

STANDARDIZED METHODOLOGY FOR COLLECTING, CURATING, AND PREPARING PENTATOMIDAE OF THE FERGANA VALLEY FOR DNA BARCODING

Mirmukhsin Makhmudov Uchqunbek o'g'li ^{1*}

Dadarahim G'aniyev Zafarjon o'g'li ²

¹ Doctoral Researcher, Fergana State University, Fergana, Uzbekistan

E-mail: mirmuhsinmahmudov1994@gmail.com; ORCID: 0000-0002-3013-6684

² Independent Researcher, Uzbekistan

E-mail: dadarahim80@gmail.com

* Corresponding author: mirmuhsinmahmudov1994@gmail.com

Abstract

Pentatomidae (Hemiptera: Heteroptera) are ecologically and economically important components of agroecosystems and natural landscapes in the Fergana Valley. This methodological paper presents a standardized workflow that integrates field collection, morphological voucher preparation, COI-based DNA barcoding, and deposition of interoperable biodiversity data. The protocol defines habitat- and season-linked sampling, measurable sampling effort for each collection method, a Darwin Core-compatible metadata package, a single-identifier architecture linking the voucher, images, tissue subsample, and sequence, and explicit quality-control checkpoints. Rather than treating one preservation regime as universally optimal, the workflow incorporates the documented trade-off between morphological integrity and DNA preservation and requires preservation strategy to be matched to the taxon, storage duration, and downstream analytical goal. For reference barcode records, the workflow follows core BOLD expectations, including a retrievable voucher, unique specimen identifier, complete collection metadata, and a quality-controlled COI sequence. The protocol provides a reproducible basis for inventorying the Fergana Valley Pentatomidae fauna, monitoring invasive and economically important species, and generating interoperable records for GBIF, BOLD, and GenBank.

Keywords: Pentatomidae; Fergana Valley; entomological monitoring; voucher; Darwin Core; COI; DNA barcoding; BOLD; GBIF; data interoperability.

Introduction

Pentatomidae Leach, 1815 is one of the largest families within the superfamily Pentatomoidea of the heteropteran suborder of Hemiptera. Modern syntheses recognize approximately 940 genera and 4,949 species (Rider et al., 2018). In addition to phytophagous pest species, the family includes predatory taxa belonging to the subfamily Asopinae. In agroecosystems, *Aelia* spp., *Nezara viridula*, *Dolycoris baccarum*, and other pentatomids may be economically important, while *Halyomorpha halys* is among the most important invasive pentatomids globally (McPherson,



2018).

The Fergana Valley is characterized by the close juxtaposition of irrigated lowlands, orchards and vineyards, riparian habitats, foothill zones, and montane landscapes. This landscape mosaic provides an opportunity to study feeding, reproduction, dispersal, and overwintering strategies of Pentatomidae under contrasting ecological conditions. *H. halys* has been recorded from Uzbekistan, including a record from the Fergana area in 2022, reinforcing the need for a standardized monitoring framework for this invasive species in the valley (Gandjaeva et al., 2022). Classical entomological manuals describe general procedures for collecting and curating insects in detail (Martin, 1977; Upton, 1991). Modern biodiversity research, however, requires three additional elements: recording sampling effort in reproducible units; linking the voucher, images, tissue subsample, and DNA sequence through a persistent identifier; and structuring biodiversity data in a form interoperable with Darwin Core and barcode platforms. In molecular workflows, preservation should not be treated as a single rigid formula; instead, the trade-off between morphology and DNA integrity must be considered (Marquina et al., 2021).

The aim of this paper is to improve and systematize a reproducible protocol adapted to the Fergana Valley that covers the full pathway from field collection to morphological voucher preparation and COI barcode records while remaining compatible with international biodiversity-data standards. The paper is methodological in nature and does not claim empirical estimates of population size or species richness; its principal result is a standardized workflow with explicit quality-control and data-linkage rules.

