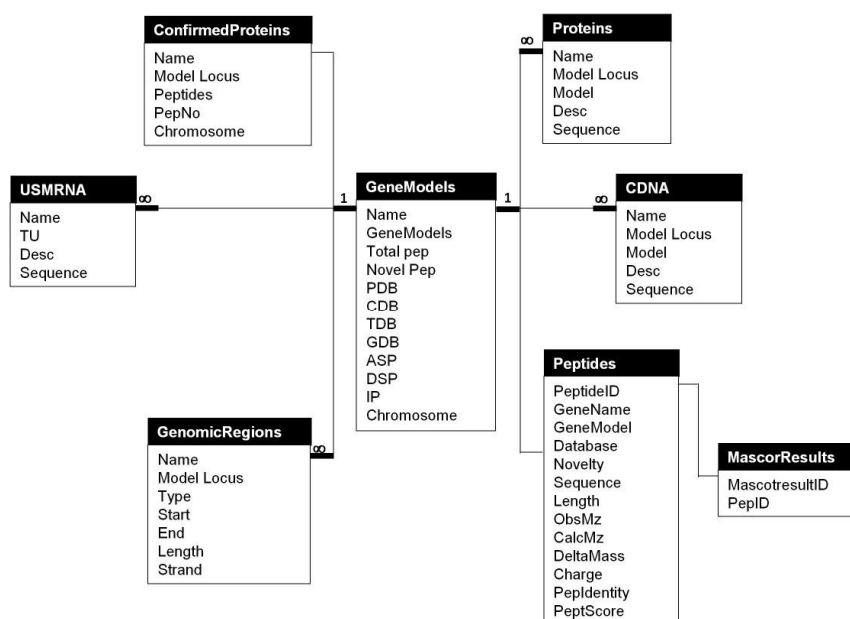
شکل ۵- جزئیات پایگاه داده MCDRP. (a) نمایش مفهوم دیجیتالی کردن داده‌های تجربی، تمام ابعاد اطلاعات در هر نقطه داده (میله‌ها) با یک اصطلاح هستی‌شناسی/ علامت اختصاری نمایش داده می‌شوند، (b) وضعیت کنونی داده‌های ذخیره شده در پایگاه داده، (c) توزیع پنج صفت برتر، شرایط محیطی و عملکرد پروتئین در پایگاه داده.

Figure 5. Details of the MCDRP database. a) Depiction of the concept for digitization of the experimental data, all aspects of the information in each data point (bars) is represented by an appropriate ontology/notation term, b) Current status of the data curated in the database, c) Distribution of the top five traits, environmental conditions and protein function in the curated database.

داده رونویسی شامل اینترون‌ها است و بنابراین می‌توان پپتیدهای اینترونیک را از این طریق شناسایی کرد. در نهایت، پایگاه داده ژنوم شامل مناطق بین ژنی است که امکان پیدا کردن ژن‌های جدید را ارایه می‌دهد. در نتیجه، این پایگاه ۱۴,۹۵۵ پپتید منحصر به فرد از پایگاه داده را از طریق پایگاه داده پروتئین (V6.1) MSU شناسایی کرده است. در این پایگاه از اصطلاحات موجود در پایگاه MSU (V6.1) برای نام‌گذاری پروتئین‌ها استفاده شده است. شناسه‌های پپتیدها با P, C, T یا G شروع می‌شوند که نشان می‌دهد منبع شناسایی پپتید به ترتیب پروتئین، cDNA، رونوشت یا پایگاه داده ژنوم است. معمولاً بعد از حروف P, C, T یا G یک شماره سریال چند رقمی می‌آید. به‌عنوان مثال P345 نشان‌دهنده پپتید شماره ۳۴۵ در میان پپتیدهای شناسایی شده از پایگاه داده پروتئین است. همان‌طور که در شکل ۶ ارایه شده است، این پایگاه داده از قسمت‌های مختلف تشکیل شده که با ورود به هر قسمت می‌توان نام، لوکوس، توالی و تعداد پپتیدها را جستجو کرد.

