3-24-2020

Phylogenetic and Transcriptomic Analyses of Vision in Two Cave Adapted Crustaceans, *Asellus aquaticus* (Isopoda: Asellidae) and *Niphargus hrabei* (Amphipoda: Niphargidae).

Jorge L. Perez Moreno
Florida International University, jpere645@fiu.edu

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

Part of the [Biology Commons](#), and the [Evolution Commons](#)

Recommended Citation

Perez Moreno, Jorge L., "Phylogenetic and Transcriptomic Analyses of Vision in Two Cave Adapted Crustaceans, *Asellus aquaticus* (Isopoda: Asellidae) and *Niphargus hrabei* (Amphipoda: Niphargidae)." (2020). *FIU Electronic Theses and Dissertations*. 4378.
<https://digitalcommons.fiu.edu/etd/4378>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

PHYLOGENETIC AND TRANSCRIPTOMIC ANALYSES OF VISION IN TWO
CAVE ADAPTED CRUSTACEANS, *ASELLUS AQUATICUS* (ISOPODA:
ASELLIDAE) AND *NIPHARGUS HRABEI* (AMPHIPODA: NIPHARGIDAE).

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOLOGY

by

Jorge Luis Pérez Moreno

2020

To: Dean Michael R. Heithaus
College of Arts, Sciences and Education

This dissertation, written by Jorge Luis Pérez Moreno, and entitled Phylogenetic and Transcriptomic Analyses of Vision in Two Cave Adapted Crustaceans, *Asellus aquaticus* (Isopoda: Asellidae) and *Niphargus hrabei* (Amphipoda: Niphargidae), having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Jose Maria Eirin-Lopez

Joel Trexler

Mauricio Rodriguez-Lanetty

Kalai Mathee

Thomas Iliffe

Heather Bracken-Grissom, Major Professor

Date of Defense: March 23, 2020.

The dissertation of Jorge Luis Pérez Moreno is approved.

Dean Michael R. Heithaus
College of Arts, Sciences and Education

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2020

© Copyright 2020 by Jorge Luis Pérez Moreno

All rights reserved.

DEDICATION

To my ancestors.

לא עליך המלאכה לגמור, ולא אתה בן חורין ליבטל ממנה.

ACKNOWLEDGMENTS

This dissertation has been possible thanks to the continuous support and excellent mentorship from my PhD advisor, Dr. Heather Bracken-Grissom. I also thank the members of my committee for all their advice, be it academic or else, and for their endless patience throughout these long years: Dr. Jose Maria Eirin-Lopez, Dr. Kalai Mathee, Dr. Thomas I. Iliffe, Dr. Joel Trexler, and Dr. Mauricio Rodriguez-Lanetty. I am also eternally grateful to my Hungarian colleagues, Dr. Gergely Balazs, Dr. David Brankovits, and Balazs Lerner who, besides becoming my close friends, were absolutely invaluable for the research undertaken in this dissertation. I also extend my most sincere thanks to FIU's UGS and the Biology Graduate Department, in particular Dr. Timothy Collins, Dr. Lidia Kos, and Yoo Kyung Song for always being there to happily assist when navigating graduate school became challenging. I would also like to thank all the amazing non-FIU scientists that I had the pleasure to work alongside with during this journey: Dr. Megan Porter, Dr. Fernan Palero, Dr. Tamara Frank, and many others that I was lucky to meet during fieldwork and conferences. I also acknowledge the computational support of Dr. Cassian D'Cunha and staff members of FIU's High Performance Computing Cluster for the resources and assistance provided. My most sincere thanks to all the institutions and funding agencies that made this research possible (the National Science Foundation, the Cave Research Foundation, The Crustacean Society, the International Society for Subterranean Biology, the Society for Integrative and Comparative Biology, and the Center for Coastal Oceans Research at FIU). Last but not least I would like to thank all my CRUSTOMICS lab mates who always provided moral support, encouragement, and ears to vent to during those difficult moments –

special thanks to my PhD mates Dr. Laura Timm, Dr. Emily Warschefsky, Robert Ditter, Carlos Varela, and Charles Golightly, all of whom I now consider to be part of my family. Also, I extend my gratitude and best wishes to my CRUSTOMICS undergrad assistants whose help in the lab was fundamental, Blake Wilkins and Joan Glenny, thank you for allowing me to share part of my PhD research journey with you all.

ABSTRACT OF THE DISSERTATION

PHYLOGENETIC AND TRANSCRIPTOMIC ANALYSES OF VISION IN TWO CAVE ADAPTED CRUSTACEANS, *ASELLUS AQUATICUS* (ISOPODA: ASELLIDAE) AND *NIPHARGUS HRABEI* (AMPHIPODA: NIPHARGIDAE).

by

Jorge Luis Pérez Moreno

Florida International University, 2020

Miami, Florida

Professor Heather Bracken-Grissom, Major Professor

The unique characteristics of aquatic caves and of their predominantly crustacean biodiversity nominate them as ideal study subjects for evolutionary biology. The present dissertation capitalizes on a perfect natural experiment, the Molnar Janos thermal cave system in Budapest, Hungary. This intricate freshwater cave system and the immediately adjacent Malom Lake present the ideal opportunity to address questions of colonization, adaptation, and evolution. Despite marked environmental differences between the cave and surface waters, both localities are inhabited by natural populations of two emerging model cave species, the isopod *Asellus aquaticus* and the amphipod *Niphargus hrabei*. In the present dissertation, I first conduct an extensive literature review to examine and discuss the role that molecular methodologies have played in the study of cave biology. Additionally, I discuss the potential of “speleogenomic” methodologies to address long-standing questions in cave and evolutionary biology in fields such as biodiversity, phylogeography, and evolution. I then investigate the phylogeographic patterns and divergence-time estimates between surface and cave populations of the aforementioned

species to elucidate mechanisms and processes driving the colonization of subterranean environments. These populations' phylogenies then serve as robust frameworks on which to evaluate the transcriptional basis behind the divergence of traits involved in troglomorphy, namely vision. RNA sequencing approaches are used to identify and evaluate differences in the transcription of photoreception genes and pathways to in subterranean vs. surface populations. To achieve so, in a scalable manner suitable for modern sequencing technologies, here I produce a bioinformatics pipeline that allows for an accurate and efficient identification of genes present in a transcriptome that are involved in photoreception and visual pathways. I then use this bioinformatics pipeline to depict, in a phylogenetically informed context, the transcriptional basis behind photoreception and vision in *A. aquaticus* and *N. hrabei*, and the role these traits play in cave adaptation, and in the evolution of troglomorphy in the subphylum Crustacea. With the findings herein, the present dissertation aims to provide a framework for the discovery of evolutionarily significant molecular mechanisms that permit the survival and evolution of life in caves and other extreme environments.

TABLE OF CONTENTS

CHAPTER	PAGE
PREFACE	1
I. INTRODUCTION	2
Literature Cited	5
II. LIFE IN THE UNDERWORLD: ANCHIALINE CAVE BIOLOGY IN THE ERA OF SPELEOGENOMICS.....	7
Abstract	8
Introduction.....	9
Ecology and Biodiversity of Anchialine Caves	12
Current Advances and Future Prospects	23
Cave Biodiversity in the Molecular Era.....	26
Phylogeography of Anchialine Cave Ecosystems	28
Evolution of Troglomorphy	32
Concluding Remarks.....	36
Acknowledgments.....	37
References	38
Figure Captions	62
Figures.....	63
III. THE ROLE OF ISOLATION ON CONTRASTING PHYLOGEOGRAPHIC PATTERNS IN TWO CAVE CRUSTACEANS	67
Abstract	68
Background	69
Methods.....	73
Results.....	79
Discussion	82
Future Propsects.....	87
References	91
Tables	97
Figure Captions	100
Figures.....	102
IV. PHYLOGENETIC ANNOTATION AND GENOMIC ARCHITECTURE OF OPsin GENES IN CRUSTACEA	111
Abstract	112
Introduction.....	113
Methods.....	117
Results.....	122
Discussion	125
Concluding remarks	131

References	133
Tables	142
Figure Captions	147
Figures.....	149
V. TRANSCRIPTOMIC INSIGHTS INTO THE LOSS OF VISION IN MOLNÁR JÁNOS CAVE’S CRUSTACEANS.....	153
Abstract	154
Introduction.....	155
Methods.....	159
Results and Discussion	163
Conclusions	169
References	171
Tables	177
Figure Captions	180
Figures.....	181
VI. CONCLUSIONS	186
VITA.....	191

LIST OF TABLES

TABLE	PAGE
CHAPTER III	
1 Specimens, locations, and type of habitat in which the crustacean populations were sampled	97
2 Divergence time estimations between Molnar Janos Cave and phylogenetically closest surface populations of the peracarid crustaceans under study (thousands of years [95% H.P.D.]).....	98
3 Pairwise Genealogical Sorting Index estimates for each population pair of <i>Asellus aquaticus</i> . P-values assess significance that exclusive ancestry observed for every population pair is greater than that which would be observed at random.	98
4 Pairwise Genealogical Sorting Index estimates for each population pair of <i>Niphargus hrabei</i> . P-values assess significance that exclusive ancestry observed for every population pair is greater than that which would be observed at random.....	99
CHAPTER IV	
1 Raw data chosen for de novo transcriptome assembly and annotation of opsin proteins in <i>Hyalella azteca</i> and <i>Daphnia pulex</i>	142
2 Summary statistics for the de novo transcriptome assemblies produced as part of this study and the corresponding genome assemblies.....	142
3 Results of transcriptome completeness assessment by Benchmarking Universal Single-Copy Orthologs (BUSCO) using OrthoDB's Arthropoda database of orthologous genes.	143
4 Number of genes and respective isoforms, as defined by Trinity, recovered for each type of opsin in <i>Hyalella azteca</i> and <i>Daphnia pulex</i>	143
5 Identification and classification of <i>Hyalella azteca</i> and <i>Daphnia pulex</i> opsins based on phylogenetic inference and HMM profile alignments. Putative visual opsins are marked in black, under the "Visual" column, when both methodologies are in agreement and in gray when they differ.	144

CHAPTER V

1 Summary statistics for the de novo transcriptome assemblies produced as part of this study	177
2 Results of transcriptome completeness assessment by Benchmarking Universal Single-Copy Orthologs (BUSCO) using OrthoDB's Arthropoda database of orthologous genes.	178
3 Phototransduction pathway genes identified by PIA (Speiser et al. 2014) in the cave transcriptomes of <i>A. aquaticus</i> and <i>N. hrabei</i>	178
4 Amino acids present in sites known to influence spectral sensitivity for each identified opsin. Position numbers depicted refer to the equivalent sites in the reference rhodopsin sequences (the bovine <i>Bos taurus</i> and the squid <i>Todarodes pacificus</i>).....	179

LIST OF FIGURES

FIGURE	PAGE
 CHAPTER II	
1 Schematic representation of an anchialine cave system. A) “Blue hole”, “Cenote” or “Sinkhole” opening to the surface; B) Meteoric freshwater lens upper stratum; C) Halocline or mixing zone – often accompanied by a layer of hydrogen sulfide by-product of microbial activity; D) Hypoxic saltwater lower stratum – devoid of sunlight, food webs in this habitats are likely to depend on chemosynthetic microbial communities.....	62
2 Anchialine systems can be found in a range of different forms including (but not limited to): A) karst cave systems (Crystal Cave, Bermuda – photo by Jill Heinerth), B) lava tubes (Jameos del Agua, Lanzarote, Canary Islands, Spain – photo by Jill Heinerth), and C) pools (Angel Pool, Bermuda – photo by Tamara Thomsen).	62
3 Examples of various crustacean taxa found in anchialine caves: A) Parhippolyte sterreri (Decapoda), B) Pseudoniphargus grandimanus (Amphipoda), C) Bahalana caicosana (Isopoda), D) Tulumella sp. (Thermosbaenacea), E) Mictocaris halope (Mictacea), F) Bermudamysis speluncola (Mysida), G) Cumella abacoensis (Cumacea), H) Ridgewayia sp. (Calanoida), I) Spelaeoecia sp. (Ostracoda), J) Cryptocorynetes sp. (Remipedia). Photographs of anchialine crustaceans by Thomas M. Iliffe, PhD.	62
 CHAPTER III	
1 <i>Asellus aquaticus</i> displays contrasting phenotypes in and out of the cave, while <i>Niphargus hrabei</i> exhibits the same phenotype in both environments.....	102
2 Schematic illustration of our three main sampling localities within Budapest, Hungary. Red circles indicate exact sites within Molnár János Cave (Rákös Rock) and surface environments (Malom Lake and Danube River [Soroksár])..	103
3 Haplotype networks of <i>Asellus aquaticus</i> : nuclear (A- PseudoND2) and mitochondrial (B- 16S, 12S, and COI) loci. Node diameter and annotation denote sample sizes. Colours represent sampling locality, as illustrated in the legend...	104
4 Divergence time estimates (x axis in thousands of years) of <i>Asellus aquaticus</i> populations (calculated with a multi-locus coalescent model in *BEAST; outgroups not shown) and the distribution of its populations with relative mtDNA haplotype frequencies throughout Hungary. Phylogenetic and population tree analyses support	

the inclusion of the cave phenotype as part of the species, but with evident population structuring and marked morphological differences as a result of the cave environment.	105
5 Haplotype networks of <i>Niphargus hrabei</i> : nuclear (A- ITS; B- NaK) and mitochondrial (A- 16S and COI) loci. Node diameter and annotation denote sample sizes (alleles in the case of nuclear genes). Colours represent sampling locality, as illustrated in the legend.....	106
6 Divergence time estimates (x axis in thousands of years) of <i>Niphargus hrabei</i> populations (calculated with a multi-locus coalescent model in *BEAST; outgroups not shown) and the distribution of its populations with relative COI haplotype frequencies in our three main Hungarian sites and neighbouring populations. Phylogenetic and population tree analyses do not support any evident genetic structuring between cave and surface populations.....	107
7 An Extended Bayesian Skyline Plot illustrates the demographic history of the sampled Hungarian <i>Asellus aquaticus</i> populations over the last 800,000 years. The x-axis represents time before present in thousands of years, while the y-axis denotes the estimated population (θ) assuming a generation time of one year.....	108
8 An Extended Bayesian Skyline Plot of <i>Niphargus hrabei</i> 's demographic history, from 100,000 years ago to today, depicts a possible genetic bottleneck at approximately 10,000 years ago. The x-axis represents time before present in thousands of years, while the y-axis denotes the estimated population (θ) assuming a generation time of one year.....	109
 CHAPTER IV	
1 Maximum-Likelihood phylogeny of opsins estimated using putative opsin proteins identified by our annotation pipeline from the de novo transcriptome assemblies of <i>Hyalomma azteca</i> and <i>Daphnia pulex</i> , along with a dataset of reference opsin sequences. Clades are annotated with opsin types contained therein and, in the case of visual opsins, with their inferred spectral sensitivities.	148
2 Expanded view of the melanopsin, Arthropod LWS, and Arthropod SWS/UV clades. Large non-crustacean clades have been collapsed for readability. Support values correspond to SH-aLRT/aBayes/UFBoot, and are not shown when bootstrap support < 75.	149
3 Intron-Exon gene structure patterns of representative <i>Hyalomma azteca</i> SWS/UV and LWS opsins	150

4 Intron-Exon gene structure patterns of representative <i>Daphnia pulex</i> opsins SWS/UV and LWS opsins, which are arranged in the genome in distinct patterns according to opsin type.	151
--	-----

5 Exon size distribution for both <i>Daphnia pulex</i> and <i>Hyalomma azteca</i> . X-axis is in base pair units (bp), while Y-axis represents proportion of transcripts in said range....	152
--	-----

CHAPTER V

1 <i>Asellus aquaticus</i> is troglomorphic only within the cave environment, while <i>Niphargus hrabei</i> 's morphotype remains constant across environments (adapted from Pérez-Moreno et al. 2017).....	181
---	-----

2 Specimens of both species were sampled from three localities within Budapest, Hungary. Red circles indicate exact sites within Molnár János Cave (Rákos Rock) and surface environments (Malom Lake and Soroksár [Danube River]; adapted from Pérez-Moreno et al. 2017).....	182
---	-----

3 Circular phylogeny of identified and reference opsins illustrating major clades (denoted in numerals) that correspond to their functional classification.	183
--	-----

4 Expanded view of the Short Wavelength Sensitive & UV opsin clade. All populations of <i>A. aquaticus</i> expressed the same SWS/UV opsin regardless of their morphotype. In the case of <i>Niphargus hrabei</i> , SWS/UV opsins were not found in their transcriptomes.....	184
---	-----

5 Expanded view of the Long Wavelength Sensitive opsin clade. Both <i>A. aquaticus</i> and <i>N. hrabei</i> express functional opsins belonging to this clade with low intraspecific differentiation.	185
--	-----

PREFACE

The following chapters have been published in their entirety in peer-reviewed journals. According to the stated policy of each journal, they can be included in this dissertation without obtaining formal copyright permission, with the corresponding citations :

CHAPTER II

Pérez-Moreno JL, Iliffe TM, Bracken-Grissom HD. 2016. *Life in the Underworld: Anchialine cave biology in the era of speleogenomics*. International Journal of Speleology, 45(2): 149-170.

CHAPTER III

Pérez-Moreno JL, Balázs G, Wilkins B, Herczeg G, Bracken-Grissom HD. 2017. *The role of isolation on contrasting phylogeographic patterns in two cave crustaceans*. BMC Evolutionary Biology, 17:247.

CHAPTER IV

Pérez-Moreno JL, DeLeo D, Palero F, Bracken-Grissom HD. 2018. *Phylogenetic annotation and genomic architecture of opsin genes in Crustacea*. Hydrobiologia, 825(1): 159-175.

CHAPTER V

Pérez-Moreno JL, Balázs G, Bracken-Grissom HD. 2018. *Transcriptomic insights into the loss of vision in Molnár János Cave crustaceans*. Integrative and Comparative Biology, 58(3):452-464.

CHAPTER I
INTRODUCTION

The underlying mechanisms and processes that prompt the colonization of extreme environments constitute major research themes of evolutionary biology and modern biospeleology (Barr & Holsinger, 1985; Juan et al., 2010; Benvenuto et al., 2015; Pérez-Moreno et al., 2016). Subterranean habitats, with their truncated food webs, anoxic environments, and permanent darkness, often require specific adaptations from their inhabitants for them to survive in such hostile environments (Benvenuto et al., 2015). These special adaptations and the geographical isolation of cave habitats nominate cave biodiversity as logical study subjects to answer long-standing questions concerning the interplay among adaptation, biogeography, and evolution (Zhang & Li, 2013; Pérez-Moreno et al., 2016). The ability of a species to successfully colonize such extreme environments, however, might be mediated by ecological filtering and requires specific exaptations to life in darkness (Gould & Vrba, 1982; Gibert & Deharveng, 2002; Benvenuto et al., 2015). Exaptations to cave life include, for example, morphological (reduced dependence on vision, elongation of body and appendages) and physiological (tolerance to oligoxic conditions) characteristics that are already present in many species inhabiting benthic and interstitial ecosystems. However, to fully understand the mechanisms behind cave colonization events it is essential to incorporate approaches that also consider the phylogenetic history, ecology and distribution of the organisms under study, and therefore the factors that ultimately drive their evolution. The present dissertation capitalizes on a perfect natural experiment, the Molnar Janos thermal cave system in Budapest, Hungary. The intricate freshwater cave system and the immediately adjacent Malom Lake present the ideal opportunity to address questions of colonization, adaptation, and evolution. Despite marked environmental differences between the cave

and surface waters, both localities are inhabited by natural populations of two emerging model cave species, the isopod *Asellus aquaticus* and the amphipod *Niphargus hrabei*. In my dissertation, I first conduct an extensive literature review to examine and discuss the role that molecular methodologies have played in the study of cave biology. Additionally, I discuss the potential of “speleogenomic” methodologies to address long-standing questions in cave and evolutionary biology in fields such as biodiversity, phylogeography, and evolution. I then investigate the phylogeographic patterns and divergence-time estimates between surface and cave populations of the aforementioned species to elucidate mechanisms and processes driving the colonization of subterranean environments. These populations’ phylogenies then serve as robust frameworks on which to evaluate the transcriptional basis behind the divergence of traits involved in troglomorphy, namely vision. I used RNA sequencing approaches to identify and evaluate differences in the transcription of photoreception genes and pathways to in subterranean vs. surface populations. To achieve this objective, in a scalable manner suitable for modern sequencing technologies, I produced a bioinformatics pipeline that allows for an accurate and efficient identification of genes present in a transcriptome that are involved in photoreception and visual pathways. I then use this bioinformatics pipeline to depict, in a phylogenetically informed context, the transcriptional basis behind photoreception and vision in *A. aquaticus* and *N. hrabei*, and the role these traits play in cave adaptation, and in the evolution of troglomorphy in the subphylum Crustacea. With the findings herein, the present dissertation aims to provide a framework for the discovery of evolutionarily significant molecular mechanisms that permit the survival and evolution of life in caves and other extreme environments.

Intellectual Merit

The molecular mechanisms by which organisms adapt to their environments have long been sought to fully understand fundamental biological processes pertaining to the realms of organismal biology, ecology, and evolution. These mechanisms ultimately underlie species resilience and their ability to adapt to environmental challenges, changing ecological interactions, and disease, and as such have been subject of attention from the scientific community. However, the molecular basis behind adaptation in natural populations yet remains poorly understood, as the majority of work has been centered on examining a limited number of a priori selected genes, preliminary detection of selection at the gene and genomic level, and on model systems not representative of their natural states. Despite great progress facilitated by recent high-throughput sequencing technologies, important questions remain to be answered concerning the mechanisms of adaptation at the transcriptomic level. The identification of genes and pathways involved in photoreception and vision in Crustacea, in an appropriate phylogenetic framework, helps provide a solid bridge between genotype-phenotype, and aid in our understanding of patterns of molecular evolution in extreme environments and their role in the adaptive divergence in analogous systems.

Literature Cited

- Barr, T. C., & Holsinger, J. R. (1985). Speciation in Cave Faunas. *Annual Review of Ecology and Systematics*, 16(1), 313–337.
<https://doi.org/10.1146/annurev.es.16.110185.001525>
- Benvenuto, C., Knott, B., & Weeks, S. (2015). Crustaceans of extreme environments. In M. Thiel & L. Watling (Eds.), *The Natural History of the Crustacea* (Vol. 1–JANUARY, pp. 379–417). Oxford University Press.

- Gibert, J., & Deharveng, L. (2002). Subterranean Ecosystems : A Truncated Functional Biodiversity. *BioScience*, 52(6), 473–481.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology*, 8(01), 4–15.
- Juan, C., Guzik, M. T., Jaume, D., & Cooper, S. J. B. (2010). Evolution in caves: Darwin’s “wrecks of ancient life” in the molecular era. *Molecular Ecology*, 19(18), 3865–3880. <https://doi.org/10.1111/j.1365-294X.2010.04759.x>
- Pérez-Moreno, J. L., Iliffe, T. M., & Bracken-Grissom, H. D. (2016). Life in the Underworld: Anchialine cave biology in the era of speleogenomics. *International Journal of Speleology*, 45(2), 149–170. <https://doi.org/10.5038/1827-806X.45.2.1954>
- Speiser, D. I., Pankey, M., Zaharoff, A. K., Battelle, B. a, Bracken-Grissom, H. D., Breinholt, J. W., Bybee, S. M., Cronin, T. W., Garm, A., Lindgren, A. R., Patel, N. H., Porter, M. L., Protas, M. E., Rivera, A. S., Serb, J. M., Zigler, K. S., Crandall, K. a, & Oakley, T. H. (2014). Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinformatics*, 15(1), 350–350. <https://doi.org/10.1186/s12859-014-0350-x>
- Zhang, Y., & Li, S. (2013). Ancient lineage, young troglobites: recent colonization of caves by *Nesticella* spiders. *BMC Evolutionary Biology*, 13(1), 183.

CHAPTER II

LIFE IN THE UNDERWORLD: ANCHIALINE CAVE BIOLOGY IN THE ERA OF SPELEOGENOMICS

Jorge Luis Pérez-Moreno, Thomas M. Iliffe, Heather D. Bracken-Grissom

This chapter was published in:

International Journal of Speleology, 45(2): 149-170, 2016.

Abstract

Anchialine caves contain haline bodies of water with underground connections to the ocean and limited exposure to open air. Despite being found on islands and peninsular coastlines around the world, the isolation of anchialine systems has facilitated the evolution of high levels of endemism among their inhabitants. The unique characteristics of anchialine caves and of their predominantly crustacean biodiversity nominate them as particularly interesting study subjects for evolutionary biology. However, there is presently a distinct scarcity of modern molecular methods being employed in the study of anchialine cave ecosystems. The use of current and emerging molecular techniques, e.g. next-generation sequencing (NGS), bestows an exceptional opportunity to answer a variety of long-standing questions pertaining to the realms of speciation, biogeography, population genetics, and evolution, as well as the emergence of extraordinary morphological and physiological adaptations to these unique environments. The integration of NGS methodologies with traditional taxonomic and ecological methods will help elucidate the unique characteristics and evolutionary history of anchialine cave fauna, and thus the significance of their conservation in face of current and future anthropogenic threats. Here we review previous contributions to our understanding of anchialine biodiversity and evolution, and discuss the potential of “speleogenomic” methods for future research in these threatened systems.

Keywords *Biospeleology, Crustacea, Evolution, Genomics, Phylogeography*

Introduction

The term anchialine, from the Greek “*anchialos*” meaning “near the sea”, is generally used in reference to ‘tidally-influenced subterranean estuaries within crevicular and cavernous karst and volcanic terrains that extend inland to the limit of seawater penetration’ (Stock, 1986; Iliffe, 1992; Bishop et al., 2015). Despite tidal influences acting through small conduits and/or the porosity of the surrounding limestone or volcanic rock, anchialine systems have restricted biological connectivity with adjacent water bodies and their associated ecosystems (Iliffe & Kornicker, 2009; Becking et al., 2011; Bishop et al., 2015). Anchialine caves are occasionally interconnected, forming extensive underground networks and giving rise to large and spatially complex habitats (e.g. cenotes in the Yucatan Peninsula, Mexico) (Beddows et al., 2007; Mylroie & Mylroie, 2011). Anchialine caves’ stratified waters often further increase their habitat complexity (Moritsch et al., 2014). This stratification involves a surface layer of meteoric freshwater, separated from underlying marine water by a halocline or mixing zone, where dissolved oxygen levels are low or absent and clouds of hydrogen sulfide occur (Fig. 1) (Sket, 1996; Humphreys, 1999; Iliffe, 2000; Seymour et al., 2007; Becking et al., 2011; Gonzalez et al., 2011). Anchialine systems are widely distributed around the world, mostly isolated from each other, and occurring on karst or volcanic coastlines of islands and peninsulas. Partially explored locations include (but are not limited to) the islands of the Bahamas, Bermuda, Galapagos (Ecuador), Hawaii (U.S.A.), the Ryukyus archipelago (Japan), Canary and Balearic Islands (Spain), the Philippines, Indonesia, Christmas Island and Barrow Island (Australia), and peninsular areas of the Yucatan (Mexico), Belize and Cape Range (Australia) (Iliffe, 1991; Jaume et al., 2001; Humphreys, 2002; Pesce &

Iliffe, 2002; Fosshagen & Iliffe, 2004; Kano & Kase, 2004; Namiotko et al., 2004; Koenemann et al., 2009a; Russ et al., 2010; Becking et al., 2011; Gonzalez et al., 2011). Anchialine habitats are locally known by a variety of names: the most notable being Australia's "sinkholes", Belize's and the Bahamas' "blue holes", and the Yucatan's "cenotes" (from the Maya word *ts'onot*) (Jaume et al., 2001; Iliffe & Kornicker, 2009; Humphreys et al., 2012). These habitats can take a variety of different forms including pools, lava tubes, faults in volcanic rock, karstic limestone cave systems, and connected groundwater (Fig. 2) (Iliffe, 1992; Namiotko et al., 2004; Becking et al., 2011; Mylroie & Mylroie, 2011), yet they all share the same characteristic patterns of stratification and limited biological connectivity with surrounding environments (Kano & Kase, 2004; Hunter et al., 2007; Porter, 2007).

Anchialine caves have a relatively young history in their current state and locations (Mylroie & Mylroie, 2011), originating when formerly dry caves were flooded by rising, post-glacial sea-levels in the early Holocene (11,650-7000 years ago) (Becking et al., 2011; Smith et al., 2011). However, anchialine habitats have existed for millions of years (Iliffe, 2000; Suárez-Morales et al., 2004; Sathiamurthy & Voris, 2006; Becking et al., 2011). Previous studies of cave geology have shown that a great number of extensive and complex caves were formed by the cyclical sea-level changes associated with the Quaternary period (~2.5 million years ago to present) (Mylroie & Mylroie, 2011), while the fossil record indicates that the putative ancestors of modern anchialine fauna were already present in marine systems hundreds of million years ago (e.g. remipedes ~328-306 mya, atyid shrimp ~145-99 mya) (Brooks, 1955; Emerson, 1991; von Rintelen et al., 2012; Moritsch et al., 2014). It is therefore possible that the colonization of anchialine

caves and similar marine crevicular habitats has been occurring since at least the late Jurassic (i.e. the thaumatocypridid ostracod *Pokornyopsis feifeli* Triebel, 1941) (Iglukowska & Boxshall, 2013; Jaume et al., 2013). The particular geochemistry that distinguishes anchialine habitats (low dissolved oxygen, stratified and oligotrophic waters) (Moore, 1999; Seymour et al., 2007; Pohlman, 2011; Neisch et al., 2012), coupled with the distributional patterns and isolation of these cave systems has allowed for a high proportion of endemism among their autochthonous fauna (Iliffe, 1993; Myers et al., 2000; Porter, 2007; Iliffe & Kornicker, 2009). Due to these circumstances, novel and complex chemosynthetically based food webs have arisen, analogous to those found in the deep seas (Sarbu et al., 1996; Engel et al., 2004; Opsahl & Chanton, 2006; Engel, 2007; Porter et al., 2009; Pohlman, 2011).

Recent improvements in scientific cave diving technology and techniques (e.g. mixed-gas rebreathers) have facilitated access and greatly contributed to sampling capabilities in anchialine cave systems (Iliffe & Bowen, 2001; Iliffe & Kornicker, 2009; Iliffe, 2012). Increased access to these systems has resulted in the description of numerous species, genera, families, orders and even a new class (Remipedia) previously unknown to science (Yager, 1981; Iliffe, 2002). However, the scarcity of modern genomic methods being employed in the study of anchialine ecosystems remains to be addressed. Although biospeleological studies that incorporate genetic methodologies have been previously conducted (Adams & Humphreys, 1993; Porter, 2007; Page et al., 2008; Juan et al., 2010), the use of modern sequencing technologies for the study of anchialine caves still lags behind their freshwater and terrestrial counterparts (e.g. Friedrich et al., 2011; Protas et al., 2011; Friedrich, 2013; Gross et al., 2013), with

perhaps the exception of some localized studies of specific taxa (e.g. Meland & Willassen, 2007; Russ et al., 2010; Neiber et al., 2012; von Reumont et al., 2014). In this contribution we examine the current state of knowledge on anchialine cave ecology, biodiversity, and evolution and also discuss the advantages and possibilities that biospeleological investigations at the genomic level, or “speleogenomics”, will provide to the understanding of these fascinating systems – with special emphasis in the areas of biodiversity, phylogeography, and molecular evolution.

Ecology and Biodiversity of Anchialine Caves

Anchialine caves display unique species assemblages with biodiversity often varying not only by location, but also in response to abiotic factors such as tidal flux, salinity, temperature, dissolved oxygen, and water stratification (e.g. haloclines) (Iliffe, 2002; Gonzalez et al., 2011). Cave food webs have been regarded as nutrient poor and dependent on external inputs of nutrients such as decaying organic matter (Dickson, 1975; Sket, 1996; Neisch et al., 2012), but recent discoveries have attributed considerable importance to the chemosynthetic activity of bacterial communities (Sarbu et al., 1996; Pohlman et al., 1997; Engel et al., 2004; Engel, 2007; Seymour et al., 2007; Gonzalez et al., 2011; Humphreys et al., 2012; Pakes et al., 2014; Pakes & Mejía-Ortíz, 2014), particularly with increasing distances from cave openings (Neisch et al., 2012). In fact, productivity of cave chemoautotrophic communities appears to correlate with diversity of heterotrophic microbes and of macro-invertebrates in higher trophic levels, which suggests that microbial diversity plays a role in mediating cave biodiversity (Engel, 2007; Porter et al., 2009). Chemosynthetic ectosymbioses between bacteria and several

invertebrate phyla have been documented in similar ecosystems (Dubilier et al., 2008; Goffredi, 2010), including freshwater caves (Dattagupta et al., 2009; Bauermeister et al., 2012). Recent studies suggest that analogous interactions occur in anchialine systems, with both ecto- and endosymbioses of chemoautotrophic bacteria having been found in two crustacean taxa (the remipede *Xibalbanus tulumensis* and the atyid shrimp *Typhlatya pearsei*) from anchialine caves (Pakes et al., 2014; Pakes & Mejía-Ortíz, 2014). Other microbiota also present in anchialine caves include microscopic eukaryotes such as fungi, protozoa, and rotifers, but documentation on their biodiversity and ecological roles in anchialine caves is limited (Engel, 2007).

Assemblages of anchialine cave fauna display unique variations and stratified ecological niches, due to thermoclines and haloclines, among and within caves. An interesting phenomenon observed in these systems is the assemblage cave “types” (e.g. Remipedia or Procaridid communities) – where similar crustacean communities of only a few different genera are found inhabiting different caves, often geographically distant from each other, and displaying predictable generic compositions (Poore & Humphreys, 1992; Jaume et al., 2001; Humphreys & Danielopol, 2006; Neiber et al., 2011). Remipede-type caves typically contain remipedes and other crustacean stygobionts (aquatic and cave-dwelling) such as cirolanid isopods, hadziid amphipods, calanoid copepods, ostracods, thermosbaenaceans, and atyid shrimps; while Procaridid-type communities are characterized by the presence of shrimp from the genus *Procaris* Chace and Manning, 1972 along with a number of species of alpheid, atyid, and barbouriid shrimps (Chace & Manning, 1972; Humphreys & Danielopol, 2006; Neiber et al., 2011). The exact reasons underlying these phenomena of community “types” and disjunct

distributions continue to be subject to investigation. The dominant hypothesis suggests that this community-type phenomenon is due to ancient geological patterns when many of these species and their ancestors originated (in the Tethys Sea during the Mesozoic), as these cave community-types tends to be associated with particular geographical features (e.g. Procaridid-type communities are more commonly located on sea-mount islands) (Humphreys, 1999, 2002; Humphreys & Danielopol, 2006; Neiber et al., 2011). The underlying mechanisms and processes that gave rise to cave biodiversity and its ecology constitute one of the major research themes for modern biospeleology (Peck & Finston, 1993; Sket, 1999; Juan et al., 2010).

Stygobitic fish, particularly eel-like fish (orders Ophidiiformes, Synbranchiformes) and eleotrids (order Perciformes), can be encountered in anchialine caves (Williams et al., 1989; Humphreys, 2001; Medina-Gonzalez et al., 2001; Wilkens, 2001; Larson et al., 2013). However, these habitats are clearly dominated by invertebrates both in terms of diversity and biomass (Iliffe, 2002). Anchialine invertebrates encompass a diverse range of taxa, e.g. annelids, arachnids, chaetognaths, echinoderms, gastropods, poriferans, turbellarians, but most importantly crustaceans (Culver & Sket, 2000; Engel, 2007; Mejía-Ortíz et al., 2007; Iliffe & Kornicker, 2009; Solís-Marín & Laguarda-Figueras, 2010; Bribiesca-Contreras et al., 2013). The reason for the high diversity of crustaceans, the endemism of higher taxa to anchialine systems, and their preponderance over other higher taxa is unknown (Stoch, 1995; Sket, 1999). The diversity, abundance, and widespread distributions of crustacean taxa in anchialine caves designate them as the ideal subjects for biospeleological, biogeographical, and evolutionary studies in these

systems. The sub-phylum Crustacea is most commonly represented in anchialine cave environments by organisms from the following taxa:

Order Decapoda (Class Malacostraca, Superorder Eucarida)

Stygobitic decapods (Fig. 3d) are broadly distributed throughout tropical and subtropical anchialine caves (Bruce & Davie, 2006; Hunter et al., 2007; Iliffe & Kornicker, 2009). Freshwater crayfish, and both brachyuran and anomuran crabs (e.g. *Munidopsis polymorpha* Koelbel, 1892) have been found inhabiting cave environments (Iliffe, 1993; Ng et al., 1996; Mejía-Ortíz et al., 2003; Cabezas et al., 2012; Álvarez et al., 2014), but the most common stygobitic decapods are the caridean shrimp (e.g. families Agostocarididae, Alpheidae, Atyidae, Barbouriidae, Hippolytidae, Palaemonidae) (Chace & Manning, 1972; Jaume & Brehier, 2005; Hunter et al., 2007; Álvarez et al., 2012), procarididean (e.g. family Procarididae) (Chace & Manning, 1972; Felgenhauer et al., 1988; Bruce & Davie, 2006; Bracken et al., 2010), stenopodidean (e.g. family Macromaxillocarididae) (Álvarez et al., 2006), and gebiidean (e.g. family Laomediidae) (Iliffe & Kornicker, 2009) representatives living in anchialine systems around the world. Decapods are also among the most studied anchialine taxa, perhaps due to their charismatic nature and larger sizes (making them easier to be located and captured). However, the life-history, biogeography, and ecology of their anchialine cave inhabiting representatives for the most part remain poorly understood. Genetic studies of anchialine decapods have resulted in valuable insights on the phylogenetic position and biogeography of some species (for example Santos et al., 2006; Hunter et al., 2007; Page et al., 2008; Bracken et al., 2010; Botello et al., 2013), but investigations at the genomic

or transcriptomic level remain scarce (Genomic Resources Development Consortium et al., 2014; Justice et al., 2015).

Order Amphipoda (Class Malacostraca, Superorder Peracarida)

Stygobitic amphipods (Fig. 3a) are small “shrimp-like” crustaceans that can be found in a variety of cave environments, including freshwater and anchialine caves, and are distributed across the world with a considerable number of species described from the Atlantic region (Southern Europe, the Mediterranean, North America, and the Caribbean) (Culver & Pipan, 2009; Iliffe & Kornicker, 2009). They are mostly represented in anchialine systems by a number of families from the suborder Senticaudata (e.g. Bogidiellidae, Hadziidae, Melitidae, Metacrangonyctidae Niphargidae, Salentinellidae) (Jaume & Christenson, 2001; Iliffe & Kornicker, 2009; Gràcia & Jaume, 2011). Recent molecular investigations have identified novel ectosymbioses between cave amphipods (*Niphargus* spp.) and sulphur-oxidizing chemosynthetic bacteria (Dattagupta et al., 2009; Flot et al., 2010; Bauermeister et al., 2012). Although such findings concerned freshwater species, the findings raise the possibility of similar symbioses occurring in these environments.

