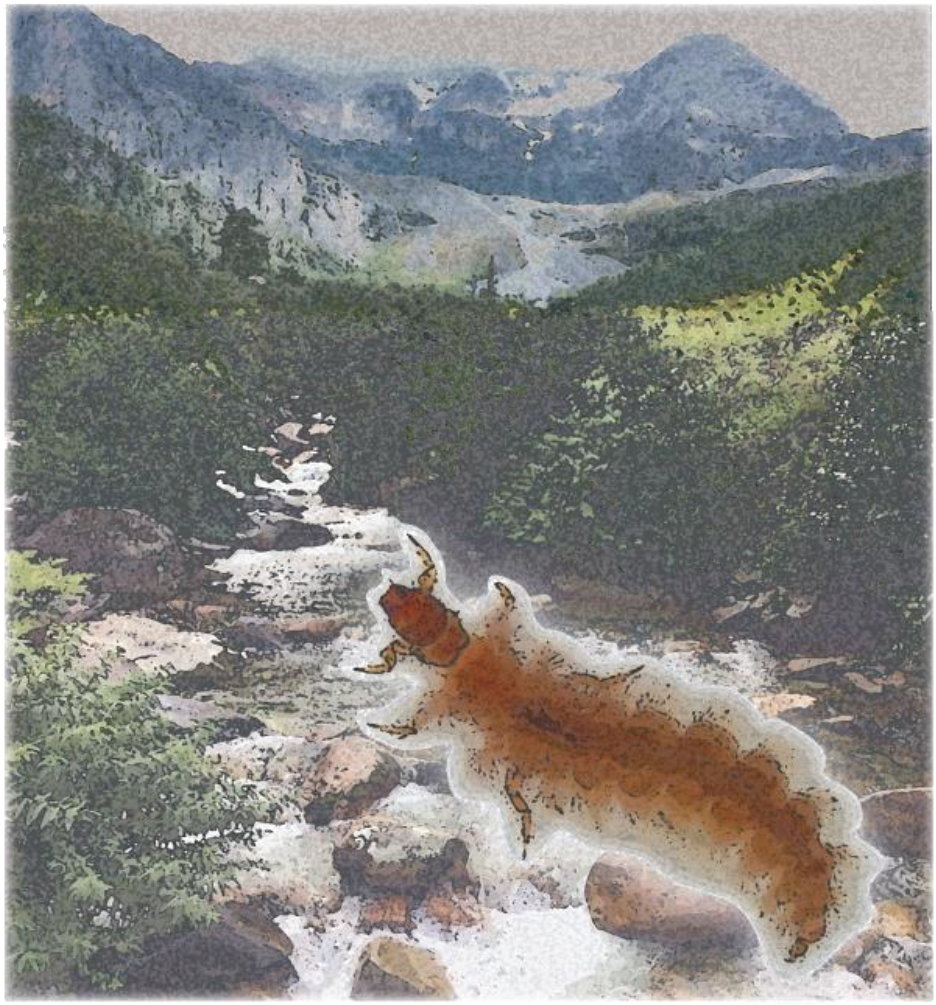
Department of Insect Biotechnology, Justus Liebig University Giessen

Supervisor: Prof. Dr. Steffen Pauls

Phylogeography of High Mountain Caddisflies (Trichoptera) in Asias Subtropical Mountains



Dissertation for the degree of Doctor rerum naturalium (Dr. rer. nat)

Faculty 09 – Agricultural Sciences, Nutritional Sciences and Environmental Management

Justus-Liebig-University Giessen

Heinrich-Buff-Ring 26-32 IFZ, 35392 Giessen

Presented by M. Sc. Xiling Deng

Frankfurt am Main, 2023

Doctoral Board

Chair

Professor Dr. Marc F. Schetelig

Department of Insect Biotechnology

Faculty 09 – Agricultural Sciences, Nutritional Sciences and Environmental Management

Justus-Liebig-University Giessen

First Reviewer

Prof. Dr. Steffen Pauls

Department of Insect Biotechnology

Faculty 09 – Agricultural Sciences, Nutritional Sciences and Environmental Management

Justus-Liebig-University Giessen

Second Reviewer

Prof. Dr. Miklós Bálint

Department of Insect Biotechnology

Faculty 09 – Agricultural Sciences, Nutritional Sciences and Environmental Management

Justus-Liebig-University Giessen

Examiner

Prof. Dr. Lutz Breuer

Department of Landscape Ecology and Resources Management

Faculty 09 – Agricultural Sciences, Nutritional Sciences and Environmental Management

Justus-Liebig-University Giessen

Examiner

apl. Prof. Dr. Christian Albrecht

Department of Animal Ecology & Systematics

Faculty 08 – Biology and Chemistry

Justus-Liebig-University Giessen

Date of disputation: 12. December 2023

Abbreviations

THR	Tibeto-Himalayan Region
HIM	Himalayas
HDM	Hengduan Mountains
QTP	Qinghai-Tibet Plateau
MGH	Mountain-Geobiodiversity Hypothesis
m a.s.l.	meters above sea level
WGS	Whole Genome Sequencing
LGM	Last Glacial Maximum
FCS	Flickering Connectivity System

Summary

As one of the most prominent mountain systems on Earth, the Tibeto-Himalayan Region (THR), is not only famous for its geographic and climatic influence at a global scale but is also well-known for harboring high levels of biodiversity and it presently comprises two global biodiversity hotspots. To explain the formation of the rich biodiversity in this area, the "Mountain-Geobiodiversity Hypothesis (MGH)" proposes that the present-day montane biodiversity patterns are derived from a combination of geology, biology, and climate change. However, validations and refinements of the hypothesis for a broader taxonomic spectrum are missing, particularly from animals.

Trichoptera, colloquially known as "caddisflies", is the largest order of primary aquatic insects. Caddisflies occur on all continents except Antarctica and have adapted to virtually all forms of freshwater ecosystems, including high-altitude streams, rivers, and lakes. Caddisflies of the genus *Himalopsyche* (Rhyacophilidae) are mainly distributed in the subtropical mountains of central and eastern Asia. Currently, there are 56 named species within this genus: Twenty-three occur in the Himalayas, 34 in the Hengduan Mountains, and four species are distributed in both mountain systems. In addition, several species also occur in far east Asia and Southeast Asia, such as Indonesia and Japan, and one species inhabits North America. Hence, the Himalayas and Hengduan Mountains are the centers of diversity for this caddisfly group. The phylogenetic relationships of the genus *Himalopsyche* were recently resolved with traditional sequencing technology and morphology, but ambiguity remains about the shallow phylogeny of several species complexes. Also, studies of intraspecific variation and population genetics are hitherto lacking.

Their high level of species diversity in subtropical mountains and the available knowledge of their phylogenetics and ecology make *Himalopsyche* a good model to study the formation of biodiversity of aquatic insects in these high mountain systems of Asia and assess the validity of the MGH. The central theme of this work is thus to derive hypotheses on the processes underlying the formation of the high biodiversity in these two mountain regions using selected caddisflies species of the genus *Himalopsyche*. Specifically, this work is focused on 1) solving the ambiguity of the shallow phylogeny of one *Himalopsyche* species complex; 2) studying intraspecific variation and population genetics of several *Himalopsyche* species. To address the open questions related to the evolutionary process, I harness the power and resolution of next-generation sequencing technologies including targeted amplicon sequencing, *de novo* sequencing, and whole genome resequencing.

To assess the value of genomic-level data in resolving phylogenetic relationships in caddisflies, we reconstructed the phylogeny and performed gene flow and network analyzes on the hitherto unresolved *H. martynovi* complex using multiple allelic datasets generated from anchored hybrid enrichment (Chapter 1). The data identified three robust lineages, which were supported by morphological evidence. The remaining morphological ambiguity within *H. martynovi* sensu stricto may have resulted from gene flow within the species complex. I hypothesized this gene flow could have been fostered by climate

oscillations and drainage re-arrangement. To verify the influence of climate and topography on species diversification, I attempted to investigate the intraspecific variation of *Himalopsyche* species by using genome-wide variants. A methodological study on how to set up a population genomic study regarding the impact of sequencing depth and relatedness of reference genomes is summarized in Chapter 2. In this technical chapter, I provided a general guideline for population genomic studies. I revealed that a high-quality reference genome closely related to the focal species would be optimal for setting up a population genomics study. In addition, the results of this study indicated that the population structure of *H. tibetana* and *H. digitata* were highly consistent with the geographic distribution of populations within the drainage and river networks. To further explore the formation of the diversity of *Himalopsyche*, especially in different mountain regions, I then conducted a comparative phylogeographic study (Chapter 3). In this study, I performed a series of population genomic inferences using individual-based genomic data on four *Himalopsyche* species that inhabit different niches in the Himalayas and the Hengduan Mountains. The population genetic patterns of the four species demonstrated that the high-elevation species showed strong local differentiation in both mountain regions. In contrast, the low-elevation species were shaped by river basins, indicating greater regional dispersal activity. In addition, caddisfly species in the Himalayas generally exhibited an East-West oriented dispersal. Species from the Hengduan Mountains showed greater connectivity on the North-South orientation, suggesting that species have a higher chance to survive in the Hengduan Mountains by both *in-situ* displacement (along the elevational gradients) and long-distance dispersal (along the latitudinal gradients) during glaciation. To better understand the processes underlying these differences in the context of historical climate and topography, I incorporated genomic and ecological evidence to reveal the demographic history and potential habitat range dating back to the last glacial maximum. These analyzes revealed a demographic expansion for all four *Himalopsyche* species linked to increased potential habitats during the LGM. Therefore, this study demonstrated that historical geodiversity and climate fluctuations interact and influence the diversification of caddisflies in the THR, thus supporting the MGH.

Zusammenfassung

Als eines der bedeutendsten Gebirgssysteme der Erde ist die Tibeto-Himalaya-Region nicht nur für ihren geografischen und klimatischen Einfluss auf globaler Ebene bekannt, sondern auch für ihre hohe biologische Vielfalt. Der Region werden daher als zwei globale Biodiversität-Hotspots zugewiesen. Zur Erklärung der Entstehung der reichen biologischen Vielfalt in diesem Gebiet wurde die "Mountain-Geobiodiversity Hypothesis (MGH)" vorgeschlagen, die davon ausgeht, dass das heutige Muster der montanen biologischen Vielfalt aus der Kombination von Geodiversitätsentwicklung und Klimawandel entstanden ist. Es fehlt jedoch eine Validierung und Verfeinerung der auf Basis von Pflanzen entwickelten Hypothese für ein breiteres taxonomisches Spektrum, insbesondere für Tiere.

Trichoptera, umgangssprachlich als "Köcherfliegen" bekannt, ist die größte Ordnung der primären Wasserinsekten. Köcherfliegen kommen auf allen Kontinenten mit Ausnahme der Antarktis vor und haben sich an praktisch alle Formen von Süßwasserökosystemen angepasst, einschließlich hoch gelegener Bäche, Flüsse und Seen. Köcherfliegen der Gattung *Himalopsyche* (Rhyacophilidae) sind hauptsächlich in den subtropischen Gebirgen Zentral- und Ostasiens verbreitet. Derzeit sind 56 *Himalopsyche*-Arten bekannt: 23 kommen im Himalaya vor, 34 im Hengduan-Gebirge, und vier Arten sind in beiden Gebirgssystemen verbreitet. Darüber hinaus kommen mehrere Arten auch in Fernost- und Südostasien vor, etwa in Indonesien und Japan, und eine Art bewohnt Nordamerika. Somit sind der Himalaya und das Hengduan-Gebirge die Zentren der Vielfalt dieser Köcherfliegengruppe. Die phylogenetischen Beziehungen der Gattung *Himalopsyche* wurden vor kurzem mithilfe der traditionellen Sequenzierungstechnologie und der Morphologie geklärt, aber es besteht weiterhin Unklarheit über die Phylogenie mehrerer Artenkomplexe. Auch fehlen bislang Studien zur innerartlichen Variation und zur Populationsgenetik.

Dank ihrer hohen Artenvielfalt in subtropischen Gebirgen und unserem gutem Hintergrundwissen über Phylogenetik und Ökologie ist *Himalopsyche* ein gutes Modell, um die Entstehung der Biodiversität aquatischer Insekten in diesen Hochgebirgssystemen Asiens zu untersuchen, insbesondere im Rahmen der MGH. Zentrales Thema dieser Arbeit ist es daher, anhand ausgewählter Köcherfliegenarten der Gattung *Himalopsyche* Hypothesen über die Prozesse abzuleiten, die der Entstehung der hohen Biodiversität in diesen beiden Gebirgsregionen zugrunde liegen. Um dies zu erreichen, konzentriert sich diese Arbeit erstens auf die Klärung der unklaren Phylogenie eines *Himalopsyche*-Artenkomplexes und zweitens auf die Untersuchung der intraspezifischen Variation und Populationsgenetik mehrerer *Himalopsyche*-Arten. Um die offenen Fragen im Zusammenhang mit dem Evolutionsprozess zu klären, wurden die Leistungsfähigkeit und die Auflösung neuester DNA Sequenzierungstechnologien genutzt, z. B. gezielte Amplikon Sequenzierung, *de novo* Sequenzierung und Sequenzierung des gesamten Genoms.

Um den Wert von Daten auf Genomebene für die Klärung der phylogenetischen Beziehungen bei Köcherfliegen zu bewerten, wurde u.a. für den bisher unaufgelösten *H. martynovi*-Artenkomplex die

Phylogenie rekonstruiert und Genfluss- und Netzwerkanalysen mittels *Anchored Hybrid Enrichment* Analyse durchgeführt (Kapitel 1). Die Daten identifizierten drei Linien, die durch morphologische Beweise gestützt wurden. Die verbleibende morphologische Unklarheit innerhalb *H. martynovy* s.s. ist möglicherweise auf Genfluss zurückzuführen, der im Zusammenhang mit Klimaschwankungen und der Umgestaltung von Einzugsgebieten stehen könnte. Um den Einfluss von Klima und Topografie auf die Diversifizierung der Arten zu überprüfen, wurde versucht, die intraspezifische Variation der *Himalopsyche*-Arten anhand genomweiter Varianten zu untersuchen. Eine erste methodische Studie widmete sich der Bedeutung von Sequenzierungstiefe und der Verwandtschaft der Referenzgenome (Kapitel 2). In diesem technischen Kapitel wurde ein allgemeiner Leitfaden für populationsgenomische Studien erstellt. Es wurde gezeigt, dass ein hochwertiges Referenzgenom einer Art, die eng mit der Zielart verwandt ist, optimal für die Durchführung einer Populationsgenomikstudie ist. Darüber hinaus zeigten die Ergebnisse dieser Untersuchung, dass die Populationsstruktur von *H. tibetana* und *H. digitata* in hohem Maße mit der geografischen Verteilung der Populationen innerhalb der Einzugsgebiete und Flussnetzwerke übereinstimmt. Um die Entstehung der Vielfalt von *Himalopsyche*, insbesondere in verschiedenen Bergregionen, weiter zu erforschen, wurde eine vergleichende phylogeografische Untersuchung durchgeführt (Kapitel 3). In dieser Studie wurde eine Reihe von populationsgenomischen Rückschlüssen anhand individuenbasierter genomischer Daten von vier *Himalopsyche*-Arten gezogen, die unterschiedliche Nischen im Himalaya und im Hengduan-Gebirge bewohnen. Die populationsgenetischen Muster der vier Arten zeigten, dass die Arten der höheren Lagen in beiden Bergregionen eine starke lokale Differenzierung aufweisen. Im Gegensatz dazu waren die Arten der niederen Lagen durch den Einfluss von Einzugsgebieten geprägt, was auf eine größere regionale Ausbreitungsaktivität hindeutet. Darüber hinaus wiesen die Köcherfliegenarten im Himalaya im Allgemeinen eine Ost-West-orientierte Ausbreitung auf. Arten aus dem Hengduan-Gebirge wiesen eine größere Konnektivität in Nord-Süd-Richtung auf, was darauf hindeutet, dass die Arten eine größere Chance haben, im Hengduan-Gebirge zu überleben, und zwar sowohl durch *in-situ*-Verschiebung (entlang der Höhengradienten) als auch durch Ausbreitung über große Entfernungen (entlang der Breitengradienten) während der Vergletscherung. Um die Prozesse, die diesen Unterschieden zugrunde liegen, besser zu verstehen, insbesondere vor dem Hintergrund des historischen Klimas und der Topographie, wurden genomische und ökologische Daten einbezogen, um die demografische Geschichte und den potenziellen Lebensraum Bereich seit dem letzten glazialen Maximum zu ermitteln. Die Ergebnisse dieser Analysen weisen auf eine demographische Expansion für alle vier *Himalopsyche*-Arten in Verbindung mit einer Vergrößerung der potenziellen Lebensräume während des LGM hin. Diese Studie hat gezeigt, dass historische Geodiversität und Klimaschwankungen zusammenwirken und die Diversifizierung der Köcherfliegen im THR beeinflussen und somit die MGH unterstützen.

Table of Contents

1. Introduction and objectives

1.1 Two biodiversity hotspots: the Himalayas and the adjacent Hengduan Mountains

The Tibeto-Himalayan Region (THR), which is located at the intersection of central, east, and south Asia, comprises the world's largest and highest plateau, the Qinghai-Tibet Plateau (QTP). Its mean elevation exceeds 4500 m, its surface area is over 2.3 million km^2 (Mosbrugger et al., 2018) and it is home to the highest mountain ranges in the world. Except for the Arctic and Antarctic polar caps, the THR also has the largest concentration of alpine glaciers and extensive high-altitude permafrost and snow cover. It is thus colloquially known as the "Third Pole" (Kang et al., 2019). Several mountain ranges are located in the surrounding area of the THR, including the Himalayas (HIM), the Hengduan Mountains (HDM), Kunlun Shan, Qilian Shan, and Tianshan. For this thesis, I took a closer look at the two most diverse mountain ranges in geology and biology – the HIM and the HDM.

As one of the largest mountain ranges in the world, the HIM is located on the southern fringe of the THR and expands from West to East across five countries: Bhutan, China, India, Nepal, and Pakistan. Nine of the world's ten highest peaks lie in the Himalayan range. The ranges altitude extends from 500 m a.s.l. to 8848 m a.s.l. (Mount Everest/Qomolangma, Sabin et al., 2020). The HDM is a group of mountains in Southwest China that sit on the south-eastern fringes of the QTP and also display drastic altitudinal variations ranging from 1000 m a.s.l. (in deep valley floors) to 7556 m a.s.l. (Mount Gongga). The HDM are generally North-South oriented mountain ranges divided by multiple parallel north-south running valleys, thus forming three of Southeast Asia's great river systems – the Yangtze, Mekong, and Salween.

All these topographic features in the THR at present-day originated from the initial continental collision between India and Asia during the Cenozoic (65–60 Ma, Ding et al., 2022). However, the uplift of the HIM and the HDM is asynchronous. Current geological evidence suggests that the HIM started to rise from ~45–40 million years ago and eventually rose to its modern elevation ~25–15 million years ago (Ding et al., 2022; Gébelin et al., 2013). In comparison, the uplift of the HDM is considered to be more recent. Previous studies revealed that the HDM experienced a major uplift after the Miocene and had rapidly reached its peak elevation shortly before the Late Pliocene (Sun et al., 2011). However, recent fossil and tectonic evidence showed that some parts of the HDM may have reached modern elevations prior to the Oligocene (Su et al., 2019). Regardless of the discordance regarding time, the different uplift histories have resulted in unique geographic features in the two mountain systems.

Furthermore, the uplift of the QTP impacted the Asian monsoon system and middle-latitude atmospheric circulation in the Northern Hemisphere (Bookhagen and Burbank, 2010; Owen and Dortch, 2014; Yang et al., 2019). Previous studies have revealed that the evolution of the Asia monsoon (including the Indian summer monsoon and the East Asian monsoon) is closely related to the topographic development of the QTP and the HIM (e.g., Tada et al., 2016). Specifically, the QTP, in terms of height and extent, acts as

a source of elevated thermal and mechanical forcing for the monsoon circulation and moisture convergence patterns (Li et al., 2016). Moreover, each regional monsoon system is intensified by different parts of the mountains uplifted in varied time periods. For example, the Indian summer monsoon intensified with the uplift of the southern and central QTP (ca. 40–35 Ma), while the East Asian monsoon intensified during the uplift of the northern QTP (ca. 25–20 Ma) and subsequent uplift of the northeastern and eastern QTP (ca. 15–10 Ma, Tada et al., 2016). Overall, the complex orogeny resulted in the current topographic and climatic features in the HIM and the HDM.

Beyond the outstanding topography, species diversity is also extraordinarily high in the two mountain regions. As outlined by Myers et al. (2000) around two decades ago, the HIM (named as Indo-Burma together with parts of Southeast Asia) and the HDM (referred to as South-Central China) harbor a large number of plant and vertebrate species. In particular, there is a high percentage of endemic species: 2.3% of the world's endemic plants and 1.9% of endemic vertebrates inhabit the HIM; 1.2% of endemic plants and 0.7% of endemic vertebrates occur in the HDM. Therefore, as part of the Indo-Burma hotspot, Myers et al. (2000) classified these two mountain systems as global biodiversity hotspots based on their respective levels of endemism, species richness, and threat to their survival. After a conservation oriented re-evaluation, Marchese (2015) redefined and expanded the HIM and the HDM as two independent biodiversity hotspots owing to their unique biodiversity patterns (Fig. 1).

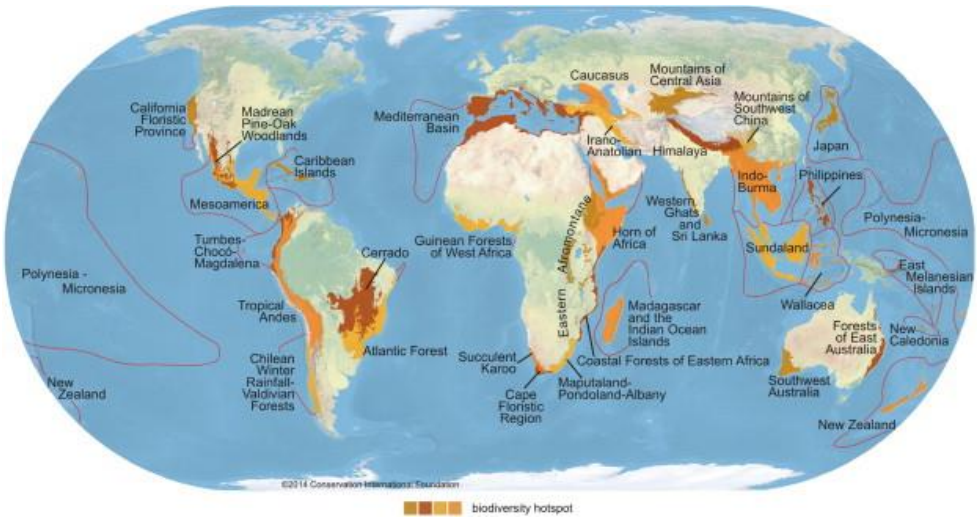


Fig. 1. The global biodiversity hotspot. Colors in orange are used to differentiate different biodiversity hotspots. Reprinted from (Marchese, 2015).

Each region has unique biodiversity features. Whereas species richness increases from West to East in the HIM (Dahal et al., 2021), for instance, in birds (Price et al., 2011), plants (Ashokan et al., 2022; Bhattarai et al., 2014; Rana et al., 2019; Yan et al., 2013) and mammals (Srinivasan et al., 2014), a North-South floristic divide is found in the HDM (Li et al., 2021). Moreover, patterns of diversity in the HDM appear to have been more dynamic through time, for example, with hotspots of a variety of montane plants having shifted from the southeastern to the central and western parts of the HDM between the last glacial maximum (LGM) to today (Liang et al., 2018). The phylogeographic complexity

of the HDM is further complicated by the presence of geographically extensive and long-lasting barriers to dispersal, such as the deeply incised valleys of the Irrawaddy, the Salween, the Mekong, and the Yangtze rivers (Muellner-Riehl, 2019). These dispersal barriers have been shown to be instrumental in delineating floristic motifs in the region (Li et al., 2021; Muellner-Riehl and Favre, 2021).

With the increasing understanding of the biodiversity pattern in mountain systems, researchers now aim to understand the underlying mechanisms of high biodiversity in mountain areas, particularly in the THR. To answer the question of "how montane biodiversity hotspots were formed", biogeographers and ecologists have increasingly scrutinized both the evolutionary processes underlying the formation of the biota and the local abiotic dynamics within mountain regions (e.g., Ding et al., 2020).

1.2 Main concepts for explaining the high levels of mountain biodiversity

Since Alexander von Humboldt initiated the principle of *Cosmos* ("unity of nature") that combines geology and biology to explain the distribution pattern of life (Von Humboldt, 1860), researchers tried to continue following this principle in biogeography, macroecology, and evolutionary biology to explain the extraordinarily high biodiversity of mountain regions (Antonelli et al., 2018; Favre et al., 2015; Rahbek et al., 2019a, 2019b). After two centuries of study, it is generally accepted that the interaction of long-term climate change and topographically dynamic landscapes in the mountain regions play key roles in generating and maintaining diversity (Rahbek et al., 2019a). Many general concepts and processes have been developed to explain the interplay among geography, climate, and biodiversity, including "sky island" biogeography (Heald, 1967, Warshall, 1995), the "flickering connectivity system (FCS, Flantua et al., 2018)", and the Mountain Geo-biodiversity Hypothesis (MGH, Fig. 2; Mosbrugger et al., 2018).

The term "sky island" was introduced to the scientific community by naturalist Weldon Heald in 1951 (Fig. 2). Subsequently, he described the changes in the landscape in the mountains of southeastern Arizona, which started at low elevation in the desert, and transitioned to grasslands, oak-pine woodland, pine forest, and finally to spruce-fir-aspen forest at high elevation (Heald, 1967). The term "sky island" was subsequently used to refer to these isolated, high-elevation, continental habitats that are separated by ecologically differing valley systems and to allow applying theories of island biogeography in studying them (Warshall, 1995). Due to different habitats along the elevational gradients, species can migrate vertically annually or during climatic events in a "sky island". This process has been observed in numerous plants (DeChaine and Martin, 2005; Zhang et al., 2019), birds (Cox et al., 2014; McCormack et al., 2008), mammals (Atwood et al., 2011; He et al., 2019), amphibian (Shepard and Burbrink, 2008), spiders (Hedin et al., 2015; Masta, 2000), and insects (Knowles, 2000; Pauls et al., 2006). And similar to the islands in the ocean, which are isolated by water, the "sky islands" are isolated by the "sea" of unsuitable habitats in the surrounding valleys (Warshall, 1995). These valleys may act as a barrier for species to disperse and consequently result in genetic drift between islands (Knowles, 2000; Lacey Knowles and Alvarado-Serrano, 2010). In general, both the various habitats and the

geographic barriers of "sky islands" can greatly influence the species. For instance, elevational gradient in mountains facilitates species expansion, migration, and colonization to new "sky islands", especially when facing an abiotic dynamic, thus forming the diversification patterns of "sky islands" (He and Jiang, 2014).

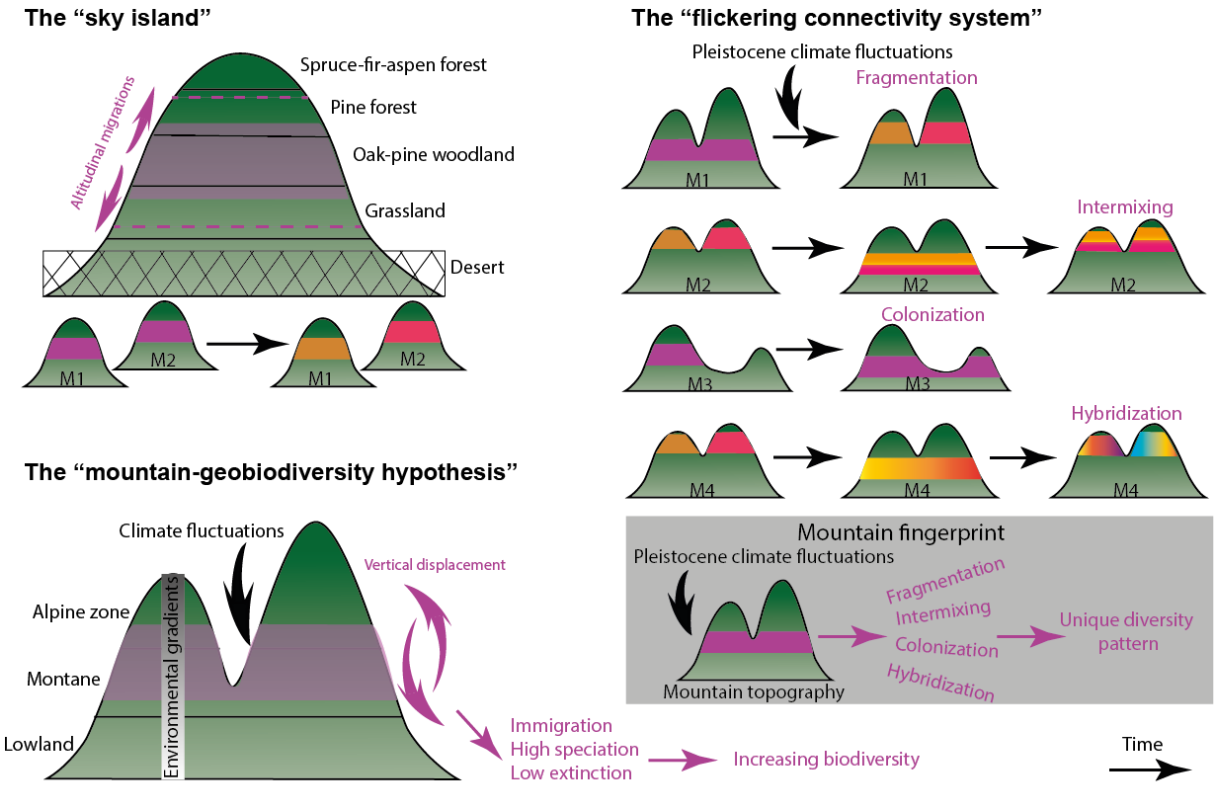


Fig. 2. Summary of the three main concepts for explaining the high levels of mountain biodiversity (Flantua et al., 2018; Mosbrugger et al., 2018; Warshall, 1995).

To better understand how abiotic dynamics shaped the distribution patterns of species in the mountains, especially in these so-called "sky islands", Flantua et al. (2018) introduced the concept of "flickering connectivity system (FCS)" (Fig. 2). By reconstructing the paleo-topographic distribution of alpine biome in the northern Andes, Flantua et al. demonstrated that the connectivity dynamics of species resulting from repetitive climatic changes during the Pleistocene forced the rapid evolutionary process and high species richness in mountain ecosystems. According to the FCS, species may react to the climatic fluctuations of the Pleistocene by shifting their geographic range in mountain areas, thus leading to habitat connectivity or isolation (depending on the climate conditions). In consequence, this dynamic connectivity may result in four processes, including fragmentation, colonization (dispersal), intermixing, and hybridization, thus promoting diversification in mountain ecosystems (Flantua et al., 2018; Muellner-Riehl, 2019). Moreover, since topography and climate conditions are distinct in different mountain systems, the diversification of species shaped by these four processes under the concept of the FCS is unique in each mountain. Thus, this procedure caused by the interaction between topography and climate is called the "fingerprint" of a mountain (Flantua et al., 2018).

Unlike the FCS that emphasizes the effects of Pleistocene climatic fluctuations, Mosbrugger et al. (2018) developed the "mountain-geobiodiversity hypothesis (MGH)", which attempts to explain the origination and evolution of mountain biodiversity throughout the whole uplift history (Fig. 2). In this hypothesis, three boundary conditions, essential to the accumulation of biodiversity in a mountain system, are postulated: (i) a full elevational zonation with the presence of lowland, montane, and alpine zones; (ii) the occurrence of a species-pump driven by climatic fluctuation; and (iii) strong environmental gradients. Within this conceptual framework, the uplift of mountains provides elevational gradients and locally diverse topography. This condition increases opportunities for local or regional taxa to adapt to a high variety of niches (i) and fosters a higher resistance to climate change via vertical displacement (iii). Meanwhile, during climate fluctuations, diversification is fostered by a species-pump effect (ii) via cyclical range fragmentation (causing divergence) and secondary contacts (involving hybridization or reinforcement, Mosbrugger et al., 2018; Muellner-Riehl, 2019).

As of today, many phylogeographic studies on various taxa have provided evidence to support these concepts or hypotheses in different mountain regions worldwide, for instance, the mountains in Asia (e.g., the HDM: Fu et al., 2022, 2020; Mu et al., 2022; Wang et al., 2022; the HIM: Ashokan et al., 2022; Rana et al., 2019; or both mountain regions: Ding et al., 2020; Rana et al., 2021; Xu et al., 2021), Europe (Carpathians, Dénes et al., 2016; Dinaric Mountains, Bartonova et al. 2018, Kutnjak et al., 2014, Previsic et al. 2014; Mediterranean Mountains, Gentili et al., 2015), Africa (the Eastern Afromontane, Lawson, 2013; the Cape Floristic Region, Tolley et al., 2006), North America (as summarized by Jaramillo-Correa et al., 2009; Shafer et al., 2010), Central America (Mexican highlands, Mastretta-Yanes et al., 2015; Talamanca mountain range, Fuchs et al., 2023), and South America (Andes, as reviewed by Flantua et al., 2018). However, most of the studies are focused on plant taxa, such that the validation and refining of the hypothesis for a broader taxonomic spectrum are limited, particularly for aquatic insects.

1.3 The genus Himalopsyche – a caddisfly group exhibiting their diversity center in the HIM and the HDM

Caddisfly (Trichoptera) is an insect order closely related to Lepidoptera (butterflies and moths). However, among the 30 orders of insects, Trichoptera is the largest order out of several that have fully adapted to aquatic environments at their immature stages, while adults are terrestrial (Wiggins, 2004). Except for a few species that inhabit terrestrial (e.g., genus *Enoicyla* and *Philocasca demita*, Anderson, 1967) or marine environments (family Chathamiidae distributed in salt water along the coasts of the south-western Pacific Ocean, Riek, 1977), the larvae of most caddisflies inhabit different kinds of freshwater habitats, including streams, rivers, springs, ponds, and lakes. In these freshwater ecosystems caddisfly larvae have an essential role. They are an important part of the trophic dynamics and energy flow, especially when considering their diverse feeding strategies (Holzenthal et al. 2007; Morse et al. 2019). Currently, these freshwater habitats and their inhabitants are largely threatened by global climate

change and pollution caused by human activities (Holzenthal et al. 2007). To monitor these effects, Trichoptera are considered good bioindicators since they have low tolerance to high sediment and nutrient concentrations (Jehamalar et al. 2010).

Himalopsyche is a genus of caddisflies that is generally distributed in mountainous areas. The larvae of this genus inhabit typically turbulent, fast-flowing rivers and streams, and even waterfalls (Hjalmarsson et al., 2019, 2018; Malicky, 2011, 2008; Schmid and Botosaneanu, 1966; Tsuruishi, 2006). They live underwater as free-living predators. According to our observations in the field, they generally have an extremely high requirement for water quality. Thus, *Himalopsyche* potentially can be a well-suited indicator organism for monitoring biological water quality (Hjalmarsson, 2020).

Ross (1963) laid the foundation for the classification of the genus *Himalopsyche*, but it was only recently that Hjalmarsson et al. (2018) conducted a comprehensive review of this group using morphological and molecular evidence. In this study, *Himalopsyche* species were segregated into five groups: the *tibetana* group, the *lepcha* group, the *kuldschensis* group, the *phryganea* group, and the *japonica* group. The most useful characteristic to distinguish these groups is the distinct specialization of their gills (Fig. 3). Additionally, differences in the setal configuration of the pronotum and anal sclerites can also be used. The gill shape of *Himalopsyche* larvae can help categorize them into different groups. However, the variations in physical characteristics are not distinct enough to differentiate between other species within the same group.

Based on both morphological and molecular evidence, 56 *Himalopsyche* species have been named so far (Hjalmarsson et al., 2019). Most of these species are distributed in the Central and East Asian mountains, including the HIM, Tian Shan, HDM, the Khasi Hills, Japan, and several other areas in East Asia (Hjalmarsson et al., 2019). The only species that inhabits the Western Nearctic is *H. phryganea* (Ross, 1941). The highest species diversity of this caddisfly group is known from the HDM (34 species), followed by the HIM (23 species). Species of *Himalopsyche* are distributed over a broad range of elevations (400 to 4700 m a.s.l.). Yet individual species often exhibit strongly differentiated elevational distributions (Hjalmarsson, 2020; Schmid and Botosaneanu, 1966), indicating a successful adaptation to various elevational habitats. In line with the high level of biodiversity and geodiversity in the HDM and HIM, as introduced in section 1.1, the diversity and distribution of *Himalopsyche* show that the genus is potentially a good model to study adaptation and diversification of aquatic insects in mountain areas and scrutinizing the MGH. Unfortunately, the evolutionary history of this genus, especially considering a time-calibrated phylogeny and biogeographic analysis is still not available due to lack of fossil data and comprehensive species sampling (Hjalmarsson et al., 2019). However, to explore the formation of high diversity of this taxonomic group in these two mountain ranges, it is feasible to investigate the evolutionary process of *Himalopsyche* among several species or within a species.

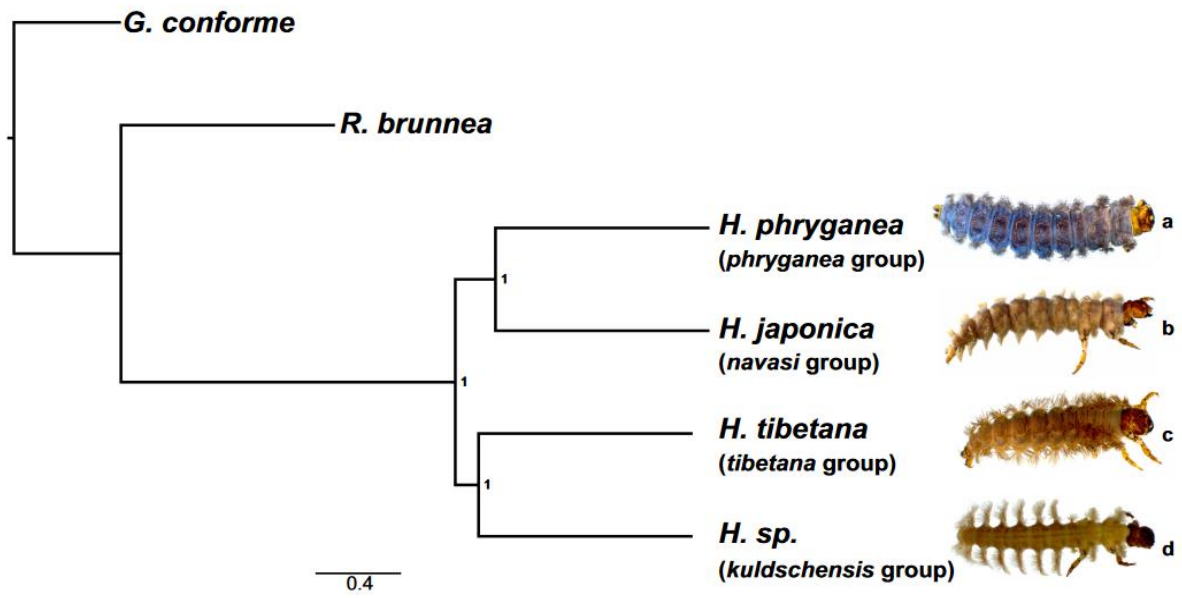


Fig. 3. Phylogenetic relationship of the four main groups of *Himalopsyche* generated from *de novo* genomes. Reprinted from Deng et al. (2022).

1.4 The application of next generation sequencing in population genetics

DNA sequencing, the process of determining the order of nucleotides in DNA, was a major breakthrough in biology research. Since Frederick Sanger sequenced the first insulin protein by deciphering and overlapping the fragments of the protein chains in 1950, sequencing technology has updated rapidly within a few decades (Shendure et al., 2017). In the first 40 years of DNA sequencing, Sanger sequencing, also known as first-generation sequencing, was the most widely used approach for generating nucleic acid sequences. Although Sanger sequencing has been regarded as the gold standard in life science for several decades (Grada and Weinbrecht, 2013; Janitz, 2008), it is superseded by next-generation sequencing (NGS, or massively parallel sequencing) technologies, especially for large-scale genomic research (Behjati and Tarpey, 2013; Grada and Weinbrecht, 2013). Compared to first-generation sequencing, NGS is able to sequence multiple fragments at once, thus making it possible to sequence a whole genome in a short time (Schadt et al., 2010). For instance, to generate an entire human genome, it took decades to deliver the final draft using Sanger sequencing, but it could be processed within one day using NGS (Behjati and Tarpey, 2013). In recent years, third-generation sequencing (also called real-time, single-molecule sequencing) has brought the technologies closer to perfection. It generates long reads, which approach 100 kb, and high accuracy and contiguity (Van Dijk et al., 2018). More importantly, DNA sequencing technology is further developing. Thus, a more efficient approach to improving time, cost, and throughput is possible in the near future.

Along with the development of DNA sequencing technology, the applications of DNA sequencing also expanded to a broader usage in the past decades. Over the years, huge progress in the field of genetics and molecular biology was achieved by applying first-generation sequencing, and NGS brought the

research field of genetics to a higher level. Currently, advanced technologies make it possible to sequence the genome of every single individual on earth, and every single cell in any tissue at every development stage (Green et al., 2017; Shendure and Aiden, 2012). This benefits research in many aspects, for instance, evolutionary biology, taxonomy, ecology, conservation, agriculture, forestry, and human health (Green et al., 2017).

Progress in evolutionary biology, including population genetics, phylogeography, and phylogenetics, was only possible with the DNA sequencing technology developed in the 1970s. In the past, researchers were aware of the limitations of the number of genes generated by Sanger sequencing, as well as the labor and cost to collect these locus-by-locus bases for the research projects (Carstens et al., 2012). However, at present, the development of NGS has solved many of these problems. In a single run of an NGS platform, data can be collected from thousands or millions of reads from multiple samples (Carstens et al., 2012). Moreover, due to the decline of cost and time, it is practicable to satisfy the need for sequencing large numbers of samples per species at a genome-wide scale, which allows analyzing high numbers of alleles across multiple loci across an individual's genome (Cross et al., 2016; Garrick et al., 2015).

In recent years, several sequencing approaches emerged for answering various biological questions in population genomic studies, including restriction-site associated DNA sequencing (RADseq, also processed to related methods, i.e., double digest RADseq and hybridization RAD), targeted sequencing (i.e., anchored hybrid enrichment (AHE)), and whole genome sequencing (WGS, Hendricks et al., 2018). These approaches generate massive datasets that can greatly improve the efficiency and accuracy in inferring the population or species history, for instance, in understanding inbreeding and outbreeding depression, the genomic basis of local adaptation and speciation, adaptive gene flow, and population demographic history (Luikart et al., 2018). Nevertheless, challenges for studies of evolutionary biology still exist, for instance, the increasingly demanding for bioinformatic tools, selection of target loci for sequencing or sequencing depth to answer a specific scientific question, and the most concerning question about reference genome for studies that work on non-model species (Deng et al., 2022; Lexer et al., 2013).

1.5 Aims of Thesis

Following the MGH, my main hypothesis is that the high species diversity of the genus *Himalopsyche* in the HIM and HDM relates to the local topography and historical climate change. Furthermore, I hypothesize that the diversity pattern of *Himalopsyche* may be consistent with drainage basins or elevation. This means species may migrate on a large scale within a certain range of drainage basins during climate change or only migrate along elevational gradients *in situ*. Thus, to explore the diversification patterns and underlying evolutionary processes of *Himalopsyche* in the HIM and HDM, three testable hypotheses will be addressed throughout the chapters of this thesis.

Hypothesis 1: The current diversity patterns of the genus *Himalopsyche* are related to the complex topographic features in the two mountain regions (Chapters 1, 2, and 3). Species or populations distributed in different environments may have different genetic diversity patterns. Mountain systems provide various environments, for instance, lowlands, mountain and alpine zone, different river networks, and varied microhabitats related to the elevational gradients. These heterogeneous environments coupled with climate dynamics may lead to differentiation and ultimately speciation of the caddisfly lineages. Since caddisflies are highly sensitive to their living environment, especially for the genus *Himalopsyche*, the diversity pattern of this group may be shaped by these topographic features, such as elevation and drainage patterns.

Hypothesis 2: The genetic diversity of *Himalopsyche* in the two mountain regions are shaped by the history of topography, climate change, and drainage re-arrangement (Chapter 1 and 3). To explain the formation of the genetic diversity pattern of *Himalopyche* species distributed in the two mountains, it is important to interpret the data in the context of a species' history. This area experienced one of the most enormous orogeny in recent times, resulting in a complex history in topography and climate. The diversification of the caddisfly group at the population level may not be caused directly by the orogeny but is more likely associated with the climate fluctuations during the glacial period, for instance, the last ice age that happened ca. 115,000 – ca. 11,700 years ago. Moreover, the drainage connections in the HDM were re-arranged several times. Therefore, I think that the extensive ice period, drainage re-arrangement, and the complex topography that resulted from the orogeny may be the main driver for the diversification of this caddisfly group.

Hypothesis 3: The MGH may sufficiently explain the diversification process of our target caddisfly group, but the implication may be different in the HIM and the HDM (Chapter 3). The topography of the HIM and the HDM is different. The HIM forms a narrow, east-west oriented mountain belt, while the HDM has a broad North-South extent and multiple parallel deep valleys. The local topography thus leads to distinct climate conditions in each mountain region, for instance, the distribution of precipitation, temperature, or vegetation. In addition, a broader extent of a mountain range is crucial to species, in particular on a north-south orientation, since it provides more potential habitat for species to disperse when facing climate change, not only along the elevational gradients but to a warmer place in the south (or to a cooler area in the north during a warm period). Thus, the biodiversity patterns may be distinct in these two mountain regions. As shown for some other plants and animals, in the HIM, the biodiversity increases from west to east; in the HDM, the highest level of biodiversity is in the central south. Thus, we assume that the influence of topography and climate change may result in different consequences for the species that inhabit the HIM and the HDM.

