turnover in evolution [13], through increased speciation being balanced by the greater extinction risk of narrow niches [12]. It is however becoming evident that symbiosis, both mutualistic and antagonistic, may tip the scales of niche evolution [14,15]. Takeover of a functional role by a second symbiont can lead either to narrower niches if genes are lost in the first symbiont owing to relaxed selection [16,17], or to ecological range expansion if a symbiont switch brings new functionalities [18]. That closely related fungi can consort with different species of alga [19] suggests that in lichens, at least, switches are common over evolution. Any given lichen symbiont pedigree may have been associated with different symbionts, or different numbers of symbionts, over evolutionary timescales, an attribute that makes them attractive systems in which to study the effects of symbiosis on niche.

The last decades have seen at least three major changes to our understanding of lichens that frame how we assess niche breadth and the symbiotic relationships that potentially affect it. The first change concerns the circumscription of the species itself. Historically, lichens were classified using a mixture of traits assumed only in symbiosis and other, purely fungal characteristics such as spore size; the totality was called a lichen species. Assumed evolutionary groupings, such as the genus *Lecidea*, included dozens of interdigitated species that were specialized for rock, bark or wood [20–22]. In 1950, in order to rectify the nomenclatural instability arising from the recognition of lichens as multi-domain symbioses, the name of a lichen was anchored to that of its presumed single fungus by a change to the code of nomenclature [23]. Molecular phylogenetic studies of the fungus have since resulted in drastically changed species circumscriptions, with some species split into many narrower 'cryptic species' and others more broadly delimited, with downstream consequences for (re-)assessing niche breadth [24,25]. The second change is another by-product of fungal molecular phylogenetics: species once thought closely related have often turned out not to be [21]. We now know that the rock-dwelling species of *Lecidea*, for instance, are only distantly related to those found on bark and wood, placed in other genera and families altogether [22,26]. The third change concerns the nature of the symbioses themselves. Lichens were long thought of as a neat twosome of a fungus and alga, but metagenomic data are unearthing evidence of additional constituent fungi, of yet unknown function [27] as well as suites of algal lineages, rather than a single alga, present in common lichens [28,29]. Evidence is likewise building that bacterial assemblages influence lichen symbioses [30].

The extraordinary substrate specificity of many lichens raises intriguing questions about how the range of substrate use evolved and under what circumstances it switches. In the present study, we ask four specific questions about niche breadth evolution in the constituent fungus in an ancient and taxonomically well-studied group of crustose lichens: (i) is substrate affinity phylogenetically conserved in the constituent fungus? (ii) what are ancestral substrate types in this group? (iii) is there evidence for specialists evolving from generalists or vice versa? and (iv) how does specialization correlate with patterns of speciation and extinction? Our study group represents a cross-section of different kinds of substrate specificity and niche breadth and may serve as a good test case for evolutionary niche studies in a lichen symbiont.

2. Material and methods

(a) Study system

We focused on lichens formed by members of the ascomycete families Trapeliaceae and Xylographaceae (hereafter: trapelioid lichens). The constituent fungi of trapelioid lichens form a monophyletic group, which has been well studied from taxonomic and phylogenetic perspectives [6,31–34]. Trapelioid fungi began diversifying about 150–200 Ma BP [35]. The lichens in which they occur are exclusively crust-forming and establish physical bonds with a wide range of mineral and organic substrates. They can occur on multiple (generalist) or only one (specialist) substrate type, which can be carbohydrate-rich (e.g. wood and bark) or carbohydrate-poor (rock).

Our taxon set consists mostly of specimens and sequences published by Resl *et al*. [32] and Schneider *et al*. [33], augmented with some new data (electronic supplementary material, table S1). In *Placopsis*, we have considered as separate species the operational taxonomic units estimated by Schneider *et al*. [33], although they have yet to be formally described as species. Sequences were generated following methods and primers described in [32].

(b) Estimating taxon sampling completeness

To account for bias in species capture, we estimated the number of known species in each group using one of the largest databases for fungal taxonomy, Index Fungorum (www.indexfungorum.org; accessed January 2018). We checked every trapelioid genus except the recently described *Ducatina* and recorded the total number of described species. Additionally, taxonomic and evolutionary knowledge on trapelioid lichens accumulated over the years [6,31–34] allows us to estimate the expected total species number of the group with confidence. We then calculated per cent ratios of total known versus included species per genus in our dataset (electronic supplementary material, figure S1). Whenever possible, we performed analyses under multiple sampling regimes.

(c) Chronogram estimation

We assembled a dataset of eight fungal loci including mitochondrial ribosomal (mtSSU), nuclear ribosomal (ITS, SSU, LSU) and nuclear protein-coding genes (RPB1, RPB2, MCM7 and EF1α; abbreviations following [32]). DNA isolation, polymerase chain reaction and Sanger sequencing were performed as in [32] and [33]. Alignments were generated for each locus using MAFFT [36] following our *phylo-scripts* pipeline [32,37]. Using BEAST 2.2.4 [38], we estimated time-calibrated phylogenies for the concatenated dataset using locus-independent site and clock models and a birth-death tree prior. We chose the best substitution models according to the Akaike information criterion (AIC) for each locus with JModelTest 2 [39].

(d) Tree selection and phylogenetic uncertainty

For downstream analyses, we consistently used either (i) a random subset of 100 trees selected from the BEAST posterior distribution for analyses using multiple trees to account for phylogenetic uncertainty, or (ii) a maximum clade credibility (MCC) tree when only single topologies could be used (Bayesian analysis of macroevolutionary mixtures (BAMM)). The MCC topology was estimated in TreeAnnotator 2.2.1 after discarding the first 15% of trees as burn-in.

