



دانشگاه علوم کشاورزی و منابع طبیعی گorgan

نشریه پژوهش در نشخوارکنندگان

جلد چهارم، شماره دوم، ۱۳۹۵

<http://ejrr.gau.ac.ir>

بررسی انباشت اطلاعات در توالی DNA اشرشیاکولی موثر در بیماری ورم پستان در گاو شیری با استفاده از نظریه اطلاعات

*مصطفی قادری زفره‌ایی^۱، علی بندی‌دستجردی^۱، محمدرضا بحرینی بهزادی^۱

فرهاد صمدیان کرج‌آباد^۱ و مهرداد معمار^۱

^۱استادیار و دانش‌آموخته کارشناسی ارشد، گروه علوم دامی، دانشکده کشاورزی و منابع طبیعی، دانشگاه یاسوج

تاریخ دریافت: ۹۵/۳/۱۹؛ تاریخ پذیرش: ۹۵/۶/۲۲

چکیده

سابقه و هدف: بیماری ورم پستان هزینه‌های زیادی را سالانه به واحدهای پرورش گاو شیری در کشورهای مختلف وارد می‌کند. ورم پستان ناشی از باکتری اشرشیاکولی، در گاوهای با تولید شیر بالا شایع بوده که همراه با تعداد کم سلول‌های بدنی در شیر است. شدت و تأثیر ورم پستان ناشی از باکتری اشرشیاکولی در بین گاوهای درون یک گله و همچنین در بین مراحل مختلف شیرواری یک گله متفاوت است. حالت شدید ورم پستان ناشی از باکتری اشرشیاکولی با کاهش میزان تولید شیر همراه است که می‌تواند منجر به مرگ گاو شیری دارای بیماری نیز شود. از آنجایی که باکتری اشرشیاکولی مهمترین نقش را در ایجاد و گسترش بیماری ورم پستان دارد، در این تحقیق سعی شد که توالی‌های قطعات DNA این باکتری که در ایجاد ورم پستان نقش دارند، از پایگاه بانک جهانی ژن استخراج و با استفاده از نظریه اطلاعات مورد بررسی قرار گیرند.

مواد و روش‌ها: در این مطالعه، ابتدا توالی‌های DNA باکتری اشرشیاکولی مرتبط با بیماری ورم پستان در گاو شیری از پایگاه بانک جهانی ژن (NCBI) انتخاب و سپس این توالی‌ها با فرمت FASTA ذخیره شدند. توالی‌های قطعه‌های DNA با یکدیگر همتراز شدند. میزان آنتروپی، اطلاعات متقابل بین نوکلئوتیدها و همچنین فاصله بین قطعه‌های DNA با استفاده از نرم‌افزار متلب محاسبه و مورد بررسی قرار گرفت.

*نویسنده مسئول: mghaderi@yu.ac.ir

یافته‌ها: نتایج پژوهش حاضر نشان داد که آنتروپی توالی‌های قطعات DNA این باکتری متفاوت بوده اما تقریباً همه آن‌ها از میزان زیادی آنتروپی برخوردار بودند ($H(x) > 1.900$). قدر مطلق فاصله کولبک - لایبلر نشان داد که توالی‌های DNA، نامشابه بودند. بنابراین چنین استنباط شد که توالی‌های DNA یاد شده در شبکه‌های متفاوت متابولیکی نقش دارند و لذا دست‌ورزی آن‌ها - نظیر خاموش کردن بیان آن‌ها می‌تواند اثرات پیش‌بینی نشده‌ایی را ایجاد کند. نتایج حاصل از اطلاعات متقابل نشان داد که نوکلئوتیدهای موجود در قطعات مختلف DNA دارای میزان ارتباط مختلفی هستند. این ارتباط غیر یکسان بین قطعات مختلف DNA می‌تواند نوعی از درجه‌ی عدم تعادل پیوستگی را بین بازهای DNA نشان دهد.

نتیجه‌گیری کلی: در یک نتیجه‌گیری کلی می‌توان چنین بیان کرد که برای مدیریت بیماری ورم پستان ناشی از باکتری اشرشیاکولی می‌توان آن دسته از قطعات DNA که بیشترین آنتروپی را دارند به عنوان اهداف زیست‌فن‌آوری و دارویی در نظر گرفت. به دلیل اینکه فرضیه اساسی و کلیدی این پژوهش آن بود که هر چه قدر یک قطعه DNA آنتروپی بیشتری در یک مجموعه ژنومی خاص را ایجاد کند، شاید نقش بیشتری در ایجاد ورم پستان در گاو شیری داشته باشد.

واژه‌های کلیدی: اشرشیاکولی، اطلاعات متقابل، آنتروپی، کولبک لایبلر، ورم پستان

مقدمه

ورم پستان از شایع‌ترین بیماری گاوها و چالش اصلی برای پرورش دهندگان و متخصصان اصلاح نژاد گاو شیری است (۳ و ۲۸). این بیماری به وسیله طیف وسیعی از عوامل بیماری‌زای باکتریایی و قارچی بوجود می‌آید (۱۷). بعضی از این عوامل عبارتند از استرپتوکوکوس آگالاکتیه^۱، استافیلوکوکوس ارتوس^۲، مایکروپلاسما بوایز^۳، کورینه‌باکتریوم بویس^۴، استرپتوکوکوس دیسگالاکتیه^۵، انتروکوکوس فیسالیس^۶، انتروکوکوس فیسیوم^۷، استرپتوکوکوس بوویس^۸، استرپتوکوکوس آبرایز^۹ و کولی‌فرم‌هایی مثل اشرشای کولی^{۱۰}، کلبسیلا پنومونا^{۱۱} و انتروباکتر اروژنیز^{۱۲} می‌باشند (۱). باکتری اشرشایکولی از اعضای شاخص خانواده انتروباکتریاسه و جزء فلور طبیعی دستگاه گوارش انسان و حیوان بوده و در آب و خاک یافت می‌شود. اشرشایکولی بیماری‌زا با منشأ گاوی شامل سه گونه اشرشایکولی انروتوکسی‌ژنیک^{۱۳}، وروتوکسی‌ژنیک^{۱۴} و نکروتوکسی‌ژنیک^{۱۵} است (۸). در گله‌هایی با تعداد سلول‌های بدنی^{۱۶} کم در شیر، میزان شیوع بیماری ورم پستان بالینی در یک سال بین ۴۵ تا ۵۰ درصد است (۲۳). به نظر می‌رسد که در ایران، ورم پستان تحت بالینی در کنار بیماری‌های تولیدمثلی، لنگش و احتمالاً برخی بیماری‌های شایع دیگر نظیر لوکوز و یون از مهم‌ترین و خسارت‌بارترین بیماری‌هایی باشد که گله‌های گاو شیری را تهدید می‌کند (۱۳، ۱۸، ۱۹، ۲۰، ۲۵). لذا مدیریت این بیماری به صورت مستقیم یا غیر مستقیم، معقول به نظر می‌آید.

1. Streptococcus agalactiae
2. Staphylococcus aureus
3. Mycoplasma bovis
4. Corynebacterium bovis
5. Streptococcus dysgalactiae
6. Enterococcus faecalis
7. Enterococcus faecium
8. Streptococcus bovis
9. Streptococcus uberis
10. Escherichia coli
11. Klebsiella pneumonia
13. Varella Coelho
14. Malinowski
12. Enterobacter aerogenes
13. Entero toxigenic E.coli
14. Vero toxigenic E.coli
15. Necro toxigenic E.coli
16. Somatic cells

