Winter (2023) 13(44): 45-52.
DOI: 10.30473/cb.2023.69786.1933
Received: 25, Aug. 2023 Accepted: 30, Dec. 2023

Crop Biotechnology Open Access

ORIGINAL ARTICLE

Analyzing the reverse transcriptase sequence of viruses using bioinformatics methods

Samira Karimi[1], Mahin Pouresmaeil[2*]

[1]M.Sc, Department of Biotechnology, Faculty of Agriculture, Azarbaijan Shahid Madani University, Tabriz, Iran.

[2]Postdoctoral Researcher, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran.

Correspondence
Mahin Pouresmaeil
Email: pouresmaeil.mahin@gmail.com

How to cite
Karimi, S., & Pouresmaeil, M. (2024). Probing the Reverse transcriptase Sequences of Viruses Using bioinformatics methods. *Crop Biotechnology*, *13*(44), 45-52.

ABSTRACT

Bioinformatics is an interdisciplinary science that utilizes information technologies to organize and analyze biological data. This science enables researchers to perform comprehensive and documented investigations on various biological problems without the need for expensive and time-consuming laboratory experiments. In this study, we acquired the reverse transcriptase (RT) sequence of eight virus strains from NCBI with the following accession numbers: NC_001497.2, NC_001648.1, NC_001839.2, NC_003977.2, AF053008.1, EF428979.1, NC_001802.1. We investigated the structural and functional characteristics, domains, and motifs. The analysis revealed that the proteins from the eight virus strains, belonging to different families, exhibited distinct properties that set them apart from one another. The analysis also showed that these proteins are found in the membrane, cytoplasm, and periplasm, and all of them contain at least one specific domain of the reverse transcriptase enzyme. Based on all the analyses performed, *Cauliflower mosaic virus*, *Cassava vein mosaic virus*, and *Soybean chlorotic mottle virus*, all belonging to the *Caulimoviridae* family, were suitable for producing RT enzymes. The ability of these viruses to adapt to different plant hosts could potentially lead to the development of more efficient and cost-effective methods for producing RT enzymes. This adaptability could also open up new possibilities for genetic engineering and biotechnology, enabling the development of more effective enzymes.

KEYWORDS

Reverse transcriptase, Bioinformatics, Protein domain, Virus

© 2023, by the author(s). Published by Payame Noor University, Tehran, Iran.
This is an open access article under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

سال سیزدهم، شماره دوم، پیاپی چهل و چهارم، زمستان ۱۴۰۲ (۵۲-۴۵)

تاریخ دریافت: ۱۴۰۲/۰۶/۰۳ تاریخ پذیرش: ۱۴۰۲/۱۰/۰۹

DOI: 10.30473/cb.2023.69786.1933

نشریه علمی

زیست‌فناوری گیاهان زراعی

«مقاله پژوهشی»

بررسی توالی ترانس کریپتاز معکوس ویروس‌ها با استفاده از روش‌های بیوانفورماتیک.

سمیرا کریمی[۱]، مهین پوراسمعیل[۲*]

[۱]دانش‌آموخته کارشناسی ارشد، گروه بیوتکنولوژی، دانشکده کشاورزی، دانشگاه شهید مدنی آذربایجان، تبریز ، ایران.

[۲]پژوهشگر پسادکتری، دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران.

نویسنده مسئول:
مهین پوراسمعیل
رایانامه: pouresmaeil.mahin@gmail.com

چکیده

بیوانفورماتیک یک علم بین‌رشته‌ای است که از فناوری اطلاعات برای سازماندهی و تجزیه و تحلیل داده-های بیولوژیکی استفاده می‌کند. این علم به پژوهشگران امکان می‌دهد که بدون نیاز به انجام آزمایش‌های زمان‌بر و پر هزینه، مطالعات مستند و جامعی در مورد مسائل مختلف علوم زیستی انجام دهند. هدف از این مطالعه بررسی و شناسایی توالی‌های مربوط به Reverse transcriptase (RT) در ویروس‌ها با استفاده از روش‌های بیوانفورماتیکی است. بدین منظور هشت سویه ویروس دارای توالی RT با شماره دسترسی NC_001497.2، NC_001648.1، NC_001839.2، NC_003977.2، AF053008.1، EF428979.1، NC_001802.1 از پایگاه NCBI استخراج گردید. در این مطالعه، خصوصیات ساختاری و عملکردی، دمین‌ها و موتیف‌های این پروتئین‌ها مورد بررسی قرار گرفت. آنالیزها نشان داد که پروتئین‌های هشت سویه ویروس، متعلق به خانواده‌های مختلف، ویژگی‌های متمایزی را از خود نشان می‌دهند که آنها را از یکدیگر متمایز می‌سازد. همچنین مشخص شد این پروتئین‌ها در غشاء، سیتوپلاسم و پری پلاسم قرار دارند و همه آنها حاوی حداقل یک دمین مربوط به آنزیم ترانس کریپتاز معکوس هستند. بر اساس آنالیزهای صورت گرفته *Cauliflower mosaic virus*، *Cassava vein mosaic virus* و *Soybean chlorotic mottle virus* که همگی متعلق به خانواده *Caulimoviridae* هستند، برای تولید آنزیم RT مناسب می‌باشند. توانایی این ویروس‌ها برای سازگاری با میزبان‌های مختلف گیاهی می‌تواند به طور بالقوه منجر به توسعه روش‌های کارآمدتر و مقرون به صرفه‌تر برای تولید آنزیم RT شود. این سازگاری همچنین می‌تواند فرصت‌های جدیدی را برای مهندسی ژنتیک و بیوتکنولوژی ایجاد کرده و امکان ایجاد آنزیم‌های با کارایی بیشتر را فراهم کند.

