



OPEN Euro-Asian hybrids of *Echinococcus multilocularis* from red foxes in northern and northeastern Poland result from secondary contact between long-isolated populations

Paweł Gładysz¹✉, Dorota Bielińska-Wąż², Piotr Wąż³, Małgorzata Samorek-Pieróg⁴, Jacek Karamon⁴ & Anna Lass¹✉

We investigated the mitochondrial genetic diversity of *Echinococcus multilocularis* in red foxes from northern and northeastern Poland, and tested the hypothesis that the presence of Asian haplotypes in Poland stems from their introduction from Northeast Asia to Eastern Europe. We targeted the cytochrome c oxidase subunit 1 gene (*cox1*) of 263 tapeworms from 59 red foxes hunted in selected districts of Pomorskie, Warmińsko-Mazurskie, and Podlaskie Voivodships. We calculated several measures of genetic variation, and constructed median-joining haplotype networks and a RelTime timetree. In addition, a novel alignment-free computational method was applied: the 4D-Dynamic Representation of DNA/RNA Sequences combined with *K*-means clustering. PCR and sequencing were successful for 252 tapeworms. We identified six haplotypes, three of which were novel. The mismatch distribution was bimodal. The fixation index (F_{ST}) between the European and Asian haplogroups was 0.97. Our findings suggest secondary contact between two previously isolated metapopulations with distinct genetic lineages. Clustering patterns appear to support the hypothesis that Asian variants of *E. multilocularis* spread into northeastern Poland as a consequence of multiple introductions of raccoon dogs from the Soviet Far East to Eastern Europe.

Keywords Alignment-free method, Alveolar echinococcosis, Asian origin, Phylogeography

Research into the phylogeography of parasites and their hosts offers insight into past modes of pathogen dispersal, possible present mechanisms at play, and potential directions of future spread. *Echinococcus multilocularis* (Leuckart, 1863) is the causative agent of alveolar echinococcosis, a severe zoonotic disease with a long incubation period and poor prognosis unless treated^{1,2}. The phylogeography of the species in Europe is shaped primarily by the refugial and post-glacial history of the red fox and the recession of the Arctic fox^{3–7}.

Echinococcus multilocularis gene flow to a large extent depends on chance. The parasite has little agency: its movement relies on the migratory behaviour of its hosts and passive egg transport. When mating, it remains limited to conspecifics in the definitive host's small intestine, most of which likely stem from a single metacestode—making geitonogamy and autogamy its predominant modes of sexual reproduction^{8,9}. It comes as no surprise, then, that detection of autochthonous extra-continental variants and cross-continental hybrids raises questions about their origin as well as their pathogenicity and virulence for local host communities, including people¹⁰.

To this day, Asian or part-Asian adult tapeworms of *E. multilocularis* west of the Urals have been observed in definitive hosts from European Russia¹¹, Poland¹², and Latvia⁵. A local Asian mitochondrial variant was first discovered in Poland in 2017¹³. Concatenated sequences of three mitochondrial markers revealed a haplotype clustering away from European samples and close to Japanese and Kazakh ones in a statistical parsimony

¹Division of Tropical Parasitology, Department of Tropical Medicine and Parasitology, Institute of Maritime and Tropical Medicine, Medical University of Gdańsk, Powstania Styczniowego 9B, 81-519 Gdynia, Poland. ²Department of Radiological Informatics and Statistics, Medical University of Gdańsk, 80-210 Gdańsk, Poland. ³Department of Nuclear Medicine, Medical University of Gdańsk, 80-210 Gdańsk, Poland. ⁴National Veterinary Research Institute, Partyzantów 57, 24-100 Puławy, Poland. ✉email: pawel.gladysz@gumed.edu.pl; anna.lass@gumed.edu.pl

network. Of 74 tapeworms collected across the country, seven (9.5%) from northeastern and central Poland belonged to the foreign haplotype.

A follow-up study using the microsatellite EmsB and the complete cytochrome c oxidase subunit 1 gene (*cox1*) revealed introgression¹². Thirty-two out of 349 Polish tapeworms (9.2%) were Asian by descent in the nuclear and/or mitochondrial genome, extending as far west as the Polish-German border. Seventeen of them (4.9%) were Euro-Asian hybrids. Later investigations demonstrated cases of Asian variants parasitising pigs¹⁴ and people¹⁵, indicating that their eggs contaminate soil and produce.

A few explanations for the presence of Asian genomic elements in the Polish *E. multilocularis* population have been proposed: (a) recent introduction by migrating canids^{12,13}; (b) human-mediated introduction with raccoon dogs transported in the 20th century from Northeast Asia to Eastern Europe⁵; and (c) ancient polymorphism present for a long time: in this scenario, hybrids result from prolonged sympatry of European and Asian populations, sufficient to allow introgression, which, in turn, would indicate the existence of a westward gradient of progressively decreasing Asian contribution¹².

In this study, we explored the *cox1* diversity of *E. multilocularis* from red foxes in northern and northeastern Poland. We tested the hypothesis that the presence of Asian haplotypes in Poland is a result of their introduction from Northeast Asia to Eastern Europe. We also compared our results with previously acquired nuclear data¹⁶.

Materials and methods

Two hundred and sixty-three tapeworms were removed from the intestines of 59 red foxes using the sedimentation and counting technique¹⁷. The foxes were hunted between 2022 and 2024 in Pomorskie Voivodship (Puck, Wejherowo, Kartuzy, and Kościerzyna Districts, treated as one area), Warmińsko-Mazurskie Voivodship (Bartoszyce, Gołdap, Kętrzyn, and Węgorzewo Districts), and Podlaskie Voivodship (Augustów District) (Figure 1). The foxes were obtained legally in collaboration with the Polish Hunting Association and the General Directorate of State Forests. The research project and all experimental protocols received approval from the Medical University of Gdańsk Bioethics Committee for Scientific Research (decision no. NKBBN/467/2021). All methods were carried out in accordance with relevant guidelines and regulations. For sampling details, see¹⁸ and¹⁶. Tapeworms were washed in physiological saline. Total genomic DNA was extracted from whole specimens using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany, cat. no. 69506), following the manufacturer's protocol.

PCR amplification and sequencing of the complete *cox1* gene

We processed up to five randomly selected tapeworms per red fox: the same specimens as in¹⁶. PCRs to amplify the complete coding sequence of the cytochrome c oxidase subunit 1 gene (1,608 bp) were performed in a 25- μ l reaction mixture containing 10–12 ng of template DNA, 12.5 μ l of DreamTaq™ Hot Start Green PCR Master Mix (Thermo Scientific™, cat. no. K9021), 0.4 μ M of primers EmcF and EmcR¹⁹, and 0.75 μ l of DMSO (final concentration 3% v/v). PCR conditions were as follows: 95 °C for 3 min; 35 cycles of 95 °C for 30 s, 56.5 °C for 30 s, 72 °C for 1.5 min; and 72 °C for 15 min.

