International Journal for Parasitology xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

Comparative analysis of Presence-Absence gene Variations in five hard tick species: impact and functional considerations

Umberto Rosani [a,*], Marco Sollitto [b,c], Nicolò Fogal [a], Cristiano Salata [d]

[a] Department of Biology, University of Padova, 35121 Padova, Italy
[b] Department of Life Science, University of Trieste, 34100 Trieste, Italy
[c] Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Glagoljaška 8, 6000 Koper, Slovenia
[d] Department of Molecular Medicine, University of Padova, 35121 Padova, Italy

ARTICLE INFO

Article history:
Received 25 January 2023
Received in revised form 6 August 2023
Accepted 31 August 2023
Available online xxxx

Keywords:
Tick
Presence Absence Variation
PAV
Tick population
Mobile genetic elements
Host-pathogen interactions

ABSTRACT

Tick species are vectors of harmful human and animal diseases, and their expansion is raising concerns under the global environmental changes' scenario. Ticks host and transmit bacteria, protozoa and viruses, making the understanding of host-pathogen molecular pathways critical to development of effective disease control strategies. Despite the considerable sizes and repeat contents of tick genomes, individual tick genomics is perhaps the most effective approach to reveal genotypic traits of interest. Presence-Absence gene Variations (PAVs) can contribute to individual differences within species, with dispensable genes carried by subsets of individuals possibly underpinning functional significance at individual or population-levels. We exploited 350 resequencing datasets of *Dermacentor silvarum, Haemaphysalis longicornis, Ixodes persulcatus, Rhipicephalus microplus* and *Rhipicephalus sanguineus* hard tick specimens to reveal the extension of PAV and the conservation of dispensable genes among individuals and, comparatively, between species. Overall, we traced 550–3,346 dispensable genes per species and were able to reconstruct 5.3–7 Mb of genomic regions not included in the respective reference genomes, as part of the tick pangenomes. Both dispensable genes and *de novo* predicted genes indicated that PAVs preferentially impacted mobile genetic elements in these tick species.

1. Introduction

Ticks (subphylum Chelicerata, suborder Ixodida) are ectoparasites of animals and vectors of infectious agents (Zhang et al., 2019). Lyme disease, tick-borne encephalitis, haemorrhagic fevers, tick-borne macular fever, Q-fever, babesiosis and tick paralysis are well-known tick-borne diseases, which threaten human and animal health with relevant medical costs and detrimental production losses (Medlock and Leach, 2015). The geographical distribution of ticks is expanding latitudinally and altitudinally, with mild winters, enlargements of the areas with tick host species and an increased use of recreational areas, further exacerbating the spread of vector-borne diseases (Randolph and Rogers, 2007; Medlock and Leach, 2015; Mysterud et al., 2017; Piotrowski and Rymaszewska, 2020; Van Gestel et al., 2021; Zhao et al., 2021). In this scenario, monitoring activities and educational norms to prevent tick bites (Mowbray et al., 2012; Beaujean et al., 2016) should be supported by approaches aiming to control tick diffusion and pathogen transmission (Ostfeld et al., 2006). Since chemical compounds and bioagents have shown limited efficacy (Eisen and Stafford, 2021; Keesing et al., 2022), the development of alternative approaches, possibly based on genomic tools, is required (Murgia et al., 2019). A better knowledge on tick transcriptomes and proteomes, and a progressive understanding of microbiome structures and interactions, will help to further pursue genomics-based control strategies (Kurokawa et al., 2020; Orozco Orozco et al., 2021; Yang et al., 2021). Recently, use of the CRISPR-Cas9-based genome manipulation technique has been tested on the black-legged tick *Ixodes scapularis*, to develop a protocol for Cas9-mediated control of tick-borne diseases such as Lyme disease (Sharma et al., 2022).

Accordingly, sequencing and analysis of individual genomes is of paramount importance to detect macro- and micro-evolutionary dynamics between tick population and species, moving beyond single nucleotide variations (SNVs) or phylogenetics (Murgia et al., 2019). In this context, comparative analyses focused on determination of the individual gene contents can further contribute to understand population or individual resistance or sus-

* Corresponding author at: via U. Bassi 58/b, 35121 Padova, Italy.
E-mail address: umberto.rosani@unipd.it (U. Rosani).

https://doi.org/10.1016/j.ijpara.2023.08.004
0020-7519/

ceptibility traits. Genome rearrangements, as large insertions and deletions, have been linked to phenotypic traits in *Arabidopsis thaliana*, yeast (Muramoto et al., 2018) and fungi (Rajeh et al., 2018), or to pathogenic traits in bacteria (Seferbekova et al., 2021). Limited to the gene regions, Presence-Absence gene Variations (PAVs) could impact the number of genes in individual genomes (Golicz et al., 2016), possibly modifying the capacity to react to the environment and to pathogenic insults (Takahashi-Kariyazono et al., 2020). Disentangling PAV consists of the identification of core genes, present in all the analysed individuals, and of dispensable genes which can be absent in a variable number of individuals. The extension of PAVs greatly varies across the tree of life, as shown for bacteria, archaea (Hwang and Girguis, 2022), plants (Song et al., 2020), and recently in invertebrate molluscs and in a stony coral (Gerdol et al., 2020; Takahashi-Kariyazono et al., 2020; Calcino et al., 2021). The latter finding suggested that, although with variable extensions, PAV contributed to definition of a pangenome in metazoans (Golicz et al., 2020).

The genome of tick species is characterized by a relevant size, complexity and repetitiveness, making whole genome resequencing projects expensive and their analysis computationally demanding (de la Fuente et al., 2016). Although the mission of sequencing thousand tick genomes has been proposed (Murgia et al., 2019), so far, a limited number of genomes have been deposited in public databases, covering 12 species of six tick genera (NCBI genome database, accessed January 2023). A total of 72 whole-genome resequencing samples which originated from a single study on *Ixodes scapularis* are listed in Vectorbase (Amos et al., 2022). Elsewhere, whole genome sequencing data of six tick species collected in China have been deposited and currently represent the major repository of tick resequencing data (Jia et al., 2020).

In this study we aimed to analyse individual tick genome resequencing data to reveal (i) if PAV is present among hard ticks; (ii) the extent of PAV in the different species and the identities of dispensable genes and (iii) if patterns of dispensable genes are present and conserved between different tick species. Overall, our study aimed to expand knowledge on tick genomics, providing data which may be useful in identifying targets for innovative biocontrol approaches, as well as to provide a first known insight into PAVs and pangenome composition in hard tick species.

2. Materials and methods

2.1. Data retrieval and preliminary analysis

Whole genome resequencing data of *Dermacentor silvarum, Haemaphysalis longicornis, Ixodes persulcatus, Rhipicephalus microplus* and *Rhipicephalus sanguineus* hard tick species were retrieved from the Chinese National Genomics Data Center database (https://bigd.big.ac.cn) under the project accession ID PRJCA002242. A total of 70 datasets per species (n = 350) were randomly selected and downloaded in order to obtain datasets of a similar size for each species (Supplementary Table S1). The reference genomes of the five tick species were retrieved from GenBank, with accession IDs JABSTQ000000000-JABSTV000000000, together with the corresponding gene annotations. The predicted coding sequences (CDS) were used to assess the completeness of genomes, using the Benchmarking Universal Single-Copy Orthologs software (BUSCO v5.2.2) (Simão et al., 2015), testing the arthropoda_odb10 dataset. Complete BUSCO genes present in single copies were extracted and used later in the analysis. The predicted tick proteomes were functionally annotated using *Interproscan* v.5.60–92.0 (Jones et al., 2014), in order to identify conserved autonomous structural unit (domains from the Pfam database, https://www.ebi.ac.uk/interpro/) domains associated with the predicted proteins.

