

Similarity Analysis of Complete Blood Count (CBC) Reference Interval Distributions Across Ethnic and Geographic Populations

Kunlin Wu¹, Abicumaran Uthamacumaran³, Hector Zenil^{1,2,3*}

¹Algorithmic Dynamics Lab, Departments of Biomedical Computing & Digital Twins, King's Institute for AI, Faculty of Life Sciences & Medicine, King's College London, UK

² Cancer Research Interest Group, The Francis Crick Institute, London, UK

³ Oxford Immune Algorithmics, Oxford University Innovation & London Institute for Healthcare Engineering, UK

Abstract

Blood reference intervals (RIs) are central to diagnosis and therapeutic monitoring across diseases and conditions worldwide and are key in current human medicine practice, yet most widely adopted health systems were developed in Western populations and assumed to be universally applicable. This not only creates a risk of misclassification in regions with diverse demographic or physiological profiles, but their clinical, biological, genetic and geographical relevance can be put into question. In this study, we focus on one of the main blood panels, the Complete Blood Count (CBC), a widely and routinely used clinical test. We compiled CBC reference interval data from 28 countries and applied a multi-stage analytical framework. We evaluated potential structural similarity using multiple clustering strategies pairing different linkage rules with Euclidean, correlation-based, and information-theoretic distance metrics. To benchmark sensitivity, we introduced a Two-Level Cohesion Score to measure continent-level grouping. Finally, UMAP embeddings with feature importance scores assessed which analytes, if any, drive geography-related separation. Using BMI as a cross-country positive control experiment, known to be limited but their results shown to be similar across physiologically similar populations, CBC reference intervals showed no reproducible clustering by geography or population genetic similarities, and weak signals were limited to MCV and HGB, not stable across methods in contrast to BMI. These findings indicate that CBC physiology is not only not coordinated but likely the arbitrary result of lab equipment capabilities or sensitivity, legacy systems or logistical decisions in national or local laboratory practices. Our findings support a transition away from one-size-fits-all global RIs that appear arbitrary when seen globally toward geography-adaptive and personalized reference frameworks. By demonstrating the absence of geography-based structure in CBC physiology, this work supports the development of personalized RIs that unites precision and predictive medicine with diagnostic equity in patient care. The analytical workflow developed here provides a transparent template for advancing precision hematology, inclusivity, and equity in clinical care.

* Corresponding author: hector.zenil@kcl.ac.uk

Introduction

Reference intervals (RIs) are widely applied in clinical laboratory testing as standard of care for disease management and screening. Their establishment directly influences diagnostic sensitivity/specificity and therapeutic monitoring [1-2] and makes the difference between labelling a person normal or abnormal when it comes to general health. Current systems largely derive these intervals from an “average population” model often incorporating sex and ages differences and pregnant status but often overlooking ethnicity, geography, and environmental factors. RIs originate from statistical summaries and lack fine-grained individual characterization [3]. Studies (e.g., NHANES) show CBC differences by race/ethnicity for HGB, WBC, and PLT, implying that “universal” RIs may not be globally applicable [4]. A harmonization view (e.g., IFCC) promotes unified international standards to improve comparability [5], whereas others warn uniform RIs can obscure risk signals and exacerbate health inequities in heterogeneous populations [6-7]. Machine learning and deep clustering have been explored to model latent variability [8-9], but without standardized QC or multinational sources these remain experimental [10]. Some argue population RIs are insufficient for precision medicine and advocate individual reference values (RCV) or longitudinal trajectories [11], consistent with integrating genomic, environmental, lifestyle, and social data for individualized prevention/treatment [12-13]. Yet RI systems still rely on Western, European-descent data, lacking cross-national/ethnic/geographical modeling, leading to reduced accuracy and inequity, especially in developing regions and minority populations [7,14-15]. BMI shows systematic ethnic/regional variation [16], suggesting even “universal” indicators can be population-specific.

In this work we explore the following hypothesis. As routine blood tests are globally massively performed on the population for diagnostics purposes driving up or informing up to 70% of diagnosis. If RIs across the globe are grounded on any objective clinical or biological feature, they should be related geographically or ethnically across countries distributed around the world. Here, we will test whether RIs of a Complete Blood Count (CBC, also known as Full Blood Count) exhibit any geography- or ethnicity(genetic)-related similarity structure across countries, and provide a transparent framework that informs movement from a “single standard” toward geo-ethnicity-adaptive and individualized systems. The CBC is the most popular medical tests with over 4 billion performed every year worldwide, one of the 2-3 main blood panels, comprised in over 90% of all blood tests, and therefore an important proxy for this investigation.

Empirical cross-national assessments for foundational tests like CBC remain scarce and lack structured, international agreements. We explore CBC RI data across 28 countries to examine variation across countries and continents to evaluate potential physiological similarities or dissimilarities—addressing debates on global standardization vs localization, population- vs individual-based modeling, arbitrary rule-based partitioning vs clinical-driven discoveries. We focus on four gaps: (1) regional bias in data sources (NHANES [18], Japan [19], NORIP [20]) with limited

representation from Africa, Middle East, South Asia, Latin America [5,7,21,22] raising equity concerns [23]; (2) coarse granularity due to summary-level data (intervals or mean $\pm$ 2SD) that mask intra-population heterogeneity and lack longitudinal tracking [24-27]; (3) methodological homogeneity with limited structural modeling and little analyte-level discriminability quantification (mutual information, permutation importance, random forest) limiting interpretability and decision support; (4) theoretical homogeneity assuming universality, despite arguments for context-specific, dynamically adjusted values and ongoing RCV debates without a bridge from population structure to individual baselines [5-7,28]. Response strategy. Build a cross-continental CBC RI database; combine variability mapping, clustering, and interpretable dimensionality reduction; compare linkage–distance combinations; quantify analyte contributions; and propose a pathway toward geo-ethnicity-adaptive and ultimately individualized RIs.

Methods

The aims of this study were to compile a multinational CBC reference interval (RI) database and incorporate BMI as a positive benchmark case; quantify cross-country variability in CBC using visualization and SDR; apply multiple clustering strategies to both CBC and BMI and benchmark their geographical sensitivity using the Two-Level Cohesion Score; use UMAP with feature-importance metrics to evaluate whether geography contributes to latent structure in CBC versus BMI; and, finally, develop a methodological framework that integrates variability mapping, clustering evaluation, and interpretable dimensionality reduction as a foundation for future geography-adaptive RI systems.

Data Collection

CBC: We first sought official national RI websites; most countries lacked standardized CBC RIs. Reasons include lack of unified standards, intra-country regional variation [29], and policies granting local autonomy (e.g., Dubai Health Authority) [30]. We adopted diversified sourcing: 12 countries from peer-reviewed RI studies; 9 from university-affiliated/teaching or public hospitals; 3 from local health authorities; 1 textbook; 1 medical-school hematology notes PDF; 1 private lab; 1 unofficial document attributed to an authoritative hospital. Where authoritative sources were absent, we substituted with geographically/demographically similar neighbors; for geographically unique cases (e.g., Greece), we used the most credible available source. We annotated source type and confidence level for transparency. Final dataset: 28 countries. We collected RIs for WBC, Platelets, MCHC, MCV, MCH, RBC, HGB, HT. Sweden discontinued routine MCHC reporting as of 2015-10-14 (Karolinska), so MCHC is blank [31]. Hospital-derived RIs for WBC, Platelets, MCHC, MCV, MCH were typically not sex-specific; academic sources were generally sex-stratified. Non-sex-specific RIs were treated as identical for males and females; when sex-specific RIs were missing for typically stratified analytes (e.g.,

HGB, HCT, RBC in Canada, Philippines), we assumed equivalence. Country-level sources are in Appendix A.

BMI (positive case): We obtained mean BMI for the same 28 countries from the WHO Global Health Observatory [32] for both sexes across six age groups (60–64, 65–69, 70–74, 75–79, 80–84, 85+), yielding 12 sex–age strata to mirror MI-based top-5 feature selection. Although latest year was 2016, BMI served as a methodological benchmark, not temporal surveillance.

Data Analysis

All analyses and figures used Python frameworks. Comparative visualization with SDR. For each analyte/sex, we visualized RIs and computed relative SDR (coefficient of variation) by taking country-level midpoints, then the mean and SD of midpoints, and SD/mean as SDR; each country was annotated with source type and confidence level 1–3.

Phylogenetic trees: We constructed univariate (per analyte, by sex) and multivariate (all eight analytes, by sex) trees. We benchmarked diverse linkage–distance pairs to test robustness across dependence structures: Euclidean for magnitude-based geometry, Manhattan for L1 robustness, Cosine/Correlation for orientation and shape similarity, and Mutual Information (MI) for nonlinear and distributional dependencies, providing complementary sensitivity profiles across analytes and sexes [33] (See Table 1). We compared Ward and Average (UPGMA) linkage [33]. For univariate Average linkage, Euclidean and Manhattan yielded identical structures; MI/Cosine/Correlation are not applicable to univariate. Multivariate trees used Ward+Euclidean; Average with Euclidean, Cosine, Manhattan, Correlation; and Average+MI. MI distance required feature selection by MI to “Continent”: top-5 features (and a top-3 sensitivity) to balance noise vs coverage. Only three analytes had nonzero MI; two zero-score analytes were added to reach five, which still altered dendrograms due to their contribution to global pairwise distances. We repeated clustering with top-3 only to assess stability. MI discretization used quantile binning with five bins; alternative binning would change outcomes.

Two-Level Cohesion Score: For each dendrogram, we computed mean cophenetic distance across within-continent country pairs, then averaged across continents; lower scores indicate stronger geographical cohesion. We summarized scores across all linkage–distance combinations for univariate and multivariate analyses. We computed matrix-level statistics including Mantel tests (10 000 permutations), Pearson/Spearman concordance between distance and similarity matrices, Receiver Operating Characteristic – Area Under the Curve (ROC-AUC) for discriminative association, and MDS stress/variance explained to quantify geometric coherence. We also computed Cohen’s *d* effect size to quantify standardized mean separation between high- and low-correlation pairs, and an Odds Ratio (Fisher’s exact test) comparing the likelihood of strong correlations among short- versus long-distance

pairs to provide interpretable effect-size validation. Additionally, we computed distance-correlation (dCor) between the flattened upper triangles of the distance matrix and (*1-correlation*) to capture nonlinear concordance, and quantified variance partitioning by the eigenvalue spectrum of the classical MDS (Gower-centered) matrix (reporting cumulative variance for 2D–5D).

UMAP & feature importance: We embedded male and female CBC data with UMAP to visualize high-dimensional structure, then quantified analyte contributions via absolute Pearson correlations with UMAP1/UMAP2 (averaged), permutation importance (projection sensitivity), and random forest importance (Gini impurity reduction). We applied the same pipeline to BMI for a positive-case benchmark.

Linkage Method	Distance Metric	Applicability	Sensitivity	Rationale
Ward	Euclidean	Univariate + Multivariate	Absolute magnitude differences	Baseline metric for variance-minimizing hierarchical structure
Average (UPGMA)	Euclidean	Univariate + Multivariate	Absolute magnitude differences	Benchmark against Ward to assess linkage dependence
Average (UPGMA)	Manhattan	Multivariate only	Robustness to large deviations	Captures additive (L1) structure and reduces influence of outliers
Average (UPGMA)	Cosine	Multivariate only	Directional similarity	Tests whether countries cluster by relative analyte <i>profiles</i> rather than raw levels
Average (UPGMA)	Correlation	Multivariate only	Shape/orientation of co-fluctuations	Detects similarity in analyte <i>patterns</i> independent of absolute values
Average (UPGMA)	Mutual Information	Multivariate only	Nonlinear/dependence structures	Captures non-Euclidean similarity potentially missed by linear metrics

Table 1. Phylogenetic Tree clustering algorithms. Overview of linkage–distance combinations evaluated for phylogenetic modeling of CBC reference intervals. Euclidean is applicable to both univariate and multivariate analyses, while Cosine, Correlation, Manhattan, and Mutual Information distances are applied only to multivariate feature spaces.

Results

Cross-national variability in CBC reference intervals

Figure 1A–D shows the between-country distribution of WBC and MCV RIs, stratified by sex and annotated with data-source confidence levels. Both analytes display modest overall dispersion and no reproducible continent-level structure, and male-female patterns are concordant across countries. The SDR analysis confirms that between-population variability is low relative to within-population variance,

indicating that these parameters are not geographically partitioned at the reference-interval level.

Other CBC analytes (RBC, HGB, HCT, MCH, MCHC, Platelets; additional panels in Figure 1) similarly exhibit heterogeneous but non-geographic scatter, where wider ranges in select countries appear attributable to local laboratory conventions, not regionally conserved physiology. Missing MCHC values for Sweden reflect a documented change in institutional reporting policy rather than biological divergence. Collectively, Figure 1 demonstrates that CBC reference intervals do not exhibit intrinsic geography-linked stratification, motivating the subsequent multivariate clustering analyses.

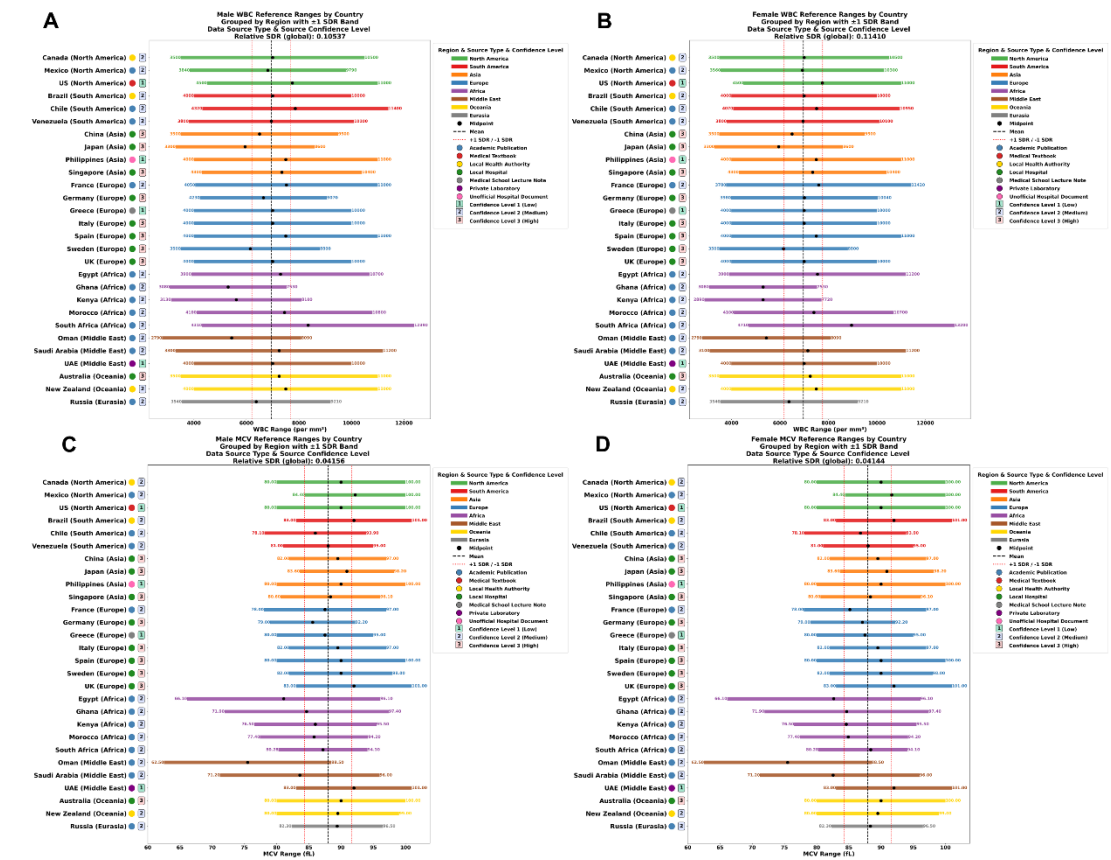


Figure 1. A) Male WBC Reference Ranges by Country and Region, with Data Source Type and Confidence Level. B) Female WBC Reference Ranges by Country and Region, with Data Source Type and Confidence Level. C) Male MCV Reference Ranges by Country and Region, with Data Source Type and Confidence Level. D) Female MCV Reference Ranges by Country and Region, with Data Source Type and Confidence Level.

Multivariate clustering of CBC versus BMI

Figure 2A–B presents multianalyte phylogenetic clustering of CBC reference intervals by sex. Across all tested linkage–distance combinations, no persistent grouping of countries by continent is observed. Occasional local neighbourhood effects collapse when the metric or feature set is altered, indicating instability of any

apparent CBC geography signal. Additional dendrograms for individual CBC analytes and alternative linkage–distance specifications are provided in the Supplementary Information. Mutual-information clustering accentuates subtle structure but fails to reproduce after re-specification (top-5 vs top-3 features), demonstrating that such structure is method-dependent rather than biologically anchored.

In contrast, BMI clustering in Figure 2C–D forms continent-coherent groupings under the same modelling framework, particularly with MI and correlation distances. This establishes BMI as a positive benchmark, confirming that the pipeline is sensitive to population-level structure when present. The contrast between CBC and BMI directly supports the interpretation that CBC lacks stable geo-physiological boundaries, whereas BMI retains them.

Mutual-information feature ranking in the BMI trees (Figure 2C–D) showed that the geography signal was driven by mid- to late-adulthood BMI intervals: for males, the highest MI scores were observed for the 60–64 and 65–69 age groups (0.551 and 0.512), while for females the strongest contributions came from the 70–74 and 65–69 groups (0.581 and 0.531). This age-anchored concentration of MI signal indicates that continent-level clustering in BMI is not noise-driven but biologically structured, emerging in life-stages when adiposity reflects long-term regional lifestyle and environmental exposures rather than transient physiological fluctuation. In contrast, no analogous concentration or replicable “driver analyte” pattern was observed for CBC, explaining the failure of MI to stabilise continent-level groupings in CBC despite its strong performance on BMI.

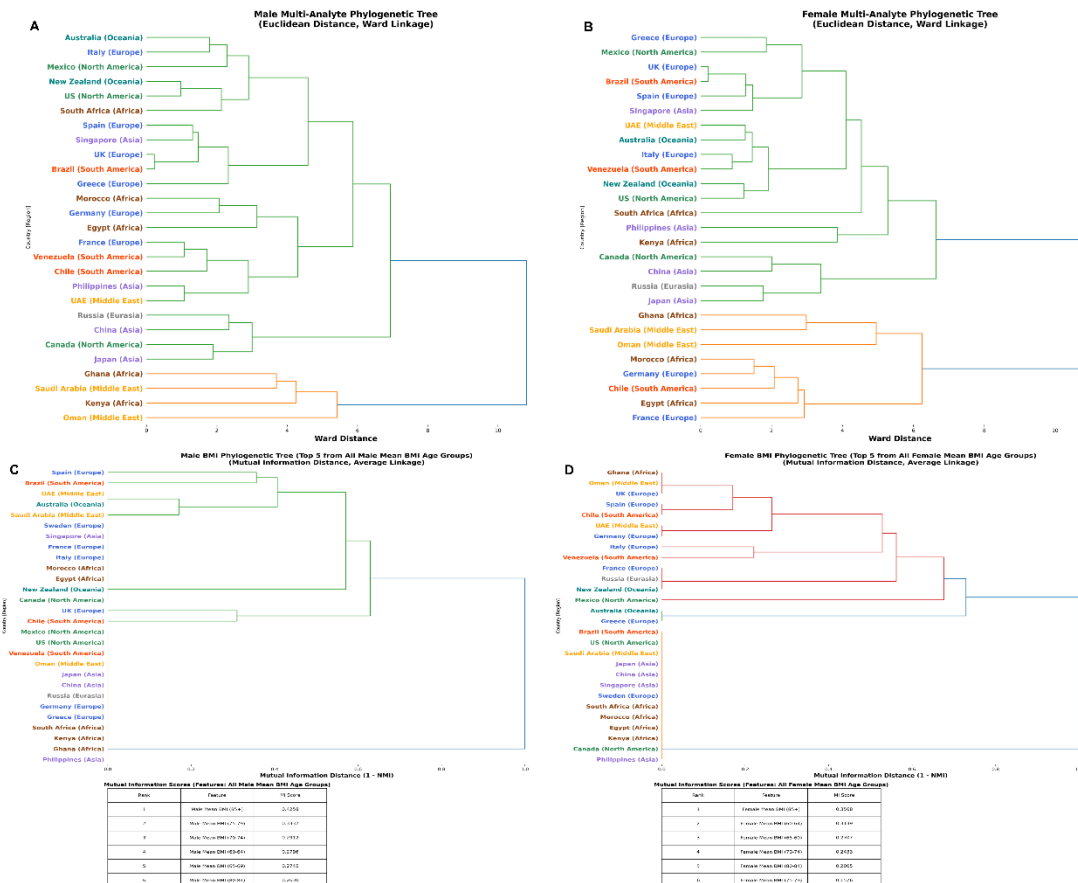


Figure 2. A) Male Multi-Analyte Phylogenetic Tree (Ward Linkage + Euclidean Distance). B) Female Multi-Analyte Phylogenetic Tree (Ward Linkage + Euclidean Distance). C) Multidimensional Phylogenetic Tree of Countries by All Male Mean BMI Age Groups (Average Linkage + Mutual Information Distance (Top 5)). D) Multidimensional Phylogenetic Tree of Countries by All Female Mean BMI Age Groups (Average Linkage + Mutual Information Distance (Top 5)).

Cohesion benchmarking

To quantify the strength of continent-level organisation in each clustering approach, Two-Level Cohesion Scores were computed (Tables 2–5). For male and female CBC (Tables 2 and 3), cohesion scores remained uniformly high across linkage–distance combinations, confirming weak within-continent similarity and the absence of reproducible geography-linked structure. Average + MI marginally lowered cohesion but did not stabilise any continent-level pattern, consistent with only weak and non-replicating geography effects in CBC.

In contrast, male and female BMI (Tables 4 and 5) exhibited substantially lower cohesion scores, particularly under information-theoretic and correlation-based distances. This concordance between topology (Figure 2) and cohesion benchmarking (Tables 2–5) demonstrates that BMI encodes genuine geographic structure, whereas the null result for CBC is a true negative rather than a methodological limitation. Complementary effect-size analyses further supported this matrix-level consistency.

High-correlation pairs were substantially closer in distance space (Cohen's $d = 1.24$), and pairs with short distances were ≈ 6 -fold more likely to exhibit strong correlation (Odds Ratio = 6.15, $p < 0.0001$) confirming that the observed CBC null pattern reflects genuine physiological homogeneity rather than measurement noise.

Quantitatively, mean cohesion scores across all non-Ward clustering strategies were 1.274 (male CBC) and 1.300 (female CBC), compared with 0.777 (male BMI) and 0.811 (female BMI), corresponding to an absolute ~ 0.49 – 0.50 reduction when geography genuinely contributes to structure. This corresponds to an approximate 38–42% relative reduction in within-continent cophenetic distance for BMI compared with CBC, representing a medium-to-large structural effect size. Notably, mutual-information distance ranked first in three of the four tables, but only yielded sustained low cohesion in the BMI case, confirming that MI functions as a true detector of structure rather than a false-positive amplifier. These findings quantitatively corroborate the dendrogram analyses and suggest that the absence of CBC clustering reflects a physiological non-association.

To further assess concordance between the distance and similarity structures, we computed additional global metrics on two 103×103 matrices summarizing all pairwise relationships among national analyte reference datasets: one capturing inter-country dissimilarity (distance) and the other profile similarity (correlation). The Mantel test yielded $r = 0.48$, $p = 0.0001$, indicating strong correspondence between the two matrices, i.e., the pairwise distance geometry and correlation structure. Pearson and Spearman coefficients (both ≈ 0.48 , $p < 0.0001$) confirmed this linear and rank-level agreement. The discriminative accuracy of distance for identifying highly correlated pairs (area under the curve; AUC = 0.86) and the 2D classical multidimensional scaling (MDS) variance explained (64 %) further demonstrate that the matrix topology is internally consistent and geometrically stable.

A nonlinear robustness check using the Székely–Rizzo distance-correlation (dCor) test yielded a high concordance between distance and similarity matrices (dCor = 0.74, $p < 0.001$). Variance partitioning based on the eigenvalue spectrum of the classical MDS analysis showed that 3D, 4D, and 5D components explained 78%, 87%, and 92% of total relational variance, respectively, indicating a low-dimensional, geometrically coherent structure. This further validates that the observed absence of continent-level CBC clustering may reflect a true physiological uniformity rather than an artifact of linear modeling or metric choice.

UMAP embeddings and feature importance

Figure 3A–B shows UMAP embeddings of CBC values for males and females. Country positions overlap extensively, and no continent-separated manifolds emerge. Feature-importance scoring (correlation, permutation, random-forest) identifies red-cell indices (most often MCV, sometimes HGB) as relatively more influential for local embedding geometry, but these signals remain non-geographic and sex-inconsistent, again confirming the absence of continent-level stratification. BMI UMAP embeddings in Figure 3C–D show clearer continent-separated gradients, reproducing the benchmark topology observed in Figure 2 and the low-cohesion scores in Tables 4–5. This further validates that the analytic workflow can detect when geography when it is encoded in the data.

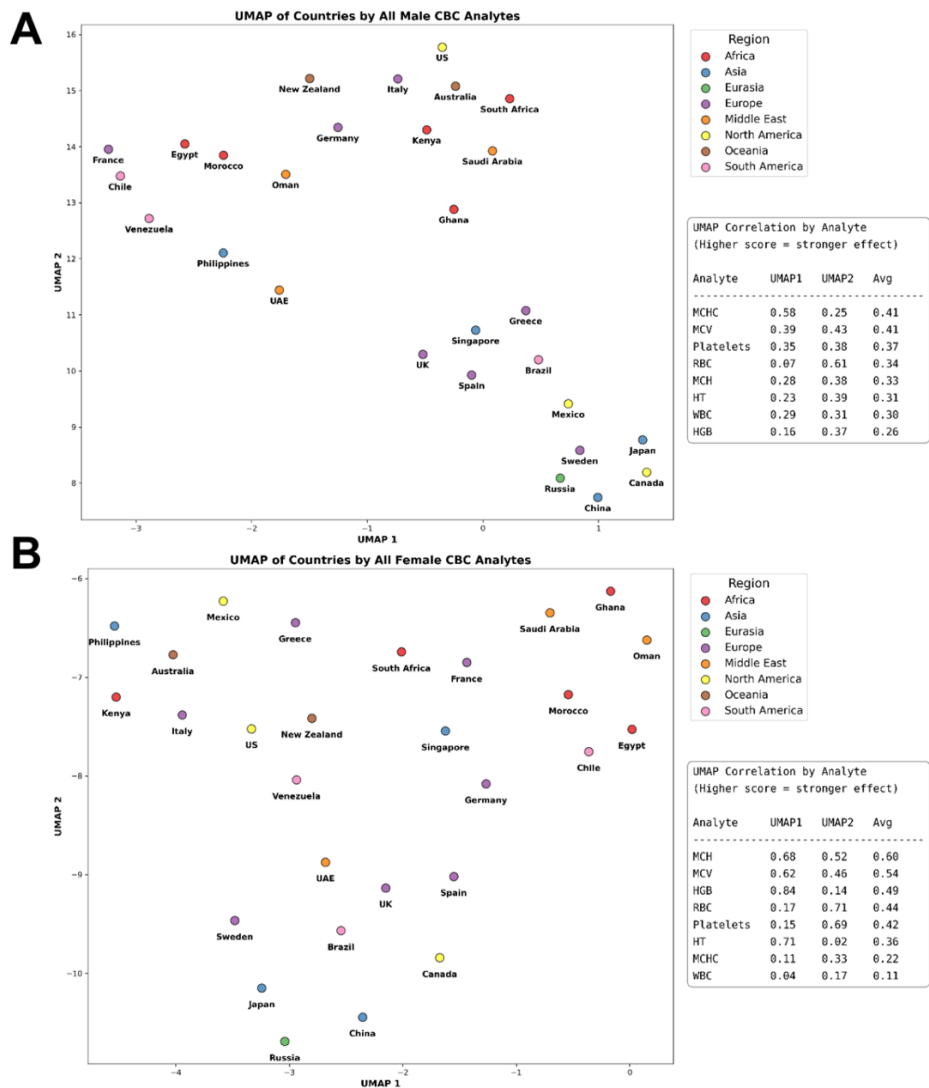


Figure 3. A) UMAP Visualization and Correlation Score of All Male CBC Analytes. B) UMAP Visualization and Correlation Score of All Female CBC Analytes.

Discussion

Key Findings and Interpretations

Across all analytic layers—including RI dispersion (Figure 1), hierarchical clustering (Figure 2), UMAP embeddings (Figure 3), and quantitative benchmarking using the Two-Level Cohesion Score (Tables 2–5)—CBC reference intervals showed **no** reproducible geography-driven structure. Mutual Information–based clustering occasionally produced weak, localised structure in red-cell indices, particularly MCV and sometimes HGB, but these effects were non-replicating across sexes, feature sets, or distance functions, indicating instability rather than a stable population signal. By contrast, BMI produced consistent continent-level organization under the same pipeline (Figure 2C–D; Figure 3C–D; Tables 4–5), validating that the negative CBC result is an absence of structure and not a modeling limitation.

These findings imply that CBC physiology is not governed by a coherent geographical pattern. Instead, as supported by recent haematological evidence, CBC values are individual-specific, not population-anchored. Foy et al. [34] demonstrated that personal hematologic setpoints persist for decades and outperform external population references in detecting deviation states. Similarly, intra-individual variability has been shown to be substantially lower than inter-individual variability, reinforcing the concept of personalized baselines as the clinically relevant anchor.

The present study extends this reasoning to the global scale: if CBC setpoints are intrinsically individual, then aggregating individuals into continents or ethnic blocks will not yield physiologically meaningful reference partitions, precisely what we observed. The SDR visualisation (Figure 1) first indicated low cross-national dispersion; clustering instability (Figure 2A–B) confirmed the absence of group structure; the cohesion benchmarking (Tables 2–5) demonstrated that even the most sensitive distance metric (MI) could not stabilise geography-based groupings for CBC; and the manifold geometry in UMAP embeddings (Figure 3A–B) showed that no latent spatial separation exists even in a nonlinear representation.

Critically, the success of BMI as a positive case (Figure 2C–D; Figure 3C–D) shows that when geography *does* drive biological structure, this pipeline detects it. The contrast therefore functions as an internal sensitivity control, supporting the conclusion that CBC lacks a continent-level phenotype.

Beyond the empirical findings, the modelling pipeline yields several methodological insights that advance RI research beyond descriptive comparison. First, weak apparent structure occasionally emerged under mutual-information clustering, but these effects

consistently disappeared once the feature set or discretisation scheme was re-specified, indicating that the signal was method-induced rather than physiologically anchored (Tables 2–5; Figure 2). Ward+Euclidean served as the most stable baseline configuration (Figure 2A–B), reinforcing that CBC deviations are not variance-cohesive in the way a geography-structured phenotype would be. By contrast, mutual-information distance was the only metric sensitive enough to detect the BMI geography signal (Figure 2C–D; Table 5), confirming that MI operates as a legitimate detector of population structure rather than merely amplifying noise. The UMAP embeddings (Figure 3) function as a nonlinear stress-test: if any underlying manifold separation existed in CBC, it would surface here; its absence therefore validates the negative dendrogram and cohesion results. Together with large effect sizes (Cohen’s $d = 1.24$) and a clinically interpretable enrichment (OR = 6.15; pairs with short distances were approximately 6× likelier to be highly correlated), the strong nonlinear concordance ($dCor = 0.74$) and low-dimensional geometry ($\geq 87\%$ variance by 4D) indicate that our matrix topology is internally consistent rather than a metric artifact, further supporting the conclusion that the absence of continent-level CBC structure reflects physiological homogeneity rather than modeling limitations.

Taken together, the lack of separation across all four analytical layers—RI dispersion (Figure 1), hierarchical topology (Figure 2), cohesion benchmarking (Tables 2–5), and nonlinear manifold structure (Figure 3), constitutes convergent evidence that CBC reference intervals do not encode continent-level structure. In short, the algorithmic result mirrors the physiological one: CBC is not globally structured, whereas BMI is.

Limitations and Future Directions

The study’s principal limitation is that CBC data are available only as reference intervals, not raw individual-level distributions. This prevents aggregation-level stability from being decomposed into within-population vs cross-population signals. Further, a second limitation is heterogeneity of national source authority, although this was explicitly annotated (Figure 1) and its effects were modeled structurally in downstream robustness checks. Mutual-information clustering illustrates a third limitation: high sensitivity to discretization choices, reinforcing the need for reproducibility safeguards when RI variance is low.

Future work should address these gaps by using large-scale individual-level CBC **datasets** with a diverse and multi-institutional cohort, as well as incorporating longitudinal measurements to reconstruct personal hematologic trajectories. In addition, prospective analyses can integrate genomic, environmental, and demographic variables to model the “individual baseline” as complex systems and

make mechanistic inferences. Lastly, we propose that formalising RI evolution toward adaptive or personalized reference systems, rather than geographically universal ones could benefit translatability to predictive medicine frameworks.

Future work should also integrate causal-mechanistic modelling to decode the latent generative structure underlying individual hematologic setpoints (i.e., attractor dynamics) and to identify minimal perturbations that shift diagnostic trajectories across populations. Methods from *algorithmic information dynamics* (AID) provide a computational/in silico framework for causal discovery, intervention modelling, and attractor reconstruction in dynamical systems lacking stable population structure [38-41]. Embedding these causal pattern discovery tools into global CBC datasets may clarify why physiology remains uniform across geography, strengthen equitable diagnostic baselines, and accelerate the translation toward personalized, adaptive reference systems in precision and predictive medicine.

Our findings carry important implications for accessible, equitable, and precision diagnostics. Demonstrating that CBC physiology remains geographically uniform reframes population reference intervals from region-specific correction factors to individualized longitudinal baselines, emphasizing that true diagnostic precision lies in tracking intra-individual trends over time. This positions CBC-derived circulatory biomarkers as translatable, low-cost (affordable) substrates for longitudinal monitoring and equitable access to personalized care.

Conclusions

Our study demonstrates that CBC RIs do not reflect continent-level or ethnicity-linked biological architecture. Rather, they align with the paradigm of individualized hematological setpoints, where personalized longitudinal stability instead of population averages, anchors interpretability. Conversely, BMI retains robust geography-linked structure, validating the modeling pipeline and confirming that the absence of CBC clustering is a physiological reality and not an analytic artefact.

These findings support a transition from population-based **to** individual-based reference systems in medicine, with CBC values interpreted against a personal baseline rather than a presumed universal norm. This transition could advance precision and diagnostic equity. Future development of dynamic, data-adaptive RI frameworks will require longitudinal and multimodal integration, eventually recalibrating precision hematology around the “human”, the individual, in personalized medicine.

REFERENCES

- [1] Ozarda Y. Reference intervals: current status, recent developments and future considerations. *Biochemia Medica*. 2016;26(1):5-16.
- [2] Jones G, Barker A. Reference intervals. *Clin Biochem Rev*. 2008;29(Suppl 1):S93.
- [3] Coskun A, Sandberg S, Unsal I, et al. Personalized reference intervals: from theory to practice. *Crit Rev Clin Lab Sci*. 2022;59(7):501-516.
- [4] Lim E, Miyamura J, Chen JJ. Racial/ethnic-specific reference intervals for common laboratory tests: a comparison among Asians, Blacks, Hispanics, and White. *Hawaii J Med Public Health*. 2015;74(9):302.
- [5] Ichihara K, Ozarda Y, Barth JH, et al. A global multicenter study on reference values: 1. Assessment of methods for derivation and comparison of reference intervals. *Clin Chim Acta*. 2017;467:70-82.
- [6] Kauh TJ, Read JG, Scheitler AJ. The critical role of racial/ethnic data disaggregation for health equity. *Popul Res Policy Rev*. 2021;40(1):1-7.
- [7] Price MA, Fast PE, Mshai M, et al. Region-specific laboratory reference intervals are important: A systematic review of the data from Africa. *PLOS Glob Public Health*. 2022;2(11):e0000783.
- [8] Velev J, LeBien J, Roche-Lima A. Unsupervised machine learning method for indirect estimation of reference intervals for chronic kidney disease in the Puerto Rican population. *Sci Rep*. 2023;13(1):17198.
- [9] Shomorony I, Cirulli ET, Huang L, et al. An unsupervised learning approach to identify novel signatures of health and disease from multimodal data. *Genome Med*. 2020;12(1):7.
- [10] Dockès J, Varoquaux G, Poline JB. Preventing dataset shift from breaking machine-learning biomarkers. *GigaScience*. 2021;10(9):giab055.
- [11] Fraser CG, Bartlett WA. Are reference change values more useful than population-based reference intervals [Internet]. Washington (DC): AACC; 2013 [cited 2018 May 24]. Available from: <https://www.aacc.org/community/aacc-academy/publications/scientificshorts/2013/are-reference-change-values-more-useful-than-population-based-reference-intervals>
- [12] Molla G, Bitew M. Revolutionizing personalized medicine: synergy with multi-omics data generation, main hurdles, and future perspectives. *Biomedicine*. 2024;12(12):2750.
- [13] Roberts MC, Fohner AE, Landry L, et al. Advancing precision public health using human genomics: examples from the field and future research opportunities. *Genome Med*. 2021;13(1):97.
- [14] Hu XL, Hassan H, Al-Dayel FH. Reference intervals for common biochemistry laboratory tests in the Saudi population by a direct a priori method. *Ann Saudi Med*. 2017;37(1):16-20.

