The genomics of phenotypically differentiated Asellus aquaticus cave, surface stream and lake ecotypes

Article *in* Molecular Ecology · May 2021

DOI: 10.1111/mec.15987

CITATION
1

READS
81

7 authors, including:

Vid Bakovic
Biology Centre CAS
13 PUBLICATIONS **46** CITATIONS
SEE PROFILE

Maria Luisa Martin Cerezo
Linköping University
9 PUBLICATIONS **34** CITATIONS
SEE PROFILE

Andrey Höglund
Linköping University
35 PUBLICATIONS **170** CITATIONS
SEE PROFILE

Jesper Fogelholm
Linköping University
31 PUBLICATIONS **120** CITATIONS
SEE PROFILE

Received: 3 March 2021 | Revised: 30 April 2021 | Accepted: 4 May 2021

DOI: 10.1111/mec.15987

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

The genomics of phenotypically differentiated *Asellus aquaticus* cave, surface stream and lake ecotypes

Vid Bakovic | Maria Luisa Martin Cerezo | Andrey Höglund | Jesper Fogelholm | Rie Henriksen | Anders Hargeby | Dominic Wright

IFM Biology, University of Linköping, Linköping, Sweden

Correspondence
Dominic Wright and Vid Bakovic, IFM Biology, University of Linköping, Linköping, Sweden.
Emails: dominic.wright@liu.se (DW); vidbakovic@gmail.com (VB)

Funding information
Science for Life Laboratory; SNIC/Uppsala Multidisciplinary Centre; LiU Neuro Framework

Abstract

Organisms well suited for the study of ecotype formation have wide distribution ranges, where they adapt to multiple drastically different habitats repeatedly over space and time. Here we study such ecotypes in a Crustacean model, *Asellus aquaticus*, a commonly occurring isopod found in freshwater habitats as diverse as streams, caves and lakes. Previous studies focusing on cave vs. surface ecotypes have attributed depigmentation, eye loss and prolonged antennae to several south European cave systems. Likewise, surveys across multiple Swedish lakes have identified the presence of dark-pigmented "reed" and light-pigmented "stonewort" ecotypes, which can be found within the same lake. In this study, we sequenced the first draft genome of *A. aquaticus*, and subsequently use this to map reads and call variants in surface stream, cave and two lake ecotypes. In addition, the draft genome was combined with a RADseq approach to perform a quantitative trait locus (QTL) mapping study using a laboratory bred F_2 and F_4 cave × surface intercross. We identified genomic regions associated with body pigmentation, antennae length and body size. Furthermore, we compared genome-wide differentiation between natural populations and found several genes potentially associated with these habitats. The assessment of the cave QTL regions in the light–dark comparison of lake populations suggests that the regions associated with cave adaptation are also involved with genomic differentiation in the lake ecotypes. These demonstrate how troglomorphic adaptations can be used as a model for related ecotype formation.

KEYWORDS
cave colonization, divergence with-gene-flow, ecotype formation, QTL mapping, reed lake habitat

Anders Hargeby and Dominic Wright contributed equally to this study.

1 | INTRODUCTION

Ecotype formation, wherein populations that are found in heterogeneous environments begin to differentiate in traits that are under divergent selection, is an early stage of the speciation continuum, which may further progress towards the formation of new species (Feder et al., 2012). Several well-known, but also newly emerging study systems are being used to elucidate the factors driving or constraining evolution and speciation (e.g., Filchak et al., 2000; Herman et al., 2018; Soria-Carrasco et al., 2014). Together with modern technological advances in DNA sequencing, genome-wide patterns of differentiation between populations spanning the speciation continuum and under various environments and selection regimes are being studied. It is now possible to identify candidate barrier loci (i.e., loci that putatively contribute to reproductive isolation between diverging populations) using F_{ST} scans, genome-wide association studies (GWAS) and quantitative trait loci (QTL) mapping (Comeault et al., 2014; Peichel et al., 2001; Soria-Carrasco et al., 2014). Factors that have been found to contribute to genomic patterns of differentiation include whether the diverging populations are found in sympatry (i.e., have ongoing gene flow during speciation) or allopatry (no or minimal gene flow during speciation) (Butlin et al., 2008). Furthermore, the genetic architecture of traits under divergent selection may have an important role in determining the pace of genome-wide divergence (Feder et al., 2014; Kautt et al., 2020). Thus, systems well suited for the study of adaptation, and in particular those that are repeated over multiple different ecotypes, can be vital in understanding this early precursor to speciation.

A few well-studied examples of early-stage ecological divergence driven by divergent selection stemming from contrasting environments include stickleback fish (Colosimo et al., 2005), *Astyanax* Mexican tetra fish (Borowsky & Cohen, 2013), stick insects (Soria-Carrasco et al., 2014), *Littorina* snails (Ravinet et al., 2016) and *Rhagoletis* fruit flies (Doellman et al., 2019). Natural replicates, wherein niche specific adaptation occurs repeatedly over geography, such as on islands, or in lakes and caves, provide additional power to infer environment-driven divergence. Moreover, we can question whether the same genetic pathways are associated with traits under divergent selection across geography and estimate the predictability of evolution (Nosil et al., 2018). In this regard, studies of ecotype evolution across geography suggest that a small but significant portion of the genome experiences parallel change in nature and is more pronounced in recently diverged populations, which would often share higher amounts of standing genetic variation (e.g., Morales et al., 2019; Ravinet et al., 2016; Soria-Carrasco et al., 2014). One example includes the Mexican tetra fish, *Astyanax mexicanus*, which has colonized several unrelated cave systems where individuals have undergone body depigmentation and eye degradation among several other changes in response to this drastically different environment (Protas et al., 2007). Cave animals that have ancestral surface counterparts represent excellent systems to study evolutionary adaptations to this extreme environment, especially when coupling several independent cave systems (e.g., Herman et al., 2018). However, the relevance that such cave adaptations have to other ecotypes is less well known. For example, are the alleles that are selected during cave adaptation also relevant to the adaptation of other ecotypes or do these cave adaptations represent entirely novel mutations/alleles?

Here, we study genome-wide divergence between morphologically differentiated populations of *Asellus aquaticus* (Linnaeus, 1758) on the island of Gotland in Sweden (Figures 1 and 2), and in particular between cave and surface stream ecotypes. Importantly, besides the repeated and parallel nature of cave colonizations, *A. aquaticus* is also known for adapting to various lake habitats, some of which can be found across lakes, causing repeated parallel adaptation. In two south Swedish lakes, it has been established that *A. aquaticus* shifted from reed to novel stonewort habitats, being accompanied by parallel shifts in body pigmentation (Eroukhmanoff, Hargeby, et al., 2009; Hargeby et al., 2004, 2005), morphology (Eroukhmanoff & Svensson, 2009), sexual behaviour (Eroukhmanoff, Hargeby, et al., 2009; Karlsson et al., 2010) and predator-induced behavioural differences (Harris et al., 2011). In the reed habitat, dense stands of *Phragmites australis* dominate and are usually found alongside the shore itself, where they form dark puddles with high amounts of debris (Hargeby et al., 2007). The stonewort habitat, on the other hand, is dominated by *Chara tomentosa* and can be found further away from the shore in open waters (Hargeby et al., 2004, 2005). In Lake Horsan, also located on the island of Gotland off the eastern Swedish coast, the open water habitat contains no plants and is here referred to as the stony bottom habitat. Here, *A. aquaticus* individuals live slightly buried within the light-coloured calcareous sand. It appears that these two broad lake ecotypes, even if separated by small spatial scales, are subject to divergent selection stemming from the environment and probably exhibit reduced gene flow (Eroukhmanoff et al., 2011; Hargeby et al., 2005; Harris et al., 2011; Karlsson et al., 2010). In addition, a strong correlation has been found between *A. aquaticus* body pigmentation and background substrate colour in 29 Swedish lakes, supporting a role for cryptic pigmentation (Hargeby et al., 2005). Eroukhmanoff, Hrageby & Svensson (2009) further illustrated an additive genetic basis for the differentiated traits between these two ecotypes through heritability estimates in common garden settings.

Troglomorphic adaptations in *A. aquaticus* have independently occurred over multiple cave systems across Europe. The most studied and perhaps best adapted cave populations are those of the Postojna-Planina Cave system in Slovenia, where individuals exhibit elongated antennae, body depigmentation and the degradation of eyes among multiple other troglomorphic characteristics (Fišer et al., 2019; Jemec et al., 2017; Konec et al., 2015). Similar troglomorphic traits can also be found in *A. aquaticus* populations from Movile Cave in Romania (Turk et al., 1996), Molnar Janos Cave in Hungary (Pérez-Moreno et al., 2018) and Trebiciano Cave in Italy (Sket, 1994). Several studies using genetic and gene expression data mapped putative candidate genes associated with these traits in *A. aquaticus* (Gross et al., 2020; Mojaddidi et al., 2018; Protas et al., 2011; Stahl et al., 2015). To date, the only QTL mapping study

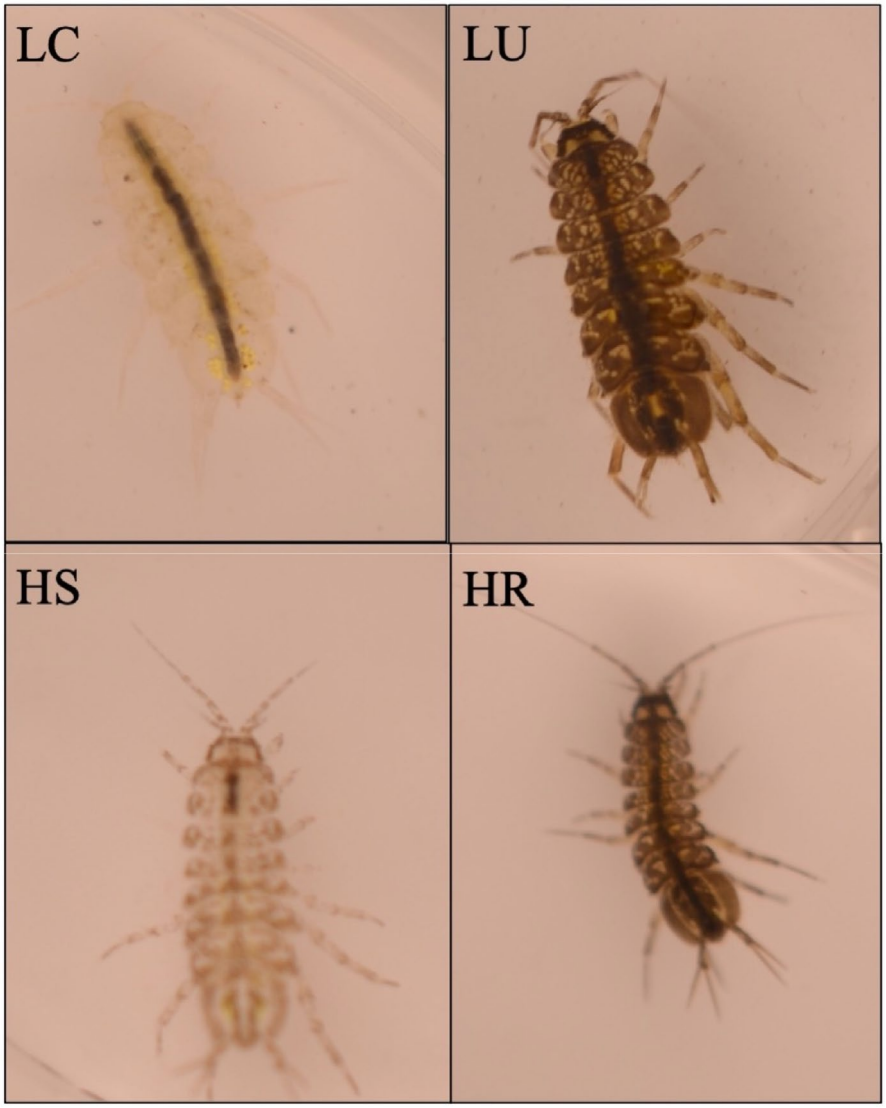
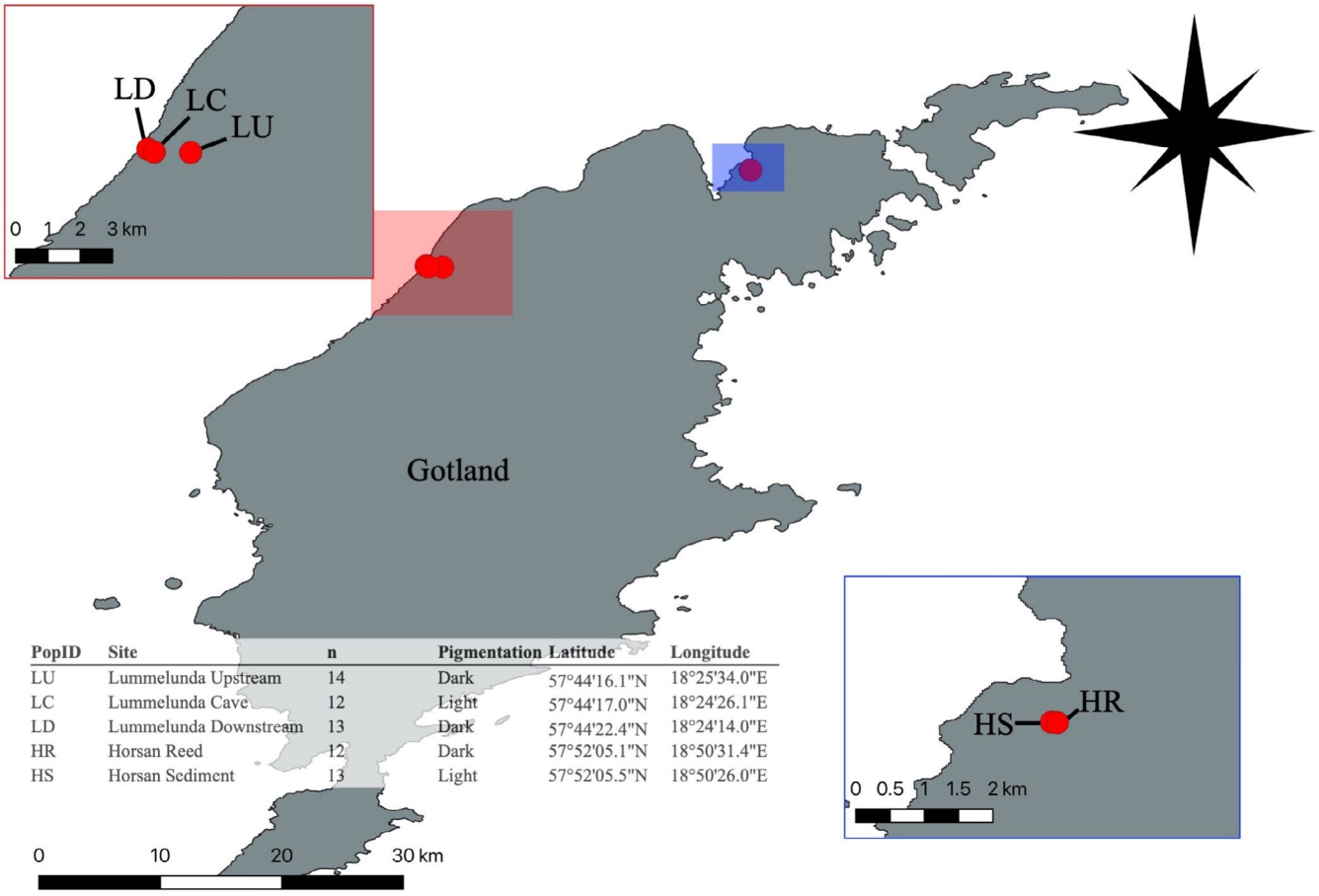


