Gene 596 (2017) 89–97

Contents lists available at ScienceDirect

Gene

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

Research paper

The evolutionary divergence of STAT transcription factor in different *Anopheles* species

Kuldeep Gupta [a], Rini Dhawan [a], Mithilesh Kajla [a], Tripti Misra [a], Sanjeev Kumar [a,b], Lalita Gupta [a,c,*]

[a] Molecular Parasitology and Vector Biology Laboratory, Department of Biological Sciences, Birla Institute of Technology & Sciences (BITS) - Pilani, Pilani 333031, Rajasthan, India
[b] Department of Biotechnology, Ch. Bansi Lal University, Bhiwani, Haryana, India
[c] Department of Zoology, Ch. Bansi Lal University, Bhiwani, Haryana, India

ARTICLE INFO

Article history:
Received 25 February 2016
Received in revised form 28 August 2016
Accepted 14 September 2016
Available online xxxx

Keywords:
Anopheles
STAT transcription factor
Genetic divergence
Phylogenetic relationship
Gene duplication
Evolution

ABSTRACT

Anopheles mosquito transmits *Plasmodium*, the malaria causing parasite. Different species of *Anopheles* mosquito dominate in a particular geographical location and are capable of transmitting specific strains of *Plasmodium*. It is important to understand the biology of different anophelines to control the parasite transmission. STAT is an evolutionary conserved transcription factor that regulates the parasite development in African malaria vector *Anopheles gambiae*. Unlike *Drosophila* and *Aedes aegypti*, where a single STAT gene plays an important role in immunity, *An. gambiae* contains one evolutionary conserved STAT-A and another retro-duplicated, introns-less STAT-B gene. To find out whether other species of *Anopheles* also have two STATs, the available genomic data of different anophelines were used to annotate their STATs through *in silico* analyses. Our results revealed that Indian malaria vector *An. stephensi* genome contains two STATs, AsSTAT-A and AsSTAT-B genes. These genes were cloned and confirmed by sequencing. Both AsSTATs were found to be expressed in different development stages of mosquito. However, the relative mRNA levels of evolutionary conserved AsSTAT-A gene were always higher than the retroduplicated AsSTAT-B gene. STAT pathway was activated upon *Plasmodium berghei* infection, indicated its role in immunity. Furthermore, comparative *in silico* analysis of eighteen *Anopheles* species revealed that five species: *An. sinensis*, *An. albimanus*, *An. darlingi*, *An. dirus and An. farauti* do not contain STAT-B gene in their genome. Interestingly, thirteen species of the subgenus *Anopheles* and *Cellia* that contain both STATs were also mutually diverged. This consequence leads to sequence variability in some significant protein motifs within the STAT-B genes. Phylogenetic analyses indicated that an independent, lineage-specific duplication occurred in the subgenus *Cellia* after the diversification of series Neomyzomyia from its last common ancestor. In *An. atroparvus* (subgenus *Anopheles*), STAT gene underwent recent lineage-specific duplication and give rise to a highly similar STAT-B gene. This suggested that the genetic divergence in various *Anopheles* species might appeared due to their adaptations to the altered environmental conditions or pathogen encounters.

1. Introduction

The Janus kinase (JAK)–Signal Transducer and Activator of Transcription (STAT) signaling pathway was initially documented in humans and now recognized in lower organisms also (Wang and Levy, 2012). This pathway is regulated by the transcription factor named as STAT that regulates the expression of several effector molecules. Thorough investigations of all the model organisms has revealed that this pathway is regulated by the homologs of at least one STAT gene, which is extremely conserved and ancestral in origin (Hurst et al., 2004). In higher vertebrates, five to seven copies of STAT genes are present while the model insect *Drosophila melanogaster* comprises only one STAT92E gene, which is an ortholog of human STAT5a. Comparative analysis of genome and proteome between *D. melanogaster* and *An. gambiae* revealed that the duplication of STAT gene occurred only in *An. gambiae*. It was also suggested that the duplication of an evolutionary conserved STAT-A gene happened by retrotransposition into an intronless gene named as STAT-B (Zdobnov et al., 2002; Gupta et al., 2009).

STAT pathway has been reported for its diverse roles in development and immune responses throughout the insect world. However, the functional role of STAT-B gene was first described in *An. gambiae*

Abbreviations: STAT, Signal Transducer and Activator of Transcription; JAK, Janus kinase; NJ, Neighbor-Joining; An., *Anopheles*; MEGA, Molecular Evolutionary Genetics Analysis; UTR, untranslated region; MYA, million years ago; *P. berghei*, *Plasmodium berghei*; SOCS, suppressor of cytokine signaling; PIAS, protein inhibitor of activated STAT.

* Corresponding author at: Department of Zoology, Ch. Bansi Lal University, Bhiwani, Haryana, India.
E-mail addresses: lalitagupta@yahoo.com, lalita@pilani.bits-pilani.ac.in (L. Gupta).

http://dx.doi.org/10.1016/j.gene.2016.09.022
0378-1119/

against bacterial challenge (Barillas-Mury et al., 1999). Several research reports also revealed the regulatory role of STAT gene in pathogenic development. This signaling pathway is activated upon *Plasmodium* infection and regulates the development of late oocysts in *An. gambiae* (Gupta et al., 2009). However, in case of *An. aquasalis*, STAT gene regulates the early stages of *Plasmodium* development (Bahia et al., 2011). In addition, the genes encoding immune effector molecules are activated through the STAT pathway and are able to reduce dengue infection in *Aedes aegypti*, thereby establishing its role in mosquito immunity (Souza-Neto et al., 2009). In *An. gambiae* the duplicated STAT-B gene also acts like an upstream regulator of the evolutionary conserved STAT-A gene (Gupta et al., 2009).