(e) Coding ecological and substrate preference characters

We coded ecological strategies as two sets of categorical variables. Specialization (GS) was treated as binary (generalist:

growing on multiple substrates; specialist: growing on single substrate), while the preferred substrate (PS) was coded as multi-state (rock, soil, bark, wood, other lichens). We derived substrate use data from our own collections as well as from herbarium collections (BG, GZU, UPS), species catalogues [40], identification keys [4,5] and recent monographs (*Placopsis*: [41], *Xylographa*: [31]). A fungal-species was considered a specialist when greater than 95% of its global occurrences were from one substrate.

(f) Testing for phylogenetic signal

We estimated phylogenetic signal of the PS variable using two simulation-based multi-tree approaches: (i) recursive use of Pagel's λ [42], and (ii) comparison of the distribution of cophenetic distances based on the assumption that closely related species are ecologically similar [43,44]. For further details, see the electronic supplementary material.

(g) Ancestral state reconstruction

We employed two maximum-likelihood (ML) approaches based on implementations in ape [45] and corHMM [46] and stochastic character mapping implemented in phytools [47] to reconstruct ancestral states of the GS and PS characters at the main 19 internal nodes of the Trapeliales phylogeny (figure 1). To provide a summary of ancestral character reconstruction while accounting for bias introduced by methods, models and tree topologies, we developed a recursive strategy. First, we fitted models with all possible parameter combinations available for each method to each tree. Then we only considered as the most probable ancestral state the one recovered most often across all 24 analyses and tree topologies which are shown with the MCC tree. For each node, we created a plot indicating the number of trees for which a particular ancestral character state was estimated under all possible models for one method.

(h) Reconstructing transitions between substrates

Transitions between the different character states were counted from unconstrained stochastic character mappings as created for ancestral state reconstructions (see above). The cumulative results of the 10 000 alternative transition histories were summarized numerically and are presented as histograms for binary ecological strategy characters (GS) and as circle plots for multi-state substrate characters (PS).

(i) Testing substrate 'no-switch' scenarios

To test whether models prohibiting certain substrate transitions are more likely given our set of trees, we created 30 transition rate matrices describing different scenarios of character change. We compared these constrained models on each of the 100 trees from the BEAST posterior distribution of trees. The tested models include all possible combinations of no-switch scenarios for our multi-state substrate character. Each model and each tree was subjected to a ML ancestral state estimation using ape [45]. We then calculated and ranked models from best to worst according to AIC score comparisons for each tree. To see which models scored best over all trees, we calculated for how many trees a specific model would be the best, second best, third best and so on. We then searched for the models for which the majority of trees were recovered in the first five ranks.

(j) Modelling of diversification rates

We characterized the diversification dynamics of the trapelioid clade with character-independent BAMM 2.6.0 [48] on the MCC tree topology as well as with character-dependent multi-state speciation and extinction (MuSSE) [49,50] models on a set of 100 tree topologies (see above). We analysed the output of the BAMM analyses with BAMMTOOLS [51]. We combined the posterior samples from all MuSSE runs and created density plots for diversification rate (speciation–extinction). To identify significantly different speciation rates, we compared the obtained probability distributions with the Mann–Whitney tests for all possible combinations of characters.

Owing to the lack of consensus on the performance and suitability of the SSE approach to model evolutionary trends [52], we tested the extent to which the modelled diversification rates respond to the phylogenetic tree alone without a further connection to the distribution of characters on the tree according to the method described in [52].

3. Results

(a) Phylogenetic reconstruction confirms previous results

Our phylogenetic results (figure 1*a*) recover the same relationships found in previous studies ([32]; fig. 4) and confirm the recently recognized two-family split between Trapeliaceae and Xylographaceae [32]. We could also confirm the paraphyly of *Trapelia* and *Placopsis* [32,33] and the monophyly of all other genera. A table with all used sequences is provided in the electronic supplementary material, table S1.

(b) Substrate association displays strong phylogenetic signal

The distribution of substrate characters displays strong phylogenetic signal according to both simulation approaches. Model fit was significantly better for the real character data compared to all randomizations under multiple scenarios of Pagel's $\lambda > 0$ (electronic supplementary material, figure S2; $p < 0.05$) except when $\lambda = 0$ and the tree is one single polytomy (electronic supplementary material, figure S2). The mean tip-to-tip distance method yielded similar results (electronic supplementary material, figure S3). In greater than 95% of simulations, the mean tip-to-tip distance between tips with the same character coding was significantly shorter for the real character distribution (electronic supplementary material, figure S3; $p < 0.05$) compared to randomizations. For bark-growing species, the mean distance of the real distribution was significantly shorter in 75 of the simulations ($p < 0.05$; electronic supplementary material, figure S3), which probably referred to the low number of tips with that character state.

(c) Ancestral substrate use and amplitude

We recovered evidence for ancestral ecological strategy and preferred substrate use of 19 nodes representing all currently recognized trapelioid genera, as well as important nodes of the tree backbone (figure 1*a*). Our approach is based on three methods imposing 30 and 20 different models for the preferred substrate and ecological strategy characters, respectively (electronic supplementary material, figures S4–S117). How many methods recovered which ancestral states in figure 1*a* are given in the electronic supplementary material, table S6. For all extant species groups, we recovered the currently preferred substrate as its ancestral substrate. We estimated rock as the ancestral substrate for *Placopsis* (node 1; 29 out of 30 methods), *Trapelia* (node 2; 29 out of 30 methods), *Rimularia* (node 9; 29