Standardized workflow for Pentatomidae fieldwork and DNA barcoding in the Fergana Valley

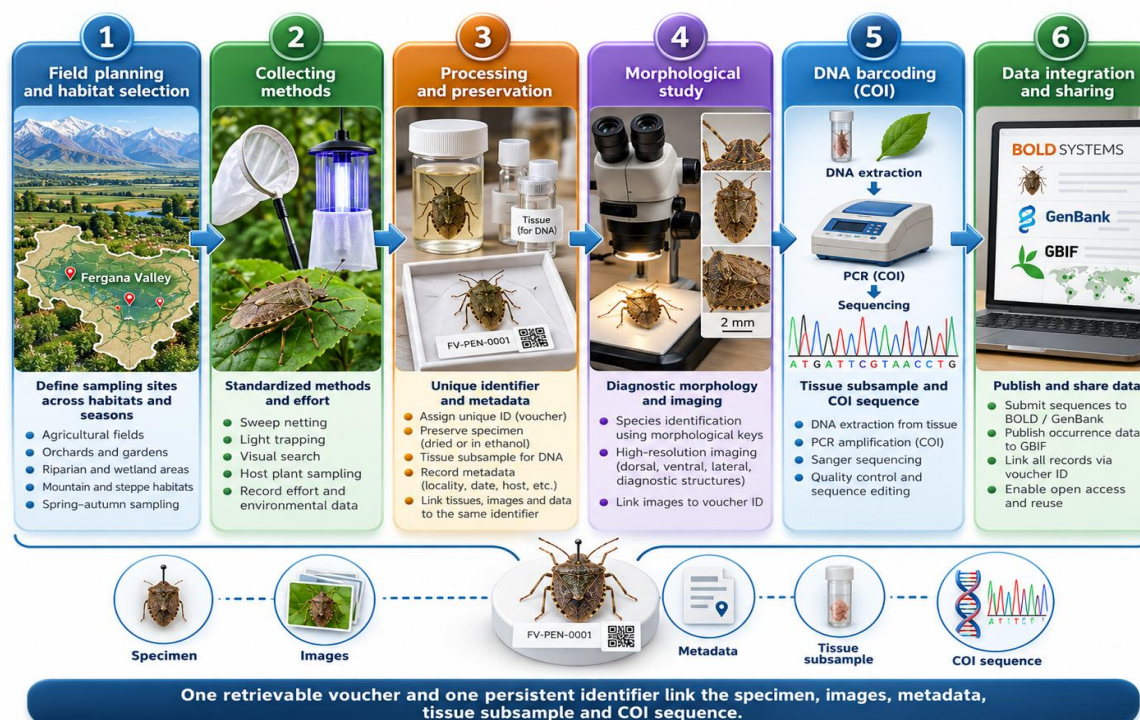


Figure 1. Conceptual workflow for managing Fergana Valley Pentatomidae material from field collection to interoperable international biodiversity data.

Source: author-developed synthesis based on Darwin Core and barcode-data integration principles.

2. Materials and Methods

The methodology was adapted to agroecosystems and natural landscapes of the Andijan, Namangan, and Fergana regions of the Fergana Valley. Orchards and vineyards, vegetable and melon fields, legume and forage crops, natural grasslands and foothills, riparian/tugai habitats, and anthropogenic environments are recorded as distinct habitat categories. For each sampling point, habitat description, host or food plant, elevation, coordinate precision, and sampling effort are documented in a standardized format. Seasonal monitoring covers the spring-summer-autumn cycle, but the calendar is treated as a phenological window rather than a fixed set of dates because altitude and annual weather conditions may shift activity periods by several weeks.

Collection methods are used in a complementary manner. For sweep-netting, the number of sweeps or transect duration is recorded; for hand collecting, person-minutes; for beating trees and shrubs, the number of plants or branches examined; for light sampling, trap-nights and active operating time; and for overwintering-site searches, search time and searched area. This approach allows later standardization of observations by effort. Photography is used as an additional method to document live coloration, behavior, and microhabitat.

