



درخت زندگی ابولاویروس

مقدمه‌ای بر بیوانفورماتیک

پاییز ۱۳۹۶

دانشگاه صنعتی شریف

طراحی و پیاده‌سازی:

الیاس حیدری - رسا قوامی

در مورد ابولاویروس:

ابولاویروس چیست؟

ابولاویروس یک نوع ویروس^۱ است که تا کنون پنج گونه^۲ آن شناخته شده است. [۱] ژنوم این ویروس شامل ۷ ژن است که یکی از این ژن‌ها به نام 24VP پس از سرکوب کردن مسیرهای^۳ انتقال سیگنال^۴ در سلول در نهایت منجر به سرکوب سیستم ایمنی^۵ شود و به این صورت مرگبار عمل کند. [۲] از ۵ گونه شناخته شده از این ویروس ۴ گونه آن سبب بیماری ویروس ابولا^۶ در آدمی می‌شوند. [۳] این بیماری به گونه‌ای است که بیمار دچار تبی خواهد شد که بسیار کشنده است. [۴] بین سالهای ۲۰۱۳ تا ۲۰۱۶ میلادی به دلیل شیوع بیش از حد^۷، ۱۱۳۱۰ نفر که بیشتر آنها ساکنین آفریقا بودند، توسط بیماری کشته شدند. [۵] ابولا ویروس اولین بار در شیوع بیش از حد در سودان جنوبی در سال ۱۹۷۶ یافت شدند. در همان سال گونه دیگری از آن ویروس در کونگو^۸ یافت شد. [۶]

اپیدمیولوژی (شیوع شناسی)

در آخرین شیوع (که کشنده‌ترین شیوع هم بوده) یک نوزاد یک ساله پسر که در سال ۲۰۱۳ در یک روستا در گینه به سبب ابتلا به بیماری کشته‌شد به عنوان اولین نقطه شیوع شناخته می‌شود و به نظر می‌رسد این نوزاد بیماری خود را به خانواده خود انتقال داده و خانواده او بیماری را به روستاهای اطراف و به همین صورت گسترش بیماری ادامه یافته‌است. [۷] محققان بر این باورند که انتقال ویروس از خفاش‌ها به انسان دلیل اصلی گسترش این بیماری است. [۸] انتقال بیماری از حیوان به انسان و سپس از انسان به انسان تنها از طریق تماس مستقیم مایعات بدنی و خونی صورت می‌گیرد. با توجه به باور محققان مبنی بر منشاگرفتن ویروس از خفاش‌ها شیوع بیماری اینگونه توجیه می‌شود که ابتدا خفاش بیمار مقداری از غذای خورده شده اش را بر زمین و جایی که دیگر

^۱ Virological taxon

^۲ Species

^۳ Pathway

^۴ Signal transduction

^۵ Immunosuppression

^۶ Ebola virus disease (EVD)

^۷ Outbreak

^۸ Zaire

پستانداران زندگی می‌کنند می‌اندازد و پستانداران دیگر از غذای باقیمانده تغذیه کرده و آلوده می‌شوند.[۹]
(شکل ۱)

ساختار و ژنوم

ابولاویروس دارای یک ژنوم تک رشته‌ای RNA است که به صورت یک ویریون^۹ (فرم بیماری‌زا و کامل ویروس خارج از سلول میزبان) وجود دارد. همانطور که در شکل ۲ دیده می‌شود ویریون به شکل استوانه‌ای^{۱۰}، لوله‌ای^{۱۱} است.[۱۱] در شکل ۳، ژنوم ابولاویروس نشان داده شده است. طول ژنوم بین ۱۸۹۵۹ و ۱۸۹۶۱ نوکلئوتید است.[۱۲] نکته جالب در ژنوم این موجود عدم وجود non-coding RNA است (زیر ۲۰۰ نوکلئوتید غیرکدینگ) و ژنوم شامل ژنهای ۶ پروتئین ساختاری و یک پروتئین غیرساختاری است. همانطور که در شکل ۳ مشاهده می‌کنید این ۷ ژن پشت سر هم قرار گرفته‌اند:

