

Comparative phylogenetic tree reconstruction algorithms based on the *BRCA1* gene in different species

M. Ghaderi-Zefrehei^{1*}

1. Department of Animal Science, Faculty of Agriculture, Yasouj University, Yasouj, Iran

(Received: 13-04-2025 – Revised: 11-09-2025 – Accepted: 13-09-2025 – Available online: 20-09-2025)

Abstract

Introduction: The *BRCA1* gene is a pivotal candidate in current research endeavors, specifically drawing significant attention in studies related to bovine mammary tumors and human breast cancer. This gene is recognized for its fundamental biological role, as it encodes a protein that primarily functions as a tumor suppressor. Understanding the intricate evolutionary trajectory and the degrees of conservation or divergence of the *BRCA1* gene across various species is of paramount importance. Such an understanding can provide profoundly crucial insights into its underlying functional mechanisms and can offer invaluable comparative perspectives directly relevant to the development and progression of diseases in both human populations and livestock. Considering the substantial and impactful economic consequences that mammary tumors exert on agricultural productivity within cattle farming, alongside the pervasive and persistent challenge posed by breast cancer in human health, a detailed phylogenetic analysis of the *BRCA1* gene across a broad and representative spectrum of species becomes an imperative scientific undertaking. This research endeavor specifically aimed to meticulously reconstruct the evolutionary history of the *BRCA1* gene across a wide array of selected species. By successfully achieving this, it sought to contribute significantly to a deeper and more comprehensive understanding of its functional preservation across evolutionary time. Moreover, it is intended to explore its potential practical applications, which notably include informing more effective breeding strategies designed to foster disease-resistant livestock and guiding the development of novel diagnostic approaches pertinent to improving human health.

Materials and methods: In this computational study, the coding sequence (CDS) region of the *BRCA1* gene was precisely extracted from the genomes of 32 mammalian species from the NCBI database. The extracted sequences were then aligned using the BLAST algorithm and MEGA software. To reconstruct the evolutionary relationships, five phylogenetic methods were employed: DNAPARS, Pars, PhyML, Dnaml, and Proml. This comprehensive approach allowed for a multifaceted analysis of the gene's evolutionary history. The current version of Dnaml is faster than its previous iterations. This algorithm accounts for unequal predicted frequencies of the four DNA bases and allows for different predicted frequencies of transitions and transversions, along with various methods for incorporating different rates of interaction at distinct DNA sites. Key assumptions of this algorithm include the independent evolution of each DNA site and different lineages. Proml, optimized for protein data, offers robust statistical inference.

Results and discussion: The application of various selected phylogenetic algorithms consistently revealed that all methods clustered closely related species, such as *Bos taurus* and *Ovis aries*, into similar phylogenetic groups. This consistent grouping strongly confirms the conserved functional role of *BRCA1* in DNA repair and immune regulation. However, significant differences in phylogenetic tree topology were observed. Maximum likelihood methods (PhyML/Dnaml) grouped humans (*Homo sapiens*) with primates, showing bootstrap values of 65-75%. In contrast, maximum parsimony methods (DNAPARS/Pars) revealed weaker associations for rodents, with

* Corresponding author: mgahderi@yu.ac.ir



bootstrap values of 50-60%. The Proml algorithm, based on Bayesian inference, also showed uncertainty in estimating branch lengths. These differences highlight the sensitivity of the methods to the heterogeneous evolutionary rate of *BRCA1* and underscore the necessity of using combined mRNA and protein data to reduce methodological biases. The results showed that most mammals are grouped into closely related clusters; however, even within a single phylogeny, maximum likelihood did not always yield consistent topologies, hinting at complex evolutionary dynamics not fully captured by single gene analyses.

Conclusions: The findings of this study not only illuminate the intricate evolutionary history of *BRCA1* but also possess practical applications in breeding disease-resistant livestock and in human breast cancer research. Future research should investigate the role of horizontal gene transfer and its influence on *BRCA1* evolution. This study ultimately provides a framework for the evolutionary analysis of disease-resistance genes. Given the crucial role of the *BRCA1* gene in mammary tumors and breast cancer, the results of this study can provide a better evolutionary perspective across different species and be utilized in diagnostic assays for both humans and livestock. The study consistently showed that in most procedures and algorithms, bovine and ovine species form a very close evolutionary cluster. Therefore, the evolutionary results for one species can be applied to another. The *BRCA1* gene, essential for maintaining genomic stability through its role in DNA repair and cell cycle regulation, exhibits a high degree of evolutionary conservation, particularly in the N-terminal RING domain and C-terminal BRCT domains. This was demonstrated in this study, indicating that these vital regions are conserved among various vertebrate species, underscoring their functional importance.

Keywords: Phylogenetic analysis, Molecular evolution, Disease resistance, Bovine mastitis

Ethics statement: This article does not contain any studies with human participants or animals performed by any of the authors.

Data availability statement: The data that support the findings of this study are available on request from the corresponding author.

Conflicts of interest: The authors declare no conflicts of interest.

Funding: The authors received no specific funding for this work.

How to cite this article:

Ghaderi-Zefrehei, M. (2025). Comparative phylogenetic tree reconstruction algorithms based on the *BRCA1* gene in different species. *Animal Production Research*, 14(4), 59-80. doi: 10.22124/ar.2025.30337.1890



مقایسه الگوریتم‌های درخت تبارزایی روی ژن *BRCA1* در گونه‌های مختلف

مصطفی قادری زفره‌ئی*

۱- گروه علوم دامی، دانشکده کشاورزی، دانشگاه یاسوج

تاریخ دریافت: ۱۴۰۴/۰۱/۲۴ - تاریخ بازنگری: ۱۴۰۴/۰۶/۲۰ - تاریخ پذیرش: ۱۴۰۴/۰۶/۲۲ - تاریخ انتشار برخط: ۱۴۰۴/۰۶/۲۹

چکیده

ژن *BRCA1* به‌عنوان یک سرکوبگر تومور در مطالعات ورم پستان گاو و سرطان سینه انسان مورد توجه است. در این پژوهش، توالی CDS این ژن از ۳۲ گونه پستاندار از پایگاه NCBI استخراج شده و با استفاده از الگوریتم BLAST و نرم‌افزار MEGA هم‌رديف‌سازی شد. روابط تکاملی با پنج روش فیلوژنتیک DNAPARS، Pars، PhyML، Dnaml و Proml بازسازی شدند. نتایج نشان داد تمام روش‌ها، گونه‌های نزدیک مانند *Bos taurus* و *Ovis aries* را در یک خوشه قرار دادند که حفاظت عملکردی *BRCA1* در ترمیم DNA و تنظیم ایمنی را تأیید می‌کند. با این حال، تفاوت‌های معنی‌داری در توپولوژی درخت‌ها مشاهده شد: روش‌های حداکثر درست‌نمایی (PhyML/Dnaml) با بوت‌استرپ ۶۵-۷۵ درصد، انسان را با نخست‌سازان گروه‌بندی کردند، در حالی که روش‌های حداکثر صرفه‌جویی (DNAPARS/Pars) با بوت‌استرپ ۶۰-۵۰ درصد، روابط ضعیف‌تری برای جوندگان نشان دادند. الگوریتم بی‌زی Proml نیز عدم قطعیت در تخمین طول شاخه‌ها را آشکار ساخت. این تفاوت‌ها، حساسیت روش‌ها به نرخ تکاملی ناهمگون *BRCA1* را نشان می‌دهد و لزوم استفاده از داده‌های ترکیبی mRNA و پروتئین را برای کاهش سوگیری‌های روش‌شناختی برجسته می‌سازد. یافته‌های این مطالعه نه تنها تاریخچه تکاملی *BRCA1* را روشن می‌کند، بلکه کاربردهای عملی در پرورش دام‌های مقاوم به ورم پستان و مطالعات سرطان سینه انسان دارد. پژوهش‌های آینده باید نقش انتقال افقی ژن و تأثیر آن بر تکامل *BRCA1* را بررسی کنند. این مطالعه، چارچوبی برای تحلیل تکاملی ژن‌های مقاومت به بیماری ارائه می‌دهد.

واژه‌های کلیدی: تحلیل فیلوژنتیک، تکامل مولکولی، مقاومت به بیماری، ورم پستان گاوی

* نویسنده مسئول: mgahderi@yu.ac.ir

مقدمه

جهش‌های ژنی *BRCA1* و *BRCA2* در سلول‌های رده زایا، باعث خطر ابتلاء به سرطان پستان و تخمدان می‌شوند. ژن‌های *BRCA1* و *BRCA2*، عامل ۲۰٪ از موارد سرطان پستان ارثی هستند. این دو ژن عضو گروهی از ژن‌ها هستند که به خانواده ژن‌های سرکوبگر چنگار معروف هستند. از دست رفتن فعالیت هر یک از این ژن‌ها، سبب بروز برخی انواع بدخیمی‌ها و سرطان می‌شود. وقوع جهش‌های ژنتیکی در ژن‌های *BRCA1* و *BRCA2* باعث افزایش خطر بروز سرطان سینه و تخمدان در زنان می‌شود. نگاه تکاملی به این ژن‌ها، می‌تواند مناطق محافظت‌شده و عناصر عملکردی آن را که در فرآیندهای حیاتی مانند ترمیم DNA و مهار چنگار نقش دارند آشکار سازد (Jabbir, 2019). نشان داده شد که بررسی روابط تکاملی *BRCA1* می‌تواند به شناسایی جهش‌های مؤثر بر حساسیت به سرطان کمک کرده و نقش این ژن را در گونه‌های مختلف، از جمله دام و انسان، روشن کند (Paull et al., 2001). افزون بر این، بررسی تنوع ژنتیکی *BRCA1* در گونه‌های مختلف، می‌تواند اطلاعات مهمی برای برنامه‌های اصلاح نژاد به منظور افزایش مقاومت به بیماری در دام‌ها فراهم کند و در نتیجه، به پایداری کشاورزی و ارتقای سلامت حیوانات کمک نماید (Xu, 1997). از آنجایی که گاوها به عنوان مدل‌های زیست‌واره‌ای ارزشمندی برای بررسی بیماری‌های انسانی، به‌ویژه سرطان سینه، مطرح هستند، تحلیل تبارزایی می‌تواند به شناسایی نشانگرهای ژنتیکی و سازوکارهای مرتبط کمک کند که در پزشکی و دامپزشکی قابل استفاده خواهند بود. در زمینه بیماری‌های دامی، *BRCA1* می‌تواند بینش‌های جدیدی درباره ورم پستان گاوی ارائه دهد، زیرا واریانت‌های این ژن ممکن است با پاسخ‌های ایمنی مرتبط باشند. پژوهش‌ها نشان داده‌اند که چندشکلی‌های تک‌نوکلئوتیدی (SNPs) در *BRCA1* با حساسیت به ورم پستان همبستگی داشته و می‌توانند در انتخاب ژنتیکی برای افزایش مقاومت گاوها به این بیماری مؤثر باشند (Ilie, 2023). به‌طور مشابه، مطالعات نشان داده‌اند که سایر ژن‌های مرتبط با پاسخ ایمنی و سلامت پستان، مانند لاکتوفرین، نیز دارای چندشکلی‌هایی هستند که با ورم پستان و تعداد سلول‌های سوماتیک شیر در گاوها، از جمله گاوهای دورگ استان گیلان، مرتبط هستند (Ayatollahi et al., 2015). بررسی حفظ تکاملی

BRCA1 به درک بهتر نقش آن در فرآیندهای سلولی مرتبط با التهاب و پاسخ ایمنی که در توسعه ورم پستان حیاتی هستند، کمک می‌کند (Deb, 2013). شناسایی نشانگرهای ژنتیکی مرتبط با ورم پستان نیز می‌تواند در تدوین استراتژی‌های اصلاح نژاد برای بهبود سلامت پستان در گاوها نقش داشته باشد (Ilie, 2023). از منظر مدیریتی، بینش‌های حاصل از پژوهش‌های تبارزایی می‌توانند راهبردهای پیشگیری و کنترل ورم پستان، از جمله واکسیناسیون و درمان‌های آنتی‌بیوتیکی، را بهینه‌سازی کنند (Deb, 2013). علاوه بر این، *BRCA1* در تنظیم مسیرهای پیام‌دهی ایمنی، به‌ویژه در پاسخ به اینترفرون-گاما (γ -IFN) نقش دارد و از راه افزایش بیان *IRF-7*، بیان *STAT1* و *STAT2* را در واکنش‌های ایمنی ذاتی تقویت می‌کند (Buckley, 2007). پژوهش‌ها همچنین نشان می‌دهند که این ژن ممکن است در تعدیل پاسخ ایمنی ذاتی به عفونت‌هایی مانند استافیلوکوکوس اورئوس، یکی از عوامل شایع ورم پستان، مؤثر باشد، به‌گونه‌ای که فراسنجه‌های ایمنی مانند تولید سیتوکین‌ها و تعداد گلبول‌های سفید خون تحت تأثیر فعالیت *BRCA1* قرار دارند (Bannerman, 2008). در مجموع، تحلیل تبارزایی *BRCA1* نه تنها موجب درک بهتر نقش آن در سرطان و ایمنی دام می‌شود، بلکه می‌تواند راهکارهای اصلاح نژاد و مدیریت بیماری را برای بهبود سلامت حیوانات و افزایش مقاومت آن‌ها در برابر بیماری‌هایی مانند ورم پستان هدایت کند.

