Systematic and highly resolved modelling of biodiversity in inherently rare groundwater amphipods

Mara Knüsel1,2 | Roman Alther1,2 | Nadine Locher2 | Arpat Özgül1 | Cene Fišer3 | Florian Altermatt1,2

Abstract

Aim: Groundwater ecosystems harbour a unique biodiversity, but remain poorly studied, mainly due to difficulties in accessibility and imperfect species detection. Consequently, knowledge on the state and change of groundwater biodiversity remains highly deficient. In the context of global warming and excessive groundwater extraction, understanding groundwater from an ecosystem-perspective, including organism diversity and distribution, is essential. This study presents the largest ever systematic assessment of groundwater amphipods, which are a key component of European groundwater biodiversity.

Location: Switzerland (41,285 km²), including data from 906 sampling sites.

Taxon: Groundwater amphipods, genera Niphargus and Crangonyx (Crustacea, Amphipoda).

Methods: We applied a highly standardized citizen science approach to collect repeated groundwater fauna samples in collaboration with municipal drinking water providers. Using detection–nondetection data of the genetically identified groundwater amphipod species, we assessed the overall species diversity of both rare and common species. The distribution of commonly found species was predicted using multispecies occupancy modelling.

Results: We retrieved 3882 samples from 906 sites, yielding 2350 groundwater amphipod individuals. We identified a remarkable species diversity, comprising few commonly and many rarely found species. Considering commonly found species, we identified distinct distribution ranges, low local species richness and a predominance of negative co-occurrences. In contrast, a major portion of species were found rarely (generally at just one or two sites each), distributed uniformly throughout the study area and unrelated to common species’ recognized hotspots. Many of these rarely found species are not yet formally described.

Main Conclusions: Our results give robust emphasis on the rare occurrence and narrow distribution of many groundwater dwellers. Our systematic and standardized sampling data of groundwater amphipods suggest that rarity is particularly prominent and inherent to groundwater organisms. We emphasize the need of systematic data...
to integrate rare groundwater species in biodiversity assessments, especially in times of global change.

KEYWORDS
Citizen science, Niphargus, occupancy model, species distribution, stygofauna, subterranean, Switzerland

1 | INTRODUCTION

In the water bodies beneath the earth’s surface, a diverse range of organisms exists, adapted to the dark and energy-limited environment (Culver & Pipan, 2019). It is assumed that the majority of this stygofauna, consisting of obligate groundwater organisms, has yet to be discovered and described (Stoch & Galassi, 2010; Zagmajster et al., 2018). Nonetheless, geographic distribution, evolutionary history and ecological functioning often remain unresolved even for described species (Mammola et al., 2020). The lack of knowledge is mostly due to the difficulty or impossibility to systematically sample groundwater habitats by humans. These habitats remain poorly explored, mapped and sampled in comparison to surface environments (so-called Rovežnian impediment, Ficetola et al., 2019). Consequently, groundwater biodiversity contributes to the least known portion of the overall biodiversity worldwide (Mammola et al., 2019; Sánchez-Fernández et al., 2021). The incomplete knowledge of the stygofauna is alarming and hampers adequate management and conservation of groundwater ecosystems (Mammola et al., 2020). It is especially concerning given that groundwater habitats are becoming increasingly affected by human activities, such as groundwater depletion through overextraction and climate change, contamination and the alteration of water temperature through geothermal usage (Famiglietti, 2014; Griebler et al., 2016; Korbel et al., 2019; Lall et al., 2020; Nanni et al., 2023).

One major challenge of groundwater biodiversity research is the shift from local, point-based knowledge to a regional or even global and spatially inclusive understanding of biodiversity patterns (Koch et al., 2024; Mammola et al., 2020). While point-based assessments are a window into local groundwater ecosystems, they are insufficient to understand biodiversity patterns on a macroecological scale (Hortal et al., 2015). For a better understanding and conservation of biodiversity at regional to global scales, solid species distribution data are required (Malard et al., 2023). Gaining such knowledge is especially challenging for groundwater ecosystems, as many species are sparsely distributed and the proportion of narrowly endemic species is very high (Mammola & Leroy, 2018; Trontelj et al., 2009). To capture the whole groundwater biodiversity on a large scale, it is therefore essential to account for both the common and many rare species.