FIGURE 1 *Asellus aquaticus* individuals from four different sites on Gotland, Sweden: Lummelunda Cave (LC), Lummelunda Upstream (LU), Horsan Stony bottom (HS) and Horsan Reed (HR). Note: the black stripe along the dorsal side is not related to body pigmentation but is due to the digestive tract being full with leaf decay

using a cave vs. surface *A. aquaticus* intercross was performed by Protas et al., (2011), who identified loci that were associated with eye development and body pigmentation. However, the study used a relatively small number of molecular markers, resulting in large confidence intervals for the detected QTL, often spanning half of the linkage group where the QTL were located and meaning candidate gene identification was not possible. Further, the transcriptomes of Slovenian and Hungarian cave and surface ecotypes have been sequenced (Gross et al., 2020; Pérez-Moreno et al., 2018; Stahl et al., 2015). Interestingly, Lürig et al., (2019) showed that body pigmentation in lake populations can be plastic and is highly dependent on diet (i.e., high vs. low nutrient availability associated with dark and light morphs, respectively. These results highlight the possibility that pigmentation polymorphisms in *A. aquaticus* may be achieved through different mechanisms (i.e., genetic vs. plastic) or that plastic responses precede evolutionary responses (Levis & Pfennig, 2016).

In this work, we first generate a draft genome build for *A. aquaticus* using a cave individual from Lummelunda Cave on the island of Gotland, off the Swedish coast. This genome was subsequently used to map either whole genome (WGS) or reduced genome representation (RAD-Seq) DNA reads from cave vs. surface streams and reed vs. stony bottom lake ecotypes. We were then able to compare light- vs. dark-pigmented individuals in two entirely different and separate sites both located on the island of Gotland, that is cave vs. surface in Lummelunda and stony bottom vs. reed in Horsan Lake. This, in turn, enables us to study ecotype formation in lake and cave habitats for this system and to examine how these processes may have occurred. Genotype–phenotype associations were estimated using a QTL mapping approach in F_2 and F_4 intercross populations derived from a cave × surface intercross from animals derived from Lummelunda Cave and surface populations, and patterns of genomic variation among ecotypes were studied using comparative genomics (RADSeq) of natural populations. We then go on to assess differences between light and dark comparisons both between the two lake ecotypes, as well as between the cave/surface ecotypes. The identification of QTL in the cave/surface intercross population further allowed us to investigate the relevance such loci have for adaptation to other ecotypes as well as identify candidate genes associated with cave adaptations.



PopID	Site	n	Pigmentation	Latitude	Longitude
LU	Lummelunda Upstream	14	Dark	57°44'16.1"N	18°25'34.0"E
LC	Lummelunda Cave	12	Light	57°44'17.0"N	18°24'26.1"E
LD	Lummelunda Downstream	13	Dark	57°44'22.4"N	18°24'14.0"E
HR	Horsan Reed	12	Dark	57°52'05.1"N	18°50'31.4"E
HS	Horsan Sediment	13	Light	57°52'05.5"N	18°50'26.0"E

FIGURE 2 Map of the island of Gotland in Sweden with red dots indicating *Asellus aquaticus* sampling sites. Population identifiers, sample sizes and locality names are presented in the inset table

2 | MATERIALS AND METHODS

2.1 | Natural populations

In June 2019, we collected cave specimens of *Asellus aquaticus* from Lummelunda Cave (LC), located on Gotland, Sweden (first noted in Gislén & Brinck, 1950). This cave was probably created before the last ice age by flowing water from Martebo mire located 11 km to the east of the cave (Lundevall, 1965). The stream flows through the cave system before reaching the surface again and draining into the Baltic Sea. We collected surface specimens of *A. aquaticus* from upstream (LU) and downstream (LD) of Lummelunda Cave (Figures 1 and 2). Both the surface populations, LU and LD, contained only dark-pigmented individuals, while the LC population consisted of mainly light-pigmented individuals with a few slightly darker-pigmented samples either washed in from the surface or found at low frequencies within the cave. Two additional differentiated surface ecotypes of *A. aquaticus* were collected from contrasting reed and stony bottom habitats in Lake Horsan (HR and HS, respectively), located ~40 km north of LC (described in Hargeby et al., 2005). Here, the two ecotypes are separated by small spatial distances (~50 m), but are, nevertheless, well differentiated with regard to body pigmentation. Again, with the exception of a few individuals, HS consisted of light-pigmented animals found within the light-coloured stony bottom (light-coloured rock environment mixed with fluffy calcareous precipitated flocks of calcium carbonate but no vegetation) of the lake in open waters. HR individuals, however, were dark-pigmented and found within dense dark-coloured puddles formed within reed stands alongside the shore. Previous studies from 29 Swedish lakes have shown that *A. aquaticus* body pigmentation often matches its substrate colour (Hargeby et al., 2005). Although not studied in this lake, reed and stonewort ecotypes from two other lakes, Krankesjön and Tåkern, show compelling evidence for adaptational crypsis and reduced gene flow (Eroukhmanoff, Hargeby, et al., 2009; Eroukhmanoff, Hargeby & Svensson, 2009; Eroukhmanoff et al., 2011; Eroukhmanoff & Svensson, 2011; Hargeby et al., 2004; Hargeby et al., 2005; Harris et al., 2011; Karlsson et al., 2010; Karlsson et al., 2016). Considering reported occurrences of the main predatory fish of this species, *Perca fluviatilis*, in Lake Horsan (Nyberg et al., 2014), it may be reasonable to assume that the open water stony bottom environment would exert higher fish predatory selective pressures similarly to the stonewort habitat in other lakes and that pigmentation may be cryptic. Additional information on the populations used in this study, including coordinate data and number of individuals, can be found in Figure 2.

2.2 | Cave × surface intercross

All intercross (F_1,F_2, F_4) offspring were derived from a single parental pair, consisting of a cave male and a surface female. A total of 26 F_1 offspring, 83 F_2 and 79 F_4 were produced. Lummelunda animals from an earlier collection in 2014, including cave (LC) and surface (LU) ecotypes, were reared in the laboratory within rearing aquaria (transparent 5-L polypropene boxes) under 15°C and a 12-hr/12-hr light–dark photoperiod. The animals were kept in natural lake water, which was routinely collected from Lake Tåkern in Sweden and filtered through 100-μm nets to avoid introducing wild invertebrates (e.g., predators). Water was stored in 25-L tanks for at least 96 hr to eliminate possible chemical cues from other animals. Water was changed every 2 weeks and decaying elm (*Ulmus* sp.) and alder (*Alnus* sp.) leaves were added in excess amounts (Graça et al., 1993). Additionally, the animals were given shoots of *Elodea canadensis*, which was reared in the laboratory (Marcus et al., 1978). Besides the food providing substrate and shelter, each rearing aquarium was supplemented with two-footed bricks 10 × 10 cm in size for additional shelter and area for periphyton. The tanks were observed daily and animals found in amplexus were isolated. Once pairs of both cave and surface stream animals were noted, the male and female were carefully separated. New pairs were constructed in which a male and a female were combined in a mixed *cave + surface* setting. If the pairs within a mixed setting formed amplexus, they were transferred to a rearing aquarium. The pairs were then monitored every second day. Once the pair separated, the female was inspected for a filled egg pouch, and if this was present, the male was removed and frozen at −20°C. The female was checked at regular intervals for eggs vs. empty pouch. If young were found in the aquaria, the female was removed and stored at −80°C. The F_1 offspring were left to interbreed, and we collected F_2 (n = 83) and F_4 (n = 16) samples. This included all F_2 offspring and the most extreme light- and dark-pigmented F_4 individuals (those that were determined to be the lightest and darkest; see pigmentation phenotyping below). These F_4 samples presumably provide higher QTL detection power through two additional recombination events in the F_4 individuals, which is expected to improve the resolution of the QTL analysis (Darvasi & Soller, 1995).

2.3 | Phenotypic measurements

Prior to cold storage, individual *A. aquaticus* samples were photographed using a Canon D3100 camera. Pictures were taken of single live animals in Petri dishes filled with 10 ml of water, but also of dry dead animals (in the case of the F_4 population). In both cases, a photograph was taken within a light tent with external illumination on both sides of the tent. A white colour checker was used to standardize mean RGB values and a scale bar was used to standardize length measurements. Furthermore, all length measurements used in the QTL interval mapping were standardized to body length. RAW format files were first converted to TIFF format and the images were processed using IMAGEJ version I.x (Schneider et al., 2012). Our primary interest was in body pigmentation, as this trait clearly differentiated the ecotypes collected in this study (Figure 1). We measured pigmentation using the selection and measure tools in IMAGEJ. Mean RGB values for the left and right side of the dorsal body were measured to avoid sampling the dark-coloured digestive tract in the centre (as in Hargeby et al., 2004). We measured additional traits that did not obviously differ among the collection sites but had been shown to differ between ecotypes in other cave and lake populations (Hargeby & Erlandsson, 2006; Konec et al., 2015). These included the surface area of the animal, length of the distalmost article of the peduncle of antenna II, body length and the width of the 1st to 7th body segments (all phenotypic measurements of the cave × surface intercross can be found in Table S1 and are schematically presented in Figure S1). Therefore, our QTL analysis primarily focused on size-related traits, body pigmentation and antennae length. As opposed to the degradative nature of pigmentation and eye loss in cave environments, antennae length may be an example of a constructive change. This was shown to be the case in comparisons between Slovenian surface and cave individuals of *A. aquaticus* (Konec et al., 2015; Mojaddidi et al., 2018). The length of the distalmost article of the peduncle of antenna II was used, as the total flagellar antennae is sensitive and prone to breaking (Maruzzo et al., 2007, 2008). Finally, the sex of animals was noted where possible, by identifying the sex-specific shape of the gonopods on the ventral side of the body, under a stereomicroscope.

