

## بررسی تنوع سیستم‌های CRISPR-Cas در باکتری کلاستریدیوم بوتولینوم با استفاده از رویکرد کاوش ژنی

نوع مقاله: پژوهشی

نعمیه شیخ الاسلامی<sup>۱</sup>، حمید میرزائی<sup>۱</sup>، جلیل خندقی<sup>۲</sup>، یوسف نامی<sup>۴</sup>، افشین جوادی<sup>۱\*</sup>

۱- گروه بهداشت مواد غذایی، دانشکده دامپزشکی، واحد علوم پزشکی تبریز، دانشگاه آزاد اسلامی، تبریز، ایران.

۲- گروه بیوتکنولوژی مواد غذایی، مرکز تحقیقات بیوتکنولوژی، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران.

۳- گروه علوم و صنایع غذایی، واحد سراب، دانشگاه آزاد اسلامی، سراب، ایران.

۴- گروه بیوتکنولوژی مواد غذایی، واحد شمال غرب و غرب پژوهشکده بیوتکنولوژی کشاورزی، سازمان تحقیقات، آموزش و ترویج

کشاورزی (AREEO)، تبریز، ایران.

hmirzaei@iaut.ac.ir

تاریخ دریافت: ۱۴۰۳/۰۲/۲۲، تاریخ پذیرش: ۱۴۰۳/۰۵/۰۱

صفحه ۳۴-۱۷

### چکیده

سیستم تکرارهای کوتاه پالیندرمیک خوشه‌ای منظم و پروتئین مرتبط با آن یا CRISPR-Cas اهداف ژنتیکی ارزشمندی برای تجزیه و تحلیل فیلوژنی باکتری‌ها در مطالعات همه‌گیری شناسی محسوب می‌شوند. در این مطالعه داده‌های ژنومی ۸۰۰ توالی سویه‌های کلاستریدیوم بوتولینوم از پایگاه داده مرکز ملی اطلاعات بیوتکنولوژی (NCBI) جمع‌آوری و حضور و تنوع سیستم‌های CRISPR-Cas کامل و توالی‌های فاصله‌انداز و تکرارهای آن در ژنوم‌های ثبت شده برای این باکتری‌ها با روش *In-silico* ارزیابی شد. همچنین ساختارهای ثانویه توالی‌های تکراری بر اساس الگوریتم حداقل انرژی آزاد (MFE) پیش‌بینی شد. بر اساس نتایج، از مجموع ۵۲۶ ژنوم با آرایه CRISPR-Cas کامل، تعداد ۳۰۶ سویه (۵۸ درصد) دارای ژن‌های *Cas* بودند. زیرگروه‌های I-B، I-D، II-C، III-D و III-B در این آرایه‌ها شناسایی شد. یافته‌ها همچنین نشان داد که 6 *Cas* پروتئین اصلی تولید شده توسط مجموعه ژن‌های بیان‌کننده در سیستم‌های CRISPR-Cas مورد مطالعه است و پروتئین‌های متنوع *Cas* توسط ماژول‌های سازگاری و تداخل در این آرایه‌ها تولید می‌شوند. طول متوسط توالی‌های فاصله‌انداز و تکرار در آرایه‌های CRISPR به ترتیب ۳۵-۳۰ و ۳۶-۳۰ جفت باز بود. به علاوه، پیش‌بینی ساختارهای ثانویه توالی‌های تکراری نشان داد که زیرگروه II-C، با ساقه‌های بلندتر نسبت به سایر زیرگروه‌ها، تمایل به تشکیل ساختارهای ثانویه RNA پایدارتری دارد.

واژه‌های کلیدی: توالی تکرار، توالی فاصله‌انداز، کاوش ژن، کلاستریدیوم بوتولینوم، CRISPR-Cas.

## مقدمه

بیماری‌زا این است که ژن‌های تولید توکسین در کلستریدیوم‌ها بر روی کروموزوم، پلاسمیدها یا فاژها شناسایی شده‌اند و این ژن‌ها می‌توانند از طریق پلاسمیدها (Smith et al. 2015) و همچنین باکتریوفاژها (Skarin et al. 2011) به روش انتقال افقی ژن (horizontal gene transfer) بین سویه‌های مختلف *C. botulinum* و همچنین برخی از گونه‌های غیر بوتولینوم جابجا شوند (Hill et al. 2015). بنابراین، دانش بیشتر در مورد سویه‌های *C. botulinum* به درک بهتر مکانیسم سم‌زایی و یافتن راهکارهای بازدارنده کمک می‌کند.

ارزیابی تنوع ژنتیکی سویه‌های دخیل در اپیدمی برای درک توزیع فیلوژنتیکی عوامل ایجادکننده، نظارت بر شیوع بیماری‌ها و اتخاذ تدابیر مدیریتی ضروری است. ردیابی و طبقه‌بندی جدایه‌های بالینی به طور معمول با روش‌هایی مانند ژل الکتروفورز میدان پالسی (pulse-field gel electrophoresis) و تعیین توالی چند لوکوس (multilocus sequence typing) انجام می‌شود (Kuijper et al. 2009) و همچنین روش استاندارد برای تفکیک نوع سویه‌ها نیز توالی‌یابی کل ژنوم جدایه‌ها است (Andersen et al. 2016)، اما این رویکردها هنوز برای اکثر آزمایشگاه‌ها غیرقابل دسترس هستند. در این میان، سیستم تکرارهای کوتاه پالیندرومیک خوشه‌ای منظم و پروتئین

کلستریدیوم بوتولینوم (*Clostridium botulinum*) باکتری گرم مثبت و بی‌هوای اجباری تولیدکننده آندوسپور است که به‌عنوان باکتری بیماری‌زا و عامل فساد غذایی شناخته می‌شود. سویه‌های این باکتری هتروژن بوده و از نظر سرولوژیکی هفت نوع از باکتری *C. botulinum* (A-G) شناخته شده‌است. سروتیپ‌های A، B، E و F باعث ایجاد بوتولیسم انسان و حیوان بوده درحالی که انواع C و D باعث بوتولیسم حیوانی می‌شوند (Lindström et al. 2010). سموم عصبی بوتولینوم قوی‌ترین سموم زیستی شناخته شده تا به امروز هستند که توسط *C. botulinum* و برخی دیگر از گونه‌های کلستریدیوم تولید می‌شوند. سم بوتولین باعث ایجاد عوارض متعددی مانند افتادگی پلک، فلجی، تهوع و استفراغ، گرفتگی شکم و اختلال در بلع یا تکلم می‌شود (Rossetto et al. 2014).

موارد اپیدمی متعدد بوتولیسم ناشی از مصرف سوسیس‌ها یا غذاهای کنسروی سبب اتخاذ اقدامات کنترلی خاصی به‌ویژه برای تولید غذاهای کنسروی شده‌است. با اینحال، به‌ویژه در مورد غذاهای آماده سرد، بوتولیسم همچنان یک چالش محسوب می‌شود، زیرا این میکروب می‌تواند در یخچال رشد کند (Hashempour-Baltork et al. 2019). یکی از نکات مهم در مورد این باکتری

## "شیخ الاسلامی و همکاران، بررسی تنوع بررسی تنوع سیستم‌های CRISPR-Cas در باکتری ..."

دادن دی.ان.ا را دارند (Karvelis et al. 2013). به‌علاوه، تنوع و پراکندگی این جایگاه‌ها که از آرایه‌های فاصله‌انداز و تکرارشونده تشکیل شده‌اند می‌تواند نشان‌دهنده مواجهه یک سویه میکروبی با عناصر ژنتیکی مهاجم بوده و در شکل‌گیری روابط خویشاوندی بین سویه‌های دارای اجداد مشترک و بنابراین نمایش مسیرهای تکاملی در طول زمان مورد استفاده قرار گیرد (Shariat and Dudley, 2014). بنابراین، ماهیت متفاوت آرایه‌های CRISPR آن‌ها را برای کاربردهای تعیین تیپ، به‌ویژه در گونه‌هایی با سیستم‌های فعال CRISPR-Cas ارزشمند می‌کند. همچنین پژوهشگران این سیستم را به‌عنوان ابزاری برای ویرایش ژن توسعه داده‌اند (Jiang et al. 2013). ابزارهای بیوانفورماتیکی روش‌هایی برای توسعه دانش در حوزه‌های مختلف، بدون نیاز به آزمایشات وقت‌گیر و گران‌قیمت آزمایشگاهی هستند (Lino et al. 2018) به‌طوری‌که در گذشته مطالعات متعددی در مورد سیستم‌های CRISPR بر اساس این رویکرد در باکتری‌های گوناگون انجام شده است (Nami et al. 2021; Pursey et al. 2022; Yadav and Singh, 2022).

