Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease

Monika Gulia-Nuss et al.

Ticks transmit more pathogens to humans and animals than any other arthropod. We describe the 2.1 Gbp nuclear genome of the tick, *Ixodes scapularis* (Say), which vectors pathogens that cause Lyme disease, human granulocytic anaplasmosis, babesiosis and other diseases. The large genome reflects accumulation of repetitive DNA, new lineages of retrotransposons, and gene architecture patterns resembling ancient metazoans rather than Pancrustacea. Annotation of scaffolds representing ~57% of the genome, reveals 20,486 protein-coding genes and expansions of gene families associated with tick-host interactions. We report insights from genome analyses into parasitic processes unique to ticks, including host ‘questing’, prolonged feeding, cuticle synthesis, blood meal concentration, novel methods of haemoglobin digestion, haem detoxification, vitellogenesis and prolonged off-host survival. We identify proteins associated with the agent of human granulocytic anaplasmosis, an emerging disease, and the encephalitis-causing Langat virus, and a population structure correlated to life-history traits and transmission of the Lyme disease agent.
Ticks (subphylum Chelicerata: suborder Ixodida) are notorious ectoparasites and vectors of human and animal pathogens, transmitting a greater diversity of infectious agents than any other group of blood-feeding arthropods. Ticks are responsible for serious physical damage to the host, including blood loss and toxicosis. Tick-borne diseases result in significant morbidity and thousands of human and animal deaths annually. The genus *Ixodes* includes multiple species of medical and veterinary importance, most notably serving as vectors of Lyme borreliosis in North America, Europe and Asia. Lyme disease is the most prevalent vector-borne disease in the northern hemisphere. In the USA, 22,014 confirmed human cases were reported in 2012 (ref. 2), with ~10-fold more infections suspected. In Europe, ~65,500 Lyme borreliosis patients are documented annually. In the USA, *Ixodes scapularis* also vectors the infectious agents that cause human babesiosis, human granulocytic anaplasmosis, tick-borne relapsing fever and Powassan encephalitis. The increased incidence and distribution of Lyme disease and other tick-borne diseases necessitates new approaches for vector control.

The first genome assembly for a tick vector of disease (*I. scapularis*) was reported in 2012 (ref. 2), with ~10-fold more infections suspected. In Europe, ~65,500 Lyme borreliosis patients are documented annually. In the USA, *Ixodes scapularis* also vectors the infectious agents that cause human babesiosis, human granulocytic anaplasmosis, tick-borne relapsing fever and Powassan encephalitis. The increased incidence and distribution of Lyme disease and other tick-borne diseases necessitates new approaches for vector control.

**Results**

The first genome assembly for a tick vector of disease. The assembly, IscaW1, comprises 570,640 contigs in 369,495 scaffolds (N₅₀ = 51,551 bp) representing 1.8 Gbp, including gaps (Table 1, Supplementary Table 2). The *ab initio* annotation of 18,385 scaffolds >10 Kbp in length and representing 1.2 Gbp (57% of the genome) predicted 20,486 protein-coding genes, and 4,439 non-coding RNA genes (Supplementary Figs 1–6 and Supplementary Table 3). Ixodid ticks typically have haploid genomes that exceed 1 Gbp (ref. 8). In contrast, the 90 Mbp genome of the two-spotted spider mite, *Tetranychus urticae*, a horticultural pest, is the smallest of any known arthropod, and contains <10% transposable elements. Repetitive DNA is estimated to comprise ~70% of the *I. scapularis* genome, reflecting an extreme case of tandem repeat and transposable element accumulation.

The *I. scapularis* genome possesses 26 acrocentric autosomes and two sex chromosomes (XX:XY). Fluorescent in situ hybridization (FISH)-based physical mapping was used to develop a karyotype and physical map (Fig. 2; Supplementary Tables 12 and 15). Mapping revealed that tandem repeat accumulation in centromeric or peri-centromeric regions, also noted in some other arthropods, is high in *I. scapularis* and comprises ~40% of genomic DNA. The low complexity tandem repeat families, ISR-1, ISR-2 and ISR-3, account for ~8% of the genome (Supplementary Text). The most abundant ISR-2 (95–99 bp; ~7% of the genome) is localized at the near-terminal heterochromatic regions of the chromosomes (Fig. 2).

The moderately repetitive fraction of the genome (~30% of genomic DNA) contains numerous copies of Class I and II transposable elements (Supplementary Tables 13 and 14 and Supplementary Text). For example, 41 well-represented elements (that is, comprising a full-length canonical and/or consensus sequence (Supplementary Figs 7 and 8)) of the long-terminal repeat (LTR) retro-transposon family, estimated to make up <1% of the genome, were identified. Thirty-seven members of the Ty3/gypsy group were identified, with the remainder being Pao/Bel-like. Two (Mag and CsRn1) of the six well-known insect Ty3/gypsy lineages were confirmed in the tick and two new clades, Squirrel and Toxo, are likely specific to the subphylum Chelicerata (Supplementary Fig. 8).

Structural characterization of elements belonging to these lineages revealed shared features that include the CCHC gag and GPF integrase domains, and two ORFs matching gag and pol. The LTRs possess the TG.CA pattern and their integration generates a duplication of 4 bp.

Non-LTR retro-transposons comprise about 6.5% of the genome. Sequence conservation and transposable element copy number suggest recent activity in the *I. scapularis* CR1, 1 and L2 clades; these elements are also abundant in birds, mammals and lizards, and the possibility of horizontal transposable element transmission warrants further investigation. The R2, RTE and LOA non-LTR retro-transposon clades found in mosquitoes and *Drosophila* were not identified in the tick. Seemingly intact *mariner* and *piggyBac* transposable elements were identified, indicating possible recent or active transposition, and 234 miniature inverted-repeat transposable elements (MITEs) were annotated. These MITEs range in copy number from 50 to 14,500 and occupy ~5% of the genome. Collectively, these findings suggest a genome permissive to high repeat accumulation.

Approximately 60% of tick genes have recognizable orthologs in other arthropods, about half of which are maintained across representative species of the major arthropod lineages (Supplementary Fig. 9). Approximately 50% of the remaining genes have homologs and ~1/5th of tick genes appear unique (*T. urticae* has a similar proportion of unique genes); these provide an important resource to understand tick-specific processes and develop highly selective interventions. Analysis of gene models and 20,901 tentative consensus sequences (the Gene Index Project; compbio.dfci.harvard.edu/cgi) compiled from 192,461 expressed sequence tags (ESTs) identified ~22% of *I. scapularis* genes as paralogs (Supplementary Note 1 and Supplementary Text). This is in line with estimates for *Homo sapiens* (15%) and the nematode, *Caenorhabditis elegans* (20%) (Supplementary Text). Complementary analyses of paralogs suggest two duplication events in *I. scapularis*, involving hundreds of genes that took place within the last 40 million years, consistent with the radiation of ticks through Europe, America and Africa. The tick mitochondrial genome retains the inferred ancestral arthropod organization as predicted by its phylogenetic position (Supplementary Fig. 10).

