scientific reports

Check for updates

OPEN

Identification of a TNF-TNFR-like system in malaria vectors (*Anopheles stephensi*) likely to influence *Plasmodium* resistance

Subhashini Srinivasan[1]✉, Chaitali Ghosh[2,4], Shrestha Das[1,4], Aditi Thakare[1,4], Siddharth Singh[1], Apoorva Ganesh[1], Harsh Mahawar[1], Aadhya Jaisimha[1], Mohanapriya Krishna[1], Aritra Chattopadhyay[1], Rishima Borah[1], Vikrant Singh[1], Soumya M[2], Naveen Kumar[2], Sampath Kumar[2], Sunita Swain[2]✉ & Suresh Subramani[3]

Identification of *Plasmodium*-resistance genes in malaria vectors remains an elusive goal despite the recent availability of high-quality genomes of several mosquito vectors. *Anopheles stephensi*, with its three distinctly-identifiable forms at the egg stage, correlating with varying vector competence, offers an ideal species to discover functional mosquito genes implicated in *Plasmodium* resistance. Recently, the genomes of several strains of *An. stephensi* of the type-form, known to display high vectorial capacity, were reported. Here, we report a chromosomal-level assembly of an intermediate-form of *An. stephensi* strain (IndInt), shown to have reduced vectorial capacity relative to a strain of type-form (IndCh). The contig level assembly with a L50 of 4 was scaffolded into chromosomes by using the genome of IndCh as the reference. The final assembly shows a heterozygous paracentric inversion, 3L*i*, involving 8 Mbp, which is syntenic to the extensively-studied 2L*a* inversion implicated in *Plasmodium* resistance in *An. gambiae* involving 21 Mbp. Deep annotation of genes within the 3L*i* region in the IndInt assembly using the state-of-the-art protein-fold prediction and other annotation tools reveals the presence of a tumor necrosis factor-alpha (TNF-alpha) like gene, which is the homolog of the Eiger gene in *Drosophila*. Subsequent chromosome-wide searches revealed homologs of Wengen (Wgn) and Grindelwald (Grnd) genes, which are known to be the receptors for Eiger in *Drosophila*. We have identified all the genes in IndInt required for Eiger-mediated signaling by analogy to the TNF-alpha system, suggesting the presence of a functionally-active Eiger signaling pathway in IndInt. Comparative genomics of the three type-forms with that of IndInt, reveals structurally disruptive mutations in Eiger gene in all three strains of the type-form, suggesting compromised innate immunity in the type-form as the likely cause of high vectorial capacity in these strains. This is the first report of the presence of a homolog of Eiger in malaria vectors, known to be involved in cell death in *Drosophila*, within an inversion region in IndInt syntenic to an inversion associated with *Plasmodium* resistance in *An. gambiae*.

Anopheles stephensi is an established urban malarial vector in India and South East Asia with expanding range into African cities. This vector quickly adapts to the local environment, survives extremely high temperatures, and establishes resistance to multiple insecticides (Vector alert: *Anopheles stephensi* invasion and spread, 2019)[1]. Based on the vectorial capacity of *An. stephensi*, there are three variants classified as type-form (high vectorial capacity), intermediate-form (moderate vectorial capacity), and mysorensis (low vectorial capacity)[2,3]. These forms are identifiable based on the number of ridges on eggs[4]. The genomes of three strains of the type-form have been reported recently[5–7]. A comparison of these three genomes revealed the presence of all three genotypes of the 2R*b* inversion among the various strains of the type-form.

[1]Institute of Bioinformatics and Applied Biotechnology, Biotech Park, Electronic City Phase I, Bengaluru 560100, India. [2]Tata Institute for Genetics and Society (TIGS), Center at inStem, Bellary Road, GKVK Campus, Bengaluru 560065, India. [3]TIGS, University of California San Diego, La Jolla, CA 92093, USA. [4]These authors contributed equally: Chaitali Ghosh, Shrestha Das and Aditi Thakare. ✉email: ssubha@ibab.ac.in; sunita.swain@tigs.res.in

	Assembly size (Mbp)	Total contigs	Longest contig (Mbp)	Shortest contig (bp)	Median contig (Kbp)	N50	L50
CANU	405	3464	16.6	1055	44.2	48.7 Kbp	115
WTDBG	205.4	789	33.08	1126	15.4	9.1 Mbp	6
FALCON	293	851	11.4	5734	87.4	1.4 Mbp	46
Flye	211.7	689	27.15	1007	14.6	3.9 Mbp	15
Flye + WTDBG	213.3	521	45.56	1018	8.3	17.04 Mbp	4
(Flye + WTDBG) + SSPACE	213.8	464	55.4	1018	5.9	36.9 Mbp	3

Table 1. Statistics of contig-level assemblies by CANU, FALCON, FLYE, and WTDBG2 tools.

Sixteen paracentric inversions in the chromosomes of *An. stephensi* have been reported by visualizing loop formation in chromosomes from large heterozygous inversions. While chromosome X has no reported inversions, chromosomal arms 2R, 2L, 3R, and 3L have many paracentric inversions characterized by cytobands[8,9]. Moreover, each of these inversions is reported to impart varying resistance to insecticides[10]. However, during the maintenance of various strains of *An. stephensi* under laboratory conditions, several inversions are lost within the first few generations, such as 2R*e*, 3R*c,* and 3L*c*, whereas other inversions, such as 2R*b*, 3R*b,* and 3L*i,* persist through generations even in the absence of any environmental pressure, and in some cases, increasing in frequency with increasing generations[8]. Based on these reports and studies on other malaria vectors, it is now well established that chromosomal inversions are dynamically used by vectors to adapt to environmental changes while retaining fitness[11].

An. gambiae, a malaria vector from Africa, is by far the most extensively-studied and interrogated using genomics. The inversions 2R*b* and 2L*a* in *An. gambiae* have been extensively characterized at the molecular and population levels. While the 2R*b* inversion in *An. gambiae,* involving 8 Mbp, is associated with resistance to DDT and adaptation to climate[11], the 2L*a* inversion, encompassing 21 Mbp, is associated with resistance to *Plasmodium* and desiccation[12,13]. The breakpoints of 2L*a* have been extensively studied by selecting BAC clones spanning breakpoints from strains with alternate arrangements of the 2L*a* configuration[14]. A systems-level study of gene expression in *An. gambiae* between different configurations of a given inversion shows that a large number of genes within the inversion loci are differentially regulated[15].

Comparative proteomics of *D. melanogaster, Ae. aegypti, and An. gambiae* revealed 91 gene trios with a total of 589 genes comprising the immune repertoire[16]. These are homologs of genes from pathways implicated in classical innate immunity or defense functions in insects. The pathways include apoptosis inhibitors (IAPs), oxidative defense enzymes [superoxide dismutases (SODs), glutathione peroxidases (GPXs), thioredoxin peroxidases (TPXs), and heme-containing peroxidases (HPXs)], and class A and B scavenger receptors (SCRs) common to all three species. Of these, more recently, homologs of 361 genes were reported in *An. stephensi*[5], which is in the same range (285, 338, and 353, respectively) of immune-repertoire genes reported for *D. melanogaster* (*Dm*), *An. gambiae* (*Ag*), and *Ae. aegypti* (*Aa*), respectively.

Deep annotation of genes within inversion regions associated with resistance to pathogens offers an opportunity to discover additional immune repertoire genes that may have diverged beyond the point of recognition by standard sequence-based annotation tools. RoseTTaFold, a state-of-the-art protein-fold prediction tool, allows deeper annotation of uncharacterized or divergent proteins, thereby aiding the identification of hidden/missed immune-related genes in *Anopheles*[17]. The idea is that the diversity in proteins far exceeds the number of folds in nature, which is estimated to be as low as one thousand[18]. This is especially important in the identification of cytokine and lymphokine homologs, which are notorious for exhibiting low or undetectable homology at the sequence level[19]. For example, members of TNF and TNFR superfamilies, which has been identified and characterized in *Drosophila*, are missing in the collation of genes implicated in innate immunity in flies and in mosquito vectors[5,16].

In order to understand the evolution of *Plasmodium* resistance in *An. stephensi*, we sequenced and assembled the genome of an isofemale line from a strain of intermediate-form from Bangalore (IndInt) displaying diminished susceptibility to *Plasmodium*[20]. We present the assembly of this strain and report an inversion not found in the genomes of the three strains of the type-form[5–7,21]. Comparative genomics with other *Anopheles* vectors and deep annotation of genes within this inversion, using fold-prediction algorithms, revealed missed or hidden genes likely to be relevant in *Plasmodium* resistance.

