5-2019

Phylogenomics and Geometric Morphometrics Define Species Flocks of Snowtrout (Teleostei: Schizothorax) in the Central Himalayas

Binod Regmi
University of Arkansas, Fayetteville

Recommended Citation

Regmi, Binod, "Phylogenomics and Geometric Morphometrics Define Species Flocks of Snowtrout (Teleostei: Schizothorax) in the Central Himalayas" (2019). *Theses and Dissertations*. 3170.
https://scholarworks.uark.edu/etd/3170

Phylogenomics and Geometric Morphometrics Define Species Flocks of Snowtrout (Teleostei: *Schizothorax*) in the Central Himalayas

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Biological Sciences

by

Binod Regmi
Tribhuvan University, Institute of Forestry, Nepal
Bachelor of Science in Forestry, 2001
Wildlife Institute of India
Post Graduate Diploma (Honors) in Wildlife Management, 2007
University of Louisiana
Master of Science in Biology, 2012
University of Arkansas
Master of Science in Statistics and Analytics, 2018

May 2019
University of Arkansas

The dissertation is approved for recommendation to the Graduate Council.

____
Marlis R. Douglas, Ph.D.
Dissertation Director

____
Michael E. Douglas, Ph.D.
Committee Member

____
Andrew J. Alverson, Ph.D.
Committee Member

____
David R. Edds, Ph.D.
Committee Member

Abstract

Schizothorax (Snowtrout) is a genus of medium-sized minnows (Cypriniformes) inhabiting glacier-fed streams, rivers, and lakes in the Himalayas. There are more than 30 species of *Schizothorax* across the region. The speciation and diversity of the Snowtrout in the vast hinterlands of the Himalayan Region has not been fully explored. Three species in Lake Rara, Western Nepal are considered a species flock, comprising endemic ecotypes that are morphologically differentiated and reproductively isolated.

My dissertation research examined the diversity of *Schizothorax* in the Central Himalayan region and evolutionary relationships among species distributed in the Tibet, Central and Southeast Asia. ***Chapter I*** describes the historical biogeography and distribution of *Schizothorax* species in the Himalayas and Tibetan Region. **In *Chapter II,*** morphological and genetic variation was examined among *Schizothorax* collected from three major drainage systems in Nepal using 18 anatomical landmarks (number of images, N=565) and mitochondrial gene (cytochrome b) sequence analysis (n=115). **In *Chapter III***, machine learning algorithms were evaluated to discriminate morphological species based on head and body shape using Procrustes aligned data generated in Chapter 2. In ***Chapter IV***, a phylogenetic tree of *Schizothorax* was constructed comprising Central (Nepal, haplotypes=14) and Eastern (Bhutan, haplotypes=18) Himalayan species to explore their evolutionary relationships within in a global species phylogeny based on GenBank data (n=51, outgroups=5). ***Chapter V*** employed a phylogenomic approach to examine fine-scale relationships amongst *Schizothorax* in Nepal and assess uniqueness of endemic forms in Lake Rara. Double digest restriction associated DNA (ddRAD)sequences were generated to extract 20,000 single nucleotide polymorphism (SNPs)

loci. These data were used to trace the selection driven phenotypic convergence among species isolated largely due to the geographical and ecological barriers.

Both species and basins were significant predictors of the shape. Classifiers, such as Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM), assigned individuals to morphological species with high accuracy. However, a strong geographic structure was reflected in the mitochondrial (cytochrome b) gene sequence data. Conversely, phylogenomic analysis of SNPs uncovered basin-specific upstream and downstream clades, as well as Lake Rara endemic species as a monophyletic group that mitochondrial gene analyses failed to resolve in previous studies.

Table of Content

Chapter II Morphological and genetic variability in a species-complex of Central Himalayan fish (Cyprinidae:Schizothorax) inhabiting Lake Rara in Nepal

Chapter III Can machine learning accurately delimit morphological species: Central Himalayan Snowtrout (Schizothorax) as a test case

Chapter IV Paleodrainages and the biogeography of Snowtrout (Cyprinidae: Schizothorax) in South and Central Asia

Chapter V Species flocks of Snowtrout (Schizothorax) in the Central Himalayas: convergent evolution along elevation gradient

Acknowledgments

This dissertation has been made possible by dedicated field staff and researchers who have a deep passion for the ichthyology and ecology of the Himalayan Region. During this research, Dr. Leo Smith and Mr. Andrew Bentley, Biodiversity Institute at Kansas University (BI-KU), made specimens and tissue samples available and provided technical support to capture digital images of vouchers. Dr. Dustin Lynch, Department of Integrative Biology, Oklahoma State University (OSU), provided access to preserved specimens maintained in the institution's Collection of Vertebrates (CV-OSU). Mr. Karma Wangchuk and other dedicated staff at the National Research Centre for Riverine and Lake Fisheries, Division of Livestock, Ministry of Agriculture and Forests, Royal Government of Bhutan, provided field assistance for sampling *Schizothorax* in Bhutan.

Special recognition is presented to Dr. David Edds, Professor, Emporia State University, Kansas, for his endless desire to explore the fishes of the Central Himalayas over the last forty years. Tissue samples and specimens came from his sampling efforts in Nepal, now vouchered at BI-KU and CV-OSU. Special thanks are due to Dr. Whitney B. J. Anthonysamy (Postdoc.) in Douglas lab for her help with the wet lab procedures.

I express my sincere gratitude to my M.Sc. (Statistics and Analytics) advising committee members: Dr. Michael Douglas, Dr. Mark Arnold and Dr. Marlis Douglas for their guidance provided to complete the coursework required for the program. The coursework was an enormous support for my dissertation research, especially with regards to the machine learning approaches and analysis of molecular data. I extend my sincere gratitude to my Ph.D. Dissertation Committee members: Dr. Andrew Alverson and Dr. David Edds for their advice from developing the proposal to finalizing my dissertation. I will be always indebted to my advisors Dr. Marlis Douglas and Dr. Michael Douglas for their guidance and inspirational support.

This research was made possible through generous endowments to the University of Arkansas: The Bruker Professorship in Life Sciences (Marlis Douglas), and the 21st Century Chair in Global Change Biology (Michael Douglas), as well as Graduate Assistantship (BR), all of which provided salaries and/or research funds for this work. Additional analytical resources were provided by the Arkansas Economic Development Commission (Arkansas Settlement Proceeds Act of 2000) and the Arkansas High Performance Computing Center (AHPCC).

Finally, I am thankful to my family members, especially my mom Bodh Kumari Regmi for her love, inspiration and, patience in my long absence. And last but not least, I am indebted to my wife Kavita Regmi and daughter Avani Regmi for their patience, and support at every step of my long academic journey.

Chapter I

Introduction

Climate and tectonism act synergistically to shape the hydrology of aquatic ecosystems (Whipple, 2009) and impact the distributions of resident freshwater fishes (Last et al., 2011; Hopken et al., 2013; Craw et al., 2016). Global climate fluctuations over the past three million years, including the major ice ages, caused dramatic shifts in species distributions, with genetic consequences of these range shifts and concomitant population size fluctuations still reverberating and traceable with appropriate markers (Hewitt, 2000). A long-term, historic drought severely constrained drainages in western North America (Cook et al., 2016) inducing the basin-wide collapse of fish populations and drastically reducing genetic diversity (Douglas et al., 2003). Despite such range-wide bottlenecks, these species have remained viable over evolutionary time. Conversely, large-scale episodic flooding (Harden et al., 2010) down-cut antecedent streams, eroded headwaters, and over-spilled basins, thus promoting large-scale dispersals. As a result, drainage reorganization (i.e., captures, diversions, beheadings; Bishop, 1995; Li et al., 2018) not only extended distributions into adjacent basins (Strange & Burr, 1997) but also confound taxonomies by facilitating hybridization (Carson & Dowling, 2006; Bangs et al., 2018). Flows within basins were frequently abridged by longitudinal barriers that subsequently became vicariant for larger-bodied species but served as filters for those smaller and more limited in their dispersal (Mims & Olden, 2013; Radinger et al., 2017). As a result, fish distributions are often characterized as a series of biogeographic ‘islands’ within and among basins, each reflecting local expansions, contractions, and eliminations (Burridge et al., 2008; Hugueny et al., 2010; Tedesco et al., 2012).

The above 'benchmarks' are particularly applicable when evaluating fishes found within remote dendritic drainage systems with similar climatic and tectonic histories. One such region is the Himalayan Mountains, a high-elevation global biodiversity hotspot located at the southern border of the Tibetan Plateau (Myers et al., 2000). As with most mountainous regions, it offers ample opportunity for resident biodiversity to differentiate, such as via elevational clines (Du et al., 2017), isolation in refugia (Li et al., 2016a), and allopatric speciation (Liu et al., 2016). All of these processes have enhanced the recognition of the Himalayas as a global biodiversity hotspot (CREP, 2017). The benchmarks (above) also serve as templates from which to gauge the potential response by fishes to rapid anthropogenic climate change. The Himalayas are extremely vulnerable in this regard (Jacobsen et al., 2012), with a warming rate 2x greater than the global average (Shrestha et al., 2012; Xu & Grumbine, 2014), yet with impacts on biodiversity that can potentially promote divergence in lieu of range contraction. For example, cold-adapted terrestrial species may be buffered by microclimates at a higher elevation (Scherrer & Körner, 2011), whereas aquatic invertebrate communities seemingly lose their species diversity as elevation increases (Jacobson et al., 2012; Li et al., 2016b). However, data for high elevation fishes remain elusive, particularly so for species found in headwater drainages of developing countries, a situation exacerbated by nascent monitoring programs and concomitant taxonomy uncertainty with regard to endemic biotas (Myers et al., 2000).

In my dissertation, I integrated different sources of information so as to identify biodiversity at multiple spatial scales within Snowtrout (*Schizothorax*), a genus of minnows (Cypriniformes) broadly distributed across the Himalayan region. I first evaluated phenotypic diversity and historic phylogeography of Snowtrout by using museum samples collected from three rivers in Nepal (Koshi, Gandaki, Karnali), as well from an isolated headwater lake (Lake

Rara, Nepal). These drainages cumulatively flow south into the Ganges River, a major tributary in northern India that drains east and south into the Bay of Bengal (Figure 1). I then expanded the phylogeographic analyses across four rivers of Bhutan, a second Himalayan country east of Nepal. These drainages also flow south but merge into the Brahmaputra River, a second major tributary that independently drains from the Tibetan Plateau into the Bay of Bengal (Figure 2). Finally, I expanded the genetic assessment across the Himalayan Region by integrating my data with sequences obtained from public databases (National Institute of Health Genetic Sequence Database GenBank®: https: //www.ncbi.nlm.nih.gov /GenBank/) so as to juxtapose diversity within Snowtrout against contemporary and historic drainage patterns. The resulting phylogenetic tree offers a hypothesis for the evolution of Tibetan Plateau basins and a framework to understand dispersal and speciation of aquatic organisms in the Central Himalayas.

Chapter I provides a general introduction and regional biogeography of the Himalayan Region. In **Chapter II**, I evaluated phenotypic variation by conducting a shape analysis using geometric morphometrics (GM) to discriminate amongst morphotypes. Morphological data are prerequisite for the formal recognition and description of species. Yet, the previous description of forms endemic to Lake Rara is qualitative, rather than quantitative, with supporting evidence derived from ecological and life history data. Here, my sample size comprised 565 preserved specimens almost equally distributed among three *Schizothorax* species endemic to Lake Rara (*S. macrophthalmus, S. nepalensis, and S. raraensis*), and two riverine species (*S. richardsonii* and *S. progastus*) more widely distributed and found in three major rivers in Nepal. Each specimen was photographed using a high-resolution digital camera anchored within a camera box designed to eliminate shadows and distortions. The lateral shape of the head and body of each image was digitized, and Cartesian coordinates captured across 18 homologous landmarks

(Type I). Data were amalgamated, superimposed, and condensed into shape variables and predictability of taxonomy (species) versus geography (basin) on these variables was evaluated. I found both ‘species’ and ‘basin’ as significant predictors of shape variables, but the taxonomic signal was stronger than the geographic patterns.

In **Chapter III** I explored machine learning approaches to discriminate amongst morphospecies as defined by shape variables derived in Chapter II. The purpose of this study was to evaluate the applicability of popular machine learning algorithms to discriminate closely related species. I used principal component analysis (PCA) with logistic regression (LG) to visualize the specimens in the shape space. Classifiers comprising linear discriminant analysis (LDA), support vector machine (SVM), and random forest (RF) were evaluated based on their accuracy and receiver operating character (ROC) curves. LDA, and SVM with linear kernel performed well discriminating morphospecies, but random forest (RF) did not.

In **Chapter IV**, I examined phylogenetic diversity in Snowtrout across the Himalayan Region. I combined 32 unique haplotypes I generated from samples representing *Schizothorax* species in the Central and Eastern Himalayan (i.e. Nepal and Bhutan) with an additional 51 haplotypes of more than 25 *Schizothorax* species distributed in Tibet, Central and South East Asia obtained from GenBank®. I obtained five additional sequences of two outgroup species from the GenBank. These data encompassed samples from the Indus and the Upper Ganges River basins (Western Himalayas), as well as major south- and southeast draining Asian rivers, to include the Salween, Mekong, Yangtze, and the Yarlung Tsangpo rivers (termed the Brahmaputra in India), each of which independently drain the Tibetan Plateau. The resulting phylogenetic tree offers a hypothesis for the evolution of Tibetan Plateau drainage patterns. In this sense, *Schizothorax* serves as a model from which to juxtapose the dispersal and speciation of aquatic organisms in the Central Himalayas. I found a strong geographic structure in the phylogeny. Species in upper catchments of the Himalayan Region have been isolated from other river basins, but there is putative gene flow with downstream species. This clearly indicates that tectonism and climate change have shaped the diversity of Snowtrout in the Himalayan Region.

In Chapter V, I tested the hypothesis that the three *Schizothorax* found in a remote and mountainous lake in Nepal (i.e., Lake Rara) represent a species flock by employing a population genomic approach. All three species have been described as distinct, based upon morphological, ecological, and life history data, yet mitochondrial DNA analyses could not diagnose them as monophyletic. I used double-digest restriction-site associated (ddRAD) sequencing of 48 samples representing five *Schizothorax* species found in Nepal, two more wide-spread riverine species (*S. richardsonii, and S. progastus*) and the three species endemic to Lake Rara (*S. macrophthalmus, S. nepalensis, S. raraensis*) to screen single nucleotide polymorphisms (SNPs) across thousands of nuclear loci. This allowed me to examine the genetic relationships among populations in Lake Rara as compared with other *Schizothorax* that reside in adjacent rivers and are presumed to be ancestral to the lacustrine forms. Additionally, genetic differences between upstream and downstream ecotypes within a river were examined to understand the selection gradient along with the elevational range. Geographic patterns revealed by the SNP data were congruent with those derived from mitochondrial DNA; genetic structure suggests a closer relationship between species within basins than among. Similar phenotypes of upstream *versus* downstream morphs appear to have evolved repeatedly across basins, with *in-situ* diversification likely driven by selection along steep elevation gradients.

In summary, I integrated phylogenetic, population genomic and shape variation data to evaluate concordance amongst these different data sets and to derive a taxonomic hypothesis that is based on statistical delimitation of five *Schizothorax* OTUs (operational taxonomic units) in Nepal (i.e., three Lake Rara endemics and two riverine species external to the lake). The drainage patterns in South Asia are of recent origin (Brookfield, 1998) leaving insufficient time for strong phylogenetic signals to emerge. However, my approach using multiple lanes of evidence, revealed that morphological data correlate more strongly with OTUs than geography (basins), with the exception of the Eastern basin in Nepal (Koshi River). Yet, molecular data support that species divergence occurred within basins, irrespective of putative morphological species. Hence, geography could be a significant predictor of biodiversity evolution in the Himalayas. Furthermore, emergency of similar phenotypes (shape, morpho-species) in each basin suggests selection gradients are defined by elevation, with upstream reaches inducing the *S. richardsonii* phenotype, and lower reaches the *S. progastus* phenotype. This, in turn, suggests that *Schizothorax* may be relatively plastic and can quickly adapt to different ecological niches. If indeed convergent evolution is the driver for repeated emergence of two riverine ecotypes, the divergence of *Schizothorax* in Lake Rara would just represent an extension of this capacity to quickly adapt to novel niches and radiate into a species flock. This has important consequences for biodiversity conservation in the region; despite similar phenotypes, species in each basin have independently evolved, are endemic to this drainage and each represents a unique evolutionary lineage in need of conservation.

My combined approach provided a template to quantitatively discriminate species in a region where taxonomy is at best unresolved, but largely uncertain due to lack of comprehensive, integrated studies. The findings of my study will aid in evaluating the potential response of freshwater biodiversity in the Himalayan region to rapid climate change.

The geographical setting

Himalayan Mountains

The Himalayas are the most extensive and recently evolved mountain system on earth (length=2,400km; width=240km; elevation=7,500-8,800m). They are bordered by rivers on the west (Indus) and east (Brahmaputra), and their breadth encompasses most of northern India, (Kashmir to the west, Assam to the east), as well as Nepal, Bhutan, and parts of China, with the extensive Tibetan Plateau to the north and the alluvial plains of India and Bangladesh to the south.

The physiographic evolution of the Himalayas occurred over a 30my span, resulting from a two-stage impact of the Indian craton with the Eurasian landmass. A preliminary 'soft' collision at 50mya preceded a much more substantial and more contiguous 'hard' collision at 25-20mya, as driven by the Indian lithosphere (van Hinsbergen et al., 2012; Spakman & Torsvik, 2012). The leading edge of the Indian craton was subducted, whereas its upper surface was essentially skimmed to produce three sequential north-south stratigraphic zones in the Himalayas (i.e., Tibetan, Greater, and Lesser; van Hinsbergen et al., 2012). The hard collision not only established the Himalayas as a topographic entity but also profoundly influenced the dendritic connectivity of regional rivers (Brookfield, 1998; Bracciali et al., 2015), and defined the evolution of their aquatic and semi-aquatic organisms. It not only drove global climate but also transformed the cultural and biological endemism of the region on a massive scale (Xu et al., 2009; Xu & Grumbine, 2014). Crustal thickening and lift continued through the Pliocene-Pleistocene, with the Lesser Himalayas uplifting between 2.5-1.0mya (Sakai et al., 2006).

Drainages of Nepal

There are three major east-to-west drainage basins in Nepal (i.e., Koshi, Gandaki, and Karnali; Figure 1), with eight major tributaries, several of which originate within the Tibet. In the west, the Karnali River provides a vicariant barrier for western Nepal and demarks a floristic composition for the region that is most similar to that of Central Asia. One study site, Lake Rara (with a surface area of 10.6 km^2), is located at 2,990 m elevation within the Karnali River basin. Central Nepal is drained by the Gandaki River, which is recognized as a vicariant barrier separating the flora of western and eastern Nepal. The Kali Gandaki River Gorge, located in the upper catchment area, is considered the deepest geomorphic feature on a global scale (>5,500m). The Koshi River to the east includes the Arun River which represents the most extensive drainage in Nepal, cutting deeply through the Himalayas between Mt. Makalu (8,481m) and Mt. Kangchenjunga (8,586m). The elevational change is more acute in eastern Nepal, as compared to western Nepal, and the impact of the monsoon is much stronger in the East. These biogeographic aspects offer numerous opportunities for biodiversity diversification in the region.

Historically, the major tributaries of these rivers flowed independently southward into the Ganges River. However, orogenesis and uplift of the Mahabharat Range (1-2mya) coalesced their paths into the three outlets that now cut through the mountain range (Gupta, 1997). The current drainage pattern thus reflects recent connectivity of these freshwater tributaries, as compared to their historic segregation documented in the historical hydrography of the region.

Drainages of Bhutan

There are five large rivers that drain Bhutan, three of which are located in the western part of the country (i.e., Amo Chhu, Wang Chhu, and Punatsang Chhu/ Sunkosh, Figure 2). Central Bhutan is drained by a large, dendritic system composed of four large rivers (Mangde Chhu, Chamkhar Chhu, Kuri Chhu, and Dangme Chhu) that join pairwise before merging into the Manas River just north of the Indian border. The Nyera Ama Chhu is less extensive in length and drains the eastern part of the country.

The major rivers of Bhutan drain south into the Brahmaputra River of India. This river originates on the Tibetan Plateau as the Tsang Pho River, traverses eastward then south along the eastern edge of Bhutan. It then flows west across northern Bangladesh and India before it curves south, eventually joining the Ganges River and discharging into the Bay of Bengal.

References:

Bangs, M. R., Douglas, M. R., Mussmann, S. M., and Douglas, M. E. (2018). Unraveling historical introgression and resolving phylogenetic discord within *Catostomus* (Osteichthys: Catostomidae). *BMC Evolutionary Biology,* 18:86. doi: 10.1186/s12862-018-1197-y

Bishop, P. (1995). Drainage rearrangement by river capture, beheading, and diversion. *Progress in Physical Geography*, 19:449-473. doi: 10.1177/030913339501900402.

Bracciali, L., Najman, Y., Parrish, R. R., Akhter, S. H., and Millar, I. (2015). The Brahmaputra tale of tectonics and erosion: Early Miocene river capture in the Eastern Himalaya. *Earth and Planetary Science Letters*, 415:25-37. doi: 10.1016/j.epsl.2015.01.022.

Brookfield, M. E. (1998). The evolution of the great river systems of southern Asia during the Cenozoic India-Asia collision: Rivers draining southwards. *Geomorphology*, 22:285-312. doi: 10.1016/S0169-555X(97)00082-2.

Burridge, C. P., Craw, D., Jack, D. C., King, T. M., and Waters, J. M. (2008). Does fish ecology predict dispersal across a river drainage divide? *Evolution*, 62:1484-1499. doi: 10.1111/j.1558-5646.2008.00377.x.

Carson, E. W., and Dowling, T. E. (2006). Influence of hydrogeographic history and hybridization on the distribution of genetic variation in the pupfishes *Cyprinodon atrorus* and *C. bifasciatus*. *Molecular Ecology*, 15:667-679.

Cook, B. I., Cook, E. R., Smerdon, J. E., Seager, R., Williams, A. P., Coats, S., Stahle, D. W., and Díaz, J. V. (2016). North American megadroughts in the Common Era: reconstructions and simulations. *Wiley Interdisciplinary Reviews: Climate Change*, 7: 411-432.

Craw, D., Upton, P., Burridge, C. P., Wallis, G. P., and Waters, J. M. (2016). Rapid biological speciation driven by tectonic evolution in New Zealand. *Nature Geoscience*, 9:140.

CREP-Critical Ecosystem Partnership Fund (2017). The biodiversity hotspots. (http://www.cepf.net/resources/hotspots/Pages/default.aspx.)

Douglas, M. R., Brunner, P. C., and Douglas, M. E. (2003). Drought in an evolutionary context: Molecular variability in Flannelmouth Sucker (*Catostomus latipinnis*) from the Colorado River Basin of western North America. *Freshwater Biology*, 48:1254-1273. doi: 10.1046/j.1365-2427.2003.01088.x.

Du, F., K., Hou, M., Wang, W., Mao, K., and Hampe, A. (2017). Phylogeography of *Quercus aquifolioides* provides novel insights into the Neogene history of a major global hotspot of plant diversity in south-west China. *Journal of Biogeography*, 44:294-307. doi: 10.1111/jbi.12836.

Gupta, S. (1997). Himalayan drainage patterns and the origin of fluvial megafans in the Ganges foreland basin. *Geology*, 25:11-14. doi: 10.1130/0091-7613(1997)025<0011:HDPATO>2.3.CO;2.

Harden, T., Macklin, M. G., and Baker, V. R. (2010). Holocene flood histories in southwestern USA. *Earth Surface Processes and Landforms*, 35:707–716. doi: 10.1002/esp.1983.

Hewitt, G., (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*, 907.

Hopken, M. W., Douglas, M. R., and Douglas, M. E. (2013). Stream hierarchy defines riverscape genetics of a North American desert fish. *Molecular Ecology*, 22:956-971.

Hugueny, B., Oberdorff, T., and Tedescco, P. A. (2010). Community ecology of fishes: A large scale perspective. *American Fisheries Society Symposium*, 73:29-62. http://www.k-state.edu/fishecology/msreprints/Gido%20and%20Jackson%202010.pdf.

Jacobson, D., Milner, A. M., Brown, L. E., and Dangles, O. (2012). Biodiversity under threat in glacier-fed river systems. *Nature Climate Change*, 2:361-364. doi: 10.1038/NCLIMATE1435.

Last, P. R., White, W. T., Gledhill, D. C., Hobday, A. J., Brown, R., Edgar, G.J., and Pecl, G. (2011). Long-term shifts in abundance and distribution of a temperate fish fauna: a response to climate change and fishing practices. *Global Ecology and Biogeography*, 20:58-72.

Li, F., Shah, D. N., Pauls, S. U., Qu, X., Cai, Q., and Shah, R. T. (2016b). Elevational shifts of freshwater communities cannot catch climate warming in the Himalaya. *Water*, 8:327. doi: 10.3390/w8080327.

Li, H., He, Y., Jiang, J., Liu, Z., and Li, C., (2018). Molecular systematics and phylogenetic analysis of the Asian endemic freshwater sleepers (Gobiiformes: Odontobutidae). *Molecular Phylogenetics and Evolution*, 121:1-11.

Li, J., McCarthy, T. M., Wang, H., Weckworth, B. V., Schaller, G. B., Mishra, C., Lua, Z., and Beissinger, S. R. (2016a). Climate refugia of snow leopards in High Asia. *Biological Conservation*, 203:188-196. doi: 10.1016/j.biocon.2016.09.026.

Liu, Y., Hu, J., Li, S-H., Duchen, P., Wegmann, D., and Schweizer, M. (2016). Sino-Himalayan mountains act as cradles of diversity and immigration centers in the diversification of parrotbills (Paradoxornithidae). *Journal of Biogeography*, 43:1488-1501. doi: 10.1111/jbi.12738.

Mims, M. C., and Olden, J. D. (2013). Fish assemblages respond to altered flow regimes via ecological filtering of life history strategies. *Freshwater Biology*, 58:50-62. doi: 10.1111/fwb.12037.

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

Radinger, J., Essl, F., Holker, F., Horky, P., Slavik, O., and Wolter, C. (2017). The future distribution of river fish: The complex interplay of climate and land use changes, species dispersal and movement barriers. *Global Change Biology*, 2017:1-17. doi: 10.1111/gcb.13760.

Sakai, H., Sakai, H., Yahagi, W., Fujii, R., Hayashi, T., and Upreti, B. N. (2006). Pleistocene rapid uplift of the Himalayan frontal ranges recorded in the Kathmandu and Siwalik basins. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 241:16-27. doi: 10.1016/j.palaeo.2006.06.017.

Scherrer, D., and Körner, C. (2011). Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *Journal of Biogeography*, 38: 406-416. doi: 10.1111/j.1365-2699.2010.02407.x.

Shrestha, U. B., Gautam, S., and Bawa, K. S. (2012). Widespread climate change in the Himalayas and associated changes in local ecosystems. *PLoS ONE*, 7, e36741. doi: 10.1371/journal.pone.0036741.

Spakman, W., and Torsvik, T. H. (2012). Greater India Basin hypothesis and a two-stage Cenozoic collision between India and Asia. *Proceedings of the National Academy of Sciences,* 109:7659-7664. doi: 10.1073/pnas.1117262109.

Strange, R. M., and Burr, B. M. (1997). Intraspecific phylogeography of North American Highland fishes: A test of the Pleistocene vicariance hypothesis. *Evolution*, 51:885-897. doi: 10.1111/j.1558-5646.1997.tb03670.x.

Tedesco, P., A., Leprieur, F., Hugueny, B., Brosse, S., Dürr, H., H., Beauchard, O., Busson, F., and Oberdorff, T. (2012). Patterns and processes of global riverine fish endemism. *Global Ecology and Biogeography*, 21:977-987. doi: 10.1111/j.1466-8238.2011.00749.x.

Van Hinsbergen, D. J., Lippert, P. C., Dupont-Nivet, G., McQuarrie, N., Doubrovine, P.V., Spakman, W., and Torsvik, T. H. (2012). Reply to Aitchison and Ali: Reconciling Himalayan ophiolite and Asian magmatic arc records with a two-stage India-Asia collision model. *Proceedings of the National Academy of Sciences*, 109:E2646-E2646.

Whipple, K. X. (2009). The influence of climate on the tectonic evolution of mountain belts. *Nature Geoscience*, 2: 97.

Xu, J., and Grumbine, R. E. (2014). Building ecosystem resilience for climate change adaptation in the Asian highlands. *WIREs Climate Change*, 5:709-718. doi: 10.1002/wcc.302.

Xu, J., Grumbine R. E., Shresta, A., Eriksson, M., Yang, X., Wang, Y., and Wilkes, A. (2009). The melting Himalayas: Cascading effects of climate change on water, biodiversity, and livelihoods. *Conservation Biology*, 23:520-530. doi: 10.1111/j.1523-1739.2009.01237.x.

Appendix

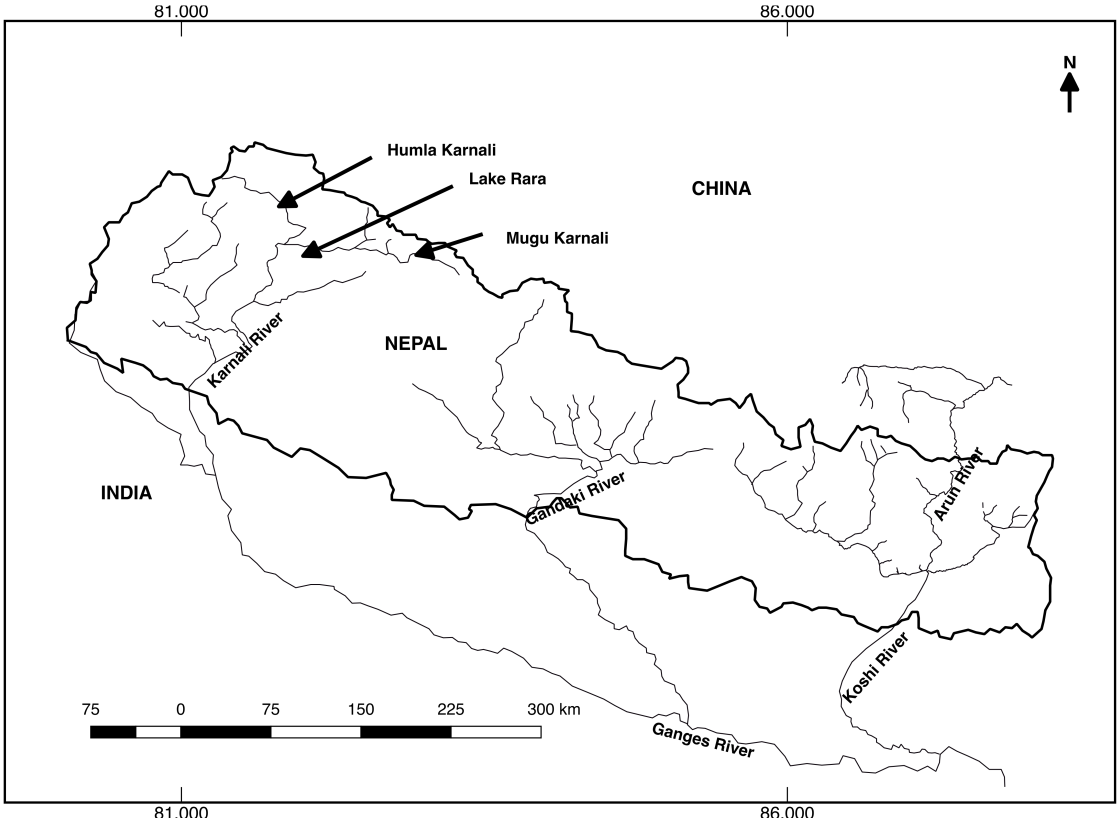

Figure 1: Three major drainage basins in Nepal, Central Himalayas. From west to east: Karnali, Gandaki, and Koshi rivers. Lake Rara is situated in the Karnali River Basin in western Nepal. Tributaries join before each river crosses the Lesser (or outer) Himalayas and drains into the Ganges River in India.

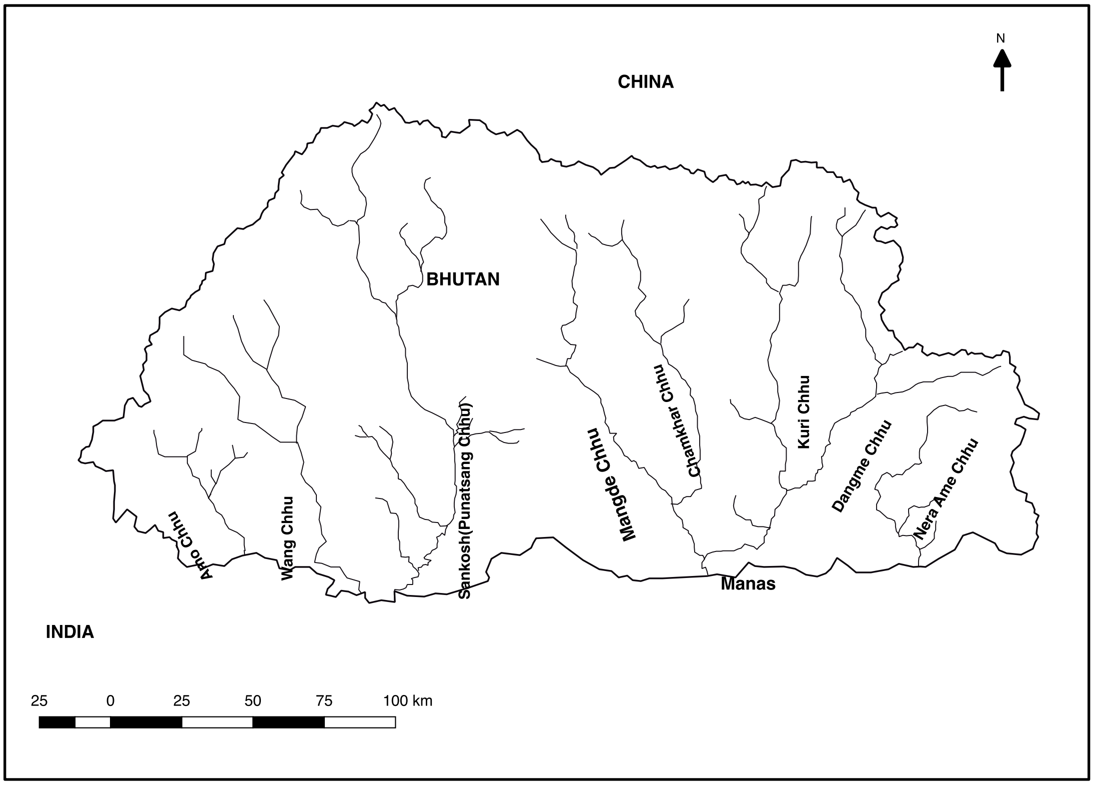

Figure 2: Five major drainage basins in Bhutan (Eastern Himalayas). From west to east: Amo Chhu, Wang Chhu, Punatsang Chhu/Sunkosh, Manas River (includes, Mangde, Chamkhar, Kuri and Dangme Chhu), Nyera Ama Chhu. Each river drains into the Lower Brahmaputra River in India.

