

پاسخ سوالات تیوری پروژه‌ی مبانی بیوانفورماتیک

میلاد آقاچوهری ، احسان سلطان‌آقایی

۸ بهمن ۱۳۹۶

۱ سوال اول

۱.۱ سوال اول بخش دوم یا اولین سوال تیوری

شناخت ژنوم ویروس ابولا:

۱. گلیکوپروتئین وظیفه متصل شدن به گیرنده‌های روی سطح سلول برای هدایت ژنوم ابولا به درون سلول را دارد و با زنجیره کربوهیدراتی پوشیده شده است که آن را از سیستم ایمنی مخفی نگه می‌دارد. گلیکوپروتئین از سطح ویروس گسترش می‌یابد و ویروس و سلول را به اندازه کافی به هم نزدیک می‌کند تا غشاها به یک دیگر متصل شوند. GP تنها پروتئین بیان شده توسط ابولا است که در غشای سطح سلول است. عوامل متعددی برای ویروس ابولا مورد نیاز است تا بتواند به سلول میزبان دسترسی پیدا کند. یکی از اجزای مورد نیاز برای ورود به سلول میزبان گلیکوپروتئین‌ها هستند که فیلوویروس‌ها را به سطح میزبان متصل می‌کنند، ویروس‌ها را به اندوسوم‌ها می‌فرستند و به هم‌جوشی بین ویروس و غشای اندوسوم کمک می‌کنند. به عبارتی این پروتئین مسئول متصل شدن ویروس و آلوده کردن سلول میزبان است و هم‌چنین در مختل کردن چسبندگی سلول‌ها موثر است. فلذا می‌تواند پروتئین منجر به بیماری قلمداد شود ولی یکی دیگر از پروتئین‌ها VP۲۴ نیز می‌تواند مقصود باشد که در ادامه آن را ذکر خواهیم کرد.

۲. پروتئین ماتریس ابولا (Protein Matrix Ebola) که با نام VP۴۰ شناخته می‌شود و در فرآیند جوانه زنی موثر است. (نقش این فرآیند در بخش بعدی سوال توضیح داده خواهد شد) نمونه‌های زیادی از این پروتئین مربوط به غشا است و به نظر می‌رسد که با هر دو غشا و نوکلئوکپسید در ارتباط است. هم‌چنین از آن جایی که ویروس‌ها درون ژنومشان تنها فضا برای کدگذاری تعداد کمی پروتئین دارند اکثر پروتئین‌ها چندین وظیفه برعهده دارند. برای مثال پروتئین ماتریس در ابولا ساختارهای کاملاً متفاوتی برای وظایف مختلفش به خود می‌گیرد: به عنوان هگزامر در ساختار ویریون (فرم ویروس در خارج از سلول میزبان) شرکت دارد، به عنوان اکتامر به RNA متصل می‌شود و رونویسی آن را تنظیم می‌کند و به عنوان دیامر در حمل پروتئین موثر است.

در مرکز ویروس مجموعه پروتئین‌های نوکلئوکپسید از ژنوم محافظت می‌کنند که با نام‌های VP۲۴ ، VP۳۰ ، VP۳۵ ، NP شناخته می‌شوند. اطلاعات ژنتیکی ویروس ابولا در نوکلئوکپسید است. زمانی که ابولا با موفقیت وارد سلول میزبان شد ترجمه و رونویسی آن را می‌ریابد.

۳. ویروس ابولا تب همراه با مرگ و میر به احتمال ۹۰ درصد را ایجاد می کند و در موارد مرگ و میر، ناشی از سرکوب اولیه سیستم ایمنی ذاتی میزبان است. یکی از پروتئین هایی که ادر این مورد تاثیر می گذارد VP۲۴ است که سیگنال اینترفرون را متوقف می کند. در نتیجه به عبارتی می توان این پروتئین را که سیستم ایمنی را مختل می کند باعث ایجاد بیماری دانست.

۴. رونویسی ویروس ابولا وابسته به VP۳۰ است و یک Factor Transcription از نوع Activator برای ویروس ابولا است و به وسیله تثبیت mRNA به رونویسی ویروس کمک می کند.