To better test these hypotheses, I will use various methods, including target capture by AHE, whole genome (re-)sequencing (WGS), and whole genome *de novo* assembly, and apply them at the species and population level. Generally, it is important to select a proper sequencing approach that efficiently

generates sufficient amounts of data at the right level of resolution for a study. In this context, the "best-fit" largely depends on the genetic distance of the focal taxonomic group, the availability of reference genome, the budget of time and cost, and the specific research question. Therefore, in this study, I will also try to address the contribution of the different datasets generated by different sequencing approaches and what specific questions they can address (Chapters 1, 2, and 3), as well as how to select a proper reference genome and sequencing depth in a population genomic study (Chapter 2).

2. Thesis overview

Chapter 1. This chapter investigated the mechanisms fostering the divergence and diversification in a species complex of a stream insect endemic to the HDM. The morphological features of adult male genitalia are commonly used as the key feature to identify species for all caddisflies, and genital variation within the species of *Himalopsyche* is generally limited. However, a high variation of genitalia was observed in the species of *H. martynovi*. To investigate the formation of these morphological variations and the evolutionary history of this group, 691 loci generated by anchored hybrid enrichment were used to reconstruct phylogenetic trees and performed network and gene flow analyzes. According to the results, there are three clades in the *H. martynovi* complex, and they presented gradient morphologies among the individuals. Multiple gene flow signals were detected among several lineages. The process of speciation in this species complex was likely aided by gene flow between a species of *H. martynovi* complex and another closely related species. Finally, a biogeographic scenario of this diversification process was presented. Overall, this study suggests that biological novelty, in this case, a novel trait variation, may have been acquired via hybridization. I hypothesized this gene flow could have been fostered by climate oscillations and drainage re-arrangement.

Chapter 2. Whole-genome sequencing is increasingly used in population genetic studies. However, obtaining high quality *de novo* genomes for massive numbers of samples is still not within the budgets of many researchers. Shallow resequencing approaches can aid in providing accurate genome-wide information based on more affordable shallow coverage short read sequencing data. In this context, it is imperative to select an appropriate reference genome and sequencing coverage to ensure the accuracy of the results for a specific research question while balancing cost and feasibility. This chapter aims to evaluate the impacts of the reference genome and sequencing depth on population genomic analyzes. I selected four populations of two caddisfly species endemic to the Himalayas (*Himalopsyche digitata* and *H. tibetana*, respectively) and used them for individual-based shallow whole genome resequencing. Five *de novo* genomes (including three novel genomes) that were confamilial with the two target species at variable relatedness levels were applied as reference genomes. Additionally, resequencing data were randomly rarefied to three levels of sequencing coverage, which generated 30 datasets in total (five reference genomes × three coverages × two species) for the analyzes. Based on the SNPs called in these 30 datasets, we estimated population genetic indices and population structure. The results show that both distantly related reference genomes and lower sequencing coverage lead to resolution degradation. However, there are differences in the impact: the more closely related the reference genome is, the more stable the estimates of population genetic indices are, while the higher the sequencing depth, the better the resolution of the population structure analyzes. However, the results also vary depending on the inherent genetic variation of the target species. This research not only provides broad genomic resources but also demonstrates the effects of the reference genome and sequencing depth in population genetic analyzes, which will be valuable in helping other researchers optimize the selection of reference genome and sequencing depth for their own projects in the field of evolution and ecology.

Beyond the technical results, this research also revealed that the population structure of *H. digitata* and *H. tibetana* is highly consistent with the geographic distribution of populations within the drainage and river networks.

Chapter 3. Mountains are known for harboring high levels of biodiversity. To explain the evolution of mountain biodiversity, the MGH explores the interaction of topography, climate, and biology. In this Chapter, I investigate the diversification process of a group of caddisfly species that are endemic to the THR in the context of the MGH using evidence of WGS of 333 individuals and ecological modelling. One pair of *Himalopsyche* species from the HIM and another one from the HDM was selected, each species pair containing a species inhabiting high elevation and a species inhabiting low elevation. The results indicated that the high-elevation species showed a strong local differentiation in both the HIM and the HDM. In comparison, the low-elevation species were shaped by drainage basins, indicating greater regional dispersal activity. Results of demographic history and species distribution modeling supported a demographic expansion for all four caddisfly species during the LGM linked to increased potential habitat. This study demonstrates that climate fluctuations during the LGM promoted the species pump effect for caddisflies in the region, thus leading to a local or regional movement along the elevational gradient. This study provides the first comparative population genomic analysis of aquatic insects from remote mountain areas and integrates ecological modelling to reveal the species' glacial/postglacial ecology and evolution. This is thus the first study using massive empirical evidence of an aquatic insect group to explain the formation of mountain biodiversity in the THR. More importantly, it demonstrates a question of broader interest in biodiversity research: why are mountain regions so biodiverse?

3. Discussion

3.1 Genetic structure and differentiation of Himalopsyche in the HIM and the HDM

Due to the high dependency on water ecosystems, aquatic species generally show a distribution pattern that is constrained by the dendritic structure of the stream network (Tonkin et al., 2018). In previous studies, based on the level of gene flow among populations throughout the drainage networks, these genetic diversity patterns are specified as the stream hierarchy model, the death valley model, the headwater model, isolation by distance, and panmixia (or also called the widespread gene flow model, Hughes, 2007). In caddisflies, dispersal is often an "along-stream" movement conducted by both larvae and adults as defined in the context of the "colonization cycle" (Collier and Smith, 1997; Müller, 1954; Petersen et al., 2004; Winterbourn et al., 2007). In addition, overland dispersal (also called lateral dispersal) is also possible at the adult stage, which promotes gene flow among caddisfly populations from different catchments (Bowler and Benton, 2005; Collier and Smith, 1997; Deng et al., 2021; Engelhardt et al., 2011; Geismar et al., 2015; Griffith et al., 1998; Malicky, 1987; Müller-Peddinghaus, 2011; Smith and Smith, 2009; Svensson, 1974; Wilcock et al., 2007).

Hypothesis 1 proposed that as a group of these aquatic insects, *Himalopsyche* species show a general pattern of genetic diversity correlated with the drainage. This hypothesis was supported by the results of Chapters 1, 2, and 3, in particular at both species level and population level. The phylogenetic analyzes in Chapter 1 revealed that speciation of the *H. martynovi* complex occurred among groups located in different river basins. A similar pattern was shown in *H. platon*: genetic differentiation was observed among populations from different basins (Chapter 3). These results suggest that low-elevation species are well connected within the same basin, indicating that dispersal within the stream network appears to be the dominant process, thus in line with the stream hierarchy model.

However, although the population genetic inferences in Chapters 2 and 3 indicated diversification among populations from different basins, regardless of major basins (HIM: Gandaki and Koshi river basin; HDM: mostly between the Mekong and Yangtze river basin) or subbasins (basins of the tributary of each major basins), or the HIM and the HDM, we observed several exceptions that populations from different basins showed non-significant genetic variance, mainly for these high-elevation species. This is common for aquatic insects that inhabit high mountains or other types of fragmented distributions (Engelhardt et al., 2011; Pauls et al., 2006). These results support that instead of being connected by stream network, *Himalopsyche* species inhabiting high elevation have a good connection within the local mountain ranges but show limited gene flow among different mountains, thus suggesting that lowlands acted as the dispersal barrier for these high-elevation species. Therefore, the genetic pattern of the high-elevation species fits a "sky island" pattern or a headwater model.

Results from Chapters 2 and 3 also emphasize that genetic structure for aquatic organisms may not always be clear-cut when identifying the prevalent models, especially for taxa with a broad distribution (Hughes et al., 2013; Sproul et al., 2014). For instance, low-elevation species that show a stream hierarchy model pattern (*H. digitata* inhabiting the HIM) could also be interpreted as exhibiting isolation by distance since the genetic diversity also increased with geographic distance (or "stream distance", the same as geographic distance in this case). In addition, except for the distinct diversity pattern of high- and low-elevation species, the genetic patterns of some *Himalopsyche* are likely species-specific. For instance, for the species inhabiting the HDM, *H. gregoryi* (low-elevation) showed both headwater and stream hierarchy models. Overall, these complex patterns may result from either the complexity of the dendritic stream network embedded in a high-gradient landscape or simply the geographic scale of the studies examined (Hughes et al., 2013).

3.2 Key abiotic factors in shaping the montane biodiversity of Himalopsyche

Environmental heterogeneity is considered the crucial determinant of species diversity because a more heterogenous area provides more ecological niches for more species with different requirements (Allouche et al. 2012). In freshwater ecosystems, the longitudinal change from headwater to the river mouth leads to various microhabitats associated with water velocity, discharge, flood disturbance frequency, temperature, riparian inputs, salinity, and others, thus allowing numerous aquatic species to inhabit different niches along the river continuum (Múrria et al., 2012). This characteristic is largely enhanced in mountains due to the steep elevational gradients (Hotaling et al. 2017). Therefore, the abundant freshwater resources in the HIM and HDM derived from the complex topography and climate conditions may make it a cradle for freshwater diversity.

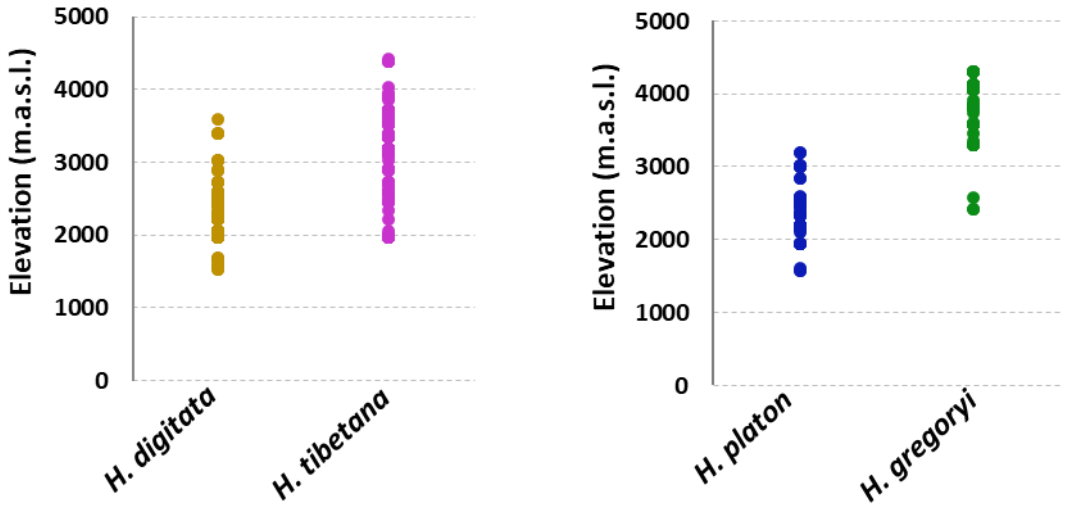


Fig. 4. Elevational distribution of the four *Himalopsyche* species in Chapters 2 and 3. Data from Hjalmarsson (2020) and Deng et al. (2023).

In line with Hjalmarsson (2020), we observed a clear diversification pattern along the altitudinal gradients among *Himalopsyche* species in the HIM and HDM (Fig. 4). As reported in previous research, caddisflies often show clear ecological preferences along elevational gradients, e.g., in the HIM and the Mediterranean area (Hoppeler et al., 2016; Múrria et al., 2012; Tonkin et al., 2018). To further assess the validity of these patterns in the HDM, we analyzed the relationship between 44 Trichoptera communities and potential environmental drivers, in particular elevation. The results revealed that the caddisfly community is significantly associated with environmental distance at the genus level (Fig. 5A). This indicates that caddisfly communities tend to be more divergent when the environmental distance between sampling sites increases (e.g., water temperature, conductivity, dissolved oxygen, substrate composition, and physical biotopes). Among these environmental factors, elevational distance significantly segregates community dissimilarity (Fig. 5B), suggesting that caddisfly communities in the HDM are differentiated along elevational gradients. Moreover, these results also show that the highest richness of caddisflies occurs between 2000 and 3000 m (Fig. 6). The elevational zonation determined here is similar to that observed in Himalayan Trichoptera communities of Nepal and may indicate a similar climate-sensitive zone in the HDM (Shah et al., 2015). Overall, the results at population, species, and community levels indicate that elevation is the primary organizational gradient of biodiversity of caddisflies in these mountain regions.

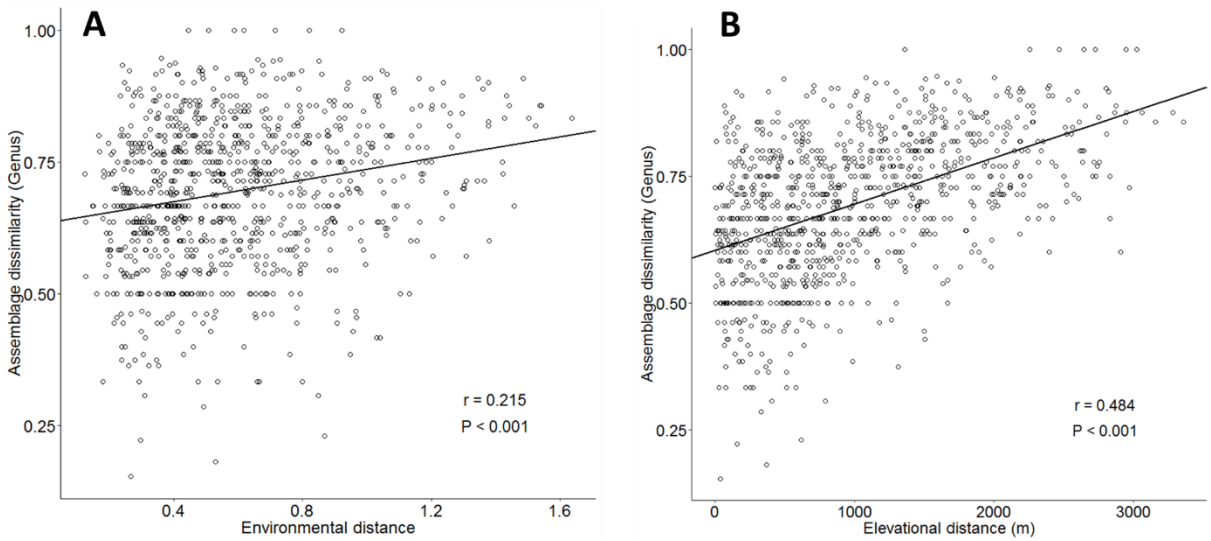


Fig. 5. Relationships between caddisfly communities dissimilarity and (A) environmental and (B) elevational distances. All samples from 44 communities were identified to a genus level. These communities were collected from all four main drainage basins (Yangtze, Mekong, Salween, and Irrawaddy) in HDM along a large elevational gradient ranging from 1000 to 4400 m a.s.l. Environmental distances between sites were calculated based on in-stream physicochemical variables (e.g., water temperature, conductivity, dissolved oxygen, composition of substrates, and physical biotopes). Relationships between assemblage dissimilarity (measured by Jaccard distance) and distance metrics (elevational and environmental) were examined with the Mantel test.[1]

[1] Data used for generating Fig. 4 and Fig. 5 is from the manuscript "Elevational patterns of Trichoptera diversity in Hengduan Mountains" which is in preparation by Deng et al.

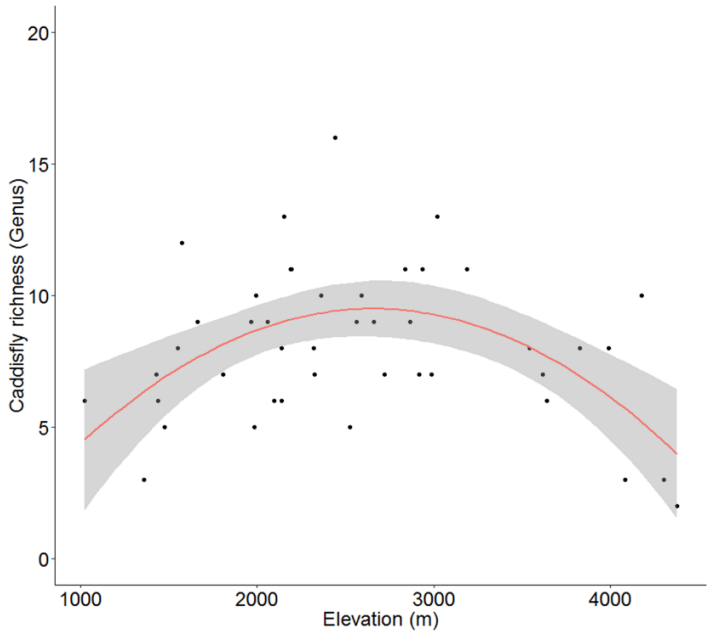


Fig. 6. Altitudinal pattern of caddisfly assemblage richness in the Hengduan Mountains. The same data was used as explained in Fig. 5. Both linear and polynomial curves were fitted to test the altitudinal richness pattern of caddisfly assemblage. The best fitted curve was selected.[1]

However, the ecological preferences of caddisflies we observed are not sufficient to fully explain the current distribution patterns of the target species. It is important to interpret the data in the context of a species' history, as proposed in hypothesis 2. The study presented in Chapters 2 and 3 provides considerable empirical evidence to support hypothesis 2. The results revealed that the evolution of climate, mountain topography, and historical drainage rearrangement are key factors for the evolution of aquatic biodiversity in the THR. These factors have interacted with each other over time and vastly shaped current diversity patterns in these aquatic insects. This is in line with most of the previous studies on flora in this mountain region (Antonelli et al., 2018; Ding et al., 2020; Favre et al., 2015; Mosbrugger et al., 2018; Muellner-Riehl, 2019; Rahbek et al., 2019a, 2019b). Yet, our study is the first empirical study using both genomic and ecological data to explain the diversification history of aquatic insects in the THR since LGM. In addition, it promotes our understanding of the origin and evolution of mountain biodiversity, not only in the THR but in all mountain systems on Earth.

3.3 Implication of mountain biodiversity concepts in different mountain systems

To explain the formation of spatially complex patterns of biodiversity in the present day in mountain areas, scientists proposed many concepts during the time, for instance, the term "sky island" (Heald, 1951), the Flickering Connectivity System (FCS, Flantua et al., 2018), and the Mountain Geo-biodiversity Hypothesis (MGH, Mosbrugger et al., 2018). These concepts were originally developed to systematically study the evolutionary history of diversification and speciation processes in the mountains. They were then tested in different mountain systems, such as southeastern Arizona ("sky island"), northern Andes (FCS), and the THR (MGH). I attempt to refer to these concepts to inspect the results of genetic diversity patterns of caddisflies throughout the study, in particular in different mountain regions (the HIM and the HDM), as proposed in hypothesis 3.

As summarized in the first two sections of the discussion, the historical evolution of abiotic conditions such as changing topography (including paleo-drainage reorganization) and climate are likely to have played dominant roles in shaping the genetic diversity patterns of caddisflies in the HIM and the HDM. This pattern can be well explained by the MGH, especially when revising it to the three boundary conditions of the MGH (i.e., elevational gradients, response to climatic fluctuations, and a high-relief terrain with environmental gradients; Mosbrugger et al., 2018). After the initial India-Eurasia collision since the Paleocene (65–60 Ma), the HIM and the HDM reached their modern height before the Pliocene (as reviewed by Ding et al. in 2022). Instead of the direct influence of mountain uplifting, the intraspecific divergence is mainly caused by the interconnectivity of habitats among and within each mountainous region (Craw et al., 2016; Rahbek et al., 2019b). Therefore, when facing climate fluctuations, such as the LGM, species may migrate along the elevational gradients to find a new habitat, causing habitat connectivity and interconnectivity and thus leading to diversification.

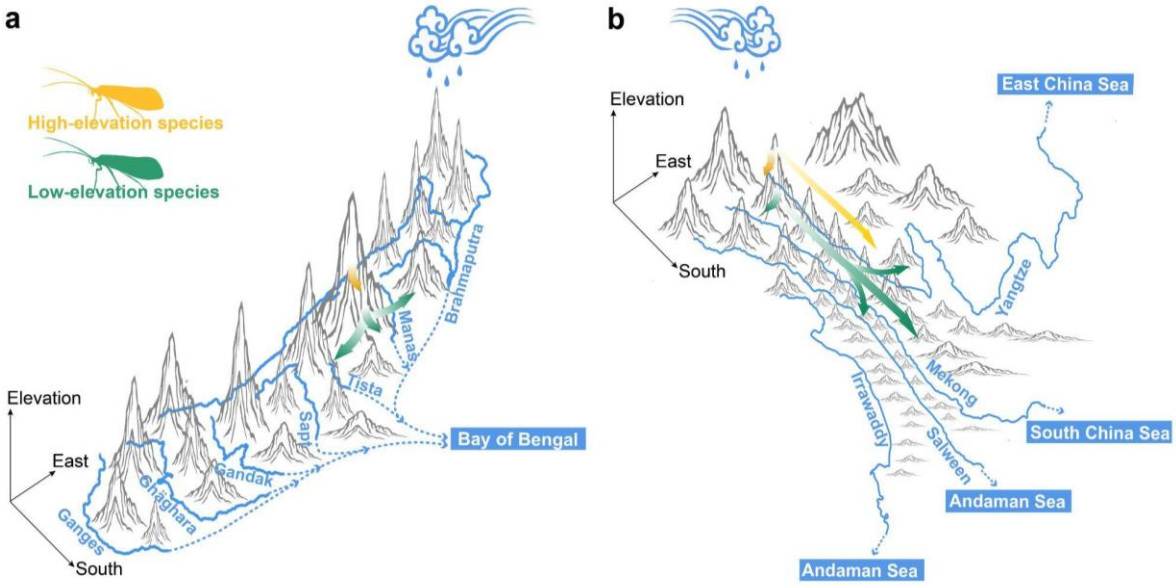


Fig. 7. Diagrammatic sketch of the topographic landscape in the Himalayas (a) and the Hengduan Mountains (b) and the potential *in-situ* displacement (short arrows) and long-distance dispersal (long arrows) of caddisflies during the LGM. Different colors of the caddisflies indicate high-elevation species (yellow) and low-elevation species (green). During the glacial period, which means a cooler climate, a vertical displacement along the elevational gradient (*in-situ* displacement) may occur on both high- and low-elevation caddisfly species. However, compared to the Himalayas, a wider extent of mountain ranges along the North-South orientation in the Hengduan Mountains gives the possibility for species to disperse further south (long-distance dispersal). During the interglacial period, the same dispersal or replacement may happen but in the opposite direction. Therefore, geographic connection and isolation may repeatedly occur during climate fluctuation, thus forming the so-called "species pump" effect and promoting speciation or diversification, according to the MGH (Mosbrugger et al., 2018).

However, the genetic diversity patterns of the *Himalopsyche* species in the HIM and the HDM are distinct. Compared to the HIM, populations in the HDM show better connectivity within or among river basins. For instance, *H. platon*, the low-elevation species distributed in the HDM, appears panmictic across all four major basins. This indicates that populations from most basins in the HDM are highly connected, and gene flow is strong within all four primary basins until recent times. This is consistent with the unique geological landscape of the HDM, for instance, the extended range from north to south and the parallel rivers inlaid in the North-South mountains. Therefore, during a cooling period, species have the chance to disperse from cold or glacial areas (north) to warmer places (south) along the long valleys, as verified by the results of SDM (Fig. 7).

Based on these insights, the results support hypothesis 3 and demonstrate that the implication of these mountain biodiversity concepts are distinct in different mountain systems because the topography and climate conditions are unique in each mountain system. This is in line with the concept of "fingerprints" as proposed in the FCS, which defines the interaction between topography and climate as the "fingerprint" of a mountain (Flantua et al., 2018). Although the "fingerprints" are not yet specified for all the mountain systems of the world, our study revealed that the "fingerprints" differ between the HIM and the HDM.

3.4 Next-generation sequencing is a powerful tool in phylogenetics and phylogeographic studies of caddisflies

By using Sanger sequencing, the phylogenetic relationship of the genus *Himalopsyche* has been greatly resolved in recent years (Hjalmarsson et al., 2019, 2018). Yet, along with the development of NGS technologies, WGS would be the next step to address unsolved research questions of *Himalopsyche*, for instance sorting the lineages of species complex, variation within species, and historical demographics. Thus, to answer these research questions, we applied a variety of sequencing strategies throughout this thesis, including whole genome *de novo* assembly by combining both long- and short-reads, WGS, and target sequencing using only short-reads.

For most of the *Himalopsyche* species, a reference genome is not available. For these non-model organisms, anchored hybrid enrichment (AHE) is a powerful and affordable approach for quickly obtaining many orthologous loci across the genome for deep and shallow phylogenetic studies (Lemmon et al., 2012). The results of Chapter 1 have shown that data generated by AHE successfully resolved the shallow phylogeny of the *H. martynovi* complex and was able to detect ancestral gene flow. Moreover, the phylogenetic results were verified by morphological evidence, thus suggesting that gene flow may be the source of morphological variation in this species complex.

Yet, whole genome resources are more and more widely used in evolutionary biology due to the decreasing cost and fast-developing technologies. A complete genome of a species would be an

essential resource in many research efforts (Ballouz et al., 2019). Genome-wide data can greatly facilitate the research scope of biology; for instance, it makes the intention of identifying all genes in a given species possible, or it allows a comprehensive investigation of genetic differences at a genomic scale both within and among species (Ballouz et al., 2019). Using a high-quality genome of target species as the reference genome will be a great benefit in a population genetic study because it helps to generate a massive number of effective information, such as the genome-wide variation, which can remarkably improve the downstream inferences (Günther and Nettelblad, 2019; Wright et al., 2019). In Chapter 2, we generated three novel *de novo* assemblies of three *Himalopsyche* species from three different morphotypes, which largely enriched the genomic database of aquatic insects. In addition, by comparing the impacts of reference genome selection and sequencing depth on population genomic analyzes, we highlighted that population genetic studies would benefit the most from closely related reference genomes, especially as the costs of obtaining a high-quality reference genome continue to decrease. Importantly, the results in this chapter revealed that when a conspecific reference genome is available, 12.5 × coverage is more than enough to do individual-based population genomics. This finding in Chapter 2 is a cornerstone for the methodology employed in the study in Chapter 3.

Based on the instruction on how to select a reference genome and the generated novel genomic resources in Chapter 2, it is thus possible to study the genetic variation and demographic history of *Himalopsyche* species at a population level. As represented in Chapters 1 and 3, compared to AHE, the number of variants discovered by WGS dramatically increased, from thousands (AHE) to millions (WGS). Therefore, the genome-wide variants successfully revealed the genetic pattern and historical demography of *Himalopsyche* species that inhabit different niches in the HIM and the HDM. Above all, by using different sequencing technologies, we addressed various scientific questions at the species and population levels in Chapters 1, 2, and 3. We emphasize that depending on the focal research scope, a proper sequencing strategy needs to be carefully considered.

3.5 Conclusions and Outlook

The central aim of this work was to better understand the formation of mountain biodiversity, particularly in the THR. By applying various NGS sequencing strategies, I addressed three hypotheses that aim to explain the speciation and diversification processes of *Himalopsyche* at species and population levels. The results shed light on the patterns of population genetic structure and diversity of aquatic insects in the THR, the interaction of strong topographic gradients with historic climate change as the driver of these patterns, and that individual-based population genomic studies are increasingly feasible for assessing these types of questions.

By investigating the current diversity patterns of the *Himalopsyche* species inhabiting diverse elevational environments (high *versus* low) and distinct mountain regions (HIM *versus* HDM), this

work revealed that the *Himalopsyche* species generally show a pattern of genetic diversity correlated with drainage basins. Yet, compared to the low-elevation species, which show a clear genetic connection within basins, high-elevation species also exhibit a connection among basins within the local mountain ranges. In addition, the genetic patterns of the examined *Himalopsyche* species appear to be species-specific and not universal. While this pattern has been observed in European aquatic insects (e.g., Engelhardt et al., 2011; Lehrian et al., 2009; Pauls et al., 2009), this is the first population genetic study of caddisflies from mountain region in the THR to show this pattern may be widespread in aquatic insects or at least Trichoptera. The findings thus enrich our knowledge of genetic diversity patterns of caddisflies throughout the drainage networks.

This work comprises the first empirical studies using both genomic and ecological data to explain the diversification history of aquatic insects in the THR since LGM. Regarding the key environmental factors that form the different diversity patterns of caddisflies, my results show that elevation seems to be the primary organizational factor for biodiversity of caddisflies in these mountain regions. Yet, historical climate change, mountain topography, and drainage rearrangement are additional factors driving the evolution of aquatic biodiversity in the THR.

The combination of the findings generally supports the MGH, and that the MGH is not only relevant for plants but also for aquatic insects and quite likely other animals as well. However, the results also demonstrate that the validity of the MGH and other mountain biodiversity concepts are distinct in different mountain systems due to differences in latitudinal extent and topographical complexity. Given the fact that species react differently to changes in their environment, more studies are needed to further explore how the entire biota responds when facing climatic challenges. And it is also important to test the MGH in other mountain regions worldwide.

Methodologically, the work includes the first interspecific gene flow study on caddisflies using genomic data captured by anchor hybrid enrichment, as well as the first individual-based genome wide resequencing studies for population genomics in caddisflies. While the approaches are clearly feasible, my results emphasize that population genetic studies using WGS would benefit more from reference genomes that are closely related to the target species.

Clearly, high-quality *de novo* genomes of more *Himalopsyche* species are needed as a basis to further study evolutionary processes in *Himalopsyche* and also more generally to fully harness the potential of population genomic studies. For example, with more reference genomes, the type resequencing data generated here could also be used for genome-wide association studies to identify specific genes that are associated with high altitude adaptation, thereby providing better insight on adaptation and diversification of *Himalopsyche* or other species in the high mountain area.

4. References

Allouche, O., Kalyuzhny, M., Moreno-Rueda, G., Pizarro, M., Kadmon, R., 2012. Area–heterogeneity tradeoff and the diversity of ecological communities. Proc. Natl. Acad. Sci. U.S.A. 109, 17495–17500. https://doi.org/10.1073/pnas.1208652109

Anderson, N.H., 1967. Life cycle of a terrestrial caddisfly, *Philocasca demita* (Trichoptera: Limnephilidae), in North America1. Ann. Entomol. Soc. Am. 60, 320–323. https://doi.org/10.1093/aesa/60.2.320

Antonelli, A., Kissling, W.D., Flantua, S.G.A., Bermúdez, M.A., Mulch, A., Muellner-Riehl, A.N., Kreft, H., Linder, H.P., Badgley, C., Fjeldså, J., Fritz, S.A., Rahbek, C., Herman, F., Hooghiemstra, H., Hoorn, C., 2018. Geological and climatic influences on mountain biodiversity. Nat. Geosci. 11, 718–725. https://doi.org/10.1038/s41561-018-0236-z

Ashokan, A., Xavier, A., Suksathan, P., Ardiyani, M., Leong-Škorničková, J., Newman, M., Kress, W.J., Gowda, V., 2022. Himalayan orogeny and monsoon intensification explain species diversification in an endemic ginger (*Hedychium*: Zingiberaceae) from the Indo-Malayan Realm. Mol. Phylogenet. Evol. 170, 107440. https://doi.org/10.1016/j.ympev.2022.107440

Atwood, T.C., Young, J.K., Beckmann, J.P., Breck, S.W., Fike, J., Rhodes, O.E., Bristow, K.D., 2011. Modeling connectivity of black bears in a desert sky island archipelago. Biol. Conserv. 144, 2851–2862. https://doi.org/10.1016/j.biocon.2011.08.002

Ballouz, S., Dobin, A., Gillis, J.A., 2019. Is it time to change the reference genome? Genome Biol. 20, 159. https://doi.org/10.1186/s13059-019-1774-4

Bartonova, A., Konvicka, M., Korb, S., Kramp, K., Schmitt, T., Faltynek Fric, Z., 2018. Range dynamics of Palaearctic steppe species under glacial cycles: the phylogeography of Proterebia afra (Lepidoptera: Nymphalidae: Satyrinae). Biological Journal of the Linnean Society. https://doi.org/10.1093/biolinnean/bly136

Behjati, S., Tarpey, P.S., 2013. What is next generation sequencing? Arch. Dis. Child. - Educ. Pract. Ed. 98, 236–238. https://doi.org/10.1136/archdischild-2013-304340

Bhattarai, K.R., Måren, I.E., Subedi, S.C., 2014. Biodiversity and invasibility: distribution patterns of invasive plant species in the Himalayas, Nepal. J. Mt. Sci. 11, 688–696. https://doi.org/10.1007/s11629-013-2821-3

Bookhagen, B., Burbank, D.W., 2010. Toward a complete Himalayan hydrological budget: spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. J. Geophys. Res. 115, F03019. https://doi.org/10.1029/2009JF001426

Bowler, D.E., Benton, T.G., 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. Biol. Rev. 80, 205–225. https://doi.org/10.1017/S1464793104006645

Carstens, B., Lemmon, A.R., Lemmon, E.M., 2012. The promises and pitfalls of next-generation sequencing data in phylogeography. Syst. Biol. 61, 713–715. https://doi.org/10.1093/sysbio/sys050

Collier, K.J., Smith, B.J., 1997. Dispersal of adult caddisflies (Trichoptera) into forests alongside three New Zealand streams. Hydrobiologia 361, 53–65. https://doi.org/10.1023/A:1003133208818

Cox, S.C., Prys-Jones, R.P., Habel, J.C., Amakobe, B.A., Day, J.J., 2014. Niche divergence promotes rapid diversification of East African sky island white-eyes (Aves: Zosteropidae). Mol. Ecol. 23, 4103–4118. https://doi.org/10.1111/mec.12840

Craw, D., Upton, P., Burridge, C.P., Wallis, G.P., Waters, J.M., 2016. Rapid biological speciation driven by tectonic evolution in New Zealand. Nat. Geosci. 9, 140–144. https://doi.org/10.1038/ngeo2618

Cross, H., Biffin, E., Van Dijk, K., Lowe, A., Waycott, M., 2016. Effective application of next-generation sequencing (NGS) approaches in systematics and population genetics: case studies in *Eucalyptus* and *Acacia*. Aust. Syst. Bot. 29, 235. https://doi.org/10.1071/SB16019

Dahal, N., Lamichhaney, S., Kumar, S., 2021. Climate change impacts on Himalayan biodiversity: evidence-based perception and current approaches to evaluate threats under climate change. J. Indian Inst. Sci. 101, 195–210. https://doi.org/10.1007/s41745-021-00237-1

DeChaine, E.G., Martin, A.P., 2005. Marked genetic divergence among sky island populations of *Sedum lanceolatum* (Crassulaceae) in the Rocky Mountains. Am. J. Bot. 92, 477–486. https://doi.org/10.3732/ajb.92.3.477

Dénes, A.-L., Kolcsár, L.-P., Török, E., Keresztes, L., 2016. Phylogeography of the micro-endemic *Pedicia staryi* group (Insecta: Diptera): evidence of relict biodiversity in the Carpathians. Biol. J. Linn. Soc. 119, 719–731. https://doi.org/10.1111/bij.12667

Deng, X., Domisch, S., Favre, A., Jähnig, S., Frandsen, P., He, F., Shah, D.N., Shah, R.D.T., Cai, Q., Pauls, S., 2023. Comparative phylogeography of *Himalopsyche* (Trichoptera, Rhyacophilidae) in the Tibeto-Himalayan Region: an assessment of the mountain-geobiodiversity hypothesis (preprint). Preprints. https://doi.org/10.22541/au.167888198.83706514/v1

Deng, X., Frandsen, P.B., Dikow, R.B., Favre, A., Shah, D.N., Shah, R.D.T., Schneider, J.V., Heckenhauer, J., Pauls, S.U., 2022. The impact of sequencing depth and relatedness of the reference genome in population genomic studies: A case study with two caddisfly species (Trichoptera, Rhyacophilidae, *Himalopsyche*). Ecol. Evol. 12. https://doi.org/10.1002/ece3.9583

Deng, X.-L., Favre, A., Lemmon, E.M., Lemmon, A.R., Pauls, S.U., 2021. Gene flow and diversification in *Himalopsyche martynovi* dpecies complex (Trichoptera: Rhyacophilidae) in the Hengduan Mountains. Biology 10, 816. https://doi.org/10.3390/biology10080816

Ding, L., Kapp, P., Cai, F., Garzione, C.N., Xiong, Z., Wang, H., Wang, C., 2022. Timing and mechanisms of Tibetan Plateau uplift. Nat. Rev. Earth Environ. 3, 652–667. https://doi.org/10.1038/s43017-022-00318-4

Ding, W.-N., Ree, R.H., Spicer, R.A., Xing, Y.-W., 2020. Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. Science 369, 578–581. https://doi.org/10.1126/science.abb4484

Engelhardt, C.H., Haase, P., Pauls, S.U., 2011. From the Western Alps across Central Europe: postglacial recolonisation of the tufa stream specialist *Rhyacophila pubescens* (Insecta, Trichoptera). Front. Zool. 8, 10. https://doi.org/10.1186/1742-9994-8-10

Favre, A., Päckert, M., Pauls, S.U., Jähnig, S.C., Uhl, D., Michalak, I., Muellner-Riehl, A.N., 2015. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. Biol. Rev. 90, 236–253. https://doi.org/10.1111/brv.12107

Flantua, S.G., Hooghiemstra, H., Hoorn, C., 2018. Historical connectivity and mountain biodiversity, in: Hoorn, C., Perrigo, A., Antonelli, A. (Eds.), Mountains, climate and biodiversity. John Wiley & Sons, Hoboken, NJ, pp. 171–185.

Fu, P.-C., Sun, S.-S., Hollingsworth, P.M., Chen, S.-L., Favre, A., Twyford, A.D., 2022. Population genomics reveal deep divergence and strong geographical structure in *gentians* in the Hengduan Mountains. Front. Plant Sci. 13, 936761. https://doi.org/10.3389/fpls.2022.936761

Fu, P.-C., Sun, S.-S., Khan, G., Dong, X.-X., Tan, J.-Z., Favre, A., Zhang, F.-Q., Chen, S.-L., 2020. Population subdivision and hybridization in a species complex of *Gentiana* in the Qinghai-Tibetan Plateau. Ann. Bot. 125, 677–690. https://doi.org/10.1093/aob/mcaa003

Fuchs, E.J., Cascante-Marín, A., Madrigal-Brenes, R., Quesada, M., 2023. Genetic diversity and phylogeographic patterns of the dioecious palm *Chamaedorea tepejilote* (Arecaceae) in Costa Rica: the role of mountain ranges and possible refugia. AoB PLANTS 15, plac060. https://doi.org/10.1093/aobpla/plac060

Garrick, R.C., Bonatelli, I.A.S., Hyseni, C., Morales, A., Pelletier, T.A., Perez, M.F., Rice, E., Satler, J.D., Symula, R.E., Thomé, M.T.C., Carstens, B.C., 2015. The evolution of phylogeographic data sets. Mol. Ecol. 24, 1164–1171. https://doi.org/10.1111/mec.13108

Gébelin, A., Mulch, A., Teyssier, C., Jessup, M.J., Law, R.D., Brunel, M., 2013. The Miocene elevation of Mount Everest. Geology 41, 799–802. https://doi.org/10.1130/G34331.1

Geismar, J., Haase, P., Nowak, C., Sauer, J., Pauls, S.U., 2015. Local population genetic structure of the montane caddisfly *Drusus discolor* is driven by overland dispersal and spatial scaling. Freshw. Biol. 60, 209–221. https://doi.org/10.1111/fwb.12489

Gentili, R., Bacchetta, G., Fenu, G., Cogoni, D., Abeli, T., Rossi, G., Salvatore, M.C., Baroni, C., Citterio, S., 2015. From cold to warm-stage refugia for boreo-alpine plants in southern

European and Mediterranean mountains: the last chance to survive or an opportunity for speciation? Biodiversity 16, 247–261. https://doi.org/10.1080/14888386.2015.1116407

Grada, A., Weinbrecht, K., 2013. Next-generation sequencing: methodology and application. J. Invest. Dermatol. 133, 1–4. https://doi.org/10.1038/jid.2013.248

Green, E.D., Rubin, E.M., Olson, M.V., 2017. The future of DNA sequencing. Nature 550, 179–181. https://doi.org/10.1038/550179a

Griffith, M.B., Barrows, E.M., Perry, S.A., 1998. Lateral dispersal of adult aquatic insects (Plecoptera, Trichoptera) following emergence from headwater streams in forested Appalachian catchments. Ann. Entomol. Soc. Am. 91, 195–201. https://doi.org/10.1093/aesa/91.2.195

Günther, T., Nettelblad, C., 2019. The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLOS Genet. 15, e1008302. https://doi.org/10.1371/journal.pgen.1008302

He, K., Gutiérrez, E.E., Heming, N.M., Koepfli, K., Wan, T., He, S., Jin, W., Liu, S., Jiang, X., 2019. Cryptic phylogeographic history sheds light on the generation of species diversity in sky-island mountains. J. Biogeogr. 46, 2232–2247. https://doi.org/10.1111/jbi.13664

He, K., Jiang, X., 2014. Sky islands of southwest China. I: an overview of phylogeographic patterns. Chin. Sci. Bull. 59, 585–597. https://doi.org/10.1007/s11434-013-0089-1

Heald, W., 1967. Sky Island. D. Van Nostrand Co., Inc, Princeton, New Jersey.

Heald, W., 1951. Sky islands of Arizona. Nat. Hist. 60, 56–63.

Hedin, M., Carlson, D., Coyle, F., 2015. Sky island diversification meets the multispecies coalescent - divergence in the spruce-fir moss spider (*Microhexura montivaga*, Araneae, Mygalomorphae) on the highest peaks of Southern Appalachia. Mol. Ecol. 24, 3467–3484. https://doi.org/10.1111/mec.13248

Hendricks, S., Anderson, E.C., Antao, T., Bernatchez, L., Forester, B.R., Garner, B., Hand, B.K., Hohenlohe, P.A., Kardos, M., Koop, B., Sethuraman, A., Waples, R.S., Luikart, G., 2018. Recent advances in conservation and population genomics data analysis. Evol. Appl. 11, 1197–1211. https://doi.org/10.1111/eva.12659

Hjalmarsson, A.E., 2020. Phylogeny and species delimitation of *Himalopsyche* (Trichoptera, Rhyacophilidae) (Doctoral dissertation). Johann Wolfgang Goethe-Universität, Frankfurt am Main.

Hjalmarsson, A.E., 2019. Delimitation and description of three new species of *Himalopsyche* (Trichoptera: Rhyacophilidae) from the Hengduan Mountains, China. Zootaxa 4638, 419–441. https://doi.org/10.11646/zootaxa.4638.3.7

Hjalmarsson, A.E., Graf, W., Jähnig, S.C., Vitecek, S., Pauls, S.U., 2018. Molecular association and morphological characterisation of *Himalopsyche* larval types (Trichoptera, Rhyacophilidae). ZooKeys 773, 79–108. https://doi.org/10.3897/zookeys.773.24319

Hjalmarsson, A.E., Graf, W., Vitecek, S., Jähnig, S.C., Cai, Q., Sharma, S., Tong, X., Li, F., Shah, D.N., Shah, R.D.T., Pauls, S.U., 2019. Molecular phylogeny of *Himalopsyche* (Trichoptera, Rhyacophilidae). Syst. Entomol. 44, 973–984. https://doi.org/10.1111/syen.12367

Holzenthal, R.W., Blahnik, R.J., Prather, A.L., Kjer, K.M., 2007. Order trichoptera kirby, 1813 (insecta), caddisflies.

Hoppeler, F., Tachamo Shah, R.D., Shah, D.N., Jähnig, S.C., Tonkin, J.D., Sharma, S., Pauls, S.U., 2016. Environmental and spatial characterisation of an unknown fauna using DNA sequencing - an example with Himalayan Hydropsychidae (Insecta: Trichoptera). Freshw. Biol. 61, 1905–1920. https://doi.org/10.1111/fwb.12824

Hotaling, S., Hood, E., Hamilton, T.L., 2017. Microbial ecology of mountain glacier ecosystems: biodiversity, ecological connections and implications of a warming climate. Environmental Microbiology 19, 2935–2948. https://doi.org/10.1111/1462-2920.13766

Hughes, J.M., 2007. Constraints on recovery: using molecular methods to study connectivity of aquatic biota in rivers and streams. Freshw. Biol. 52, 616–631. https://doi.org/10.1111/j.1365-2427.2006.01722.x

Hughes, J.M., Huey, J.A., Schmidt, D.J., 2013. Is realised connectivity among populations of aquatic fauna predictable from potential connectivity? Freshw. Biol. 58, 951–966. https://doi.org/10.1111/fwb.12099

Janitz, M., 2008. Next generation genome sequencing: towards personalized medicine, 1st ed. Wiley. https://doi.org/10.1002/9783527625130

Jaramillo-Correa, J.P., Beaulieu, J., Khasa, D.P., Bousquet, J., 2009. Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. Can. J. For. Res. 39, 286–307.

Jehamalar, E.E., Gloda, D., Kiruba, S., Das, S., 2010. Trichopterans as a bioindicators of a stream ecosystem. J Basic Applied Biol 4, 86–90.