واژه‌های کلیدی

آنزیم رونویسی معکوس، بیوانفورماتیک، دمین پروتئینی، ویروس.

استناد به این مقاله:

کریمی، سمیرا و پوراسمعیل، مهین (۱۴۰۲). بررسی توالی ترانس کریپتاز معکوس در ویروس‌ها با استفاده از روش‌های بیوانفورماتیک. فصلنامه علمی زیست فناوری گیاهان زراعی، ۱۳(۴۴), ۴۵-۵۲.

مقدمه

ویروس‌ها عامل بیماری‌زای میکروسکوپی هستند که از یک مولکول اسید نوکلئیک (DNA یا RNA) تشکیل شده‌اند که توسط یک پوشش پروتئینی احاطه شده است. ویروس‌ها برای تکمیل چرخه زندگی خود به سلول‌های میزبان وابسته هستند. هنگامی که ویروس‌ها وارد سلول میزبان می‌شوند، ماشین‌های سلولی میزبان را در اختیار گرفته و تکثیر می‌یابند و در نهایت موجب ایجاد بیماری در میزبان خود می‌شوند (Harper, 2011). طبقه‌بندی و نام‌گذاری رسمی ویروس‌ها توسط کمیته بین‌المللی طبقه‌بندی ویروس‌ها (ICTV) صورت می‌گیرد. بر اساس طبقه بندی این کمیته، ویروس‌های دارای مکانیسم رونویسی معکوس به پنج خانواده تقسیم می‌شوند که عبارتنداز: *Caulimoviridae*، *Hepadnaviridae*، *Metaviridae*، *Pseuviridae*، *Retroviridae*. آنزیم رونوشت بردار معکوس (Reverse Transcription) نقش مهمی در روند تکثیر این پنج خانواده ویروس دارد (Menéndez-Arias, Sebastián-Martín, & Álvarez, 2017)، به طوری که آلوده‌سازی میزبان را از طریق تولید آنزیم رونویسی معکوس درطی فرآیند تکثیر انجام می‌دهند (Krupovic *et al*., 2018).

پیشینه پژوهش

رونویسی معکوس فرآیندی است که تبدیل RNA به مولکول‌های DNA مکمل (cDNA) را امکان‌پذیر می‌سازد و ابزار ارزشمندی را برای مطالعه الگوهای بیان RNA در اختیار محققان قرار می‌دهد. اولین مرحله شامل استفاده از ترانس کریپتاز معکوس است، آنزیمی که cDNA را از یک الگوی RNA سنتز می کند. این آنزیم به عنوان یک کاتالیزور بسیار خاص و کارآمد برای سنتز رشته‌های DNA با استفاده از RNA به عنوان یک الگو عمل می‌کند (Haddad & Baldwin, 2010). آنزیم رونوشت بردار معکوس با کاربردهای گسترده خود در رشته‌های مختلف علمی، نقش اساسی در روشن کردن فرآیندهای پیچیده بیولوژیکی و پیشرفت تحقیقات زیست پزشکی ایفا می‌کند (Menéndez-Arias & Delgado, 2022). این آنزیم در سال ۱۹۷۰ توسط دیوید بالتیمور و تیمین هاوارد کشف شد و موجب انقلاب در زیست‌شناسی مولکولی گردید (Coffin & Fan, 2016).

