Bioeksperimen: Jurnal Penelitian Biologi
Volume 12, Issue 1, March 2026, pp. 01 – 13
p-ISSN: 2460-1365, e-ISSN: 2527-2799
https://journals2.ums.ac.id/bioeksperimen

Phylogenetic Analysis of Representative Mammalian MUC16 Supported by Comparative SEA Domain and Tandem Repeat Variation

Yudi Gebri Foenna*[1)], Ardhana Yulisma[2)], Nilüfer Şahin Calapoğlu[3)]
[1]Department of Biology, Faculty of Sciences and Technology, Universitas Medan Area, Jl. H. Agus Salim Siregar, Medan Tembung, Medan, 20223, Indonesia
[2]Department of Pharmacy, Faculty of Health Sciences, Universitas U'budiyah Indonesia, Jl. Alue Naga Kec. Syiah Kuala Desa Tibang, Banda Aceh, Aceh, Indonesia.
[3]Department of Medical Biology, Faculty of Medicine, Süleyman Demirel Üniversitesi, Süleyman Demirel Cd. Çünür Mahallesi, 32260, Isparta, Isparta, Türkiye

*Corresponding e-mail: yudigebrifoenna@staff.uma.ac.id; yudigebri97@gmail.com

How to cite:
Foenna, Y. G., Yulisma, A., & Calapoğlu, N. Şahin. (2026). The Phylogenetic Analysis of Representative Mammalian MUC16 Supported by Comparative SEA Domain and Tandem Repeat Variation. *Bioeksperimen: Jurnal Penelitian Biologi*, *12*(1), 1–13. https://doi.org/10.23917/bioeksperimen.v12i1.14233.

Article info

Article History:
Received: 12 December 2025, Revised: 6 Februari 2026, Available Online: 31 March 2026

Keywords:
MUC16, SEA domain, tandem repeat, mammalian evolution, phylogenetics, comparative genomics

Abstract

MUC16 is one of the largest mammalian mucins and exhibits substantial evolutionary variation in both sequence composition and structural modularity. Comparative analysis of SEA domain composition and tandem repeat architecture is therefore essential for understanding the evolutionary diversification of this gene across mammals. This study investigates the phylogenetic relationships of mammalian MUC16 and examines how variation in SEA domains and tandem repeats contributes to lineage-specific structural divergence. MUC16 nucleotide and protein sequences from 20 mammalian species representing Primates and Rodentia were retrieved from public databases. Multiple sequence alignment and phylogenetic reconstruction were conducted using the Neighbor-Joining method with 1,000 bootstrap replicates. SEA domains were annotated using the SMART database, while tandem repeats were identified with Tandem Repeats Finder. Structural features were evaluated using descriptive statistics, hierarchical clustering, and Spearman's rank correlation analysis. Phylogenetic reconstruction revealed a clear molecular separation between Primates and Rodentia with strong bootstrap support. Primate species generally exhibited conserved sequences and expanded SEA domain and tandem repeat architectures, whereas rodents displayed higher sequence divergence and reduced structural complexity. A moderate positive association between SEA domain number and tandem repeat count ($\rho = 0.44$) was observed, although this relationship did not reach statistical significance and is therefore interpreted as a biologically suggestive trend rather than evidence of coordinated evolution. Overall, the results indicate that MUC16 evolution follows lineage-dependent patterns shaped by both sequence divergence and domain-level remodeling. This comparative framework provides an evolutionary context for understanding structural diversity in mammalian mucins and offers a foundation for future functional and genomic investigations.

Introduction

Recent advances in comparative genomic and proteomic have revealed substantial evolutionary diversity within mucin gene families, particularly MUC16, one of the largest and most structurally complex transmembrane glycoproteins in mammals (Faruque *et al.*, 2025; Zhang *et al.*, 2024). MUC16 represents a compelling model for studying the evolutionary dynamics of large, repeat-rich genes that combine extensive

sequence divergence with conserved functional architecture (Pajic et al., 2022). Comparative analyses across mammals indicate that MUC16 exhibits pronounced interspecies variation in gene length, repeat organization, and domain composition, suggesting that its evolution has been shaped by lineage-specific selective pressures (Pajic et al., 2022; Faruque *et al.*, 2025).

In human, the MUC16 gene is located on chromosome 19 (19p13.2) and spans more than 430 kb (Perez & Gipson, 2008; Aithal *et al.*, 2018; Zhang *et al.*, 2024), encoding a protein exceeding 22,000 amino acids (Perez & Gipson, 2008; Zhang *et al.*, 2024; Das *et al.*, 2015). Its extraordinary size is largely attributable to extensive tandem repeat regions and the presence of multiple SEA (sea urchin sperm protein, enterokinase, agrin) domains, which are characteristic features of membrane-associated mucins. These structural modules are functionally important for proteolytic cleavage, membrane tethering, and regulation of extracellular interactions (Faruque *et al.*, 2025; White *et al.*, 2022). Comparative genomic studies have shown that while chromosomal synteny of MUC16 is broadly conserved among mammals, repeat-rich regions and domain copy numbers vary substantially, indicating recurrent expansion and contraction events during evolution (Gipson *et al.*, 2014).

Functionally, MUC16 contributes to epithelial barrier integrity, cell-cell interactions, and modulation of immune responses (Kufe, 2022). In specific biological context, including reproduction and cancer, altered MUC16 expression or processing has been associated with changes in cell adhesion (Perez & Gipson, 2008; Gipson *et al.*, 2008), signaling pathways (Gipson *et al.*, 2014), and immune recognition (Zhang *et al.*, 2024). However, most previous studies have emphasized these biomedical roles, whereas the evolutionary diversification of MUC16 across mammalian lineages has received comparatively limited attention. As a result, the extent to which structural variation in MUC16 reflects broader evolutionary patterns remains insufficiently characterized (Faruque *et al.*, 2025; White *et al.*, 2022).

Among the defining features of MUC16, SEA domains and tandem repeat arrays are particularly informative from an evolutionary perspective. SEA domains are evolutionarily conserved modules present in several mucins and are implicated in protein cleavage and structural stability. Although their core structure is conserved, both the number and sequence composition of SEA domains vary among species, suggesting lineage-dependent duplication and divergence (Faruque *et al.*, 2025; Duraisamy *et al.*, 2007; White *et al.*, 2022). Tandem repeat regions, by contrast, evolve rapidly through mechanisms such as replication slippage and unequal crossing-over, generating species-specific architectures that influence glycosylation density and surface properties. Variation in repeat number (VNTR) and organization has therefore been widely used as an indicator of molecular adaptation in mucin genes (Sulovari *et al.*, 2019).

