

Applied molecular biology

Vol.3, issu.5, Winter. 2025

P.P 72-86

An overview of distance matrix phylogeny analysis

Masoomeh Hasanbarani¹, Shadi Hajrasouliha^{2*}

1. Department of Biology, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

2. Associate Professor, Pharmacology Research Center of Medicinal Plants, Tehran Islamic Azad University of Medical Sciences, Tehran, Iran (Corresponding Author; mh_plantbiology@yahoo.com).

Article history:

Received:13/02/2025

Revised:24/02/2025

Accepted:10/03/2025

Abstract

In phylogenetic analyses, evolutionary relationships between genetic materials are considered. These evolutionary relationships are determined by studying substitution, deletion, multiplication, and rearrangement mutations that are subject to natural selection. In fact, phylogeny is the concept of evolutionary relationships between organisms or their genes that have been derived from their common ancestors over time. Undoubtedly, by creating a phylogenetic tree, it will be easier to examine the evolutionary relationships between organisms. In phylogenetic analyses, there are methods for drawing trees. One of these methods is the distance-based method, which is influenced by Sokal and Sneath cluster algorithms. The general idea of this method is in a way that does not seem very practical. Among the most famous distance-based methods are UPGMA, Fitch-margoliash, Minimum evolution and Neighbor joining.

Keywords: Evolutionary, UPGMA, Minimum evolution, Neighbor-joining.

مروری بر آنالیز تبارزایی: ماتریکس فاصله (Distance matrix)

معصومه حسن بارانی^۱، شادی حاج‌رسولیه‌ها^{۲*}

۱. گروه زیست‌شناسی، واحد تهران مرکزی، دانشگاه آزاد اسلامی، تهران، ایران.
 ۲. استادیار، مرکز تحقیقات فارماکولوژی گیاهان دارویی، دانشگاه علوم پزشکی آزاد اسلامی تهران، تهران، ایران (نویسنده مسئول؛ hajrasouliha@iautmu.ac.ir).
- تاریخ بارگزاری: ۱۴۰۳/۱۱/۲۵ تاریخ بازنگری: ۱۴۰۳/۱۲/۰۶ تاریخ پذیرش: ۱۴۰۳/۱۲/۲۰

چکیده

در تجزیه‌های تبارزایی، روابط تکاملی بین مواد ژنتیکی مد نظر است. این روابط تکاملی به‌وسیله مطالعه جهش‌های جایگزینی، حذف، ازدیاد و جهش با آرایش مجدد تعیین می‌شود که در معرض انتخاب طبیعی قرار می‌گیرند. در واقع تبارزایی، به مفهوم روابط تکاملی بین موجودات یا ژن‌های آن‌ها می‌باشد که در طول زمان با فرایندهایی از اجداد مشترک خود مشتق شده‌اند. بی‌شک با ایجاد درخت تبارزایی امکان بررسی روابط تکاملی بین موجودات راحت‌تر خواهد بود. در تجزیه‌های تبارزایی برای ترسیم درخت‌ها روش‌هایی وجود دارد. یکی از این روش‌ها، روش‌های مبتنی بر فاصله (Distance based method) می‌باشد که متأثر از الگوریتم‌های خوشه‌ای Sokal و Sneath است. از معروف‌ترین روش‌های مبتنی بر فاصله UPGMA، Fich- margoliash، Minimum evolution و Neighbour joining می‌باشد.

کلمات کلیدی: روابط تکاملی، UPGMA، Distance، Minimum evolution و Neighbour joining.

