Tick Genome Assemblies: Overcoming biological limitations through advances in sequencing and assembly

Katie C. Dillon[*1], Julia C. Frederick[2], Hein Sprong[3], Travis C. Glenn[*1,2] and Isobel Ronai[4-6]

[1]Institute of Bioinformatics, University of Georgia, Athens, GA, USA
[2]Department of Environmental Health Science, University of Georgia, Athens, GA, USA
[3]Centre for Infectious Diseases, National Institute of Public Health and Environment, Bilthoven, The Netherlands
[4]University of Sydney, Sydney, NSW, Australia
[5]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA
[6]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA
[*]Corresponding authors

Further author information

Katie C. Dillon	katie.dillon@uga.edu / 0009-0005-7125-5793
Julia Frederick	julia.frederick@uga.edu / 0000-0003-4499-8688
Hein Sprong	hein.sprong@rivm.nl / 0000-0002-0218-4320
Travis Glenn	travisg@uga.edu / 0000-0001-7725-3637
Isobel Ronai	ironai@fas.harvard.edu / 0000-0002-1658-857X

RUNNING TITLE: Tick Genomes

ABSTRACT

Ticks are blood-feeding arthropods with approximately 1,000 species, however, only 24 species currently have a genome assembly. These genome assemblies are important resources to advance tick biology and control of tick-associated diseases. Generating tick genome assemblies is challenging due to their small body size (low DNA input), large genome size (approximately 2.4 Gbp for hard ticks), DNA contamination (from microbiota and host bloodmeals), and abundant transposable elements. Advances in sequencing technologies and assembly software have facilitated an increasing number of tick assemblies from 2011 to 2025. We characterize and assess the 54 tick genome assemblies within public genome databases using QUAST-LG and BUSCO compleasm. Then we evaluate the impact on these tick genome assemblies of biological source material, sequencing platform, and gene and repetitive element annotation. We identify 34 high-quality assemblies from 21 species that are well-suited for a variety of downstream analyses. Future tick genome assemblies should use long-read sequencing and Hi-C scaffolding to improve genomic resources for these unique blood-feeding parasites.

KEY WORDS: Ixodida, Ixodidae, Argasidae, vector biology, genetics, RNA-seq

THE IMPORTANCE OF TICKS

Ticks are hematophagous mites with a global health impact on humans, domesticated animals including livestock, and wildlife. As blood-feeders, ticks are disease vectors, transmitting a diverse range of disease agents, including viruses (for example, tick-borne *Flaviviridae*, *Dabie bandavirus*, and *Orthonairovirus*), bacteria (for example, *Borrelia* spp., *Anaplasma* spp. and Spotted Fever Group *Rickettsiae*), protozoa (for example, *Babesia* spp., *Theileria* spp.), and filarial nematodes[1-5]. The occurrence and distribution of these tick-associated diseases are increasing due to urbanization and climate changes[6,7]. Furthermore, ticks are developing resistance to current control strategies[8-11].

Ticks (order Ixodida) are comprised of three extant families: hard ticks (Ixodidae), soft ticks (Argasidae), and Nuttalliellidae. While there are approximately 1,000 valid tick species[12,13], few have genome assemblies available on databases such as the National Center for Biotechnology Information (NCBI). This lack of whole genome sequencing, which determines the complete DNA sequence for all chromosomes, is in part due to the unique biology of ticks.

Tick Biology Considerations for Genome Sequencing

The number of autosomes varies across tick species from 12 to 28 autosomes (diploid)[14-20] and the sex chromosome system is typically XY or XO[18]. Key aspects of tick biology should be taken into consideration prior to whole-genome sequencing and assembly.

Ticks have large genomes that are more similar in size to mammalian genomes than other vector species. We identified 17 genome assemblies of human disease vector species, spanning six taxonomic orders, from the World Health Organization's list of major human vector-borne diseases globally (https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases). For these 17 vector assemblies, the five largest all belong to ticks (Fig. 1). These tick assemblies have an average size that is up to 19 times larger than the other vector species, such as the human body louse (Fig. 1). The large genome-size of ticks has hindered the generation of genome assemblies, as this requires large amounts of sequencing and bioinformatic resources to achieve high

quality genomes with sufficient genome coverage (the number of times a single nucleotide has been sequenced).

Tick genomes contain numerous highly repetitive regions, with transposable elements (mobile DNA sequences that move around the genomes[21-23]) making up almost 70% of some genomes[23]. This proportion is amongst the highest observed in animals[12,24,25]. The high proportion of transposable elements in ticks increases genome size and complexity, greatly complicating genome assembly.

The body size of a tick is relatively small for an animal; the largest life stage (adult) and larger sex (female) is typically only a few millimeters when unengorged[26]. This small amount of tissue leads to a low total DNA yield (typically less than 6,000 ng) from a single individual[23,27-29]. As tick genomes are large and highly repetitive, relatively large amounts of high molecular weight DNA are required to meet optimal sample (library) preparation minimums for genome sequencing that were originally established for the generation of vertebrate genomes. Tick assemblies from individuals initially needed to rely on amplification-based sample preparation[29]. In 2025, the minimum amount of high molecular weight genomic DNA needed for non-amplification-based sample preparation can be met by single-tick DNA extractions, such as Oxford Nanopore 1,000 ng (https://store.nanoporetech.com/us/ligation-sequencing-kit-v14.html), and PacBio Standard HiFi Sequencing 500 ng (https://www.pacb.com/revio). Therefore, whole-genome sequencing of individual ticks has only become technically feasible within the last few years[23].

DNA extractions from ticks are particularly high in contaminants, as they contain DNA from their hosts and microbes[30]. Most tick species consume three blood meals throughout their lifecycle, often from multiple host species. In addition, these bloodmeals initiate the transfer of microbes from host vertebrate to tick[31]. Furthermore, tick species generally possess facultative and sometimes also obligatory endosymbionts[32]. This contaminating DNA from other organisms, as well as environmental contaminants means that genome sequencing of ticks generates many reads of non-tick origin[23,33] (for example, ~14% total reads for an *I. ricinus* assembly[23]). In summary, generating tick genome assemblies is challenging due to their unique biology.

108

109 **Survey of tick genome assemblies and quality assessment**

110 We conducted a survey for tick whole-genome assemblies (keyword Ixodida) in the
111 publicly available sequence databases: United States National Center for Biotechnology
112 Information (NCBI); European Nucleotide Archive (ENA); and China National Center for
113 Bioinformation (CNCB). A total of 70 tick genome assemblies: found on NCBI (65
114 assemblies) or ENA (one assembly), as of April 30, 2025; found on CNCB (one
115 assembly), as of February 28, 2025; or came from in-house (three assemblies), which
116 were not yet publicly available. We identify 16 assemblies (Table S1) on NCBI that have
117 been mistakenly classified as tick genomes; they are endosymbiont genomes[34,35] or
118 transcriptomes[36]. Therefore, in this study, we examine 54 genome assemblies,
119 representing 24 species of tick (Table S2). Species designations were obtained from
120 tick genome assembly publications, NCBI, and personal communications. We assessed
121 quality metrics for these tick genome assemblies using QUAST-LG v.5.2.0[37] and
122 BUSCO (benchmarking universal single-copy orthologs) lineage arthropoda_odb10,
123 using compleasm v.0.2.5[38]. Here, we evaluated how biological source material,
124 sequencing platforms, bioinformatic assembly methods, and gene and repetitive
125 element annotation impact the quality of these tick genome assemblies.

126

127 **BIOLOGICAL SOURCE MATERIAL FOR GENOME SEQUENCING OF TICKS**

128 We compiled the metadata of the biological material for the 54 tick genome assemblies.
129 For the zoogeographic origin of the tick material[39], the highest number of assemblies
130 (~35%) are generated from ticks collected in the Nearctic zoogeographic region (North
131 America), followed by ~31% Palearctic (Europe and China) and ~26% Afrotropical
132 (Africa) (Fig. 2A, Table S3). We also note that some tick species are invasive and have
133 been collected outside their native zoogeographic region and a genome assembly
134 generated from this material, such as *Haemaphysalis longicornis* (see [40,41]).
135 Currently, there are no genome assemblies from ticks collected in the Oriental region
136 (South Asia and Southeast Asia) (Fig. 2A, Table S3), which has the second highest
137 number of hard tick species, nor the Neotropical region (Central America and South
138 America) (Fig. 2A, Table S3) which has the fourth highest number of hard tick

species[39]. We recommend the Neotropical and Oriental zoogeographic regions be prioritized in future tick genome projects.

For the collection source of the tick material, the highest number of assemblies are generated from tick colonies (Fig. 2B, Table S3). These ticks from colonies were sourced ~24% from the Deutch colonies[42] followed by ~14% Oklahoma State University colonies[29,43,44] (Fig. 2B, Table S3). The second highest number of assemblies are from ticks caught in the field (Fig. 2B, Table S3). The third highest number of tick assemblies are from laboratory reared ticks that were caught in the field and maintained in the laboratory until a desired life stage had been obtained, and the smallest number of assemblies are from tick cell lines (Fig. 2B, Table S3). These cell lines are typically created from heterogenous embryonic cells and maintained indefinitely in a laboratory, which leads to large-scale genome duplications, rearrangements, and deletions[45]. We recommend the biological source material for assemblies be obtained from laboratory colonies, laboratory reared, or field caught.