این مطالعه با هدف بررسی حضور و ویژگی‌های سیستم‌های CRISPR-Cas در سویه‌های مختلف *C. botulinum* با روش *In-silico* طراحی شده است. اطلاعاتی که در مورد سیستم‌های CRISPR-

مرتبط با آن یا CRISPR-Cas (Clustered regularly interspersed short palindromic repeats-CRISPR associated protein) اهداف ژنتیکی ارزشمندی در مطالعات اپیدمیولوژیک برای تجزیه و تحلیل فیلوژنی باکتری‌ها (از جمله بیماری‌زاهای انسانی) با دقت بالا محسوب می‌شوند (Shariat and Dudley, 2014).

سیستم‌های CRISPR-Cas در اصل یکی از سیستم‌های ایمنی میکروبی است که در دفاع باکتری در برابر انواع عناصر ژنتیکی متحرک مانند فاژها و پلاسمیدها نقش دارد. این راهکارها با برش عناصر ژنتیکی مهاجم خارجی از ایجاد عفونت جلوگیری می‌کنند. آرایه‌های CRISPR نشان‌دهنده حافظه ایمنی عفونت‌های قبلی است که درون فاصله‌اندازهای جداگانه رمزگذاری شده‌اند و با تکرارهای حفظ‌شده از هم جدا می‌شوند (Panahi et al. 2023). به‌عبارت بهتر پروتئین‌های Cas اسید نوکلئیک خارجی را که وارد سلول پروکاریوتی شده است، شناسایی کرده و آن را به یک توالی جدید پردازش می‌کنند که به‌صورت یک رونوشت آر.ان.ا از ناحیه CRISPR موسوم به crRNA نگهداری می‌شود. این فرآیند به سلول باکتری اجازه می‌دهد تا در صورت مواجهه مجدد، اسید نوکلئیک خارجی را هدف قرار دهند. همچنین پروتئین‌های Cas به‌عنوان نوکلئاز و هلیکاز عمل کرده و توانایی باز کردن و برش

فاصله‌اندازه‌های طولانی، به چندین مورد تقسیم شود که برای جلوگیری از این حالت، کریسپرهایی که توالی‌های تکراری مشابه‌ای را در یک مکان به اشتراک می‌گذارند و نزدیک به یکدیگر هستند (۱۰۰۰۰ جفت باز)، به‌عنوان یک CRISPR واحد در نظر گرفته شدند. در قدم بعدی، تکرارهای حفاظت شده برای هر آرایه CRISPR تعیین و با استفاده از CD-HIT-EST بر اساس دستورالعمل و گزینه‌های استاندارد خوشه‌بندی شد (Li and Godzik, 2006). سپس از hmmscan برای جستجوی پروتئین‌های Cas مرتبط با سیستم‌های شناسایی شده با بهره‌گیری از داده پایگاه pfam استفاده شد و بر این اساس پروتئین‌های Cas یکی از دسته‌های پروتئینی universal, Cas1, Cas5, Cas7, Cas8, Cas9, Cas10, csf1 و cpf1 طبقه‌بندی شدند (Panahi et al. 2023).

#### تعیین نوع و زیرشاخه CRISPR-Cas

جایگاه CRISPR همراه با ژن‌های Cas مرتبط با آن به‌عنوان یک جایگاه CRISPR-Cas تعریف می‌شوند. برای پیش‌بینی نوع سیستم CRISPR از پایگاه داده CRISPR-Cas<sup>++</sup> (<https://crisprcas.i2bc.paris-saclay.fr/>) استفاده شد (Nami et al. 2023). جایگاه‌های CRISPR، تکرارهای مستقیم و فاصله‌اندازها در سویه‌های باکتریایی *C. botulinum*، در این داده پایگاه تعیین

Cas به دست خواهد آمد، می‌تواند در درک بهتر زمینه‌های تکاملی این باکتری بیماری‌زا و بررسی اهداف فاژی و یا پلاسمیدی باکتری و مکانیسم‌های دفاعی ذاتی آن در برابر عناصر متحرک ژنتیکی در مطالعات آینده مورد استفاده قرار گیرد.

### مواد و روش‌ها

#### جمع‌آوری داده‌های ژنومی

به‌منظور جمع‌آوری داده‌های سویه‌های مختلف *C. botulinum* در این پژوهش، لیست ۸۰۰ توالی ژنومی و ژنوم‌های سرهم‌بندی شده با متاژنوم یا MAGs (Metagenome-assembled genomes) برای این باکتری استفاده شد. توالی‌های ژنوم کامل به همراه اطلاعات مرتبط و همچنین MAG برای کلاستریدایوم بوتولینوم (در قالب FASTA) از پایگاه داده مرکز ملی اطلاعات بیوتکنولوژی (NCBI) جمع‌آوری و استخراج شد.

#### شناسایی سیستم‌های CRISPR-Cas در ژنوم

##### باکتری‌ها

در قدم اول، تشخیص تکرارهای جزئی در انتهای آرایه‌های CRISPR در ژنوم باکتری‌ها با استفاده از Meta CRT انجام شد (Moller and Liang, 2017). یک آرایه CRISPR طولانی ممکن است به دلیل تکرار با جهش‌های بیش از حد یا

## "شیخ الاسلامی و همکاران، بررسی تنوع بررسی تنوع سیستم‌های CRISPR-Cas در باکتری ..."

جایگزین شده و سپس ساختار ثانویه RNA بر اساس الگوریتم حداقل انرژی آزاد (minimum free energy) در پایگاه RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) پیش‌بینی شد (Hofacker, 2003). لازم به ذکر است در این نرم‌افزار پیش‌فرض تعیین شده در نرم‌افزار ارائه بهترین تاخوردگی است.

### نتایج و بحث

#### سیستم‌های CRISPR-Cas شناسایی شده

در این مطالعه، از مجموع ۵۲۶ ژنوم با آرایه CRISPR-Cas کامل متعلق به سویه‌های *C. botulinum* که از پایگاه داده NCBI جمع‌آوری شد، بر اساس جستجوی سیستم‌های CRISPR تأیید شده ۳۰۶ جدایه دارای ژن‌های *Cas* بودند که ۵۸ درصد از سویه‌های مورد بررسی را شامل می‌شود. تا به امروز، تنها یک مطالعه در رابطه با نقش و ارزیابی سیستم‌های CRISPR در باکتری *C. botulinum* منتشر شده است (Negahdaripour et al. 2017). در مطالعه مذکور که بر روی تعداد ۱۰۰ سویه *C. botulinum* انجام شده بود ۸۰ درصد سویه‌های مورد بررسی دارای سیستم‌های CRISPR تأیید شده بودند.

برای تعیین نوع و زیرگروه آرایه‌های کامل CRISPR-Cas سویه‌های *C. botulinum*

شدند. در مرحله بعد، ۱۰۰۰۰ جفت باز در بالادست و پایین دست CRISPR‌های به دست آمده انتخاب و به وب سرور CRISPR-Cas Finder (<https://crisprcas.i2bc.paris-saclay.fr/>) منتقل شد تا ژن‌های *Cas* مرتبط را شناسایی کند (Panahi et al. 2023). در این مرحله، نوع سیستم‌های دفاعی در یکی از کلاس‌های I, II, III, IV و V تعیین گروه و طبقه‌بندی خواهند شد.

به‌علاوه، برای اعتبارسنجی بیشتر و نیز به منظور شناسایی نوع و زیرشاخه آرایه‌های CRISPR شناسایی‌شده، از پایگاه داده CRISPRone (<https://omics.informatics.indiana.edu/CRISPRone/>) استفاده شد و فقط مکان‌های تأیید شده برای تنوع سیستم CRISPR مورد ارزیابی قرار گرفت (Zhang and Ye, 2017). پایگاه داده CRISPRone ابزاری برای پیش‌بینی مکان‌های CRISPR و تصویرسازی آرایه CRISPR است و اطلاعاتی مانند مشخصات آرایه، مکان و جهت *tracrRNA*، طول ناحیه ضد تکرار (anti-repeat) و عدم تطابق (mismatch) را ارائه می‌دهد.