مقدمه

فنتیک، روشی محبوب در اواسط قرن بیستم بود که در حال حاضر تا حد زیادی منسوخ شده است. تبارزایی یا تبارزایی شاخه‌ای در علم زیست‌شناسی است که به بررسی ارتباط تکاملی گروه‌های مختلف جانداران نظیر گونه‌ها یا جمعیت‌ها می‌پردازد که از داده‌های توالی‌یابی مولکولی و ماتریس داده‌های ریخت‌شناسی به دست می‌آید. [5]

واژه تبارزایی از ریشه یونانی phyll به معنای تبار و geneticus به معنی مربوط به زایش یا زایشی گرفته شده است. در آرایه‌شناسی، طبقه‌بندی، شناسایی و نام‌گذاری جانداران بسیار از تبارزایی کمک گرفته‌اند ولی همچنان از نظر روش‌شناسی و منطق از آن مجزا هستند. [2]

در تجزیه‌های تبارزایی، روابط تکاملی بین مواد ژنتیکی مد نظر است. هنگامی که روابط تکاملی بین موجودات مختلف به وسیله روش‌های زیست مولکولی مطالعه می‌شود، تبارزایی مولکولی نام دارد. تبارزایی مولکولی را می‌توان به وسیله روش‌های ایمونولوژیکی، الکتروفورز پروتئین، هیبریداسیون DNA ژنومی، توالی‌یابی نوکلئوتیدها و اسیدهای آمینه مطالعه کرد. روش‌های ایمونولوژیکی مبتنی بر اثرات متقابل آنتی‌ژن - آنتی‌بادی است و یک روش قوی برای جداسازی موجودات نزدیک به هم است؛ اگرچه تجزیه کمی آن مشکل است. به این واسطه امروزه از جایگزینی‌های آمینواسیدی و نوکلئوتیدی برای تعیین روابط ژنتیکی و بررسی‌های تبارزایی پروتئین‌ها و ژن‌ها استفاده می‌شود. از نظر تاریخی لینه اولین کسی است که به طبقه‌بندی گیاهان پرداخت و اصول طبقه‌بندی او براساس صفات ظاهری گیاهان بود. به طوری که سلسله‌مراتب طبقه‌بندی او شامل: سلسله، شاخه، رده، راسته، جنس و گونه گیاه بود. [13] دانشمندان سده هجدهم و اوایل سده نوزدهم این سلسله‌مراتب را درخت زندگی نامیدند. پس از آن طبقه‌بندی صورت گرفته به وسیله لینه به نام تبارزایی بیان شد که نه تنها شباهت بین گونه‌ها، بلکه روابط تکامل بین آن‌ها را نیز نشان می‌داد. [13]

ارتباط تکاملی بین موجودات به وسیله درخت تبارشناسی نمایش داده می‌شود از آنجا که تکامل در طول دوره‌های زمانی که به طور مستقیم قابل مشاهده نیست اتفاق می‌افتد

زیست‌شناسان مجبورند تبارزایی‌ها را با استنباط روابط تکاملی میان جانداران امروزی بازسازی کنند. سنگواره‌ها می‌توانند به بازسازی تبارزایی‌ها کمک کنند اما رکوردهای فسیلی اغلب بیش از حد کمیاب هستند و کمک خوبی نیستند؛ بنابراین زیست‌شناسان معمولاً به تجزیه و تحلیل موجودات امروزی برای شناسایی روابط تکاملی آن‌ها محدود هستند. روابط تبارزایی در گذشته از روی ویژگی‌های فنوتیپ که اغلب شاخصه‌های کالبدشناختی می‌باشند بازسازی می‌شدند. امروزه داده‌های مولکولی شامل پروتئین و رشته‌های DNA برای ساخت درخت تبارزایی استفاده می‌شوند. [14]