Data scarcity, uneven sampling and presence-only data are common to virtually all ecosystems, yet may be particularly prevalent for groundwater systems (Mammola & Leroy, 2018; Wynne et al., 2019; Zagmajster et al., 2018). Presence-only datasets have the drawback that they are missing the systematic collection of nondetection data, where a species was searched for, but not found (Kéry & Royle, 2016). The lack of repeated data including nondetections is problematic, as the detection and occurrence processes cannot be separated (Kéry & Schaub, 2012). Was a species found infrequently because of imperfect detection (despite occurring frequently) or because of true underlying infrequent occurrence? For ecosystems with a high number of rare and narrowly distributed species, such as the groundwater (Malard et al., 2023), this question is central, because failure to account for imperfect detection and varying sampling effort can lead to biased inference about species distributions (Guillera-Arroita, 2017; Kéry & Schmidt, 2008).

Integrating detection and occurrence in species distribution modelling is thus paramount, in particular when the assumption of perfect detection cannot be held. Such modelling approaches usually require a large amount of data (Fernandes et al., 2018; Stockwell & Peterson, 2002), which—as stated above—has been a prevalent limitation in groundwater systems (Ficetola et al., 2019). One possible solution to resolve this limitation and access the required data involves novel citizen science approaches. Highly standardized citizen science approaches allow scalable and unprecedented access to large amount of data especially for systems otherwise hard to cover (Altwegg & Nichols, 2019; Feldman et al., 2021), and a proof of principles of their suitability in groundwater systems has recently been made (Alther et al., 2021; Raghavan et al., 2023).

Here, we scaled a novel, structured citizen science approach to understand the biodiversity of groundwater amphipods at a fine resolution across all of Switzerland. We established an unprecedented network of local drinking water providers to obtain repeated detection–nondetection data, giving one of the globally most coherent coverages of groundwater amphipods assessed in a systematic approach. Our main questions were as follows: (i) What patterns of groundwater biodiversity emerge when systematically sampled? (ii) How do both common and rare species shape these biodiversity patterns? (iii) What distribution and co-occurrence patterns exist, and how do they influence conservation and management strategies?

2 | MATERIALS AND METHODS

2.1 | Citizen science approach

The groundwater fauna samples were collected countrywide across Switzerland in a systematic approach from sites used for drinking water provisioning. Thus, citizen scientists in this study correspond...
to drinking water providers and infrequent private well owners. As drinking water provisioning is structured at the municipal level in Switzerland, we used a regular grid to select municipalities for participation in our study. In addition, we had an increased sampling density in one region of Switzerland (Tös Valley, see also Couton, Hürlemann, Studer, et al., 2023; Couton, Hürlemann, Locher, et al., 2023; Studer et al., 2022). We contacted the corresponding local drinking water providers and asked for their participation in the citizen science sampling. Participating drinking water providers were equipped with a sampling kit, containing the sampling material, instructions and protocols needed for data collection (see Appendix S1, Figure S1.1a). Water providers are well trained to take, for example, standardized water quality measurements. We consider the observer bias negligible, as Studer et al. (for example, standardized water quality measurements. We consider further processing.

data on repeatability of detection. Samples were stored at 4°C until pepods are not captured. In general, the week-long sampling was collecting macroinvertebrates, while smaller organisms such as co- molecular grade ethanol. The sampling protocol is designed for organisms were preserved again in separate tubes containing 80%

filter nets remained attached for weekly intervals and all retained Sefiltec AG, Höri, Switzerland) at one or multiple inlets of the drain- grade ethanol. Second, they attached filter nets (mesh size 0.8 mm, JBL GmbH & Co. KG, Neuhofen, Germany) and preserved all organ- isms that were visually observed by the naked eye in 80% molecular grade ethanol. Second, they attached filter nets (mesh size 0.8 mm, Sefiltec AG, Höri, Switzerland) at one or multiple inlets of the drain- age pipe(s) into the spring box (Appendix S1, Figure S1.1b,c). These filter nets remained attached for weekly intervals and all retained organisms were preserved again in separate tubes containing 80% molecular grade ethanol. The sampling protocol is designed for collecting macroinvertebrates, while smaller organisms such as copepods are not captured. In general, the week-long sampling was repeated multiple times (1–29 times, median 3) per inlet pipe, to gain data on repeatability of detection. Samples were stored at 4°C until further processing.