Results

Assembly of the genome of IndInt strain. Fourteen generations of isofemale lines from a parent T6 lab-colony of the intermediate-form of *An. stephensi* (IndInt) were grown, both to achieve homogeneity required for high-quality assembly from sequencing of pooled DNA from multiple insects and to enrich for *Plasmodium* resistant phenotype[20]. Multiple tools were used to obtain contig-level assembly of IndInt from 60X coverage of reads sequenced using PacBio Sequel II technologies. The assembly statistics obtained from the various tools are shown in Table 1. In the first round, the best assembly statistics were obtained using the WTDBG2 tool with a L50 value of 6 and N50 9.1 Mbp. The next best statistics was shown by the tool Flye assembler with a L50 of 15 and N50 value of 3.9 Mbp.

To improve the contiguity of the assembly, the four different contig-level assemblies from CANU, FALCON, FLYE, and WTDBG2[22–25] were merged in different combinations using the tool Quickmerge[26]. Assembly statistics

were computed for each of the merged combinations to determine the best merged assembly (Table 1). During the second round, the assembly with the best statistics was obtained by merging assemblies from FLYE and WTDBG2, with a L50 of 4 and a high N50 of 17 Mbp. The computed genome size of this merged assembly was comparable to the actual size of *An. stephensi* genome, and was used for scaffolding. The scaffolds of the error-corrected contigs were built using the SSPACE assembly tool with simulated mate-pairs of varying sizes from the assembly of IndCh strain as reference. The L50 value for the scaffold level assembly was 3 and the N50 value significantly increased from 17 to 37 Mbp (Table 1, last row).

In order to check any heterozygosity from potential inversions, haplotype phasing of the assembly of the IndInt strain was done using a set of primary and associated contigs from FALCON and FALCON-unzip tools, respectively[23]. In the presence of heterozygosity, the mapped reads will be split between two contigs that represent the two forms of the heterozygous allele, resulting in reduced coverage. This creates two peaks in the histogram of read counts, one for the haploid level coverage from heterozygous regions and the other for the diploid level coverage from homozygous regions. A hump in the histogram indicated the presence of potential inversion heterozygosity in the genome of IndInt (Fig. 1a).

Synteny between the assemblies of IndInt and IndCh strains is shown in Fig. 1b. The block view clearly suggests that IndInt is homozygous for the standard form of 2R*b* ($2R^{+b}/2R^{+b}$) inversion similar to the IndCh strain. However, in the 3L arm of chromosome 3, IndInt shows the clear presence of an inversion not seen in the IndCh assembly. The locus of the physical markers revealed that the markers from 42A to 44C are reversed in the assembly, which matched with a reported breakpoint for the 3L*i* inversion in *An. stephensi*[9]. Based on the photomaps from several individuals of IndInt strain (see Fig. 1e), it is clear that 48% of the individuals are heterozygous for the 3L*i* inversion ($3L^{+i}/3Li$), with the rest (52%) being homozygous for the standard form of 3L*i* ($3L^{+i}/3L^{+i}$) (Supplementary Fig. 1). It should be mentioned that the percentage heterozygosity reported here is from an outgrown generation maintained in the lab after the generation used for assembly. Markers 41A, 41B, 41C, and 46D shown in the standard form reported for 3L*i* (top Fig. 1c) are not detectable in the IndInt assembly using the BLASTN tool, perhaps because of limitations of the assembly tools in handling noise generated by reads representing the two arms near the inversion breakpoints from the heterozygous arms. Figure 1d suggests that the IndInt assembly is comparable to the high-quality genomes of other *Anopheles* species, *Drosophila*, and humans at both the contig and scaffold levels.

Validation of the 2R*b* breakpoint region in IndInt. The IndInt strain is homozygous for the standard form of the 2R*b* inversion, making it the second assembled genome for one of the three possible 2R*b* genotypes, such as $2R^{+b}/2R^{+b}$. This provides an opportunity to further refine and validate the breakpoint regions described previously by our group[6]. A BLASTN of the sequences from breakpoint regions from IndCh including 1000 bases to the left and right of both breakpoints, against the IndInt assembly, shows 100% identity for all the 2079 bases from the centromere-proximal breakpoint and 99.8% identity for all the 2009 bases from the centromere-distal breakpoint. We also provide the actual alignment with 50± nucleotides around the breakpoints in Supplementary Fig. 2 This confirms the 2R*b* genotype to be $2R^{+b}/2R^{+b}$ for IndInt.

Interrogation of 3L*i* breakpoints in IndInt. In order to define the breakpoints for the 3L*i* inversion, a contact map was generated by mapping the HiC reads from the UCI strain onto the final IndInt assembly. Since UCI is homozygous for the standard 3L*i* ($3L^{+i}/3L^{+i}$) configuration and IndInt is heterozygous ($3L^{+i}/3Li$), two expected butterfly-like structures in the chromosome 3L were observed in the contact map (Fig. 2a) defining the location of both the proximal and distal breakpoints of the 3L*i* inversion in the IndInt assembly. By magnifying each of the butterfly structures to a resolution of 5 Kbp per dot, the proximal and distal breakpoints of the 3L*i* inversion were estimated to be between bp 22,545,000 to 22,550,000, and between bp 31,555,000 to 31,560,000 respectively.

Synteny of 3L*i* inversion with 2L*a* of An. gambiae. Synteny of the final IndInt assembly was also plotted against the *An. gambiae* PEST strain[27]. Figure 2b shows that chromosome X, 2R, and 3R of *An. gambiae* are homologous to that of IndInt, respectively. However, the synteny between chromosomal arms 2L and 3L of *An. gambiae* are switched to chromosomal arms 3L and 2L of IndInt, respectively, as reported elsewhere[6]. Figure 2c shows the synteny between the 2L*a* inversion of *An. gambiae* with the 3L*i* region of IndInt. The 2L*a* region of *An. gambiae* was taken from a previous study[28]. The circle view (Fig. 2c) of this synteny reveals that the entire 3L*i* inversion region of IndInt consisting of 8 Mbp is syntenic to a big part of the 2L*a* inversion region of *An. gambiae* comprising 21 Mbp. Since the 2L*a* inversion of *An. gambiae* is associated with *Plasmodium* sensitivity[12,13], by analogy, we infer that the 3L*i* of IndInt may also be associated with the *Plasmodium* resistance observed for IndInt strain. Furthermore, it is reported that the strains of *An. gambiae* that are homozygous for the standard form of 2L*a* ($2L^{+a}/2L^{+a}$) show higher vectorial capacity than the corresponding strains that are heterozygous for the inverted form. Similarly, the strain IndCh with higher vectorial capacity is homozygous for the standard form of 3L*i* ($3L^{+i}/3L^{+i}$) and IndInt is heterozygous for the 3L*i* inversion ($3L^{+i}/3Li$), suggesting the potential role of the 3L*i* inversion in *Plasmodium* sensitivity.

Functional characterization of genes within the 3L*i* region. Within the 2L*a* region of the *An. gambiae* PEST strain spanning 21 Mbp, there are 1243 genes of which there are 494 genes that are homologous/syntenic to the 3L*i* region of IndInt. Of these, 464 are functionally annotated in IndInt based on sequence level annotation tools, and the remaining 30 genes are uncharacterized due to a lack of sequence homology to known functional genes. Among functionally-characterized genes, we found 36 isoforms of cuticle-forming genes in IndInt. Interestingly, cuticle thickening is associated with pyrethroid & insecticide resistance in other *Anopheles*