Chapter II

Morphological and genetic variability in a species-complex of Central-Himalayan fish (Cyprinidae: *Schizothorax*) inhabiting Lake Rara in Nepal

Abstract

Snowtrout *(Schizothorax*) is a genus of minnow native to the Himalayan Region. Out of five species native to the Central Himalayas, three are endemic to Lake Rara in northwestern Nepal. I examined the morphological and genetic variation of *Schizothorax* in Lake Rara and compared these lacustrine species with two more widely distributed riverine species in adjacent drainages. To evaluate shape differences, I digitized Cartesian coordinates of 18 type I landmarks from images of 565 museum specimens and conducted a principal component analysis (PCA) and multivariate analysis of variance (MANOVA) to explore variation in morpho-space. A morphological distance matrix (MDM) was estimated across species to evaluate relationships between taxonomy (species) and geography (basins). To quantify genetic variation, 1,140 base pairs of the mitochondrial cytochrome b gene were sequenced from 115 archived tissue samples and a haplotype network constructed to examine genetic relationships within a geographic framework. There was partial concordance between genetic and morphometric data. The two riverine species in the Koshi River Basin diverged in morpho-space from other riverine species, a pattern consistent with their separate grouping in a haplotype network. Samples from the Gandaki and Karnali rivers grouped by morpho-type, independent of geographic origin, but clustered by basin within the haplotype network. Among the lacustrine species, *S. raraensis* was the most morphologically-diverged, but nestled among other samples from the Karnali River in the haplotype network and could not be genetically distinguished from the other Lake Rara endemics. While phylogenetic relationships of *Shizothorax* from the Central Himalayas remain

unresolved, these data provide insight into the variation of endemic fishes in Lake Rara and their relationship with species distributed in adjoining rivers.

Introduction

The Himalayas exhibit unparalleled ecological variation associated with the elevational range, shaped by rapidly elevating mountains and characterized by fast flowing rivers and pristine habitats in isolated mountain chains. These diverse ecological conditions along rivers and streams could be the strong driver towards morphological and genetic differentiation (He & Chen, 2006; Bhatt et al., 2012; Liang et al., 2017; Pellissier et al., 2017). Separation of small communities in remote physiographic settings often promotes endemism in terrestrial species and could be even more pronounced in freshwater species due to isolation by longer river distances and dispersal barriers such as deep gorges, falls and rapids (Douglas et al., 2016; Guo et al., 2016). In fact, morphological diversity among closely related fish species in Himalayan streams has been associated with environmental variables such as habitat type and its complexity (Li et al., 2009). However, diversity of freshwater organisms associated with fast flowing headwater streams and steady flowing lowland rivers, as well as lake habitats, has largely been left uncovered, particularly in the Himalayan Region. As a consequence, endemic biodiversity may be profoundly underestimated and not recognized taxonomically, impeding effective conservation across the region.

Conventionally, taxonomy heavily relied on comparative morphology and anatomy. Over the past two decades, taxonomic studies have increasingly incorporated molecular approaches, partially due to the ease with which genetic data can be generated (Avise, 2004). Approaches have morphed from more expensive and time-consuming techniques such as allozyme analysis to gene sequencing technologies that provide higher resolution (e.g., Sanger sequencing). Recently, the advent of massif-parallel sequencing technologies (NGS: Next Generation Sequencing) has further accelerated data generation, but also facilitates the generation of unprecedented amounts of information about genetic diversity. The advantage of molecular approaches is their capacity for characterizing evolutionary relationships and delimit diversity across all taxonomic hierarchical levels, from order and family to populations (Blaxter, 2004). Although both morphological and molecular data can convey phylogenetic signals, the juxtaposition of the two approaches offer unique perspectives, yet has been adopted reluctantly (Lefébure, 2006; Douglas et al., 2007; Davis et al., 2016).

Geometric morphometrics (GM) has recently gained popularity with new analytical methods offering comparative analyses of morphological shape variation with molecular data at high statistical power (Rohlf, 1998; Zelditch et al., 2012). High-resolution images from both microscopic and macroscopic organisms serve as the source for extracting two or three-dimensional quantitative data that encapsulate shape variation. The shape reflects adaptive convergence or divergence in response to particular habitat condition among various taxonomic groups (Barlow et al., 1997; Jones, 2008). In particular head and body shape are critical drivers for niche separation and reproductive isolation and reflect adaptive divergence in feeding, mating, and locomotion (Russo, 2008; Daniele et al., 2017).

There are few studies on the morphological variation of Snowtrout in the Himalayan Region, and those that are available have been non-synthetic, in that specimens were not assessed concerning distinct ecological niches or variable habitats, such as upstream *versus* downstream habitats or lotic *versus* lacustrine systems. Instead, analyses were more prosaic in that characters were simply sought to potentially discriminate species. For example, anal fin length seemingly distinguished between two Snowtrout species (*S. richardsonii* and *S. progastus*) in the Upper Ganges River of northwest India (Pandey & Nautiyal, 1997). Similarly, Mir et al. (2013) applied a truss network analysis involving 14 landmarks to identify significant phenotypic variability in a single species (*S. richardsonii*) distributed in the central and western Himalayan rivers of India. A similar truss evaluation (Pradhan & Wagle, 2013) quantified differences in the same species but distributed across three tributaries of the Gandaki River, Central Himalayas. Terashima (1984) published meristic data on three Snowtrout species endemic to Lake Rara in northwestern Nepal (Figure 1). A re-analysis of these data (B. Regmi, unpubl.) indicated that the three endemic species could only be discriminated according to the number of outer gill rakers**.**

Here, my goal was to test the predictability of the morphological species and geographic association of their shape. In this context, I acquired GM shape data across five *Schizothorax* OTUs (operational taxonomic units) using 18 anatomical landmarks encapsulating head and body shape variation. I applied multivariate morphometric analyses as a means to potentially distinguish OTUs in pairwise comparisons, one from another. Samples represented three OTUs endemic to a lentic system (Lake Rara) as well as two OTUs distributed in adjacent lotic systems. Shape variation was juxtaposed with molecular data to evaluate character congruence, quantify the extent of diversification and speciation of Snowtrout in Lake Rara and to provide a blueprint with regard to species delimitation within the Central Himalayas. I addressed two fundamental questions: (1) How morphologically distinct are endemic *Schizothoax* in Lake Rara from the species distributed in rivers and streams in the Central Himalayan Region based on shape variation? and (2) Is shape variation consistent with the genetic variation?

Methods

Study species

Snowtrout (Cypriniformes, Cyprinidae, subfamily Schizothoracinae, genus *Schizothorax)* is a minnow widely distributed in the Himalayan waters, but not related to coldwater fishes generally known as 'trout' (Salmoniformes). Superficially similar in shape to a trout with its elongated body and broad peduncle (but lacking an adipose fin and a subterminal rather than a terminal mouth) it is named Snowtrout because of its distribution in the Himalayan and Tibetan cold-water rivers that are primarily fed by snowmelt (Terashima, 1984; Jayaram, 1999; Shrestha, 1999; Dimmick & Edds, 2002). Snowtrout are found at elevations ranging from 100m-5,000m above mean sea level in Central Asia, Tibet, South and South-east Asia (Talwar & Jhingran, 1991; Edds, 1993; Jayaram 1999; Shrestha & Edds, 2012).

The genus *Schizothorax* is species-rich comprising more than 30 species. Although the Cyprinidae is recognized as the largest and most diverse family of fishes, very few of cyprinids have been considered species flocks (Dimmick & Edds, 2002). A noteworthy exception is three lacustrine species of Snowtrout, *S. macrophthalmus*, *S. nepalensis, and S. raraensis,* recognized as a cyprinid species flock restricted to a single lake (Lake Rara) in the western part of the Central Himalayan Region in Nepal (Terashima, 1984). Two other species incorporated in this study, the more common, riverine *S. richardsonii and S. progastus*, are distributed more widely in river drainages of the Central Himalayas (Terashima, 1984; Dimmick & Edds, 2002; Edds, 2007).

Data generation

The mitochondrial (mt)DNA cytochrome b gene is widely used for reconstructing phylogenetic relationships and evaluating genetic variation within a geographic context (i.e., phylogeography; Avise, 2000). A total of 115 tissue samples representing five species of *Schizothorax* were obtained from the University of Kansas Natural History Museum. Genomic DNA was extracted using the Qiagen DNeasy kit using manufacturers protocol with minor modification. Species and sample sizes as follows: *S. richardsonii* (N=50); *S. progastus* (N=20); *S. macrophthalmus* (N=15); *S nepalensis* (N=16); *S. raraensis (*N=14) (Table 1). PCR-amplifications were performed using published primers (Unmack et al., 2009; Houston et al., 2012) and conditions, visualized on agarose gels and enzymatically purified. Sanger sequencing was conducted using BigDye (ver.3.1) chemistry [Applied Biosystems Inc. (ABI), Forest City, CA] and recorded on an ABI Prism 3730 Genetic Analyzer (W.M. Keck Center for Comparative and Functional Genomics, University of Illinois, Urbana/Champaign). Sequences were aligned manually using SEQUENCHER (ver. 5.0, Gene Codes, Ann Arbor, MI). Five published sequences representing *S. macrophthalmus* (N=1), *S. nepalensis* (N=2), and *S. richardsonii* (N=2) (Table 2) were added from GenBank (Yang et al., 2015).

To quantify shape variation, digital images were captured using a high-resolution camera with optimum lighting condition to ascertain consistent and accurate recording of morphological data. A total of 565 images was captured from museum specimens to obtain shape data for all five species. Specimens had been sampled in 1984-1986 from three major river drainages in Nepal (i.e., Gandaki, Koshi, and Karnali rivers) and 1996 from Lake Rara (Edds, 1989) (Figure 1). Image capture was performed using a Canon (T3 Rebel) digital camera with a 300mm lens. A photographic box, designed for optimal light and exposure, was used to house the specimens as they were photographed (LED Pro Box Plus 1419 Photography Lighting Studio, MK Digital Direct, San Diego, CA 92154). The left lateral view was recorded by placing specimens in a V-shape photarium filled with water to standardize positioning, and distorted specimens and juveniles were not employed in the study.

Determining homologous landmarks is an important step in the analysis of morphological shape variation. Eighteen type I landmarks representing homologous anatomical features (Figure 2) were digitized using the software TPDDIG (Rohlf, 2004) to obtain two dimension Cartesian coordinates. Landmarks were selected to adequately capture head morphology, i.e., snout elongation, head depth, and placement of the eyes and nostrils (Figure 2). The body shape was primarily defined by origin/ insertion of fins (dorsal, pelvic, and pectoral) and upper / lower base of the caudal fin, with the urostyle designating the posterior most point (Figure 2). The program TPSUTIL (F.J. Rohlf, http://life.bio.sunysb.edu/ee/rohlf/software.html) was used to prepare the tps file (i.e. specific file format for storing morphometric data, Adams et al., 2013a) containing the x/y-coordinates of all landmark for each image, extract the list of images and append the images. GEOMORPH v3.05 (Adams et al., 2013a), an R-based software package, was employed to read and superimpose the coordinate data using Generalized Procrustes Analysis (GPA: Adams et al., 2013b).

Biases due to digitization errors and the effect of sex

Digitization errors could swamp the signal if shape variation associated with species and/or basins is subtle. To estimate the size of the error, I digitized nine individuals of each species from three basins 10 times (n=90) and implemented Hotelling's *T-square* statistics (HT^2), where I standardized t with the number of features (18x2 landmark co-ordinates) and replicates (number of specimens). After standardization, the data followed the *F*-distribution. The test revealed that the variance attributable to differences among species or basins is larger than variance introduced due to digitization errors.

The gender of most museum specimen was unknown, and it was not possible to determine sex based on internal anatomy or gonad development. To eliminate potential biases due to sexual dimorphism, I classified individuals with breeding tubercles as ‘male’ (M) and others as ‘sex unknown’ (SU). Anecdotal observations indicate that female Snowtrout do not develop breeding tubercules (D. Edds, pers. comm.). I implemented Hotelling's *T-square* test to evaluate the significance of gender-based shape variation between M and SU.

Data analysis

My data analyses involved four fundamental approaches. First, I used principal component analysis (PCA) to explore the distribution of total morphological variance in morpho-space (i.e., a set of data points across mutually orthogonal axes). PCA extracts several sets of data that contribute to morphological variation independent of predetermined hypotheses (e.g., taxonomic identification of a specimen). In contrast, multivariate analysis of variance (MANOVA) tests the statistical significance of morphological variation in the context of taxonomic (species) or geographic (basins) designations. Third, hierarchal clustering of species based on the Euclidean distance of group mean vectors was implemented to estimate and visualize the differences among species in morpho-space. Finally, I constructed a haplotype network based on the cytochrome b sequence data and assigned each specimen to a haplogroup to compare morphological differences within the context of phylogeographic patterns (i.e., genetically distinct lineages due to taxonomic and/or geographical parameters). Although genetic data could only be obtained for specimens with archived tissue (N=115), all specimens representing a particular species from a location were assigned to the same haplogroup for morphometric analyses.

To visualize morphological variation in tangent space I used GEOMORPH to generate deformation grids (Thin Plate Splines) reflecting shape variation along the x/y-axes incorporating multiple principal components (Figure 3). These deformation grids (DG) were examined to infer shape differences among species and basins and assess the contribution of each variable on major components.

Implementing principal component analysis at multiple groupings provides detailed visualization of how variables are loaded onto these components. Therefore, PCA was conducted for three groupings: (A) across all specimens, considering taxonomic designation (species) independent of location (basins); (B) for each species separately across locations (basin); and (C) across all specimens, considering both taxonomic designation (species) and location (basin). Since each grouping used a different set of specimens, GPA was implemented before visualization of each analysis.

To examine whether shape variation is attributable to taxonomy or location, I partitioned data by species, basins, or both, and conducted a multivariate analysis of variance (MANOVA) on the GPA aligned data. I implemented MANOVA for Goodall's *F*-test. The significance of the *F*-ratio was tested with 1,000 permutations. The MANOVA for the multivariate dataset is similar to distance-based Procrustes ANOVA (Anderson, 2001). Since there is no guarantee of obtaining a positive definite matrix in morphometric data, MANOVA that requires an estimate of the covariance matrix is not practical.

Obtaining the multivariate distance matrix based on putative species and their geographic distribution helps to obtain their relationship in standardized shape space. I used the *mshape* function (Claude, 2008) in GEOMORPH to calculate the mean vector of GPA-aligned coordinates for each species and *dist* function in R to estimate the Euclidian distance matrix. Based on this distance matrix, I hierarchically clustered 'species' and 'species by basin' (Figure 4) based on their shape variation in a cladogram using the *hclust* with the *single* link function in R.

A haplotype network helps to visualize genetic variation among species reflecting their geographical distribution. The cytochrome b gene sequences were aligned using MUSCLE V.3.8 (Edgar, 2004) and sequences collapsed into haplotypes in DNASP V. 6 (Rozas et al., 2017). A haplotype network was constructed using median-joining (Bandelt et al., 1999) and visualized using POPART (Leight & Bryant, 2015) to depict the distribution of the five putative morphological species across haplogroups.

Results

Out of approximately 800 digital images taken of museum vouchers, 565 images were selected for geometric morphometric analysis after removing distorted, ill-preserved and juvenile specimens. Table 1 represents the number of museum specimens analyzed for each species. The two common riverine species distributed in three major watersheds were represented by larger sample sizes compared to lacustrine species endemic to Lake Rara (Figure 1).

Biases due to digitization errors and sex

The Hotelling's *T*-test was not significant between two mean vectors of classes ['male' (M) *versus* 'sex unknown' (SU), p=0.39, df=36]. An analysis conducted to evaluate shape variation due to digitization errors showed higher variance among species (between groups) than among individuals within groups (replicate digitization of the same images) indicating that variation due to digitization errors was negligible.

Principal components of shape variation across basins and species

Shape data comprised 2-dimensional Cartesian coordinates (x/y) representing 18 landmarks (Figure 2), resulting in 36 Principal Components (PC). Of these, only 32 PC had non-zero eigenvalues; however, the four zero eigenvalue PCs were not removed from the data matrix because their contribution to computational complexity is negligible, and more importantly, a landmark coordinate with computationally zero eigenvalues for one axis may have non-zero eigenvalue in another axis.

The first principal component explained 30% and the second 14.4% of the shape variation (Figure 5) and 74% of the variation was loaded onto the first six PCs. The majority of variables loaded with equal proportion onto PC1, whereas loading for PC2-PC6 were significantly associated with a subset of traits. Loadings for PC2 primarily reflected landmarks associated with head shape (i.e., tip of the mouth, ventral end of operculum, left end of the mouth opening), for PC3 were dominated by traits associated with the placement of fins (i.e., origin and insertion of the dorsal and pectoral fin), for PC4 with the placement of the eye and urostyle (posterior end of vertebrae column), PC5 with orbital edges, neurocranial end, and mouth opening, and PC6 the insertion of the anal fin with a significant proportion.

A plot contrasting shape variation along PC1 and PC2 among all species shows broad overlap, with the two riverine species separating along PC1 (Figure 5A). Contrasting only the three species endemic to Lake Rara shows clear separation of *S. raraensis* and S.

macrophthalmus, whereas *S. nepalensis* is embedded into *S. macrophthalmus* (Figure 5B). Plots contrasting shape variation along PC1 and PC2 within riverine forms among the three basins show clear separation of samples from the Koshi River compared to Gandaki, and Karnali River samples in *S. richardsonii* (Figure 5C) and some distinction of *S. progastus* samples from the Gandaki River compared to Karnali and Koshi rivers (Figure 5D). The deformation grid (DG) of a lacustrine species (*S. nepalensis,* Lake Rara) and riverine species (*S. progastus,* Koshi River Basin) along PC1 and PC2 are presented in Figure 3. The first DG (PC1, variation=0.40) shows the shape change from the reference form (average configuration Figure 3A) along most of the landmarks. The second DG (PC2, variation=0.18) reflects shape variation between the two species along the landmarks associated with the upper end of the operculum, the mouth opening and placement of pectoral fins (Figure 3B).

Multivariate analysis of shape variation

Testing models reflecting either taxonomic (species) and/or geographic (basins) hypotheses to explain shape variation, MANOVA revealed low but significant regression coefficients for both parameters. However, the strength of the dependency of shape on 'species' was stronger than on 'basins' (Table 3).

Mean vectors of the Euclidian distance (average configuration, Claude et al., 2008) derived from pairwise comparisons of either species alone or species partitioned by basins are provided in Table 4 and Table 5, respectively. Hierarchal clustering of these mean vectors is shown in Figure 4. The cluster dendrogram of species (Figure 4A) shows that the riverine *S. richardsonii* and *S. nepalensis* cluster together, as do the lacustrine *S. macrophthalmus* and *S. progastus*. Surprisingly, the third Lake Rara endemic, *S. raraensis* clusters as the basal clade in the morphological space (Figure 4A). When partitioning the riverine *S. richardsonii* and *S. progastus* by drainage to examine morphological variation in a geographic context, a different pattern emerged (Figure 4B). In this more nuanced perspective that tests for potential effects of geographic isolation in the Himalayan mountains, the Lake Rara endemic *S. nepalensis* clustered with *S. richardsonii* from the Gandaki and Karnali River basins. Similarly, the other riverine species, *S. progastus,* reflected shape variation to be similar among specimens from the Gandaki and Karnali rivers, but clustered with another Lake Rara endemic, *S. macrophthalmus*. The third species endemic to Lake Rara, *S. raraensis,* clustered with specimens from the eastern-most basin, the Koshi River, although shape among these three groups appeared to be more divergent from one another.

Genetic variation

All cytochrome b sequences used in this study were 1,140 base pairs long. The nucleotide composition was (T=29.1%, C=27.2%, A=27.0%, G=16.7%). A total of 167 polymorphic sites were detected, with 20 sites representing singletons (only found in a single specimen), and 147 parsimony-informative sites. Nucleotide diversity was relatively high (π=0.012). A total of 89 haplotypes were identified (Table 7), that segregated into three major haplogroups associated with the three basins (Figure 6). All Lake Rara endemic species clustered together with haplotypes from the Karnali River basin, the drainage within which the lake is located. Interestingly, the same species in different basin associated with different haplogroups (Figure 6). There was a strong geographic structure in haplogroups. Importantly, one haplotype was shared among all lacustrine (Lake Rara) and riverine samples from Karnali Basin.

Discussion

Fishes are known to exhibit high intra-specific morphological variation that could have both ecological and evolutionary significance (Robinson & Wilson, 1996; Clabaut et al., 2007). Often, this variation is associated with sexual dimorphism, with well-known examples including pronounced differences in body shape between males and females in livebearing fishes (Hassel et al., 2012). Although sexual dimorphism among *Schizothorax* species in Lake Rara has been described (Terashima, 1984), none of those divergent traits were related to head or body shape. I tested for patterns of sexual dimorphism in my data set but was unable to discriminate sex based on the shape variables and thus decided to analyze morphological variation without considering sex as a source of bias.

Head morphology is well known to be associated with feeding performance and modifications to head shape, mouth position, among many other head features, have been described as drivers of trophic divergence (Ornelas-García et al., 2018). Snowtrout species endemic to Lake Rara occupy different food niches: *S. macrophthalmus* is planktivorous, *S. nepalensis* is herbivorous and *S. raraensis* is insectivorous (Terashima, 1984). This trophic diversification is reflected in their head morphology: *S. macrophthalmus* and *S. raraensis* have oblique mouths, whereas *S. nepalensis* has a subterminal mouth, considered an adjustment for scraping sessile algae. However, these observations are based on qualitative assessments (Terashima, 1984) and have not been quantified. My statistical analysis of shape variation in *Schizothorax* of Lake Rara and adjacent rivers (Figure 5) revealed that while PC1 loadings were distributed across all landmarks, PC2 loadings were strongly associated with the landmarks describing head shape. Based on shape analysis of a substantial number of specimens (Table 1) I uncovered a distinct morphological trajectory for *S. raraensis* compared to other Snowtrout distributed across the Central Himalayas (Figure 4). In contrast, the other two species endemic to Lake Rara were similar in shape to riverine species, with *S. nepalensis* closer to the blunt-nosed *S. richardsonii* and *S. macrophthalmus* closer to *S. progastus*, both with more elongated snouts.

The riverine *S. richardsonii,* a detritophagic bottom feeder of algae growing on gravel and boulders, is distributed upstream in headwaters of Himalayan rivers (Sharma, 1984; Wagle, 2016). The other riverine species, *S. progastus*, is found in downstream reaches and relies on insects and diatoms suspended on the water. This difference in feeding mode is clearly reflected in their head shape, and similar diet preferences are likely responsible for trophic morphology and head shape variation in the lacustrine species endemic to Lake Rara.

Other traits strongly associated with PC loadings were the placement of dorsal and pectoral fins, traits associated with locomotion. This suggests that adaptation to high gradient streams in upper catchment areas *versus* gently flowing rivers in the foothill region of the Himalayas drove morphological divergence of the two riverine species, *S. richardsonii* and *S. progastus*. Although a MANOVA of multivariate shape variation based on both taxonomy (species) and geography (basin) was significant, ‘species’ was a stronger predictor compared to the ‘basin’ (Table 3).

To date, no study has been published that combined morphological and molecular data to delimit freshwater fish diversity in isolated headwaters of Tibetan Plateau and the Himalayan Region. A few comprehensive studies examined phylogeography and genetic variation of Snowtrout in the region. Using sequence analysis of cytochrome b data, He and Chen (2006) examined phylogenetic relationships of 30 species of *Schizothorax* distributed in nine rivers originating on the Tibet Plateau and draining into the different oceans. Genetic clades were inconsistent with traditional morphology-based taxonomy but instead reflected geographic associations with basins, leading authors to speculate that incomplete lineage sorting may be responsible for the discrepancies between taxonomy and molecular data. In a more localized study based on sequence analysis of the mitochondrial D-loop region and tRANs, Dimmick & Edds (2002) reported substantial differentiation of *Schizothorax* among Koshi, Gandaki, and Karnali River basins in Nepal.

My study of morphological variation of *Schizothorax* dispersed in the Central Himalayan Region showed partial concordance with phylogeographic signals derived from sequence analysis of cytochrome b data. Strong geographic structuring of haplotypes in a network analysis (Figure 6) supports previous findings that *Schizothorax* species in isolated headwaters of Tibetan and Himalayan rivers are genetically differentiated based on drainage basin. Furthermore, sharing of haplotypes among four of my study species, including three lacustrine (*S. macrophthalmus, S. nepalensis, and S. raraensis*) and one riverine form (*S. richardsonii)* in the Karnali Region of western Nepal, indicates potential hybridization or incomplete lineage sorting. In contrast, populations in the Koshi River Basin in eastern Nepal appear to be more divergent both in morphology and with regards to genetic diversity (Figures 3B, 5D and 6).

Conclusion

Although species endemic to Lake Rara appear to be morphological differentiated, their distinctness was not supported by molecular data analyzed in this study. This warrants further investigation with approaches that allow distinguishing hybridization, incomplete lineage sorting or reticulated evolution (Douglas et al., 2010; Bangs et al., 2018) as potential causes of incongruence between morphological and genetic diversity of lacustrine and riverine fishes in Nepal. An important outcome of this study is that the substantial differentiation of *Schizothorax* species in Koshi River basin merits its recognition as significant evolutionarily unit (ESU; Avise, 1989; Mortiz, 1994; Douglas & Brunner, 2002). This should be considered in future management plans and help guide conservation efforts to ensure persistence of this intra-specific diversity of minnows uniquely adapted to Himalayan rivers in Nepal.

Acknowledgments

Museum specimens and tissue samples for this study were obtained from the University of Kansas, Natural History Museum. We thank Curator Leo Smith and Collection Manager Andrew Bentley for their prompt response and support for tissue loans and allowing me to use their imaging facility in Lawrence, Kansas. Dr. Mark Davis provided guidance for getting started with morphometric data analysis. The endless effort of Dr. David Edds during his ichthyological explorations in Himalayan streams of Nepal and systematic collection and vouchering of museum specimens and tissue samples made this study possible. This research was made possible through generous endowments to the University of Arkansas: The Bruker Professorship in Life Sciences (MRD), and the 21st Century Chair in Global Change Biology (MED), as well as a Graduate Assistantship (BR), all of which provided salaries and/or research funds for this work.

References:

Adams, D. C., and Otárola-Castillo, E. (2013a). GEOMORPH: An R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4: 393-399. doi: 10.1111/2041-210X.12035.

Adams, D. C., Rohlf, F. J., and Slice, D. (2013b). A field comes of age: Geometric morphometrics in the 21st century. *Hystrix*, 24:7-14. doi: 10.4404/hystrix-24.1-6283.

Anderson, M. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol*ogy, 26:32-46.

Avise, J. C. (1989). A role for molecular genetics in the rec- ognition and conservation of endangered species. *Trends in Ecology and Evolution,* **4**:279-281.

Avise J. C. (2000). *Phylogeography. The history and formation of species*. Harvard University Press, Cambridge, MA.

Avise, J. C. (2004). *Molecular Markers, Natural History, and Evolution*. Sunderland MA: Sinauer and Associates, Inc.

Bandelt, H., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*,16:37-48.

Bangs, M. R., Douglas, M. R., Mussmann, S. M., and Douglas, M. E. (2018). Unraveling historical introgression and resolving phylogenetic discord within *Catostomus* (Osteichthys: Catostomidae). *BMC Evolutionary Biology*, 18, 86. doi:10.1186/s12862-018-1197-y.

Barlow, K. E., Jones, G., and Barratt, E. M. (1997). Can skull morphology be used to predict ecological relationships between bat species? A test using two cryptic species of *Pipistrelle*. *Proceedings of the Royal Society of London B*, 264:1695-1700. doi: 10.1098/rspb.1997.0235.

Bhatt J. P., Kumar, M., and Pandit, M. K. (2012). Elevational gradients in fish diversity in the Himalaya: water discharge is the key driver of distribution patterns. *PLoS One* 7: e46237.

Blaxter, M. L. (2004). The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359:669-679.

Clabaut, C., Bunje, P. M., Salzburger, W., and Meyer, A. (2007**).** Geometric morphometric analyses provide evidence for the adaptive character of the Tanganyikan cichlid fish radiations. *Evolution*, 61:560-578.

Claude, J. (2008). *Morphometrics with R*. Springer, New York.

Daniele, V., Bonhomme, V., Colangelo, P., Bonifazi, A., Lasinio, A. J., and Ardizzone, G. (2017). Does morphology predict trophic niche differentiation? The relationship between feeding habits and body shape in four co-occurring juvenile species (Pisces: Perciformes, Sparidae). *Estuarine, Coastal and Shelf Science,* 191: 84-95.

Davis, M. A., Douglas, M. R., Collyer, M. L., and Douglas, M. E., (2016). Deconstructing a species-complex: geometric morphometric and molecular analyses define species in the Western Rattlesnake (*Crotalus viridis*). *PloS One*, *11*, p.e0146166.

Dimmick, W. W., and Edds, D. R. (2002). Evolutionary genetics of the endemic schizorathicine (Cypriniformes: Cyprinidae) fishes of Lake Rara, Nepal. *Biochemical Systematics and Ecology*, 30:919-929. doi: 10.1016/S0305-1978(02)00030-3.

Douglas, M. E., Douglas, M. R., Schuett, G. W., Porras, L. W., and Thomason, B. L. (2007). Genealogical concordance between mitochondrial and nuclear DNAs support species recognition of the Panamint Rattlesnake (*Crotalus mitchellii stephensi*). *Copeia*, 2007:920-932. doi: 10.1643/0045- 8511(2007)7[920:GCBMAN]2.0.CO;2.

Douglas, M. R., and Brunner, P. C. (2002). Biodiversity of Central Alpine *Coregonus* (Salmoniformes): Impact of one-hundred years management. *Ecological Applications,* 12:154-172.

Douglas, M. R., and Douglas, M. E. (2010). Molecular approaches to stream fish ecology, pp. 157-195. *In: Community Ecology of Stream Fishes*. K. Gido and D. Jackson (eds.). *American Fisheries Society Symposium,* 73.

Douglas, M. R., Davis, M. A., Amarello, M., Smith, J. J., Schuett, G. W., Herrmann, H-W., Holycross, A. T., and Douglas M. E. (2016). Anthropogenic impacts drive niche and conservation metrics of a cryptic rattlesnake on the Colorado Plateau of western North America. *Royal Society Open Science*, 3, 160047. doi: 10.1098/rsos.160047.

Edds, D. R., (1989). Multivariate analysis of fish assemblage composition *and environmental correlates in a Himalayan River-Nepal's Kali Gandaki/Narayani*. Doctoral dissertation, Oklahoma State University.

Edds, D. R. (1993). Fish assemblage structure and environmental correlates in Nepal's Gandaki River. *Copeia*, 1993:48-60.

Edds, D. R. (2007). Fishes in Nepal: ichthyofaunal surveys in seven nature reserves. *Ichthyological Exploration of Freshwaters*, 18:277-287.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32:1792-97 doi:10.1093/nar/gkh340. PMC 390337

Guo, X. Z., Zhang, G. R., Wei, K. J., Yan, R. J., Ji, W., Yang, R. B., Wei, Q. W., and Gardner, J. P. (2016). Phylogeography and population genetics of *Schizothorax o'connori*: strong subdivision in the Yarlung Tsangpo River inferred from mtDNA and microsatellite markers. *Scientific Reports*, 6:29821.

Hassell, E. M., Meyers, P. J., Billman, E. J., Rasmussen, J. E., and Belk, M. C. (2012). Ontogeny and sex alter the effect of predation on body shape in a livebearing fish: sexual dimorphism, parallelism, and costs of reproduction. *Ecology and Evolution*, 2:1738-1746.

He, D., and Chen, Y. (2006). Biogeography and molecular phylogeny of the genus *Schizothorax* (Teleostei: Cyprinidae) in China inferred from cytochrome b sequences. *Journal of Biogeography*, 33:1448-1460. doi:10.1111/j.1365-2699.2006.01510.x

Houston, D. D., Evans, R. P., and Shiozawa, D. K. (2012). Evaluating the genetic status of a Great Basin endemic minnow: the relict dace (*Relictus solitarius*). *Conservation Genetics,* 13:727-742.

Jayaram, K. C. (1999). *The Freshwater Fishes of the Indian Region*. Narendra Publishing House, Delhi.

Jones, M. H. (2008). Skull shape and feeding strategy in *Sphenodon* and other Rhynchocephalia (Diapsida: Lepidosauria). *Journal of Morphology,* 269: 945-966. doi: 10.1002/jmor.10634.

Lefébure, T., Douady, C. J., Gouy, M., and Gibert, J. (2006). Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*, 40:435-447.

Leigh, J. W., and Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. *Methods Ecology and Evolution,* 6:1110-1116.

Li, J., He, Q., Hua, X., Zhou, J., Xu, H., Chen, J., and Fu, C. (2009). Climate and history explain the species richness peak at mid-elevation for *Schizothorax* fishes (Cypriniformes: Cyprinidae) distributed in the Tibetan Plateau and its adjacent regions. *Global Ecology and Biogeography*, 18:264-272. doi:10.1111/j.1466-8238.2008.00430.x

Liang, Y., He, D., Jia, Y., Sun, H., and Chen, Y. (2017). Phylogeographic studies of schizothoracine fishes on the central Qinghai-Tibet Plateau reveal the highest known glacial microrefugia. *Scientific Reports*, 7:10983.

