Freshwater Biology

Freshwater Biology (2016) doi:10.1111/fwb.12824

Environmental and spatial characterisation of an unknown fauna using DNA sequencing – an example with Himalayan Hydropsychidae (Insecta: Trichoptera)

FELICITAS HOPPELER*, RAM DEVI TACHAMO SHAH*†, DEEP NARAYAN SHAH*†, SONJA C. JÄHNIG†‡, JONATHAN D. TONKIN†§, SUBODH SHARMA¶ AND STEFFEN U. PAULS*

*Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt, Germany

†Senckenberg Research Institute and Natural History Museum Frankfurt, Department of River Ecology and Conservation, Gelnhausen, Germany

‡Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Department of Ecosystem Research, Berlin, Germany

§Department of Integrative Biology, Oregon State University, Corvallis, OR, U.S.A.

¶Aquatic Ecology Centre (AEC), Department of Environmental Science & Engineering, School of Science, Kathmandu University, Dhulikhel, Kathmandu, Nepal

SUMMARY

1. Ecological studies of freshwater biota in understudied regions are often obstructed by poor taxonomic knowledge. We posit that molecular tools can help alleviate this issue and present an example where we combined molecular tools, environmental data and ecological statistics to investigate the distribution and community ecology of an unknown fauna of hydropsychid caddisflies along altitudinal gradients in four Himalayan river systems of Central and Eastern Nepal.
2. A total of 484 larval specimens from 34 tributaries were examined. Phylogenetic analysis of the mitochondrial cytochrome c oxidase I (COI) and the nuclear ribosomal RNA 28S were used to delineate molecular operational taxonomic units (MOTUs) applying three analytical methods: general mixed Yule-coalescent (GMYC) model, Automatic Barcode Gap Discovery (ABGD) and Bayesian Phylogenetics and Phylogeography (BPP). Spatial distributional patterns and potential differences in ecological niches among MOTUs were statistically tested using regression and correlation approaches. Further, we examined the data for signs of non-random structure in MOTU communities.
3. MOTU diversity within the family of Hydropsychidae was generally high but varied across evaluated gene fragments and slightly among delineation methods. Yet, the subsequent evaluation of environmental and spatial drivers and resulting distributional patterns were highly consistent among the different MOTU estimates.
4. Within each river system, we found community composition varied greatly along the altitudinal gradients, with many MOTUs associated with specific altitudinal ranges. Prevalent MOTU turnover at the river system scale indicated high β-diversity in the hydropsychid community leading to high degrees of regional endemism. In the Langtang river system, we found fewer MOTU co-occurrences than expected by chance.
5. These results highlight the utility of DNA-based approaches using variable genetic markers (mitochondrial or ribosomal nuclear) for primary biodiversity assessment of poorly studied groups or regions. Our study further shows that DNA-based biodiversity measures are suitable for downstream applications, such as exploring fundamental questions in stream ecology.

Keywords: altitudinal gradient, coalescent, community structure, genetic diversity, headwater tributaries

Correspondence: Steffen U. Pauls, Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, D-60325 Frankfurt, Germany. E-mail: steffen.pauls@senckenberg.de

© 2016 John Wiley & Sons Ltd

Introduction

Studying the ecology of freshwater ecosystems requires detailed knowledge about the environment and the biota of interest. In poorly studied regions, however, taxonomic information is rarely sufficient to assess the biota at species or even genus level. This lack of taxonomic resolution can inhibit ecological research. Thus, DNA is increasingly used to characterise biodiversity in such unknown faunas (e.g., Pons *et al*., 2006; Monaghan *et al*., 2009; Vuataz *et al*., 2011). While many researchers have used DNA solely to discover (or identify) new species, only a few have additionally explored fundamental ecological questions (reviewed by Valentini, Pompanon & Taberlet, 2009; Joly *et al*., 2014). These include microbial ecologists and community ecologist working with taxonomically difficult groups or cryptic species (e.g., Craft *et al*., 2010; Obertegger, Fontaneto & Flaim, 2012). To our knowledge, no attempts have been made to investigate stream biotas in remote and understudied regions in this manner. Yet, a holistic understanding of distributional patterns of freshwater species requires knowledge about freshwater ecosystems from all ecoregions of the world, stressing the need for linking molecular and ecological approaches to expedite the generation of this knowledge.

Little is known about the diversity, distribution and structure of biotic communities of rivers in the Hindu Kush-Himalaya. In light of the wide altitudinal range and varied climatic conditions present in this region, taxonomic richness and diversity are expected to be high for many aquatic groups, such as hydropsychid caddisflies, the focal organism of this study. To date, 53 species of Hydropsychidae have been recorded from Nepal (Malicky, 2006). Identification of species is entirely based on adults due to the lack of available taxonomic resources for immature larval stages (Graf, 2006). Moreover, the absence of diagnostic morphological characters for larvae, as well as the presence of cryptic species, within the family of Hydropsychidae has been reported earlier (e.g., Zhou, Kjer & Morse, 2007; Statzner & Mondy, 2009; Geraci *et al*., 2010; Pauls *et al*., 2010). We posit that timely characterisation of Hydropsychidae communities in understudied regions will benefit from using a molecular approach to recognise species within this caddisfly family.

Hydropsychidae display an interesting serial succession along the river continuum (e.g., Hildrew & Edington, 1979; Ross & Wallace, 1982; Tachet *et al*., 1992). Studies from Europe and North America suggest that this pattern is mainly driven by altitude, and more precisely temperature. Other environmental factors proposed to be of influence include the following: substrate composition and current velocity, food availability and feeding habits, competition for net-spinning sites, life-cycles, hydraulic stress, conductivity and anthropogenic impact (Hildrew & Edington, 1979; Garcıa & Ferreras-Romero, 2008; Statzner & Dolédec, 2011). The large gradients in chemical, physical and biotic parameters given in the Himalayan Mountains present an interesting opportunity to study distributional patterns of larval hydropsychids across an unprecedented altitudinal gradient, and compare them with other mountainous Hydropsychidae communities.

We pursue two main aims in this study. First, we present an approach for studying community ecology in poorly known faunas using DNA-based diversity hypotheses. Second, we exemplify this approach by testing ecological hypotheses on spatial and environmental patterns of hydropsychid diversity in Nepalese streams. To achieve these aims, we assess Hydropsychidae diversity using the mitochondrial cytochrome c oxidase I (COI) gene, which represents the most sampled protein-coding mtDNA fragment and the nuclear ribosomal RNA 28S. Both genes can effectively distinguish closely related species in Hydropsychidae (Zhou *et al*., 2007; Geraci *et al*., 2010). To specifically recognise molecular operational taxonomic units (MOTUs; i.e. independent evolutionary units or presumptive species), we employed three commonly used analytical approaches: two coalescent-based methods, General mixed Yule-coalescent model (GMYC; Pons *et al*., 2006) and Bayesian Phylogenetics and Phylogeography (BPP; Yang, 2015), and the distance-based Automatic Barcode Gap Discovery (ABGD; Puillandre *et al*., 2012a) method. The combination of multiple approaches was intended to lend confidence to proposed MOTU hypotheses (Carstens *et al*., 2013). We expected to find high levels of biodiversity based on existing knowledge on Hydropsychidae (De Moor & Ivanov, 2008) and by sampling expansive altitudinal gradients across a relatively large region in the Himalaya. Next, we linked the hypothesised MOTUs to prevailing environmental and spatial variables to explore distributional patterns and potential differences in ecological niches among evolutionary units. Specifically, we evaluated the influence of altitude (also as a proxy of temperature), river system, network centrality (i.e. a measure of site connectedness), river network distance (i.e. direct and along-stream distance between sites), and conductivity on hydropsychid communities using statistical regression and correlation methods. We expected communities to show relatively high turnover both along the altitudinal gradient and among sites as

previously observed in headwater streams (Clarke *et al.*, 2008; Finn *et al.*, 2011; Múrria *et al.*, 2013). Finally, we explored community structuring by evaluating randomness of MOTU co-occurrences among and within sites using two common methods in community ecology. These analyses evaluate if community patterns are influenced by competition (Gotelli, 2000; Sanders *et al.*, 2007) or environmental processes (Gotelli, 2000; Gotelli & Mccabe, 2002). Since competition is considered typical in hydropsychid caddisfly larvae (Hildrew & Edington, 1979; Statzner & Dolédec, 2011), we expected low numbers of co-occurring MOTUs at the site scale, and that this pattern would indicate competitive structuring.

Methods

Sampling

Headwater streams in the Himalayan regions of Central and Eastern Nepal were sampled from April to June 2012 and 2013 as part of a larger macroinvertebrate survey (Fig. 1; Tachamo Shah *et al.*, 2015; Li *et al.*, 2016). Hydropsychidae were collected from 34 remote first-and second-order tributaries of the Langtang, Yangri, Indrawati and Arun Rivers between 1400 and 3200 m a.s.l.. Environmental data were obtained on site using a GPS device (Garmin Etrex, Schaffhausen) for altitude and geographical position and a portable field meter (Multi 3430 SET F; WTW, Weilheim) for water temperature and conductivity.