Order Isopoda (Class Malacostraca, Superorder Peracarida)

Several families of isopods (e.g. Anthuridae, Asellidae, Atlantasellidae, Cirolanidae, Microcerberidae, Stenasellidae, Sphaeromatidae) (Fig. 2e) are also found inhabiting cave systems, and their distributions tend to be relatively widespread. Isopods have been described from anchialine caves in Africa (Canary Islands), Asia (India,

Indonesia, Japan, Malaysia), Europe (Mediterranean), North America (The Bahamas, Bermuda, Mexico and the Caribbean), Central and South America (Galapagos Islands), and Oceania (Australia and Polynesian Islands) (Bruce & Humphreys, 1993; Botosaneanu & Iliffe, 2006; Iliffe & Botosaneanu, 2006; Iliffe & Kornicker, 2009). Cirolanids and sphaeromatid isopods are thought to have a marine origin, and are prevalent in anchialine systems, in contrast with other stygobitic families (e.g. Asellidae, Stenasellidae, Microcerberidae) that are likely to be product of colonizations from epigean freshwater habitats (Culver & Pipan, 2009). A limited number of anchialine isopods have been included in genetic studies (for example, molecular phylogeny of Cirolanidae in Baratti et al., 2010), but none of these have been conducted in the context of anchialine systems, nor at the genomic/transcriptomic level.

Orders Bochusacea and Thermosbaenacea (Class Malacostraca, Superorder Peracarida)

Bochusaceans are small (1.2 – 1.6 mm) swimming peracarids that display several morphological regressive adaptations to cave life (lack of pigmentation and visual organs) (Guțu & Iliffe, 1998; Iliffe & Kornicker, 2009). Only two species of Bochusacea are known, inhabiting anchialine and submarine caves from the Bahamas and Cayman Islands (Guțu & Iliffe, 1998). Two other species are also known to be found in deep-sea environments (Guțu & Iliffe, 1998; Jaume et al., 2006). There is presently only a single Bochusacean DNA sequence available online (small-subunit ribosomal RNA gene for *Thetispelecaris remex*), which resulted from a study that investigated peracarid monophyly (Spears et al., 2005). Thermosbaenaceans (Fig. 3j) are small (< 5 mm) and

enigmatic stygobitic swimming crustaceans. They tend to live in the water column in proximity to the halocline, where they feed off organic matter and microbial communities that inhabit these density interphases (Gràcia & Jaume, 2011). They are globally distributed with some species found in Australia, Cambodia, the Mediterranean, and the Caribbean (Poore & Humphreys, 1992; Iliffe & Kornicker, 2009). Although they are believed to have originated from marine ancestors, no extant epigean marine species have been found (Sket, 1996). Interestingly, thermosbaenaceans brood their young in a dorsal pouch, as opposed to a ventral marsupium as in the case of other extant peracarids (Olesen et al., 2015), and their brain's olfactory lobe seems to be less developed than in other blind cave-dwelling crustaceans (Stegner et al., 2015). Similarly to bochusaceans, genetic resources for the order Thermosbaenacea are severely lacking. Of the four thermosbaenacean DNA sequences deposited in Genbank (National Center for Biotechnology Information), only one is from an anchialine representative (*Tethysbaena scabra*). Furthermore, this sequence for the 18S rRNA gene from *T. scabra* was simply used as an out-group for an asellote isopod phylogeny (Wägele et al., 2003). Despite recent innovations and examinations of thermosbaenacean morphology (Olesen et al., 2015; Stegner et al., 2015), genetic and genomic/transcriptomic studies yet remain to be conducted.

Orders Mictacea, Mysida, and Stygiomysida (Class Malacostraca, Superorder Peracarida)

Mictaceans (Fig. 3f) are relatively small (~3 mm) swimming peracarid crustaceans with only a single species in the order, *Mictocaris halope* (Bowman & Iliffe, 1985). This single representative of the order inhabits anchialine caves of Bermuda, primarily in the deeper and harder to access areas (Bowman & Iliffe, 1985). Stygobitic mysids (Fig. 3g) have a wide distribution with species endemic to anchialine caves in Africa, the Caribbean, Mediterranean, and India (Pesce & Iliffe, 2002; Iliffe & Kornicker, 2009). The Mysidacea has been split into two new orders, Mysida and Lophogastrida (Martin & Davis, 2001; Spears et al., 2005; Porter et al., 2007), with stygobitic mysids belonging to the former. However, more recent molecular analyses have concluded that the order “Mysidacea” actually consists of three monophyletic groups and strongly suggest classifying some stygobitic mysids from the Caribbean and Mediterranean in the proposed order of “Stygiomysida” (Meland & Willassen, 2007; Porter et al., 2007).

Orders Cumacea and Tanaidacea (Class Malacostraca, Superorder Peracarida)

Cumaceans (Fig. 3c) are peracarid crustaceans that can be found globally distributed with many species inhabiting areas as varied as the Australian Indo-Pacific to the Western Atlantic Ocean (Tafe & Greenwood, 1996a, 1996b; Petrescu, 2003; Petrescu & Iliffe, 2009). In the Western Atlantic region, cumaceans can be encountered both in oceanic basins (Petrescu et al., 1993; Petrescu, 1995) as well as in anchialine cave systems (Petrescu & Iliffe, 1992, 2009). The physiology, life history, and ecology of most cumacean species are poorly understood (Gnewuch & Croker, 1973; Corey, 1981;

Duncan, 1984; Corbera et al., 2008;), especially when concerning that of stygobitic species. Tanaidaceans are another group of anchialine crustaceans found across the globe, with specimens having been recovered from caves in the Western Atlantic (the Bahamas Islands) and the tropical Indo-Pacific (Fiji Islands and Palau) (Guțu & Iliffe, 1989a; Guțu & Iliffe, 1989b; Guțu & Iliffe, 2011). They are small dorso-ventrally flattened crustaceans with generally highly chitinized bodies, although some cave species with softer bodies have been found (Guțu & Iliffe, 1989a; Guțu & Iliffe, 1989b). Both cumaceans and tanaids are underrepresented in genetic studies in general (Xin et al., 2015), and especially in anchialine systems where these investigations are yet to be undertaken.

Suborder Nebaliacea (Class Malacostraca, Order Leptostraca) and Subclass Tantulocarida (Superclass Multicrustacea)

Nebaliaceans are small shrimp-like benthic crustaceans typically from 5 to 15 mm long. Although they are mostly marine, an anchialine cave species of nebaliacean, known from the Turks and Caicos Islands, shares with its marine counterparts the ability to tolerate low-oxygen environments (Bowman et al., 1985; Walker-Smith & Poore, 2001). There are no genetic resources available for anchialine Nebaliacea. Tantulocarids are small crustacean ectoparasites usually associated with other crustacean hosts (Boxshall & Huys, 1989; Huys, 1990). Stygobitic tantulocarids have been described parasitizing harpacticoid copepods in anchialine caves of the Canary Islands (Boxshall & Huys, 1989; Iliffe & Kornicker, 2009). Recent molecular phylogenetic investigations have suggested a close relation between tantulocarids and the subclass Thecostraca, and that Tantulocarida

might in fact belong within this subclass as a sister group to Cirripedia (barnacles) (Petrunina et al., 2014). However, the precise phylogenetic position of Tantulocarida still awaits further investigation (Petrunina et al., 2014).

Orders Calanoida, Cyclopoida, Harpacticoida, Misophrioida, Platycopioidea (Superclass Multicrustacea, Subclass Copepoda)

Copepods (Fig. 3b) are amongst the most abundant and widely distributed taxa of aquatic animals, and exist in a wide range of environments across the globe (Boxshall & Defaye, 2008). Not surprisingly, several orders from the subclass Copepoda can be found inhabiting most anchialine caves (Rouch, 1994; Gràcia & Jaume, 2011). They are typically encountered in the water column where they filter feed, except for a number of benthic bio-film grazers (e.g. cyclopoids & harpacticoids), and predatorial (e.g. cyclopoids & epacteriscids) species (Rouch, 1994; Fosshagen et al., 2001; Suárez-Morales et al., 2004; Suárez-Morales et al., 2006; Suárez-Morales & Iliffe, 2007; Suárez-Morales & Iliffe, 2005a, 2005b; Iliffe & Kornicker, 2009). Stygobitic copepods often present troglomorphies such as the reduction or absence of eyes and enlargement of eggs (Rouch, 1968). Genetic studies of copepods from anchialine caves are rare, with only a few studies having sequenced them for molecular phylogenetic purposes (Huys et al., 2006; Figueroa, 2011).

Orders Halocyprida, Myodocopida, Platycopida, Podocopida (Class Ostracoda)

Ostracods (Fig. 3h) are a very diverse and abundant group, with approximately 980 species described from caves and other subterranean habitats (Martens, 2004; Iliffe

& Kornicker, 2009; Hobbs, 2012). These small (~1 mm) bivalved crustaceans are active swimmers and as such are commonly found in the water column, which may be a contributing factor to their dispersal capabilities (Humphreys & Danielopol, 2006; Kornicker et al., 2009). Ostracods are distributed across anchialine habitats in both hemispheres, with some genera (e.g. *Humphreysella*) having representatives on opposite sides of the planet (Humphreys & Danielopol, 2006; Kornicker et al., 2008, 2009; Iglukowska & Boxshall, 2013). Stygobitic ostracods are easily distinguishable from epigean representatives by the morphological differences associated with their adaptations to cave life (i.e. smaller size, lack of eyes and pigmentation) (Danielopol, 1981). Even though anchialine ostracods have not received much attention from molecular biologists, genetic and genomic/transcriptomic studies of ostracods in other environments have been conducted with great success (Oakley & Cunningham, 2002; Oakley, 2005; Oakley et al., 2012). These studies provide a great basis on which to build upon future investigations of anchialine cave ostracods, which are likely to yield interesting evolutionary insights.

Order Nectiopoda (Class Remipedia)

Remipedes (Fig. 3i) are an unusual class of blind crustaceans with extensive body segmentation and lateral biramous swimming appendages that superficially resemble polychaete worms. Characteristics such as their cephalic anatomy warranted their classification in the subphylum Crustacea (Yager, 1981), which was later confirmed by molecular studies (von Reumont et al., 2012). Remipedes follow similar distribution patterns to those of halocyprid ostracods (Kornicker et al., 2007), and can be found

exclusively in anchialine caves throughout the globe in a seemingly disjunct range of locations such as the Western Atlantic and Caribbean (Bahamas, Belize, Cuba, Dominican Republic, Yucatan), Africa (Canary Islands), and Western Australia (Sket, 1996; Yager & Humphreys, 1996; Koenemann et al., 2003, 2004, 2007a, 2007c, 2009a; Lorentzen et al., 2007; Daenekas et al., 2009; Neiber et al., 2011, 2012; Hoenemann et al., 2013; Koenemann & Iliffe, 2013). Although at first sight remipedes may appear morphologically primitive (Yager, 1994), they possess an advanced nervous system (Stemme et al., 2013), highly specialized feeding mouthparts for capturing prey (von Reumont et al., 2014), and they are the top predatory crustaceans in the low-oxygen anchialine systems they inhabit (Koenemann et al., 2007c; Iliffe & Kornicker, 2009). Remipede larvae are so far only known from a single species in one cave (Koenemann et al., 2007b; 2009b; Olesen et al., 2014). Recent investigations of the remipede *Xibalbanus tulumensis* (Yager, 1987) have found that in addition to feeding from particulate matter in the water column, this species harbors chemosynthetic bacteria in ectosymbiosis that allow for the uptake of inorganic carbon as a supplement to their diet (Pakes & Mejía-Ortíz, 2014). Furthermore, *X. tulumensis* has been shown to employ venom to capture and digest atyid shrimp, which makes it the first venomous crustacean ever documented (von Reumont et al., 2014).

Current Advances and Future Prospects

Despite difficulties and dangers of sampling in anchialine caves (Iliffe & Bowen, 2001; Iliffe, 2002, 2012), previous studies have made monumental contributions to the field and an extraordinary amount of novel diversity from these habitats has been

described to present day. Although traditional sanger DNA sequencing technologies (Glossary Box 1) have provided valuable insights to biospeleology (including but not limited to species identification, phylogenetics, and estimates of genetic diversity) (Juan et al., 2010), high-resolution molecular data from cave systems have the potential to greatly expand the depth and breadth of knowledge to be gained from these types of studies. “Next-generation” DNA sequencing technologies (NGS), which allow for the sequencing of thousands of loci and/or hundreds of samples at a time, have scarcely been used by biospeleologists (Juan et al., 2010; Friedrich et al., 2011; Friedrich, 2013; Tierney et al., 2015). Previous biospeleological studies that incorporate genetic data to their investigation efforts have mainly focused on a single locus (for examples see: phylogeography – Caccone & Sbordoni, 2001; Buhay & Crandall, 2005; population genetics – Russ et al., 2010; phylogenetics – Neiber et al., 2011, 2012) or a limited number of loci at a time (for examples see: phylogeography – Villacorta et al., 2008; Trontelj et al., 2009; Zakšek et al., 2009; phylogenetics – Leys et al., 2003; Hunter et al., 2007; Lefébure et al., 2007; Zakšek et al., 2007; Page et al., 2008; von Rintelen et al., 2012; Hoenemann et al., 2013), with only a small portion of those studies employing four or more loci in their analyses (for examples see: phylogenetics – Bracken et al., 2010; Botello et al., 2013; population genetics – Adams & Humphreys, 1993). Employing a limited number of loci is suitable for the specific purposes that have been addressed so far, nevertheless the continuous development and improvement of molecular techniques offers an enormous potential for answering long-standing questions in biospeleology (Juan et al., 2010). These technologies open the way for analyses of a much higher resolution at an accelerated pace, and facilitate work on whole genomes and

transcriptomes (Shendure & Ji, 2008; Metzker, 2010; Lemmon et al., 2012; Friedrich, 2013). Additionally, NGS has the ability to provide researchers with vast amounts of data in a cost-effective manner (Metzker, 2010). NGS has also permitted the development of techniques that target many loci and/or many samples at once (Lemmon et al., 2012), such as “Targeted Sequencing” (Glossary Box 1) (Meyer et al., 2007; Mamanova et al., 2010; Bybee et al., 2011; Ekblom & Galindo, 2011; Hedges et al., 2011; Cronn et al., 2012; Grover et al., 2012; Hancock-Hanser et al., 2013; Stull et al., 2013), “Anchored Hybrid Enrichment” (Glossary Box 1) (Lemmon et al., 2012), and other high-throughput methods (Binladen et al., 2007; Miller et al., 2007; Lemmon & Lemmon, 2012; Rohland & Reich, 2012; Peñalba et al., 2014). These methods, some of which have already been employed successfully for pancrustacean phylogenetics (Bybee et al., 2011), are easily adaptable for other purposes where massively parallel sequencing would be advantageous (e.g. multi-locus phylogenetics, metagenomics, DNA barcoding, biodiversity assessments, etc.) (Glossary Box 1). In combination with non-destructive tissue sampling techniques, the high-throughput nature of NGS paves the way for studies with large sample sizes with a minimal impact on natural populations. Minimizing the impact of sampling is of particular importance when working with rare and endemic cave species, especially those with small population sizes such as many anchialine cave dwellers. These methodologies can be employed for biological research in caves and similar environments to answer questions in a diverse array of areas such as biogeography/phylogeography (Porter, 2007; Juan et al., 2010; Lemmon & Lemmon, 2012; McCormack et al., 2013), ecology (Mock & Kirkham, 2012), phylogenetics/phylogenomics (Bybee et al., 2011; Lemmon & Lemmon, 2012;

McCormack et al., 2013), cryptic speciation and evolution (Juan et al., 2010). The potential of next-generation sequencing has so far been demonstrated by the relatively few biospeleological studies that have successfully incorporated these modern techniques (e.g. Humphreys et al., 2012; Gross et al., 2013; O’Quin et al., 2013; von Reumont et al., 2012, 2014).

Cave biodiversity in the Molecular Era

Current molecular tools, such as DNA barcoding, allow us to identify species by using a DNA sequence in a specific genomic region as an identifier or “barcode” (Savolainen et al., 2005; Shokralla et al., 2014). DNA barcoding can be useful to discern species complexes that would otherwise go unnoticed due to morphological similarities or dissimilarities within a single species at different life-stages (Puillandre et al., 2011; Bracken-Grissom et al., 2012; Neiber et al., 2012). This is of special importance in anchialine caves and other subterranean systems where the possibility that troglomorphy and convergent evolution of morphological traits obscure phylogenetic relationships is significant (Wiens et al., 2003; Wilcox et al., 2004; Buhay & Crandall, 2005; Porter, 2007; Trontelj et al., 2009). For example, Zakšek et al. (2009) investigated the seemingly widespread distribution of a common species of freshwater cave shrimp from the Balkan Peninsula (*Troglocaris anophthalmus*) and concluded that they should be considered distinct evolutionary significant units for conservation purposes. The study thus provides an example of how molecular tools can contribute to the delimitation of species with extensive convergent morphologies, which in turn could have important conservation implications. Molecular tools, such as DNA sequencing, will undoubtedly continue to be

of importance for resolving cryptic species complexes that are pervasive in cave environments (Lefébure et al., 2007; Trontelj et al., 2009; Neiber et al., 2012). Similarly, morphological differences between life-stages within a species are commonplace among crustaceans, and in many instances pose important challenges for organism identification and taxonomic classification. This is especially common in poorly studied or rare species, where adult-larval linkages have not been determined experimentally due to logistical difficulties in obtaining samples or difficulty of larval rearing. DNA barcoding has proven useful to link morphologically distinct adults and larvae of the same species. For example, Bracken-Grissom et al. (2012) employed DNA barcoding regions and molecular systematics to show that the mid-water species *Cerataspis monstrosa* was in fact the larval stage of the deep-sea shrimp *Plesiopenaeus armatus*. The high-throughput capabilities of NGS can substantially benefit DNA barcoding efforts by targeting specific amplicons over hundreds of samples at a time (Glossary Box 2) (Floyd et al., 2002; Wu et al., 2009; Puillandre et al., 2011; Shokralla et al., 2014), making the sequencing and processing of numerous samples more efficient and cost-effective than with traditional Sanger DNA sequencing. These high-throughput capabilities can be especially useful for applications such as species identification, creation of species inventories (and large scale projects, such as the Barcode of Life initiative), detection of cryptic species complexes, and species delimitation (Savolainen et al., 2005; Bickford et al., 2007; Hajibabaei et al., 2007; Ratnasingham & Hebert, 2007; Trontelj et al., 2009; Niemiller et al., 2013; Shokralla et al., 2014), all of which would be of benefit to research in anchialine caves (i.e. Bribiesca-Contreras et al., 2013).

Phylogeography of Anchialine Cave Ecosystems

The vast amounts of genomic data that are possible to obtain with current technologies can be used to investigate evolutionary rates, diversification, and speciation among anchialine cave fauna, as well as enabling the investigation of population structure and gene-flow patterns at an unprecedented resolution (Leys et al., 2003; Porter, 2007; McCormack et al., 2013). Furthermore, these kinds of molecular data can be used to answer questions regarding the intriguing distribution patterns of cave fauna, such as the determination of species origins aligning with the climatic-relic or adaptive-shift hypotheses (Leys et al., 2003). In biogeographical terms, anchialine fauna have provided a very interesting source of debate, where several models have been proposed to explain their origins (Suárez-Morales & Iliffe, 2005a; Porter, 2007; Culver et al., 2009; Iliffe & Kornicker, 2009). The vicariance hypothesis states that the distribution of present-day anchialine fauna can be explained by plate tectonics, whereas the dispersal models suggest that stygobitic species dispersed to their present location where non-cave sister species invaded and adapted to cave environments (Jaume et al., 2001; Iliffe & Kornicker, 2009). The actual mechanisms that gave rise to contemporary anchialine fauna are likely to be a more complex combination of the previously mentioned models (Culver et al., 2009). Molecular studies provide the opportunity to test these hypotheses (Page et al., 2008; Juan et al., 2010). A number of comparative phylogeography studies have already been undertaken to explain the evolutionary origins and distributional patterns of cave fauna (Caccone & Sbordoni, 2001; Espinasa & Borowsky, 2001; Hunter et al., 2007; Ribera et al., 2010; von Rintelen et al., 2012). Although in the case of most

taxa, the evidence of their origins remains inconclusive at best (Phillips et al., 2013), the incorporation of modern molecular techniques with datasets at the genomic scale will undoubtedly shape the future of this research area (e.g. with use of phylogenomic approaches (Leaché et al., 2014)).

One method that can be applied to fine-scale questions of phylogeography (i.e. population to species) is Restriction-site associated DNA sequencing (i.e. RAD-Seq) (Glossary Box 2). This is a methodology that allows for the sequencing, identification, and use of thousands of genetic markers, such as Single Nucleotide Polymorphisms (SNPs), distributed across hundreds of loci (Ekblom & Galindo, 2011; McCormack et al., 2012, 2013). Restriction-site associated DNA sequencing reduces the complexity of the genome to be investigated with the use of restriction enzymes, which allows for genome-wide analyses to be performed without the computational and financial requirements of working with whole genomes (Davey & Blaxter, 2010; Davey et al., 2011; Toonen et al., 2013). RAD-Seq provides high resolution data that enable the identification of potentially thousands of these genetic markers across individuals and populations that can be employed for further analyses (Davey & Blaxter, 2010; Peterson et al., 2012). For example, Coghill et al. (2014) used RAD-Seq to trace the colonization of caves by the blind Mexican cavefish *Astyanax mexicanus*. This methodology enabled them to find over 2,000 SNPs across the examined populations and provided evidence for at least four independent colonization events from surface populations to the caves, which suggests parallel evolution of the cave phenotypes observed in these stygobitic fish.

Cave-inhabiting organisms can be used as a proxy for investigating the connectivity of intricate cave systems, by looking at patterns of gene flow and population

connectivity. Many submerged cave systems form underground web-like tunnels that extend for several hundreds of kilometers (e.g. the Yucatan cave systems) (Iliffe, 2000; Beddows et al., 2007; Mylroie & Mylroie, 2011; Moritsch et al., 2014). The complexity of these cave systems makes them extremely challenging to be explored using traditional cave-diving methods, mainly due to technological and physiological constraints. Several studies have used stygobiont genetics to assess present or historical hydrological connectivity of cave systems (e.g. Culver et al., 1995; Verovnik et al., 2004; Krejca, 2005). Culver et al (1995), while examining cave-adapted populations of *Gammarus minus* in West Virginia (USA), found congruent patterns between genetic differentiation and hydrology even when accounting for the possible selective pressures of different habitats. Krejca (2005) compared mitochondrial DNA phylogenies of two lineages of aquatic isopods (Cirolanidae and Asellidae) to examine the evolution of aquifers in Texas (USA) and northern Mexico. Despite finding differences between the two species examined, which could be explained by their individual ecologies and life-histories, Krejca (2005) found congruency between the crustacean phylogenies and the hydrogeological history of the examined systems. The molecular examination of these two cave-dwelling isopod species allowed her to test a priori biogeographical hypotheses and investigate the evolution of the aquifers studied (Krejca, 2005). Further, Verovnik et al. (2004) also used molecular data (mtDNA) (Glossary Box 2) of a crustacean species (*Asellus aquaticus*), in combination with paleogeographical information, to reveal possible scenarios of hydrological history of the Dinaric karst in the Balkan Peninsula. A study in the Pilbara region of Western Australia uncovered similar patterns amongst subterranean amphipods (Finston et al., 2007), where the mitochondrial haplotypes

(Glossary Box 2) found were congruent with the hydrology of the tributaries examined as previously hypothesized (Humphreys, 2001b). Anchialine cave system hydrology can be similar to that of freshwater karstic cave systems, with the added complexity of underground connections to marine waters. Santos (2006) investigated the population genetics and connectivity patterns of the iconic Hawaiian anchialine shrimp *Halocaridina rubra*. Amongst his findings, he determined that there appears to be strong population subdivisions and a clear genetic structure particularly when surface distances between anchialine pools exceeded 30 km. Santos' (2006) results also suggest that dispersal through subterranean conduits between anchialine pools is of more importance for this species than oceanic dispersal. These results contrast with Kano and Kase's (2004) findings of extensive oceanic dispersal by anchialine gastropods, further illustrating the importance of meticulous consideration of study species for cave connectivity purposes – where the chosen species' dispersal abilities should correspond to the geographical scales under investigation. Coupled with NGS technologies, these could offer a compelling alternative for the investigation of cave connectivity, by using population genomics as a proxy via methodologies such as RAD-Seq. Reduced-representation genome sequencing methodologies offer an unprecedented resolution (even compared to microsatellites) to genotype populations of cave organisms by sampling thousands of genomic regions at a time (Bradbury et al., 2015). The population structure and gene-flow patterns of those stygobiont populations could then be employed for a fine-scale evaluation of the connectivity of the anchialine caves under investigation, and complement traditional exploration efforts (e.g. scientific cave diving, Iliffe & Bowen, 2001; dye-tracing, Beddows & Hendrickson, 2008) of these spatially complex habitats.

Evolution of Troglomorphy

RNA sequencing (RNA-Seq) (Glossary Box 2) can provide invaluable resources for evolutionary studies of cave biota. The term RNA-Seq refers to the high-throughput sequencing of RNA from a specific tissue or organism at a discrete point in time (Wang et al., 2009; De Wit et al., 2012). This is achieved by reverse transcribing extracted RNA to cDNA, followed by high-throughput sequencing by an NGS platform (e.g., 454 pyrosequencing, Illumina, PacBio), and subsequent *de novo* assembly of the sequenced reads or the alignment of these reads to reference genomes (Wang et al., 2009; Deyholos, 2010; Martin & Wang, 2011; Zhang et al., 2011; De Wit et al., 2012). The resulting transcriptome assembly can then be characterized to identify the transcripts that are being expressed in that tissue, organism, and/or life-stage (Ekblom & Galindo, 2011; De Wit et al., 2012). Albeit being purely descriptive, a characterized transcriptome provides a base on which to build further analyses. The characterized transcriptome assembly can be used as a reference sequence and both the original and additional sequenced reads (for other treatments, for example) can be mapped back to the assembly to obtain quantitative data of gene expression and genetic variation (Ellegren, 2008; Deyholos, 2010; Ekblom & Galindo, 2011). These data can be further utilized for a variety of applications such as the development of molecular markers and even the identification of events associated with speciation processes (i.e. alternative splicing) (Harr & Turner, 2010; Ekblom & Galindo, 2011). The small size of RNA sequence datasets, in comparison with whole-genome data, can also be valuable for the identification of new molecular markers and of novel proteins from non-model organisms in a computationally efficient manner. Additionally, transcriptomic data can be used for studies in a wide range of areas such as evolution

(Harr & Turner, 2010; Friedrich et al., 2011; Rehm et al., 2011; Wong et al., 2015), development (Zeng et al., 2011; Ichihashi et al., 2014), physiology (Dassanayake et al., 2009; Harms et al., 2013; Groh et al., 2014), adaptations to changing environments (Deyholos, 2010; Friedrich, 2013; Harms et al., 2013), and responses to physicochemical challenges (e.g. biomonitoring & ecotoxicogenomics) (Watanabe et al., 2008; Suárez-Ulloa et al., 2013a, 2013b).

RNA-Seq (Wang et al., 2009) can also be used to address more basic questions of cave evolution, by investigating the “speleotranscriptome” – the transcriptomic profile of stygobitic fauna's physiological and morphological adaptations (Gross et al., 2013). In addition, such investigations can set the stage for addressing broader questions regarding natural selection and the evolution of phenotypic diversity, novel molecular functions, and complex organismal features (Christin et al., 2010). Animals inhabiting cave environments usually undergo various distinct physiological, morphological, and behavioral changes, which together are commonly referred to as “troglomorphy” (Desutter-Grandcolas, 1997; Porter & Crandall, 2003; Mejía-Ortíz et al., 2006). Troglomorphic modifications can be classified in either progressive (constructive) or regressive (reductive) adaptations (Porter & Crandall, 2003; Mejía-Ortíz & Hartnoll, 2006; Mejía-Ortíz et al., 2006). In anchialine cave environments, stygobitic (aquatic and cave-limited) fauna typically present a combination of both types of troglomorphy. Examples of progressive adaptations may include cases such as those of enlarged sensory and ambulatory appendages, increased numbers of chemoreceptor setae, or enhancement of spatial orientation capabilities (Turk et al., 1996; Li & Cooper, 2001, 2002; Mejía-Ortíz & Hartnoll, 2006;). Regressive modifications involve the decrease or loss of

features present in their epigean (surface) counterparts, e.g. reduced pigmentation, reduction or loss of visual functions, or decreased metabolism (Sket, 1985; Wilkens, 1986; Mejía-Ortíz & López-Mejía, 2005; Mejía-Ortíz et al., 2006; Bishop & Iliffe, 2012). Troglomorphy is a perfect example of convergent morphological evolution where analogous traits have evolved in different lineages to adapt to similar environments (Caccone & Sbordoni, 2001; Wilcox et al., 2004; Protas et al., 2007; Bishop & Iliffe, 2012; Mejía-Ortíz et al., 2013). Species from a variety of crustacean taxa have been documented to have convergent characters (e.g. pigmentation, Beatty, 1949; Anders, 1956; body-size, Hobbs et al., 1977) by seemingly analogous mechanisms as adaptations to their subterranean life. This phenomenon poses the question on whether the underlying mechanisms of troglomorphy in cave crustaceans are also convergent at the molecular level. Although morphological and physiological convergence is well documented (Arendt & Reznick, 2008), particularly in the case of adaptations to extreme environments (including caves) (Wiens et al., 2003; Wilcox et al., 2004; Protas et al., 2006, 2007; Dassanayake et al., 2009), cases of convergent molecular evolution remain elusive (Tierney et al., 2015). Nevertheless, it has been suggested that this seemingly rare occurrence may be simply a product of the low-resolution genetic sampling that has been prevalent in the last few decades (Castoe et al., 2010). Recent investigations at the genomic and transcriptomic levels have indeed revealed evidence of convergent molecular evolution associated to phenotypic convergence (see Foote et al., 2015 for genomic convergence in marine mammals, Pankey et al., 2014 for transcriptomic convergence in bioluminescent squid, and Tierney et al., 2015 for transcriptomic convergence in subterranean beetles). A combination of transcriptomic and genomic

approaches can help elucidate the strategies and mechanisms of adaptation to extreme environments (Benvenuto et al., 2015), as well as evaluate the prevalence of molecular convergence and the patterns it might follow in anchialine caves, where strong selective pressures could prompt for homologous mechanisms of genetic adaptation across different taxa.

Molecular studies of the evolution of special adaptations to extreme environments have been undertaken in a wide array of taxa; although to date most of these have focused on prokaryotes (Lauro & Bartlett, 2008; Sahl et al., 2011; Bonilla-Rosso et al., 2012; Lesniewski et al., 2012; Baker et al., 2012, 2013; Orsi et al., 2013; Iwanaga et al., 2014), plants (Gidekel et al., 2003; Dassanayake et al., 2009; Deyholos, 2010; Champigny et al., 2013; Liu et al., 2013; Torales et al., 2013), and vertebrates (Wilcox et al., 2004; Protas et al., 2006; Qiu et al., 2012; Gross et al., 2013). However, recent NGS efforts that specifically target crustaceans in extreme environments have been embarked upon with very promising results (for examples see: Clark et al., 2011, Antarctic waters; Protas et al., 2011, freshwater caves; Harms et al., 2013, Arctic waters; von Reumont et al., 2014, anchialine caves; Wong et al., 2015, deep sea). For instance, Hinaux et al. (2013) used RNA-Seq to show that the loss of vision in the Mexican cavefish *Astyanax fasciatus* is probably due to relaxed selective pressures on their visual genes, which showed numerous deleterious mutations. A similar occurrence was reported by Tierney et al. (2015), who analyzed the transcriptomes of three cave-dwelling beetles and found evidence of convergent loss of opsin photoreceptor transcription by neutral processes. Likewise, von Reumont et al. (2014) pioneered one of the first examinations of an anchialine crustacean transcriptome, and revealed that the remipede *Xibalbanus*

tulumensis (Yager, 1987) is capable of producing and utilizing venom proteins for predation. This discovery not only provides evidence for the first and only venomous crustacean documented, but also illustrates the potential that NGS technologies offer to the biological and evolutionary study of anchialine cave ecosystems.

Concluding Remarks

Anchialine caves are unique ecosystems with highly specialized inhabitants, which are often endemic (Iliffe, 2002). As such, these unique ecosystems function as natural laboratories (Mejía-Ortíz & Hartnoll, 2006; Gonzalez et al., 2011) that allow us to test numerous hypotheses concerning adaptation, speciation, and evolution. Furthermore, cave ecosystems present us with the opportunity to study organisms existing in habitats and conditions perhaps analogous to those of our planet many millions of years ago (Por, 2007). The special adaptations and evolutionary processes that gave rise to extant extremophiles, including some cave organisms, grant us the ability to examine questions regarding the origin and early evolution of life on our planet, and applications relating to these (i.e. astrobiology) (Christin et al., 2010; Czyżewska, 2011; Gonzalez et al., 2011; Protas et al., 2011; Bonilla-Rosso et al., 2012). The unique processes and characteristics of anchialine caves (distribution, biogeochemistry and habitat stratification, chemosynthetic food-webs) and their biodiversity make them important communities to conserve in face of current anthropogenic threats (Myers et al., 2000; Iliffe, 2002; Porter, 2007; Mercado-Salas et al., 2013). Unfortunately, anchialine caves are often found in conflict with the impacts of anthropogenic forces such as tourism-driven habitat loss,

pollution by sewage, overexploitation of aquifers, climate change, and others (Iliffe et al., 1984; Sket, 1999; Iliffe, 2002). The distribution of these coastal caves in ‘desirable’ locations in the tropics often places them at a considerable disadvantage (Iliffe, 2002). Numerous stygobiont species follow patterns of regional and even single-cave endemism (Sket, 1999; De Grave et al., 2007), making them more prone to be severely impacted and becoming extinct as a result of pollution and habitat destruction. The opportunity to document and study anchialine cave biodiversity and evolution is a fleeting one (Wilson, 1985; Iliffe, 2002) and the potential for substantial discoveries is under threat of rapid decline and eventual disappearance.

Even though biological research in caves has seen significant advances in recent decades, new and emerging genomic technologies have just begun to scratch the surface of the underworld’s deepest mysteries. The adoption of these technologies not only will considerably expand the breadth of scientific questions that can be addressed and the depth with which these can be answered, but will surely provide us with necessary knowledge and tools to manage and conserve these intriguing and threatened habitats and their unique biodiversity.

Acknowledgments

We would like to thank Dr. José María Eirín-López and Dr. Kalai Mathee (Florida International University) for their valuable comments during the preparation of this manuscript. We are also grateful to Joseph Song-López, Jill Heinerth, and Tamara Thomsen for their diagrams and photographs of anchialine systems. This research was made possible in part by a grant from The Gulf of Mexico Research Initiative. Iliffe’s

anchialine cave research has been supported by grants from the National Science Foundation, NOAA, and National Geographic Society, among others.

References

- Adams M. & Humphreys W., 1993 – *Patterns of genetic diversity within selected subterranean fauna of the Cape Range peninsula, Western Australia: systematic and biogeographic implications*. Records of the Western Australian Museum, Supplement, **45**: 145-164.
- Álvarez F., Iliffe T.M., Gonzalez B. & Villalobos J.L., 2012 – *Triacanthoneus akumalensis, a new species of alpheid shrimp (Crustacea: Caridea: Alpheididae) from an anchialine cave in Quintana Roo, Mexico*. Zootaxa, **68**: 61-68.
- Álvarez F., Iliffe T.M. & Villalobos J.L., 2006 – *Macromaxillocarididae, a new family of stenopodidean shrimp from an anchialine cave in the Bahamas, with the description of Macromaxillocaris bahamaensis, n. gen.* Journal of Crustacean Biology, **26** (3): 366-378.
- Álvarez F., Villalobos J.L., Hendrickx M.E., Escobar-Briones E., Rodríguez-Almaraz G. & Campos E., 2014 – *Biodiversidad de crustáceos decápodos (Crustacea: Decapoda) en México*. Revista Mexicana de Biodiversidad, **85**: S208-S219.
- Anders F., 1956 – *Über Ausbildung und Vererbung der Körperfarbe bei Gammarus pulex ssp. subterraneus (Schneider), einer normalerweise pigmentlosen Höhlenform des gemeinen Bachflohkrebses*. Zeitschrift für Induktive Abstammungs- und Vererbungslehre, **87**: 567-579.
- Arendt J. & Reznick D., 2008 – *Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation?* Trends in Ecology & Evolution, **23** (1): 26-32.
- Baker B.J., Lesniewski R.A. & Dick G.J., 2012 – *Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume*. The ISME Journal, **6** (12): 2269-2279.
- Baker B.J., Sheik C.S., Taylor C.A., Jain S., Bhasi A., Cavalcoli J.D. & Dick G.J., 2013 – *Community transcriptomic assembly reveals microbes that contribute to deep-sea carbon and nitrogen cycling*. The ISME Journal, **7** (10): 1962-1973.
- Baratti M., Filippelli M., Nardi F., Messana G., 2010 – *Molecular phylogenetic relationships among some stygobitic cirolanid species (Crustacea, Isopoda)*. Contributions to Zoology, **79** (2): 57-67.

- Bauermeister J., Ramette A. & Dattagupta S., 2012 – *Repeatedly evolved host-specific ectosymbioses between sulfur-oxidizing bacteria and amphipods living in a cave ecosystem*. PLoS ONE, **7** (11): e50254.
- Beatty R., 1949 – *The pigmentation of cavernicolous animals. III. The carotenoid pigments of some amphipod crustacea*. Journal of Experimental Biology, **26**: 125-130.
- Becking L.E., Renema W., Santodomingo N.K., Hoeksema B.W., Tuti Y. & Voogd N.J., 2011 – *Recently discovered landlocked basins in Indonesia reveal high habitat diversity in anchialine systems*. Hydrobiologia, **677** (1): 89-105.
- Beddows P. & Hendrickson M., 2008 – *When the survey is not enough: temperature, salinity, and dye tracing reveal flow paths*. Proceedings of the 2007 National Cave and Karst Management Symposium, p. 198-203.
- Beddows P.A., Smart P.L., Whitaker F.F. & Smith S.L., 2007 – *Decoupled fresh-saline groundwater circulation of a coastal carbonate aquifer: Spatial patterns of temperature and specific electrical conductivity*. Journal of Hydrology, **346**: 18-32.
- Benvenuto C., Knott B. & Weeks S., 2015 – *Crustaceans of extreme environments*. In: Thiel M. & Watling L. (Eds.) *The Natural History of the Crustacea*. Oxford University Press, New York, p. 379-417.
- Bickford D., Lohman D.J., Sodhi N.S., Ng P.K.L., Meier R., Winker K., Ingram K.K. & Das I., 2007 – *Cryptic species as a window on diversity and conservation*. Trends in Ecology & Evolution, **22** (3): 148-155.
- Binladen J., Gilbert M.T.P., Bollback J.P., Panitz F., Bendixen C., Nielsen R. & Willerslev E., 2007 – *The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing*. PLoS ONE, **2** (2): e197.
- Bishop R.E., Humphreys W.F., Cukrov N., Žic V., Boxshall G.A., Cukrov M., Iliffe T.M., Kršinić F., Moore W.S., Pohlman J.W. & Sket B., 2015 – *“Anchialine” redefined as a subterranean estuary in a crevicular or cavernous geological setting*. Journal of Crustacean Biology, **35** (4): 511-514.
- Bishop R.E. & Iliffe T.M., 2012 – *Ecological physiology of the anchialine shrimp Barbouria cubensis: a comparison of epigeal and hypogean populations*. Marine Biodiversity, **42** (3): 303-310.
- Bonilla-Rosso G., Peimbert M., Alcaraz L.D., Hernández I., Eguiarte L.E., Olmedo-Alvarez G. & Souza V., 2012 – *Comparative metagenomics of two microbial mats at Cuatro Ciénegas Basin II: community structure and composition in oligotrophic environments*. Astrobiology, **12** (7): 659-73.