Tick material has varied in sample size for assemblies, with ~56% individual ticks and ~44% of pooled ticks (Fig 2C, Table S3). For assemblies generated from individuals, all are adult ticks, apart from three nymphs (Fig. 2D, Table S3). For assemblies generated from pooled tick material, the highest number of assemblies came from eggs (~46%) (Fig. 2D, Table S3). For the assemblies generated from individuals, ~78% were from females (Fig. 2E, Table S3), most likely because they generally contain larger amounts of DNA than males[46]. Until 2020, pooling DNA from multiple individuals (Table S2-S3) was needed exclusively to achieve the minimum amount of DNA required for sequencing. Many of the most recent assemblies use DNA from a single individual to generate the data used for primary (contig) assembly, then use pooled DNA to generate sequence data for high-throughput chromosome conformation capture (Hi-C) assembly scaffolding[29]. Pooling individual ticks or eggs provides sufficient DNA yield for a variety of sample preparation and sequencing approaches, however, the high amount of heterozygosity among individuals can lead to significant difficulty in producing highly contiguous assemblies[21]. We recommend prioritizing tick assemblies generated from single, individual ticks, which can achieve a high-quality haploid assembly.

170

171 **TICK WGS AND ASSEMBLY DEVELOPMENTS FROM 2011 – 2025**

172 We traced the history of the 54 tick genome assemblies. The first tick assembly,
173 JCVI_ISG_i3_1.0 for *I. scapularis*, was released in 2011 (Fig. 3, Table S2). This
174 assembly project[33] resulted from a 2005 community white paper, with this species
175 being prioritized due to its medical importance in the United States as the key species
176 associated with Lyme disease[47]. The JCVI_ISG_i3_1.0 assembly is the only tick
177 assembly to use Sanger dideoxy chain termination on capillary instruments (first-
178 generation sequencing)[48,49]. This method costs millions of dollars in reagents per tick
179 genome, due to the large genome size[48]. Use of pooled individuals from an inbred
180 colony, combined with modest sequence read lengths[50] was unable to produce a
181 highly contiguous assembly due to the highly repetitive and heterozygous tick genomes.

182 The next tick genome assemblies used 454-pyrosequencing (second generation
183 sequencing), CCG_Rmi1.0[50] and Rmi2.0[51] *Rhipicephalus microplus*. This
184 sequencing platform obtains millions of ~500 nt reads at a time, thus reducing costs, but
185 has similar limitations in handling repetitive and heterozygous assemblies as Sanger
186 sequencing[52]. By 2016, published tick assemblies were using the Illumina (second-
187 generation) short-read sequencing platform (Fig. 3, Table S2). Illumina reads are 50-
188 600 nt in total, most commonly paired-end 100 or 150 nt reads (i.e., 200 or 300 nt total)
189 and sequencing is massively paralleled for high throughput, providing higher genome
190 coverage[52]. As Illumina uses short reads, it cannot be used to reliably map or
191 assemble long repetitive regions present in tick genomes.

192 After 2017, tick genomes were assembled using the data from the Pacific
193 Biosciences (PacBio) (third-generation sequencing) sequencing platform (Fig. 3, Table
194 S2). PacBio uses specialized library preparation (double-stranded DNA is circularized
195 via SMRTbell adapters) and single-molecule sequencing to create long-read reads (up
196 to ~100 kb for single-pass CLR reads or modal length of ~20 kb when sequenced
197 repeatedly for HiFi reads) to handle the high proportion of repetitive elements in ticks.
198 Initially, PacBio CLR reads were used to create tick assemblies with much better
199 contiguity[53] and Illumina reads were often used to correct uncalled or mis-called
200 bases (Fig. 3, Table S2). Now the standard output for PacBio is HiFi reads, which have

201 greatly improved base-calling accuracy (trading-off reads >25kb) for tick
202 assemblies[29,54].
203 From 2020 to 2023, 22 tick assemblies were released and more than half used
204 the PacBio sequencing platform (Fig. 3, Table S2). In 2023, the Oxford Nanopore
205 (another third-generation sequencing) sequencing platform was first used for tick
206 genome sequencing (Fig. 3, Table S2). This platform uses long-reads (reads 50 bp to >
207 4 Mbp) generated in real time via flow cells with nanopores[55]. While Oxford Nanopore
208 base-calling accuracy has greatly improved, ~33% of Nanopore tick assemblies have
209 been generated in combination with Illumina short-reads to correct base-calling errors
210 (Fig. 3, Table S2). In 2024 the most genome assemblies were released (20 assemblies)
211 with ~40% using Nanopore and the remaining using combinations of Nanopore or
212 PacBio, and Illumina (Fig. 3, Table S2). By April 2025, six assemblies were released,
213 with four using Nanopore (Fig. 3). The accuracy of Oxford Nanopore reads has now
214 improved sufficiently that Illumina reads are often no longer necessary for error
215 correction[56] or are only needed in scenarios where near perfect sequence accuracy is
216 required[57].
217 The development of new generations of sequencing platforms has been critical
218 for overcoming the limitations of tick biology (large genome size and high proportion of
219 repeats) leading to an increasing number of tick genome assemblies. Standalone first-
220 generation sequencing generated ~4% of tick assemblies, second-generation
221 sequencing ~20%, and third-generation sequencing ~39% (Fig. 3, Table S2). By 2018,
222 first generation sequencing was no longer being used to generate tick assemblies, while
223 third-generation sequencing was first used for ticks in 2017 (Fig. 3, Table S2). In
224 addition, ~37% of tick genome assemblies have been generated using combinations of
225 different generation sequencing (Fig. 3, Table S2).
226 For the 54 tick assemblies, ~37% used one or more integrations (Hi-C or RNA-
227 sequencing) to the sequencing platform. Hi-C is a molecular approach[58,59], which
228 improves contiguity of an assembly and is useful for ticks with their large, repetitive,
229 heterozygous genomes. Alternatively, RNA-seq provides direct evidence of
230 transcription, as well as the sequence of genes and isoforms[60]. Of the 20 assemblies
231 using integrations, 55% used a combination of Hi-C and RNA-sequencing, 30% RNA-

232 sequencing alone, and 15% Hi-C alone (Fig. 3, Table S2). Hi-C was first used to
233 generate a tick assembly in mid-2020, since then, ~26% of assemblies have integrated
234 Hi-C (Fig. 3, Table S2).
235

236 **MAJOR CHALLENGES IN STANDARDIZATION OF ANNOTATION FOR TICK**
237 **GENOME ASSEMBLIES**

238 Annotation of genes and repetitive elements (in particular, the transposable elements) in
239 tick genome assemblies has not been standardized. These non-standardized
240 annotations currently limit our ability to compare the genes and repetitive elements
241 between tick assemblies, even for the same species.

242 While genome assemblies commonly include gene annotation, tick genome
243 assemblies have not had a standardized approach. In particular, some assemblies
244 report the number of protein-coding genes and others the total number of genes (Table
245 S2). Gene annotation for tick genome assemblies has four common approaches. First,
246 custom gene annotation pipelines individually created for each assembly. Second,
247 taking a custom pipeline annotation and aligning RNA-seq data (either pre-existing or
248 generated alongside the genome assembly) using gene modeling programs (such as
249 AUGUSTUS[61] and BRAKER[61,62]). Third, manual pipeline including aligning RNA-
250 seq and proteome data. Fourth, requesting NCBI to perform a RefSeq annotation,
251 running the NCBI Eukaryotic Genome Annotation Pipeline[63], which incorporates
252 multiple data types (including RNA-seq and proteome) from within NCBI.

253 Gene annotation has been performed for 30 tick assemblies, representing 18
254 species (Figure S1, Table S2). A custom gene annotation pipeline was used for seven
255 tick assemblies, with a median of ~27,100 total genes and ~26,400 protein-coding
256 genes (Figure S1, Table S2). Also, a custom gene annotation pipeline with RNAseq
257 data has been used for 17 tick assemblies with a median of ~28,500 total genes and
258 ~27,000 protein-coding genes (Figure S1, Table S2). A manual pipeline was used for
259 four tick assemblies, with a median of ~23,700 protein-coding genes (Figure S1, Table
260 S2). The NCBI Pipeline has been used for seven tick assemblies with a median of
261 ~27,000 total genes and ~26,900 protein-coding genes (Figure S1, Table S2). While
262 four alternative approaches have been used to annotate the genes in tick assemblies

263 and resulted in variable gene number estimates, the median number for total genes (15
264 assemblies) is between ~27,000 to ~28,500 and protein-coding genes (24 assemblies)
265 is between ~23,700 to ~27,000. We recommend future studies that generate tick
266 genome assemblies achieve high quality gene annotation by including a transcriptome,
267 ideally via PacBio Iso-Seq, and submitting their assembly to NCBI for annotation.
268 The highly repetitive genomes of ticks mean that these repetitive elements
269 should also be annotated. The common approach to annotating repetitive elements is
270 using an automated pipeline, such as one or a combination of RepeatModeler[64] and
271 RepeatMasker[65]. Repetitive elements have been annotated in 25 tick genome
272 assemblies, representing 15 species (Figure S2, Table S2). An automated pipeline for
273 repetitive elements has been performed on 23 tick assemblies with a median of ~62%
274 repetitive DNA, composed of ~1.5% tandem repeats and ~59% transposable elements
275 (Figure S2A, Table S2). When these transposable elements have been further
276 classified, the median is ~24% class I, ~6% class II, and ~2.4% unclassified (Figure
277 S2B, Table S2). A manual pipeline for repetitive elements has been performed on three
278 *Ixodes* assemblies from one study[23] with a median of 73% repetitive DNA, composed
279 of ~4.3% tandem repeats and ~68% transposable elements (Figure S2A, Table S2).
280 Further classification gave a median ~32% class I and ~36% class II (Figure S2B, Table
281 S2). We recommend future tick genome assemblies perform manual curation of the
282 transposable elements until a standardized, automated pipeline is established.
283
284 **QUALITY ASSESSMENT OF TICK GENOME ASSEMBLIES**
285 We comprehensively assessed the impact of sequencing platform and Hi-C scaffolding
286 on tick assembly quality, using metrics for contiguity (auN, N50, L50, and L90) and
287 BUSCO complete scores (a higher BUSCO score indicates higher assembly
288 completeness), which are common assessments of genome assembly quality[66]. The
289 complete BUSCO score is the sum of the single copy and duplicated genes that can be
290 completely aligned to the assembly with one or more than one gene copy present,
291 respectively. Sanger and 454-FLX sequencers were used for the first two tick
292 assemblies (Table S2), generating highly fragmented assemblies (Fig. 4A-D, Table S4).

