

به نام خدا

درخت زندگی ابولا و یروس

گزارش تئوری پروژه درس مبانی یوانفورماتیک

زهرابابی

۹۳۱۰۵۴۸۳

ترم پاییز ۹۶

کدام پروتئین ها ساختاری و کدام غیرساختاری هستند؟

پروتئین های ساختاری آن دسته از پروتئین های ویروسی هستند که ملحق به ساختار خود ویروس می شوند، حال آنکه پروتئین های غیر ساختاری در سلول آلوده بیان میشوند. حال در این ویروس این پروتئین ها به این شرح اند:

NP : این پروتئین که نوکلئوپروتئین هم نام دارد از ژنوم محافظت میکند، به این صورت که دور RNA یک پوشش می شود. پس این پروتئین ساختاری است.

VP35 : این پروتئین جزو RNA binding هاست و ساختاری است.

VP40 : این ژن پروتئین ماتریکس را تولید میکند که وظیفه شکل دهی به ویروس را برعهده دارد. پس واضح است که پروتئین ساختاری است.

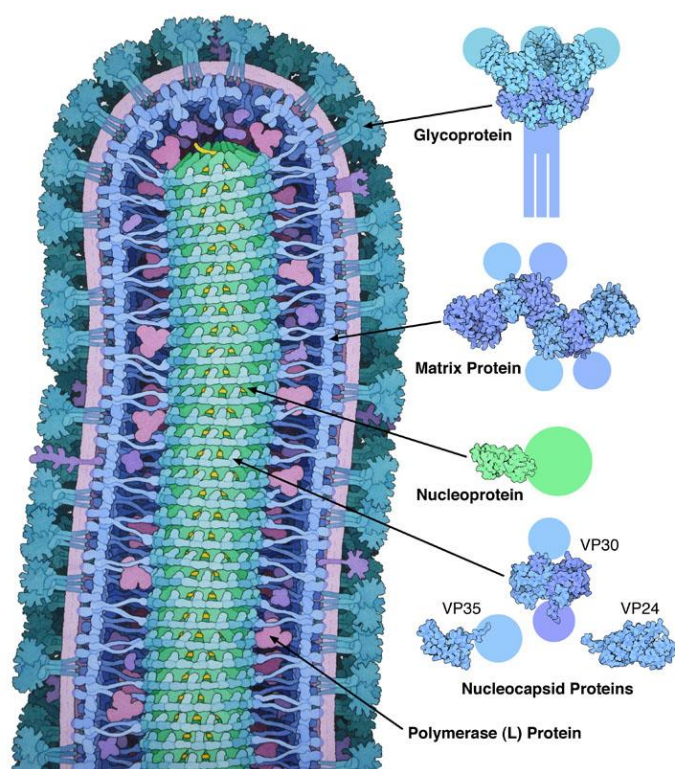
GP : این پروتئین یک گلیکوپروتئین است که از ویروس ترشح می شود و به سلول های میزبان می چسبد و در نتیجه غیر ساختاری است.

VP30 : این پروتئین هم ساختاری و جزو پروتئین های ملخ شونده به RNA می باشد.

VP24 : این پروتئین در بدن بیمار به سلول های ایمنی ملحق شده و آن ها را مهار میکند و در نتیجه غیر ساختاری است.

L : این پروتئین غیر ساختاری است، زیرا پلیمرز است و ویروس ها در سلول میزبان آلوده RNA خود را تکثیر میکنند.

این شکل تاحدی میتواند ساختاری بودن برخی از پروتئین ها نمایش دهد:



کدام پروتئین باعث ایجاد بیماری می شود؟

پروتئین های L و VP30 و VP35 که در تکثیر ویروس نقش دارند سبب گسترش ویروس در بدن فرد و در نتیجه بیماری میشوند. پروتئین VP24 به طور خاص به علت از کار انداختن سلول های ایمنی نقش مهمی در بیماری دارد.