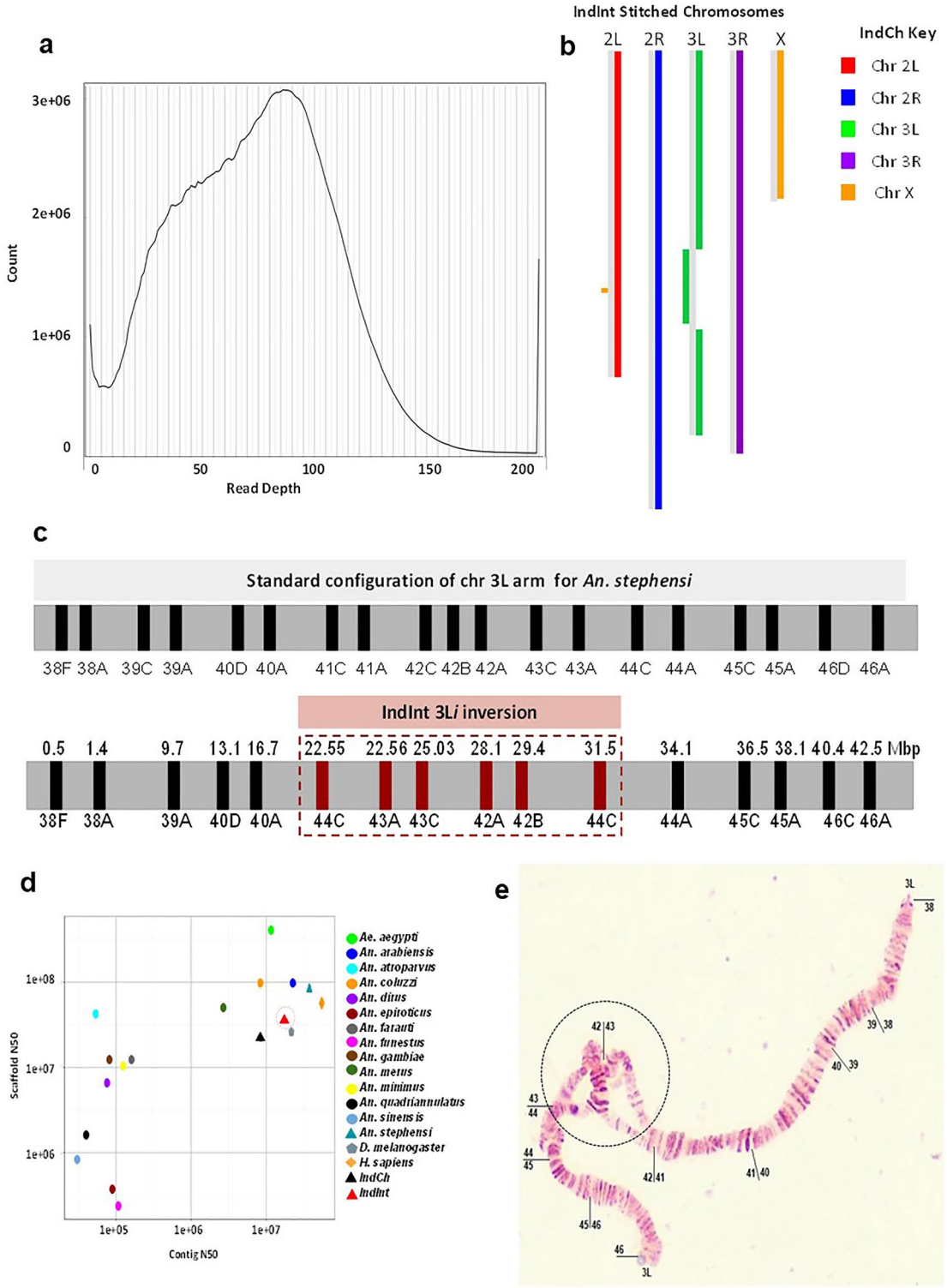


Figure 1. (**a**) Histogram for the genome assembly of IndInt showing evidence of heterozygosity by Purge Haplotigs tool. (**b**) Block diagram of synteny between IndCh and IndInt assembly showing an inversion block in chromosome 3L of IndInt. (**c**) Top—Orientation of the physical markers as expected in the standard configuration for *An. stephensi*. Bottom—BLASTN analysis of the assembled IndInt strain revealing the orientation of the physical markers consistent with the 3L*i* inversion. (**d**) Quality of IndInt assembly as compared to that of other *Anopheles* species and completed gold-standard genomes of human and fly. (**e**) A sample photograph of polytene chromosomes from the IndInt strain showing the loop from the heterozygous inverted configuration of the 3Li.

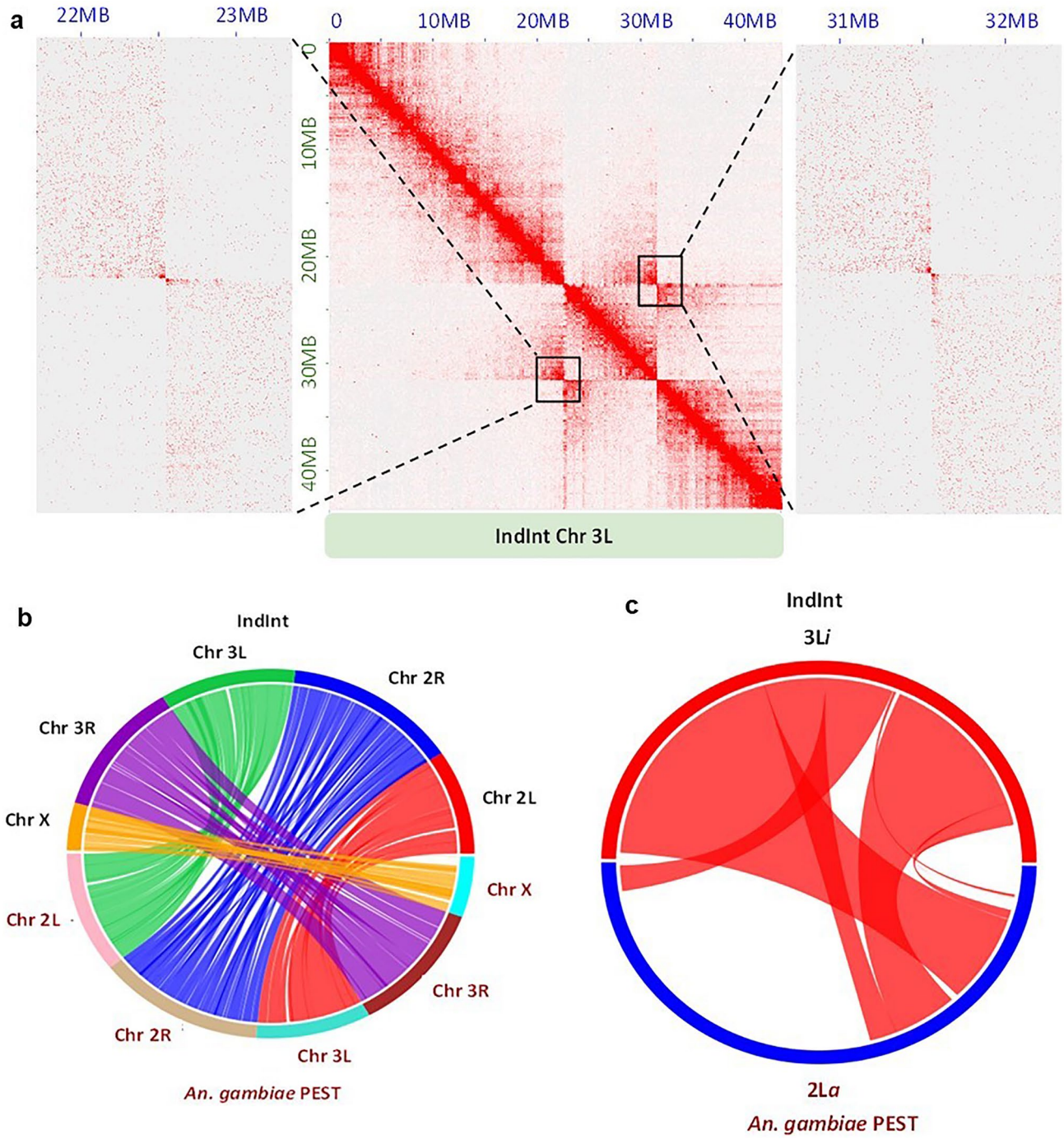


Figure 2. (**a**) Proximal and distal breakpoints obtained by mapping UCI HiC reads against the chromosome 3L of IndInt. The black boxes represent the centromere-proximal (left) and centromere-distal (right) breakpoints located between the ranges of bp 22,545,000 to 22,550,000 and bp 31,555,000 to 31,560,000, respectively, in the IndInt assembly. Adjacent to the contact map are the enlargements of the butterfly structures, for a resolution of 5 Kbp per dot. (**b**) Synteny between *An. gambiae* and IndInt assemblies. (**c**) Synteny between the *An. gambiae* 2L*a* and the IndInt 3L*i* regions.

species[29], while also resisting *Plasmodium* infection[30]. We found 99 missense mutations in the IndInt strain within 23 copies of the cuticle genes, as compared to the IndCh strain of the type form, which has a homozygous configuration for the standard form of 3L*i*, like the UCI strain.

In order to find genes implicated in immune response within the 3L*i* region of IndInt, we intersected the 494 genes within this region with a list of 589 genes implicated in immunity in insects[16]. Out of these 589 genes, we found only 17 genes within the 3L*i* region (Table 2) including NFkB, a transcription factor implicated in cytokine production in humans.