293 While the complete BUSCO score for the Sanger assembly was close to 90%, the 454-
294 FLX partial assembly was close to 0% (Fig. 4A-D, Table S5).
295 Illumina sequencing has been used for 12 tick assemblies (Table S2). Only one
296 Illumina assembly integrated Hi-C scaffolding, *I. ricinus* (IXRI_v2)[27], which generated
297 the highest quality Illumina tick assembly with a ~95% complete BUSCO score (Fig. 4A-
298 D, Table S4-S5). The other 11 Illumina assemblies are more fragmented and seven did
299 not produce sufficient reads to cover the genome, resulting in lower complete BUSCO
300 scores of 0 to ~65% (Fig. 4A-D, Table S4-S5). Therefore, standalone short read
301 platforms that are of insufficient depth and do not integrate Hi-C scaffolding are not
302 sufficient to handle the genome complexity of ticks.
303 PacBio sequencing has been used for eight tick assemblies, two of which
304 implemented Hi-C scaffolding (Table S2). Interestingly, one of the PacBio assemblies
305 with Hi-C had a similar quality to the Illumina assembly that integrated Hi-C (Fig. 4A-D,
306 Table S4-S5). The remaining six PacBio assemblies without Hi-C varied in quality (Fig.
307 4A-D, Table S4-S5).
308 A combination of PacBio sequencing error-corrected with Illumina short-reads
309 has been used for 13 tick assemblies (Table S2). Twelve of these assemblies have
310 complete BUSCO scores over 92%, indicating sufficient sequencing coverage of the
311 genome (Fig. 4A-D, Table S4-S5). Ten of these 13 assemblies implemented Hi-C
312 scaffolding and the three without Hi-C were lower in quality regarding contiguity and
313 coverage (Fig. 4A-D, Table S4-S5). Our findings indicate that Hi-C positively impacts
314 assembly contiguity and quality for PacBio tick assemblies with or without additional
315 Illumina reads.
316 Nanopore sequencing has been used for 12 tick assemblies (Table S2). These
317 assemblies overlap in quality with the Illumina, PacBio, and PacBio+Illumina tick
318 assemblies (Fig. 4A-D, Table S4-S5). None of the standalone Nanopore assemblies
319 implemented Hi-C (Fig. 4A-D). The quality and contiguity of Nanopore assemblies is
320 moderate when compared to other sequencing platforms.
321 A combination of Nanopore error-corrected with Illumina short-reads has been
322 used for six tick assemblies (Table S2). For all Nanopore+Illumina assemblies, the
323 complete BUSCO values were higher than 83%, which is unexpectedly lower than the

standalone Nanopore assemblies, which were ~94% (Fig. 4A-D, Table S4-S5). One of these six assemblies integrated Hi-C scaffolding, which had higher quality and contiguity compared to Nanopore+Illumina tick assemblies without Hi-C (Fig. 4A-D, Table S4-S5). The quality of the remaining five Nanopore+Illumina assemblies without Hi-C were similar in contiguity, but lower in quality in regard to BUSCO complete scores compared to the standalone Nanopore tick assemblies (Fig. 4A-D, Table S4-S5). Therefore, the highest quality Nanopore assemblies were generated without error-correction using Illumina reads, with exception of the Nanopore+Illumina assembly that integrated Hi-C.

Across all the sequencing platforms, on average, assemblies integrating Hi-C data improved quality metrics and complete BUSCO scores, resulting in assemblies with higher contiguity, quality, and completeness (Fig. 4A-D, Table S4-S5). Additionally, assemblies generated by a combination of long-read platforms (PacBio and Nanopore) then error-corrected with Illumina reads had, on average, better contiguity but lower genome completeness compared to those generated by long-read platforms without error-correction, suggesting a trade-off in contiguity versus errors (Fig. 4A-D, Table S4-S5). In summary, we recommend future tick genome assembly projects use long-read sequencing platforms with Hi-C integration to produce high-quality assemblies.

CHARACTERIZATION OF HIGH-QUALITY TICK ASSEMBLIES

We assessed the quality of the tick genome assemblies using number of contigs, contig N50, and BUSCO complete score to identify 34 high quality tick genome assemblies that are suitable for downstream analyses (Fig. 5). These high-quality assemblies have fewer than 100,000 contigs, a contig N50 greater than 45 Kbp, and a complete BUSCO score higher than 83% (Fig. 5, Table S4-S6). These 34 high-quality assemblies were all generated since 2020 (Fig. 3) and used ~15% Illumina, ~18% Nanopore, ~18% Nanopore+Illumina, ~21% PacBio, and ~29% PacBio+Illumina (Fig. 5, Table S4-S6). Approximately 41% of the high-quality assemblies integrated Hi-C (Fig. 5, Table S4-S6).

We also identify a cluster of 14 highest quality tick genome assemblies which have fewer than 7,500 contigs, a contig N50 greater than 84 Mbp, and a complete

355 BUSCO score greater than 93% (Fig. 5, Table S4-S6). These 14 highest-quality
356 assemblies were all generated since 2020 (Fig. 3) and ~7% used Illumina, ~7%
357 Nanopore+Illumina, ~21% PacBio, and ~64% PacBio+Illumina (Fig. 5, Table S4-S6). All
358 14 assemblies integrated Hi-C, except for *Rhipicephalus microplus* (USDA_Rmic)[15]
359 (Fig. 5). Our use of comprehensive metrics for the highest-quality tick assemblies
360 means future tick assembly projects now have metrics to benchmark against.
361

362 **TICK SPECIES WITH HIGH QUALITY GENOME ASSEMBLIES**

363 The 34 high-quality genome assemblies cover 21 species of tick (Fig. 6). As some
364 species have multiple high-quality assemblies, we identify the best quality genome
365 assembly for each of the 21 tick species (Fig. 6, Table S6).
366 For the 21 best quality assemblies, we find the hard ticks have an average
367 assembly length of ~2,390 Mbp and soft ticks ~1,176 Mbp (Fig. 6, Table S4). The hard
368 tick *Ixodes* spp. (Prostriata) have an average assembly length of ~2,292 Mbp and the
369 non-*Ixodes* spp. (Metastriata) ~2,425 Mbp (Fig. 6, Table S4). Across all tick species
370 *Rhipicephalus* spp. have the largest assembly length, with an average of ~2,689 Mbp,
371 whereas the two *Hyalomma* spp. have the smallest assembly lengths for the hard ticks
372 (average ~1,840 Mbp) (Fig. 6, Table S4). The two soft tick *Ornithodoros* spp. have the
373 smallest assembly lengths for all ticks with an average of ~1,176 Mbp (Fig. 6, Table S4).
374 Assembly size (length of gapped DNA sequence) is not equivalent to genome size. The
375 average genome size for soft ticks is 1.3 Gbp and for hard ticks is 2.7 Gbp using flow
376 cytometry[67]. The average percent difference of the estimated genome size to both the
377 NCBI assembly size and assembly total length is 95.5% (Table S7). We hypothesize
378 that hard ticks have larger genomes compared to soft ticks due to larger proportions of
379 repetitive elements.
380 The 21 species with high-quality assemblies are two soft ticks (Argasidae) and
381 19 hard ticks (Ixodidae), or five Ixodes and 14 non-Ixodes species (Table S7). This
382 makes up 0.9% (2/216) of the known soft tick species and 2.4% (19/786) of the known
383 hard tick species[12,13]. We recommend that generating genome assemblies for
384 additional soft ticks should be a priority. So far, tick species sequenced are those of
385 medical and veterinary importance[47]. Therefore, we also recommend future tick

genome projects prioritize capturing the diversity of tick species, particularly those of medical, veterinary, and phylogenetic importance[68,69].

CONCLUDING REMARKS

We have assessed 54 genome assemblies for ticks that have been generated, identifying 34 high-quality assemblies for 21 tick species. Our assessment of the tick genome assemblies has provided an opportunity to make recommendations for future tick genome assembly projects. We recommend future assemblies better represent the biodiversity of tick species, particularly the soft ticks and from unrepresented zoogeographic regions. In addition, to obtain the highest quality tick genome assemblies, future projects should prioritize obtaining individual, adult ticks for DNA extractions and for sequencing platforms use a combination of PacBio HiFi with Hi-C Illumina scaffolding (or Oxford Nanopore with Hi-C scaffolding) with transcriptomes. High-quality annotations should include transcriptomes from as many life-stages as possible, plus TE manual curation. With these recommendations, future tick genome assemblies will improve in quality and representation of tick species.