Kang, S., Zhang, Q., Qian, Y., Ji, Z., Li, C., Cong, Z., Zhang, Y., Guo, J., Du, W., Huang, J., You, Q., Panday, A.K., Rupakheti, M., Chen, D., Gustafsson, Ö., Thiemens, M.H., Qin, D., 2019. Linking atmospheric pollution to cryospheric change in the Third Pole region: current progress and future prospects. Natl. Sci. Rev. 6, 796–809. https://doi.org/10.1093/nsr/nwz031

Knowles, L.L., 2000. Test of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western north America. Evolution 54, 1337–1348. https://doi.org/10.1111/j.0014-3820.2000.tb00566.x

Kutnjak, D., Kuttner, M., Niketić, M., Dullinger, S., Schönswetter, P., Frajman, B., 2014. Escaping to the summits: phylogeography and predicted range dynamics of *Cerastium dinaricum*, an endangered high mountain plant endemic to the western Balkan Peninsula. Mol. Phylogenet. Evol. 78, 365–374. https://doi.org/10.1016/j.ympev.2014.05.015

Lacey Knowles, L., Alvarado-Serrano, D.F., 2010. Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. Mol. Ecol. 19, 3727–3745. https://doi.org/10.1111/j.1365-294X.2010.04702.x

Landeiro, V.L., Bini, L.M., Melo, A.S., Pes, A.M.O., Magnusson, W.E., 2012. The roles of dispersal limitation and environmental conditions in controlling caddisfly (Trichoptera) assemblages: Environmental and spatial factors controlling caddisfly distributions. Freshw. Biol. 57, 1554–1564. https://doi.org/10.1111/j.1365-2427.2012.02816.x

Lawson, L.P., 2013. Diversification in a biodiversity hot spot: landscape correlates of phylogeographic patterns in the African spotted reed frog. Mol. Ecol. 22, 1947–1960. https://doi.org/10.1111/mec.12229

Lehrian, S., Pauls, S.U., Haase, P., 2009. Contrasting patterns of population structure in the montane caddisflies *Hydropsyche tenuis* and *Drusus discolour* in the Central European highlands. Freshw. Biol. 54, 283–295. https://doi.org/10.1111/j.1365-2427.2008.02107.x

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst. Biol. 61, 727–744. https://doi.org/10.1093/sysbio/sys049

Lexer, C., Mangili, S., Bossolini, E., Forest, F., Stölting, K.N., Pearman, P.B., Zimmermann, N.E., Salamin, N., 2013. 'Next generation' biogeography: towards understanding the drivers of species diversification and persistence. J. Biogeogr. 40, 1013–1022. https://doi.org/10.1111/jbi.12076

Li, Q., Sun, H., Boufford, D.E., Bartholomew, B., Fritsch, P.W., Chen, J., Deng, T., Ree, R.H., 2021. Grade of membership models reveal geographical and environmental correlates of floristic structure in a temperate biodiversity hotspot. New Phytol. 232, 1424–1435. https://doi.org/10.1111/nph.17443

Li, Z., Lau, W.K. -M., Ramanathan, V., Wu, G., Ding, Y., Manoj, M.G., Liu, J., Qian, Y., Li, J., Zhou, T., Fan, J., Rosenfeld, D., Ming, Y., Wang, Y., Huang, J., Wang, B., Xu, X., Lee, S. -S., Cribb, M., Zhang, F., Yang, X., Zhao, C., Takemura, T., Wang, K., Xia, X., Yin, Y., Zhang, H., Guo, J., Zhai, P.M., Sugimoto, N., Babu, S.S., Brasseur, G.P., 2016. Aerosol and monsoon climate interactions over Asia. Rev. Geophys. 54, 866–929. https://doi.org/10.1002/2015RG000500

Liang, Q., Xu, X., Mao, K., Wang, M., Wang, K., Xi, Z., Liu, J., 2018. Shifts in plant distributions in response to climate warming in a biodiversity hotspot, the Hengduan Mountains. J. Biogeogr. 45, 1334–1344. https://doi.org/10.1111/jbi.13229

Luikart, G., Kardos, M., Hand, B.K., Rajora, O.P., Aitken, S.N., Hohenlohe, P.A., 2018. Population genomics: advancing understanding of nature, in: Rajora, O.P. (Ed.), Population genomics, population genomics. Springer International Publishing, Cham, pp. 3–79. https://doi.org/10.1007/13836_2018_60

Malicky, H., 2011. Neue Trichopteren aus Europa und Asien. Braueria 23–43 (in German).

Malicky, H., 2008. Beschreibungen von neuen Trichopteren aus Asien. Braueria 35, 45–57 (in German).

Malicky, H., 1987. Anflugdistanz und Fallenfangbarkeit von Köcherfliegen (Trichoptera) bei Lichtfallen. Jahresber. Biol. Stn. Lunz 10, 140–157 (in German).

Marchese, C., 2015. Biodiversity hotspots: A shortcut for a more complicated concept. Glob. Ecol. Conserv. 3, 297–309. https://doi.org/10.1016/j.gecco.2014.12.008

Martynov, A.B., 1935. On a collection of Trichopiera from the Indian Museum. Part I. Annulipalpia. Rec. Zool. Surv. India 37, 93–209.

Martynov, A.B., 1930. On the Trichopterons Fauna of China and Eastern Tibet. Proc. Zool. Soc. Lond. 100, 65–112.

Masta, S.E., 2000. Phylogeography of the jumping spider *Habronattus pugillis* (Araneae: Salticidae): recent vicariance of sky island populations? Evolution 54, 1699–1711.

Mastretta-Yanes, A., Moreno-Letelier, A., Piñero, D., Jorgensen, T.H., Emerson, B.C., 2015. Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. J. Biogeogr. 42, 1586–1600. https://doi.org/10.1111/jbi.12546

McCormack, J.E., Bowen, B.S., Smith, T.B., 2008. Integrating paleoecology and genetics of bird populations in two sky island archipelagos. BMC Biol. 6, 28. https://doi.org/10.1186/1741-7007-6-28

Morse, J.C., Frandsen, P.B., Graf, W., Thomas, J.A., 2019. Diversity and Ecosystem Services of Trichoptera. Insects 10, 125. https://doi.org/10.3390/insects10050125

Mosbrugger, V., Favre, A., Muellner-Riehl, A.N., Päckert, M., Mulch, A., 2018. Cenozoic evolution of geo-biodiversity in the Tibeto-Himalayan region, in: Hoorn, C., Perrigo, A., Antonelli, A. (Eds.), Mountains, climate and biodiversity. John Wiley & Sons, Hoboken, NJ, pp. 429.

Mu, Q.-Y., Yu, C.-C., Wang, Y., Han, T.-S., Wang, H., Ding, W.-N., Zhang, Q.-Y., Low, S.L., Zheng, Q.-J., Peng, C., Hu, Z.-Y., Xing, Y.-W., 2022. Comparative phylogeography of *Acanthocalyx* (Caprifoliaceae) reveals distinct genetic structures in the Himalaya–Hengduan Mountains. Alp. Bot. 132, 153–168. https://doi.org/10.1007/s00035-021-00262-x

Muellner-Riehl, A.N., 2019. Mountains as evolutionary arenas: patterns, emerging approaches, paradigm shifts, and their implications for plant phylogeographic research in the Tibeto-Himalayan Region. Front. Plant Sci. 10, 195. https://doi.org/10.3389/fpls.2019.00195

Muellner-Riehl, A.N., Favre, A., 2021. Mountain biogeography coming full circle: a new '3D' floristic approach provides units for reconstructing evolutionary trajectories. New Phytol. 232, 964–966. https://doi.org/10.1111/nph.17645

Müller, K., 1954. Investigations on the organic drift in north Swedish streams. Rep. Inst. Freshw. Res. 35, 133–148.

Müller-Peddinghaus, E., 2011. Flight-morphology of Central European caddisflies (Insecta: Trichoptera) in relation to their ecological preferences (Doctoral dissertation). Universität Duisburg-Essen, Duisburg, Essen.

Múrria, C., Bonada, N., Arnedo, M.A., Zamora-Muñoz, C., Prat, N., Vogler, A.P., 2012. Phylogenetic and ecological structure of Mediterranean caddisfly communities at various spatio-temporal scales: Evolution of aquatic insect assemblages. J. Biogeogr. 39, 1621–1632. https://doi.org/10.1111/j.1365-2699.2012.02729.x

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858. https://doi.org/10.1038/35002501

Owen, L.A., Dortch, J.M., 2014. Nature and timing of Quaternary glaciation in the Himalayan–Tibetan orogen. Quat. Sci. Rev. 88, 14–54. https://doi.org/10.1016/j.quascirev.2013.11.016

Pauls, S.U., Lumbsch, H.T., Haase, P., 2006. Phylogeography of the montane caddisfly *Drusus discolor*: evidence for multiple refugia and periglacial survival. Mol. Ecol. 15, 2153–2169. https://doi.org/10.1111/j.1365-294X.2006.02916.x

Pauls, S.U., Theissinger, K., Ujvarosi, L., Balint, M., Haase, P., 2009. Patterns of population structure in two closely related, partially sympatric caddisflies in Eastern Europe: Historic introgression, limited dispersal, and cryptic diversity. J. North Am. Benthol. Soc. 28, 517–536. https://doi.org/10.1899/08-100.1

Petersen, I., Masters, Z., Hildrew, A.G., Ormerod, S.J., 2004. Dispersal of adult aquatic insects in catchments of differing land use. J. Appl. Ecol. 41, 934–950. https://doi.org/10.1111/j.0021-8901.2004.00942.x

Previšić, A., Schnitzler, J., Kučinić, M., Graf, W., Ibrahimi, H., Kerovec, M., U. Pauls, S., 2014. Microscale vicariance and diversification of Western Balkan caddisflies linked to karstification. Freshwater Science 33, 250–262. https://doi.org/10.1086/674430

Price, T.D., Mohan, D., Tietze, D.T., Hooper, D.M., Orme, C.D.L., Rasmussen, P.C., 2011. Determinants of northerly range limits along the Himalayan bird diversity gradient. Am. Nat. 178, S97–S108. https://doi.org/10.1086/661926

Rahbek, C., Borregaard, M.K., Antonelli, A., Colwell, R.K., Holt, B.G., Nogues-Bravo, D., Rasmussen, C.M.Ø., Richardson, K., Rosing, M.T., Whittaker, R.J., Fjeldså, J., 2019a. Building mountain biodiversity: geological and evolutionary processes. Science 365, 1114–1119. https://doi.org/10.1126/science.aax0151

Rahbek, C., Borregaard, M.K., Colwell, R.K., Dalsgaard, B., Holt, B.G., Morueta-Holme, N., Nogues-Bravo, D., Whittaker, R.J., Fjeldså, J., 2019b. Humboldt's enigma: what causes global patterns of mountain biodiversity? Science 365, 1108–1113. https://doi.org/10.1126/science.aax0149

Rana, S.K., Luo, D., Rana, H.K., O'Neill, A.R., Sun, H., 2021. Geoclimatic factors influence the population genetic connectivity of *Incarvillea arguta* (Bignoniaceae) in the Himalaya–

Hengduan Mountains biodiversity hotspot. J. Syst. Evol. 59, 151–168. https://doi.org/10.1111/jse.12521

Rana, S.K., Price, T.D., Qian, H., 2019. Plant species richness across the Himalaya driven by evolutionary history and current climate. Ecosphere 10. https://doi.org/10.1002/ecs2.2945

Riek, E., 1977. The marine caddisfly family Chathamiidae (Trichoptera). Aust. J. Entomol. 15, 405–419.

Ross, H.H., 1963. Evolution and classification of the mountain caddisflies. Miscel· Lània Zoològica 94–114.

Ross, H.H., 1941. Descriptions and records of North American Trichoptera. Trans. Am. Entomol. Soc. 1890- 67, 35–126.

Sabin, T.P., Krishnan, R., Vellore, R., Priya, P., Borgaonkar, H.P., Singh, B.B., Sagar, A., 2020. Climate change over the Himalayas, in: Krishnan, R., Sanjay, J., Gnanaseelan, C., Mujumdar, M., Kulkarni, A., Chakraborty, S. (Eds.), Assessment of climate change over the Indian Region. Springer Singapore, Singapore, pp. 207–222. https://doi.org/10.1007/978-981-15-4327-2_11

Schadt, E.E., Turner, S., Kasarskis, A., 2010. A window into third-generation sequencing. Hum. Mol. Genet. 19, R227–R240. https://doi.org/10.1093/hmg/ddq416

Schmid, F., Botosaneanu, L., 1966. Le genre *Himalopsyche* Banks (Trichoptera: Rhyacophilidae). Ann Ent Soc Quebec 11, 123–176 (in French).

Shafer, A.B.A., Cullingham, C.I., Côté, S.D., Coltman, D.W., 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Mol. Ecol. 19, 4589–4621. https://doi.org/10.1111/j.1365-294X.2010.04828.x

Shah, R.D.T., Sharma, S., Haase, P., Jähnig, S.C., Pauls, S.U., 2015. The climate sensitive zone along an altitudinal gradient in central Himalayan rivers: a useful concept to monitor climate change impacts in mountain regions. Clim. Change 132, 265–278. https://doi.org/10.1007/s10584-015-1417-z

Shendure, J., Aiden, E.L., 2012. The expanding scope of DNA sequencing. Nat. Biotechnol. 30, 1084–1094. https://doi.org/10.1038/nbt.2421

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., Waterston, R.H., 2017. DNA sequencing at 40: past, present and future. Nature 550, 345–353. https://doi.org/10.1038/nature24286

Shepard, D.B., Burbrink, F.T., 2008. Lineage diversification and historical demography of a sky island salamander, *Plethodon ouachitae*, from the Interior Highlands. Mol. Ecol. 17, 5315–5335. https://doi.org/10.1111/j.1365-294X.2008.03998.x

Smith, P.J., Smith, B.J., 2009. Small-scale population-genetic differentiation in the New Zealand caddisfly *Orthopsyche fimbriata* and the crayfish *Paranephrops planifrons*. N. Z. J. Mar. Freshw. Res. 43, 723–734. https://doi.org/10.1080/00288330909510037

Sproul, J.S., Houston, Derek.D., Davis, N., Barrington, E., Oh, S.Y., Evans, R.P., Shiozawa, D.K., 2014. Comparative phylogeography of codistributed aquatic insects in western North America: insights into dispersal and regional patterns of genetic structure. Freshw. Biol. 59, 2051–2063. https://doi.org/10.1111/fwb.12406

Srinivasan, U., Tamma, K., Ramakrishnan, U., 2014. Past climate and species ecology drive nested species richness patterns along an east-west axis in the Himalaya: nestedness in Himalayan fauna. Glob. Ecol. Biogeogr. 23, 52–60. https://doi.org/10.1111/geb.12082

Su, T., Spicer, R.A., Li, S.-H., Xu, H., Huang, J., Sherlock, S., Huang, Y.-J., Li, S.-F., Wang, L., Jia, L.-B., Deng, W.-Y.-D., Liu, J., Deng, C.-L., Zhang, S.-T., Valdes, P.J., Zhou, Z.-K., 2019. Uplift, climate and biotic changes at the Eocene–Oligocene transition in south-eastern Tibet. Natl. Sci. Rev. 6, 495–504. https://doi.org/10.1093/nsr/nwy062

Sun, B.-N., Wu, J.-Y., Liu, Y.-S. (Christopher), Ding, S.-T., Li, X.-C., Xie, S.-P., Yan, D.-F., Lin, Z.-C., 2011. Reconstructing Neogene vegetation and climates to infer tectonic uplift in western Yunnan, China. Palaeogeogr. Palaeoclimatol. Palaeoecol. 304, 328–336. https://doi.org/10.1016/j.palaeo.2010.09.023

Svensson, B.W., 1974. Population movements of adult Trichoptera at a South Swedish stream. Oikos 25, 157. https://doi.org/10.2307/3543638

Tada, R., Zheng, H., Clift, P.D., 2016. Evolution and variability of the Asian monsoon and its potential linkage with uplift of the Himalaya and Tibetan Plateau. Prog. Earth Planet. Sci. 3, 4. https://doi.org/10.1186/s40645-016-0080-y

Tolley, K.A., Burger, M., Turner, A.A., Matthee, C.A., 2006. Biogeographic patterns and phylogeography of dwarf chameleons (*Bradypodion*) in an African biodiversity hotspot. Mol. Ecol. 15, 781–793. https://doi.org/10.1111/j.1365-294X.2006.02836.x

Tonkin, J.D., Altermatt, F., Finn, D.S., Heino, J., Olden, J.D., Pauls, S.U., Lytle, David.A., 2018. The role of dispersal in river network metacommunities: patterns, processes, and pathways. Freshw. Biol. 63, 141–163. https://doi.org/10.1111/fwb.13037

Tsuruishi, T., 2006. Life cycle of *Himalopsyche japonica* (Morton) (Trichoptera: Rhyacophilidae) in two high mountain streams in Nagano, Central Japan. Hydrobiologia 563, 493–499. https://doi.org/10.1007/s10750-006-0197-x

Ulmer, G., 1932. Aquatic insects of China. Article III. Neue Chinesishe Trichopteren, nebst ubersicht uber die bischer aus China bekannten Arten. Peking Nat. Hist. Bull. 7, 39–70.

Van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., Thermes, C., 2018. The third revolution in sequencing technology. Trends Genet. 34, 666–681. https://doi.org/10.1016/j.tig.2018.05.008

Von Humboldt, A., 1860. Cosmos: a sketch of a physical description of the universe. Harper & brothers.

Wang, K., Zhou, X.-H., Liu, D., Li, Y., Yao, Z., He, W.-M., Liu, Y., 2022. The uplift of the Hengduan Mountains contributed to the speciation of three *Rhododendron* species. Glob. Ecol. Conserv. 35, e02085. https://doi.org/10.1016/j.gecco.2022.e02085

Warshall, P., 1995. The Madrean sky island archipelago: a planetary overview. US Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station.

Wiggins, G.B., 2004. Caddisflies: the underwater architects. University of Toronto Press, Toronto, Buffalo.

Wilcock, H.R., Bruford, M.W., Nichols, R.A., Hildrew, A.G., 2007. Landscape, habitat characteristics and the genetic population structure of two caddisflies. Freshw. Biol. 52, 1907–1929. https://doi.org/10.1111/j.1365-2427.2007.01818.x

Winterbourn, M.J., Chadderton, W.L., Entrekin, S.A., Tank, J.L., Harding, J.S., 2007. Distribution and dispersal of adult stream insects in a heterogeneous montane environment. Fundam. Appl. Limnol. 168, 127–135. https://doi.org/10.1127/1863-9135/2007/0168-0127

Wright, B., Farquharson, K.A., McLennan, E.A., Belov, K., Hogg, C.J., Grueber, C.E., 2019. From reference genomes to population genomics: comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. BMC Genomics 20, 453. https://doi.org/10.1186/s12864-019-5806-y

Xu, W., Dong, W.-J., Fu, T.-T., Gao, W., Lu, C.-Q., Yan, F., Wu, Y.-H., Jiang, K., Jin, J.-Q., Chen, H.-M., Zhang, Y.-P., Hillis, D.M., Che, J., 2021. Herpetological phylogeographic analyses support a Miocene focal point of Himalayan uplift and biological diversification. Natl. Sci. Rev. 8, nwaa263. https://doi.org/10.1093/nsr/nwaa263

Yan, Y., Yang, X., Tang, Z., 2013. Patterns of species diversity and phylogenetic structure of vascular plants on the Qinghai-Tibetan Plateau. Ecol. Evol. 3, 4584–4595. https://doi.org/10.1002/ece3.847

Yang, Y.-M., Lee, J.-Y., Wang, B., 2019. The Tibetan Plateau uplift is crucial for eastward propagation of Madden-Julian oscillation. Sci. Rep. 9, 15478. https://doi.org/10.1038/s41598-019-51461-w

Zhang, D., Hao, G., Guo, X., Hu, Q., Liu, J., 2019. Genomic insight into “sky island” species diversification in a mountainous biodiversity hotspot. J. Syst. Evol. 57, 633–645. https://doi.org/10.1111/jse.12543

5. Portfolio of publications

Chapter 1

Gene flow and diversification in *Himalopsyche martynovi* species complex (Trichoptera: Rhyacophilidae) in the Hengduan Mountains

Xi-Ling Deng, Adrien Favre, Emily Moriarty Lemmon, Alan R. Lemmon, Steffen U. Pauls

This article was published by MDPI in Biology on 23 August 2021. available online:

https://doi.org/10.3390/biology10080816

SUP designed the study and acquired funding for the study. EML and ARL performed laboratory work and formal analyzes. I conducted bioinformatic data analysis. I interpret results together with AF and SUP. I wrote the manuscript with inputs from AF and SUP.

biology

MDPI

Article

Gene Flow and Diversification in *Himalopsyche martynovi* Species Complex (Trichoptera: Rhyacophilidae) in the Hengduan Mountains

Xi-Ling Deng [1,2,3,*], Adrien Favre [1], Emily Moriarty Lemmon [4,5], Alan R. Lemmon [4,5] and Steffen U. Pauls [1,2,3,*]

[1] Entomology III, Senckenberg Research Institute and Natural History Museum, Senckenberganlage 25, 60325 Frankfurt am Main, Germany; adrien.favre@senckenberg.de
[2] Institute of Insect Biotechnology, Justus-Liebig-University Gießen, Heinrich-Buff-Ring 26, 35392 Gießen, Germany
[3] LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[4] Department of Scientific Computing, 400 Dirac Science Library, Florida State University, Tallahassee, FL 32306-4102, USA; chorusfrog@bio.fsu.edu (E.M.L.); alemmon@fsu.edu (A.R.L.)
[5] Department of Biological Science, Florida State University, 89 Chieftan Way, Tallahassee, FL 32306-4295, USA
* Correspondence: xiling.deng@senckenberg.de (X.-L.D.); steffen.pauls@senckenberg.de (S.U.P.)

Simple Summary: *Himalopsyche* is a group of aquatic insects endemic to the Hengduan Mountains, of which species are usually easily identifiable based on male genitalia, except for a few morphologically variable groups including the *Himalopsyche martynovi* complex. In order to clarify species boundaries within this complex, we investigated its evolutionary history (phylogenetics and gene flow analyses) using a large genomic dataset (~500,000 sites). We found three clades in the *Himalopsyche martynovi* complex, one of which being very variable morphologically while being involved in gene flow with other related lineages. When interpreted in the light of past geological, climatic and palaeohydrologi- cal changes, our study suggests that biological novelty—here, trait variation and recombination—may have been acquired via hybridization and represent a source of mountain biodiversity.

Abstract: The Hengduan Mountains are one of the most species-rich mountainous areas in the world. The origin and evolution of such a remarkable biodiversity are likely to be associated with geological or climatic dynamics, as well as taxon-specific biotic processes (e.g., hybridization, polyploidization, etc.). Here, we investigate the mechanisms fostering the diversification of the endemic *Himalopsyche martynovi* complex, a poorly known group of aquatic insects. We used multiple allelic datasets generated from 691 AHE loci to reconstruct species and RaxML phylogenetic trees.We selected the most reliable phylogenetic tree to perform network and gene flow analyses. The phylogenetic reconstructions and network analysis identified three clades, including *H. epikur*, *H. martynovi* sensu stricto and *H.* cf. *martynovi*. *Himalopsyche martynovi* sensu stricto and *H.* cf. *martynovi* present an intermediate morphology between *H. epikur* and *H. viteceki,* the closest known relative to the *H. martynovi*-complex. The gene flow analysis revealed extensive gene flow among these lineages. Our results suggest that *H. viteceki* and *H. epikur* are likely to have contributed to the evolution of *H. martynovi* sensu stricto and *H.* cf. *martynovi* via gene flow, and thus, our study provides insights in the diversification process of a lesser-known ecological group, and hints at the potential role of gene flow in the emergence of biological novelty in the Hengduan Mountains.

Keywords: Hengduan Mountains; *Himalopsyche martynovi*; gene flow; morphology; phylogeny; speciation; target enrichment

check for updates

Citation: Deng, X.-L.; Favre, A.; Lemmon, E.M.; Lemmon, A.R.; Pauls, S.U. Gene Flow and Diversification in *Himalopsyche martynovi* Species Complex (Trichoptera: Rhyacophilidae) in the Hengduan Mountains. *Biology* **2021**, *10*, 816. https://doi.org/10.3390/biology10080816

Academic Editor: Jukka Finne

Received: 14 July 2021
Accepted: 17 August 2021
Published: 23 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The distribution of species diversity is globally uneven [1–3], and areas with an exceptional concentration of species are often qualified as 'biodiversity hotspots' depending

on their current degradation and need for conservation [4,5]. One of the most outstanding mountainous hotspots of diversity is located in the Hengduan Mountains (Hengduan Mts), Southwest China, a region flanking the Qinghai–Tibetan Plateau to the East [5,6]. There, the process of cataloguing life is still ongoing, and a considerable number of species are probably still unknown. Two of the challenges encountered by taxonomists in describing this diversity are detecting cryptic species [7,8], and identifying species boundaries in the face of gene flow within species complexes [9]. To address these issues, genomic data have become increasingly popular because they allow detecting intra- and interspecific gene flow with more accuracy, and thus shed light on when and how boundaries formed among diverging populations and species.

The origin and evolution of biodiversity in the region of the Qinghai–Tibet Plateau (QTP) (i.e., incl. the Hengduan Mts) is an intricate process, involving multiple local and global changes, resulting in a remarkable accumulation of species. In the literature, the diversification of most taxa is attributed to abiotic changes, either climatic or geologi- cal [10], and indeed such environmental modifications could act as dynamic drivers for the speciation process, for instance by modifying geographic barriers. As suggested in the "mountain-geobiodiversity hypothesis" [11], patterns of increased diversification in mountain areas (and specifically in the region of the QTP) are associated with three boundary conditions, including a full elevational zonation (i), a high ruggedness of the terrain providing environmental gradients (ii), as well as climate oscillations which facilitate mountain systems to act as "species-pumps", i.e., they serve as the background for repeated pulses of elevational migration leading to fragmentation and ultimately speciation followed by species expansions over large time scales [11] (iii). Conditions (i) and (ii) were probably realized early on in the orogeny of the Hengduan Mts, as alpine vegetation evolved during the Oligocene in this region [12]. Many radiations, however strongly overlap temporally with climate oscillations (iii) of the last few million years, at least in plants [10]. During these more recent times, cyclical climatic modifications may have fostered alternate phases of species' range fragmentation (allopatric differentiation) and secondary contact, upon which gene flow among species or populations with incomplete reproductive isolation may have occurred. Thus, the hypothesis provides a good overview on how dynamic abiotic conditions affect the evolution of mountain biota. In the Hengduan Mts, climate changes and orogenic movements profoundly changed the relative frequency and the distribution of available habitats, affecting species movement and distributions and ultimately speciation and patterns of diversity [13]. Species complexes are common in the Hengduan Mts (e.g., plant [14], mammal [15], fungus [16], caddisflies [17]) and represent good models for studying evolutionary processes and mechanisms of speciation in the context of the mountain-geobiodiversity hypothesis.

Himalopsyche is a common taxon in the region of the QTP with strong affinity to high elevation streams and rivers. The genus has its center of diversity in the Hengduan Mts [18]. Currently, the genus encompasses 56 species, mostly described on the basis of stable morphological features of adult male genitalia that display a diverse morphology but very little intraspecific variation. Thus, the morphology of male genitalia is usually sufficient for reliable taxonomic identification [19]. However, for a few species, a larger intraspecific variability of this trait renders species delineation challenging. To cope with these cases, several species complexes were defined, including the *triloba*-complex, the *japonica*-complex, the *excise*-complex and the *martynovi*-complex [17,20–22]. The *martynovi*-complex is endemic to the Hengduan Mts and was originally described as the *H. alticola-martynovi*-complex by Ross on 1956 [21], and later split as *H. alticola* and *H. martynovi* by Schmid and Botosaneanu [22]. Then, Malicky [23] assigned *H. epikur* to the *martynovi*-complex as a new species based on morphological evidence. Recently, Hjalmarsson [24] attempted to delineate the species of this complex more precisely, using both morphological and genetic approaches. Using five genetic markers, Hjalmarsson [24] showed that *H. epikur* forms a monophyletic group with stable morphological characters and concluded that *H. epikur* is genetically, morphologically and ecologically distinct. However, in that

study, *H. martynovi* (as sister clade to *H. epikur*) contained a perplexing morphological variation with some diagnostic traits showing a gradient of intermediate forms between the typical morphologies of the two species. The intraspecific morphological variation within the *H. martynovi* clade is a very unusual case in *Himalopsyche*. Given climate cycles and associated range expansion-regression dynamics of a "species-pump" in the last few million years, some extent of hybridization upon secondary contact may be envisaged as an origin of such intraspecific morphological variation. In fact, Malicky and Pauls [25] found that hybridization between two closely related species of Trichoptera lead to explicitly intermediate morphologies consistent with co-dominant inheritance patterns. Therefore, detecting intraspecific and interspecific gene exchange among species of the *H. martynovi* complex may shed light on how morphological variation arose, whether or not gene flow played a role in its evolution, and help taxonomic delineation among diverging populations and species.

In this study, we thus investigate the different clades of the *H. martynovi*-complex, using phylogenetic reconstructions based upon molecular data from 691 loci captured by anchored hybrid enrichment (AHE; [26]). We aim to compare the resulting phylogenies with the morphological features of male genitalia of adult specimens. In order to improve species delineation, we test (i) whether AHE loci provide enough resolution, (ii) whether gene flow occurred among the different clades and (iii) whether the direction of gene flow may explain the morphological variation observed in the *H. martynovi*-complex.

2. Materials and Methods

2.1. Taxon Sampling

We included 11 specimens representing five closely related species. In this case, 10 of these samples were adult males and one was a larva. All specimens were determined using Hjalmarsson [24]. *Himalopsyche immodesta* was chosen as outgroup, *H. viteceki*, *H. martynovi* and *H. epikur* were regarded as ingroup species. Specimens were collected inthe wild between 2010 to 2014, preserved in >75% ethanol and deposited in the collections of the Senckenberg Research Institute and Natural History Museum. The genitalia of all the adults were cleared in either KOH or lactic acid, then photographed using a DP2 digital camera mounted on an Olympus SZX7 steromicroscope. All specimen and voucher information are shown in File S1 (Table S1) and Figure 1. *H. epikur* is distributed in southern region of the Hengduan Shan, *H. martynovi* further north-east. The distribution area of *H. viteceki* and the outgroup *H. immodesta* overlap with that of *H. epikur*.

2.2. DNA Extraction, Library Preparation, Enrichment and Sequencing

Prior to DNA extraction, wings (in adults), heads and terminal abdominal segments were removed as specimen vouchers. The thoracic or abdominal muscle were then used for the genomic DNA extraction. DNA was extracted with hot sodium hydroxide as described in Truett et al. [27]. Extracted DNA was visualized on an agarose gel to estimate the mean size of the fragments. Concentration of double stranded DNA was fluorometrically estimated using a Qubit assay kit (Thermo Fisher Scientific, Waltham, MA, USA). The DNA pellet was washed 2×with 70% EtOH. Following the wash, the EtOH was discarded and the pellet dried. The pellet was sent dried to the Center for Anchored Phylogenomics at Florida State University (http://www.anchoredphylogeny.com, accessed on 22 April 2016) for quality check and library preparation. The library preparation was conducted by 4 steps: (1) sonicated extracted DNA to a size range of approximately 200–500 bp using a Covaris Ultrasonicator (Covaris, Woburn, MA, USA). (2) ligated common Illumina adapters. (3) indexed with 8 bp (single) indexing adapters. (4) pooled the libraries in equal concentration. In addition, the libraries were enriched using probes designed for Trichoptera (as described in the next section). Afterwards samples were sequenced on one lane of Illumina Hiseq sequencer with a paired-end 150 bp protocol following Lemmon et al. [26] and Prum et al. [28]. The Florida State University College of Medicine Translational lab performed the sequencing.

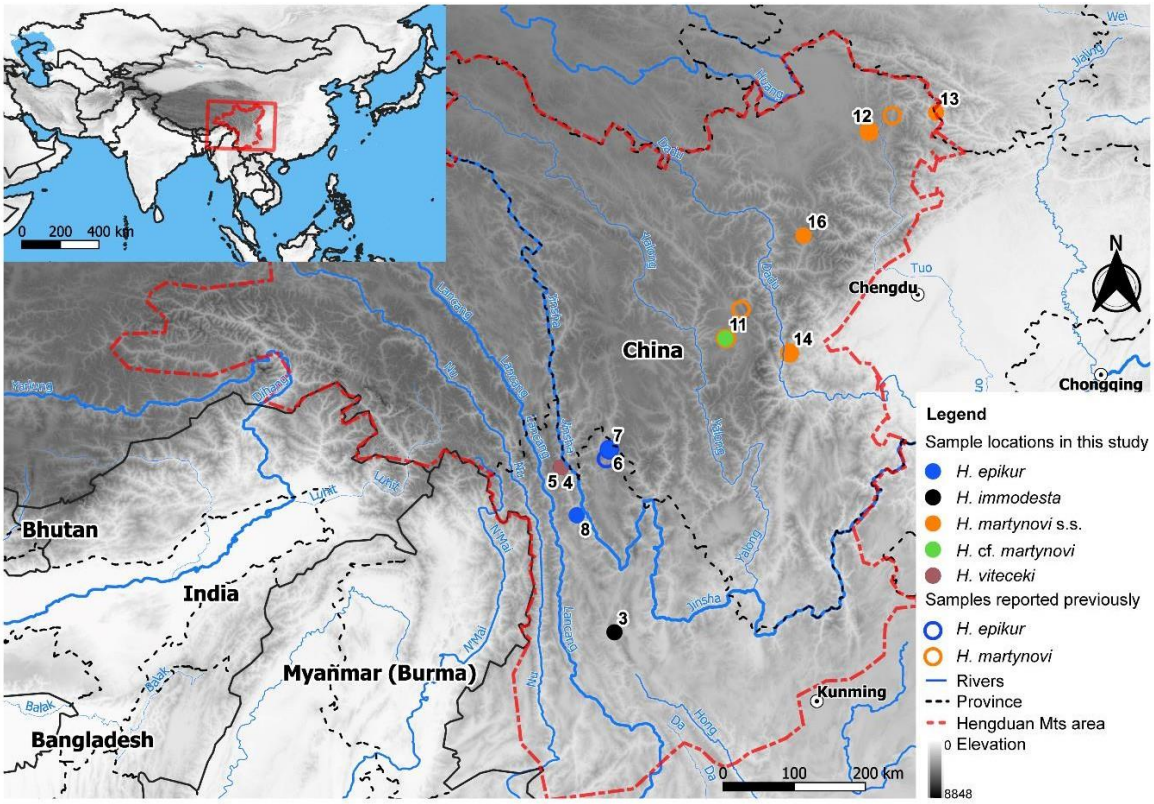


Figure 1. Sample locations of *Himalopsyche* species used in this study (solid dots) and compiled from the literature (hollow dots). *H. martynovi* s.s. is the abbreviation of *H. martynovi* sensu stricto.

2.3. Probe Design

The probes were designed in collaboration with the Center for Anchored Phylogenomics. Target AHE loci in common among Trichoptera and other Holometabola were identified by scanning 15 trichopteran 1-Kite transcriptomes (see Table S2 in File S1 for details) for the Lepodoptera AHE loci identified by Breinholt et al. [29]. The transcripts identified were aligned in MAFFT v7.023b [30] by target locus, then trimmed to well aligned regions and finally manually inspection in Geneious R9 [31]. A total of 989 target loci (averaging 232 bp in length) remained after masking/removing regions identified to be repetitive using kmer distribution profiling (see [32] for details). For each of these loci, probes were tiled uniformly across each of the 16 reference sequences at with 4×2 coverage, to produce 57,094 probes. Probes were produced by Agilent as a SureSelect XP kit. The final alignments used for probe tiling, as well as the probe design itself is given as supplemental material.

2.4. Raw Data Processing and Assembly

Raw read pairs passing the CASAVA high-chastity filter were merged following Rokyta et al. [33] and library adapters were trimmed. Reads were then assembled to the reference *Rhyacophila fasciata* (probe region sequences) using the quasi-de novo assembly procedure described by Hamilton et al. [32]. Consensus sequences were derived from assembly clusters containing 10 or more reads and ambiguities were called when site patterns could not be explained by a 1% sequencing error. Orthology among consensus sequences was determined using a neighbor-joining approach based on a pairwise distance matrix

(see Hamilton et al. 2017 for details). Alleles were then phased following Pyron et al. [34] to produce two sequences per individual.

2.5. Alignment and Trimming

Each locus containing two allelic sequences for each individual was aligned using MAFFT v. 7.023b [30] separately. To generate a high-quality dataset, we then performed data filtering for each locus as follows: (1) we removed the columns which representedless than 75% of individuals on two ends of each alignment with trimAl [35]. (2) Weidentified the randomness section in each alignment with Aliscore and then removed itwith Alicut [36]. (3) We discarded loci with a length <400 bp. (4) Finally, we also discarded alignments that contained less than 16 individuals. The remaining 691 loci were thenanalyzed both in concatenation and individually (File S2, File S3).

2.6. Phylogenetic Analyses

Prior to phylogenetic analyses, we generated four datasets based on the allele alignments after data filtering in order to estimate the phylogenetic accuracy at allelic level. Indeed, the multiple sequence alignment based on alleles is known to be the optimal data type for phylogenetic analyses because it contains additional heterozygous information and represents the smallest units in evolution [37]. These four datasets were: (1) multiple sequence alignments (MSAs) of each locus and a concatenated alignment containing only allele 1 (allele 1 alignment); (2) MSAs of each locus and a concatenated alignment contain- ing only allele 2 (allele 2 alignment); (3) MSAs of each locus and a concatenated alignment containing both alleles (bi-allelic alignment); (4) MSAs of each locus and a concatenated alignment containing merged alleles based on the International Union of Pure and Applied Chemistry (IUPAC) ambiguity codes (merged-allele alignment). Then, we conducted the phylogenetic analyses on these four datasets separately.

We inferred the maximum-likelihood (ML) tree with the concatenated loci in RAxML v8.2.X [38]. The ML tree was generated with 1000 bootstrap replicates and partitioned by each locus. We conducted a model test with the concatenated sequence using jModelTest 2.1.4 [39,40], which identified the GTR+I+G model as the most appropriate (File S1: Table S3). However, according to the opinion of Stamatakis [38], GTRGAMMA is to be preferred over GTR+I+G for small datasets. Thus, we selected GTRGAMMA model instead of GTR+I+G in this case. We used the same approach to construct the ML tree for each single locus but using 100 bootstrap replicates instead of 1000. Since the single locus only contained 2–7% informative sites, these single gene trees showed a highly incongruent phylogenetic pattern. Finally, we generated a summary-based species tree based on the 691 unrooted gene trees using ASTRAL-III [41].

To better visualize the genealogical concordance and discordance, as well as getting an initial assessment of potential gene flow, we conducted a network analysis using SplitsTree4 ver.4.16.2 [42] of 571 gene trees which contains all individuals selected from the 691 trees. We used ConsensusNetwork as the tree transformation and three different thresholds (3%, 5% and 20%) in SplitsTree v4.

2.7. Gene Flow Analysis Using ExDFOIL

Usually, hybridization is defined as the interbreeding between two species or two genetically distinct lineages [43], and introgression as gene transfer between species or differentiated population by backcrossing, potentially leading to a permanent incorporation of some genes (and traits) of one species into the other [44]. In parallel, gene flow is a collective term that includes all mechanisms resulting in the movement of genes from one population to another [45,46]. In our case, as we deal with a species complex, we will use the terms “introgression” for gene exchange between two well-supported species, and “gene flow” within the species complex itself.

A SNP dataset was called from two allele alignments using SNP-sites [47] (https://github.com/sanger-pathogens/snp-sites, accessed on 12 December 2020) and then used

to detect potential introgression among species. To better grasp the speciation mechanisms, we employed ExD$_{FOIL}$ which uses allelic patterns to detect interspecific introgression [48,49] (https://github.com/SheaML/ExDFOIL, accessed on 10 November 2020). ExD$_{FOIL}$ test is an extension of the D$_{FOIL}$ analysis, which we applied to a symmetric five-taxon phylogeny as (((P1, P2), (P3, P4)), O) [48]. Typically, the program is able to detect introgression between lineages, and suggest its direction. It can also estimate whether introgression is director indirect, current or ancestral. We used *H. immodesta* as outgroup and the species tree calculated by bi-allelic alignment as the tested topology because of its high reliability. Each allele was regarded as one individual in the approach. In total we tested 1755 suitable combinations for reconstructing introgression. Afterwards the predicted introgression between individuals was summarized as introgression between lineages.

3. Results

3.1. AHE Target Characteristics

The target locus set contained 989 loci averaging 232 bp per locus (range: 150–1721 bp). Note however that 37 of the loci were missing a *Rhyacophila* sequence as outgroup. Consequently, the maximum number of loci that could be obtained in this study is 952. The total target size was 367,184 sites. Pairwise sequence identity averaged 73.39% (range: 87.2–50.3%). The taxon coverage was very good, with 97% of the loci containing sequences for at least 10 of the 11 species. As a result, the alignments were very complete; only 3.9%of the alignment characters were missing or ambiguous.

3.2. Read and Assembly Characteristics

The sequencing effort produced 95 million raw read pairs, averaging 1.6 Gb per sample (File S4). Reads mapped to the reference sequence with a 24.6 on-target percentage. On average, 738 loci (with consensus sequences at least 250 bp in length) were obtained per sample. Consensus sequences, constructed from an average of 3446 reads per locus, averaged 630 bp in length. Due to the high coverage per locus, the PCR duplicate rate was moderate (approximately 50%).

3.3. Sequence Characteristics after Filtering

After filtering, the alignments contained a total of 509,113 sites including 691 loci, of which 65.8% were identical sites and 4% were missing characters (File S1: Figure S1, File S2, File S3). Individual alignments of each locus ranged from 401 bp to 2293 bp in length, the mean length being 737 bp. Each of these loci were recovered for more than eight specimens included in this study and were used for the subsequent phylogenetic analyses.

3.4. Phylogeny

We reconstructed a species tree based on 691 loci, and a RaxML tree based on concatenated loci for four different alignments (allele 1 alignment, allele 2 alignment, bi-allelic alignment and merged-allele alignment). All analyses recovered almost identical topologies regardless of the datasets used. Most nodes were highly supported (BS > 95), especially those forming the backbone of the phylogeny, whereas some uncertainty occurred between species tree and RaxML tree for some datasets. The species relationship of *Himalopsyche* using two approaches based on four allelic procedures were highly consistent, and all phylogenetic trees showed four monophyletic groups: *H. viteceki*, main clade of *H. martynovi*, one individual lineage of *H. martynovi* and *H. epikur*. For convenience, we will refer to the main clades as *H. martynovi* sensu stricto (*H. martynovi* s.s.) and the single individual lineage as *H.* cf. *martynovi*, respectively. *Himalopsyche* cf. *martynovi* was identified as *H. martynovi* based on morphological and molecular evidence in earlier studies [24], but clustered as a sister clade to *H. epikur* in our analyses. Finally, *H. viteceki* was sister to the clade containing *H. martynovi* s.s., *H.* cf. *martynovi* and *H. epikur*.

The coalescent species trees and concatenated RaxML trees were consistent for most of the clades, except for a few shallow nodes in *H. epikur* and *H. martynovi* s.s. when

calculated using the allele 1 alignment and the bi-allelic alignment (Figure 2). Notably, the topological structure of *H. martynovi* s.s. were incongruent comparing the species tree with the ML tree using the three alignments which contained only one allele (allele 1 alignment, allele 2 alignment, merged-allelic alignment) (Figure 2A–C), but congruent when using the bi-allelic alignment. In all the species trees calculated by one single allele alignment, the individuals of *H. martynovi* s.s. are clustered as four monophyletic sister lineages with some low support values. All trees showed two pairs of sister branches in all four ML trees and the one Astral tree based on the bi-allelic alignment. In addition, the phylogenies based on bi-alleles' alignments comprise the most complete data set and showed greatest consistency between species tree and RaxML tree in all the key nodes (densitree in Figure 2D). Hence, we chose the species tree calculated with the bi-allelic alignment for the subsequent gene flow analysis, as these required a stable and accurate topology.

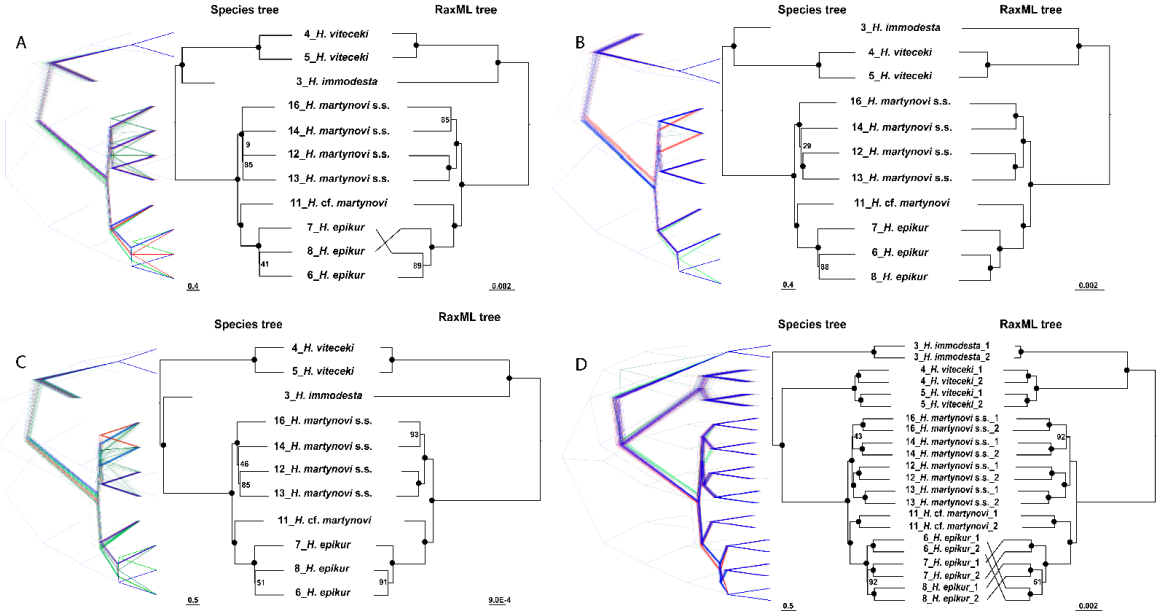


Figure 2. Species tree with 691 loci (left) and RaxML tree with the concatenated loci (right) for *Himalopsyche martynovi* complex based on four datasets used in this study: (**A**) with allele 1; (**B**) with allele 2; (**C**) with the merged allele based on IUPAC code. On each figure, the left part shows a plot of the complete bootstrap species tree distribution using DensiTree (**D**). The black circle on the branch node represents bootstrap scores above 96, otherwise is shown with numbers. *H. martynovi s.s.* is the abbreviation of *H. martynovi* sensu stricto.