2.2 | Species identification

We focus here on amphipods, which are the most common and most diverse of groundwater macroinvertebrates in Switzerland (Schneider et al., 2023). We sorted all groundwater amphipods from the remaining organisms and then applied DNA barcoding to identify each individual to the species level. We amplified a 658-bp long fragment of the mtDNA COI gene (Folmer fragment, Folmer et al., 1994), using the primers LCO1490-JJ and HCO2198-JJ (Aastrin & Stüben, 2008). For the PCR reactions, we used a Multiplex PCR Kit (QIAGEN, Basel, Switzerland). Sanger sequencing was conducted by Microsynth AG (Balgrach, Switzerland). We aligned and visually checked the forward and reverse sequences for each individual in CodonCode Aligner v-10.0.1 (CodonCode Corporation, Centerville, United States). The sequences were then assigned to a lab-internal database using a BLAST-like algorithm from the R package ‘blaster’ (Steven et al., 2018; Tamminen et al., 2021), or alternatively compared to BOLD (package ‘bold’, Chamberlain, 2021) or NCBI's GenBank using Blast®. A detailed protocol is provided in Appendix S1. All distinct groundwater amphipod taxa (including formally described species, well-known phylogenetic lineages and newly found yet undescribed species) are in the following referred to as 'species'.

2.3 | Environmental and sampling variables

We aggregated all species detection–nondetection data based on day and site (unique date and unique spring box based on coordi- nates). Specifically, if multiple drainage pipe inlets and/or overflow basins were sampled inside one spring box at a given day, we aggregated the corresponding findings. To account for differences in sampling efforts arising from this aggregation, we extracted two parameters: the number of subsamples (sum of sampled drainage pipe inlets and overflow basins) that were collected at a given day and site, and the sampling method (hand net only or full method includ- ing at least one filter net sample, because hand net samples from the overflow basin were performed in a less standardized manner than filter net samples from drainage pipe inlets).

For each sampling site, we extracted variables related to topogra- phy, hydrogeology, glaciation history and biogeographic region, covering a variety of environmental predictors relevant to subterranean ecosystems (Dole-Olivier, Malard, et al., 2009; Knüsel et al., 2024; Mammola & Leroy, 2018; Weber et al., 2020). As a measure for topography, we included the altitude (m.a.s.l.) of each sampling site. To obtain hydrogeological variables, we used a Swiss-wide map of the predominant groundwater aquifer types, which consist of karstic, fissured and unconsolidated aquifers (the latter being an abbreviation for aquifers in unconsolidated sediments) (BAFU, 2017). We then calculated the great-circle distance (m) from each sam- pling site to the closest karstic and unconsolidated aquifer borders. Distances of sites within the given aquifer were assigned a negative value, 12% of sites were located within karstic aquifers. For glaciation history, we computed two variables. First, we extracted the predicted glacier cover time for the last glacial cycle (120–0 ka BP) for each site. We used the numerical modelling predictions from Seguinot et al. (2018). Second, we used a field-based reconstruction (Bini et al., 2009) of the last glacial maximum (LGM, approx. 24 ka BP) and determined for each site whether it had been covered by ice or not at the LGM. Switzerland intersects several biogeographic regions including Jura mountains, central plateau and Alps (BAFU, 2022), and main river catchments, which likely affect species’ distributions and dispersal in addition to the before-mentioned environmental co- variates. For example, the Alps might function as a dispersal barrier, and consequently, different species assemblages are expected north and south of the Alps, even under similar environmental conditions.
Using a principal component analysis, we computed two continuous variables (to reduce the number of categorical variables for the subsequent occupancy modelling), based on the river catchments and biogeographic regions. As input, we used categorical data on the main river catchment (Aare, Adige, Inn, Limmat, Po, Reuss, Rhein or Rhone) and an approximation of the coordinates of each sampling site in respect to the Alps. For the latter, we created a new coordinate system, which was tilted 25° anticlockwise in comparison to the original projected coordinate system (CH1903/LV03), so that the x-axis was parallel and the y-axis orthogonal to the orientation of the Alps within Switzerland. Thus, it better approximated the biogeographic regions of Switzerland (BAFU, 2022) which are aligned in parallel to the Alps. We then applied a factor analysis of mixed data (FAMD), which is a principal component technique suitable for analysing datasets with qualitative and quantitative variables (Pagès, 2015). For each sampling site, we extracted the two corresponding FAMD axis values, based on the sites’ location in respect to the main river catchment and the coordinate system relative to the Alps. These two values per site were then used as covariates for the further occupancy modelling (hereafter referred to as ‘biogeography’), with axis 1 representing predominantly a west–east orientation along the Alps and axis 2 representing a south–north orientation relative to the Alps (Appendix S1, Figure S1.2).

