



دانشگاه گیلان  
دانشکده علوم کشاورزی

تحقیقات غلات

سال چهارم / شماره اول / ۱۳۹۳ (۸۷-۷۷)

مقاله علمی کوتاه

## بیوانفورماتیک و پایگاه‌های اطلاعاتی ژنوم غلات (مطالعه موردی برنج)

نرجس طبخ‌کار<sup>۱</sup> و بابک ربیعی<sup>۲\*</sup>

۱ و ۲- به ترتیب دانشجوی دکتری و استاد گروه زراعت و اصلاح‌نباتات دانشکده علوم کشاورزی دانشگاه گیلان

(تاریخ دریافت: ۱۳۹۲/۷/۲۲ - تاریخ پذیرش: ۱۳۹۲/۱۰/۲۳)

### چکیده

پیشرفت فناوری‌های نوین و سرعت تحقیقات علم ژنتیک منجر به تولید حجم بالای اطلاعات ژنتیکی در زمینه‌های مختلف نظیر توالی نوکلئوتیدها و اسیدهای آمینه، عملکرد ژن‌ها و پروتئین‌ها، جهش‌یافته‌ها و فنوتیپ‌های آن‌ها، نشانگرها و استفاده از آن‌ها و موضوعات مرتبط شده است. علاوه بر حجم زیاد اطلاعات ژنتیکی، این اطلاعات پیچیده و دارای تفاوت‌های ظریف در روش‌های آزمایشی هستند که می‌توانند در کیفیت و تفسیر اطلاعات تأثیر بگذارند. به این ترتیب، توصیف و طبقه‌بندی اطلاعات مرتبط با ژن‌ها و اطلاعات فنوتیپی آن‌ها به طریقی که امکان مقایسه سریع و آسان آن‌ها فراهم باشد، اهمیت بسیاری دارد. استفاده از پایگاه‌های اطلاعات ژنتیکی می‌تواند در تکمیل و به روز شدن تحقیقات و آگاهی محققین از آخرین یافته‌های علمی کمک بسیاری نماید. استفاده از پایگاه‌های اطلاعات ژنتیکی در برنامه‌های اصلاحی غلات در سال‌های اخیر، امری متداول و اجتناب ناپذیر بوده است. در این مطالعه، ضمن بررسی پایگاه‌های اطلاعاتی ژنوم برنج، یکی از کاربردی‌ترین پایگاه‌های اطلاعاتی غلات معرفی می‌شود.

**واژه‌های کلیدی:** اطلاعات ژنتیکی، بیوانفورماتیک، پایگاه‌های اطلاعاتی غلات، گرامنه

کلمات اختصاری:

IRGSP, International Rice Genome Sequencing Project; RAP, Rice Annotation Project; MSU-RGAP, Michigan State University- Rice Genome Annotation Project; RIS, Rice Information System; OMAP, *Oryza* Mapping Alignment Project; SSR, Simple Sequence Repeats; QTL, Quantitative Trait Loci.

\* نویسنده مسئول: rabiei@guilan.ac.ir

## مقدمه

معرفی ارقام پر محصول و رشد تکنولوژی باعث شده که تولید جهانی برنج در سه دهه گذشته به میزان قابل توجهی افزایش یابد. اما این افزایش تولید نمی‌تواند پاسخگوی نیاز رو به رشد مصرف‌کنندگان برنج باشد. سازمان خوار و بار جهانی (FAO) اعلام نموده است که تا سال ۲۰۳۰ میلادی تولید جهانی برنج باید تا دو برابر مقدار کنونی افزایش یابد تا پاسخگوی نیاز مصرف‌کنندگان باشد (FAO, 2010). فناوری زیستی و مهندسی ژنتیک پتانسیل بالایی برای افزایش کارایی روش‌های اصلاح غلات دارند و با تلفیق این روش‌ها با روش‌های اصلاح سنتی، اصلاح کیفیت، کمیت و مقاومت به تنش‌های زنده و غیر زنده با سرعت و کارایی بیشتری انجام خواهد شد. این در حالی است که مقدار بالای اطلاعات حاصل از ژنوم‌های توالی‌یابی شده، مطالعات بیان ژن‌ها و تنوع ژنتیکی باعث شده تا محققین با چالش‌هایی چون ذخیره، تجزیه و بازیابی کارآمدتر اطلاعات ژنتیکی رو به رو باشند (Youens-Clark *et al.*, 2011). به منظور صرفه جویی در وقت و استفاده بهینه از این حجم بالای اطلاعات، استفاده از پایگاه‌های اطلاعاتی امری ضروری و اجتناب‌ناپذیر است. دسترسی سریع به منابع و اطلاعات علمی و معتبر از طریق پایگاه‌های اطلاعاتی، امکان داده‌کاوی نامحدود و حتی کشف یافته‌های جدید را برای محققین فراهم می‌کند. برنج به دلیل داشتن ویژگی‌های منحصر به فردی چون مجموعه ژرم‌پلاسمی با بیش از ۱۲۰۰۰۰ نمونه، اندازه ژنوم نسبتاً کوچک حدود ۴۳۰ Mb، سطح چند شکلی نسبتاً بالا و دیپلوئید بودن ژنوم، گیاهی مناسب برای مطالعه به عنوان گیاه مدل است (Jaiswal *et al.*, 2002). دسترسی به اطلاعات ژنتیکی گیاه برنج نه تنها برای محققین برنج، بلکه برای تحقیقات در مورد سایر گیاهان خانواده غلات نیز مفید است، زیرا به دلیل اندازه ژنوم بزرگ و تعداد زیاد گونه‌های خانواده غلات، توالی‌یابی کامل ژنوم بسیاری از این گونه‌ها مدنظر نمی‌باشد. از طرف دیگر، درجه بالای حفاظت ژن در گونه‌های غلات و عملکرد حفاظتی ژن بین گونه‌های با خویشاوندی نزدیک باعث شده که اطلاعات ژنوم یکی از اعضای این خانواده برای گرفتن تصمیمات آگاهانه در مورد ساختار ژنوم، عملکرد ژن و پتانسیل فنوتیپ‌ها در گونه‌های وابسته مفید باشد (Ware *et al.*, 2002). به این دلیل است که امروزه

