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

Subhashini Srinivasan[1,*], Chaitali Ghosh[2!], Shrestha Das[1!], Aditi Thakare[1!], 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,*], and Suresh Subramani[3]

[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

[!] these authors contributed equally to this paper

* sunita.swain@tigs.res.in and ssubha@ibab.ac.in (Corresponding Authors)

Keywords: *TNFSF, TNFRSF, ADAM17, cerebellin, LITAFs, TNF-alpha signaling pathway, eiger, wengen, grindelwald, 2La inversion in An. gambiae, IndInt, IndCh, UCI strain*

ABSTRACT

Identification of *Plasmodium*-resistance genes in malaria vectors remains an elusive goal despite the recent availability of high-quality genomes of several mosquito vectors. *An. 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, have been 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 IndInt assembly using the state-of-the-art protein-fold prediction and other annotation tools reveals the presence of a TNF-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 in IndInt, 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 TNF-TNFR system, suggesting the presence of a functionally active eiger signaling pathway present in IndInt. Comparative genomics of high-quality genome assemblies of the three type-forms with that of IndInt, reveals structurally disruptive mutations in eiger gene in all three strains of the type-form, alluding to compromised innate immunity in the type-form as the cause of high vectorial capacity in these strains. This is

the first report of the presence of an intact evolutionarily-conserved TNF-TNFR signaling system in malaria vectors, with a potential role in *Plasmodium* resistance.

INTRODUCTION

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). 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)[1] [2]. These forms are identifiable based on the number of ridges on eggs[3]. The genomes of three strains of the type-form have been reported recently[4][5][6]. 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.

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 based on cytobands[7][8]. Moreover, each of these inversions are reported to impart varying resistance to insecticides[9]. 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[7]. Based on these reports and studies

3

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[10].

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[10], the 2L*a* inversion*,* encompassing 21 Mbp, is associated with resistance to *Plasmodium* and desiccation[11][12]. 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[13]. 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[14].

Comparative proteomics of *D. melanogaster, Ae. aegypti, and An. gambiae* revealed 91 gene trios with a total of 589 genes comprising the immune repertoire[15]. 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[4]* (https://github.com/mahulchak/stephensi_genome/blob/master/immune.gff), 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[16]*. 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[17]. This is especially important in the identification of cytokines and lymphokine homologs, which are notorious for exhibiting low or undetectable homology at the sequence level (https://www.amazon.com/Homology-Folding-Proteins-Biotechnology-Intelligence/dp/3540636056). For example, members of TNF and TNFR superfamilies, recently identified and characterized in *Drosophila*, are missing in the collation of genes implicated in innate immunity in flies and in mosquito vectors[4][15].

In order to understand the evolution of *Plasmodium* resistance in *An. stephensi*, we have sequenced and assembled the genome of an isofemale line from a strain of intermediate-form from Bangalore (IndInt) displaying diminished susceptibility to *Plasmodium*. We present the assembly of this strain and report an inversion not found in the genomes of the three strains of the type-form[4][5][6][18]. 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 Iso-female lines from a parent T6 lab-colony of the intermediate-form of *An. stephensi* (IndInt) were grown, both to achieve homogenity required for high-quality assembly from sequencing of pooled DNA from multiple insects and to enrich for *Plasmodium* resistant phenotype. 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 was obtained from 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.

	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.
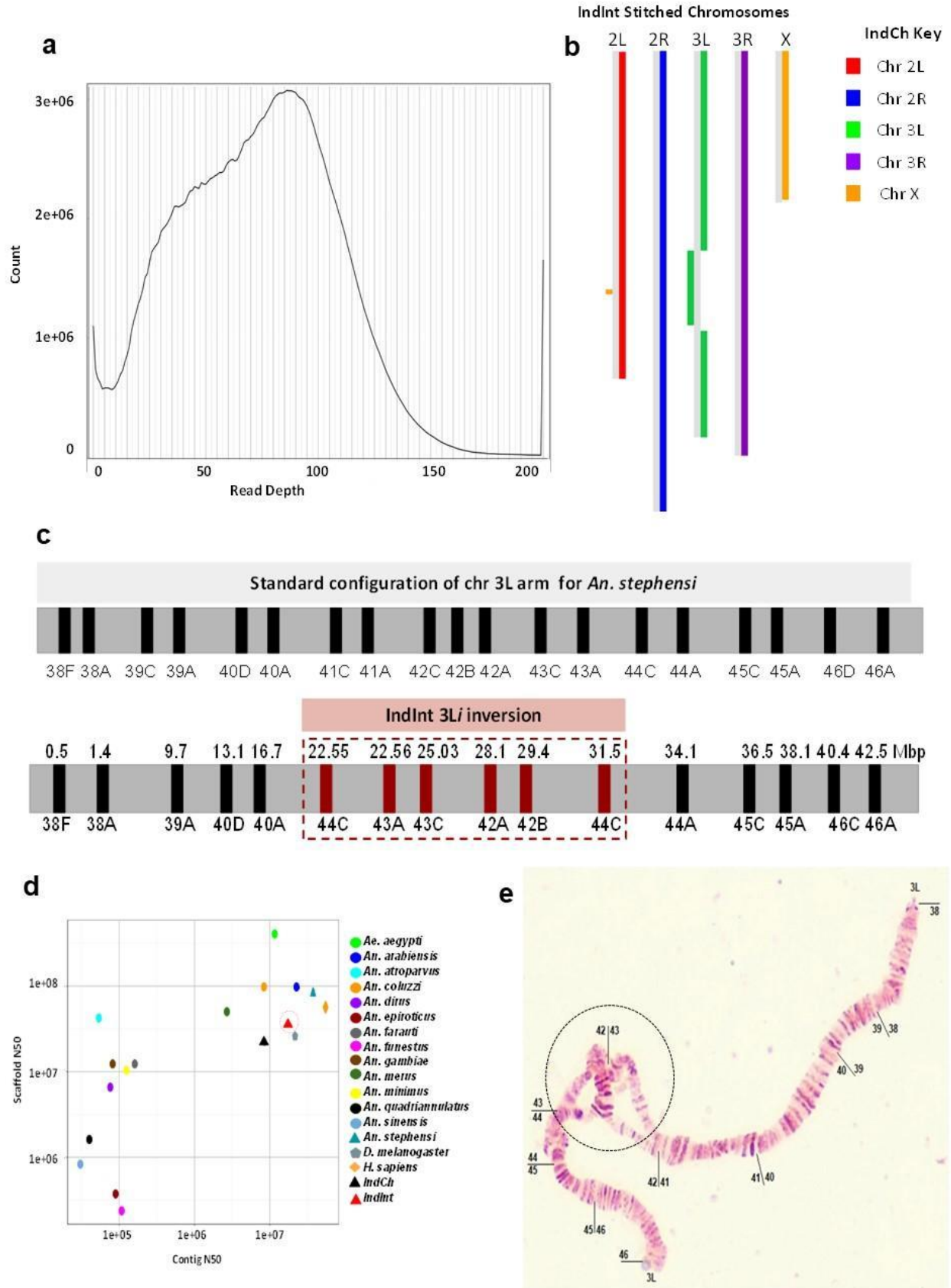
To improve the contiguity of the assembly, the four different contig-level assemblies from CANU, FALCON, FLYE, and WTDBG2[19][20][21][22] were merged in different combinations using the tool Quickmerge[23]. Assembly statistics were computed for each of the merged combinations to determine the best merged 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 Mbp 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[20]. 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. Fig. 1a with a hump indicates the presence of potential inversion heterozygosity in the genome of IndInt.

