

January 28, 2018

گزارش پروژه

۱۸.۲

پروتئین‌های NP، VP30، VP35 و L ساختاری بوده و سایر پروتئین‌ها غیرساختاری هستند.

پروتئین L در ویروس ابولا که یک RNA-Polymerase است وظیفه‌ی کپی کردن ژنوم ویروس را بر عهده

دارد.

در چندین پژوهش که درباره‌ی نحوه‌ی ورود این ویروس به سلول‌ها انجام شده است، به نظر می‌رسد که

پروتئین NP باعث ورود ویروس به سلول‌ها می‌شود.

پروتئین GP باعث بیماری می‌شود.

۱۸.۳

ویروس ابولا به اشکال مختلفی به سلول‌های ایمنی بون آسیب می‌زند. برای مثال پروتئین VP24 این

ویروس عامل بازدارنده‌ای برای جابه‌جایی مولکول interferon در سطح سلول‌های ایمنی بدن می‌باشد. این

مولکول‌ها برای مقابله سلول‌ها با رشد ویروس‌ها حیاتی هستند.²

¹ source: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2829775/>

² [http://www.cell.com/cell-host-microbe/abstract/S1931-3128\(14\)00263-7](http://www.cell.com/cell-host-microbe/abstract/S1931-3128(14)00263-7)

علاوه بر این ابولا باعث خونریزی داخلی و بعضاً خارجی در بیماران می‌شود. با آلوده ساختن سلول‌های ایمنی ماکروفاژ بدن، این سلول‌ها به تولید نیتریک اسید و پروتئین‌های سیگنال دهنده‌ای می‌پردازند که به رگ‌های خونی آسیب زده و باعث خونریزی می‌شوند.³

۲.۲

برای این کار از پکیج Biostrings و تابع pairwiseAlignment استفاده نمودیم. جزئیات آرگومان‌هایی که در این بخش به تابع داده شد به شرح زیر است.

به عنوان ماتریس substitution امتیاز یک برای match و -1 برای mismatch در نظر گرفته شد. همچنین هزینه‌ی ایجاد و افزودن gap نیز برابر یک بود. هم‌ترازی نیز از نوع local-global قرار داده شد تا به این نکته که یک ژن کوتاه را داریم با یک ژنوم بلند هم‌تراز می‌کنیم توسط الگوریتم مورد توجه قرار گیرد.

به جز ژن اول در ویروس ماربورگ، مکان ژن‌های فرضی در سایر انواع ویروس ابولا از دویست باز عقب‌تر از مکانشان در ماربورگ قرار گرفتند. همچنین طول هر ژن نیز سه برابر ژن در ماربورگ فرض شد.

پس از اجرای الگوریتم هم‌ترازی برای هر ژن از هر نوع ویروس ابولا، از مکان شروع و پایان هم‌ترازی در ژنوم برای استخراج هر ژن استفاده شد.

۲.۳

پروژه‌ی بیوانفورماتیک

در این جا باز مانند قسمت قبل عمل کردیم. با این تفاوت که همترازی global بوده و اکنون امتیاز

match صفر و سایر هزینه‌ها همه منفی یک قرار داده شدند. با این تغییرات منفی امتیاز همترازی‌ها برابر

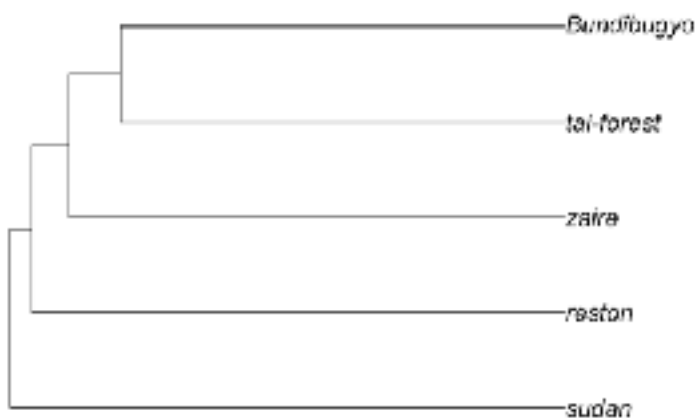
فاصله‌ی ویرایشی بین ژن‌ها می‌شود. پس از این که تمام هفت ماتریس فاصله محاسبه شدند همگی آنها با دستور

write.table با فرمت csv ذخیره شدند.

