دانشکده کشاورزی

گروه زراعت واصلاح نباتات

پایان نامه جهت اخذ مدرک کارشناسی ارشد در رشته بیوتکنولوژی کشاورزی

عنوان

روش های محاسباتی برای پیش بینی آلترناتیو اسپلایسینگ

اساتید راهنما

دکتر علی حق نظری

دکتر مهدی صادقی

تدوین

وحید اصلان زاده

اسفند ۸۸

تقدیم به

پدر و مادر عزیزتر از جانم

خواهر و برادران مهربانم

و اساتید راهنمای ارجمندم

تشکر و قدردانی

سپاس بیکران یزدان پاک را، همو که قدرت اندیشیدن را به انسان عطا نمود و شکر هزاران هزار باره که موهبت علم آموزی را به بنده خود ارزانی داشت.

وظیفه خود می دانم که به حکم ادب و احترام، مراتب سپاس و قدردانی خود را در حق همه عزیزانی که مرا در به سرانجام رساندن این تحقیق همراهی نموده اند به جا آورم.

از استاد راهنمای ارجمندم جناب آقای دکتر علی حق نظری که تفکر علمی و تواضع در اوج بزرگی و قدرت را از ایشان آموختم و از راهنمایی‌های ایشان در طول دوران تحصیل بهره بی پایان برده‌ام سپاسگزارم. از استاد راهنمای دیگرم جناب آقای دکتر مهدی صادقی، استاد علم و اخلاق، به پاس صبر و متانت و راهنمایی های ارزشمندی که در انجام تحقیق و رفع مشکلات این پایان نامه داشته‌اند بی‌نهایت سپاسگزارم. از خانم دکتر مریم زارع میرک آباد، دکترای بیوانفورماتیک دانشگاه تهران، که تجربیات ارزنده‌شان را بی منت رهگشای کارهایم قرار دادند نهایت سپاس و قدردانی را دارم. از دوست بسیار عزیز و ارجمندم جناب آقای میثم حافظ پرست، دانشجوی دکترای دانشگاه مالایای مالزی، که همکاری‌ها و کمک‌های بی دریغشان در ابتدای شروع پروژه بسیار مفید و سازنده بود سپاسگزاری می‌کنم. از آقای دکتر رسولی استاد آمار دانشکده ریاضی نیز به پاس همکاری‌های ایشان در راستای پیش‌برد تحقیق اینجانب قدردانی می‌کنم. از جناب آقای دکتر بهرام ملکی زنجانی و آقای دکتر رضا فتوت که زحمت بازخوانی پایان نامه را بر عهده گرفتند و از آقای دکتر جلال صبا، مدیر گروه محترم و نماینده تحصیلات تکمیلی سپاسگزارم. از دوستان بسیار عزیزم آقایان زینالی، قادریان، نوروزیان، حسینی، رشید آبادی، فرهی، عیوضی، ایمانی، آقاپور، پناهی، نوایی، میرخلیل زاده و همکلاسی‌های عزیزم خانم‌ها اکبری، محمدی، لطفی و عسگری سپاسگزارم.