Mir, F., Mir, J. I., and Chandra, S. (2013). Phenotypic variation in the Snowtrout *Schizothorax richardsonii* (Gray, 1832) (Actinopterygii Cypriniformes: Cyprinidae) from the Indian Himalayas. *Contributions to Zoology,* 82-3.

Moritz, C. (1994). Defining ‘‘evolutionary significant units’’ for conservation. *Trends in Ecology and Evolution*, 9:373– 375.

Ornelas-García, C. P., Córdova-Tapia, F., Zambrano, L., Bermúdez-González, M. P., Mercado-Silva, N., Mendoza-Garfias, B., and Bautista, A. (2018). Trophic specialization and morphological divergence between two sympatric species in Lake Catemaco, Mexico. *Ecology and Evolution*, 8*:*4867-4875.

Pandey, S. K., and Nautiyal, P. (1997). Statistical evaluation of some meristic and morphometric characters of taxonomic significance in *Schizothorax richardsonii* (Gray) and *Schizothorax plagiostomus* (Heckel). *Indian Journal of Fisheries,* 44:75-79.

Pellissier, L., Heine, C., Rosauer, D. F., and Albouy, C. (2017). Are global hotspots of endemic richness shaped by plate tectonics? *Biological Journal of the Linnean Society*, 123:247-261.

Pradhan, N., and Wagle, S. K. (2013). Landmark-based morphometric variation among three populations of Asla (*Schizothorax richardsonii*) in rivers of Nepal. *Livestock and Fisheries Research,* 30:15.

Robinson, B. W., and Witson, D. S. *(1996).* Genetic variation and phenotypic plasticity in a trophically polymorphic population of pumpkinseed sunfish (*Lepomis gibbosus*). *Evolutionary Ecology,* 10:631-652. doi: 10.1007/BF01237711.

Rohlf, F. J. (1998). On applications of geometric morphometrics to studies of ontogeny and phylogeny. *Systematic Biology,* 47:147-158, doi: 10.1080/106351598261094.

Rohlf, F. J. (2004). *TpsDig*. Department of Ecology and Evolution, State University of New York, Stony Brook, NY.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., and Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Molecular Biology and Evolution,* 34:3299-3302. doi: 10.1093/molbev/msx248.

Russo, T., Pulcini, D., O'Leary, Á., Cataudella, S., and Mariani, S. (2008). Relationship between body shape and trophic niche segregation in two closely related sympatric fishes. *Journal of Fish Biology*, 73:809-828.

Sharma, R. C. (1984). Trophic dynamics of snowtrout, *Schizothorax richardsonii* (Gray) of Garhwal Himalaya. *Indian Journal of Animal Science,* 54:666-9.

Shrestha, J. (1999). Coldwater fish and fisheries in Nepal. In T. Petr (Ed.), *Fish and Fisheries at Higher Altitudes: Asia* (13-40). Rome, Italy: *FAO Fisheries Technical Paper,* 385.

Shrestha, O. M., and Edds, D. R. (2012). Fishes of Nepal: Mapping distributions based on voucher specimens. *Emporia State Research Studies,* 48:14-21.

Talwar, P. K., and Jhingran, A. G. (1991). *Inland Fishes of India and Adjacent Countries*. Oxford and IBH Publishing Co. Pvt. Ltd., New Delhi.

Terashima, A. (1984). Three new species of the cyprinid genus *Schizothorax* from Lake Rara, northwestern Nepal. *Japanese Journal of Ichthyology*, 31:122-135. doi: 10.11369/jji1950.31.122.

Unmack, P.J., Bennin, A., Habit, E. M., Victoriano, P. F., and Johnson, J. B. (2009) Impact of ocean barriers, topography, and glaciation on phylogeography of the catfish *Trichomycterus areolatus* (Teleostei: Trichomycteridae) in Chile. *Biological Journal of the Linnean Society* 97:876-892. doi:10.1111/j.1095-8312.2009.01224.x

Wagle, S. K., Pradhan, N., and Shrestha, M. K. (2016). Morphological divergence of Snowtrout (*Schizothorax richardsonii*, Gray 1932) from rivers of Nepal with insights from a morphometric analysis. *International Journal of Applied Sciences and Biotechnology,* 3:464-473.

Yang, L., T., Sado, M. Vincent Hirt, E. Pasco-Viel, M. Arunachalam, J. Li, X. Wang, J. Freyhof, K. Saitoh, A. M. Simons, M. Miya, S. H., and Mayden, R. L. (2015). Phylogeny and polyploidy: resolving the classification of cyprinine fishes (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution,* 85:97-116.

Zelditch, M. L., Swiderski, D. L., and Sheets, H. D. (2012). *Geometric Morphometrics for Biologists: A Primer, 2nd Edition*. New York: Academic Press, Inc.

Appendix

Table 1: Overview of *Schizothorax* specimens collected from three drainages in Nepal and analyzed in this study for morphological and genetic variation. Samples were identified as one of five species based on phenotype and geographic location, with *S. richardsonii* and S. *progastus* collected from rivers across three drainages, and *S. macrophthalmus, S. nepalensis,* and *S. raraensis* collected from Lake Rara (Figure 1). Specimens: number of museum vouchers examined for morphological variation using shape analysis; Tissues: number of DNA samples analyzed for sequence variation across the mitochondrial cytochrome b gene. Details on collection information listed in Table 6.

Species	**Vouchers**	**Tissues**
Schizothorax richardsonii	219	50
Schizothorax progastus	178	20
Schizothorax nepalensis	27	16
Schizothorax raraensis	59	14
Schizothorax macrophthalmus	82	15
Total	565	115

Table 2: List of five mitochondrial cytochrome b gene sequences obtained from the GenBank to include in a phylogeographic analyses of 115 *Schizothorax* from Nepal.

Species	**GenBank accession #**	**Location**	**Basin**
Schizothorax macrophthalmus	KP712253	Lake Rara	Karnali
Schizothorax nepalensis	NC031537	Lake Rara	Karnali
Schizothorax nepalensis	KP712254	Lake Rara	Karnali
Schizothorax richardsonii	KP712220	Melamchi	Koshi
Schizothorax richardsonii	KP712249	Indrawati River	Koshi

Table 3: Results of Multivariate Analysis of Variance (MANOVA) of shape variation across 565 *Schizothorax* collected from three drainages in Nepal. Data were based on Cartesian coordinates of 18 anatomical landmarks aligned within each group using Generalized Procrustes Analyses. Models tested whether variation was partitioned by (1) taxonomy. (Species only), (2) geography (Basin only), or both taxonomy and geography (Species + Basin). Geographic location of Basins is shown in Figure 1, with KOS=Koshi, GAN=Gandaki, KAR=Karnali and RAR=Lake Rara. Landmark configuration is depicted in Figure 2. Statistical significance levels were tested with 999 permutations.

Model	**DF**	**R-square**	**F-ratio**	**p-value (>F)**
1. Species	4	0.16	28.23	0.001
2. Basin				
(KOS, GAN, KAR, RAR)	3	0.08	17.69	0.001
3. Species + Basin				
Species	4	0.16	28.76	0.001
Basin	3	0.01	6.27	0.001

Table 4: Pairwise Euclidean distance matrix between five species of *Schizothorax* from Nepal. Distances were derived between average configurations of General Procrustes Analysis aligned landmark coordinates. The average configuration was estimated through minimizing the sum of squared norms of pairwise differences between the shapes of all samples within a group. MAC= *Schizothorax macrophthalmus*, NEP= *S. nepalensis*, PRO= *S. progastus*, RAR= *S. raraensis,* RIC= *S. richardsonii*.

	MAC	NEP	PRO	RAR
MAC				
NEP	0.02442958			
PRO	0.02326648	0.02850848		
RAR	0.03556793	0.04338788	0.04331892	
RIC	0.02380394	0.0194934	0.0278966	0.03999885

Table 5: Pairwise Euclidean distance matrix of average configuration of GPA aligned landmark coordinates for five species of *Schizothorax* sampled from three basins in Nepal. The average configuration is estimated through minimizing the sum of squared norms of pairwise differences between all shapes of all samples within a group. PRO_KOS= *S. progastus* from the Koshi River basin, PRO_GAN= *S. progastus* from the Gandaki River Basin, PRO_KAR= *S._progastus* from the Karnali River basin, RIC_KOS= *S. richardsonii* from the Koshi River basin, RIC_GAN= *S. richardsonii* from the Gandaki River basin, RIC_KAR= *S. richardsonii* from the Karnali River basin, RAR= *S. raraensis*, MAC= *S. macrophthalmus*, NEP= *S. nepalensis*.

	PRO_KO	**PRO_GA**	**PRO_KA**	**RIC_KO**	**RIC_GA**	**RIC_KA**	**MAC**	**RAR**
PRO_GA	0.0319							
PRO_KA	0.0301	0.0242						
RIC_KO	0.0376	0.0352	0.0404					
RIC_GA	0.0495	0.0292	0.0341	0.0332				
RIC_KA	0.0496	0.0333	0.0376	0.0271	0.0154			
MAC	0.0368	0.0257	0.0242	0.0308	0.0267	0.0273		
RAR	0.0466	0.0423	0.0419	0.0338	0.0417	0.0361	0.0297	
NEP	0.0494	0.0266	0.0359	0.0331	0.0204	0.0225	0.0241	0.0384

Table 6: List of specimens used to capture digital images for geometric morphometric analysis of shape variation among five species of *Schizothorax*. The accession number is the museum catalog number of University of Kansas (KU) or Oklahoma State University (OSU) collection.

Accession #	Museum collection	Species	Basin	Latitude	Longitude	River	N
29036	KU	*S. progastus*	Gandaki	27.9355	83.4399	Kaligandaki	3
28998	KU	*S. progastus*	Gandaki	27.92	83.68	Kaligandaki	2
27815	KU	*S. progastus*	Gandaki	28.3403	83.5692	Kaligandaki	1
15608	OSU	*S. progastus*	Gandaki	27.7037	84.4248	Gandaki	6
15709	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	7
15722	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	5
29018	KU	*S. progastus*	Gandaki	28.1533	83.6367	Kaligandaki	6
29024	KU	*S. progastus*	Gandaki	28.064	83.5607	Kaligandaki	2
29028	KU	*S. progastus*	Gandaki	27.9783	83.585	Kaligandaki	11
15731	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	19
15892	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	10
16349	OSU	*S. progastus*	Gandaki	28.3403	83.5692	Kaligandaki	7
16755	OSU	*S. progastus*	Gandaki	28.1753	83.6396	Kaligandaki	15
16928	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	5
16948	OSU	*S. progastus*	Gandaki	28.1753	83.6396	Kaligandaki	5
16763	OSU	*S. progastus*	Gandaki	28.3403	83.5692	Kaligandaki	1
16733	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	2
17464	OSU	*S. progastus*	Gandaki	27.7037	84.4248	Gandaki	2
16693	OSU	*S. progastus*	Gandaki	27.9783	83.585	Kaligandaki	1
16733	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	4
17505	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	2
16758	OSU	*S. richardsonii*	Gandaki	28.2026	83.6699	Modi	1
29023	KU	*S. richardsonii*	Gandaki	28.064	83.5607	Kaligandaki	5
29039	KU	*S. richardsonii*	Gandaki	27.9355	83.4399	Kaligandaki	1
28989	KU	*S. richardsonii*	Gandaki	27.8996	83.6333	Kaligandaki	1
16758	OSU	*S. richardsonii*	Gandaki	28.2026	83.6699	Modi	1
16766	OSU	*S. richardsonii*	Gandaki	28.2026	83.6699	Modi	2
15710	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	3
15732	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	4
16758	OSU	*S. richardsonii*	Gandaki	28.2026	83.6699	Modi	1
15710	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	3
15732	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	4
15891	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	1
16929	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	3
29624	KU	*S. richardsonii*	Gandaki	28.0333	84.0667	setigandaki	1

Table 6 (count..)

Accession #	**Museum collection**	**Species**	**Basin**	**Latitude**	**Longitude**	**River**	**N**
29043	KU	*S. progastus*	Gandaki	28.3403	83.5692	Kaligandaki	3
27887	KU	*S. richardsonii*	Gandaki	28.4	83.5667	Kaligandaki	5
27886	KU	*S. richardsonii*	Gandaki	28.3403	83.5692	Kaligandaki	2
29042	KU	*S. richardsonii*	Gandaki	28.3403	83.5692	Kaligandaki	4
16779	OSU	*S. richardsonii*	Gandaki	28.6841	83.6148	Kaligandaki	3
16780	OSU	*S. richardsonii*	Gandaki	28.6841	83.6148	Kaligandaki	2
29614	KU	*S. progastus*	Karnali	29.028	81.1819	Lower Karnali	1
17109	OSU	*S. progastus*	Karnali	29.028	81.1819	Lower Karnali	7
28780	KU	*S. progastus*	Karnali	28.3999	81.1999	Lower Karnali	1
28666	KU	*S. progastus*	Karnali	29.0816	80.1417	Mahakali	4
17109	OSU	*S. progastus*	Karnali	29.028	81.1819	Karnali	7
27884	KU	*S. progastus*	Karnali	29.5217	81.78	Karnali	21
27810	KU	*S. richardsonii*	Karnali	29.5216	81.93	Nijar	10
27885	KU	*S. richardsonii*	Karnali	29.5533	82.1699	MuguKarnali	12
29560	KU	*S. richardsonii*	Karnali	29.6483	81.805	Lower Karnali	11
29595	KU	*S. progastus*	Koshi	29.028	81.1819	Lower Koshi	1
29527	KU	*S. progastus*	Koshi	27.2716	87.2082	Arun	2
29595	KU	*S. progastus*	Koshi	26.93	87.3133	Arun	1
29527	KU	*S. progastus*	Koshi	27.2716	87.2082	Arun	2
27808	KU	*S. progastus*	Koshi	27.2716	87.2082	Arun	9
29539	KU	*S. progastus*	Koshi	27.2133	87.2399	Arun	2
29521	KU	*S. richardsonii*	Koshi	27.3199	87.2082	Arun	2
29528	KU	*S. richardsonii*	Koshi	27.2716	87.2082	Arun	2
29233	KU	*S. richardsonii*	Koshi	27.5517	87.3016	Khokta Khola	11
16779	OSU	*S. richardsonii*	Gandaki	28.6841	83.6148	Kaligandaki	3
16780	OSU	*S. richardsonii*	Gandaki	28.6841	83.6148	Kaligandaki	2
29614	KU	*S. progastus*	Karnali	29.028	81.1819	Lower Karnali	1
17109	OSU	*S. progastus*	Karnali	29.028	81.1819	Lower Karnali	7
28780	KU	*S. progastus*	Karnali	28.3999	81.1999	Lower Karnali	1
28666	KU	*S. progastus*	Karnali	29.0816	80.1417	Mahakali	4
17109	OSU	*S. progastus*	Karnali	29.028	81.1819	Karnali	7
27884	KU	*S. progastus*	Karnali	29.5217	81.78	Karnali	21
27810	KU	*S. richardsonii*	Karnali	29.5216	81.93	Nijar	10
27885	KU	*S. richardsonii*	Karnali	29.5533	82.1699	MuguKarnali	12
29560	KU	*S. richardsonii*	Karnali	29.6483	81.805	Lower Karnali	11
29595	KU	*S. progastus*	Koshi	29.028	81.1819	Lower Koshi	1
29527	KU	*S. progastus*	Koshi	27.2716	87.2082	Arun	2

Table 6 (cont…)

Accession #	**Museum collection**	**Species**	**Basin**	**Latitude**	**Longitude**	**River**	**N**
29595	KU	*S. progastus*	Koshi	26.93	87.3133	Arun	1
29527	KU	*S. progastus*	Koshi	27.2716	87.2082	Arun	2
27808	KU	*S. progastus*	Koshi	27.2716	87.2082	Arun	9
29539	KU	*S. progastus*	Koshi	27.2133	87.2399	Arun	2
29521	KU	*S. richardsonii*	Koshi	27.3199	87.2082	Arun	2
29528	KU	*S. richardsonii*	Koshi	27.2716	87.2082	Arun	2
29233	KU	*S. richardsonii*	Koshi	27.5517	87.3016	Arun	11
40564	KU	*S. richardsonii*	Koshi	27.8284	85.5769	Melamchi	38
27813	KU	*S. macropthalmus*	Karnali	29.5333	82.0667	Lake Rara	23
27812	KU	*S. macropthalmus*	Karnali	29.5333	82.0667	Lake Rara	24
29237	KU	*S. macropthalmus*	Karnali	29.5333	82.0667	Lake Rara	35
27811	KU	*S. nepalensis*	Karnali	29.5333	82.0667	Lake Rara	18
29234	KU	*S. nepalensis*	Karnali	29.5333	82.0667	Lake Rara	9
27807	KU	*S. nepalensis*	Karnali	29.5333	82.0667	Lake Rara	12
27814	KU	*S. nepalensis*	Karnali	29.5333	82.0667	Lake Rara	41
29238	KU	*S. raraensis*	Karnali	29.5333	82.0667	Lake Rara	6
27807	KU	*S. raraensis*	Karnali	29.5333	82.0667	Lake Rara	8

Table 7: Haplotypes derived from 1,140 bp of the cytochrome b gene sequenced across 115 tissue samples of *Schizothorax* from three major basins from Nepal in the Central Himalayas.

Haplotype	Species	Basin	Latitude	Longitude
[Hap_1	*S. progastus*	Koshi	27.34	87.19
[Hap_2	*S. progastus*	Koshi	27.34	87.19
[Hap_3	*S. progastus*	Koshi	27.34	87.19
[Hap_4	*S. progastus*	Koshi	27.34	87.19
[Hap_5	*S. progastus*	Gandaki	28.4	83.57
[Hap_6	*S. progastus*	Gandaki	28.4	83.57
[Hap_7	*S. progastus*	Gandaki	28.4	83.57
[Hap_8	*S. progastus*	Gandaki	28.4	83.57
[Hap_9	*S. progastus*	Gandaki	28.4	83.57
[Hap_10	*S. progastus*	Gandaki	28.4	83.57
[Hap_11	*S. progastus*	Gandaki	28.4	83.57
[Hap_12	*S. progastus*	Gandaki	28.4	83.57
[Hap_13	*S. progastus*	Gandaki	28.34	83.57
[Hap_14	*S. progastus*	Gandaki	28.34	83.57
[Hap_15	*S. progastus*	Gandaki	28.34	83.57
[Hap_16	*S. progastus*	Gandaki	28.34	83.57
[Hap_17	*S. richardsonii*	Gandaki	27.55	87.30
[Hap_18	*S. richardsonii*	Gandaki	29.52	81.78
[Hap_19	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_20	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_21	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_22	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_23	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_24	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_25	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_26	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_27	*S. richardsonii*	Gandaki	29.55	82.17
[Hap_28	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_29	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_30	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_31	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_32	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_33	*S. richardsonii*	Gandaki	28.4	83.57
[Hap_34	*S. richardsonii*	Gandaki	28.34	83.57
[Hap_35	*S. richardsonii*	Gandaki	28.68	83.61
[Hap_36	*S. richardsonii*	Gandaki	28.68	83.61
[Hap_37	*S. richardsonii*	Gandaki	28.68	83.61
[Hap_38	*S. richardsonii*	Gandaki	28.68	83.61

Table 7 (cont..)

Haplotype	Species	Basin	Latitude	Longitude
[Hap_39	*S. richardsonii*	Gandaki	28.34	83.57
[Hap_40	*S. richardsonii*	Koshi	27.82	85.57
[Hap_41	*S. richardsonii*	Koshi	27.82	85.57
[Hap_42	*S. richardsonii*	Karnali	29.55	82.17
[Hap_43	*S. richardsonii*	Karnali	29.55	82.17
[Hap_44	*S. richardsonii*	Karnali	29.55	82.17
[Hap_45	*S. richardsonii*	Karnali	29.52	81.78
[Hap_46	*S. richardsonii*	Karnali	29.52	81.78
[Hap_47	*S. richardsonii*	Karnali	29.52	81.78
[Hap_48	*S. richardsonii*	Karnali	29.52	81.93
[Hap_49	*S. richardsonii*	Karnali	29.55	82.17
[Hap_50	*S. richardsonii*	Karnali	29.55	82.17
[Hap_51	*S. richardsonii*	Karnali	29.55	82.17
[Hap_52	*S. richardsonii*	Koshi	29.55	82.17
[Hap_53	*S. richardsonii*	Koshi	27.55	87.30
[Hap_54	*S. richardsonii*	Koshi	27.55	87.30
[Hap_55	*S. richardsonii*	Koshi	27.55	87.30
[Hap_56	*S. richardsonii*	Koshi	27.55	87.30
[Hap_57	*S. richardsonii*	Koshi	27.55	87.30
[Hap_58	*S. richardsonii*	Koshi	27.55	87.30
[Hap_59	*S. richardsonii*	Koshi	27.55	87.30
[Hap_60	*S. richardsonii*	Koshi	27.55	87.30
[Hap_74	*S. nepalensis*	Karnali	29.53	82.08
[Hap_75	*S. nepalensis*	Karnali	29.53	82.08
[Hap_76	*S. nepalensis*	Karnali	29.53	82.08
[Hap_77	*S. nepalensis*	Karnali	29.53	82.08
[Hap_78	*S. nepalensis*	Karnali	29.53	82.08
[Hap_79	*S. raraensis*	Karnali	29.53	82.08
[Hap_80	*S. raraensis*	Karnali	29.53	82.08
[Hap_81	*S. raraensis*	Karnali	29.53	82.08
[Hap_82	*S. raraensis*	Karnali	29.53	82.08
[Hap_83	*S. raraensis*	Karnali	29.53	82.08
[Hap_84	*S. raraensis*	Karnali	29.53	82.08
[Hap_85	*S. raraensis*	Karnali	29.53	82.08
[Hap_86	*S. raraensis*	Karnali	29.53	82.08
[Hap_87	*S. raraensis*	Karnali	29.53	82.08
[Hap_88	*S. raraensis*	Karnali	29.53	82.08
[Hap_89	*S. raraensis*	Karnali	29.53	82.08

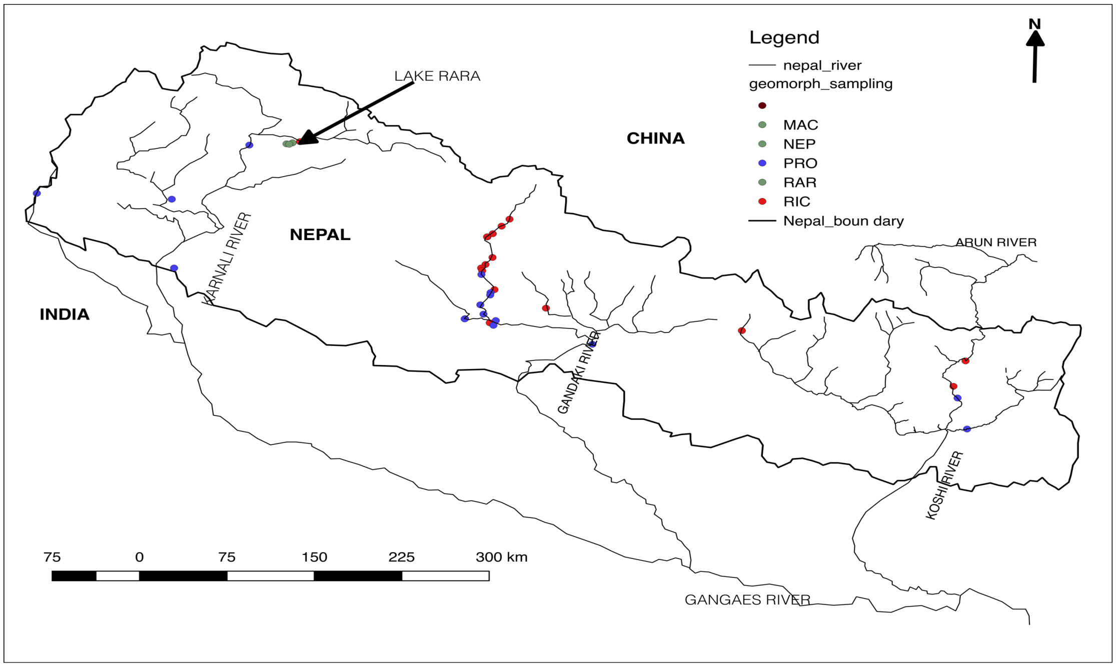

Figure 1: Sampling sites of *Schizothorax* specimens analyzed for morphological and genetic variation. Samples of two riverine species were collected in 1984-1986 from three river drainages (Karnali, Gandaki, and Koshi), with red indicating upstream sites for RIC= *S. richardsonii* and blue downstream sites for PRO= *S. progastus*. Three lacustrine species were collected in 1996 and are endemic to Lake Rara (green dot) in northwestern Nepal: MAC= *S. macrophthalmus,* NEP= *S. nepalensis,* RAR= *S. raraensis*). Sample sizes for each species are provided in Table 1.

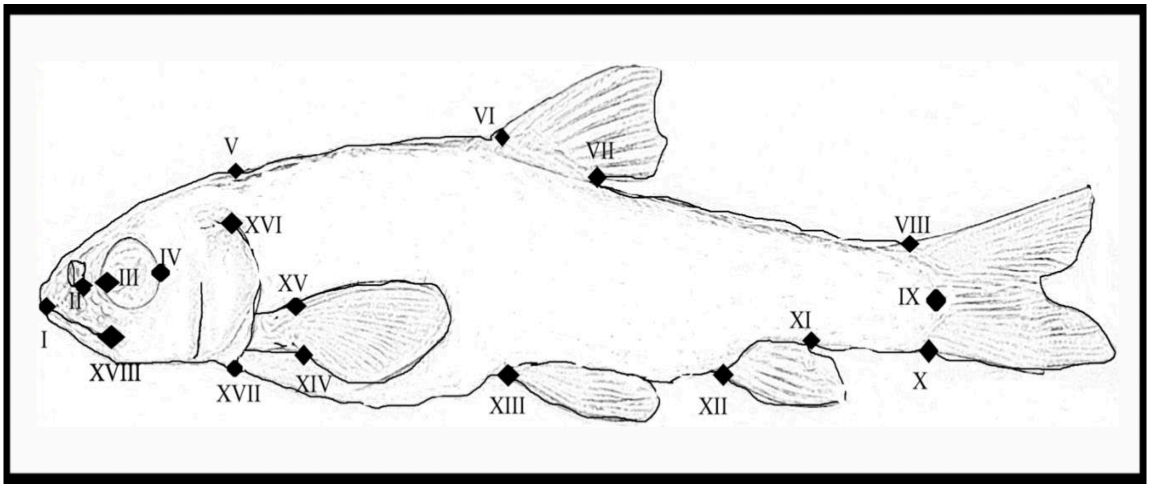

Figure 2: Eighteen anatomical landmarks (type I) selected to capture the head and body shape of the specimens. I: anterior mouth, II: nostril, III: orbital circle (anterior) point, IV: orbital circle posterior point, V: neuro-cranial boarder, VI: origin of dorsal fin, VII: insertion of dorsal fin, VIII: dorsal base of caudal fin, IX: urostyle, X: ventral base of caudal fin, XI: insertion of anal fin, XII: origin of anal fin, XIII: origin of pelvic fin, XIV: insertion of pectoral fin, XV: origin of pectoral fin, XVI: anterior end of operculum, XVII: posterior end of operculum, XVIII: left end of mouth opening.

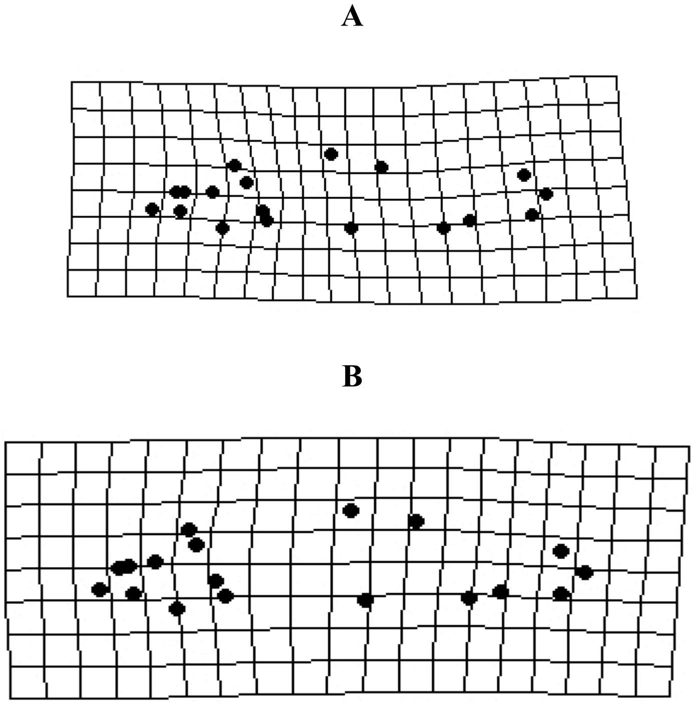

Figure 3: Deformation grid to demonstrate the shape change from reference form (average configuration) along the first (A, variance=0.40%) and second (B, variance=0.18%) principal components between *Schizothorax progastus* (Koshi Basin, riverine system) and *S. nepalensis* (Lake Rara, lacustrine system). The vertical line on very right end of the grid represents the average configuration.

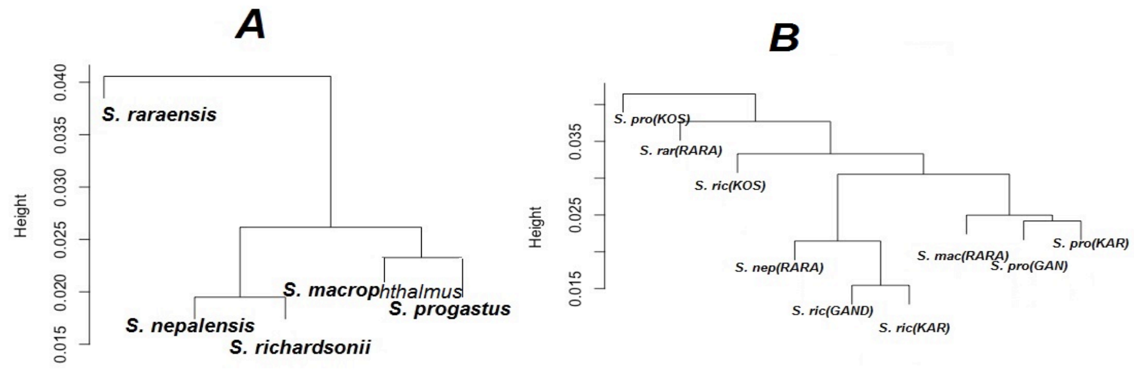

Figure 4: Dendrogram depicting relationships among *Schizothorax* morpo-types derived from a multivariate Euclidian distance matrix. Panel A: samples grouped by five morphological species, and Panel B: samples grouped by morphological species and split into basins. Acronyms for basins are: KOS= Koshi River, GAN= Gandaki River, KAR= Karnali River, RAR= Lake Rara (in Karnali River basin). Acronyms for species are: *S.pro*= S. *progastus*, *S. ric*= *S. richardsonii*, *S. mac*= S. macrophthalmus, *S. nep*= *S. nepalensis*, and *S. rar*= S. raraensis,

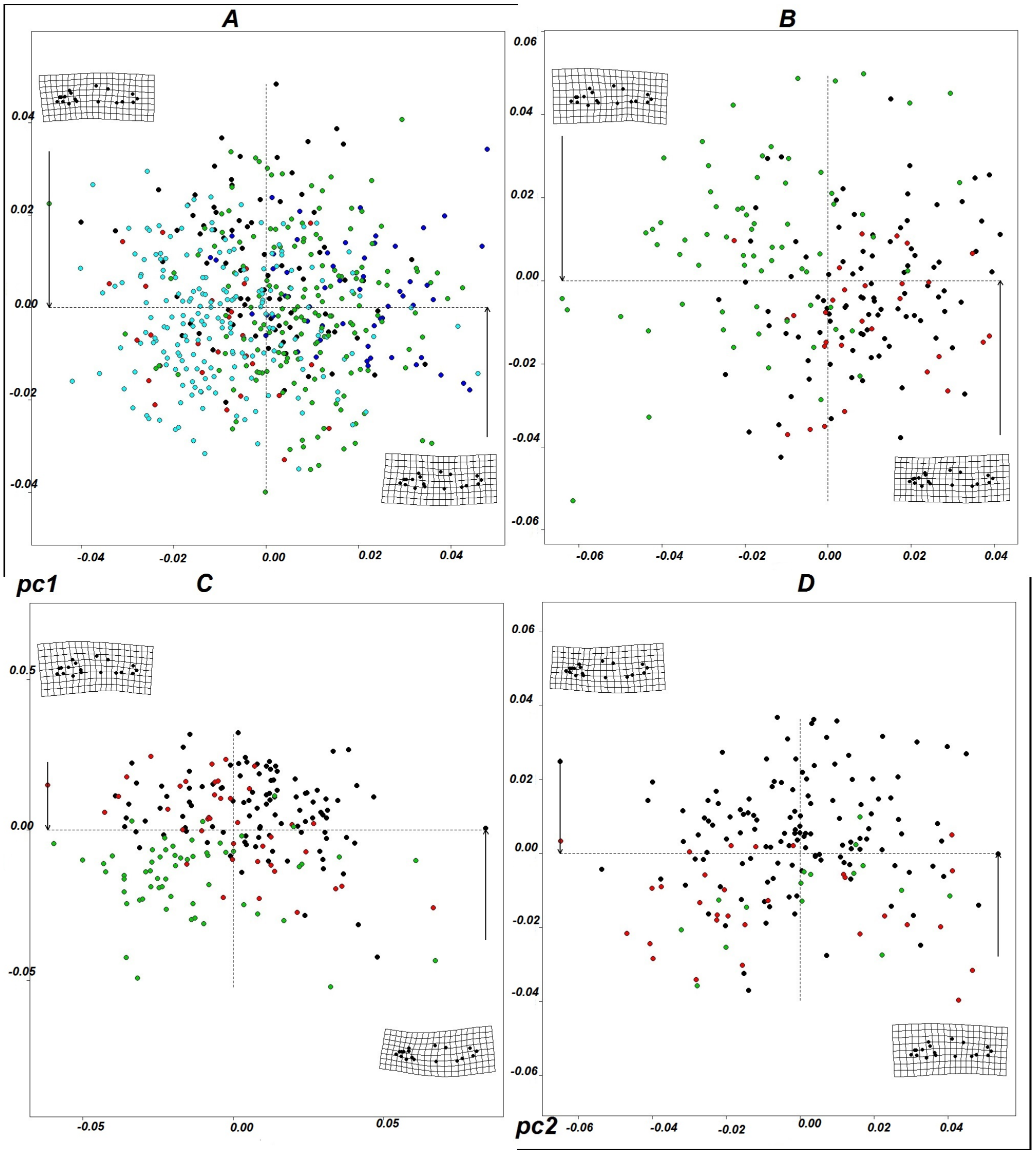

Figure 5: Shape variation along Principal Component 1 and 2 for five *Schizothorax* morpho-species sampled from three basins in Nepal. Panel A: all species (black = *S. macrophthalmus*, red=*S. nepalensis*, green= *S. raraensis*, blue= *S. progastus*, cyan= *S. richardsonii*), Panel B: Rara endemic fishes (black= *S. macrophthalmus*, red= *S. nepalensis*, green= *S. raraensis*), Panel C: *S. richardsonii* for all three basins (black= Gandaki, red=Karnali, green= Koshi), and Panel D: *S. progastus* (black= Gandaki, red= Karnali, green= Koshi) for all three basins. Deformation grid is presented on top left and bottom right corner of each figure.