علاوه بر اینکه بسیاری از محققین، برنج را موضوع اصلی تحقیق خود قرار داده‌اند، بسیاری دیگر برنج را به عنوان یک گیاه مدل برای سایر گیاهان خانواده غلات می‌دانند و نیازمند دسترسی آسان و سریع به اطلاعات و تحقیقات انجام شده روی این گیاه می‌باشند.

در این پژوهش، پایگاه‌های اطلاعاتی حاوی اطلاعات جامعی از ژنوم برنج از نظر نوع اطلاعات، موارد استفاده و کیفیت، مقایسه و بررسی شده و سپس یکی از معتبرترین و کاربردی‌ترین پایگاه‌های اطلاعاتی غلات در امر به‌نژادی یعنی Gramene، معرفی و شیوه استفاده از آن شرح داده شده است. به امید اینکه مقاله حاضر بتواند در جهت شناسایی و استفاده از پایگاه‌های اطلاعات ژنتیکی برای استادان، محققان و دانشجویان تحصیلات تکمیلی رشته‌های بیوتکنولوژی، ژنتیک، بیولوژی و کشاورزی مفید واقع شود.

## مواد و روش‌ها

در این پژوهش پایگاه‌های اطلاعاتی حاوی اطلاعات حاصل از پژوهش‌های جدید در زمینه‌های مختلف ژنتیکی مرتبط با گونه‌های غلات (با تاکید بر برنج به عنوان یک گیاه مدل برای سایر گیاهان خانواده غلات) معرفی شده‌اند (جدول ۱). در ابتدا پایگاه‌های مرتبط با توالی‌های ژنوم برنج و همچنین پایگاه‌هایی که امکان جستجوی BLAST را برای محققین فراهم می‌نمایند توضیح داده شده است (جدول ۲). همچنین ضمن تعریف روش پیشگویی ژن، وبگاه‌های حاوی برنامه‌های برخط برای پیشگویی ژن در برنج معرفی می‌شود (جدول ۳). در انتها ضمن معرفی پایگاه اطلاعاتی گرامنه (Gramene) به عنوان یکی از معتبرترین پایگاه‌های اطلاعاتی غلات به توضیح روش‌های مختلف جستجو در این پایگاه پرداخته می‌شود. در این بخش محققین می‌توانند با شیوه جستجو، نوع و منابع اطلاعاتی وبگاه گرامنه در زمینه‌های مختلفی چون ژنوم، تنوع ژنتیکی، مسیرهای متابولیکی، پروتئین، ژن، نقشه-های ژنتیکی و فیزیکی، نقشه‌های مبتنی بر توالی، نشانگرها، QTLها و منابع و مستندات مختلف در هر یک از این زمینه‌ها، آشنا شوند.

جدول ۱- پایگاه‌های اطلاعاتی مرتبط با ژنوم برنج و بعضی از گونه‌های غلات  
Table 1. Rice and some cereal species genome databases