آنزیم رونوشت بردار معکوس کدشده توسط ویروس‌های مختلف اگر چه بر اساس ساختاری باهم متفاوتند اما همه آن‌ها توانایی تبدیل RNA را به DNA دو رشته‌ای از طریق فعالیت DNA پلیمراز وابسته به RNA، ریبونوکلئاز H (RNase H) و فعالیت DNA پلیمراز وابسته به DNA را دارند (Sluis-Cremer, 2021). این آنزیم بعنوان یک RNA polymerase وابسته به RNA که فاقد دنباله اگزونوکلئازی کنترل کننده خطا بوده و قادر به سنتز DNA از روی RNA می‌باشد (Bhagavan & Ha, 2015; Ellefson *et al*., 2016). آنزیم رونوشت بردار معکوس یکی از اجزای ضروری در تحقیقات علم ژنتیک در زمینه‌های مختلف محسوب می‌شود، و برای کلون کردن mRNA به صورت cDNA ، تولید پروب، آنالیز رونوشت به وسیله گسترش پرایمر، جهت تعیین مقدار کمی RNA، تهیه کتابخانه‌ی cDNA و بررسی بیان ژن مورد استفاده قرار می‌گیرد (Brown, 2020).

خانواده *Caulimoviridae* شامل ۸ جنس که بر اساس مورفولوژی ویریون در دو گروه ویروس‌های دارای ویریون کروی و ویروس‌های باسیلی شکل قرار می‌گیرند. این دو گروه دارای ژنوم DNA دورشته‌ای (dsDNA) بوده که فقط گیاهان را آلوده می‌کنند (Schoelz & Adhab, 2021; Teycheney *et al*., 2020). خانواده *Hepadnaviridae* دارای ۵ جنس از ویروس‌های کوچک با ژنوم DNA دو رشته‌ای که پرندگان و پستانداران را آلوده می‌کنند. همانندسازی این گروه از ویروس‌ها شامل رونویسی معکوس در سیتوپلاسم سلول‌های کبدی است (Magnius *et al*., 2020). اعضای خانواده *Pseudoviridae* و *Metaviridae* ویروس‌های DNAدار و دارای ساختار ۲۰ وجهی که نقش مهمی در چرخه عفونت آن‌ها ایفا می‌کند، همچنین این ویروس‌ها دارای رتروترانسپوزون‌های تکراری طولانی می‌باشند (Leonetti *et al*., 2020; Llorens *et al*., 2021). خانواده *Retroviridae* گروهی از ویروس‌های دارای دو نسخه از RNA تک رشته‌ای مثبت و شامل دو زیر خانواده *Orthoretrovirinae* و *Spumaretrovirinae* که طیف وسیعی از حیوانات و مهره‌داران را آلوده می‌کنند (Passos *et al*., 2021; Sluis-Cremer, 2021).

ویروس‌ها نقش مهمی در تنظیم اکوسیستم ایفا می‌کنند از این رو امروزه فناوری‌های توالی‌یابی به همراه ابزارهای جدید بیوانفورماتیکی زمینه‌های مناسبی برای مطالعه ویروس‌ها فراهم کرده است (Auslander *et al*., 2021; Hufsky *et al*., 2022). بیوانفورماتیک یک دانش بین رشته‌ای که شامل استفاده

از ابزارها و تکنیک‌های محاسباتی برای تجزیه و تحلیل داده‌های بیولوژیکی می‌باشد. بیوانفورماتیک نقش مهمی در درک سیستم‌های بیولوژیکی پیچیده و پیشرفت تحقیقات در زمینه‌هایی مانند ژنومیک، پروتئومیکس و زیست‌شناسی تکاملی دارد. با ادغام روش‌های محاسباتی و آماری با دانش بیولوژیکی، بیوانفورماتیک به دانشمندان اجازه می‌دهد تا مجموعه داده‌های بزرگ و متنوع را درک کنند و در نهایت منجر به اکتشافات و بینش‌هایی در مورد فرآیندهای اساسی زندگی شود (Gauthier *et al*., 2019; Rhee *et al*., 2006). در مطالعه حاضر توالی‌های نوکلئوتیدی کدکننده (CDs) و پروتئینی آنزیم RT در ۸ سویه ویروس از ۵ خانواده *Caulimoviridae*، *Hepaviridae*، *Metaviridae*، *Pseuviridae*، *Retroviridae* (به جز خانواده *Metaviridae* که سویه آن فاقد CDs کدکننده آنزیم RT می باشد.) با بهره‌گیری از ابزارهای بیوانفورماتیکی مورد مطالعه قرارگرفت.