Specimens originate from semi-quantitative Multi-Habitat Sampling (MHS; Moog, 2007) of 10 pooled sampling units, taken from all habitat types along a 100-m river stretch. This standardised method increases repeatability and comparability among sites (Hering *et al.*, 2004). Benthic samples were collected by kick-sampling (stationary, sampling from down- to upstream over 10 different habitat patches) using a hand net (standard 25 × 25 cm^2 frame; mesh size 500 μm). All organisms were preserved in 96% ethanol. In the laboratory,

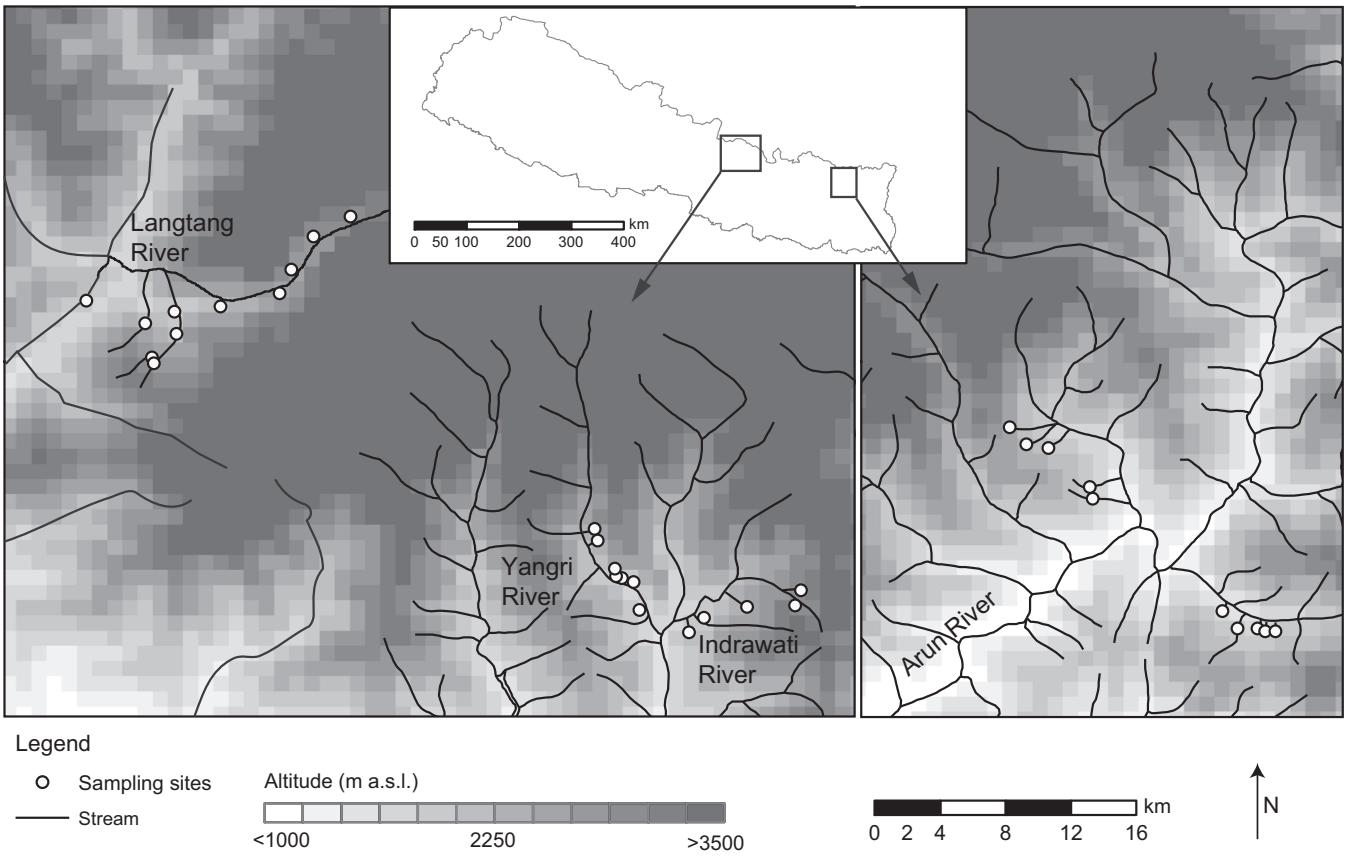


Fig. 1 Location of the 34 headwater streams sampled along the Langtang, Yangri and Indrawati river systems in Central Nepal (left) and the Arun River in Eastern Nepal (right). Number of sites and altitudinal ranges for the individual river systems are as follows: 11 sites from 1556 to 3107 m a.s.l. in Langtang; seven sites from 1485 to 2650 m a.s.l. in Yangri; five sites from 1750 to 3148 m a.s.l. in Indrawati; 11 sites from 1557 to 2307 m a.s.l. in Arun.

© 2016 John Wiley & Sons Ltd,

Hydropsychidae larvae were identified to genus following Graf (2006).

DNA extraction, amplification and sequencing

At each site, up to 20 individuals covering all collected size classes of each hydropsychid genus found were selected for DNA analysis. We sequenced nearly 50% of all available specimens (*Hydromanicus*: 35%, *Hydropsyche*: 63%, *Arctopsyche*: 76% and *Diplectrona*: 100%).

Whole genomic DNA was extracted from abdominal tissue or individual legs using Qiagen DNeasy Tissue or QIAmp DNA Micro Kits (Qiagen, Hilden) following the manufacturer's protocol. Tissue was first placed in Proteinase K and ATL lysis buffer for at least 1 h.

We analysed the mitochondrial cytochrome c oxidase I (COI) and the nuclear ribosomal RNA 28S. Due to repeated difficulties amplifying the common COI 'barcode' region (from here on referred to as COI-5P), we focussed our efforts for this study on the more variable downstream region of the COI gene (from here on referred to as COI-3P). COI-3P has been widely and successfully used in numerous studies on aquatic insects (e.g., Monaghan *et al.*, 2005; Hjalmarsson, Bergsten & Monaghan, 2014; Vitecek *et al.*, 2015). The following primer sets were used for PCR amplification and sequencing: LCO1490/HCO2198 (Folmer *et al.*, 1994) and LepF1/LepR1 (Hebert *et al.*, 2004) for COI-5P (658 bp), Jerry/S20 (Pauls, Lumbsch & Haase, 2006) for COI-3P (541 bp), and D1-3up1/ D3-TRIC-DN (K. Kjer, unpubl. data) and D2-UP-4/ D2DN-B (Zhou *et al.*, 2007) for 28S (D1–D3 regions; 1627 bp).

PCR mixes for COI-5P and COI-3P contained 1 μL of genomic DNA, 1 μL 10x Taq S-Buffer (Peqlab, Erlangen), 1 μL dNTPs (2 μM each), 0.25 μL of each primer (10 μmol), 0.2 μL Hot Taq DNA Polymerase (Peqlab) and 6.3 μL of sterile H_2O in 10 μL. COI-5P PCR amplification included 35 cycles of 95 °C for 45 s, 50–57 °C (Zhou *et al.*, 2007) for 30 s and 72 °C for 60 s. COI-3P PCR amplification included 35 cycles of 95 °C for 60 s, 41 °C for 60 s, and 72 °C for 120 s. PCR mixes for the 28S D1–D3 regions contained 1 μL of DNA template, 1 μL 10x Taq Y-Buffer, 1 μL dNTPs (2 mM each), 0.5 μL DMSO, 0.2 μL BSA, 0.2 μL of each primer (10 μmol), 0.2 μL Hot Taq DNA Polymerase (Peqlab) and 5.7 μL of sterile H_2O. 28S PCR amplification included 35 cycles of 95 °C for 45 s, 60 °C for 45 s and 72 °C for 60 s.

PCR products were visualised on 1.5% agarose gels and purified in case of the 28S (Qiagen Gel Extraction; Qiagen). PCR products were sequenced in both directions on an ABI 3730XL sequencer (Applied Biosystems, Waltham) at the Biodiversity and Climate Research Centre, Frankfurt.

Molecular data analyses

Sequence alignment. All sequences were assembled and edited in Geneious 6.1.6 (Drummond *et al.*, 2010). Mitochondrial sequences were aligned with default settings using the Geneious alignment algorithm; the open reading frame was then verified. 28S sequences were pre-aligned using MAFFT within Geneious (Katoh *et al.*, 2002) and manually adjusted using the universally conserved RNA secondary structure (Kjer, Roshan & Gillespie, 2009). We included only sequences with <10% missing data in the analyses. Alignments are available in MetaCat (Senckenberg data repository: http://dataportal-senckenberg.de/database/). Sequence information and detailed specimen records are available from Barcode of Life Database (SPHYD project; BOLD, http://www.barcodinglife.org).

To examine the utility of genetic data to delineate hydropsychid MOTUs, we constructed six data sets: COI-3P, COI-5P and 28S with (i) all individual sequences successfully amplified per gene fragment and (ii) only specimens for which all three gene fragments were available.