نظریه‌ی اطلاعات^۱ مدلی ریاضی از شرایط و عوامل مؤثر در انتقال و پردازش داده‌ها و اطلاعات است که در سال ۱۹۴۸ میلادی توسط شانون^۲ ارائه شد (۱۴). واژه آنروپی^۳ حاکی از تمایل سامانه‌ها به بی‌نظمی است. آنروپی، درهم ریختگی^۴، آشفتگی^۵، میزان تصادفی بودن یا میزان پیشامدی^۶ را گویند که در نظریه‌ی اطلاعات، به‌عنوان سنجی^۷ اطلاعات به کار می‌رود. نظریه اطلاعات به عنوان ابزاری مهم برای جستجوی الگوهایی در توالی‌های DNA و پروتئین استفاده شده است (۲۶). آنروپی شانون معیاری برای نظم و اختلال در توالی‌های DNA است (۱۴). هرزل و همکاران (۱۹۹۴) نشان دادند که آنروپی توالی‌های DNA نزدیک به بیشینه است (۱۱). شروین (۲۰۱۰) مزایای استفاده از اندازه‌گیری تنوع ژنتیکی مبتنی بر آنروپی در سطوح بیان ژن را نمایان ساخت و فرمول‌های ریاضی جهت محاسبه آنروپی و استفاده از آن جهت بررسی تنوع ژنی را ارائه کرد (۲۷). حبیبی و همکاران (۲۰۰۷) الگوریتم جدیدی برای تعیین ساختمان دوم پروتئین بر پایه‌های هندسی آنروپی ارائه نموده و نشان دادند که قطعه‌های دارای آنروپی پایین‌تر، نظم ساختاری بیشتری دارند (۱۰). با در دست بودن توالی قطعه‌های DNA اشرشیاکولی مؤثر بر ورم پستان این امکان وجود دارد که میزان اطلاعات ذخیره شده در بخش‌های مختلف آن‌ها را با استفاده از نظریه اطلاعات بررسی و با تفسیر زیستی آن‌ها، رهیافت جدیدی برای افزایش مقاومت به این بیماری ایجاد کرد. این پژوهش بر این فرض استوار شد که محاسبه‌ی میزان آنروپی توالی قطعه‌های DNA اشرشیاکولی مرتبط با ورم پستان، می‌تواند به طور غیرمستقیم به مدیریت این بیماری کمک کند. بنابراین هدف پژوهش حاضر، بررسی میزان آنروپی در توالی قطعه‌های DNA باکتری اشرشیاکولی وابسته به ورم پستان در گاو شیری و استخراج تاثیرگذارترین توالی قطعه‌های DNA این باکتری در وقوع ورم پستان بر اساس نظریه اطلاعات بود. بدین منظور میزان شباهت و تفاوت توالی قطعه‌های DNA باکتری اشرشیاکولی بر اساس معیار کولبک – لایبلر^۸ و محاسبه اطلاعات متقابل^۹ بین توالی قطعه‌های DNA مورد بررسی قرار گرفت. این‌گونه

۱. Information theory

۲. Shanon

۳. Entropy

۴. Clutter

۵ Disorderliness

۶ Shaffledness

۷ Messures

۸ Kulback – Leibler

۹ Mutual information

فرض شد که آن دسته از قطعات DNA اشرشیاکولی که از آنترپپی بالاتری برخوردار هستند، احتمالاً در وقوع ورم پستان نقش مهم‌تری خواهند داشت. از آنجائیکه تاکنون مطالعه‌ای از طریق کنکاش اطلاعات ژنوم باکتری اشرشیاکولی و بررسی میزان آنترپپی در توالی نوکلئوتیدهای DNA آن انجام نشده است، لذا انجام اینگونه پژوهش‌ها می‌تواند در مدیریت ورم پستان گاو شیری ارزشمند باشد.

مواد و روش‌ها

در این مطالعه، ابتدا توالی‌های DNA باکتری اشرشیاکولی مرتبط با بیماری ورم پستان در گاو شیری از پایگاه بانک ژن^۱ (NCBI, <http://www.ncbi.nlm.nih.gov>) انتخاب و سپس این توالی‌ها با فرمت FASTA ذخیره شدند. به دلیل در دسترس نبودن توالی مشابه از روش جستجوی کلمه کلیدی^۲ در پایگاه NCBI استفاده شد. همچنین اطلاعات این ژن‌ها از جمله شماره دست‌یابی یا شاخص کانتینگ ژن^۳، طول توالی ژنومی، طول توالی پروتئینی و طول توالی کدکننده^۴ از این پایگاه گرفته شدند. در این راستا توالی ژنومی و پروتئینی برای ۲۰ قطعه DNA استخراج شد (جدول ۱). به علت وجود توالی‌های ناحیه غیرترجمه‌ای^۵ در برخی از توالی‌های پروکاریوتیک، توالی ناحیه کدکننده این ۲۰ قطعه DNA نیز استخراج شد. به منظور تسهیل محاسبات، توالی‌هایی که هم‌اندازه بودند استخراج شدند تا همترازی یا هم‌ردیفی^۶ راحت‌تر انجام شود.

متناسب با هر قطعه DNA استخراج شده، یک توالی DNA هم‌اندازه به طور کاملاً تصادفی و بر اساس اصل استقلال مکان هر نوکلوتید، ایجاد شد. سپس توالی ۲۰ قطعه DNA با استفاده از نرم افزار MEGA هم‌تراز شدند. محاسبات دیگر به صورت زیر انجام شد.

الف) محاسبه آنترپپی برای هر قطعه DNA

$$H(X) = - \sum_{x \in X} p(x) \log_b p(x)$$

۱. Gene bank
۲. Key world search
۳. Conting Gene Index
۴. Coding Sequence
۵. Untranscribed region
۶. Alignment

در این محاسبه فرض شد که مقدار $b = 2$ است، $x \in \{A, T, C, G\}$ و $p(x)$ به عنوان تابع جرم احتمال^۱ DNA در نظر گرفته شد. $H(X)$ نیز مقدار آنتروپی است که واحد اندازه‌گیری آن بیت می‌باشد. از نظر ریاضی $H(X)$ به صورت $H(X) = E[-\log_b p(x)] = E[I_H(x)]$ قابل محاسبه است، که در آن $I_H(x)$ به مفهوم محتوی اطلاعات^۲ است و ثابت می‌شود که $I_H(x) = -\log_b p(x)$ است. (۶)

(ب) فاصله کولبک - لایبیلر (آنتروپی نسبی)

$$D(p||q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} = - \sum_{x \in X} p(x) \log \frac{q(x)}{p(x)} = -E \left[\log \frac{q(x)}{p(x)} \right]$$

$$\geq D(p||q) = -\log \left(E \left[\frac{q(x)}{p(x)} \right] \right) = -\log \left(\sum_{x \in X} p(x) \log \frac{q(x)}{p(x)} \right) = 0$$

از یک نگاه $D(p||q)$ عبارت از مقایسه آنتروپی بین دو توزیع برای یک متغیر تصادفی است و $q(x)$ تعریفی مانند $p(x)$ را دارد که روی قطعه دیگر DNA محاسبه می‌شود. بنابراین این معیار قطعه‌های DNA را دو به دو مقایسه و فاصله بین آن‌ها را اندازه‌گیری می‌کند.

(ج) اطلاعات متقابل بین قطعات DNA

$$I(X; Y) = D(p(x, y) || p(x)p(y)) = \sum_{\substack{x \in X \\ y \in Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

$$= \sum_{\substack{x \in X \\ y \in Y}} p(x, y) \log p(x, y) - \sum_{\substack{x \in X \\ y \in Y}} p(x, y) \log p(x) - \sum_{\substack{x \in X \\ y \in Y}} p(x, y) \log p(y)$$

$$= -H(X, Y) + H(X) + H(Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

برای محاسبه اطلاعات متقابل، از ماتریس همتراز شده قطعات DNA استفاده شد. در محاسبه فوق امکان ایجاد شکاف^۳ احتمالی در ماتریس همتراز شده نیز منظور شد. همانطور که مشاهده می‌شود، اطلاعات متقابل بین بازهای مختلف DNA در واقع اختلاف بین آنتروپی مطلق و آنتروپی شرطی است. تمامی محاسبات با استفاده از نرم افزار متلب^۴ نسخه ۹ انجام شد.