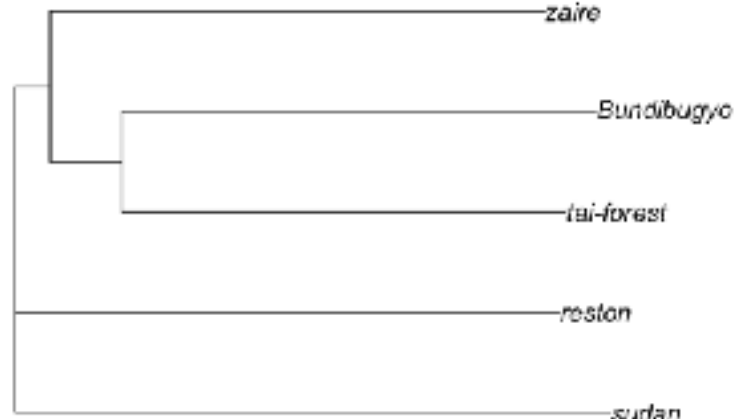
۳

۳.۱

UPGMA_geneNP



NJ_geneNP.png



اولا برای ساخت درخت از توابع upgma و nj پکیج phangorn بهره بردم. این دو تابع یک ماتریس

فاصله به عنوان ورودی گرفته و درخت مربوطه را می‌سازند.

کلیت دو درخت برای هر الگوریتم در ژن‌های مختلف تقریباً یکسان است. یعنی شکل درخت upgm‌های

در ژن‌های مختلف تفاوت‌های جزئی با یکدیگر دارند. تفاوت اصلی‌ای که بین درخت‌های upgma و nj مشاهده

می‌شود در این است که upgma اجداد مشترک بیشتری را بین انواع ویروس در نظر می‌گیرد. یعنی راس‌های

پروژه‌ی بیوانفورماتیک

میانی بیشتری در درخت‌های upgma مشاهده می‌کنیم. این خود می‌تواند یک دلیل برای بهتر بودن upgma

نسبت به nj در مطالعات مربوط به تاریخچه‌ی ابولا باشد.

یک دلیل تفاوت درخت‌های تولید شده توسط دو الگوریتم این است که به طور کلی upgma تولید

درخت‌های ریشه‌دار و NJ درخت‌های بدون ریشه تولید می‌کند.

یک اشکال upgma این است که فرض می‌کند تمام شاخه‌های این درخت با یک روند تکامل پیدا می‌کنند.

در حالی که NJ این فرض را ندارد.

اگر ریشه‌دار بودن درخت برای ما اهمیت داشته باشد یا فرض اینکه تمام شاخه‌های درخت زندگی با یک