IndInt Augustus ID	IndInt Chr 3L start	IndInt Chr 3L end	Waterhouse et. al [16] Immune DB ID	Alignment percentage	Function
g22033.t1	22194979	22196095	AGAP005552	76.24%	Peptidoglycan recognition protein
g22086.t1	22559888	22560559	AGAP007020	92.83%	Thioredoxin peroxidase 5
g22107.t1	22657471	22661294	AGAP006939	72.71%	Armitage protein
g22121.t1	22702737	22707661	AAEL001675	60.41%	Clip domain, serine protease family A
g22273.t1	24373199	24379261	AGAP006631	88.68%	Class A scavenger receptor (SRCR domain) with serine protease domain
g22562.t1	26855939	26862215	AGAP006914	61.09%	Fibrinogen-related protein 1
g22567.t1	26922106	26923713	AGAP006911	91.98%	Serine protease inhibitor—Serpin 2
g22568.t1	26925252	26932268	AGAP006909	83.54%	Serine protease inhibitor—Serpin 1
g22799.t1	28434986	28435854	AGAP006790	72.73%	Fibrinogen C-terminal domain-containing protein
g22840.t1	28809510	28812635	AGAP006761	83.96%	3-Glucan binding protein
g22865.t1	28987304	28992526	AGAP006747	79.51%	IMD pathway signaling NF-kappaB relish-like transcription factor
g22870.t1	29022137	29026993	AGAP006743	64.44%	Unspecified product BLASTP: similar to ficolin 1
g22921.t1	29370252	29370571	AGAP006722	81.25%	Unspecified product BLASTP: similar to Cecropin B1
g23080.t1	30464940	30470336	AGAP005672	94.10%	Staphylococcal nuclease domain-containing protein 1
g23109.t1	30680813	30689983	AGAP005652	86.48%	ATP-dependent RNA helicase DDX5/DBP2
g23151.t1	30998165	31003426	AGAP005625	78.15%	Class A scavenger receptor (SRCR domain) with serine protease domain
g23220.t1	31575970	31576777	CPIJ004568	68.64%	Similar to Cu-chaperone

Table 2. Genes within the 3Li region of IndInt that are reported to be implicated in immune response by Waterhouse et al. [16].

The genes implicated in *Plasmodium* resistance in *An. gambiae*[31] LRIM1, APL1C, and TEP1 are not within the 3L*i* region in IndInt. The TEP1 protein in IndInt is located on chromosomal arm 2L. The LRIM1 gene is outside the 3L*i* region, while the gene APL1C is located immediately outside the distal breakpoint of the 3L*i* region. The LRIM1 gene has leucine-rich repeats and forms a disulfide-bonded complex with APL1C (PDB ID: 3OJA[31]), which stabilizes the mature TEP1 and promotes its binding to parasites to trigger pathogen destruction. Also, more recently, a knockout of LRIM1 in *An. stephensi* was shown to play a role in vector competence[32]. Furthermore, LRIM1 and APL1C are within the 2L*a* region in *An. gambiae,* which is three times larger than the 3L*i* region in IndInt. In IndInt, the LRIM1 has four and six missense mutations compared to IndCh and UCI strains of type-form, respectively. However, since the gene APL1C shows no missense mutation in IndInt compared to IndCh, we expected other unannotated paralogs of LRIM1 within the 3L*i* region of IndInt. Deep annotation of the 30 functionally uncharacterized genes within the 3L*i* region shown in Supplementary Fig. 3 with fold-prediction algorithms, revealed two novel leucine-rich proteins, g22432 and g22212, similar in structure to LRIM1 (Supplementary Fig. 3a and 3b). These two LRIM1-like genes could potentially work with the intact APL1C gene in IndInt to launch a host defense system to antagonize the malaria parasite. The high conservation of these, otherwise uncharacterized genes across *Anopheles* species, shown in the multiple sequence alignments of g22432 and g22212, suggests functional conservation across species (Supplementary Fig. 4a and 4b).

Recently, members of LPS-induced TNF-alpha factors, LITAFs, have been shown to regulate anti-*Plasmodium* immunity in *An. Gambiae*[33,34]. A search in the assembled genome of *An. stephensi* identified four members of LITAF-like transmembrane genes (Supplementary Text 1). LITAFs are known to send signals to the nucleus to transcribe TNF-alpha molecules in humans[35]. In fact, LPS-induced animal models were used in developing drugs against TNF-alpha to treat auto-immune disorders. We were also able to find homologs of all the genes involved in this signaling pathway in IndInt (Supplementary Text 2).

We found an Eiger homolog, g22826, which is within the 3L*i* region of IndInt and is missing in the reported list of 361 genes linked to the immune repertoire in *An. stephensi*[5]. A crystal structure of the complex of Eiger and its receptor from *Drosophila* has revealed that Eiger belongs to the TNF superfamily and its two receptors, Wengen and Grnd, are members of the TNFR superfamily (PDB: 6ZT0[36])[37–39]. A search for an active TNFR-like gene in *An. stephensi,* using the HMM model (PF00020) of the cysteine-rich domain (CRD) from PFAM database, identified two potential hits g1129 and g18030 in IndInt. It should be mentioned that while Wengen displays detectable homology with TNFRSF, predicted Grnd homologs from many sequenced species, including malaria vectors, remain uncharacterized by sequence-based annotation tools. The gene, g1129, from chromosome X of the IndInt strain encodes a Wengen homolog of *Drosophila*, with a well-defined extracellular CRD, and a transmembrane region. The Grnd homolog (g18030) in the IndInt strain consists of an intact CRD region sandwiched between a predicted signal sequence and transmembrane region (Supplementary Fig. 5c and 5d). We found that the Eiger gene has a frame-shift mutation at the tail-end of the very conserved folding domain in the IndCh strain. Also, in the UCI and STE2 strains, the predicted Eiger homologs are missing the critical

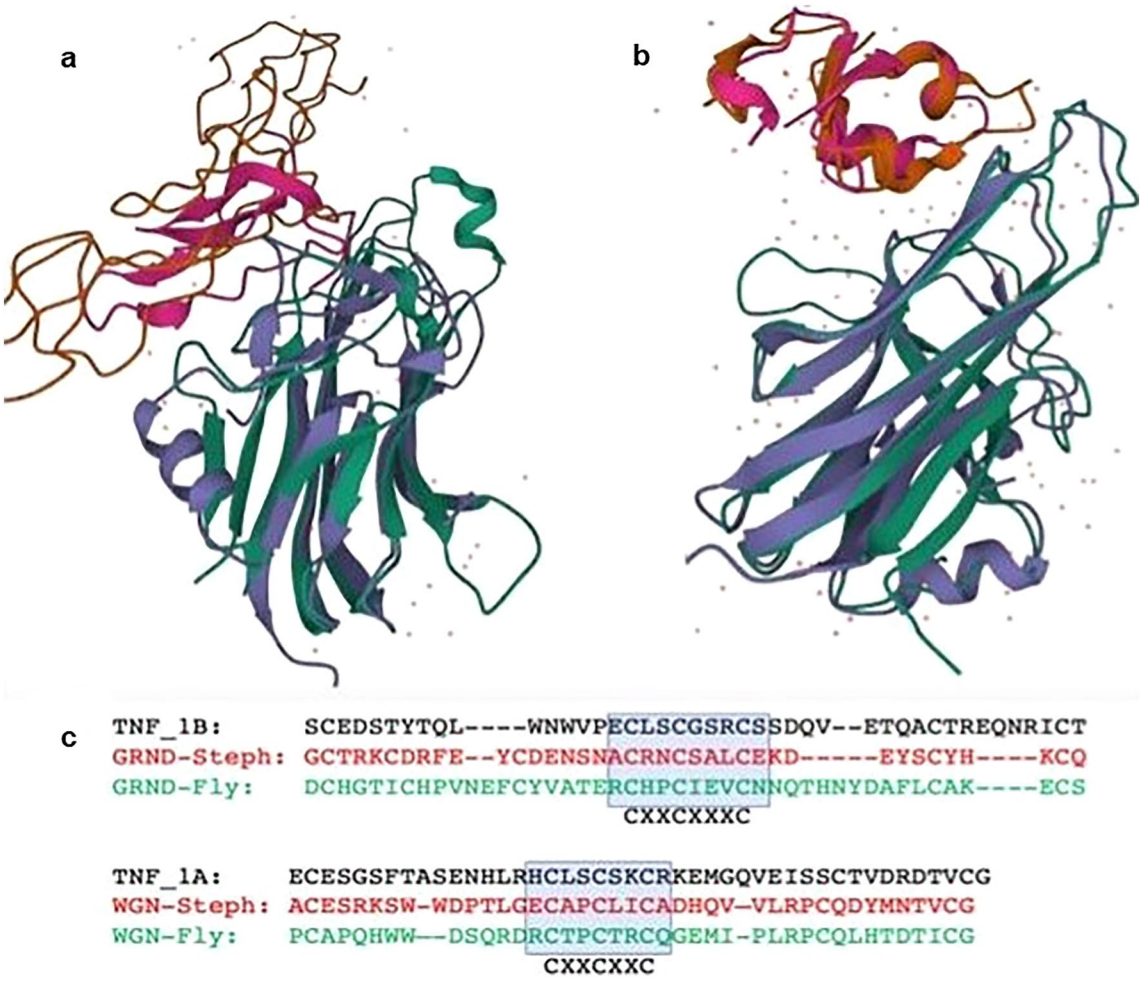
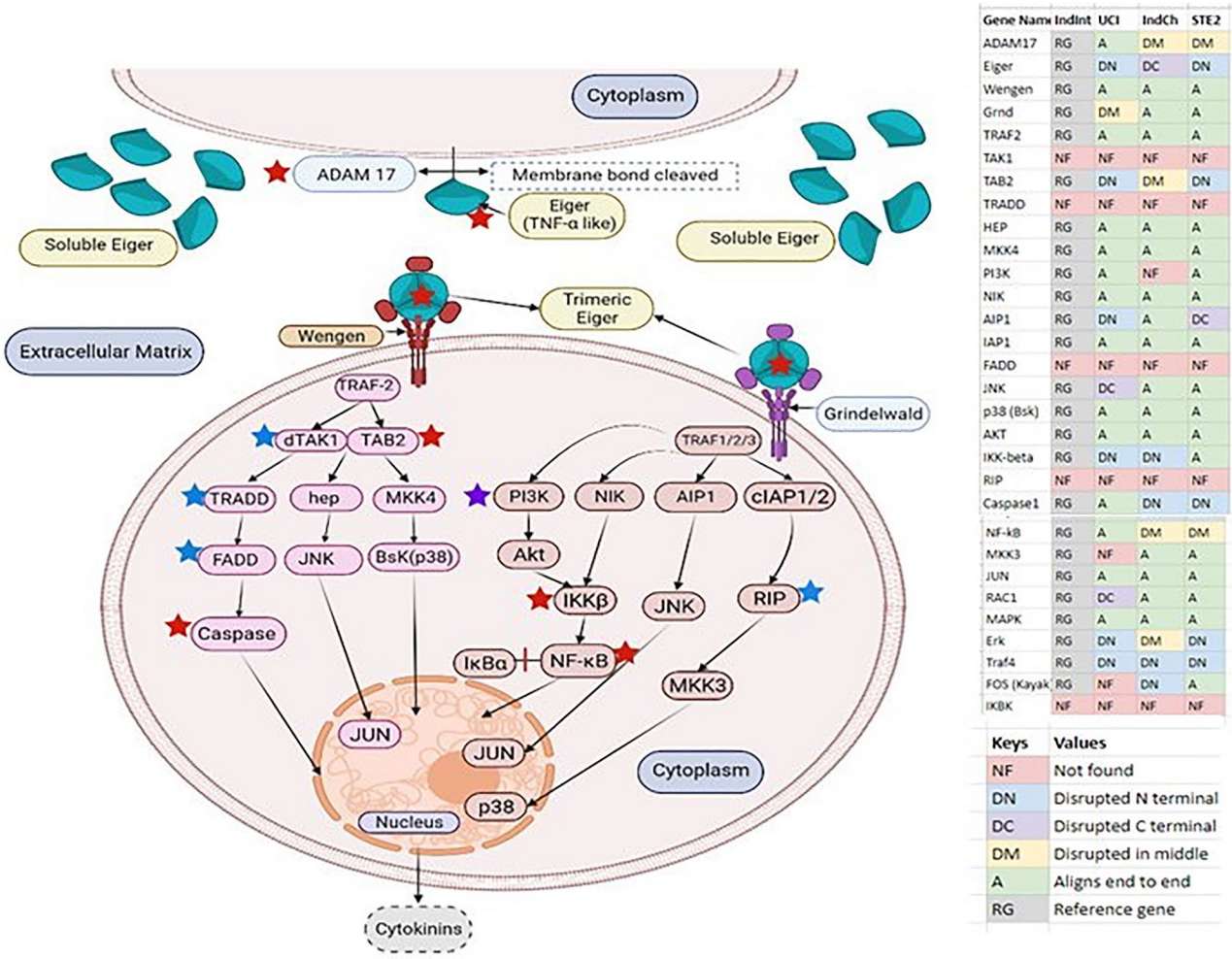