STAT function has been characterized only in a few *Anopheles* species, however, 465 species of *Anopheles* mosquitoes are known and only three dozen are able to transmit human malaria parasite (Harbach, 2004). Certain species of *Anopheles* are endemic in particular zoogeographical regions of the world where they flourish due to the favorable climatic conditions (Sinka et al., 2012). Among them, afrotropical major malaria vector *An. gambiae* is the most studied species throughout the world, however, other *Anopheles* species are equally important to understand the host parasite interactions. The lack of genome sequences from other endemic malaria vectors indubitably impeded the clarification of the unique biology associated with the particular mosquito species. For better understanding of the vectorial competence of mosquitoes, *Anopheles* Genome Cluster Committee (AGCC) released genome assemblies of sixteen worldwide distributed *Anopheles* species (Neafsey et al., 2013). Their whole genome shotgun (WGS) sequences are available in unannotated form only, which can be employed to solve the mystery of vast genetic variations that exist within the same genus. Comparative study of different genes in the available genome sequences of anophelines will be helpful to understand the similarity or divergence among them. For that purpose, we identified, cloned and sequenced two STATs from the major Indian malaria vector *An. stephensi* and also analyzed their mRNA expression in different tissues. Further, we performed a genome-wide analysis to annotate the STAT genes in the available genome assemblies of sixteen *Anopheles* species with the help of *An. gambiae* and *An. stephensi* STATs. The objective of the current study was to analyze the evolutionary pattern and phylogenetic relationships of STATs among the members of anopheline subgenus *Cellia*, *Anopheles* and *Nyssorhynchus*. These findings will be certainly helpful to understand the STAT gene diversity among different *Anopheles* species and its interactions with the malaria parasites.

2. Methods

2.1. Amplification and cloning of An. stephensi STAT genes

Initially, at the starting of this study, genome sequence of *An. stephensi* was not available. Therefore, degenerate primer amplification approach was used to amplify the partial segment of STAT gene. For that, degenerate primers were designed from the conserved regions of *An. gambiae*, *Ae. aegypti* and *Culex tritaeniorhynchus* STAT sequences. To obtain complete gene sequences of both STATs, we used the genome data (then available) of *An. stephensi* for designing gene specific primers. The majority of the designed primers were 19–22 nucleotides long, with fewer than seven contiguous identical bases between two STATs. The gene-based primers were used to amplify the full length gene including 5′ and 3′ untranslated regions (UTRs). All amplicons were purified through Qiaquick PCR purification kit (Qiagen Cat No. 28104) and sequenced through Sanger's dideoxy sequencing method. Whenever required, PCR products were cloned into PCR-II TOPO TA-Vector® (Invitrogen Cat No K46001-01) following the manufacturer's instructions. Primer sequences (5′ → 3′) of AsSTAT-A and AsSTAT-B and other genes used in the study are provided in Table 1.

Table 1
Different primers used in the present study.

S. no.	Primer set	Primer sequence (5′–3′)	Size of PCR product using cDNA template	Purpose
STAT-A gene				
1.	D1-F	GARAADCARCCRCCRCARGTSATG	900 bp	Cloning with degenerate primers
	D2-R	GTRAAYGGCTGRATGTGYAGDAYCTG		
2.	1F	TCGCTTCATTAGAGCGGGAC	2067 bp	Cloning and sequencing
	7R	ACCGAGCTCACTGTCGGA		
3.	4F	CAGGTATGGACTATCTCGC	862 bp	Cloning and sequencing
	8R	CGGGCACCTGACCGTAATTG		
4.	2F	CAATCCGGCCCAGCTGTAC	150 bp	Sequencing
	3R	GTGCGCACCATTATCTGCAG		
5.	5F	CCTGTGTGAAAAAGCGTTCA	220 bp	Real time PCR
	6R	ATAATCCTCCGCCTTCGACT		
6.	5F	CCTGTGTGAAAAAGCGTTCA	915 bp	PCR product to obtain full length sequence
	9R	CGTAAACCTTAAGCTACCGTGC		
STAT-B gene				
7.	1F	TAGTGTTTTCCCCGCTGCAT	878 bp	To obtain full length sequence
	3R	TGCCGTATTGATCGTTGCCT		
8.	2F	GCTTTGCAGTCATCGACCTG	983 bp	Cloning and sequencing
	5R	CGAAGAATTTGGGGCGGTTG		
9.	4F	TATCGGTGCGTATGCCAAGC	1312 bp	Cloning and sequencing
	7R	CAGCTCCGTCGCGATATAGT		
10.	6F	GCCAGTTGTAAGCCGGGCACA	276 bp	Real time PCR
	7R	CAGCTCCGTCGCGATATAGT		
11.	6F	GCCAGTTGTAAGCCGGGCACA	563 bp	PCR product to obtain full length sequence
	8R	CTTTTCATCATCGTAAGCTCCG		
Others primers				
12.	SOCS-F	CGTCGTACGTCGTATTGCTC	240 bp	Positive control for STAT pathway
	SOCS-R	CGGAAGTACAATCGGTCGTT		
13.	S7-F	GGCGATCATCATCTACGT	487 bp	PCR internal loading control
	S7-R	GTAGCTGCTGCAAACTTCGG		
14.	M13-F	TGTAAAACGACGGCCAGT	Vector internal primers (PCR-II TOPO TA) used for confirming the clones and sequencing	
	M13-R	CAGGAAACAGCTATGAC		