Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predominantly Anthroponotic Spread and Divergent Giardia Assemblages Drive Co-Circulating Lineages in Pediatric infection from Egypt

Doaa Naguib
Beltsville Agricultural Research Center, Agricultural Research Service, United States Department of Agriculture

Benjamin M. Rosenthal
Beltsville Agricultural Research Center, Agricultural Research Service, United States Department of Agriculture

Asis Khan
asis.khan@usda.gov

Beltsville Agricultural Research Center, Agricultural Research Service, United States Department of Agriculture

Article

Keywords: Giardia, Assemblage A, Assemblage B, SSU rRNA, tpi, bg, gdh, population genetic, genetic diversity, and recombination

Posted Date: June 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9772554/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on September 27th, 2026. See the published version at https://doi.org/10.1038/s41598-026-72912-1.

Abstract

Giardia duodenalis is a zoonotic protozoan causing substantial waterborne diarrheal disease, yet its genetic diversity and population structure remain under-characterized in many endemic regions. This study examined 451 diarrheic children using qPCR targeting the *SSU rRNA* gene and multi-locus genotyping of the *tpi*, *gdh, and bg* loci to assess prevalence, assemblage distribution, and population-genetic patterns. *G. duodenalis* was detected in 10.64% of children, with significantly greater prevalence in those aged 1–5 years. Assemblages A and B occurred at nearly equal frequencies, and all assemblage A isolates belonged to the human-adapted AII sub- assemblage, while assemblage B comprised diverse BIII/BIV variants. No mixed infections were detected, and animal contact did not elevate infection risk, suggesting predominantly anthroponotic transmission. Multi-locus analyses revealed striking contrasts between assemblages: assemblage A exhibited extremely low nucleotide and haplotype diversity, complete linkage disequilibrium, absence of recombination, and a highly clonal structure, whereas assemblage B displayed substantially greater polymorphism, deeper sub-structuring, elevated haplotype richness, and reticulated phylogenetic networks. AMOVA demonstrated near-complete genetic differentiation between assemblages. These findings show that assemblages A and B follow distinct evolutionary trajectories and that giardiasis transmission in this region is driven overwhelmingly by human-to-human spread. The strong anthroponotic signal, coupled with the high burden in children aged 1–5 and the absence of any association with animal contact, indicates that public health efforts should prioritize interrupting transmission among young children, particularly in household and childcare settings.

Introduction

Giardiasis is a waterborne disease caused by the flagellated protozoan *Giardia duodenalis* (syn. *G. intestinalis*, *G. lamblia*), one of the most common parasitic enteric pathogens infecting humans and animals worldwide. These parasites pose a persistent public health concern, particularly in areas with inadequate sanitation, and are a major contributor to global waterborne disease outbreaks [1]. Clinical presentations range from asymptomatic infection to acute or chronic gastrointestinal symptoms such as diarrhea, abdominal pain, flatulence, and bloating [2]. Asymptomatic infections are also common, particularly in developing countries [3]. In the United States, approximately 1.2 million cases occur each year [4]. Globally, an estimated 280 million infections occur annually, with prevalence ranging from 0.4–7.5% in high-income countries to 8–30% in low- and middle-income regions [5]. *G. duodenalis* is typically transmitted through the fecal–oral route via contaminated water or food, or through direct contact with infected individuals or animals [6]. Children, travelers, immunocompromised individuals, and older adults are particularly vulnerable [1]. Importantly, *Giardia* is a well-recognized cause of traveler's diarrhea, contributing to infections among individuals returning from endemic areas [1,7]. In young children, infection can impair growth and cognitive development, with long-lasting effects [8]. Despite its widespread impact, surveillance systems, including those in the United States, remain limited in their ability to detect outbreaks, monitor shifts in transmission, and assess changes in the parasite's population structure [7].

G. duodenalis exhibits extensive genetic diversity, comprising eight assemblages (A–H). Assemblages A and B are zoonotic, accounting for most human infections; the remaining assemblages show strong host-specificity [1,9]. Molecular markers commonly used for detection and genotyping include the small subunit ribosomal RNA (*SSU-rRNA*), triose phosphate isomerase (*tpi*), glutamate dehydrogenase (*gdh*), and β-giardin (*bg*) genes [5]. Although *SSU-rRNA* is conserved and multicopy, its short length and high degree of conservation limit its usefulness in differentiating among isolates. In contrast, single-copy loci such as *tpi, gdh, and bg*, used in multi-locus sequence typing (MLST), provide higher resolution for phylogenetic analysis [10]. MLST typing method further subdivides the Assemblage A into AI, AII, and AIII, each comprising 9–12 sequence types or genotypes per locus [5,11]. AI is found predominantly in animals, AII is mainly associated with human infections, and AIII is primarily detected in wild ruminants [12,13]. Assemblage B, which includes BIII and BIV, displays extensive allelic heterogeneity and frequent ambiguous chromatogram peaks, complicating subtype classification and molecular epidemiological efforts [14]. This high allelic sequence heterozygosity (ASH) remains a major challenge for outbreak investigations and genotyping [15].

Population-level analyses have revealed distinct evolutionary patterns between assemblages A and B. Assemblage A tends to form a conserved, compact cluster encompassing only limited sequence variation, while assemblage B exhibits deep branching, high haplotype diversity, and reticulated haplotype networks (suggestive of recombination or mixed infections) [14]. AMOVA results frequently report high ΦST values between assemblages, indicating substantial genetic differentiation consistent with deep evolutionary divergence [1]. Despite these findings, the population structure of *G. duodenalis* remains only partly resolved.

Population genetics provides a powerful framework for understanding how evolutionary forces shape parasite diversity [16]. Natural parasite populations are often subdivided across geographic regions, host species, and even individual hosts. Genetic structure reflects and enables recognition of such epidemiological subdivisions [17–19]. However, muti-locus datasets remain limited in many endemic regions, including destinations for US travelers, such as Egypt, constraining efforts to fully characterize *Giardia* transmission and evolutionary patterns.

To address these gaps, we integrated prevalence assessment with multi-locus genotyping, phylogenetic reconstruction, Bayesian clustering, haplotype network analysis, and AMOVA to examine genetic differentiation, sub-structuring, admixture, and population-level variation in *Giardia*-positive samples from children in northeastern Egypt. Specifically, we hypothesized that infections in this setting are shaped predominantly by anthroponotic rather than zoonotic transmission, leading to distinct lineage patterns among young children. By combining molecular epidemiology with population genetics, we sought to advance understanding of *G. duodenalis* evolutionary dynamics to support more effective public health strategies, particularly by strengthening surveillance and traveler-related outbreak detection in the United States.

Results

Prevalence of G. duodenalis in diarrheic children

Based on qPCR targeting the *SSU rRNA* gene, amplified in 48 of 451 children (10.64%) of children examined in northeastern Egypt (Table 1). Prevalence was significantly higher among children 1–5 years of age (14.86%) than those aged 5–10 years (5.45%) ($\chi^2 = 14.86$, $P = 0.0001$). No significant differences were observed between males (11.21%) and females (10.09%) ($\chi^2 = 0.302$, $P = 0.582$), or between rural (10.75%) and urban residents (10.47%) ($\chi^2 = 0.043$, $P = 0.835$). Children without animal contact had a slightly higher, but non-significant, prevalence (11.86%) compared with those reporting animal contact (9.09%) ($\chi^2 = 1.271$, $P = 0.260$). Prevalence varied significantly by governorate, with the highest rates observed in Sharqia (17.24%), followed by Damietta (14.67%), Dakahliya (8.72%), and Gharbia (4.23%) ($\chi^2 = 10.23$, $P = 0.017$). Significant seasonal variation was also detected, with the highest prevalence in November (26.92%) and the lowest in May (2.70%) ($\chi^2 = 42.23$, $P = 0.001$).

Table 1

Occurrence of *G. duodenalis* in diarrheic children using the *SSUrRNA* [55] by age, sex, residency, animal contact, locality, and month of collection

Factor	Category	No. of samples	No. of positive samples (%)	X^2	*P* - value
Age	1–5 years	249	37 (14.86)	14.86	0.0001
	5–10 years	202	11 (5.45)		
Sex	Female	228	23 (10.09)	0.302	0.582
	Male	223	25 (11.21)		
Residency	Rural	279	30 (10.75)	0.043	0.835
	Urban	172	18 (10.47)		
Animal contact	No	253	30 (11.86)	1.271	0.260
	Yes	198	18 (9.09)		
Locality	Dakahliya	218	19 (8.72)	10.23	0.017
	Sharqia	87	15 (17.24)		
	Damietta	75	11 (14.67)		
	Gharbia	71	3 (4.23)		
Month of collections	May	37	1 (2.70)	42.23	0.001
	June	74	3 (4.05)		
	July	169	8 (4.73)		
	August	50	8 (16.00)		
	September	27	7 (25.93)		
	October	41	7 (17.07)		
	November	26	7 (26.92)		
	December	27	7 (25.93)		
Total	-	451	48 (10.64)		

G. duodenalis assemblages

Presence of two G. duodenalis assemblages in northeastern Egypt

Out of the 48 *SSU rRNA* qPCR-positive samples, 47 were successfully amplified at one or more loci (**Table S1**). At the *tpi* locus, 20 isolates (42.6%) were classified as assemblage A and 27 isolates (57.4%) as assemblage B. Assemblage A sequences belonged to sub-assemblage AII and clustered with the A2

reference strain (U57897), whereas assemblage B sequences grouped with BIII/BIV references (KM926535, AY368165, MT427472) (**Table S2**). At the *bg* locus, sequences were obtained from 46 samples, with 20 isolates (43.5%) identified as assemblage A (classified as sub-assemblage AII, including 5 A2 and 15 A3) and 26 isolates (56.5%) defined as assemblage B, clustering with reference sequences: AY647266, AY072727, and AY072728. Two samples failed to amplify (**Table S3**). At the *gdh* locus, 44 samples yielded sequences; 20 isolates (45.5%) were classified as assemblage A (sub-assemblage AII, subtype A2), clustering with reference sequence AY178737, while 24 isolates (54.5%) were classified as assemblage B, grouping with reference sequences MK952603, MG746607, and LC184466. Four samples did not amplify. No mixed infections between assemblages A and B were detected at any locus (**Table S4**).