With the availability of high-quality tick genome assemblies, there are opportunities to move beyond sequencing a single individual per species to better understand tick biology. Genomic resources that are well established in other non-model arthropods but are needed in ticks include a chromosome-anchored reference genome using fluorescent *in situ* hybridization (FISH)[70] , sequencing multiple individuals per species to increase phylogeographic representation[71], and cross-species whole genome alignments[71]. New approaches to make even more advances include new sequencing technologies, surveys for transposase-accessible chromatin sequencing (ATAC-seq)[72] to characterize regulatory elements in the genome, and telomere-to-telomere (T2T) gapless sequencing[73] to understand tick chromosome structures. These additional approaches will greatly improve our understanding of tick genomes and tick biology.

Over the past decade, major developments in sequencing technologies have brought about a renaissance for tick genomics. These genome assemblies are critical for advancing tick research from vectorial capacity[74] to anti-tick vaccines[75]. The tick

genome assemblies have so far been referred to or used in more than 130 follow up studies (Table S8) since 2011. These follow up studies have four key themes: identifying gene targets for anti-tick vaccination studies, tick control, and acaracide resistance; annotation of repetitive elements; intra-species and inter-species comparative genomics studies; transcriptomics and proteomics studies. Therefore, the tick genome assemblies have enabled major advances in our understanding of tick biology.

We look forward to new tick genome assemblies providing more insights into the fascinating biology of these medically and veterinary important blood-feeding parasites.

Ethics approval

None.

Availability of data and materials

All supporting datasets for this study are available in the supplementary material.

Funding

KCD and TCG have been supported in part with federal funds from the Centers of Disease Control and Prevention Pathogen Genomics Center of Excellence (Contract No. NU50CK000626). JCF was partially supported by Grant No. DGE-1545433 from the National Science Foundation and the IDEAS programme at the University of Georgia. HS was financially supported by the Dutch Ministry of Health, Welfare and Sport (VWS). IR is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation.

Acknowledgements

We thank all the researchers who have pushed the boundaries of tick genomics and shared with us additional information. In particular, we thank the following researchers for their willingness and transparency to provide information on unpublished tick genome assemblies that they have released on genome databases, in alphabetical order: Joshua Benoit, Seemay Chou, Christopher Faulk, Jan Forth, Melissa Klein,

Michel Labuschagne, Susan Noh, Pia Olafson, Perot Saelao, Ryan O. M. Rego, Claude Rispe, Hailey Schau, and Pavlina Vechtova. We also acknowledge Roslyn Hickson for facilitating the resources that made the *O. gurneyi* genome assembly possible. The use of the unpublished genome assemblies in this paper does not constitute their announcement, as details related to preparation, assembly and gene annotation processes have yet to be described.

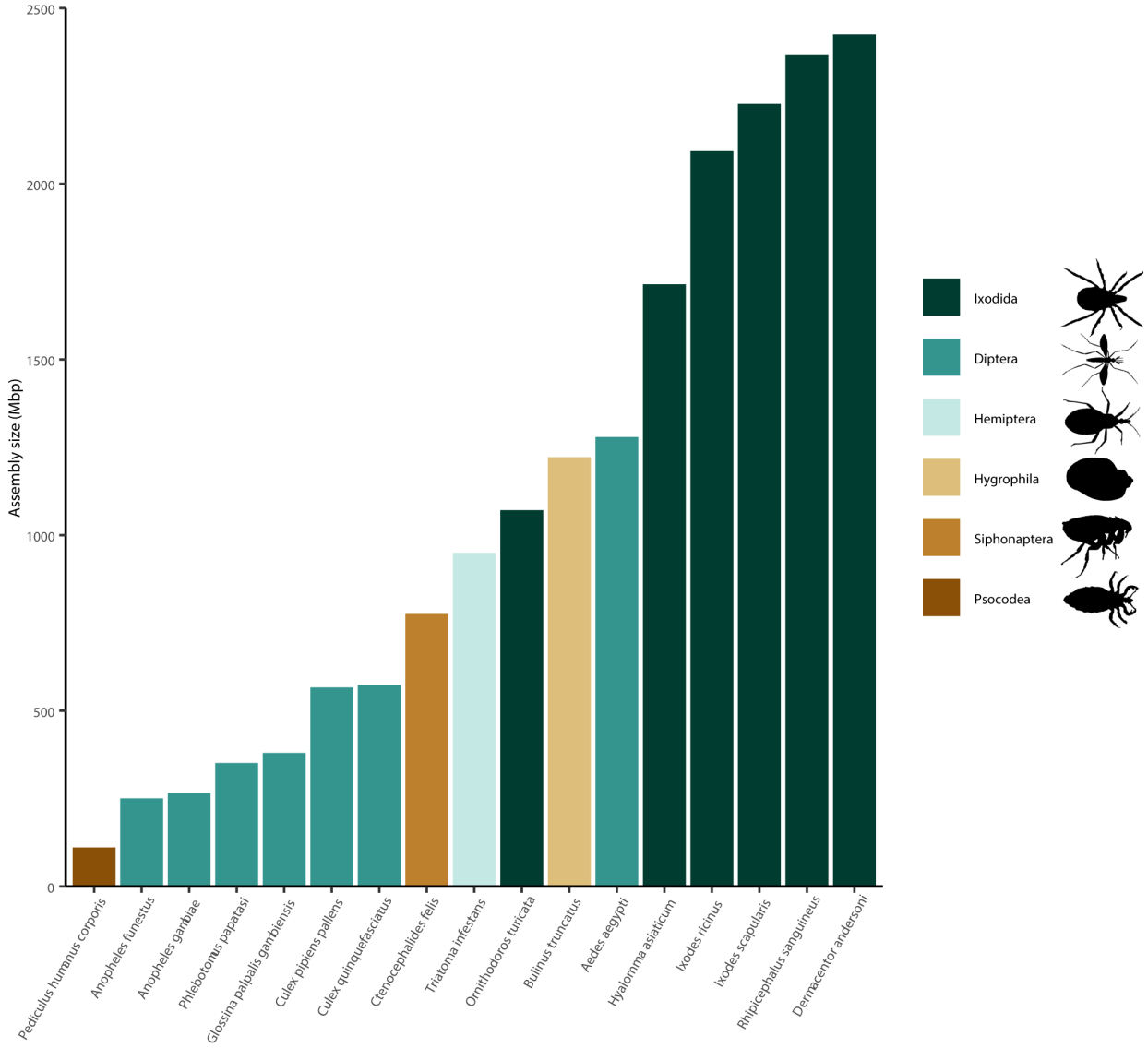
Author contributions

Conceptualization: KCD, IR, TCG, JF
Validation: KCD
Formal analysis: KCD
Investigation: JF, KCD
Resources: HS
Data Curation: KCD
Writing - Original Draft: KCD, IR
Writing - Review & Editing: KCD, IR, TCG, HS, JF
Visualization: KCD, IR
Supervision: IR, TCG
Project administration: IR, TCG
Funding acquisition: HS, IR, TCG

Declaration of interests

None declared.

