مقاله علمی پژوهشی/نشریه پژوهش‌های علوم دامی/ جلد ۲۹ شماره ۱/ سال۱۳۹۸/ صفحات ۱۰۷ تا ۱۲۳

کنکاش ساختار، جزایر CPG، متیله شدن DNA و پروتئین حاصل از ژن‌های کاندیدای موثر بر ورم پستان در گاو شیری

مصطفی قادری زفره‌ای[1*]، محمد امیری[2]، لیلا هاشمی[2] و فرهاد صمدیان[1]

تاریخ دریافت: ۹۶/۹/۲ تاریخ پذیرش: ۹۷/۱۱/۲

[1] استادیار گروه علوم دامی، دانشکده کشاورزی، دانشگاه یاسوج

[2] به ترتیب دانشجو و فارغ التحصیل کارشناسی ارشد گروه علوم دامی، دانشکده کشاورزی، دانشگاه یاسوج

[*] مسئول مکاتبه: Email: mghaderi@yu.ac.ir

چکیده

زمینه مطالعاتی: متیلاسیون DNA اساساً در دی‌نوکلئوتیدهای CpG اتفاق می‌افتد و در تنظیم بیان ژنی درگیر می‌باشد. در پستانداران، مشخص شده است که بیش‌متیلاسون در جزایر CpG با پیری و بیماری‌های مختلف در دام مرتبط است. میزان نیمرخ متیلاسیون ژن‌ها می‌تواند به متخصص اصلاح نژاد کمک کند که برای مدیریت بیماریی که این ژن‌ها در آن درگیر هستند از روش‌های ژنتیک یا وراژنتیک استفاده کند. **هدف**: کنکاش غیر مستقیم متیلاسیون DNA در سطح ژن‌های کاندیدای موثر در ورم پستان گاوشیری [ژن‌های اینترلوکین ۱ تا ۱۳ (به جز اینترلوکین ۸ و ۹)، ژن TNF و ژن اینترفرون گاما (در کل ۱۳ ژن کاندید)] و پروتئین حاصل از این ژن‌ها هدف اصلی پژوهش حاضر است. امید بر آن که بتوان ژن‌های درگیر در این بیماری را از نظر میزان متیلاسیون گروه‌بندی کرد و به اهمیت پدیده ژنتیکی متیلاسیون ژن‌ها در مدیریت بیماری ورم پستان در گاو شیری پی برد. **روش کار**: ویژگی‌هایی مثل مکان ژن، تعداد اگزون، طول کلی اکسون پیرایش یافته و حاشیه‌نویسی شده، درصد گوانین، تعداد گوانین، درصد سیتوزین، تعداد سیتوزین و طول ژن‌ها استخراج شدند. برای استخراج ویژگی‌های وراژنتیک ژن‌های مورد بررسی در این پژوهش، از ترکیبی از انواع نرم افزارهای برخط DBCAT، SMS و Geneinfinity و یک نرم افزار متکی به C++ استفاده شد. بعد از استخراج ویژگی‌های ژنی، مشاهده گردید که ژن‌های مورد بررسی از نظر اندازه، تعداد آکسون و درصد GC تفاوت زیادی با هم دارند. با توجه به نوع الگوریتم استفاده شده برای شناسایی جزایر CpG، تعداد و طول جزایر CpG یافت شده متفاوت بود. CpGهای پیش‌بینی شده به عنوان معیاری غیر مستقیم برای بررسی میزان متیلاسیون در سطح DNA ژن‌ها به کار رفت. **نتایج**: نتایج نشان داد که به طور میانگین کمتر از ۵۰ درصد جزایر CpG در ناحیه راه‌انداز ژن‌ها قرار داشتند و جزایر CpG متیله شده بیشتر در نواحی بین ژنی توزیع شده بودند. بیشترین میزان متیلاسیون در اینترلوکین‌های ۱، ۲ و ژن TNF مشاهده شد که از مهم‌ترین ژن‌های سرکوب ورم پستان به شمار می‌روند. درجه متیلاسیون در پروتئین‌های حاصل از ژن‌های مورد بررسی، با نتایج متیلاسیون رخ داده در سطح دی‌ان‌ایِ (DNA) این ژن‌ها همخوانی و مطابقت نداشت. پروتئین حاصله از اینترلوکین ۱۱ در قسمت‌های زیادی از ساختار اولیه خود دچار متیلاسیون می‌شود. در سطح DNA نیز این ژن، بیشترین تعداد CpG را در قسمت راه‌انداز خود داشت. **نتیجه‌گیری نهایی**: ژن اینترلوکین ۱۱ می‌تواند از کاندیدهای مناسب برای انجام آزمایش‌های دمتیلاسیون، تهیه داروهای زیستی و بررسی اثر آن بر سرکوب عفونت ورم پستان باشد.

واژگان کلیدی: جزایر CPG، گاو شیری، متیلاسیون پروتئین، ورم‌پستان

مقدمه

وراژنتیک[1] چالشی جدید در عصر پساژنومی حیوانات اهلی به شمار می‌رود. ژنتیک حیوانی در حال عبور از عصر ژنومی است و شواهدی دال بر وجود تفاوت‌های وراژنتیکی بین حیوانات اهلی وجود دارد که قابل انتقال به نسل‌های بعد هستند (گودارد و وایتلو ۲۰۱٤). از شناخته شده‌ترین سازوکارهای وراژنتیکی می‌توان به متیلاسیون DNA اشاره کرد. نتایج پژوهش‌های دو گروه کاملاً مستقل در سال ۱۹۵۷ نشان دادند که تغییرات شیمیایی DNA (متیلاسیون سیتوزین) در تنظیم بیان ژن درگیر می‌باشد (لوئو ۲۰۱۲). متیلاسیون DNA تنها تغییر وراژنتیکی است که به طور مستقیم برروی DNA تأثیر می‌گذارد. متیلاسیون DNA بر روی باز سیتوزین (۵ متیل سیتوزین) ژنوم بیشتر یوکاریوت‌ها و بعضی از باکتری‌ها اتفاق می‌افتد. در پستانداران متیلاسیون DNA برای تکامل جنینی و تمایز سلول‌های بنیادی ضروری است (رئیک و همکاران ۲۰۰۱). همچنین در پستانداران، متیلاسیون DNA اساساً درون دی‌نوکلئوتیدهای CpG (C متصل به G) اتفاق می‌افتد. دی‌نوکلئوتیدهای CpG تمایل به تشکیل خوشه‌هایی در ژنوم دارند که جزایر CpG[2] نامیده می‌شوند. جزایر CpG به عنوان توالی‌ایی از DNA با طول بیش از ۲۰۰ جفت باز و دارای محتوی GC بیش از ۵۰ درصد تعریف می‌شود که نسبت مقادیر مشاهده شده به مقادیر مورد انتظارِ دی‌نوکلئوتید CpG[3] در آن بیش از ٦۰ درصد باشد (گاردینر-گاردن و فورمر ۱۹۷۸). نسبت یاد شده به صورت ذیل بیان می‌شود: $Obs_{CpG}/Exp_{CpG}=\frac{N_{CpG}/N}{N_C/N\times N_G/N}$ که در این معادله N_C، N_G، N_{CpG} و N به ترتیب تعداد باز سیتوزین، تعداد باز گوانین و تعداد دی‌نوکلئوتید CpG در قطعه ژنومی به طول N هستند. جزایر CpG اغلب در نواحی راه‌انداز ژن‌ها قرار دارند (ایلینگورس و همکاران ۲۰۰۸، و مدودِنا ۲۰۱۱). الگوی متیلاسیون DNA در طی همانندسازی DNA در جایگاه CpG از بین نمی‌رود و در واقع متیلاسیون DNA قابل توارث است. با وجود این، فرآیند متیلاسیون ممکن است بلوکه شود، یعنی آنزیم‌هایی هستند که قادر هستند گروه‌های متیل را از توالی‌های DNA بر دارند که دمتیلاز نامیده می‌شوند. الگوی متیلاسیون در ژنوم پستانداران به این صورت است که دی‌نوکلئوتیدهای CpG در جزایر CpG به صورت غیر متیله هستند که بدین شکل محل اتصال پروتئین‌های تنظیمی ویژه به ژن می‌باشند، ولی در بدنه ژنوم به صورت متیله می‌باشند. تغییر این الگو یعنی هایپومتیلاسیون وسیع بدنه ژنومی منجر به ناپایداری کروموزومی و تغییر بیان ژن‌ها می‌شودو تغییر الگو به شکل هایپرمتیلاسیون جزایر CpG منجر به بروز سرطان‌ها و یا بیماری‌های دیگر می‌شود (شامس و همکاران ۲۰۰۷). به طور کلی متیلاسیون DNA در فرایندهای تمایز سلولی، نشانه‌گذاری ژنومی، غیرفعال شدن تصادفی کروموزوم X، مرگ برنامه‌ریزی شده سلولی، سرطان‌ها، پیری و غیره نقش تعیین‌کننده‌ای دارد (جیانگ و همکاران ۲۰۰۷، ورزودک و همکاران ۲۰۱۲ و تابرلای و همکاران ۲۰۱٤). متیلاسیون DNA ممکن است سبب ایجاد فنوتیپ‌های غیرطبیعی در حیوانات شود (اوهانگ و همکاران ۲۰۰٤). متیلاسیون DNA یک تغییر و تبدیل مهم وراژنتیکی در سطح تنظیمات رونویسی است که به طور قابل توجهی وابسته به جزایر CpG می‌باشد (بیرد ۲۰۰۲ و دیو و همکاران ۲۰۱۲). به تازگی پژوهش‌های زیادی روی متیلاسیون ژنوم گونه‌های اهلی انجام شده است. هو و همکاران (۲۰۱۳) نشان دادند که کمترین تراکم متیلاسیون در جزایر CpG واقع در راه‌انداز ژن‌ها است، اما در بدنه ژن‌ها تراکم متیلاسیونی در جزایر CpG واقع در اینترون بیش‌تر از اگزون و UTRs[4] بود. کوآک و همکاران