Genetic diversity within assemblages

Genetic diversity analyses across the multi-locus dataset revealed consistent and pronounced differences between assemblages A and B at all loci. For example, the *tpi* locus in assemblage A exhibited complete genetic uniformity (S = 0, π = 0, Hd = 0, h = 1), whereas assemblage B showed markedly higher diversity, with seven variable sites, three parsimony-informative sites, five haplotypes, and elevated nucleotide diversity (π = 0.00164; Hd = 0.339).

A similar pattern was observed at the *gdh* locus: assemblage A again displayed no detectable variation (S = 0, π = 0, Hd = 0, h = 1), while assemblage B demonstrated moderate diversity (S = 6, π = 0.00193, Hd = 0.692, h = 5).

At the *bg* locus, assemblage A showed low but measurable diversity (S = 2, π = 0.00109, Hd = 0.395, h = 2), in contrast to assemblage B, which exhibited higher polymorphism (S = 5, π = 0.00130, Hd = 0.569, h = 6) (Table 2).

Table 2
Diversity indices for the *tpi*, *gdh*, and *bg* loci in *G. duodenalis* assemblages A and B.

Loci	Assemblage	Sequence length (bp)	number of sequences analyzed (N)	Variable sites (S)	Parsimony -informative sites (PI)	nucleotide diversity (π)	Haplotypes (h)
tpi	A	503	20	0	0	0	1
	B	503	27	7	3	0.00164	5
gdh	A	735	20	0	0	0	1
	B	735	24	6	4	0.00193	5
bg	A	728	20	2	2	0.00109	2
	B	728	26	5	5	0.0013	6

Thus, across loci, heterozygosity differences (ΔH) consistently indicated reduced genetic diversity in assemblage A relative to assemblage B, with values of -0.339 for *tpi*, -0.692 for *gdh*, and − 0.174 for *bg* (Table 3). Collectively, these multi-locus patterns demonstrate extremely low genetic diversity in

assemblage A compared with the higher, structured diversity observed in assemblage B, supporting a predominantly clonal or near-clonal population structure in assemblage A.

Table 3
Haplotype diversity and heterozygosity differences (ΔH) across *tpi*, *gdh*, and *bg* loci in assemblages A and B

Loci	Haplotype diversity (Hd)		Heterozygosity differences
	Assemblage A	Assemblage B	(ΔH = H(A) – H(B))
tpi	0	0.339	-0.339
gdh	0	0.692	-0.692
bg	0.395	0.569	-0.174

Haplotype network analysis

The haplotype networks for the *tpi, gdh, and bg* loci revealed a strong genetic separation between *G. duodenalis* assemblage A and assemblage B (Fig. 1). Assemblage A formed a compact cluster dominated by a single major haplotype and a few closely related variants, reflecting low sequence divergence across all three loci. In contrast, assemblage B exhibited substantially higher haplotype diversity, comprising multiple haplotypes each demarcated by several mutational steps (Fig. 1). No haplotypes were shared among assemblages, and the extensive mutational distances separating their clusters underscored their deep evolutionary divergence.

Genetic differentiation and AMOVA

AMOVA results demonstrated extremely high and consistent genetic differentiation between assemblages A and B across all loci and the concatenated dataset (Table 4). ΦST values were close to unity (0.997–0.999) and were statistically significant ($P < 0.001$), indicating near-complete genetic structuring between these two assemblages. In all cases, variance among assemblages greatly exceeded within-population (error) variance, further supporting minimal gene flow between A and B and strong genetic separation. The concatenated analysis reinforced this pattern, yielding the highest ΦST estimate (0.9989), confirming robust multi-locus differentiation between assemblages.

Table 4
Genetic differentiation between assemblages A and B based on Analysis of Molecular Variance (AMOVA) across individual and concatenated loci

Loci	PhiST	P_value*	Var_Assemblage**	Var_Error***
tpi	0.9992087	< 0.001	5057.932	4.0058
gdh	0.9976083	< 0.001	4079.671	9.7805
bg	0.9970599	< 0.001	665.211	1.9615
Concatenated	0.9988802	< 0.001	25924.448	29.0625
* PhiST= Fixation index based on molecular variance among populations				
** Var_Assemblage = Variance among assemblages				
*** Var_Error =Variance within populations (error component)				

Phylogenetic reconstruction

Maximum likelihood phylogenetic analyses of the *tpi, gdh, and bg* loci revealed contrasting patterns of genetic diversity between *G. duodenalis* assemblages A and B. For assemblage A, the *tpi* and *gdh* single-locus trees showed all isolates clustering within one compact clade, indicating minimal divergence (**Figures S1 and S3**). The *bg* locus provided slightly higher resolution, separating assemblage A isolates into two subclusters (**Figure S2**). Assemblage B trees showed substantially greater variability, with isolates distributed across multiple well-supported subclusters (**Figure S1, S2, and S3**). Concatenated multi-locus phylogenies reinforced these patterns: assemblage A retained a cohesive topology with limited divergence (Fig. 2A), whereas assemblage B exhibited extensive branching and deeper phylogenetic structure (Fig. 2B). Overall, multi-locus analyses highlighted the conserved structure of assemblage A versus pronounced heterogeneity in assemblage B.

Nucleotide-level diversity for *tpi*, *bg*, and *gdh* is presented in **Tables S2–S4**. At *tpi*, 47 sequences were characterized. All 20 assemblage A isolates belonged to AII and were identical to U57897 (**Figure S1**). Among 27 assemblage B isolates, one matched BIV with a single SNP relative to KM926535, and 26 belonged to BIII (**Figure S1**). Most BIII sequences showed 1–6 SNPs relative to AY368165, with recurrent hotspots at G39, T91, A483, and C504, and several sequences showed double peaks (**Table S2**).

At *bg*, 46 sequences were obtained. All 20 assemblage A isolates were AII: five subtype A2 identical to AY545643, and fifteen A3, fourteen identical to AY072724, and one differing by one SNP (**Figure S2**). Among 26 assemblage B sequences, multiple subtypes were detected, showing 1–5 SNPs relative to AY647266, AY072727, AY647265, MG746612, and AY072728, with several displaying double peaks (**Table S3**).

At *gdh*, 44 sequences were characterized. All assemblage A isolates were AII; nineteen matched AY178737, and one differed by three SNPs (**Figure S3**). Assemblage B included eight BIV and others showing 4–11 SNPs relative to MG746607, AY178754, MK952603, MG746610, and MG746604, with recurrent hotspots (**Table S4**). No mixed A/B infections were detected.

Neutrality test analysis of G. duodenalis assemblages across three loci

Neutrality test statistics indicated that most loci in both assemblages A and B did not deviate significantly from neutrality, with Tajima's D, Fu & Li's D*, and Fu & Li's F* values generally close to zero and non-significant (P > 0.10), suggesting a pattern consistent with neutral evolution and demographic equilibrium (Table 5). Assemblage A showed largely neutral signals across all loci, whereas assemblage B exhibited mostly negative but non-significant values for the neutrality tests, particularly at the *tpi* locus, where Tajima's D showed a negative deviation (D = − 1.72; 0.10 > P > 0.05), indicating a tendency toward an excess of rare variants. Overall, the results did not provide strong evidence for selection or recent demographic expansion in either assemblage.

Table 5
Neutrality test statistics for *G. duodenalis* assemblages A and B

Loci	Assemblage	Tajima's D	p-value	Fu & Li's D*	p-value	Fu & Li's F*	p-value
tpi	A	0	> 0.10	0	> 0.10	0	> 0.10
	B	− 1.71918	0.10 − 0.05	-1.43465	> 0.10	-1.76817	> 0.10
gdh	A	0	> 0.10	0	> 0.10	0	> 0.10
	B	− 0.88971	> 0.10	-0.23775	> 0.10	-0.49469	> 0.10
bg	A	+ 0.93886	> 0.10	+ 0.86615	> 0.10	+ 1.01631	> 0.10
	B	− 0.83078	> 0.10	+ 1.15807	> 0.10	+ 0.67409	> 0.10

Clonal versus non-clonal expansion of G. duodenalis assemblages A and B

Multi-locus genotype (MLG) and linkage disequilibrium (LD) analyses

Linkage disequilibrium (LD) analyses of the concatenated three-locus dataset revealed strong contrasts between *G. duodenalis* assemblages A and B. Assemblage A showed extremely low genetic variability, with only two polymorphic sites detected (Fig. 3A). Site-based LD statistics indicated a complete non-random association among segregating sites (Za = 1.0, ZnS = 1.0, ZZ = 0), and all pairwise comparisons were significant (Fisher's exact test, P = 0.001). Multi-locus LD metrics further supported full linkage among loci, with maximal values for both the index of association (I_A = 1.0) and the standardized index ($\bar{r}D$ = 1.0; P = 0.001). LD-decay analysis showed no decline in D' or R^2 with increasing nucleotide distance (Table 6), consistent with an absence of detectable recombination. MLG analysis identified only two multi-locus genotypes among 19 isolates, one of which was dominant, indicating an extremely clonal population structure.