- ساختار کلی رشته : 3'-NP-VP35-VP40-GP-VP30-VP24-L
- ابتدا و انتهای هر ژن شامل این زیر رشته است : 3'-UAAUU
- بین NP و 53VP یک زیررشته ۳ نوکلئوتیدی (GAU)
- بین 40VP و GP یک زیررشته ۳ نوکلئوتیدی (AUG)
- بین 30VP و 24VP یک توالی ۱۴۲ تایی از نوکلئوتیدها

طبق مروری^{۱۲} بر اپیدمی بین سال‌های ۲۰۱۳ و ۲۰۱۶ تخمین زده می‌شود که هر جایگاه از ژنوم ابولاویروس در سال با نرخ 10×10^9 جهش می‌یابد.[۱۳]

انواع گونه‌های شناخته شده و ویروس ماربرگ

همانطور که گفته شد تا به حال ۵ گونه از ابولاویروس شناخته شده‌اند

- Zaire
- Sudan
- Reston
- Tai Forest
- Bundibugyo

^۹ Virion

^{۱۰} Cylindrical

^{۱۱} Tubular

^{۱۲} Review paper

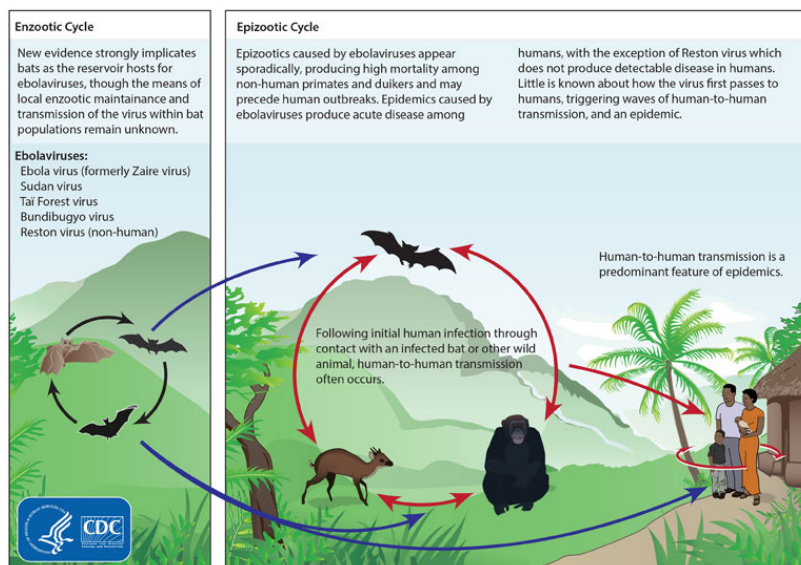
هر یک از این‌ها در مکان خاصی یافت شده اند و شیوه انتقال مرسوم در آنها متفاوت است، علاوه بر این با توجه به تفاوت در ژنوم، میزان کشندگی و عملکرد هر یک از این ویروس‌ها برای انسان متفاوت است. ابولاویروس در واقع یک سرده¹³ که خود از خانواده Filoviridae¹⁴ است. در این خانواده یک سرده دیگر هم موجود است و آن سرده ماربرگ ویروس¹⁵ است که شباهت زیادی به سرده ابولاویروس دارد. هر دو این سرده‌ها باعث ایجاد تب همراه با خونریزی¹⁶ در پستانداران و به خصوص انسان می‌شوند که علایم بسیار نزدیکی ایجاد می‌کنند. هر دو اولین بار در آفریقا مشاهده شده‌اند و بارها باعث به راه افتادن اپیدمی‌های گسترده در آن منطقه شده‌اند. [۱۴] علاوه بر این‌ها هر دو در ساختار ویرون شبیه به هم دارند که حاصل ژنوم مشابه آنها و پروتئین‌های ساختاری مشابه است. [۱۵]

