Freshwater Biology

Freshwater Biology (2011) doi:10.1111/j.1365-2427.2010.02566.x

APPLIED ISSUES

Loss of genetic diversity in the North American mayfly *Ephemerella invaria* associated with deforestation of headwater streams

LAURIE C. ALEXANDER*, DAVID J. HAWTHORNE†, MARGARET A. PALMER† AND WILLIAM O. LAMP†

*U.S. Environmental Protection Agency, Washington, DC, U.S.A.

†Department of Entomology, University of Maryland, College Park, Maryland, U.S.A.

SUMMARY

1. Terrestrial dispersal by aquatic insects increases population connectivity in some stream species by allowing individuals to move outside the structure of the stream network. In addition, individual survival and reproductive success (as well as dispersal) are tightly linked to the quality of the terrestrial habitat.

2. In historically forested catchments, deforestation and altered land use have the potential to interfere with mayfly dispersal or mating behaviours by degrading the quality of the terrestrial matrix among headwater streams. We hypothesised that loss of tree cover in first-order catchments would be associated with an increase in population substructure and a decrease in genetic diversity of mayfly populations.

3. To test this hypothesis, we investigated spatial patterns of genetic variation in the common mayfly *Ephemerella invaria* across a gradient of deforestation in the central piedmont region of eastern United States. Intraspecific genetic diversity and population substructure were estimated from data obtained using fluorescent amplified fragment length polymorphism (AFLP) markers.

4. We found that mayfly populations had low population substructure within headwater stream networks and that genetic diversity was strongly negatively correlated with mean deforestation of the first-order catchments. The large-scale pattern of population substructure followed a pattern of isolation by distance (IBD) in which genetic differentiation increases with geographical distance, but assignment tests placed a few individuals into populations 300 km away from the collection site.

5. Our results show that loss of genetic diversity in this widespread aquatic insect species is co-occurring with deforestation of headwater streams.

6. Most arguments supporting protection of headwater streams in the United States have centred on the role of these streams as hydrological and biogeochemical conduits to downstream waters. Our work suggests that headwater stream land use, and specifically tree cover, may have a role in the maintenance of regional genetic diversity in some common aquatic insect species.

Keywords: amplified fragment length polymorphism (AFLP), aquatic insects, deforestation, Ephemeroptera, headwater streams

Correspondence: Laurie C. Alexander, U.S. Environmental Protection Agency, 1200 Pennsylvania Avenue, NW (8623-P), Washington, DC 20460 U.S.A. E-mail: alexander.laurie@epa.gov

Introduction

Recent reviews have highlighted the effects of land use in headwater stream catchments on water quality and ecosystem function in downstream rivers, estuaries and seas (Alexander *et al.*, 2007; Craig *et al.*, 2008; Brakebill, Ator & Schwarz, 2010). The contributions of headwater streams to biodiversity in river networks, reviewed by Meyer *et al.* (2007), also are affected by local land use. Lower diversity in stream assemblages has been associated with deforestation (Harding, 2003; Death & Collier, 2010) and urbanisation (e.g. Lenat & Crawford, 1994; Moore & Palmer, 2005) of small stream catchments. Deforestation leads to higher stream temperatures, increased runoff, altered water chemistry and reduced allochthonous inputs (Allan, 2004). Sweeney (1992) argues that in the historically forested small streams of the central piedmont region of North America, 'the presence or absence of trees adjacent to stream channels is one of the single most important factors altered by humans that affects the structure and function of stream ecosystems'. Catchment restoration programmes in temperate zones have placed high priority on replanting and conserving trees, especially along stream corridors where vegetated riparian buffers help reduce the movement of pollutants to waterways (Hassett *et al.*, 2005).

Aquatic insects are sensitive to conditions in the adjacent landscape as well as in the stream, because many species emerge from streams as adults to feed, find mates, colonise new headwaters or immigrate into existing headwater populations (Sweeney, 1993; Bilton, Freeland & Okamura, 2001; Smith, Alexander & Lamp, 2009). Degradation of terrestrial environments adjacent to headwater streams that serve as adult habitat or corridors for migration and gene flow may be expected to influence genetic diversity and population substructure in aquatic insect species specialised on headwater habitats (Malmqvist, 2002; Hughes, Schmidt & Finn, 2009).

Here, we investigate patterns of genetic variation within and among populations of the mayfly *Ephemerella invaria* (Walker), 1853 (Ephemeroptera, Ephemerellidae) in forested and partially deforested headwater catchments in the Mid-Atlantic region of the eastern United States, where whole streams are being lost (USEPA, 2005; Elmore & Kaushal, 2008) and only 18% of wadeable streams are classified by the EPA as being in 'good' condition (USEPA, 2006). The genus *Ephemerella* is common in the study area (MDDNR, 2005) and is tolerant of agricultural land uses (Moore & Palmer, 2005; Utz, Hilderbrand & Boward, 2009). Predicting the effects of land use change on flight-capable aquatic insects depends in part on understanding how qualitative differences in the landscape affect terrestrial movement of individuals between aquatic habitat patches (Ricketts, 2001; Kupfer, Malanson & Franklin, 2006). In a review of genetic estimation of dispersal among populations of freshwater fauna, Hughes *et al.* (2009) identified four factors influencing genetic connectivity: position in the hierarchical structure of the stream network, geographical distance among populations, species life history and dispersal traits. Models of gene flow based on the interaction and relative importance of these factors provide a conceptual framework for predicting and evaluating effects of landscape alterations on population connectivity in streams.