Although prior studies have mapped the genomic structure of MUC16 across several species, most focus narrowly on its biomedical or diagnostic roles rather than its evolutionary context (Duraisamy *et al.*, 2007; Faruque *et al.*, 2025). Phylogenetic investigations are scarce and often limited by incomplete sequence annotation or by neglecting repetitive regions (Duraisamy *et al.*, 2007; Maeda *et al.*, 2004). Consequently, the evolutionary relationships among mammalian MUC16 orthologs and the contribution of domain-level remodeling to lineage divergence remain incompletely resolved.

In order to address this gap, the present study conducts a comparative phylogenetic analysis of MUC16 across representative mammalian species, with particular emphasis on the relationship between evolutionary divergence, SEA domain architecture, and tandem repeat variation. By integrating sequence-based phylogenetic reconstruction with structural annotation and quantitative analysis, this study aims to clarify how modular features of MUC16 have diversified among major mammalian lineages. Rather than testing functional mechanisms directly, the analysis provides an evolutionary framework that generates biologically informed hypotheses regarding the structural evolution of MUC16 and its potential implications for mammalian adaptation.

Materials and methods

1. Study Design and Workflow

This study employed a comparative molecular evolutionary framework to investigate phylogenetic relationships and structural variation of the MUC16 gene across representative mammalian species. The analytical workflow consisted of sequence retrieval, multiple sequence alignment, phylogenetic reconstruction, structural annotation of SEA domains and tandem repeats, and quantitative statistical

analysis. All analyses were conducted between September and November 2025 at the Ecology and Physiology Laboratory, Universitas Medan Area.

2. Hardware and Software

All computational analyses were performed on a Lenovo LOQ Essential 15IAX9E labtop (Intel® CoreTM i5-12450HX (2,4 GHz), 12 GB RAM, RTX3050 GPU, 512 GB SSD and a Windows 11 operating system. Sequence retrieval, alignment, and phylogenetic reconstruction were carried out using MEGA version 12. SEA domain identification was performed using the SMART (Simple Modular Architecture Research Tool) web server, while tandem repeat detection was conducted using Tandem Repeats Finder (TRF). Statistical analyses were performed using IBM SPSS Statistics version 20, and clustering heatmaps were generated using RStudio. Figures were finalized using Microsoft Excel.

3. Data Set

MUC16 nucleotide sequences from twenty mammalian species representing the orders Primates and Rodentia were retrieved from the NCBI Nucleotide Database using the Nucleotide Search interface. Species were selected based on the availability of complete or high-quality partial MUC16 gene sequences. The nucleotide accession numbers obtained were as follows: *Homo sapiens* (NM 001401501.2), *Pan paniscus* (XM 055102885.2), *Gorilla gorilla gorilla* (XM 055371568.2), *Symphalangus syndactylus* (XM 055237836.1), *Nomascus leucogenys* (XM 030821778.1), *Hylobates moloch* (XM 058425754.1), *Papio anubis* (XM 031659988.1), *Theropithecus gelada* (XM 025367835.1), *Cercocebus atys* (XM 012077369.1), *Macaca mulatta* (XM 028840242.1), *Chlorocebus sabaeus* (XM 073016992.1), *Colobus angolensis palliatus* (XM 011954968.1), *Trachypithecus francoisi* (XM 033195391.1), *Rhinopithecus roxellana* (XM 030936547.1), *Cebus imitator* (XM 037728349.1), *Sapajus apella* (XM 032251910.1), *Myodes glareolus* (XM 048434638.1), *Mus musculus* (NM 001428752.1), *Apodemus sylvaticus* (XM 052189402.1), dan *Rattus norvegicus* (NM 001428753.1). All nucleotide datasets were downloaded in FASTA format and stored locally for downstream analyses.

Corresponding MUC16 protein sequences for each species were also retrieved from the NCBI Protein Database through the same NCBI platform. The protein accession numbers were as follows: *Homo sapiens* (NP_001388430.1), *Pan paniscus* (XP_054958860.2), *Gorilla gorilla gorilla* (XP_055227543.1), *Symphalangus syndactylus* (XP_055093811.1), *Nomascus leucogenys* (XP_030677638.1), *Hylobates moloch* (XP_058281737.1), *Papio anubis* (XP_031515848.1), *Theropithecus gelada* (XP_025223620.1), *Cercocebus atys* (XP_011932759.1), *Macaca mulatta* (XP_028696075.1), *Chlorocebus sabaeus* (XP_072873093.1), *Colobus angolensis palliatus* (XP_011810358.1), *Trachypithecus francoisi* (XP_033051282.1), *Rhinopithecus roxellana* (XP_030792407.1), *Cebus imitator* (XP_037584277.1), *Sapajus apella* (XP_032107801.1), *Myodes glareolus* (XP_048290595.1), *Mus musculus* (NP_001415681.1), *Apodemus sylvaticus* (XP_052045362.1), and *Rattus norvegicus* (NP_001415682.1).

Both nucleotide and amino acid datasets were also used to evaluate variation in sequence length, number of SEA domains, and tandem repeat organization, forming the basis for comparative structural and phylogenetic analyses.

4. Multiple Sequence Alignment and Quality Control

Multiple sequence alignment (MSA) of MUC16 nucleotide sequences was performed using the ClustalW algorithm implemented in MEGA 12. Given the repeat-rich architecture of MUC16, alignment trimming was conducted conservatively to remove ambiguously aligned regions while retaining homologous segments shared across taxa. Tandem repeat regions were retained to preserve biologically meaningful variation; however, their contribution to the phylogenetic signal was interpreted cautiously. Manual inspection focused on repeat-rich segments exhibiting excessive gap density or uncertain homology to ensure alignment reliability. Preliminary inspection indicated that exclusion of highly repetitive segments did not alter the higher-level topology separating Primates and Rodentia, suggesting that the inferred phylogenetic relationships are robust to repeat-associated alignment variability. The final alignments were exported in MEGA format for downstream phylogenetic reconstruction.