¹³ genus

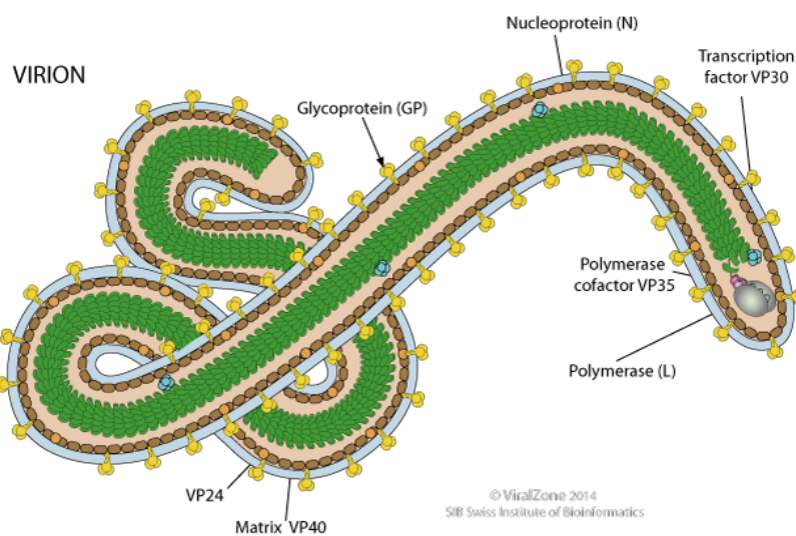
¹⁴ family

¹⁵ Marburgvirus

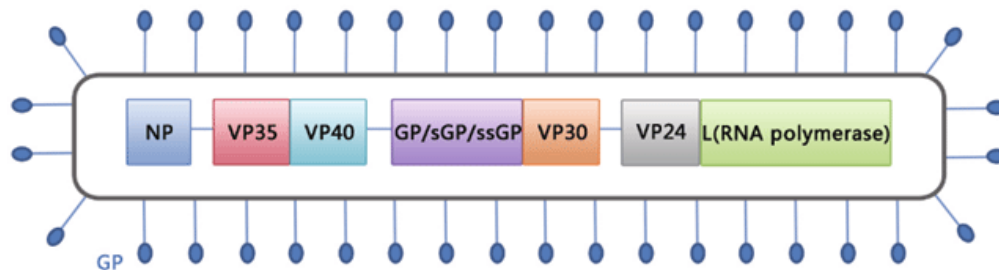
¹⁶ hemorrhagic fever



شکل ۱: اکولوژی ابولا ویروس [۱۰]



شکل ۲: ساختار ابولا ویروس [۱۶]



شکل ۳: ژنوم ابولا ویروس [۱۷]

۱. مقدمه

در این پروژه قصد داریم تا با استفاده از داده‌های ژنوم ویروس ابولا¹⁷ و انجام هم‌ترازی¹⁸ پس از استخراج ژن‌های این ویروس، درخت زندگی را برای آن به دست آوریم. سپس با در دست داشتن اطلاعات کافی می‌خواهیم نقطه آغازین این سرده را حدس بزنیم. در آخر می‌خواهیم با در دست داشتن نرخ جهش در ژنوم این ویروس ضمن تخمین زدن فاصله زمانی جدا شدن گونه‌های مختلف پیش‌بینی کنیم احتمالاً چه زمانی گونه‌ی جدیدی حاصل خواهد شد.

۱.۲. شناخت ژنوم ابولاویروس (۲٪)

در مورد کارکرد هر کدام از پروتئین‌های ابولا توضیح دهید، بطور خاص بگویید

- کدام پروتئین‌ها ساختاری و کدام غیرساختاری هستند؟
- و کدام پروتئین باعث ایجاد بیماری می‌شود؟

۱.۳. ایجاد بیماری توسط ابولا ویروس (۳٪)

تحقیق کنید که ابولاویروس دقیقاً به چه صورت باعث ایجاد بیماری می‌شود و چه مسیرهایی را تحت تاثیر قرار می‌دهد؟

۲. توالی

۲.۱. داده‌های ژنوم

داده‌های ژنوم گونه‌های مختلف ویروس ابولا را از [اینجا](#) بارگیری کنید.