پایگاه اطلاعاتی Database	آدرس وبگاه URL	نوع اطلاعات Information category
Gramene	<a href="http://www.gramene.org/">http://www.gramene.org/</a>	دامنه‌ی وسیعی از اطلاعات ژنتیکی شامل توالی‌ها، زن‌ها، نشانگرها، QTL‌ها، نقشه‌های کروموزمی، پروتئین‌ها و ... امکان جستجوی BLAST و تجزیه مقایسه‌ای بین توالی‌های ژنومی غلات Widespread range of genetic information include sequences, genes, markers, QTLs, chromosomal maps, proteins and etc, search BLAST and comparative analysis between cereal genomic sequences امکان جستجوی BLAST BLAST search
RiceGE	<a href="http://signal.salk.edu/cgi-bin/RiceGE">http://signal.salk.edu/cgi-bin/RiceGE</a>	بانک اطلاعات DNA، امکان جستجوی BLAST، تجزیه ژنوم و پروتئین DNA database, BLAST search, genome and protein analysis اطلاعات عمومی DNA، RNA، امکان جستجوی BLAST General information about DNA and RNA, BLAST search اطلاعات توالی DNA و لینک به سایر پایگاه‌ها
DDBJ	<a href="http://www.ddbj.nig.ac.jp/Welcome-e.html">http://www.ddbj.nig.ac.jp/Welcome-e.html</a>	DNA sequence information and link to other databases
EMBL	<a href="http://www.ebi.ac.uk/embl/">http://www.ebi.ac.uk/embl/</a>	اطلاعات کلون‌های cDNA Full length و توالی‌های رمز اسیدهای آمینه برنج
Osa1	<a href="http://www.tigr.org/tdb/e2k1/osa1/">http://www.tigr.org/tdb/e2k1/osa1/</a>	اطلاعات ریزارایه‌های برنج، پروتوکل‌های مرکز ریزارایه، لینک به سایر پایگاه‌ها
KOME	<a href="http://cdna01.dna.affrc.go.jp/cDNA/">http://cdna01.dna.affrc.go.jp/cDNA/</a>	Rice microarrays information, microarrays center protocols, link to other databases اطلاعات متنوع، نشانگرهای مولکولی برنج Diverse information, rice molecular markers
RMOS	<a href="http://cdna01.dna.affrc.go.jp/RMOS/main_en.html">http://cdna01.dna.affrc.go.jp/RMOS/main_en.html</a>	اطلاعات عمومی DNA، نشانگرهای مولکولی و امکان جستجوی BLAST General information about DNA, molecular markers and BLAST search نقشه‌های ژنتیکی، نشانگرها و QTL‌ها Genetic maps, markers and QTLs
OryzaSNP	<a href="http://oryzasnp.org/">http://oryzasnp.org/</a>	نشانگرهای ژنتیکی، SNP، SSR Genetic markers, SNP, SSR
OsGDB	<a href="http://www.plantgdb.org/OsGDB/index.php">http://www.plantgdb.org/OsGDB/index.php</a>	عوامل رونویسی ژنوم برنج (زیرگونه ژاپونیکا و ایندیکا) Rice genome transcription Factors (subspecies <i> japonica</i> and <i> indica</i> ) تظاهر ژن در خانواده غلات و ارتباط آن با صفات زراعی
PlantGM	<a href="http://www.niab.go.kr/nabic/PlantGM">http://www.niab.go.kr/nabic/PlantGM</a>	Gene expression in cereal family and its relation with agronomy traits
PlantMarkers	<a href="http://markers.btk.fi">http://markers.btk.fi</a>	
DRTF	<a href="http://dtrf.cbi.pku.edu.cn/">http://dtrf.cbi.pku.edu.cn/</a>	
GRASSIUS	<a href="http://grassius.org/">http://grassius.org/</a>	

Table 1. Continued

ادامه جدول ۱	RED	اطلاعات آزمایش‌های تظاهر زن در برنج
OryzaBase	<a href="http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp">http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp</a>	Information about gene expression experiments in rice اطلاعاتی در مورد ذخایر ژنتیکی، موتانت‌ها، نقشه‌های فیزیکی و مقایسه‌ای، توالی DNA، واژه‌نامه ژنی برای برنج و امکان جستجوی BLAST
<i>Tos17</i> insertion mutants	<a href="http://tos.nias.affrc.go.jp/">http://tos.nias.affrc.go.jp/</a>	Information about genetic resources, mutants, physical and comparative maps, DNA sequencing, gene ontology for rice and BLAST search اطلاعاتی در مورد نقاط درج و اطلاعات فنوتیپی لاین‌های جهش‌یافته <i>Tos17</i> برنج، امکان جستجوی BLAST
RetrOryza	<a href="http://www.retroryza.org/">http://www.retroryza.org/</a>	Information about insertion regions and phenotypic information of rice <i>Tos17</i> mutant lines, BLAST search پایگاه اطلاعاتی ترانسپوزون‌های معکوس برنج Rice retrotransposons database
Gene Ontology	<a href="http://www.geneontology.org/">http://www.geneontology.org/</a>	کلمات کلیدی مشترک مرتبط با ژن‌ها که در پایگاه‌های مختلف استفاده می‌شود، به همراه معانی آن‌ها Common keywords related to genes which used in different databases with their means
IRIS	<a href="http://www.iris.irri.org/">http://www.iris.irri.org/</a>	شجره ژن‌پلاسمی، ارزیابی‌های مزه‌ای، نقشه‌های ژنتیکی، اطلاعات ساختاری و کارکردی ژنوم برنج به همراه لینک به سایر پایگاه‌ها، اطلاعات محیطی
MetaCrop	<a href="http://meta.crop.ipk-atersleben.de">http://meta.crop.ipk-atersleben.de</a>	Germplasm pedigree, field evaluations, genetic maps, structural and functional information of rice genome with link to other databases, environmental information اطلاعات مسیرهای متابولیکی غلات شامل نمودارها، واکنش‌ها، موقعیت‌ها، فرآیند انتقال و مقالات مرتبط Cereal metabolic pathways information include graphs, reactions, positions, transport process and related articles

