CrossMark

ORIGINAL PAPER

Conservation of aquatic insect species across a protected area network: null model reveals shortfalls of biogeographical knowledge

Jason L. Robinson[1] · James A. Fordyce[2] · Charles R. Parker[3]

Abstract The effective conservation of species requires some understanding of where populations occur in a landscape. Gaps in this knowledge base (the "Wallacean Shortfall" of some authors) may coincide with hotspots of diversity for different plant and animal species, requiring the cooperation of a number of different federal, state, local and non-governmental agencies for effective conservation. In this example, the distribution and abundance of benthic macroinvertebrates are widely used as metrics for water quality monitoring, but far less is known about these organisms *qua* species (taxonomic orders EPT—Ephemeroptera, Plecoptera and Trichoptera). In this study, we inventoried a network of individual US National Park units for species in these orders. These parks are located in geological, ecological and historical places of interest across the states of Alabama, Georgia, Kentucky, North Carolina, South Carolina, Tennessee and Virginia. We sampled these parks in a multi-year intensive inventory in order to determine the composition of the aquatic insect fauna in each park. Since there are no comprehensive accounts of the geographic ranges of these species, we compiled published accounts of species occurrences in these and adjacent states (Arkansas, Florida, Louisiana, Mississippi, West Virginia) to construct a potential species pool for each state. This pool comprised our best estimate of the EPT species that might potentially occur in each state. We used these source pools to test null hypotheses on whether parks disproportionately under- or over-protect species in different categories of risk of imperilment. We find that parks have fewer rare (G1) species than expected from a null model, and parks over-protect some of the most common (G5) species in the network. This pattern would be expected if the actual landscape distributions of the most imperiled (G1) species are small and/or disjunct and tend to occur outside of the national parks in the region. Interactions between park shape (and size) and individual species geographic ranges are likely to influence the precision of estimates of the potential species pool within a protected area. More research is needed on the distribution of imperiled species across the entire geographic range of species, and the traditional practice of compilation and reporting of occurrence records by state is not sufficient for informed conservation practice. State natural heritage programs and biodiversity conservation database efforts (e.g. NatureServe) implicitly recognize the importance of species ranges, but our analysis demonstrates the need to assess these patterns at a finer spatial grain in order for these state lists to serve as meaningful expectations of the composition of species assemblages. Our analysis considers only a tiny fraction of the protected lands in the region, and an enormous additional area of protected lands exists where many of these rare species occur. More precise and accurate reporting of EPT species occurrences in this region will allow resource managers to target the

✉ Jason L. Robinson
jrob@illinois.edu

James A. Fordyce
fordyce@utk.edu

Charles R. Parker
cddis@att.net

[1] Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL 61820, USA

[2] Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA

[3] Sevierville, TN, USA

conservation of particular species within single parks, or across protected area networks.

Introduction

In theory, the successful sustained conservation of aquatic biodiversity will require the perpetual systematic protection of large-scale (e.g. watershed) dynamic patterns and processes, demanding strategies above and beyond those aimed at protecting single sites, individuals or populations of a species. In practice, our efforts to conserve large taxonomic groups of animals (e.g., invertebrates) are still limited by gaps in our knowledge of the existence, ecology and biogeography of individual species (Margules and Pressey 2000; Cardoso et al. 2011; Hortal et al. 2015). Through enormous effort to accumulate target field collections by taxonomic experts and captured data from regional museum and institutional collections, simple models or predictions of species distributions can estimate partial geographic distributions of aquatic insect species (DeWalt et al. 2012, 2013; Cao et al. 2013), but these models have not yet been built for the full geographic range of species in North America. In species distribution models, the mismatch between the spatial grain of reporting records and the spatial grain (at which "ecology" and "biogeography" or other phenomena) occur has a potential to bias quantitative measures of range size and other macroecological parameters (Cao et al. 2013). Since the EPT insect orders constitute a large fraction of the target taxa in benthic macroinvertebrate water quality assessments (Kenney et al. 2009), more finely tuned expectations of the regional species pool might improve the accuracy of estimates of stream condition at sites based on the expected species assemblage.

The large area and variety of conservation lands administered across North America span many political and ecological boundaries, creating excellent opportunities for cooperative efforts on specific conservation goals across protected area networks (PANs).

When individual management units embedded within a PAN share administrative hierarchies (e.g. the US National Park Service), resource managers can coordinate conservation and management objectives across large geographic areas, perhaps even larger than the geographic range of many species within the network. In such a network, systematic inventories of protected area units generate knowledge applicable to the performance of the entire network (e.g. Nichols and Langdon 2007; Parker et al. 2007).

Popular measures of the conservation performance of a PAN are derived from the composition of species assemblages within the units of the network. Within each assemblage, some species can have higher or lower conservation significance than others (e.g. some are locally rare or in low abundance), so the distribution of rarity across these assemblages is a second common measure of PAN conservation performance (Cabeza 2013). Conservation rankings are a standardized tool to organize the geographic distribution of species and the faunal expertise of specialists to estimate the extinction risk (imperilment) faced by plant and animal species. For example, NatureServe (2015) conservation rankings are made on the basis of habitat specificity, number or size of populations, connectivity among populations, sub-species designations and other biological or systematic criteria (NatureServe 2015). At the global level, a given species may be relatively common across its geographic range and therefore be ranked at a low risk of imperilment (e.g., G5, the least imperiled category). However, within a state where there are only a few populations (perhaps at the margin of this species range) the same species could be ranked by that state as most imperiled (S1). These methods of prioritizing distributional data are descended from a long history of "gap analyses" that assess the geographic intersection of stacked layers of predicted species occurrences or distributions and the boundaries of geographic regions or protected areas (Brooks et al. 2004).

In this study we summarize a multi-year, multi-seasonal inventory of EPT (insects in taxonomic orders Ephemeroptera, Plecoptera and Trichoptera) species in seventeen national parks distributed across the southern highlands of the USA, widely known as a hotspot of biological diversity for EPT and many other aquatic organisms (Lydeard and Mayden 1995; Morse et al. 1993, 1997). Since parks vary in size and environmental conditions, we ordinated the parks onto dominant geographic and climatic gradients to explore patterns of species and environmental diversity among parks. We use these data, and estimates of the regional species pool for each national park, to construct a null model that can answer the following sets of questions:

Q1. Are there significant species-area relationships among assemblages in parks? What about the species pools of EPT assemblages in the states?
Q2. Does species richness vary with spatial metrics of park configurations?
Q3. How well do climate similarity and geographic distance predict the rate of turnover of species between parks (or states)?
Q4. How does the prevalence of rarity vary among orders in the state lists?
Q5. Are aquatic insect assemblages in national parks comprised of species at a greater risk of imperilment

than would be expected by chance alone (e.g. do parks differentially protect rare species?).

Methods

Occurrence records

Field sampling

Over a 3-year period we sampled aquatic insect communities from lentic, lotic and madicolous habitats in sixteen southeastern national parks (Fig. 1). We collected immature aquatic insects by hand, kick nets and seines. Adult insects were collected with black lights, beat sheets, aerial nets, and by rearing live larvae and pupae in the laboratory via the metamorphotype method (Etnier et al. 2010). We did not systematically sample the seventeenth national park (Great Smoky Mountains National Park (GRSM) in North Carolina and Tennessee) but used previously published results for a comprehensive list of all species occurrence records from GRSM (Parker et al. 2007; unpublished data).

Site selection

We located sampling locations from maps, prior collections, Park Service staff, literature records and by exploration of the parks by the investigators. Many parks in this study are small (median area = 21 km^2) and are limited in available aquatic habitats. Our aim was to representatively sample the diversity of the available habitat types within the park, but the occurrence of those habitat types varied a great deal across the PAN units. For example, the Blue Ridge Parkway is an extremely long transect across many headwater streams at high elevations. Similarly, large parks that capture many large watersheds, or parks arranged along a single watershed, can capture many different habitat types. In order to completely characterize the assemblages in these habitats, we selected sites that we could visit multiple times in different seasons. Our sampling efforts were designed as an attempt to inventory all EPT species present in each park; larger parks with more sampling sites were necessarily sampled more frequently and intensely than smaller parks with fewer total habitats.

Specimen identification

For this study, we only include specimens identified to species, relying on the expertise of outside taxonomic experts for some Plecoptera and Ephemeroptera identifications (BK, ED, LJ; see acknowledgments). Some species records we obtained from a DNA barcoding program that associated sequences from immature specimens with sequences from confirmed adult identifications (Zhou et al. 2011). We obtained global conservation rankings for EPT species using NatureServe Explorer (http://www.natureserve.org/explorer/), where species are ranked on a spectrum of very rare (G1) to very common (G5) (NatureServe 2015, accessed for this analysis 11/1/2011). At the time of this analysis, a small number of EPT species were not ranked by NatureServe (33 of 656; see Supplementary Data): these species we assigned ranks of G5, a conservative estimate since no better data are available. All 33 of these post-ranked species are known from Great Smoky Mountains National Park, but none these 33 species were collected during our surveys of the remaining 16 parks.

Sampling effort in biodiversity inventory

The aquatic insect fauna of the Great Smoky Mountains National Park is particularly well studied, and prior records have previously been compiled for all three orders of aquatic insects considered here (Parker et al. 2007). The area of this park is several orders of magnitude larger than several of the other parks and has been the target of long term sampling efforts for many plant and animal groups (e.g. Sharkey 2001; Nichols and Langdon 2007). These species occurrence records are derived from many different collection events, made under differing environmental conditions and by multiple investigators and using a range of types of collecting gear and methods. The complexity of these interacting factors precludes a simple estimate of the actual effort applied to each park or even to a single site. Blacklight traps may yield tens of thousands of adult insects in the span of a few hours, but total catch can vary with local environmental conditions. Hand collecting by careful and diligent manual searching of some habitats might yield only a few dozen specimens, but of species not captured in other habitats. As a proxy for the sampling effort applied to each system, we calculated the total number of individual specimens examined from each park during all collection efforts. To test whether the relative over-sampling of GRSM biased our analyses, we removed GRSM records from the assemblage data and analyzed these reduced datasets.

