



# Gill Structure Linked to Ecological and Species Diversification in a Clade of Caddisflies

Steffen U. Pauls<sup>1,2,3</sup>, Wolfram Graf<sup>4</sup>, Anna E. Hjalmarsson<sup>1</sup>, Alan Lemmon<sup>5</sup>, Emily Moriarty Lemmon<sup>6</sup>, Malte Petersen<sup>7</sup>, Simon Vitecek<sup>1,4,8</sup>, Paul B. Frandsen<sup>2,9,10</sup>

1 Senckenberg Research Institute and Natural History Museum Frankfurt, Senckenberganlage 25, 60325 Frankfurt/Main, Germany

2 LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt/Main, Germany

3 Institute of Insect Biotechnology, Justus-Liebig-University Gießen, Heinrich-Buff-Ring 26-32, 35392 Gießen

4 Universität für Bodenkultur Wien, Institut für Hydrobiologie und Gewässermanagement, Gregor-Mendel-Straße 33, 1180 Vienna, Austria

5 Department of Scientific Computing, Florida State University, 400 Dirac Science Library, Tallahassee, FL, 32306-4120, USA

6 Department of Biological Science, Florida State University, 89 Chieftain Way, Biology Unit 1, Tallahassee, FL 32306-4295, USA

7 High Performance Computing & Analytics Lab, University of Bonn, Friedrich-Hirzebruch-Allee 8, 53115 Bonn, Germany

8 WasserCluster Lunz – Biol. Station GmbH, Dr. Carl Kupelwieser Promenade 5, 3293 Lunz am See, Austria

9 Brigham Young University, Department of Plant and Wildlife Sciences, Provo, UT 84602, USA

10 Smithsonian Institution, Data Science Lab, Office of the Chief Information Officer, PO Box 37012, Washington, D.C. 20013-7012, USA

<https://zoobank.org/F0CB276A-AE14-46B0-BC40-396E19934956>

Corresponding author: Steffen U. Pauls ([steffen.pauls@senckenberg.de](mailto:steffen.pauls@senckenberg.de))

Received 31 July 2023  
Accepted 28 October 2023  
Published 8 December 2023

Academic Editor Anna Hundsdörfer

**Citation:** Pauls SU, Graf W, Hjalmarsson AE, Lemmon A, Lemmon EM, Peterson M, Vitecek S, Frandsen PB (2023) Gill Structure Linked to Ecological and Species Diversification in a Clade of Caddisflies. *Arthropod Systematics & Phylogeny* 81: 917–929. <https://doi.org/10.3897/asp.81.e110014>

## Abstract

Streams represent a special case of directional environmental gradients where ecological opportunity for diversification may be associated with upstream and downstream dispersal into habitats that differ in selective pressures. Temperature, current velocity and variability, sediment erosion dynamics and oxygen saturation are key environmental parameters that change in predictable ways from springs to river mouth. Many aquatic insects occupy specific longitudinal regions along these gradients, indicating a high degree of adaptation to these specific environmental conditions. In caddisflies, the evolution of tracheal gills in larval and pupal stages may be a major driver in oxygen uptake efficiency and ecological diversification. Here we study the evolution of larval gill structure in the *Rhyacophila vulgaris* species group using phylogenomic methods. Based on anchored hybrid enrichment, we sequenced 97 kbp of data representing 159 independent nuclear protein coding gene regions to infer the phylogeny of the *R. vulgaris* species group, whose species exhibit both high diversity of gill types and varied longitudinal preferences. We find that the different gill types evolved independently as derived characters in the genus and that gill structure is linked to the longitudinal habitat preference, thereby serving as a possible ecological key innovation in the *R. vulgaris* group.

## Keywords

Anchored hybrid enrichment, phylogenomics, Trichoptera, Rhyacophilidae, gill evolution

## 1. Introduction

Along strong environmental gradients, ecological opportunity for diversification may be associated with dispersal into habitats that differ in selective pressures (Yoder et al. 2010). Streams represent a special case of such gradients where physico-chemical conditions change longitudinally along the river, i.e. from the headwaters downstream (e.g., Illies 1961). These longitudinal gradients make streams and their diverse invertebrate radiations excellent models for studying ecological and ecomorphological diversification (Dijkstra et al. 2014).

A prominent feature of stream ecosystems is the differential availability of oxygen in different, hierarchically structured river stretches and (micro)habitats which is linked to current velocity, water temperature, nutrient load (N, P, and C), sediment structures, and aquatic and riparian vegetation (Illies 1961; Vannote et al. 1980; Minshall et al. 1985, Leopold 1994). Habitat suitability in downstream river sections may be limited for organisms with high oxygen demand, e.g. stoneflies, or salmonids, because increased water temperature and reduced turbulence in water flow, coupled with high biological oxygen demand can lead to reduced oxygen availability (Schönborn & Risse-Buhl 2013).

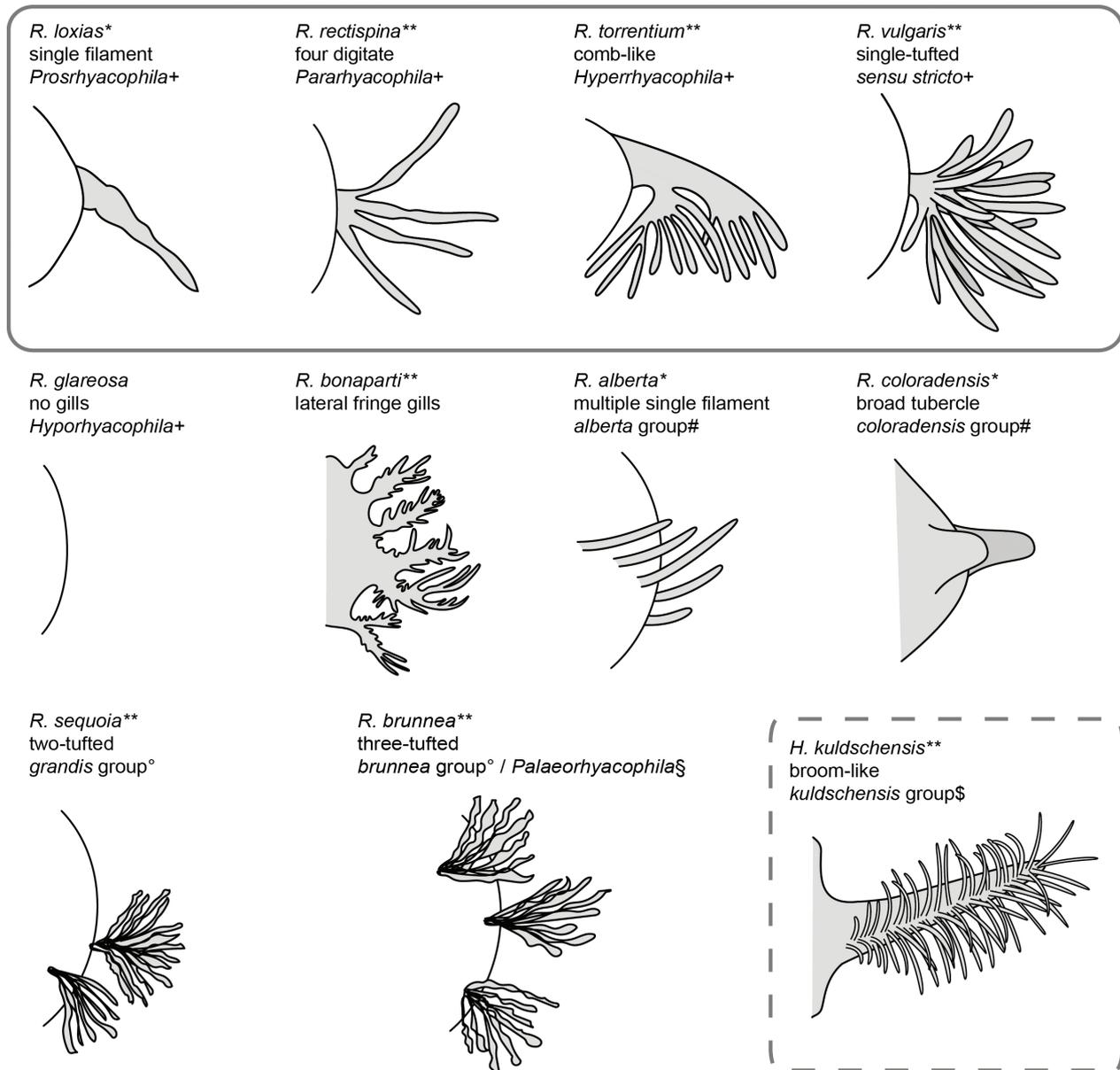
In contrast to salmonids or stoneflies, caddisflies have evolved to occupy virtually all types freshwater habitat outside of Antarctica, from highly oxygenated headwater streams to stagnant, ephemeral ponds. One hypothesis associates the basis of this ecological diversity with case and pupal building variation among caddisflies (Mackay & Wiggins 1979, Dijkstra et al. 2014). In some families of caddisflies, portable tube cases are used to enhance respiratory efficiency (Dodds and Hisaw, 1924; Wiggins, 1996) via abdominal undulation that causes water to be pulled in through the front opening of the larval case and pushed out through the rear, thereby improving oxygen uptake which is additionally facilitated by tracheal abdominal gills (reviewed in Holzenthal et al. 2015). Pupae in these groups furthermore spin porous cocoons that allow the entrance of water to carry oxygen directly to the body during metamorphosis (Wiggins and Wichard, 1989). Ross (1956) proposed that these species were better adapted to survival in both well and poorly oxygenated habitats. In other, putatively more primitive groups like Annulipalpia or Rhyacophilidae, larvae are caseless and/or pupate in enclosed semi-permeable cocoons. Both characteristics have been postulated to limit these groups to cool, fast flowing, well-oxygenated water (Ross 1956, Wiggins 1996).