۵. پروتئین VP۳۵ همانطور که در بالا گفته شد چند منظوره است و به عنوان یکی از اجزای مجموعه RNA پلیمراز، عامل مونتاژ ویروس و مهارکننده تولید اینترفرون میزبان عمل می کند.

۶. نوکلئوپروتئین (NP) دور RNA می پیچید و یک مجموعه مارپیچی تشکیل می دهد. این پروتئین فراوان ترین پروتئین در سلول آلوده و در مجموعه نوکلئوکسپید است. هم چنین به شدت با ژنوم ویروس در ارتباط است و در رونویسی ویروس (Transcription)، تکثیر RNA، بسته بندی ژنوم و سر هم کردن نوکلئوکسپید قبل از کپسوله سازی غشا مورد نیاز و موثر است.

۷. پروتئین L که یک RNA پلیمراز وابسته به RNA است و بسیاری از نمونه های جدید ژنوم RNA را می سازد. (RNA Replication)

پروتئین های ماتریس و VP۲۴ و VP۳۰ و VP۳۵ و نوکلئوپروتئین و L پروتئین های ساختاری هستند و خود پروتئین GP نیز ساختاری است ولی پروتئین های GP، ssGP، delta غیرساختاری هستند که حاصل ویرایش و پردازش پس از ترجمه ژن GP و محصولات ژن هستند. SGP می تواند نقش مهمی در فرار از پاسخ ایمنی هومورال با جذب آنتی بادی های ایجاد شده داشته باشد.

۲.۱ سوال اول بخش سوم

نحوه وارد شدن ویروس ابولا به درون سلول و آلوده کردن آن:

ویروس ابولا از ماکروپینوسیتوز (macropinocytosis) به منظور دسترسی به سلول های میزبان هدف خود و انتقال خود به داخل سلول استفاده می کند. ماکروپینوسیتوز فرآیندی است که در آن بخش های رفع غشای پلازما خارج از سلول به صورت به خود برگشته ای قرار میگیرند که حفره هایی ایجاد می شود (ماکروپینوزوم) و ویروس به کمک گلیکوپروتئین به سطح غشای پلازما متصل می شود. شکل گیری ماکروپینوزوم به صورت خود به خودی به علت فعال شدن عوامل مختلف رشد یا به طور همزمان با مصرف مولکول های سلولی یا مایع خارج سلولی اتفاق می افتد. سپس غشا روی ویروس بسته می شود و ویروس همراه با مولکول های دیگر و مایعات خارج سلولی، تشکیل یک کیسه کوچک می دهند و به درون سلول منتقل می شوند.

NPC۱ پروتئین حمل کننده کلسترول اندوسوم و لیزوزوم است که درغشای لیزوزوم یافت می شود و برای آلوده کردن سلول میزبان مورد نیاز است. (در صورت فقدان این پروتئین ویروس نمی تواند سلول را آلوده کند ولی خود سلول نیز نمی تواند در فقدان این پروتئین به حیات خود ادامه دهد) هنگامی که ویروسی درون سلول قرار دارد، NPC۱ به همجوشی فیبر ویروس با اندوسوم و لیزوزوم کمک می کند و پس از آن ویروس می تواند کیسه (vesicle) خود را ترک کند و تکثیر شود. (علاوه بر NPC۱ مجتمع HOPS نیز به عنوان عامل مهمی در روند عفونت ویروسی شناخته شده است.)

پس از ترک کیسه ویروس دستگاه رونویسی سلول را می ربايد تا بتواند خود را تکثیر کند و در این صورت دیگر سلول نمی تواند اجزای مورد نیاز خود را از طریق رونویسی به دست آورد و منجر به مرگ

سلول و یا ناتوانی آن در عملکرد مناسب می‌شود. زمانی که تمام اجزای مورد نیاز برای تشکیل یک ویروس جدید فراهم شدند اجزای ویروس از سلول جوانه می‌زنند و به وسیله غشای سلول برای خود کپسول درست می‌کنند تا بتوانند سفر امنی به سلول مجاور داشته باشند و عملیات آلوده کردن سلول‌ها را ادامه دهند.