5. Phylogenetic Reconstruction

Phylogenetic relationships among mammalian MUC16 sequences were reconstructed using the Neighbor-Joining (NJ) method (Saitou & Nei, 1987) as implemented in MEGA 12 (Kumar *et al.*, 2024). NJ was selected due to its computational efficiency and robustness when applied to extremely large, low-

complexity, and repeat-rich genes such as MUC16, for which model-based approaches may be computationally prohibitive or sensitive to alignment uncertainty (Kurt *et al.*, 2024; Yoshida & Nei, 2016). Evolutionary distances were calculated using the p-distance model, and gaps or missing data were treated using pairwise deletion to reduce bias associated with low-complexity regions. Node support was assessed using bootstrap resampling with 1,000 replicates (Felsenstein, 1985). The resulting topology was interpreted primarily for comparative and topological consistency with established mammalian taxonomy rather than for model-rich evolutionary parameter estimation. Accordingly, phylogenetic inference in this study is restricted to major clade-level relationships, and future analyses incorporating Maximum Likelihood or Bayesian frameworks may further refine evolutionary estimates for MUC16.

6. SEA Domain Annotation

SEA domains were identified using the SMART database based on full-length MUC16 protein sequences. All sequences were analyzed using the SMART "Normal Mode" search, which applies hidden Markov models to detect conserved domain architectures. Annotated SEA domains were recorded. Only confidently annotated SEA domains were included in downstream quantitative analyses.

7. Tandem Repeat Identification

Tandem repeat regions were identified using Tandem Repeats Finder (TRF) with default parameters (match = 2, mismatch = 7, indel = 7, minimum alignment score = 50, maximum period size = 500). Analyses were performed on nucleotide sequences, as TRF operates at the DNA level. Repeat regions exhibiting clear repeat units were retained, while overlapping or degenerate repeats were evaluated manually. Only non-everlapping tandem repeat segments were counted to ensure consistency across species.

8. Statistical and Comparative Analysis

Descriptive statistics were calculated for SEA domain counts and tandem repeat numbers across species. Because normality assumptions were not met for one or more variables, non-parametric statistical approaches were applied. Differences between Primates and Rodentia were evaluated using appropriate non-parametric tests, and Spearman's rank correlation coefficient was used to assess the association between SEA domain number and tandem repeat variation. In order to explore multivariate patterns, hierarchical clustering was performed using Z-score-standardized domain features, and results were visualized as heatmaps.

Results and discussion

1. Phylogenetic Reconstruction of Mammalian MUC16

The evolutionary relationships and cross-taxonomic conservation of MUC16 were examined through phylogenetic reconstruction based on orthologous nucleotide sequences derived from twenty representative mammalian species. This comparative approach enables an assessment of how MUC16 has diversified across major mammalian lineages, while situating observed sequence variation within a broader evolutionary context. By applying a standardized phylogenetic framework, the resulting topology highlights patterns of divergence and conservation that are broadly consistent with establish mammalian taxonomy. The reconstructed phylogeny, presented in Figure 1, serves as a reference framework for interpreting lineage-specific variation and evolutionary separation of MUC16 among the examined taxa.

Figure 1. Neighbor-Joining phylogenetic tree of MUC16 from 20 mammalian species.

The Neighbor-Joining phylogenetic analysis of MUC16 across twenty mammalian species produced a well-resolved topology that largely mirrors accepted mammalian evolutionary relationships, while also revealing lineage-specific divergence within the gene. The inferred tree shows a major bifurcation separating Primates and Rodentia, accompanied by pronounced differences in branch lengths that reflect variation in evolutionary rates among lineages.

Within primate, Old World monkeys (including *Papio anubis, Theropithecus gelada, Cercopithecus aethiops, Macaca mulatta,* and *Chlorocebus sabaeus*) form a compact and strongly supported cluster, with bootstrap values consistently approaching or reaching 100. The short branch lengths observed within this group indicate high sequence conservation and relatively slow divergence of MUC16 between Cercopithecidae. A closely related Colobinae subgroup (*Colobus angolensis palliatus*, *Trachypithecus francoisi*, and *Rhinopithecus roxellana*) also exhibits maximal bootstrap support and minimal branch distances, suggesting recent common ancestry and a stable MUC16 sequence configuration across these taxa.

The Hominoidea clade (including *Homo sapiens, Pan paniscus, Gorilla gorilla*, *Hylobates moloch, Symphalangus syndactylus*, and *Nomascus leucogenys*) shows a clearly defined branching pattern with robust bootstrap support (predominantly 98-100), indicating strong topological confidence. Great apes cluster closely with short internal branches, reflecting minimal divergence among their MUC16 orthologs, whereas lesser apes form a sister lineage characterized by moderately longer branches. This pattern suggests slightly higher divergence rates in lesser apes relative to great apes, though still lower than those observed in other primate groups.

New World monkeys (*Cebus imitator* and *Sapajus apella*) occupy a distinct position within the primate clade, supported by high bootstrap values (>90) but characterized by longer branch lengths compared with Catarrhini. This pattern indicates increased sequence divergence and potential lineage-specific modifications of MUC16 within Platyrrhini.

Rodents species form a distinct and well-supported clade comprising *Muc musculus, Rattus norvegicus, Apodemus sylvaticus,* and *Myodes glareolus*. This clade is characterized by markedly longer branch lengths relative to primates, suggesting higher substitution rates and accelerated molecular evolution of MUC16 within Rodentia. In particular, the extended external branches observed for *Apodemus sylvaticus* and *Myodes glareolus* indicate pronounced lineage-specific divergence. All rodent nodes are strongly supported (bootstrap values ranging from 92 to 100), reinforcing the reliability of the inferred relationships.

Overall, bootstrap analysis based on 1000 replicates demonstrates high confidence across major nodes, particularly within primate subclades where values frequently reach 100, indicating strong topological stability. In contrast, deeper ancestral nodes separating major mammalian lineages show moderately lower

support, a pattern expected given the substantial evolutionary distances involved. Branch lenghts, representing estimated substitutions per site, vary considerably among lineages. Short branch lengths within Hominoidea and Cercopithecidae suggest relatively conservative evolution of MUC16 in these groups, consistent with functional constraints acting on core regions of the gene associated with epithelial protection, immune modulation, and cell-cell interactions (Gipson *et al.*, 2014; Perez & Gipson, 2008). Conversely, the longer branches observed in rodents are indicative of accelerated molecular evolution and greater lineage-specific divergence, a trend previously reported for mucin genes under selective pressures related to reproduction, immune challenges, and environmental exposure (Duraisamy *et al.*, 2007; White *et al.*, 2022).