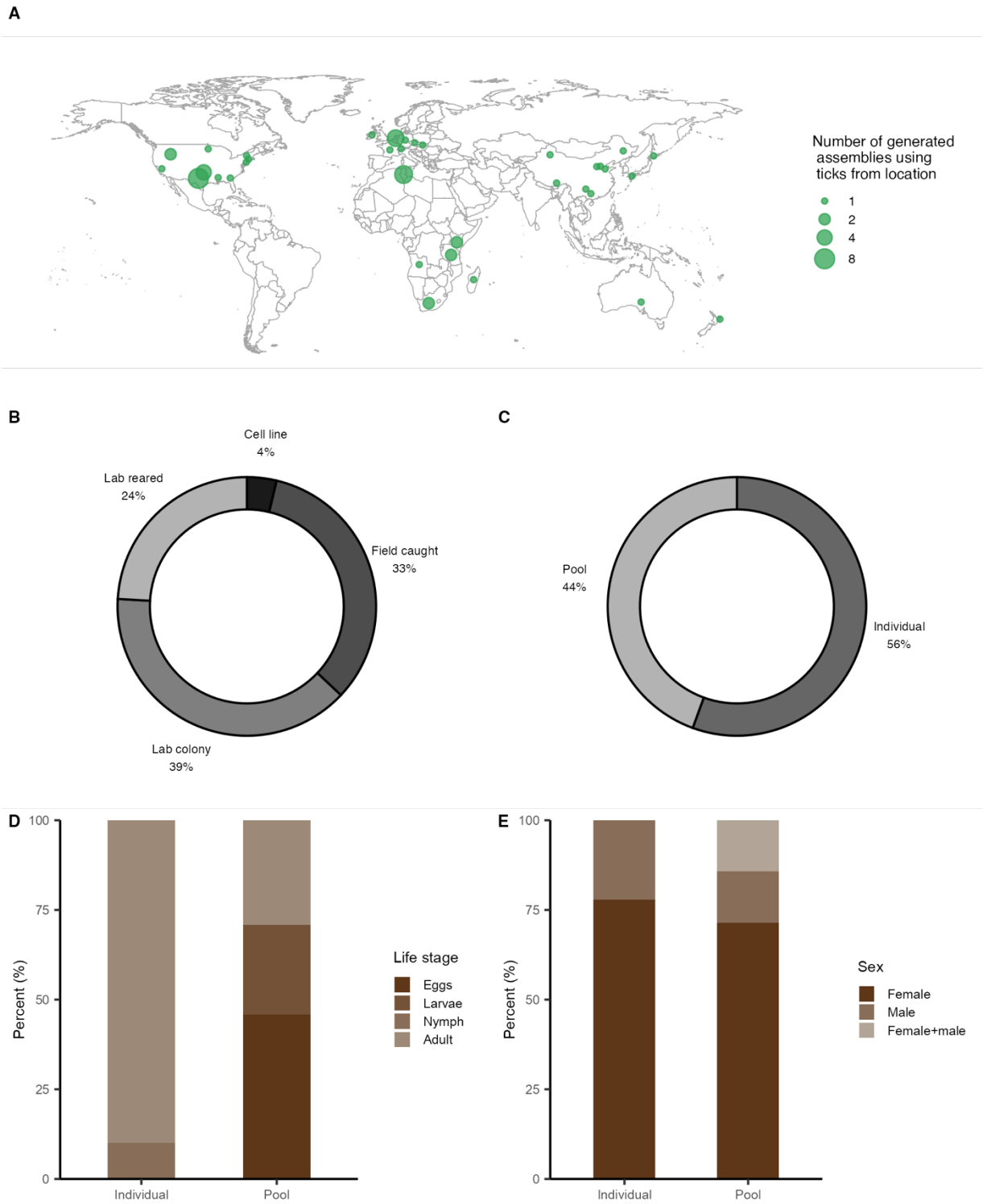
472 **FIGURES**



473

474 **Figure 1. Ticks (Ixodida) have the largest genome assembly sizes when compared**
475 **with other major human disease vectors**. Genome assembly size for 17 disease
476 vectors, spanning six taxonomic orders, from the major human vector-borne diseases
477 global list (https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases).
478 For human vector-borne diseases associated with multiple vector species, the species
479 associated with the highest number of disease cases and has an available assembly is
480 shown. When a vector species has multiple assemblies, the current reference assembly
481 from NCBI is used. For each assembly, the NCBI assembly size is reported.

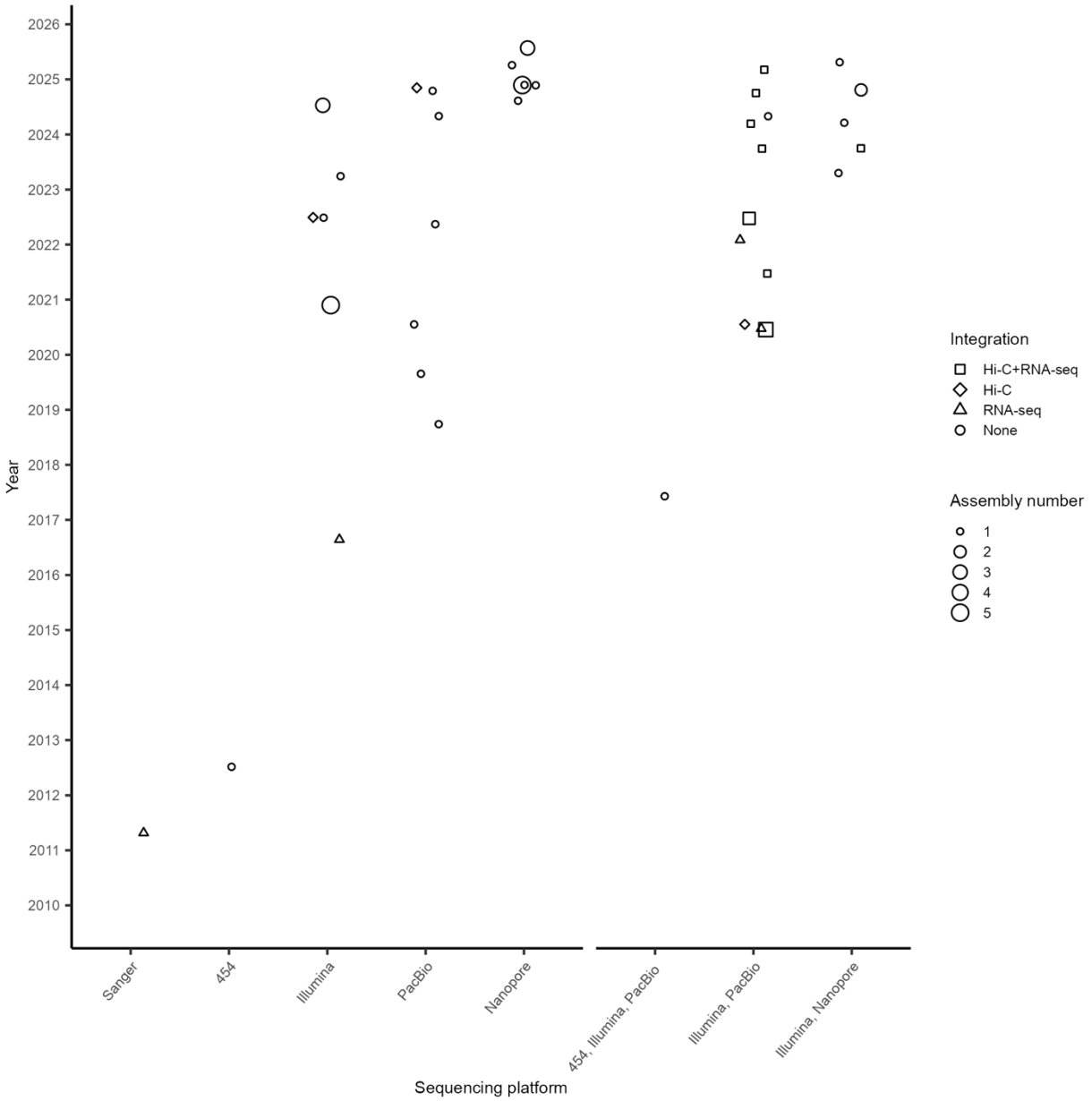
482



483

484 **Figure 2. Biological source material for 54 tick genome assemblies. A)** Geographic
485 origin of the tick material. Two assemblies were generated using ticks from three
486 locations. **B)** Biological sources of the tick material. A laboratory colony is defined as

487 ticks being maintained across generations in a lab environment. We define laboratory
488 reared ticks as those collected from the field and maintained in the laboratory until a
489 desired life stage has been obtained, not across generations. Field caught ticks are
490 collected from either the environment or hosts. Tick cell lines are preserved cells,
491 typically from embryos. **C)** Sample size of the tick material, either is an individual
492 specimen or pooled specimens. **D)** Life stage of the tick in individual versus pooled
493 material. **E)** The sex of the adult ticks in individual versus pooled material. See Table S3
494 for further details.
495



496

497 **Figure 3. Timeline of assembly release for 54 tick genome assemblies, first**
498 **release was in 2011.** Each point is the release date on a genome database of a
499 genome project; a project has included up to five tick assemblies. Sequencing platforms
500 are in chronological order, with first generation (Sanger or 454-FLX), second generation
501 (Illumina) and third generation (PacBio or Nanopore). Followed by combinations of
502 sequencing platforms. Integrations to the sequencing platforms were none, RNA-seq
503 alone, Hi-C alone or Hi-C with RNA-seq. See Table S2 for further details.