Based on current knowledge, most caseless caddisflies are limited to such well-oxygenated habitats. A few representatives of caseless caddisflies, most notably gilled Hydropsychidae and gillless Polycentropodidae, are, however, exceptional in that they comprise species that occur in poorly oxygenated stretches along the stream continuum, including stagnant pools. Rhyacophilidae are caseless predators that primarily occupy cold, turbulent and well-oxygenated streams. They are Holarctic and

Oriental in distribution with major centers of diversity in the Oriental region (~500 species) as well as Nearctic and European mountain ranges (Parey et al. 2023; Schmid 1970; Sun 2016). Intriguingly they also boast an unparalleled diversity of tracheal gill conformations (Figure 1): The genera *Himalopsyche* (56 species known to date, Hjalmarsson et al. 2019, Hjalmarsson 2019) and especially *Rhyacophila* (over 800 species) exhibit four and ten different known types of tracheal abdominal gills of varying surface area and structure (Smith 1995, Doepler 1950; Lepneva 1970; Waringer & Graf 2011). Döhler (1950) defined species groups based on gill configuration of larvae. Because these groupings turned out to not be phylogenetically relevant (Schmid 1970), we follow Schmid's (1970) terminology of species groups and only use Döhler's definitions to indicate larval gill configuration (Fig. 1). Gills are present in several *Rhyacophila* species groups, either as simple short, broad tubercles (*R. coloradensis* group), or simple single fingerlike projections (*R. alberta* group in North America; *Prosrhyacophila* as defined by Doepler (1950) in Europe), or complex, branched gills with numerous filaments (*R. grandis* and *R. brunnea* groups in North America (Smith 1995), *Palaeorhyacophila* as defined by Lepneva (1970) in Asia, or *Rhyacophila sensu stricto* as defined by Döhler 1950 in Europe).

Diversity in gill structure is not uncommon in caddisflies. For example, final instars of Limnephilidae generally have abdominal gills that can be made up of single filaments, multiple filaments, a mixture of these and/or have varying configurations on different abdominal segments (Waringer & Graf 2011). In *Himalopsyche*, four distinct gill types are known, all of which are complex (Hjalmarsson et al. 2018). Although these gill types exhibit striking differences, they appear to be early features in the genus and have remained relatively stable within groups over long periods of time (Hjalmarsson et al. 2018). In contrast, the extent of variation observed within the genus *Rhyacophila* and, in particular, within the *R. vulgaris* species group is not known from other groups of caddisflies, and appears to have evolved both repeatedly and, in some cases, as very derived character conditions. The fact that we observe independent development of structurally identical single filament gills or multiple convergent evolution of tufted gills (either with single or multiple tufts) suggests a flexible evolutionary background for gill configuration in *Rhyacophila*. Considering the large number of known species for which larvae remain unknown, it is possible and likely that additional gill types will be discovered, particularly in understudied regions of Asia.

While species of *Rhyacophila* are restricted to fast-flowing water in North America and Asia, some European species occur in moderately flowing stream sections. In this context the *R. vulgaris* group is particularly interesting. This species group currently comprises 70 species in Europe. Among these 70 species we find a striking diversity of gill types (Fig. 1). And, whereas many species in this group inhabit fast flowing sections of highland streams, a few species like *R. pascoei* can even occur in the potamal sections of lowland rivers. The *R. vulgaris*



**Figure 1.** Known gill configurations in *Rhyacophila* and *Himalopsyche kuldschensis* groups, shown as a latero-apical depiction of the gill filaments on the side of larval abdominal segments. Grey box indicates *R. vulgaris* group species; dashed line segregates the representative of *Himalopsyche*. — \* For the purpose of this manuscript we define simple gills of single unbranched filaments as “simple” gills. — \*\* For the purpose of this manuscript we define “complex” gills as those gills that exhibit a branching pattern; + sensu Döhler (1950); sensu Giersch 2002; ° sensu Giersch & Wissemann 2012; § sensu Lepneva (1970); \$ sensu Hjalmarsson et al. (2018).

group is thus at the center of our analysis linking phylogeny of the group with gill structure and river zonation preference.

In this study, we analyze ecological traits and phylogenetic data to help shed light on how morphological traits evolved and what ecological function they may have. Specifically, we ask 1) if individual gill types in *Rhyacophila* evolved once or repeatedly, 2) if structurally complex gill types are derived or ancestral in the *R. vulgaris* group, and 3) if there is a link between gill structure and distribution of *Rhyacophila* along the stream zonation gradient. To address these questions, we infer phylogenetic relationships in the genus *Rhyacophila* using an anchored hybrid enrichment-based multi-locus phylogenomic approach, with a particular focus on the *R. vulgaris*

group. We estimate ancestral character states for different gill types along the phylogeny, particularly within the *R. vulgaris* group where numerous gill types of varying structure are known (Fig. 1). Additionally, we assess the river zonation distribution of European *Rhyacophila* with different gill types as river zonation information is too limited in other regions. Because the genus *Rhyacophila* is too diverse to be sampled entirely and gill types are unknown for the majority of species outside of Europe, the sampling strategy for this study had three priorities. First, we put an extensive effort on deeply sampling the *R. vulgaris* group to best address the above questions in a species group with high gill type and ecological variation. Second, we sampled two or more representatives from all branches of *Rhyacophila sensu* Schmid (1970). Third,

we sampled all gill types known to us. This resulted in 60 species analyzed in the data set (see Supplementary Material A). Our sampling is suited to assess the placement of the *vulgaris* group within *Rhyacophila* and the relationships within that group. For deeper level relationships within Rhyacophilidae our sampling of 60 species of this 700+ species genus is limited and we refrain from discussing the results in detail, as they are not pertinent to the main focus and objectives of this study, i.e. assessing if different gill types evolved multiple times within genus *Rhyacophila* and assessing the evolution of gill types within *R. vulgaris* species group.

## 2. Material and Methods

### 2.1. Specimen collection and DNA extraction

Specimens were collected with light traps or sweep nets and preserved in either 70% or 95% ethanol. Prior to DNA extraction, wings (in adults), heads and terminal abdominal segments were removed as specimen vouchers. DNA was extracted with hot sodium hydroxide as described in Truett et al. (2000). Extracted DNA was visualized on an agarose gel to estimate the mean size of the fragments. Concentration of double stranded DNA was fluorometrically estimated using a Qubit assay kit (Thermo Fisher Scientific). The DNA pellet was washed 2x with 70% EtOH. Following the wash, the EtOH was discarded and the pellet dried. The pellet was sent dried to the Center for Anchored Phylogenomics at Florida State University (<http://www.anchoredphylogeny.com>) for quality check and library preparation, sequencing and raw data handling following Lemmon et al. (2012) and Prum et al. (2015).

