


RESEARCH

Open Access



Molecular forms of *Anopheles subpictus* and *Anopheles sundaicus* in the Indian subcontinent

Ankita Sindhania¹, Manoj K. Das², Gunjan Sharma¹, Sinnathamby N. Surendran³, B. R. Kaushal⁴, Himanshu P. Lohani⁴ and Om P. Singh^{1*} 

Abstract

Background: *Anopheles subpictus* and *Anopheles sundaicus* are closely related species, each comprising several sibling species. Ambiguities exist in the classification of these two nominal species and the specific status of members of these species complexes. Identifying fixed molecular forms and mapping their spatial distribution will help in resolving the taxonomic ambiguities and understanding their relative epidemiological significance.

Methods: DNA sequencing of Internal Transcribed Spacer-2 (ITS2), 28S-rDNA (D1-to-D3 domains) and cytochrome oxidase-II (*COII*) of morphologically identified specimens of two nominal species, *An. subpictus* sensu lato (*s.l.*) and *An. sundaicus s.l.*, collected from the Indian subcontinent, was performed and subjected to genetic distance and molecular phylogenetic analyses.

Results: Molecular characterization of mosquitoes for rDNA revealed the presence of two molecular forms of *An. sundaicus s.l.* and three molecular forms of *An. subpictus s.l.* (provisionally designated as Form A, B and C) in the Indian subcontinent. Phylogenetic analyses revealed two distinct clades: (i) subpictus clade, with a single molecular form of *An. subpictus* (Form A) prevalent in mainland India and Sri Lanka, and (ii) sundaicus clade, comprising of members of Sundaicus Complex, two molecular forms of *An. subpictus s.l.* (Form B and C), prevalent in coastal areas or islands in Indian subcontinent, and molecular forms of *An. subpictus s.l.* reported from Thailand and Indonesia. Based on the number of float-ridges on eggs, all *An. subpictus* molecular Form B were classified as Species B whereas majority (80%) of the molecular Form A were classified as sibling species C. Fixed intragenomic sequence variation in ITS2 with the presence of two haplotypes was found in molecular Form A throughout its distribution.

Conclusion: A total of three molecular forms of *An. subpictus s.l.* and two molecular forms of *An. sundaicus s.l.* were recorded in the Indian subcontinent. Phylogenetically, two forms of *An. subpictus s.l.* (Form B and C) prevalent in coastal areas or islands in the Indian subcontinent and molecular forms reported from Southeast Asia are members of Sundaicus Complex. Molecular Form A of *An. subpictus* is distantly related to all other forms and deserve a distinct specific status.

Keywords: *Anopheles subpictus*, *Anopheles sundaicus*, Sibling species, Species complex, Phylogenetics, Ribosomal DNA

Background

Anopheles subpictus and *Anopheles sundaicus* are closely related species belonging to Pyrethophorus Series [1] and each has been reported to be comprised of several sibling

*Correspondence: singh@nimr.org.in

¹ National Institute of Malaria Research, Sector 8 Dwarka, New Delhi 110077, India

Full list of author information is available at the end of the article



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

members. *Anopheles subpictus* sensu lato (*s.l.*) is widely distributed species prevalent throughout Oriental and Australasian Zones; mainly in Afghanistan, Australia, Bangladesh, Cambodia, China, India, Indonesia, Iran, Malaysia, Maldives, Mariana Islands, Myanmar (Burma), Nepal, New Guinea (Island)-Papua New Guinea, Pakistan, Philippines, Sri Lanka, Thailand and Vietnam [2]. In India, *An. subpictus s.l.* is the most common species occurring in all the mainland zones [3] and is found up to 1900 above msl and in many parts of Himalayas, though in small numbers. *Anopheles subpictus* gradually declines in abundance proceeding eastwards in India [4]. Other nominal taxa, *An. sundaicus*, has been recorded mainly from the coastal areas of north-eastern India, Andaman & Nicobar (A&N) Islands, Peninsular Malaysia, Malaysian Borneo (Miri, Sarawak), northern Sumatra & Java, and Indonesia [5]. Currently, the distribution of *An. sundaicus* in India is limited to the A&N Islands and Kutch (western coast of India) [6, 7]. *Anopheles sundaicus* has been reported to be disappeared from Chilka (Odisha, India) [8], which has been reported earlier [9].

In India, *An. sundaicus* is considered as a potent malaria vector, but *An. subpictus* has not been recognized yet as a malaria vector by the national malaria control programme [10]. However, *An. subpictus* is considered a potential vector in Southeast Asian countries [11–14]. In India, there are increasing evidences supporting the role of *An. subpictus* as a potential malaria vector, especially in coastal areas of India [15, 16] and Sri Lanka [17]. The varying vectorial potential in different geographical areas is most likely due to the presence of different biological species (cryptic species) present in *An. subpictus*. In India, four sibling species of *An. subpictus*, provisionally designated as species A, B, C and D have been recognized based on paracentric inversions present on the chromosome X [18]. However, the status of sibling species in other countries, except Sri Lanka (where species A and B were identified based on chromosomal inversions) [19], remains obscure. Though *An. subpictus* has been incriminated as a malaria vector in several places in India, mostly in coastal areas, the differential role of their sibling species in malaria transmission is not well understood. Apparently, species B is an efficient vector of malaria, which is prevalent in coastal areas of India [19] and Sri Lanka [20]. *Anopheles sundaicus* comprises of four sibling species [5], designated as *An. sundaicus* sensu stricto (*s.s.*), *Anopheles epiroticus s.s.* (formerly, *An. sundaicus* species A) [21], *An. sundaicus* species D [22, 23] and *An. sundaicus* species E [24]. All of them act as predominant malaria vectors depending upon the location [25].

Morphologically, *An. subpictus* and *An. sundaicus* are almost similar and are distinguished based on the

presence/absence of speckling on legs [26]. However, the taxonomic status of members of *An. subpictus* seems to be ambiguous based on the molecular phylogenetic studies which revealed that the majority of the morphologically attributed species B of *An. subpictus* is closely related to *An. sundaicus* complex and far distantly related to other members of *An. subpictus s.l.* [27, 28]. Therefore, Surendran et al. [27] recognized species B of Subpictus Complex as a member of Sundaicus Complex. Interestingly, species B of Subpictus Complex [29] and members of the Sundaicus Complex are prevalent mainly in coastal areas and islands [14], although ecological plasticity in the breeding preference has also been noted [5].

Molecular characterization of *An. subpictus* from different geographical regions is important in identifying different biological forms, their distribution and role in malaria epidemiology. Studies carried out in Sri Lanka revealed the presence of two distinct molecular forms. Very recently, three additional molecular forms in respect of ITS2 has been reported from Thailand and Indonesia [30]. However, reliable published data on molecular forms of *An. subpictus* in Indian territory is not available. Few published and unpublished data from GenBank showed a high degree of polymorphism in ITS2 sequences, in the Indian population, to the extent that every individual was different [31–33], which raises suspicion in the quality of sequences. The verification of such data is warranted. Therefore, a study was carried out to characterize the rDNA of *An. subpictus* population from wide geographical areas in the Indian subcontinent in addition to closely related species *An. sundaicus*.

For molecular characterization, nuclear ribosomal DNA (rDNA), particularly ITS2 and 28S rDNA was selected, which have been extensively used in species identification due to their high uniformity in sequence in an interbreeding population, while differences exist between species [34]. Such uniformity of sequence in a species is believed to be due concerted evolution acting on rDNA, which tends to homogenize sequence [35]. The homogenization of rDNA is thought to be due to unequal crossing over of rDNA copies arranged in a tandem fashion, however, the phenomenon has not been fully understood [36]. On the other hand mitochondrial DNA, *COI* in particular has also been successfully used in barcoding of life where sufficient gap, known as 'barcoding gap' (an arbitrary threshold), exists between intraspecific variation and interspecific divergence. In closely related species, such gap often does not exist, but a high percentage of well-differentiated species has similar or even identical *COI* sequences [37], particularly in dipteran [38]. Such overlap may be due to the retention of ancestral polymorphism and introgression [36]. Moreover, rDNA is thought to be advantageous over the mtDNA in terms of

the speed of lineage sorting of its multicopy rDNA array [36].

Methods

Mosquito collection

Adult female *An. subpictus* and *An. sondaicus* were collected from different parts of the Indian subcontinent, mainly from mainland India, A&N Islands (India) and Sri Lanka. From mainland India, mosquitoes were collected from Ranchi (Jharkhand, 23° 21' N, 85° 18' E), Nuh (Haryana, 28° 06' N, 76° 59' E), Goa (15° 38' N, 73° 45' E), Alwar (Rajasthan, 27° 37' N, 76° 35' E), Jodhpur (Rajasthan, 26° 18' N, 73° 00' E), Gadchiroli (Maharashtra, 20° 10' N, 80° 00' E), Delhi (28° 34' N, 77° 01' E), villages surrounding Chilka lake (Odisha) and Puducherry. The villages surveyed surrounding Chilka lake, which is a large brackish water lagoon covering an area of over 1100 km² in Orissa state, were Panasapada (19° 43' N 85° 31' E), Satapada (19° 40' N 85° 26' E), Brahmigiri (19° 47' N 85° 36' E), Sipakuda (19° 23' N 85° 04' E), Minsa (19° 18' N 84° 47' E), Siara (19° 44' N 85° 32' E) and Gambhari (19° 72' N, 85° 46' E). In Puducherry, mosquitoes were collected from Kaliakuppam (12° 05' N 79° 55' E), Munjalkuppam (12° 02' N 79° 76' E), Sedarapet (12° 00' N 79° 44' E), Pillayarkuppam (11° 48' N 79° 46' E). From A&N Islands, mosquitoes were collected from Kimios, Car Nicobar (9° 15' N 92° 76' E) and Sippighat, Port Blair (11° 59' N 92° 68' E). In Sri Lanka, mosquitoes were collected from Kallady (7° 42' N 81° 42' E), Muthoor (8° 27' N 81° 16' E), and Chenkalady (7° 47' N 81° 35' E). Few *An. sondaicus* collected from Thanbyuzayat Township, Mon State, Myanmar (15° 58' N 97° 44' E) were also included in the study. The geographical location of mosquito collection sites is depicted in Fig. 1. Mosquitoes were either preserved in isopropanol or kept dried in microcentrifuge containing a piece of silica gel crystal. Where feasible, live female mosquitoes were transported to the laboratory for oviposition and identification of sibling species based on the number of float-ridges on eggs. The adult female mosquitoes were identified morphologically using keys by Christophers [26] and Reid [39].

Egg float-ridges count

Live *An. subpictus* mosquitoes were transported from mainland India only. The transported mosquitoes from the field were allowed to feed overnight on a rabbit. Eggs were obtained from individual morphologically identified gravid-females in a plastic bowl with the inner side lined with blotting paper and containing a small amount of water for keeping the blotting paper moist. The mosquitoes, which successfully laid their egg, were persevered in a microcentrifuge tube containing isopropanol for molecular studies. Eggs from individual mosquitoes were

transferred on to microslides and the number of ridges on egg-floats obtained were counted under a transmission microscope using 10× objective lens. At least 10 eggs from each mosquito were examined. Based on the mode number of ridges, the mosquitoes were classified for sibling species following Suguna et al. [18].