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 2Rb (2R^{+b}/2R^{+b})

7

inversion similar to the IndCh strain. However, in the 3L arm of chromosome 3 IndInt shows a 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 inverted in the assembly, which matched with a reported breakpoint for the 3Li inversion in *An. stephensi[8]*. 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 3Li inversion (3L^{+i}/3Li), with the rest (52%) being homozygous for the standard form of 3Li *(*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 3Li (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.

a

b IndInt Stitched Chromosomes

IndCh Key
- Chr 2L
- Chr 2R
- Chr 3L
- Chr 3R
- Chr X

c

Standard configuration of chr 3L arm for *An. stephensi*

38F 38A 39C 39A 40D 40A 41C 41A 42C 42B 42A 43C 43A 44C 44A 45C 45A 46D 46A

IndInt 3L*i* inversion

0.5 1.4 9.7 13.1 16.7 22.55 22.56 25.03 28.1 29.4 31.5 34.1 36.5 38.1 40.4 42.5 Mbp

38F 38A 39A 40D 40A 44C 43A 43C 42A 42B 44C 44A 45C 45A 46C 46A

d

- Ae. aegypti
- An. arabiensis
- An. atroparvus
- An. coluzzi
- An. dirus
- An. epiroticus
- An. farauti
- An. funestus
- An. gambiae
- An. merus
- An. minimus
- An. quadriannulatus
- An. sinensis
- An. stephensi
- D. melanogaster
- H. sapiens
- IndCh
- IndInt

e

9

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 3Li 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.

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[5]. 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 ($3L^{+i}/3L^{+i}$) configuration, 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.
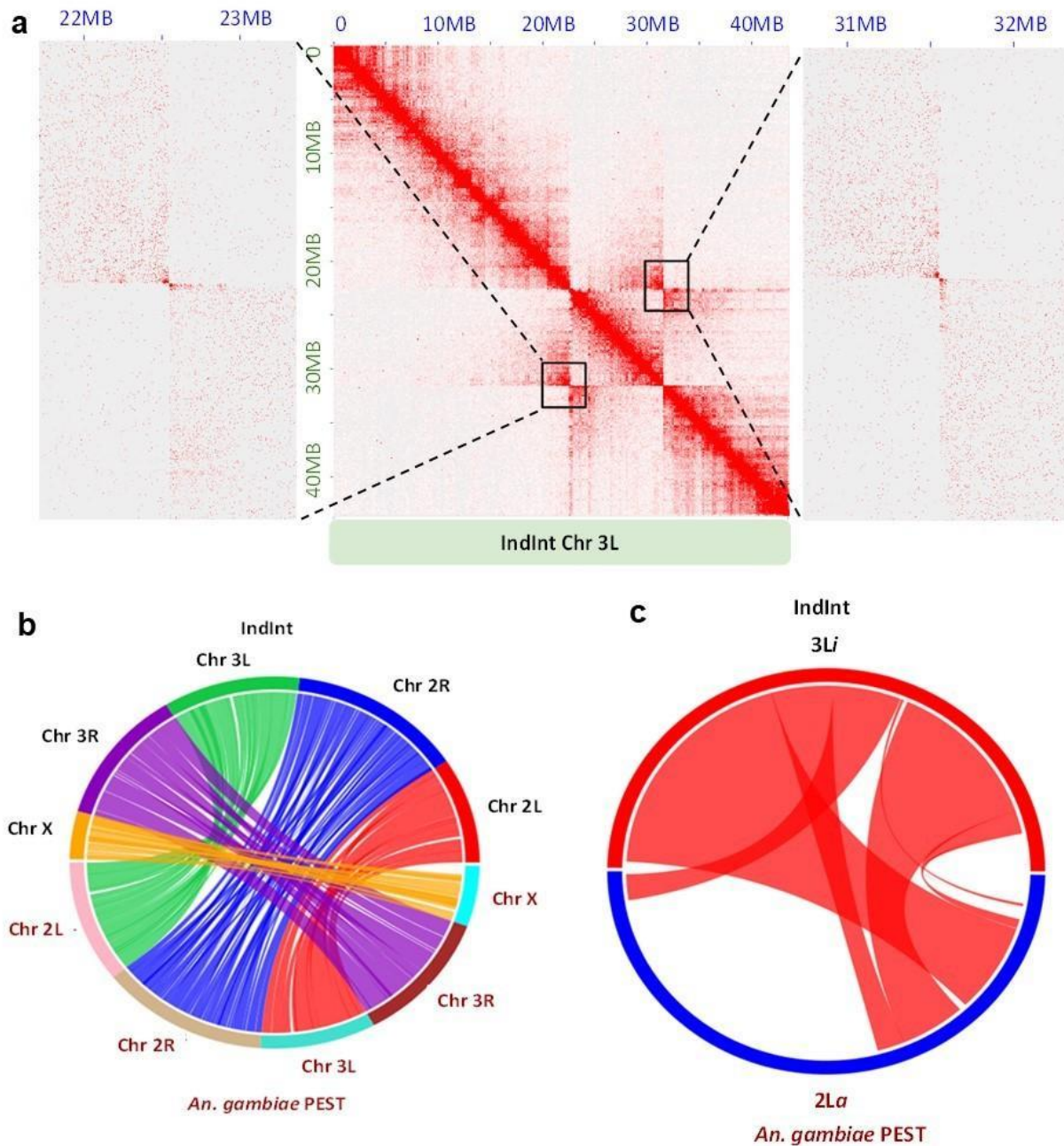
Figure 2: *(a) Proximal and distal breakpoints obtained by mapping UCI HiC reads against the chromosome 3L of IndInt. The black boxes represent the centromere-proximal (left) and centromere-distal (right) breakpoints located between the ranges of bp 22,545,000 to 22,550,000 and bp 31,555,000 to 31,560,000, 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 2La and the IndInt 3Li regions.