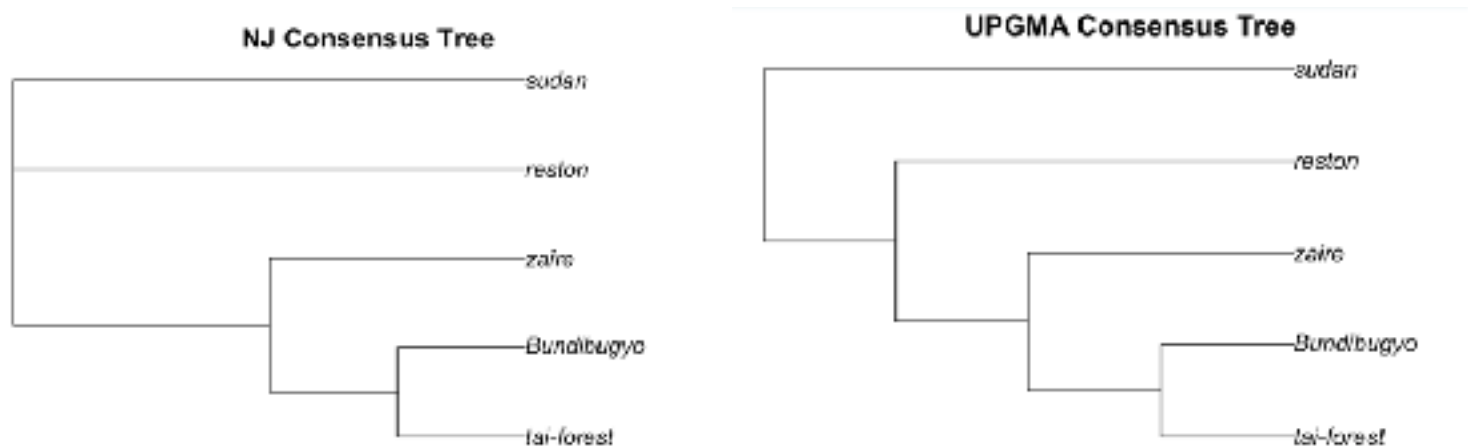
روند و سرعت به تکامل رسیده‌اند برای مساله‌ی پیش رویمان قابل قبول باشد، الگوریتم UPGMA می‌تواند انتخاب

مناسبی باشد. اما در غیر این صورت الگوریتم NJ با توجه به تضمین درستی و سرعت محاسبه الگوریتم بهتری

است.

از آنجایی که در این پروژه تمرکز روی یافتن منشا و اجداد و فاصله‌ی زمانی انواع ویروس ابولا است من

در ادامه پروژه از الگوریتم UPGMA استفاده می‌کنم.



۳.۲

برای ترکیب درخت‌ها از روش consensus که الگوریتم آن در پکیج phangorn نیز پیاده‌سازی شده استفاده می‌کنیم. در این روش یک سری درخت به عنوان ورودی به الگوریتم داده می‌شود و نوعی از رای‌گیری برای ترکیب درخت‌ها فرض می‌شود. این رای‌گیری با در توجه یک مقدار threshold بین صفر و یک انجام می‌شود که در کد به آن P گفته می‌شود. P بیان می‌کند که هر شاخه باید در در چه درصدی از درخت‌های داده شده به الگوریتم مشاهده شود تا در درخت نهایی و ترکیبی نیز حضور داشته باشد. با توجه به مقداری که به P می‌دهیم الگوریتم اسم‌های مختلفی می‌گیرد. برای مثال اگر P برابر ۰.۵ باشد یعنی مقداری که در کد نیز از آن استفاده کرده‌ایم، اصطلاحاً Majority Rule Consensus Tree را ساخته‌ایم که فقط شاخه‌هایی را شامل می‌شود که در بیش از نیمی از درخت‌های ورودی وجود داشته باشد.

Global Genome Alignment UPGMA Tree



در این بخش تمام پنج ژنومی اصلی

ویروس ابولا را با یکدیگر همترازی سراسری

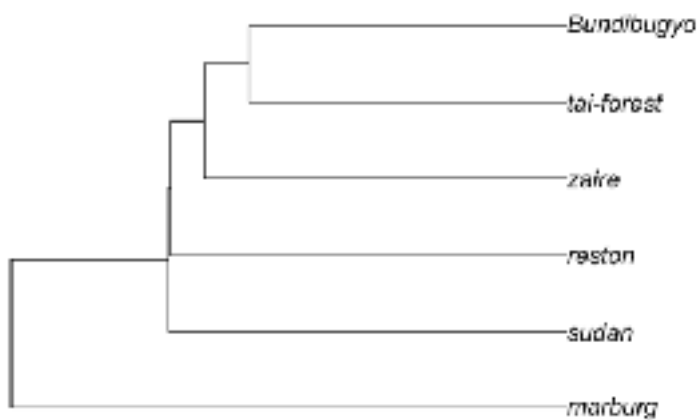
می‌کنیم و ماتریس فاصله‌ی آن را محاسبه کرده و

سپس با upgma درخت فیلورژنی آن را رسم

می‌کنیم. مشاهده می‌کنیم که درخت به دست آمده

با دو درخت consensus تقریباً یکسان است.