- Botello A., Iliffe T.M., Alvarez F., Juan C., Pons J. & Jaume D., 2013 – *Historical biogeography and phylogeny of Typhlatya cave shrimps (Decapoda: Atyidae) based on mitochondrial and nuclear data*. Journal of Biogeography, **40** (3): 594-607.
- Botosaneanu L. & Iliffe T.M., 2006 – *A new species of stygobitic cirolanid (Isopoda: Cirolanidae) from an anchialine cave on Abaco, the Bahamas*. Bulletin de L'Institut Royal des Sciences Naturelles de Belgique, **76**: 27-31.
- Bowman T. & Iliffe T.M., 1985 – *Mictocaris halope, a new unusual peracaridan crustacean from marine caves on Bermuda*. Journal of Crustacean Biology, **5** (1): 58-73.
- Bowman T., Yager J. & Iliffe T.M., 1985 – *Speonebalia cannoni, n. gen., n. sp., from the Caicos Islands, the first hypogean leptostracan (Nebaliacea: Nebaliidae)*. Proceedings of the Biological Society of Washington, **98** (2): 439-446.
- Boxshall G.A. & Defaye D., 2008 – *Global diversity of copepods (Crustacea: Copepoda) in freshwater*. Hydrobiologia, **595** (1): 195-207.
- Boxshall G.A. & Huys R., 1989 – *New tantulocarid, Stygotantulus stocki, parasitic on harpacticoid copepods, with an analysis of the phylogenetic relationships within the Maxillopoda*. Journal of Crustacean Biology, **9**: 126-140.
- Bracken H.D., De Grave S., Toon A., Felder D.L. & Crandall K.A., 2010 – *Phylogenetic position, systematic status, and divergence time of the Procarididea (Crustacea: Decapoda)*. Zoologica Scripta, **39** (2): 198-212.
- Bracken-Grissom H.D., Felder D.L., Vollmer N.L., Martin J.W. & Crandall K.A., 2012 – *Phylogenetics links monster larva to deep-sea shrimp*. Ecology and Evolution, **2** (10): 2367-73.
- Bradbury I.R., Hamilton L.C., Dempson B., Robertson M.J., Bourret V., Bernatchez L. & Verspoor E., 2015 – *Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure*. Molecular Ecology, **24** (20): 5130-5144.
- Bréhier F. & Jaume D., 2009 – *A new species of Pseudoniphargus (Crustacea, Amphipoda, Melitidae) from an anchialine cave on the French Mediterranean coast*. Zoosystema, **31** (1): 17-32.
- Bribiesca-Contreras G., Solís-Marín F.A., Laguarda-Figueras A. & Zaldívar-Riverón A., 2013 – *Identification of echinoderms (Echinodermata) from an anchialine cave in Cozumel Island, Mexico, using DNA barcodes*. Molecular Ecology Resources, **13** (6): 1137-1145.
- Brooks H.K., 1955 – *A crustacean from the Tesnus Formation (Pennsylvanian) of Texas*. Journal of Paleontology, **29** (5): 852-856.

- Bruce A.J. & Davie P., 2006 – *new anchialine shrimp of the genus Procaris from Christmas Island: the first occurrence of the Procarididae in the Indian Ocean (Crustacea: Decapoda: Caridea)*. Zootaxa, **1238**: 23-33.
- Bruce N.L. & Humphreys W.F. 1993 – *Haptolana pholeta, sp. nov., the first subterranean flabelliferan isopod crustacean (Cirolanidae) from Australia*. Invertebrate Taxonomy, **7**: 875-884
- Buhay J.E. & Crandall K.A., 2005 – *Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes*. Molecular Ecology, **14**: 4259-4273.
- Bybee S.M., Bracken-Grissom H.D., Haynes B.D., Hermansen R.A., Byers R.L., Clement M.J., Udall J.A., Wilcox E.R. & Crandall K.A., 2011 – *Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics*. Genome Biology and Evolution, **3**: 1312-1323.
- Bybee S.M., Bracken-Grissom H.D., Hermansen R.A., Clement M.J., Crandall K.A. & Felder D.L., 2011 – *Directed next generation sequencing for phylogenetics: An example using Decapoda (Crustacea)*. Zoologischer Anzeiger - A Journal of Comparative Zoology, **250 (4)**: 497-506.
- Cabezas P., Alda F., Macpherson E. & Machordom A., 2012 – *Genetic characterization of the endangered and endemic anchialine squat lobster Munidopsis polymorpha from Lanzarote (Canary Islands): management implications*. ICES Journal of Marine Science, **69 (6)**: 1030-1037.
- Caccone A. & Sbordoni V., 2001 – *Molecular biogeography of cave life: a study using mitochondrial DNA from bathysciine beetles*. Evolution, **55 (1)**: 122-130.
- Castoe T.A., de Koning A.P.J. & Pollock D.D., 2010 – *Adaptive molecular convergence. Molecular evolution versus molecular phylogenetics*. Communicative & Integrative Biology, **3 (1)**: 67-69.
- Chace F.A. & Manning R.B., 1972 – *Two new caridean shrimps, one representing a new family, from marine pools on Ascension Island (Crustacea: Decapoda: Natantia)*. Smithsonian Contributions to Zoology, **131**: 1-18.
- Champigny M.J., Sung W.W., Catana V., Salwan R., Summers P.S., Dudley S.A., Provart N.J., Cameron R.K., Golding G.B. & Weretilnyk E.A., 2013 – *RNA-Seq effectively monitors gene expression in Eutrema salsugineum plants growing in an extreme natural habitat and in controlled growth cabinet conditions*. BMC Genomics, **14**: 578.
- Christin P.A., Weinreich D.M. & Besnard G., 2010 – *Causes and evolutionary significance of genetic convergence*. Trends in Genetics, **26**: 400-405.

- Clark M.S., Thorne M.A.S., Toullec J.-Y., Meng Y., Guan L.L., Peck L.S. & Moore S., 2011 – *Antarctic krill 454 pyrosequencing reveals chaperone and stress transcriptome*. PLoS ONE, **6** (1): e15919.
- Coghill L.M., Darrin Hulsey C., Chaves-Campos J., García de Leon F.J. & Johnson S.G., 2014 – *Next generation phylogeography of cave and surface Astyanax mexicanus*. Molecular Phylogenetics and Evolution, **79** (1): 368-374.
- Corbera J., Segonzac M. & Cunha M.R., 2008 – *A new deep-sea genus of Nannastacidae (Crustacea, Cumacea) from the Lucky Strike hydrothermal vent field (Azores Triple Junction, Mid-Atlantic Ridge)*. Marine Biology Research, **4** (3): 180-192.
- Corey S., 1981 – *Comparative fecundity and reproductive strategies in seventeen species of the Cumacea (Crustacea: Peracarida)*. Marine Biology, **62** (1): 65-72.
- Cronn R., Knaus B.J., Liston A., Maughan P.J., Parks M., Syring J. V & Udall J., 2012 – *Targeted enrichment strategies for next-generation plant biology*. American Journal of Botany, **99** (2): 291-311.
- Culver D., Kane T.C. & Fong D., 1995 – *Adaptation and Natural Selection in Caves: The Evolution of Gammarus minus*. Harvard University Press, Cambridge, Massachusetts.
- Culver D. & Sket B., 2000 – *Hotspots of subterranean biodiversity in caves and wells*. Journal of Cave and Karst Studies, **62** (1): 11-17.
- Culver D.C. & Pipan T., 2009 – *The Biology of Caves and Other Subterranean Habitats*. Oxford University Press.
- Culver D.C., Pipan T. & Schneider K., 2009 – *Vicariance, dispersal and scale in the aquatic subterranean fauna of karst regions*. Freshwater Biology, **54** (4): 918-929.
- Czyżewska U.K., 2011 – *Exploring knowledge about microbes living in the extreme environments – the resources review*. ELBA Bioflux, **3** (1): 27-37.
- Daenekas J., Iliffe T.M., Yager J. & Koenemann S., 2009 – *Speleonectes kakuki, a new species of Remipedia (Crustacea) from anchialine and sub-seafloor caves on Andros and Cat Island, Bahamas*. Zootaxa, **2016**: 51-66.
- Danielopol D., 1981 – *Distribution of ostracods in the groundwater of the North Western Coast of Euboea (Greece)*. International Journal of Speleology, **11**: 91-103.
- Dassanayake M., Haas J.S., Bohnert H.J. & Cheeseman J.M., 2009 – *Shedding light on an extremophile lifestyle through transcriptomics*. The New Phytologist, **183** (3): 764-775.
- Dattagupta S., Schaperdoth I., Montanari A., Mariani S., Kita N., Valley J.W. & Macalady J.L., 2009 – *A novel symbiosis between chemoautotrophic bacteria and a freshwater cave amphipod*. The ISME Journal, **3** (8): 935-943.

- Davey J.W. & Blaxter M.L., 2010 – *RADSeq: next-generation population genetics*. Briefings in Functional Genomics, **9** (5): 416-423.
- Davey J.W., Hohenlohe P.A., Etter P.D., Boone J.Q., Catchen J.M. & Blaxter M.L., 2011 – *Genome-wide genetic marker discovery and genotyping using next-generation sequencing*. Nature Reviews Genetics, **12** (7): 499-510.
- Deyholos M.K., 2010 – *Making the most of drought and salinity transcriptomics*. Plant, Cell and Environment, **33**: 648-654.
- Dickson G.W., 1975 – *A preliminary study of heterotrophic microorganisms as factors in substrate selection of troglobitic invertebrates*. The NSS Bulletin, **37** (4): 89-93.
- Dubilier N., Bergin C. & Lott C., 2008 – *Symbiotic diversity in marine animals: the art of harnessing chemosynthesis*. Nature Reviews Microbiology, **6** (10): 725-740.
- Duncan T.K., 1984 – *Life history of Almyracuma proximoculo Jones and Burbanck, 1959 (Crustacea: Cumacea) from intertidal fresh-water springs on Cape Cod, Massachusetts*. Journal of Crustacean Biology, **4** (3): 356-374.
- Ekblom R. & Galindo J., 2011 – *Applications of next generation sequencing in molecular ecology of non-model organisms*. Heredity, **107** (1): 1-15.
- Ellegren H., 2008 – *Sequencing goes 454 and takes large-scale genomics into the wild*. Molecular Ecology, **17** (7): 1629-1635.
- Emerson M.J., 1991 – *Remipedia: Part 2 Paleontology*. San Diego Society of Natural History.
- Engel A.S., 2007 – *Observations on the biodiversity of sulfidic karst habitats*. Journal of Cave and Karst Studies, **69** (1): 187-206.
- Engel A.S., Porter M.L., Stern L.A., Quinlan S. & Bennett P.C., 2004 – *Bacterial diversity and ecosystem function of filamentous microbial mats from aphotic (cave) sulfidic springs dominated by chemolithoautotrophic "Epsilonproteobacteria"*. FEMS Microbiology Ecology, **51** (1): 31-53.
- Espinasa L. & Borowsky R.B., 2001 – *Origins and relationship of cave populations of the blind Mexican tetra, Astyanax fasciatus, in the Sierra de El Abra*. Environmental Biology of Fishes, **62** (1-3): 233-237.
- Felgenhauer B.E., Abele L.G. & Kim W., 1988 – *Reproductive morphology of the anchialine shrimp Procaris ascensionis (Decapoda: Procarididae)*. Journal of Crustacean Biology, **8** (3): 333-339.
- Figueroa D.F., 2011 – *Phylogenetic analysis of Ridgewayia (Copepoda: Calanoida) from the Galapagos and of a new species from the Florida Keys with a reevaluation of the phylogeny of Calanoida*. Journal of Crustacean Biology, **31** (1): 153-165.

- 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*. *Molecular Ecology*, **16** (2): 355-365.
- Flot J.-F., Wörheide G. & Dattagupta S., 2010 – *Unsuspected diversity of Niphargus amphipods in the chemoautotrophic cave ecosystem of Frasassi, central Italy*. *BMC Evolutionary Biology*, **10**: 171.
- Floyd R., Abebe E., Papert A. & Blaxter M., 2002 – *Molecular barcodes for soil nematode identification*. *Molecular Ecology*, **11** (4): 839-850.
- Foot A.D., Liu Y., Thomas G.W.C., Vinař T., Alföldi J., Deng J., Dugan S., Elk C.E. Van, Hunter M.E., Joshi V., Khan Z., Kovar C., Lee S.L., Lindblad-toh K., Mancina A., Nielsen R., Qin X., Qu J., Raney B.J., et al., 2015 – *Convergent evolution of the genomes of marine mammals*. *Nature Genetics*, **47** (3): 272-275.
- Fosshagen A., Boxshall G. & Iliffe T.M., 2001 – *The Epacteriscidae, a cave-living family of calanoid copepods*. *Sarsia*, **86**: 245-318.
- Fosshagen A. & Iliffe T.M., 2004 – *A new species of cave-living calanoid copepod from Grand Bahama*. *Sarsia*, **89** (5): 346-354.
- Friedrich M., 2013 – *Biological Clocks and Visual Systems in Cave-Adapted Animals at the Dawn of Speleogenomics*. *Integrative and Comparative Biology*, **53** (1): 50-67.
- Friedrich M., Chen R., Daines B., Bao R., Caravas J., Rai P.K., Zgmaajster M. & Peck S.B., 2011 – *Phototransduction and clock gene expression in the trogloliont beetle Ptomaphagus hirtus of Mammoth cave*. *The Journal of Experimental Biology*, **214**: 3532-3541.
- Genomic Resources Development Consortium, Havird J.C. & Santos S R., 2014 – *Genomic Resources Notes accepted 1 June 2014 – 31 July 2014*. *Molecular Ecology Resources*, **14** (6): 1322.
- Gidekel M., Destefano-Beltrán L., García P., Mujica L., Leal P., Cuba M., Fuentes L., Bravo L.A., Corcuera L.J., Alberdi M., Concha I. & Gutiérrez A., 2003 – *Identification and characterization of three novel cold acclimation-responsive genes from the extremophile hair grass Deschampsia antarctica Desv*. *Extremophiles*, **7** (6): 459-469.
- Gnewuch W. & Croker R., 1973 – *Macroinfauna of northern New England marine sand. I. The biology of Mancocuma stellifera Zimmer, 1943 (Crustacea: Cumacea)*. *Canadian Journal of Zoology*, **51** (10): 1011-1020.
- Goffredi S.K., 2010 – *Indigenous ectosymbiotic bacteria associated with diverse hydrothermal vent invertebrates*. *Environmental Microbiology Reports*, **2** (4): 479-488.

- Gonzalez B.C., Iliffe T.M., Macalady J.L., Schaperdoth I. & Kakuk B., 2011 – *Microbial hotspots in anchialine blue holes: initial discoveries from the Bahamas*. Hydrobiologia, **677** (1): 149-156.
- Gràcia F. & Jaume D., 2011 – *La fauna aquàtica dels hàbitats anquihalins i dolçaquícules de les cavitats balears*. ENDINS, **35**: 257-268.
- De Grave S., Cai Y. & Anker A., 2007 – *Global diversity of shrimps (Crustacea: Decapoda: Caridea) in freshwater*. Hydrobiologia, **595** (1): 287-293.
- Groh K.C., Vogel H., Stensmyr M.C., Grosse-Wilde E. & Hansson B.S., 2014 – *The hermit crab's nose — antennal transcriptomics*. Frontiers in Neuroscience, **7**: 266.
- Gross J.B., Furterer A., Carlson B.M. & Stahl B.A., 2013 – *An integrated transcriptome-wide analysis of cave and surface dwelling Astyanax mexicanus*. PLoS ONE, **8** (2): e55659.
- Grover C.E., Salmon A. & Wendel J.F., 2012 – *Targeted sequence capture as a powerful tool for evolutionary analysis*. American Journal of Botany, **99** (2): 312-319.
- Guțu M. & Iliffe T.M., 1989a – *Description of two new species of Tanaidacea (Crustacea) from the marine water caves of the Palau Islands (Pacific Ocean)*. Travaux du Muséum National d'Histoire Naturelle "Grigore Antipa." **30**: 169-180.
- Guțu M. & Iliffe T.M., 1989b – *Apseudes orghidani, a new species of Tanaidacea (Crustacea) from an anchialine cave on Bermuda*. Travaux du Muséum National d'Histoire Naturelle "Grigore Antipa." **30**: 161-167.
- Guțu M. & Iliffe T.M., 1998 – *Description of a new hirsutiid (ng, n. sp.) and reassignment of this family from order Mictacea to the new order, Bochusacea (Crustacea, Peracarida)*. Travaux du Muséum National d'Histoire Naturelle "Grigore Antipa." **40**: 93-120.
- Guțu M. & Iliffe T.M., 2011 – *Leptochelia vatulelensis (Crustacea: Tanaidacea), a new species from anchialine caves of the South-Western Pacific*. Travaux du Muséum National d'Histoire Naturelle "Grigore Antipa." **54** (2): 351-363.
- Hajibabaei M., Singer G.A.C., Hebert P.D.N. & Hickey D.A., 2007 – *DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics*. Trends in Genetics, **23** (4): 167-172.
- Hancock-Hanser B.L., Frey A., Leslie M.S., Dutton P.H., Archer F.I. & Morin P. a, 2013 – *Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics*. Molecular Ecology Resources, **13** (2): 254-268.
- Harms L., Frickenhaus S., Schiffer M., Mark F.C., Storch D., Pörtner H.-O., Held C. & Lucassen M., 2013 – *Characterization and analysis of a transcriptome from the*

- boreal spider crab* *Hyas araneus*. *Comparative Biochemistry and Physiology*, Part D, **8** (4): 344-351.
- Harr B. & Turner L.M., 2010 – *Genome-wide analysis of alternative splicing evolution among Mus subspecies*. *Molecular Ecology*, **19** (1): 228-239.
- Hedges D.J., Guettouche T., Yang S., Bademci G., Diaz A., Andersen A., Hulme W.F., Linker S., Mehta A., Edwards Y.J.K., Beecham G.W., Martin E.R., Pericak-Vance M. a, Zuchner S., Vance J.M. & Gilbert J.R., 2011 – *Comparison of three targeted enrichment strategies on the SOLiD sequencing platform*. *PLoS ONE*, **6** (4): e18595.
- Hinaux H., Poulain J., Da Silva C., Noirot C., Jeffery W.R., Casane D. & Rétaux S., 2013 – *De novo sequencing of Astyanax mexicanus surface fish and Pachón cavefish transcriptomes reveals enrichment of mutations in cavefish putative eye genes*. *PLoS ONE*, **8** (1): e53553.
- Hobbs H.H., 2012 – *Crustacea*. In: White W.B. & Culver D.C. (Eds.) *Encyclopedia of Caves*. Elsevier Academic Press, Amsterdam, Netherlands, p. 141-153.
- Hobbs H.H. & Daniel M.A., 1977 – *A review of the troglobitic decapod crustaceans of the Americas*. *Smithsonian Contributions to Zoology*, **244**: 1-183.
- Hoenemann M., Neiber M.T., Humphreys W., Iliffe T.M., Li D., Schram F.R. & Koenemann S., 2013 – *Phylogenetic analysis and systematic revision of Remipedia (Nectiopoda) from Bayesian analysis of molecular data*. *Journal of Crustacean Biology*, **33** (5): 603-619.
- Humphreys W., 1999 – *Physico-chemical profile and energy fixation in Bundera Sinkhole, an anchialine remiped habitat in north-western Australia*. *Journal of the Royal Society of Western Australia*, **82**: 89-98.
- Humphreys W., 2001a – *Milyeringa veritas (Eleotridae), a remarkably versatile Cave Fish from the Arid Tropics of Northwestern Australia*. *Environmental Biology of Fishes*, **62** (1): 297-313.
- Humphreys W., 2001b – *Groundwater calcrete aquifers in the Australian arid zone: the context to an unfolding plethora of stygal biodiversity*. *Records of the Western Australian Museum, Supplement*, **64**: 63-83.
- Humphreys W., 2002 – *The subterranean fauna of Barrow Island, northwestern Australia, and its environment*. *Mémoires de biospéologie*, **28**: 107-127.
- Humphreys W. & Danielopol D., 2006 – *Danielopolina (Ostracoda, Thaumatoctyprididae) on Christmas Island, Indian Ocean, a sea mount island*. *Crustaceana*, **78** (11): 1339-1352.

- Humphreys W., Tetu S., Elbourne L., Gillings M., Seymour J., Mitchell J. & Paulsen I., 2012 – *Geochemical and microbial diversity of Bundera sinkhole, an anchialine system in the eastern Indian ocean*. *Natura Croatica*, **21** (1): 59-63.
- Hunter R.L., Webb M.S., Iliffe T.M. & Alvarado Bremer J.R., 2007 – *Phylogeny and historical biogeography of the cave-adapted shrimp genus Typhlatya (Atyidae) in the Caribbean Sea and western Atlantic*. *Journal of Biogeography*, **35**: 65-75.
- Huys R., 1990 – *Campyloxiphos dineti gen. et spec. nov. from off Namibia and a redefinition of the Deoterthridae Boxshall and Lincoln (Crustacea: Tantulocarida)*. *Journal of Natural History*, **24** (2): 415-432.
- Huys R., Llewellyn-Hughes J., Olson P.D. & Nagasawa K., 2006 – *Small subunit rDNA and Bayesian inference reveal Pectenophilus ornatus (Copepoda incertae sedis) as highly transformed Mytilicolidae, and support assignment of Chondracanthidae and Xarifiidae to Lichomolgoidea (Cyclopoida)*. *Biological Journal of the Linnean Society*, **87** (3): 403-425.
- Ichihashi Y., Aguilar-Martínez J.A., Farhi M., Chitwood D.H., Kumar R., Millon L. V., Peng J., Maloof J.N. & Sinha N.R., 2014 – *Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape*. *Proceedings of the National Academy of Sciences of the United States of America*, **111** (25): E2616-E2621.
- Iglukowska A. & Boxshall G. a., 2013 – *Danielopolina revised: Phylogenetic relationships of the extant genera of the family Thaumatoocyprididae (Ostracoda: Myodocopa)*. *Zoologischer Anzeiger*, **252** (4): 469-485.
- Iliffe T.M., 1991 – *Anchialine fauna of the Galápagos Islands*. In: James M.J. (Ed.) *Galapagos Marine Invertebrates*. Springer US, p. 209-231.
- Iliffe T.M., 1992 – *Anchialine cave biology*. In: Camacho A.I. (Ed.) *The Natural History of Biospeleology*. Museo Nacional de Ciencias Naturales. Madrid, p. 614-636.
- Iliffe T.M., 1993 – *Fauna troglobia acuática de la Península de Yucatán*. In: Salazar-Vallejo S. & Gonzalez N. (Eds.) *Biodiversidad Marina Y Costera de México*. Com. Nal. Biodiversidad y CIQRO. Mexico, p. 673-686.
- Iliffe T.M., 2000 – *Anchialine cave ecology*. In: Wilkens H., Culver D.C. & Humphreys W.F. (Eds.) *Ecosystems of the World. Subterranean Ecosystems*. Elsevier Science, Amsterdam, Netherlands, p. 59-76.
- Iliffe T.M., 2002 – *Conservation of anchialine cave biodiversity*. *Proceedings of the Symposium Karst Frontiers*, **7**: 99-102.
- Iliffe T.M., 2012 – *Diving investigations of Bermuda's deep water caves*. *Natura Croatica*, **21** (1): 64-67.

- Iliffe T.M. & Botosaneanu L., 2006 – *The remarkable diversity of subterranean Cirolanidae (Crustacea: Isopoda) in the peri-Caribbean and Mexican Realm*. Bulletin de L'Institut Royal des Sciences Naturelles de Belgique, **76**: 5-26.
- Iliffe T.M. & Bowen C., 2001 – *Scientific cave diving*. Marine Technology Society Journal, **35** (2): 36-41.
- Iliffe T.M., Jickells T. & Brewer M., 1984 – *Organic pollution of an inland marine cave from Bermuda*. Marine Environmental Research, **12**: 173-189.
- Iliffe T.M. & Kornicker L.S., 2009 – *Worldwide diving discoveries of living fossil animals from the depths of anchialine and marine caves*. In: Lang M.A., Macintyre I.G. & Rützler K. (Eds.) *Proceedings of the Smithsonian Marine Science Symposium*. Smithsonian Contributions to the Marine Sciences, p. 269-280.
- Koelbel K., 1892 – *Beiträge zur Kenntnis der Crustaceen der Canarischen Inseln*. Annalen Des Naturhistorischen Museums in Wien, **7**(3): 105-116.
- Iwanaga N., Ide K., Nagashima T., Tomita T., Agari Y., Shinkai A., Kuramitsu S., Okada-Hatakeyama M., Kuzuyama T. & Nishiyama M., 2014 – *Genome-wide comprehensive analysis of transcriptional regulation by ArgR in Thermus thermophilus*. Extremophiles, **18** (6): 995-1008.
- Jaume D., Boxshall G.A. & Bamber R.N., 2006 – *A new genus from the continental slope off Brazil and the discovery of the first males in the Hirsutiidae (Crustacea: Peracarida: Bochusacea)*. Zoological Journal of the Linnean Society, **148**: 169-208.
- Jaume D., Boxshall G.A. & Humphreys W., 2001 – *New stygobiont copepods (Calanoida; Misophrioida) from Bundera Sinkhole, an anchialine cenote in north-western Australia*. Zoological Journal of the Linnean Society, **133**: 1-24.
- Jaume D. & Brehier F., 2005 – *A new species of Typhlatya (Crustacea: Decapoda: Atyidae) from anchialine caves on the French Mediterranean coast*. Zoological Journal of the Linnean Society, **144**: 387-414.
- Jaume D. & Christenson K., 2001 – *Amphi-Atlantic distribution of the subterranean amphipod family Metacrangonyctidae (Crustacea, Gammaridea)*. Contributions to Zoology, **70** (2): 99-125.
- Jaume D., Pinardo-Moya E. & Boxshall G.A., 2013 – *A presumed spelaeogriphacean crustacean from an upper Barremian wetland (Las Hoyas; Lower Cretaceous; Central Spain)*. Palaeontology, **56** (1): 15-28.
- Juan C., Guzik M.T., Jaume D. & Cooper S.J.B., 2010 – *Evolution in caves: Darwin's "wrecks of ancient life" in the molecular era*. Molecular Ecology, **19** (18): 3865-3880.
- Justice J.L., Weese D.A., Santos S.R., 2015 – *Phylogenetic utility, and variability in structure and content, of complete mitochondrial genomes among genetic lineages of*

- the Hawaiian anchialine shrimp* Halocaridina rubra Holthuis 1963
(Atyidae:Decapoda). Mitochondrial DNA. DOI: 10.3109/19401736.2015.1046161
- Kano Y. & Kase T., 2004 – *Genetic exchange between anchialine cave populations by means of larval dispersal: the case of a new gastropod species* Neritilia cavernicola. Zoologica Scripta, **33** (5): 423-437.
- Koenemann S., Bloechl A. & Martínez A., 2009a – *A new, disjunct species of Speleonectes (Remipedia, Crustacea) from the Canary Islands*. Marine Biodiversity, **39** (3): 215-225.
- Koenemann S. & Iliffe T.M., 2013 – *Remipedia*. In: von Vaupel Klein J., Charmantier-Daures M., Schram F. (Eds.) *Treatise on Zoology - Anatomy, Taxonomy, Biology. The Crustacea, Vol, 4A*. Brill. Leiden, Netherlands, p. 125-177.
- Koenemann S., Iliffe T.M. & Ham J. Van Der, 2003 – *Three new sympatric species of Remipedia (Crustacea) from Great Exuma Island, Bahamas Islands*. Contributions to Zoology, **72** (4): 227-252.
- Koenemann S., Iliffe T.M. & van der Ham J.L., 2007a – *Micropacteridae, a new family of Remipedia (Crustacea) from the Turks and Caicos Islands*. Organisms Diversity & Evolution, **7** (1): 52.e1-52.e14.
- Koenemann S., Iliffe T.M. & Yager J., 2004 – *Kaloketos pilosus, a new genus and species of Remipedia (Crustacea) from the Turks and Caicos Islands*. Zootaxa, **12** (618): 1-12.
- Koenemann S., Olesen J., Alwes F., Iliffe T.M., Hoenemann M., Ungerer P., Wolff C. & Scholtz G., 2009b – *The post-embryonic development of Remipedia (Crustacea) - Additional results and new insights*. Development Genes and Evolution, **219** (3): 131-145.
- Koenemann S., Schram F.R., Bloechl A., Iliffe T.M., Hoenemann M. & Held C., 2007b – *Post-embryonic development of remipede crustaceans*. Evolution and Development, **9** (2): 117-121.
- Koenemann S., Schram F.R., Hoenemann M. & Iliffe T.M., 2007c – *Phylogenetic analysis of Remipedia (Crustacea)*. Organisms Diversity & Evolution, **7** (1): 33-51.
- Kornicker L.S., Iliffe T.M. & Harrison-Nelson E., 2007 – *Ostracoda (Myodocopa) from anchialine caves and ocean blue holes*. Zootaxa, **1565**: 1-151.
- Kornicker L.S., Iliffe T.M. & Harrison-Nelson E., 2008 – *Ostracoda of Moss Town Blue Hole, Great Exuma Island, Great Bahama Bank (Crustacea: Ostracoda: Myodocopa)*. Proceedings of the Biological Society of Washington, **121** (2): 237-268.
- Kornicker L.S., Iliffe T.M. & Harrison-Nelson E., 2009 – *Ontogeny of Deeveya medix Kornicker, 1990, collected in anchialine and marine caves in the Bahama Islands*

- (*Ostracoda: Halocypridae: Deeveyinae*). Proceedings of the Biological Society of Washington, **122** (2): 171-205.
- Krejca J.K., 2005 – *Stygobite phylogenetics as a tool for determining aquifer evolution*. Unpublished PhD dissertation, The University of Texas at Austin.
- Larson H.K., Foster R., Humphreys W.F. & Stevens M.I., 2013 – *A new species of the blind cave gudgeon Milyeringa (Pisces: Gobioidae, Eleotridae) from Barrow Island, Western Australia, with a redescription of M. veritas Whitley*. Zootaxa, **3616** (2): 135-150.
- Lauro F.M. & Bartlett D.H., 2008 – *Prokaryotic lifestyles in deep sea habitats*. Extremophiles, **12** (1): 15-25.
- Leaché A.D., Wagner P., Linkem C.W., Böhme W., Papenfuss T.J., Chong R.A., Lavin B.R., Bauer A.M., Nielsen S. V., Greenbaum E., Rödel M.O., Schmitz A., LeBreton M., Ineich I., Chirio L., Ofori-Boateng C., Eniang E.A., Baha El Din S., Lemmon A.R., et al., 2014 – *A hybrid phylogenetic-phylogenomic approach for species tree estimation in african agama lizards with applications to biogeography, character evolution, and diversification*. Molecular Phylogenetics and Evolution, **79** (1): 215-230.
- Lefébure T., Douady C.J., Malard F. & Gibert J., 2007 – *Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (Niphargus rhenorhodanensis)*. Molecular Phylogenetics and Evolution, **42** (3): 676-686.
- Lemmon A.R., Emme S. A & Lemmon E.M., 2012 – *Anchored hybrid enrichment for massively high-throughput phylogenomics*. Systematic Biology, **61** (5): 727-744.
- Lemmon A.R. & Lemmon E.M., 2012 – *High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography*. Systematic Biology, **61** (5): 745-761.
- Lesniewski R.A., Jain S., Anantharaman K., Schloss P.D. & Dick G.J., 2012 – *The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs*. The ISME Journal, **6** (12): 2257-2268.
- Leys R., Watts C.H., Cooper S.J. & Humphreys W., 2003 – *Evolution of subterranean diving beetles (Coleoptera: Dytiscidae: Hydroporini, Bidessini) in the arid zone of Australia*. Evolution, **57** (12): 2819-2834.
- Li H. & Cooper R.L., 2001 – *Spatial familiarity in the blind cave crayfish, Orconectes australis packardii*. Crustaceana, **74** (5): 417-433.
- Liu S., Wang N., Zhang P., Cong B., Lin X., Wang S., Xia G. & Huang X., 2013 – *Next-generation sequencing-based transcriptome profiling analysis of Pohlia nutans reveals insight into the stress-relevant genes in Antarctic moss*. Extremophiles, **17** (3): 391-403.

- Lorentzen D., Koenemann S. & Iliffe T.M., 2007 – *Speleonectes emersoni*, a new species of *Remipedia* (Crustacea) from the Dominican Republic. *Zootaxa*, **1543**: 61-68.
- Mamanova L., Coffey A.J., Scott C.E., Kozarewa I., Turner E.H., Kumar A., Howard E., Shendure J. & Turner D.J., 2010 – *Target-enrichment strategies for next-generation sequencing*. *Nature Methods*, **7** (2): 111-118.
- Martens K., 2004 – *Crustacea: Ostracoda*. In: Gunn J. (Ed.) *Encyclopedia of Caves and Karst Science*. Fitzroy Dearborn, New York, p. 267-268.
- Martin J. & Davis G., 2001 – *An Updated Classification of the Recent Crustacea*. Natural History Museum of Los Angeles County, Los Angeles, California.
- Martin J.A. & Wang Z., 2011 – *Next-generation transcriptome assembly*. *Nature Reviews Genetics*, **12** (10): 671-682.
- McCormack J.E., Hird S.M., Zellmer A.J., Carstens B.C. & Brumfield R.T., 2013 – *Applications of next-generation sequencing to phylogeography and phylogenetics*. *Molecular Phylogenetics and Evolution*, **66** (2): 526-38.
- McCormack J.E., Maley J.M., Hird S.M., Derryberry E.P., Graves G.R. & Brumfield R.T., 2012 – *Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences*. *Molecular Phylogenetics and Evolution*, **62** (1): 397-406.
- Medina-Gonzalez R., Proudlove G., Chumba-Segura L. & Iliffe T.M., 2001 – *Threatened fishes of the world: Ophisternon infernale* (Hubbs, 1938)(Synbranchidae). *Environmental Biology of Fishes*, **62**: 170.
- Mejía-Ortíz L.M. & Hartnoll R., 2006 – *A new use for useless eyes in cave crustaceans*. *Crustaceana*, **79** (5): 593-600.
- Mejía-Ortíz L.M., Hartnoll R.G. & López-Mejía M. 2006 – *Progressive troglomorphy of ambulatory and sensory appendages in three Mexican cave decapods*. *Journal of Natural History*, **40** (5-6): 255-264.
- Mejía-Ortíz L.M., Hartnoll R.G. & Viccon-Pale J.A., 2003 – *A new stygobitic crayfish from Mexico, Procambarus cavernicola* (Decapoda: Cambaridae), with a review of cave-dwelling crayfishes in Mexico. *Journal of Crustacean Biology*, **23** (2): 391-401.
- Mejía-Ortíz L.M., López-Mejía M., Pakes J., Hartnoll R.G. & Zarza-González E., 2013 – *Morphological adaptations to anchialine environments in species of five shrimp families* (*Barbouria yanezi*, *Agostocaris bozanici*, *Procaris mexicana*, *Calliasmata nohochi* and *Typhlatya pearsei*). *Crustaceana*, **86** (5): 578-593.
- Mejía-Ortíz L.M., Yáñez G. & López-Mejía M., 2007 – *Echinoderms in an anchialine cave in Mexico*. *Marine Ecology*, **28**: 31-34.

- Meland K. & Willassen E., 2007 – *The disunity of “Mysidacea” (Crustacea)*. Molecular Phylogenetics and Evolution, **44** (3): 1083-1104.
- Mercado-Salas N.F., Morales-Vela B., Suárez-Morales E. & Iliffe T.M., 2013 – *Conservation status of the inland aquatic crustaceans in the Yucatan Peninsula, Mexico: shortcomings of a protection strategy*. Aquatic Conservation: Marine and Freshwater Ecosystems, **23** (6): 939-951.
- Metzker M.L., 2010 – *Sequencing technologies - the next generation*. Nature Reviews Genetics, **11** (1): 31-46.
- Meyer M., Stenzel U., Myles S., Prüfer K. & Hofreiter M., 2007 – *Targeted high-throughput sequencing of tagged nucleic acid samples*. Nucleic Acids Research, **35** (15): e97.
- Miller M.R., Dunham J.P., Amores A., Cresko W.A. & Johnson E.A., 2007 – *Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) marker*. Genome Research, **17** (2): 240-248.
- Mock T. & Kirkham A., 2012 – *What can we learn from genomics approaches in marine ecology? From sequences to eco-systems biology*. Marine Ecology, **33** (2): 131-148.
- Moore W.S., 1999 – *The subterranean estuary: a reaction zone of ground water and sea water*. Marine Chemistry, **65** (1-2): 111-125.
- Moritsch M.M., Pakes M.J. & Lindberg D.R., 2014 – *How might sea level change affect arthropod biodiversity in anchialine caves: a comparison of Remipedia and Atyidae taxa (Arthropoda: Altocrustacea)*. Organisms Diversity & Evolution, **14** (2): 225-235.
- Myers N., Mittermeier R.A., Mittermeier C.G., Da Fonseca G.A. & Kent J., 2000 – *Biodiversity hotspots for conservation priorities*. Nature, **403**: 853-858
- Mylroie J.E. & Mylroie J.R., 2011 – *Void development on carbonate coasts: creation of anchialine habitats*. Hydrobiologia, **677** (1): 15-32.
- Namiotko T., Wouters K., Danielopol D.L. & Humphreys W.F., 2004 – *On the origin and evolution of a new anchialine stygobitic Microceratina species (Crustacea, Ostracoda) from Christmas Island (Indian Ocean)*. Journal of Micropalaeontology, **23**: 49-59.
- Neiber M.T., Hansen F.C., Iliffe T.M., Gonzalez B.C. & Koenemann S., 2012 – *Molecular taxonomy of Speleonectes fuchscockburni, a new pseudocryptic species of Remipedia (Crustacea) from an anchialine cave system on the Yucatán Peninsula, Quintana Roo,.* Zootaxa, **3190**: 31-46.
- Neiber M.T., Hartke T.R., Stemme T., Bergmann A., Rust J., Iliffe T.M. & Koenemann S., 2011 – *Global biodiversity and phylogenetic evaluation of Remipedia (Crustacea)*. PLoS ONE, **6** (5): e19627.