۱. Probability mass function

۲. Information content

۳. Gap

۴. Matlab

نشریه پژوهش در نشخوارکنندگان (۴)، شماره (۲) ۱۳۹۵

جدول ۱- اطلاعات مربوط به توالی‌های DNA اشرشیاکولی به کار رفته در این پژوهش (استخراج شده از NCBI)

Table 1. Information pertaining to Escherichia Coli DNA sequences used in this study(extracted from NCBI)

شماره	شاخص کانتیگ ژن	طول توالی پروتئین (اسید آمینه)	طول توالی کدکنده (نوکلئوتید)	طول توالی ژنومی (نوکلئوتید)
1	NZ_JTFD01000048.1 Gene 01	467	1404	1404
2	NZ_JTFD01000048.1 Gene 02	310	933	933
3	NZ_JTFD01000048.1 Gene 03	348	1047	1047
4	NZ_JTFD01000048.1 Gene 04	949	2850	2850
5	NZ_JTFD01000048.1 Gene 05	381	1146	1146
6	NZ_JTFD01000048.1 Gene 06	569	1710	1710
7	NZ_JTFD01000048.1 Gene 07	383	1152	1152
8	NZ_JTFD01000049.1 Gene 01	396	1191	1191
9	NZ_JTFD01000049.1 Gene 02	551	1656	1656
10	NZ_JTFD01000049.1 Gene 03	382	1149	1149
11	NZ_JTFD01000049.1 Gene 04	637	1914	1914
12	NZ_JTFD01000049.1 Gene 05	378	1137	1137
13	NZ_JTFD01000049.1 Gene 06	312	939	939
14	NZ_JTFD01000050.1 Gene 01	614	1845	1845
15	NZ_JTFD01000050.1 Gene 02	366	1101	1101
16	NZ_JTFD01000050.1 Gene 03	466	1401	1401
17	NZ_JTFD01000050.1 Gene 04	457	1374	1374
18	NZ_JTFD01000050.1 Gene 05	344	1005	1005
19	NZ_JTFD01000050.1 Gene 06	383	1152	1152
20	NZ_JTFD01000047.1 Gene 01	339	1020	1020

نتایج و بحث

آنترویی محاسبه شده برای تمامی قطعه‌های DNA، تعداد بازهای آلی و درصد نوکلئوتیدهای جفتی گوانین - سیتوزین در جدول ۲ نشان داده شده است.

جدول ۲- ترکیب نوکلئوتیدی قطعه‌های DNA به کار رفته و میزان آنروپی محاسبه شده آن‌ها

Table 2. Nucleotide composition of DNA Segments used in this study and their calculated entropies

شماره	کاتالوگ شماره	تعداد کل باز	(A) آدنین	(T) تیمین	(G) گوانین	(C) سیتوزین	%G-C	محتوای %G-C	آنروپی
1	NZ_JTFD01000048.1 Gene 01	1409	241	392	428	343	54.9	1.9695	
2	NZ_JTFD01000048.1 Gene 02	933	175	195	315	248	60.3	1.9619	
3	NZ_JTFD01000048.1 Gene 03	1047	226	231	284	306	56.4	1.9877	
4	NZ_JTFD01000048.1 Gene 04	2850	573	417	971	889	65.3	1.9480	
5	NZ_JTFD01000048.1 Gene 05	1146	275	243	334	294	54.8	1.9905	
6	NZ_JTFD01000048.1 Gene 06	1710	228	350	588	544	66.2	1.9112	
7	NZ_JTFD01000048.1 Gene 07	1152	197	206	367	382	65	1.9336	
8	NZ_JTFD01000049.1 Gene 01	1191	278	250	294	369	55.7	1.9847	
9	NZ_JTFD01000049.1 Gene 02	1656	454	289	461	452	55.1	1.9763	
10	NZ_JTFD01000049.1 Gene 03	1149	256	292	282	320	52.3	1.9540	
11	NZ_JTFD01000049.1 Gene 04	1914	463	420	494	537	53.9	1.9420	
12	NZ_JTFD01000049.1 Gene 05	1137	268	269	319	281	52.8	1.9630	
13	NZ_JTFD01000049.1 Gene 06	939	213	279	205	242	47.6	1.9892	
14	NZ_JTFD01000050.1 Gene 01	1845	465	453	437	490	50.2	1.9987	
15	NZ_JTFD01000050.1 Gene 02	1101	289	251	292	269	51	1.9984	
16	NZ_JTFD01000050.1 Gene 03	1401	355	320	358	368	51.8	1.9980	
17	NZ_JTFD01000050.1 Gene 04	1374	292	306	340	436	56.5	1.9814	
18	NZ_JTFD01000050.1 Gene 05	1005	241	226	260	278	53.5	1.9956	
19	NZ_JTFD01000050.1 Gene 06	1152	285	271	292	304	51.7	1.9988	
20	NZ_JTFD01000047.1 Gene 01	1020	235	290	258	237	48.5	1.9947	

از آنجایی که آنروپی تمام قطعه‌های DNA تقریباً به بیشینه آنروپی یعنی عدد ۲ نزدیک بود، می‌توان از مفاهیم نظریه اطلاعات به این نتیجه رسید که اطلاعات آنروپی قطعه‌های DNA تقریباً با هم برابر است. انتظار می‌رفت که میزان آنروپی قطعه‌های DNA به هم نزدیک باشند زیرا در استخراج این قطعه‌های DNA، سعی شد آن قسمتی از ژنوم اشرشیاکولی در نظر گرفته شود که بیشترین نقش را در ایجاد ورم پستان دارد. با این وجود تفاوت‌هایی در میزان آنروپی مشاهده شد. کمترین مقدار آنروپی مربوط به قطعه NZ_JTFD01000048.1|Gene 06 با $1/9112$ بیت و بیشترین مقدار مربوط به قطعه NZ_JTFD01000050.1|Gene 01 با مقدار $1/9987$ بیت بود. از آنجا که مقدار آنروپی هر قطعه DNA بین مقدار (صفر تا ۲) است و آنروپی همه قطعات DNA در این پژوهش نزدیک به ۲ بودند، استخراج مسیرهای زیستی و متابولیکی حاصل از این قطعه‌های DNA و نقش محصولات نهایی این مسیرها روی چگونگی ایجاد بیماری ورم پستان در گاو شیری باید مورد بررسی بیشتر قرار گیرند. کیم و همکاران (۲۰۰۹) از نظریه اطلاعات جهت آنالیز DNA کروموزوم ۱۶ مخمر استفاده نمودند

(۱۴). پژوهش آن‌ها نشان داد که ساختار دستوری^۱ آمینواسیدهای حاصل از کروموزوم شماره ۱۶ مخمر ارتباط عمیقی با ساختمان دوم پروتئین دارد (۱۴). رویز- مارین و همکاران (۲۰۱۰) از آزمون آنتروپی برای تجزیه و تحلیل پیوستگی ژنتیکی تک-لوکوس^۲ استفاده نمودند (۲۴). آن‌ها نشان دادند که انجام آزمون انتقال عدم تعادل^۳ مبتنی بر آنتروپی برای انجام مطالعات کل ژنوم امکان پذیر است (۲۴). بولشوی (۲۰۰۸) یک بررسی بر اساس آنتروپی شانون روی اطلاعات کروموزومی ۲۵ گونه از جمله پستانداران، ماهی‌ها، پرندگان، حشرات، نماتدها، قارچ‌ها و یک گیاه انجام داد و الگوریتمی ارائه داد که در این الگوریتم اطلاعات هر کروموزوم به یک توزیع دو بعدی^۴ تبدیل شد و نقشه‌های حاصل به وسیله آنتروپی مورد تجزیه و تحلیل قرار گرفتند (۲). نشان داده شد که به این وسیله اطلاعات کلی ژنومی^۵ را می‌توان پیش‌بینی و مجسم نمود و در نهایت می‌توان یک الگوریتم مقایسه ژنومی برای این گونه‌ها ایجاد کرد (۲). نیاگو و همکاران (۲۰۱۴) چند روش محاسبه آنتروپی را برای آنالیز نواحی از ژنوم گذشته ۶۰gp زیرگونه‌های انگل کریپتواسپوریوم^۶ به کار گرفتند (۲۲). با استفاده از تفاوت‌های حاصل از اندازه‌گیری با آنتروپی شانون- جنسن^۷ و رینی- جنسن^۸، داده‌های موجود این انگل را مورد مقایسه قرار دادند و مقایسات آماری به وسیله قطعه‌بندی DNA و تفاوت‌های آشکار شده توسط آنتروپی‌های رینی و شانون، به‌طور دقیق نواحی با تغییرپذیری با وضوح بالا (نواحی ریزماهورایی و نواحی با نرخ تغییرپذیری بالا) را نشان داد (۲۲). لی و وانگ (۲۰۰۵) با انجام بررسی کاربرد آنتروپی نسبی در توالی‌های DNA، یک بردار ۱۲ جزئی بر اساس نظریه اطلاعات برای توصیف توالی DNA جهت تمایز توالی پروتئینی کدگذار و غیرکدگذار در ژنوم مخمر ارائه دادند (۱۶). داماسویکیوس (۲۰۱۰) در مطالعه خود روی کروموزوم ۲۲ انسان، پیچیدگی توالی‌های DNA را با استفاده از نظریه اطلاعات و آنالیز فراوانی بررسی نمود (۷). در پژوهش یاد شده، پیچیدگی توالی‌های DNA با محاسبه آنتروپی شانون، آنتروپی رینی و پیچیدگی کولماگرافی نسبی^۹ برآورد شدند. این برآوردها توانست تصادفی بودن، واژگونی، تکرارپذیری و نواحی شبیه به هم (با پیچیدگی پایین) را تعیین نماید.