Table 6
Linkage disequilibrium (LD) and multi-locus association indices in *G. duodenalis* assemblages A and B (concatenated three loci)

Parameter	Assemblage A	Assemblage B
Polymorphic sites (S)	2	16
Za	1.000	0.1458
ZnS	1.000	0.2194
ZZ	0	-0.0735
Index of association (I$_A$)	1.000	2.718
Standardized index of association ($\bar{r}D$)	1.000	0.0623
Significant Fisher's comparisons	1	24
P-value	0.001	0.001
No. of MLGs	2	21

Za = index of association based on variance of pairwise genetic distances; ZnS = average standardized linkage disequilibrium; ZZ = standardized deviation of LD; I$_A$ = index of association; $\bar{r}D$ = standardized index of association. Significant associations were assessed using Fisher's exact test across all pairwise comparisons. Higher I$_A$ and $\bar{r}D$ values indicate stronger multi-locus linkage disequilibrium and deviation from random recombination.

In contrast, assemblage B exhibited substantially higher genetic diversity, with 16 polymorphic sites detected across the concatenated dataset (Fig. 3B). Despite this increased variability, LD remained significant. Site-based LD statistics (ZnS = 0.2194, Za = 0.1458, ZZ = − 0.0735) and 24 significant pairwise comparisons (P = 0.001) indicated persistent but weaker non-random associations among segregating sites. Multi-locus LD values also demonstrated significant deviation from random allelic assortment (I$_A$ = 2.718; $\bar{r}D$ = 0.062; P = 0.001). LD-decay analysis revealed only minor and inconsistent reductions in R^2 with increasing distance, without a clear decay trend, suggesting infrequent recombination within assemblage B (Table 6). MLG analysis identified 21 multi-locus genotypes among 22 isolates, reflecting high genotypic diversity driven by accumulated mutations and limited genetic exchange.

Split network analysis

The concatenated-loci network revealed a clear and pronounced separation between assemblages A and B, with no reticulation connecting the two groups. Within assemblage A, the split-network analysis showed a strongly tree-like topology among Egyptian samples, mirroring the maximum-likelihood phylogeny (Fig. 3C). Reticulation signals were essentially absent, network complexity was minimal, and all samples clustered into two well-defined groups except for isolate 319. This pattern is consistent with the low haplotype diversity and the lack of recombination signatures detected by both RDP5 and the Phi test. The network also indicated that circulating Egyptian strains are distinct from canonical sub-assemblage AI and representing sub-assemblage AII (A2/A3).

In contrast, assemblage B exhibited a highly reticulated network, indicative of extensive gene flow and higher sequence heterogeneity (Fig. 3D). Statistical analyses further confirmed significant recombination events, consistent with the presence of admixed mosaic genomes in which different genomic regions follow different evolutionary histories. Collectively, these results support a predominantly clonal population structure for assemblage A. In contrast, assemblage B displays substantially greater internal variability and reticulation, characteristic of a hybridized population (Fig. 4A).

Recombination analysis

Recombination screening was performed independently on the concatenated multi-locus datasets for assemblages A and B. In assemblage A, none of the RDP5 algorithms detected recombination, and the PHI test returned a non-significant result ($P > 0.05$). These findings, together with the low sequence variability and complete linkage observed across loci, indicate a highly homogeneous genetic background with no detectable evidence of genetic exchange. In assemblage B, a weak putative recombination signal was identified in a single isolate (sample ID: 423) by the MaxChi method; however, other RDP5 algorithms did not corroborate this signal, and the inferred breakpoints lacked statistical support. Consistent with this, the PHI test did not detect significant phylogenetic incompatibility, suggesting an overall absence of robust or reproducible recombination signals in the dataset. Because the analysis relies on a limited set of unlinked genetic markers (*tpi*, *bg*, and *gdh*), the resolution for detecting recombination, particularly in genetically heterogeneous populations, is inherently constrained. The observed patterns, therefore, align more closely with mutation-driven diversification, although the possibility of infrequent recombination events cannot be fully excluded.

Population structure analysis

To further resolve the population structure of *G. duodenalis*, we applied the Bayesian clustering method STRUCTURE to infer shared ancestry among isolates using the concatenated multi-locus dataset. STRUCTURE partitions individuals into ancestry components, where color blocks represent the proportional contribution of inferred ancestral populations, and mixed patterns indicate admixed genomes. We evaluated a range of K values and used the Evanno method to identify the most likely number of founder populations underlying the observed variation.

When all isolates were analyzed together, including both assemblages A and B, the strongest ΔK signal occurred at $K = 2$ ($\Delta K = 2284.39$), revealing two well-defined clusters that correspond precisely to the major genetic division between assemblages A (blue) and B (red) (Fig. 4B, **Figure S4A, and Table S5)**. This bipartition reflects the deep evolutionary split between the two assemblages and aligns with the phylogenetic and network-based analyses.

Separate STRUCTURE analyses of each assemblage uncovered additional sub-structuring. For assemblage A, the Evanno method supported $K = 4$ ($\Delta K = 35.44$) as the most likely number of ancestral groups. The bar plot reveals four distinct ancestry components with minimal admixture; nearly all isolates were assigned cleanly to a single-color block, with only one isolate (319) showing mixed ancestry (Fig. 4C, **Figure S4B, and**

Table S6). This pattern is consistent with the highly tree-like network topology observed for assemblage A and its lack of detectable recombination. The strong within-group homogeneity and limited mixed ancestry suggest a predominantly clonal structure with stable, well-defined lineage boundaries.

In contrast, assemblage B exhibited a markedly more heterogeneous population structure. The strongest ΔK peak supported K = 9 (ΔK = 67.23), indicating a substantially more complex ancestral composition (Fig. 4D, **Figure S4C, and Table S7**). At this higher K value, isolates displayed numerous color combinations with admixture, reflecting the reticulated network structure and recombination signature observed. Minor secondary peaks at lower K values were present but did not improve model support, reinforcing K = 9 as the best representation of the current population configuration. The ancestry patterns within assemblage B highlight its elevated sequence diversity, recombination, and deeper historical population mixing relative to assemblage A.

Together, the STRUCTURE analyses corroborate results from phylogenetic and network approaches, revealing a clear dichotomy between the population architectures of the two assemblages: a predominantly clonal and internally homogeneous assemblage A versus a subdivided and genetically diverse assemblage B. Although these patterns are consistent with greater sequence diversity and the possibility of past genetic exchange, the extent and biological significance of recombination within assemblage B remain uncertain. Overall, assemblage B displays deeper sub-structuring and greater heterogeneity than assemblage A, reflecting a more dynamic evolutionary history.

Discussion

Giardiasis remains one of the most important protozoal diarrheal diseases affecting children worldwide. This study presented a high-resolution multi-locus analysis of *G. duodenalis* in diarrheic children in northeastern Egypt. The findings revealed a pronounced contrast in the evolutionary and population-genetic structure of assemblages A and B. Using phylogenetic reconstruction, haplotype network analysis, linkage disequilibrium patterns, recombination screening, and Bayesian clustering, we showed that assemblage A exhibits extreme genetic restriction and clonal stability. In contrast, assemblage B displayed markedly greater genetic diversity, deeper sub-structuring, and signatures of a deeper and more complex evolutionary history.

G. duodenalis appears less prevalent in our study (10.64%) than previously reported in Egypt (19–21%) [20–22] and in neighboring or regional countries such as Sudan (25%) [23], Ethiopia (41%) [24], and Uganda (87%) [25], while exceeding rates documented in parts of Asia, including China (1.3%) [26]. Such variability likely reflects differences in diagnostic approaches, sample sizes, population characteristics, and environmental or socio-economic conditions that influence exposure, sanitation, and water quality. Infection appears significantly more common in children under 5 years, consistent with earlier Egyptian reports [20,27], and likely attributable to immature immunity, limited hygiene practices, and greater environmental contact (although some studies have shown higher infection in older children [28]). No significant sex-related difference was observed, aligning with most local data [20,27], despite occasional reports of higher male infection rates likely driven by behavioral exposures [28]. Seasonal patterns remain inconsistent across studies; although our

findings showed a peak in November, other Egyptian investigations reported increases in winter or summer [20,27], underscoring the need for more comprehensive national data. Collectively, these findings highlight the importance of strengthening water, sanitation, and hygiene interventions, prioritizing preventive measures for young children, improving diagnostic surveillance, and incorporating regional epidemiological patterns into public-health strategies to reduce transmission and disease burden.

This study provides a multi-locus characterization of *G. duodenalis* infections in children from northeastern Egypt, revealing that assemblages A and B occur at nearly equal frequencies. This balanced distribution mirrors previous findings in Egyptian children [20] and contrasts with earlier work in Egypt and across Africa that documented a dominance of assemblage B [29–36]. Similar B predominance has also been reported across Asia, Europe, and Latin America [37]. In contrast, recent studies in China and Algeria observed a higher prevalence of assemblage A [26,38], highlighting the substantial geographic variability in assemblage distribution.

Importantly, the absence of assemblages C or E, despite their occasional detection in Egyptian children [21,39], reinforces the primarily anthroponotic nature of infections in this region. Variation in assemblage distribution across populations is influenced by environmental exposures, sanitation conditions, host susceptibility, parasite traits, and transmission routes, including waterborne, foodborne, zoonotic, and human-to-human pathways [40]. Globally, assemblage B is more strongly associated with human infections than with animals, suggesting predominantly anthroponotic transmission [11,37].

Additionally, in this study, all assemblage A isolates belonged to the human-specific AII sub-assemblage, with no evidence of AI, which is more commonly detected in animals. This exclusive presence of AII is consistent with minimal zoonotic contribution and strongly supports human-to-human transmission as the primary route. Epidemiological analysis also indicated no significant association between infection status and animal contact, further reinforcing that domestic or livestock reservoirs likely played little role.