2.4 | DNA extraction and sequencing, genome assembly, RADseq data processing, gene annotation and contamination

A detailed description of these methods is presented in the Supporting Information. In brief, chromium linked reads generated in this study were assembled using SUPERNOVA version 2.1.1 (Weisenfeld et al., 2017) (the assembly report of this run can be found in Table S2 and the final genome assembly statistics in Table S3). BLAST screens for contamination and gene annotation outputs can be found in Tables S4–S7. For comparative purposes, multiple genome builds were conducted using the WGS data (Table S8). Finally, processing statistics of the RADSeq data, including sequence qualities and amount of reads mapping to the genome, can be found in Table S9.

2.5 | Mitochondrial *COI* phylogenies

To establish the relationship between *A. aquaticus* populations from Gotland with previously studied European populations (Konec et al., 2015), we extracted a fragment of the *COI* gene from the intercross founder male and female DNA sequences (WGS data). Further, we used a subset of samples from the Slovenian and Romanian cave and surface populations to show that the cave colonization on Gotland is probably another independent event in this species. Maximum-likelihood trees using a simple UPGMA tree as a starting point were

constructed using the Lummelunda Cave male, Lummelunda surface female and sequences retrieved from NCBI (listed in supplements of Konec et al., 2015). All trees were built using the APE, PHANGORN and PHYTOOLS packages (Paradis & Schliep, 2019; Revell, 2012) in R version 3.1.4 (R Core Team, 2014) and visualized using GGTREE (Yu, 2020). Only the male and female parents of our cave × surface intercross, from which we generated WGS data, were used for the *COI* comparison. The consensus sequences of the male and female were established by using the de novo assembly and the variant call file (VCF) with VCF-CONSENSUS (Danecek et al., 2011).

2.6 | Population genomic parameter estimates

The 1,187 single nucleotide polymorphisms (SNPs) generated from *A. aquaticus* natural populations were used to calculate population genomics summary statistics and pairwise F_{ST} values among our populations. An AMOVA F_{ST} estimate was obtained using the *populations* module in STACKS version 1.4 (Catchen et al., 2013). Using F_{ST} and F_{IS} values for each SNP within each population, 95% confidence intervals were calculated using the *groupwiseMean* function of the RCOMPANION package (Mangiafico & Mangiafico, 2017) in R version 3.1.4. As a measure of genetic variation, we calculated nucleotide diversity (π) and expected heterozygosity for each population. A one-way ANOVA was used to assess significance for nucleotide diversity differences among populations, followed by a post-hoc analysis, where populations were assigned significance groupings using Tukey's honestly significant difference (HSD) test within the AGRICOLAE R package (Mendiburu & Simon, 2015).

2.7 | Population structure

To establish the relationships among our natural populations based on the 1,187 SNPs, we conducted two separate clustering analyses. First, we ran a Bayesian implementation in STRUCTURE version 2.3.4 (Pritchard et al., 2000) to determine how many clusters are best supported by our data. An admixture model was used with 10 replicates for each *K* value (number of groups) in the range from *K* = 2 to *K* = 10, and a burn-in period of 100,000 and 100,000 Markov chain Monte Carlo (MCMC) repetitions after initial burn-in. The number of clusters was estimated using the delta *K* method (Evanno et al., 2005) implemented in STRUCTURE HARVESTER (Earl, 2012). Replicate runs using Bayesian statistics commonly produce slightly variable output (Jakobsson & Rosenberg, 2007). To utilize the 10 replicates, we aligned the output using the *LargeKGreedy* model in CLUMPP version 1.1.2 (Jakobsson & Rosenberg, 2007), with 1,000 random input orders taken before calculating mean values of membership probabilities. Graphical output was generated using DISTRUCT version 1.1 (Rosenberg, 2004).

Second, a principal component analysis (PCA) was conducted on the same data set as implemented within the R package ADEGENET using the function *dudi.pc* (Jombart & Ahmed, 2011). Allele frequencies were scaled to mean zero with the *ScaleGen* function and missing genotypes were replaced by zeros. Population sizes used for testing population structure are given in Figure 2.

2.8 | Testing A. aquaticus migration models between habitats

We looked for evidence of ongoing gene flow between the surface and cave samples by testing several migration models in the three populations sampled at Lummelunda using MIGRATE-N version 4.4.3 (Beerli & Palczewski, 2010). A further subset of SNPs was extracted from the Lummelunda populations with VCFTOOLS as follows: minGQ: 20, max missingness: 0.1, inter-SNP distances of at least 1,000 bp, which resulted in 114 higher quality and well-represented SNPs among our samples (sample sizes are given in Figure 2). This software uses Bayesian inference of the Kingman's coalescent to estimate model parameters such as effective population size and migration rate. Due to the large homogeneity in our genetic data, we did not expect to estimate accurate parameters, but rather we were interested in which general model is best supported (i.e., gene flow vs. no gene flow). We compared the following models: (1) directional migration going from LU to LC to LD, (1a) same as 1 except that LD receives migrants from both LC and LU, (2) directional migration in the opposite direction, that is from LD to LC to LU, (3) no gene flow model emphasizing that LC diverged from LU and that LD diverged from LC, (3a) same as 3 except that LD diverged from LU and not LC, (3b) including both divergence events from 3 and 3a, and (4) each population sends and receives migrants from each of the other populations with model 4a denoting the same as model 4 but with emphasized symmetrical gene flow between all populations. These models are graphically presented in Figure S2. For each model four heated chains were used with the following scheme: 1, 1.5, 3 and 1,000,000. The sampling prior increment was 20 with a burn-in of 1,000. The number of steps analysed was 10,000. Prior migration rates were generated from a uniform distribution ranging from 2,500 to 3,500, while prior thetas were generated from a uniform distribution ranging from 0 to 0.1.

2.9 | Differentiated loci among A. aquaticus ecotypes

To compare genomic differences between light- and dark-pigmented *A. aquaticus* in Horsan and Lummelunda, we computed allele frequency differences in each site separately (sample sizes in Figure 2). Significance was assigned using a randomization test by randomly shuffling population IDs and calculating allele frequency differences 10,000 times to produce a null distribution. Estimated allele frequency differences falling above the 0.975 or below the 0.025 quantile of the null distribution were considered statistically significant. Next, we correlated allele frequency differences between light and dark individuals in Lummelunda (LC vs. LU) with allele frequency

differences between light and dark individuals in Horsan (HS vs. HR) using (i) all SNPs, (ii) all SNPs with significant allele frequency differences and (iii) only SNPs with significant allele frequency differences that are shared by the two localities. We additionally computed the percentage of loci whose allele frequency differences are in the same direction in both localities (i.e., both negative and both positive). The PL (genotype probability) field from our GATK-generated VCF file was extracted with the R package VCFR version 1.12.0 (Knaus and Grünwald, 2017) and allele frequencies were computed as:

$$P = \frac{\sum_{i=1}^{n} \sum_{j=1}^{3} (gjxp(gj|Di))}{2xn}$$

where n = number of individuals, $g1{:}3$ = [0,1,2] (number of alternative allele copies in a homozygote, heterozygote and alternativee homozygote, respectively), and $P(gi|Di)$ is the probability of genotype j given the sequence data D at locus i. Genotype probabilities provide the advantage of incorporating genotype uncertainty as opposed to genotype calls (Nielsen et al., 2011).

Allele frequencies, as estimated above, were also used to compare the cave population to two neighbouring surface populations independently. Specifically, we computed allele frequency differences between LC and LU, and between LC and LD. We reasoned that shared differences between these two comparisons may be attributed to the cave habitat through a genotype–environment association.

2.10 | Genetic map build and QTL analysis

First, a genetic map using F_2 and F_4 intercross progeny was generated. After filtering for missing data, identical genotypes and loci exhibiting segregation distortion (not conforming to 1:3:1, $p < 1e^{-10}$), 506 SNPs in 67 individuals were used for the construction of the genetic map. Linkage groups (LGs) were formed based on the recombination fraction and LOD scores of markers using the *FormLinkageGroups* function in RQTL (thresholds used: min lod: 5.7, max rf: 0.35). This resulted in 43 LGs but only the first 14 were retained, while the rest of the markers were grouped into an LG named "un." These LGs contained three or fewer markers. The 14 LGs contained 435 markers and spans 3,997 centimorgans (cM) with on average 9 cM spacing in between markers (Table S10 and Figure S3).

For the QTL interval mapping, we included F_2 (n = 83, 42 males and 32 females) and F_4 (n = 16, two males and nine females) individuals. Sex and batch were included as covariates and models both with and without sex interaction were tested. F_2 and F_4 were assigned two different batch IDs to account for fixed differences that may be present between the two generations, as they were by definition reared in separate batches. Genotype–phenotype associations were quantified using LOD scores and LOD significance thresholds were established by running 1,000 permutations. QTL p values for the trait of interest <.05 were considered significant and a 1.8 LOD drop method was used to determine QTL confidence intervals. An RQTL cross-file, including genotypes and phenotypes of F_2 and F_4 progeny, is given in Table S1.

Finally, to verify if QTL regions associated with the different *A. aquaticus* traits are differentiated in the wild, we overlapped LOD scores from within the QTL intervals for each trait with F_{ST} values of these same markers between natural populations. To this end, SNPs from the genetic map were extracted and ordered within the natural populations' data set. These SNPs showed lower sample coverage in the natural populations (a minimum of six individuals per population) but were, nevertheless, used to calculate F_{ST}.

3 | RESULTS

3.1 | Mitochondrial *COI* phylogenies

Maximum-likelihood trees created using a fragment of the *COI* gene from Lummelunda with other publicly available European populations showed that the samples from Gotland clustered most with Swedish and Danish populations (Figure 3a). Although bootstrap support values for some of the external nodes were low, they tended to increase drastically within the internal nodes. As the phylogeny of these populations has been extensively studied elsewhere, we were only interested in which populations cluster closest to the Gotland populations studied here. Both our cave male and surface female *COI* sequences most closely resembled the UP3 *COI* haplotype, which was found in previous studies from Swedish samples collected near Uppsala (Figure S4). Further, when using only a subset of samples which compared cave and surface populations from Slovenia and Romania with those studied here, cave and surface populations clustered together by geography and not by habitat (Figure 3b). However, our Swedish populations clustered closer to the Romanian populations than the Slovenian ones.

3.2 | Population genomic parameter estimates

Genetic variation of *Asellus aquaticus* from Lummelunda and Horsan was estimated through expected heterozygosity, nucleotide diversity (π), inbreeding coefficient (F_{IS}) and F_{ST} (Tables 1 and 2). Genetic diversity was consistently the lowest in LC (Lummelunda Cave) and the highest in HS (Horsan Stony bottom). Mean F_{ST} values based on these 1,187 SNPs were generally low (Table 2). The two most genetically differentiated populations were LC and HS (between-site F_{ST}: 0.051), while the two least differentiated were LD and LC (within-site F_{ST}: 0.038).