فهرست مطالب

عنوان **صفحه**

فصل اول: مقدمه و کلیات ۱

۱.۱. مقدمه ۲

۱.۲. توالی های حفاظت شده مرز اینترون اگزون ۴

۱.۳. مکانیسم پیرایش ۵

۱.۳.۱. اسپلایسوزوم ۵

۱.۳.۲. مسیر واکنش ۶

۱.۴. پیرایش متناوب ۷

۱.۴.۱. مدل های مختلف پیرایش متناوب ۸

۱.۵. موتیف های پیرایش ۹

۱.۵.۱. جایگاه های پیرایش ۹

۱.۵.۲. نقطه انشعاب ۱۲

۱.۶. تنظیم کننده های پیرایشی ۱۳

۱.۷. پیرایش متناوب و بیماری های انسان ۱۷

۱.۸. اهمیت پیرایش متناوب در گیاهان ۱۸

۱.۹. جهش اگزونی و اهمیت آن ۲۰

۱.۱۰. ساختار دوم RNA و پیرایش ۲۱

۱.۱۱. هدف تحقیق حاظر ۲۳

فصل دوم: مروری بر منابع ۲۴

۲.۱. روش های مطالعه پیرایش متناوب ۲۵

۲.۲. مطالعه پیرایش متناوب به کمک اطلاعات حاصل از داده های EST و mRNA ۲۵

۲.۳. تکنولوژی ریزآرایه ۳۰

۲.۴. مطالعات مقایسات ژنومیکسی ۳۷

۲.۵. پیشگوئی پیرایش متناوب با روش های Ab initio ۳۹

۲.۵.۱. الگوریتم های کلاس بندی اگزون ۴۰

۲.۵.۲. الگوریتم های پیدا کننده اگزون های جدید ۴۲

فصل سوم: مواد و روش ها ۴۶

۳.۱ . استخراج داده ها ۴۷

۳.۱.۱. روش ساختن دیتابیس ۴۷

۳.۲. Mutual information ۴۹

۳.۳. Information content ۵۰

۳.۴. سینرژی دو متغییره ۵۱

۳.۴.۱. برآورد $MI_{\psi}(X;Y;Z)$ ۵۱

۳.۵. پیشگوئی اگزون های کاست و دائمی ۵۳

۳.۵.۱. ساخت پروفایل امتیاز دهی ۵۳

۳.۵.۲. امتیاز دهی به توالی ها ۵۴

۳.۵.۳. روش تست کردن ۵۴

۳.۵.۳.۱. بدون انجام Cross validation (CV) ۵۴

۳.۵.۳.۲. با انجام CV ۵۷

فصل چهارم: بحث و نتیجه گیری ۶۰

۴.۱. وابستگی جایگاه های پیرایشی ۶۱

۴.۱.۱. وابستگی داخل جایگاه های پیرایشی ۶۱

۴.۱.۲. نحوه عمل احتمالی ساختارها ۶۳

۴.۱.۳. وابستگی بالا در جایگاه های پیرایش بخشنده نسبت به گیرنده ۶۴

۴.۱.۴. وابستگی بین جایگاه های پیرایش ′۳ و ′۵ ..... ۶۴

۴.۲. مدل پیشنهادی برای پدیده جهش اگزون ۶۶

۴.۳. سینرژی ۶۸

۴.۴. وابستگی در داخل مجموعه توالی های با طول ۱۲۰ نوکلئوتید ۶۹

۴.۵. پیشگوئی ۷۰

۴.۶. مقایسه با کارهای قبلی ۷۵

۴.۷. دلیل استفاده از داده های انسانی و موش در این تحقیق ۷۷

۴.۸. پیشنهادات ۷۹

فهرست منابع ۸۰

فهرست تصاویر

عنوان **صفحه**

شکل ۱.۱. توالی های مرز اینترون - اگزون ۵

شکل ۳.۱. شماتیکی از یک اگزون و اینترون های اطراف آن ۴۸

شکل ۴.۱. مدل پیشنهادی برای پدیده جهش اگزون ۶۷

فهرست نمودارها و جداول

عنوان **صفحه**

نمودار ۴.۱. وابستگی در جایگاه های پیرایش ′۳ و ′۵ اگزون های کاست (ca) و دائمی (co) انسان و موش ۶۱

نمودار ۴.۲. وابستگی بین مناطق مختلف جایگاه های پیرایش ′۳ و ′۵ اگزون های کاست و دائمی ۶۵

نمودار ۴.۳. نمودار سینرژی ۶۸

نمودار ۴.۴. نمودار میانگین وابستگی (Avg) هر موقعیت با ۱۱۹ موقعیت دیگر در اگزون های کاست نسبت به اگزون های دائمی ۷۰