1. Grammar Structure
2. Single-Locus Genetic Association Analysis
3. Transmission Disequilibrium Test
4. Bidimensional
5. Global genomic information
6. Criptosporium
7. Shannon - Jentsin
8. Reny-Jentsin
9. Relative kolmogorov complexity

همچنین ساختار پیچیدگی بر اساس فراوانی با استفاده از تابع خود همبستگی^۱ و تقابل اطلاعات تاخیری^۲ مورد تجزیه و تحلیل قرار گرفت. نتایج نشان داد که هم نواحی کدشونده و هم نواحی غیرکدشونده، می‌توانند پیچیدگی پایین داشته باشند (۷). زای و همکاران (۲۰۱۰) با استفاده از اطلاعات پیچیدگی در توالی DNA و پروتئین ژن DNMT1 را مورد بررسی قرار دادند (۲۹). ژن DNMT1، یک ژن مسئول حفاظت متیل ترانسفراز برای کپی کردن الگوهای متیلاسیون DNA به رشته‌های خواهری در همانندسازی در گونه‌های مختلف است (۲۹). آنتروپی ژن DNMT1 در گونه‌های مختلف بر اساس وضعیت بازهای DNA محاسبه شد و نشان داده شد که پیچیدگی DNA در نواحی ایترون پایین‌تر از نواحی کد شده است. آن‌ها همچنین اثرات آنتروپی روی نواحی دامین و غیردامین ژن DNMT1 را اثبات نمودند. در این راستا بر اساس نتایج حاصل از توالی پروتئین‌ها و DNA نشان داده شد که تکامل DNA تمایل به سوی پیچیدگی دارد (۲۹). مونگ و کرسپو (۲۰۱۴) با اندازه‌گیری پیچیدگی DNA و تجزیه و تحلیل آن، مقادیر آنتروپی شانون، پیچیدگی کولموگراف^۳ - تخمین زده شده بر اساس فشردگی الگوریتم لمپل و زیو^۴ - به صورت آماری پیچیدگی آن‌ها را مقایسه کردند (۲۱). نتایج نشان داد که میزان پیچیدگی نواحی مربوط به ژن‌های مرتبط به هم از ژن‌های غیرمرتبط به هم متفاوت است (۲۱). از دیگر نتایج پژوهش مذکور، ارتباط بین پیچیدگی DNA و محتوای اطلاعات بود. به طوریکه با کاهش پیچیدگی DNA محتوای اطلاعات کمتری نیز جهت تجزیه و تحلیل در دسترس قرار می‌گرفت. لذا پیشنهاد شد که پژوهش‌ها در آینده به سمت مطالعه روی تنوع تک-نوکلئوتیدی تمرکز یابد (۲۱). در مطالعه‌ای نشان داده شده است که میانگین آنتروپی آگزون‌ها بیشتر از ۷۳ درصد و آنتروپی ایترون‌ها بیشتر از ۸۰ درصد است (۴). مشخص شده است که آنتروپی توالی ژن‌ها برای پروتئین‌ها در مقایسه با دیگر بخش‌های DNA بیشتر است و از آن جایی که آگزون و ایترون تحت تاثیر فرایندهای مختلف قرار می‌گیرند، آنتروپی آن‌ها نیز متفاوت است (۴ و ۱۵).

نتایج اطلاعات متقابل بین بازهای قطعه‌های DNA در جدول‌های ۳ و ۴ ارایه شده است. منظور از اطلاعات متقابل، میزان اشتراک اطلاعات دو منبع است. برای مثال اگر اطلاعات متقابل یک قطعه

1. Autocorrelation function
 2. Time delayed mutual information
 3. Kolmogorov Complexity
 4. Lempel Ziv

DNA با خودش را در نظر بگیریم. معلوم است که اطلاعات یک قطعه DNA کاملاً با اطلاعات خودش برابر است. در این صورت اطلاعات متقابل یک قطعه DNA با خودش حداکثر مقدار اطلاعات متقابل را دارد. از لحاظ عددی مقدار اطلاعات یک قطعه DNA بین صفر و ۲ است. در فرمول محاسبه اطلاعات متقابل، علاوه بر اینکه به تابع جرم احتمال هر یک از قطعات DNA نیاز داریم، به تابع توزیع جرم احتمال مشترک دو قطعه DNA نیز نیاز داریم. از آنجاییکه تابع احتمال مشترک دو قطعه DNA در دسترس نیست لذا برای محاسبه‌ی تابع توزیع جرم احتمال مشترک، روش‌های مختلفی ابداع شده است که هر کدام به صورت تجربی تابع توزیع جرم احتمال مشترک را محاسبه می‌کنند. در این پژوهش برای محاسبه‌ی تابع توزیع جرم احتمال مشترک، یک ماتریس تشکیل داده شد که بالای ردیف اول آن قطعه اول DNA و کنار ستون اول آن قطعه دوم DNA قرار داده شد که نتیجه آن یک ماتریس هم‌تراز بود. برای بدست آوردن مقدار هر درایه‌ی ماتریس به این صورت عمل شد که اگر دو باز بالا و سمت چپ آن درایه، که مربوط به دو قطعه DNA است با هم برابر بودند، آن درایه برابر مقدار ۱ و اگر نابرابر بودند برابر مقدار صفر در نظر گرفته شد. در پایان نیز تمام درایه‌ها بر کل تعداد درایه‌های ماتریس تقسیم شد. نتایج ابتدایی نشان داد که اگر به این صورت اطلاعات متقابل محاسبه شود، باید یک ماتریس بسیار بزرگ، که تعداد درایه‌هایش بسیار زیاد است را مورد استفاده قرار دهیم که با نرم افزارهای عادی قابل محاسبه نیست. این پژوهش نشان داد که به روش بسیار ساده‌تر و مناسب می‌شود اطلاعات متقابل را محاسبه نمود. در این روش بعد از هم‌ردیف کردن قطعات DNA، به ازای بازهای نامعلوم (شکاف‌ها که در اثر هم‌ردیفی ممکن است ایجاد شوند)، مقدار ۱/۴ قرار داده شد (زیرا هر یک از ۴ باز می‌تواند با احتمال ۱/۴ در آن شکاف قرار گیرد) و به ازای بازهای معلوم مقدار ۱ قرار داده شد و در نهایت این اعداد که برای تمام بازهای هر دو قطعه DNA هم‌تراز شده بدست می‌آید را با هم جمع کرده و بر تعداد کل بازها تقسیم شد تا درصد اطلاعات متقابل بدست آید. در این مطالعه برای ۲۰ قطعه DNA مورد بررسی، دامنه متفاوتی (۲-۱/۱) از اطلاعات متقابل بین نوکلئوتیدها به دست آمد. مقدار صفر نشان دهنده‌ی عدم ارتباط نوکلئوتیدها با یکدیگر بود و به عبارتی از لحاظ زیستی نیز این امکان وجود دارد که ارتباطی بین این دو نوکلئوتیدها وجود نداشته باشد. در نتیجه امکان دارد نوع پروتئینی که این نوکلئوتیدها در تولید آن دخالت دارند، از لحاظ زیستی نقش مشابه نداشته باشند. بیشینه مقدار اطلاعات متقابل محاسبه شده ۲ و کمترین آن ۱/۱ بود. مقدار اطلاعات متقابل برابر صفر را از لحاظ ژنتیک جمعیت و کمی به مفهوم نبود تعادل

پیوستگی^۱ تلقی کردیم. بنابراین به عنوان یک ایده جدید محاسبه اطلاعات متقابل می‌تواند ارزش بسیار بالایی در اصلاح نژاد داشته باشد. همچنین اساس تشخیص جایگاه‌های صفات کمی^۲ (QTL) بر پایه عدم تعادل پیوستگی استوار است. بنابراین هر چقدر اطلاعات متقابل بین نوکلئوتیدها بالاتر باشد، میزان نبود تعادل پیوستگی نیز بالاتر خواهد بود.