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[24]. Fig. 2b shows that chromosome X, 2R, and 3R of *An. gambiae* is 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[5]. Fig. 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[25]. 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[11][12], by analogy, one can 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 ($3L^{+i}/3L i$), suggesting the potential role of the 3L*i* inversion in *Plasmodium* sensitivity.

Functional characterization of genes within the 3L*i* region

13

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* species[26], while also resisting *Plasmodium* infection[27]. 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 reported to be implicated in immunity in insects [15]. 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* [15] Immune DB ID	Function
g22033.t1	22194979	22196095	AGAP005552	Peptidoglycan recognition protein
g22086.t1	22559888	22560559	AGAP007020	Thioredoxin peroxidase 5
g22107.t1	22657471	22661294	AGAP006939	Armitage protein
g22121.t1	22702737	22707661	AAEL001675	Clip domain, serine protease family A
g22273.t1	24373199	24379261	AGAP006631	Class A Scavenger Receptor (SRCR domain) with Serine Protease domain

14

g22562.t1	26855939	26862215	AGAP006914	Fibrinogen-related protein 1
g22567.t1	26922106	26923713	AGAP006911	Serine protease inhibitor - Serpin 2
g22568.t1	26925252	26932268	AGAP006909	Serine protease inhibitor - Serpin 1
g22799.t1	28434986	28435854	AGAP006790	Fibrinogen C-terminal domain-containing protein
g22840.t1	28809510	28812635	AGAP006761	3-glucan binding protein
g22865.t1	28987304	28992526	AGAP006747	IMD pathway signaling NF-kappaB Relish-like transcription factor
g22870.t1	29022137	29026993	AGAP006743	Unspecified product BLASTP : Similar to ficolin 1
g22921.t1	29370252	29370571	AGAP006722	Unspecified product BLASTP : Similar to Cecropin B1
g23080.t1	30464940	30470336	AGAP005672	Staphylococcal nuclease domain-containing protein 1
g23109.t1	30680813	30689983	AGAP005652	ATP-dependent RNA helicase DDX5/DBP2
g23151.t1	30998165	31003426	AGAP005625	Class A Scavenger Receptor (SRCR domain) with Serine Protease domain
g23220.t1	31575970	31576777	CPIJ004568	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. [15]. Highlighted in cyan is the only gene implicated in TNF-TNFR signaling pathway.

The genes implicated in *Plasmodium* resistance in *An. gambiae[28]* 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 [28]), 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[29]. 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[30][31].* 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 [32]. 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 with immune repertoire in *An. stephensi [4]*. 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(6ZT0 [33]). 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 assembly. Also, in the UCI and STE2 strains, the predicted eiger homologs are missing the critical N-terminal transmembrane region (Supplementary Fig. 6). In Fig. 3a, a 3D model of the eiger gene from IndInt, built using RoseTTaFold tool[16], is shown superimposed on the crystal structure of human TNF-TNFR complex (PDB: 1TNR,[34]) 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 [33]), along with a model of grnd

17

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.


```
c   TNF_1B:     SCEDSTYTQL----WNWVPECLSCGSRCSSDQV--ETQACTREQNRICT
    GRND-Steph: GCTRKCDRFE--YCDENSNACRNCSALCEKD-----EYSCYH----KCQ
    GRND-Fly:   DCHGTICHPVNEFCYVATERCHPCIEVCNNQTHNYDAFLCAK----ECS
                                    CXXCXXXC

    TNF_1A:     ECESGSFTASENHLRHCLSCSKCRKEMGQVEISSCTVDRDTVCG
    WGN-Steph:  ACESRKSW-WDPTLGECAPCLICADHQV-VLRPCQDYMNTVCG
    WGN-Fly:    PCAPQHWW--DSQRDRCTPCTRCQGEMI-PLRPCQLHTDTICG
                                CXXCXXC
```

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.

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[33]. 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[33]. 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 form in humans [35]. 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 [36].

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-TNFR signaling pathway from KEGG (Supplementary Text 2). Fig. 4 shows genes

19

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 Supplementray table 2. The homologs for genes marked by red and purple stars are disrupted and absent, respectively, in the IndCh genome, potentially disrupting eiger-mediated signaling pathway. Genes with blue stars are missing in all 4 assembled genomes, suggesting that these pathways may have evolved in vertebrates after the invertebrate-vertebrate split.