In the absence of strong selection, dispersal (gene flow) increases the proportion of genetic variation within populations relative to the proportion of genetic variation among populations (population substructure). With gene flow, spatially subdivided populations become genetically similar and, if levels of gene flow are high enough, function as the same effective population. Without the homogenising effects of gene flow, isolated populations diverge genetically through time. Population substructure therefore is an inverse indicator of population interconnectivity. One process driving divergence of isolated populations is genetic drift, in which alleles (units of genetic variation) randomly drift to fixation (or loss). Different alleles may become fixed (or lost) in different populations, increasing the level of population substructure (Wright, 1931). Drift also reduces overall levels of genetic variation or genetic diversity, an indicator of species resilience to changing environmental conditions.

Under a model of dispersal in which local mayfly populations in different headwater streams are connected primarily by adult flight through forested landscapes, we expected genetic variation within stream populations to be higher, and genetic variation among stream populations lower, in forested catchments compared with partially deforested catchments, where barriers to dispersal may effectively isolate mayfly populations in different headwater

streams. Deforestation, an indicator of early land use change, affects in-stream habitat quality for mayfly nymphs and terrestrial habitat quality for mayfly adults. Therefore, we also expect genetic diversity to decrease with higher rates of deforestation in a headwater stream network, as reduced aquatic habitat and barriers to terrestrial dispersal decrease the size and number of populations, and increase their spatial fragmentation (Lowe, 2002; Lowe & Bolger, 2002; Grant, Lowe & Fagan, 2007).

Methods

Study system

E. invaria is widely distributed across eastern and central North America and is common in our study area in first to third-order streams with good water quality (L. Alexander, unpubl. data). Nymphs cling to root wads or other complex vegetation at the stream margins, although in rapid flow they also are abundant in gravel substrata. The life cycle is univoltine, with loosely synchronous emergence in the study area starting in mid-April and continuing through May. Subimagoes emerge at the stream surface and fly directly into high branches of riparian trees. In the laboratory, subimago and adult stages last 24–36 h each, for a total winged stage of 48–72 h (L. Alexander, unpubl. data). Females with extruded egg masses were frequently seen descending at dusk from tree canopies >12 m above the stream surface, but mating swarms were not observed. Oviposition behaviours are easily observed, and individual females may fly up or downstream before releasing egg masses into small riffles by dipping the tip of the abdomen into the stream surface. Eggs (300–500 per mass) sink quickly and adhere to the substratum, where they enter diapause until nymphs hatch in early autumn.

Based on its preference for small streams and brief terrestrial adult stage, we predicted that *E. invaria* would be especially sensitive to isolating effects of recent (<100 years) deforestation and degradation of small streams in this region. A DNA barcoding study using the mitochondrial cytochrome *c* oxidase subunit I (mtDNA COI) gene revealed 11.6% divergence among samples of *E. invaria* taken across eastern North America, but <0.01% divergence among samples from streams separated by 300 km in our study area (Alexander *et al.*, 2009). Given the wide distribution of a single haplotype, effects of historical gene flow are expected to be consistent with a model isolation by distance (IBD) in which geographical distance is the best predictor of genetic distance between mayfly populations in the study area.

Site selection and catchment analysis

In 2001, we surveyed mayfly communities in forested and deforested first- to third-order streams in the piedmont region of Maryland and Virginia. From that survey, we selected nine headwater stream networks (Table 1 and Fig. 1) that, in addition to having *E. invaria* present in at least one stream, collectively met the following study criteria: (i) tree cover spanning a gradient of deforestation (1–80%, Table 2) and (ii) inclusion of headwater stream networks in five 6-digit Hydrologic Unit Code (Seaber, Kapinos & Knapp, 1987) river catchments in Maryland and Virginia (Table 1). A headwater stream network is defined here as a set of co-located headwater streams connected to each other or to the same higher-order stream (Grant *et al.*, 2007).

Sample sites were nested within catchments of different sizes to determine the scale at which genetic population substructure existed (headwater stream, headwater stream network, 8- or 6- digit HUC catchment). For analysis of population substructure and IBD, the sampling unit is a headwater stream (n = 16, Table 1). Because population substructure was found to be low or absent within headwater streams networks, the sampling unit was adjusted to a headwater stream network (n = 9) for regression of genetic diversity on catchment land cover. To distinguish between the two scales at which samples were sampled and analysed, the term 'local population' will be used when referring to samples from one headwater stream. Unless otherwise specified, the term 'population' will refer to a headwater stream network (Table 1).

ArcView 3.3 (ESRI Inc., Redlands, CA, U.S.A.) was used to delineate first-order catchments and estimate values for catchment characteristics predicted to influence the distribution of genetic variation in resident mayfly populations. Spatial grids of flow direction and accumulation derived from 30 m U.S. Geological Survey (USGS) digital elevation models were used to estimate catchment area and stream altitude. Forest cover and impervious surface area

Table 1 Locations of sample sites

County and state	6-Digit watershed (HUC#)	8-Digit watershed (HUC#)	HW stream network ID (from Fig. 1) and name	HW stream ID and name	No. individuals analysed (pre-drought, post-drought)
Baltimore Co., MD	Gunpowder River (021308)	Pretty Boy Reservoir (02130806)	(1) Gunpowder State Park	(1.1) Slip Stream	17
	Patapsco River (021309)	Patapsco River Lower North Branch (02130906)	(2) Patapsco State Park	(2.1) Daniels Creek B (2.2) Daniels Creek C	16, 16 12,16
Howard Co., MD	Patuxent River (021311)	Middle Patuxent River (02131106)	(3) Middle Patuxent Environmental Area	(3.1) Little Creek (3.2) Right Stream (3.3) T-West Branch	16, 24 16, 16 15
			(4) UMD Dairy Farm	(4.1) Folly Quarter Stream (4.2) South Stream	16 17,8
		Rocky Gorge Dam (02131107)	(5) Rocky Gorge South	(5.1) Rocky Gorge Tributary	16, 16
		Brighton Dam (02131108)	(6) Cattail Creek	(6.1) Hunt Valley Stream (6.2) Miller's Mill Stream	16, 8 16, 16
Montgomery Co., MD	Middle Potomac River (021402)	Seneca Creek (02140208)	(7) Seneca Creek State Park	(7.1) Schaeffer Stream (7.2) Black Rock Stream	8 20, 16
Appomattox Co., VA	James River (020802)	Appomattox River (02080207)	(8) Holliday Lake State Park	(8.1) Saunders Creek (8.2) Mossy Oak Stream	17 16, 8
Buckingham Co., VA	James River (020802)	Slate River (0208020)	(9) AppBuck. State Forest, Jamison Creek	(9.1) Big Jamie Creek	12, 12