The consensus network analysis showed unambiguous clusters of species (Figure 3). Samples of *H. epikur* formed a cluster with simple branch structure at all the thresholds. *Himalopsyche martynovi* s.s. clustered together and formed more internal reticulation within the clade at a lower threshold. *Himalopsyche* cf. *martynovi* clustered as sister to *H. epikur* at the thresholds of 3% and 5%, but clustered together with *Himalopsyche martynovi* s.s. at the thresholds of 20%. The outgroups and *H. viteceki* were recovered as two independent lineages without any reticulation toward any of the other species regardless of the threshold.

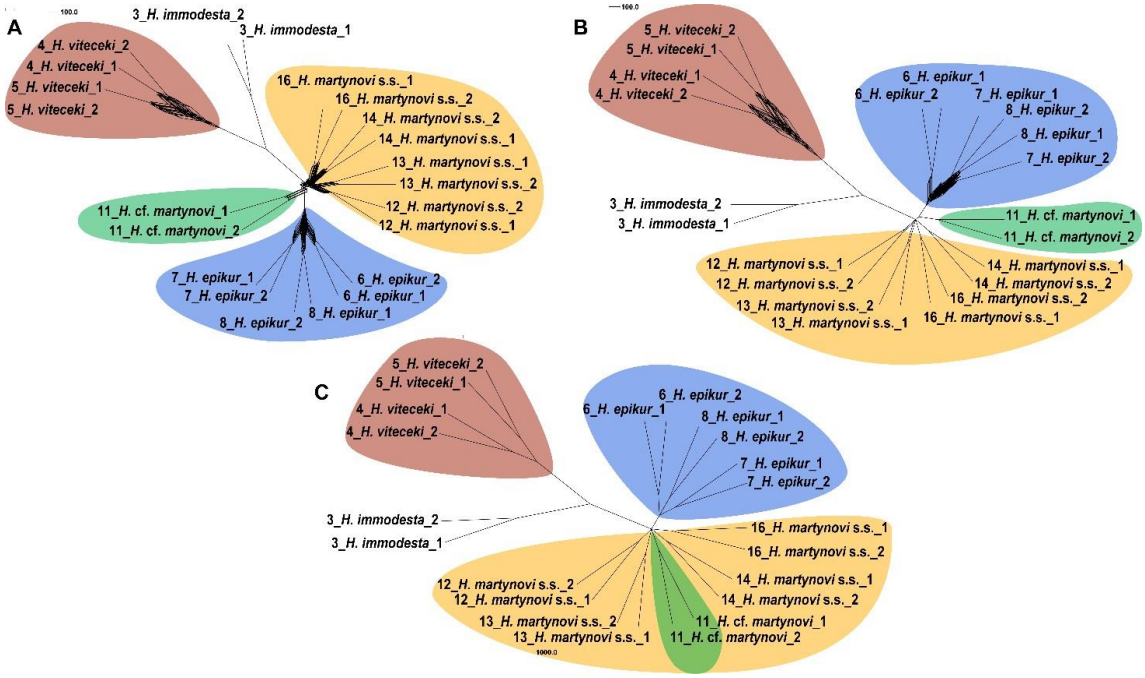


Figure 3. Network of *Himalopsyche martynovi* complex based on all the RaxML gene tree using SplitsTree. The thresholds were set to 3% (**A**), 5% (**B**) and 20% (**C**), respectively. *H. martynovi s.s.* is the abbreviation of *H. martynovi* sensu stricto.

3.5. Introgression

The ExD$_{FOIL}$ analyses revealed numerous signals of gene flow among the four taxa (excluding *H. immodesta*, the outgroup) (Figure 4, File S1: Table S4, File S5). For example, the analysis detected frequent bidirectional gene flow (counts = 10) between different individuals of *H. epikur* and one subclade of *H. martynovi* s.s.. Introgression is also likely to have occurred between *H. martynovi* s.s. and *H.* cf. *martynovi* (counts = 11). Finally, there were some unidirectional signals (counts = 2) of gene flow from one subclade of *H. martynovi* s.s. into *H.* cf. *martynovi*. In addition, there was one bidirectional gene flow signal between *H. viteceki* and *H.* cf. *martynovi*.

3.6. Morphological Sorting

Figure 4 associates the morphology of adult male's genitalia with the phylogenetic context. We found uniform morphological characteristics within *H. viteceki* and *H. epikur* based upon two and three individuals, respectively. In contrast, the morphology for *H. martynovi* s.s. and *H.* cf. *martynovi*, was more complex and variable within *H. martynovi* s.s. (based upon four individuals). However, these two lineages were not only separated by molecular evidence, but also present distinct morphologies. The main morphological differences between these two lineages were: (i) the incised curving in the middle of the superior appendages in lateral view was slightly incised in some of the individuals of *H. martynovi* s.s. but strongly concave in other individuals of *H. martynovi* s.s. and *H.* cf. *martynovi* (ii) the distal margin of the superior appendages in lateral view, had a protruding shape in *H. martynovi* s.s. and a concave shape in *H.* cf. *martynovi*. We found that individuals of the *H. martynovi*-complex often displayed intermediate morphological features in the superior appendages, thus partly mirroring the phylogenetic relationships. Specifically, *H.* cf. *martynovi* appeared as an intermediate form between *H. martynovi* s.s. and *H. epikur*, and *H.* cf. *martynovi* and *H. martynovi* s.s. had an intermediate morphology

between *H. viteceki* and *H. epikur*. In addition, the intraspecific morphological variation of *H. martynovi* s.s. was on par with interspecific morphological variation in the complex.

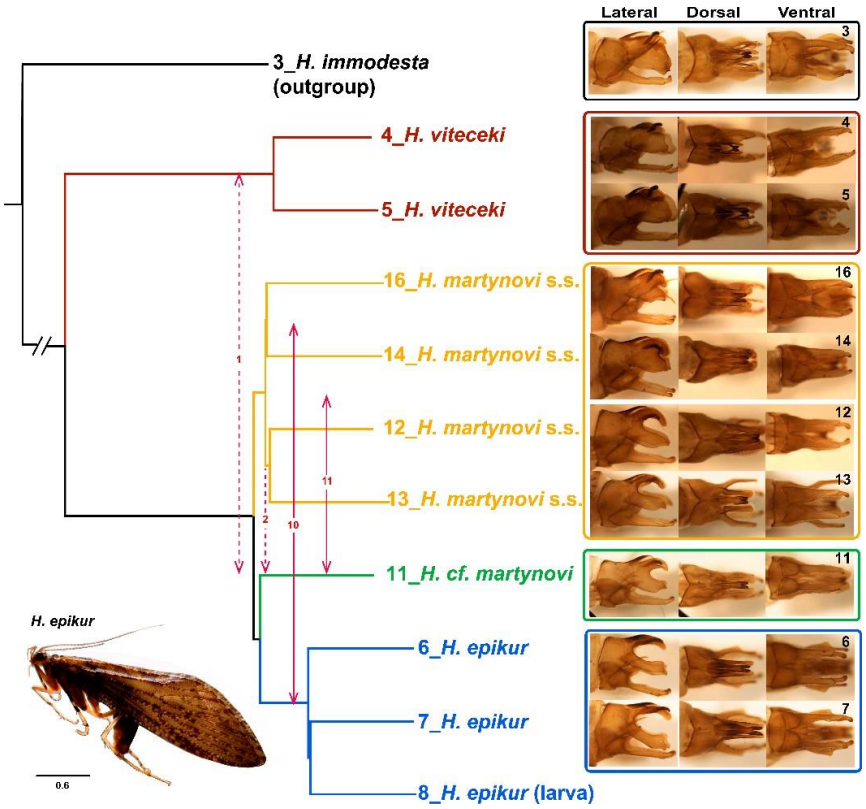


Figure 4. Gene flow and morphological variation of the genitals among the *Himalopsyche* species showing on the species tree. The species tree is calculated by ASTRAL-III based on 691 gene trees constructed with maximum likelihood. Different colors show the phylogenetic position among the same species. Gene flow are calculated by ExoDfoil. Two-way arrows depict bidirectional gene flow signal, one-way arrows depict unidirectional gene flow signal, numbers inside arrows represent the frequency of significant gene flow signals. Different thickness of the solid arrows shows different gene flow frequency, dashed red lines indicate low gene flow frequency. *H. martynovi s.s.* is the abbreviation of *H. martynovi* sensu stricto. Numbers on the images are corresponding with the sample ID.

4. Discussion

Species complexes are taxonomically challenging because they are usually characterized by a constellation of morphological traits which are not clade-specific, and often traditional molecular markers do not suffice to resolve phylogenetic relationships convincingly. In this study, we used genomic data associated with morphological investigations in order to clarify species delineation within the *H. martynovi*-complex. Endemic to the Hengduan Mts, this species complex is characterized by an unusually high intraspecific morphological variation and interspecific morphological gradients [24]. We thus investigated whether gene flow may be the source of this morphological variation. Our study is the first to use an advanced molecular methodology (AHE) to investigate a species complex in caddisflies and more generally to link the resulting pattern with gene flow as a mechanism potentially responsible for the morphological ambiguity and biological novelty.

4.1. Genomic and Morphological Evidence and Their Taxonomic Implications

In order to test the reliability of phylogenetic reconstructions, we used four different datasets based on different allele sequences, namely alignment including either allele 1 or 2, as well as a bi-allelic and a merged-allelic alignment. Our results on allele phasing are in line with the study of Andermann et al. [37] and reveal that the data set including two alleles generally recovers a more accurate estimate of tree topology in terms of less ambiguity and high consistency in the species tree approaches. This was particularly true for phylogenetic relationships among the study's focal clades. Moreover, the maintree topologies were consistent across both species trees and ML trees using concatenated alignments within each dataset. Thus, the reduced genomic datasets we generated by AHE appear to improve our understanding of ambiguous taxonomic relationships within this caddisfly species complex as previously shown for other arthropods [29,32,50–52].

As expected, we identified three robustly supported (100 BP) main clades, corresponding to *H. martynovi* s.s., *H. epikur* and *H. viteceki*, respectively. However, we uncovered a fourth lineage which we called *H.* cf. *martynovi*. This lineage appears as sister to *H. epikur* in all trees, as well as in the network analyses. *Himalopsyche* cf. *martynovi* might represent another enigmatic species called *H. alticola*, which was originally described as one of three main species of the *H. martynovi*-complex (with *H. martynovi* and *H. epikur*). However, the description of this species is rather incomplete, providing only coarse explanations and a single drawing of its morphology [20,22]. Thus, although there is a morphological resemblance between *H.* cf. *martynovi* and *H. alticola* on the basis of the rudimentary species description, the morphological information available for this latter species is too unreliable to conclusively attribute *H.* cf. *martynovi* to *H. alticola*, especially in the presence of the unusual morphological variation in the *H. martynovi*-complex. Further morphological and molecular evidence would be needed to evaluate the validity of *H. alticola*, in particular from its holotype or at least from its type locality. While the sampling of our study is limited in the number of individuals, but it includes all hitherto known morphological variation and most of the distribution range of this group. Moreover, our study mainly focuses on the origin and intraspecific variation pattern in this species complex, which can be readily assessed with our data.

From our results it is clear that there are at least three lineages in the *H. martynovi* complex even though there is unusually high morphological variation among the specimens of *H. martynovi* s.s. Interestingly, *H.* cf. *martynovi* presents an intermediate morphology between *H. martynovi* s.s. and *H. epikur* (Figure 4). For example, the incised curving in the middle of the superior appendages of *H.* cf. *martynovi* is concave to an intermediate degree compared with specimens of *H. martynovi* s.s. and *H. epikur*. Moreover, the distal margin of the superior appendages in *H.* cf. *martynovi* shows an intermediate form between the protruding shape in *H. martynovi* s.s. and the concave shape in *H. epikur*. The concordance of the molecular and morphological evidence suggests that either a stepwise evolution of traits occurred among these three lineages, or alternatively that intermediate morphologies in *H.* cf. *martynovi* were acquired via gene exchange between or with *H. martynovi* s.s. and *H. epikur*.

4.2. Gene Flow and Speciation of H. martynovi Complex

We found that species of the *H. martynovi*-complex, and particularly *H. martynovi* s.s., cannot easily be distinguished morphologically because of the high versatility of trait morphology in the male genitalia, whereas our phylogeny unequivocally recovers three robust lineages in this species complex. In parallel, we find evidence for gene flow or introgression not only within lineages (network: within *H. martynovi* s.s. and *H. epikur*), but also among lineages/species (ExD_{foil} analysis: among the three lineages and a faint introgression between *H. viteceki* and *H. martynovi* complex). This line of evidence, coupled with the intermediate morphology of *H.* cf. *martynovi* and the variable nature of morphological traits in *H. martynovi* s.s. suggests that these lineages may bear the morphological signature of gene flow. This would not be an isolated case in caddisflies, since it has already been

shown in Limnephilidae that hybrid males would present an intermediate morphology for their genitalia [25].

In fact, it is increasingly accepted that hybridization, gene flow and introgression may play a supporting role during speciation and diversification, because they foster new gene combinations and the transfer of adaptive genes [53–55]. Even though this concept is more accepted for plants, some evidence indicates it is also the case for animals [56–60]. For example, hybridizing lineages of salamanders have been shown to have significantly greater net-diversification rates than non-hybridizing lineages [61]. Our analyses showed that *H. martynovi* s.s. is likely to have been involved in introgression with other lineages, possibly making it the recipient of genes corresponding to contrasting trait values, which today still co-exist in this lineage, explaining its remarkable morphological diversity. However, *H. epikur*, which was one of the known partners involved in gene flow, does not display such morphological variation. One can only assume that this discrepancy could be due to a bidirectional but asymmetric gene flow (predominantly towards *H. martynovi* s.s.), or that selective pressures have eliminated parts of the morphological variation resulting from gene exchange in *H. epikur* and not in *H. martynovi* s.s.

Considering the current distribution ranges of species of the *H. martynovi*-complex and our knowledge on range displacement during climate oscillations for organisms of the Hengduan Mts, we hypothesize that our results match with a general species-pump scenario. For example, it is often reported that species or populations have diverged allopatrically in the northern and southern Hengduan Mts, respectively [15,62,63]. This might have been the case for *H. epikur* (southern Hengduan Mts) and *H. martynovi* s.s. (northern Hengduan Mts), since these two species occur in drainage systems that were historically distinct [64]. Indeed, the upper Yangtze River (the distribution area of *H. epikur*, Figure 1) and the Yalong-Dadu river (the distribution area of *H. martynovi* s.s.) remained separated until ca. 1.3 Ma [64], suggesting a possible minimum divergence time for these two species. After their allopatric divergence, re-organization of drainage systems coupled with climate oscillations may have fostered repeated encounters between these two species, upon which gene flow may have occurred, as is predicted by the "*mountain geo-biodiversity hypothesis*" [11]. We believe it should come as no surprise that *H.* cf. *martynovi*, phylogenetically closely related to *H. epikur* but morphologically more similar to *H. martynovi* s.s., has been found at the southwestern edge of the distribution range of the latter, closest to known populations of the former (see Figure 1). The area between the Dadu and Yalong rivers may thus have acted as a contact zone between the two species. If this were formally verified, our study would be one of the very few to document the role of gene flow as a source of biological novelty under the impulse of a species-pump effect. However, sympatric species such as *H. viteceki* and *H. epikur* do not seem to have a history of gene flow, which would counterbalance our hypothesis. We argue that not only a stricter ecological differentiation may have evolved between these two species (as is often the case in sympatry), or alternatively, that their more ancient divergence promotes more genetically-based incompatibilities today, hence protecting *H. viteceki* and *H. epikur* from hybridization. Admittedly speculative, our hypothesis is nevertheless testable, for example by extending the sampling to represent all secondary drainage systems in the area, and by dating the divergence of the lineages of the *Himalopsyche martynovi* species complex.

5. Conclusions

The phylogenetic analyses based on genomic data captured by AHE identified three lineages in the *H. martynovi*-complex, one of them unequivocally identified as *H. epikur*. The other two lineages, which we named *H. martynovi* s.s. and *H.* cf. *martynovi*, were similar to each other and morphologically despite trait variation, and intermediate between *H. viteceki* and *H. epikur*. We showed that gene flow occurred frequently between *H. martynovi* s.s. and *H.* cf. *martynovi*, as well as between *H. martynovi* s.s. and *H. epikur*. We argue that gene flow may in fact be the source of morphological variation in *H. martynovi* s.s., and intermediate morphotypes in *H.* cf. *martynovi*. Climate oscillations and re-arrangement of local drainage

systems may have fostered ancient and recent introgression upon secondary contact in this species. Finally, our study shows that in genera where diagnostic traits are considered to be stable and thus excellent for the determination of most species (as the male genitalia in caddisflies), morphological attributes may not be stable in all species. Therefore, an integrative approach such as ours should be taken into consideration for taxonomy as it avoids over-splitting morphologically versatile lineages, especially for taxa characterized by limited diagnostic features [65–68].

Supplementary Materials: Materials can be found at: https://www.mdpi.com/article/10.3390/biology10080816/s1. File S1: Table S1: List of all the samples used in this study, Table S2: Probe design resources, Table S3: Best model selected by jModelTest2 based on the concatenated sequence, Table S4: The gene flow frequency between different species summarized by ExD_{foil}, Figure S1: Length distribution of anchored hybrid enrichment loci after removing gaps and removing loci coverclusters less than 75% individuals. File S2: Alignments of each locus individually after filtering and trimming. The files are in FASTA format. File S3: The concatenated alignment of all loci. The file is in FASTA format. File S4: Assembly statistics of *Himalopsyche Martynovi* complex. File S5: The D_{foil} Statistic results on all the combinations of sequences from the bi-allelic dataset by using ExD_{foil}.

Author Contributions: Conceptualization, S.U.P.; data curation, X.-L.D., E.M.L. and A.R.L.; formal analysis, E.M.L. and A.R.L.; funding acquisition, S.U.P.; investigation, X.-L.D., A.F. and S.U.P.; methodology, X.-L.D., E.M.L., A.R.L. and S.U.P.; project administration, S.U.P.; resources, S.U.P.; supervision, A.F. and S.U.P.; validation, A.F. and S.U.P.; visualization, X.-L.D.; writing—original draft, X.-L.D.; writing—review and editing, X.-L.D., A.F., E.M.L., A.R.L. and S.U.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG), grant number PA 1617/2-1 and PA 1617/2-2 awarded to Steffen Pauls, FA 1117/1-2 awarded to Adrien Favre. The APC was funded by PA 1617/2-2.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data of Anchored hybrid enrichment in this study has been deposited at the NCBI SRA database under the Bioproject ID PRJNA744478 (accessed on 7 July 2021).

Acknowledgments: We thank Anna Hjalmarsson for providing materials used in this study, and for laying the groundwork for this study in her morphological treatments of *Himalopsyche*. We are thankful to Jürgen Otte extracting the DNA for AHE work, and to Michelle Kortyna who carried out the AHE data collection from Florida State University. We thank Axel Janke who helped us set up and interpret the gene flow analysis as well as provided valuable comments on the manuscript. We thank Anna Hjalmarsson, Ram Devi Tachamo Shah, Deep Narayan Shah, Fengqing Li and Sonja Jähnig for collecting these specimens in the field, and Qinhua Cai for obtaining the field work permit from the Institute of Hydrobiology, Chinese Academy of Sciences. Meanwhile, we appreciate the insightful comments offered by all the reviewers which help improving this study a lot. This study was financially supported by and reports results of the Deutsche Forschungsgemeinschaft (DFG) grants PA 1617/2-1 and PA 1617/2-2 awarded to Steffen Pauls, FA 1117/1-2 awarded to Adrien Favre.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meier, R.; Dikow, T. Significance of specimen databases from taxonomic revisions for estimating and mapping the global species diversity of invertebrates and repatriating reliable specimen data. *Conserv. Biol.* **2004**, *18*, 478–488. [CrossRef]
2. Kier, G.; Mutke, J.; Dinerstein, E.; Ricketts, T.H.; Küper, W.; Kreft, H.; Barthlott, W. Global patterns of plant diversity and floristic knowledge. *J. Biogeogr.* **2005**, *32*, 1107–1116. [CrossRef]
3. Bastida, F.; Eldridge, D.J.; Abades, S.; Alfaro, F.D.; Gallardo, A.; García-Velázquez, L.; García, C.; Hart, S.C.; Pérez, C.A.; Santos, F. Climatic vulnerabilities and ecological preferences of soil invertebrates across biomes. *Mol. Ecol.* **2020**, *29*, 752–761. [CrossRef] [PubMed]
4. Myers, N.; Mittermeier, R.A.; Mittermeier, C.G.; Da Fonseca, G.A.; Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **2000**, *403*, 853–858. [CrossRef] [PubMed]
5. Marchese, C. Biodiversity hotspots: A shortcut for a more complicated concept. *Glob. Ecol. Conserv.* **2015**, *3*, 297–309. [CrossRef]

6. Boufford, D.E. Biodiversity hotspot: China's Hengduan Mountains. *Arnoldia* **2014**, *72*, 24–35.
7. Li, J.; Milne, R.I.; Ru, D.; Miao, J.; Tao, W.; Zhang, L.; Xu, J.; Liu, J.; Mao, K. Allopatric divergence and hybridization within *Cupressus chengiana* (Cupressaceae), a threatened conifer in the northern Hengduan Mountains of western China. *Mol. Ecol.* **2020**, *29*, 1250–1266. [CrossRef]
8. Liu, C.; Fischer, G.; Garcia, F.H.; Yamane, S.; Liu, Q.; Peng, Y.Q.; Economo, E.P.; Guénard, B.; Pierce, N.E. Ants of the Hengduan Mountains: A new altitudinal survey and updated checklist for Yunnan Province highlight an understudied insect biodiversity hotspot. *ZooKeys* **2020**, *978*, 1. [CrossRef] [PubMed]
9. Petit, R.J.; Excoffier, L. Gene flow and species delimitation. *Trends Ecol. Evol.* **2009**, *24*, 386–393. [CrossRef]
10. Muellner-Riehl, A.N.; Schnitzler, J.; Kissling, W.D.; Mosbrugger, V.; Rijsdijk, K.F.; Seijmonsbergen, A.C.; Versteegh, H.; Favre, A. Origins of global mountain plant biodiversity: Testing the 'mountain-geobiodiversity hypothesis'. *J. Biogeogr.* **2019**, *46*, 2826–2838. [CrossRef]
11. Mosbrugger, V.; Favre, A.; Muellner-Riehl, A.N.; Päckert, M.; Mulch, A. Cenozoic evolution of geo-biodiversity in the Tibeto-Himalayan region. *Mt. Clim. Biodivers.* **2018**, *429*, 448.
12. Ding, W.-N.; Ree, R.H.; Spicer, R.A.; Xing, Y.-W. Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science* **2020**, *369*, 578–581. [CrossRef] [PubMed]
13. Clark, M.K.; House, M.; Royden, L.; Whipple, K.; Burchfiel, B.; Zhang, X.; Tang, W. Late Cenozoic uplift of southeastern Tibet. *Geology* **2005**, *33*, 525–528. [CrossRef]
14. Yu, W.-B.; Randle, C.P.; Lu, L.; Wang, H.; Yang, J.-B.; de Pamphilis, C.W.; Corlett, R.T.; Li, D.-Z. The hemiparasitic plant Phtheirospermum (Orobanchaceae) is polyphyletic and contains cryptic species in the Hengduan Mountains of southwest China. *Front. Plant. Sci.* **2018**, *9*, 142. [CrossRef] [PubMed]
15. Ge, D.; Lu, L.; Cheng, J.; Xia, L.; Chang, Y.; Wen, Z.; Lv, X.; Du, Y.; Liu, Q.; Yang, Q. An endemic rat species complex is evidence of moderate environmental changes in the terrestrial biodiversity centre of China through the late Quaternary. *Sci. Rep.* **2017**, *7*, 46127. [CrossRef] [PubMed]
16. Feng, B.; Zhao, Q.; Xu, J.; Qin, J.; Yang, Z.L. Drainage isolation and climate change-driven population expansion shape the genetic structures of *Tuber indicum* complex in the Hengduan Mountains region. *Sci. Rep.* **2016**, *6*, 21811. [CrossRef] [PubMed]
17. Hjalmarsson, A.E.; Graf, W.; Jähnig, S.C.; Vitecek, S.; Pauls, S.U. Molecular association and morphological characterisation of *Himalopsyche* larval types (Trichoptera, Rhyacophilidae). *ZooKeys* **2018**, *773*, 79. [CrossRef]
18. Hjalmarsson, A.E.; Graf, W.; Vitecek, S.; Jähnig, S.C.; Cai, Q.; Sharma, S.; Tong, X.; Li, F.; Shah, D.N.; Shah, R.D.T. Molecular phylogeny of *Himalopsyche* (Trichoptera, Rhyacophilidae). *Syst. Entomol.* **2019**, *44*, 973–984. [CrossRef]
19. Shapiro, A.M.; Porter, A.H. The lock-and-key hypothesis: Evolutionary and biosystematic interpretation of insect genitalia. *Annu. Rev. Entomol.* **1989**, *34*, 231–245. [CrossRef]
20. Banks, N. Report on Certain Groups of Neuropteroid Insects from Szechwan, China. *Proc. U. S. Natl. Mus.* **1940**. Available online: https://repository.si.edu/bitstream/handle/10088/16322/1/USNMP-88_3079_1940.pdf (accessed on 17 August 2021). [CrossRef]
21. Ross, H.H. *Evolution and Classification of the Mountain Caddisflies*; The University of Illinois Press: Urbana, IL, USA, 1956; p. 213.
22. Schmid, F. Le genre *Himalopsyche* Banks (Trichoptera: Rhyacophilidae). *Ann. Ent. Soc. Que.* **1966**, *11*, 123–176.
23. Malicky, H. Neue Trichopteren aus Europa und Asien. *Braueria* **2011**, *38*, 23–43.
24. Hjalmarsson, A.E. Delimitation and description of three new species of *Himalopsyche* (Trichoptera: Rhyacophilidae) from the Hengduan Mountains, China. *Zootaxa* **2019**, *4638*, 419–441. [CrossRef] [PubMed]
25. Malicky, H.; Pauls, S.U. Cross-breeding of *Chaetopteryx morettii* and related species, with molecular and eidonomical results (Trichoptera, Limnephilidae). *Ann. Limnol. Int. J. Limnol.* **2012**, *48*, 13–19. [CrossRef]
26. Lemmon, A.R.; Emme, S.A.; Lemmon, E.M. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* **2012**, *61*, 727–744. [CrossRef]
27. Truett, G.; Heeger, P.; Mynatt, R.; Truett, A.; Walker, J.; Warman, M. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques* **2000**, *29*, 52–54. [CrossRef]
28. Prum, R.O.; Berv, J.S.; Dornburg, A.; Field, D.J.; Townsend, J.P.; Lemmon, E.M.; Lemmon, A.R. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **2015**, *526*, 569–573. [CrossRef]
29. Breinholt, J.W.; Earl, C.; Lemmon, A.R.; Lemmon, E.M.; Xiao, L.; Kawahara, A.Y. Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for anchored phylogenomics. *Syst. Biol.* **2018**, *67*, 78–93. [CrossRef]
30. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
31. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]
32. Hamilton, C.A.; Lemmon, A.R.; Lemmon, E.M.; Bond, J.E. Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evol. Biol.* **2016**, *16*, 212. [CrossRef]
33. Rokyta, D.R.; Lemmon, A.R.; Margres, M.J.; Arnow, K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genom.* **2012**, *13*, 312. [CrossRef] [PubMed]

34. Pyron, R.A.; Hendry, C.R.; Hsieh, F.; Lemmon, A.R.; Lemmon, E.M. Integrating phylogenomic and morphological data to assess candidate species-delimitation models in Brown and Red-bellied snakes (*Storeria*). *Zool. J. Linn. Soc.* **2016**, *177*, 937–949. [CrossRef]
35. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef]
36. Misof, B.; Misof, K. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: A more objective means of data exclusion. *Syst. Biol.* **2009**, *58*, 21–34. [CrossRef]
37. Andermann, T.; Fernandes, A.M.; Olsson, U.; Töpel, M.; Pfeil, B.; Oxelman, B.; Aleixo, A.; Faircloth, B.C.; Antonelli, A. Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. *Syst. Biol.* **2019**, *68*, 32–46. [CrossRef]
38. Stamatakis, A. The RAxML v8.2.X Manual. Heidleberg Institute for Theoretical Studies. 2016. Available online: https://cme.h-its.org/exelixis/resource/download/NewManual.pdf (accessed on 10 August 2021).
39. Guindon, S.; Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **2003**, *52*, 696–704. [CrossRef]
40. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [CrossRef] [PubMed]
41. Zhang, C.; Rabiee, M.; Sayyari, E.; Mirarab, S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **2018**, *19*, 153. [CrossRef] [PubMed]
42. Huson, D.H.; Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **2006**, *23*, 254–267. [CrossRef]
43. Barton, N.H.; Hewitt, G.M. Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* **1985**, *16*, 113–148. [CrossRef]
44. Twyford, A.; Ennos, R. Next-generation hybridization and introgression. *Heredity* **2012**, *108*, 179–189. [CrossRef]
45. Slarkin, M. Gene flow in natural populations. *Annu. Rev. Ecol. Syst.* **1985**, *16*, 393–430. [CrossRef]
46. Slatkin, M. Gene flow and the geographic structure of natural populations. *Science* **1987**, *236*, 787–792. [CrossRef] [PubMed]
47. Page, A.J.; Taylor, B.; Delaney, A.J.; Soares, J.; Seemann, T.; Keane, J.A.; Harris, S.R. SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* **2016**, *2*, 4. [CrossRef]
48. Pease, J.B.; Hahn, M.W. Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Syst. Biol.* **2015**, *64*, 651–662. [CrossRef]
49. Lambert, S.M.; Streicher, J.W.; Fisher-Reid, M.C.; Mendez de la Cruz, F.R.; Martinez-Mendez, N.; Garcia-Vazquez, U.O.; Nieto Montes de Oca, A.; Wiens, J.J. Inferring introgression using RADseq and DFOIL: Power and pitfalls revealed in a casestudy of spiny lizards (*Sceloporus*). *Mol. Ecol. Resour.* **2019**, *19*, 818–837. [CrossRef]
50. Haddad, S.; Shin, S.; Lemmon, A.R.; Lemmon, E.M.; Svacha, P.; Farrell, B.; ŚLIPIŃSKI, A.; Windsor, D.; McKenna, D.D. Anchored hybrid enrichment provides new insights into the phylogeny and evolution of longhorned beetles (Cerambycidae). *Syst. Entomol.* **2018**, *43*, 68–89. [CrossRef]
51. Buenaventura, E.; Szpila, K.; Cassel, B.K.; Wiegmann, B.M.; Pape, T. Anchored hybrid enrichment challenges the traditional classification of flesh flies (Diptera: Sarcophagidae). *Syst. Entomol.* **2020**, *45*, 281–301. [CrossRef]
52. Dowdy, N.J.; Keating, S.; Lemmon, A.R.; Lemmon, E.M.; Conner, W.E.; Scott Chialvo, C.H.; Weller, S.J.; Simmons, R.B.; Sisson, M.S.; Zaspel, J.M. A deeper meaning for shallow-level phylogenomic studies: Nested anchored hybrid enrichment offersgreat promise for resolving the tiger moth tree of life (Lepidoptera: Erebidae: Arctiinae). *Syst. Entomol.* **2020**, *45*, 874–893. [CrossRef]
53. Nosil, P. Speciation with gene flow could be common. *Mol. Ecol.* **2008**, *17*, 2103–2106. [CrossRef]
54. Fitzpatrick, B.; Fordyce, J.; Gavrilets, S. Pattern, process and geographic modes of speciation. *J. Evol. Biol.* **2009**, *22*, 2342–2347. [CrossRef]
55. Fu, P.-C.; Gao, Q.-B.; Zhang, F.-Q.; Xing, R.; Wang, J.-L.; Liu, H.-R.; Chen, S.-L. Gene flow results in high genetic similarity between Sibiraea (Rosaceae) species in the Qinghai–Tibetan Plateau. *Front. Plant. Sci.* **2016**, *7*, 1596. [CrossRef]
56. Mullen, S.P.; Dopman, E.B.; Harrison, R.G. Hybrid zone origins, species boundaries, and the evolution of wing-pattern diversity in a polytypic species complex of North American admiral butterflies (Nymphalidae: *Limenitis*). *Evol. Int. J. Org. Evol.* **2008**, *62*, 1400–1417. [CrossRef]
57. Kumar, V.; Lammers, F.; Bidon, T.; Pfenninger, M.; Kolter, L.; Nilsson, M.A.; Janke, A. The evolutionary history of bears is characterized by gene flow across species. *Sci. Rep.* **2017**, *7*, 1–10. [CrossRef]
58. Árnason, Ú.; Lammers, F.; Kumar, V.; Nilsson, M.A.; Janke, A. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Sci. Adv.* **2018**, *4*, eaap9873. [CrossRef]
59. Malinsky, M.; Svardal, H.; Tyers, A.M.; Miska, E.A.; Genner, M.J.; Turner, G.F.; Durbin, R. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* **2018**, *2*, 1940–1955. [CrossRef]
60. Nilsson, M.A.; Zheng, Y.; Kumar, V.; Phillips, M.J.; Janke, A. Speciation generates mosaic genomes in kangaroos. *Genome Biol. Evol.* **2018**, *10*, 33–44. [CrossRef]
61. Patton, A.H.; Margres, M.J.; Epstein, B.; Eastman, J.; Harmon, L.J.; Storfer, A. Hybridizing salamanders experience accelerated diversification. *Sci. Rep.* **2020**, *10*, 1–12. [CrossRef]
62. Feng, B.; Liu, J.W.; Xu, J.; Zhao, K.; Ge, Z.W.; Yang, Z.L. Ecological and physical barriers shape genetic structure of the Alpine Porcini (*Boletus reticuloceps*). *Mycorrhiza* **2017**, *27*, 261–272. [CrossRef]
63. Li, Y.; Ludwig, A.; Peng, Z. Geographical differentiation of the Euchiloglanis fish complex (Teleostei: Siluriformes) in the Hengduan Mountain Region, China: Phylogeographic evidence of altered drainage patterns. *Ecol. Evol.* **2017**, *7*, 928–940. [CrossRef]
64. Deng, B.; Chew, D.; Mark, C.; Liu, S.; Cogné, N.; Jiang, L.; O'Sullivan, G.; Li, Z.; Li, J. Late Cenozoic drainage reorganization of the paleo-Yangtze river constrained by multi-proxy provenance analysis of the Paleo-lake Xigeda. *GSA Bull.* **2021**, *133*, 199–211. [CrossRef]
65. Dayrat, B. Towards integrative taxonomy. *Biol. J. Linn. Soc.* **2005**, *85*, 407–417. [CrossRef]
66. Fujita, M.K.; Leaché, A.D.; Burbrink, F.T.; McGuire, J.A.; Moritz, C. Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol. Evol.* **2012**, *27*, 480–488. [CrossRef]
67. Oláh, J.; Kiss, O. Splitting by adaptive traits in the *Rhyacophila obscura* species group (Trichoptera, Rhyacophilidae). *Opusc. Zool.* **2018**, *49*, 152–161. [CrossRef]
68. Oláh, J.; Kovács, T.; Ibrahimi, H. *Agaphylax*, a new limnephilid genus (Trichoptera) from the Balkan: Lineage ranking by adaptive paramere. *Opusc. Zool.* **2018**, *49*, 77–89. [CrossRef]

Chapter 2

The impact of sequencing depth and relatedness of the reference genome in population genomic studies: A case study with two caddisfly species (Trichoptera, Rhyacophilidae, *Himalopsyche*)

Xi-LingDeng, Paul B. Frandsen, Rebecca B. Dikow, Adrien Favre, Deep Narayan Shah, Ram Devi Tachamo Shah, Julio V. Schneider, Jacqueline Heckenhauer, Steffen U. Pauls

This article was published by WILEY in Ecology and Evolution on 12 December 2022. available online:

https://doi.org/10.1002/ece3.9583

SUP, PBF, JH designed the work. SUP acquired funding for the project. DNS, RDTS and SUP collected the samples from the field. JVS and I performed laboratory work. JH, PBF, RBD and I analyzed the data. JH, PBF, SUP and I interpreted the results. JH and I wrote the manuscript with the support of all my coauthors.

Received: 22 September 2022 | Revised: 10 November 2022 | Accepted: 16 November 2022

DOI: 10.1002/ece3.9583

RESEARCH ARTICLE

Ecology and Evolution Open Access WILEY

The impact of sequencing depth and relatedness of the reference genome in population genomic studies: A case study with two caddisfly species (Trichoptera, Rhyacophilidae, *Himalopsyche*)

Xi-Ling Deng[1,2,3] | **Paul B. Frandsen**[3,4,5] | **Rebecca B. Dikow**[5] | **Adrien Favre**[1,6] | **Deep Narayan Shah**[7] | **Ram Devi Tachamo Shah**[8,9] | **Julio V. Schneider**[1] | **Jacqueline Heckenhauer**[1,3] | **Steffen U. Pauls**[1,2,3]

[1]Senckenberg Research Institute and Natural History Museum, Frankfurt/Main, Germany

[2]Institute of Insect Biotechnology, Justus-Liebig-University Gießen, Gießen, Germany

[3]LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt/Main, Germany

[4]Department of Plant & Wildlife Sciences, Brigham Young University, Provo, Utah, USA

[5]Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, DC, USA

[6]Regional Nature Park of the Trient Valley, Salvan, Switzerland

[7]Central Department of Environmental Science, Tribhuvan University, Kirtipur, Nepal

[8]Aquatic Ecology Centre, School of Science, Kathmandu University, Dhulikhel, Nepal

[9]Department of Life Sciences, School of Science, Kathmandu University, Dhulikhel, Nepal

Correspondence
Xi-Ling Deng, Senckenberg Research Institute and Natural History Museum, Senckenberganlage 25, D-60325 Frankfurt/Main, Germany.
Email: xiling.deng@outlook.com

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: AF 1117/1-2, PA 1617/2-1 and PA 1617/2-2; LOEWE — "Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz" of Hesse's Ministry of Higher Education, Research and the Arts

Abstract

Whole genome sequencing for generating SNP data is increasingly used in popula- tion genetic studies. However, obtaining genomes for massive numbers of samplesis still not within the budgets of many researchers. It is thus imperative to select an appropriate reference genome and sequencing depth to ensure the accuracy of the results for a specific research question, while balancing cost and feasibility. To evalu- ate the effect of the choice of the reference genome and sequencing depth on down- stream analyses, we used five confamilial reference genomes of variable relatedness and three levels of sequencing depth (3.5×, 7.5× and 12×) in a population genomic study on two caddisfly species: *Himalopsyche digitata* and *H. tibetana*. Using these 30 datasets (five reference genomes × three depths × two target species), we estimated population genetic indices (inbreeding coefficient, nucleotide diversity, pairwise F_{ST}, and genome-wide distribution of F_{ST}) based on variants and population structure (PCA and admixture) based on genotype likelihood estimates. The results showed that both distantly related reference genomes and lower sequencing depth lead to degradation of resolution. In addition, choosing a more closely related reference

Jacqueline Heckenhauer and Steffen U. Pauls equal contribution authors.

genome may significantly remedy the defects caused by low depth. Therefore, we conclude that population genetic studies would benefit from closely related refer- ence genomes, especially as the costs of obtaining a high-quality reference genome continue to decrease. However, to determine a cost-efficient strategy for a specific population genomic study, a trade-off between reference genome relatedness and sequencing depth can be considered.

KEYWORDS
aquatic insects, de novo genomes, population genomics, reference genomes, sequencing depth, whole genome resequencing

TAXONOMY CLASSIFICATION
Entomology, Evolutionary ecology, Genetics, Genomics, Population genetics

1 | INTRODUCTION

As high-throughput sequencing (HTS) technologies and bioinformatic tools are rapidly becoming more accurate and increasingly affordable, it is possible to generate whole genome resequencing (WGR) data for almost any species. High-quality WGR data provide a remarkable amount of information, including a vast number of loci as well as a large number of genetic variants, thus enabling powerful population genomics analyses (Goodwin et al., 2016). Today, whole genome resequencing with low read depth, which indicates a low average number of reads that are aligned to a base in the reference genome, is widely applied in population studies (Lou et al., 2021; Nielsen et al., 2011; Sims et al., 2014). When initiating a project on population genetics using WGR, there are two prerequisites: (i) the availability of a reference genome representing the focal species (Ellegren, 2014) and (ii) estimating the necessary sequencing depth to support accurate results (Meisner & Albrechtsen, 2018).

The selection of reference genome and sequencing depth are therefore two important features in a population genetic study. However, with a fixed budget, it is important to find a balance between data quality and sequencing costs by compromising on the reference genome or sequencing depth. For every study, the reference genome needs to be carefully chosen to avoid bias in mapping and variant calling, considering the amount of sequence identity between the reference genome and the data resulting from resequencing (Nielsen et al., 2011). Ideally, such studies should include a high-quality species-specific reference genome, which is often not available for nonmodel organisms. Given thecosts and time associated with generating a de novo reference ge- nome, it can be more realistic to use an existing one, yet more dis- tantly related, as is done most often in population genomic studies (Duchen & Salamin, 2021). Currently, several empirical studies have examined the impact of nonconspecific reference genomesin population genomics. For example, Gopalakrishnan et al. (2017) compared the use of dog and wolf genomes as reference genomes for either domestic dog or wolf populations. Their results showed that the selection of the reference genome only had a minor influence on downstream analyses, probably because of the close relatedness of these two taxa, which diverged approx. 30,000 years ago (Skoglund et al., 2015; Wang et al., 2013). By contrast, other studies have found that the use of distantly related reference genomes biases the results of resequencing analyses in bacteria (Valiente-Mullor et al., 2021), fungi (Garcia-Rubio et al., 2018), and mammals (Yang et al., 2019), but no studies are yet available for insects.

In comparison to choosing the reference genome based upon relatedness, sequencing depth needs more preliminary knowledge to determine. The sequencing depth needs to be tailored to each particular study based on many aspects, for instance genome size of the target species, the availability of funding, and of course the research question. Using the strategy of low sequencing depthin a population genetic study may result in data loss, thus caus-ing statistical uncertainty during genotype and variant calling (Crawford & Lazzaro, 2012; Meisner & Albrechtsen, 2018; Nielsen et al., 2012). This statistical uncertainty is mainly due to the limited information provided by the low amount of reads, leading to poor discrimination between sequencing error and real variation, that is, SNPs (Meisner & Albrechtsen, 2018). To improve the accuracy of estimates in a cost-limited population genetic study, especially using low sequencing depth, some researchers therefore demon- strated that employing a large sample size provides more accurate results, for instance, more than 50 individuals from a popula- tion with a sequencing depth of two (Buerkle & Gompert, 2013; Fumagalli, 2013; Han et al., 2014; Sims et al., 2014). However, unlike these studies that are based on simulated data, it is chal- lenging to obtain a large number of samples for taxa such as the aquatic insect genus *Himalopsyche*. These species are apex pred- ators in the benthic invertebrate community and naturally havesmall population sizes. Many other taxa are similarly rare. Thus, it is pragmatic to consider how we might improve population-based inference through the optimization of reference genome choice or sequencing depth for these sample-limited studies. Therefore, comprehensively exploring the influence of reference choice and

sequencing depth on downstream analyses is urgently needed, particularly for insects. In this study, we will evaluate the effects of the reference genome and sequencing depth on a genus of caddisflies, for which we already have significant molecular and taxonomic data (Deng et al., 2021; Heckenhauer et al., 2022).

Himalopsyche is a genus of caddisflies (Insecta: Trichoptera) that is primarily distributed in the mountainous areas of central and east Asia (Hjalmarsson et al., 2018). Their larvae live as free-roaming predators in cool fast-flowing rivers and streams and are regarded as bioindicators of water quality (Hjalmarsson, 2019; Morse et al., 2019; Tsuruishi et al., 2006). There are currently 56 named species of *Himalopsyche* (Hjalmarsson, 2019), whichare divided into five species groups: the *tibetana* group, the *lep- cha* group, the *kuldschensis* group, the *phryganea* group, and the *navasi* group (Hjalmarsson et al., 2019). The species *H. digitata* and *H. tibetana* both belong to the *tibetana* group, which is dis- tributed in the Himalayas. This area is characterized by a numberof parallel north–south running river systems (such as the Karnali, the Narayani, and the Koshi) with sharp elevational gradients. Currently, climate change is causing a number of cascading ef-fects on river flow via rapidly receding glaciers greatly affecting aquatic biodiversity (Xu et al., 2009). It is therefore crucial to in- vestigate the patterns of genetic diversity of aquatic insects in the region to understand the current and past ecological processes,in order to promote the conservation of freshwater biodiversity(Geist, 2011). Genome-wide analysis is an important conserva- tion tool that can provide novel insights essential for identifying,for example, hotspots or reservoirs of genetic diversity, dispersal routes, ecological corridors, and stepping stone habitats (Barbosa et al., 2018; Brandies et al., 2019; Hohenlohe et al., 2021; Jasperet al., 2019). However, the quality and quantity of aquatic in-sect genomes are still relatively low compared with terrestrial insects (Hotaling et al., 2020, 2021). Despite Trichoptera covering ~275 million years of evolution (Thomas et al., 2020) and comprising ~16,300 known species (Morse et al., 2019), only 29 Trichoptera genome assemblies (26 species) have been published to date (Heckenhauer et al., 2019, 2022; Luo et al., 2018; Olsen et al., 2021; Ríos-Touma et al., 2022). This represents <0.15% of all known species, which limits progress in genomics-based research of this ecologically relevant group.