Another key finding is the absence of mixed A/B infections. Although mixed infections have been documented in India, Algeria, and China [38,41], none were detected here across all three loci. This may reflect lower simultaneous exposure to multiple assemblages, competitive exclusion within the host, or transmission patterns dominated by human-restricted lineages. Methodological sensitivity is unlikely to be a primary factor given our robust multi-locus design.

The multi-locus characterization of assemblage A showed striking consistency, with all isolates assigned to AII at each locus, with no novel polymorphisms detected. Only a single isolate showed heterogeneity at *gdh* and *bg*. This high degree of homogeneity echoes findings in prior studies that describe AII as a genetically stable, arguably recently evolved human-adapted lineage [20,31,42].

In contrast, sub-assemblage assignment for assemblage B remained challenging. While *tpi* sequences clustered with BIII/BIV, *bg* sequences could not be confidently assigned due to sparse reference data, and *gdh* indicated a mixture of BIII, BIV, and unassigned variants. These patterns reflect the globally recognized classification difficulties arising from allele sequence heterogeneity in assemblage B [14,43]. Despite these

uncertainties, the overall result mirrors global trends and regional reports from Angola, Ethiopia, and Mozambique [31,33,44], reinforcing greater diversity and sub-structuring of assemblage B.

Haplotype networks constructed from the multi-locus dataset further emphasized the deep separation between assemblages A and B. Assemblage A formed a compact cluster centered on a dominant haplotype, whereas assemblage B contained numerous distinct haplotypes separated by multiple mutational steps. This absence of shared haplotypes between assemblages, combined with large mutational distances, reflects their long-standing evolutionary divergence. Similar patterns have been documented globally, where sub-assemblage AII shows minimal haplotype variation, while assemblage B displays pronounced heterogeneity [15,43,45,46]. Phylogenetic analyses also demonstrated consistent separation of assemblages A and B across single-locus and concatenated datasets. Assemblage A formed uniform, well-supported clades consistent with its low diversity, while assemblage B exhibited complex branching and deeper structure, reflecting its higher polymorphism. Concatenated phylogenies improved resolution, particularly for assemblage A, confirming the value of multi-locus sequence typing for resolving evolutionary patterns in *Giardia* [46,47]. Split-network analysis supported these findings: assemblage A showed a strictly tree-like topology with minimal reticulation, whereas assemblage B exhibited greater internal reticulation, with statistically robust infrequent recombination signals.

Overall, the assemblage distribution observed in this study reflects predominantly anthroponotic transmission in Egyptian children. The consistency of AII, absence of AI and mixed infections, and the distinct clustering of BIII/BIV subtypes all point toward human-restricted cycles. These findings have important implications for public health, emphasizing the need for targeted interventions focusing on sanitation, water infrastructure, and hygiene in childcare and household settings.

This study uncovered strong contrasts in the evolutionary characteristics and population structures of *G. duodenalis* assemblages A and B. Assemblage A demonstrated extremely low genetic diversity across all loci, reflected in complete linkage disequilibrium, limited haplotype variation, and absence of detectable recombination. Multi-locus genotypes were few, with one dominant genotype representing most isolates. These patterns suggest a highly clonal population structure consistent with findings from earlier studies examining AII populations [42,46,48].

Assemblage B exhibited substantially greater polymorphism, with more variable sites, higher haplotype richness, and multiple SNP hotspots across different loci. Its multi-locus genotype richness far exceeded that of assemblage A, and LD values, while significant, were lower. This indicates persistent non-random allele associations, but with more genetic variation. Despite these indicators of complexity, we observed infrequent recombination, reflecting prior observations that assemblage B often shows patterns consistent with mutation accumulation and allele sequence heterogeneity rather than true meiotic recombination [49–52].

AMOVA results showed near-complete genetic differentiation between assemblages A and B, with ΦST values approaching 1 across all loci. The extreme level of divergence supports long-term independent evolution and minimal gene flow between assemblages, aligning with findings from other regions [53,54]. Neutrality tests did not indicate strong selection in either assemblage. The slightly negative Tajima's D for

assemblage B at *tpi* suggests an excess of rare variants, potentially reflecting population expansion or sub-structuring, though the deviation was not statistically significant.

Bayesian clustering provided further insight into population structure. When both assemblages were analyzed together, K = 2 was the strongest model, corresponding to the deep split between A and B. Within assemblage A, K = 4 indicated some fine-scale subdivision but with minimal admixture, consistent with strong clonality. Assemblage B required K = 9 to model its heterogeneity, reflecting its deeper sub-structuring and elevated diversity.

Conclusion

The exclusive detection of human-adapted AII and BIII/BIV subtypes, combined with epidemiological data showing no association with animal exposure, indicates that transmission in this setting is predominantly anthroponotic. These findings refine our understanding of *Giardia* population biology and clarify the limited role of domestic animals and livestock within the local transmission cycle. To further advance high-resolution molecular epidemiology, future investigations should adopt whole-genome sequencing directly from clinical specimens using emerging techniques such as targeted capture-enrichment sequencing and adaptive sequencing, which enable efficient recovery of parasite genomes from samples with high host-DNA background. Integration of these approaches with longitudinal sampling could resolve microevolutionary dynamics, detect emerging genotypes, and improve subtyping precision beyond traditional multi-locus methods.

From a One Health perspective, the assemblage-specific patterns observed here contribute valuable regional data that enhance the broader understanding of travelers' diarrhea risks. The predominance of human-adapted lineages underscores the importance of human-to-human and fecal-oral transmission pathways, rather than zoonotic sources, in both local and travel-associated giardiasis. These insights support more targeted public health strategies, improved risk assessments for travelers, and strengthened cross-border surveillance of enteric pathogens.

Methods

Ethics approval

All procedures followed the International Ethical Guidelines for Research Involving Human Subjects in Biological Research. The study was approved by the Research Ethics Committee of the Faculty of Veterinary Medicine, Mansoura University, Egypt (Approval Code: MU-ACUC [VM.R.24.05.166]). Written informed consent was obtained from all participants before sample collection. All stool samples and associated epidemiological and clinical data were anonymized to ensure confidentiality.

Study area and sample collection

Between May and December 2023, 451 fresh stool samples were collected from diarrheic children aged 1–10 years across four governorates in Northeastern Egypt: Dakahliya (n = 218), Sharqia (n = 87), Damietta (n = 75), and Gharbia (n = 71) (Table S1). Epidemiological information, including age, sex, residence (rural or urban), and history of animal contact, was collected using a structured questionnaire (**Table S8**). Approximately 2 g of stool sample from each child was transferred into sterile tubes containing 2 mL of 2.5 percent potassium dichromate (Sigma Life Sciences, Saint Louis, MO, USA), homogenized, labeled, and stored at 4°C at Mansoura University until processing. Samples were subsequently shipped to the Animal Parasitic Diseases Laboratory (Maryland, USA) for molecular analyses.

DNA extraction

Approximately 500 mg of each fecal sample was washed two to three times with Dulbecco's phosphate-buffered saline to remove potassium dichromate. DNA was extracted from the remaining pellet using the DNeasy PowerSoil Pro Kit (QIAGEN, Germantown, MD, USA) following the manufacturer's instructions. Extracted DNA was stored at − 20°C until further analysis.

Giardia detection, genotyping, and subtyping

All 451 samples were screened for *G. duodenalis* using a real-time PCR (qPCR) assay targeting the *SSU rRNA* gene with the 18SJVGF and 18SJVGR primers and the GEMS-G_18S-P probe as described by Jothikumar et al (**Table S9**) [55]. The hydrolysis probe was labeled at the 5' end with a Tex 615 fluorescent dye and utilized a double-quencher system consisting of a 3' Iowa Black quencher and an internal ZEN quencher. qPCR assays were performed using the PrimeTime® Gene Expression Master Mix (IDT; AB TaqMan FAST Advanced, Applied Biosystems) in a final reaction volume of 20 µL. Each reaction contained 10 µL of 2× Master Mix, 0.5 µL of each primer (final concentration 0.25 µM), 0.3 µL of probe (final concentration 0.15 µM), 6.7 µL of nuclease-free water, and 2 µL of DNA template. Amplification was carried out on a QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems) using the following cycling conditions: an initial denaturation at 95°C for 3 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Fluorescence signals were recorded at the end of each cycle. Positive and negative controls were included in each qPCR run. Samples with threshold cycle (CT) values < 40 were considered positive.

qPCR-positive samples were subjected to nested PCR targeting the *tpi* (530 bp) [56] [56], *bg* (511 bp) [57], and *gdh* (599 bp) [58] loci using published primer sets (**Table S9**). PCR amplification was performed in a total reaction volume of 25 µL, consisting of 12.5 µL of Platinum™ Hot Start PCR Master Mix (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA), 6.7 µL of molecular-biology–grade water, 0.5 µL of each primer (final concentration 10 µM), and 5 µL of template DNA. Thermal cycling began with an initial denaturation at 94°C for 2 minutes, followed by 40 cycles of 94°C for 15 seconds, 55°C for 15 seconds, and 68°C for 1.5 minutes. A final extension step was carried out at 68°C for 5 minutes. Each reaction included positive and negative controls. Amplified products were visualized on 1.5 percent agarose gels stained with GelRed (Millipore Sigma, Burlington, MA, USA). Secondary PCR products were purified with ExoSAP-IT Express (Thermo Fisher Scientific, Waltham, MA, USA) and sequenced bidirectionally at Psomagen (Psomagen, Rockville, MD, USA).

Sequence analysis and assemblage identification

Chromatograms were assembled and edited in Geneious Prime v2024.0.2 (https://www.geneious.com). Consensus sequences were aligned and compared with sequences in GenBank (https://www.ncbi.nlm.nih.gov/genbank/) using BLASTn to confirm species identity and assign isolates to *G. duodenalis* assemblages and sub-assemblages.