پایگاه داده (OryzaPG-DB ProteoGenomic Database)

این پایگاه داده برای مطالعه پروتئوم برنج بر اساس تکنیک شاتگان پروتئوژنومیکس است. از طریق آدرس <http://oryzapg.iab.keio.ac.jp> می‌توان به پایگاه داده OryzaPG-DB دسترسی پیدا کرد. برخلاف پایگاه‌های داده پروتئینی موجود در برنج که در حال حاضر اطلاعات پروتئوم مبتنی بر ژل دوبعدی آکرلامید را دارند، OryzaPG-DB حاوی پپتیدهای حاصل از تکنیک شاتگان پروتئوژنومیکس است که دارای تفسیر جامع پپتید و اطلاعات به‌روز پروتئین، cDNA، رونوشت و توالی ژنوم مربوطه می‌باشد. در پایگاه داده پروتئین، اطلاعات تمامی پروتئین‌های بالقوه و پپتیدها فراهم شده است. پایگاه داده cDNA شبیه به پایگاه داده پروتئین است، اما اجازه می‌دهد تا محقق انواع فرم‌های ترجمه توالی نوکلئوتیدی را جستجو و بنابراین، بتواند پپتیدهای اتصال‌دهنده آگزون-آگزون و وقایع پرش آگزون را از دو پایگاه داده جستجو کند. پایگاه



شکل ۶- الگوی پایگاه داده OryzaPG
Figure 6. OryzaPG database scheme

کیفی تمامی متابولیت‌ها و محصولات متابولیک فعال در مسیرهای بیوشیمیایی سلول، بافت یا موجود زنده است. پایگاه‌های عمده مطالعه متابولوم برنج شامل EXPath Chien *et al.*, 2015; <http://expa>)

پایگاه‌های داده متابولوم برنج

علم متابولومیکس علم پیشرو و قدرتمندی است که قادر به بررسی هزاران متابولیت در بافت و اندام موجودات زنده است. این علم در ارتباط با تعیین و اندازه‌گیری‌های کمی و

جزئیات مسیر متابولیک (Pathway Detail Page)

با کلیک روی این لینک وارد یک صفحه جدید برای هر مسیر متابولیک خواهیم شد که دارای یک فرمت گرافیکی قابل بزرگ‌نمایی است و در کنار آن یک لینک مرتبط با اطلاعات واکنش، نوع ماده واکنش‌دهنده، آنزیم‌ها و ژن‌های مربوطه وجود دارد.

نمودار مسیر متابولیک (Pathway Diagram)

هر مسیر متابولیک شامل برخی از اطلاعات رنگی زیر است (شکل ۷):

- واکنش‌ها با خطوط آبی، آنزیم‌ها با رنگ زرد، ژن‌ها با رنگ بنفش و ترکیبات حاصله با رنگ قرمز نشان داده شده است.
- فلش‌های سبز مسیرهای مربوطه را نشان می‌دهند (این مسیرها می‌توانند مستقل یا به مسیر دیگری وابسته باشند)؛ با کلیک روی هر فلش می‌توان به صفحه جزئیات مسیر وارد شد.

- هر داده در نمودار مسیر متابولیک با صفحات جزئیات خود مرتبط است و با کلیک روی آن می‌توان به صفحه جزئیات مربوطه وارد شد.

RiceCyc (th.itps.ncku.edu.tw)،

<http://pathway.gtramene.org/RICE/organism->

OryzaCyc (summary?)،

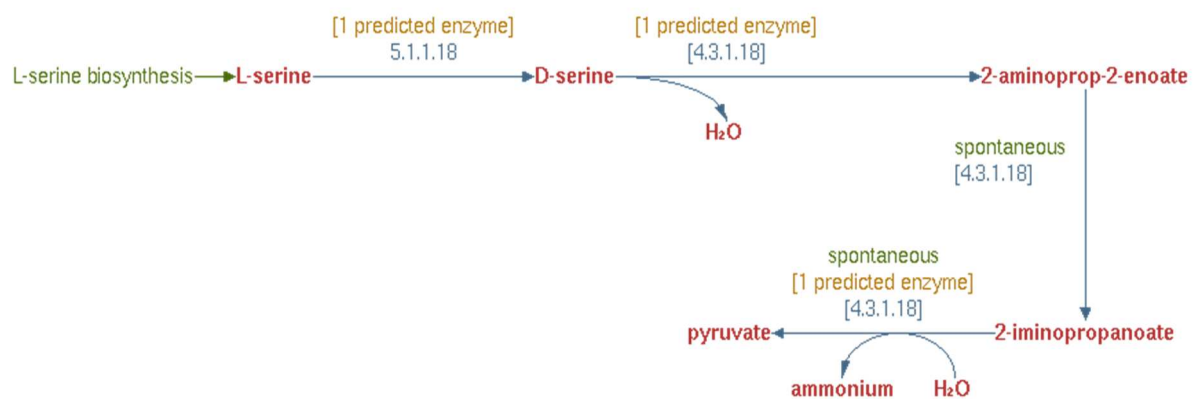
(<http://www.plantcyc.org/databases/oryzacyc>).