مواد و روش‌ها

برای بررسی تنوع ژنتیکی و روابط فیلوژنتیک بین ۸ سویه ویروسی که دارای توالی کدکننده آنزیم رونویسی معکوس هستند ابتدا توالی‌ها از وب سایت (National Center for Biotechnology Information) NCBI با فرمت FASTA استخراج شد (جدول ۱). مقایسه توالی‌های نوکلئوتیدی ویروس‌ها به فرمت FASTA توسط ابزارMultAlin انجام شد، این سرور قابلیت مقایسه چندین توالی را دارد و در نهایت درخت فیلوژنتیک توسط نرم‌افزارMEGA6 با استفاده از روش (N.J) Neighbor joining و بوت استراپ (Bootstrap) 1000 رسم گردید (Allahi, Sohani, & Hasani Kumleh, 2017).

امروزه در بیوانفورماتیک، پیشرفت روش‌های محاسباتی برای شناسایی و بررسی موتیف‌ها یکی از چالش‌های بزرگ محسوب می‌شود. موتیف‌ها پپتیدهایی هستند که به عملکرد پروتئین‌ها و برهمکنش‌های پروتئین- پروتئین کمک ویژه‌ای می‌کنند. ویروس‌ها از موتیف‌ها برای ورود به میزبان، تعامل با پروتئین‌های سلولی یا خروج از سلول‌های میزبان استفاده می‌کنند. در این راستا مطالعه موتیف‌ها نقش احتمالی پروتئین‌ها را آشکار کرده و به درمان علیه بیماری‌ها ویروسی کمک شایانی می‌کند (Sobhy, 2016).

موتیف‌ها و دمین‌ها از واحدهای مهم و عملکردی توالی‌های اسیدآمینه می‌باشند و بیشتر توسط ابزارهایی مانند Pfam/NCBI-CDD و SMART شناسایی می‌شوند. شناسایی و مطالعه موتیف پروتئین‌های مربوط به هشت سویه ویروس از طریق سرور Motif search با کمک پایگاه داده Pfam و NCBI-CDD انجام شد. در این پژوهش بررسی دمین نیز توسط سرور SMART صورت گرفت.

جدول ۱. شماره دسترسی توالی‌های ۸ سویه ویروس

خانواده ویروسی	سویه ویروس	شماره دستیابی توالی نوکلئوتیدی	شماره دستیابی توالی اسیدآمینه
Caulimoviridae	*Cauliflower mosaic virus*	NC_001497.2	NP_056728.1
Caulimoviridae	*Cassava vein mosaic virus*	NC_001648.1	NP_056848.1
Caulimoviridae	*Petunia vein clearing virus*	NC_001839.2	NP_127504.1
	Soybean chlorotic mottle virus	NC_001739.2	NP_068729.1
Hepaviridae	*Hepatitis B virus*	NC_003977.2	YP_009173866.1
Metaviridae	------	-----	-----
Pseuviridae	*Glycine max SIRE1 virus*	AF053008.1	AAC64917.1
Retroviridae	*Walleye dermal sarcoma virus*	EF428979.1	ABO25842.2
	Human immunodeficiency virus 1	NC_001802.1	NP_057849.4

در این مطالعه از سرور PIR و ابزار Analysis برای تعیین وزن مولکولی پروتئین‌ها استفاده شد. همچنین از سرور targetP برای مشخص کردن وظیفه پروتئین‌ها استفاده گردید. محل پروتئین موردنظر نیز ازطریق سایت Cello مورد بررسی قرار گرفت. در جدول شماره ۲ اسامی پایگاه داده‌های استفاده شده در این پژوهش اشاره شده است.

جدول ۲. نام و آدرس پایگاه داده‌های استفاده شده در پژوهش

Website	URL
Blast	https://blast.ncbi.nlm.nih.gov/
MultAlin	http://multalin.toulouse.inra.fr/multalin/
Motif search	https://www.genome.jp/tools/motif
https://pfam.xfam.org/	https://pfam.xfam.org/
NCBI-CDD	https://www.ncbi.nlm.nih.gov/cdd
SMART	http://smart.embl-heidelberg.de/
PIR	https://proteininformationresource.org/pirwww/
targetP	http://www.cbs.dtu.dk/services/TargetP/)
Cello	http://cello.life.nctu.edu.tw/
Expasy	http://web.expasy.org/compute_pi/

بحث و نتایج

شناسایی سویه‌های مشابه به ۸ سویه ویروسی تولیدکننده آنزیم RT

مطالعه توالی‌های مشابه نقش مهمی در درک و مطالعه توالی هدف ما دارد، ازاین‌رو در این مطالعه توالی‌هایی با بیشترین شباهت به توالی‌های نوکلئوتیدی هدف در ۸ سویه ویروسی ازطریق بلاست استخراج و نتیجه آن درجدول ارائه شده هست (جدول۳).