### 2.2. Probe Design

The probes were designed in collaboration with the Center for Anchored Phylogenomics following Breinholt et al. (2018) and Deng et al. (2022). In short, the transcripts identified as homologous with the existing Lepidoptera probe set from the 1Kite Project (Misof et al. 2014, Breinholt et al. 2018) were aligned in MAFFT v7.023b (Katoh & Standley 2013) by target locus, then trimmed to conserved regions (Lemmon et al. 2012), and finally manually inspected in Geneious R9, (Biomatters Ltd., Kearse et al. 2012). A total of 960 target loci (averaging 232 bp in length) remained after masking/removing regions identified to be repetitive using *k*-mer distribution profiling (see Hamilton et al. 2016 for details). For each of these loci, probes were tiled uniformly across each of the sixteen reference sequences with 4.2x coverage, to produce 57094 probes. Probes were produced by Agilent as a SureSelect XP kit. Methods for probe tiling are described in Deng et al. (2022).

### 2.3. Library Preparation and Enrichment

Extracted DNA was sonicated using the Covaris Ultrasonicator to a size distribution of 150–400 bp. Libraries were then prepared following Lemmon et al. (2012) and indexed with single 8-bp indexes chosen for being different at two or more sites. Libraries were then combined in pools of 16 samples and enriched using the Agilent® SureSelect XP probe kit. The enriched libraries were sequenced on two lanes of an Illumina HiSeq 2500 sequencer with a paired end 150 bp sequencing protocol and onboard cluster generation. The raw data totaled 88 gbp.

### 2.4. Read assembly

Raw read pairs passing the CASAVA high-chastity filter were merged following Rokyta et al. (2012) and library adapters were trimmed using TrimGalore! V. 0.4.0 (<https://github.com/FelixKrueger/TrimGalore>). Reads were then assembled to the reference *R. fasciata* (probe region sequences) using the assembly procedure described by Prum et al. (2015) with the same parameters as Hamilton et al. (2016). Consensus sequences were derived from assembly clusters containing ten or more reads and ambiguities were called when site patterns could not be explained by a 1% sequencing error. Orthology among consensus sequences was determined using a neighbor-joining approach based on a pairwise distance matrix (see Hamilton et al. 2016 for details). Orthologous sequences were aligned using MAFFT v7.023b (Katoh & Standley 2013) in Geneious R9 (Biomatters Ltd., Kearse et al. 2012).

### 2.5. Data Filtering

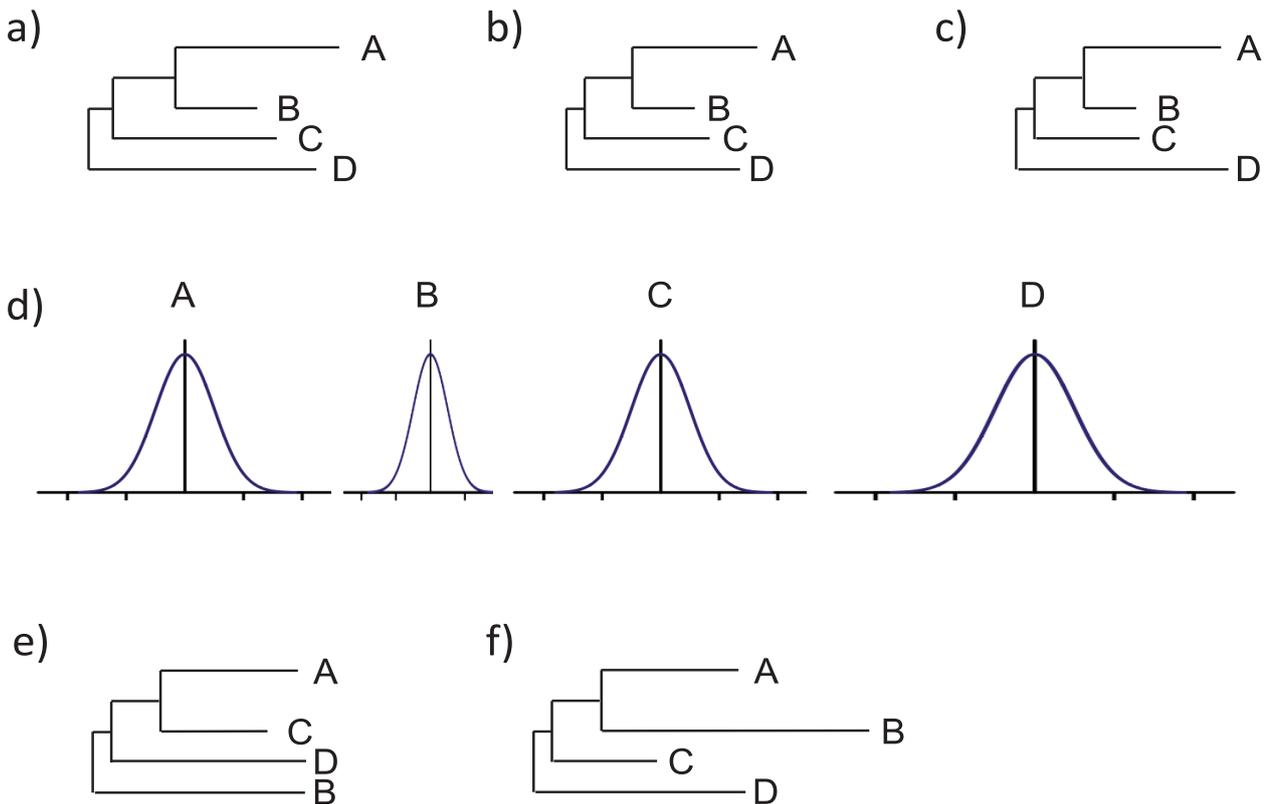
High-throughput sequencing data have been shown to be susceptible to contamination (e.g. Lusk 2014). Also, poor alignment of individual taxa or entire regions within aligned loci impact the phylogenetic inference derived from those alignments. We therefore performed multiple filtering steps to obtain our final data set for analysis.

### 2.6. Removing poorly aligned data

We used Gblocks (Castresana 2000) to remove poorly aligned regions of regions containing gaps as we expected invariant length and gap free alignment across the exons we sequenced. We used default settings to obtain a stringent alignment without indels.

### 2.7. Identification of potential contaminant sequences

We then used a relative branch length approach to identify potential contaminants (Fig. 2). The rationale for this is that if a taxon is constrained in a phylogeny to a certain



**Figure 2.** The absolute branch lengths vary between a) and b) because the evolutionary rates of the two loci differ. But the relative branch length of each terminal taxon is constant. There is some variation in the relative branch length depending on the rate of variation regarding each taxon in each locus (compare b) and c)). This „normal“ variation leads to normal distribution of relative terminal branch lengths for each taxon (d)). In case a sequence is affected by contamination, e.g. cross-contamination of DNA from a different organism, or chimera assembly of different organisms, the placement of the contaminated taxon may be different in a new phylogenetic inference (e.g. taxon B in e)). But especially using many short loci, most single loci phylogenetic inferences are poor and not very precise. It is thus difficult to assess if the different topology is the product of limited phylogenetic signal in the locus, a contamination or true gene tree discordance. If, however we constrain the topology of the data set in e), the odd taxon will sit on an exceptionally long terminal branch (f)), that can be identified as an outlier in the distribution of relative terminal branch lengths for taxon B. In the case of true gene tree discordance we can expect the relative branch lengths for several taxa to be flagged as being outliers. The approach should be independent of the constraint tree used, as the distribution of branch lengths for a misplaced taxon will generally be too long, but an outlier can still be detected.

position, its relative branch length in each locus should vary following a normal distribution. In the case that a sequence represents a contaminant and is forced into a phylogenetic position it would not take based on the data, that branch should be unusually long. Thus, the branch will be an outlier to the normal distribution and could be interpreted as the result of contamination. We used the outlier definition according to Tukey (1997): it uses the first and third quartile ( $Q1$ ,  $Q3$ ) of the branch length distribution and the interquartile range  $IQR = Q3 - Q1$ . An outlier is defined as a value that is larger than  $Q3 + IQR * k$  or smaller than  $Q1 - IQR * k$ , with  $k = 1.5$ . Because this criterion is very conservative and can lead to false positives, we used  $k = 3$ . The implementation also allows us to flag only extreme outliers with  $k = 9$ , and the option to use a flexible  $\sigma$  multiplier instead, with outliers being outside of  $[Q1 - k * \sigma, Q3 + k * \sigma]$ .