2.2. PAV analysis

To trace the set of 'dispensable genes' in individual tick genomes we adopted the pipeline described elsewhere (https://github.com/Carmen-Tuc/PAV_pipeline) (Sollitto et al., 2022). Briefly, raw reads were trimmed for quality and to remove adapter sequences using *fastp* v0.23.1 (Chen et al., 2018) with the following parameters: -V --detect_adapter_for_pe -w 16 -x -g -n 2–5 −3 -p -l 75. Trimmed reads were mapped on the corresponding reference genome using *bwa mem* (Li and Durbin, 2009), allowing read multi-mapping. For each mapping file, the per base coverage along the genome was determined using *samtools depth* (Li et al., 2009) and compared with the expected coverage determined, based on the BUSCO single copy complete genes previously identified. PAV events, defined as genes with coverage below 1/8 of the median coverage, were listed for each dataset and counted per species. To remove less relevant hits, only the PAV events present in at least 10% of the samples of a given species were called dispensable genes and further considered. The number and percentage of mapped reads per sample were extracted and used to detect sample outliers, which were removed from the subsequent analysis. As a result, we considered 70 samples for *D. silvarum*, 52 samples for *H. longicornis*, 51 samples for *I. persulcatus*, 60 samples *for R. microplus* and 41 samples for *R. sanguineous*. The reads that could not be mapped on the reference genome were used to reconstruct the *de novo* pangenome for each species by using the RePaRe pipeline (https://github.com/IlFog/RePaRe). First the unmapped genomic reads were recovered from the mapping files in bam format using *samtools*, using the command *samtools view -b -f 4*, then the reads were recursively assembled per sample with *SPAdes v. 3.15.5* (Nurk et al., 2017), applying the *–isolate* flag. Scaffolds were filtered as follows: minimal allowed length 1,000 bp; GC content in a range between the 5th and the 95th quantiles of the GC content in the reference genome; coverage depth between 65% of the hemizygous peak and 135% of the homozygous peak. Moreover, a taxonomic filter was also applied by using *diamond v. 2.0.6* (Buchfink et al., 2015) *blastx* searches against the NCBI *nr* database and selecting only scaffolds with a "bilateria" taxonomic classification. The selected scaffolds for each sample were iteratively added to the pangenome, indexed and used for read mapping in order to recover and analyse only the unmapped reads of the following samples. The *FasTE* pipeline (Bell et al., 2022) was used for *de novo* transposable element (TE) library generation and subsequent TE screening using the filtered pangenomes as input files. This pipeline exploits semi-automated techniques to identify repetitive sequences and regions of low complexity, employing tools such as *EDTA* v1.9.9 (Ou et al., 2022) and *DeepTE* (Yan et al., 2020). We followed executed *RepeatMasker* v4.1.2 (http://www.repeatmasker.org/RepeatMasker/) to soft-mask the genome, using the previous resultant library. The masked pangenome scaffolds were subjected to annotation of coding genes using BRAKER2, exploiting a combination of gene prediction tools such as *GeneMark* (Brůna et al., 2020) and *Augustus* (Stanke et al., 2006), and *ProtHint* (Brůna et al., 2020), a tool which analyses protein homology information using the *OrthoDB* v. 10.0 reference proteome of Arthropoda (Kriventseva et al., 2019).

2.3. Analysis of dispensable genes and data visualization

An enrichment test analysis was computed on the Pfam domains using an in-house python script available at https://gitlab.com/54mu/enrichment_test and based on an hypergeometric test (Falcon and Gentleman, 2008). Enriched annotations were filtered based on an arbitrary threshold of difference between observed and expected counts (>2.5) for any annotation in the dispensable gene subset. Only annotations displaying a False Discovery Rate (FDR) corrected *P*-value (Benjamini and Hochberg, 1995)

lower than 0.05 were considered significantly enriched. Data were analysed and rendered with RStudio 2022.07.2 + 576, using the *heatmaply*, *tidyverse*, *ggplot2*, *ggpubr* and *smplot2* packages.

2.4. Analysis of a putative I. persulcatus endosymbiont scaffold

To assess the taxonomic classification of a putative endosymbiont-derived scaffold identified in the *I. persulcatus* genome (GWHAMMH00005687.1), CAT v5.0.4 was used (von Meijenfeldt et al., 2019), based on prodigal v2.6.3 (Hyatt et al., 2010) and DIAMOND v2.0.6 (Buchfink et al., 2015), with databases downloaded on 1st January 2022.

2.5. Data accessibility

The relevant data associated with this paper are available as indicated, either publicly available or deposited as supplementary materials. The used codes are available in *github*, at the following links: https://github.com/Carmen-Tuc/PAV_pipeline and https://github.com/IlFog/RePaRe. Additional supplementary data have been deposited in MendeleyData (Rosani, Umberto; Sollitto, Marco; Fogal, Nicolò; Salata, Cristiano (2023), "Comparative analysis of Presence-Absence gene Variations in five hard tick species: impact and functional considerations", Mendeley Data, V2, https://doi.org/10.17632/by5kgmy637.2). Supplementary Data S1-S5, interactive heatmaps, are available in Mendeley Data, V1, https://doi.org/10.17632/by5kgmy637.1).

3. Results

3.1. PAVs are present in the five studied tick species

A total of 350 whole genome resequencing datasets of five hard-tick species were used to comparatively evaluate the extension of PAVs among species and to identify dispensable genes. These DNA sequencing samples included more than 100 million reads each (Fig. 1A), resulting in an average genome coverage between 4.9X for *R. sanguineus* and 9.6X for *I. persulcatus* (Fig. 1B). The percentages of reads mapped on the five reference genomes indicated that most of these samples referred to clean tick preparations, with low contamination. A limited number of *D. silvarum*, *H. longicornis* and *I. persulcatus* samples showed, however, significantly lower percentages of mapped reads (Fig. 1C).

PAV analysis revealed 286,189 events, impacting 1186–9522 genes per species for a total of 29,499 genes in the five species. In *H. longicornis* the PAV phenomenon appeared more evident, both considering the total number of events and the distribution of events per individual (Table 1, Fig. 2A–B). Outliers, namely individuals with significantly higher numbers of PAV events, were present for *H. longicornis* and *R. sanguineus*. Notably, most of the PAV events occurred with low frequencies in the analysed individuals (Fig. 2C), and setting a minimal frequency of 10% to consider a PAV event as a dispensable gene, we obtained 550–3346 dispensable genes per species (Table 1, Fig. 2D). *Dermacentor silvarum* and *R. microplus* possessed the highest values of individuals impacted by dispensable genes, with a median of 30, followed by *H. longicornis* (23), *I. persulcatus* (19) and *R. sanguineus* with 11.

3.2. De novo reconstruction of tick pangenomes

Next, we exploited the reads not mapped to the reference genomes to reconstruct tick pangenomes. Samples with a significantly lower mapping rate compared with the species average were excluded from the analysis to avoid, as much as possible, the inclusion of the contribution of possible hybrids in the pangenome. After read assemblies, filters based on the expected coverage, GC content and taxonomic composition were used to remove non-tick scaffolds. Interestingly, the majority of non-tick scaffolds appeared to have likely originated from co-sequenced organisms, primarily Proteobacteria and Ascomycota (Supplementary Fig. S1, Table 2). Only a fraction of the assembled scaffolds were therefore selected, resulting in 5.3–7.0 Mb of genomic region, generally highly fragmented, considered as the genuine pangenomes of these

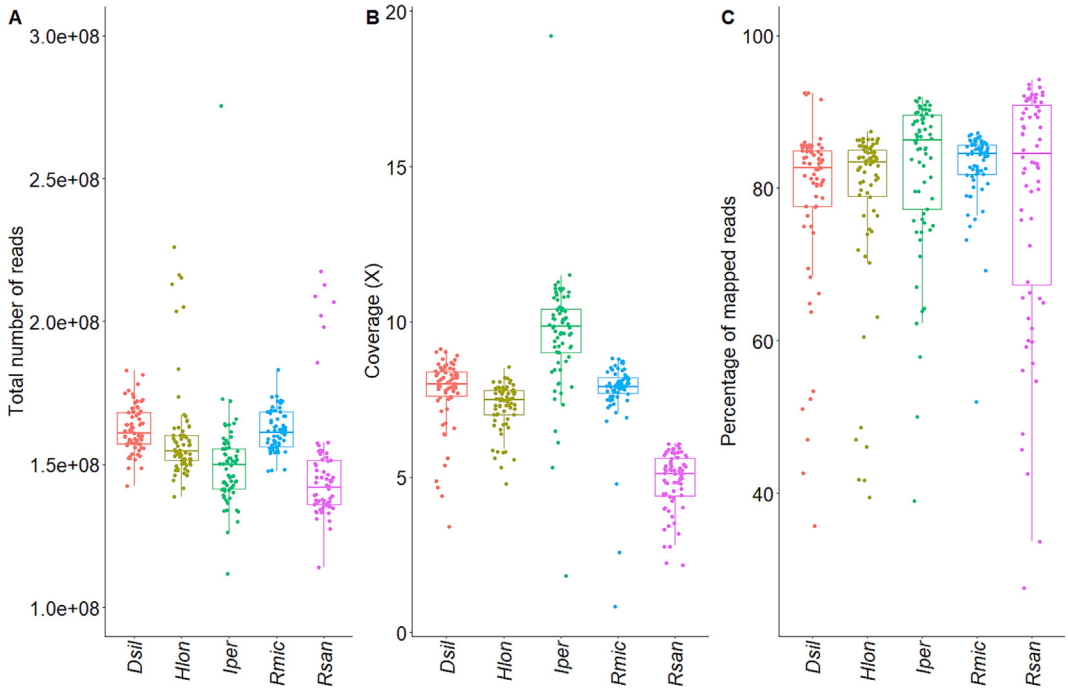

Fig. 1. Read mapping on the studied tick reference genomes. The distribution of the number of reads per sample (A), of the genome coverage (B) and of the percentages of mapped reads (C) were reported per species. *Dsil*, *Dermacentor silvarum; Hlon, Haemaphysalis longicornis; Iper, Ixodes persulcatus; Rmic, Rhipicephalus microplus* and *Rsan, Rhipicephalus sanguineus*.