KEGG rice (http://www.genome.jp/kegg-bin/show_organism?menu_type=pathway_map&s&org=dosa) اشاره کرد (جدول ۳). از میان آن‌ها، پایگاه داده OryzaCyc اهمیت بیش‌تری دارد که در ادامه به شرح آن می‌پردازیم.

این پایگاه داده از طریق آدرس

<http://www.plantcyc.org/databases/oryzacyc> در

دسترس است و شامل گیاهان مختلف از جمله برنج می‌باشد. هر نوع داده (مسیر متابولیک، نوع واکنش، نوع آنزیم و ترکیب حاصله) در پایگاه داده OryzaCyc دارای صفحه جزئیات است که توسط توابع جستجوی مختلف (و یا از طریق لینک‌های دیگر از صفحات جزئیات اطلاعات) قابل دسترسی می‌باشد. این بخش جزئیات مسیر متابولیک، جزئیات واکنش، جزئیات دقیق آنزیم و جزئیات ترکیب حاصله را شرح می‌دهد.



شکل ۷- مسیر متابولیک دی-سیرین در گیاه *Oryza brachyantha*
Figure 7. The metabolic pathway of D-serine in *Oryza brachyantha*

جنس یکسان است. اگر این مسیر در گونه‌های مختلف یک جنس متفاوت باشد، گزینه Variants در کنار آن قرار دارد که با کلیک روی آن می‌توان مسیرها را نیز مشاهده کرد.

خلاصه (Summary)

این صفحه، شامل یک چکیده مختصر است که توسط متخصصان پایگاه داده تهیه شده است.

شواهد و مدارک (Evidence)

هر مسیر ویژه با یک کد مرتبط است که نشان می‌دهد داده‌ها، شواهد و مدارک لازم برای حمایت از این مسیر وجود دارد.

نوع گونه موردنظر (Species Data Available for)

در این پایگاه داده فرض بر این است که کل مسیر متابولیک ارایه‌شده در نمودار در تمام گونه‌های موجود یک

باعث می‌شود نتوان آن‌ها را در یک مکان ادغام کرد. در یک وضعیت ایده‌آل، باید یک پایگاه داده واحد در یک دامنه منحصر به فرد وجود داشته باشد تا در آن کاربر بتواند تمام پایگاه‌های داده مربوطه را با یک بررسی فقط با استفاده از APIها و اطلاعات موجود (Ontologies) جستجو و نتایج را با هم مقایسه کند (مثل پایگاه داده EMBL-EBI Gene Expression Atlas). در حال حاضر، بعضی از پایگاه‌های داده به منظور افزایش اعتبار داده‌های خود با دیگر پایگاه‌های داده پیوندهایی را انجام داده‌اند (مثل پایگاه داده GENEVESTIGATOR و IC4R). این کار اولین گام در ارائه یک ایدئولوژی و اصول کاری (Platform) متحد است. چنین کاری امکان استفاده از اطلاعات موجود در منابع فعلی را به حداکثر می‌رساند و ممکن است در اجتناب از دوباره کاری کمک کند. این امر قابلیت نمایش بهتری را برای پایگاه‌های داده کوچک فراهم می‌کند، زیرا پایگاه‌های داده کوچک به طور کلی روی یک جنبه خاص تمرکز می‌کنند و اطلاعات دقیق تری ارائه می‌دهند. علاوه بر این، موجب افزایش اعتبار پایگاه‌های داده بزرگ نیز خواهد شد.

منابع (References)

برای مسیریایی که در این پایگاه ارائه شده‌اند، مقالات استفاده شده به عنوان منابع در این قسمت هایپرلینک شده‌اند.