- Neisch J., Pohlman J. & Iliffe T.M., 2012 – *The use of stable and radiocarbon isotopes as a method for delineating sources of organic material in anchialine systems*. *Natura Croatica*, **21**: 83-85.
- Ng P., Guinot D. & Iliffe T.M., 1996 – *Revision of the anchialine varunine crabs of the genus Orcovita Ng & Tomascik, 1994 (Crustacea: Decapoda: Brachyura: Grapsidae), with descriptions of four new species*. *The Raffles Bulletin of Zoology*, **44** (1): 109-134.
- Niemiller M.L., Graening G.O., Fenolio D.B., Godwin J.C., Cooley J.R., Pearson W.D., Fitzpatrick B.M. & Near T.J., 2013 – *Doomed before they are described? The need for conservation assessments of cryptic species complexes using an amblyopsid cavefish (Amblyopsidae: Typhlichthys) as a case study*. *Biodiversity and Conservation*, **22** (8): 1799-1820.
- Oakley T.H., 2005 – *Myodocopa (Crustacea: Ostracoda) as models for evolutionary studies of light and vision: multiple origins of bioluminescence and extreme sexual dimorphism*. *Hydrobiologia*, **538**: 179-192.
- Oakley T.H. & Cunningham C.W., 2002 – *Molecular phylogenetic evidence for the independent evolutionary origin of an arthropod compound eye*. *Proceedings of the National Academy of Sciences of the United States of America*, **99**: 1426-1430.
- Olesen J., Boesgaard T. & Iliffe T.M., 2015 – *The Unique Dorsal Brood Pouch of Thermosbaenacea (Crustacea, Malacostraca) and Description of an Advanced Developmental Stage of Tulumella unidens from the Yucatan Peninsula (Mexico), with a Discussion of Mouth Part Homologies to Other Malacostraca*. *PLoS ONE*, **10** (4): e0122463.
- Oakley T.H., Wolfe J.M., Lindgren A.R., Zaharoff A.K., 2012 – *Phylotranscriptomics to bring the understudied into the fold: monophyletic ostracoda, fossil placement, and pancrustacean phylogeny*. *Molecular Biology and Evolution*, **30**: 215-233.
- Olesen J., Martinsen S.V., Iliffe T.M. & Koenemann S., 2014 – *Remipedia*. In: Martin J., Olesen J. & Høeg J. (Eds.) *Atlas of Crustacean Larvae*. John Hopkins University Press, Baltimore, Maryland, USA, p. 84-89.
- Opsahl S.P. & Chanton J.P., 2006 – *Isotopic evidence for methane-based chemosynthesis in the Upper Floridan aquifer food web*. *Oecologia*, **150** (1): 89-96.
- O'Quin K.E., Yoshizawa M., Doshi P. & Jeffery W.R., 2013 – *Quantitative genetic analysis of retinal degeneration in the blind cavefish Astyanax mexicanus*. *PLoS ONE*, **8** (2): e57281.
- Orsi W.D., Edgcomb V.P., Christman G.D. & Biddle J.F., 2013 – *Gene expression in the deep biosphere*. *Nature*, **499**: 205-208.

- Page T.J., Humphreys W. & Hughes J.M., 2008 – *Shrimps down under: evolutionary relationships of subterranean crustaceans from Western Australia (Decapoda: Atyidae: Stygiocaris)*. PLoS ONE, **3** (2): e1618.
- Pakes M.J. & Mejía-Ortíz L.M., 2014 – *Chemosynthetic ectosymbiosis reported in the predatory anchialine cave endemic, Xibalbanus tulumensis (Yager, 1987) (Remipedia)*. Crustaceana, **87**(14): 1657-1667.
- Pakes M.J., Weis A.K. & Mejía-Ortiz L., 2014 – *Arthropods host intracellular chemosynthetic symbionts, too: cave study reveals an unusual form of symbiosis*. Journal of Crustacean Biology, **34** (3): 334-341.
- Pankey M.S., Minin V.N., Imholte G.C., Suchard M.A. & Oakley T.H., 2014 – *Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid*. Proceedings of the National Academy of Sciences of the United States of America, **111** (44): E4736-E4742.
- Peck S.B. & Finston T.L., 1993 – *Galapagos islands troglobites: the questions of tropical troglobites, parapatric distributions with eyed-sister-species, and their origin by parapatric speciation*. Mémoires de Biospéologie, **20**: 19-37.
- Peñalba J. V, Smith L.L., Tonione M. A, Sass C., Hykin S.M., Skipwith P.L., McGuire J. A, Bowie R.C.K. & Moritz C., 2014 – *Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms*. Molecular Ecology Resources, **14** (15): 1000-1010.
- Pesce G.L. & Iliffe T.M., 2002 – *New records of cave-dwelling mysids from the Bahamas and Mexico with description of Palaumysis bahamensis n. sp. (Crustacea: Mysidacea)*. Journal of Natural History, **36** (3): 265-278.
- Peterson B.K., Weber J.N., Kay E.H., Fisher H.S. & Hoekstra H.E., 2012 – *Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species*. PLoS ONE, **7** (5): e37135.
- Petrescu I., 1995 – *Cumaceans (Crustacea: Peracarida) from the South American coasts collected by the R/V “Vema”*. Travaux du Muséum d’Histoire Naturelle Grigore Antipa, **35**: 49-86.
- Petrescu I., 2003 – *Cumacea (Crustacea: Peracarida) from western tropical Atlantic*. Travaux du Muséum National d’Histoire Naturelle Grigore Antipa, **45**: 117-128.
- Petrescu I. & Iliffe T.M., 1992 – *Contributions to the knowledge of the Cumacea species from the British Blue Holes, Andros Island (Bahamas)*. Travaux du Muséum National d’Histoire Naturelle Grigore Antipa, **32**: 283-301.
- Petrescu I. & Iliffe T.M., 2009 – *New species of Cumacea (Crustacea: Peracarida: Cumacea) from Bahamas*. Travaux du Muséum National d’Histoire Naturelle Grigore Antipa, **52**: 127-141.

- Petrescu I., Iliffe T.M. & Sarbu S., 1993 – *Contributions to the knowledge of Cumacea (Crustacea) from the littoral waters of Jamaica Island, including the description of three new species (I)*. Travaux du Muséum National d'Histoire Naturelle Grigore Antipa, **33**: 1-3.
- Petrunina A.S., Neretina T.V., Mugué N.S., Kolbasov G.A., 2014 – *Tantulocarida versus Thecostraca: inside or outside? First attempts to resolve phylogenetic position of Tantulocarida using gene sequences*. Journal of Zoological Systematics and Evolutionary Research, **52** (2): 100-108.
- Phillips M.J., Page T.J., De Bruyn M., Huey J.A., Humphreys W.F., Hughes J.M., Santos S.R., Schmidt D.J. & Waters J.M., 2013 – *The linking of plate tectonics and evolutionary divergence*. Current Biology, **23** (14): R603-R605
- Pohlman J.W., 2011 – *The biogeochemistry of anchialine caves: progress and possibilities*. Hydrobiologia, **677** (1): 33-51.
- Pohlman J.W., Iliffe T.M. & Cifuentes L.A., 1997 – *A stable isotope study of organic cycling and the ecology of an anchialine cave ecosystem*. Marine Ecology Progress Series, **155**: 17-27.
- Poore G. & Humphreys W., 1992 – *First record of Thermosbaenacea (Crustacea) from the Southern Hemisphere: a new species from a cave in tropical Western Australia*. Invertebrate Systematics, **6** (3): 719-725.
- Por F., 2007 – *Ophel: a groundwater biome based on chemoautotrophic resources. The global significance of the Ayyalon cave finds, Israel*. Hydrobiologia, **592** (1): 1-10.
- Porter M.L., 2007 – *Subterranean biogeography: what have we learned from molecular techniques*. Journal of Cave and Karst Studies, **69** (1): 179-186.
- Porter M.L., Engel A.S., Kane T.C. & Kinkle B.K., 2009 – *Productivity-Diversity Relationships from Chemolithoautotrophically Based Sulfidic Karst Systems*. International Journal of Speleology, **38** (1): 27-40.
- Porter M.L., Meland K. & Price W., 2007 – *Global diversity of mysids (Crustacea-Mysida) in freshwater*. Hydrobiologia, **595** (1): 213-218.
- Protas M., Conrad M., Gross J.B., Tabin C. & Borowsky R., 2007 – *Regressive evolution in the Mexican cave tetra, Astyanax mexicanus*. Current Biology, **17** (5): 452-454.
- Protas M.E., Hersey C., Kochanek D., Zhou Y., Wilkens H., Jeffery W.R., Zon L.I., Borowsky R. & Tabin C.J., 2006 – *Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism*. Nature Genetics, **38** (1): 107-11.
- Protas M.E., Trontelj P. & Patel N.H.N., 2011 – *Genetic basis of eye and pigment loss in the cave crustacean, Asellus aquaticus*. Proceedings of the National Academy of Sciences of the United States of America, **108** (14): 5702-5707.

- Puillandre N., Macpherson E., Lambourdière J., Cruaud C., Boisselier-Dubayle M.C. & Samadi S., 2011 – *Barcoding type specimens helps to identify synonyms and an unnamed new species in Eumunida Smith, 1883 (Decapoda: Eumunididae)*. Invertebrate Systematics, **25** (4): 322.
- Qiu Q., Zhang G., Ma T., Qian W., Wang J., Ye Z., Cao C., Hu Q., Kim J., Larkin D.M., Auvil L., Capitanu B., Ma J., Lewin H. a, Qian X., Lang Y., Zhou R., Wang L., Wang K., et al., 2012 – *The yak genome and adaptation to life at high altitude*. Nature Genetics, **44** (8): 946-9.
- Ratnasingham S. & Hebert P., 2007 – *BOLD: the barcode of life data system* (www.barcodinglife.org). Molecular Ecology Notes, **7**: 355-364.
- Rehm P., Borner J., Meusemann K., von Reumont B.M., Simon S., Hadrys H., Misof B. & Burmester T., 2011 – *Dating the arthropod tree based on large-scale transcriptome data*. Molecular Phylogenetics and Evolution, **61** (3): 880-887.
- von Reumont B.M., Blanke A., Richter S., Alvarez F., Bleidorn C. & Jenner R. A, 2014 – *The first venomous crustacean revealed by transcriptomics and functional morphology: remipede venom glands express a unique toxin cocktail dominated by enzymes and a neurotoxin*. Molecular Biology and Evolution, **31** (1): 1-32.
- von Reumont B.M., Jenner R. A, Wills M. A, Dell'ampio E., Pass G., Ebersberger I., Meyer B., Koenemann S., Iliffe T.M., Stamatakis A., Niehuis O., Meusemann K. & Misof B., 2012 – *Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda*. Molecular Biology and Evolution, **29** (3): 1031-1045.
- Ribera I., Fresneda J., Bucur R., Izquierdo A., Vogler A.P., Salgado J.M. & Cieslak A., 2010 – *Ancient origin of a Western Mediterranean radiation of subterranean beetles*. BMC Evolutionary Biology, **10**: 29.
- von Rintelen K., Page T.J., Cai Y., Roe K., Stelbrink B., Kuhajda B.R., Iliffe T.M., Hughes J. & von Rintelen T., 2012 – *Drawn to the dark side: a molecular phylogeny of freshwater shrimps (Crustacea: Decapoda: Caridea: Atyidae) reveals frequent cave invasions and challenges current taxonomic hypotheses*. Molecular Phylogenetics and Evolution, **63** (1): 82-96.
- Rohland N. & Reich D., 2012 – *Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture*. Genome Research, **22** (5): 939-946.
- Rouch R., 1968 – *Contribution à la connaissance des Harpacticides hypogés (Crustacés-Copépodes)*. Annales de Speleologie, **23**: 9-167.
- Rouch R., 1994 – *Copepoda*. In: Juberthie C. & Decu V. (Eds.) *Encyclopaedia Biospeologica, Tome I*. Societe Internationale de Biospeologie. Moulis, France, p. 105-111.

- Russ A., Santos S.R. & Muir C., 2010 – *Genetic population structure of an anchialine shrimp, Metabetaeus lohena (Crustacea: Alpheidae), in the Hawaiian Islands*. Revista de Biología Tropical, **58** (1): 159-70.
- Sahl J.W., Gary M.O., Harris J.K. & Spear J.R., 2011 – *A comparative molecular analysis of water-filled limestone sinkholes in north-eastern Mexico*. Environmental Microbiology, **13** (1): 226-240.
- Santos S.R., 2006 – *Patterns of genetic connectivity among anchialine habitats: a case study of the endemic Hawaiian shrimp Halocaridina rubra on the island of Hawaii*. Molecular Ecology, **15** (10): 2699-2718.
- Sarbu S.M., Kane T.C. & Kinkle B., 1996 – *A chemoautotrophically based cave ecosystem*. Science, **272**: 1953-1955.
- Sathiamurthy E. & Voris H.K., 2006 – *Maps of Holocene sea level transgression and submerged lakes on the Sunda Shelf*. The Natural History Journal of Chulalongkorn University, **2**: 1-44.
- Savolainen V., Cowan R.S., Vogler A.P., Roderick G.K. & Lane R., 2005 – *Towards writing the encyclopedia of life: an introduction to DNA barcoding*. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences, **360**: 1805-1811.
- Seymour J., Humphreys W. & Mitchell J., 2007 – *Stratification of the microbial community inhabiting an anchialine sinkhole*. Aquatic Microbial Ecology, **50**: 11-24.
- Shendure J. & Ji H., 2008 – *Next-generation DNA sequencing*. Nature Biotechnology, **26** (10): 1135-1145.
- Shokralla S., Gibson J.F., Nikbakht H., Janzen D.H., Hallwachs W. & Hajibabaei M., 2014 – *Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens*. Molecular Ecology Resources, **14**: 892-901.
- Sket B., 1996 – *The ecology of anchihaline caves*. Trends in Ecology & Evolution, **11** (5): 221-225.
- Sket B., 1999 – *The nature of biodiversity in hypogean waters and how it is endangered*. Biodiversity and Conservation, **8** (10): 1319-1338.
- Smith D.E., Harrison S., Firth C.R. & Jordan J.T., 2011 – *The early Holocene sea level rise*. Quaternary Science Reviews, **30** (15-16): 1846-1860.
- Solís-Marín F.A. & Laguarda-Figueras A., 2010 – *A new species of starfish (Echinodermata: Asteroidea) from an anchialine cave in the Mexican Caribbean*. Revista Mexicana de Biodiversidad, **81** (3): 663-668.

- Spears T., Debry R.W., Abele L.G. & Chodyla K., 2005 – *Peracarid monophyly and interordinal phylogeny inferred from nuclear small-subunit ribosomal DNA sequences (Crustacea: Malacostraca: Peracarida)*. Proceedings of the Biological Society of Washington, **118** (1): 117-157.
- Stegner M.E., Stemme T., Iliffe T.M., Richter S. & Wirkner C.S., 2015 – *The brain in three crustaceans from cavernous darkness*. BMC Neuroscience, **16** (1): 19.
- Stemme T., Iliffe T.M., von Reumont B.M., Koenemann S., Harzsch S. & Bicker G., 2013 – *Serotonin-immunoreactive neurons in the ventral nerve cord of Remipedia (Crustacea): support for a sister group relationship of Remipedia and Hexapoda?*. BMC Evolutionary Biology, **13** (1): 119.
- Stoch F., 1995 – *The ecological and historical determinants of crustacean diversity in groundwaters, or: why are there so many species*. Mémoires de Biospéologie, **22**: 139-160.
- Stock J.H., 1986 – *The concept “anchialine” reconsidered*. Stygologia, **2**: 90-92.
- Stull G.W., Moore M.J., Mandala V.S., Douglas N.A., Kates H.-R., Qi X., Brockington S.F., Soltis P.S., Soltis D.E. & Gitzendanner M.A., 2013 – *A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes*. Applications in Plant Sciences, **1** (2): 1200497.
- Suárez-Morales E., Ferrari F.D. & Iliffe T.M., 2006 – *A new epacteriscid copepod (Calanoida: Epacteriscidae) from the Yucatan Peninsula, Mexico, with comments on the biogeography of the family*. Proceedings of the Biological Society of Washington, **119** (2): 222-238.
- Suárez-Morales E. & Iliffe T.M., 2005a – *A new Exumella (Crustacea: Copepoda: Ridgewayiidae) from anchialine waters of the western Caribbean, with comments on regional biogeography*. Bulletin of Marine Science, **77** (3): 409-423.
- Suárez-Morales E. & Iliffe T.M., 2005b – *A new Stygonitocrella Petkovski (Copepoda: Harpacticoida) from a cave in Northern Mexico with comments on the taxonomy of the genus*. Hydrobiologia, **544** (1): 215-228.
- Suárez-Morales E. & Iliffe T.M., 2007 – *A new genus of Ridgewayiidae (Copepoda: Calanoida) from a karstic cave of the western Caribbean*. Journal of Crustacean Biology, **27** (2): 339-350.
- Suárez-Morales E., Reid J.W., Fiers F. & Iliffe T.M., 2004 – *Historical biogeography and distribution of the freshwater cyclopine copepods (Copepoda, Cyclopoida, Cyclopinae) of the Yucatan Peninsula, Mexico*. Journal of Biogeography, **31**: 1051-1063.
- Suárez-Ulloa V., Fernández-Tajes J., Aguiar-Pulido V., Rivera-Casas C., González-Romero R., Ausio J., Méndez J., Dorado J. & Eirín-López J.M., 2013a – *The*

- CHROMEVALOA database: a resource for the evaluation of Okadaic Acid contamination in the marine environment based on the chromatin-associated transcriptome of the mussel Mytilus galloprovincialis*. *Marine Drugs*, **11** (3): 830-841.
- Suárez-Ulloa V., Fernández-Tajes J., Manfrin C., Gerdol M., Venier P. & Eirín-López J.M., 2013b – *Bivalve omics: state of the art and potential applications for the biomonitoring of harmful marine compounds*. *Marine Drugs*, **11**: 4370-4389.
- Tafe D.J. & Greenwood J.G., 1996a – *A new species of Schizotrema (Cumacea: Nannastacidae) from Moreton Bay, Queensland*. *Memoirs of The Queensland Museum*, **39**: 381-389.
- Tafe D.J. & Greenwood J.G., 1996b – *The Bodotriidae (Crustacea: Cumacea) of Moreton Bay, Queensland*. *Memoirs of The Queensland Museum*, **39**: 391-482.
- Tierney S.M., Cooper S.J.B., Saint K.M., Bertozzi T., Hyde J., Humphreys W.F. & Austin A.D., 2015 – *Opsin transcripts of predatory diving beetles: a comparison of surface and subterranean photic niches*. *Royal Society open science*, **2** (1): 140386.
- Toonen R.J., Puritz J.B., Forsman Z.H., Whitney J.L., Fernandez-Silva I., Andrews K.R. & Bird C.E., 2013 – *ezRAD: a simplified method for genomic genotyping in non-model organisms*. *PeerJ*, **1**: e203.
- Torales S.L., Rivarola M., Pomponio M.F., Gonzalez S., Acuña C. V, Fernández P., Lauenstein D.L., Verga A.R., Hopp H.E., Paniego N.B. & Poltri S.N.M., 2013 – *De novo assembly and characterization of leaf transcriptome for the development of functional molecular markers of the extremophile multipurpose tree species Prosopis alba*. *BMC Genomics*, **14**: 705.
- Trontelj P., Douady C.J., Fišer C., Gibert J., Gorički Š., Lefébure T., Sket B. & Zakšek V., 2009 – *A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts?*. *Freshwater Biology*, **54** (4): 727-744.
- Turk, S., Sket B., & Sarbu Ș., 1996 – *Comparison between some epigean and hypogean populations of Asellus aquaticus (Crustacea: Isopoda: Asellidae)*. *Hydrobiologia*, **337**: 161-170.
- Verovnik R., Sket B. & Trontelj P., 2004 – *Phylogeography of subterranean and surface populations of water lice Asellus aquaticus (Crustacea: Isopoda)*. *Molecular Ecology*, **13**: 1519-1532.
- Villacorta C., Jaume D., Oromí P. & Juan C., 2008 – *Under the volcano: phylogeography and evolution of the cave-dwelling Palmorchestia hypogaea (Amphipoda, Crustacea) at La Palma (Canary Islands)*. *BMC Biology*, **6**: 7.
- Wägele J.W., Holland B., Dreyer H., Hackethal B., 2003 – *Searching factors causing implausible non-monophyly: ssu rDNA phylogeny of Isopoda Asellota (Crustacea:*

- Peracarida*) and faster evolution in marine than in freshwater habitats. *Molecular Phylogenetics and Evolution*, **28**: 535-551.
- Walker-Smith G.K. & Poore G.C.B., 2001 – *A phylogeny of the Leptostraca (Crustacea) with keys to families and genera*. *Memoirs of Museum Victoria*, **58** (2): 383-410.
- Wang Z., Gerstein M. & Snyder M., 2009 – *RNA-Seq: a revolutionary tool for transcriptomics*. *Nature Reviews Genetics*, **10** (1): 57-63.
- Watanabe H., Kobayashi K., Kato Y., Oda S., Abe R., Tatarazako N. & Iguchi T., 2008 – *Transcriptome profiling in crustaceans as a tool for ecotoxicogenomics: Daphnia magna DNA microarray*. *Cell Biology and Toxicology*, **24** (6): 641-647.
- Wiens J.J., Chippindale P.T. & Hillis D.M., 2003 – *When are phylogenetic analyses misled by convergence? A case study in Texas cave salamanders*. *Systematic Biology*, **52** (4): 501-514.
- Wilcox T.P., García de León F.J., Hendrickson D. a & Hillis D.M., 2004 – *Convergence among cave catfishes: long-branch attraction and a Bayesian relative rates test*. *Molecular Phylogenetics and Evolution*, **31** (3): 1101-1113.
- Wilkens H., 2001 – *Convergent adaptations to cave life in the Rhamdia laticauda catfish group (Pimelodidae, Teleostei)*. *Environmental Biology of Fishes*, **62** (1-3): 251-261.
- Williams J.E., Johnson J.E., Hendrickson D.A., Contreras-Balderas S., Williams J.D., Navarro-Mendoza M., McAllister D.E. & Deacon J.E., 1989 – *Fishes of North America endangered, threatened, or of special concern: 1989*. *Fisheries*, **14** (6): 2-20.
- Wilson E., 1985 – *The biological diversity crisis*. *BioScience*, **35** (11): 700-706.
- De Wit P., Pespeni M.H., Ladner J.T., Barshis D.J., Seneca F., Jaris H., Therkildsen N.O., Morikawa M. & Palumbi S.R., 2012 – *The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis*. *Molecular Ecology Resources*, **12** (6): 1058-1067.
- Wong J.M., Pérez-Moreno J.L., Chan T.-Y., Frank T.M. & Bracken-Grissom H.D., 2015 – *Phylogenetic and transcriptomic analyses reveal the evolution of bioluminescence and light detection in marine deep-sea shrimps of the family Oplophoridae (Crustacea: Decapoda)*. *Molecular Phylogenetics and Evolution*, **83**: 278-292.
- Wu T., Ayres E., Li G., Bardgett R.D., Wall D.H. & Garey J.R., 2009 – *Molecular profiling of soil animal diversity in natural ecosystems: Incongruence of molecular and morphological results*. *Soil Biology and Biochemistry*, **41** (4): 849-857.
- Shen X., Tian M., Yan B., Chu K., 2015 – *Phylomitogenomics of Malacostraca (Arthropoda: Crustacea)*. *Acta Oceanologica Sinica*, **34** (2): 84-92.

- Yager J., 1981 – *Remipedia, a new class of Crustacea from a marine cave in the Bahamas*. Journal of Crustacean Biology, **1**: 328-333.
- Yager J., 1987 – *Speleonectes tulumensis, n. sp. (Crustacea, Remipedia) from 2 anchialine cenotes of the Yucatan Peninsula, Mexico*. Stygologia, **3**: 161-166.
- Yager J., 1994 – *Remipedia*. In: Juberthie C., Decu V. (Eds.) *Encyclopaedia Biospeologica, Tome I*. Societe Internationale de Biospeologie. Moulis, France, p. 87-90.
- Yager, J. & Humphreys, W., 1996 – *Lasionectes exleyi, sp. nov., the first remipede crustacean recorded from Australia and the Indian Ocean, with a key to the world species*. Invertebrate Systematics, **10** (1): 171-187.
- Zakšek V., Sket B., Gottstein S., Franjević D. & Trontelj P., 2009 – *The limits of cryptic diversity in groundwater: phylogeography of the cave shrimp Troglolaris anophthalmus (Crustacea: Decapoda: Atyidae)*. Molecular Ecology, **18** (5): 931-946.
- Zakšek V., Sket B. & Trontelj P., 2007 – *Phylogeny of the cave shrimp Troglolaris: Evidence of a young connection between Balkans and Caucasus*. Molecular Phylogenetics and Evolution, **42** (1): 223-35.
- Zeng V., Villanueva K.E., Ewen-Campen B.S., Alwes F., Browne W.E. & Extavour C.G., 2011 – *De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean Parhyale hawaiiensis*. BMC Genomics, **12** (1): 581.
- Zhang W., Chen J., Yang Y., Tang Y., Shang J. & Shen B., 2011 – *A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies*. PLoS ONE, **6** (3): e17915.

Figure Captions

Figure 1. Schematic representation of an anchialine cave system. A) “Blue hole”, “Cenote” or “Sinkhole” opening to the surface; B) Meteoric freshwater lens upper stratum; C) Halocline or mixing zone – often accompanied by a layer of hydrogen sulfide by-product of microbial activity; D) Hypoxic saltwater lower stratum – devoid of sunlight, food webs in this habitats are likely to depend on chemosynthetic microbial communities. Diagram by J.M. Song-López.

Figure 2. Anchialine systems can be found in a range of different forms including (but not limited to): A) karst cave systems (Crystal Cave, Bermuda – photo by Jill Heinerth), B) lava tubes (Jameos del Agua, Lanzarote, Canary Islands, Spain – photo by Jill Heinerth), and C) pools (Angel Pool, Bermuda – photo by Tamara Thomsen).

Figure 3. Examples of various crustacean taxa found in anchialine caves:

A) *Parhippolyte sterreri* (Decapoda), B) *Pseudoniphargus grandimanus* (Amphipoda), C) *Bahalana caicosana* (Isopoda), D) *Tulumella* sp. (Thermosbaenacea), E) *Mictocaris halope* (Mictacea), F) *Bermudamysis speluncola* (Mysida), G) *Cumella abacoensis* (Cumacea), H) *Ridgewayia* sp. (Calanoida), I) *Spelaeoecia* sp. (Ostracoda), J) *Cryptocorynetes* sp. (Remipedia). Photographs of anchialine crustaceans by Thomas M. Iliffe, PhD.

Figures

Figure 1.

Figure 2.

Figure 3.

Sanger DNA Sequencing: A methodology for sequencing DNA molecules based on in-vitro replication with the incorporation of labeled chain-terminating dideoxynucleotides. Sanger sequencing allows for the sequencing of longer DNA reads (typically up to ~1000 contiguous bases) in a single reaction. Despite its limitations of one sequence per reaction, it is still useful for smaller-scale applications. Its relatively longer reads are also of utility for the validation of Next-generation sequencing data.

Next-generation DNA Sequencing (NGS): A term used to describe a variety of modern high-throughput DNA sequencing technologies, including but not limited to: the Illumina platform, Roche 454 pyrosequencing, IonTorrent, Pacific Biosciences. They are more cost-effective than Sanger DNA sequencing, and in recent years their use has demonstrated their enormous potential for studies at the genomic and transcriptomic scales.

Targeted/Directed Sequencing: Refers to a type of sequencing where only a specific region of interest in the genome is sequenced for a particular application. It can be used in conjunction with next-generation sequencing technologies for cost-effectiveness, which also allows for projects of a much larger scale than with Sanger DNA sequencing technologies.

Metagenomics: It refers to the sequencing and study of genes across whole communities in an environmental sample. It is especially useful as it allows for the examination of microbes that are typically uncultured in laboratory settings.

DNA Barcoding: The use of a given genetic sequence as an identifying marker for a given species. The best loci to use for this purpose may vary among taxa, however most recent efforts have focused on the mitochondrial cytochrome c oxidase subunit I or COI (animals and most eukaryotes), the nuclear ribosomal internal transcribed spacer or ITS (fungi), and the chloroplast *rbcL* and *matK* genes (plants).

Glossary Box 1.

High-throughput sequencing: Refers to sequencing technologies that are able to generate vast amounts of data in a timely and cost-effective manner.

mtDNA: Mitochondrial DNA is the DNA contained within the mitochondria organelles in eukaryotic organisms. mtDNA is derived from bacterial genomes early in eukaryotic evolution, and thus has different evolutionary origins than nuclear DNA. In most organisms it is exclusively maternally inherited.

Haplotypes: Refers to a set of genetic variations in a DNA sequence that share common inheritance. The scale of these variations and determination of haplotypes can be from Single Nucleotide Polymorphisms (SNPs) in a particular locus to groups of alleles on the same chromosome that are inherited together.

RNA-Seq: The term refers to the high-throughput sequencing of RNA from a specific tissue or organism at a discrete point in time. This provides the research with a snapshot of what is occurring in terms of transcription in that precise moment. Transcriptomic data can be used for studies in a wide range of areas such as evolution, development, physiology, adaptations to changing environments, and responses to physicochemical challenges.

Glossary Box 2.

CHAPTER III

THE ROLE OF ISOLATION ON CONTRASTING PHYLOGEOGRAPHIC
PATTERNS IN TWO CAVE CRUSTACEANS

Jorge Luis Pérez-Moreno, Gergely Balázs, Blake Wilkins, Gabor Herczeg, Heather D.

Bracken-Grissom

This chapter was published in:

BMC Evolutionary Biology, 17:247, 2017.

Abstract

Background. The underlying mechanisms and processes that prompt the colonisation of extreme environments constitute major research themes of evolutionary biology and biospeleology. The special adaptations required to survive in these extreme habitats (low food availability, hypoxic waters, permanent darkness), and the geographical isolation of caves, nominate cave biodiversity as ideal subjects to answer long-standing questions concerning the interplay amongst adaptation, biogeography, and evolution. The present project aims to examine the phylogeographic patterns exhibited by two sympatric species of surface and cave-dwelling peracarid crustaceans (*Asellus aquaticus* and *Niphargus hrabei*), and in doing so elucidate the possible roles of isolation and exaptation in the colonisation and successful adaptation to the cave environment.

Results. Specimens of both species were sampled from freshwater hypogean (cave) and epigean (surface) habitats in Hungary, and additional data from neighbouring countries were sourced from Genbank. Sequencing of mitochondrial and nuclear loci revealed, through haplotype network reconstruction (TCS) and phylogenetic inference, the genetic structure, phylogeographic patterns, and divergence-time estimates of *A. aquaticus* and *N. hrabei* surface and cave populations. Contrasting phylogeographic patterns were found between species, with *A. aquaticus* showing strong genetic differentiation between cave and surface populations and *N. hrabei* lacking any evidence of genetic structure mediated by the cave environment. Furthermore, *N. hrabei* populations show very low levels of genetic differentiation throughout their range, which suggests the possibility of recent expansion events over the last few thousand years.

Conclusions. Isolation by cave environment, rather than distance, is likely to drive the genetic structuring observed between immediately adjacent cave and surface populations of *A. aquaticus*, a predominantly surface species with only moderate exaptations to subterranean life. For *N. hrabei*, in which populations exhibit a fully ‘cave-adapted’ (troglomorphic) phenotype, the lack of genetic structure suggests that subterranean environments do not pose a dispersal barrier for this surface-cave species.

Keywords. Adaptation; biospeleology; exaptation; evolution; phylogenetics; subterranean; troglomorphy.

Background

One of the major recurring themes in evolutionary biology and ecology is discerning the drivers of genetic differentiation and diversity among populations, and their interplay with the environment. These patterns can often be associated to a wide array of geographic and environmental factors that influence population differentiation by means of both adaptive (i.e. selection) and non-adaptive (i.e. genetic drift) processes. As it is often observed in natural systems, reduced gene flow due to geographic distance can often result in distinct patterns of genetic differentiation across a spatial continuum [1]. However, other factors besides geographic distance often affect gene-flow between populations. Biotic and abiotic interactions can impact gene-flow through a variety of mechanisms (such as local adaptation, selection against immigrants, and biased dispersal), which result in populations being isolated by environmental differences [2–4].

Isolation by environment can be identified as a driver for observed patterns of genetic differentiation when there is a positive correlation between genetic and environmental differentiation, but no correlation between genetic differentiation and spatial distances between populations (the latter being an indication of isolation by distance) [4,5]. It is important to note, however, that isolation by distance and isolation by environment are not mutually exclusive and their effects might be particularly challenging to disentangle when environmental variables and geographic distances co-vary [3]. Caves and other subterranean habitats show marked environmental differences with adjacent surface ecosystems with a sharp boundary, most notably the absence of light and associated ecological and biogeochemical conditions. Such habitat differences can constitute significant barriers to gene flow and population connectivity, which in turn lead to high levels of genetic differentiation even at relatively small spatial scales [6,7].

In addition to how genetic diversity is distributed across distributional space, the underlying mechanisms and processes that prompt the colonisation of extreme environments, and more specifically caves, constitute one of the major research themes of evolutionary biology [8–10]. There are two major hypotheses generally regarded to explain the transition from surface to subterranean habitats. The adaptive shift hypothesis suggests that the colonisation of subterranean habitats is a result of founder populations actively expanding into and colonising new niches [11], rather than by accidental stranding and persistence in the aphotic zones [12,13]. The ability of a species to successfully colonise these extreme environments, however, might be mediated by ecological filtering and thus requires specific exaptations to life in darkness [9,14–16]. Possible exaptations to cave life include morphological (e.g. reduced dependence on

vision, elongation of body and appendages) and physiological (e.g. tolerance to oligoxic conditions) characteristics that are already present in numerous species inhabiting benthic and interstitial ecosystems [17–19]. On the other hand, the climatic relict hypothesis states that a species may be forced to adapt to cave life as a result of environmental change that results in uninhabitable conditions on the surface (e.g. glaciation events) [18,20,21]. The actual mechanisms that gave rise to contemporary cave populations are likely to be a complex combination of both processes, and continue to be a subject of investigation. The estimation of phylogenetic relationships from genetic data of cave-dwellers offers the possibility of elucidating the mechanisms and processes that eventually lead to cave colonisations and the persistence of cave populations. This is especially the case when genetic data of both surface and cave-dwelling organisms are coupled with their present-day geographic distributions to infer ancient events (e.g. [22]). However, to fully understand the mechanisms behind cave colonisation events through present-day phylogeographic patterns it is imperative to incorporate approaches that consider the environment and ecology of the species under study, and therefore the underlying factors that ultimately drive their evolution.

The isopod *Asellus aquaticus* and the amphipod *Niphargus hrabei* are two aquatic crustacean species that serve as ideal models to explore questions regarding the colonisation, barriers to gene-flow, and evolution of cave fauna. *Asellus aquaticus* is a widespread species of freshwater isopod commonly found in surface waters throughout Europe [23]. This species is also known to occasionally colonise caves where its populations exhibit “troglomorphic” phenotypes [23,24]. Troglomorphy can be defined as the set of morphological, physiological, and behavioural characteristics associated with a

species transition to life in caves (e.g. enlarged sensory and ambulatory appendages, lack of pigmentation, loss of vision, etc. [25–28]. Contrastingly, the amphipod species *N. hrabei* is an atypical representative of an almost exclusively cave-dwelling genus that has escaped the confines of the subterranean environment to colonise surface habitats. Its distribution spans an extensive area of central and eastern Europe with ranges of up to 1,300 km [29], where it lives in sympatry with *A. aquaticus* (e.g. in the Danube River and its floodplains). Observations suggests that *N. hrabei* populations are troglomorphic throughout its range in both caves and surface waters (Figure 1; [28]), perhaps due to the ancient cave-origin of the genus *Niphargus*. The disposition to inhabit both surface and cave environments, geographical distributions, and life-history characteristics of these two crustacean species make them ideal study organisms to disentangle the effects of isolation by distance and/or isolation by environment and to reveal the mechanisms and processes at play during cave colonisation.

The present study examines the phylogeographic patterns exhibited by sympatric surface and cave-dwelling populations of *A. aquaticus* and *N. hrabei*. We aim to test the hypothesis that isolation by environment drives the patterns of genetic differentiation of surface-exapted *A. aquaticus*, but not those of cave-exapted *N. hrabei*, for which isolation by distance is the expected mechanism. The Molnár János thermal cave system and the immediately adjacent Malom Lake (Budapest, Hungary) provide a perfect natural experiment to address questions of cave colonisation, adaptation, and population differentiation. Despite marked environmental differences between hypogean and epigean habitats, both localities are inhabited by the species under study and are hydrologically interconnected with the Danube River Basin, one of the largest and most

species-rich natural floodplains in Europe [30,31]. Therefore, the patterns of genetic differentiation that emerge from this system will allow for a better understanding of the effects of isolation (distance and/or environment) and possible roles of exaptations in the evolution of these cave and surface populations.

Methods

Sample collection

Specimens were sampled from three main sites in Budapest, Hungary: The Molnár János thermal cave system, the adjacent thermal Malom Lake, and the Soroksár branch of the Danube River (Figure 2). The three sites are interconnected hydrologically and the two study species (*Asellus aquaticus* and *Niphargus hrabei*) inhabit all the sites. Additional specimens of *A. aquaticus* were sourced from other locations in Hungary (Table 1) and sequence data for *N. hrabei* from neighbouring countries were obtained from GenBank to aid in the analyses [29]. All of the samples were collected using a “Sket bottle” [32] and preserved individually in 99% ethanol for subsequent molecular analyses. All of the samples employed by this study are housed in the Florida International Crustacean Collection (FICC; North Miami, FL, USA). Additional metadata associated with each specimen is securely stored in the collection’s curated electronic database.

DNA extraction and amplification

Genomic DNA was extracted from each specimen's pereopods and/or antennae using the commercially available QIAGEN DNeasy Blood and Tissue Kit (Cat. No. 69506). Several mitochondrial and nuclear loci were selected in order to maximise the resolution at the scale of interest (population level). Specifically, for *A. aquaticus* the loci chosen were: two mitochondrial ribosomal genes (*12S* and *16S*), a mitochondrial protein-coding gene (cytochrome c oxidase subunit I, *COI*), and a 'Numt' (nuclear mitochondrial DNA segment) [33],[34] of NADH dehydrogenase 2 (hereby referred to as *PseudoND2*). For *N. hrabei* the sequenced loci included: a mitochondrial ribosomal gene (*16S*), a mitochondrial protein-coding gene (cytochrome c oxidase subunit I, *COI*), a nuclear ribosomal gene (internal transcribed spacer, *ITS*), and a nuclear protein-coding gene (*NaK*). These loci have proved to be useful in inferring intra- and interspecific relationships across the subphylum Crustacea (*12S*, *16S*, *COI*, *ITS*, *NaK*) [35–39] or were specifically targeted to increase population-level resolution (*PseudoND2*) [34]. Polymerase Chain Reaction (PCR) amplifications were performed in reactions containing DNA template, sterile non-DEPC treated water, forward and reverse primers, and of GoTaq® Green Master Mix (Promega, M712). The thermal cycling profiles consisted of an initial denaturing step of 1 minute at 94 °C, followed by 35-40 cycles of 30 seconds at 94 °C, annealing for 30s at 48°-62 °C (depending on primer set and species, Supplementary Materials), 1 min extension at 72°C, and then a final extension of 7 min at 72 °C. PCR products were sent to Beckman Coulter Genomics (Danvers, MA, USA) for amplicon purification using solid-phase reversible immobilisation (SPRI) beads, and subsequent sequencing reactions using BigDye Terminator v3.1. Post reaction dye

terminator removal was done using Agencourt CleanSEQ, after which both forward and reverse strands were sequenced on an Applied Biosystems PRISM 3730xl DNA Analyzer.