Table 1
Summary of the Presence Absence Variation (PAV) analysis for the five species studied.

PAV analysis	*Dermacentor silvarum*	*Haemaphysalis longicornis*	*Ixodes persulcatus*	*Rhipicephalus microplus*	*Rhipicephalus sanguineus*
Total number of PAV events	19,735	110,836	33,015	68,816	53,787
No. of impacted genes	1186	8441	4019	9522	6331
Dispensable genes (frequency >10%)	550	3346	1081	1749	1977

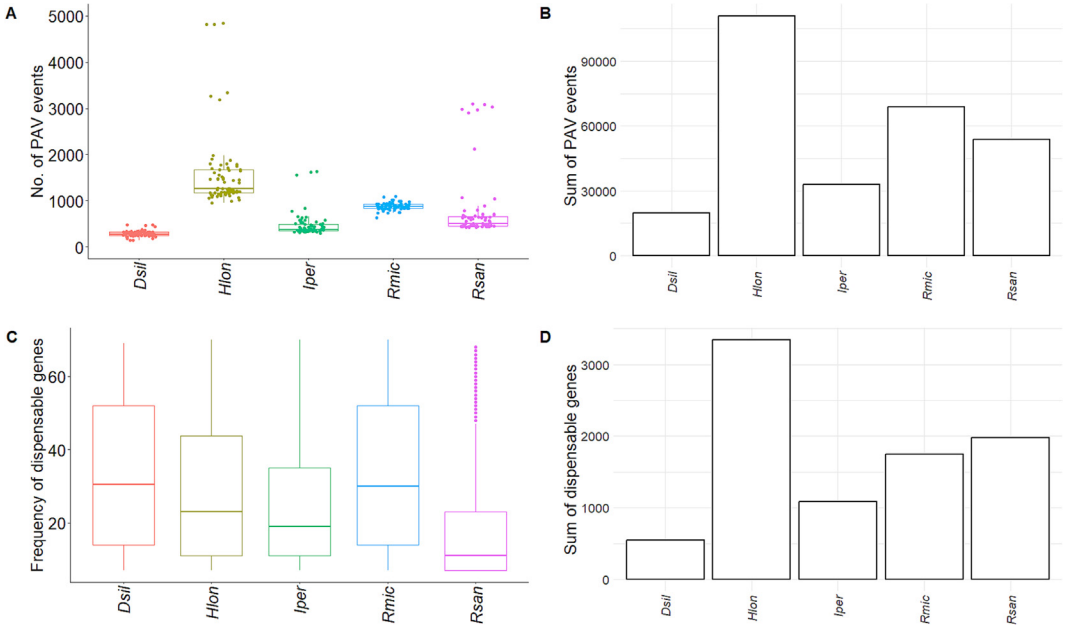

Fig. 2. Presence-Absence gene Variation (PAV) analysis of five tick species. (A) Distribution of the PAV events per individual among the five tick species. (B) Totals of the PAV events per species. (C) Distribution of the number of dispensable genes per individual. (D) Sum of dispensable genes. *Dsil*, *Dermacentor silvarum; Hlon, Haemaphysalis longicornis; Iper, Ixodes persulcatus; Rmic, Rhipicephalus microplus* and *Rsan, Rhipicephalus sanguineus.*

Table 2
Summary of the tick pangenomes for the five species studied.

Species	Contigs pre-filtration	Length pre-selection (kb)	Contigs post-selection	Pangenome size (kb)	N50 (bp)	No. of predicted genes	No. of genes with conserved domains
Dermacentor silvarum	39,681	65,643	3685	5635	1489	1277	578
Haemaphysalis longicornis	42,059	66,559	3260	5302	1573	1401	552
Ixodes persulcatus	25,128	44,849	4146	7019	1678	1841	841
Rhipicephalus microplus	49,833	67,399	4093	6657	1569	1157	645
Rhipicephalus sanguineus	26,898	41,261	3003	4757	1552	745	355

N50, length of the contig at 50% of the total assembly length.

tick species (Table 2). The plot showing the contributions of the individuals to the pangenome indicated that for *H. longicornis* and *R. microplus* the number of analysed individuals was appropriate to reach a plateau, with an homogenous incremental contribution provided by most of the samples (Fig. 3A). However for the other species, the prevalent contribution of certain samples was evident, suggesting high genomic heterogenicity of the analysed samples (Fig. 3B). Gene prediction identified a range of 745–1841 genes on tick pangenomes, a fraction of them encoding conserved domains (Table 2).

3.3. PAVs mostly impacted mobile genetic elements

A total of 528 different Pfam domains are encoded by the five sets of dispensable genes. The correlation between the abundance of Pfam domains in the proteomes and in the corresponding list of dispensable genes is very low, except for *H. longicornis, R. sanguineous* and *I. persulcatus* (R^2 = 0.54, 0.41 and 0.60, respectively, Supplementary Fig. S2). The most abundant Pfam domains, both considering the per species distribution and the overall abundance, are characteristic of mobile genetic elements (Fig. 4A–B). These included reverse transcriptases, transposases and endonucleases (RVT_1, Tnp_P_element, DDE_1, DDE_Tnp_4, Exo_endo_phos_2, HTH_Tnp_4 domains), domain of unknown function (DUF1759), which function is linked to retrotransposons, and DNA binding motifs (THAP, CENP-B_N, Myb_DNA-bind_5 and HTH_Tnp_Tc5). Considering the *de novo* predicted genes in the five pangenomes (Table 2), we identified 1434 unique Pfam domains, 50% of them represented only once. Considering the distribution per species, *D. silvarum* and *I. persulca-*