جدول ۴.۱. فاکتور های *FP*، *FN*، *TP* و *TN* محاسبه شده توسط پروفایل های ProMI و ProIC بدون انجام CV ۷۲

نمودار ۴.۵. پیشگوئی اگزون های کاست بدون انجام CV ۷۲

نمودار ۴.۶. پیشگوئی اگزون های دائمی بدون انجام CV ۷۳

جدول ۴.۲. فاکتور های *FP*، *FN*، *TP* و *TN* محاسبه شده توسط پروفایل های ProMI و ProIC به کمک روش CV ۷۳

نمودار ۴.۷. پیشگوئی اگزون های کاست به کمک روش CV ۷۴

نمودار ۴.۸. پیشگوئی اگزون های دائمی به کمک روش CV ۷۴

روش های محاسباتی برای پیش بینی آلترناتیو اسپلایسینگ

چکیده

آلترناتیو اسپلایسینگ فرآیندی است که به موجب آن چندین ایزوفورم mRNA بالغ از یک mRNA نابالغ ایجاد می شود. آلترناتیو اسپلایسینگ به صورت های مختلف دیده می شود که رایج ترین فرم آن کاست اگزون می باشد. مطالعات نشان داده اند که توالی های اطراف جایگاه های پیرایش بیشترین اطلاعات را در مورد فرآیند جهش اگزون فراهم می کنند. به این دلیل یک سری از اگزون های کاست و دائمی موش و انسان برای پیشگوئی اگزون های کاست مورد آنالیز قرار گرفتند. در این پژوهش برای مقایسه توالی های اطراف جایگاه های پیرایش از روش Mutual information استفاده شد و مشاهده شد که جایگاه های متعلق به اگزون های کاست و دائمی با هم الگوی وابستگی متفاوتی نشان می دهند. در نهایت از این وابستگی برای پیشگوئی کاست اگزون استفاده شد. مقدار حساسیت و اختصاصیت حاصل از این پیشگوئی به ترتیب برابر ۵۸ و ۶۳ درصد بود. از این روش می توان برای پیشگوئی اگزون های کاست در ژن های مختلف استفاده کرد.

واژگان کلیدی: آلترناتیو اسپلایسینگ، اگزون کاست، اگزون دائمی، پیشگوئی، حساسیت، اختصاصیت

فصل اول

مقدمه و کلیات

۱.۱. مقدمه

توالی کد کننده یک ژن، توالی هایی از کدون های سه نوکلئوتیدی است که این توالی ها در نهایت توالی خطی آمینو اسیدی را تعیین می کنند. کدون کد کننده یک آمینو اسید بلافاصله در مجاورت کدون اسید آمینه بعدی در توالی کد کننده قرار دارد. این نحوه آرایش کدون ها در باکتری ها و فاژها صادق می باشد. اما در مورد ژن های یوکاریوتی همیشه این حالت صادق نمی باشد. در مورد ژن های یوکاریوتی، توالی کد کننده به طور متناوب با رشته های غیر کننده از هم جدا می شوند لذا بسیاری از ژن های یوکاریوتی حالت موزائیکی دارند که شامل بلوک هایی از توالی های کد کننده هستند که توسط بلوک هایی از توالی های غیر کد کننده از هم جدا شده اند، توالی های کد کننده را اگزون[۱] و توالی هائی که اگزون ها را از هم جدا می کنند را اینترون[۲] می نامند. بعد از رونویسی از ژن های یوکاریوتی، mRNA نا بالغ[۳] حاصل از رونویسی که حاوی اگزون و اینترون می باشد توسط کمپلکس پروتئینی به نام اسپلایسوزوم پیرایش می شود. در فرآیند پیرایش، اینترون ها از mRNA نابالغ حذف شده و اگزون های مجاور به هم دیگر متصل شده و mRNA بالغ بوجود می آید. شناسایی جایگاه های پیرایشی[۴] توسط اسپلایسوزوم[۵]، بوسیله توالی های حفاظت شده ای هدایت می شود که در انتهای اینترون ها قرار دارند.

