

## آموزش آشنایی با بانک ژنی مرکز ملی اطلاعات بیوتکنولوژی: فرایند ثبت و پذیرش سکوانس ژنها

عبدالحسن کاظمی\*، عباس علی جعفری، امین موسوی آیت‌اللهی

مجله ایرانی آموزش در علوم پزشکی / بهار و تابستان ۱۳۸۶؛ ۷(۱): ۱۷۵ تا ۱۷۸.

http://www.expasy.org (of Bioinformatic) به آدرس DNA Database of Japan (DDBJ) و http://www.ddbj.nig.ac.jp با الگویی ویژه، پذیرش و نگهداری شده و در اختیار پژوهشگران قرار می‌گیرد. این مراکز اطلاعاتی هر سال نسبت به نوسازی سیستم و اطلاعات خود اقدام می‌نمایند (۲۰۱ و ۱۴ تا ۱۴).

NCBI در چهارم نوامبر ۱۹۸۸ در ایالات متحده آمریکا، با درک اهمیت روش‌های رایانه‌ای پردازش اطلاعات پزشکی و حیاتی، به عنوان بخشی از کتابخانه ملی پزشکی (National Library of Medicine-NLM) در مجموعه تشکیلات انستیتو ملی بهداشت (National Institute of Health-NIH) ایالات متحده آمریکا تأسیس گردید و هم اکنون با فاصله زیادی از دیگر مراکز مشابه، گسترده‌ترین مرکز اطلاعات پزشکی و بیولوژی مولکولی در دنیا محسوب می‌شود (۱۰ تا ۱۰).

در ایران، در سال ۱۳۷۸ شمسی، انستیتو بیوشیمی و بیوفیزیک (IBB) دانشگاه تهران، یک مرکز اطلاعات بیوانفورماتیکی تحت عنوان مرکز بیوانفورماتیک ایران (Iran's Bioinformatic Center-IBC) به آدرس http://www.abc.ut.ac.ir تأسیس نموده است که دربردارنده تسهیلات و امکاناتی در مورد سکوانس‌های پروتئینی و نوکلئوتیدی با اتصال به مراکز بیوانفورماتیکی متعددی است (۱).


حوزه‌ای جدید از دانش بشری که عرصه اجتماع زیست‌شناسی مولکولی، علوم رایانه‌ای و تکنولوژی اطلاعاتی می‌باشد، بیوانفورماتیک نام گرفته است و پایگاه‌های ثبت و پذیرش، نگهداری، پردازش و عرضه خام و یا پردازش شده سکوانس‌های مولکول‌های حیاتی، یعنی بانک‌های ژنی، در واقع پایگاه‌های علمی هستند که محصول علمی برای یک کاسه نمودن یافته‌های علمی و پژوهشی در زمینه زیست‌شناسی مولکولی، و ایجاد تسهیلات برای دسترسی جوامع علمی و گروه‌های تخصصی به سکوانس‌های مولکول‌های حیاتی به صورت همسان شده، می‌باشند (۶ تا ۶).

اطلاعات مربوط به سکوانس‌های مولکول‌های حیاتی شناسایی و ثبت شده در مراکز اطلاعات بیوانفورماتیکی مانند NCBI (National Center for Biotechnology Information) به آدرس http://www.ncbi.nlm.nih.gov و EBI (European Bioinformatic Institute) به آدرس http://www.ebi.ac.uk و EXPASY (Swiss Institute)

\* آدرس مکاتبه. دکتر عبدالحسن کاظمی (دانشیار) مرکز تحقیقات بیوتکنولوژی، مرکز تحقیقات اندوکراین و بیماری‌های متابولیک، مرکز تحقیقات بیماری‌های عفونی و گرمسیری، NPMC و مرکز تحقیقات انکولوژی و هماتولوژی دانشگاه علوم پزشکی تبریز، تبریز. E-mail: hassan5628@yahoo.com

دکتر عباس علی جعفری، استادیار گروه انگل‌شناسی دانشکده پزشکی دانشگاه علوم پزشکی شهید صدوقی یزد و دکتر امین موسوی آیت‌اللهی، استادیار گروه انگل‌شناسی دانشکده پزشکی دانشگاه علوم پزشکی کرمان.