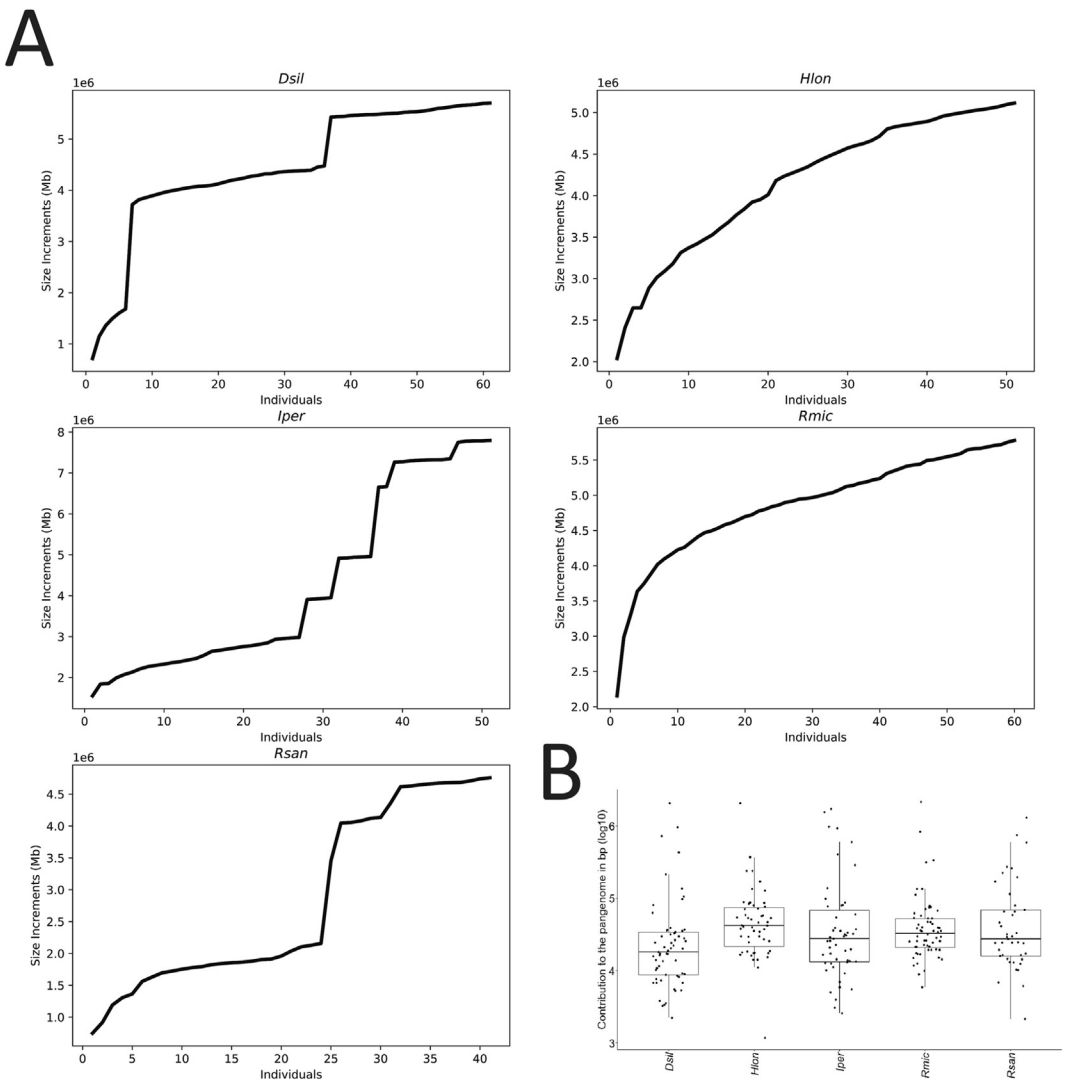

Fig. 3. Pangenome size of five tick species. (A) The plot depicted the incremental contribution of each individual to the species pangenome, for the five species. (B) The distribution of the sizes of the pangenome scaffolds is reported in a log10 scale. *Dsil*, *Dermacentor silvarum; Hlon, Haemaphysalis longicornis; Iper, Ixodes persulcatus; Rmic, Rhipicephalus microplus* and *Rsan, Rhipicephalus sanguineus*.

tus showed a similar domain composition, with several endonucleases of the DDE superfamily, HTH elements, THAP and YqaJ (Fig. 4C). For *H. longicornis and Rhipicephalus* spp., the abundances of the top 10 domains appeared more widely distributed and involved different domains. However, the most abundant domains found in the *de novo* predicted genes mirrored the ones found on dispensable genes (Fig. 4D). Next, we ran a hypergeometric analysis to identify Pfam domains enriched in the set of dispensable genes and in the *de novo* predicted genes compared with the reference proteomes. As result, we identified 19, 21, 14, 11 and 11 domains enriched in *D. silvarum*, *H. longicornis, I. persulcatus, R. microplus* and *R. sanguineus,* respectively, for a total of 45 unique domains (Supplementary Table S2). ‘‘Exo_endo_phos_2” was the only domain shared in all five datasets, whereas most domains were species-specific (Fig. 5A). Considering the novel genes predicted from the pangenomes, slightly more enriched Pfam domains were present, supporting the same trend with mostly species-specific domains and only the ‘‘Reverse transcriptase (RNA-dependent DNA polymerase)” being shared by all five datasets (Fig. 5B and Supplementary Table S2).

3.4. Investigation of dispensable genomic regions revealed a putative endosymbiont in I. persulcatus

Finally, we produced heatmap-like visualizations of the five sets of dispensable genes, depicting the presence or absence of the dispensable genes among individuals (horizontal clustering) and the co-occurrence of those (vertical clustering, Fig. 6 and Supplementary Data S1–S5). In accordance with the total number of dispensable genes and their frequencies among individuals, heatmaps possess different colour patterns. In *R. sanguineus*, the dispensable genes clearly divided into three subsets of the samples, whereas a more homogenous distribution was detectable in *R. microplus*. The separate clustering of four *D. silvarum* and six *H. longicornis* samples matched their higher or lower percentages of genomic mapped reads and, consequently, their numbers of dispensable genes. In the heatmap of *I. persulcatus*, we identified a cluster of co-occurring dispensable genes in 22–25 individuals, which also maintained their genomic order (Fig. 6B). This is indicative of a larger genomic region absent in these genomes, namely a 1.45 Mb scaffold with 14 annotated genes (GWHAMMH00005687.1, Fig. 6C). A search of open reading frames with prodigal revealed

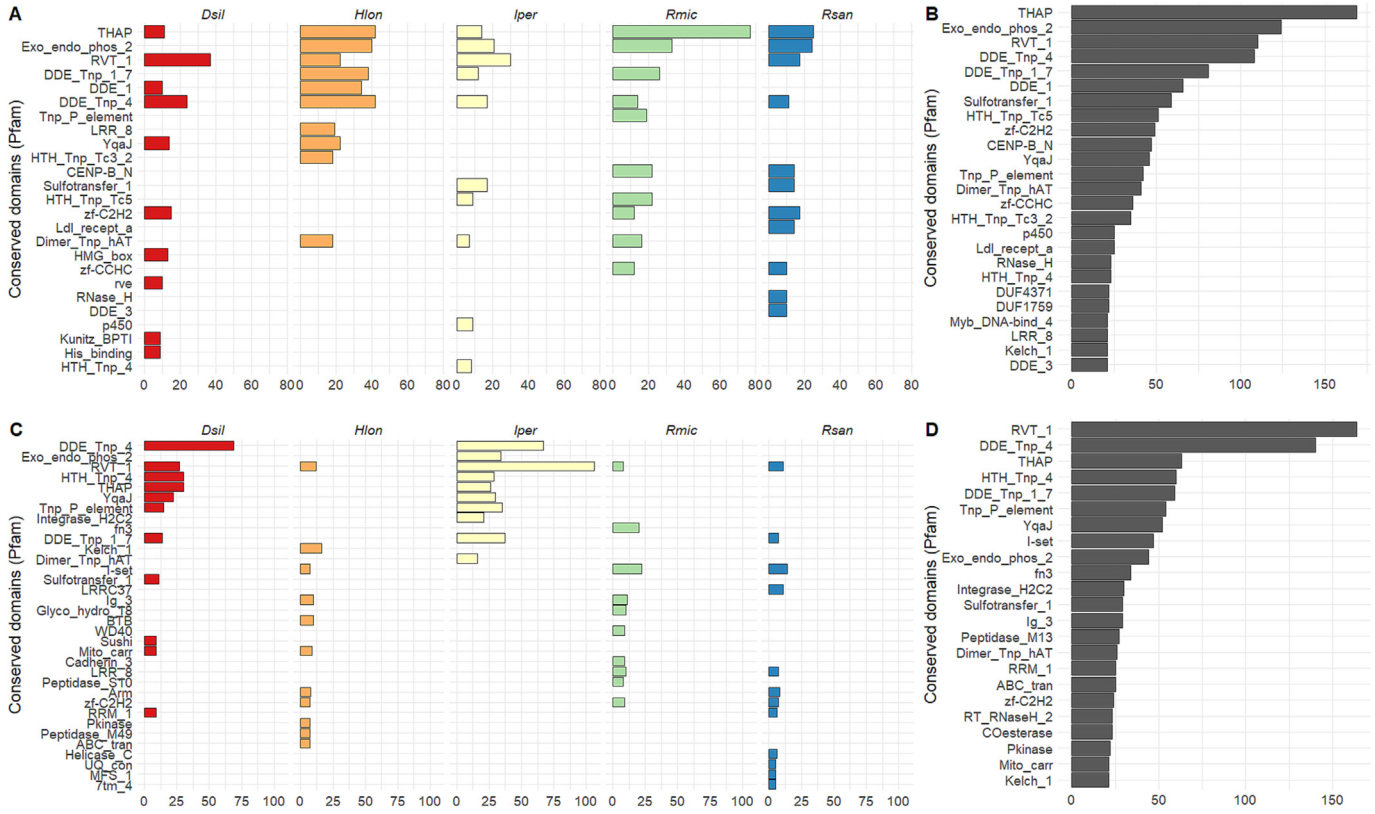

Fig. 4. Conserved protein domains (Pfam) encoded by dispensable and de novo genes of five tick species. The plot depicted the counts of the top10 domains per species encoded by the dispensable genes (A) and the total abundance (B), as well as the top10 domains per species encoded by the de novo genes (C) and the total abundance (D). *Dsil*, *Dermacentor silvarum; Hlon, Haemaphysalis longicornis; Iper, Ixodes persulcatus; Rmic, Rhipicephalus microplus* and *Rsan, Rhipicephalus sanguineus*.