Field data are harmonized with Darwin Core terms. A traditional short date format may be retained on a physical label, but eventDate is stored in ISO 8601 format in the digital database. decimalLatitude and decimalLongitude are accompanied by geodeticDatum and coordinateUncertaintyInMeters to document positional precision. occurrenceID or catalogNumber serves as the primary specimen identifier and is cross-referenced with tissueSampleID, image filenames, molecular results, and the BOLD Process ID.

Preservation and curation strategy is selected according to the final analytical purpose. Experimental evidence shows that regimes optimized for DNA preservation may increase morphological fragility in some insects, while regimes favorable for morphology may provide less reliable long-term DNA integrity (Marquina et al., 2021). Therefore, the protocol does not designate a single preservation regime as universally correct. Material intended for molecular analysis and morphological vouchers are managed separately, and specific laboratory storage conditions or chemical reagents are used only under validated institutional SOPs and qualified supervision.

At the laboratory stage, adults and nymphs are separated, sex is recorded, and major morphological characters are examined under a stereomicroscope. For Pentatomidae, the head, antennal segments, pronotum, scutellum, connexivum, rostrum, and ventral characters are photo-documented. When necessary, genital characters are studied according to validated taxonomic preparation protocols, while the voucher body and any derived preparation remain linked by the same identifier. Important vouchers are photographed in dorsal, lateral, and ventral views with a scale.

For molecular identification, mitochondrial cytochrome c oxidase subunit I (COI) is used as the reference barcode marker (Folmer et al., 1994; Hebert et al., 2003). The molecular subsample is linked to the voucher with a separate tissueSampleID, while PCR and sequencing parameters follow the laboratory's validated SOP. In BOLD reference-barcode practice, a retrievable voucher, a unique specimen identifier, a complete collection record, and a quality-controlled COI sequence are essential. Platform guidance also provides quality criteria for reference records, including sufficient sequence length and a very low proportion of ambiguous bases (BOLD Systems, 2026).



Sequences are submitted to BOLD and GenBank, while occurrence and ecological metadata are prepared for GBIF-compatible publication.

Table 1. Adapted seasonal sampling calendar for Pentatomidae monitoring in the Fergana Valley.

Period	Main phenological condition	Monitoring objective
March-April	Adults emerging from overwintering	Early-season assemblage and phenology of overwintering emergence
May-August	Feeding, mating, oviposition; nymphal stages	Habitat composition, host plants, and generation dynamics
September-October	New-generation adults and migration	End-of-season composition and preparation for overwintering
Early November	Search for overwintering sites	Identification of overwintering microhabitats

Note: the windows are not fixed dates and may shift with altitude and annual climatic conditions.