درخت‌های تبارزایی یک نمودار انشعابی و یا درختی است که روابط تکاملی در میان گونه‌های مختلف زیستی را نشان می‌دهند. تبارزایی به پژوهشگران اجازه می‌دهد تا ژن‌های مرتبط با ویژگی‌های مطلوب مانند بهره‌وری، مقاومت در برابر بیماری‌ها و کیفیت محصولات دامی را شناسایی کنند. بنابراین، این اطلاعات می‌تواند برای اصلاح نژاد و تولید دام‌هایی با ویژگی‌های بهبود یافته، استفاده شود. به‌عنوان مثال، شناسایی ژن‌های مرتبط با مقاومت در برابر بیماری‌ها می‌تواند به کاهش استفاده از آنتی‌بیوتیک‌ها و افزایش سلامت دام‌ها کمک کند. در مواردی که جمعیت‌های دامی زیر فشار انتخاب مصنوعی قرار دارند، ایجاد درخت‌های تبارزایی می‌تواند به حفظ تنوع ژنتیکی جمعیت‌ها در سازگاری با تغییرات محیطی و چالش‌های جدید کمک کند. ساخت و ایجاد درخت تبارزایی بر پایه

(al., 2001). با این حال، این روش نیاز به منابع محاسباتی قابل توجه دارد و ممکن است با مشکلات همگرایی مواجه شود، و اجزای بد همگرا شده ممکن است منجر به نتایج ناقابل اعتماد شوند (Ronquist et al., 2012). روش حداکثر صرفه‌جویی سعی در یافتن درختی دارد که کمترین تغییرات تکاملی را برای توضیح داده‌های توالی مشاهده شده نیاز داشته باشد، که آن را از نظر مفهومی، ساده و از نظر محاسباتی، برای مجموعه‌های داده‌های کوچک کارآمد می‌کند (Farris, 1970). با وجود سادگی آن، حداکثر صرفه‌جویی با مشکل جذب شاخه‌های بلند دست و پنجه نرم می‌کند و فاقد مدل‌سازی صریح نرخ‌های جایگزینی است، که دقت آن را برای سناریوهای تکاملی پیچیده محدود می‌کند (Felsenstein, 1978). در مقایسه این روش‌ها، روش‌های مبتنی بر فاصله از نظر کارایی محاسباتی برتری دارند، اما در دقت برای مجموعه‌های داده‌های پیچیده، کاستی دارند. روش حداکثر درست‌نمایی تعادلی بین دقت و عملی بودن محاسباتی برقرار می‌کند، که آن را برای بیشتر تحلیل‌ها مناسب می‌کند. استنتاج بیزی، بالاترین دقت و اندازه‌گیری‌های جزئی عدم قطعیت را ارائه می‌دهد، اما نیاز به منابع محاسباتی قابل توجه دارد و حداکثر صرفه‌جویی برای مجموعه‌های داده‌های کوچک بدون مدل‌های تکاملی واضح مناسب است. برای توالی‌های DNA، روش حداکثر صرفه‌جویی یا روش‌های مبتنی بر فاصله ممکن است برای مجموعه‌های داده‌های کوچک با توالی‌های نزدیک به هم کافی باشند، در حالی که روش درست‌نمایی یا استنتاج بیزی برای مجموعه‌های داده‌های بزرگ یا روابط تکاملی عمیق ترجیح داده می‌شوند (Yang, 2006). نشان داده شده است که برای توالی‌های AA، روش حداکثر درست‌نمایی یا استنتاج بیزی غالباً بهترین انتخاب هستند (Whelan & Goldman, 2001). دامنه سوال‌هایی که در حوزه تبارزایی مطرح می‌شود می‌تواند بسیار پیچیده باشد و نیاز به روش‌های خاصی داشته باشد که به سختی در حوزه بالا قرار گیرد. جدول ۱ تعدادی از این الگوریتم‌ها را نشان می‌دهد. در پژوهش حاضر برای درک ارتباط تکاملی گونه‌های مختلف بر اساس توالی ژن *BRAC1*، از دامنه‌ای از الگوریتم‌های مختلف استفاده شد.

الگوریتم TreeKO در دسته‌بندی روش‌های مبتنی بر فاصله قرار می‌گیرد. سایر الگوریتم‌ها و چارچوب‌ها (درختان تبارزایی تخصصی، درختان تبارزایی مبتنی بر سنتز، تحلیل

تحلیل روابط تکاملی بین گونه‌ها یا موجودات زنده استوار است. این فرآیند با جمع‌آوری داده‌های ژنتیکی یا ریخت‌شناسی آغاز می‌شود و پس از هم‌ردیف‌سازی توالی‌ها و انتخاب مدل تکاملی مناسب، به ساخت درخت می‌رسد. به‌طور کلی، روش‌های اصلی ساخت درخت شامل روش‌های مبتنی بر فاصله، روش‌های مبتنی بر صفات (مانند حداکثر صرفه‌جویی و حداکثر درست‌نمایی) و روش‌های بیزی هستند که هر یک با توجه به نوع داده‌ها و اهداف مورد نظر انتخاب می‌شوند. درخت‌های تبارزایی به دو شکل تبارشاخه (بدون در نظر گرفتن طول شاخه‌ها) و فیلوگرام (با نمایش طول شاخه‌ها به‌عنوان زمان یا میزان تغییرات تکاملی) و همچنین، به‌صورت ریشه‌دار یا بدون ریشه ارائه می‌شوند. ارزیابی آماری درخت‌ها با روش‌هایی مانند بوت‌استرپ انجام می‌شود تا دقت و اعتبار نتایج تضمین شود.

روش‌های مبتنی بر فاصله، فواصل جفتی بین توالی‌های زیستی را محاسبه کرده و از نظر محاسباتی سریع و قابل توسعه هستند، که آن‌ها را برای مجموعه‌های داده‌های بزرگ مناسب می‌کند (Saitou & Nei, 1987; Swofford et al., 1996). با این حال، این روش‌ها، نرخ‌های تکاملی ثابت را در بین توالی‌ها فرض می‌کنند، که ممکن است در بسیاری از موارد این مهم صادق نباشد و منجر به کاهش دقت شود، به‌ویژه در روابط تکاملی عمیق یا زمانی که توالی‌ها دارای نرخ‌های تکاملی بسیار متفاوت باشند (Felsenstein, 2004). روش حداکثر درست‌نمایی، توپولوژی درخت و طول شاخه‌ها را تخمین می‌زند که احتمال رصد توالی‌های داده‌شده را زیر مدل تکاملی مشخصی به حداکثر برساند و دقت بالا و انعطاف‌پذیری برای هر دو توالی‌های DNA و AA را با ادغام مدل‌های جایگزینی پیچیده مانند Jukes-Cantor، Kimura 2-parameter و GTR ارائه می‌دهد (Yang, 2006). با این حال، روش حداکثر درست‌نمایی از نظر محاسباتی پرهزینه است، به‌ویژه برای مجموعه‌های داده‌های بزرگ، و دقت آن به‌شدت به انتخاب مدل تکاملی بستگی دارد، و انتخاب نادرست مدل ممکن است منجر به نتایج پیش‌داوری‌شده شود (Posada & Buckley, 2004). استنتاج بیزی از روش‌های زنجیره مارکف مونت کارلو برای تخمین احتمالات پسین توپولوژی‌های درخت تبارزایی استفاده می‌کند، و تخمین‌های قوی از عدم قطعیت و مدیریت عالی مدل‌های تکاملی پیچیده را ارائه می‌دهد (Huelsenbeck et

در گونه‌های مختلف پستاندار و ارزیابی مقایسه‌ای الگوریتم‌های فیلوژنتیک مختلف شامل DNAPARS, Pars, PhyML, Dnaml و Proml در تعیین روابط تبارزایی بر اساس توالی CDS این ژن است. این پژوهش همچنین به دنبال روشن ساختن حفاظت عملکردی ژن *BRCA1* در ترمیم DNA و تنظیم ایمنی در گونه‌های مختلف، و بررسی کاربردهای عملی این یافته‌ها در بهبود سلامت دام (به‌ویژه افزایش مقاومت به ورم پستان گاوی) و پیشرفت مطالعات سرطان سینه انسان است.

مواد و روش‌ها

استخراج داده از بانک جهانی ژن *NCBI*: در این پژوهش ابتدا با استفاده از بانک جهانی ژن توالی‌های مربوط به ژن *BRCA1* گونه گاو استخراج شد. در نهایت و بعد از جستجوی ژن‌های هم‌تا مرتبط با این ژن، تعداد ۳۳ توالی

تبارزایی چندمرحله‌ای، سیستم مختصات شبکه‌ای برای درختان، الگوریتم جدید برای گروه‌های متقاطع و چارچوب iPhyloC (به‌طور انحصاری به یک روش تبارزایی خاص محدود نمی‌شوند و می‌توانند بسته به اهداف و داده‌های تحلیل، به‌همراه روش‌های مختلفی استفاده شوند. در کل، نرم‌افزارهای متعددی برای تحلیل ژنتیکی و رایانش روی داده‌های DNA از جمله: Convert, Formatomatic, GenAIEx, Fstat, GENEPop, MICROCHECK, Arlequin, STRUCTURE, CERVUS, TESS, SAMOVA, Bottleneck و LDNe وجود دارند که این ابزارها امکان انجام رایانش مختلفی و گاه‌ا شبیه به هم را در زمینه تحلیل DNA فراهم می‌کنند. بنابراین، هنر در تحلیل DNA می‌تواند این باشد که بتوان از برونادهای نرم‌افزارهای مختلف به‌صورت درون‌داد نرم‌افزارهای مختلف دیگر استفاده کرد و متکی به یک نرم‌افزار نبود. هدف مطالعه حاضر، بازسازی دقیق تاریخچه تکاملی ژن *BRCA1*

جدول ۱- گزیده‌ای از الگوریتم‌های نسبتاً جدید ایجاد درخت تبارزایی

Table 1. A selection of relatively recent phylogenetic tree reconstruction algorithms

Algorithm	Limitations	Strengths	Description
Specialized phylogenetic trees	Requires significant effort and expertise (Li, 2018)	High accuracy and specificity for the research in question	Constructed based on gene sequence data
Synthesis-based phylogenetic trees	May overestimate phylogenetic diversity; less designed for specific datasets (Li, 2018)	Requires less effort; useful for broad comparisons	Derived from existing phylogenetic trees
Multiphasic phylogenetic analysis	Complexity in understanding polynomial notation (Liu, 2022)	Provides a comprehensive and quantitative comparison of tree shapes	Uses polynomials to describe tree shapes
Network coordinate system for trees	May be challenging to apply to unlabeled trees (Liu, 2022)	Facilitates distance-based methods for clustering and parameter estimation	A framework for representing and comparing tree shapes and branch lengths
New algorithm for overlapping taxa	Limited to trees with overlapping taxa (Koshkarov & Tahiri, 2024)	Efficient and considers both branch lengths and topology	Compares trees with different but overlapping taxa
TreeKO algorithm	Complexity in implementation and interpretation (Marcet-Houben & Gabaldón, 2011) May require internet access	Manages incomplete mappings and provides consistency measures	Compares tree topologies in the presence of duplication
iPhyloC framework	May not work efficiently with very large datasets (Hammoud et al., 2021)	User-friendly, scalable, and supports non-binary trees	An interactive web-based tool for side-by-side tree comparison

مرحله هم‌ردیف‌سازی به‌طور دقیق مدیریت شود تا کلیت هم‌ترازی DNA حفظ شود. توالی‌های استخراج‌شده ابتدا با استفاده از الگوریتم Clustal W در نرم‌افزار MEGA هم‌ردیف‌سازی شدند. سپس، جهت اصلاح دقیق‌تر هم‌ردیف‌سازی و حذف خطاهای احتمالی، توالی‌ها در نرم‌افزار BioEdit ویرایش شدند. پس از اطمینان از کیفیت هم‌ردیف‌سازی، انواع روش‌های ساخت درخت تبارزایی در تارگاه (http://www.trex.uqam.ca/) TREX مقایسه شدند. در این راستا، دامنه‌ای از الگوریتم‌های مختلف (DNAPARS, Pars, PhyML, Dnaml, Dnamlk, Proml) و دو فراروایت حداکثر درست‌نمایی و حداکثر صرفه‌جویی مورد استفاده قرار گرفتند. همچنین، به‌منظور تحلیل دقیق‌تر محتوای توالی‌های به‌دست آمده از ابزار برخط استفاده FaBox (https://birc.au.dk/~palle/php/fabox) شد. این ابزار امکان بررسی و ویرایش دقیق توالی‌ها، از جمله شناسایی نواحی خاص و حذف بخش‌های غیرضروری، را فراهم می‌کند. لازم به ذکر است که در سراسر متن این مقاله، نام گونه‌ها به‌صورت فارسی و انگلیسی ذکر شده است. در صورتی که ترجمه به فارسی باعث ایجاد ابهام شود، بلادرنگ نام انگلیسی ارائه شده است تا خواننده به‌راحتی بتواند متن را درک کند.

نتایج و بحث

همان‌طور که در مقدمه اشاره شد، ژن *BRCAl* در تنظیم چندین سیتوکین کلیدی مرتبط با پاسخ‌های ایمنی گاوها، به‌ویژه در شرایط التهابی مانند ورم پستان، نقش مهمی ایفا می‌کند. این ژن با اینترفرون-گاما (*IFN- γ*) از راه همکاری با *STAT1*، رونویسی ژن‌های هدف را فعال کرده و نقش مهمی در تقویت پاسخ‌های ایمنی دارد (Ouchi et al., 2000). علاوه بر این، *BRCAl* موجب افزایش فعالیت رونویسی فاکتور نکروز تومور آلفا (*TNF- α*)، که یک سیتوکین پیش‌التهابی کلیدی در واکنش‌های ایمنی است، می‌شود (Benzra et al., 2003). همچنین، این ژن در تنظیم اینترلوکین-۶ (*IL-6*)، که بخشی از پاسخ سیستمیک در طول ورم پستان است، نقش دارد. برای درک بهتر نتایج، هر یک از الگوریتم‌های به‌کار رفته در این پژوهش، بعد از توضیح ضمنی هر یک از آن‌ها به‌همراه نتایج مورد نظر مورد کنکاش قرار می‌گیرد.

از گونه‌های مختلف زیر به‌صورت زیر فرمت FASTA استخراج و ذخیره شد: گاو اهلی (*Bos taurus*)، انسان (*Homo sapiens*)، موش خانگی (*Mus musculus*)، موش صحرایی نروژی (*norvegicus Rattus*)، مرغ خانگی (*Gallus gallus*)، سگ اهلی (*Canis lupus familiaris*)، میمون رزوس (*Macaca mulatta*)، شامپانزه (*Pan troglodytes*)، صاریغ کوتوله خاکستری (*Monodelphis domestica*)، خوک وحشی (*Sus scrofa*)، میمون خرچنگ‌خوار (*Macaca fascicularis*)، بابون آنوبیس (*Papio Anubis*)، بونوبو (*Pan paniscus*)، گیبون سفیدچانه (*Nomascus leucogenys*)، قورباغه پنجه‌دار غربی (*Xenopus tropicalis*)، راسو اهلی یا فِرِت (*Mustela putorius furo*)، همستر چینی (*Cricetulus griseus*)، خفاش ناتال (*Miniopterus natalensis*)، خفاش میوه‌خوار مصری (*Rousettus aegyptiacus*)، بلدرچین ژاپنی (*Coturnix japonica*)، مار زنگی فلسدار (*Protobothrops mucrosquamatus*)، چرخریسک بزرگ (*Parus major*)، مارموت آلپ (*Marmota marmota marmota*)، گکو ژاپنی (*Gekko japonicus*)، ماهی پاپی (*Cyprinodon variegatus*)، یوزپلنگ (*Acinonyx jubatus*)، ماهی مولی باله‌باله (*Poecilia latipinna*)، ماهی مولی مکزیکی (*Poecilia mexicana*)، آچلیک جنگنده (*Calidris pugnax*)، ماهی سالمون اطلسی (*Salmo salar*)، تیلاپیا نیل (*Oreochromis niloticus*) و گوسفند اهلی (*Ovis aries*).