Effective sampling effort

Since the effort applied to sites within parks often varied with weather, season, and available personnel available, direct estimates of effort among sample sites are problematic. Additionally, blacklight traps actively attract insects and (as mentioned above) offer many potential confounders to any direct comparisons of the catches among individual trapping occasions within a park. Our sampling approach was designed to maximize our ability to
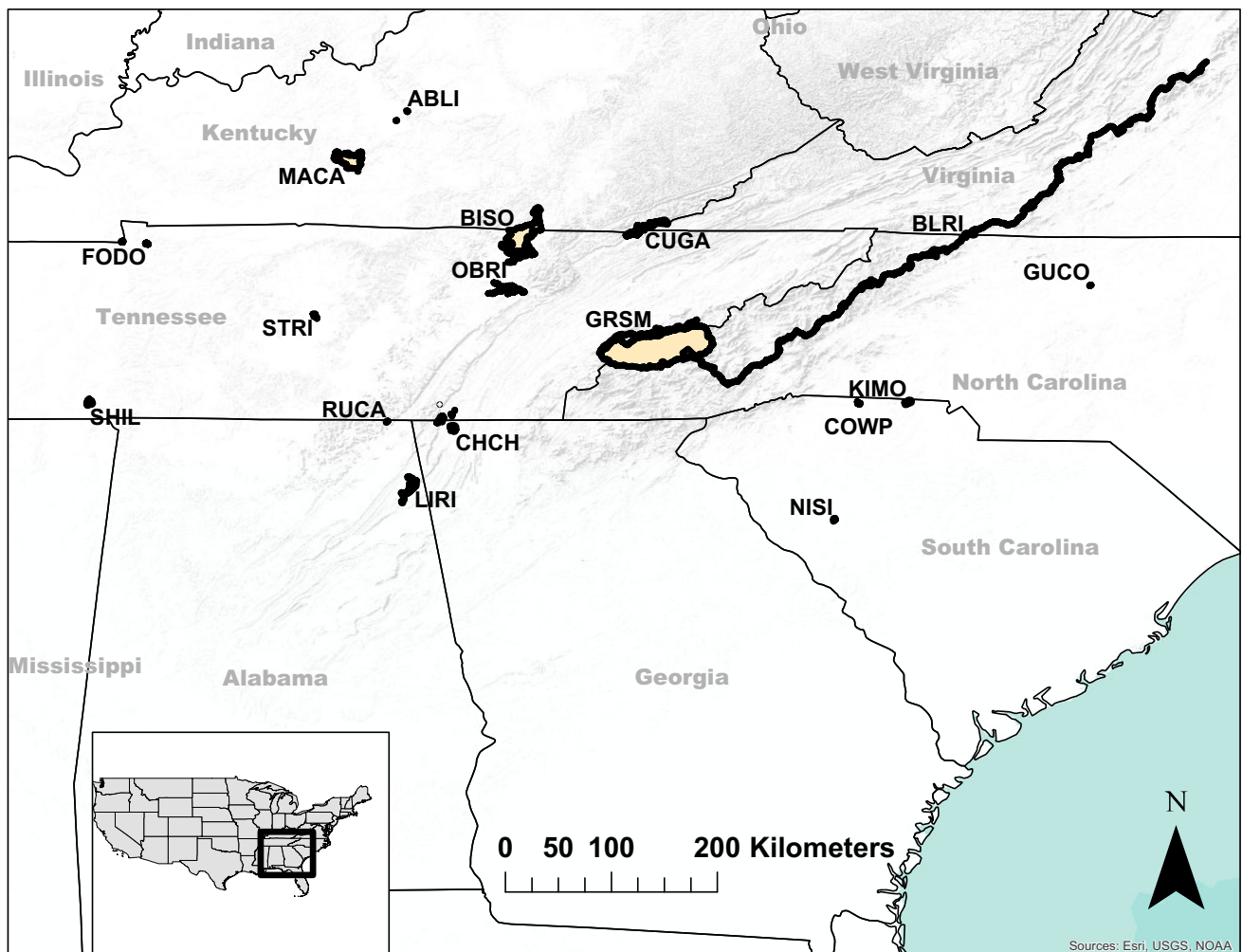
Fig. 1 Location of the 17 US National Parks sampled in this study. *ABLI* Abraham Lincoln Boyhood Home (Hodgenville, KY), *BISO* Big South Fork National River and Recreation Area (Oneida, TN), *BLRI* Blue Ridge Parkway (Asheville, NC), *CHCH* Chickamauga and Chattanooga National Military Park (Fort Oglethorpe, GA), *COWP* Cowpens National Battlefield (Gaffney, SC), *CUGA* Cumberland Gap National Historic Park (Middlesboro, KY), *FODO* Fort Donelson National Battlefield (Dover, TN), *GRSM* Great Smoky Mountains National Park (Gatlinburg, TN), *GUCO* Guilford Courthouse National Military Park (Greensboro, NC), *KIMO* Kings Mountain National Military Park (Blacksburg, SC), *LIRI* Little River Canyon National Preserve (Fort Payne, AL), *MACA* Mammoth Cave National Park (Mammoth Cave, KY), *NISI* Ninety Six National Historic Site (Ninety Six, SC), *OBRI* Obed Wild and Scenic River (Wartburg, TN), *RUCA* Russell Cave National Monument (Bridgeport, AL), *SHIL* Shiloh National Military Park (Shiloh, TN), *STRI* Stone's River National Battlefield (Murfreesboro, TN)

detect species within parks, at the expense of the detection of all species at all sites.

Spatial analysis of park shapes

Protected area units in PANs are often designated for many different reasons, including the commemoration of important historical events, geologic features and unique landscapes or other factors that may not correlate with local or regional patterns of biodiversity. These factors can cause variation in the "core", or the proportion of area within a protected area that is well buffered from human disturbance, connected to other patches or that might influence

patterns of biodiversity (McGarigal et al. 2002). In order to describe how our PAN units vary in these spatial descriptors of patch shape and complexity, we first used a GIS to rasterize the PAN unit boundaries, from a pixel size of approximately 0.025 ha, then used the PatchStat function (in the *R* package SDMTools; VanDerWal et al. 2014) to calculate landscape patch statistics for each of the 17 PAN units. We plotted the "shape.index" against the logarithm of park area to illustrate how longer, linear PAN units can be distinguished. As a heuristic to visually portray the variation in shape and size among the PAN, we then plotted EPT richness against the fractal dimension index, (approximately twice the logarithm of patch perimeter

divided by the logarithm of patch area), the shape index (the sum of the perimeter of the park divided by the square root of the park area), and the core area index, (the ratio of the internal non-edge influenced area: total park area).

Estimating regional species pools

No comprehensive, species-specific, source of detailed data on the entire geographic range of North American EPT species has been published, although some excellent recent regional summaries of state-wide occurrences exist (e.g. McCafferty et al. 2010). We used literature records and consultation with agency experts to generate lists of species that occur in each state, then used that list to construct species pools for parks in each state (Table S1). We relied heavily upon the North American Plecoptera list (Stark et al. 2009) and a recent review of mayfly records in the southeastern US (McCafferty et al. 2010). Trichoptera records are derived from published reviews (Frazer et al. 1991; Harris et al. 1991; Etnier et al. 1998, 2010; Flint et al. 2004, 2008, 2009; Lenat et al. 2012; Floyd et al. 2012) and scattered literature records compiled by CRP and JLR. Parks that overlapped the boundary of two states were given a regional source pool of both states.

Distance measures and regression on the decay of similarity

Spatial analyses of similarity necessarily require some measures of distance. Our questions utilize these analyses to ask questions about how patterns of similarity among units (i.e. states, parks) are structured in geographic or multivariate space. Similarity measures could include geographic distance, environmental similarity or the taxonomic turnover of ecological assemblages. If species assemblages are strongly structured in space (if there is high local endemism, or if species assemblages are strongly associated with environmental gradients) this would have particular management implications that are not entailed by other spatial configurations (if species occurrences in PAN units are random draws from widespread populations, or there is low autocorrelation in species occurrences). Comparative analyses of the strength of the DDR can elucidate the relative contribution of environmental gradients and species ranges to turnover in organismal assemblages.

Geographic and climate distances

Using spatial analysis and mapping software (ESRI 2011), we found the geographic centroid of each national park and state, and then used R (R Development Core Team 2011, fields package, v. 6.6.3, Furrer et al. 2012) to compute all pairwise great circle distances between park and state

centroids, respectively. To obtain a proxy distance metric to assess the climatic properties of parks, we used down-scaled bioclimatic variables from WORLDCLIM (Hijmans et al. 2005). For all 1 km^2 raster cells at least partially occupied by a park we extracted the annual mean temperature, mean diurnal temperature range, maximum temperature of the warmest month, minimum temperature of the coldest month, annual precipitation, precipitation of the wettest month, and precipitation of the driest month variables. We used principal components analysis (PCA) to summarize the variation among parks and to calculate a mean score for each park along each principal component axis. We then used these PCA scores for the first two axes to calculate the Euclidean climate distances for each pair of parks. We then replicated the PCA analysis after removing the annual precipitation variable.

