DNA barcoding of selected short-horned grasshoppers (Orthoptera: Acrididae) from Indian Himalayan region

Shantanu Kundu, Hirdesh Kumar, Kaomud Tyagi, Kailash Chandra and Vikas Kumar

Centre for DNA Taxonomy, Molecular Systematics Division, Zoological Survey of India, Kolkata, India

ABSTRACT
In the context of Indian zoogeography, the DNA barcode data of short-horned grasshoppers (family Acrididae) are limited in global databases. Hence, the present study was aimed to collect selected Acridid species from the Indian Himalayan regions and generate DNA barcode data to enrich the global database. The estimated K2P genetic distances, Bayesian analysis (BA) topology and multiple species delimitation methods (ABGD, bPTP, and GMYC) clearly discriminate all the studied species. Based on high genetic distance (7.5%), multiple clades, and more than one molecular operational taxonomic unit, the present study elucidates the allopatric speciation and presence of possible cryptic diversity of Oxya japonica within India, China, and Russia. The present study suggests the collection of multiple specimens from different geographical locations and the generation of more DNA barcode data would facilitate the actual diversity of this insect group.

Introduction
Orthoptera is one of the oldest living insect lineages with 28,418 described species and contributes almost half of the total arthropods diversity globally (Grimaldi and Engel 2005; Cigliano et al. 2020). They are generally inhabits in the grass-land ecosystems and play an important role in maintaining the food chain (Belovsky and Slade 1993; Joern et al. 2006). Due to their high mobility across different geographical locations and ecological sensibilities, this group of insects also acts as a bio-indicator (Samways and Sergeev 1997; Bazelet and Samways 2014). Among the extant diversity, most of the species were considered as major pests in agriculture, animal husbandry, and forest ecosystem (Uvarov 1966; COPR 1982). In recent past, the United Nations’ Food and Agriculture Organization (FAO) declared that a single Acridid species has made upsurge invasion and worst outbreak as well as damaged several crops in India and neighboring countries. However, few of them were used as a bio-control agent; e.g. the semi-aquatic grasshopper species, Cornops aquaticum has been used as a bio-control agent on aquatic plant (Bownes et al. 2013) and the meadow purple-striped grasshopper, Hesperotettix viridis feed the noxious snakeweeds which can harm cattle and other livestock (Thompson and Richman 1993). Besides its ecological and economic significance, the family Acrididae also has a biogeographic history and assumed to have originated in the early Cenozoic Era and diversified through mid to late Cenozoic (Song et al. 2015).

The short-horned grasshopper (family Acrididae) is the largest and most diverse lineage within the suborder Caelifera and comprised around 6700 species within 26 subfamilies worldwide (Cigliano et al. 2020). So far, India shares 2.4% of the global orthopteran diversity, and the family Acrididae is represented by 285 species under 135 genera (Shishodia et al. 2018). However, majority of the studies have been focused on three subfamilies, Melanoplinae (Chapco and Contreras 2011; Song et al. 2018), Oedipodinae (Chapco et al. 2003), and Oedipodinae (Song et al. 2019). As of now, several studies have been conducted to elucidate the accurate species identification, estimation of genetic divergence, phylogenetic relationship, detection of cryptic diversity, and species complexes of this insect group (Wang et al. 2008; Chapco and Contreras 2011; Song et al. 2018). However, majority of the studies have been focused on three subfamilies, Melanoplinae (Chapco et al. 2001; Amédégénoto et al. 2003), Oedipodinae (Chapco et al. 1997; Fries et al. 2007), and Gomphocerinae...
(Bugrov et al. 2005; Chapco and Contreras 2011). Nevertheless, the generation of DNA barcode data of several Acrididae species is still underway from different geographical regions especially from India, which restrict the assumption on their actual diversity. Hence, the present study was attempted to generate 20 partial mitochondrial cytochrome c oxidase subunit I (mtCOI) sequences of six morphologically identified Acridid species from the Indian Himalayan regions (IHRs). We used genetic distance matrix, Bayesian analysis (BA), and multiple species delimitation methods to elucidate their genetic diversity. This is the first and preliminary attempt which advocated for further large-scale effort to illuminate the extant orthopteran diversity from the IHRs.