نحوه آلوده و عفونی کردن بدن و از بین بردن میزبان:
ویروس ابولا پس از وارد شدن به بدن سلول‌های خاصی را مورد هدف قرار می‌دهد از جمله سلول‌های کبدی، سلول‌های سیستم ایمنی بدن و سلول‌های اندوتلیال که داخل رگ‌های خونی قرار دارد. ابولا انواع سلول‌ها را می‌تواند آلوده کند و این قابلیت آن را یک عامل عفونی بسیار موفق می‌کند.

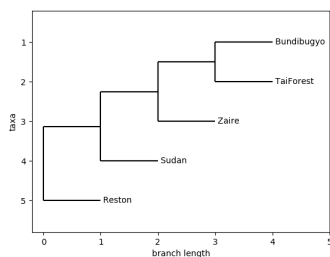
ویروس ابولا درون سلول به کمک گلیکوپروتئین چسبندگی سلول را مختل می‌کند و در نتیجه سلول‌ها در چسبیدن به هم و تشکیل ماتریس خارج سلولی (که در بافت‌های سالم کمک به کنار هم نگه داشته شدن سلول‌ها کنار یک دیگر می‌شود) دچار مشکل می‌شوند. از دست دادن چسبندگی سلول برای هر بافت جامدی زیان آور است. ویروس ابولا با آلوده کردن سلول‌های رگ‌های خونی باعث ایجاد سوراخ در آن و خونریزی داخلی می‌شود. هم چنین با هدف قرار دادن سلول‌های کبدی، توانایی بدن برای پاک کردن سموم از جریان خون به خطر می‌افتد و با آلوده سازی سیستم ایمنی بدن که سلول‌های آن در همه جا در بدن حرکت می‌کنند، ابولا می‌تواند به سرعت در حال افزایش سطح عفونت خود باشد. با گذشت زمان، عفونت سلول‌ها در سراسر بدن می‌تواند باعث نارسایی اندام شود و تب، خونریزی داخلی، اسهال و استفراغ می‌تواند باعث کاهش شدید الکترولیت‌ها، پلاسمای خون و مایع شود و در نهایت، نارسایی اندام و شوک ناشی از خونریزی داخلی منجر به مرگ می‌شود. و رفرنس‌ها:

[۱] [۲] [۳] [۴] [۵] [۶] [۷] [۸] [۹] [۱۰]

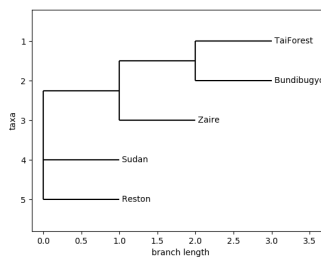
۳.۱ تشکیل درخت زندگی برای هر ژن

۱.۳.۱ مقایسه‌ی نتایج الگوریتم برای ۵ ژن

در ابتدا تمامی نتایج حاصله را روبروی هم مشاهده می‌کنیم:

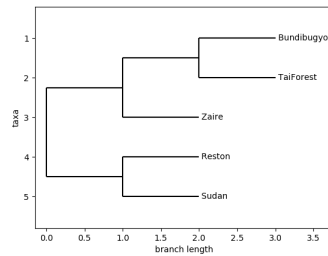


NP_UPGMA.png (ب)

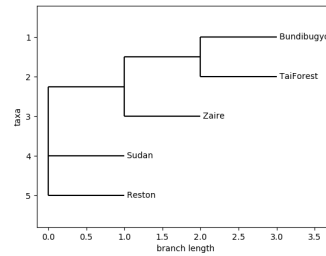


NP_NJ.png (آ)

برای این ژن درخت‌های حاصل یکی هستند.

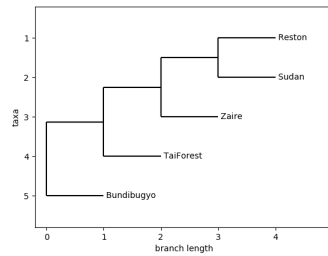


VP35 _ UPGMA.png (ب)

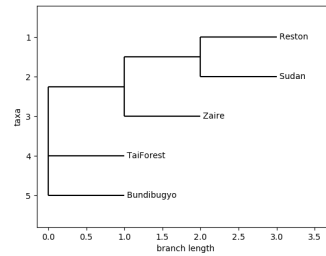


Vp35 _ NJ.png (آ)

برای این ژن درخت‌های حاصل یکی هستند (در واقع یک مفهوم را می‌رسانند).