Tree estimation. As a prerequisite for GMYC and BPP, we performed Bayesian phylogenetic reconstructions. We favoured using individual gene trees over a multilocus phylogeny due to the limited overlap between all three data sets at the level of individuals. Sequences were collapsed into unique haplotypes using a custom Perl script (Chesters, 2013). The best model of sequence evolution (Table S1) for each data set and partition (COI: first, second and third position) was estimated in MEGA 5.2 (Tamura *et al.*, 2011) and selected based on the Akaike information criterion (AIC). Ultrametric single gene trees were calculated in BEAST v.2.3.0 (Bouckaert *et al.*, 2014) running three independent analyses per data set for 100 million generations. We unlinked substitution rates among partitions, linked clock models (lognormal uncorrelated clock, coalescent prior) and topology, and used default settings for all priors. We plotted log-likelihood scores in Tracer v.1.6 (Rambaut & Drummond, 2007) to determine stable convergence of model parameters. TreeAnnotator v.1.7.1. was used to generate consensus trees (maximum clade credibility; 25% burn-in). We combined independent runs in LogCombiner v.1.7.1 (25% burn-in) and visualised consensus trees in FigTree v.1.4.0.

MOTU delineation. To formulate a primary MOTU hypothesis, we employed the coalescent-based GMYC and the sequence-based ABGD methods. They were chosen because they do not require any *a priori* estimate on the number of expected units. Because of their fundamentally different statistical approach comparing the results of both methods can increase or limit confidence in those results (e.g., Ratnasingham & Hebert, 2013). Based on simulations and empirical data, GMYC is considered a robust delimitation tool (Fujisawa & Barraclough, 2013; Schwarzfeld & Sperling, 2015), but some studies indicate that GMYC has problems with over splitting evolutionary lineages in cases where fewer than 20% of demes are sampled (Lohse, 2009; Papadopoulou *et al.*, 2009). Thus, we used the Bayesian species delineation approach BPP to validate the MOTU hypothesis resulting from GMYC and ABGD analyses.

Overall, we generated six MOTU hypotheses with different loci and delineation methods, i.e. COI-3P, COI-5P and 28S for both GMYC and ABGD (for details see below). We assessed congruence among MOTU hypotheses and calculated a concordance score for each MOTU. The score shows how often a specific MOTU was recovered and ranges from one to six (one indicates that a MOTU was only found once; six indicates that a MOTU was recovered in all analyses).

General mixed Yule-coalescent model. General mixed Yule coalescent estimates species boundaries in a maximum likelihood framework by identifying the transition point between population-level (coalescent) and species-level (diversification, i.e., speciation and extinction) processes (Pons *et al.*, 2006). The transition is estimated based on the branching pattern in an ultrametric tree. Samples are divided into *n* independently evolving units (MOTUs) with confidence intervals given for each MOTU. We applied the single-threshold GMYC model that assumes equal transition from coalescent to speciation for all MOTUs (Monaghan *et al.*, 2009) to the ultrametric trees for each individual gene data set (see Sequence alignment). GMYC analyses were performed in R Statistical Software v.2.15.3 (R Core Team, 2012) using the splits package (Species Limits by Threshold Statistics) (Ezard, Fujisawa & Barraclough, 2009).

Automatic Barcode Gap Discovery and Barcode Index Number. The Automatic Barcode Gap Discovery uses a clustering algorithm to distinguish partitions in genetic distance among individuals, to find a 'barcode gap' (Puillandre *et al.*, 2012a), where intraspecific variation is less than and discernible from interspecific variation. Initial partitioning is recursively evaluated until all partitions possess unimodal distributions. The analysis was performed online (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html) using default parameters (range of *a priori* intraspecific divergence = 0.01–0.1; steps = 10; relative gap width = 1.5; Nb bins for distance distribution = 20; Kimura 2-parameter distances). A similar application to delineate species based on genetic pairwise distances is implemented at the BOLD web interface. COI barcode sequences (COI-5P) were automatically grouped into Barcode Index Numbers (BINs, analogous to MOTUs), and later compared to our ABGD and GMYC hypotheses.

Multi-locus delineation using BPP. The Bayesian programme BPP analyses molecular sequence alignments under the multispecies coalescent model. A reversible-jump Markov Chain Monte Carlo (rjMCMC) algorithm moves between different species delimitation models by collapsing and resolving nodes throughout a pre-specified guide tree, sampling a posterior distribution of speciation probabilities for each split (Rannala & Yang, 2013). Importantly, the model evaluates if merging of pre-defined MOTUs improves the model (i.e. leads to a better MOTU hypothesis). We chose the highest MOTU count for validation, using both loci including identical specimens for COI (both fragments) and 28S. Two replicates that varied by rjMCMC algorithm (0 or 1; Rannala & Yang, 2013) were run with different values for the gamma shape parameters α and β of effective population sizes (θ) and species divergence times (τ) (Giarla, Voss & Jansa, 2014). Prior '3' that assigns equal probabilities for MOTUs was chosen because of its suitability for a large number of species/populations.

Sampling effort. We wanted to assess if our sampling was adequate to better evaluate the statistical results of subsequently applied environmental and spatial analyses. We calculated site-based and individual-based richness estimators and confidence intervals in EstimateS (Colwell *et al.*, 2012) to assess sampling completeness regarding MOTUs. Resampling curves reaching saturation were assumed to indicate that sampling sites, and individuals chosen for sequencing were comprehensively sampled and representative of the present hydropsychid communities.

Multivariate environmental and spatial analyses

Regression methods were used to analyse relationships between MOTU occurrences and the environment.

Specifically, we applied generalised linear models (GLMs), multivariate regression trees (MRTs) and Mantel tests to determine the main environmental and spatial drivers of MOTU distribution. Explanatory variables used in both GLM and MRT analyses were altitude, river system, network centrality, and conductivity (Table 1). Altitude and conductivity were assumed to reflect environmental factors at the local scale, whereas river system and network centrality reflected broader spatial characteristics. Moreover, altitude was used as a proxy for temperature, supported by a negative correlation between point measurements of temperature and altitude (pairwise Pearson correlation test: $r^2 = -0.753$, $P < 0.0001$). Network centrality was defined as the mean river network distance of a site to all other sites within a respective river system (calculation after Carrara *et al.*, 2012). We ran multivariate GLMs with the R package mvabund (Wang *et al.*, 2012) on both presence/absence and abundance data. We fit separate models to each MOTU (manyglm function) using the negative binomial distribution (best fitting family, uses a log-link function). ANOVA was applied to test for significant effects of explanatory variables (anova.manyglm function; likelihood ratio test, 1000 resamples). Pairwise Pearson correlation tests were performed between explanatory variables to check for collinearity.

We calculated MRTs using the R package mvpart (Therneau & Atkinson, 2013). Sites with similar species assemblages are clustered. Then, each cluster is defined by a set of environmental values that allow differentiation from the sister cluster. The dependence of clusters on associated habitats is graphically represented by a tree. Selection of the tree followed the method described by De'ath & Fabricius (2000), where the most complex tree within one standard error of the best tree (based on cross validation) was selected.

Mantel tests (based on Pearson's product–moment correlation) were performed by correlating community dissimilarity (Bray–Curtis and Soerensen indices for abundance-based and presence/absence-based measures) with three environment/space distance metrics: altitudinal, direct-line and river network distance (Table 1). Distances were calculated for each river system individually. Altitudinal distance was the pairwise difference in altitude between sites. Direct-lines were calculated using Euclidean distances between site coordinates. River network distance was manually assessed using Google Earth following the river network between sites (because GIS layers were lacking for these first- and second-order streams). We performed tests with the R package ecodist (Goslee & Urban, 2007) using the mantel function (1000 permutations). Spatial distance matrices were additionally tested to account for auto-correlation.

Community structure analyses

Examining co-occurrence patterns among habitats provide information on the abiotic or biotic processes determining species assemblages. Checkerboard scores (Stone & Roberts, 1990) and the Mean Nearest Taxon Distance (MNTD; Webb *et al.*, 2002) algorithm were used to test if MOTU co-occurrences among and within sites were non-random using an MOTU by site presence/absence matrix. The checkerboard score statistic (*C*-score) tests for co-occurrences in the data matrix, referred to as checkerboard patterns (Gotelli, 2000). Observed *C*-scores that are greater than expected by chance, with greater effect size (*z*-score), indicate community segregation (i.e., there are fewer co-occurrences than expected). Observed *C*-scores less than expected by chance (negative *z*-score) indicate community aggregation (i.e. there is greater co-occurrence than expected). The *C*-score was assessed using the 'nestedchecker' function and examined in relation to a null model using the 'oecosimu' function, both implemented in the R package vegan (Oksanen *et al.*, 2013). The 'quasiswap' null model algorithm of Miklós & Podani (2004) was employed where MOTU occurrences and sites are initially fixed and two by two matrices are swapped based on the 'quasiswap' algorithm (burnin = 50 000; nsimul=10 000). *C*-score metrics were calculated both among overall community samples and within individual river systems.

Table 1 Environmental and spatial parameters used in statistical regression analyses given as range for each river system.