تبارزایی مولکولی در ابتدا به وسیله نوتال در سال ۱۹۰۴ به کار رفت. یکی از اهداف او قرار دادن انسان در یک موقعیت تکاملی مناسب نسبت به دیگر موجودات نخستین بود. اما از این روش خیلی استقبال نشد، که بیشتر به خاطر محدودیت‌های روشی بود. تا این که در اواخر دهه ۱۹۵۰ دوباره داده‌های مولکولی برای بررسی روابط تکاملی موجودات استفاده شد. امروزه درختان تبارزایی براساس اطلاعات حاصل از توالی‌های DNA و پروتئین ساخته می‌شوند تا اطلاعات مورفولوژیکی و یا حتی ایمونولوژیکی؛ زیرا: (۱) امتیاز دادن آن‌ها ساده‌تر است؛ (۲) هر نوکلئوتید یا آمینواسید را می‌توان به‌عنوان یک خصوصیت برای بررسی روابط تبارزایی در نظر گرفت؛ یعنی برای مثال با مقایسه دو توالی ۱۰۰۰ نوکلئوتیدی، ۱۰۰۰ نقطه قابل‌مقایسه در دسترس است، که در مقایسه با اطلاعات مورفولوژیکی این میزان بسیار زیاد است. در حقیقت داده‌های مورفولوژیکی پتانسیل اطلاعاتی کمتری نسبت به داده‌های مولکولی دارند؛ (۳) توالی‌های DNA و پروتئین، اطلاعاتی را هم برای حالت‌های انشعاب‌یافته و هم برای حالت‌های اجدادی در اختیار قرار می‌دهند. [1]

به علت توسعه تکنیک‌های پیشرفته استخراج توالی در بیولوژی مولکولی امکان جمع‌آوری مقدار زیادی اطلاعات جهت استنتاج فرضیه‌های تبارزایی فراهم شده است. به‌عنوان مثال پیدا کردن مقالاتی در مورد میتوکندری (با حدود ۱۶۰۰۰ نوکلئوتید در اکثر حیوانات) که بر پایه ماتریس کاراکتر هستند ساده است. به‌هرحال افزایش تاکسا خیلی اهمیت دارد؛ زیرا افزایش تاکساها به معنی افزایش دقت و در نتیجه ایجاد درخت تبارزایی قوی‌تر است. در تبارزایی ما انتظار دنباله‌های نسبتاً طولانی داریم و این یک

دلیل محکم برای این است که در تبارزایی داده‌های به‌دست آمده از سنگواره‌ها را هر جا که ممکن باشد استفاده می‌کنیم. البته داده‌های تبارزایی فسیل‌ها بیشتر شامل اطلاعات ریخت‌شناسی است تا اطلاعات مربوط به DNA [15].

Hillis و Zwickl در سال ۲۰۰۲ با استفاده از شبیه‌سازی نشان دادند نمونه‌های آرایه‌شناسی در استنتاج تبارزایی تأثیر مثبتی دارد و دقت آنالیز تبارزایی را افزایش می‌دهد. [19] بی‌شک با ایجاد درخت تبارزایی امکان بررسی روابط تکاملی بین موجودات راحت‌تر خواهد بود. در حقیقت زیست‌شناس‌ها با ایجاد درخت تبارزایی درصدد رسیدن به اهداف زیر هستند:

الف) بررسی و آزمون فرضیه‌های تکامل: تجزیه ساده داده‌ها نشان داده که موجودات خیلی نزدیک به هم توالی‌های مشابه دارند و موجودات خیلی دور از هم توالی‌های ناشناسی دارند. در نتیجه با بررسی توالی‌های مختلف می‌توان گره ژنی بین دو یا چند موجود را پیدا کرد و روابط تکاملی بین گونه‌ها را ترسیم کرد؛

ب) کسب اطلاعاتی درباره گونه‌ها و نسل‌های اجدادی منقرض شده؛

ج) طبقه‌بندی دقیق‌تر موجودات؛

د) برآورد شروع زمان انشعاب دو موجود. [1]

علم رده‌بندی در آغاز بر پایه مشاهدات ریخت‌شناسی (Morphologic) و شباهت ظاهری استوار بوده است که اولین نوع آن سیستم رده‌بندی مصنوعی بود که در آن از یک یا چند صفت ویژه بدون توجه به اهمیت این صفات برای رده‌بندی استفاده می‌شد. این نوع رده‌بندی از زمان ارسطو آغاز شد و لینه گیاه‌شناس سوئدی آن را بر مبنای تعداد تغییر داد. او گیاهان را بر حسب تعداد و وضع قرار گرفتن پرچم‌ها تقسیم کرد. از کارهای مهم او نامگذاری گیاهان و جانوران با سیستم دو نامی بود. در سیستم بعدی سیستم رده‌بندی طبیعی به بسیاری از همانندی‌های ریخت‌شناسی جانوران و با در نظر گرفتن میزان اهمیت هر یک از صفات توجه می‌شود؛ پس صفات بیشتری مورد استفاده قرار می‌گیرد. پس از آن رده‌بندی‌های جدیدتر براساس نظریه‌های لامارک و داروین صورت گرفت. سیستم رده‌بندی تبارزایی یا تبارزایی که جانداران را برحسب میزان تکامل رده‌بندی می‌کنند. در واقع در این رده‌بندی میزان ارتباط‌های ژنتیکی و