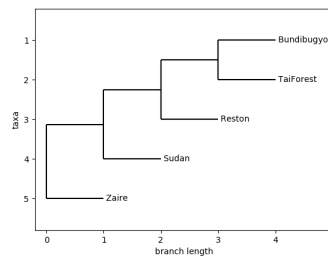


VP40 _ UPGMA.png (ب)

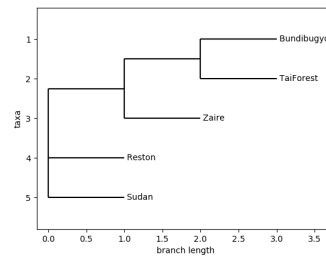


Vp40 _ NJ.png (آ)

برای این ژن درخت‌های حاصل تفاوت جزئی دارند، البته معنایشان تقریباً یکسان است، صرفاً UPGMA اطلاعات بیشتری داده است.

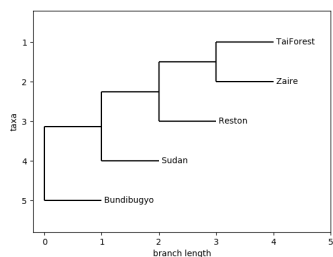


GP _ UPGMA.png (ب)

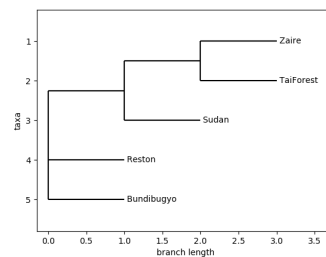


GP _ NJ.png (آ)

درخت‌های حاصل در این جا بسیار متفاوت هستند و تنها در مورد و TaiForest Bundibugyo یکسان هستند.

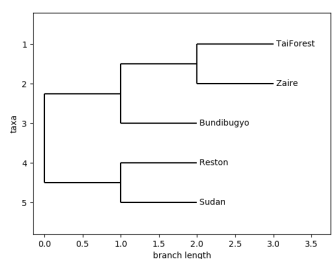


VP30 _ UPGMA.png (ب)

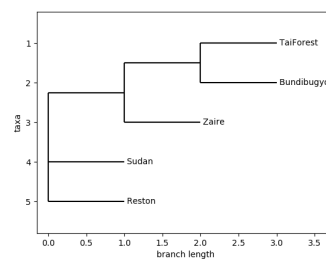


VP30 _ NJ.png (آ)

در این جا درخت‌های حاصل تفاوت زیادی دارند.

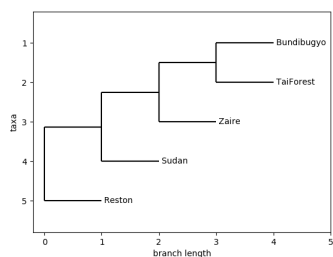


VP24 _ UPGMA.png (ب)

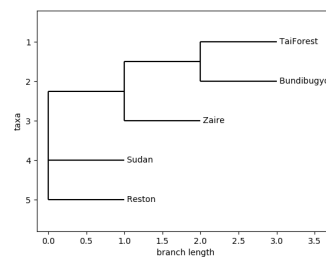


VP24 _ NJ.png (آ)

تنها در قسمت Zaire و Bundibugyo تفاوت در درخت‌ها هست.



L _ UPGMA.png (ب)



L _ NJ.png (آ)

در اینجا هر دو درخت یک مفهوم را می‌رسانند اگرچه ظاهراً تفاوت جزئی مشاهده می‌شود.

۴.۱ مقایسه‌ی نتایج

نمی‌توان گفت نتایج تفاوت‌های چشم‌گیر دارند. مخصوصاً اینکه در NJ در چهار ژن به یک درخت رسیده‌ایم که در سه عدد از UPGMA ها هم همین درخت حاصل شده‌است اما به هر حال نتایج در مورد بعضی ژن‌ها تفاوت دارند.