Data preparation and analyses of selection

The obtained sequences were visually inspected, quality trimmed, and cleaned manually with Geneious 8.0 [40]. Sequences from specimens heterozygous at nuclear loci were phased with PHASE v2.1 [41,42] and SeqPHASE [43]. In instances where haplotype reconstruction during phasing resulted in more than one pair of possible sequences, pairs with the highest posterior probabilities were retained for subsequent analyses. To further verify the appropriateness of the chosen loci for phylogenetic inference, tests for selection in protein-coding loci were conducted with MEGA 7 [44]. Due to the small number of substitutions in most of the dataset, Fisher's exact test of neutrality was chosen as the preferred option to detect evidence of selection [44,45]. For this purpose, synonymous and non-synonymous substitutions were estimated using the Nei-Gojobori method [46].

Haplotype network reconstruction

Haplotype networks were subsequently built in PopArt 1.7 [47] using TCS (Templeton-Crandall-Sing) Networks [48]. Haplotype networks allow for informative visualisation of genealogical information at shallow divergence levels. Although several

methodologies exist for constructing these networks, TCS was chosen for its effectiveness at recovering accurate population-scale phylogeographic patterns even when genetic differentiation is low (e.g. [49–52]). Subsequent to haplotype network reconstruction, the relative frequencies of the mitochondrial haplotypes identified were plotted on maps to visualise their geographic distributions in an intuitive manner.

Phylogenetic analyses

Individual and concatenated gene trees were estimated using Maximum Likelihood (GARLI 2.01 [53]) and Bayesian inference (MrBayes 3.2.6 [54]) methods as implemented in the CIPRES portal [55], after using PartitionFinder v1.1.1 [56] to identify the best-fit models of molecular evolution and partitioning schemes for the dataset (Supplementary Materials). Maximum Likelihood phylogenetic trees were reconstructed with an initial search for the best tree, using 10 parallel runs via GARLI 2.01. Additionally, 10,000 bootstrap replicates were generated in 40 independent runs to assess nodal support of the best tree. All ML trees were then summarised with a 95% consensus rule and annotated using the SumTrees.py python script from the DendroPy library [57]. Bayesian phylogenetic trees were inferred with the same partitioning scheme as in the ML analyses. MrBayes 3.2.6 was executed with two independent runs, each consisting of 4 chains running for 10M generations. The MCMC run was sampled every 1,000 generations, and a relative burn-in frequency of 25% was set for accurate posterior sampling. After assessing for convergence (Tracer v1.6 [58]), the SumTrees.py script was again invoked to extract the Maximum Clade Credibility Tree (MCCT) and to annotate

the phylogenetic trees' nodal support as posterior probabilities. Further, population trees were estimated under a multi-locus coalescent model using *BEAST (BEAST 2.4.0 [59]). Intraspecific divergence times of *A. aquaticus* and *N. hrabei* populations were concurrently estimated using molecular clock rate calibrations for peracarid crustaceans' *COI* (1.25% of substitutions per site per million years [60]; and between 0.34% and 0.76% of substitutions per site per Myr [61]), which were previously estimated for taxa closely related to our species of study (Stenasellidae and Niphargiidae). All of *BEAST and BEAST analyses were run in triplicate at Florida International University's High Performance Computing cluster (Panther) for 200M generations after which they were assessed for convergence using Tracer v1.6. The *BEAST speciation models for which there was no evidence for convergence were discarded. The runs using a Yule model of speciation as a tree prior with a strict molecular clock calibration were retained for subsequent analyses. After discarding 25% of the sampled trees as burn-in, nodal support was annotated as posterior probabilities on the MCCT of each population tree analysis (TreeAnnotator; [59]). For each species, population trees that included divergence time estimates were plotted using the *geoscale.phylo* function of the R package 'strap' [62]. These were then georeferenced on precomputed maps with custom scripts that made use of the R package 'phytools' [63].

Genealogical Sorting Indices (*gsi*) were calculated using the R package 'genealogicalSorting' to quantify intraspecific lineage divergence, which allowed for the evaluation of monophyly in each population [64,65]. Lastly, a modified approach to calculate *gsi* (pairwise-*gsi* or *pwgsi*) was used to independently quantify the divergence of every population-pair [66]. This modified approach thus prevented possible bias and

false-positives that could have arisen as by-products of the trees' topologies outside the main groups of interest [66]. Outgroup populations of *N. hrabei* were excluded from this final analysis due to sample size requirements of the *pwgsi* approach. Nonetheless, this exclusion bears no impact on the comparisons amongst the target populations (Molnár János Cave, Malom Lake, and the Danube River's Soroksár).

Demographic history

Additionally, we sought to better understand the demographic histories of *A. aquaticus* and *N. hrabei*, as reflected in their sequence data, by estimating changes in their population sizes over time. For this purpose, we conducted Extended Bayesian skyline (EBS) analyses [67]) as implemented in BEAST [59]. Extended Bayesian skyline analyses allow for the incorporation of multi-locus datasets to estimate population history over time along with an assessment of the estimations' uncertainty [67]. The parameters employed for these analyses were maintained as in the previous successful BEAST runs, with the exception of the priors associated to the species tree, which was set to "Coalescent Extended Bayesian Skyline". These analyses were also run in triplicate at Florida International University's High Performance Computing cluster (Panther) for 200M generations after which they were assessed for convergence using Tracer v1.6. EBS run logs were subsequently combined, after discarding 25% as burn-in, and the demographic histories of both species were plotted using custom R scripts for ease of visualisation and further inferences.

Results

DNA sequences and data deposition

A total of ~1,690 and ~2,757 base pairs (bp) of nucleotide sequence data were recovered for *A. aquaticus* (81 sequences for *12S*, 83 for *16S*, 76 for *COI*, and 84 for *PseudoND2*) and *N. hrabei* (55 sequences for *16S*, 54 for *COI*, 51 for *ITS*, and 58 for *NaK*) respectively. All sequence data from this project were curated, annotated with their respective metadata, and deposited in the NCBI's Genbank database to allow for their dissemination and future use by other researchers (See Supplementary Materials for accession numbers).

Testing for neutrality of selected loci

Fisher's exact test of neutrality was employed to determine the suitability of the selected loci for phylogeographic inference by ensuring that they are not being subject to selective pressures. The probability (P) of rejecting the null hypothesis of strict-neutrality in favor of the alternative hypothesis of positive selection was larger than 0.05 for all loci in both species, and therefore not considered significant at an alpha value (α) of 5%. The chosen loci were therefore deemed suitable for subsequent analyses.

Haplotype network reconstruction

Analyses of haplotype networks display evident genetic structuring in surface and cave *A. aquaticus*, with particularly distinct haplotypes differentiating Molnár János

Cave's population from Malom Lake's despite their spatial proximity (Figure 3). The haplotypes found in Molnár János Cave and Malom Lake are exclusive to each locality, with the exception of a single "surface-phenotype" cave specimen that shared a haplotype found in both Malom Lake and the western population of Lipót (Figures 3, 4). This strong differentiation and lack of shared haplotypes between the cave's and adjacent surface populations is not the case for *N. hrabei*, where Malom Lake's and Molnár János Cave's mtDNA haplotypes are the same and no genetic structuring is evident (Figure 5). Mitochondrial haplotypes (*COI* and *16S*) of *N. hrabei* are not only closely related, but also found widely distributed throughout its range (Figures 5, 6).

Phylogeographic and genealogical sorting analyses

The patterns observed in the haplotype network reconstructions are reflected in and confirmed by the concatenated phylogenetic trees (tree-files are available in the TreeBASE repository at: <http://purl.org/phylo/treebase/phyloids/study/TB2:S21599>). The final Maximum Likelihood and Bayesian trees for each species are nearly identical, with the exception of a few unsupported nodes, and are concordant with the *BEAST population trees estimated with all the sequenced loci (*A. aquaticus*, Figure 4; *N. hrabei*, Figure 6; see Supplementary Materials for evolutionary model selection details). Furthermore, these population trees show that cave and surface populations of *A. aquaticus* diverged from each other at least 60k years ago (Table 2). There is no evidence for genetic structuring between cave and surface populations of *N. hrabei*, and the phylogenetic split between these is not supported. The Pairwise Genealogical Sorting

Index (*pwgsi*) estimates follow patterns similar to those of the population tree topologies recovered (*A. aquaticus*, Table 3; *N. hrabei*, Table 4). In *A. aquaticus*, the distinction between the Molnár János Cave population and the adjacent Malom Lake is clear despite its proximity (*pwgsi* = 1.00, $p < 0.001$), with the former having higher affinities to the south-western population of Balatonfenyves (*pwgsi* = 0.22, $p < 0.001$). The *pwgsi* between Molnár János Cave's and Malom Lake's *N. hrabei* populations on the other hand, provides no evidence of genealogical differentiation between these (*pwgsi* = 0.14, $p = 0.52$). Nevertheless, both populations display a statistically significant but modest degree of reciprocal monophyly when compared to the next geographically closest population, the Soroksár branch of the Danube River (*pwgsi* = 0.46 and 0.48 respectively, both $p < 0.001$).

Demographic history

We conducted Extended Bayesian Skyline (EBS) analyses [67]), as implemented in BEAST [59], to evaluate the demographic history of our two study species and investigate if there is any evidence of possible climate-associated population changes. The EBS plot for *A. aquaticus* shows a gradual population contraction reaching a minimum approximately between 100 – 200 thousand years ago and a gradual recovery thereafter (Figure 7). Contrastingly, EBS analyses for *N. hrabei* point to a sharper decline beginning at a later date (~ 60 thousand years ago). The 95% H.P.D. interval suggests that an evident population bottleneck followed by a rapid expansion occurred

approximately 10 thousand years before present (Figure 8), roughly corresponding to the end of the Würm glaciation (~ 11,700 years ago).

Discussion

The most intriguing finding of the present study is the population differentiation between cave and surface populations of *Asellus aquaticus*, a pattern which is not reflected in *Niphargus hrabei*. Here, we will discuss the phylogeographic patterns in the light of alternate isolation mechanisms (geographic distance *vs.* environment). Second, we focus on the Molnár János Cave system, and discuss its potential role as a climatic refugium together with the role of exaptation in successful cave colonisation. Lastly, we conclude by illustrating future possibilities and directions for research in this emergent study system.

Contrasting phylogeographies: isolation by distance, environment or both?

Our first objective was to resolve the phylogeographic patterns between sympatric surface and cave populations of two crustaceans in order to investigate if the cave environment is acting as a barrier for dispersal and connectivity of populations. The *A. aquaticus* populations throughout Hungary are genetically diverse with each population being comprised mostly by distinct, but closely related, mitochondrial haplotypes exclusive to their respective localities. Phylogenetic and population tree analyses do provide strong statistical support for the genetic differentiation between Molnár János Cave's subterranean population and its epigeal counterparts. Our results further suggest

that the observed genetic and morphological differentiation between cave and surface isopods result from the cave acting as an isolating environment. Nevertheless, despite this differentiation, the Molnár János Cave *A. aquaticus* population still falls well within the species range for *A. aquaticus* in a phylogenetic context (Figure 4). Furthermore, the presence of a single haplotype (mtDNA H01) in all of the western Hungarian populations sampled for this study suggests that movement over large distances does occur, suggesting a smaller role for geographical distance (vs. environment) as a driver of genetic differentiation (Figure 4). The relatively high haplotypic diversity found in the Danube River (Soroksár, in comparison to the other localities examined) further supports its role as a dispersal avenue for isopods inhabiting surface waters. Effective dispersal in surface environments, but not into the cave, is further evident in our phylogenetic analyses by the lack of nodal support for the differentiation of the surface populations.

Interestingly, movement of individuals from Malom Lake into the cave was detected by the sampling of a single isopod with both surface haplotype (mtDNA H14) and phenotype (Figure 3 & 4). Personal observations confirm that, in rare instances, surface isopods can be found in the aphotic zone well within Molnár János Cave. However, surface isopods do not persist in the cave habitat and the lack of shared mitochondrial and nuclear haplotypes suggests infrequent or non-existent admixture between these two genetically close but morphologically distinct populations (Figure 3 & 4). A similar pattern was observed in Slovenian and Romanian *A. aquaticus* where no population connectivity was found between troglomorphic cave isopods and other nearby populations from surface waters [24]. This recurrent pattern could be explained by a variety of mechanisms ultimately driven by the environmental differences between

subterranean and epigean habitats [2–4]. It is thus feasible that surface individuals who wander into the cave are outcompeted by the troglomorphic resident population (i.e. competitive exclusion) before successful breeding takes place, and/or that hybrid individuals are at a significant fitness disadvantage that prevents their genes from persisting in the cave population [3]. Investigating and understanding which exact mechanisms might be at play in the Molnár János Cave system undoubtedly constitutes an important question to address in future studies.

Unlike that of *A. aquaticus*, haplotype network reconstruction and population tree analyses of *N. hrabei* show no evidence of genetic structuring between surface and cave populations (Figure 5 & 6). In fact, haplotypic diversity seems to be relatively low throughout its range (Figure 6). The only phylogeographic pattern recovered is the segregation of Hungarian populations as a distinct clade (Figure 6). However, geography does not explain the inclusion of Danube River (Soroksár) individuals within the clade comprised by the Austrian, Serbian, and Romanian specimens. Low genetic diversity, weak genetic structuring despite large geographical distances, and lack of statistical support for any of the populations sampled, suggest that this species' modern populations resulted from a recent expansion event and do not provide any clear evidence for isolation by distance nor environment. This is further evidence by the results of our EBS analyses that suggest a population bottleneck for *N. hrabei* at approximately 10 thousand years ago followed by a rapid expansion (Figure 8). It is important to note that sampling for some of the localities examined was limited and this could have a negative effect on statistical support for those populations. Nevertheless, the three main target populations of the present study (Molnár János cave, Malom Lake and Danube River [Soroksár])

were sufficiently sampled to adequately evaluate the population structure of surface vs. adjacent cave populations. Therefore, for *N. hrabei*, whose populations exhibit a fully ‘troglomorphic’ phenotype in all environments, the lack of genetic structure at this scale suggests that subterranean environments do not pose a barrier for this species. Further analyses employing high-resolution data (e.g. genome-wide SNPs) from next-generation sequencing methodologies would undoubtedly be of advantage to clarify whether this lack of genetic structure is truly due to unimpeded movement in and out of the cave or a by-product of *N. hrabei*’s recent colonisation of the habitats under investigation [10].

The Molnár János thermal cave system: a climatic refugium and a possible role for exaptation

Divergence-time estimates, calibrated with peracarid *COI* molecular clock rates [60,61], place the divergence of *A. aquaticus* populations from Molnár János Cave and Malom Lake at approximately 60,000 to 140,000 years ago (Figure 4). This relatively recent split falls within the Pleistocene, a period of time during which severe climatic changes associated with glaciation events impacted the geographical distributions of numerous taxa across the globe [21,68]. Even though Hungary was not directly glaciated, hydrological changes in the region would have altered the viability of this crustacean’s surface populations and shifted their latitudinal distributions [21,68].

Caves have been known to act as refuges during rapid environmental changes on the surface [21,23,69–71]. Molnár János Cave’s waters are exclusively fed by thermomineral springs with a constant water temperature (~ 24°C), which increases its suitability as a refugium for organisms with exaptations that help them subsist in cave

environments. Being a moderately exapted species, it is possible that *A. aquaticus* would have been able to take refuge in the cave where it remained isolated from surface populations. This isolation eventually resulted in the emergence of troglomorphic phenotypes via adaptive and neutral processes. Upon cessation of this isolation, it is possible that competitive exclusion prevented new and/or returning surface populations from successfully invading the cave and vice versa. This mechanism would be in accordance with the observed phylogeographic patterns. Extended Bayesian Skyline Plot analyses illustrate a population decline for *A. aquaticus* and the possibility of the aforementioned events occurring approximately 100-200 thousand years ago (Figure 7). It is also possible that *A. aquaticus* may have had a constant food source independent from the surrounding surface environment, as bacterial communities in Molnár János Cave, upon which *A. aquaticus* feeds (pers. obs.), have been shown to thrive via chemoautotrophic processes [72].

Niphargus hrabei is likely to have colonised Molnár János Cave thousands of years later, as suggested by the divergence-time estimates (Figure 6). *Niphargus hrabei* cave and surface populations appear to be panmictic and show no evidence of isolation by the cave environment or of competitive exclusion within the cave. They have successfully colonised the cave from surface populations and appear to have no limitations with dispersing from and into cave environments. *Niphargus hrabei*'s facility for dispersal and its exceptional adaptability to markedly different habitats is reflected by an atypical large distributional range [29]. This adaptability is also evidenced by its unimpeded presence despite putative competitors in Molnár János Cave: the isopod *A. aquaticus* and two other species of *Niphargus* that are yet to be described (pers. obs.).

However, an alternative explanation would be that older populations of *N. hrabei* once inhabited the Molnár János Cave, but were outcompeted by the obligate cave-dwellers in absence of migrants from surface waters. In fact, our EBS analyses show an abrupt population decline with a bottleneck and subsequent expansion at approximately 10 thousand years ago (Figure 8) that roughly coincides with the end of the Würm glaciation, event which took place approximately between 115,000 and 11,700 years ago [73]. The end of the Würm glaciation is known for great climatic variability [74] with major temperature fluctuations that played a significant role in shaping the modern distributional patterns of other European species ([73,75,76]). Nevertheless, niphargiid coexistence in other caves has also been previously explained by evolutionary and ecological processes such as niche differentiation [77]. Understanding the exact mechanisms by which these processes take place continues to be an important research theme in evolutionary biology and biospeleology.

Future Prospects

The Molnár János Cave system and its inhabitants serve as ideal models for phylogeographic and biospeleological studies in an evolutionary context. While the present study has provided significant insights into the phylogeographic histories of two species and their transition into and out of caves, important questions remain to be answered. Further analyses will greatly aid in the understanding of the exact causes of the observed patterns, as well as in the elucidation of the mechanisms by which exaptations have helped them thrive in such extreme environments. An integrative approach incorporating different sources of molecular data (e.g. genomic, transcriptomic, epigenetic, etc.) has been initiated and will be of definitive advantage to address these

outstanding questions [10]. Advances in modern molecular methodologies will undoubtedly enable future high-resolution studies of the adaptive processes that underlie the contrasting phylogeographic patterns revealed by this study.

Abbreviations

DEPC: Diethylpyrocarbonate

DNA: Deoxyribonucleic acid

EBS: Extended Bayesian Skyline

FICC: Florida International Crustacean Collection

gsi: Genealogical Sorting Index

HPD: Highest Posterior Density

MCCT: Maximum Clade Credibility Topology

MCMC: Markov chain Monte Carlo

ML: Maximum Likelihood

mtDNA: Mitochondrial DNA

NCBI: National Center for Biotechnology Information

Numt: Nuclear mitochondrial DNA segment

PCR: Polymerase Chain Reaction

pwgsi: Pairwise Genealogical Sorting Index

SPRI: Solid Phase Reversible Immobilization

TCS: Templeton-Crandall-Sing

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Availability of data and materials

The dataset supporting the conclusions of this article is available in the NCBI's Genbank database. Accession numbers are detailed in the Supplementary Materials included with this publication. Phylogenetic trees resulting from each step of the study are available in the TreeBASE repository at: <http://purl.org/phylo/treebase/phylows/study/TB2:S21599>

Competing interests

No competing interests to declare.

Funding

J.P.M. was supported by the Philip M. Smith Graduate Research Grant for Cave and Karst Research from the Cave Research Foundation and The Crustacean Society Scholarship in Graduate Studies. This work was partially funded by the National Science Foundation (Doctoral Dissertation Improvement Grant #1701835) awarded to J.P.M and

H.B.G. G.B. and G.H. were supported by the Hungarian Scientific Research Fund (#K-105517).

Authors' contributions

J.P.M. and G.B. conceived the project; J.P.M., G.B., G.H., and H.B.G. designed the experiments; J.P.M., G.B., and B.W. collected the data; J.P.M. analysed the data and led the writing; G.H. and H.B.G. funded the project and provided conceptual advice. J.P.M., G.B., B.W., G.H, and H.B.G. discussed the results and implications, interpreted the data, participated in the writing, and provided critical contributions towards the final version of the manuscript.

Acknowledgments

We would like to thank Dr. Péter Borza for his assistance in the collection of specimens of *Asellus aquaticus* from Hungarian localities. We also extend our gratitude to Attila Hosszú, Szabolcs Storozynski, Péter Zsoldos, and Zoltán Bauer for their help as support divers during the collection of specimens in Molnár János Cave. The authors would like to acknowledge the Instructional & Research Computing Center (IRCC) at Florida International University for providing High Performance Computing resources that have contributed to the research results reported within this article. This is contribution #69 of the Marine Education and Research Center of the Institute for Water and the Environment at Florida International University.

References

1. Wright S. Isolation by Distance. *Genetics*. 1943;28:114–38.
2. Nosil P, Egan SP, Funk DJ. Heterogeneous Genomic Differentiation Between Walking-Stick Ecotypes: “Isolation By Adaptation” and Multiple Roles for Divergent Selection. *Evolution*. 2008;62:316–36.
3. Wang I, Bradburd G. Isolation by Environment. *Ecology Letters*. 2014;5649–62.
4. Manthey JD, Moyle RG. Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. *Molecular Ecology*. 2015;24:3628–38.
5. Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*. 2013;22:5983–99.
6. Mathieu J, Jeannerod F, Hervant F, Kane TC. Genetic differentiation of *Niphargus rhenorhodanensis* (Amphipoda) from interstitial and karst environments. *Aquatic Sciences*. 1997;59:39–47.
7. Tobler M, Dewitt TJ, Schlupp I, García de León FJ, Herrmann R, Feulner PGD, et al. Toxic hydrogen sulfide and dark caves: phenotypic and genetic divergence across two abiotic environmental gradients in *Poecilia mexicana*. *Evolution*. 2008;62:2643–59.
8. Juan C, Guzik MT, Jaume D, Cooper SJB. Evolution in caves: Darwin’s “wrecks of ancient life” in the molecular era. *Molecular Ecology*. 2010;19:3865–80.
9. Benvenuto C, Knott B, Weeks S. Crustaceans of extreme environments. In: Thiel M, Watling L, editors. *The Natural History of the Crustacea*. New York: Oxford University Press; 2015. p. 379–417.
10. Pérez-Moreno JL, Iliffe TM, Bracken-Grisson HD. Life in the Underworld: Anchialine cave biology in the era of speleogenomics. *International Journal of Speleology*. 2016;45:149–70.
11. Rouch R, Danielopol D. L’origine de la faune aquatique souterraine, entre le paradigme du refuge et le modèle de la colonisation active. *Stygologia*. 1987;3:345–72.
12. Kane TC, Culver DC, Jones RT. Genetic structure of morphologically differentiated populations of the amphipod *Gammarus minus*. *Evolution*. 1992;46:272–8.

13. Kano Y, Kase T. Genetic exchange between anchialine cave populations by means of larval dispersal: the case of a new gastropod species *Neritilia cavernicola*. *Zoologica Scripta*. 2004;33:423–37.
14. Gould SJ, Vrba ES. Exaptation—a missing term in the science of form. *Paleobiology*. 1982;8:4–15.
15. Gibert J, Deharveng L. Subterranean Ecosystems : A Truncated Functional Biodiversity. *BioScience*. 2002;52:473–81.
16. Gilgado JD, Ledesma E, Cuesta E, Arrechea E, Zapata de la Vega JL, Sánchez-Ruiz A, et al. *Dima assoi* Pérez Arcas 1872 (Coleoptera: Elateridae): from montane to hypogean life. An example of exaptations to the subterranean environment? *Annales de la Société entomologique de France (N.S.)*. 2014;50:264–71.
17. Humphreys W, Eberhard S. Subterranean fauna of Christmas Island, Indian Ocean. *Helictite*. 2001;37:59–74.
18. Leys R, Watts CH, Cooper SJ, Humphreys W. Evolution of subterranean diving beetles (Coleoptera: Dytiscidae: Hydroporini, Bidessini) in the arid zone of Australia. *Evolution*. 2003;57:2819–34.
19. Suárez-Morales E, Iliffe TM. A new *Exumella* (Crustacea: Copepoda: Ridgewayiidae) from anchialine waters of the western Caribbean, with comments on regional biogeography. *Bulletin of Marine Science*. 2005;77:409–23.
20. Barr TC, Holsinger JR. Speciation in Cave Faunas. *Annual Review of Ecology and Systematics*. 1985;16:313–37.
21. Bryson RWJ, Prendini L, Savary WE, Pearman PB. Caves as microrefugia: Pleistocene phylogeography of the troglomorphic North American scorpion *Pseudouroctonus reddelli*. *BMC Evolutionary Biology*. 2014;14.
22. Villacorta C, Jaume D, Oromí P, Juan C. Under the volcano: phylogeography and evolution of the cave-dwelling *Palmorchestia hypogaea* (Amphipoda, Crustacea) at La Palma (Canary Islands). *BMC Biology*. 2008;6:7.
23. Verovnik R, Sket B, Trontelj P. The colonization of Europe by the freshwater crustacean *Asellus aquaticus* (Crustacea: Isopoda) proceeded from ancient refugia and was directed by habitat connectivity. *Molecular Ecology*. 2005;14:4355–69.
24. Konec M, Prevorsek S, Sarbu SM, Verovnik R, Trontelj P. Parallels between two geographically and ecologically disparate cave invasions by the same species, *Asellus aquaticus* (Isopoda, Crustacea). *Journal of Evolutionary Biology*. 2015;28:n/a–n/a.

25. Turk S, Sket B, Sarbu S. Comparison between some epigean and hypogean populations of *Asellus aquaticus* (Crustacea: Isopoda: Asellidae). *Hydrobiologia*. 1996;337:161–70.
26. Li H, Cooper R. Spatial familiarity in the blind cave crayfish, *Orconectes australis packardii*. *Crustaceana*. 2001;74:417–33.
27. Li H, Cooper RL. The effect of ambient light on blind cave crayfish: social interactions. *Journal of Crustacean Biology*. 2002;22:449–58.
28. Mejía-Ortíz LM, Hartnoll R. A new use for useless eyes in cave crustaceans. *Crustaceana*. 2006;79:593–600.
29. Copilas-Ciocianu D, Fišer C, Borza P, Balázs G, Angyal D, Petrusek A. Low intraspecific genetic divergence and weak niche differentiation despite wide ranges and extensive sympatry in two epigean *Niphargus* species (Crustacea: Amphipoda). *Zoological Journal of the Linnean Society*. 2017;
30. Schiemer F. Conservation of biodiversity in floodplain rivers. *River Systems*. 1999;11:423–38.
31. Danielopol DL, Pospisil P. Hidden biodiversity in the groundwater of the Danube flood plain national park (Austria). *Biodiversity & Conservation*. 2001;10:1711–21.
32. Chevaldonné P, Sket B, Marschal C, Lejeusne C, Calado R. Improvements to the “Sket bottle”: a simple manual device for sampling small crustaceans from marine caves and other cryptic habitats. *Journal of Crustacean Biology*. 2008;28:185–8.
33. Lopez JV, Yuhki N, Masuda R, Modi W, O’Brien SJ. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution*. 1994;39:174–90.
34. Miraldo A, Hewitt GM, Dear PH, Paulo OS, Emerson BC. Numts help to reconstruct the demographic history of the ocellated lizard (*Lacerta lepida*) in a secondary contact zone. *Molecular Ecology*. 2012;21:1005–18.
35. Bracken HD, De Grave S, Toon A, Felder DL, Crandall KA. Phylogenetic position, systematic status, and divergence time of the Procarididea (Crustacea: Decapoda). *Zoologica Scripta*. 2010;39:198–212.
36. Bracken H, Grave SD, Felder D. Phylogeny of the infraorder Caridea based on mitochondrial and nuclear genes (Crustacea: Decapoda). In: Martin J, Crandall K, Felder DL, editors. *Decapod Crustacean Phylogenetics*. Boca Raton: Taylor and Francis/CRC Press; 2009. p. 274–98.
37. Bracken-Grissom HD, Enders T, Jara CG, Crandall KA. Molecular diversity of two freshwater anomuran crab species in Southern Chile (Decapoda: Anomura:

- Aeglidae) compared to associated morphometric differences. In: Held C, Schubart C, Koenemann S, editors. *Phylogeography and population genetics in Crustacea*. Boca Raton: Taylor and Francis/CRC Press; 2011. p. 305–22.
38. Bracken-Grissom HHD, Cannon ME, Cabezas P, Feldmann RM, Schweitzer CE, Ah Yong ST, et al. A comprehensive and integrative reconstruction of evolutionary history for *Anomura* (Crustacea: Decapoda). *BMC Evolutionary Biology*. 2013;13:128–128.
 39. Bybee SM, Bracken-Grissom HD, Hermansen RA, Clement MJ, Crandall KA, Felder DL. Directed next generation sequencing for phylogenetics: An example using Decapoda (Crustacea). *Zoologischer Anzeiger - A Journal of Comparative Zoology*. 2011;250:497–506.
 40. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9.
 41. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*. 2001;68:978–89.
 42. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *The American Journal of Human Genetics*. 2003;73:1162–9.
 43. Flot J-F. SEQPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources*. 2010;10:162–6.
 44. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*. 2016;33:1870–4.
 45. Zhang J, Kumar S, Nei M. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. *Molecular Biology and Evolution*. 1997;14:1335–8.
 46. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*. 1986;3:418–26.
 47. Leigh JW, Bryant D. POPART : full-feature software for haplotype network construction. Nakagawa S, editor. *Methods in Ecology and Evolution*. 2015;6:1110–6.
 48. Templeton AR, Crandall KA, Sing CF. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*. 1992;132:619–33.

49. Gerber AS, Templeton AR. Population Sizes and Within-Deme Movement of *Trimerotropis saxatilis* (Acrididae), a Grasshopper with a Fragmented Distribution. *Oecologia*. 1996;105:343–50.
50. Gomez-Zurita J, Petitpierre E, Juan C. Nested cladistic analysis, phylogeography and speciation in the *Timarcha goettingensis* complex (Coleoptera, chrysomelidae). *Mol Ecol*. 2000;9:557–70.
51. Johnson JB, Jordan S. Phylogenetic divergence in leatherside chub (*Gila copei*) inferred from mitochondrial cytochrome b sequences. *Mol Ecol*. 2000;9:1029–35.
52. Turner TF, Trexler JC, Harris JL, Haynes JL. Nested cladistic analysis indicates population fragmentation shapes genetic diversity in a freshwater mussel. *Genetics*. 2000;154:777–85.
53. Zwickl DJ. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas at Austin; 2006. Available from: <http://molevol.lysine.umiacs.umd.edu/molevolfiles/garli/zwicklDissertation.pdf>
54. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572–4.
55. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), 2010. IEEE; 2010. p. 1–8.
56. Lanfear R, Calcott B, Ho SYW, Guindon S. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular Biology and Evolution*. 2012;29:1695–701.
57. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. *Bioinformatics*. 2010;26:1569–71.
58. Rambaut A, Suchard M, Drummond A. Tracer v1.6 [Internet]. 2014. Available from: <http://beast.bio.ed.ac.uk/Tracer>
59. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*. 2007;7:214.
60. Ketmaier V, Argano R, Caccone a. Phylogeography and molecular rates of subterranean aquatic Stenasellid Isopods with a peri-Tyrrhenian distribution. *Molecular Ecology*. 2003;12:547–55.
61. Lefébure T, Douady CJ, Gouy M, Trontelj P, Briolay J, Gibert J. Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments. *Molecular Ecology*. 2006;15:1797–806.

62. Bell MA, Lloyd GT. strap: an R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence. *Palaeontology*. 2015;58:379–89.
63. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things): phytools: R package. *Methods in Ecology and Evolution*. 2012;3:217–23.
64. Cummings MP, Neel MC, Shaw KL. A genealogical approach to quantifying lineage divergence. *Evolution; international journal of organic evolution*. 2008;62:2411–22.
65. R Development Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2008. Available from: <http://www.R-project.org>
66. Winter DJ, Trewick SA, Waters JM, Spencer HG. The genealogical sorting index and species delimitation [Internet]. 2016 Jan. Report No.: biorxiv;036525v1. Available from: <http://biorxiv.org/lookup/doi/10.1101/036525>
67. Ho SYW, Shapiro B. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*. 2011;11:423–34.
68. Hewitt GM. The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology*. 2004;1.
69. Howarth FG. Ecology of cave arthropods. *Annual Review of Entomology*. 1983;28:365–89.
70. Prendini L, Francke OF, Vignoli V. Troglomorphism, trichobothriotaxy and typhlochactid phylogeny (Scorpiones, Chactioidea): more evidence that troglobitism is not an evolutionary dead-end. *Cladistics*. 2010;26:117–42.
71. Allegrucci G, Todisco V, Sbordoni V. Molecular phylogeography of *Dolichopoda* cave crickets (Orthoptera, Rhaphidophoridae): A scenario suggested by mitochondrial DNA. *Molecular Phylogenetics and Evolution*. 2005;37:153–64.
72. Borsodi AK, Knáb M, Krett G, Makk J, Márialigeti K, Erőss A, et al. Biofilm Bacterial Communities Inhabiting the Cave Walls of the Buda Thermal Karst System, Hungary. *Geomicrobiology Journal*. 2012;29:611–27.
73. Schmitt T, Seitz A. Intraspecific allozymatic differentiation reveals the glacial refugia and the postglacial expansions of European *Erebia medusa* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society*. 2001;74:429–58.
74. Ashworth A. A late-glacial insect fauna from Red Moss, Lancashire, England. *Insect Systematics & Evolution*. 1992;3:211–24.
75. Coope GR. The response of insect faunas to glacial-interglacial climatic fluctuations. *Philosophical Transactions: Biological Sciences*. 1994;344:19–26.

76. Hewitt GM. Post-glacial re-colonization of European biota. *Biological journal of the Linnean Society*. 1999;68:87–112.
77. Trontelj P, Blejec A, Fišer C. Ecomorphological convergence of cave communities. *Evolution*. 2012;66:3852–65.

Tables

Table 1. Specimens, locations, and type of habitat in which the crustacean populations were sampled.

Species	N	Locality	Coordinates	Habitat
<i>Asellus aquaticus</i>	20	Soroksár, Budapest, Hungary	47.4360697 N, 19.0878143 E	Epigean
		Molnar Janos Cave, Budapest,		
	20	Hungary	47.5181846 N, 19.036064 E	Hypogean
		Malom Lake, Budapest,		
	20	Hungary	47.5181167 N, 19.036075 E	Epigean
	4	Lipót, Hungary	47.86316 N, 17.458875 E	Epigean
	6	Polgár, Hungary	47.869443 N, 21.200598 E	Epigean
<i>Niphargus hrabei</i>	6	Balatonfenyves, Hungary	46.65515 N, 17.498538 E	Epigean
	10	Cserdi, Hungary	46.06575 N, 17.991012 E	Epigean
	20	Soroksár, Budapest, Hungary	47.4360697 N, 19.0878143 E	Epigean
		Molnar Janos, Budapest		
	18	Hungary	47.5181846 N, 19.036064 E	Hypogean
<i>*Niphargus</i> sp. nov. 1	20	Malom Lake, Budapest Hungary	47.5181167 N, 19.036075 E	Epigean
	13	Molnar Janos, Budapest,	47.5181846 N, 19.036064 E	Hypogean

		Hungary		
		Diabaz Cave, Nagyvisnyo,		
<i>*Niphargus forroi</i>	2	Hungary	48.08809 N, 20.46627 E	Hypogean

*Used as outgroups for the analyses.

Table 2. Divergence time estimations between Molnar Janos Cave and phylogenetically closest surface populations of the peracarid crustaceans under study (thousands of years [95% H.P.D.]).

Species	Molecular Clock Rate	
	Ketmaier et al. (2003)	Lefébure et al. (2006)
<i>Asellus aquaticus</i>	60.81 (28.68, 96.53)	139.21 (67.07, 222.26)
<i>Niphargus hrabei</i>	Surface/cave split not supported	Surface/cave split not supported

Table 3. Pairwise Genealogical Sorting Index estimates for each population pair of *Asellus aquaticus*. P-values assess significance that exclusive ancestry observed for every population pair is greater than that which would be observed at random.

Population 1	Population 2	Bayesian Phylogeny		Maximum Likelihood Phylogeny	
		<i>pwgsi</i>	p-value	<i>pwgsi</i>	p-value
Molnar Janos Cave	Soroksár (Danube)	0.71	< 0.001	0.68	< 0.001
Molnar Janos Cave	Malom Lake	1.00	< 0.001	1.00	< 0.001
Molnar Janos Cave	Lipót	0.46	< 0.001	0.35	< 0.001

Molnar Janos Cave	Polgar	0.89	< 0.001	0.89	< 0.001
Molnar Janos Cave	Balatonfenyves	0.22	< 0.001	0.22	< 0.001
Molnar Janos Cave	Cserdi	0.56	< 0.001	0.57	< 0.001
Soroksár (Danube)	Malom Lake	1.00	< 0.001	1.00	< 0.001
Soroksár (Danube)	Lipot	0.49	< 0.001	0.43	< 0.001
Soroksár (Danube)	Polgar	0.89	< 0.001	0.89	< 0.001
Soroksár (Danube)	Balatonfenyves	0.17	0.003	0.13	0.001
Soroksár (Danube)	Cserdi	0.46	< 0.001	0.43	< 0.001
Malom Lake	Lipot	1.00	< 0.001	1.0	< 0.001
Malom Lake	Polgar	1.00	< 0.001	1.0	< 0.001
Malom Lake	Balatonfenyves	1.00	< 0.001	1.0	< 0.001
Malom Lake	Cserdi	1.00	< 0.001	1.0	< 0.001
Lipot	Polgar	1.00	< 0.001	1.0	< 0.001
Lipot	Balatonfenyves	0.38	< 0.001	0.23	< 0.001
Lipot	Cserdi	0.47	< 0.001	0.47	< 0.001
Polgar	Balatonfenyves	0.78	< 0.001	0.78	< 0.001
Polgar	Cserdi	0.83	< 0.001	0.83	< 0.001
Balatonfenyves	Cserdi	0.15	0.012	0.15	< 0.001

Table 4. Pairwise Genealogical Sorting Index estimates for each population pair of *Niphargus hrabei*. P-values assess significance that exclusive ancestry observed for every population pair is greater than that which would be observed at random.

Population 1	Population 2	Bayesian Phylogeny		Maximum Likelihood Phylogeny	
		<i>pwgsi</i>	p-value	<i>pwgsi</i>	p-value
Soroksár (Danube)	Molnar Janos Cave	0.46	<0.001	0.15	<0.001
Soroksár (Danube)	Malom Lake	0.48	<0.001	0.18	<0.001
Molnar Janos Cave	Malom Lake	0.14	0.521	0.06	0.005

Figure Captions

Figure 1. *Asellus aquaticus* displays contrasting phenotypes in and out of the cave, while *Niphargus hrabei* exhibits the same phenotype in both environments.