## پایگاه‌های تفسیری ژنوم برنج

برنج اولین غله‌ای است که توالی ژنومی آن به طور کامل رمزگشایی شد. پیش‌نویس توالی ژنوم برنج هر دو زیرگونه جاپونیکا (*japonica*) و ایندیکا (*indica*) در سال ۲۰۰۲ منتشر شد (Goff et al., 2002; Yu et al., 2002) و به دنبال آن در سال ۲۰۰۵ میلادی توالی کامل ژنوم برنج ژاپونیکا (رقم Nipponbare) توسط پروژه بین‌المللی توالی‌یابی ژنوم برنج (IRGSP) منتشر شد. پایگاه‌های اطلاعاتی تفسیری (Annotation databases)، اطلاعاتی در مورد توالی‌ها، مکان‌های ژنی، مدل‌های ژنی و توضیحاتی در مورد خصوصیات توالی‌ها و بعضاً توضیحات کاربردی برای مدل‌های ژنی را در قالب یک وبگاه اطلاعاتی ارائه می‌کنند (Childs, 2009). با داشتن اطلاعات توالی کامل ژنوم، محققین می‌توانند اطلاعاتی در مورد خصوصیات ژنوم مورد نظر به دست آورند. این اطلاعات، شامل تعیین ژن‌های رمزگذار و غیر رمزگذار پروتئین‌ها و ساختن چنین خانواده‌های ژنی، عوامل تنظیم کننده، توالی‌های تکراری، توالی‌های تکراری ساده (SSR) و مقدار گوانین-سیتوزین (GC) است. اطلاعات فوق می‌توانند برای طراحی ریزآرایه‌ها و نشانگرهای مولکولی نیز استفاده شوند (Mochida and Shinozaki, 2010). در تمام این پایگاه‌ها، اطلاعات کلی در مورد توالی DNA، ساختار و تفسیر ژن‌ها فراهم است. از آنجا که BLAST به عنوان یکی از بهترین برنامه‌ها برای جستجوی تشابه شناخته شده است، امکان جستجو و تطابق توالی مورد نظر در ژنوم با BLAST نیز در این پایگاه‌ها وجود دارد (Altschul et al., 1997). فهرست منابع اطلاعاتی که

اطلاعات توالی ژنوم برنج و امکان تفسیر آن را دارند، در جدول ۲ ارائه شده است. سه منبع اصلی تفسیر ژنوم برنج (Rice genome annotation)، شامل پروژه تفسیر ژنوم برنج یا RAP (Rice Annotation Project)، پروژه تفسیر ژنوم برنج دانشگاه ایالتی میشیگان یا MSU-RGAP (Michigan State University-Rice Genome Annotation Project) و سیستم اطلاعات برنج یا RIS (Rice Information System) می‌باشند (Childs, 2009). این سه منبع، اطلاعات توالی ژنوم برنج را به ترتیب در پایگاه‌های RAP-DB، RGAP و RISE ارائه می‌نمایند. در این بین، فقط پایگاه RISE است که اطلاعات توالی ژنوم برنج گونه ایندیکا را ارائه می‌کند و دو پایگاه دیگر حاوی اطلاعات ژنوم برنج جاپونیکا هستند. در نسخه فعلی پایگاه گرامنه (Gramene, #38, August 2013) اطلاعات توالی ژنوم ۲۱ گونه غلات (شامل ۱۱ گونه برنج) ارائه شده است (www.gramene.org). برای این کار گرامنه از نسخه ۵۸ Ensembl استفاده می‌کند (Flicek et al., 2009). علاوه بر نقشه‌های کاملاً توالی‌یابی شده، گرامنه با پروژه تطابق‌سنجی نقشه‌برداری برنج (OMAP) (Wing et al., 2005)، برای به تصویر کشیدن نقشه فیزیکی *O. rufipogon* و بازوی کوتاه کروموزوم ۳ *O. brachyantha*، *O. barthii*، *O. rufipogon*، *O. nivara*، *O. officinalis*، *O. minuta* CC *glaberrima* و *O. punctata* نیز همکاری می‌کند (Youens-Clark et al., 2011).

جدول ۲- پایگاه‌های اطلاعاتی تفسیری که اطلاعات توالی ژنوم برنج را ارائه می‌کنند

Table 2. Rice annotation databases that represent sequenced rice genome information

نام پایگاه اطلاعاتی	آدرس وبگاه
Database name	URL
RAP-DB	<a href="http://rapdb.dna.affrc.go.jp/">http://rapdb.dna.affrc.go.jp/</a>
RISe	<a href="http://rise.genomics.org.cn/rice/index2.jsp">http://rise.genomics.org.cn/rice/index2.jsp</a>
RGAP	<a href="http://rice.plantbiology.msu.edu/">http://rice.plantbiology.msu.edu/</a>
Gramene*	<a href="http://www.gramene.org/Oryza_sativa/index.html">http://www.gramene.org/Oryza_sativa/index.html</a>
NCBI*	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
Phytozome*	<a href="http://www.phytozome.net/Phytozome_info.php">http://www.phytozome.net/Phytozome_info.php</a>

\* علاوه بر برنج اطلاعات ژنوم سایر گونه‌ها را نیز ارائه می‌کنند.

\* Represent genome information of rice and other species.

پایگاه‌های اطلاعاتی غلات بی‌همتا است که این به دلیل گونه‌های زیاد و اطلاعات جامعی است که در این پایگاه برای هر یک از این گونه‌ها وجود دارد (Childs, 2009; Mochida and Shinozaki, 2010). جستجوی اطلاعات در گرامنه از طریق وارد کردن کلمات کلیدی در بخش جستجوی سریع (Quick Search) که در صفحه اصلی قرار دارد، امکان‌پذیر است. همچنین با وارد شدن به منوی جستجو (Search) و انتخاب بخش مورد نظر (مثلاً بخش نشانگرها) و وارد کردن کلمات کلیدی نیز می‌توان جستجو را انجام داد. اخیراً برای سهولت دسترسی کاربران به اطلاعات وبگاه، طراحی صفحه اصلی به گونه‌ای تغییر داده شده است که به راحتی می‌توان در صفحه اصلی بخش مورد نظر را انتخاب نمود (شکل ۱). برای جستجو در هر بخش باید از کلمات کلیدی مخصوص آن بخش استفاده نمود. جدول ۵ کلمات مجاز برای هر بخش را ارائه نموده است.