The genome-scale quantitative molecular species phylogeny inferred from single-copy orthologs from OrthoDB, confirms the expected position of Chelicerata as basal to crustaceans and insects (Fig. 3a). The rate of molecular evolution of *I. scapularis* genes is slightly slower than that of other representative arthropods, and considerably slower than the rapidly evolving dipterans. Quantification of shared intron positions (Fig. 3b) and lengths (Fig. 3c) among orthologs...
Figure 1 | Genes associated with the unique parasitic lifestyle of Ixodes scapularis. (a) Host detection. Ticks spend long periods off-host and locate hosts by ‘questing’ from vegetation. The Haller’s organ, located on the first pair of tarsi, is the major sensory appendage. The tick has a relatively small repertoire of visual and chemosensory genes and an expansion of detoxification genes, presumably to counteract environmental toxicants. (b) Attachment and blood feeding. The tick creates a wound cavity and injects saliva containing cement, vasodilators, pain inhibitors, anticoagulants and immune-suppressing factors to facilitate long periods of attachment and blood feeding. (c) Engorgement. Blood engorgement takes place over days to weeks and includes slow and rapid phases (dotted lines indicate increase in body volume). New cuticle is putatively synthesized to accommodate ingestion of the large (~100-fold increase in body weight) blood meal. The tick has an expansion of neuropeptide receptors to regulate diuresis and concentrate the blood meal. (d) Digestion. The processes of haemoglobin digestion in intracellular vesicles of midgut cells and haem sequestration involving specialized storage proteins are unique to ticks. Haemolyzed erythrocytes are absorbed by midgut epithelial cells by pinocytosis. Digestion is accomplished by fusion with lysosomes containing digestive enzymes (see text) and sequential breakdown of proteins (1) liberating haem and 8–11kDa peptide fragments, (2) ~5–7 kDa fragments, (3) 3–5 kDa peptides and finally (4) dipeptides and free amino acids. Amino acids are transcytosed from the digestive cells into haemolymph and haem is transported by haem-binding proteins to haemosomes for detoxification. Absorbed nutrients are converted to storage proteins (CP) throughout development or to vitellogenin in adult females for yolk provisioning of the egg just before oviposition. AMP, antimicrobial peptide; CAT, cathepsin; CP, 5–7 kDa fragments; CPs, 3–5 kDa peptides and finally (4) dipeptides and free amino acids. Amino acids are transcytosed from the digestive cells into haemolymph and haem is transported by haem-binding proteins to haemosomes for detoxification. Absorbed nutrients are converted to storage proteins (CP) throughout development or to vitellogenin in adult females for yolk provisioning of the egg just before oviposition. AMP, antimicrobial peptide; CAT, cathepsin; CP, haemlipoglyco-carrier protein; CYP450, cytochrome P450; GR, gustatory receptor; IR/iGluR, ionotropic receptor/ionotropic glutamate receptor; LAP, lysosomal aspartic protease; OBp, odorant binding protein; OR, odorant receptor; SCP, serine cysteine protease; Vg, vitellogenin.

Tick saliva is exceptionally diverse compared with that of modern mammals. The tick genome, therefore, supports an intron-rich gene architecture at the base of the arthropod radiation and more similar to that of ancestral metazoans than extant pancrustaceans.

**Ticks as parasites.** Tick mouthparts (chelicerae and barbed hypostome) attach to and create a feeding lesion in the dermis of the host (Fig. 1b). Tick saliva consists of a complex mixture of peptides and other compounds that facilitate attachment and disarm host haemostasis, inflammation and immunity, thereby enabling prolonged blood feeding. Antimicrobials in the saliva presumably prevent bacterial overgrowth within the ingested blood and/or feeding lesion. Transcriptome analyses indicate that tick saliva is exceptionally diverse compared with that of
Table 1 | Summary of the *Ixodes scapularis* genome assembly and annotation statistics.

| IsoW1 assembly statistics                  |                  |
|-------------------------------------------|------------------|
| Total number of sequence reads            | 17.4 M           |
| Estimated fold coverage of the assembly   | 3.8-fold         |
| Number of scaffolds                       | 369,495          |
| N50 scaffold length                       | 51,551 bp        |
| Number of contigs used in assembly        | 570,637          |
| N50 contig length                         | 2,942 bp         |
| Total length of combined contigs          | 1.4 Gb           |
| Total length of combined scaffolds (including gaps) | 1.8 Gb |
| Estimated genome size                     | 2.1 Gb           |

| Annotation release 1.2 statistics         |                  |
|-------------------------------------------|------------------|
| Total number of genes                     | 20,486           |
| Mean gene length                          | 10,589 bp        |
| Mean coding DNA sequence (CDS) length     | 855 bp           |

Ticks have evolved a novel mechanism for haemoglobin digestion. Haemolysis of host erythrocytes occurs in the midgut but the digestion of blood meal proteins takes place within specialized vesicles of midgut epithelial cells following internalization by pinocytosis (Fig. 1d). Haemoglobin digestion occurs via a cascade of proteolytic enzymes resulting in dipeptides and free amino acids that are transcytosed into the haemolymph (Supplementary Text and Supplementary Table 21). Orthologs of *Ixodes ricinus* haemoglobinolytic enzymes were identified in the *I. scapularis* genome that contains multiple genes for cathepsin D (three genes), cathepsin L (three genes), and serine carboxypeptidase (four genes), suggesting the relative importance of these enzymes in haemoglobin digestion. Haemoglobinolytic enzymes have also been identified in other tick species suggesting that this mode of haemoglobin digestion is widespread throughout the Ixodida. Liberated haem is transported from the digestive vesicles by transport proteins to haemosomes, unique storage vesicles where haem is detoxified by formation of haematin-like aggregates. Thus, haemoglobinolyis in ticks is similar to that in endoparasitic flatworms and nematodes. However, tick-specific intracellular digestion in midgut epithelial vesicles and haem detoxification in specialized haemosomes could offer novel acaricide targets (Supplementary Text and Supplementary Table 21).

Haem is associated with multiple essential functions as it complexes with proteins that perform oxygen transport and sensing, enzyme catalysis and electron transfer. However, ticks are incapable of *de novo* haem synthesis, and it has been proposed that they rely on haem recovery from the diet. The identification of orthologous genes in *I. scapularis* for the
from a tick to date, and two Vg genes were identified in the *I. scapularis* genome (Supplementary Fig. 16 and Supplementary Table 22).

The genome contains orthologs for at least 39 invertebrate neuropeptide genes (Supplementary Tables 25–28), including peptides that regulate ecdisis, cuticle synthesis, hardening and tanning. Orthologs involved in insect moulting34, that is, corazonin, eclosion hormone, cardioactive peptide and buriscon \( \alpha \) and \( \beta \), were identified (Fig. 4). Additional novel putative neuropeptide genes were identified based on the presence of tandem repeats in conserved C-terminal sequences, including the canonical sequences for amidation and dibasic (or monobasic) cleavage signals (Supplementary Table 25). ESTs matching corazonin, eclosion hormone and buriscon \( \alpha \) and \( \beta \) were found in the synganglion transcriptome of adult *Dermacenior variabilis*35, which do not moult, suggesting previously unrecognized roles for these neuropeptide hormones. Companion analyses36 identified major differences in gene expression between *I. scapularis* and the soft tick, *Ornithodoros turicata* (Argasidae) in response to feeding that may explain how synganglion neuropeptides regulate different life styles of the two tick families. The identification of orthologs of neuropeptides known to regulate insect moulting provides a much needed starting point to understand the regulation of development in ticks and in the modification of cuticle to accommodate the approximately 100-fold increase in size that occurs during blood feeding (Fig. 4).