#### پیش‌بینی ساختار ثانویه توالی‌های تکراری

برای مطالعه ساختارهای ثانویه توالی‌های تکراری بر اساس توالی‌های حفاظت شده در هر نوع از آرایه‌های CRISPR، نوکلئوتیدهای T با U

شناسایی شده در پژوهش حاضر از الگوریتم Blast پروتئین استفاده شد. در جدول ۱، جایگاه‌های CRISPR شناسایی شده کامل که حاوی ژن *Cas* بودند نشان داده شده است. همانطوری که مشاهده می‌شود زیرگروه‌های I-B، I-D، II-C، III-D و III-B در این آرایه‌ها وجود دارد (جدول ۱).

جدول ۱- آرایه‌های CRISPR-Cas شناسایی شده در سویه‌های *C. botulinum*

ردیف	نوع	نام	ردیف	فرآیند	نوع	نام	ردیف	فرآیند	نوع	نام	ردیف
۱	I-D	IZSVe-TV_1899_2_14	۲۰۵	-	I-B	Su1935	۱۰۳	-/+	I/III	Okra	۱
۲	I-D	IZSVe-TV_1671_2_11	۲۰۶	+	I-B	Su1918	۱۰۴	+	I-B	BF	۲
۳	I-D	IZSVe-TV_7863_07	۲۰۷	-	I-B	AM526	۱۰۵	+	I-B	Eklund	۳
۴	I-D	IZSVe-TV_6025_15	۲۰۸	+	I-B	Su0850	۱۰۶	+	I-B	Loch Maree	۴
۵	I-D	IZSVe-TV_3522_4_13	۲۰۹	+	I-B	Man111	۱۰۷	+	I-B	BoNT E Beluga	۵
۶	I-D	IZSVe-TV_2667_2_14	۲۱۰	+	I-B	BrQuSJ	۱۰۸	+	I-B	1873	۶
۷	I-D	IZSVe-TV_2454_7_14	۲۱۱	-	I-B	Su0110	۱۰۹	+	I-B	V891	۷
۸	I-D	IZSVe-TV_3524_3_13	۲۱۲	+/+	I-B/III-B	CDC59947	۱۱۰	-	I-B	BkT028 387	۸
۹	I-D	IZSVe-TV_1585_18_11	۲۱۳	+/+	I-B/III-B	CDC48611	۱۱۱	+	I-B	BKT01 5925	۹
۱۰	I-D	IZSVe-TV_5262_1_14	۲۱۴	+/+	I-B/III-B	CCUG7969	۱۱۲	+	I-B	CB11.1 -1	۱۰
۱۱	I-D	IZSVe-TV_2454_6_14	۲۱۵	-/-	I-B/III-B	62-A-KM138-1	۱۱۳	+	I-B	Af84	۱۱
۱۲	I-D	IZSVe-TV_3522_5_13	۲۱۶	+/+	I-B/III-B	Cam11A-KM36-1	۱۱۴	+	I-B	CDC54 085	۱۲
۱۳	II-C	CCUG7971	۲۱۷	+/+	I-B/III-B	JP6-A-KM22-1	۱۱۵	-	I-B	CDC54 075	۱۳
۱۴	II-C	1277_DC	۲۱۸	+/+	I-B/III-B	78A-KM34-1	۱۱۶	+	I-B	16868	۱۴
۱۵	III-D	F1425	۲۱۹	+	I-B	DFPST0043	۱۱۷	+	I-B	BKT75 002	۱۵
۱۶	III-D	B305	۲۲۰	+	I-B	KM28	۱۱۸	-/+	I-B	BKT28 73	۱۶
۱۷	III-D	B609	۲۲۱	-	I-B	KM78-1-VHE	۱۱۹	-/-	I-B/III	BKT12 695	۱۷
۱۸	III-D	B515	۲۲۲	+/+	I-B/III-B	DFPST009	۱۲۰	-	I-B	Sp77	۱۸
۱۹	III-D	Su1036	۲۲۳	+	I-B	KM125-1-BelugaE	۱۲۱	+	I-B	B2 433	۱۹
۲۰	III-B	ATCC3502	۲۲۴	+/+	I-B/III-B	DFPST008	۱۲۲	-/-	I-B/III-B	A2B387	۲۰
۲۱	III-B	Hall	۲۲۵	+	I-B	DFPST010	۱۲۳	-	I-B	B2 331	۲۱

"شیخ الاسلامی و همکاران، بررسی تنوع سیستم‌های CRISPR-Cas در باکتری ..."

ردیف	نوع	نام	ردیف	فرارگری	نوع	نام	ردیف	فرارگری	نوع	نام	ردیف
۲۲	III-B	langeland	۲۲۶	+/+	I-B/III-B	69-A-KM145-1	۱۲۴	+	I-B	A2B792	۲۲
۲۳	III-B	ATCC19397	۲۲۷	+	I-B	SCPM-O-B-7189	۱۲۵	+	I-B	KAPB-3	۲۳
۲۴	III-B	657	۲۲۸	-	I-B	SCPM-O-B-8707	۱۲۶	-/+	I-B/III-B/I-D	DC5	۲۴
۲۵	III-B	230613	۲۲۹	-/+	I-B	IZSVe-TV_9877_1_12	۱۲۷	-/-	I-B/III	111	۲۵
۲۶	III-B	H04402065	۲۳۰	+	I-B	IZSVe-TV_6537_1_13	۱۲۸	-	I-B	Prevot_594	۲۶
۲۷	III-B	CFSAN002369	۲۳۱	+	I-B	IZSVe-TV_6433_3_13	۱۲۹	-/+	I-B/III-B	CDC_1436	۲۷
۲۸	III-B	CFSAN002367	۲۳۲	+	I-B	IZSVe-TV_10018_12	۱۳۰	+	I-B	B2 450	۲۸
۲۹	III-B	CFSAN002368	۲۳۳	+/+/-	I-B/III-B/I-D	IZSVe-TV_7573_3_12	۱۳۱	-	I-B	CDC68016	۲۹
۳۰	III-B	Osaka05	۲۳۴	-/+	I-B/I-D/III-B	IZSVe-TV_7573_4_12	۱۳۲	-/+	I-B/III-B	CDC41370	۳۰
۳۱	III-B	CDC54088	۲۳۵	-	I-B	IZSVe-TV_900_3_11	۱۳۳	-	I-B	48212-CD	۳۱
۳۲	III-B	B2_267	۲۳۶	-	I-B	IZSVe-TV_9877_3_12	۱۳۴	+	I-B	38028-CD	۳۲
۳۳	III-B	B2 275	۲۳۷	+	I-B	IZSVe-TV_6433_2_13	۱۳۵	+	I-B	12LNR13-CD	۳۳
۳۴	III-B	CDC-297	۲۳۸	+	I-B	IZSVe-TV_6433_4_13	۱۳۶	-	I-B	49511-CD	۳۴
۳۵	III-B	Prevot25NCASE	۲۳۹	-	I-B	IZSVe-TV_6186_1_16	۱۳۷	+/+	I-B/I-D/III-B	51714-DC	۳۵
۳۶	III-B	CDC41370NT	۲۴۰	-	I-B	IZSVe-TV_7608_3_14	۱۳۸	-/+	I-B/I-D/III-B	47295-DC	۳۶
۳۷	III-B	Walls8G	۲۴۱	-	I-B	IZSVe-TV_6606L_3_14	۱۳۹	+	I-B	71840-CD	۳۷
۳۸	III-B	ATCC17862	۲۴۲	-	I-B	IZSVe-TV_6188_5_16	۱۴۰	+	I-B	58272-CD	۳۸
۳۹	III-B	AM1295	۲۴۳	-	I-B	IZSVe-TV_5694_2_13	۱۴۱	+/+	I-B/III-B	ATCC25763	۳۹
۴۰	III-B	SU 1575 NT	۲۴۴	-/-	I-B/III-B	IZSVe-TV_10380_20_14	۱۴۲	-	I-B	ATCC17786	۴۰
۴۱	III-B	Prevot 697B	۲۴۵	-	I-B	IZSVe-TV_8103_09	۱۴۳	+	I-B	SU1064	۴۱
۴۲	III-B	SU 0729	۲۴۶	-	I-B	IZSVe-TV_2659_2_10	۱۴۴	-/+	I-B/III-B	211	۴۲
۴۳	III-B	McCLung 844	۲۴۷	-	I-B	IZSVe-TV_6433_6_13	۱۴۵	+	I-B	SU0994	۴۳
۴۴	III-B	ATCC 449	۲۴۸	-	I-B	IZSVe-TV_2659_1_10	۱۴۶	-	I-B	SU1891	۴۴
۴۵	III-B	Walls 8G	۲۴۹	+/-	I-B/III-B	IZSVe-TV_4045_5_13	۱۴۷	-	I-B	SU1274	۴۵
۴۶	III-B	VPI 7124	۲۵۰	-/-	I-B/III-B	IZSVe-TV_10380_18_14	۱۴۸	-	I-B	SU1054	۴۶
۴۷	III-B	KFMeyer 33	۲۵۱	-/-	I-B/III-B	IZSVe-TV_10380_17_14	۱۴۹	-	I-B	SU1072	۴۷