Figure 2. Schematic illustration of our three main sampling localities within Budapest, Hungary. Red circles indicate exact sites within Molnár János Cave (Rákos Rock) and surface environments (Malom Lake and Danube River [Soroksár]).

Figure 3. Haplotype networks of *Asellus aquaticus*: nuclear (A- PseudoND2) and mitochondrial (B- *16S*, *12S*, and *COI*) loci. Node diameter and annotation denote sample sizes. Colours represent sampling locality, as illustrated in the legend.

Figure 4. Divergence time estimates (x axis in thousands of years) of *Asellus aquaticus* populations (calculated with a multi-locus coalescent model in *BEAST; outgroups not shown) and the distribution of its populations with relative mtDNA haplotype frequencies throughout Hungary. Phylogenetic and population tree analyses support the inclusion of the cave phenotype as part of the species, but with evident population structuring and marked morphological differences as a result of the cave environment.

Figure 5. Haplotype networks of *Niphargus hrabei*: nuclear (A- *ITS*; B- *NaK*) and mitochondrial (A- *16S* and *COI*) loci. Node diameter and annotation denote sample sizes (alleles in the case of nuclear genes). Colours represent sampling locality, as illustrated in the legend.

Figure 6. Divergence time estimates (x axis in thousands of years) of *Niphargus hrabei* populations (calculated with a multi-locus coalescent model in *BEAST; outgroups not shown) and the distribution of its populations with relative *COI* haplotype frequencies in our three main Hungarian sites and neighbouring populations. Phylogenetic and population tree analyses do not support any evident genetic structuring between cave and surface populations.

Figure 7. An Extended Bayesian Skyline Plot illustrates the demographic history of the sampled Hungarian *Asellus aquaticus* populations over the last 800,000 years. The x-axis represents time before present in thousands of years, while the y-axis denotes the estimated population (θ) assuming a generation time of one year.

Figure 8. An Extended Bayesian Skyline Plot of *Niphargus hrabei*'s demographic history, from 100,000 years ago to today, depicts a possible genetic bottleneck at approximately 10,000 years ago. The x-axis represents time before present in thousands of years, while the y-axis denotes the estimated population (θ) assuming a generation time of one year.

Figures

Figure 1.

Figure 2.

***Asellus aquaticus* (Linnaeus, 1758)**

A.

B.

Figure 3.

Asellus aquaticus (Linnaeus, 1758)

Figure 4.

***Niphargus hrabei* S. Karaman, 1932**

Figure 6.

Niphargus hrabei S. Karaman, 1932

Figure 5.

***Asellus aquaticus* (Linnaeus, 1758)**

Figure 7.

***Niphargus hrabei* S. Karaman, 1932**

Figure 8.

Additional files

Supplementary_Materials.docx

Supplementary Materials

File includes additional tables with information regarding evolutionary models selected for phylogenetic analyses, the genetic loci sequence and primers employed, accession numbers for data deposited in GenBank, and Genealogical Sorting Index estimates for all populations of *Asellus aquaticus* and *Niphargus hrabei*.

CHAPTER IV

PHYLOGENETIC ANNOTATION AND GENOMIC ARCHITECTURE OF OPSIN GENES IN CRUSTACEA

Jorge Luis Pérez-Moreno, Danielle DeLeo, Ferran Palero, Heather D. Bracken-Grissom

This chapter was published in:

Hydrobiologia, 825(1): 159-175, 2018.

Abstract

A major goal of evolutionary biology is to understand the role of adaptive processes on sensory systems. Visual capabilities are strongly influenced by environmental and ecological conditions, and the evolutionary advantages of vision are manifest by its complexity and ubiquity throughout Metazoa. Crustaceans occupy a vast array of habitats and ecological niches, and are thus ideal taxa to investigate the evolution of visual systems. A comparative approach is taken here for efficient identification and classification of opsin genes, photoreceptive pigment proteins involved in color vision, focusing on two crustacean model organisms (*Hyalella azteca* and *Daphnia pulex*). Transcriptomes of both species were assembled *de novo* to elucidate the diversity and function of expressed opsins within a robust phylogenetic context. For this purpose, we developed a modified version of the Phylogenetically Informed Annotation tool to filter and identify visual genes from transcriptomes in a scalable and efficient manner. Additionally, reference genomes of these species were used to validate our pipeline while characterizing the genomic architecture of the opsin genes. Next-generation sequencing and phylogenetics provide future venues for the study of sensory systems, adaptation, and evolution in model and non-model organisms.

Keywords: evolution; phototransduction; protein; rnaseq; transcriptomics; vision

Introduction

Opsins are photoreceptor molecules that play a crucial role in animal vision and can be found across metazoans (Terakita, 2005). As membrane-associated, G-protein-coupled receptors (GPCRs), opsins can function in both visual and non-visual phototransduction, and in some instances as photoisomerases (Shichida & Matsuyama, 2009). Previous studies have classified opsins in three primary categories according to the type of G-protein to which they couple namely, “ciliary” (c-opsins), “rhabdomeric” (r-opsins), and RGR/Go opsins (Terakita, 2005; Feuda et al., 2014, 2016). Ciliary and rhabdomeric opsins diversified prior to the protostome-deuterostome split and are found in both invertebrates and vertebrates, which suggests that they co-occurred in a common ancestor (Kojima et al., 1997; Shichida & Matsuyama, 2009; Hering & Mayer, 2014; Ramirez et al., 2016). Opsin categories can additionally be further subdivided into subfamilies based on molecular phylogenetics and functional classifications (Terakita, 2005). These subfamilies share less than 20 percent amino acid identity (Fryxel & Meyerowitz, 1991) and comprise c-opsins (visual and non-visual); tmt/encephalopsins; Go- and Gs- coupled, (r-opsins) Gq-coupled/melanopsins; and photoisomerases/neuropsins. Photoreceptive opsins can be of either the ciliary-type, found largely in vertebrates (for exceptions see Arendt et al., 2004; Passamaneck et al., 2011; Bok et al., 2017; Tsukamoto et al., 2017), or the rhabdomeric-type found in mollusks, annelids, and the compound eyes of arthropods (Arendt et al., 2002; Shichida & Matsuyama, 2009; Gühmann et al., 2015) , with the last being the focus of the present study.

Opsins form visual pigments capable of absorbing photons when bound to a chromophore, generally a vitamin A1 derivative (11-cis retinal). These visual pigments trigger conformational changes that activate G-proteins (Nathans, 1987) and elicit phototransduction signaling cascades. Key biological processes such as the regulation of circadian clocks, phototaxis, and vision have been shown to be linked to the phototransduction cascade (e.g. Arendt et al., 2004; review Shichida & Matsuyama, 2009). The set of amino acid residues that interact with the chromophore produce an environment suitable for the absorption of light with distinct wavelengths (Imai et al., 1997; Kuwayama et al., 2002) and thus influence spectral tuning (e.g. Porter et al., 2007; Katti et al., 2010). As the absorption spectrum of the photopigment is influenced by the amino acid composition of the opsin protein, slight variations can alter its physical and chemical properties and lead to visual pigments maximally sensitive to different wavelengths of light. This in turn would allow organisms to perceive and distinguish between light of particular wavelengths. The direct association between amino acid composition of photoreceptive opsins and their spectral sensitivity make them amenable to functional classification by sequence analysis (Mirzadegan et al., 2003; Matsumoto & Ishibashi, 2016).

Three main approaches have been employed to characterize opsins from transcriptomic data: I) Sequence similarity searches via pairwise alignments (i.e. BLAST); II) Protein structure prediction through Hidden Markov Model (HMM) profile alignments; and III) Phylogenetic inference. Functional annotation by means of sequence

similarity is typically based on heuristic algorithms that search for matching nucleotide and/or amino acid sequences in curated databases (e.g. BLAST; Altschul et al., 1990). Sequences are locally or globally aligned and subsequently annotated based on inferred homology with statistically significant matches (Pearson, 2013). These comparisons, however, can rapidly become computationally expensive as the number of query and/or reference sequences increases (Suzuki et al., 2012). Although similarity searches via pairwise alignments are capable of identifying homologous sequences, their shortcomings are notorious when the queries consist of protein families with low sequence similarities, as it is the case for opsins and other GPCRs (Pearson, 2013). Hidden Markov Model (HMM) methods offer an enticing alternative to pairwise alignments at similar computational costs (Eddy, 2011; Pearson, 2013). HMM profiles can also contain relevant information regarding protein structure, which translates to more accurate identification, classification, and annotation of proteins even when overall sequence similarity is low (Krogh et al., 1994; Yoon, 2009; Pearson, 2013). However, the efficacy of HMMs is intrinsically dependent on the quality of the training data, which is a non-trivial process in the case of understudied taxa or protein families (Rasmussen & Krink, 2003; Pearson, 2013). Therefore, the use of HMMs for annotation of GPCRs is hindered when independently verified sequences are not readily available. The robustness and suitability of phylogenetic approaches for functional annotation of opsins (and other proteins) is unparalleled, as it can readily overcome many of the deficiencies of other homology-based methodologies (Engelhardt et al., 2009; Gaudet et al., 2011; Speiser et al., 2014). The placement of proteins on a phylogenetic tree not only enables a rapid assessment of homology and efficient discrimination of false-positives, but also allows

for the inference of putative functions and roles within an evolutionary context (Engelhardt et al., 2009). This approach has been successful in classifying novel opsins (and other GPCRs) despite their characteristic low sequence similarities and, in the case of non-model organisms, scarce genomic resources (Porter et al., 2007, 2012; Speiser et al., 2014). The main drawbacks of phylogenetic reconstruction as an efficient functional annotation method are possible difficulties aligning distantly related sequences, its propensity to be time-consuming (obtaining adequate references, computation of trees, etc.) and the steep learning curve to master these analyses, which might result in subjectivity and misinterpretations (Crisp & Cook, 2005).

Efforts to characterize opsins from high-throughput sequencing data in non-model Crustacea have primarily focused on transcriptomes, but without genomic validation (Porter et al., 2013; Wong et al., 2015; Biscontin et al., 2016). When available, genomes can provide valuable information regarding opsin gene duplication in an organism, as well as the relative locations of those genes. Gene locations allow for intra- and interspecific comparisons (Nordström et al., 2004) and to make inferences about the evolutionary history of opsin diversification (review Shichida & Matsuyama, 2009).

In this study we modified the Phylogenetically Informed Annotation tool's pipeline (Speiser et al., 2014) to conduct a robust and scalable phylogenetic annotation of visual opsins from transcriptomes of two crustacean model organisms, *Hyalella azteca* and *Daphnia pulex*. *Hyalella azteca* is a freshwater epibenthic amphipod, commonly used as a bioindicator species, which has one pair of pigmented compound eyes (Gonzalez & Watling, 2002). *Daphnia pulex* is a freshwater cladoceran that has a single but movable

cyclopean, compound, and pigmented eye. Specifically, we made modifications for PIA to run on the command-line rather than on Galaxy’s GUI and wrote wrapper scripts to facilitate the analyses. This resulted in a scalable and automated platform to annotate visual genes and pathways, while minimizing possible biases and subjectivity from manual curation. As genomes are available for both species, they were used to validate the annotations and make inferences about the genomic architecture and the opsin intron-exon gene structure within these species.

Methods

Data, Quality Control, and Transcriptome Assembly

Raw RNA sequencing data of the freshwater amphipod *Hyaella azteca* and the model branchiopod *Daphnia pulex* were downloaded from the NCBI’s Sequence Read Archive (SRA). In order to facilitate *de novo* transcriptome assembly and accurate detection of complete opsin isoforms, the read files were trimmed taking into consideration factors such as length and quality of the sequencing reads, sequencing depth, and tissue type (Table 1).

Prior to the assembly process, quality of the raw sequencing reads was evaluated via FastQC (Andrews, 2010). The FastQC output was subsequently used to inform stringent quality and adaptor trimming with Trimmomatic 0.36 (parameters: “ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 CROP:140 HEADCROP:20 LEADING:15 TRAILING:15 SLIDINGWINDOW:4:20 MINLEN:36”; Bolger et al., 2014). Clean

sequencing reads were then assembled into a *de novo* transcriptome with the Trinity pipeline (version 2.4.0; Grabherr et al., 2011; Haas et al., 2013) using default parameters, a minimum contig length of 200bp, and a *kmer* size of 23. Assembly summary statistics were calculated using Transrate 1.0.3 (Smith-Unna et al., 2016). BUSCO 3.0.2 (Benchmarking Universal Single-Copy Orthologs; (Simão et al., 2015) was employed to assess the quality and completeness of the resulting transcriptomes. The latter method provides an accurate evaluation of transcriptomes in an evolutionary informed context by assessing the presence and completeness of universal single copy orthologs (Simão et al., 2015). BUSCO analyses were conducted with the Arthropoda database of orthologous groups (n= 1066) sourced from OrthoDB (Waterhouse et al., 2013).

Identification and Annotation of Crustacean Opsins

Identification and functional classification of putative opsin transcripts was achieved through the use of our modified version of the existing Phylogenetically-Informed Annotation (PIA) tool (Speiser et al., 2014). While phylogenetic confirmation of BLAST similarity hits is becoming routine in model systems, PIA allows for the identification of proteins involved in visual pathways for non-model organisms in a computationally efficient manner (Speiser et al., 2014). This informative tool places putative visual gene transcripts (e.g. opsins), previously identified via BLAST searches against a custom database, in pre-computed phylogenies of such genes. The resulting phylogenies can then be used to discriminate BLAST false-positives and/or paralogous sequences from the transcripts of interest. While PIA has been used in previous studies to

annotate genes in a phylogenetic context, it was originally designed as a workflow for the Galaxy Project (Afgan et al., 2016) and as such is dependent on a Graphical User Interface (Speiser et al., 2014). This workflow can become inefficient when conducting concurrent analyses of numerous transcriptomes. Further, curation of the phylogenetic gene trees produced by PIA for each input transcriptome is typically undertaken manually, which inevitably makes it sensitive to potential biases. Tree branch length cutoff values for a given gene (i.e. opsins) can however be determined empirically through a series of manual tree curation comparisons. The pipeline presented here is a modification to PIA's source code in which the authors wrote a wrapper script to enable its use as a command-line/automated workflow, which effectively increases its scalability allowing for the identification of visual opsins from multiple transcriptomic datasets through simple scripting. Although the pipeline was designed for analyses of visual pathways, it is possible to create custom databases and phylogenies for customized analyses of other genes/pathways. We refer the reader to the original publication of PIA for additional information regarding included genes and pathways (Speiser et al., 2014). The modified Phylogenetically Informed Annotation pipeline employed in this study, along with usage examples, will be made available at: <http://xibalbanus.github.io>

Once the transcriptome assembly was completed, our *de novo* assemblies were scanned with Biopython's *get_orfs_cds.py* script (Cock et al., 2009) to translate each transcript into its corresponding amino acid sequence. Open Reading Frames (ORFs) were then extracted via the same script to facilitate the PIA annotation process. After conclusion of PIA's main component (BLAST, MAFFT alignment, and phylogenetic

placement via RAxML; Altschul et al., 1990; Stamatakis, 2014; Yamada et al., 2016) , a script adapted from the Osiris Phylogenetics toolkit (*long_branch_finder.py*; Oakley et al., 2014) was used to identify transcripts that exceeded 4x the Mean Absolute Deviation of the tree's branch lengths. This simple threshold proved effective at removing spurious BLAST hits in an unbiased manner. Subsequently, the previously identified false-positives were pruned from our phytab-formatted hit-list (part of PIA's output) with the *prune_phytab_using_list.py* script, also adapted from Osiris (Oakley et al., 2014). The resulting list of putative opsins was then converted to FASTA format and sequence redundancy was reduced by removing identical protein sequences with UCLUST (Edgar, 2010). The multiple sequence aligner MAFFT (Yamada et al., 2016) was then invoked to align our filtered putative opsins to a large opsin dataset (n= 910) compiled by the Porter Lab (University of Hawaii at Manoa), which includes representatives of the main opsin subfamilies. MAFFT alignment parameters were chosen to prioritize accuracy over speed and to allow for large unalignable regions that can be pervasive with divergent GPCRs ("*--ep 0 --genafpair --maxiterate 1000*"). Following the alignment procedure, a final phylogenetic reconstruction was undertaken with IQ-tree (Nguyen et al., 2015) for characterization and annotation of our PIA-identified putative opsins. IQ-tree compares favorably to alternatives (e.g. RAxML, FastTree, etc.) in recent benchmarks (Zhou et al., 2017), while also providing a more extensive choice of evolutionary models for phylogenetic inference. After proper consideration, IQ-tree was selected given that evolutionary model choice is important and its choice would be limited in alternative software. Choosing an appropriate model is especially relevant when inferring phylogenetic relationships in protein families with both highly conserved domains and

hypervariable regions (e.g. opsins). The IQ-tree analysis was run with a LG general amino acid replacement matrix under a FreeRate model with 10 rate categories and empirical base frequencies (LG+R10+F; Le & Gascuel, 2008; Soubrier et al., 2012) as suggested by ModelFinder (Kalyaanamoorthy et al., 2017). Branch support was assessed in tripartite by Ultra-fast bootstrap approximation (UFBoot; 10,000 replicates), a Shimodaira–Hasegawa–like approximate likelihood ratio test (SH-aLRT; 10,000 replicates), and an approximate Bayes test (Guindon et al., 2010; Anisimova et al., 2011; Minh et al., 2013).

Finally, the tool HHBlits ‘HMM-HMM–based lightning-fast iterative sequence search’ (Remmert et al., 2012) was used to confirm opsin identity based on profile HMMs using Uniclust30 (Mirdita et al., 2017) as the reference database. HHBlits was chosen as it incorporates highly sensitive sequence search methods (HMMs) in a fast and more accurate manner compared to other sequence search tools like PSI-BLAST (Remmert et al., 2012).

Genomic Architecture of Annotated Opsins

Proteins encoded in the transcriptomes analyzed may not have corresponding annotations in public databases. Therefore, to validate our pipeline, the exon-intron architecture of the opsin genes obtained from transcriptomic data (see above) was annotated *de novo* using the recently assembled genomes of *H. azteca* (GCA_000764305.2; accession date: 20-07-2017) and *D. pulex* (GCA_000187875.1;

accessed on 20-07-2017). The Benchmarking set of Universal Single-Copy Orthologs (BUSCO version 3; Simão et al., 2015) was used to ensure an adequate completeness of the genomes used for transcriptome/genome comparison. BUSCO provides quantitative measures for the assessment of genome assembly based on evolutionarily-informed expectations of gene content from near-universal single-copy orthologs selected from OrthoDB v9. The tblastn algorithm v2.2.29+ was then used with default parameters in order to discriminate between exonic and intronic regions along the genomic scaffolds. When a significant blast hit was found (similarity > 80%; e-value < 10^{-8}), the corresponding genomic region was annotated as exonic, or protein coding/expressed region. DNA regions located between two consecutive exons in the same genomic scaffold (chromosome) but with no corresponding counterpart in the expressed RNA were considered as introns. In addition, the nucleotide coding sequence of each putative opsin were mapped to their respective genomes using the spliced aligner HISAT2 (Kim et al., 2015). The mapping was done without penalties for non-canonical splicing using the following command and arguments: “*hisat2 -f -x index input.cds.fasta --score-min L,0,-4 --pen-noncansplice 3 -S output.sam*”. Plots of the gene architecture and the exon length distribution were subsequently completed using the Integrated Genome Browser (Freese et al., 2016) and the software package Mathematica v.11.1 (Wolfram Inc., USA).

Results

Hyalella azteca's transcriptome assembly recovered 243,398 contigs with a mean sequence length of 1033.04 base pairs (Table 2). Of these, 61,401 sequences contained

Open Reading Frames (ORFs) designating them as putative protein-coding genes. Similarly, our *de novo* transcriptome for *D. pulex* was comprised of 187,310 contigs with a mean sequence length of 848.76. Despite the higher number of assembled contigs and the lower mean sequence length, *D. pulex*'s transcriptome contained 38,157 sequences with ORFs. Additional metrics for our *de novo* transcriptomes, as well as for the reference genomes, are given in Table 2.

Completeness assessment of our *de novo* transcriptomes by Benchmarking Universal Single-Copy Orthologs (BUSCO) was favorable for both species. In *H. azteca* we were able to find 990 (92.48%) complete sequences of the 1066 arthropod genes used for benchmarking. An additional 49 (4.6%) were also present as fragmented sequences, and only 27 (2.6%) were not found. Similarly, *D. pulex*'s transcriptome was found to be nearly complete with 1048 (98.4%) full-length BUSCO genes, 16 (1.5%) fragmented, and a marginal 2 (0.1%) missing. The reference genomes selected for validation were rather complete as well, with over 90% of the BUSCO genes being found complete. Interestingly, the proportion of missing BUSCOs was slightly higher for the genomic data compared to the transcriptomic data (Table 3).

Our custom version of the Phylogenetically Informed Annotation tool outputs a single FASTA file of amino acid sequences per transcriptome. This file contains the transcripts that remain after the removal of spurious BLAST hits and the merging/removal of duplicated and fragmented sequences, and should only contain those that are closely related to the gene of interest (i.e. opsins). This output can then be piped

to a final step for functional annotation by phylogenetic inference. In our case, putative opsin sequences for both species were aligned to a curated dataset of different opsin types. This final step resulted in a large phylogeny (Fig. 1) where opsins are classified based on their phylogenetic position.

Following Trinity's definition of assembled genes/isoforms, *Hyalella azteca*'s transcriptome contained 1 SWS/UV opsin (2 isoforms), 3 LWS opsins (3 isoforms), and 1 opsin-like GPCRs (Fig. 2; Table 4) that were identified by our pipeline. On the other hand, *Daphnia pulex* transcriptome 2 different SWS opsins (4 isoforms), 6 LWS opsins (35 isoforms), 2 melanopsins (4 isoforms), and 1 opsin-like transcript that was placed within the outgroup clade (Fig. 2; Table 4).

Identity results of the HHBlits search using iterative pairwise alignments and profile HHMs are summarized in Table 5, along with the inferred classification of each putative opsin transcript based on their respective placement in the phylogeny (Figs. 12). Additionally, each sequence entry was annotated as visual or non-visual based on the sequence homology inferred by both methods (represented by a black box when both methods are in agreement; and a gray box when HHMs fail to identify them as a visual opsin; Table 5).

Genomic structure of annotated opsins

Every protein sequence predicted using the modified PIA pipeline gave at least one significant TBLASTN hit both in the *D. pulex* and *H. azteca* reference genomes. The observed distribution of introns within opsin genes appears to be variable both within and between genomes. To illustrate this variation, the Intron-Exon gene structure patterns of representative opsins were further characterized. The genomic region encoding for the SWS/UV opsin gene spanned about 4 kb in *Hyalella* and presented an extremely partitioned structure formed by at least 7 different exons (Figure 3). Interestingly, some LWS opsins within the *H. azteca* genome were located on the antisense strand and appear to be duplicated retrogenes (Figure 3; see Discussion). Both SWS/UV and LWS opsins were also arranged following disparate gene architectures in the *D. pulex* genome (Figure 4). LWS opsins presented slightly shorter introns on average than SWS/UV opsins, but the presence of gene duplications and genes with numerous introns were identified in most cases. Exon size distribution had a similar shape in *D. pulex* and *H. azteca*, being multimodal for both genomes (Figure 5). Nevertheless, *H. azteca* had a larger average exon size (Mean = 410 bp; Median = 235 bp) than *D. pulex* (Mean = 225 bp; Median = 164).

Discussion

Our results demonstrate the power of incorporating phylogenetic annotation towards the characterization and interpretation of large transcriptomic datasets of non-

model organisms. Annotations via simple sequence similarity based methods like BLAST alone can result in false positives including, but not limited to, functional diversification following gene duplication events, domain shuffling, or even existing database errors (review Sjölander, 2004) . Using the modified version of PIA has allowed for the rapid and automated identification of false positives among the putative visual opsins curated for the two species of crustaceans, *Hyalella azteca* and *Daphnia pulex*. The modified Phylogenetically Informed Annotation pipeline was able to successfully identify and filter opsins from the *de novo* transcriptomes in a fully automated manner with minimal manual curation. This automation is made possible mainly by the modifications and wrapper scripts that converted PIA from a Galaxy workflow to a command-line one, which effectively increases the scalability of the pipeline allowing for the identification of opsins (and other genes) from multiple transcriptomic datasets through simple scripting. Theoretically, this would allow for the annotation of dozens if not hundreds of transcriptomes at a time without the need of the excessive time-costs that a graphical user interface and manual curation of hundreds of phylogenetic trees would imply. Further improvements are certainly possible, particularly in terms of parallelization for its use in High Performance Computing environments for even greater computing speeds. Nevertheless, the current pipeline is dependent on its individual components and would thus require those to be made compatible with parallelization.

The initial hits recovered by a BLAST search using the original PIA opsin dataset recovered 11 putative opsin isoforms for *H. azteca* and 51 for *D. pulex*. Our pipeline removed 54.5% and 23.5% of those as spurious hits (Table 4) based on the chosen branch

length thresholds. These thresholds can easily be adjusted for increased/decreased conservativeness if deemed necessary, which should be assessed on a gene-to-gene basis. The final phylogenetic inference took a step further by classifying these opsins in statistically supported functional clades (Figs. 1-2), which allowed for the determination of their putative photoreceptive roles. Both *H. azteca* and *D. pulex* transcriptomes were generated from whole organism RNA extractions. As opsins are known to function in various cells and tissues of arthropods, as well as the retina (e.g. Lampel et al., 2005; Tong et al., 2009), it is likely that the opsin groups identified here are expressed across several tissue types. Non-visual opsins can be readily identified via phylogenetic inference provided that appropriate reference sequences are included in the multiple sequence alignments (for example, Fig. 3). HMM alignments were also used to confirm protein identities as well as to compare with the results of the phylogenetic annotation. HHMs were able to pair most putative visual opsins to the lateral compound eye opsins of arthropods for both species and, in the case of *D. pulex*, specifically to *Daphnia* class A rhodopsins. While there were a few discrepancies among annotation methods with regard to visual opsins (r-opsins) and melanopsins, this could be explained by their common origin (Porter et al., 2012). Melanopsins are very similar to the r-opsins found in invertebrates (Provencio et al., 1998, 2000) and can couple to similar signaling cascades (Isoldi et al., 2005; Panda et al., 2005; Qiu et al., 2005), likely contributing to the observed contradictory results depending on the methods. In fact, the similarities between these opsin types are evident in our phylogenetic trees (Figs. 1 & 2), showing a well-supported clade of arthropod UV opsins nested within the melanopsin clade. Partial sequences or existing database errors could also be a contributing factor to which BLAST

and HMM approaches are more sensitive. Even though the HMM searches were not able to determine the functional classification of the opsins in terms of spectral sensitivity, they were confirmed as visual opsins (Table 5). Our results further support the notion that integrated annotation methods are advantageous and recommended to confirm the robustness of findings and annotations.

Opsin repertoire and spectral sensitivities

There are several subgroups of rhabdomeric visual opsins responsible for vision in crustaceans, each with characteristic absorption spectra when bound to a chromophore (Kashiyama et al., 2009; Henze & Oakley, 2015). The number and type of opsins found throughout Crustacea can range greatly, partially owing to differences in methodologies—with no present homologs found in freshwater *Bathynellacea* (Kim et al., 2017), one or two SWS visual opsins found in species of deep-sea shrimp (Wong et al., 2015), and brachyuran crabs (Sakamoto et al., 1996), and as many as 33 identified in stomatopods (Porter et al., 2009, 2013). The number of opsins and corresponding spectral sensitivity of an organism appears to correlate with its life-history, habitat, and the ecological niche it may occupy (Marshall et al., 2015; Stieb et al., 2017). This study represents the first transcriptomic exploration of *H. azteca*'s opsin repertoire, which revealed several putative visual opsins (Fig. 2; Table 4). *Hyalella azteca* is a freshwater epibenthic amphipod commonly used as a bioindicator species. Though further evidence is required to make inferences regarding the expression and functionality of these putative opsins, the ability to differentiate between short and long wavelengths would allow *H. azteca* to

discern between direct and reflected light from the benthos. Direct sunlight (or moonlight) tends to be abundant in short-wavelengths (<450nm) whereas reflected light from sources like leaves and sediment tends to be shifted towards longer (>450nm) wavelengths (Menzel, 1979). Our analyses revealed four distinct opsin genes (one SWS/UV and three LWS) expressed in its transcriptome, suggest that *H. azteca* may be capable of discriminating between the aforementioned light sources. The authors hypothesize that if *H. azteca* does possess functional SWS and LWS visual opsins, this distinction could serve as an important environmental cue influencing their response to a variety of abiotic and biotic factors (e.g. refugia, vegetation, predators, prey). However, the authors note additional studies incorporating electroretinographic analyses are needed to confirm whether *H. azteca* can indeed discriminate between different wavelengths of light as the transcriptomic data suggests. Fewer opsins were found to be expressed in *H. azteca* compared to *D. pulex*, which is not surprising given the expansive opsin repertoire previously described for *Daphnia* (Colbourne et al., 2011; Brandon et al., 2017). *Daphnia* has both a simple and compound eyes, which may contribute to the relatively large number of opsin isoforms expressed. Differences have been found in the number and type of opsin genes expressed among eye forms within the ostracod *Skogsberia leneri* (Oakley & Huber, 2004) and similarly hypothesized for *Daphnia* (Brandon et al., 2017).

The subset of identified rhabdomeric opsins expressed in the *D. pulex* transcriptome allows for comparisons to prior studies characterizing the range of opsin types found in the *D. pulex* genome (Colbourne et al., 2011; Brandon et al., 2015). Colbourne et al. (2011) reported 25 medium- (MWS) and long-wavelength sensitive

(LWS) opsin genes as present in the *D. pulex* genome, but only 3 LWS opsin genes (and 24 isoforms) were identified in our analyses. While it is possible that the additional opsin classes identified in previous genomic investigations were not expressed in the current *D. pulex* dataset, it is also possible these discrepancies are due to differences in classification schemes across Arthropoda, with ‘blue-green’ wavelengths currently grouped under SWS. An alternative explanation is that separate genes are being considered isoforms of each other by Trinity during assembly process. Considering the large number of “isoforms” and low number of “genes” identified in the *D. pulex* transcriptome, in contrast with previous genomic investigations (i.e. Brandon et al., 2017), this is likely a contributing factor to this observed discrepancy.

Genomic architecture and opsin gene duplications

Gene duplications play a fundamental role in genome evolution (Ohno, 1970; Kondrashov et al., 2002), with replicates occasionally evolving new biological functions (Zhang, 2003; Pegueroles et al., 2013). Some of these genome duplications are considered as pseudogenes, loci whose nucleotide sequences are similar to a normal gene but that do not produce a functional product when translated. The “unprocessed” pseudogenes, have all the normal parts of a protein-coding gene, but generally are non-functional due to coding errors (Lynch & Force, 1999). The so-called “processed” pseudogenes or retrogenes lack the non-coding introns present in the original gene, and are thought to arise from mRNA reinserted into the genome by reverse transcription (Betrán & Long, 2002). Some retrogenes have been found to be actively transcribed and

the RNA product can be further processed to give two different molecules of RNA of smaller size that form elaborate secondary structures. These RNA regulatory molecules can control a variety of key genes involved in the regulation of the cell cycle and in cell growth (Tutar, 2012; Wen et al., 2012). Intron-less opsin genes, such as the LWS opsins our analyses identified in *H. azteca* (Fig. 3), have evolved in various metazoans (including crustaceans) and are thought to be functional photoproteins (Morris et al., 1993; Fitzgibbon et al., 1995; Porter et al., 2007; Liegertová et al., 2015), though it has been postulated that the expression of retrotransposed opsins is a form of transcriptional noise and a byproduct of transcriptional activity in the new genomic region (Xu et al., 2016). Opsin diversification and photopigment evolution seems to have been driven by duplicated opsin genes (e.g. Frentiu et al., 2007; Briscoe et al., 2010), as is the case of both ocular and extraocular cnidarian photoreceptors (Liegertová et al., 2015). Likewise, a functional LWS retrogene was recently found in the arthropod *Helicoverpa armigera*, though expression was believed to be under temporal compartmentalization and primarily expressed in larval stages (Xu et al., 2016). Our results provide further evidence supporting the importance of retrogenes in the evolution of the opsins gene family.

Concluding Remarks

Our results support the use of integrative phylogenetic annotation rather than just similarity-based approaches. This is an often overlooked but especially important consideration for the study of protein families (e.g. GPCRs) known for having large numbers of isoforms, multiple duplication events, low sequence similarities, and various

combinations of highly conserved domains with hypervariable regions. Phylogenetic approaches are not only able to robustly evaluate homology in an evolutionary context, but they can also provide valuable functional information based on recovered branching patterns. In the case of opsins, this functional information can be insightful from a variety of perspectives, and aid in the formulation and testing of organismal, ecological, and evolutionary hypotheses. Many of these will be put to the test in the present and forthcoming genomic era, for which efficient and scalable methodologies and pipelines will be paramount.

Acknowledgments

The authors would like to thank Megan Porter for access to her compilation of reference opsin data, as well as Daniel Speiser and Todd Oakley for allowing us to modify the original PIA tool. JPM was supported by the Philip M. Smith Graduate Research Grant for Cave and Karst Research from the Cave Research Foundation and The Crustacean Society Scholarship in Graduate Studies. This work was partially funded by two grants awarded from the National Science Foundation: Doctoral Dissertation Improvement Grant (#1701835) awarded to JPM and HBG and Division of Environmental Biology Bioluminescence and Vision grant (DEB-1556059) awarded to HBG. FP acknowledges project CHALLENGEN (CTM2013-48163) of the Spanish Government and a post-doctoral contract funded by the Beatriu de Pinós Programme of the Generalitat de Catalunya (2014-BPB-00038). The authors would like to acknowledge the Instructional & Research Computing Center (IRCC) at Florida International University for providing

High Performance Computing resources that have contributed to the research results reported within this article.

References

- Afgan, E., D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Grüning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, & J. Goecks, 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research* 44: W3–W10.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, & D. J. Lipman, 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
- Andrews, S., 2010. FastQC A Quality Control tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Anisimova, M., M. Gil, J.-F. Dufayard, C. Dessimoz, & O. Gascuel, 2011. Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-based Approximation Schemes. *Systematic Biology* 60: 685–699.
- Arendt, D., K. Tessmar, M.-I. M. de Campos-Baptista, A. Dorresteyn, & J. Wittbrodt, 2002. Development of pigment-cup eyes in the polychaete *Platynereis dumerilii* and evolutionary conservation of larval eyes in Bilateria. *Development* 129: 1143–1154.
- Arendt, D., K. Tessmar-Raible, H. Snyman, A. W. Dorresteyn, & J. Wittbrodt, 2004. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science* 306: 869–871.
- Betrán, E., & M. Long, 2002. Expansion of genome coding regions by acquisition of new genes. *Genetica* 115: 65–80.
- Biscontin, A., E. Frigato, G. Sales, G. M. Mazzotta, M. Teschke, C. De Pittà, S. Jarman, B. Meyer, R. Costa, & C. Bertolucci, 2016. The opsin repertoire of the Antarctic krill *Euphausia superba*. *Marine Genomics* 29: 61–68.
- Bok, M. J., M. L. Porter, & D.-E. Nilsson, 2017. Phototransduction in fan worm radiolar eyes. *Current Biology* 27: R681–R701.
- Bolger, A. M., M. Lohse, & B. Usadel, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

- Brandon, C. S., M. J. Greenwold, & J. L. Dudycha, 2017. Ancient and Recent Duplications Support Functional Diversity of *Daphnia* Opsins. *Journal of Molecular Evolution* 84: 12–28.
- Briscoe, A. D., S. M. Bybee, G. D. Bernard, F. Yuan, M. P. Sison-Mangus, R. D. Reed, A. D. Warren, J. Llorente-Bousquets, & C.-C. Chiao, 2010. Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. *Proceedings of the National Academy of Sciences* 107: 3628–3633.
- Cock, P. J. A., T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, & M. J. L. de Hoon, 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25: 1422–1423.
- Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker, T. H. Oakley, S. Tokishita, A. Aerts, G. J. Arnold, M. K. Basu, D. J. Bauer, C. E. Caceres, L. Carmel, C. Casola, J.-H. Choi, J. C. Detter, Q. Dong, S. Dusheyko, B. D. Eads, T. Frohlich, K. A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E. V. Kriventseva, D. Kultz, C. Laforsch, E. Lindquist, J. Lopez, J. R. Manak, J. Muller, J. Pangilinan, R. P. Patwardhan, S. Pitluck, E. J. Pritham, A. Rechtsteiner, M. Rho, I. B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzy, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y. I. Wolf, H. Yamagata, T. Yamada, Y. Ye, J. R. Shaw, J. Andrews, T. J. Crease, H. Tang, S. M. Lucas, H. M. Robertson, P. Bork, E. V. Koonin, E. M. Zdobnov, I. V. Grigoriev, M. Lynch, & J. L. Boore, 2011. The ecoresponsive genome of *Daphnia pulex*. *Science* 331: 555–561.
- Crisp, M., & L. Cook, 2005. Do early branching lineages signify ancestral traits? *Trends in Ecology & Evolution* 20: 122–128.
- Eddy, S. R., 2011. Accelerated Profile HMM Searches. *PLoS computational biology* 7: e1002195–e1002195.
- Edgar, R. C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460–2461.
- Engelhardt, B. E., M. I. Jordan, S. T. Repo, & S. E. Brenner, 2009. Phylogenetic molecular function annotation. *Journal of Physics: Conference Series* 180: 012024.
- Feuda, R., O. Rota-Stabelli, T. H. Oakley, D. Pisani, 2014. The Comb Jelly Opsins and the Origins of Animal Phototransduction. *Genome Biology and Evolution* 6: 1964–1971.
- Feuda, R., F. Marlétaz, M. A. Bentley, P. W.H. Holland, 2016. Conservation, Duplication, and Divergence of Five Opsin Genes in Insect Evolution. *Genome Biology and Evolution* 8:579–587.

- Fitzgibbon, J., A. Hope, S. J. Slobodyanyuk, J. Bellingham, J. K. Bowmaker, & D. M. Hunt, 1995. The rhodopsin-encoding gene of bony fish lacks introns. *Gene* 164: 273–277.
- Freese, N. H., D. C. Norris, & A. E. Loraine, 2016. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32: 2089–2095.
- Frentiu, F. D., G. D. Bernard, M. P. Sison-Mangus, A. Van Zandt Brower, & A. D. Briscoe, 2007. Gene Duplication Is an Evolutionary Mechanism for Expanding Spectral Diversity in the Long-Wavelength Photopigments of Butterflies. *Molecular Biology and Evolution* 24: 2016–2028.
- Fryxel, K. J., & E. M. Meyerowitz, 1991. The evolution of rhodopsins and neurotransmitter receptors. *Journal of Molecular Evolution* 33: 367–378.
- Gaudet, P., M. S. Livstone, S. E. Lewis, & P. D. Thomas, 2011. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Briefings in Bioinformatics* 12: 449–462.
- Gonzalez, E. R., L. Watling, 2002. Redescription of *Hyaella azteca* from Its type locality, Vera Cruz, Mexico (Amphipoda:Hyalellidae). *Journal of Crustacean Biology* 22: 173–183.
- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, & A. Regev, 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
- Gühmann, M., H. Jia, N. Randel, C. Verasztó, L. A. Bezares-Calderón, N. K. Michiels, S. Yokoyama, & G. Jékely, 2015. Spectral Tuning of Phototaxis by a Go-Opsin in the Rhabdomeric Eyes of *Platynereis*. *Current Biology* 25: 2265–2271.
- Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, & O. Gascuel, 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* 59: 307–321.
- Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. Williams, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, & A. Regev, 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494–1512.
- Henze, M. J., & T. H. Oakley, 2015. The Dynamic Evolutionary History of Pancrustacean Eyes and Opsins. *Integrative and Comparative Biology* 55: 830–842.