In ticks, over-hydration from large blood meals is counterbalanced by hormonally controlled salivary secretion into the host, presumably regulated by neuropeptides and their G-protein-coupled receptors (GPCRs) (Fig. 1c). The homologs of many insect neuropeptides, protein hormones, biogenic amines and associated GPCRs37 (Supplementary Tables 25–28) that steer processes such as diuresis, behaviour, reproduction and development38, were identified in *I. scapularis*. Some of the neuropeptide genes identified encode multiple neuropeptides. Of note is the extreme number of copies (19) of the kinin gene, which ranges from one to eight in other arthropods38 (Supplementary Table 28), suggesting that high peptide copy number is also needed for effective diuresis. In accordance, four kinin GPCRs are present (Supplementary Table 28). The tick has 20 GPCRs for five biogenic amines, a number similar to that for all other sequenced arthropods37, suggesting an early evolutionary origin of these molecules and a core set of highly conserved arthropod signalling molecules. Typically in insects, each neuropeptide interacts with one, or at most two, GPCRs37. Remarkably, the numbers of some neuropeptide GPCRs have expanded significantly (up to 10-fold) in *I. scapularis* (Supplementary Tables 26 and 28). This includes the GPCRs for AKH/corazonin-related peptide, allatostatin-A, diuretic hormones (calcitonin- and CRF-like), inotocin, kinin, pigmentation-dispersing-factor, sulfakinin, and tachykinin (Supplementary Table 28)37. In insects, these GPCRs are involved in regulating meal size (kinin), satiety (sulfakinin) and diuresis (kinin, tachykinin and calcitonin-like diuretic hormone)38. In ticks, the increased efficacy and fine regulation of diuresis may be accomplished through an increased repertoire of diuretic GPCRs rather than via corresponding neuropeptides, emphasizing their potential as targets for tick control.

Blood feeding is essential for reproduction in adult female ticks (Fig. 4). In lower insects, reproduction is largely regulated by juvenile hormone III. Biochemical evidence suggests that ticks do not synthesize juvenile hormone III and instead employ ecdy steroids to initiate vitellogenesis (Fig. 4, reviewed in39). In insects, the final hydroxylations for the synthesis of ecdysteroids are performed sequentially by cytochrome P450s (CYP450s)
encoded by the *Halloween* genes (Supplementary Fig. 17 and Supplementary Table 19). Genes for all steroidogenic CYP450s except for *phantom* were identified in the *I. scapularis* genome and putative gene duplications were identified for *disembodied* and the *spook/spookier* clades, suggesting conservation of ecdysteroid regulated processes between ticks and insects. Genes for seven of the nine enzymes in the insect mevalonate–farnesal pathway that produces the juvenile hormone precursor, farnesylpyrophosphate (farnesyl-PP), were identified in the tick genome (Supplementary Fig. 18 and Supplementary Table 18). There are five insect enzymes involved in the conversion of farnesyl-PP to juvenile hormone III. Only the gene for farnesol oxidase in the juvenile hormone branch was found in the *I. scapularis* genome (Supplementary Table 18) and is transcribed in the synganglion of *I. scapularis* and *D. variabilis*. The tick genome reveals a striking expansion of the methyl transferase family (44 genes) and EST data indicate that at least 26 of these are transcribed (Supplementary Fig. 19). However, the *I. scapularis* methyl

Figure 4 | Model of neuroendocrine processes controlling mating and egg production in *Ixodes scapularis*. (1) Mating takes place off or on the host (before or during blood feeding), but is required for rapid blood feeding. The male attaches to the genital pore of the female via its mouthparts (evidence suggests the potential involvement of female specific cuticular lipids and a non-volatile mounting pheromone in *I. scapularis*), then transfers sperm and gonadotropins (unidentified at present), among other seminal components, including the spermatophore. (2) Gonadotropins initiate the synganglion to release EDTH, stimulate rapid engorgement, initiate synthesis of neuropeptides which in insects regulate molting and synthesis of new cuticle (tick functions unknown), and release of allatostatins and allatotropins (which may stimulate or inhibit the mevalonate-farnesal pathway). (3) EDTH initiates production of ecdysteroids by the epidermis, (4) High ecdysteroid titres activate transcription factors for VgR in the ovaries, are stored in developing eggs and, as 20-E, activates transcription factors for Vg in the fat body and specialized cells of the midgut. (5) Vg is taken up via VgR-receptor mediated endocytosis by developing oocytes and incorporated into the yolk as vitellin, and (6) The female produces a single batch of ~3,000 mature eggs from the genital pore that are passed forward to the mouthparts for coating with wax released from the Gene’s organ. Biochemical and genomic evidence suggests that *I. scapularis* do not make JH III although the genes for the preceding mevalonate and parts of the farnesal pathway were identified. Dashed lines indicate proposed pathways and factors. 20-E, 20-hydroxyecdysone; CAP, cardioactive peptide; EDTH, hypothesized epidermal trophic hormone; Vg, vitellogenin (yolk protein in haemolymph before egg uptake); VgR, vitellogenin receptor.
transferases studied so far lack the juvenile hormone binding motif. An ortholog of the insect cytochrome P450 (CYP15A1) that adds the epoxide to methyl farnesoate to produce juvenile hormone III was not found in either the tick genome (Supplementary Table 18) or synganglion transcriptomes. The neuropeptides, allatostatin and allatotropin, which perform a variety of functions in insects, including the regulation of juvenile hormone biosynthesis, were also identified in the tick (Fig. 4). Important questions remain as to the role of the mevalonate-farnesal pathway in tick reproduction and development. In a complementary study, transcripts for genes in the mevalonate-farnesal pathway were identified from the synganglion of two hard and one soft tick species. The I. scapularis genome reflects a parasitic lifestyle requiring detoxification of multiple xenobiotic factors (Fig. 1a). We identified a record 206 CYP450 (Supplementary Table 23) and 75 carboxylesterase/cholinesterase-like genes, including five putative acetylcholinesterase genes (Supplementary Table 24). CYPs are haem-containing enzymes that catalyse biological oxidation reactions, many of which detoxify xenobiotics, including acaricides. In contrast, the body louse, Pediculus humanus, also an obligate blood-feeding ectoparasite, has 36 CYPs, the fewest known in an animal, while the plant feeding mite, T. urticae has 81 (ref. 9). Carboxylesterases are also associated with metabolic detoxification in animals. While the function of these enzymes is not known, the abundance of these genes in I. scapularis may reflect the need to detoxify large blood meals from diverse hosts and toxicants encountered during off-host stages.