"مجله ایمنی زیستی، دوره ۱۶، شماره ۴، زمستان ۱۴۰۲"

ردیف	نوع	نام	ردیف	نتیجه	ردیف	نوع	ردیف	نتیجه	ردیف	نوع	نام	ردیف
-	III-B	CDC_69096	۲۵۲	-	I-B	IZSVe-TV_5313_4_14	۱۵۰	-/-	I-B/III-B	Prevot Dewpin g	۴۸	
-	III-B	GiorgioA-KM21-1	۲۵۳	+	I-B	IZSVe-TV_6315_2_15	۱۵۱	+	I-B	SU 1304	۴۹	
+	III-B	KM123	۲۵۴	+/-	I-B/III-B	IZSVe-TV_7574_5_12	۱۵۲	+	I-B	SU 1917	۵۰	
+	III-B	CCUG7968	۲۵۵	+/-	I-B/III-B	IZSVe-TV_1585_19_11	۱۵۳	-	I-B	SU 1937	۵۱	
-	III-B	CDC21601	۲۵۶	-	I-B	IZSVe-TV_6660_7_15	۱۵۴	-/-	I-B/III-B	KF Meyer 126	۵۲	
-	III-B	B742	۲۵۷	-	I-B	IZSVe-TV_66061_3_14	۱۵۵	+/+	I-B/III-B	Hall 80	۵۳	
+	III-B	CFSAN034200	۲۵۸	-	I-B	IZSVe-TV_5792_11	۱۵۶	-/-	I-B/III-B	An436	۵۴	
-	III-B	CDC69043	۲۵۹	+	I-B	IZSVe-TV_5313_3_14	۱۵۷	+/+	I-B/III-B	Prevot 910	۵۵	
+	III-B	CDC69057	۲۶۰	-	I-B	IZSVe-TV_3650_3_13	۱۵۸	+	I-B	CDC KA-95B	۵۶	
+	III-B	CDC_67086	۲۶۱	-	I-B	IZSVe-TV_3650_1_13	۱۵۹	-	I-B	CDC_6 9094	۵۷	
+	III-B	CDC68158	۲۶۲	-	I-B	IZSVe-TV_7494_7_14	۱۶۰	-	I-B	AM282	۵۸	
+	III-B	NCTC13319	۲۶۳	-	I-B	IZSVe-TV_6503_1_13	۱۶۱	+	I-B	F634	۵۹	
+	III-B	DFPST011	۲۶۴	+/-	I-B/III-B	IZSVe-TV_5262_6_14	۱۶۲	-/+	I-B/I-D/II-C	1276_D C	۶۰	
-	III-B	X58540	۲۶۵	+/-	I-B/III-B	IZSVe-TV_3859_5_11	۱۶۳	-/+	I-B/I-D/II-C	1274_D C	۶۱	
+	III-B	H20	۲۶۶	+/-	I-B/III-B	IZSVe-TV_1899_2_14	۱۶۴	+/+	I-B/III-B	BS-A-KM7-1	۶۲	
+	III-B	P85	۲۶۷	+/-	I-B/III-B	IZSVe-TV_1671_2_11	۱۶۵	+	I-B	DFPST 010	۶۳	
-	III-B	H11	۲۶۸	+/-	I-B/III-B	IZSVe-TV_7863_07	۱۶۶	+/+	I-B/III-B	CCUG 7969	۶۴	
-	III-B	2015006193-DFPST	۲۶۹	+/-	I-B/III-B	IZSVe-TV_6025_15	۱۶۷	-	I-B	BrDura	۶۵	
-	III-B	KM87-1-86F	۲۷۰	+/-	I-B/III-B	IZSVe-TV_3522_4_13	۱۶۸	+	I-B	A634	۶۶	
-	III-B	KM84	۲۷۱	+/-	I-B/III-B	IZSVe-TV_2667_2_14	۱۶۹	+	I-B	AM105 1	۶۷	
-	III-B	DFPST0028	۲۷۲	+	I-B	IZSVe-TV_4863_4_13	۱۷۰	-/-	I-B/III-B	Food_2 0	۶۸	
+	III-B	GS1342-KM11-1	۲۷۳	-	I-B	IZSVe-TV_9659_15	۱۷۱	+/+	I-B/III-B	RF5	۶۹	
+	III-B	2015006195-DFPST	۲۷۳	+	I-B	IZSVe-TV_5792_2_15	۱۷۲	+/+	I-B/III-B	MAP5	۷۰	
-	III-B	DFPST0033	۲۷۵	-/-	I-B/III-B	IZSVe-TV_2454_7_14	۱۷۳	-/-	I-B/III-B	CFSAN 064329	۷۱	
+	III-B	Cam2A-KM19-1	۲۷۶	-	I-B	IZSVe-TV_6289_14	۱۷۴	+/+	I-B/III-B	ATCC1 7841	۷۲	

"شیخ الاسلامی و همکاران، بررسی تنوع سیستم‌های CRISPR-Cas در باکتری ..."

ردیف	نوع	نام	ردیف	فرارگری	نوع	نام	ردیف	فرارگری	نوع	نام	ردیف
۷۳	III-B	GiorgioA-KM21-1	۲۷۷	-	I-B	IZSVe-TV_6422_2_14	۱۷۵	+/+	I-B/III-B	CDC69068	۷۳
۷۴	III-B	CJ5-1	۲۷۸	+	I-B	IZSVe-TV_5850_15	۱۷۶	-/+	I-B/III-B	CDC37496	۷۴
۷۵	III-B	2015006187-DFPST	۲۷۹	+/-	I-B/III-B	IZSVe-TV_3524_3_13	۱۷۷	+/+	I-B/III-B	BAC-04-07755	۷۵
۷۶	III-B	DFPST0032	۲۸۰	-	I-B	IZSVe-TV_4863_2_13	۱۷۸	+	I-B	AM922	۷۶
۷۷	III-B	Cam3A-KM20-1	۲۸۱	+/-	I-B/III-B	IZSVe-TV_1585_18_11	۱۷۹	+/+	I-B/III-B	ZBS3_16-257-02	۷۷
۷۸	III-B	KM52-1-383B	۲۸۲	-/+	I-B/III-B	IZSVe-TV_5262_1_14	۱۸۰	+/+	I-B/III-B	ZBS3_16-240-01	۷۸
۷۹	III-B	97A-KM35-1	۲۸۳	+	I-B	IZSVe-TV_5792_1_15	۱۸۱	+/+	I-B/III-B	ZBS3_16-257-05	۷۹
۸۰	III-B	Cbot23-KM43-1	۲۸۴	-/+	I-B/III-B	IZSVe-TV_2454_6_14	۱۸۲	+/+	I-B/III-B	ZBS3_16-257-04	۸۰
۸۱	III-B	73A-KM33-1	۲۸۵	-	I-B	IZSVe-TV_4879_20_10	۱۸۳	-/-	I-B/III-B	ATCC25763	۸۱
۸۲	III-B	DFPST0024	۲۸۶	+/-	I-B/III-B	IZSVe-TV_3522_5_13	۱۸۴	+	I-B	2015006229-DFPST	۸۲
۸۳	III-B	DFPST0025	۲۸۷	-/-	I-B/III-B	IZSVe-TV_10380_13_14	۱۸۵	-/-	I-B/III-B	DFPST0041	۸۳
۸۴	III-B	DFPST0027	۲۸۸	-	I-B	FE9504ACG	۱۸۶	-	I-B	KM25	۸۴
۸۵	III-B	KM123	۲۸۹	-/+	I-B/III-B	OKra_BUMBAC_00010	۱۸۷	-/+	I-B/III-B	DFPST012	۸۵
۸۶	III-B	KM88-1-PC-F	۲۹۰	-	I-B	Beluga_EUMBA_C_00009	۱۸۸	+/+	I-B/III-B	DFPST007	۸۶
۸۷	III-B	DFPST0029	۲۹۱	+/+	I-B/III-B	69_AUMBAC_00004	۱۸۹	+/+/+	I-B/III-B/III-B	DFPST002	۸۷
۸۸	III-B	DFPST0031	۲۹۲	-/-	I-B/III-B	62_AUMBAC_00003	۱۹۰	+/-	I-B/III-B	DFPST0006	۸۸
۸۹	III-B	2015006204-DFPST	۲۹۳	+	I-B	CDC61126	۱۹۱	+	I-B	DFPST005	۸۹
۹۰	III-B	DFPST0026	۲۹۴	+	I-B	3213087	۱۹۲	+	I-B	DFPST0017	۹۰
۹۱	III-B	KM90-1-8G-F	۲۹۵	+	I-B	3213029	۱۹۳	+	I-B	DFPST011	۹۱
۹۲	III-B	KM53	۲۹۶	+	I-B	3213041	۱۹۴	-	I-B	SBPRU-0556	۹۲
۹۳	III-B	KM57-1-TJ980B	۲۹۷	+	I-B	FE9508BPD	۱۹۵	+/+	I-B/III-B	3859 11	۹۳
۹۴	III-B	CC1-A-KM44-1	۲۹۸	-/+	I-B/III-B	FE9508BRB	۱۹۶	-	I-B	E-1331	۹۴
۹۵	III-B	CC4A-KM61-1	۲۹۹	-/+	I-B/III-B	PA9508B	۱۹۷	-	I-B	P559	۹۵
۹۶	III-B	ST7B	۳۰۰	-/-	I-D/II-C	LNC5-DC	۱۹۸	+	I-B	P81	۹۶
۹۷	III-B	90_AUMBAC_00005	۳۰۱	+/+	I-D/III-B	Stockholm	۱۹۹	+	I-B	H4	۹۷