The selected marker is notorious for the presence of a poly-T region (positions 207–218), which causes polymerase slippage that leads to read corruption during sequencing^{5,14}. To overcome this issue, we designed

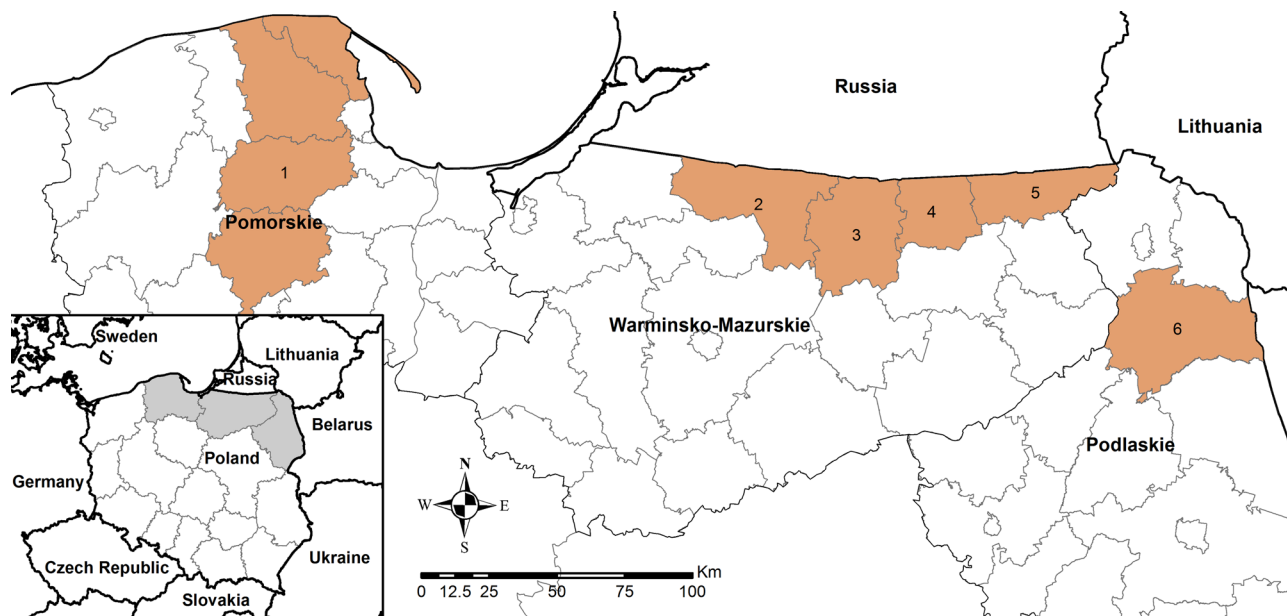


Fig. 1. Map of the study area. Districts where red foxes were hunted are shown in orange: 1 – Puck, Wejherowo, Kartuzy, Kościerzyna; 2 – Bartoszyce; 3 – Kętrzyn; 4 – Węgorzewo; 5 – Gołdap; 6 – Augustów.

the reverse sequencing primer 10068R (5'-CCGTCTTCACATCCAACCCA-3'), which binds mid-gene and facilitates assembly of the complete coding sequence.

Amplicons were purified with Clean-Up Concentrator (A&A Biotechnology, Gdańsk, Poland, cat. no. 021-250C). Bidirectional sequencing was outsourced to MacroGen Europe (Amsterdam). Regions on both ends with more than a 1% chance of an error per base were automatically trimmed in Geneious Prime 2025.0.3 (<https://www.geneious.com/>). Forward and reverse strands were assembled into consensus sequences and checked for internal stop codons to ensure that orthologs were amplified. Species identity was confirmed with BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Population genetic analysis

The following measures of genetic variation were calculated for the sampled regions: number of segregating sites (S), number of haplotypes (h), haplotype diversity (H_d ; the probability that two randomly chosen haplotypes are different²⁰), average number of nucleotide differences between two randomly chosen sequences (K), and nucleotide diversity (P_i ; the average number of nucleotide differences per site between two randomly chosen sequences²¹). Haplotypes were defined when at least one mutation occurred between two samples.

To quantify genetic differentiation, the fixation index (F_{ST}) was calculated between two haplogroups, European and Asian. To assess whether the observed genetic differentiation was greater than expected by chance under panmixia, haplogroup labels were shuffled among sequences 1,000 times, and genetic differentiation statistics (H_S/H_{ST} , K_S/K_{ST} , K_S^*/K_{ST}^* , Z , Z^* , and S_{nn}) were calculated for each permutation and compared with their observed values. All calculations were done in DnaSP 6.12.03²².

A histogram of nucleotide site differences between pairs of individuals (mismatch distribution) was generated using the packages 'adegenet'^{23,24} and 'pegas'²⁵ in R version 4.3.2²⁶.

Two median-joining (MJ) haplotype networks ($\epsilon = 0$)²⁷ were created in PopART (<https://popart.maths.otago.ac.nz/>) and graphically refined in Inkscape (<https://inkscape.org/about/>): one for the PCR dataset only, and one for the PCR dataset combined with complete *cox1* sequences of Asian descent and records from North America and the Arctic deposited in GenBank (Supplementary File 1).

Timetree

To estimate the divergence time of the European and Asian haplogroups in Poland, a timetree was constructed using MEGA12 version 12.0.9²⁸. First, haplotypes observed in the PCR dataset were combined with other haplotypes detected in Poland^{13,14} and complete coding sequences of the *cox1* gene from *Echinococcus* species other than *E. multilocularis*, with *Taenia solium* serving as an outgroup²⁹. All sequences were aligned using MUSCLE by translating codons to amino acids, performing the alignment, and then replacing the amino acids with the original codons.

The best nucleotide substitution model for constructing a Maximum Likelihood (ML) tree was determined using the ML statistical method and selected based on the Akaike Information Criterion and Bayesian Information Criterion. The robustness of the tree was tested by bootstrapping with 1000 replicates.

The topology of the ML tree was used to estimate a timetree by employing the RelTime method. Molecular dating was calibrated using the species pair *Echinococcus oligarthrus* and *Echinococcus vogeli*. The time to the most recent common ancestor (TMRCA) of these species was modelled as a normal distribution centred at 3.0 million years ago with a standard deviation of 0.3 million years (95% confidence interval: 2.41–3.59 million years), as in³⁰.

This divergence time calibration constraint corresponds to the Great American Biotic Interchange, an event during which canid and felid ancestors of the modern endemic *E. oligarthrus* and *E. vogeli* definitive hosts migrated from North America to South America, following the formation of the Panamanian land bridge^{29,31}.

Comparison with nuclear data

In our previous article, we investigated the same set of 263 adult tapeworms using a nuclear marker, the microsatellite EmsB¹⁶. The study aimed to explore the diversity of EmsB profiles of *E. multilocularis* from red foxes and humans, with a particular focus on Asian variants. The profiles were hierarchically clustered into European and Asian units.

The results obtained in the current study were compared to those reported in¹⁶ to determine the continent of origin of the mitochondrial and nuclear genomic elements of each tapeworm. For the microsatellite profiling results used in this comparison, see Additional File 3 appended to¹⁶.

Alignment-free approach

We introduce a novel computational approach: the 4D-Dynamic Representation of DNA/RNA Sequences combined with K -means clustering. The 4D-Dynamic Representation of DNA/RNA Sequences is an alignment-free bioinformatics method recently developed by us³². DNA or RNA sequences are transformed into sets of 'material points' in a 4D space, forming 4D-dynamic graphs. These graphs are treated as 'rigid bodies' and characterised by descriptors analogous to those used in classical dynamics, such as centres of mass and principal moments of inertia. For visualization, 2D and 3D projections of these graphs are employed^{32,33}.

This bioinformatics method was integrated with K -means clustering, an unsupervised machine learning algorithm widely used in artificial intelligence. Previously, K -means clustering was combined with another of our bioinformatics approaches³⁴. Here, we present cluster plots generated by applying K -means clustering to the descriptors derived from the 4D-Dynamic Representation of DNA/RNA Sequences.

Results

PCR amplification and sequencing of the complete *cox1* gene

We obtained complete coding sequences of the *cox1* gene for 252 out of 263 tapeworms from 59 red foxes. PCR was repeated for 11 samples but yielded no product.