504

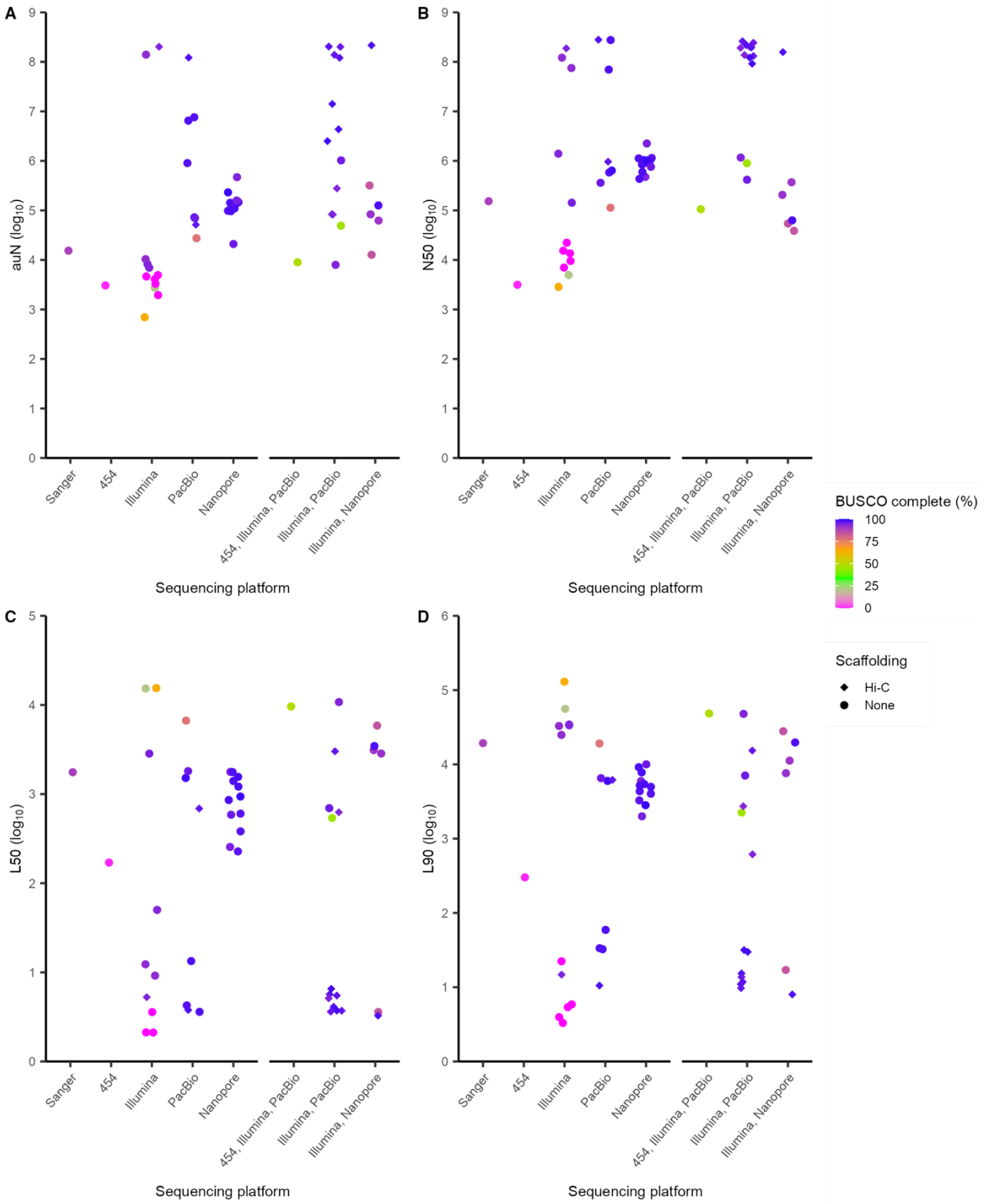
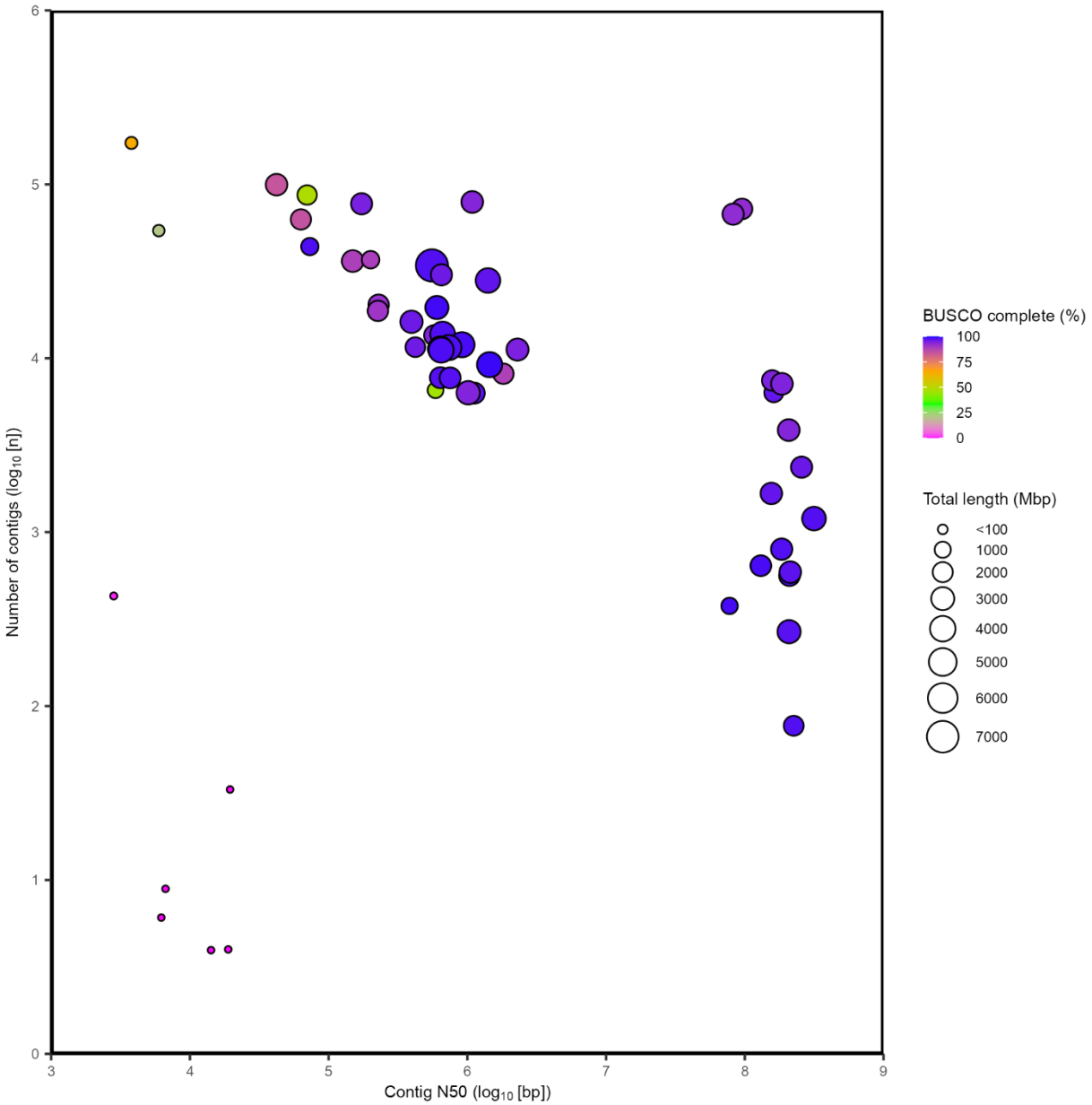


Figure 4. Quality metrics for 54 tick genome assemblies across sequencing platforms. A) auN (comprehensive metric for contiguity), as auN increases so does assembly quality. **B)** Contig N50 (the length of the shortest contig that must be included

509 to cover half an assembly), as N50 increases, fewer longer contigs are used to cover
510 half of the assembly, leading to a higher quality assembly. **C)** L50 (minimal number of
511 contigs that cover half of the assembly), as L50 decreases assembly quality can
512 increase, however, the BUSCO score needs to be included in the assessment. **D)** L90
513 (minimal number of contigs that cover 90% of the assembly), as L90 decreases
514 assembly quality can increase, however the BUSCO score needs to be included in the
515 assessment. Each point is a genome assembly. Sequencing platforms are ordered by
516 first generation (Sanger and 454-FLX), second generation (Illumina) and third
517 generation (PacBio and Nanopore). Followed by sequencing platform combinations.
518 Scaffolding for the sequencing platforms were none or Hi-C. See Table S2, S4, and S5
519 for further details.
520



521

522 **Figure 5. Quality assessment of 54 tick genome assemblies.** Identified 34 high-
523 quality assemblies (lower contig number, higher N50, and higher BUSCO score),
524 including 21 highest quality assemblies. Each dot is a genome assembly. Contig N50 is
525 the length of the shortest contig that must be included to cover half an assembly, a
526 higher N50 means less small contigs which is a more complete, less fragmented
527 assembly. For the number of contigs, fewer contigs typically result in a more contiguous
528 assembly as there is less fragmentation. However, few contigs associated with a low
529 N50 and small assembly length can lead to poor assembly quality. As not all assemblies

530 were obtained from NCBI, we used a metric that could be calculated for all 54
531 assemblies: the total length of the assembly (generated by QUAST) which includes
532 uncalled bases. However, QUAST total length more closely represents the assembly
533 size, not genome size. BUSCO complete score, a higher BUSCO score indicates higher
534 assembly completeness and a higher quality assembly. Estimated haploid assembly
535 size from NCBI. See Table S4-S6 for further details.
536

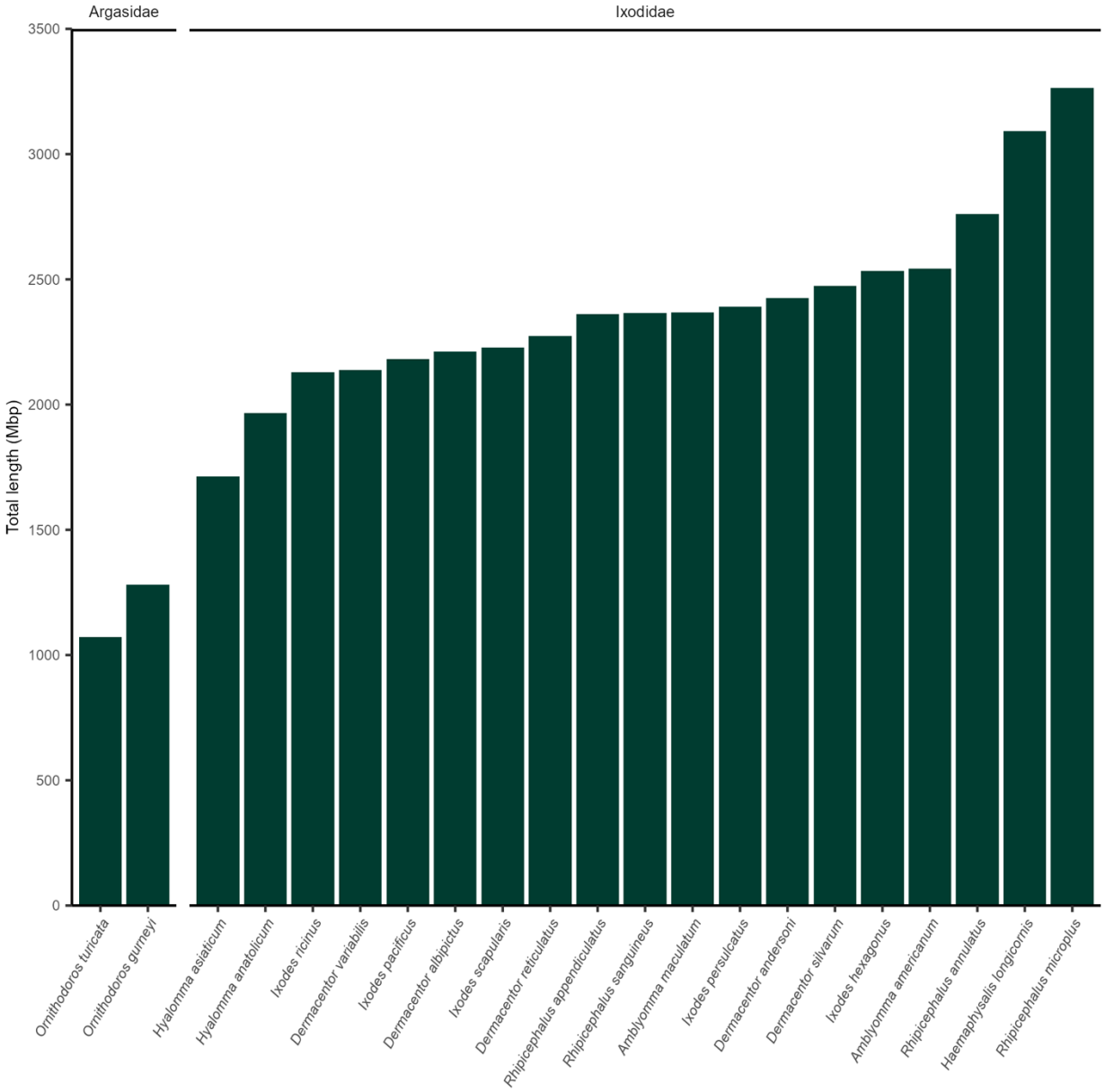


Figure 6. Assembly total length of the best quality genome assembly for 21 tick species. The best assembly for each tick species was identified from the high and highest quality assemblies. Assembly length (Mbp) of the best quality assembly for each species. Tick species are ordered by assembly total length, from smallest to largest. See Table S6 for further details.