جدول ۳- برآورد اطلاعات متقابل قطعه‌های DNA باکتری اشرشیا کولی موثر در ایجاد ورم پستان

Table 3. Predicted mutual information of DNA segments in Escherichia Coli affecting mastitis

قطعه DNA	1	2	3	4	5	6	7	8	9	10
1	2	1.500	1.618	1.489	1.610	1.610	1.636	1.659	1.420	-
2	1.5	2.00	1.457	1.625	1.692	1.688	1.522	1.685	1.688	1.660
3	1.618	1.457	2.00	1.456	1.541	1.566	1.802	1.594	1.403	1.599
4	1.489	1.625	1.456	2.00	1.736	1.659	1.518	1.688	1.756	1.815
5	1.610	1.692	1.541	1.736	2	1.921	1.641	1.869	1.598	1.574
6	1.610	1.688	1.566	1.659	1.921	2.00	1.649	1.870	1.601	1.562
7	1.636	1.522	1.802	1.518	1.641	1.649	2.00	1.701	1.445	1.704
8	1.659	1.685	1.594	1.688	1.869	1.870	1.701	2.00	1.566	1.611
9	1.420	1.686	1.403	1.756	1.598	1.601	1.445	1.566	2.00	1.412
10	-	1.660	1.599	1.815	1.574	1.562	1.704	1.611	1.412	2.00
11	1.669	1.593	1.643	1.612	1.773	1.780	1.764	1.834	1.509	1.672
12	1.712	1.416	1.630	1.428	1.508	1.514	1.718	1.442	1.375	1.809
13	1.520	1.621	1.448	1.890	1.754	1.748	1.532	1.706	1.736	1.478
14	1.385	1.586	1.286	1.587	1.503	1.492	1.389	1.473	1.753	1.863
15	1.427	1.696	1.385	1.760	1.601	1.587	1.434	1.561	1.934	1.406
16	1.531	1.627	1.491	1.763	1.817	1.761	1.551	1.758	1.685	1.498
17	1.596	1.625	1.657	1.626	1.785	1.825	1.769	1.869	1.508	1.668
18	1.594	1.377	1.734	1.387	1.449	1.455	1.609	1.469	1.350	1.689
19	1.598	1.677	1.562	1.745	1.864	1.921	1.645	1.764	1.614	1.566
20	1.433	1.684	1.401	1.766	1.588	1.598	1.439	1.568	1.946	1.409

۱. Linkage disequilibrium

۲. Quantitative trait loci

جدول ۴- برآورد اطلاعات متقابل قطعه‌های DNA باکتری اشرشیا کولی موثر در ایجاد ورم پستان - ادامه

Table 4. Predicted mutual information of DNA segments in Escherichia Coli affecting mastitis-continued

قطعه DNA	11	12	13	14	15	16	17	18	19	20
1	1.669	1.712	1.520	1.385	1.427	1.531	1.596	1.594	1.598	1.423
2	1.593	1.416	1.621	1.586	1.696	1.627	1.625	1.377	1.677	1.684
3	1.643	1.630	1.448	1.286	1.385	1.491	1.657	1.734	1.562	1.401
4	1.612	1.428	1.890	1.587	1.760	1.763	1.626	1.387	1.745	1.766
5	1.773	1.508	1.754	1.503	1.601	1.817	1.782	1.449	1.864	1.588
6	1.780	1.514	1.748	1.492	1.587	1.761	1.825	1.455	1.921	1.598
7	1.764	1.718	1.532	1.389	1.434	1.551	1.769	1.609	1.645	1.439
8	1.834	1.442	1.706	1.473	1.561	1.758	1.869	1.469	1.764	1.568
9	1.509	1.375	1.736	1.753	1.934	1.685	1.508	1.350	1.614	1.964
10	1.672	1.809	1.478	1.863	1.406	1.498	1.668	1.689	1.566	1.409
11	2.00	1.584	1.617	1.439	1.503	1.635	1.864	1.521	1.785	1.611
12	1.584	2.00	1.434	1.341	1.377	1.456	1.569	1.815	1.507	1.375
13	1.617	1.434	2.00	1.604	1.725	1.870	1.632	1.395	1.751	1.718
14	1.439	1.341	1.604	2.00	1.732	1.566	1.441	1.321	1.508	1.387
15	1.503	1.377	1.725	1.732	2.00	1.667	1.520	1.374	1.603	1.934
16	1.635	1.456	1.870	1.566	1.667	2.00	1.678	-	1.819	1.685
17	1.864	1.569	1.632	1.441	1.520	1.678	2.00	1.515	1.791	1.521
18	1.521	1.815	1.395	1.321	1.374	-	1.515	2.00	1.446	1.347
19	1.785	1.507	1.751	1.808	1.603	1.819	1.719	1.446	2.00	1.611
20	1.611	1.375	1.718	1.387	1.934	1.685	1.521	1.347	1.611	2.00

فاصله کولبک-لایبلر (آنتروپی نسبی) بین ۲۰ قطعه DNA باکتری اشرشیاکولی ایجاد کننده ورم پستان در جداول ۵ و ۶ نشان داده شده است. ساختار نوکلئوتیدی مشابه تا حد زیادی فعالیت زیستی مشابه دو ژن یا دو قطعه DNA را بیان می‌کند. برای تعیین شباهت زیستی دو قطعه DNA، استخراج فاصله کولبک-لایبلر (آنتروپی نسبی) بین دو قطعه DNA لازم است. همانطور که در جدول ۵ نشان داده شده است فاصله کولبک-لایبلر به صورت یک ماتریس مربع نمایان می‌شود اما ماتریس یاد شده نامتقارن با علامت‌های مختلف است. به عبارت دقیق‌تر، قدر مطلق فاصله بین قطعات DNA NZ_JTFD01000048.1|Gene 02 و NZ_JTFD01000048.1|Gene 01 با هم برابر نیستند. با توجه به اینکه نام کانتیگ‌ها (قطعات DNA) بزرگ بود، از اعداد ۱ تا ۲۰ برای نشان دادن نام کانتیگ‌ها در ستون سمت چپ استفاده شد. محاسبه فاصله‌ی کولبک - لایبلر در قطعه‌ی ۱ نسبت به ۲ و ۲ نسبت به ۱ به ترتیب ۱/۰۱۳- و ۰/۰۱۳ محاسبه شد. علت این اختلاف به ماهیت فرمول کولبک-لایبلر بر می‌گردد. همانطور که مشاهده می‌شود، این فرمول در واقع نسبت دو آنتروپی را در ارتباط به هم نشان می‌دهد. اگر آنتروپی قطعه DNA اولی بیشتر از قطعه DNA دو می‌باشد، آنگاه کولبک-لایبلر آن‌ها

مثبت می‌گردد و برعکس. تفاوت اعداد حاصل شده نیز وابسته به ماهیت فرمول است که نشان می‌دهد مقدار لگاریتم دو قطعه در چه مقداری باید ضرب شود. در کل می‌توان نتیجه گرفت اگر آنترپی دو قطعه DNA مشابه به هم باشند، آنگاه، فاصله کولبک-لایبلا آن‌ها صفر خواهد شد و انتظار می‌رود آن دسته از قطعات DNA که چنین خاصیتی را داشته باشند یا فاصله کولبک-لایبلا آن‌ها به صفر نزدیک باشد، احتمالاً یک نقش زیستی مشابه دارند.