#### ژنوم (Genomes)

همان‌طور که گفته شد، گرامنه کامل‌ترین اطلاعات ژنوم برنج را در خود جای داده است. در پایگاه گرامنه، بررسی میزان حفاظت ساختار کروموزوم و ژن با روش تطابق کل ژنوم (Whole genome alignments) با استفاده از Ensemble برای حدود ۲۰ گونه گیاهی انجام شده و امکان دسترسی به آن از طریق مسیر <http://www.gramene.org/info/docs/compara/analyses.html#blastz> فراهم شده است. همچنین، برای درک صحیح روابط تکاملی، امکان ترسیم درخت ژنی با استفاده از اطلاعات ژنوم گونه‌های برنج *Oryza glaberrima* و *O. sativa* (indica و japonica) و برای چند گیاه دیگر وجود دارد (Youens-Clark et al., 2011).

#### تنوع ژنتیکی (Genetic Diversity)

بخش تنوع ژنتیکی حاوی اطلاعات آلی نشانگرهای SSR، AFLP، RFLP و SNP و نیز اطلاعات QTLها می‌باشد که توضیحاتی در مورد ژرم‌پلاسم‌های برنج، ذرت، گندم، آرابیدوپسیس و سورگوم ارائه کرده و همچنین اطلاعات فنوتیپی برای برنج و ذرت نیز در آن موجود می‌باشد. در حقیقت، این اطلاعات شامل اطلاعات آلی نشانگرهای SSR، RFLP، AFLP و اطلاعات فنوتیپی برای برنج، اطلاعات فنوتیپی و SNP برای ذرت و اطلاعات

پایگاه اطلاعاتی NCBI نیز منبع کاملی از اطلاعات ژنومی می‌باشد. این پایگاه، پروژه ژنومی تعداد زیادی از گونه‌های جانوری و گیاهی را نمایه کرده است که از بین گیاهان عالی، تنها به ارایه پروژه‌های ژنومی آرابیدوپسیس، برنج (*O. sativa japonica*) و ذرت اکتفا نموده است. همچنین Phytozome یک پایگاه اطلاعاتی به منظور تسهیل مطالعات مقایسه‌ای و تکاملی گیاهان سبز است. در Phytozome v 7.0 امکان دسترسی به ژنوم ۲۵ گیاه سبز که ژنوم آن‌ها توالی‌یابی و تفسیر شده است، وجود دارد و این گیاهان از نظر خانواده‌های ژنی در ۱۱ گروه ژنی گروه‌بندی شده‌اند.

#### پیشگویی ژن

کشف مکان‌های ژنی و پیشگویی ساختارهای دقیق اینترون-اگزون آن‌ها، کشف یا پیشگویی ژن نام دارد (Itoh, 2007). ژن‌های کشف شده، اساس مطالعات آزمایشگاهی و محاسباتی بیشتری مانند تجزیه‌های تکامل مولکولی، برچسب زدن ترانسپوزون‌ها (Transposon tagging) و آزمایش‌های ریز آرایه‌ها خواهند بود. یکی از راه‌های پیشگویی ژن، استفاده از اطلاعات آماری به دست آمده از ژن‌های شناخته شده قبلی است (Hirano et al., 2008). فهرست برنامه‌هایی که به صورت برخط (Online) پیشگویی ژن در برنج را انجام می‌دهند، در جدول ۳ ارائه شده است.

#### پایگاه اطلاعات گرامنه

گرامنه (<http://www.gramene.org>) یک پایگاه اطلاعاتی برخط است که امکان دسترسی آسان به منابع دقیق و معتبر، نقشه‌های ژنتیکی و فیزیکی، نقشه‌های مبتنی بر توالی، نشانگرها، ژنوم‌ها، پروتئین‌ها، QTLها و همچنین ابزارهای تجزیه مقایسه‌ای را فراهم می‌کند (Jaiswal et al., 2005). در این پایگاه، امکان مقایسه نقشه‌های ژنومی غلات مهم خانواده گرامینه شامل برنج، ذرت، گندم، جو، چاودار، ارزن، یولاف و سورگوم وجود دارد. علاوه بر این، گرامنه یک منبع جامع و کامل اطلاعاتی برای انواع گونه‌های برنج است (جدول ۴). در حال حاضر، گرامنه پروژه‌ای گروهی است که با همکاری آزمایشگاه Cold Spring Harbor، دانشگاه کرنل (Cornell University) و انجمن برنج (The rice community) اداره می‌شود (Jaiswal et al., 2005). به اعتقاد بسیاری از محققین، این پایگاه در بین تمام

جدول ۳- وبگاه‌های حاوی برنامه‌های برخط برای پیشگویی ژن در برنج

(Hirano *et al.*, 2008)

Table 3. Websites include online programs for gene prediction in rice (Hirano *et al.*, 2008)

وبگاه Website	آدرس وبگاه URL
FGENESH+	http://www.softberry.com/
JIGSAW	http://www.cbcb.umd.edu/software/jigsaw/
RiceGAAS	http://ricegaas.dna.affrc.go.jp/usr/
Agene	http://servers.binf.ku.dk/agene/
Augustus	http://augustus.gobics.de/
GeneZilla	http://www.genezilla.org/
GlimmerHMM	http://cbcb.umd.edu/software/glimmerhmm/

جدول ۴- مقدار اطلاعات موجود در پایگاه گرامنه برای برنج

Table 4. Information content available in Gramene database for rice

Markers	نشانگرها	82776
Genes	ژن‌ها	2713
QTL		8646
Proteins	پروتئین‌ها	158224
Maps	نقشه‌ها	102
Pathways	مسیرها	۳۵۰ مسیر برای ۴۵۴۲۰ ژن 350 Rice Pathways for 45420 genes
Diversity	تنوع	۱۶۱۷ نشانگر برنج و ۸۴۱ ژرم پلاسم 1617 Rice Markers and 841 Germplasm

Adapted from ver. #39, 2014.