ایجاد درخت تبارزایی: ابتدا، قسمت ژنومیک و مناطق کدکننده (CDS) ژن *BRCAl* مربوط به ژنوم گاوسانان از پایگاه داده NCBI با فرمت FASTA ذخیره شدند. از روی این دو نوع داده در نهایت تصمیم گرفته شد رایانش‌های بعدی روی توالی‌های CDS ژن *BRCAl* صورت گیرد. در ادامه، استفاده از ابزار BLAST و در نظر گرفتن مقادیر *E-value* معقول و پایین (از نظر آماری معنی‌دار)، توالی‌های مشابه به‌دست‌آمده شناسایی و ذخیره شدند. انتخاب توالی‌ها بر اساس چند معیار صورت گرفت: وجود توالی CDS در پایگاه NCBI، هم‌سانی نسبی طول توالی‌ها و امکان استخراج توالی از گونه‌های نزدیک به گاوسانان. در مواردی که طول توالی‌ها به‌دلیل حضور جهش‌های بزرگ حذف و درج متفاوت بود، تلاش شد تا این اختلافات در

از گونه‌ها، شامل ماهی‌ها، دوزیستان، پرندگان و پستانداران را نشان می‌دهد. درخت تبارزایی با گره‌هایی که نشان‌دهنده اجداد مشترک هستند و شاخه‌هایی که نشانگر دودمان‌های تکاملی‌اند، ایجاد شده است. مقادیر بوت‌استرپ در گره‌ها، نشان‌دهنده سطح اطمینان از این الگوهای شاخه‌بندی است. به‌طور خاص، بخش بالایی درخت، گونه‌های مختلف ماهی و دوزیست، مانند "*SalmoSalar*، *PoeciliaLa*" و "*XenopusTro*" را در یک گروه قرار می‌دهد که نشان‌دهنده پیوندهای تکاملی نزدیک‌تر آن‌هاست. پرندگانی مانند "*GallusGall*" و "*CoturnixJa*" یک تبارشاخه متمایز را تشکیل می‌دهند که نیای مشترک آنها را برجسته می‌کند. بخش پایینی عمدتاً شامل پستانداران است و تمایز روشنی از نخستی‌ها، از جمله "*HomoSapien*، "*PanTroglod*" و "*NomascusLe*" را نشان می‌دهد که نیای مشترک اخیر آن‌ها را به نمایش می‌گذارد. این درخت تبارزایی به‌طور مؤثر هم واگرایی‌های باستانی، مانند بین "*GallusGall*" و "*XenopusTro*" و هم واگرایی‌های جدیدتر، مانند بین انسان‌ها و شامپانزه‌ها را نشان می‌دهد و نمای کلی جامعی از تاریخ تکاملی و روابط ژنتیکی بین این گونه‌ها ارائه می‌دهد.

الگوریتم *Pars* این الگوریتم، باز در فراروایت حداکثر صرفه‌جویی، گروه‌بندی می‌شود. این الگوریتم روش صرفه-جویی واگنر را با چندین حالت انجام می‌دهد. روش صرفه-جویی واگنر امکان تغییرات را در بین همه حالت‌ها فراهم می‌کند. مانند روش قبل، هدف یافتن درختی است که به حداقل تعداد تغییرات نیاز دارد (Kluge & Farris, 1969). این الگوریتم بر اساس مفروضات زیر بنا شده است: حالات اجدادی ناشناخته است، جایگاه‌های DNA مختلف به‌طور مستقل تکامل می‌یابند، دودمان‌های مختلف به‌طور مستقل تکامل می‌یابند، تغییرات در تمام حالت‌های دیگر به یک اندازه محتمل است، این تغییرات از همان ابتدا، نسبت به بازه‌های زمانی تکاملی که در تمایز گروه مورد نظر دخیل هستند، غیرمحتمل هستند، سایر انواع رویدادهای تکاملی مانند حفظ چندشکلی، بسیار کمتر از این تغییرات، حالت محتمل هستند، نرخ تکامل در دودمان‌های مختلف به اندازه کافی کم است که دو تغییر در یک بخش بلند از درخت تبارزایی بسیار کمتر از یک تغییر در یک بخش کوتاه است. شکل ۲، درخت تبارزایی حاصل از این الگوریتم در داده‌های این پژوهش را نشان می‌دهد، و روابط تکاملی بین گونه‌های

الگوریتم *DNAPARS* این الگوریتم در زمره الگوریتم‌های فراروایت حداکثر صرفه‌جویی که درختان تبارزایی بدون ریشه را از روی توالی‌های DNA ایجاد می‌کند، قرار می‌گیرد. در کل، هدف این الگوریتم پیدا کردن توپولوژی درخت تبارزایی برای یک مجموعه توالی‌های هم‌ردیف شده است که با کمترین تعداد تغییر صفت (یعنی جهش‌ها) همراه است. الگوریتم یاد شده مانند سایر الگوریتم‌های حداکثر صرفه‌جویی با در نظر گرفتن یک توپولوژی خاص درخت تبارزایی، شروع به جستجو می‌کند. سپس این الگوریتم، کمترین تعداد تغییر صفت که برای توضیح همه گره‌های درخت تبارزایی در هر موقعیت توالی مورد نیاز است را محاسبه می‌کند و در ادامه، توپولوژی دیگر، ارزیابی می‌شود. وقتی همه توپولوژی‌های قابل قبول ارزیابی شد، درخت تبارزایی که نیاز به کمترین تعداد تغییر دارد به-عنوان بهترین درخت تبارزایی انتخاب می‌شود. الگوریتم‌های زیادی برای استخراج درخت تبارزایی در این حوزه وجود دارند که مفروضات آن‌ها عبارتند از: هر مکان DNA به‌طور مستقل تکامل می‌یابد، دودمان‌های مختلف به‌طور مستقل تکامل می‌یابند، احتمال جایگزینی ریشه در یک مکان معین در طول زمان در شاخه‌ای از درخت تبارزایی اندک است، مقادیر مورد انتظار تغییر در شاخه‌های مختلف تبارزایی آنقدر تغییر نمی‌کند که دو تغییر در یک شاخه با نرخ بالا محتمل‌تر از یک تغییر در یک شاخه کم نرخ باشد و مقادیر مورد انتظار تغییر در بین مکان‌های DNA به اندازه کافی متفاوت نیست که دو تغییر در یک مکان DNA بیشتر از یک تغییر در مکان *DNA* دیگر باشد. در شکل ۱ مشخص است که گونه گاوسانان *Bos taurus* با گونه گوسفندسانان *Ovis aries* تشکیل خوشه می‌دهند. به‌عبارت دیگر، این گونه‌ها کاملاً یک روال تکاملی مشابه‌ای را تجربه کرده‌اند. حتی در کنار این سازواره‌ها، بلافاصله گونه خوک *Sus scrofa* ظاهر شده است که نشان می‌دهد این گونه‌های اهلی، تشابه یکسانی از نظر تکاملی در ارتباط با ژن مورد بررسی دارند. در همین راستا، در کنار این گونه، بلافاصله گونه اسبان *Equus* ظاهر شده‌اند که باز دلالت بر نزدیکی این گونه‌ها از نظر تکاملی دارند. در فاصله بسیار دورتر، گونه‌هایی مثل *Gekko Japan* قرار دارند که نشان از این دارد این نوع الگوریتم توانسته است به‌طرز نسبتاً درستی، گونه‌ها را از نظر تکاملی گروه‌بندی کند. همچنین، این درخت تبارزایی، روابط تکاملی بین مجموعه‌ای متنوع

می‌کند. به‌طور کلی، این درخت تبارزایی، نمای کلی جامعی از تاریخ تکاملی و روابط ژنتیکی بین مجموعه‌ای متنوع از گونه‌ها ارائه می‌دهد و مقادیر بوت‌استرپ نشان‌دهنده اعتبار این روابط است. توپولوژی حاصل از الگوریتم Pars مانند توپولوژی حاصل از الگوریتم هم‌خانواده آن یعنی DNAPARS است که این مورد در شکل ۲ مشخص است. به عبارت دیگر، ساختار الگوریتم در این خانواده، تغییر خاصی در نتایج فراهم نیاورده است. لذا تا حد زیادی می‌توان نتایج حاصل از الگوریتم Pars را در الگوریتم DNAPARS مشاهده کرد. آنچه که بیشتر چشم می‌خورد این است که انسان گرچه یک پستاندار محسوب می‌شود، اما در هیچکدام از الگوریتم‌ها با سایر پستانداران از نظر تکامل، هم‌خوشه نشده است. موضوعی که نیاز به داده بیشتر برای بررسی علت این مهم دارد استفاده از اصل صرفه‌جویی، به‌عنوان یک ابزار شناختی، تضمین می‌کند که با توجه به داده‌ها، قوی‌ترین نتیجه تجربی را داشته باشیم. اتخاذ خلاصه‌ای از داده‌ها با توجه به مقایسه‌های برون‌گروهی تضمین می‌کند که قوی‌ترین نتیجه را می‌توان به‌صورت تبارزایی تفسیر کرد. به همین ترتیب، برازش داده‌ها با یک مدل، توضیحاتی را ارائه می‌دهد، اما تنها در صورتی که مدل شناخته شده یا درست فرض شود، یعنی هیچ وسیله‌ای وجود ندارد که روش‌های مبتنی بر مدل بتوانند صحت هر مدل معین یا مفروض‌های آن را آزمایش کنند. پژوهشگران معمولاً مدلی را برای مجموعه داده‌های معینی انتخاب می‌کنند که تنها بر اساس ترجیحات موردی برای سادگی هستی‌شناختی (هستی‌شناختی) است. با اضافه شدن داده‌های بیشتری به پژوهش، روش‌های احتمال، از جمله تحلیل درستیابی بیزی، عموماً با سرعت بیشتری به سمت پیچیده‌ترین مدل حرکت می‌کنند. اصل حداکثر آنتروپی در نظریه اطلاعات بیزی به‌وضوح نشان می‌دهد که پیچیده‌ترین مدل همه احتمالات را به‌صورت پیشینی اجازه می‌دهد، در حالی که اصل کمکی هیچ چیزی را از پیش منع نمی‌کند. با اضافه شدن داده‌های بیشتر به تجزیه و تحلیل، سه دسته از روش‌های کمی بازسازی تبارزایی شروع به همگرایی در همان پاسخ می‌کنند. اگر هر سه روش نتایج یکسانی داشته باشند، هیچ راهی یا حتی دلیلی برای انتخاب بین آنها وجود ندارد. نتایج متناقض در بخشی از درخت، ویژگی‌های مجموعه داده را برجسته می‌کند (به‌عنوان مثال، جذب شاخه طولانی، تنوع سریع و

مختلف، از جمله ماهی‌ها، دوزیستان، پرندگان و پستانداران را ارائه می‌دهد. درخت با گره‌هایی که نشان‌دهنده اجداد مشترک هستند و شاخه‌هایی که نشانگر دودمان‌های تکاملی‌اند، ساختار یافته است. مقادیر بوت‌استرپ در گره‌ها، نشان‌دهنده سطح اطمینان از این الگوهای شاخه‌بندی است. در بالای درخت، گونه‌های ماهی و دوزیست مانند "*PoeciliaLa*"، "*Cyprinodon*"، "*Nothobranc*"، "*Oreochromi*"، "*Oreochromi*"، "*SalmoSalar*"، "*Protobothr*" و "*GekkoJapon*" و "*RousettusA*" با هم گروه‌بندی شده‌اند که نشان‌دهنده پیوندهای تکاملی نزدیک‌تر آن‌ها است. دسته‌بندی این گونه‌ها، دودمان مشترک متمایز از سایر گروه‌ها در درخت را نشان می‌دهد. به‌سمت پایین درخت، پرندگانی مانند "*GallusGall*" (مرغ)، "*CoturnixJa*" (بلدرچین) و "*CalidrisPp*" مشاهده می‌شوند که یک تبارشاخه متمایز را تشکیل می‌دهند و دودمان مشترک آن‌ها را برجسته می‌کنند. دوزیست "*XenopusTro*" (قوریباغه) و پرند "*ParusMajor*" نیز نشان داده شده‌اند که واگرایی تکاملی آن‌ها را از سایر گونه‌های پرند نشان می‌دهد. بخش پایینی درخت عمدتاً شامل پستانداران است. گونه‌هایی مانند "*BosTaurus*" (گاو)، "*OvisAriesB*" (گوسفند)، "*SusScrofa*" (خوک)، "*EquusPrzew*" (اسب پرژوالسکی)، "*Miniopteru*" (خفاش)، "*CanisLupus*" (سگ)، "*MustelaPut*" (راسوی اهلی) و "*AcinonyxJu*" (یوزپلنگ) با هم گروه‌بندی شده‌اند که نشان‌دهنده روابط تکاملی نزدیک‌تر آن‌ها است. در میان نخستین‌ها، "*PanTroglood*" (شامپانزه)، "*PanPaniscu*" (بونوبو)، "*HomoSapien*" (انسان)، "*NomascusLe*" (گیبون)، "*MacacaMmu*" (میمون رزوس)، "*MacacaFasc*" (میمون ماکاک خورچنگ‌خوار) و "*PapioAnub*" (بابون) یک تبارشاخه متمایز را تشکیل می‌دهند که دودمان مشترک اخیر آن‌ها را به نمایش می‌گذارد. علاوه بر این، درخت شامل پستانداران دیگری مانند "*MarmotaMar*" (موش خرمایی)، "*MusMusculus*" (موش)، "*RattusNorv*" (موش صحرایی) و "*Cricetulus*" (همستر) است که با هم گروه‌بندی شده‌اند و روابط تکاملی آنها را نشان می‌دهند. "*Monodelphi (opossum)*" به‌عنوان یک گروه بیرونی نشان داده شده است که رابطه تکاملی دورتر آن را با سایر گونه‌ها در درخت برجسته

می‌کنند. دوزیست "*XenopusTro*" (قورباغه) و خزنده "*Protothru*" و "*GekkoJapon*" نیز نشان داده شده‌اند که واگرایی تکاملی آن‌ها را از سایر گونه‌ها نشان می‌دهد. بخش پایینی درخت‌تبارزایی عمدتاً شامل پستانداران است. گونه‌هایی مانند "*PanTroglod*" (شامپانزه)، "*PanPaniscu*" (بونوبو)، "*HomoSapien*" (انسان)، "*NomascusLe*" (گیبون)، "*MacacaMmu*" (میمون رزوس)، "*MacacaFasc*" (میمون ماکاک خورچنگ‌خوار) و "*PapioAnub*" (بابون) یک تبارشاخه متمایز را تشکیل می‌دهند که دودمان مشترک اخیر آن‌ها را به نمایش می‌گذارد.