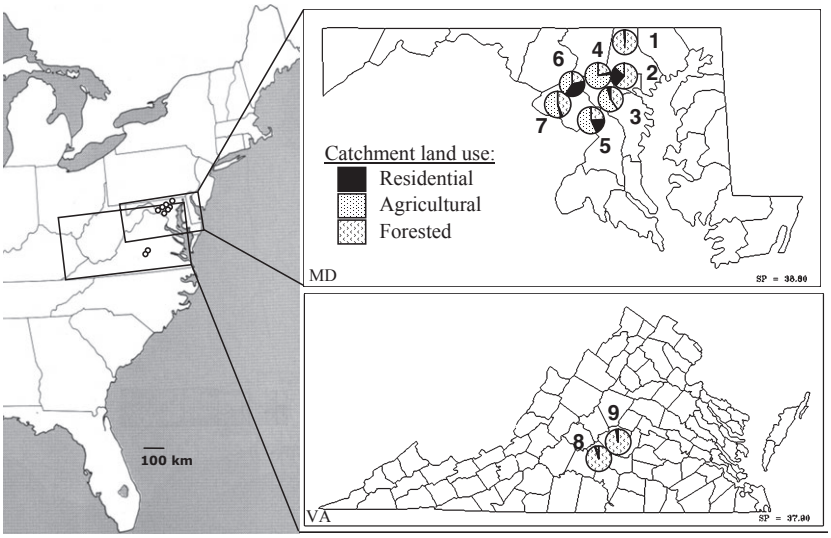

Fig. 1 Map of sample sites in (a) Maryland and (b) Virginia, U.S.A. Circles represent approximate locations of sampled headwater stream networks. The proportions of catchment area land uses, estimated from GIS coverages, are shown in pie charts (residential = black, agricultural = dotted, forested = arrow pattern). Location data are given in Table 1.

(ISA) (e.g. roads, parking lots, rooftops or compacted soils) within each delineated catchment were estimated from the 2001 National Landcover Database (Homer *et al.*, 2004). Deforestation was calculated at 1 – the proportion of forested area in the catchment. Headwater stream networks were selected to be as similar as possible except for the amount of deforestation, which ranged from 1 to 80% of first-order catchment area. Deforestation, a surrogate for stream impairments commonly found in agricultural and

Table 2 Catchment characteristics

HW stream network	Deforest (0–1)	%ISA (0–100)	Alt (m a.s.l.)	Area (km^2)
(1) Gunpowder State Park	0.01	0.30	438	1.04
(2) Patapsco State Park	0.39	8.30	359	0.93
(3) Middle Patuxent Environmental Area	0.46	16.20	408	0.79
(4) UMD Dairy Farm	0.80	1.25	375	0.67
(5) Rocky Gorge South	0.68	6.90	523	1.52
(6) Cattail Creek	0.78	10.25	522	2.45
(7) Seneca Creek State Park	0.60	1.15	392	1.43
(8) Holliday Lake State Park	0.03	0.10	558	0.81
(9) AppBuck. State Forest, Jamison Creek	0.05	0.05	574	2.23

'HW Stream Network' refers to Table 1. Variables: Deforest, average proportion of deforested area; %ISA, average percentage of impervious surface area; Alt, average altitude; Area, average catchment area.

low-density residential systems (e.g. sedimentation, nutrient loading, channelisation) and a direct estimate of the density of tree cover between headwater streams, is the land use/land cover variable of primary interest for this study. ISA, a common surrogate for urbanisation of stream catchments, is included here as a covariate.

Sampling

We sampled *E. invaria* from reaches within 16 streams in nine headwater stream networks over a period of 4 years (2001–2004). In 2001, 2002 and 2004, mayfly samples were collected using moss-packs (colonising samplers) consisting of 2.5 g dried moss enclosed in plastic mesh bags and tied with string to roots or stakes along the stream margin for 3 weeks in March and April, when late instar nymphs are present in the stream margins. Individuals were counted and sexed as male or female based on dimorphisms in eye colour and structure visible in late instar nymphs. Moss-packs are designed to move freely with streamflow to imitate natural moss or root-wad habitats. They are readily colonised by *E. invaria* and other aquatic invertebrate taxa. Eight moss-packs were placed in each stream, positioned in pairs along a 75-m reach so that a total of four sub-samples were taken in each stream. The pairing of moss-packs was done to provide redundancy in case one moss-pack was buried, lost or moved out of the flow. The actual location of a moss-pack pair was selected arbitrarily within the reach. Reach location was recorded with a GPS waypoint.

Samples were bagged in stream water and processed alive. When a stream sample contained fewer than 16 individuals total, that stream was re-sampled with a D-frame net, in an attempt to increase the size of the sample available for genetic analysis. These samples were labelled as D-frame samples and stored separately from the moss-pack samples.

In 2003, nymph samples were collected with a D-frame net. A starting point along the stream was selected at random to define the start of a 150 m reach, divided into six sections 25 m in length. Of these sections, three were selected at random for sampling. All habitats suitable for *E. invaria* were sampled exhaustively within the three randomly selected sections.