Gene Name	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

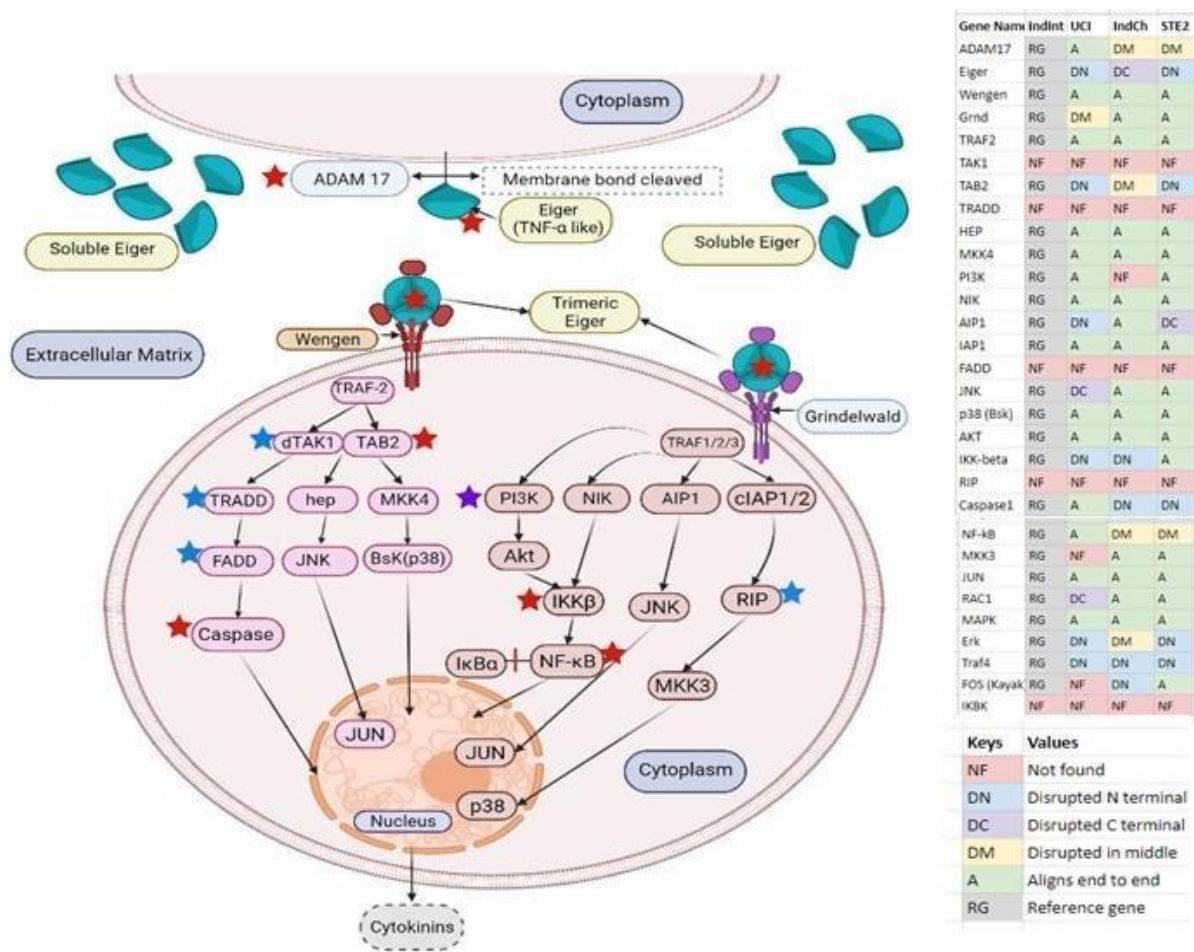| Keys	Values
NF	Not found
DN	Disrupted N terminal
DC	Disrupted C terminal
DM	Disrupted in middle
A	Aligns end to end
RG	Reference gene

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 (https://www.genome.jp/pathway/hsa04668) and validated using other sources [37][38]. 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. (Template for designing the pathway diagram was adapted from https://app.biorender.com/biorender-templates)

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[15] and, more specifically, missing from a list of 361 immune related genes in *An. stephensi* reported recently in the UCI strain[4].

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 (manuscript under preparation). A comparative analysis of this assembly with those of type-forms, displaying higher vectorial capacity, 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*, which is shown to be associated with *Plasmodium* sensitivity[11]; thus providing an opportunity to identify intact genes that may be offering resistance to *Plasmodium*. Here, for the first time, we report an intact TNF-TNFR system in malaria vectors with eiger, a member of the TNF superfamily, present within this inversion region (Figure 3b). Considering that the eiger gene in the IndCh strain harbors a frameshift mutation within the structurally folding domain, it is tempting to predict a role for this gene in the innate

21

immune response of *An. stephensi* against *Plasmodium* in IndInt strain. Similarly, the predicted eiger gene in STE2 and UCI strains are both disrupted in the N-terminal lacking transmembrane region and several genes in the eiger pathway are disrupted in 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[39]*, demands an exploration of the role of this complex in *Plasmodium* resistance in *An. stephensi* and in other malaria vectors, in general.

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 -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). We have found homologs of just about every gene required for eiger-based signaling similar to TNF-alpha in the IndInt assembly of *An. stephensi* (Fig. 4). Considering that many genes in 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 [15] (Supplementray Table 1).

It is known that human TNF-alpha reduces *Plasmodium falciparum* growth and activates calcium channels in human malarial parasites[40]. 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. It is well established that Ca^{+2} mediated signaling is correlated with the developmental stages of the parasite life cycle[41]. The eiger and its receptor homologs in *Drosophila* uses JNK pathway in a manner similar to human TNF-alpha. For example, in *Drosophila,* a grnd null mutant completely abolishes cell-death[33]. Furthermore, the eiger mediated cell-death in flies could be reversed by JNK inhibitors[33]. 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-TNFR system-based innate immunity may have evolved before this point of inflection. 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 3 fold 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 [42]. Since the TNF-TNFR system is conserved in all insects and vertebrates, missing/lost only in *nematodes*, innate immunity by the TNF-TNFR system may predate the invertebrate-vertebrate split and TNF-like fold evolution may even be older.

CONCLUSION

23

The three forms of *An. stephensi* (type, intermediate and mysorensis)*,* displaying varying vector competence, which are clearly identifiable by their number of egg ridges, offer an ideal system to discover genes implicated in *Plasmodium* resistance by interrogating genomes of strains with low vectorial capacity. Here, by assembling the genome of an intermediate-form (IndInt) and comparing it to that of a type-form (IndCh), we report a unique 3L*i* inversion in IndInt, which is syntenic to the 2L*a* inversion in *An. gambiae* previously implicated in *Plasmodium* resistance. This has allowed, for the first time, identification of TNF-alpha homolog, eiger, in malaria vectors along with homologs of genes required for downstream-signaling to offer an innate immune response. Most strikingly, IndCh strain, with high vectorial capacity for *Plasmodium*, harbors disruptive mutations in one or more genes of the eiger-grnd-wengen pathway. These observations strongly suggest that disruption of eiger-based signaling in the type-forms of *An. stephensi* may account for their high vectorial capacity predicated upon a reduction in innate immunity to combat *Plasmodium* infections.

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®)[43]. 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 to 48 h and transferred to the rearing trays (L39 x B30 cm; Polylab, Catalog no. 81702) with larval food.

Establishment of IndInt iso-female line of *An. stephensi*

About 10-15 gravid females were separated from the intermediate line[44]. 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 iso-female 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 iso-female 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 iso-female lines as per the method of Ghosh & Shetty, which was adapted from our recent publication[5]. 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[45].

Assembly

PacBio reads from the IndInt strain were assembled independently using FLYE[21] and WTDBG2[22] assemblers. The resulting assemblies were combined using Quickmerge[23] 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 100x 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 (DOI: 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[46] tool. DNA physical marker sequences for each chromosome arm published by Jiang *et al[18].* 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[20] 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[5].

Purge haplotigs

The tool 'Purge Haplotigs'[47] 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*i* breakpoints in IndInt

The tool HiC-Pro[48] 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[49]. 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)[50], 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[5].

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.