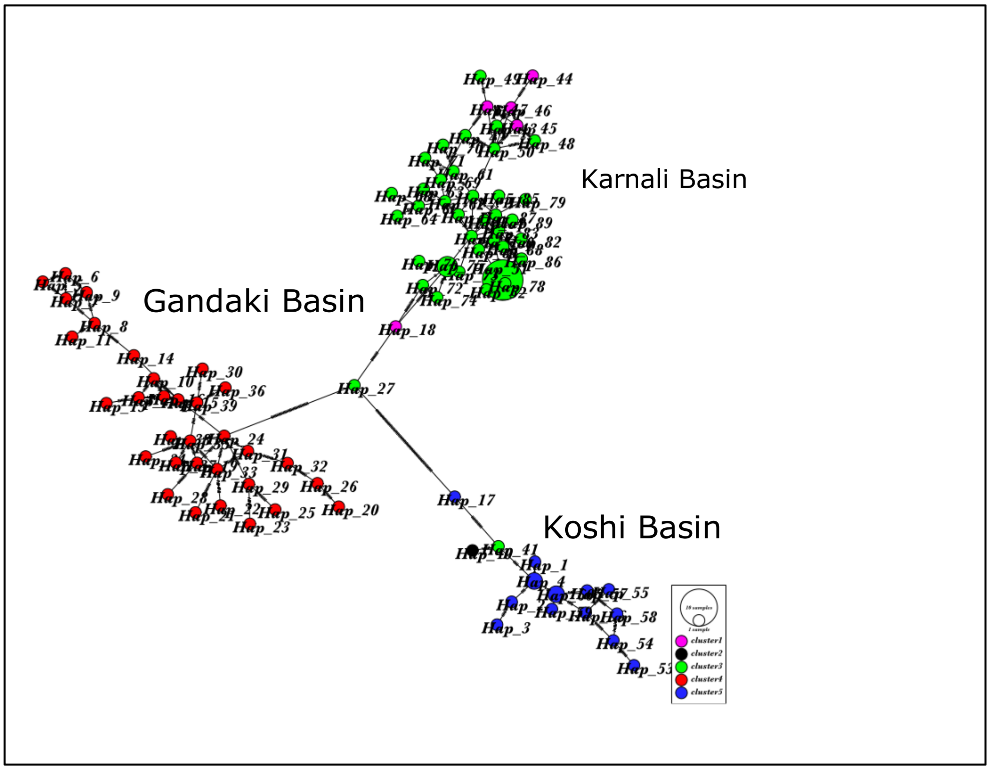

Figure 6: Haplotype network derived from 1,140 base pairs of the cytochrome b gene sequenced across 115 samples of *Schizothorax* collected from three basins in Nepal, Central Himalayas. Geographic location of basins is shown in Figure 1. Haplotypes are listed in Table 7. Colors represent unsupervised (K=5) clusters.

Chapter III

Can machine learning accurately delimit morphological species: Central Himalayan Snowtrout (*Schizothorax*) as a test case

Abstract

Machine learning algorithms (MLAs) have been implemented to parse biological diversity, primarily to identify distinct evolutionary lineages (i.e., higher-level taxa). However, the potentiality of MLAs to discriminate closely related species has not been fully explored yet. For this study, 565 museum specimens of five species (Snowtrout, genus *Schizothorax*) distributed in the Central Himalayas (Nepal) were obtained. Two species are common in major rivers in the region and the other three appear to be restricted to Lake Rara in Western Nepal. The shape variations across the species were captured using 18 anatomical landmarks. Effectiveness of four analytical approaches (principal component analysis with logistic regression, linear discriminant analysis, support vector machine, and random forest) were compared to discriminate five putative species based on the shape variation. Support vector machine and linear discriminant analysis performed well in discriminating the five species. Poor discrimination of one of the endemic species (*S. nepalensis*) when validated through Receiving Operating Characteristics (ROC) curve, but clustering with the two other species (*S. macrophthalmus* and *S. richardsonii*) is evidence of morphological intergradation, presumably stemming from hybridization between them. Analyses presented in this study provide insights into the morphological variation of the endemic species and their phenotypic affiliation with species in adjacent rivers. Furthermore, this study demonstrates the prospect of developing automatized endangered species recognition systems *for live or dead individuals being shipped* at important entry points of international trade.

Background

Machine learning has gained traction in a variety of computationally intensive applications, largely because of its ability to parse large datasets, but also its flexibility to apply accurate statistical models from which to develop predictions regarding those data. Its algorithms are trained to solve a given problem based on the knowledge derived from the data available (Baldi et al., 2001). Powerful, open source libraries now offer novel ways for researchers to interrogate, analyze, and process data, and therefore advance particular disciplines, be it molecular biology, physiology, electronic health records, or biomedicine in general (Baldi et al., 2001; Beam & Kohane, 2018).

Classification problems are commonly encountered in biological systems. A prime example is identification and classification of biodiversity (i.e., species) based on morphological or molecular data. Traditionally, species were distinguished by comparing morphological and anatomical traits, a method established in 1735 (Linneaus, 1735) and globally adopted and standardized (http://iczn.org/). The advent of low-cost molecular techniques triggered a bloom of the mitochondrial and nuclear genomic data (Springer et al., 2015) to characterize distinct evolutionary lineages. Today, biologists generally rely on molecular rather than morphological data to discriminate biodiversity and characterize evolutionary (phylogenetic) relationships amongst species, largely due to the size and extent of the genomic data, the spectrum of genetic markers with varying evolutionary rates, and applicability across various groups of organisms (Hillis, 1987). However, the plethora of information for phylogenetic studies comes from fossils (deep evolutionary history), or formalin preserved museum samples (historic record), both limited in the application of molecular techniques due to lack or degradation of suitable DNA to generate genetic data. As a compromise, many studies juxtapose molecular and morphological data to derive hypotheses for meaningful evolutionary relationships (Davis et al., 2016; Yang et al., 2016).

Reversely, trait based approaches such as geometry and morphometry are gaining popularity integrating functional as well as evolutionary signals (Adams, 2013a). Geometric morphometrics has enticed biologists because of its ability to convert the shape variation of an organism into a quantitative trait (Douglas et al., 2001; Adams et al., 2004). Freshwater fishes demonstrate higher levels of morphological disparity due fundamentally to differential habitat conditions, predation, reproduction as well as feeding behaviors (Langerhans et al., 2003). This variation is reflected in mouth shape, placement of eyes or fins, and can be captured using several techniques. The first approach taken by many authors is size related morphometry (Trapani, 2003). Measuring the standard length, total length, body depth, eyes diameter and estimating the ratios and applying statistics on these data give a comparable result for fish communities and their morphological variation. The serious disadvantage of this approach is reliability on the size that has been affected by individuals and their development stage (González-Castro & Ghasemzadeh, 2016). Remarkably, most of these measurements are taken in the longitudinal direction of the body representing the same feature in multiple measurements. In a different scenario, the object lacking homologous landmarks can be dealt with digitizing the outline and extracting variance with Fourier transformation (Caillon et al., 2018). However, the outline method is less applicable when there are adequate landmarks available to capture the shape variation. Considering the outline of the whole head and body shape could introduce more noise swamping small signals of shape variation particularly in closely related, and hence morphological similar species. Employing Cartesian coordinates of homologous anatomical landmarks (Adams et al., 2004) to capture the shape variation could be the best approach for studying species diverged very recently.

Study organism and system

Snowtrout (Cyprinidae, subfamily Schizothoracinae, genus *Schizothorax)* is a group of freshwater fish widely distributed in Central Asia, the Tibetan Plateau, as well as South and South East Asia (Terashima, 1984; Jayaram, 1999; Dimmick & Edds, 2002). Also known as Snow Minnow (Dimmick & Edds, 2002), Snowtrout are minnows (order Cypriniformes), a different taxonomic category than the fish commonly referred to as trout (order Salmoniformes). Both groups of fishes prefer cold water, and Snowtrout are found in upper drainages of Himalayan and Tibetan rivers that are fed by snow and glacier melts. Snowtrout are recorded from elevations as low as 100masl (meters above sea level) to as high as 5,000masl on the Tibetan Plateau (Edds, unpubl.).

The genus *Schizothorax* comprises a diverse array of more than 30 species (Sharma, 2008; Yang et al., 2012). Three species, *S. macrophthalmus, S. nepalensis, and S. raraensis* are restricted to a single lake (Rara Lake) in the Central Himalayas of northwestern Nepal, considered ‘endemic’ species, which reflects that they are found nowhere else on earth, and are considered as a species flock – indicating that they may have diverged from a single common ancestor that diversified into three distinct species through evolutionary adaptation (Tarashima, 1984; Dimmick & Edds, 2002). Two other species of Snowtrout, *S. richardsonii* and *S. progastus* are more widely distributed in streams and rivers in the Central Himalayas (Dimmick & Edds, 2002).

For assessment of biological diversity based on morphological traits, standardized measurements (i.e., morphometric) and counts (i.e., meristic) techniques are predominantly used. A variation, geometric morphometrics, quantifies morphological shape variation based on Cartesian coordinates derived from anatomical landmarks in 2-dimensional space using Procrustes transformation and has become very popular in ecological and evolutionary studies (Adams, 2014, Davis et al., 2016). In morphological classification (taxonomy), however, species are identified based on several dimensions, including qualitative characters such as color, texture, as well as quantitative data (e.g., shape, depth of the body). In reality, projecting the 3-dimensional shape of living organisms merely in 2-dimensional space and measuring the variation as described mathematically by removing the effect of the size is a daunting task; but there is no other option available for shape-based morphological studies. The subtle morphological variation of cryptic or recently diverged species is often impossible to detect through these traditional morphometric approaches. Statistical analysis of such data requires the estimation of a covariance matrix and matrix inversion often provoke matrix singularity and computationally zero eigenvalues (Ledoit & Wolf, 2004) that deter the use of larger feature space with low morphological variation.

In this study, I examined the potential use of machine learning algorithms (MLAs) to delimit five morphological species of Snowtrout based on shape variation. I examined and classified the species based on more traditional algorithms, such as Linear discriminant analysis (LDA), principal component analysis (PCA) with logistic regression (LG) as well as algorithms based on geometric optimization called support vector machine (SVM, Vapnik, 1999) and an ensemble approach referred to as random forest (RF). The main objective of this study was to evaluate the applicability of machine learning algorithms in species delimitation based on multivariate shape data. The study species are morphologically distinct, and their divergence is likely driven by ecological adaptations (Terashima, 1984), such as differences in feeding behavior and habitat preferences. I also examined the detectability of putative morphological species based on feature space derived only from head or body shape. Use of MLAs for species delimitation is a new approach; but it should be noted that fitting shape data to species defined by traditional morphological approaches is for interpretive purposes only, rather than a predictive tool.

Methods

Digital images of approximately 800 museum specimens representing the five *Schizothorax* species were captured. Individuals were sampled from Lake Rara, and three major rivers in Nepal (i.e. Gandaki, Koshi, and Karnali, Figure1) and are now vouchered as preserved specimens at the Oklahoma State University Museum of Vertebrate Zoology, Stillwater, USA, and the University of Kansas, Natural History Museum, Lawrence, USA.

Image capture was performed using a Canon (T3 Rebel) digital camera with a 300mm lens. A photographic box, designed for optimal light and exposure, was used to house the specimens as they were photographed (LED Pro Box Plus 1419 Photography Lighting Studio, MK Digital Direct, San Diego, CA 92154). Distorted specimens were not used in the study. Based on extensive literature review, I selected eighteen anatomical (type I) landmarks to capture the head and body shape variation of each specimen (Figure 2, Douglas et al., 2001; Clabaut et al., 2007; Dieleman et al., 2015).

To capture the head morphology, landmarks were selected that encapsulate snout elongation, head depth, and placement of the eyes and nostrils in the head (Figure 2). To describe body shape, landmarks representing origin and insertion of fins (dorsal, pelvic and pectoral) were designated, as were the upper and lower base of the caudal fin and the urostyle (bone structure forming the posterior-most end of the vertebra column). To extract Cartesian coordinates of each specimen after digitization a standard text file, tps (Rohlf, 2010) was generated using the program TPSUTIL (F.J. Rohlf, http://life.bio.sunysb.edu/ee/rohlf/software.html).

An R-based software package (GEOMORPH v.2.0, Adams et al., 2013b) was employed to read and superimpose the coordinate data using Generalized Procrustes Analysis. GPA transfers coordinates of all specimens to a common origin, scales them down into a unit-centroid size and rotates them optimally using a least square technique so that all the corresponding landmarks align as close as possible (Gower 1975; Rohlf & Slice 1990; Adams et al., 2004; Mitteroecker & Gunz 2009).

Potential biases due to sexual dimorphism and digitization error

In most fish species, gender is difficult to determine because they lack external traits that would distinguish males from females (i.e., sexually dimorphic characters). In this study, sex of specimens was not recorded at time of capture and could not be identified based on gonads or internal anatomy, because that would require a destructive process and destroyed the museum specimens. Instead, I classified the specimens as 'male' if breeding tubercles were present (knob-like structures appearing on the body and head during the breeding season); specimens lacking such tubercles were classified 'sex unknown.' Hotelling's T-test was employed to evaluate the significance of the means between these two classifications (i.e. certain 'male' *versus* 'sex unknown') to examine the potential effect of gender on shape.

To gauge the extent of digitization error, I randomly selected one specimen for each species and digitized its landmarks ten times (n=50). After obtaining Procrustes aligned data, I implemented a multivariate analysis of variance (MANOVA) in GEOMORPH to test the significance of the variance due to digitization error (within-group variance). Analyzing the variance of morphometric data is implemented through SS (sum of squares distance in morpho-space) rather than by estimating the covariance matrix due to potential complication posed by matrix singularity (Adams, 2014). The significance of square distances was evaluated through permutations with randomization of the matrix along rows (Goodall, 1991; Anderson, 2001).

Learning algorithms and classifications

For training and validation of the shape data, I used PYTHON 3.5 with following libraries: PANDA 0.15.2 (//pandas.sourceforge.nt); SCIPY 0.14.0 (Jones et al., 2001); NUMPY1.9.1 (Walt et al., 2011); and SCIKIT-LEARN 0.15.2 (Pedregosa et al., 2011; Buitinck et al., 2013). To minimize the effect of variable sampling size, the data were analyzed in two stages; one included all species, whereas another partitioned specimen into species endemic to Lake Rara *versus* species distributed in rivers. Additionally, I implemented stratified random sampling based on species before assigning 70% of specimens as the training and the remainder as the testing set. After splitting into the training and testing sets, features were standardized. I used multiple classifiers for analyses: principal component analysis (PCA; for visualization) with logistic regression (LR), linear discriminant analysis (LDA), support vector machine (SVM), and random forest (RF). The built-in pipeline function was implemented in SCIKIT-LEARN to automatize splitting and standardizing the data, defining the model parameters, and fitting the model for better control of the workflow.

Principal component analysis (PCA) is a popular unsupervised technique for machine learning. It is a dimension reduction technique that extracts the major components that maximize the variation in the reduced dimension of the feature space. Plotting the species on two top-loaded principal components helps to visualize the separability of species based on shape data. I also implemented the logistic regression (LR) model in the dataset and visualized the decision region of LR on the PCA plot.

The purpose of LDA is to determine the feature subspace in the training set that can be optimized as a class separator while preserving the within-class variation (Nasser, 2007). The class separators have different mean vectors (μ_1, μ_2) of the Generalized Procrustes Aligned (GPA) landmark coordinates, but the same covariance matrix ($\sum$) and Gaussian's distribution. The projection vectors (β_0, β_{px1}) were obtained by maximizing the between-class covariance while minimizing it within-class. Unlike the logistic regression, linear discriminant analysis places an observation (X^i) into one or another category (i.e. species y_1 or species y_2), but for a classification function ($L(X)$), small negative or positive values close to 0 indicate lower certainty that observation (X^i) falls into the category. An important advantage of LDA over SVM is that it does not require tuning parameters like cost function. I used SVD (Singular Value Decomposition) to solve the multiclass LDA to circumvent the situation of computationally zero eigenvalues in the low variance feature space. Any real matrix can be factorized into singular vectors and singular values using SVD even if eigen-decomposition is not possible.

Unlike discriminant analysis, SVM is less sensitive to the underlying distribution of the training dataset. Many machine learning algorithms, such as Artificial Neural Networks, update the weights during the training phase, and consequently, the predictability of the network for the small number of observation (n) could plummet. SVM overcomes this pitfall by estimating the decision boundaries from the training data set where the separating margins are maximized in the feature space (Nayak et al., 2015). SVM provides a good opportunity to get insight into the uncertainty zone between morpho-spaces of the species. The SVM classifier estimates an optimized separating hyperplane ($\beta_0 + \beta^T X = 0$, where β_0 is a constant and β^T is a projection vector), where a vector of the new observation (X) with a negative or positive value falls into one or the other category.

Support vector machine (SVM) is effective in high dimensional spaces. Since it uses a subset of the training points (support vectors) for building a decision function in the p-dimension space, it is computationally effective for a large number of features such as morphometric data. I implemented the grid search algorithm in SCIKIT-LEARN to search the best-cost function and kernel type for the classifier. SVM was implemented using Python library sklearn.svm (Pedregosa et al., 2011).

Random forest (RF), an ensemble approach of decision trees, averages multiple decision trees constructed through multiple iterations by either bootstrapping the samples with the replacement or randomly selecting a certain number of features. Based on majority rule consensus, an observation is assigned to a class. I implemented Random Forest with a Gini diversity function as a measure to evaluate the goodness of fit of dividing trees and estimated the tree 100 times. The RF was implemented with four parallel runs.

Model accuracy cross-validation

There is a risk of model overfitting when data comprise too many landmarks. A simple tool for diagnosing a potential over-fit is to compare its performance with regards to training *versus* test datasets (Raschka, 2015); if the model performs excellently with the training data, but poor with the test data, overfitting could be the cause. I assessed the performance of each classifier based on prediction accuracy in training as well as test sets. Here, the accuracy of each classifier is measured as the ratio of true positive classification and total observations in the data. To estimate the classifier's accuracy merely due to random chance, a random array was created with values 0 to 4 to assign species randomly for each row and implemented SVM and LDA algorithms.

The ROC (receiver operating characteristics) curve shows the relationship between true and false positive classification at various threshold settings. The classifiers' performance was also gauged using the ROC curve. The area under the curve (AUC) in the ROC plot, on the right-hand side from the dotted diagonal, represents the probability of classification based on random chances (area a=0.5). The performance of a classifier is gauged based on the area in between 0.5 and 1. A nearly perfect classifier has an AUC value closer to 1.

Results

Data generation, digitization error, and bias due to sexual dimorphism

A total of 565 specimens was used to extract the shape data (Table 1); distorted (bent or wrinkled), poorly preserved and juvenile specimens were excluded.

I did not find a significant difference (p=0.39, df=36) between the two mean vectors of specimens representing 'males' (i.e., with breeding tubercles) *versus* those of 'sex unknown' (i.e., without breeding tubercles). When estimating digitization error, the variance across species was significantly higher than within species (F=31.16, p=0.010), indicating that variation in selecting landmark points on digitized specimen did not majorly contribute to observed variance.

Training, prediction, and cross-validation

Before splitting the data into training and testing sets, I implemented a principal component analysis (PCA) to examine the possibility of dimension reduction. The first component explained 29.9% and the second 14.37% of the total variance, respectively. Thirty-two, out of 36 components had non-zero eigenvalues. After manually analyzing the loading matrix, I did not reduce the components of the data matrix simply because it did not make big difference in computational complexity and more importantly the components of a landmark co-ordinate with computationally zero eigenvalues had another axis that had the non-zero eigenvalue.

Plotting major principal components with a decision boundary derived from logistic regression allows visualizing the data within a space between two major components and reveals the single species most separable from all other species (Figure 3). Comparing all species (Figure 3A), the PCA showed separation of *S. raraensis, S. richardsonii,* and *S. progastus* from one another, while *S. macrophthalmus* and *S. nepalensis* were less well defined based on PC1 and PC2. More interestingly, two Lake Rara endemic species, *S. macrophthalmus* and *S. nepalensis* clustered together with the riverine *S. richardsonii*. When overall five species were divided into two classes using LR (Figure 3A) *S. raraensis* was defined by a blatant decision boundary from the rest of the species. Comparing only Lake Rara endemic species (Figure 3B), the major principal components again revealed separation of *S. raraensis* from the other two species, with *S. macropthalmus* and *S. nepalensis* broadly overlapping in the morpho-space defined by PC1 and PC2. Comparing the two riverine species (Figure 3C), *S. progastus* and *S. richardsonii* also showed good separation with a distinct decision boundary partitioning the morpho-space along PC1 and PC2.

Accuracy is a commonly used measure of classifier performance, yet it is sensitive to an imbalance among class sizes. To evaluate potential model overfit, I examined model performance based on test accuracy (Table 2: TeA). A 'null species model' that measures accuracy obtained by random chance was estimated TeA=22% for LDA (Model #1: T) and TeA=24% for SVM (Model #2). Both SVM and LDA performed well in separating all five species (Model # 3: SVM TeA=89%; Model #6 LDA,TeA=90%). Implementing a separate analysis including only the three Lake Rara endemic species revealed only a minor change in performance for both classifiers (Model #4 and #7). An analysis of the two riverine species showed machine learning classifiers performed very well in separating the two (Model #5: LDA TeA=97%; Model #8 TeA=94). Accuracy for the test (TeA) *versus* the training (TrA) data sets did not differ much (<10%) for any model and either classifier (Table 2).

In contrast, the Random Forest (RF) approach revealed a considerable gap between training and test accuracy for the overall model implemented (Model #14: TeA=86%, TrA=100%); this discrepancy could indicate model overfitting. However, RF performed well when the two riverine species were evaluated (Model#16: TrA=100%, TeA=96%). When reducing the feature space by only considering landmarks defining either head shape (k=12) or body shape (k=24) the LDA showed diminished accuracy for both training and testing data sets (head only Model #9: TeA=76%, TrA=79%; body only Model #10: TeA=77%, TrA=82%).

The receiver operating character (ROC) curve provides a robust estimate of model performance as well as separability of each class and the technique is insensitive to label imbalance (Figure 4). ROC curve to multi-class evaluation revealed very weak predictability of *S. nepalensis* (Figure 4: AUC, a<0.5) for both LDA and SVM.

Since *S. nepalensis* as a species underperformed in models as indicated by the ROC curves, I tested modified models (Table 2) by merging *S. nepalensis* with *S. macrophthalmus* (Model #11) or *S. richardsonii* (Model #12). Merging S. *nepalensis* with either species did improve the classification accuracy using SVM compared to the null-species model (Model #2) when considering all five species but was consistently better using the former (Model #11: TeA=90%) than the latter (Model #12: TeA=93%). Interestingly, species detectability was poor when all three species, *S. richardsonii, S. macrophthalmus* and *S. nepalensis* were merged (Model #13: TeA=89%).

Discussion

There are few studies on the application of machine learning for identification of fishes, birds, and amphibians (Trucco et al., 2000; Chen et al., 2005; Acevedo et al., 2009; Pittman et al., 2009; Joo et al., 2013). Most of these studies focus on either ecological informatics and modeling abundance of fishes in natural habitat (Pittman et al., 2011) or predicting the higher-level taxa based on digital images (Chen et al., 2005; Chen et al., 2008; Joo et al., 2013).

Recently, methods became available for building and testing algorithms for discriminating species. An R package VARSEDIG (Guisande et al., 2016) allows selecting features to discriminate biodiversity at broader categories (i.e., higher taxonomic level such as families). Nevertheless, the algorithm can only be implemented to distinguish two classes. The software package *IPez* (Guisande et al., 2010) implements a classification regression tree in multiple logistic regression to classify a specimen starting from broader taxonomic categories (e.g., an order) to more specific (e.g., species) incorporating an experts' knowledge. It has less discrimination ability for closely related species or morpho-types within a species.

In this study, I tested traditional classifiers, such as linear discriminant analysis (LDA) to more recently developed classifiers as support vector machine (SVM) or random forest (RF) to differentiate closely related species or morpho-types using subtle differences such as the placement of fins or orbital edge, traits that are not easily perceived, but appear to be consistent and likely biologically relevant. To the best of my knowledge, no study has to date examined the potential of machine learning algorithms for delimiting closely related species based on body shape.

When classifying organisms into ecologically or evolutionarily distinct groups dissociating the effect of sexual dimorphism (consistent differences between females and males) from variation among groups is an important first step in processing shape data. Depending on the organism, sexual dimorphism can be well pronounced in the variation of body shape (Vodounnou et al., 2017). In this study, I did not detect shape variation based on gender within the five species of *Schizothorax*. Furthermore, variance across groups was significantly greater than within groups, indicating that digitization error was too small to swamp the morphological variation across the study species. The distribution of variance in the principal components and their corresponding eigenvectors showed a balanced contribution of features (landmarks) to capture the shape across all species.

The linear discriminant analysis relies on the assumption that feature vectors for each class follow a multivariate normal distribution and covariance matrices of all classes are homogeneous. In contrast, support vector machines are free from such assumptions that rely on geometric optimization of the data points that are within or nearby the margin that separates the classes in the multidimensional hyperplane. In my dataset, I did not find a significant effect on accuracy between these two algorithms (Table 2).

A random forest is an ensemble approach of decision trees that average multiple trees to estimate the most robust tree that has a better generalization and less susceptibility to model overfitting. It averages the probabilistic prediction in each iteration instead of voting for each class. In my dataset, the prediction accuracy of RF was slightly lower than the SVM or LDA. This underperformance of RF could be due to the too small threshold set by the classifier itself for each iteration or, alternatively the sample size. High accuracy in the training set and lower accuracy in test set in my evaluations indicates model overfit (Table 2, Models #14-17).

The logistic regression analysis showed *S. raraensis* to be the most diverged species in morphological space with respect to the remaining four species (Figure 4A). This finding is corroborated by other morphological differences characterizing *S. raraensis*. Minnows such as Snowtrout do not have teeth along their jaws; instead they have pharyngeal 'teeth' on the fifth-gill arch (i.e., a tooth-bearing surface deep in their 'throat'), an outstanding adaptation to feeding that allows them to macerate larger food items before swallowing (Howes, 1991; Golubtsov et al., 2005). Most minnows have 1-3 rows of such pharyngeal teeth, but *S. raraensis* is one among few species bearing four rows (Golubtsov et al., 2005). Other morphological differences that make *S. raraensis* more distinguishable from the other four species is an advanced insertion of the pelvic fin, slightly posterior to the dorsal fin and halfway along the body length, as well as an elongated snout (Tarashima, 1985). Both the pharyngeal teeth and elongated snout are characteristics relevant for feeding, and divergence in the type of food exploited is an important driver of diversifying selection in fishes (i.e., character displacement).

The poor separability of *S. nepalensis* from other species by both SVM and LDA could reflect the origin and diversification of phenotypes among lake species (mechanisms of adaptive radiation). Genetically, *S. nepalensis* is distinct from the other species (Dimmick & Edds, 2002;

Tarashima 1985), indicating potential reproductive isolation (i.e., no interbreeding with the other species). In contrast, *S nepalensis* is morphologically more similar to *S. macrophthalmus* and *S. richardsonii.* This could indicate the three species have a direct ancestral relationship. In other words, a potential hypothesis for the adaptive radiation of *Schizothorax* in Lake Rara could be that *S. macrophthalmus* and *S. nepalensis* are derived from *S. richardsonii* once they became land-locked and isolated in Lake Rara. However, the underperformance of the models using a reduced number of features (landmarks for head shape only or body shape only) showed that variation in trophic morphology or body shape are weak and not adequate to differentiate the species.

Conclusion

Organisms can be perceived as separate species based on a variety of characteristics and humans have the visual ability to detect very subtle differences by eye. They also can judge differences in multiple morphometric and qualitative dimensions, such as color or texture. Other distinguishing characteristics can be geographical occurrence or ecological traits such as diet or reproductive habits. In this research, I only evaluated one component (body shape) by limiting the distinguishing characteristics to 36 features merely from anatomical landmark coordinates encapsulating only 2-dimensional variation in shape. I successfully trained the computer (i.e., machine learning) to recognize these morphological species based on subtle differences. The huge advantage of machine learning approaches over traditional comparative morphology is that the former can recognize and record the small signals of shape difference, for example, a small variation on the placement of eyes or fins. Combined, these signals can confound into significant class differences (separate areas in the multidimensional morpho-space) that help machine learning classifiers to distinguish one class *versus* another. In my data set, both support vector machine and linear discriminate analysis performed better as compared to the random forest approach. This illustrates that the two algorithms are useful to delineate species with subtle differences using generalized Procrustes aligned shape data.

While fish species flocks are among the best known and most intensely studied in African cichlids (Greenwood, 1974) and African barbs (Berrebi & Valiushok, 1998), it is rare among minnows and the Lake Rara *Shizothorax* could be one of them. Understanding the process of diversification (evolution) is essential for effective conservation and management of biodiversity, including this potential species flock of minnows in the Central Himalayas. This study also provides a template for designing a geometric morphometric workflow to evaluate subtle differences among similar species, a commonly encountered problem in conservation.

My findings generated in this study can now be applied to help guide conservation planners and policymakers to make informed decisions on how best to protect these unique fishes in the Central Himalayas. While this study extracted data from museum specimens, the findings open the prospect of developing an automatized intelligence system to recognize several endangered species and could be helpful to implement legal enforcement to protect them from regional, national and international trade.

Abbreviations

AUC: area under curve; LDA: linear discriminant analysis; GEOMORPH: geometry and morphometry; GPA: generalized Procrustes alignments; LG: logistic regression; MANOVA: multivariate analysis of variance; ML: machine learning; RF: random forest; ROC: receiving operating character; SVM: support vector machine; TrA: training accuracy; TeA: test accuracy

Acknowledgments

I express my thanks to curator Leo Smith and Collection Manager Andrew Bentley, Kansas Natural History Museum for their support to take images of Snowtrout collected from Nepal. My data also comes from Oklahoma State University and Dustin Lynch deserves special thanks to releasing the valuable museum specimens to the University of Arkansas for imaging. And last but not least, my thanks go to Dr. David Edds and his endless desire to study Himalayans fishes. These specimens were collected during his several expeditions in Nepal.

Funding

This research was made possible through generous endowments to the University of Arkansas: The Bruker Professorship in Life Sciences (MRD), and the 21st Century Chair in Global Change Biology (MED), as well as Graduate Assistantship (BR), all of which provided salaries and/or research funds for this work. Additional analytical resources were provided by the Arkansas Economic Development Commission (Arkansas Settlement Proceeds Act of 2000) and the Arkansas High Performance Computing Center (AHPCC). Samples were originally sampled by David Edds (Emporia State University) funded under Fulbright Scholarship, Explorer Club, National Science Foundation and National Geographic Society Committee for Research.

References

Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., and Aide, T. M. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4:206-214.

Adams, D. C. (2014). A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution*, 68:2675-2688.

Adams, D. C., and Otárola-Castillo, E. (2013b). GEOMORPH: An R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*. 4:393-399.

Adams, D. C., Rohlf, F. J., and Slice, D. E. (2004). Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology*, 71:5-16.

Adams, D. C., Rohlf, F. J., and Slice, D. A. (2013a). Field comes of age: Geometric morphometrics in the 21st century. *Hystrix*, 24:7-14.

Anderson, M. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Eco*logy, 26:32-46.

Baldi, P., Brunak, S., and Bach, F., (2001). *Bioinformatics: the machine learning approach*. MIT Press.

Beam, A. L., and Kohane, I. S. (2018). Big data and machine learning in health care. *Journal of American Medical Association*, 319:1317-1318.

Berrebi, P., and Valiushok, D. (1998). Genetic divergence among morphotypes of Lake Tana (Ethiopia) barbs.*Biological Journal of the Linnean Society*, 64:369-384.

Buitinck, L., Louppe G., Blondel, M. et al. (2013). API design for machine learning software: experiences from the scikit-learn project. *arXiv preprint arXiv*:1309.0238.

Caillon, F., Bonhomme, V., Möllmann, C,. and Frelat, R. (2018). A morphometric dive into fish diversity. *Ecosphere*, 9:p.e02220.

Chen, H., Henry, L. B., and Shuqing, H. (2008). Integrated feature selection and clustering for taxonomic problems within fish species complexes. *Journal of Multimedia,* 3.

Chen, Y., Bart, H. L., Huang S., and Chen, H. (2005). A computational framework for taxonomic research: diagnosing body shape within fish species complexes. *Fifth IEEE International Conference on Data Mining*, 27:4- .

Clabaut, C., Bunje, P. M., Walter, S., and Meyer, A. (2007). Geometric morphometric analyses provide evidence for the adaptive character of the Tanganyikan cichlid fish radiations. *Evolution*, 61:560-578.

Davis, M. A., Douglas, M. R., Collyer, M. L., and Douglas, M. E. (2016). Deconstructing a species-complex: geometric morphometric and molecular analyses define species in the Western Rattlesnake (*Crotalus viridis*). PloS One, 11:e0146166.