ردیف	ردیف	ردیف	ردیف	ردیف	ردیف	ردیف	ردیف	ردیف	ردیف	ردیف	ردیف
۹۸	H41	I-B	+	۲۰۰	IZSVe-TV_4045_5_13	I-D	+	۳۰۲	56_AUMBAC_00002	III-B	+
۹۹	H49	I-B	-	۲۰۱	IZSVe-TV_7574_5_12	I-D	+	۳۰۳	TMW2.991	III-B	-
۱۰۰	H61	I-B	-	۲۰۲	IZSVe-TV_1585_19_11	I-D	-	۳۰۴	FE9508BPD	III-B	+
۱۰۱	P551	I-B	-	۲۰۳	IZSVe-TV_5262_6_14	I-D	+	۳۰۵	DFPST005	III-B	-
۱۰۲	H34	I-B	-	۲۰۴	IZSVe-TV_3859_5_11	I-D	+	۳۰۶	DFPST0017	III-B	-

ج) سیستم کریسپر نوع III که برای فعالیت به پروتئین‌های زیادی از جمله Cas2 و Cas6 نیاز داشته و به دو زیرشاخه (III-A و III-B) تقسیم شده است. در مطالعه‌ی قبلی انجام شده بر روی *C. botulinum* تنها زیرگروه‌های I-B و III-B در ۱۰۰ سویهٔ این باکتری شناسایی شد (Negahdaripour et al. 2017).

شایع‌ترین زیرگروه آرایه‌های CRISPR-Cas در سویه‌های *C. botulinum* در پژوهش حاضر نیز همین دو زیر گروه هستند.

#### تنوع ساختار پروتئینی در سیستم‌های CRISPR-Cas

بر اساس بررسی‌های صورت گرفته، پروتئین‌های ساختاری و عملکردی تشکیل‌دهندهٔ سیستم‌های CRISPR-Cas از نظر ترکیب پروتئینی، ترتیب قرارگیری پروتئین‌ها و نیز جهت قرارگیری برای ایجاد یک واحد عملکردی CRISPR-Cas با هم متفاوت بودند. ژن *Cas* نوع مکان CRISPR/Cas را در هر آرایه تعیین می‌کند

در بین جدایه‌ها ۱۲۰ مورد (۳۹/۲۱ درصد) در زیرگروه I-B، پنج مورد (۱/۶۳ درصد) در زیرگروه II-C، ۱۷ مورد (۵/۵۶ درصد) در زیرگروه I-D، پنج مورد (۱/۶۳ درصد) در زیرگروه III-D و ۸۳ مورد (۲۷/۱۲ درصد) در زیرگروه III-B قرار گرفتند. همان‌طوری‌که در جدول ۱ دیده می‌شود نزدیک به یک پنجم از سویه‌های *C. botulinum* بررسی شده به‌صورت هم‌زمان دو یا چند نوع آرایهٔ CRISPR-Cas در ژنوم خود داشتند. سه نوع سیستم CRISPR باکتریایی شناسایی شده تا کنون، عبارتند از الف) سیستم کریسپر نوع I که برای برش از پروتئین‌های متعدد مانند ریونوکلئوپروتئین‌های Cas1، Cas5 و Cas6 استفاده می‌کند و در حال حاضر به هفت زیر شاخه (I-A، I-B، I-C، I-D، I-E، I-F و I-U) تقسیم‌بندی شده است. ب) سیستم کریسپر نوع II نیز برای فعالیت به پروتئین‌های کمتری از جمله Cas9 نیاز داشته و به سه زیرشاخه (II-A، II-B و II-C) تقسیم شده است.

## "شیخ الاسلامی و همکاران، بررسی تنوع بررسی تنوع سیستم‌های CRISPR-Cas در باکتری ..."

III-D و III-B مشاهده نشد (شکل ۱). مجموعه ژن‌های سازگاری در آرایه‌های CRISPR-Cas به باکتری‌ها اجازه می‌دهد تا با متصل شدن به قطعات دی.ان.ای خارجی، نسبت به عناصر ژنتیکی خاصی مصونیت پیدا کنند (Faure et al. 2019). به علاوه، تجزیه و تحلیل ژنومی سویه‌های *C. botulinum* انواع متنوعی از ماژول‌های تداخل (interference module) را شناسایی کرد. این ژن‌ها در زیرگروه I-B، پروتئین‌های Cas3، Cas5، Cas7 و Cas8 را تولید می‌کردند که Cas5 در همه سویه‌ها وجود داشت. آرایه‌های سویه‌های نوع I-D فقط Cas3 را در ماژول تداخل در خود جای دادند و در زیرگروه II-D، همه‌ی سویه‌ها Cas9 را به‌عنوان جزء اصلی ماژول تداخل نشان دادند. همچنین، در زیرگروه‌های III-B و III-D، به جز سویه‌های A2B387 و CCUG7968، همه باکتری‌ها حاوی Cas10 به‌عنوان جزء اصلی ماژول تداخل برای این زیرگروه بودند (شکل ۱). مشخص شده است که ماژول تداخل به شناسایی و برش قطعات ژنومی هدف کمک می‌کنند (Zhang et al. 2020).

ماژول تداخل آرایه‌های CRISPR-Cas کلاس I غالباً از ژن‌های سازنده پروتئین‌های Cas3، Cas5، Cas7 و Cas8، Cas3HD تشکیل شده است (Bhatia & Yadav, 2023) و ژن‌های ماژول تداخل در آرایه‌های کلاس II نیز معمولاً شامل