Specimen collection and preservation

Fresh samples were preserved in 100% ethanol, stored at ambient temperature during transit and put into long-term storage −20 °C. Heads were removed for DNA extraction and bodies (thoraces + abdomens) were labelled and stored as vouchers. Prior to extraction, the heads were frozen in liquid nitrogen and pulverised. DNA was extracted using the DNEasy kit and protocol (Qiagen, Chatsworth, CA, U.S.A.).

AFLP fragment construction and amplification

Amplified fragment length polymorphism (AFLP) (Vos *et al.*, 1995) fragments for polymerase chain reactions (PCR) were constructed by mixing 10 μL genomic DNA, Fermentas Buffer 'O' (Fermentas, Burlington, ON, U.S.A.), 10 mM ATP, 10 units PstI enzyme, 10 units EcoRI enzyme, 2 units T4 DNA ligase and 5 pmols of each double-stranded adapter (Table 3). The mixture was incubated overnight at 37 °C in a shaker oven. The restriction and ligation steps of the AFLP reaction were performed

Table 3 Amplified fragment length polymorphism adapter and primer sequences

Adapter or primer	Sequence	No. of polymorphic loci (Total = 902)
*Eco*RI adapters	5′-AAT TGG TAC GCA GTC-3′	
	5′-CTC GTA GAC TGC GTA CC-3′	
*Pst*I adapters	5′-TGT ACG CAG TCT TAC-3′	
	5′-CTC GTA GAC TGC GTA CAT GCA-3′	
**6-FAM labelled* *Eco*RI primer	(*E1) 5′-GAC TGC GTA CCA ATT CAG-3′	
*Pst*I primers	(P1) 5′-GAC TGC GTA CAT GCA GAC A-3′	471
	(P2) 5′-GAC TGC GTA CAT GCA GAG A-3′	431

simultaneously in a total reaction volume of 25 μL. The first amplification reaction (pre-selective amplification) was run with primers that are complementary to the adapter sequence only. The second amplification (selective amplification) was done using primers that had two or three overhanging nucleotides at the 3′ end. Both reactions were run in a standard PCR cocktail [20 mm Tris-HCl (pH 8.4), 50 mm KCl, 1.5 mm $MgCl_2$, 0.2 mm dNTPs, 1 unit Taq DNA polymerase] that included 5 pmols of each primer (Table 3). The pre-amplification reaction contained 5 μL of the AFLP construct (diluted 1 : 2 with ultrapure water) as template and was cycled 20 times for 1 min at 94 °C, 1 min at 56 °C and 1.5 min at 72 °C. The selective amplification contained 2 μL of 10 : 1 diluted pre-amplification product as template and was run as a touchdown-PCR in which the annealing temperature was high (65 °C) for the first round and then reduced 0.7 °C for each of the next 12 cycles. The denaturing and extension stages for each cycle were 94 °C for 10 s and 72 °C for 90 s, respectively. This ramping-down of the annealing temperature was followed by 25 cycles of 94 °C for 10 s, 56 °C for 40 + 1 s per cycle and 72 °C for 90 s.

AFLP fragment detection and analysis

Fluorescent dye-labelled selective amplification products were diluted 10 : 1 with ultrapure water prior to capillary electrophoresis on an ABI 3730 DNA Sequencer (Applied Biosystems, Foster City, CA, U.S.A.). Electrophoresis mixtures consisted of 2 μL of 10 : 1 diluted PCR product and 8 μL of deionized formamide and X-rhodamine-labelled MapMarker®1000 size standard (BioVentures Inc., Murfreesboro, TN, U.S.A.) in a 150 : 1 ratio. Data collection, processing, fragment sizing and pattern analysis were done using the AFLP functions provided by ABI in GeneMapper version 3.7 software (Applied Biosystems, Foster City, CA, U.S.A.), with factory default settings except as described here. Only fragments in the range from 50 to 600 bp were analysed. Markers were selected by initially setting the peak detection threshold to 50 relative fluorescence units (rfu) and the Genemapper 3.7 'allele-calling' threshold to 300 rfu. The project panel and bin set produced were then applied to the same set of individuals with the allele-calling threshold reset to 100 rfu for genotyping. To further reduce the number of spurious and low-frequency peaks, loci at which the presence of a fragment occurred in fewer than 5% of the total number samples were deleted from the data matrix. Positive and negative controls were included on each plate, and positive controls were compared for consistency. Two primer pairs, each with >400 polymorphic loci (Table 3), were used to generate and analyse genotypes for all individuals.

Population genetic substructure and genetic diversity

The AFLP presence/absence matrix was analysed with the software programs Hickory 1.0.5 (beta version) (Holsinger, Lewis & Dey, 2002) and Arlequin 3.0 (Excoffier, Laval & Schneider, 2005). Average panmictic heterzygosity (*hs*), a Bayesian estimate of average heterozygosity within local populations, and population genetic substructure (θ_B) (specifically the statistic θ_{II}, Bayesian analog to Fst based on Weir & Cockerham (1984)) estimates were obtained from Hickory using the 'f-free' model and uniform priors. The f-free model is recommended for dominant markers by the authors of Hickory because it incorporates uncertainty about the value of *f* [Bayesian analog to Wright's F_{is}, the average inbreeding coefficient (Wright, 1951)] into the

estimation of genetic diversity and population substructure. Because estimation of heterozygosity from dominant markers may be misleading, even without assumptions about the inbreeding coefficient, as a cautionary step we included a second estimate of local population diversity, mean pairwise distance (PD) between individuals, calculated in Arlequin (mean pairwise difference, π).