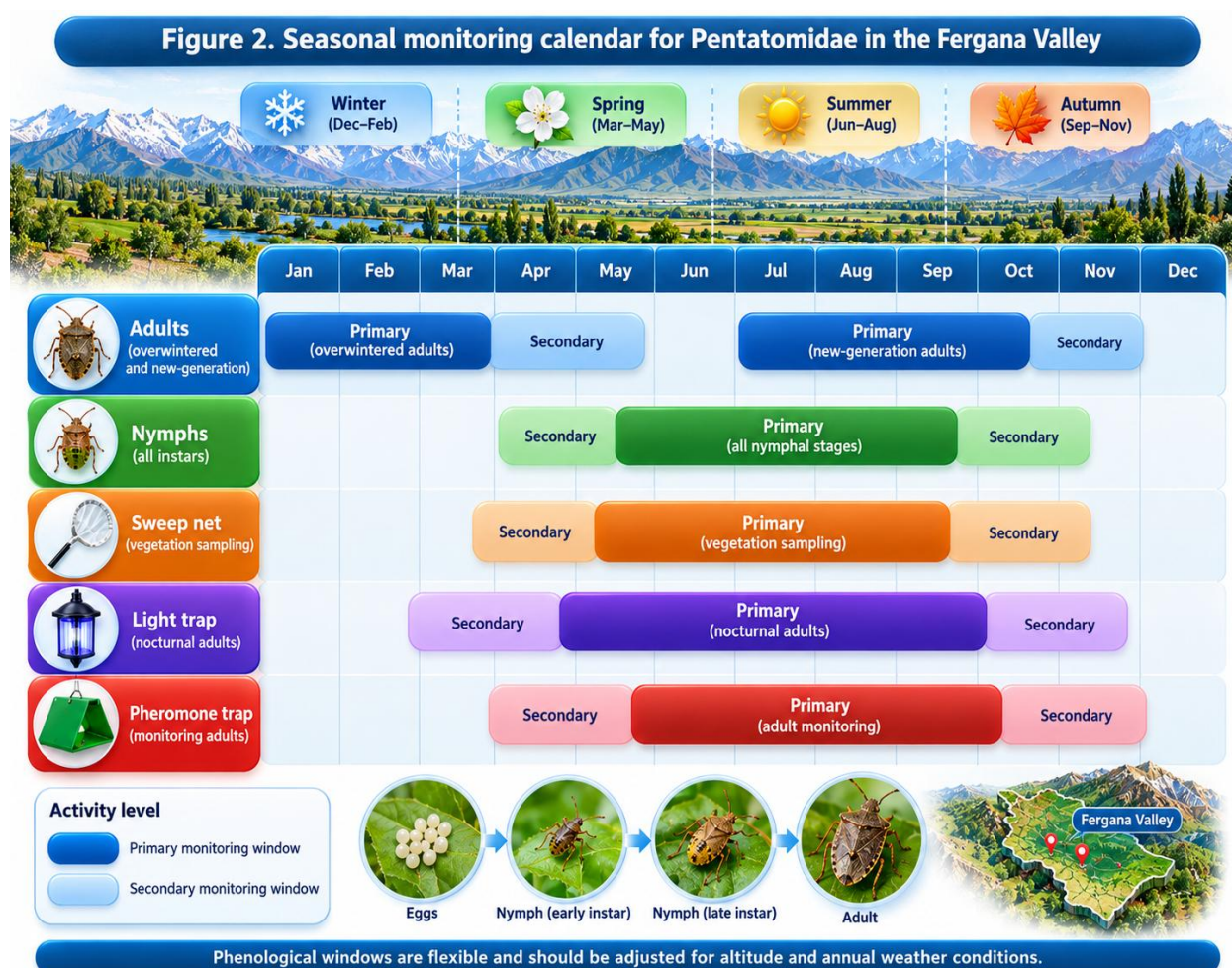


Figure 2. Visual model of recommended seasonal monitoring windows.

Source: methodological synthesis based on Table 1.



Table 2. Standardization of sampling effort when collecting Pentatomidae material.

Collection method	Main habitat / target	Standard effort unit	Primary quality-control requirement
Sweep net	Grasslands, alfalfa, cereals, and low vegetation	Number of sweeps or transect-time	Use the same operator technique and equivalent effort
Hand collecting	Large and visible individuals	Person-minutes	Record selectivity and observer effect
Beating	Trees and shrubs	Number of plants / branches examined	Use the same beating sheet or screen and equivalent effort
Light attraction	Nocturnally active groups	Trap-night and active operating time	Record device type, operating period, and weather
Overwintering-site search	Leaf litter, bark, dry vegetation, structures	Search time or searched area	Describe microhabitat precisely
Photography	Live coloration and microhabitat	Photo event	Record scale, view, and image file ID