این مقاله در تاریخ ۸۴/۲/۴ به دفتر مجله رسیده، در تاریخ ۸۶/۳/۲ اصلاح شده و در تاریخ ۸۶/۳/۱۵ پذیرش گردیده است.

GenBank submission, enter the size in nucleotides of your contiguous sequence here and press  شروع می‌شود. با فشار دادن

تکمه فعال  در این قسمت، به صفحه دیگری وارد می‌شویم که مجدداً دارای عنوان BankIt: GenBank Submissions by WWW بوده و در آن یک شماره اولیه ثبت سکوانس معرفی شده به بانک ژنی NCBI به عنوان شماره مرجع به کاربر داده می‌شود. این صفحه دارای بخش‌هایی متعدد برای ادامه مراحل بعدی ثبت سکوانس است.

در مرحله اول، مشخصات فردی و آدرس فرد معرفی‌کننده سکوانس به بانک ژنی در قسمتی با عنوان Multiple Submissions Information و زیر مجموعه Reference Information خواسته می‌شود و سپس نظر فرد معرفی‌کننده سکوانس برای تاریخ انتشار و معرفی سکوانس در پایان مراحل پردازش علمی و پذیرش سکوانس، با الگوی:

#### Release Information

When may we release your sequence record?

Immediately After Processing

Release Date

January 1 2005

درخواست می‌گردد. ارائه اطلاعات علمی مربوط به آدرس آکادمیک مؤلفین (معرفی‌کنندگان) اطلاعات، تعلق ژن به نوعی مشخص از موجودات زنده، دستکاری و یا عدم دستکاری ژنتیکی موجود زنده، منبع ژن و... درخواست‌های بعدی بانک ژنی است. سپس باید توالی کامل ژن همراه با اطلاعات مربوط به خود ژن در پنجره موجود در صفحه وارد گردد. در آخرین قسمت از این صفحه، امکان درج هر گونه اطلاعات اضافی و اختیاری مورد نظر معرفی‌کنندگان سکوانس فراهم شده است.

با انجام مراحل پیش‌گفت، ثبت اولیه سکوانس در بانک ژنی NCBI خاتمه یافته و با کلیک بر روی تکمه فعال موجود در صفحه با عنوان Validate the submission

هر یک از مراکز اطلاعات بیوانفورماتیکی، معمولاً با اتصال به سایر مراکز اطلاعات بیوانفورماتیکی، به صورت یک شبکه بسیار گسترده، امکان دسترسی پژوهشگر وارد شده به یک مرکز را به اطلاعات مراکز اطلاعات بیوانفورماتیکی دیگر فراهم می‌کند (۱ تا ۳). هدف از این نوشتار، آموزش و آشنایی با فرایند ثبت و پذیرش سکوانس‌های ژن‌ها در بانک ژنی NCBI است.

توالی مولکول‌های تشکیل‌دهنده ژن‌های آنزیم‌های فسفولیپاز D1 و B1، B2، B3 در هسته سلول‌های قارچ کپکی بیماریزا به نام اسپرژیلوس فومیگاتوس با استفاده از روش‌های مختلف مولکولار بیولوژی و مهندسی ژنتیک برای اولین بار در سطح جهانی شناسایی شدند.