2. Comparative Structural Organization of MUC16

In order to contextualize the phylogenetic patterns observed for MUC16, a comparative analysis of domain-level structural organization was conducted across representative mammalian lineages. While phylogenetic reconstruction elucidates evolutionary relatedness, examination of structural architecture provides complementary insight into how specific regions of MUC16 are conserved or diversified among species. Owing to the exceptional size and repetitive nature of MUC16, schematic comparisons were restricted to three representative taxa (*H. sapiens, C. atys,* and *M. glareolus*) to facilitate clear visualization of lineage-specific structural patterns (Figure 2).

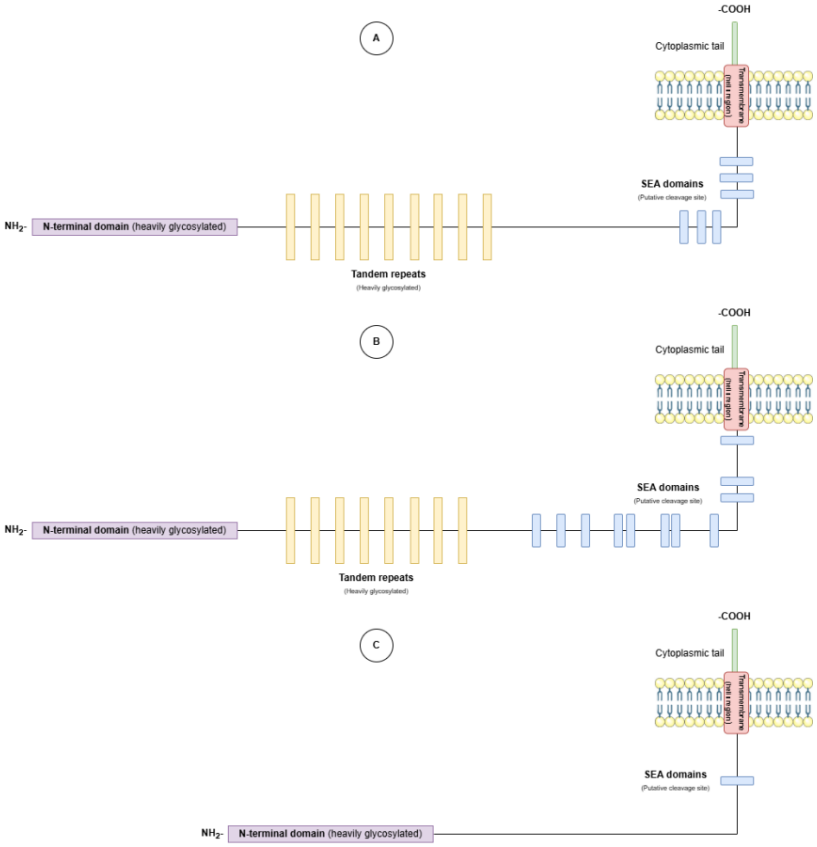

Figure 2. Schematic comparison of MUC16 structural organization in three mammalian species: (A) *Homo sapiens*, (B) *Cercocebus atys*, and (C) *Myodes glareolus*. Diagrams illustrate the major extracellular domains, including the tandem repeat region, SEA domains proximal to the transmembrane segment, and the conserved cytoplasmic tail.

As shown in Figure 2, MUC16 exhibits pronounced interspecies variation in modular architecture despite sharing a conserved overall organization consisting of an extensive extracellular domain, tandem repeat, SEA domains, and a C-terminal transmembrane-cytoplasmic module. In *H. sapiens*, both the N-terminal region and tandem repeat are markedly expanded, consistent with the documented structural complexity and functional specialization of human MUC16 in mucosal protection and disease-associated

processes. *Cercocebus atys* displays an intermediate architectural profile, retaining substantial tandem repeat regions but exhibiting fewer SEA domains, reflecting moderate divergence among non-human primates. In contrast, the rodent *M. glareolus* shows a compact configuration characterized by reduced tandem repeat content and a limited number of SEA domains, a pattern consistent with streamlined mucin architectures commonly reported in Rodentia (Maeda *et al.*, 2004).

When interpreted alongside the phylogenetic reconstruction, these structural differences align closely with evolutionary relationships among species. Primate lineages generally exhibit expanded SEA domain clusters and larger tandem repeat, supporting previous observations that mucins in higher primates have undergone modular proliferation, potentially enhancing glycosylation diversity and epithelial barrier complexity (Faruque *et al.*, 2025; Zhang *et al.*, 2024). The particularly elaborate architecture of human MUC16 has been associated with its established roles in immune modulation, reproductive biology, and tumor progression (Kufe, 2022). Conversely, the reduced structural organization observed in rodents supports the view that selective pressures in this lineage may favor genomic streamlining rather than extensive expansion of mucin modularity (Maeda *et al.*, 2004).

The three illustrated species thus represent a biologically informative cross section of mammalian MUC16 diversity, capturing structurally elaborate, intermediate, and compact architectures across evolutionary distance. By emphasizing proportional domain organization rather than absolute sequence length, the schematic provides a conceptual framework for visualizing lineage-dependent structural variation in MUC16, complementing the sequence-based phylogenetic analysis.

3. Quantitative Comparison of SEA Domains and Tandem Repeats

A quantitative comparison was conducted to characterize interspecific variation in MUC16 structural architecture by examining its two principal modular components (tandem repeats and SEA domains) across twenty mammalian species representing Primates and Rodentia. This analysis enables assessment of structural expansion or reduction within and between major lineages, providing a complementary perspective to sequence-based phylogenetic inference. Variation in tandem repeat number and SEA domain composition has long been recognized as a defining feature of mucin evolution and is shaped by lineage-specific selective pressures acting on mucosal defense, immune modulation, and reproductive function (Duraisamy *et al.*, 2007). The distribution patterns are summarized in Figure 3.