Genetic diversity analysis

Genetic diversity was assessed using DnaSP v5 [59] (http://www.ub.edu/dnasp), calculating several standard indices: the number of polymorphic sites (S), representing nucleotide positions with variation among sequences; parsimony-informative sites (PI), which are informative for phylogenetic inference; nucleotide diversity (π), reflecting the average number of nucleotide differences per site between sequences; the number of haplotypes (h); and haplotype diversity (Hd), indicating the probability that two randomly selected sequences differ. Analyses were conducted separately for each assemblage (A and B) to evaluate intra-assemblage diversity, and positions with missing or ambiguous nucleotides were excluded. These metrics were used to summarize genetic variability and inform subsequent neutrality and phylogenetic analyses.

Haplotype network construction

Median-joining haplotype networks were constructed using PopART v1.7 [60] under default parameters (ε = 0). Node sizes represent haplotype frequencies, and edges indicate mutational steps, allowing visualization of genealogical relationships within and between assemblages.

Analysis of molecular variance (AMOVA)

Genetic differentiation among *G. duodenalis* assemblages was evaluated using AMOVA implemented in Arlequin [61].The analyses were conducted independently for each assemblage (A and B) as well as for the concatenated multi-locus dataset. Genetic variation was partitioned into three hierarchical components: variation between assemblages, variation among populations within assemblages, and variation within populations. The fixation index ΦST (PhiST) was calculated to assess the degree of genetic differentiation. The significance of the variance components and fixation indices was evaluated using 10,000 permutations based on pairwise nucleotide differences, with p-values < 0.05 considered statistically significant.

Phylogenetic reconstruction

Phylogenetic trees were inferred using the maximum likelihood method in MEGA v11 [62]. Trees were constructed separately for each locus and for the concatenated multi-locus dataset, incorporating reference sequences from GenBank and representative reference genomes (https://www.ncbi.nlm.nih.gov/genbank/).

Pairwise P-distances were calculated after aligned the sequences followed by removing insertions and deletions. Branch support was assessed using 1,000 bootstrap replicates. Consensus trees were drawn with an arbitrary root and a scale length of 0.0005 substitutions per site, providing a measure of evolutionary divergence. All analyses followed default ML model settings unless otherwise specified.

Neutrality tests

Neutrality of the *G. duodenalis* sequences was evaluated using Tajima's D and Fu and Li's D* and F* statistics, as implemented in DnaSP v6 [63]. These tests compare different estimators of nucleotide diversity to detect deviations from expectations under neutral evolution. Tajima's D assesses the discrepancy between the number of segregating sites and the average pairwise nucleotide differences, whereas Fu and Li's D* and F* incorporate information from singleton mutations to increase sensitivity to recent demographic events. Sites containing missing or ambiguous nucleotides were excluded from the analyses. Statistical significance was assessed using 10,000 coalescent simulations. Positive values may indicate balancing selection or historical population bottlenecks, while negative values suggest population expansion or purifying selection; values near zero are consistent with neutrality. Neutrality tests were performed for each locus individually as well as for the concatenated multi-locus dataset to infer potential demographic patterns across assemblages.

Multi-locus genotype (MLG) and linkage disequilibrium (LD) analyses

LD patterns were interpreted in conjunction with MLG distributions to distinguish between clonal propagation and potential recombination. Multi-locus genotypes (MLGs) were identified using the poppr package in R v4.3 [64] based on the concatenated alignment of the *tpi*, *bg, and gdh* loci. Each MLG was defined as a unique combination of alleles across all polymorphic sites. MLG richness and genotype frequencies were calculated for each assemblage, and repeated MLGs were interpreted as evidence of clonal expansion. Linkage disequilibrium (LD) was assessed to quantify non-random associations among polymorphic sites within each assemblage and across the concatenated multi-locus dataset. LD statistics (Z□, ZnS, and ZZ) were calculated in DnaSP v6 to estimate the average and standardized associations among segregating sites [63]. Statistical significance was evaluated using Fisher's exact test applied to all pairwise comparisons of polymorphic sites. Multi-locus LD and potential recombination were further examined using the index of association (I□) and the standardized index of association ($\bar{r}d$), computed with the poppr package in R v4.3. These metrics quantify deviations of multi-locus genotype distributions from expectations under linkage equilibrium. Significance was assessed by comparing observed values to a null distribution generated from 999 random permutations of alleles among individuals, with $P < 0.05$ indicating significant multi-locus LD. To investigate the relationship between physical distance and genetic linkage, LD decay was evaluated using two complementary approaches. First, D' was calculated for all pairs of polymorphic sites and plotted against nucleotide distance; sustained high D' values across long genomic distances were interpreted as evidence of strong linkage and a predominantly clonal population structure.

Second, R^2 was plotted as a function of physical distance to assess the strength of allelic association across loci. LD decay patterns were interpreted as follows: a stable R^2 across increasing distances indicated limited or absent recombination and consistent clonal inheritance, whereas a gradual decline in R^2 suggested historical recombination or mutation-driven diversification. Sample 319 (Assemblage A) and sequences exhibiting ambiguous nucleotide signals in Assemblage B were excluded from all concatenated multi-locus datasets before downstream analyses.

Split network analysis

The MEGA v11 software program [62] was used to align the sequences using default settings. Neighbor-net networks were generated after importing the aligned Nexus sequence file in SplitsTree5 [65] to visualize conflicting phylogenetic signals and possible reticulate evolutionary patterns based on concatenated multi-locus datasets.

Recombination detection

Recombination was assessed using RDP5 (http://web.cbio.uct.ac.za/~darren/rdp.html) [66], which employs algorithms including RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, and 3Seq [66,67]. Only recombination events detected by at least three independent methods and supported by statistically significant p-values ($p < 0.05$) after Bonferroni correction were considered reliable. Recombination was further evaluated using the Pairwise Homoplasy Index (PHI) test in SplitsTree5, with p-values < 0.05 interpreted as evidence of recombination. Analyses were performed separately for each locus, for concatenated datasets, and within assemblage A, assemblage B, and combined samples to evaluate both intra- and inter-assemblage recombination patterns.

Population structure analysis

Population structure was assessed using a Bayesian clustering algorithm implemented in the STRUCTURE v2.3.4 [68] under the admixture model with correlated allele frequencies. Five independent runs were performed for each K value (1–10), with a burn-in of 50,000 and a Markov chain Monte Carlo (MCMC) length of 100,000. The most likely number of clusters was identified using the Evanno method implemented in Structure Harvester [69] (structureHarvester.py; Dent Earl), which provided mean LnP(K), standard deviations, and ΔK values across replicates. Cluster alignment and visualization of ancestry coefficients were performed using CLUMPAK (https://clumpak.evolseq.net/) [70]. Individual membership coefficients (Q values) were used to assign isolates to clusters based on the highest posterior probability, and the distribution of cluster sizes was examined to assess the stability and biological relevance of the inferred subpopulations.

Split network analysis

The MEGA v11 software program [62] was used to align the sequences using default settings. Neighbor-net networks were generated after importing the aligned Nexus sequence file in SplitsTree5 [65] to visualize

conflicting phylogenetic signals and possible reticulate evolutionary patterns based on concatenated multi-locus datasets.

Recombination detection

Recombination was assessed using RDP5 (http://web.cbio.uct.ac.za/~darren/rdp.html) [66], which employs algorithms including RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, and 3Seq [66,67]. Only recombination events detected by at least three independent methods and supported by statistically significant p-values ($p < 0.05$) after Bonferroni correction were considered reliable. Recombination was further evaluated using the Pairwise Homoplasy Index (PHI) test in SplitsTree5, with p-values < 0.05 interpreted as evidence of recombination. Analyses were performed separately for each locus, for concatenated datasets, and within assemblage A, assemblage B, and combined samples to evaluate both intra- and inter-assemblage recombination patterns.

Population structure analysis

Population structure was assessed using a Bayesian clustering algorithm implemented in the STRUCTURE v2.3.4 [68] under the admixture model with correlated allele frequencies. Five independent runs were performed for each K value (1–10), with a burn-in of 50,000 and a Markov chain Monte Carlo (MCMC) length of 100,000. The most likely number of clusters was identified using the Evanno method implemented in Structure Harvester [69] (structureHarvester.py; Dent Earl), which provided mean LnP(K), standard deviations, and ΔK values across replicates. Cluster alignment and visualization of ancestry coefficients were performed using CLUMPAK (https://clumpak.evolseq.net/) [70]. Individual membership coefficients (Q values) were used to assign isolates to clusters based on the highest posterior probability, and the distribution of cluster sizes was examined to assess the stability and biological relevance of the inferred subpopulations.

Statistical analysis

Associations between *G. duodenalis* infection and categorical variables (age, sex, residence, animal contact, governorate, and sampling month) were tested using the Chi-square test. Statistical significance was defined as $p < 0.05$. Analyses were conducted in SPSS v20.0 (IBM Corp., Armonk, NY, USA).

Declarations

Acknowledgments

We acknowledge GiardiaDB (https://giardiadb.org/giardiadb/app) for providing a publicly available repository for Giardia genomic data. GiardiaDB is part of the Eukaryotic Pathogen, Vector, and Host Informatics Resources (VEuPathDB). Additional computing resources were provided by the SciNet HPC Consortium (https://scinet.usda.gov/). We also acknowledge Mr. Valsin Fournet, Mr. Matthew Valente, and Dr. Randi Turner for their helpful discussion.

Author contributions

D.N. and A.K. conceptualized the project. D.N. performed the experiments and generated the data used for analysis, figures, and tables. D.N., B.M.R., and A.K. provided guidance throughout the project. A.K. supervised the work. D.N., B.M.R., and A.K. wrote the manuscript, and B.M.R. and A.K. provided funding.