- [15] Kelem A, Engidaye G, Addisu B, et al. Hematological reference intervals for adult population of Debre Berhan town, North East Ethiopia. *Sci Rep.* 2025;15(1):14121.
- [16] Karnes JH, Arora A, Feng J, et al. Racial, ethnic, and gender differences in obesity and body fat distribution: An All of Us Research Program demonstration project. *PLoS One.* 2021;16(8):e0255583.
- [17] Angelo A, Derby G, Demtse A, et al. Umbilical cord blood hematological parameters reference interval for newborns from Addis Ababa, Ethiopia. *BMC Pediatr.* 2021;21(1):275.
- [18] Cheng CKW, Chan J, Cembrowski GS, et al. Complete blood count reference interval diagrams derived from NHANES III: stratification by age, sex, and race. *Lab Hematol.* 2004;10(1):42-53.
- [19] Takami A, Watanabe S, Yamamoto Y, et al. Reference intervals of white blood cell parameters for healthy adults in Japan. *Int J Lab Hematol.* 2021;43(5):948-958.
- [20] Rustad P, Felding P, Franzson L, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. *Scand J Clin Lab Invest.* 2004;64(4):271-284.
- [21] Al-Mawali A, Pinto AD, Al-Busaidi R, et al. Comprehensive haematological indices reference intervals for a healthy Omani population: First comprehensive study in Gulf Cooperation Council (GCC) and Middle Eastern countries based on age, gender and ABO blood group comparison. *PLoS One.* 2018;13(4):e0194497.
- [22] Shaikh MS, Ahmed S, Khalid A, et al. Establishment of population specific reference intervals in healthy Pakistani adults for 21 routine and special haematology analytes. *Ejifcc.* 2022;33(3):220.
- [23] Mendez KM, Reinke SN, Kelly RS, et al. A roadmap to precision medicine through post-genomic electronic medical records. *Nat Commun.* 2025;16(1):1700.
- [24] Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat Med.* 2017;36(5):855-875.
- [25] Veroniki AA, Seitidis G, Tsivgoulis G, et al. An introduction to individual participant data meta-analysis. *Neurology.* 2023;100(23):1102-1110.
- [26] Pyrkov TV, Avchaciov K, Tarkhov AE, et al. Longitudinal analysis of blood markers reveals progressive loss of resilience and predicts human lifespan limit. *Nat Commun.* 2021;12(1):2765.
- [27] Schüssler-Fiorenza Rose SM, Contrepois K, Moneghetti KJ, et al. A longitudinal big data approach for precision health. *Nat Med.* 2019;25(5):792-804.
- [28] Fraser CG. Reference change values. *Clin Chem Lab Med.* 2012;50(5):807-812.
- [29] Elderderly AY, Alshaiban AS. Reference value profile for healthy individuals from the Aljouf region of Saudi Arabia. *J Hematol.* 2017;6:6-11.
- [30] Dubai Health Authority. Standards for clinical laboratory services. Code:

- DHA/HRS/HPSD/ST-28, Issue 2. Dubai: Dubai Health Authority; 2023 [cited 2025 Aug 22]. Available from:
<https://www.dha.gov.ae/uploads/052023/Standards%20for%20Clinical%20Laboratory%20Services2023552664.pdf>
- [31] Karolinska Universitetssjukhuset. Blodstatus [Blood Status] [Internet]. Solna: Karolinska Universitetssjukhuset; 2015 [cited 2025 Jul]. Available from:
<https://www.karolinska.se/KUL/Alla-anvisningar/Anvisning/9021>.
- [32] World Health Organization. Mean body mass index [Internet]. Geneva: World Health Organization; 2016 [cited 2025 Aug 22]. Available from:
<https://platform.who.int/data/maternal-newborn-child-adolescent-ageing/indicator-explorer-new/MCA/mean-body-mass-index>
- [33] Felsenstein J. Inferring Phylogenies. Sunderland (MA): Sinauer Associates; 2004.
- [34] Foy BH, Petherbridge R, Roth MT, et al. Haematological setpoints are a stable and patient-specific deep phenotype. *Nature*. 2025;637(8045):430-8.
- [35] Ganie MA, Chowdhury S, Suri V, et al. Variation in normative values of major clinical biochemistry analytes in healthy reproductive-age women in India: A subset of data from a National Indian Council of Medical Research-Polycystic Ovary Syndrome task force study. *Indian J Pharmacol*. 2023;55(2):76-88.
- [36] Omuse G, Maina D, Mwangi J, et al. Complete blood count reference intervals from a healthy adult urban population in Kenya. *PLoS One*. 2018;13(6):e0198444.
- [37] Singh M. Medical benchmarks and the myth of the universal patient. *The New Yorker* [Internet]. 2025 Mar 24 [cited 2025 Aug 22]. Available from:
<https://www.newyorker.com/magazine/2025/03/31/medical-benchmarks-and-the-myth-of-the-universal-patient>
- [38] Zenil H, Kiani NA, Marabita F, Deng Y, Elias S, Schmidt A, Ball G & Tegnér J. *An Algorithmic Information Calculus for Causal Discovery and Reprogramming Systems*. iScience. 2019;19:1160–1172.
- [39] Uthamacumaran A & Zenil H. *A Review of Mathematical and Computational Methods in Cancer Dynamics*. *Front Oncol*. 2022;12:850731.
- [40] Zenil H, Kiani NA & Tegnér J. *Algorithmic Information Dynamics: A Computational Approach to Causality with Applications to Living Systems*. Cambridge University Press; 2023.
- [41] Abrahão FS & Zenil H. *Emergence and algorithmic information dynamics of systems and observers*. *Phil Trans R Soc A*. 2022;380:20200429.

SUPPLEMENTARY INFORMATION

Tables

Rank	Type	Selected Variables	Linkage	Metric	Two-Level Cohesion Score
1	Multi-D	Top-5 by MI	Average	Mutual Information	0.357
2	1D	Male_MCH_Midpoint	Average	Euclidean	0.829
3	Multi-D	All Male Midpoint Variables	Average	Cosine	0.873
4	1D	Male_WBC_Midpoint	Average	Euclidean	0.937
5	1D	Male_MCV_Midpoint	Average	Euclidean	0.971
6	Multi-D	All Male Midpoint Variables	Average	Correlation	0.989
7	1D	Male_RBC_Midpoint	Average	Euclidean	1.120
8	1D	Male_HGB_Midpoint	Average	Euclidean	1.141
9	1D	Male_MCHC_Midpoint	Average	Euclidean	1.147
10	1D	Male_Platelets_Midpoint	Average	Euclidean	1.157
11	1D	Male_MCH_Midpoint	Ward	Euclidean	2.041
12	1D	Male_MCV_Midpoint	Ward	Euclidean	2.691
13	1D	Male_WBC_Midpoint	Ward	Euclidean	2.983
14	1D	Male_RBC_Midpoint	Ward	Euclidean	3.132
15	Multi-D	All Male Midpoint Variables	Average	Euclidean	3.327
16	1D	Male_HGB_Midpoint	Ward	Euclidean	3.611
17	1D	Male_Platelets_Midpoint	Ward	Euclidean	3.655
18	1D	Male_MCHC_Midpoint	Ward	Euclidean	4.044
19	Multi-D	All Male Midpoint Variables	Ward	Euclidean	5.894
20	Multi-D	All Male Midpoint Variables	Average	Cityblock	7.065

Table 2. Two-Level Cohesion Score of Male CBC

Rank	Type	Selected Variables	Linkage	Metric	Two-Level Cohesion Score
1	Multi-D	Top-5 by MI	Average	Mutual Information	0.394
2	1D	Female_MCH_Midpoint	Average	Euclidean	0.881
3	1D	Female_MCV_Midpoint	Average	Euclidean	0.894
4	1D	Female_WBC_Midpoint	Average	Euclidean	0.909
5	Multi-D	All Female Midpoint Variables	Average	Cosine	0.939
6	Multi-D	All Female Midpoint Variables	Average	Correlation	1.083
7	1D	Female_HGB_Midpoint	Average	Euclidean	1.201
8	1D	Female_HT_Midpoint	Average	Euclidean	1.214
9	1D	Female_Platelets_Midpoint	Average	Euclidean	1.241
10	1D	Female_MCHC_Midpoint	Average	Euclidean	1.241
11	1D	Female_MCV_Midpoint	Ward	Euclidean	2.193
12	1D	Female_MCH_Midpoint	Ward	Euclidean	2.416
13	1D	Female_WBC_Midpoint	Ward	Euclidean	2.505
14	Multi-D	All Female Midpoint Variables	Average	Euclidean	3.274
15	1D	Female_HT_Midpoint	Ward	Euclidean	3.523
16	1D	Female_HGB_Midpoint	Ward	Euclidean	3.606
17	1D	Female_MCHC_Midpoint	Ward	Euclidean	4.114
18	1D	Female_Platelets_Midpoint	Ward	Euclidean	4.387
19	Multi-D	All Female Midpoint Variables	Ward	Euclidean	6.651
20	Multi-D	All Female Midpoint Variables	Average	Cityblock	7.207

Table 3. Two-Level Cohesion Score of Female CBC

Rank	Type	Selected Variables	Linkage	Metric	Two-Level Cohesion Score
1	Multi-D	ALL Male BMI Variables	Average	Cosine	0.551
2	Multi-D	ALL Male BMI Variables	Average	Correlation	0.612
3	1D	Male Mean BMI (65–69)	Average	Euclidean	0.645
4	1D	Male Mean BMI (60–64)	Average	Euclidean	0.668
5	1D	Male Mean BMI (75–79)	Average	Euclidean	0.674
6	Multi-D	Top-5 by MI	Average	Mutual Information	0.701
7	1D	Male Mean BMI (85+)	Average	Euclidean	0.877
8	1D	Male Mean BMI (80–84)	Average	Euclidean	0.886
9	1D	Male Mean BMI (70–74)	Average	Euclidean	0.901
10	1D	Male Mean BMI (65–69)	Ward	Euclidean	1.699
11	1D	Male Mean BMI (75–79)	Ward	Euclidean	1.815
12	1D	Male Mean BMI (60–64)	Ward	Euclidean	1.898
13	Multi-D	ALL Male BMI Variables	Average	Euclidean	2.142
14	1D	Male Mean BMI (85+)	Ward	Euclidean	2.453
15	1D	Male Mean BMI (70–74)	Ward	Euclidean	2.530
16	1D	Male Mean BMI (80–84)	Ward	Euclidean	2.577
17	Multi-D	ALL Male BMI Variables	Ward	Euclidean	4.704
18	Multi-D	ALL Male BMI Variables	Average	Cityblock	5.120

Table 4. Two-Level Cohesion Score of Male BMI

Rank	Type	Selected Variables	Linkage	Metric	Two-Level Cohesion Score
1	Multi-D	Top-5 by MI	Average	Mutual Information	0.581
2	Multi-D	ALL Female BMI Variables	Average	Cosine	0.672
3	1D	Female Mean BMI (80–84)	Average	Euclidean	0.744
4	1D	Female Mean BMI (75–79)	Average	Euclidean	0.747
5	1D	Female Mean BMI (85+)	Average	Euclidean	0.768
6	1D	Female Mean BMI (70–74)	Average	Euclidean	0.804
7	1D	Female Mean BMI (65–69)	Average	Euclidean	0.818
8	1D	Female Mean BMI (60–64)	Average	Euclidean	0.880
9	Multi-D	ALL Female BMI Variables	Average	Correlation	1.089
10	1D	Female Mean BMI (85+)	Ward	Euclidean	1.984
11	Multi-D	ALL Female BMI Variables	Average	Euclidean	2.097
12	1D	Female Mean BMI (80–84)	Ward	Euclidean	2.144
13	1D	Female Mean BMI (75–79)	Ward	Euclidean	2.258
14	1D	Female Mean BMI (65–69)	Ward	Euclidean	2.388
15	1D	Female Mean BMI (70–74)	Ward	Euclidean	2.409
16	1D	Female Mean BMI (60–64)	Ward	Euclidean	2.916
17	Multi-D	ALL Female BMI Variables	Average	Cityblock	4.883
18	Multi-D	ALL Female BMI Variables	Ward	Euclidean	5.167

Table 5. Two-Level Cohesion Score of Female CBC

Data Sources for CBC Reference Intervals

	Country	Source (Title/Organization)	Link
North America			
1	US	Walker HK, Hall WD, & Hurst JW. <i>Clinical Methods</i> (1990).	NCBI Bookshelf
2	Canada	MCC. Normal Lab Values.	MCC
3	Mexico	<i>Patología. Revista Latinoamericana</i> (2012).	PDF
South America			
4	Brazil	PNCQ. Reference hematological values (2020).	PDF
5	Venezuela	Fernández LE et al. <i>Revista de la Facultad de Medicina (Caracas)</i> (2006).	SciELO
6	Chile	Gutiérrez LA et al. <i>Advances in Laboratory Medicine</i> (2025).	PMC
Asia			
7	Japan	University of Tokyo Hospital. Reference intervals for adult CBC (2024).	PDF
8	China	Peking Union Medical College Hospital. Laboratory reference intervals (2013).	PDF
9	Philippines	Philippine General Hospital. Normal Values (PGH).	Scribd
10	Singapore	NUH. Full Blood Count (FBC) Adult Reference Intervals (2022).	NUH
Europe			
11	France	Troussard X et al. <i>Journal of Clinical Pathology</i> (2014).	PubMed
12	Italy	Azienda Ospedaliera di Niguarda. CBC reference values.	Hospital
13	Spain	Hospital Clínic de Barcelona. Hemograma reference intervals.	Hospital
14	Greece	University of Patras. General blood count (2021).	PDF
15	Sweden	Karolinska University Hospital. Blodstatus (2024).	Karolinska
16	Germany	Klinikum Braunschweig. Hämatologische Referenzbereiche.	Hospital

	Country	Source (Title/Organization)	Link
17	UK	Medway NHS Foundation Trust. Haematology reference ranges (2020).	PDF
Africa			
18	Egypt	Radwan EH et al. <i>Journal of Medical and Life Science</i> (2022).	PDF
19	Kenya	Omuse G et al. <i>PLoS ONE</i> (2018).	PLoS ONE
20	Morocco	Bakrim S et al. <i>Clinical Laboratory</i> (2023).	PubMed
21	South Africa	De Koker A et al. <i>SAMJ</i> (2021).	PubMed
22	Ghana	Bawua AS et al. <i>Int J Lab Hematology</i> (2020).	PMC
Middle East			
23	UAE	FML Dubai. CBC – Complete Blood Count with differential.	FML Dubai
24	Oman	Al-Mawali A et al. <i>PLoS ONE</i> (2018).	PLoS ONE
25	Saudi Arabia	Shaheen NA et al. <i>Frontiers in Medicine</i> (2022).	PMC
Eurasia			
26	Russia	Ruzhanskaya A et al. <i>PLOS ONE</i> (2024).	PMC
Oceania			
27	Australia	NSW Health Pathology. Adult Reference Ranges (2019).	PDF
28	New Zealand	Auckland DHB, LabPLUS. Full Blood Count reference intervals.	LabPLUS

1.1.1 Platelets

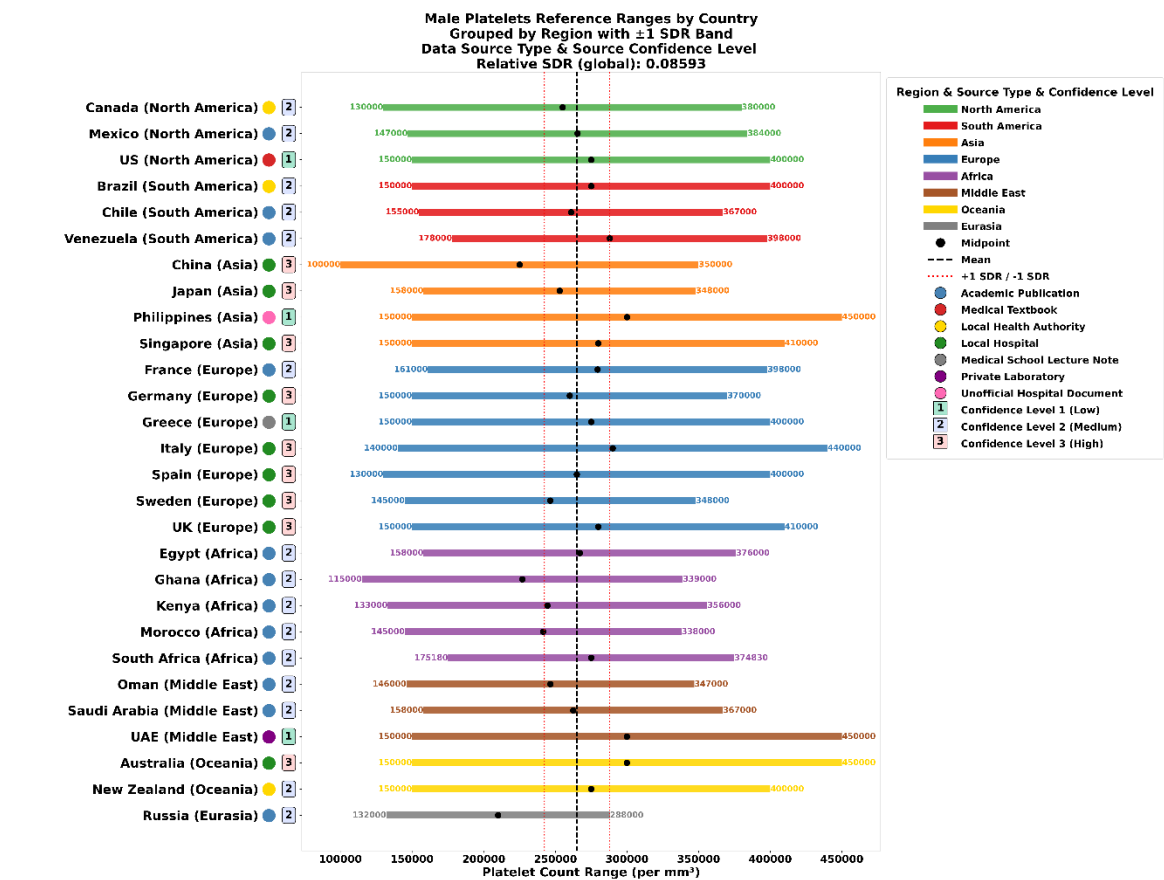


Figure 1.1.1-1. Male Platelets Reference Ranges by Country and Region, with Data Source Type and Confidence Level

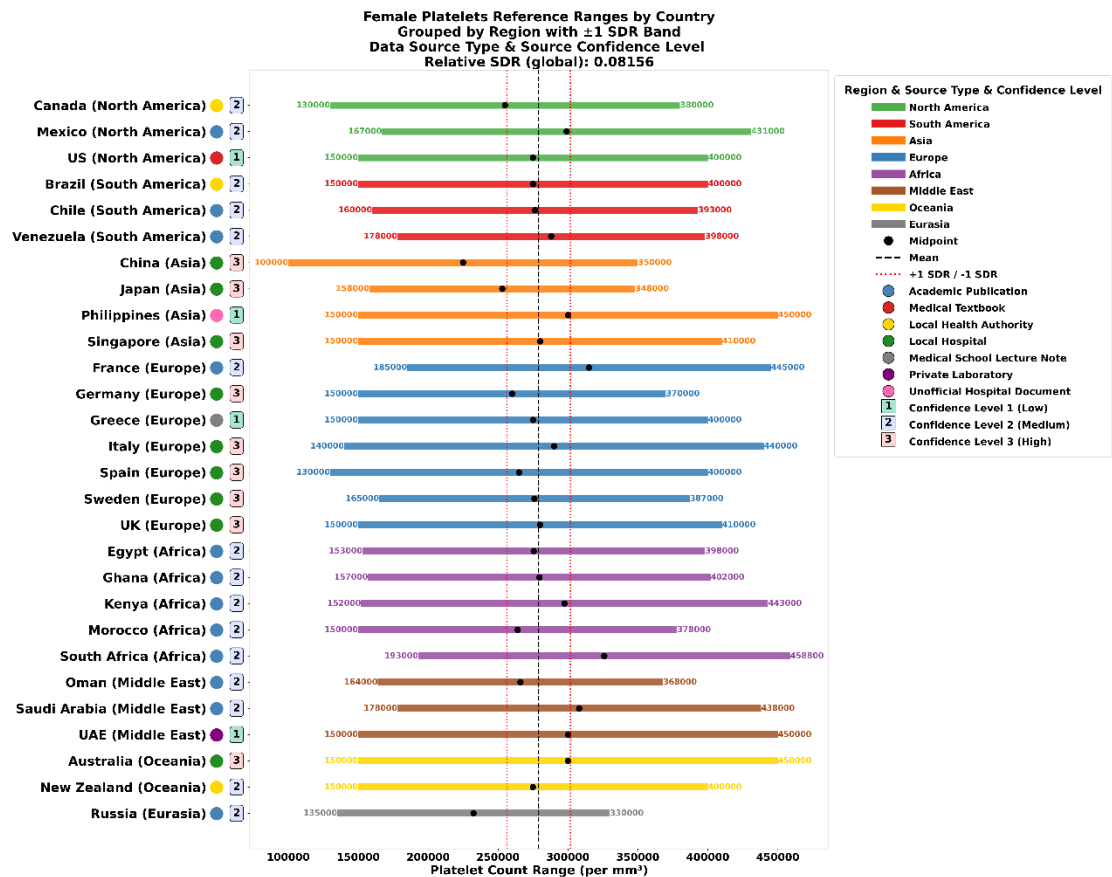


Figure 1.1.1-2. Female Platelets Reference Ranges by Country and Region, with Data Source Type and Confidence Level

1.1.2 MCHC (Mean Corpuscular Hemoglobin Concentration)

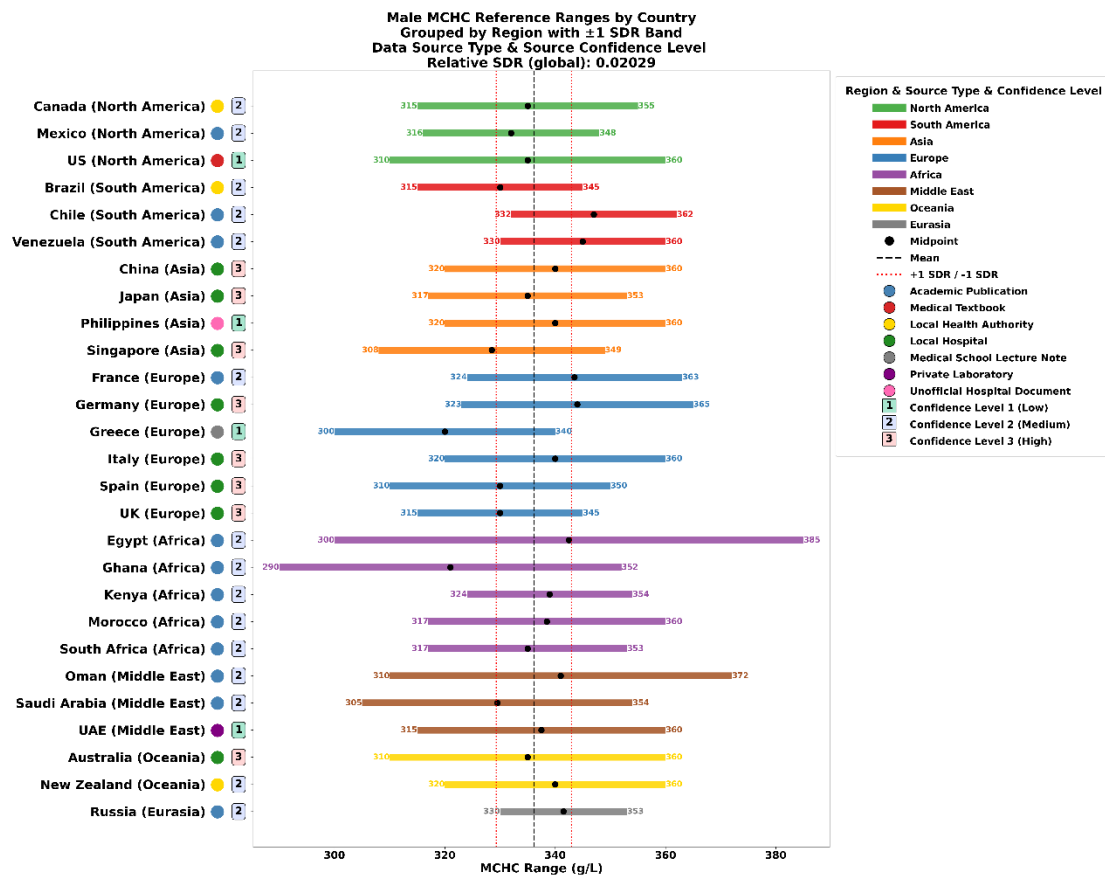


Figure 1.1.2-1. Male MCHC Reference Ranges by Country and Region, with Data Source Type and Confidence Level

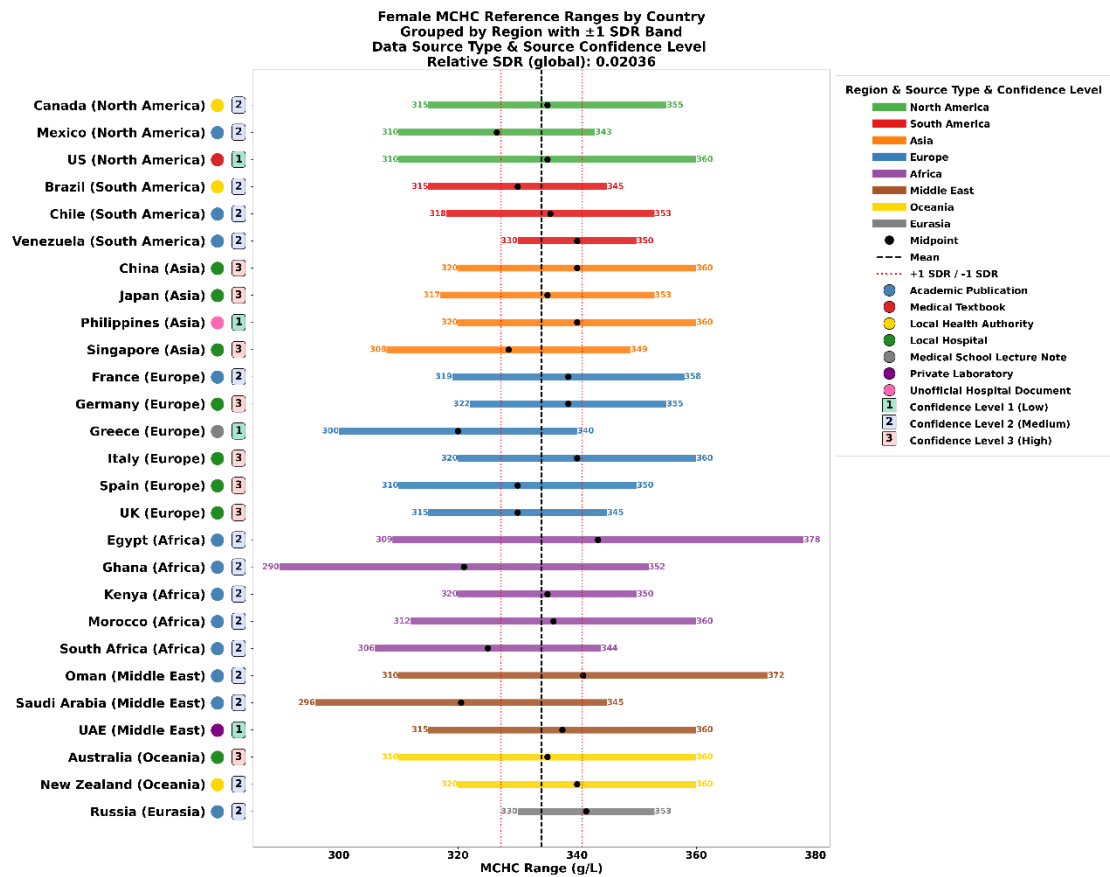


Figure 1.1.2-2. Female MCHC Reference Ranges by Country and Region, with Data Source Type and Confidence Level

1.1.3 MCH (Mean Corpuscular Hemoglobin)

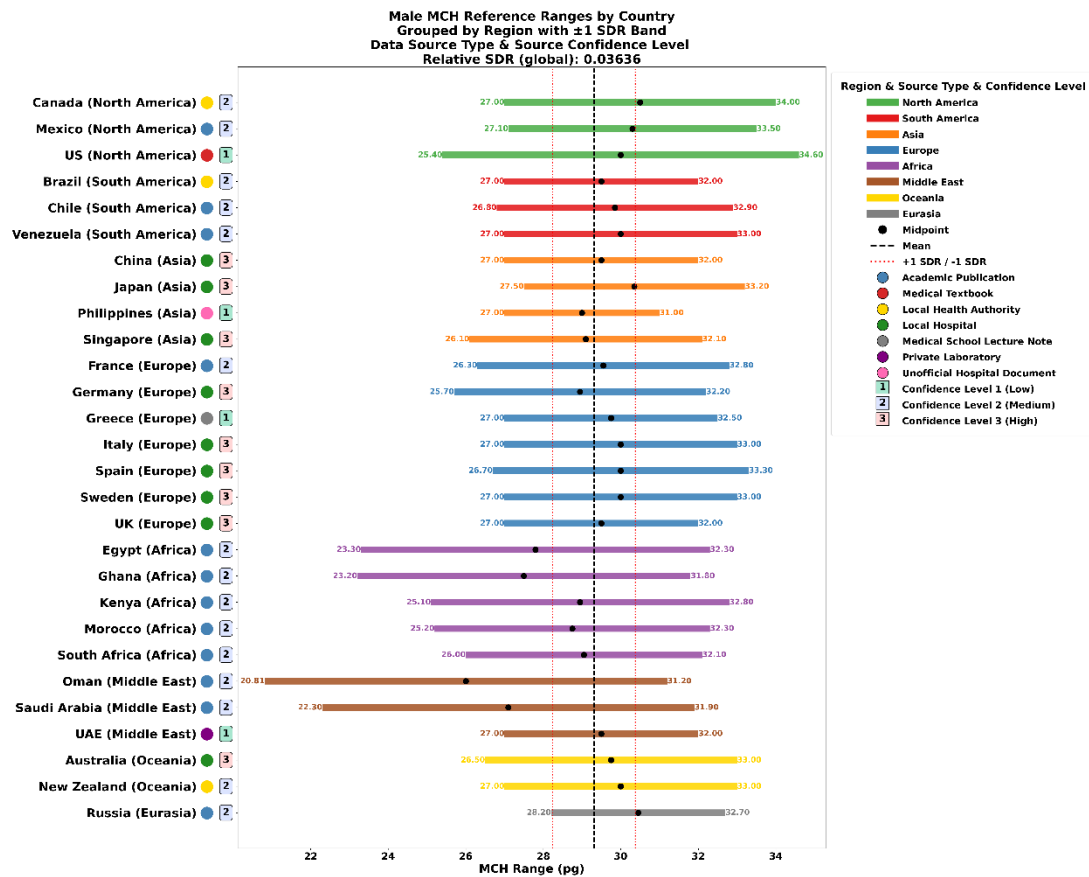


Figure 1.1.3-1. Male MCH Reference Ranges by Country and Region, with Data Source Type and Confidence Level

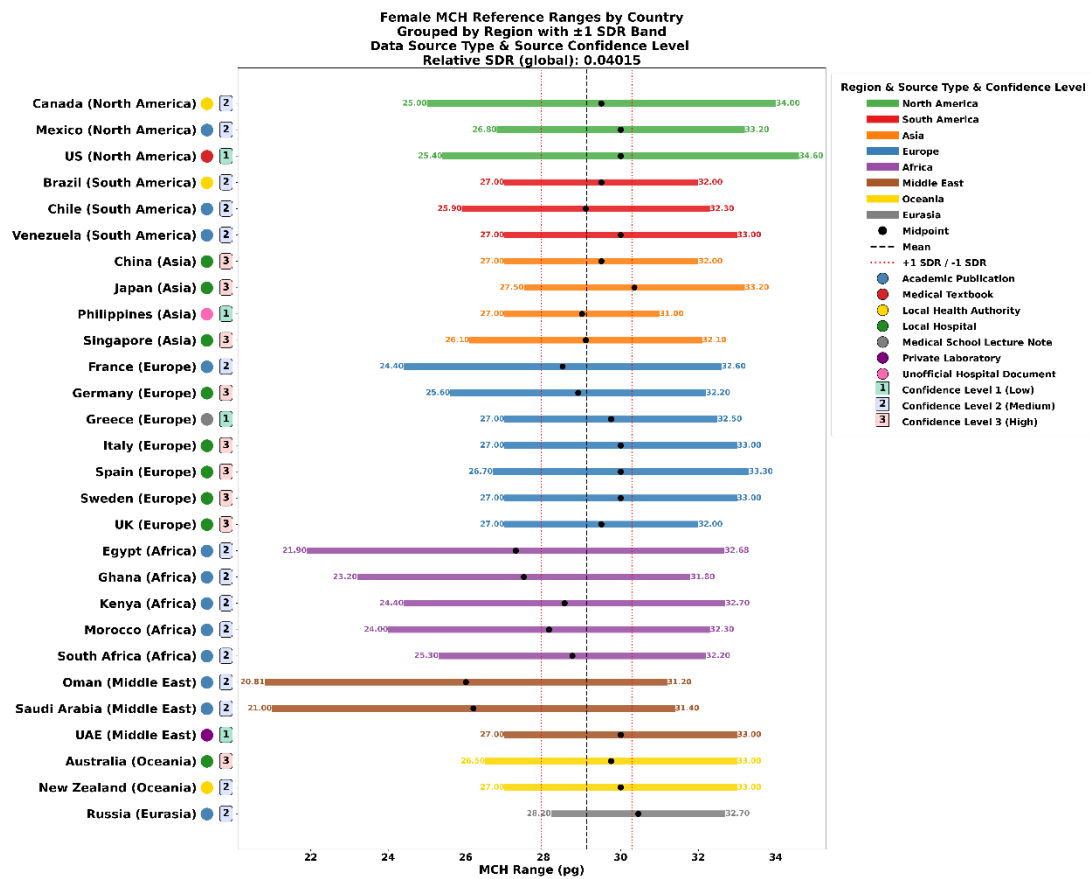


Figure 1.1.3-2. Female MCH Reference Ranges by Country and Region, with Data Source Type and Confidence Level

Confidence Level

1.1.1 RBC (Red Blood Cells)

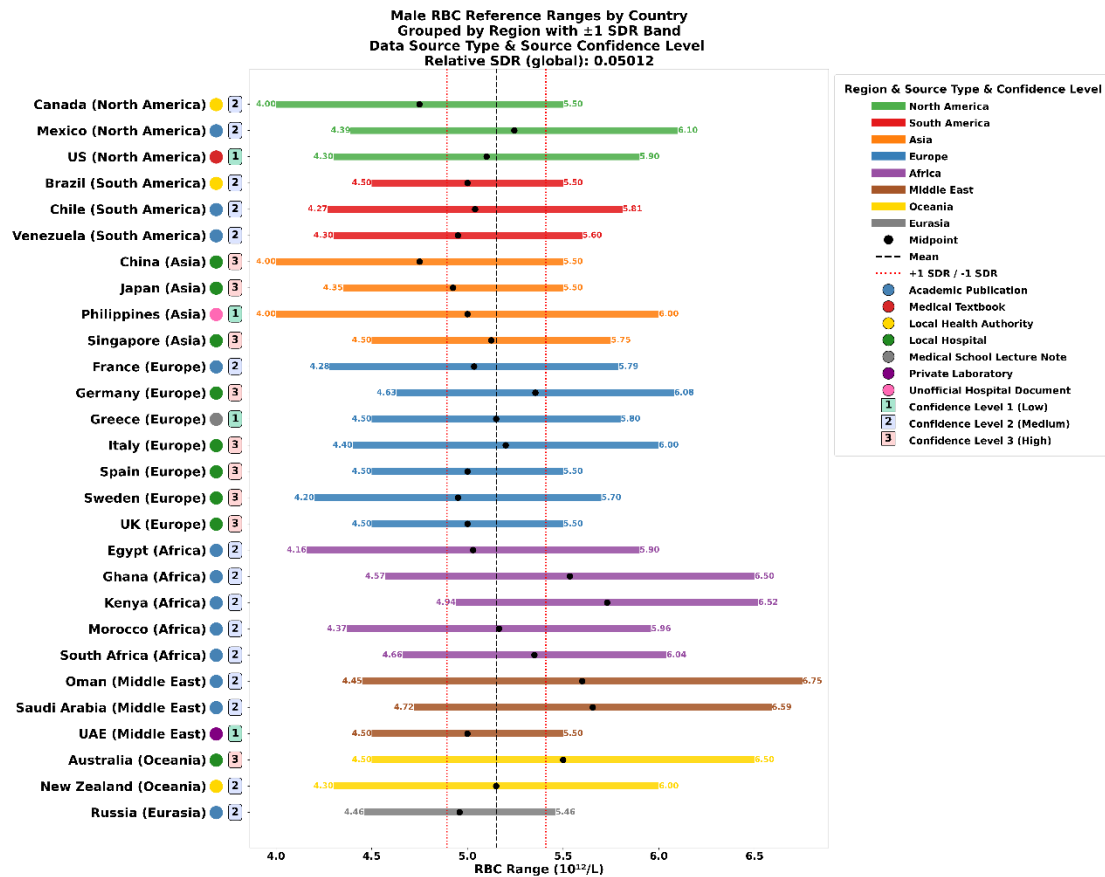


Figure 1.1.1-1. Male RBC Reference Ranges by Country and Region, with Data Source Type and Confidence Level

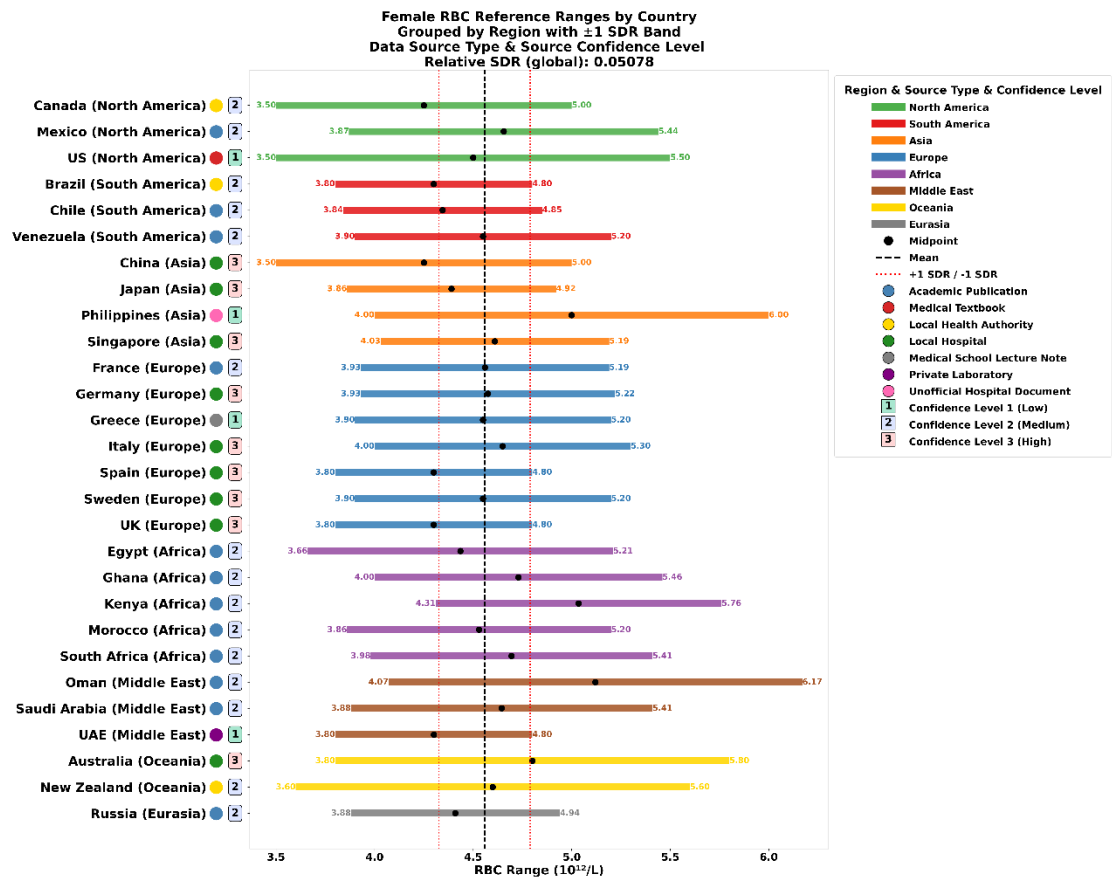


Figure 1.1.1-2. Female RBC Reference Ranges by Country and Region, with Data Source Type and Confidence Level