۱.۴.۱ بیان دلیل تفاوت‌ها

بستگی دارد منظور کدام نوع از تفاوت‌ها باشد. اگر منظور تفاوت‌های بین نتایج دو الگوریتم است، که این دو الگوریتم فرض‌های متفاوتی در مورد درخت اصلی دارند که به آن‌ها در قسمت‌های قبلی اشاره شده و منطقی است که نتایج متفاوت باشد. مثلاً UPGMA تاکید دارد که فاصله‌ی تمام برگ‌ها از درخت ریشه یکسان باشد.

اگر منظور تفاوت‌های بین درخت‌های حاصل برای ژن‌هاست، الزاماً جهش‌های ژنی روند درختی دقیق ایجاد ویروس‌های جدید را دنبال نمی‌کنند، مثلاً ممکن است یک نوع از ویروس در یک جایگاه مثلاً جایگاه پنجاهم که در ژن سوم است جهش کند اما انواع دیگر نکنند، در این‌جا روش‌های ما این ویروس را جدا از بقیه ترسیم می‌کند و حتی او را از تمام برادران و پدران‌ش در یک دسته‌ی جدا می‌گذارد. پس الزاماً بر حسب یک ژن نمی‌توان تصمیم گرفت که روند جهش‌ها و تولید گونه‌های جدید ویروس چه بوده است اما با ترکیب آن‌ها می‌توان دقت بیشتری در این امر به دست آورد.

۲.۴.۱ مقایسه‌ی دو الگوریتم NJ ، UPGMA

مزیت بزرگ الگوریتم UPGMA محاسبه‌ی درخت‌های ریشه‌دار است. این الگوریتم با فرض یکسان بودن فاصله‌ی تمام برگ‌ها از ریشه درخت را محاسبه می‌کند، این فرضی است که در بسیاری از موارد نادرست است. فرض دیگر این الگوریتم ثابت بودن نرخ تکامل است (که من متوجه نمی‌شوم این فرض در کجای این الگوریتم هست ناشی از عدم تسلط من به ریاضی پشت منطق این الگوریتم است). ثابت فرض کردن نرخ تکامل را نمی‌توان یک نقطه‌ی ضعف در نظر گرفت، زیرا طبق مطالبی که در ویکی‌پدیا خواندم اکثر کاربردهای این الگوریتم در مواقعی است که می‌خواهند بدون در نظر گرفتن نرخ تکاملی بین توالی‌ها و تنها با دقت به شباهت آن‌ها، آن‌ها را گروه‌بندی کند که در شاخه‌ای به نام Phenetics که هدف آن دسته‌بندی میان موجودات بر حسب مشاهدات کلی است، روشی مناسب است.

به نظر من حتی در همین پروژه دیدیم که این الگوریتم در زمینه‌ی دسته‌بندی بر حسب ژن‌ها چندان هم خوب عمل نکرد و درخت‌های نسبتاً متفاوتی را برای هر ژن نتیجه می‌داد البته اجماع تمام آن‌ها به درختی صحیح منتهی شد که نتیجه منطبق با نتیجه‌ی ترکیب‌شده‌ی NJ بود و البته طبق بررسی‌های یواشکی: ما در اینترنت با نتیجه‌ی درست هم منطبق است.

نکته‌ی منفی الگوریتم NJ تولید درخت‌های بدون ریشه است، اما نکته‌ی مثبت آن عدم فرض یکسان بودن نرخ تکامل است (البته من متوجه نمی‌شوم که این فرض در کجای الگوریتم هست که ناشی از عدم تسلط من به ریاضی پشت منطق این الگوریتم است). همین فرض مثبت این الگوریتم آن را برای بررسی داده‌های توالی بسیار مناسب کرده است. نقطه‌ی مثبت دیگر این الگوریتم سریع بودن آن در قیاس با روش‌های دیگر است که آن را برای بررسی داده‌های در مقیاس بزرگ و استفاده