Population genetic analysis

Nine segregating sites in total allowed us to identify six haplotypes, three of which were absent from previous studies (Table 1). The highest values of all genetic diversity estimators were observed in the districts of Warmińsko-Mazurskie Voivodship, with the most haplotypes detected in Kętrzyn District and the highest haplotype diversity in Bartoszyce District. The lowest overall genetic diversity was noted in Augustów District (Table 2). The mismatch distribution was bimodal (Figure 2). The value of the fixation index (F_{ST}) between two haplogroups: European ($n = 221$, excluding haplotypes Hap_2 and Hap_6) and Asian ($n = 31$, composed of haplotypes Hap_2 and Hap_6), equalled 0.97057. The observed genetic differentiation between the two haplogroups was greater than expected by chance under panmixia, as assessed by permutation tests ($\alpha = 0.05$, all $P < 0.001$).

Hap_1 was the most numerous and widespread haplotype (77%). It formed a starburst pattern (a central haplotype surrounded by derivatives) with Hap_3, Hap_4, and Hap_5. The second most-represented variant was Hap_2 (12%), present in all four districts of Warmińsko-Mazurskie Voivodship. Hap_2 and Hap_6 formed a satellite haplogroup separated from other haplotypes by five mutations (Figure 3).

In the intercontinental MJ haplotype network of a total of 433 sequences, Polish haplotypes blended into a reticulated pattern of Asian haplotypes represented by *E. multilocularis* from the Chinese province of Xinjiang (Figure 4), filling the five-mutation distance between Hap_2 and Hap_3 depicted in Figure 3. A Mongolian haplogroup was formed 18 mutations apart from the closest, Chinese (Xinjiang), haplotype. Some of these parasites originated from red and corsac foxes, a wolf, and a human from Hulunbuir (Inner Mongolia), Ulaanbaatar, and an unspecified area of Mongolia. Others came from Russian flat-headed voles collected on Olkhon Island on Lake Baikal (Irkutsk Oblast) and in the Altai Republic (Figure 4).

Haplogroups centred around Hap_16 and Hap_2 created a starburst pattern, with Hap_6 from two red foxes hunted in Kętrzyn District forming one of the rays of Hap_2. Hap_16 contained mainly Japanese, Russian, and Kyrgyz samples. Hap_2 comprised predominantly Polish samples—including tapeworms from previous studies^{12,13}—as well as four Chinese samples (Qinghai, Sichuan, Xinjiang), seven Japanese adult tapeworms, and one Russian metacestode from the Altai region (Figure 4).

Sequences from Anabarsky and Nizhnekolymsky Districts (Russia, Yakutia), Svalbard (Norway), Alaska/St. Lawrence Island (USA)—regarded as Arctic regions, but also the states of South Dakota and Indiana (USA), were indirectly connected to haplotypes from Xinjiang (China), through missing haplotypes. Six Canadian haplotypes, from the provinces of British Columbia, Alberta, and Saskatchewan, formed rays stemming from the predominantly Polish Hap_3, which also included two samples from the state of Missouri. The remaining three samples from Canada and four samples from St. Lawrence Island belonged to Hap_16 (Figure 4).

Timetree

European and Asian haplotypes formed distinct lineages in the ML tree (Figure 5). The divergence time of the European and Asian haplogroups was estimated to be 26,050 years ago (Figure 6).

Comparison with nuclear data

Results per tapeworm and per red fox are presented in Supplementary File 2.

Results per tapeworm: Eighty per cent of the 252 tapeworms were of European origin, as indicated by the nuclear microsatellite marker *EmsB* and the mitochondrial marker *cox1*. Tapeworms of Asian descent constituted 6% of the total worm pool. Thirty-five parasites were of mixed origin (14%), i.e., exhibited different origin in mitochondrial and nuclear DNA. The proportion of Asian tapeworms in northern and northeastern Poland (15/252) was comparable to the proportion calculated for the whole country (13/349 from¹²) (Fisher's exact test for count data, $\alpha = 0.05$, $P = 0.2402$). However, the proportion of Euro-Asian hybrids was substantially higher in the study area (35/252) than at the national level (17/349 from¹²) (Fisher's exact test for count data, $\alpha = 0.05$, $P = 0.0002$).

Results per red fox: Five worms per fox were both genotyped and profiled for 41 red foxes. Twenty-three red foxes (56%) harboured European tapeworms with the most common *EmsB* pattern and haplotype, Pol-B/PH-2 and Hap_1, respectively. Two red foxes from Bartoszyce District contained only Asian tapeworms (5%). All five tapeworms from four red foxes—two from Gołdap District, one from Węgorzewo District, and one from Pomorskie Voivodship—were Euro-Asian hybrids (10%). Two red foxes from Bartoszyce District and two from Kętrzyn District were infected with European and hybrid tapeworms. One red fox from Gołdap District and one from Kętrzyn District were infected with European and Asian tapeworms. One red fox from Kętrzyn District contained tapeworms of European, Asian, and mixed origin.

4D-Dynamic representation of DNA/RNA Sequences combined with K-means clustering

Biological sequences are complex objects, making it uneasy to represent their structure and to define a similarity measure. The similarity space for such objects is multidimensional; only simple, one-dimensional entities can be classified uniquely using a single similarity metric. Complex objects may appear similar in one respect yet exhibit significant differences in another. Descriptors reflecting different aspects of similarity can be treated separately or combined into a single composite measure. A unified measure often results from either averaging across multiple aspects or disregarding most of them. Each bioinformatics method highlights distinct facets of similarity, and novel approaches continue to emerge. Standard alignment-based methods can overlook subtle,

Haplotype	Origin	Frequency [no.]	Frequency [%]	Reference	139	182	220	237	289	364	368	482	527	675	688	742	822	1432	1492	1514
Hap_1 (EmPL_cox_A)	European	194/252	77	13,14	C	T	A	C	C	G	C	G	A	G	T	C	A	G	A	G
Hap_4 (EmPL_cox_A2)	European	4/252	2	This study	G	.
Hap_5 (EmPL_cox_A3)	European	4/252	2	This study	T
Hap_3 (EmPL_cox_B)	European	19/252	8	13,14	A
EmPL_cox_B2	European	Not detected	Not detected	14	.	.	G	A
EmPL_cox_C	European	Not detected	Not detected	13	.	G	A
EmPL_cox_D	European	Not detected	Not detected	13	A	A
Hap_2 (EmPL_cox_E)	Asian	29/252	12	13	.	.	.	T	T	T	C	.	G	.	.	A
EmPL_cox_E2*	Asian	Not detected	Not detected	14	.	.	.	T	T	T	C	.	G	A	.	A
Hap_6 (EmPL_cox_E3)	Asian	2/252	1	This study	.	.	.	T	T	T	C	A	G	.	.	A
EmPL_cox_F	European	Not detected	Not detected	13,14	T	A
EmPL_cox_G	European	Not detected	Not detected	13,14	G	A
EmPL_cox_G2	European	Not detected	Not detected	14	T	G	A

Table 1. Haplotypes of the cytochrome c oxidase subunit 1 gene (1,608 bp) observed in Poland, from this and previous studies, with their frequencies in the PCR dataset from the present study. Conventional haplotype labels are given in parentheses to maintain consistency across reports^{13,14}. *The 1401-bp sequence EmPL_cox_E(pig), accession OQ874676, from¹⁴.

District(s)	Voivodship	No. of Sequences	S	h	H_d	K	Pi
Augustów	Podlaskie	45	0	1	0.00000	0.00000	0.00000
Bartoszyce	Warmińsko-Mazurskie	54	6	3	0.60098	2.45982	0.00153
Goldap	Warmińsko-Mazurskie	53	7	4	0.44267	1.47750	0.00092
Kętrzyn	Warmińsko-Mazurskie	65	8	5	0.35481	1.27115	0.00079
Węgorzewo	Warmińsko-Mazurskie	18	6	2	0.42484	2.54902	0.00159
Puck, Wejherowo, Kartuzy, Kościerzyna	Pomorskie	17	1	2	0.22059	0.22059	0.00014
Total		252	9	6	0.38939	1.48103	0.00092

Table 2. Measures of *Echinococcus multilocularis* genetic variation in the cytochrome c oxidase subunit 1 locus (1,608 bp), calculated for the sampled districts. S – number of segregating sites, h – number of haplotypes, H_d – haplotype diversity, K – average number of nucleotide differences, Pi – nucleotide diversity.