1.1.2 HGB (Hemoglobin)

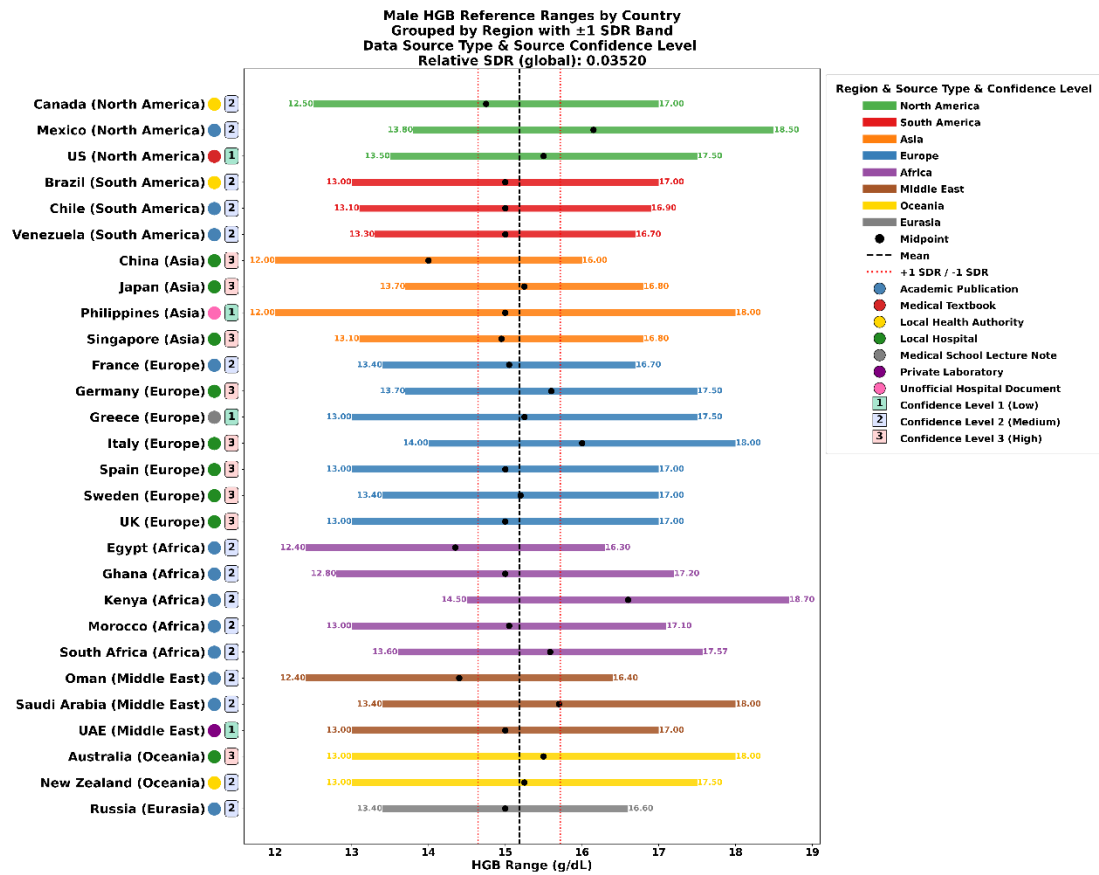


Figure 1.1.2-1. Male HGB Reference Ranges by Country and Region, with Data Source Type and Confidence Level

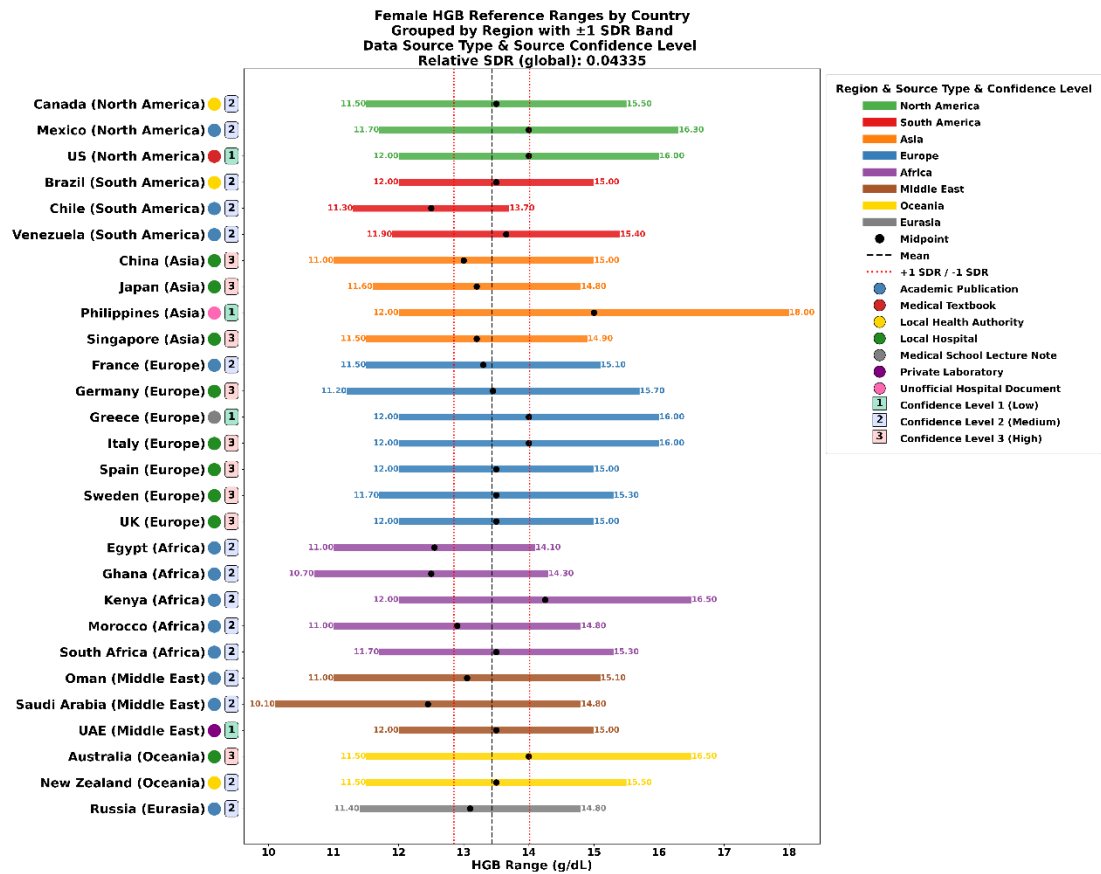


Figure 1.1.2-2. Female HGB Reference Ranges by Country and Region, with Data Source Type and Confidence Level

1.1.3 HT (Hematocrit)

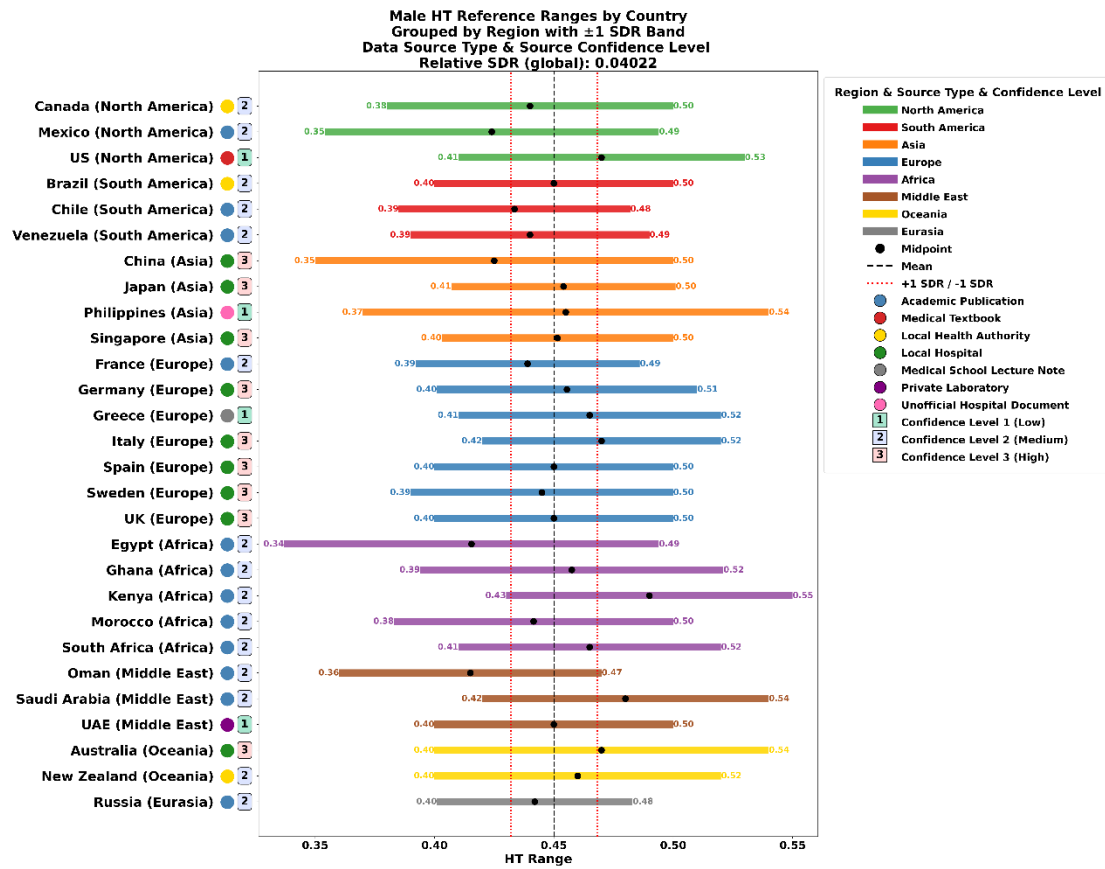


Figure 1.1.3-1. Male HT Reference Ranges by Country and Region, with Data Source Type and Confidence Level

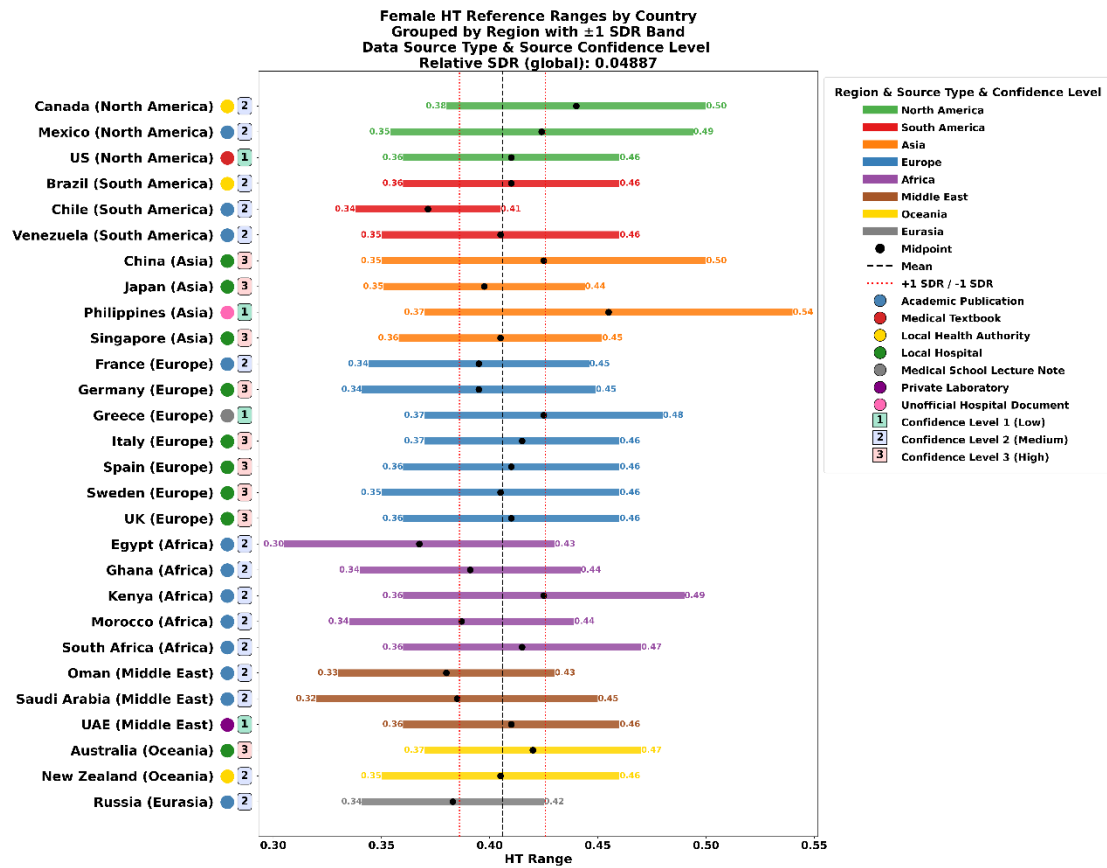


Figure 1.1.3-2. Female HT Reference Ranges by Country and Region, with Data Source Type and Confidence Level

One-Dimensional Dendrograms by Sex for Individual CBC Analytes

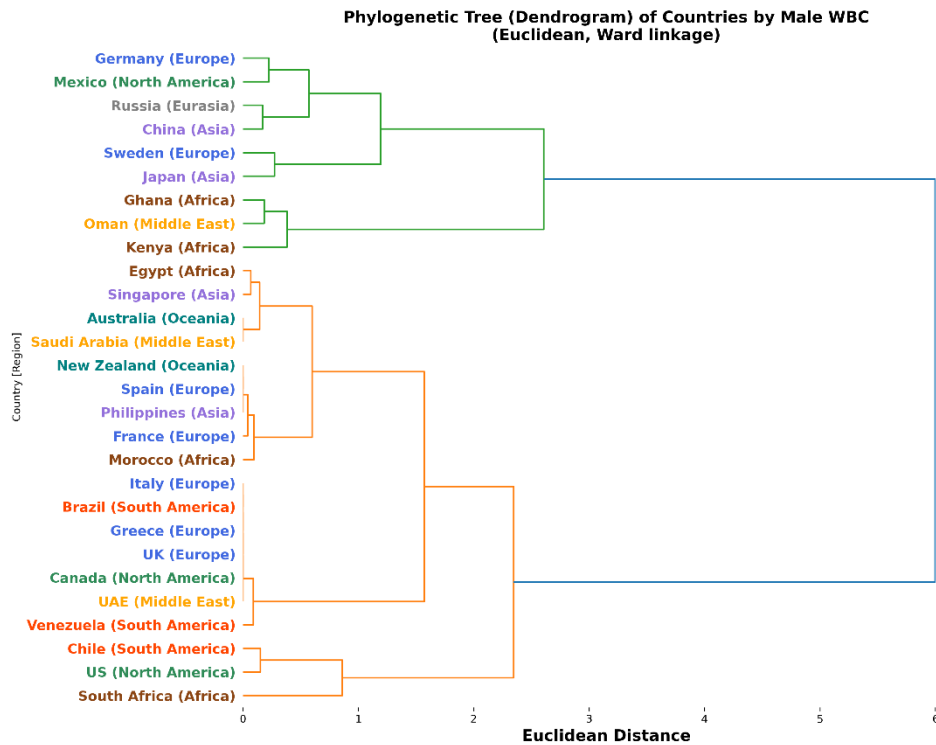


Figure 1.1.3-3. Phylogenetic Tree of Countries by Male WBC (Ward Linkage + Euclidean Distance)

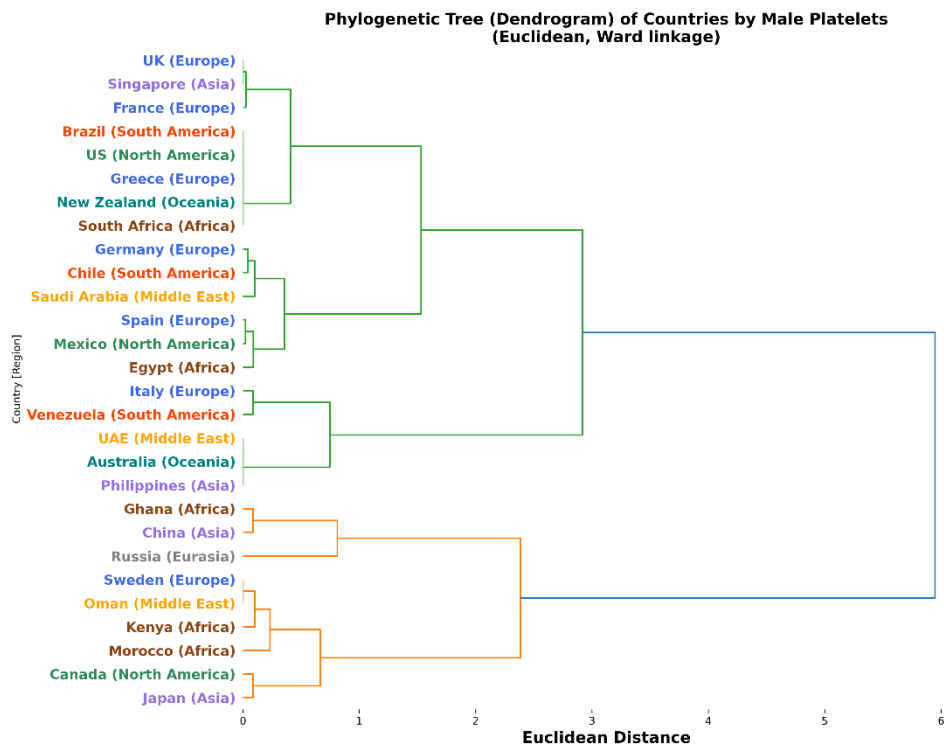


Figure 1.1.3-4. Phylogenetic Tree of Countries by Male Platelets (Ward Linkage + Euclidean Distance)

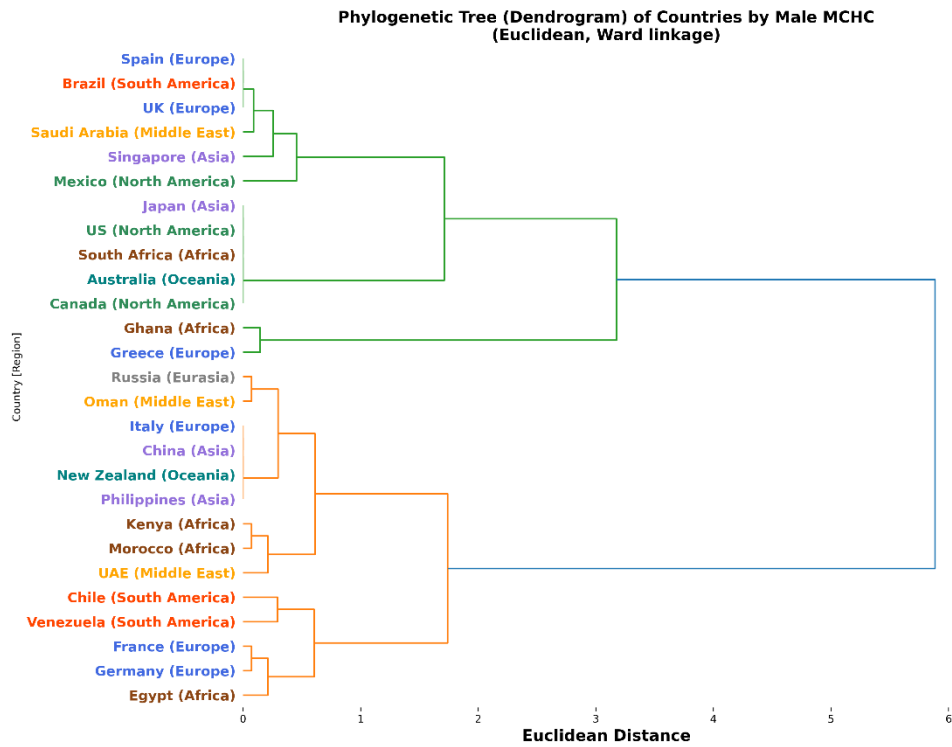


Figure 1.1.3-5. Phylogenetic Tree of Countries by Male MCHC (Ward Linkage + Euclidean Distance)

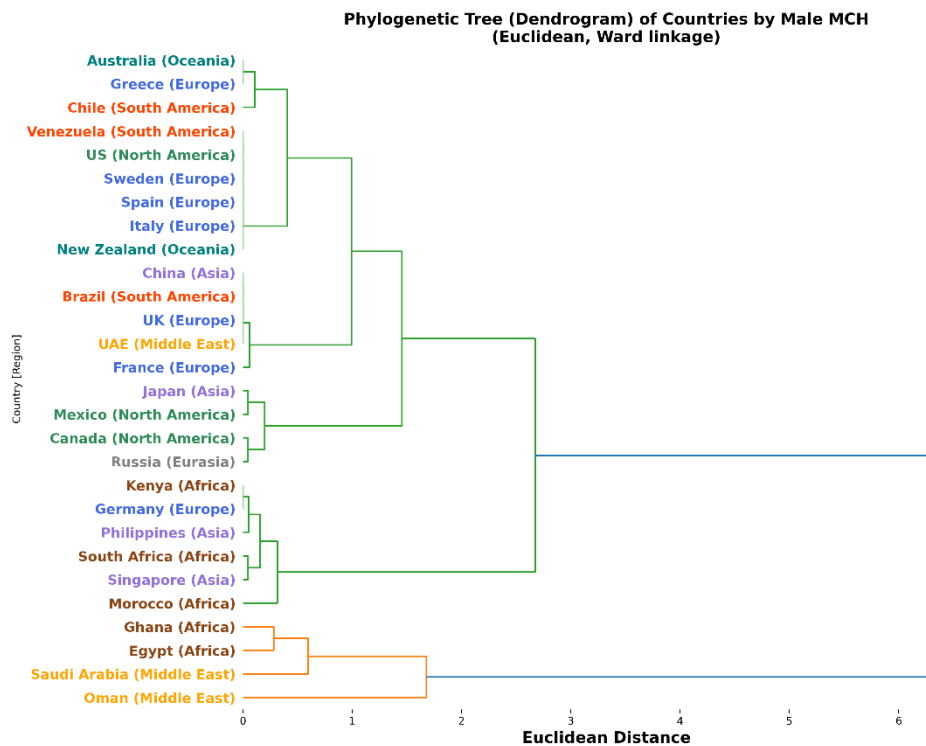


Figure 1.1.3-6. Phylogenetic Tree of Countries by Male MCH (Ward Linkage + Euclidean Distance)

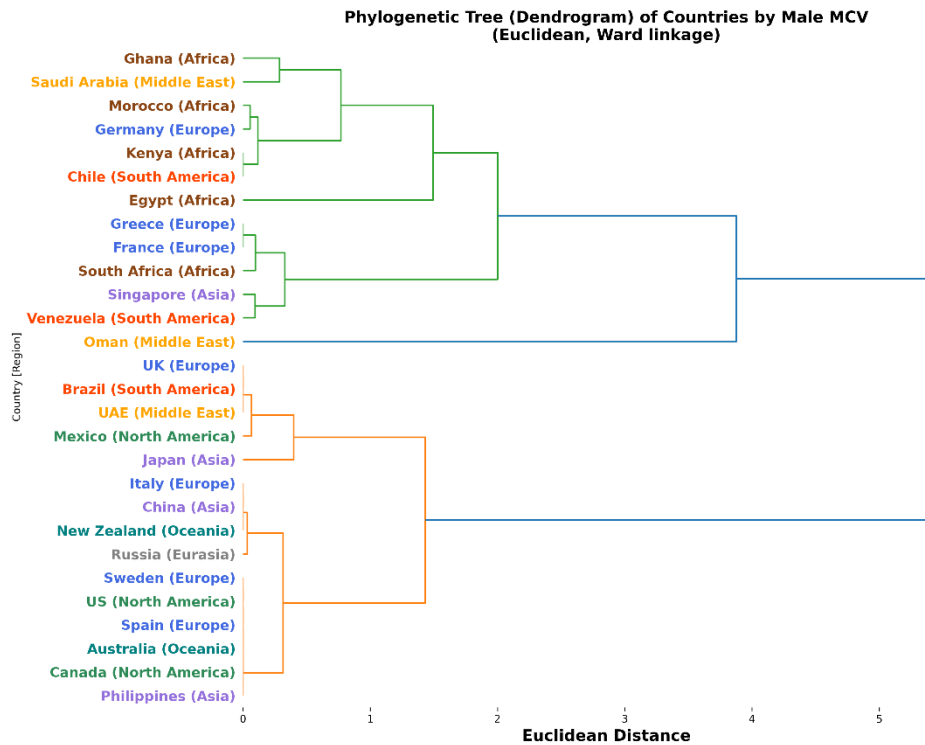


Figure 1.1.3-7. Phylogenetic Tree of Countries by Male MCV (Ward Linkage + Euclidean Distance)

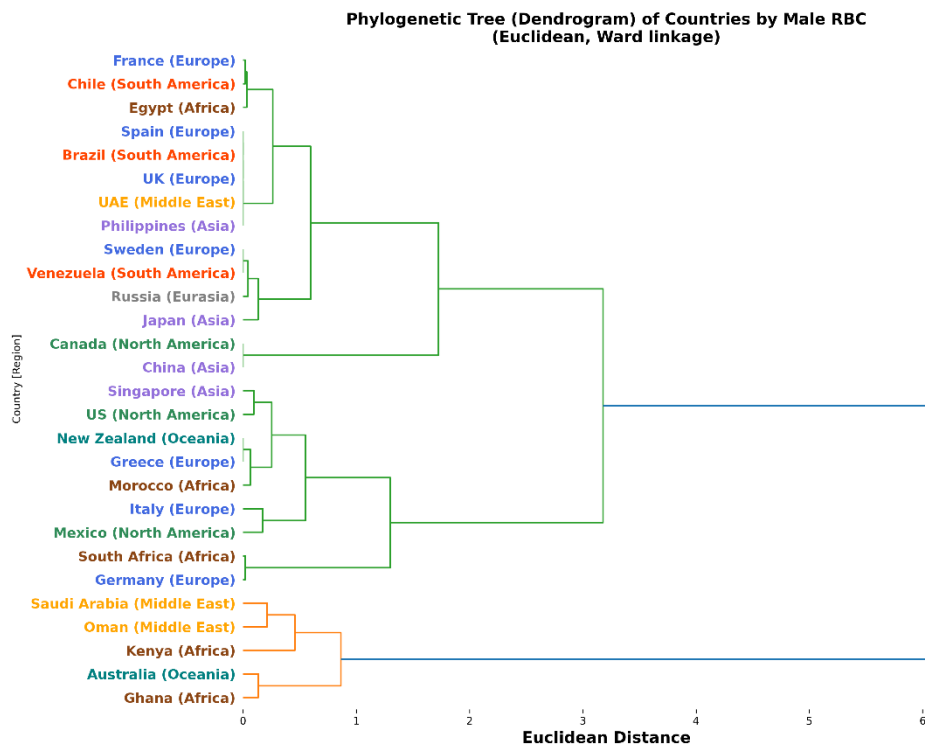


Figure 1.1.3-8. Phylogenetic Tree of Countries by Male RBC (Ward Linkage + Euclidean Distance)

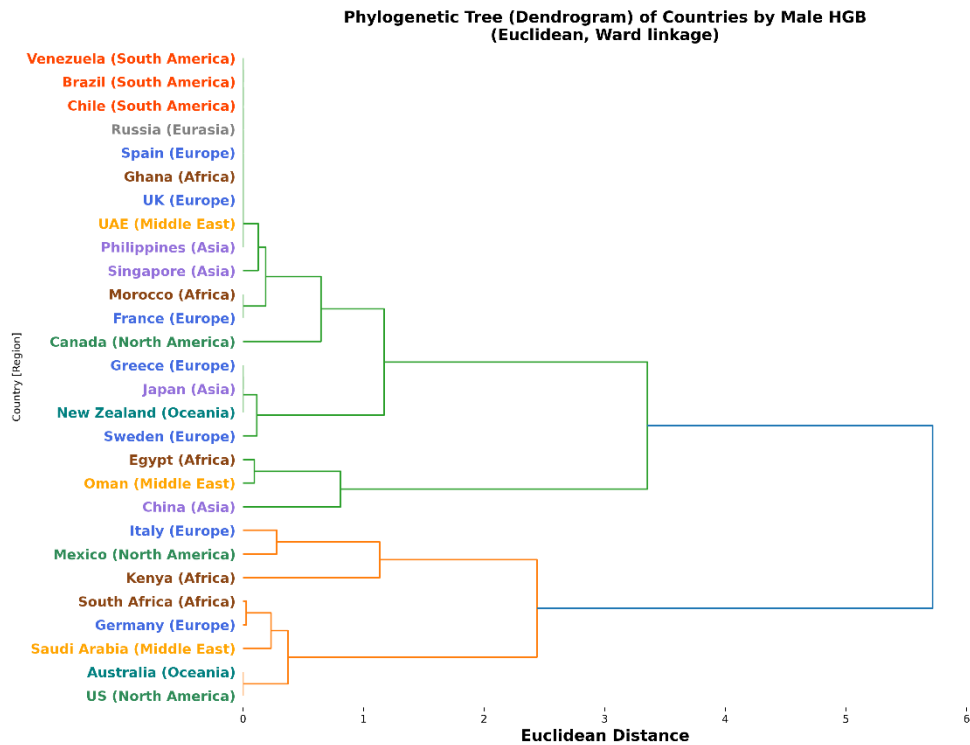


Figure 1.1.3-9. Phylogenetic Tree of Countries by Male HGB (Ward Linkage + Euclidean Distance)

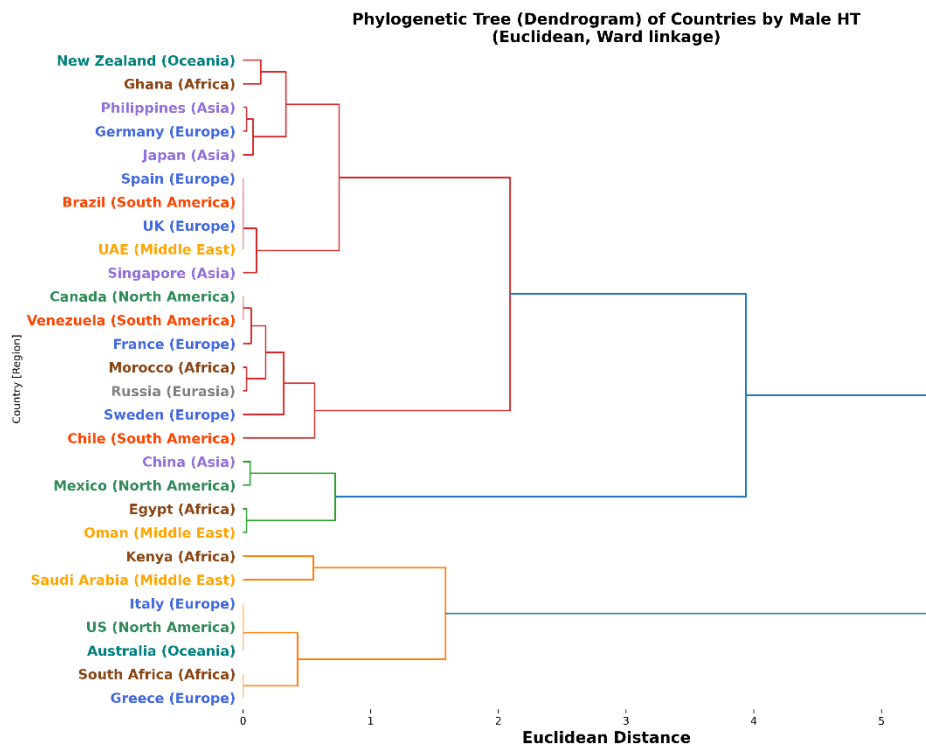


Figure 1.1.3-10. Phylogenetic Tree of Countries by Male HT (Ward Linkage + Euclidean Distance)

1.1.3.1.1 Female

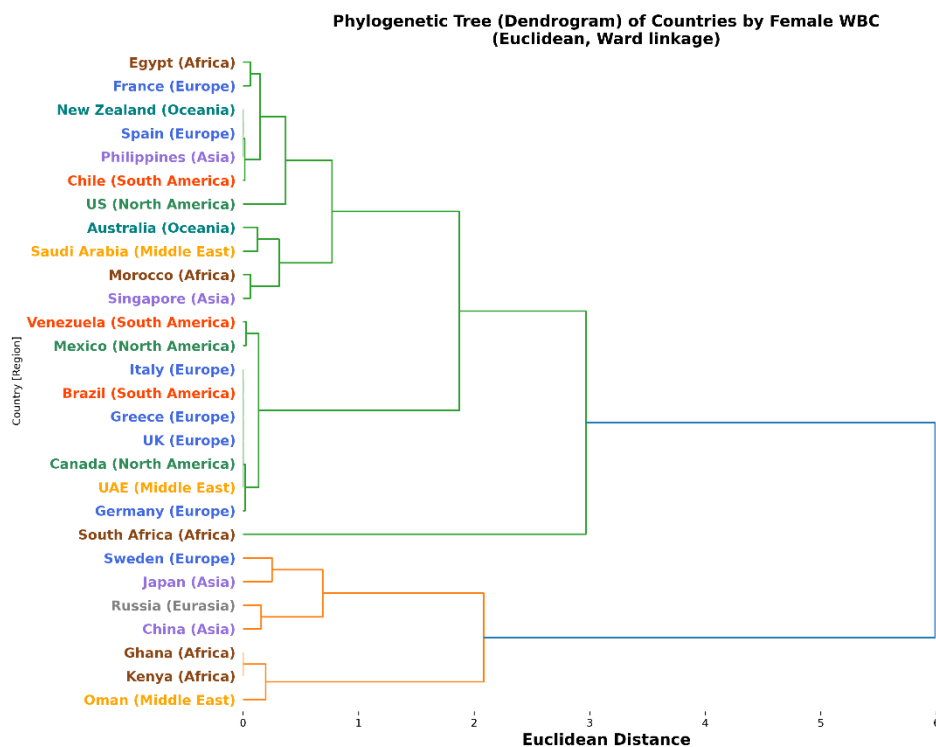


Figure 1.1.3-11. Phylogenetic Tree of Countries by Female WBC (Ward Linkage + Euclidean Distance)

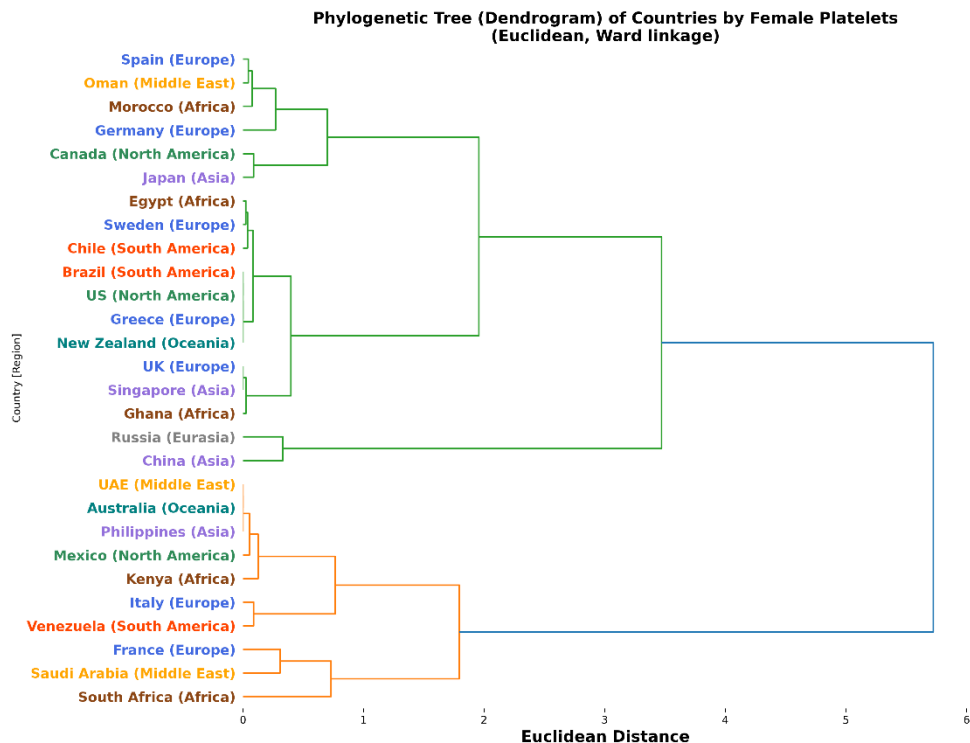


Figure 1.1.3-12. Phylogenetic Tree of Countries by Female Platelets (Ward Linkage + Euclidean Distance)