Study Consent

The study protocol was reviewed and approved by the Ethics Committee of the Faculty of Veterinary Medicine, Mansoura University, under approval number (MU-ACUC [VM.R.24.05.166]). Written informed consent was obtained from the parents or legal guardians of all participating children before sample collection. All procedures were performed in accordance with relevant ethical guidelines and regulations.

Data availability

Representative sequences were deposited in GenBank under accession numbers PX778845-PX778891 for the *tpi* locus, PX778892-PX778937 for the *bg* locus, and PX782280-PX782323 for the *gdh* locus.

Competing interest

The authors declare no competing interests.

Funding information

This work was financially supported by USDA CRIS Project 8042-32420-007-00D, 8042-31000-112-000D.

References

1. Cacciò, S. M., Lalle, M. & Svärd, S. G. Host specificity in the Giardia duodenalis species complex. *Infect. Genet. Evol.* **66**, 335–345 (2018).
2. Asghari, A. et al. Molecular epidemiology and multi-locus genotyping of Giardia duodenalis in individuals attending major public hospitals in Shiraz, southwestern Iran: A public health concern. *Parasite Epidemiol. Control*. **25**, e00354 (2024).
3. Cacciò, S. M. & Ryan, U. Molecular epidemiology of giardiasis. *Mol. Biochem. Parasitol.* **160**, 75–80 (2008).
4. Coffey, C. M. et al. Evolving Epidemiology of Reported Giardiasis Cases in the United States, 1995–2016. *Clin. Infect. Dis.* **72**, 764–770 (2021).
5. Cai, W., Ryan, U., Xiao, L. & Feng, Y. Zoonotic giardiasis: an update. *Parasitol. Res.* **120**, 4199–4218 (2021).
6. Feng, Y. & Xiao, L. Zoonotic Potential and Molecular Epidemiology of Giardia Species and Giardiasis. *Clin. Microbiol. Rev.* **24**, 110–140 (2011).
7. Zhao, W. et al. Occurrence and multi-locus genotyping of Giardia duodenalis in diarrheic and asymptomatic children from South of Zhejiang province in China. *Acta Trop.* **258**, 107341 (2024).

8. Gutiérrez, L. & Bartelt, L. Current Understanding of Giardia lamblia and Pathogenesis of Stunting and Cognitive Deficits in Children from Low- and Middle-Income Countries. *Curr. Trop. Med. Rep.* **11**, 28–39 (2024).
9. Dixon, B. R. Giardia duodenalis in humans and animals – Transmission and disease. *Res. Vet. Sci.* **135**, 283–289 (2021).
10. Ryan, U. M., Feng, Y., Fayer, R. & Xiao, L. Taxonomy and molecular epidemiology of Cryptosporidium and Giardia – a 50 year perspective (1971–2021). *Int. J. Parasitol.* **51**, 1099–1119 (2021).
11. Xiao, L. & Feng, Y. Molecular epidemiologic tools for waterborne pathogens Cryptosporidium spp. and Giardia duodenalis. *Food Waterborne Parasitol.* **8–9**, 14–32 (2017).
12. Ankarklev, J. et al. A novel high-resolution multi-locus sequence typing of Giardia intestinalis Assemblage A isolates reveals zoonotic transmission, clonal outbreaks and recombination. *Infect. Genet. Evol.* **60**, 7–16 (2018).
13. Solarczyk, P. et al. Multi-locus genotyping of Giardia duodenalis isolates from red deer (Cervus elaphus) and roe deer (Capreolus capreolus) from Poland. *Folia Parasitol.* **59**, 237–240 (2012).
14. Seabolt Matthew, H., Konstantinidis Konstantinos, T. & Roellig Dawn, M. Hidden Diversity within Common Protozoan Parasites as Revealed by a Novel Genomotyping Scheme. *Appl. Environ. Microbiol.* **87**, e02275–e02220 (2021).
15. Woschke, A. et al. Suitability of current typing procedures to identify epidemiologically linked human Giardia duodenalis isolates. *PLoS Negl. Trop. Dis.* **15**, e0009277 (2021).
16. Hartl, D. L. & Clark, A. G. *Principles of population genetics* 2nd edn, 682 (Sinauer Associates., 1989).
17. Combes, C. *Parasitism: the ecology and evolution of intimate interactions* (University of Chicago Press, 2001).
18. de Meeûs, T. et al. Population genetics and molecular epidemiology or how to débusquer la bête. *Infect. Genet. Evol.* **7**, 308–332 (2007).
19. Huyse, T., Poulin, R. & Théron, A. Speciation in parasites: a population genetics approach. *Trends Parasitol.* **21**, 469–475 (2005).
20. Naguib, D. et al. Molecular characterization of Cryptosporidium spp. and Giardia duodenalis in children in Egypt. *Parasites Vectors*. **11**, 403 (2018).
21. Helmy, Y. A. et al. Epidemiology of Giardia duodenalis infection in ruminant livestock and children in the Ismailia province of Egypt: insights by genetic characterization. *Parasit. Vectors*. **7**, 321 (2014).
22. Ibrahim, A., El-Alfy, E. S. N., Darwish, A., Naguib, D. & Gad, M. E. Genetic Diversity of Cryptosporidium Causing Infections from Diarrheic Cases in Egypt and Co-infections with Other Intestinal Protozoan Parasites. *Egyptian J. Veterinary Sciences*, 1–13 (2025).
23. Tamomh, A. G. et al. Intestinal Parasitic Infections and Associated Risk Factors Among Children in an Internally Displaced Center (IDC) in Kosti City of the White Nile State, Sudan. *J Parasitol Res* 5542680 (2025). (2025).
24. Alemu, Y. et al. Prevalence and assemblage of Giardia duodenalis in a case-control study of children under 5 years from Jimma, Southwest Ethiopia. *Parasitol. Res.* **123**, 38 (2023).

25. Al-Shehri, H., James LaCourse, E., Klimach, O., Kabatereine, N. B. & Stothard, J. R. Molecular characterisation and taxon assemblage typing of giardiasis in primary school children living close to the shoreline of Lake Albert, Uganda. *Parasite Epidemiol. Control*. **4**, e00074 (2019).
26. Zhao, W. et al. Occurrence and multi-locus genotyping of Giardia duodenalis in diarrheic and asymptomatic children from South of Zhejiang province in China. *Acta Trop.* **258**, 107341 (2024).
27. Elmahallawy, E. K. et al. Microscopy detection and molecular characterisation of Giardia duodenalis infection in outpatients seeking medical care in Egypt. *Front. Public. Health*. **12**, 1377123 (2024).
28. Helmy, M. M., Abdel-Fattah, H. S. & Rashed, L. Real-time PCR/RFLP assay to detect Giardia intestinalis genotypes in human isolates with diarrhea in Egypt. *J Parasitol* **95**, (2009).
29. El Fatni, C., Olmo, F., Fatni, E., Romero, H., Rosales, M. J. & D. & First genotyping of Giardia duodenalis and prevalence of enteroparasites in children from Tetouan (Morocco). *Parasite* **21**, 48 (2014).
30. Mbae, C. et al. Molecular Characterization of Giardia duodenalis in Children in Kenya. *BMC Infect. Dis.* **16**, 135 (2016).
31. Wegayehu, T. et al. Multi-locus genotyping of Giardia duodenalis isolates from children in Oromia Special Zone, central Ethiopia. *BMC Microbiol.* **16**, 89 (2016).
32. Tembo, S. J. et al. Prevalence and genotypic characterization of Giardia duodenalis isolates from asymptomatic school-going children in Lusaka, Zambia. *Food Waterborne Parasitol.* **19**, e00072 (2020).
33. Messa, A. Jr. et al. Molecular diversity of Giardia duodenalis in children under 5 years from the Manhiça district, Southern Mozambique enrolled in a matched case-control study on the aetiology of diarrhoea. *PLoS Negl. Trop. Dis.* **15**, e0008987 (2021).
34. Tijani, M. K. et al. High Diversity of Giardia duodenalis Assemblages and Sub-Assemblages in Asymptomatic School Children in Ibadan, Nigeria. *Trop Med. Infect. Dis* **8** (2023).
35. Fahmy, H. M. et al. Giardia duodenalis assemblages in Egyptian children with diarrhea. *Eur. J. Clin. Microbiol. Infect. Dis.* **34**, 1573–1581 (2015).
36. El-Wahsh, H. M. et al. Antiparasitic Activity of Hedera helix Extract-Loaded Chitosan Nanoparticles in Experimentally Induced Giardiasis. *Veterinary sciences* 13 (2026).
37. Ryan, U. & Cacciò, S. M. Zoonotic potential of Giardia. *Int. J. Parasitol.* **43**, 943–956 (2013).
38. Baroudi, D. et al. Distribution of Giardia duodenalis genotypes in diarrheic children from Algiers, Algeria. *Parasitol. Int.* **110**, 103132 (2026).
39. Soliman, R. H., Fuentes, I. & Rubio, J. M. Identification of a novel Assemblage B subgenotype and a zoonotic Assemblage C in human isolates of Giardia intestinalis in Egypt. *Parasitol. Int.* **60**, 507–511 (2011).
40. Cotton, J. A., Beatty, J. K. & Buret, A. G. Host parasite interactions and pathophysiology in Giardia infections. *Int. J. Parasitol.* **41**, 925–933 (2011).
41. Laishram, S., Kannan, A., Rajendran, P., Kang, G. & Ajjampur, S. S. R. Mixed Giardia duodenalis assemblage infections in children and adults in South India. *Epidemiol. Infect.* **140**, 2023–2027 (2012).
42. Minetti, C. et al. Determination of Giardia duodenalis assemblages and multi-locus genotypes in patients with sporadic giardiasis from England. *Parasit. Vectors*. **8**, 444 (2015).