We evaluated the relationship of genetic diversity to five environmental variables (catchment area, deforestation, ISA, stream altitude, latitude) for each headwater stream network by means of multiple regression in R (R Development Core Team, 2009). Regression models for dependent variables *hs* and *PD* were selected using the function *step* (in R package *stats*).

Partial Mantel tests (Mantel, 1967) to evaluate the correlation of genetic distances and geographical distances among local populations (after controlling for the effect of deforestation) were run in R using the package *vegan* and online using the IBD Web Service (Jensen, Bohonak & Kelley, 2005). Three distance matrices were correlated: pairwise Fst among 16 local populations; estimated in AFLP-SURV (Vekemans, 2002), pairwise geographical distance among headwater streams, estimated as the Euclidean distance between pairs of geographical coordinates; and as a modifying factor, pairwise differences in deforestation within first-order catchments. The deforestation distance matrix was generated using the *dist* function in R (package *stats*).

Structure version 2.3.3 (Pritchard, Stephens & Donnelly, 2000), a model-based clustering method, was used to examine admixture in individuals. Structure assigns individuals to ancestral gene pools characterised by allele frequencies at each locus, without prior information about geographical location or genetic substructure from other analyses. Under a model of admixture, one individual may be fractionally assigned to multiple ancestral groups; thus, Structure can be used to identify putative migrants or descendents of migrants in each local population. We ran five iterations of the Structure analysis for $K = 2$ to $K = 10$, with 50 000 burn-in cycles and 100 000 Markov chain Monte Carlo (MCMC) replications each.

Results

A total of 402 individuals were analysed across streams and sample years. Sex ratios were 1 : 1 in all nymph samples, and we found no evidence of parthenogenesis in local populations. The AFLP method produced a total of 902 useable polymorphic loci from both primer pairs (Table 3).

We found high population substructure among headwater streams separated by >100 km ($\theta_B = 0.189$ with 95% credible interval = 0.174–0.205), but moderate population substructure between streams separated by 10–100 km ($\theta_B = 0.090$ with interval = 0.080–0.100). Within headwater stream networks (<10 km), population substructure was low ($\theta_B = 0.045$ with interval 0.028–0.063), suggesting that observations within networks (i.e. local populations) may not be independent. Therefore, for regression analysis, *hs* and *PD* for each headwater stream network were estimated as the average of local populations across sampling years (Table 1). Catchment area, deforestation, ISA and stream altitude were estimated as the average of first-order catchments in that network (Table 2).

Mayflies were absent from some streams in some years, in part because the sampling period spanned a severe drought during which surface flow in some streams stopped (Alexander & Lamp, 2008). However, estimates of local population heterozygosity in streams sampled before and after the drought did not differ significantly (paired *t*-test, $P = 0.46$, $n = 9$ headwater streams), so mayfly samples taken in different years were treated as sub-samples of each local population.

Average polymorphism was high (65%) in all local populations. Heterozygosity (*hs*) ranged from a low of 0.2185 at an agricultural site in Maryland to a high of 0.2517 in a fully forested site in Virginia. Pairwise difference (*PD*) followed the same pattern as *hs* and ranged from a low of 105 to a high of 126.

Model selection for regression analysis of genetic diversity (*hs, PD)* and catchment-scale influences evaluated five predictor variables: catchment area, deforestation, ISA, stream altitude and latitude. The final models were selected from competing models using forward and backward selection with Akaike's information criterion (AIC). The final model for *PD* retained deforestation as the only predictor variable. Owing to nonlinear trends in the residuals, the final model for *hs* was a second-order polynomial with deforestation as the only predictor variable.

Both measures of genetic diversity were significantly higher in forested headwater stream networks

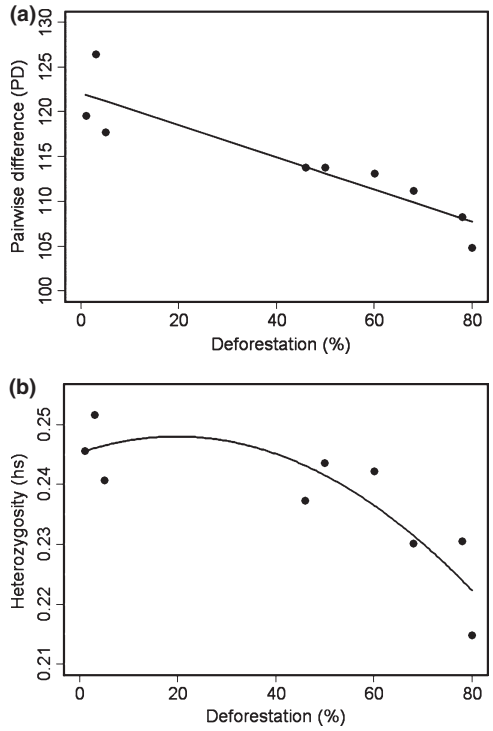

Fig. 2 Linear regressions of genetic diversity (PD, hs) on catchment deforestation (*def*). A, pairwise distance (*PD*): adj.R^2 = 0.81, $P < 0.001$; $\hat{Y}_{PD}$ = 122.09 − 17.93**def*. B, heterozygosity (*hs*): adj.R^2 = 0.67, P = 0.01; $\hat{Y}_{hs}$ = 0.2453 + 0.000279**def* − 0.000007**def*^{2}.