¹⁷ Ebolavirus

¹⁸ alignment

۲.۲. استخراج ژن‌ها (۱۰٪)

برای استخراج ژن‌های هر گونه از ابولاویروس قصد داریم از ژن‌های استخراج شده‌ی نزدیک‌ترین سرده‌ی شناخته شده به ابولاویروس، یعنی ماربرگ، استفاده کنیم. برای این کار لازم است هر کدام از ژن‌های ماربرگ به طور جداگانه با کل ژنوم هر کدام از گونه‌های ابولا هم‌تراز شوند تا قسمتی از ژنوم که بیشترین شباهت را به هر کدام از ژن‌ها دارد به عنوان همان ژن در آن گونه فرض شود. با توجه به این که طول یک ژن با طول ژنوم قابل مقایسه نیست، لازم است الگوریتم هم‌ترازی سراسری طوری تغییر داده شود که هزینه‌ی شکاف‌های ابتدایی و انتهایی فقط برای رشته‌ی مربوط به ژن 0 در نظر گرفته شود (در واقع الگوریتمی مابین هم‌ترازی سراسری و محلی). برای این کار به نکات زیر توجه کنید:

- با توجه به جایگاه ژن‌ها در ژنوم ابولا و طول هر کدام، لازم نیست بین هر کدام از ژن‌ها را با کل ژنوم هم‌ترازی انجام دهید و کافی است تنها بخشی از ژنوم را برای این کار در نظر بگیرید.
- با توجه به اینکه ژن‌های استخراج شده دقیقاً ژن‌های ابولاویروس نیستند و ممکن است جایگاهشان در حد چند صد ریبونوکلیک اسید جابه‌جا تشخیص داده شده باشد لازم است از فرض دقیقاً پشت سر هم بودن ژن‌ها صرف نظر کنید و هر بار برای جستجوی ژن بعدی برای اطمینان از یک بازه‌ای قبل از انتهای ژن قبلی یافت شده آغاز کنید نه دقیقاً انتهای آن.

۲.۳. هم‌ترازی ژن‌ها و به دست آوردن ماتریس فاصله (۱۰٪)

حال ژن‌های متناظر از هر ۵ گونه را به صورت سراسری¹⁹ هم‌تراز کنید و فاصله ویرایش²⁰ را برای دو به دو آن‌ها بدست آورید. خروجی شما باید ۷ ماتریس ۵×۵ باشد که هر ماتریس باید به نام همان ژن و در یک فایل CSV ذخیره شود. برای این کار از الگوریتم نیدلمن-وونش²¹ استفاده کنید. **(استفاده از ابزارهای آنلاین مجاز نیست.)**

¹⁹ global

²⁰ Edit distance

²¹ Needleman-Wunsch

۳. درخت زندگی

۳.۱. تشکیل درخت زندگی برای هر ژن (۱۰٪)

حال با استفاده از دو الگوریتم UPGMA و neighbor-joining برای هر یک از ماتریس‌ها دو درخت زندگی²² بسازید. نتیجه را در فایل‌ها png ذخیره کنید و نامگذاری شما باید به این صورت باشد که UPGMA_geneX و NJ_geneX که به ترتیب برای الگوریتم‌های UPGMA و neighbor-joining به کار می‌روند و geneX نشان‌دهنده نام ژن مورد نظر است.