The number of studies using population genomics is rapidly increasing. With it, the need to test the effect of reference ge- nome's selection and sequencing depth on the results. Indeed,such studies will allow to find a balance between data quality and sequencing costs by compromising on the reference genome or sequencing depth. Therefore, we conducted a case study usingan empirical dataset to evaluate how reference genome selection, that is, degree of relatedness, and sequencing depth affect down- stream population genetic analyses of the species *H. digitata* and *H. tibetana*. In addition, we tried to reveal the correlation between the inferences from population genetics and drainage network and provided insights for local biodiversity conservation of these enigmatic species.

2 | MATERIALS AND METHODS

2.1 | Study design

As illustrated in Figure 1, first, we generated three new de novo whole genome assemblies for *H. tibetana* (*tibetana* group), *H. japon- ica* (*navasi* group), and *H.* sp. (*kuldschensis* group) sensu Hjalmarsson et al., 2019, in addition to two previously published genomes (i.e., *H. phryganea* (*phryganea* group) and *R. brunnea*; Heckenhauer et al., 2022). Since the specimens of *Himalopsyche sp.* was collected as a larva that cannot be identified to species level by morphologic diagnosis, and the CO1 sequence of this specimen differed slightly from all hitherto known sequences of named species, we can only classify this specimen as a species belonging to the *kuldschensis* group. These five genomes represent a gradient of genetic relatedness with respect to our target species, which were used as reference genomes. We used populations of two *Himalopsyche* species (*H. digitata* and *H. tibetana*), each species including four populations with each population containing six individuals except for one popu- lation from *H. digitata* and *H. tibetana*, respectively, which contained only five individuals. The reads of the populations were subsampled into three separate datasets with an average depth of 12.5×, 7.5× and 3.5×, respectively. Reads from each dataset were mapped to the five different reference genomes separately. Afterward, variants were called from all datasets using two strategies: Genotype calling with GATK and direct estimation of the genotype likelihood using ANGSD. Variants identified with the first strategy were used to calculate the population genetic indices including inbreeding coefficient (F), nucleotide diversity (π), and pairwise fixation index (F_{ST}); variants estimated from the second strategy were used in the principal component analysis (PCA) and admixture analysis. Finally, we compared population genetic indices and population structure with different references and sequencing depths.

2.2 | De novo genomes of three reference species

We used the genome assemblies of four species of *Himalopsyche* (*H. tibetana*, *H.* sp., *H. phryganea*, and *H. japonica*) and of one species from the closely related genus *Rhyacophila* (*R. brunneae*, both family Rhyacophilidae) as reference genomes. The *Himalopsyche* species represent the four major taxonomic groups in the genus according to Hjalmarsson et al. (2018): the *tibetana* group (*H. tibetana*), the *kuldschensis* group (*H.* sp.), the *phryganea* group (*H. phryganea*), and the *navasi* group (*H. japonica*). Unfortunately, we could not obtain a sample of *H. lepcha* (the only species in the *lepcha* group). *Rhyacophila* is the most closely related genus to *Himalopsyche* (Thomas et al., 2020). The genomes of *H. phryganea* (JAGVSL000000000) and *Rhyacophila brunnea* (previous version of JAGYXB000000000, available at https://doi.org/10.6084/m9.figshare.c.6033011.v1) were previously sequenced and assembled (Heckenhauer et al., 2022). We generated new de novo assemblies of *H. tibetana* (collected from the Ê Ghunsa River, Nepal), *H.* sp. (*kuldschensis* group, collected from the Ê Ghunsa

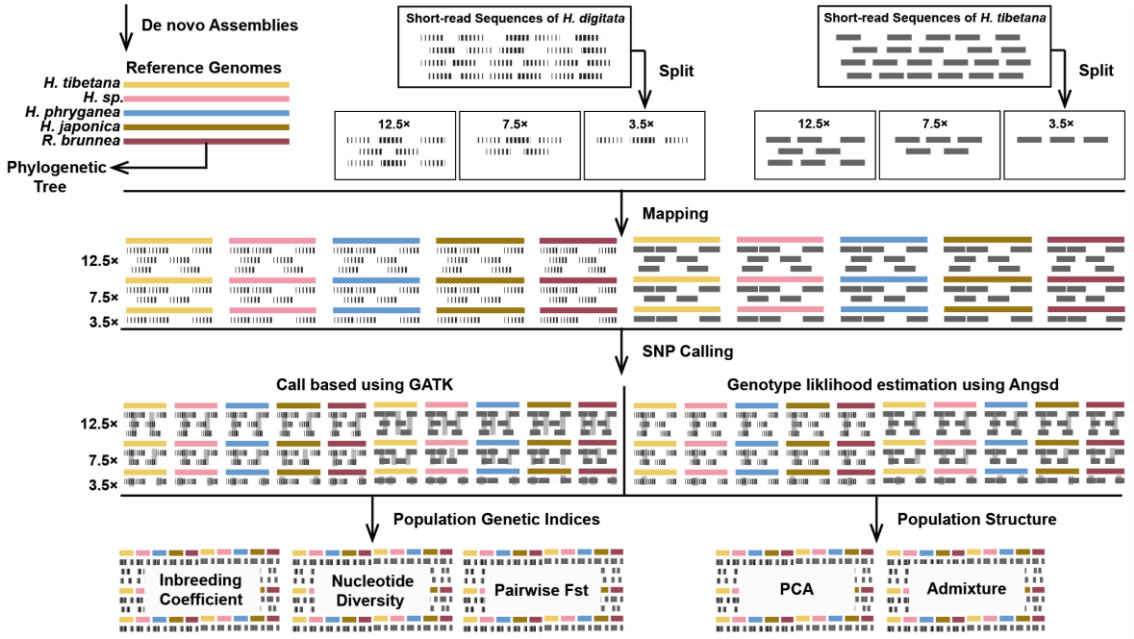


FIGURE 1 Workflow of data processing in this study showing the treatment of short-read resequencing data from the two target species and subsequent mapping and variant calling with different reference genomes for assessing genetic diversity and population genetic structure (see details in Section 2.1).

River, Nepal), and H. japonica (collected from the Nogami River, Kisomachi, Nagano Prefecture, Japan). Tissue of abdomen and thorax segments were used for DNA extraction after removal of the intestinal tract. We extracted high molecular weight genomic DNA using a salting-out protocol adapted from Miller et al. (1988), as described in Heckenhauer et al. (2019). We quantified the DNA using a Qubit 4.0 fluorometer with the dsDNA Broad Range Kit (ThermoFisher Scientific) and checked its purity with a DS11 spectrophotometer (DeNovix). We used a low-cost sequencing strategy that has been shown to produce contiguous genome assemblies, that is, employing a combination of short (Illumina) and long-read (Oxford Nanopore) technologies to sequence the three new reference genomes, as described in Appendix S1 (Section 1.1).

We conducted a long-read assembly of the Oxford Nanopore Technology sequencing reads with wtdbg2 v2.4 (Ruan & Li, 2020), followed by mapping and polishing with long reads with Minimap2 v14 (Li, 2018) and Racon v1.3.1 (Vaser et al., 2017). We then performed another round of long-read polishing with nanopolish 0.11.1 (Loman et al., 2015) by first indexing the signal-level data in the FAST5 files using nanopolish index, realigning the long reads to the Racon-polished assembly with minimap2, and then sorting and indexing the bam file with samtools. We used nanopolish_makerange.py to split our draft genome assembly into 50-kb segments and generated a consensus for each segment in parallel with nanopolish variants (–-consensus --min-candidate-frequency 0.1). We generated the polished genome in FASTA format using nanopolish vcf2fasta. Because noisy long reads can suffer from indel errors, even with polishing, we further polished the assembly with high-quality short-read data with Pilon v1.22 (Walker et al., 2014). To do this, we mapped quality trimmed Illumina reads to the "nanopolished" assembly with bwa-mem and sorted the read alignments by leftmost coordinates using the sort options in SAMtools v1.9 (Li et al., 2009). Finally, we used Pilon v1.22 (option --fix indels) to polish the assembly. Following polishing, we used purge_dups 1.2.3 (Roach et al., 2018) to purge haplotigs and overlaps in the assembly based on read depth.

For H. japonica, this pipeline did not meet the expected quality regarding contiguity and BUSCO completeness. Thus, we conducted a de novo hybrid assembly with the raw Illumina data together with the long reads using MaSuRCA v.3.1.1 (Zimin et al., 2013, 2017). In the config file, we specified the insert size and a standard deviation (10% of insert size) for the Illumina reads, as well as jellyfish hash size (estimated_genome_size* ~long-read coverage (equal to depth in this scenario)). All other parameters were left as defaults. We used purge_dups 1.2.3 to purge haplotigs and overlaps in the assembly based on read depth.

We calculated assembly statistics with QUAST v5.0.2 (Gurevich et al., 2013) and examined completeness with BUSCO v4.1.4 (Simão et al., 2015; Waterhouse et al., 2018) using the Endopterygota odb10 dataset with the options --long, –m = genome and -sp = fly. A summary of the assembly statistics and BUSCO completeness is given in Table 1. The final genome assemblies were screened for potential contaminations with taxon-annotated GC-coverage plots (TAGC plots) using BlobTools v1.0 (Laetsch & Blaxter, 2017). For this

TABLE 1 Assembly statistics of reference genomes used in this study.

Species	Accession number	Sequencing platform (depth)[d]	Assembly length bp	N50 (bp)	No of contigs	N's per 100 kb	BUSCOS % [c]	Number of proteins
Himalopsyche japonica[a]	JAHFWJ000000000	Nanopore +Illumina (18× +170 ×)	546,840,812	2,150,202	847	0	C:97.2% [S:96.6%, D:0.6%], F:0.8%, M:2.0%, n:2124	9983
Himalopsyche sp.[a]	JAHFWI000000000	Nanopore +Illumina (26× +200 ×)	592,402,457	7,599,818	528	0	C:96.7% [S:96.3%, D:0.4%], F:1.0%, M:2.3%, n:2124	10,049
Himalopsyche phryganea	JAGVSL000000000	Nanopore +Illumina (36.8 × +170 ×)	633,785,554	4,634,010	710	0	C:97.0% [S:96.5%, D:0.5%], F:1.0%, M:2.0%, n:2124	10,994
Himalopsyche tibetana[a]	JAHFWH000000000	Nanopore +Illumina (24× +150 ×)	665,312,086	949,059	1853	0	C:96.4% [S:95.6%, D:0.8%], F:1.0%, M:2.6%, n:2124	10,994
Rhyacophila brunnea	JAGYXB000000000 [b]	Nanopore +Illumina (19× +154 ×)	1,086,872,538	1,030,560	2125	0.36	C:96.0% [S:93.3%, D:2.7%], F:1.1%, M:2.9%, n:2124	10,846

[a]This study.
[b]In this study, we used a previous version of this assembly for SNP calling.
[c]Based on the Endopterygota odb10 dataset (2124 genes), C: complete, S: single, D: duplicated, F: fragmented, M: missing.
[d]Based on Genomescope2 genome size estimation.

purpose, all preprocessed Illumina reads of the respective species were mapped against the final genome assemblies using BWA- MEM v0.7.17- r1188 (Li, 2013) and taxonomic assignment for BlobTools was done with blastn using -task megablast and -e-value 1 e-25. Details are given in Appendix S1 (Section 1.2).

Genome size estimation and profiling was conducted from the short-read sequence data with GenomeScope 2.0 (Ranallo-Benavidez et al., 2020; Vurture et al., 2017) as described in Appendix S1 (Section 1.3).

2.3 | Annotation

We identified and classified repetitive elements de novo and generated a library of consensus sequences for each genome using RepeatModeler 2.0 (Flynn et al., 2020). We then annotated and masked repeats in each assembly with RepeatMasker 4.1.0 (http://www.repea tmasker.org) using the custom repeat library for the species generated in the previous step. After masking repeats, genes were predicted using the homology-based gene prediction tool GeMoMa v1.6.4 (Keilwagen et al., 2016, 2019) and the two previously published species (H. phryganea = RG 1 and R. brunnea = RG 2) as reference organisms as follows: GeMoMa -Xmx50G GeMoMaPipeline threads=$SLURM_NPROCS outdir=<out_dir> GeMoMa.Score =ReAlign AnnotationFinalizer.r =NO o=true t=<genome to be annotated> s=own i=<name of RG 1>a=<RG 1.gff >g=<RG 1 assembly.fasta >s=own i=i=<name of RG 2>a=a=<RG 2.gff >g=g=<RG 2 assembly.fasta >.

For functional annotation of predicted genes, we first split each amino acid FASTA file into multiple files with 50 sequences each using awk and then used blastp to search against the ncbi-blast2.9.0 + nr database with an e-value cutoff of 10– 4, –max_target_seqs set to 10 and -out format 6. The resulting xml files were merged using cat and then functionally annotated using the command-line version of Blast2GO (Götz et al., 2008).

2.4 | Species tree reconstruction

To determine the phylogenetic relatedness among the five species, we estimated a species tree using single-copy orthologs resulting from the BUSCO analyses of the five genome assemblies with an additional species, Glossosoma conforme, as an outgroup (Heckenhauer et al., 2022). For each single-copy ortholog, we generated an unaligned FASTA file with sequences from each species. We then aligned each ortholog with the MAFFT L-INS- i algorithm (Katoh & Standley, 2013). We selected the best-fit substitution model for each alignment using ModelFinder (option -m mfp, (Kalyaanamoorthy et al., 2017)) in IQtree v.2.0.6 (Minh et al., 2020) and estimated a maximum-likelihood tree with 1000 ultrafast bootstrap replicates (Hoang et al., 2018) with the BNNI correction (options -bb 1000 -bnni). We then generated a multispecies coalescent species tree in ASTRAL- III (Zhang et al., 2018) using the best maximum-likelihood

tree from each ortholog as input. We visualized the trees using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

2.3 | Cactus alignment and Hal

To further characterize the variation among the five genomes, we computed a whole genome alignment using Cactus v1.0.0 (Armstrong et al., 2020) with a star tree of the five genomesas input: ((Rhyacophila_brunnea, Himalopsyche_kuldschensis, Himalopsyche_phryganea, Himalopsyche_japonica, Himalopsyche_tibetana)mr).

We used HALtools (Hickey et al., 2013) to obtain global alignment information with the halStats option. We then used "halSummarizeMutations" to generate mutation statistics, including the length of each mutation and the number of substitutions, transi- tions, transversions, gaps, insertions, deletions, inversions, duplica- tions, and transpositions among genomes.

2.4 | Population genomic analyses

2.4.1 | Taxon sampling

We conducted whole genome resequencing on a total of 46 individuals from four *H. tibetana* populations and four *H. digitata* populations, with six individuals from each population except for pop 1 of *H. digitata*, which included five individuals. All 46 samples were collected as larvae in April 2018 and March 2019 in Nepal. All four *H. digitata* populations and two populations of *H. tibetana* were collected in the headwaters of the Gandaki basin with two additional populations of *H. tibetana* collected in the headwaters of Koshi basin (Figure 6). Since it is presently not possible to identify *Himalopsyche* larvae to species level solely based on morphological characters, we ensured correct identification of these samples based on two molecular markers: the mitochondrial COI and the nuclear CAD, using the methods outlined in Hjalmarsson et al. (2018).

All the samples were preserved in 95% ethanol and archived in the collections of the Senckenberg Research Institute and Natural History Museum (SMF). Specimen and voucher information is shown in Appendix S2 and Figure 6.

2.4.2 | DNA extraction, library preparation, and sequencing

Genomic DNA was extracted following a modified salting-out protocol adapted from Miller et al. (1988), as described in Heckenhauer et al. (2019). In case of low purity DNA (A260/A230 purity ratio below 1.4 and DNA concentration higher than 100 μg/μl that was measured by DeNovix DS11 spectrophotometer), the template DNA was subject to an additional cleanup using magnetic beads as described in Appendix S1 (Section 2.1). Afterward, all the samples were sent to Novogene Co., Ltd. for DNA library preparation and sequencing. 150-bp paired-end reads were generated on an Illumina HiSeq 2000 platform.

2.4.3 | Quality control, data processing of population sequence data, and variant calling

We assessed the quality of Illumina reads before and after each step using FastQC v0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC v1.7 (Ewels et al., 2016), details are described in Appendix S1 (Section 2.2). We trimmed overrepresented *k-mers* using autotrim.pl v0.6.1 (Waldvogel et al., 2018) with Trimmomatic v0.38 (Bolger et al., 2014), and we removed adapters with Cutadapt v2.23 (Martin, 2011). After quality filtering, the average sequence depth was 18× and the minimal depth was 12.5× for all the samples. To maintain even depth of all the individuals and to simulate subsets with varying depth levels, we chose 12.5×, 7.5× and 3.5× as our test depth levels. We randomly subsampled the sequencing reads to a specified depth using rasusa 0.3.0. (Hall, 2022) with a random seed of 1 (−s 1), the target depth (−-coverage 12.5, −-coverage 7.5, −-coverage 3.5) and the estimated genome size (−-genome-size 550 m) based on GenomeScope2 estimation (see below; Ranallo-Benavidez et al., 2020; Vurture et al., 2017).

For population genomic analyses, FASTA files of the reference genomes were indexed with the function bowtie2-build of bowtie2 2.3.5 (Langmead & Salzberg, 2012). We then mapped the reads of the three different depth datasets (12.5×, 7.5× and 3.5×) of the *H. digitata* and *H. tibetana* populations to each of the five different reference genomes separately using Bowtie2, which resulted in 30 datasets (2 species × 3 depth × 5 reference genome, Figure 1). After marking duplicate reads for each dataset using Picard v2.20.8 (Picard Tools – By Broad Institute), we conducted variant calling for downstream population genetic analysis using two strategies. In the first strategy, we called the haplotype of each individual separately, by running GATK v4.1.7.0 (GATK, broadinstitute.org) HaplotypeCaller on each bam file. Then, we called genotypes with GATK GenotypeGVCFs across all resulting vcf files. Ultimately, we selected the variants and filtered missing sites using VCFtools v0.1.17 (Danecek et al., 2011). With the second strategy, we esti- mated genotype likelihoods with appropriate filtering (including reads/sites/alleles/depth filtering, nonmissing individual, and SNP filtering) using ANGSD v0.931 (Korneliussen et al., 2014). For details and parameters used, see Appendix S1 (Sections 2.3 and 2.4).

2.6.4 | Genetic diversity and population structure analysis

The variants called by GATK with genotype calling were used to estimate π, individual *F*, and pairwise F_{ST} (Weir & Cockerham, 1984). These are commonly used measurements for population genetics. *F* is used to directly quantify the alleles inherited from common

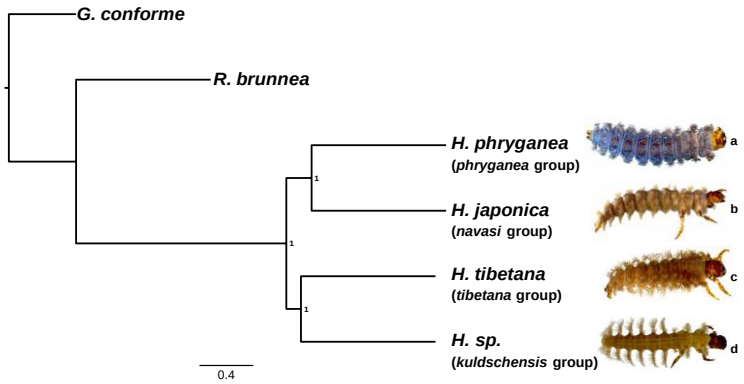


FIGURE 2 Astral tree of the five reference genomes generated from BUSCO genes. Numbers on the nodes indicate local posterior probabilities. *G. conforme* was used as an outgroup. Images taken from Hjalmarsson et al. (2019), (a) *H. phryganea*, (b) *H. japonica*, (c) *H. gregoryi* (identical with *H. tibetana* morphologically), (d) *H. sylvicola* (identical with *H.* sp. (*kuldschensis* group) morphologically).

ancestors in an individual's lineage, π is an index for population-level genetic diversity quantification, and F_{ST} provides a primary description of population genetic differentiation. In this study, π and F_{ST} were calculated using VCFtools with a 50-kb window size. To better understand F_{ST} within different datasets (species × reference genome × depth), we applied pairwise F_{ST} estimates (pairwise populations among the four populations) and global F_{ST} estimates (including all the four populations).

We used the genotype likelihoods estimated by ANGSD to generate a PCA with PCAngsd (Meisner & Albrechtsen, 2018) and individual admixture proportions estimating using NgsAdmix (Skotte et al., 2013) after Linkage pruning with ngsLD (Fox et al., 2019). Further details about tools, parameters, and commands were described in Appendix S1 (Section 2.4). Most plots were generated in Rstudio v3.6.1 (http://www.rstudio.org) (scripts for plotting were included in Appendix S1 (Section 3)).

2 | RESULTS

2.1 | New genomic resources

We combined long- and short-read sequencing technologies to generate three new de novo genome assemblies for the genus *Himalopsyche* (Rhyacophilidae): *H. japonica*, *H.* sp. (*kuldschensis* group), and *H. tibetana*. For each species, we obtained ~150–200× Illumina and ~18–26× Oxford Nanopore sequencing depth. All three assemblies are of high quality with respect to the number of contigs and contig N50 (Table 1). We identified >96% of the Endopterygota BUSCO gene set in the assemblies. BlobTools detected no contamination (Appendix S1: Figures 1–3). Remapping the Illumina reads back to the assemblies revealed more than 98% could be unambiguously placed (Appendix S1: Figures 1–3).

The estimated genome sizes resulting from the *k-mer*-based estimation with Genomescope2 were 481 Mb (*H. japonica*), 495 Mb (*H.* sp. (*kuldschensis* group)), and 568 Mb (*H. tibetana*; Appendix S1: Figures 4–6). Between 31% (*H. japonica*) and 44% (*H. tibetana*) of the genome assemblies were identified as repeats. A high percentage of the repeats were classified as interspersed repeats (approx. 28.5–40.3%). More than half of the interspersed repeats remain unclassified and therefore may be specific to Trichoptera. Details on repeat classes are given in Appendix S1 (Tables S1–S3). The annotation of the genomes resulted in the prediction of 9983 (*H. japonica*), 10,049 (*H.* sp. (*kuldschensis* group)), and 10,994 (*H. tibetana*) proteins. Most of the annotated proteins had functional Blast2GO annotations, were verified by BLAST, or were mapped to GO terms. GO Distributions were similar to previously annotated caddisfly genomes, that is, the major biological processes were cellular processes. Catalytic activity was the largest subcategory in molecular function, and the cell membrane subcategories were the largest cellular component (Appendix S1: Figures 7–12).

The Cactus alignment of the five reference genomes showed that *Rhyacophila brunnea* had the longest sequence length and largest number of contigs (Appendix S4). It also showed a higher level of mutations compared with the four *Himalopsyche* genomes, for instance gaps, insertions, inversions, and duplications. The number of mutations among the four *Himalopsyche* genomes was at similar levels in the Cactus alignment over all.

2.2 | Phylogenetic relationships of the reference genomes—resolving the *Himalopsyche* backbone

We built a species tree for the five reference genomes to estimate their evolutionary history and to verify the phylogenetic relationship between the reference genomes and the two population-level target species, *H. digitata* and *H. tibetana*. Phylogenetic relationships were strongly supported (Figure 2). Our phylogenetic tree showed two pairs of sister species (*H. tibetana* + *H.* sp. (*kuldschensis* group); *H. phryganea* + *H. japonica*) with *R. brunnea* forming the sister clade of all four *Himalopsyche* species. According to Hjalmarsson et al. (2018, 2019), *H. digitata* and *H. tibetana* both belong to the *tibetana* group, *H.* sp. to the *kuldschensis* group, *H. phryganea* to the *phryganea* group and *H. japonica* to the *navasi* group. In contrast to our results, Hjalmarsson et al., 2019 recovered the *navasi* group sister to all other Himalopsyche and a sister relationship between the *phryganea* and

kuldschensis groups. Differences could be related to the different sampling strategies. Here, we present much more genome-wide data per taxon, but reduced taxon sampling. For the purpose of assessing the impact of reference genome relatedness in SNP calling, H. tibetana and H. digitata (both belonging to tibetana group) are most closely related to H. tibetana, followed by H. sp. (kuldschensis group), H. phryganea and H. japonica, with R. brunnea being the most distantly related species.

3.3 | The impacts of sequencing depth and phylogenetic relatedness of the reference genome on population genetic studies

To better understand the impacts of sequencing depth and reference genome on population genomic analyses, we compared the number of quality-filtered variants, genetic diversity and population structure of the species H. digitata and H. tibetana among the datasets based on different reference genomes and varied sequencing depth. The results revealed that, perhaps unsurprisingly, the most distantly related reference genome, or the lowest sequencing depth, resulted in the least accurate downstream analysis.

We observed that reference genome selection has a measurable impact on the number of variants (Figure 3). With both strategies (variant calling using GATK and genotype likelihood estimation using Angsd), the number of variants sharply decreased when the reference genome was more distantly related, but remained similar with decreasing sequencing depth, especially when estimated by genotype likelihood. This was particularly striking for the H. tibetana populations when using ANGSD to call the variants, which dropped from millions to thousands when changing the reference genome from H. tibetana to the others (Figure 3). Consequently, the massive loss of information resulting from choosing a distant reference genome is likely to lead to poor performance in the downstream analyses.

We observed that the choice of reference genome influenced population genetic values such as pairwise F_{ST} (Appendix S3) and nucleotide diversity estimates (Appendix S1: Figure 14), the number of outliers in the inbreeding coefficient estimates (Appendix S1: Figure 13), as well as the resolution of the final outcome in genome-wide F_{ST} estimates (Figure 4), PCA, and admixture (Figure 5). For example, in the case of pairwise F_{ST} estimates, when changing the reference genome from H. tibetana to R. brunnea (with the same sequencing depth, e.g., 12.5 ×), the F_{ST} value between pop 1 and pop 3 of H. tibetana reduced from 0.17 to 0.06 (Appendix S3). The number of outliers ($F < 0$) in inbreeding coefficient estimates increased when using a reference genome that is distantly related to the target species, while the inbreeding coefficient values of individuals tended to fluctuate for both species, which might result in an ineffective or misleading conclusion (Appendix S1: Figure 13). The results of the PCA, admixture analysis, and the genome-wide F_{ST} clearly showed notable decreases in the resolution of the plots when selecting a more distantly related reference genome (Figure 5). For instance, for populations of H. digitata, regardless of the depth, the cluster of pop 1 was well defined when mapped to H. tibetana in the PCA plot and admixture analysis, but no distinguishable structure was discernible among all four populations when mapped to R. brunnea (Figure 5a). In addition, using a conspecific reference genome significantly improved the accuracy of genome-wide F_{ST} estimates, including both the global F_{ST} value (the global F_{ST} value of H. tibetana populations doubled when mapped to H. tibetana) and the distribution of F_{ST} values from sliding windows. This effect was particularly pronounced with the abnormal value (highly qualified resolution, Figure 5b).

We also observed extensive influence of sequencing depth on the downstream analyses. In addition to decreasing number of variants, low sequencing depth also affected estimates of nucleotide diversity, the inbreeding coefficient, pairwise and global F_{ST}, and population structure. In most cases, low sequencing depth reduced the accuracy of these inferences; however, the influence proved to be negligible or limited in some treatments. For example, the population structure of H. tibetana was highly differentiated regardless of depth when mapped to H. tibetana and was less differentiated when sequencing depth decreased while mapped to H. phryganea, but still sufficient to obtain a reliable result (Figure 5b). Likewise, the F value of all individuals changed when the sequencing depth decreased from 12.5 × to 3.5 ×, but the differentiation among populations remained comparable (Appendix S1: Figure 13).

Furthermore, the results revealed that reference genome and sequencing depth had variable impacts between the populations of H.

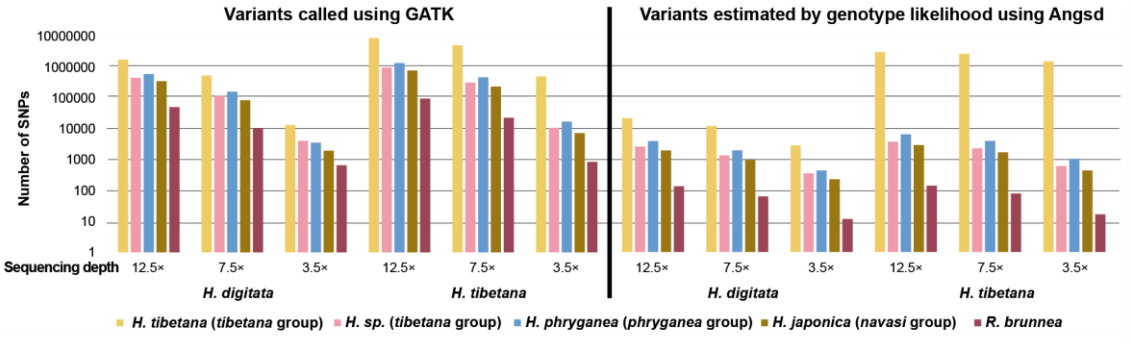


FIGURE 3 Number of variants called/estimated by two different strategies. Numbers of the Y-axis were logarithmically scaled.

digitata and *H. tibetana*. More specifically, the same treatment (same reference genome and sequencing depth) might be sufficient to generate a reliable result for the populations of *H. tibetana*, but not for *H. digitata*. For instance, when mapped to *H. japonica*, the population structure of the *H. tibetana* populations was distinguishable regardless of sequencing depth, but it was barely detectable for the *H. digitata* populations even with 12.5× depth (Figure 5). Considering the inherent genetic variation of the two species, it is not surprising that more distantly related reference genome or lower sequencing depth were more tolerable for the populations with higher genetic diversity.

2.1 | Genetic diversity and population structure of *H. digitata* and *H. tibetana*

Each population of *H. digitata* had a similar level of nucleotide diversity, but less consistent *F* values: pop 4 was slightly higher (0.31), whereas the other three were similar (0.22–0.27, Figure 6d). Population 1, which is located in the main river of Gandaki basin in central western Nepal, formed a distinct cluster in the PCA analyses (Figure 6a,b). Moreover, no admixture signal was detected in pop 1. As a population located in the middle of the basin, but more closely connected with pop3 and pop 4 by catchment, pop 2 formed a distinct cluster and adjoined pop3 and pop4 in the PCA analysis. Meanwhile, pop 2 was represented as genetic mixtures regardless of *K* values. Population 3 and 4 were mixed together in the PCA analyses, which was in line with the fact that they were located closely together in the most northeastern tributaries of the Gandaki basin. In addition, pop 3 and pop 4 were homogeneous when $K = 2$, but mixed when $K = 3$ or 4. This population structure was further supported by the pairwise F_{ST} estimates using the base-called variants (Figure 6f).

Following the pattern of *H. digitata*, the populations of *H. tibetana* had a very similar level of both nucleotide diversity and inbreeding coefficient (Figure 6e) and a more distinct population structure congruent with catchments (Figure 6c,f). Moreover, compared with *H. digitata*, *H. tibetana* showed a greater population diversity among the four populations, including higher nucleotide diversity (~12-fold) and F_{ST} (both pairwise and globalwise, Figures 6f and 4), lower inbreeding coefficient (~ 0.5-fold), as well as a more distinct population structure (Figure 6b,c). This is in accordance with the fact that the geographic locality and catchment connection of the *H. tibetana* populations were further apart from each other compared with the ones of the *H. digitata* populations.

3 | DISCUSSION

3.1 | Reference genomes

In terms of BUSCO completeness and contiguity, the threede novo genomes provided in this study as references are of comparable quality than the other Trichoptera genomes published previously (Heckenhauer et al., 2019, 2022; Luo et al., 2018; Olsen et al., 2021; Ríos-Touma et al., 2022). They contained 96%–97% of an Endopterygota core gene collection indicating an almost complete coverage of known single-copy orthologs in the assembly. The back-mapping rate of Illumina reads to the assemblies ranged between 96 and 98%, which also indicates the high quality of the assemblies.

Previous genomic studies have focused on sequencing a wide range of different families of Trichoptera and investigating variations across the order (Heckenhauer et al., 2022). These three new genome assemblies provide important new genomic resources for the scientific community, especially since Trichoptera and other aquatic insects are in general underrepresented in genomic research (Hotaling et al., 2020, 2021). Together with previously published ones (*H. phryganea* and *R. brunnea*), these newly available genomes adequately provide an initial perspective about the phylogenetic relationship of the four main taxonomic groups of *Himalopsyche*, as well as implying the genetic relatedness between the target species and the five different reference genomes, respectively, for this study.

3.2 | Impacts of reference genome and sequencing depth on population genetic inferences

High-throughput sequencing of individual specimens is poised to become the state-of-the-art in population genetics studies. However, despite falling prices in sequencing, this approach can be prohibitively costly for large number of samples in species with large genomes. Many researchers may thus face the choice of either using an existing nonconspecific reference genome or lowering sequencing depth of individual samples. Previous studies have demonstrated that either the reference genome or the sequencing depth has a direct influence on population genetic estimates (Garcia-Rubioet al., 2018; Gopalakrishnan et al., 2017; Valiente-Mullor et al., 2021;Yang et al., 2019). Studies assessing the joint effects of the refer- ence genome and sequencing depth are rare and hitherto lacking in insects. Here, we evaluated population diversity and structure based on a range of reference genomes and varied levels of low sequencing depth in the first case study on insects. The results revealed that the choice of the reference genome and sequencing depth both had an influence on estimating population genetic indices, including *F*, π and F_{ST}, as well as the genetic structure inferred from genotype likelihoods. To some extent, the general trends are that (a) the more closely related the reference genome is, the more stable the estimates of population genetic indices are, and (b) the higher the sequencing depth, the better the resolution of the population structure analyses. However, the results also vary depending on the inherent genetic variation of the target species.

The choice of reference genome has a stronger influence on downstream analyses than resequencing depth, which is consistent with previous studies (Garcia-Rubio et al., 2018; Günther & Nettelblad, 2019; Valiente-Mullor et al., 2021). This is mainly due to the massive loss of reads while mapping to a distantly related

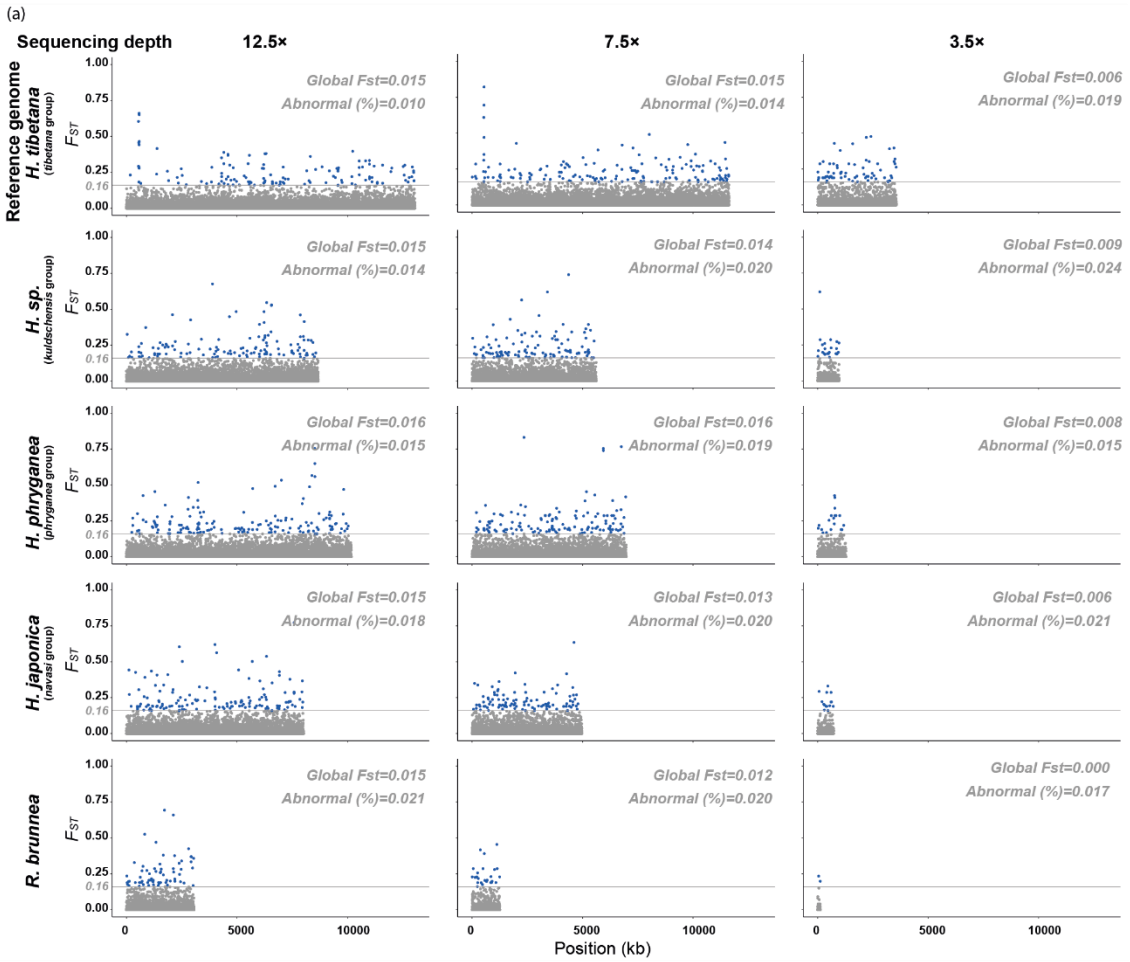


FIGURE 4 Genome-wide distribution of F_{ST} values (weighted) of (a) the H. digitata populations and (b) H. tibetana populations. F_{ST} values were calculated in 50-kb windows across contigs obtained from each reference genome; thus, the X-axis (position) is not comparable among datasets. The horizontal gray lines indicate a threshold that is used for selecting the abnormal F_{ST} windows for each species. The threshold is the minimum value of the top 1% F_{ST} windows on the whole genome when using H. tibetana as reference genome and 12.5× as sequencing depth (0.16 for H. digitata and 0.28 for H. tibetana). The percentage of abnormal F_{ST} windows is the number of F_{ST} windows above the threshold over the total number of F_{ST} windows. The global F_{ST} is the genome-wide weighted Fst value estimated based on Weir and Cockerham (1984).

reference genome. For example, in H. tibetana, mapping rates decrease from ~98% when using a conspecific reference genome, to 10% with a reference genome of a species of another genus (Rhyacophila brunnea; Appendix S5). This leads to a dramatic decline of variants (Figure 3). Moreover, increasing mismatches may also occur due to alternative alleles, thus increasing the so-called "reference bias" (Günther & Nettelblad, 2019). Consequently, reference bias can impact variant calling by missing alternative alleles or by incorrectly calling heterozygous sites and therefore lead to an underestimation of variants, including rare/private variants (Günther & Nettelblad, 2019; Taub et al., 2010). Considering these two aspects, the effects related to the choice of reference genome may propagate to a certain degree to subsequent downstream analyses in a population genetic study, for example when investigating heterozygosity and genetic diversity, gene flow, as well as ancestry proportions (Brandt et al., 2015; Günther & Nettelblad, 2019). We have observed these in our results: the number of variants decreases logarithmically with genetic relatedness of the reference genome, especially for the genotype likelihood dataset called by ANGSD, which includes a strict filtering on base level, read level, sequencing depth, and other levels during the genotype calling. Moreover, when the genetic relatedness between reference genomes and target species decreases, the population genetic indices including F, π, and pairwise F_{ST} tend to become less accurate, especially at a low depth

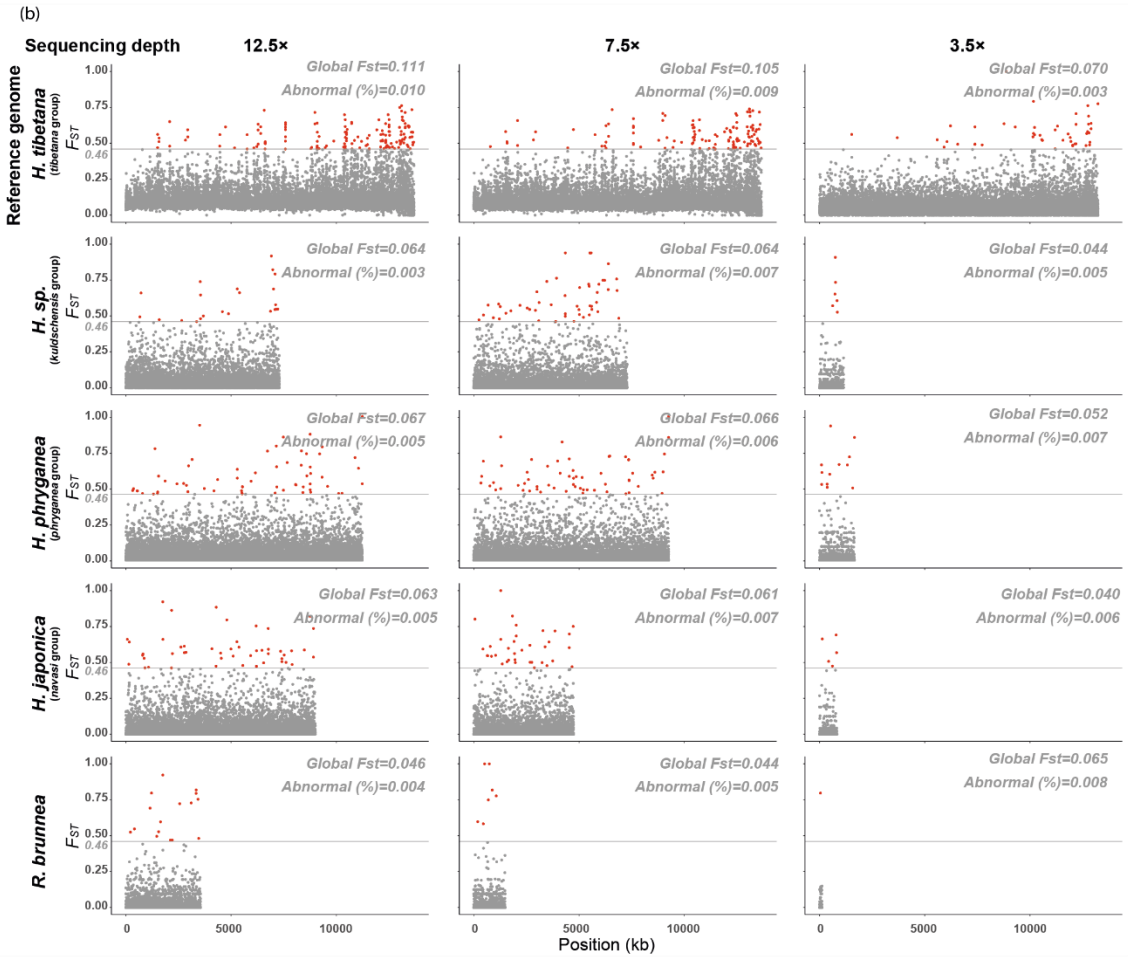


FIGURE 4 (Continued)

(3.5×). In addition, estimates of population structure, including PCA and admixture, are strongly affected, resulting in unreliable, weakly supported findings.

Notably, even though population genetic analyses of both H. tibetana and H. digitata show the greatest accuracy when using H. tibetana as the reference genome, we have observed a remarkable improvement in the results when using a conspecific reference genome, for instance, in the high resolution of the population structure, especially with the low sequencing depth (Figure 5b). For example, when mapping H. tibetana population samples to the H. tibetana reference genome, we observed a mapping rate of 98%. However, mapping success decreases rapidly with decreased relatedness. When mapping H. digitata to H. tibetana, the mapping rate is only 35% and decreases to 20% when mapped to H. sp. kuldschensis and 18% when mapped to H. phryganea. As a consequence, even though the genome size of the two species is similar (Table 1), the number of variants called from the H. digitata population dataset is much lower than variants called from the H. tibetana population dataset, especially when estimated with genotype likelihood methods (Figure 3). Although H. digitata and H. tibetana are recovered in the same clade in previous phylogenetic analyses (Hjalmarsson et al., 2019), the poor mapping success suggests that genetic distance may still be high. Unfortunately, there are no established divergence times within the genus of himalopsyche at present. The most recent study (Thomas et al., 2020) shows that the divergence between Rhyacophila and Himalopsyche was approx. 90 Ma. Considering the distinct ecological niches of H. tibetana (inhabits high altitude) and H. digitata (inhabits a lower altitude) in the same distribution range (both endemic to the Himalayas), we hypothesize that the divergence between these two species may be associated with the uplift of the Qinghai–Tibet Plateau, which began ca. 45 Ma ago (Ding et al., 2022). In summary, considering the long evolutionary history of caddisflies in general (ca. 280 Ma), and the rapid radiation of caddisflies (Thomas et al., 2020), a high level of divergence between two closely related

caddisflies species is not entirely unexpected. Therefore, we suggest that when selecting a closely related species as a reference genome in a population genomics study, it is important to consider genetic relatedness.