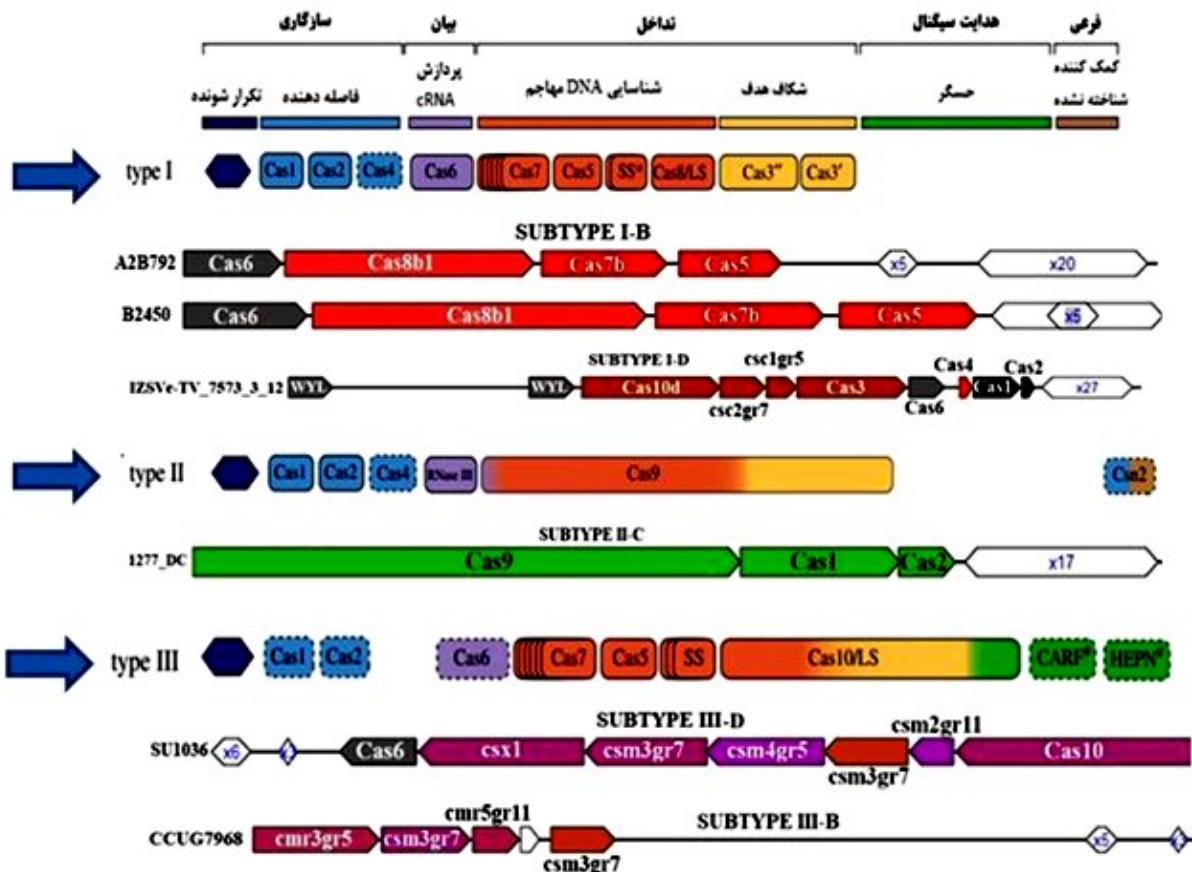
و شناسایی یک پروتئین Cas زمانی مطمئن است که حداقل دارای سه ژن *Cas* سازگار باشد (به جز نوع V) زیرا برای نوع V تنها یک ژن منفرد *cpf1* گزارش شده است (Makarova et al. 2011).

ارزیابی ما نشان داد (شکل ۱) که مجموعه ژن‌های بیان‌کننده پروتئین‌های Cas (expression module) در آرایه‌های CRISPR-Cas متعلق به پروتئین Cas 6 است که تقریباً در تمام سیستم‌های پیش‌بینی شده وجود دارد و اهمیت عملکردی Cas6 را در مدولاسیون ایمنی سویه‌های *C. botulinum* برجسته می‌کند. Cas 6 جزء اصلی ماژول بیان در آرایه‌های CRISPR-Cas باکتریایی است (Makarova et al. 2020). ژن‌های بیان‌کننده، رونویسی، پردازش، و بارگذاری موثر crRNA بر روی پروتئین‌های Cas را تضمین کرده و سیستم CRISPR-Cas را قادر می‌سازد تا به‌طور موثر عناصر ژنتیکی خارجی را شناسایی کرده و در برابر آن‌ها دفاع کند (Munawar and Ahmad, 2021).

یافته‌ها همچنین نشان دادند که مجموعه ژن‌های سازگاری (adaptation module) در سیستم‌های کریسپر زیرگروه‌های I-D و I-B سویه‌های *C. botulinum* شامل ژن‌های تولید پروتئین Cas1، Cas2 و Cas4 است، درحالی‌که در زیرگروه II-C این پروتئین‌ها شامل Cas1 و Cas2 است. همچنین، هیچ ماژول سازگاری در زیرگروه‌های

ماژول تداخل در آرایه‌های CRISPR-Cas کلاس III باکتری‌ها هستند (Tamulaitis et al. 2017).

ژن‌های ایجادکننده پروتئین Cas9 است (Makarova et al. 2020). همچنین، پروتئین‌های Cas5، Cas7 و Cas10 هم اغلب محصول ژن‌های



شکل ۱- شماتیک آرایه CRISPR-Cas کامل متعلق به زیرگروه‌های I-B، I-D، II-C، III-B و III-D شناسایی شده در باکتری *C. botulinum*

### توالی‌های فاصله‌انداز و تکرارها

یک چهارم سوم (third quartile) و بیشترین مقدار (upper whisker) برای این توالی‌ها نشان داده شد (جدول ۲، ۳ و ۴). گزارش شده است که طول فاصله‌اندازها و تکرارها در سیستم‌های CRISPR-Cas نقش مهمی در تعیین ویژگی، پایداری، کارایی و پویایی تکاملی پاسخ ایمنی

بررسی توالی‌های فاصله‌انداز و تکرارها در سیستم‌های کامل CRISPR-Cas در سویه‌های *C. botulinum* انجام و خلاصه‌ی پنج ویژگی عددی شامل کمترین مقدار (lower whisker)، یک چهارم اول (first quartile)، میانه (median)،

## "شیخ الاسلامی و همکاران، بررسی تنوع بررسی تنوع سیستم‌های CRISPR-Cas در باکتری ..."

سویه‌های *C. botulinum* شناسایی شده نشان داد که زیرگروه‌های I-D و II-C دارای میانگین طول تکرار ۳۶ جفت باز بودند، در حالی که زیرگروه‌های I-B، III-D، III-B و III-B میانگین طول تکرار ۳۰ جفت باز را نشان دادند (جدول ۴). ثابت شده است که توالی‌های تکرارهای طولانی‌تر، ادغام فاصله‌اندازهای جدید در آرایه CRISPR را تسهیل کرده و بنابراین منجر به کارایی بالاتر در اکتساب ایمنی جدید باکتری میزبان در برابر عناصر ژنتیکی مهاجم می‌شود (Heler et al. 2017).

علاوه بر این، توالی‌های طولانی به پایداری ساختاری آرایه CRISPR کمک می‌کند. این ثابت برای حفظ یکپارچگی آرایه در طول زمان و در نسل‌های متعدد سلول‌های باکتری میزبان مهم است (Levy et al. 2015).

ایجاد شده توسط این سیستم‌ها ایفا می‌کنند (Nussenzweig and Marraffini, 2020). همانطوری که در جدول ۲ نشان داده شده است، در زیرگروه‌های I-B، I-D، III-D، III-B و III-B طول متوسط فاصله‌اندازها ۳۵ جفت باز بود، در حالی که در آرایه نوع II-C طول ۳۰ جفت باز مشاهده شد. همچنین مشخص شد بالاترین تعداد توالی‌های فاصله‌انداز سیستم‌های CRISPR، در زیرگروه II-C (با میانگین ۱۷ فاصله‌انداز) وجود دارد (جدول ۳). فاصله‌اندازهای طولانی‌تر به دلیل توانایی بیشتر در هدف قرار دادن عناصر ژنتیکی، ایمنی بیشتری را ایجاد می‌کنند (Heler et al. 2017). از سوی دیگر، فاصله‌اندازهای کوتاه‌تر ممکن است با هدف‌گیری کمتر اختصاصی، خطر برش ناخواسته دی.ان.ای میزبان را افزایش دهند (Doench et al. 2014).