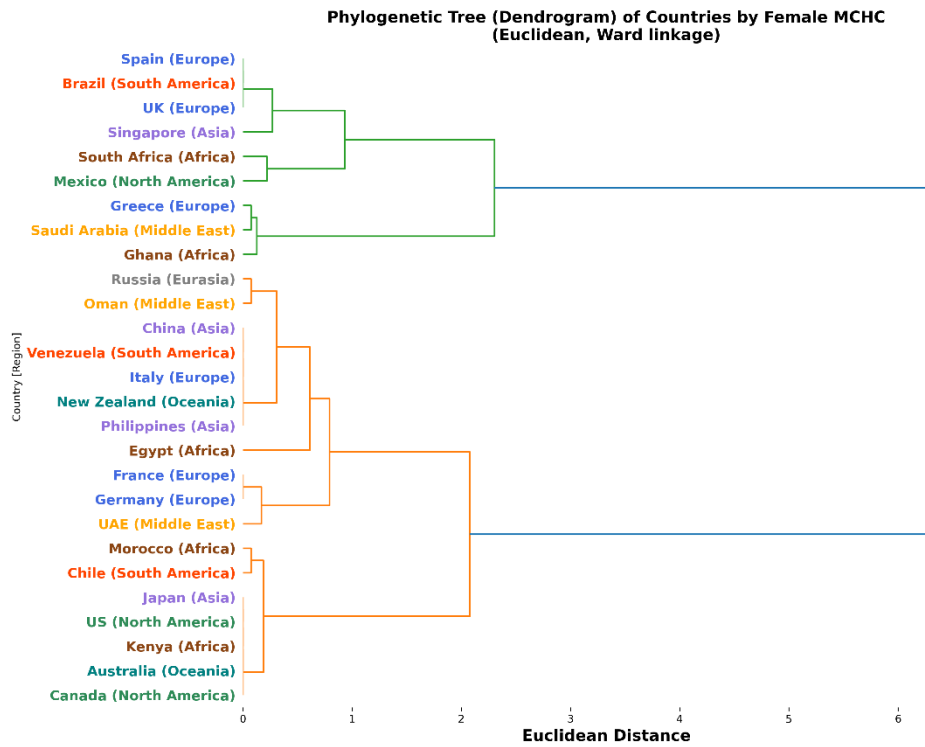


Figure 1.1.3-13. Phylogenetic Tree of Countries by Female MCHC (Ward Linkage + Euclidean Distance)

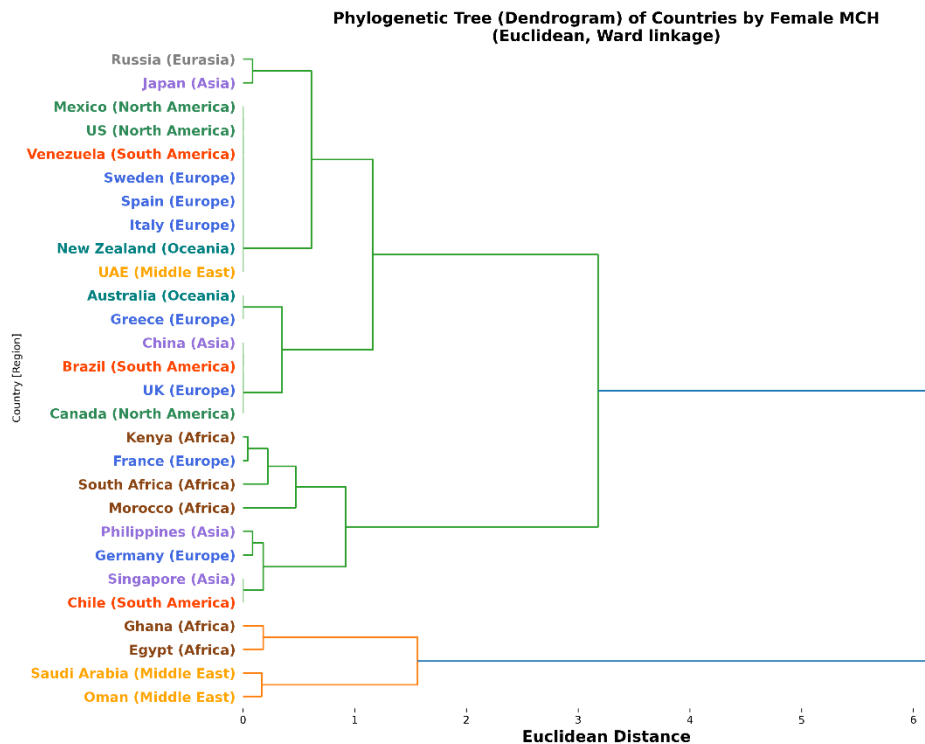


Figure 1.1.3-14. Phylogenetic Tree of Countries by Female MCH (Ward Linkage + Euclidean Distance)

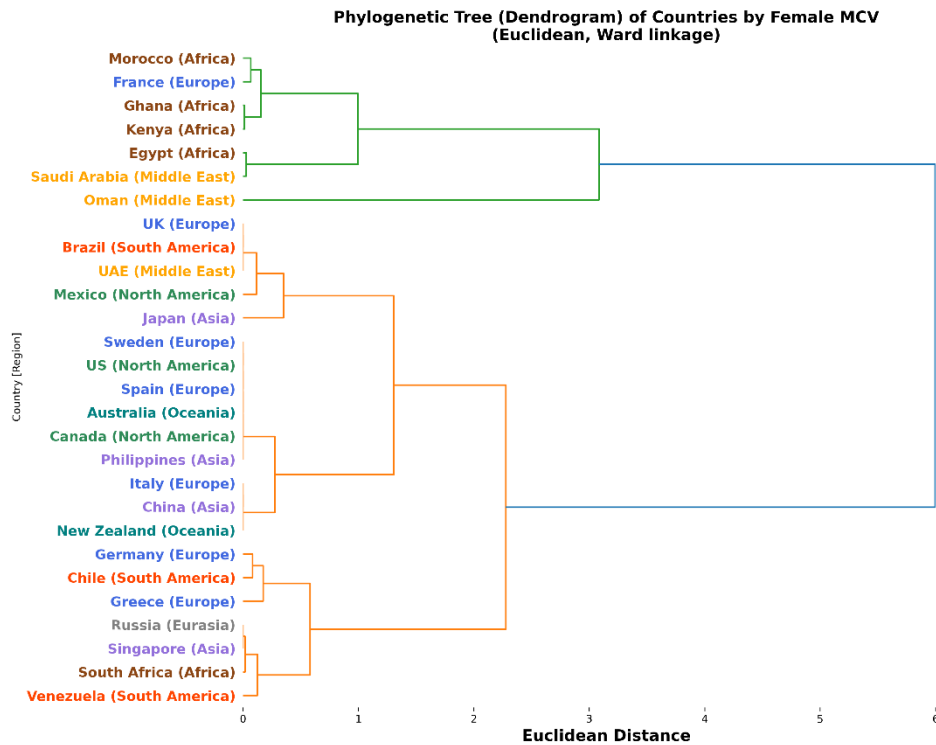


Figure 1.1.3-15. Phylogenetic Tree of Countries by Female MCV (Ward Linkage + Euclidean Distance)

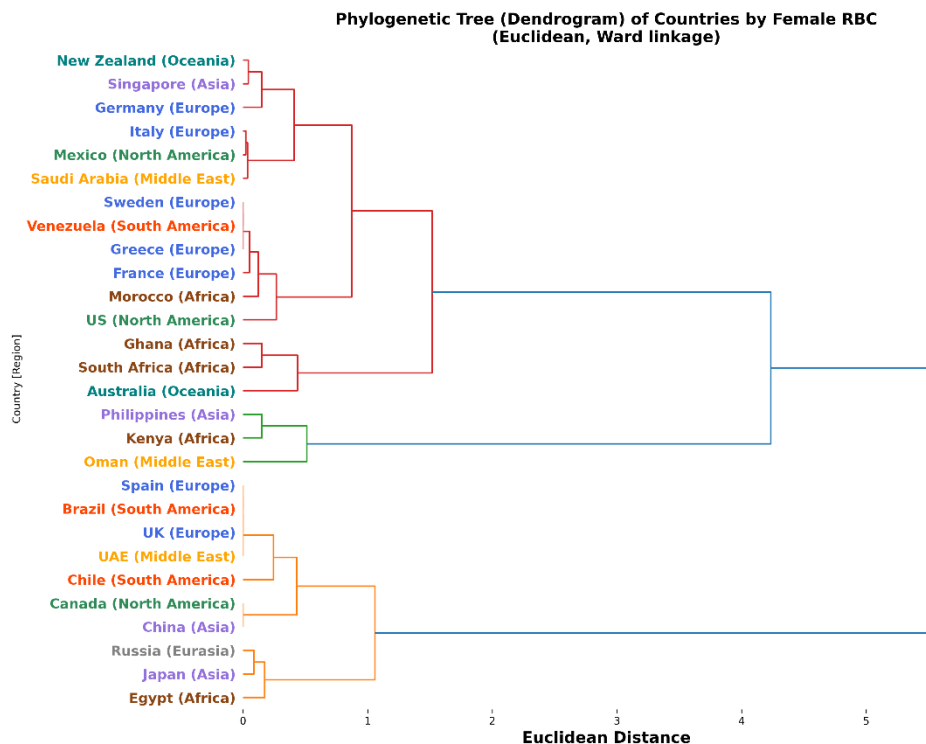


Figure 1.1.3-16. Phylogenetic Tree of Countries by Female RBC (Ward Linkage + Euclidean Distance)

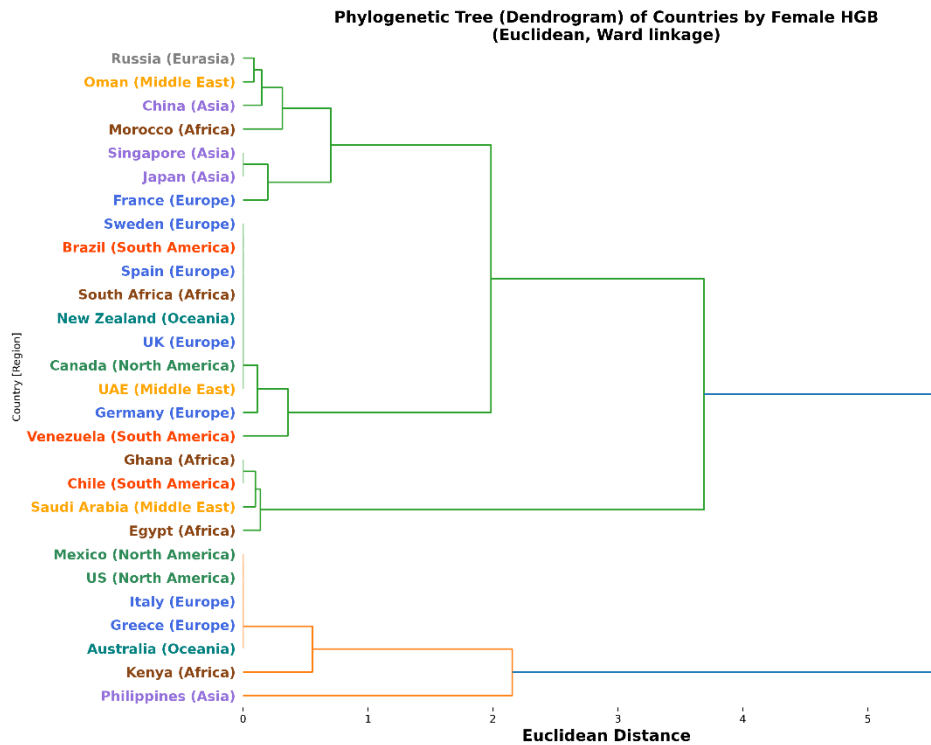


Figure 1.1.3-17. Phylogenetic Tree of Countries by Female HGB (Ward Linkage + Euclidean Distance)

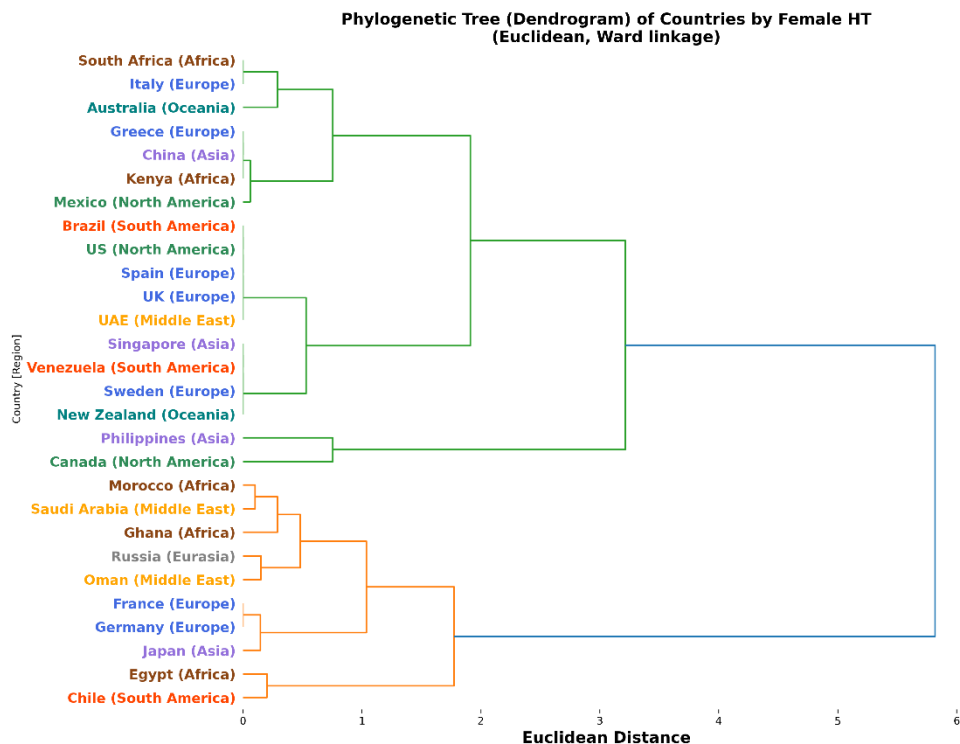


Figure 1.1.3-18. Phylogenetic Tree of Countries by Female HT (Ward Linkage + Euclidean Distance)

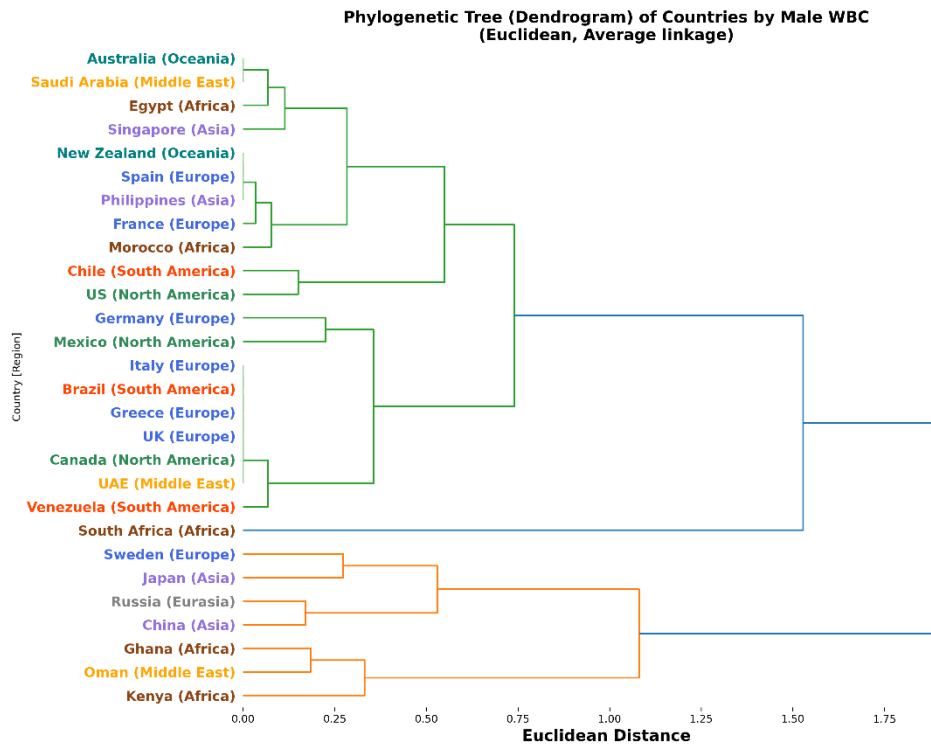


Figure 1.1.3-19. Phylogenetic Tree of Countries by Male WBC (Average Linkage + Euclidean Distance)

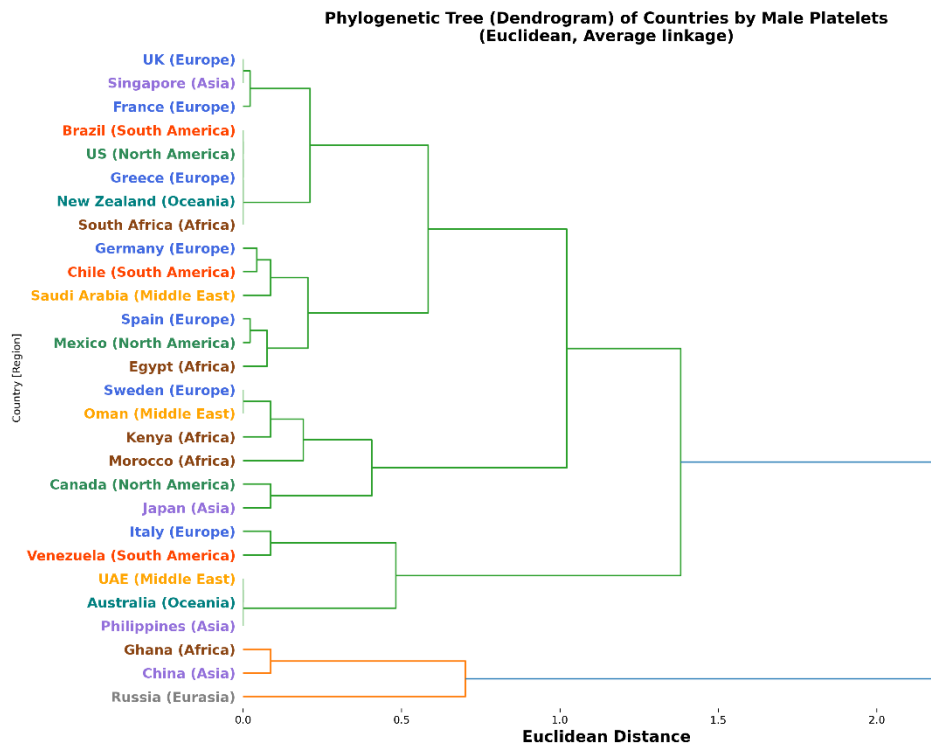


Figure 1.1.3-20. Phylogenetic Tree of Countries by Male Platelets (Average Linkage + Euclidean Distance)

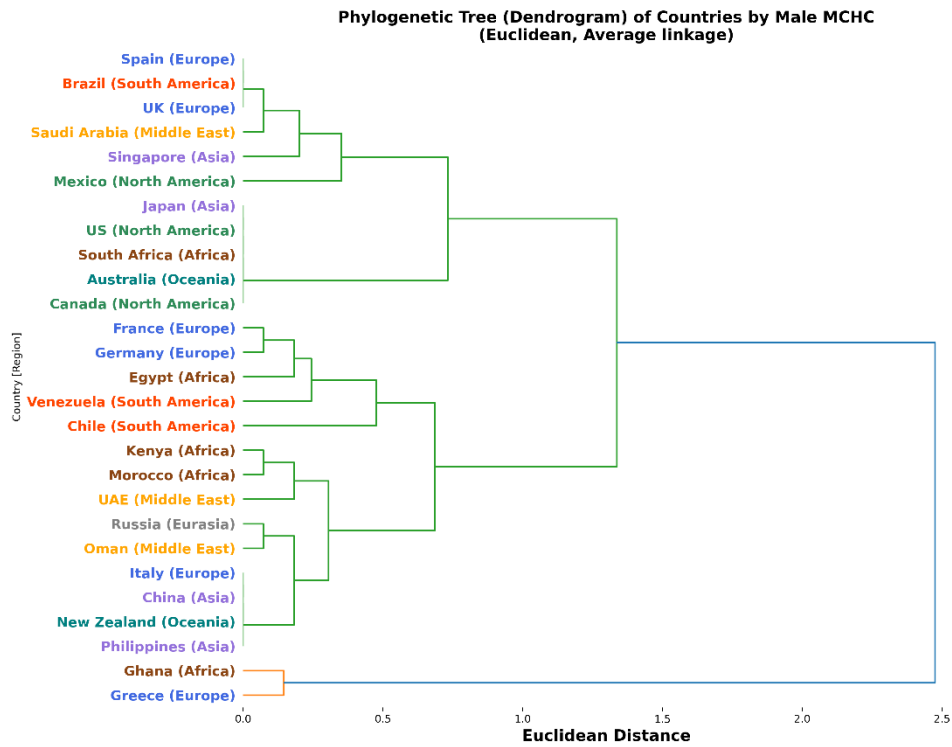


Figure 1.1.3-21. Phylogenetic Tree of Countries by Male MCHC (Average Linkage + Euclidean Distance)

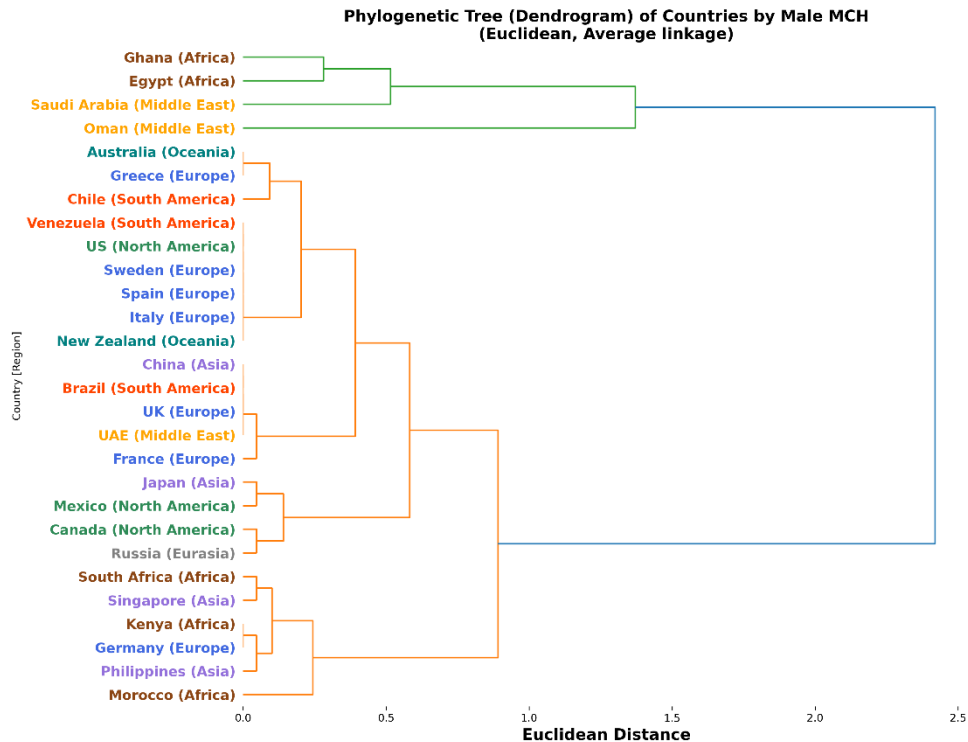


Figure 1.1.3-22. Phylogenetic Tree of Countries by Male MCH (Average Linkage + Euclidean Distance)

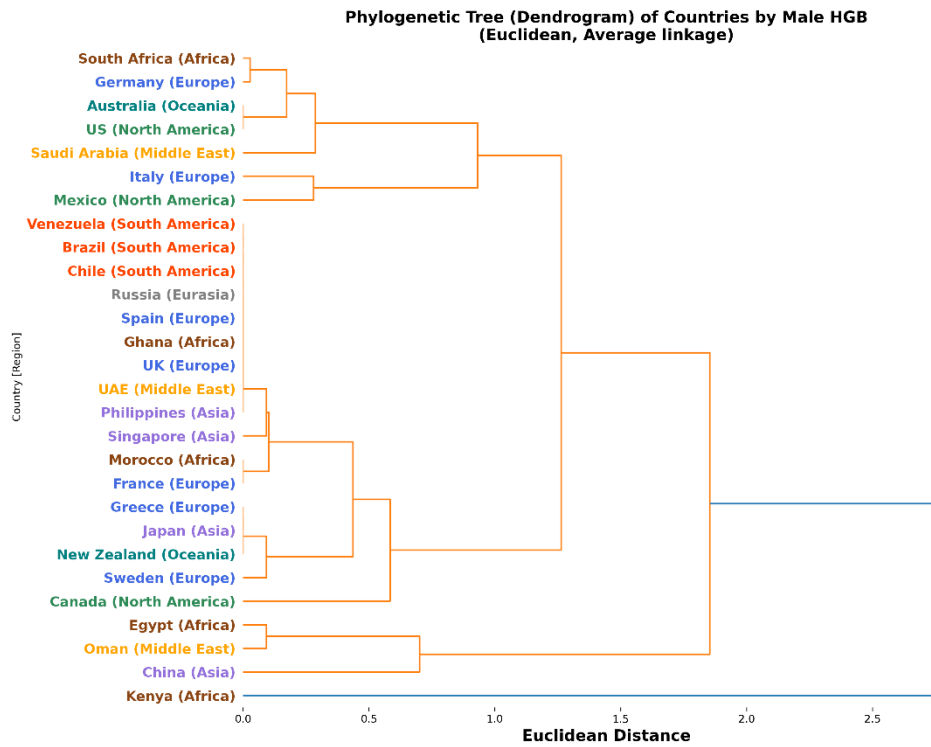


Figure 1.1.3-25. Phylogenetic Tree of Countries by Male HGB (Average Linkage + Euclidean Distance)

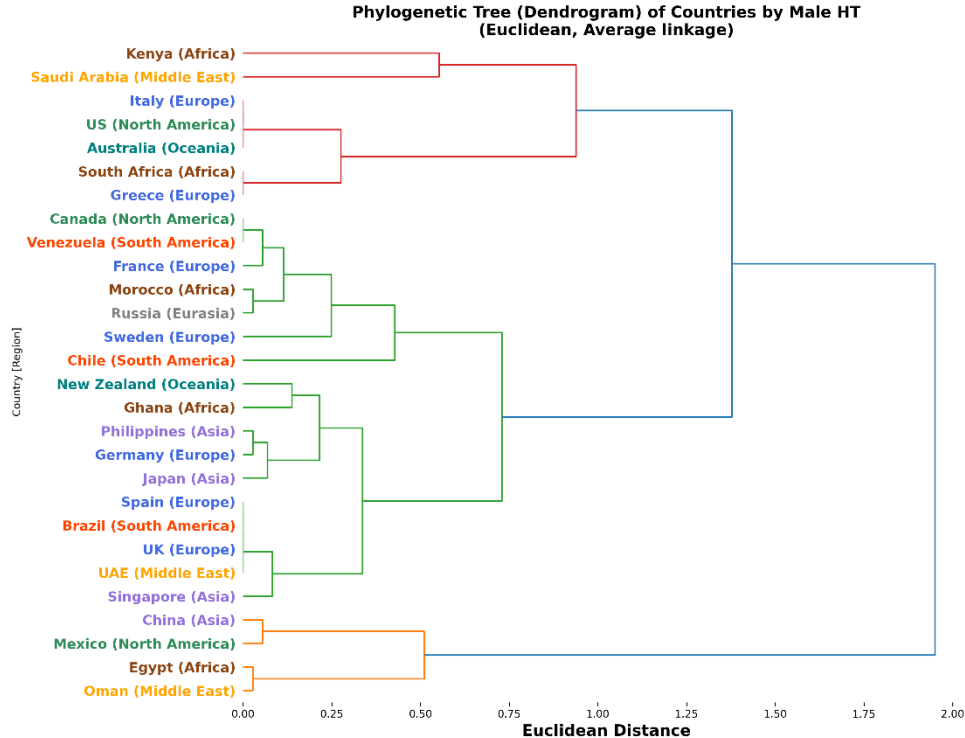


Figure 1.1.3-26. Phylogenetic Tree of Countries by Male HT (Average Linkage + Euclidean Distance)

1.1.3.1.2 Female

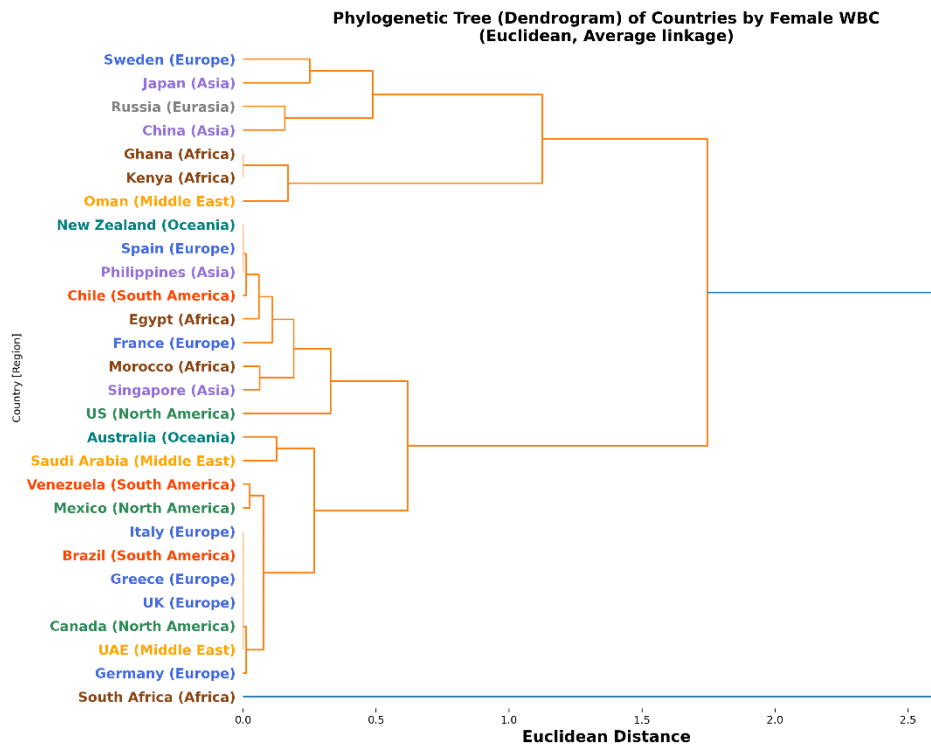


Figure 1.1.3-27. Phylogenetic Tree of Countries by Female WBC (Average Linkage + Euclidean Distance)

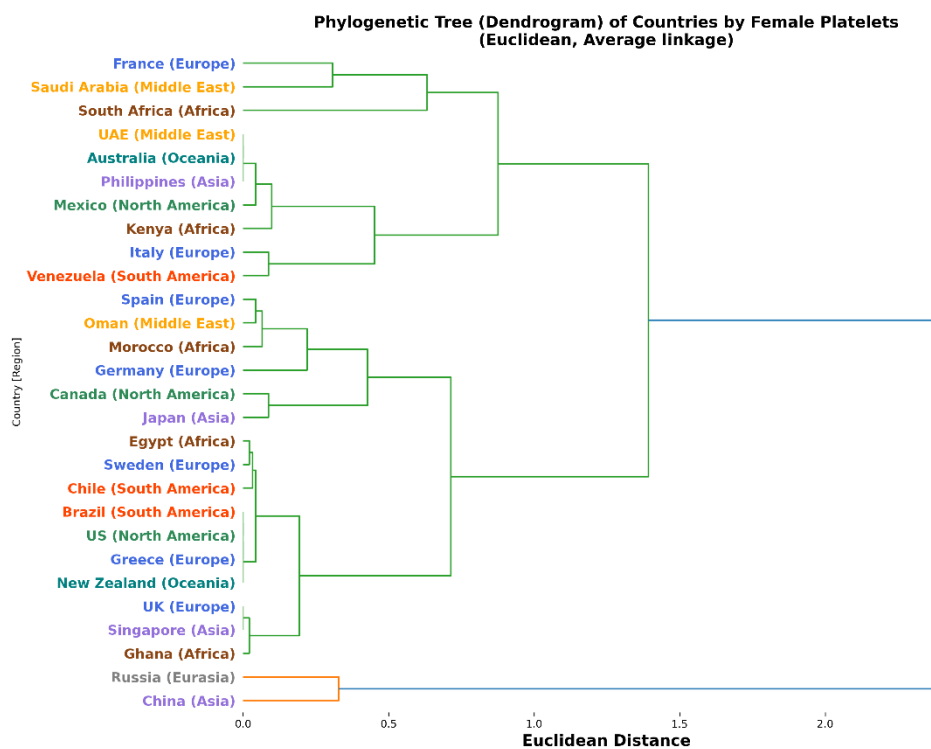


Figure 1.1.3-28. Phylogenetic Tree of Countries by Female Platelets (Average Linkage + Euclidean Distance)

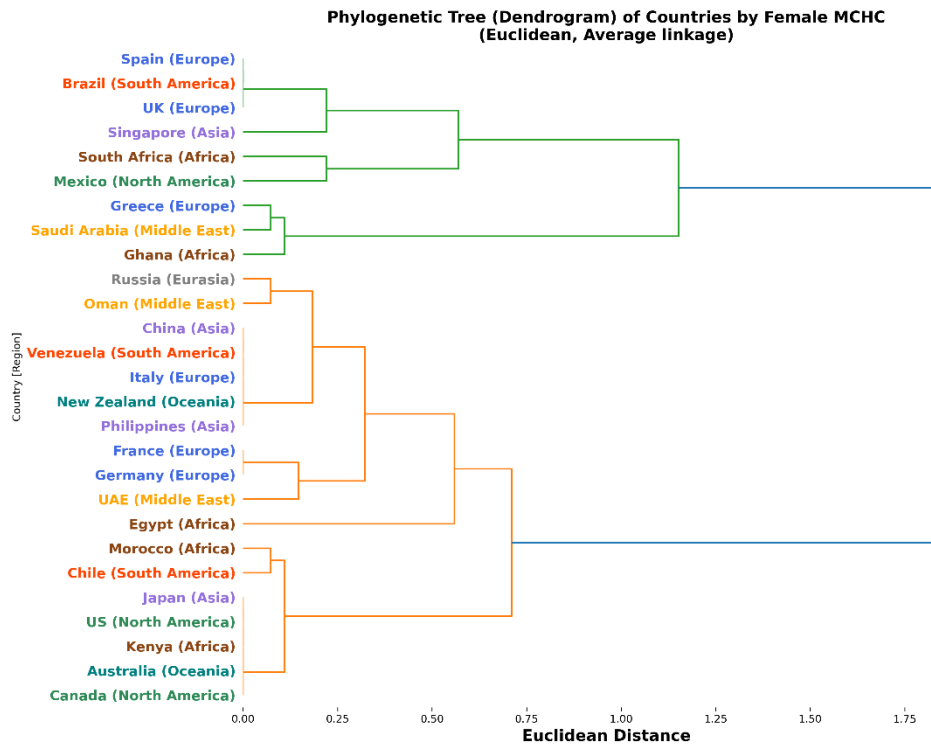


Figure 1.1.3-29. Phylogenetic Tree of Countries by Female MCHC (Average Linkage + Euclidean Distance)

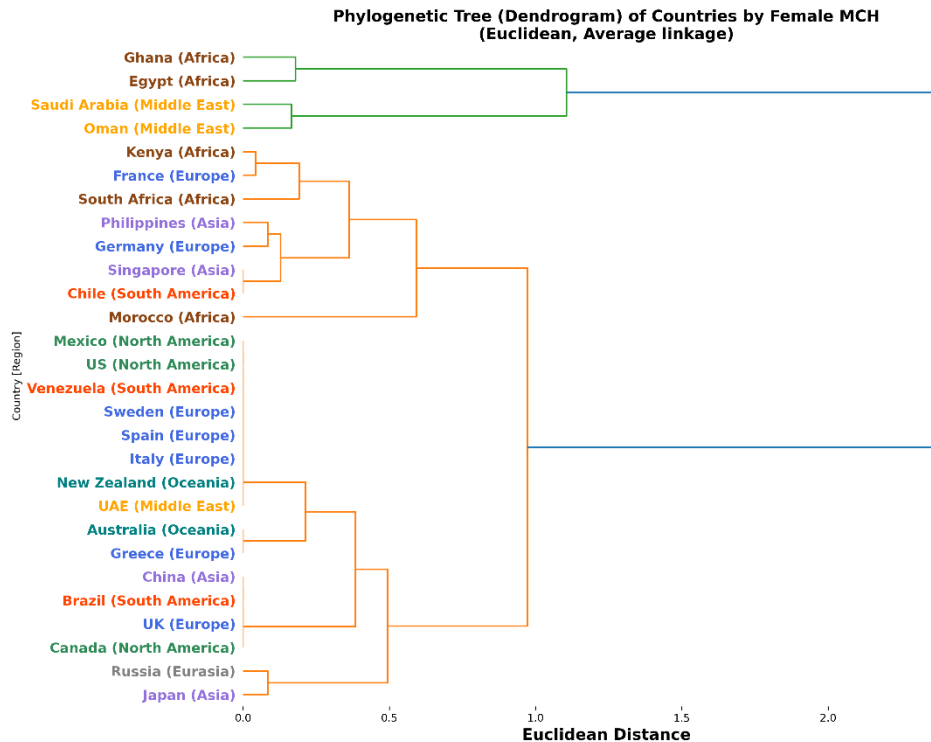


Figure 1.1.3-30. Phylogenetic Tree of Countries by Female MCH (Average Linkage + Euclidean Distance)

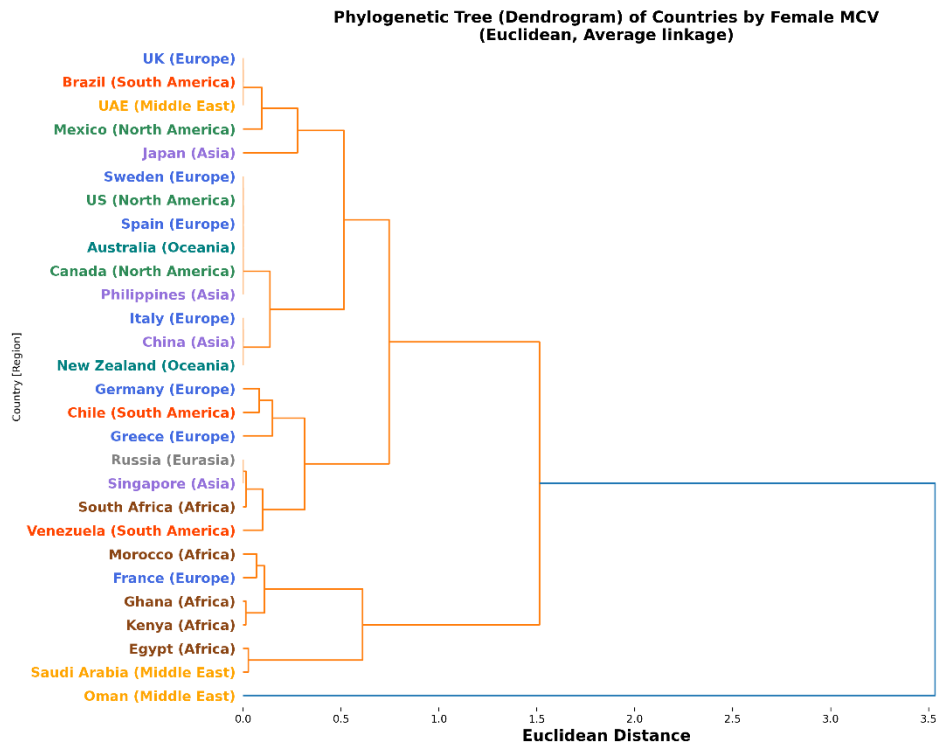


Figure 1.1.3-31. Phylogenetic Tree of Countries by Female MCV (Average Linkage + Euclidean Distance)