AUTHOR CONTRIBUTION

SSri for overseeing bioinformatics work, setting the tone and writing of the manuscript

SD for writing the first draft, taming the assembly and discovery of 3L*i* inversion

CG for creating the IndInt isofemale line and generating photomaps

SSingh for contig level assemblies using various methods

AT for characterizing candidate genes

AG characterizing breakpoints and TNF-TNFR pathway work

HM, AJ, MK, AC, RB, VS for deep annotation of 3L*i* region using RoseTTaFold prediction

SM Insectary work

NK, SK insectary work and for vectorial capacity in both standard and heterozygous 3L*i* inverted individuals

SSwa for overseeing the entire insect work

SS for conceptualization and contributions to the manuscript

ACKNOWLEDGEMENT

for funding researchers involved in this work and for sequencing and the Government of Karnataka for funding the computational infrastructure at IBAB.

FUNDING

REFERENCES

1.   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. 2013;8: e63753. doi:10.1371/journal.pone.0063753

2.   Basseri HR, Mohamadzadeh Hajipirloo H, Mohammadi Bavani M, Whitten MMA. Comparative susceptibility of different biological forms of Anopheles stephensi to Plasmodium berghei ANKA strain. PLoS One. 2013;8: e75413. doi:10.1371/journal.pone.0075413

3.   Chavshin AR, Oshaghi MA, Vatandoost H, Hanafi-Bojd AA, Raeisi A, Nikpoor F. Molecular characterization, biological forms and sporozoite rate of Anopheles stephensi in southern Iran. Asian Pac J Trop Biomed. 2014;4: 47–51. doi:10.1016/S2221-1691(14)60207-0

4.   Chakraborty M, Ramaiah A, Adolfi A, Halas P, Kaduskar B, Ngo LT, et al. Author

Correction: Hidden genomic features of an invasive malaria vector, Anopheles stephensi, revealed by a chromosome-level genome assembly. BMC Biol. 2022;20: 96. doi:10.1186/s12915-022-01314-2

5.   Thakare A, Ghosh C, Alalamath T, Kumar N, Narang H, Whadgar S, 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. 2022;12: 3610. doi:10.1038/s41598-022-07462-3

6.   Chida AR, Ravi S, Jayaprasad S, Paul K, Saha J, Suresh C, et al. A Near-Chromosome Level Genome Assembly of. Front Genet. 2020;11: 565626. doi:10.3389/fgene.2020.565626

7.   Mahmood F, Sakai RK. Inversion polymorphisms in natural populations of Anopheles stephensi. Can J Genet Cytol. 1984;26: 538–546. doi:10.1139/g84-086

8.   Kamali M, Sharakhova MV, Baricheva E, Karagodin D, Tu Z, Sharakhov IV. An integrated chromosome map of microsatellite markers and inversion breakpoints for an Asian malaria mosquito, Anopheles stephensi. J Hered. 2011;102: 719–726. doi:10.1093/jhered/esr072

9.   Shetty NJ, Hariprasad TPN, Sanil D, Zin T. Chromosomal inversions among insecticide-resistant strains of Anopheles stephensi Liston, a malaria mosquito. Parasitol Res. 2013;112: 3851–3857. doi:10.1007/s00436-013-3575-0

10.   Ayala D, Ullastres A, González J. Adaptation through chromosomal inversions in Anopheles. Front Genet. 2014;5: 129. doi:10.3389/fgene.2014.00129

11.   Riehle MM, Bukhari T, Gneme A, Guelbeogo WM, Coulibaly B, Fofana A, et al. The 2La chromosome inversion is associated with susceptibility to in Africa. Elife. 2017;6. doi:10.7554/eLife.25813

12.   Gray EM, Rocca KAC, Costantini C, Besansky NJ. Inversion 2La is associated with enhanced desiccation resistance in Anopheles gambiae. Malar J. 2009;8: 215. doi:10.1186/1475-2875-8-215

13.   Sharakhov IV, White BJ, Sharakhova MV, Kayondo J, Lobo NF, Santolamazza F, et al. Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the Anopheles gambiae complex. Proc Natl Acad Sci U S A. 2006;103: 6258–6262. doi:10.1073/pnas.0509683103

14.   Cheng C, Tan JC, Hahn MW, Besansky NJ. Systems genetic analysis of inversion polymorphisms in the malaria mosquito. Proc Natl Acad Sci U S A. 2018;115: E7005–E7014. doi:10.1073/pnas.1806760115

15.   Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, et al. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science. 2007;316: 1738–1743. doi:10.1126/science.1139862

16.   Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373: 871–876. doi:10.1126/science.abj8754

17.   Chothia C. Proteins. One thousand families for the molecular biologist. Nature. 1992;357: 543–544. doi:10.1038/357543a0

18.   Jiang X, Peery A, Hall AB, Sharma A, Chen X-G, Waterhouse RM, et al. Genome analysis of a major urban malaria vector mosquito, Anopheles stephensi. Genome Biol. 2014;15: 459. doi:10.1186/s13059-014-0459-2

19.   Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive -mer weighting and repeat separation. Genome Res. 2017;27: 722–736. doi:10.1101/gr.215087.116

20.   Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods. 2016;13: 1050–1054. doi:10.1038/nmeth.4035

21.   Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37: 540–546. doi:10.1038/s41587-019-0072-8

22.   Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 2020;17: 155–158. doi:10.1038/s41592-019-0669-3

23.   Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res. 2016;44: e147. doi:10.1093/nar/gkw654

24.   Sharakhova MV, Hammond MP, Lobo NF, Krzywinski J, Unger MF, Hillenmeyer ME, et al. Update of the Anopheles gambiae PEST genome assembly. Genome Biol. 2007;8: R5. doi:10.1186/gb-2007-8-1-r5

25.   Corbett-Detig RB, Said I, Calzetta M, Genetti M, McBroome J, Maurer NW, et al.

Fine-mapping complex inversion breakpoints and investigating somatic pairing in the *Anopheles gambiae* species complex using proximity-ligation sequencing. doi:10.1101/662114

26.   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. 2010;3: 67. doi:10.1186/1756-3305-3-67

27.   Alout H, Djègbè I, Chandre F, Djogbénou LS, Dabiré RK, Corbel V, et al. Insecticide exposure impacts vector–parasite interactions in insecticide-resistant malaria vectors. Proceedings of the Royal Society B: Biological Sciences. 2014. p. 20140389. doi:10.1098/rspb.2014.0389