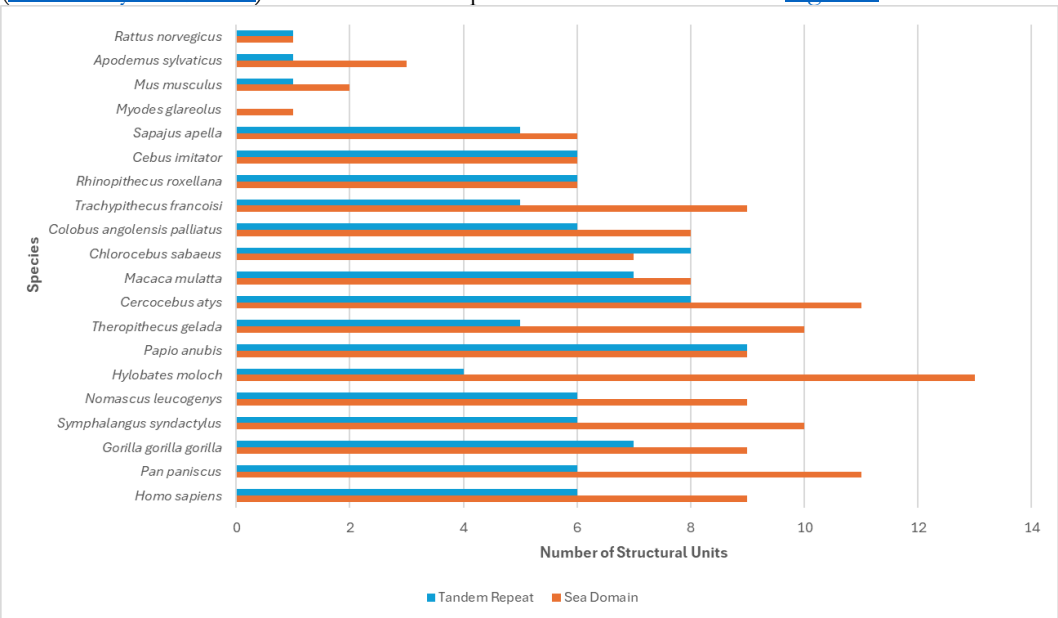

Figure 3. Comparative distribution of MUC16 structural features across selected mammalian species. Bar chart illustrating the number of tandem repeats and SEA domains identified in MUC16 from 20 species representing Primates and Rodentia.

As shown in Figure 3, primate species generally exhibit higher numbers of both tandem repeats and SEA domains than rodents, indicating greater structural complexity of MUC16 within the primate lineage.

This pattern is consistent with previous genomic analyses of mucin genes, including MUC5AC and MUC5B, which reported increased protein length variability in primates driven largely by variation in tandem repeat number (Plender *et al.*, 2024). Several primates, including *H. moloch, P. anubis,* and *P. paniscus*, show pronounced SEA domain expansion (≥10 domains), a feature also reported for other membrane-associated mucins such as MUC1 and MUC4 (Kufe, 2022; Ganguly *et al.*, 2020; Pei & Grishin, 2017; Dhar & McAuley, 2019).

Despite substantial variation in SEA domain counts, tandem repeat numbers among primates remain relatively conserved, with most species exhibiting six to eight repeat units. This pattern aligns with earlier findings indicating that tandem repeat arrays, while highly variable across mammals, tend to be more conserved within closely related primate lineages (Ahmad *et al.*, 2020; Arabfard *et al.*, 2022; Faruque *et al.*, 2025). Notable exceptions include *T. gelada* and *T. francoisi*, which display extensive SEA domain expansion while maintaining moderate tandem repeat counts. *Homo sapiens* exhibits a balanced structural profile characterized by both high SEA domain numbers and an expanded tandem repeat region, consistent with the documented structural elaboration of human MUC16 and its established roles in mucosal defense and tumor biology (Kufe, 2022).

In contrast, rodent species (including *R. norvegicus, A. sylvaticus, M. musculus*, and *M. glareolus*) display markedly reduced structural architectures. Tandem repeat counts are minimal, often ≤2, and SEA domain numbers remain low, reflecting contraction of mucin gene families commonly observed in Rodentia (Pajic *et al.*, 2022). This pattern is consistent with reported genomic streamlining and reduced selective pressure for mucin diversification in rodents, particularly for genes associated with mucosal surface specialization (Pajic *et al.*, 2022; Roycroft *et al.*, 2021).

To complement the absolute feature counts shown in Figure 3, hierarchical clustering based on standardized tandem repeat and SEA domain features was performed to examine relational patterns among species (Figure 4). This multivariate approach highlights relative similarity rather than magnitude and has been widely applied in comparative analyses of mucin modular organization (Shen & Li, 2016; Adams & Collyer, 2019).

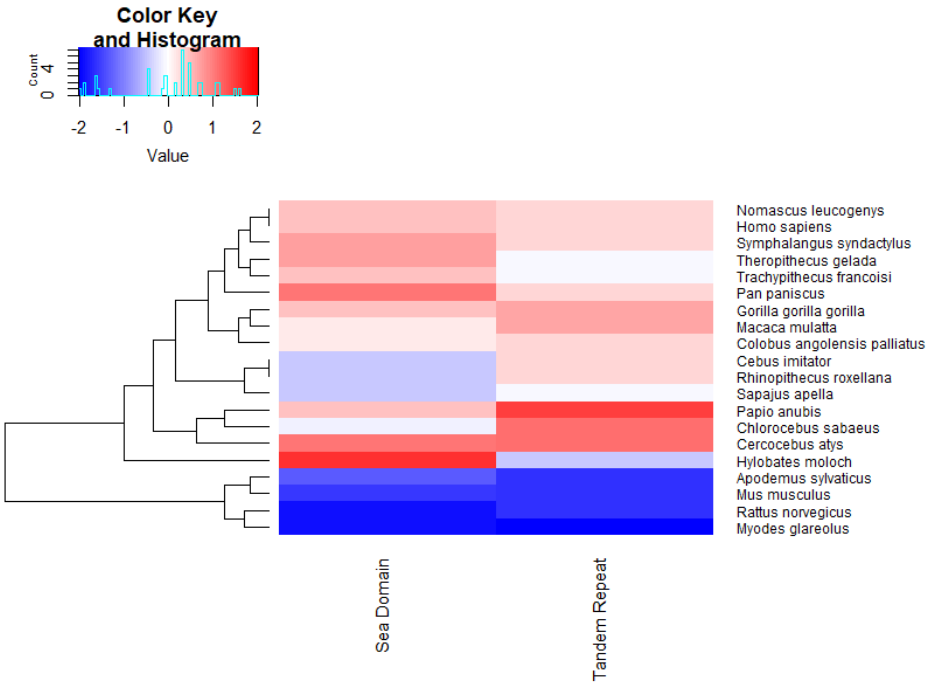

Figure 4. Hierarchical clustering heatmap of MUC16 structural features derived standardized tandem repeat and SEA domain counts across 20 mammalian species.