مقایسه میانگین طول توالی‌های تکرار در میان پنج زیرگروه سیستم‌های CRISPR-Cas در

جدول ۲- تنوع طول فاصله‌اندازها در جایگاه‌های CRISPR شناسایی شده در سویه‌های *C. botulinum*

نوع سیستم CRISPR	بیشترین مقدار	یک چهارم سوم	میان	یک چهارم اول	کمترین مقدار
I-B	۳۱	۳۰	۳۰	۲۹	۲۸
I-D	۳۶	۳۶	۳۶	۳۶	۳۶
II-C	۳۶	۳۶	۳۶	۳۶	۳۶
III-D	۳۱	۳۰	۳۰	۲۹	۲۸
III-B	۳۴	۳۲	۳۰	۳۰	۲۸

جدول ۳- تنوع تعداد فاصله اندازها در جایگاه های CRISPR شناسایی شده در سویه های *C. botulinum*

نوع سیستم CRISPR	بیشترین مقدار	یک چهارم سوم	میانه	یک چهارم اول	کمترین مقدار
I-B	۳۴	۱۹	۱۱	۷	۱
I-D	۱۲	۱۱	۱۱	۶	۳
II-C	۱۷	۱۷	۱۷	۱۷	۱۷
III-D	۶	۶	۳/۵	۳	۳
III-B	۱۱	۷	۶	۴	۳

جدول ۴- تنوع طول تکرارهای حفاظت شده در جایگاه های CRISPR شناسایی شده در *C. botulinum*

نوع سیستم CRISPR	بیشترین مقدار	یک چهارم سوم	میانه	یک چهارم اول	کمترین مقدار
I-B	۳۷	۳۶	۳۵	۳۵	۳۴
I-D	۳۶	۳۶	۳۵	۳۵	۳۵
II-C	۳۰	۳۰	۳۰	۳۰	۳۰
III-D	۳۵	۳۵	۳۵	۳۵	۳۵
III-B	۳۷	۳۵	۳۵	۳۱	۲۶

### پیش بینی ساختار ثانویه تکرارهای مستقیم

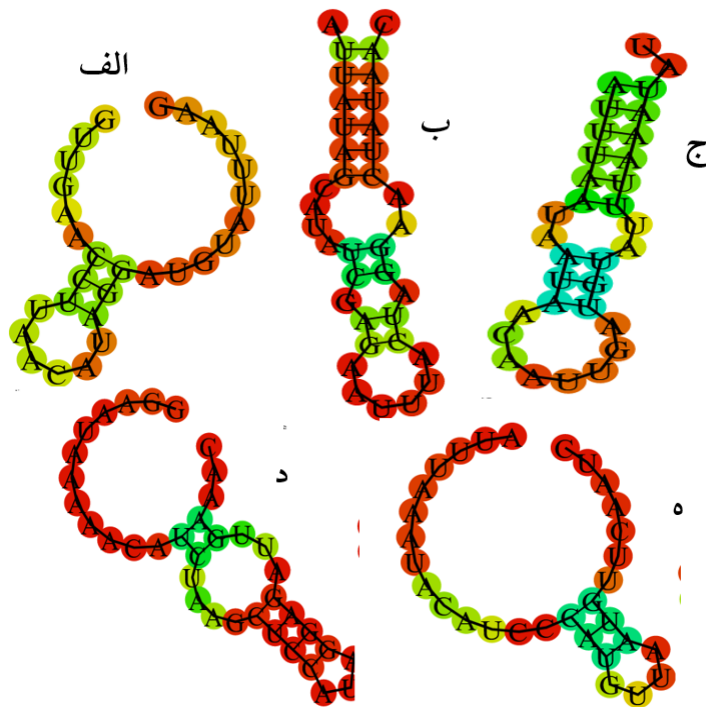
در شکل ۲ ساختار مربوط به هر زیر گروه آرایه CRISPR/Cas باکتری های *C. botulinum* نشان داده شده است و همانطوری که مشاهده می شود، ساختارهای ثانویه آر.ان.ای متنوع بوده و همگی ساقه هایی را در وسط تشکیل می دهند. بر اساس پارامترهای پیش فرض سیستم، در هر ساختار رنگ های قرمز و سبز به ترتیب نشان دهنده ی نقاط با احتمال بالا و پایین کانفورماسیون هستند. توالی های تکرار مستقیم در جایگاه های CRISPR مرتبط با ژن های *Cas* زیرگروه های I-D و I-B، به ترتیب طول ساقه های ۵ و ۶ جفت باز با حداقل انرژی آزاد ۰/۲۰- و ۴/۷۳- کیلوکالری بر مول را

نشان دادند، درحالی که در زیرگروه II-C، طول ساقه ۷ جفت باز با حداقل انرژی آزاد ۰/۱۰- کیلوکالری بر مول مشاهده شد. این پیش بینی نشان می دهد که زیرگروه II-C، با ساقه های بلندتر در مقایسه با زیرگروه های I-D و I-B، تمایل به تشکیل ساختارهای ثانویه آر.ان.ای پایدارتری دارد. کوتاه ترین ساقه های ساختار ثانویه تکرارهای مستقیم با ۴ جفت باز (با حداقل انرژی آزاد ۲/۰۰- کیلوکالری بر مول) نیز در زیرگروه III-B از آرایه های CRISPR/Cas سویه های *C. botulinum* مطالعه شده وجود داشت (شکل ۲). تجزیه ساختار ثانویه تکرارهای مستقیم، اطلاعاتی در مورد سیستم CRISPR/Cas به دست

## "شیخ الاسلامی و همکاران، بررسی تنوع سیستم‌های CRISPR-Cas در باکتری ..."

ساختار ثانویه آر.ان.ا پایدار، همیشه یک ارزش انرژی منفی بر حسب کیلوکالری بر مول دارد (Ghaffarian and Panahi, 2023). اثبات شده است ساختارهایی با محتوای GC بیشتر و ساقه (stem) بلندتر پایدارتر از ساختارهایی با محتوای GC کمتر و طول کوتاه‌تر هستند و همچنین اثبات شده است ساختارهای ثانویه با محتوای بازهای گوانین و سیتوزین بیشتر و ساقه بلندتر، پایدارتر از ساختارهایی با محتوای GC کمتر و طول کوتاه‌تر هستند (Yang et al. 2020).

می‌دهد به طوری که ساختار مناسب که منجر به آزاد شدن crRNAهای منحصر به فرد می‌شود، برای شناسایی و از بین بردن اهداف ژنومی توسط پروتئین‌های Cas میزبان، بسیار مهم است (O'Connell. 2019). پیوندهایی که بین جفت بازهای CG و AU در ساختار سه بعدی مولکول آر.ان.ا وجود دارد باعث پایداری آن شده و این مفهوم به عنوان مدل پایین‌ترین سطح انرژی نامیده می‌شود چرا که برای باز کردن آن باید مقداری انرژی (به صورت گرما) به کار برد. بنابراین هر



شکل ۲- ساختارهای ثانویه توالی‌های تکرار در آرایه‌های CRISPR باکتری *C. botulinum*. الف) آرایه نوع III-B متعلق به سویه CDC\_1436 با انرژی  $-2/00$  کیلوکالری بر مول، ب) آرایه نوع II-C متعلق به سویه DC\_1277 با انرژی  $-1/00$  کیلوکالری بر مول، ج) آرایه نوع I-B متعلق به سویه DFPST010 با انرژی  $-0/20$  کیلوکالری بر مول، د) آرایه نوع I-D متعلق به سویه IZSVe-TV\_7573\_4\_12 با انرژی  $-4/73$  کیلوکالری بر مول و ه) آرایه نوع III-D متعلق به سویه B515 با انرژی  $-0/10$  کیلوکالری بر مول.

پژوهش بینشی عمیق از تنوع و عملکرد سیستم CRISPR/Cas در باکتری *C. botulinum* ارائه می‌کند و اطلاعات پایه‌ای با ارزشی را برای مطالعات پیش رو فراهم می‌کند، به‌طوریکه تنوع و ویژگی سیستم‌های CRISPR-Cas، می‌تواند درک ما از ژنتیک میکروبی و توسعه ابزارهای مهندسی ژنتیکی و راهکارهای کنترل زیستی علیه این باکتری را افزایش دهد.

در مجموع، مطالعه حاضر تنوع قابل‌توجه‌ای از سیستم‌های CRISPR-Cas را در سویه‌های *C. botulinum* نشان داد که شامل پنج زیرگروه مجزا است. این تنوع نقش طبیعی این سیستم‌ها را به‌عنوان دفاع ایمنی فعال در برابر فاژها و دی.ان.ا. خارجی برجسته می‌کند. بین گونه‌های مورد مطالعه در تعداد و توالی فاصله‌اندازها و توالی‌های تکرارهای حفاظت شده، همچنین ساختار ثانویه تکرارهای مستقیم تنوع بالایی مشاهده شد. این

## References

## فهرست منابع

- Andersen JM, Shoup M, Robinson C, Britton R, Olsen KE, Barrangou R. 2016. CRISPR diversity and microevolution in *Clostridium difficile*. *Genome biology and evolution* 8(9): 2841-2855.
- Bhatia S, Yadav SK. 2023. CRISPR-Cas for genome editing: Classification, mechanism, designing and applications. *International Journal of Biological Macromolecules*. 124054.
- Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature biotechnology*. 32(12): 1262-1267.
- Faure G, Makarova KS, Koonin EV. 2019. CRISPR-Cas: complex functional networks and multiple roles beyond adaptive immunity. *Journal of molecular biology*. 431(1): 3-20.
- Ghaffarian S, Panahi B. 2023. Investigating the diversity of phage and plasmid targets of the CRISPR/Cas systems in the *Leuconostoc* genus using the insilico approach. *Journal of Biosafety*. 16(1). (In Farsi with English abstract)
- Hashempour-Baltork F, Hosseini H, Shojaee-Aliabadi S, Torbati M, Alizadeh AM, Alizadeh M. 2019. Drug resistance and the prevention strategies in food borne bacteria: An update review. *Advanced pharmaceutical bulletin*. 9(3): 335.
- Heler R, Wright AV, Vucelja M, Bikard D, Doudna JA, Marraffini LA. 2017. Mutations in Cas9 enhance the rate of acquisition of viral spacer sequences during the CRISPR-Cas immune response. *Molecular cell*. 65(1): 168-175.
- Hill KK, Xie G, Foley BT, Smith TJ. 2015. Genetic diversity within the botulinum neurotoxin-producing bacteria and their neurotoxins. *Toxicon*. 107: 2-8.
- Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic acids research*. 31(13): 3429-3431.
- Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology*. 31(3): 233-239.