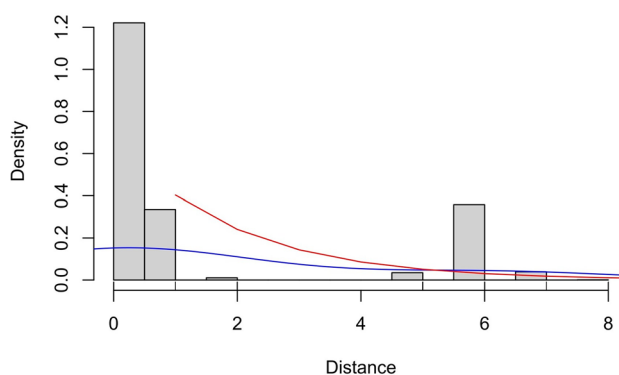


Fig. 2. Observed distribution of pairwise distances between the 252 complete coding sequences of the cytochrome c oxidase subunit 1 gene (1,608 bp) (mismatch distribution). The blue line shows the empirical density estimate, and the red line shows the expected distribution under a stable population model⁴⁶.

global patterns or topological distributions of the nucleobases within sequences. Our method is designed to quantify these higher-order properties, which are not directly accessible through standard similarity measures. Our goal is to uncover potential structural and organisational similarities that might be conserved even in the absence of strong sequence alignment.

The relationships between sequences are visualised in a non-standard manner in Figures 6 and 7. Figure 7 displays a cluster plot generated using the coordinates of the centres of mass of the 4D-dynamic graphs as input for the *K*-means clustering algorithm. Figure 8 presents an analogous cluster plot based on the principal moments of inertia tensor elements of the graphs (see³² for mathematical details).

Both plots reveal six distinct point clusters. Sequence relationships can be analysed by comparing the Euclidean distances between these clusters. Notably, the distances between three point clusters—Hap_1, Hap_3, and Hap_4—are small in both visualisations.

The new method enables examination of different aspects of sequence similarity. For instance, Hap_5 exhibits varying distances from the three aforementioned clusters in the two plots, highlighting distinct relational patterns. To align with conventional bioinformatics approaches, the two descriptors—the centre of mass and the principal moment of inertia—can be combined into a single metric, as demonstrated in³⁴, yielding an averaged similarity measure.

In the present work, to integrate the two descriptors, we extended the input dataset for the *K*-means clustering algorithm to include both the centres of mass coordinates and the principal moments of inertia tensor elements of the 4D-dynamic graphs representing the sequences. The resulting cluster plot is shown in Figure 9.

Additionally, the method enables the visual inspection of sequence similarity and the identification of regions with significant or minimal variation. As the method operates in a multidimensional (4D) space, the results are graphically represented as two- or three-dimensional projections of the 4D-dynamic graphs. Since the analysed sequences are highly similar, their differences are difficult to discern in 3D projections. We have therefore computationally identified specific sequence fragments where these distinctions are most pronounced. These critical regions are shown as 2D projections of the 4D-dynamic graphs and presented in Figure 10.

Discussion

In this study, we explored the *cox1* diversity of *E. multilocularis* from red foxes in northern and northeastern Poland, tested the hypothesis that the presence of Asian haplotypes in Poland is a consequence of their introduction from Northeast Asia to Eastern Europe, and compared the results with previously acquired EmsB

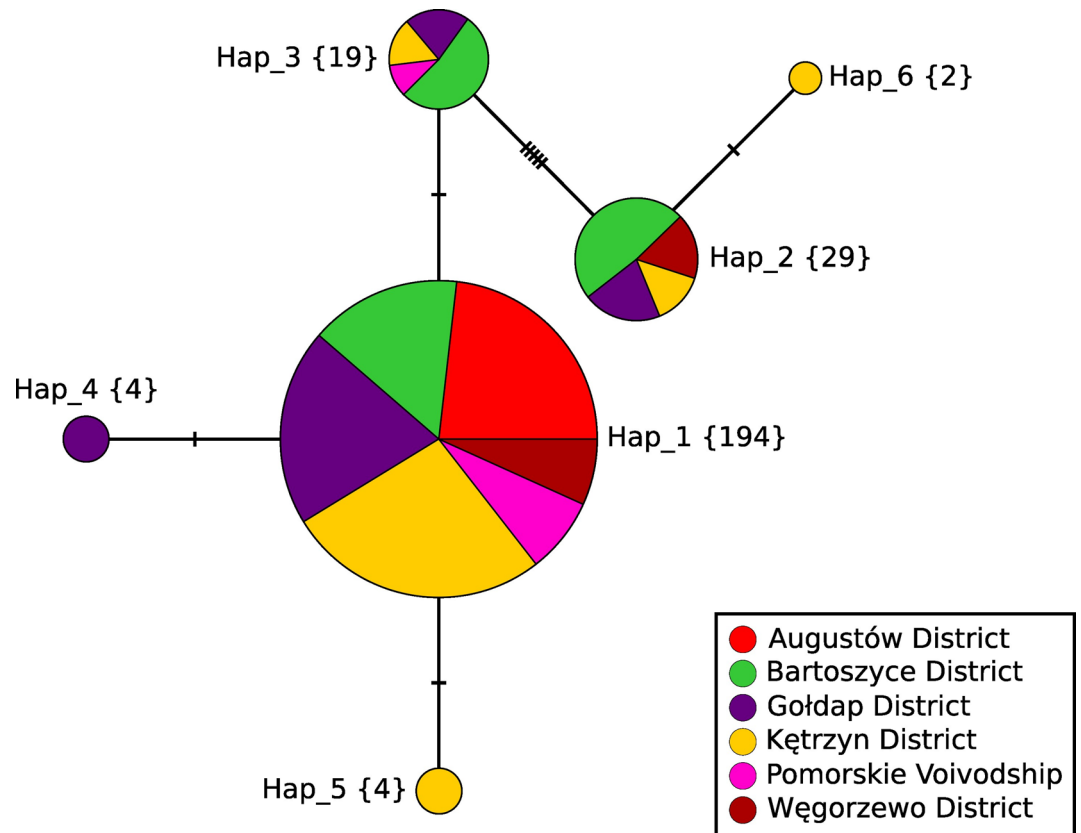


Fig. 3. Median-joining haplotype network ($\epsilon = 0$)²⁷ generated from the PCR dataset of the 252 complete coding sequences of the cytochrome c oxidase subunit 1 gene (1,608 bp). Haplotypes were defined when at least one mutation occurred between two samples. The number of samples representing each haplotype is given in curly brackets. The number of mutations is shown as hatch marks.

data¹⁶. We used standard population genetic and phylogenetic tools and a novel approach to sequence analysis, the 4D-Dynamic Representation of DNA/RNA Sequences combined with K -means clustering.

The latter successfully demonstrated differences between the six observed haplotypes using physical descriptors of DNA sequences. Within this model, DNA sequences are represented as clouds of material points in a four-dimensional space (referred to as 4D-dynamic graphs), with each sequence yielding a unique spatial distribution. This abstract representation enables their description using formulas analogous to those of classical mechanics. To numerically characterise these 4D-dynamic graphs, we employ quantities equivalent to dynamical descriptors. In classical mechanics, common measures for describing the mass distribution of a rigid body include the centre of mass and the moment of inertia. The moment of inertia about a given axis quantifies how mass is distributed relative to that axis; the moments calculated about the principal axes are known as the principal moments of inertia. We extend this concept to describe distributions of other quantities: by assigning a conceptual ‘mass’ to each point in the 4D-dynamic graph, we define a graph moment of inertia analogous to its physical counterpart.