The heatmap reveals distinct clustering patterns that broadly correspond to taxonomic affiliations, with primates forming groups separate from rodents. Species with similar domain architectures cluster closely, whereas others occupy more isolated positions, indicating unique structural profiles. These multivariate patterns reinforce the evolutionary relationships inferred from both sequence-based phylogeny and absolute structural feature comparisons (Pajic *et al.*, 2022; Roycroft *et al.*, 2021).

4. Association Between SEA Domains and Tandem Repeats

Structural variation in MUC16 across the examined mammalian species was quantitatively characterized by summarizing the distribution of SEA domain and tandem repeat numbers. These features represent key components of mucin modular architecture and are widely recognized as major contributors to interspecies differences in glycoprotein structure and evolutionary diversification (Perez & Gipson, 2008; Faruque *et al.*, 2025; White *et al.*, 2022). Descriptive statistics provide an essential quantitative framework for evaluating the extent of structural expansion and contraction prior to inferential assessment. Summary metrics are presented in Table 1.

Table 1. Summary statistics of SEA domain and tandem repeat counts across 20 mammalian species

Variable	Mean	SD	Median	Min	Max
SEA domain	7.40	3.41	8.50	1	13
Tandem repeat	5.015	2.54	6.00	0	9

SEA domain numbers exhibit substantial heterogeneity, ranging from 1 to 13 units (mean = 7.40; SD = 3.41). The median value (8.50) exceeds the mean, indicating a right-skewed distribution driven primarily by species with expanded SEA domain arrays. Such patterns are consistent with prior observations that membrane-associated mucins undergo lineage-specific modulation of SEA domain architecture (Faruque *et al.*, 2025; Zhang *et al.*, 2024).

Tandem repeat counts display a mean of 5.02 and a standard deviation of 2.54, reflecting moderate dispersion relative to SEA domains. The observed range extends from the absence of detectable repeats in several rodents to pronounced repeat amplification in multiple primate taxa. Variation in tandem repeat number is consistent with established models of mucin evolution, in which repeat expansion contributes to increased glycan density and surface complexity (Sulovari *et al.*, 2019; Ganguly *et al.*, 2020; Verbiest *et al.*, 2022; Taniguchi *et al.*, 2017). Together, these descriptive patterns indicate that MUC16 structural diversification is highly lineage dependent.

In order to examine whether variation in SEA domain number is associated with corresponding changes in tandem repeat content, a nonparametric Spearman's rank correlation analysis was performed. This method is appropriate for datasets that deviate from normality and for evolutionary traits shaped by repetitive sequence dynamics (Fotsing *et al.*, 2019; Pajic *et al.*, 2022; Wang *et al.*, 2024). The relationship between the two structural features is illustrated in Figure 5.

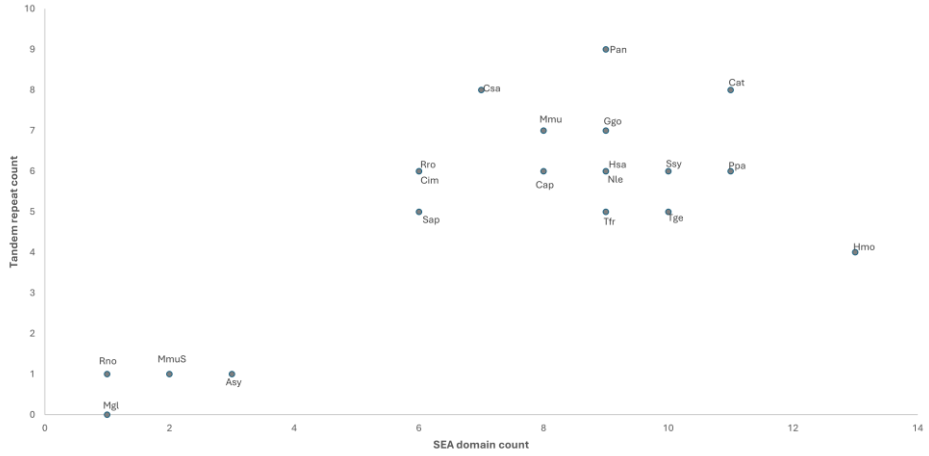

Figure 5. Scatterplot showing the association between SEA domain and tandem repeat count across 20 mammalian species. Abbreviations: Hsa (*Homo sapiens*), Ppa (*Pan paniscus*), Ggo (*Gorilla gorilla*

***gorilla*), Ssy (*Symphalangus syndactylus*), Nle (*Nomascus leucogenys*), Hmo (*Hylobates moloch*), Pan (*Papio anubis*), Tge (*Theropithecus gelada*), Cat (*Cercocebus atys*), Mmu (*Macaca mulatta*), Csa (*Chlorocebus sabaeus*), Cap (*Colobus angolensis palliatus*), Tfr (*Trachypithecus francoisi*), Rro (*Rhinopithecus roxellana*), Cim (*Cebus imitator*), Sap (*Sapajus apella*), Mgl (*Myodes glareolus*), MmuS (*Mus musculus*), Asy (*Apodemus sylvaticus*), and Rno (*Rattus norvegicus*).**

The analysis reveals a moderate positive association (ρ = 0.44) between SEA domain and tandem repeat counts; however, this relationship does not reach conventional statistical significance (p = 0.052). Accordingly, the result should be interpreted as a biologically suggestive trend rather than evidence of coordinated evolution. Similar tendencies toward parallel expansion of mucin structural modules have been reported in comparative genomic studies, although such patterns are often shaped by lineage-specific constraints and stochastic variation (Duraisamy *et al.*, 2007; Sulovari *et al.*, 2019).

Visual inspection of the scatterplot indicates lineage-associated clustering. Primate species tend to occupy regions corresponding to higher values of both structural features, whereas rodents cluster at lower values, reflecting compact MUC16 architectures. These distributions are consistent with previously reported contrasts in mucin gene organization between Primates and Rodentia (Maeda *et al.*, 2004; Pajic *et al.*, 2022). Intermediate positions observed in several primate taxa further suggest that MUC16 structural evolution proceeds along a continuum rather than discrete categorical states.