DNA extraction, PCR amplification and sequencing

DNA was extracted from individual mosquitoes following Coen et al. [40]. Two adjacent regions of rDNA, i.e., ITS2 and partial 28S rDNA (D2-to-D3 domains), were PCR-amplified separately following Singh et al. [41]. In addition, to obtain a continuous stretch of rDNA covering partial 5.8S, full ITS2 and partial 28S rDNA (up to d3 domain) an approximately 2 kb region was also amplified using primers ITS2A and D3B (Table 1) from at least two representative samples of each molecular form of *An. subpictus* and *An. sondaicus*, and one sample of *An. epiroticus*. The PCR conditions were as described by Singh et al. [41], except extension time, which was increased to 1.5 min. The PCR products were cleaned using Exo-Sap IT (Thermo-Scientific) and sequenced using Big Dye Terminator v3.1 (Thermo-Scientific). The products were cleaned using the vendor's protocol and electrophoresed in ABI-Prism 3770xl DNA sequencer. The sequencing of ITS2 PCR-products was done with the primers that were used for PCR amplification, i.e., ITS2A and ITS2B. Larger PCR products were sequenced using a primer walking strategy. The primers used for primer walking sequencing were ITS2A, ITS2B, D2A, D2B, D3A and D3B (sequences are provided in Table 1).

Five individuals of each molecular form, except *An. epiroticus* (n=1), were also sequenced for cytochrome oxidase II (*COII*) mitochondrial DNA. An approximately 750 bp region of the *COII* gene was PCR-amplified in a 20 µL reaction mixture using GOTaq Master Mix (Promega), 0.5 µM of each primer COIIF and COIIR (Table 1) and 0.5 µL of DNA. The PCR thermal cycling conditions were 95 °C for 2 min followed by 35 cycles each at 95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min and a final extension at 72 °C for 7 min. The PCR products were cleaned using Exo-Sap IT (Thermo Scientific) and sequenced using Big Dye Terminator v3.1 (Thermo Scientific).

Cloning, sequencing and phasing of haplotypes

During the sequencing of ITS2 of a molecular form of *An. subpictus* (Form A) collected from mainland India and Sri Lanka, it was observed that the presence of mixed bases in DNA sequence chromatogram starting from a specific point due to the presence of indel in one of the two haplotypes present in the mosquito (Fig. 2). To phase out the sequence of two haplotypes, PCR product amplified for the ITS2 region of a total of five

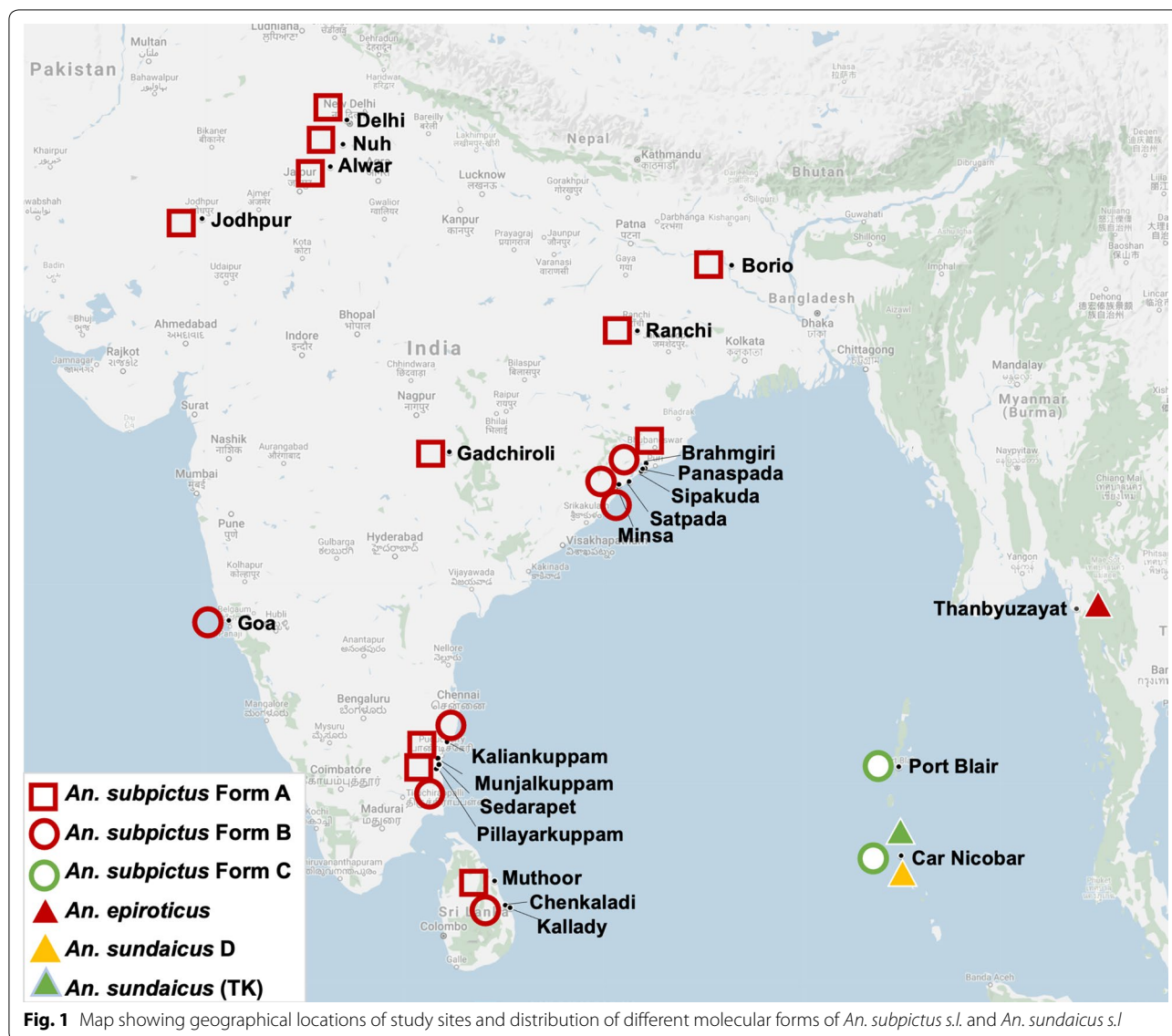


Fig. 1 Map showing geographical locations of study sites and distribution of different molecular forms of *An. subpictus* s.l. and *An. sundaicus* s.l

Table 1 List of primers used for DNA sequencing

Primers' name	Nucleotide sequence (5'-3')	Reference
ITS2A	TGT GAA CTG CAG GAC ACA T	Beebe and Saul [60]
ITS2B	TAT GCT TAA ATT CAG GGG GT	Beebe and Saul [60]
D2A	AGT CGT GTT GCT TGA TAG TGC AG	Campbell et al. [61]
D2B	TTG GTC CGT GTT TCA AGA CGG G	Campbell et al. [61]
D3A	GAC CCG TCT TGA AAC ACG GA	Litvaitis et al. [62]
D3B	TCG GAA GGA ACC AGT TAC TA	Litvaitis et al. [62]
COIIF	TCT AAT ATG GCA GAT TAG TGC A	Sharpe et al. [63]
COIIR	ACT TGC TTT CAG TCA TCT AAT G	Sharpe et al. [63]
SP6	ATTTAGGTGACACTATAG	Universal primer
T7	TAATACGACTCACTATAGGG	Universal primer

samples (three samples from mainland India, one each from Delhi, Ranchi and Kutch, and two samples from Sri Lanka) were sequenced after cloning. For cloning, briefly, PCR products were purified using Qiaquick PCR Purification Kit, cloned in pGEM-T Vector System (Promega) following vendor's protocol and transformed into chemically competent *Escherichia coli* DH5α. The transformants were plated on LB-Agar containing 100 µg/ml ampicillin. White colonies were picked up and plasmid DNA was isolated by boiling them in 50 µl of TE buffer. At least five clones from each mosquito were sequenced using universal primers T7 and SP6 (sequences provided in Table 1).

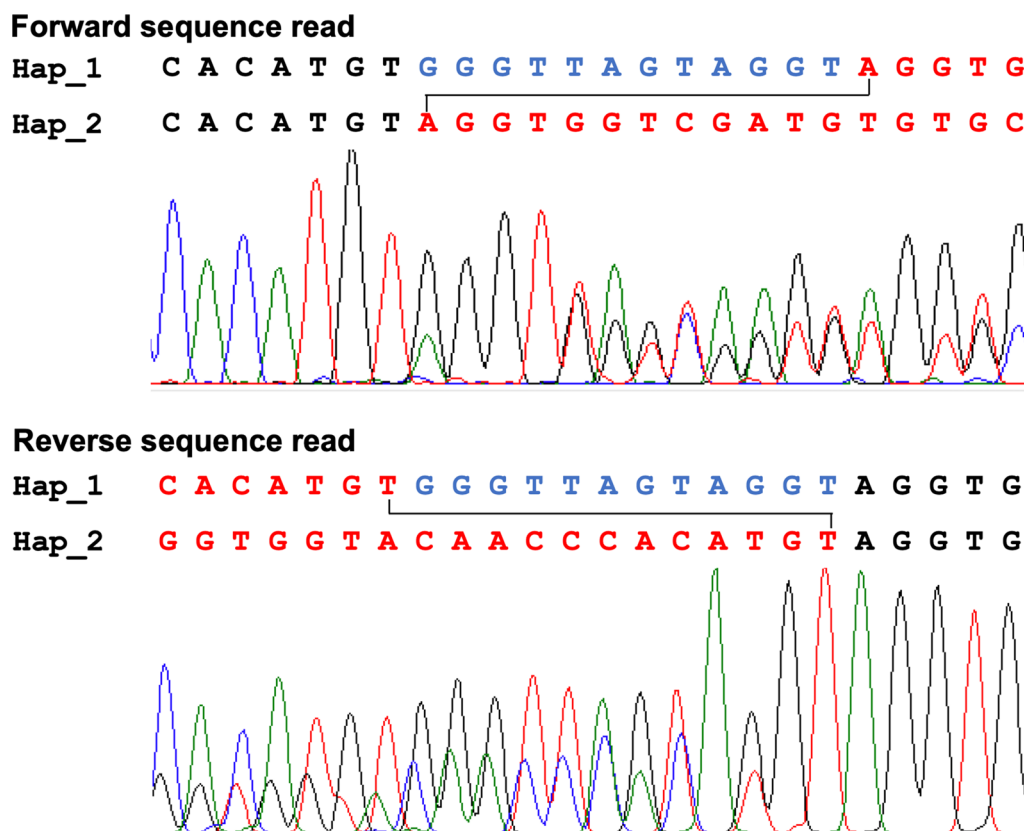


Fig. 2 DNA sequence chromatograms of a portion of ITS2 (Forward and reverse) of *An. subpictus* form A, having indel in one haplotype. Hap_1 and Hap_2 in this figure represent haplotype 1 and 2, respectively

Sequence analyses and phylogenetic inferences

The DNA sequence chromatograms were edited using Finch TV (Geospiza, Inc.) and aligned using an online tool ClustalW 1.4 [42] available at <https://www.genome.jp/tools-bin/clustalw>. Since *An. subpictus* Form A had two haplotypes, arising due to 12 bp indel in a haplotype in ITS2 region, and sequence chromatograms were ambiguous from the point of indel, identification of two haplotypes were done on the basis of combining reads from forward sequences and reverse sequence (1X) up to the point of start of indel. The determination of sequence in the indel region was made empirically, as illustrated in Fig. 2.