Materials and methods

Total 20 short-horned grasshoppers specimens were sampled from five different locations (27.49N 96.39E, 27.54N 96.44E, 27.53N 91.67E, 27.71N 91.73E, 27.53N 91.72E) in Arunachal Pradesh state in the eastern Himalayan region (Figure 1(A)). The expedition was conducted under the National Mission on Himalayan Landscape’ project ID: NMHS/2015-16/LG-05; project grant number: NMHS/LG2016/0011/8509 at Zoological Survey of India (ZSI) (Kolkata, India). The hind leg of each specimen was dissected through surgical blade and preserved in 70% ethanol at –20 °C for further molecular analysis. The specimens were dried, pinned, preserved, and identified up to species level by reviewing the available literatures (Bolivar 1914; Kirby 1914; Ramme 1940; Hollis 1971; Jago 1984; Hollis 2009; Swaminathan et al. 2018) (Table S1). Nikon digital camera (D-7000) was used to acquire the photograph of the representative species. The voucher specimens were stored in the National Zoological Collections of the Orthoptera section, ZSI (Kolkata, India).

Total genomic DNA was extracted from each morphologically identified specimen in 500 μl of tissue lysis buffer containing 50 mM Tris–HCl, 25 mM EDTA, 150 mM NaCl, and proteinase K (200 μg/ml) through phenol–chloroform extraction method (Sambrook and Russell 2006). To generated the barcode data, degenerative primer pair: COBU (5’- TyTcaaacaycayaargatatgg-3’) and COBL (5’- TaacttggWGrTgWCCAARATCA-3’) was used to amplify the partial fragment of mitochondrial cytochrome c oxidase subunit I (COI) gene (Pan et al. 2006). The amplification was performed in 30 μl reaction mixture of 100 ng of DNA template, 1× PCR buffer, 2 mM MgCl2, 10 pmol of each primer, 0.25 mM of each dNTPs, 0.25 U of high-fidelity Taq DNA polymerase (Applied Biosystems Inc., Foster City, CA), and nuclease free water. The amplification was carried out in VeritiVR Thermal Cycler (Applied Biosystems, Foster City, CA) with the following thermal profile: an initial denaturation step at 95°C for 5 min, followed by 30–35 cycles of denaturation at 94°C for 45 s, annealing at 48°C for 45 s, and extension at 70°C for 1 min 30 s, and a final extension at 72°C for 10 min and then held at 4°C. The PCR products were purified by QiAguickR Gel extraction kit (Qiagen Inc., Germantown, MD) with manufactures protocol. The cycle-sequencing was performed by using BigDye® Terminator ver. 3.1 Cycle Sequencing Kit (Applied Biosystems, Inc., Foster City, CA) and 3.2 pmol of each primer on the same thermal cycler. The cycle-sequencing products were purified by using BigDye X-terminator kit (Applied Biosystems Inc., Foster City, CA, USA) and sequenced bi-directionally through 48 capillary ABI 3730 Genetic Analyzer housed at ZSI (Kolkata, India).

Both forward and reverse contigs of each sample were screened by Sequence Scanner (Applied Biosystems, Foster City, CA, USA) to generate the consensus sequences. Each sequence was examined through the nucleotide BLAST program (https://blast.ncbi.nlm.nih.gov) and ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/) to diminish the indels, mismatches, and start-stop codons. The generated sequences were submitted in the GenBank database through Bankit submission tool (https://www.ncbi.nlm.nih.gov/WebSub/?tool=genbank) to acquire the accession numbers. Further, to make an analytical dataset, total 26 barcode sequences of same or related species under four targeted subfamilies were downloaded from GenBank database. The barcode sequence of Dericorys annulata: accession no. NC_046555 (Family: Dericorythidae) was also acquired from GenBank and used as an out-group in the present topology analysis to test the monophyletic criterion. The final dataset were aligned using ClustalX program (Thompson et al. 1997) and the genetic divergences were estimated by Kimura 2 parameter (K2P) in MEGA7 (Kumar et al. 2016). The best fit model for all three codon positions ‘GTR + I + G’ was estimated by observing the lowest BIC value through PartitionFinder version 2.1.1 (Lanfear et al. 2012). To check the monophyletic clustering of the studied taxa, BA was followed by using Mr Bayes 3.1.2 (Ronquist and Huelsenbeck 2003). The dataset was run for 50,000,000 generations with 25% burn-in with trees saving at every 100 generation. The Markov Chain Monte Carlo (MCMC) analysis was used to generate the convergence metrics, till the standard deviation of split frequencies reached 0.01 and the potential scale reduction factor for all parameters bordered on 1.0. To visualize the topology with better representation, the tree was modified through web based iTOL tool (https://itol.embl.de/) (Letunic and Bork 2007). In addition to understand the below species-level diversity, three species delimitation methods: Automatic Barcode Gap Discovery (ABGD) (Puillandre et al. 2012), the General Mixed Yule-coalescent (GMYC) (Fujisawa and Barraclough 2013), and Poisson-Tree-Processes (bPTP) (Zhang et al. 2013) were executed to estimate the molecular operational taxonomic units (MOTUs). The ABGD analysis was carried out on the web interface (www.abi.snv.jussieu.fr/public/abgd/) with Simple, K80 Kimura, and JC69 Jukes-Cantor Distance with relative gap width (X = 1.5). The ultrametric tree was generated by BEAST program (Drummond and Rambaut 2007 ) with GTR + I + G model and the Tree Annotator was used for visualizing the output tree. For GMYC analysis, multiple threshold was carried out in RStudio (https://www.r-project.org/) using packages like ‘ape’ and ‘splits’ (Paradis et al. 2004; Ezard et al. 2009). The maximum-likelihood tree was further built in RAxML (Stamatakis 2006) to execute the bPTP analysis on web interface (http://species.h-its.org/ptp/).
Results and discussion