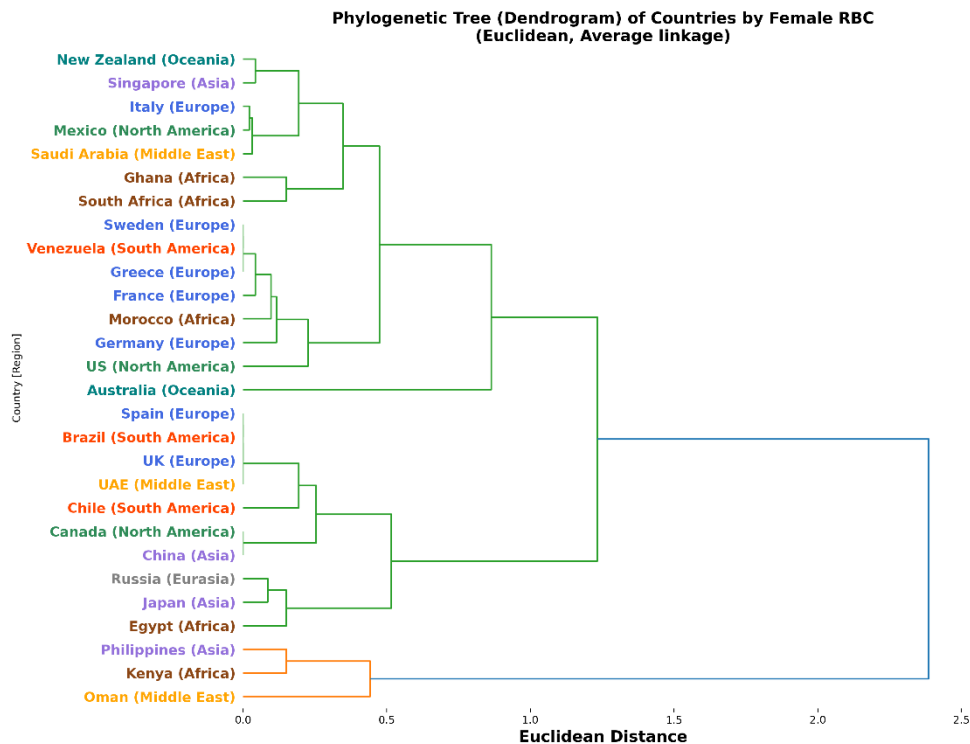


Figure 1.1.3-32. Phylogenetic Tree of Countries by Female RBC (Average Linkage + Euclidean Distance)

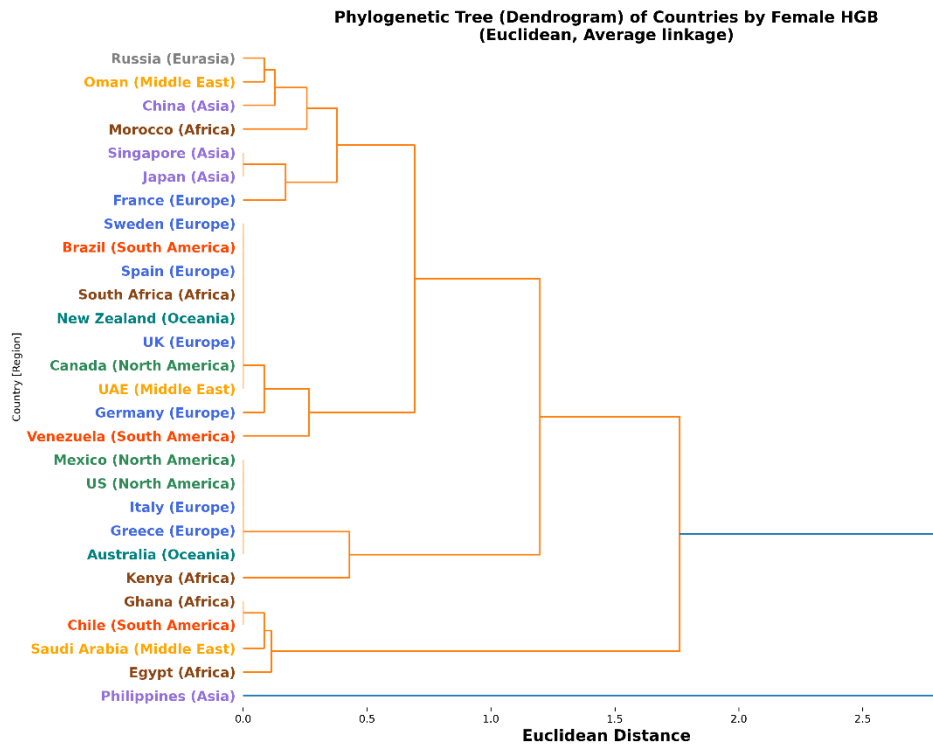


Figure 1.1.3-33. Phylogenetic Tree of Countries by Female HGB (Average Linkage + Euclidean Distance)

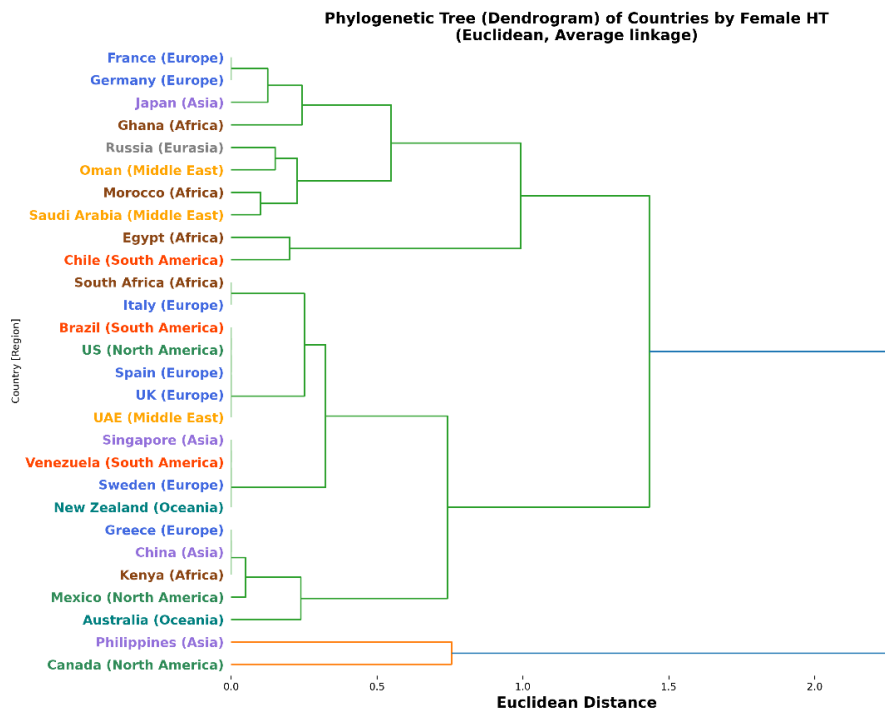


Figure 1.1.3-34. Phylogenetic Tree of Countries by Female HT (Average Linkage + Euclidean Distance)

MULTIVARIATE PHYLOGENETIC TREE

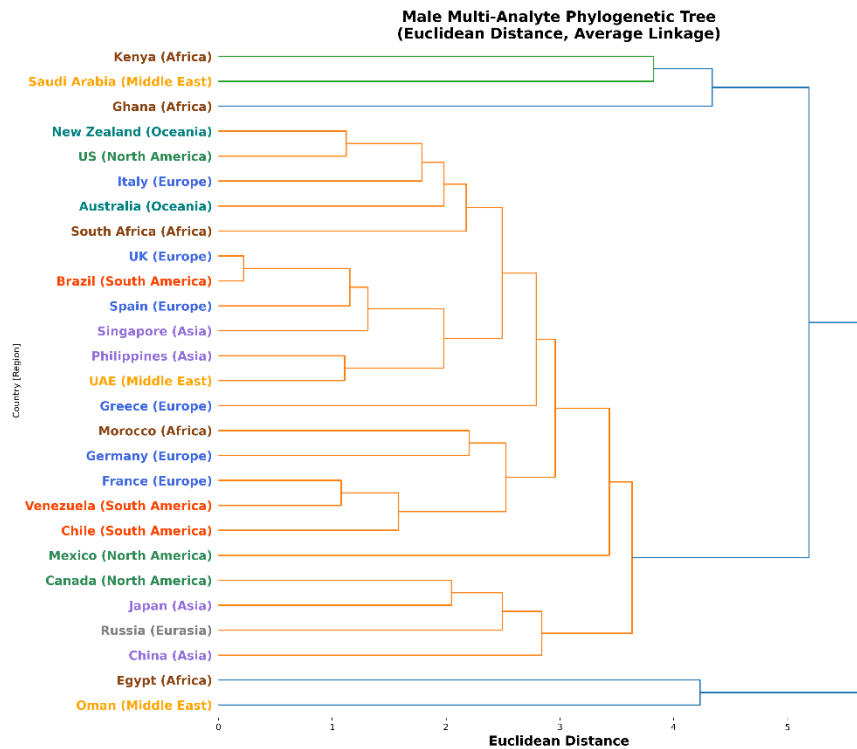


Figure 1.1.3-35. Male Multi-Analyte Phylogenetic Tree (Average Linkage + Euclidean Distance)

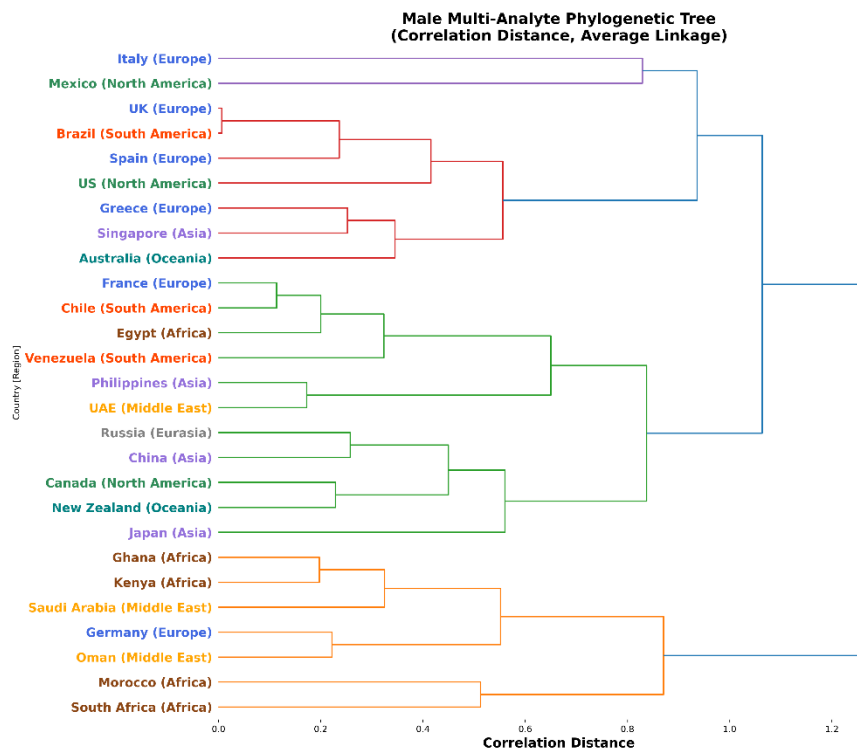


Figure 1.1.3-36. Male Multi-Analyte Phylogenetic Tree (Average Linkage + Correlation Distance)

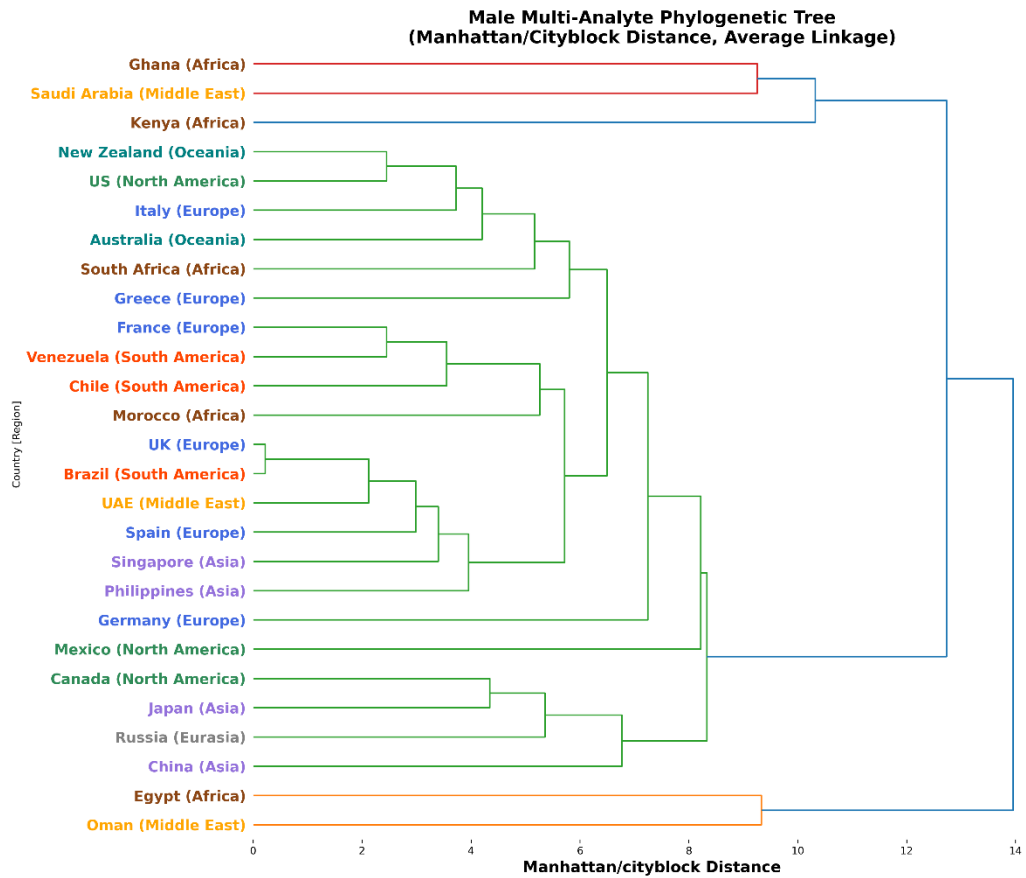


Figure 1.1.3-37. Male Multi-Analyte Phylogenetic Tree (Average Linkage + Manhattan Distance)

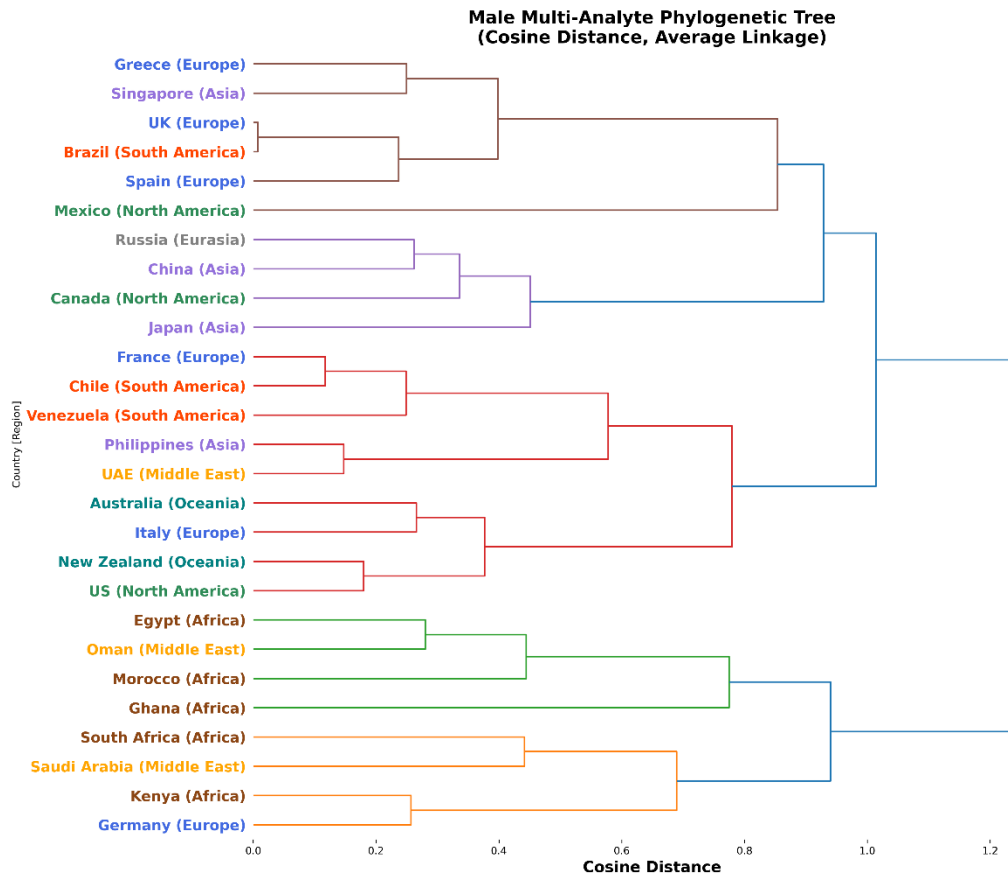


Figure 1.1.3-38. Male Multi-Analyte Phylogenetic Tree (Average Linkage + Cosine Distance)

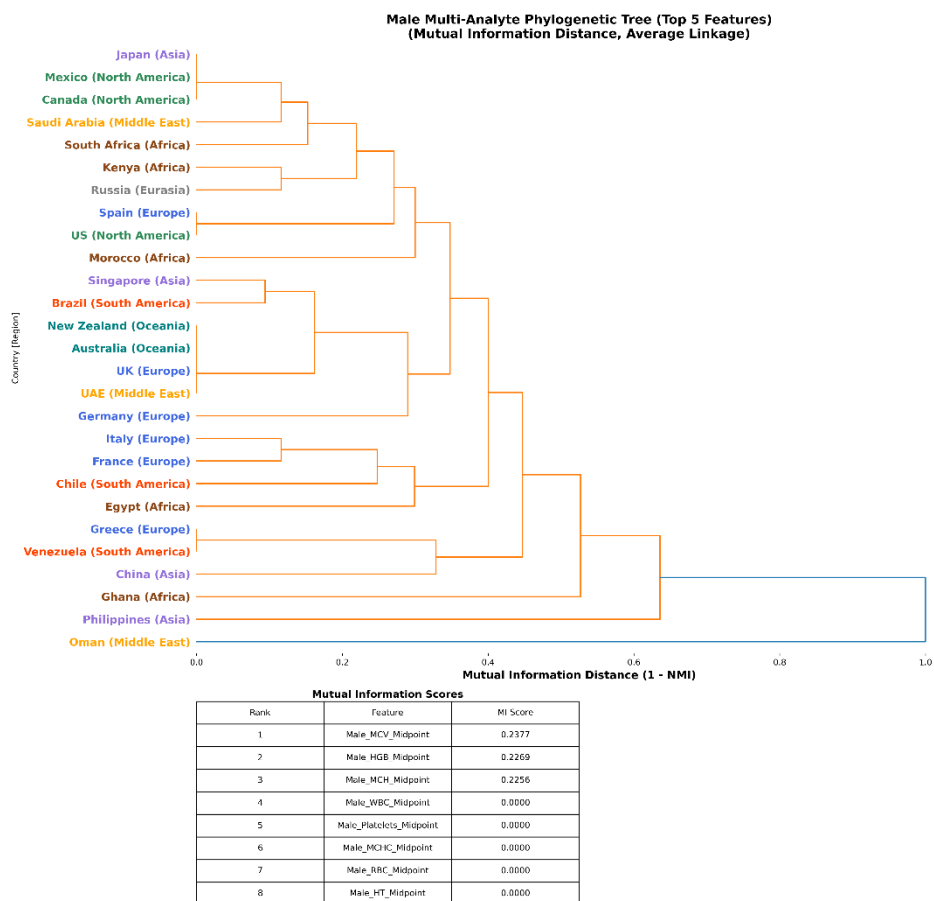


Figure 1.1.3-39. Male Multi-Analyte Phylogenetic Tree (Average Linkage + Mutual Information Distance (Top 5))

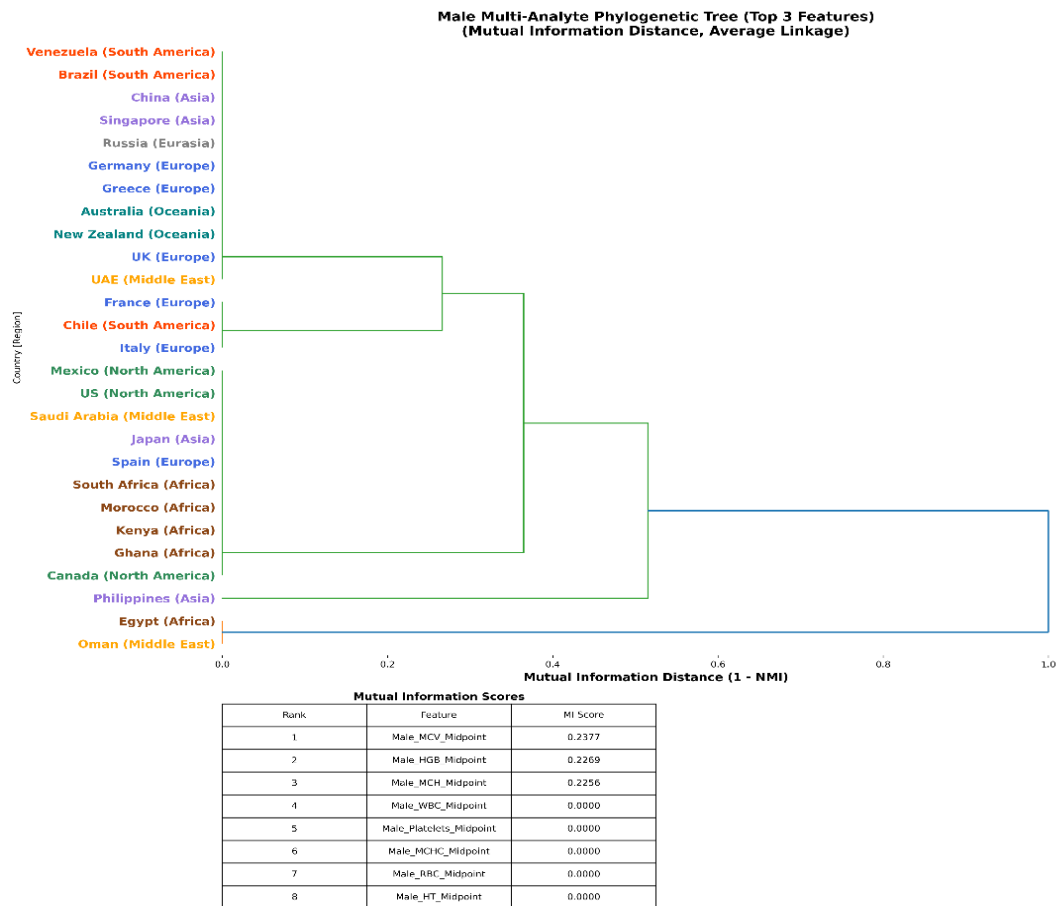


Figure 1.1.3-40. Male Multi-Analyte Phylogenetic Tree (Average Linkage + Mutual Information Distance (Top 3))

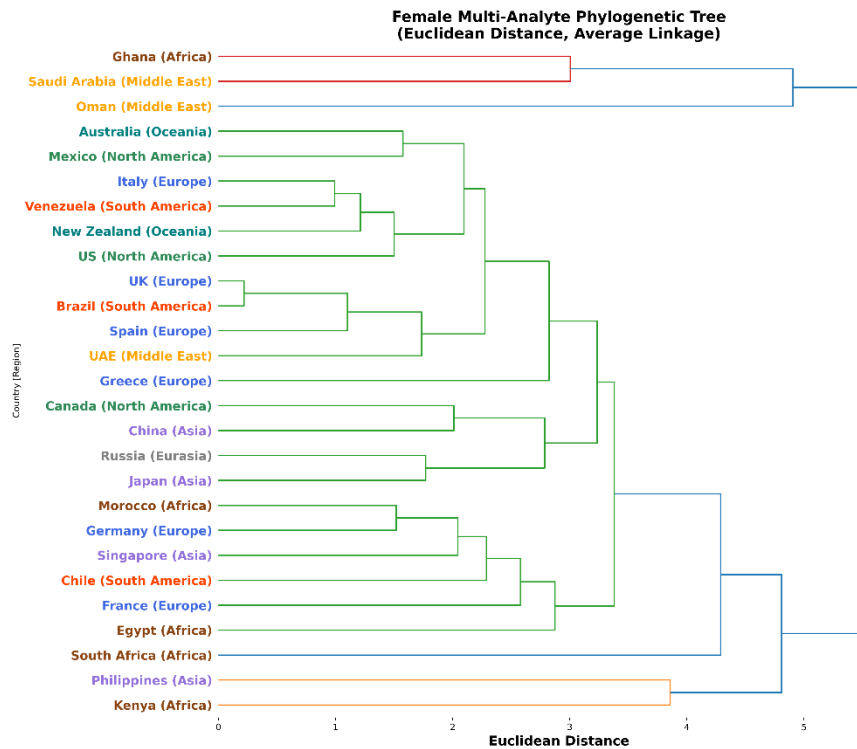


Figure 1.1.3-41. Female Multi-Analyte Phylogenetic Tree (Average Linkage + Euclidean Distance)

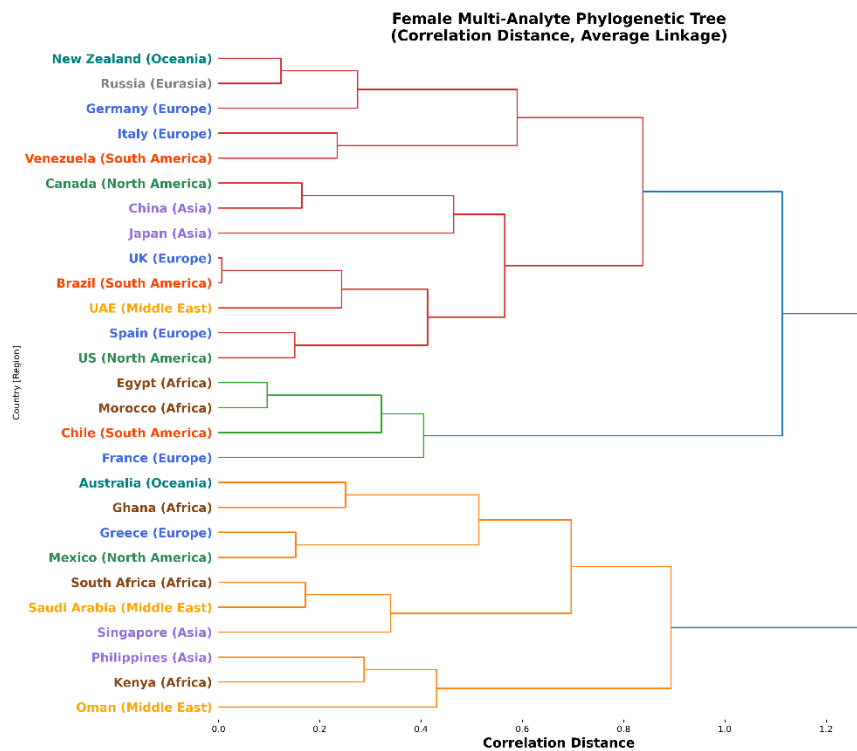


Figure 1.1.3-42. Female Multi-Analyte Phylogenetic Tree (Average Linkage + Correlation Distance)

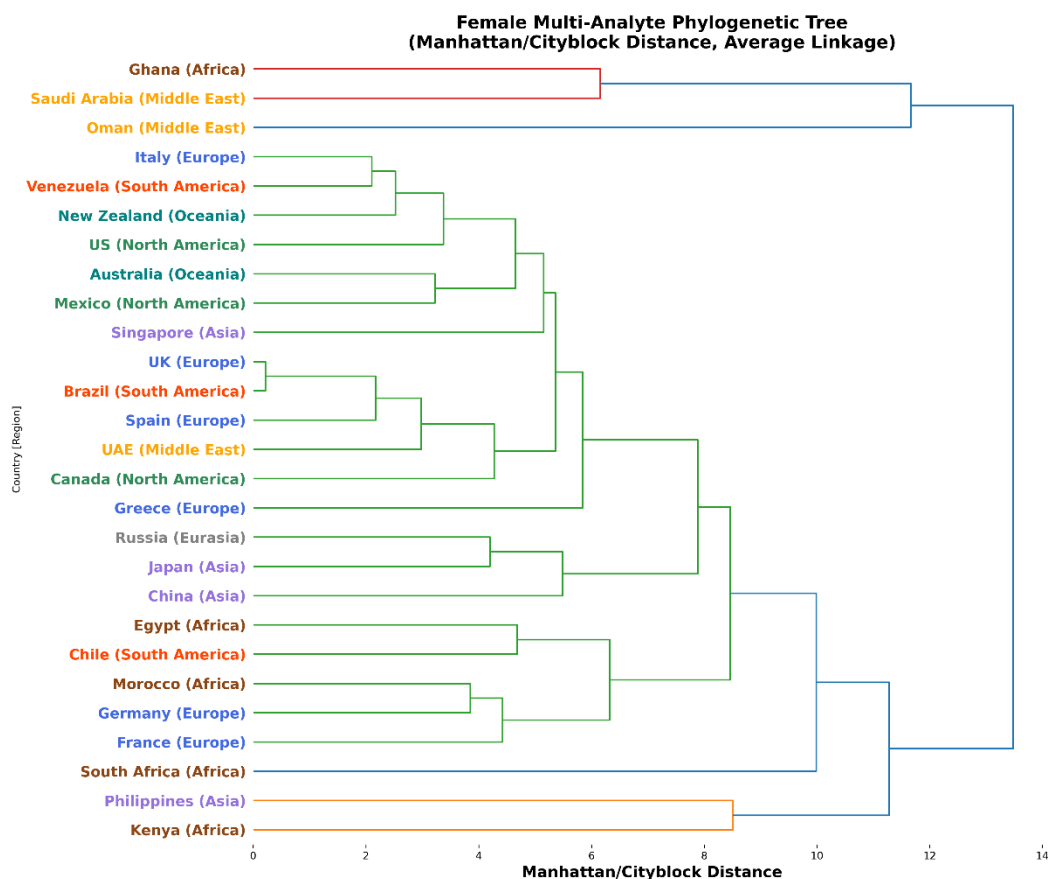


Figure 1.1.3-43. Female Multi-Analyte Phylogenetic Tree (Average Linkage + Manhattan Distance)

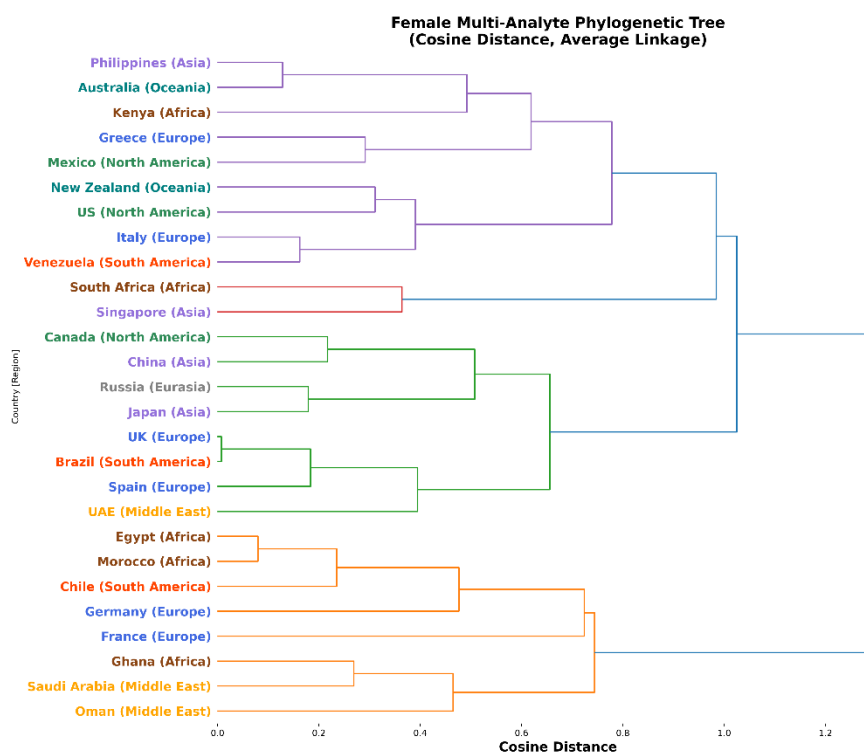


Figure 1.1.3-44. Female Multi-Analyte Phylogenetic Tree (Average Linkage + Cosine Distance)

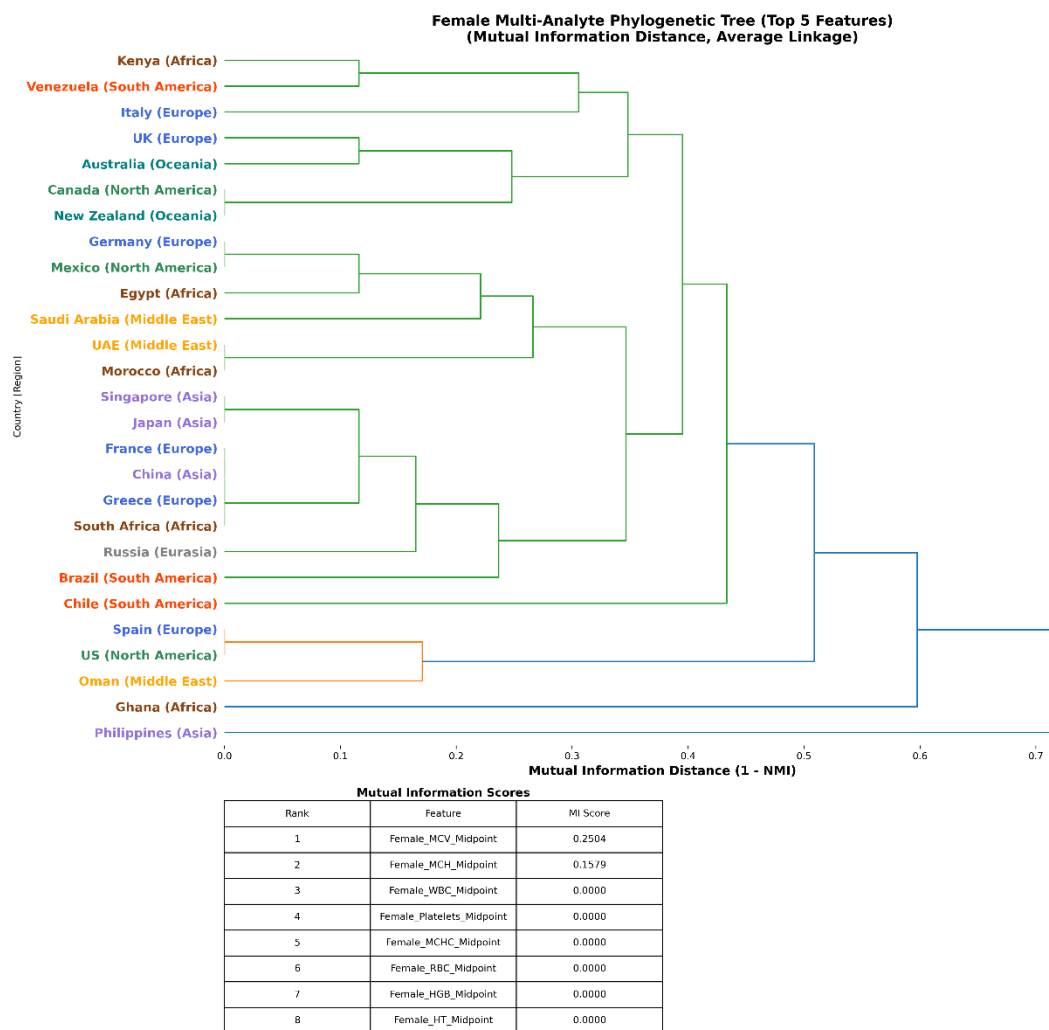


Figure 1.1.3-45. Female Multi-Analyte Phylogenetic Tree (Average Linkage + Mutual Information Distance (Top 5))