For consecutive construction of continuous countrywide occurrence predictions, we extracted the same environmental variables based on a 1 x 1 km cell grid across Switzerland, using the centre of the cells.

### 2.4 Analysis

First, we calculated the geographic range for each species by using a cell size of 20 x 20 km, averaged over 100 different placements of the grid. The here chosen cell size is commonly used in European studies of subterranean diversity and thus allows direct comparability of groundwater amphipods relative to our study duration. All variables were standardized to have a mean of 0 and a standard deviation of 1. We applied a spatial explicit model and included the site coordinates (projected coordinate system in metres, CH1903/LV03). The spatial process was modelled using one latent spatial factor (to avoid overfitting), an exponential correlation function and a Nearest Neighbour Gaussian Process with 15 nearest neighbours (following recommendations by Datta et al., 2016; Doser et al., 2023). We specified relatively vague priors (Appendix S1; Doser et al., 2023), but restricted the spatial decay parameter to a minimum of 1 km to avoid overfitting based on small-scale spatial heterogeneity. We ran the model using Pólya-Gamma data augmentation, three chains of 120,000 iterations each, a burn-in period of 20,000 iterations and a thinning rate of 40. Additional details are provided in Appendix S1. Convergence of the chains was assessed by using the Gelman–Rubin diagnostic, the effective sample size and the trace plots of the posterior samples. To assess goodness of fit and any impacts from possible violations of underlying assumptions, we computed the Freeman–Tukey and chi-squared fit statistics as posterior predictive checks, grouping each fit statistic once by sites and once by replicates and assessed the Bayesian p-value. The posterior predictive checks did not indicate any lack of fit (Appendix S1).
To predict the occurrence of the 12 focal species, we used a 1\times1 km grid across Switzerland and predicted per-cell occurrence probabilities of each species and the local species richness per cell, resulting in spatially explicit maps with associated uncertainties. We validated the predicted species richness by comparing it to the observed species richness for all 1\times1 km grid cells that contained at least one sampling site (see Appendix S1 for results).

Based on the 1\times1 km predicted latent occurrence values for each species, we then inferred the mean regional species richness per 20\times20 km grid cell across all posterior samples.

In addition, we assessed the contribution of unmodelled, rarely found species to regional alpha and beta diversity (using the same 20\times20 km grid). We therefore included the occupancy model output of the commonly found species and computed the difference in diversity indices when raw occurrence data from unmodelled rarely found species were added. We used species richness per cell to calculate alpha diversity, and LCBD (local contribution to beta diversity; Legendre & De Cáceres, 2013) to calculate the uniqueness of each cell’s species composition from overall beta diversity.

Co-occurrences of commonly found species were investigated using two approaches. First, we used the raw detection–nondetection dataset based on our sampling data. Thereby, we aggregated the data by site and species. If a species was detected at least once at a given site, we classified it as present at that site and absent otherwise. All sites with at least one species present were included. Second, we used the output data from the occupancy model prediction. For each posterior sample, we computed the co-occurrence probabilities between all species pairs using the latent occurrence values (including all sites with at least one species present) and then averaged across all posterior samples. The frequencies of co-occurrences were calculated with the R package ‘cooccur’ (version 1.3; Griffith et al., 2016).

All statistical analyses for this study were conducted in R version 4.2.2 (R Core Team, 2022) and RStudio version 2022.07.2 (RStudio Team, 2022).

3 | RESULTS

3.1 | Citizen Science data

Overall, we received 3882 samples from 906 sites (Figure 1). In total, 2294 (59%) of the samples contained a wide range of macroinvertebrates (see Schneider et al., 2023), including obligate groundwater taxa (stygobites), taxa that can inhabit groundwater environments for part of their life cycle (stygophiles) and organisms from adjacent soil and surface water habitats (stygoxenes). A total of 2350 obligate groundwater amphipods were found at 298 sites (33% of the sampling sites).