As a parasite that lives largely off-host, I. scapularis has developed unique mechanisms for host detection that are reflected in the genome (Fig. 1a). The sensory system in ticks includes setiform sensilla for chemo-, mechano-, thermo- and reflected in the genome (Fig. 1a). The sensory system in ticks developed unique mechanisms for host detection that are

The tick possesses a small repertoire of photon-sensitive receptors compared with most insects. Genes for three opsins GPCRs were identified (Fig. 1a, Supplementary Table 26) and include orthologs of the insect putative long-wavelength sensitive ‘visual’ opsins, the honey bee ‘non-visual’ pteropsin likely involved in extracellular light detection and regulation of circadian rhythm, as well as the D. melanogaster Rh7 opsin. Orthologs of the insect UV and short wavelength receptors were not identified. This indicates a reduced visual system as compared with other blood-feeding arthropods (Supplementary Text) that rely heavily on visual processes during flight for location of mates, hosts and oviposition sites. During host detection, olfactory, mechano- and thermoreception may offset limited visual acuity and wavelength detection in the tick.

Ticks as vectors of pathogens and parasites. Ticks are biological vectors of viruses, bacteria and protozoa that are typically acquired via the blood meal and transmitted through saliva during feeding (Fig. 5). The tick immune system has several mechanisms to fend off pathogen invasion. Most components of the Toll, IMD (Immunodeficiency), JAK-STAT (Janus Kinase/Signal Transducers and Activators of Transcription) immune pathways and the RNA interference-antiviral signalling pathways were identified in the tick genome (Supplementary Figs 22 and 23 and Supplementary Table 17). The repertoire of immunity-related genes also includes aikirins, antimicrobial peptides, caspases, defensins, oxidases, the fibrinogen-related protein family of ixodid, lysozymes, thi-o-ester containing proteins and peptidoglycan-recognition proteins (Supplementary Table 17).

Multiple infection factors facilitate transmission of the Lyme disease pathogen, Borrelia burgdorferi (Fig. 5). These include the tick salivary gland proteins Salp15, Salp20, Salp25D, tick salivary lectin pathway inhibitor and tick histamine-release factor, as well as the tick receptor for OsPA and tick protein tre31, and the Borrelia lipoprotein BBE31 (ref. 51). Increasingly, research is focused on interactions with Anaplasma phagocytophilum (Rickettsiales: Anaplasmataceae), the causative agent of human granulocytic anaplasmosis prevalent in the USA and Europe. The I. scapularis proteins P11, SALP16, x1, 3-fucosyltransferases and the X-linked inhibitor of apoptosis E3 ubiquitin ligase are required for A. phagocytophilum infection and transmission, and modification of the tick cytoskeleton by A. phagocytophilum increases infection. To establish infection, A. phagocytophilum inhibits apoptosis in midgut and salivary gland cells through the JAK/STAT and intrinsic pathways. In response, the extrinsic apoptosis pathway is induced in tick salivary glands. All known components of these pathways were identified in the tick with the exception of the Perforin ortholog (Supplementary Table 17). Systems biology analyses revealed that the generalized responses
of tick cells to *A. phagocytophilum* infection include changes in protein processing in the endoplasmic reticulum and glucose metabolism. Protein misfolding is increased in infected tick cells, a possible strategy by which *A. phagocytophilum* evades the cellular response to infection. The subsequent activation of protein targeting and degradation, reduces endoplasmic reticulum stress and prevents cell apoptosis, and may also benefit the pathogen through provision of raw materials critical for an obligatory intracellular parasite with reduced biosynthetic and metabolic capacity. In addition, *A. phagocytophilum* can induce an increase in expression of antifreeze glycoproteins, enhancing *I. scapularis* survival in cold temperatures, and downregulate Porin expression to inhibit apoptosis, increasing tick colonization and parasite derived factors known to facilitate these processes are shown. IMD, Immunodeficiency; JAK-STAT, Janus Kinase/Signal Transducers and Activators of Transcription; OspA, *Borrelia* outer surface protein A; Salp15/16, salivary gland protein 15/16; tHRF, tick histamine-release factor; TROPSA, tick receptor for OspA; TSLP1, tick salivary lectin pathway inhibitor.

We used quantitative proteomics to further characterize tick–*Anaplasma* interactions, and identify differential protein expression in an *I. scapularis* ISE6 cell line in response to infection; 735 unique peptides assigned to 424 different *I. scapularis* proteins, were identified (Supplementary Tables 32–35). In total, 83 proteins were differentially represented (11–17% infected cells at 3 days post-inoculation). Most were also represented as infection advanced when the number of under- and over-represented proteins increased to 50 and 31, respectively (56–61% infected cells; 10 days post-inoculation). Analysis of protein ontology demonstrated differences between under- and over-represented proteins in both early and late infections for cell growth (adducin, spectrin and β-tubulin) and transport (Na+/K+ ATPase, voltage-dependent anion-selective channel or mitochondrial porin and fatty acid-binding protein; Supplementary Tables 32–34).

The genome of a *Rickettsia* (Alphaproteobacteria: *Rickettsiales*) species, *Rickettsia* endosymbiont of *Ixodes scapularis* (REIS), was assembled from both bacterial artificial chromosome clones and recruited whole-genome shotgun reads (available at GenBank, NZ_ACLC00000000). Phylogenomics analysis of the REIS genome, which comprises a single 1.82 Mbp chromosome and four plasmids, indicates a novel non-pathogenic species that is ancestral to all Spotted Fever Group *Rickettsia* species, providing a valuable resource for understanding the evolution of symbiosis versus pathogenicity.

Much less is known about the molecular mechanisms involved with viral interactions in ticks. Research suggests the RNA interference pathway provides an important defense against virus infection in tick cells, with a significant expansion of *Ago* genes in comparison with insects. In a companion proteomics study of
the *I. scapularis* ISE6 cell line following infection with the Langat virus\(^3\), 266 differentially expressed tick proteins were identified. Functional analyses suggest perturbations in transcription, translation and protein processing, carbohydrate and amino acid metabolism, transport and catabolism responses. The majority of differentially expressed proteins were downregulated, similar to the proteomics profile described above. Interestingly, 121 differentially expressed proteins lacked homology to known orthologs, suggesting these may be unique to *I. scapularis*.

**Population structure of *Ixodes scapularis* in North America.** The restriction-site-associated DNA sequencing (RADseq) technique was employed for genome-wide discovery of single-nucleotide polymorphisms (SNPs) and examination of genetic diversity within and among eight *I. scapularis* populations from the north-east, mid-west and south-east regions of the USA and the Wikel reference colony. F-statistics were used to assess genetic distance as evidence of selection. *F*\(_{ST}\) values (range 0.003–0.012; Supplementary Table 36) suggest random mating or low levels of inbreeding among members comprising each population. Further supporting this hypothesis, among all populations, the average observed heterozygosity (Ho) per variable SNP was comparable (range 0.013–0.016) to expected heterozygosity (He) (range 0.013–0.018) and the nucleotide diversity among SNP loci (\(\pi\)) (range 0.015–0.019) was comparable among samples. *F*\(_{ST}\) values (range 0.03–0.16; Supplementary Table 37), support a single species classification for *I. scapularis* across North America as previously reported\(^6\). Low/moderate genetic variation (*F*\(_{ST}\) = 0.03–0.06) was observed among northern tick populations from Indiana, Maine, Massachusetts, New Hampshire and Wisconsin, and moderate variation (*F*\(_{ST}\) = 0.07–0.09) among southern populations from Florida, North Carolina and Virginia. *F*\(_{ST}\) analyses revealed signatures of north–south structure in *I. scapularis* populations. Moderate-to-high genetic variation was observed between northern versus southern populations (*F*\(_{ST}\) = 0.10–0.15). Interestingly, low genetic variation (*F*\(_{ST}\) = 0.03–0.06) was observed between populations from the mid-west (Indiana and Wisconsin) versus the north-east (Maine, Massachusetts and New Hampshire), two areas associated with a high prevalence of human Lyme disease cases. As expected, moderate-to-high genetic variation was observed between the reference Wikel colony and field populations (*F*\(_{ST}\) = 0.07–0.16).

