



پروژه شماره 2 درس بیوانفورماتیک.

عنوان پروژه: شناسایی ژن های آنزیم زایلاناز از متازنوم شکمبه نشخوارکنندگان

اهمیت اکتشاف آنزیم های جدید از متازنوم

اکتشاف آنزیم های جدید از متازنوم ها، دریچه ای به سوی گنجینه ای عظیم از تنوع زیستی میکروبی و پتانسیل های بالقوه آن ها می گشاید. متازنوم، به عنوان مخازن ژنتیکی میکروارگانیسم های محیط های مختلف، حاوی اطلاعات ژنتیکی بی شماری هستند که می توانند منجر به کشف آنزیم های جدید با ویژگی های منحصربه فرد شوند. این آنزیم ها می توانند در صنایع مختلف، از جمله صنایع غذایی، دارویی، کشاورزی و بیوتکنولوژی، کاربردهای فراوانی داشته باشند. به عنوان مثال، آنزیم های تجزیه کننده پلیمرهای پیچیده مانند سلولز و همی سلولز، می توانند در تولید سوخت های زیستی و مواد شیمیایی تجدیدپذیر نقش مهمی ایفا کنند. همچنین، آنزیم های مقاوم به دما و pH بالا، می توانند در فرآیندهای صنعتی سخت و پرهزینه، کارایی و صرفه جویی قابل توجهی را به ارمغان آورند.

اهمیت آنزیم های Xylanase میکروبی

آنزیم های زایلاناز میکروبی، به عنوان درشت مولکول های قدرتمند در تجزیه زایلان (دومین پلی ساکارید فراوان در طبیعت)، نقش حیاتی در صنایع مختلف ایفا می کنند. زایلان، به عنوان یکی از اجزای اصلی دیواره سلولی گیاهان، پتانسیل بالایی در تولید سوخت های زیستی، مواد شیمیایی و خوراک دام دارد. آنزیم های زایلاناز، با شکستن زایلان به قندهای ساده تر، امکان استفاده بهینه از این منبع تجدیدپذیر را فراهم می کنند. در صنعت خمیر و کاغذ، زایلانازها می توانند با کاهش مصرف مواد شیمیایی و انرژی، فرآیند سفیدسازی را بهبود بخشند. همچنین، در تولید خوراک دام، زایلانازها با افزایش قابلیت هضم خوراک، می توانند عملکرد و سلامت دام را بهبود بخشند. با توجه به اهمیت روزافزون منابع تجدیدپذیر و توسعه پایدار، آنزیم های زایلاناز میکروبی به عنوان ابزارهای کلیدی در این مسیر، توجه روزافزونی را به خود جلب کرده اند.

مقدمه و هدف:

این پروژه با هدف شناسایی ژن های کدکننده آنزیم زایلاناز از متازنوم شکمبه نشخوارکنندگان طراحی شده است. زایلانازها آنزیم های تجزیه کننده زایلان (دومین پلی ساکارید فراوان در زیست کره) هستند و در صنایع مختلف از جمله تولید سوخت های زیستی، خوراک دام و کاغذ کاربرد دارند. متازنوم شکمبه نشخوارکنندگان به عنوان منبعی غنی از تنوع میکروبی، پتانسیل بالایی برای کشف زایلانازهای جدید با ویژگی های منحصربه فرد را داراست.

داده ها:

یک فایل حاوی توالی های کانتینگ اسمبل شده از متازنوم شکمبه نشخوارکنندگان در گوگل درایو به آدرس زیر قرار داده شده است:

<https://drive.google.com/file/d/14PGwsGuL2ouY-fv0yrzjGnBMSjREU6/view>



گام اول: شناسایی توالی‌های بالقوه زایلاناز (50 درصد از امتیاز کل)

در این گام، هدف شناسایی توالی‌هایی است که احتمالاً ژن‌های زایلاناز thermostabe را کد می‌کنند. برای این منظور، از رویکرد مبتنی بر شباهت توالی استفاده خواهد شد. توالی‌های کانتینگ با 11 آنزیم زایلاناز شناخته شده با ویژگی thermostability مقایسه می‌شوند. این توالی‌ها از طریق لینک زیر در دسترس هستند:

https://drive.google.com/file/d/1-dDyldAXKUK97rxw_n4zE_gzh2fztFRU/view?usp=sharing

اگر یک کانتینگ شباهت قابل توجهی با یک یا چند زایلاناز شناخته شده thermostabe داشته باشد، به عنوان یک توالی بالقوه زایلاناز پایدار در دماهای بالا در نظر گرفته می‌شود. تعیین معیار شباهت بر عهده دانشجویان است.