- نتایج دو الگوریتم را برای هر یک از پنج ژن مقایسه کنید.
- آیا تفاوت چشمگیری را در نتایج مشاهده می‌کنید؟
- به نظر شما تفاوت‌ها به چه دلیل دیده می‌شوند؟
- آیا می‌توانید بگویید هر الگوریتم برای چه دسته‌ای از مسائل مناسب است؟

۳.۲. ترکیب درخت‌ها و ارائه درخت نهایی (۱۵٪)

برای اینکه موجودات در نهایت با هم قابل مقایسه شوند نیازمند یک درخت نهایی هستیم. (توجه کنید که شما در واقع باید یک ماتریس فواصل جدید بدست آورید و سپس از آن درخت را بسازید.)

- روشی ارائه دهید که از تمام نتایج یک درخت به دست بدهد.
- از روش خود دفاع کنید.

توجه کنید که استفاده از همترازی سراسری برای رسیدن به درخت ترکیبی در این بخش مجاز نیست و باید از روش دیگری استفاده کنید.

²² Phylogenetic tree

۳.۳. مقایسه درخت ترکیبی و درخت حاصل از همترازی سراسری (۱۵٪)

در این بخش باید درخت حاصل از بخش ۳.۲ را با یک درخت که از فاصله ویرایش موجودات در همترازی سراسری بدست می آید مقایسه کنید. پیشنهاد ما برای انجام این بخش استفاده از سریعترین الگوریتم موجود برای همگرایی سراسری است که package های مناسبی برای آن به زبانهای برنامه نویسی مختلف موجود است. در غیر این صورت ممکن است با محدودیت منابع محاسباتی روبرو شوید. این الگوریتمها را می توانید تحت نام myers's bit vector algorithm یا پیاده سازی موازی آن پیدا کنید.

۳.۴. تعیین نقطه شروع (۱۰٪)

برای اینکه بدانیم اولین ابولا ویروس کدام بوده است، نیازمند کمک گرفتن از یک موجود دیگر هستیم که به لحاظ تکاملی نزدیک این ابولا ویروس باشد، کاندید ما برای این کار، ماربرگ ویروس است. ژنوم ماربرگ ویروس را از [اینجا](#) بگیرید و با هر روشی که میدانید یک درخت زندگی برای این ۶ ویروس (۵ گونه ابولا ویروس و این یک گونه از ماربرگ ویروس) بدست آورید. (برای بدست آوردن درخت از روشی که از بخش ۳.۱ بهتر میدانید استفاده کنید).

۴. تخمین گذشته، پیش بینی آینده!

۴.۱. چه زمانی از هم جدا شده اند؟ (۱۰٪)

همان طور که پیش تر گفته شد نرخ جهش برای هر جایگاه²³ از ابولا ویروس در طول یک سال، 1.9×10^{-3} است. با این اطلاعات

- فاصله زمانی دو به دو گونه ها را بدست آورید.
 - تخمین بزنید که گونه ها چه زمانی از جد خود جدا شده اند. (جد مشترک چه زمانی می زیسته است؟)
- (در این کار باید شرح بدهید از چه روش آماری و به چه دلیل استفاده کرده اید.)

راهنمایی: برای پیدا کردن فاصله ویرایش موجودات مختلف می توانید از نتایج همترازی سراسری استفاده کنید.

²³ site

۴.۲. چه زمانی یک گونه جدید به وجود خواهد آمد؟ (+۱۰٪)

به الگوهای تکامل که باعث شده است دو گونه از هم جدا شده اند دقت کنید. ساده ترین معیار میتواند تعداد جهش‌ها باشد. اما قطعا الگوهای دقیقتر از قبیل مکان جهش‌ها و ژنها هم پر اهمیت هستند. حال با نتایجی که بدست آوردید، یک توزیع احتمال از زمان ایجاد موجود بعدی ارائه دهید.

۵. ارائه گزارش

یکی از مهمترین قسمت‌های هر پروژه بیوانفورماتیک ارائه گزارش به شکلی قابل فهم و جامع برای طیف بزرگ محققین این حوزه به طور خاص زیست‌شناسان است. در این پروژه هم در آخرین گام شما باید یک گزارش جامع شامل روند کار، چالش‌ها و نتایج ارائه دهید. هر قدر قابلیت بازتولید²⁴ و فهم گزارش شما بیشتر باشد، کار شما ارزشمندتر خواهد بود.