We also show that sequencing depth must be considered when designing population genomic analyses. Unlike de novo genome assembly which demands high sequencing depth, highly accurate results can be achieved with lower sequencing depth in population genomics by combining information from a large number of individuals either during SNP calling or other processes (Buerkle & Gompert, 2013 ; Fumagalli, 2013 ; Han et al., 2014). Even though the accuracy of population genetic inferences can be improved by increasing the number of samples, the bias caused by low sequencing depth cannot be ignored, especially since it can often be difficult to obtain many samples for some populations of rare animals. As revealed by previous studies, low sequencing depth may cause erroneous SNP calls, due to the errors introduced and amplified during PCR during library prep (Sims et al., 2014). Moreover, it may also produce ambiguous reads during mapping to the reference genome (Taub et al., 2010). Such biases may lead to incorrect conclusions in population genetics inferences, including population genetic differentiation, population structure, and demography (Crawford & Lazzaro, 2012 ; Fumagalli, 2013 ; Han et al., 2014 ; Jiang et al., 2019 ; Korneliussen et al., 2013). Our results show that decreasing depth massively reduced the number of informative sites, which may largely result from the variant calling step, especially when applying a series of filtering approaches, thus leading to a less accurate result in downstream analyses. However, sequencing depth may have limited impact in some other cases: The results are not affected by depth when the reference genome is either very closely or very distantly related to the target species. For example, when using H. tibetana as reference genome with H. digitata (the most closely related reference), population structures are distinguishable regardless of sequencing depth (Figure 5a). However, when using H. japonica or R. brunnea as the reference genome (most distantly related), population structures are indistinguishable regardless of the sequencing depth. Therefore, increasing the sequencing depth may not improve the results in such cases where the reference genome is too distantly related.

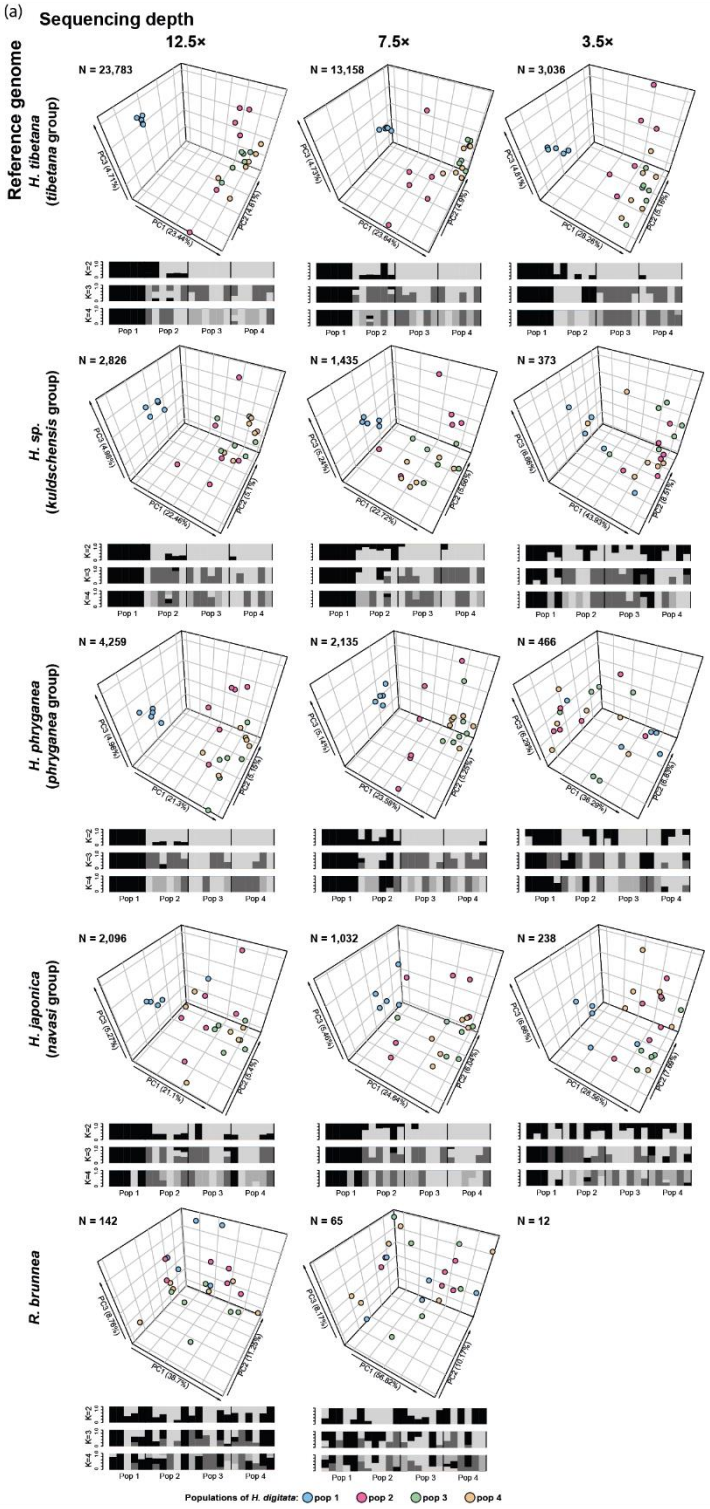
To conclude, both reference genome and sequencing depth have various degrees of influence on downstream analyses, whereas their respective impact is different for each target species. Our results imply that populations with a higher genetic diversity are less affected by the relatedness of the reference genome and the sequencing depth in population structure analyses. In general, the results for H. tibetana appeared more robust despite the variation of reference genome and sequencing depth in comparison to those obtained for H. digitata. This may result from (1) different pairwise relatedness between reference genome and target species and (2) inherent population variation (or expected level of differentiation). Even though H. tibetana and H. digitata are very closely related and both belong to the tibetana group (Hjalmarsson et al., 2019), populations of H. tibetana show a higher level of population variation compared to those of H. digitata, as, for instance, shown by higher F_{ST} value and nucleotide diversity of populations. This is likely caused by different distribution patterns with H. tibetana populations being more isolated at high elevations than H. digitata. A detailed investigation of this was not intended with this study, and our sampling is insufficient and thus inconclusive regarding the underlying biological reasons for the differences in species-specific population structures.

Similar to the trade-off between sequencing depth and sample size demonstrated by previous studies (Buerkle & Gompert, 2013 ; Fumagalli, 2013), we reveal that the roles of reference genome and sequencing depth in a study of population genetics could also be considered as a trade-off. We suggest that population genetic study using genomic data may benefit from applying a more closely related reference genome. Undoubtedly, the optimal option would be a conspecific reference genome even with a low sequencing depth. Our results showed that a conspecific reference genome can significantly improve the accuracy and reliability of all kinds of analyses, in particular WGR-based SNP imputation which will be promising in other genomic investigations like genome-wide association studies. However, if resource limitations exist (in terms of funding, available biological material, time (i.e., wet- and dry lab efforts)), or the research is focused on populations with high interpopulation variation, a trade-off between a more distant reference genome and a higher sequencing depth can be considered. In other words, in a population genetic study, the trade-off between reference genome and sequencing depth is dictated by the focus of research. Although our case study is carried out on caddisflies and thus may not be universally applicable to other organisms, we believe that our results do provide a valuable example that enhances developing roadmaps involved in the choice of appropriate reference genome and sequencing depth in population genomic studies.

4.3 | Concordance between genetic patterns and biogeography of H. digitata and H. tibetana

Irrespective of sequencing depth and reference genome, the results of the population genetic analyses for H. tibetana and H. tibetana are highly consistent with the geographic distribution of populations within the drainage and river networks. For example, in H. tibetana, the PCA plots show that pop1 and pop 2 form two distinct

FIGURE 5 3D scatter plots of all individuals derived from PCA by using PCAngsd and population structure (k = 2, 3, 4, respectively) inferred from NgsAdmix depending on different reference genomes and different sequencing depth of (a) H. digitata and (b) H. tibetana populations. The explained variances are shown as percentages. Numbers on the top left of each plot show the number of SNPs used for the structure estimating. PCA and admixture were not able estimate when using R. brunnea as reference genome with 3.5× depth due to the limited number of variants. Colors in the plots represent the four populations of H. digitata and H. tibetana, respectively.



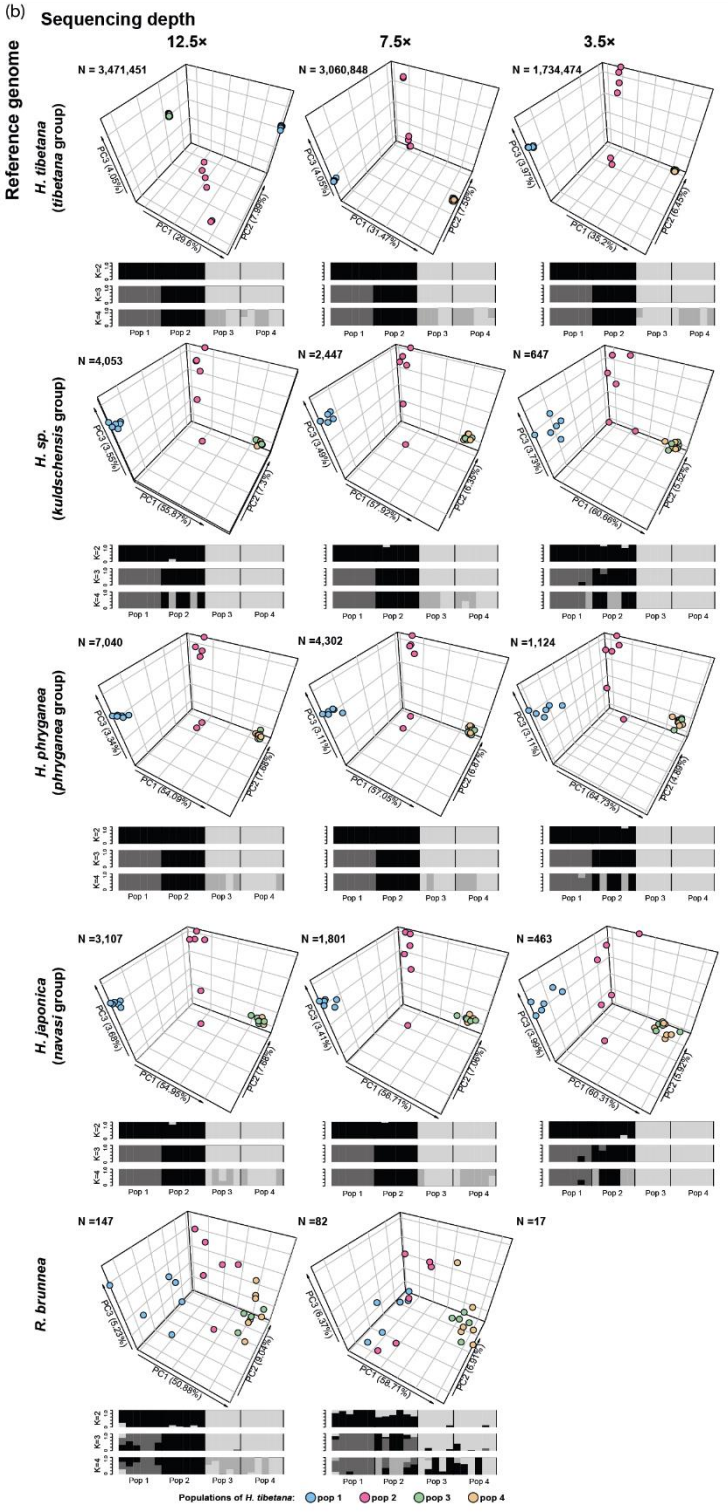


FIGURE 5 (Continued)

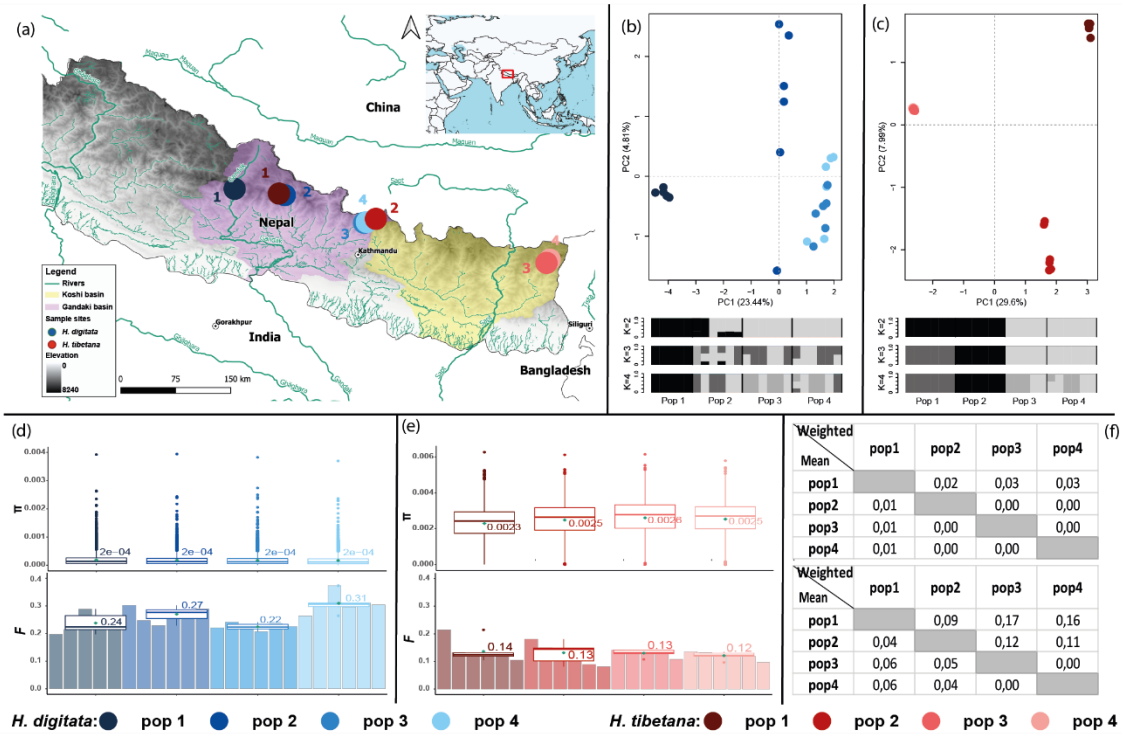


FIGURE 6 Location of populations, genetic diversity, and population structure of H. digitata and H. tibetana. (a) Map showing the eight population sites of the two species across two main drainage basins in Nepal. Population numbers are signed on the sample sites; (b) PCA plots and admixture proportions (k = 2, 3, 4) of H. digitata; (c) PCA plots and admixture proportions (k = 2, 3, 4) of H. tibetana; (d) nucleotide diversity (upper) and inbreeding coefficient (lower) of H. digitata, the mean values of each population were labeled with green dots and numbers; (e) nucleotide diversity (upper) and inbreeding coefficient (lower) of H. tibetana, the mean values of each population were labeled with green dots and numbers; (f) pairwise F_{ST} of H. digitata (upper) and H. tibetana.

clusters, respectively, while pop 3 and pop 4 cluster together and are separated from pop1 and pop 2. This is also consistent with their geographical location: pop 3 and pop 4 are both located in the Kanchenjunga region in far eastern Nepal; pop 2 is located in the most northeastern tributaries of the Gandaki Basin, and pop 1 is located in the east of the Annapurna circuit, which is in the center of the Gandaki Basin. In the admixture plots, pop 1 and 2, as well as pop 3 and 4, share the same structure, respectively. These results are consistent with the basin structure of these populations (pop 1 and pop 2 in Gandaki basin; pop 3 and pop 4 in Koshi basin). Moreover, compared with H. digitata populations, which are all located in one basin, H. tibetana populations show greater genetic diversity and clearer population structure. Due to the dependency of larvae on the aquatic habitat (De Moor & Ivanov, 2007), the limited dispersal capabilities of the adults (Griffith et al., 1998; Petersen et al., 2004), and the unique geographic feature of the Himalayan region, it is not surprising that the population genetics of the two target species show a notable correlation with drainages and river network, like in other Trichoptera studied in the region (Hoppeler et al., 2016) and elsewhere (Altermatt et al., 2013; Engelhardt et al., 2011). The re-sequencing approach, even with low sequencing depth, appears to be a suitable methodological avenue to study population genomics of insect populations when reference genomes of at least moderate relatedness are available.

AUTHOR CONTRIBUTIONS

Xiling Deng: Data curation (equal); formal analysis (equal); methodology (equal); software (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). Paul B. Frandsen: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); resources (equal); software (equal); supervision (equal); validation (equal); writing – review and editing (equal). Rebecca B Dikow: Data curation (equal); methodology (equal); resources (equal); software (equal); writing – review and editing (equal). Adrien Favre: Writing – review and editing (equal). Deep Narayan Shah: Investigation (equal); resources (equal); writing – review and editing (equal). Ram Devi Tachamo Shah: Data curation (equal); resources (equal); writing – review and editing (equal). Julio V. Schneider: Data curation (equal); methodology (equal); resources (equal); writing – review and editing (equal). Jacqueline Heckenhauer: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); resources

(equal); software (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Steffen Pauls:** Conceptualization (equal); data curation (equal); funding acquisition (equal); investigation (equal); project administration (lead); resources (equal); supervision (equal); validation (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

We are thankful to Raphael Coimbra providing for valuable advice in population genetic analyses. We thank Koji Tojo and Robert Wisseman for providing specimens and numbers of Nepalese students for supporting our collection efforts in the field. We thank Subodh Sharma for obtaining the fieldwork permit from Kathmandu University. This research was funded by the Deutsche Forschungsgemeinschaft (DFG), grant number PA 1617/2-1 and PA 1617/2-2 awarded to Steffen Pauls, and AF 1117/1-2 awarded to Adrien Favre. The sequencing of the reference genomes was funded through the LOEWE Center for Translational Biodiversity Genomics with a grant from the Hessen State Ministry of Higher Education, Research and the Arts (HMWK). Analyses were conducted on the BYU Fulton Supercomputer, the TBG Deepfry clusters, and the Smithsonian High-Performance Cluster (SI/HPC), Smithsonian Institution (https://doi.org/10.25572/SIHPC). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

All the COI and CADdata for this research are available in the GenBank of National Center for Biotechnology Information (NCBI), and the accession codes of each individual are provided in Appendix S2. The raw data and genome assemblies of the three novel references have been deposited at NCBI under the Bioproject ID PRJNA728835, and the raw data of populations have been deposited at NCBI under the Bioproject ID PRJNA749154. All BUSCO genes, annotation gffs, and predicted proteins resulting from GEMOMA, blastp, and BLAST2GO results, as well as repeatmodeler and -masker results are available at: https://doi.org/10.6084/m9.figshare.c.6033011.v1 [dataset]. Thetwo reference genome of *H. phryganea* and *R. brunnea* were previ- ously published by Heckenhauer et al. (2022) and has been depos- ited at NCBI under the BioProject ID PRJNA558902.

ORCID

Xi-Ling Deng https://orcid.org/0000-0002-0806-1428
Adrien Favre https://orcid.org/0000-0001-6132-2992
Deep Narayan Shah https://orcid.org/0000-0001-8436-7560

REFERENCES

Altermatt, F., Seymour, M., & Martinez, N. (2013). River network properties shape α-diversity and community similarity patterns of aquatic insect communities across major drainage basins. *Journal of Biogeography*, *40*(12), 2249–2260.

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., Fang, Q., Xie, D., Feng, S., & Stiller, J. (2020). Progressive cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, *587*(7833), 246–251.

Barbosa, S., Mestre, F., White, T. A., Paupério, J., Alves, P. C., & Searle, J. B. (2018). Integrative approaches to guide conservation decisions: Using genomics to define conservation units and functional corridors. *Molecular Ecology*, *27*(17), 3452–3465.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Brandies, P., Peel, E., Hogg, C. J., & Belov, K. (2019). The value of reference genomes in the conservation of threatened species. *Genes*, *10*(11), 846.

Brandt, D. Y., Aguiar, V. R., Bitarello, B. D., Nunes, K., Goudet, J., & Meyer, D. (2015). Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 genomes project phase I data. *G3: Genes*, *Genomes*, *Genetics*, *5*(5), 931–941.

Buerkle, C. A., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: How low should we go? *Molecular Ecology*, *22*(11), 3028–3035.

Crawford, J. E., & Lazzaro, B. P. (2012). Assessing the accuracy and power of population genetic inference from low-pass next-generation sequencing data. *Frontiers in Genetics*, *3*, 66.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., & Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

De Moor, F., & Ivanov, V. (2007). Global diversity of caddisflies (Trichoptera: Insecta) in freshwater. In *Freshwater animal diversity assessment* (pp. 393–407). Springer.

Deng, X. L., Favre, A., Lemmon, E. M., Lemmon, A. R., & Pauls, S. U. (2021). Gene flow and diversification in *Himalopsyche martynovi* species complex (Trichoptera: Rhyacophilidae) in the Hengduan Mountains. *Biology*, *10*(8), 816.

Ding, L., Kapp, P., Cai, F., Garzione, C. N., Xiong, Z., Wang, H., & Wang, C. (2022). Timing and mechanisms of Tibetan plateau uplift. *Nature Reviews Earth & Environment*, *3*, 1–16.

Duchen, P., & Salamin, N. (2021). A cautionary note on the use of genotype callers in Phylogenomics. *Systematic Biology*, *70*(4), 844–854.

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, *29*(1), 51–63.

Engelhardt, C. H., Haase, P., & Pauls, S. U. (2011). From the Western Alps across Central Europe: Postglacial recolonisation of the tufa stream specialist Rhyacophila pubescens (Insecta, Trichoptera). *Frontiers in Zoology*, *8*(1), 1–14.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences USA*, *117*(17), 9451–9457.

Fox, E. A., Wright, A. E., Fumagalli, M., & Vieira, F. G. (2019). ngsLD: Evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics*, *35*(19), 3855–3856.

Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS One*, *8*(11), e79667.

Garcia-Rubio, R., Monzon, S., Alcazar-Fuoli, L., Cuesta, I., & Mellado, E. (2018). Genome-wide comparative analysis of aspergillus fumigatus strains: The reference genome as a matter of concern. *Genes*, *9*(7), 363.

Geist, J. (2011). Integrative freshwater ecology and biodiversity conservation. *Ecological Indicators*, *11*(6), 1507–1516.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333–351.

Gopalakrishnan, S., Castruita, J. A. S., Sinding, M.-H. S., Kuderna, L. F., Räikkönen, J., Petersen, B., Sicheritz-Ponten, T., Larson, G., Orlando, L., & Marques-Bonet, T. (2017). The wolf reference genome sequence (Canis lupus lupus) and its implications for Canis spp. population genomics. *BMC Genomics*, *18*(1), 1–11.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435.

Griffith, M. B., Barrows, E. M., & Perry, S. A. (1998). Lateral dispersal of adult aquatic insects (Plecoptera, Trichoptera) following emergence from headwater streams in forested Appalachian catchments. *Annals of the Entomological Society of America*, *91*(2), 195–201.

Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genetics*, *15*(7), e1008302.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075.

Hall, M. (2022). Rasusa: Randomly subsample sequencing reads to a specified coverage. *Journal of Open Source Software*, *7*(69), 3941. https://doi.org/10.21105/joss.03941

Han, E., Sinsheimer, J. S., & Novembre, J. (2014). Characterizing bias in population genetic inferences from low-coverage sequencing data. *Molecular Biology and Evolution*, *31*(3), 723–735.

Heckenhauer, J., Frandsen, P. B., Gupta, D. K., Paule, J., Prost, S., Schell, T., Schneider, J. V., Stewart, R. J., & Pauls, S. U. (2019). Annotated draft genomes of two caddisfly species Plectrocnemia conspersa CURTIS and Hydropsyche tenuis NAVAS (Insecta: Trichoptera). *Genome Biology and Evolution*, *11*(12), 3445–3451.

Heckenhauer, J., Frandsen, P. B., Sproul, J. S., Li, Z., Paule, J., Larracuente, A. M., Maughan, P. J., Barker, M. S., Schneider, J. V., Stewart, R. J., & Pauls, S. U. (2022). Genome size evolution in the diverse insect order Trichoptera. *GigaScience*, *11*, giac011. https://doi.org/10.1093/gigascience/giac011

Hickey, G., Paten, B., Earl, D., Zerbino, D., & Haussler, D. (2013). HAL: A hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics*, *29*(10), 1341–1342.

Hjalmarsson, A. E. (2019). Delimitation and description of three new species of Himalopsyche (Trichoptera: Rhyacophilidae) from the Hengduan Mountains, China. *Zootaxa*, *4638*(3), 419–441.

Hjalmarsson, A. E., Graf, W., Jähnig, S. C., Vitecek, S., & Pauls, S. U. (2018). Molecular association and morphological characterisa- tion of Himalopsyche larval types (Trichoptera, Rhyacophilidae). *ZooKeys*, *773*, 79–108.

Hjalmarsson, A. E., Graf, W., Vitecek, S., Jähnig, S. C., Cai, Q., Sharma, S., Tong, X., Li, F., Shah, D. N., & Shah, R. D. T. (2019). Molecular phylogeny of Himalopsyche (Trichoptera, Rhyacophilidae). *Systematic Entomology*, *44*(4), 973–984.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522.

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, *30*(1), 62–82.

Hoppeler, F., Tachamo Shah, R. D., Shah, D. N., Jähnig, S. C., Tonkin, J. D., Sharma, S., & Pauls, S. U. (2016). Environmental and spatial characterisation of an unknown fauna using DNA sequencing–an example with Himalayan Hydropsychidae (Insecta: Trichoptera). *Freshwater Biology*, *61*(11), 1905–1920.

Hotaling, S., Kelley, J. L., & Frandsen, P. B. (2020). Aquatic insects are dramatically underrepresented in genomic research. *Insects*, *11*(9), 601.

Hotaling, S., Sproul, J. S., Heckenhauer, J., Powell, A., Larracuente, A. M., Pauls, S. U., Kelley, J. L., & Frandsen, P. B. (2021). Long-reads are revolutionizing 20 years of insect genome sequencing. *Genome Biology and Evolution*, *13*(8), evab138.

Jasper, M., Schmidt, T. L., Ahmad, N. W., Sinkins, S. P., & Hoffmann, A. A. (2019). A genomic approach to inferring kinship reveals limited intergenerational dispersal in the yellow fever mosquito. *Molecular Ecology Resources*, *19*(5), 1254–1264.

Jiang, Y., Jiang, Y., Wang, S., Zhang, Q., & Ding, X. (2019). Optimal sequencing depth design for whole genome re-sequencing in pigs. *BMC Bioinformatics*, *20*(1), 1–12.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Keilwagen, J., Hartung, F., & Grau, J. (2019). GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-seq data. In *Gene prediction* (pp. 161–177). Springer. http://www.jstacs.de/index.php/GeMoMa

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using intron position conservation forhomology-based gene prediction. *Nucleic Acids Research*, *44*(9), e89.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 1–13.

Korneliussen, T. S., Moltke, I., Albrechtsen, A., & Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics*, *14*(1), 1–14.

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*(1287), 1287.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods*, *9*(4), 357–359.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (Version 2). *arXiv preprint arXiv: 1303.3997*. https://doi.org/10.48550/ARXIV.1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, *12*(8), 733–735.

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology*, *30*(23), 5966–5993.

Luo, S., Tang, M., Frandsen, P. B., Stewart, R. J., & Zhou, X. (2018). The genome of an underwater architect, the caddisfly Stenopsyche tienmushanensis Hwang (Insecta: Trichoptera). *Gigascience*, *7*(12), giy143.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, *17*(1), 10–12.

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, *210*(2), 719–731.

Miller, S., Dykes, D., & Polesky, H. (1988). A simple salting out proce- dure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, *16*(3), 1215.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530–1534.

Morse, J. C., Frandsen, P. B., Graf, W., & Thomas, J. A. (2019). Diversity and ecosystem services of Trichoptera. *Insects*, *10*(5), 125.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., & Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS One*, *7*(7), e37558.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, *12*(6), 443–451.

Olsen, L. K., Heckenhauer, J., Sproul, J. S., Dikow, R. B., Gonzalez, V. L., Kweskin, M. P., Taylor, A. M., Wilson, S. B., Stewart, R. J., & Zhou, X. (2021). Draft genome assemblies and annotations of Agrypnia vestita Walker, and Hesperophylax magnus Banks reveal substantial repetitive element expansion in tube case-making caddisflies (Insecta: Trichoptera). *Genome Biology and Evolution*, *13*(3), evab013.

Petersen, I., Masters, Z., Hildrew, A., & Ormerod, S. J. (2004). Dispersal of adult aquatic insects in catchments of differing land use. *Journal of Applied Ecology*, *41*(5), 934–950.

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11*(1), 1–10.

Ríos-Touma, B., Holzenthal, R. W., Rázuri-Gonzales, E., Heckenhauer, J., Pauls, S. U., Storer, C. G., & Frandsen, P. B. (2022). De novo genome assembly and annotation of an Andean caddisfly, Atopsyche davidsoni Sykora, 1991, a model for genome research of high elevation adaptations. *Genome Biology and Evolution*, *14*(1), evab286.

Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assem- blies. *BMC Bioinformatics*, *19*(1), 1–10.

Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, *17*(2), 155–158.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212.

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews Genetics*, *15*(2), 121–132.

Skoglund, P., Ersmark, E., Palkopoulou, E., & Dalén, L. (2015). Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Current Biology*, *25*(11), 1515–1519.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693–702.

Taub, M. A., Bravo, H. C., & Irizarry, R. A. (2010). Overcoming bias and systematic errors in next generation sequencing data. *Genome Medicine*, *2*(12), 1–5.

Thomas, J. A., Frandsen, P. B., Prendini, E., Zhou, X., & Holzenthal, R. W. (2020). A multigene phylogeny and timeline for Trichoptera (Insecta). *Systematic Entomology*, *45*(3), 670–686.

Tsuruishi, T., Ketavan, C., Suwan, K., & Sirikajornjaru, W. (2006). Importance of water flow on larval growth and pupation of Himalopsyche acharai, (Malicky and Chantaramongkol, 1989) (Trichoptera: Rhyacophilidae). *Hydrobiologia*, *563*(1), 537–540.

Valiente-Mullor, C., Beamud, B., Ansari, I., Francés-Cuesta, C., García-González, N., Mejía, L., Ruiz-Hueso, P., & González-Candelas, F. (2021). One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLoS Computational Biology*, *17*(1), e1008678.

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746.

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204.

Waldvogel, A. M., Wieser, A., Schell, T., Patel, S., Schmidt, H., Hankeln, T., Feldmeyer, B., & Pfenninger, M. (2018). The genomic footprint of climate adaptation in Chironomus riparius. *Molecular Ecology*, *27*(6), 1439–1456.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., & Young, S. K. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, *9*(11), e112963.

Wang, G.-D., Zhai, W., Yang, H.-C., Fan, R.-X., Cao, X., Zhong, L., Wang, L., Liu, F., Wu, H., & Cheng, L.-G. (2013). The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nature Communications*, *4*(1), 1–9.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene predic- tion and phylogenomics. *Molecular Biology and Evolution*, *35*(3), 543–548.

Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Xu, J., Grumbine, R. E., Shrestha, A., Eriksson, M., Yang, X., Wang, Y., & Wilkes, A. (2009). The melting Himalayas: Cascading effects of climate change on water, biodiversity, and livelihoods. *Conservation Biology*, *23*(3), 520–530.

Yang, X., Lee, W.-P., Ye, K., & Lee, C. (2019). One reference genome is not enough. *Genome Biology*, *20*(1), 1–3.

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(6), 15–30.

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*(21), 2669–2677.

Zimin, A. V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marçais, G., Yorke, J. A., Dvořák, J., & Salzberg, S. L. (2017). Hybrid assembly of the large and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Research*, *27*(5), 787–792.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Deng, X.-L., Frandsen, P. B., Dikow, R. B., Favre, A., Shah, D. N., Shah, R. D. T., Schneider, J. V., Heckenhauer, J., & Pauls, S. U. (2022). The impact of sequencing depth and relatedness of the reference genome in population genomic studies: A case study with two caddisfly species (Trichoptera, Rhyacophilidae, *Himalopsyche*). *Ecology and Evolution*, *12*, e9583. https://doi.org/10.1002/ece3.9583

Chapter 3

Comparative phylogeography of *Himalopsyche* (Trichoptera, Rhyacophilidae) in the Tibeto-Himalayan Region: An assessment of the mountain-geobiodiversity hypothesis

Xi-Ling Deng, Sami Domisch, Adrien Favre, Sonja C. Jähnig, Paul B. Frandsen, Fengzhi He, Deep Narayan Shah, Ram Devi Tachamo Shah, Qinghua Cai, Steffen U. Pauls

This manuscript is currently under review.

SUP, SD, and SCJ designed the study and acquired funding. SUP, PBF, DNS, RDTS, QC, FH and I collected sample material from the field. I performed the laboratory work. I performed the population genetic analyzes. Species distribution modeling was reconstructed by SD and me. I submitted the repositories. SUP, SD, and I interpreted the results. I wrote the manuscript with input from all other authors.

Comparative phylogeography of *Himalopsyche* (Trichoptera, Rhyacophilidae) in the Tibeto-Himalayan Region: An assessment of the mountain-geobiodiversity hypothesis

Phylogeography of caddisflies in the THR

Xi-Ling Deng[1,2,3,4,*], Sami Domisch[4], Adrien Favre[5], Sonja C. Jähnig[4,6], Paul B. Frandsen[3,7,8], Fengzhi He[4,6,9], Deep Narayan Shah[10], Ram Devi Tachamo Shah[11,12], Qinghua Cai[13], Steffen U. Pauls[1,2,3]

[1] Senckenberg Research Institute and Natural History Museum, Frankfurt am Main, Germany

[2] Institute of Insect Biotechnology, Justus-Liebig-University Gießen, Gießen, Germany

[3] LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt am Main, Germany

[4] Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany

[5] Regional nature park of the Trient Valley, Salvan, Switzerland

[6] Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany

[7] Department of Plant & Wildlife Sciences, Brigham Young University, Provo, USA

[8] Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, USA

[9] Center for Biodiversity Dynamics in a Changing World (BIOCHANGE) and Section for Ecoinformatics and Biodiversity, Department of Biology, Aarhus University, Aarhus C, Denmark

[10] Central Department of Environmental Science, Tribhuvan University, Kirtipur, Nepal

[11] Aquatic Ecology Centre, School of Science, Kathmandu University, Dhulikhel, Nepal

[12] Department of Life Sciences, School of Science, Kathmandu University, Dhulikhel, Nepal

[13] State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

* Corresponding author

Abstract

The Tibeto-Himalayan Region is famous for its geography, climatic influence, and exceptional and immense biodiversity. The "mountain-geobiodiversity hypothesis (MGH)" explores the interaction of topography, climate, and biology in the evolution of mountain biodiversity. We tested this hypothesis in the Himalayas and the Hengduan Mountains on a group of caddisflies that are endemic to this region. We investigated one caddisfly species pair from each mountain respectively, each pair containing one species inhabiting high elevation and one inhabiting low elevation. We incorporated genomic and ecological evidence to reveal population structure, demographic history, and potential habitat range dating back to the last glacial maximum (LGM) of each species. The results indicated that in both mountains, the high-elevation species showed strong local differentiation, while the low-elevation species were shaped by hydro-morphology indicating greater regional dispersal activity. Results of demographic history and species distribution modelling supported demographic expansions for all species during the LGM linked to an increase in potential habitats. Caddisfly species in the Himalayas generally exhibited an East-West oriented dispersal. Species from the Hengduan Mountains showed greater connectivity on the North-South orientation, suggesting that species have a higher chance to survive in the Hengduan Mountains by both *in-situ* displacement (along the elevational gradients) and long-distance dispersal (along the latitudinal gradients) during glaciation. Our study demonstrates that historical geodiversity and climate fluctuations interact and influence the diversification of caddisflies in the Tibeto-Himalayan Region, thus supporting the MGH.

Keywords

caddisfly, population genomics, demographic history, species distribution modelling, Tibeto-Himalayan Region, mountain biodiversity

1 Introduction

Mountains exhibit extraordinarily heterogeneous environments and host a remarkable diversity of (endemic) terrestrial and aquatic species (Rahbek et al., 2019; Rahbek et al., 2019; Perrigo et al., 2020). Since Alexander von Humboldt initiated the principle of *Cosmos* ("unity of nature") that combines geology and biology to explain the distribution patterns of life (Von Humboldt, 1860), researchers have attempted to explore how geophysical modifications over time, such as orogeny and climate change, could have influenced biological processes involved in speciation and diversification (Ding et al., 2020). To better understand the origin and evolution of biodiversity in mountains, we investigate four caddisfly species living in the highest and largest mountain system in the world: the Tibeto-Himalayan Region.

The Himalayas and the adjacent Hengduan Mountains (HM) have drawn increasing interest from biogeographers and ecologists (Favre et al., 2015; Hoorn et al., 2018; Muellner-Riehl, 2019; Rahbek et al., 2019), especially after Myers et al. (2000) classified these mountain systems as two of the global biodiversity hotspots. In recent years, biogeographic studies on the Tibeto-Himalayan Region have revealed that the evolution of species was profoundly shaped by changes in geomorphology and climate over millions of years (e.g., Favre et al., 2015; Xing & Ree, 2017; Mosbrugger et al., 2018; Muellner-Riehl et al., 2019; Rana et al., 2019; Ding et al., 2020; Rana et al., 2022). For instance, Ding et al. (2020) revealed that *in situ* speciation, diversification, and colonization in the alpine flora inhabiting the HM, the Himalayas, and the Qinghai-Tibetan Plateau were jointly driven by mountain uplift and intensification of the Asian monsoon system. Nonetheless, different topographic relief, orogenic activity, and climate history in the Himalayas and the HM led to distinct biodiversity patterns (Ding et al., 2020). For example, species richness increases from west to east in birds (Price et al., 2011), plants (Yan et al., 2013; Bhattarai et al., 2014; Rana et al., 2019), and mammals (Srinivasan et al., 2014). This parallels the increase of precipitation towards the east in the Himalayas. In contrast, a North-South floristic divide was revealed in the HM, which may be associated with both climate (separation by the line of regular freezing) and topography (divided by the Jinsha River, Li et al., 2021). Moreover, patterns of biodiversity in the HM appear to have been more dynamic through time: diversity hotspots of montane plants have shifted from the southeastern to the central and western parts of the HM between the last glacial maximum (LGM) to today (Liang et al., 2018). The phylogeographic history of the HM is further complicated by the presence of geographically extensive and long-lasting barriers to dispersal such as the deeply incised valleys of the Irrawaddy, the Salween, the Mekong and the Yangtze rivers, which have been shown to be instrumental in delineating floristic motifs in the region (Li et al., 2021; Muellner-Riehl & Favre, 2021). Therefore, biogeographers and ecologists have increasingly viewed evolutionary processes in these two mountain systems as relatively independent despite their biogeographical interconnection (e.g., Ding et al., 2020).

The "mountain-geobiodiversity hypothesis (MGH)" conceptualizes the link between geophysical changes and the origin and evolution of biodiversity based on the Tibeto-Himalayan Region (Mosbrugger et al. 2018). In this hypothesis three boundary conditions are deemed essential to the

accumulation of biodiversity in mountains: (i) full elevational zonation with lowland, montane, and alpine zones; (ii) the occurrence of a species-pump driven by climatic fluctuation; and (iii) strong environmental gradients. Within this conceptual framework, mountain uplift provides elevational gradients and locally diverse topography. This condition increases opportunities for local or regional taxa to adapt to a high variety of niches (i), and fosters a higher resistance to climate change via vertical displacement (iii). Meanwhile, during climate fluctuations, for instance in the Quaternary, diversification is fostered by a species-pump effect (ii) via cyclical range fragmentation (causing divergence) and secondary contacts (involving hybridization or reinforcement) (Mosbrugger et al., 2018; Muellner-Riehl, 2019). This hypothesis was partially verified on a few taxa from the HM (e.g., Fu et al., 2020, 2022; Wang et al., 2022), while some global-scale meta-analyzes also support it (Muellner-Riehl et al., 2019). However, case studies have so far been limited to plant taxa, such that the validation and refining of the hypothesis for a broader taxonomic spectrum is missing. Following Favre et al. (2015), which provided a generalized overview of the origin and evolution of mountain biodiversity, we investigate the MGH in the context of the diversification of species of the aquatic insect genus *Himalopsyche* (Trichoptera, Rhyacophilidae).

Himalopsyche is a genus of aquatic caddisflies that inhabit mountains. Most *Himalopsyche* species are distributed in Central and East Asia (Hjalmarsson et al., 2019) except for the Nearctic *H. phryganea* (Ross, 1941). Like all caddisflies, species of *Himalopsyche* have a merolimnic life cycle and are considered important bioindicators (Resh & Unzicker, 1975; Tsuruishi et al., 2006; Hjalmarsson, 2019; Morse et al. 2019). The larvae of this taxon generally inhabit turbulent, fast-flowing rivers and streams where they live as ferocious predators (Hjalmarsson et al., 2018). Currently, there are 56 described species in the genus (Hjalmarsson et al., 2019), with 23 occurring in the Himalayas and 34 occurring in the HM (some of them distributed in both). The center of diversity of the genus *Himalopsyche* is located in these two mountain regions. But regionally, different species exhibit strongly differentiated niches usually associated with elevational gradients (Schmid & Botosaneanu, 1966). For example, some species inhabit lower elevations between 1500–2500 m.a.s.l., as in the case of *H. digitata* (Martynov, 1935, in the Himalayas) or *H. platon* (Malicky, 2011, in the HM), whereas other species prefer higher elevation ranges between 2000–4500 m.a.s.l., such as *H. tibetana* (Martynov, 1930, in the Himalayas) and *H. gregoryi* (Ulmer, 1932, in the HM; summarized in Hjalmarsson 2020). As reported by Lehrian et al. (2009), montane caddisflies that inhabit different elevation ranges but have similar geographic distributions may exhibit distinct population structures putatively associated with varying dispersal capabilities, habitat specificity or differences in phylogeographic history. Because species of *Himalopsyche* inhabit different elevations, they are a good model for investigating how geography has shaped their genetic diversity over time.

For aquatic species, distribution patterns and dispersal among habitats are constrained by the dendritic structure of the stream network (Tonkin et al., 2018). To assess and interpret differing patterns of population structure in aquatic insects, Finn et al. (2007) and Hughes et al. (2013) proposed process-

based models of population genetic diversity patterns that account for the structure of drainage systems: the stream hierarchy model, the death valley model, the headwater model, isolation by distance and panmixia (or also called as the widespread gene flow model, Hughes, 2007). These models are assigned to a given species primarily by their population connectivity, defined by the level of gene flow among populations throughout the drainage networks. The population connectivity generally depends on (1) traits that determine dispersal ability (dispersal ability and behaviour, life cycle, oviposition, and the spatial distribution of source populations) (Smith & Smith, 2009; Parkyn & Smith, 2011); (2) the distance between populations; (3) the suitability of the new habitat; and (4) the permeability of the landscape (Rader et al., 2019). In caddisflies, dispersal is often an "along-stream" movement as defined in the context of the "colonization cycle" (Müller, 1954; Collier & Smith, 1997; Petersen et al., 2004; Winterbourn et al., 2007). Current-induced downstream movements often occur in the larval stage. Part of this downstream movement is then compensated by adult females (and also males) flying upstream prior to mating or egg-laying. However, adults also fly perpendicular to the stream (lateral dispersal) and thus disperse overland between catchments allowing gene flow between caddisfly populations from different catchments (Svensson, 1974; Malicky, 1987; Collier & Smith, 1997; Griffith et al., 1998; Bowler & Benton, 2005; Wilcock et al., 2007; Smith & Smith, 2009; Engelhardt et al., 2011; Müller-Peddinghaus, 2011; Deng et al., 2021). A genomic study on a species complex of the *Himalopsyche* revealed that continuous gene flow can be maintained over millions of years between two basins (Deng et al., 2021). Hence, in addition to elevational zonation (i) and environmental gradients (iii) as proposed in the MGH, and because of the unique dispersal ability of caddisflies, drainage systems may also be crucial topographical features that drive distribution ranges of caddisflies, e.g. through the species-pump effect (ii).

To examine these patterns across both the Himalayas and HM, we studied four *Himalopsyche* species from the two mountains. For each mountain range, we chose a species that inhabits high elevation streams, *H. tibetana* (Himalayas) and *H. gregoryi* (HM), and another species that inhabits lower elevations, *H. digitata* (Himalayas) and *H. platon* (HM). We combined population genomics analysis of 333 individuals across the four species with species distribution model (SDM) to (1) reveal the genetic pattern of species inhabiting high-elevation versus low-elevation, and species from the Himalayas versus the HM, (2) evaluate the role of environmental factors including climate change, mountain topography and drainage rearrangement on the evolution of aquatic biodiversity in the Tibeto-Himalayan Region, and (3) assess the implication of the MGH in different mountain systems such as the Himalayas and the HM.

2 Materials and Methods

2.1 Sample collection and DNA sequencing

We analyzed 333 individuals across the four *Himalopsyche* species. Specimens were collected from 2010 to 2020 in the Himalayas (*H. tibetana*: N_{ind} (number of individuals) = 97, N_{pop} (number of

populations) = 28; *H. digitata*: N_{ind} = 105; N_{pop} = 23) and the HM (*H. gregoryi*: N_{ind} = 71, N_{pop} = 18; *H. platon*: N_{ind} = 60; N_{pop} = 12, Fig. 1). All specimens were preserved in 95% ethanol and archived in the collections of the Senckenberg Research Institute and Natural History Museum (Supplementary Table 1).

We extracted DNA from abdominal and/or thoracic tissue after removing the intestinal tract following Miller et al. (1988). We purified low-quality DNA samples with magnetic beads as described by Deng et al. (2022). After quality check, all DNA samples were sent to Novogene Co., Ltd. (Hongkong, China) for DNA library preparation and 2×150 bp whole genome resequencing with a coverage of 12× on an Illumina HiSeq 2000 platform.

