






RESEARCH ARTICLE

Evolution of endoglucanase genes in subterranean and surface isopod crustaceans from Central Western Australia

Mohammad Javidkar^{1,2}  | Steven J. B. Cooper^{1,3}  | Nahid Shokri Bousjein⁴  |
William F. Humphreys^{5,6} | Rachael A. King^{1,3}  | Andrew D. Austin^{1,3} 

¹Department of Ecology and Evolutionary Biology, School of Biological Sciences, and The Environment Institute, University of Adelaide, Adelaide, South Australia, Australia

²The Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center, Houston, Texas, USA

³South Australian Museum, Adelaide, South Australia, Australia

⁴Faculty of Biological Sciences, Flinders University, Bedford Park, South Australia, Australia

⁵Western Australian Museum, Welshpool, Western Australia, Australia

⁶School of Animal Biology, University of Western Australia, Crawley, Western Australia, Australia

Correspondence

Mohammad Javidkar and Steven J. B. Cooper, Department of Ecology and Evolutionary Biology, School of Biological Sciences, and The Environment Institute, University of Adelaide, SA 5005, Australia. Email: m.javidkar@gmail.com and steve.cooper@samuseum.sa.gov.au

Funding information

Bennelongia Pty Ltd; Biota Environmental Sciences Pty Ltd; Curtin University of Technology; Flinders University; Minara Resources Ltd; South Australian Museum; The University of Adelaide; WA Department of Biodiversity, Conservation and Attractions; Western Australian Museum; Australian Research Council Linkage, Grant/Award Number: LP140100555 and LP100200494

Abstract

Recent studies have identified a significant number of endogenous cellulase genes in various arthropods, including isopods, allowing them to process hydrocarbons efficiently as a food source. While this research has provided insight into underlying gene-level processes in cellulose decomposition by arthropods, little is known about the existence and expression of cellulase genes in species from cave environments where carbohydrates are sparse. To investigate whether endogenous cellulase genes are maintained in subterranean species, we sequenced the transcriptomes of two subterranean paraplatharid isopod species from calcrete (carbonate) aquifers of central Western Australia and a related surface isopod species. Seven protein-coding open-reading frames associated with endoglucanase genes were identified in all species. Orthology inference analyses, using a wide range of cellulase sequences from available databases, supported the endogenous origin of the putative endoglucanase genes. Selection analyses revealed that these genes are primarily subject to purifying selection in most of the sites for both surface and subterranean isopod species, indicating that they are likely to encode functional peptides. Furthermore, evolutionary branch models supported the hypothesis of an adaptive shift in selective pressure acting on the subterranean lineages compared with the ancestral lineage and surface species. Branch-site models also revealed a few amino acid sites on the subterranean branches to be under positive selection, suggesting the acquisition of novel adaptations to the subterranean environments. These findings also imply that hydrocarbons exist in subsurface aquifers, albeit at reduced levels, and have been utilized by subterranean isopods as a source of energy for millions of years.

KEYWORDS

calcrete aquifers, carbohydrates, lignocellulose, oniscidea

TAXONOMY CLASSIFICATION

Evolutionary ecology

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Subterranean realms are renowned for their aphotic and low variability of abiotic conditions (e.g., temperature, humidity) and are also known for being energy-limited (oligotrophic) and reliant on the photosynthetic surface environment. As a result, subterranean species are not only characterized by troglomorphies such as reduced or loss of eyes and body pigmentation due to regressive evolution but also the acquisition of special behavioral and physiological features to cope with oligotrophic environments through adaptive processes. Food scarcity in subterranean environments is considered a significant selective force in the evolution of adaptive features in subterranean animals such as smaller body size, higher food utilization efficiency, and reduction in growth rate, metabolic rate, and energy demand (Culver & Pipan, 2009; Hüppop, 2012). Exogenic organic matter (OM), including cellulose produced by plants and phototrophic bacteria, is regarded as the main source of energy to subterranean environments from surface aquatic and terrestrial zones by water flow (Humphreys, 2006).