River system	Altitude (m a.s.l.)	Network centrality	Conductivity ($\mu S\ cm^{-1}$)	Altitudinal distance (m)	River network distance (km)	Euclidean distance (km)
Langtang	1556–3107	0.121–0.330	23–134	14–1551	0.550–15.924	0.106–12.772
Yangri	1485–2650	0.155–0.495	31–54	78–1165	0.493–6.667	0.248–5.203
Arun	1557–2307	0.089–0.359	10–78	7–593	0.614–25.330	0.216–18.830
Indrawati	1745–3148	0.097–0.145	21–38	127–1387	2.152–12.022	0.385–7.183

Within sites, observed phylogenetic relatedness was compared to the pattern expected under a null model of phylogeny and community randomisation by calculating the MNTD. Phylogenetic overdispersion (co-existence of distantly related MOTUs) is indicated by positive and underdispersion (co-existence of closely related MOTUs) by negative effect size. Analysis of phylogenetic over- and underdispersion was performed in R using the package picante (function ses.mntd; Kembel *et al.*, 2010), evaluating relatedness across all MOTUs and within individual genera. The analysis requires an inference of phylogenetic relationships among MOTUs. We used *BEAST in BEAST v.2.3.0 (Bouckaert *et al.*, 2014) to calculate a phylogenetic species tree based on all specimens available for COI-3P and 28S (173 individuals plus two outgroup taxa). Sequences were assigned to putative species based on the COI-3P MOTU count as our best MOTU hypothesis (see Results). Nucleotide substitution rates were estimated as described earlier (Table S1) and unlinked across the two loci and mtDNA partitions. We ran two independent analyses for one billion generations under a relaxed lognormal uncorrelated clock (linked) using a Yule species tree prior (default settings for all priors). Stable convergence of parameters and congruence of independent runs was assessed as described above.

Results

MOTU delineation

Overall, the sequencing success varied greatly among gene fragments. We analysed 484 individuals for a total of 471 COI-3P, 128 COI-5P and 173 28S sequences (84 individuals sequenced successfully for all three gene fragments; Table 2).

Primary species delimitation with GMYC and ABGD resulted in very similar (±1–2) or identical MOTU counts within loci (Table 2), whereby fewer MOTUs were found with the nuclear marker. Congruence between the mitochondrial and nuclear marker, i.e. the placement of individuals into MOTUs, was nearly perfect: mitochondrial MOTUs were generally clumped together in 28S and only one single individual was differently placed. Within single fragments, GMYC and ABGD were highly congruent, i.e. clumped MOTUs but never assigned specimens to different MOUTs. Identical results were obtained with the ABGD and BIN algorithm (BIN is based on COI-5P only). Delineation results and concordance scores for the MOTUs derived from COI-3P are summarised in Fig. 2. For more details on

Table 2 Mitochondrial and nuclear sequence variation and results of ABGD and GMYC species delimitation methods. *N*, total number of sequences; bp, aligned matrix length; K_n, number of haplotypes; Pi, parsimony informative sites (%); (N_{ABGD}), the number of MOTUs recovered with ABGD method; *P*, limit to intraspecific divergence (barcode gap); *T*, threshold genetic distance from the branch tips where transition occurred; (N_{GMYC}), the number of MOTUs recovered with the GMYC method; CI, confidence intervals. Likelihoods are presented for null (L_0) and GMYC (L_{GMYC}) models. Significance of the likelihood ratio (LR) was evaluated using a chi-square test to compare GMYC and null models (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

				Variable		ABGD			GMYC					
Data set	*N*	bp	K_n	sites (%)	Pi (%)	N_{ABGD}	*P*	Singletons	*T*	N_{GMYC}	CI	L_0	L_{GMYC}	LR	Singletons
COI-3P$_{84ind}$	84	541	36	40.67	38.82	14	0.036	4	−0.0098	17	16–18	145.591	165.209	39.235***	8
COI-5P$_{84ind}$	84	685	44	41.61	38.25	15	0.008	7	−0.0146	15	15–17	189.778	213.139	46.724***	7
28S$_{84ind}$	84	1627	48	14.44	11.99	6	0.022	1	−0.0039	12	3–16	346.111	351.702	11.181**	3
COI-3P	471	541	85	43.25	41.59	28	0.022	6	−0.0152	28	28–32	491.137	511.775	41.274***	6
COI-5P	128	685	50	42.63	40.88	16	0.008	5	−0.0097	17	16–18	260.679	284.368	47.377***	4
28S	173	1627	81	18.75	14.94	13	0.005	2	−0.0087	11	8–36	636.541	641.742	10.401**	1

GMYC model fits and single gene trees, see Figs S1–S4. Within the reduced data sets (84 specimens), the highest recovered MOTU count was 17 (COI-3P, GMYC method; Table 2). Validation with BPP supported this mitochondrial-based MOTU hypothesis; posterior probability for 18 units (17 MOTUs and one outgroup taxa) was highest in the majority of tested models using different parameter values (Table S2). Based on this result, and the fact that we covered the most specimens with the mitochondrial COI-3P marker, we were confident to consider the COI-3P-based MOTU count of 28 as our best estimate of diversity. We will, thus, illustrate distributional patterns on the basis of this MOTU hypothesis (see results below). Overall, *Hydromanicus* (11 MOTUs) and *Hydropsyche* (10) were found to be the most diverse genera followed by *Arctopsyche* (four) and *Diplectrona* (three). Resampling curves estimating individual-based MOTU richness reached saturation (Fig. S5a), but site-based estimates did not show a stable plateau (Fig. S5b).

Multivariate environmental and spatial analyses

Regression methods were applied to all community data sets ($N = 10$) resulting from the analysis of different gene fragments and delineation methods. Test statistics were very similar across data sets despite variation in MOTU communities. Significant effects of altitude and river system were found in all evaluated data sets, and network centrality was additionally important in most cases (GLM; Tables S3 & S4 for coefficients). Explanatory variables were not significantly correlated (Table S5).

The best predictive MRT models had little explanatory power in distinguishing hydropsychid communities in general. To improve ecological relevance, we chose the most complex trees within one standard error of the best predictive trees. Resulting trees varied in the number of leaves (2–6), but primary splits were always explained by river system or altitude (Table S6).

Mantel tests revealed contrasting patterns in the most important distance metrics linked with communities within individual river systems (Table S7). In the Langtang river system, altitudinal distance was significantly correlated with community dissimilarity in all data sets. None of the distances were important in Yangri. In the Indrawati and Arun river systems, significant distances were altitude and/or river network. However, the strength of correlations varied greatly across data sets. Significant auto-correlation of distance matrices within these two river systems was likely influencing these heterogeneous results (Table S8).

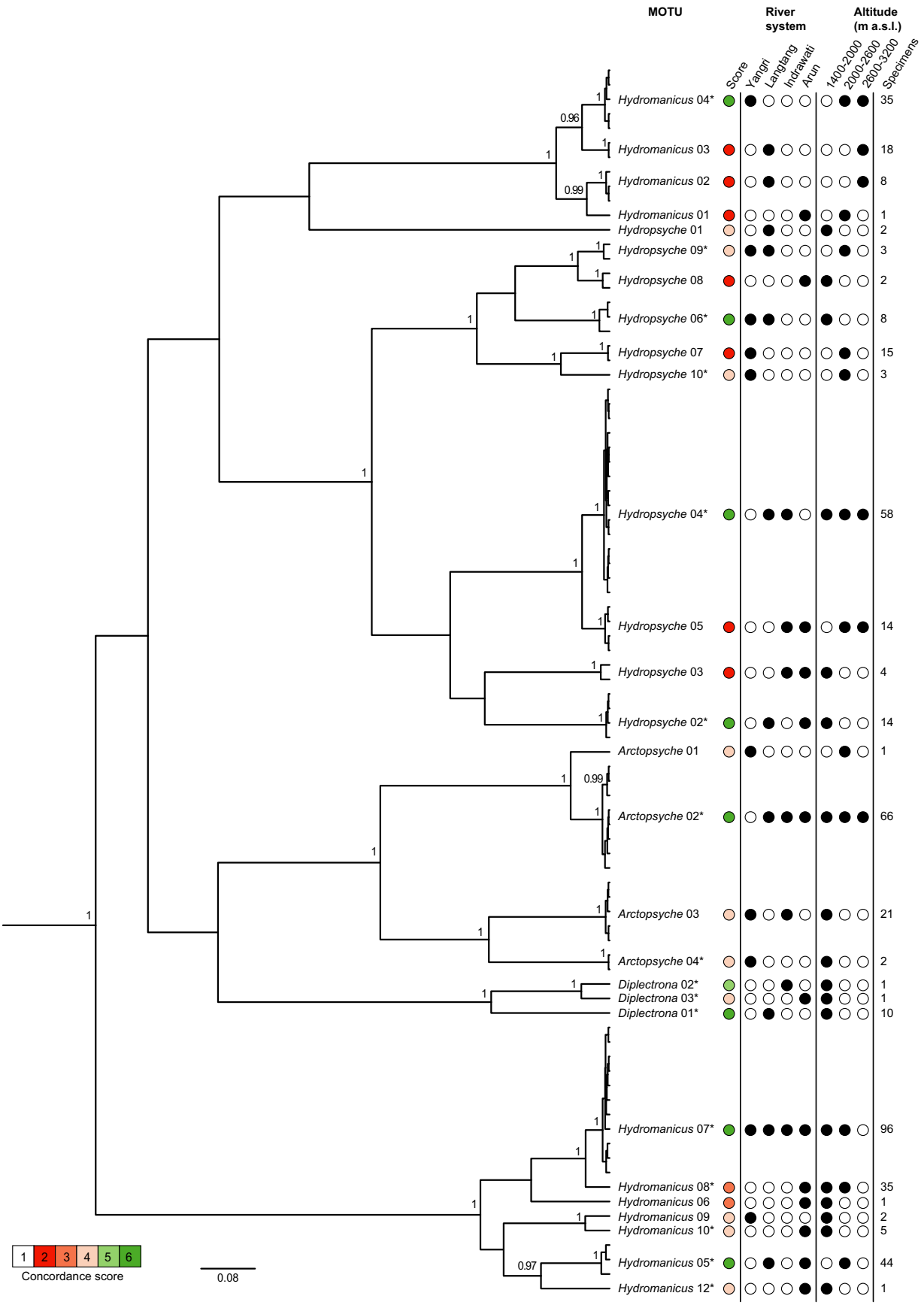
Distributional patterns were further illustrated using our best estimate of diversity (COI-3P: 28 MOTUs) to visualise the importance of statistically significant variables (Figs 2, 3, S6 & S7). This choice was considered well justified based on the coherent ecological patterns revealed by environmental and spatial analyses. Generally, community composition varied along the altitudinal gradient, with many MOTUs displaying restricted altitudinal ranges (Figs 2 & 3). Numbers of regionally overlapping MOTUs were relatively low across all data sets. That is, high levels of MOTUs (up to 70%) are restricted to certain river systems (Figs 3 & S6; see Fig. S7 for community composition in relation to river network position). Altitudinal range and regional overlap did not seem to be a function of sample size, i.e. the more abundant MOTUs did not tend to be the more widely spread (Fig. 2).