بررسی روابط تکاملی و فیلوژنتیکی

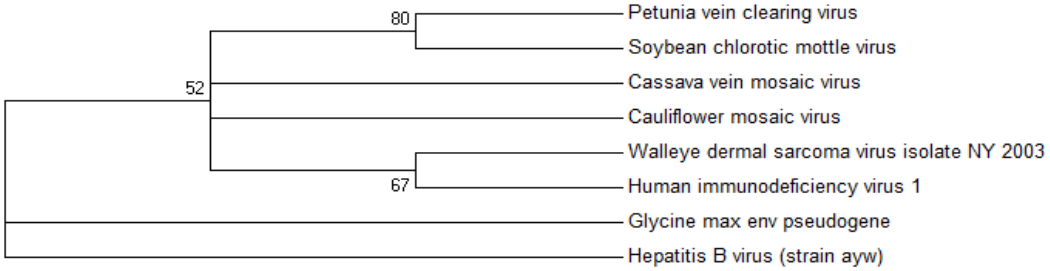
مقایسه چندین توالی یک راهکار مفید برای مطالعه تکامل ملکولی و تجزیه تحلیل روابط ساختاری در توالی‌ها می‌باشد (Naznin *et al*., 2011). نتایج بررسی روابط فیلوژنتیک در این پژوهش نشان داد که توالی‌های نوکلئوتیدی کدکننده RT در این سویه‌های ویروسی دارای شباهت‌های زیادی هستند که از نوکلئوتید شماره ۱۸۴۳ تا نوکلئوتید شماره ۵۱۵۰ مشهود است. این شباهت‌ها می‌توانند نشان‌دهنده انتقال ژنی یا تکامل مشترک بین این سویه‌های ویروسی باشند. این پژوهش می‌تواند برای شناسایی منشأ و تکامل ویروس‌های دارای توالی آنزیم RT مفید باشد.

جدول ۳. نتایج بلاست توالی‌های نوکلئوتیدی

شماره دسترسی توالی نوکلئوتیدی (سویه ویروس)	شماره دسترسی نتیجه بلاست	درصد شباهت
NC_001497.2	KY703614.1	97.01%
NC_001648.1	HQ694978.1	97.47%
NC_001839.2	AY228106.1	98.55%
NC_001739.2	MH718847.1	94.17%
NC_003977.2	MFG18341.1	99.97%
AF053008.1	AC235444.1	98.77%
EF428979.1	AF033822.1	99.81%
NC_001802.1	KU521529.1	98.99%

برای فهم بیشتر روابط فیلوژنتیکی و بررسی تنوع زیستی داده خام توالی‌ها نیاز به مطالعه مقایسه‌ای دارند (Kumar *et al*., 2018). بنابراین درخت فیلوژنتیک برای ژن‌های کد کننده آنزیم RT در ۸ سویه ویروس به روش NJ رسم و بوت استراپ‌های زیر ۵۰ به دلیل ارزش کمتر حذف گردید (شکل ۱).

نکته جالب توجه در درخت فیلوژنتیکی قرارگرفتن ژن‌های با توالی مشابه دریک کلاستر و همچنین تشکیل یک گروه پارانیایی بین اعضای خانواده (*Caulimoviridae* و *Retroviridae*) که دارای بوت استرپ ۵۲ و نشان‌دهنده شباهت زیاد توالی هدف، در این دو خانواده می‌باشد. براساس درخت فیلوژنتیکی سویه‌های ویروس در دو خانواده *Caulimoviridae* و *Retroviridae* ارتباط تکاملی نزدیکتری نسبت به اعضای دو گروه دیگر یعنی *Hepaviridae* و *Pseudoviridae* دارند، و همچنین سویه‌های *Hepatitis B virus* و *Glycine max SIRE1 virus* تغییرات بیشتری را در ژن مورد مطالعه نسبت به حالت اجدادی نشان می‌دهند (شکل۱).



شکل ۱. رابطه فیلوژنی ژن‌های کدکننده آنزیم رونویسی معکوس در ۸ سویه ویروس.درخت فیلوژنتیکی توسط نرم افزار MEGA6 با روش NJ و آزمون بوت استراپ رسم شده است.

بررسی اعضا ژنی کدکننده آنزیم RT

مطالعه ویژگی‌های فیزیکوشیمیایی ژن‌های هدف و همچنین محصول رونویسی آن‌ها (پروتئین) حاکی ازآنست که این ژن‌ها نقطه ایزوالکتریک و وزن ملکولی متفاوتی دارند، نقطه ایزوالکتریک، pH هست که پروتئین در آن فاقد بارالکتریکی می‌باشد و نقش مهمی در خالص‌سازی و رسوب پروتئین‌ها و طیف‌سنجی جرمی دارد (Kozlowski, 2017). بر طبق نتایج بدست‌آمده از ابزار تحلیل پروتئین (Expasy)، پروتئین به شماره دسترسی YP_009173866.1 بیشترین نقطه ایزوالکتریک و پروتئین NP_127504.1 وزن مولکولی بیشتری نسبت به سایر پروتئین‌ها دارد (جدول ۴).