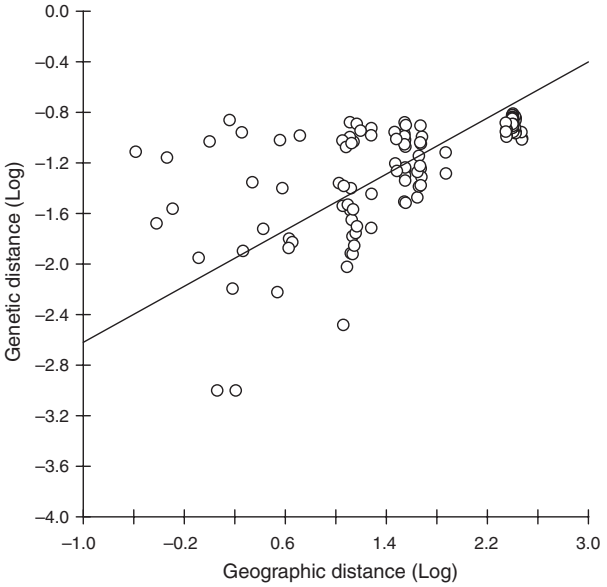

Fig. 3 Reduced major axis (RMA) regression of log-transformed Fst over log-transformed geographical distance (km) for 16 local populations.

compared with those flowing through partially deforested agricultural and residential areas: Estimated heterozygosity (*hs*) and mean *PD* of mayfly populations were both negatively correlated with the percentage of deforestation (Fig. 2, adj.R^2 = 0.67, P = 0.01 and adj.R^2 = 0.81, $P < 0.001$, respectively). Regression of local population heterozygosity (not averaged) produced a similar trend (n = 16, adj.R^2 = 0.60, P = 0.001), confirming that this result is not an artefact of averaging different estimates of genetic diversity within networks.

Isolation by distance

The large-scale pattern of population substructure for *E. invaria* across the 300-km study range is consistent with a process of IBD in which genetic differentiation increases with geographical distance as the influence of genetic drift becomes gradually stronger than the homogenising influence of gene flow (Fig. 3). The correlation of geographical distance and genetic distance (here, pairwise Fst) was significant after controlling for local effects of deforestation (r = 0.56, $P < 0.001$). IBD slopes and intercepts may be used to visualise differences; here, Fig. 3 shows a uniform rate of change with distance across the sampled populations.

Assignment tests

Structure 2.3.3 identified two gene pools, one in Maryland and the other in Virginia, with limited admixing in most individuals (Fig. 4). The proportion of Maryland ancestry in individuals collected in Virginia ranged from 0.003 to 0.132, with an average of 0.02. Excluding 5 Maryland individuals with >80% Virginia ancestry (see populations 1 and 7 in Fig. 4), the proportion of Virginia ancestry assigned to the individuals collected in Maryland ranged from 0.002 to 0.504, with an average of 0.11. Admixed individuals are putative migrants carrying alleles from the Virginia gene pool into the Maryland populations.

Populations in the intermediate (50–80% forested) land use streams are of particular interest in this analysis of population genetic variation and distribution because the AFLP data show that these intermediate populations are as genetically diverse (Fig. 2) as the populations in fully forested catchments.

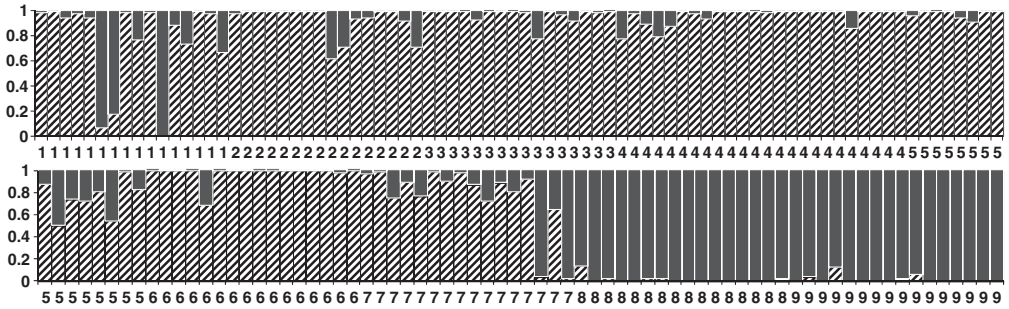

Fig. 4 Individual assignment to inferred gene pool clusters. Each vertical bar represents one individual. The proportion (0–1) of an individual's Virginia ancestry is shown in dark gray; the proportion (0–1) of Maryland ancestry is shown in cross-hatched pattern. Numbers on the *x*-axis refer to headwater stream network ID (Table 1).

Assignment tests that placed individuals in the 'Virginia' or 'Maryland' gene pool based on Bayesian analysis of estimated allele frequencies at each AFLP locus revealed that the intermediate and forested catchments in Maryland contain the highest proportion of individuals of admixed ancestry and that local populations at the low end of the scale (deforested and less diverse) in Maryland had no strongly admixed individuals (Fig. 4).

Discussion

Based on its preference for small streams with good water quality and brief terrestrial adult stage, we predicted that *E. invaria* would be especially sensitive to isolating effects of recent deforestation and degradation of small streams in the study region. The IBD pattern agrees with a model of widespread gene flow for *E. invaria* (Hughes *et al*., 2009), in which potential limitations including the species' univoltine life cycle, short adult stage and specialisation on headwater habitats are countered by landscape factors (e.g. density and distribution of stream channels and terrestrial vegetation, shallow channels, patterns of wind flow) that facilitate movement of individuals.

We found that both measures of genetic diversity were strongly negatively related to the average area of deforestation in first-order catchments (Fig. 2). A reduction in the number or quality of headwater streams could lead to smaller effective population sizes by temporary but substantial reductions in census population sizes, or the occasional extinction and recolonisation of streams, causing the observed regional loss of intra-specific diversity through genetic drift. Larger populations support more diverse assemblages of genotypes, including rare alleles found at lower frequencies; and connectivity across small and large spatial and temporal scales enables the transport of alleles among populations in semi-isolated stream and river catchments. Cumulative habitat loss may be limiting dispersal as well, erecting barriers to dispersal and introducing isolation among surviving populations.