علاوه بر این، درخت‌تبارزایی شامل پستانداران دیگری مانند "*MusMusculus*" (موش)، "*RattusNorv*" (موش صحرایی)، "*Cricetulus*" (همستر)، "*MarmotaMar*" (موش خرمايي)، "*CanisLupus*" (سگ)، "*MustelaPut*" (راسوی اهلی)، "*AcinonyxJu*" (یوزپلنگ)، "*EquusPrzew*" (اسب پرژوالسکی)، "*BosTaurus*" (گاو)، "*OvisAriesB*" (گوسفند)، "*SusScrofa*" (خوک) و "*Miniopertu*" (خفاش) است که با هم گروه‌بندی شده‌اند و روابط تکاملی آنها را نشان می‌دهند. "*Monodelphi*" (*opossum*) به‌عنوان یک گروه بیرونی نشان داده شده است که رابطه تکاملی دورتر آن را با سایر گونه‌ها در درخت برجسته می‌کند. به‌طور کلی، این درخت‌تبارزایی، نمای کلی جامعی از تاریخ تکاملی و روابط ژنتیکی بین مجموعه‌ای متنوع از گونه‌ها ارائه می‌دهد و مقادیر بوت‌استرپ نشان‌دهنده اعتبار این روابط است.

الگوریتم *Dnam* این الگوریتم در گروه فراروایت حداکثر درست‌نمایی گروه‌بندی می‌شود. نسخه فعلی *Dnaml*، سریع‌تر از نسخه‌های قبلی آن است (Felsenstein & Churchill, 1996). این الگوریتم اجازه می‌دهد تا فراوانی‌های پیش‌بینی‌شده چهار باز DNA نابرابر باشند، فراوانی‌های پیش‌بینی‌شده انتقال‌ها و جابه‌جایی بازها نابرابر باشند و راه‌های مختلفی برای اجازه دادن به نرخ‌های مختلف تعامل در مکان‌های مختلف DNA فراهم می‌کند. از مفروض‌های این الگوریتم می‌توان به: هر جایگاه در توالی DNA به‌طور مستقل تکامل می‌یابد، دودمان‌های مختلف به‌طور مستقل تکامل می‌یابند، هر جایگاه DNA با نرخ مورد انتظار جایگزین می‌شود که از یک سری نرخ‌ها (هر کدام با احتمال وقوع) که مشخص می‌کنیم، انتخاب می‌شود، همه

غیره). در این شرایط، انتخاب روش بستگی به مقاومت هر روش در برابر تغییرات خاص داده‌ها دارد. تجزیه و تحلیل معرفت‌شناختی، بهترین روش کشف درخت‌تبارزایی است زیرا نتایج آن به کمینه فرض‌های پیشینی وابسته است و بنابراین، دامنه اکتشافات بالقوه نشان داده شده به‌وسیله داده‌ها بیشتر از هر رویکرد صرفه‌جویی هستی‌شناختی است. در این راستا، رویکردهای مختلف حداکثر درست‌نمایی و بیزی به‌طور تحسین‌برانگیزی به‌عنوان روش‌های ارزیابی مناسب هستند، یعنی روش‌هایی که این امکان را می‌دهند که فرآیندهای بالقوه زیربنای الگوهای تبارزایی را بررسی کنیم. بنابراین، کامل‌ترین توضیح برای تبارزایی، توضیحی است که اطلاعاتی را از هر دو رویکرد داده و احتمال ترکیب می‌کند. PhyML یک الگوریتم جدید برای جستجوی فضای درخت‌تبارزایی با استفاده از هرس زیر درخت‌تبارزایی و پیوند مجدد است (Guindon et al., 2010). همان‌طور که در شکل ۳ آشکار است، مجموعه پستانداران مورد بررسی در این پژوهش از نظر ژن *BRAC1* در چند خوشه نزدیک به‌هم از نظر تکاملی هم‌گروه شده‌اند. از این نظر، این روال رایانشی نیز مانند روش حداکثر صرفه‌جویی جواب یکسانی را ایجاد کرده است. از نظر ساختار و شکل و توپولوژی، تفاوت فاحشی با الگوریتم‌های حداکثر درست‌نمایی در این الگوریتم مشاهده می‌شود. این درخت‌تبارزایی، روابط تکاملی بین مجموعه‌ای متنوع از گونه‌ها، شامل ماهی‌ها، دوزیستان، خزندگان، پرندگان و پستانداران را نشان می‌دهد. درخت‌تبارزایی با گره‌هایی که نشان‌دهنده اجداد مشترک هستند و شاخه‌هایی که نشانگر دودمان‌های تکاملی‌اند، ایجاد شده است.

مقادیر بوت‌استرپ در گره‌ها، نشان‌دهنده سطح اطمینان از این الگوهای شاخه‌بندی است. در بالای درخت‌تبارزایی، گونه‌های ماهی و دوزیست مانند "*PoeciliaLa*"، "*PoeciliaMe*"، "*Cyprinodon*"، "*Nothobranc*"، "*Oreochromi*" و "*SalmoSalar*" با هم گروه‌بندی شده‌اند که نشان‌دهنده پیوندهای تکاملی نزدیک‌تر آن‌ها است. دسته‌بندی این گونه‌ها، دودمان مشترک متمایز از سایر گروه‌ها در درخت‌تبارزایی را نشان می‌دهد. به‌سمت پایین درخت، پرندگانی مانند "*GallusGall*" (مرغ)، "*CoturnixJa*" (بلدرچین)، "*CalidrisPp*" و "*ParusMajor*" مشاهده می‌شوند که یک تبارشاخه متمایز را تشکیل می‌دهند و دودمان مشترک آنها را برجسته

"*Miniopeteru*" (خفاش) است که با هم گروه‌بندی شده‌اند و روابط تکاملی آنها را نشان می‌دهند. "*Monodelphi*" (پوسوم) به‌عنوان یک گروه بیرونی نشان داده شده است که رابطه تکاملی دورتر آن را با سایر گونه‌ها در درخت تبارزایی برجسته می‌کند. به‌طور کلی، این درخت تبارزایی، نمای کلی جامعی از تاریخ تکاملی و روابط ژنتیکی بین مجموعه‌ای متنوع از گونه‌ها ارائه می‌دهد و مقادیر بوت‌استرپ نشان‌دهنده اعتبار این روابط است. این درخت تبارزایی به‌طور مؤثر هم واگرایی‌های باستانی، مانند بین "*GallusGall*" و "*XenopusTro*" و هم واگرایی‌های جدیدتر، مانند بین انسان‌ها و شامپانزه‌ها را نشان می‌دهد و بینش‌هایی را در مورد مسیرهای تکاملی این موجودات ارائه می‌دهد. در شکل ۴، با تأکید ویژه بر پستانداران، درخت تبارزایی از مهره‌داران پایه مانند ماهی‌ها (به‌عنوان مثال، *Salmo salar*, *Oreochromis*, *Cyprinodon*, *Poecilia* دوزیستان (به‌عنوان مثال، *Xenopus*) مشتق شده و به اجداد اولیه پستانداران و پرندگان (به‌عنوان مثال، *Calidris*، *Parus*) می‌رسد.

بیشتر درخت تبارزایی بر تنوع پستانداران متمرکز است و تبارشاخه و گونه‌های کلیدی را نشان می‌دهد. میمون‌های بزرگ، از جمله انسان‌ها (*HomoSapien*)، شامپانزه‌ها (*PanPanisc*، *PanTroglod*) و گوریل‌ها، دودمان مشترک نزدیکی دارند که تاریخ تکاملی مشترک ما را برجسته می‌کند. میمون‌های دنیای قدیم (به عنوان مثال، *Macaca*، *Nomascus*) و جوندگان (به عنوان مثال، *RattusNorv*، *MusMusculu*) شاخه‌های متمایزی را تشکیل می‌دهند که مسیرهای تکاملی جداگانه آن‌ها را نشان می‌دهد. گوشتخواران (به‌عنوان مثال، *CanisLupus*، *MustelaPut*) و سم‌داران (به‌عنوان مثال، *EquusPrzew*، *BosTaurus*) نیز شاخه‌های متمایزی را نشان می‌دهند که واگرایی آنها را از یک دودمان مشترک نشان می‌دهد. این درخت، نمای جامعی از روابط تکاملی بین این گونه‌ها ارائه می‌دهد و بر روابط نزدیک بین میمون‌های بزرگ، شاخه‌های متمایز جوندگان، گوشتخواران و سم‌داران، و فاصله تکاملی بین پستانداران و مهره‌داران پایه مانند ماهی‌ها و دوزیستان تأکید می‌کند.

جایگاه‌های DNA مرتبط در توالی DNA گنجانده می‌شوند، نه فقط آن‌هایی که تغییر کرده‌اند یا آن‌هایی که از نظر تبارزایی قابل اتکا هستند، اشاره کرد. شکل ۴، درخت تبارزایی و روابط تکاملی بین مجموعه‌ای متنوع از گونه‌ها، شامل ماهی‌ها، دوزیستان، خزندگان، پرندگان و پستانداران را نشان می‌دهد. درخت تبارزایی با گره‌هایی که نشان‌دهنده اجداد مشترک هستند و شاخه‌هایی که نشانگر دودمان‌های تکاملی‌اند، ایجاد شده است. مقادیر بوت‌استرپ در گره‌ها، نشان‌دهنده سطح اطمینان از این الگوهای شاخه‌بندی است در بالای درخت. گونه‌های ماهی و دوزیست مانند "*PoeciliaLa*"، "*PoeciliaMe*"، "*Cyprinodon*"، "*Nothobranc*"، "*Oreochromi*" و "*SalmoSalar*" با هم گروه‌بندی شده‌اند که نشان‌دهنده پیوندهای تکاملی نزدیک‌تر آن‌ها است. دسته‌بندی این گونه‌ها، دودمان مشترک متمایز از سایر گروه‌ها در درخت تبارزایی را نشان می‌دهد. به‌سمت پایین درخت، پرندگانی مانند "*GallusGall*" (مرغ)، "*CoturnixJa*" (بلدرچین)، "*CalidrisPp*" و "*ParusMajor*" مشاهده می‌شوند که یک تبارشاخه متمایز را تشکیل می‌دهند و دودمان مشترک آنها را برجسته می‌کنند. دوزیست "*XenopusTro*" (قورباغه) و خزنده "*Protobothr*" نیز نشان داده شده‌اند که واگرایی تکاملی آن‌ها را از سایر گونه‌ها نشان می‌دهد. بخش پایینی درخت عمدتاً شامل پستانداران است. گونه‌هایی مانند "*PanTroglod*" (شامپانزه)، "*PanPaniscu*" (بونوبو)، "*HomoSapien*" (انسان)، "*NomascusLe*" (گیبون)، "*MacacaMmu*" (میمون رزوس)، "*MacacaFasc*" (میمون ماکاک خورچنگ‌خوار) و "*PapioAnub*" (بابون) یک تبارشاخه متمایز را تشکیل می‌دهند که جد مشترک اخیر آنها را به نمایش می‌گذارد. علاوه بر این، درخت شامل پستانداران دیگری مانند "*MusMusculus*" (موش)، "*RattusNorv*" (موش صحرایی)، "*Cricetulus*" (همستر)، "*MarmotaMar*" (موش خرمایی)، "*CanisLupus*" (سگ)، "*MustelaPut*" (راسوی اهلی)، "*AcinonyxJu*" (یوزپلنگ)، "*EquusPrzew*" (اسب پرژوالسکی)، "*BosTaurus*" (گاو)، "*OvisAriesB*" (گوسفند)، "*SusScrofa*" (خوک) و

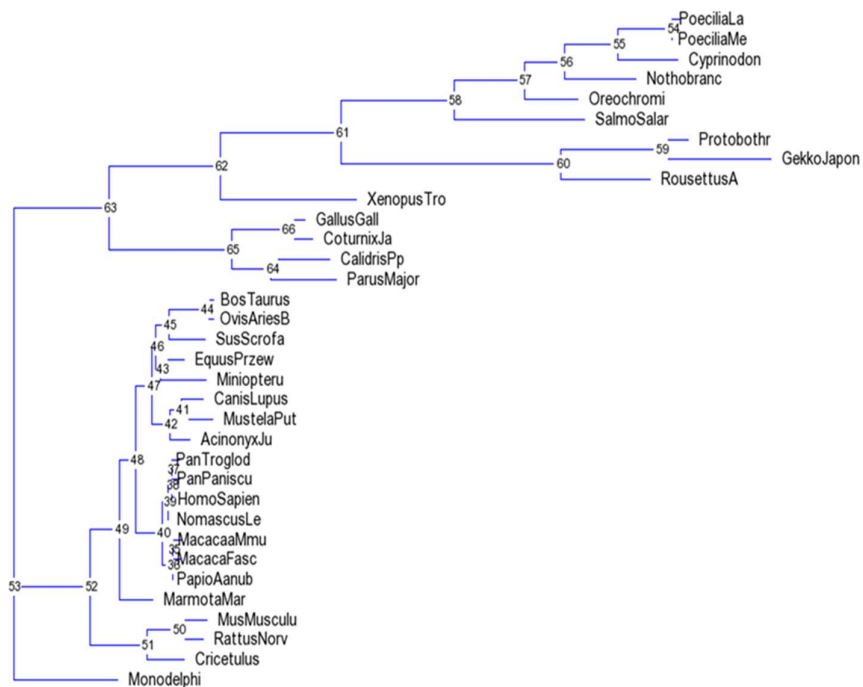


Fig. 1. Phylogenetic tree reconstruction of the studied species using the maximum parsimony algorithm of DNAPARS