- Hering, L., & G. Mayer, 2014. Analysis of the Opsin Repertoire in the Tardigrade *Hypsibius dujardini* Provides Insights into the Evolution of Opsin Genes in Panarthropoda. *Genome Biology and Evolution* 6: 2380–2391.
- Imai, H., D. Kojima, T. Oura, S. Tachibanaki, A. Terakita, & Y. Shichida, 1997. Single amino acid residue as a functional determinant of rod and cone visual pigments. *Proceedings of the National Academy of Sciences* 94: 2322–2326.
- Isoldi, M. C., M. D. Rollag, A. M. de Lauro Castrucci, & I. Provencio, 2005. Rhabdomeric phototransduction initiated by the vertebrate photopigment melanopsin. *Proceedings of the National Academy of Sciences* 102: 1217–1221.
- Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, & L. S. Jermini, 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.
- Kashiyama, K., T. Seki, H. Numata, & S. G. Goto, 2009. Molecular Characterization of Visual Pigments in Branchiopoda and the Evolution of Opsins in Arthropoda. *Molecular Biology and Evolution* 26: 299–311.
- Katti, C., K. Kempler, M. L. Porter, A. Legg, R. Gonzalez, E. Garcia-Rivera, D. Dugger, & B.-A. Battelle, 2010. Opsin co-expression in *Limulus* photoreceptors: differential regulation by light and a circadian clock. *Journal of Experimental Biology* 213: 2589–2601.
- Kim, D., B. Langmead, & S. L. Salzberg, 2010. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12: 357–360.
- Kim, B.-M., S. Kang, D.-H. Ahn, J.-H. Kim, I. Ahn, C.-W. Lee, J.-L. Cho, G.-S. Min, & H. Park, 2017. First insights into the subterranean crustacean *Bathynellacea* transcriptome: transcriptionally reduced opsin repertoire and evidence of conserved homeostasis regulatory mechanisms. *PloS One* 12: e0170424.
- Kojima, D., A. Terakita, T. Ishikawa, Y. Tsukahara, A. Maeda, & Y. Shichida, 1997. A novel Go-mediated phototransduction cascade in scallop visual cells. *Journal of Biological Chemistry* 272: 22979–22982.
- Kondrashov, F. A., I. B. Rogozin, Y. I. Wolf, & E. V. Koonin, 2002. Selection in the evolution of gene duplications. *Genome Biology* 3: research0008–1.
- Krogh, A., M. Brown, I. S. Mian, K. Sjolander, & D. Haussler, 1994. Hidden Markov Models in Computational Biology. *Molecular Biology* 235: 1501–1531.
- Kuwayama, S., H. Imai, T. Hirano, A. Terakita, & Y. Shichida, 2002. Conserved Proline Residue at Position 189 in Cone Visual Pigments as a Determinant of Molecular Properties Different from Rhodopsins. *Biochemistry* 41: 15245–15252.

- Lampel, J., A. D. Briscoe, & L. T. Wasserthal, 2005. Expression of UV-, blue-, long-wavelength-sensitive opsins and melatonin in extraretinal photoreceptors of the optic lobes of hawkmoths. *Cell and Tissue Research* 321: 443–458.
- Le, S. Q., & O. Gascuel, 2008. An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution* 25: 1307–1320.
- Liebertová, M., J. Pergner, I. Kozmíková, P. Fabian, A. R. Pombinho, H. Strnad, J. Pačes, Č. Vlček, P. Bartůňek, & Z. Kozmik, 2015. Cubozoan genome illuminates functional diversification of opsins and photoreceptor evolution. *Scientific Reports* 5: 11885.
- Lynch, M., & A. Force, 1999. The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics* 154: 459–473.
- Marshall, J., K. L. Carleton, & T. Cronin, 2015. Colour vision in marine organisms. *Current Opinion in Neurobiology* 34: 86–94.
- Matsumoto, T., & Y. Ishibashi, 2016. Sequence analysis and expression patterns of opsin genes in the longtooth grouper *Epinephelus bruneus*. *Fisheries Science* 82: 17–27.
- Menzel, R., 1979. Spectral Sensitivity and Color Vision in Invertebrates In Autrum, H. (ed), *Comparative Physiology and Evolution of Vision in Invertebrates*. Springer Berlin Heidelberg, Berlin, Heidelberg: 503–580.
- Minh, B. Q., M. A. T. Nguyen, & A. von Haeseler, 2013. Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution* 30: 1188–1195.
- Mirdita, M., L. von den Driesch, C. Galiez, M. J. Martin, J. Söding, & M. Steinegger, 2017. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Research* 45: D170–D176.
- Mirzadegan, T., G. Benkő, S. Filipek, & K. Palczewski, 2003. Sequence Analyses of G-Protein-Coupled Receptors: Similarities to Rhodopsin. *Biochemistry* 42: 2759–2767.
- Morris, A., J. K. Bowmaker, & D. M. Hunt, 1993. The molecular basis of a spectral shift in the rhodopsins of two species of squid from different photic environments. *Proceedings of the Royal Society B: Biological Sciences* 254: 233–240.
- Nathans, J., 1987. Molecular biology of visual pigments. *Annual review of neuroscience* 10: 163–194.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, & B. Q. Minh, 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 32: 268–274.

- Nordström, K., T. A. Larsson, & D. Larhammar, 2004. Extensive duplications of phototransduction genes in early vertebrate evolution correlate with block (chromosome) duplications. *Genomics* 83: 852–872.
- Oakley, T. H., M. A. Alexandrou, R. Ngo, M. S. Pankey, C. K. Churchill, W. Chen, & K. B. Lopker, 2014. Osiris: accessible and reproducible phylogenetic and phylogenomic analyses within the Galaxy workflow management system. *BMC Bioinformatics* 15: 230.
- Oakley, T. H., & D. R. Huber, 2004. Differential Expression of Duplicated Opsin Genes in Two EyeTypes of Ostracod Crustaceans. *Journal of Molecular Evolution* 59: 239–249.
- Ohno, S., 1970. *Evolution by gene duplication*. George Allen and Unwin, London.
- Panda, S., S. K. Nayak, B. Campo, J. R. Walker, J. B. Hogenesch, & T. Jegla, 2005. Illumination of the Melanopsin Signaling Pathway. *Science* 307: 600–604.
- Passamaneck, Y. J., N. Furchheim, A. Hejnl, M. Q. Martindale, & C. Lüter, 2011. Ciliary photoreceptors in the cerebral eyes of a protostome larva. *EvoDevo* 2: 6.
- Pearson, W. R., 2013. An Introduction to Sequence Similarity (“Homology”) Searching In Baxevanis, A. D., G. A. Petsko, L. D. Stein, & G. D. Stormo (eds), *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc., Hoboken, NJ, USA.
- Pegueroles, C., S. Laurie, & M. M. Albà, 2013. Accelerated Evolution after Gene Duplication: A Time-Dependent Process Affecting Just One Copy. *Molecular Biology and Evolution* 30: 1830–1842.
- Porter, M. L., J. R. Blasic, M. J. Bok, E. G. Cameron, T. Pringle, T. W. Cronin, & P. R. Robinson, 2012. Shedding new light on opsin evolution. *Proceedings of the Royal Society B: Biological Sciences* 279: 3–14.
- Porter, M. L., M. J. Bok, P. R. Robinson, & T. W. Cronin, 2009. Molecular diversity of visual pigments in Stomatopoda (Crustacea). *Visual Neuroscience* 26: 255–265.
- Porter, M. L., T. W. Cronin, D. A. McClellan, & K. A. Crandall, 2007. Molecular Characterization of Crustacean Visual Pigments and the Evolution of Pancrustacean Opsins. *Molecular Biology and Evolution* 24: 253–268.
- Porter, M. L., D. I. Speiser, A. K. Zaharoff, R. L. Caldwell, T. W. Cronin, & T. H. Oakley, 2013. The Evolution of Complexity in the Visual Systems of Stomatopods: Insights from Transcriptomics. *Integrative and Comparative Biology* 53: 39–49.
- Provencio, I., G. Jiang, W. J. De Grip, W. PÄR HAYES, & M. D. Rollag, 1998. Melanopsin: An opsin in melanophores, brain, and eye. *Proceedings of the National Academy of Sciences of the United States of America* 95: 340–345.

- Provencio, I., I. R. Rodriguez, G. Jiang, W. P. Hayes, E. F. Moreira, & M. D. Rollag, 2000. A novel human opsin in the inner retina. *Journal of Neuroscience* 20: 600–605.
- Qiu, X., T. Kumbalasiri, S. M. Carlson, K. Y. Wong, V. Krishna, I. Provencio, & D. M. Berson, 2005. Induction of photosensitivity by heterologous expression of melanopsin. *Nature* 433: 745–749.
- Ramirez, M. D, A. N. Pairett, M. S. Pankey, J. M. Serb, D. I. Speiser, A. J. Swafford, T. H. Oakley, 2016. The Last Common Ancestor of Most Bilaterian Animals Possessed at Least Nine Opsins. *Genome Biology and Evolution* 8: 3640–3652.
- Rasmussen, T. K., & T. Krink, 2003. Improved Hidden Markov Model training for multiple sequence alignment by a particle swarm optimization—evolutionary algorithm hybrid. *Biosystems* 72: 5–17.
- Remmert, M., A. Biegert, A. Hauser, & J. Söding, 2012. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods* 9: 173–175.
- Sakamoto, K., O. Hisatomi, F. Tokunaga, & E. Eguchi, 1996. Two opsins from the compound eye of the crab *Hemigrapsus sanguineus*. *Journal of Experimental Biology* 199: 441–450.
- Shichida, Y., & T. Matsuyama, 2009. Evolution of opsins and phototransduction. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 2881–2895.
- Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, & E. M. Zdobnov, 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.
- Sjölander, K., 2004. Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics* 20: 170–179.
- Smith-Unna, R., C. Boursnell, R. Patro, J. M. Hibberd, & S. Kelly, 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Research* 26: 1134–1144.
- Soubrier, J., M. Steel, M. S. Y. Lee, C. Der Sarkissian, S. Guindon, S. Y. W. Ho, & A. Cooper, 2012. The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates. *Molecular Biology and Evolution* 29: 3345–3358.
- Speiser, D. I., M. Pankey, A. K. Zaharoff, B. a Battelle, H. D. Bracken-Grissom, J. W. Breinholt, S. M. Bybee, T. W. Cronin, A. Garm, A. R. Lindgren, N. H. Patel, M. L. Porter, M. E. Protas, A. S. Rivera, J. M. Serb, K. S. Zigler, K. a Crandall, & T. H. Oakley, 2014. Using phylogenetically-informed annotation (PIA) to search for

- light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinformatics* 15: 350–350.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Stieb, S. M., F. Cortesi, L. Sueess, K. L. Carleton, W. Salzburger, & N. J. Marshall, 2017. Why UV vision and red vision are important for damselfish (Pomacentridae): structural and expression variation in opsin genes. *Molecular Ecology* 26: 1323–1342.
- Suzuki, S., T. Ishida, K. Kurokawa, & Y. Akiyama, 2012. GHOSTM: A GPU-Accelerated Homology Search Tool for Metagenomics. *PLoS ONE* 7: e36060.
- Terakita, A., 2005. The opsins. *Genome biology* 6: 213.
- Tong, D., N. S. Rozas, T. H. Oakley, J. Mitchell, N. J. Colley, & M. J. McFall-Ngai, 2009. Evidence for light perception in a bioluminescent organ. *Proceedings of the National Academy of Sciences* 106: 9836–9841.
- Tsukamoto, H., I.-S. Chen, Y. Kubo, & Y. Furutani, 2017. A ciliary opsin in the brain of a marine annelid zooplankton is ultraviolet-sensitive, and the sensitivity is tuned by a single amino acid residue. *Journal of Biological Chemistry* 292: 12971–12980.
- Tutar, Y., 2012. Pseudogenes. *Comparative and Functional Genomics* 2012: 1–4.
- Waterhouse, R. M., F. Tegenfeldt, J. Li, E. M. Zdobnov, & E. V. Kriventseva, 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Research* 41: D358–D365.
- Wen, Y.-Z., L.-L. Zheng, L.-H. Qu, F. J. Ayala, & Z.-R. Lun, 2012. Pseudogenes are not pseudo any more. *RNA Biology* 9: 27–32.
- Wong, J. M., J. L. Pérez-Moreno, T.-Y. Chan, T. M. Frank, & H. D. Bracken-Grissom, 2015. Phylogenetic and transcriptomic analyses reveal the evolution of bioluminescence and light detection in marine deep-sea shrimps of the family Oplophoridae (Crustacea: Decapoda). *Molecular Phylogenetics and Evolution* 83: 278–292.
- Xu, P., R. Feuda, B. Lu, H. Xiao, R. I. Graham, & K. Wu, 2016. Functional opsin retrogene in nocturnal moth. *Mobile DNA* 7: 18.
- Yamada, K. D., K. Tomii, & K. Katoh, 2016. Application of the MAFFT sequence alignment program to large data—reexamination of the usefulness of chained guide trees. *Bioinformatics* 32: 3246–3251.

- Yoon, B.-J., 2009. Hidden Markov models and their applications in biological sequence analysis. *Current genomics* 10: 402–415.
- Zhang, J., 2003. Evolution by gene duplication: an update. *Trends in Ecology & Evolution* 18: 292–298.
- Zhou, X., X.-X. Shen, C. T. Hittinger, & A. Rokas, 2017. Evaluating Fast Maximum Likelihood-Based Phylogenetic Programs Using Empirical Phylogenomic Data Sets. *bioRxiv* 142323.

Tables

Table 1. Raw data chosen for *de novo* transcriptome assembly and annotation of opsin proteins in *Hyalomma azteca* and *Daphnia pulex*.

<i>Species</i>	<i>Megabytes</i>	<i>Megabases</i>	<i>Read lengths</i>	<i>Sequencing Platform</i>	<i>Tissue type</i>	<i>SRA BioProject</i>
<i>Hyalomma azteca</i>	15543	33160	2x150 bp	Illumina HiSeq	Whole	PRJNA312414
<i>Daphnia pulex</i>	16134	39280			organism	PRJNA380400

Table 2. Summary statistics for the *de novo* transcriptome assemblies produced as part of this study and the corresponding genome assemblies.

Metric	<i>Hyalomma azteca</i>		<i>Daphnia pulex</i>	
	Genome	Transcriptome	Genome	Transcriptome
Number of sequences/contigs	23426	243398	18989	187310
Longest sequence/contig (bp)	2207822	16780	528830	27096
Number of bases	550886000	251440760	197206000	158981525
Mean transcript/contig length (bp)	23404	1033.04	8352	848.76
Number of transcripts/contigs > 1,000 bp long	14563	73869	16743	42717
Number of transcripts/contigs > 10,000 bp long	7614	157	2854	179
Number of transcripts with		61401		38157

ORFs				
Mean ORF percent		45.73		50.22
N50	114415	1929	49250	1404
N30		3213		2588
N10		5560		5122
GC content	0.38	0.42	0.40	0.39

Table 3. Results of transcriptome completeness assessment by Benchmarking Universal Single-Copy Orthologs (BUSCO) using OrthoDB's Arthropoda database of orthologous genes.

Species	Dataset	Complete BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCOs Searched
<i>Hyalella azteca</i>	Genome	970 (91.0%)	29 (2.7%)	67 (6.3%)	1066
	Transcriptome	990 (92.8%)	49 (4.6%)	27 (2.6%)	
<i>Daphnia pulex</i>	Genome	1038 (97.3%)	9 (0.8%)	19 (1.9%)	
	Transcriptome	1048 (98.4%)	16 (1.5%)	2 (0.1%)	

Table 4. Number of genes and respective isoforms, as defined by Trinity, recovered for each type of opsin in *Hyalella azteca* and *Daphnia pulex*.

Species	Short-wavelength sensitive / UV		Long-wavelength sensitive		Peropsins / Neuropsins / Encephalopsins		Melanopsins		Opsin-like transcripts (Outgroups)	
	Genes	Isoforms	Genes	Isoforms	Genes	Isoforms	Genes	Isoforms	Genes	Isoforms
<i>Hyalella azteca</i>	1	2	3	3	0	0	0	0	1	1
<i>Daphnia pulex</i>	5	15	3	24	0	0	2	4	1	1

Table 5. Identification and classification of *Hyalella azteca* and *Daphnia pulex* opsins based on phylogenetic inference and HMM profile alignments.

Putative visual opsins are marked in black, under the “Visual” column, when both methodologies are in agreement and in gray when they differ.

Species	Sequence ID	Phylogeny Clade	Visual	HHBlits w/ Uniclust30			
				Prob.	E-value	P-value	Hit ID
<i>Hyalella azteca</i>	Haz_DN255_c1_g1_i1	OUTGROUP		100.0	1.2E-66	4.3E-72	A0A0P5Y4K7_9CRUS Putative Tachykinin peptides receptor 99D (Fragment)
	Haz_DN34101_c1_g4_i2	SWS		100.0	1.2E-77	4.3E-83	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
	Haz_DN34101_c1_g6_i4	SWS		100.0	1.9E-76	6.9E-82	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
	Haz_DN31505_c2_g2_i1	LWS		100.0	8.8E-80	3.1E-85	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
	Haz_DN31505_c2_g4_i1	LWS		100.0	3.8E-80	1.4E-85	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
	Haz_DN44548_c0_g1_i1	LWS		100.0	6.9E-72	2.5E-77	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
<i>Daphnia pulex</i>	Dpu_DN105553_c0_g1_i1	MELANOPSIN		100.0	8E-52	2.8E-57	H0UT82_CAVPO Uncharacterized protein OS=Cavia porcellus GN=GALR1
	Dpu_DN16067_c1_g1_i1	LWS		100.0	6.3E-74	2.2E-79	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
	Dpu_DN16472_c1_g1_i1	OUTGROUP		100.0	1.8E-73	6.2E-79	A0A0P5Y4K7_9CRUS Putative Tachykinin peptides receptor 99D (Fragment)
	Dpu_DN21293_c0_g1_i1	MELANOPSIN		100.0	3.4E-40	1.3E-45	A0A0P5EI05_9CRUS Class a rhodopsin g-protein coupled receptor gprop2
	Dpu_DN21293_c0_g2_i1	MELANOPSIN		100.0	3.6E-62	1.3E-67	A0A1A6GZZ2_NEOLE Uncharacterized protein OS=Neotoma lepida
	Dpu_DN21293_c0_g2_i2	MELANOPSIN		100.0	3.9E-38	1.4E-43	A0A0F8BM2_LARCR Melanopsin-B OS=Larimichthys crocea GN=EH28_08950
	Dpu_DN22657_c1_g4_i1	SWS/UV		100.0	1.7E-79	6.1E-85	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
	Dpu_DN23719_c0_g2_i1	SWS/UV		100.0	1.9E-78	6.8E-84	A0A0P5RD30_9CRUS Class a rhodopsin g-protein coupled receptor gprop2
	Dpu_DN23719_c0_g2_i2	SWS/UV		100.0	4.5E-78	1.6E-83	A0A0P5RD30_9CRUS Class a rhodopsin g-protein coupled receptor gprop2

Dpu_DN23719_c0_g3_i1	SWS/UV		100.0	2E-78	7.4E-84	A0A0P5RD30_9CRUS Class a rhodopsin g-protein coupled receptor gprop2
Dpu_DN25251_c0_g2_i1	SWS/UV		100.0	2.6E-77	9.3E-83	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
Dpu_DN25251_c0_g2_i3	SWS/UV		100.0	3.4E-78	1.2E-83	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
Dpu_DN25251_c0_g2_i4	SWS/UV		100.0	1.9E-77	6.8E-83	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
Dpu_DN25251_c0_g2_i5	SWS/UV		100.0	8.2E-77	2.9E-82	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
Dpu_DN25251_c0_g3_i1	SWS/UV		100.0	2.6E-77	9.1E-83	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
Dpu_DN25293_c2_g1_i1	SWS/UV		100.0	5.7E-48	2.2E-53	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
Dpu_DN25293_c2_g1_i3	SWS/UV		100.0	2E-78	7.1E-84	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
Dpu_DN25293_c3_g1_i1	SWS/UV		100.0	5.7E-35	2E-40	D0E2W5_CHICK Uncharacterized protein OS=Gallus gallus GN=OPNVA PE=2 SV=1
Dpu_DN25293_c4_g1_i1	SWS/UV		100.0	1.1E-75	3.9E-81	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
Dpu_DN25293_c7_g1_i1	SWS/UV		100.0	1.4E-33	5.1E-39	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
Dpu_DN25377_c0_g1_i1	LWS		100.0	6.3E-75	2.2E-80	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g1_i10	LWS		100.0	3.1E-62	1.1E-67	A0A0P5KGY5_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g1_i11	LWS		100.0	4.1E-76	1.4E-81	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g1_i12	LWS		100.0	1.4E-75	4.9E-81	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g1_i4	LWS		100.0	1E-56	3.6E-62	A0A0P5KGY5_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g1_i5	LWS		100.0	5.3E-82	1.8E-87	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g1_i6	LWS		100.0	8E-77	2.8E-82	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g1_i7	LWS		100.0	2E-76	6.8E-82	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g1_i8	LWS		100.0	1.1E-80	3.7E-86	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g1_i9	LWS		100.0	4E-77	1.4E-82	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g2_i1	LWS		100.0	4.5E-76	1.6E-81	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g2_i11	LWS		99.9	8.6E-29	3.3E-34	E9FX22_DAPPU Octopamine receptor beta-2-like protein OS=Daphnia pulex

Dpu_DN25377_c0_g2_i2	LWS		100.0	1.7E-75	6E-81	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g2_i3	LWS		100.0	1.2E-75	4.2E-81	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g2_i5	LWS		100.0	4.4E-76	1.5E-81	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g2_i6	LWS		100.0	4.2E-77	1.4E-82	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g2_i8	LWS		100.0	9.3E-76	3.2E-81	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25377_c0_g2_i9	LWS		100.0	4.3E-76	1.5E-81	A0A0P5USN7_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25506_c1_g1_i1	LWS		100.0	8.9E-35	3.2E-40	E6ZHB1_DICLA Beta-1 adrenergic receptor OS=Dicentrarchus labrax GN=ADRB1
Dpu_DN25506_c1_g1_i2	LWS		100.0	1.1E-34	3.9E-40	E6ZHB1_DICLA Beta-1 adrenergic receptor OS=Dicentrarchus labrax GN=ADRB1
Dpu_DN25506_c3_g1_i1	LWS		100.0	4.3E-80	1.4E-85	A0A0P4XHY0_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25506_c3_g1_i3	LWS		100.0	1.1E-60	4E-66	A0A0P5KGY5_9CRUS Class a rhodopsin g-protein coupled receptor gprop1
Dpu_DN25506_c4_g1_i1	LWS		100.0	2.3E-48	8.2E-54	OPSL_LIMPO Lateral eye opsin OS=Limulus polyphemus PE=1 SV=1
Dpu_DN27815_c0_g1_i1	SWS/UV		100.0	9.5E-71	3.4E-76	A0A0P5RD30_9CRUS Class a rhodopsin g-protein coupled receptor gprop2

Figure Captions

Figure 1. Maximum-Likelihood phylogeny of opsins estimated using putative opsin proteins identified by our annotation pipeline from the *de novo* transcriptome assemblies of *Hyalella azteca* and *Daphnia pulex*, along with a dataset of reference opsin sequences. Clades are annotated with opsin types contained therein and, in the case of visual opsins, with their inferred spectral sensitivities.

Figure 2. Expanded view of the melanopsin, Arthropod LWS, and Arthropod SWS/UV clades. Large non-crustacean clades have been collapsed for readability. Support values correspond to SH-aLRT/aBayes/UFBoot, not shown when bootstrap support < 75.

Figure 3. Intron-Exon gene structure patterns of representative *Hyalella azteca* SWS/UV and LWS opsins.

Figure 4. Intron-Exon gene structure patterns of representative *Daphnia pulex* opsins SWS/UV and LWS opsins, which are arranged in the genome in distinct patterns according to opsin type.

Figure 5. Exon size distribution for both *Daphnia pulex* and *Hyalella azteca*. X-axis is in base pair units (bp), while Y-axis represents proportion of transcripts in said range.

Figure 1.

Figure 2.

Hyaella azteca

Figure 3.

Daphnia pulex

Dpu_DN22657_c1_g4_i1

Dpu_DN22657_c1_g4_i1

Scaffold: GL733199.1

Short Wavelength Sensitive (SWS)

Dpu_DN25506_c3_g1_i1

Dpu_DN25506_c1_g1_i2

Dpu_DN25506_c4_g1_i1

Dpu_DN25506_c3_g1_i3

Scaffold: GL732600.1

Long Wavelength Sensitive (SWS)

Figure 4

Figure 5.

CHAPTER V

TRANSCRIPTOMIC INSIGHTS INTO THE LOSS OF VISION IN MOLNÁR JÁNOS CAVE'S CRUSTACEANS

Jorge Luis Pérez-Moreno, Gergely Balázs, Heather D. Bracken-Grissom

This chapter was published in:

Integrative and Comparative Biology, 58(3):452-464, 2018.

Abstract

Animals that inhabit subterranean environments often undergo various distinct physiological, morphological, and behavioral modifications (referred to as “troglomorphy”) as they transition to life in perpetual darkness. However, the molecular basis behind these troglomorphic changes remains poorly understood. Important questions remain to be answered concerning the mechanisms involved in the loss of traits at the transcriptomic level, and the role of gene expression in driving adaptive divergence and/or convergence in these systems. In this study we investigate the transcriptional basis behind vision loss in populations of cave-dwelling crustaceans. To do so, we employ phylogenetic and transcriptomic methods on surface and cave-adapted populations of an emerging model cave species, the isopod *Asellus aquaticus* (Linnaeus, 1758), and the amphipod *Niphargus hrabei* S. Karaman, 1932. These two species show contrasting directionality in the surface-cave transition that positions them as ideal study subjects; *A. aquaticus* is common in surface waters and is only occasionally found in caves, whereas *N. hrabei* has successfully colonized surface environments despite belonging to an almost exclusively cave-dwelling genus. By sequencing and assembling robust *de novo* transcriptomes we characterized differences in visual genes and pathways among surface and cave populations of the aforementioned species. Our results indicate that despite having reduced eyes, recent cave colonizer *A. aquaticus* is still capable of expressing functional visual opsins and other components of the phototransduction pathway within the cave. *Niphargus hrabei*, a species with an ancient cave origin, showed no indication of being capable of sight with an incomplete phototransduction pathway in all populations. However, the expression of functional visual opsins was maintained in this

species for a yet unknown function. With the present study, we aim to bring forth the Molnár János cave system as a promising research avenue to improve our understanding of patterns of reduction and loss of vision in caves and other aphotic environments.

Keywords: adaptation; biospeleology; evolution; opsin; phototransduction; rnaseq

Introduction

As populations of animal species colonize and adapt to life in the perpetual darkness of caves, they are often subject to a suite of morphological, physiological, and behavioral changes that are collectively described as “troglomorphic” (Porter and Crandall 2003; Mejía-Ortíz et al. 2006; Pérez-Moreno et al. 2016). Cave-dwelling animals typically exhibit dichotomous troglomorphic traits, classified as either progressive/constructive or regressive/reductive depending on the assumed directionality of these changes (Porter and Crandall 2003; Mejía-Ortíz et al. 2006; Pérez-Moreno et al. 2016). Examples of progressive troglomorphic changes include improved spatial orientation and olfaction as well as larger sensory and ambulatory appendages (Turk et al. 1996; Li and Cooper 2001, 2002; Mejía-Ortíz and Hartnoll 2006). In contrast, regressive troglomorphic modifications typically feature a decrease or loss of characters in comparison with their surface-inhabiting counterparts (Pérez-Moreno et al. 2016). Among these regressive traits one may find those that are more commonly associated with subterranean life such as slower metabolism, little or no pigmentation, and the reduction or loss of vision and visual structures (Sket 1985; Wilkens 1986; Mejía-Ortíz and López-Mejía 2005; Bishop and Iliffe 2009).

Trogomorphic phenotypes, with reduced or absent visual structures, have arisen repeatedly and independently across the phylogenetic breadth of the animal kingdom in response to analogous environmental conditions (Caccone and Sbordoni 2001; Protas et al. 2007; Pérez-Moreno et al. 2016). To this day, most comprehensive studies of cave fauna's regressive evolution have focused on freshwater fish, particularly the Mexican Cave Tetra *Astyanax mexicanus* (Porter and Crandall 2003). This species has served an important role as a model organism for evolutionary and developmental biology, facilitated by the presence of “normal” surface populations, numerous independent cave colonization events, and varying degrees of eye loss (Jeffery and Martasian 1998; Yamamoto and Jeffery 2000; Jeffery 2001). This has led to significant insights regarding the diverse mechanisms by which eyes and vision can be lost in vertebrates. In comparison, invertebrates have received little attention, despite often comprising the majority of the animal diversity found in caves (Protas et al. 2011; Pérez-Moreno et al. 2016). Given the global diversity of cave invertebrates and phenotypic convergence that has arisen from a multitude of independent colonization events; these organisms can be considered ideal subjects to study the predictability of evolution, particularly in regards to the loss of complex traits such as vision (Pérez-Moreno et al. 2016).

Recent advances in cave diving technology and in the field of genomics now offer an unprecedented opportunity to investigate the molecular basis and evolution of vision loss in subterranean organisms (Pérez-Moreno et al. 2016). Likewise, molecular techniques have been pivotal for increasing our understanding of invertebrate vision and the genes and pathways underlying this trait (Cronin and Porter 2014). However, to date

there are relatively few empirical studies of loss of function in visual genes and pathways in subterranean environments (Niemiller et al. 2013). This absence is especially noted for those studies that incorporate high-throughput sequencing techniques (Pérez-Moreno et al. 2016). Genomic and transcriptomic methods, such as RNA sequencing, have proven useful tools to examine the expression of pathways that determine an organism's ability to visually detect light (i.e. the phototransduction pathway; Porter et al. 2012). This pathway is activated through the detection of specific wavelengths of light by visual pigments composed of an opsin protein bound to a chromophore (e.g. 11-cis retinal, a vitamin A derivative; Nathans 1987; Tierney et al. 2015). Opsins are a diverse group of photoreceptor proteins found in metazoans that are essential for animal vision (Terakita 2005). Belonging to the G-protein-coupled receptors family (GPCRs), these membrane-associated proteins play important roles in both visual and non-visual phototransduction (Shichida and Matsuyama 2009). When bound to a chromophore, opsins form photon-absorbing pigments that activate G-proteins when exposed to light (Nathans 1987). This exposure thus initiates the phototransduction signaling cascade, which is tightly linked to regulation of circadian rhythms, phototactic behaviors, and most importantly, vision (e.g., Arendt et al. 2004; Shichida and Matsuyama 2009; Pérez-Moreno et al. 2018;) . The specific amino acid residues that interact with the chromophore to form a functional photopigment have a direct influence on the opsin's spectral sensitivity by permitting the absorption of distinct wavelengths of light (Imai et al. 1997; Kuwayama et al. 2002; Porter et al. 2007; Katti et al. 2010). This direct correlation between opsins' amino acid compositions with specific spectral sensitivities not only allows for fine scale spectral tuning at both ecological and evolutionary timescales, but also makes opsins tractable to

functional classification by sequence similarity and phylogenetic approaches (Mirzadegan et al. 2003; Matsumoto and Ishibashi 2016; Pérez-Moreno et al. 2018).

The isopod *Asellus aquaticus* (Linnaeus, 1758) and the amphipod *Niphargus hrabei* S. Karaman, 1932 are two crustacean species that function as ideal and complementary systems to explore questions regarding the loss of eyes and vision. *Asellus aquaticus* is a widespread species of freshwater isopod found in surface waters throughout Europe that occasionally colonizes caves, upon which they exhibit distinct troglomorphic morphotypes with varying degrees of eye reduction (Protas et al. 2011; Konec et al. 2015; Stahl et al. 2015; Pérez-Moreno et al. 2017; Fig. 1). *Niphargus hrabei*, on the other hand, is an almost exclusively cave-dwelling genus that is hypothesized to have ancient cave-origins as it exhibits troglomorphic morphotypes and lacks eyes in both surface and cave populations (Balázs et al. 2015; Copilas-Ciocianu et al. 2017; Pérez-Moreno et al. 2017; Fig. 1). Furthermore, these two species, which are distributed across a wide-area of Central and Eastern Europe, are sympatric above and below ground.

The present study aims to investigate the molecular basis behind the loss of vision, with an emphasis on the phototransduction pathway, via transcriptomic and phylogenetic approaches. Specifically, we aimed to characterize visual opsins in surface and cave populations to elucidate any differences that may be present, as well as assessing the presence of other phototransduction genes that are indicative of light-detection abilities. This investigation was undertaken by capitalizing on natural

experimental conditions generated by the presence of *A. aquaticus* and *N. hrabei* in the Molnár János Cave system (Budapest, Hungary). Despite having markedly different selective pressures, Molnár János Cave, the adjacent Malom Lake, and the Danube River are all inhabited by both *A. aquaticus* and *N. hrabei* (Pérez-Moreno et al. 2017). As such, this system facilitates comprehensive and intricate investigations regarding the adaptation and evolution of cave fauna.

Methods

Specimen Collection

Specimens of *A. aquaticus* and *N. hrabei* were sampled from Rakos Rock in Molnár János Cave (Fig. 2) using a “Sket bottle” (Chevaldonné et al. 2008) following the methods of Pérez-Moreno et al. (2017). To minimize stress to the animals and exposure to light, cave specimens were preserved in situ immediately upon resurfacing from the dive (within the tunnel connecting to the Kessler Hubert air chamber). Surface specimens of both species were additionally sampled from surface populations (Malom Lake and the Soroksár branch of the Danube River; Fig. 2). As with the cave specimens, surface individuals were also preserved in situ and immediately upon collection. Following the aforementioned preservation in RNAlater® (ThermoFisher Scientific), the samples were frozen at -80 °C to prevent nucleic acid degradation and thus allow for subsequent RNA isolation, sequencing, and transcriptomic analyses.

Data and Quality Control

Total RNA was individually extracted from two whole specimens per species (*A. aquaticus* and *N. hrabei*) per population (Molnár János Cave, Malom Lake, Soroksár) using TRIzol[®] reagent (ThermoFisher Scientific). Each specimen's RNA was prepared for sequencing in individual reactions with an initial rDNase[®] (Macherey Nagel) treatment, and subsequent mRNA isolation, cDNA synthesis, and barcode/adaptor ligation with a NEBNext[®] Ultra™ II Directional RNA Library Prep Kit for Illumina[®]. The resulting barcoded libraries were size-selected with a Pippin Prep[®] instrument (Sage Science), pooled in equimolar concentrations, and sent for sequencing on an Illumina[®] HiSeq 4000 lane.

Quality of the raw sequencing reads was evaluated via FastQC (Andrews 2010) prior to the assembly process to inform quality and adaptor trimming with Trimmomatic 0.36 (Bolger et al. 2014). Trimmed and quality-filtered reads were then piped to *k*-mer based Rcorrector (Song and Florea 2015) for correction of random sequencing errors and to BBNorm for read depth normalization prior to assembly.

Transcriptome Assembly and Post-processing

The clean and normalized sequencing reads were assembled into *de novo* transcriptomes for each population and species with the Trinity pipeline (version 2.5.0; Grabherr et al. 2011; Haas et al. 2013), specifying a minimum contig length of 200bp and a *k*-mer size of 23. Post-processing of the *de novo* assemblies consisted of removal of duplicate transcripts and rRNA sequences (dedupe.sh and BBduk from the BBTools

suite). Transrate 1.0.3 and BUSCO 3.0.2 (Benchmarking Universal Single-Copy Orthologs; Simão et al. 2015; Smith-Unna et al. 2016) were then respectively employed to calculate summary statistics and to assess the completeness of the transcriptome assemblies. BUSCO evaluates transcriptome completeness in an evolutionary informed context by evaluating the presence and/or fragmentation of universal single copy orthologs (Simão et al. 2015). BUSCO analyses were conducted using OrthoDB's Arthropoda database of orthologous groups (n= 1066; Waterhouse et al. 2013) as a reference dataset.

Phylogenetic Identification and Annotation of Visual Genes

A suite of phylogenetic tools were employed to identify and functionally classify expressed putative opsins in our *de novo* transcriptomes as part of a previously described pipeline (Pérez-Moreno et al. 2018). In essence, this pipeline extracts open-reading frames (ORFs) from the entire transcriptome and then runs these data through a modified command-line version of the Phylogenetically-Informed Annotation (PIA) tool (Speiser et al. 2014). Briefly, PIA allows for the identification of visual genes by initial BLAST searches against a database of known visual genes, and subsequent alignment and placement of BLAST hits in pre-computed phylogenies to differentiate false-positives from the genes of interest. Our pipeline extends PIA's functionality by integrating additional steps of unbiased automated pruning of tree branches exceeding certain threshold (in this case, 4x mean absolute deviation of branch lengths), and redundancy reduction, which results in more concise and cleaner candidate gene lists (Pérez-Moreno et al. 2018).

Finally, opsins were further investigated as they are initiators and essential components of the phototransduction cascade. The amino acid composition of opsin proteins can determine their spectral sensitivity and thus permit an organism to detect specific wavelengths. This correlation between composition and absorption spectrum makes phylogenetic approaches particularly suitable for the identification and functional classification of opsins (Mirzadegan et al. 2003; Matsumoto and Ishibashi 2016; Pérez-Moreno et al. 2018). As such, the previously identified putative opsins were aligned with PROMALS3D (Pei and Grishin 2014) using a large opsin reference dataset (n=910; Porter et al. 2012) that includes the main types of opsin subfamilies and representatives of a variety of spectral sensitivities from an extensive taxonomical range. A final opsin phylogenetic tree was then reconstructed with IQ-tree (Nguyen et al. 2015), a maximum-likelihood based phylogenetics software package that has performed favorably in recent benchmarks when compared to more established alternatives (e.g. RAxML, PhyML, FastTree, etc.; Zhou et al. 2017). The IQ-tree phylogenetic reconstruction was undertaken using an LG general amino acid replacement matrix, under a FreeRate model with 9 rate categories, and empirical base frequencies (LG+R9+F; Le and Gascuel 2008; Soubrier et al. 2012) as suggested by ModelFinder (Kalyaanamoorthy et al. 2017). Subsequently, split support was assessed using three different methodologies: an Ultra-fast bootstrap approximation (UFBoot; 10,000 replicates), a Shimodaira–Hasegawa–like approximate likelihood ratio test (SH-aLRT; 10,000 replicates), and an approximate Bayes test (Guindon et al. 2010; Anisimova et al. 2011; Minh et al. 2013).