The emerging patterns of mayfly genetic diversity in partially deforested headwater streams are not the result simply of IBD, but rather appear to be a response to changes in the catchment vegetation. This may be due in part to a mismatch in the geometries of dispersal and terrestrial habitat structure, mating behaviour and terrestrial habitat structure, or both. Historically, stream networks in the central piedmont of the United States were dendritic and forests were broadly two dimensional. Today, surviving and restored forest patches largely follow stream corridors as riparian buffers. The re-arrangement of forests into dendritic structures confined to stream corridors could interact with the dispersal abilities and behaviours of stream insects with aerial life stages, imposing new constraints on movement (see review by Smith *et al*., 2009) and shifting populations towards a headwater model of gene flow described by Finn, Blouin & Lytle (2007) and reviewed by Hughes *et al*. (2009). Grant *et al*. (2007) defined specific properties of the dendritic ecological network that cause resident populations to respond differently to disturbance than those in habitats with other types (e.g. linear or lattice) of structures. For example, Fagan (2002) found that dendritic habitat structure increases fragmentation effects, with negative consequences for population persistence in desert fishes. Similarly, salamander

occupancy in headwater streams is influenced by the stream network configuration and landscape context, with higher occupancy in undisturbed streams that are connected to other small streams (Grant, Green & Lowe, 2009).

Deforestation (or tree cover) is a surrogate for multiple environmental disturbances co-occurring in developing catchments that contribute to stream degradation, habitat loss and fragmentation (Snyder, Goetz & Wright, 2005; Walsh *et al*., 2007). The finding that reduced genetic diversity in a common stream insect is linked to deforestation and habitat loss associated with agriculture and urban development adds to the growing body of scientific evidence that human-induced changes to the landscape have negative consequences for many species (Dirzo & Raven, 2003). Additionally, our findings may be particularly important given the recent controversies concerning the extent of protection of U.S. headwater streams under the Clean Water Act (Leibowitz *et al*., 2008). Most arguments supporting protection of headwater streams have centred on the hydrological and biogeochemical contributions of small stream to larger, 'navigable' waters (Nadeau & Rains, 2007). Our work suggests that headwater streams and their attendant riparian and upland communities of trees may also have a role in maintaining regional genetic diversity. Since low genetic diversity is known to contribute to extinction risk (Frankham, 2005), the loss of small streams from river networks may be contributing to the declining biodiversity documented for stream insects in developed catchments.

Acknowledgments

We thank the University of Maryland Center for Biodiversity and the Gahan Fellowship, Department of Entomology, University of Maryland for funding this project. We also thank David H. Funk, Stroud Water Research Center, and Steven Burian, Southern Connecticut State University, for information on the systematics and ecology of ephemerellid mayflies; Glenn Moglen, Virginia Polytechnic Institute and State University, for help with watershed analysis; the Middle Patuxent Valley Association in Columbia, Maryland, and Virginia Department of Forestry for access to field sites; and Merrill Peterson at Western Washington University and two anonymous referees for comments on the manuscript.

References

Alexander L.C. & Lamp W.O. (2008) Mayfly population density, persistence, and genetic structure in fragmented headwater habitats. In: *International Advances in the Ecology, Zoogeography and Systematics of Mayflies and Stoneflies* (Eds F.R. Hauer , J.A. Stanford & R.L. Newell), pp. 39–50. University of California Press, Berkeley, CA.

Alexander R.B., Boyer E.W., Smith R.A., Schwarz G.E. & Moore R.B. (2007) The role of headwater streams in downstream water quality. *Journal of the American Water Resources Association*, **43**, 41–59.

Alexander L.C., Delion M., Hawthorne D.J., Lamp W.O. & Funk D.H. (2009) Mitochondrial lineages and DNA barcoding of closely related species in the mayfly genus *Ephemerella* (Ephemeroptera: Ephemerellidae). *Journal of the North American Benthological Society*, **28**, 584–595.

Allan J.D. (2004) Landscape and riverscapes: the influence of land use on river ecosystems. *Annual Review of Ecology, Evolution and Systematics*, **35**, 257–284.

Bilton D.T., Freeland J.R. & Okamura B. (2001) Dispersal in freshwater invertebrates. *Annual Review of Ecology and Systematics*, **32**, 159–181.

Brakebill J.W., Ator S.W. & Schwarz G.E. (2010) Sources of suspended-sediment flux in streams of the Chesapeake Bay Watershed: a regional application of the SPARROW Model1. *Journal of the American Water Resources Association*, **46**, 757–776.

Craig L.S., Palmer M.A., Richardson D.C., Filoso S., Bernhardt E.S., Bledsoe B.P. *et al.* (2008) Stream restoration strategies for reducing river nitrogen loads. *Frontiers in Ecology and the Environment*, **6**, 529–538.

Death R.G. & Collier K.J. (2010) Measuring stream macroinvertebrate responses to gradients of vegetation cover: when is enough enough? *Freshwater Biology*, **55**, 1447–1464.

Dirzo R. & Raven P.H. (2003) Global state of biodiversity and loss. *Annual Review of Environment and Resources*, **28**, 137–167.

Elmore A.J. & Kaushal S.S. (2008) Disappearing headwaters: patterns of stream burial due to urbanization. *Frontiers in Ecology and the Environment*, **6**, 308–312.

Excoffier L., Laval G. & Schneider S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.

Fagan W.F. (2002) Connectivity, fragmentation, and extinction risk in dendritic metapopulations. *Ecology*, **83**, 3243–3249.

Finn D.S., Blouin M.S. & Lytle D.A. (2007) Population genetic structure reveals terrestrial affinities for a headwater stream insect. *Freshwater Biology*, **52**, 1881–1897.

Frankham R. (2005) Genetics and extinction. *Biological Conservation*, **126**, 131–140.