28.   Baxter RHG, Steinert S, Chelliah Y, Volohonsky G, Levashina EA, Deisenhofer J. A heterodimeric complex of the LRR proteins LRIM1 and APL1C regulates complement-like immunity in Anopheles gambiae. Proc Natl Acad Sci U S A. 2010;107: 16817–16822. doi:10.1073/pnas.1010575107

29.   Inbar E, Eappen AG, Alford RT, Reid W, Harrell RA, Hosseini M, et al. Knockout of Anopheles stephensi immune gene LRIM1 by CRISPR-Cas9 reveals its unexpected role in reproduction and vector competence. PLoS Pathog. 2021;17: e1009770. doi:10.1371/journal.ppat.1009770
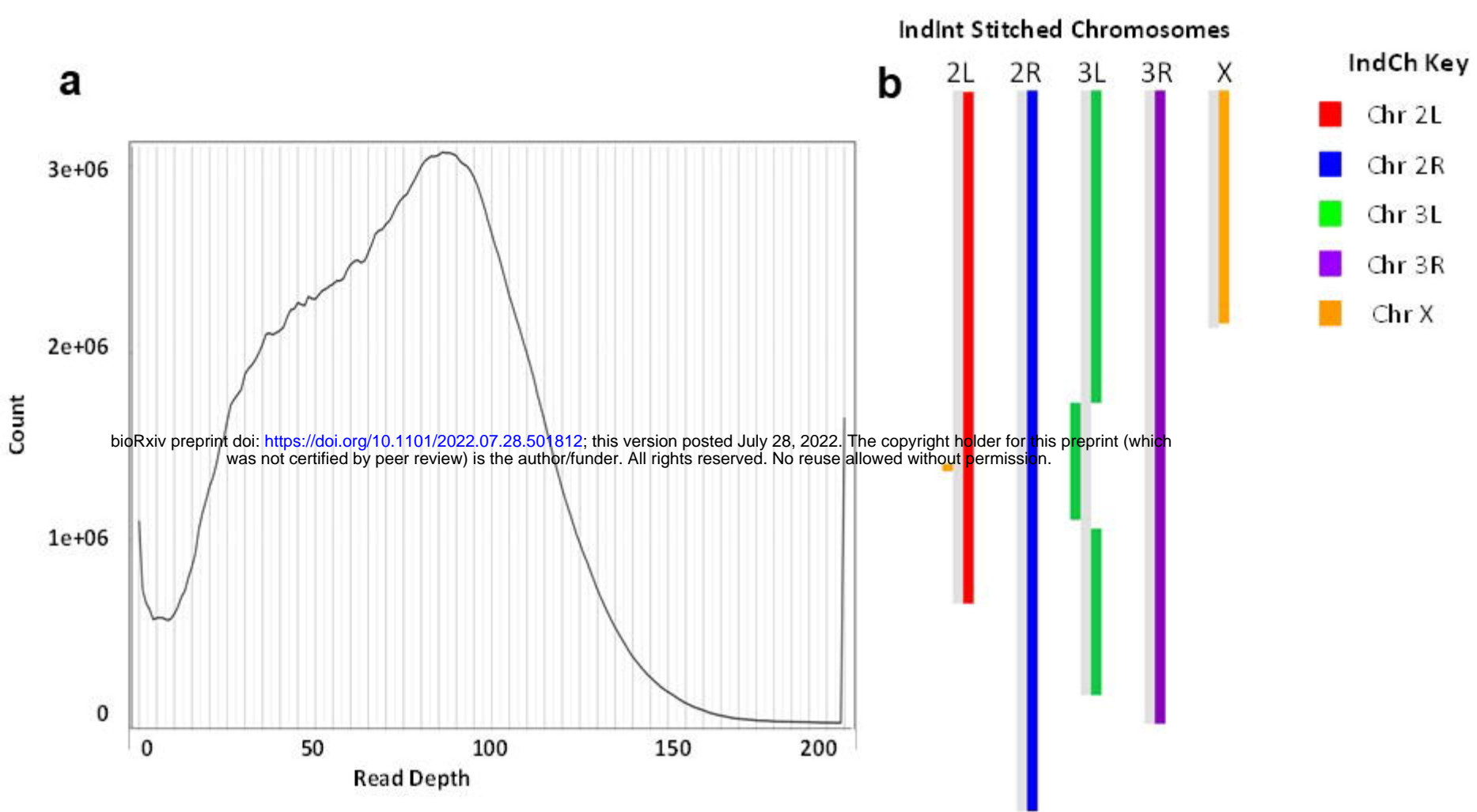
30.   Lombardo F, Christophides GK. Novel factors of Anopheles gambiae haemocyte immune response to Plasmodium berghei infection. Parasites & Vectors. 2016. doi:10.1186/s13071-016-1359-y

31.   Smith RC, Eappen AG, Radtke AJ, Jacobs-Lorena M. Regulation of Anti-Plasmodium Immunity by a LITAF-like Transcription Factor in the Malaria Vector Anopheles gambiae. PLoS Pathogens. 2012. p. e1002965. doi:10.1371/journal.ppat.1002965

32.   Bushell KN, Leeman SE, Gillespie E, Gower AC, Reed KL, Stucchi AF, et al. LITAF Mediation of Increased TNF-α Secretion from Inflamed Colonic Lamina Propria Macrophages. PLoS ONE. 2011. p. e25849. doi:10.1371/journal.pone.0025849

33.   Palmerini V, Monzani S, Laurichesse Q, Loudhaief R, Mari S, Cecatiello V, et al. Drosophila TNFRs Grindelwald and Wengen bind Eiger with different affinities and promote distinct cellular functions. Nature Communications. 2021. doi:10.1038/s41467-021-22080-9

34.   Banner DW, D'Arcy A, Janes W, Gentz R, Schoenfeld HJ, Broger C, et al. Crystal structure of the soluble human 55 kd TNF receptor-human TNF beta complex: implications for TNF receptor activation. Cell. 1993;73: 431–445. doi:10.1016/0092-8674(93)90132-a

35.   Black RA. Tumor necrosis factor-alpha converting enzyme. Int J Biochem Cell Biol. 2002;34: 1–5. doi:10.1016/s1357-2725(01)00097-8

36.   Mohan MJ, Seaton T, Mitchell J, Howe A, Blackburn K, Burkhart W, et al. The tumor necrosis factor-alpha converting enzyme (TACE): a unique metalloproteinase with highly defined substrate selectivity. Biochemistry. 2002;41: 9462–9469. doi:10.1021/bi0260132