Dieleman, J., Bert, Van B., Claudia, M., Dorothy, W. N., Dominique, A., and Dirk V. (2015). Tracing functional adaptation in African cichlid fishes through morphometric analysis of fossil teeth: exploring the methods. *Hydrobiologia,* 755:73-88.

Dimmick, W. W., and Edds, D. R. (2002). Evolutionary genetics of the endemic schizorathicine (Cypriniformes: Cyprinidae) fishes of Lake Rara, Nepal. *Biochemical Systematics and Ecology*, 30:919-929.

Douglas, M. E, Douglas, M. R., Lynch J. M., and Douglas, M. M. (2001). Use of geometric morphometrics to differentiate Gila (Cyprinidae) within the upper Colorado River basin. *Copeia,* 2:389-400.

Golubtsov, A. S., Dzerjinskii, K. F., and Prokofiev, A. M. (2005). Four rows of pharyngeal teeth in an aberrant specimen of the small African barb *Barbus paludinosus* (Cyprinidae): novelty or atavistic alteration? *Journal of fish biology,* 67:286-291.

González-Castro, M., and Ghasemzadeh, J. (2016). Morphology and morphometry based taxonomy of Mugilidae. *Biology Ecology and Culture of Grey Mullet (Mugilidae).* CRC Press Taylor & Francis Group, Boca Raton, 1-21.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society.* Series B (Methodological):285-339.

Gower, J. C. (1975). Generalized Procrustes analysis. Psychometrika 40:33-51

Greenwood, P. H. (1974). Cichlid fishes of Lake Victorira, East Africa: the biology and evolution of species flock. *Bulletin of the British Museum of Natural History (Zoology),* Supplement, 6:1-134.

Guisande, C., Manjarrés-Hernández, A., Pelayo-Villamil, P., Granado-Lorencio, C., Riveiro, I., Acuña, A., Prieto-Piraquive, E., Janeiro, E., Matías, J.M., Patti, C., and IPez, P. A. (2010). An expert system for the taxonomic identification of fishes based on machine learning techniques. *Fisheries Research*, 102:240-247.

Guisande, C., Vari, R. P., Heine, J. Ü., Garcia-Rosello, E., Gonzalez-Dacosta, J., Perez-Schofield, B. G., Gonzalez-Vilas, L., and Pelayo-Villamil, P. (2016). VARSEDIG: an algorithm for morphometric characters selection and statistical validation in morphological taxonomy. *Zootaxa*, 4162:571-580.

Hillis, D. M. (1987). Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematics*, 18:23-42.

Howes, G. J. (1991). Systematics and biogeography: an overview. *In Cyprinid Fishes: Systematics, Biology and Exploitation (Winfield, I. J. & Nelson, J. S., eds)*, 1-33.

Jayaram, K. C. (1999). *The Freshwater Fishes of the Indian Region*. Narendra Publishing House, Delhi.

Jones, E., Oliphant, E., Peterson, P. *et al. (2001).* SciPy: Open Source Scientific Tools for Python**,** 2001-, http://www.scipy.org/.

Joo, D., Ye-seul, K., Jongwoo, S., Catarina, P., Jody, H., and Yong-Jin, W. (2013). Identification of cichlid fishes from Lake Malawi using computer vision. *PLoS One,* 8:e77686.

Langerhans, R. B., Layman, C. A., Langerhans, A. K., and Dewitt, T. J. (2003). Habitat-associated morphological divergence in two Neotropical fish species. *Biological Journal of the Linnean Society*, 80:689-698.

Ledoit, O., and Wolf, M. A. (2004). Well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88: 365-411.

Linneaus, C. (1735). *Systema naturæ, sive regna tria naturæ systematice proposita per classes, ordines, genera, & species*. *Lugduni Batavorum.*

Mitteroecker, P., and Gunz, P. (2009). Advances in geometric morphometrics. *Evolutionary Biology*, 36:235-247.

Mitteroecker, P., Gunz, P., Bernhard, M., Schaefer, K., and Bookstein, F. L. (2004). Comparison of cranial ontogenetic trajectories among great apes and humans. *Journal of Human Evolution,* 46:679-697.

Nasser, M. N. (2007). Pattern Recognition and Machine Learning. *Journal of Electronic Image*, 16:049901.doi.org/10.1117/1.2819119

Nayak, J., Naik, B. and Behera, H., (2015). A comprehensive survey on support vector machine in data mining tasks: applications and challenges. *International Journal of Database Theory and Application*, 8:169-186.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research,* 12:2825-30.

Pittman, S. J., Costa, B. M., and Battista, T. A. (2009). Using lidar bathymetry and boosted regression trees to predict the diversity and abundance of fish and corals. *Journal of Coastal Research,* 11:27-38.

Pittman, S. J., and Kerry, A. B. (2011). Multi-scale approach for predicting fish species distributions across coral reef seascapes. *PloS One*, 6:e20583.

Raschka, S. (2015). *Python machine learning*. Packt Publishing Ltd.

Rohlf, F. J., and Slice, D. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *System Biology*, 39:40-59.

Rohlf, F. J., (2010). tpsRelw, relative warps analysis, version 1.49. *Department of Ecology and Evolution, State University of New York at Stony Brook.*

Sharma, C. M. (2008). Freshwater fishes, fisheries, and habitat prospects of Nepal. *Aquatic Ecosystem Health & Management*, 11:289-297.

Springer, M. S., Anthony, V. S., Johanna, L. A. et al. (2015). Interordinal gene capture, the phylogenetic position of Steller's sea cow based on molecular and morphological data, and the macroevolutionary history of Sirenia. *Molecular Phylogenetics and Evolution,* 91:178-193.

Terashima, A. (1984). Three new species of the cyprinid genus *Schizothorax* from Lake Rara, northwestern Nepal. *Japanese Journal of Ichthyology*, 1984; 31:122-135.

Trapani, J. (2003). Geometric morphometric analysis of body-form variability in *Cichlasoma minckleyi*, the Cuatro Cienegas cichlid. *Environmental Biology of Fishes*, 68:357-369.

Trucco, E., Ouzounis, G., Odone, F., and Verri, A. (2000). Fish grading using machine vision and statistical learning. *Proceedings of Oceanology International*, 287-96.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10:988-999.

Vodounnou, D. V., Kpogue, D. S., Akpo, Y., Lagnika, M., and Fiogbe, E. D. (2017). Determination of sexual dimorphism of African snakehead (*Parachanna obscura*): morphometric and meristic parameters, weight-length relationship and condition factor. *International Journal of Biological and Chemical Sciences*, *11*:1742-1752.

Walt, S. V., Colbert, S. C., and Varoquaux, G. (2011). The NumPy Array: A structure for efficient numerical computation. *Computing in Science & Engineering*, 13:22-30. doi: 10.1109/MCSE.2011.37.

Yang, J., Yang, J. X., and Chen, X. Y. (2012). A re-examination of the molecular phylogeny and biogeography of the genus *Schizothorax* (Teleostei: Cyprinidae) through enhanced sampling with emphasis on the species in the Yunnan-Guizhou plateau, China. *Journal of Zoological Systematics and Evolutionary Research*, 50:184-191.

Yang, K., Wansheng, J., Xiaoyong, C., Wei, Z., and Junxing, Y. (2016). An integrative approach to reappraising species validity in *Pseudexostoma* (Teleostei: Sisoridae). *Zootaxa,* 4158:352-366.

Appendix

Table 1: Numbers of museum specimens analyzed for each of the five study species. Details on accession numbers and collection information provided in Table 3.

Species	**Number of specimens**
Schizothorax richardsonii	219
Schizothorax progastus	178
Schizothorax nepalensis	27
Schizothorax raraensis	59
Schizothorax macrophthalmus	82
Total	565

Table 2: Accuracy test of Machine Learning classifiers to delimit morpho-types of *Schizothorax* from three basins in Nepal. Listed are: Model description (pooling of samples), model number, number of features (x/y coordinates of landmarks), classifier, test, and training accuracy. Single star (*) represents test accuracy (TeA) < 90% and double stars (**) represent test accuracy (TeA) <80%. c indicates the cost function. Acronyms for morpho-species are: MAC= *S. macrophthalmus,* NEP= *S. nepalensis*, PRO= *S. progastus*, RAR= *S. raraensis*, and RIC= S. *richardsonii*. Samples sizes for each morpho-species are listed in Table 1.

Model description	**Model #**	**Features**	**Classifier**	**Accuracy (Test)**	**Accuracy (Training)**
Null species model	1	36	LDA	0.22	0.38
Null species model	2	36	SVM	0.24	0.39
All five species	3	36	SVM	0.89	0.97, c=0.2
Rara endemic species	4	36	SVM	0.92	0.97, c=0.2
RIC + PRO (riverine species)	5	36	SVM	0.94	0.98, c=0.2
All five species	6	36	LDA	0.90	0.92
Rara endemic species	7	36	LDA	0.88*	0.93
RIC + PRO (riverine species	8	36	LDA	0.97	0.98
All five species ***(Head part only)***	9	12	LDA	0.76**	0.79
All five species ***(Body part only)***	10	24	LDA	0.77**	0.82
RAR + ***merge*** (RIC+NEP) +MAC +PRO	11	36	SVM	0.90	1.00, c=0.2
RAR + ***merge*** *(*MAC+NEP)+RIC+PRO.	12	36	SVM	0.93	0.96
RAR + ***merge*** (MAC+NEP+RIC) +PRO	13	36	SVM	0.89*	0.93
All five species	14	36	RF	0.86*	1.00
RAR + ***merge*** *(*RIC+NEP*)+MAC+PRO.*	17	36	RF	0.87*	1.00

Table 3: List of voucher specimens used to capture images for geometric morphometric analysis. Accession number represents the catalog number of University of Kansas (KU) and Oklahoma State University (OSU). Listed are: Species= morpho-type as identified in the collection, N= number of samples per Accession.

Accession	**Museum**	**Species**	**Basin**	**Latitude**	**Longitude**	**River**	**N**
29036	KU	*S. progastus*	Gandaki	27.9355	83.4399	Kaligandaki	3
28998	KU	*S. progastus*	Gandaki	27.9200	83.6800	Kaligandaki	2
27815	KU	*S. progastus*	Gandaki	28.3403	83.5692	Kaligandaki	1
15608	OSU	*S. progastus*	Gandaki	27.7037	84.4248	Gandaki	6
15709	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	7
15722	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	5
29018	KU	*S. progastus*	Gandaki	28.1533	83.6367	Kaligandaki	6
29024	KU	*S. progastus*	Gandaki	28.0640	83.5607	Kaligandaki	2
29028	KU	*S. progastus*	Gandaki	27.9783	83.5850	Kaligandaki	11
15731	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	19
15892	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	10
16349	OSU	*S. progastus*	Gandaki	28.3403	83.5692	Kaligandaki	7
16755	OSU	*S. progastus*	Gandaki	28.1753	83.6396	Kaligandaki	15
16928	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	5
16948	OSU	*S. progastus*	Gandaki	28.1753	83.6396	Kaligandaki	5
16763	OSU	*S. progastus*	Gandaki	28.3403	83.5692	Kaligandaki	1
16733	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	2
17464	OSU	*S. progastus*	Gandaki	27.7037	84.4248	Gandaki	2
16693	OSU	*S. progastus*	Gandaki	27.9783	83.5850	Kaligandaki	1
16685	OSU	*S. progastus*	Gandaki	27.8768	83.6616	Kaligandaki	1

Table 3 (cont..)

Accession	**Museum**	**Species**	**Basin**	**Latitude**	**Longitude**	**River**	**N**
16733	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	4
17505	OSU	*S. progastus*	Gandaki	27.9372	83.4423	Kaligandaki	2
16758	OSU	*S. richardsonii*	Gandaki	28.2026	83.6699	Modi	1
29023	KU	*S. richardsonii*	Gandaki	28.064	83.5607	Kaligandaki	5
29039	KU	*S. richardsonii*	Gandaki	27.9355	83.4399	Kaligandaki	1
28989	KU	*S. richardsonii*	Gandaki	27.8996	83.6333	Kaligandaki	1
16758	OSU	*S. richardsonii*	Gandaki	28.2026	83.6699	Modi	1
16766	OSU	*S. richardsonii*	Gandaki	28.2026	83.6699	Modi	2
15710	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	3
15732	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	4
16758	OSU	*S. richardsonii*	Gandaki	28.2026	83.6699	Modi	1
15710	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	3
15732	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	4
15891	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	1
16929	OSU	*S. richardsonii*	Gandaki	27.9372	83.4423	Kaligandaki	3
29624	KU	*S. richardsonii*	Gandaki	28.0333	84.0667	setigandaki	1
29043	KU	*S. progastus*	Gandaki	28.3403	83.5692	Kaligandaki	3
27887	KU	*S. richardsonii*	Gandaki	28.4	83.5667	Kaligandaki	5
27886	KU	*S. richardsonii*	Gandaki	28.3403	83.5692	Kaligandaki	2
29042	KU	*S. richardsonii*	Gandaki	28.3403	83.5692	Kaligandaki	4

Table 3(cont…)

Accession	**Museum**	**Species**	**Basin**	**Latitude**	**Longitude**	**River**	**N**
16765	OSU	*S. richardsonii*	Gandaki	28.3403	83.5692	Kaligandaki	1
16774	OSU	*S. richardsonii*	Gandaki	28.4961	83.6545	Kaligandaki	3
16775	OSU	*S. richardsonii*	Gandaki	28.4961	83.6545	Kaligandaki	3
16348	OSU	*S. richardsonii*	Gandaki	28.3403	83.5692	Kaligandaki	2
16766	OSU	*S. richardsonii*	Gandaki	28.3403	83.5692	Kaligandaki	1
16348	OSU	*S. richardsonii*	Gandaki	28.3403	83.5692	Kaligandaki	2
16766	OSU	*S. richardsonii*	Gandaki	28.3403	83.5692	Kaligandaki	1
16769	OSU	*S. richardsonii*	Gandaki	28.3754	83.575	Kaligandaki	4
16351	OSU	*S. richardsonii*	Gandaki	28.4961	83.6545	Kaligandaki	1
16774	OSU	*S. richardsonii*	Gandaki	28.4961	83.6545	Kaligandaki	2
16775	OSU	*S. richardsonii*	Gandaki	28.4961	83.6545	Kaligandaki	2
29050	KU	*S. richardsonii*	Gandaki	28.684	83.6123	Kaligandaki	6
29044	KU	*S. richardsonii*	Gandaki	28.4318	83.6007	Kaligandaki	6
29050	KU	*S. richardsonii*	Gandaki	28.684	83.6123	Kaligandaki	6
15923	OSU	*S. richardsonii*	Gandaki	28.6841	83.6148	Kaligandaki	22
15924	OSU	*S. richardsonii*	Gandaki	28.6841	83.6148	Kaligandaki	8
15922	OSU	*S. richardsonii*	Gandaki	28.782	83.725	Kaligandaki	4
15921	OSU	*S. richardsonii*	Gandaki	28.8457	83.7859	Kaligandaki	1
16778	OSU	*S. richardsonii*	Gandaki	28.7129	83.6567	Kaligandaki	1
16353	OSU	*S. richardsonii*	Gandaki	28.6841	83.6148	Kaligandaki	2

Table 3(cont..)

Accession	**Museum**	**Species**	**Basin**	**Latitude**	**Longitude**	**River**	**N**
16779	OSU	*S. richardsonii*	Gandaki	28.6841	83.6148	Kaligandaki	3
16780	OSU	*S. richardsonii*	Gandaki	28.6841	83.6148	Kaligandaki	2
29614	KU	*S. progastus*	Karnali	29.028	81.1819	Lower Karnali	1
17109	OSU	*S. progastus*	Karnali	29.028	81.1819	Lower Karnali	7
28780	KU	*S. progastus*	Karnali	28.3999	81.1999	Lower Karnali	1
28666	KU	*S. progastus*	Karnali	29.0816	80.1417	Mahakali	4
17109	OSU	*S. progastus*	Karnali	29.028	81.1819	Karnali	7
27884	KU	*S. progastus*	Karnali	29.5217	81.78	Karnali	21
27810	KU	*S. richardsonii*	Karnali	29.5216	81.93	Nijar	10
27885	KU	*S. richardsonii*	Karnali	29.5533	82.1699	MuguKarnali	12
29560	KU	*S. richardsonii*	Karnali	29.6483	81.805	Lower Karnali	11
29595	KU	*S. progastus*	Koshi	29.028	81.1819	Lower Koshi	1
29527	KU	*S. progastus*	Koshi	27.2716	87.2082	Arun	2
29595	KU	*S. progastus*	Koshi	26.93	87.3133	Arun	1
29527	KU	*S. progastus*	Koshi	27.2716	87.2082	Arun	2
27808	KU	*S. progastus*	Koshi	27.2716	87.2082	Arun	9
29539	KU	*S. progastus*	Koshi	27.2133	87.2399	Arun	2
29521	KU	*S. richardsonii*	Koshi	27.3199	87.2082	Arun	2
29528	KU	*S. richardsonii*	Koshi	27.2716	87.2082	Arun	2
29233	KU	*S. richardsonii*	Koshi	27.5517	87.3016	Arun	11

Table 3(cont..)

Accession	**Museum**	**Species**	**Basin**	**Latitude**	**Longitude**	**River**	**N**
40564	KU	*S. richardsonii*	Koshi	27.8284	85.5769	Melamchi	38
27813	KU	*S. macrophthalmus*	Karnali	29.5333	82.0667	Lake Rara	23
27812	KU	*S. macrophthalmus*	Karnali	29.5333	82.0667	Lake Rara	24
29237	KU	*S. macrophthalmus*	Karnali	29.5333	82.0667	Lake Rara	35
27811	KU	*S. nepalensis*	Karnali	29.5333	82.0667	Lake Rara	18
29234	KU	*S. nepalensis*	Karnali	29.5333	82.0667	Lake Rara	9
27807	KU	*S. nepalensis*	Karnali	29.5333	82.0667	Lake Rara	12
27814	KU	*S. nepalensis*	Karnali	29.5333	82.0667	Lake Rara	41
29238	KU	*S. raraensis*	Karnali	29.5333	82.0667	Lake Rara	6
27807	KU	*S. raraensis*	Karnali	29.5333	82.0667	Lake Rara	8

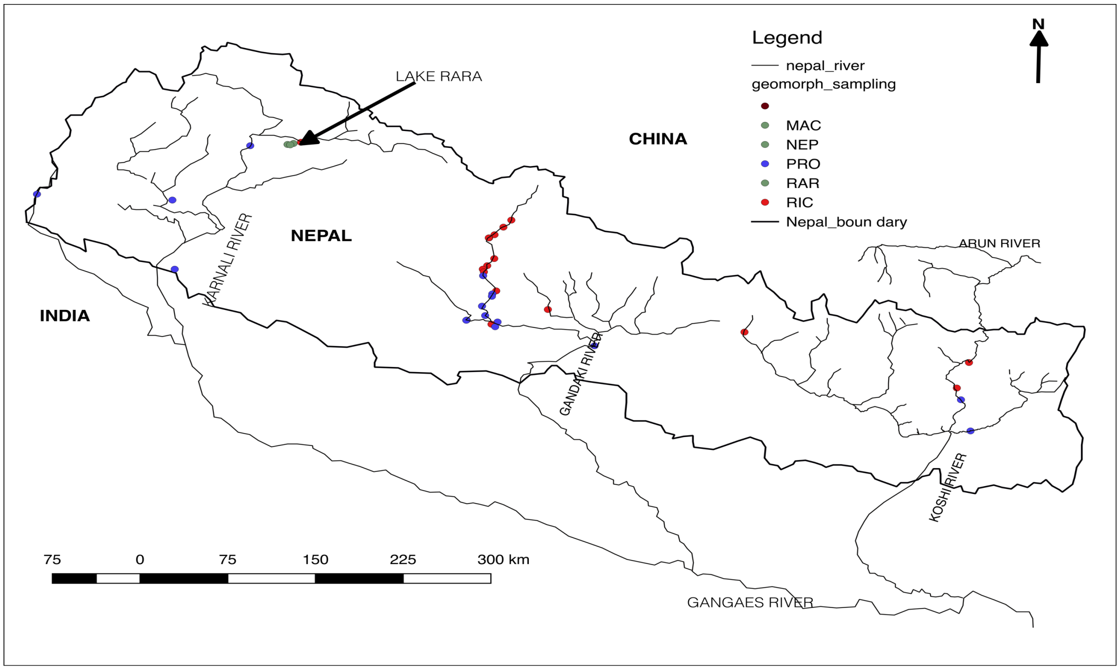

Figure 1: Collection sites of the 565 specimens representing five species of *Schizothorax* (Snowtrout) evaluated for shape variation using a geometric morphometric analysis. Red dots represent collection sites for *S. richardsonii* (RIC) and blue dots sites for *S. progastus* (PRO). *S. macrophthalmus* (MAC), *S. nepalensis* (NEP), and *S. raraensis* (RAR)are endemic to Lake Rara (green dot) in northwestern Nepal. Details on sample sizes are provided in Table 1, and details on collection sites in Table 3.

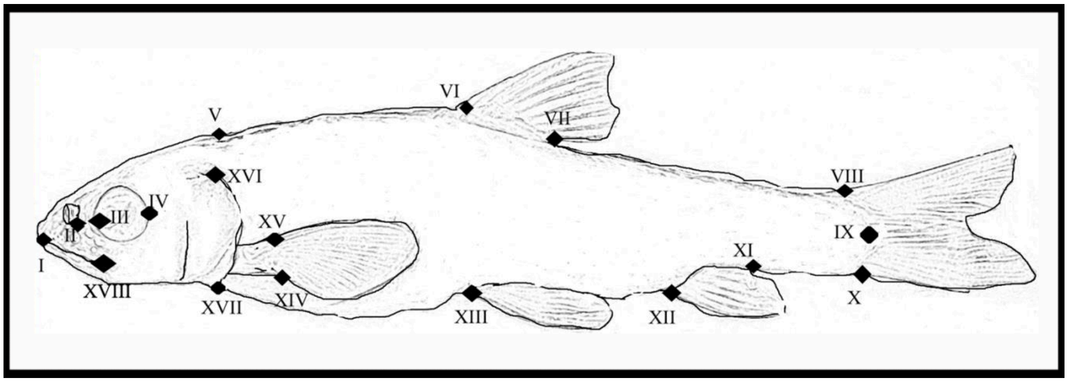

Figure 2: Configuration of 18 anatomical landmarks (type I) selected to capture head and body shape of museum specimens of five *Schizothorax* species collected from Nepal. Landmarks are I: anterior mouth, II: nostril, III: orbital circle anterior point, IV: orbital circle posterior point, V: neuro-cranial border, VI: origin of dorsal fin, VII: insertion of dorsal fin, VIII: dorsal base of caudal fin, IX: pygostyle, X: ventral base of caudal fin, XI: insertion of anal fin, XII: origin of anal fin, XIII: origin of pelvic fin, XIV: insertion of pectoral fin, XV: origin of pectoral fin, XVI: anterior end of operculum, XVII: posterior end of operculum, XVIII: left end of mouth opening.

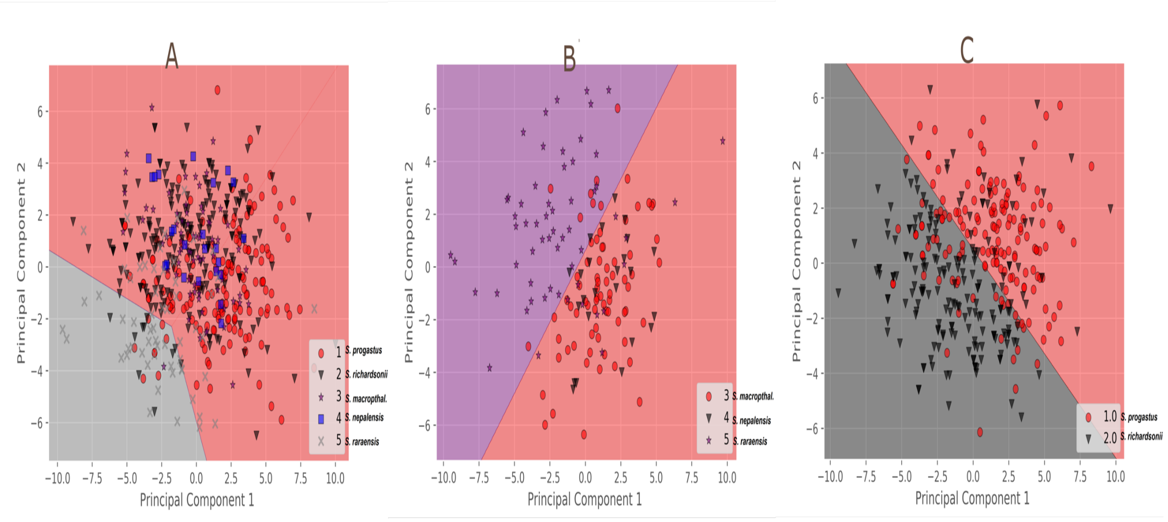

Figure 3: Principal component analysis of shape variation in five *Schizothorax* species collected in Nepal, Central Himalayas. Shape variation was derived from Cartesian coordinates of 18 anatomical landmarks (see Figure 2). The decision boundary reflects the most prominent demarcation among all species. Panel A: Comparison of all five species; the gray area represents the *S. raraensis* (x marks) with *S. progastus* (red dots) clustering on the right. The other three species, *S. richardsonii* (triangles), *S. macrophthalmus* (star) and *S. nepalensis* (blue squares) cluster together on the space between PC 1 and PC2. Panel B: Comparison of three species endemic to Lake Rara in northwestern Nepal, with *S. nepalensis* (triangles) and *S. macrophthalmus* (red dots) clustering together in the lower right (red area), with apparent separation of *S raraensis* (stars). Panel C: Comparison of the two riverine species more widely distributed in the Central Himalayas, with *S. progastus* (red dots) in the upper right, and *S. richardsonii* (triangles) in the lower left.

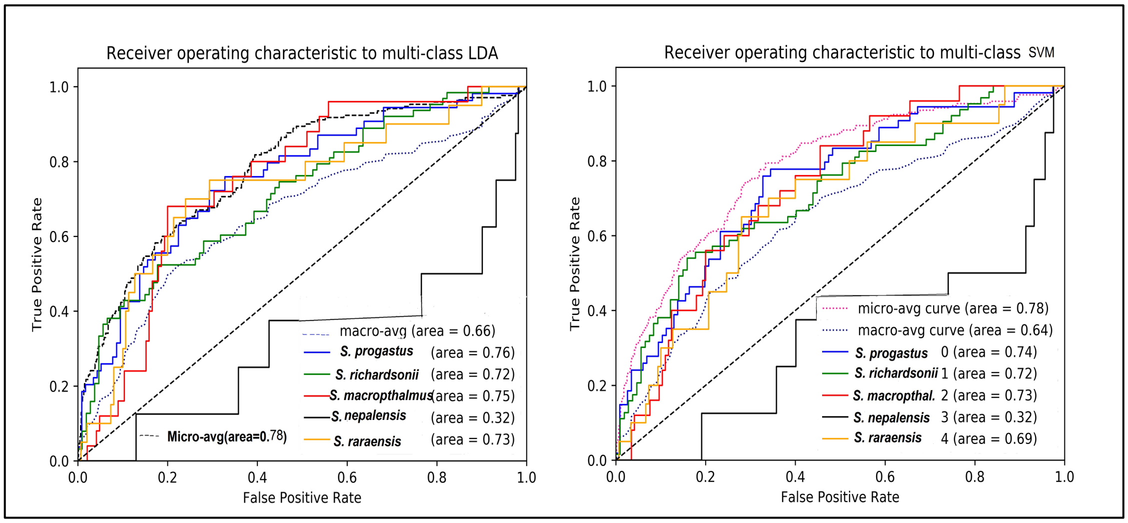

Figure 4: Receiver Operating Characteristic (ROC) curves to visualize model performance using two classifiers: LDA = Linear Discriminant Analysis, and SVM = Support Vector Machine. Specimens of five *Schizothorax species* (see Table 1) were classified based on shape. The ROC shows the relationship between true and false positive classifications at various threshold settings.

Chapter IV

Paleodrainages and the biogeography of Snowtrout (Cyprinidae: *Schizothorax*) in South and Central Asia

Abstract

Remote, tectonically active areas that have undergone substantial geological and climatic change are often diagnosed as global biodiversity hotspots. One such region, the Himalayan Mountains, was defined by the gradual collision of the Indian plate and the Eurasian landmass, with the uplift of the Tibetan Plateau to the north allowing drainages to subsequently downcut with fossil-poor aquatic taxa disseminating and diversifying. The tectonic evolution of these dendritic rivers provides a biogeographic template upon which phylogeographic relationships of endemic fishes can be juxtaposed, with divergence times providing a comparative framework to infer likely colonization routes and vicariant events.

To acquire such data, I sequenced 1,140 base pair of mitochondrial (mt)DNA cytochrome-b (cytb) for *Schizothorax* (Snowtrout) species endemic to the Central Himalayas and collapsed redundant haplotypes (Nepal: N=115 samples, 14 haplotypes; Bhutan: N=46 samples, 18 haplotypes). I placed these data in a larger geographic context by acquiring an additional 51 sequences from GenBank, representing 31 *Schizothorax* species across trans-Himalayan and Central Asian rivers, as well as sequences of five outgroup taxa. Data comprising a total of 88 haplotypes, with 83 representing *Schizothorax* and five outgroup taxa, were evaluated using maximum likelihood (ML) and Bayesian (BA) phylogenetic analyses, with divergence times estimated among clades.

My molecular analyses point to a monophyletic Central Himalayan clade that subsequently diverged following its dispersal south and west from the Tibetan Plateau. An east-

to-west pattern coupled with a close phylogenetic relationship between the Indus River and western tributaries of the Ganges River suggests the Central Himalayan clade diverged prior to the capture of the Jhelum River by the Indus. Two distinct, monophyletic clades were recovered from drainages in the Eastern Himalayas (Bhutan) and likely represent distinct dispersal events from Central Tibet and repeated colonization of tributaries in the Lower Brahmaputra.

My analyses further suggest that morphological species among basins are genetically more divergent than those distributed within basins. This, in turn, suggests the potential for undiagnosed diversity in Himalayan *Schizothorax* and, as such, requires additional data from which to diagnose potential taxonomic units. These results argue for more extensive morphological and molecular studies that can adequately diagnose these taxa as discrete conservation units.

Introduction

Climate and tectonism act synergistically to impact the hydrology of aquatic ecosystems (Whipple, 2009) and the distributions of resident freshwater fishes (Hopken et al., 2013). Drainages in western North America, for example, have been severely constrained by long-term, historic drought (Cook et al., 2016), such that genetic diversity of fishes has eroded over evolutionary time as populations repeatedly collapsed into refugia that subsequently served as a source for recolonization as more pluvial times returned (Douglas et al., 2003). Conversely, large-scale episodic flooding (Harden et al., 2010) has downcut antecedent streams, eroded headwaters, and over-spilt basins, thus promoting large-scale dispersals (Douglas et al., 2016). As a result, drainage reorganization (i.e., captures, diversions, beheadings; Bishop, 1995) not only extended the distributions of fishes into adjacent basins (Strange & Burr, 1997), but also confounded their taxonomies by facilitating hybridization (Bangs, 2016, 2018). Flows within basins were frequently abridged by longitudinal barriers that subsequently became vicariant structures for large-bodied species, yet served as filters for those smaller and more limited in their dispersal (Mims & Olden, 2013; Radinger et al., 2017). Consequently, fish distributions are often characterized as a series of biogeographic "islands" within and among basins, each reflecting local expansions, contractions, and eliminations (Burridge et al., 2008, Hugueny et al., 2010; Tedesco et al., 2012).

The above scenarios are particularly applicable when evaluating fishes within dentritic ecosystems subjected to similar climatic and tectonic histories. One such region is the Himalayan Mountains, a high-altitude global biodiversity hotspot at the southern border of the Tibetan Plateau (Myers et al., 2000). As with most mountainous regions, the sharp elevational clines provide ample opportunity for diversification (Du et al., 2017) and riverscape refugia (Li et al., 2016a). These, in turn, have contributed to allopatric speciation (Liu et al., 2016), thus enhancing its recognition as a hotspot for global biodiversity (CREP, 2017).

These benchmarks are also templates from which to gauge the potential response by fishes to rapid anthropogenic climate change. The Himalayas are extremely vulnerable in this regard, with a warming rate 2x greater than the global average (Shrestha et al., 2012; Xu & Grumbine, 2014). Impacts include increased precipitation (Wang et al., 2012), glacial retreat (Yao et al., 2012a), decline in snow cover (Yao et al., 2012b; Xu et al., 2017), expansion of glacial lakes (Sun et al., 2018), and a concomitant loss of permafrost (Ran et al., 2018). Yet, these impacts often elicit surprisingly divergent effects on biodiversity. For example, aquatic invertebrates seemingly lose species-diversity as elevation increases (Jacobsen et al., 2012; Li et al., 2016b), whereas cold-adapted terrestrial species may instead be buffered by microclimates at higher altitude (Scherrer & Körner, 2011). However, data for high elevation fishes remain elusive, particularly so for developing countries, a consideration that looms large with regard to potential oversights or misidentifications that often emerge from incomplete or non-systematic monitoring (Myers et al., 2000; Lovett et al., 2007).

Here I integrate numerous, disparate sources of information so as to determine the potential response by Himalayan freshwater biodiversity to rapid climate change. I first sequenced Snowtrout samples from three river basins in Nepal (Koshi, Gandaki, Karnali), as well as endemic species found in a restricted headwater (Lake Rara; Figure 1). I also sampled four basins in Bhutan (N=19 tributaries; Figure 2, Appendix 1). I obtained additional sequences from public databases (GenBank®) representing *Schizothorax* from the Indus and the Upper Gangetic River basins as well as major south-east draining Asian rivers, to include the Salween, Mekong, Yangtze, and the Upper Brahmaputra in Tibet (termed the Yarlung Tsangpo; Figure 3).