Community structure analyses

Overall, sampling sites were characterised by relatively low levels of co-occurring MOTUs. We used our best estimate of MOTU diversity (COI-3P; 24 MOTUs within the 173 individuals successfully sequenced for both COI-3P and 28S) to statistically evaluate community structuring. Checkerboard analysis showed non-random structuring at the within-catchment scale for Langtang river system (Table 3). Greater observed *C*-scores than simulated indicated less co-occurrence than expected by chance (MOTU segregation). Co-occurrence was non-significant in the remaining individual river systems as well as in the combined analysis of all river systems.

Taking phylogenetic relatedness into account, co-existing MOTUs showed a mixed picture of over- and underdispersion (positive and negative effect sizes). However, only in the Indrawati river system were co-occurring MOTUs significantly more closely related than expected by chance (Table S9). Analysing only

Fig. 2 Bayesian maximum clade credibility tree of 28 Hydropsychidae MOTUs (85 haplotypes) recovered with the mitochondrial COI-3P marker. Each MOTU is coded with a concordance score, i.e. the number of markers (COI-3P, COI-5P, 28S) and delineation methods (GMYC, ABGD) supporting a specific MOTU, ranging from one to six. MOTUs without missing data (included in all data sets) are marked with an asterisk. River system and altitudinal range is indicated with filled circles presenting presence and empty circles absence; number of specimens per MOTU is indicated in the last column. Values on branches indicate posterior $P \geq 0.95$.


MOTU
River system
Altitude (m a.s.l.)
Score
Yangri
Langtang
Indrawati
Arun
1400-2000
2000-2600
2600-3200
Specimens
Hydromanicus 04* 35
Hydromanicus 03 18
Hydromanicus 02 8
Hydromanicus 01 1
Hydropsyche 01 2
Hydropsyche 09* 3
Hydropsyche 08 2
Hydropsyche 06* 8
Hydropsyche 07 15
Hydropsyche 10* 3
Hydropsyche 04* 58
Hydropsyche 05 14
Hydropsyche 03 4
Hydropsyche 02* 14
Arctopsyche 01 1
Arctopsyche 02* 66
Arctopsyche 03 21
Arctopsyche 04* 2
Diplectrona 02* 1
Diplectrona 03* 1
Diplectrona 01* 10
Hydromanicus 07* 96
Hydromanicus 08* 35
Hydromanicus 06 1
Hydromanicus 09 2
Hydromanicus 10* 5
Hydromanicus 05* 44
Hydromanicus 12* 1
1 2 3 4 5 6
Concordance score
0.08

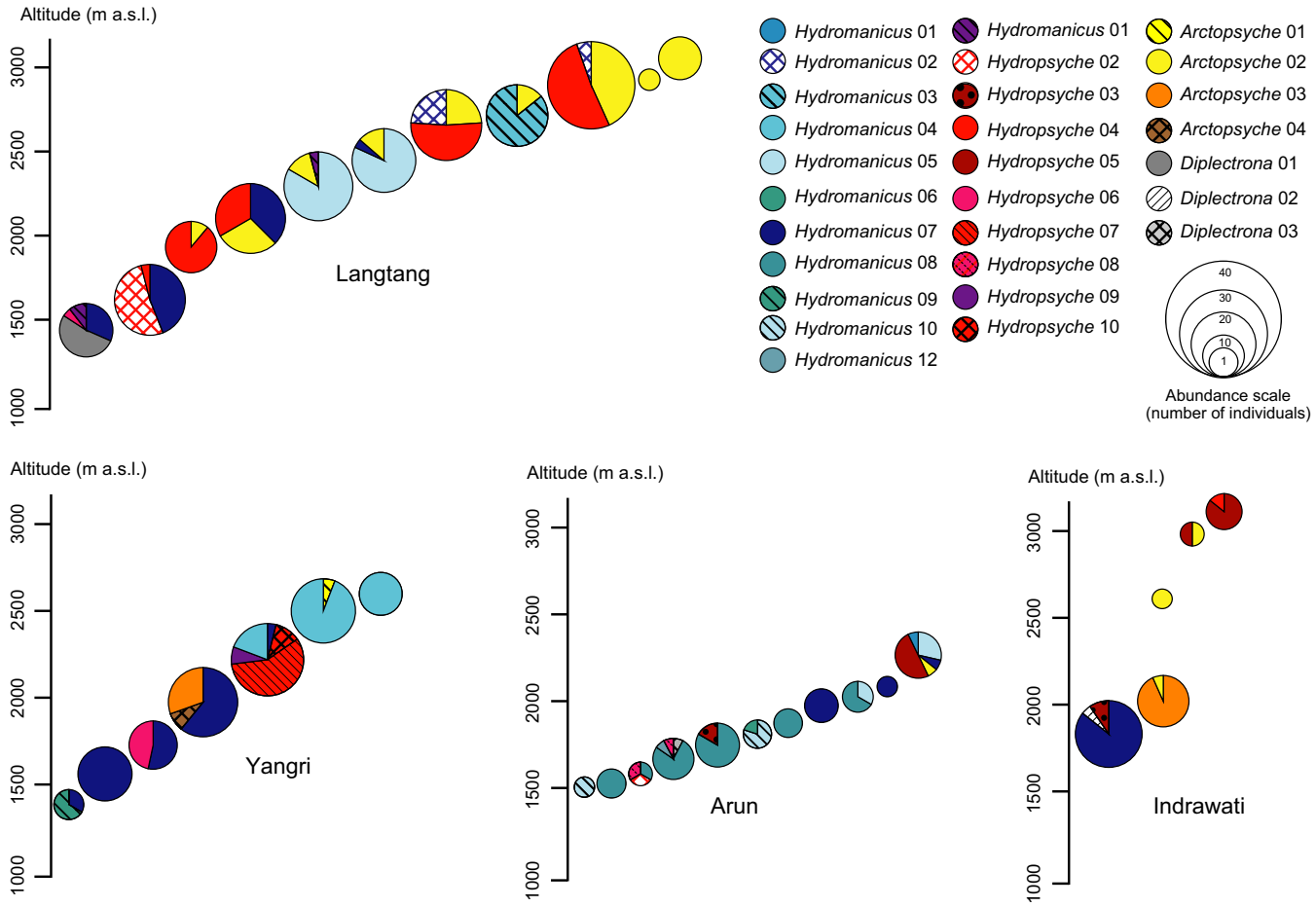


Fig. 3 MOTU composition changes among 34 Hydropsychidae communities along the altitudinal gradient in the Langtang, Yangri, Arun and Indrawati river system based on the COI-3P data set (28 MOTUs composed of 471 individuals).

Table 3 The null model analysis of MOTU co-occurrences recovered in the COI-3P data set. Positive and negative effect size indicates species segregation and aggregation, respectively. Checkerboard units, observed and mean *C*-scores, *P*-values and effect size are shown. Results are based on 10 000 permutations. Significant *P*-values are indicated using bold font.