Prior to using this approach on our study data, we examined its utility and reliability on simulated DNA sequence data. First, we used Seq-Gen v.1.3.3 (Rambaut & Cassidy 1997) to simulate DNA sequences for 100 loci

(500 bp, HKY model, nucleotide frequencies A: 0.294, T: 0.246, G: 0.241, C: 0.221) based on the characteristics of our own concatenated alignment and the resulting guide tree. To simulate contamination, we manipulated the simulated data set in three ways: 1) by randomly replacing 10 complete loci with others from within the simulated data set to imitate cross contamination within a study; 2) by randomly inserting 10 120-bp-fragments from within the simulated data to other species sequences to imitate chimera production in the original sequence assembly; and 3) by randomly inserting 10 120-bp-long fragments of random data to imitate contamination from an unknown source in the original sequence assembly. Randomization of loci, taxa and insert position were based on Random Thing Picker (<http://andrew.hedges.name/experiments/random/pickone.html>, accessed on May 12, 2016). We then performed phylogenetic analysis on all 4 simulated data sets (each form of contamination individually and all contaminations combined) under the topological constraint and estimated relative terminal branch lengths for each taxon individually across all loci. We used the

approach described above to identify if the integrated contaminants resulted in outlier branch lengths. Under the topological constraint of our guide tree, our approach identified 25 of the 30 contaminations (all random inserts, all cross contaminations, five of the 10 chimeras) from our simulated data set. We also flagged two false positive outliers. These results show that the approach is useful in detecting contaminants, including small random inserts and potentially even chimeras with congeners. However, half of the chimeras were not flagged and it is possible that some false positives could be flagged and uncritically removed from a data set. We view the approach as a tool for flagging potential contaminants that allows us to further scrutinize the flagged sequences by hand, since it is time-consuming and error-prone to scrutinize all sequences of all loci manually.

## 2.8. Identification and deletion of contaminant sequences

We applied our contamination detection approach to our original 61 taxon data set. For the input constraint tree, we used a maximum likelihood tree generated in RAxML v. 8 (Stamatakis 2014) using the concatenated data set under a GTRCAT model with 100 fast bootstrap replicates. We then calculated the distribution of relative maximum likelihood branch lengths of each taxon for all 451 loci and flagged all sequences that represented statistical outliers in relative branch length for each taxon. The procedure was straightforward for the 66 loci that had sequence data for all 61 taxa. For 385 loci we had some taxa with missing data (from 1 to 23 taxa missing). For each locus with missing data, we generated a new constraint tree by dropping the terminals for which we had no data and used the locus-specific reduced constraint trees to identify potential contaminants using the method described above.

Our approach flagged 33 sequences from the 66 loci that had sequence data for all 61 taxa, and 230 sequences from the 385 loci with missing data, resulting in 263 total flagged sequences. We removed these sequences unless the outlier branch length was the result of the sister taxon to the flagged taxon having been removed from the single locus constraint tree due to missing data. In this case, the terminal branch length of the flagged specimen is necessarily longer than usual because the closest taxon is missing from the single locus constraint tree. Using these guidelines we removed 126 potential contaminant sequences.

Additionally, we removed 116 potential cross-contaminated sequences that were identical among non-sister taxa (based on the constraint tree). In total 242 individual sequences were removed (1.03%).

## 2.9. Preparation of final data sets

The most important predictor of phylogenetic informativeness of a particular locus is simply sequence length (Kjer et al. 2007). We thus concatenated remaining

loci into “super loci”, *i.e.*, we concatenated all loci that mapped to the same gene, but different exons in *Bombyx mori*. Following this, we removed all (super) loci that were <100 bp in length as well as all (super) loci and taxa with >35% missing data.

This filtering resulted in a final data set comprising 60 taxa, 159 (super) loci ranging from 302–1922 bp in length ( $\bar{O}$  570.5 bp), and a total concatenated alignment length of 90,717 bp. The molecular data sets are available at the Senckenberg (Meta) Data Portal under <https://dataportal.senckenberg.de/dataset/cf47cddb-7384-47e4-a89d-cb1abcf65963>.

## 2.10. Phylogenetic analysis

We followed three general approaches to infer the phylogenetic relationships in our data set. First, we generated a maximum likelihood tree on the concatenated alignment (90,717 bp). We used the PartitionFinder 2.0 relaxed clustering algorithm with default settings to select an optimal partitioning scheme (Lanfear et al. 2017). Then we selected nucleotide substitution models for each subset of the optimal partitioning scheme using ModelFinder as implemented in IQtree v. 1.6 (Kalyaanamoorthy et al. 2017, Nguyen et al. 2015). We then used to IQtree v. 1.6 to estimate 50 maximum likelihood trees, 25 replicates with a parsimony tree and 25 with a random starting tree, and chose the topology with the best likelihood score. We evaluated node support using 1000 ultrafast bootstrap replicates in IQtree v1.6 (Hoang et al. 2018).

Second, we generated a species tree in ASTRAL v. 5.6.1 (Zhang et al. 2018), which accounts for incomplete lineage sorting by considering conflict among gene trees. For each (super) locus, we selected the best fit model in IQtree using ModelFinder, and then generated 10 maximum likelihood trees and chose the topology with the best maximum likelihood score. We then used the best topology for each (super) locus as input into ASTRAL to generate the species tree (Supplementary Material B).

Third, we identified the best partitioning scheme and nucleotide substitution models via ModelFinder. For each of the 159 (super)loci, we used independent, variable substitution rates assuming an average substitution rate of 1 across all (super)loci and independent character state frequencies, transition/transversion rate ratios, gamma parameter shapes and proportions of invariable characters. Based on this model we generated a Bayesian tree with MrBayes v. 3.1 (Ronquist & Huelsenbeck 2003). We ran Bayesian inference for  $10 \times 10^6$  generations in 2 independent runs with 8 chains each, sampling each 5000<sup>th</sup> generation. We used average standard deviations of split frequencies as reported by MrBayes and visualizations of MrBayes parameter files in combination with ESS estimates as available in Tracer v. 1.6 (Rambaut et al. 2018) to determine when analyses had converged and reached stationarity. After roughly  $4.5 \times 10^6$  generations, average deviation of split frequencies remained under 0.01 while visualizing MrBayes parameters indicated stationary distribution with ESS estimates generally above 200

for the majority of parameters when using a burn-in of 50%. Accordingly, we discarded the first 1000 trees as burn-in and used the remaining trees for ancestral character state estimations (see below) as well as to construct a strict consensus tree.

## 2.11. Ancestral Character State Estimations of Gill Types

In total, we differentiated the eleven known gill types in Rhyacophilidae that were included in our phylogeny. Species where larval gill types are not known were coded as “unknown” and treated as missing data. We estimated the ancestral character state of gill type at each supported node along the backbone of our phylogeny as well as all nodes within the *R. vulgaris* group. Ancestral character estimation (ACE) was performed in BayesTraits V3.0 (available from <http://www.evolution.rdg.ac.uk>), using the BayesMultistate method (Pagel et al. 2004). The multistate method estimates trait values at ancestral nodes for traits that adopt a finite number of discrete states using Markov chain Monte Carlo (MCMC) or Maximum Likelihood (ML) methods. This method fits a continuous-time Markov model to the discrete character data. In this model, trait values can change along the branches at infinitesimally small time intervals, and transition rates from state *i* to *j* are defined by a transition rate  $q_{ij}$ .

We performed MCMC ancestral character estimation on 16 nodes with a uniform prior on a sample of trees from the MrBayes output at two levels: first with a binary coding of gills/no gills, and second for all eleven character states (Figs 1, 3).

## 2.12. Ecological Statistics

Abdominal tracheal gills in caddisflies have been associated with osmoregulatory and breathing functions (Badcock et al. 1987). We thus assessed the interaction between gill types and the river zonation preference of larval *Rhyacophila* for the European species only. Taking advantage of the longitudinal river zonation as a general proxy for oxygen, current, and conductivity conditions (Moog et al. 2002), Ofenböck et al. (2010) developed the longitudinal river zonation index (LZI) to categorize European benthic invertebrate species according to their longitudinal occurrence within European river systems. The ecological analysis was necessarily restricted to European species of *Rhyacophila* because detailed species-level information on gill types and ecological traits are not available or known in other regions. We used the data from freshwater ecology.info (Graf et al. 2015, Schmidt-Kloiber & Hering 2015) for this analysis. Of the 116 species listed, the gill types and river zonation are known for 50 species (Supplementary Material C). We applied analysis of variance (ANOVA) to assess if river zonation occurrence differed between the different gill types, the null hypothesis being that all gill groups have the same river zonation index distribution.