خویشاوندی جانداران با یکدیگر مورد استفاده قرار می‌گیرد. آخرین سیستم رده‌بندی سیستم رده‌بندی مدرن یا جدید است که در آن از مجموعه دانش‌ها (ریخت‌شناسی، تشریح، مقایسه‌های مولکولی تبارزایی) برای پی بردن به روابط خویشاوندی جانداران استفاده می‌شود. علم رده‌بندی به تنهایی اصل نسل مشترک را دربر نمی‌گیرد اما به همراه دیگر شواهد، اهمیت تکاملی رده‌بندی انکارناپذیر است. برای مثال تجزیه و تحلیل‌های ژنتیکی مشخص می‌کند که گونه‌های بسیاری که براساس ویژگی‌های تشریحی و دیگر صفات رده‌بندی شده‌اند، خویشاوندان خونی با زمینه وراثتی مشترکی هستند. مطالعه علمی گوناگونی موجودات زنده و ارتباط تکاملی آن‌ها با یکدیگر علم رده‌بندی (Systematic) نامیده می‌شود. یک جنبه مهم این علم آرایه‌شناسی (Taxonomy) است که علم نامگذاری، توصیف و طبقه‌بندی موجودات زنده است. اصطلاح طبقه‌بندی (Classification) به معنی رده‌بندی موجودات زنده به گروه‌ها، براساس همانندی‌ها یا خویشاوندی آن‌ها است. تاکسون (Taxon) (آرایه) که جمع آن Taxa می‌باشد، گروه‌بندی منظمی از موجودات زنده است که در سطحی معین نظیر گونه، جنس یا شاخه در نظر گرفته می‌شود. در تجزیه‌های تبارزایی برای ترسیم درخت‌ها دو روش متکی بر خصوصیات (Character-based methods) و روش فاصله‌ای (distance method) وجود دارد. به روش Distance based methods یا روش مبتنی بر فاصله در این تحقیق پرداخته می‌شود. [6]

روش‌های Distance matrix

Distance matrix، یکی از روش‌های اصلی آنالیز تبارزایی است که توسط Cavali- (1967) Sforza، (1967) Margoliash&Fitch و (1967) Horne ارائه شده است. ایده کلی این روش به نحوی است که زیاد کاربردی به نظر نمی‌آید. محاسبه فاصله میان هر جفت از گونه‌ها، و سپس یافتن درختواره‌ای که فواصل بین گروه‌ها را تا حد امکان نزدیک به هم پیش‌بینی می‌کند. این روش از تمامی اطلاعاتی که در مورد راسته‌های بالاتر، و خصوصیات ترکیبی آن‌ها وجود دارد، صرف‌نظر کرده و توجهی به آن‌ها ندارد. علاوه بر آن موجب می‌شود تا ماتریس داده‌ها، تا حد یک جدول دو به دو ساده

فاصله‌ای تنزل پیدا کند. در نتیجه خیلی‌ها تصور می‌کنند که این امر موجب خواهد شد تا زیرمجموعه‌هایی از داده‌های مفید ما بلااستفاده مانده و عملاً امکان بررسی و برآوردی تبارزایی صحیح از میان برود. باین حال، مطالعاتی که با استفاده از شبیه‌سازی رایانه‌ای صورت گرفته نشان می‌دهند، شمار داده‌هایی که در این روش از دست می‌روند فوق‌العاده کم و ناچیز است. برآوردهای تبارزایی انجام شده، بسیار دقیق بوده‌اند. [۳،۷،۹،۶]