Data set	Checkerboard units	Observed *C*-score	Mean *C*-score	*P*-value	Effect size
Overall	1750	6.341	6.109	0.063	2.040
Langtang	113	0.409	0.319	**<0.001**	4.595
Yangri	37	0.134	0.144	0.870	−0.688
Arun	143	0.518	0.486	0.235	1.337
Indrawati	29	0.105	0.095	0.093	2.144

MOTUs from genus *Hydropsyche* revealed the same signal. The analogous analysis based on genus *Hydromanicus* MOTUs showed non-significant structuring at all sites. The phylogenetic species tree could not resolve basal relationships between genera (Fig. S8). However, tip nodes within clusters and nodes immediately separating clusters were mostly well supported. Effective sample sizes (ESS) >100 were obtained for likelihoods and continuous parameters relevant to the topology in the combined traces of the two independent MCMC chains.

Discussion

Estimation of MOTU diversity

We used molecular sequence data to describe the diversity of an unknown fauna of larval hydropsychids in Nepalese headwater tributaries. We expected to find high levels of diversity given the large altitudinal gradients covered with our sampling and the known species richness within Hydropsychidae in general (De Moor & Ivanov, 2008). Our findings supported this hypothesis indicating the presence of a diverse community. Moreover, site-based richness estimation suggested the existence of unsampled diversity within sites. Many MOTUs

were found to be rare, a common feature of stream communities when assessed at high taxonomic resolution, such as species level (Woodward *et al.*, 2002; Wagner *et al.*, 2011; Múrria *et al.*, 2015). For the purpose of MOTU delineation, however, rare taxa could increase artificial splitting of evolutionary lineages due to insufficient taxonomical and geographical sampling (Lohse, 2009; Papadopoulou *et al.*, 2009).

Generally, species boundaries defined based on phylogeny and morphology have been shown to correlate relatively well in various taxonomic groups (including Trichoptera: Zhou *et al.*, 2007; Pauls *et al.*, 2010; Salokannnel, Rantala & Wahlberg, 2010; Statzner *et al.*, 2010; Previšić *et al.*, 2014). Here, we used a species concept, where closer relationships between specimens belong to the same species/MOTUs than to specimens from other species/MOTUs. This concept was tested using both mitochondrial and nuclear markers with different levels of sequence variation. As expected, this led to differing levels of MOTU resolution (Ahrens, Monaghan & Vogler, 2007; Monaghan *et al.*, 2009). However, both markers were highly congruent, with only a single specimen placed into different MOTUs by the various different analyses. Thus, we conclude both mitochondrial and the nuclear loci to be suitable for establishing meaningful MOTU boundaries in Hydropsychidae in line with findings by Zhou *et al.* (2007). It should be noted, however, that MOTU confidence intervals based on 28S were wide. This suggested the transition from speciation to coalescent events to be rather indistinct, potentially resulting from recent and rapid divergence relative to the rate of evolution in this genetic marker (Reid & Carstens, 2012).

The combination of multiple delineation approaches provided confidence in proposed MOTU estimates. Within single loci, we found perfect congruence between the coalescent-based GMYC and the ABGD method, and, in line with other studies, only slight differences in levels of diversity (e.g., Puillandre *et al.*, 2012b; Hendrixson *et al.*, 2013; Schwarzfeld & Sperling, 2015). The Bayesian method BPP further supported our best MOTU hypothesis but was unable to rule out uncertainty as the model entailed specification of a pre-defined species tree (Satler, Carstens & Hedin, 2013). Now, a logical next step would be to assess the congruence of the recovered MOTUs with existing species hypotheses based on the larval-adult associations using integrative taxonomy (Zhou *et al.*, 2007, 2016; Pauls *et al.*, 2010). Unfortunately, a sufficient collection of adult male specimens is currently not available to us.

Environmental and spatial drivers of distribution

We explored environmental and spatial drivers of distribution among different MOTU hypotheses. Perhaps most importantly, we found ecological results to be highly consistent across evaluated markers and delineation methods. The fact that differences in COI and 28S communities did not translate to different ecological patterns is highly interesting and somewhat unexpected. It is conceivable that the size and resolution of the environmental data might play a role, i.e. larger, fine-scale data records could be more sensitive to species resolution. However, the congruence lends confidence to our interpretation of distributional patterns and ecological niches and shows that ecological analysis of MOTU-based communities is robust independent of the gene fragment(s) or delimitation methods used, when dealing with the available data resolution in remote areas.

We initially hypothesised MOTU communities would show relatively high turnover along altitudinal gradients given the large elevation range sampled. Note that our sampling focused on first- and second-order tributaries over a wide range of elevations and does not cover samples along the river continuum. Altitude was found to be strongly linked to MOTU distribution in the entire study region and specifically so in the Langtang river system. These findings highlight another key structuring gradient for hydropsychids (altitudinal structure) in addition to the well-studied successional gradient of hydropsychids along a river course (e.g., Ross & Wallace, 1982), potentially reflecting adaptation to different temperature regimes (Hildrew & Edington, 1979). Temperature has generally been viewed as an important factor shaping distribution of aquatic invertebrates by influencing physiological responses (Vannote & Sweeney, 1980). In addition, dissolved oxygen, current velocity or other altitude-dependent parameters, for which we were unable to collect enough data, were previously observed to drive successional patterns in hydropsychids (e.g., Hildrew & Edington, 1979; Tachet *et al.*, 1992). Besides environmental characteristics, we included the spatial parameters river network distance and network centrality in our analyses to describe habitat connectivity, given their importance in shaping stream community composition (e.g., Altermatt *et al.*, 2013). Within river systems, we observed close sites to be faunistically similar and probably connected by dispersal. However, this pattern was not consistent across statistical analyses or data sets, complicating conclusive statements on the level of influence that network position has in different headwater streams in Nepal.

At the whole basin scale, we expected headwater communities to show high turnover among sites. This reflects the findings of Finn *et al.* (2011), who compared population-genetic and whole-community diversity between headwaters and downstream sections, reporting higher levels of ß-diversity in upstream reaches. They concluded headwaters to contribute strongly to basin-wide biodiversity through high species turnover attributed to habitat heterogeneity and landscape barriers (Brown & Swan, 2010; Múrria *et al.*, 2013). We found high turnover among sampled Himalayan streams leading to high levels of endemism. In comparison to the study by Múrria *et al.* (2013) on Mediterranean *Hydropsyche*, we observed turnover at much smaller geographical scales. Even close-by river systems like Langtang, Indrawati and Yangri feature unique, possibly endemic MOTUs. This could potentially result from the different topographical situation in the two study regions, where the higher mountains in the Himalayas may restrict migration more strongly than the mountains of Europe. However, regional restriction was not always the case: *Hydromanicus 07* was found across the entire study region showing no signs of regional or altitudinal structuring. Such widespread (e.g., Graf *et al.*, 2008) and genetically uniform (Lehrian, Pauls & Haase, 2009) montane hydropsychid taxa are present in other regions, such as central Europe and the Alps.

Drivers of community structure

Within our hydropsychid communities, we expected to find low levels of MOTU co-existence with non-random structures mainly driven by competition (Hildrew & Edington, 1979; Statzner & Dolédec, 2011). We found evidence for a segregated pattern of co-occurrence in the Langtang river system that is consistent with either competitively (Gotelli, 2000; Sanders *et al.*, 2007) or environmentally (Gotelli, 2000; Gotelli & Mccabe, 2002) structured communities. Both mechanisms are conceivable, but difficult to disentangle. At the level of individual sites, hydropsychids are expected to occupy different micro-niches (e.g. different net-spinning sites with varying flow conditions) to ensure stable co-existence (Hildrew & Edington, 1979; Statzner & Dolédec, 2011). To assess the influence of competition and habitat filtering at this scale, we evaluated phylogenetic relatedness of MOTUs. The species tree was unable to resolve all relevant nodes so that findings need to be interpreted with caution, but the analysis revealed random phylogenetic structure in the majority of cases. Interestingly, the majority of sites were occupied by MOTUs from different genera that are likely to have distinct habitat preferences. However, we found no strong statistical evidence of competition in structuring Nepalese communities of Hydropsychidae, contrary to our expectations.

Ultimately, the goal of our study was to present an example that unknown faunas can be characterised ecologically based on a DNA sequence-derived assessment of diversity and community composition. We posit that this approach offers a better approximation of diversity at the ecologically relevant taxonomic level of species (or an approximation thereof), than working at family or genus level. DNA sequencing, particularly meta-barcoding, has the potential to readily generate large data sets at these lowest taxonomic levels, which are a prerequisite for asking many ecological questions. Given their reliance on detailed biodiversity data, researchers in isolated regions with poorly understood taxonomy are faced with great challenges when addressing ecological questions. A practical approach for studying the ecology of understudied regions is needed and could include DNA sequence-derived community data. The robustness of the approach presented here was emphasised by high concordance of environmental and spatial drivers of communities when applying different marker systems and species recognition methods. We do not see this as a reason for bypassing integrative taxonomy to progress in ecological research; on the contrary, we strongly promote 'mutualism of ecologists and taxonomists' (Sheldon, 2016) to further our understanding of key ecological and evolutionary processes in hitherto understudied ecosystems.