Assemblage similarity measures

Regional patterns of species endemism or patchiness create spatial auto-correlation in the geographic ranges of species, as well as in the similarity of assemblages of either (or both) parks and states. Using function vegdist in the R package vegan (v. 2.0-2, Oksanen et al. 2011), we calculated pairwise dissimilarities of the presence-absence assemblages of both parks and states to test for distance decay of similarity with simple regressions. Some small parks did not have confirmed species-level records (in some insect orders); therefore we removed these parks from the relevant pairwise distance measures for those taxon assemblages (Ephemeroptera: RUCA, Plecoptera: FODO, STRI). The Jaccard dissimilarity index is appropriate for testing DDR hypotheses on presence-absence analysis data.

Distance-decay of similarity along spatial and environmental gradients

There are many ways to evaluate beta diversity of species assemblages, depending on the nature of the question (Anderson et al. 2011). Our question was, precisely, whether species assemblages of parks with similar climates or in close geographical proximity are more similar than species assemblages in parks at greater distances (i.e., do proximal parks capture species from the same regional species pool). Therefore, we evaluated beta diversity as the turnover in parkwise or statewise EPT insect assemblages. To test for DDR relationships among states or parks, we regressed Jaccard dissimilarities onto distances and tested whether the slope of the best-fit regression line was significantly different from zero (positive slopes are evidence for a distance-decay of assemblage similarity, or DDR). We tested for a distance decay of environmental similarity

using the climate distance and geographic methods and these methods. Finally, we ignored the taxonomic order of insects and grouped assemblages by Natureserve rankings (G1–G2, G3–G4, G5) to test for a DDR relationship among assemblages of species in each category of rarity (i.e. are DDR of rare species different, among parks, than common species).

Spatial gradients of environmental similarity

Species turnover along gradients can occur purely as a function of space (e.g., across non-overlapping species ranges) or also along environmental gradients (e.g., across differing environments). These gradients are frequently auto-correlated, which complicates analyses designed to tease apart the relative contribution of space and environment to turnover in assemblages. Partial Mantel tests have been used to compare correlation in climatic, geographic distance factors and assemblage dissimilarities, but technical issues preclude strong confidence in the Type II error rates of this method (Guillot and Rousset 2013). As a precautionary example, we compare the strength of EPT assemblage DDR along the geographic and environmental gradients where parks are located to demonstrate the limitations on inference on these hypothesis tests.

Null model tests of effective imperiled species conservation

Since parks could systematically over- or underprotect rare (or common species), the "conservation performance" of a protected area we characterize by species richness metrics, but also measures that assess the observed PAN unit assemblages on the basis of rarity of those species in the known regional assemblages or lists. In general, larger areas contain more species of plants and animals (Arrhenius 1921), but since many protected areas are constructed for historical commemoration of events within developed landscapes, they may still potentially protect small islands of quality habitat. Patterns of species assemblages can be generated by many different non-exclusive phenomena, but our analyses are not aimed at elucidating the ecological processes that determine the distributions of individual species, and are aimed instead at comparings the conservation performance of the protected areas across the network.

To test the hypothesis that park EPT assemblages have disproportionate numbers of rare or common members, relative to the regional species pool, we built a null model that randomly drew samples from the regional species pool and compared these random pseudoassemblages to the observed assemblages in each park. For each aquatic insect order, the resampling procedure provided a null

distribution of assemblages, where the composition of rarity is determined solely by a random draw from each species pool. This assesses the occurrence of rare elements relative to the prevalence in the regional species pool. This randomization was repeated 99,999 times for each park, while holding species richness fixed (i.e., the row sums of our presence-absence matrix were constant). Thus, each of these null assemblages has the same species richness as observed within the park, but with a random distribution of rarity values (G1–G5). For each observed park EPT assemblage and within each category of rarity (G1–G5), we calculated the probability of the observed frequency of species in each category of rarity, based on the 100,000 assemblages.

Results

Patterns in species richness and rarity

Species richness of EPT insect orders is highly variable across both the parks in our inventory (Table 1) and the southeastern USA, varying by a factor of 2–4× within each taxonomic order across the states (Fig. 2a; full matrix given as Table S1 in Supplemental Materials). No species-area relationship is evident among the southeastern US states; no slope estimates were significantly different from zero (Table 2). The fraction of the state EPT fauna ranked as "most imperiled" (G1) varied among southeastern states, with Florida reporting the highest proportion of G1 taxa at 10 % (state G1 mean 3.9 %). The least imperiled taxa, G5, comprised at least 59 % of the species of each states (mean 71.6 %) (Fig. 3).

Species richness of park assemblages varied more than the species richness of states, by more than two orders of magnitude within the network (Fig. 2a, b; full matrix given as Table S2 in Supplemental Materials). Park assemblages demonstrate clear positive species richness-area relationships for all three insect orders (Fig. 2b). More specimens were collected from larger parks, which also were sampled more intensively (specimens per unit area[2]) than small parks (Fig. 2c, d.) More individuals of Trichoptera species were observed in the collections from each park than Ephemeroptera and Plecoptera, particularly in larger collections (Fig. 2c; Table 1, full species list in Table S1). The largest parks were also the most heavily sampled (Fig. 2c, d). There was no relationship among log transformed park area and the proportion of G1 species, but the proportions of G2–G4 species were positively related to park area. The proportion of G5 species in park assemblages had a negative relationship with area, so that common species formed a smaller fraction of the assemblages contained within larger parks (Table S3).

Table 1 Patterns of species richness in national parks

	Ephemeroptera	Plecoptera	Trichoptera	Area (km^2)
Total	129	136	359	3493.2
ABLI	3	4	23	1.4
BISO	29	23	119	507.1
BLRI	48	57	213	385.1
CHCH	11	4	83	36.6
COWP	4	7	44	3.4
CUGA	15	16	94	95.7
FODO	5	0	28	4.1
GRSM	104	94	202	2114.2
GUCO	4	4	22	1
KIMO	8	17	66	16
LIRI	17	26	108	61.9
MACA	12	11	92	213.8
NISI	15	5	59	4.1
OBRI	10	26	120	20.5
RUCA	0	3	13	1.3
SHIL	12	6	76	24.1
STRI	5	0	17	2.9

See Fig. 1 caption for key to park abbreviations

Our estimates of the strength of the abundance-area and species-area relationships could be positively skewed by the Great Smoky Mountains National Park (GRSM), since that park has a total area several times greater than the next largest park in our study and has been extensively surveyed for decades beyond the scope of our project (Fig. 2b, c; Table 1). Omitting GRSM from analyses did reduce the slope estimates of species-area regressions, outside of the 95 % confidence intervals of estimates including GRSM, for Ephemeroptera and Plecoptera. However omitting GRSM from the analysis increased the slope of the species-area regression, for Trichoptera, to the upper limit of the 95 % confidence interval for the full regression (Table 2). Park species richness was correlated with the richness of the regional species pool (Pearson's *r*; E 0.57, P 0.46, T 0.48) (See Table S2 for state source pools of EPT species). EPT assemblages observed within each park were dominated by common species with the lowest rank of imperilment (Fig. 4). *Shape, fractal dimension and core area indices.*

The shape index clearly distinguishes two PAN units, the BLRI (Blue Ridge Parkway) and the OBRI (Obed Wild and Scenic River) (Fig. 5a). Each of these units has a longer and more linear shape than the other PAN units, although the parkway is much longer, more narrow and complex. The BLRI was also the most self-similar at two scales (Fig. 5b). The GRSM, a large and mostly contiguous patch, scored relatively low on the fractal dimension and

shape indices (Fig. 5c). All park boundaries scored highly on the core area index (Fig. 5d). Shape metrics did not have clear relationships with EPT richness in this PAN.

Spatial patterns of faunal and climatic similarity

Temperature and precipitation gradients

Principal component analysis revealed that parks are distributed along a strong precipitation gradient: the first principal component axis most heavily weighted annual precipitation and accounted for 98.4 % of the variance among parks (Table S4, Figure S1). Ordination of parks on principal component axes calculated without annual precipitation yielded a similar topology (Figure S2) but this algorithm explained less cumulative variance among parks than analyses that included annual temperature and precipitation data (first axis accounting for 78.5 % of the variation; Table S5), so we retained the full ordination for distance regressions.

Distance decay of similarity relationships (DDRs)

We measured a strong decay of assemblage similarity with geographic distance across all the states of the southeastern USA, for all three aquatic insect orders (Table 3). The strength and statistical significance of these regional trends of similarity diminished when we restricted this regression to exclude states without parks sampled for Plecoptera and Trichoptera, but for Ephemeroptera the significance of the distance decay relationship collapsed altogether (Table 3). Among all of the park assemblages, DDRs were detected in Plecoptera and Trichoptera, but not Ephemeroptera. After removing GRSM from analyses, distance-decay in Trichoptera assemblages was only marginally significant ($p = 0.051$) and did not explain much of the variation in park Trichoptera or Plecoptera assemblages (Table 3).

