Genomic data reveal similar genetic differentiation between invertebrates living under and on a riverine floodplain

Steve Jordan, Amanda DelVecchia, Brian K. Hand, Laney Hayssen, Clark Nissley, Gordon Luikart, Jack Stanford

sdjordan@bucknell.edu
amanda.delvecchia@mso.umt.edu
brian.hand@flbs.umt.edu
ech015@bucknell.edu
can@udel.edu
gordon.luikart@mso.umt.edu
jack.stanford@flbs.umt.edu

Abstract
Historical and ecological factors influencing population connectivity of stygobiontic and amphibiontic organisms have long been a topic of intense research and speculation. Here we offer RAD-seq data to show that levels of connectivity of a winged stonefly with groundwater larvae and an obligate groundwater amphipod show similar levels of connectivity between floodplains separated by 40 km in the Flathead River basin of NW Montana, USA. This may be due to deep phreatic water connections that enhance movement of organisms between these populations.

1. Introduction

In a landmark 1974 paper, Stanford and Gaufin identified several species of large stonefly larvae from a shallow alluvial aquifer supplying domestic water to a Montana community. Over the ensuing decades, researchers have documented diverse communities of macroinvertebrates, meiofauna, and microbes in shallow aquifers in Montana and worldwide [1,2]. These communities include not only insects that spend some time above ground (amphibionts), but crustaceans, oligochaets, mites, and other taxa that never leave interstitial spaces deep below ground (stygobionts). These animals have been documented beneath riverine floodplains at depths of up to 10 meters and up to 1.5 km from the main river channels [3]. The stoneflies spend 1-3 years maturing in the aquifer before emerging as winged, reproductive adults with an aerial lifespan of only a few days [4].

Shallow aquifers are bounded geologically, and both stygobionts and amphibionts have limited above-ground dispersal capabilities, though to different extents. In addition to isolation, groundwater habitats present many other challenges to stygobiontic taxa, including lack of light and reduced autochthonous carbon fixation, variable water flow rates, and reduced nutrient and oxygen availability [5].

Recent stygobiont research has focused on the ecology of the shallow groundwater environment, noting the variable influence of many abiotic and biotic factors on their spatial distribution, including bedrock geology, soil permeability, water chemistry and quality, groundwater levels, adjacent surface flows, riparian vegetation, and climate, among

others [6–9]. Furthermore, several recent genetic studies suggest that long-term barriers to dispersal by groundwater species are widespread, even within drainages and in spite of potentially linking floods [10–13]. The impact of these barriers likely varies according to species' vagility [9].

In spite of recent work, the ability of these organisms to move within and between adjacent watersheds has remained stubbornly unclear. Life history may play a big role in population connectivity, with the retention of ephemeral, winged life stages by amphibionts offering dispersal advantages.

Cutting-edge genetic tools offer the possibility to infer the current and historical extent of organismal movement within and between floodplains. An understanding of this dispersal would contribute to a better biological conceptualization of connectivity along the river corridor, a major theme in river ecology and management (e.g. SHM paper). Here we address this issue with a state-of-the-art RAD-seq genomic dataset for two charismatic groundwater species from an exceptionally well-characterized system, two floodplains of the Flathead River in northwestern Montana.

2. Material and Methods

Study sites. The Nyack Floodplain Research Area is a 9x2 km floodplain located on the Middle Fork of the Flathead River. The Kalispell Floodplain is located approximately 40 km downstream, on the main stem of the Flathead River. Both of these locations are underlain by high-porosity alluvial aquifers that are entirely recharged by river water. The aquifers are likely not connected in the subsurface [14], but each is known to contain a diverse array of meiofauna and macroinvertebrates [15].

Taxa. We considered two taxa exemplifying the various life history types: *Paraperla frontalis*, a hyporheic stonefly with a winged adult stage, and *Stygobromus* spp., a stygobiont crustacean with an entirely hypogean life history that is so adapted to the aquifer environment that it is blind and lacks pigment. We hypothesized that *Paraperla* would display higher gene flow and less genetic structure between floodplains than *Stygobromus* because only *Paraperla* has an epigean stage that might allow it to disperse.

Genotyping and population genomics analysis. We prepared libraries for RAD sequencing from genomic DNA from *Paraperla* and *Stygobromus* individuals according to standard protocols using the restriction enzyme SbfI and unique 6bp barcodes for each sample [16]. We sequenced single-end reads on 192 samples in 2 sequencing lanes on an Illumina HiSeq machine. Raw sequences were filtered by quality where 90% of the bases in a sequence were required to have a quality score ≥ 20. *Stygobromus* were furthered trimmed to 80 base pairs due to observed quality score drop-off. We used process "radtags" from Stacks software package, version 1.19 [17] to sort reads by barcode and attempted to rescue barcodes and RAD tags and removing any reads with an uncalled base.