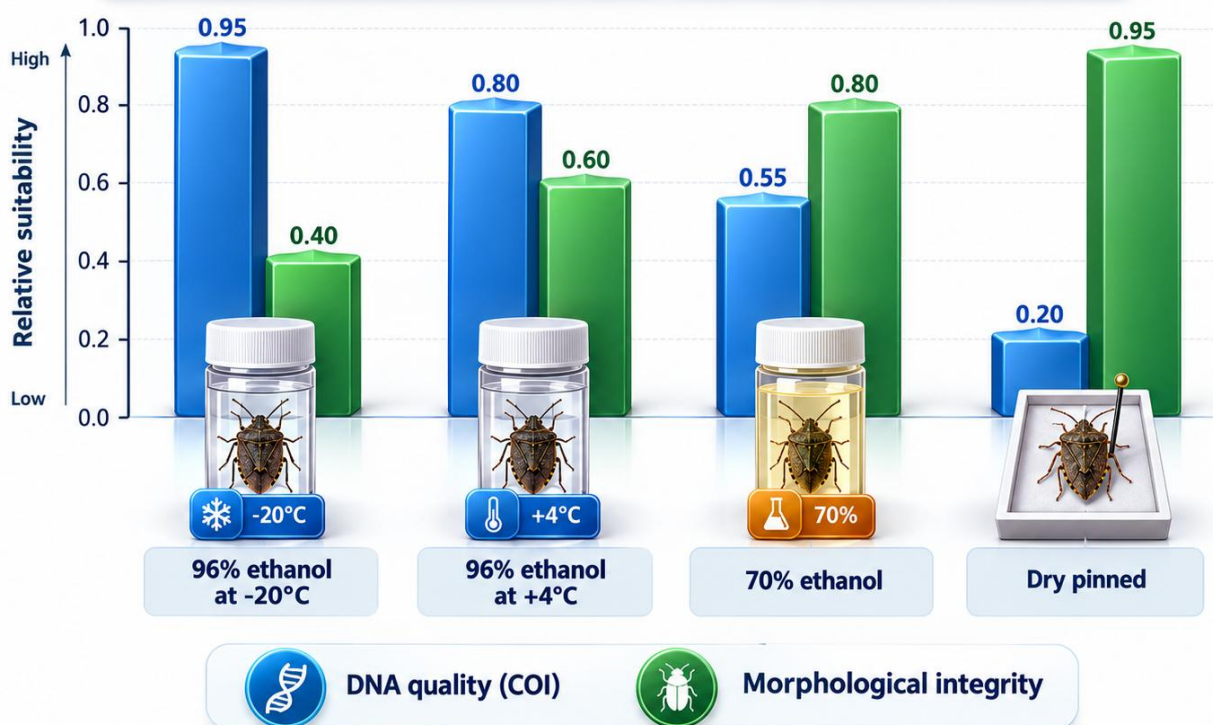
Source: improved from Martin (1977), Upton (1991), and contemporary monitoring principles.

Table 3. Minimal metadata package for aligning a field record with Darwin Core.

Data block	Recommended Darwin Core field	Comment
Unique record ID	occurrenceID / catalogNumber	Primary link for the specimen and all derivative data
Date	eventDate	Store in ISO 8601 format in the digital database
Coordinates	decimalLatitude; decimalLongitude	Decimal degrees
Coordinate precision	geodeticDatum; coordinateUncertaintyInMeters	Document both datum and coordinate uncertainty
Locality	countryCode; stateProvince; locality	Country code + administrative region + precise locality
Collector	recordedBy	Standardized personal name or persistent identifier
Method and effort	samplingProtocol; samplingEffort	Record collection method together with effort
Biological status	lifeStage; sex	Adult/nymph and sex
Host / habitat	habitat; associatedTaxa	Scientific name of host plant and microhabitat
Voucher	institutionCode; collectionCode; catalogNumber	Retrievable physical specimen

Source: based on Wiczorek et al. (2012) and TDWG Darwin Core Terms (2026).

Figure 3. Preservation trade-off between DNA quality and morphological integrity



Preservation methods involve a trade-off between DNA quality and morphological integrity. Cold storage in 96% ethanol maximizes DNA quality, while dry pinning best preserves morphology.

Figure 3. Relative suitability of preservation strategies for morphology and DNA-barcoding objectives.

Source: conceptual synthesis based on Marquina et al. (2021). The 3D bar heights and numerical labels are illustrative suitability values used for visualization and are not direct empirical measurements.

3. Results: Standardized Protocol and Quality Control

Methodological synthesis produced a six-stage workflow for managing Pentatomidae material from the Fergana Valley: (1) habitat and seasonal planning; (2) standardized collection with measurable effort; (3) a persistent identifier and Darwin Core metadata; (4) morphological voucher preparation and photography; (5) a voucher-linked COI barcode record; and (6) integration with BOLD, GenBank, and GBIF. Progression to the next stage requires completion of the relevant quality-control criterion.