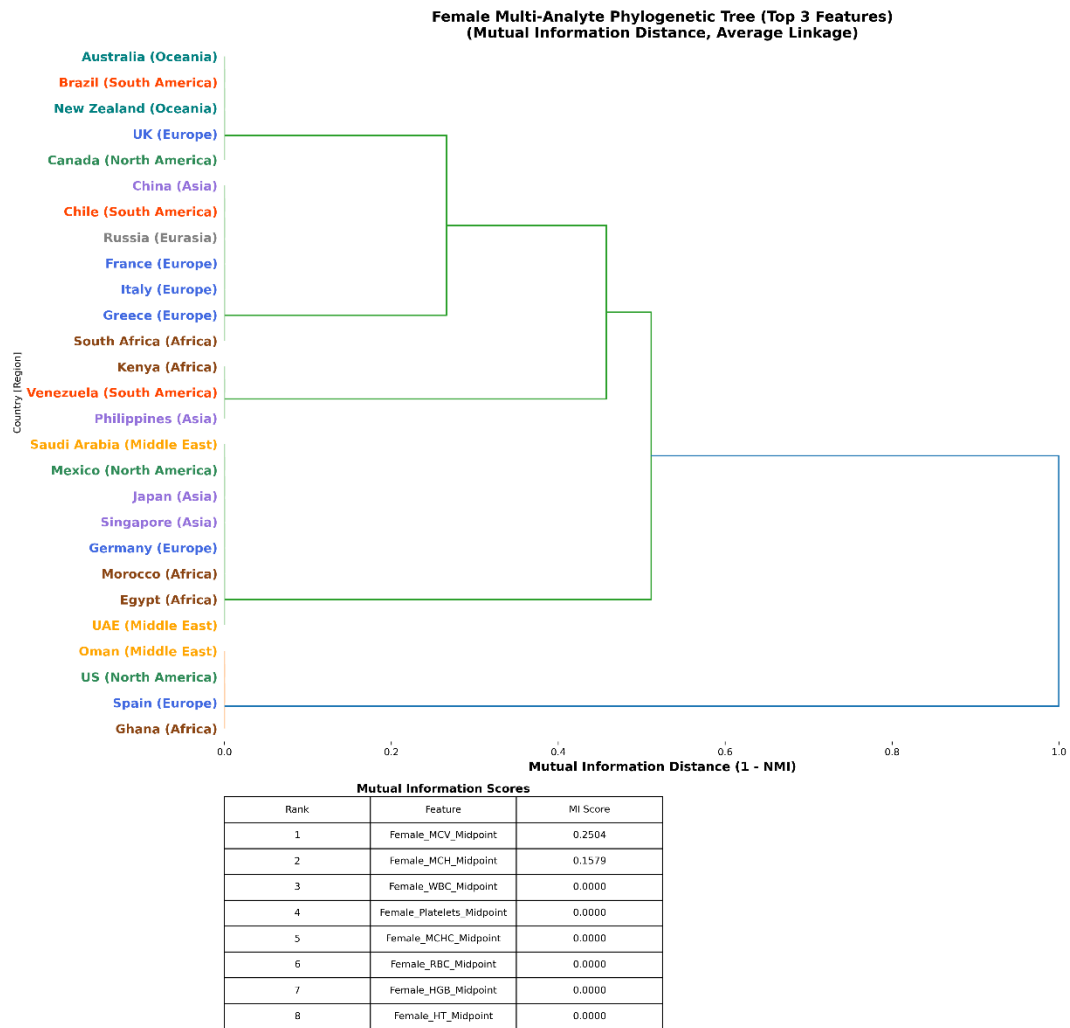


Figure 1.1.3-46. Female Multi-Analyte Phylogenetic Tree (Average Linkage + Mutual Information Distance (Top 3))

Comparative Analysis Using the Same Clustering Combinations for BMI

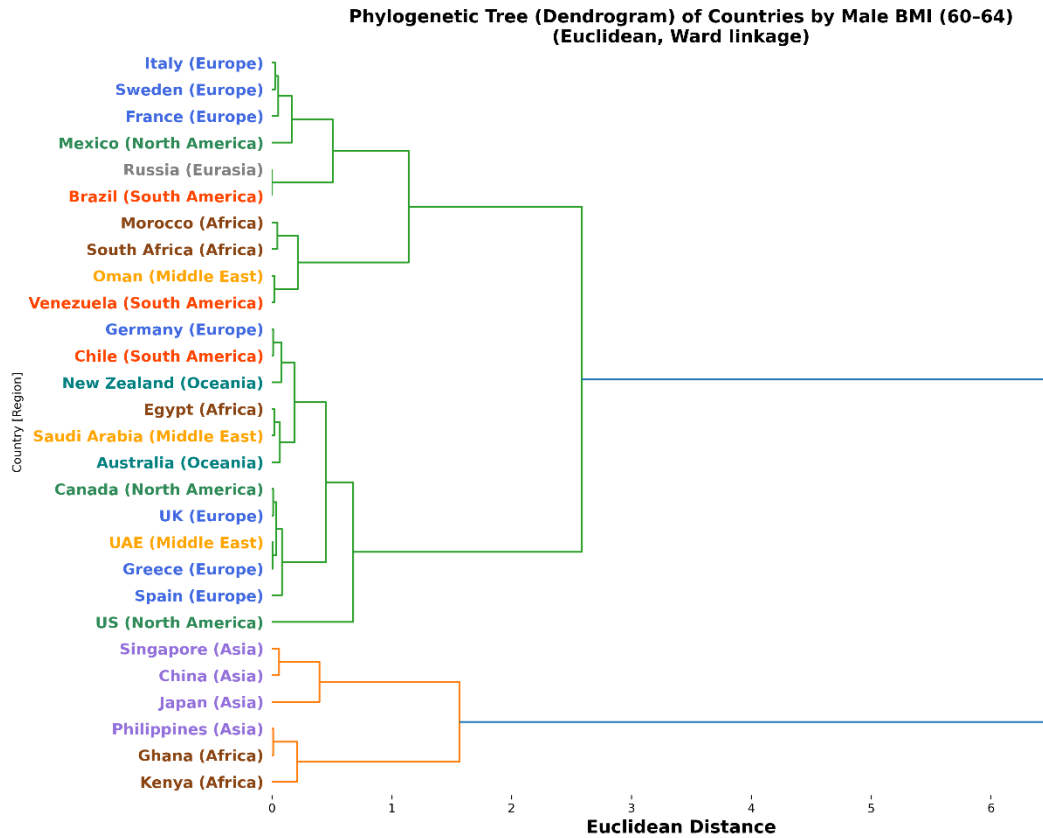


Figure 1.1.3-47. Phylogenetic Tree of Countries by Male Mean BMI (60-64) (Ward Linkage + Euclidean Distance)

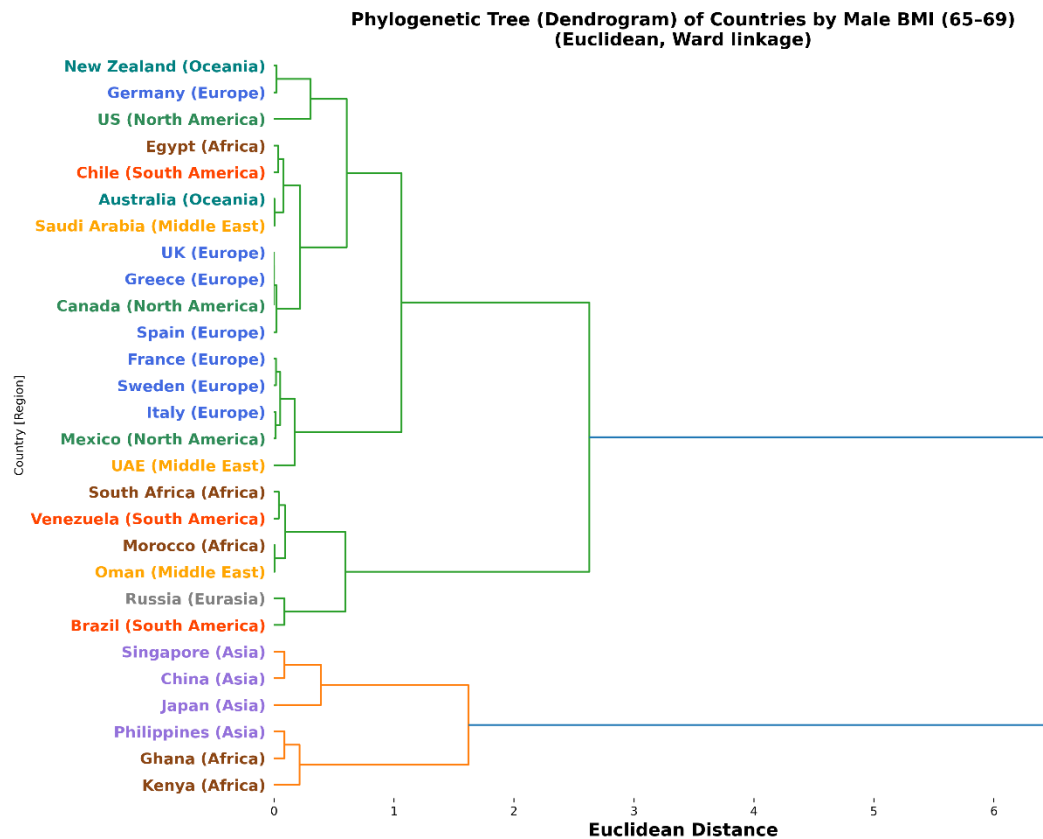


Figure 1.1.3-48. Phylogenetic Tree of Countries by Male Mean BMI (65-69) (Ward Linkage + Euclidean Distance)

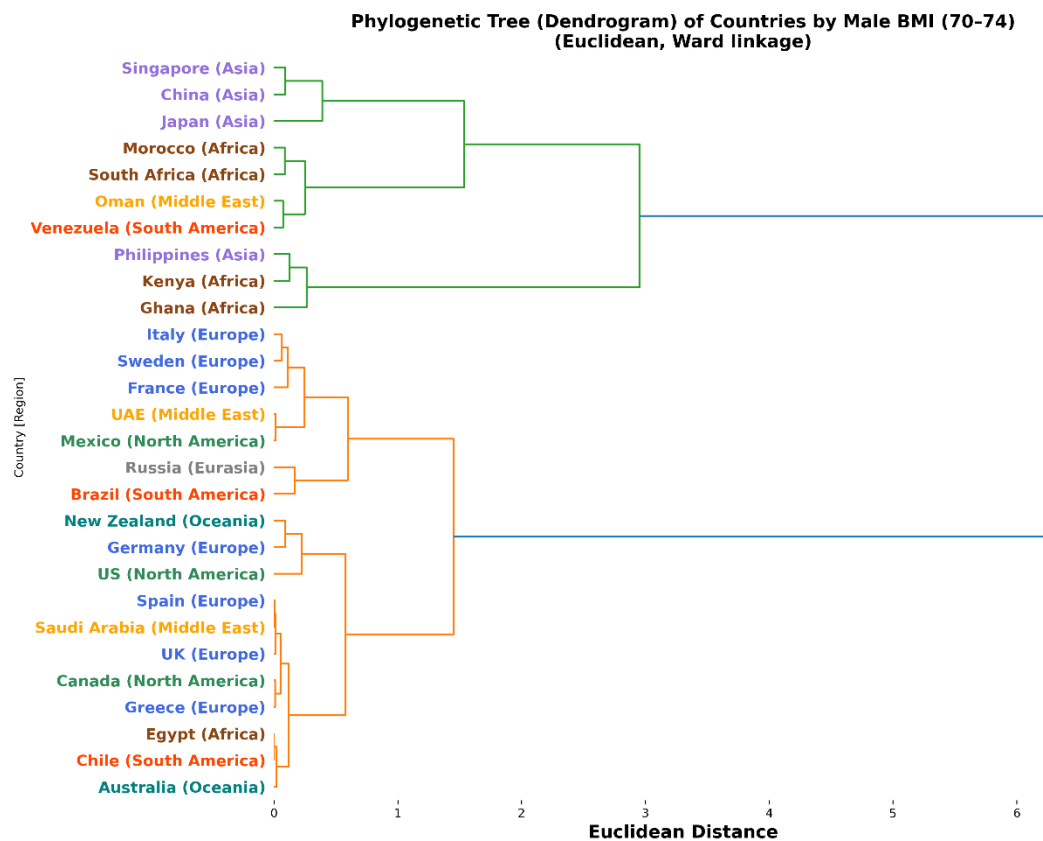


Figure 1.1.3-49. Phylogenetic Tree of Countries by Male Mean BMI (70-74) (Ward Linkage + Euclidean Distance)

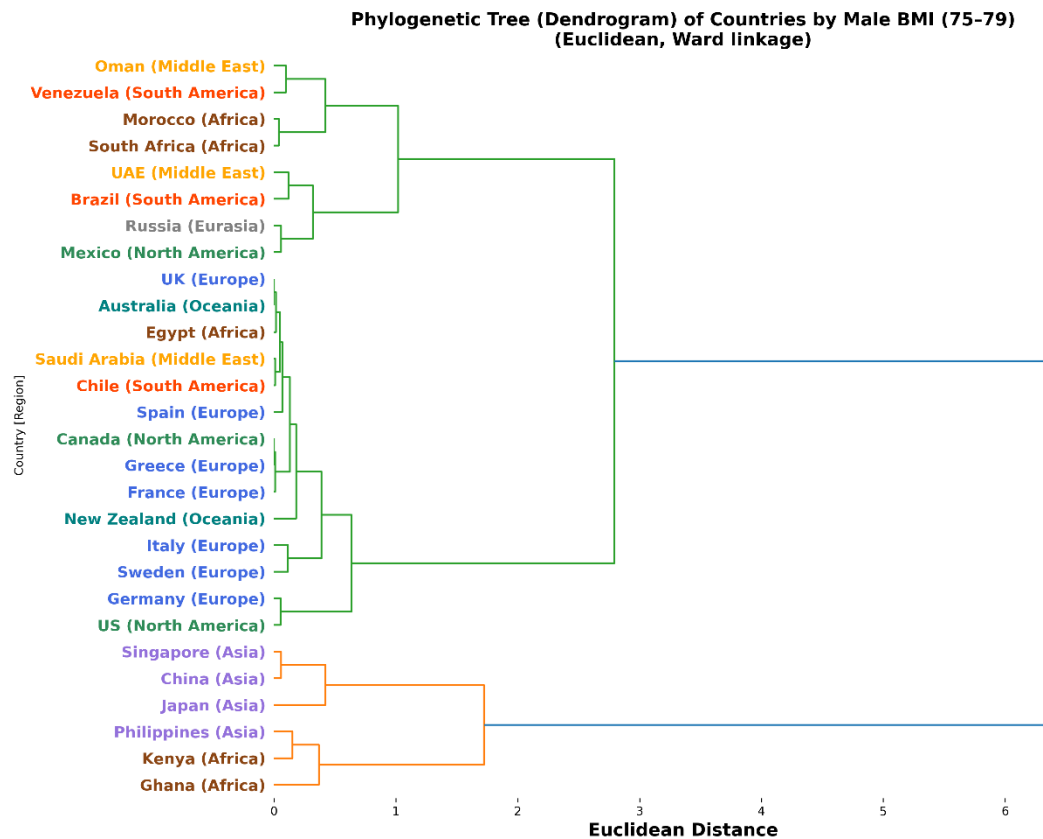


Figure 1.1.3-50. Phylogenetic Tree of Countries by Male Mean BMI (75-79) (Ward Linkage + Euclidean Distance)

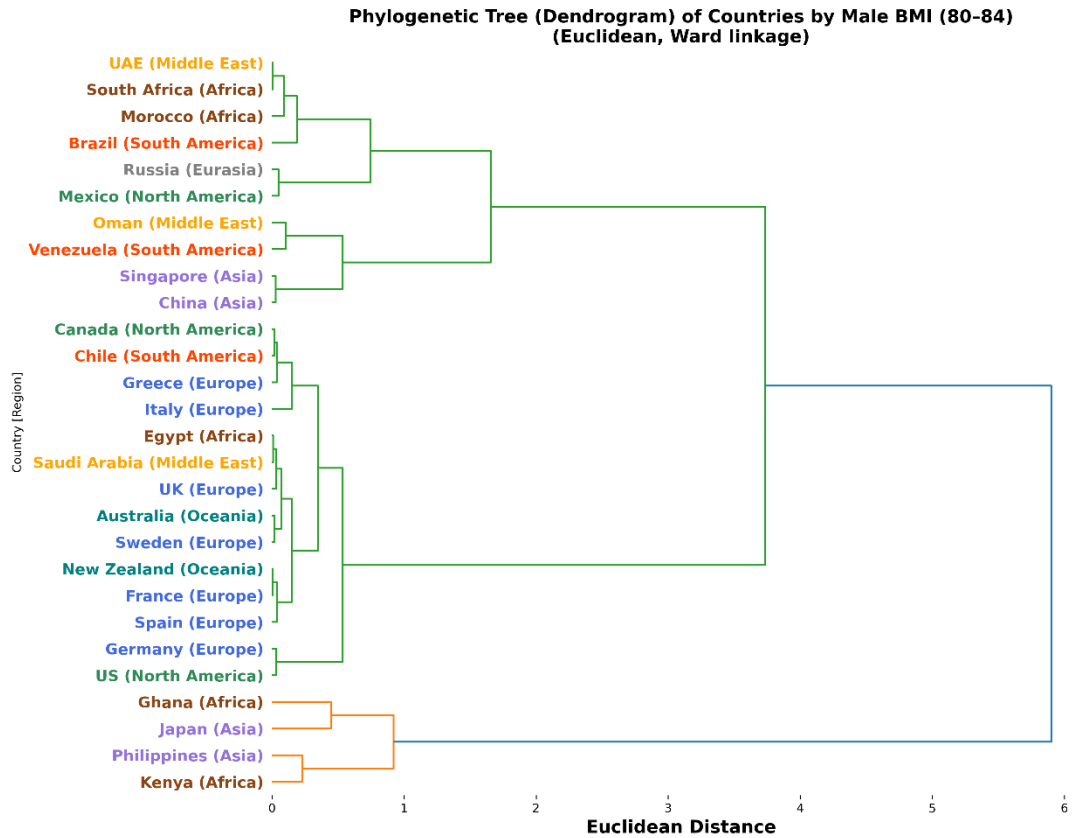


Figure 1.1.3-51. Phylogenetic Tree of Countries by Male Mean BMI (80-84) (Ward Linkage + Euclidean Distance)

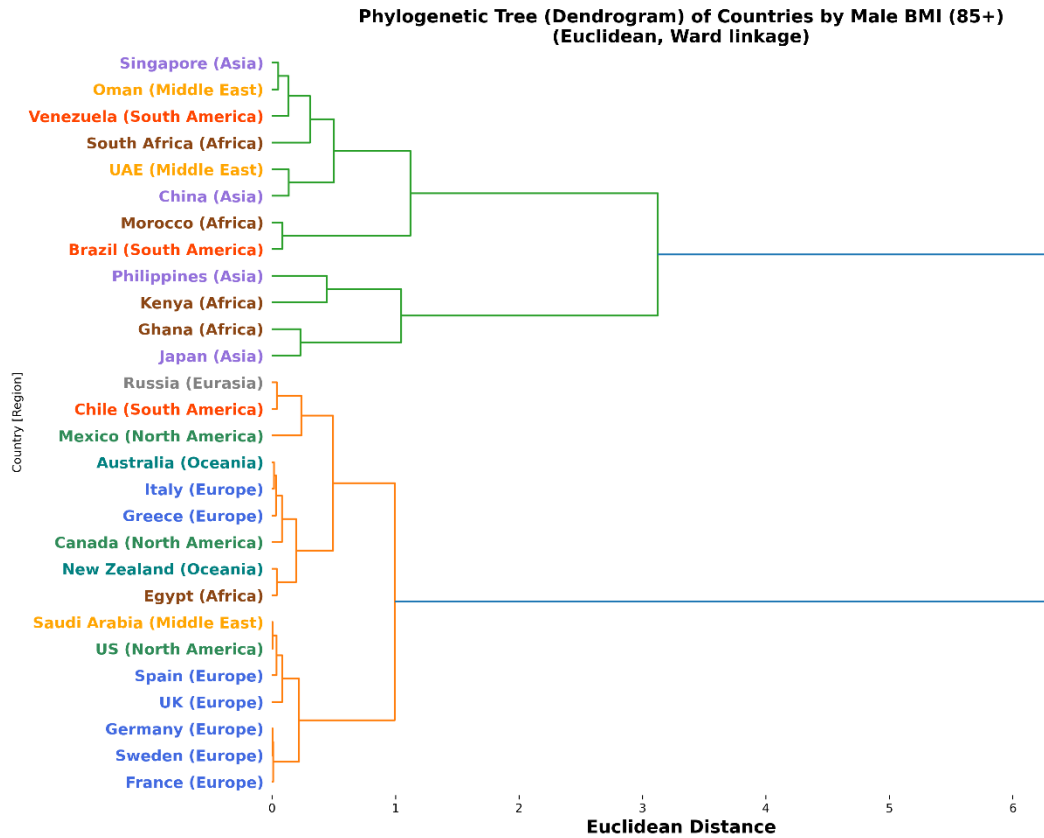


Figure 1.1.3-52. Phylogenetic Tree of Countries by Male Mean BMI (85+) (Ward Linkage + Euclidean Distance)

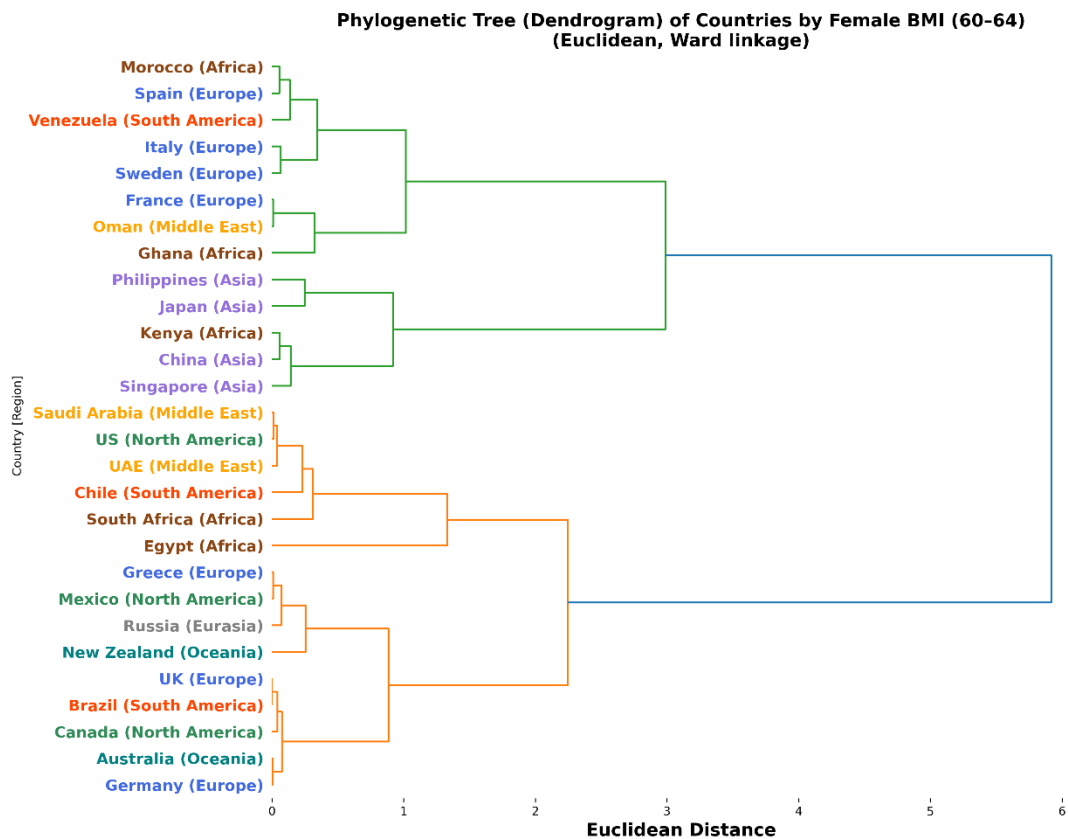


Figure 1.1.3-53. Phylogenetic Tree of Countries by Female Mean BMI (60-64) (Ward Linkage + Euclidean Distance)

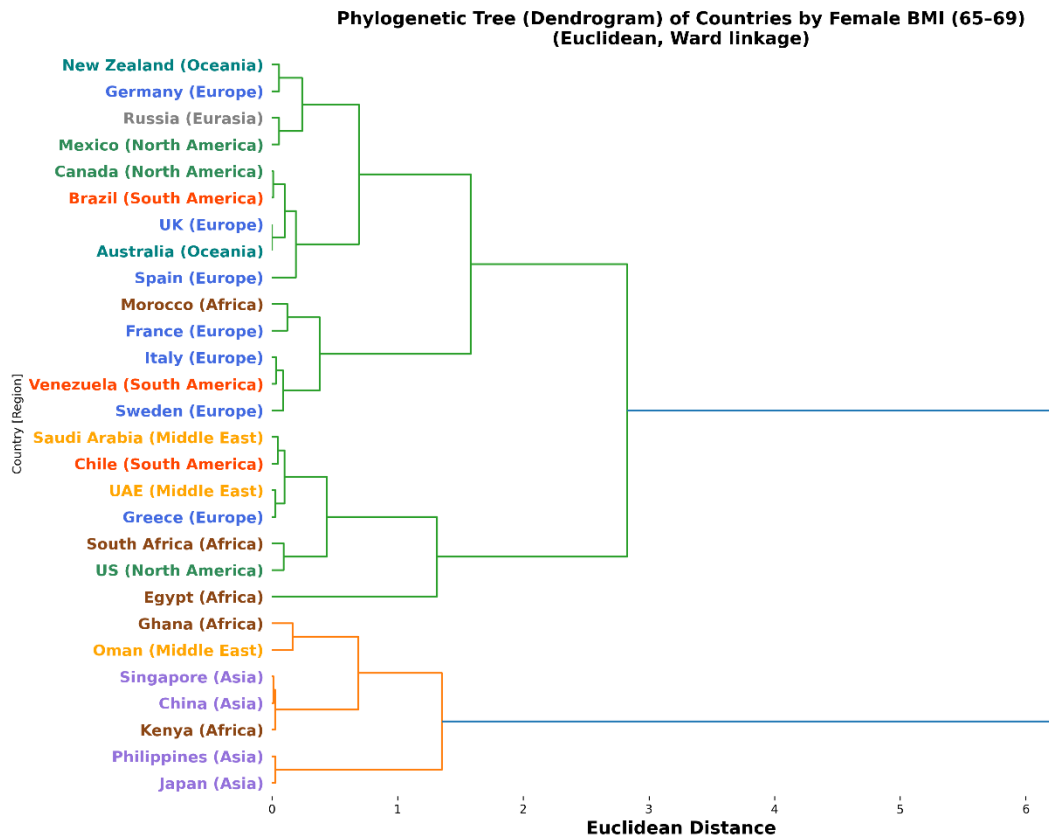


Figure 1.1.3-54. Phylogenetic Tree of Countries by Female Mean BMI (65-69) (Ward Linkage + Euclidean Distance)

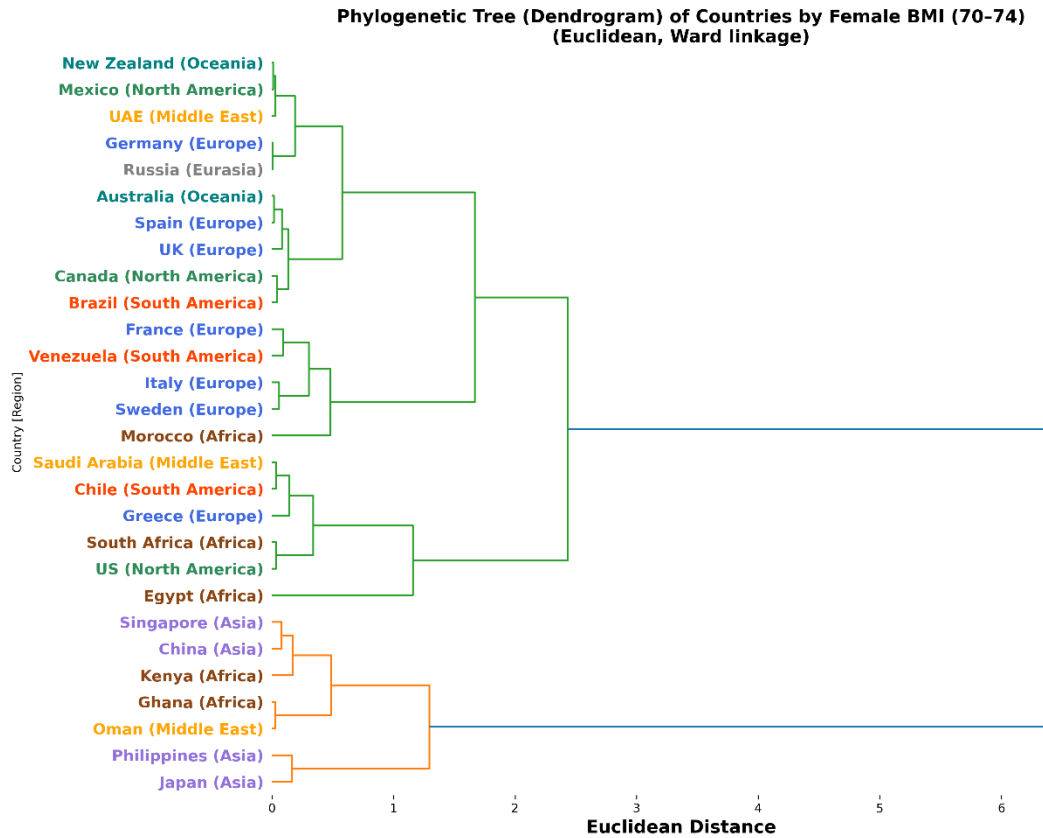


Figure 1.1.3-55. Phylogenetic Tree of Countries by Female Mean BMI (70-74) (Ward Linkage + Euclidean Distance)

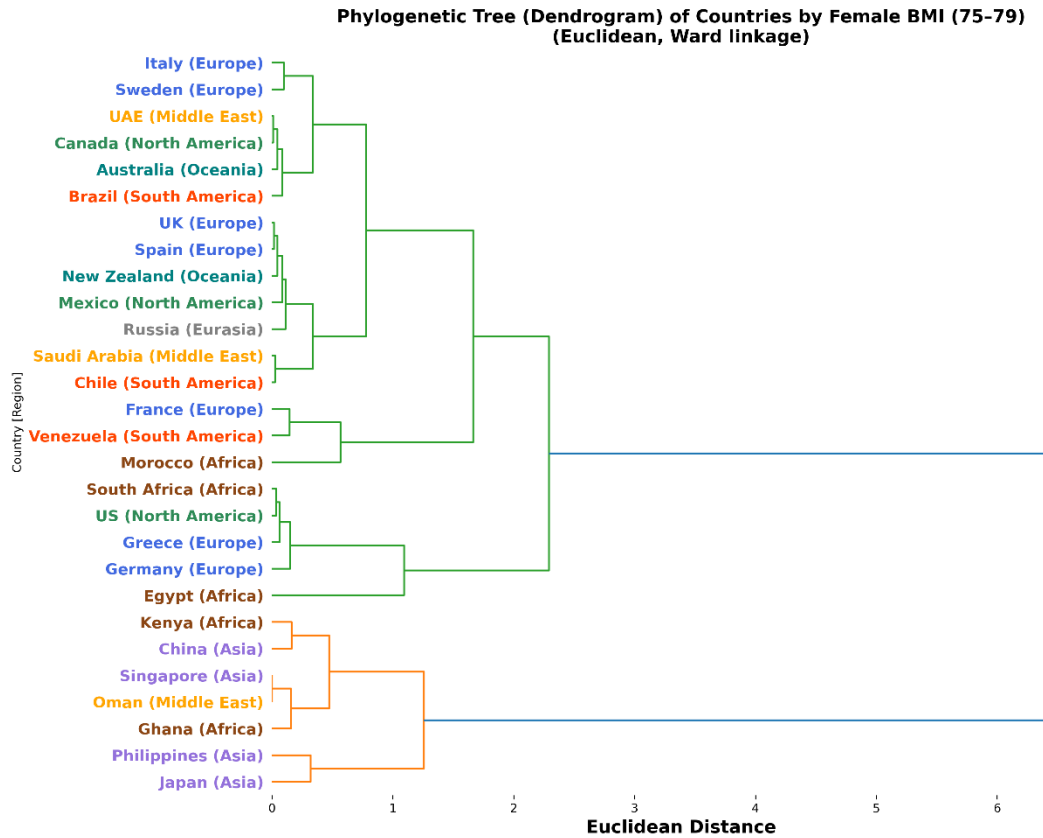


Figure 1.1.3-56. Phylogenetic Tree of Countries by Female Mean BMI (75-79) (Ward Linkage + Euclidean Distance)

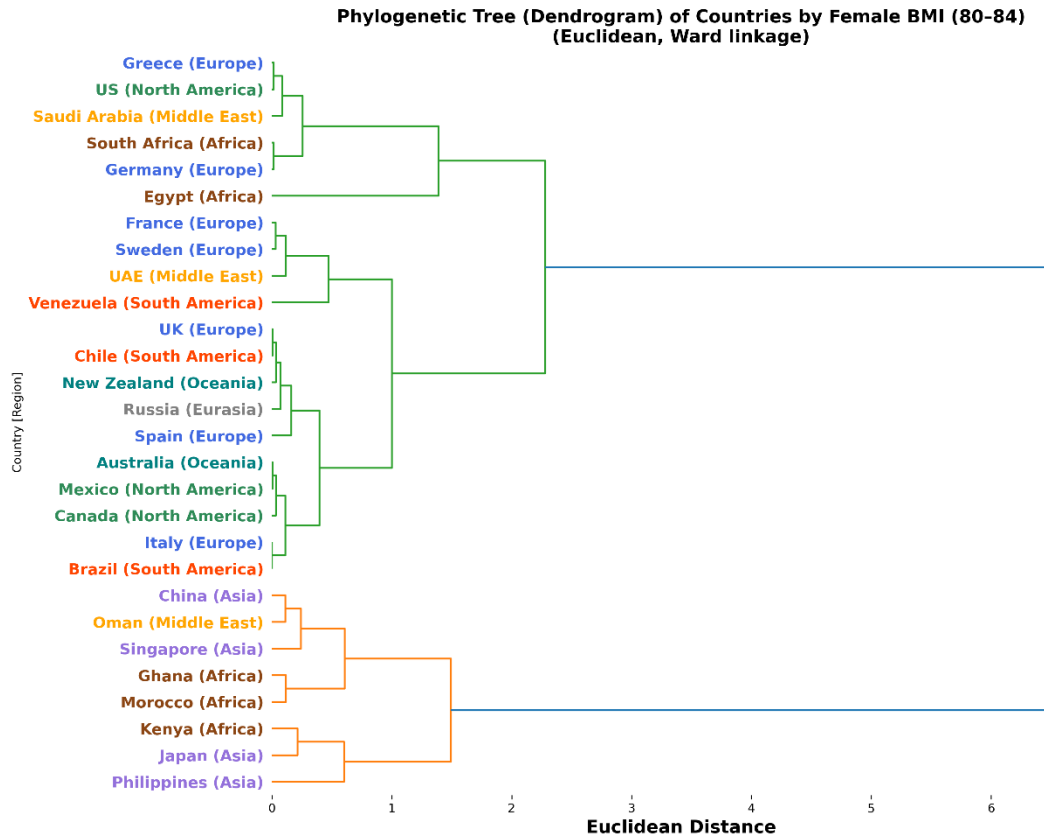


Figure 1.1.3-57. Phylogenetic Tree of Countries by Female Mean BMI (80-84) (Ward Linkage + Euclidean Distance)

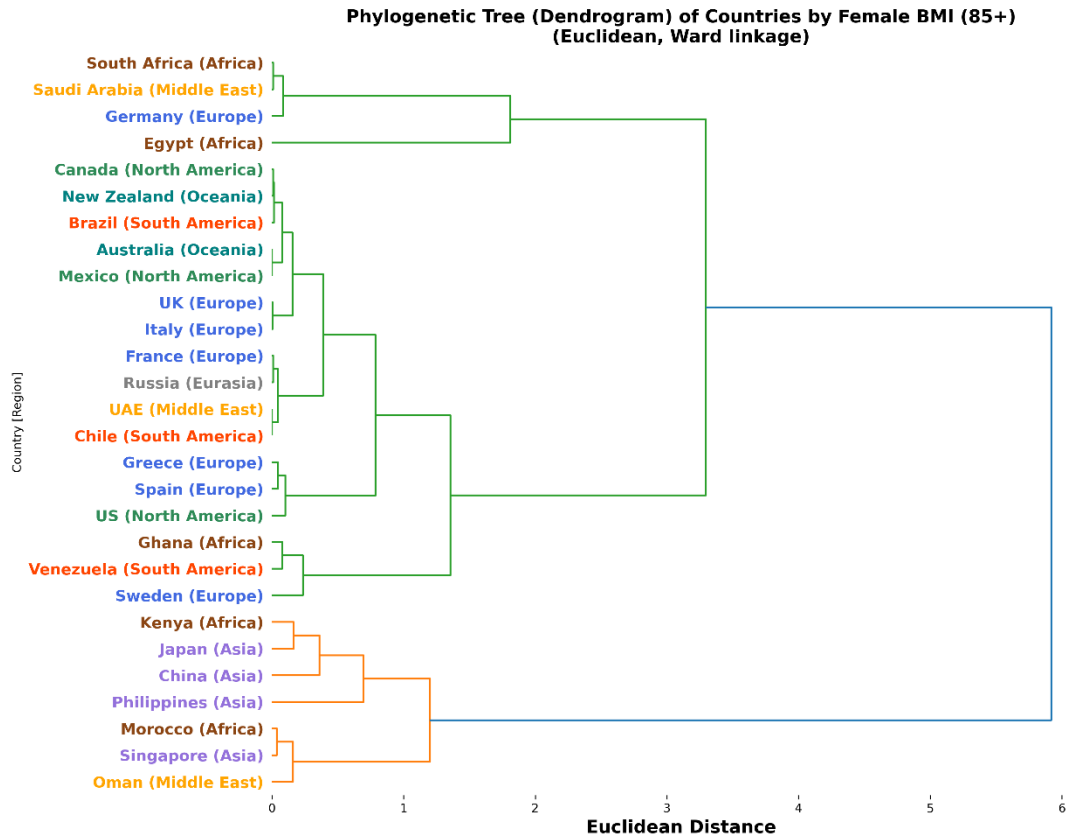


Figure 1.1.3-58. Phylogenetic Tree of Countries by Female Mean BMI (85+) (Ward Linkage + Euclidean Distance)