Grant E.H.C., Lowe W.H. & Fagan W.F. (2007) Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters*, **10**, 165–175.

Grant E.H.C., Green L.E. & Lowe W.H. (2009) Salamander occupancy in headwater stream networks. *Freshwater Biology*, **54**, 1370–1378.

Harding J.S. (2003) Historic deforestation and the fate of endemic invertebrate species in streams. *New Zealand Journal of Marine and Freshwater Research*, **37**, 333–345.

Hassett B., Palmer M., Bernhardt E., Smith S., Carr J. & Hart D. (2005) Restoring watersheds project by project: trends in Chesapeake Bay tributary restoration. *Frontiers in Ecology and the Environment*, **3**, 259–267.

Holsinger K.E., Lewis P.O. & Dey D.K. (2002) A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology*, **11**, 1157–1164.

Homer C., Huang C., Yang L., Wylie B. & Coan M. (2004) Development of a 2001 National Landcover Database for the United States. *Photogrammetric Engineering and Remote Sensing*, **70**, 829–840.

Hughes J.M., Schmidt D.J. & Finn D.S. (2009) Genes in streams: using DNA to understand the movement of freshwater fauna and their riverine habitat. *BioScience*, **59**, 573–583.

Jensen J.L., Bohonak A.J. & Kelley S.T. (2005) Isolation by distance, web service. *BMC Genetics*, **6**, v.3.16, Available at http://ibdws.sdsu.edu/.

Kupfer J.A., Malanson G.P. & Franklin S.B. (2006) Not seeing the ocean for the islands: the mediating influence of matrix-based processes on forest fragmentation effects. *Global Ecology and Biogeography*, **15**, 8–20.

Leibowitz S.G., Wigington P.J., Rains M.C. & Downing D.M. (2008) Non-navigable streams and adjacent wetlands: addressing science needs following the Supreme Court's Rapanos decision. *Frontiers in Ecology and the Environment*, **6**, 366–373.

Lenat D.R. & Crawford J.K. (1994) Effects of land-use on water quality and aquatic biota of 3 North-Carolina piedmont streams. *Hydrobiologia*, **294**, 185–199.

Lowe W.H. (2002) Landscape-scale spatial population dynamics in human-impacted stream systems. *Environmental Management*, **30**, 225–233.

Lowe W.H. & Bolger D.T. (2002) Local and landscape-scale predictors of salamander abundance in New Hampshire headwater streams. *Conservation Biology*, **16**, 183–193.

Malmqvist B. (2002) Aquatic invertebrates in riverine landscapes. *Freshwater Biology*, **47**, 679–694.

Mantel N. (1967) Detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

MDDNR (2005) *Maryland Department of Natural Resources, 2000–2004 Maryland Biological Stream Survey*, Volume 9: Aquatic Biodiversity. MDDNR, Annapolis, MD, USA.

Meyer J.L., Strayer D.L., Wallace J.B., Eggert S.L., Helfman G.S. & Leonard N.E. (2007) The contribution of headwater streams to biodiversity in river networks. *Journal of the American Water Resources Association*, **43**, 86–103.

Moore A.A. & Palmer M.A. (2005) Invertebrate biodiversity in agricultural and urban headwater streams: implications for conservation and management. *Ecological Applications*, **15**, 1169–1177.

Nadeau T.L. & Rains M.C. (2007) Hydrological connectivity between headwater streams and downstream waters: how science can inform policy. *Journal of the American Water Resources Association*, **43**, 118–133.

Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ricketts T.H. (2001) The matrix matters: effective isolation in fragmented landscapes. *American Naturalist*, **158**, 87–99.

Seaber P.R., Kapinos F.P. & Knapp G.L. (1987) *Hydrologic Unit Maps: U.S. Geological Survey Water-Supply Paper 2294*. U.S. Geological Survey.

Smith R.F., Alexander L.C. & Lamp W.O. (2009) Dispersal by terrestrial stages of stream insects in urban watersheds: a synthesis of current knowledge. *Journal of the North American Benthological Society*, **28**, 1022–1037.

Snyder M.N., Goetz S.J. & Wright R.K. (2005) Stream health rankings predicted by satellite derived land cover metrics. *Journal of the American Water Resources Association*, **41**, 659–677.

Sweeney B.W. (1992) Streamside forests and the physical, chemical, and trophic characteristics of piedmont streams in eastern North-America. *Water Science and Technology*, **26**, 2653–2673.

Sweeney B.W. (1993) Effects of streamside vegetation on macroinvertebrate communities of White Clay Creek in eastern North-America. *Proceedings. Academy of Natural Sciences of Philadelphia*, **144**, 291–340.

USEPA (2005) *Final Programmatic Environmental Impact Statement on Mountaintop Mining/*Valley Fills in Appalachia. U.S. Environmental Protection Agency, Region 3, Philadelphia, PA.

USEPA (2006) *Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams*. Vol. EPA 841-B-06-002. United States Environmental Protection Agency, Washington, DC.

Utz R.M., Hilderbrand R.H. & Boward D.M. (2009) Identifying regional differences in threshold responses of aquatic invertebrates to land cover gradients. *Ecological Indicators*, **9**, 556–567.

Vekemans X. (2002) *AFLP-SURV Version 1.0*. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Vos P., Hogers R., Bleeker M., Reijans M., Vandelee T., Hornes M. *et al.* (1995) AFLP – A new technique for DNA-fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Walsh C.J., Waller K.A., Gehling J. & Mac Nally R. (2007) Riverine invertebrate assemblages are degraded more by catchment urbanisation than by riparian deforestation. *Freshwater Biology*, **52**, 574–587.

Weir B.S. & Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wright S. (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Wright S. (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

(Manuscript accepted 28 December 2010)