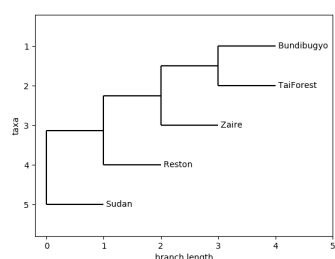
از روش‌های آماری که بارها از داده نمونه‌برداری می‌کنند (مانند bootstrap مناسب کرده‌است. این روش در صورتی که ماتریس فاصله‌ها درست باشد، پاسخ صحیح را تولید می‌کند، اما حتی اگر ماتریس فاصله‌ها درست نباشد و حدودا درست باشد، این الگوریتم به طرز معجزه‌آسایی به احتمال خوبی باز هم درست را تولید می‌کند (شهودی در مورد این که این اتفاق چرا می‌افتد پیدا نکردم، حتی ویکیپدیای انگلیسی از لفظ *but neighbor joining often constructs the correct tree topology anyway*) استفاده کرده‌است و به یک مقاله (اینجا) لینک داده است که فرصت نشد مطالعه کنم. این الگوریتم به بعضی از شاخه‌ها وزن منفی می‌دهد و این نامطلوب است (این در ویکیپدیا گفته شده، اما ما در این تمرین مشاهده نکردیم).

۵.۱ ترکیب درخت‌ها و ارائه‌ی درخت نهایی

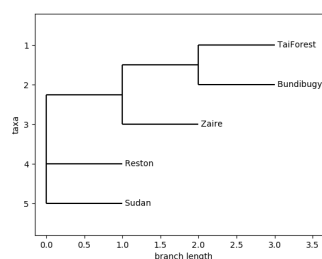
ما در اینجا از پکیج ape متد consensus استفاده کردیم. در این روش ما از احتمال برابر با ۰.۵ استفاده کردیم به این صورت است که یال‌هایی را که در ۵۰ درصد یا بیش‌تر از درخت‌ها آمده است با همدیگر ترکیب می‌کند و یک majority rule consensus tree حاصل می‌کند. این روش مناسبی به نظر می‌رسد چرا که به هر یال به اندازه‌ای که تکرار شده وزن و اهمیت می‌دهد و در واقع روش مبتنی بر اهمیت دادن به یال‌های آمده در روش‌هاست. (البته واژه‌ی یال در اینجا بار معنایی مناسب و دقیق را ندارد و منظور ما در اینجا از یال clade است. گویا این الگوریتم یال‌ها را بر حسب تعداد تکرار مرتب می‌کند و سعی می‌کند درختی تولید کند که با اکثر آن‌ها که بالای ۵۰ هستند بخواند.

۶.۱ مقایسه‌ی درخت ترکیبی و درخت حاصل از همترازی سراسری

ابتدا درخت حاصل از همترازی سراسری را مشاهده می‌کنیم.

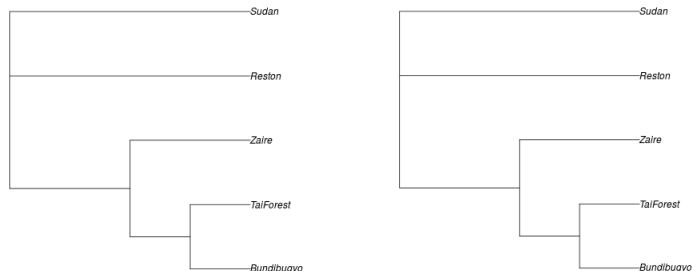


genome_UPGMA.png (ب)



genome_NJ.png (آ)

و سپس درخت‌های ترکیبی را مشاهده می‌کنیم:



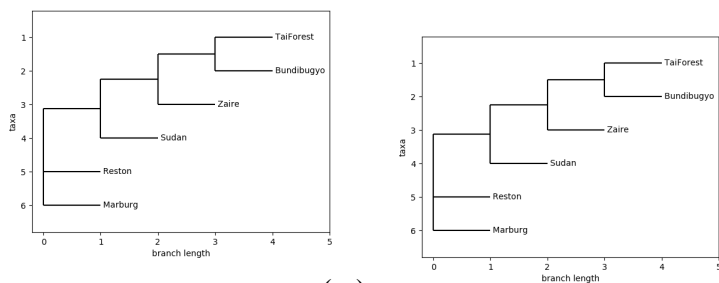
mix_UPGMA.png (ب)

mix_NJ.png (آ)

همانطور که مشاهده می‌شود تمام این نتایج یکسان هستند و چندان مقایسه‌ای نداریم:).