۱.۳

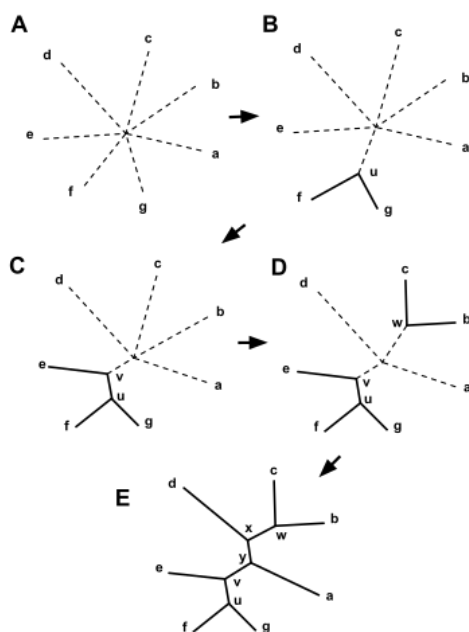
ابتدا ویروس وارد بدن فرد می شود و به سرعت شروع به تکثیر خود می کند. در این مدت اولیه هنوز بدن از حضور ویروس آگاه نشده و در نتیجه ویروس بدون مزاحمت تکثیر میشود. همچنین ویروس با ترشح گلیکوپروتئین همانطور که در بخش قبل گفته شده سلول های ایمنی بدن را از کار می اندازد. بعد از حدود یک هفته پاسخ ایمنی شروع میشود. در پایان ویروس در بافت های بدن پخش میشود و بافت ها را تخریب میکند و موجب خونریزی و مرگ فرد میشود. علائم بیماری با تب و .. شروع میشود و آخرین علامت خونریزی است. پس بدین ترتیب این ویروس مسیر سیستم ایمنی بدن را هدف قرار داده و از کار می اندازد.

۳.۱

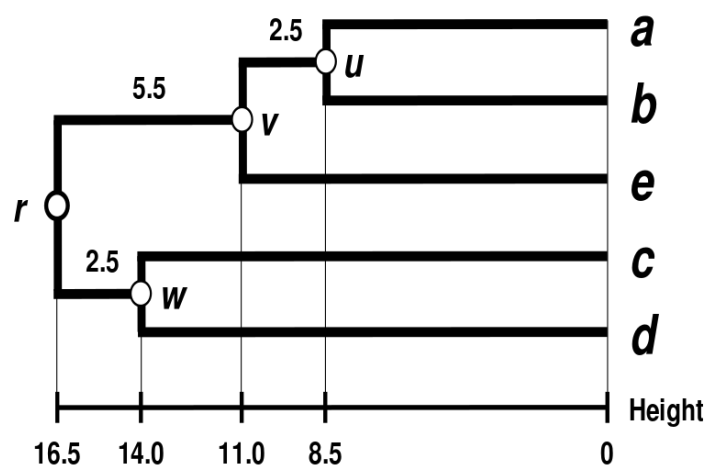
نتایج حاصل از دو الگوریتم برای ژنهای مختلف متفاوت است. اما NJ دسته بندی بهتری انجام میدهد. به طور کلی NJ قابل اعتمادتر است، خصوصا در درخت هایی که از ماتریس فواصل به دست می آیند زیرا داده ورودی ماتریس برای الگوریتم upgma باید ultra metric باشد که در اینجا اینطور نیست. از طرفی برای NJ باید ماتریس additive باشد اما تقریبا همیشه اگر ماتریس اینگونه نباشد نیز خوب جواب میدهد.

من در درخت ها تفاوت چشم گیر مشاهده کردم درحالی که انتظار نداشتم. چندبار ژن های پیدا شده را چک کردم و مشکلی پیدا نکردم. در نهایت به این نتیجه رسیدم که شاید این تفاوت بین ژن های مختلف به علت این است که جهش ها در طول زمان به ژن وابسته بوده است. یعنی اینطور نبوده که کل ژنوم به صورت یونیفرم تغییر کند. بنابراین ژنهای مختلف درخت های مختلفی دارند. (اینها نظرات شخصی است).