I examined phylogenetic relationships among *Schizothorax* species across this extended distributional range to juxtapose dispersal and speciation of the genus in the Central Himalayas against historic drainage patterns of the Tibetan Plateau. To examine geographic patterns within the phylogeny, I partitioned *Schizothorax* samples into six large geographic regions, with Southeast Asia including the drainage basins of Yangtze, Mekong, Salween and Irrawaddy rivers (Figure 3). The Eastern Himalayas included the drainage basins of the Lower Brahmaputra River (Bhutan) whereas the Central Himalayas encompassed three major tributaries of Ganges River in Nepal (Koshi, Gandaki, and Karnali rivers). The Western Himalayas included the Upper Ganges River Basin in northwestern India and the Indus River Basin in Pakistan. Central Tibet comprised the upper Brahmaputra River (Tsangpo), and Central Asia the Tarim River and its tributaries, an endorheic basin in the Central Asiatic Desert. In comparison to previous studies (He & Chen, 2006), my re-examination of phylogenetic relationships in *Schizothorax* across this large geographic region included more taxa from the Himalayan Region. Specifically, I examined how *Schizothorax* species distributed into the Himalayan Region, and other species endemic to Lake Rara are monophyletic.

Methods

Himalayan mountains

The Himalayas are the most extensive and recently evolved mountain system on Earth (length=2,400km; width=2,40km; elevation=75-8,800m). They are bordered by rivers on the west (Indus) and east (Brahmaputra), and their breadth encompasses most of northern India, (Kashmir to the west, Assam to the east), as well as Nepal, Bhutan, and parts of China, with the extensive Tibetan Plateau to the north and the alluvial plains of India and Bangladesh to the south (Figure 3). The Himalayas evolved over a 30my span, the result of a two-stage impact of the Indian craton with the Eurasian landmass. A preliminary 'soft' collision at 50mya preceded a much more substantial and more contiguous 'hard' collision at 25-20mya, as driven by the Indian lithosphere (van Hinsbergen et al., 2012). The leading edge of the Indian craton was subducted, with its upper surface essentially being ‘skimmed’ so as to form three sequential north-south Himalayan stratigraphic zones (i.e., Tibetan, Greater, and Lesser; Van Hinsbergen et al., 2012).

The hard collision not only established the Himalayas as a topographic entity, but also profoundly influenced the dendritic connectivity of its regional rivers (Brookfield, 1998; Bracciali et al., 2015), to include the evolution of their aquatic and semi-aquatic organisms. Crustal thickening and lift continued through the Pliocene-Pleistocene, with the Lesser Himalayas uplifting between 2.5-1.0mya (Sakai et al., 2006). These tectonic and geomorphic events not only drove global climate, but also transformed on a massive scale the cultural and biological endemism of the region (Li, et al., 2009; Xu et al., 2009; Xu & Grumbine, 2014).

Drainages of Nepal and Bhutan

Three major drainage basins are found east-to-west in Nepal (i.e., Koshi, Gandaki, and Karnali rivers; Figure 1), supporting eight major tributaries, several of which originate on the Tibetan Plateau. The Arun River to the east is the most extensive, cutting deeply through the Himalayas between Mt. Makalu (8,481m) and Mt. Kangchenjunga (8,586m), before merging into the Koshi River. The Eastern Himalayas are the recipients of a stronger monsoonal impact compared to western Himalayas (Kansakar et al., 2004), and mountain slopes are more eroded with more abrupt elevational gradients.

The Gandaki River drains Central Nepal, and the Kali Gandaki River is a vicariant barrier separating the flora of eastern and western Nepal (Jackson et al., 1994; Vetaas & Grytnes, 2002).

The upper catchment of the Kali Gandaki represents the deepest geomorphic gorge on the globe (Edds, 1989). The Karnali River demarcates Western Nepal, with a floristic composition more similar to that of Central Asia (Jackson et al., 1994). One study site, Lake Rara (with a surface area of 10.6km^2), is located within the Karnali River Basin at 2,990m elevation.

The study genus

Snowtrout (*Schizothorax*) is an ideal organism for deciphering the geologic processes that have controlled historic and contemporary fish distributions in the Himalayas. The genus contains a diverse array of species distributed across deserts, lakes, and rivers of the Tibetan Plateau, as well as the snow-fed tributaries of the Himalayas (Edds, 2007; Yang et al., 2012). In general, freshwater habitat along the elevational gradient of the Himalayas can be broadly divided into: High mountain, high-hill, low-hill, and lowland (Shrestha & Edds, 2012). The high mountain fish assemblage is dominated by *S. richardsonii*, with co-occurring genera being *Euchiloglanis* and *Noemaceilus* (Shrestha, 1999). At lower elevations (i.e., high-hill through lowlands), *S. richardsonii* is replaced by *S. progastus,* with associated species being *Tor tor*, *T. putitora*, *Barilius* and *Noemaceilus* species (Edds, 1989). Some 28 *Schizothorax* are recorded in the Himalayan, sub-Himalayan, and Tibetan regions (Sharma, 2008). Three species are endemic to Lake Rara in northwestern Nepal (Terashima, 1984; Dimmick & Edds, 2002), with two (i.e., *S. raraensis* and *S. nepalensis*) listed as critically endangered (IUCN, 2017; http://www.iucnredlist.org/details/168564/0/).

Several hypotheses have been offered to explain the origin and dispersal of *Schizothorax* across the Himalayas and Asia. Das & Subla (1963) suggested the subfamily *Schizothoraciane* initially dispersed into the Tibetan plateau from southeastern Asia, and subsequently became isolated by orogeny and climate change. Similarly, Cao et al., (1981) proposed that tectonism isolated a primitive barbin clade on the Tibetan Plateau that subsequent differentiated into *Schizothorax*. He & Chen (2006) invoked tectonism as the driver for phylogenetic relationships of *Schizothorax* in Southeast Asian rivers. The evidence supports strong monophyly within drainages and molecular data are demonstrably incongruent with current taxonomy.

Three morphologically differentiated and reproductively isolated *Schizothorax* (i.e., *S. macrophthalmus, S. nepalensis* and *S. raraensis*) are endemic to Lake Rara in the Karnali River Basin of northwestern Nepal (Terashima, 1984), and occupy separate trophic niches. Dimmick & Edds (2002) evaluated these taxa using mitochondrial gene sequence data, but could only differentiate *S. nepalensis*. Similarly, they were unable to discriminate two morphologically distinctive Snowtrout in the upper Karnali River (i.e., *S. richardsonii* and *S. progastus*). This prompted the speculation that *S. macrophthalmus* and *S. raraensis* shared mtDNA haplotypes as a result of hybridization. A similar situation was predicted for *Schizothorax* in the upper Karnali River.

Sampling, amplification, and sequencing of mtDNA

Tissue samples, representing five *Schizothorax* species (N=115), were obtained from the tissue collection at the University of Kansas Natural History Museum. Three species, *S. macrophthalmus, S. nepalensis,* and *S. raraensis,* represent a landlocked species-flock within Laka Rara, whereas *S. progaster* and *S. richardsonii* are riverine species. After removing redundant haplotypes, 14 sequences were used for phylogenetic analyses (Table 1).

I supplemented the samples from Nepal with *Schizothorax* (N=46) collected from four drainage basins in Bhutan (East to West: Mangde Chhu/Manas River, Punatsang Chhu/Sunkosh

River, Wang Chhu, and Amo Chhu); these collapsed into 18 non-redundant haplotypes (Table 1). In addition, from public databases (National Institute of Health Genetic Sequence Database, GenBank®: https://www.ncbi.nlm.nih.gov/genbank/) I downloaded 51 cytochrome b sequences of *Schizothorax*, representing 31 species, as well as sequences representing outgroup taxa to include: *Gymnocypris* (2 species, N=2); *Barbus* (three species, N=3) (Table 1). All sequences represented 1,140 bp without insertions or deletions.

Genomic DNA was extracted using the Qiagen DNeasy kit, per manufacturer's instructions. PCR-amplification of the cytochrome b gene was accomplished using published primers (Glu31, Unmack et al., 2009; Houston et al., 2012). Amplified products were enzymatically purified, sequenced using BigDye (ver.3.1) chemistry (Applied Biosystems Inc. [ABI], Forest City, CA), and analyzed on an ABI Prism 3700 Genetic Analyzer (W.M. Keck Center for Comparative and Functional Genomics, University of Illinois, Urbana/Champaign).

Sequence divergence and phylogenetic analyses

Sequences were edited using SEQUENCHER v.5.3 (Gene Codes Inc., Ann Arbor, MI), aligned using MUSCLE v.3.8 (Edgar, 2004), and nucleotide composition and uncorrected p-distances derived among groups using MEGA 6 (Tamura et al., 2013). I tested for nucleotide saturation at all positions using DAMBE5 (Xia, 2013), and employed PARTITIONFINDER 2 (Lanfear et al., 2016) to estimate the rate of sequence evolution. I used 56 models, each with four options: Model only, the model with invariant sites (I), gamma distributed rates across sites (G), or both invariant and gamma-distributed rates (I + G).

I derived a Bayesian phylogeny using a parallel version of MRBAYES v.3.2 (Ronquist et al., 2012), with each independent run on a separate node, and with heated and cold chains distributed among processors (Altekar et al., 2004). The Markov Chain Monte Carlo (MCMC) analysis was run for 100 million generations, with sampling every 1,000. Trees were summarized from two independent runs, with the best consensus tree retained. A parallel version of RAxML v.8.0 (Stamakakis, 2014) was also used to estimate a maximum likelihood (ML) tree, using a rapid bootstrap procedure.

An XML input file was prepared using BEAUTY and divergence times were estimated using BEAST v.1.7 (Drummond et al., 2012), using a relaxed molecular clock with uncorrelated lognormal distribution and a Yule tree prior. Calibration points consisted of *Barbus* species from Marocco and Spain that were separated 5mya by the opening of the Strait of Gibraltar (Zardoya & Doadrio, 1999). Likewise, two species of *Gymnocypris* were estimated to have diverged approximately 0.15mya in eastern Tibet (Li & Fang, 1998). Each analysis ran for 50 million generations and was sampled every 1,000. Effective sample size (ESS) was checked and MCMC chain convergence visually inspected using TRACER (Rambaut & Drummond, 2007). Trees from three independent runs were combined using LOGCOMBINER, and maximum clade credibility (MCC) derived with TREEANNOTATOR (Rambaut & Drummond, 2007) using 10% burn-in and >0.5 posterior probability. The resulting trees were visualized using FIGTREE (Rambaut, 2012).

Results

Sequence data

The phylogenetic analysis comprised 88 cytochrome b sequences including 34 *Schizothorax* species and 5 outgroup taxa (Table 1). Among all sequences, 436 sites were phylogenetically informative and 79 represented singletons. Based on the Akiaki Information Criteria (AICc), HKY+G+I was the best fitting model of sequence evolution.

Phylogeography of Schizothorax across the region

Clear geographic patterns were evident, with similar topographies depicted for ML (Figure 4) and BA trees (Figure 5). Four major geographic clades were apparent: Central Asia, Central Tibet, Southeast Asia, and Himalayas. The Central Asian clade was sister to all other clades, with the Tibetan clade sister to a clade comprising the Southeast Asian and Himalayan clades. An exception was the clustering of *S. argentatus* (a Central Asian species), placed as a separate clade in the ML tree, yet grouped with other Central Asian species in the BA tree. *Schizothorax* from Southeast Asian fell into two clades; in the ML tree they grouped together and were placed as sister to the Himalayan clades (Figure 4). In contrast, in the BA tree one Southeast Asian clade clustered as sister to the Central Tibetan clade, whereas the other grouped as sister to the Himalayan clades. Samples from the Eastern Himalayas (Lower Brahmaputra) also showed a unique pattern (described below).

Phylogeny of Himalayan Schizothorax

Himalayan species clustered as a monophyletic clade (Figures 4, 5), with the exception of some samples from the Lower Brahmaputra River Basin (Eastern Himalayas). These samples were all collected in Bhutan and formed two distinct, monophyletic clades; one (Bhutan-1) placed as sister to the Central Tibetan clade, whereas the other (Bhutan-2) clustered as sister to the Koshi River Basin clade (eastern-most drainage in the Central Himalayas). This pattern emerged in both the ML (Figure 4) and BA (Figure 5) trees.

The Himalayan clade comprised several sub-clades consistent with major basins and their clustering reflecting a distinct east-west trend. *Schizothorax* from the eastern-most basin from Nepal (i.e., Koshi River – Central Himalayas) formed a clade and clustered as sister to a clade representing the Lower Brahmaputra River Basin (Eastern Himalayas - Bhutan-2). Samples from the Gandaki and the Karnali rivers (Central Himalayas) each formed a monophyletic clade, as did samples from the Western Himalayas (Upper Ganges and Indus rivers). Interestingly, the latter formed a monophyletic and was placed as sister to western-most Central Himalayan clade (Karnali River). The Indus River Basin (Pakistan) drains to the Mediterranean Sea, whereas the Upper Ganges drains to the Bay of Bengal (India), with the Koshi, Gandaki, and Karnali rivers (Central Himalayas) draining into the Ganges once they reach the lowlands in India (Figure 3).

Divergence time

Seven major phylogenetic events were apparent in the time-calibrated phylogeny (Figure. 5). The average divergence time for the Central Asian clade was 8.5mya. The Tibetan clade diverged approximately 7mya, followed by one of the two Southeast Asian clades at 6.37mya. Several recent divergence events were recorded within the Himalayan clade, ranging from 5.31-0.5mya.

Within the Himalayas, the Koshi River clade (eastern-most drainage in the Central Himalayas) separated approximately 4.13mya, followed by the Gandaki River clade dated at 2.65mya, and the Karnali River clade separating from rest of the Himalayan clade at ~1.19mya. The Western Himalayan clade showed no clear separation between the Indus River and Upper Ganges clades (dated at about 0.5mya).

Discussion

Historical biogeography of the Himalayan Region is rooted at approximately 60mya when the Indian and Eurasian plates collided. Asian cichlids, arowwana and aplocheiloid fishes are Gondwanian relics now distributed in Asia (Azuma et al., 2008; Karanth, 2006). The biotic interchange between peninsular India and South East Asia preceded the hard hit of peninsular India with Eurasia (Gower et al., 2002; Renner, 2016). The Indian origin of some Southeast Asian freshwater fishes and gastropods (Grismer et al., 2016) corroborates the existence of a landbridge facilitating the interchange.

Current species richness, endemism and diversity of the Himalayan and Tibetan plateaus reflect the three-phase uplift of the region (He et al., 2001; Favre et al., 2015; Liang et al., 2017; Pellissier et al., 2017). More recent Pleistocene climate oscillations have also played a part (Zhao et al., 2009; Liu et al., 2015; Ma et al., 2015; Klaus at al., 2016; Meng et al., 2017). Most of the regional freshwater fish divergence occurred in late-Miocene when the Tibetan Plateau rose more than 3,000m in elevation (Zhou et al., 2016; Deng et al., 2019). This altered climatic conditions as well as drainage systems and accelerated the accumulation of high-altitude specialists that can subsist in harsh conditions (Schluter, 2016). Reconstructing ancestral states from the genealogical relationships among various freshwater fishes in the region suggested that plateau uplift had a major impact on phylogenetic events, with more recent glacial cycles being somewhat subsidiary (Liu et al., 2015; Ma et al., 2015; Klaus at al., 2016; Meng et al., 2017, Li et al., 2017).

There are limited studies on the phylogeography of freshwater organisms in the Himalayan Region. Much research on the evolutionary phylogeny of freshwater species in the region focused on the mountains of Tibet, Hengdung, and Qinghai (Liu et al., 2015; Ma et al., 2015; Klaus at al., 2016; Chi et al., 2017; Wanghe et al., 2017). Herein I focused on those species that potentially originated on the Tibetan Plateau and subsequently dispersed south into the Himalayas and Southeast Asia.

Phylogeny

My re-examination of the *Schizothorax* phylogeny employed publicly available data as well as my data generated for additional Himalayan taxa. My results underscored four major events that defined taxa in Central Asia, Tibet, Southeast Asia, and the Himalayas. The Central Asian species form a distinct clade sister to the rest of the Schizothoracine. Within these, a clade comprising species from Tibet and the Eastern Himalayas (Lower Brahmaputra River, Bhutan) groups as sister to remaining *Schiozothorax* species from Southeast Asia and the Himalayas (including Nepal). While the Tibetan/Eastern Himalayan (Bhutan) clade is monophyletic, species from the Tibet are paraphyletic with regards to a monophyletic Eastern Himalayan clade nested within. Species from Southeast Asia and the Himalayan regions each form a clade and are sisters to each other (He & Chen 2006; Yang et al., 2012; Figure 4).

The split of samples from the Lower Brahmaputra River into two distinct clades (Eastern Himalayas - Bhutan) is an important phylogenetic signature and represents an evolutionary transition between Tibetan Plateau species and those more recently diverged within the Himalayas and Southeast Asia. This may be stem from two separate dispersal events that extended from the Upper to the Lower Brahmaputra River Basin. Interestingly, there is no clear geographic segregation with regards to between the two clades representing Lower Brahmaputra River samples; both clades contain samples from each major drainage in Bhutan (Figure 2).

However, I detected a pattern of divergence in the Himalayan clade extending from east to west (Figures 4 and 5), with Koshi River Basin *Schizothorax* clustering as sister to the clade representing Lower Brahmaputra River (Eastern Himalayas - Bhutan) samples. Within the remaining Himalayan river drainages, the east-west trend continues, with *Schizothorax* from the Gandaki River forming the sister clade to samples collected further west in the Karnali River

Basin (Central Himalayas) the Indus and Upper Ganges (both Western Himalayas), with each of these basins forming a clade. The lacustrine species from Lake Rara cluster firmly within the Gandaki River clade.

The uplift of the Tibetan Plateau and the connections among paleorivers that resulted were crucial to speciation events that shaped the distribution of *Schizothorax* in Southeast Asian rivers (He & Chen, 2006). These patterns also established a strong pattern of monophyly. Few species are distributed across the numerous drainage systems in Southeast Asia, despite the fact that these river basins originated in a comparatively reduced Qinghai-Tibet Plateau, then separated as they flowed southward and drained into the sea. However, haplotype sharing occurs frequently among phenotypically diverse *Schizothora*x*,* such as *S. gongshanensis, S. lissolabiatus, and S. nukiangensis*, with retention of ancestral polymorphism a better explanation than introgressive hybridization (Chen et al., 2017). The population genetics of *S. o'conneri* from the Yarlung Tsangpo River (Central Tibet) revealed the role of Tibetan Plateau uplift and Quaternary climatic oscillations shaping genetic divergence (Guo et al., 2016).

In a previous study, *S. progastus* and *S. richardsonii* clustered according to drainage basin, regardless of the morpho-type (Dimmick & Edds, 2002). My results are consistent with this pattern. The Lake Rara endemics did not form a monophyletic clade, but clustered instead with individuals from the Karnali River that drains the lake. Dimmick & Edds (2002) suggested that haplotypes of *S. richardsonii* and *S. progastus* were not significantly different in the Karnali River Basin, but differed significantly when compared to conspecifics from the Koshi and Gandaki rivers; again, this was confirmed by my analyses.

There are very few taxonomic studies on fishes of the Central Himalayas and even fewer of them use molecular phylogenetics. My examination of the phylogenetic relationships of

Schizothorax species in the Himalayan region suggests that the current classification of species based on morphological traits does not reflect genetic diversity; it should be reassessed including molecular data.

Increasing divergence from east to west in the Himalayas

My analysis did not support the dispersal of *Schizothorax* species from the Himalayan region. The strong east-west pattern of divergence among clades could be associated with the evolution of the drainage systems in the Himalayan and South Asian region. Recent studies have suggested that current distribution of Snowtrout was shaped by uplift of the Tibetan Plateau in late Miocene that prompted vicariant events in headwaters of several East Asian rivers in south-eastern Tibet (He & Chen, 2006; Qi et al., 2015; Li et al., 2017). My estimate of approximately 6mya divergence between the Himalayan and Tibetan clades corresponds closely to that derived by He & Chen (2006), suggesting expansion from the Brahmaputra and Arun rivers westwards to the Ganges and Indus river basins. The Brahmaputra River originates on the Tibetan Plateau, flows eastwards then turns abruptly towards the south as it enters the Namche Barwa Syntaxis (Brookfield, 1998). Given this, the Brahmaputra River corridor may represent an important dispersal route for freshwater species of the Tibetan highlands.

Tectonic forces have driven orogenesis, with subsequent climate change, erosion, and glaciation having deeply influenced the evolution of the Himalayan drainage system. Based on isotope and seismic reflection data, Clift & Blusztajn (2005) reported that the Jhelum River capture by Indus River was very recent (<5mya); the Jhelum River was a tributary of the Ganges Basin before it drained into the Indus River in north-western Pakistan (Figure 3). The westernmost distribution of *Schizothorax* within the Upper Ganges Basin shows less divergence to the species of Indus Basin than the eastern tributaries of the Lower Ganges River. Similar low divergence patterns have been suggested between Indus *(Platanista gangetica minor)* and Gangetic *(P. gangetica gangetica)* dolphins (Braulik et. al., 2015). Based on these patterns, I argue that *Schizothorax* was already isolated in the Central Himalayan rivers (i.e., Koshi, Gandaki, and Karnali) even before the Jhelum River was captured by the Indus River.

The age of Lake Rara was estimated at approx. 60,000ya based on surface exposure dating of moraines deposited at the lake (Zakaria, 2018). Speciation in Karnali River Basin could be at an incipient stage with strong selection driving higher morphological diversification in the lacustrine environment, yet retention of ancestral alleles likely obscures a genetic signal.

Conclusion

Himalayan *Schizothorax* forms a monophyletic clade and phylogenetic relationships reflect an east-west divergence indicating their origin in Tibet and subsequent dispersal through rivers that originated in these highlands. These species are isolated in headwaters of Himalayan rivers and currently recognized morphological species appear to be substantially diverge from one basin to another. My findings suggest that detailed molecular studies of these species could help conservation planning at a regional scale but will require the inclusion of geographically isolated and endemic species across the Himalayan region.

Acknowledgments

Tissue samples for this study were obtained from the University of Kansas Natural History Museum. We thank Curator Leo Smith and Collection Manager Andrew Bentley for their prompt response and support for tissue loan. I would like to acknowledge the dedicated staff in the

National Research Centre for Riverine and Lake Fisheries, Bhutan who provided field assistance for sampling. Analyses were completed at the Arkansas High-Performance Computing Center (AHPCC) – University of Arkansas. This research was made possible through generous endowments to the University of Arkansas: The Bruker Professorship in Life Sciences (MRD), and the 21st Century Chair in Global Change Biology (MED), as well as Graduate Assistantship (BR), all of which provided salaries and/or research funds for this work.

References:

Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., and Ronquist, F. (2004). Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*, 20: 407-415. doi:10.1093/bioinformatics/btg427.

Azuma, Y., Kumazawa, Y., Miya, M., Mabuchi, K., and Nishida, M. (2008). Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC Evolutionary Biology*, 8, 215. doi:10.1186/1471-2148-8-215.

Bangs, M. R., Douglas, M. R., Thompson, P., and Douglas, M. E. (2016). Anthropogenic impacts facilitate native fish hybridization in the Bonneville Basin of Western North America. *Transactions of the American Fisheries Society*, 146:16-21. doi:10.1080/00028487.2016.1235611)

Bangs, M. R., Douglas, M. R., Mussmann, S. M., and Douglas, M. E. (2018). Unraveling historical introgression and resolving phylogenetic discord within *Catostomus* (Osteichthys: Catostomidae). *BMC Evolutionary Biology*, 18, 86. doi:10.1186/s12862-018-1197-y.

Bishop, P. (1995). Drainage rearrangement by river capture, beheading and diversion. *Progress in Physical Geography*, 19:449-473. doi:10.1177/030913339501900402.

Bracciali, L., Najman, Y., Parrish, R. R., Akhter, S. H., and Millar, I. (2015). The Brahmaputra tale of tectonics and erosion: Early Miocene river capture in the Eastern Himalaya. *Earth and Planetary Science Letters*, 415:25-37. doi:10.1016/j.epsl.2015.01.022.

Braulik, G. T., Barnett, R., Odon, V., Islas-Villanueva, V., Hoelzel, A. R., and Graves, J. A. (2015). One species or two? Vicariance, lineage divergence and low mtDNA diversity in geographically isolated populations of South Asian River dolphin. *Journal of Mammalian Evolution*, 22:111-120. doi:10.1007/s10914-014-9265-6

Brookfield, M. E. (1998). The evolution of the great river systems of southern Asia during the Cenozoic India-Asia collision: Rivers draining southwards. *Geomorphology*, 22:285-312. doi:10.1016/S0169-555X(97)00082-2.

Burridge, C. P., Craw, D., Jack, D. C., King, T. M., and Waters, J. M. (2008). Does fish ecology predict dispersal across a river drainage divide? *Evolution*, 62:1484-1499. doi:10.1111/j.1558-5646.2008.00377.x.

Cao, W. X., Chen, Y. Y., Wu, Y. F., and Zhu, S. Q. (1981). Origin and evolution of Schizothoracine fishes in relation to the upheaval of the Qinghai-Xizang Plateau. In Q. F. Li (Ed.), *Studies on the Period, Amplitude and Type of the Uplift of the Qinghai-Xizang Plateau* (118-129). Beijing, China: Science Press.

Chen, W., Yue, X., and He, S. (2017). Genetic differentiation of the *Schizothorax* species complex (Cyprinidae) in the Nujiang River (upper Salween). *Scientific Reports*, 7, 5944. doi:10.1038/s41598-017-06172-5

Chi, W., Ma, X., Niu, J., and Zou, M. (2017). Genome-wide identification of genes probably relevant to the adaptation of schizothoracins (Teleostei: Cypriniformes) to the uplift of the Qinghai-Tibet Plateau. *BMC Genomics*, 18, 310. doi:10.1186/s12864-017-3703-9

Clift, P. D., and Blusztajn, J. (2005). Reorganization of the western Himalayan river system after five million years. *Nature*, 438:1001-1003. doi:10.1038/nature04379.

Cook, B. I., Cook, E. R., Smerdon, J. E., Seager, R., Williams, A. P., Coats, S., Stahle, D. W., and Díaz J. V. (2016). North American megadroughts in the Common Era: Reconstructions and simulations. *WIREs Climate Change*, 7:411-432. doi:10.1002/wcc.394.

CREP-Critical Ecosystem Partnership Fund (2017). The biodiversity hotspots. (http://www.cepf.net/resources/hotspots/Pages/default.aspx).

Das, S. M., and Subla, B. A. (1964). The ichthyofauna of Kasmir: Part II The speciation of Kasmir fishes with two new records of species. *Ichthyologica*, 3:57-62.

Deng, T., Wang, X., Wu, F., Wang, Y., Li, Q., Wang, S., and Hou, S. (2019). Review: Implications of vertebrate fossils for paleo-elevations of the Tibetan Plateau. *Global and Planetary Change,* 174:58-69. doi:10.1016/j.gloplacha.2019.01.005

Dimmick, W. W., and Edds, D. R. (2002). Evolutionary genetics of the endemic schizorathicine (Cypriniformes: Cyprinidae) fishes of Lake Rara, Nepal. *Biochemical Systematics and Ecology*, 30:919-929. doi:10.1016/S0305-1978(02)00030-3.

Douglas, M. R., Brunner, P. C., and Douglas, M. E. (2003). Drought in an evolutionary context: Molecular variability in Flannelmouth Sucker (*Catostomus latipinnis*) from the Colorado River Basin of western North America. *Freshwater Biology*, 48:1254-1273. doi:10.1046/j.1365-2427.2003.01088.x.

Douglas, M. R., Davis, M. A., Amarello, M., Smith, J. J., Schuett, G. W., Herrmann, H-W., Holycross, A. T., and Douglas, M. E. (2016). Anthropogenic impacts drive niche and conservation metrics of a cryptic rattlesnake on the Colorado Plateau of western North America. *Royal Society Open Science*, 3, 160047. doi:10.1098/rsos.160047.

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUTI and the BEAST 1.7. *Molecular Biology and Evolution*, 29:1969-1973. doi:10.1093/molbev/mss075.

Du, F., K., Hou, M., Wang, W., Mao, K., and Hampe, A. (2017). Phylogeography of *Quercus aquifolioides* provides novel insights into the Neogene history of a major global hotspot of plant diversity in south-west China. *Journal of Biogeography*, 44:294-307. doi:10.1111/jbi.12836.

Edds, D. R. (1989). *Multivariate analysis of fish assemblage composition and environmental correlates in a Himalayan River-Nepal's Kali Gandaki/Narayani* (Doctoral dissertation, Oklahoma State University).

Edds, D. R. (2007). Fishes in Nepal: ichthyofaunal surveys in seven nature reserves. *Ichthyological Exploration of Freshwaters*, 18:277-287.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32:1792-1797. doi:10.1093/nar/gkh340

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

Gower, D. J., Kupfer, A., Oommen, O. V., Himstedt, W., Nussbaum, R. A., Loader, S. P., Presswell, B., Müller, H., Krishna, S.B., Boistel, R., and Wilkinson, M. (2002). A molecular phylogeny of ichthyophiid caecilians (Amphibia: Gymnophiona: Ichthyophiidae): out of India or out of South East Asia? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269:1563-1569. doi:10.1098/rspb.2002.2050

Grismer, J. L., Schulte, J. A., Alexander, A., Wagner, P., Travers, S. L., Buehler, M. D., Welton, L. J., and Brown, R. M. (2016). The Eurasian invasion: Phylogenomic data reveal multiple Southeast Asian origins for Indian Dragon Lizards. *BMC Evolutionary Biology*, 16, 43. doi:10.1186/s12862-016-0611-6

Guo, X. Z., Zhang, G. R., Wei, K. J., Yan, R. J., Ji, W., Yang, R. B., Wei, Q. W., and Gardner, J. P. (2016). Phylogeography and population genetics of *Schizothorax o'connori*: Strong subdivision in the Yarlung Tsangpo River inferred from mtDNA and microsatellite markers. *Scientific Reports*, 6, 29821. doi:10.1038/srep29821

Harden, T., Macklin, M. G., and Baker, V. R. (2010). Holocene flood histories in the southwestern USA. *Earth Surface Processes and Landforms*, 35:707-716. doi:10.1002/esp.1983.

He, D., and Chen, Y. (2006). Biogeography and molecular phylogeny of the genus *Schizothorax* (Teleostei: Cyprinidae) in China inferred from cytochrome b sequences. *Journal of Biogeography*, 33:1448-1460. doi:10.1111/j.1365-2699.2006.01510.x

He, S., Cao, W., and Chen, Y. (2001). The uplift of Qinghai-Xizang (Tibet) Plateau and the vicariance speciation of glyptosternoid fishes (Siluriformes: Sisoridae). *Science in China Series C: Life Sciences*, 44:644-651. doi:10.1007/BF02879359.

Hopken, M. W., Douglas, M. R., and Douglas, M. E. (2013). Stream hierarchy defines riverscape genetics of a North American desert fish. *Molecular Ecology*, 22:956-971. doi:10.1111/mec.12156.

Houston, D. D., Evans, R. P., and Shiozawa, D. K. (2012). Evaluating the genetic status of a Great Basin endemic minnow: The relict dace (*Relictus solitarius*). *Conservation Genetics,* 13:727-742. doi:10.1007/s10592-012-0321-6

Hugueny, B., Oberdorff, T., and Tedescco, P. A. (2010). Community ecology of fishes: A large scale perspective. *American Fisheries Society Symposium*, 73:29-62.

IUCN. (2017). The IUCN red list of threatened species. Version 2017.2, Gland, Switzerland.

Jackson, J. K., Stapleton, C. M., and Jeanrenaud, J. P. (1994). *Manual of afforestation in Nepal*.

Jacobsen, D., Milner, A. M., Brown, L. E., and Dangles, O. (2012). Biodiversity under threat in glacier-fed river systems. *Nature Climate Change*, 2:361-364. doi:10.1038/NCLIMATE1435.

Kansakar, S. R., Hannah, D. M., Gerrard, J., and Rees, G. (2004). Spatial pattern in the precipitation regime of Nepal. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *24*:1645-1659.

Karanth, P. K. (2006). Out-of-India Gondwanan origin of some tropical Asian biota. *Current Science*, 90:789-792

Klaus, S., Morley, R. J., Plath, M., Zhang, Y. P., and Li, J. T. (2016). Biotic interchange between the Indian subcontinent and mainland Asia through time. *Nature Communications*, *7,* 12132. doi:10.1038/ncomms12132

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2016). PARTITIONFINDER 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34:772-773. doi:10.1093/molbev/msw260.

Li, J. J., Fang, X. M. (1998). Uplift of the Tibetan Plateau and environmental changes. *Chinese Science Bulletin*, 44:2117-2125

Li, F., Shah, D. N., Pauls, S. U., Qu, X., Cai, Q., and Shah, R. T. (2016a). Elevational shifts of freshwater communities cannot catch climate warming in the Himalaya. *Water*, 8, 327. doi:10.3390/w8080327.

Li, J., He, Q., Hua, X., Zhou, J., Xu, H., Chen, J., and Fu, C. (2009). Climate and history explain the species richness peak at mid-elevation for *Schizothorax* fishes (Cypriniformes: Cyprinidae) distributed in the Tibetan Plateau and its adjacent regions. *Global Ecology and Biogeography*, 18:264-272. doi:10.1111/j.1466-8238.2008.00430.x

Li, J., McCarthy, T. M., Wang, H., Weckworth, B. V., Schaller, G. B., Mishra, C., Lua, Z., and Beissinger, S. R. (2016b). Climate refugia of snow leopards in High Asia. *Biological Conservation*, 203:188-196. doi:10.1016/j.biocon.2016.09.026.

Li, Y., Ludwig, A., and Peng, Z. (2017). Geographical differentiation of the *Euchiloglanis* fish complex (Teleostei: Siluriformes) in the Hengduan Mountain Region, China: Phylogeographic evidence of altered drainage patterns. *Ecology and Evolution, 7*:928-940. doi:10.1002/ece3.2715

Liang, Y., He, D., Jia, Y., Sun, H., and Chen, Y. (2017). Phylogeographic studies of schizothoracine fishes on the central Qinghai-Tibet Plateau reveal the highest known glacial microrefugia. *Scientific Reports,* 7, 10983. doi:10.1038/s41598-017-11198-w

Liu, J., Wang, C., Fu, D., Hu, X., Xie, X., Liu, P., Zhang, Q., and Li, M. H. (2015). Phylogeography of *Nanorana parkeri* (Anura: Ranidae) and multiple refugia on the Tibetan Plateau revealed by mitochondrial and nuclear DNA. *Scientific Reports,* 5, 9857. doi:10.1038/srep09857

Liu, Y., Hu, J., Li, S-H., Duchen, P., Wegmann, D., and Schweizer, M. (2016). Sino-Himalayan mountains act as cradles of diversity and immigration centers in the diversification of parrotbills (Paradoxornithidae). *Journal of Biogeography*, 43:1488-1501. doi:10.1111/jbi.12738.