This method provides a fundamentally different and complementary framework for analysing sequence relationships. A key outcome of this analysis is that our approach independently confirms the major global patterns observed in networks constructed using standard methods. This mutual validation both strengthens the established results presented in this work and underscores the relevance of our novel descriptors.

A considerably larger sampling size per area than in other studies^{12,13,35} allowed us to detect rare haplotypes, such as Hap_4 (4/252, 2%), Hap_5 (4/252, 2%), and Hap_6 (2/252, 1%). It also elucidated the intestinal genetic diversity of *E. multilocularis* in local red foxes. We observed the highest worm diversity per individual fox near Kaliningrad Oblast, likely due to the prevalence of *E. multilocularis* in red foxes, ranging from 26.0% to 52.5% along the Polish-Russian border¹⁸. Small mammals being the main source of infection for canids, this finding raises the question of metacestode diversity and availability in prey in local transmission hotspots.

The proportion of Euro-Asian hybrids was higher in northern and northeastern Poland than at the national level. This difference could be attributed to a relatively high metacestode diversity in available prey, resulting in more frequent mixed infections among red foxes. Because wildlife does not respect administrative boundaries, the influence of adjacent Russian and Lithuanian populations needs to be considered to account for the observed genetic richness and hybrid frequency.

The median-joining haplotype network in Figure 3 revealed the demographic history of the investigated metapopulation. The starburst pattern—with the most-represented core haplotype Hap_1 and three less-

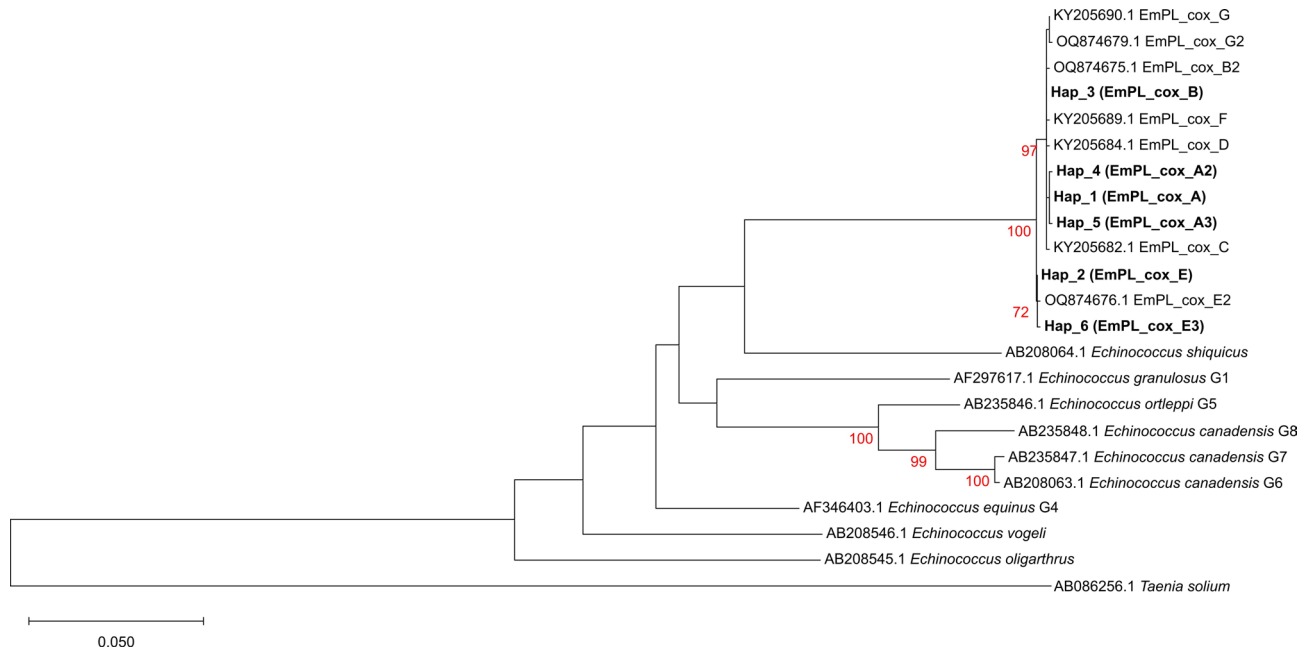


Fig. 5. Phylogeny inferred using the Maximum Likelihood method and the Tamura-Nei model of nucleotide substitutions⁴⁷. The tree with the highest log likelihood (-6,104.75) is shown. Bootstrap support values greater than 70% (1,000 replicates) are shown in red below the branches⁴⁸. The initial tree for the heuristic search was selected based on superior log-likelihood between a Neighbour-Joining (NJ) tree⁴⁹ and a Maximum Parsimony (MP) tree. The NJ tree was generated using a matrix of pairwise distances computed using the Tamura-Nei model⁴⁷. The MP tree had the shortest length among 10 MP tree searches, each initiated with a randomly generated starting tree. Evolutionary rate variation among sites was modelled using a discrete Gamma distribution with 5 categories (+G, parameter = 0.8023), assuming that 50.02% of sites were evolutionarily invariant (+I). The analysis included 23 coding nucleotide sequences (total length 1,683 bp) and involved 1st, 2nd, and 3rd codon positions, as well as non-coding positions. For details on haplotype labelling, see Table 1.

of the species into Eastern Europe⁵, or by other wild canids after earlier deposition of Asian variants further east by raccoon dogs. Current European raccoon dog populations are descendants of southeastern Siberian individuals of the subspecies *Nyctereutes procyonoides ussuriensis* Matschie, 1907, transported to European parts of the Soviet Union between the late 1920s and the late 1950s^{39,40}. Their predecessors inhabited mainly the Amur and Ussuri river valleys, the shores of the Sea of Japan, and eastern China. Initially bred for their thick fur, the animals were released into the wild. They spread rapidly, and in 1955, after crossing the Polish-Belarusian border, a raccoon dog was spotted in Narewka Forest District, in what is now Podlaskie Voivodship. The same year, another individual was observed in Strzelce Forest District, over the Polish-Ukrainian border, in modern-day Lubelskie Voivodship. It took raccoon dogs little time to enter inland and reach as far as East Germany, where they were first spotted in the early 1960s⁴⁰.

If this explanation is correct, introduced *E. multilocularis* lineages still closely resemble their ancestral populations in Asia. Tapeworm clustering in phenograms and networks appears to support this view in light of historical records but with limitations. Formally, the available Chinese reference sequences come from West-Central China, and we lack genetic information from Northeast Asia. Still, the general undersampling of the Asian continent leaves us in the dark about the geographical range of Hap_2 and Hap_6, both of which may be widespread enough to be present in Northeast Asia.

Human-mediated introduction of *E. multilocularis* is not unprecedented. Major and minor haplogroups in Hokkaido likely result from animal introductions from St. Lawrence Island and Sichuan, respectively, as supported by historical evidence and the aforementioned mitogenomic study³⁸. Similarly, multiple events of transportation of red foxes and/or domestic dogs from continental Europe are the most likely reason for the invasion of European genotypes into Canada^{41,42} and the USA^{43–45}, as reflected by the composition and configuration of Hap_3 and Hap_60–64 in Figure 4. With more regional research like ours in the future, preferably based on whole mitogenomes, and insights from historical records, it may be possible to trace geographically aberrant lineages back to their source and explain their journey.