Acknowledgments

We thank G. Regmi, K. Khatiwada, M. Prajapati, B. Tamang, D. Tamang, R. Lama, K. Nayaju, P. Sherpa, R.K. Rai, K. Tamang and T.K. Tamang for their help in the field. We thank Miklos Bálint for fruitful discussions on the study. The project was funded by the Federal Ministry of Education and Research – International Postgraduate Studies in Water Technologies (IPS11/36P) and the research funding programme 'LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-Ökonomischer Exzellenz' of Hesse's Ministry of Higher Education, Research, and the Arts. Field work was also supported by DFG grant PA1617/2-1 to SUP and SCJ. We thank four anonymous reviewers, Brian Golding, Tim Vines, Michael Monaghan and Thomas Mehner whose comments helped improve the manuscript.

Author contributions

SUP, FH, SJ and SS conceived the project. FH, RDTS, DNS and SUP conducted fieldwork. FH generated sequence data. FH and SUP performed the molecular analyses. FH and JDT performed statistical analyses. FH and SUP wrote the manuscript; all authors edited and approved the manuscript.

References

Ahrens D., Monaghan M.T. & Vogler A.P. (2007) DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae). *Molecular Phylogenetics and Evolution*, **44**, 436–449.

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

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D. *et al*. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Brown B. & Swan C. (2010) Dendritic network structure constrains metacommunity properties in riverine ecosystems. *Journal of Animal Ecology*, **79**, 571–580.

Carrara F., Altermatt F., Rodriguez-Iturbe I. & Rinaldo A. (2012) Dendritic connectivity controls biodiversity patterns in experimental metacommunities. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 5761–5766.

Carstens B.C., Pelletier T.A., Reid N.M. & Satler J.D. (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.

Chesters D. (2013) *collapsetypes.pl* [computer software available at http://sourceforge.net/projects/collapsetypes/].

Clarke A., Mac Nally R., Bond N. & Lake P. (2008) Macroinvertebrate diversity in headwater streams: a review. *Freshwater Biology*, **53**, 1707–1721.

Colwell R.K., Chao A., Gotelli N.J., Lin S.-Y., Mao C.X., Chazdon R.L. *et al*. (2012) Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology*, **5**, 3–21.

Craft K.J., Pauls S.U., Darrow K., Miller S.E., Hebert P.D., Helgen L.E. *et al*. (2010) Population genetics of ecological communities with DNA barcodes: an example from New Guinea Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 5041–5046.

De Moor F. & Ivanov V. (2008) Global diversity of caddisflies (Trichoptera: Insecta) in freshwater. *Hydrobiologia*, **595**, 393–407.

De'ath G. & Fabricius K.E. (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, **81**, 3178–3192.

Drummond A., Ashton B., Buxton S., Cheung M., Cooper A., Heled J. *et al*. (2010) *Geneious v6.1.6*. Available at http://www.geneious.com.

Ezard T., Fujisawa T. & Barraclough T. (2009) *Splits: SPecies' LImits by Threshold Statistics, R Package Version 1.0-11/r29*. Available at http://R-Forge.R-project.org/projects/splits/.

Finn D.S., Bonada N., Múrria C. & Hughes J.M. (2011) Small but mighty: headwaters are vital to stream network biodiversity at two levels of organization. *Journal of the North American Benthological Society*, **30**, 963–980.

Folmer O., Black M, Hoeh W., Lutz R. & Vrijenhoek R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fujisawa T. & Barraclough T.G. (2013) Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent (GMYC) approach: a revised method and evaluation on simulated datasets. *Systematic Biology*, **62**, 707–724.

Garcıa A.R. & Ferreras-Romero M. (2008) Distribution patterns of Hydropsychids and Rhyacophilids species (Trichoptera) in a not regulated Mediterranean river (SW Spain). *Limnetica*, **27**, 227–238.

Geraci C.J., Zhou X., Morse J.C. & Kjer K.M. (2010) Defining the genus *Hydropsyche* (Trichoptera: Hydropsychidae) based on DNA and morphological evidence. *Journal of the North American Benthological Society*, **29**, 918–933.

Giarla T.C., Voss R.S. & Jansa S.A. (2014) Hidden diversity in the Andes: comparison of species delimitation methods in montane marsupials. *Molecular Phylogenetics and Evolution*, **70**, 137–151.

Goslee S.C. & Urban D.L. (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

Gotelli N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606–2621.

Gotelli N.J. & Mccabe D.J. (2002) Species co-occurrence: a meta-analysis of JM Diamond's assembly rules model. *Ecology*, **83**, 2091–2096.

Graf W. (2006) *Ecology and Identification of Trichoptera in the HKH Region*. Regional Capacity Building Workshop on the Macro-Invertebrates Taxonomy and Systematics for Evaluating the Ecological Status of Rivers in the Hindukush-Himalayan (HKH) Region, Dhulikhel.

Graf W., Murphy J., Dahl J., Zamora-Munoz C. & López-Rodríguez M.J. (2008) *Distribution and Ecological Preferences of European Freshwater Organisms. Volume 1. Trichoptera*. Pensoft Publishing, Sofia, Moscow.

Hebert P.D., Stoeckle M.Y., Zemlak T.S. & Francis C.M. (2004) Identification of birds through DNA barcodes. *PLoS Biology*, **2**, e312.

Hendrixson B.E., Derussy B.M., Hamilton C.A. & Bond J.E. (2013) An exploration of species boundaries in turret-building tarantulas of the Mojave Desert (Araneae, Mygalomorphae, Theraphosidae, Aphonopelma). *Molecular Phylogenetics and Evolution*, **66**, 327–340.

Hering D., Moog O., Sandin L. & Verdonschot P.M. (2004) Overview and application of the AQEM assessment system. *Hydrobiologia*, **516**, 1–20.

Hildrew A.G. & Edington J.M. (1979) Factors facilitating the coexistence of hydropsychid caddis larvae (Trichoptera) in the same river system. *Journal of Animal Ecology*, **48**, 557–576.

Hjalmarsson A.E., Bergsten J. & Monaghan M.T. (2014) Dispersal is linked to habitat use in 59 species of water beetles (Coleoptera: Adephaga) on Madagascar. *Ecography*, **38**, 732–739.

Joly S., Davies T.J., Archambault A., Bruneau A., Derry A., Kembel S.W. *et al.* (2014) Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Molecular Ecology Resources*, **14**, 221–232.

Katoh K., Misawa K., Kuma K.I. & Miyata T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059–3066.

Kembel S.W., Cowan P.D., Helmus M.R., Cornwell W.K., Morlon H., Ackerly D.D. *et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463–1464.

Kjer K.M., Roshan U. & Gillespie J.J. (2009) Structural and evolutionary considerations for multiple sequence alignment of RNA, and the challenges for algorithms that ignore them. In: *Sequence Alignment* (Ed. M.S Rosenberg), pp. 105–150. Methods, Models, Concepts, and Strategies. University of California Press, Berkeley, California.

Lehrian S., Pauls S.U. & Haase P. (2009) Contrasting patterns of population structure in the montane caddisflies *Hydropsyche tenuis* and *Drusus discolor* in the Central European highlands. *Freshwater Biology*, **54**, 283–295.

Li F., Shah D.N., Pauls S.U., Qu X., Cai Q. & Tachamo Shah R.D. (2016) Elevational shifts of freshwater communities cannot catch up climate warming in the Himalaya. *Water*, **8**, 327.

Lohse K. (2009) Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). *Systematic Biology*, **58**, 439–442; discussion 442–434.

Malicky H. (2006) Caddisflies from Bardia National Park, Nepal, with a preliminary survey of Nepalese species. *Entomofauna*, **27**, 241–264.

Miklós I. & Podani J. (2004) Randomization of presence-absence matrices: comments and new algorithms. *Ecology*, **85**, 86–92.

Monaghan M.T., Balke M., Gregory T.R. & Vogler A.P. (2005) DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **360**, 1925–1933.

Monaghan M.T., Wild R., Elliot M., Fujisawa T., Balke M., Inward D.J. *et al.* (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology*, **58**, 298–311.

Moog O. (2007) *Manual on Pro-Rata Multi-Habitat-Sampling of Benthic Invertebrates from Wadeable Rivers in the HKH-Region*. Deliverable 8, part 1 for ASSESS-HKH, European Commission, Brussels.

Múrria C., Bonada N., Arnedo M.A., Prat N. & Vogler A.P. (2013) Higher β- and γ-diversity at species and genetic levels in headwaters than in mid-order streams in Hydropsyche (Trichoptera). *Freshwater Biology*, **58**, 2226–2236.

Múrria C., Rugenski A.T., Whiles M.R. & Vogler A.P. (2015) Long-term isolation and endemicity of Neotropical aquatic insects limit the community responses to recent amphibian decline. *Diversity and Distributions*, **21**, 938–949.