This study provides a comparative evolutionary framework that integrates phylogenetic reconstruction with quantitative and domain-level structural analyses of MUC16 across representative mammalian lineages. Unlike previous studies that primarily emphasized biomedical relevance or limited taxonomic scope (McLemore & Aouizerat, 2005; Faruque *et al.*, 2025; Zhang *et al.*, 2024; White *et al.*, 2022), the present work highlights lineage-specific variation in SEA domains and tandem repeats as key contributors to MUC16 diversification. Representative structural models of *H. sapiens, C. atys,* and *M. glareolus* further demonstrate how sequence divergence translates into distinct architectural patterns.

Several limitations should be acknowledged. The present analysis is restricted to representative species from Primates and Rodentia, and therefore the inferred evolutionary patterns should be interpreted within this taxonomic scope rather than as general features of all mammals. Domain annotation and tandem repeat detection depend on the quality of publicly available genomic resources, and the Neighbor-Joining method provides limited modeling resolution compared with likelihood-based approaches. In addition, SEA domain identification relied on a single annotation platform (SMART), which has been widely applied for domain annotation in mucin proteins. While cross-validation using complementary databases such as Pfam or InterPro would further strengthen confidence in domain assignment.

Future studies expanding taxonomic sampling, applying model-based phylogenetic inference, integrating multiple annotation platforms and incorporating glycoproteomic or functional data will be valuable for testing the biological implications of MUC16 modular variation, which should be regarded as hypotheses rather than direct outcomes of the present analysis.

Conclusion

This study presents a comparative evolutionary analysis of MUC16 across mammalian lineages by integrating phylogenetic reconstruction with quantitative assessments of SEA domain and tandem repeat variation. The clear molecular separation between Primates and Rodentia reflects lineage-specific evolutionary trajectories, with primates exhibiting expanded domain architectures and rodents retaining comparatively compact structures. Although a moderate positive association between SEA domain number and tandem repeat count was observed, this relationship should be interpreted as a biologically suggestive trend rather than evidence of coordinated evolution. Collectively, these findings demonstrate that MUC16 diversification follows structured, lineage-dependent patterns rather than random variation. This evolutionary framework provides a foundation for future studies examining the functional and adaptive significance of mucin modularity across mammals..

Author Statements

Acknowledgements and funding statements: The authors express their sincere appreciation to the Department of Biology, Faculty of Sciences and Technology, Universitas Medan Area, for providing

research facilities and institutional support during the study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing of interest: The authors declare that there are no financial or personal relationships that could be construed as a potential conflict of interest.

Author's contributions: YGF was responsible for conceptualization, study design, data curation, formal analysis, methodology, visualization, and manuscript drafting. AY performed data analysis, validation, literature review, and critical revision of the manuscript. NŞC provided supervision, methodological guidance, interpretation of results, and final approval of the manuscript.

Generative AI: Not applicable

Data availability: All data used in this study were obtained from publicly available databases, including the NCBI Nucleotide and Protein repositories. The accession number of all sequences analyzed are provided within the Materials and Methods section. Processed data and additional analytical outputs supporting the findings of this study are available from the corresponding author upon reasonable request.

References

Adams, D., & Collyer, M. (2019). Phylogenetic Comparative Methods and the Evolution of Multivariate Phenotypes. Annual Review of Ecology, Evolution, and Systematics, 50(1): 1-21. https://doi.org/10.1146/annurev-ecolsys-110218-024555.

Ahmad, S., Singchat, W., Jehangir, M., Suntronpong, A., Panthum, T., Malaivijitnond, S., & Srikulnath, K. (2020). Dark Matter of Primate Genomes: Satellite DNA Repeats and Their Evolutionary Dynamics. Cells, 9(12): 2714. https://doi.org/10.3390/cells9122714.

Aithal, A., Rauth, S., Kshirsagar, P., Shah, A., Lakshmanan, I., Junker, W., Jain, M., Ponnusamy, M., & Batra, S. (2018). MUC16 as a novel target for cancer therapy. Expert Opinion on Therapeutic Targets, 22: 675-686. https://doi.org/10.1080/14728222.2018.1498845.

Arabfard, M., Salesi, M., Nourian, Y., Arabipour, I., Maddi, A., Kavousi, K., & Ohadi, M. (2022). Global abundance of short tandem repeats is non-random in rodents and primates. BMC Genomic Data, 23(1): 77. https://doi.org/10.1186/s12863-022-01092-4.

Das, S., Majhi, P., Al-Mugotir, M., Rachagani, S., Sorgen, P., & Batra, S. (2015). Membrane proximal ectodomain cleavage of MUC16 occurs in the acidifying Golgi/post-Golgi compartments. Scientific Reports, 5: 9759. https://doi.org/10.1038/srep09759.

Dhar, P., & McAuley, J. (2019). The Role of the Cell Surface Mucin MUC1 as a Barrier to Infection and Regulator of Inflammation. Frontiers in Cellular and Infection Microbiology, 9: 117. https://doi.org/10.3389/fcimb.2019.00117.

Duraisamy, S., Ramasamy, S., Kharbanda, S., & Kufe, D. (2007). Distinct evolution of the human carcinoma-associated transmembrane mucins, MUC1, MUC4 AND MUC16. Gene. 373: 28-34. https://doi.org/10.1016/j.gene.2005.12.021.

Faruque, M., Medha, M.M., Mahfuz, A.M.U.B., Islam, M.M. and Siraj, M.A. (2025), Deciphering Deleterious nsSNPs in MUC16's SEA Domain: Structural and Functional Implications in Cancer Metastasis via Computational Analysis. J Cell Mol Med, 29: e70633. https://doi.org/10.1111/jcmm.70633

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution, 39: 783-791.

Fotsing, S., Margoliash, J., Wang, C., Saini, S., Yanicky, R., Shleizer-Burko, S., Goren, A., & Gymrek, M. (2019). The impact of short tandem repeat variation on gene expression. Nature genetics, 51: 1652-1659. https://doi.org/10.1038/s41588-019-0521-9.

Ganguly, K., Rauth, S., Marimuthu, S., Kumar, S., & Batra, S. K. (2020). Unraveling mucin domains in cancer and metastasis: when protectors become predators. Cancer metastasis reviews, 39(3): 647–659. https://doi.org/10.1007/s10555-020-09896-5

Gipson, I., Spurr-Michaud, S., Tisdale, A., & Menon, B. (2014). Comparison of the Transmembrane Mucins MUC1 and MUC16 in Epithelial Barrier Function. PLoS ONE, 9(6): e100393. https://doi.org/10.1371/journal.pone.0100393.