In addition to the temporo-spatial limitation of food distribution in hypogean environments (though not all are), the ability of organisms to break down OM is highly variable. It has long been assumed that metazoans can only degrade cellulose via cellulase enzymes derived from their symbiotic protozoa and bacteria in their digestive system. Recent research has provided increasing evidence for the presence of endogenous cellulase enzymes in various arthropods including crustaceans (Bredon et al., 2018; Kostanjsek et al., 2010). For instance, Bredon et al. (2019) using transcriptome data from various marine, freshwater, and terrestrial isopods, identified 40 endogenous lignocellulose degrading CAZymes, with some cellulase and hemicellulase genes found in more than 95% of transcriptomes. Unlike the identification of endoglucanase genes (Glycoside hydrolase family 9 (GH9), subgroup E₂) in many arthropods, the endogenous exoglucanases (GH7) in eukaryotes and prokaryotes are uncommon (Lombard et al., 2014) and have been reported in a few crustacean species. To the best of our knowledge, research on isopod cellulases has been largely concentrated on the digestive tract and hepatopancreas. However, studies of termites have shown that cephalic tissues are likely to provide an important source of cellulases (Watanabe et al., 1997).

Although the current evidence has shed light on the ability of arthropods to secrete endogenous cellulases, information on the type and origin of these enzymes in subterranean arthropods, which live in habitats where the organic supply is scarce, is limited to a few subterranean termites (Zhang et al., 2011). The arid central Western Australia (WA), which contains hundreds of isolated groundwater calcrete bodies, harbors a high diversity of subterranean invertebrates, occurring both in and above the aquifer, with species generally confined to individual calcrete bodies (Cooper et al., 2002, 2007, 2008; Guzik et al., 2008; Figure 1). Despite extensive research on the biodiversity and evolutionary history of this subterranean fauna, knowledge of organic supply to these subsurface habitats remains incomplete. Saccò et al. (2022), using

radiocarbon fingerprinting and metabarcoding of the gut contents of stygobiotic crustaceans occurring in the subterranean aquifers of the Yilgarn region of WA (Sturt Meadows), reported evidence of organic matter flow to the aquifers from surface ecosystems, implying a link between the subterranean and above-ground environments. The mean concentrations of dissolved organic carbon in the same region range from 0.39 to 1.94 mg/L (Saccò et al., 2020) during low and high rainfall regimes, respectively, which is in the range of oligotrophic groundwater systems (0.2 to 2 mg/L; Thurman & Malcolm, 1981).

To determine the existence and expression of endogenous cellulase genes in subterranean isopods and their evolution, we examined subterranean isopods in an oligotrophic subterranean ecosystem associated with the Yilgarn calcrete aquifers. The primary aim of our study was to identify putative endogenous endo-beta-1,4-glucanase (endoglucanase) genes involved in cellulolytic activities in two subterranean oniscidean isopods (*Paraplatyarthus subterraneus* and *Paraplatyarthus* sp.) and one surface species, *Porcellionides pruinosus*, using transcriptome data from cephalon tissues. *Paraplatyarthus subterraneus* possesses well-defined troglomorphies (e.g., lack of eyes, no pigmentation) with evidence for genetic isolation within the calcretes (Javidkar et al., 2016) over a long evolutionary time scale. Therefore, it represents a troglomorphic species with populations strongly bound to subterranean habitats. In contrast, *Paraplatyarthus* sp., with partial eyes, a semi-pigmented body, and molecular evidence for dispersal between calcretes, is likely to be a troglophile (subterranean species able to live and reproduce underground as well as in the surface environment). Hence, we anticipate that the long evolutionary history of these subterranean isopods in a potentially low-food environment should promote the acquisition of adaptive features at the amino acid level to efficiently use cellulose as an energy source.

2 | MATERIALS AND METHODS

2.1 | Fieldwork

The specimens of *P. subterraneus* ($n = 7$), *Paraplatyarthus* sp. ($n = 12$), and *P. pruinosus* ($n = 2$) were collected from the Laverton Downs calcrete aquifer (Yilgarn Region, Western Australia; Figure 1). More specifically, the subterranean isopod specimens were sampled from Laverton Downs Windarra calcrete aquifers (S28.4989, E122.1798) using leaf litter traps. The traps were left underground and suspended above the water table for 3–12 months to be colonized by invertebrates and the recovery of the traps was carried out 2–3 times per year (between April and October). After retrieving the traps, the contents were examined to identify any live isopod specimens. These specimens were then preserved in the RNAlater solution (Qiagen) to facilitate the extraction of intact RNA for transcriptome analyses. The surface (epigeal) isopods were collected by hand under/between crevices of fallen tree branches and preserved in RNAlater (See Javidkar et al., 2016 for more details on the approach).