![Map of Switzerland](https://onlinelibrary.wiley.com/doi/10.1111/jbi.14975)
3.2 | Groundwater amphipod diversity

We barcoded 2100 specimens, of which 1954 specimens (93%) had an adequate sequence quality and could be assigned a species. We identified 27 different groundwater amphipod species (Figure 2a). Among them, 26 belonged to the genus Niphargus and one to Crangonyx. Some species were detected frequently, such as the most common species Niphargus luchhoffmanni, of which 562 individuals were collected from 88 sampling sites. A large portion of species, however, was detected at a very limited number of sites only, spanning very small geographical ranges (Figure 2a). Despite having sampled more than 900 sites in total, almost half of the species were detected at only 1–3 sites (i.e. <0.5% of all sampled sites) and 37% of the species were found at a single site only. Among these rarely found species, a high proportion consisted of species that have been found for the first time and that are not formally described yet. Contrastingly, most of the commonly found species are formally described or represent well-known phylogenetic lineages (belonging to the Niphargus rhenorhodanensis complex, see Delić et al., 2023; Lefébure et al., 2007).

The species accumulation curve did not reach an asymptote when all species were included (Figure 2b), indicating that even with an unprecedented density in sampling efforts, more diversity remains to be discovered. This finding was supported by the computed species richness estimators (Appendix S1, Table S1.1). The accumulation curve for ‘uniques’ (species found at a single site only) increased linearly with a slope similar to the accumulation curve built from the entire dataset (after ca 300 sampling sites).

3.3 | Occupancy modelling

3.3.1 | Sampling effects on detection probability

Based on the occupancy model results, we found a substantial variation in the detection probabilities between species (Appendix S1, Table S1.2 and Figure S1.2). The highest detection probability was found for N. luchhoffmanni (0.41±0.08 for one filternet sample) and the lowest for Crangonyx cf. subterraneus (0.023±0.017 for one filternet sample). On the community level, the handnet method...
resulted in significantly lower detection probabilities compared to the full method including one or more filternets (Appendix S1, Table S1.2). Crangonyx cf. subterraneus was the only species for which the handnet method was predicted to increase the detection probability (although not statistically significant). This might be related to its small body size (Studer et al., 2022); it might have been easier to capture and to visually detect the C. cf. subterraneus specimens with the hand net because of the slightly narrower mesh size and the darker mesh colour compared to the filter net. The number of subsamples increased the detection probability on the community level significantly, with relatively small variation among species (Appendix S1, Figure S1.2).

3.3.2 | Environmental effects on species occurrence

The number of occurrence records varied for the 12 focal species from 5 (Niphargus rufofi) to 88 sampling sites (N. luchosofmanni) with a mean of 31 ± 28. The spatial factor multispecies occupancy model predicted different effects of the environmental variables on species' occurrence probabilities (Figure 3 and Appendix S1, Table S1.2). Altitude and its quadratic term had a limited number of significant effects, but indicated a possible nonlinear relationship for most of the species, with highest occurrence probabilities predicted at lower elevations or peaking midrange. The estimated effect of distance to karst was overall stronger than the effect of distance to unconsolidated aquifers. Some species, such as N. luchosofmanni and N. rhenorhodanensis lineage FG, were predicted to have higher occupancy probabilities in areas inside or close to karst, while other species (e.g. N. auerbachii, N. tonywhitteni) were predicted the opposite. Late Pleistocene glaciation cover time did not have a significant effect on any of the species due to large confidence intervals, but tendencies towards positive or negative effects were visible for most of the species. LGM glaciation had a significant effect on one species (Niphargus fontanus), predicting higher occurrence probabilities at locations that remained ice-free during the LGM. Among all environmental parameters included in the model, the two biogeographic parameters had the strongest effects on the occurrence probabilities. The direction and magnitude of these effects varied strongly between species, but there was an overall trend of negative estimates for the first biogeography axis (higher occurrence probabilities towards central and western Switzerland) and positive estimates for the second biogeography axes (higher occurrence probabilities towards northern Switzerland and north of the Alps). The modelled species had distinct and different predicted distribution patterns, linked to geology, elevation and ice coverage during the last glaciation cycle (Figure 4). For a discussion of species-level results, see Appendix S1. Residual spatial correlation was estimated to drop to 0.05 at a distance of 46 km between sampling sites, which might indicate a residual spatial pattern. However, we refrain from further interpretation, as the spatial decay parameter was only weakly identifiable.