Phylogenetic analyses of ITS2 and 28S-D2-D3 were performed individually using software PAUP* version 4.0 beta 10 [43]. For phylogenetic analysis of ITS2, sequences of four *An. sundaicus* variants, i.e., *An. epiroticus* s.s. (variant I, GenBank accession no. AY768540), variant II (AY768541), *An. sundaicus* cytotype D (variant III, AY768542), *An. sundaicus* s.s. (variant IV, AY768543) [24], *Anopheles indefinitus* (GQ870332) and sequences of *An subpictus* variants from Thailand (MT068425 and

MT068434) and Indonesia (MT068436) [30] were also used. Sequences of *Anopheles fluviatilis* (DQ345964) [41] and *Anopheles gambiae* (JN994138) were taken as outgroup taxa for ITS2, and *An. gambiae* (KC177663) and *An. fluviatilis* (DQ665846) [41] were taken as outgroup taxa for 28S-D2-D3 phylogenetic analysis. *Anopheles gambiae* was chosen as outgroup due to its phylogenetic proximity with *An. subpictus* and *An. sundaicus* complexes, whereas *An. fluviatilis* (belongs to Myzomyia series of subgenus *Cellia*) was chosen as it is phylogenetically relatively distant from *An. subpictus* and *An. sundaicus* complexes. ITS2 sequences of the Indian and Sri Lankan *An. subpictus* (published or submitted in GenBank) were not included in the analysis due to significant error in the sequences (to the extent that every individual was different [31–33]), resulting from quality issues mostly arising due to the presence of indel in one of the two haplotypes present in Form A, which is prevalent throughout mainland India and Sri Lanka. Maximum Likelihood (ML) methods were used for the construction of phylogenetic trees. The ML analyses were carried out using the heuristic searches and TBR branch-swapping

algorithm with ten random taxon addition sequences. For ML-tree construction, the best evolutionary model of nucleotide substitution that fits the data were obtained using software Modeltest 3.7 [44]. The robustness of inference was calculated by the bootstrap method (1000 times) within PAUP. The estimated bootstrap values were reported as Maximum Likelihood bootstrap percentages (MLB).

Allele-specific PCR (ASPCR) for the identification of molecular forms of *Anopheles subpictus* in mainland India

Since the sequencing of a large number of samples was not feasible, a PCR assay was developed for the quick identification of two molecular forms, *An. subpictus* Form A and Form B, only forms prevalent in mainland India. A universal primer Sub F (5'-ACT GCA GGA CAC ATG AAC ACC G-3') was designed from upstream of ITS2A sequence and two reverse primers namely SUBP-A (5'-CGT TAC ACG CAA CAA GCG AC-3') and SPB (5'-GCC GAC ACC ACC AAC TG-3'), specific to Form A and B of *An. subpictus*, respectively, were designed. The expected amplicon sizes for Form A and B of *An. subpictus* are 563 and 444 bp, respectively. The PCR reaction (15 µl) was carried out using GOTaq

Master Mix (Promega) and 0.2 µM of each primer. The PCR thermal cycling conditions were 95 °C for 2 min, followed by 35 cycles each at 95 °C for 30 s, 58 °C for 30 s, 72 °C for 45 s, and a final extension at 72 °C for 7 min. The PCR products were visualized on 2% agarose gel containing ethidium bromide under the gel documentation system. Numbers of samples on which ASPCR was performed have been given in Table 2.

Results

DNA sequences analyses

Details of samples sequenced for 28S-D2-D3 and ITS2 are provided in Table 2. Analysis of sequences revealed a total of five distinct molecular forms (in respect to ITS2 and 28S rDNA combined), of which three forms were from morphologically identified *An. subpictus* and two from *An. sundaicus*. Sequences obtained from representative samples of each molecular forms (and haplotypes) after amplifying complete stretch of partial 5.8S rRNA, complete ITS2 and partial 28S rRNA (D1-D3 domain) have been shown in the aligned form in Additional file 1: Fig. S1. The sequences are also available from GenBank (accessions numbers: MW078484 to MW078490). The three forms of *An. subpictus* have been designated as molecular forms A, B and C. Form A of *An. subpictus*

Table 2 Number of mosquitoes identified into molecular forms of *An. subpictus* and *An. sundaicus* based on the rDNA (ITS2 and 28S-D2-D3) sequences and AS-PCR

Locality	Molecular forms	DNA sequencing		AS-PCR ^a
		ITS2	28S-D2-D3	
Nuh, India	<i>An. subpictus</i> Form A	4	4	86
Ranchi, India	<i>An. subpictus</i> Form A	63	19	15
Borio, India	<i>An. subpictus</i> Form A	2	2	10
Alwar, India	<i>An. subpictus</i> Form A	3	3	12
Jodhpur, India	<i>An. subpictus</i> Form A	2	7	7
Gadchiroli, India	<i>An. subpictus</i> Form A	4	4	25
Chilka, India	<i>An. subpictus</i> Form A	3	3	62
	<i>An. subpictus</i> Form B	8	8	618
Goa, India	<i>An. subpictus</i> Form B	2	2	5
Puducherry, India	<i>An. subpictus</i> Form A	8	8	59
	<i>An. subpictus</i> Form B	5	5	73
Sri Lanka	<i>An. subpictus</i> Form A	22	13	16
	<i>An. subpictus</i> Form B	6	6	44
Portblair, A&N Island	<i>An. subpictus</i> Form C	5	5	ND
Carnicobar, A&N Island	<i>An. subpictus</i> Form C	5	5	ND
	<i>An. sundaicus</i> D (variant III ^b)	12	12	ND
	<i>An. sundaicus</i> (TK)	1	1	ND
Myanmar	<i>An. epiroticus</i> (variant I ^b)	4	4	ND

ND not done

^a Assignment of Form B or C of *An. subpictus* in a specific locality was done on the basis of molecular form present in the area as determined by DNA sequencing

^b Designation of variants as per Dusfour et al. [24]

were recorded from mainland India and Sri Lanka, form B was found restricted to the coastal areas of mainland India (Chilka, Goa, Puducherry) and Sri Lanka island while form C from A&N Islands (Port Blair and Car Nicobar). The two molecular forms of *An. sundaicus* were identified as *An. epiroticus s.s.* and *An. sundaicus* species D based on homology with published ITS2 sequences [21, 23, 24]. The characterization of species was done based on the polymorphism at three nucleotide positions of ITS2, i.e., 479 (T/C), 538 (G/T) and 603 (C/indel) (nucleotide position as described by Linton et al. (2005) [21]). Nucleotide bases at these three sites for *An. epiroticus*, *An. sundaicus* D and *An. sundaicus s.s.* were TGC, CGC and CT- (dash denotes indel), respectively. The *An. epiroticus s.s.* was recorded from Myanmar and Car Nicobar (A&N Island) and *An. sundaicus* species D from Port Blair and Car Nicobar (A&N Island). The sequence of *An. epiroticus s.s.* recorded from Myanmar had 100% homology to *An. epiroticus s.s.* [21] (variant I [24]). One sample of *An. sundaicus* from Car Nicobar also had 100% homology with *An. epiroticus*, but had mixed nucleotide bases (G+T) at the nucleotide position 538 (nucleotide position as per Linton et al. [21], where 'G' peak was substantially prominent as compared to 'T' (Fig. 3). This molecular form was identical to haplotype 'TK' designated by Zarowiecki et al. [57] and referred hereafter as *An. sundaicus* (TK). The ITS2 of *An. sundaicus* D was identical to variant III [24]. However, there was no difference in the 28S-D1-D3 sequence among all molecular forms of *An. sundaicus*.

Intragenomic sequence variation in *Anopheles subpictus* molecular Form A

DNA sequencing of ITS2 of *An. subpictus* (Form A) resulted in mixed nucleotide peaks from a specific position of nucleotide sequence (from base position 450 in Additional file 1: Fig. S1) resulting in an ambiguous sequence (Fig. 2) from this point due to the presence of mixed haplotypes where one haplotype had indel.

Phasing of haplotypes through sequencing of cloned PCR product revealed the presence of two haplotypes (referred to as haplotype 1 and haplotype 2 in Additional file 1: Fig. S1), which differ by indel/insert of 12. The presence of these two haplotypes was found fixed in Form A throughout its distribution from northern India to Sri Lanka.

Genetic distance and phylogenetic analyses

The genetic distances and number of nucleotide substitutions, considering ITS2 and 28S-D2-D3 sequences, between three molecular forms of *An. subpictus* (A, B and C), *An. epiroticus* and *An. sundaicus* D are shown in Table 3. It was observed that *An. subpictus* Form A is distantly related to all other molecular forms of *An. subpictus* and *An. sundaicus*, with nucleotide substitutions 100 to 106 and genetic distances 0.063 to 0.066 (Kimura-2-parameter). However, two molecular forms of *An. subpictus*, i.e., Form B and C, *An. epiroticus* and *An. sundaicus* species D are closely related (nucleotide substitution 1 to 10; Kimura-2-parameter genetic distance 0.001 to 0.006).

Recently two new molecular forms of *An. subpictus* have been reported from Thailand and Indonesia with respect to ITS2 [30]. Therefore, a separate

Table 3 Pairwise distances (Kimura-2 parameter model) (below diagonal) and nucleotide substitution (above diagonal) between different molecular forms of *An. subpictus* and *An. sundaicus* based on ITS2 and 28S-D2-D3

	Sub_A	Sub_B	Sub_C	Epi	Sun_D
Sub_A	–	104	106	100	101
Sub_B	0.065	–	9	9	10
Sub_C	0.066	0.005	–	8	9
Epi	0.063	0.005	0.005	–	1
Sun_D	0.063	0.006	0.005	0.001	–

Sub_A: *An. subpictus* Form A; Sub_B: *An. subpictus* Form B; Sub_C: *An. subpictus* Form C; Epi: *An. epiroticus* (variant I); Sun_D: *An. sundaicus* D (variant III)

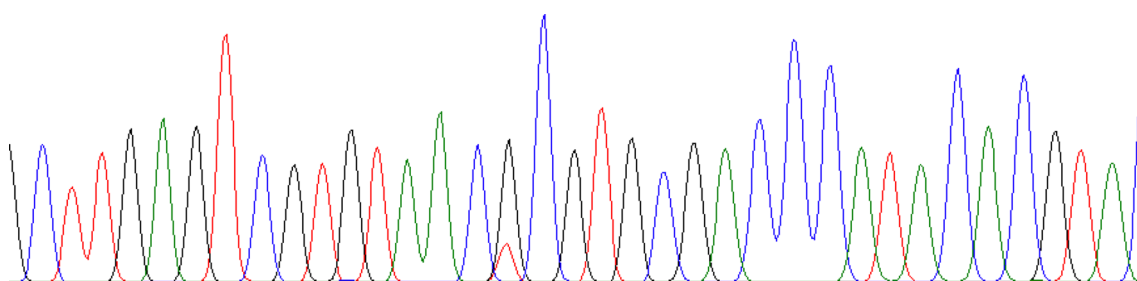


Fig. 3 Portion of DNA chromatogram of ITS2 sequence of *An. sundaicus* (TK) from Car Nicobar island showing mixed bases (highlighted) at nucleotide base position 518 (base position as per Zarowiecki et al. 2014 [57])

genetic distance analysis was performed, after including molecular forms reported from Thailand and Indonesia (representative GenBank sequences MT068425 and MT068436) [30] and *An. indefinitus* (GQ870332). The genetic distance and number of nucleotide substitutions are displayed in Table 4. It was observed that there is no genetic distance between Form C (India) and MT068436 (Indonesia) Table 4, but these sequences differed by 2 bp indel (Additional file 1: Fig. S1). *Anopheles subpictus* Forms B, C, molecular forms from Thailand and Indonesia (MT068425, MT068434, MT068436), *An. epiroticus s.s.*, *An. sundaicus* species D and *An. indefinitus* are closely related (nucleotide substitution 0 to 10; Kimura-2-parameter genetic distance 0.0 to 0.018) and are distantly related to Form A (nucleotide substitution 48 to 52; Kimura-2-parameter genetic distance 0.094 to 0.105).