3.3 | Spatial genetic population structure

Although the delta K method best supported K = 7. After sequentially increasingly K, however, we found that $K \geq 4$ was in accordance with the geographic distribution of the samples and minimum

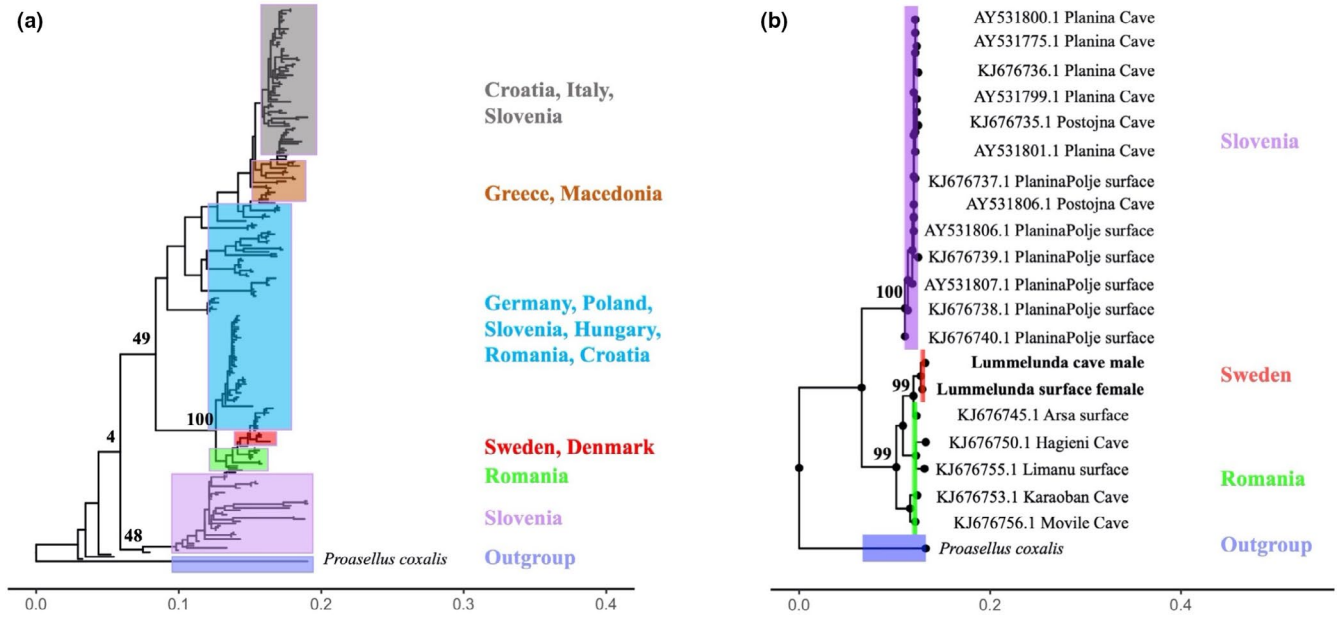


FIGURE 3 Maximum-likelihood trees based on a fragment of the *Asellus aquaticus* mitochondrial *COI* gene. (a) Comparison of populations from Gotland with previously studied European populations; (b) comparison between surface and cave populations among Slovenia, Sweden and Romania. *Proasellus coxalis* was used as the outgroup and bootstrap support percentages based on 1,000 iterations are shown for the main nodes

TABLE 1 F-statistics for *Asellus aquaticus* populations from Gotland based on 1,187 SNPs obtained using single enzyme RADSeq

PopID	Polymorphic sites	% Polymorphic loci	Obs Het	Exp Het	π	F_{IS}	95% CI Fis
HR	515	43.980	0.103^{ab}	0.105^{b}	0.112^{c}	0.043	0.0265–0.0586
HS	701	59.863	0.122^{a}	0.136^{a}	0.144^{a}	0.102	0.081–0.123
LC	432	36.892	0.099^{b}	0.096^{b}	0.101^{c}	0.027	0.0117–0.0428
LD	523	44.663	0.098^{b}	0.111^{b}	0.118^{bc}	0.079	0.0602–0.0976
LU	591	50.470	0.110^{ab}	0.129^{a}	0.136^{ab}	0.111	0.0892–0.132

Under Obs Het, Exp Het and π, letters a–c denote statistical groupings of populations with alpha = .05 using a Tukey-based approach. Population abbreviations are as in Figure 2.

TABLE 2 Pairwise F_{ST} values among *Asellus aquaticus* populations from Gotland based on 1,187 SNPs (upper diagonal) and 95% CI (lower diagonal)

	HR	HS	LC	LD	LU
HR		0.040	0.039	0.042	0.048
HS	0.0371–0.0436		0.051	0.047	0.047
LC	0.0357–0.0425	0.046–0.0561		0.038	0.050
LD	0.0385–0.0455	0.0423–0.0508	0.0349–0.0414		0.042
LU	0.0437–0.0521	0.0429–0.0509	0.0453–0.056	0.0382–0.0456	

Population abbreviations are as in Figure 2.

additional information was gained by further increasing *K* (Figure S5). In all cases, the different populations showed large homogeneity. The two Horsan populations are distinguishable from the three populations found in Lummelunda (Figure 4a). Within Lummelunda, however, LC and LD grouped closer together and were spatially closer to one another. A similar trend was found when conducting PCAs on the same data set (Figure 4b). Horsan and Lummelunda form the two main clusters, but within- and between-locality relationships were also present in this global PCA conducted on all five populations. To further resolve the relationship between populations within each site separately, additional PCAs confirmed differential clustering between HR and HS in Horsan, and between LC, LD and LU in Lummelunda (Figure S6). Depending on the PC axes being used, we find support for both the clustering of LD–LC and LU–LD, that is the two spatially closer sites and the two surface sites, respectively.

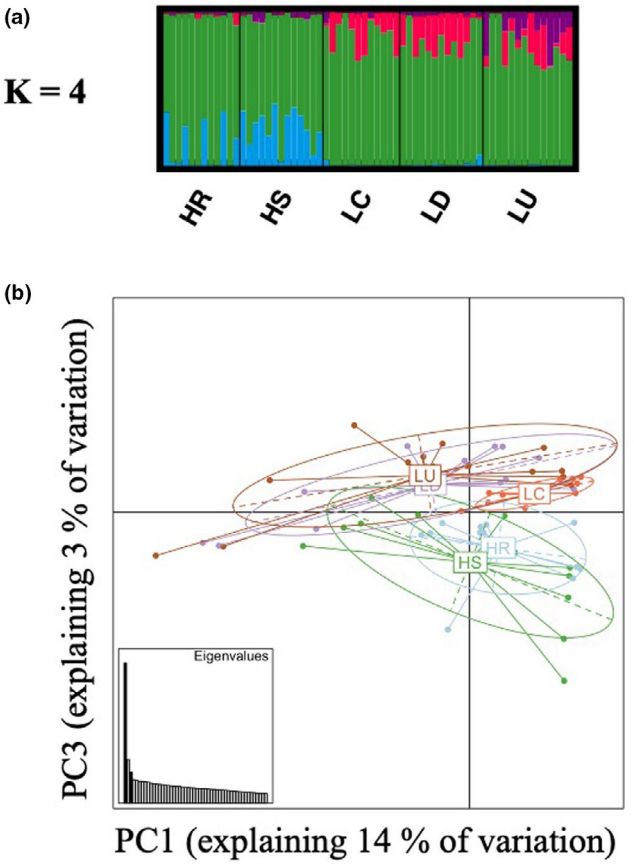


FIGURE 4 The spatial genetic population structure of *Asellus aquaticus* samples collected on the island of Gotland based on 1,187 SNPs. (a) STRUCTURE analysis showing $K = 4$ (five sampling sites grouped into four clusters); (b) PCA scatter plot of five sampling sites; population abbreviations are as in Figure 2

3.4 | Migration models

The comparison of migration models revealed that the full gene flow models 4 and 4a best fit our data (Table 3). Interestingly, the second best model is model 1a, directional gene flow from LU to both LC and LD, which makes the most sense given the direction in which these streams flow. The "no gene flow" models, 3, 3a and 3b, showed the weakest support.

3.5 | Ecotype formation within two separate locations

We were interested in how genetic differences between light and dark populations compared between Lummelunda and Horsan. Individuals from HS and LC are both light-pigmented, but they live in drastically different environments, with biotic and abiotic selective pressures probably differing largely between the cave and open water lake environments. The correlation between allele frequency differences between light- and dark-pigmented *A. aquaticus* from Horsan and Lummelunda was strongly negative when using all 1,187 SNPs ($r = -.42$, $p = 2.2e\text{-}16$) (Figure 5a). Allele frequency differences for 34% of the SNPs went in the same direction in both locations. Two additional subsets of these SNPs were used to evaluate allele frequency difference correlations and direction: (1) all SNPs that were significantly different in one and the other location (Figure 5b), and (2) all SNPs that were significantly different in both locations only (Figure 5c). The correlation was strongest when using subset 2 ($r = -.79$, $p = 2.3e\text{-}15$), which also resulted in the fewest number of SNPs exhibiting allele frequency differences in the same direction at both localities (9%). In total, 128, 255 and 66 SNPs exhibiting significant allele frequency differences between light and dark populations were found in Horsan, Lummelunda and in both localities, respectively. Gene and scaffold information can be found in Table S11.

In an attempt to associate genotype to environment, we additionally compared allele frequency differences between LC–LU and LC–LD within Lummelunda, with the expectation that shared differences may reflect genomic regions associated with the cave environment. When correlating LC–LU with LC–LD allele frequency differences, we found a strong correlation using all three SNP sets; all SNPs ($r = .7$, $p = 2.2e\text{-}16$), all significant SNPs ($r = .71$, $p = 2.2e\text{-}16$) and only shared significant SNPs ($r = .89$, $p = 2.2e\text{-}16$) (Figure 5d–f). Percentages of these differences going in the same direction were 82%, 87% and 100%, respectively. In total, 321, 286 and 187 SNPs exhibiting significant allele frequency differences between light and dark Lummelunda populations were found in LC–LU, LC–LD and in both comparisons, respectively. This highlights the presence of a shared genetic component between the two surface populations which is not shared in the cave population. Gene and scaffold information can be found in Table S12.

TABLE 3 *Asellus aquaticus* MIGRATE-N results obtained using 114 SNPs found in three Lummelunda collecting sites: 1, LC; 2, LD; 3, LU

Model	Model parameters	Log (ml)	LBF	Model-ranking
1	*0* **0 00*	1,562.58	−388.15	4
1a	*0* *** 00*	1,703.30	−247.43	3
2	**0 0*0 *0*	1,462.35	−488.38	5
3	*0d d*0 00*	257.18	−1,693.55	8
3a	*0d 0*d 00*	298.19	−1,652.54	6
3b	*0d d*d 00*	263.19	−1,687.54	7
4	*** *** ***	1,950.73	0.00	1
4a	*ss s*s ss*	1,751.30	−199.43	2

Model parameters: as used to run the software; Log (ml): marginal likelihoods; LBF: natural log Bayes factors.

3.6 | QTL interval mapping

We examined *A. aquaticus* trait–genotype associations using a non-parametric interval mapping approach in RQTL. The interval mapping

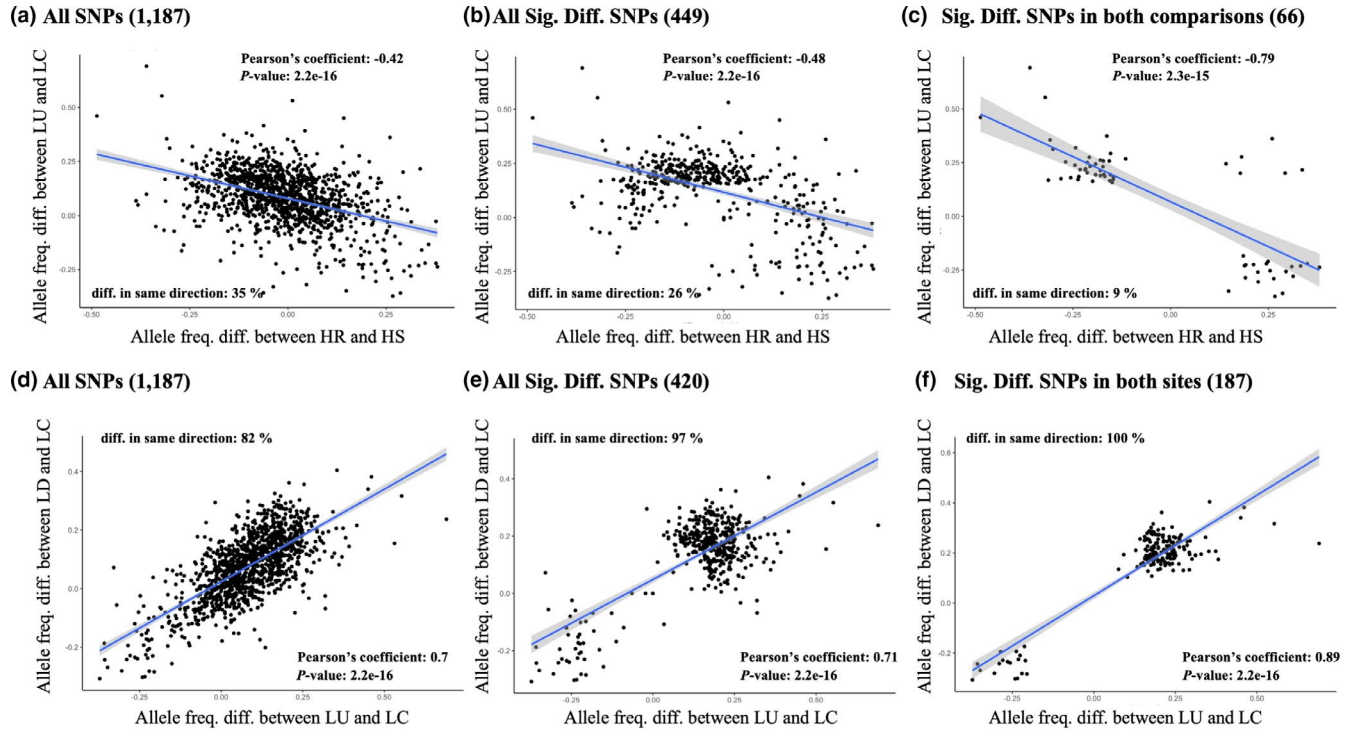


FIGURE 5 Corelations of allele frequency differences between light- and dark-pigmented populations of *Asellus aquaticus* in Horsan (HR–HS) and Lummelunda (LU–LC). (a–c) Comparisons between allele frequency differences in Horsan and Lummelunda; (d–f) comparisons between allele frequency differences between LC–LU and LC–LD. Comparisons were made using all SNPs (a, d), SNPs with significantly different allele frequencies in total (b, e), and SNPs with significantly different allele frequencies but in both comparisons only (c, f). Shaded grey area represents the standard error of a linear regression