"شیخ الاسلامی و همکاران، بررسی تنوع بررسی تنوع سیستم‌های CRISPR-Cas در باکتری ..."

**Karvelis T, Gasiunas G, Miksys A, Barrangou R, Horvath P, Siksnys V. 2013.** crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. RNA biology. 10(5): 841-851.

**Kuijper EJ, van den Berg RJ, Brazier JS. 2009.** Comparison of molecular typing methods applied to *Clostridium difficile*. Molecular Epidemiology of Microorganisms: Methods and Protocols. 159-171.

**Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G, Edgar R, Qimron U, Sorek R. 2015.** CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature. 520(7548): 505-510.

**Li W, Godzik A. 2006.** Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22(13): 1658-1659.

**Lindström M, Myllykoski J, Sivelä S, Korkeala H. 2010.** *Clostridium botulinum* in cattle and dairy products. Critical reviews in food science and nutrition. 50(4): 281-304.

**Lino CA, Harper JC, Carney JP, Timlin JA. 2018.** Delivering CRISPR: a review of the challenges and approaches. Drug delivery. 25(1): 1234-1257.

**Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJ, Charpentier E, Cheng D, Haft DH, Horvath P. 2020.** Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nature Reviews Microbiology. 18(2): 67-83.

**Makarova KS, Aravind L, Wolf YI, Koonin EV. (2011).** Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biology direct. 6: 1-27.

**Moller AG, Liang C. 2017.** MetaCRASST: reference-guided extraction of CRISPR spacers from unassembled metagenomes. PeerJ. 5: e3788.

**Munawar N, Ahmad A. 2021.** CRISPR/Cas system: an introduction. CRISPR Crops: The Future of Food Security. 1-35.

**Nami Y, Imeni N, Panahi B. 2021.** Application of machine learning in bacteriophage research. BMC microbiology. 21(1): 1-8.

**Nami Y, Rostampour M, Panahi B. 2023.** CRISPR-Cas systems and diversity of targeting phages in *Lactobacillus johnsonii* strains; insights from genome mining approach. Infection, Genetics and Evolution. 114: 105500.

**Negahdaripour M, Nezafat N, Hajighahramani N, Rahmatabadi SS, Ghasemi Y. 2017.** Investigating CRISPR-Cas systems in *Clostridium botulinum* via bioinformatics tools. Infection, Genetics and Evolution. 54: 355-373.

**Nussenzweig PM, Marraffini LA. 2020.** Molecular mechanisms of CRISPR-Cas immunity in bacteria. Annual review of genetics. 54: 93-120.

**O'Connell MR. 2019.** Molecular mechanisms of RNA targeting by Cas13-containing type VI CRISPR-Cas systems. Journal of molecular biology. 431(1): 66-87.

**Panahi B, Dehganzad B, Nami Y. 2023.** CRISPR-Cas systems feature and targeting phages diversity in *Lactocaseibacillus rhamnosus* strains. Frontiers in microbiology. 14: 1281307.

**Pursey E, Dimitriu T, Paganelli FL, Westra ER, van Houte S. 2022.** CRISPR-Cas is associated with fewer antibiotic resistance genes in bacterial pathogens. Philosophical Transactions of the Royal Society B. 377(1842): 20200464.

**Rossetto O, Pirazzini M, Montecucco C. 2014.** Botulinum neurotoxins: genetic, structural and mechanistic insights. Nature Reviews Microbiology. 12(8): 535-549.

**ShariatN., Dudley EG. 2014.** CRISPRs: molecular signatures used for pathogen subtyping. Applied and environmental microbiology. 80(2): 430-439.

**Skarin H, Håfström T, Westerberg J, Segerman B. 2011.** *Clostridium botulinum* group III: a group with dual identity shaped by plasmids, phages and mobile elements. BMC genomics. 12(1): 1-13.

**Smith TJ Hill KK, Raphael BH. 2015.** Historical and current perspectives on *Clostridium botulinum* diversity. *Research in Microbiology*. 166(4): 290-302.

**Tamulaitis G, Venclovas Č, Siksnyš V. 2017.** Type III CRISPR-Cas immunity: major differences brushed aside. *Trends in microbiology*. 25(1): 49-61.

**Yadav G, Singh R. 2022.** In silico analysis reveals the co-existence of CRISPR-Cas type I-F1 and type I-F2 systems and its association with restricted phage invasion in *Acinetobacter baumannii*. *Frontiers in microbiology*. 3079.

**Yang L, Li W, Ujiroghene OJ, Yang Y, Lu J, Zhang S, Pang X, Lv J. 2020.** Occurrence and diversity of CRISPR loci in *Lactobacillus casei* group. *Frontiers in microbiology*. 11: 624.

**Zhang H, Li Z, Xiao R, Chang L. 2020.** Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nature structural & molecular biology*. 27(11): 1069-1076.

**Zhang Q, Ye Y. 2017.** Not all predicted CRISPR-Cas systems are equal: isolated cas genes and classes of CRISPR like elements. *BMC bioinformatics*. 18: 1-12.

## Investigation of CRISPR-Cas Systems Diversity in *Clostridium botulinum* via Genome Mining Approach

**Naiymeh Sheikholeslami<sup>1</sup>, Hamid Mirzaei<sup>1,2</sup>, Jalili Khandaghi<sup>2,3</sup>, Yousef Nami<sup>4</sup>, Afshin Javadi<sup>1,2</sup>**

1- Department of Food Hygiene, Faculty of Veterinary Medicine, Tabriz Medical Sciences, Islamic Azad University, Tabriz, Iran.

2- Department of food Biotechnology, Biotechnology Research Center, Tabriz Branch, Islamic Azad University, Tabriz, Iran.

3- Department of Food Science and Technology, Sarab Branch, Islamic Azad University, Sarab, Iran.

4- Department of Food Biotechnology, Branch for Northwest & West region, Agricultural Biotechnology Research, Institute of Iran, Agricultural Research, Education and Extension Organization (AREEO), Tabriz, Iran.

hmirzaei@iaut.ac.ir

### Abstract

Clustered regularly interspaced short palindromic repeats and its associated protein or CRISPR-Cas systems are valuable genetic targets for bacterial phylogeny analysis in epidemiological studies. In this study, the genomic data of 800 sequences of *Clostridium botulinum* strains from the National Center for Biotechnology Information (NCBI) database was collected and the occurrence and variety of the complete CRISPR-Cas systems and its spacer and repeat sequences evaluated in the genomes recorded for these bacteria with the *in-silico* method. Also, the secondary structures of repeat sequences were predicted based on the minimum free energy (MFE) algorithm. According to the results, from the 526 genomes with complete CRISPR-Cas array, 306 strains (58%) had *Cas* genes. Subtypes I-B, I-D, II-C, III-D, and III-B were identified in these arrays. The findings also revealed that Cas 6 is the main protein produced by expression module genes in the studied CRISPR-Cas systems, and diverse *Cas* proteins are produced by the adaptation and interference modules in these arrays. The average length of spacer and repeat sequences in CRISPR arrays was 30-35 and 30-36 bp, respectively. Moreover, the prediction of secondary structures of repeat sequences showed that subtype II-C, with longer stems than other subtypes, tends to form more stable RNA secondary structures.

**Keywords:** *Clostridium botulinum*, CRISPR-Cas, Genome Mining, Repeat Sequence, Spacer Sequence.