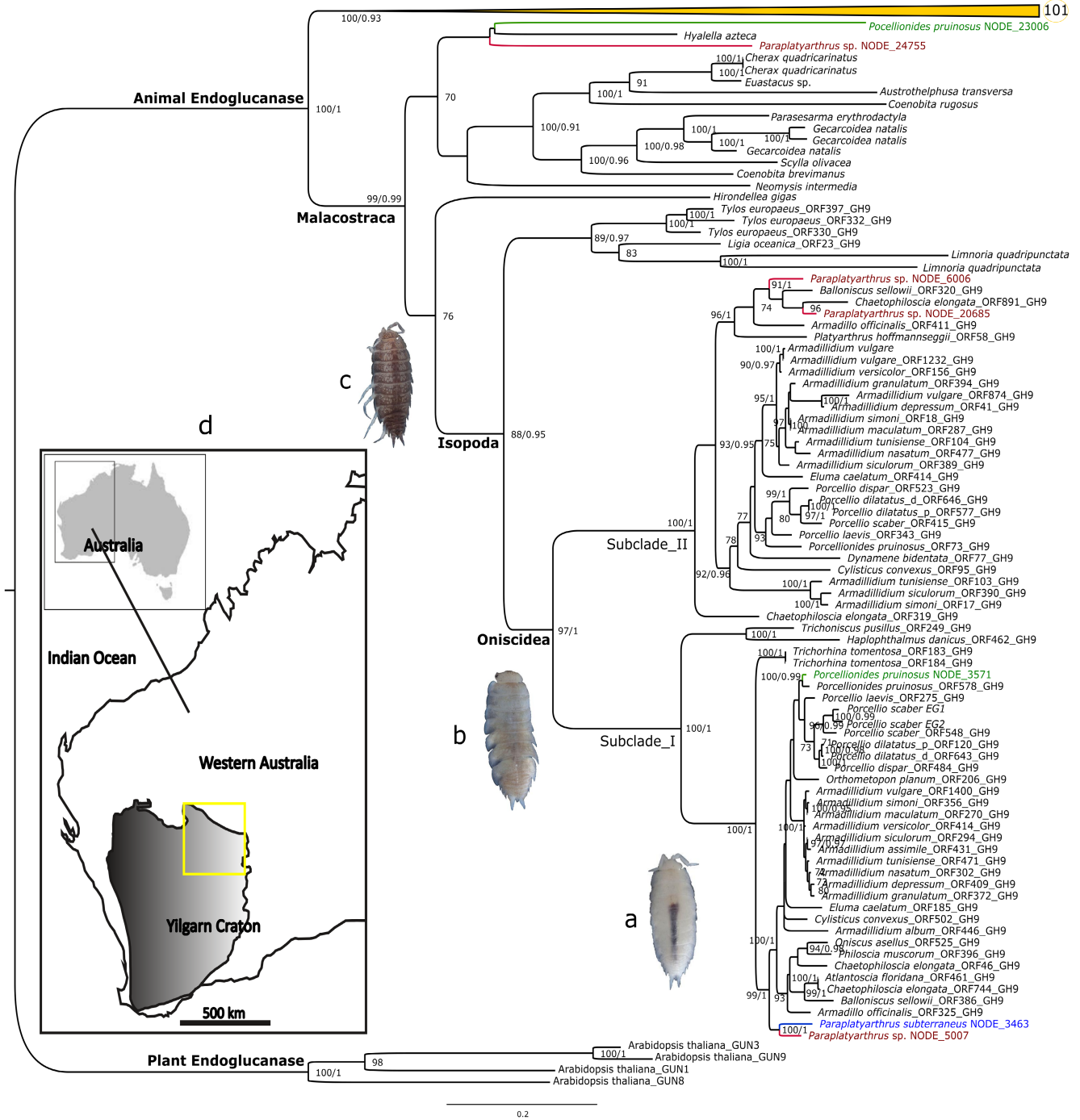


FIGURE 1 Maximum likelihood protein phylogeny of GH9-endoglucanase across animal and plant species. Numbers next to the nodes are ML bootstrap and BI posterior probability values, respectively. ML and BI support values less than 70 and 0.95, respectively, are not shown. Branches color-coded with “red,” “green,” and “blue” represent *Paraplatyarthrus* sp. (troglobite), *Porcellionides pruinosus* (surface), and *Paraplatyarthrus subterraneus* (troglobite) respectively. The collapsed clade shown in yellow comprises other arthropods, some annelid and mollusk species; the circle next to the collapsed clade indicates the number of terminal branches. Images (a–c) show typical troglobite, troglophile, and surface species, respectively. (d) The Yilgarn region of Western Australia; the yellow square indicates the location of calcrete aquifers subjected to sampling.