was carried out on a cave (LC) × surface (LU) intercross only. In total, we observed 10 significant QTL for the various traits. Our main focus was on traits that were differentiated among the ecotypes but also on traits that are potentially under different regimes of selection in these environments, such as antennae length in cave and surface environments. The two strongest QTL found were associated with body pigmentation and surface area. Pigmentation QTL were located on linkage groups LG1, LG2 and LG7 (Table 4). The marker found on LG2 was significantly associated with mean white-standardized RGB values with and without sex interaction, explaining 37% of the phenotypic variation (Table 4; Table S13). Furthermore, one QTL associated with relative antennae II segment length was found on LG11, explaining 35% of the phenotypic variation (Table 4; Table S13). In total, we identified 20 genes within the confidence intervals of our QTL. Of these, three were associated with body pigmentation, two with antennae length and six with size-related traits (Table 4).

3.7 | Differentiation of QTL regions in wild populations

In comparing LOD scores with F_{ST} values within QTL confidence intervals, we found that regions associated with most of the analysed traits were also differentiated in natural populations (at least two-fold higher than the average across the entire genetic map; average F_{ST} in Horsan: 0.08, average F_{ST} in Lummelunda: 0.14). Our results indicate that QTL regions associated with body pigmentation, antennae length and a part of size-related traits showed peaks of differentiation in both Lummelunda and Horsan (Figure 6). QTL regions on LG2 and LG5 that were found associated with surface area, however, showed differentiation peaks only in Lummelunda. A few traits from Table 4 are not present in Figure 6 either due to having a small number of markers or high amounts of missing F_{ST} data in natural populations.

4 | DISCUSSION

Asellus aquaticus has evolved distinct ecotypes in numerous heterogeneous environments and over small spatial scales (Hargeby et al., 2005; Protas et al., 2011). We studied genome-wide differences among ecotypes collected from surface stream and cave habitats in Lummelunda, and from stony bottom and reed habitats in Lake Horsan. Population structure analyses indicated that our collection sites formed a relatively homogeneous cluster with embedded signals of spatial genetic structure. Furthermore, several genes related to traits previously shown to be differentiated between these ecotypes in other caves and lakes were found using a QTL mapping approach and comparing allele frequencies among the studied populations. In particular, we found that body pigmentation and antennae length QTL were differentiated between the ecotypes in both Lummelunda and Horsan, indicating that these loci appear to at least partly regulate ecotype formation in these populations. Allele

TABLE 4 QTL interval mapping results showing significant QTL for *Asellus aquaticus* phenotypic traits

Trait	Chr	Pos	LOD	r2	Lower CI	Upper CI	Interaction	Genes
Surface area	2	783	4.3	7.7	774	807	Sex	NA
Surface area	5	66	15	36.4	64	69	5@66.0:11@2.0, 5@66.0:1@625.0	g001010(VCP_APIME), g001020, g001030, g001040(Interleukin2)
Surface area	11	2	9.5	19.6	0	13	5@66.0:11@2.0	g002700, g002150(bloc1_2)
Surface area	1	625	10.1	21	620	629	5@66.0:1@625.0	g002520, g000720, g000730(NOP58_MOUSE), g000740
Antennae II	4	18	4.7	17.3	5	35	4@18.0:11@31.4	g001140(PF04106.7)
Antennae II	11	31	8.6	35.6	26	37	4@18.0:11@31.4	g002150(PF10046.4), g000650(SALM_DROME)
Pigmentation	1	616	7.8	15.6	605	622	1@616.1:2@119.0	g002520, g000720, g000730(NOP58_MOUSE), g000740
Pigmentation	2	119	11.9	26.9	115	124	1@616.1:2@119.0	g001880, g001890,
Pigmentation	7	81	4.8	8.9	73	89	Sex	NA
Segments 1–7	4	249	3.7	23.6	234	258	None	g001880, g001890, g002020

All length and area measurements were standardized to full body length, while mean RGB values were standardized to a white colour checker. Chr, chromosome; Pos, position; LOD, logarithm of odds indicating strength of association; r2, total phenotypic variation explained; Lower CI, start position of confidence interval; Upper CI, end position of confidence interval; Interaction, markers and/or variables with which the QTL is interacting.

frequency differences between light- and dark-pigmented populations in Lummelunda and Horsan add further evidence that the genetic basis of ecotype formation potentially occurs in the same regions between both ecotype pairs.

4.1 | Genetic variation in *Asellus aquaticus* on the island of Gotland

Genetic diversity and divergence among the *A. aquaticus* populations on Gotland were generally weak, as identified using common population structure approaches, as well as F_{ST} and nucleotide diversity comparisons. The F_{ST} values between cave and surface individuals obtained in this study (LC–LU F_{ST}: 0.05) fit the notion of early stages of divergence with ongoing gene flow well. The striped vs. unstriped ecotypes of the famous *Timema* stick insects show median F_{ST} measures of 0.007–0.015 using WGS data (Soria-Carrasco et al., 2014). Similarly, multiple stream vs. lake stickleback comparisons using two subsets of markers (one under selection and the other neutral) showed F_{ST} values ranging from 0.045 to 0.192 for neutral markers and 0.071 to 0.395 for selected markers (Kaeuffer et al., 2012). Our test of migration models further confirms the connected nature of cave and surface populations in Lummelunda. Gene flow with surface counterparts but also with other cave populations presumably facilitated the adaptation to extreme environments such as caves in *Astyanax* cave fish (Herman et al., 2018) and may play an equally important role in *A. aquaticus*.

Fuller et al., (2019) compared *A. aquaticus* populations from multiple lakes near Chernobyl using RADSeq data, with the results presented here being comparable, in that genetic differentiation of populations from different locations was lower when found within proximity (F_{ST} = 0.085) but increased with distance (F_{ST} > 0.2) (Fuller et al., 2019). In Lummelunda, we found that the two spatially closer locations, downstream (LD) and cave (LC), are more genetically similar than cave (LC) and upstream (LU). Similarly, our data are in line with low genetic differentiation identified between ecotypes found within the same lakes (F_{ST} ≈ 0.003–0.03 based on AFLP [amplified fragment length polymorphism] markers; Eroukhmanoff, Hargeby, et al., 2009; Eroukhmanoff, Hrageby & Svensson, 2009), but cannot explain the low differentiation we find between localities in Horsan and Lummelunda. Low differentiation among populations on the island of Gotland (F_{ST} range: 0.038–0.051) may be the consequence of several events: (i) a recent and/or single introduction onto the island; or (ii) alternative force homogenizing the genomes of these populations may be admixture. While the three collection sites in Lummelunda seem to be connected, with surface streams (LU) flowing into the cave (LC) and back out again (LD), it is less clear what would facilitate gene flow between Horsan and Lummelunda.

While some ongoing gene flow probably exists between light and dark populations found in Lummelunda and Horsan, it is either restricted or the selective pressures differentiating the ecotypes are strong. In support of restricted gene flow, we detected overall spatial genetic population structure, albeit with F_{ST} values among the

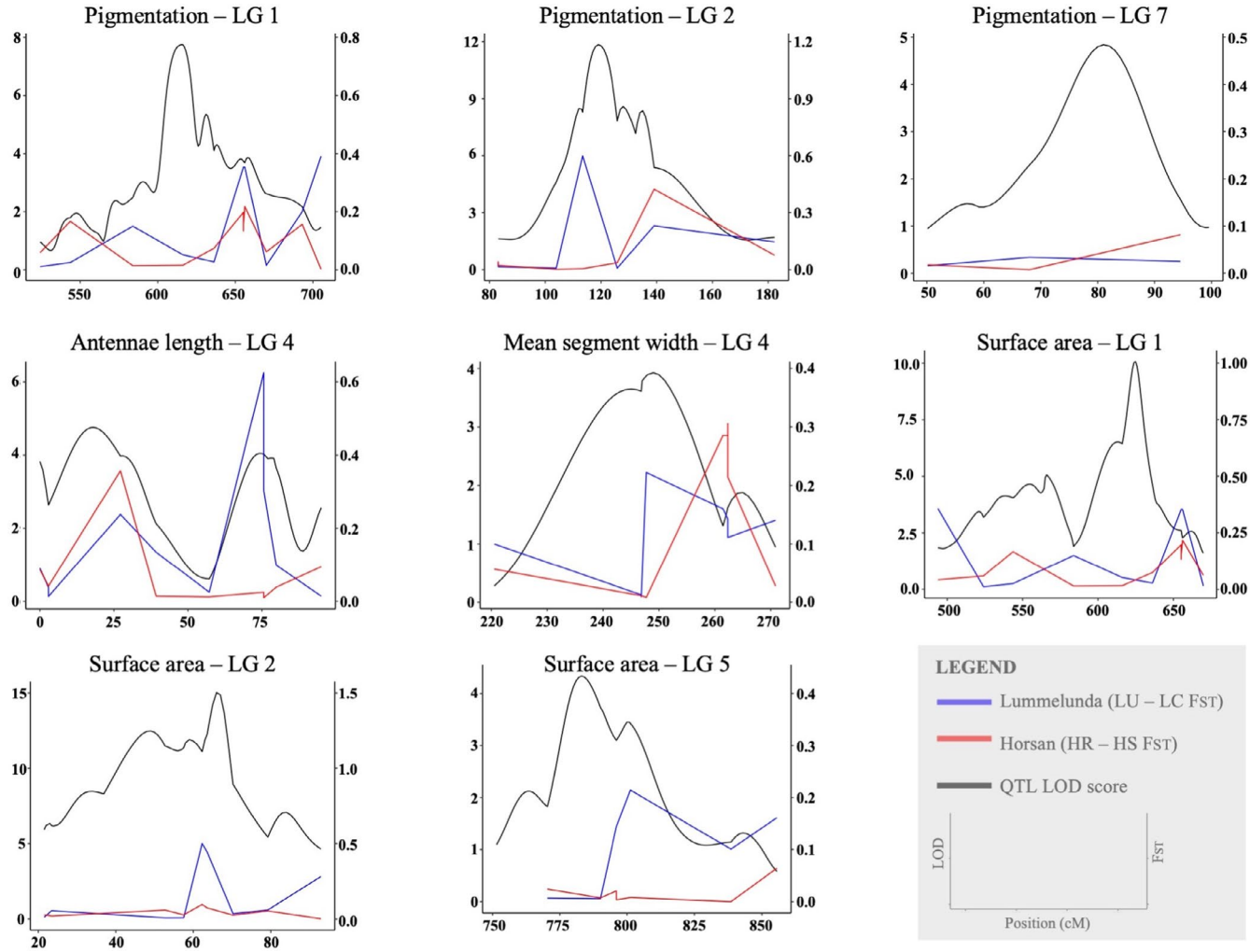


FIGURE 6 A comparison of LOD scores obtained from a QTL interval mapping approach using a *cave* × *surface Asellus aquaticus* intercross with the differentiation index (F_{ST}) of natural populations in these same genomic regions. Blue and red lines denote F_{ST} values between light and dark populations in Lummelunda and Horsan, respectively, while the black line denotes LOD scores for that particular trait

collection sites being low. Further support for restricted gene flow stems from the presence of distinct body pigmentation and lower nucleotide diversity found in LC (indicating that LC consists of smaller population sizes and/or strong selective pressures but is not being fully genetically replenished by gene flow from the surface). We expect that at least some degree of differential adaptation has occurred to the LC and HS habitats even if there were ongoing gene flow. This would limit admixture through reinforcement (Nosil et al., 2003) and would place this system in a speciation-with-gene-flow context, with the selection/migration ratio governing divergence until sufficient reproductive isolation is achieved, at which point other forces such as drift may begin to aid in genome-wide divergence (Feder et al., 2012).