[1] Epigenetics
[2] CpG Islands
[3] Observed CpG/Expected CpG
[4] Untranslated Region

(۲۰۱٤) نیز در پژوهشی رخ‌نمای[1] متیلاسیون را در ژنوم خوک بررسی نمودند که نتایج آن‌ها مشابه نتایج به دست آمده در ژنوم جوجه بود. در پژوهشی که کلدری و همکاران (۲۰۱٤) بر روی گوسفند انجام دادند، نشان داده شد که نواحی نزدیک TSS[2] ژن‌ها دارای متیلاسیون پایین‌تری هستند. در دو پژوهش روی ژنوم گاو (سو و همکاران ۲۰۱٤؛ هوآنگ و همکاران ۲۰۱٤) نتایجی مشابه با سایر پژوهش‌ها به دست آمد، به این صورت که سطح متیلاسیون در بدنه ژن‌ها نسبتاً بالا و در راه‌انداز ژن‌ها پایین بود. اکثر جزایر CpG در بخش راه‌انداز ژن‌ها غیرمتیله بودند و جزایر CpG متیله در اینترون و بدنه ژن‌ها قرار داشتند. گرچه کنکاش متیلاسیون ژنوم حیوانات مزرعه‌ای با عنایت به فناوری‌های رو به گسترش NGS، رو به گسترش است، اما تا کنون کنکاش همزمان متیلاسیون DNA و پروتئین حاصل از ژن‌های کاندید موثر بر ورم پستان گاو شیری انجام نشده است. ورم پستان به عنوان دومین عامل گاوهای هلشتاین کشور عنوان شده است (شریعتی و همکاران ۲۰۱۷). بنابراین میزان نیمرخ متیلاسیون ژن‌ها می‌تواند به متخصص اصلاح نژاد کمک کند که برای مدیریت بیماری ورم پستان که ژن‌ها در آن درگیر هستند از روش‌های ژنتیک یا وراژنتیک استفاده کند. میزان بالای متیلاسیون، استفاده از روش‌های وراژنتیک را طلب می‌کند. در پژوهش حاضر به روش غیر مستقیم و متکی به برآورد جزایر CpG و جایگاه آن‌ها روی این ژن‌ها، میزان متیلاسیون در سطح DNA بررسی شد. همچنین برای اولین بار، کنکاشی روی میزان متیلاسیون پروتئین‌های حاصله از این ژن‌ها صورت گرفت تا شباهت و تفاوت این ژن‌ها از نظر میزان متیلاسیون در دو سطح یاد شده آشکار شود؛ به طوری که بتوان این ژن‌ها را از نظر میزان متیلاسیون نیز گروه‌بندی کرد و به اهمیت پدیده ژنتیکی متیلاسیون ژن‌ها در مدیریت بیماری ورم پستان در گاو شیری پی برد.

مواد و روش‌ها

استخراج توالی

در این پژوهش، ابتدا منابعی که در بردارنده ژن‌های کاندید موثر بر ورم پستان در گاو شیری بودند، کنکاش شدند و تعدادی از آن‌ها که متعلق به گروه اینترلوکین‌ها بودند (اینترلوکین ۱ تا ۱۳ به جز اینترلوکین ۸ و۹)، به همراه دو ژن دیگر (TNF و گاما اینتروفرن) انتخاب شد (در مجموع ۱۳ ژن) و توالی DNA آن‌ها از بانک ژن NCBI[3] استخراج و به صورت فایل FASTA ذخیره شدند (جدول۲). برای استخراج ویژگی‌های ژن‌ها (جدول۲) از نرم‌افزار متکی به C++ استفاده شد (جدول۱). جهت شناسایی راه‌انداز در ژن‌ها، از نرم‌افزار Promoter2.0 Prediction Server[4] استفاده شد. این نرم‌افزار جهت پیش‌بینی رونویسی توالی DNA مهره‌داران استفاده می‌شود. برای شناسایی جزایر CpG توالی هر ژن، از ۳ نرم افزار (جدول ۱) استفاده شد.

در ادامه ساختار اولیه پروتئین‌های ژن‌های مورد بررسی، از پایگاه NCBI استخراج شد و در نهایت میزان متیلاسیون پروتئین هر ژن توسط نرم‌افزار PMeS (جدول۱) به دست آمد.

نتایج و بحث

تعداد ۱۳ ژن در این پژوهش مورد کنکاش قرار گرفت. اطلاعات حاصل از استخراج ویژگی‌های این ژن‌ها (جدول ۲) نشان داد که بین طول ژن‌ها اختلافات زیادی وجود دارد، به طوری که بیشترین طول ژنی به ترتیب مربوط به ژن‌های اینترلوکین ٤، ٥، ۱۲، ۱۳و اینترفرون گاما بود و کمترین طول ژنی مربوط به ژن TNF بود. همچنین بیشترین میزان سیتوزین مشاهده شده به ترتیب در اینترلوکین‌های ۲، ۱۰، ۱۳، گاما و ۳ و

[1] Profile
[2] Transcription Start Site
[3] https://www.ncbi.nlm.nih.gov/
[4] http://www.cbs.dtu.dk/services/Promoter/

بیشترین میزان گوانین مشاهده شده به ترتیب در اینترلوکین‌های ۱۰، TNF، ۱۳، ٦ و ٤ بود. جایگاه کروموزومی ژن‌ها نیز متفاوت بود. اینترلوکین‌های ٤، ٥، ۱۲ و ۱۳ روی کروموزوم شماره‌ی ۷ و اینترلوکین‌های ۱ و ۷ روی کروموزوم شماره‌ی ١٤ قرار داشتند (جدول ۲).

جدول ۱- فهرست نرم‌افزارها و الگوریتم‌های مورد استفاده در این پژوهش

Table 1- The list of software and algorithms used in this research

Software Name*	Performance	Internet Address
	Extraction of gene features	It was written in C++
Geneinfinity	Calculation of CPG islands	http://www.geneinfinity.org/sms/sms_cpgislands.html
DBCAT	Calculation of CPG islands	http://dbcat.cgm.ntu.edu.tw/
SMS	Calculation of CPG islands	http://www.bioinformatics.org/sms2/cpg_islands.html
PMeS	Calculation of protein methylation	http://bioinfo.ncu.edu.cn/inquiries_PMeS.aspx

* This software was first developed by Dehganzadeh & Ghaderi in the C ++ and was used in present study. This software has been noted in a paper published in Online Journal of Veterinary Research: Dehghanzadeh, H., Mirhoseini, S. Z., Ghaderi-Zefrehei, M., Tavakoli, H., & Esmaeilkhaniyan, S. (2017). Clustering dairy cattle genes by Kullback-Leibler divergence. Online Journal of Veterinary Research, 21(9), 600-615. Volume 21 (9): 600-615, 2017); if necessary, the readers can contact the corresponding author of this article.