در ظاهر، مرسوم نیست که برای فرایندهای تکاملی (یا حداقل برای مدل‌های ساده‌ای که ما از آن‌ها استفاده می‌کنیم) خصوصیات ترکیبی مربوط به راسته‌های بالاتر را حذف کنند، درحالی‌که همان ویژگی‌ها و داده‌ها را برای جفت فاصله‌های مابین گونه‌ها، مد نظر قرار دهند. بهترین راه درک و فهم روش Distance matrix این است که فاصله‌ها را به‌صورت تابعی از طول شاخه‌هایی در نظر بگیریم، که جفت‌گونه‌ها را از هم جدا می‌سازند. هر فاصله، بهترین درخت را برای هر جفت‌گونه استنتاج می‌کند. در واقع ما در اینجا تعداد زیادی از درختواره‌های متشکل از جفت‌گونه‌ها را خواهیم داشت، و سعی خواهیم کرد تا بر این اساس درختواره‌ای با n تعداد گونه را بیابیم. مشکل این کار آن است که هر یک از فاصله‌ها، دقیقاً همان طولی نیست که مابین هر جفت از گونه‌ها، در درخت n گونه‌ای وجود دارد. نتیجتاً باید گفت که انحرافی از فاصله واقعی وجود داشته و سعی ما باید این باشد که درختواره کامل و جامعه‌ای را پیدا کنیم که نزدیک‌ترین فاصله بین جفت‌گونه‌ها را با تقریب بالا داشته باشد. [۷]

از انواع روش‌های ایجاد ماتریس فاصله‌ای می‌توان به روش UPGMA، NeighborMatrix، Fitch-Margoliash اشاره کرد.

معرفی روش UPGMA

UPGMA یک درخت فرامتریک (درختی که طول تمام مسیرها از ریشه به سرشاخه‌ها یکسان باشد) را مد نظر می‌گیرد. اگر نرخ تکامل در تمام نسل‌های موجود در نمونه یکسان باشد (یک ساعت مولکولی) و اگر درخت به‌طور کامل متوازن باشد آنگاه UPGMA نباید یک نتیجه را تولید کند. این انتظارات در اکثر مجموعه‌های داده دیده

نمی‌شود و با وجود این که UPGMA تا حدی برای برطرف کردن این نقایص توانایی دارد اما معمولاً آن را برای تخمین تبارزایی به کار نمی‌برند. مزیت UPGMA سریع بودن آن است و می‌تواند با بسیاری از توالی‌ها کار کند.

مراحل ساخت درخت به روش UPGMA

الف) ساخت ماتریس فاصله؛

ب) تشکیل خوشه‌بین دو تاکسونی که کوتاه‌ترین فاصله را دارند؛

ج) محاسبه مجدد میانگین فاصله بین دو خوشه جدید و هر یک از تاکسون‌های دیگر و ایجاد یک ماتریس فاصله جدید؛

د) تکرار این فرایندها تا زمانی که همه تاکسون‌های مورد مطالعه در یک خوشه منفرد گروه‌بندی شوند. [۱۷]

معرفی Neighbor Joining

روش اتصال همسایه (Neighbour joining) یکی از پرکاربردترین روش‌های مبتنی بر فاصله است. [۱۸]

روش‌های اتصال مجاور (N-J) تکنیک‌های عمومی خوشه‌بندی داده‌ها را با استفاده از فاصله ژنتیکی به‌عنوان یک متریک خوشه‌بندی، در مورد آنالیز توالی‌ها استفاده می‌کند. روش ساده اتصال مجاور یک درخت بدون ریشه تولید می‌کند اما در این روش یک نرخ تکامل ثابت (ساعت مولکولی) در طی نسل‌های مختلف در نظر گرفته نمی‌شود. اما در روش UPGMA (روش جفت گروه بدون وزن با میانگین حسابی) درخت‌های ریشه‌دار تولید می‌شود و فرض را بر ثابت بودن نرخ تکامل می‌گذارد؛ در واقع درختی فرامتریک تولید می‌شود که در آن فاصله از ریشه تا هر سرشاخه یکسان است. [۱۸]