بر گرفته از نسخه #39, 2014.

[http://www.gramene.org/species/oryza/rice\\_stat.html](http://www.gramene.org/species/oryza/rice_stat.html)

شکل ۱- صفحه اصلی گرامنه. جستجوی اطلاعات در گرامنه از طریق وارد کردن کلمات کلیدی در بخش جستجوی سریع و یا از طریق منوی جستجو امکان پذیر است.

Figure 1. Home page of Gramene. Searching information in Gramene may via entering keywords in "Quick Search" section or "Search" menu.



جدول ۵- کلمات کلیدی برای جستجو در هر بخش از پایگاه گرامنه

Table 5. Key words for searching in each section of gramene database

زمینه جستجو Search category	کلمات کلیدی مجاز Permissive key words
نشانگر Marker QTL	نام نشانگر (مانند: RM237 یا C104) و نوع گونه Marker name (example: RM237 or C104) and species type نام صفت (مانند: Amylose Content)، سمبل صفت (مانند: AMCYN)، نوع صفت (مانند: Quality)، کلمات مترادف صفت (مانند: Apparent Amylose Content)، گروه لینکاژی، شماره کروموزوم، سمبل QTLهای منتشر شده و QTL accession ID (مانند: AQB001) Trait name (example: Amylose Content), trait symbol (example: AMCYN), trait type (example: Quality), trait synonym terms (example: Apparent Amylose Content), linkage group, chromosome number, published QTLs symbols and QTL accession ID (example: AQB001)
تنوع ژنتیکی Genetic diversity	نام ژرم پلاسِم (مانند: IR36)، شماره ژرم پلاسِم (مانند: RA4968)، نام نشانگر (مانند: RM22) یا مکان ژنی Germplasm name (example: IR36), germplasm number (example: RA4968), marker name (example: RM22) or gene loci
ژن Gene پروتئین	نام یا سمبل ژن، شماره کروموزوم، صفت و... Gene name or symbol, chromosome number, trait and etc نام پروتئین (مانند: granule-bound starch synthase)، نام ژن (مانند: waxy)، نام گونه (مانند: indica) یا <i>japonica</i> ، نام رقم (مانند: IR36)، شماره مخصوص آن پروتئین در پایگاه اطلاعاتی SWISSPROT (مانند: BAA01272.1)، نوع ID پایگاه اطلاعاتی SPTreMBL (مانند: BAA01272.1)، Protein_ID و Genbank GI شماره (مانند: UGST-ORYSA)
Protein	Protein name (example: granule-bound starch synthase), gene name (example: waxy), species name (example: indica or japonica), cultivar name (example: IR36), Specific number of protein in SWISSPROT database (SWISSPROT accession number) (example: BAA01272.1), SPTreMBL database ID type (example: UGST-ORYSA), Genbank GI number and Protein_ID
ژرم پلاسِم Germplasm مقالات و منابع Literatures	نام گونه (مانند: japonica)، شماره ژرم پلاسِم یا ژنوتیپ Species name (example: japonica), germplasm or genotype number نام نویسنده، عنوان ژورنال یا PubMed ID Author's name, journal title or PubMed ID
هستی‌شناسی Ontology	عبارت اسمی (مانند: days to heading, defense response و Zea mays)، accession یا Ontology ID مورد نظر (مانند: TO:0000303) یا اسامی مترادف (مانند: Culm length یا panicle). (توجه: نام و سمبل ژن را نمی‌توان به عنوان کلمات کلیدی در این بخش وارد نمود) Nominal term (example: days to heading, defense response and Zea mays), accession or Ontology ID (example: TO:0000303) or synonym names (example: Culm length or panicle). (Note: gene name and symbol can not be entered as keywords in this section)

را از این درگاه دریافت کنند. امکان مقایسه بین این مسیره‌ها و ژن‌های همراه آن‌ها نیز وجود دارد.

### پروتئین‌ها (Proteins)

بخش پروتئین پایگاه اطلاعاتی گرامنه حاوی اطلاعات معتبری از پروتئین‌های کشف شده از خانواده گرامینه می‌باشد که از پایگاه اطلاعاتی SWISSPROT- Tremble و پایگاه‌های مهم دیگر گردآوری شده است. این اطلاعات شامل کارکرد مولکولی محصولات ژن‌ها یا پروتئین‌ها، فرایندهای بیولوژیکی که هر پروتئین در آن درگیر است و اجزای سلولی که محل قرارگیری پروتئین‌ها است، می‌باشد.

برای گندم می‌باشد. بخش تنوع ژنتیکی این امکان را به کاربران می‌دهد تا تغییرات آلی را در مکان‌های ژنی مشخص در بین یک یا چند ژرم پلاسِم از یک آزمایش خاص مشاهده نمایند.