شکل ۱- استخراج درخت‌تبارزایی برای گونه‌های مورد بررسی با استفاده از الگوریتم حداکثر صرفه‌جویی DNAPARS

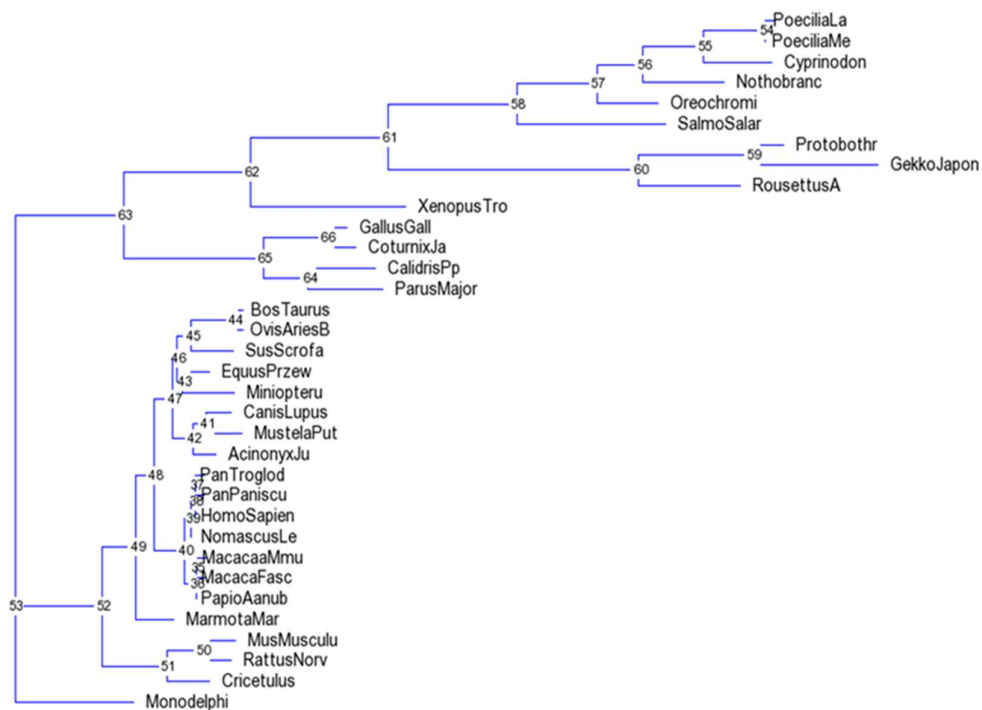


Fig. 2. Phylogenetic tree reconstruction of the studied species using the Maximum Parsimony algorithm

شکل ۲- استخراج درخت‌تبارزایی برای گونه‌های مورد بررسی با استفاده از الگوریتم حداکثر صرفه‌جویی Maximum Parsimony

تشکیل می‌دهند، اما با سطوح پشتیبانی متفاوت (مقادیر بوت‌استرپ از ۵۶ تا ۶۵). پستانداران، به‌ویژه نخست‌ها مانند "*HomoSapien*"، "*PanTroglod*"، "*PanPaniscu*"، "*NomascusLe*" و "*MacacaFas*"، "*MacacaMmu*"، "*PapioAnub*"، در همه درخت‌ها با هم گروه‌بندی شده‌اند، هرچند با مقادیر بوت‌استرپ متفاوت (دامنه ۴۱ تا ۴۹). سایر پستانداران، از جمله "*MusMusculus*" (موش)، "*RattusNorv*" (موش صحرایی)، "*Cricetulus*" (همستر)، "*MarmotaMar*" (موش خرمایی)، "*CanisLupus*" (سگ)، "*MustelaPut*" (راسوی اهلی)، "*AcinonyxJu*" (یوزپلنگ)، "*EquusPrzew*" (اسب پرژوالسکی)، "*BosTaurus*" (گاو)، "*OvisAriesB*" (گوسفند)، "*SusScrofa*" (خوک) و "*Miniopteru*" (خفاش)، به‌طور مشابه گروه‌بندی شده‌اند، اما سطوح پشتیبانی متفاوتی از نظر مقادیر بوت‌استرپ دارند.

الگوریتم *Dnaml*: درخت تبارزایی تصویر شده (شکل ۵) نمایای از روابط تکاملی بین گونه‌های مختلف است که به‌وسیله الگوریتم *Dnaml* و بر اساس فراروایت حداکثر درست‌نمایی ساخته شده است. این درخت تبارزایی از یک ریشه مشترک آغاز می‌شود که نشان‌دهنده دودمان مشترک

در درخت تبارزایی مهره‌داران پایه مانند ماهی‌ها (به‌عنوان مثال، *Salmo*، *Oreochromis*، *Cyprinodon*، *Poecilia*، *salar*) و دوزیستان (به‌عنوان مثال، *Xenopus*) اجداد اولیه پستانداران و پرندگان (به‌عنوان مثال، *Parus*، *Calidris*) را شکل می‌دهند. گوشتخواران (به‌عنوان مثال، *CanisLupus*، *MustelaPut*) و سم‌داران (به‌عنوان مثال، *EquusPrzew*، *BosTaurus*) نیز شاخه‌های متمایزی را نشان می‌دهند که واگرایی آنها را از یک دودمان مشترک نشان می‌دهد. درخت‌های تبارزایی ارائه شده، به‌طور کلی، روابط تکاملی بین مجموعه‌ای متنوع از گونه‌ها، شامل ماهی‌ها، دوزیستان، خزندگان، پرندگان و پستانداران را نشان می‌دهند، با این حال، گروه‌بندی گونه‌ها و مقادیر بوت‌استرپ که نشان‌دهنده میزان اطمینان به این روابط هستند، تغییرات قابل توجهی دارند. در همه درخت‌ها، گونه‌های ماهی و دوزیست مانند "*PoeciliaLa*"، "*PoeciliaMe*"، "*Cyprinodon*"، "*Nothobranc*"، "*Oreochromi*" و "*SalmoSalar*" با هم گروه‌بندی شده‌اند، اگرچه مقادیر بوت‌استرپ متفاوت است و از ۳۶ تا ۵۵ متغیر است. پرندگانی مانند "*GallusGall*" (مرغ)، "*CoturnixJa*" (بلدرچین)، "*CalidrisPp*" و "*ParusMajor*" به‌طور مداوم یک تبارشاخه متمایز را

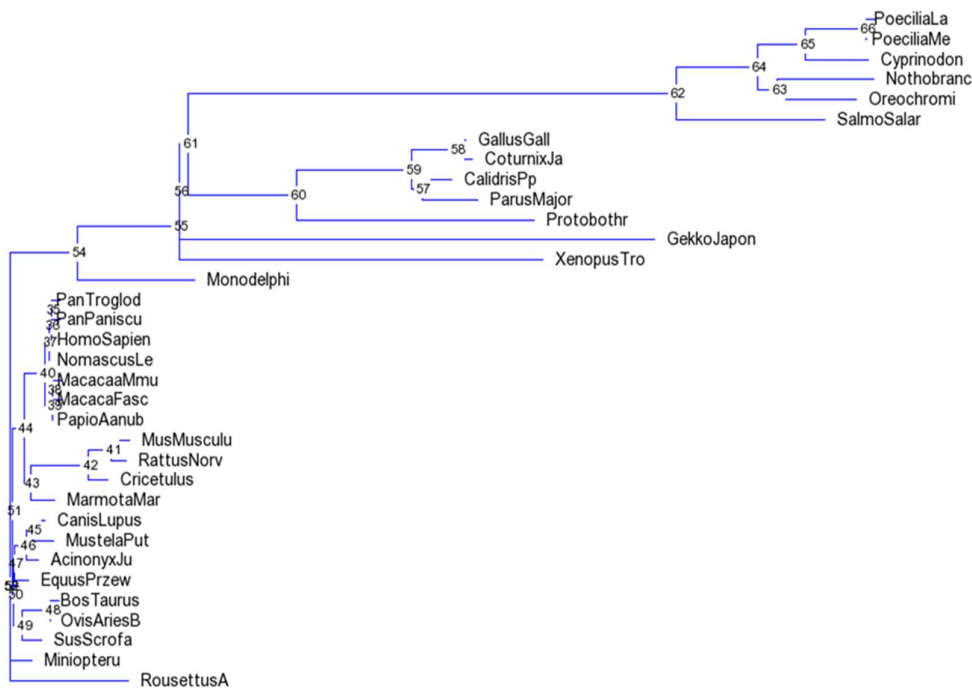


Fig. 3. Phylogenetic tree reconstruction of studied species using the PhyML algorithm

شکل ۳- استخراج درخت تبارزایی گونه‌های مورد مطالعه با استفاده از الگوریتم PhyML

متمایزی را تشکیل می‌دهند که تاریخ تکاملی مشترک آن‌ها را برجسته می‌کند.

جوندگان (مانند *MusMusculu* و *RattusNorv*)، گوشتخواران (مانند *CanisLupus* و *MustelaPut*)، سم‌داران (مانند *BosTaurus* و *OvisAriesB*) و خفاش‌ها (مانند *Miniopteru* و *RousettusA*) نیز شاخه‌های متمایزی را نشان می‌دهند که مسیرهای تکاملی جداگانه آن‌ها را نشان می‌دهد. این درخت بر روابط نزدیک بین نخستین‌ها، شاخه‌های متمایز جوندگان، گوشتخواران، سم‌داران و خفاش‌ها، و فاصله تکاملی بین پستانداران و مهره‌داران پایه مانند ماهی‌ها و دوزیستان تأکید می‌کند. الگوریتم *Proml* درخت‌تبارزایی حاصل از الگوریتم *Proml*، روابط تکاملی بین گونه‌های مختلف مهره‌دار، به‌ویژه پستانداران را نشان می‌دهد. درخت‌تبارزایی با مهره‌داران پایه مانند ماهی‌ها (از جمله *Poecilia*، *Cyprinodon*، *Oreochromis*، *Salmo salar*) و دوزیستان (*Xenopus*) ریشه‌گذاری شده و به اجداد اولیه پستانداران و پرندگان (مانند *Gallus* و *Coturnix*) می‌رسد. بخش اعظم درخت‌تبارزایی بر تنوع پستانداران متمرکز است و تبارشاخه و گونه‌های کلیدی را نشان می‌دهد. میمون‌های بزرگ، شامل انسان (*HomoSapien*)، شامپانزه‌ها (*PanTroglod*) و *PanPanisc* و میمون‌های دنیای قدیم (مانند *Nomascus* و *Macaca*)، شاخه‌های متمایزی را تشکیل می‌دهند که تاریخ تکاملی مشترک آن‌ها را برجسته می‌کند.

همه گونه‌های موجود در درخت است. هر شاخه از درخت، نشان‌دهنده یک خط تکاملی است که از دودمان مشترک به گونه‌های امروزی منتهی می‌شود. اعداد روی شاخه‌ها، مقادیر بوت استراپ را نشان می‌دهند. در این درخت‌تبارزایی، گونه‌هایی مانند "*PanTroglody*" و "*PanPaniscus*" در یک تبارشاخه قرار دارند که نشان‌دهنده رابطه نزدیک‌تر آن‌ها با یکدیگر نسبت به سایر گونه‌ها است. همچنین، گونه‌هایی مانند "*HomoSapiens*" و "*GorillaGorilla*" نیز در یک گروه قرار دارند که نشان‌دهنده رابطه نزدیک آن‌ها است. گونه‌هایی مانند "*Monodelphi*" که در یک شاخه طولانی‌تر قرار دارند، نشان‌دهنده تکامل قابل توجه از سایر گونه‌ها هستند. درخت همچنین گونه‌هایی مانند "*GallusGallus*" و "*XenopusTro*" را به‌عنوان گروه‌های خارجی استفاده می‌کند که به‌عنوان مرجع برای تعیین جهت تکامل به‌کار می‌روند. درخت‌تبارزایی با مهره‌داران پایه مانند ماهی‌ها (مانند *Salmo*، *Oreochromis*، *Cyprinodon*، *Poecilia*) و دوزیستان (مانند *Xenopus*) ریشه‌گذاری شده و به اجداد اولیه پستانداران و پرندگان (مانند *Gallus* و *Coturnix*) می‌رسد. بخش اعظم درخت بر تنوع پستانداران متمرکز است و تبارشاخه‌ها و گونه‌های کلیدی را نشان می‌دهد. میمون‌های بزرگ، شامل انسان (*HomoSapien*)، شامپانزه‌ها (*PanTroglod*) و *PanPanisc* و میمون‌های دنیای قدیم (مانند *Nomascus* و *Macaca*)، شاخه‌های

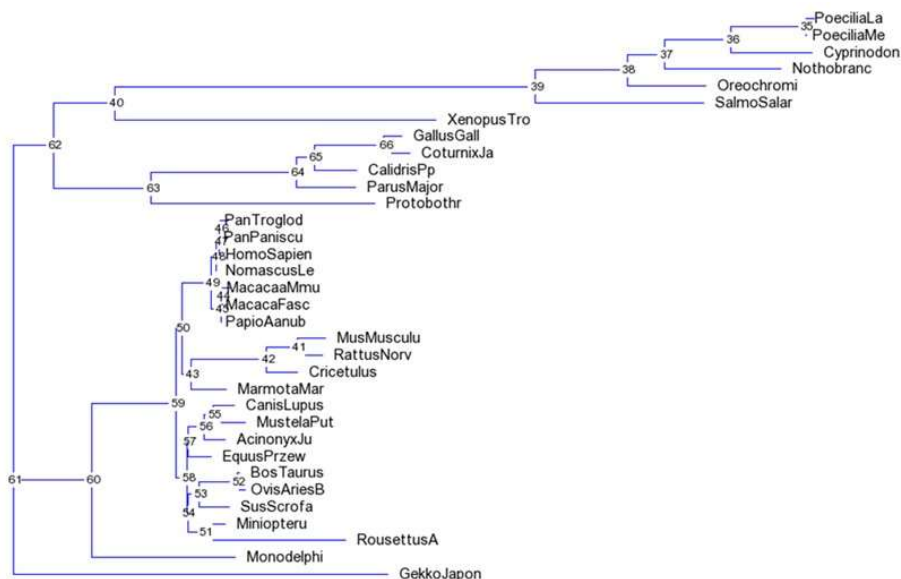


Fig. 4. Phylogenetic tree reconstruction of studied species using the DNAML algorithm