The population structure of *I. scapularis* was separately analysed using a subset of representative SNPs. Membership probabilities, interpreted as proximities of individuals belonging to each cluster, revealed five clades (Fig. 6), with clear separation of the Wikel colony from field populations. Clustering of Indiana and New Hampshire, and Massachusetts, Maine and Wisconsin populations, indicates significant shared alleles, while the Virginia, Florida and North Carolina populations may share a small number of alleles. Interestingly, the population structure suggests a genetic component associated with differences in the natural history of northern and southern *I. scapularis* and a correlation to the prevalence of human Lyme disease cases. The incidence of Lyme disease is greatest in the upper mid-west and north-east where *I. scapularis* populations feed predominantly on deer as adults and complete the life cycle over 2 years. In contrast,
Genome-based interventions to control tick-borne disease.

Prevaling methods of tick control rely heavily on the use of repellents and acaricides. Resistance to currently applied pesticides that disrupt neural signalling and tick development has prompted the search for novel targets. GPCRs represent a source of candidate targets for development of novel interventions. High-throughput target-based approaches have been employed to discover new mode-of-action chemicals that selectively inhibit the *I. scapularis* dopamine receptors. The ligand-gated ion channels (LGICs) offer another rich source of targets. iGluRs play a major role in neurotransmission and chemosensory signalling within arthropods. Twenty-nine putative *iGluR* genes and 32 putative cys-loop receptors were identified in the genome (Fig. 7, Supplementary Table 31). Among the *iGluR* genes, 14 encode members of the three principal subclasses of synaptic *iGluRs* (AMPA, Kainate and NMDA; Supplementary Fig. 21 and Supplementary Tables 30 and 31), while the remaining 15 more divergent sequences likely belong to the chemosensory ionotropic receptor subfamily (see above). The cys-loop LGIC family also contains six candidate glutamate-gated *Cl*− channels (*Glucls*), 12 nicotinic acetylcholine receptor subunits, and four GABA-gated chloride channels. One histamine-gated *Cl*− channel and one pH-gated *Cl*− channel gene were also identified. Both the *iGluRs* and cys-loop LGIC families contain tick-specific genes with no apparent insect ortholog. This striking divergence may contribute to the apparent ineffectiveness of some insecticides on acaridial targets.

Classifying LGIC candidates by functional expression is underway and an example is shown for a GluCl (Fig. 7, Supplementary Fig. 25). Selective targeting of tick LGICs and GPCRs may offer routes to new, safe and effective acaricides.

**Discussion**

The genome sequence of *I. scapularis*, the first for a medically important chelicerate, offers insights into the molecular processes that underpin the remarkable parasitic lifestyle of the tick and its success as a vector of multiple disease-causing organisms. Foundational studies of genome organization and population structure will advance research to determine the genetic basis of tick phenotypes, and efforts are ongoing to discover novel chemosensory characters that selectively disrupt molecular targets mined from the genome. This study is a pioneering project for genome research on ticks and mites of public health and veterinary importance, with efforts proposed to expand genomic resources across this phyletic group. In 2011, the National Institutes of Health approved the sequencing of additional species of hard ticks, including *European and Asian Ixodes* species, the soft tick *Orientalodoros moubata* (Family Argasidae) and the *Leptotrombidium* mite vector of scrub typhus (Superorder Acrariformes) (Supplementary Table 38). The *I. scapularis* genome offers a roadmap for research on tick-host–pathogen interactions to achieve the goals of the One Health Initiative and improve human, animal and ecosystem health on a global scale.

**Methods**

**Genomic sequencing, assembly and annotation.** The genome of *I. scapularis* Wikel strain was sequenced in a joint effort by the Broad Institute and the JCVI and funded by the National Institute of Allergy and Infectious Diseases, National Institutes of Health. The *I. scapularis* Wikel strain (Quinnipiac University, Hamden, CT) genome was sequenced to approximately 3.8-fold coverage using Sanger sequencing and assembled using the Celera Assembler configured to accommodate high repeat content within the genome and heterozygosity in the donor population (Supplementary Table 1). The assembly and raw reads are available at GenBank under the project accession ABJ010000000, consisting of contig accessions ABJ010000000-ABJ011141594 and VectorBase as IscaW1, 3 May 2012. The annotation of the *I. scapularis* genome was performed via a joint effort between the I.CVI and Vectorbase. The genome sequence was also submitted to Vectorbase (IscaW1.4) is available at VectorBase (https://www.vectorbase.org/) and GenBank (accession ID: ABJ010000000). Forty-five bacterial artificial chromosome clones, ~183,834 ESTs and 45 microRNAs were also sequenced and annotated (Supplementary Figs 4–6 and Supplementary Tables 4–6).

**Proteomics of Ixodes-Anaplasma interactions.** The *I. scapularis* IS66 cells were inoculated with *A. phagocytophilum* (human NY18 isolate) or left uninfected. Uninfected and infected cultures (n = 5 independent cultures each) were sampled at early infection (11–17% infected cells (Avg ± S.D., 13 ± 2)) and late infection (56–61% infected cells (Avg ± S.D., 58 ± 2)) and used for proteomics. Protein extracts from the four experimental conditions, control uninfected early, infected early, control uninfected late and infected late (100 μg each) were gel-concentrated, digested overnight at 37 °C with 60 ng μl−1 trypsin (Promega, Madison, WI, USA) and the resulting tryptic peptides from each proteome were extracted and iTRAQ labelled for analysis. The samples were fractionated by gel-filtration chromatography and each fraction analysed by liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) using a Surveyor LC system coupled to a linear ion trap mass spectrometer model LTQ (Thermo Finnigan, San Jose, CA, USA) and protein identification was carried out using SEQUEST algorithm (Bioworks 3.2 package, Thermo Finnigan). Amino-terminal (Methionine oxidation) and fixed modifications (Cysteine carboxamidomethylation, Lysine and N-terminal modification of +144.1020 Da). The MS/MS raw files were searched against the alphaproteobacteria combined with the arachnida Swissprot database (Uniprot release 15.5, 7 July 2009) supplemented with porcine trypsin and human keratins. This joint database contains 638,408 protein sequences. False discovery rate of identification was controlled by searching the same collections of MS/MS spectra against inverted databases constructed from the same target databases. The alphaproteobacteria Swissprot database was used to identify *Anaplasma* and discard possible symbiotic bacterial sequences from further analyses.