### مسیره‌ها (Pathways)

در بخش مسیره‌ها، پایگاه اطلاعاتی گرامنه، اطلاعات مسیره‌های متابولیکی هشت گونه از جمله برنج را از سه پایگاه تخصصی تر EcoCyc (Keseler et al., 2009)، PlantCyc (Zhang et al., 2010) و MetaCyc (Caspi et al., 2009) تأمین می‌کند. کاربران قادر خواهند بود، فهرست ژن‌های درگیر در هر مسیر متابولیکی



**ژن‌ها (Genes)**

بخش ژن‌ها در گرامنه منبع معتبری است که در آن می‌توان به اطلاعات جامعی در مورد ژن‌های کشف شده غلات دسترسی یافت. این اطلاعات شامل توضیحاتی در مورد ژن‌ها و آلل‌های همراه با فنوتیپ‌های مهم صفات مورفولوژیکی، رشد و زراعی و همچنین خصوصیات مختلف فیزیولوژیکی، عملکردهای بیوشیمیایی و آیزوایم‌ها می‌باشد. از دیگر اطلاعاتی که در این بخش می‌توان به آن دسترسی پیدا کرد، موقعیت روی نقشه، توالی‌ها، محصولات ژن، ژرم‌پلاسمی که ژن در آن کشف شده است، منابع و کلمات مرتبط می‌باشد.

**هستی‌شناسی (Ontology)**

این بخش شامل یک واژه‌نامه از کلمات کلیدی است که به صورت یک ساختار منظم از مفاهیم زیست‌شناسی در کنار هم قرار گرفته‌اند. با افزایش تعداد آزمایش‌های ژنومیکس و در پی آن تولید اطلاعات زیاد در مورد تظاهر ژن‌ها و تجزیه و تحلیل‌های فنوتیپی، نیاز به استفاده از لغات و کلمات کنترل شده (Ontologies) ضرورت بیشتری یافته است، زیرا این حجم عظیم اطلاعات، نام‌ها و کلمات مختلفی را برای یک مفهوم مشترک به همراه داشته‌اند و مقایسه بین گونه‌ای را مشکل نموده‌اند. این بخش این امکان را به کاربران می‌دهد تا ژن‌ها یا QTL های مرتبط را از طریق آناتومی یا صفات مرتبط پیدا نمایند و با استفاده از کلمات کلیدی مشترک موجود در این بخش می‌توان در پایگاه‌های اطلاعاتی مختلف به جستجوی اطلاعات پرداخت. این لغات در پایگاه اطلاعاتی گرامنه شامل: Plant Ontology (PO), Trait, Gene Ontology (GO), Ontology (TO).

**Gramene's و Environment Ontology (EO)**

taxonomy ontology (GR\_tax) می‌باشند.

**نشانگرها و نقشه‌ها (Markers and Maps)**

بخش نشانگرها این امکان را به کاربران می‌دهد تا با استفاده از کلمات کلیدی به اطلاعات جامعی از نشانگرهای مورد استفاده برای نقشه‌یابی ژنتیکی دسترسی پیدا کنند. برای همه نشانگرها، نام نشانگر، منشأ گونه‌ای و فهرستی از موقعیت‌های نقشه ارائه شده است، اما اطلاعات جزئی‌تر و کامل‌تر بسته به نوع نشانگر متفاوت می‌باشد.

**QTL**

بخش QTL دارای اطلاعات جامع و مفیدی در مورد QTL‌های شناسایی شده برای بسیاری از صفات زراعی در ده گونه غلات می‌باشد. تأکید بیشتر روی QTL‌های حاوی هر دو نوع اطلاعات در زمینه صفات و مکان‌های ژنی نقشه‌یابی شده روی نقشه ژنتیکی می‌باشد (Ni et al., 2009). موقعیت‌های QTL‌های برنج روی ژنوم گونه جاپونیکا کاملاً به وسیله تطابق نشانگرهای همراه با آن‌ها استنباط شده است.

**منابع و مستندات (Literature)**

این بخش در پایگاه اطلاعاتی گرامنه امکان دسترسی آسان به مقالات و منابع استناد شده را فراهم می‌نماید. پس از جستجو، علاوه بر ارائه مشخصات مربوط به منابع مربوطه، می‌توان مطالب مرتبط با موضوعاتی مثل پروتئین، QTL، ژن، آلل، جهش یافته‌ها، نقشه‌ها و نشانگرها را نیز مشاهده نمود. از دیگر امکانات این بخش این است که مسیرهایی برای دسترسی به سایر پایگاه‌های اطلاعاتی مثل PubMed و وبگاه مجلات مربوطه، ارائه شده است که در صورت تمایل می‌توان منابع مورد نظر را از پایگاه اصلی بارگذاری و مشاهده نمود.

**References**

- Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., Kaipa, P., Karthikeyan, A. S., Kothari, A. and Krummenacker, M. 2009. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research* 38 (Database issue): D473–D479.
- Childs, K. L. 2009. Genomic and genetic database resources for the grasses. *Plant Physiology* 142: 132–136.
- FAO. 2010. FAO Stat Yearbook. Rome, Italy.
- Flicek, P., Aken, B. L., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Coates, G. and Fairley, S. 2009. Ensembl's 10<sup>th</sup> year. *Nucleic Acids Research* 38 (Database issue): D557–D562.
- Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R. and Dunn, M. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296: 92–100.