شکل ۴- استخراج درخت‌تبارزایی گونه‌های مورد مطالعه با استفاده از الگوریتم DNAML

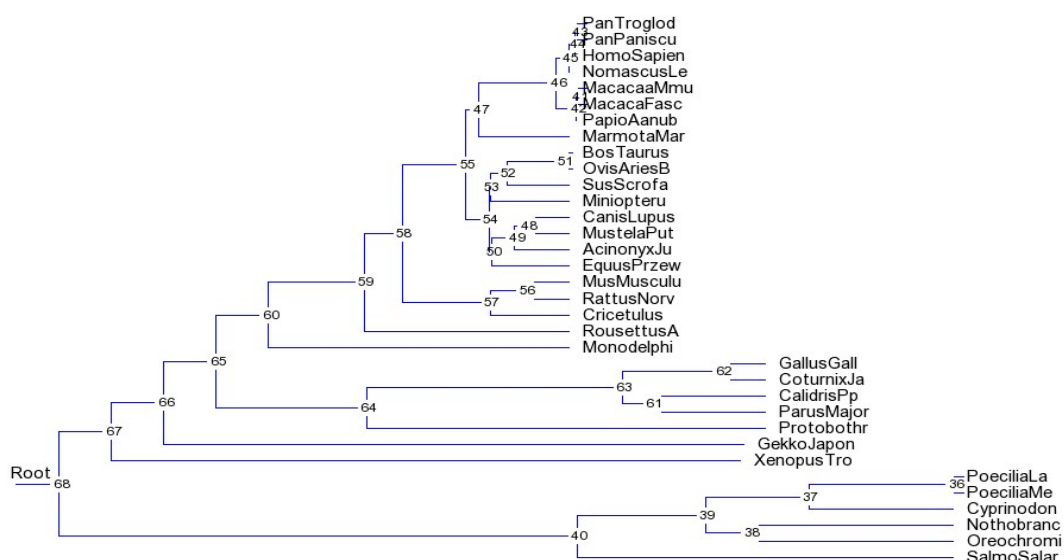


Fig. 5. Phylogenetic tree reconstruction of studied species using the Maximum Parsimony algorithm (DNAMLK)

شکل ۵- استخراج درخت تبارزایی گونه‌های مورد مطالعه با استفاده از الگوریتم حداکثر صرفه‌جویی (DNAMLK)

کنند. پژوهش‌های مقایسه‌ای که به وسیله این درخت‌های تبارزایی امکان‌پذیر شده است، می‌تواند جهش‌ها یا چندشکلی‌های منحصر به فردی را در ژن *BRCAl* شناسایی کند که ممکن است بر پیامدهای ورم پستان تأثیر بگذارد.

با استفاده از این اطلاعات ژنتیکی، پژوهشگران می‌توانند راهبردهای هدفمندی مانند غربالگری ژنتیکی و اصلاح نژاد انتخابی را برای بهبود سلامت و بهره‌وری گله با کاهش تأثیر ورم پستان در گاوها توسعه دهند. تشخیص زود هنگام ورم پستان در پیشگیری از ضررهای اقتصادی ناشی از ورم پستان مهم است. در سال‌های اخیر علاقه فزاینده‌ای به توسعه گله‌های مقاوم به ورم پستان ژنتیکی شده است. پژوهش‌های مختلف نشان داده است که بین برخی از ژن‌ها و مقاومت به ورم پستان، رابطه قوی وجود دارد. بنابراین، علاقه فزاینده‌ای به توسعه گله‌های مقاوم به ورم پستان وجود دارد. (Yang et al., 2004; Roehe et al., 2007; Asif, Khan, & Behrouzi, 2016). ژن *BRCAl* یکی از آن ژن‌هایی است که گزارش شده است با خطر سرطان پستان ارثی در انسان (Friedman & Hastie, 1994) مرتبط است.

جوندگان (مانند *MusMusculu* و *RattusNorv*), گوشتخواران (مانند *CanisLupus* و *MustelaPut*), سم‌داران (مانند *OvisAriesB* و *BosTaurus*) و خفاش‌ها (مانند *Miniopeteru* و *RousettusA*) نیز شاخه‌های متمایزی را نشان می‌دهند که مسیرهای تکاملی جداگانه آن‌ها را نشان می‌دهد.

تحلیل تبارزایی توالی‌های DNA ژن *BRCAl* از گونه‌های مختلف، از جمله گاوها، بینش‌های مهمی را در مورد مبانی ژنتیکی حساسیت و مقاومت به ورم پستان در گاو ارائه می‌دهد. گروه‌بندی همیشگی پستانداران، به‌ویژه سم‌داران مانند "*BosTaurus*" (گاو)، در سراسر درخت‌های تبارزایی نشان می‌دهد که ژن *BRCAl* از نظر تکاملی حفظ شده است و نقش‌های عملکردی مشابهی را در ترمیم DNA و مقاومت در برابر بیماری پیشنهاد می‌کند. تنوع ژنتیکی مشاهده شده در ژن *BRCAl*، همانطور که به‌وسیله مقادیر بوت‌استرپ متفاوت و گروه‌بندی گونه‌ها مشهود است، پتانسیل سطوح مختلف حساسیت به بیماری را در گاوسانان برجسته می‌کند. انواع ژنتیکی خاص *BRCAl* ممکن است درجات مختلفی از مقاومت یا حساسیت به ورم پستان را ایجاد کنند و آنها را به اهداف ارزشمندی برای برنامه‌های اصلاح نژاد با هدف افزایش مقاومت در برابر بیماری تبدیل

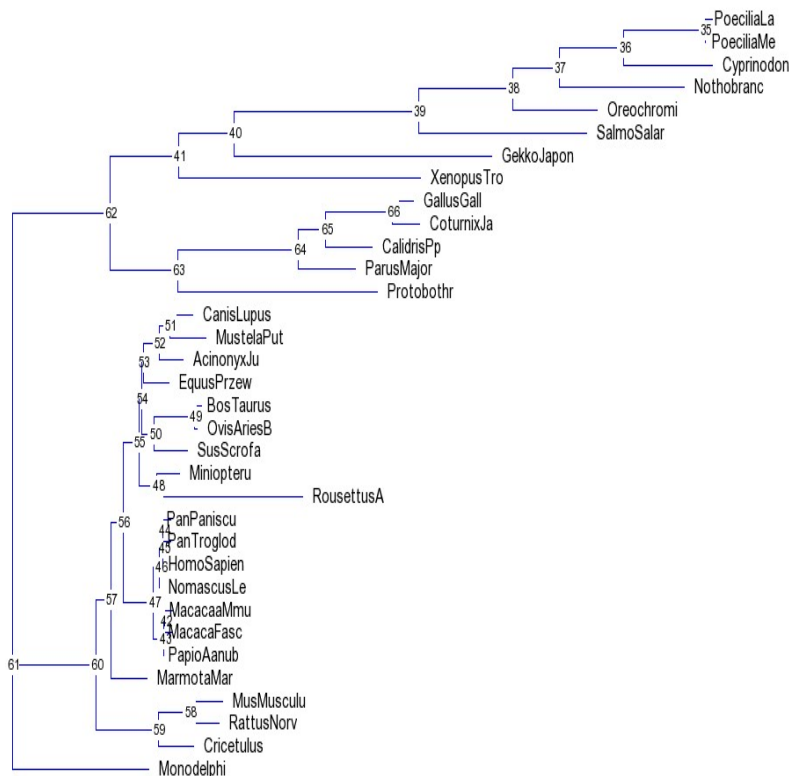


Fig. 6. Phylogenetic tree reconstruction of studied species using the PROML algorithm

شکل ۶- استخراج درخت تبارزایی گونه‌های مورد مطالعه با استفاده از الگوریتم PROML

می‌تواند انتخاب مدل را بهبود بخشد و درک تکامل صفات را افزایش دهد. ترکیب منابع داده‌ای مختلف، ماهیت چندوجهی فرآیندهای تکاملی را نشان داده و به استنتاج‌های دقیق‌تر کمک می‌کند (Slater, 2012). در مجموع، انتخاب و اعمال مدل‌های مناسب برای افزایش دقت تحلیل تبارزایی و کاهش خطاهای احتمالی ضروری است. مقایسه مدل‌ها و روش‌های مختلف برای ساخت درختان تبارزایی، تفاوت‌های کلیدی بین روش‌های مبتنی بر فاصله و مبتنی بر خصوصیت را آشکار می‌سازد. روش‌های مبتنی بر فاصله، مانند Neighbor-Joining و UPGMA، با استفاده از یک ماتریس فاصله‌های جفتی، روابط تکاملی را تخمین می‌زنند، اما این رویکردها ممکن است اطلاعات مربوط به حالت‌های خصوصیت منفرد را از دست داده و تحت تأثیر خواص غیرمتریک ماتریس فاصله قرار گیرند (Huson & Steel, 2004; Scott & Gras, 2012). در مقابل، روش‌های مبتنی بر خصوصیت، از جمله حداکثر پارسیمونی و رویکردهای بیزی، مستقیماً تغییرات نوکلئوتیدی یا اسیدهای آمینه را تجزیه و تحلیل کرده و در صورت سازگاری مدل تکاملی، درخت‌های دقیق‌تری ارائه می‌دهند

مدل‌های تکاملی، نقش اساسی در استنتاج روابط تبارزایی دارند. انتخاب مدل مناسب بسیار مهم است، زیرا مدل‌های مختلف، نرخ‌ها و الگوهای جایگزینی نوکلئوتید را به شیوه‌های متفاوتی در نظر می‌گیرند. مدل‌های ساده‌تر ممکن است پیچیدگی تکامل توالی را به درستی منعکس نکنند و منجر به ایجاد درخت‌های نادرست شوند (Posada & Crandall, 2001). علاوه بر این، دقت برآوردهای تبارزایی به شدت تحت تأثیر مدل انتخابی قرار دارد. مطالعات نشان داده‌اند که استفاده از مدلی که با داده‌ها بیشترین سازگاری را دارد، قابلیت اطمینان استنتاج‌های تکاملی را افزایش می‌دهد، در حالی که مدل‌های نامناسب می‌توانند باعث نتیجه‌گیری‌های گمراه‌کننده شوند (Rohlf, 1990). از نظر پیچیدگی، مدل‌های پیچیده‌تر ممکن است فرآیندهای تکاملی را بهتر منعکس کنند، اما همچنین تعداد زیادی فراسنجه را وارد تحلیل می‌کنند که می‌تواند بر دقت برآورد تأثیر بگذارد. ایجاد تعادل میان پیچیدگی مدل و برازش داده‌ها برای دستیابی به یک تحلیل تبارزایی قوی ضروری است (Caldart, 2018). علاوه بر این، یکپارچه‌سازی داده‌های فسیلی و سایر اطلاعات تاریخی

می‌دهند. مدل جکس-کانتور (Jukes-Cantor) یکی از ساده‌ترین مدل‌ها است که نرخ جایگزینی یکنواختی را بین تمام نوکلئوتیدها فرض می‌کند و معمولاً در تحلیل‌های مقدماتی کاربرد دارد (Fabreti & Höhna, 2022). در مقابل، مدل کیمورا دو فراسنجه‌ای (Kimura 2-Parameter) با تمایز قائل شدن بین نرخ‌های انتقال (تغییر بین پورین یا بین پیریمیدین‌ها) و ترانسورژن (تغییر بین پورین و پیریمیدین)، دقت بیشتری نسبت به مدل جکس-کانتور ارائه می‌دهد (Nwezeobi et al., 2020). مدل زمان برگشت‌پذیر عمومی (GTR)، به‌عنوان یک مدل پیچیده‌تر، تمامی نرخ‌های جایگزینی بین جفت‌های نوکلئوتیدی را به‌طور مستقل در نظر می‌گیرد و به‌دلیل انعطاف‌پذیری بالا، در بسیاری از تحلیل‌های تبارزایی به‌کار گرفته می‌شود (Bollback, 2002). علاوه بر این، مدل‌های بیزی با استفاده از توزیع‌های پیشین، امکان تحلیل مجموعه داده‌های پیچیده را فراهم کرده و در برآورد روابط تبارزایی به‌طور فزاینده‌ای محبوب شده‌اند (Wright, 2019). از سوی دیگر، قانون دولو (Dollo's Law)، که برای بررسی تکامل خصوصیات پیچیده طراحی شده است، فرض می‌کند که صفات از دست رفته به‌سختی مجدداً به‌دست می‌آیند و بنابراین برای مطالعه مسیرهای تکاملی برگشت‌ناپذیر بسیار مفید است (Farris, 1977).