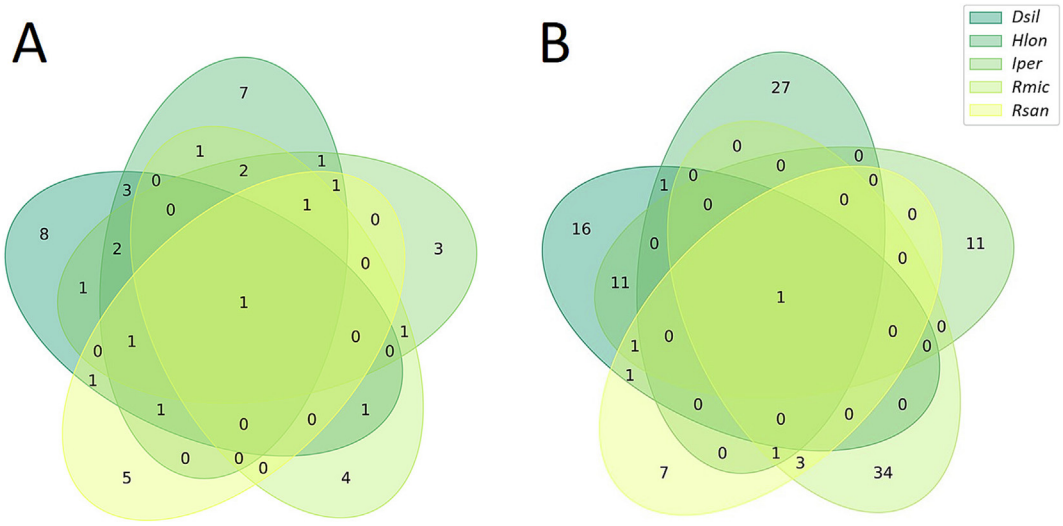

Fig. 5. Venn diagram depicting the shared and exclusive enriched conserved protein domains (Pfam database) found in the dispensable genes (A) and in the de novo genes (B) of five tick species. *Dsil*, *Dermacentor silvarum; Hlon, Haemaphysalis longicornis; Iper, Ixodes persulcatus; Rmic, Rhipicephalus microplus* and *Rsan, Rhipicephalus sanguineus.*

1,412 genes, and their taxonomic annotation with CAT showed their similarity with Rickettsiales (Fig. 6D).

4. Discussion

We report the first known comparative investigation of the pangenomes of five hard tick species. By exploiting 350 whole genome resequencing datasets, we were able to trace PAVs involving coding genes and to evaluate the extent of this phenomenon. Moreover, we tentatively reconstructed tick pangenomes, composed of 5.3–7 Mb of DNA regions not present in the corresponding reference genomes. We found clear evidence that, although to different extents among species, PAV is present in all five species, by identification of hundreds of dispensable genes. Among arthropods, pangenome analysis has been carried out for the silkworm (Tong et al., 2022), for the butterfly *Heliconius charithonia* (Ruggieri et al., 2022), and a pangenome-like analytical approach has been used in *Aedes aegypti* to reveal structural genomic differences among strains (Fisher et al., 2022). More recently, a

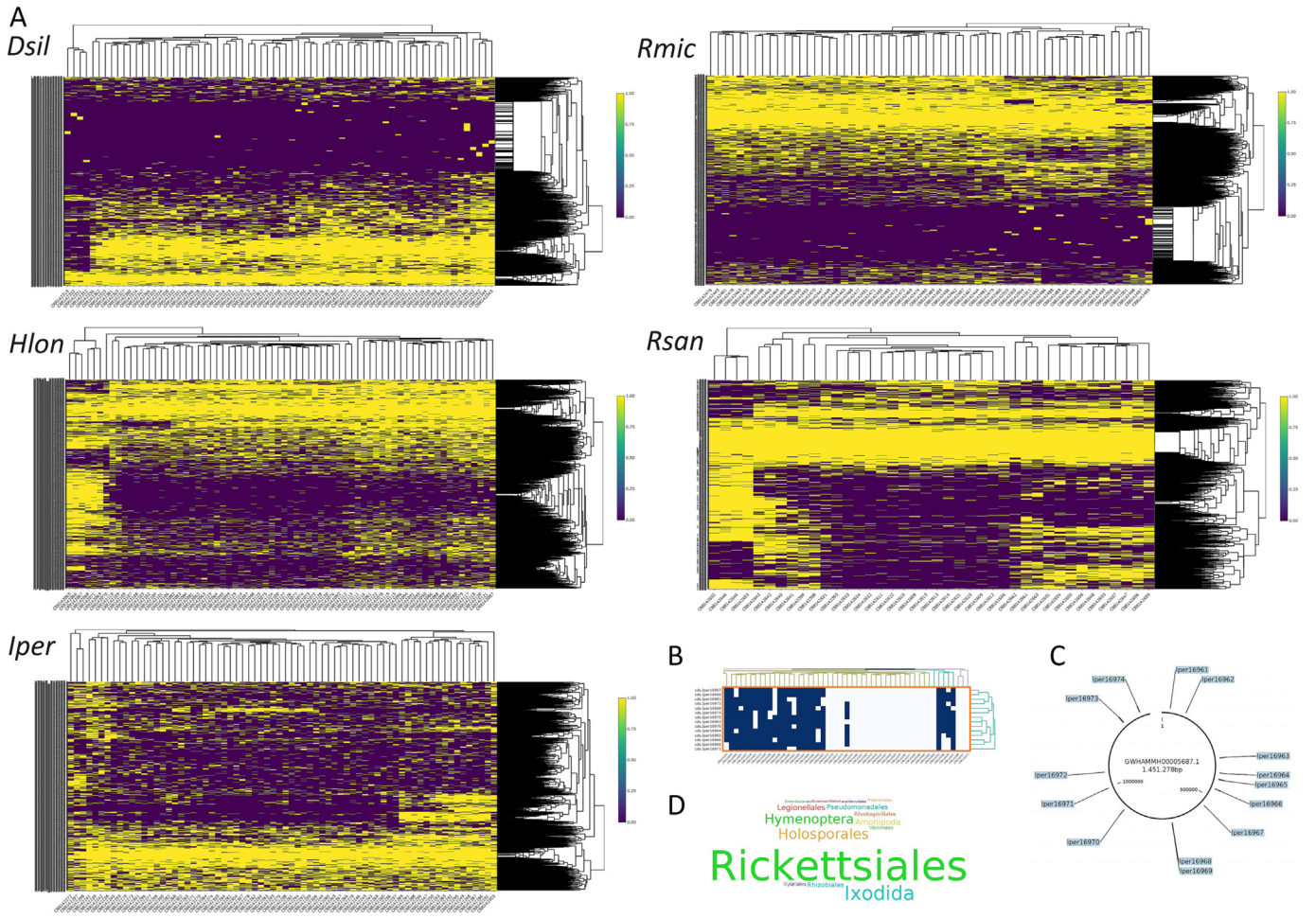

Fig. 6. Heatmap of dispensable genes in five tick species. (A) The heatmaps show the distribution of dispensable genes in the tested individuals per species. The absent genes are reported in yellow and in violet if present. Hierarchical clustering has been applied to the horizontal axis (sample) and to the vertical axis (genes). In orange is highlighted a set of dispensable genes shared in 22 individuals, magnified (B). These genes are located on a single scaffold (GWHAMMH00005687.1) (C). (D) Word graph depicting the most frequent taxonomic terms when all the newly predicted genes identified on GWHAMMH00005687.1 were subjected to taxonomic classifications. *Dsil*, *Dermacentor silvarum; Hlon, Haemaphysalis longicornis; Iper, Ixodes persulcatus; Rmic, Rhipicephalus microplus* and *Rsan, Rhipicephalus sanguineus.*

pangenome analysis has been carried out for *Apis cerana*, revealing more than 100 Mb of pangenome regions not included in the reference sequence, suggesting a possible role in environmental adaptation (Li et al., 2023). However, the published literature does not include pangenome analysis for tick species. For these five tick species, most of the dispensable genes are individual- or species-specific, i.e. present in less than 10% of the tested individuals or only in a given species, respectively. The former likely represented random individual genetic variations, whereas species-specific dispensable genes can underpin functional significance. Although the frequency of dispensable genes among the 70 tested individuals per species did not greatly differ, we observed a different extent of PAV among single species, particularly evident for *H. longicornis,* resulting in a considerably higher number of observed events. No difference in the extent of PAV was observed between one-host (R. microplus) and three-host tick species. Of note, a certain bias might be introduced by the presence of hybrids in the sequencing pool, evident from the lower mapping rates and/or extremely high number of PAV events. A similar situation characterized six *H. longicornis* samples, which had a lower mapping percentage and so clustered separately in the dispensable gene heatmap. Differently, four samples of *D. silvarum* displayed higher mapping percentages, leading to a new cluster in the dispensable gene heatmap. Although we aimed to remove these samples from the analysis by setting a minimal mapping rate threshold, the graph showing the individual contribution to each pangenome clearly indicated that specimens with marked genomic differences are still present, perhaps except for *H. longicornis* and *R. microplus*.