جدول ۲- ویژگی‌های ژن‌های کاندید موثر بر ورم پستان مورد بررسی در این پژوهش

Table 2- Characteristics of candidate genes affecting on mastitis in this research

Gene name	Gene length	Cytosine number	Cytosine %	Guanine number	Guanine%	Total Annotated Spliced Exon Length	Exon number	Locus
IL-1	397637	20.29	23.84	1877	22.05	8509	21	AC_000161.1 (40580268..40643381, complement
IL-2	260017	792	51.16	811	16.91	4795	8	AC_000170.1 (17540757..17586322)
IL-3	210639	526	23.93	485	22.06	2198	3	AC_000165.1 (87878331..87893188)
IL-4	660714	702	20.21	785	22.60	3472	4	AC_000164.1 (22993178..23001067, complement)
IL-5	541559	407	16.95	432	17.99	2401	4	AC_000164.1 (23184382..23186782)
IL-6	272996	364	22.60	1371	22.70	6099	6	AC_000161.1 (31576590..31582688)
IL-7	1158917	6834	20.91	6189	18.94	32672	6	AC_000171.1 (44095419..44187924, complement)
IL-10	209807	2755	34.45	2825	35.32	7997	5	AC_000173.1 (4402474..4406412, complement)
IL-11	24318	1619	22.23	1536	21.09	7280	4	AC_000175.1 (62560634..62565444)
IL-12	738823	4029	20.98	4029	20.98	19202	9	AC_000164.1 (73002003..73021204, complement)
IL-13	324388	611	30.44	506	25.21	2007	4	AC_000164.1
TNF	17623	985	25.19	1095	28.01	3909	4	AC_000180.1 (27533899..27537807, complement)
IFNγ	673150	815	16.89	958	19.85	4824	4	AC_000162.1 (45830158..45834981)

A: Interleukin 8 has been reported as an effective candidate for dairy cattle mastitis, but its DNA sequence was not present in the CBI, therefore its study was excluded in this research.

اطلاعات مربوط به راه‌اندازهای پیش‌بینی شده در ژن‌ها، درصد جزایر CPG موجود در راه‌اندازها، نسبت ObsCpG/ExpCpG و جایگاه احتمالی جزایر CPG در راه‌اندازها در جدول ۳ ارائه شده است. با توجه به پژوهش گانگ و فایفر (۲۰۱۵) جایگاه‌های DNA یافت شده که دارای نمره[1]بالاتر از ۵۰٪ باشند، جهت انتخاب مکان راه‌انداز قابل اعتماد هستند. جدول ۳ مشخصات نقطه شروع راه‌انداز برآورد شده برای هر ژن را نشان می‌دهد. همان طور که نشان داده شده است، نمره راه‌انداز تمام ژن‌ها بالاتر از ۵/. است و بیشترین و کمترین نمره راه‌انداز به ترتیب مربوط به ژن اینترلوکین ٦ و اینترلوکین ۱۲ است. راه‌انداز بیش از ٦۰ درصد از ژن‌ها دارای نمره بیش از ۱ بود و همچنین نقطه‌ی شروع بیش از ۵۵ درصد ژن‌ها در یک سوم ابتدایی هر ژن بود. همچنین در این پژوهش، بلندترین توالی پیش از راه‌انداز در اینترلوکین‌های ٦ و ۱۰ مشاهده شد. در این پژوهش به هر نرم‌افزار، طول کل ژن معرفی شد و کل جزایر CpG برای هر ژن به دست آمد. سپس با توجه به طول راه‌انداز هر ژن، CpG های به دست آمده به دو گروه تقسیم شدند: آن‌هایی که در راه‌انداز قرار داشتند و آن‌هایی که خارج از آن قرار داشتند. جدول شماره ۳ نشان می‌دهد که استفاده از نرم‌افزارهای مختلف، جواب متفاوتی را حاصل کرده است. مثلاً برای ژن اینترلوکین ۳ تمامی جزایر CpG تشخیص داده شده با استفاده از نرم‌افزارهای SMS و Geneinfinity در ناحیه راه‌انداز بودند. اما برای ژن‌های اینترلوکین ۱۳، فاکتور نکروزه کننده تومور(TNF) و اینترفرون گاما (Interferon-gamma) هیچ گونه CpG تشخیص داده نشد. که این شاید به قوت الگوریتم‌های به کار رفته بر گردد و نیاز است با الگوریتم‌های اختصاصی‌ترِ توسعه یافته برای ژنوم گاو بررسی شود. چندین الگوریتم جهت شناسایی جزایر CpG وجود دارد و در تعداد زیادی از پژوهش‌ها به کار گرفته شده است. این الگوریتم‌ها قابل تقسیم به دو گروه سنتی و جدید هستند. الگوریتم‌های سنتی متکی بر سه معیار توالی (محتوی[2]GC، نسبت مقادیر مشاهده شده بر مورد انتظار دی نوکلئوتید CpG و طول توالی[3]) می‌باشند ولی الگوریتم‌های جدید بر اساس ویژگی‌های آماری توالی تعریف می‌شوند و متکی بر سه معیار الگوریتم‌های سنتی نمی‌باشند (هان و ژائو ۲۰۰۸).

الگوریتم‌های جدید براساس خوشه‌بندی دی‌نوکلئوتید CpG در طول توالی ژنوم، جزایر CPG را به عنوان یک ویژگی آماری شناسایی می‌کنند و در نتیجه بر آستانه‌های محتوی GC، نسبت مقادیر مشاهده شده بر مورد انتظار دی‌نوکلئوتید CpG و طول توالی تکیه ندارند (ایریزاری و همکاران ۲۰۰۹). پژوهش‌ها نشان داده است که توالی پیش از راه‌انداز ژن در اتصال RNA پلیمراز، جذب آنزیم و شکل‌گیری مضاعف مارپیچ DNA نقش بسیار مهمی دارد و در صورتی که این توالی برداشته شود میزان رونویسی ژن تا ۱۰ برابر کاهش می‌یابد (آبوله و همکاران ۱۹۹۸). میزان رونویسی در این منطقه می‌تواند توسط فرایند متیلاسیون تحت تاثیر قرار گیرد. انتظار است که با توجه به تراکم بالای جزایر CpG در ژن‌های اینترلوکین ۳، ۷، ۱۰ و ۱۱، این ژن‌ها بیشتر تحت تاثیر متیلاسیون قرار گیرند. به طور کلی به دلیل متفاوت بودن الگوریتم مورد استفاده در نرم‌افزار DBCAT با نرم‌افزار SMS و Geneinfinity، نتایج حاصل از محاسبات با هم متفاوت بود. یافته‌های پژوهش نشان داد که جزایر CpG یافت شده در الگوریتم SMS نسبت به جزایر CPG یافت شده در نرم‌افزار DBCAT، دارای طول بیشتر و تعداد کمتر بودند. الگوریتم SMS یک چارچوب ۲۰۰ جفت نوکلئوتیدی را انتخاب و با گامه‌های ۲۰۰ جفت نوکلئوتیدی در طول ژنوم حرکت می‌کند.

جهت جلوگیری از تولید جزایر CpG ریاضی[4](یعنی اگر مثلاً یک توالی به طول ۳۰۰ جفت نوکلئوتید داشته باشیم که ۱۵۰ تا از این نوکلئوتیدها سیتوزین باشند و فقط یک