۵.۱. نتایج نظری (+۵٪، ۵٪)

در طول پروژه سوالاتی مطرح شد که پاسخ آنها باید به صورت نظری داده شود. پاسخ این سوالات باید به شکل یک سند جداگانه تایپ شده و با نام theoritical_report.pdf در بخش نتایج شما قرار گیرد.

- لطفا اگر از منبعی استفاده می‌کنید در سند خود به آن ارجاع دهید.
- اگر با دانشجویی همکاری می‌کنید حتما نام آن دانشجو را ذکر کنید.
- اگر از منبعی استفاده میکنید به هیچ عنوان کپی نکنید و سعی کنید دوباره جمله‌بندی را انجام دهید.
- استفاده از لاتک برای ارائه گزارش دارای نمره اضافه است.
- مفهوم بودن متن و استفاده از المان‌های مناسب برای فهم بیشتر خواننده از جمله نمودارها، شکلها و فلوچارت‌ها دارای نمره اضافه است.
- حتما به چالش‌هایی که در حین پروژه به آنها برخوردید اشاره کنید و راهکارتان را برای رفع آن بگویید. این کار به بازتولید پروژه کمک می‌کند.

²⁴ Reproducibility

۵.۲. کدها و نتایج عملی (۵%+۱۰%)

امروزه ابزارهای بسیار مناسبی برای تولید یک گزارش عملی شامل کدها و نتایج وجود دارند که دو تا از کاملترین ابزارها [Rmarkdown](#) و [iPython notebook](#) هستند. با استفاده از این ابزارها میتوان کدهای نوشته شده (حتی به زبانهای متفاوت) را در کنار نتایج و توضیحات لازم آورد که نتیجه آن ارائه گزارشی بسیار جذاب و کامل است.

- یکی از دو ابزار فوق را برای ارائه گزارش انتخاب کنید. در مورد کدها توضیح دهید و نتایج را به همراه شکل‌های مناسب ارائه دهید.
- از کدهای آماده به هیچ وجه استفاده نکنید.
- از ابزارهای آنلاین استفاده نکنید.
- اگر با دانشجویی همکاری می‌کنید حتما نام آن دانشجو را ذکر کنید.
- اگر از کد خودتان که در تمرین دیگری از آن استفاده کردید استفاده می‌کنید این استفاده را ارجاع دهید.
- حتما به چالش‌هایی که در حین پروژه به آنها برخوردید اشاره کنید و راهکارتان را برای رفع آن بگویید. این کار به بازتولید پروژه کمک می‌کند.
- استفاده از کامنت‌های مناسب برای فهم کد توسط خواننده نمره اضافه دارد.
- به اشتراک گذاشتن پروژه روی مخازن عمومی اعم از گیت‌هاب نمره اضافه دارد.
- اگر در حین پروژه پکیجی توسط شما تولید شود و این پکیج کارایی مناسب را داشته باشد و شبیه آن موجود نباشد، نمره اضافه ویژه ای به این قسمت تعلق خواهد گرفت.