The primary informational unit of the protocol is not merely the physical specimen, but a linked data package centered on a persistent occurrence/voucher identifier. This package brings together the voucher body, high-quality images, field metadata, tissue subsample, and COI sequence. The approach preserves provenance and increases the ability to re-check molecular results against a physical specimen.

The roles of BOLD and GBIF are separated within the protocol. BOLD and GenBank manage molecular sequence and barcode-related records, while GBIF serves as a major platform for

dissemination of occurrence and ecological metadata. Thus, the same biological specimen may be represented across several platforms as cross-referenced scientific objects rather than as isolated duplicate records, which supports international reuse and later taxonomic revision.

Figure 4. Unique identifier architecture linking voucher, images, metadata, tissue, and COI sequence

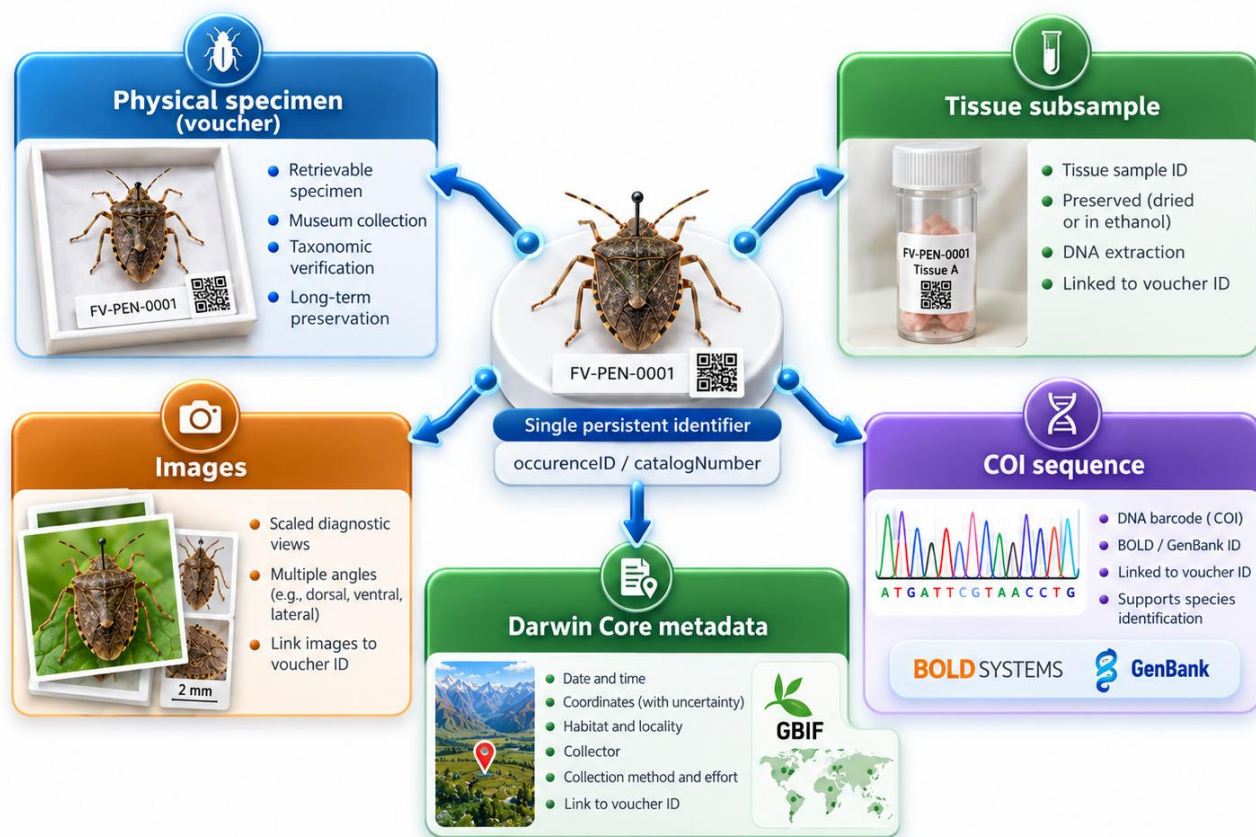


Figure 4. Voucher-metadata-DNA linkage architecture built around a persistent specimen identifier.