Many large-scale attempts have been made to generate the DNA barcode data of orthopterans from diverse geographical regions around the world. However, a single attempt has been made from India focusing on the family Tettigoniidae so far (Muhammedali et al. 2017). This restricts the knowledge on the actual diversity of orthopterans from the IHRs. The present study intended to enrich the global barcode library of the orthopteran insect through an integrated approach from the IHRs.
The 20 studied samples were identified as six morphospecies under four Acrididae subfamilies (Acridinae, Catantopinae, Oedipodinae, Oxyinae) (Figure 1(B)). Among the generated DNA barcodes, seven sequences of *Phlaeoba sikkimensis* are novel contributions to the global database. To confirm the molecular identification of the studied taxa and monophyletic criteria, the genetic divergence and topological clustering were tested for the final dataset (46 DNA barcode sequences of 20 species under four Acridid subfamilies). The overall mean genetic distance of the studied dataset was 13.9%. Excluding the singleton taxa, the mean intra-species genetic distance was ranging from 0% (*Ceracris kiangsu, Phlaeoba tenebrosa*) to 7.5% (*Oxya japonica*).

The highest inter-species genetic distance (20.8%) was depicted between *Xenocatantops humilis* and *Heteropternis couloniana*. Further, the significantly low genetic distances were observed in few Oxyinae species: *Oxya chinensis* and *Oxya velox* (0.3%), *Oxya sinuosa* and *O. velox* (0.2%), as well as *O. sinuosa* and *O. chinensis* (0.5%). The low genetic distance was also detected for two Acridinae species, *Phlaeoba infumata* and *Phlaeoba albonema* (0.5%). Due to the low sample coverage and use of single gene analysis, we suggest multiple specimens and additional molecular information is required to validate the systematics status of these species. The present BA analysis showed distinct clades for all the studied species with high posterior probability support which is congruent with the earlier cladistics hypothesis (Song et al. 2018). The multiple species delimitation methods: ABGD, GMYC, and bPTP yielded 17 (Table S2), 19 (Table S3), and 22 (Table S4) MOTUs, respectively (Figure 1(C)).