We *denovo* aligned filtered reads from each individual using the denovo_map.pl perl script in Stacks [17]. We used a minimum of 5 identical raw reads to create a stack and allowed 2 mismatches between loci when processing a single individual and between loci when

building the catalog. We also removed or broke-up highly repetitive RAD-Tags. We assigned diploid genotypes at each nucleotide position in each individual using the bounded maximum-likelihood method described by Catchen et al. [17] with an upper bound on sequencing error rate set to 0.1, a lower bound of 0.001, and a likelihood ratio significance level of α = 0.05. Finally, we used the populations program from Stacks to filter out any RAD tags with a read depth of less than 8 and a minor allele frequency (MAF) < 0.05.

We performed additional SNP pruning in the PLINK pipeline [18] by removing, for both species, all individuals with >50% missing data, and all loci genotyped in < 60% of individuals. We also removed all loci that were not present in 50% of individuals in each population for *Paraperla* (7) and *Stygobromus* (5). We also removed all tags that included more than 4 SNPs or were out of Hardy-Weinburg equilibrium (P < 0.001) in more than 50% of populations. Finally, we calculated pairwise FST values using GENEPOP [19,20] for population pairs and by using all individuals in each floodplain.

3. Results

Sequencing and genotyping. Using RAD-seq, we identified final sets of 806 SNPs for 90 *P. frontalis* individuals (7 populations) and 314 SNPs for 50 *Stygobromus* sp. individuals (5 populations). We found genetic differentiation between floodplains to be low for both species (FST = 0.004 for P. frontalis and FST = 0.000 for *Stygobromus*). These low FSTs were likewise reflected in low population pairwise FST values.

For *Paraperla* we found a significant number of reads with poor quality and removed 48% of our initial reads, for *Stygobromus* we removed 12% of the raw reads. For *Stygobromus*, we identified two distinct species, one that was predominatly on the Nyack and one that predominantly appeared on the Kalispell floodplain. We also observed low depth of coverage and a lower final SNP count despite a much larger (e.g., 2-3 fold) SNP catalog identified in Stacks for *Stygobromus* versus *Paraperla*. This is mostly likely because while the genome size of *Stygobromus* is still unknown, it may be quite large: several amphipods have some of the largest genomes sizes known [21]. For all subsequent analyses we used only those samples from the species that was present predominantly on the Nyack, with 37 individuals on the Nyack and 13 on the Kalispell floodplains. For *Paraperla* we identified a final set of 806 SNPs for 90 individuals (7 populations) and for *Stygobromus* we identified a final set of 314 SNPs for 50 individuals (5 populations).

Population genomics analysis. We calculated pairwise FST values within and between floodplains for both species. However, for *Stygobromus* we only kept 13 individuals from the Kalispell floodplain so we did not conduct within population pairwise FST tests for *Stygobromus* in this floodplain. We found genetic differentiation between floodplains to be low for both species (FST = 0.004 for *Paraperla* and FST = 0.000 for *Stygobromus*). These low FST were likewise reflected in low population pairwise FST values.

4. Discussion

There is a great deal that is still unknown about the life history of stygobionts and amphibionts in shallow alluvial aquifers. Information that might be easily gathered for epigean taxa is hard to come by for groundwater species. We have only begun to understand the complete timing of life-history patterns, including such fundamental questions as whether winged species are constrained to above-ground reproduction, or whether underground reproduction is possible. We also know little about organismal responses to challenging environmental conditions, and are assembling basic information on sources of organic carbon, and its flow through trophic levels.

This study represents one of the first applications of state-of-the-art genomic techniques to an aquatic invertebrate system in an effort to better understand local and regional population structure. Using hundreds of loci, we have shown a lack of regional structure in two species that is especially striking given their widely divergent life histories.

The stygobiontic life history of the amphipod genus *Stygobromus* suggests that it has limited dispersal capacity. It is an obligate groundwater taxon, with no epigean (above-ground) stages and *Stygobromus* individuals have never been found in surface waters (data not shown). And yet, we found a pairwise Fst value of 0.0 for comparisons between the Nyack and Kalispell floodplains, which are separated by 40 river km. This value is actually nominally lower than the Fst for winged *Paraperla* (0.004). Given that the Nyack floodplain is bounded by bedrock nickpoints on both its upper and lower end, it is surprising to see similar Fst estimates for an obligate groundwater taxon and a species with winged reproductives.

There are several possible explanations for this result, and we intend to explore each further. First, we have identified deep bedrock phreatic water channels underlying this region, and have long suspected that such connections may be linking distant floodplains. Such a habitat might be highly unfavorable for groundwater taxa, and yet our current results are consistent dispersal through it. Second, it is possible that Stygobromus individuals irregularly enter the drift, and pass bedrock nickpoints to become established in downstream floodplains.

In summary, we have applied powerful new genomics techniques to a fundamental biological question in a system that our group has studied for 40 years, and found a novel, surprising result. Ongoing work will extend these results with more robust taxonomic, genomic, and geographic sampling. We expect to better understand the roles of historical and ecological drivers of ground-water genomic and phylogenetic diversity in coming years.