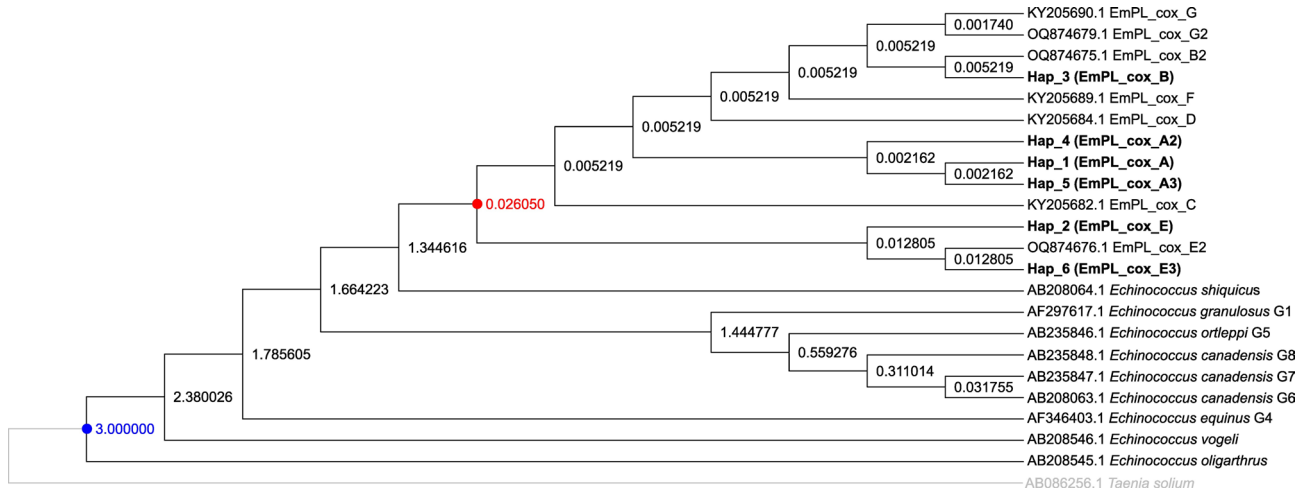


Fig. 6. Timetree inferred by the RelTime method^{50,51} using the Tamura-Nei model of nucleotide substitutions⁴⁷. Branch lengths were computed using the Maximum Likelihood method⁵². Evolutionary rate variation among sites was modelled using a discrete Gamma distribution across five categories (+G, parameter = 0.8023), assuming that 50.02% of sites were evolutionarily invariant (+I). The RelTime analysis incorporated a single calibration to derive minimum and maximum bounds for the calibrated node (marked in blue in the tree)⁵³. Divergence time estimates in millions of years are shown next to nodes in the tree. Notably, times are not estimated for nodes in the outgroup because the RelTime method does not assume that evolutionary rates are the same for ingroup and outgroup lineages⁵¹. The divergence time of the European and Asian haplogroups is marked in red. The analysis included 23 coding nucleotide sequences (total length 1,683 bp) and involved 1st, 2nd, and 3rd codon positions, as well as non-coding positions. For details on haplotype labelling, see Table 1.

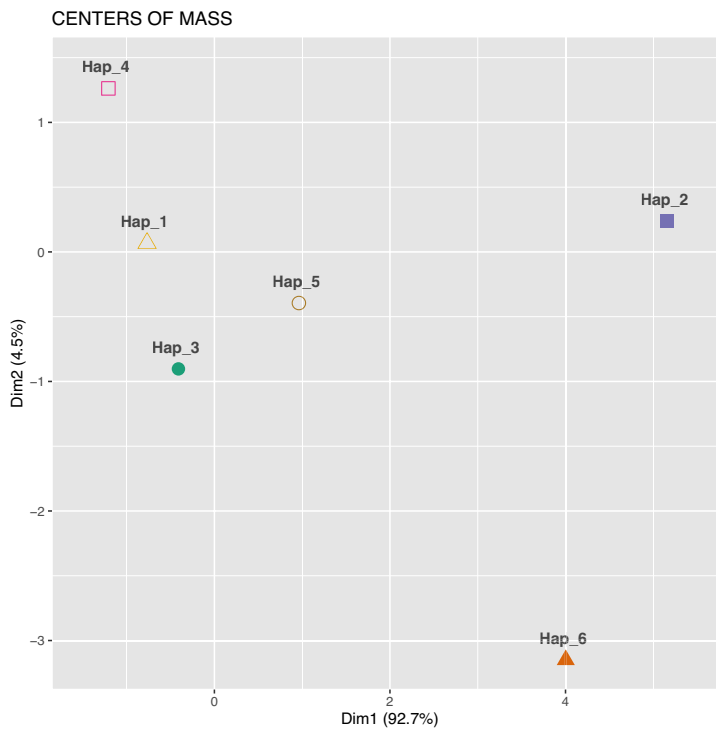


Fig. 7. Cluster plot obtained using *K*-means clustering, with the centre of mass coordinates of the 4D-dynamic graphs as input data. (The computational framework for the 4D-dynamic method was developed by our team; the implementation and *K*-means clustering were performed in R⁵⁴).

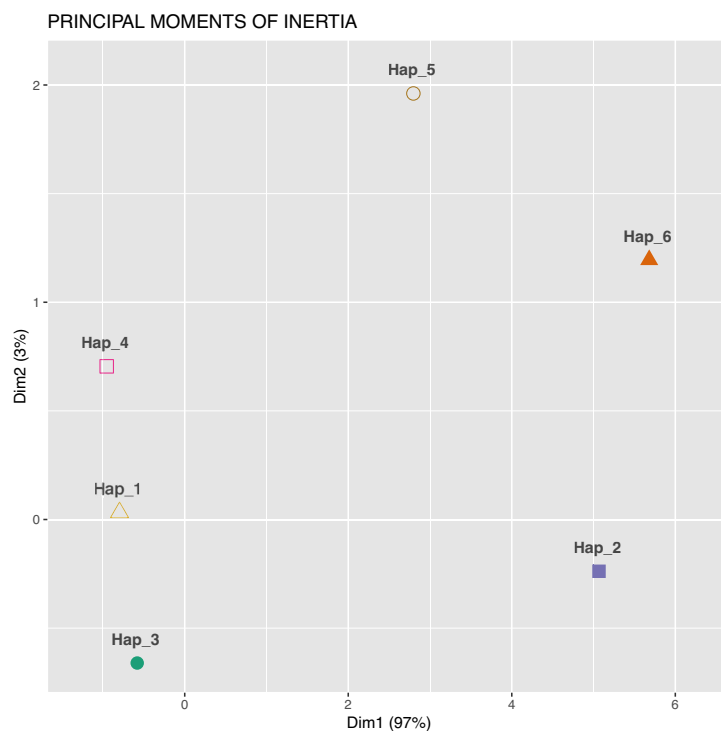


Fig. 8. Cluster plot obtained using K -means clustering, with the principal moments of inertia tensor elements of the 4D-dynamic graphs as input data.