4.2 | Genetic variation associated with cave vs. surface habitats in Lummelunda

The interesting convergent feature of body pigment and eye loss in cave environments spans across the tree of life, including numerous species of millipedes (e.g., Liu & Wynne, 2019; Vahtera et al., 2020), crustaceans (e.g., Carlini & Fong, 2017), insects (overview provided in Howarth, 2009), fish (e.g., Protas et al., 2006), amphibians (e.g., Hervant et al., 2001) and mammals (e.g., Simoes et al., 2019). Similarly, body pigment and eye loss have been found in several *A. aquaticus* cave populations including Lummelunda (Konec et al., 2015; Protas et al., 2011). Here, we identified six genes within our body pigmentation QTL confidence intervals. However, only one was functionally annotated: *NOP58*. This gene also fell into the confidence interval of a size-related QTL (surface area) and is involved in ATPase binding and ribosome biogenesis in mice (Abel et al., 2021). The detection of our body pigmentation QTL is in accordance with Protas et al., (2011), who also found a main effect pigmentation QTL on LG2 using a cave × surface *A. aquaticus* intercross from Slovenia. Furthermore, we found three genes associated with antennae length, one of which was annotated as *SALM*. Biological processes associated with *SALM* are numerous, with several potentially interesting functions in the context of cave adaptations. For example, in *Drosophila melanogaster*, it has been shown to be involved in antennal

joint development (Dong et al., 2002) and compound eye photoreceptor cell differentiation (Mollereau et al., 2001). Prolonged antennae length has also been reported in more than one *A. aquaticus* cave population as compared to their surface counterparts (Prevorcnik et al., 2004; Turk et al., 1996). In the Lummelunda Cave population (LC), around 10% of individuals show no visible eye pigmentation (our personal observations). Finally, *VCP*, *InterLeukin2* and the *BLOC-1* protein complex were found within confidence intervals of surface area-associated QTL. Among these, *InterLeukin2* was found to be involved in adaptive immune responses in humans (Weinberg & Parkman, 1990) and deficiency of the *BLOC-1* protein complex resulted in eye pigmentation defects in *D. melanogaster* (Cheli et al., 2010).

Our approach towards finding genomic regions associated with the cave environment, essentially by taking significant and shared allele frequency differences between LC–LU and LC–LD, resulted in the detection of 74 annotated genes. Some of these show potential functions that may be important for life in caves: *EMS*, *GRH*, *SPT6*, *LUSH*, *TLE4*, *CUPA2*, *CUPA3* and *CHH4*. The *EMS* gene is involved in brain development as well as head and brain segmentation (Hirth et al., 1995). Specifically, it controls antennal and mandibular segment identity in *D. melanogaster* (Peel, 2004). Next, *GRH* is involved in chitin-based cuticle development (Kim & McGinnis, 2011) and nervous system development (Cenci & Gould, 2005) through regulation of dopamine synthesis (*DDC* gene; De Luca et al., 2003). The neuronal dopamine pathway has been shown to be associated with temperature adaptations in *Drosophila simulans* (Jakšić et al., 2020) and is well known to be involved in melanin synthesis and pigmentation (Lemonds et al., 2016). Similarly, *CUPA2* and *CUPA3* are responsible for structural constituents of the cuticle in a rock crab species (Andersen, 1999). Temperatures in the Lummelunda cave habitat are low, on average 8°C (Lundevall, 1965). Interestingly, *SPT6* is required for the transcriptional induction of heat shock response genes in *D. melanogaster* (Ardehali et al., 2009). The *LUSH* gene is interesting with regard to sensory perception, as, in *D. melanogaster*, it is an odorant-binding protein required for olfactory behaviour and activity of pheromone-sensitive neurons (Xu et al., 2005). Further, regressive evolution of eye degradation in the Mexican tetra fish has been associated with the *PAX6* gene (Protas et al., 2007). In mouse, *TLE4* inhibits *PAX5* (Zhu et al., 2002), and is involved in several neuronal development processes (e.g., Ohtsuka et al., 2013). Finally, *CHH4* is a hormone specifically found in the sinus gland of isopods, which controls haemolymph sugar levels (Davey et al., 2000).

4.3 | Genetic variation associated with reed and stony bottom habitats in Lake Horsan

Similar to the above, the comparison of allele frequencies between light and dark lake ecotypes can also potentially identify heterogeneous patterns of differentiation along the genome, some of which may be attributed to their ecology and ecotype formation (Ferchaud & Hansen, 2016; Nosil et al., 2009). These SNPs were located in scaffolds containing 45 annotated genes that are potentially related to the light–dark ecotype differentiation. Some genes found on scaffolds exhibiting differences between HR and HS included *FERH*, *NRT*, *L2CC*, *TUT4*, *TRIM50*, *DHX37*, *EIF4g1* and *NINAC*. These genes were associated with a diverse range of functions, including oocyte maturation (*TUT4*), testis development (*DHX37*), starvation response (*TRIM50*) (Fusco et al., 2018; McElreavey et al., 2020; Morgan et al., 2017), resistance to iron toxicity (*FERH*; Geiser et al., 2003), larval metabolism, cuticle formation and responses to hypoxia (*L2CC*; Eveleth & Marsh, 1986; Lee et al., 2008), fear responses (*EIF4G1*; Ramírez-Valle et al., 2008) and visual perception (*NINAC*; Stephenson et al., 1983).

4.4 | Cave adaptation QTL and their role in lake ecotype formation

Given the potential for either parallel adaptation or the selection of low-frequency alleles that are common to both cave/surface and lake populations, one of the aims of this study was to identify whether some of the same loci underpin these different ecotypes. We find that in the case of the QTL identified in the cave (LC) × surface stream (LU) intercross, these QTL regions appear to show significant F_{ST} differences in both the cave (LC) × upstream (LU) and the stony bottom (HS) × reed (HR) lake comparisons. In particular, similar F_{ST} patterns in the QTL regions are seen in both the cave/surface and stony bottom/reed lake comparisons. This indicates first that the loci identified in the intercross are indicative of the loci that are segregating in the wider populations from which they were sampled, and second that these loci also appear to be under selection in stony bottom/reed lake ecotype formation. Although it is impossible to know if the same alleles are being selected in these cases from the current data, it does indicate that the same genomic regions are being selected on during the adaptation of these different ecotypes. The analysis of pairwise allele frequency differences between the two light–dark comparisons showed a negative correlation (i.e., an SNP allele with a higher frequency in the light (HS) lake population had a lower frequency in the light cave (LC) population). However, this analysis utilized relatively few SNPs and is thus not identifying the actual haplotype under selection and certainly not the causal variant. Therefore, any correlation (positive or negative) is more indicative that these regions appear to be under selection and gives no information with regards to which actual alleles are being selected at these loci.

4.5 | Concluding remarks

Previous work has demonstrated how *A. aquaticus* is well suited as a model for studies of habitat adaptation and in parallel contexts (Eroukhmanoff, Hargeby, et al., 2009; Hargeby et al., 2005; Karlsson et al., 2010; Konec et al., 2015; Re et al., 2018). Our work expands this potential by adding a genetic analysis of uncharacterized cave

and lake populations on Gotland. Further, although they are largely homogeneous, we demonstrate that cave and lake ecotypes have distinct genetic differentiation patterns at least in allele frequencies, which resulted in the identification of several interesting candidate genes associated with these habitats. We found several QTL associated with body pigmentation, antennae length and animal size, and showed that these same regions appear differentiated between natural populations.

One of the unique features of the *A. aquaticus* system that we present here is that the same trait (i.e., body pigmentation) is possibly under different selection regimes in the different populations (i.e., either predator induced or due to the absence of light in lake populations and cave populations, respectively). Given this differential selection, it is possible to address whether the same loci are targeted by selection in these separate ecotype formation events. Although the QTL regions identified in the cave intercross population here do appear to show F_{ST} signatures that could indicate ongoing selection in the lake ecotype comparison, the study lacks the resolution to identify whether these are the exact same loci being selected upon. Future studies will elucidate this and help to address whether parallel selection can still select the same loci even with a different underlying root selection pressure. In particular, genome-wide association and artificial selection studies using WGS data can more precisely test how the genetic architecture of this species responds to selection on a genome-wide level.

CONFLICT OF INTEREST

The authors have no conflict of interest.

ACKNOWLEDGEMENTS

The authors would like to especially thank the County Administrative Board of Gotland for permission to collect the isopods in the Lummelunda and Horsan nature preserves, and Hanne Ödin and Lummelundagrottan AB for giving us permission to enter the caves and use their equipment. The authors acknowledge support from the National Genomics Infrastructure in Genomics Production Stockholm funded by Science for Life Laboratory, and the SNIC/Uppsala Multidisciplinary Centre for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. Sequencing was performed by the Stockholm and Uppsala Sequencing Centers. Bioinformatic support was provided by NBIS (National Bioinformatics Infrastructure Sweden). Funding and support were provided by the LiU Neuro Framework and the Swedish Research Council (VR). We additionally thank Serra Helene, Jukka Palm and Jari Arkko for help with collecting samples in the field and providing with pictures and videos of cave animals. Finally, we thank Robin Abbey-Lee for her assistance in the laboratory.

AUTHOR CONTRIBUTIONS

D.W., R.H. and A.H.A. conceived the study. D.W. provided the funding. A.H.A. and V.B. collected samples from the field. V.B. conducted the genetics laboratory work with assistance from J.F., A.H.A. and M.L.M.C. A.H.A. and D.W. generated the intercross. A.H.A., D.W. and V.B. photographed the samples. V.B. led the data analysis and manuscript writing with supervision from D.W., R.H. and A.H.A. A.H.O. and M.L.M.C. assisted with the bioinformatics. All authors participated in the interpretation of results and manuscript editing.

DATA AVAILABILITY STATEMENT

The following data sets have been uploaded to public repositories: raw DNA reads from 10× genomics data set (NCBI Bioproject: PRJNA727065); raw DNA reads from WGS data set (NCBI Bioproject: PRJNA727065); raw DNA reads from RadSeq data set (NCBI Bioproject: PRJNA727065); DNA sequence of *COI* fragment of male and female intercross founders (NCBI: MW995949 and MW995950); assembled genome of *Asellus aquaticus* (Dryad: https://doi.org/10.5061/dryad.2547d7wqj); raw VCF file containing all single nucleotide variants (Dryad: https://doi.org/10.5061/dryad.2547d7wqj).