## 3. Results

### 3.1. Phylogenetic analysis

Our final data set comprised 60 taxa (Supplementary Material A). Of these, 19 species belonged to the *vulgaris*-species group *sensu* Schmid (1970). Thirty-nine additional species of *Rhyacophila* were sampled from all evolutionary “branches” proposed by Schmid (1970) covering 11, 5, and 10 additional species groups from North America, Europe, and Asia, respectively. In addition, our data set included one specimen of *Himalopsyche kuldschensis* (Rhyacophilidae) and one Glossosomatidae (*Agapetus hessi*) as outgroups. Within our taxon sampling the gill configuration of last instar larvae is known for 43 species (~72%). In our analyses we included all known gill configurations.

### 3.2. Phylogenetic Relationships

Briefly, all phylogenetic analyses recovered almost congruent tree topologies: only minor changes in branch length were returned and almost all relationships received high support values (Fig. 3). We thus describe the results on the basis of the Bayesian Inference, which was also used for ancestral character state reconstruction. Phylogenetic groupings, *i.e.*, ‘branches’, ‘veins’ and ‘species groups’ were introduced by Schmid (1970) as hierarchical categories with branches forming large clades, veins forming groups within branches, and species groups forming smaller groups within veins. Unless stated otherwise, we use this terminology, and compare our results with the branches and species groups hypothesized by Schmid (1970). Regarding the deep splits within Rhyacophilidae, *R. rickeri* is sister to all other *Rhyacophila* and the nested *Himalopsyche kuldschensis*. The *Rhyacophila stigmatica* group follows as sister to all remaining *Rhyacophila*. At this point the phylogeny splits *Rhyacophila* into two main clades. Clade A (pp = 1.0) comprises species from the *divarica*, *naviculata*, *philopotamoides* and *vulgaris* branches, and Clade B (pp = 1.0) comprises species from the *naviculata*, *philopotamoides* and *vulgaris* branches, rendering these branches as defined by Schmid (1970) polyphyletic. The deeper nodes in clades A and B resolve the relationships within these clades but exhibit very short internodes. At the species level, six of seven species groups sampled with more than one species were inferred as being monophyletic (except the *naviculata* group).

Of the *vulgaris* branch species in Clade A, *R. oreata*, *R. dongkyapa* and *R. petersorum* form a monophylum. The remaining *vulgaris* branch species form a clade (Clade C; pp = 1.0) within Clade B. In Clade C, *R. fuscula* of the *fuscula* group is sister to the supported *vulgaris* species group clade. Within the *vulgaris* group (Clade D-G; all pp = 1.0) all relationships are resolved. However, some internodes are very short.

Both structurally simple and complex gill types are distributed across the phylogeny, with only a few gill

types representing apomorphies for specific branches, veins or species groups: *H. kuldschensis*, *R. alberta*, *R. bonaparti* and *R. coloradensis* have unique gill types in the phylogeny, while gill types observed in *R. alberta*, *R. coloradensis* and *H. kuldschensis* are also known from other species that we did not include in our study (Giersch 2002, Hjalmarsson et al. 2018). Single filament gills (*Prosrhyacophila sensu* Döhler (1950)) occur in different clades within Clade B. Gill tufts occur in several taxa of Clade B and can be differentiated depending on the configuration in which they occur on each abdominal segment: in *R. brunnea* (*brunnea* group *sensu* Smith 1995) there are three independent tufts, in *R. sequoia* (*grandis* group *sensu* Smith 1995) there are two independent tufts and the *vulgaris* group exhibits a single tuft (*Rhyacophila sensu stricto* as defined by Döhler 1950). Four digitate gills (*Pararhyacophila* as defined by Döhler (1950) and comb-like gills (*Hyperrhyacophila* as defined by Döhler (1950)) are limited to the *vulgaris* group.

In the *vulgaris* group (Fig. 3, clade D), *R. loxias* (single filament gill) is sister to all other taxa. *Rhyacophila ferox* subsequently follows as sister to the remaining *vulgaris* group species. *Rhyacophila ferox* has four digitate gills on each abdominal segment. In our analysis it is placed within the *vulgaris* group, though its placement was previously considered to be isolated or distantly related to *R. fasciata* on the basis of preanal appendages, anal sclerites and the aedeagus (Graf 2006; Graf et al. 2009).

The *vulgaris* group then splits into three clades that correspond to different gill configurations. The *Pararhyacophila* clade (Clade E) comprises *R. rectispina*, *R. intermedia*, *R. lusitanica* and *R. munda*, whose larvae exhibit four digitate gills on each abdominal segment. The *Pararhyacophila* clade is sister to the *Hyperrhyacophila* + *sensu stricto* clades. The *Hyperrhyacophila* clade (Clade F) comprises all sampled species with larvae exhibiting comb-like gills (*R. torrentium*, *R. evoluta*, *R. occidentalis*) as well as the *sensu stricto* species *R. moscaryi* at its base. All other species with tufted gills in the data set fall into the *sensu stricto* clade (Clade G). Based on the polarization of the tree with the outgroup, it appears that the clades within the *vulgaris* group follow an evolutionary trajectory where deeper nodes exhibit ancestors with simple gill configurations and shallow nodes bear ancestors with complex gills.

The IQTree ML analysis resulted in a topology that showed the same topology as the Bayesian Inference. We highlight differences in support values on those nodes with bootstrap supports (bs) lower than 70 (Fig. 3). The ASTRAL species tree of the 159 loci (each > 300 bp in length; Supplementary Material B), shows almost identical relationships throughout. The only conflict of supported nodes (bs > 70) with the Bayesian and ML inference relates to the placement of *H. kuldschensis* and *R. rickeri*: In the species tree, these form a supported clade that is sister to the *stigmatica* group, and nested within *Rhyacophila*. In the Bayesian inference *R. rickeri* is sister to all other *Rhyacophila*, while *H. kuldschensis* remains nested within *Rhyacophila*. Additionally, *R. vibox* was not placed with support in the species tree (bs = 0.39).

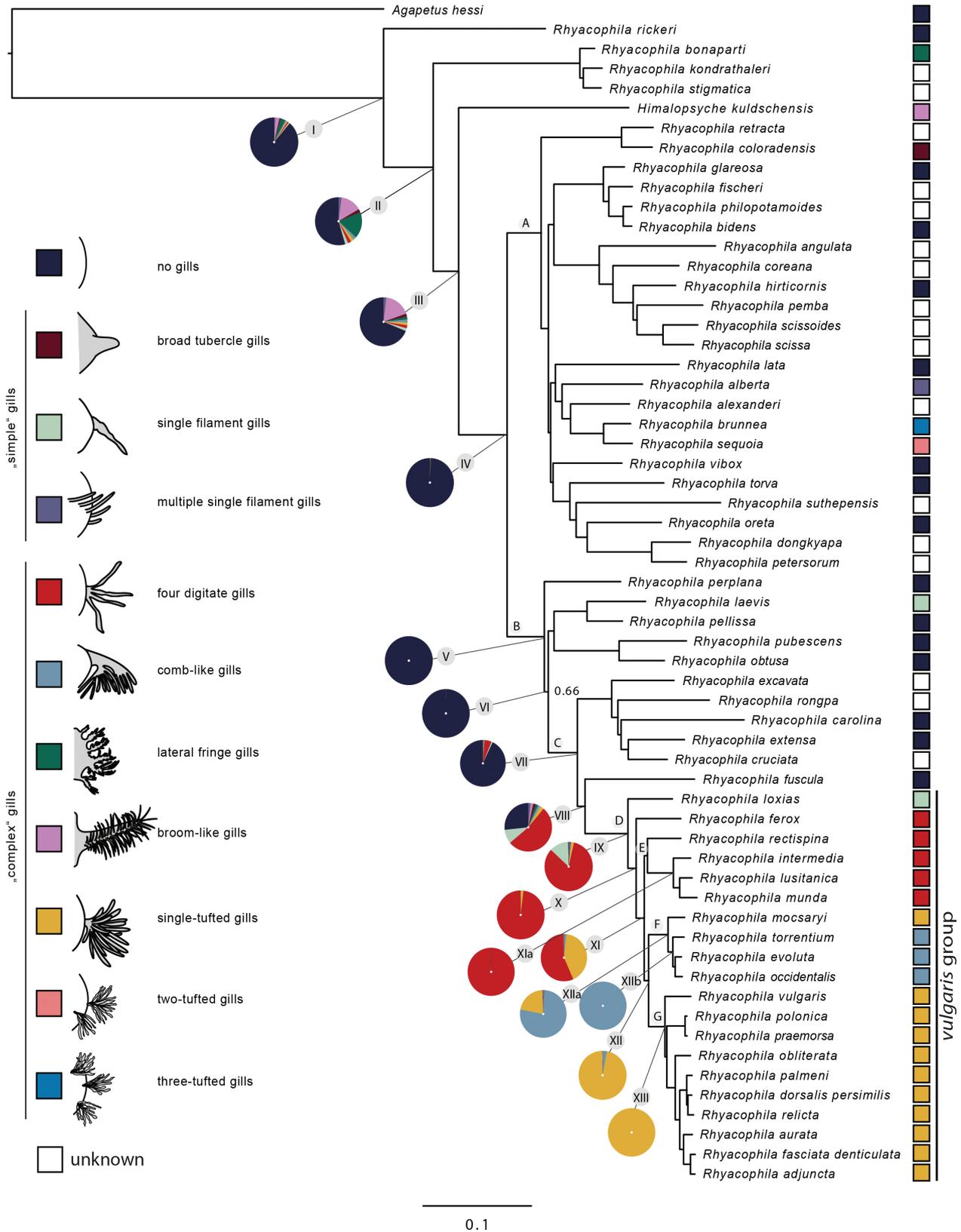
### 3.3. Ancestral Character State Estimations