The phylogenetic analysis of all molecular forms of *An. subpictus* and *An. sundaicus* from the Indian subcontinent along with four variants in *An. sundaicus*, *An. indefinitus* and *An. subpictus* variants from Thailand and Indonesia were performed using ITS2 sequences. The best-fit model inferred for ITS2 sequences with outgroup taxa TPM2uf+G was taken for maximum likelihood (ML) analysis. Construction of phylogenetic tree resulted in two distinct monophyletic clades supported by high bootstrap value (100), i.e., subpictus clade comprising of the single molecular form of *An. subpictus* (Form A) and sundaicus clade comprising of *An. subpictus* Form B, *An. subpictus* Form C, molecular forms of *An. subpictus* from Thailand and Indonesia, all variants of *An. sundaicus* and *An. indefinitus* (Fig. 4).

Phylogenetic tree was also constructed for 28S rDNA (D2-D3) sequences of all molecular forms of *An. subpictus* (Forms A, B and C), *An. sundaicus* D and *An. epiroticus s.s.* prevalent in the Indian subcontinent. The best-fit model of nucleotide substitution TIM3+G was implemented in the construction of ML trees. Cladogram

derived from 28S rDNA sequences also resulted in the generation of two distinct monophyletic clades similar to ITS2, supported with high bootstrap value (100) (Fig. 5).

Analysis of limited numbers of *COII* sequences of all representative molecular forms (based on rDNA sequences) revealed the presence of multiple haplotypes in each form. A total of 3/5 (number of haplotypes/number sequenced) haplotypes in *An. subpictus* Form A, 3/5 in Form B, 1/5 in Form C, 2/5 in *An. sundaicus* D and 1/1 in *An. epiroticus* were recorded. The Maximum Likelihood phylogeny estimate based on *COII* data for all molecular forms of Indian *An. subpictus* and *An. sundaicus* complex was constructed using software PAUP 4.0 beta. The best fit model used was GTR+I+G, as estimated by the Modeltest 3.7 [44] (Fig. 6). ML bootstrap values based on 1000 replicates were calculated. ML tree generated two distinct monophyletic clades similar to the phylogenetic tree based on rDNA. Based on *COII* phylogeny, *An. subpictus* form C was much closer to *An. epiroticus* than any members of *An. subpictus*.

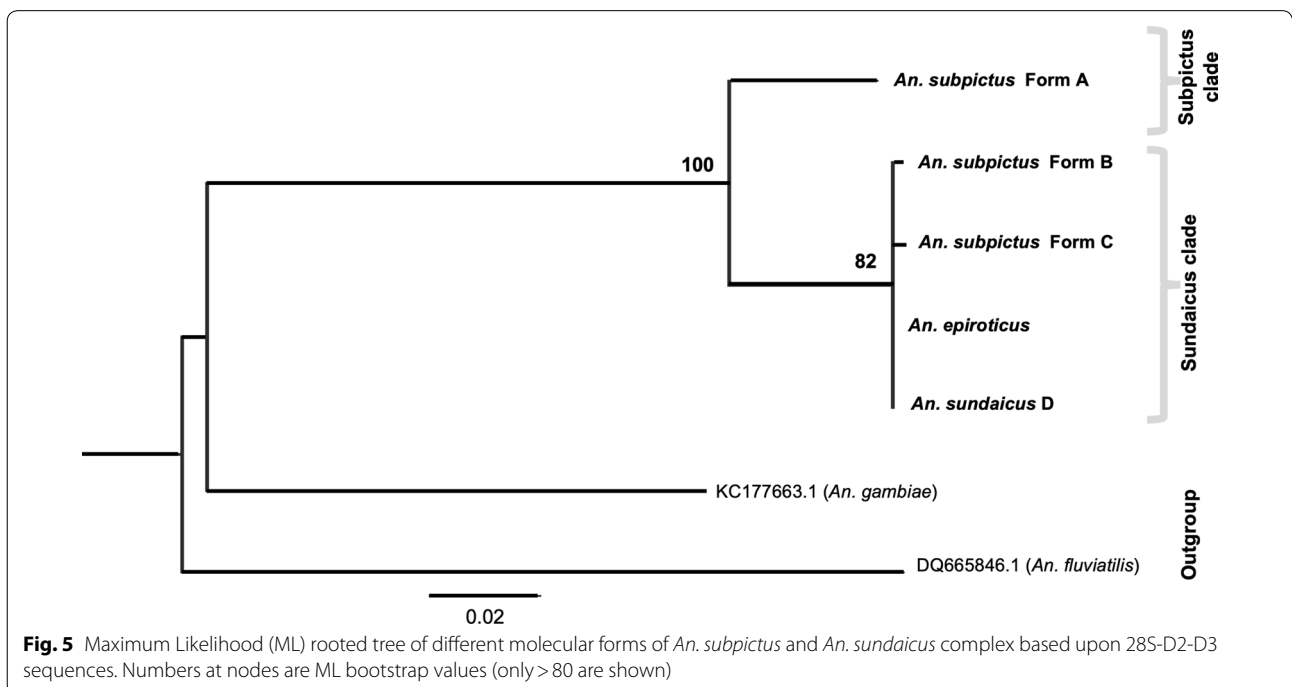
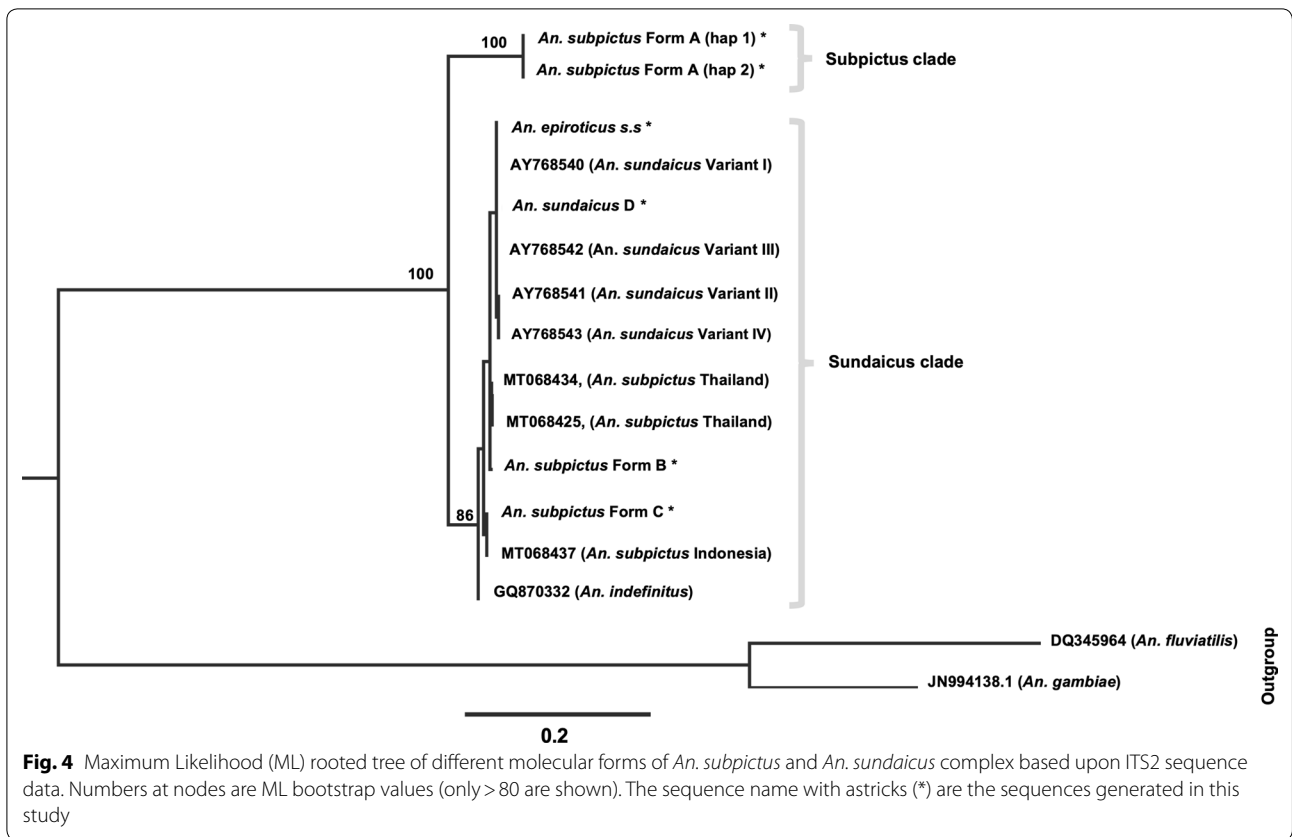
Allele-specific PCR