ابزارها و روش‌ها:

برای انجام این گام، می‌توان از ابزارهای بیوانفورماتیکی مختلفی استفاده کرد. به عنوان مثال Blast+ که از طریق لینک زیر در دسترس است:

<https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html#downloadblastdata>

ابزارهای دیگری هم به انتخاب دانشجویان قابل استفاده هستند.

مانند BLAST، DIAMOND یا HMMER استفاده کرد. این ابزارها امکان مقایسه سریع و کارآمد توالی‌ها را فراهم می‌کنند. همچنین، می‌توان از روش‌های مبتنی بر یادگیری ماشین برای بهبود دقت شناسایی توالی‌های زایلاناز استفاده کرد.

خروجی:

خروجی این گام، فهرستی از توالی‌های کانتینگ است که به عنوان توالی‌های بالقوه زایلاناز شناسایی شده‌اند. برای هر توالی لازم است ترجمه پروتئینی آن نیز به دست آید. این توالی‌ها می‌توانند در گام‌های بعدی برای بررسی‌های بیشتر و تایید عملکرد زایلانازی مورد استفاده قرار گیرند.

گام دوم: خوشه‌بندی و انتخاب نماینده (30 درصد از امتیاز کل)

پس از شناسایی توالی‌های بالقوه زایلاناز در گام اول، ممکن است تعدادی از این توالی‌ها بسیار شبیه به هم باشند. این شباهت بالا می‌تواند ناشی از افزونگی در داده‌های متاژنومی و یا وجود ایزوفرم‌های مختلف یک ژن باشد. برای کاهش افزونگی و ساده‌سازی تحلیل‌های بعدی، توالی‌های بسیار مشابه در خوشه‌هایی گروه‌بندی می‌شوند.

روش خوشه‌بندی:

در این پروژه، از الگوریتم‌های خوشه‌بندی توالی مانند CD-HIT استفاده می‌شود. CD-HIT یک ابزار سریع و کارآمد برای خوشه‌بندی توالی‌های DNA و پروتئین بر اساس شباهت آن‌ها است. این ابزار، توالی‌ها را بر اساس یک آستانه شباهت (در اینجا 97٪) در خوشه‌ها گروه‌بندی می‌کند.

انتخاب نماینده:



پس از خوشه‌بندی، از هر خوشه یک توالی به‌عنوان نماینده انتخاب می‌شود. نماینده معمولاً طولانی‌ترین توالی در خوشه یا توالی با بیشترین شباهت به سایر اعضای خوشه است. انتخاب نماینده به این دلیل انجام می‌شود که تحلیل‌های بعدی را بر روی یک مجموعه کوچک‌تر و غیر افزونه از توالی‌ها متمرکز کنیم.

در گام سوم لازم است که یک مدل برای طولانی‌ترین ناحیه conserved یک زیر خانواده از زایلانازها ساخته شود. انتخاب این زیرخانواده بر عهده دانشجویان است. به عنوان مثال زایلانازهای ترموفیل، فعال در pH های اکستریم اسیدی یا بازی، زایلانازهای قارچی، ...

این مدل می‌تواند بر مبنای روش دلخواه دانشجویان ساخته شود. به عنوان مثال PSSM، مدل‌های مخفی مارکوف، عبارات منظم، ... و یا ترکیبی از آنها. این مدل باید برای فیلتر توالی‌ها استفاده شود تا به توالی‌های محدود تری برسیم.

گام سوم: مدل‌سازی ناحیه حفاظت‌شده و فیلتر کردن توالی‌ها (20 درصد از امتیاز کل)

در این گام، هدف ساخت یک مدل برای ناحیه حفاظت‌شده (conserved region) طولانی‌ترین توالی زیرخانواده thermostable از زایلانازها است.