- Hirano, H. Y., Hirai, A., Sano, Y. and Sasaki, T. 2008.** Rice biology in the genomics era. Springer-Verlag, Berlin, Heidelberg.
- Itoh, T. 2007.** Rice genome annotation: Beginnings of functional genomics. In: Upadhyaya, N. M. (Eds.), Rice functional genomics: Challenges, progress and prospects. Springer, Berlin Heidelberg, New York. pp: 21–31.
- Jaiswal, P., Ni, J., Yap, I., Ware, D., Spooner, W., Younes-Clark, K., Ren, L., Liang, C., Hurwitz, B., Zhao, W., Ratnapu, K., Faga, B., Canaran, P., Fogleman, M., Hebbard, C., Avraham, S., Schmidt, S., Casstevens, T., Buckler, E. S., Stein, L. and McCouch, S. 2005.** Gramene: a genomics and genetics resource for rice. **Rice Genetic Newsletter** 22 (1): 9-16.
- Jaiswal, P., Ware, D., Ni, J., Chang, K., Zhao, W., Schmidt, S. C., Pan, X., Clark, K. Y., Teytelman, L., Cartinhour, S., Stein, L. D. and McCouch, S. 2002.** Gramene: Development and integration of trait and gene ontologies for rice. **Comparative and Functional Genomics** 3: 132–136.
- Keseler, I. M., Bonavides-Martinez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R. P., Johnson, D. A., Krummenacker, M., Nolan, L. M., Paley, S. and Paulsen, I. T. 2009.** EcoCyc: A comprehensive view of *Escherichia coli* biology. **Nucleic Acids Research** 37 (Database issue): D464–D470.
- Mochida, K. and Shinozaki, K. 2010.** Genomics and bioinformatics resources for crop improvement. **Plant Cell Physiology** 51 (4): 497–523.
- Ni, J., Pujar, A., Youens-Clark, K., Yap, I., Jaiswal, P., Tecle, I., Tung, C. W., Ren, L., Spooner, W. and Wei, X. 2009.** Gramene QTL database: Development, content and applications. **Database**, doi: 10.1093/bap005.
- Ware, D. H., Jaiswal, P., Ni, J., Yap, I. V., Pan, X., Clark, K. Y., Teytelman, L., Schmidt, S. C., Zhao, W., Chang, K., Cartinhour, S., Stein, L. D. and McCouch, S. 2002.** Gramene: a tool for grass genomics. **Plant Physiology** 130: 1606-1613.
- Wing, R. A., Ammiraju, J. S., Luo, M., Kim, H., Yu, Y., Kudrna, D., Goicoechea, J. L., Wang, W., Nelson, W. and Rao, K. 2005.** The *oryza* map alignment project: The golden path to unlocking the genetic potential of wild rice species. **Plant Molecular Biology** 59: 53–62.
- Youens-Clark, K., Buckler, E., Casstevens, T., Chen, Ch., DeClerck, G., Derwent, P., Dharmawardhana, P., Jaiswal, P., Kersey, P., Karthikeyan, A. S., Lu, J., McCouch, S. R., Ren, L., Spooner, W., Stein, J. C., Thomason, J., Wei, S. and Ware, D. 2011.** Gramene database in 2010: Updates and extensions. **Nucleic Acids Research** 39: D1085–D1094.
- Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S. and Liu, B. 2002.** A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). **Science** 296: 79 – 92.
- Zhang, P., Dreher, K., Karthikeyan, A., Chi, A., Pujar, A., Caspi, R., Karp, P., Kirkup, V., Latendresse, M. and Lee, C. 2010.** Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and creation of metabolic pathways for plants. **Plant Physiology** 153: 1479–91.

**Short Communication****Bioinformatics and cereal genome databases: A case study in rice****Narjes Tabkhkar<sup>1</sup> and Babak Rabiei<sup>2\*</sup>**

1 and 2. Ph. D. Student and Prof., respectively, Dept. of Agronomy and Plant Breeding,  
Faculty of Agricultural Sciences, University of Guilan

---

(Received: October 14, 2013- Accepted: January 13, 2014)

---

**Abstract**

Advances of technology and the speed of genetic researches have enhanced genetic data in different fields like nucleotides and amino acids sequences, genes and proteins functions, mutants and their phenotypes, markers and etc. In addition to its volume, genetic information is complex and has subtle differences in the experimental manners which can influence on the quality and interpretation of the data. Thus it is crucial to describe and categorize the data related to genes and its phenotype information in the way that rapid and easy comparison of them be available. The use of genetic databases can help researchers to complete and update their researches and it awards them of the latest acquisitions of the science. Application of genetic databases in research for cereal is prevalent in recent years. This article will review databases which offer comprehensive information of rice genome and then one of the most usage cereal databases will be introduced.

**Keywords:** Bioinformatics, Cereal databases, Genetic information, Gramene

**Abbreviations:** IRGSP, International Rice Genome Sequencing Project; RAP, Rice Annotation Project; MSU-RGAP, Michigan State University- Rice Genome Annotation Project; RIS, Rice Information System; OMAP, *Oryza* Mapping Alignment Project; SSR, Simple Sequence Repeats; QTL, Quantitative Trait Loci.

---

\*Corresponding author: rabiei@guilan.ac.ir