تفاوت بین درخت های دو الگوریتم نیز واضحا به علت تفاوت روند آن دو است. الگوریتم NJ به این صورت عمل میکند که ابتدا یک ماتریس Q می سازد و سپس به صورت حریصانه هربار کمترین مقدار را انتخاب میکند و فاصله اش با قبلیها پیدا میکند و آن را به آنها اضافه میکند:



الگوریتم upgma به این نحو است که در هر قدم دو خوشه ای که نزدیکترین هستند به هم جویین میشوند و این روند ادامه پیدا میکند تا درخت ساخته شود. همچنین همه برگ ها را در یک سطح می آورد:



الگوریتم NJ درخت ریشه دار نمی سازد. بنابراین وقتی در مسئله ای مثلا به دنبال جد مشترک هستیم مناسب نیست حال آنکه upgma برای این دست مسائل به دلیل ریشه دار بودن مناسب تر است. از طرفی اگر داده های ماتریس ultra metric نباشند خوب کار نمیکند و قابل اعتماد نیست و بهتر است از NJ استفاده کنیم. اگر ماتریس فواصلمان additive نباشد هم بهتر است از NJ استفاده نکنیم. اگر داده های evaluation از یک strict molecular clock پیروی میکنند الگوریتم upgma مناسب است و برعکس.

کتابخانه بایوپایتون که من از آن برای ترکیب درخت ها استفاده کردم سه متد برای این کار دارد، adam، strict majority. در این میان تنها majority است که علاوه بر درخت ها یک متغیر cutoff هم میگیرد که به صورت دیفالت صفر است. این متغیر که عددی بین صفر و یک است مشخص میکند از کدام خوشه ها استفاده کنیم. اگر یک باشد یعنی فقط خوشه هایی که در همه درخت ها آمده باشند و با تغییر ان میتوان این را تغییر داد. من از مقدار 0.5 استفاده کردم.

درخت من در این بخش در الگوریتم NJ اندکی با درخت حاصل از بخش قبلی تفاوت داشت. این تفاوت در zaire و sudan بود که در این دو درخت جابه جا بودند. (الگوریتم معرفی شده بسیار جالب بود و از نحوه کار و سرعت آن شگفت زده شدم 😊)

میخواهیم نقطه شروع به دست آوریم پس نیاز به درخت ریشه دار داریم و از این رو upgma کارتر است.

بعد از بستن گزارش کد متوجه شدم این بخش آخر ذخیره نشده است و دیگر اجازه ویرایش نداد. به همین جهت در اینجا گزارش کد را هم می آورم.

حالا می خواهیم فاصله زمانی تا جد مشترک را به دست آوریم. در بخش قبل با اضافه کردن ماربرگ به گونه ها توانستیم درخت ریشه دار بسازیم که ریشه درخت همان جد مشترک گونه های ماست. در این بخش بار دیگر با الگوریتم upgma درخت ریشه دار می سازیم اما این بار به جای فاصله ویرایش از فاصله های زمانی به دست آمده در بخش قبلی استفاده میکنیم. از upgma استفاده میکنیم، به علت ریشه دار بودن درخت و همچنین evolutionary بودن داده ها.

حالا که درخت ساخته شد با به دست آوردن طول یالهای بین گره های آن میتوانیم فاصله ها را تخمین بزنیم. برای اینکار متد دریافت فاصله وجود دارد. کافی است از آن استفاده کنیم. سوال از ما خواسته که جد مشترک چه زمانی می زیسته است. از آنجا که ماربرگ و گونه های ابولا اول از هم جدا شده اند کافی است فاصله ماربرگ تا ریشه را به دست آوریم. ابتدا درخت را رسم کردم تا نام ریشه را بیابم و همچنین بررسی مسئله شهودی تر شود. فاصله ماربرگ تا ریشه حدود 194 سال بود پس حدودا 194 سال پیش این گونه ها از هم جدا شده اند.

- [1] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5003320/>
- [2] <http://jvi.asm.org/content/82/24/12569.full>
- [3] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1951390/>
- [4] <https://www.ncbi.nlm.nih.gov/pubmed/16698997>
- [5] https://en.wikipedia.org/wiki/Ebola_viral_protein_24
- [6] <http://pdb101.rcsb.org/motm/178>
- [7] https://en.wikipedia.org/wiki/Ebola_virus
- [8] <http://biopython.org/DIST/docs/api/Bio.Phylo.Consensus-module.html>
- [9] https://www.researchgate.net/post/What_is_the_difference_between_UPGMA_and_NEJ_method_while_constructing_a_tree_using_a_MEGA_4_software
- [10] <https://en.wikipedia.org/wiki/UPGMA>
- [11] https://en.wikipedia.org/wiki/Neighbor_joining