The results from the Bayesian multistate analysis showed that acceptance rates of proposed rate changes stabilized around 38% indicating sufficient mixing of the MCMC chain. The resulting probabilities for ancestral character states in the binary coding (no gills/gills) were the same as those for the sum of all individual gill states vs no gills. In accordance with our sampling design and hypotheses we focused on the ancestral character state estimations in the *R. vulgaris* group (*sensu* Schmid 1970; Fig. 3, Clade D). The *R. vulgaris* group comprises four different gill types, making it the most diverse in terms of gill types known to date. However, we also estimated ancestral character states along the phylogeny backbone leading to the *vulgaris* group. The analyses indicate a gillless ancestor for the *Rhyacophilidae* and the genus *Rhyacophila* with *Himalopsyche* nested within (Fig. 3, nodes I–IV). The ancestor for *Himalopsyche* cannot be estimated based on the single representative of the genus. Nodes V–VII also show gillless larvae as the most likely ancestral state. This pattern changes in clade C, when the estimate of ancestral states along the backbone first shifts to four digitate gills at nodes VIII–XI and then to single-tufted gills at nodes XII–XIII (Fig. 3). Comb-like gills are estimated to arise with node XIIa. Based on ancestral state estimations, the evolution of gill shapes for the *vulgaris* group would thus be no gills > four digitate gills > single-tufted gills. Under these estimates the simple single digitate gills would have arisen either from a gillless ancestor or an ancestor with four digitate gills (Fig. 3, nodes VIII–IX), whereas the complex comb-like gills either arose from an ancestor with four digitate gills or single-tufted gills (Fig. 3, nodes XI–XII, XIV).

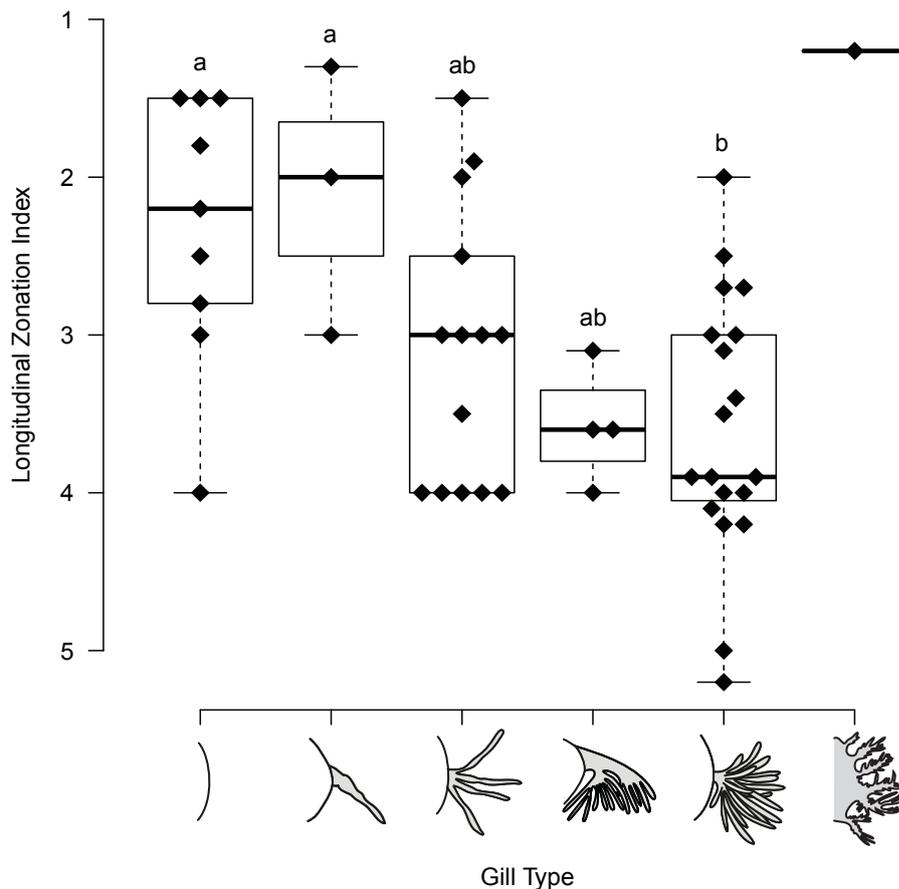
Of the 10 recognized gill types in *Rhyacophila* it appears that only single digit gills (*R. loxias* and *R. laevis*) evolved more than once. While tufted gills also evolved multiple times, they did so with different configurations in the *vulgaris* group (single tuft), *brunnea* group (three tufts) and *grandis* group (two tufts) (Giersch & Wissemann 2012). Ancestral character estimations indicate convergent evolution in both cases.

### 3.4. Phenotype-Environment Interactions

The distribution of river zonation occurrences for individual gill types is presented in Figure 4. The data used are summarized in Supplementary Material C. Species with single-tufted gills have the broadest distribution and extend furthest downstream. Their distribution is significantly different from that of species without gills or species with single digitate gills (ANOVA,  $F_{4,4} = 5.175$ ;  $p = 0.00167$ ; Tukey HSD: *R. sensu stricto* vs *Hyporhyacophila*  $p = 0.0034$ ; *R. sensu stricto* vs *Prosrhyacophila*  $p = 0.042$ ). While there appears to be a trend of species with increasing structural complexity and gill surface area to occur further downstream, differences between other gill types are not significant.



**Figure 3.** Strict consensus tree from MrBayes output with BayesTraits ancestral state reconstructions for gill types. Gill types are illustrated on the left and assigned a color, which are assigned to taxa on the right-hand side of the tree. The posterior probability is shown on those nodes with a posterior probability below 1.0. Unlabelled nodes have a posterior probability of 1.0. In nodes with bootstrap support <70 in the IQTree ML inference, the bootstrap values are shown after the posterior probability for the respective node. Pie charts with the probabilities for the various gill types ancestral states are connected to their assigned node via numbered lines.



**Figure 4.** Box and whisker plots of longitudinal occurrence of 50 European *Rhyacophila* species based on the longitudinal zonation index (see Supplementary Material C). The species are grouped by the six gill types known to occur in the European fauna. Groups with significantly different longitudinal distribution ranges are indicated by letters above the box and whisker plots. Boxes represent the 50<sup>th</sup> percentiles, the whiskers extend to the 100<sup>th</sup> percentile. Horizontal lines within boxes indicate median values.

## 4. Discussion

Our results show that structurally complex gill types, i.e. the four digitate gills, comb-like gills and single-tufted gills evolved as derived character states in the *R. vulgaris* group (Fig. 3). We also show that the prevalence of these gill types is associated with stream sections downstream of those where species without gills or with structurally simple gill types that do not greatly increase surface area predominate. And while the deeper level phylogeny is not the focus of this study, a few key points are of interest. We will discuss these three aspects in turn.

Tracheal gills in aquatic insects have been shown to improve oxygen uptake (e.g., Eriksen & Moeur 1990). Jacobsen (2000) observes greater proportions of gills surface to total body surface in larger species of caddisflies. There are several studies that indicate that, with increasing size, larvae have increasing metabolic demands on oxygen uptake that are in some species compensated by an increase in gill surface area relative to body size. Jacobsen (2000) argues that this is in part because the efficiency for oxygen uptake through diffusion is physically greater in gills than the remaining body surface because the gill cuticle is thinner and the diffusive oxygen flux per unit surface area increases with decreasing cylinder diameter (Nobel 1991). This may also be the reason why early instars of *R. ferox* and *R. intermedia* larvae have fewer gill filaments than final instars (Graf et al. 2005).