مکان‌یابی پروتئین هدف در ۸ سویه ویروس نشان می‌دهد، پروتئین موردنظر (آنزیم) باتوجه به میزبان هدف ویروس‌های موردبحث در غشاها، سیتوپلاسم، پری‌پلاسم تولید و فعال می‌باشد و از همین طریق موجب اختلال در فعالیت‌های سلولی میزبان می‌شوند (جدول۴).

شناسایی دمین و موتیف مربوط به پروتئین‌های هدف

براساس مطالعات صورت گرفته دمین‌ها و موتیف توالی‌های هدف شناسایی شدند، طبق نتایج بدست آمده پروتئین شماره ABO25842.2 دارای بیشترین تعداد موتیف و پروتئین شماره NP_127504.1 دارای بیشترین تعداد دمین در توالی اسیدآمینه ای می باشند، در بین پروتئین‌های مورد بررسی همه آن‌ها شامل حداقل یک دمین احتصاصی آنزیم Reverse transcriptase بوده با این حال پروتئین شماره NP_057849.4 بیشترین دمین مربوط به آنزیم RT (RVT) دارد و همین امر نقش بسزایی در تولید این نوع آنزیم از ویروس مربوطه دارد (جدول ۵).

مطالعه دمین‌ها برخلاف موتیف‌ها درک مارا از اطلاعات ساختاری و عملکردی ژنوم افزایش می‌دهد، دمین‌ها به طور طبیعی تعامل و عملکرد پروتئین را تعیین و با تجزیه، تحلیل آن‌ها در توالی ژنوم نقش پروتئین مشخص می‌شود (Kozlowski, 2017; Vogel *et al*., 2004). موتیف‌ها می‌توانند در ایجاد ساختار پروتئین، تعامل و اتصال آن‌ها نقش مهمی داشته باشند (Kaiser *et al*., 2015).

جدول ٤. برخی ویژگی‌های پروتئین‌های هدف بر اساس شماره دسترسی در NCBI

جایگاه سلولی	میزبان ویروس	pH ایزوالکتریک	وزن ملکولی (کیلودالتون)	طول پروتئین (اسیدآمینه)	شماره دستیابی پروتئین
سیتوپلاسم	گل کلم	9.21	78629.04	679	NP_056728.1
سیتوپلاسم	مانیوک (کاساوا)	8.96	77054.06	652	NP_056848.1
غشا پری‌پلاسم	پتونیاهیبرید cv	8.68	252213.98	2179	NP_127504.1
سیتوپلاسم	سویا	9.20	80373.94	692	NP_068729.1
غشا	پستانداران	9.80	93676.73	832	YP_009173866.1
سیتوپلاسم	سویا	8.11	176659.19	1550	AAC64917.1
پری‌پلاسم	آبزیان	9.05	196238.96	1752	ABO25842.2
سیتوپلاسم	پستانداران، انسان	8.88	162041.90	1435	NP_057849.4

جدول ۵. دمین‌ها و تعداد موتیف‌های پروتئینی

شماره پروتئین	اسامی دمین	کدون آغاز	کدون پایان	تعداد موتیف
NP_056728.1	Peptidase-A3	20	24	17
	RVT-1	293	452	
	Low complexity	535	544	
NP_056848.1	Peptidase-A3	2	206	20
	RVP	3	104	
	RVT-1	250	413	
	Low complexity	458	472	
NP_127504.1	Mp	85	237	26
	Coiled coil	569	592	
	Low complexity	643	658	
	Low complexity	705	716	
	Low complexity	783	806	
	Low complexity	831	845	
	Low complexity	1084	1098	
	Znf – C_2HC	1111	1127	
	RVT-1	1428	1591	
	Low complexity	1661	1673	
	Low complexity	1822	1840	
	Low complexity	2121	2141	

ادامه جدول ۵. دمین‌ها و تعداد موتیف‌های پروتئینی

شماره پروتئین	اسامی دمین	کدون آغاز	کدون پایان	تعداد موتیف
NP_068729.1	Peptidase-A3	16	210	20
	RVT-1	246	411	
	Low complexity	549	557	
YP_009173866.1	DNApol- N ter	1	341	15
	Low complexity	400	417	
	RVT-1	490	589	
	DNApol- C ter	590	832	
AAC64917.1	Retrotran– gag- 2	52	191	31
	Znf– C_2HC	274	290	
	Low complexity	304	316	
	Low complexity	326	343	
	Coiled coil	353	409	
	Znf– C2HC	477	493	
	rve	722	840	
	RVT-2	1068	1311	
ABO25842.2	Low complexity	146	186	70
	Low complexity	484	495	
	Znf– C2HC	502	518	
	Low complexity	521	537	
	RVP	603	703	
	RVT-1	812	977	
	RNase- H	1223	1368	
	rve	1483	1596	
	Low complexity	1734	1743	