[1] Exon
[2] Intron
[3] Pre-mRNA
[4] Splice site
[5] Spliceosome

اگزون های mRNA نابالغ در نتیجه پیرایش می توانند به حالت های مختلف به همدیگر متصل شوند که به این پدیده پیرایش متناوب[۱] گفته می شود. مطالعات به کمک داده های EST و cDNA نشان داده اند که بیش از ۳۵ درصد از ژن های آرابیدوپسیس و برنج (ونگ و برندل[۲]، ۲۰۰۶؛ چن[۳] و همکاران، ۲۰۰۷؛ خیو[۴] و همکاران، ۲۰۰۵) و بیش از ۷۵ درصد از ژن های انسان (جانسون[۵] و همکاران، ۲۰۰۳) پیرایش متناوب نشان می دهند. پیرایش متناوب منجر به تولید تعداد زیادی از ایزوفورم های mRNA بالغ و در نهایت ایزوفورم های مختلف پروتئینی از یک ژن می شود. این پدیده یکی از مهمترین عواملی است که منجر به تنوع پروتئینی در ارگانیسم ها می شود. اهمیت پیرایش متناوب از آنجا روشن می شود که حداقل ۱۵ درصد و شاید بیش از ۵۰ درصد از بیماری های ژنتیکی به دلیل جهشی که در توالی های حفاظت شده جایگاه های پیرایشی و یا توالی های تنظیمی پیرایشی[۶] صورت می گیرد ایجاد می شوند (کارتگنی[۷] و همکاران، ۲۰۰۲؛ کاسرس و کورنبلیت[۸]، ۲۰۰۲؛ فاستینو و کوپر[۹]، ۲۰۰۳؛ ماتلین و کلارک[۱۰]، ۲۰۰۵؛ کالنینا و همکاران[۱۱]، ۲۰۰۵). همچنین مطالعات اخیراً نشان داده است که این پدیده در پاسخ به بسیاری از فرآیند های زیستی و غیر زیستی در گیاهان و افزایش کیفیت و کمیت محصولات گیاهی نقش مهمی ایفا می کند (ردی[۱۲] و همکاران، ۲۰۰۷). توالی های حفاظت شده جایگاه های پیرایشی و عناصر تنظیمی اینترون ها

[1] Alternative splicing
[2] Wang and Brendel
[3] Chen
[4] Xiao
[5] Johnson
[6] Splicing regulatory elements
[7] Cartegni
[8] Caceres and Kornblihtt
[9] Faustino and Cooper
[10] Matlin and clark
[11] Kalnina
[12] Reddy

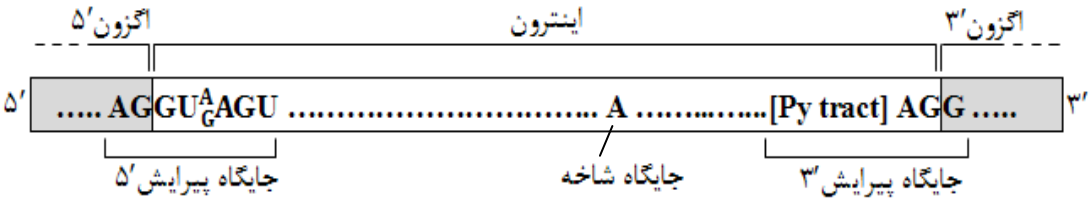
و اگزون ها در تمایز اگزون های متناوب از دائمی[1] برای انجام پیرایش متناوب صحیح نقش مهمی ایفا می کنند. به دلیل اهمیت بالای پدیده پیرایش متناوب در موجودات زنده، تعیین الگوی پیرایشی ژن ها می تواند نقش مهمی در شناسایی ژن های جدید و درک بهتر مکانیسم ها و صفاتی شود که با این پدیده درگیر هستند.