۷.۱ تعیین نقطه‌ی شروع

روش ما در این بخش همان روش همترازی سراسری است. کافی است زن تمام ۵ گونه‌ی ویروس ابولا را با ویروس ماربرگ همترازی سراسری کنیم و از فاصله‌ی ویرایش آن استفاده کرده و درخت را تشکیل دهیم، خواهیم داشت:



(ب)

all_and_marburg_UPGMA.png all_and_marburg_NJ.png (آ)

پس نتیجه می‌گیریم که در ابتدا (یا طبق عبارت صورت سوال نقطه‌ی شروع) ویروسی بوده که به سه گونه‌ی marburg و Restone و گونه‌ی پدر تمام ویروس‌ها Sudan و Zaire و Taiforest و Bundibugyo جهش پیدا کرده. آن گونه‌ی پدر آخری به Sudan و گونه‌ی پدر گونه‌های Zaire و Bundibugyo جهش پیدا کرده. آن گونه‌ی پدر گونه‌های Zaire و Bundibugyo و Bundibugyo و Bundibugyo جهش پیدا کرده. در نهایت و Zaire به TaiForest و گونه‌ی پدر Bundibugyo و TaiForest جهش پیدا کرده. این گونه‌ی پدر آخری به TaiForest و Bundybugyo جهش پیدا کرده.

۲ بخش چهارم

۱.۲ چه زمانی از هم جدا شده‌اند؟

در این‌جا ما از مدل Jukes_Cantor استفاده کرده‌ایم. علت استفاده از این روش این است که تنها روشی است که بلد بودیم و البته ساده است و قدرت محاسباتی زیادی به ما می‌دهد (با فرض بازگشت‌پذیر بودن مارکوف‌ها در زمان و غیره). حقیقت این است که چون با روش‌های دیگر آشنا نیستیم نمیتوانم بگویم که انتخابی داشته‌ام. اما این روش در کل ساده و منطقی به نظر می‌رسد. مثلاً دارد نرخ خطای دی ان ای پلیمراز را مدل می‌کند و منطقی است که این نرخ خطا چندان تغییر نکند. مخصوصاً این که ویروس و گونه‌هایش در چیزی حدود چند صد سال اخیر به وجود آمده‌اند. در این روش یک ماتریس احتمال به صورت

$$P(t) = \begin{matrix} & \begin{matrix} A & T & G & C \end{matrix} \\ \begin{pmatrix} 1-3f(t) & f(t) & f(t) & f(t) \\ f(t) & 1-3f(t) & f(t) & f(t) \\ f(t) & f(t) & 1-3f(t) & f(t) \\ f(t) & f(t) & f(t) & 1-3f(t) \end{pmatrix} & \begin{matrix} A \\ T \\ G \\ C \end{matrix} \end{matrix}$$

شکل ۱۱: ماتریس احتمال مارکوف

داریم که احتمال تبدیل یک نوکلئوتید به دیگر نوکلئوتیدها را مدل می‌کند. با این مدل‌سازی و استفاده از تقریب‌هایی که در کلاس استاد مطهری به آن‌ها اشاره شده است می‌توان به این دو نتیجه رسید:

$$p_{ij}(t) = f(t) = \frac{1}{4} - \frac{e^{-4\alpha t}}{4}$$

$$p_{ii}(t) = 1 - 3f(t) = \frac{1}{4} + \frac{3}{4}e^{-4\alpha t}$$

شکل ۱۲: نتایج حاصله که در آن‌ها آلفا برابر یک سوم است
 $\alpha = \frac{1}{3}$

پس با توجه به تعداد تعداد جایگاه‌هایی که دو ژن در آن‌ها به هم شبیه هستند (آن را b بنامید) و تعداد جایگاه‌هایی که در دو ژن متفاوت هستند (آن‌ها را a بنامید). پس احتمال رسیدن دو ویروس به هم را می‌توان در صورت دانستن فاصله‌ی بین آن دو با فرمول:

$$\left(\frac{1}{4}(1 - e^{-\alpha t})\right)^a \times \left(\frac{1}{4}(1 + 3e^{-\alpha t})\right)^b$$