3.3.3 | Species richness patterns across space

The 12 commonly found focal species included in the occupancy model could be used to predict species richness across the study area. Overall, the predicted local species richness was low, with 0.2–3.3 (mean 0.9) predicted species per 1 × 1 km cell, despite having included 12 species in the model. The highest species richness was predicted in the region along the western border of Switzerland (Jura) and around the catchment of the river Töss in the eastern Swiss plateau (Figure 5a, associated standard deviations: Appendix S1, Figure S1.5). Interestingly, these two regions had contrasting spatial sampling patterns: The Töss catchment was sampled more densely than the Jura, indicating that sampling density alone did not drive the pattern of high local richness. Regional species richness based on 20 × 20 km cells indicated a similar pattern, with higher species richness of commonly found species in the northern and northwestern parts of Switzerland, in comparison to the southern, Alpine areas (Figure 5b).

Another component of species diversity consisted of species only found at a very limited number of sites and small geographical ranges each (generally less than 0.5% of sites sampled). They could not be included in the occupancy modelling and spatial species richness estimation due to low occurrence. These rarefied species displayed a scattered distribution across the study region and were detected in different regions and biogeographic areas (Figure 5c). The occurrence of rare species did not depend on local species richness of more commonly found species (Figures 5a–c and Appendix S1, Figure S1.6). Also, the rarely found species contributed up to 43% to regional alpha diversity of 20 × 20 km cells (Figure 5d) and increased the uniqueness of cells' species compositions from overall beta diversity up to 28% (Appendix S1, Figure S1.7).

3.4 | Species co-occurrences

In the co-occurrence matrix based on the occupancy model prediction, many pairs of the 12 focal species showed negative co-occurrence estimates (Figure 6a). For two species, namely Niphargus thienemanni and Niphargus arolaensis, exclusively negative or random co-occurrences were predicted. Between species pairs of C. cf. subterraneus, N. fontanus, Niphargus tonywhitteni, Niphargus auerbachii and Niphargus puteanus predominantly positive co-occurrences were predicted. A smaller number of positive and negative co-occurrences could be identified when using the raw sampling data (Figure 6b), because of the limited numbers of presence sites for each species and of sites where multiple species were found. Nevertheless, the two species co-occurrence models produced concordant results.

4 | DISCUSSION

Using a highly replicated and standardized citizen science approach, we conducted one of the most systematic assessments on the
FIGURE 3 Parameter estimates from the occurrence part of the spatial factor multispecies occupancy model. Points indicate the mean of the posterior samples with 50% (thick line) and 95% (thin line) confidence intervals.
diversity and distribution of groundwater organisms, with a focus on amphipods. Our study revealed a remarkable and substantially higher species diversity than previously known for the area. Nearly half of all species were detected at less than 0.5% of sampled sites each, highlighting the significance of rare species in groundwater biodiversity studies. Even the more commonly found species exhibited limited co-occurrences, indicating low local richness and high spatial turnover. These findings underscore the importance of highly resolved, systematic data for groundwater biodiversity assessments from regional to global scales, especially given the increasing anthropogenic threats on groundwater ecosystems such as warming, pollution and over-extraction.

**Figure 4** Model predictions for the occurrence probability of each focal species across Switzerland. Sites where the species were detected are marked in orange and remaining sampling sites are marked in black. Lakes are shown in white. Standard deviations are provided in Appendix S1, Figure S1.3.
4.1 | Remarkable and substantially higher subterranean biodiversity than previously known for the area

In total, 27 species (some not yet formally described) were found in this study. As such, subterranean amphipod diversity by far exceeds the native amphipod diversity known for surface freshwaters in Switzerland (see Altermatt et al., 2019). Despite the high sampling effort, with more than 900 sites systematically and repeatedly sampled, the species accumulation curve and species richness estimates indicate an even higher number of species occurring in Switzerland than already found, with some species remaining to be discovered. This pattern is likely driven by the high number of uniques (species that were only found at a single location) as the slopes of the accumulation curves including all species and including only ‘focal’ species were highly comparable. Only a decade ago, the total number of known groundwater amphipod species still remain undiscovered and undescribed (Borko et al., 2021; Borko et al., 2022).