[1] Score

[2] GC Content

[3] Length

[4] Mathematical CpG

گوانین داشته باشیم، آن‌گاه یک دی‌نوکلئوتید CpG به لحاظ ریاضی معیارهای جزایر CpG را دارد، ولی در واقع یک جزیره CpG نیست)، یک آستانه اضافی در الگوریتم برای تعداد دی‌نوکلئوتیدهای CpG در نظر گرفته شده است (حداقل تعداد ۷ دی‌نوکلئوتید CpG باید در ۲۰۰ جفت باز وجود داشته باشد). در این الگوریتم اگر فاصله دو جزیره CpG کمتر از ۱۰۰ جفت نوکلئوتید باشد آن‌ها را با هم ادغام می‌کند. طراحان این الگوریتم، که به الگوریتم تاکایی نیز معروف است، جهت پیدا کردن جزایر CpG مرتبط با راه‌اندازهای ژن‌های کد کننده پروتئین‌ها و جلوگیری از جزایر CpG مرتبط با توالی‌های تکراری Alu آستانه‌های[1] سخت‌گیرانه‌تری را برای سه معیار طول قطعه (۵۰۰ جفت باز)، محتوی GC (۵۵٪) و نسبت دی‌نوکلئوتید مشاهده شده به مورد انتظار (۰/۶۵) در نظر گرفته‌اند (تاکایی و جونز ۲۰۰۲). بنابراین، به نتایج این الگوریتم می‌توان بیشتر اعتماد داشت. در دو نرم‌افزار SMS و Geneinfinity برای هر جایگاه یک امتیاز با استفاده از روش جمع اجرا[2] محاسبه می‌شود و از ابتدای توالی شروع و در انتهای توالی ختم می‌شود. اگر در یک جایگاه دی‌نوکلئوتید CpG وجود نداشته باشد، امتیاز کاهش و اگر یک دی‌نوکلئوتید CpG وجود داشته باشد امتیاز با ارزش ثابتی افزایش می‌یابد. اگر امتیاز محاسبه شده از مقدار آستانه‌ای بیشتر باشد، یک جزیره فرضی شناسایی می‌شود. نواحی دارای امتیاز بالای مقدار آستانه به صورت برگشتی جستجو می‌شوند. جزایر نهایی کشف شده به وسیله این الگوریتم با دی‌نوکلئوتید CpG شروع و ختم می‌شوند و ضرورتی برای دستیابی به معیارهای تعریف شده توسط گاردین-گاردن و فرومر (۱۹۸۷) وجود ندارد.در این روش تعداد بسیار زیادی جزیره CpG با طول کم و محتوی GC بالا و دی‌نوکلئوتید CpG فراون مشاهده می‌شود که منجر به پیش‌بینی بیش از حد می‌گردد. با توجه به این امر، در صورتی که نمره‌ی مشخص شده توسط نرم‌افزار برای هر جزیره CPG، بزرگ‌تر از ۶/. باشد، نتایج حاصل معنی‌دار است. هرآینه، اساس الگوریتم‌های به کار رفته در این پژوهش از تعریف الگوریتم‌های سنتی پیروی می‌کرد. با توجه به جدول ۳ تشخیص وجود جزایر CPG در سه الگوریتم به طور متوسط در بیشتر موارد مشابه بود. وجود جزایر CPG در دو مورد دارای تفاوت بود که این امر می‌تواند ناشی از نوع الگوریتم مورد استفاده باشد. همچنین ۶ ژن اینترلوکین ۲، ۳، ٤، ٥، و ۱۳ از ژن‌های مورد بررسی در پژوهش حاضر فاقد جزیره CPG بودند که این امر در سه نرم‌افزار به کار رفته مورد تأیید است. در نرم‌افزار DBCAT در صورتی که نسبت مقادیر مشاهده شده بر مورد انتظار دی‌نوکلئوتید CpG محاسبه شده توسط نرم افزار بیشتر از ٥/. باشد، وجود جزیره CPG قابل تأیید است. در این الگوریتم جزایر CpG برای ۸ ژن اینترلوکین ۲، ۳، ٤، ٥، و ۱۳، TNF و اینترفرون گاما شناسایی نشد. نسبت CpG مشاهده شده به مورد انتظار بین ٥٠-٦٥ درصد بود که برای تمامی موارد قابل تأیید است. به طور کلی با توجه به جدول ۳، ٤٦ درصد جزایر CpG یافت شده در این پژوهش بیشتر نزدیک راه‌انداز و در نواحی راه‌انداز بود که با یافته‌های لارسن و همکاران (۱۹۹۲) و ایلینگورس و همکاران (۲۰۰۸) تطابق خوبی را نشان داد. در این دو نرم افزار تعداد زیادی جزیره CPG یافت شدند که با توجه به نمره هر ژن تمام جزایر یافت شده معنی‌دار هستند و این امر می‌تواند ناشی از پیش‌بینی بیش از حد نرم‌افزار باشد. در این نرم‌افزار جزایر CpG برای اینترلوکین‌های ۲، ۳، ٤، ٥، ٦، ۸، ۱۳ و اینترفرون گاما شناسایی نشدند. نسبت CpG مشاهده شده به مورد انتظار بین ٦٠-٩٢ درصد بود که برای تمامی موارد قابل تأیید است. به طور کلی صحت یافته‌های این پژوهش با یافته‌های دیو و همکاران (۲۰۱۲) مطابقت داشت.

در این پژوهش احتمال متیله شدن پروتئین هر یک از

[1] Threshold
[2] Running Sum

ژن‌های مورد بررسی با استفاده از نرم‌افزار pMes (شی و همکاران ۲۰۱۲) بررسی شد. در این راستا توالی پروتئین پیش‌آلفا مربوط به هر یک از ژن‌ها از پایگاه داده‌ی NCBI استخراج شد و سطح متیلاسیون پروتئین هر ژن محاسبه گردید.

جدول ۳- مشخصات راه‌اندازها و جزایر CpG پیش بینی شده در نرم‌افزارهای مختلف

Table 3- Characteristics of CPG islands and predicted promoters in various software

	SMS و Geneinfinity			**DBCAT**				
Gene Name	CpG islands location	ObsCpG/ExpCpG ratio	% of CpG islands in promoter	CpG islands location	ObsCpG/ExpCpG ratio	% of CpG islands in promoter	**Start of the promoters**	**Score**
IL-1	-	-	-	433 - 4543 8288 -8508	50 50	50	4800	1.102
IL-2		-	-	-	-	-	2700	1.165
IL-3	346-545 347-546 351-550 353-552 354-553 355-554 356-555 357-556	0.60 0.60 0.60 0.60 0.62 0.63 0.63 0.63	100	-	-	-	500	0.503
IL-4	-	-	-	-	-	-	900	1.165
IL-5	-	-	-	-	-	-	554	0.96
IL-6	-	-	-	1796 - 2182	60	10	4100	1.193
IL-7	2495-4249 2838-4283 3313-4333 3921-3392 2001-2986	0.77 0.72 0.70 0.63 0.90	60	9594 -9814	55	0	4900	1.164
IL-10	2299-2498 34-233 440-639 530-729 63-262 236-435 2319-2518	0.68 0.62 0.80 0.82 0.71 1.07 1.01	0	116 -1019 2250 -2605 3295 - 3515 5398 - 5801 6915 - 7259 845 - 8815	63 60 5 63 63 62	33	6000	1.010
IL-11	1-200 50-249 100-299 150-349 2338-2537	0.61 0.67 0.71 0.75 0.92	20	27 - 509	59	0	2600	0.690
IL-12	5683-5882 5691-5890 5693-5892 5705-5904 12211-12410	0.61 0.63 0.64 0.65 0.66	80	-	-	-	4100	1.188
IL-13	-	-	-	-	-	-	900	0.615
TNF	-	-	-	-	-	-	1300	1.192
IFNγ	-	-	-	-	-	-	2532	0.99

جدول ٤ اطلاعات متیلاسیون پروتئین‌های مربوط به ژن‌های مورد بررسی را نشان می‌دهد که در واقع به تصحیحات پروتئین‌های ژن‌ها اشاره دارد. تصحیحات بعد از ترجمه پروتئین‌ها اشاره به وقایع بیوشیمیایی دارد که پروتئین‌ها بعد از ترجمه شدن متحمل می‌شوند. از این وقایع می‌توان به متیلاسیون، فسفوریلاسیون، گلیکوزیلاسیون سولفاسیون، سومولاسیون، یوبیکوئیتین لاسیون، پیوندهای دی‌سولفید و غیره اشاره کرد. نرم‌افزار PMeS از الگوریتم ماشین بردار پشتیبان (SVM) – که از الگوریتم‌های مشهور یادگیری ماشین می‌باشد –برای بررسی احتمال متیلاسیون ساختار اولیه پروتئین‌های ژن‌های مورد بررسی- استفاده می‌کند. بر اساس تعریفی که PMeS ارایه کرده است، هر چه احتمال SVM در این الگوریتم برای جایگاهی بیشتر باشد احتمال این که آن جایگاه اسید آمینه متیله شود، زیادتر است. در جدول ٤ منظور از اسید آمینه‌ی کناری همان توالی‌های پپتیدی دو طرف ناحیه متیلاسیون می‌باشد. حرف‌های دوطرف ناحیه متیلاسیون با حروف بزرگ نوشته شده است.

معمولا متیلاسیون پروتئینی بیشتر روی دو اسید آمینه لیزین (K) و آرژنین (R) اتفاق می‌افتد (شی و همکاران ۲۰۱۲). جستجوی متیلاسیون برای هر دو متیلاسیون (چه متیلاسیون لیزین و چه متیلاسیون آرژنین) می‌تواند انجام گیرد. اگر ناحیه متیله شده در موقعیت آمینوی اسید آرژنین باشد در آن صورت حرف R نشان دهنده متیلاسیون می‌باشد (قسمت اسید آمینه کناری حرف وسط R می‌باشد). مکان متیلاسیون همان موقعیت دقیق آمینو اسید متیله شده را نشان می‌دهد. همان طور که جدول ٤ نشان می‌دهد، پروتئین ژن اینترلوکین ۱۱ در جایگاه‌های مختلف خود، میزان بالایی از متیلاسیون را نشان می‌دهد. با توجه به خروجی نرم‌افزار تمام مکان‌های متیله شده مربوط به اسید آمینه‌ی آرژنین می‌باشد که این می‌تواند از نظر زیستی دارای اهمیت باشد. در قسمت بالا نیز مشخص شد که ژن اینترلوکین ۱۱ از جمله ژن‌هایی بود که میزان قابل‌توجهی از جزایر CpG را در قسمت‌های مختلف راه‌اندازش داشت. بنابراین می‌شود گفت که ژن یاد شده احتمالاً در سطوح علمکردی مختلف خود تحت تاثیر سازوکارها و ابزارهای مختلف وراژنتیک قرار می‌گیرد.