نتیجه‌گیری کلی

پیشرفت‌های گسترده علم زیست‌شناسی و کثرت داده‌های موجود در این علم، ابزارهای موجود در حوزه بیوانفورماتیک را تقریباً به بخش جدایی‌ناپذیر مطالعات زیستی تبدیل کرده است به طوری که دسترسی محققان را به اطلاعات و آنالیزهای مربوط به DNA و پروتئین موجودات را بیش از پیش سرعت و دقت بخشیده است. از این‌رو استفاده از داده‌ها و اطلاعات بیوانفورماتیکی قبل از انجام مطالعات پژوهشی کمترین هزینه را برای محقق دربر خواهد داشت و این نقطه مثبت مهمی در پیشبرد تحقیقات می‌باشد. رونویسی از روی RNA یکی از اقدامات کاربردی در کارهای تحقیقاتی و تشخیصی بخصوص در علم پزشکی می‌باشد. امروزه انواعی از پرکاربردترین آنزیم‌های M-MLV RT و AMV RT بصورت تجاری تولید و استفاده می‌شوند که هر یک به نوبه خود مزایا و معایبی دارند، بنابراین پژوهش حاضر می‌تواند به عنوان راه‌گشای پایه و علمی برای تولید آنزیم RT در مراکز تحقیقاتی، سودمند و کارساز واقع شود و توجه محققان این عرصه را به تولید آنزیم‌های رونویسی معکوس (RT) متنوع با قابلیت و کارایی بهتر و بیشتر سوق دهد. بر اساس نتایج حاصله از آنالیزهای بیوانفورماتیکی در پژوهش حاضر پروتئین تولیدی (آنزیم RT) در سویه *Cauliflower mosaic virus* و *Cassava vein mosaic virus* و *Soybean chlorotic mottle virus* دارای تعداد اسیدآمینه کمتر و نقطه ایزوالکتریک بیشتری می‌باشد که این دو پارامتر در خالص‌سازی پروتئین تولیدی نقش مهمی دارد. در مقابل طول ژن هدف در این سویه‌ها کمتر هست که این ویژگی هم در اهداف مربوط به PCR و کلونینگ ژن، نقطه مثبت و قابل توجهی می‌باشد. با توجه به شباهت زیاد توالی مربوط به ژن RT در سه سویه موردنظر و سویه *Human immunodeficiency virus* از خانواده *Retroviridae* گزینه بعدی برای تولید آنزیم هدف می‌تواند باشد از جمله این که توالی هدف در این ویروس دارای بیشترین دمین مربوط به آنزیم RT بوده و بالطبع قابلیت بهتری خواهد داشت ولی از آنجایی که میزبان این ویروس، انسان است ادامه مطالعات عملی با محدودیت و هزینه‌های بالایی روبرو خواهد شد. برخلاف این خانواده، سه سویه ویروس ذکر شده مربوط به خانواده *Caulimoviridae* هستند که دارای میزبان گیاهی می‌باشند بنابراین براحتی می‌توان ویروس موردنظر را در محیط کنترل شده القا و مطالعات بعدی را انجام داد. نتایج این مطالعه می‌تواند به عنوان یک منبع مفید برای شناخت و استفاده از این سویه‌های ویروسی در تولید آنزیم رونویسی معکوس در شرایط آزمایشگاهی (*in vitro*) برای پژوهشگران علوم زیستی باشد.

References

Allahi, S., Sohani, M. M., & Hasani Kumleh, H. (2017). In silico identification of the pld gene family and analysis of their expression pattern in response to salt stress in medicago truncatula. *Genetic Engineering and Biosafety Journal, 6*(1), 143-156.

Auslander, N., Gussow, A. B., & Koonin, E. V. (2021). Incorporating machine learning into established bioinformatics frameworks. *International journal of molecular sciences, 22*(6), 2903.

Bhagavan, N., & Ha, C.-E. (2015). DNA replication, repair, and mutagenesis. *Essentials of Medical Biochemistry, 2*, 401-417.

Brown, T. A. (2020). *Gene cloning and DNA analysis: An introduction*: John Wiley & Sons.

Coffin, J. M., & Fan, H. (2016). The discovery of reverse transcriptase. *Annual review of virology, 3*, 29-51.