Moreover, the previous studies suggested that >3% intra-species genetic distances can be used as a threshold for delimiting arthropod species (Hebert et al. 2003; Huang et al. 2013). However, this barcode gap may vary among different taxa due to the unlike molecular evolutionary rate within the mitochondrial genes (Wiemers and Fiedler 2007; Meier et al. 2008). Adding to this, a recent molecular study also implied different species delimitation methods for resolving cryptic species and species complex among different Orthopteran families (Zhou et al. 2019). However, the present study detected considerable high intra-species genetic distance (7.5%) and more than one MOTU in *O. japonica*, which indicates the presence of possible cryptic diversity (Figure 1(C)). This Oxyinae species is commonly known as rice grasshopper and originally described from Japan. The species act as a pest on rice and widely distributed throughout Asia, Africa, Northern Africa, and Algeria (Cigiano et al. 2020). A total of two specimens of *O. japonica* were collected from Tawang, Arunachal Pradesh in the eastern Himalayan region. The studied specimens of *O. japonica* along with the available database sequences (KJ261378 collected from Russia and MF125299 collected from China) showed polytomy in BA tree and more than one MOTU in ABGD, bPTP, and GMYC analyses. By superimposing all species delimitation methods, the present molecular-based investigation presumed the presence of cryptic diversity of *O. japonica* in India, China, and Russia. Further, the collection localities of *O. japonica* from distant geographical areas suggest the occurrence of allopatric speciation of this species within its range-distribution. In conclusion, the present study contributed first DNA barcode data of morphologically identified selected Acridid species from the IHRs. We encourage more rigorous sampling of this taxonomic group from different geographical regions and generation of more barcode data to elucidate their actual species diversity.

**Acknowledgements**

The authors are thankful to the Director, Zoological Survey of India (ZSI), Ministry of Environment, Forests and Climate Change, Govt. of India for providing necessary facilities, constant support, and encouragement throughout the study. The second author (SK) acknowledges the fellowship grant received from the Council of Scientific and Industrial Research (CSIR) Senior Research Associateship (Scientists’ Pool Scheme) Pool No. 9072-A. We thank Devkant Singh and Avas Pakrashi for analytical help and Pronomay Karmakar for helping in improving the language.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

The research is funded by National Mission on Himalayan Studies (NMHS), Ministry of Environment and Climate Change, New Delhi large grant research project (Ref. No. NMHS/LG-2016/0011) entitled ‘Biodiversity Assessment through long-term monitoring plots in the Indian Himalayan Landscape’. The second author (SK) acknowledges the fellowship grant received from the Council of Scientific and Industrial Research (CSIR) Senior Research Associateship (Scientists’ Pool Scheme) Pool No. 9072-A.

**ORCID**

Shantanu Kundu http://orcid.org/0000-0002-5488-4433
Hirdesh Kumar http://orcid.org/0000-0003-3152-2496
Koamd Tyagi http://orcid.org/0000-0003-1064-9826
Kailash Chandra http://orcid.org/0000-0001-9076-5442
Vikas Kumar http://orcid.org/0000-0002-0215-0120

**Data availability statement**

The data that support the findings of this study are openly available in NCBI GenBank database at https://www.ncbi.nlm.nih.gov with the accession numbers MT325817, MT325819–MT325837 which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**References**

Amédégnato C, Chapco W, Litzenberger G. 2003. Out of South America? Additional evidence for a southern origin of melanopline grasshoppers. Mol Phylogenet Evol. 29(1):115–119.
Bazelet CS, Samways MJ. 2014. Habitat quality of grassland fragments affects dispersal ability of a mobile grasshopper, *Orinthacris cyanea* (Orthoptera: Acrididae). Afr Entomol. 22(4):714–725.
Bellmann H. 2006. Der Kosmos-HeuschreckenFührer. Stuttgart: Kosmos. Bolovsky GE, Slade JB. 1993. The role of vertebrate and invertebrate predators in a grasshopper community. Oikos. 68(2):193–201.
Bolivar I. 1914. Estudios entomológicos. Segunda parte. I. El grupo de los Euprepocnemes. II. Los Truxalinos del antiguo mundo. Trabajos del Museo de Ciencias Naturales (Serie Zoológica). 20:110.
Wang NX, Feng X, Jiang GF, Fang N, Xuan WJ. 2008. Molecular phylogenetic analysis of five subfamilies of the Acrididae (Orthoptera: Acridoidea) based on the mitochondrial cytochrome b and cytochrome c oxidase subunit I gene sequences. Acta Entomol Sin. 51: 1187–1195.

Wiemers M, Fiedler K. 2007. Does the DNA barcoding gap exist? A case study in blue butterflies (Lepidoptera: Lycaenidae). Front Zool. 4:8.

Zhang J, Kapli P, Pavlidis P, Stamatakis AA. 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics. 29(22):2869–2876.

Zhou Z, Guo H, Han L, Chai J, Che X, Shi F. 2019. Singleton molecular species delimitation based on COI-5P barcode sequences revealed high cryptic/undescribed diversity for Chinese katydids (Orthoptera: Tettigoniidae). BMC Evol Biol. 19(1):79.