متیلاسیون DNA یک تصحیح مهم وراژنتیکی در سطح تنظیمات رونویسی است که به طور قابل توجهی وابسته به جزایر CpG می‌باشد (بیرد ۲۰۰۲؛ دو و همکاران ۲۰۱۲). در این راستا حدود ۵۰ درصد از ژن‌های انسانی در راه‌انداز خود دارای جزایر CpG هستند (وانگ و لیونگ ۲۰۰٤). بر اساس الگوریتم‌های مختلف حدود ۱۵ تا ۳۵ درصد از جزایر CpG در راه‌انداز ژن‌ها قرار دارند (هان و ژائو ۲۰۰۸). ارتباط بین متیلاسیون جزایر CpG مرتبط با راه انداز ژن‌ها و بیان ژن بسیار پیچیده است. چندین پژوهش همبستگی منفی بین متیلاسیون جزایر CpG مرتبط با راه‌انداز ژن‌ها و بیان ژن را گزارش کرده‌اند (فیوتچر و همکاران ۲۰۰۲؛ سونگ و همکاران ۲۰۰۵)، اما برخی دیگر عدم همبستگی را گزارش نموده‌اند (وبر و همکاران ۲۰۰۷؛ ایلینگورس و همکاران ۲۰۰۸). این تفاوت‌ها ممکن است به دلیل پویایی و طبیعت پیچیده متیلاسیون در سامانه‌های سلولی باشد (کانگاسپسکا و همکاران ۲۰۰۸؛ جیا و همکاران ۲۰۱٦).

در پژوهش دیو و همکاران (۲۰۱۲) ویژگی‌های جزایر CpG پیش‌بینی شده توسط الگوریتم‌های مختلف در ژنوم انسان با هم مقایسه شد. در این پژوهش جزایر پیش‌بینی شده توسط الگوریتم تاکایی در مقایسه با جزایر پیش‌بینی شده توسط الگوریتم CpGcluster (با معیار قوی)، دارای طول بیشتر، محتوی GC کمتر، نسبت CpG مشاهده شده به مورد انتظار کمتر و سطح متیلاسیون بالاتری بودند. همچنین در بررسی این پژوهش‌گران بر روی رخ‌نمای متیلاسیون در دو لاین سلولی مشخص شد که ژن‌های دارای سطح متیلاسیون کمتر در جزایر CpG موجود در راه‌انداز خود، در بافت‌های بیشتری بیان شده و همچنین بیان قوی‌تری دارند.

جدول ۴- اطلاعات متیلاسیون پروتئین‌های مربوط به ژن‌های مورد بررسی

Table 4- Data related to protein methylation on the examined genes

Gene	Interleukin 1					Interleukin 2		Interleukin 3	Interleukin 4			Interleukin 5		
Protein code	776517	776517	776517	776517	776517	851340	851340	776345	776346	776346	776346	776347	776347	776347
Methylation sites	44	84	85	128	185	58	113	122	71	77	123	103	109	111
Methylated amino acid location	DASYEPL-R-EDQMNKF	SGKILKK-R-RLSLNQF	GKILKKR-R-LSLNQFI	NVKYNFM-R-VIHQECI	SQLPVTL-R-ISKTQLF	PENLKLS-R-MHTFDFY	DSMDNIK-R-IVLELQG	NKNLGDF-R-MKLEEYL	TEKETFC-R-VGIELRR	CRVGIEL-R-RIYRSHT	TLKDLLE-R-LKTIMKE	EYIDLQK-R-KCGGERW	KRKCGGE-R-WRVKQFL	KCGGERW-R-VKQFLDY
The highest probability of methylation	0.520904	0.50	0.642332	0.539867	0.734988	0.740089	0.611601	0.50000	0.666564	0.583436	0.528942	0.539097	0.709896	0.58435

Gene	Interleukin 6	Interleukin 7		Interleukin 10						Interleukin gamma		
Protein code	776348	776348	776349	776513	776513	776513	776513	776513	776513	776511	776511	776511
Methylation sites	144	192	82	42	45	120	122	124	125	150	152	154
Methylated amino acid location	KEASFLN-R-ASRKLRQ	CEAAVLS-R-MSTELRC	PKFIKEL-R-VIESGPH	TSLPHML-R-ELRAAFG	PHMLREL-R-AAFGEAK	GEKLKTL-R-LRLRRCH	KLKTLRL-R-LRRCHRF	KTLRLRL-R-RCHRFLP	TLRLRLR-R-CHRFLPC	LSPKSNL-R-KRKRSQN	PKSNLRK-R-KRSQNLF	SNLRKRK-R-SQNLFRG
The highest probability of methylation	0.657575	0.500	0.515543	0.56376	0.687545	0.603755	0.642478	0.661306	0.528756	0.729943	0.572697	0.746623

ادامه جدول ۴

Table 4- Continue

Gene	**Interleukin 11**												
Protein code	15323	15323	15323	15323	15323	15323	15323	15323	15323	15323	15323	15323	15323
Methylation sites	18	36	47	54	85	93	101	111	137	143	151	177	195
Methylated amino acid location	ALSLWPD-R-AAAPGPP	PRASPDP-R-AELDSAV	DSAVLLT-R-SLLADTR	RSLLADT-R-QLAAQLV	LLGYQGG-R-DRGGKGV	DRGGKGV-R-GLPGVLT	GLPGVLT-R-LRADLFS	ADLFSYL-R-HVQWLRR	ELGTLQA-R-LDRLLRR	ARLDRLL-R-RLQLLMS	RLQLLMS-R-LALPQVP	ASAWGGI-R-AAHAILG	LTLDWAV-R-GLLLLKT
،The highest probability of methylation	0.509383	0.564159	0.568024	0.585233	0.597982	0.598713	0.523918	0.638152	0.619633	0.632115	0.604458	0.656821	0.616765

Gene	**Interleukin 12**			**Interleukin 13**		**TNF**
Protein code	776780	776780	776780	776514	776514	347561
Methylation sites	58	205	207	81	120	8
Methylated aminoacid location	SNTLQKA-R-QTLEFYS	CILLHAF-R-IRAVTID	LLHAFRI-R-AVTIDRM	SNCSVIQ-R-TKKMLNA	QFLKDLL-R-HSRIVFR	MSTKSMI-R-DVELAEE
The highest probability of methylation	0.513102	0.606984	0.678525	0.551305	0.500000	0.515523