References Cited

1. Madison-Antenucci, S. *et al.* (2020) Emerging Tick-Borne Diseases. *Clinical Microbiology Reviews* 33, 10.1128/cmr.00083–00018. https://doi.org/10.1128/cmr.00083-18
2. Parola, P. *et al.* (2013) Update on tick-borne rickettsioses around the world: a geographic approach. *Clinical microbiology reviews* 26, 657–702. https://doi.org/10.1128/cmr.00032-13
3. Kim, E.-H. and Park, S.-J. (2023) Emerging tick-borne *Dabie bandavirus*: virology, epidemiology, and prevention. *Microorganisms* 11, 2309. https://doi.org/10.3390/microorganisms11092309
4. Bishop, R. *et al.* (2004) *Theileria*: intracellular protozoan parasites of wild and domestic ruminants transmitted by ixodid ticks. *Parasitology* 129, S271–S283. https://doi.org/10.1017/S0031182003004748
5. Ajileye, O.D. *et al.* (2025) A review of filarial nematodes parasitizing tick vectors: unraveling global patterns in species diversity, host associations, and interactions with tick-borne pathogens. *Parasites & Vectors* 18, 50. https://doi.org/10.1186/s13071-025-06690-6
6. Kilpatrick, A.M. and Randolph, S.E. (2012) Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *The Lancet* 380, 1946–1955. https://doi.org/10.1016/S0140-6736(12)61151-9
7. Saker, L. *et al.* (2004) Globalization and infectious diseases, a review of the linkages. *UNDP/World Bank/WHO Special Programme on Tropical Diseases Research, Geneva.* https://researchonline.lshtm.ac.uk/id/eprint/15642
8. Burtis, J.C. *et al.* (2021) Susceptibility of *Ixodes scapularis* (Acari: Ixodidae) to Permethrin Under a Long-Term 4-Poster Deer Treatment Area on Shelter Island, NY. *Journal of Medical Entomology* 58, 1966–1969. https://doi.org/10.1093/jme/tjab054
9. Vudriko, P. *et al.* (2016) Emergence of multi-acaricide resistant *Rhipicephalus* ticks and its implication on chemical tick control in Uganda. *Parasites & Vectors* 9, 4. https://doi.org/10.1186/s13071-015-1278-3
10. Ziapour, S. *et al.* (2016) Susceptibility status of field populations of *Rhipicephalus bursa* (Acari: Ixodidae) to pyrethroid insecticides. *Tropical Biomedicine* 33, 446–461
11. Lempereur, L. *et al.* (2025) Research-led innovation for integrated tick management and acaricide resistance mitigation in livestock. *Trends in Parasitology* 0. https://doi.org/10.1016/j.pt.2025.09.002
12. Mans, B.J. *et al.* (2019) Argasid and ixodid systematics: implications for soft tick evolution and systematics, with a new argasid species list. *Ticks and Tick-borne Diseases* 10, 219–240. https://doi.org/10.1016/j.ttbdis.2018.09.010
13. Robbins, R.G. *et al.* (2025) Type specimens of the world's hard tick species (Acari: Ixodida: Ixodidae): collection data and depositories for all valid names and the current status of invalid names. *Zootaxa* 5606, 1–287. https://doi.org/10.11646/zootaxa.5606.1.1

14. Wysoki, M. and Bolland, H.R. (1978) Spermatogenesis, chromosomes and sex determination of four *Rhipicephalus* species (Acari: Ixodidae) from East Africa. *Genetica* 48, 233–238. https://doi.org/10.1007/BF00155576
15. Tidwell, J.P. *et al.* (2024) Identifying the sex chromosome and sex determination genes in the cattle tick, *Rhipicephalus (Boophilus) microplus*. *G3 Genes|Genomes|Genetics* 14. https://doi.org/10.1093/g3journal/jkae234
16. Gunn, S.J. and Hilburn, L.R. (1990) Cytosystematics of Five North American *Dermacentor* (Acari: Ixodidae) Species. *Journal of Medical Entomology* 27, 620–627. https://doi.org/10.1093/jmedent/27.4.620
17. Yilmazer, N. (2018) The genome size of some species of *Hyalomma* ticks from Turkish Thrace. *Systematic and Applied Acarology* 23, 2122–2128. https://doi.org/10.11158/saa.23.11.5
18. Oliver, J.H. (1977) Cytogenetics of Mites and Ticks. *Annual Review of Entomology* 22, 407–429. https://doi.org/10.1146/annurev.en.22.010177.002203
19. Kahn, J. (1964) Cytotaxonomy of ticks. *Journal of Cell Science* 3, 123–137. https://doi.org/10.1242/jcs.s3-105.69.123
20. Gunn, S.J. and Hilburn, L.R. (1989) Differential Staining of Tick Chromosomes: Techniques for C-Banding and Silver-Staining and Karyology of *Rhipicephalus sanguineus* (Latreille). *The Journal of Parasitology* 75, 239–245. https://doi.org/10.2307/3282772
21. Petersen, M. *et al.* (2019) Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Ecology and Evolution* 19, 11. https://doi.org/10.1186/s12862-018-1324-9
22. Wu, C. and Lu, J. (2019) Diversification of transposable elements in arthropods and its impact on genome evolution. *Genes* 10, 338. https://doi.org/10.3390/genes10050338
23. Ronai, I. *et al.* (2024) The repetitive genome of the *Ixodes ricinus* tick reveals transposable elements have driven genome evolution in ticks. *bioRxiv*, 2024.2003.2013.584159. https://doi.org/10.1101/2024.03.13.584159
24. Osmanski, A.B. *et al.* (2023) Insights into mammalian TE diversity through the curation of 248 genome assemblies. *Science* 380, eabn1430. https://doi.org/10.1126/science.abn1430
25. Shao, C. *et al.* (2023) The enormous repetitive Antarctic krill genome reveals environmental adaptations and population insights. *Cell* 186, 1279–1294. e1219. https://doi.org/10.1016/j.cell.2023.02.005
26. Keirans, J.E. and Durden, L.A. (2005) Tick systematics and identification. *Tick-borne Diseases of Humans*, 123–140. https://doi.org/10.1128/9781555816490.ch7
27. Cerqueira de Araujo, A. *et al.* (2025) Genome sequences of four *Ixodes* species expands understanding of tick evolution. *BMC biology* 23, 17. https://doi.org/10.1186/s12915-025-02121-1
28. Ribeiro, J.M. *et al.* (2023) A draft of the genome of the Gulf Coast tick, *Amblyomma maculatum*. *Ticks and tick-borne diseases* 14, 102090. https://doi.org/10.1016/j.ttbdis.2022.102090