References

1. Stanford, J. A. & Ward, J. V 1993 An ecosystem perspective of alluvial rivers: connectivity and the hyporheic corridor. **12**, 48–60.
2. Boulton, A. J., Datry, T., Kasahara, T., Mutz, M. & Stanford, J. A. 2010 Ecology and management of the hyporheic zone: stream–groundwater interactions of running waters and their floodplains. *J. North Am. Benthol. Soc.* **29**, 26–40. (doi:10.1899/08-017.1)
3. Stanford, J. A. & Ward, J. V 1988 The hyporheic habitat of river ecosystems. *Nature* **335**, 64–66. (doi:10.1038/335064a0)

4. Stewart, K. W. & Stark, B. P. 2002 *Nymphs of North American Stonefly Genera (Plecoptera). Second Edition.* Stewart, Kenneth W.; University of North Texas, Denton, USA, USA.: Caddis Press.
5. Tockner, K., Pusch, M., Borchardt, D. & Lorang, M. S. 2010 Multiple stressors in coupled river-floodplain ecosystems. *Freshw. Biol.* **55**, 135–151. (doi:10.1111/j.1365-2427.2009.02371.x)
6. Larned T., S., Unwin, M. J. & Boustead, N. C. 2015 Ecological dynamics in the riverine aquifers of a gaining and losing river. *Freshw. Sci.* **34**, 245–262. (doi:10.1086/678350)
7. Korbel, K. L. & Hose, G. C. 2015 Habitat, water quality, seasonality, or site? Identifying environmental correlates of the distribution of groundwater biota. *Freshw. Sci.* **34**, 329–343. (doi:10.1086/680038)
8. Stubbington, R., Boulton, A. J., Little, S. & Wood, P. J. 2015 Changes in invertebrate assemblage composition in benthic and hyporheic zones during a severe supraseasonal drought. *Freshw. Sci.* **34**, 344–354. (doi:10.1086/679467)
9. Johns, T., Jones, J. I., Knight, L., Maurice, L., Wood, P. & Robertson, A. 2015 Regional-scale drivers of groundwater faunal distributions. *Freshw. Sci.* **34**, 316–328. (doi:10.1086/678460)
10. Finston, T. L., Johnson, M. S., Humphreys, W. F., Eberhard, S. M. & Halse, S. A. 2007 Cryptic speciation in two widespread subterranean amphipod genera reflects historical drainage patterns in an ancient landscape. *Mol. Ecol.* **16**, 355–365. (doi:10.1111/j.1365-294X.2006.03123.x)
11. Lefebure, T., Douady, C. J., Gouy, M., Trontelj, P., Briolay, J. & Gibert, J. 2006 Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments. *Mol. Ecol.* **15**, 1797–1806. (doi:10.1111/j.1365-294X.2006.02888.x)
12. Cooper, S. J. B., Bradbury, J. H., Saint, K. M., Leys, R., Austin, A. D. & Humphreys, W. F. 2007 Subterranean archipelago in the Australian arid zone: mitochondrial DNA phylogeography of amphipods from central Western Australia. *Mol. Ecol.* **16**, 1533–1544. (doi:10.1111/j.1365-294X.2007.03261.x)
13. Cooper, S. J. B., Saint, K. M., Taiti, S., Austin, A. D. & Humphreys, W. F. 2008 Subterranean archipelago: mitochondrial DNA phylogeography of stygobitic isopods (Oniscidea : Haloniscus) from the Yilgarn region of Western Australia. *Invertebr. Syst.* **22**, 195–203. (doi:10.1071/IS07039)
14. Hauer, Fr., Stanford A., J. & Lorang S., M. 2007 Pattern and process in Northern Rocky Mountain headwaters: ecological linkages in the headwaters of the Crown of the Continent. *J. Am. Water Resour. Assoc.* **43**, 104–117.
15. Stanford, J. A., Ward, J. V & Ellis, B. K. 1994 Ecology of the alluvial aquifers of the Flathead River, Montana. In *Groundwater Ecology* (ed J. D. Gibert Dan L. Stanford,Jack A.), pp. 367–390. San Diego, California: Academic Press.
16. Miller, M. R., Brunelli, J. P., Wheeler, P. A., Liu, S., Rexroad ,III, C. E., Palti, Y., Doe, C. Q. & Thorgaard, G. H. 2012 A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Mol. Ecol.* **21**, 237–249. (doi:10.1111/j.1365-294X.2011.05305.x)
17. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. 2013 Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140.

(doi:10.1111/mec.12354)
18. Purcell, S. et al. 2007 PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575. (doi:10.1086/519795)
19. Raymond, M. & Rousset, F. 1995 GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248–249.
20. Rousset, F. 2008 GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106.
21. Rees, D. J., Dufresne, F., Glemet, H. & Belzile, C. 2007 Amphipod genome sizes: first estimates for Arctic species reveal genomic giants. *Genome* **50**, 151–158. (doi:10.1139/G06-155)