حساب می‌شود. حال باید t ای را انتخاب کنیم که احتمال بالا را بیشینه کند. با گرفتن لوگاریتم برای ساده‌تر شدن و انجام محاسبات خواهیم داشت:

$$t = -\frac{3}{4} \ln\left(1 - \frac{4}{3}p\right)$$

شکل ۱۳: نتیجه‌ی نهایی
 که در آن $p = \frac{a}{a+b}$

البته باید دقت کنیم که در این تصاویر در واقع $t = \lambda \times time$ است که در آن λ برابر با نرخ تحول است. پس باید t حاصل در فرمول بالا را بر λ که اطلاعات داده شده در صورت پروژه برابر با $10^{-3} \times 1/9$ در نظر گرفته شده، تقسیم کنیم. در این قسمت البته از این منبع استفاده کردیم. حال ما از distance edit به عنوان یک تقریب برای a و از طول یکی از ژنوم‌ها (با توجه به نزدیک بودن طول تمامی ژنوم‌ها) به عنوان یک تقریب برای $a + b$ یا در واقع طول ژنوم استفاده کردیم پس داریم:

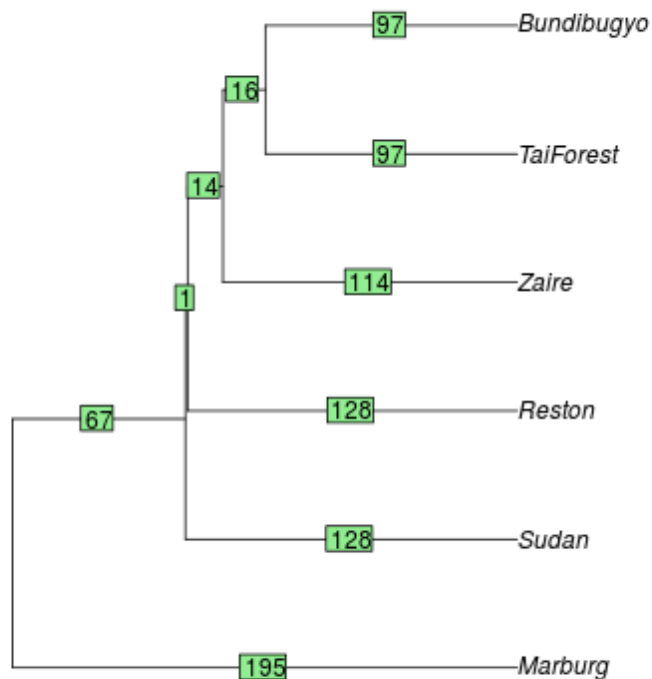
$$p \approx \frac{\text{edit_distance}}{\text{genome_length}}$$

و سپس یک ماتریس محاسبه کردیم که شامل این فاصله‌های زمانی بود

	Reston	Sudan	TaiForest	Zaire	Bundibugyo	Marburg
Reston	0	258	257	253	258	393
Sudan	258	0	259	255	260	398
TaiForest	257	258	0	227	195	400
Zaire	252	254	227	0	229	395
Bundibugyo	257	258	195	229	0	401
Marburg	385	390	394	389	395	0

شکل ۱۴: فاصله‌های زمانی بین گونه‌ها

و آن را به الگوریتم UPGMA دادیم و طول یال‌ها را برای ما حساب کرد و این نتیجه خروجی شد:



شکل ۱۵: درخت زمانی حاصله

که در آن این اعداد نوشته شده در جداول سبز در واقع سال‌های تخمین زده شده هستند، مثلاً طبق این جدول marburg ۱۹۵ سال با پدر این ویروس‌ها فاصله دارد و sudan ۱۲۸ سال با پدر شاخه‌ی ابولا ویروس فاصله دارد و پدر ابولا ویروس‌ها ۶۷ سال با پدر مشترکشان با ماربرگ ویروس فاصله دارد. پس در کل حدس این است که جد مشترک ابولا ویروس‌ها ۱۲۸ سال قبل میزیسته و پدر مشترک تمام آن‌ها با ماربرگ در ۱۹۵ سال پیش میزیسته (خصوصیت جالب UPGMA که فرض می‌کند تمام برگ‌ها در یک زمان هستند را نیز می‌توانید در تصویر مشاهده کنید).