٦. منابع

1. Kuhn, J. H.; Becker, S.; Ebihara, H.; Geisbert, T. W.; Johnson, K. M.; Kawaoka, Y.; Lipkin, W. I.; Negredo, A. I.; Netesov, S. V.; Nichol, S. T.; Palacios, G.; Peters, C. J.; Tenorio, A.; Volchkov, V. E.; Jahrling, P. B. (2010). "Proposal for a revised taxonomy of the family Filoviridae: Classification, names of taxa and viruses, and virus abbreviations". *Archives of Virology*. **155** (12): 2083–103. doi:10.1007/s00705-010-0814-x. PMC 3074192. PMID 21046175.
2. Shabman, R. S.; Gulcicek, E. E.; Stone, K. L.; Basler, C. F. (2011-11-01). "The Ebola Virus VP24 Protein Prevents hnRNP C1/C2 Binding to Karyopherin 1 and Partially Alters its Nuclear Import". *Journal of Infectious Diseases*. **204** (suppl 3): S904–S910. doi:10.1093/infdis/jir323.
3. Zheng, H., Yin, C., Hoang, T., He, R. L., Yang, J., & Yau, S. S.-T. (2015). Ebolavirus Classification Based on Natural Vectors. *DNA and Cell Biology*, **34**(6), 418–428. <http://doi.org/10.1089/dna.2014.2678>.
4. <http://www.who.int/mediacentre/factsheets/fs103/en/>
5. <https://www.cdc.gov/vhf/ebola/outbreaks/history/chronology.html> | <https://www.cdc.gov/vhf/ebola/outbreaks/history/distribution-map.html>
6. Kalra S, Kelkar D, Galwankar SC, Papadimos TJ, Stawicki SP, Arquilla B, Hoey BA, Sharpe RP, Sabol D, Jahre JA. The emergence of Ebola as a global health security threat: From 'lessons learned' to coordinated multilateral containment efforts. *J Global Infect Dis* [serial online] 2014 [cited 2015 Mar 1]; 6:164–77.
7. Joel G. Breman, David L. Heymann, Graham Lloyd, Joseph B. McCormick, Malonga Miatudila, Frederick A. Murphy, Jean-Jacques Muyembé-Tamfun, Peter Piot, Jean-François Ruppel, Pierre Sureau, Guido van der Groen, Karl M. Johnson; Discovery and Description of Ebola Zaire Virus in 1976 and Relevance to the West African Epidemic During 2013–2016, *The Journal of Infectious Diseases*, Volume 214, Issue suppl_3, 15 October 2016, Pages S93–S101, <https://doi.org/10.1093/infdis/jiw207>.
8. Suresh Rewar, Dashrath Mirdha, Transmission of Ebola Virus Disease: An Overview, In *Annals of Global Health*, Volume 80, Issue 6, 2014, Pages 444–451, ISSN 2214-9996, <https://doi.org/10.1016/j.aogh.2015.02.005>.
9. Groseth, A., Feldmann, H., & Strong, J. E. (2007). The ecology of Ebola virus. *Trends in Microbiology*, **15**(9), 408–416. <https://doi.org/10.1016/j.tim.2007.08.001>
10. <https://www.cdc.gov/vhf/ebola/resources/virus-ecology-fr.html>

11. Lee, J. E., & Saphire, E. O. (2009). Ebolavirus glycoprotein structure and mechanism of entry. *Future Virology*, 4(6), 621–635. <https://doi.org/10.2217/fvl.09.56>
12. Sanchez, A., Kiley, M. P., Holloway, B. P., & Auperin, D. D. (1993). Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. *Virus Research*, 29(3), 215–240. [https://doi.org/10.1016/0168-1702\(93\)90063-s](https://doi.org/10.1016/0168-1702(93)90063-s)
13. Holmes, E. C., Dudas, G., Rambaut, A., & Andersen, K. G. (2016). The evolution of Ebola virus: Insights from the 2013–2016 epidemic. *Nature*, 538(7624), 193–200. <https://doi.org/10.1038/nature19790>
14. <https://www.mayoclinic.org/diseases-conditions/ebola-virus/symptoms-causes/syc-20356258>
15. Mühlberger, E. (2007). Filovirus replication and transcription. *Future Virology*, 2(2), 205–215. <https://doi.org/10.2217/17460794.2.2.205>
16. http://viralzone.expasy.org/207?outline=all_by_species
17. <https://www.bio-connect.nl/ebola-virus/cnt/page/4198>

موفق باشید،

الیاس حیدری، رسا قوامی، مجتبی نیک‌عهد، فاطمه اصلان بیگی