Lovett, G. M., Burns, D. A., Driscoll, C. T., Jenkins, J. C., Mitchell, M. J., Rustad, L., Haeuber, R. (2007). Who needs environmental monitoring? *Frontiers in Ecology and the Environment*, 5:253-260. doi:10.1890/1540-9295(2007)5[253:WNEM]2.0.CO;2)

Ma, X., Kang, J., Chen, W., Zhou, C., and He, S. (2015). Biogeographic history and high-elevation adaptations inferred from the mitochondrial genome of Glyptosternoid fishes (Sisoridae, Siluriformes) from the southeastern Tibetan Plateau. *BMC Evolutionary Biology*, 15, 233. doi:10.1186/s12862-015-0516-9

Meng, H. H., Su, T., Gao, X. Y., Li, J., Jiang, X. L., Sun, H., and Zhou, Z. K. (2017). Warm–cold colonization: Response of oaks to the uplift of the Himalaya–Hengduan Mountains. *Molecular Ecology*, 26:3276-3294. doi:10.1111/mec.14092

Mims, M. C., & Olden, J. D. (2013). Fish assemblages respond to altered flow regimes via ecological filtering of life history strategies. *Freshwater Biology*, 58:50-62. doi:10.1111/fwb.12037.

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

Pellissier, L., Heine, C., Rosauer, D. F. and Albouy, C. (2017). Are global hotspots of endemic richness shaped by plate tectonics? *Biological Journal of the Linnean Society*, 123:247-261. doi:10.1093/biolinnean/blx125

Qi, D., Guo, S., Chao, Y., Kong, Q., Li, C., Xia, M., Xie, B., and Zhao, K. (2015). The biogeography and phylogeny of schizothoracine fishes (*Schizopygopsis*) in the Qinghai-Tibetan Plateau. *Zoologica Scripta*, 44:523-533. doi:10.1111/zsc.12116.

Radinger, J., Essl, F., Holker, F., Horky, P., Slavik, O., and Wolter, C. (2017). The future distribution of river fish: The complex interplay of climate and land use changes, species dispersal and movement barriers. *Global Change Biology*, 2017:1-17. doi:10.1111/gcb.13760.

Rambaut, A., 2012. FigTree v1. 4. Molecular evolution, phylogenetics and epidemiology. *Edinburgh: The University of Edinburgh, Institute of Evolutionary Biology. software*. (http://tree.bio.ed.ac.uk/software/figtree/)

Rambaut, A., and Drummond, A. J. (2007). http://beast.bio.edu.ac.uk/Tracer. *Tracer v1*, *4*.

Ran, Y., Li, X., and Cheng, G. (2018). Climate warming over the past half-century has led to thermal degradation of permafrost on the Qinghai–Tibet Plateau. *The Cryosphere*, 12:595-608. doi:10.5194/tc-12-595-2018

Renner, S. S. (2016). Available data point to a 4-km-high Tibetan Plateau by 40 Ma, but 100 molecular-clock papers have linked supposed recent uplift to young node ages. *Journal of Biogeography*, 43:1479-1487. doi:10.1111/jbi.12755.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhn, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012). MRBAYES 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61:539-542. doi:10.1093/sysbio/sys029.

Sakai, H., Sakai, H., Yahagi, W., Fujii, R., Hayashi, T., and Upreti, B. N. (2006). Pleistocene rapid uplift of the Himalayan frontal ranges recorded in the Kathmandu and Siwalik basins. *Paleogeography, Palaeoclimatology, Palaeoecology*, 241:16-27. doi:10.1016/j.palaeo.2006.06.017.

Scherrer, D., and Körner, C. (2011). Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *Journal of Biogeography*, 38: 406-416. doi:10.1111/j.1365-2699.2010.02407.x.

Schluter, D. (2016). Speciation, ecological opportunity, and latitude. *The American Naturalist*, *187*:1-18. doi:10.1086/684193

Sharma, C. M. (2008). Freshwater fishes, fisheries, and habitat prospects of Nepal. *Aquatic Ecosystem Health & Management*, 11:289-297. doi:10.1080/14634980802317329.

Shrestha, J. (1999). Coldwater fish and fisheries in Nepal. In T. Petr (Ed.), *Fish and Fisheries at Higher Altitudes: Asia* (pp. 13-40). Rome, Italy: FAO Fisheries Technical Paper. No. 385. (http://www.fao.org/docrep/003/X2614E/x2614e03.htm).

Shrestha, O. M, & Edds, D. R. (2012). Fishes of Nepal: Mapping distributions based on voucher specimens. *Emporia State Research Studies*, 48:14-21. (http://academic.emporia.edu/esrs/vol48/shrestha_edds.pdf).

Shrestha, U. B., Gautam, S., & Bawa, K. S. (2012). Widespread climate change in the Himalayas and associated changes in local ecosystems. *PLoS ONE*, 7, e36741. doi:10.1371/journal.pone.0036741.

Stamakakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30:1312-1313. doi:10.1093/bioinformatics/btu033.

Strange, R. M., and Burr, B. M. (1997). Intraspecific phylogeography of North American Highland fishes: A test of the Pleistocene vicariance hypothesis. *Evolution*, 51:885-897. doi:10.1111/j.1558-5646.1997.tb03670.x.

Sun, J. Zhou, T., Liu, M., Chen, Y., Shang, H., Zhu, L., Ali Shedayi, A., Yu, H., Cheng, G., Liu, G., Xu, M., Deng, W., Fan, J., Lu, X., and Sha, Y. (2018). Linkages of the dynamics of glaciers and lakes with the climate elements over the Tibetan Plateau. *Earth-Science Reviews*, 185:308-324. doi:10.1016/j.earscirev.2018.06.012

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30:2725-2729. doi:10.1093/molbev/mst197.

Tedesco, P., A., Leprieur, F., Hugueny, B., Brosse, S., Dürr, H., H., Beauchard, O.,Busson, F., and Oberdorff, T. (2012). Patterns and processes of global riverine fish endemism. *Global Ecology and Biogeography*, 21:977-987. doi:10.1111/j.1466-8238.2011.00749.x.

Terashima, A. (1984). Three new species of the cyprinid genus *Schizothorax* from Lake Rara, northwestern Nepal. *Japanese Journal of Ichthyology*, 31:122-135. doi:10.11369/jji1950.31.122.

Unmack, P.J., Bennin, A., Habit, E. M., Victoriano, P. F., and Johnson, J. B. (2009) Impact of ocean barriers, topography, and glaciation on phylogeography of the catfish *Trichomycterus areolatus* (Teleostei: Trichomycteridae) in Chile. *Biological Journal of the Linnean Society* 97:876-892. doi:10.1111/j.1095-8312.2009.01224.x

van Hinsbergen, D. J., Lippert, P. C., Dupont-Nivete, G., McQuarrie, N., Doubrovine, P. V., Spakman, W., and Torsvik, T. H. (2012). Greater India Basin hypothesis and a two-stage Cenozoic collision between India and Asia. *Proceedings of the National Academy of Sciences U.S.A.*, 109:7659-7664. doi:10.1073/pnas.1117262109.

Vetaas, O. R., and Grytnes, J. A. (2002). Distribution of vascular plant species richness and endemic richness along the Himalayan elevation gradient in Nepal. *Global Ecology and Biogeography*, *11*:291-301.

Wang, B., Bao, Q., Hoskins, B., Wu, G., and Liu, Y. (2012). Tibetan Plateau warming and precipitation changes in East Asia. *Geophysical Research Letters*, 35, L14702. doi:10.1029/2008GL034330

Wanghe, K., Tang, Y., Tian, F., Feng, C., Zhang, R., Li, G., Liu, S., and Zhao, K. (2017). Phylogeography of *Schizopygopsis stoliczkai* (Cyprinidae) in Northwest Tibetan Plateau area. *Ecology and Evolution,* 7:9602-9612. doi:10.1002/ece3.3452

Whipple, K. X. (2009). The influence of climate on the tectonic evolution of mountain belts. *Nature Geoscience*, 2:97-99. doi:10.1038/ngeo413.

Xia, X. (2013). DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, 30:1720-1728. doi:10.1093/molbev/mst064.

Xu, J., and Grumbine, R. E. (2014). Building ecosystem resilience for climate change adaptation in the Asian highlands. *WIREs Climate Change*, 5:709-718. doi:10.1002/wcc.302.

Xu, J., Grumbine R. E., Shresta, A., Eriksson, M., Yang, X., Wang, Y., and Wilkes, A. (2009). The melting Himalayas: Cascading effects of climate change on water, biodiversity, and livelihoods. *Conservation Biology*, 23:520-530. doi:10.1111/j.1523-1739.2009.01237.x.

Xu, W., Ma, L., Ma, M., Zhang, H., and Yuan W. (2017). Spatial-temporal variability of snow cover and depth in the Qinghai–Tibetan Plateau. *Journal of Climatology*, 30:1521-1533. doi:10.1175/JCLI-D-15-0732.1

Yang, J., Yang, J. X., and Chen, X. Y. (2012). A re-examination of the molecular phylogeny and biogeography of the genus *Schizothorax* (Teleostei: Cyprinidae) through enhanced sampling with an emphasis on the species in the Yunnan-Guizhou plateau, China. *Journal of Zoological Systematics and Evolutionary Research*, 50:184-191. doi:10.1111/j.1439-0469.2012.00661.x.

Yao, T., Thompson, L. G., Mosbrugger, V., Zhang, F., Ma, Y., Luo, T., Xu, B., Yang, X., Joswiak, D. R., Wang, W., Joswiak, M. E., Devkota, L. P., Tayal, S., Jilani, R., and Fayziev, R. (2012a). Third pole environment (TPE). *Environmental Development*, 3:52-64. doi:10.1016/j.envdev.2012.04.002

Yao, T., Thompson, L., Yang, W., Yu, W., Gao, Y., Guo, X., Yang, X., Duan, K., Zhao, H., Xu, B., Pu, J., Lu, A. Xiang, Y., Kattel, D.B., and Joswiak, D. (2012b). Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings. *Nature Climate Change*, 2:663-667. doi:10.1038/nclimate1580

Zakaria G. (2018). Late Quaternary Seismicity and Climate in Western Nepal: Himalaya. Applied geology. Unpubl. Ph.D. dissertation. Université Ghent, Department of Geology, Grenoble, France, Ghent, Belgium. (https://tel.archives-ouvertes.fr/tel-01935412/document)

Zardoya, R., and Doadrio, I. (1999). Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *Journal of Molecular Evolution*, 49: 227-237. doi:10.1007/PL00006545

Zhao, K. I., Duan, Z. Y., Peng, Z. G., Guo, S. C., Li, J. B., He, S. P., and Zhao, X. Q. (2009). The youngest split in sympatric Schizothoracine fish (Cyprinidae) is shaped by ecological adaptations in a Tibetan Plateau glacier lake. *Molecular Ecology*, 18:3616-3628. doi:10.1111/j.1365-294X.2009.04274.x

Zhou, C., Wang, X., Gan, X., Zhang, Y., Irwin, D. M., Mayden, R. L. and He, S., (2016). Diversification of Sisorid catfishes (Teleostei: Siluriformes) in relation to the orogeny of the Himalayan Plateau. *Science Bulletin*, *61*:991-1002. doi:10.1007/s11434-016-1104-0

Appendix

Table 1: List of species/ sequences of *Schizothorax* analyzed in this study to evaluate phylogenetic patterns, Listed are Species Name= scientific name, Accession#= tissue or sequence identifier, representing either GenBank accession number (starting with two letters), KU tissue voucher number (four digits) or Douglas Lab DNA code (starting with 58), Location= geographic region or major river basin, and Area= specific locality information.

Species Name	**Accession #**	**Location**	**Area**
G. eckloni	DQ309350	China: Madoi	outgroup
G. przewalskii	DQ309362	China: Qinghai	outgroup
S. richardsonii	KP712249	Nepal: Indrawati river	Koshi
S. richardsonii	1964	Nepal:Arun river	Koshi
S. esocinus	KP712250	Nepal: Melamchi	Koshi
S. progastus	1974	Nepal:Arun River	Koshi
S. richardsonii	1972	Nepal:Arun River	Koshi
S. biddulphi	FJ931466	China: Kezi River	Central Asia
S. intermedius	AY954273	China: Taxkorga	Central Asia
S. biddulphi	FJ931464	China: Taxkorga	Central Asia
S. intermedius	AY954272	China: Uqturpan	Central Asia
S. eurystomus	AY954275	China: Taxkorga	Central Asia
S. biddulphi	FJ931467	China: Kezi River	Central Asia
S. pseudoaksaiensis	AY954270	China: Ili	Central Asia
S. argentatus	AY954269	China: Ili	Central Asia
S. progastus	1956	Nepal:Beni	Gandaki
S. progastus	1955	Nepal: Beni	Gandaki
S. progastus	1957	Nepal: Beni	Gandaki
S. richardsonii	8820	Nepal: Beni	Gandaki
S. macrophthalmus	1932	Nepal: Rara Lake	Lake Rara
S. richardsonii	1935	Nepal Shrikot	Karnali

Table 1(cont…)

Species Name	**Accession #**	**Location**	**Area**
S. nepalensis	1794	Nepal: Rara Lake	Lake Rara
S. plagiostomus	KR232372	Pakisthan: Jhelum	Indus
S. labiatus	AY954281	China: Bangong Co	Southeast Asia
S. richardsonii	JX485902	India: Upper Ganges	Upper Ganges
S. richardsonii	JX485901	India: Upper Ganges	Upper Ganges
S. richardsonii	JX485884	India: Upper Ganges	Upper Ganges
S. richardsonii	JX485883	India: Upper Ganges	Upper Ganges
S. richardsonii	JX485882	India: Upper Ganges	Upper Ganges
S. molesworthi	DQ126129	China: Zayu, Tibet	Southeast Asia
S. yunnanensis	KP796171	China: Nujiang River	Southeast Asia
S. yunnanensis	KP796169	China: Nujiang River	Southeast Asia
S. dulongensis	AY954284	China: Tengchong	Southeast Asia
S. rotundimaxillaris	AY954283	China: Tengchong	Southeast Asia
S. yunnanensis	AY954286	China: Tengchong	Southeast Asia
S. meridionalis	AY954287	China: Tengchong	Southeast Asia
S. malacanthus	AY954277	China: Diantian	Southeast Asia
S. oconnori	GQ495883	China: Tsangpo River	Central Tibet
S. oconnori	GQ495873	China: Tsangpo River	Central Tibet
S. wangchiachii	HQ198881	China: Tsangpo River	Central Tibet
S. oconnori	GQ495871	Yalung Tsangpo River	Central Tibet
S. oconnori	JQ082339	China: Lhasa River	Central Tibet
S. oconnori	AY463519	China: Lahsa, Tibet	Central Tibet
S. oconnori	GQ495881	China: Bangong, Tibet	Central Tibet

Table 1(cont..)

Species Name	**Accession #**	**Location**	**Area**
S. oconnori	GQ495889	China: Tsangpo R.	Central Tibet
S. macropogon	AY463517	China: Lahsa	Central Tibet
S. waltoni	HM536804	China: Xizang	Central Tibet
S. davidi	DQ126113	China: Ya'an	Southeast Asia
S. kozlovi	DQ126112	China: Jiulong	Southeast Asia
S. prenanti	AY954262	China: Fenjie	Southeast Asia
S. chongi	DQ126118	China: Jiulong	Southeast Asia
S. dolichonema	DQ126117	China: Jiulong	Southeast Asia
S. prenanti	AY954261	China: Judian	Southeast Asia
S. wangchiachii	DQ126121	China: Jiulong	Southeast Asia
S. griseus	AY954253	China: Yongping	Southeast Asia
S. yunnanensis	AY954252	China: Yongping	Southeast Asia
S. lissolabiatus	KP796159	China: YangBi River	Southeast Asia
S. lissolabiatus	AY954251	China: Yongping	Southeast Asia
S. lantsangensis	DQ646882	China: Lancang River	Southeast Asia
S. nukiangensis	DQ126125	China: Chalong	Southeast Asia
S. gongshanensis	AY954279	China: Fugong	Southeast Asia
S. lissolabiatus	KP796167	China: Nujiang River	Southeast Asia
S. gongshanensis	AY954280	China: Fugong	Southeast Asia
B. callensis	AF045974	Algeria: Kabir river	outgroup
B. sclateri	AF045970	Spain:Guadalquivir ri.	outgroup
B. guiraon	AF045972	Spain: Buyent river	outgroup

Table 1(cont…)

Species Name	Accession #	Location	Area
S. species	58DAK009	Dakpai Chhu	Lower Brahmaputra
S. species	58DAKP09	Dakpai Chhu	Lower Brahmaputra
S. species	58DIKC03	Dik Chhu	Lower Brahmaputra
S. species	58DIKC04	Dik Chhu	Lower Brahmaputra
S. species	58KAME05	Kame Chhu	Lower Brahmaputra
S. species	58KAME06	Kame Chhu	Lower Brahmaputra
S. species	58MABI03	Bibigang Chhu	Lower Brahmaputra
S. species	58MABIO08	Bibigang Chhu	Lower Brahmaputra
S. species	58MANG03	Mangde Chhu	Lower Brahmaputra
S. species	58MANG04	Mangde Chhu	Lower Brahmaputra
S. species	58NYAC02	Nya Chhu	Lower Brahmaputra
S. species	58NYAC01	Amo Chhu	Lower Brahmaputra
S. species	58NYACO08	Amo Chhu	Lower Brahmaputra
S. species	58PACH01	Amo Chhu	Lower Brahmaputra
S. species	58PAIPP01	Wang Chhu	Lower Brahmaputra
S. species	58TOEB02	Punatsang Chhu	Lower Brahmaputra
S. species	58TOEB017	Punatsang Chhu	Lower Brahmaputra
S. species	58TOEB18	Punatsang Chhu	Lower Brahmaputra

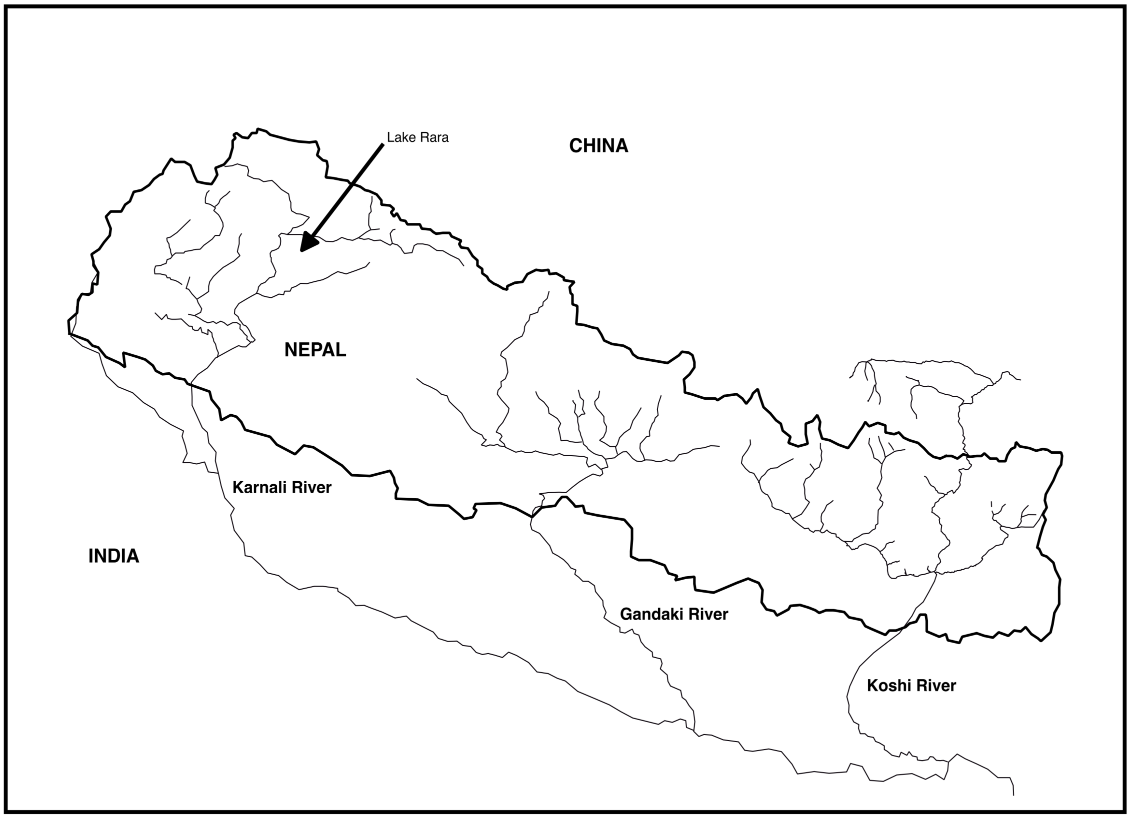

Figure 1: Drainage patterns of the Central Himalayas; Three major river basins drain Nepal before merging with the Lower Ganges River in the lowlands of India. Lake Rara (northeast) drains into the Karnali River. Details on collection localities of *Schizothorax* samples are provided in Table 1.

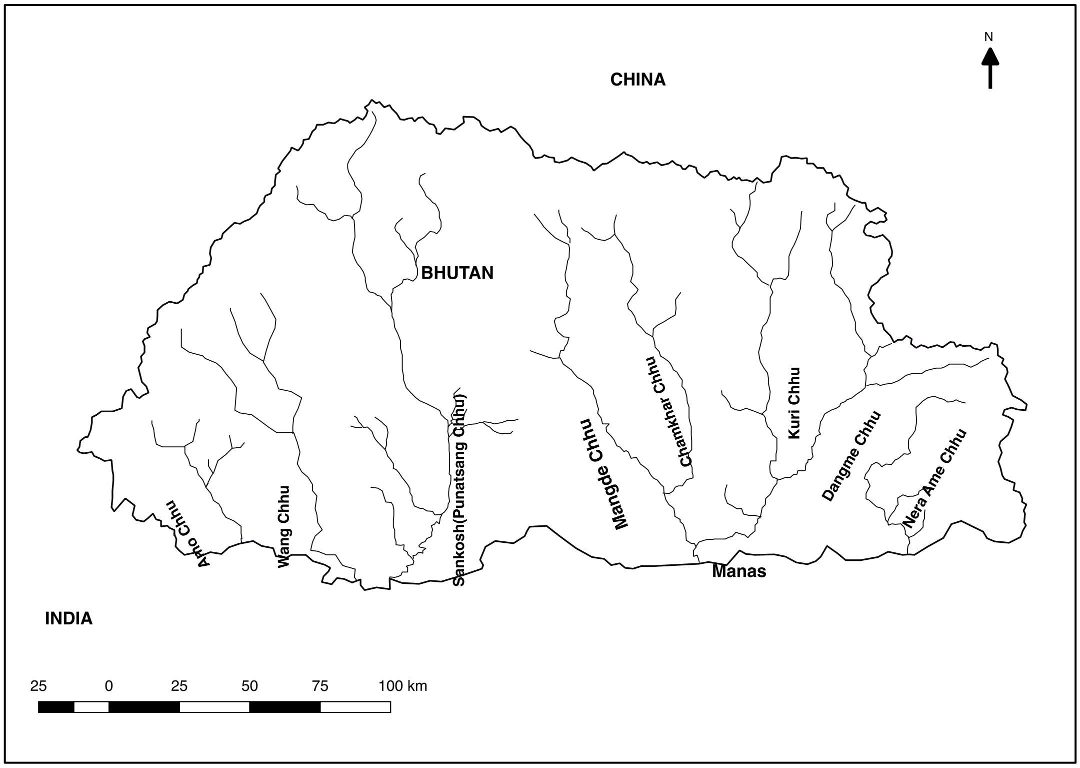

Figure 2: Drainage patterns of the Eastern Himalayas in Bhutan. Six major rivers drain into the Brahmaputra River to the south, in India. *Schizothorax* samples included in this study were collected in the Amo Chhu, Wang Chhu, Punatsang Chhu/Sunkosh, and Mangde Chhu/Manas. Details on collection localities is provided in Table 1.

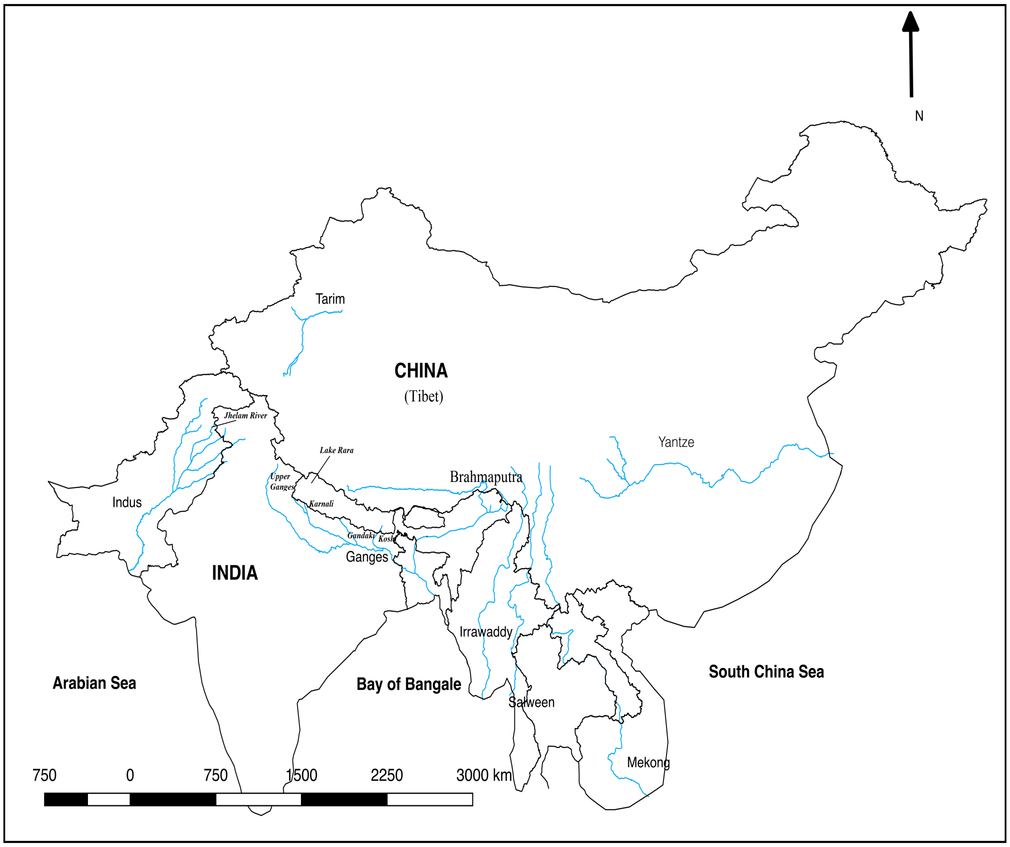

Figure 3: Drainage system of Asia (simplified) showing the major basins from which *Schizothorax* samples were obtained. Regions from northwest to southeast include the following major basins and their tributaries: **Central Asia**: Tarim River; **Central Tibet**: Upper Brahmaputra; **Western Himalayas**: Indus and Upper Ganges rivers; **Central Himalayas**: Lower Ganges; **Eastern Himalayas**: Lower Brahmaputra; **Southeast Asia**: Irrawady, Salween, Mekong and Yangtze rivers.

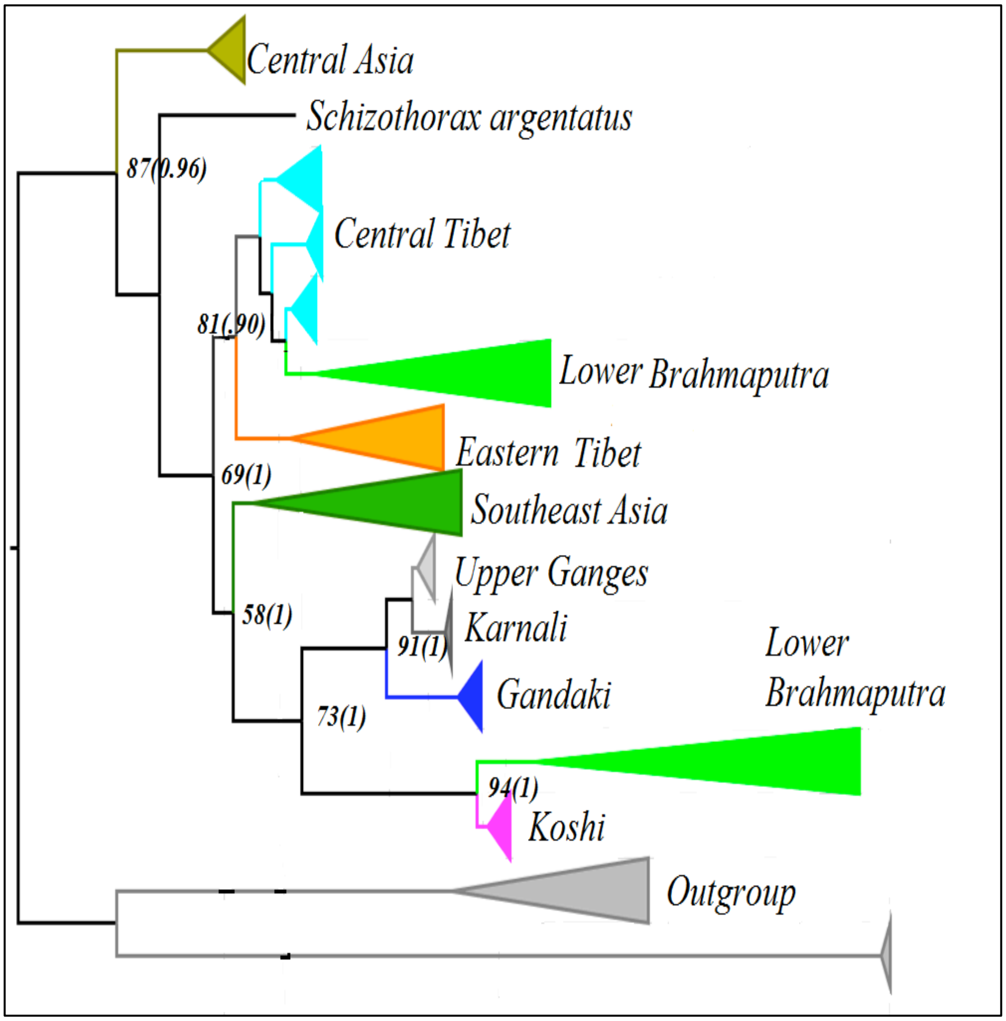

Figure 4: Maximum likelihood (RAxML) tree reflecting phylogenetic relationships among *Schizothorax* species. Data were derived from sequence analysis of the cytochrome b mitochondrial gene (1140 bp, 32 new plus approx. 200 GenBank sequences). Redundant sequences were removed for analyses, leaving 83 ingroup and 5 outgroup haplotypes. Clades were collapsed, and values denote bootstrap support and posterior probability obtained via Bayesian analyses. The list of sequences used for phylogenetic analysis is presented in Table 1. Regions and drainage basins are depicted in Figures 1-3.

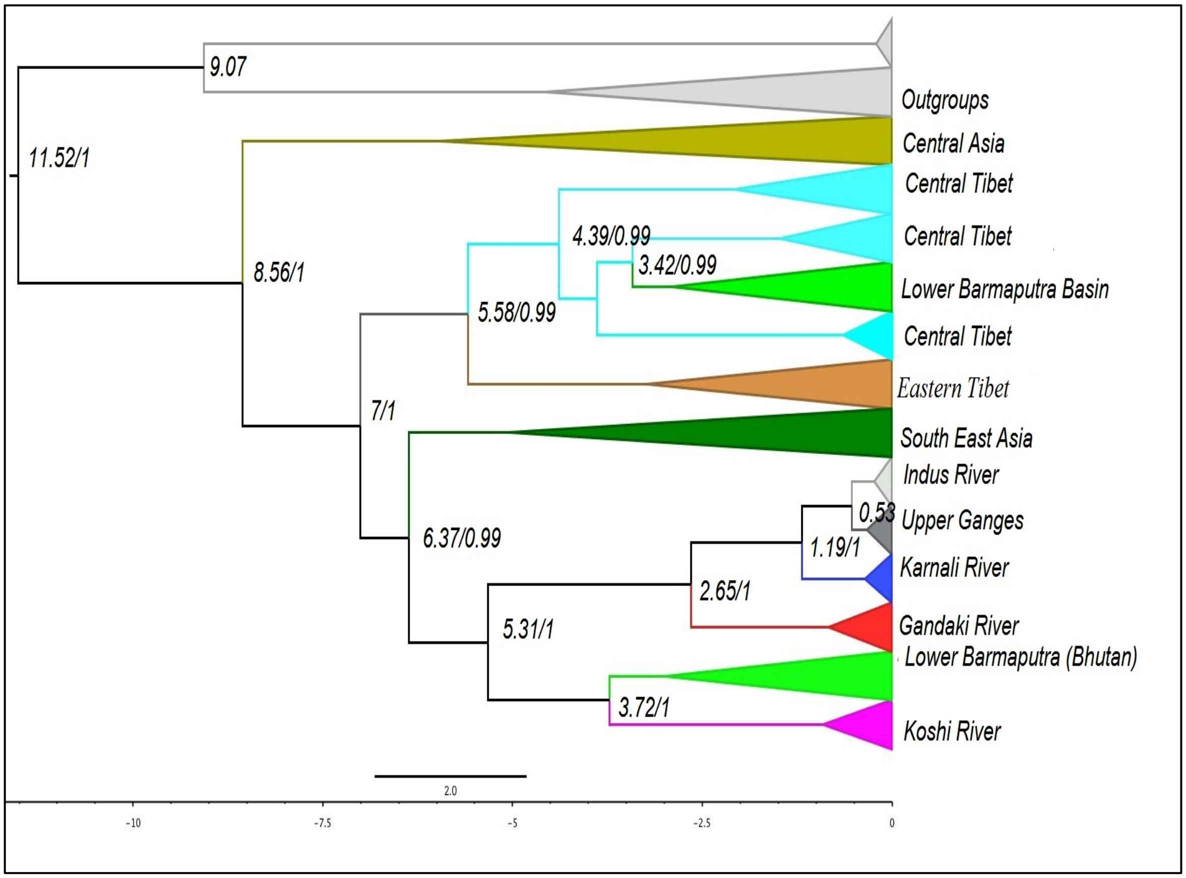

Figure 5: Time-calibrated Bayesian tree generated in BEAST reflecting phylogenetic relationships among *Schizothorax* species. Data were derived from sequence analysis of the cytochrome b mitochondrial gene (1140 bp, 32 new plus approx. 200 GenBank sequences). Redundant sequences were removed for analyses, leaving 83 ingroup and 5 outgroup haplotypes. Clades were collapsed. Node support represent divergence time and posterior probability. The list of sequences used for phylogenetic analysis is presented in Table 1. Regions and drainage basins are depicted in Figures 1-3.