**Ixodes scapularis genetic diversity and population structure.** 74 RADseq libraries were produced from female *I. scapularis* representing nine ‘populations’ from the states of Florida, Indiana, Maine, Massachusetts, North Carolina, New Hampshire, Virginia and Wisconsin and the Wikel reference colony. RADseq libraries were constructed using 1 μg genomic DNA from individual ticks, separately digested with the SfiI restriction enzyme. Adaptor ligated libraries were pooled and sequenced at the Purdue Genomics Core Facility on the Illumina HiSeq 2500 in Rapid run mode. Further analysis was performed by the Bioinformatics Core at Purdue University. Illumina reads were corrected for restriction site, clustered and de-multiplexed (sorted by barcode) using the ‘process_radtags.pl’ script of STACKS. For SNP identification, reads from each sample were separately aligned to the IscaW1 assembly using the end-to-end mode and default parameters of Bowtie2 v 2.1.0. Genetic diversity within and between *I. scapularis* populations was calculated using 745,760 SNPs across 35,460 polymorphic loci. F-statistics were used to assess genetic distance or differentiation as evidence of selection where Fst is the inbreeding coefficient of an individual (I) relative to the subpopulation (S) and FST is the difference in allele frequency between subpopulations (S) compared with the total population (T). The population structure of *I. scapularis* across North America was separately analysed using a subset of 34,693 representative SNPs (1 SNP per polymorphic locus). The ‘population’ step from STACKS was used to analyse genetic diversity and fastStructure (beta release) was used to analyse population structure. Detailed methods are available in Supplementary Text. All variation data are available at NCBI SRA (SRP065406), VectorBase and via BioMart: http://biomart.vectorbase.org.

**Functional expression of tick LGICs.** Expression studies were performed on mature oocytes extracted from anaesthetised female *Xenopus laevis* and *A. phagocytophilum* (human NY18 isolate). Briefly, complementary RNA encoding IscaGluCl1 was injected at 1 mg ml−1 using a Drummond Nanoject injector into oocytes that had been treated for 20–40 min in a 2 mg ml−1 solution of collagenase type 1A (Sigma UK) in calcium-free saline. Following 3–5 days incubation at 18 °C in saline supplemented with penicillin (100 units per ml) and streptomycin (100 μg ml−1), gentamycin (50 μg ml−1) and 2.5 mM sodium pyruvate, oocytes were secured individually in a Perspex chamber (~90 μl) and perfused continually in saline at 5 ml min−1. They were impaled by two glass microelectrodes filled with 3 M KCl (resistance 1–5 MOhm in saline), with which the oocytes were voltage clamped at −100 mV using an Axoclamp 2A amplifier. Solutions were applied in the perfusing saline. The saline consisted of (in mM): NaCl 100, KCl 2, CaCl2, 1.8, MgCl2 1, HEPES 5, adjusted to pH 7.6 with 10 M NaOH.
References

1. Lindgren, E. & Jaenson, T. G. T. Lyme borreliosis in Europe: influences of climate and climate change, epidemiology, ecology and adaptation measures (World Health Organization, 2006). http://www.euro.who.int/__data/assets/pdf_file/0006/98219/E89522.pdf.

2. MMWR. Morbidity and Mortality Weekly Report. Report No. 61, 1-124 (Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA, 2014) http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6153a1.htm.

3. Walker, D. T. Tick-transmitted infectious diseases in the United States. Annu. Rev. Public Health 19, 237–269 (1998).

4. Rizzoli, A. et al. Lyme borreliosis in Europe. Eurosurveillance 16: pii 19906 (2011).

5. Worner, G. P. et al. The clinical assessment, treatment, and prevention of Lyme disease, human granulocytic anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious Diseases Society of America. Clin. Infect. Dis. 43, 1089–1134 (2006).

6. Rehm, P., Meusemann, K., Borner, J., Misof, B. & Burmester, T. Phylogenetic position of Myriapoda revealed by 454 transcriptome sequencing. Mol. Phylogenet. Evol. 77, 25–33 (2014) Erratum 80, 340 (2014).

7. Kaufman, W. R. In Integument and ecdysis. Ch. 5 Volume I (eds Sonenshine, D. E. & Roe, R. M.) pp 416–448 (Oxford University Press, 2014).

8. Geraci, N. S., Johnston, J. S., Robinson, J. P., Wikel, S. K. & Hill, C. A. Variation in genome size of arsaíd and ixodid ticks. Insect Biochem. Mol. Biol. 37, 399–408 (2007).

9. Grbic, M. et al. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature 479, 487–492 (2011).

10. Ullmann, A. J., Lima, C. M., Guererro, F. D., Piesman, J. & Black, IV W. C. Adaptations against heme toxicity in blood-feeding arthropods. Insect Biochem. Mol. Biol. 40, 79–90 (2010).

11. Meyer, J. M., Kurtti, T. J., Van Zee, J. P. & Hill, C. A. Genome organization genes. Mol. Biol. Evol. 23, D283–D288 (2011).

12. Shao, R. & Barker, S. C. Mitochondrial genomes of parasitic arthropods: a novel salivary antimicrobial peptide from the tick, Ixodes scapularis. J. Proteomics 82, 86–90 (2013).

13. Ugarkovic, D., Podnar, M. & Plohl, M. Satellite DNA of the red flour beetle, Tribolium castaneum-comparative study of satellites from the genus Tribolium. Mol. Biol. Evol. 13, 1059–1066 (1996).

14. Tubio, J. M., Naveira, H. & Costas, J. Structural and evolutionary analyses of the ty3/gypsy group of LTR retrotransposons in the genome of Anopheles gambiae. Mol. Biol. Evol. 22, 29–39 (2005).

15. Li, W.-H., Gu, Z., Wang, H. & Nekrutenko, A. Evolutionary analyses of the human genome. Genome Res. 19, 409, 847–850 (2009).

16. Lynch, M. & Conery, J. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155 (2000).

17. Van Zee, J. P. et al. Paralog analyses reveal gene duplication events and genes under positive selection in Ixodes scapularis and other ixodid ticks BMC. Genomics. doi:10.1186/1427-4157-9-231 (2008).

18. Shao, R. & Barker, S. C. Mitochondrial genomes of parasitic arthropods: implications for studies of population genetics and evolution. Parasitology 134, 153–167 (2007).

19. Waterhouse, R. M., Zdobnov, E. M., Tegenfeldt, F., Li, J. & Kriventseva, E. V. OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. Nucleic Acids Res. 39, D283–D288 (2011).

20. Csu,ros, M., Rogozin, I. B. & Koonin, E. V. A detailed history of intron-rich microphil larvae and females. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 157, 326–335 (2010).

21. Pichu, S., Ribeiro, J. M. C. & Mather, T. N. Purification and characterization of a novel salivary antimicrobial peptide from the tick, Ixodes scapularis. Biochem. Biophys. Res. Commun. 390, 511–515 (2009).

22. Chartrand, J., Copley, R. R., Doerks, T., Ponting, C. P. & Bork, P. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. 28, 231–234 (2000).

23. Sandmanner, D., Sances, G. C., Slovak, M. & Nattall, P. A high affinity serotonin- and histamine-binding lipocalin from tick saliva. Insect Mol. Biol. 11, 79–86 (2002).