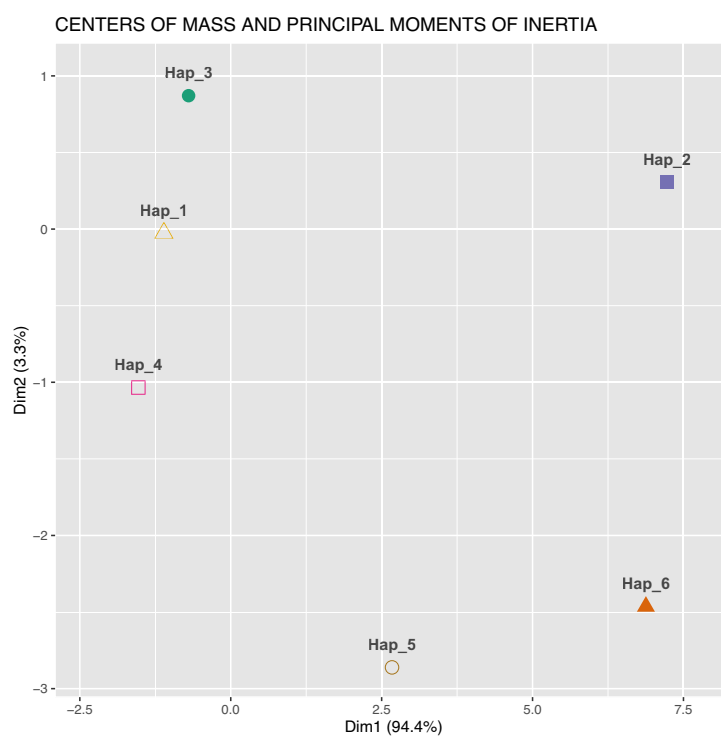


Fig. 9. Cluster plot obtained using K -means clustering, with both the centres of mass coordinates and the principal moments of inertia tensor elements of the 4D-dynamic graphs as input data.

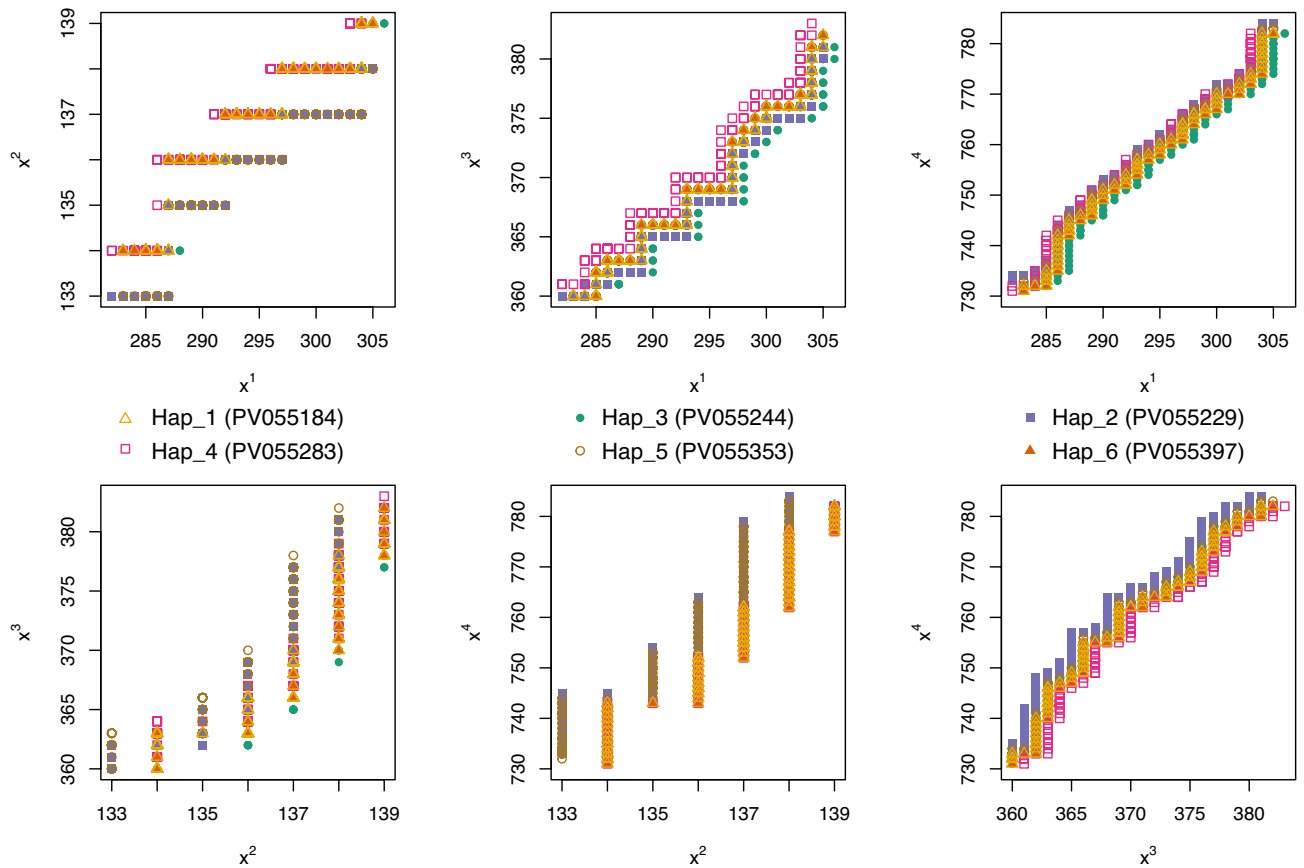


Fig. 10. 2D projections of the 4D-dynamic graphs representing fragments of the sequences.

Data availability

The 252 complete coding sequences of the cytochrome c oxidase subunit 1 gene (1,608 bp) obtained in this study were submitted to GenBank and received accessions PV055184–PV055435.

Received: 16 October 2025; Accepted: 12 February 2026

Published online: 20 February 2026

References

- Sulima, M. et al. Liver transplantation for incurable alveolar echinococcosis: an analysis of patients hospitalized in department of tropical and parasitic diseases in gdynia. *Transplant. Proc.* **48**, 1708–1712 (2016).
- Wen, H. et al. Echinococcosis: advances in the 21st Century. *Clin. Microbiol. Rev.* **32**, 10 (2019).
- Knapp, J. et al. Genetic diversity of the Cestode *Echinococcus multilocularis* in red foxes at a continental scale in Europe. *PLoS Negl. Trop. Dis.* **3**, e452 (2009).
- Laurimaa, L. et al. First report of the zoonotic tapeworm *Echinococcus multilocularis* in raccoon dogs in Estonia, and comparisons with other countries in Europe. *Vet. Parasitol.* **212**, 200–205 (2015).
- Santoro, A. et al. Mitochondrial genetic diversity and phylogenetic relationships of *Echinococcus multilocularis* in Europe. *Int. J. Parasitol.* **54**, 233–245 (2024).
- McDevitt, A. D. et al. Next-generation phylogeography resolves post-glacial colonization patterns in a widespread carnivore, the red fox (*Vulpes vulpes*). *Europe. Mol. Ecol.* **31**, 993–1006 (2022).
- Dalén, L. et al. Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox. *Proc. Natl. Acad. Sci.* **104**, 6726–6729 (2007).
- Nakao, M., Sako, Y. & Ito, A. Isolation of polymorphic microsatellite loci from the tapeworm *Echinococcus multilocularis*. *Infect. Genet. Evol.* **3**, 159–163 (2003).
- Haag, K., Araújo, A., Gottstein, B. & Zaha, A. Selection, Recombination and History in a Parasitic Flatworm (*Echinococcus*) Inferred from Nucleotide Sequences. *Mem. Inst. Oswaldo Cruz* **93**, 695–702 (1998).
- Bartel, M., Seesee, F. & Worley, D. Comparison of Montana and Alaska isolates of *Echinococcus multilocularis* in gerbils with observations on the cyst growth, hook characteristics, and host response. *J. Parasitol.* **78**, 529–532 (1992).
- Konyaev, S. V. et al. Genetic diversity of *Echinococcus* spp. in Russia. *Parasitol.* **140**, 1637–1647 (2013).
- Uhang, G. et al. Asian admixture in European *Echinococcus multilocularis* populations: new data from Poland comparing EmsB microsatellite analyses and mitochondrial sequencing. *Front. Vet. Sci.* **7**, 620722 (2021).
- Karamon, J. et al. Genetic diversity of *Echinococcus multilocularis* in red foxes in Poland: The first report of a haplotype of probable Asian origin. *Folia Parasitol (Praha)* **64**, 10 (2017).
- Karamon, J. et al. *Echinococcus multilocularis* genetic diversity based on isolates from pigs confirmed the characteristic haplotype distribution and the presence of the Asian-like haplotype in Central Europe. *J. Vet. Res.* **67**, 567–574 (2023).