The rate of change along gradients of temperature and precipitation similarity, in this network, is several orders of magnitude greater than change in assemblage similarity along the same geographic distances, suggesting that many species occupy broad climate across our study area (Table 3). Distance decay regression of park climate similarity revealed a significant geographic decay of climate similarity with geographic distance, among all parks. When the large, centrally located and environmentally heterogeneous GRSM was omitted from this regression, the slope estimate increased along with the fit of this regression. Regressions of assemblage distance measures onto temperature and precipitation dissimilarities yielded contrasting results; no DDR was detected among Plecoptera assemblages using all parks, but all three insect orders had significant climate distance decay when the large and

Fig. 2 Biodiversity of EPT orders in 17 national parks in the southern highlands USA. For plots **a**, **b** and **d**, the heavily sampled Great Smoky Mountains National Park (GRSM) is in the *upper right corner*. **a** EPT species richness of southeastern US states plotted against the natural log of the state area (km²). **b** Observed national park EPT species richness plotted against the natural log of the park area (km²). Total number individuals examined from each national park. **c** Species richness—abundance relationships of EPT specimens from national parks, plotted against park area (log–log scale). **d** Effort-area relationship for EPT species in 17 national parks, where effort is indexed by the number of individual specimens identified to species
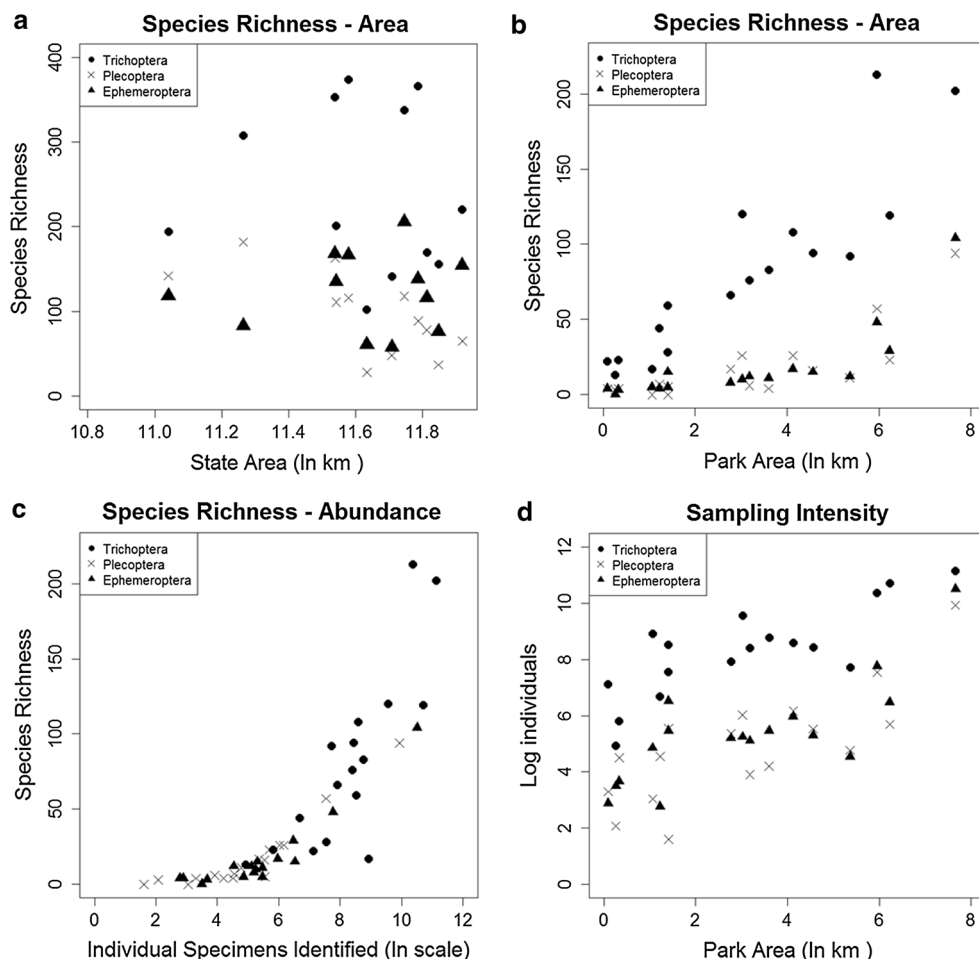
Table 2 Parameter estimates (and 95 % CI around slope) for area and richness relationships among three insect orders

	Ephemeroptera	Plecoptera	Trichoptera
17 Parks			
Slope	2.4 (2–2.8)	2.2 (1.7–2.7)	4.7 (3.2–6.2)
p	<0.0005	<0.0005	<0.0005
r^2	0.89	0.83	0.72
F (*df*)	136.4 (1, 15)	76.7 (1, 15)	43.1 (1, 15)
No GRSM			
Slope	1.5 (0.9–2.1)	1.4 (0.5–2.4)	6.2 (3.5–8.8)
p	<0.0005	0.007	<0.0005
r^2	0.69	0.42	0.62
F (*df*)	32.6 (1, 13)	10.5 (1, 12)	25.7 (1, 14)
States			
Slope	−0.2	−0.5	−0.3
p	0.62	0.12	0.73
r^2	−0.07	0.15	−0.09
F (*df*)	0.27 (1,10)	2.9 (1,10)	0.1 (1,10)

Omitting GRSM alters slope, strength of relationship. States show no significant species-area relationship
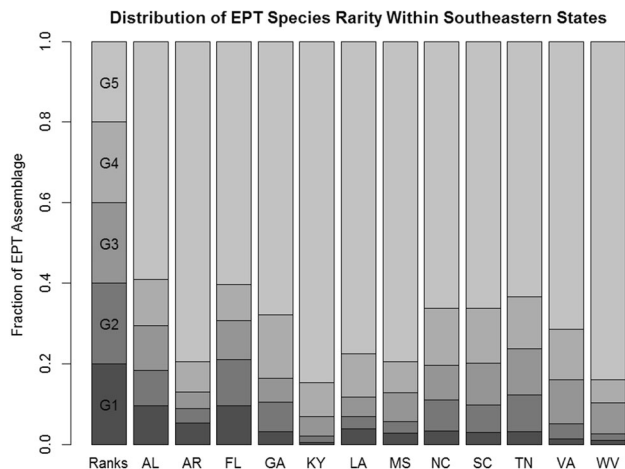
Fig. 3 Frequency of species in each imperilment ranking, across all EPT species recorded from the southeastern US

climatically complex GRSM was excluded. Species assemblages containing members ranked as "least imperiled" (G5) showed no significant DDR among parks, but
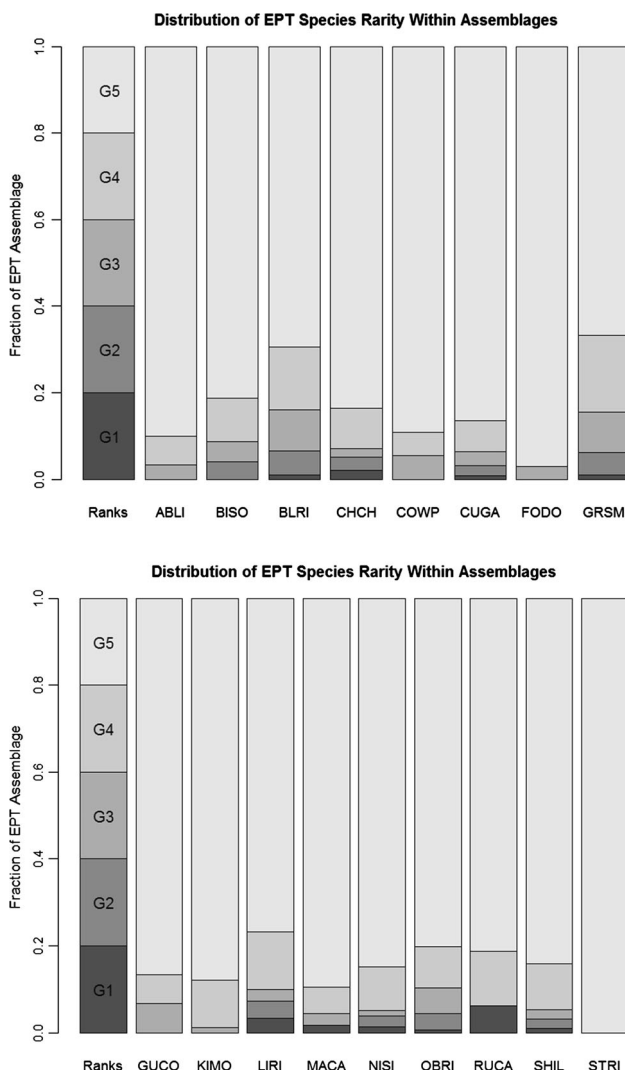
Fig. 4 Frequency of species in each imperilment ranking, across all EPT species observed in each of the 17 national parks

the similarity of more imperiled assemblages (G1 and G2, G3 and G4) did significantly decay with geographic distance; removing the GRSM had little impact on the fit or parameter estimates for this regression (Table 4).

Null model assembly of faunal composition from regional source pools

We created 10,000 random EPT assemblages for each park, drawing from our estimated regional species pool for each unique park. The distribution of rarity in these random assemblages was used to assess the statistical significance of the observed frequency of rarity in park EPT assemblages (Fig. 4). Null model results are overwhelmingly unequivocal for all three insect orders: park assemblages are disproportionately composed of the most common species (rank G5) than would be expected from random draws of the known species pool. There are not

disproportionately more severely imperiled species in this PAN, relative to our random expectations from the null assemblages, because these units do not more effectively protect the most imperiled species in these regional species pools (Table 5a–c). Two parks, GRSM and LIRI, had significantly more G4 Plecoptera species than predicted by the randomized species pool ($p = 0.02$ and $p = 0.009$ respectively, Table 5b), but no parks differentially over-protected G4 Ephemeroptera or Trichoptera (e.g. there were not more G4 species present than would be expected from a random draw from the regional source pool). The assemblage-level prevalence of rarity observed in park Trichoptera assemblages was not more frequent in parks than in the source pool for any of the taxa except G5 (Table 5c). Many PAN units had a higher incidence of G5 taxa than predicted by the null model (Ephemeroptera: 9 parks $p \leq 0.1$, Plecoptera: 5 parks $p \leq 0.1$, Trichoptera: 17 parks $p < 0.08$).