Regarding the identity of the PAV genes, both considering the dispensable genes and the *de novo* genes, it becomes evident that there is a preponderance of transposable elements. Intriguingly, the *de novo* genes showed different domain compositions, with *D. silvarum* and *I. persulcatus* genes mostly encoding domains of endonuclease and transposable genes, whereas the other species showed more broad and homogenous domain distributions. Due to the fact that the adopted analytical pipeline allowed multiple matches during read mapping, we excluded that the abundant presence of multi-copy genes in the list of dispensable genes can be linked to a technical bias related to their possible nucleotide similarity. However, we suggest that this mirrored the rapid and still active spread and diversification of these gene families in tick genomes. Indeed, a considerable presence and activity of mobile genetic elements has been reported in different tick genomes, including the ones we have considered (Gulia-Nuss et al., 2016; Jia et al., 2020), with the *I. scapularis* genome taken as an example of permissiveness in repeat accumulation (Gulia-Nuss et al., 2016). In this context, DUF1759 is the dispensable Pfam domain with the higher frequencies. This domain is related to LTR polyproteins or

retrotransposons and, together with THAP-domain containing proteins, catalyse the hydrolysis of genomic DNA within the cells that synthesize them and are implicated in driving genomes' dynamics (Rappoport and Linial, 2015). Notably, we revealed other dispensable genes not related to mobile genetic elements, suggesting a possible role of PAV in differentiating the individual ability to interact with pathogens and environmental factors, an hypothesis that requires more dedicated analysis.

Intriguingly, by analysing co-occurring dispensable genes in *I. persulcatus*, we identified one MutL DNA mismatch repair protein as dispensable in 31% of the individuals. This gene co-localized with 13 other genes in the heatmap and revealed a 1.45 Mb genomic region possibly impacted by PAV, although taxonomic analysis revealed similarity to a rickettsia-like organism. In fact, this scaffold is likely an almost complete rickettsia-like genome, confirming their abundance if considering the one recently reconstructed due to long-read sequencing in the last update of the *I. scapularis* genome (De et al., 2023). Improvements of the PAV analytical pipeline could be, therefore, planned to extend the analysis to the detection of endosymbionts and proviruses. However, the coverage to identify and close exogenous genomes (i.e. non-host genome) should be higher than the one applied for these samples that appeared to be borderline for performing PAV analysis.

Methodologically, the availability of genome resequencing datasets with an appropriate coverage allows the detection of dispensable genes by applying a subtractive approach, namely identifying the genes listed in the reference gene models which are absent in a given individual. On the other hand, to reveal genes not present in the reference genome, *de novo* pangenome reconstruction is necessary. The approach that we have adopted aimed to reconstruct pangenome regions based on unmapped short reads, revealing intrinsic limitations related to high scaffold fragmentation. Haplotype-resolved genomes based on long-read sequencing are probably more suitable to resolve tick pangenomes (Eizenga et al., 2020). In future, this will also allow the application of other available tools and pipelines tested for pangenome reconstruction in different species (Hu et al., 2017; Wang et al., 2018; Duan et al., 2019) and can contribute to reveal a more accurate picture of tick (pan)genomes.

Rapid advancement of tick genomics can support strategies for controlling tick-borne diseases (Murgia et al., 2019). In this framework, the analysis of individual tick genomes can contribute to untangling genotypic traits associated with vectorial capacity, host-pathogen interactions and susceptibility in tick populations as well as, comparatively, among tick species. Considering the above-discussed limitations, we successfully reported the existence of PAV in tick species. The impact of PAV can be seen as further evidence of the dynamically evolving genomes of ticks, although gene-centred studies can reveal the importance of specific dispensable genes. Overall, the analysis of PAV can contribute to better understanding of micro- and macro-evolutionary dynamics between and within tick populations, and indicates that individual tick genomic research should be prioritized.

Acknowledgements

This research was partially supported by EU funding within the MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2023.08.004.

References

Amos, B., Aurrecoechea, C., Barba, M., Barreto, A., Basenko, E.Y., Bażant, W., Belnap, R., Blevins, A.S., Böhme, U., Brestelli, J., Brunk, B.P., Caddick, M., Callan, D., Campbell, L., Christensen, M.B., Christophides, G.K., Crouch, K., Davis, K., DeBarry, J., Doherty, R., Duan, Y., Dunn, M., Falke, D., Fisher, S., Flicek, P., Fox, B., Gajria, B., Giraldo-Calderón, G.I., Harb, O.S., Harper, E., Hertz-Fowler, C., Hickman, M.J., Howington, C., Hu, S., Humphrey, J., Iodice, J., Jones, A., Judkins, J., Kelly, S.A., Kissinger, J.C., Kwon, D.K., Lamoureux, K., Lawson, D., Li, W., Lies, K., Lodha, D., Long, J., MacCallum, R.M., Maslen, G., McDowell, M.A., Nabrzyski, J., Roos, D.S., Rund, S.S.C., Schulman, S.W., Shanmugasundram, A., Sitnik, V., Spruill, D., Starns, D., Stoeckert Jr, C.J., Tomko, S.S., Wang, H., Warrenfeltz, S., Wieck, R., Wilkinson, P.A., Xu, L., Zheng, J., 2022. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. Nucleic Acids Res. 50, D898–D911. https://doi.org/10.1093/nar/gkab929.

Beaujean, D.J.M.A., Gassner, F., Wong, A., Steenbergen, J.E., Crutzen, R., Ruwaard, D., 2016. Education on tick bite and Lyme borreliosis prevention, aimed at schoolchildren in the Netherlands: comparing the effects of an online educational video game versus a leaflet or no intervention. BMC Public Health 16, 1163. https://doi.org/10.1186/s12889-016-3811-5.

Bell, E.A., Butler, C.L., Oliveira, C., Marburger, S., Yant, L., Taylor, M.I., 2022. Transposable element annotation in non-model species: the benefits of species-specific repeat libraries using semi-automated EDTA and DeepTE de novo pipelines. Mol. Ecol. Resour. 22, 823–833. https://doi.org/10.1111/1755-0998.13489.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B. Methodol. 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Brůna, T., Lomsadze, A., Borodovsky, M., 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics Bioinforma. 2, lqaa026. https://doi.org/10.1093/nargab/lqaa026.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176.

Calcino, A.D., Kenny, N.J., Gerdol, M., 2021. Single individual structural variant detection uncovers widespread hemizygosity in molluscs. Philos. Trans. R. Soc. Lond. B Biol. Sci. 376, 20200153. https://doi.org/10.1098/rstb.2020.0153.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinforma. Oxf. Engl. 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

De, S., Kingan, S.B., Kitsou, C., Portik, D.M., Foor, S.D., Frederick, J.C., Rana, V.S., Paulat, N.S., Ray, D.A., Wang, Y., Glenn, T.C., Pal, U., 2023. A high-quality *Ixodes scapularis* genome advances tick science. Nat. Genet. 1–11. https://doi.org/10.1038/s41588-022-01275-w.

de la Fuente, J., Waterhouse, R.M., Sonenshine, D.E., Roe, R.M., Ribeiro, J.M., Sattelle, D.B., Hill, C.A., 2016. Tick genome assembled: new opportunities for research on tick-host-pathogen interactions. Front. Cell. Infect. Microbiol. 6, 103. https://doi.org/10.3389/fcimb.2016.00103.

Duan, Z., Qiao, Y., Lu, J., Lu, H., Zhang, W., Yan, F., Sun, C., Hu, Z., Zhang, Z., Li, G., Chen, H., Xiang, Z., Zhu, Z., Zhao, H., Yu, Y., Wei, C., 2019. HUPAN: a pan-genome analysis pipeline for human genomes. Genome Biol. 20, 149. https://doi.org/10.1186/s13059-019-1751-y.

Eisen, L., Stafford, K.C., 2021. Barriers to Effective Tick Management and Tick-Bite Prevention in the United States (Acari: Ixodidae). J. Med. Entomol. 58, 1588–1600. https://doi.org/10.1093/jme/tjaa079.

Eizenga, J.M., Novak, A.M., Sibbesen, J.A., Heumos, S., Ghaffaari, A., Hickey, G., Chang, X., Seaman, J.D., Rounthwaite, R., Ebler, J., Rautiainen, M., Garg, S., Paten, B., Marschall, T., Sirén, J., Garrison, E., 2020. Pangenome Graphs. Annu. Rev. Genomics Hum. Genet. 21, 139–162. https://doi.org/10.1146/annurev-genom-120219-080406.