جدول ۵- برآورد فاصله کولبک-لایبلا بین قطعه‌های DNA باکتری اشرشیاکولی موثر در ایجاد ورم پستان

Table 5. Predicted Kullback – Leibler of DNA segments in Escherichia Coli affecting mastitis

قطعه DNA	1	2	3	4	5	6	7	8	9	10
1	0.0	-1.013	0.084	0.0	-0.011	0.005	-0.082	0.018	-0.013	0.021
2	0.013	0.0	0.098	0.013	0.009	0.018	-0.021	0.031	0.0	0.034
3	-0.081	0.093	0	-0.081	-0.09	-0.076	-0.073	-0.063	-0.094	-0.060
4	0	-0.013	0.084	0	-0.011	-0.046	-0.083	0.018	-0.013	0.021
5	0.011	-0.002	0.096	0.011	0	0.016	-0.019	0.032	-0.002	0.032
6	-0.005	-0.017	0.079	0.046	-0.016	0	-0.004	0.014	0.018	0.016
7	0.008	0.021	0.073	0.008	0.02	0.004	0	0.01	-0.022	0.014
8	-0.018	-0.031	0.065	-0.018	-0.03	-0.014	-0.01	0	-0.031	0.003
9	0.013	0	0.098	0.013	0.002	0.018	0.022	0.014	0	0.004
10	-0.021	-0.034	0.062	-0.021	-0.032	-0.016	-0.013	-0.003	-0.004	0
12	0.092	-0.004	0.094	0.091	-0.002	0.014	0.002	0.028	-0.004	0.03
12	0.017	-0.03	0.067	-0.002	-0.003	-0.012	-0.009	0.001	-0.03	0.004
13	0.001	-0.012	0.085	0.001	-0.01	0.006	0.01	0.02	-0.012	0.022
14	0.009	-0.004	0.093	0.008	-0.003	0.013	0.017	0.027	0.005	0.03
15	0.001	0.007	0.098	0.014	0.003	0.018	0.022	0.032	0.001	0.035
16	-0.003	0.016	0.081	-0.03	-0.004	0.002	0.005	-0.015	0.006	-0.018
17	0.007	0.101	0.092	0.072	-0.004	0.018	0.016	0.026	-0.006	0.025
18	-0.016	0.99	0.068	-0.016	-0.027	-0.011	-0.007	0.026	-0.029	0.005
19	0.011	-0.5	0.095	0.01	-0.001	0.015	0.012	0.029	-0.003	0.032
20	0.007	0.59	0.091	0.007	-0.004	0.012	0.015	0.026	-0.006	0.029

جدول ۶- برآورد فاصله کولبک- لایبلر بین قطعات DNA باکتری اشرشیاکولی موثر در ایجاد ورم پستان - ادامه

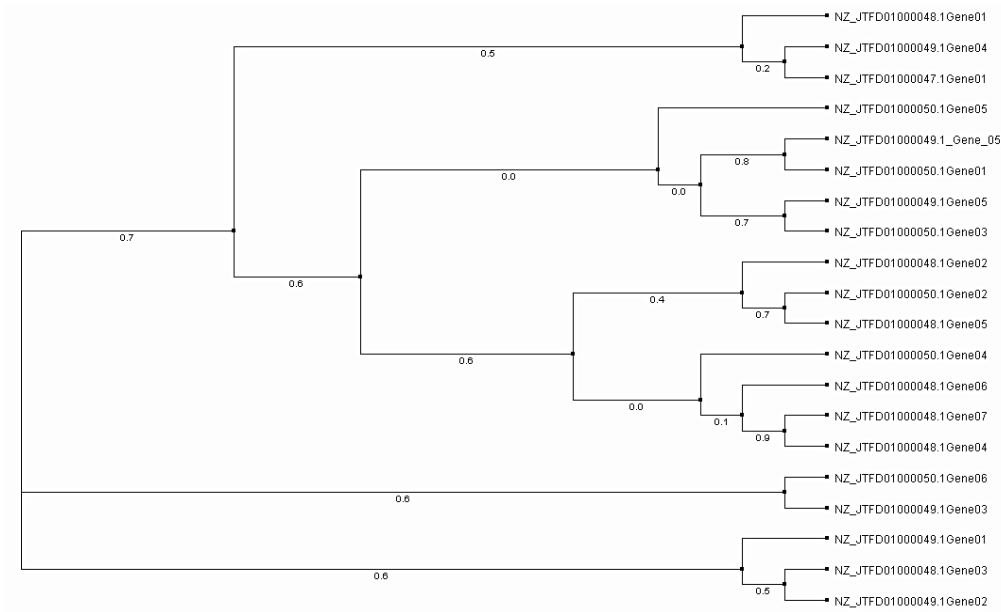
Table 6. Predicted Kullback – Leibler of DNA segments in Escherichia Coli affecting mastitis-continued

قطعه DNA	11	12	13	14	15	16	17	18	19	20
1	-0.009	0.017	-0.001	-0.009	-0.001	0.003	-0.007	0.016	-0.011	-0.007
2	0.003	0.030	0.016	0.006	-0.007	-0.016	-0.006	-0.028	0.003	0.006
3	-0.089	-0.064	-0.082	-0.087	-0.934	-0.077	-0.086	-0.066	-0.091	-0.087
4	-0.091	0.002	-0.002	-0.008	-0.013	0.029	-0.072	0.016	-0.010	-0.007
5	0.002	0.003	-0.010	0.003	-0.003	0.036	0.004	0.027	0.007	0.004
6	-0.014	0.012	0.006	-0.013	-0.018	-0.002	-0.019	0.011	-0.015	-0.012
7	-0.016	0.009	-0.010	-0.017	-0.022	-0.005	-0.015	0.007	-0.019	-0.015
8	-0.027	-0.001	-0.020	-0.027	-0.032	-0.015	-0.025	-0.003	-0.029	-0.025
9	0.004	0.030	0.001	0.005	-0.001	0.006	0.006	0.029	0.003	0.006
10	-0.030	-0.004	-0.022	-0.029	-0.034	-0.018	-0.028	-0.005	-0.031	-0.028
12	0.0	0.026	0.008	0.001	-0.005	0.012	-0.002	0.025	-0.001	0.002
12	-0.026	0.0	-0.008	-0.025	-0.031	-0.014	-0.024	-0.001	-0.027	-0.024
13	-0.004	0.018	0.0	-0.007	-0.012	0.004	-0.006	-0.019	-0.009	-0.006
14	-0.001	0.026	0.007	0.0	-0.005	0.011	0.001	0.024	-0.002	0.001
15	0.005	0.031	0.012	0.005	0.0	0.017	0.006	0.030	0.003	0.007
16	-0.012	0.014	-0.004	-0.011	-0.016	0.0	-0.010	0.028	-0.013	-0.010
17	0.002	0.024	0.006	-0.001	-0.007	-0.010	0.0	0.023	-0.003	0.001
18	-0.025	0.001	-0.017	-0.024	-0.029	-0.027	-0.023	0.0	-0.026	-0.026
19	0.001	0.028	0.009	0.002	-0.003	0.013	0.003	0.026	0.0	0.003
20	-0.002	0.024	0.006	-0.001	-0.007	0.010	0.002	0.023	-0.003	0.0

جدول‌های ۵ و ۶ ماتریس فاصله بین قطعات DNA از نظر معیار کولبک - لایبلر را به نمایش می‌گذارد. این ماتریس می‌تواند ورودی الگوریتم‌های نظارت نشده^۱ مثل خوشه‌بندی سلسله مراتبی^۲ باشد. در آن صورت قطعه‌های DNA که خود را به صورت خوشه نشان می‌دهند به راحتی قابل تشخیص خواهند بود. بر این اساس احتمالاً قطعاتی که در داخل یک خوشه قرار می‌گیرند در یک مسیر زیستی مشترک فعالیت دارند. در پژوهش حاضر سعی شد که نتایج حاصل از محاسبه‌ی فاصله کولبک - لایبلر روی قطعات DNA با درختچه حاصل روی توالی قطعات DNA مقایسه شود که شکل ۱ نتایج حاصل را نشان می‌دهد. لازم به ذکر است درختچه فیلوژنی توسط نرم افزار PhyML (Phylogeny.lirmm.fr) v3.0 ترسیم شد.