ویژگی‌های روش UPGMA و NJ

این دو روش در بین سایر روش‌های رسم درخت بر مبنای Distance هر دو ساده و سریع هستند. روش UPGMA که به‌صورت ساده و گسترده مورد استفاده قرار می‌گیرد

برای رسم درخت تبارزایی تاکسون‌هایی که دارای میزان ثابتی تکامل هستند مفید است و همچنین اجداد تکاملی را نشان نمی‌دهد. در روش NJ برخی از اطلاعات از دست می‌روند. به‌ویژه اطلاعاتی که درباره هویت اجداد و نوکلئوتیدهای مشتق شده در هر موقعیت می‌باشد. در این روش امکان محاسبه طول شاخه‌ها و ترکیب اطلاعات توالی‌ها وجود دارد. این روش وابسته به مدل تکاملی به کار رفته است. [۱۵]

روش Fitch- Margoliagh (FM)

در این روش نیز نیاز است که ابتدا با استفاده از ماتریس‌های مختلف امتیازدهی فاصله دو به دو بین دو توالی یا تاکسون را تعیین کنیم. این روش براساس معیار کوچک‌ترین مجذور عمل می‌کند. [۱۶]

روش Minimum Evolution (ME)

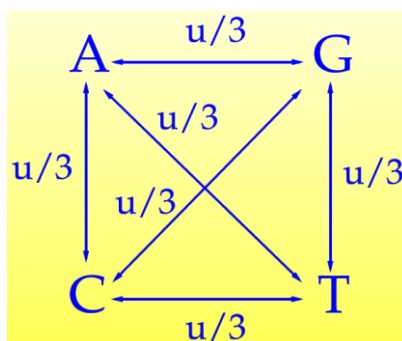
روش Minimum evolution، اولین بار توسط دو محقق کید و زونتا (kidd و Zonta) در سال ۱۹۷۱؛ مورد استفاده قرار گرفت. [۱۲] این روش درختی را جست‌وجو می‌کند که کوتاه‌ترین طول درخت را دارد، روشش مشابه FM است، که از طریق روش کوچک‌ترین مجذور عمل می‌کند. برخلاف FM، ME همه فواصل دو به دو احتمالی را به کار نمی‌گیرد. طول شاخه‌های داخلی براساس اندازه‌های حداقل بهینه می‌شود. [۸]

روش‌های مبتنی بر Distance برای رسم درخت

روش Fitch- Margoliagh و Minimum Evolution به‌عنوان بهترین روش‌ها ظاهر شده‌اند، و تقریباً در مطالعات به کار می‌روند. ME به‌طور گسترده در برنامه‌های کامپیوتری METREE و PAUP قابل اجرا است. روش NJ بسیار سریع است و به‌طور کلی یک درخت ارائه می‌دهد که نزدیک به درخت ME است. با توجه به ساختار اطلاعاتی ممکن است بی‌شمار درخت ارائه شود که بهتر و معنی‌دارتر از تنها درخت NJ باشند. [۸]

مدل تکاملی Jukes-cantor

ساده‌ترین مدل ممکن برای تغییر و تکامل توالی‌های DNA، مدلی است که جوکس و کانتور (Jukes و Cantor) در سال ۱۹۶۹، ارائه کردند. در این مدل، هر یک از بازهای موجود در توالی‌ها، از شانس یکسانی برای تغییر یافتن برخوردارند. زمانی که یک باز تغییر پیدا می‌کند، این امکان را دارد تا به هر یک از سه باز دیگر، (با احتمال مساوی) تبدیل شود. نتیجه مورد انتظار ما قطعاً به شکلی خواهد بود که از هر ۴ باز، با فراوانی یکسان، در DNA وجود داشته باشند. [۱۰]



مدل تکاملی Jukes-cantor (برگرفته از کتاب جوکس و کنتور (Jukes و Cantor)، ۱۹۶۹)

مقدار عددی u که میزان تغییرات را نشان داده و در کنار همه فلش‌ها به نمایش درآمده‌اند، در واقع میزان تعویض و جانشینی است که مابین تمامی جفت بازها صورت می‌گیرد؛ رخدادی است که در آن یک نوکلئوتید در یک جمعیت، جایگزین نوکلئوتید دیگری می‌شود [۱۲].