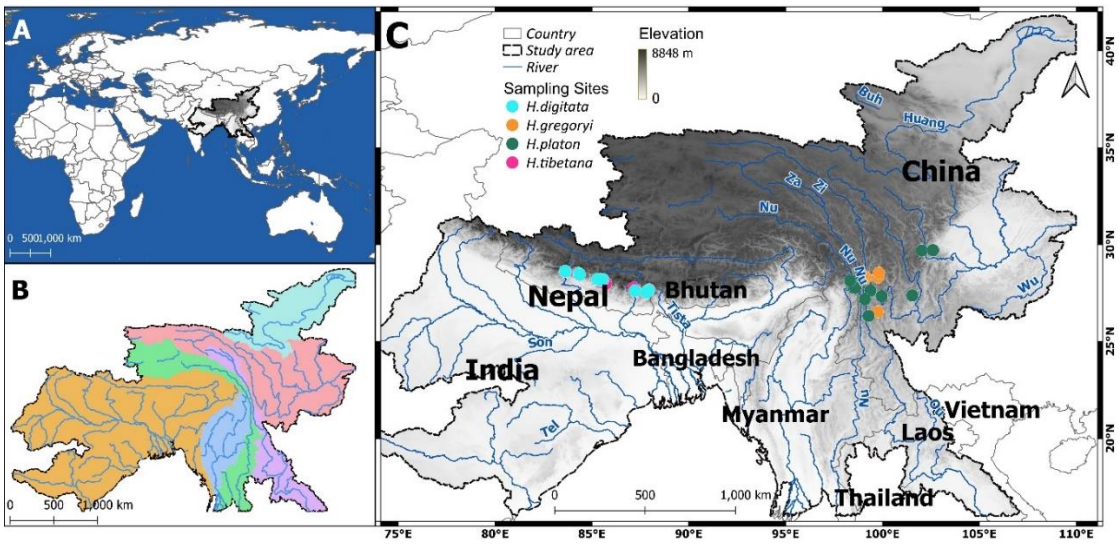


Fig. 1 Study area and sampling sites in this study. The study area of this research was limited to the main river basins of the species distribution based on previous knowledge, more details are described in section 2.11. A: overview of the study area; B: primary basins in the study area. C: sampling sites of the four target species, for more detail see Fig. 3.

2.2 Data preparation

We assessed the quality of Illumina reads using FastQC v0.11.8 (Babraham Institute) and MultiQC v1.7 (Ewels et al., 2016). Following Deng et al. (2022), we trimmed overrepresented *k-mers* using autotrim.pl v0.6.1 (Waldvogel et al., 2018) with Trimmomatic v0.38 (Bolger et al., 2014) and removed adapters with Cutadapt v2.23 (Martin, 2011). After cleaning, the average sequence coverage was 14× (see Fig. S1 for details). We mapped the reads to a published *de novo* genome of *H. tibetana* (Deng et al., 2022, NCBI accession number: JAHFWH000000000) using Bowtie2 (Fig. S2). Duplicate reads were marked and indexed using Picard v2.20.8 (Picard Tools – By Broad Institute) and SAMtools (v1.1, Danecek et al., 2021).

We used ANGSD (v0.931, Korneliussen et al., 2014) to estimate genotype likelihoods and GATK (v4.1.7.0, Danecek et al., 2011) for genotype calling (see Supplementary material 1, section 4 for details and parameter settings). Since the genotype likelihood algorithm in ANGSD is designed for low to medium-depth data (Korneliussen et al., 2014), we used it for all of the population genomic inferences

except for the TreeMix analysis since the program cooperated better with GATK for calling the SNPs without missing data (Pickrell & Pritchard, 2012). With the GATK pipeline, we obtained approximately 4,643,429; 68,261; 221,016; 245,170 SNPs for *H. tibetana*, *H. digitata*, *H. gregoryi*, and *H. platon* respectively. The genotype likelihood estimation with ANGSD produced 1,668,494; 6,756; 88,947: 12,842 high-quality variants for *H. tibetana*, *H. digitata*, *H. gregoryi*, and *H. platon*, respectively. The quality of the genotype was shown in Fig. S3, S4, and S5.

2.3 Genetic structure

Before performing the principal component analysis (PCA), we pruned linked sites within our SNP dataset estimated by genotype likelihoods using ngsLD (Fox et al., 2019). We calculated a covariance matrix for PCA using PCAngsd (Meisner & Albrechtsen, 2018) and admixture proportions using NGSadmix (Skotte et al., 2013) based on the linkage pruned genotype likelihoods. The optimal value of ancestral components (K) was determined by CLUMPAK (Kopelman et al., 2015).

We measured the genome-wide nuclear heterozygosity (π) of each sample based on a site frequency spectrum (SFS) following the workflow suggested by the authors of ANGSD v0.931 (Korneliussen et al., 2014). To quantify the alleles inherited from common ancestors in an individual's lineage, we estimated individual inbreeding coefficients (*F*) based on the filtered genotype likelihoods (same filter as used for PCA analyzes) using ngsF v1.2.0 (Vieira et al., 2013) following de Jager et al. (2021).

To better understand the relationships among populations of each species, we reconstructed a maximum-likelihood tree with migration events among populations of each species respectively using TreeMix version 1.13 (Pickrell & Pritchard, 2012) following Dahms et al. (2022) and software guidelines. We additionally calculated a median-joining (MJ) network of each species based on the COI sequences using Network 10 (Fluxus-engineering.com, Bandelt et al., 1999).

To better characterize the genetic structure, we assigned the populations to subbasins delineated by Lehner and Grill (2013) at level 10 (http://www.hydrosheds.org/page/hydroatlas). The four-digit code of each subbasin used in this study is replicated from the source layer; additionally, we use descriptive names for the subbasins (see Table S9).

2.4 Historic effective population size

To gain insights into the history of population-size change, especially during the LGM, we inferred the demographic history of the four target species by applying the pairwise sequentially Markovian coalescent (PSMC) model to each genome (Li & Durbin, 2011). We prepared the input files for PSMC and ran the analyzes following the software documentation (more details see Supplementary material 1, section 8). For the final visualization, we plotted the inverse instantaneous coalescent rate (IICR) as estimated by PSMC following Humble et al. (2020) and de Jager et al. (2021) in R. Notably, for the plot we used a generation time of 1 year (Tsuruishi, 2003, 2006) and a per generation mutation rate for

caddisflies of 2.9e−9 as estimated for a Lepidoptera species (Keightley et al., 2015), though the mutation rate was proposed to be consistent with most of the insects (Liu et al., 2017).

2.5 Ecological variables statistics and species distribution modelling (SDM)

We collected *in-situ* environmental variables and GIS resources at each sampling site (all variables see Table S1, S2). We conducted a PCA for each species pair from the Himalayas and the HM respectively, using all environmental variables as mentioned above (Fig. S6, S7).

To better understand the influence of different climatic conditions on *Himalopsyche*, we compared the potential suitable area of each species between the LGM and the present day by applying SDMs. We used all known locations of each species from our previous work and the current study as validated occurrence points in the models (Supplementary Table 1). According to the current distribution ranges of the species in references and field observation, we defined the extent of our study area as the main drainage basins of all four species' occurrences (Fig. 1, details see Supplementary material 1, section 10). Considering the high dependence of caddisflies on water generally and flight ability at the adult stage, we attempted to include hydrological connectivity in the modelling. Thus, we applied a sub-sampled layer that was calculated from the DEM with a stream-initialization threshold of 200 upstream 90 m grid cells, resulting in 2771220 sub-catchments and approximately 10 km^2 for each sub-catchment.

We downloaded topographic variables from the EarthEnv database and the bioclimatic variables from the CHELSA database (hereafter 1 km, Table S3). All variables were chosen based on our supposition of their importance for the distribution of caddisflies (Graf et al., 2008; Morse et al., 2019). We used seven bioclimatic datasets (CCSM4, MRI, CNRM, FGOALS, CM5A, MIROC, and MPI download date: 10.12.2019) provided by CHELSA for the LGM. All variable layers were cropped to our study area first and buffered approximately 500 km around the study area to dissolve and avoid errors caused by the coast. Then we calculated the univariate statistics of each environmental layer that was assigned to the sub-sampled layer as described above. Before building the SDMs, we removed highly correlated variables (correlation coefficient 'r' cutoff = 0.75). The same variables were used for all species at different time scenarios.

We applied the ensemble method implemented in the R package biomod2 (Thuiller et al., 2009) to construct the SDM for each species at different time scenarios. This method enables combining the predictive outputs of different algorithms, thus improving the accuracy of the modelling (Hao et al., 2019; Stone & Wolfe, 2021). The details of the setting were described in Supplementary material 1, section 10. To compare the distribution range of each species between the LGM and at present, we obtained a consensus binary map for the LGM prediction by using the committee averaging method (Araujo & New, 2007; Gallien et al., 2012; Gillard et al., 2017).

The final results of the range size change from the LGM to the present for each species were deposited on the IGB-GeoNode as raster layers (https://fred.igb-berlin.de/data/package/812). All intermediate

results of the predictions using different models on different time scenarios were presented in supplementary material 1 (Table S4–S8, Fig. S8–S21).

3 Results

3.1 Population-level genomic diversity

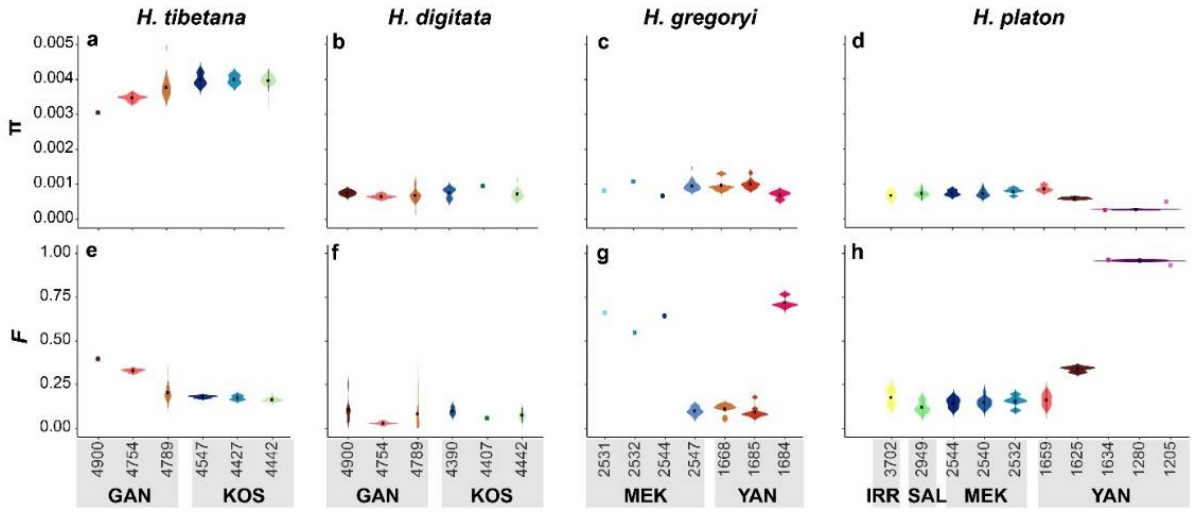


Fig. 2 Population genetic diversity (genome-wide heterozygosity (π, a–d) and inbreeding coefficient (*F*, e–h)) of the four target species. The x-axes show the value of each index that is associated with the primary basins (GAN: Gandaki Basin; KOS: Koshi Basin; MEK: Mekong Basin; YAN: Yangtze Basin; IRR: Irrawaddy Basin; SAL: Salween Basin) and subbasins (Lehner and Grill 2013, http://www.hydrosheds.org/page/hydroatlas). The colours of the box plots are consistent with the colours shown in the map in Fig. 3.

In the Himalayas, the high-elevation *H. tibetana* showed genome-wide heterozygosity between 0.03 to 0.04 (Fig. 2a) with populations in the Koshi Basin exhibiting higher heterozygosity compared with those from the Gandaki Basin. We generally observed increasing heterozygosity from the western to the eastern Himalayan populations, while the *F* decreased along the same geography. Populations of the low-elevation *H. digitata* exhibited a comparatively lower level of genome-wide heterozygosity and *F* (π = ~0.01, *F* = ~0.13). Moreover, there was no significant difference among populations from different basins.

Both HM species, *H. gregoryi* (high-elevation) and *H. platon* (low-elevation) showed similarly low levels of heterozygosity (π = ~0.001). However, populations from some of the HM basins exhibited an extremely high *F* while the others had a low *F*. For instance, the *F* of most populations of *H. platon* were around 0.13, but the populations from the three basins located in the remote area of the Yangtze basin (ID 1634, 1280, 1205, Fig. 3) showed *F* approaching 1.0 (Fig. 2h). Moreover, the heterozygosity of these populations was relatively lower than that of the remaining populations. Similar to *H. platon*, populations of *H. gregoryi* also showed fluctuating levels of *F* across basins: populations from some of the subbasins of Mekong (ID 2531, 2532, 2544) had a high level of *F* (0.5–0.7), as did populations from one subbasin of the Yangtze (ID 1684). In particular, these basins were geographically isolated, at least at the level of subbasins.

3.2 Population structure

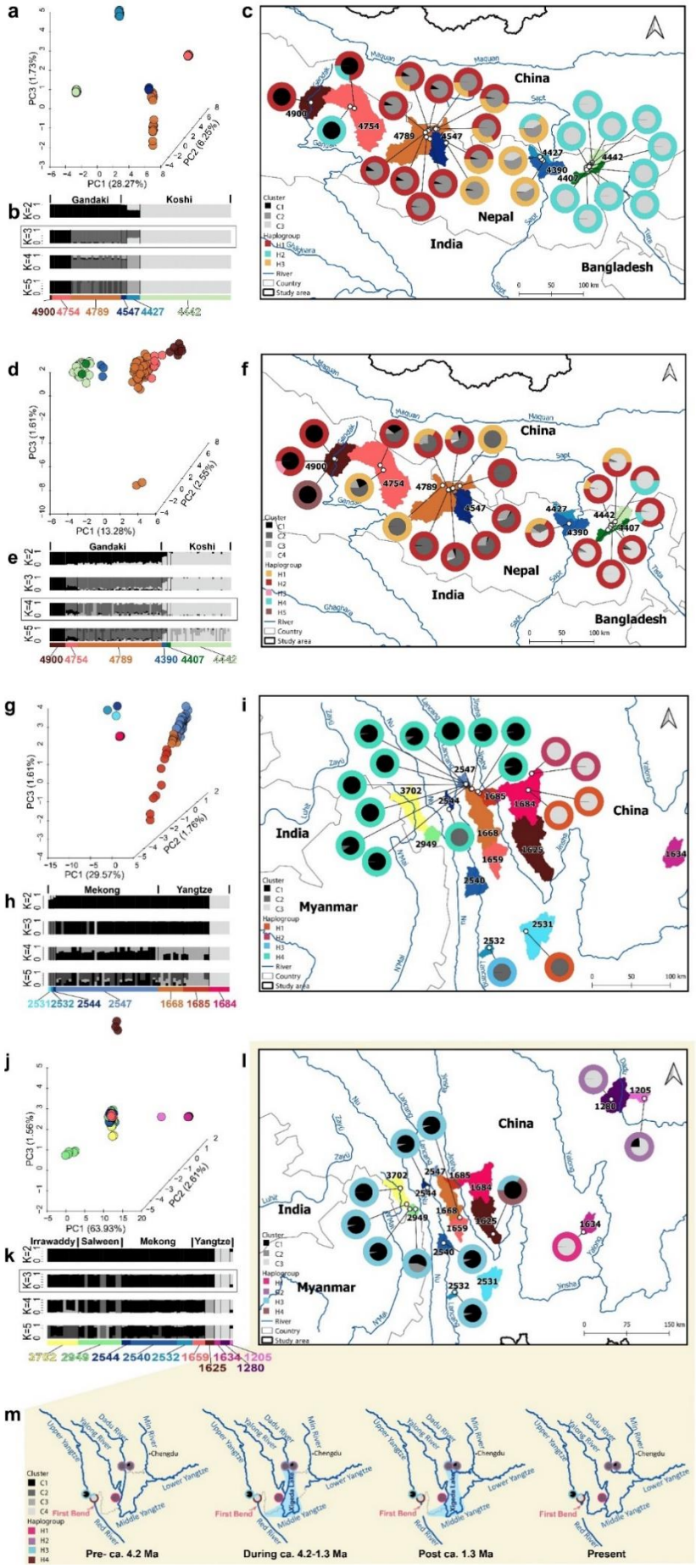


Fig. 3 Population structure of *H. tibetana* (a, b, c), *H. digitata* (d, e, f), *H. gregoryi* (g, h, i) and *H. platon* (j, k, l, m), including results of PCA (a, d, g, j) and admixture (b, e, h, k) analyses based on genomic data. Best fit K in the admixture estimates was indicated with black boxes except for *H. gregoryi*, for which the methods outlined in Evanno et al. (2005) did not lead to conclusive results. The colour and the code represented the minimal basins that the sample was assigned to. (c, f, i, l) Sampling sites of the four target species in this study. The inner pie chart indicates the distribution of inferred admixture clusters with best fit K (except for *H. gregoryi*, for which K3 was selected) of each species at the given sites; Outer ring pie charts indicate the distribution of haplogroups calculated by the MJ network based on mtCOI (more details see Fig. S22). Colourful patches on the map show the minimal basins as categorized in this study. (m) Paleo-drainage rearrangement of the Yangtze Basin and sampling location of *H. platon* in this area. The inner pie chart shows the distribution of inferred admixture clusters with K4. Outer ring pie charts indicated the distribution of haplogroups calculated by the MJ network based on mtCOI. The Paleo-drainage maps were reproduced from (Deng et al. 2021).

Interestingly, the PCA using the *in situ* and GIS environmental variables successfully separated *H. gregoryi* and *H. platon* in the HM (Fig. S6 and S7), which indicated a different microhabitat of the two species. However, *H. tibetana* and *H. digitata* in the Himalayas were not sufficiently separated from each other in the PCA, suggesting that the two species might inhabit a similar microhabitat.

We found four distinct clusters in the PCA plot for the high-elevation *H. tibetana* (Fig. 3a, b, c): (i) West and Middle Gandaki (subbasin ID 4900 and 4754); (ii) East Gandaki (ID 4789) and West Koshi (ID 4547); (iii) Middle Koshi (ID 4427); (iv) East Koshi (ID 4442). The MJ network of mtCOI sequence data, also supported four distinct clades, except that several individuals of cluster i shared mtDNA haplotypes with cluster iv, which may indicate a preservation of these old haplotypes at Middle Gandaki (ID 4754) or an invasion of individuals from East Koshi (ID 4442, Fig. S22, Fig. 3c). This same result was supported by the maximum likelihood estimation of individual ancestries using NGSadmix with the best K (K = 3), and with the taxonomic units delineated by TreeMix (Fig. S23). Analyzes of PCA, admixture (with the best K = 4) and TreeMix revealed congruent structuring patterns for populations of *H. digitata*: populations were grouped into 5 clusters representing populations from (i) West Gandaki (ID 4900); (ii) Middle Gandaki (ID 4754); (iii) East Gandaki (ID 4789); (iv) Middle Koshi (ID 4390); (v) East Koshi (ID 4407 and 4442, Fig. 3d, e, f, Fig. S23). In addition, the MJ network of the mtCOI sequence data showed evidence of secondary contact between populations from cluster i and populations from other groups (Fig. S22).

When compared with the two species distributed in the Himalayan region, populations of *H. gregoryi* and *H. platon* showed more complex structures. Three distinct groups were recognized within the samples of *H. gregoryi* from (i) Mekong (including subbasin ID 2531, 2532 and 2544); (ii) West Yangtze (ID 1684); (iii) East Mekong and West Yangtze (ID 2547, 1668 and 1685, Fig. 3g, h, i). Although the optimal K could not be determined following the methods outlined in Evanno et al. (2005), we found congruence between the output of admixture and PCA: the samples from group iii formed a cluster in the PCA that simultaneously exhibits a gradient among populations of different elevations and subbasins. This cluster was assigned to an admixed group and one clade in the TreeMix but split into multiple lineages (Fig. S23). It was notable that samples within group iii were located in the border region of the three subbasins located in the same mountain area. The population structure of *H. gregoryi* estimated by the genomic data was partially congruent with the results of the MJ network: samples of group iii were assigned to the same haplogroup, yet, samples of groups i and ii were segregated into four haplogroups (Fig. 3i, Fig. S22). The 12 populations of low-elevation species *H. platon* were grouped into four distinct clusters: (i) Middle and East Yangtze (ID 1634, 1280 and 1205); (ii) West Yangtze (ID 1625); (iii) Salween (one population from subbasin ID 2949); (iv) populations from remaining basins (including Irrawaddy, Salween, Mekong and Yangtze, Fig. 3j, k, l). Similar patterns were also observed in the mtCOI MJ network, except that one individual from West Yangtze (ID 1625) fused with the near-panmictic haplogroup (iv), indicating a secondary contact (Fig. S22).

3.3 Demographic history and SDM

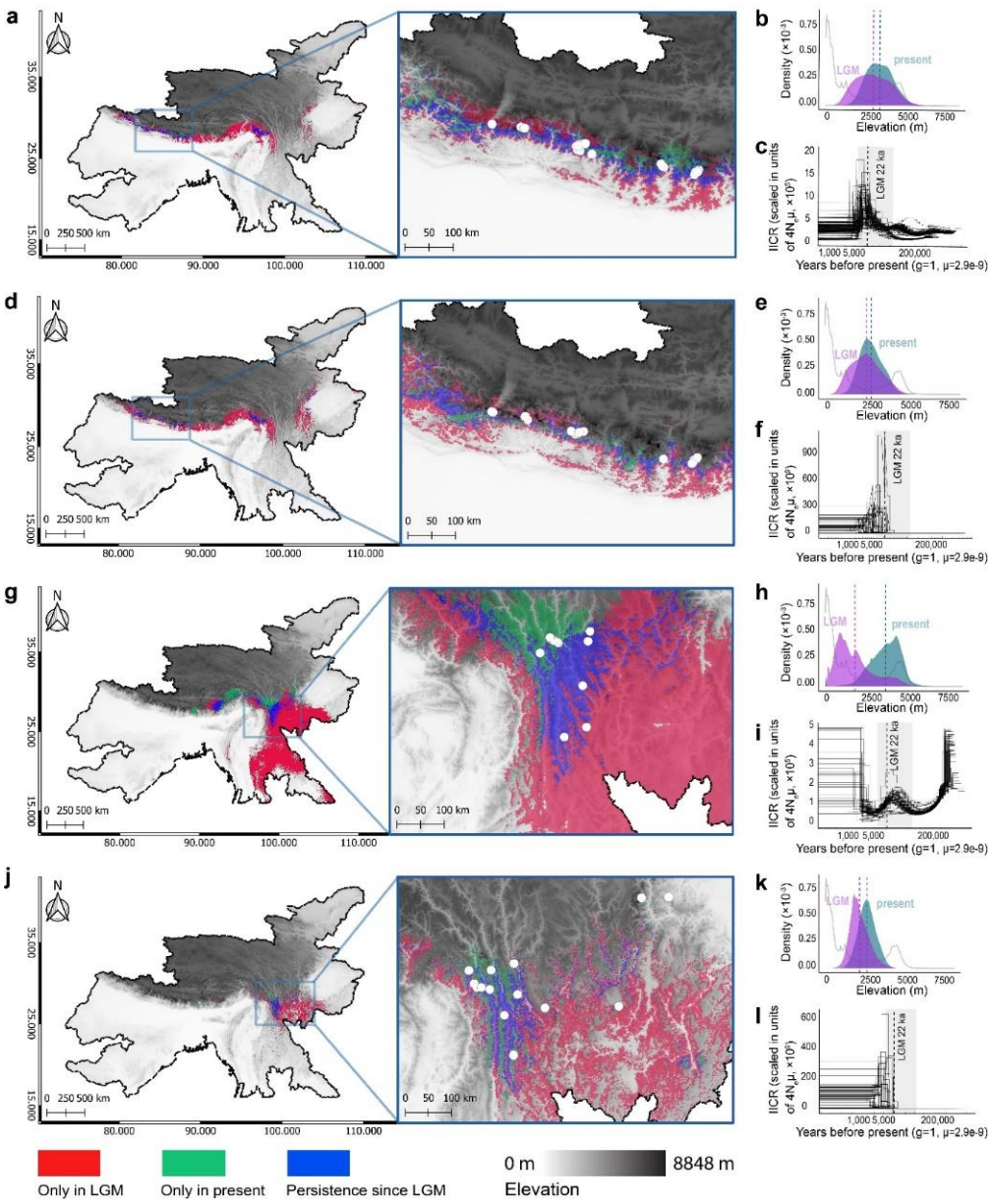


Fig. 4 Changes in distribution range from the LGM (minority prediction) to present day (a, d, g, j), predicted elevational range during the LGM and at the present day (b, e, h, k) and historic effective population size (c, f, i, l) of *H. tibetana* (a, b, c), *H. digitata* (d, e, f), *H. gregoryi* (g, h, i), *H. platon* (j, k, l). The white dots in the distribution range size map show the actual observation of each population. In the predicted elevational range plots (b,e,h,k), the grey line presents the background elevation range of the whole study area, and the dashed line indicates the mean elevation of the suitable area of species at a time scenario (purple: during the LGM; cyan: at present). The maps can be viewed online at the original spatial resolution at https://fred.igb-berlin.de/data/package/812. The y-axis of the PSMC plots for *H. gregoryi* (i) was re-scaled to filter the noise closer to the present time (~5,000 years before the present).

The demographic history of each species was investigated by PSMC analysis of each genome and as represented by the inverse instantaneous coalescent rate (IICR; Fig. 4 c, f, i, l). Although the IICR estimates became less reliable at more recent time scales, which led to a great variance of the effective population size (*Ne*) between individuals, the results revealed a general trend of an expansion-decline event during or by the end of the LGM (22 ka) for all of the four target species. In addition, the *Ne* of

low-elevation species were 10,000–100,000 times greater than those of the high-elevation distributed species. When comparing the mountain ranges, the *Ne* of the HM and the Himalayan species were more or less at a similar level at their respective peaks.

Estimation of the effects of climatic versus topographic variables using ensemble models produced clear results to explain the potential distribution of species. The highest percentage contribution of the ensemble model was the seasonal precipitation for the two species in the Himalayas, and the seasonal temperature, especially the annual range of temperature for the two species in the HM (Table S5). In addition, for the ensemble modelling of all four target species, the ROC and TSS scores are very high (>0.99, Table S4), indicating a very high model performance.

Coinciding with the maximum *Ne* during the LGM and subsequent population decline, the predicted climatically suitable area of all these four target species was greatest during the LGM and contracted to the present day (Fig. 4 a, d, g, j). For instance, during the LGM, the predicted suitable area of *H. tibetana* and *H. digitata* comprised most of the Himalayas, north of Khasi Hills and part of the HM. However, the present-day suitable area shrunk to a small region in the middle of the Himalayas, which aligns with the known present-day distribution range. Interestingly, except for the middle of the Himalayas, the modelling identified another area on the border of the Khasi Hills and HM that has been persistent for the existence of the two species since the LGM. However, no actual presence of the two species (Fig. 4 a, d) was observed there. Similar to the species distributed in the Himalayas, species distributed in HM also experienced a contraction of climatically suitable area from the LGM to the present, especially for *H. gregoryi*. The suitable area of *H. gregoryi* during the LGM spread all over the southern and eastern part of the HM or further, as well as the eastern Himalayas, but dramatically shrunk to the central part of HM, which still vastly extends the known present-day distribution range. In addition, the suitable areas for *H. gregoryi* and *H. platon* shifted northward over time. Furthermore, the predicted elevational range for all four species during the LGM is lower than at present, implying a tendency for species to move uphill in the Holocene (Fig. 4 b, e, h, k).

4 Discussion

To further understand the origin and evolution of mountain biodiversity, we investigated the genetic diversity and distribution of a caddisfly group in the Himalayas and the HM. For the first time in aquatic insects, our results are derived from whole genome re-sequencing of populations of four congeneric species and fine-grain environmental GIS data. And, while we interpreted our results carefully, our study provides evidence that changes in geodiversity and climate affect the evolutionary history of closely related species occupying varying ecological niches, thus supporting the MGH.

4.1 Genetic patterns of caddisfly in the Himalayas and the HM

Present-day genetic diversity patterns offer important insights on the historic interconnectivity of populations and can help inform the effect of past environmental change on the studied species. Our

results indicate that in most of the cases, the high-elevation species of both mountain regions (*H. tibetana* and *H. gregoryi*) align best with the headwater model pattern *sensu* Hughes et al. (2007). For instance, populations of *H. tibetana* (high-elevation species, Himalayas) in the West Gandaki (subbasin ID 4900) and Middle Gandaki (ID 4754) cluster together and share the same homogeneous ancestry (Fig. 3a, b, c). Instead of being connected by a stream network, these populations are connected by a series of local mountain ranges (including Tilitso Himal, Anna Dakshin, Machhapuchhara and Lamjung Kailas), as defined in the headwater model pattern. Similarly in *H. gregoryi* (HM): populations in the Baima Snow Mountains (occupied in North Mekong (ID 2547) and West Yangtze (ID 1668 and 1685)) originated from an admixed ancestral group and formed one haplogroup based on mtCOI (Fig. S22, Fig. 3g, h, i), in addition to a low level of inbreeding coefficient (F = ~ 0.12), indicating a good connection within this mountain range. Outside the Baima Snow Mountains, the *H. gregoryi* populations located in the North and South Mekong River (containing subbasin ID 2531, 2532, and 2544) were associated with a stream hierarchy model since they show non-significant genetic variance among each other. Nevertheless, the ancestral group of these populations is distinct from others, suggesting limited gene flow between this group and populations from other mountains. Therefore, the populations of high-elevation species are generally restricted by local mountain ranges and show a stronger "insularity" among the different mountain ranges, or a "Sky Island" pattern. This pattern implies some degree of overland dispersal among headwater streams. Caddisflies can disperse overland during their adult stage, although the associated dispersal distance seems to be restricted (Geismar et al., 2015). Nonetheless, high-elevation species likely disperse along local mountain ranges within or among basins, using suitable habitats (so-called "montane archipelagos") as stepping stones, thus leading to better connectivity within these mountain ranges. Moreover, due to their strong association with specific habitats, lowland areas, especially deep and large valleys, may act as dispersal barriers for these species, thus leading to limited gene flow among mountains (Myers et al. 2001; Pauls et al. 2006). In fact, freshwater organisms, in particular headwater specialists, are often endemic to specific mountain ranges due to their low tolerance to elevated temperature or other habitat factors characterizing the surrounding lowland (Finn et al., 2007; Hughes et al., 2009; Bálint et al., 2011; Vitecek et al., 2015, 2017).

In comparison, for the low-elevation species, a significant genetic diversity was shared among populations across basins, thus providing support for the stream hierarchy model pattern. It suggests gene flow mostly occurred among populations from the same basins, indicating that dispersal within the stream network appears to be the dominant process. The divergence among dispersal strategies of high-elevation and low-elevation species is evidenced by the pattern we observed where populations of *H. digitata* and *H. tibetana* in the West and Middle Gandaki (ID 4900, 4754): although populations of these two species partly overlap, populations of the *H. digitata* appeared as two distinct groups associated with the two subbasins, while populations of *H. tibetana* grouped together (Fig. 3c, f). In addition, the population connectivity of *H. digitata* reflects the dendritic nature of the stream as described by Hughes et al. (2013): minimal differentiation was observed for populations within the West Gandaki subbasin

(e.g., individuals from subbasin ID 4900, Fig. 3d, e, f), higher differentiation among populations from different subbasins (e.g., ID 4900 and 4754) and the highest among populations from different primary basins (e.g., Gandaki and Koshi). The stream hierarchy model pattern can also be observed in *H. platon* in the HM. Several populations located in the Yangtze River (subbasin ID 1625, 1634, 1280, 1205) formed distinct groups, suggesting higher connectivity among populations within subbasins (Fig. 3j, k, l). In addition, these populations exhibit low genomic heterozygosity ($\pi < 0.001$) and an extreme inbreeding coefficient (F = ~0.75 – ~0.1, Fig. 2d, h), indicating isolation and future allopatric speciation. The rest of the populations of *H. platon* that extend across all four primary basins (Irrawaddy, Salween, Mekong, and Yangtze) demonstrate panmixia, accompanied by low genetic diversity and low inbreeding coefficient (Fig. 2d, h). This pattern suggests that all of these populations are highly connected and gene flow is strong within all four primary basins until recent times, which may be due to their very well-developed dispersal abilities as adults or continuous habitats (Finn et al., 2007; Baggiano et al., 2011).

Our results highlight that identifying the prevalent model of genetic structure for aquatic organisms may not always be clear-cut, especially for taxa with a broad distribution (Hughes et al., 2013; Sproul et al., 2014). In our case some species even appear to follow multiple patterns: Stream hierarchy model patterns observed in the low-elevation species could also be interpreted as isolation by distance since the genetic diversity also increased with geographic distance (or "stream distance", the same as geographic distance in this case).

In line with previous studies that suggested that a general pattern of population structure is challenging to identify for caddisflies (Lehrian et al., 2009; Pauls et al., 2009; Engelhardt et al., 2011; Previšić et al., 2014), our study also reveals that the patterns in *Himalopsyche* are likely to be species-specific: high-elevation species are associated with a headwater model pattern, whereas the low-elevation species tend to be more similar to stream hierarchy model or panmixia patterns, indicating that habitat specialization is a crucial factor for shaping the population structure. Further, species in the HM show a better-connected gene flow within the mountain range than species distributed in the Himalayas (panmixia appears in the HM), implying that the effects of topography are strong for the dispersal of species with an aquatic life history in this region. Overall, these complex patterns may result from either the complexity of the dendritic stream network embedded in a high-gradient landscape or simply the geographic scale of the studies examined (Hughes et al., 2013).

4.2 Key abiotic factors in shaping the montane biodiversity of caddisflies

Ecological preferences and dispersal abilities cannot explain present-day distribution patterns without interpreting the data in the context of a species' history (Avise 2000). In particular, topography and historical climate changes may have a strong impact on the current genetic diversity pattern of caddisflies. Our study, combined with the current literature (e.g., Favre et al., 2015; Antonelli et al., 2018; Mosbrugger et al., 2018; Muellner-Riehl et al., 2019; Muellner-Riehl, 2019), suggests that the evolution of climate, mountain topography, as well as historical drainage rearrangement, are key factors

for the evolution of aquatic biodiversity in the Tibeto-Himalayan Region. These factors have interacted with each other through time and vastly contributed to shaping current diversity patterns in these aquatic insects.

4.2.1 Genetic patterns versus climate and topographical features

The influence of climate on population diversity of caddisflies is significant, especially in the Himalayan region. In line with other organisms in previous studies (e.g., Wikramanayake et al., 2002; Price et al., 2011; Yan et al., 2013; Rana et al., 2019), we find an increased genetic diversity of our target caddisfly species from west to east in the Himalayan region at the population level, especially for the high-elevation species. There may be a link between this climatic gradient and the genetic diversity of caddisflies across the whole mountain region, leading to patterns associated with topography and longitude (Bookhagen & Burbank, 2006, 2010): species richness increased approximately threefold from the northwest to the east of the Himalayas, especially at higher elevations, which is in accordance with the fact that the lowland receives about three times more precipitation in the subtropical East than the Northwest (Anders et al., 2006; Kreft & Jetz, 2007; Bookhagen & Burbank, 2010; Rana et al. 2019, 2022). This pattern is strongly supported by studies on regional plant diversity (Yan et al., 2013; Muellner-Riehl, 2019; Rana et al., 2022). Precipitation may be the direct or indirect factor resulting in the current distribution of caddisfly species (Collier & Smith, 1997). Furthermore, since both larval and adult stages of caddisflies are reliant on flowing waters, surface water availability from rainfall and snowmelt in the east of the Himalayas may significantly benefit them (Bookhagen & Burbank, 2010).

Compared to the straightforward genetic diversity pattern of caddisflies in the Himalayas, a common distribution pattern of genetic diversity is difficult to identify for caddisflies in the HM. The complex terrain characterizing the HM region offers highly heterogeneous microclimates, making it difficult to evaluate the impacts of climate change in general (Yin et al., 2020). Yet, a variety of genetic diversity patterns of caddisflies were observed in the HM, which is strongly correlated with topographic complexity, in particular those associated with elevation and hydromorphology. Generally, species in the HM show better connectivity all over the region, but there are plenty of allopatric populations observed for both species. For the high-elevation species, the main distribution pattern is the “Sky Island”, thus the geographic barrier made by certain lower elevations. However, for the low-elevation species, the mechanism of formation of the allopatric populations is different. Unlike the alpine species, the low-elevation species generally show better connectivity within basins, especially in *H. platon* which shows near-panmixia across four primary basins. Furthermore, the fact that both strongly connected (the upper Yangtze: Jinsha River) and allopatric populations (the middle Yangtze: Yalong River and Dadu River) occurred in the Yangtze Basin, may derive from ancient drainage re-organizations (e.g., via a so-called “river capture”, Fig. 3m). Firstly, the upper Yangtze River used to flow southward as a tributary of the paleo-Red River until the late Eocene (Clark et al., 2004; Zheng, 2015; He et al., 2021; Zheng et al., 2021), which was not the case of its middle part (see Zhang et al., 2017). Fuelled by regional orogeny and erosion (Zheng, 2015; Wang et al., 2018), a so-called river capture event occurred, changing the

course of the upper Yangtze away from the paleo-Red River and into its current course (Clark et al., 2004; Zheng, 2015; Deng et al., 2021). Our results reveal patterns consistent with an ancient disconnection: populations located in the Yalong (ID 1634) and Dadu River (ID 1280 and 1205) showed a distinct genetic ancestry with any other populations with any K values, including those located at the upper Yangtze River (ID 1625 and 1659). Secondly, although the populations located in the Yanglong and Dadu River represented the same ancestral, they are distinct clusters in population structure, indicating a recent divergence. It is in line with the interpretation that the paleo-Yalong and paleo-Dadu Rivers drained together with the middle Yangtze to the paleo Xigeda lake before ca. 1.3 Ma, but segregate after ca. 1.3 Ma since the paleo-Dadu River was integrated into the lower Yangtze directly (Deng et al., 2021). Until ca. 1.3 Ma, the middle Yangtze captured the upper Yangtze and drained into the lower Yangtze drainage basin (Kong et al., 2009; Deng et al., 2021). Thus, populations of *H. platon* potentially experience some gene flow along the connected river networks, as shown in our TreeMix analysis: gene flow has occurred among populations from the upper Yangtze (ID 1625) and middle Yangtze (ID 1634). However, the genetic connection seems to be very limited, since the populations located in the middle Yangtze showed extremely high inbreeding coefficient.

Interestingly, the population of *H. platon* located in subbasin ID 1625 shares the same ancestral source with the near-panmictic populations when K = 3. However, when increasing the value of K, it is assigned to a unique ancestry and the near-panmictic populations are inferred to have a certain level of ancestry from this unique population with relatively high inbreeding (Fig. 3, Fig. 2). Notably, this population is located at the middle of the First Bend of the Yangtze River (as shown in Fig. 3m). The First Bend is the place where upper Yangtze makes a dramatic turn to the North and starts flowing eastward to the East China Sea (Zhang et al., 2017). It is regarded as the capture point of the upper Yangtze away from the Red River, thus resulting in a key period for geo(hydro)morphology in this area (Clark et al., 2004; Li et al., 2022). The processes leading to the present-day structure and diversity patterns remain unclear, however. Very limited information is available on the genetic diversity patterns of plants or animals here. It would be worthwhile to investigate species of this specific region in future studies.

4.2.2 Distributional and demographic history of the four caddisfly species

Repeated climatic fluctuations in the Pleistocene promoted range contractions and expansions of numerous plant and animal species, which is believed to result in high diversification rates in the Tibeto-Himalayan Region (Xing & Ree, 2017; Mosbrugger et al., 2018; Ding et al., 2020). However, the effects of Pleistocene glacial cycles on the distribution and demography of species may vary across regions (Hewitt, 2000). For example, in addition to ecological preferences, life-history traits, and dispersal ability (Tonzo & Ortego, 2021), local topography and latitude likely have strong effects on rates of dispersal, colonization, and extinction of a species during climate fluctuations.

Our ensemble model results show that in comparison to the present, suitable habitats for all four target species were more extensive during the LGM but shifted from their current core distribution (Fig. 4).

Nevertheless, SDMs depict similar suitable habitats for species from the same mountain region, but divergent patterns for species from the Himalayas and the HM. In the Himalayas, the core potential habitat of both species (at intermediate elevations), seems to have only shifted a little between the LGM and today, indicating minor *in-situ* vertical displacement along the elevational gradient (Fig. 4). Dispersal between the East and the West of the range seems also to have been possible, as indicated by the results of the ancestral components. However, suitable habitats at the eastern and westernmost part of the Himalayan range were likely only available during the LGM, indicating potential local extinction or long-distance migration if species occurred there before the LGM. In comparison, for both species in the HM, a latitudinal displacement (from South at warm times to North at colder times) and a vertical displacement (move up to a higher elevation) occurred after the LGM, which was not the case in the Himalayas. Latitudinal displacement was aided by the mountain ranges long North-South extent (Fig. 5b), whereas the Himalayas are latitudinally strongly constrained by the high plateau to the North and the tropical lowland to the South (Fig. 5a).

In line with the SDM results, evidence of the population size history based on genomic data showed that all species experienced an abrupt demographic expansion during the LGM, followed by a population size decline from the end of the LGM until recent times (Fig. 4). Previous studies combining genetic data and ecological niche modelling have also detected such demographic expansions during the LGM, particularly in cold-tolerant/adapted species, for example in grasshoppers (Tonzo & Ortego, 2021), Vipers (Yousefi et al., 2015), caribou (Taylor et al., 2021), birds (Garg et al., 2020), and plants (Manthey et al., 2017). The *Himalopsyche* species we investigated are cold-tolerant, and their ability to disperse overland (to some extent) may have allowed them to persist *in situ* regionally throughout cold snaps. In fact, this was the case for many mountain caddisfly species in Central Europe (Pauls et al., 2006; Kubow et al., 2010; Lehrian et al., 2010; Previšić et al., 2014), despite a much stronger impact of glaciation on this continent (Višnjević et al., 2020) than in the Tibeto-Himalayan Region (Mao et al., 2021). The limited formation of inland ice during the LGM in the Tibeto-Himalayan Region may have represented an opportunity for cold-adapted species to expand their range rather than being moved far away by massive ice sheets.

4.3 Implication of MGH in different mountain systems

In line with the MGH, we have shown that the historical evolution of abiotic conditions such as changing topography (including paleo-drainage reorganization) and climate are likely to have played dominant roles in shaping the genetic diversity pattern of caddisflies in the Himalayas and the HM. The three MGH boundary conditions (i.e., extensive elevational zonation, response to climatic fluctuations, and a high-relief terrain with environmental gradients; Mosbrugger et al., 2018) are met in both the Himalayas and the HM. Yet, we observe distinct genetic diversity patterns of species in each mountainous region, indicating that although the conditions of the MGH are generally realized opportunities and constraints for species vary across different mountain systems. To explain the formation of different genetic diversity patterns in each mountain region, it seems crucial to compare the geological and climatic

conditions in each region and their effects on the evolution process of species under the concept of MGH.

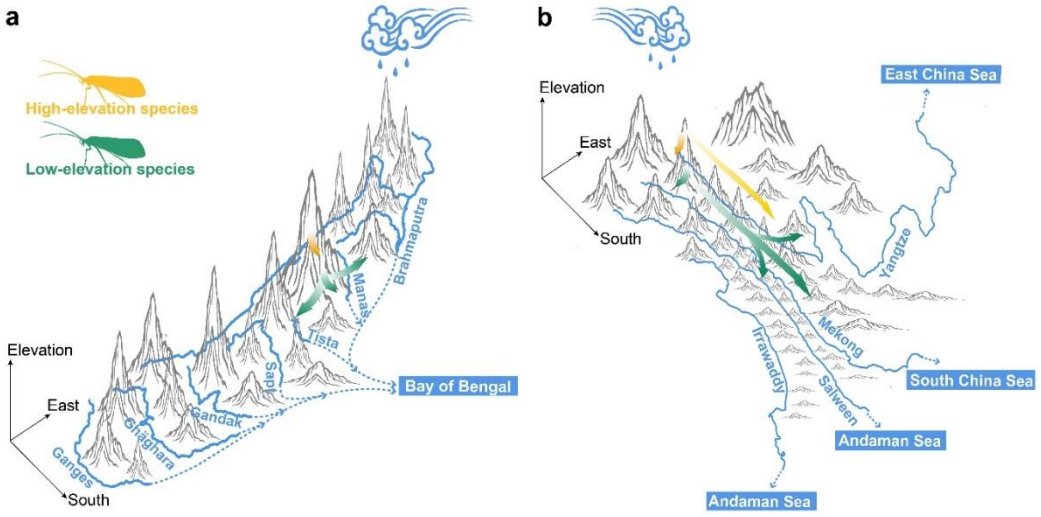


Fig. 5 Diagrammatic sketch of the topographic landscape in the Himalayas (a) and the Hengduan Mountains (b) and the potential *in-situ* displacement (short arrows) and long-distance dispersal (long arrows) of caddisflies during the LGM. Different colours of the caddisflies indicate high-elevation species (yellow) and low-elevation species (green). During the glacial period, which means a cooler climate, a vertical displacement along the elevational gradient (*in-situ* displacement) may occur on both high- and low-elevation caddisfly species. However, compared to the Himalayas, a wider extent of mountain ranges along the North-South orientation in the Hengduan Mountains gives the possibility for species to disperse further south (long-distance dispersal). During the interglacial period, the same dispersal or replacement may happen but in the opposite direction. Therefore, geographic connection and isolation may repeatedly occur during climate fluctuation, thus forming the so-called "species pump" effect and promoting speciation or diversification, according to the MGH (Mosbrugger et al. 2018).