Source: author-developed scheme based on Darwin Core and BOLD data models.

Table 4. Quality-control checkpoints in the standardized protocol.

Stage	Quality-control question	Acceptance criterion
Planning	Do habitat and season match the research objective?	Sampling design and effort are defined in advance
Collection	Was effort recorded?	Method + effort + date + locality are present
Metadata	Can the specimen be unambiguously traced?	Persistent ID, coordinates, uncertainty, collector, and habitat are complete
Voucher	Are diagnostic characters documented?	Voucher, scaled images, and collection location are available
Molecular record	Is the sequence linked to the voucher?	tissueSampleID and sequence ID are cross-referenced
Publication	Are platform records interoperable?	BOLD/GenBank sequence IDs and GBIF occurrenceID are cross-referenced

4. Discussion

The first major improvement of the revised methodology is the explicit maintenance of taxonomic boundaries. A Pentatomidae protocol should use only taxa that belong to this family as examples; for instance, *Eurygaster integriceps* belongs to Scutelleridae rather than Pentatomidae and should not be included in a pentatomid reference list. Although this distinction may appear minor, it is fundamental for the integrity of morphological keys, voucher collections, and barcode reference libraries.

The second improvement is the rejection of overly categorical statements about preservation regimes. Marquina et al. (2021) showed that the trade-off between DNA preservation and morphological fragility varies among insect taxa. A lower-concentration preservation medium may be suboptimal for long-term molecular storage, but universal claims such as “always unsuitable for DNA analysis” or “COI amplification will inevitably fail” are not scientifically justified. Preservation strategy should therefore be considered jointly with taxon, storage duration, temperature control, and the intended downstream analysis.

The third improvement is closer alignment of field metadata with current Darwin Core practice. Recording eventDate in ISO 8601 format, documenting coordinateUncertaintyInMeters and geodeticDatum together with coordinates, and recording samplingProtocol and samplingEffort in parallel improves machine readability and supports later meta-analysis. The 2026 Darwin Core terms continue to recognize coordinateUncertaintyInMeters as a dedicated field for documenting spatial uncertainty (TDWG, 2026).

A fourth methodological point is to treat BOLD, GenBank, and GBIF as interoperable platforms with different primary functions rather than as a single generic database. BOLD manages barcode specimen and sequence information and emphasizes retrievable vouchers and unique identifiers for reference records. GenBank archives nucleotide sequences, whereas GBIF is a major platform for disseminating occurrence data. A cross-referencing model centered on occurrenceID/catalogNumber is therefore one of the strongest components of the proposed methodology.

The methodology also has limitations. The recommended seasonal windows are biologically plausible for the Fergana Valley, but they require calibration through multi-year phenological observations. COI is an effective identification marker for many taxa, yet closely related, introgressed, or otherwise problematic species may require morphology and additional markers. The protocol has not yet undergone formal empirical validation for inter-observer repeatability, DNA success rate, or metadata completeness. Future work should quantify these indicators directly.

From an applied perspective, the record of *H. halys* from the Fergana area (Gandjaeva et al., 2022) increases the relevance of standardized monitoring. Consistent sampling effort, complete metadata, and voucher-linked barcode records can improve temporal tracking of an invasive species, reduce misidentification, and allow national observations to be compared with international surveillance systems.