ORCID

Vid Bakovic https://orcid.org/0000-0001-9506-5816
Dominic Wright https://orcid.org/0000-0003-2329-2635

REFERENCES

Abel, Y., Paiva, A. C. F., Bizarro, J., Chagot, M.-E., Santo, P. E., Robert, M.-C., Quinternet, M., Vandermoere, F., Sousa, P. M. F., Fort, P., Charpentier, B., Manival, X., Bandeiras, T. M., Bertrand, E., & Verheggen, C. (2021). NOPCHAP1 is a PAQosome cofactor that helps loading NOP58 on RUVBL1/2 during box C/D snoRNP biogenesis. *Nucleic Acids Research*, *49*, 1094–1113. https://doi.org/10.1093/nar/gkaa1226

Andersen, S. O. (1999). Exoskeletal proteins from the crab, *Cancer pagurus*. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *123*, 203–211. https://doi.org/10.1016/S1095-6433(99)00051-3

Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W., & Lis, J. T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. *The EMBO Journal*, *28*, 1067–1077. https://doi.org/10.1038/emboj.2009.56

Beerli, P., & Palczewski, M. (2010). Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, *185*, 313–326. https://doi.org/10.1534/genetics.109.112532

Borowsky, R., & Cohen, D. (2013). Genomic consequences of ecological speciation in *Astyanax* cavefish. *PLoS One*, *8*, e79903. https://doi.org/10.1371/journal.pone.0079903

Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Sympatric, parapatric or allopatric: The most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 2997–3007. https://doi.org/10.1098/rstb.2008.0076

Carlini, D. B., & Fong, D. W. (2017). The transcriptomes of cave and surface populations of *Gammarus minus* (Crustacea: Amphipoda) provide evidence for positive selection on cave downregulated transcripts. *PLoS One*, *12*, e0186173. https://doi.org/10.1371/journal.pone.0186173

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140. https://doi.org/10.1111/mec.12354

Cenci, C., & Gould, A. P. (2005). *Drosophila* Grainyhead specifies late programmes of neural proliferation by regulating the mitotic activity

and Hox-dependent apoptosis of neuroblasts. *Development*, *132*, 3835–3845. https://doi.org/10.1242/dev.01932

Cheli, V. T., Daniels, R. W., Godoy, R., Hoyle, D. J., Kandachar, V., Starcevic, M., Martinez-Agosto, J. A., Poole, S., DiAntonio, A., Lloyd, V. K., Chang, H. C., Krantz, D. E., & Dell'Angelica, E. C. (2010). Genetic modifiers of abnormal organelle biogenesis in a *Drosophila* model of BLOC-1 deficiency. *Human Molecular Genetics*, *19*, 861–878. https://doi.org/10.1093/hmg/ddp555

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., & Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, *307*, 1928–1933. https://doi.org/10.1126/science.1107239

Comeault, A. A., Soria-Carrasco, V., Gompert, Z., Farkas, T. E., Buerkle, C. A., Parchman, T. L., & Nosil, P. (2014). Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in *Timema cristinae*. *The American Naturalist*, *183*, 711–727. https://doi.org/10.5061/dryad.ck2cm

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Darvasi, A., & Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics*, *141*, 1199–1207. https://doi.org/10.1093/genetics/141.3.1199

Davey, M. L., Hall, M. R., Willis, R. H., Oliver, R. W. A., Thurn, M. J., & Wilson, K. J. (2000). Five crustacean hyperglycemic family hormones of *Penaeus monodon*: complementary DNA sequence and identification in single sinus glands by electrospray ionization-Fourier transform mass spectrometry. *Marine Biotechnology*, *2*, 80–91. https://doi.org/10.1007/s101269900011

De Luca, M., Roshina, N. V., Geiger-Thornsberry, G. L., Lyman, R. F., Pasyukova, E. G., & Mackay, T. F. (2003). Dopa decarboxylase (Ddc) affects variation in *Drosophila* longevity. *Nature Genetics*, *34*, 429–433. https://doi.org/10.1038/ng1218

Doellman, M. M., Egan, S. P., Ragland, G. J., Meyers, P. J., Hood, G. R., Powell, T. H., Lazorchak, P., Hahn, D. A., Berlocher, S. H., Nosil, P., & Feder, J. L. (2019). Standing geographic variation in eclosion time and the genomics of host race formation in *Rhagoletis pomonella* fruit flies. *Ecology and Evolution*, *9*, 393–409. https://doi.org/10.1002/ece3.4758

Dong, P. S., Dicks, J. S., & Panganiban, G. (2002). Distal-less and homothorax regulate multiple targets to pattern the *Drosophila* antenna. *Development*, *129*, 1967–1974.

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Eroukhmanoff, F., Hargeby, A., Arnberg, N. N., Hellgren, O., Bensch, S., & Svensson, E. I. (2009). Parallelism and historical contingency during rapid ecotype divergence in an isopod. *Journal of Evolutionary Biology*, *22*, 1098–1110. https://doi.org/10.1111/j.1420-9101.2009.01723.x

Eroukhmanoff, F., Hargeby, A., & Svensson, E. I. (2011). The role of different reproductive barriers during phenotypic divergence of isopod ecotypes. *Evolution: International Journal of Organic Evolution*, *65*, 2631–2640. https://doi.org/10.1111/j.1558-5646.2011.01327.x

Eroukhmanoff, F., Hargeby, A., & Svensson, E. I. (2009). Rapid adaptive divergence between ecotypes of an aquatic isopod inferred from FST-QST analysis. *Molecular Ecology*, *18*, 4912–4923. https://doi.org/10.1111/j.1420-9101.2011.02322.x

Eroukhmanoff, F., & Svensson, E. I. (2009). Contemporary parallel diversification, antipredator adaptations and phenotypic integration in an aquatic isopod. *PLoS One*, *4*, e6173. https://doi.org/10.1371/journal.pone.0006173

Eroukhmanoff, F., & Svensson, E. I. (2011). Evolution and stability of the G-matrix during the colonization of a novel environment. *Journal of Evolutionary Biology*, *24*, 1363–1373. https://doi.org/10.1111/j.1420-9101.2011.02270.x

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Eveleth, D. D., & Marsh, J. L. (1986). Sequence and expression of the Cc gene, a member of the dopa decarboxylase gene cluster of *Drosophila*: possible translational regulation. *Nucleic Acids Research*, *14*, 6169–6183. https://doi.org/10.1093/nar/14.15.6169

Feder, J. F., Egan, S. E., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*, 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Feder, J. L., Nosil, P., Wacholder, A. C., Egan, S. P., Berlocher, S. H., & Flaxman, S. M. (2014). Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *Heredity*, *105*, 810–820. https://doi.org/10.1093/jhered/esu038

Ferchaud, A. L., & Hansen, M. M. (2016). The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: Three-spine sticklebacks in divergent environments. *Molecular Ecology*, *25*, 238–259. https://doi.org/10.1111/mec.13399

Filchak, K. E., Roethele, J. B., & Feder, J. L. (2000). Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature*, *407*, 739–742. https://doi.org/10.1038/35037578

Fišer, Ž., Prevorčnik, S., Lozej, N., & Trontelj, P. (2019). No need to hide in caves: shelter-seeking behavior of surface and cave ecomorphs of *Asellus aquaticus* (Isopoda: Crustacea). *Zoology*, *134*, 58–65. https://doi.org/10.1016/j.zool.2019.03.001

Fuller, N., Ford, A. T., Lerebours, A., Gudkov, D. I., Nagorskaya, L. L., & Smith, J. T. (2019). Chronic radiation exposure at Chernobyl shows no effect on genetic diversity in the freshwater crustacean, *Asellus aquaticus*, thirty years on. *Ecology and Evolution*, *9*, 10135–10144. https://doi.org/10.1002/ece3.5478

Fusco, C., Mandriani, B., Di Rienzo, M., Micale, L., Malerba, N., Cocciadiferro, D., Sjøttem, E., Augello, B., Squeo, G. M., Pellico, M. T., Jain, A., Johansen, T., Fimia, G. M., & Merla, G. (2018). TRIM50 regulates Beclin 1 proautophagic activity. *Molecular Cell Research*, *1865*, 908–919. https://doi.org/10.1016/j.bbamcr.2018.03.011

Geiser, D. L., Chavez, C. A., Flores-Munguia, R., Winzerling, J. J., & Pham, D. Q. D. (2003). *Aedes aegypti* ferritin: a cytotoxic protector against iron and oxidative challenge? *European Journal of Biochemistry*, *270*, 3667–3674. https://doi.org/10.1046/j.1432-1033.2003.03709.x

Gislén, T., & Brinck, P. (1950). Subterranean waters on Gotland with special regard to the Lummelunda current 2. Environmental conditions, plant and animal life, immigration problems. *Acta Universitatis Lundensis*, *46*, 1–81.

Graça, M. A. S., Maltby, L., & Calow, P. (1993). Importance of fungi in the diet of *Gammarus pulex* and *Asellus aquaticus*. *Oecologia*, *96*, 304–309. https://doi.org/10.1007/BF00317498

Gross, J. B., Sun, D. A., Carlson, B. M., Brodo-Abo, S., & Protas, M. E. (2020). Developmental transcriptomic analysis of the cave-dwelling crustacean, Asellus Aquaticus. *Genes*, *11*, 42. https://doi.org/10.3390/genes11010042

Hargeby, A., Blindow, I., & Andersson, G. (2007). Long-term patterns of shifts between clear and turbid states in Lake Krankesjön and Lake Tåkern. *Ecosystems*, *10*, 29–36. https://doi.org/10.1007/s10021-006-9008-5

Hargeby, A., & Erlandsson, J. (2006). Is size-assortative mating important for rapid pigment differentiation in a freshwater isopod? *Journal of Evolutionary Biology*, *19*, 1911–1919. https://doi.org/10.1111/j.1420-9101.2006.01170.x

Hargeby, A., Johansson, J., & Ahnesjö, J. (2004). Habitat-specific pigmentation in a freshwater isopod: adaptive evolution over a small

spatiotemporal scale. *Evolution*, *58*, 81–94. https://doi.org/10.1111/j.0014-3820.2004.tb01575.x

Hargeby, A., Stoltz, J., & Johansson, J. (2005). Locally differentiated cryptic pigmentation in the freshwater isopod *Asellus aquaticus*. *Journal of Evolutionary Biology*, *18*, 713–721. https://doi.org/10.1111/j.1420-9101.2004.00837.x

Harris, S., Eroukhmanoff, F., Green, K. K., Svensson, E. I., & Pettersson, L. B. (2011). Changes in behavioural trait integration following rapid ecotype divergence in an aquatic isopod. *Journal of Evolutionary Biology*, *24*, 1887–1896. https://doi.org/10.1111/j.1420-9101.2011.02322.x

Herman, A., Brandvain, Y., Weagley, J., Jeffery, W. R., Keene, A. C., Kono, T. J., Bilandzija, H., Borowski, R., Espinasa, L., O'Quin, K., & Omelas-Garcia, C. P. (2018). The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, *Astyanax mexicanus*. *Molecular Ecology*, *27*, 4397–4416. https://doi.org/10.1111/mec.14877

Hervant, F., Mathieu, J., & Durand, J. (2001). Behavioural, physiological and metabolic responses to long-term starvation and refeeding in a blind cave-dwelling (*Proteus anguinus*) and a surface-dwelling (*Euproctus asper*) salamander. *Journal of Experimental Biology*, *204*, 269–281. https://doi.org/10.1242/jeb.204.2.269

Hirth, F., Therianos, S., Loop, T., Gehring, W. J., Reichert, H., & Furukubo-Tokunaga, K. (1995). Developmental defects in brain segmentation caused by mutations of the homeobox genes orthodenticle and empty spiracles in *Drosophila*. *Neuron*, *15*, 769–778. https://doi.org/10.1016/0896-6273(95)90169-8

Howarth, F. G. (2009). Cave insects. In V. H. Resh, & R. T. Carde (Eds.), *Encyclopedia of insects* (pp. 139–143). Academic Press.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*, 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jakšić, A. M., Karner, J., Nolte, V., Hsu, S. K., Barghi, N., Mallard, F., Otte, K. A., Svečnjak, L., Senti, K. A., & Schlötterer, C. (2020). Neuronal function and dopamine signaling evolve at high temperature in *Drosophila*. *Molecular Biology and Evolution*, *37*, 2630–2640. https://doi.org/10.1093/molbev/msaa116

Jemec, A., Škufca, D., Prevorčnik, S., Fišer, Ž., & Zidar, P. (2017). Comparative study of acetylcholinesterase and glutathione S-transferase activities of closely related cave and surface *Asellus aquaticus* (Isopoda: Crustacea). *PLoS One*, *12*, e0176746. https://doi.org/10.1371/journal.pone.0176746

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*, 3070–3071. https://doi.org/10.1093/bioinformatics/btn129

Kaeuffer, R., Peichel, C. L., Bolnick, D. I., & Hendry, A. P. (2012). Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution: International Journal of Organic Evolution*, *66*, 402–418. https://doi.org/10.1111/j.1558-5646.2011.01440.x

Karlsson, K., Eroukhmanoff, F., Härdling, R., & Svensson, E. I. (2010). Parallel divergence in mate guarding behaviour following colonization of a novel habitat. *Journal of Evolutionary Biology*, *23*, 2540–2549. https://doi.org/10.1111/j.1420-9101.2010.02102.x

Karlsson, K., Eroukhmanoff, F., Harris, S., Pettersson, L. B., & Svensson, E. I. (2016). Rapid changes in genetic architecture of behavioural syndromes following colonization of a novel environment. *Journal of Evolutionary Biology*, *29*, 144–152. https://doi.org/10.1111/jeb.12769

Kautt, A. F., Kratochwil, C. F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., Torres-Dowdall, J., Härer, A., Hulsey, C. D., Franchini, P., Pippel, M., Myers, E. W., & Meyer, A. (2020). Contrasting signatures of genomic divergence during sympatric speciation. *Nature*, *588*, 106–111. https://doi.org/10.1038/s41586-020-2845-0

Kim, M., & McGinnis, W. (2011). Phosphorylation of Grainy head by ERK is essential for wound-dependent regeneration but not for development of an epidermal barrier. *Proceedings of the National Academy of Sciences*, *108*, 650–655. https://doi.org/10.1073/pnas.1016386108

Knaus, B. J., & Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*, 44–53. https://doi.org/10.1111/1755-0998.12549