The uplift of the Himalayas and HM are both caused by the India-Eurasia collision since the Paleocene (65–60 Ma). The ranges reached their modern height before the Pliocene (as reviewed by Ding et al., 2022). Although Muellner-Riehl (2019) indicates that the major uplift of the Qinghai-Tibet plateau proper would not have had an impact at the level of intraspecific divergence (as it is vastly older), the distinctive topographic features of these two mountain systems that resulted from this orogenic process has influenced the interconnectivity of habitats among and within each mountainous region (Craw et al., 2016; Rahbek et al., 2019). Compared to the Himalayas, the HM provides a more extensive area in the North-South orientation, thus implying better geographic connectivity for species to shift their distribution towards the south during the cooling period (Xing & Ree, 2017). Our results have verified that caddisfly populations in the HM show better connectivity and a broader potential habitat during the LGM. For this particular scenario that is associated with Pleistocene climatic fluctuations, Flantua et al. (2018) recommended a concept called "flickering connectivity system (FCS)" to explain the rapid diversification and high species richness in mountain ecosystems. Unlike the MGH, which attempts to explain the origination and evolution of mountain biodiversity throughout the whole uplift history, the FCS emphasizes the connectivity dynamics resulting from repetitive climatic changes during the Pleistocene. As reviewed by (Muellner-Riehl, 2019), many previous phylogeographic studies have indicated that the late Pleistocene is a crucial time period for intraspecific diversification in the Tibeto-

Himalayan Region. Therefore, the FCS concept is applicable to explain the formation of intraspecific diversification of caddisflies that occurred in a more recent time period (the LGM).

Caddisflies of the HM and the Himalayas may have reacted to the climatic fluctuations of the Pleistocene by shifting their geographic range within mountain areas. Consequently, the dynamically changing connectivity levels may lead to fragmentation, colonization (dispersal), intermixing, and hybridization, thus promoting diversification in mountain ecosystems (Flantua et al., 2018; Muellner-Riehl, 2019). Moreover, since topography and climate conditions are distinct in different mountain systems, the diversification of species shaped by these four processes under the concept of the FCS is unique in each mountain. The resulting interaction between topography and climate is considered the “fingerprint” of a mountain (Flantua et al., 2018). Although the “fingerprints" are not yet specified for all the mountain systems of the world, our study revealed that the “fingerprints” differ between the Himalayas and the HM. Hence, in agreement with Muellner-Riehl (2019), we think that together with climate fluctuations as one driver of diversification in mountain systems *sensu* MGH, the FCS is relevant for explaining the montane biodiversity patterns of caddisflies. However, given the fact that species react differently to changes in their environment, more studies are needed in the future to explore how the entire biota reacts when facing climatic challenges.

Acknowledgments

We thank Subodh Sharma for obtaining the fieldwork permit from Kathmandu University. We are thankful to Xianfu Li and Lu Tan for their help with the fieldwork permit. We thank Jaime R. Garcia Marquez for providing valuable advice in species distribution modelling. We thank Oskar Schröder, Jacqueline Heckenhauer, and Shenglin Liu for giving professional suggestions on data analyses, results interpretation and comments on the manuscript. We thank Angelika Losey Kurthen, Jiatang Li, and a number of Nepalese students for supporting our fieldwork.

This research was funded by the Deutsche Forschungsgemeinschaft (DFG), grant number PA 1617/2-1 and PA 1617/2-2 awarded to SUP, JA 1827/6-2 awarded to SCJ, DO 1880/3-2 to SD, and AF 1117/1-2 awarded to AF. Reference genome sequencing was supported through the LOEWE–“Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz” of Hesse's Ministry of Higher Education, Research and the Arts. We acknowledge funding through German Federal Ministry of Education and Research (BMBF grant 033W034A). SD acknowledges funding by the Leibniz-Competition (J45/2018). SCJ acknowledges funding through the Leibniz Competition project “Freshwater Megafauna Futures”.

Data Accessibility

All the COI and CAD data for this research are available in the GenBank of National Center for Biotechnology Information (NCBI), and the accession codes of each individual are provided in Supplementary Table 1. The raw data of populations have been deposited at NCBI under the Bioproject ID PRJNA749154. The reference genome of H. tibetana was previously published by Deng et al. (2022)

and has been deposited at NCBI under the BioProject ID PRJNA728835. Scripts used to generate the analyses presented in the paper were deposited in GitLab (Project ID: 31757987).

Author Contributions

SUP, SD and SJ designed the study. RDTS, DNS, QC and FH identified study areas and designed fieldwork. SUP, RDTS, DNS, XD, FH, and PBF performed the fieldwork. XD conducted the laboratory work. XD, SD performed the data analysis. XD and AF drafted the manuscript. All authors contributed to the writing of the manuscript. All authors approved the final draft of the manuscript.

References

Anders, A. M., Roe, G. H., Hallet, B., Montgomery, D. R., Finnegan, N. J., & Putkonen, J. (2006). Spatial patterns of precipitation and topography in the Himalaya. In S. D. Willett, N. Hovius, M. T. Brandon, & D. M. Fisher, *Tectonics, Climate, and Landscape Evolution* (pp. 39–53). Geological Society of America. https://doi.org/10.1130/2006.2398(03)

Antonelli, A., Kissling, W. D., Flantua, S. G. A., Bermúdez, M. A., Mulch, A., Muellner-Riehl, A. N., Kreft, H., Linder, H. P., Badgley, C., Fjeldså, J., Fritz, S. A., Rahbek, C., Herman, F., Hooghiemstra, H., & Hoorn, C. (2018). Geological and climatic influences on mountain biodiversity. *Nature Geoscience*, *11*(10), 718–725. https://doi.org/10.1038/s41561-018-0236-z

Araujo, M., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, *22*(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Avise, J. (2000). *Phylogeography: The History and Formation of Species*. Harvard University Press.

Baggiano, O., Schmidt, D. J., Sheldon, F., & Hughes, J. M. (2011). The role of altitude and associated habitat stability in determining patterns of population genetic structure in two species of *Atalophlebia* (Ephemeroptera: Leptophlebiidae). *Freshwater Biology*, *56*(2), 230–249. https://doi.org/10.1111/j.1365-2427.2010.02490.x

Bálint, M., Domisch, S., Engelhardt, C. H. M., Haase, P., Lehrian, S., Sauer, J., Theissinger, K., Pauls, S. U., & Nowak, C. (2011). Cryptic biodiversity loss linked to global climate change. *Nature Climate Change*, *1*(6), 313–318. https://doi.org/10.1038/nclimate1191

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Bhattarai, K. R., Måren, I. E., & Subedi, S. C. (2014). Biodiversity and invasibility: Distribution patterns of invasive plant species in the Himalayas, Nepal. *Journal of Mountain Science*, *11*(3), 688–696. https://doi.org/10.1007/s11629-013-2821-3

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bookhagen, B., & Burbank, D. W. (2006). Topography, relief, and TRMM-derived rainfall variations along the Himalaya. *Geophysical Research Letters*, *33*(8), L08405. https://doi.org/10.1029/2006GL026037

Bookhagen, B., & Burbank, D. W. (2010). Toward a complete Himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *Journal of Geophysical Research*, *115*(F3), F03019. https://doi.org/10.1029/2009JF001426

Bowler, D. E., & Benton, T. G. (2005). Causes and consequences of animal dispersal strategies: Relating individual behaviour to spatial dynamics. *Biological Reviews*, *80*(2), 205–225. https://doi.org/10.1017/S1464793104006645

Clark, M. K., Schoenbohm, L. M., Royden, L. H., Whipple, K. X., Burchfiel, B. C., Zhang, X., Tang, W., Wang, E., & Chen, L. (2004). Surface uplift, tectonics, and erosion of eastern Tibet from large-scale drainage patterns. *Tectonics*, *23*(TC1006). https://doi.org/10.1029/2002TC001402

Collier, K. J., & Smith, B. J. (1997). Dispersal of adult caddisflies (Trichoptera) into forests alongside three New Zealand streams. *Hydrobiologia*, *361*(1/3), 53–65. https://doi.org/10.1023/A:1003133208818

Craw, D., Upton, P., Burridge, C. P., Wallis, G. P., & Waters, J. M. (2016). Rapid biological speciation driven by tectonic evolution in New Zealand. *Nature Geoscience*, *9*(2), 140–144. https://doi.org/10.1038/ngeo2618

Dahms, C., Kemppainen, P., Zanella, L. N., Zanella, D., Carosi, A., Merilä, J., & Momigliano, P. (2022). Cast away in the Adriatic: Low degree of parallel genetic differentiation in three-spined sticklebacks. *Molecular Ecology*, *31*(4), 1234–1253. https://doi.org/10.1111/mec.16295

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

de Jager, D., Glanzmann, B., Möller, M., Hoal, E., van Helden, P., Harper, C., & Bloomer, P. (2021). High diversity, inbreeding and a dynamic Pleistocene demographic history revealed by African buffalo genomes. *Scientific Reports*, *11*(1), 4540. https://doi.org/10.1038/s41598-021-83823-8

Deng, B., Chew, D., Mark, C., Liu, S., Cogné, N., Jiang, L., O'Sullivan, G., Li, Z., & Li, J. (2021a). Late Cenozoic drainage reorganization of the paleo-Yangtze river constrained by multi-proxy provenance analysis of the Paleo-lake Xigeda. *GSA Bulletin*, *133*(1–2), 199–211.

Deng, X.-L., Frandsen, P. B., Dikow, R. B., Favre, A., Shah, D. N., Shah, R. D. T., Schneider, J. V., Heckenhauer, J., & Pauls, S. U. (2022). The impact of sequencing depth and relatedness of the reference genome in population genomic studies: A case study with two caddisfly species (Trichoptera, Rhyacophilidae, *Himalopsyche*). *Ecology and Evolution*, *12*(12). https://doi.org/10.1002/ece3.9583

[dataset] Deng, X.-L., Frandsen, P. B., Dikow, R. B., Favre, A., Shah, D. N., Shah, R. D. T., Schneider, J. V., Heckenhauer, J., & Pauls, S. U.; 2022; Trichoptera, *Himalopsyche*; NCBI; PRJNA728835

Deng, X.-L., Favre, A., Lemmon, E. M., Lemmon, A. R., & Pauls, S. U. (2021b). Gene Flow and Diversification in *Himalopsyche martynovi* Species Complex (Trichoptera: Rhyacophilidae) in the Hengduan Mountains. *Biology*, *10*(8), 816. https://doi.org/10.3390/biology10080816

Ding, L., Kapp, P., Cai, F., Garzione, C. N., Xiong, Z., Wang, H., & Wang, C. (2022). Timing and mechanisms of Tibetan Plateau uplift. *Nature Reviews Earth & Environment*, *3*(10), 652–667. https://doi.org/10.1038/s43017-022-00318-4

Ding, W.-N., Ree, R. H., Spicer, R. A., & Xing, Y.-W. (2020). Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. *Science*, *369*(6503), 578–581. https://doi.org/10.1126/science.abb4484

Engelhardt, C. H., Haase, P., & Pauls, S. U. (2011). From the Western Alps across Central Europe: Postglacial recolonisation of the tufa stream specialist *Rhyacophila pubescens* (Insecta, Trichoptera). *Frontiers in Zoology*, *8*(1), 10. https://doi.org/10.1186/1742-9994-8-10

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Favre, A., Päckert, M., Pauls, S. U., Jähnig, S. C., Uhl, D., Michalak, I., & Muellner-Riehl, A. N. (2015). The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews*, *90*(1), 236–253. https://doi.org/10.1111/brv.12107

Finn, D. S., Blouin, M. S., & Lytle, D. A. (2007). Population genetic structure reveals terrestrial affinities for a headwater stream insect. *Freshwater Biology*, *52*(10), 1881–1897. https://doi.org/10.1111/j.1365-2427.2007.01813.x

Fitak, R. R. (2021). OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols*, *6*(1), bpab017. https://doi.org/10.1093/biomethods/bpab017

Flantua, S. G., Hooghiemstra, H., & Hoorn, C. (2018). Historical connectivity and mountain biodiversity. In C. Hoorn, A. Perrigo, & A. Antonelli, *Mountains, Climate and Biodiversity* (pp.171–185). John Wiley & Sons, Ltd.

Fox, E. A., Wright, A. E., Fumagalli, M., & Vieira, F. G. (2019). NgsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics*, *35*(19), 3855–3856. https://doi.org/10.1093/bioinformatics/btz200

Fu, P.-C., Sun, S.-S., Hollingsworth, P. M., Chen, S.-L., Favre, A., & Twyford, A. D. (2022). Population genomics reveal deep divergence and strong geographical structure in gentians in the Hengduan Mountains. *Frontiers in Plant Science*, *13*, 936761. https://doi.org/10.3389/fpls.2022.936761

Fu, P.-C., Sun, S.-S., Khan, G., Dong, X.-X., Tan, J.-Z., Favre, A., Zhang, F.-Q., & Chen, S.-L. (2020). Population subdivision and hybridization in a species complex of *Gentiana* in the Qinghai-Tibetan Plateau. *Annals of Botany*, *125*(4), 677–690. https://doi.org/10.1093/aob/mcaa003

Gallien, L., Douzet, R., Pratte, S., Zimmermann, N. E., & Thuiller, W. (2012). Invasive species distribution models - how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography*, *21*(11), 1126–1136. https://doi.org/10.1111/j.1466-8238.2012.00768.x

Garg, K. M., Chattopadhyay, B., Koane, B., Sam, K., & Rheindt, F. E. (2020). Last Glacial Maximum led to community-wide population expansion in a montane songbird radiation in highland Papua New Guinea. *BMC Evolutionary Biology*, *20*(1), 82. https://doi.org/10.1186/s12862-020-01646-z

Geismar, J., Haase, P., Nowak, C., Sauer, J., & Pauls, S. U. (2015). Local population genetic structure of the montane caddisfly *Drusus discolour* is driven by overland dispersal and spatial scaling. *Freshwater Biology*, *60*(1), 209–221. https://doi.org/10.1111/fwb.12489

Gillard, M., Thiébaut, G., Deleu, C., & Leroy, B. (2017). Present and future distribution of three aquatic plants taxa across the world: Decrease in native and increase in invasive ranges. *Biological Invasions*, *19*(7), 2159–2170. https://doi.org/10.1007/s10530-017-1428-y

Graf, W., John, M., Joakim, D., Carmen, Z.-M., & Manuel, J. L.-R. (2008). Distribution and ecological preferences of European freshwater organisms. Volume 1. Trichoptera. In A. Schmidt-Kloiber, & D. Hering, *Distribution and ecological preferences of European freshwater organisms*. Pensoft Publishing.

Griffith, M. B., Barrows, E. M., & Perry, S. A. (1998). Lateral Dispersal of Adult Aquatic Insects (Plecoptera, Trichoptera) following Emergence from Headwater Streams in Forested Appalachian Catchments. *Annals of the Entomological Society of America*, *91*(2), 195–201. https://doi.org/10.1093/aesa/91.2.195

Hao, T., Elith, J., Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, *25*(5), 839–852. https://doi.org/10.1111/ddi.12892

He, M., Zheng, H., Clift, P. D., Bian, Z., Yang, Q., Zhang, B., & Xia, L. (2021). Paleogene Sedimentary Records of the Paleo-Jinshajiang (Upper Yangtze) in the Jianchuan Basin, Yunnan, SW China. *Geochemistry, Geophysics, Geosystems*, *22*(6). https://doi.org/10.1029/2020GC009500

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907–913. https://doi.org/10.1038/35016000

Hjalmarsson, A. E. (2019). Delimitation and description of three new species of *Himalopsyche* (Trichoptera: Rhyacophilidae) from the Hengduan Mountains, China. *Zootaxa*, *4638*(3), 419–441. https://doi.org/10.11646/zootaxa.4638.3.7

Hjalmarsson, A. E. (2020). *Phylogeny and species delimitation of himalopsyche (trichoptera, rhyacophilidae)* [Doctoral dissertation]. Johann Wolfgang Goethe-Universität.

Hjalmarsson, A. E., Graf, W., Jähnig, S. C., Vitecek, S., & Pauls, S. U. (2018). Molecular association and morphological characterisation of *Himalopsyche* larval types (Trichoptera, Rhyacophilidae). *ZooKeys*, *773*, 79–108. https://doi.org/10.3897/zookeys.773.24319

Hjalmarsson, A. E., Graf, W., Vitecek, S., Jähnig, S. C., Cai, Q., Sharma, S., Tong, X., Li, F., Shah, D. N., Shah, R. D. T., & Pauls, S. U. (2019). Molecular phylogeny of *Himalopsyche* (Trichoptera, Rhyacophilidae). *Systematic Entomology*, *44*(4), 973–984. https://doi.org/10.1111/syen.12367

Hoorn, C., Perrigo, A., & Antonelli, A. (2018). *Mountains, climate and biodiversity*. John Wiley & Sons.

Hughes, J. M. (2007). Constraints on recovery: Using molecular methods to study connectivity of aquatic biota in rivers and streams. *Freshwater Biology*, *52*(4), 616–631. https://doi.org/10.1111/j.1365-2427.2006.01722.x

Hughes, J. M., Huey, J. A., & Schmidt, D. J. (2013). Is realised connectivity among populations of aquatic fauna predictable from potential connectivity? *Freshwater Biology*, *58*(5), 951–966. https://doi.org/10.1111/fwb.12099

Hughes, J. M., Schmidt, D. J., & Finn, D. S. (2009). Genes in Streams: Using DNA to Understand the Movement of Freshwater Fauna and Their Riverine Habitat. *BioScience*, *59*(7), 573–583. https://doi.org/10.1525/bio.2009.59.7.8

Humble, E., Dobrynin, P., Senn, H., Chuven, J., Scott, A. F., Mohr, D. W., Dudchenko, O., Omer, A. D., Colaric, Z., Lieberman Aiden, E., Al Dhaheri, S. S., Wildt, D., Oliaji, S., Tamazian, G., Pukazhenthi, B., Ogden, R., & Koepfli, K. (2020). Chromosomal-level genome assembly of the scimitar-horned oryx: Insights into diversity and demography of a species extinct in the wild. *Molecular Ecology Resources*, *20*(6), 1668–1681. https://doi.org/10.1111/1755-0998.13181

Keightley, P. D., Pinharanda, A., Ness, R. W., Simpson, F., Dasmahapatra, K. K., Mallet, J., Davey, J. W., & Jiggins, C. D. (2015). Estimation of the Spontaneous Mutation Rate in *Heliconius melpomene*. *Molecular Biology and Evolution*, *32*(1), 239–243. https://doi.org/10.1093/molbev/msu302

Kong, P., Granger, D., Wu, F., Caffee, M., Wang, Y., Zhao, X., & Zheng, Y. (2009). Cosmogenic nuclide burial ages and provenance of the Xigeda paleo-lake: Implications for evolution of the Middle Yangtze River. *Earth and Planetary Science Letters*, *278*(1–2), 131–141. https://doi.org/10.1016/j.epsl.2008.12.003

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Kreft, H., & Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences*, *104*(14), 5925–5930. https://doi.org/10.1073/pnas.0608361104

Kubow, K. B., Robinson, C. T., Shama, L. N. S., & Jokela, J. (2010). Spatial scaling in the phylogeography of an alpine caddisfly, *Allogamus uncatus*, within the central European Alps. *Journal of the North American Benthological Society*, *29*(3), 1089–1099. https://doi.org/10.1899/09-084.1

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10.1002/hyp.9740

Lehrian, S., Bálint, M., Haase, P., & Pauls, S. U. (2010). Genetic population structure of an autumn-emerging caddisfly with inherently low dispersal capacity and insights into its phylogeography. *Journal of the North American Benthological Society*, *29*(3), 1100–1118. https://doi.org/10.1899/09-100.1

Lehrian, S., Pauls, S. U., & Haase, P. (2009). Contrasting patterns of population structure in the montane caddisflies *Hydropsyche tenuis* and *Drusus discolour* in the Central European highlands. *Freshwater Biology*, *54*(2), 283–295. https://doi.org/10.1111/j.1365-2427.2008.02107.x

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Li, Q., Sun, H., Boufford, D. E., Bartholomew, B., Fritsch, P. W., Chen, J., Deng, T., & Ree, R. H. (2021). Grade of Membership models reveal geographical and environmental correlates of floristic structure in a temperate biodiversity hotspot. *New Phytologist*, *232*(3), 1424–1435. https://doi.org/10.1111/nph.17443

Li, Y., Chen, J., Yan, J., Zhou, F., Wang, Q., Li, Z., & Zhang, Y. (2022). Formation and evolution of a giant old deposit in the First Bend of the Yangtze River on the southeastern margin of the Qinghai-Tibet Plateau. *CATENA*, *213*, 106138. https://doi.org/10.1016/j.catena.2022.106138

Liang, Q., Xu, X., Mao, K., Wang, M., Wang, K., Xi, Z., & Liu, J. (2018). Shifts in plant distributions in response to climate warming in a biodiversity hotspot, the Hengduan Mountains. *Journal of Biogeography*, *45*(6), 1334–1344. https://doi.org/10.1111/jbi.13229

Liu, H., Jia, Y., Sun, X., Tian, D., Hurst, L. D., & Yang, S. (2017). Direct Determination of the Mutation Rate in the Bumblebee Reveals Evidence for Weak Recombination-Associated Mutation and an Approximate Rate Constancy in Insects. *Molecular Biology and Evolution*, *34*(1), 119–130. https://doi.org/10.1093/molbev/msw226

Malicky, H. (1987). Anflugdistanz und Fallenfangbarkeit von Köcherfliegen (Trichoptera) bei Lichtfallen. *Jahresberichte Der Biologischen Station Lunz*, *10*, 140–157.

Malicky, H. (2011). Neue Trichopteren aus Europa und Asien. *Braueria*, 23–43.

Manthey, J. D., Moyle, R. G., Gawin, D. F., Rahman, M. A., Ramji, M. F. S., & Sheldon, F. H. (2017). Genomic phylogeography of the endemic Mountain Black-eye of Borneo (*Chlorocharis emiliae*): Montane and lowland populations differ in patterns of Pleistocene diversification. *Journal of Biogeography*, *44*(10), 2272–2283. https://doi.org/10.1111/jbi.13028

Mao, K., Wang, Y., & Liu, J. (2021). Evolutionary origin of species diversity on the Qinghai–Tibet Plateau. *Journal of Systematics and Evolution*, *59*(6), 1142–1158. https://doi.org/10.1111/jse.12809

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Martynov, A. B. (1930). On the Trichopterons Fauna of China and Eastern Tibet. *In Proceedings of the Zoological Society of London*, *100*(1), 65–112.

Martynov, A. B. (1935). On a Collection of Trichopiera from the Indian Museum. Part I. Annulipalpia. *Records of the Zoological Survey of India*, *37*(2), 93–209.

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, *210*(2), 719–731. https://doi.org/10.1534/genetics.118.301336

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, *16*(3), 1215.

Morse, J. C., Frandsen, P. B., Graf, W., & Thomas, J. A. (2019). Diversity and Ecosystem Services of Trichoptera. *Insects*, *10*(5), 125. https://doi.org/10.3390/insects10050125

Mosbrugger, V., Favre, A., Muellner-Riehl, A. N., Päckert, M., & Mulch, A. (2018). Cenozoic evolution of geo-biodiversity in the Tibeto-Himalayan region. In C. Hoorn, A. Perrigo, & A. Antonelli, *Mountains, climate, and biodiversity* (p. 429). John Wiley & Sons.

Muellner-Riehl, A. N. (2019). Mountains as Evolutionary Arenas: Patterns, Emerging Approaches, Paradigm Shifts, and Their Implications for Plant Phylogeographic Research in the Tibeto-Himalayan Region. *Frontiers in Plant Science*, *10*, 195. https://doi.org/10.3389/fpls.2019.00195

Muellner-Riehl, A. N., & Favre, A. (2021). Mountain biogeography coming full circle: A new '3D' floristic approach provides units for reconstructing evolutionary trajectories. *New Phytologist*, *232*(3), 964–966. https://doi.org/10.1111/nph.17645

Muellner-Riehl, A. N., Schnitzler, J., Kissling, W. D., Mosbrugger, V., Rijsdijk, K. F., Seijmonsbergen, A. C., Versteegh, H., & Favre, A. (2019). Origins of global mountain plant biodiversity: Testing the 'mountain-geobiodiversity hypothesis'. *Journal of Biogeography*, *46*(12), 2826–2838. https://doi.org/10.1111/jbi.13715

Müller, K. (1954). Investigations on the organic drift in north Swedish streams. *Report of the Institute of Freshwater Research, Drottningholm*, *35*, 133–148.

Müller-Peddinghaus, E. (2011). *Flight-morphology of Central European caddisflies (Insecta: Trichoptera) in relation to their ecological preferences* [Doctoral dissertation]. Universität Duisburg-Essen.

Myers, M. J., Sperling, F. A. H., & Resh, V. H. (2001). Dispersal of Two Species of Trichoptera from Desert Springs: Conservation Implications for Isolated vs Connected Populations. *Journal of Insect Conservation*, *5*(3), 207–215. https://doi.org/10.1023/A:1017998513721

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., & Wang, J. (2012). SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. *PLoS ONE*, *7*(7), e37558. https://doi.org/10.1371/journal.pone.0037558

Parkyn, S. M., & Smith, B. J. (2011). Dispersal Constraints for Stream Invertebrates: Setting Realistic Timescales for Biodiversity Restoration. *Environmental Management*, *48*(3), 602–614. https://doi.org/10.1007/s00267-011-9694-4

Pauls, S. U., Lumbsch, H. T., & Haase, P. (2006). Phylogeography of the montane caddisfly *Drusus discolour*: Evidence for multiple refugia and periglacial survival. *Molecular Ecology*, *15*(8), 2153–2169. https://doi.org/10.1111/j.1365-294X.2006.02916.x

Pauls, S. U., Theissinger, K., Ujvarosi, L., Balint, M., & Haase, P. (2009). Patterns of population structure in two closely related, partially sympatric caddisflies in Eastern Europe: Historic introgression, limited dispersal, and cryptic diversity. *Journal of the North American Benthological Society*, *28*(3), 517–536. https://doi.org/10.1899/08-100.1

Perrigo, A., Hoorn, C., & Antonelli, A. (2020). Why mountains matter for biodiversity. *Journal of Biogeography*, *47*(2), 315–325. https://doi.org/10.1111/jbi.13731

Petersen, I., Masters, Z., Hildrew, A. G., & Ormerod, S. J. (2004). Dispersal of adult aquatic insects in catchments of differing land use. *Journal of Applied Ecology*, *41*(5), 934–950. https://doi.org/10.1111/j.0021-8901.2004.00942.x

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, *40*(7), 887–893. https://doi.org/10.1111/ecog.03049

Pickrell, J., & Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings*. https://doi.org/10.1038/npre.2012.6956.1

Previšić, A., Schnitzler, J., Kučinić, M., Graf, W., Ibrahimi, H., Kerovec, M., & U. Pauls, S. (2014). Microscale vicariance and diversification of Western Balkan caddisflies linked to karstification. *Freshwater Science*, *33*(1), 250–262. https://doi.org/10.1086/674430

Price, T. D., Mohan, D., Tietze, D. T., Hooper, D. M., Orme, C. D. L., & Rasmussen, P. C. (2011). Determinants of Northerly Range Limits along the Himalayan Bird Diversity Gradient. *The American Naturalist*, *178*(S1), S97–S108. https://doi.org/10.1086/661926

Rader, R. B., Unmack, P. J., Christensen, W. F., & Jiang, X. (2019). Connectivity of two species with contrasting dispersal abilities: A test of the isolated tributary hypothesis. *Freshwater Science*, *38*(1), 142–155. https://doi.org/10.1086/701671

Rahbek, C., Borregaard, M. K., Antonelli, A., Colwell, R. K., Holt, B. G., Nogues-Bravo, D., Rasmussen, C. M. Ø., Richardson, K., Rosing, M. T., Whittaker, R. J., & Fjeldså, J. (2019). Building mountain biodiversity: Geological and evolutionary processes. *Science*, *365*(6458), 1114–1119. https://doi.org/10.1126/science.aax0151

Rahbek, C., Borregaard, M. K., Colwell, R. K., Dalsgaard, B., Holt, B. G., Morueta-Holme, N., Nogues-Bravo, D., Whittaker, R. J., & Fjeldså, J. (2019). Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science*, *365*(6458), 1108–1113. https://doi.org/10.1126/science.aax0149

Rana, S. K., Price, T. D., & Qian, H. (2019). Plant species richness across the Himalaya driven by evolutionary history and current climate. *Ecosphere*, *10*(11). https://doi.org/10.1002/ecs2.2945

Rana, S. K., White, A. E., Price, T. D., & Meireles, J. E. (2022). Key roles for the freezing line and disturbance in driving the low plant species richness of temperate regions. *Global Ecology and Biogeography*, *31*(2), 280–293. https://doi.org/10.1111/geb.13427

Resh, V. H., & Unzicker, J. D. (1975). Water quality monitoring and aquatic organisms: The importance of species identification. *Journal (Water Pollution Control Federation)*, 9–19.

Ross, H. H. (1941). Descriptions and records of North American Trichoptera. *Transactions of the American Entomological Society (1890-)*, *67*(1/2), 35–126.

Schmid, F., & Botosaneanu, L. (1966). Le genre Himalopsyche Banks (Trichoptera: Rhyacophilidae). *Ann Ent Soc Quebec*, *11*(2), 123–176.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics*, *195*(3), 693–702. https://doi.org/10.1534/genetics.113.154138

Smith, P. J., & Smith, B. J. (2009). Small-scale population-genetic differentiation in the New Zealand caddisfly *Orthopsyche fimbriata* and the crayfish *Paranephrops planifrons*. *New Zealand Journal of Marine and Freshwater Research*, *43*(3), 723–734. https://doi.org/10.1080/00288330909510037

Sproul, J. S., Houston, Derek. D., Davis, N., Barrington, E., Oh, S. Y., Evans, R. P., & Shiozawa, D. K. (2014). Comparative phylogeography of codistributed aquatic insects in western North America: Insights into dispersal and regional patterns of genetic structure. *Freshwater Biology*, *59*(10), 2051–2063. https://doi.org/10.1111/fwb.12406

Srinivasan, U., Tamma, K., & Ramakrishnan, U. (2014). Past climate and species ecology drive nested species richness patterns along an east-west axis in the Himalaya: Nestedness in Himalayan fauna. *Global Ecology and Biogeography*, *23*(1), 52–60. https://doi.org/10.1111/geb.12082

Stone, B. W., & Wolfe, A. D. (2021). Phylogeographic analysis of shrubby beardtongues reveals range expansions during the Last Glacial Maximum and implicates the Klamath Mountains as a hotspot for hybridization. *Molecular Ecology*, *30*(15), 3826–3839. https://doi.org/10.1111/mec.15992

Svensson, B. W. (1974). Population Movements of Adult Trichoptera at a South Swedish Stream. *Oikos*, *25*(2), 157. https://doi.org/10.2307/3543638

Taylor, R. S., Manseau, M., Klütsch, C. F. C., Polfus, J. L., Steedman, A., Hervieux, D., Kelly, A., Larter, N. C., Gamberg, M., Schwantje, H., & Wilson, P. J. (2021). Population dynamics of caribou shaped by glacial cycles before the last glacial maximum. *Molecular Ecology*, *30*(23), 6121–6143. https://doi.org/10.1111/mec.16166

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, *32*(3), 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Tonkin, J. D., Altermatt, F., Finn, D. S., Heino, J., Olden, J. D., Pauls, S. U., & Lytle, David. A. (2018). The role of dispersal in river network metacommunities: Patterns, processes, and pathways. *Freshwater Biology*, *63*(1), 141–163. https://doi.org/10.1111/fwb.13037

Tonzo, V., & Ortego, J. (2021). Glacial connectivity and current population fragmentation in sky islands explain the contemporary distribution of genomic variation in two narrow-endemic montane grasshoppers from a biodiversity hotspot. *Diversity and Distributions*, *27*(9), 1619–1633. https://doi.org/10.1111/ddi.13306

Tsuruishi, T. (2003). Life cycle of a giant carnivorous caddisfly, *Himalopsyche japonica* (Morton) (Trichoptera: Rhyacophilidae), in the mountain streams of Nagano, Central Japan. *Limnology*, *4*(1), 11–18. https://doi.org/10.1007/s10201-003-0091-4

Tsuruishi, T. (2006). Life Cycle of *Himalopsyche japonica* (Morton) (Trichoptera: Rhyacophilidae) in Two High Mountain Streams in Nagano, Central Japan. *Hydrobiologia*, *563*(1), 493–499. https://doi.org/10.1007/s10750-006-0197-x

Tsuruishi, T., Ketavan, C., Suwan, K., & Sirikajornjaru, W. (2006). Importance of Water Flow on Larval Growth and Pupation of *Himalopsyche acharai*, (Malicky and Chantaramongkol, 1989) (Trichoptera: Rhyacophilidae). *Hydrobiologia*, *563*(1), 537–540. https://doi.org/10.1007/s10750-006-0198-9

Ulmer, G. (1932). Aquatic insects of China. Article III. Neue Chinesishe Trichopteren, nebst ubersicht uber die bischer aus China bekannten Arten. *Peking Natural History Bulletin*, *7*, 39–70.

Vieira, F. G., Fumagalli, M., Albrechtsen, A., & Nielsen, R. (2013). Estimating inbreeding coefficients from NGS data: Impact on genotype calling and allele frequency estimation. *Genome Research*, *23*(11), 1852–1861. https://doi.org/10.1101/gr.157388.113

Višnjević, V., Herman, F., & Prasicek, G. (2020). Climatic patterns over the European Alps during the LGM derived from inversion of the paleo-ice extent. *Earth and Planetary Science Letters*, *538*, 116185. https://doi.org/10.1016/j.epsl.2020.116185

Vitecek, S., Graf, W., Previšić, A., Kučinić, M., Oláh, J., Bálint, M., Keresztes, L., Pauls, S. U., & Waringer, J. (2015). A hairy case: The evolution of filtering carnivorous Drusinae (Limnephilidae, Trichoptera). *Molecular Phylogenetics and Evolution*, *93*, 249–260. https://doi.org/10.1016/j.ympev.2015.07.019

Vitecek, S., Vinçon, G., Graf, W., & Pauls, S. (2017). High cryptic diversity in aquatic insects: An integrative approach to study the enigmatic Leuctra inermis species group (Plecoptera). *Arthropod Systematics & Phylogeny*, *75*, 497–521.

Von Humboldt, A. (1860). *Cosmos: A sketch of a physical description of the universe: Vol. 5*. Harper & brothers.

Waldvogel, A.-M., Wieser, A., Schell, T., Patel, S., Schmidt, H., Hankeln, T., Feldmeyer, B., & Pfenninger, M. (2018). The genomic footprint of climate adaptation in *Chironomus riparius*. *Molecular Ecology*, *27*(6), 1439–1456. https://doi.org/10.1111/mec.14543

Wang, K., Zhou, X.-H., Liu, D., Li, Y., Yao, Z., He, W.-M., & Liu, Y. (2022). The uplift of the Hengduan Mountains contributed to the speciation of three *Rhododendron* species. *Global Ecology and Conservation*, *35*, e02085. https://doi.org/10.1016/j.gecco.2022.e02085

Wang, P., Zheng, H., Liu, S., & Hoke, G. (2018). Late Cretaceous drainage reorganization of the Middle Yangtze River. *Lithosphere*, *10*(3), 392–405. https://doi.org/10.1130/L695.1

Wikramanayake, E. D., Dinerstein, E., & Loucks, C. J. (2002). *Terrestrial ecoregions of the Indo-Pacific: A conservation assessment* (Vol. 3). Island Press.

Wilcock, H. R., Bruford, M. W., Nichols, R. A., & Hildrew, A. G. (2007). Landscape, habitat characteristics and the genetic population structure of two caddisflies. *Freshwater Biology*, *52*(10), 1907–1929. https://doi.org/10.1111/j.1365-2427.2007.01818.x

Winterbourn, M. J., Chadderton, W. L., Entrekin, S. A., Tank, J. L., & Harding, J. S. (2007). Distribution and dispersal of adult stream insects in a heterogeneous montane environment. *Fundamental and Applied Limnology*, *168*(2), 127–135. https://doi.org/10.1127/1863-9135/2007/0168-0127

Xing, Y., & Ree, R. H. (2017). Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences*, *114*(17). https://doi.org/10.1073/pnas.1616063114

Yan, Y., Yang, X., & Tang, Z. (2013). Patterns of species diversity and phylogenetic structure of vascular plants on the Qinghai-Tibetan Plateau. *Ecology and Evolution*, *3*(13), 4584–4595. https://doi.org/10.1002/ece3.847

Yin, L., Dai, E., Zheng, D., Wang, Y., Ma, L., & Tong, M. (2020). What drives the vegetation dynamics in the Hengduan Mountain region, southwest China: Climate change or human activity? *Ecological Indicators*, *112*, 106013. https://doi.org/10.1016/j.ecolind.2019.106013

Yousefi, M., Ahmadi, M., Nourani, E., Behrooz, R., Rajabizadeh, M., Geniez, P., & Kaboli, M. (2015). Upward Altitudinal Shifts in Habitat Suitability of Mountain Vipers since the Last Glacial Maximum. *PLOS ONE*, *10*(9), e0138087. https://doi.org/10.1371/journal.pone.0138087

Zhang, Z., Daly, J. S., Li, C., Tyrrell, S., Sun, X., & Yan, Y. (2017). Sedimentary provenance constraints on drainage evolution models for SE Tibet: Evidence from detrital K-feldspar: Pb Isotope of the SE Tibet. *Geophysical Research Letters*, *44*(9), 4064–4073. https://doi.org/10.1002/2017GL073185

Zheng, H. (2015). Birth of the Yangtze River: Age and tectonic-geomorphic implications. *National Science Review*, *2*(4), 438–453. https://doi.org/10.1093/nsr/nwv063

Zheng, H., Clift, P. D., He, M., Bian, Z., Liu, G., Liu, X., Xia, L., Yang, Q., & Jourdan, F. (2021). Formation of the First Bend in the late Eocene gave birth to the modern Yangtze River, China. *Geology*, *49*(1), 35–39. https://doi.org/10.1130/G48149.1

6. Appendix

Publication List

1. **Deng, Xi-Ling**; Domisch, Sami; Favre, Adrien; Jähnig, Sonja C.; Frandsen, Paul B.; He, Fengzhi; Narayan Shah, Deep; Devi Tachamo Shah, Ram; Cai, Qinghua; Pauls, Steffen U.; Comparative phylogeography of *Himalopsyche* (Trichoptera, Rhyacophilidae) in the Tibeto-Himalayan Region: An assessment of the mountain-geobiodiversity hypothesis, Authorea. March 15, 2023. (Preprints)
2. Grigoropoulou, Afroditi; Hamid, Suhaila Ab; Acosta, Raúl; Akindele, Emmanuel Olusegun; Al-Shami, Salman A.; … **Deng, Xi-Ling**; … Domisch, Sami; The global EPTO database: Worldwide occurrences of aquatic insects, Global Ecology and Biogeography, 32(5), 642-655, 2023.
3. **Deng, Xi-Ling**; Frandsen, Paul B.; Dikow, Rebecca B.; Favre, Adrien; Narayan Shah, Deep; Devi Tachamo Shah, Ram; Schneider, Julio V.; Heckenhauer, Jacqueline; and Pauls, Steffen U.; The impact of sequencing depth and relatedness of the reference genome in population genomic studies: A case study with two caddisfly species (Trichoptera, Rhyacophilidae, *Himalopsyche*), Ecology and Evolution, 12(12), e9583, 2022.
4. **Deng, Xi-Ling**; Favre, Adrien; Lemmon, Emily Moriarty; Lemmon, Alan R; Pauls, Steffen U; Gene Flow and Diversification in *Himalopsyche martynovi* Species Complex (Trichoptera: Rhyacophilidae) in the Hengduan Mountains, Biology, 10(8), 816, 2021.
5. **Deng, Xi-Ling**; Yang, Kong; Favre, Adrien; Building an Integrative and Testable Hypothesis on How Edaphic Factors Ultimately Influence the Occurrence and Quality of *Ophiocordyceps Sinensis*, the Vegetable Caterpillar, Research Square, 2021. (Preprints)
6. Yuan, Li-Mei; **Deng, Xi-Ling**; Jiang, De-Chun; Klaus, Sebastian; Orlov, Nikolai L; Yang, Kong; Li, Jia-Tang; Geographical range evolution of the genus *Polypedates* (Anura: Rhacophoridae) from the Oligocene to present, Zoological research, 42(1), 116, 2021.
7. **Deng, Xi-Ling**; Xu, Gao-Wie; Liu, Wie; Yang, Ting-Yong; He Jian; Xie, Hongqi; Yang, Kong; Li, Xi-Dong; Density of *Myospalax baileyi* and Effects of Attractant and Compound Poison Bait with D-type Kreotoxin, Sichuan Journal of Zoology, 36 (2), 203-207, 2017.
8. Yang, Kong; Liu, Wie; **Deng, Xi-Ling**; A Survey on the Herpetological Resources and Species Diversity in Western Wuling Mountains, Sichuan Journal of Zoology, 36 (6), 708-717, 2017.

Conference Contributions

1. Conference of the the Society for Molecular Biology and Evolution (2023)
   Ferrara, Italy
   Poster presentation: "Comparative phylogeography of alpine/subalpine *Himalopsyche* of the Tibeto-Himalayan region: An assessment of mountain geobiodiversity hypothesis"
2. Symposium for European Freshwater Sciences (2023)
   Newcastle Upon Tyne, England
   Oral presentation: "Species distribution modelling and molecular analyzes indicate climate change promoted population expansion during the last glacial maximum"
3. 24th Conference of our Society of Biological Systematics (2023)
   Online
   Oral presentation: "Comparative phylogeography of alpine/subalpine *Himalopsyche* of the Tibeto-Himalayan region: An assessment of mountain geobiodiversity hypothesis"
4. XVII. International Symposium on Trichoptera (2022)
   Lunz am See
   Oral presentation: "Comparative phylogeography of alpine/subalpine *Himalopsyche* species revealed distinct genetic structures in the Himalayas and Hengduan Mountains"
5. 36th Congress of the International Society of Limnology (2022)
   Berlin, Germany
   Oral presentation: "Comparative phylogeography of alpine/subalpine *Himalopsyche* species revealed distinct genetic structures in the Himalayas and Hengduan Mountains"
6. XXVI International Congress of Entomology (2022)
   Helsinki, Finland
   Oral presentation: "Applying anchored hybrid enrichment to study gene flow in the *Himalopsyche martynovi* species complex (Trichoptera Rhyacophilidae)"
7. Senckenberg Biodiversity Genomics Symposium (2021)
   Online
   Poster presentation: "The effects of reference genome and sequencing coverage of short-reads on downstream population genetic analyzes, a study on two caddisfly species (Trichoptera, Rhyacophilidae, *Himalopsyche*)"
8. 9th Biennial Conference of the International Biogeography Society (2019)
   Málaga, Spain
   Poster presentation: "Elevational patterns of Trichoptera diversity in Hengduan Mountains"

Declaration/ Erklärung

Erklärung gemäß der Promotionsordnung des Fachbereichs 09 vom 07. Juli 2004 § 17 (2), geändert am 29. Mai 2019

English

"I declare that I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and cited all text passages that are derived verbatim from or are based on the content of published work of others, and all information relating to verbal communications. I consent to the use of an antiplagiarism software to check my thesis. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University Giessen „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" in carrying out the investigations described in the dissertation."

Deutsch

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-LiebigUniversität Gießen zur Sicherung guter wissenschaftlicher Praxis" niedergelegt sind, eingehalten.

_____ _____

Ort, Datum Unterschrift

7. Acknowledgments

First and foremost, I sincerely thank my mentor and doctor father, Prof. Dr. Steffen Pauls, for accepting me as his student, which allowed me to continue working on science. Throughout my doctoral study, he gave me enormous support, inspiration, and guidance. Under his supervision, I have gained much knowledge on professions, such as scientific thinking, communication, and dependency. All these skills I have learned from him helped me become a better and wiser person than I was years ago. I greatly appreciate the time I have in his group.

Further, I would extend my gratitude to Dr. Sami Domisch and Professor Dr. Sonja Jähnig from the Leibniz Institute for Freshwater Ecology and Inland Fisheries (IGB) for their great contribution to this project. Especially, I would like to thank Dr. Sami Domisch for helping me on this project and providing me with the excellent opportunity to work with the IGB. And his expert knowledge and advice on spatial biodiversity informatics and species distribution modeling are crucial for this study. It is a great pleasure to work with him and his team! In addition, I am deeply grateful to Dr. Adrien Favre. He managed to supervise me in the everyday work and taught me things with attentive guidance. I am especially thankful for his great patience in helping to edit my manuscript draft over and over. And his authority for mountain biodiversity greatly improves my research.

Here I would also like to thank all my colleagues during my PhD. Thank Dr. Julio Schneider, Dr. Jacqueline Heckenhauer, Dr. Oskar Schröder, Dr. Ernesto Rázuri-Gonzales, Alexander Schneider, Dr. Bina Perl, Dr. Daniela Warzecha, Dr. Darko Cotoras, Patricia Peters, Angela Röhner, and Jennifer Steppler for all your help and support! I thank Dr. Simon Vitecek for teaching me knowledge of insects. Next, I want to thank Dr. Paul Frandsen from Brigham Young University, Dr. Fengzhi He, Angelika Kurthen from IGB, Professor Qinghua Cai, Xianfu Li, and Lu Tan from the Institute of Hydrobiology, Chinese Academy of Sciences, Dr. Deep Narayan Shah, Dr. Ram Devi Tachamo Shah and several Nepalese students from Tribhuvan University and Kathmandu University for their help with the fieldwork and providing advice for my research. Thank Dr. Jaime R. Garcia Marquez from IGB for giving valuable advice in species distribution modeling. I sincerely thank Hans Malicky and Wolfram Mey for providing research specimens of *Himalopsyche*.

I would also like to thank my committee members and reviewers of this thesis. Thank you for your effort and time in reviewing this paper, giving valuable suggestions, and agreeing to examine my dissertation. Furthermore, I want to acknowledge the support from the German Research Foundation (DFG) for funding my project and the transitional funding for women postdocs and female doctoral students from Justus-Liebig-University Gießen, as well as support from Senckenberg Research Institute and Natural History Museum, LOEWE-TBG and IGB.

In the end, I want to thank all my family members. It is a long journey full of challenges, but you all make it easier and enjoyable. Thank you all!

8. Curriculum Vitae