Figure 3. (**a**) De novo model of Eiger and comparative model of Wengen superimposed on the crystal structure of the complex of TNFb-TNFR-1A (1tnr). (**b**) De novo model Eiger and comparative model of Grnd superimposed on the crystal structure of the complex of Eiger and Grnd from *Drosophila* (6zt0). (**c**) Sequence alignment of Wengen and Grnd against TNFR-1A, -1B and Grnd from *Drosophila* used in comparative modeling.

N-terminal transmembrane region (Supplementary Fig. 6). In Fig. 3a, a 3D model of the Eiger gene from IndInt, built using RoseTTaFold tool[17], is shown superimposed on the crystal structure of human TNF-TNFR complex (PDB: 1TNR[40]) on the left, along with a Wengen model built using homology to the second domain of TNFR-1A. Similarly, the Eiger model is shown superimposed on the crystal structure of the Eiger-Grnd complex (PDB: 6ZT0[36]), along with a model of Grnd built using homology to Grnd of *Drosophila* in Fig. 3b. The choice of templates for comparative modeling of the CRDs of the two receptors is based on the conserved cysteine patterns of the ligand binding regions, which are different for the two receptors Wengen and Grnd. Interestingly, the Wengen binding domain has a cysteine pattern similar to TNFR-1A of humans and that of Grnd is similar to that from TNFR-1B, suggesting an evolutionary-conserved link between Eiger and TNF-alpha mediated signaling in arthropods and vertebrates, respectively.

A functional TNF-alpha/TNFR pathway in malaria vectors with lower vectorial capacity. It is now well established that the *Drosophila* genome harbors two functional TNFR receptors and a single ligand, Eiger, belonging to the TNF-TNFR system[36]. In order to predict whether the homologs of the Eiger and its cognate receptor genes in *An. stephensi* (IndInt) are potentially functional, we used signal peptide and transmembrane predictions on the respective protein sequences derived from IndInt (Supplementary Fig. 5). The Eiger homolog shows a transmembrane region at its N-terminus (Supplementary Fig. 5a) consistent with type II configuration common to all members of the TNF superfamily across organisms. The two receptor homologs show clear transmembrane regions (Supplementary Fig. 5b and 5c) following the cysteine-rich folding domain. However, only Grnd, but not Wengen, has a predicted signal peptide signature (Supplementary Fig. 5d) N-terminal to the CRD, consistent with the localization of these two genes in *Drosophila,* where Wengen is almost always localized in intracellular vesicles[36]. We also found the homolog of the TACE gene in IndInt, which is a TNF-alpha processing metalloprotease that converts the type-II membrane-anchored ligand into its soluble/



Gene Nam	IndInt	UCI	IndCh	STE2
ADAM17	RG	A	DM	DM
Eiger	RG	DN	DC	DN
Wengen	RG	A	A	A
Grnd	RG	DM	A	A
TRAF2	RG	A	A	A
TAK1	NF	NF	NF	NF
TAB2	RG	DN	DM	DN
TRADD	NF	NF	NF	NF
HEP	RG	A	A	A
MKK4	RG	A	A	A
PI3K	RG	A	NF	A
NIK	RG	A	A	A
AIP1	RG	DN	A	DC
IAP1	RG	A	A	A
FADD	NF	NF	NF	NF
JNK	RG	DC	A	A
p38 (Bsk)	RG	A	A	A
AKT	RG	A	A	A
IKK-beta	RG	DN	DN	A
RIP	NF	NF	NF	NF
Caspase1	RG	A	DN	DN
NF-kB	RG	A	DM	DM
MKK3	RG	NF	A	A
JUN	RG	A	A	A
RAC1	RG	DC	A	A
MAPK	RG	A	A	A
Erk	RG	DN	DM	DN
Traf4	RG	DN	DN	DN
FOS (Kayak	RG	NF	DN	A
IKBK	NF	NF	NF	NF

Keys	Values
NF	Not found
DN	Disrupted N terminal
DC	Disrupted C terminal
DM	Disrupted in middle
A	Aligns end to end
RG	Reference gene

Figure 4. Pathways leading to Eiger-Grnd/Wgn-mediated signaling collated from TNF-TNFR and Eiger-Wgn-mediated signaling pathways obtained from the KEGG database[43–45] and validated using other sources[46,47]. Red stars represent the disrupted genes in the IndCh genome, purple star is the gene missing in IndCh but present in IndInt, blue stars are the missing pathway genes in all four assembled *An. stephensi* genomes. The inset table shows the status of the genes in all 4 strains (IndInt, IndCh, UCI and STE2) of *An. stephensi* for which a high-quality genome is available.

active form in human[41]. In the Eiger gene from IndInt, we also found a cleavage site (AQ) N-terminal to the folding domain that is known to be specific to TACE[42]. The gene TACE itself is rendered active by genes called iRhom1 and iRhom2, which are actively being developed as potential targets to inhibit TNF-alpha processing. Interestingly, there are homologs of iRhoms in IndInt, albeit with a much longer N-terminal region than human iRhom1 and 2.