۱.۲. توالی های حفاظت شده مرز اینترون - اگزون

همانطور که گفته شد در عمل پیرایش، اینترون ها از mRNA نابالغ حذف شده و اگزون ها به هم دیگر می چسبند. این عمل توسط کمپلکس پروتئینی به نام اسپلایسوزوم که یک کمپلکس ریبونوکلئوپروتئینی[2] است صورت می گیرد. هدایت اسپلایسوزوم به منظور اتصال صحیح بر روی mRNA نابالغ توسط توالی های حفاظت شده صورت می گیرد. این توالی ها محل وقوع پیرایش را تعیین می کنند. بدهد. آنچنان که در شکل ۱.۱ نشان داده شده است مرز اگزون – اینترون که در انتهای ′۵ اینترون قرار دارد، جایگاه پیرایش ′۵[3] (دهنده[4]) و مرز مرز اینترون – اگزون در انتهای ′۳ اینترون نیز جایگاه پیرایش ′۳[5] (گیرنده[6]) نامیده می شوند. توالی سومی که برای عمل پیرایش ضروری می باشد نقطه انشعاب[7] نامیده می شود. این توالی در داخل اینترون و نزدیک به جایگاه پیرایش ′۳ واقع شده است. بعد از این جایگاه، توالی پلی پیریمیدینی[8] قرار دارد. توالی توافقی برای هر یک از این عناصر در شکل ۱.۱ نشان داده شده است. توالی حفاظت شده در جایگاه پیرایش ′۵، GU، در جایگاه پیرایش ′۳،

[1] Constitutive
[2] Ribonucleoprotein
[3] 3′ Splice site
[4] Donor
[5] 5′ Splice site
[6] Acceptor
[7] Branch point site
[8] Polypyrimidine tract

AG و در نقطه انشعاب، A می باشد. همه این نوکلئوتیدهای حفاظت شده در داخل توالی های اینترونی واقع شده اند.



شکل ۱.۱. توالی های مرز اینترون – اگزون: توالی های توافقی در هر دو جایگاه پیرایشی ′۵ و ′۳ و همچنین A حفاظت شده در نقطه انشعاب نشان داده شده است.

۱.۳. مکانیسم پیرایش

۱.۳.۱. اسپلایسوزوم

بیان شد که عمل پیرایش بوسیله کمپلکس پروتئینی بزرگی به نام اسپلایسوزوم صورت می گیرد. این کمپلکس متشکل از حدود ۱۵۰ پروتئین و پنج RNA می باشد و از لحاظ اندازه بسیار شبیه ریبوزوم می باشد. در انجام حتی یک واکنش ساده پیرایش، اسپلایسوزم چندین ملکول ATP هیدرولیز می کند. قابل توجه است که بسیاری از اعمال اسپلایسوزوم توسط اجزای RNA آن انجام می گیرد تا پروتئین ها که این نیز خود یک شباهت دیگر اسپلایسوزوم با ریبوزوم می باشد. این کمپلکس از پنج RNA کوچک هسته‌ای[1] (U1, U2, U4, U5, U6) تشکیل شده است. این RNA ها توالی های خاصی دارند که مشابه توالی های مرز اینترون – اگزون می باشند و بدین ترتیب در کاتالیز واکنش پیرایش دخالت دارند. هر یک از این RNA ها بین ۱۰۰ تا ۳۰۰ نوکلئوتید طول دارند و با چندین پروتئین

[1] Small nuclear RNAs (SnRNAs)

کمپلکس تشکیل می دهند. این کمپلکس های RNA – پروتئین را ریبونوکلئوپروتئین های کوچک هسته ای[1] می نامند (بلک[2]، ۲۰۰۳).