نتیجه‌گیری کلی

در این تحقیق تلاش شد که پایگاه‌های داده‌های عمده و اصلی برای بررسی بیان ژن در گیاهان و به طور ویژه برنج ارائه شوند. بعضی از آن‌ها به خوبی شناخته شده هستند و به طور گسترده استفاده می‌شوند، در حالی که تعدادی دیگر منابع جدید و کوچکی هستند. با افزایش تعداد پایگاه‌های داده، واضح است که مقدار زیادی اطلاعات موجود در بستر وب وجود خواهد داشت که به هر جنبه‌ای از تحقیقات کمک شایان توجهی خواهد کرد. با اینکه مقدار زیادی اطلاعات وجود دارد، اما این اطلاعات به طور موثر مورد بررسی قرار نگرفته‌اند. به عنوان مثال، بسیاری از محققان زیست‌شناسی هنوز به تمام منابع ممکن برای جستجو و تجزیه و تحلیل داده‌ها ناآشنا هستند. اغلب پایگاه‌های داده مختلف، فرمت‌ها و دستورالعمل‌های مختلفی برای تبادل داده‌ها دارند که

References

- Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, H. M. and Xiao, H. 1991. Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* 252: 1651-1656.
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M. and Yefanov, A. 2012. NCBI GEO: Archive for functional genomics data sets-update. *Nucleic Acids Research* 41 (D1): 991-995.
- Barrett, L. W., Fletcher, S. and Wilton, S. D. 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences* 69 (21): 3613-3634.
- Belacel, N., Wang, Q. and Cuperlovic-Culf, M. 2006. Clustering methods for microarray gene expression data. *Omic: A Journal of Integrative Biology* 10 (4): 507-531.
- Bennet, M. D. and Smith, J. B. 1991. Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society* B334: 309-345.
- Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N. and Oezcimen, A. 2003. ArrayExpress-A public repository for microarray gene expression data at the EBI. *Nucleic Acids Research* 31 (1): 68-71.
- Callard, D., Lescure, B. and Mazzolini, L. 1994. A method for the elimination of false positives generated by the mRNA differential display technique. *BioTechniques* 16: 1096-1103.
- Cao, P., Jung, K. H., Choi, D., Hwang, D., Zhu, J. and Ronald, P. C. 2012. The rice oligonucleotide array database: An atlas of rice gene expression. *Rice* 5 (1): 17.
- Chen, C., Huang, H. and Wu, C. H. 2017. Protein bioinformatics databases and resources. In: Protein bioinformatics. Humana Press, New York, NY. pp: 3-39.
- Chien, C. H., Chow, C. N., Wu, N. Y., Chiang-Hsieh, Y. F., Hou, P. F. and Chang, W. C. 2015. EXPath: A database of comparative expression analysis inferring metabolic pathways for plants. *BMC Genomics* 16: S6.