Several empirical examples support the relationship between environmental oxygen conditions and gill surface area. Wichard (1974a) showed that the number and length of the gill filaments in different limnephilid species is greater in habitats with poorer oxygenation. Badcock et al. (1987) showed that individuals living in habitats with higher temperature and thus lower oxygen concentrations also had more gill filaments than conspecifics in lower temperatures. Wichard (1974b) further showed that within two species of Limnephilidae the number of gill filaments increases significantly more in the fourth larval instar in specimens subjected to lower oxygen habitat conditions. Based on these data it seems reasonable to assume that the occurrence and structure of gills as well as gill surface area, i.e. filament number, are also relevant for *Rhyacophila* to persist in habitats with varying levels of oxygenation.

Phenotype-environment correlations and trait evolution in adaptive radiations have been widely studied to gain insight into the dynamics underpinning rapid species diversification (Wilson et al. 2015), and can generally improve our understanding how diversity is formed and maintained. Our study shows that phylogenetic analysis of ecological diversification is a useful approach to study such questions, particularly when phylogenies are based on sufficient amounts of data (Trautwein et al. 2012). Aquatic insects and the strong ecological gradients in aquatic ecosystems set an excellent stage for studies on eco-evolutionary processes (Dijkstra et al. 2014). Yoder et

al. (2010) posited that “ecological opportunity for diversification can arise by means of an acquired key innovation, dispersal into a new habitat, or the extinction of an antagonist.” Gill structure in *Rhyacophila* seems to have either evolved in concert with downstream expansion of the genus or, more complex gill types evolved in certain groups allowing them to subsequently expand their range into more downstream habitats. In the first case, the transition from headwater to midstream sections and the associated strong environmental gradients could create a situation where dispersal into adjacent habitats could be associated with changes in selective pressures which ultimately can lead to parapatric diversification (Yoder et al. 2010).

If this were the case, we would, however, expect to see a similar pattern evolve multiple times across the genus because the transition of physical properties from headwater to midstream to downstream habitat conditions is a ubiquitous feature of running waters. It thus seems more likely that structurally complex gills evolved progressively in the *vulgaris* group, and subsequently allowed these species to cope better with low oxygen conditions that are more frequently encountered in relatively downstream conditions (Illies 1961, Vannote et al. 1980). Under this scenario, the evolution of gill structure could have enabled ecological opportunities for diversification and thus represent a key ecological innovation for the *R. vulgaris* group.

The taxon sampling in our phylogeny is clearly insufficient to draw expansive conclusions on the phylogeny or extensive biogeographic analyses of the family. Yet some observations suggest that such a study would be very worthwhile. Where our sampling comprises more than one relevant taxon, many hypotheses on the relatedness of species, i.e. the composition and distinction of species groups, seem supported. However, regarding the deeper relationships, our analysis did not recover most of the groupings suggested by Schmid (1970). For example, the placement of *Himalopsyche* within *Rhyacophila* questions the validity of the current generic concept in Rhyacophilidae. This is in line with Ross’s (1956) hypothesis that *Himalopsyche* is nested within *Rhyacophila* based on sclerotization of the aedeagus. Perhaps this feature is systematically more relevant than the wing venation used to separate the otherwise very similar genera (*Rhyacophila* wing R4 apical and R5 posterior to wing tip; *Himalopsyche* R veins apical and M veins posterior to wing tip). And while the presented data improves our understanding of basal relationships compared to the most recent 2-gene molecular phylogenetic inference (McLaughlin et al. 2019), the differences between the species tree and Bayesian/ML inferences related to placement of three species (*H. kuldschensis*, *R. rickeri*, *R. vibox*) indicates that basal relationships are not yet fully resolved. It is a crucial next step to assess the evolution of the group as whole on the basis of a much broader taxon sampling and potentially revise the taxonomic family. Considering the interesting eco-morphological results our study produced it is likely that such an endeavour would uncover many additional aspects of caddisfly evolution and ecological diversification.

## 5. Author contributions

SUP and WG designed the study. AL performed the wet lab methodology. PF, MP, SUP, SV and AEH performed the analyses and prepared the figures; AEH and SV illustrated the gill types. SUP and PF drafted the paper. All authors edited the manuscript and approved of the final version.

## 6. Conflict of interest

The authors declare that they do not have any conflicts of interest.

## 7. Acknowledgements

We thank Karl Kjer for important discussions on this study, and Anna Hundsdörfer and three anonymous reviewers for helpful comments on an earlier version of this manuscript. Jürgen Otte (Senckenberg, Frankfurt) is thanked for assisting in DNA extractions. We thank Christine Frandsen for assisting with the figures. This study was supported by the LOEWE-Centre for Translational Biodiversity Genomics funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK).

## 8. References

- Badcock RM, Bales MT, Harrison JD (1987) Observation on gill number and respiratory adaptation in caddis larvae. In: Bournaud M, Tachet H (Eds) Proceedings of the 5th International Symposium on Trichoptera. Dr W Junk, Dordrecht, The Netherlands, 175–178.
- Breinholt JW, Earl C, Lemmon AR, Lemmon EM, Xiao L, Kawahara AY (2018) Resolving the relationships among the megadiverse butterflies and moths with a novel pipeline for Anchored Phylogenomics. *Systematic Biology* 67: 78–93. <https://doi.org/10.1093/sysbio/syx048>
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17: 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- Deng X, Frandsen PB, Dikow R, Favre A, Narayan Shah D, Tachamo Shah RD, Schneider J, Heckenhauer J, Pauls SU (2022) The impact of sequencing depth and relatedness of the reference genome in population genomic studies: a case study with two caddisfly species (Trichoptera, Rhyacophilidae, *Himalopsyche*). *Ecology and Evolution*, e9583. <https://doi.org/10.1002/ece3.9583>
- Dijkstra K-DB, Monaghan MT, Pauls SU (2014) Freshwater Biodiversity and Insect Diversification. *Annual Review of Entomology* 59: 143–163. <https://doi.org/10.1146/annurev-ento-011613-161958>
- Dodds GS, Hisaw FL (1924) Ecological studies in aquatic insects. II. Size of respiratory organs in relation to environmental conditions. *Ecology* 5: 262–271. <https://doi.org/10.2307/1929452>
- Döhler W (1950) Zur Kenntnis der Gattung *Rhyacophila* im mitteleuropäischen Raum (Trichoptera). *Archiv für Hydrobiologie* 44: 271–293.
- Eriksen CH, Mæur JE (1990) Respiratory Functions of Motile Tracheal Gills in Ephemeroptera Nymphs, As Exemplified by *Siphonurus Occidentalis* Eaton. In: Campbell IC (Ed) *Mayflies and Stoneflies:*