به‌طور کلی، مقایسه کلی درخت تبارزایی ایجاد شده روابط تکاملی بین مجموعه‌ای متنوع از گونه‌های مهره‌دار، شامل ماهی‌ها، دوزیستان، خزندگان، پرندگان و پستانداران را نشان دادند. همه درخت‌های تبارزایی، ماهی و دوزیست مانند "*PoeciliaLa*"، "*PoeciliaMe*"، "*Cyprinodon*"، "*Nothobranch*"، "*Oreochromis*" و "*SalmoSalar*" را با هم گروه‌بندی کردند، اگرچه با مقادیر بوت‌استرپ متفاوت از ۳۶ تا ۵۵، که نشان‌دهنده سطوح مختلف اطمینان در این روابط است. پرندگانی مانند "*GallusGall*" (مرغ)، "*CoturnixJa*" (بلدرچین)، "*CalidrisPp*" و "*ParusMajor*" در همه درخت‌ها یک تبارشاخه متمایز را تشکیل می‌دهند، اگرچه سطوح پشتیبانی با مقادیر بوت‌استرپ از ۵۶ تا ۶۵ متفاوت است. پستانداران، به‌ویژه نخستی‌ها از جمله "*PanTroglod*"، "*PanPaniscu*"، "*HomoSapien*"، "*NomascusLe*"، "*MacacaMmu*"، "*MacacaFasc*" و "*PapioAnub*"، به‌طور مداوم با هم گروه‌بندی می‌شوند و مقادیر بوت‌استرپ آنها از ۴۱ تا ۴۹

(Charleston, 1994; Keating, 2020). علاوه بر این، این روش‌ها توانایی بالاتری در کنترل هم‌سانی داشته و اغلب در سناریوهای تکاملی پیچیده ترجیح داده می‌شوند (Zou, 2011). در مجموع، در حالی که روش‌های مبتنی بر فاصله به‌دلیل سرعت و سادگی محاسباتی مزیت دارند، روش‌های مبتنی بر خصوصیت، به‌طور کلی درختان تبارزایی قابل اعتمادتر و دقیق‌تری را تولید می‌کنند، به‌ویژه در شرایطی که فرآیندهای تکاملی پیچیده در نظر گرفته شوند. روش‌های مختلف تبارزایی را می‌توان بر اساس ارتباط آن‌ها با رویکردهای مبتنی بر فاصله و مبتنی بر خصوصیت طبقه‌بندی کرد. درختان تبارشناسی اختصاصی، که بر پایه داده‌های توالی یک ژن خاص ساخته می‌شوند، عمدتاً از روش‌های مبتنی بر خصوصیت مانند حداکثر احتمال یا رویکردهای بیزی استفاده می‌کنند و دقت بالاتری نسبت به درختان مبتنی بر سنتز دارند، اما به تخصص بیشتری نیز نیاز دارند (Li, 2018). در مقابل، درختان تبارشناسی مبتنی بر سنتز، که از ترکیب چندین درخت تبارزایی موجود ساخته می‌شوند، می‌توانند روش‌های مبتنی بر فاصله و خصوصیت را در بر بگیرند و در حالی که برآوردهای گسترده‌تری از تنوع تبارزایی ارائه می‌دهند، ممکن است در برخی موارد، بیش‌برآورد داشته باشند (Li, 2019). تحلیل چندجمله‌ای تبارزایی به‌عنوان یک رویکرد مبتنی بر خصوصیت، با استفاده از خصوصیت‌های سلسله‌مراتبی و الگوریتم‌های چندجمله‌ای، دقت برآوردهای تبارزایی را افزایش می‌دهد (Hopkins & John, 2021). سیستم مختصات شبکه درخت، با ارائه یک چارچوب مقایسه‌ای برای شکل درخت و طول شاخه‌ها، امکان تحلیل‌های مبتنی بر فاصله را بدون وابستگی به روش‌های سنتی مبتنی بر خصوصیت فراهم می‌کند (Liu, 2022). علاوه بر این، الگوریتم‌های جدیدی مانند TreeKO، که برای تاکسون‌های هم‌پوشانی طراحی شده‌اند، با ترکیب روش‌های مبتنی بر فاصله و خصوصیت، دقت برآوردهای تبارزایی را در سناریوهای داده‌ای پیچیده افزایش می‌دهند (Wiens, 1998). به‌طور کلی، این روش‌ها نشان‌دهنده ترکیب راهبردهای مختلف برای تحلیل تبارزایی هستند، به‌طوری که هر کدام، نقاط قوت و کاربردهای ویژه‌ای در استنتاج روابط تکاملی دارند. مدل‌های تکاملی متعددی در تحلیل تبارزایی مورد استفاده قرار می‌گیرند که هر یک، سطحی از پیچیدگی و دقت را در استنباط روابط تکاملی ارائه

موقعیت‌های نوکلئوتیدی خاص که برای تحلیل مورد استفاده قرار می‌گیرند، تأثیر چشمگیری بر درخت حاصل دارد و باید اطمینان حاصل شود که این خصوصیات، آموزنده و مرتبط با پرسش‌های زیستی و تکاملی هستند (Gonnet & Scholl, 2009). کاربردهای روش‌های تبارزایی شامل درک سازمان ژنوم، علت شیوع بیماری، پیش‌بینی عملکرد پروتئین و تصمیم‌گیری برای تجزیه و تحلیل ژن‌ها در پژوهش‌های تطبیقی است. به‌طور معمول، چنین کاربرانی، علاقه کمتری به درک روابط تاریخی بین گونه‌ها دارند تا اینکه به درک روابط بین توالی ژن‌ها و پروتئین‌هایی که در معرض تجزیه تبارزایی قرار می‌گیرند. با آگاهی از انواع روش‌های تبارزایی موجود، با مشکل دشوار انتخاب مناسب‌ترین روش برای اهداف خود مواجه می‌شوند. کاربران به دقت نسبی روش‌ها و همچنین به معاوضه بین دقت، سرعت محاسباتی و سهولت استفاده از برنامه‌هایی که روش‌های مختلف را پیاده‌سازی می‌کنند، علاقه‌مند هستند. اگرچه منابع علمی قابل توجهی در مورد مقایسه روش‌های مختلف وجود دارد، اما در آن منابع علمی اطلاعات اندکی وجود دارد که راهنمایی روشنی را برای استفاده‌کنندگان معمولی تجزیه تبارزایی ارائه کند. مقایسه روش‌های مختلف مستلزم تولید یک مجموعه داده از یک درخت شناخته شده یا واقعی، اعمال روش‌های مختلف تبارزایی برای آن مجموعه داده، و سپس مقایسه درختان تبارزایی تخمین زده شده با درخت واقعی برای تعیین دقت تخمین هر روش درخت است. قابل اعتمادترین درختان واقعی از سامانه‌های تکامل تجربی می‌آیند که در آن‌ها، جمعیتی از سازواره‌ها، معمولاً با کتریوفاژها، به‌طور دوره‌ای به دودمان تقسیم می‌شوند (Hillis et al., 1992; Bull et al., 1997; Huelsenbeck et al., 2001). شکل تبارزایی، یعنی ترتیب رویدادهای انشعاب و زمان بین شاخه‌ها، به‌وسیله پژوهشگر تعیین می‌شود، اما تغییرات تکاملی به ویژگی‌های خود سامانه مورد بررسی بستگی دارد. تعداد و ماهیت تغییرات از توالی ژن(های) مورد نظر در هر نقطه شاخه تعیین می‌شود. چنین پژوهش‌های تجربی، منابع فشرده‌ای هستند و بیشتر آن‌ها به تعداد اندکی از دودمان محدود می‌شوند. در پژوهشی، یک روش واقعی زیستی برای شبیه‌سازی درختان تبارزایی استفاده شد. این روش از یک توالی کدکننده DNA واقعی به‌عنوان نقطه شروع استفاده می‌کند، جهش را با توجه به طیف جهش اشرشیاکلی از جمله جایگزینی،

متغیر است. سایر پستانداران، مانند جوندگان ("*MusMusculus*", "*RattusNorv*", "*Cricetulus*")، گوشتخواران ("*CanisLupus*", "*MustelaPut*")، سم‌داران ("*BosTaurus*", "*AcinonyxJu*")، خفاش‌ها ("*OvisAriesB*", "*SusScrofa*") و خفاش‌ها ("*Miniopteru*") نیز به‌طور مداوم گروه‌بندی می‌شوند، اما با سطوح پشتیبانی متفاوت. "*Monodelphi (opossum)*" به‌عنوان یک گروه بیرونی ثابت در همه درخت‌ها عمل می‌کند. این اختلافات در مقادیر بوت‌استرپ و تغییرات جزئی در گروه‌بندی گونه‌ها احتمالاً نشان‌دهنده تغییرات در مجموعه داده‌ها، روش‌های تحلیلی یا فرضیه‌های تکاملی مورد استفاده در ساخت هر درخت است و چشم‌انداز دقیق‌تری از تاریخ تکاملی و روابط ژنتیکی بین این گونه‌ها ارائه می‌دهد.

در این پژوهش از توالی CDS ژن *BRCA1* استفاده شد. پیشنهاد می‌شود که از توالی‌های دیگر این ژن مثل توالی mRNA، توالی ژنومی و توالی پروتئینی آن نیز استفاده شود تا نتیجه درست‌تری حاصل شود. پیشنهاد می‌شود برای گونه‌های مورد بررسی در این پژوهش، تکامل افقی همزمان با تکامل عمودی (مانند آن چیزی که در این پژوهش به‌کار رفت) استفاده شود تا رابط بهتری بین گونه‌ها حاصل شود. در تحلیل تبارزایی ژن *BRCA1* در گاوها و گونه‌های مرتبط، چندین عامل باید در نظر گرفته شود تا نتایج دقیق و معتبری حاصل شود. تراز توالی، اولین گام حیاتی است، زیرا عدم هم‌ترازی مناسب می‌تواند به استنتاج‌های نادرست منجر شود؛ ابزارهایی مانند ClustalW برای دستیابی به تراز دقیق استفاده می‌شوند. انتخاب مدل تکاملی مناسب نیز از اهمیت بالایی برخوردار است، زیرا مدل‌های مختلف می‌توانند منجر به توپولوژی‌های متفاوتی در درخت تبارزایی شوند و انتخاب مدلی که بیشترین تطابق را با داده‌ها داشته باشد، ضروری است (Otu & Sayood, 2003). علاوه بر این، کیفیت داده‌های DNA باید تضمین شود؛ داده‌های دارای کیفیت پایین یا دارای مقادیر مفقود می‌توانند تحلیل را منحرف کرده و به نتایج نامعتبر منجر شوند (Li, 2000). از نظر روش‌های محاسباتی، استفاده از رویکردهای پیشرفته مانند مدل‌های بیزی و شبیه‌سازی زنجیره مارکوف مونت کارلو (MCMC) می‌تواند به بهبود دقت استنتاج‌های تبارزایی کمک کند (Cheon & Liang, 2008). در نهایت، انتخاب مناسب خصوصیات تکاملی، مانند

از نقش این ژن در بیماری‌هایی مانند ورم پستان در گاو شیری و سرطان سینه در انسان کمک کند. این موضوع از جنبه‌های مختلفی قابل توجیه است. نخست، شباهت‌های تکاملی و عملکردی بین گونه‌ها نشان می‌دهد که ژن *BRCA1* در حفاظت از سلول‌ها در برابر آسیب‌های DNA و تنظیم چرخه سلولی نقش کلیدی دارد. این ژن به‌ویژه در ترمیم آسیب‌های دو رشته‌ای DNA و جلوگیری از تجمع جهش‌های سرطان‌زا اهمیت دارد.

با بررسی تکاملی این ژن در گونه‌های مختلف، می‌توان الگوهای حفظ شده را شناسایی کرد که نشان‌دهنده اهمیت عملکردی خاصی در همه گونه‌ها هستند. این الگوها می‌توانند به درک سازوکارهای زمینه‌ای بیماری‌هایی مانند سرطان سینه در انسان و ورم پستان در گاو کمک کنند. مقایسه توالی‌های ژن *BRCA1* در گونه‌های مختلف امکان شناسایی جهش‌های مرتبط با بیماری را فراهم می‌کند. این جهش‌ها شامل جهش‌های نقطه‌ای و حذف‌ها یا اضافه‌های بزرگ می‌شوند که می‌توانند منجر به اختلال در عملکرد پروتئین و افزایش خطر ابتلا به بیماری‌هایی مانند سرطان سینه و ورم پستان شوند. بررسی تکاملی این ژن می‌تواند به درک سازوکارهای ایمنی و التهابی کمک کند. به‌عنوان مثال، در گاو، ورم پستان به‌دلیل عفونت باکتریایی رخ می‌دهد و سامانه ایمنی نقش کلیدی در مقابله با این عفونت دارد. شناسایی تنوع ژنتیکی در ژن *BRCA1* می‌تواند به شناسایی گاوهایی که به‌طور ژنتیکی مقاوم‌تر به عفونت‌ها هستند، کمک کند که این اطلاعات می‌تواند در برنامه‌های اصلاح نژاد و مدیریت گاو‌داری مورد استفاده قرار گیرد.

درج و حذف نوکلئوتید-شبییه‌سازی می‌کند و فرآیندهای جهش و انتخاب را از هم جدا می‌کند (Hall, 2005). درختان ۸، ۱۶، ۳۲ و ۶۴ گونه با میانگین طول شاخه‌های ۵۰، ۱۰۰، ۱۵۰، ۲۰۰ و ۲۵۰ تغییر در هر شاخه، شبیه‌سازی شدند. توالی‌های به‌دست‌آمده با ClustalX هم‌تراز شدند و درخت‌های تبارزایی با روش‌های فاصله‌ای، حداکثر درستنمایی، حداکثر صرفه‌جویی و بی‌زی از هر دو توالی DNA و توالی‌های پروتئین مربوطه برآورد شدند. درختان تبارزایی تخمین زده شده با درختان تبارزایی واقعی مقایسه شدند و دقت توپولوژیکی و طول شاخه به‌دست آمد. طی انواع شرایط آزمایش شده، درختان بی‌زی تخمین زده شده از توالی‌های DNA که بر اساس هم‌ترازی توالی‌های پروتئینی مربوطه تراز شده بودند، دقیق‌ترین بودند و پس از آن، درختان تبارزایی حداکثر درستنمایی تخمین زده شده از توالی‌های DNA و درختان تبارزایی حداکثر صرفه‌جویی تخمین زده شده از توالی‌های پروتئین قرار گرفتند. همان‌گونه که در این پژوهش مشاهده می‌شود، الگوریتم‌های مختلف درخت‌های تکاملی متفاوتی استفاده شدند. یک علت شاید به آرایه‌ای از گزینه‌ها و پیش فرض‌هایی بر می‌گردد که برای ایجاد درخت تبارزایی استفاده شدند. مقایسه‌ای از الگوریتم‌های به‌کار رفته در این پژوهش در جدول ۲ نشان داده شده است. انتخاب مناسب مدل تکاملی از اهمیت بالایی برخوردار است، زیرا دقت درخت‌های تبارزایی و درک فرآیندهای زیربنایی تکامل توالی‌ها را بهبود می‌بخشد. ایجاد درخت فیلوژنی برای ژن *BRCA1* در گونه‌های مختلف وحشی و اهلی می‌تواند به درک عمیق‌تری

جدول ۲- مقایسه الگوریتم‌های به‌کاررفته در این پژوهش

Table 2. Comparison of the phylogenetic algorithms used in this study

Algorithm	Type	Data type	Strengths	Weaknesses
DNAPARS	Maximum Parsimony	DNA	Simple and fast, suitable for small datasets	May not explore complex evolutionary processes
Pars	Maximum Parsimony	DNA/AA	Versatile, can process different data types.	Computationally expensive for large datasets
PhyML	Maximum Likelihood	DNA/AA	Statistically robust, capable of processing large datasets and complex models	Computationally expensive
Dnaml	Maximum Likelihood	DNA	Provides statistically robust results for DNA data	Slower than maximum parsimony methods
Dnamlk	Maximum Likelihood	DNA	Improved capabilities over Dnaml	Computationally expensive
Proml	Maximum Likelihood	AA	Optimized for protein data, strong statistical inference	Computationally expensive for large protein datasets

داریم از روش‌های فاصله‌ای نیز برای بازسازی درخت تبارزایی استفاده کنیم. این کار این امکان را می‌دهد تا روابط تکاملی بین سازواره‌ها را از دیدگاه‌های مختلف بررسی کنیم و نتایج را با یکدیگر مقایسه نماییم. با استفاده از روش‌های فاصله‌ای، می‌توانیم درخت‌های تبارزایی را با توجه به ماتریس فاصله‌های دو به دو بازسازی کنیم و به دنبال یافتن بهترین توپولوژی باشیم. این رویکرد مکمل این امکان را فراهم می‌کند تا نتایج خود را تأیید کنیم و درک دقیق‌تری از روابط تکاملی بین ارگانسیم‌ها به دست آوریم. همچنین، در این پژوهش مشخص شد که ژن *BRCA1* که نقش مهمی در تقسیم سلولی دارد در طول تکامل دچار جهش‌های متفاوتی شده است که این جهش‌ها بعد از انباشته شدن و پذیرفته شدن منجر شده تا شکل تکاملی متفاوتی برای گونه‌های مختلف دور یا نزدیک ایجاد شود. همچنین، این پژوهش نشان داد که نه فقط روال‌های مختلف ایجاد درخت‌های تبارزایی برای یک ژن می‌تواند نتایج متفاوتی را ایجاد کند بلکه الگوریتم‌های متفاوت یک رول نیز جواب متفاوتی را ایجاد می‌کنند. در این پژوهش مشخص شد که در اغلب روال‌ها و الگوریتم‌های داخل روال-ها، گونه‌های گاوسانان و گوسفندسانان کنار هم خوشه تکاملی بسیار نزدیک به هم تشکیل می‌دهند. بنابراین، نتایج تکامل یک گونه می‌تواند برای گونه دیگر استفاده شود.