In order to show that all the genes in the cytoplasm necessary for Eiger-Wgn-Grnd mediated signaling are encoded by the genome of IndInt, we identified homologs of all the pathway genes from TNF-alpha and eiger signaling pathways from KEGG[43–45] (Supplementary Text 2). Figure 4 shows genes implicated in Wengen-based (left) and Grnd-based (right, hypothesis) based on the homology of the receptors to TNFR-1A and -1B, respectively. Several genes are disrupted in the other strains of type-form viz IndCh, UCI and STE2, as shown in Fig. 4, Supplementary table 1 and Supplementary table 2. The homologs for genes marked by red and purple stars are disrupted and absent, respectively, in the IndCh genome, potentially disrupting the Eiger-mediated signaling pathway. These genes are Eiger, ADAM17, Tab2, IKKb, NFkB, and Caspase. Interestingly, genes with blue stars are missing in all 4 assembled genomes.

It should be mentioned that a majority of these genes including Eiger, Grnd, and Wengen are missing in the collated list of 589 immune related genes in insects[16] and, more specifically, missing from a list of 361 immune related genes in *An. stephensi* reported recently in the UCI strain[5].

Discussion

We report a high-quality genome assembly of an isofemale line of the intermediate-form of *An. stephensi*, IndInt, displaying low vectorial capacity relative to the strain IndCh of type-form[20]. A comparative analysis of this assembly with those of type-forms reveals a 3L*i* inversion (in a heterozygous configuration) that is uniquely

present only in the IndInt strain. As shown in Fig. 2c, this inversion is syntenic to the 2L*a* inversion region in *An. gambiae*, the heterozygous genotype of which is shown to be associated with *Plasmodium* resistance[12]; thus, providing an opportunity to zero-in on genes that may be offering resistance to *Plasmodium*.

Here, for the first time, we report the presence of an Eiger homolog, a member of the TNF family, within this inversion region (Fig. 3b). Considering that Eiger has been shown to cause cell death in *Drosophila*[37], it is tempting to hypothesize apoptotic role for Eiger in defence of parasites. Furthermore, considering that the Eiger gene in the IndCh strain harbors a frameshift mutation within the structurally folding domain and considering the extensive role of TNF-alpha in human immunity, it is tempting to predict a role for the Eiger gene in the innate immune response of *An. stephensi* against *Plasmodium* in IndInt strain. Similarly, the predicted Eiger genes in the STE2 and UCI strains are both disrupted in the N-terminal segments lacking a transmembrane region and several genes in the Eiger pathway are disrupted in the IndCh strain (Fig. 4, Supplementary table 2). These observations suggest a testable correlation between high vectorial capacity in *An. stephensi* strains with a dysfunctional Eiger-mediated signaling pathway, which has a lower vectorial capacity for *Plasmodium* with a potentially functional Eiger-signaling pathway. Furthermore, the increased expression of Eiger-Wgn genes in response to infection by Nora virus in *Drosophila*[48] demands an exploration of the role of Eiger mediated signalling in *Plasmodium* resistance in *An. stephensi* and in other malaria vectors, in general. More recently, in support of this, it has been shown that inhibitors of JNK[49], a downstream gene of the Eiger/TNF-alpha pathway, abolish resistance to *Plasmodium* in both *An. gambiae*[50] and *An. stephensi*[51].

The gene TNF-alpha and its two cognate receptors in humans, TNFR-1A and -1B, are implicated both in adaptive and innate immune response. Indeed, drugs inhibiting this pathway are now available to treat various autoimmune diseases such as rheumatoid arthritis, Crohn's disease, and psoriasis. The two cognate receptors of Eiger, Wengen and Grnd, show homology in their cysteine patterns to TNFR-1A and TNFR-1B respectively (Fig. 3c), suggesting evolutionary conservation of this signaling pathway. Consistent with the ability of the Eiger gene to participate in innate immune response analogous to signaling by TNF-alpha, we found genes in *An. stephensi* (IndInt) that are homologous to genes implicated in TNF-alpha mediated signaling (Fig. 4). Most interestingly, Eiger is not only a type-2 membrane protein like TNF-alpha, but the presence of homologs of ADAM17 and those of iRhoms in IndInt suggests that Eiger is rendered soluble/active similar to TNF-alpha, thus strengthening the evolutionary conservation of this entire pathway. Considering that many genes downstream of the TNF-alpha pathway are involved in multiple processes, it is surprising that a majority of these are missing from the list of genes with immune repertoire reported for insects[16] (Supplementary Table 1).

It is known that human TNF-alpha reduces *Plasmodium falciparum* growth and activates calcium channels in human malarial parasites[52]. It is reasonable to hypothesize that mosquito vectors may also elicit similar responses via the Eiger-Wengen-Grnd signaling pathway in parasites to control their growth. Furthermore, the Eiger mediated cell-death in *Drosophila* could be reversed by JNK inhibitors[36]. Most importantly, JNK inhibitors reverse *Plasmodium* resistance in both *An. gambiae*[50] and *An. stephensi*[51]. These observations support the hypothesis that a functional Eiger pathway in malaria vectors may reduce the growth of *Plasmodium falciparum,* similar to the action of TNF-alpha in the human host.

Drosophila and mosquitoes diverged some 350 million years ago, suggesting that the TNF-alpha based innate immunity may have evolved before this point of inflection between vertebrate and invertebrates. A search for TNF and TNFR-like molecules in the Alpha-fold database of *C. elegans* reveals ~ 14 TNF-like molecules without any TNFR-like homologs. The TNF-like molecules in *C. elegans* are actually cerebellin homologs. Interestingly, cerebellins are also both type 2-membrane proteins and form threefold functional oligomers just like members of the TNF superfamily. As proof of this, there are homologs of glutamate receptors and neurexin genes encoded by *C. elegans*, which are involved in forming synaptic structures in rats[53]. Since the TNF-alpha signaling system is conserved in flies, mosquitoes and vertebrates, missing/lost in *nematodes*, innate immunity by the TNF-alpha mediated signaling may predate the invertebrate-vertebrate split and TNF-like fold evolution may even be older.

Conclusion

The three forms of *An. stephensi* (type, intermediate and mysorensis) displaying varying vector competence, which are clearly identifiable by their number of egg ridges, offered an ideal system to discover genes implicated in *Plasmodium* resistance by interrogating genomes of strains with low vectorial capacity. Here, we have demonstrated that a bottom-up approach starting from phenotype to gene discovery, is more efficient in identifying hidden/divergent genes implicated in *Plasmodium* sensitivity. We have identified several genes in *An. stephensi* from gene families known to be involved in innate immunity including LITAFs, novel homologs of LRIM1 within the 3L*i* region, a homolog of TNF-alpha within the 3L*i* region, and majority of the genes from TNF-alpha signaling pathway. In particular. this is the first report of a homolog of Eiger in vectors with demonstrated role in innate immunity particularly apoptosis. This approach has opened up several novel avenues/molecular targets for advancing our understanding of the evolution of *Plasmodium* resistance in vectors and, perhaps, offer effective ways to manage malaria in the near future.

Materials and methods

Maintenance of An. stephensi in insectary. Larvae of *An. stephensi* were collected from Sriramanahalli area of Bangalore rural (13.4310° N, 77.3310° E) from the state of Karnataka, India, and successfully colonized in the insectary. Larvae were provided with larval food prepared by mixing Brewer's yeast and dog biscuits at a ratio of 30:70. Pupae were bleached with 1% sodium hypochlorite for 1 min and kept in adult rearing cages (Bugdorm-4S3030) (W32.5 × D32.5 × H32.5 cm). The adults were fed on a mixture of 8% sucrose, 2% glucose solution mixed with 3% multivitamin syrup (Polybion LC®)[54]. Adults were maintained in the insectary at 28 ± 1 °C, RH 75 ± 5% and 12:12-h day and night photoperiod cycles. Adult females were provided with blood on day 7 or 8

post-emergence to obtain eggs for continuing next progeny. Eggs were collected on day 4 in an ovicup containing water and lined with filter paper at the inner margin. Larvae were hatched out after 36–48 h and transferred to the rearing trays (L39 × B30 cm; Polylab, Catalog no. 81702) with larval food.

Establishment of IndInt isofemale line of An. stephensi. About 10–15 gravid females were separated from the intermediate line[55]. On day 3, each gravid female was transferred carefully to a single ovicup. Each ovicup was covered with a nylon net. A cotton ball soaked with sugar solution was kept on the top of each ovicup. The line from single female (G_0) was selected to generate the isofemale lines that laid the highest number of eggs and had highest percent hatchability with egg ridge numbers 15–17. The eggs thus collected from the single female were allowed to hatch, and the larvae were provided with larval food and reared to adults as per the protocol mentioned earlier. Emerged adult siblings (G_1) were kept in a mosquito rearing cage and allowed to inbreed (sibling mating). Mated females were blood fed, and the same procedure was followed as mentioned earlier and continued till further generations. As of July 2022, the IndInt isofemale colony has been maintained in our insectary for 55 generations.