۱.۳.۲. مسیر واکنش

ابتدا جایگاه پیرایش ′۵ توسط U1 شناسایی می شود. سپس یک زیر واحد از U2AF به بخش پلی پیریمیدین و بقیه به جایگاه پیرایش ′۳ متصل می شوند. زیر واحد اول با پروتئین متصل شونده به نقطه انشعاب [3] (BBP) برهم کنش نشان داده و به آن پروتئین کمک می کند تا به نقطه انشعاب متصل شود. این آرایش پروتئین ها و RNA تحت عنوان کمپلکس E نامیده می شود. سپس U2 به نقطه انشعاب متصل می شود این کار به کمک U2AF و با تغییر مکان BBP انجام می شود این آرایش تحت عنوان کمپلکس A نامیده می شود. جفت شدن باز ها بین U2 و نقطه انشعاب به گونه ای است که نوکلئوتید A در جایگاه شاخه نقطه ای از رشته RNA به صورت یک برآمدگی کوچک بیرون می زند طوری که این نوکلئوتید به راحتی در دسترس قرار می گیرد تا با جایگاه پیرایش ′۵ واکنش نشان دهد. در مرحله بعد U4 و U6 به همراه U5 به کمپلکس متصل می شوند با ورود این سه جزء، کمپلکس A به کمپلکس B تبدیل می شود. در مرحله بعدی U1 از کمپلکس خارج می شود و U6 جای آن را در جایگاه پیرایش ′۵ می گیرد. این مراحل، مسیر همایش[4] پروتئین ها را کامل می کنند. در مرحله بعدی U4 از کمپلکس آزاد می شود و به U6 اجازه می دهد تا با U2 بر هم کنش نشان دهد. این آرایش که کمپلکس C نامیده می شود یک جایگاه فعال[5] را ایجاد می کند.

[1] Small nuclear ribonuclear proteins (snRNPs)
[2] Black
[3] Branch point binding protein (BBP)
[4] Assembly pathway
[5] Active site

تشکیل جایگاه فعال، جایگاه پیرایش ′۵ در mRNA نابالغ را در مقابل جایگاه شاخه نقطه ای قرار می دهد و اولین واکنش ترانس استریفیکاسیون[۱] را تسهیل می کند و واکنش دوم بین جایگاه پیرایش ′۳ و جایگاه پیرایش ′۵ به کمک U5 انجام می شود که این عمل باعث می شود تا دو اگزون در کنار هم دیگر قرار گرفته و به هم متصل شوند. در نهایت mRNA بالغ حاصل آزاد شده و اینترون نیز به صورت افسار اسب[۲] آزاد خواهد شد. snRNP ها به ساختار افسار اسب متصل باقی می مانند تا زمانی که قطعه اینترون حاصل تجزیه شود. بعد از آن دوباره snRNP ها باز چرخ خواهند شد (بلک[۳] ۲۰۰۳).

۱.۴. پیرایش متناوب

پیرایش متناوب فرآیندی است که در آن اگزون های mRNA نابالغ حاصل از رونویسی DNA در نتیجه پیرایش به چندین حالت مختلف می توانند به همدیگر متصل شوند نتیجه این عمل بوجود آمدن mRNA های بالغ مختلفی می باشد که ممکن است به ایزوفورم های پروتئینی متفاوتی ترجمه شوند بنابراین یک ژن می تواند چندین نوع پروتئین را کد کند (بلک، ۲۰۰۳). حدود ده سال پیش تصور بر این بود که ژن ها تنها یک ایزوفورم mRNA بالغ کد می کنند اما امروزه مشخص شده است که پیرایش متناوب به عنوان یک پدیده رایج در یوکاریوت ها رخ می دهد و تنوع پروتئین هایی را که توسط ژنوم کد می شوند را در سطح گسترده ای افزایش می دهد (گرولی[۴]، ۲۰۰۹). مدل های مختلفی از پدیده پیرایش متناوب مشاهده شده است که رایج ترین آنها جهش اگزونی[۵] می باشد. در این مدل یک اگزون خاص