- Craigon, D. J., James, N., Okyere, J., Higgins, J., Jotham, J. and May, S. 2004.** NASCArrays: A repository for microarray data generated by NASC's transcriptomics service. **Nucleic Acids Research** 32 (suppl_1): D575-D577.
- Clark R. M., Wagler T. N., Quijada, P. and Doebley, J. 2006.** A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. **Nature Genetics** 38: 594-597.
- Cochrane, G., Karsch-Mizrachi, I., Takagi, T. and Sequence Database Collaboration, I.N. 2015.** The international nucleotide sequence database collaboration. **Nucleic Acids Research** 44 (D1): D48-D50.
- Deans, C. and Maggert, K. A. 2015.** What do you mean, "epigenetic"? **Genetics** 199 (4): 887-896.
- Dharmawardhana, P., Ren, L., Amarasinghe, V., Monaco, M., Thomason, J., Ravenscroft, D. and Jaiswal, P. 2013.** A genome scale metabolic network for rice and accompanying analysis of tryptophan, auxin and serotonin biosynthesis regulation under biotic stress. **Rice** 6 (1): 15.
- Dodd, A. N., Salathia, N., Hall, A., Kévei, E., Tóth, R., Nagy, F. and Webb, A. A. 2005.** Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. **Science** 309 (5734): 630-633.
- Garcia-Hernandez, M., Berardini, T., Chen, G., Crist, D., Doyle, A., Huala, E. and Mundodi, S. 2002.** TAIR: A resource for integrated *Arabidopsis* data. **Functional and Integrative Genomics** 2 (6): 239-253.
- Ghildiyal, M. and Zamore, P. D. 2009.** Small silencing RNAs: An expanding universe. **Nature Reviews Genetics** 10 (2): 94.
- Geertz, M. and Maerkl, S. J. 2010.** Experimental strategies for studying transcription factor–DNA binding specificities. **Briefings in Functional Genomics** 9 (5-6): 362-373.
- Gour, P., Garg, P., Jain, R., Joseph, S. V., Tyagi, A. K. and Raghuvanshi, S. 2013.** Manually curated database of rice proteins. **Nucleic Acids Research** 42 (D1): D1214-D1221.
- Grigoriev, A. 2001.** A relationship between gene expression and protein interactions on the proteome scale: Analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. **Nucleic Acids Research** 29 (17): 3513-3519.
- Gruber, A. R., Martin, G., Keller, W. and Zavolan, M. 2014.** Means to an end: Mechanisms of alternative polyadenylation of messenger RNA precursors. **Wiley Interdisciplinary Reviews: RNA** 5 (2): 183-196.
- Gu, H., Zhu, P., Jiao, Y., Meng, Y. and Chen, M. 2011.** PRIN: A predicted rice interactome network. **BMC Bioinformatics** 12 (1): 161.
- Hamada, K., Hongo, K., Suwabe, K., Shimizu, A., Nagayama, T., Abe, R. and Tsuchida, H. 2011.** OryzaExpress: An integrated database of gene expression networks and omics annotations in rice. **Plant and Cell Physiology** 52 (2): 220-229.
- Helmy, M., Tomita, M. and Ishihama, Y. 2011.** OryzaPG-DB: Rice proteome database based on shotgun proteogenomics. **BMC Plant Biology** 11 (1): 63.
- Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L. and Zimmermann, P. 2008.** GENEVESTIGATOR v3: A reference expression database for the meta-analysis of transcriptomes. **Advances in Bioinformatics** 2008: 1-5.
- Hudson, M. E. and Quail, P. H. 2003.** Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. **Plant Physiology** 133 (4): 1605-1616.
- Jombart, T. 2008.** ADEGENET: A R-package for the multivariate analysis of genetic markers. **Bioinformatics** 24 (11): 1403-1405.
- Jung, K. H., Cao, P., Sharma, R., Jain, R. and Ronald, P. C. 2015.** Phylogenomics databases for facilitating functional genomics in rice. **Rice** 8 (1): 26.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. 2013.** Data, information, knowledge and principle: Back to metabolism in KEGG. **Nucleic Acids Research** 42 (D1): D199-D205.
- Karsch-Mizrachi, I., Nakamura, Y. and Cochrane, G. 2011.** The international nucleotide sequence database collaboration. **Nucleic Acids Research** 40 (D1): D33-D37.
- Katari, M. S., Nowicki, S. D., Aceituno, F. F., Nero, D., Kelfer, J., Thompson, L. P., Cabello, J. M., Davidson, R. S., Goldberg, A. P., Shasha, D. E., Coruzzi, G. M. and Gutierrez, R. A. 2010.**