37.   Al-Lamki RS, Mayadas TN. TNF receptors: signaling pathways and contribution to renal dysfunction. Kidney Int. 2015;87: 281–296. doi:10.1038/ki.2014.285

38.   Hoareau L, Bencharif K, Rondeau P, Murumalla R, Ravanan P, Tallet F, et al. Signaling pathways involved in LPS induced TNFalpha production in human adipocytes. J Inflamm . 2010;7: 1. doi:10.1186/1476-9255-7-1

39.   Lopez W, Page AM, Carlson DJ, Ericson BL, Cserhati MF, Guda C, et al. Analysis of immune-related genes during Nora virus infection of using next generation sequencing. AIMS Microbiol. 2018;4: 123–139. doi:10.3934/microbiol.2018.1.123

40.   Cruz LN, Wu Y, Ulrich H, Craig AG, Garcia CRS. Tumor necrosis factor reduces Plasmodium falciparum growth and activates calcium signaling in human malaria parasites. Biochim Biophys Acta. 2016;1860: 1489–1497. doi:10.1016/j.bbagen.2016.04.003

41.   Oliveira LS de, de Oliveira LS, Alborghetti MR, Carneiro RG, Bastos IMD, Amino R, et al. Calcium in the Backstage of Malaria Parasite Biology. Frontiers in Cellular and Infection Microbiology. 2021. doi:10.3389/fcimb.2021.708834

42.   Cheng S, Seven AB, Wang J, Skiniotis G, Özkan E. Conformational Plasticity in the Transsynaptic Neurexin-Cerebellin-Glutamate Receptor Adhesion Complex. Structure. 2016;24: 2163–2173. doi:10.1016/j.str.2016.11.004

43.   Ghosh SK, Ghosh C (2019). New ways to tackle malaria. IntechOpen, DOI:10.5772/intechopen.89467. Available from: https://www.intechopen.com/online-first/new-ways-to-tackle-malaria

44.   Ghosh C, Shetty NJ. Mode of inheritance of fenitrothion resistance in Anopheles stephensi Liston. Journal of Cytology and Genetics July-December. 1999;342: 141–146. Available: https://eurekamag.com/research/021/355/021355564.php

45.   Sharakhova MV, Xia A, Tu Z, Shouche YS, Unger MF, Sharakhov IV. A physical map for an Asian malaria mosquito, Anopheles stephensi. Am J Trop Med Hyg. 2010;83: 1023–1027. doi:10.4269/ajtmh.2010.10-0366

46.   Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27: 578–579. doi:10.1093/bioinformatics/btq683

47.   Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018;19: 460. doi:10.1186/s12859-018-2485-7

48.   Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 2015;16: 259. doi:10.1186/s13059-015-0831-x

49.   Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. 2016;3: 99–101. doi:10.1016/j.cels.2015.07.012

50.   Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005;33: W465–7. doi:10.1093/nar/gki458
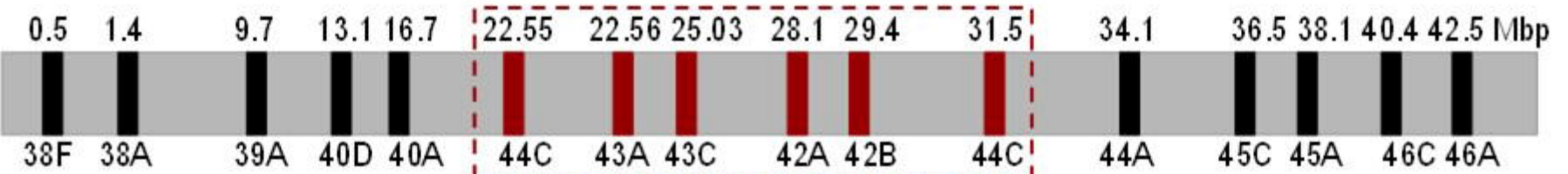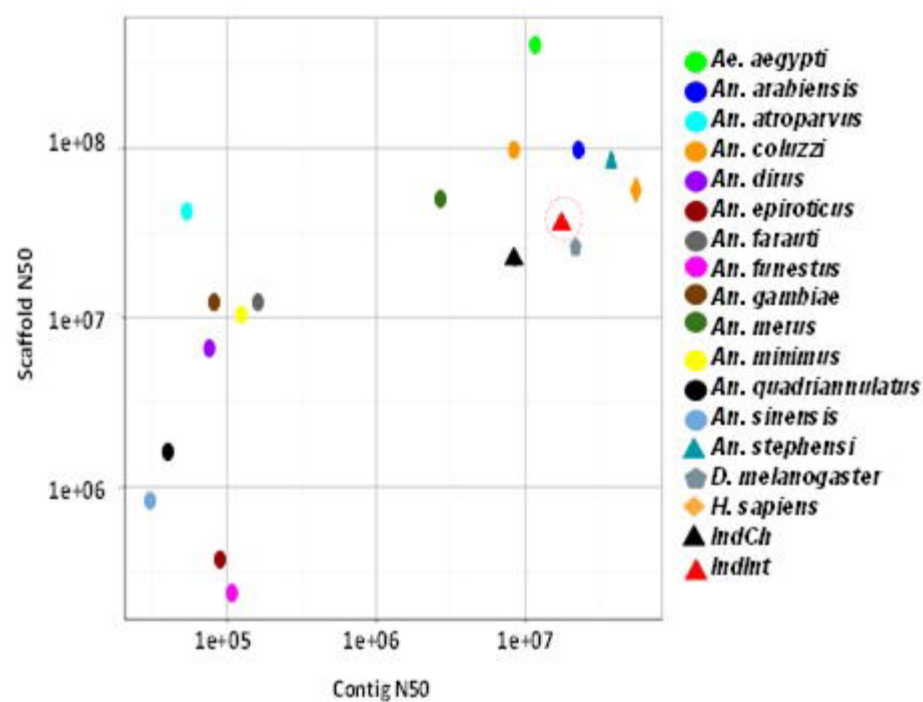
a

b

IndInt Stitched Chromosomes

IndCh Key

- Chr 2L
- Chr 2R
- Chr 3L
- Chr 3R
- Chr X

c

Standard configuration of chr 3L arm for *An. stephensi*

38F 38A 39C 39A 40D 40A 41C 41A 42C 42B 42A 43C 43A 44C 44A 45C 45A 46D 46A

IndInt 3L*i* inversion

0.5 1.4 9.7 13.1 16.7 22.55 22.56 25.03 28.1 29.4 31.5 34.1 36.5 38.1 40.4 42.5 Mbp