Results and Discussion

The present study aimed to identify and characterize phototransduction genes, with an emphasis on visual opsins, to enable intra- and inter-specific comparisons amongst surface and cave populations of *A. aquaticus* and *N. hrabei*. For this purpose, we have contributed additional cave & surface *A. aquaticus* transcriptomes from previously unstudied populations, in addition to the first transcriptomic resources available for the speciose and wide-spread genus *Niphargus*. Here, we present and discuss this study's *de novo* transcriptome assemblies, their summary statistics, and completeness. Secondly, we compare and contrast the cave and surface transcriptomes in search of visual genes and those involved in the phototransduction pathway. In light of our results, we infer the functional status of their visual opsins and discuss the future prospects for this emerging study system.

Transcriptome sequencing and assembly

Over 380M reads and approximately 115 gigabases of data were obtained for our 12 samples, respectively yielding approximately 32M reads and 9.5 gigabases per individual. The quality of our *de novo* transcriptome assemblies was evaluated by examining the summary statistics as reported by Transrate (Smith-Unna et al. 2016) and the results of the BUSCO (Simão et al. 2015) assessment of completeness. *Asellus aquaticus* transcriptome assemblies recovered 86903–123735 contigs with a mean sequence length of 938–1076 base pairs (Table 1). Of these, 25562–30264 sequences contained Open Reading Frames (ORFs) designating them as putative protein-coding

genes. Transcriptome contiguity was evaluated by calculating N50 values, a commonly employed metric to assess the contiguity of a *de novo* transcriptome assembly. *Asellus aquaticus de novo* transcriptomes had N50 values ranging from 1737 bp to 2022 bp, which are more than twice as long as the only previous transcriptomic study for this species (Stahl et al. 2015). Similarly, the *de novo* transcriptomes for *N. hrabei* were comprised of 107451–134592 contigs with a mean sequence length of 874–883 bp. The N50 values for *N. hrabei* transcriptomes range from 1598 bp to 1648 bp, which also compare favorably with previous amphipod transcriptome studies (Lan et al. 2017). Additional metrics for our *de novo* transcriptomes, as well as for the reference genomes, are given in Table 1.

Transcriptome completeness of our *de novo* transcriptomes, as assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO), was favorable for both species (Table 2). In *A. aquaticus* we were able to find 90.1–93.6% complete sequences of the 1066 genes employed for benchmarking in Arthropoda. An additional 3.3–6.2% were present but fragmented, and only 2.7–3.9% were not found in the transcriptome. Similarly, *N. hrabei*'s transcriptome was found to be nearly complete but slightly more fragmented with 82–90.5% full-length BUSCO genes, 5.3–7.9% fragmented, and 4.2–10.5% missing. *Niphargus hrabei*'s larger proportion of missing BUSCOs, despite equivalent sequencing depth, might reflect ancestral gene-loss patterns due to reductive troglomorphy. Future explorations of speleotranscriptome completeness in Niphargidae and other ancient stygobiont/troglobiont lineages are needed to confirm this hypothesis.

Subterranean expression of opsins and the phototransduction pathway

Phylogenetic analyses revealed an almost complete phototransduction pathway, as defined in the PIA tool (Speiser et al. 2014) for the cave transcriptome of *A. aquaticus* (Table 3), which suggests that they might retain the ability to detect light despite their reduced eyes. However, further analyses are required to determine if these are potentially down-regulated in the cave versus the surface populations or if alternate mechanisms are at play. It is possible that the expression of phototransduction genes could be maintained at baseline levels, while vision loss is occurring by alternate mechanisms. For example, previous studies have shown a reduction or loss of optical ganglia and respective brain regions in other cave-adapted crustaceans (Stegner et al. 2015). *Niphargus hrabei* cave transcriptome contained notably fewer phototransduction genes (Table 3), which could be attributed to the cave origin of this ancient genus leading to pseudogenization events. However, the lack of expression of these genes suggests that *N. hrabei* is likely to remain blind in the darkness of caves and when exposed to light in surface environments.

Following identification and annotation of putative opsins using our customized PIA pipeline, these were subsequently aligned with a reference dataset for a final step of phylogenetic functional annotation. As a result, a large phylogeny (Fig. 3) was produced where putative opsins were assigned functional identities and spectral sensitivities were inferred based on their phylogenetic position. The transcriptome for the cave morphotype of *A. aquaticus* contained a single Short Wavelength Sensitive/UV sensitive opsin and one Long Wavelength Sensitive opsin, which did not differ significantly from the surface

individuals (Figs. 4 & 5). *Niphargus hrabei* on the other hand was found to express no SWS/UV opsins, but a single LWS opsin that was also identical in all populations (Fig. 5). Pairwise sequence alignments of each species and morphotype's opsins were compared with reference sequences, bovine (*Bos taurus*) and squid (*Todarodes pacificus*) rhodopsins (Accessions NP_001014890 and CAA49906 respectively) to further elucidate any amino acid substitutions differences among populations and their effects on function and inferred spectral sensitivities (Table 4; Palczewski et al. 2000; Takahashi and Ebrey 2003; Katti et al. 2010)

All of the opsins identified by our analyses appear to be functional, including those in the cave populations. The opsin annotated as SWS/UV for *A. aquaticus* contains a lysine residue in the site corresponding to position 90 in bovine rhodopsin, further supporting its putative UV sensitivity (Salcedo et al. 2003). It is thus plausible that the expression of this opsin has not been fully lost within the cave, despite the lack of light, due to the recent divergence of *A. aquaticus* surface and cave populations. On the other hand, both *A. aquaticus* and *N. hrabei* have LWS opsins that lack a lysine residue in this position, and in fact appear to be more similar to each other than to the SWS/UV found in *A. aquaticus*. These LWS opsins only differ between the two species at bovine rhodopsin site 126 and 269, the significance of which yet remains to be investigated. However, LWS opsins for both species show a serine residue at site 292 instead of an alanine, which has been previously shown to be able to shift invertebrate spectral sensitivity in either direction (Salcedo et al. 2009; Katti et al. 2010).

Asellus aquaticus transcriptome analyses show that major key components of the phototransduction pathway are expressed within the cave population. All populations express the same opsin genes as surface individuals (Figure 4 & 5), with only minor differences. The SWS/UV opsin found in *A. aquaticus* is identical in all populations, except for their total length. However, there is no other indication that they consist of different isoforms, which suggests that the length is due to fragmentation during assembly. *Asellus aquaticus* LWS opsin in Malom Lake shows an aspartic acid to glutamic acid substitution at site 12, position which suggests that the mutation is inconsequential and possibly attributed to allelic variation. A similar occurrence is observed in *N. hrabei*, where the Soroksár (Danube River) population show a single serine to arginine substitution. There were no indications of pseudogenization in neither opsins nor other expressed phototransduction genes, as has been the case in other subterranean animals (Yokoyama et al. 1995; Kim et al. 2011; Hinaux et al. 2013; Rétaux and Casane 2013). The sequence similarities and lack of significant differentiation between the cave and surface population's expressed opsins are perhaps not that surprising, given recent divergence-time estimates based on molecular clocks (circa 139,000 years ago; (Pérez-Moreno et al. 2017). Similarly, an *A. aquaticus* study by Protas et al. (2011) found a candidate gene (*lim1*) that appeared to be linked with loss of eyes, by employing reduced-representation DNA sequencing techniques and QTL analyses. However, this gene was not thought to be responsible for the loss of this trait, but rather a gene located in its genomic vicinity (Protas et al. 2011). Future research with the use of whole genome sequencing will undoubtedly help answer if perhaps drift or selective

forces are acting in areas upstream of the opsin gene, leading to reduced but detectable transcription.

Alternatively, some subterranean diving beetles from Australia have been found to be negatively phototactic despite their blindness and the complete absence of light in their environments (Tierney et al. 2015; Langille et al. 2018). These findings suggest that the ability to detect light may persist in some taxa, perhaps as an advantage to early cave colonizers or to serve for a yet unknown purpose once the organism is fully darkness-adapted (Langille et al. 2018). Selection tests on a LWS opsin in one of these subterranean beetles suggest the latter, as they did not result in any significant differentiation between habitats, and in fact strongly suggested that purifying selection was acting on the gene instead (Tierney et al. 2015). Similar observations have been made in the cave model species *Astyanax mexicanus*, where transcription of opsins are evident during development but not necessarily in adult individuals (Langecker et al. 1993). These findings could very well be indicative of pleiotropic roles for LWS opsins in cave organism that deserve further attention. The possibility for the later occurrence is also reflected in our results for the amphipod *N. hrabei*. Both surface and cave populations of this species do not appear to express most components of the phototransduction pathway (Table 3), with the notable exception of identical opsin genes being transcribed in both cave and surface individuals. Due to the lack of functional visual pathways and visual structures, it is hypothesized that the presence of opsin proteins in this species is due to pleiotropy and could perhaps even relate to extraocular photoreception through alternate pathways (see Kelley and Davies [2016] for a review on

non-visual light detection). Previous studies investigating amphipod cave and surface populations in the USA (*Gammarus minus*) also found that functional visual opsins were still present in aphotic cave populations, once again suggesting that visual opsins can be co-opted for other purposes (Carlini et al. 2013). However, the aforementioned study solely investigated opsins and did not verify whether the remainder of the phototransduction pathway had been transcribed. Although several studies have attributed the expression of functional visual opsins to pleiotropy, the possible roles that these opsins might be playing has not yet been subject of attention (Tierney et al. 2015). As with *A. aquaticus*, future prospects for the elucidation of these possibly pleiotropic roles are promising as whole genome sequencing continues to become more affordable and a real possibility for non-model cave organisms.

Conclusions

The present study provides first insights into the evolution of vision loss in subterranean crustaceans from Molnár János Cave. We hypothesize that vision loss in recent cave colonizer *A. aquaticus* is likely due to down-regulation of visual opsins, which directly leads to reduced activation of the pathway that is still present in the transcriptome. Vision loss in recent surface colonizer *N. hrabei* was likely due to repressed expression or complete loss of several phototransduction genes, rather than by the down-regulation of the cascade's activator as in the case of *A. aquaticus*. It is hypothesized that the lack of detectable expression of these genes is a result of pseudogenization or complete gene loss given the ancient cave origin of the genus *Niphargus* Schiödte, 1847 (circa 30-50 mya). Although, functional visual opsins were

found being expressed in *N. hrabei*, determining whether these form part of an alternate photoreceptive pathway or play a role in a yet unknown function is still not possible. Finally, we found no convincing evidence of transcriptomic convergence in these two species, as they appear to have lost their vision by different mechanisms. Whole methylome and genome sequencing projects currently being undertaken in this system will undoubtedly provide additional insights into the mechanisms involved in vision loss within this system and other extreme aphotic environments.

Funding

JPM was supported by the Philip M. Smith Graduate Research Grant for Cave and Karst Research from the Cave Research Foundation, The Crustacean Society Scholarship in Graduate Studies, Florida International University's Dissertation Year Fellowship, and the National Science Foundation's Doctoral Dissertation Improvement Grant (#1701835) awarded to JPM and HBG. HBG was additionally supported by NSF's Division of Environmental Biology Bioluminescence and Vision grant (#1556059). GB was supported by the Hungarian Scientific Research Fund (#K-105517), the Hungarian Ministry of Human Capacities (#ÚNKP-17-3), and by the National Research, Development and Innovation Fund for International Cooperation (#SNN 125627).

Acknowledgments

The authors would like to thank Drs. Megan Porter and Lauren Sumner-Rooney for organizing the SICB symposium "Evolution in the dark: unifying understanding of

eye loss”. This symposium was generously sponsored by the Company of Biologists (<http://www.biologists.com>), the Palaeontological Association (PA-GA201707), the American Microscopical Society, the Crustacean Society, and the SICB divisions DEDB, DEE, DIZ, DNB, and DPCB. The authors would also like to acknowledge the Instructional & Research Computing Center (IRCC) at Florida International University for providing High Performance Computing resources that have contributed to the research results reported within this article.

References

- Andrews S. 2010. FastQC A Quality Control tool for High Throughput Sequence Data <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. 2011. Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-based Approximation Schemes. *Syst Biol* 60:685–99.
- Arendt D, Tessmar-Raible K, Snyman H, Dorresteijn AW, Wittbrodt J. 2004. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science* 306:869–71.
- Balázs G, Angyal D, Kondorosy E. 2015. Niphargus (Crustacea: Amphipoda) species in Hungary: literature review, current taxonomy and the updated distribution of valid taxa. *3974:361–76*.
- Bishop RE, Iliffe TM. 2009. Metabolic rates of stygobiontic invertebrates from the Túnel de la Atlántida, Lanzarote. *Mar Biodivers* 39:189–94.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–20.
- Caccone A, Sbordoni V. 2001. Molecular biogeography of cave life: a study using mitochondrial DNA from bathysciine beetles. *Evolution* 55:122–30.
- Carlini DB, Satish S, Fong DW. 2013. Parallel reduction in expression, but no loss of functional constraint, in two opsin paralogs within cave populations of *Gammarus minus* (Crustacea: Amphipoda). *BMC Evol Biol* 13:89–89.

- Chevaldonné P, Sket B, Marschal C, Lejeune C, Calado R. 2008. Improvements to the “Sket bottle”: a simple manual device for sampling small crustaceans from marine caves and other cryptic habitats. *J Crustac Biol* 28:185–88.
- Copilas-Ciocianu D, Fişer C, Borza P, Balázs G, Angyal D, Petrusek A. 2017. Low intraspecific genetic divergence and weak niche differentiation despite wide ranges and extensive sympatry in two epigean *Niphargus* species (Crustacea: Amphipoda). *Zool J Linn Soc*.
- Cronin TW, Porter ML. 2014. The Evolution of Invertebrate Photopigments and Photoreceptors. In: *Evolution of Visual and Non-visual Pigments* Boston, MA: Springer.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson D a, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–52.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* 59:307–21.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512.
- Hinaux H, Poulain J, Da Silva C, Noirot C, Jeffery WR, Casane D, Rétaux S. 2013. De novo sequencing of *Astyanax mexicanus* surface fish and Pachón cavefish transcriptomes reveals enrichment of mutations in cavefish putative eye genes. *PloS ONE* 8:e53553–e53553.
- Imai H, Kojima D, Oura T, Tachibanaki S, Terakita A, Shichida Y. 1997. Single amino acid residue as a functional determinant of rod and cone visual pigments. *Proc Natl Acad Sci* 94:2322–26.
- Jeffery WR. 2001. Cavefish as a Model System in Evolutionary Developmental Biology. *Dev Biol* 231:1–12.
- Jeffery WR, Martasian DP. 1998. Evolution of eye regression in the cavefish *Astyanax*: apoptosis and the Pax-6 gene. *Am Zool* 38:685–96.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–89.

- Katti C, Kempler K, Porter ML, Legg A, Gonzalez R, Garcia-Rivera E, Dugger D, Battelle B-A. 2010. Opsin co-expression in *Limulus* photoreceptors: differential regulation by light and a circadian clock. *J Exp Biol* 213:2589–2601.
- Kelley JL, Davies WIL. 2016. The Biological Mechanisms and Behavioral Functions of Opsin-Based Light Detection by the Skin. *Front Ecol Evol* 4.
- Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, Marino SM, Sun X, Turanov AA, Yang P, Yim SH, Zhao X, Kasaikina MV, Stoletzki N, Peng C, Polak P, Xiong Z, Kiezun A, Zhu Y, Chen Y, Kryukov GV, Zhang Q, Peshkin L, Yang L, Bronson RT, Buffenstein R, Wang B, Han C, Li Q, Chen L, Zhao W, Sunyaev SR, Park TJ, Zhang G, Wang J, Gladyshev VN. 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* 479:223–27.
- Konec M, Prevorčnik S, Sarbu SM, Verovnik R, Trontelj P. 2015. Parallels between two geographically and ecologically disparate cave invasions by the same species, *Asellus aquaticus* (Isopoda, Crustacea). *J Evol Biol* 28:n/a–n/a.
- Kuwayama S, Imai H, Hirano T, Terakita A, Shichida Y. 2002. Conserved Proline Residue at Position 189 in Cone Visual Pigments as a Determinant of Molecular Properties Different from Rhodopsins [†]. *Biochemistry (Mosc)* 41:15245–52.
- Lan Y, Sun J, Tian R, Bartlett DH, Li R, Wong YH, Zhang W, Qiu J-W, Xu T, He L-S, Tabata HG, Qian P-Y. 2017. Molecular adaptation in the world's deepest-living animal: Insights from transcriptome sequencing of the hadal amphipod *Hirondellea gigas*. *Mol Ecol* 26:3732–43.
- Langecker TG, Schmale H, Wilkens H. 1993. Transcription of the opsin gene in degenerate eyes of cave-dwelling *Astyanax fasciatus* (Teleostei, Characidae) and of its conspecific epigeal ancestor during early ontogeny. *Cell Tissue Res* 273:183–92.
- Langille BL, Tierney SM, Austin AD, Humphreys WF, Cooper SJB. 2018. How blind are they? Phototactic responses in stygobiont diving beetles (Coleoptera: Dytiscidae) from calcrete aquifers of Western Australia: Negative phototaxis in subterranean diving beetles. *Austral Entomol*.
- Le SQ, Gascuel O. 2008. An Improved General Amino Acid Replacement Matrix. *Mol Biol Evol* 25:1307–20.
- Li H, Cooper R. 2001. Spatial familiarity in the blind cave crayfish, *Orconectes australis packardii*. *Crustaceana* 74:417–33.
- Li H, Cooper RL. 2002. The effect of ambient light on blind cave crayfish: social interactions. *J Crustac Biol* 22:449–58.

- Matsumoto T, Ishibashi Y. 2016. Sequence analysis and expression patterns of opsin genes in the longtooth grouper *Epinephelus bruneus*. *Fish Sci* 82:17–27.
- Mejía-Ortíz LM, Hartnoll R. 2006. A new use for useless eyes in cave crustaceans. *Crustaceana* 79:593–600.
- Mejía-Ortíz LM, Hartnoll RG, López-Mejía M. 2006. Progressive troglomorphism of ambulatory and sensory appendages in three Mexican cave decapods. *J Nat Hist* 40:255–64.
- Mejía-Ortíz LM, López-Mejía M. 2005. ARE THERE ADAPTATION LEVELS TO CAVE LIFE IN CRAYFISH? *J Crustac Biol* 25:593–97.
- Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol Biol Evol* 30:1188–95.
- Mirzadegan T, Benkő G, Filipek S, Palczewski K. 2003. Sequence Analyses of G-Protein-Coupled Receptors: Similarities to Rhodopsin. *Biochemistry (Mosc)* 42:2759–67.
- Nathans J. 1987. Molecular biology of visual pigments. *Annu Rev Neurosci* 10:163–94.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* 32:268–74.
- Niemiller ML, Fitzpatrick BM, Shah P, Schmitz L, Near TJ. 2013. EVIDENCE FOR REPEATED LOSS OF SELECTIVE CONSTRAINT IN RHODOPSIN OF AMBLYOPSID CAVEFISHES (TELEOSTEI: AMBLYOPSIDAE): LOSS OF FUNCTION IN RHODOPSIN OF CAVEFISHES. *Evolution* 67:732–48.
- Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Trong IL, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. 2000. Crystal Structure of Rhodopsin: A G Protein–Coupled Receptor. *Science* 289:739–45.
- Pei J, Grishin NV. 2014. PROMALS3D: Multiple Protein Sequence Alignment Enhanced with Evolutionary and Three-Dimensional Structural Information. *Mult Seq Alignment Methods* 1079:263–71.
- Pérez-Moreno JL, Balázs G, Wilkins B, Herczeg G, Bracken-Grissom HD. 2017. The role of isolation on contrasting phylogeographic patterns in two cave crustaceans. *BMC Evol Biol* 17.
- Pérez-Moreno JL, DeLeo D, Palero F, Bracken-Grissom HD. 2018. Phylogenetic annotation and genomic architecture of opsin genes in Crustacea. *Hydrobiologia* Accepted.

- Pérez-Moreno JL, Iliffe TM, Bracken-Grissom HD. 2016. Life in the Underworld: Anchialine cave biology in the era of speleogenomics. *Int J Speleol* 45:149–70.
- Porter ML, Blasic JR, Bok MJ, Cameron EG, Pringle T, Cronin TW, Robinson PR. 2012. Shedding new light on opsin evolution. *Proc R Soc B Biol Sci* 279:3–14.
- Porter ML, Crandall KA. 2003. Lost along the way: the significance of evolution in reverse. *Trends Ecol Evol* 18:541–47.
- Porter ML, Cronin TW, McClellan DA, Crandall KA. 2007. Molecular Characterization of Crustacean Visual Pigments and the Evolution of Pancrustacean Opsins. *Mol Biol Evol* 24:253–68.
- Protas M, Conrad M, Gross JB, Tabin C, Borowsky R. 2007. Regressive evolution in the Mexican cave tetra, *Astyanax mexicanus*. *Curr Biol* 17:452–54.
- Protas ME, Trontelj P, Patel NHN. 2011. Genetic basis of eye and pigment loss in the cave crustacean, *Asellus aquaticus*. *Proc Natl Acad Sci U S A* 108:5702–7.
- Rétaux S, Casane D. 2013. Evolution of eye development in the darkness of caves: adaptation, drift, or both? *EvoDevo* 4:26.
- Salcedo E, Farrell DM, Zheng L, Phistry M, Bagg EE, Britt SG. 2009. The Green-absorbing *Drosophila* Rh6 Visual Pigment Contains a Blue-shifting Amino Acid Substitution That Is Conserved in Vertebrates. *J Biol Chem* 284:5717–22.
- Salcedo E, Zheng L, Phistry M, Bagg EE, Britt SG. 2003. Molecular basis for ultraviolet vision in invertebrates. *J Neurosci* 23:10873–78.
- Shichida Y, Matsuyama T. 2009. Evolution of opsins and phototransduction. *Philos Trans R Soc B Biol Sci* 364:2881–95.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–12.
- Sket B. 1985. Why all cave animals do not look alike--a discussion on adaptive value of reduction processes. *NSS Bull* 47:78–85.
- Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res* 26:1134–44.
- Song L, Florea L. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* 4.

- Soubrier J, Steel M, Lee MSY, Der Sarkissian C, Guindon S, Ho SYW, Cooper A. 2012. The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates. *Mol Biol Evol* 29:3345–58.
- Speiser DI, Pankey M, Zaharoff AK, Battelle B a, Bracken-Grissom HD, Breinholt JW, Bybee SM, Cronin TW, Garm A, Lindgren AR, Patel NH, Porter ML, Protas ME, Rivera AS, Serb JM, Zigler KS, Crandall K a, Oakley TH. 2014. Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinformatics* 15:350–350.
- Stahl BA, Gross JB, Speiser DI, Oakley TH, Patel NH, Gould DB, Protas ME. 2015. A Transcriptomic Analysis of Cave, Surface, and Hybrid Isopod Crustaceans of the Species *Asellus aquaticus*. *PloS One* 10:e0140484.
- Stegner ME, Stemme T, Iliffe TM, Richter S, Wirkner CS. 2015. The brain in three crustaceans from cavernous darkness. *BMC Neurosci* 16.
- Takahashi Y, Ebrey TG. 2003. Molecular Basis of Spectral Tuning in the Newt Short Wavelength Sensitive Visual Pigment. *Biochemistry (Mosc)* 42:6025–34.
- Terakita A. 2005. The opsins. *Genome Biol* 6:213.
- Tierney SM, Cooper SJB, Saint KM, Bertozzi T, Hyde J, Humphreys WF, Austin AD. 2015. Opsin transcripts of predatory diving beetles: a comparison of surface and subterranean photic niches. *R Soc Open Sci* 2:140386–140386.
- Turk S, Sket B, Sarbu S. 1996. Comparison between some epigean and hypogean populations of *Asellus aquaticus* (Crustacea: Isopoda: Asellidae). *Hydrobiologia* 337:161–70.
- Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV. 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res* 41:D358–D365.
- Wilkens H. 1986. The tempo of regressive evolution - studies of the eye reduction in stygobiont fishes and decapod crustaceans of the gulf-coast and West Atlantic region. *Stygologia* 2:130–43.
- Yamamoto Y, Jeffery WR. 2000. Central role for the lens in cave fish eye degeneration. *Science* 289:631–33.
- Yokoyama S, Meany A, Wilkens H, Yokoyama R. 1995. Initial mutational steps toward loss of opsin gene function in cavefish. *Mol Biol Evol* 12:527–32.
- Zhou X, Shen X-X, Hittinger CT, Rokas A. 2017. Evaluating Fast Maximum Likelihood-Based Phylogenetic Programs Using Empirical Phylogenomic Data Sets. *bioRxiv* 142323.

Tables

Table 1. Summary statistics for the *de novo* transcriptome assemblies produced as part of this study.

Metric	<i>Asellus aquaticus</i>			<i>Niphargus hrabei</i>		
	Molnár János Cave	Malom Lake	Soroksár (Danube River)	Molnár János Cave	Malom Lake	Soroksár (Danube River)
Number of sequences/contigs	123,735	86,903	98,343	112,658	107,451	134,592
Longest sequence/contig (bp)	26,685	23,685	29,698	16,763	13,022	24,002
Number of bases	125,035,326	93,533,167	92,258,922	98,447,524	94,883,864	118,581,253
Mean transcript/contig length (bp)	1,011	1076	938	874	883	881
Number of transcripts/contigs >1,000 bp long	34,381	27,588	26,257	27,492	27,582	32,435
Number of transcripts/contigs >10,000 bp long	272	122	117	66	17	157
Number of transcripts with ORFs	30,264	25,562	26,989	20,669	21,547	27,159
Mean ORF percent	50.42	53.29	55.98	46.08	46.32	48.53
N50	2,003	2,022	1,737	1,624	1,598	1,648
N30	3,410	3,280	2,932	2,776	2,669	2,870
N10	6,398	5,868	5,552	5,180	4,699	5,626
GC content	0.35	0.36	0.37	0.41	0.41	0.42

Table 2. Results of transcriptome completeness assessment by Benchmarking Universal Single-Copy Orthologs (BUSCO) using OrthoDB's Arthropoda database of orthologous genes.

Species	Transcriptome	Complete BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCOs Searched
<i>Asellus aquaticus</i>	Molnár János Cave	93.6%	3.7%	2.7%	1066
	Malom Lake	92.8%	3.3%	3.9%	
	Soroksár (Danube River)	90.1%	6.2%	3.7%	
<i>Niphargus hrabei</i>	Molnár János Cave	82%	7.5%	10.5%	
	Malom Lake	84.1%	7.9%	8.9%	
	Soroksár (Danube River)	90.5%	5.3%	4.2%	

Table 3. Phototransduction pathway genes identified by PIA (Speiser et al. 2014) in the cave transcriptomes of *A. aquaticus* and *N. hrabei*.

Gene	<i>Asellus aquaticus</i> Molnár János Cave	<i>Niphargus hrabei</i> Molnár János Cave
Arr		
DAGK		
GPRK1		
GPRK2		
Gq_alpha		
Gq_beta		
Gq_gamma		
r_opsin		
PKC		

PLC	
rdgB	
rdgC	
trp	

Table 4. Amino acids present in sites known to influence spectral sensitivity for each identified opsin. Position numbers depicted refer to the equivalent sites in the reference rhodopsin sequences (the bovine *Bos taurus* and the squid *Todarodes pacificus*).

Species / Opsin Type	Population	Bovine G ⁹⁰	Bovine E ¹¹³	Bovine T ¹¹⁸	Bovine E ¹²²	Bovine W ¹²⁶	Bovine A ¹⁶⁴	Bovine H ²¹¹	Bovine A ²⁶⁹	Bovine A ²⁹²	Bovine A ²⁹⁵
		Squid G	Squid Y	Squid G	Squid F	Squid M	Squid S	Squid G	Squid A	Squid V	Squid A
<i>Asellus</i> <i>aquaticus</i> SWS/UV	Molnár János Cave Malom Lake Soroksár (Danube)	K	Y	A	P	V	P	V	A	A	C
<i>Asellus</i> <i>aquaticus</i> LWS	Molnár János Cave Malom Lake Soroksár (Danube)	M	Y	S	C	Y	S	V	C	S	A
<i>Niphargus</i> <i>hrabei</i> LWS	Molnár János Cave Malom Lake Soroksár (Danube)	M	Y	S	C	W	S	Y	L	S	A

Figure Captions

Figure 1. *Asellus aquaticus* is troglomorphic only within the cave environment, while *Niphargus hrabei*'s morphotype remains constant across environments (adapted from Pérez-Moreno et al. 2017).

Figure 2. Specimens of both species were sampled from three localities within Budapest, Hungary. Red circles indicate exact sites within Molnár János Cave (Rákos Rock) and surface environments (Malom Lake and Soroksár [Danube River]; adapted from Pérez-Moreno et al. 2017).

Figure 3. Circular phylogeny of identified and reference opsins illustrating major clades (denoted in numerals) that correspond to their functional classification.

Figure 4. Expanded view of the Short Wavelength Sensitive & UV opsin clade. All populations of *A. aquaticus* expressed the same SWS/UV opsin regardless of their morphotype. In the case of *Niphargus hrabei*, SWS/UV opsins were not found in their transcriptomes.

Figure 5. Expanded view of the Long Wavelength Sensitive opsin clade. Both *A. aquaticus* and *N. hrabei* express functional opsins belonging to this clade with low intraspecific differentiation.

Figures

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure 5.

CHAPTER VI
CONCLUSIONS

The work herein was undertaken with the objective of increasing our understanding of evolution in extreme environments, namely subterranean ecosystems, in Crustacea. I begin the dissertation by reviewing the literature on the biology of cave ecosystems and the use of modern molecular techniques for their study, I continue in the second chapter with a comparative phylogeographic study to elucidate possible mechanisms of cave colonization by two cave crustaceans, the isopod *Asellus aquaticus* and the amphipod *Niphargus hrabei*. In Chapters III and IV, I further investigate the adaptation of these organisms to their environment, by focusing on the genetic changes in pathways that allow for light-detection and vision. These studies emphasize the importance of incorporating modern molecular methodologies for the elucidation of evolutionary processes at differing scales and its usefulness for solving biospeleology's long-standing questions.

As previously mentioned, Chapter 1 resulted in a literature review on the use of next-generation sequencing methodology for cave organisms and has been published in the *International Journal of Speleology*. The review highlighted the use of current and emerging molecular techniques, e.g., next-generation sequencing, as an exceptional opportunity to answer a variety of long-standing ecological and evolutionary questions in cave environments. We argue the integration of modern molecular methodologies with traditional taxonomic and ecological methods will help elucidate the unique characteristics and evolutionary history of cave fauna, and thus the significance of their conservation in face of current and future anthropogenic threats. In Chapter 2 we reviewed previous contributions to our understanding of anchialine biodiversity and

evolution, and discussed the potential of next-generation molecular methods for future research in these fascinating systems.

The second data chapter investigated the phylogeographic patterns and divergence times of surface and cave dwelling amphipods and isopods in the Molnar Janos Cave System in Budapest, Hungary. The work was published in *BMC Evolutionary Biology* in 2017. Our project examined the phylogeographic patterns exhibited by two sympatric species of surface and cave-dwelling peracarid crustaceans (*Asellus aquaticus* and *Niphargus hrabei*), and in doing so elucidated the possible roles of isolation and exaptation in the colonisation and successful adaptation to the cave environment. Contrasting phylogeographic patterns were found between species, with *A. aquaticus* showing strong genetic differentiation between cave and surface populations and *N. hrabei* lacking any evidence of genetic structure mediated by the cave environment. Furthermore, *N. hrabei* populations showed very low levels of genetic differentiation throughout their range, which suggests the possibility of recent expansion events over the last few thousand years. We concluded isolation by cave environment, rather than distance, is likely to drive the genetic structuring observed between immediately adjacent cave and surface populations of *A. aquaticus*, a predominantly surface species with only moderate exaptations to subterranean life. For *N. hrabei*, in which populations exhibit a fully ‘cave-adapted’ (troglomorphic) phenotype, the lack of genetic structure suggests that subterranean environments do not pose a dispersal barrier for this surface-cave species.

In chapter IV, I examined the transcriptomic basis of troglomorphy's adaptive divergence across surface and cave populations of *Niphargus hrabei* and *Asellus aquaticus* using RNAseq. The work was published in several parts due to unforeseen complications with RNA extractions. Because of these complications, a side project developed which resulted in a methods paper. This work titled "RNA profile diversity across Arthropoda: Guidelines, methodological artifacts, and expected outcomes" describes the "best" methods for working with and interpreting RNA extractions. The paper was published in *Biology Methods and Protocols* in 2018. Once I optimized RNA extractions for these two species the second part of the study was to investigate the transcriptional basis behind vision loss in populations of cave-dwelling crustaceans. To do so, I employed phylogenetic and transcriptomic methods on surface and cave-adapted populations of an emerging model cave species, the isopod *Asellus aquaticus* and the amphipod *Niphargus hrabei*. Results illustrated that these two species show contrasting directionality in the surface-cave transition which positioned them as ideal study subjects; *A. aquaticus* is common in surface waters and is only occasionally found in caves, whereas *N. hrabei* has successfully colonized surface environments despite belonging to an almost exclusively cave-dwelling genus. By sequencing and assembling robust *de novo* transcriptomes we characterized differences in visual genes and pathways among surface and cave populations of the aforementioned species. Our results indicated that despite having reduced eyes, recent cave colonizer *A. aquaticus* is still capable of expressing functional visual opsins and other components of the phototransduction pathway within the cave. *Niphargus hrabei*, a species with an ancient cave origin, showed no indication of being capable of sight with an incomplete phototransduction

pathway in all populations. However, the expression of functional visual opsins was maintained in this species for a yet unknown function. With the present study, we suggested the Molnár János cave system is a promising research avenue to improve our understanding of patterns of reduction and loss of vision in caves and other aphotic environments. This work was published in *Integrative and Comparative Biology* in 2018.

VITA

JORGE LUIS PEREZ MORENO

Born, Mérida, Yucatán, México

- 2011 B.MarSt. (Hons.), Marine Biology and Ecology
The University of Queensland
St. Lucia, Queensland, Australia
- 2017 M.S., Biology
Florida International University
Miami, Florida, U.S.A.
- 2020 Doctoral Candidate
Florida International University
Miami, Florida, U.S.A.

PUBLICATIONS AND PRESENTATIONS

- 2019 Pârvulescu L, Pérez-Moreno JL, Panaiotu C, Drăguț L, Schrimpf A, Zaharia C, Weiperth A, Gál B, Popovici ID, Schubart CD, Bracken-Grissom HD. *A journey on plate tectonics sheds light on European crayfish phylogeography*. Ecology and evolution, DOI: 10.1002/ece3.48
- 2019 DeLeo DM, Pérez-Moreno JL, Vázquez-Miranda H, Bracken-Grissom HD. *RNA profile diversity across arthropoda: guidelines, methodological artifacts, and expected outcomes*. Biology Methods and Protocols, 3(1):1-10.
- 2018 Pérez-Moreno JL, Balázs G, Bracken-Grissom HD. *Transcriptomic insights into the loss of vision in Molnár János Cave crustaceans*. Integrative and Comparative Biology, 58(3):452-464.
- 2018 Pérez-Moreno JL, Balázs G, Bracken-Grissom HD. *Transcriptomic analyses illuminate the molecular basis of troglomorphy in the Molnár János cave system*. International Crustacean Congress IX. May 22nd – 25th, Washington, DC, USA. *Invited oral presentation*.
- 2018 Pérez-Moreno JL, Balázs G, Bracken-Grissom HD. *Transcriptomic and epigenetic insights into the evolution of vision loss in cave-dwelling crustaceans*. Society for Integrative and Comparative Biology 2018 Annual Meeting. January 3rd – 8th, San Francisco, California, USA. *Invited oral presentation*.
- 2018 Pérez-Moreno JL, DeLeo D, Palero F, Bracken-Grissom HD. *Phylogenetic annotation and genomic architecture of opsin genes in Crustacea*. Hydrobiologia, 825(1): 159-175.

- 2017 DeLeo D, Pérez-Moreno JL, Vazquez-Miranda H, Bracken-Grissom HD. *Diversification of crustacean visual opsins revealed using phylotranscriptomics*. Evolution 2017. June 23rd -27th, Portland, Oregon, USA.
- 2017 Glenny J, Pérez-Moreno JL, Bracken-Grissom HD. *Metagenomic analyses reveal the microbiome of crustaceans from oil-affected waters*. 19th annual Biomedical and Comparative Immunology (BCI) Symposium. Miami, FL, USA.
- 2017 Pérez-Moreno JL, DeLeo D, Vazquez-Miranda H, Bracken-Grissom HD. *Phylotranscriptomic analyses illuminate the evolution and diversity of visual opsins in Crustacea*. The Crustacean Society Mid-Year Meeting 2017. June 19th - 22th, Barcelona, Spain. *Invited oral presentation*.
- 2017 Pérez-Moreno JL, Balázs G, Wilkins B, Herczeg G, Bracken-Grissom HD. *The role of isolation on contrasting phylogeographic patterns in two cave crustaceans*. BMC Evolutionary Biology, 17:247.
- 2017 Pérez-Moreno JL, Balázs G, Wilkins B, Herczeg G, Bracken-Grissom HD. *Contrasting phylogeographic patterns in two sympatric species of surface and cave-dwelling crustaceans (Crustacea: Peracarida)*. 2017 Florida International University Biosymposium. North Miami, FL, USA. February 4th. *Poster presentation*.
- 2017 Wilkins B, Golightly C, Morgan A, Ramos L, Varela C, Pérez-Moreno JL, Bracken-Grissom HD. *DEEPEND: DNA barcoding of deep-sea crustaceans in the Gulf of Mexico*. Gulf of Mexico Oil Spill & Ecosystem Science Conference. New Orleans, Louisiana, USA.
- 2016 Bracken-Grissom HD, Pérez-Moreno JL, Vázquez-Miranda H. *The evolution of bioluminescence and light detection in deep-sea decapods*. Society for Molecular Biology and Evolution Conference. Gold Coast, Queensland, Australia.
- 2016 Paterson AT, Iliffe TM, Bracken-Grissom HD, Pérez-Moreno JL, Porter M, Gonzalez BC, Engel AS. *Niche bacterial and archaeal community compositions as indicators of ecosystem processes and health in Bahamian and Mexican anchialine caves*. The 23rd International Conference on Subterranean Biology. Fayetteville, Arkansas, USA.
- 2016 Pérez-Moreno JL, Iliffe TM, Bracken-Grissom HD. *Life in the Underworld: Anchialine cave biology in the era of speleogenomics*. International Journal of Speleology, 45(2): 149-170.
- 2015 Wong JM, Pérez-Moreno JL, Chan TY, Frank TM, Felder DL, Bracken-Grissom HD. *Phylogenetic and transcriptomic analyses reveal the evolution of bioluminescence and vision in marine deep-sea shrimp (Decapoda: Oplophoridae)*. Molecular Phylogenetics and Evolution, 83: 278-292.