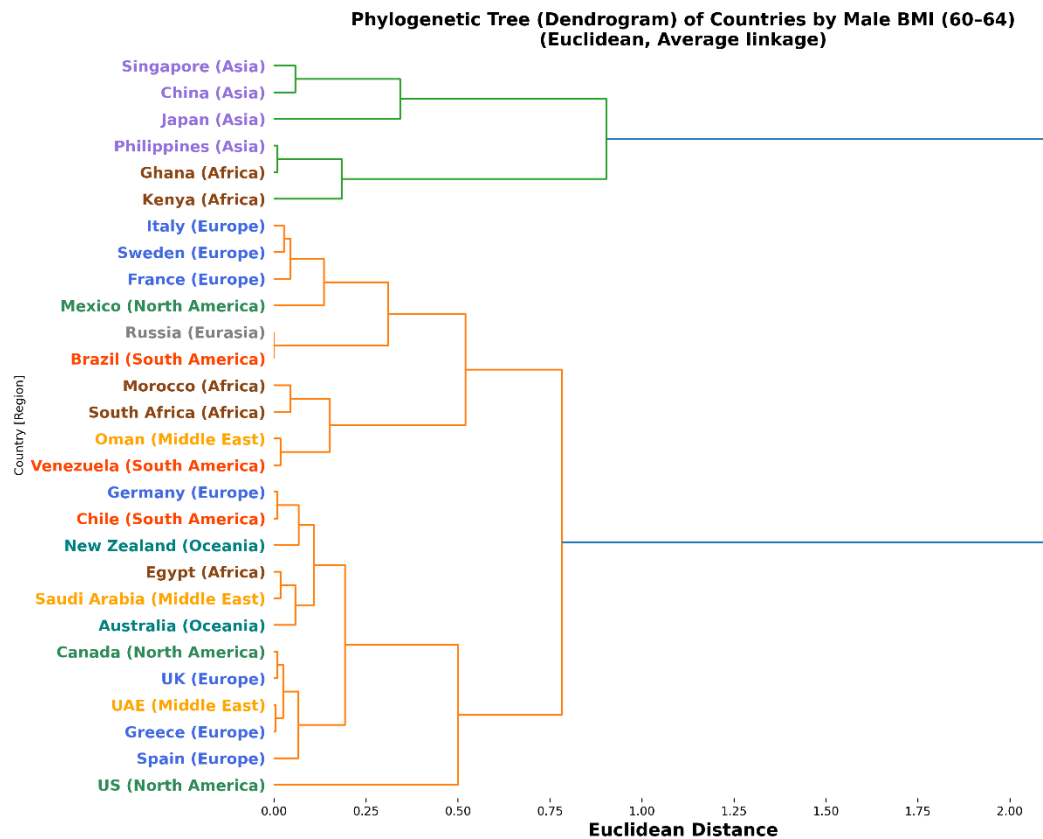


Figure 1.1.3-59. Phylogenetic Tree of Countries by Male Mean BMI (60-64) (Average Linkage + Euclidean Distance)

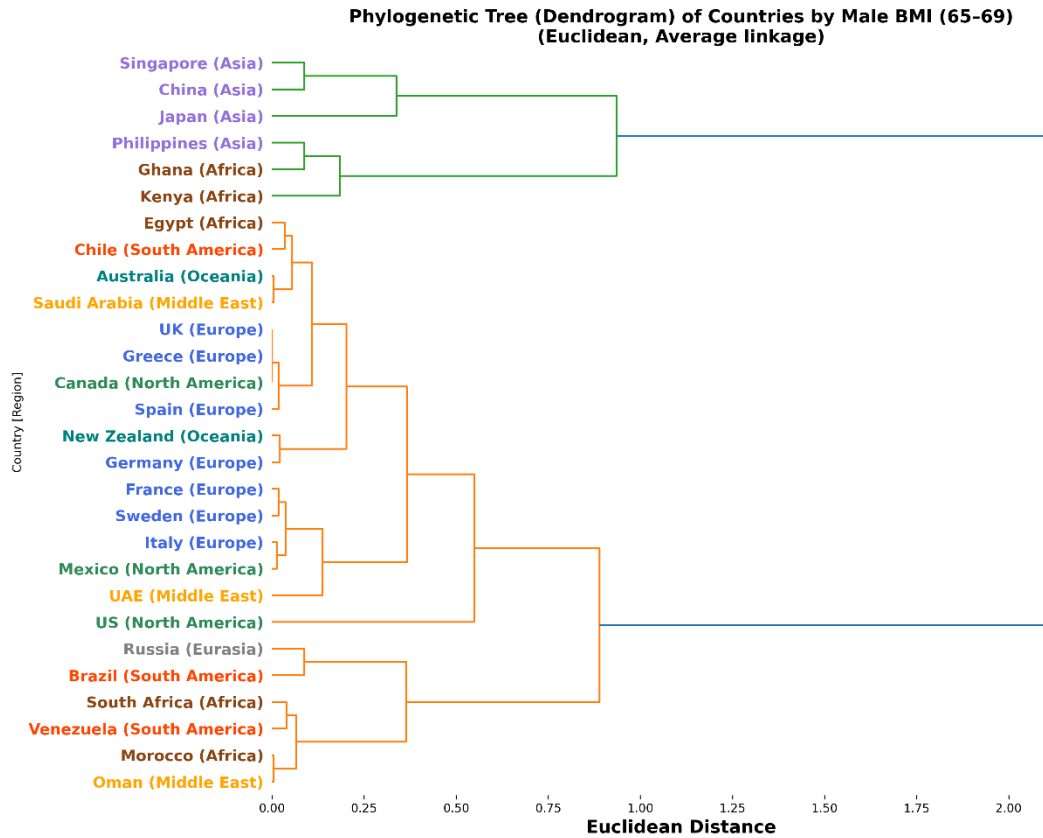


Figure 1.1.3-60. Phylogenetic Tree of Countries by Male Mean BMI (65-69) (Average Linkage + Euclidean Distance)

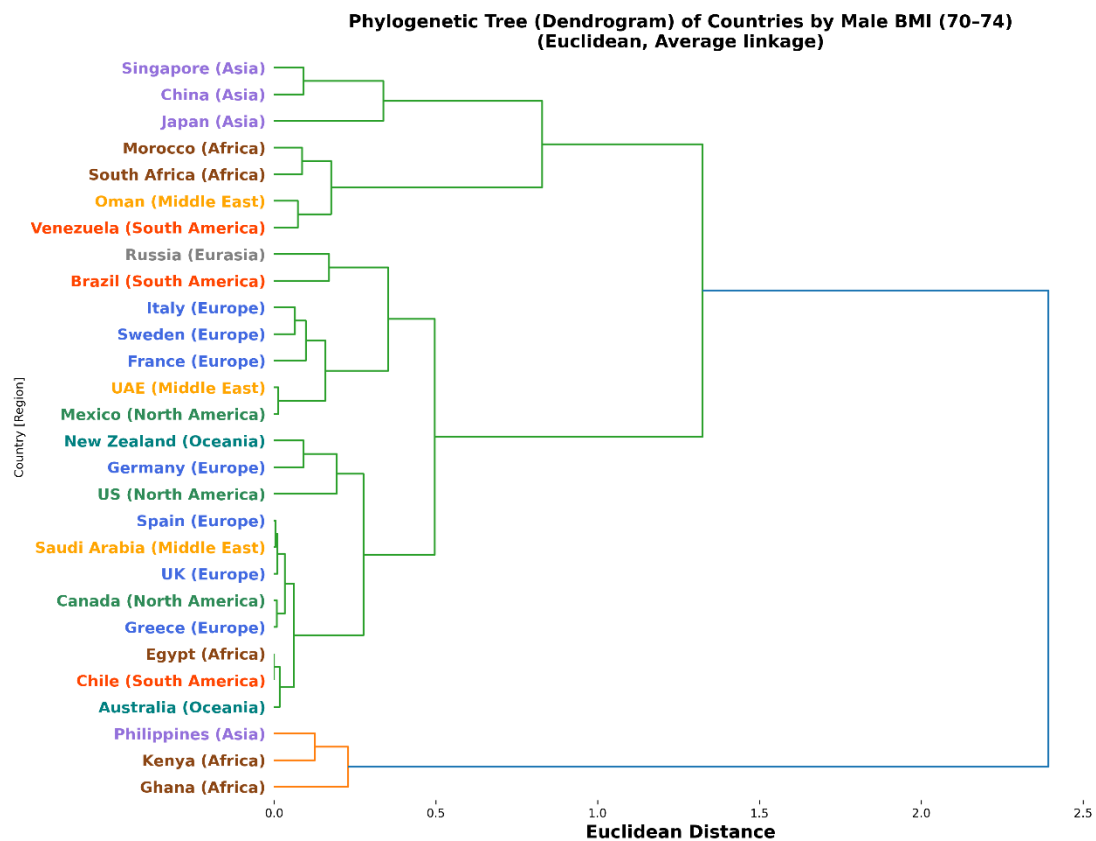


Figure 1.1.3-61. Phylogenetic Tree of Countries by Male Mean BMI (70-74) (Average Linkage + Euclidean Distance)

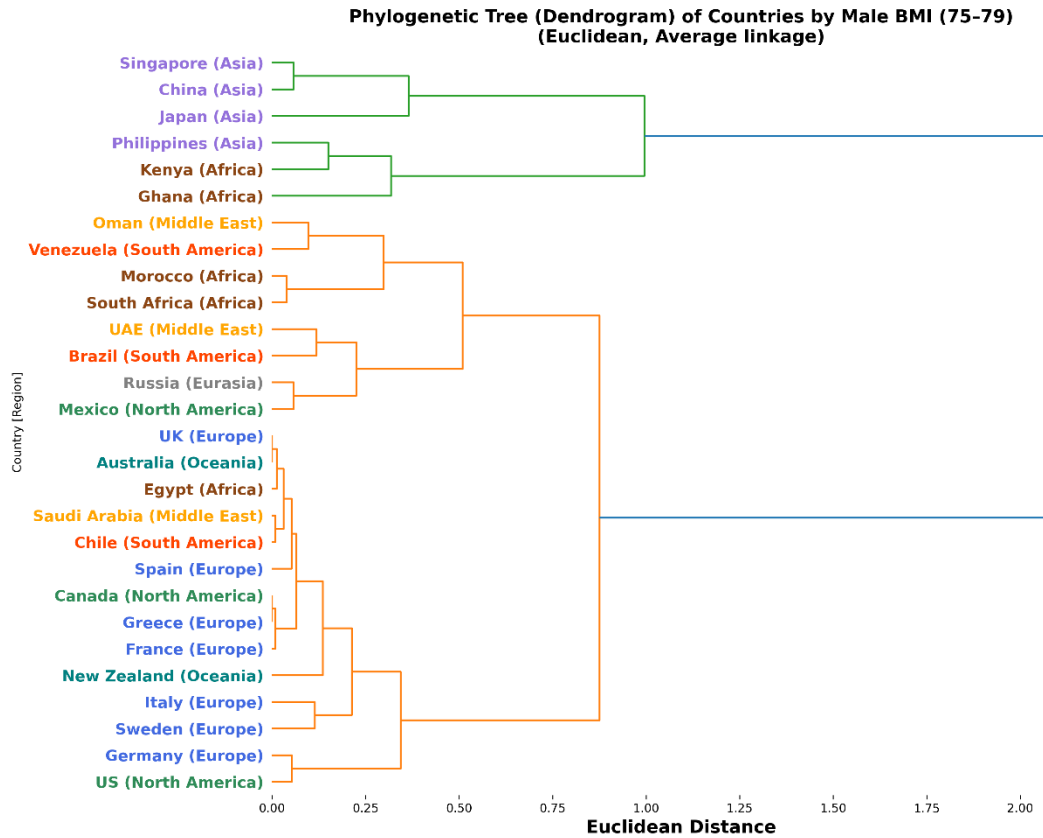


Figure 1.1.3-62. Phylogenetic Tree of Countries by Male Mean BMI (75-79) (Average Linkage + Euclidean Distance)

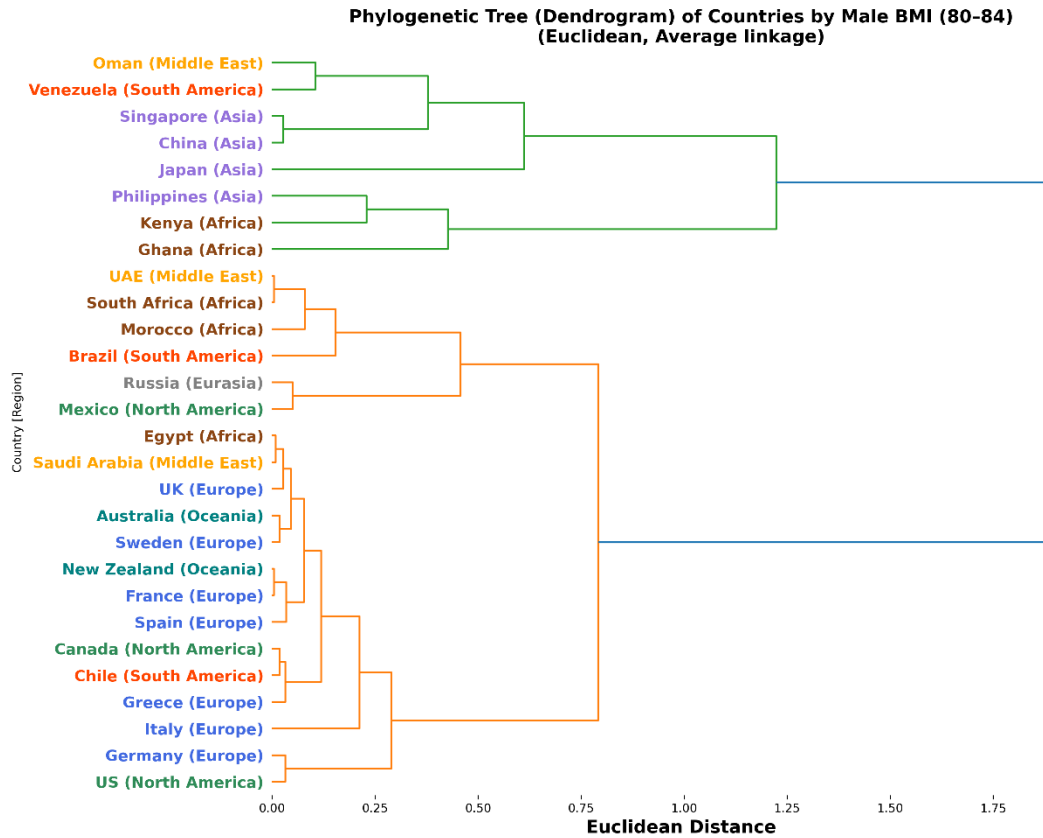


Figure 1.1.3-63. Phylogenetic Tree of Countries by Male Mean BMI (80-84) (Average Linkage + Euclidean Distance)

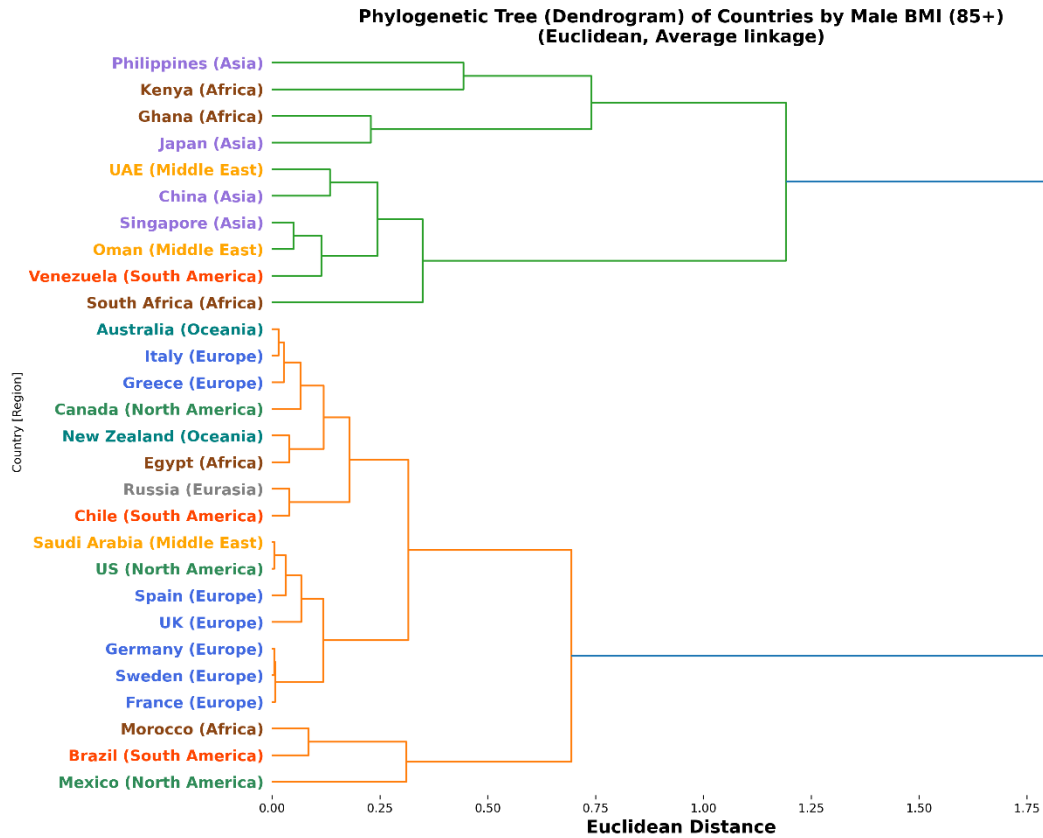


Figure 1.1.3-64. Phylogenetic Tree of Countries by Male Mean BMI (85+) (Average Linkage + Euclidean Distance)

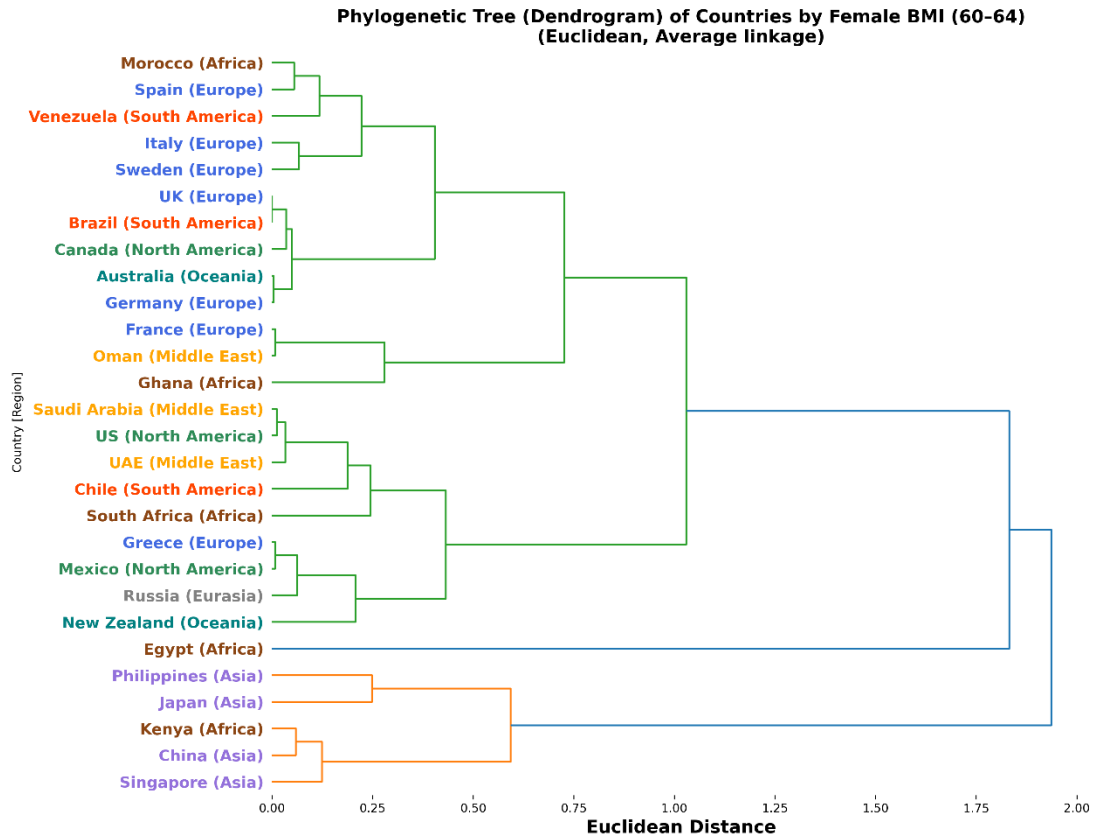


Figure 1.1.3-65. Phylogenetic Tree of Countries by Female Mean BMI (60-64) (Average Linkage + Euclidean Distance)

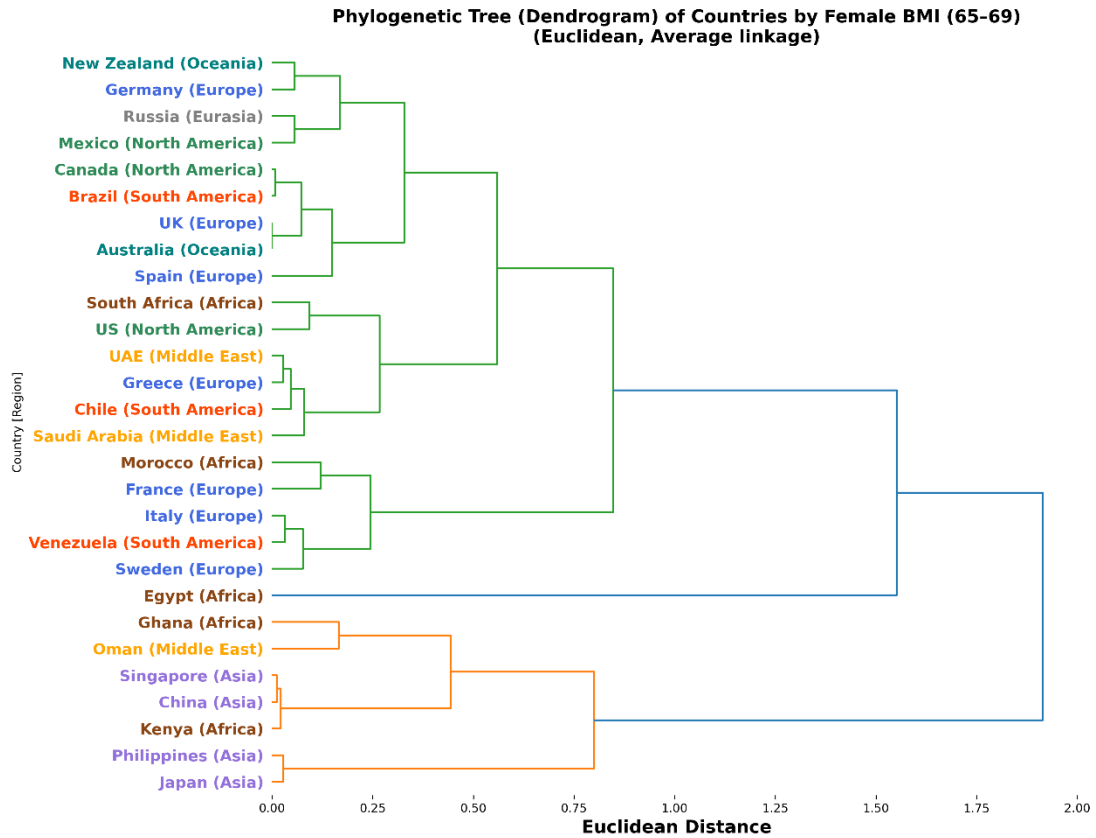


Figure 1.1.3-66. Phylogenetic Tree of Countries by Female Mean BMI (65-69) (Average Linkage + Euclidean Distance)

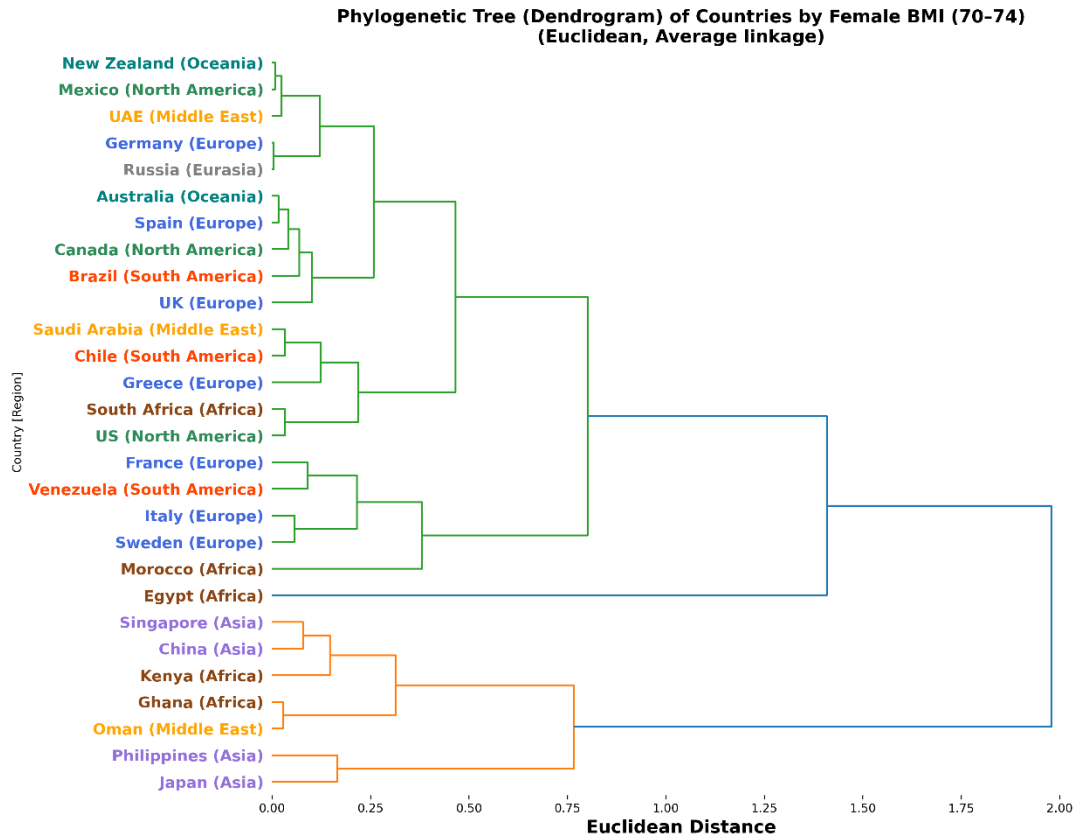


Figure 1.1.3-67. Phylogenetic Tree of Countries by Female Mean BMI (70-74) (Average Linkage + Euclidean Distance)

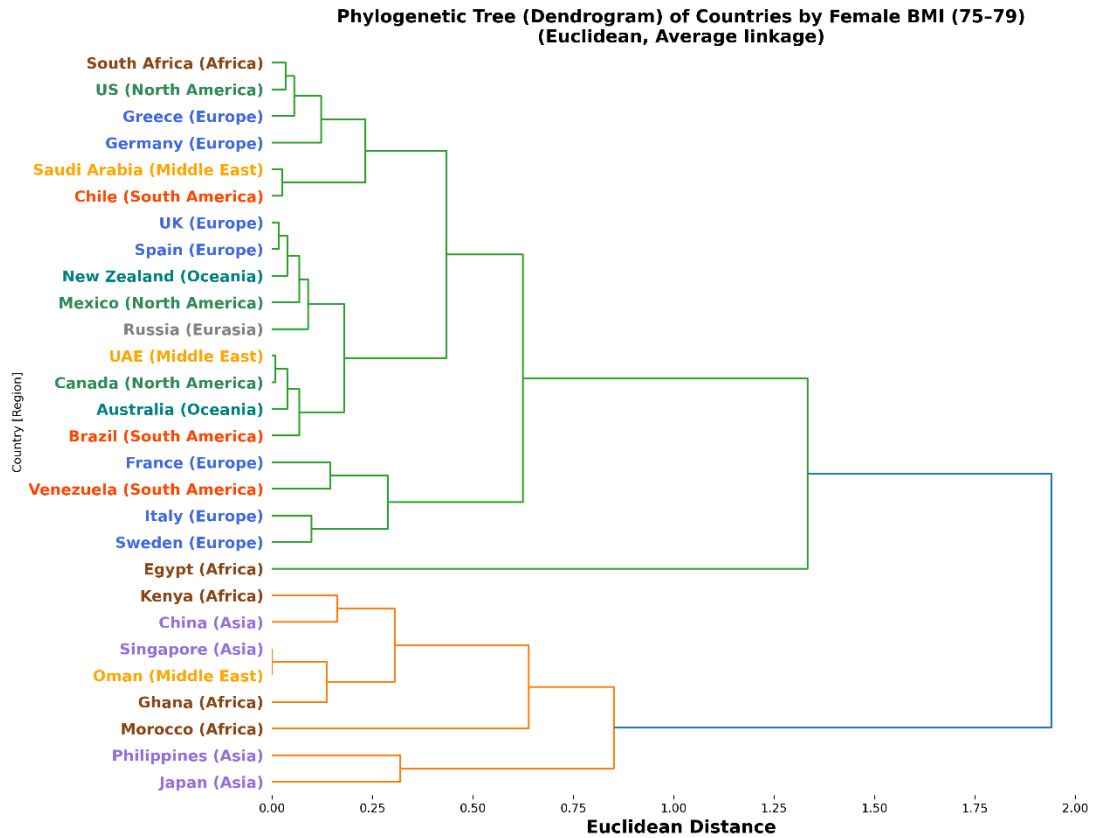


Figure 1.1.3-68. Phylogenetic Tree of Countries by Female Mean BMI (75-79) (Average Linkage + Euclidean Distance)

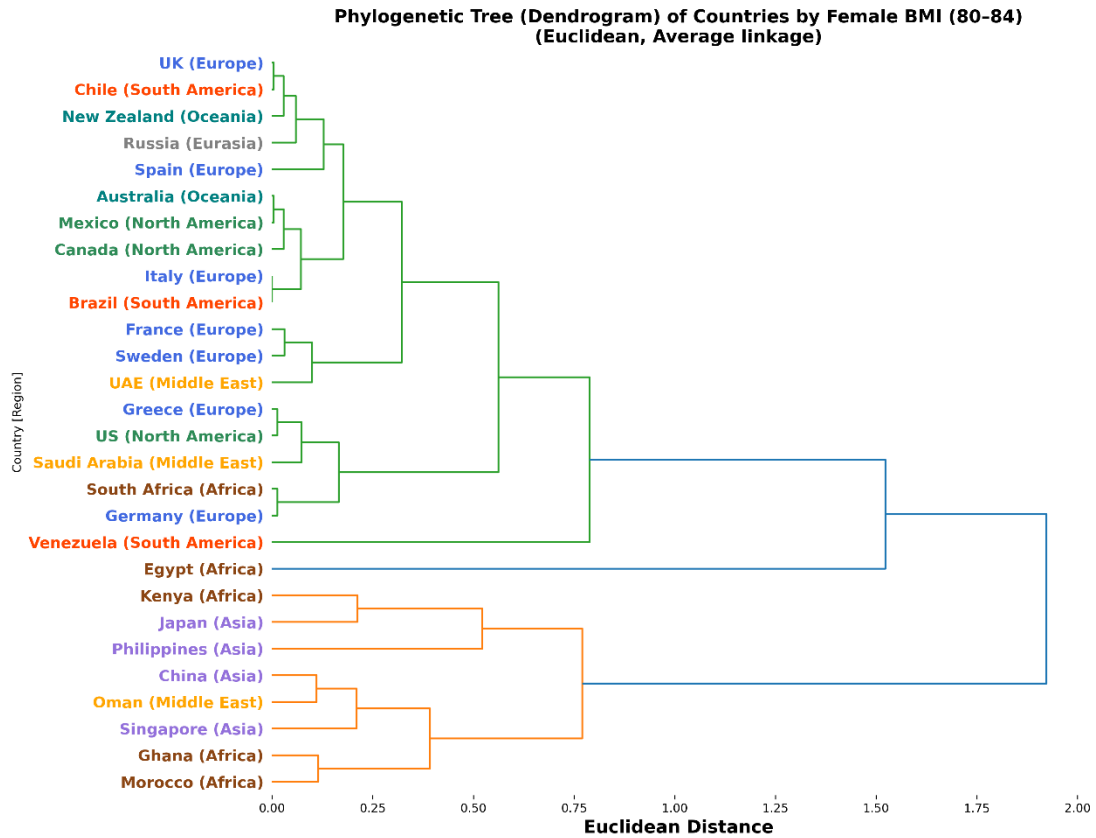


Figure 1.1.3-69. Phylogenetic Tree of Countries by Female Mean BMI (80-84) (Average Linkage + Euclidean Distance)

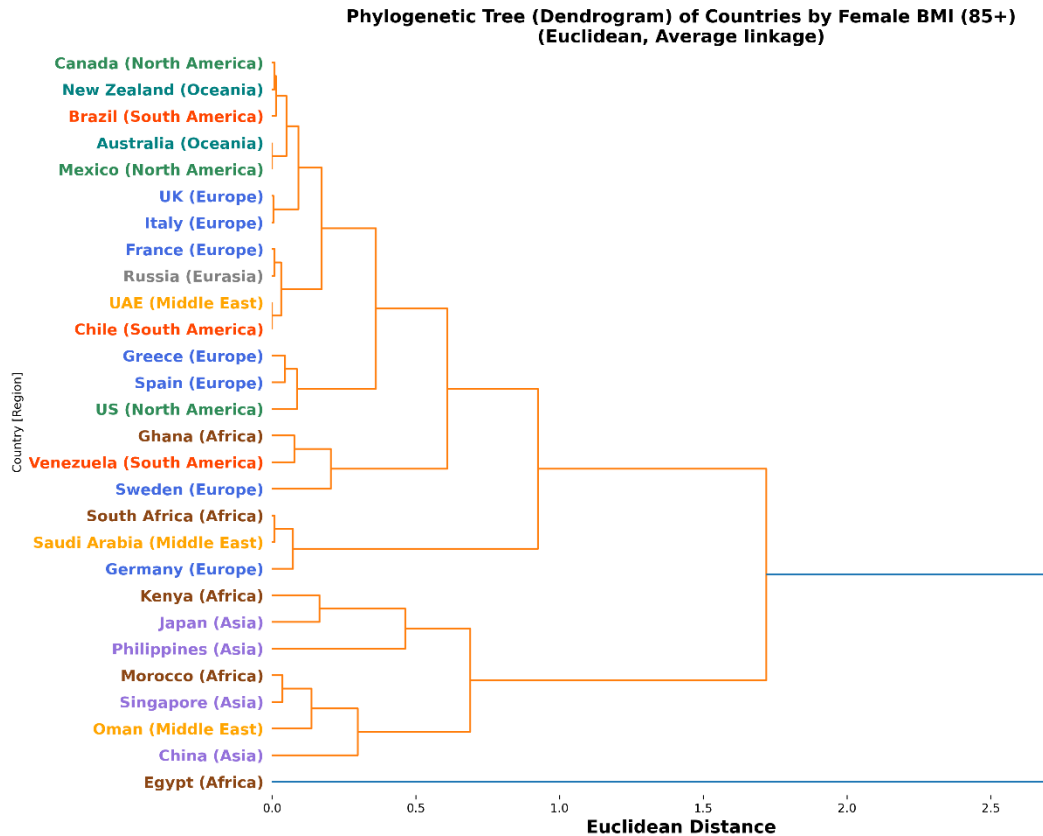


Figure 1.1.3-70. Phylogenetic Tree of Countries by Female Mean BMI (85+) (Average Linkage + Euclidean Distance)

Multidimensional Dendrograms by Sex for All BMI Age Groups

Settings. Ward + Euclidean; Average + Euclidean; Average + Correlation; Average + Manhattan (Cityblock); Average + Cosine; Average + Mutual Information Distance

(Top-5).