- VirtualPlant: A software platform to support systems biology research. **Plant Physiology** 152 (2): 500-515.
- Kawahara, Y., Oono, Y., Wakimoto, H., Ogata, J., Kanamori, H., Sasaki, H. and Itoh, T. 2015.** TENOR: Database for comprehensive mRNA-seq experiments in rice. **Plant and Cell Physiology** 57 (1): e7-e7.
- Kudo, T., Akiyama, K., Kojima, M., Makita, N., Sakurai, T. and Sakakibara, H. 2013.** UniVIO: A multiple omics database with hormone and transcriptome data from rice. **Plant and Cell Physiology** 54 (2): e9-e9.
- Lee, T., Oh, T., Yang, S., Shin, J., Hwang, S., Kim, C. Y. and Lee, I. 2015.** RiceNet v2: An improved network prioritization server for rice genes. **Nucleic Acids Research** 43 (W1): W122-W127.
- Liang, P. and Pardee, A. B. 1992.** Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. **Science** 257: 967-971.
- Lu, T., Huang, X., Zhu, C., Huang, T., Zhao, Q., Xie, K. and Han, B. 2008.** RICD: A rice indica cDNA database resource for rice functional genomics. **BMC Plant Biology** 8 (1): 118.
- Mockler, T. C., Michael, T. P., Priest, H. D., Shen, R., Sullivan, C. M., Givan, S. A. and Chory, J. 2007.** The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. Cold Spring Harbor Symposia on Quantitative Biology. Vol. 72. Clocks and Rhythms. Cold Spring Harbor Laboratory Press. pp: 353-363.
- Moody, D. E. 2001.** Genomics techniques: An overview of methods for the study of gene expression. **Journal of Animal Science** 79 (E. Suppl.): E128-E135.
- Narsai, R., Devenish, J., Castleden, I., Narsai, K., Xu, L., Shou, H. and Whelan, J. 2013.** Rice DB: An *Oryza* information portal linking annotation, subcellular location, function, expression, regulation, and evolutionary information for rice and *Arabidopsis*. **The Plant Journal** 76 (6): 1057-1073.
- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F. and Duesbury, M. 2013.** The MIntAct project-IntAct as a common curation platform for 11 molecular interaction databases. **Nucleic Acids Research** 42 (D1): D358-D363.
- Petryszak, R., Keays, M., Tang, Y. A., Fonseca, N. A., Barrera, E., Burdett, T. and Mannion, O. 2015.** Expression atlas update-An integrated database of gene and protein expression in humans, animals and plants. **Nucleic Acids Research** 44 (D1): D746-D752.
- Priya, P. and Jain, M. 2013.** RiceSRTFDB: A database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. **Database** 2013: 1-7.
- Rajaram, S. and Oono, Y. 2010.** NeatMap-non-clustering heat map alternatives in R. **BMC Bioinformatics** 11 (1): 45.
- Rhee, S. Y., Beavis, W., Berardini, T. Z., Chen, G., Dixon, D., Doyle, A. and Miller, N. 2003.** The *Arabidopsis* information resource (TAIR): A model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. **Nucleic Acids Research** 31 (1): 224-228.
- Riechmann, J. L. and Ratcliffe, O. J. 2000.** A genomic perspective on plant transcription factors. **Current Opinion in Plant Biology** 3 (5): 423-434.
- Sakurai, T., Kondou, Y., Akiyama, K., Kurotani, A., Higuchi, M., Ichikawa, T. and Sakakibara, H. 2010.** RiceFOX: A database of *Arabidopsis* mutant lines overexpressing rice full-length cDNA that contains a wide range of trait information to facilitate analysis of gene function. **Plant and Cell Physiology** 52 (2): 265-273.
- Sapkota, A., Liu, X., Zhao, X. M., Cao, Y., Liu, J., Liu, Z. P. and Chen, L. 2011.** DIPOS: Database of interacting proteins in *Oryza sativa*. **Molecular BioSystems** 7 (9): 2615-2621.
- Sargent, T. D. and Dawid, I. B. 1983.** Differential gene expression in the gastrula of *Xenopus laevis*. **Science** 222: 135-139.
- Sato, Y., Namiki, N., Takehisa, H., Kamatsuki, K., Minami, H., Ikawa, H. and Nagamura, Y. 2013.** RiceFRIEND: A platform for retrieving coexpressed gene networks in rice. **Nucleic Acids Research** 41 (D1): D1214-D1221.
- Sato, Y., Takehisa, H., Kamatsuki, K., Minami, H., Namiki, N., Ikawa, H. and Nagamura, Y. 2013.** RiceXPro version 3.0: Expanding the informatics resource for rice transcriptome. **Nucleic Acids Research** 41 (D1): D1206-D1213.