Discussion

We set out to describe the composition of aquatic insect assemblages among a network of protected areas, US National Parks in the southeastern USA. This resulted a number of new state records and the discovery of some undescribed species (Robinson unpublished data). We sought to compare how well parks "capture" rare species from the regional species pool, using a null model to account for differences in the species pool specific to each park. Our questions are motivated by the reasoning that parks might have a disproportionate number of rare (or common) species due to their size, location or by capturing high quality environments, and that the potential pool of species which might occur in a particular park varies with the geographic distribution of species. We conceptualize the boundaries of a park as delimiting a geographic sample of the regional fauna, where species occurrences are a function of the location of species geographic ranges (i.e. presence of individuals in metapopulations) within park boundaries. Null model tests compare our observed species assemblages to hypothetical assemblages constructed by random assembly.

Regional aquatic insect biodiversity and biogeographic context

We interpret our results as consistent with the ideas of earlier workers that the unglaciated highland regions of eastern North America harbor a large reserve of phylogenetically significant, geographically structured taxonomic and ecological insect diversity (Allen 1990). Explanations offered for the consilience of distributional patterns among

Fig. 5 PAN units area and shape metrics do not consistently track EPT species richness. **a** Shape index distinguishes linear shapes from larger contiguous regions. **b** Fractal dimension index does not track EPT richness. **c** The shape index distinguishes linear PAN units but does not predict EPT richness. **d** Core area index metrics distinguishes smaller parks
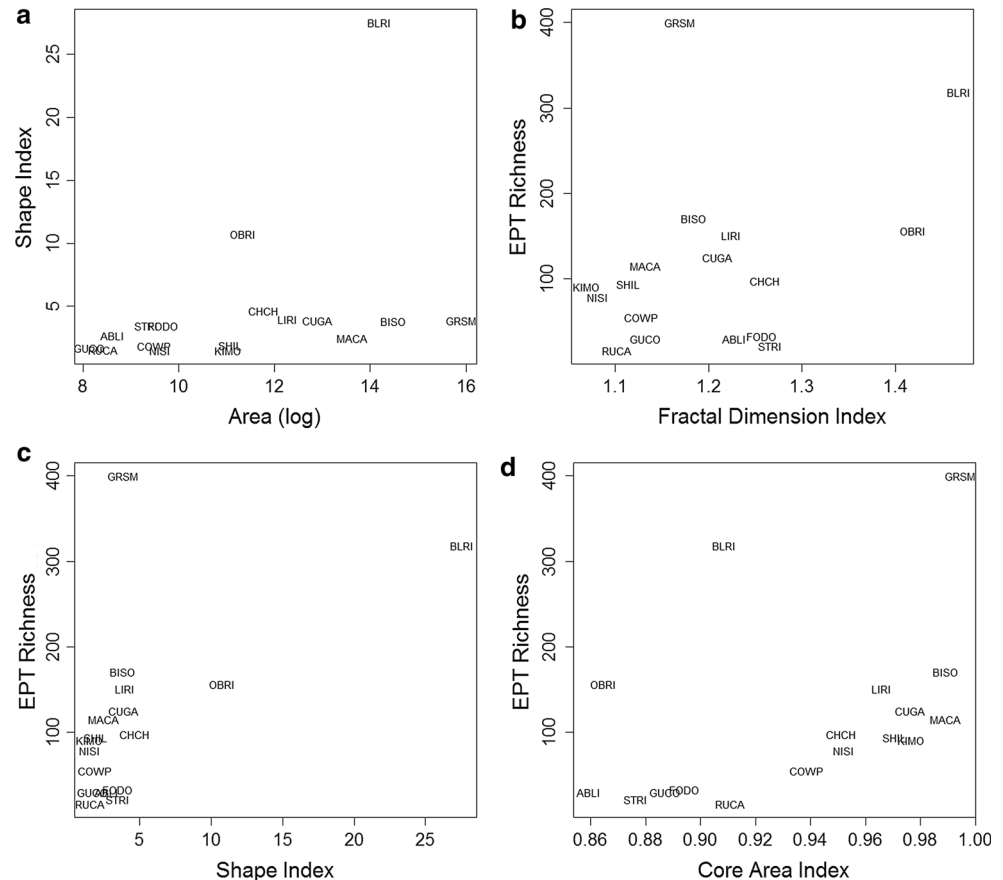
phylogenetic groups have generally assumed that the current geographic ranges of species across recently glaciated regions reflect historical dynamics associated (at least in part) with dispersal from historical unglaciated refugia (Ross 1953, 1956, 1965; Ross and Ricker 1971; McCafferty 1977; Allen 1990; Hamilton and Morse 1990). Yet, there is broad consensus among aquatic ecologists that environmental heterogeneity is an important driver of aquatic insect diversity and abundance at more immediate spatial and temporal scales (Wallace and Merritt 1980; Ward and Stanford 1982; Vinson and Hawkins 1998; Brown and Swan 2010). Our analyses of species presence-absence within PAN units were aimed at an intermediate scale, where detections of presence in a park were minimally influenced by variation in the occupancy of each biotope.

More species in the mountains

Demonstrating the distribution of species within the spatial units of our analysis is beyond the scope of this paper, but we do emphasize that our hypothesis of "more species in the mountains" is also supported by many accounts of state-wide species occurrences in the literature. Lenat et al.

(2012) described many species restricted in distribution, within North Carolina, to the western mountainous region (and some to the eastern blackwater coastal plain). Frazer et al. (1991) provided excellent lists of temporal species occurrences from many different streams within a single protected area in a mountainous region of the northeastern corner of Alabama. Etnier et al. (1998) did not extensively summarize occurrences across the state for each species, but many species in Tennessee are reported as restricted to the eastern mountains within the state. Floyd et al. (2012) described the reported species richness by ecoregions within Kentucky as consistent with the hypothesis of greater species richness in the mountainous coalfields and interior plateaus of southeastern corner of the state, near the borders of NC, VA, WV and TN.

Species-area and effort

In general, in this study we found more species in larger parks. In contrast, we found no species-area relationships in the state lists, the regional source pools, from which these park assemblages are constructed. Our null model results support our intuitions about the regional fauna, in that faunal assemblages are likely to vary substantially across

Table 3 Distance-decay regression estimates for southeastern aquatic insect assemblages

	Slope	r^2	F
Southeastern states			
Ephemeroptera	**2.93E−04*** **	0.36	37.78 (1, 64)
Plecoptera	**2.67E−04*** **	0.43	48.94 (1, 64)
Trichoptera	**2.43E−06*** **	0.4	43.79 (1, 64)
Study states			
Ephemeroptera	1.51E−04	0.06	2.49 (1, 19)
Plecoptera	1.85E−04	0.14	4.2 (1, 19)
Trichoptera	2.18E−04	0.1	3.21 (1, 19)
All parks			
Ephemeroptera	7.48E−05	0	0.59 (1, 118)
Plecoptera	**1.15E−04** **	0.04	5.19 (1, 103)
Trichoptera	**1.29E−04** **	0.03	5.42 (1, 134)
Parks (No GRSM)			
Ephemeroptera	1.04E−04	0	1.35 (1, 103)
Plecoptera	**1.69E−04** **	0.03	4.06 (1, 89)
Trichoptera	**1.75E−04** *	0.02	3.89 (1, 118)
Climate			
Ephemeroptera	**4.72E−04*** **	0.15	21.92 (1, 118)
Plecoptera	1.39E−05	0.01	2.68 (1, 103)
Trichoptera	**3.06E−04** **	0.06	10.11 (1, 134)
Climate (No GRSM)			
Ephemeroptera	**3.34E−04** **	0.06	7.53 (1, 103)
Plecoptera	**2.13E−04** **	0.04	4.46 (1, 89)
Trichoptera	**3.60E−04** **	0.07	10.2 (1, 118)
Climate similarity			
All parks	**0.137** **	0.05	8.306 (1, 134)
DDR			
GRSM excluded	**0.298*** **	0.14	21.030 (1, 118)

Bold font denotes slope estimates significantly different from zero (* significant at $p < 0.1$; ** significant at $p < 0.05$; *** significant at $p < 0.001$)

the spatial units (states) for which we constructed source pools. The proliferation of new state records and new species descriptions across the region, from our efforts

alone, confirms that there remain large gaps in our understanding of these species distributions.

The species-area relationship observed among parks could be due to larger parks having greater total number of individual organisms in all habitats. The two largest parks in this study, the Blue Ridge Parkway (BLRI) and Great Smoky Mountains National Park (GRSM), capture many high elevation mountain ranges (>1800 m asl) with populations of narrowly restricted or endemic plants and animals. These are high quality aquatic environments and protect rare species, but those habitats and fauna only occur in what happen to be the largest parks (and in the BLRI along a several hundred kilometer transect). Small parks in our study do not capture such vast expanses of rare or high quality landscapes or hydrologic conditions, but may protect formerly high quality habitats that now occur within a spatial context of greatly altered land cover conditions. Similarly, these PAN units may experience lower rates of colonization by EPT species, as a function of surrounding land uses or connectivity. If PAN units protect species now restricted to protected areas, parks may contribute individuals to source new local breeding populations dispersing to highly altered surrounding environments. These questions require more sampling effort, as well as more precise estimates of the composition of local and regional species pools.