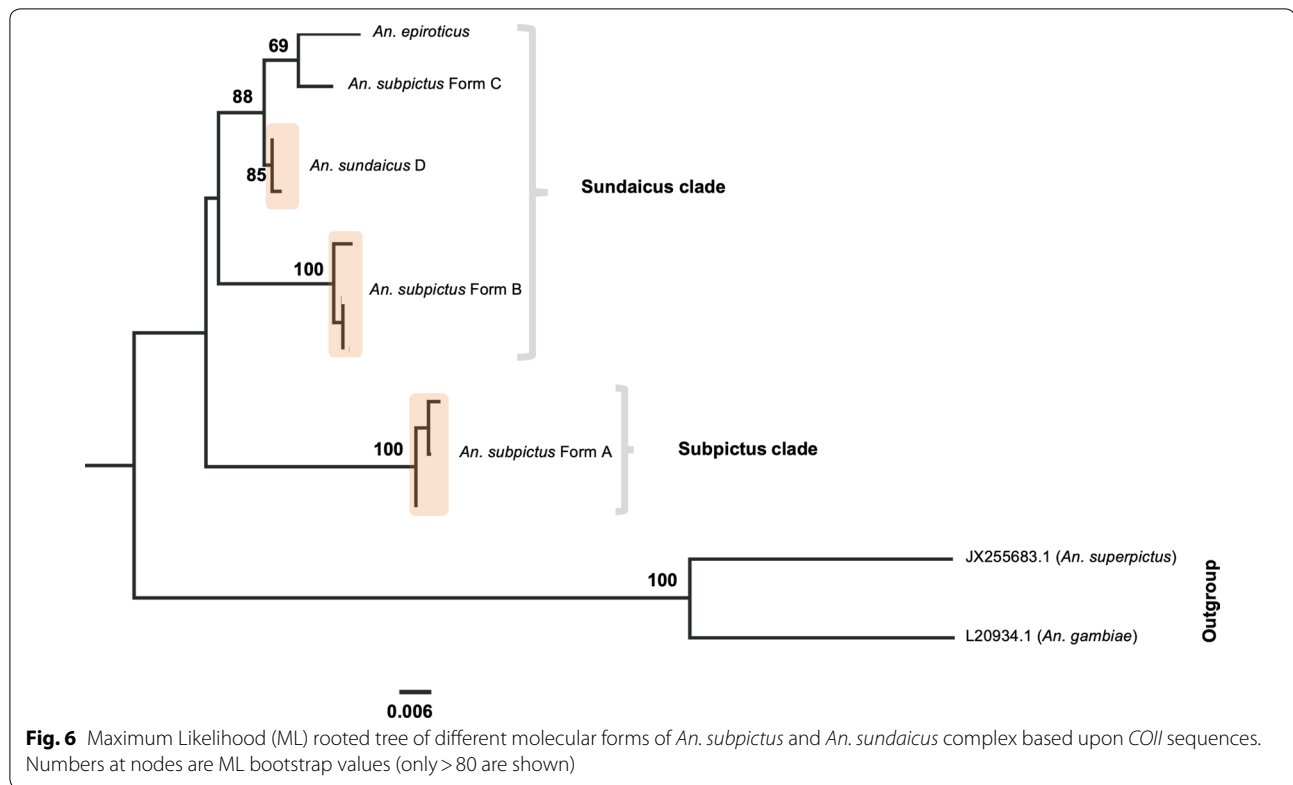
Though this PCR was developed with an intension to differentiate the two forms of morphologically identified *An. subpictus* (*An. subpictus* Form A and Form B) in the Indian mainland and Sri Lanka, but upon discovery of additional species under the sundaicus clade (i.e., *An. subpictus* Form C) from A&N Islands, it was found that this PCR cannot differentiate *An. subpictus* Form B from *An. subpictus* Form C. However, this PCR can discriminate members of subpictus clade (*An. subpictus* A) and sundaicus clade (*An. subpictus* Form B, C, *An. sundaicus* D and *An. epiroticus*) (Fig. 7). This PCR was, therefore, used to differentiate molecular forms A and B of *An. subpictus* in areas where Form C or *An. sundaicus* was absent (mainland India and Sri Lanka).

Table 4 Nucleotide substitution (above diagonal) and pairwise distances (Kimura-2 parameter model) (below diagonal) and between different molecular forms of *An. subpictus* and *An. sundaicus* based on ITS2 sequences

	Sub_A	Sub_B	MT068425	Sub_C	MT068436	Epi	Sun_D	Ind
Sub_A	–	51	49	52	52	47	48	48
Sub_B	0.103	–	2	4	4	5	6	3
MT068425	0.099	0.004	–	4	4	3	4	1
Sub_C	0.105	0.008	0.008	–	0	7	8	5
MT068436	0.105	0.008	0.008	0	–	7	8	5
Epi	0.094	0.010	0.006	0.013	0.013	–	1	4
Sun_D	0.097	0.011	0.008	0.015	0.015	0.002	–	5
Ind	0.097	0.006	0.002	0.009	0.009	0.008	0.010	–

Sub_A: *An. subpictus* Form A; Sub_B: *An. subpictus* Form B; MT068425 (Thailand), Sub_C: *An. subpictus* Form C, MT068436 (Indonesia); Epi: *An. epiroticus* (variant I); Sun_D: *An. sundaicus* D (variant III); Ind: *An. indefinitus*





Distribution of different molecular forms of *An. subpictus* and *An. sundaicus*

The distribution of different molecular forms of *An. subpictus* and *An. sundaicus* in different localities as determined by DNA sequencing as well as by AS-PCR has been shown in Table 2 and Fig. 1. It was observed that *An. subpictus* Form A is prevalent throughout mainland

India (found exclusively in inland areas) and Sri Lanka, whereas Form B was prevalent only in coastal areas or A&N islands. *Anopheles subpictus* Form C is prevalent only in the A&N Islands. *Anopheles sundaicus* D and *An. epiroticus* were found in Car Nicobar of A&N Islands.

Assignment of sibling species based on the mode number of float-ridges on eggs

The designation of sibling species of *An. subpictus* collected from mainland India was assigned based on the mode number of ridges on their egg's float following Suguna et al. [18]. The distribution of sibling species based on the number of ridges and their molecular forms have been provided in Table 5. It was observed that all individuals (n=51) belonging to molecular Form B were correctly assigned to species B based on the mode number of float-ridges. All molecular Form A (n=97) which were identified as species A, C or D based on the mode number of float-ridges, of which majority (80%) fall under the species C category (Fig. 8). The mode number of float-ridges on eggs for form B ranged between 15 and 19, and for Form A between 22 and 33 (Fig. 9a). The number of float-ridges on individual eggs for Form B ranged between 14 and 20, and for Form A ranged between 21 and 35 (Fig. 9b).

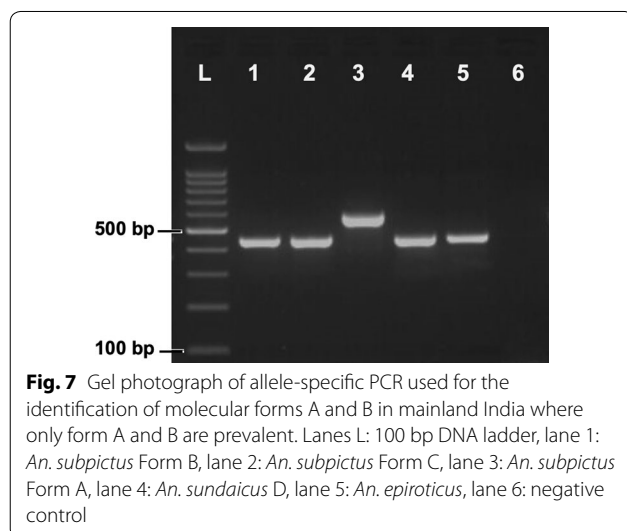


Table 5 Distribution of *An. subpictus* sibling species based on number of float-ridges on eggs and their molecular forms

Localities	Assigned sibling species based on mode number of float-ridges (shown in parenthesis) and their molecular forms									
	Species B (16–20)		Species D (21–24)		Species C (25–29)		Unassigned (30)		Species A (31–36)	
	Form A	Form B	Form A	Form B	Form A	Form B	Form A	Form B	Form A	Form B
Nuh	–	–	7	–	46	–	1	–	–	–
Delhi	–	–	1	–	5	–	4	–	4	–
Jodhpur	–	–	–	–	4	–	2	–	–	–
Ranchi	–	–	–	–	1	–	–	–	–	–
Chilka	–	49	4	–	11	–	–	–	–	–
Puducherry	–	2	4	–	10	–	–	–	–	–

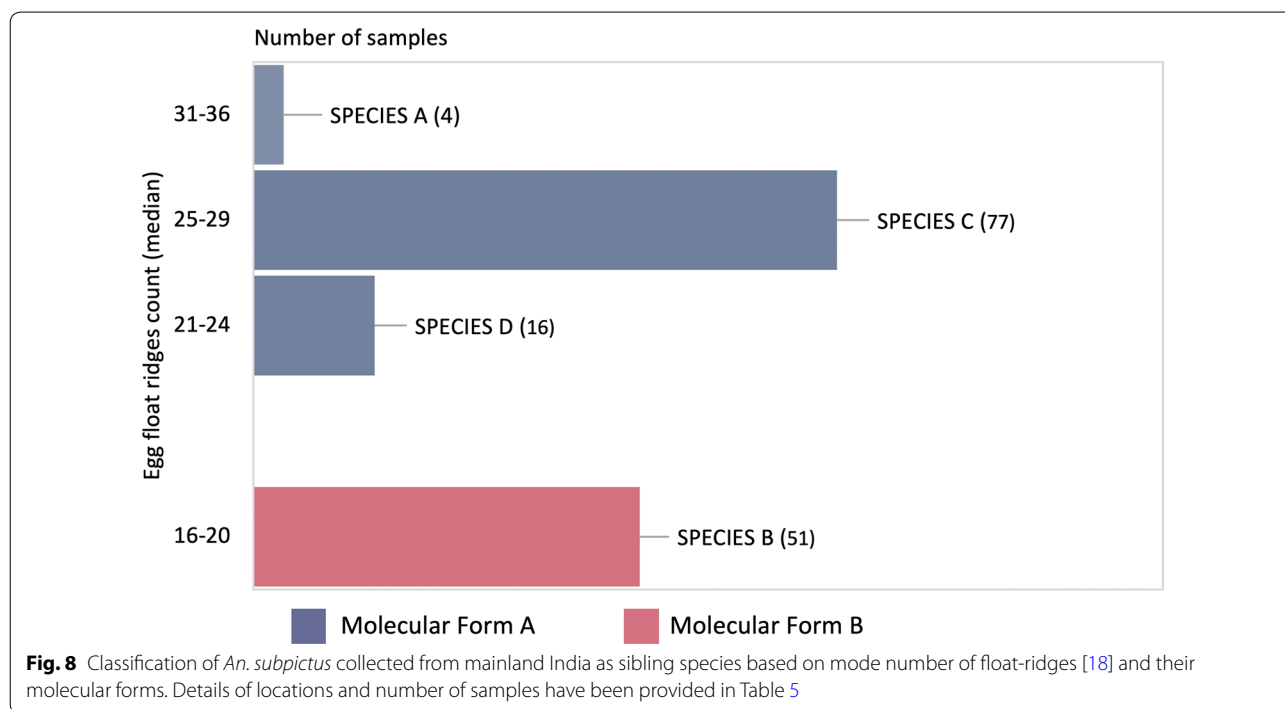


Fig. 8 Classification of *An. subpictus* collected from mainland India as sibling species based on mode number of float-ridges [18] and their molecular forms. Details of locations and number of samples have been provided in Table 5

The frequency distribution of mode number of float-ridges from individual mosquito progeny and number of float-ridges counts from individual egg (pooled data) for molecular Form A and B are shown in Fig. 9a and b, respectively. The figures clearly depict two independent curves, each showing normal distribution, one corresponding to molecular Form B (species B) and another for Form A (comprising of species A, C and D). Based on the distribution pattern of the number of floats-ridges that correlate with molecular forms, it is suspected that species A, C and D might be conspecific while species B is a distinct species.