2.2 | Laboratory experiments and sequencing

The specimens of *P. subterraneus*, *Paraplatyarthrus* sp., and *Porcellionides pruinosus* were used for RNA extractions. The cephalothorax for

each species was dissected in sterile Petri dishes within RNAlater to avoid RNA degradation during handling. The head samples of each species, with identical COI haplotypes verified using PCR experiments and DNA sequencing, were then pooled within sterile/RNAase-DNAase

safe 2 mL vials. Total mRNA was extracted using a QIAGEN RNeasy Plus Micro Kit (www.qiagen.com). To generate full-length cDNA libraries, the Clontech SMARTER PCR cDNA Synthesis Kit protocol was followed according to the manufacturer's manual (www.clontech.com). The cDNA libraries for each species were sequenced using an Illumina HiSeq 2000 platform with 100bp paired-end reads.

2.3 | Bioinformatics

FastQC v0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to assess the quality of the Illumina reads. Trimmomatic v0.22 (Lohse et al., 2012) was applied to remove primers, adapters, and other Illumina-specific sequences from the reads. The de novo transcriptome assembly analyses were carried out using rnaSPAdes (Bankevich et al., 2012). The assembly metrics were calculated using TransRate v1.0.1 (Smith-Unna et al., 2016) and Samtools v1.7 (Danecek et al., 2021). BUSCO v5.3.1 (Simão et al., 2015) was utilized to assess the completeness of the transcriptome assemblies with single-copy orthologs, using arthropods' orthologous groups (number of genomes: 90, number of BUSCOs: 1013). To predict protein-coding open reading frames (ORFs) within transcripts across the species, Transdecoder V5.5.0 (Haas et al., 2013), with at least 100 amino acids long as minimum criteria to avoid increasing the false-positive discovery rate, was used. The translated ORFs were then searched against an in-house database of endoglucanase (15,602 seqs), exoglucanase (5182 seqs) protein sequences obtained from UniProtKB and the published isopod CAZymes (37,481 seqs; Bredon et al., 2019), using NCBI BlastP (Johnson et al., 2008) with an E -value cutoff of $1e^{-6}$. The putative cellulase-related ORFs (cORFs) were searched against the UniProtKB/Swiss-Prot non-redundant protein sequence database using BLASTP for cross-validation. The cellulase-related DNA transcripts identified in the three species, as queries, were searched against the combined transcriptome dataset using BlastN (E -value cutoff: $1e^{-6}$) for detecting potential cellulase pseudogenes and partial cellulase transcripts not picked up by BlastP. dbCAN3 (<https://bcb.unl.edu/dbCAN2/index.php>; Zheng et al., 2023), which is an automated carbohydrate-active enzyme and substrate annotation server, was used to determine the family and activity of the identified enzymes.

2.4 | Phylogenetic-oriented orthology inference

A dataset of 194 metazoan endoglucanase amino acid sequences was aligned using the MAFFT multiple aligner implemented in Geneious 9.1.4 (<https://www.geneious.com>). To further verify the identity of putative cellulase orthologs, both Bayesian inference (BI) and maximum likelihood (ML) phylogenetic methods were conducted, using the same amino acid dataset (495 aa), including arthropods, mollusks, and annelids. The endoglucanase sequences of *Arabidopsis thaliana* (plant) were used as the outgroup. WAG+I+G4 (LogL: -38128.063, AICc: 79836.780) was determined to be the best-fit model of protein evolution using AICc criteria with ModelFinder

(Kalyaanamoorthy et al., 2017). The ML analysis (1000 bootstraps) was carried out using IQ-TREE v1.6.12 (Hoang et al., 2018). The BI analysis was performed using two independent runs, 11 million generations, subsampling trees/parameters every 100 generations, and a 10% burn-in, as implemented in MrBayes v3.2.0 (Huelsenbeck & Ronquist, 2005). The final standard deviation of split frequencies was 0.009, suggesting convergence had occurred.