روش‌های مدل‌سازی:

برای مدل‌سازی ناحیه حفاظت‌شده، می‌توان از روش‌های مختلفی استفاده کرد، از جمله:

- **ماتریس امتیازدهی موقعیت-ویژه (PSSM):** PSSM یک ماتریس است که احتمال حضور هر اسیدآمینه را در هر موقعیت از یک ناحیه حفاظت‌شده نشان می‌دهد. این ماتریس‌ها معمولاً از طریق هم‌تراز کردن چندگانه توالی‌های مرتبط ساخته می‌شوند.
- **مدل‌های مخفی مارکوف (HMM):** HMMها مدل‌های احتمالاتی هستند که می‌توانند برای نمایش نواحی حفاظت‌شده در توالی‌های پروتئینی استفاده شوند. HMMها نسبت به PSSMها پیچیده‌تر هستند، اما می‌توانند اطلاعات بیشتری را در مورد ساختار و عملکرد ناحیه حفاظت‌شده ارائه دهند.
- **عبارات منظم (Regular Expressions):** عبارات منظم الگوهایی هستند که می‌توانند برای توصیف توالی‌های زیستی استفاده شوند. عبارات منظم می‌توانند برای شناسایی نواحی حفاظت‌شده ساده و کوتاه استفاده شوند.
- **سایر روش‌ها یا ترکیبی از روش‌های فوق:** می‌توان از ترکیبی از روش‌های فوق برای ساخت یک مدل قوی‌تر و دقیق‌تر برای ناحیه حفاظت‌شده استفاده کرد.

فیلتر کردن توالی‌ها:

پس از ساخت مدل، می‌توان از آن برای فیلتر کردن توالی‌های نماینده خوشه‌ها استفاده کرد. توالی‌هایی که با مدل مطابقت خوبی دارند، به‌عنوان کاندیدهای قوی‌تر برای ژن‌های زایلاناز در نظر گرفته می‌شوند.

خروجی:



خروجی این گام، مجموعه‌ای محدودتر از توالی‌های کاندید زایلاناژ است که با مدل ناحیه حفاظت‌شده زیرخانواده thermostabe مطابقت دارند. این توالی‌ها می‌توانند در گام‌های بعدی برای بررسی‌های دقیق‌تر و تایید عملکرد زایلاناژی مورد استفاده قرار گیرند.

اطلاعات تکمیلی پروژه

- **کار گروهی:** پروژه به صورت گروهی انجام می‌شود. هر گروه می‌تواند حداکثر ۳ عضو داشته باشد.
- **سرگروه:** هر گروه باید یک سرگروه داشته باشد.
- **اسامی اعضا:** سرگروه موظف است اسامی اعضای گروه را در گروه اسکایپ قرار دهد.
- **نمره‌دهی:** نمره پروژه به صورت یکسان به همه اعضای گروه تعلق می‌گیرد.
- **زمان تحویل:** زمان تحویل پروژه به احتمال زیاد روز یکشنبه 14 بهمن ماه خواهد بود. تایید نهایی زمان تحویل متعاقباً اعلام خواهد شد.
- **گزارش:** هر گروه باید یک گزارش مکتوب شامل جزئیات انجام پروژه به همراه فایل‌های تولید شده ارائه کند.

جزئیات گزارش:

گزارش باید شامل بخش‌های زیر باشد:

1. **مقدمه:** توضیح مختصری در مورد اهمیت زایلاناژها و اهداف پروژه.
2. **روش‌ها:** شرح دقیق مراحل انجام پروژه (گام‌های ۱ تا ۴).
3. **نتایج:** ارائه نتایج حاصل از هر گام به همراه جداول، نمودارها و تصاویر مرتبط.
4. **بحث:** تفسیر نتایج، مقایسه با مطالعات قبلی، و بیان محدودیت‌ها و پیشنهادات برای تحقیقات آینده.
5. **نتیجه‌گیری:** خلاصه نتایج اصلی و دستاوردهای پروژه.
6. **منابع:** فهرست منابع مورد استفاده در پروژه.

فایل‌های تولید شده:

فایل‌های زیر باید به همراه گزارش ارائه شوند:
 فایل‌های حاوی توالی‌های شناسایی‌شده در هر گام
 فایل‌های حاوی مدل‌های ساخته‌شده (PSSM)، HMM، یا عبارت منظم

توجه:

رعایت اصول اخلاق پژوهشی (عدم سرقت ادبی، ارجاع صحیح به منابع) در گزارش ضروری است.
 استفاده از Gemini، ChatGPT، و مانند آنها، با ذکر دلیل و نحوه استفاده مجاز است!
 گزارش باید به صورت واضح، مختصر، و علمی نوشته شود.

موفق باشید!

کاوه کاوسی