همچنین نتایج نهایی این پژوهش‌گران (دیو و همکاران ۲۰۱۲) نشان داد که متیلاسیون جزایر CpG مرتبط با راه‌انداز ژن‌ها منجر به سرکوب بیان ژن می‌شود. هو و همکاران (۲۰۱۳) در پژوهشی رخ‌نمای متیلاسیون کل ژنوم جوجه‌های گوشتی را بررسی نمودند. آن‌ها نشان دادند که کمترین تراکم متیلاسیون در جزایر CpG واقع در راه‌انداز ژن‌ها است، اما در بدنه ژن‌ها[1] تراکم متیلاسیون جزایر واقع در اینترون بیش‌تر از اگزون و UTRs بود. جزایر CpG متیله شده به‌طور قابل ملاحظه‌ای در نواحی بین ژنی[2]توزیع شده و اکثرا دارای طول ۲۰۰ تا ۳۰۰ جفت باز بودند. این پژوهش‌گران اذعان کردند که جزایر غیرمتیله بیشتر در نواحی راه‌انداز قرار دارند. کوآک و همکاران (۲۰۱٤) نیز در پژوهشی رخ‌نمای متیلاسیون را در ژنوم خوک بررسی نمودند که نتایج آنها مشابه نتایج به دست آمده در ژنوم جوجه گوشتی (هو و همکاران ۲۰۱۳) بود. همچنین در پژوهشی که کلدری و همکاران (۲۰۱٤) بر روی گوسفند انجام دادند، نواحی نزدیک TSS ژن‌ها دارای متیلاسیون پایین‌تری بود. رخ‌نمای کلی متیلاسیون در اکثر پستانداران و حتی پرندگان مشابه هم می‌باشد و به نظر می‌رسد که بین گونه‌های مختلف به صورت محافظت شده است (ساتی و همکاران ۲۰۱۲؛ هوآنگ و همکاران ۲۰۱٤؛ کلدری و همکاران ۲۰۱۵). در پژوهش هان و همکاران (۲۰۰۸) در ژنوم ۱۰ پستاندار شامل انسان، شامپانزه، میمون ماکاک، موش، موش صحرایی قهوه‌ای، سگ، گاو، اسب، صاریغ و پلاتی‌پوس (نوک اردکی) با استفاده از الگوریتم تاکایی و جونز رخ‌نمای کلی متیلاسیون پیش‌بینی و با هم مقایسه شدند. این محققین دریافتند با وجود این که تعداد ژن‌های کدکننده پروتئین در اکثر این ژنوم‌ها مشابه است، اما تعداد جزایر CpG و تراکم جزایر CpG در بین این ژنوم‌ها بسیار متغیر است. ژنوم سگ دارای بیشترین تعداد جزایر CpG 58327 در بین ژنوم‌ها بود و تراکم جزایر CpG در ژنوم پلاتی‌پوس (CpG islands/Mb 9/35) بیشترین بود. در پژوهش هان و همکاران (۲۰۰۸) تعداد ژن در این ژنوم‌ها بین ۲۰۰۰۰ تا ۳۰۰۰۰ پیش‌بینی شده است ولی تعداد جزایر CpG از ۱٤۰۰۰ تا ۵۹۰۰۰ متغیر بود و این امر می‌تواند بیان‌گر این حقیقت باشد که در ژنوم پستانداران توزیع جزایر CpG در قسمت‌های مختلف ژنوم (مثلا ژن‌ها و نواحی غیر کد کننده) متفاوت است. در پژوهش جداگانه‌ای هان و ژائو (۲۰۰۸) به بررسی جزایر CpG در ژنوم چهار ماهی تترداون، استیکِل بَک، مداکا و زبرافیش پرداختند. آن‌ها دریافتند که تعداد و تراکم جزایر CpG در بین ژنوم این چهار ماهی بسیار متغیر است. تعداد جزایر دامنه‌ای از ۲۱۵۲۲ (مداکا) تا ٦۱۱۷٦۸ (استیکِل بَک) داشت. تغییر در تراکم جزایر CpG در ماهی‌ها بسیار بیشتر از پستانداران است. نسبت ObsCpG/ExpCpG در ژنوم ماهی از۰/٤۷۹ تا ۰/٦٦۲ برآورد شد که نسبت به ژنوم پستانداران بیشتر بود. متیلاسیون و در نتیجه دی‌آمینوسیون رویه اصلی کاهش CpG در ژنوم مهره‌داران خون‌گرم است. نسبت ObsCpG/ExpCpG در ژنوم ماهی‌ها ممکن است نشان‌دهنده تأثیر مشابه متیلاسیون و دی‌آمینوسیون در ژنوم مهره‌داران خونسرد با میزان کمتر باشد. در پژوهش هان و ژائو (۲۰۰۹) تعداد جزایر CpG و تراکم جزایر CpG در ژنوم سگ بیشتر از ژنوم انسان و موش بود. اما بیشتر این جزایر در نواحی غیر کد کننده (اینترون و نواحی بین ژن) قرار داشتند و ژنوم سگ نسبت به انسان و موش، تعداد جزایر CpG مرتبط با نواحی راه‌انداز ژنی کمتری را داشت. بررسی‌های بیشتر ژن‌های همولوگ سگ، انسان و موش نشان داد که ممکن است ژنوم سگ نسبت به انسان و موش دارای نرخ کاهش و فرسایش بیشتر در جزایر مرتبط با راه‌انداز باشد (هان و ژائو ۲۰۰۹). در پژوهش هان و ژائو (۲۰۰۸) همبستگی بین تراکم جزایر CpG و سایر ویژگی‌های ژنومی بررسی گردید. آن‌ها در پستانداران

1 Gene Body
2 Intergenic Regions

یک همبستگی مثبت و معنی داری بین تراکم جزایر CpG و تعداد جفت کروموزوم (r=۰/۸۸, p=۷/۹×۱۰-٤) مشاهده کردند. همچنین همبستگی مثبت بین تراکم جزایر CpG و تعداد بازوهای کروموزومی[1](p=۰/۰۳۷, r=۰/٦۲) مشاهده گردید. همبستگی مثبت و معنی‌داری نیز بین تراکم جزایر CpG و نسبت ObsCpG/ExpCpG (r=۰/٦۳, p=۰/۰۳۵) برآورد شد. آن‌ها مشاهده کردند که همبستگی بین تراکم جزایر CpG با اندازه ژنوم و محتوی GC معنی‌دار نیست. همچنین همبستگی تراکم جزایر CpG با لگاریتم در مبنای ۱۰ اندازه کروموزوم[2] منفی و بسیار معنی‌دار (r=-۰/۵۱ ,p=۲/٦×۱۰-۲٦) برآورد گردید و همبستگی مثبت و بسیار معنی‌دار بین تراکم جزایر CpG با محتوی GC و نسبت ObsCpG/ExpCpG در سطح کروموزوم (به ترتیب ۲۸-۱۰×۳/۵=p ,r=۰/٦۵ و ٤۱-۱۰×۲/۸=p ,r=۰/۷۵) مشاهده شد. هان و ژائو (۲۰۰۸) کروموزوم‌ها را براساس اندازه گروه‌بندی نمودند و مشاهده شد که با افزایش سایز کروموزوم تراکم جزایر CpG کاهش می‌یابد. همچنین جزایر CpG که در نواحی بین ژنی ژنوم‌ها قرار داشتند استخراج شد و همبستگی بین تراکم CpG در این جزایر با لگاریتم در مبنای ۱۰ اندازه کروموزوم، محتوی GC و نسبت ObsCpG/ExpCpG در سطح کروموزوم به ترتیب ۰/۵۵-، ۰/۳۹ و ۰/٦۷ پیش‌بینی شد. آن‌ها نشان دادند که همبستگی بین جزایر CpG داخل ژنی و نسبت ObsCpG/ExpCpG (۰/۹۲) بیشتر از همبستگی بین جزایر CpG مرتبط با نواحی بین ژنی و نسبت ObsCpG/ExpCpG (۰/۹۱) است و همچنین همبستگی بین نرخ نوترکیبی[3](۵ Mb window) و تراکم جزایر CpG در انسان، موش و موش‌خرما معنی‌دار و به ترتیب ۰/۳۳، ۰/۲٤ و ۰/۱۷ برآورد شد. پژوهش‌گران (هان و ژائو، ۲۰۰۸) مشاهده نمودند که نرخ نوترکیبی از سانترومر به سمت تلومر در سطح کروموزوم و همچنین تراکم جزایر CpG از سانترومر به سمت تلومر افزایش می‌یابد و این موضوع دلیل احتمالی همبستگی مثبت بین نرخ نوترکیبی و تراکم جزایر CpG عنوان گردید. بررسی همبستگی بین تراکم جزایر CpG و تعداد جفت کروموزوم‌ها در غیر پستانداران غیر معنی‌دار مشاهده گردید(هان و ژائو ۲۰۰۸). هان و ژائو (۲۰۰۸) در پایان همبستگی بین تراکم جزایر CpG و محتوی GC در سطح کروموزوم پستانداران و غیر پستانداران را بررسی کرده و الگوی توزیع مجزایی را بین غیر پستانداران، به‌ویژه ماهی‌ها مشاهده نمودند. علاوه بر موارد ذکر شده، همبستگی بین تراکم جزایر CpG با درجه حرارت بدن و طول عمر توسط هان و ژائو (۲۰۰۸) بررسی گردید. همبستگی (p<۰/۰۵) بین تراکم جزایر CpG و درجه حرارت بدن[4] ۰/٦۵ برآورد گردید. اما همبستگی بین تراکم جزایر CpG و طول عمر[5] معنی دار نبود. در این بررسی ژنوم هر کدام از ماهی‌ها دارای توزیع جداگانه و منحصر به فرد بود . همبستگی بین تراکم جزایر CpG با لگاریتم در مبنای ۱۰ اندازه کروموزوم (۲۳-۱۰×۵/۵=p، ۰/۸۱-=r)، با محتوی GC (۱۰-۵۰ × ۷/۹=p، ۰/۹٦=r) و با نسبت ObsCpG/ExpCpG (۱۰-۲۸ × ۲/٤=p، ۰/۸٦=r) در سطح کروموزوم برآورد شد. ارتباط بسیار معنی‌دار و قوی بین تراکم جزایر CpG و محتوی GC در سطح کروموزوم بیان‌گر آن است که محتوی GC فاکتور ژنتیکی مهم و تاثیرگذاری بر تراکم جزایر CpG است (برازنده ۲۰۱٦). وانگ و همکاران (۲۰۱۳) چگونگی سطوح متیلاسیون CpG در راه‌انداز ژن CD4 در گاوهای هلشتاین چینی مبتلا به ورم پستان بالینی (CM) را مورد بررسی قرار داده و با گروه شاهد سالم مقایسه کردند. پژوهش نشان داد که ناحیه راه‌انداز در ژن CD4 گاوهای مبتلا، ۱٦٪ گروه متیل بیشتر در مقایسه با گاو های گروه شاهد داشتند. کاهش سطوح