We relied on the number of individuals per km^2 sampled, as well as the species richness per number of individuals, as indices of our collection effort in these analyses. Thus, we have not directly considered variation among sites within a park, or the number of sites visited in a park. Our questions were directed at composition of assemblages and the distribution of rarity, relative to the background species pool. In instances where we may have failed to observe the presence of some species, the most obvious consequence to our results would be a smaller repeated draw from the species pool we used to create each null assemblage (i.e. the species richness of the park determined the size of the null model draw). Since common species dominate those regional species pools, drawing one more taxa from each null assemblage would most likely increase

Table 4 Distance-decay regression estimates for aquatic insects in different categories of perceived threat

	Ranking	Slope	r^2	F
All parks	**G1 and G2**	**2.54E−04** **	0.08	5.54 (1, 53)
	G3 and G4	**2.04E−04*** **	0.11	15.79 (1, 118)
	G5	**9.78E−05** *	0.01	3.18 (1, 134)
Parks (No GRSM)	**G1 and G2**	**2.64E−04** **	0.1	5.63 (1, 43)
	G3 and G4	**2.01E−04*** **	0.12	15.2 (1, 103)
	G5	8.92E−05	0.01	2.63 (1, 118)

Bold font denotes slope estimates significantly different from zero (* significant at $p < 0.1$; ** significant at $p < 0.05$; *** significant at $p < 0.001$)

Table 5 (a) Ephemeroptera, (b) Plecoptera, (c) Trichoptera

	G1	G2	G3	G4	G5
(a)					
ABLI	1	1	1	1	0.509
BISO	1	0.773	0.81	0.865	0.174
BLRI	1	1	1	0.902	**0.001**
CHCH	1	1	1	0.947	**0.053**
COWP	1	1	1	1	0.213
CUGA	1	1	1	0.835	**0.071**
FODO	1	1	1	1	0.195
GRSM	0.834	0.996	0.92	0.885	**0.005**
GUCO	1	1	1	1	0.192
KIMO	1	1	1	1	**0.044**
LIRI	1	1	0.611	0.987	**0.088**
MACA	1	1	1	1	**0.06**
NISI	1	1	1	0.864	**0.087**
OBRI	1	1	1	1	**0.037**
RUCA	1	1	1	1	1
SHIL	1	1	1	0.923	0.109
STRI	1	1	1	1	0.193
(b)					
ABLI	1	1	1	0.667	0.505
BISO	1	0.819	1	0.414	**0.033**
BLRI	1	0.772	0.705	**0.039**	0.587
CHCH	1	1	0.561	0.575	0.68
COWP	1	1	0.743	0.794	0.379
CUGA	1	1	0.829	0.743	**0.049**
FODO	1	1	1	1	1
GRSM	1	0.88	0.985	**0.02**	**0.039**
GUCO	1	1	1	1	**0.069**
KIMO	1	1	0.972	0.435	0.193
LIRI	0.829	0.754	0.974	**0.009**	0.853
MACA	1	1	0.702	0.956	0.132
NISI	1	1	0.615	0.673	0.658
OBRI	1	0.926	0.967	0.565	**0.034**
RUCA	1	1	1	0.381	0.709
SHIL	1	0.418	0.681	1	0.372
STRI	1	1	1	1	1
(c)					
ABLI	1	1	0.863	0.94	**0.048**
BISO	1	0.993	0.994	0.999	**0**
BLRI	0.989	0.75	0.9	0.995	**0**
CHCH	0.879	0.995	1	0.964	**0**
COWP	1	1	0.962	0.992	**0**
CUGA	0.982	0.995	1	1	**0**
FODO	1	1	0.971	1	**0**
GRSM	0.999	0.987	0.924	0.592	**0.001**
GUCO	1	1	0.582	0.848	**0.079**
KIMO	1	1	1	0.928	**0**
LIRI	0.998	0.983	1	0.748	**0**

Table 5 continued

	G1	G2	G3	G4	G5
MACA	0.379	1	0.999	0.973	**0**
NISI	0.857	0.937	1	0.938	**0**
OBRI	0.992	0.995	0.996	0.976	**0**
RUCA	0.737	1	1	0.801	**0.048**
SHIL	0.938	1	1	0.674	**0**
STRI	1	1	1	1	**0**

Null model probabilities that observed proportions of EPT species in any global imperilment ranks are larger than would be expected from a random draw from the regional species pool. p values are one-sided; the lower the p value the more likely that the observed frequency is larger than expected by random draws. Bold font indicates $p < 0.1$

the representation of common species in the null assemblages rather than rare species.

Distance-decay of similarity of insect assemblages

Patterns of the distance decay of EPT assemblage similarities are consistent with several interesting biogeographical hypotheses. We detected a DDR, for all three insect orders, across all the states of the southeastern US. However, this DDR disappeared when we only analyzed the subset of states with parks in the PANs we inventoried (i.e., excluding the states AR, FL, LA, MS and WV). We interpret this as evidence to be consistent with the hypothesis of a core Appalachian fauna, locally interacting with the sub-tropical, Midwestern and northeastern faunal assemblages on the regions peripheral to the uplifted areas (Hamilton and Morse 1990). Although Trichoptera and Plecoptera assemblages within the parks showed a statistical significant decline in similarity with distance ($p < 0.05$; Table 3), with similar slope estimates in each regression, distance decay of similarity explained more variation among EPT assemblages across the entire southeastern USA than it did among the units of the PAN in our study. We interpret this as affirming both the limitations of our actual species distribution data for estimating the true, real underlying patterns of species diversity across the region at finer biogeographic scales.

Interestingly, slope estimates from both the state and park DDRs are much weaker than the estimated slope of the distance decay of climate similarity. This PAN is not a random sample from the strong climatic gradients of temperature and precipitation across the southeastern United States, but we still observed a strong decay of climate similarity among parks (with geographic distance). Turnover in the regional fauna does not appear to strongly track these environmental gradients, since many species were collected in many of the parks and certainly occupy even

broader bioclimatic envelopes than we have observed in our survey. It is certainly true that the national parks in this protected area nextwork occupy a range of very different environments, with very different forests, geology, watershed geomorphology, local endemism and other factors creating spatial autocorrelation in species distributions and distance measures of similarity.

Size and shape of protected area units along environmental gradients

Increasingly, managers of protected area networks spanning natural environmental gradients are seeking to exploit these features for planning, prediction and experimentation (Peters and Darling 1985; Hannah 2008, 2011; Ackerly et al. 2010; Davison et al. 2012). In our study, some parks are large intact land holdings, while others are long narrow strips (parkways, river systems) or historic markers in plots of various sizes. Relative to parks in suburban areas selected mainly on historical significance, parkways and parks encompassing river and tributary corridors are examples of protected area configurations that may "capture" more of the local species pool by conserving different biotopes present along longer hydrological gradients from headwaters to mainstem environments (Clarke et al. 2008, 2010). In our study, one large and one intermediately sized park had much more linear shapes than the other units of the PAN. However, no obvious relationships with species richness were observed within these shape metrics.

These differences in spatial configuration of PANs, and the attendant consequences to the ecology of species occurring within those networks, are understudied topics in aquatic insect conservation. However, these topics have been explored in the conservation literature with examples from terrestrial and marine systems. Recommendations for the design of marine protected areas have included suggestions that managers minimize the ratio of reserve perimeter to area, to minimize edge effects and fishing mortality (McLeod et al. 2009). Other studies have recommended long, linear shaped reserves to facilitate enforcement and to simplify navigation (Friedlander et al. 2003). There have been fewer studies on the influence of the shape of PAN units in terrestrial systems, but McKinney (2002) found no evidence for a relationship between the shape of 77 protected area units and the frequency of alien plant species.

Among the parks in our study, the linear BLRI in particular is known to traverse breaks in the geographic distribution of several species (some species are apparently only known from habitats along the Parkway) and crosses a variety of differently sized streams, along an elevational gradient of 200–2000 meters asl. A similar elevation gradient occurs across the GRSM, but instead of a linear

highway transect, this gradient spans an enormous, contiguous tract of forests. Long linear transects may intersect many metapopulations and offer the possibility of protecting specific populations of particular species, while large contiguous parks protect forests that form the template for ecological processes within watersheds. Many of these possibilities are not mutually exclusive and may interact; more research on interactions between protected area shapes and size and metapopulation dynamics are necessary to improve aquatic insect conservation.

State lists are the best data available, but they just aren't good enough

For many orders of insects and other invertebrates, information on the geographic distribution of species is incomplete, if available at all (Lomolino 2004; Cardoso et al. 2011). Global conservation rankings (viz. NatureServe) utilize estimates of the geographic distribution, habitat associations and population dynamics of each individual species. Even when these data are available, NatureServe distributions are reported at the spatial scale of states and each species is given a global (G) and state (S) ranking of relative risk of imperilment. For EPT species, state rankings have not been assigned across all states and species in our analysis, and we have demonstrated that there are large gaps in our knowledge base on species occurrences within the states. State ranking criteria have been applied to datasets with variable effort by different methods, which make those rankings less helpful for our study questions. State lists have been important in the history of modern entomology, where systematists working with these three orders of insects summarized species distributions in faunal treatments at broad geographic scales or geopolitical units. In the USA, early regional taxonomic experts and systematists at federal and state surveys and universities had implicit interests in compiling their data by state boundaries (e.g. Betten et al. 1934; Frison 1935; Ross 1944; Burks 1953, Etnier et al. 1998). The practice of reporting occurrence data aggregated to the spatial grain of states may have interesting original influences, but as we have demonstrated the use of this reporting unit introduces bias which can obscure known gradients of species richness and diversity at ecologically relevant spatial scales across a region (e.g. an ecoregion). Some excellent recent examples of faunal lists of Trichoptera species (Harris et al. 1991; Houghton 2012) include comprehensive information on the distribution or rarity of species within states or species groups. Etnier et al. (2010), reviewing and describing all southeastern species of Agapetus (Trichoptera: Glossosomatidae), provide an excellent example for publishing comprehensive

(range-wide) regional geo-referenced species occurrence records for each species.

Alternative species pools for PAN units

We acknowledge several limitations to our estimate of the potentially colonizing pool of species in the regions where each PAN units occur. Our regressions of the distance decay of similarity support one general interpretation of the underlying spatial structure of the geographic ranges of this assemblage of individual species. Species richness is greater in mountainous areas, consistent with known patterns of the biogeography of other species groups. Using state lists for the regional species source pools is the best possible approximation for several reasons, which we explain here.