Konec, M., Prevorčnik, S., Sarbu, S. M., Verovnik, R., & Trontelj, P. (2015). Parallels between two geographically and ecologically disparate cave invasions by the same species, *Asellus aquaticus* (Isopoda, Crustacea). *Journal of Evolutionary Biology*, *28*, 864–875. https://doi.org/10.1111/jeb.12610

Lee, S. J., Feldman, R., & O'Farrell, P. H. (2008). An RNA interference screen identifies a novel regulator of target of rapamycin that mediates hypoxia suppression of translation in *Drosophila* S2 cells. *Molecular Biology of the Cell*, *19*, 4051–4061. https://doi.org/10.1091/mbc.e08-03-0265

Lemonds, T. R., Liu, J., & Popadić, A. (2016). The contribution of the melanin pathway to overall body pigmentation during ontogenesis of *Periplaneta americana*. *Insect Science*, *23*, 513–519. https://doi.org/10.1111/1744-7917.12356

Levis, N. A., & Pfennig, D. W. (2016). Evaluating 'plasticity-first' evolution in nature: key criteria and empirical approaches. *Trends in Ecology & Evolution*, *31*, 563–574. https://doi.org/10.1016/j.tree.2016.03.012

Liu, W., & Wynne, J. J. (2019). Cave millipede diversity with the description of six new species from Guangxi, China. *Subterranean Biology*, *30*, 57. https://doi.org/10.3897/subtbiol.30.35559

Lundevall, C. F. (1965). Karst phenomena in the Lummelunda area of Gotland. *Geografiska Annaler: Series A, Physical Geography*, *47*, 45–60. https://doi.org/10.1080/04353676.1965.11879712

Lürig, M. D., Best, R. J., Svitok, M., Jokela, J., & Matthews, B. (2019). The role of plasticity in the evolution of cryptic pigmentation in a freshwater isopod. *Journal of Animal Ecology*, *88*, 612–623. https://doi.org/10.1111/1365-2656.12950

Mangiafico, S., & Mangiafico, M. S. (2017). Package 'rcompanion'. *Cran Repos*, *20*, 1–71.

Marcus, J. H., Sutcliffe, D. W., & Willoughby, L. G. (1978). Feeding and growth of *Asellus aquaticus* (Isopoda) on food items from the littoral of Windermere, including green leaves of *Elodea canadensis*. *Freshwater Biology*, *8*, 505–519. https://doi.org/10.1111/j.1365-2427.1978.tb01473.x

Maruzzo, D., Egredzija, M., Minelli, A., & Fusco, G. (2008). Segmental pattern formation following amputation in the flagellum of the second antennae of *Asellus aquaticus* (Crustacea, Isopoda). *Italian Journal of Zoology*, *75*, 225–231. https://doi.org/10.1080/11250000701885588

Maruzzo, D., Minelli, A., Ronco, M., & Fusco, G. (2007). Growth and regeneration of the second antennae of *Asellus aquaticus* (Isopoda) in the context of arthropod antennal segmentation. *Journal of Crustacean Biology*, *27*, 184–196. https://doi.org/10.1651/S-2756.1

McElreavey, K., Jorgensen, A., Eozenou, C., Merel, T., Bignon-Topalovic, J., Tan, D. S., Houzelstein, D., Buonocore, F., Warr, N., Kay, R. G. G., Peycelon, M., Siffroi, J.-P., Mazen, I., Achermann, J. C., Shcherbak, Y., Leger, J., Sallai, A., Carel, J.-C., Martinerie, L., … Bashamboo, A. (2020). Pathogenic variants in the DEAH-box RNA helicase DHX37 are a frequent cause of 46, XY gonadal dysgenesis and 46, XY testicular regression syndrome. *Genetics in Medicine*, *22*, 150–159. https://doi.org/10.1038/s41436-019-0606-y

Mendiburu, F., & Simon, R. (2015). Agricolae-Ten years of an open source statistical tool for experiments in breeding, agriculture and biology (No. e1748). *PeerJ*, *3*, e1404v1. https://doi.org/10.7287/peerj.preprints.1404v1

Mojaddidi, H., Fernandez, F. E., Erickson, P. A., & Protas, M. E. (2018). Embryonic origin and genetic basis of cave associated phenotypes

in the isopod crustacean *Asellus aquaticus*. *Scientific Reports*, *8*, 1–12. https://doi.org/10.1038/s41598-018-34405-8

Mollereau, B., Dominguez, M., Webel, R., Colley, N. J., Keung, B., De Celis, J. F., & Desplan, C. (2001). Two-step process for photoreceptor formation in *Drosophila*. *Nature*, *412*, 911–913. https://doi.org/10.1038/35091076

Morales, H. E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A. M., & Butlin, R. K. (2019). Genomic architecture of parallel ecological divergence: beyond a single environmental contrast. *Science*. *Advances*, *5*, eaav9963. https://doi.org/10.1126/sciadv.aav9963

Morgan, M., Much, C., DiGiacomo, M., Azzi, C., Ivanova, I., Vitsios, D. M., Pistolic, J., Collier, P., Moreira, P. N., Benes, V., Enright, A. J., & O'Carroll, D. (2017). mRNA 3′ uridylation and poly (A) tail length sculpt the mammalian maternal transcriptome. *Nature*, *548*, 347–351. https://doi.org/10.1038/nature23318

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, *12*, 443–451. https://doi.org/10.1038/nrg2986

Nosil, P., Crespi, B. J., & Sandoval, C. P. (2003). Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*, 1911–1918. https://doi.org/10.1098/rspb.2003.2457

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*, 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x

Nosil, P., Villoutreix, R., de Carvalho, C. F., Farkas, T. E., Soria-Carrasco, V., Feder, J. L., Crespi, B. J., & Gompert, Z. (2018). Natural selection and the predictability of evolution in *Timema* stick insects. *Science*, *359*, 765–770. https://doi.org/10.1126/science.aap9125

Nyberg, E., Danielsson, S., Eriksson, U., Faxneld, S., Miller, A., & Bignert, A. (2014). Spatio-temporal trends of PCBs in the Swedish freshwater environment 1981–2012. *Ambio*, *43*, 45–57. https://doi.org/10.1007/s13280-014-0561-4

Ohtsuka, N., Badurek, S., Busslinger, M., Benes, F. M., Minichiello, L., & Rudolph, U. (2013). GABAergic neurons regulate lateral ventricular development via transcription factor Pax5. *Genesis*, *51*, 234–245.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633

Peel, A. (2004). The evolution of arthropod segmentation mechanisms. *BioEssays*, *26*, 1108–1116. https://doi.org/10.1002/bies.20097

Peichel, C. L., Nereng, K. S., Ohgi, K. A., Cole, B. L., Colosimo, P. F., Buerkle, C. A., Schluter, D., & Kingsley, D. M. (2001). The genetic architecture of divergence between threespine stickleback species. *Nature*, *414*, 901–905. https://doi.org/10.1038/414901a

Pérez-Moreno, J. L., Balázs, G., & Bracken-Grissom, H. D. (2018). Transcriptomic insights into the loss of vision in Molnár János Cave's crustaceans. *Integrative and Comparative Biology*, *58*, 452–464. https://doi.org/10.1093/icb/icy071

Prevorcnik, S., Blejec, A., & Sket, B. (2004). Racial differentiation in *Asellus aquaticus* (L.)(Crustacea: Isopoda: Asellidae). *Archiv Für Hydrobiologie*, *160*, 193–214. https://doi.org/10.1127/0003-9136/2004/0160-0193

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959. https://doi.org/10.1093/genetics/155.2.945

Protas, M., Conrad, M., Gross, J. B., Tabin, C., & Borowsky, R. (2007). Regressive evolution in the Mexican cave tetra, *Astyanax mexicanus*. *Current Biology*, *17*, 452–454. https://doi.org/10.1016/j.cub.2007.01.051

Protas, M., 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*, 107–111. https://doi.org/10.1038/ng1700

Protas, M., Trontelj, P., & Patel, N. H. (2011). Genetic basis of eye and pigment loss in the cave crustacean, *Asellus aquaticus*. *Proceedings of the National Academy of Sciences*, *108*, 5702–5707. https://doi.org/10.1073/pnas.1013850108

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Ramírez-Valle, F., Braunstein, S., Zavadil, J., Formenti, S. C., & Schneider, R. J. (2008). eIF4GI links nutrient sensing by mTOR to cell proliferation and inhibition of autophagy. *The Journal of Cell Biology*, *181*, 293–307. https://doi.org/10.1083/jcb.200710215

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C., & Panova, M. (2016). Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology*, *25*, 287–305. https://doi.org/10.1111/mec.13332

Re, C., Fišer, Ž., Perez, J., Tacdol, A., Trontelj, P., & Protas, M. E. (2018). Common genetic basis of eye and pigment loss in two distinct cave populations of the isopod crustacean *Asellus aquaticus*. *Integrative and Comparative Biology*, *58*, 421–430. https://doi.org/10.1093/icb/icy028

Revell, L. J. (2012). Phytools: phylogenetic tools for comparative biology (and other things). *Methods in Ecology and Evolution*, *3*, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, *4*, 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*, 671–675. https://doi.org/10.1038/nmeth.2089

Simoes, B. F., Foley, N. M., Hughes, G. M., Zhao, H., Zhang, S., Rossiter, S. J., & Teeling, E. C. (2019). As blind as a bat? opsin phylogenetics illuminates the evolution of color vision in bats. *Molecular Biology and Evolution*, *36*, 54–68. https://doi.org/10.1093/molbev/msy192

Sket, B. (1994). Distribution of *Asellus aquaticus* (Crustacea: Isopoda: Asellidae) and its hypogean populations at different geographic scales, with a note on *Proasellus istrianus*. *Hydrobiologia*, *287*, 39–47. https://doi.org/10.1007/BF00006895

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., Buerkle, C. A., Feder, J. L., Bast, J., Schwander, T., Egan, S. P., Crespi, B. J., & Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, *344*, 738–742. https://doi.org/10.1126/science.1252136

Stahl, B. A., Gross, J. B., Speiser, D. I., Oakley, T. H., Patel, N. H., Gould, D. B., & Protas, M. E. (2015). A transcriptomic analysis of cave, surface, and hybrid isopod crustaceans of the species *Asellus aquaticus*. *PLoS One*, *10*, e0140484. https://doi.org/10.1371/journal.pone.0140484

Stephenson, R. S., O'Tousa, J., Scavarda, N. J., Randall, L. L., & Pak, W. L. (1983). Drosophila mutants with reduced rhodopsin content. *Symposia of the Society for Experimental Biology*, *36*, 477-501.

Turk, S., Sket, B., & Sarbu, Ş. (1996). Comparison between some epigean and hypogean populations of *Asellus aquaticus* (Crustacea: Isopoda: Asellidae). *Hydrobiologia*, *337*, 161–170. https://doi.org/10.1007/BF00028517

Vahtera, V., Stoev, P., & Akkari, N. (2020). Five million years in the darkness: A new troglomorphic species of Cryptops Leach, 1814 (Chilopoda, Scolopendromorpha) from Movile Cave, Romania. *Zookeys*, *1004*, 1. https://doi.org/10.3897/zookeys.1004.58537

Weinberg, K., & Parkman, R. (1990). Severe combined immunodeficiency due to a specific defect in the production of interleukin-2. *New England Journal of Medicine*, *322*, 1718–1723. https://doi.org/10.1056/NEJM199006143222406

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Research*, *27*, 757–767. https://doi.org/10.1101/gr.214874.116

Xu, P., Atkinson, R., Jones, D. N., & Smith, D. P. (2005). *Drosophila* OBP LUSH is required for activity of pheromone-sensitive neurons. *Neuron*, *45*, 193–200. https://doi.org/10.1016/j.neuron.2004.12.031

Yu, G. (2020). Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics*, *69*, e96. https://doi.org/10.1002/cpbi.96

Zhu, C. C., Dyer, M. A., Uchikawa, M., Kondoh, H., Lagutin, O. V., & Oliver, G. (2002). Six3-mediated auto repression and eye development requires its interaction with members of the Groucho-related family of co-repressors. *Development*, *129*, 2835–2849. https://doi.org/10.1242/dev.129.12.2835

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Bakovic, V., Martin Cerezo, M. L., Höglund, A., Fogelholm, J., Henriksen, R., Hargeby, A., & Wright, D. The genomics of phenotypically differentiated *Asellus aquaticus* cave, surface stream and lake ecotypes. *Mol Ecol*. 2021;00:1–18. https://doi.org/10.1111/mec.15987