Karyotyping of chromosome 3 of IndInt strain. Polytene chromosomes were prepared from the ovarian nurse cells collected from semi-gravid females of the IndInt isofemale lines as per the method of Ghosh & Shetty, which was adapted from our recent publication[6]. The semi-gravid females were anesthetized and placed on a microslide in a drop of phosphate buffer. The ovaries were pulled out gently and fixed in modified Carnoy's fixative (methanol: acetic acid, 3:1) for 2–3 min. The remaining mosquito body was preserved in an Eppendorf tube with a few drops of phosphate buffer solution and kept at − 20 °C for PCR analysis. After fixation, the ovaries were stained with lacto acetic orcein for 15–20 min. After staining, 60% acetic acid was added to it and a clean coverslip was placed on top of the stained sample. Gentle pressure was applied on the cover glass for squashing. The edges of the coverslip were sealed with nail polish to avoid evaporation. The slides were examined under a microscope using a 40X objective lens for chromosome inversions. The inversion nomenclature and their frequency were recorded[56].

Assembly. PacBio reads from the IndInt strain were assembled independently using FLYE[24] and WTDBG2[25] assemblers. The resulting assemblies were combined using Quickmerge[26] where the FLYE assembly was taken as the query and the WTDBG2 assembly served as the reference. Two rounds of Arrow polishing (PacBio gcpp v2.0.2) using PacBio reads, followed by two rounds of Pilon polishing (v1.22) using 100 × Illumina reads, were carried out on the merged assembly. In order to proceed with reference based de novo assembly, simulated mate-pair reads were generated from the IndCh standard assembly (https://doi.org/10.1038/s41598-022-07462-3) using the tool WGSIM (https://github.com/lh3/wgsim) for the sizes 50 Kbp, 100 Kbp, 250 Kbp, 500 Kbp, 1 Mbp, 2.5 Mbp and 5 Mbp. These simulated mate-pair reads along with the contig level assembly from FLYE + WTDBG2 served as input to obtain a scaffold level assembly using the SSPACE[57] tool. DNA physical marker sequences for each chromosome arm published by Jiang et al.[21]. were downloaded and aligned against the scaffolds using BLAST, in order to assign the chromosomal positions to the scaffolds. The scaffolds were stitched based on the order and orientation of the markers into pseudo-chromosomes. 100 N's were added between two scaffolds during stitching. Physical marker sequences were realigned to the stitched chromosomes for validation.

Haplotype phasing. The assemblers FALCON and FALCON-Unzip[23] were used to phase the IndInt genome into haplotypes. Raw IndInt PacBio reads were used by the FALCON assembler to produce a set of primary contig files and an associate contig file representing the divergent allelic variants. The output of primary and associate contig files were then used by FALCON-Unzip to produce partially phased primary contigs (all_p_ctg.fa) and fully phased haplotigs (all_h_ctg.fa), which represent divergent haplotypes. Polishing was performed on the phased contigs by FALCON-Unzip using Arrow. This procedure was adapted from our recent publication[6].

Purge haplotigs. The tool 'Purge Haplotigs'[58] was used to determine the degree of heterozygosity in the IndInt strain. PacBio reads of IndInt strain were mapped to the consensus primary assembly (cns_p_ctg.fasta) obtained from FALCON-Unzip to obtain an aligned BAM file. Coverage analysis was performed on the BAM file to generate a read-depth histogram.

Identification of 3L breakpoints in IndInt. The tool HiC-Pro[59] was used to generate a contact map by mapping the HiC reads from the UCI strain onto the genome of IndInt with different inversion genotype to produce a butterfly-like structure representing the breakpoints, which were visualized and extrapolated from the contact map of the tool Juicebox[60]. The breakpoints were validated by mapping the PacBio and Illumina reads onto the respective genomes. Alignment using 'unimap' was performed between the genomes of the IndInt and UCI strains, to determine the precise breakpoints for the 3L*i* inversion to consolidate with the findings from HiC-Pro.

Gene annotation. AUGUSTUS (version 3.2.3)[61], an eukaryotic gene prediction tool, was used to find protein-coding genes for all the assembled genomes. The model organism closest to *An. stephensi, Ae. aegypti,* which was made available by AUGUSTUS, was used for gene prediction. This procedure was adapted from our recent publication[6].

Structural prediction of genes of unknown functions. The genes obtained from AUGUSTUS were used as an input for RoseTTAFold (https://robetta.bakerlab.org/, version August 2021), a deep learning-based protein modeling method. High confidence models obtained from Robetta were used as an input for PyMOL(Version 2.0 Schrödinger, LLC). Regions conserved across the models were identified after structural alignment in PyMol as putative protein domains. Protein contact maps were generated using BIOVIA Discovery Studio Visualizer (v21.1.0.20298) to visually confirm the conserved regions and check for short and long-range interactions in the identified domains. The identified domains were given as input in both sequence and structure-based comparison softwares. Structural comparison was done using DALI (DaliLite.v5) and sequence comparison was done using online utilities such as SwissModel Expasy, Conserved Domain Database from NCBI, InterPro, HMMER Homology Search, and Pfam Sequence Search. Default parameters were used for all the tools. The DALI result showing optimal RMSD value, and seen as hits in the sequence-based tools as well, was chosen as the possible domain.

Data availability

The raw PacBio/Illumina reads used in the assembly and the IndInt chromosome level assembly are uploaded onto the NCBI server under the BioProject ID: PRJNA757559.

Received: 12 August 2022; Accepted: 4 November 2022
Published online: 09 November 2022

References

1. Programme GM. Vector alert: Anopheles stephensi invasion and spread. World Health Organization; 26 Aug 2019 [cited 30 Sep 2022]. Available: https://www.who.int/publications/i/item/WHO-HTM-GMP-2019.09.
2. Shinzawa, N., Ishino, T., Tachibana, M., Tsuboi, T. & Torii, M. Phenotypic dissection of a plasmodium-refractory strain of malaria vector *Anopheles stephensi*: The reduced susceptibility to *P. berghei* and *P. yoelii*. *PLoS One.* **8**, e63753 (2013).
3. Basseri, H. R., Mohamadzadeh Hajipirloo, H., Mohammadi Bavani, M. & Whitten, M. M. A. Comparative susceptibility of different biological forms of *Anopheles stephensi* to *Plasmodium berghei* ANKA strain. *PLoS ONE* **8**, e75413 (2013).
4. Chavshin, A. R. *et al.* Molecular characterization, biological forms and sporozoite rate of *Anopheles stephensi* in southern Iran. *Asian Pac. J. Trop. Biomed.* **4**, 47–51 (2014).
5. Chakraborty, M. *et al.* Author Correction: Hidden genomic features of an invasive malaria vector, *Anopheles stephensi*, revealed by a chromosome-level genome assembly. *BMC Biol.* **20**, 96 (2022).
6. Thakare, A. *et al.* The genome trilogy of *Anopheles stephensi*, an urban malaria vector, reveals structure of a locus associated with adaptation to environmental heterogeneity. *Sci. Rep.* **12**, 3610 (2022).
7. Chida, A. R. *et al.* A near-chromosome level genome assembly of *Anopheles stephensi*. *Front. Genet.* **11**, 565626 (2020).
8. Mahmood, F. & Sakai, R. K. Inversion polymorphisms in natural populations of *Anopheles stephensi*. *Can. J. Genet. Cytol.* **26**, 538–546 (1984).
9. Kamali, M. *et al.* An integrated chromosome map of microsatellite markers and inversion breakpoints for an Asian malaria mosquito, *Anopheles stephensi*. *J. Hered.* **102**, 719–726 (2011).
10. Shetty, N. J., Hariprasad, T. P. N., Sanil, D. & Zin, T. Chromosomal inversions among insecticide-resistant strains of *Anopheles stephensi* Liston, a malaria mosquito. *Parasitol. Res.* **112**, 3851–3857 (2013).
11. Ayala, D., Ullastres, A. & González, J. Adaptation through chromosomal inversions in Anopheles. *Front. Genet.* **5**, 129 (2014).
12. Riehle, M. M. *et al.* The 2La chromosome inversion is associated with susceptibility to in Africa. *Elife.* https://doi.org/10.7554/eLife.25813 (2017).
13. Gray, E. M., Rocca, K. A. C., Costantini, C. & Besansky, N. J. Inversion 2La is associated with enhanced desiccation resistance in *Anopheles gambiae*. *Malar. J.* **8**, 215 (2009).
14. Sharakhov, I. V. *et al.* Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the *Anopheles gambiae* complex. *Proc. Natl. Acad. Sci. USA* **103**, 6258–6262 (2006).
15. Cheng, C., Tan, J. C., Hahn, M. W. & Besansky, N. J. Systems genetic analysis of inversion polymorphisms in the malaria mosquito. *Proc. Natl. Acad. Sci. USA* **115**, E7005–E7014 (2018).
16. Waterhouse, R. M. *et al.* Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* **316**, 1738–1743 (2007).
17. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
18. Proteins, C. C. One thousand families for the molecular biologist. *Nature* **357**, 543–544 (1992).
19. Srinivasan, S. *Homology Folding of Proteins: Application to Cytokine Engineering* (Springer, 1998).
20. Ghosh, C. *et al.* Enrichment of phenotype among biological forms of *Anopheles stephensi* Liston through establishment of isofemale lines. *bioRxiv.* https://doi.org/10.1101/2022.09.28.509862 (2022).
21. Jiang, X. *et al.* Genome analysis of a major urban malaria vector mosquito, *Anopheles stephensi*. *Genome Biol.* **15**, 459 (2014).
22. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive -mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
23. Chin, C.-S. *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods.* **13**, 1050–1054 (2016).
24. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
25. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods.* **17**, 155–158 (2020).
26. Chakraborty, M., Baldwin-Brown, J. G., Long, A. D. & Emerson, J. J. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res.* **44**, e147 (2016).
27. Sharakhova, M. V. *et al.* Update of the *Anopheles gambiae* PEST genome assembly. *Genome Biol.* **8**, R5 (2007).
28. Corbett-Detig, R. B. *et al.* Fine-mapping complex inversion breakpoints and investigating somatic pairing in the *Anopheles gambiae* species complex using proximity-ligation sequencing. *Genetics* https://doi.org/10.1101/662114 (2019).
29. Wood, O., Hanrahan, S., Coetzee, M., Koekemoer, L. & Brooke, B. Cuticle thickening associated with pyrethroid resistance in the major malaria vector *Anopheles funestus*. *Parasit. Vectors.* **3**, 67 (2010).
30. Alout, H. *et al.* Insecticide exposure impacts vector–parasite interactions in insecticide-resistant malaria vectors. *Proc. R. Soc. B Biol. Sci.* https://doi.org/10.1098/rspb.2014.0389 (2014).
31. Baxter, R. H. G. *et al.* A heterodimeric complex of the LRR proteins LRIM1 and APL1C regulates complement-like immunity in *Anopheles gambiae*. *Proc. Natl. Acad. Sci. USA* **107**, 16817–16822 (2010).
32. Inbar, E. *et al.* Knockout of *Anopheles stephensi* immune gene LRIM1 by CRISPR-Cas9 reveals its unexpected role in reproduction and vector competence. *PLoS Pathog.* **17**, e1009770 (2021).