[1] Chromosome arms
[2] Log10(chromosome size)
[3] Recombination Rate
[4] Body Temperature
[5] lifespan

بیان ژن CD4 در گاو های مبتلا ممکن است به علت افزایش سطح متیلاسیون DNA در راه‌انداز ژن CD4 باشد. تجزیه و تحلیل دو بعدی خوشه‌ای نشان داد که تفاوت‌های بزرگ در متیلاسیون راه‌انداز CD4 بین گاوهای مبتلا به ورم پستان و گاو سالم وجود دارد. در همه مقایسه‌ها سطح متیلاسیون DNA ژن CD4 به شدت از وضعیت ورم پستان تأثیر پذیرفته بود. پژوهش‌ها نشان می‌دهد که سطح متیلاسیون DNA در راه‌انداز CD4 می‌تواند به عنوان یک نشانگر مولکولی برای ورم پستان بالینی در گاوهای شیری استفاده شود (وانگ و همکاران ۲۰۱۳).

نتیجه‌گیری

از آن جایی که روش‌های سنتی اندازه‌گیری متیلاسیون DNA وقت‌گیر و پر زحمت است، مناسب پژوهش‌های پویش کل ژنوم نمی‌باشد. از طرف دیگر، به خاطر این که فن‌آوری‌های زیست ملکولی جدید نظیر ریزآرایه و توالی‌یابی نسل بعدی نیز هنوز گران هستند، استفاده عمومی از آن‌ها امکان‌پذیر نیست. در این پژوهش با استفاده از منابع موجود، جزایر CpG در ژن‌های کاندید موثر بر ورم پستان گاو شیری بررسی شدند. از میان سه نرم‌افزار مورد برررسی، نتایج دو نرم‌افزار SMS و Geneinfinity تقریبا مشابه بودند و احتمالا از یک معماری مشابه الگوریتمی استفاده می‌کنند. لازم به ذکر است که الگوریتم‌های به کار رفته در این پژوهش برای شناسایی جزایر CpG اختصاصی روی ژنوم گاو توسعه نیافته بودند. ژن‌های مورد بررسی در سطوح زیادی از پروتئین‌های خود دچار فرایند متیله شدن قرار می‌گرفتند. این مهم می‌تواند مورد کنکاش بیشتری در پژوهش آینده قرار گیرد. نتایج حاصله در این پژوهش به فراسنجه‌های الگوریتمی مختلفی وابسته هستند و تغییر این فراسنجه‌ها، می‌تواند نتایج به دست آمده را دستخوش تغییرات کند. آن چه که شایان اهمیت است این است که از آن جایی که وراثت‌پذیری ورم پستان در کل پایین است، این گونه بررسی‌ها، می‌تواند، نقش بهتری از عملکرد تکی و گروهی ژن‌ها را از نظر فرایند فراژنتیک آشکار کند.

منابع مورد استفاده

Auble DT and DeHaseth PL, 1998. Promoter recognition by *Escherichia coli* RNA polymerase: Influence of DNA structure in the spacer separating the− 10 and− 35 regions. Journal of Molecular Biology 202: 471-482

Barazandeh A, 2016. Whole genome Investigation of Camelidae genome to detect CpG Islands. PhD thesis.

Bird A, 2002. DNA methylation patterns and epigenetic memory. Genes and Development 16: 6-21.

Couldrey C, Brauning R, Bracegirdle J, Maclean P, Henderson HV and Mcewan JC, 2014. Genome-Wide DNA methylation patterns and transcription analysis in sheep muscle. PLoS ONE 9: 1-7.

Couldrey C, Brauning R, Henderson HV and McEwan JC, 2015. Genome-wide DNA methylation analysis: no evidence for stable hemimethylation in the sheep muscle genome. Animal Genetics 46: 185-189.

Du X, Han L, Guo AY and Zhao Z, 2012. Features of methylation and gene expression in the promoter-associated CpG islands using human methylome data. Comparative and Functional Genomics 2012.

Faulk C and Dolinoy DC, 2011. Timing is everything: the when and how of environmentally induced changes in the epigenome of animals. Epigenetics 6(7): 791-797.

Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL, Duan H and Domann FE, 2002. Role for DNA methylation in the control of cell type–specific maspin expression. Nature Genetics 31: 175-179.

Gardiner-Garden M and Frommer M, 1987. CpG islands in vertebrate genomes. Journal of Molecular Biology 196 (2): 261-282.

Goddard ME and Whitelaw E, 2014. The use of epigenetic phenomena for the improvement of sheep and cattle. Frontiers in genetics 5: 247.

Gung M and Pfeifer GP, 2015. Aging and DNA methylation. BMC Biology 13: 1-8.

Han L, Su B, Li WH and Zhao Z, 2008. CpG island density and its correlations with genomic features in mammalian genomes. Genome Biology 9: R79.

Han L and Zhao Z, 2008. Comparative analysis of CpG islands in four fish genomes. Comparative and Functional Genomics Article ID 565631, 6 pages.

Han L and Zhao Z, 2009. Contrast features of CpG islands in the promoter and other regions in the dog genome. Genomics 94: 117-124.

Head JA, 2014. Patterns of DNA methylation in animals: an ecotoxicological perspective. American Zoologist 54(1): 77-86.

Hu Y, Xu H, Li Z, Zheng X, Jia X, Nie Q and Zhang X, 2013. Comparison of the genome-wide DNA methylation profiles between fast-growing and slow-growing broilers. PLoS One 8 (2): e56411.

Huang YZ, Sun JJ, Zhang LZ, Li CJ, Womack JE, Li ZJ, Lan XY, Lei CZ, Zhang CL and Zhao X, 2014. Genome-wide DNA methylation profiles and their relationships with mRNA and the microRNA transcriptome in bovine muscle tissue (*Bos taurine*). Scientific Reports 4: 6546.

Illingworth R, Kerr A, DeSousa D, Jørgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R and Rogers J, 2008. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLOS Biology 6 (1): e22.

Irizarry RA, Wu H and Feinberg AP, 2009. A species-generalized probabilistic model-based definition of CpG islands. Mammalian Genome 20: 674-680.

Jia M, Gao X, Zhang Y, Hoffmeister M and Brenner H, 2016. Different definitions of CpG island methylator phenotype and outcomes of colorectal cancer: a systematic review. Clinical Epigenetics 8: 25.

Jiang C, Han L, Su B, Li WH and Zhao Z, 2007. Features and trend of loss of promoter-associated CpG islands in the human and mouse genomes. Molecular Biology and Evolution 24: 1991-2000.

Kangaspeska S, Stride B, Métivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F and Reid G, 2008. Transient cyclical methylation of promoter DNA. Nature 452: 112-115.

Kwak W, Kim J, Kim D, Hong JS and Jeong JH, 2014. Genome-wide DNA methylation profiles of small intestine and liver in fast-growing and slow-growing weaning piglets. Asian-Australasian Journal of Animal Sciences 27: 1532-1539.

Larsen F, Gundersen G, Lopez R and Prydz H, 1992. CpG islands as gene markers in the human genome. Genomics 13: 1095-1107.

Medvedeva YA, 2011. Algorithms for CpG islands search: New advantages and old problems. In: Mahdavi MA (Ed.), Bioinformatics - Trends and Methodologies, InTech, pp. 449-470.

Murdoch BM, Murdoch GK, Greenwood S, McKay S, 2016. Nutritional influence on epigenetic marks and effect on livestock production. Frontiers in Genetics, 7: 182.

Ohgane J, Wakayama T, Senda S, Yamazaki Y, Inoue K, Ogura A, Marh J, Tanaka S, Yanagimachi R and Shiota K, 2004. The Sall3 locus is an epigenetic hotspot of aberrant DNA methylation associated with lacentomegaly of cloned mice. Genes Cells 9: 253-260.

Reik W, Dean W and Walter J, 2001. Epigenetic reprogramming in mammalian development. Science 293 (5532): 1089-1093.