24. Frankischetti, I. M., Mather, T. N. & Ribeiro, J. M. Cloning of a salivary gland metalloproteinase and characterization of gelatinase and fibrinogen/lytic activities in the saliva of the Lyme disease tick vector Ixodes scapularis. Biochem. Biophys. Res. Commun. 305, 869–875 (2003).

25. Horn, M. et al. Hemoglobin digestion in blood-feeding ticks: mapping a multiprepitode pathway by functional proteomics. Chem. Biol. 16, 1053–1063 (2009).

26. Yamazaki, K. et al. Hemoglobin digestion activity of a cysteine protease from the ixodid tick Haemaphysalis longicornis. Parasitol. Int. 58, 232–237 (2009).

27. Estrella, A. B., Seixas, A., Teixeira Vde, O., Pinto, A. F. & Termignoni, C. Vitellin- and hemoglobin-digesting enzymes in Rhipicephalus (Boophilus)
56. Aylón, N. et al. Systems biology of tissue-specific response to Anaplasma phagocytophilum reveals differentiated apoptosis in the tick vector Ixodes scapularis. PLoS Negl. Trop. Dis. 11, e0015210 (2015).
57. Villar, M. et al. Integrated metabolomics, transcriptomics and proteomics identifies metabolic pathways affected by Anaplasma phagocytophilum infection in tick cells. Mol. Cell. Proteomics 14, 3154–3172 (2015).
58. Huang, H., Wang, X., Kikuchi, T., Kumagai, Y. & Rikihisa, Y. Porin activity of Anaplasma phagocytophilum outer membrane fraction and purified P44. J. Bacteriol. 189, 1998–2006 (2007).
59. Neelakanta, G., Sultana, H., Fish, D., Anderson, J. F. & Fikrig, E. Anaplasma phagocytophilum induces Ixodes scapularis ticks to express an antifreeze glycoprotein gene that enhances their survival in the cold. J. Clin. Invest. 120, 3179–3190 (2010).
60. Busby, A. T. et al. Expression of heat-shock proteins and subeolin affects stress responses, Anaplasma phagocytophilum infection and questing behavior in the tick, Ixodes scapularis. Med. Vet. Entomol. 26, 92–102 (2012).
61. Gillespie, J. J. et al. A Rickettsia genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. J. Bacteriol. 194, 376–394 (2012).
62. Schnettler, E. et al. Induction and suppression of tick cell antiviral RNAi responses by tick-borne flaviviruses. Nucleic Acids Res. 42, 9436–9446 (2014).
63. Grabowski, J. M. et al. Changes in the proteome of Langat-infected Ixodes scapularis ISE6 cells: metabolic pathways associated with flavivirus infection PLoS Negl. Trop. Dis. doi:10.1371/journal.pntd.0004180.
64. Oliver, Jr. J. H. et al. Conspecificity of the ticks Ixodes scapularis and I. dammini (Acari: Ixodidae). J. Med. Entomol. 30, 54–63 (1993).
65. Diuk-Wasser, M. A. et al. Spatiotemporal patterns of host-seeking Ixodes scapularis nymphs (Acari: Ixodidae) in the United States. J. Med. Entomol. 43, 166–176 (2006).
66. Nuss, A. B. et al. Dopamine receptor antagonists as new mode-of-action insecticide leads for control of Aedes and Culex mosquitoes to resist malaria. PLoS Negl. Trop. Dis. 9, e0003515 (2015).
67. Lees, K. & Bowman, A. S. Tick neurobiology: recent advances and the post-genomic era. Invert. Neurosci. 7, 183–198 (2007).
68. Hill, C. A. Genome analysis of the major tick and mite vectors of human pathogens. NIH-NIAID-NHGRI Task Force report. Preprint at http://www.genome.gov/Pages/Research/DEREPathogensVectors/Tick_Mite_Genomes_Cluster_White_Paper_12Jan2011.pdf (2010).
69. King, L. J. et al. Executive summary of the AVMA One Health Initiative Task Force report. J. Am. Vet. Med. Assoc. 233, 259–261 (2008).

Acknowledgements

This project has been funded in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services (NIAID, NIH, DHHS) under contract numbers N01-AI-08071, HHSN27220090007C, HHSN266200400001C and 5R01GM77117-5. Its contents are those solely the responsibility of the authors and do not represent official views of the NIH. Additional grants and contracts supporting work described in this manuscript were from the NIH-NIAID (HHSN266200400005C and HHSN272200900039C) to F.H.C. and a subcontract under HHSN272200900039C to C.A.H. and J.F., the Australian Research Council Discovery Project (DP120100240) to S.C.B. and R.S., the Ministerio de Ciencia e Innovacion of Spain (BFU2007-6292; BFU2010-15484) to J.R., BIO2009-07990 and BIO2012-37926 to J.V. NIH-R01AI096268 and NSF IOS-0949194 to R.M.R., the Xunta de Galicia de Spain (10PXB19816057PR) to J.M.C.T. and M.T., BFU2011–23896 and EU FP7 ANTIGONE (278976) to J.F., the USDA-NRI/CREES (2008-35302-18820) and Texas AgriLife Research Vector Biology grant to V.P.V. and European Research Council Starting Independent Researcher Grant (205202) to R.B., J.M.R. was supported by the intramural program of the NIAID, R.M.W. by a Marie Curie International Outgoing Fellowship PIOF-GA-2011–303312, E.M.Z. by Swiss National Science Foundation awards 31003A-125350 and 3100A3-14936, J.M.G. by an NIH-NCATS award TL1 TR001062 and NSF Graduate Research Fellowship (DGE 133468), V.C. by a Boehringer Ingelheim Ph.D. Fellowship, F.G.V. by a Fundación para la Ciencia e a Tecnología, Portugal fellowship (SFRH/BD/22360/2005), C.J.P. and F.H. by the Lundbeck Foundation (Denmark), and J.G. by NIH awards HHSN272200900040C, R01AI07828 and R01AI043006. Support from the Broad Genomics Platform is gratefully acknowledged.

Author contributions

C.A.H., V.M.N. and S.K.W. wrote the genome sequencing proposal. C.A.H. and S.K.W. generated DNA and RNA for sequencing. E.C.L., Y.V.Q., C.A.H., A.B.N, J.M.M., D.B.S., D.E.S., R.M.R., J.R. and R.M.W. coordinated manuscript preparation. All other authors are members of the Ixodes scapularis genome sequencing consortium and contributed annotation, analyses or data to the genome project.