33. Lombardo, F. & Christophides, G. K. Novel factors of *Anopheles gambiae* haemocyte immune response to *Plasmodium berghei* infection. *Parasit. Vectors* https://doi.org/10.1186/s13071-016-1359-y (2016).
34. Smith, R. C., Eappen, A. G., Radtke, A. J. & Jacobs-Lorena, M. Regulation of anti-plasmodium immunity by a LITAF-like transcription factor in the malaria vector *Anopheles gambiae*. *PLoS Pathogens.* https://doi.org/10.1371/journal.ppat.1002965 (2012).
35. Bushell, K. N. *et al.* LITAF mediation of increased TNF-α secretion from inflamed colonic lamina propria macrophages. *PLoS ONE.* https://doi.org/10.1371/journal.pone.0025849 (2011).
36. Palmerini, V. *et al.* Drosophila TNFRs Grindelwald and Wengen bind Eiger with different affinities and promote distinct cellular functions. *Nat. Commun.* https://doi.org/10.1038/s41467-021-22080-9 (2021).
37. Igaki, T. *et al.* Eiger, a TNF superfamily ligand that triggers the Drosophila JNK pathway. *EMBO J.* **21**, 3009–3018 (2002).
38. Kanda, H., Igaki, T., Kanuka, H., Yagi, T. & Miura, M. Wengen, a member of the Drosophila tumor necrosis factor receptor superfamily, is required for Eiger signaling. *J. Biol. Chem.* **277**, 28372–28375 (2002).
39. Andersen, D. S. *et al.* The Drosophila TNF receptor Grindelwald couples loss of cell polarity and neoplastic growth. *Nature* **522**, 482–486 (2015).
40. Banner, D. W. *et al.* Crystal structure of the soluble human 55 kd TNF receptor-human TNF beta complex: implications for TNF receptor activation. *Cell* **73**, 431–445 (1993).
41. Black, R. A. Tumor necrosis factor-alpha converting enzyme. *Int. J. Biochem. Cell Biol.* **34**, 1–5 (2002).
42. Mohan, M. J. *et al.* The tumor necrosis factor-alpha converting enzyme (TACE): A unique metalloproteinase with highly defined substrate selectivity. *Biochemistry* **41**, 9462–9469 (2002).
43. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
44. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
45. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
46. Al-Lamki, R. S. & Mayadas, T. N. TNF receptors: Signaling pathways and contribution to renal dysfunction. *Kidney Int.* **87**, 281–296 (2015).
47. Hoareau, L. *et al.* Signaling pathways involved in LPS induced TNFalpha production in human adipocytes. *J. Inflamm.* **7**, 1 (2010).
48. Lopez, W. *et al.* Analysis of immune-related genes during Nora virus infection of using next generation sequencing. *AIMS Microbiol.* **4**, 123–139 (2018).
49. Moreno, E., Yan, M. & Basler, K. Evolution of TNF signaling mechanisms: JNK-dependent apoptosis triggered by Eiger, the Drosophila homolog of the TNF superfamily. *Curr. Biol.* **12**, 1263–1268 (2002).
50. Garver, L. S., de Almeida, O. G. & Barillas-Mury, C. The JNK pathway is a key mediator of *Anopheles gambiae* antiplasmodial immunity. *PLoS Pathog.* **9**, e1003622 (2013).
51. Souvannaseng, L. *et al.* Inhibition of JNK signaling in the Asian malaria vector *Anopheles stephensi* extends mosquito longevity and improves resistance to *Plasmodium falciparum* infection. *PLoS Pathog.* **14**, e1007418 (2018).
52. Cruz, L. N., Wu, Y., Ulrich, H., Craig, A. G. & Garcia, C. R. S. Tumor necrosis factor reduces *Plasmodium falciparum* growth and activates calcium signaling in human malaria parasites. *Biochim. Biophys. Acta.* **1860**, 1489–1497 (2016).
53. Cheng, S., Seven, A. B., Wang, J., Skiniotis, G. & Özkan, E. Conformational plasticity in the transsynaptic neurexin-cerebellin-glutamate receptor adhesion complex. *Structure.* **24**, 2163–2173 (2016).
54. Ghosh, S. K., & Ghosh, C. *New Ways to Tackle Malaria* (IntechOpen, 2019). https://doi.org/10.5772/intechopen.89467. Available from: https://www.intechopen.com/online-first/new-ways-to-tackle-malaria.
55. Ghosh, C. & Shetty, N. J. Mode of inheritance of fenitrothion resistance in *Anopheles stephensi* Liston. *J. Cytol. Genet.* **342**, 141–146 (1999).
56. Sharakhova, M. V. *et al.* A physical map for an Asian malaria mosquito, *Anopheles stephensi*. *Am. J. Trop. Med. Hyg.* **83**, 1023–1027 (2010).
57. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
58. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinform.* **19**, 460 (2018).
59. Servant, N. *et al.* HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
60. Durand, N. C. *et al.* Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
61. Stanke, M. & Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).

Acknowledgements

The authors thank Nucleome Informatics Pvt. Ltd. for generating Pacbio reads and Illumina reads for the IndInt strain. The authors thank Tata Institute for Genetics and Society (TIGS) India for funding researchers involved in this work and for sequencing and the Government of Karnataka for funding the computational infrastructure at IBAB. The authors would like to acknowledge one of the authors Ms. Apoorva Ganesh for drawing figure 4 using the Biorender tool https://app.biorender.com/biorender-templates.

Author contributions

S.Sri. for overseeing bioinformatics work, setting the tone and writing of the manuscript. S.D. for writing the first draft, taming the assembly and discovery of 3L*i* inversion. C.G. for creating the IndInt isofemale line and generating photomaps. SSingh for contig level assemblies using various methods. A.T. for characterizing candidate genes. A.G. characterizing breakpoints and TNF-TNFR pathway work. H.M., A.J., M.K., A.C., R.B., V.S. for deep annotation of 3L*i* region using RoseTTaFold prediction. S.M. Insectary work. N.K., S.K. insectary work and for vectorial capacity in both standard and heterozygous 3L*i* inverted individuals. S.Swa. for overseeing the entire insect work. S.S. for conceptualization and contributions to the manuscript.

Funding

This research is funded by internal grants from both the Institute of Bioinformatics and Applied Biotechnology and the Tata Institute for Genetics and Society in Bangalore, India.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-23780-y.

Correspondence and requests for materials should be addressed to S.S. or S.S.

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

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