We showed a significant decay of similarity in the assemblages of EPT species in states across the entire southeastern US. This pattern disappeared when we only included states where parks occur, since our expected pattern is within the spatial grain of these units. This is consistent with the hypothesis that higher species richness occurs in the southern mountains and highland areas. Within the region of the study states, the mountainous areas tend to be on the borders of the states. Several of the units of this PAN cross state lines, with the consequence that the species pools we estimate for those units include all the species from both states. This we believe to be an overestimate, but erring on the side of commission, since we have no better estimate of the distributions of individual species. We considered using the distance to centroids of states as the best predictor of the species pool of a PAN unit near the edge of a state, but this had the unintended consequence of omitting observed species from the estimated source pools. We argue that there is no better estimate of what species are likely to occur in a state than the set of species that have been reported from that state.

We have argued that we need finer-scaled estimates of the distribution of species, in order to build better regional species pools, which is the Wallacean shortfall. In our analysis, adding more species to the regional source pool has no effect on our results. The individual units in our PAN protect many species, but these species are overwhelmingly (on average) the more common and widespread species of the assemblages. The most imperiled species (G1, G2) do occur in the units, but not at a frequency greater than the rate at which they occur in the entire state lists. The decay of similarity of the G1–G2 and G3–G4 species assemblages in the PAN is consistent with the finer-scaled hypothesis of regional species distributions we have discussed above, and the absence of distance-decay of similarity in the G5s to be consistent with our null model results of park "over-protection" of these species.

Improving our estimates of the regional source pool for PAN units would be an exercise in trimming the list of the expected species, particularly species that are endemic or range restricted and thus not realistically expected in some PAN units. When the underlying pattern of species distributions is underdetermined at finer spatial scales, state lists remain the most defensible estimate of the regional source pool for a PAN unit. Some PAN unit inventories discovered new state records for many species in each of the three orders, as well as species new to science. Our understanding of the biology and biogeography of aquatic insect species is far from complete, and we emphasize that these PAN units contain many EPT species, within and across the network. Our results do not impugn the significant result that national parks contain many species, but they do suggest that they are not enough to protect all of the species in this biogeographical hotspot of diversity.

Objective assessment of the threats facing poorly understood insect species will probably continue to be a challenge for conservation managers, particularly when those species occurrences are compared across broad geographic regions. In this analysis of the ecological performance of a protected area network, we relied on NatureServe estimates of the risks of imperilment faced by EPT species. These estimates are based on expert assessment of 8 criteria, including number of observations or populations, geographic distribution, threats faced by populations and trends in population size of habitat availability (Faber-Langendoen et al. 2012). The quantity of observed occurrences may not necessarily be an unbiased estimate of the imperilment risk faced by a particular species, but in many instances is the only or best estimate available. Putatively rare EPT species can be far more common (or occupy a larger geographic range) than previously believed, when additional collecting efforts or re-identification of previously collected specimens provide more species occurrence records. This should be expected when the state of knowledge of species distributions and ecological profiles is incomplete and species ranges are undetermined (Cardoso et al. 2011). However, when properly estimated, EPT species occurrence records can not only provide robust evidence of range size contraction or expansion across smaller geographic areas (e.g., DeWalt et al. 2005), but also (in theory) provide a rigorous quantitative basis for assigning species to categories of rarity. However, few datasets of this scope and taxonomic quality exist at the moment.

Protected area networks and shortfalls of biodiversity knowledge

To be clear, we insist that compilations of state lists probably remain the best available expectation of the EPT

species which we might collect from any unit in our protected area network. It is important to remember that national parks are far from the only PANs in the southeastern US; many other state, federal and NGO entities administer lands managed for conservation objectives. Thus, our analysis cannot be construed as an assessment of the adequacy of imperilment designations for EPT taxa in general, or to offer suggestions of how PANs might more effectively conserve rare species. Using the language of "shortfalls" of previous authors (Cardoso et al. 2011; Hortal et al. 2015), we suggest some ways to address these shortfalls for future studies.

The Linnean shortfall is the gap in our full discovery and description of species. Even in our surveys, we discovered several undescribed species of insects, some of which have since been described (Curler and Moulton 2010; Etnier et al. 2010) and others that are in preparation. Further investigation of large and diverse units of this PAN, and the surrounding landscapes in which they occur, are likely to produce more undescribed species of insects in these taxonomic orders. The Wallacean shortfall, or geographic distributions, is clearly limiting in our study, but we hope to have contributed to reducing those knowledge gaps here. Darwinian shortfalls, or gaps in our understanding of the evolution of species and higher level groupings, may be reduced by broad collaborations of researchers and taxonomic experts working both regionally and worldwide (Zhou et al. 2011, 2016).

Almost nothing is known about what sorts of interactions occur between species of EPT, *qua* species. However, a great deal is known about these insects that can fill the Raunkiæran, Prestonian and Hutchinsonian shortfalls, namely gaps in the knowledge of functional traits, phenology and ecological functions and abiotic tolerances (Hortal et al. 2015) of species. In fact, because of benthic water quality monitoring (Kenney et al. 2009), we suggest that more is generally known about these properties within these insect orders than in nearly all other broad groups of insects. The caveat "broadly" is necessary, since very little of this knowledge is directly tied to information on specific *species* but instead is summarized by genus, trophic group, family or other non-evolutionary units of analysis (a Linnean shortfall).

The limitations to diagnosing the immatures forms of many of these insect species are many, and the incongruence between these units and knowledge based on information about *species* will continue to impair our ability to directly transfer most biomonitoring data directly into conservation assessments. However, benthic sampling can provide added value to systematic adult sampling by providing researchers additional information about what species might be potentially collected at a site as adults during a different time of the year. Phenology, trophic categories

and tolerance/intolerance metrics were all useful during the course of this study as guides for distributing sampling efforts at sites to capture adult forms.

Vast holdings of national forests across the southeastern US are likely, in sum, to capture a larger fraction of the expected regional aquatic insect biodiversity than this relatively tiny network of national parks, but these lands experience many different types of land use and do not all share the high level of protection afforded by the PANs we considered here. Our observations of patterns of occurrence and rarity, and our knowledge of the biology of these insects, make it clear that the protection or conservation of all species in these states will require more than this small network of national parks. What is not clear is the role that other protected areas (large networks of National Forest, other holdings) play in supporting regional EPT species diversity. Since a substantial number of EPT species are only known from a small handful of localities or collection events, more research is needed to determine whether our hypothetical patterns of regional species richness are driven by variation in sampling effort (or other factors) or truly reflect narrow geographic extents of occurrence. Integrating sampling techniques and spatial analyses that maximize the utility of published museum, specimen and literature data to derive predictive distribution of species should facilitate the effective conservation of EPT species in this biodiversity hotspot by providing a clearer estimate of the most imperiled species and by providing more realistic expectations of the composition of local species assemblages.

References

Ackerly DD, Loarie SR, Cornwell WK, Weiss SB, Hamilton H, Branciforte R, Kraft NJB (2010) The geography of climate change: implications for conservation biogeography. Divers Distrib 16:476–487

Allen RT (1990) Insect endemism in the interior highlands of North America. Fla Entomol 73(4):539–569

Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, Sanders NJ, Cornell HV, Comita LS, Davies KF, Harrison SP, Kraft NJB, Stegen JC, Swenson NG (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. Ecol Lett 14(1):19–28. doi:10.1111/j.1461-0248.2010.01552.x

Arrhenius O (1921) Species and area. J Ecol 9(1):95–99

Betten C, Kjellgren BL, Orcutt AW, Davis MB (1934) The caddis flies or Trichoptera of New York state (No. 292). The University of the State of New York

Brooks TM, Bakarr MI, Boucher T, Da Fonseca GAB, Hilton-Taylor C, Hoekstra NM, Moritz T, Olivieri S, Parrish SJ, Pressey RL, Rodrigues ASL, Sechrest S, Stattersfield A, Strahm W, Stuart SN (2004) Coverage provided by the global protected-area system: is it enough? Bioscience 54(12):1081–1091

Brown BL, Swan CM (2010) Dendritic network structure constrains metacommunity properties in riverine ecosystems. J Anim Ecol 79:571–580

Burks BD (1953) The mayflies, or Ephemeroptera, of Illinois. Bull Nat Hist Surv Div 26(1):1–216

Cabeza M (2013) Knowledge gaps in protected area effectiveness. Anim Conserv 16(4):381–382

Cao Y, DeWalt RE, Robinson JL, Tweddale T, Hinz L, Pessino M (2013) Using Maxent to model the historic distributions of stonefly species in Illinois streams: the effects of regularization and threshold selections. Ecol Model 259:30–39

Cardoso P, Erwin TL, Borges PAV, New TR (2011) The seven impediments in invertebrate conservation and how to overcome them. Biol Conserv 144(11):2647–2655

Clarke A, MacNally R, Bond NR, Lake PS (2008) Macroinvertebrate diversity in headwater streams: a review. Freshw Biol 53:1707–1721

Clarke A, MacNally R, Bond NR, Lake PS (2010) Conserving macroinvertebrate diversity in headwater streams: the importance of knowing the relative contributions of α and β diversity. Divers Distrib 16:725–736

Curler GR, Moulton JK (2010) Contributions to Nearctic Stupkaiella Vaillant (Diptera: Psychodidae). Zootaxa 2397:48–60

Davison JE, Graumlich LJ, Rowland EL, Pederson GT, Breshears DD (2012) Leveraging modern climatology to increase adaptive capacity across protected area networks. Glob Environ Change 22:268–274

DeWalt RE, Favret C, Webb DW (2005) Just how imperiled are aquatic insects? A case study of stoneflies (Plecoptera) in Illinois. Ann Entomol Soc Am 98(6):941–950