Falcon, S., Gentleman, R., 2008. Hypergeometric Testing Used for Gene Set Enrichment Analysis. In: Hahne, F., Huber, W., Gentleman, R., Falcon, S. (Eds.), Bioconductor Case Studies. Use R! Springer, New York, NY, pp. 207–220. https://doi.org/10.1007/978-0-387-77240-0_14.

Fisher, C.R., Wilson, M., Scott, J.G., 2022. A chromosome-level assembly of the widely used Rockefeller strain of *Aedes aegypti*, the yellow fever mosquito. G3 Bethesda Md 12, jkac242. https://doi.org/10.1093/g3journal/jkac242.

Gerdol, M., Moreira, R., Cruz, F., Gómez-Garrido, J., Vlasova, A., Rosani, U., Venier, P., Naranjo-Ortiz, M.A., Murgarella, M., Greco, S., Balseiro, P., Corvelo, A., Frias, L., Gut, M., Gabaldón, T., Pallavicini, A., Canchaya, C., Novoa, B., Alioto, T.S., Posada, D., Figueras, A., 2020. Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. Genome Biol. 21, 275. https://doi.org/10.1186/s13059-020-02180-3.

Golicz, A.A., Batley, J., Edwards, D., 2016. Towards plant pangenomics. Plant Biotechnol. J. 14, 1099–1105. https://doi.org/10.1111/pbi.12499.

Golicz, A.A., Bayer, P.E., Bhalla, P.L., Batley, J., Edwards, D., 2020. Pangenomics comes of age: from bacteria to plant and animal applications. Trends Genet. TIG 36, 132–145. https://doi.org/10.1016/j.tig.2019.11.006.

Gulia-Nuss, M., Nuss, A.B., Meyer, J.M., Sonenshine, D.E., Roe, R.M., Waterhouse, R.M., Sattelle, D.B., de la Fuente, J., Ribeiro, J.M., Megy, K., Thimmapuram, J., Miller, J.R., Walenz, B.P., Koren, S., Hostetler, J.B., Thiagarajan, M., Joardar, V.S., Hannick, L.I., Bidwell, S., Hammond, M.P., Young, S., Zeng, Q., Abrudan, J.L., Almeida, F.C., Ayllón, N., Bhide, K., Bissinger, B.W., Bonzon-Kulichenko, E., Buckingham, S.D., Caffrey, D.R., Caimano, M.J., Croset, V., Driscoll, T., Gilbert, D., Gillespie, J.J., Giraldo-Calderón, G.I., Grabowski, J.M., Jiang, D., Khalil, S.M.S., Kim, D., Kocan, K.M., Koči, J., Kuhn, R.J., Kurtti, T.J., Lees, K., Lang, E.G., Kennedy, R.C., Kwon, H.,

Perera, R., Qi, Y., Radolf, J.D., Sakamoto, J.M., Sánchez-Gracia, A., Severo, M.S., Silverman, N., Šimo, L., Tojo, M., Tornador, C., Van Zee, J.P., Vázquez, J., Vieira, F. G., Villar, M., Wespiser, A.R., Yang, Y., Zhu, J., Arensburger, P., Pietrantonio, P.V., Barker, S.C., Shao, R., Zdobnov, E.M., Hauser, F., Grimmelikhuijzen, C.J.P., Park, Y., Rozas, J., Benton, R., Pedra, J.H.F., Nelson, D.R., Unger, M.F., Tubio, J.M.C., Tu, Z., Robertson, H.M., Shumway, M., Sutton, G., Wortman, J.R., Lawson, D., Wikel, S.K., Nene, V.M., Fraser, C.M., Collins, F.H., Birren, B., Nelson, K.E., Caler, E., Hill, C.A., 2016. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. Nat. Commun. 7, 10507. https://doi.org/10.1038/ncomms10507.

Hu, Z., Sun, C., Lu, K., Chu, X., Zhao, Y., Lu, J., Shi, J., Wei, C., 2017. EUPAN enables pan-genome studies of a large number of eukaryotic genomes. Bioinformatics 33, 2408–2409. https://doi.org/10.1093/bioinformatics/btx170.

Hwang, Y., Girguis, P.R., 2022. Differentiated evolutionary strategies of genetic diversification in Atlantic and Pacific Thaumarchaeal populations. mSystems 7, , e0147721.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11, 119. https://doi.org/10.1186/1471-2105-11-119.

Jia, N., Wang, J., Shi, W., Du, L., Sun, Y., Zhan, W., Jiang, J.-F., Wang, Q., Zhang, B., Ji, P., Bell-Sakyi, L., Cui, X.-M., Yuan, T.-T., Jiang, B.-G., Yang, W.-F., Lam, T.-T.-Y., Chang, Q.-C., Ding, S.-J., Wang, X.-J., Zhu, J.-G., Ruan, X.-D., Zhao, L., Wei, J.-T., Ye, R.-Z., Que, T.C., Du, C.-H., Zhou, Y.-H., Cheng, J.X., Dai, P.-F., Guo, W.-B., Han, X.-H., Huang, E.-J., Li, L.-F., Wei, W., Gao, Y.-C., Liu, J.-Z., Shao, H.-Z., Wang, X., Wang, C.-C., Yang, T.-C., Huo, Q.-B., Li, W., Chen, H.-Y., Chen, S.-E., Zhou, L.-G., Ni, X.-B., Tian, J.-H., Sheng, Y., Liu, T., Pan, Y.-S., Xia, L.-Y., Li, J., Zhao, F., Cao, W.-C., 2020. Large-Scale comparative analyses of tick genomes elucidate their genetic diversity and vector capacities. Cell 182, 1328–1340.e13. https://doi.org/10.1016/j.cell.2020.07.023.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Keesing, F., Mowry, S., Bremer, W., Duerr, S., Evans, A.S., Fischhoff, I.R., Hinckley, A.F., Hook, S.A., Keating, F., Pendleton, J., Pfister, A., Teator, M., Ostfeld, R.S., 2022. Effects of tick-control interventions on tick abundance, human encounters with ticks, and incidence of tickborne diseases in residential neighborhoods, New York. USA. Emerg. Infect. Dis. 28, 957–966. https://doi.org/10.3201/eid2805.211146.

Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A., Zdobnov, E.M., 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 47, D807–D811. https://doi.org/10.1093/nar/gky1053.

Kurokawa, C., Lynn, G.E., Pedra, J.H.F., Pal, U., Narasimhan, S., Fikrig, E., 2020. Interactions between *Borrelia burgdorferi* and ticks. Nat. Rev. Microbiol. 18, 587–600. https://doi.org/10.1038/s41579-020-0400-5.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 (1000 Genome Project Data Processing Subgroup).

Li, Y., Yao, J., Sang, H., Wang, Q., Su, L., Zhao, X., Xia, Z., Wang, F., Wang, K., Lou, D., Wang, G., Waterhouse, R.M., Wang, H., Luo, S., Sun, C., 2023. Pan-genome analysis highlights the role of structural variation in the evolution and environmental adaptation of Asian honeybees. Biorxiv, pre-print. https://doi.org/10.1101/2023.06.15.545041.

Medlock, J.M., Leach, S.A., 2015. Effect of climate change on vector-borne disease risk in the UK. Lancet Infect. Dis. 15, 721–730. https://doi.org/10.1016/S1473-3099(15)70091-5.

Mowbray, F., Amlôt, R., Rubin, G.J., 2012. Ticking all the boxes? A systematic review of education and communication interventions to prevent tick-borne disease. Vector Borne Zoonotic Dis. 12, 817–825. https://doi.org/10.1089/vbz.2011.0774.

Muramoto, N., Oda, A., Tanaka, H., Nakamura, T., Kugou, K., Suda, K., Kobayashi, A., Yoneda, S., Ikeuchi, A., Sugimoto, H., Kondo, S., Ohto, C., Shibata, T., Mitsukawa, N., Ohta, K., 2018. Phenotypic diversification by enhanced genome restructuring after induction of multiple DNA double-strand breaks. Nat. Commun. 9, 1995. https://doi.org/10.1038/s41467-018-04256-y.

Murgia, M.V., Bell-Sakyi, L., de la Fuente, J., Kurtti, T.J., Makepeace, B.L., Mans, B., McCoy, K.D., Munderloh, U., Plantard, O., Rispe, C., Valle, M.R., Tabor, A., Thangamani, S., Thimmapuram, J., Hill, C.A., 2019. Meeting the challenge of tick-borne disease control: a proposal for 1000 Ixodes genomes. Ticks Tick-borne Dis. 10, 213–218. https://doi.org/10.1016/j.ttbdis.2018.08.009.