این اطلاعات با هدف دسترسی سایر پژوهشگران به این یافته‌ها، به بانک ژنی NCBI با استفاده از تسهیلات موجود در سایت NCBI معرفی شدند. برای معرفی اولیه هر سکوانس به بانک ژنی NCBI، از بخش ویژه موجود در سایت این مرکز با عنوان Submitting Sequence Data to GenBank <http://www.ncbi.nlm.nih.gov/Genbank/index.html> استفاده می‌شود (۷ تا ۱۰). مسیر BankIt برای معرفی اولیه سکوانس‌های نوکلئوتیدی انتخاب گردید. این مسیر با عنوان For quick and simple submissions مشخص شده، امکانات متفاوتی برای معرفی و ثبت سکوانس‌های بیولوژیکی به صورت الکترونیکی با عنوان BankIt: GenBank Submissions by WWW همراه با راهنمایی‌های اساسی برای معرفی سکوانس در اختیار کاربر قرار می‌گیرد که با عنوان GenBank Direct Submission Options مشخص شده است. برای توجیه بهتر کاربر نیز مثالی با عنوان annotation examples and descriptions نیز وجود دارد.

اولین مرحله عملی برای معرفی و ثبت سکوانس مورد نظر به بانک ژنی NCBI در این قسمت با وارد کردن طول سکوانس به قسمتی مشخص شده به صورت: To prepare a New

نمی سازد، مضافاً اینکه اطلاعات خام موجود در این سکوانس‌ها نیز همه قابلیت‌های اطلاعاتی و حیاتی این مولکول‌ها را نشان نمی‌دهد. پردازش بیوانفورماتیکی این اطلاعات، جزئیات بسیار زیادی را در اختیار پژوهشگران قرار می‌دهد که استنباط این اطلاعات ثانویه از متن سکوانس‌های اولیه به سادگی مقدور نیست.

نرم‌افزارهای متعدد تعبیه شده و دارای قابلیت‌های متفاوت در سایت‌های بیوانفورماتیکی این امکان را برای پژوهشگران زمینه‌های مختلف علوم حیاتی و پزشکی فراهم می‌آورد که نه فقط همه قابلیت‌های اطلاعات حیاتی این سکوانس‌ها را به دست بیاورند، بلکه حتی با مقایسه تشابهات و تفاوت‌های یک و یا چند ژن با ژن‌های دیگر، به جستجو در عرصه‌هایی از روابط حیاتی بپردازند که جز در سایه وجود بانک‌های ژنی، نمی‌توان به چنین عرصه‌هایی وارد شد (۱۲ تا ۹).

علاوه بر این، در این صفحات، اطلاعاتی دیگر به صورت فعال وجود دارند که با انتخاب این اطلاعات، می‌توان به طور اتوماتیک به شبکه‌ها، پایگاه‌ها، مقالات، و مراجع علمی حاوی اطلاعات وسیع جنبی مربوط به موضوع محوری متصل شد. همچنین وجود بانک‌های ژنی ویژه برای اطلاعات بیولوژیکی خاص، مثل بانک ژنی باکتری‌های بیماریزا (۱۰)، طبقه‌بندی باکتری‌ها (۱۱)، شناسایی ژن‌های ویروس‌های حیوانی (۱۲)، ژن‌های موش آزمایشگاهی (۱۳)، جستجوی سکوانس ژنوم انسان (۵) و حتی تفاوت‌های ژنتیکی بین افراد یک جامعه (۱۴) امکان دسترسی به اطلاعاتی وسیع با سرعتی فوق‌العاده را فراهم می‌آورد.

and correct other errors، پیامی به صورت اتوماتیک از طرف بانک ژنی NCBI برای معرفی‌کننده سکوانس با محتوای دریافت و پذیرش اولیه فرستاده می‌شود و متعاقب آن گردش مکاتبات الکترونیکی به صورت تعامل علمی بین دو طرف در مورد موضوعات مربوط به سکوانس بطور متقابل تا حصول توافق طرفینی ادامه می‌یابد که در مواردی تا مدت قابل توجهی استمرار دارد. در ایام نزدیک به قرارگیری نهایی سکوانس بیولوژیکی در پایگاه بانک ژنی NCBI برای دسترسی عمومی، مجدداً از طرف بانک ژنی نظر معرفی‌کنندگان ژن در مورد متن آماده شده و ایرادات احتمالی آن پرسیده می‌شود.