Seasonal prevalence of *An. subpictus* Form A and B in areas surrounding brackish water lagoon and near coastal area

A survey was carried out in areas surrounding Chilka lake (a brackish water lagoon) and Puducherry (a coastal area) in monsoon and post-monsoon season, to know the relative prevalence of *An. subpictus* Form A and B in monsoon and post-monsoon season. In the Chilka lake area, only *An. subpictus* Form B was found in the dry season (post-monsoon) when there is no freshwater breeding source available except a large brackish water lagoon. In the monsoon season, when fresh water breeding habitat is prevalent beside

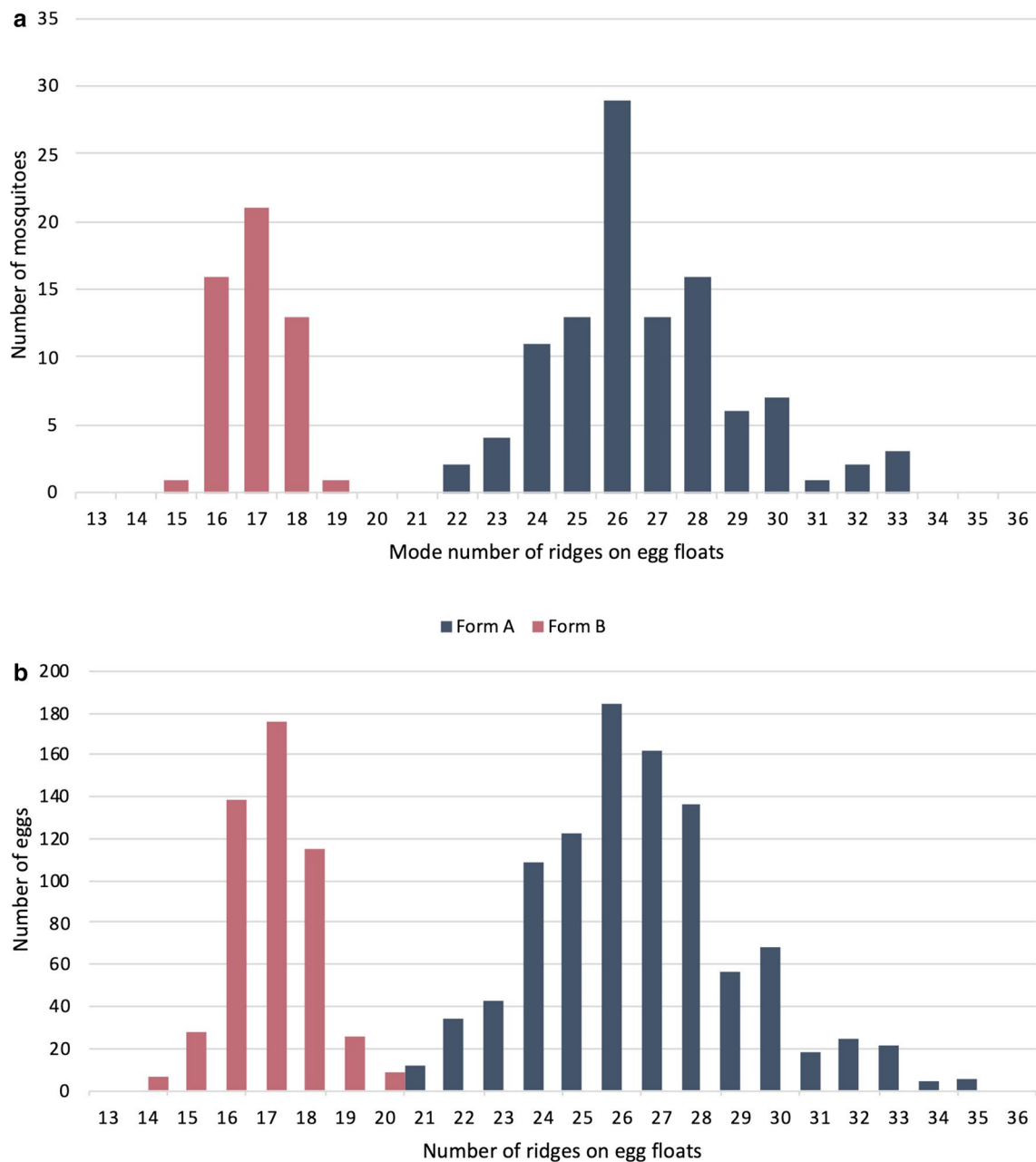


Fig. 9 The frequency distribution of mosquitoes with mode number of floats-ridges (a) and eggs with number of float-ridges (b), among molecular form A and B prevalent in mainland India

brackish water lagoon, Form A was also present (4.7 to 7.6%) alongside Form B. In another area, Puducherry (a coastal area), the proportion of species B was much higher in post-monsoon as compared to autumn. The relative proportion of *An. subpictus* Form A and B in the area surrounding Chilka Lake and coastal area of Puducherry in monsoon and post-monsoon season has given in Table 6. The data indicate that the prevalence

of Form A of *An. subpictus* determined by the availability of freshwater breeding sources.

Discussion

Correct identification of anophelines and their sibling species is crucial for the success of any malaria control programme because of possible inherent variations in the epidemiologically-important biological characteristics

Table 6 Relative prevalence of *An. subpictus* Form A and B in monsoon and post-monsoon season in areas surrounding Chilka Lake (brackish water lagoon) and Puducherry (coastal area)

Location	Season	Date of collection	Number of samples identified as	
			Form A ^a	Form B ^a
Chilka Lake, Odisha, India	Post-monsoon	Feb 2017	0	84 (100)
		Feb 2018	0	94 (100)
	Monsoon	Sept–Oct 2017	4 (4.76)	80 (95.24)
		Sept 2018	16 (6.58)	227 (93.42)
Puducherry, Tamil Nadu, India	Post-monsoon	July–Aug 2019	11 (7.64)	133 (92.36)
		Feb–Mar 2020	8 (11.76)	60 (88.24)
	Autumn	Nov 2019	51 (79.68)	13 (20.31)

^a Figures in parenthesis represent percentage proportion

among sibling species, e.g., vectorial competence [45, 46], insecticide resistance [47–50] and host preference [51, 52]. In India, *An. sundaicus* is considered as an efficient malaria vector, whereas *An. subpictus* has not been recognized so far as a malaria vector by the national malaria control programme [10]. However, published evidences suggest that at least coastal form of *An. subpictus* is a malaria vector [15, 16]. On the other hand, *An. subpictus* is considered as an important malaria vector in coastal Malaysia [14] and Indonesia [12, 13]. These geographically isolated populations differ in breeding preference, where the Malayan population prefers to breed in brackish water, unlike the majority of Indian *An. subpictus* populations that prefer to breeds in freshwater [14]. In view of the presence of such difference in biological characteristics, recognition of sibling species or ‘fixed’ molecular forms in geographically widely distributed *An. subpictus* population is an important consideration in understanding their relative epidemiological significance.

So far, four sibling species have been described from a south Indian population of *An. subpictus* s.l. (provisionally designated as species A, B, C and D) based on the arrangement of inversions present on two loci on the polytene chromosome X [18]. To the best of our knowledge, this technique has not been effectively used for the identification of sibling species in *An. subpictus*, except in a neighbouring country, Sri Lanka [20], where inversions on a single locus (X^{+a} and X^a), out of two diagnostic loci ($X^{a/+a}$ and $X^{b/+b}$), were examined (which preclude the identification of species C and D). Identification of sibling species in malaria control programme is challenging due to the non-availability of simpler technique, as an alternative to the cytological technique that requires a highly specialized skilled personnel to read banding patterns. Moreover, the cytological technique can be applied only on a fraction of a population, i.e., only live semi-gravid female mosquitoes. As an alternative, molecular tools

can be used for the identification of sibling species which are easier and can be applied to dead or alive mosquitoes of either sex. Identification of ‘fixed’ molecular form of rDNA, particularly in ITS2, is a primary step toward the recognition of sibling species. In this direction, ITS2 of *An. subpictus* populations have been characterized in Sri Lankan [27], Thai and Indonesian population [30], however, reliable data on molecular forms from India is lacking. The present study is an attempt to characterize all the molecular forms of morphologically identified *An. subpictus* and *An. sundaicus* prevalent in the Indian subcontinent to uncover hidden cryptic species, distribution pattern and to establish their phylogenetic relationship. Characterization of molecular forms will help in identifying sibling species and developing a reliable and simpler method for their identification. Characterization of *An. sundaicus* was included in this study in view of the report of incongruence in molecular taxonomy with formal (morphological) taxonomy, where one molecular form of *An. subpictus* (mostly species B) was found to be a close relative of *An. sundaicus* [27]. As a result, Surendran et al. [27] recognized them as a member of the Sundaicus Complex. Earlier phylogenetic study on Sri Lankan *An. subpictus* [27, 28] was based on partial ITS2 and D3 domain of 28S rDNA. In this study, full length ITS2 sequence and 28S-D1-D3 was used to identify molecular forms of both *An. subpictus* and *An. sundaicus* from the Indian subcontinent.

This study revealed the presence of a total of five molecular forms in Sundaicus-Subpictus complex in the Indian subcontinent (both by ITS2 as well as 28S-D1-D3): three molecular forms present in morphologically identified *An. subpictus* (Form A, B and C) and two molecular forms in *An. sundaicus* complex. Phylogenetic analysis using two regions of rDNA revealed two distinct clades where *An. subpictus* Form B and C, *An. epiroticus*, all members of Sundaicus Complex and *An. indefinitus* falls

under one clade (Sundaicus) and one molecular form, Form A, falls under distantly related clade (Subpictus). Thus, the phylogenetic relationship showed incongruence with the morphological delimitation of the species. In an earlier study, Surendran et al. [27] showed that most of the *An. subpictus* species B (Form B in this study) from Sri Lanka fall under Sundaicus complex and, therefore, considered as a member of Sundaicus Complex. This study include another molecular form of *An. subpictus*, i.e., Form C, found in A&N Island as a member of Sundaicus Complex. The two molecular forms very recently reported from Indonesia and Thailand [30] also fall in the same clade and closely related to Indian molecular forms Form B and C.

The designation of molecular forms in this study is based on ribosomal DNA (ITS2 and 28SrDNA), where no intraspecific variation in a specific population (due to concerted evolution) was observed, as seen in mitochondrial DNA (maternally inherited and mutations are passed through generations). The absence of individual variation in DNA sequences in molecular forms of *An. subpictus* across wide geographical areas is contrary to the findings by Kaura et al. [31], Chattopadhyay et al. [32] and Chhilar et al. [33], where they recorded individual variation in ITS2 sequences in samples collected from inland India. These sequences are very similar to Form A and it is suspected that individual differences in ITS2 sequences presented in above-mentioned publications is due to errors in the sequence-read of Form A, possibly arising due to the 12 bp indel in one haplotype leading to ambiguous sequence beyond the point of indel in sequence read. Interestingly, the intragenomic variation in Form A, with the presence of two haplotypes, was found fixed throughout its distribution, from north India to Sri Lanka. The fixed molecular forms based on rDNA, which are sympatric, deserve distinct specific status, but the specific status of closely related molecular forms, which are allopatric, is obscure. Thus the two molecular forms A and B found in sympatry in southern Asia (India and Sri Lanka) are genetically isolated but the specific status of other molecular forms found in South-east Asia (A&N Islands, Thailand and Indonesia) is not certain.