43. Ankarklev, J., Svärd, S. G. & Lebbad, M. Allelic sequence heterozygosity in single Giardia parasites. *BMC Microbiol.* **12**, 65 (2012).
44. Dacal, E. et al. Prevalence and molecular characterization of Strongyloides stercoralis, Giardia duodenalis, Cryptosporidium spp., and Blastocystis spp. isolates in school children in Cubal, Western Angola. *Parasites Vectors*. **11**, 67 (2018).
45. Lebbad, M. et al. From mouse to moose: Multi-locus genotyping of Giardia isolates from various animal species. *Vet. Parasitol.* **168**, 231–239 (2010).
46. Klotz, C. et al. Extensive testing of a multi-locus sequence typing scheme for Giardia duodenalis assemblage A confirms its good discriminatory power. *Parasit. Vectors*. **15**, 489 (2022).
47. Higuera, A. et al. Development of a Multi-locus Sequence Typing Scheme for Giardia intestinalis. *Genes (Basel)* 11 (2020).
48. Wielinga, C. M. & Thompson, R. C. Comparative evaluation of Giardia duodenalis sequence data. *Parasitology* **134**, 1795–1821 (2007).
49. Cacciò, S. M. & Sprong, H. Giardia duodenalis: Genetic recombination and its implications for taxonomy and molecular epidemiology. *Exp. Parasitol.* **124**, 107–112 (2010).
50. Takumi, K. et al. Population-based analyses of Giardia duodenalis is consistent with the clonal assemblage structure. *Parasites Vectors*. **5**, 168 (2012).
51. Xu, F., Jerlström-Hultqvist, J. & Andersson, J. O. Genome-Wide Analyses of Recombination Suggest That Giardia intestinalis Assemblages Represent Different Species. *Mol. Biol. Evol.* **29**, 2895–2898 (2012).
52. Cooper, M. A., Adam, R. D., Worobey, M. & Sterling, C. R. Population genetics provides evidence for recombination in Giardia. *Curr. biology: CB*. **17**, 1984–1988 (2007).
53. Choy, S. H., Mahdy, M. A., Al-Mekhlafı, H. M., Low, V. L. & Surin, J. Population expansion and gene flow in Giardia duodenalis as revealed by triosephosphate isomerase gene. *Parasit. Vectors*. **8**, 454 (2015).
54. Thompson, R. C. A. & Ash, A. Molecular epidemiology of Giardia and Cryptosporidium infections - What's new? *Infect. Genet. Evol.* **75**, 103951 (2019).
55. Jothikumar, N., Murphy, J. L. & Hill, V. R. Detection and identification of Giardia species using real-time PCR and sequencing. *J. Microbiol. Methods*. **189**, 106279 (2021).
56. Sulaiman, I. M. et al. Triosephosphate isomerase gene characterization and potential zoonotic transmission of Giardia duodenalis. *Emerg. Infect. Dis.* **9**, 1444–1452 (2003).
57. Cacciò, S. M., De Giacomo, M. & Pozio, E. Sequence analysis of the beta-giardin gene and development of a polymerase chain reaction-restriction fragment length polymorphism assay to genotype Giardia duodenalis cysts from human faecal samples. *Int. J. Parasitol.* **32**, 1023–1030 (2002).
58. Read, C. M., Monis, P. T. & Thompson, R. C. Discrimination of all genotypes of Giardia duodenalis at the glutamate dehydrogenase locus using PCR-RFLP. *Infect. Genet. Evol.* **4**, 125–130 (2004).
59. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
60. Leigh, J. W. & Bryant, D. popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).

61. Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinf. online*. **1**, 47–50 (2007).
62. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
63. Rozas, J. et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
64. Kamvar, Z. N., Tabima, J. F. & Grünwald, N. J. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281 (2014).
65. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
66. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).
67. Martin, D. P., Williamson, C. & Posada, D. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260–262 (2005).
68. Hubisz, M. J., Falush, D., Stephens, M. & Pritchard, J. K. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9**, 1322–1332 (2009).
69. Knapp, E. W., Irausquin, S. J., Friedman, R. & Hughes, A. L. PolyAna: analyzing synonymous and nonsynonymous polymorphic sites. *Conserv. Genet. Resour.* **3**, 429–431 (2011).
70. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).

Figures

Figure 1

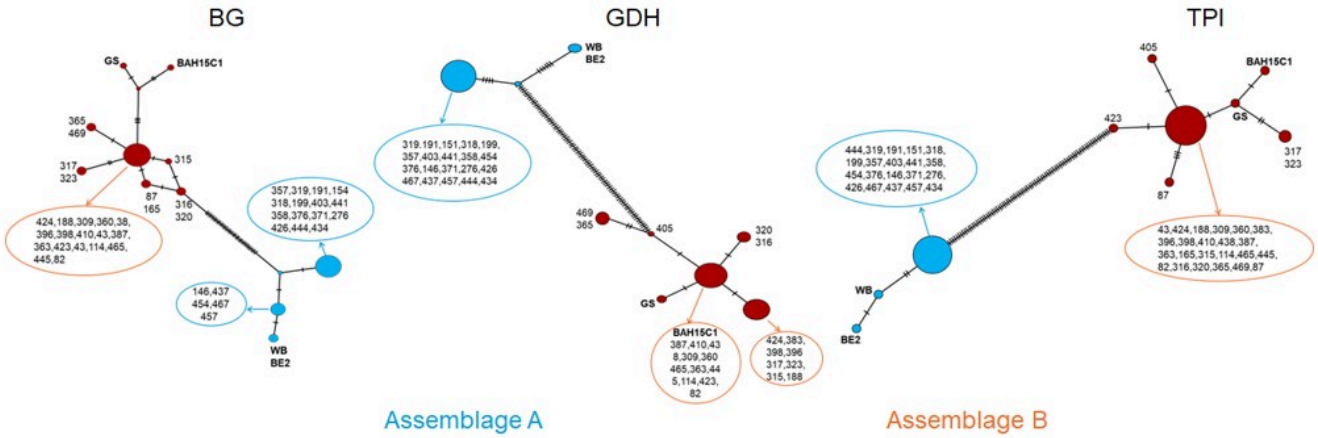


Figure 1

Haplotype network shows two distinct, deeply separated clusters of *G. duodenalis*. Median-joining haplotype networks for *G. duodenalis* assemblages A and B were constructed using multi-locus sequence data generated from the *bg*, *gdh*, and *tpi* loci. Networks were generated in PopART v1.7 [60] using default parameters (ε = 0) to visualize multi-locus genetic variation among isolates. Each node represents a unique haplotype, with node size proportional to the number of samples sharing that haplotype. Edges denote mutational steps between haplotypes. Assemblage A (blue) and assemblage B (red) form clearly separated clusters, reflecting their distinct evolutionary lineages. Reference strains WB and BE2 (assemblage A) and GS and BAH15C1 (assemblage B) are included to anchor the networks. Sample identifiers are displayed adjacent to nodes to indicate haplotype membership. The networks highlight lower haplotype diversity and tighter clustering within assemblage A, in contrast to the greater heterogeneity and more complex haplotype structure observed in assemblage B.

Figure 2

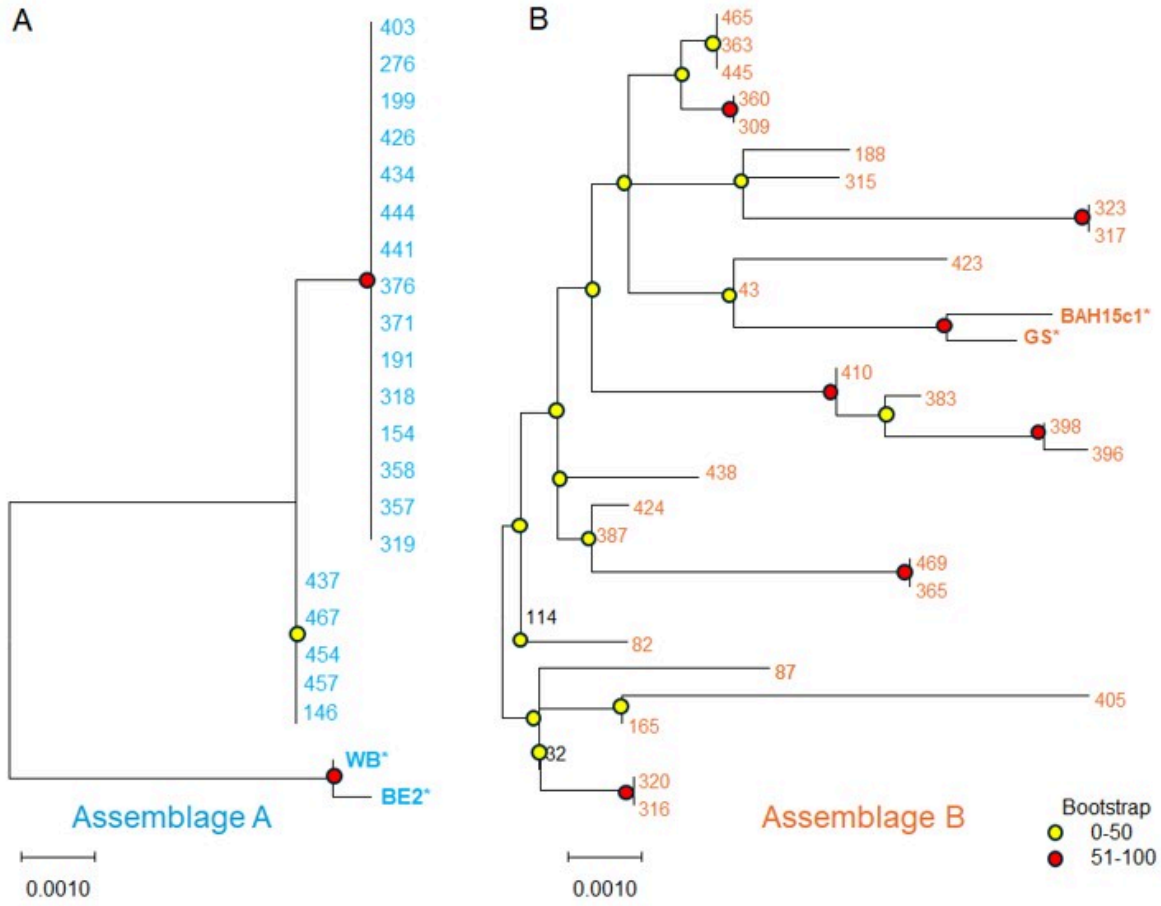


Figure 2

Maximum likelihood phylogenetic analysis of concatenated multi-locus sequences from *G. duodenalis* assemblages A and B. Phylogenetic trees were generated from concatenated *tpi*, *bg*, and *gdh* loci to resolve