Shahdadi AA, Shariati MM, Nassiri MR, Zereh Daran S and Saghi DA, 2017. The study of factors affecting culling and survival distribution function in Iranian Holstein cows. Journal of Animal Science Researches (Agricultural Science) 28(1): 58-64. (In Persian)

Sati S, Tanwar VS, Kumar KA, Patowary A, Jain V, Ghosh S, Ahmad S, Singh M, Reddy SU and Chandak GR, 2012. High resolution methylome map of rat indicates role of intragenic DNA methylation in identification of coding region. PLoS One 7: e31621.

Shi SP, Qiu JD, Sun XY, Suo SB, Huang SY and Liang RP, 2012. PMeS: Prediction of methylation sites based on enhanced feature encoding scheme. PloS One 7(6): e38772.

Shames DS, Minna JD and Gazdar AF, 2007. DNA methylation in health, disease and cancer. Current Molecular Medicine 7: 85-102.

Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H and Held WA, 2005. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proceeding of National Academic Science 102: 3336-3341.

Su J, Wang Y, Xing X, Liu J and Zhang Y, 2014. Genome-wide analysis of DNA methylation in bovine placentas. BMC Genomics 15: 12.

Taberlay PC, Statham AL, Kelly TK, Clark SJ and Jones PA 2014. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. Genome Research 24: 1-13.

Takai D and Jones PA, 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proceedings of the National Academy of Sciences 99 (6): 3740-3745.

Wang XS, Zhang Y, He YH, Ma PP, Fan LJ, Wang YC, Zhang YI, Sun DX, Zhang SL, Wang CD, Song JZ and Yu Y, 2013. Aberrant promoter methylation of the CD4 gene in peripheral blood cells of mastitic dairy cows. Genetics and Molecular Research 12 (4): 6228-6239.

Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M and Schübeler D, 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genetics 39: 457-466.

Wrzodek C, Büchel F, Hinselmann G, Eichner J, Mittag F and Zell A, 2012. Linking the epigenome to the genome: Correlation of different features to DNA methylation of CpG islands. PLoS One 7: e35327.

Exploring structure, CPG islands, DNA methylation and protein of candidate genes influencing mastitis in dairy cows

M Ghaderi-Zefrehei[1*], MAmiri[2], L Hashemi[3] and F Samadian[1]

Received: November 23, 2017 Accepted: January 22, 2019
[1]Assistant Professor, Department of Animal Science, Faculty of Agriculture, Yasouj University, Yasouj, Iran
[2]MSc Student, and MSc Graduated, Department of Animal Science, Respectively, Faculty of Agriculture, Yasouj University, Yasouj, Iran
[*]Corresponding Author: mghaderi@yu.ac.ir

Introduction: Epigenetics is assumed as the most complex field of biological sciences. Generally, epigenetics includes mechanisms that not only affect gene expression, but also bring about functional changes in the next generation of mitotic cell division (autosomal inheritance) and meiotic cell division (generational inheritance), though DNA nucleotide structure is not changed (Barazandeh, 2016). DNA methylation refers to the enzymatic addition of a CH3 group to a cytosine residue in DNA strand, which occurs almost particularly at CpG dinucleotides (i.e., a cytosine located 5′ of a guanine) in animals. The enzymatic machinery responsible for DNA methylation includes a family of DNA methyltransferases (DNMTs), including the maintenance methyltransferase DNMT1 (responsible for copying pre-existing DNA methylation patterns to the new strand during mitosis) and the de novo methyltransferases DNMT3A/3B. DNA methylation is considered as epigenetics factor. In mammalian, DNA methylation is necessary for both embryonic development and stem cell differentiation and this phenomenon is basically occurring in CpG dinucleotide. CpG dinucleotide intends to make cluster over genome that is called CpG island. CpG islands are considered gene markers and represent an important feature of mammalian genomes. CpG islands vary greatly among mammalian genomes (Medvedeva, 2011). Some factors such as recombination rate and chromosome size might have influenced the evolution of CpG islands in the course of mammalian evolution. Whenever CpG island in gene's promoter is not methylated and suitable transcriptional factors are accessible, gene expression machine is activated and therefore gene is expressed. However, methylation of CpG islands in gene promoter with compact form of chromatin could lead to inhibition of gene expression (Shi et al., 2012). Hypermethlation of CpG islands could cause a range of diseases such as cancer. Epigenetic process could modify proteins after translation that actually refers to biochemical events inflicting on proteins. Some of these events include: methylation, phosphorylation, glycosylation, sulfation, sumoylation, and ubiquitinylation disulfide bonding.

Nutrition is seen as one of the greatest environmental determinants of an individual’s health. Although nutrient quantity and quality could impart direct effects, the interaction of nutrition with genetic and epigenetic is often overlooked despite being shown to influence different biological variation in mammals (Murdoch et al., 2016). Understating and unrevealing complex traits, such as those that are nutrition-related, to determine the genetic and epigenetic contributions toward a phenotype would be a formidable job to get it done. Similar to other epigenetic endpoints, patterns of DNA methylation could be susceptible to alterations by exposing to environmental treatments, including contaminants (Head, 2014). These sort of alterations may persist in the absence of the initial treatments as cells divide, and can even be inherited over generations provided that they occur in the germ line. Although our knowledge concerning patterns of DNA methylation in animals is increasing, a big gap remains in the literature, especially when it does come to species outside of those generally used for biomedical researches. Unlike DNA, epigenetic markers could be directly influenced by the environment factors; therefore, epigenetic markers have been shown to be pivotal mediators of phenotypic responses to environmental signals. To understand how

epigenetic modifications facilitate and affect gene expression, it is crucial to understand how and when the epigenome is established (Faulk and Dolinoy, 2011). While we do not fully understand how all of the complex epigenetic genome changes occur, yet some epigenetic information is sought to retain and transmit to the next generation.

Materials and methods: In this study, first of all, some articles and resources that describe candidate genes affecting mastitis were explored and some of the genes which were considered candidate genes in bovine mastitis were chosen. These genes included 14 interleukins (in exception of interleukin 8 and interleukin 9), interferon-gamma, and tumor necrosis factor (in total 13 genes). Their DNA sequences and other features of selected genes (for example exon numbering, gene length, guanine percentage and cytosine percentage) were obtained from NCBI database and saved as FASTA format. The C++ based software and Promoter 2.0 Prediction Server were used to extract characteristics of genes and detecting gene promoters, respectively. This software is generally used to predict transcription factor-biding sites of DNA sequences in vertebrata. In order to detect CpG islands, DBCAT, SMS, and Geneinfinity softwares were used. Also, primary structure of proteins derived from these genes was extracted from NCBI database and finally protein methylation was obtained by using PMeS software. Depending on the type of algorithm used to identify the CpG islands, the number and length of the CpG islands found to be different. Predicted CpGs were used as measurements to indirectly investigate the level of methylation over DNA level of genes.

Results and discussion: The investigated genes in terms of some features like total size based on number of nucleotides, number of exons, total annotated spliced exon length, and percent of GC content were quite different. The essence of gene heterogeneity from different features made grouping of these genes a cumbersome job. In this research, our overall objective was to locate the position and number of CpG in the promoter region of genes. Referencing of applied software to detect CpG islands, number and length of detected CpGs was different. Results shown that methylation was not evenly/ uniformly occurring in considered genes and in average, less than 50 % of CpG islands were in promoter region of genes and methylated CpG was located in intergenic regions. Also, the maximum amount of methylation was seen in interleukin 1, 2, and TNF gene, which are among the most important genes in circumventing mastitis. The degree of methylation in the proteins derived from the investigated genes showed fairly non-concurrence results with their DNA methylation. However, protein derived from interleukin 11 undergoes methylation in large portions of its original protein structure. It should be noted that different algorithm architectures may give different results (Barazandeh, 2016; Medvedeva, 2011).

At the DNA level, this gene also had the highest CpG in its promoter. This may indicate the complexity of epigenetic process with this gene. Therefore, this gene could be considered as a suitable candidate for performing demethylation experiments and proving biological drugs to circumvent mastitis. An integrative view to gene structure, DNA methylation and methylation of proteins derived from the investigated genes in this research would provide the better idea pertaining to methylation function in circumventing the mastitis in dairy cattle.

Conclusion: Out of the three software tested, the results of the two SMS and Geneinfinity software were almost similar and probably use a similar algorithmic architecture. It should be noted that the algorithms used in this study were not developed to identify specific CpG islands on the bovine genome. The genes under study had undergone a methylation process at many levels of their proteins. The results of this study depend on different algorithmic parameters and changing these parameters were diverse. As matter of fact that mastitis heritability is low in general, such studies could reveal a better role for single and group genes in the epigenetics process.

Key words: CpG Island, Dairy cow, Mastitis, Protein methylation