Wilai et al. [30] reported that *An. subpictus* from Thailand and Indonesia are different from India and Sri Lanka. They recorded two types of ITS2 sequences from Thailand which were identical except in one type of sequence. There was a mixed nucleotide base at one position. It is unlikely that these two forms are reproductively isolated. The specific status of allopatric molecular forms cannot be ascertained using population genetic analysis, which can be verified using crossing experiments only. Recently, Wilai et al. [30] has carried out crossing experiments in two forms having different mtDNA sequence

(but same molecular forms with respect to ITS2) and found genetic compatibility between them. These two mtDNA forms may be considered as intraspecific variation because such variation is common in mtDNA being inherited maternally. Similar intraspecific variation were also found in this study where multiple haplotypes of *COII* mtDNA were recorded, even in small numbers of samples studied. The molecular forms reported from Thailand and Indonesia were very close to Indian molecular forms B and C with a very low genetic distance ranging from 0.0 to 0.018 (with respect to ITS2). The lack of genetic distance between Indonesian *An. subpictus* and Form C was due to 100% similarity in sequences except presence of two bp indel in Form C.

Interestingly two molecular forms of *An. subpictus s.l.*, Form B and C, which falls under *sundaicus* clade, are found in coastal areas or islands. This may be due to their breeding preference is brackish water, similar to most of the members of *An. sunaicus* complex. This is complemented by the fact that, in Chilka Lake (a brackish lagoon), Form B was found exclusively in Post monsoon season (October and February), while this was present along with Form A in Monsoon season (August–September) when freshwater breeding habitats are abundantly present beside brackish water habitat. Suguna [29] and Panicker et al. [15] also found species B of Subpictus Complex along with species A in the rainy season. These observations indicate that Form B and Form C of *An. subpictus*, found in coastal areas or islands, has a preference to breed in brackish water. Another form of *An. subpictus*, Form A, which is widely distributed throughout inland India, appears to breed in freshwater.

Review of literature shows that species B of *An. subpictus*, found mainly in coastal areas [19, 20], is a potential malaria vector [15, 53]. Recently *An. subpictus* was reported be a potential malaria vector from Goa (coastal area) [16]. This study confirms presence of molecular form B in Goa, It appears that molecular Form B (species B) is playing role in malaria transmission, however, further study is warranted. However, in some inland areas, *An. subpictus* has been found to have sporozoite in their salivary glands [53–55]. The differential role of these molecular forms in malaria transmission need to be confirmed.

An attempt was made to classify sibling species of Form A and B of the Subpictus Complex, prevalent in mainland India based on the mode number of float-ridges. All the molecular Form B were correctly classified as species B while the majority of Form A were characterized as species C. Frequency distribution pattern of float-ridges clearly shows two distinct peaks, one peak corresponds to molecular Form A and another peak corresponds to Form B. It appears that all molecular Form A is a single

species. Because majority of species C identified based on ridges on egg-floats belong to molecular Form A, it is erroneous to assign them as 'Species A' [27, 56]. It is recommended that they may be described as molecular Form A instead of Species A. However, molecular Form B may be designated as 'Species B'. It is evident from our data that molecular method as well as mode number of float ridges can correctly identify species B, and can differentiate them from molecular Form A. However, morphological identification based on egg's ridges has certain limitations. It cannot differentiate species B from *An. sundaicus* due to overlapping range of ridges (range 16 to 20). Moreover, morphological identification based on the number of ridges is a cumbersome process and can be used with only live and gravid female mosquitoes, whereas molecular method can be applied on dead or alive mosquitoes irrespective of life stages and sex.

Anopheles sundaicus is widely distributed species and a malaria vector in the coastal region of Southeast Asia [25]. In India, this is the most prevalent and important malaria vector in the A&N Islands. Earlier studies by Nanda et al. [22] and Alam et al. [23] revealed the presence of only one sibling species in A&N Islands, i.e., *An. sundaicus* D (variant III). This study confirms the presence of *An. sundaicus* D in A&N islands (Port Blair and Car Nicobar). This study also recorded the presence of an additional molecular form in the Sundaicus Complex, *An. sundaicus* (TK), in Car Nicobar island, though just one specimen out of small sample size (n = 52). Further studies are required to explore the presence of this molecular form or new cryptic species in this area. In addition, the presence of *An. epiroticus* s.s. was recorded in Myanmar, which is in agreement with Dusfour et al. [24]. *An. sundaicus* (TK) recorded from Car Nicobar island was identical to *An. epiroticus* but had mixed haplotypes of ITS2, unlike *An. epiroticus* from Myanmar or classic *An. epiroticus*. A similar variant has been reported in Indonesia, Thailand and Vietnam [57, 58]. Dusfour et al. [24] reported combinations of variants I and III, II and III, and I, II and III in continental Southeast Asia. Syafrudin et al. [58] hypothesized that such mixed bases may be due to possible hybridization/introgression between sympatric sibling species.

In this study, a new PCR based assay was developed for the differentiation of members of *sundaicus* and *subpictus* clade. Initially, it was designed to differentiate Form A and B of *Subpictus* Complex but, later upon discovery of Form C, it was found that this could not differentiate Form B and C. However, this PCR can differentiate unambiguously members of *Sundaicus* (*An. subpictus* Form A) and *subpictus* clade (*An. subpictus* Form B, Form C and *An. sundaicus*). A similar ASPCR assay was developed by Surendran et al. [28], but a new assay

was developed because it was noticed that this ASPCR has cross-reactivity with other Indian malaria vector *Anopheles stephensi*, which has superficial resemblance with *An. sundaicus* [26] and often misidentified in old specimens with lost palpal ornamentation and scales.

This study has certain limitation. In absence of data on inversion genotypes of the polytene chromosome of *An. subpictus*, association of molecular forms with previously designated sibling species could not be correlated. Recently one new species *Anopheles pseudosundaicus* belonging to the Pyretophorus series [59] has been designated from Kerala, southwest India, which unlike *An. sundaicus* does not have speckling on legs and closer to *An. subpictus* based on the DNA sequence of cytochrome oxidase C subunit 1. However, access to this molecular data has not been made public. Therefore, it could not be ascertained if this species is one of the molecular forms described in this study. A more elaborate study on molecular characterization of all cytological forms of *An. subpictus* is desirable.

Conclusions

Based on nuclear rDNA sequences, three fixed molecular forms (Form A, B and C) were identified in the morphologically classified *An. subpictus* s.l. and two forms in *An. sundaicus* s.l. in the Indian subcontinent. Molecular phylogenetic analysis revealed two diverse clades of mosquitoes where two molecular forms (B and C) of morphologically identified *An. subpictus* found mainly in coastal areas or islands along with molecular forms reported from Thailand and Indonesia were found to be close relative of *An. sundaicus*. A molecular form of *An. subpictus*, Form A, prevalent in mainland India and Sri Lanka, has fixed intragenomic sequence variation in ITS2 throughout its distribution with the presence of two haplotypes differing by 12 bp indel.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12936-020-03492-2>.

Additional file 1: Fig. S1. Alignment of rDNA sequences (ITS2 to D3 domain of 28S) of different molecular forms of *An. subpictus* s.l., and *An. sundaicus* complex s.l.

Abbreviations

A&N: Andaman & Nicobar; ASPCR: Allele-specific PCR; Bp: Base pair; COI: Cytochrome oxidase I; COII: Cytochrome oxidase II; ITS2: Internal Transcribed Spacer-2; Min: Minute; PCR: Polymerase chain reaction; rDNA: Ribosomal DNA; s: Second; TE: Tris-EDTA.

Acknowledgements

Authors are thankful to Dr. U Pe Htun, Medical Entomology Research Division, Yangon, Myanmar for providing *An. sundaicus* s.l. samples. Thanks are due to Mr. Uday Prakash, Mr. NS Bhakuni and Mrs. Alka Kapoor for processing of

samples and performing PCR reactions; Mr. Kanwar Singh, Mr. Bhupal Ram, Mr. Sadruddin for sample collections.

Authors' contributions

OPS conceived and designed study, designed PCR and sequencing strategies; AS performed PCR, DNA sequencing, phylogenetic analysis; GS performed cloning experiments; MKD and SNS supervised mosquito sampling from field and provided critical suggestions to the manuscript; BRK and HPL supervised the study and contributed to the manuscript. OPS and AS wrote the first draft of the manuscript and all authors provided critical input to the manuscript. All authors read and approved the final manuscript.

Funding

Financial support to this study was provided by the Department of Science and Technology (DST), Government of India, under Women Scientist Scheme-B (Grant No. SR/WOS-B/698/2016).

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ National Institute of Malaria Research, Sector 8 Dwarka, New Delhi 110077, India. ² National Institute of Malaria Research, Field Unit, Itki, Ranchi 835301, India. ³ Department of Zoology, University of Jaffna, Jaffna 40000, Sri Lanka. ⁴ Department of Zoology, Kumaun University, Nainital, India.