Additional information

Accession codes: The data reported in this paper are archived at GenBank under the project accession ABJ010000000, consisting of contig accessions ABJ010000001-ABJ011141594, and at VectorBase (IscaW1, 3 May 2012). The genome annotation release (IscaW1.4) is available at GenBank (accession ID: ABJ010000000) and VectorBase (https://www.vectorbase.org/) and RADseq data have been deposited in the NCBI Sequence Read Archive (SRA) under accession code SRP065406.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Gulia-Nuss, M. et al. Genomic insights into the Ixodes scapularis tick vector of Lyme disease. Nat. Commun. 7:10507. doi:10.1038/ncomms10507 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/
Adam R. Wespiser19, Yunlong Yang27, Jiwei Zhu3, Peter Arensburger27, Patricia V. Pietrantonio27, Stephen C. Barker38, Renfu Shao39, Evgeny M. Zdobnov4,5, Frank Hauser40, Cornelis J.P. Grimmelikhuijzen40, Yoonseong Park28, Julio Rozas17, Richard Benton21, Joao H.F. Pedra33,37, David R. Nelson41, Maria F. Unger16, Jose M.C. Tubio42,43, Zhijian Tu25, Hugh M. Robertson44, Martin Shumway14,†, Granger Sutton14, Jennifer R. Wortman14,†, Daniel Lawson12, Stephen K. Wikel45, Vishvanath M. Nene14,†, Claire M. Fraser46, Frank H. Collins16, Bruce Birren7, Karen E. Nelson14, Elisabet Caler14,† & Catherine A. Hill1

1Department of Entomology, Purdue University, West Lafayette, Indiana 47907, USA. 2Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23529, USA. 3Department of Entomology, North Carolina State University, Raleigh, North Carolina 27695, USA. 4Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1211, Switzerland. 5Swiss Institute for Bioinformatics, Geneva 1211, Switzerland. 6Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. 7The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. 8Centre for Respiratory Biology, UCL Respiratory Department, Division of Medicine, University College London, Rayne Building, 5 University Street, London WClE 6JF, UK. 9SaBio, Instituto de Investigación en Recursos Cinegéticos, IREC-CSIC-UCLM-JCCM, Ronda de Toledo sn, Ciudad Real 13005, Spain. 10Department of Veterinary Pathobiology, Center for Veterinary Health Sciences, Oklahoma State University, 250 McElyea Hall, Stillwater, Oklahoma 74078, USA. 11Laboratory of Malaria and Vector Research, NIAID, Rockville, Maryland 20852, USA. 12VectorBase/EMBL-EBI, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK. 13Bioinformatics Core, Purdue University, West Lafayette, Indiana 47907, USA. 14j. Craig Venter Institute, Rockville, Maryland 20850, USA. 15Genome Sequencing and Analysis Program, Broad Institute, Cambridge, Massachusetts 02142, USA. 16Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, USA. 17Department de Genètica & Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona E-08028, Spain. 18Vascular Physiopathology, Centro Nacional de Investigaciones Cardiovasculares, Madrid 28029, Spain. 19Department of Medicine, Division of Infectious Diseases, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. 20Department of Medicine, University of Connecticut Health Center, Farmington, Connecticut 06030, USA. 21Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, Lausanne CH-1015, Switzerland. 22Genetics, Bioinformatics, and Computational Biology Program, Virginia Bioinformatics Institute at Virginia Tech, Blacksburg, Virginia 24061, USA. 23Department of Biology, Indiana University, Bloomington, Indiana 47405, USA. 24Department Biological Sciences, Markey Center for Structural Biology, Purdue University, West Lafayette, Indiana 47907, USA. 25Department of Biochemistry, Virginia Tech, Blacksburg, Virginia 24061, USA. 26Department of Microbial Molecular Biology, Agricultural Genetic Engineering Research Institute, Giza 12619, Egypt. 27Department of Entomology, Texas A&M University, College Station, Texas 77843, USA. 28Department of Entomology, Kansas State University, Manhattan, Kansas 66506, USA. 29Department of Entomology, University of Minnesota, St Paul, Minnesota 55108, USA. 30Department of Neurosystems, Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK. 31Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California 94143, USA. 32Department of Entomology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. 33Department of Entomology, Center for Disease Vector Research, University of California, Riverside, California 92506, USA. 34Department of Pathology, Cambridge Genomic Services, University of Cambridge, Cambridge CB2 1QP, UK. 35Department of Physiology, School of Medicine-CIMUS-Instituto de Investigaciones Sanitarias, University of Santiago de Compostela, Santiago de Compostela 15782, Spain. 36Department of Experimental and Health Sciences, University Pompeu Fabra, Barcelona 08003, Spain. 37Department of Biomedical Sciences, California State Polytechnic University, Pomona, California 91768, USA. 38Parasitology Section, School of Chemistry & Molecular Biosciences, University of Queensland, Brisbane, Queensland 4072, Australia. 39GeneCology Research Centre, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore, Queensland 4556, Australia. 40Department of Biology, Center for Functional and Comparative Insect Genomics, University of Copenhagen, Copenhagen DK-2100, Denmark. 41Department of Microbiology, Immunology & Biochemistry, University of Minnesota, Minneapolis, Minnesota 55455, USA. 42Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. 43Department of Biochemistry, Genetics and Immunology, University of Vigo, Vigo 36310, Spain. 44Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. 45Department of Medical Genetics, Frank H. Netter MD School of Medicine at Quinnipiac University, Hamden, Connecticut 06518, USA. 46Institute for Genome Sciences, University of Maryland, School of Medicine, Baltimore, Maryland 21201, USA. *These authors contributed equally to this work. †Present addresses: Department of Biochemistry and Molecular Biology, University of Nevada, Reno, Nevada 89503, USA (M.G.-N); Department of Veterinary Science, University of Nevada, Reno, Nevada 89557, USA (A.B.N); Department of Biotechnology, Monsanto Company, Chesterfield, Missouri 63017, USA (J.M.M.); Department of Haematology, University of Cambridge, NHSBT Building, Long Road, Cambridge CB2 0PT, UK (K.M.); Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA (B.P.W or S.K.); Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA (J.B.H.); Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, USA (M.T. or L.I.H.); National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA (V.S.I. or S.B. or M.S.); Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, Nevada, 89154, USA (J.L.A.); AgBiome, Inc., Research Triangle Park, North Carolina 27709, USA (B.W.B.); Centre for Neural Circuits and Behaviour, University of Oxford, Oxford OX1 3SR, UK (V.C.); Department of Biology, West Virginia University, Morgantown 26505, West Virginia (T.D.); Department of Microbiology & Immunology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA (J.J.G or J.H.F.P); Rocky Mountain Laboratories, Biology of Vector-Borne Viruses Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana 59840, USA (J.M.G.); Department of Entomology, Kansas State University, Manhattan, Kansas 66506, USA (D.K.); Department of Veterinary Medicine, University of Maryland, School of Medicine, Baltimore, Maryland 21201, USA (J.K.); Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119074 (K.L.); Department of Entomology, Iowa State University, Ames, Iowa 50011, USA (H.K.); Department of Microbiology, Immunology, and Pathology, Arthropod-borne & Infectious Diseases Laboratory, Colorado State University, Fort Collins, Colorado 80523, USA (R.P.); Department of Vector Biology, Max-Planck-Institut für Infektionsbiologie, Charitéplatz 1, 10117 Berlin, Germany (M.S.S.); French National Institute of Agricultural Research, UMR-BIPAR INRA-ANSES-ENVA, Maisons-Alfort, 94700 France (L.S.); Seres Therapeutics, Cambridge, MA 02142, USA (J.R.W.); International Livestock Research Institute, Nairobi 00100, Kenya (V.M.N.); National Heart, Lung, and Blood Institute, Division of Lung Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA (E.C.). ‡Deceased.