برای محاسبه فاصله‌ها در این مدل، لازم است تا احتمال هر گونه جایگزینی (Transition) را به دست آوریم. مفهوم احتمال transition (جایگزینی پورین‌ها با پورین‌ها و پیریمیدین‌ها با پیریمیدین‌ها) را نباید با جایگزینی متقاطع (transversion) (جهش از نوکلئوتید نوع پورین به پورین یا پیریمیدین و بالعکس) اشتباه گرفت و به معنای احتمال جابه‌جایی از یک موقعیت (مثلاً C) به موقعیت دیگر (مثلاً A) است. ساده‌ترین راه برای محاسبه این پدیده، آن است که کمی در این مدل دست ببریم. به جای آن که دامنه وسیعی از تغییرات (u) یک باز را به هر یک از سه باز دیگر داشته باشیم، بگذارید

این‌گونه تصور کنیم که دامنه‌ای از تغییرات یک باز ($4/3u$) را داریم که ممکن است به‌طور تصادفی هر ۴ باز موجود را دربر گیرد. در نتیجه دقیقاً حالتی پیش خواهد آمد که در آن احتمال تبدیل شدن یک باز، به سه باز دیگر، معادل $u/3$ خواهد بود. [۱]

اگر مدت زمان سپری شده و در واقع طول عمر شاخه را، معادل t در نظر بگیریم، آن‌گاه احتمالات مطرح در مدل دستکاری شده، به‌صورت رابطه $4/3ut$ خواهد شد.

شاخه‌های متعلق هر زمان (t) خود از تعداد پرشماری از زمان‌های کوتاه کوتاه (dt) تشکیل شده، که برای هر یک از آن‌ها نیز به‌خودی‌خود یک احتمال بروز رویداد جداگانه ($4/3ut dt$) وجود خواهد داشت.

احتمال این‌که هیچ تغییری در یک سایت در طول زمان رخ ندهد برابر است با $e^{-4/3ut}$. احتمال این‌که حداقل یک رویداد در یک سایت رخ دهد برابر است با $1 - e^{-4/3ut}$.

احتمال وقوع یک رویداد در یک نوکلئوتید خاص، یک‌چهارم خواهد بود؛ در نتیجه می‌توان گفت که احتمال وقوع یک رویداد (مثلاً C) در انتهای شاخه‌ای که با A آغاز شده است به‌صورت $\text{Prob}(C \text{ A. } u.t) = 1/4(1 - e^{-4/3ut})$ است. از آنجایی که ۳ نوکلئوتید وجود دارد که A می‌تواند به هر یک از آن‌ها تبدیل شود؛ پس احتمال $3/4(1 - e^{-4/3ut})$ وجود دارد. فواصل مورد انتظار (d) یک فاصله ژنتیکی حقیقی از ut است که به‌صورت

$$D = ut = -3/4 \ln(1 - 4/3d) \quad \text{است.} \quad d = 3/4(1 - e^{-4/3ut})$$

می‌شود. در واقع نسبت نوکلئوتیدهایی است که بین دو توالی متفاوت‌اند. تفاوت مورد انتظار در جایگاه بین دو توالی ژنی، در مدل Jukes-Cantor، به‌صورت تابعی از طول شاخه می‌باشد (در واقع حاصل میزان تغییرات و زمان تغییرات تکاملی). [۵]

روش‌های فاصله‌گذاری به‌طور عمده در بین گروهی از متخصصان سیستم‌های مولکولی رایج است اما با رشد فزاینده روش‌های کاراکتر پایه برخی مزایای روش فاصله‌گذاری رو به کمرنگ شدن دارند؛ با این وجود مزایای چون پیاده‌سازی تقریباً هم‌زمان NJ، توانایی همراه کردن یک مدل تکاملی با یک تحلیل نسبتاً سریع، روش‌های تخمین شبکه‌ای و مزایای دیگر تضمین می‌کند که همچنان برای مدتی طولانی روش‌های فاصله‌گذاری در کانون توجه باشند.