632 29. De, S. *et al.* (2023) A high-quality *Ixodes scapularis* genome advances tick
633 science. *Nature Genetics* 55, 301–311. https://doi.org/10.1038/s41588-022-
634 01275-w
635 30. Díaz-Sánchez, S. *et al.* (2019) Evolutionary insights into the tick hologenome.
636 *Trends in Parasitology* 35, 725–737. https://doi.org/10.1016/j.pt.2019.06.014
637 31. Anderson, J.F. and Magnarelli, L.A. (2008) Biology of ticks. *Infectious disease*
638 *clinics of North America* 22, 195–215. https://doi.org/10.1016/j.idc.2007.12.006
639 32. Kolo, A.O. and Raghavan, R. (2023) Impact of endosymbionts on tick physiology
640 and fitness. *Parasitology* 150, 859–865.
641 https://doi.org/10.1017/S0031182023000793
642 33. Gulia-Nuss, M. *et al.* (2016) Genomic insights into the *Ixodes scapularis* tick
643 vector of Lyme disease. *Nature Communications* 7.
644 https://doi.org/10.1038/ncomms10507
645 34. Buysse, M. *et al.* (2021) A dual endosymbiosis supports nutritional adaptation to
646 hematophagy in the invasive tick Hyalomma marginatum. *elife* 10, e72747.
647 https://doi.org/10.7554/eLife.72747
648 35. Floriano, A.M. *et al.* (2025) Comparative genomics of *Rickettsiella* bacteria reveal
649 variable metabolic pathways potentially involved in symbiotic interactions with
650 arthropods. *Peer Community Journal* 5. https://doi.org/10.24072/pcjournal.633
651 36. Forth, J.H. *et al.* (2020) Identification of African swine fever virus-like elements in
652 the soft tick genome provides insights into the virus' evolution. *BMC biology* 18,
653 136. https://doi.org/10.1186/s12915-020-00865-6
654 37. Mikheenko, A. *et al.* (2018) Versatile genome assembly evaluation with QUAST-
655 LG. *Bioinformatics* 34, i142–i150. https://doi.org/10.1093/bioinformatics/bty266
656 38. Huang, N. and Li, H. (2023) compleasm: a faster and more accurate
657 reimplementation of BUSCO. *Bioinformatics* 39, btad595.
658 https://doi.org/10.1093/bioinformatics/btad595
659 39. Guglielmone, A.A. *et al.* (2023) Geographic distribution of the hard ticks (Acari:
660 Ixodida: Ixodidae) of the world by countries and territories. *Zootaxa* 5251, 1–274.
661 https://doi.org/10.11646/zootaxa.5251.1.1
662 40. Guerrero, F.D. *et al.* (2019) The Pacific Biosciences de novo assembled genome
663 dataset from a parthenogenetic New Zealand wild population of the longhorned
664 tick, *Haemaphysalis longicornis* Neumann, 1901. *Data in Brief* 27, 104602.
665 https://doi.org/10.1016/j.dib.2019.104602
666 41. Moustafa, M.A.M. *et al.* (2025) Genome of the invasive North American
667 *Haemaphysalis longicornis* tick as a template for bovine anti-tick vaccine
668 discovery. *BMC Genomics* 26, 307. https://doi.org/10.1186/s12864-025-11477-1
669 42. Davey, R. *et al.* (1980) Ovipositional biology of the cattle tick, *Boophilus*
670 *annulatus* (Acari: Ixodidae), in the laboratory. *Journal of Medical Entomology* 17,
671 287–289. https://doi.org/10.1093/jmedent/17.3.287
672 43. Nuss, A.B. *et al.* (2023) The highly improved genome of *Ixodes scapularis* with X
673 and Y pseudochromosomes. *Life science alliance* 6.
674 https://doi.org/10.26508/lsa.202302109
675 44. Chou, S. *et al.* (2023) De novo assembly of a long-read *Amblyomma*
676 *americanum* tick genome. *Arcadia Science*. https://dx.doi.org/10.57844/arcadia-
677 9b6j-q683

678 45. Kotsarenko, K. *et al.* (2020) Karyotype changes in long-term cultured tick cell
679 lines. *Scientific reports* 10, 13443. https://doi.org/10.1038/s41598-020-70330-5
680 46. Ammazzalorso, A.D. *et al.* (2015) To beat or not to beat a tick: comparison of
681 DNA extraction methods for ticks (*Ixodes scapularis*). *PeerJ* 3, e1147.
682 https://doi.org/10.7717/peerj.1147
683 47. Hill, C.A. and Wikel, S.K. (2005) The *Ixodes scapularis* Genome Project: an
684 opportunity for advancing tick research. *Trends in parasitology* 21, 151–153.
685 https://doi.org/10.1016/j.pt.2005.02.004
686 48. Kircher, M. and Kelso, J. (2010) High-throughput DNA sequencing–concepts and
687 limitations. *Bioessays* 32, 524–536. https://doi.org/10.1002/bies.200900181
688 49. Sanger, F. *et al.* (1977) DNA sequencing with chain-terminating inhibitors.
689 *Proceedings of the national academy of sciences* 74, 5463–5467.
690 https://doi.org/10.1073/pnas.74.12.5463
691 50. Guerrero, F.D. *et al.* (2010) Reassociation kinetics-based approach for partial
692 genome sequencing of the cattle tick, *Rhipicephalus (Boophilus) microplus*. *BMC*
693 *genomics* 11, 374. https://doi.org/10.1186/1471-2164-11-374
694 51. Barrero, R.A. *et al.* (2017) Gene-enriched draft genome of the cattle tick
695 *Rhipicephalus microplus*: assembly by the hybrid Pacific Biosciences/Illumina
696 approach enabled analysis of the highly repetitive genome. *International journal*
697 *for parasitology* 47, 569–583. https://doi.org/10.1016/j.ijpara.2017.03.007
698 52. Satam, H. *et al.* (2023) Next-generation sequencing technology: current trends
699 and advancements. *Biology* 12, 997. https://doi.org/10.3390/biology12070997
700 53. Miller, J.R. *et al.* (2018) A draft genome sequence for the *Ixodes scapularis* cell
701 line, ISE6. *F1000Research* 7, 297.
702 https://doi.org/10.12688/f1000research.13635.1
703 54. Benoit, J. (in prep.),
704 55. Branton, D. *et al.* (2008) The potential and challenges of nanopore sequencing.
705 *Nature Biotechnology* 26, 1146–1153. https://doi.org/10.1038/nbt.1495
706 56. Mostafa, H.H. (2024) An evolution of Nanopore next-generation sequencing
707 technology: implications for medical microbiology and public health. *Journal of*
708 *clinical microbiology* 62, e00246–00224. https://doi.org/10.1128/jcm.00246-24
709 57. Luan, T. *et al.* (2024) Benchmarking short and long read polishing tools for
710 nanopore assemblies: achieving near-perfect genomes for outbreak isolates.
711 *BMC genomics* 25, 679. https://doi.org/10.1186/s12864-024-10582-x
712 58. Dekker, J. *et al.* (2002) Capturing Chromosome Conformation. *Science* 295,
713 1306–1311. https://doi.org/10.1126/science.1067799
714 59. Lieberman-Aiden, E. *et al.* (2009) Comprehensive Mapping of Long-Range
715 Interactions Reveals Folding Principles of the Human Genome. *Science* 326,
716 289–293. https://doi.org/10.1126/science.1181369
717 60. Guizzo, M.G. *et al.* (2023) A comparison of Illumina and PacBio methods to build
718 tick salivary gland transcriptomes confirms large expression of lipocalins and
719 other salivary protein families that are not represented in available tick genomes.
720 *Ticks and tick-borne diseases* 14, 102209.
721 https://doi.org/10.1016/j.ttbdis.2023.102209

722 61. Stanke, M. *et al.* (2008) Using native and syntenically mapped cDNA alignments
723 to improve de novo gene finding. *Bioinformatics* 24, 637–644.
724 https://doi.org/10.1093/bioinformatics/btn013
725 62. Stanke, M. *et al.* (2006) Gene prediction in eukaryotes with a generalized hidden
726 Markov model that uses hints from external sources. *BMC bioinformatics* 7, 62.
727 https://doi.org/10.1186/1471-2105-7-62
728 63. Thibaud-Nissen, F. *et al.* (2013) Eukaryotic genome annotation pipeline. *The*
729 *NCBI Handbook* 2,
730 64. Flynn, J.M. *et al.* (2020) RepeatModeler2 for automated genomic discovery of
731 transposable element families. *Proceedings of the National Academy of Sciences*
732 117, 9451–9457. https://doi.org/10.1073/pnas.1921046117
733 65. Smit, A. *et al.* (2015). RepeatMasker Open-4.0. 2013–2015. Seattle, USA
734 66. Feron, R. and Waterhouse, R.M. (2022) Assessing species coverage and
735 assembly quality of rapidly accumulating sequenced genomes. *GigaScience* 11,
736 giac006. https://doi.org/10.1093/gigascience/giac006
737 67. Geraci, N.S. *et al.* (2007) Variation in genome size of argasid and ixodid ticks.
738 *Insect Biochemistry and Molecular Biology* 37, 399–408.
739 https://doi.org/10.1016/j.ibmb.2006.12.007
740 68. Van Zee, J.P. *et al.* (2007) Tick genomics: the *Ixodes* genome project and
741 beyond. *International journal for parasitology* 37, 1297–1305.
742 https://doi.org/10.1016/j.ijpara.2007.05.011
743 69. Murgia, M.V. *et al.* (2019) Meeting the challenge of tick-borne disease control: a
744 proposal for 1000 *Ixodes* genomes. *Ticks and Tick-borne Diseases* 10, 213–218.
745 https://doi.org/10.1016/j.ttbdis.2018.08.009
746 70. Sharakhova, M.V. *et al.* (2018) Physical genome mapping using fluorescence in
747 situ hybridization with mosquito chromosomes. In *Insect Genomics: Methods and*
748 *Protocols*, pp. 177–194, Springer
749 71. Kim, B.Y. *et al.* (2024) Single-fly genome assemblies fill major phylogenomic
750 gaps across the Drosophilidae Tree of Life. *PLoS Biology* 22, e3002697.
751 https://doi.org/10.1371/journal.pbio.3002697
752 72. Buenrostro, J.D. *et al.* (2013) Transposition of native chromatin for fast and
753 sensitive epigenomic profiling of open chromatin, DNA-binding proteins and
754 nucleosome position. *Nature methods* 10, 1213–1218.
755 https://doi.org/10.1038/nmeth.2688
756 73. Nurk, S. *et al.* (2022) The complete sequence of a human genome. *Science* 376,
757 44–53. https://doi.org/10.1126/science.abj6987
758 74. Frederick, J. *et al.* (2022) Phylogeography of the blacklegged tick (*Ixodes*
759 *scapularis*) throughout the USA identifies candidate loci for differences in
760 vectorial capacity. *Molecular Ecology* 32. https://doi.org/10.1111/mec.16921
761 75. Willadsen, P. (2004) Anti-tick vaccines. *Parasitology* 129, S367–S387.
762 https://doi.org/10.1017/S0031182003004657

763