Ellefson, J. W., Gollihar, J., Shroff, R., Shivram, H., Iyer, V. R., & Ellington, A. D. (2016). Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science, 352*(6293), 1590-1593.

Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019). A brief history of bioinformatics. *Briefings in bioinformatics, 20*(6), 1981-1996.

Gupta, A., Gangotia, D., & Mani, I. (2021). Bioinformatics tools and software. *Advances in Bioinformatics*, 15-35.

Haddad, F., & Baldwin, K. M. (2010). Reverse transcription of the ribonucleic acid: The first step in rt-pcr assay. *RT-PCR Protocols: Second Edition*, 261-270.

Harper, D. (2011). *Viruses: Biology, applications, and control*: Garland Science.

Hufsky, F., Abecasis, A., Agudelo-Romero, P., Bletsa, M., Brown, K., Claus, C., ... Gismondi, M. I. (2022). Women in the european virus bioinformatics center. *Viruses, 14*(7), 1522.

Kaiser, F., Eisold, A., & Labudde, D. (2015). A novel algorithm for enhanced structural motif matching in proteins. *Journal of Computational Biology, 22*(7), 698-713.

Kozlowski, L. P. (2017). Proteome-pi: Proteome isoelectric point database. *Nucleic acids research, 45*(D1), D1112-D1116.

Krupovic, M., Blomberg, J., Coffin, J. M., Dasgupta, I., Fan, H., Geering, A. D., . . . Johnson, W. (2018). Ortervirales: New virus order unifying five families of reverse-transcribing viruses. *Journal of virology, 92*(12), 10.1128/jvi. 00515-00518.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). Mega x: Molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution, 35*(6), 1547.

Kuznetsov, A., & Bollin, C. J. (2021). Ncbi genome workbench: Desktop software for comparative genomics, visualization, and genbank data submission. *Multiple Sequence Alignment: Methods and Protocols*, 261-295.

Leonetti, P., Miesen, P., van Rij, R. P., & Pantaleo, V. (2020). Viral and subviral derived small rnas as pathogenic determinants in plants and insects. *Advances in Virus Research, 107*, 1-36.

Llorens, C., Soriano, B., Krupovic, M., & Consortium, I. R. (2021). Ictv virus taxonomy profile: Pseudoviridae. *Journal of General Virology, 102*(3), 001563.

Magnius, L., Mason, W. S., Taylor, J., Kann, M., Glebe, D., Dény, P., ... Consortium, I. R. (2020). Ictv virus taxonomy profile: Hepadnaviridae. *Journal of General Virology, 101*(6), 571-572.

Menéndez-Arias, L., & Delgado, R. (2022). Update and latest advances in antiretroviral therapy. *Trends in pharmacological sciences, 43*(1), 16-29.

Menéndez-Arias, L., Sebastián-Martín, A., & Álvarez, M. (2017). Viral reverse transcriptases. *Virus research, 234*, 153-176.

Naznin, F., Sarker, R., & Essam, D. (2011). Vertical decomposition with genetic algorithm for multiple sequence alignment. *BMC bioinformatics, 12*(1), 1-26.

Passos, D. O., Li, M., Craigie, R., & Lyumkis, D. (2021). Retroviral integrase: Structure, mechanism, and inhibition. In *The enzymes* (Vol. 50, pp. 249-300): Elsevier.

Rhee, S. Y., Dickerson, J., & Xu, D. (2006). Bioinformatics and its applications in plant biology. *Annu. Rev. Plant Biol., 57*, 335-360.

Schoelz, J. E., & Adhab, M. (2021). Caulimoviruses (caulimoviridae).

Sluis-Cremer, N. (2021). Retroviral reverse transcriptase: Structure, function and inhibition. In *The enzymes* (Vol. 50, pp. 179-194): Elsevier.

Sobhy, H. (2016). A review of functional motifs utilized by viruses. *Proteomes, 4*(1), 3.

Teycheney, P.-Y., Geering, A. D., Dasgupta, I., Hull, R., Kreuze, J. F., Lockhart, B., . . . Pooggin, M. M. (2020). Ictv virus taxonomy profile: Caulimoviridae. *Journal of General Virology, 101*(10), 1025-1026.

Vogel, C., Berzuini, C., Bashton, M., Gough, J., & Teichmann, S. A. (2004). Supra-domains: Evolutionary units larger than single protein domains. *Journal of molecular biology, 336*(3), 809-823.

Zhang, Y., Zhang, Q., Zhou, J., & Zou, Q. (2022). A survey on the algorithm and development of multiple sequence alignment. *Briefings in bioinformatics, 23*(3), bbac069.