[1] Transestrification
[2] Lariate
[3] Black
[4] Graveley
[5] Exon skipping

خاص ممکن است در mRNA بالغ در یک شرایط و یا بافت خاصی حضور داشته باشد و در سایر حالت ها حضور نداشته باشد (ماتلین[1] و همکاران، ۲۰۰۵). تولید mRNA های بالغ پیرایش شده متناوب بوسیله یک سری پروتئین های تنظیمی که به مناطق تنظیمی روی mRNA نابالغ متصل می شوند کنترل می شود. این پروتئین ها شامل فعال کننده های پیرایش[2] که استفاده از یک جایگاه پیرایش خاص را تشدید می کند و بازدارنده های پیرایش[3] که استفاده از یک جایگاه پیرایش خاص را کاهش می دهند، می باشند. مکانیسم های کنترلی پیرایش متناوب دارای تنوع بالایی می باشند و نمونه های جدیدی همچنان به دلیل استفاده از تکنیک های پیشرفته همانند روش ریزآرایه[4] معرفی می شوند. دانشمندان امیدوارند به طور کامل مکانیسم های تنظیمی که در پیرایش درگیر هستند را شناسایی کنند به طوریکه محصولات پیرایش متناوب یک ژن خاص تحت یک سری شرایط ویژه بوسیله کد پیرایش قابل پیشگویی باشد (بلک[5]، ۲۰۰۳).

۱.۴.۱. مدلهای مختلف پیرایش متناوب:

- جهش اگزون[6]: در این مدل یک اگزون ممکن است با عمل پیرایش روی mRNA نابالغ در mRNA بالغ حاصل از پیرایش حضور داشته باشد و یا حضور نداشته باشد.

[1] Matlin
[2] Splicing enhancers
[3] Splicing silencers
[4] Microarray
[5] Black
[6] Exon skipping or Cassette exon

- اگزونهای مانعه الجمع[1]: یکی از دو اگزون در mRNA بالغ بعد از عمل پیرایش حضور نخواهد داشت نه هر دو آنها.
- جایگاه پیرایش ′۵ متناوب[2]: یک جایگاه پیرایش ′۵ متناوبی مورد استفاده قرار می گیرد. این پدیده مرز جایگاه پیرایش ′۳ بالا دست را تغییر می دهد.
- جایگاه پیرایش ′۳ متناوب[3]: یک جایگاه پیرایش ′۳ متناوبی مورد استفاده قرار می گیرد. این پدیده مرز جایگاه پیرایش ′۵ بالا دست را تغییر می دهد.
- ابقاء اینترونی[4]: یک توالی اینترونی می تواند توسط عمل پیرایش حذف شود و یا در mRNA نابالغ حاصل از پیرایش حضور داشته باشد. این پدیده از مدل جهش اگزون به دلیل اینکه توالی ابقاء شده توسط اینترون ها محصور نشده است متمایز می شود (سامت[5]، ۲۰۰۸).

۱.۵. موتیف های پیرایش:

۱.۵.۱. جایگاه های پیرایش:

ترکیب نوکلئوتیدی مجاور جایگاه های پیرایش تصادفی نیست و جایگاه های پیرایشی متشکل از توالی های توافقی حفاظت شده می باشند. این توالی ها شامل ۹ باز برای جایگاه پیرایش ′۵ و حدود ۱۵ باز برای جایگاه پیرایش ′۳ می باشند. ماتریس PSSM[6] تشکیل شده شده از هزاران اینترون، ترکیبات نسبی هر باز در هر موقعیت را نشان می دهد و اجازه

[1] Mutually exclusive
[2] Alternative 5′ splice site
[3] Alternative 3′ splice site
[4] Intron retention
[5] Sammeth
[6] Position specific scoring matrices