- Life Histories and Biology. Series Entomologica, 44. Springer, Dordrecht. [https://doi.org/10.1007/978-94-009-2397-3\\_14](https://doi.org/10.1007/978-94-009-2397-3_14)
- Giersch JJ (2002) Revision and phylogenetic analysis of the *verrula* and *alberta* species groups of *Rhyacophila* Pictet 1834 with description of a new species (Trichoptera: Rhyacophilidae). Entomology, Montana State University, Bozeman, Montana.
- Giersch JJ, Wisseman RW (2012) Annotated list of *Rhyacophila* of North America with larval key and descriptions. Southwest Association of Freshwater Invertebrate Taxonomists Workshop: The Trichoptera of Western North America, Larval, Pupal and Adult Taxonomy, plus Distribution and Biology. February 18–19, 2012, University of California-Davis. 136 pp.
- Graf W (2006) A new brachypterous species of *Rhyacophila* (Trichoptera: Rhyacophilidae) from the Eastern Alps (Carinthia, Austria). *Braueria* 33: 22.
- Graf W, Murphy J, Dahl J, Zamora-Muñoz C, López-Rodríguez MJ, Schmidt-Kloiber A (2015) Dataset “Trichoptera”. [www.freshwaterecology.info](http://www.freshwaterecology.info) – the taxa and autecology database for freshwater organisms, version 7.0.
- Graf W, Pauls SU, Waringer J (2009) The Larva of *Rhyacophila ferox* Graf, 2006 (Trichoptera: Rhyacophilidae) from the Eastern Alps (Carinthia, Austria). *Aquatic Insects* 31: 111–117. <https://doi.org/10.1080/01650420802627159>
- Hamilton CA, Lemmon AR, Lemmon EM, Bond JE (2016) Expanding Anchored Hybrid Enrichment to resolve both deep and shallow relationships within the spider Tree of Life. *BMC Evolutionary Biology* 16: 212. <https://doi.org/10.1186/s12862-016-0769-y>
- Hjalmarsson AE, Graf W, Jähnig SC, Vitecek S, Pauls SU (2018) Molecular association and morphological characterization of *Himalopsyche* larval types (Trichoptera, Rhyacophilidae). *ZooKeys* 773: 79–108. <https://doi.org/10.3897/zookeys.773.24319>
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522. <https://doi.org/10.1093/molbev/msx281>
- Holzenthal RW, Thompson RE, Rios-Touma B (2015) Chapter 38, Order Trichoptera. In: Thorp JH, Rogers DC 2015. Thorp and Covich’s Freshwater Invertebrates: Ecology and General Biology. Volume 1. Elsevier. 965–1002.
- Illies J (1961) Versuch einer allgemeinen biozönotischen Gliederung der Fließgewässer. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* 46: 205–213. <https://doi.org/10.1002/iroh.19610460205>
- Jacobsen D (2000) Gill Size of Trichopteran Larvae and Oxygen Supply in Streams along a 4000m Gradient of Altitude. *Journal of the North American Benthological Society* 19: 329–343.
- Jacobsen D, Rostgaard S, Vasconez JJ (2003) Are macroinvertebrates in high altitude streams affected by oxygen deficiency? *Freshwater Biology* 48: 2025–2032. <https://doi.org/10.1046/j.1365-2427.2003.01140.x>
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kjer KM, Swigonova Z, LaPolla JS, Broughton RE (2007) Why weight? Molecular Phylogenetics and Evolution 43: 999–1004. <https://doi.org/10.1016/j.ympev.2007.02.028>
- LANFEAR R, FRANDSEN PB, WRIGHT AM, SENFELD T, 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. <https://doi.org/10.1093/molbev/msw260>
- Lemmon AR, Emme S, Lemmon EM (2012) Anchored hybrid enrichment for massively high-throughput phylogenetics. *Systematic Biology* 61: 721–744. <https://doi.org/10.1093/sysbio/sys049>
- Lusk RW (2014) Diverse and Widespread Contamination Evident in the Unmapped Depths of High Throughput Sequencing Data. *PLoS ONE* 9, e110808. <https://doi.org/10.1371/journal.pone.0110808>
- Mackay RJ, Wiggins GB (1979) Ecological diversity in Trichoptera. *Annual Review of Entomology* 24: 185–208. <https://doi.org/10.1146/annurev.en.24.010179.001153>
- Mclaughlin JE, Frandsen PB, Mey W, Pauls SU (2019) A Preliminary Phylogeny of Rhyacophilidae with Reference to *Fansipangana* and the Monophyly of *Rhyacophila*. *Zoosymposia* 14: 189–192. <https://doi.org/10.11646/zoosymposia.14.1.20>
- Moog O, Hartmann A (Eds.) (2017) Fauna Aquatica Austriaca, 3. Edition 2017. BMLFUW, Vienna.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274. <https://doi.org/10.1093/molbev/msu300>
- Ofenböck T, Moog O, Hartmann A, Stubauer I (2010) Leitfaden Zur Erhebung Der Biologischen Qualitätselemente. Teil A2 – Makrozoobenthos. BMLFUW, Vienna.
- Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53: 673–684. <https://doi.org/10.1080/10635150490522232>
- Parey SH, Ali T, Pandher MS (2023) Rhyacophilidae of India. Jenny Stanford Singapur. 1–309. <https://doi.org/10.1201/9781003410959>
- Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR (2015) A Fully Resolved, Comprehensive Phylogeny of Birds (Aves) using Targeted Next Generation DNA Sequencing. *Nature* 526: 569–573. <https://doi.org/10.1038/nature15697>
- Rambaut A, Grass NC (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Bioinformatics* 13: 235–238. <https://doi.org/10.1093/bioinformatics/13.3.235>
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67: 901–904. <https://doi.org/10.1093/sysbio/syy032>
- Rokyta DR, Lemmon AR, Margres MJ, Arnow K (2012) The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics* 13: 312. <https://doi.org/10.1186/1471-2164-13-312>
- Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180>
- Schmid F (1970) Le genre *Rhyacophila* et la famille des Rhyacophilidae (Trichoptera). *Mémoires de la Société Entomologique du Canada* 66: 1–230. <https://doi.org/10.4039/entm10266fv>

- Schmidt-Kloiber A, Hering D (2015) www.freshwaterecology.info – an online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences. *Ecological Indicators* 53: 271–282. <https://doi.org/10.1016/j.ecolind.2015.02.007>
- Smith SD (1995) Revision of the genus *Rhyacophila* (Trichoptera, Rhyacophilidae). Central Washington University, Ellensburg, Washington.
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Sun CH (2016) New species, new records, and new collection data of *Rhyacophila* from China (Trichoptera: Rhyacophilidae). *Zootaxa* 4189: 134–144. <https://doi.org/10.11646/zootaxa.4189.1.6>
- Trautwein MD, Wiegmann BM, Beutel R, Kjer KM, Yeates DK (2012) Advances in Insect Phylogeny at the Dawn of the Postgenomic Era. *Annual Review of Entomology* 57: 449–468. <https://doi.org/10.1146/annurev-ento-120710-100538>
- Truett GE, Mynatt RL, Truett AA, Walker JA, Warman ML (2000) Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and Tris (HotSHOT). *BioTechniques* 29: 52–54. <https://doi.org/10.2144/00291bm09>
- Tukey JW (1977) *Exploratory data analysis*. Addison-Wesley, Reading, Massachusetts.
- Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE (1980) The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37: 130–137. <https://doi.org/10.1139/f80-017>
- Waringer J, Graf W (2011) *Atlas of Central European Trichoptera Larvae*. Dinkelscherben: Erik Mauch Verlag.
- Wichard W (1974a) Morphological Adaptations of Tracheal Gills of Larval Limnephilini Kol. (Insecta, Trichoptera) I. Autecological Studies in the Lake District near Eggstitt in Chiemgau. *Oecologia* 15: 159–167.
- Wichard W (1974b) Morphological Adaptations of Tracheal Gills of Larval Limnephilini Kol. (Insecta, Trichoptera) II. Adaptations in Larval Development under Different Oxygen Conditions. *Oecologia* 15: 169–175.
- Wiggins GB (1996a) *Larvae of the North American Caddisfly Genera (Trichoptera)*, 2<sup>nd</sup> ed. University of Toronto Press, Toronto. 1–457. <https://doi.org/10.3138/9781442623606>
- Wiggins GB, Wichard W (1989) Phylogeny of Pupation in Trichoptera, with Proposals on the Origin and Higher Classification of the Order. *Journal of the North American Benthological Society* 8: 260–276. <https://doi.org/10.2307/1467330>
- Wilson L, Colombo M, Sánchez-Villagra M, Salzburger W (2015) Evolution of opercle shape in cichlid fishes from Lake Tanganyika – adaptive trait interactions in extant and extinct species flocks. *Scientific Reports* 5: 16909. <https://doi.org/10.1038/srep16909>
- Yoder JB, Clancey E, Des Roches S, Eastman JM, Gentry L, Godsoe W, Hagey TJ, Jochimsen D, Oswald BP, Robertson J, Sarver BAJ, Schenk JJ, Spear SF, Harmon LJ (2010) Ecological opportunity and the origin of adaptive radiations. *Journal of Evolutionary Biology* 23: 1581–1596. <https://doi.org/10.1111/j.1420-9101.2010.02029.x>
- Zhang C, Rabiee M, Sayyari E, Mirarab S (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153. <https://doi.org/10.1186/s12859-018-2129-y>

## Supplementary Material 1

### Supplementary Material A–C

**Authors:** Pauls SU, Graf W, Hjalmarsson AE, Lemmon A, Lemmon EM, Peterson M, Vitecek S, Frandsen PB (2023)

**Data type:** .pdf

**Explanation note:** **A.** Specimens used in the phylogenomic analysis in this study. — **B.** ASTRAL species tree based on 60 taxa and 159 individual loci of > 302 bp in 8 length. — **C.** *Rhyacophila* gill type and longitudinal habitat preference. — Molecular Data is available at the Senckenberg (Meta) Data Portal under <https://dataportal.senckenberg.de/dataset/cf47cddb-7384-47e4-a89d-cb1abcf65963>

**Copyright notice:** This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

**Link:** <https://doi.org/10.3897/asp.81.e110014.suppl1>