DeWalt RE, Cao Y, Tweddale T, Grubbs SA, Hinz L, Pessino M, Robinson JL (2012) Ohio USA stoneflies (Insecta, Plecoptera): species richness estimation, distribution of functional niche traits, drainage affiliations, and relationships to other states. ZooKeys 178:1–26

DeWalt RE, Neu-Becker U, Steuber G (2013) Plecoptera species file online. Version 5(5.0), 6

ESRI (2011) ArcGIS desktop: release 10. Environmental Systems Research Institute, Redlands

Etnier DA, Baxter JT Jr, Fraley SJ, Parker CR (1998) A checklist of the Trichoptera of Tennessee. J Tenn Acad Sci 73(1–2):53–72

Etnier DA, Parker CR, Baxter JT Jr, Long TM (2010) A review of the genus Agapetus Curtis (Trichoptera: Glossosomatidae) in eastern and central North America, with description of 12 new species. Insecta Mundi 0149:1–77

Faber-Langendoen D, Nichols J, Master L, Snow K, Tomaino A, Bittman R, Hammerson G, Heidel B, Ramsay L, Teucher A, Young B (2012) NatureServe conservation status assessments: methodology for assigning ranks. NatureServe, Arlington

Floyd MA, Moulton JK, Schuester GA, Parker CR, Robinson JL (2012) An annotated checklist of the caddisflies (Insecta: Trichoptera) of Kentucky. J Ky Acad Sci 73(1):4–40

Flint OS Jr, Hoffman RL, Parker CR (2004) An annotated list of the caddisflies (Trichoptera) of Virginia: part I, introduction and families of Annulipalpia and Spicipalpia. Banisteria 24:23–46

Flint OS Jr, Hoffman RL, Parker CR (2008) An annotated list of the caddisflies (Trichoptera) of Virginia: Part II. Families of Integripalpia. Banisteria 31:3–23

Flint OS Jr, Hoffman RL, Parker CR (2009) An annotated list of the caddisflies (Trichoptera) of Virginia: part III. Emendations and biogeography. Banisteria 34:3–16

Frazer KS, Harris SC, Ward GM (1991) Survey of the Trichoptera of the Little River Drainage of Northeastern Alabama. Bull Alabama Mus Nat Hist 15(11):17–22

Friedlander A, Nowlis JS, Sanchez JA, Appeldoorn R, Usseglio P, McCormick C, Bejarano S, Mitchell-Chui A (2003) Designing effective marine protected areas in Seaflower Biosphere Reserve, Colombia, based on biological and sociological information. Conserv Biol 17(6):1769–1784

Frison TH (1935) The stoneflies, or Plecoptera, of Illinois. Bull Ill Nat Hist Surv 20(4):281–471

Furrer R, Nychka D, Sain S (2012) Fields: tools for spatial data. Version 6.6.3. http://cran.r-project.org/

Guillot G, Rousset F (2013) Dismantling the mantel tests. Methods Ecol Evol 4(4):336–344

Hamilton SW, Morse JC (1990) Southeastern caddisfly fauna: origins and affinities. Fla Entomol 73(4):587–600

Hannah L (2008) Protected areas and climate change. Ann NY Acad Sci 1134:201–212

Hannah L (2011) Climate change, connectivity and conservation success. Conserv Biol 25(6):1139–1142

Harris SC, O'Neil PE, Lago PK (1991) Caddisflies of Alabama. Bull Geol Surv Ala 142:1–442

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25(15):1965–1978

Hortal J, de Bello F, Diniz-Filho JAF, Lewinson TAM, Lobo JM, Ladle RJ (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. Annu Rev Ecol Syst 46:523–549

Houghton DC (2012) Biological diversity of the Minnesota caddisflies (Insecta, Trichoptera). Zookeys 189:1–389

Kenney MA, Sutton-Grier AE, Smith RF, Gresens SE (2009) Benthic macroinvertebrates as indicators of water quality: the intersection of science and policy. Terr Arthropod Rev 2:99–128

Lenat DR, Ruiter DE, Parker CR, Robinson JL, Beaty SR, Flint OS Jr (2012) Caddisfly (Trichoptera) records for North Carolina. Southeast Nat 9(2):201–236

Lomolino MV (2004) Conservation biogeography. In: Lomolino MV, Heaney LR (eds) Frontiers of biogeography: new directions in the geography of nature. Sinauer Associates, Sunderland, pp 293–296

Lydeard C, Mayden RL (1995) A diverse and endangered aquatic ecosystem of the Southeast United States. Conserv Biol 9(4):800–805

Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405:243–253

McCafferty WP (1977) Biosystematics of Dannella and related subgenera of Ephemerella (Ephemeroptera: Ephemerellidae). Ann Entomol Soc Am 70(6):881–889

McCafferty WP, Lenat DR, Jacobus LM, Meyer MD (2010) The mayflies (Ephemeroptera) of the southeastern United States. Trans Am Entomol Soc 136(3–4):221–233

McGarigal KA, Cushman SA, Neel MC, Ene E (2002) FRAGSTATS: spatial pattern analysis program for categorical maps. University of Massachusetts, Amherst. http://www.umass.edu/landeco/research/fragstats/fragstats.html

McKinney ML (2002) Influence of settlement time, human population, park shape and age, visitation and roads on the number of

alien plant species in protected areas in the USA. Divers Distrib 8(6):311–318

McLeod E, Salm R, Green A, Almany J (2009) Designing marine protected area networks to address the impact of climate change. Front Ecol Environ 7:362–370

Morse JC, Stark BP, McCafferty WP (1993) Southern Appalachian streams at risk: implications for mayflies, stoneflies, caddisflies and other aquatic biota. Aquat Conserv 3:293–303

Morse JC, Stark BP, McCafferty WP, Tennessen KJ (1997) Southern Appalachian and other southeastern streams at risk: implications for mayflies, dragonflies, stoneflies and caddisflies. In: Bentz GW, Collins DE (eds) Aquatic fauna in peril: the southeastern perspective. Special Publication I, Southeastern Aquatic Research Institute. Lenz Design and Communications, Decatur, pp 17–42

NatureServe (2015) NatureServe explorer: an online encyclopedia of life [web application]. Version 7.1. NatureServe, Arlington, Virginia. http://www.natureserve.org/explorer. State Lists Accessed: 7 Dec 2011

Nichols BJ, Langdon KR (2007) The Smokies all taxa biodiversity inventory: history and progress. Southeast Nat 6(sp2):27–34

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2011) Vegan: community ecology package. Version 2.0-1. http://cran.r-project.org/

Parker CR, Flint OS Jr, Jacobus LM, Kondratieff BC, McCafferty WP, Morse JC (2007) Ephemeroptera, Megaloptera, Plecoptera and Trichoptera of Great Smoky Mountains National Park. Southeast Nat 6(S1):159–174

Peters RL, Darling JDS (1985) The greenhouse effect and nature reserves. Bioscience 35:707–717

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/

Ross HH (1944) The caddisflies, or Trichoptera, of Illinois. Bull Ill Nat Hist Surv 23(1):1–326

Ross HH (1953) On the origin and composition of the Nearctic insect fauna. Evolution 7:145–158

Ross HH (1956) Evolution and classification of the mountain caddisflies. University of Illinois Press, Urbana

Ross HH (1965) Pleistocene events and insects. In: Wright HE Jr, Frey DG (eds) The quaternary of the United States. VII Congress of the International Association for Quaternary Research, Princeton University Press, New Jersey, pp 583–596

Ross HH, Ricker WE (1971) The classification, evolution and dispersal of the winter stonefly genus *Allocapnia*. Ill Biol Monogr 45:1–166

Sharkey MJ (2001) The all taxa biological inventory of the Great Smoky Mountains National Park. Fla Entomol 84(4):556–564

Stark BP, Baumann RW, DeWalt RE (2009) Valid stonefly names for North America. http://plsa.inhs.uiuc.edu/plecoptera/validnames.aspx

VanDerWal J, Falconi L, Januchowski S, Shoo L, Storlie C (2014) SDMTools: species distribution modelling tools: tools for processing data associated with species distribution modelling exercises. Version 1.1-221. http://www.rforge.net/SDMTools/

Vinson MR, Hawkins CP (1998) Biodiversity of stream insects: variation at local basin and regional scales. Ann Rev Entomol 43:271–293

Wallace JB, Merritt RW (1980) Filter feeding ecology of aquatic insects. Ann Rev Entomol 25:103–132

Ward JV, Stanford JA (1982) Thermal responses in the evolutionary ecology of aquatic insects. Ann Rev Entomol 27:97–117

Zhou XL, Frandsen PB, Holzenthal RW, Beet CR, Bennett KR, Blahnik RJ, Bonada N, Cartwright D, Chuluunbat S, Cocks GV, Collins GE, deWaard J, Dean J, Flint OS Jr, Hausmann A, Hendrich L, Hess M, Hogg ID, Kondratieff BC, Malicky H, Milton MA, Moriniere J, Morse JC, Mwangi FN, Pauls SU, Gonzalez MR, Rinne A, Robinson JL, Salokannel J, Shackleton M, Smith B, Stamatakis A, StClair R, Thomas JA, Zamora-Munoz C, Ziesmann T, Kjer KM (2016) The Trichoptera barcode initiative: a strategy for generating a species-level Tree of Life. Philos Trans R Soc B. doi:10.1098/rstb.2016.0025

Zhou X, Robinson JL, Geraci CJ, Parker CR, Flint OS Jr, Etnier DA, DeWalt RE, Jacobus LM, Hebert PDN (2011) Accelerated construction of a regional DNA-barcode reference library: caddisflies (Trichoptera) in the Great Smoky Mountains National Park. J N Am Benthol Soc 30:131–162