Gipson, I., Blalock, T., Tisdale, A., Spurr-Michaud, S., Allcorn, S., Stavreus-Evers, A., & Gemzell, K. (2008). MUC16 Is Lost from the Uterodome (Pinopode) Surface of the Receptive Human Endometrium: In Vitro Evidence That MUC16 Is a Barrier to Trophoblast Adherence1, 78(1): 134 - 142. https://doi.org/10.1095/biolreprod.106.058347.

Kufe, D. W. (2022). Chronic activation of MUC1-C in wound repair promotes progression to cancer stem cells. J Cancer Metastasis Treat. 8: 12. http://dx.doi.org/10.20517/2394-4722.2022.03

Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., and Tamura. (2024). Molecular Evolutionary Genetics Analysis Version 12 for adaptive and green computing. Molecular Biology and Evolution. 41(12): msae263. https://doi.org/10.1093/molbev/msae263

Kurt, S., Bouchard-Côté, A., & Lagergren, J. (2024). Sparse Neighbor Joining: rapid phylogenetic inference using a sparse distance matrix. Bioinformatics, 40(12): btae701. https://doi.org/10.1093/bioinformatics/btae701.

Maeda, T., Inoue, M., Koshiba, S., Yabuki, T., Aoki, M., Nunokawa, E., Seki, E., Matsuda, T., Motoda, Y., Kobayashi, A., Hiroyasu, F., Shirouzu, M., Terada, T., Hayami, N., Ishizuka, Y., Shinya, N., Tatsuguchi, A., Yoshida, M., Hirota, H., Matsuo, Y., Tani, K., Arakawa, T., Carninci, P., Kawai, J., Hayashizaki, Y., Kigawa, T., & Yokoyama, S. (2004). Solution Structure of the SEA Domain from the Murine Homologue of Ovarian Cancer Antigen CA125 (MUC16)*. Journal of Biological Chemistry, 279(13): 13174 - 13182. https://doi.org/10.1074/jbc.m309417200.

McLemore, M. R., & Aouizerat, B. (2005). Introducing the MUC16 gene: implications for prevention and early detection in epithelial ovarian cancer. Biological Research for Nursing, 6(4): 262–267. https://doi.org/10.1177/1099800404274445

Pajic, P., Shen, S., Qu, J., May, A., Knox, S., Ruhl, S., & Gokcumen, O. (2022). A mechanism of gene evolution generating mucin function. Science Advances, 8: eabm8757. https://doi.org/10.1126/sciadv.abm8757.

Pei, J. and Grishin, N.V. (2017), Expansion of divergent SEA domains in cell surface proteins and nucleoporin 54. Protein Science, 26(3): 617-630. https://doi.org/10.1002/pro.3096

Perez, B., & Gipson, I. (2008). Focus on Molecules: human mucin MUC16. Experimental Eye Research, 87(5): 400-401. https://doi.org/10.1016/j.exer.2007.12.008.

Plender, E., Prodanov, T., Hsieh, P., Nizamis, E., Harvey, W., Sulovari, A., Munson, K., Kaufman, E., O'Neal, W., Valdmanis, P., Marschall, T., Bloom, J., & Eichler, E. (2024). Structural and genetic diversity in the secreted mucins MUC5AC and MUC5B. American Journal of Human Genetics, 111(8): 1700-1716. https://doi.org/10.1016/j.ajhg.2024.06.007.

Saito, N. and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phhylogenetic trees. Molecular Biology and Evolution, 4(4): 406-425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Roycroft, E., Achmadi, A., Callahan, C., Esselstyn, J., Good, J., Moussalli, A., & Rowe, K. (2021). Molecular Evolution of Ecological Specialisation: Genomic Insights from the Diversification of Murine Rodents. Genome Biology and Evolution, 13(7): evab103. https://doi.org/10.1093/gbe/evab103.

Shen, W., & Li, Y. (2016). A novel algorithm for detecting multiple covariance and clustering of biological sequences. Scientific Reports, 6: 30425. https://doi.org/10.1038/srep30425.

Sulovari, A., Li, R., Audano, P., Porubsky, D., Vollger, M., Logsdon, G., Warren, W., Pollen, A., Chaisson, M., & Eichler, E. (2019). Human-specific tandem repeat expansion and differential gene expression during primate evolution. Proceedings of the National Academy of Sciences, 116(46): 23243-23253. https://doi.org/10.1073/pnas.1912175116.

Taniguchi, T., Woodward, A., Magnelli, P., McColgan, N., Lehoux, S., Jacobo, S., Mauris, J., & Argüeso, P. (2017). N-Glycosylation affects the stability and barrier function of the MUC16 mucin. The Journal of Biological Chemistry, 292(26): 11079-11090. https://doi.org/10.1074/jbc.m116.770123.

Verbiest, M., Maksimov, M., Jin, Y., Anisimova, M., Gymrek, M., & Sonay, T. (2022). Mutation and selection processes regulating short tandem repeats give rise to genetic and phenotypic diversity across species. Journal of Evolutionary Biology, 36(2): 321-336. https://doi.org/10.1111/jeb.14106.

Wang, C., Weaver, S., Boonpattrawong, N., Schuster-Little, N., Pantakar, M., & Whelan, R. (2024). A Revised Molecular Model of Ovarian Cancer Biomarker CA125 (MUC16) Enabled by Long-read

Sequencing. Cancer Research Communications, 4(1): 253-263. https://doi.org/10.1158/2767-9764.crc-23-0327.

White, B., Patterson, M., Karnwal, S., & Brooks, C. (2022). Crystal structure of a human MUC16 SEA domain reveals insight into the nature of the CA125 tumor marker. Proteins: Structure. 90(5): 1210-1218. https://doi.org/10.1002/prot.26303.

Yoshida, R., & Nei, M. (2016). Efficiencies of the NJp, Maximum Likelihood, and Bayesian Methods of Phylogenetic Construction for Compositional and Noncompositional Genes. Molecular biology and evolution, 33(6): 1618-1624 . https://doi.org/10.1093/molbev/msw042.

Zhang, X., Hong, L., & Ling, Z. (2024). MUC16: clinical targets with great potential. Clinical and Experimental Medicine, 24: 101. https://doi.org/10.1007/s10238-024-01365-5.