Received: 31 July 2020 Accepted: 11 November 2020

Published online: 19 November 2020

References

1. Harbach RE. The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. *Bull Entomol Res.* 2004;94:537–53.
2. Grassi GB. Artcora sulla malaria. *Rendiconti dell' Accademia Lincei.* 1899;8:559–61.
3. Chandra G, Bhattacharjee I, Chatterjee S. A review on *Anopheles subpictus* Grassi, a biological vector. *Acta Trop.* 2010;115:142–54.
4. Rao TR. The Anophelines of India (revised edition). Malaria Research Centre (Indian Council of Medical Research), Delhi. 1984;596 p.
5. Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH, et al. The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. *Parasit Vectors.* 2011;4:8.
6. Dash AP, Adak T, Raghavendra K, Singh OP. The biology and control of malaria vectors in India. *Curr Sci.* 2007;92:1571–8.
7. Singh N, Nagpal BN, Sharma VP. Mosquitoes of Kutch, Gujarat. *Indian J Malariol.* 1985;22:17–20.
8. Dash AP, Hazra RK, Mahapatra N, Tripathy HK. Disappearance of malaria vector *Anopheles sunandaicus* from Chilika Lake area of Orissa State in India. *Med Vet Entomol.* 2000;14:445–9.
9. Iyengar MOT. The distribution of *Anopheles ludlowii* in Bengal and its importance in malaria epidemiology. *Indian J Med Res.* 1931;19:499–524.
10. National Vector Borne Disease Control Programme. Operational manual for integrated vector management in India. 2016, 99 p. (https://nvbdc.gov.in/WriteReadData/1892s/IVM10_March_2016.pdf) Accessed 2 Oct 2020.
11. Collins RT, Jung RK, Anoez H, Sutrisno RH, Putut D. A study of the coastal Malaria vectors, *Anopheles sunandaicus* (Rodenwaldt) and *Anopheles subpictus* Grassi, in South Sulawesi, Sulawesi, Indonesia. Geneva: World Health Organization; 1979.
12. Cooper RD, Edstein MD, Frances SP, Beebe NW. Malaria vectors of Timor-Leste. *Malar J.* 2010;9:40.
13. Ndoen E, Wild C, Dale P, Sipe N, Dale M. Relationships between anopheline mosquitoes and topography in West Timor and Java, Indonesia. *Malar J.* 2010;9:242.
14. Reid JA. A note on *Anopheles subpictus* Grassi and *A indefinitus* Ludlow (Diptera: Culicidae). *J Med Entomol.* 1966;3:327–31.
15. Panicker KN, Bai MG, Rao USB, Viswam K, Suryanarayanamurthy U. *Anopheles subpictus*, vector of malaria in coastal village of South-East India. *Curr Sci.* 1981;50:694–5.
16. Kumar A, Hosmani R, Jadhav S, Sousa TD, Mohanty A, Naik M, et al. *Anopheles subpictus* carry human malaria parasites in an urban area of Western India and may facilitate perennial malaria transmission. *Malar J.* 2016;15:124.
17. Amerasinghe PH, Amerasinghe FP, Wirtz RA, Indrajith NG, Somapala W, Pereira LR, Rathnayake AMC. Malaria transmission by *Anopheles subpictus* Grassi in a new irrigation project in Sri Lanka. *J Med Entomol.* 1992;29:577–81.
18. Suguna SG, Rathinam GK, Rajavel AR, Dhanda V. Morphological and chromosomal description of new species in the *Anopheles subpictus* complex. *Med Vet Entomol.* 1994;8:88–94.
19. Reuben R, Kalyanasundaram M, Suguna SG. Salinity tolerance of sibling species in the taxon *Anopheles subpictus* Grassi 1899. *Indian J Med Res.* 1984;80:67–70.
20. Abhayawardana TA, Wijesuriya SR, Dilrukshi RR. *Anopheles subpictus* complex: distribution of sibling species in Sri Lanka. *Indian J Malariol.* 1996;33:53–60.
21. Linton YM, Dusfour I, Howard TM, La Ruiz F, Manh ND, Dinh TH, et al. *Anopheles (Cellia) epiroticus* (Diptera: Culicidae), a new malaria vector species in the Southeast Asian Sundaicus Complex. *Bull Entomol Res.* 2005;95:329–39.
22. Nanda N, Das MK, Wattal S, Adak T, Subbarao SK. Cytogenetic characterization of *Anopheles sunandaicus* (Diptera: Culicidae) population from Car Nicobar Island, India. *Ann Entomol Soc Am.* 2004;97:171–6.
23. Alam MT, Das MK, Ansari MA, Sharma YD. Molecular identification of *Anopheles (Cellia) sunandaicus* from the Andaman and Nicobar Islands of India. *Acta Trop.* 2005;97:10–8.
24. Dusfour I, Michaux JR, Harbach RE, Manguin S. Speciation and phylogeography of the Southeast Asian *Anopheles sunandaicus* complex. *Infect Genet Evol.* 2007;7:484–93.
25. Manguin S, Garros C, Dusfour I, Harbach RE, Coosemans M. Bionomics, taxonomy, and distribution of the major malaria vector taxa of *Anopheles* subgenus *Cellia* in Southeast Asia: an updated review. *Infect Genet Evol.* 2008;8:489–503.
26. Christophers SR. The fauna of British India, including Ceylon and Burma, Vol IV., Diptera, Family Culicidae, Tribe Anophelini London: Taylor and Francis. 1933;371 p.
27. Surendran SN, Singh OP, Jude PJ, Ramasamy R. Genetic evidence for malaria vectors of the *Anopheles sunandaicus* complex in Sri Lanka with morphological characteristics attributed to *Anopheles subpictus* species B. *Malar J.* 2010;9:343.
28. Surendran SN, Sarma DK, Jude PJ, Kempainen P, Kanthakumaran N, Gajapathu K, et al. Molecular characterization and identification of members of the *Anopheles subpictus* complex in Sri Lanka. *Malar J.* 2013;12:304.
29. Suguna SG. Cytological and morphological evidences for sibling species in *Anopheles subpictus* Grassi. *J Commun Dis.* 1982;14:1–8.
30. Wilai P, Ali RSM, Saingamsook J, Saeung A, Junkum A, Walton C, et al. Integrated systematics of *Anopheles subpictus* (Diptera: Culicidae) in the Oriental Region, with emphasis on forms in Thailand and Sulawesi, Indonesia. *Acta Trop.* 2020;208:1–8.
31. Kaura T, Sharma M, Chaudhry S, Chaudhry A. Sequence polymorphism in spacer ITS2 of *Anopheles (Cellia) subpictus* Grassi (Diptera: Culicidae). *Caryologia.* 2010;63:124–33.
32. Chattopadhyay A, Paul S, Banerjee PK. Sequence variation in spacer ITS 2 of *An. subpictus* in rural and urban populations of West Bengal. *J Int Acad Res Multidiscipl.* 2013;9:502–12.
33. Chhilar JS, Chaudhry S. Phylogenetic analysis of *Anopheles (Cellia) subpictus* Grassi using rDNA-ITS2 Sequence. *Proc Zool Soc.* 2012;65:1–10.

34. Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics*. 2007;175:477–85.
35. Hwang UW, Kim W. General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. *Korean J Parasitol*. 1999;37(4):215–28.
36. Beebe NW. DNA barcoding mosquitoes: advice for potential prospectors. *Parasitology*. 2018;145:622–33.
37. Wiemers M, Fiedler K. Does the DNA barcoding gap exist?—a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front Zool*. 2007;4:8.
38. Meier R, Shiyang K, Vaidya G, Ng PK. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst Biol*. 2006;55:715–28.
39. Reid JA. Anopheline Mosquitoes of Malaya and Borneo. Studies from the Institute for Medical Research. Malaysia, No.31, 1968;520 pp.
40. Coen ES, Strachan T, Dover G. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the melanogaster species subgroup of *Drosophila*. *J Mol Biol*. 1982;158:17–35.
41. Singh OP, Chandra D, Nanda N, Htun PT, Adak T, Subbarao SK, Dash AP. On the conspecificity of *Anopheles fluviatilis* S with *Anopheles minimus* C. *J Biosci*. 2006;31:671–7.
42. Thompson JD, Higgins DG, Gibson TJ, Clustal W. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22:4673–80.
43. Swofford DL. PAUP: Phylogenetic Analysis Using Parsimony (and Others Methods) Version 4.0b10.D.L. Sunderland: Sinauer Associates; 2002.
44. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 1998;14:817–8.
45. Adak T, Kaur S, Singh OP. Comparative susceptibility of different members of *Anopheles culicifacies* complex to *Plasmodium vivax*. *Trans R Soc Trop Med Hyg*. 1999;93:573–7.
46. Kaur S, Singh OP, Adak T. Susceptibility of species A, B, C of *Anopheles culicifacies* complex to *Plasmodium yoelii yoelii* and *Plasmodium vinckei petteri* infections. *J Parasitol*. 2000;86:1345–8.
47. Raghavendra K, Vasantha K, Subbarao SK, Pillai MK, Sharma VP. Resistance in *Anopheles culicifacies* sibling species B and C to malathion in Andhra Pradesh and Gujarat States, India. *J Am Mosq Control Assoc*. 1991;7:255–9.
48. Raghavendra K, Subbarao SK, Vasantha K, Pillai MK, Sharma VP. Differential selection of malathion resistance in *Anopheles culicifacies* A and B (Diptera: Culicidae) in Haryana State, India. *J Med Entomol*. 1992;29:183–7.
49. Subbarao SK, Vasantha K, Sharma VP. Responses of *Anopheles culicifacies* sibling species A and B to DDT and HCH in India: implications in malaria control. *Med Vet Entomol*. 1988;2:219–23.
50. Surendran SN, Jude PJ, Weerarathne TC, Karunaratne SHPP, Ramasamy R. Variations in susceptibility to common insecticides and resistance mechanisms among morphologically identified sibling species of the malaria vector *Anopheles subpictus* in Sri Lanka. *Parasit Vectors*. 2012;5:34.
51. Joshi H, Vasantha K, Subbarao SK, Sharma VP. Host feeding pattern of *Anopheles culicifacies* species A and B. *J Am Mosq Control Assoc*. 1988;4:248–51.
52. Nanda N, Joshi H, Subbarao SK, Yadav RS, Shukla RP, Dua VK, Sharma VP. *Anopheles fluviatilis* complex: host feeding pattern of species S, T and U. *J Am Mosq Control Assoc*. 1996;12:147–9.
53. Kumari S, Das S, Mahapatra N. *Anopheles subpictus* B and its role in transmission of malaria in Odisha, India. *Trop Biomed*. 2013;30:710–7.
54. Chatterjee S, Chandra G. Role of *Anopheles subpictus* as primary vector of malaria in an area in India. *Japan J Trop Med Hyg*. 2000;28:177–81.
55. Kulkarni SM. Detection of sporozoites in *Anopheles subpictus* in Baster district, Madhya Pradesh. *Indian J Malariol*. 1983;20:159–60.
56. Singh OP, Dykes CL, Sharma G, Das MK. L1014F-kdr mutation in Indian *Anopheles subpictus* arising from two alternative transversions in the voltage gated sodium channel and a single PIRA-PCR for their Detection. *J Med Entomol*. 2015;52:24–7.
57. Zarowiecki M, Linton YM, Post RJ, Bangs MJ, Htun PT, Hlaing T, et al. Repeated landmass reformation limits diversification in the widespread littoral zone mosquito *Anopheles sundaicus* sensu lato in the Indo-Oriental Region. *Mol Ecol*. 2014;23:2573–89.
58. Syafruddin D, Lestari YE, Permana DH, Asih PBS, Laurent B, Zubaidah S, et al. *Anopheles sundaicus* complex and the presence of *Anopheles epiroticus* in Indonesia. *PLoS Negl Trop Dis*. 2020;14:e0008385.
59. Tyagi BK, Hiriyani J, Tewari SC, Ayanar K, Samuel P, Paramasivan R. Description of a new species, *Anopheles pseudosundaicus* (Diptera: Culicidae) from Kerala, India. *Zootaxa*. 2009;2219:49–60.
60. Beebe NW, Saul A. Discrimination of all members of the *Anopheles punctulatus* complex by polymerase chain reaction—restriction fragment length polymorphism analysis. *Am J Trop Med Hyg*. 1995;53:478–81.
61. Campbell BC, Steffen-Campbell JD, Werren JH. Phylogeny of the *Nasonia* species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed spacer (ITS2) and 28S rDNA sequences. *Insect Mol Biol*. 1993;2:225–37.
62. Litvaitis MK, Nunn G, Thomas WK, Koche TD. A molecular approach for the identification of *Meiofaunal turbellarians* (Platyhelminthes, Turbellaria). *Marine Biol*. 1994;120:437–42.
63. Sharpe RG, Harbach RE, Butlin RK. Molecular variation and phylogeny of members of the Minimus group of *Anopheles* subgenus *Cellia* (Diptera: Culicidae). *Syst Entomol*. 2000;25:263–72.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