15. Gładysz, P. & Lass, A. Detection of Asian genetic components in autochthonous human *Echinococcus multilocularis* infections from endemic Warmia-Masuria (north-eastern Poland). *One Health* **17**, 100623 (2023).
16. Gładysz, P. et al. Genetic diversity of *Echinococcus multilocularis* from red foxes and humans in northern and northeastern Poland investigated using the microsatellite EmsB. *Infect. Genet. Evol.* **135**, 105843 (2025).
17. Hofer, S. et al. High prevalence of *Echinococcus multilocularis* in urban red foxes (*Vulpes vulpes*) and voles (*Arvicola terrestris*) in the city of Zürich Switzerland. *Parasitology* **120**, 135–142 (2000).
18. Gładysz, P., Świątańska, A., Samorek-Pieróg, M., Karamon, J. & Lass, A. An update on the prevalence of *Echinococcus multilocularis* in red foxes (*Vulpes vulpes*) in a Central European focus: Northern and northeastern Poland (2022–2024). *One Health* **21**, 101151 (2025).
19. Nakao, M. et al. Geographic pattern of genetic variation in the fox tapeworm *Echinococcus multilocularis*. *Parasitol Int.* **58**, 384 (2009).
20. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, 1987). <https://doi.org/10.7312/nei-92038>.
21. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci.* **76**, 5269–5273 (1979).
22. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
23. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
24. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
25. Paradis, E. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420 (2010).
26. R Core Team. R: A Language and environment for statistical computing. Preprint at <https://www.R-project.org/> (2023).
27. Bandelt, H. J., Forster, P. & Rohlf, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
28. Kumar, S. et al. MEGA12: molecular evolutionary genetic analysis version 12 for adaptive and green computing. *Mol. Biol. Evol.* **41**, msae263 (2024).
29. Nakao, M., McManus, D. P., Schantz, P. M., Craig, P. S. & Ito, A. A molecular phylogeny of the genus *Echinococcus* inferred from complete mitochondrial genomes. *Parasitology* **134**, 713–722 (2006).
30. Knapp, J. et al. Phylogenetic relationships within *Echinococcus* and *Taenia* tapeworms (Cestoda: Taeniidae): An inference from nuclear protein-coding genes. *Mol. Phylogenet. Evol.* **61**, 628–638 (2011).
31. Marshall, L. G., Webb, S. D., Sepkoski, J. J. & Raup, D. M. Mammalian Evolution and the Great American Interchange. *Science* **197**(215), 1351–1357 (1982).
32. Bielińska-Wąż, D. & Wąż, P. Non-standard bioinformatics characterization of SARS-CoV-2. *Comput. Biol. Med.* **131**, 104247 (2021).
33. Bielińska-Wąż, D., Wąż, P., Lass, A. & Karamon, J. 4D-dynamic representation of DNA/RNA sequences: studies on genetic diversity of *Echinococcus multilocularis* in red foxes in Poland. *Life* **12**, 877 (2022).
34. Bielińska-Wąż, D., Wąż, P. & Błaczowska, A. 20D-dynamic representation of protein sequences combined with K-means clustering. *Comb. Chem. High Throughput Screen* <https://doi.org/10.2174/0113862073359729250220131623> (2025).
35. Umhang, G. et al. A step forward in the understanding of the presence and expansion of *Echinococcus multilocularis* in Eastern Europe using microsatellite EmsB genotyping in Poland. *Infection Genet. Evol.* **54**, 176–182 (2017).
36. Avise, J. C. *Phylogeography: the history and formation of species* (Harvard University Press, UK, 2000).
37. Guo, B. et al. The genetic variation of mitochondrial sequences and pathological differences of *Echinococcus multilocularis* strains from different continents. *Spectr. Microbiol.* <https://doi.org/10.1128/spectrum.01318-24> (2025).
38. Hayashi, N. et al. Mitogenomic exploration supports the historical hypothesis of anthropogenic diffusion of a zoonotic parasite *Echinococcus multilocularis*. *iScience* **26**, 107741 (2023).
39. Kauhala, K. & Kowalczyk, R. Invasion of the raccoon dog *Nyctereutes procyonoides* in Europe: History of colonization, features behind its success, and threats to native fauna. *Curr. Zool.* **57**, 584–598 (2011).
40. Nowak, E. & Pielowski, Z. Die Verbreitung des Marderhundes in Polen im Zusammenhang mit seiner Einbürgerung und Ausbreitung in Europa. *Acta Theriol. (Warsz)* **9**, 81–110 (1964).
41. Gesy, K., Hill, J. E., Schwantje, H., Liccioli, S. & Jenkins, E. J. Establishment of a European-type strain of *Echinococcus multilocularis* in Canadian wildlife. *Parasitology* **140**, 1133–1137 (2013).
42. Massolo, A. et al. European *Echinococcus multilocularis* Identified in Patients in Canada. *N. Engl. J. Med.* **381**, 384–385 (2019).
43. Santa, M. A., Musiani, M., Ruckstuhl, K. E. & Massolo, A. A review on invasions by parasites with complex life cycles: the European strain of *Echinococcus multilocularis* in North America as a model. *Parasitology* **148**, 1532–1544 (2021).
44. Conlon, C. L., Schuler, K. L., Lejeune, M. & Whipps, C. M. Novel Report of the European Variant of *Echinococcus multilocularis* in Coyotes (*Canis latrans*) in New York State. *Journal of Parasitology* **109**, 357 (2023).
45. Conlon, C. L. et al. *Echinococcus multilocularis* in New York Wildlife: distribution and genetic diversity of an emerging pathogen. *J. Parasitol.* **110**, 697 (2024).
46. Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569 (1992).
47. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).
48. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (N Y)*. **39**, 783–791 (1985).
49. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Evol. Biol.* <https://doi.org/10.1093/oxfordjournals.molbev.a040454> (1987).
50. Tamura, K. et al. Estimating divergence times in large molecular phylogenies. *Proc. Natl. Acad. Sci.* **109**, 19333–19338 (2012).
51. Tamura, K., Tao, Q. & Kumar, S. Theoretical foundation of the reltime method for estimating divergence times from variable evolutionary rates. *Mol. Biol. Evol.* **35**, 1770–1782 (2018).
52. Nei, M. & Kumar, S. *Molecular evolution and phylogenetics* (Oxford University Press, 2000).
53. Tao, Q., Tamura, K., Mello, B. & Kumar, S. Reliable confidence intervals for reltime estimates of evolutionary divergence times. *Mol. Biol. Evol.* **37**, 280–290 (2020).
54. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at <https://www.R-project.org/> (2024).

Acknowledgements

The authors would like to express their utmost gratitude to Adrianna Kilikowska, WSB Merito University in Gdańsk, for insightful comments and suggestions that enabled us to improve the original draft. Many thanks to Ioannis Kontogeorgos for creating Figure 1.

Author contributions

P. Gładysz : Conceptualisation, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualisation, Writing – Original Draft, Writing – Review & Editing. D. Bielińska-Wąż : Formal Analysis, Methodology, Software, Visualisation, Writing – Review & Editing. P. Wąż : Formal Analysis, Methodology, Software,

Visualisation, Writing – Review & Editing. M. Samorek-Pieróg : Resources. J. Karamon : Resources, Writing – Review & Editing. A. Lass : Conceptualisation, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Review & Editing.

Funding

This work was supported by the National Science Centre grant no. 2020/37/B/NZ7/03934.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-40313-z>.

Correspondence and requests for materials should be addressed to P.G. or A.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026