Chapter V:

Species flocks of Snowtrout (*Schizothorax*) in the Central Himalayas: convergent evolution along elevation gradients

Abstract

The goal of this study was to re-examine fine-scale phylogeographic patterns in five species of *Schizothorax* among three drainages in Central Himalayas (Koshi, Gandaki, and Karnali River basins in Nepal). Two riverine species, *S. richardsonii* and S. *progastus*, are more widely distributed, whereas three lacustrine species, *S. macrophthalmus*, *S. nepalensis*, and *S. raraensis*, are endemic to Lake Rara in northwestern Nepal (Karnali River Basin). Double-digest restriction associated DNA (ddRAD) sequencing was used to screen 20,000 single nucleotides polymorphic (SNPs) loci across 48 samples. PYRAD pipeline was used to demultiplex, filter, align and annotate SNP loci and a phylogenetic tree was constructed using RAxML. Phylogeographic patterns showed a strong association of samples within-basins, independent of species identification. However, recognized morpho-species segregated into monophyletic clades within basins. The three species endemic to Lake Rara also formed distinct clades, with S. *nepalensis* as sister to a paraphyletic *S. macrophthalmus* and a monophyletic *S. raraensis.* Three monophyletic clades in the Gandaki River Basin suggest the potential for a currently un-recognized, high-elevation ecotype. This phylogenomic study gives new insights into the evolutionary history of morphological species along elevation gradients in Nepal, where mitochondrial DNA failed to resolve fine-scale genetic structure within basins.

Introduction

One of the most common approaches to identify and delimit biological diversity is the application of phylogenetic methods to reconstruct the evolutionary history of different phenotypes (Hunt & Slater 2016). Traditionally, phylogenetic analyses were conducted by sequencing one or a few genes, with genotypes based on the nucleotide composition of a single locus or concatenated genes. Because of the relatively slow evolutionary rate of most commonly sequenced genes (i.e., mitochondrial(mt)DNA or nuclear(n)DNA loci), these data had limited power to resolve recently diverged clades, as might be the case with *Schizothorax* (Snowtrout) endemic to Lake Rara in the Central Himalayas.

The advent of massive-parallel sequence technologies, referred to as Next Generation Sequencing (NGS), has brought efficient and relatively cost-effective approaches to sequence entire genomes or a subset of selected genes (i.e., reduced-representation libraries). These methods allow the generation of massive amounts of genetic information across hundreds or thousands of individuals and generate data amenable for phylogenetic and population genetic assays of biodiversity (Peterson et al., 2012).

One type of marker, Single Nucleotide Polymorphisms (SNPs), the most common mutation in the genome, has been widely adopted because of the relative ease at which data can be generated. Screening of genome-wide variation of SNPs are screened is accomplished by double-digest restriction site associated DNA sequencing (ddRADseq). This approach subsamples the genome in a repeatable manner across many individuals using the restriction cut sites to and size selection of resulting fragments to ‘reduce’ the genome in a consistent manner (Peterson et al., 2012). Subsampling a defined set of loci is an important aspect because genomes of fishes are relatively large at 1-2GBs, and up to 4+GBs in case of genome duplications or polyploidization (Volff, 2005).

These NGS approaches provide unprecedented resolution not only for phylogenetic reconstructions but also landscape/riverscape genetics to examine biodiversity within a temporal and spatial framework via integration of various types of data (Hudson, 2008) (e.g., geologic, ecologic, climate). This allows examining historical and contemporary diversification and inferences about micro- as well as macro-evolutionary processes.

The goal of this study was to, re-examine fine-scale phylogeographic patterns in *Schizothorax* among drainages in Central Himalayas using ddRADseq. My previous evaluations of *Schizothorax* across the region used geometric morphometrics (Chapter 2) and mitochondrial DNA sequence analysis (Chapter 4). These analyses revealed distinct evolutionary lineages defined by drainage patterns (major basins) and showed that currently recognized taxa poorly reflected these phylogenetic patterns.

To reconstruct a more detailed phylogenetic history of *Schizothorax* species distributed in lakes and rivers of the Central Himalayas I examined five species from the three major drainages in Nepal (Koshi, Gandaki, and Karnali River basins; Figure 1). Two riverine species, *S. richardsonii* and *S. progastus,* are common and more widely distributed in the region, but segregate along elevational gradients, with S. *richardsonii* distributed at high elevation, lower order streams and *S. progastus* found in lower elevation, higher order reaches of rivers (Edds, 1989). Three lacustrine species, *S. macrophthalmus, S. nepalensis,* and *S. raraensis*, are morphologically differentiated and reproductively isolated (Terashima, 1984). These putative species are endemic to the small temperate lacustrine system, Lake Rara in the Karnali River basin in northwestern Nepal (Terashima, 1984; Dimmick & Edds, 2002). The striking differences in morphology, behavior, and diet of these sympatric, lake-endemic species are remarkable and might represent a case of recent radiation (i.e., a species flock) (Terashima, 1984). I hypothesized that differential food and habitat conditions could have maintained genetic differences among these riverine (i.e. downstream and upstream) and lacustrine species.

Dimmick & Edds (2002) evaluated these taxa using mitochondrial (mt)DNA sequence analysis, but could only differentiate *S. nepalensis* from the other two lacustrine species. Similarly, mtDNA was unable to discriminate two morphologically distinctive riverine species in the upper Karnali River (i.e., *S. richardsonii* and *S. progastus*). This inconsistency between geographic and morphological patterns warrants evaluating putative species with a more in-depth analysis. The single mitochondrial gene (cytochrome b) used in the Chapter 2 may have insufficient resolution for gauging the extent of diversity in *Schizothorax* from the Central Himalayas. More importantly, genetic patterns revealed by the single mitochondrial gene could reflect ancestral polymorphism (retained ancient alleles), a likely scenario in recently diverged species such as *Schizothorax* (Dimmick & Edds, 2002; He & Chen, 2006; Chen et al., 2017).

Methods

Sampling, filtering, and alignment of genomic DNA

Restriction site associated DNA sequencing (RADseq) is a relatively new and cost-effective method to assay thousands of molecular loci in non-model organisms (Puritz et al., 2014). The approach is suitable for phylogenetics, phylogeography, and population genomic studies (Peterson et al., 2012; Cavender-Bares et al., 2015; Bangs et al., 2018). The digestion of genomic DNA with restriction enzymes is the first step, followed by the sequencing of genomic regions flanking the cut sites. The genetic variation screened with this approach primarily represents single nucleotide polymorphisms (SNPs) accrued across thousands of loci (Baird et al., 2008).

DNA was extracted with DNeasy® Tissue Kit (Qiagen Inc., Valencia CA) and stored in DNA hydrating solution (same kit). Libraries for ddRAD sequencing were generated following laboratory protocols, to include digesting with *Pst*I (5'-CTGCAG-3') and *Msp*I (5'-CCGG-3'). Digests were visualized on 2% agarose gels, cleaned using AMPure XP beads, and quantified using a Qubit fluorometer. Approximately 0.1µg of DNA was then ligated with barcoded Illumina adaptors, using custom oligos (Peterson et al., 2012). All barcodes differed by at least two bases so as to avoid fragment misassignment.

Samples for each reference species and location were randomly distributed across several libraries and lanes so as to reduce the potential for library preparation or lane bias. Ligations were pooled in sets of 48, cleaned/ concentrated using AMPure XP beads, then size-selected at 350-400 bps using a Pippin'Prep automated size fractionator (Sage Sciences). Size-selected DNA served as the template for Phusion high-fidelity DNA polymerase reactions using indexed primers and 10 cycles, following the manufacturer's protocol (New England BioSciences). Reactions were again cleaned with AMPure XP beads and visualized on the Agilent 2200 TapeStation to confirm successful amplification. Sequencing was performed at the University of Oregon Genomics & Cell Characterization Core Facility (GC3F; Eugene, OR).

All analyses were conducted on the Arkansas High-Performance Computing Cluster (AHPCC) at the University of Arkansas. Illumina reads were filtered and aligned (per lab protocols) using PYRAD v.3.0.5 (Eaton & Ree, 2013). This included a clustering threshold of 80% and removal of restriction site sequence and barcodes. In addition, loci were removed if they displayed: 1) <5 reads per individual), 2) >10 heterozygous sites within a consensus, 3) >2 haplotypes for an individual, 4) >75% heterozygosity for a site among individuals, and 5) <50% of individuals at a given locus (per Leaché et al., 2015). Individuals with more than 80% missing data were discarded.

Phylogenetic methods

Loci produced in PYRAD was used to generate phylogenies based on concatenated SNPs. A maximum likelihood (ML) phylogeny (RAxML v.7.3.2; Stamatakis, 2014) was developed using GTRCAT with 1,000 bootstraps. The resulting trees were visualized using FIGTREE (Rambaut, 2012).

Results

I recovered approximately 20,000 SNPs through the implementation of the PYRAD pipeline. The RAXML phylogenetic tree reflected a very strong geographic signal with samples clustering by basins into three highly supported monophyletic clades (Figure 2). However, within basins, samples clustered by morphological species, with *S. richardsonii* and *S. progastus* each forming monophyletic clades within basins, with few exceptions. Furthermore, samples from Lake Rara clustered as monophyletic clade sister to a clade formed by *S. richardsonii* from the Karnali River Basin.

Within the Lake Rara clade, samples again clustered by morphological species. *Schizothorax nepalensis* formed a clade sister to samples representing *S. macrophthalmus* and *S. raraensis*, with the latter forming a monophyletic clade sister to a paraphyletic *S. macrophthalmus*. These results clearly support *in-situ* radiation of Lake Rara endemics forming a lacustrine species flock comprising of distinct ecotypes adapted to different niches.

As noted above, there were some exceptions from these distinct patterns that reflect divergence by basin, and within basins by morphological species. Three samples from the Gandaki River that were morphologically identified as *S. richardsonii,* formed a separate clade sister to the other samples from the basin. Interestingly, these samples were collected from the most upstream sites, at higher elevation than the remaining *S. richardsonii* and hence could represent another elevational-selected ecotype.

Two samples from the Karnali River Basin, one identified morphologically as *S. richardsonii* and the other as *S. macrophthalmus*, formed a distinct clade sister to the remaining Lake Rara samples, with the former collected about 20km downstream from Lake Rara, whereas the latter was sampled near the lake outlet. Introgressive hybridization could be one possibility, but further analyses would be required to test this hypothesis. I implemented multiple clustering thresholds for extracting the SNP data, but resulting tree topologies using stringent clustering were consistent across minor changes in filtering parameters. This suggests that the presented tree topology is robust and accurately reflects genetic diversity represented in *Schizothorax* from Nepal.

Discussion

A strong geographic pattern was apparent in *Schizothorax* from the Central Himalayas, clearly reflected in phylogenies derived from mitochondrial gene sequence analysis (Chapter 4) and genome-wide SNP data (this Chapter). However, while the mitochondrial data could not distinguish among different morphological species from the Karnali River Basin, the genomic data clearly separated morpho-species within basins.

The shallow biogeographic history and recent drainage evolution of the Central Himalayas are consistent with these patterns. Mitochondrial DNA simply lacks the resolution to distinguish recently diverged species, as seems to be the case for *Schizothorax* in Nepal. A major drawback of ddRADseq is the accumulation of restriction site polymorphisms as sequences diverge over time (Dacosta & Sorenson, 2016). Previous studies indicate that the *Schizothorax* species radiated in the late Miocene in Southeastern China or Tibet (He &Chen, 2006), suggesting that decay of restriction sites among the species distributed in the eastern to western parts of the Central Himalayan drainages may not be an issue.

The current drainage patterns in Himalayan rivers were shaped by uplift of the Himalayan and Tibetan Plateau followed by tectonic activities in late Miocene to Quaternary period and subsequent changes in monsoonal precipitation that eroded the landscape into steep drainages confined in narrow gorges Previously, the Central Himalayans rivers flowed independently into the Gangetic plain, but the rise of the Lesser Himalayas (<5mya) and tectonic activities around the Frontal Himalayan Thrust (FHT) pushed tributaries of major rivers (i.e. Koshi, Gandaki, and Karnali) into common drainage before crossing the Lesser Himalayan range. This newly evolved drainage pattern might have increased connectivity between tributaries within the major basins, but basins remained separated before reaching the lower elevation of the Gangetic Plane in India (Figure 2).

Schizothorax species originated in the Tibet and dispersed along Himalayan rivers, but were isolated within basins following the tectonic changes described above. Central Himalayan rivers comprise a series of very distinct ecosystems as turbulent rivers cascade in steep gradients from higher elevations to drastically different slow-flowing lowland rivers. Strong selection gradients are associated with these distinct ecosystems and were apparently drivers of diversification in *Schizothorax*. While separation into isolated headwaters occurred in recent geologic time, gene flow between upstream and downstream was likely impeded by the strong selection for local adaptation, promoting the emergence of morphologically similar phenotypes in each basin. Redundant convergent evolution of morphotypes likely confounded the taxonomic diagnosis of morphological species, similar to well-known examples of recent species flocks, such as Central Alpine *Coregonus* (Douglas et al., 1999).

Conclusion

Genomic patterns revealed by SNP analysis in this study suggest that biodiversity among *Schizothorax* in the Himalayas is likely much higher than currently recognized. Specimens from the Gandaki River lend support to this notion. Sampling in the Gandaki River was more extensive than in the other drainages (Edds, 1989) extending to reaches at much higher elevations. Interestingly, while SNP data clearly cluster specimens from other drainages into two distinct clades, comprising samples from upper-reaches *versus* lower-reaches, respectively, *Schizothorax* in the Gandaki River segregated into three distinct monophyletic clades corresponding to an elevational gradient. This raises the possibility that presence of a third, high-elevation ecotype distinct from the currently recognized *S. richardsonii* and *S. progastus* morphotypes. Additional analyses are required to further clarify diversity in *Schizothorax* in the Central Himalayas.

References:

Baird N. A., Etter P. D., Atwood T. S. et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:1-7. doi: 10.1371/journal.pone.0003376.

Bangs, M. R., Douglas, M. R., Mussmann, S. M., & Douglas, M. E. (2018). Unraveling historical introgression and resolving phylogenetic discord within *Catostomus* (Osteichthys: Catostomidae). *BMC Evolutionary Biology*, 18, 86. doi:10.1186/s12862-018-1197-y.

Cavender-Bares, J., González-Rodríguez, A., Eaton, D. A., Hipp, A. A., Beulke, A. and Manos, P. S. (2015). Phylogeny and biogeography of the American live oaks (*Quercus* subsection Virentes): a genomic and population genetics approach. *Molecular Ecology*, 24:3668-3687.

Chen, W., Yue, X. & He, S. (2017). Genetic differentiation of the Schizothorax species complex (Cyprinidae) in the Nujiang River (upper Salween). *Scientific Reports*, 7, 5944.

DaCosta, J. M. and Sorenson, M. D. (2016). ddRAD-seq phylogenetics based on nucleotide, indel, and presence-absence polymorphisms: Analyses of two avian genera with contrasting histories. *Molecular Phylogenetics and Evolution*, 94:122-135.

Dimmick, W. W. and Edds, D. R. (2002). Evolutionary genetics of the endemic schizorathicine (Cypriniformes: Cyprinidae) fishes of Lake Rara, Nepal. *Biochemical Systematics and Ecology*, *30*:919-929.

Douglas, M. R., Brunner, P. C. and Bernatchez, L. (1999). Do assemblages of *Coregonus* (Teleostei: Salmoniformes) in the Central Alpine region of Europe represent species flocks?. *Molecular Ecology*, 8:589-603.

Eaton, D. A. and Ree, R. H. (2013). Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, 62:689-706.

Edds, D. R. (1989). Multivariate analysis of fish assemblage composition and environmental correlates in a Himalayan River-Nepal's Kali Gandaki/Narayani. Doctoral dissertation, Oklahoma State University.

He, D., and Chen, Y. (2006). Biogeography and molecular phylogeny of the genus *Schizothorax* (Teleostei: Cyprinidae) in China inferred from cytochrome b sequences. *Journal of Biogeography*, 33:1448-1460. doi: 10.1111/j.1365-2699.2006.01510.x

Hudson, M. E. (2008). Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources*, *8*:3-17.

Hunt, G., and Slater, G. (2016). Integrating paleontological and phylogenetic approaches to macroevolution. *Annual Review of Ecology, Evolution, and Systematics,* 47.

Leaché A. D., Banbury B. L., Felsenstein J, et al. (2015). Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, 64:1032-1047. doi: 10.1093/sysbio/syv053.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS ONE*, 7, p.e37135.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I. and Bird, C. E. (2014). Demystifying the RAD fad. *Molecular Ecology*, 23:5937-5942.

Rambaut, A. (2012). FigTree v1. 4. Molecular evolution, phylogenetics, and epidemiology. *Edinburgh: The University of Edinburgh, Institute of Evolutionary Biology. software*.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30:1312-1313.

Terashima, A. (1984). Three new species of the cyprinid genus Schizothorax from Lake Rara, Northwestern Nepal. *Japanese Journal of Ichthyology*, 31:122-135.

Volff, J. N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity*, 94, 280.

Appendix

Table1: List of the tissue samples sequenced for SNPs data generation. The first column is the name of sequence recorded in Douglas Lab, Department of Biology, University of Arkansas. Column 3 presents the catalog number of tissue deposit in the University of Kansas, Natural History Museum, Lawrence, Kansas

Sample name (Douglas Lab)	Species	KU (Catalog	Locality	Basin
58CSHIZ085	*S. richardsonii*	1959	Larjung	Gandaki
58CSHIZ087	*S. richardsonii*	1961	Larjung	Gandaki
58CSHIZ099	*S. richardsonii*	1984	Galeshowar	Gandaki
58CSHIZ113	*S. richardsonii*	8817	Galeshowar	Gandaki
58CSHIZ117	*S. richardsonii*	8828	Galeshowar	Gandaki
58CSHIZ119	*S. richardsonii*	8831	Galeshowar	Gandaki
58PSHIZ076	*S. progastus*	8824	Galeshowar	Gandaki
58PSHIZ107	*S. progastus*	1993	Galeshowar	Gandaki
58PSHIZ097	*S. progastus*	1980	Beni	Gandaki
58PSHIZ037	*S. progastus*	1954	Beni	Gandaki
58PSHIZ077	*S. progastus*	8829	Galeshowar	Gandaki
58CSHIZ078	*S. richardsonii*	1938	Srikot	Karnali
58MSHIZ019	*S. macrophthalmus*	1933	Rara	Rara
58NSHIZ034	*S. nepalensis*	1930	Rara	Rara
58MSHIZ008	*S. macrophthalmus*	1902	Rara	Rara
58RSHIZ058	*S. raraensis*	2008	Rara	Rara
58NSHIZ016	*S. nepalensis*	1929	Rara	Rara
58NSHIZ002	*S. nepalensis*	1793	Rara	Rara
58NSHIZ028	*S. nepalensis*	1796	Rara	Rara
58NSHIZ012	*S. nepalensis*	1907	Rara	Rara
58NSHIZ010	*S. nepalensis*	1905	Rara	Rara
58NSHIZ011	*S. nepalensis*	1906	Rara	Rara
58NSHIZ009	*S. nepalensis*	1904	Rara	Rara
58MSHIZ067	*S. macrophthalmus*	2017	Rara	Rara
58MSHIZ068	*S. macrophthalmus*	2018	Rara	Rara
58MSHIZ066	*S. macrophthalmus*	2016	Rara	Rara

Table 1 (cont…)

Sample name (Douglas Lab)	**Species**	**KU (Catalog number)**	**Locality**	**Basin**
58MSHIZ031	*S. macropthalmus*	1903	Rara	Rara
58RSHIZ029	*S. raraensis*	1798	Rara	Rara
58RSHIZ007	*S. raraensis*	1900	Rara	Rara
58RSHIZ014	*S. raraensis*	1927	Rara	Rara
58RSHIZ055	*S. raraensis*	2005	Rara	Rara
58RSHIZ005	*S. raraensis*	1797	Rara	Rara
58RSHIZ057	*S. raraensis*	2007	Rara	Rara
58RSHIZ054	*S. raraensis*	2004	Rara	Rara
58MSHIZ018	*S. macrophthalmus*	1932	Rara	Rara
58MSHIZ017	*S. macrophthalmus*	1931	Rara	Rara
58CSHIZ111	*S. richardsonii*	2001	Gargari, Mugu	Karnali
58CSHIZ109	*S. richardsonii*	1999	Kaligandaki	Gandaki
58CSHIZ080	*S. richardsonii*	1945	Jhugala	Karnali
58CSHIZ084	*S. richardsonii*	1949	Jhugala	Karnali
58PSHIZ050	*S. progastus*	1978	Tumlingtar	Koshi
58PSHIZ048	*S. progastus*	1976	Tumlingtar	Koshi
58PSHIZ024	*S. progastus*	1974	Tumlingtar	Koshi
58CSHIZ093	*S. richardsonii*	1972	Num	Koshi
58CSHIZ092	*S. richardsonii*	1971	Num	Koshi
58CSHIZ042	*S. richardsonii*	1964	Num	Koshi
58CSHIZ090	*S. richardsonii*	1969	Num	Koshi
58CSHIZ074	*S. richardsonii*	8820	Beni	Gandaki

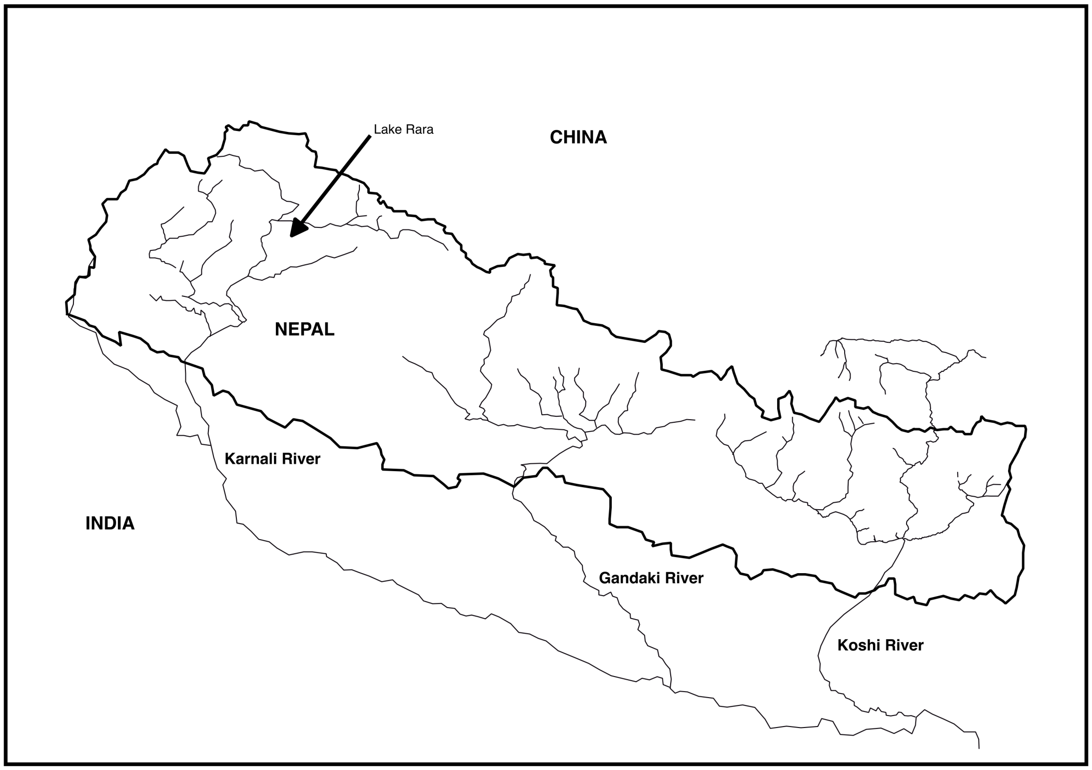

Figure 1: Geographic location of three major drainage basins in the Central Himalayas (Nepal). Tributaries of each major river basin converge before crossing the Lesser Himalayas.

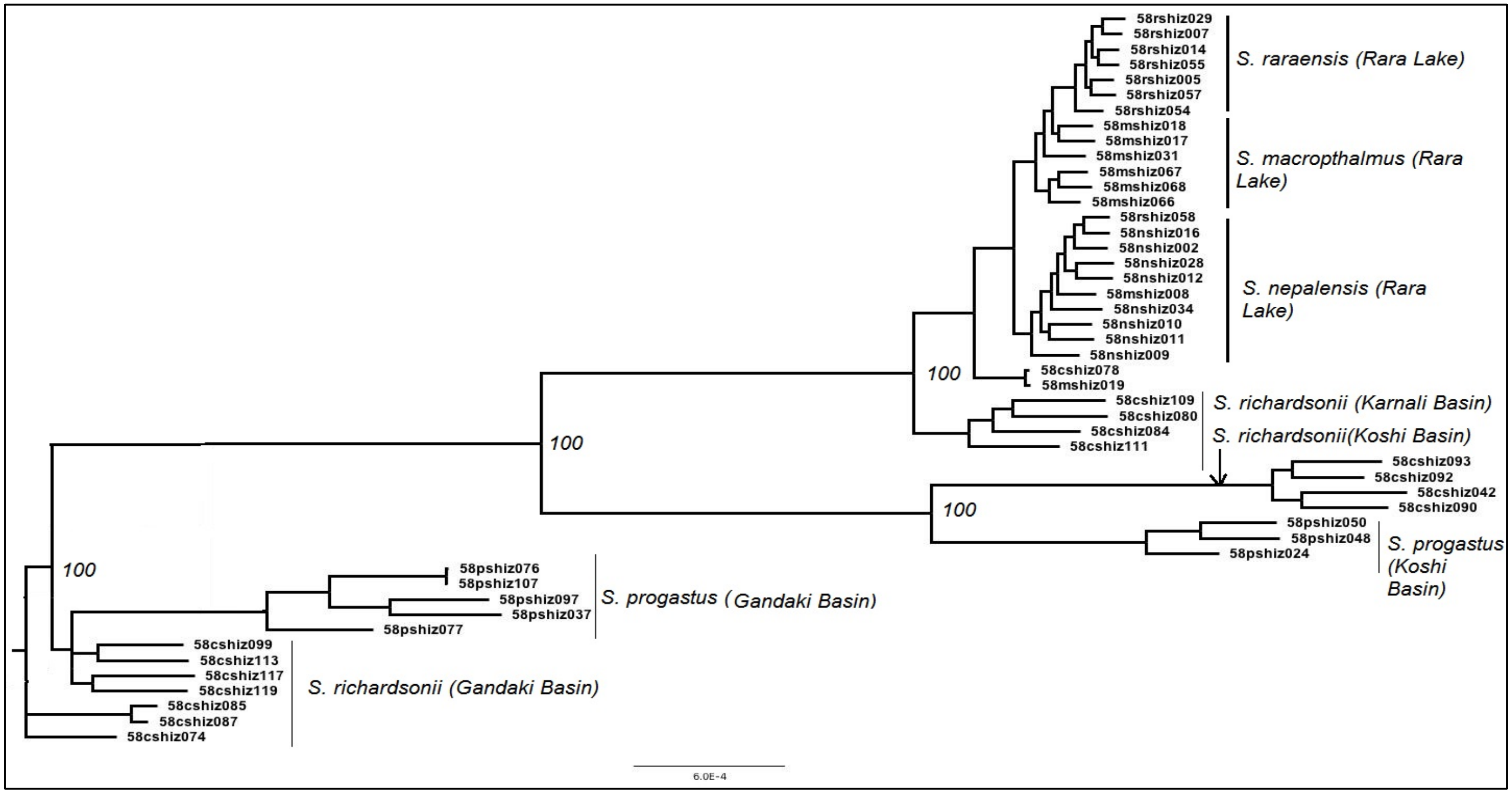

Figure 2: Phylogenetic tree constructed from 20,000 SNP loci assayed across 48 individuals of *Schizothorax* samples from three major drainages in the Central Himalayas of Nepal: Karnali, Gandaki and Koshi rivers. The tips labels represent the sample ID (University of Arkansas, Biological Science Dept., Douglas Lab) and species and basins are presented next to the major clusters. Bootstrap values are presented at important nodes. The geographic location of drainages in Nepal is depicted in Figure 1, and locality details of samples are provided in Table1.

Chapter VI

Conclusion

The current drainage pattern in the southern Himalayas evolved recently (<5mya, Brookfield, 1998). The rapidly elevating mountain range and fast flowing rivers draining into the Gangetic Flood Plain provide unparalleled habitat diversity and an ecological theater for the evolutionary play of freshwater diversification (Hutchinson, 1965). Due to their vastness and remoteness, the biologically rich and ecologically diverse Himalayan freshwater habitats are not explored and documented adequately, leaving aquatic diversity unresolved.

Snowtrout (*Schizothorax*) can serve as a model for deciphering the evolutionary process that affects speciation and contemporary distribution of species (Cao et al., 1981; Chen et al., 2017; Chi et al., 2017). Previous studies of Snowtrout in the Central Himalayas and Southeast Asia showed higher divergence of species across the basin than within (Terashima, 1984; Dimmick & Edds, 2002; He & Chen, 2006). My analysis based on a single gene (mitochondrial cytochrome b) reflected this pattern (*Chapter IV*).

Morphometric analysis of shape variation in *Schizothorax* from riverine (i.e. *S. richardsonii and S. progastus*) and lacustrine habitats (i.e. *S. macrophthalmus, S. nepalensis,* and *S. raraensis*) in Nepal demonstrated that the morphological variation can be predicted based on both species and the basin. This result leads me to hypothesize that the differential habitat conditions in the upstream and downstream part of river could have differential selection maintaining these species within the basin, but geographical and ecological barriers created by recent changes dispose them into more diverged forms from species across basins. In this scenario, the morphometric similarity of species in upstream and downstream habitat among basin likely reflects phenotypic convergence. I tested this hypothesis by generating approximately 20,000 of single nucleotide polymorphism (SNPs) loci using next generation sequencing technology (ddRAD). The species distributed in Lake Rara and the Central Himalayan rivers were clearly separated in my phylogenomic analysis, results that stand in contrast to the previously unsolved relationships based on analyses of a single mitochondrial gene. The *Schizothorax* species distributed in three river basins in the Central Himalayas are likely in an incipient stage of speciation, reflecting the recency of drainage pattern evolution in this region (Brookfield, 1998; Clark et al., 2004; Clift et al., 2001) and the geologically ‘young’ age of Lake Rara (estimated at ~60,000ya; Zakaria et al., 2018).

References:

Brookfield, M. E. (1998). The evolution of the great river systems of southern Asia during the Cenozoic India-Asia collision: Rivers draining southwards. *Geomorphology*, 22:285-312. doi: 10.1016/S0169-555X(97)00082-2.

Cao, W. X., Chen, Y. Y., Wu, Y. F., and Zhu, S. Q. (1981). Origin and evolution of Schizothoracine fishes in relation to the upheaval of the Qinghai-Xizang Plateau. In Q. F. Li (Ed.), *Studies on the Period, Amplitude and Type of the Uplift of the Qinghai-Xizang Plateau* (118-129). Beijing, China: Science Press.

Chen, W., Yue, X. and He, S. (2017) Genetic differentiation of the *Schizothorax* species complex (Cyprinidae) in the Nujiang River (upper Salween). *Scientific Reports*, 7, 5944.

Chi, W., Ma, X., Niu, J. and Zou, M. (2017). Genome-wide identification of genes probably relevant to the adaptation of schizothoracins (Teleostei: Cypriniformes) to the uplift of the Qinghai-Tibet Plateau. *BMC Genomics*, 18, 310.

Clark, M.K., Schoenbohm, L.M., Royden, L.H., Whipple, K.X., Burchfiel, B.C., Zhang, X., Tang, W., Wang, E. and Chen, L. (2004). Surface uplift, tectonics, and erosion of eastern Tibet from large-scale drainage patterns. *Tectonics*, 23(1).

Clift, P.D., Shimizu, N., Layne, G.D. and Blusztajn, J. (2001). Tracing patterns of erosion and drainage in the Paleogene Himalaya through ion probe Pb isotope analysis of detrital K-feldspars in the Indus Molasse, India. *Earth and Planetary Science Letters*, 188:475-491.

Dimmick, W.W. and Edds, D.R. (2002). Evolutionary genetics of the endemic schizorathicine (Cypriniformes: Cyprinidae) fishes of Lake Rara, Nepal. *Biochemical Systematics and Ecology*, *30*:919-929.

He, D., and Chen, Y. (2006). Biogeography and molecular phylogeny of the genus *Schizothorax* (Teleostei: Cyprinidae) in China inferred from cytochrome b sequences. *Journal of Biogeography*, 33:1448-1460. doi: 10.1111/j.1365-2699.2006.01510.x

Hutchinson, G. E. (1965). *The Ecological Theater and the Evolutionary Play*. Yale University Press, New Haven, Connecticut.

Terashima, A. (1984). Three new species of the cyprinid genus *Schizothorax* from Lake Rara, northwestern Nepal. *Japanese Journal of Ichthyology*. 31:122-135.

Zakaria G. (2018). Late Quaternary seismicity and climate in western Nepal: Himalaya. Applied geology. Unpubl. Ph.D. dissertation. Université Ghent, Department of Geology, Grenoble, France, Ghent, Belgium. (https://tel.archives-ouvertes.fr/tel-01935412/document)