استفاده از گاو به‌عنوان مدل حیوانی برای بررسی بیماری‌های انسانی مانند سرطان سینه می‌تواند به توسعه درمان‌های هدفمند و پیش‌بینی خطر ابتلا کمک کند. همچنین، بررسی تنوع ژنتیکی و تأثیر عوامل محیطی بر عملکرد این ژن می‌تواند به بهبود مدیریت سلامت گاوها و کاهش هزینه‌های درمانی منجر شود. بنابراین، ایجاد درخت فیلوژنی برای ژن *BRCA1* نه تنها به شناسایی جهش‌های مرتبط با بیماری کمک می‌کند، بلکه می‌تواند نقش مهمی در بهبود برنامه‌های اصلاح نژاد و مدیریت سلامت در گاو‌داری‌ها ایفا کند. به‌طور کلی، ژن *BRCA1* که برای حفظ پایداری ژنومی از راه نقش آن در ترمیم DNA و تنظیم چرخه سلولی ضروری است، درجه بالایی از حفاظت تکاملی، به‌ویژه در ناحیه حلقوی N-ترمینال و نواحی BRCT ترمینال C را نشان می‌دهد. این مورد در این پژوهش و در سایر روش‌های حداکثر درست‌نمایی یا استنتاج بی‌زی نشان داده شده است که این نواحی حیاتی در میان گونه‌های مختلف مهره‌دار محافظت شده‌اند که از اهمیت عملکردی آن‌ها سخن می‌گوید (Lai & Sarkar, 2020).

نتیجه‌گیری کلی

در این پژوهش، از تارنمای T-REX برای ایجاد درخت تبارزایی روی ژن *BRCA1* استفاده شد. این درخت بر اساس روش‌های صفت پایه ساخته شده است، اما در آینده قصد

فهرست منابع

- Asif, M., & Khan, R. A. (2014a). Genetic markers associated with resistance to gastrointestinal nematodes in sheep. *Tropical Animal Health and Production*, 46(4), 807-814.
- Asif, M., & Khan, R. A. (2014b). Genetic resistance to parasitic worms in livestock: A review. *Journal of Animal Science and Biotechnology*, 5(1), 1-10.
- Ayatollahi, M., Hosseini Moghaddam, S. H., Mirhosseini, S. Z., & Ghavi Hossein-Zadeh, N. (2015). Study of Lactoferrin gene single nucleotide polymorphism and its relation with milk somatic cells in crossbred cows of Guilan province. *Animal Production Research*, 4(2), 87-94. [In Persian]
- Bannerman, D. D., Springer, H. R., Paape, M. J., Kauf, A. C., & Goff, J. P. (2008). Evaluation of breed-dependent differences in the innate immune responses of Holstein and Jersey cows to *Staphylococcus aureus* intramammary infection. *Journal of Dairy Research*, 75(3), 291-301.
- Benezra, M., Chevallier, N., Morrison, D. J., MacLachlan, T. K., El-Deiry, W. S., & Licht, J. D. (2003). BRCA1 augments transcription by the NF- κ B transcription factor by binding to the Rel domain of the p65/RelA subunit. *Journal of Biological Chemistry*, 278(29), 26333-26341.
- Bollback, J. P. (2002). Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution*, 19(7), 1171-1180.
- Buckley, N. E., Hosey, A. M., Gorski, J. J., Purcell, J. W., Mulligan, J. M., Harkin, D. P., & Mullan, P. B. (2007). BRCA1 regulates IFN- γ signaling through a mechanism involving the type I IFNs. *Molecular Cancer Research*, 5(3), 261-270.
- Bull, J. J., Huelsenbeck, J. P., & Cunningham, C. W. (1997). Partitioning and combining data in phylogenetic analysis. *Systematic Biology*, 42, 384-397.

- Caldart, E. T., Mata, H., Canal, C. W., & Ravazzolo, A. P. (2016). Phylogenetic analysis: basic concepts and its use as a tool for virology and molecular epidemiology. *Acta Scientiae Veterinariae*, 44, 20-20.
- Charleston, M. A., Hendy, M. D., & Penny, D. (1994). The effects of sequence length, tree topology, and number of taxa on the performance of phylogenetic methods. *Journal of Computational Biology*, 1(2), 133-151.
- Cheon, S., & Liang, F. (2008). Phylogenetic tree construction using sequential stochastic approximation Monte Carlo. *BioSystems*, 91(1), 94-107.
- Deb, R., Kumar, A., Chakraborty, S., Verma, A. K., Tiwari, R., Dhama, K., ... & Kumar, S. (2013). Trends in diagnosis and control of bovine mastitis: a review. *Pakistan Journal of Biological Sciences*, 16(23), 1653-1661.
- Fabreti, L. G., & Höhna, S. (2022). Bayesian inference of phylogeny is robust to substitution model over-parameterization. *bioRxiv*, 2022-02.
- Farris, J. S. (1977). Phylogenetic analysis under Dollo's Law. *Systematic Biology*, 26(1), 77-88.
- Farris, J. S. (1970). Methods for computing Wagner trees. *Systematic Biology*, 19(1), 83-92.
- Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, 27(4), 401-410.
- Felsenstein, J. (2004). Inferring phylogenies Sinauer Associates Inc. Sunderland, MA.
- Felsenstein, J., & Churchill, G. A. (1996). A Hidden Markov Model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution*, 13(1), 93-104.
- Friedman, L. S., & Hastie, N. D. (1994). The BRCA1 gene in hereditary breast and ovarian cancer. *Nature Genetics*, 7(3), 309-310.
- Gonnet, G., & Scholl, R. (2009). *Scientific Computation: Phylogenetic tree construction*. doi: 10.1017/CBO9780511815027.009
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307-321.
- Hall, B. G. (2005). Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. *Molecular Biology and Evolution*, 22(3), 792-802.
- Hammoud, M., Santos, C. M. D., & Gois, J. P. (2021). Visual Comparison of phylogenetic trees through iPhyloC, a new interactive web-based framework. *bioRxiv*, 2021-05.
- Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R., & Molineux, I. J. (1992). Experimental phylogenetics: generation of a known phylogeny. *Science*, 255(5044), 589-592.
- Hopkins, M. J., & St. John, K. (2021). Incorporating hierarchical characters into phylogenetic analysis. *Systematic Biology*, 70(6), 1163-1180.
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, 294(5550), 2310-2314.
- Huson, D. H., & Steel, M. (2004). Distances that perfectly mislead. *Systematic Biology*, 53(2), 327-332.
- Ilie, D. E., Gavojdian, D., Kusza, S., Neamț, R. I., Mizeranschi, A. E., Mihali, C. V., & Csiszter, L. T. (2023). Kompetitive allele specific pcr genotyping of 89 SNPs in romanian spotted and romanian brown cattle breeds and their association with clinical mastitis. *Animals*, 13(9), 1484.
- Jabbar, F., Irfan, M., Mustafa, G., & Ahmad, H. I. (2019). Bioinformatics approaches to explore the phylogeny and role of BRCA1 in breast cancer. *Critical Reviews™ in Eukaryotic Gene Expression*, 29(6).
- Keating, J. N., Sansom, R. S., Sutton, M. D., Knight, C. G., & Garwood, R. J. (2020). Morphological phylogenetics evaluated using novel evolutionary simulations. *Systematic Biology*, 69(5), 897-912.
- Kluge, A. G., & Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Systematic Biology*, 18(1), 1-32.
- Koshkarov, A., & Tahiri, N. (2024). Novel algorithm for comparing phylogenetic trees with different but overlapping taxa. *Symmetry*, 16(7), 790.
- Lai, J., & Sarkar, I. N. (2021). A phylogenetic approach to analyze the conservativeness of BRCA1 and BRCA2 mutations. In *AMIA Annual Symposium Proceedings (Vol. 2020, p. 677)*.
- Li, D., Trotta, L., Marx, H. E., Allen, J. M., Sun, M., Soltis, D. E., ... & Baiser, B. (2019). For common community phylogenetic analyses, go ahead and use synthesis phylogenies. *Ecology*, 100(9), e02788.
- Li, D. (2019). For common community phylogenetic analyses, go ahead and use synthesis phylogenies. *Ecology*, 100. doi: 10.1002/ecy.2788
- Li, S., Pearl, D. K., & Doss, H. (2000). Phylogenetic tree construction using Markov chain Monte Carlo. *Journal of the American Statistical Association*, 95(450), 493-508.
- Liu, P., Biller, P., Gould, M., & Colijn, C. (2022). Analyzing phylogenetic trees with a tree lattice coordinate system and a graph polynomial. *Systematic Biology*, 71(6), 1378-1390.
- Marcet-Houben, M., & Gabaldón, T. (2011). TreeKO: a duplication-aware algorithm for the comparison of phylogenetic trees. *Nucleic Acids Research*, 39(10), e66-e66.

- Nwezeobi, J., Onyegbule, O., Nkere, C., Onyeka, T., van Brunschot, S., Seal, S., & Colvin, J. (2020). *Phylogenetic analysis v1*. doi: 10.17504/protocols.io.bd44i8yw
- Otu, H. H., & Sayood, K. (2003). A new sequence distance measure for phylogenetic tree construction. *Bioinformatics*, 19(16), 2122-2130.
- Ouchi, T., Lee, S. W., Ouchi, M., Aaronson, S. A., & Horvath, C. M. (2000). Collaboration of signal transducer and activator of transcription 1 (STAT1) and BRCA1 in differential regulation of IFN- γ target genes. *Proceedings of the National Academy of Sciences*, 97(10), 5208-5213.
- Paull, T. T., Cortez, D., Bowers, B., Elledge, S. J., & Gellert, M. (2001). Direct DNA binding by Brca1. *Proceedings of the National Academy of Sciences*, 98(11), 6086-6091.
- Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, 53(5), 793-808.
- Posada, D., & Crandall, K. A. (2001). Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). *Molecular Biology and Evolution*, 18(6), 897-906.
- Roehe, R., Georgieva, S., Zhang, W., & Shaw, S. E. (2007). Genetic variation in resistance to gastrointestinal nematodes in sheep. *International Journal for Parasitology*, 37(12), 1331-1340.
- Rohlf, F. J., Chang, W. S., Sokal, R. R., & Kim, J. (1990). Accuracy of estimated phylogenies: effects of tree topology and evolutionary model. *Evolution*, 44(6), 1671-1684.
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539-542.
- Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425.
- Scott, R., & Gras, R. (2012, July). Comparing distance-based phylogenetic tree construction methods using an individual-based ecosystem simulation, EcoSim. In Artificial Life Conference Proceedings (pp. 105-110). One Rogers Street, Cambridge, MA 02142-1209, USA journals-info@mit.edu: MIT Press.
- Slater, G. J., Harmon, L. J., & Alfaro, M. E. (2012). Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution*, 66(12), 3931-3944. doi: 10.1111/j.1558-5646.2012.01723.x
- Swofford, D. L. (1996). *Phylogenetic inference*. Molecular Systematics, 2nd edition. Pp. 407-514.
- Tolouei, E., & Behrouzi, M. (2016). Genetic resistance to parasitic worms in livestock: A review. *Journal of Applied Animal Research*, 44(3), 345-352.
- Whelan, S., & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*, 18(5), 691-699.
- Wiens, J. J. (1998). Testing phylogenetic methods with tree congruence: phylogenetic analysis of polymorphic morphological characters in phrynosomatid lizards. *Systematic Biology*, 47(3), 427-444.
- Wright, A. M. (2019). A systematist's guide to estimating Bayesian phylogenies from morphological data. *Insect Systematics and Diversity*, 3(3), 2.
- Xu, C. F., Brown, M. A., Nicolai, H., Chambers, J. A., Griffiths, B. L., & Solomon, E. (1997). Isolation and characterisation of the NBR2 gene which lies head-to-head with the human BRCA1 gene. *Human Molecular Genetics*, 6(7), 1057-1062.
- Yang, R., Hu, S., & Gasbarre, L. C. (2004). Genetic basis of resistance to gastrointestinal nematodes in sheep. *Animal Genetics*, 35(4), 301-310.
- Yang, Z. (2006). *Computational molecular evolution*. OUP Oxford.
- Zou, S., Li, Q., Kong, L., Yu, H., & Zheng, X. (2011). Comparing the usefulness of distance, monophyly and character-based DNA barcoding methods in species identification: a case study of Neogastropoda. *PLoS One*, 6(10), e26619.