متعاقب پذیرش نهایی سکوانس‌های نوکلئوتیدی معرفی شده و ثبت آنها به مشخصات توافق شده در بانک ژنی NCBI، این سکوانس‌ها به صورت قابل دسترس در سایت NCBI قرار گرفتند که با کلید واژه‌های مناسبی نظیر عنوان ژن‌ها و نام مؤلفین و ثبت‌کنندگان، سکوانس‌های مربوطه در این سایت، قابل جستجو می‌باشند. باید توجه کرد که برای دسترسی به سکوانس نوکلئوتیدی ژن‌های ثبت شده، در صورت استفاده از پایگاه اطلاعاتی PubMed به آدرس <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>

باید گزینه Nucleotide را انتخاب نموده و سپس با درج کلید واژه‌های مناسب، ژن‌های مورد نظر را جستجو کرد. این سکوانس‌ها همچنین از طریق سایت‌های سایر بانک‌های ژنی هم به آسانی قابل دسترسی است.

ماهیت اطلاعات موجود در سکوانس‌های مولکول‌های حیاتی، بهره‌گیری کامل از اطلاعات موجود در این سکوانس‌ها را در صورت انتشار و طبع سنتی آنها مقدور

## منابع

1. Goliaei B. Bioinformatics (Computer application in molecular biology) IBB Uni of Tehran Press. 2003.
2. Berman HM, Westbrook J, Feng Z, Iype L, Schneider B, Zardecki C. The nucleic acid database. *Methods Biochem Anal* 2003; 44: 199-216.
3. Philippi S. Light-weight integration of molecular biological databases. *Bioinformatics* 2004 Jan 1; 20(1): 51-7.

4. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. Gen Bank. *Nucleic Acids Res* 2002 Jan 1; 30(1): 17-20.
5. Hiraakawa M. HOWDY: an integrated database system for human genome research. *Nucleic Acids Res* 2002 Jan 1; 30(1): 152-7.
6. Chim SS, Tsui SK, Chan KC, Au TC, Hung EC, Tong YK, et al. Genomic characterisation of the severe acute respiratory syndrome coronavirus of Amoy Gardens outbreak in Hong Kong. *Lancet* 2003 Nov 29; 362(9398): 1807-8.
7. Wheeler DL, Church DM, Lash AE, Leipe DD, Madden TL, Pontius JU, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2001 Jan 1; 29(1): 11-6.
8. Wheeler DL, Church DM, Lash AE, Leipe DD, Madden TL, Pontius JU, et al. Database resources of the National Center for Biotechnology Information: 2002 update. *Nucleic Acids Res* 2002 Jan 1; 30(1): 13-6.
9. Wheeler DL, Church DM, Edgar R, Federhen S, Helmberg W, Madden TL, et al. Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res* 2004 Jan 1; 32(Database issue): D35-40.
10. Harmsen D, Rothganger J, Frosch M, Albert J. RIDOM: Ribosomal Differentiation of Medical Microorganisms Database. *Nucleic Acids Res* 2002 Jan 1; 30(1): 416-7.
11. Watanabe K, Nelson J, Harayama S, Kasai H. ICB database: the gyrB database for identification and classification of bacteria. *Nucleic Acids Res* 2001 Jan 1; 29(1): 344-5.
12. Alba MM, Lee D, Pearl FM, Shepherd AJ, Martin N, Orengo CA, et al. VIDA: a virus database system for the organization of animal virus genome open reading frames. *Nucleic Acids Res* 2001 Jan 1; 29(1): 133-6.
13. Divina P, Forejt J. The Mouse SAGE Site: database of public mouse SAGE libraries. *Nucleic Acids Res* 2004 Jan 1; 32(Database issue): D482-3.
14. Hiraakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y. JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res* 2002 Jan 1; 30(1): 158-62.