منابع

۱. رشیدی، س.، ملیوبی، م.، و نقوی، م. (۱۳۸۸). *بیوانفورماتیک*، مؤسسه انتشارات دانشگاه تهران.
2. Edwards, A.W.F., L.L. (1964). **Cavalli-Sforza Phylogenetics is that branch of life science, which deals with the study of evolutionary relation among various groups of organisms, through molecular sequencing data.** *Systematics Assoc. Publ. No. 6: Phenetic and Phylogenetic Classification (ed.). Reconstruction of evolutionary trees.* pp. 67-76.
3. Cavalli-Sforza, L.L. and Edwards, A.W.F. (1967). **Phylogenetic analysis: Models and estimation procedures.** *American Journal of Human Genetics* 19: 233-257. *Evolution* 21: 550-570.
4. Day, W H.E. (1986). **Computational complexity of inferring phylogenies from dissimilarity matrices.** *Bulletin of Mathematical Biology* 49: 461-467.
5. Edwards, A W.F. and Cavalli-Sforza, L.L. (1964). **Reconstruction of evolutionary trees.** pp. 67-76 in Phenetic and Phylogenetic Classification, ed. V. H. Heywood and J. McNeill. Systematics Association Publ. No.6, London.
6. Farris, J.S. (1972). **Estimating phylogenetic trees from distance matrices.** *American Naturalist* 106: 645-668.
7. Fitch, W. M. and Margoliash. E. (1967). **Construction of phylogenetic trees.** *Science* 155: 279-284.
8. Gascuel, O. (1997). **Concerning the NJ algorithm and its unweighted version, UNJ.** pp. 149-170 in *Mathematical Hierarchies and Biology*, ed. B. Mirkin, F. R. McMorris, and A Rzhetsky. American Mathematical Society, Providence, Rhode Island.
9. Horne, S.L. (1967). **Comparisons of primate catalase tryptic peptides and implications for the study of molecular evolution.** *Evolution* 21: 771-786.
10. Farris, J.S., Albert, V A., Kallersjo, M, Lipscomb, D., and Kluge, A.G. (1996). **Parsimony jackknifing outperforms neighbor-joining.** *Cladistics* 12: 99-124.
11. Jukes, T H. and Cantor, C. R. (1969). **Evolution. of protein molecules.** pp. 21-132 in *Mammalian Protein Metabolism*, Vol. III, ed. M. N. Munro. Academic Press, New York.
12. Kidd, K K., and Sgaramella-Zonta, L.A. (1971). **Phylogenetic analysis: Concepts and methods.** *American Journal of Human Genetics* 23: 235-252.
13. Linnaeus, C. (1735). **Systemae Naturae, sive regna tria naturae, systematics proposita per classes, ordines, genera & species.**
14. Pierce, Benjamin A. (2007-12-17). **Genetics: A conceptual Approach** (3rd ed.). W. H. Freeman. ISBN 978-0-7167-7928-5.
15. Saitou, N. and Neil, M. (1987). **The neighbor-joining method: A new method for reconstructing phylogenetic trees.** *Molecular Biology and*

Evolution 4: 406-425.

16. Sarich, V.M. (1969). **Pinniped phylogeny**. *Systematic Zoology* 18: 416-422.
17. Sokal, R.R. and Sneath, P.H.A. (1963). **Numerical Taxonomy**. W. H. Freeman, San Francisco.
18. Studier, J.A and Keppler, K.J. (1988). **A note on the neighbor-joining algorithm of Saitou and Nei**. *Molecular Biology and Evolution* 5: 729-731.
19. Zwickl DJ, Hillis, DM (2002). **"Increased taxon sampling greatly reduces phylogenetic error"**. *Systematic Biology*. **51** (4): 588-598. doi:10.1080/10635150290102339. PMID 12228001.