Mysterud, A., Jore, S., Østerås, O., Viljugrein, H., 2017. Emergence of tick-borne diseases at northern latitudes in Europe: a comparative approach. Sci. Rep. 7, 16316. https://doi.org/10.1038/s41598-017-15742-6.

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 27, 824–834. https://doi.org/10.1101/gr.213959.116.

Orozco Orozco, M., Gómez, G.F., Alzate, J.F., Isaza, J.P., Gutiérrez, L.A., 2021. Virome analysis of three Ixodidae ticks species from Colombia: a potential strategy for discovering and surveying tick-borne viruses. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 96, 105103. https://doi.org/10.1016/j.meegid.2021.105103.

Ostfeld, R.S., Price, A., Hornbostel, V.L., Benjamin, M.A., Keesing, F., 2006. Controlling ticks and tick-borne zoonoses with biological and chemical agents. Bioscience 56, 383–394. https://doi.org/10.1641/0006-3568(2006)056[0383:CTATZW]2.0.CO;2.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C.N., Hufford, M.B., 2022. Author correction: benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 23, 76. https://doi.org/10.1186/s13059-022-02645-7.

Piotrowski, M., Rymaszewska, A., 2020. Expansion of tick-borne rickettsioses in the world. Microorganisms 8, 1906. https://doi.org/10.3390/microorganisms8121906.

Rajeh, A., Lv, J., Lin, Z., 2018. Heterogeneous rates of genome rearrangement contributed to the disparity of species richness in Ascomycota. BMC Genomics 19, 282. https://doi.org/10.1186/s12864-018-4683-0.

Randolph, S.E., Rogers, D.J., 2007. Ecology of Tick-Borne Disease and the Role of Climate. In: Ergonul, O., Whitehouse, C.A. (Eds.), Crimean-Congo Hemorrhagic Fever: A Global Perspective. Springer, Netherlands, Dordrecht, pp. 167–186. https://doi.org/10.1007/978-1-4020-6106-6_14.

Rappoport, N., Linial, M., 2015. Trends in genome dynamics among major orders of insects revealed through variations in protein families. BMC Genomics 16, 583. https://doi.org/10.1186/s12864-015-1771-2.

Ruggieri, A.A., Livraghi, L., Lewis, J.J., Evans, E., Cicconardi, F., Hebberecht, L., Ortiz-Ruiz, Y., Montgomery, S.H., Ghezzi, A., Rodriguez-Martinez, J.A., Jiggins, C.D., McMillan, W.O., Counterman, B.A., Papa, R., Van Belleghem, S.M., 2022. A butterfly pan-genome reveals that a large amount of structural variation underlies the evolution of chromatin accessibility. Genome Res. 32, 1862–1875. https://doi.org/10.1101/gr.276839.122.

Seferbekova, Z., Zabelkin, A., Yakovleva, Y., Afasizhev, R., Dranenko, N.O., Alexeev, N., Gelfand, M.S., Bochkareva, O.O., 2021. High rates of genome rearrangements and pathogenicity of *Shigella spp*. Front. Microbiol. 12, 628622. https://doi.org/10.3389/fmicb.2021.628622.

Sharma, A., Pham, M.N., Reyes, J.B., Chana, R., Yim, W.C., Heu, C.C., Kim, D., Chaverra-Rodriguez, D., Rasgon, J.L., Harrell, R.A., Nuss, A.B., Gulia-Nuss, M., 2022. Cas9-mediated gene editing in the black-legged tick, *Ixodes scapularis*, by embryo injection and ReMOT Control. iScience 25 (3), 103781. https://doi.org/10.1016/j.isci.2022.103781.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Sollitto, M., Kenny, N.J., Greco, S., Tucci, C.F., Calcino, A.D., Gerdol, M., 2022. Detecting Structural VariantsStructural variants and Associated Gene Presence-Absence Variation Phenomena in the Genomes of Marine Organisms. In: Verde, C., Giordano, D. (Eds.), Marine Genomics: Methods and Protocols, Methods in Molecular Biology. Springer, US, New York, NY, pp. 53–76. https://doi.org/10.1007/978-1-0716-2313-8_4.

Song, J.-M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., Xie, W.-Z., Cheng, Y., Zhang, Y., Liu, K., Yang, Q.-Y., Chen, L.-L., Guo, L., 2020. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. Nat. Plants 6, 34–45. https://doi.org/10.1038/s41477-019-0577-7.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34, W435–W439. https://doi.org/10.1093/nar/gkl200.

Takahashi-Kariyazono, S., Sakai, K., Terai, Y., 2020. Presence–absence polymorphisms of single-copy genes in the stony coral *Acropora digitifera*. BMC Genomics 21, 158. https://doi.org/10.1186/s12864-020-6566-4.

Tong, X., Han, M.-J., Lu, K., Tai, S., Liang, S., Liu, Y., Hu, H., Shen, J., Long, A., Zhan, C., Ding, X., Liu, S., Gao, Q., Zhang, B., Zhou, L., Tan, D., Yuan, Y., Guo, N., Li, Y.-H., Wu, Z., Liu, L., Li, C., Lu, Y., Gai, T., Zhang, Y., Yang, R., Qian, H., Liu, Y., Luo, J., Zheng, L., Lou, J., Peng, Y., Zuo, W., Song, J., He, S., Wu, S., Zou, Y., Zhou, L., Cheng, L., Tang, Y., Cheng, G., Yuan, L., He, W., Xu, J., Fu, T., Xiao, Y., Lei, T., Xu, A., Yin, Y., Wang, J., Monteiro, A., Westhof, E., Lu, C., Tian, Z., Wang, W., Xiang, Z., Dai, F., 2022. High-resolution silkworm pan-genome provides genetic insights into artificial selection and ecological adaptation. Nat. Commun. 13, 5619. https://doi.org/10.1038/s41467-022-33366-x.

Van Gestel, M., Verheyen, K., Matthysen, E., Heylen, D., 2021. Danger on the track? Tick densities near recreation infrastructures in forests. Urban For. Urban Green. 59, 126994. https://doi.org/10.1016/j.ufug.2021.126994.

von Meijenfeldt, F.A.B., Arkhipova, K., Cambuy, D.D., Coutinho, F.H., Dutilh, B.E., 2019. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome Biol. 20, 217. https://doi.org/10.1186/s13059-019-1817-x.

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R., Zhang, F., Mansueto, L., Copetti, D., Sanciangco, M., Palis, K.C., Xu, J., Sun, C., Fu, B., Zhang, H., Gao, Y., Zhao, X., Shen, F., Cui, X., Yu, H., Li, Z., Chen, M., Detras, J., Zhou, Y., Zhang, X., Zhao, Y., Kudrna, D., Wang, C., Li, R., Jia, B., Lu, J., He, X., Dong, Z., Xu, J., Li, Y., Wang, M., Shi, J., Li, J., Zhang, D., Lee, S., Hu, W., Poliakov, A., Dubchak, I., Ulat, V.J., Borja, F.N., Mendoza, J.R., Ali, J., Li, J., Gao, Q., Niu, Y., Yue, Z., Naredo, M.E.B., Talag, J., Wang, X., Li, J., Fang, X., Yin, Y., Glaszmann, J.-C., Zhang, J., Li, J., Hamilton, R.S., Wing, R.A., Ruan, J., Zhang, G., Wei, C., Alexandrov, N., McNally, K.L., Li, Z., Leung, H., 2018. Genomic variation

in 3,010 diverse accessions of Asian cultivated rice. Nature 557, 43–49. https://doi.org/10.1038/s41586-018-0063-9.

Yan, H., Bombarely, A., Li, S., 2020. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. Bioinforma. Oxf. Engl. 36, 4269–4275. https://doi.org/10.1093/bioinformatics/btaa519.

Yang, Z., Zhang, J., Yang, S., Wang, X., Shen, Q., Sun, G., Wang, H., Zhang, W., 2021. Virome analysis of ticks in a forest region of Liaoning, China: characterization of a novel hepe-like virus sequence. Virol. J. 18, 163. https://doi.org/10.1186/s12985-021-01632-x.

Zhang, G., Zheng, D., Tian, Y., Li, S., 2019. A dataset of distribution and diversity of ticks in China. Sci. Data 6, 105. https://doi.org/10.1038/s41597-019-0115-5.

Zhao, G.-P., Wang, Y.-X., Fan, Z.-W., Ji, Y., Liu, M., Zhang, W.-H., Li, X.-L., Zhou, S.-X., Li, H., Liang, S., Liu, W., Yang, Y., Fang, L.-Q., 2021. Mapping ticks and tick-borne pathogens in China. Nat. Commun. 12, 1075. https://doi.org/10.1038/s41467-021-21375-1.