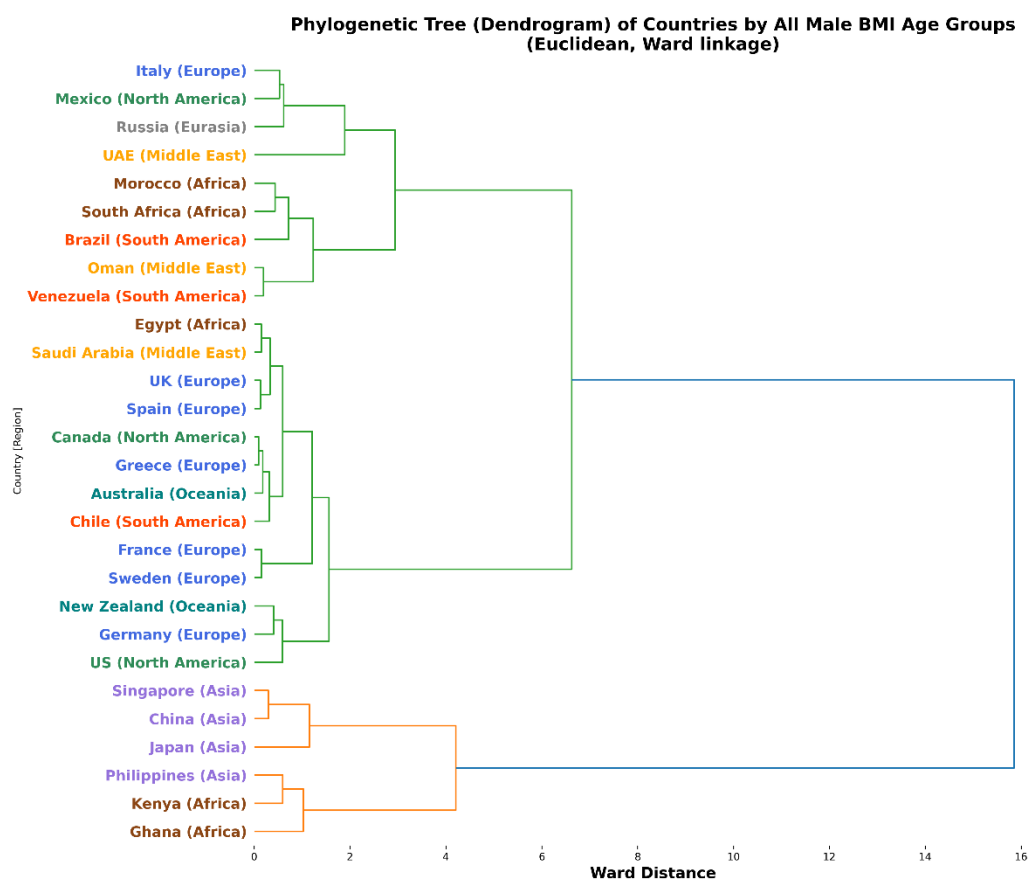


Figure 1.1.3-71. Phylogenetic Tree of Countries by All Male Mean BMI Age Groups (Ward Linkage + Euclidean Distance)

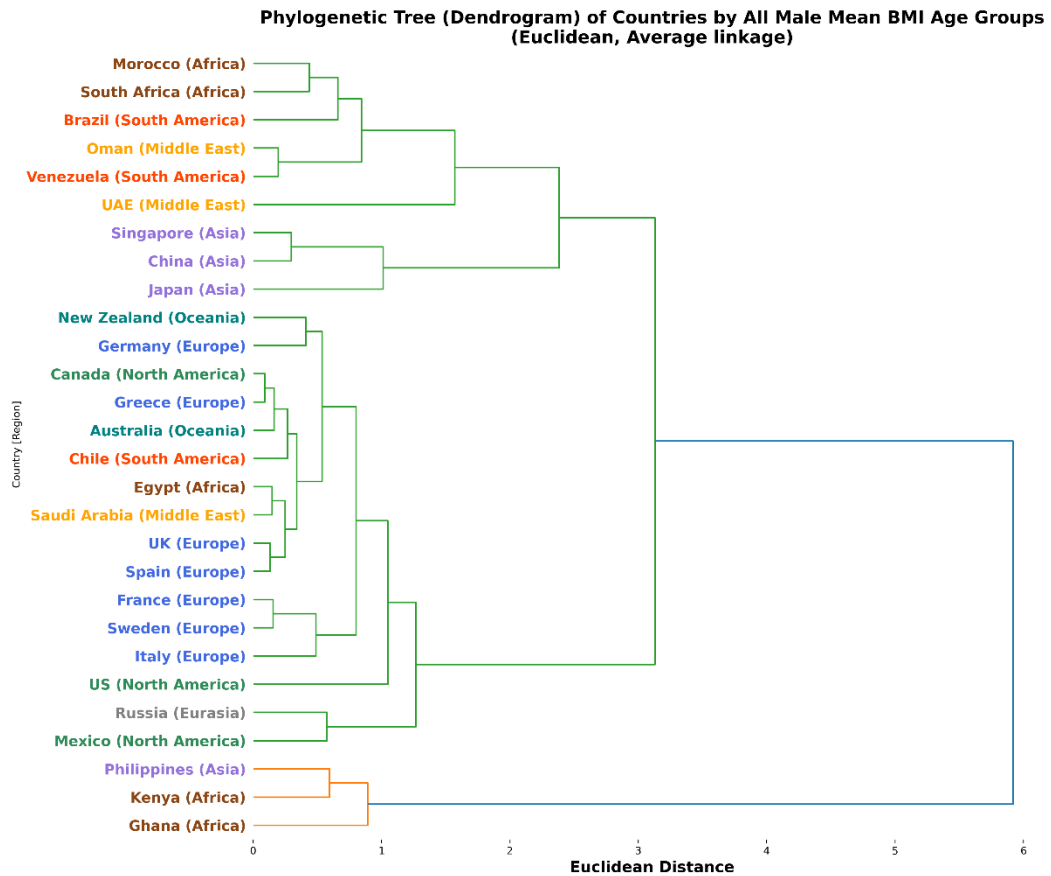


Figure 1.1.3-72. Phylogenetic Tree of Countries by All Male Mean BMI Age Groups (Average Linkage + Euclidean Distance)

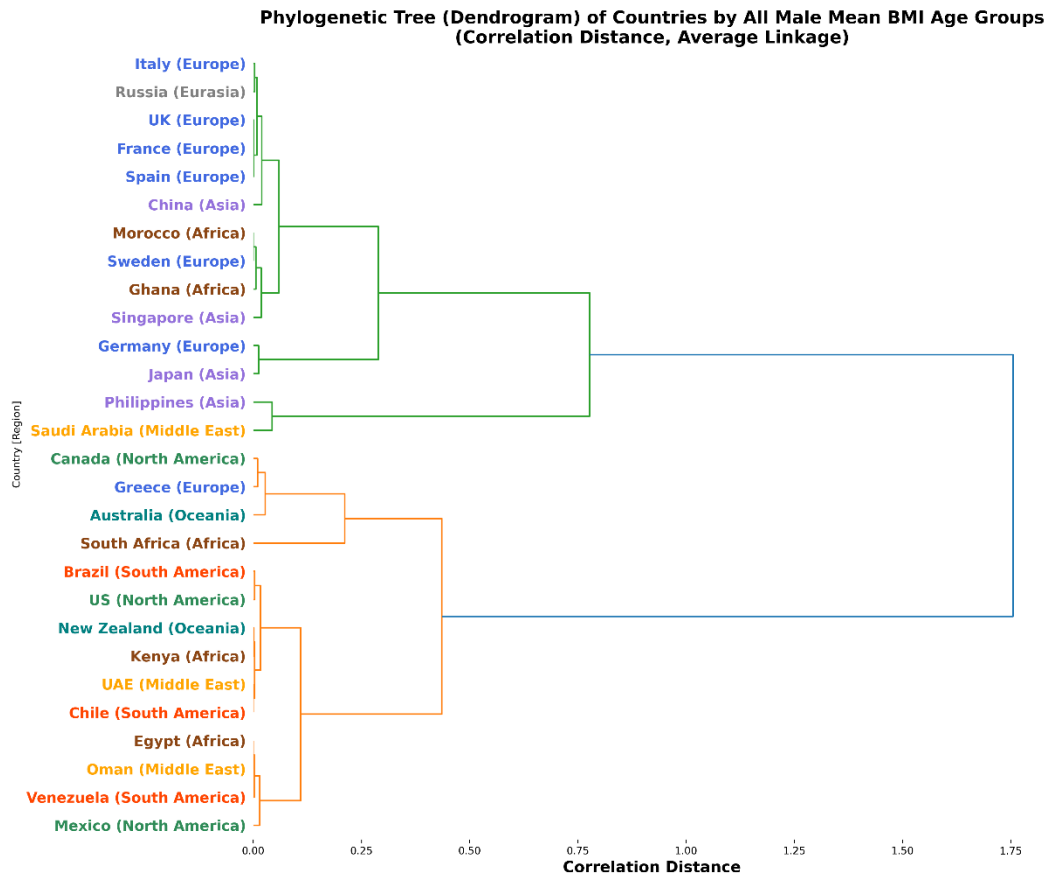


Figure 1.1.3-73. Phylogenetic Tree of Countries by All Male Mean BMI Age Groups (Average Linkage + Correlation Distance)

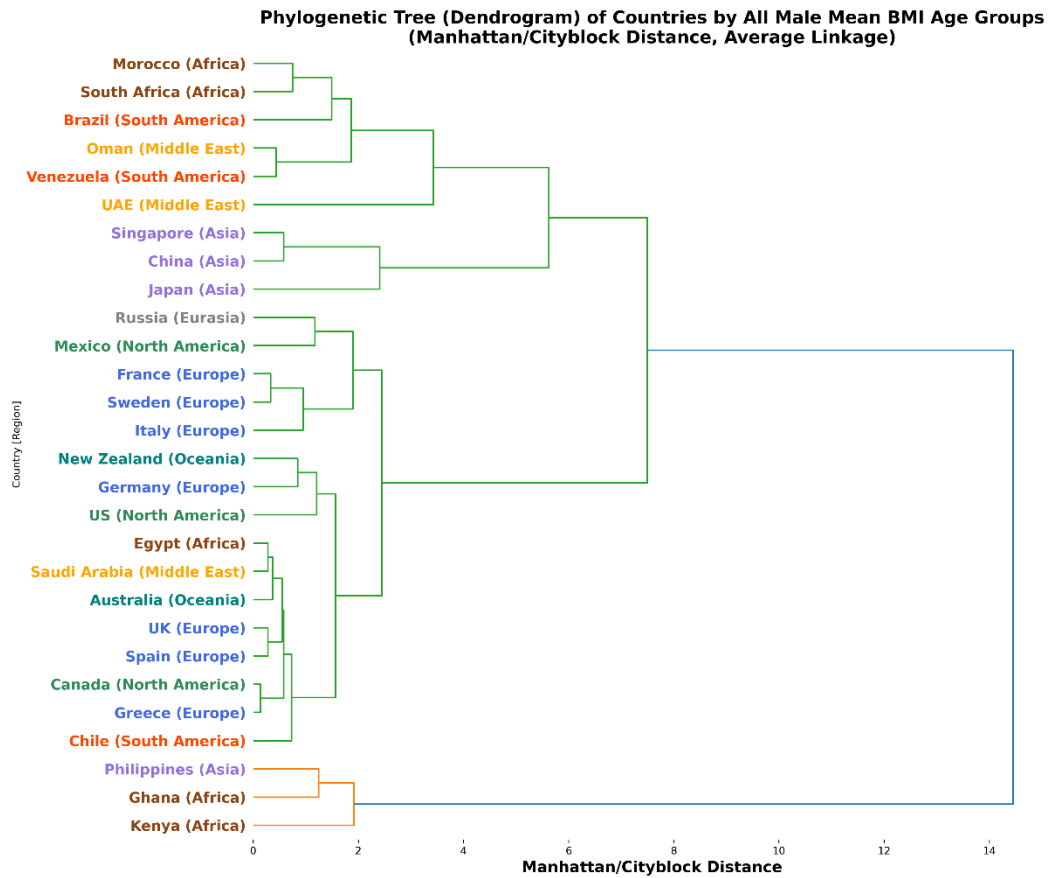


Figure 1.1.3-74. Phylogenetic Tree of Countries by All Male Mean BMI Age Groups (Average Linkage + Manhattan Distance)

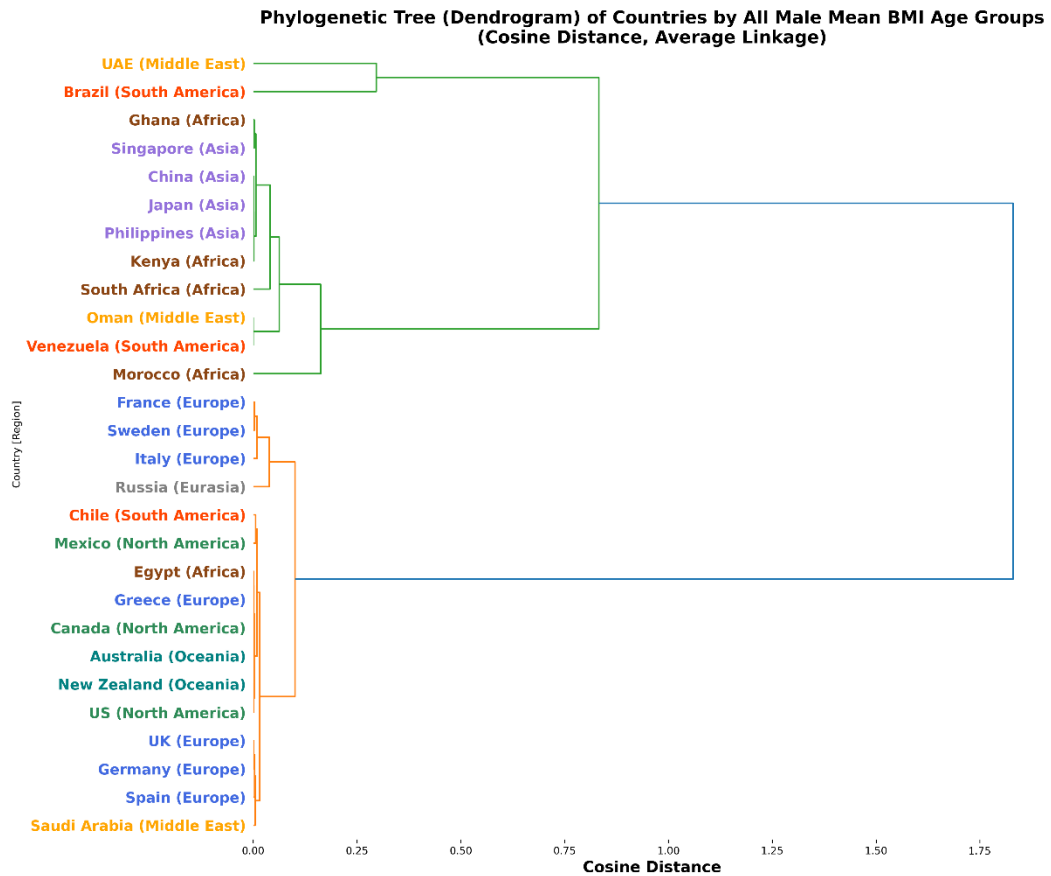


Figure 1.1.3-75. Phylogenetic Tree of Countries by All Male Mean BMI Age Groups (Average Linkage + Cosine Distance)

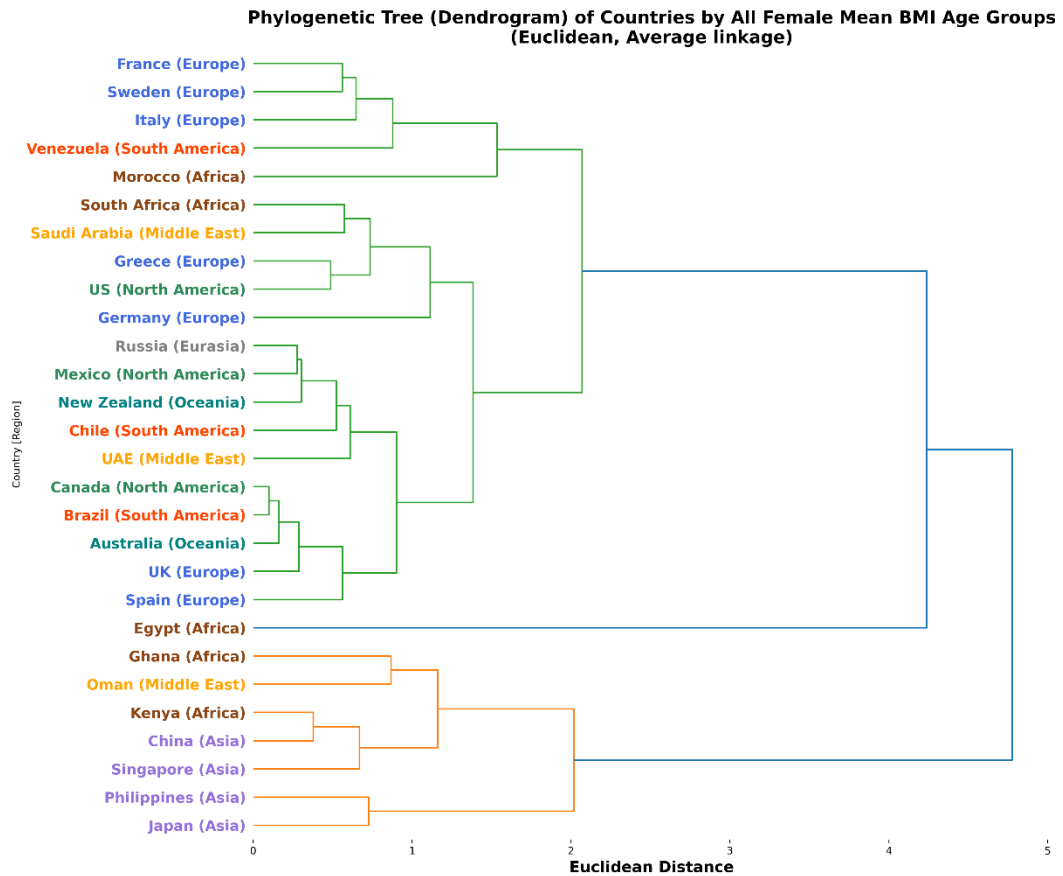


Figure 1.1.3-77. Phylogenetic Tree of Countries by All Female Mean BMI Age Groups (Average Linkage + Euclidean Distance)

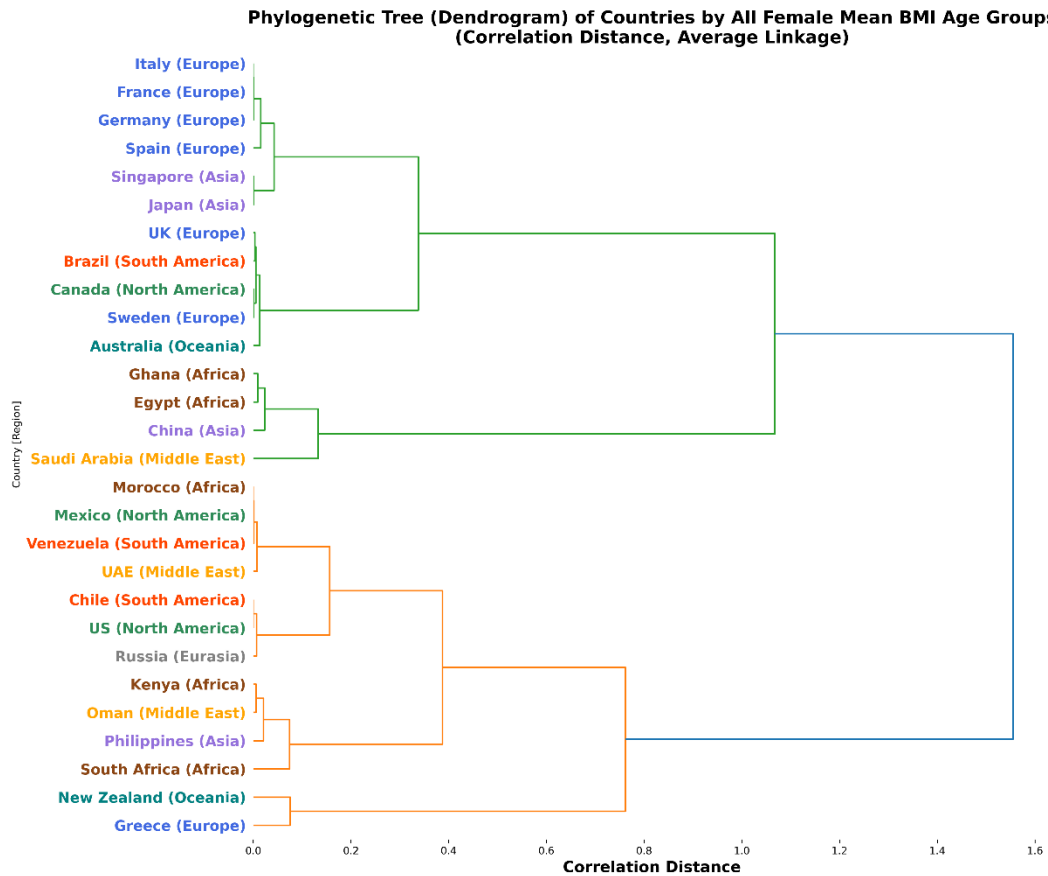


Figure 1.1.3-78. Phylogenetic Tree of Countries by All Female Mean BMI Age Groups (Average Linkage + Correlation Distance)

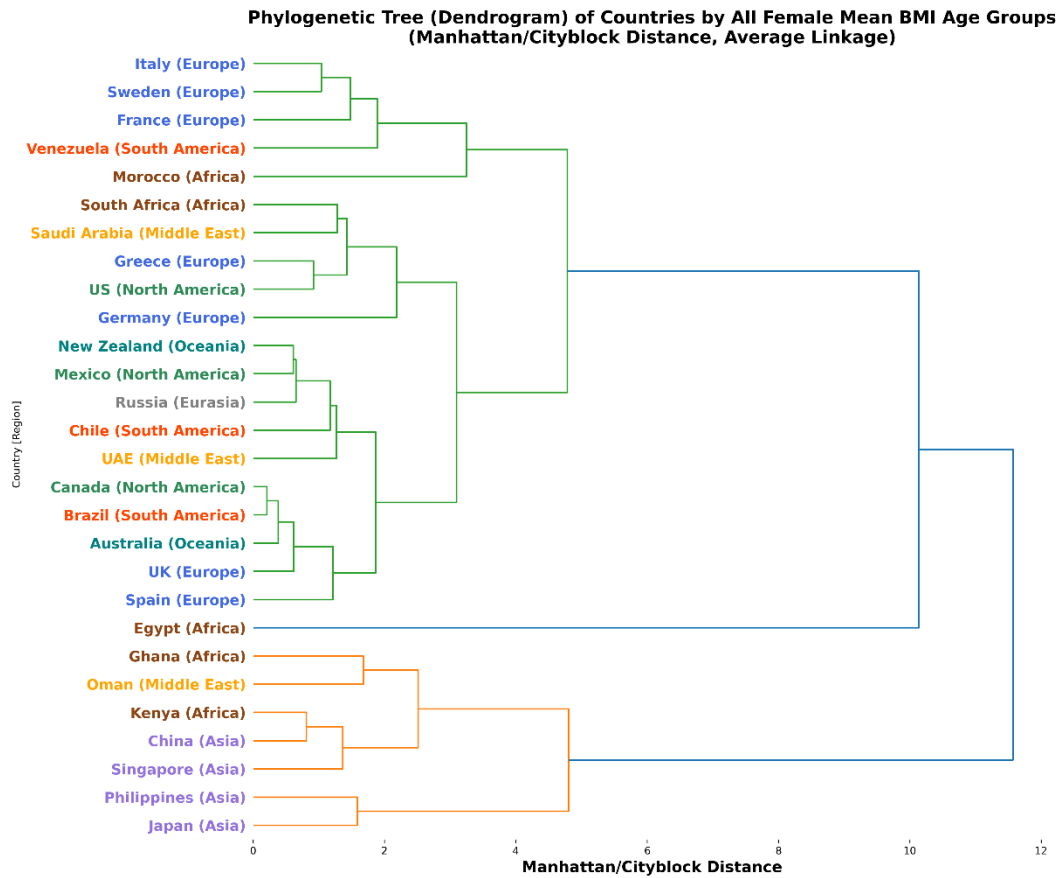


Figure 1.1.3-79. Phylogenetic Tree of Countries by All Female Mean BMI Age Groups (Average Linkage + Manhattan Distance)

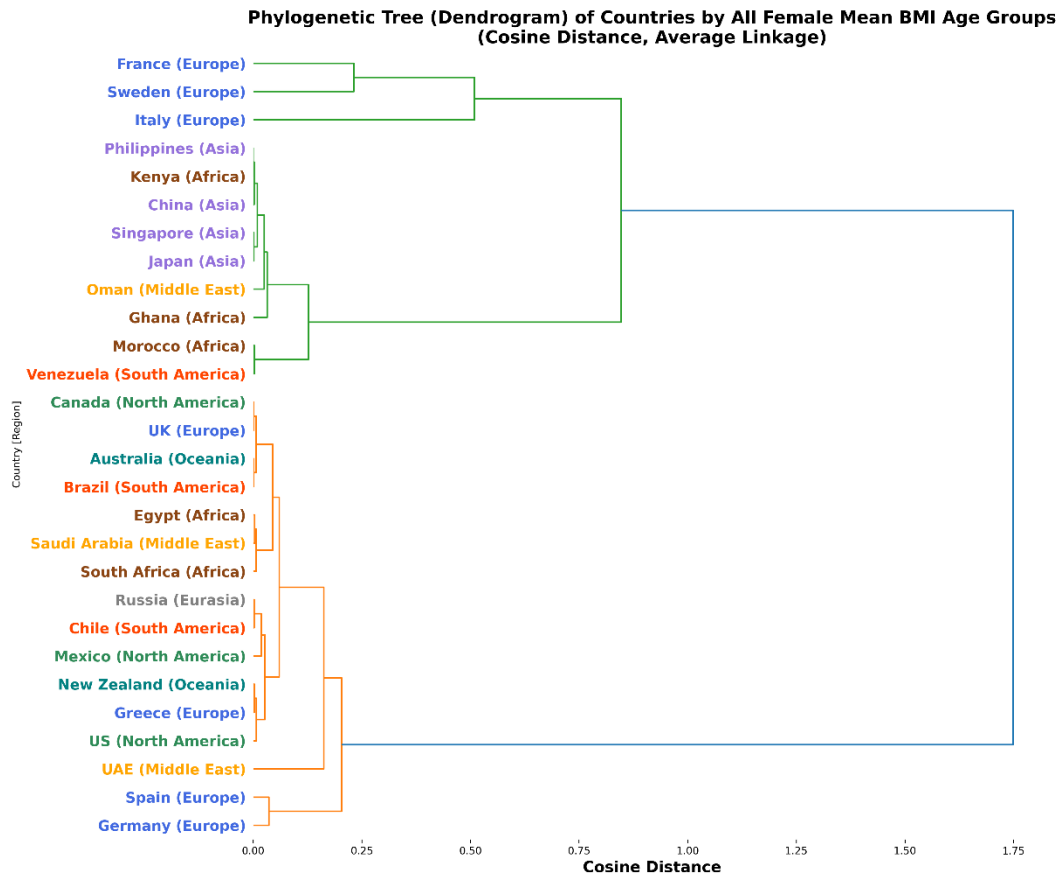


Figure 1.1.3-80. Phylogenetic Tree of Countries by All Female Mean BMI Age Groups (Average Linkage + Cosine Distance)

UMAP Visualizations: Male vs. Female

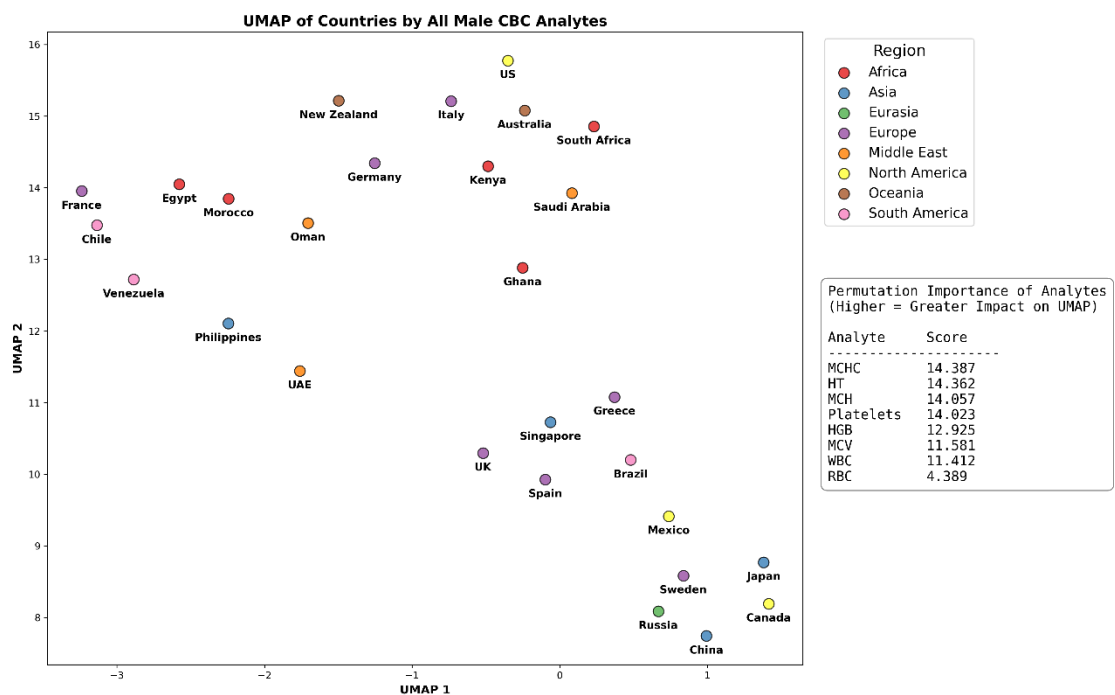


Figure 1.1.3-81. UMAP Visualization and Permutation Importance of All Male CBC Analytes

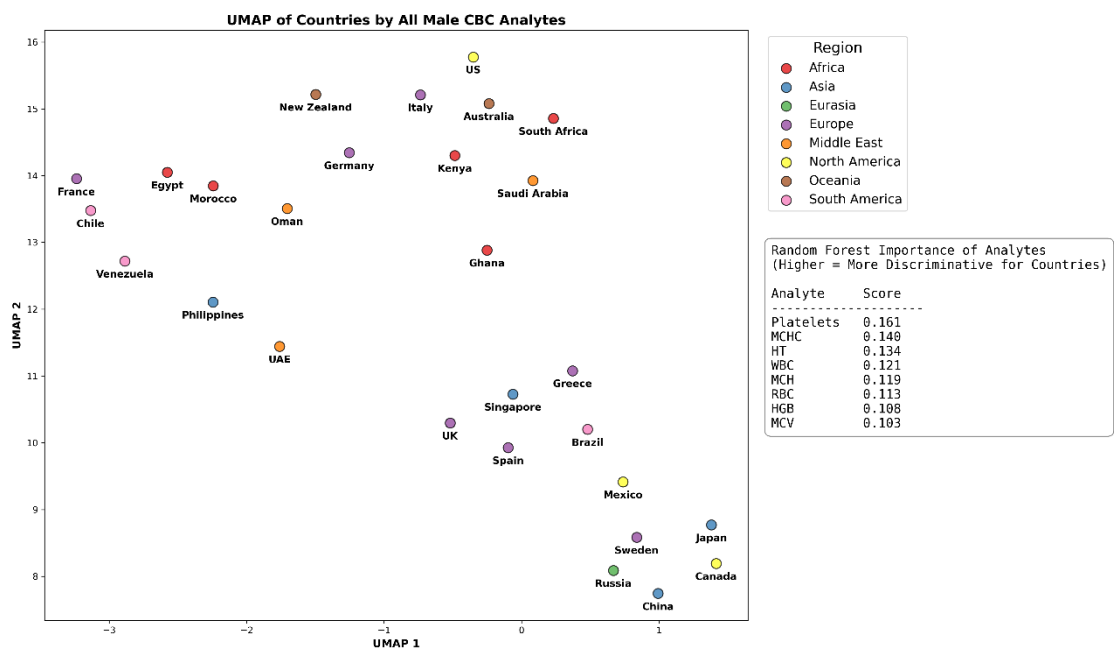


Figure 1.1.3-82. UMAP Visualization and Random Forest Importance of All Male CBC Analytes

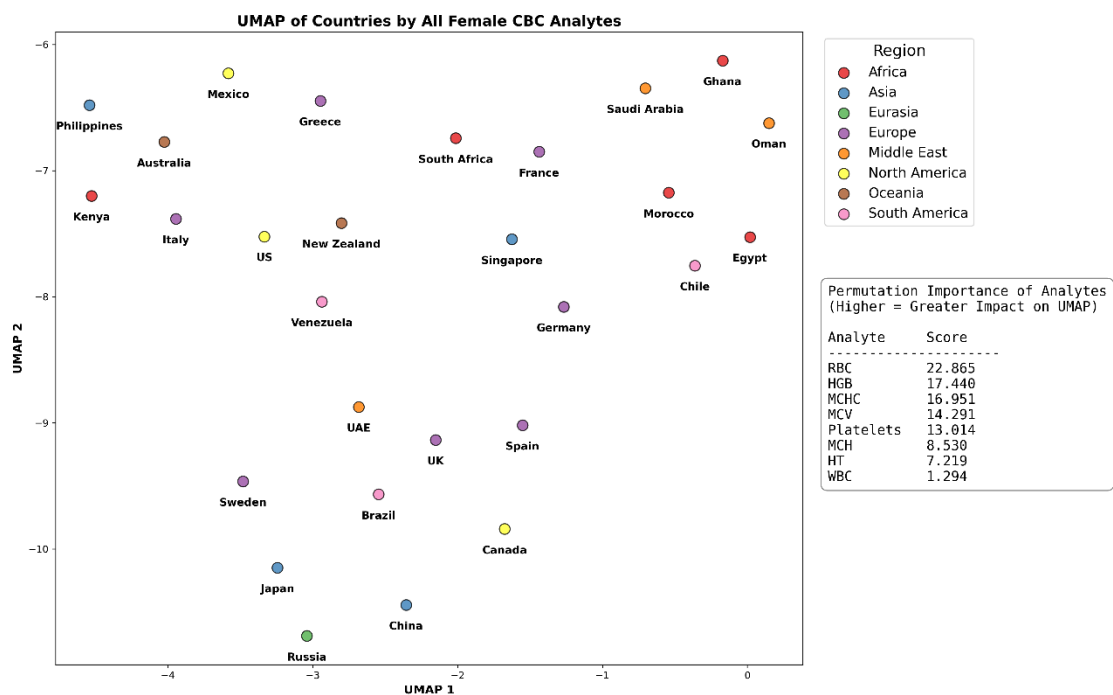


Figure 1.1.3-83. UMAP Visualization and Permutation Importance of All Female CBC Analytes

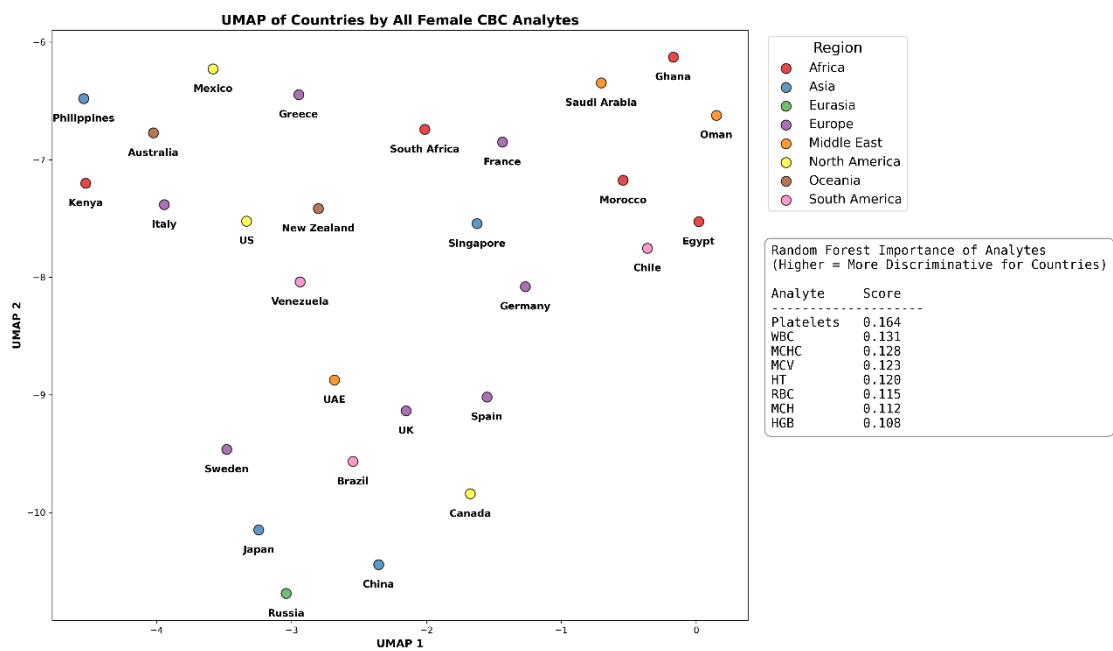


Figure 1.1.3-84. UMAP Visualization and Random Forest Importance of All Female CBC Analytes

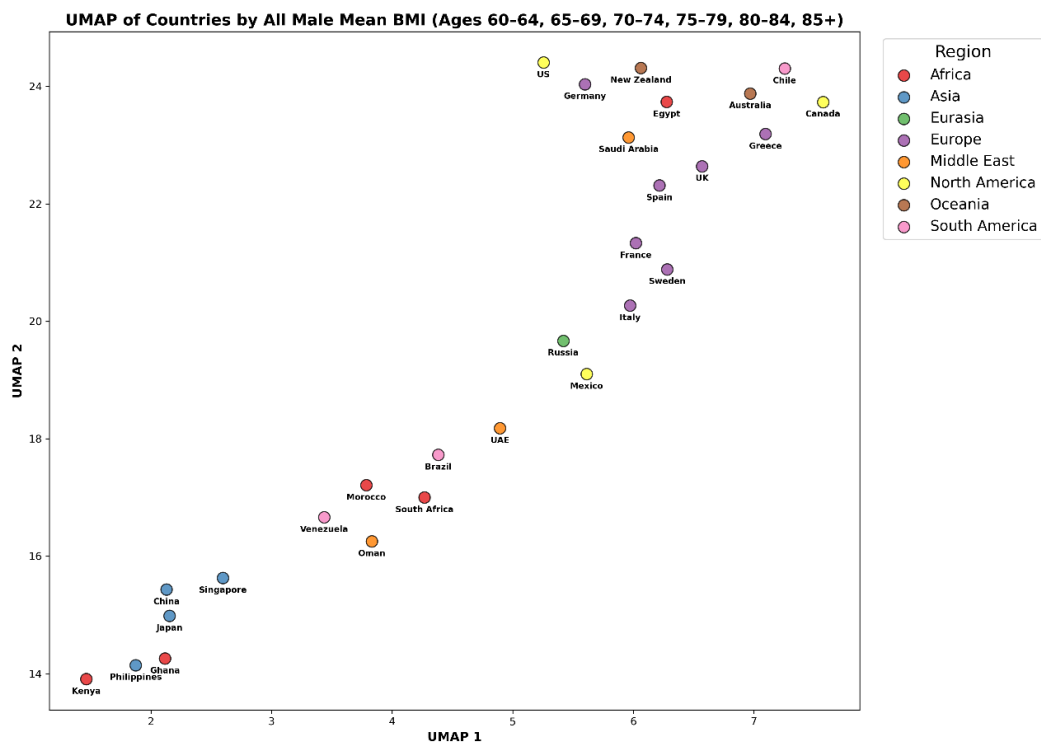


Figure 1.1.3-85. UMAP Visualization of All Male Mean BMI Age Groups

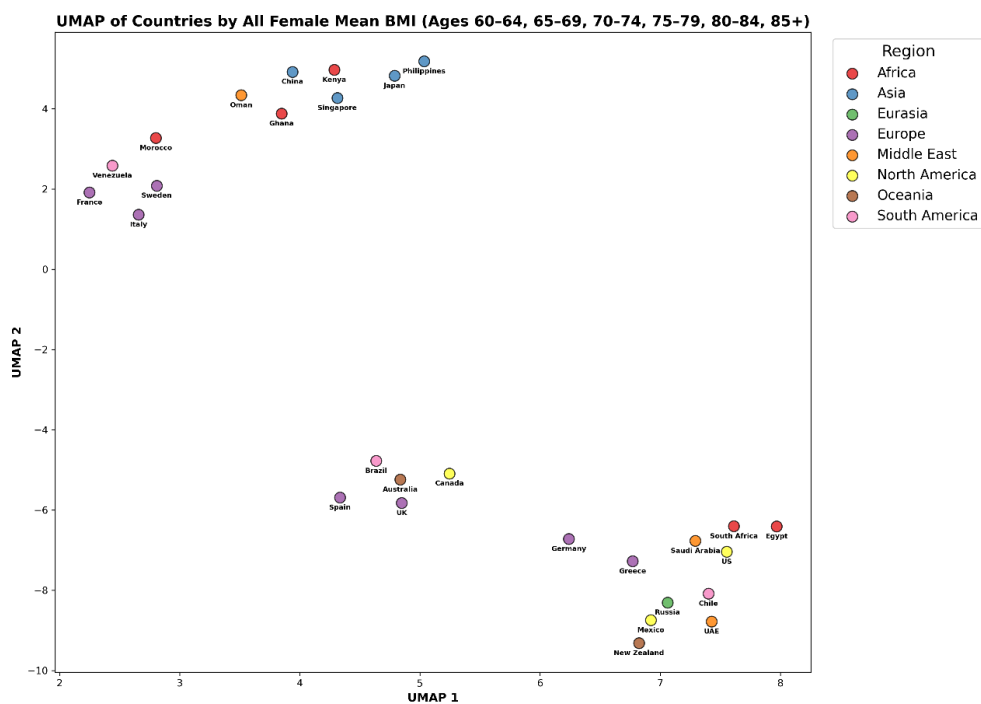


Figure 1.1.3-86. UMAP Visualization of All Female Mean BMI Age Groups