- Schläpfer, P., Zhang, P., Wang, C., Kim, T., Banf, M., Chae, L., Dreher, K., Chavali, A. K., Nilo-Poyanco, R., Bernard, T., Kahn, D. and Rhee, S. Y. 2017. Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. **Plant Physiology** 173 (4): 2041-2059.
- Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M. and Lohmann, J. U. 2005. A gene expression map of *Arabidopsis thaliana* development. **Nature Genetics** 37 (5): 501.
- Second, G. 1991. Molecular markers in rice systematics and the evaluation of genetic resources. **Biotechnology in Agriculture and Forestry** 4: 468-490.
- Shen, L., Gong, J., Caldo, R. A., Nettleton, D., Cook, D., Wise, R. P. and Dickerson, J. A. 2005. BarleyBase-An expression profiling database for plant genomics. **Nucleic Acids Research** 33 (suppl_1): D614-D618.
- Smith, T. F. 1990. The history of the genetic sequence databases. **Genomics** 6 (4): 701-707.
- Strasser, B. J. 2008. GenBank-Natural history in the 21st century? **Science** 322 (5901): 537-538.
- Tabkhkar, N. and Rabiei, B. 2014. Bioinformatics and cereal genome databases: A case study in rice. **Cereal Research** 4 (1): 77-87. (In Persian with English Abstract).
- Tardieu, F. 2013. Plant response to environmental conditions: Assessing potential production, water demand, and negative effects of water deficit. **Frontiers in Physiology** 4: 17.
- The IC4R Project Consortium, 2016. Information commons for rice (IC4R). **Nucleic Acids Research** 44: D1172-D1180.
- Thijs, G., Lescot, M., Marchal, K., Rombauts, S., De Moor, B., Rouze, P. and Moreau, Y. 2001. A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. **Bioinformatics** 17 (12): 1113-1122.
- Thijs, G., Marchal, K., Lescot, M., Rombauts, S., De Moor, B., Rouzé, P. and Moreau, Y. 2002. A Gibbs sampling method to detect over-represented motifs in the upstream regions of co-expressed genes. Proceedings of the 5th Annual International Conference on Computational Biology. pp: 305-312.
- Toufighi, K., Brady, S. M., Austin, R., Ly, E. and Provart, N. J. 2005. The botany array resource: E-northern, expression angling, and promoter analyses. **The Plant Journal** 43 (1): 153-163.
- Tollefsbol, T. O. 2011. Advances in epigenetic technology. In: Tollefsbol, T. O. (Ed.). Epigenetics protocols. Second Edition. Humana Press. pp: 1-10.
- Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. 1995. Serial analysis of gene expression. **Science** 270: 484-487.
- Wang, D., Xia, Y., Li, X., Hou, L. and Yu, J. 2013. The rice genome knowledgebase (RGKbase): An annotation database for rice comparative genomics and evolutionary biology. **Nucleic Acids Research** 41 (D1): D1199-D1205.
- Wittkopp, P. J. and Kalay, G. 2012. Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. **Nature Reviews Genetics** 13 (1): 59.
- Woelfle, M. A., Ouyang, Y., Phanvijhitsiri, K. and Johnson, C. H. 2004. The adaptive value of circadian clocks: An experimental assessment in cyanobacteria. **Current Biology** 14 (16): 1481-1486.
- Zhang, B., Whiteaker, J. R., Hoofnagle, A. N., Baird, G. S., Rodland, K. D. and Paulovich, A. G. 2018. Clinical potential of mass spectrometry-based proteogenomics. **Nature Reviews Clinical Oncology** 16: 256-268.
- Zhang, Z., Hu, S. N., He, H., Zhang, H. Y., Chen, F., Zhao, W. M., Xiao, J. F., Chen, L. L., Xue, Y. and Wang, X. F. 2016. Information commons for rice (IC4R). **Nucleic Acids Research** 44: D1172-D1180.
- Zhao, S., Guo, Y., Sheng, Q. and Shyr, Y. 2014. Advanced heat map and clustering analysis using heatmap3. **BioMed Research International** 2014: 1-6.
- Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W. 2004. GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. **Plant Physiology** 136 (1): 2621-2632.



University of Guilan
Faculty of Agricultural
Sciences

(Review Paper)
Cereal Research
Vol. 9, No. 1, Spring 2019 (83-101)

Major databases to study genes expression in plants (A case study: Rice)

Mojtaba Kordrostami¹ and Babak Rabiei^{2*}

Received: January 16, 2019

Accepted: May 5, 2019

Abstract

Rice, as one of the most important crops, has the smallest genome among cereals and is considered as a model plant for genetic studies. The small size of this plant's genome has led to comprehensive studies on it, resulting in a large amounts of data to be obtained. Properly collecting and storing these data and managing data from various experiments in a database for accessing researchers to them to avoid re-work and to compare their results with the results of other researchers is very important. To achieve these, bioinformatics can be great help. Therefore, the creation and development of specialized databases and the use of bioinformatics tools for data processing, efficient organization, analysis and visualization are necessary. In this paper, the major databases for studying gene expression at three levels, RNA, protein and metabolome in rice are reviewed and the characteristics of these databases for each level are discussed.

Keywords: Bioinformatics, Metabolic pathway, Web based databases

1. Graduated Ph. D., Dept. of Agricultural Biotechnology, Faculty of Agricultural Science, University of Guilan, Rasht, Iran

2. Prof., Dept. of Plant Production and Genetic Engineering, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran

* Corresponding author: rabiei@guilan.ac.ir