evolutionary relationships within assemblages A and B. (A) Phylogenetic tree of assemblage A isolates, including reference strains WB and BE2. (B) Phylogenetic tree of assemblage B isolates, with GS and BAH15C1 included as reference strains. Sample identifiers are shown at terminal branches. Bootstrap support values are represented by colored circles, with red indicating higher support (51–100) and yellow indicating lower support (0–50). Scale bars indicate the number of nucleotide substitutions per site. All trees were inferred using the maximum likelihood method with 1,000 bootstrap replicates, illustrating the genetic structure, divergence patterns, and lineage separation characteristic of assemblages A and B.

Figure 3

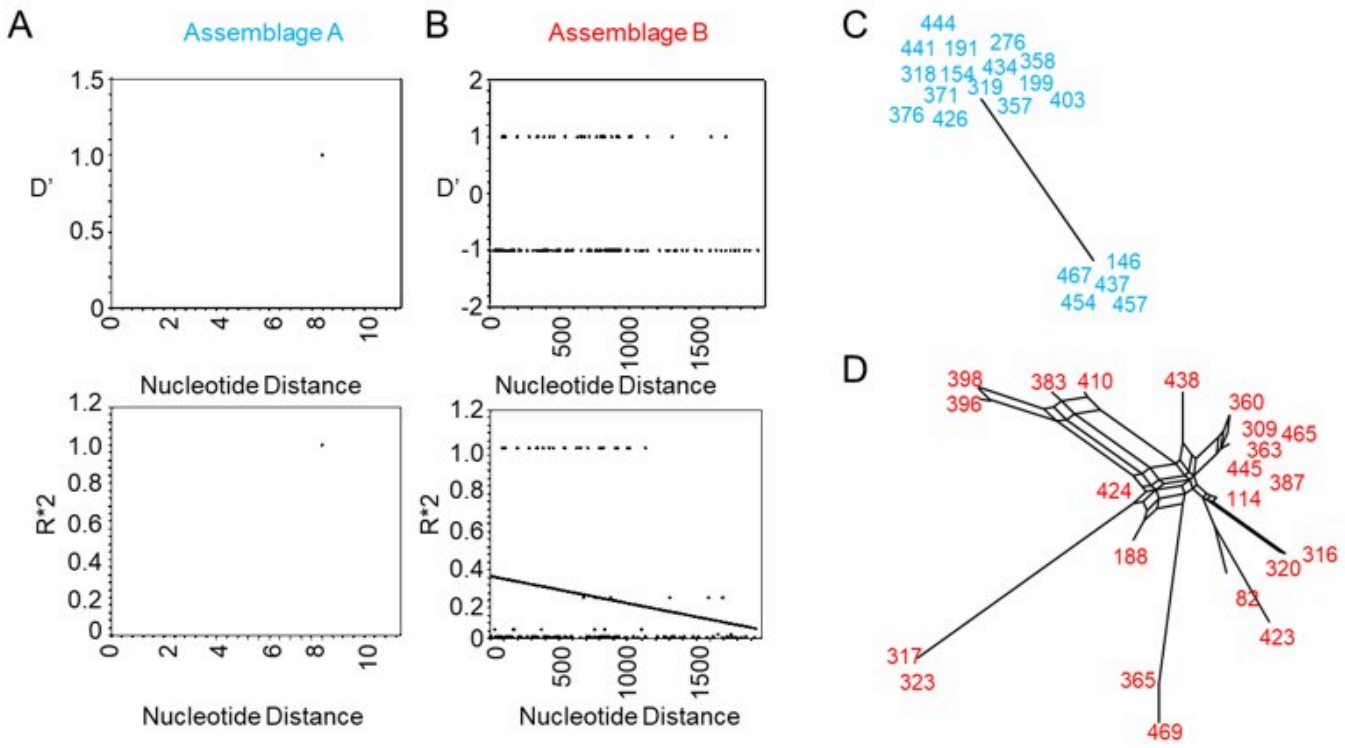


Figure 3

Linkage disequilibrium (LD) decay and Neighbor-Net phylogenetic network analyses of *G. duodenalis*assemblages A and B. (**A**) LD decay in assemblage A showing D' and R^2 values consistently close to 1.0 across all pairwise nucleotide distances, indicating strong linkage disequilibrium and absence of detectable decay. (**B**) LD decay in assemblage B showing variable D' and R^2 values across nucleotide distances without a clear decay pattern, indicating persistent linkage disequilibrium despite higher genetic variability. (**C**) Neighbor-Net phylogenetic network of assemblage A inferred using the Neighbor-Net algorithm in Splits Tree, displaying a highly tree-like topology with no evidence of reticulation, consistent with low genetic diversity and a strongly structured population. (**D**) Neighbor-Net phylogenetic network of

assemblage B inferred using the Neighbor-Net algorithm in Splits Tree, showing increased genetic diversity and reticulation, reflecting a more heterogeneous population structure.

Figure 4

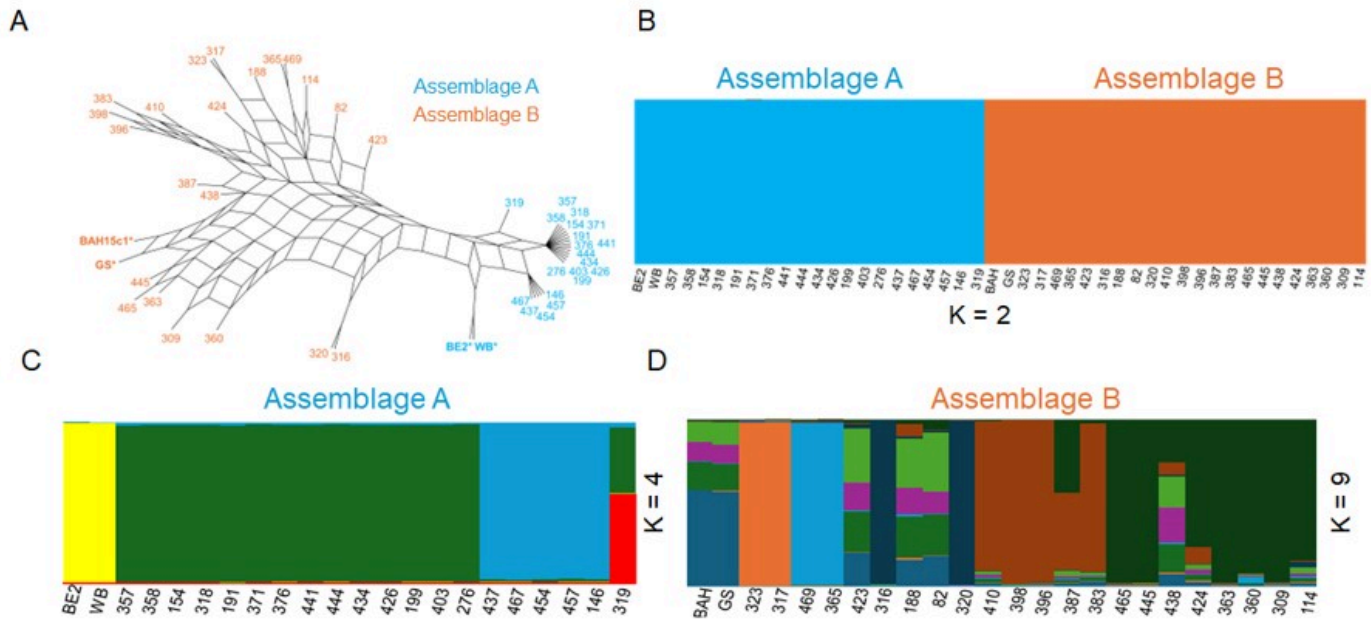


Figure 4

Neighbor-Net phylogenetic network analyses and population structure of *G. duodenalis* assemblages A and B. (A) Phylogenetic network generated in SplitsTree using the NeighborNet algorithm, based on concatenated multi-locus sequences from assemblages A and B. The network illustrates genetic relationships and reticulation patterns among isolates, with WB and BE2 included as reference strains for assemblage A and GS and BAH15C1 for assemblage B. (B) Population structure analysis at K = 2, showing the primary division between assemblage A and assemblage B. (C) Sub-structuring within assemblage A at K = 4, demonstrating genetic heterogeneity and the presence of multiple sub-clusters within assemblage A. (D) Sub-structuring within assemblage B at K = 9, reflecting the higher genetic diversity and more complex clustering patterns characteristic of assemblage B. Together, these analyses reveal both broad-scale assemblage differentiation and fine-scale population structure within *G. duodenalis*.