Obertegger U., Fontaneto D. & Flaim G. (2012) Using DNA taxonomy to investigate the ecological determinants of plankton diversity: explaining the occurrence of Synchaeta spp. (Rotifera, Monogononta) in mountain lakes. *Freshwater Biology*, **57**, 1545–1553.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R. *et al.* (2013) *vegan: Community Ecology Package. Package vegan, Verion 2.2-1*. Available at: https://cran.r-project.org/package=vegan.

Papadopoulou A., Monaghan M.T., Barraclough T.G. & Vogler A.P. (2009) Sampling error does not invalidate the Yule-Coalescent Model for species delimitation. A response to Lohse (2009). *Systematic Biology*, **58**, 442–444.

Pauls S.U., Blahnik R.J., Zhou X., Wardwell C.T. & Holzenthal R.W. (2010) DNA barcode data confirm new species and reveal cryptic diversity in Chilean *Smicridea* (Smicridea) (Trichoptera: Hydropsychidae). *Journal of the North American Benthological Society*, **29**, 1058–1074.

Pauls S.U., Lumbsch H.T. & Haase P. (2006) Phylogeography of the montane caddisfly *Drusus discolor*: evidence for multiple refugia and periglacial survival. *Molecular Ecology*, **15**, 2153–2169.

Pons J., Barraclough T., Gomez-Zurita J., Cardoso A., Duran D., Hazell S. *et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

Previšić A., Graf W., Vitecek S., Kučinić M., Bálint M., Kerezstes L. *et al.* (2014) Cryptic diversity of caddisflies in the Balkans: the curious case of *Ecclisopteryx* species (Trichoptera: Limnephilidae). *Arthropod Systematics & Phylogeny*, **72**, 309.

Puillandre N., Lambert A., Brouillet S. & Achaz G. (2012a) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, **21**, 1864–1877.

Puillandre N., Modica M., Zhang Y., Sirovich L., Boisselier M.C., Cruaud C. *et al.* (2012b) Large-scale species

delimitation method for hyperdiverse groups. *Molecular Ecology*, **21**, 2671–2691.

R Core Team. (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org/.

Rambaut A. & Drummond A. (2007) *Tracer Version 1.4*. Available at http://beast.bio.ed.ac.uk/Tracer.

Rannala B. & Yang Z. (2013) Improved reversible jump algorithms for Bayesian species delimitation. *Genetics*, **194**, 245–253.

Ratnasingham S. & Hebert P.D. (2013) A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS ONE*, **8**, e66213.

Reid N.M. & Carstens B.C. (2012) Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. *BMC Evolutionary Biology*, **12**, 196.

Ross D.H. & Wallace J.B. (1982) Factors influencing the longitudinal distribution of larval Hydropsychidae (Trichoptera) in a southern Appalachian stream system (USA). *Hydrobiologia*, **96**, 185–199.

Salokannnel J., Rantala M.J. & Wahlberg N. (2010) DNA-barcoding clarifies species definitions of Finnish *Apatania* (Trichoptera: Apataniidae). *Entomologica Fennica*, **21**, 1–11.

Sanders N.J., Gotelli N.J., Wittman S.E., Ratchford J.S., Ellison A.M. & Jules E.S. (2007) Assembly rules of ground-foraging ant assemblages are contingent on disturbance, habitat and spatial scale. *Journal of Biogeography*, **34**, 1632–1641.

Satler J.D., Carstens B.C. & Hedin M. (2013) Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, Aliatypus). *Systematic Biology*, **62**, 805–823.

Schwarzfeld M.D. & Sperling F.A. (2015) Comparison of five methods for delimitating species in Ophion Fabricius, a diverse genus of parasitoid wasps (Hymenoptera, Ichneumonidae). *Molecular Phylogenetics and Evolution*, **93**, 234–248.

Tachamo Shah R.D., Sharma S., Haase P., Jähnig S.C. & Pauls S.U. (2015) The climate sensitive zone along an altitudinal gradient in central Himalayan rivers: a useful concept to monitor climate change impacts in mountain regions. *Climatic Change*, **132**, 1–14.

Sheldon A.L. (2016) Mutualism (carpooling) of ecologists and taxonomists. *Biodiversity and Conservation*, **25**, 187–191.

Statzner B. & Dolédec S. (2011) Phylogenetic, spatial, and species-trait patterns across environmental gradients: the case of *Hydropsyche* (Trichoptera) along the Loire River. *International Review of Hydrobiology*, **96**, 121–140.

Statzner B., Douady C.J., Konecny L. & Doledec S.S. (2010) Unravelling phylogenetic relationships among regionally co-existing species: *Hydropsyche* species (Trichoptera: Hydropsychidae) in the Loire River. *Zootaxa*, **2556**, 51–68.

Statzner B. & Mondy N. (2009) Variation of colour patterns in larval *Hydropsyche* (Trichoptera): implications for species identifications and the phylogeny of the genus. *Limnologica-Ecology and Management of Inland Waters*, **39**, 177–183.

Stone L. & Roberts A. (1990) The checkerboard score and species distributions. *Oecologia*, **85**, 74–79.

Tachet H., Pierrot J.P., Roux C. & Bournaud M. (1992) Net-building behaviour of six *Hydropsyche* species (Trichoptera) in relation to current velocity and distribution along the Rhône River. *Journal of the North American Benthological Society*, **11**, 350–365.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Therneau T. & Atkinson B. (2013) *Package 'mvpart'. R Package Version 1.6–1*. Available at http://cran.r-project.org/web/packages/mvpart/mvpart.pdf.

Valentini A., Pompanon F. & Taberlet P. (2009) DNA barcoding for ecologists. *Trends in Ecology & Evolution*, **24**, 110–117.

Vannote R.L. & Sweeney B.W. (1980) Geographic analysis of thermal equilibria: a conceptual model for evaluating the effect of natural and modified thermal regimes on aquatic insect communities. *The American Naturalist*, **115**, 667–695.

Vitecek S., Graf W., Previšić A., Kučinić M., Oláh J., Bálint M. *et al.* (2015) A hairy case: the evolution of filtering carnivorous Drusinae (Limnephilidae, Trichoptera). *Molecular Phylogenetics and Evolution*, **93**, 249–260.

Vuataz L., Sartori M., Wagner A. & Monaghan M.T. (2011) Toward a DNA taxonomy of Alpine *Rhithrogena* (Ephemeroptera: Heptageniidae) using a mixed Yule-coalescent analysis of mitochondrial and nuclear DNA. *PLoS ONE*, **6**, e19728.

Wagner R., Marxsen J., Zwick P. & Cox E.J. (Eds.) (2011) *Central European Stream Ecosystems: The Long Term Study of the Breitenbach*. Wiley-VCH, Weinheim.

Wang Y., Naumann U., Wright S.T. & Warton D.I. (2012) mvabund – an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution*, **3**, 471–474.

Webb C.O., Ackerly D.D., McPeek M.A. & Donoghue M.J. (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.

Woodward G., Jones J.I. & Hildrew A.G. (2002) Community persistence in Broadstone Stream (UK) over three decades. *Freshwater Biology*, **47**, 1419–1435.

Yang Z. (2015) *The BPP Program for Species Tree Estimation and Species Delimitation, Current Zoology*, **61**, 854–865.

Zhou X., Frandsen P., Holzenthal R., Beet C.R., Bennett K.R., Blahnik R. *et al.* (2016) The Trichoptera barcode initiative: strategies for integrative taxonomy and generating a species-level tree of life. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **371**, 20160025.

Zhou X., Kjer K.M. & Morse J.C. (2007) Associating larvae and adults of Chinese Hydropsychidae caddisflies

(Insecta:Trichoptera) using DNA sequences. *Journal of the North American Benthological Society*, **26**, 719–742.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Fit of the GMYC model to the molecular data for maximum clade credibility trees.

Figure S2. Bayesian maximum clade credibility tree of 50 COI-5P Hydropsychidae haplotypes.

Figure S3. Bayesian maximum clade credibility tree of 85 COI-3P Hydropsychidae haplotypes.

Figure S4. Bayesian maximum clade credibility tree of 81 28S Hydropsychidae haplotypes.

Figure S5. Estimated richness of MOTUs based on samples and individuals.

Figure S6. Number of overlapping MOTUs in the four different river systems recovered in all data sets.

Figure S7. Map showing MOTU composition changes among 34 Hydropsychidae communities.

Figure S8. The most probable species tree based on the mitochondrial cytochrome c oxidase subunit I (COI-3P) and the nuclear large subunit ribosomal ribonucleic acid (28S).

Table S1. The best model of sequence evolution (based on AIC) for each data set.

Table S2. Results of the Bayesian BPP method.

Table S3. Results of generalised linear model (GLM) tests for the effect of spatial and environmental predictors.

Table S4. Coefficients of GLM regressions for all different data sets.

Table S5. Pearson product-moment correlation tests between explanatory variables.

Table S6. Results of multivariate regression tree (MRT) tests.

Table S7. Results of mantel test for distance matrices for individual river systems and each MOTU data set.

Table S8. Pearson product-moment correlation tests between distance metrics.

Table S9. Analysis of phylogenetic structure of Hydropsychidae communities.

(Manuscript accepted 1 August 2016)