4.2 | Rare species remain challenging

While we established detailed and well-resolved species distribution models for commonly found species, providing a comprehensive understanding of their distribution and diversity patterns across Switzerland, many rarely found species could not yet be covered in our analysis. As a result, distribution predictions for over 50% of all species (likely more, considering the presence of additional undetected rare species) were not possible, despite the rarely found species contributing substantially to regional (20×20km) alpha diversity and community uniqueness patterns. Thus, the groundwater fauna is twofold in the dark from a conservation and biodiversity perspective: First, these rare species were found too infrequently and/or too locally restricted to be integrated into species distribution models (see also Jeliazkov et al., 2022). Second, in many cases, their phylogenetic species status is not yet resolved and many species remain to be formally described. These two aspects, and given the limited groundwater accessibility, result in a substantial portion of the groundwater biodiversity remaining...
Species co-occurrence matrices of the 12 focal species based on (a) 1 × 1 km latent occurrence values predicted by the occupancy model and (b) raw sampling data.

undetected, undescribed and as a consequence, not integrated in national conservation and management strategies.

That some species are rare while others are common is a well-known phenomenon, discussed already in pioneering works by ecologists and biogeographers (e.g. Hanski, 1982; MacArthur & Wilson, 1967; Preston, 1948). Yet, there is evidence that, in groundwater systems, species rarity may be particularly pronounced because of the poor dispersal ability and highly disjunct distributions (Foulquier et al., 2008; Gibert & Deharveng, 2002). Studies from other European regions found comparably high levels of narrowly distributed or even single-locality endemics and low occurrence frequencies of many stygobitic species (Bregović et al., 2019; Dole-Olivier, Castellarini, et al., 2009; Hahn & Fuchs, 2009; Martin et al., 2009). Species ranges extending >200 km are rare and have been largely attributed to the presence of cryptic lineages (Trontelj et al., 2009). However, it is often difficult to assess if rarity patterns in the observed species (in terms of range size, abundance and occurrence; Crisfield et al., 2024) have arisen from ecological drivers or by sampling design. A major part of today’s knowledge on groundwater fauna is based on opportunistic sampling data (e.g. Altermatt et al., 2014). While various sampling methods and specific habitats can be covered, opportunistic data tend to lack robust effort quantification and species absence records. Thus, it remains unresolved, if species are found infrequently due to insufficient sampling effort or due to ‘true’ ecological rarity (Kéry & Royle, 2016). We here present a complementary approach. The filtration methodology in spring catchment boxes enabled the quantification and standardization of the sampling effort, making samples highly comparable. At the same time, species from ecosystems like caves or hyporheic zones may only sporadically appear in drinking water samples, adding to the count of ‘uniques’. For example, Niphargus virei and Niphargus brixianus, previously documented in multiple caves (Altermatt et al., 2019), were each found at only one site in this study. We acknowledge that our sampling method introduces a bias, but it is distinct from the one of other sampling approaches. Interestingly, and despite using a unique approach, we did find groundwater fauna occurrence patterns that are highly comparable to previous studies (e.g. Castellarini et al., 2007; Hahn & Fuchs, 2009; Martin et al., 2009). While no single approach can eliminate all sampling biases, combining data collected through different (systematic and opportunistic) sampling methodologies might be the most effective way forward.

4.3 | Co-occurrence patterns

While subterranean ecosystems typically show reduced variability in abiotic conditions (Mammola, 2019; Sánchez-Fernández et al., 2018), we found strikingly dense patterns of groundwater amphipod occurrence (amphipods expected for nearly all 1 × 1 km cells considered), yet consistently low local species richness (typically around one species per cell, despite a high gamma diversity) and high turnover. Our findings indicate possible mechanisms driving the observed patterns. Competition for limited resources may hinder the co-occurrence of multiple groundwater amphipod species. In particular, if resources are scarce and habitats already occupied by one species, the establishment and persistence of additional species may be impeded when they are covering similar niches. From cave ecosystems, it is known that amphipods exhibit variations in morphology and trophic niche, consequently facilitating local co-occurrence (Premate et al., 2021). However, we speculate that amphipods from the here sampled shallow groundwater aquifers might be less ecologically differentiated, resulting in higher niche overlap and competitive exclusion on a local scale (see also Fiser et al., 2015).