5. Conclusion

The proposed methodology for studying Pentatomidae in the Fergana Valley integrates field collection, sampling effort, metadata, morphological vouchers, COI barcodes, and repository

integration into a single scientific chain. In the revised protocol, taxonomic boundaries were clarified, universal and overly rigid preservation statements were replaced with a balanced interpretation based on contemporary experimental evidence, Darwin Core fields were updated, and the respective roles of BOLD, GenBank, and GBIF were distinguished. The central principle is to link every observation to a retrievable voucher and make all derivative information traceable through a persistent identifier. This approach provides a robust basis for inventorying the Fergana Valley Pentatomidae fauna, monitoring invasive and economically important species, and building a reusable biodiversity-data resource that meets international standards.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Dillon N., Austin A.D., Bartowsky E. Comparison of preservation techniques for DNA extraction from hymenopterous insects. *Insect Molecular Biology*. 1996;5(1):21–24.
2. Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*. 1994;3(5):294–299.
3. Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B*. 2003;270(1512):313–321. doi:10.1098/rspb.2002.2218.
4. Mahmudov M.U. Fauna and ecology of bugs of Western Fergana (Heteroptera: Pentatomoidea): abstract of PhD dissertation [in Uzbek]. Fergana: Fergana State University; 2025. 47 p.
5. Martin J.E.H. Collecting, Preparing, and Preserving Insects, Mites, and Spiders. *The Insects and Arachnids of Canada, Part 1*. Ottawa: Agriculture Canada; 1977.
6. McPherson J.E., ed. *Invasive Stink Bugs and Related Species (Pentatomoidea): Biology, Higher Systematics, Semiochemistry, and Management*. Boca Raton: CRC Press; 2018.
7. Nagy Z.T. A hands-on overview of tissue preservation methods for molecular genetic analyses. *Organisms Diversity & Evolution*. 2010;10(1):91–105. doi:10.1007/s13127-010-0012-4.
8. Ratnasingham S., Hebert P.D.N. BOLD: The Barcode of Life Data System. *Molecular Ecology Notes*. 2007;7(3):355–364. doi:10.1111/j.1471-8286.2007.01678.x.
9. Rider D.A., Schwertner C.F., Vilímová J., Rédei D., Kment P., Thomas D.B. Higher systematics of the Pentatomoidea. In: McPherson J.E., ed. *Invasive Stink Bugs and Related Species*. CRC Press; 2018:25–201.
10. Schuh R.T., Slater J.A. *True Bugs of the World (Hemiptera: Heteroptera): Classification and Natural History*. Ithaca: Cornell University Press; 1995.
11. Schuh R.T., Weirauch C. *True Bugs of the World (Hemiptera: Heteroptera): Classification and Natural History*. 2nd ed. Manchester: Siri Scientific Press; 2020.
12. Upton M.S. *Methods for Collecting, Preserving, and Studying Insects and Allied Forms*. 4th ed. Brisbane: Australian Entomological Society; 1991.
13. Wiczorek J., Bloom D., Guralnick R., Blum S., Döring M., Giovanni R., Robertson T., Vieglaiss D. Darwin Core: an evolving community-developed biodiversity data standard. *PLoS ONE*. 2012;7(1):e29715. doi:10.1371/journal.pone.0029715.





14. Xoliqov R.Y. Natural Geography of the Fergana Valley [in Uzbek]. Tashkent: Navro‘z; 2020. 168 p.
15. Marquina D., Buczek M., Ronquist F., Łukasik P. The effect of ethanol concentration on the morphological and molecular preservation of insects for biodiversity studies. PeerJ. 2021;9:e10799. doi:10.7717/peerj.10799.
16. Gandjaeva L.A., Hudayberdieva M.O., Abdullaev I.I., Mirzayeva G.S., Yusupboev E.K. First record of Halyomorpha halys (Heteroptera: Pentatomidae) from Uzbekistan. Zoosystematica Rossica. 2022;31(2):329–331. doi:10.31610/zsr/2022.31.2.329.
17. BOLD Systems. Barcoding Standards and Specimen Data Submission documentation. 2026. <https://boldsystems.org/resource-hub/standards/> (accessed 4 October 2026).
18. TDWG. Darwin Core Quick Reference Guide and List of Terms. 2026. <https://dwc.tdwg.org/terms/> (accessed 4 October 2026).