Ebola Strains UPGMA Tree



در این بخش نیز مانند دو قسمت عمل کرده و

درخت را با UPGMA می‌کشیم.

درخت به دست آمده با دانسته‌های ما

همخوانی دارد. ماربورگ نزدیکترین موجود به

ابولاویروس‌هاست. برای همین در درخت نیز به جد مشترک پنج نوع اصلی ابولا متصل است.

در این قسمت از تابع dist.dna استفاده شد. این تابع یک سری رشته‌ی DNA با طول یکسان می‌گیرد و

فاصله‌ی هر دو رشته را با توجه به مدل‌های تکاملی مختلفی می‌تواند حساب کند.

من از مدل Jukes-Cantor برای این تابع استفاده کردم.

پروژهی بیوانفورماتیک

dist.dna	reston	zair	sudan	tai forest	Bundibugyo	marburg
reston	0.000000	2.358350	3.210921	2.742155	3.175120	3.229482
zair	2.358350	0.000000	2.927865	3.028672	3.032277	3.201810
sudan	3.210921	2.927865	0.000000	3.145135	3.183912	2.772217
tai forest	2.742155	3.028672	3.145135	0.000000	2.811570	2.870376
Bundibugyo	3.175120	3.032277	3.183912	2.811570	0.000000	3.277997
marburg	3.229482	3.201810	2.772217	2.870376	3.277997	0.000000

از آن جا که اعداد به دست آمده چندان منطقی به نظر نمی‌رسیدند و اینکه مجبور شده بودم با پر کردن

رشته‌ها آن‌ها را هم طول کنم تا تابع dist.dna آنها را قبول کند، به دنبال تابع دیگری گشتم و به تابع

cophenetic.phylo از پکیج ape رسیدم که یک درخت(در این جا درخت به دست آمده از ۳.۴) را میگیرد و یک

ماتریس فاصله پس می‌دهد. جدول را در پایین مشاهده می‌کنید.

cophenetic.phylo	reston	zair	sudan	tai forest	Bundibugyo	marburg
reston	0.000	7462.667	7508.25	7462.667	7462.667	10453.4
zair	7462.667	0.000	7508.25	6817.500	6817.500	10453.4
sudan	7508.250	7508.250	0.00	7508.250	7508.250	10453.4
tai forest	7462.667	6817.500	7508.25	0.000	5971.000	10453.4
Bundibugyo	7462.667	6817.500	7508.25	5971.000	0.000	10453.4
marburg	10453.40	10453.40	10453.40	10453.400	10453.40	0.000000

پروژهی بیوانفورماتیک

در صورت پروژه گفته شد که نرخ جهش در سال برای این ویروس‌ها $1.9e-3$ تخمین زده می‌شود. اگر

فرض کنیم اعداد جدول بالا تعداد جهش یا تفاوت را بین هر دو ژنوم نشان بدهد، با ضرب تک تک درایه‌های جدول

در معکوس نرخ جهش می‌توانیم فاصله هر دو سرده ویروس ابولا را برحسب زمان بیابیم. باز مقادیر به دست آمده

در جدول زیر قابل مشاهده است. از سایر درخت‌هایی که در بخش‌های قبل ساختیم مانند درخت‌های

Consensus نیز برای تولید این ماتریس‌ها استفاده شد ولی مقادیر به دست آمده همگی نزدیک به صفر بودند.

diff in years	reston	zair	sudan	tai forest	Bundibugyo	marburg
reston	0.000	3927719	3951711	3927719	3927719	5501789
zair	3927719	0.000	3951711	3588158	3588158	5501789
sudan	3951711	3951711	0.00	3951711	3951711	5501789
tai forest	3927719	3588158	3951711	0.000	3142632	5501789
Bundibugyo	3927719	3588158	3951711	3142632	0.000	5501789
marburg	5501789	5501789	5501789	5501789	5501789	0.000000