Alternatively, the limited dispersal and low colonization rates of groundwater amphipods in these highly fragmented ecosystems may restrict distribution ranges and co-occurrence. Exploring the relative influence of random processes (Hubbell, 2001) and species-specific ecological requirements would help to better understand assembly mechanisms shaping the biodiversity patterns in subterranean ecosystems.
4.4 Opportunities for stygofauna research and conservation

There is an increasing call towards continental and global wide assessments of groundwater biodiversity (Zagmajster et al., 2018). We here show how such data can be acquired at a highly resolved scale, and how the understanding of the robustness of the emerging patterns is a prerequisite for such objectives. In the light of the high portion of narrowly distributed species, systematically collected detection–nondetection datasets are of paramount importance. Our dataset allowed drawing inferences also from absences of species, which can be helpful for assessing rare and endemic species. Our findings have important implications for the conservation of groundwater biodiversity. Since stygobites from a point-based sample (such as a spring) inhabit the underlying aquifer, the hydrological dynamics of that ecosystem, for example, the groundwater flow, connectivity and extent of the recharge area need to be considered (Aldous & Gannett, 2021). Conservation efforts focusing on the local extent of the aquifer (instead of the point-based occurrence site) might be effective for the protected target of narrowly distributed or even single-site endemics. However, due to the high species turnover across space and limited local co-occurrences, localized efforts would fail to integrate the overall biodiversity of a larger area. Consequently, there is a need for spatially inclusive conservation strategies on a larger geographical scale. Especially in times of global change, adequate management and conservation strategies of groundwater ecosystems are needed. As groundwater faces increasing threats from warming, pollution and over-extraction (Mammola et al., 2019), our findings underscore the imperative for systematic and extensive research to guide conservation efforts.

ACKNOWLEDGEMENTS

We thank all participating water providers for collecting data for this project and showing a lot of interest and enthusiasm. Many thanks to Nicole Bongni, Angela Studer, Anouk Petitpierre, Ramona Petrig and Jessica Colombo for their dedicated support in data acquisition. We would also like to thank Marjorie Couton for inputs on data analysis and manuscript writing, the members of the SubBioLab for collaboration and support concerning species identification and molecular work and Stephen B. Heard for advice on manuscript focus and writing. Finally, we are thankful for constructive feedback by Diana Galassi and two anonymous reviewers on a previous version of the manuscript.

FUNDING INFORMATION

Funding was provided by the Swiss Federal Office for the Environment FOEN/BAFU (project ‘AmphiWell’ to FA and RA) and the Swiss National Science Foundation (grants nr. PP00P3_150698 and 31BD30_209583/Biodiversa DarCo to FA) and the University of Zurich Research Priority Programme on Global Change and Biodiversity (URPP GCB to FA). CF was funded by the Slovenian Agency for Research and Innovation through the core program P1-0184 and project J1-2464. No permits were required for this study.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article and on Zenodo (https://doi.org/10.5281/zenodo.12570591). Due to the sensitivity of drinking water well locations and corresponding data protection laws, sampling coordinates will not be provided. Sequence data is uploaded to GenBank.

ORCID

Mara Knüsel https://orcid.org/0000-0006-2346-984X
Roman Alther https://orcid.org/0000-0001-7582-3966
Arpat Ozgul https://orcid.org/0000-0001-7477-2642
Cene Fišer https://orcid.org/0000-0003-1982-8724
Florian Altermatt https://orcid.org/0000-0002-4831-6958

REFERENCES


BIOSKETCH

Mara Knüsel is broadly interested in community ecology and how species assemble across time and space. This study represents a component of her PhD work at the University of Zurich and Eawag. She and the other authors collaborate on fundamental and applied questions in ecology and biodiversity (see Altermatt lab, https://www.altermattlab.ch/).

Author contributions: MK, RA and FA conceived the ideas; MK conducted the fieldwork and collected the data with additional support from collaborators; NL conducted the laboratory work; MK analysed the data and led the writing with assistance from RA, FA, AO and CF. All authors contributed with comments and editing to the manuscript and gave final approval for publication.

SUPPORTING INFORMATION

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