۱. Unsupervised

۲. Hierarchical Clustering



شکل ۱- نتایج حاصل از ایجاد درختچه فیلوژنی روی قطعه‌های DNA مورد بررسی. اعداد روی بازوها یا شاخه‌های درختچه میزان Bootstrapping را نشان می‌دهد.

Figure 1. The results of creation of phylogenetic tree on investigated DNA segments. The numbers on the branches of tree shows amount of Bootstrapping.

فراوانی نوکلئوتیدها در کل ماتریس هم‌ردیفی به صورت $f(G)=0.27947$, $f(C)=0.22181$, $f(T)=0.21966$ و $f(A)=0.27907$ بود. در ایجاد درختچه فیلوژنی از مدل جاننشینی DNA زمان برگشت‌پذیر تعمیم یافته^۱ استفاده شد. در این مدل احتمال جاننشینی نوکلئوتیدها به صورت زیر بود: $P(A \leftrightarrow C) = 0.83883$, $P(A \leftrightarrow G) = 3.09923$, $P(A \leftrightarrow T) = 1.26449$, $P(C \leftrightarrow G) = 0.50273$, $P(C \leftrightarrow T) = 0.91546$, $P(G \leftrightarrow T) = 1.00000$. مقدار فراسنجه شکل^۲ مدل برابر با 2.731 بود. درختچه فیلوژنی با تاریخ تکاملی موجودات زنده و گونه‌زایی سروکار دارد. بنابراین مفروضات این روش با ورودی‌های این مطالعه می‌تواند در تضاد باشد چرا که توالی‌های DNA در این پژوهش مربوط به یک سازواره بود. علت مشاهده چنین نتیجه‌ایی در این مطالعه تا حد زیادی به مفروضات ما بر می‌گردد. فرض بر این بود که

۱. Generalized time reversible

۲. Shape

آن دسته از توالی‌های DNA که ساختار یکسانی دارند، احتمالاً یک نوع پروتئین را کد می‌کند و در نتیجه نقش عملکردی زیستی یکسانی خواهند داشت. بنابراین امکان استخراج شبکه متابولیتی بین توالی‌های زیستی یک سازواره به طور نسبی وجود دارد. البته این در صورتی درست است که هیچگونه اطلاعات زیستی دیگری موجود نباشد. در مطالعه‌ای که توسط فورست و همکاران در سال ۲۰۰۱ انجام شد، اساس نظریه درختچه حیات برای تحلیل مسیرهای متابولیتی به کار رفت (۹). این مطالعه از اولین مطالعاتی محسوب می‌شود که ترکیب داده‌های سطح DNA و مسیرهای متابولیکی با استفاده از درختچه حیات انجام شد. در مطالعه دیگری هیمنز و سینگ (۲۰۰۳) با استفاده از نظریه گراف تئوری درختچه فیلوژنی را بر اساس شباهت متابولیتی ایجاد کردند که با نتایج آزمایشگاهی همخوانی داشت (۱۲). همچنین کلمته و همکاران (۲۰۰۷) با توجه به انباشت داده‌های متابولیکی، ایده‌ای را مطرح کردند که در آن استفاده هر چه بیشتر داده‌های متابولیکی برای درک بهتر ارتباط تکاملی گونه‌های مختلف، مورد بررسی قرار گرفت (۵). در مدل یاد شده هیچگونه داده ژنومی استفاده نشد اما از درختچه تکاملی برای ایجاد آن استفاده شد. بایستی خاطر نشان کرد که ماهیت پژوهش‌های یاد شده با پژوهش حاضر متفاوت بود زیرا داده‌های ورودی متعلق به گونه‌های مختلف بودند. با این وجود نشان می‌دهد که بنیاد نظری ایجاد کننده درختچه تکاملی را می‌توان برای ارتباط متابولیکی به کار برد.

نتیجه‌گیری کلی

نتایج نشان داد که آنترופی‌های قطعه‌های DNA تفاوت چندانی با هم نداشتند و یکسان بودن آنترופی، نشان می‌دهد که قطعه‌های DNA به احتمال زیاد، دارای اطلاعات زیادی هستند. همچنین الگوهای مشابه آنترופی‌ها، نشان می‌دهد که قطعه‌های DNA دارای اطلاعات ناشناخته زیادی هستند که شاید با عملکردهای مستقل در ارتباط باشند. هر چه آنترופی یک قطعه DNA بالاتر باشد احتمال آنکه در آن بخش، یک الگوی خاص DNA مثل نشانگر وجود داشته باشد نیز بیشتر است و می‌توان برای آن‌ها آغازگر طراحی نمود. با مقایسه پراکندگی مقدار آنترופی ژن‌ها نتیجه می‌گیریم که هرچقدر آنترופی توالی کل ژن به مقدار بیشینه مقدار خود عدد ۲ نزدیک شود؛ آن ژن از اهمیت بیشتری در بروز بیماری ورم پستان برخوردار است و لذا اپرون آن نیز از آنترופی بیشتری برخوردار است. دلیل آن این است که اپرون‌ها قسمت معنی‌دار ژن و نقش اصلی را بر عهده دارند و در بردارنده‌ی اطلاعات

غالب توالی هستند. برای ۲۰ قطعه مورد بررسی، دامنه متفاوتی از اطلاعات متقابل بین نوکلوتیدها بدست آمد که صفر نشان دهنده نبود ارتباط جایگاه‌ها با یکدیگر و به عبارتی از لحاظ زیستی ارتباطی بین این دو جایگاه وجود ندارد. پس نوع پروتئینی که می‌تواند توسط این جایگاه‌ها ایجاد شود با یکدیگر ارتباطی ندارند. بیشینه مقدار اطلاعات متقابل محاسبه شده (۲) و کمترین آن (۱/۱) بود. مقدار صفر از لحاظ ژنتیک کمی را می‌توان به مانند نبود تعادل پیوستگی در یک گروه یا ساختار ژنتیکی نسبت داد که می‌تواند ارزش بسیار بالایی در اصلاح نژاد داشته باشد. بنابراین هر چقدر مقدار اطلاعات متقابل بین نوکلوتیدها بالاتر باشد، میزان نبود تعادل پیوستگی نیز بالاتر خواهد بود. ژن‌های بسیاری در حیوانات مسئول مقاومت در برابر بیماری‌ها هستند. اگر چه سامانه فیزیولوژیک و تکامل موجودات، مشابه هم نیست ولی عملکرد تمامی موجودات متأثر از ژن‌هایی است که محیط نیز در بیان آن‌ها دخالت دارد. در این مطالعه از زنجیره مارکف با مرتبه صفر (یعنی استقلال نوکلئوتیدها از همدیگر) استفاده شد که این نمی‌تواند از لحاظ ژنتیک ملکولی کاملاً پشتیبانی شود. لذا پیشنهاد می‌شود که ارتباط زنجیره‌ی مارکف با درجه‌های ۱ و ۲ در توالی قطعه‌های DNA نیز مورد بررسی قرار گیرد و سپس میزان آنتروپی توالی DNA محاسبه شود. نتایج حاصل از این پژوهش می‌تواند مبنایی برای پژوهش‌های آینده در پایه‌گذاری یک مطالعه شبیه‌سازی از مفهوم معمول عدم تعادل پیوستگی باشد.

منابع

1. Biswas, D., Itoh, K., and Sasakawa, C. 2003. Role of microfilaments and microtubules in the invasion of INT-407 cells by *Campylobacter jejuni*. *Microbiol. Immunol.*, 47: 469–473.
2. Bolshoy, A. 2008. Revisiting the relationship between compositional sequence complexity and periodicity. *Comput. Biol. Chem.*, 32: 17–28.
3. Bradley, A. 2002. Bovine mastitis: an evolving disease. *Vet. J.*, 164: 116–128.
4. Chiang, D., Joshi, A.K. and Searls, D.B. 2006. Grammatical representations of macromolecular structure. *Comput. Biol.*, 13: 1077–1100.
5. Clemente, J.C., Satou, K., and Valiente, G. 2007. Phylogenetic reconstruction from non-genomic data. *Bioinformatics*, 23: 110–115.
6. Cover, T.M. and Thomas, J.A. 2006. *Elements of Information Theory*. 2nd Edition, Wiley–Interscience Publication, New York. Pp: 42-52.
7. Damasevicius, R. 2010. Complexity estimation of genetic sequences using information-theoretic and frequency analysis methods. *Informatica*, 21: 13–30.