38F 38A 39A 40D 40A 44C 43A 43C 42A 42B 44C 44A 45C 45A 46C 46A

d

- Ae. aegypti
- An. arabiensis
- An. atroparvus
- An. coluzzi
- An. dirus
- An. epiroticus
- An. farauti
- An. funestus
- An. gambiae
- An. merus
- An. minimus
- An. quadriannulatus
- An. sinensis
- An. stephensi
- D. melanogaster
- H. sapiens
- IndCh
- IndInt

e

a

22MB　23MB　　0　10MB　20MB　30MB　40MB　　31MB　32MB

0

10MB

20MB

30MB

40MB

IndInt Chr 3L

b

IndInt

Chr 3L

Chr 3R

Chr X

Chr 2L

Chr 2R

Chr 3L

An. gambiae PEST

Chr 2R

Chr 2L

Chr X

Chr 3R

c

IndInt
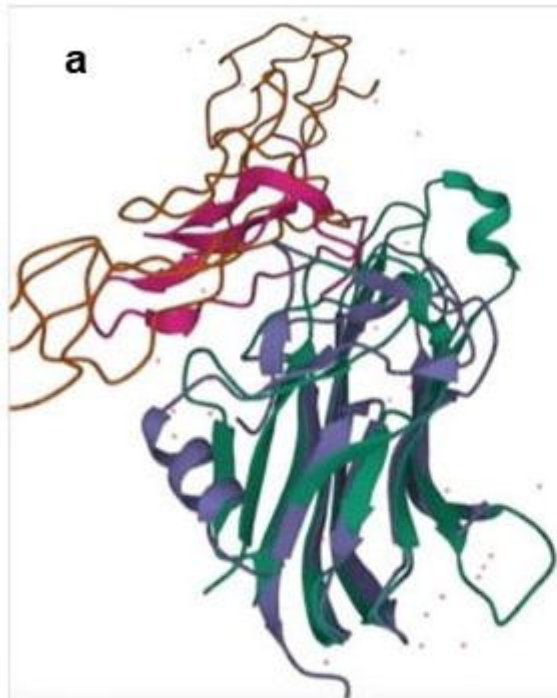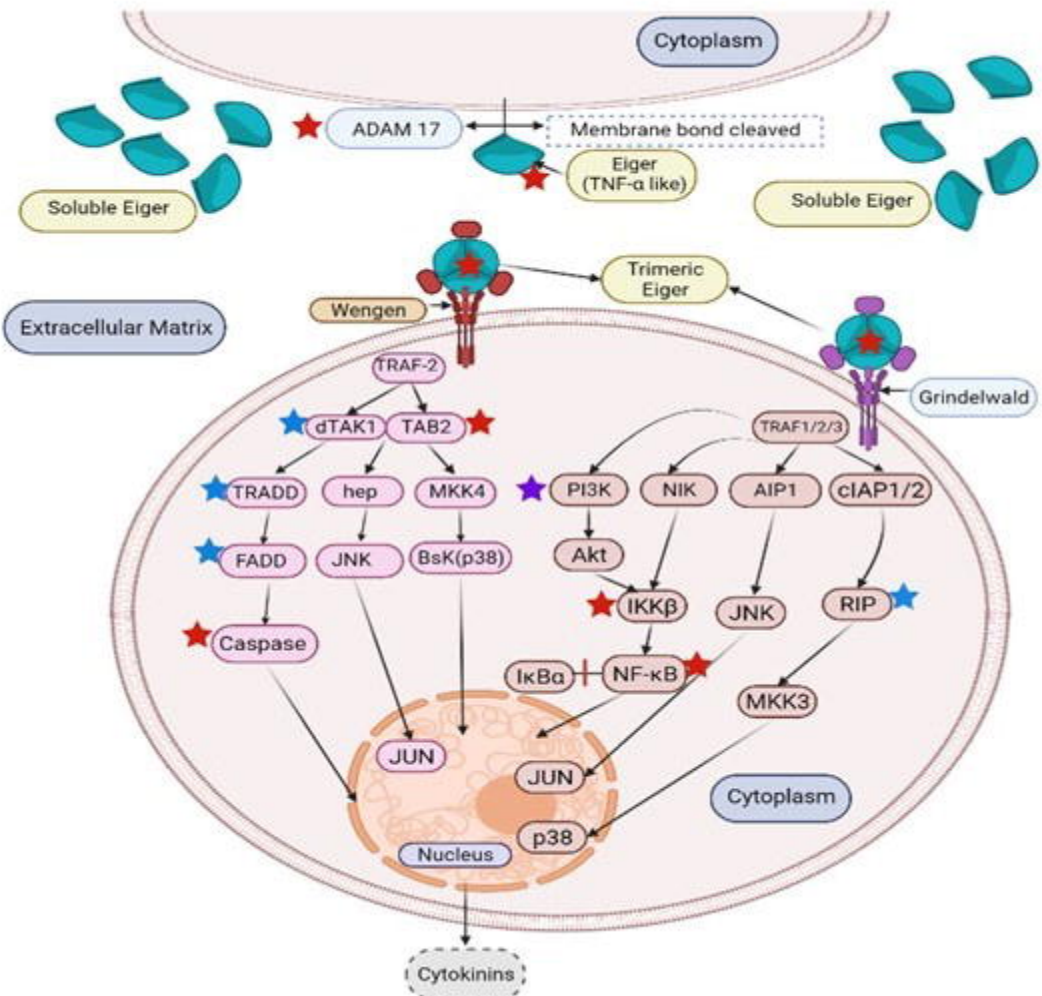
3L*i*

2L*a*

An. gambiae PEST

a

b

c

```
TNF_1B:     SCEDSTYTQL----WNWVPECLSCGSRCSSDQV--ETQACTREQNRICT
GRND-Steph: GCTRKCDRFE--YCDENSNACRNCSALCEKD-----EYSCYH---KCQ
GRND-Fly:   DCHGTICHPVNEFCYVATERCHPCIEVCNNQTHNYDAFLCAK---ECS
                                  CXXCXXXC

TNF_1A:     ECESGSFTASENHLRHCLSCSKCRKEMGQVEISSCTVDRDTVCG
WGN-Steph:  ACESRKSW-WDPTLGECAPCLICADHQV-VLRPCQDYMNTVCG
WGN-Fly:    PCAPQHWW--DSQRDRCTPCTRCQGEMI-PLRPCQLHTDTICG
                                 CXXCXXC
```

Gene Name	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