8. Dopfer, D., Almeida, R.A., Lam, T.J., Nederbragt, H., Oliver, S.P., and Gastra, W. 2000. Adhesion and invasion of *Escherichia coli* from single and recurrent clinical cases of bovine mastitis in vitro. *J. Vet. Microbiol.* 74: 331–343.
9. Forst, C.V. and Schulten, K. 2001. Phylogenetic analysis of metabolic pathways. *J. Mol. Evol.*, 52: 471–489.
10. Habibi, M., Pezeshk, H., Eslahchi, C., and Sadegi, M. 2007. Allocation of protein secondary structure using entropy. Iran's fifth largest biotechnology conference. Tehran, Iran. P 33-39. (In Persian)
11. Herzel, H., Ebelling, W., and Schmitt, A.O. 1994. Entropies of biosequences: The role of repeats. *Phys. Rev.*, 50: 5061–5071.
12. Heymans, M., and Singh. A.K. 2003. Deriving phylogenetic trees from the similarity analysis of metabolic pathways. *Bioinformatics*, 19(1): 138–146.
13. Kharrati koopaei, H., Mohammadabadi, M.R., Tarang, A., Kharrati koopaei, M., and Esmailzadeh Koshkoiyeh, A. 2012. Study of the association between the allelic variations in DGAT1 gene with mastitis in Iranian Holstein cattle. *J. Modern Gen.* 7 (1): 101-104 (In Persian).
14. Kim, J., Kim, S., Lee, K., and Kwon, Y. 2009. Entropy analysis in yeast DNA. *Chaos Soliton. Fract.*, 39: 1565–1571.
15. Lender, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., and Baldwin, J. 2001. Initial sequencing and analysis of the human genome. *Nature*, 40: 9860–9921.
16. Li, C., and Wang, J. 2005. Relative entropy of DNA and its application. *J. Physa.*, 347: 465–471.
17. Malinowski, E., Lassa, H., Kłossowska, A., Smulski, S., Markiewicz, H., and Kaczmarowski, M. 2006. Etiological agents of dairy cows' mastitis in western part of Poland. *J. Vet. Sci.*, 9: 191–194.
18. Mohammadabadi, M.R., Shaikhaev, G.O., Sulimova, G.E., Rahman, O., and Mozafari, M.R. 2004. Detection of bovine leukemia virus proviral DNA in Yaroslavl, Mongolian and black pied cattle by PCR. *Cell. Mol. Biol. Lett.*, 9(4A): 766-768.
19. Mohammadabadi, M.R., Soflaei, M., Mostafavi, H., and Honarmand, M. 2011. Using PCR for early diagnosis of bovine leukemia virus infection in some native cattle. *J. Genet. Mol. Res.*, 10(4): 2658-2663.
20. Mohammadi, A., Nassiry, M.R., Mosafer, J., Mohammadabadi, M.R., and Sulimova, G.E. 2009. Distribution of BoLA-DRB3 allelic frequencies and identification of a new allele in the Iranian cattle breed Sistani (*Bos indicus*). *Russ. J. genet.*, 45(2): 198-202.
21. Monge, R.E., and Crespo, J.L. 2014. Analysis of data complexity in human DNA for gene-containing zone prediction. *Entropy*, 17: 1673–1689.
22. Neagoe, I.M., Popescu, D., and Nicule, V.I.R. 2014. Applications of entropic divergence measures for DNA segmentation into high variable regions of

- cryposporidium* spp. GP60 gene. Rom.J. Rep. Phys. 66: 1078–1087.
23. Priestnall, S.L., Wiinberg, B., Spohr, A., Neuhaus, B., Kuffer, M., Wiedmann, M., and Simpson, K.W. 2004. Evaluation of “*Helicobacter heilmannii*” subtypes in the gastric mucosas of cats and dogs. J. Clin. Microbiol., 42: 2144–2151.
24. Ruiz-Marin, M., Matilla-Garcia, M., Cordoba, J.A.G., Susillo-Gonzalez, J.L., Romo-Astorga, A., Gonzalez-Pérez, A., Ruiz, A., and Gayan, J. 2010. Anentropytest for single-locus genetic association analysis. BMC Genet., 11: 19–27.
25. Ruzina, M.N., Shtyfurko, T.A., Mohammadabadi, M.R., Gendzhieva, O.B., Tsendsuren, T., and Sulimova, G.E. 2010. Polymorphism of the BoLA-DRB3 gene in the Mongolian, Kalmyk, and Yakut cattle breeds. Russ. J. genet., 46 (4): 456-463.
26. Schneider, T.D. 1996. Reading of DNA sequence logos: prediction of major groove binding by information theory. Meth. Enzymol., 274: 445–455.
27. Sherwin, B.W. 2010. Entropy and information approaches to genetic diversity and its expression: Genomic geography. Entropy, 12: 1765–1798.
28. Varella Coelho, M.L., Santos Nascimento, J.D., Fagundes, P.C., Madureira, D.J., Oliveira, S.S., Vasconcelos de Paiva Brito, M.A., and Freire Bastos, M.C. 2007. Activity of staphylococcal bacteriocins against staphylococcus aureus and streptococcus agalactiae involved in bovine mastitis. Res. Microbiol., 158: 625–630.
29. Xie, X., Yu, Y., Liu, G., Yuan, Z., and Song, J. 2010. Complexity and entropy analysis of DNA methyltransferase. J. Data Mining Genomics Proteomics, 1: 105–111.



Gorgan University of Agricultural
Sciences and Natural Resources

J. of Ruminant Research, Vol. 4(2), 2016
<http://ejrr.gau.ac.ir>

Investigation of Information Accumulation in *Escherichia Coli*'s DNA Sequence Affecting Mastitis in Dairy Cow Using Information Theory

*M. Ghaderi-Zefrehei¹, A. Bandi Dastjerdi², M.R. Bahreini Behzadi¹,
F. Samadian¹, M. Meamar¹

¹Assistant Prof., and ²M.Sc. Graduated, Dept. of Animal sciences, Faculty of Agriculture,
Yasouj University, Iran

Received: 06/08/2016; Accepted: 09/12/2016

Abstract

Background and objectives: Mastitis causes lots of cost annually in dairy farm enterprises of different countries. Mastitis caused by *Escherichia coli* is common in high-producing cows with low milk somatic cell count. The severity and effect of *Escherichia coli* mastitis vary between cows of the same herd and between different lactation stages in the same individual. The severe form of *Escherichia coli* mastitis is associated with loss of milk production and can outcome in death of the cow. As matter of fact that *Escherichia coli* bacterium plays enormous role in causing and expanding mastitis, in this research it was tried those DNA sequence segments of this bacterium involving in mastitis to be investigated using information theory.

Materials and methods: In this study, first, *Escherichia coli*'s DNA sequences related to mastitis in dairy cattle were downloaded from Genbank (NCBI) and saved in FASTA format. The DNA segment sequences were aligned. The amount of entropy, mutual information among nucleotides and Kullback – Leibler distance among DNA segment sequences were calculated using MATLAB software.

Results: The results shown that entropy of DNA sequences were different in this bacterium but almost all of them had high amount of entropy ($H(x) > 1.900$). The absolute values of Kullback – Leibler distance indicated that DNA sequences investigated in this study were not similar. This might indicates that these DNA sequences could be involved in different metabolic networks, therefore,

*Corresponding author; mghaderi@yu.ac.ir

manipulating them - like silencing –might bring about unpredicted consequences. The results of mutual information turned up that DNA bases in different DNA sequence segments have different degree of association, which could transparent a type of linkage disequilibrium among different DNA bases.

Conclusion: In general, it was concluded that for managing mastitis, it is possible to single out those DNA sequences segments bearing up high amount of entropy as drug and biotechnological targets; since in this study it was assumed that those DNA sequence segments that exert high amount of entropy, would have higher degree of influence on mastitis in dairy cows.

Keywords: *Escherichia coli*, Mutual Information, Entropy, Kullback– Leibler, Mastitis