2.5 | Selection pressure

To investigate potential selective pressures on endoglucanase ORFs (enORFs), using the site (M0 (one-ratio), M1a (NearlyNeutral), M2a (Selection), M3 (Discrete), M7 (beta), M8 (beta and ω)), branch and branch-site models, PAML v4.9j was used (Yang, 2007). For the branch-site model (Model A), both Naive Empirical Bayes (NEB) and Bayes Empirical Bayes (BEB) were used for calculating the posterior probability of site classes. Five evolutionary hypotheses to examine potential adaptive trends were investigated using branch models, allowing ω (d_N/d_S ; nonsynonymous rate to synonymous rate ratio) variations along the target branches (Figure 2): H_0 : Homogenous selective pressure over the phylogenetic tree; H_1 : Homogenous adaptive shift in selective pressure acting on the troglobite (E_1) and troglophile (E_2) species different from the ancestral lineage (E_0 , probably a surface ancestor) and surface species (E_3); H_2 : Homogenous long-term shift in selective pressure acting on E_0 and subterranean species (E_1 & E_2), different from E_3 ; H_3 : Heterogeneous adaptive shift in selection acting on subterranean species with E_1 and E_2 subject to differential selective pressures, and different from E_0/E_3 ; H_4 : Heterogeneous selective pressures across most of the phylogeny. Twice the log-likelihood difference ($2\Delta L = 2(L_1 - L_0)$) between the nested models was compared using a chi-square (χ^2) distribution with $df = 1$ at a 5% significance level.

Furthermore, tests of FEL to detect selection pressure for each site along the entire phylogeny, the branch-site model, aBSREL, to discover the proportion of branches under positive pressure, BUSTED to evaluate gene-wide episodic diversifying positive selection, and relaxed selection (RELAXED) were carried out via the datamonkey server (Weaver et al., 2018). For the selection tests, to obtain resolved branching patterns and boost the power of analyses, only canonical enORFs (with at least two conserved domains (Table S2) for the predicted peptides; Davison & Blaxter, 2005) were used. Exoglucanase was excluded from the selection and phylogenetic analyses as it was detected in just one species.

3 | RESULTS

3.1 | Assembly, blast search, and orthology inference

Following the de novo transcriptome assembly of the cDNA libraries, a total of 32,245, 30,014, and 30,087 transcripts were generated

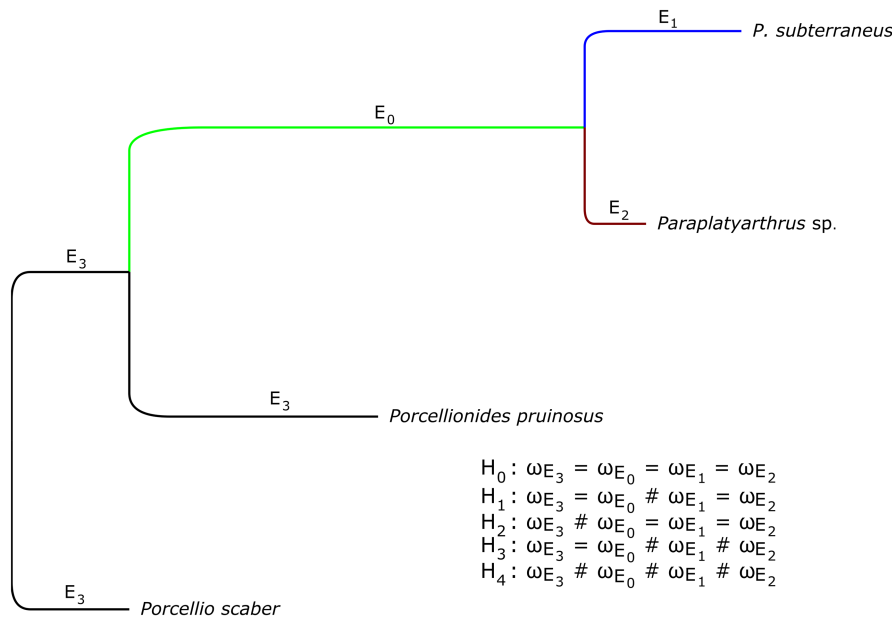


FIGURE 2 Branch models of ω variations across the phylogeny. H_0 – H_4 represent different hypotheses on the evolutionary trends of selective pressures acting on the ancestral lineage (E_0), subterranean (E_1 , E_2), and surface (E_3) branches.

for *P. subterraneus*, *Paraplatyarthus* sp., and *P. pruinosis*, respectively (Table S1). Based on the BUSCO analyses, of the total 1013 BUSCO groups searched across the arthropod genomes, 79.6%, 79.3%, and 75.7% complete-and-fragmented BUSCOs were recovered for *P. subterraneus*, *Paraplatyarthus* sp., and *P. pruinosis*, respectively (See Table S1). Based on the BlastP analyses, eight canonical and partial cORFs with significant homology (E value $< 10^{-20}$) to endoglucanase and exoglucanase (the latter, just in *Paraplatyarthus* sp.) were identified (Table 1). No additional cellulase-related transcripts or pseudogenes were identified using BlastN. According to the BlastP analyses, one, two, and four endoglucanase genes were identified as belonging to the troglobite, surface, and troglophile species, respectively. The sum of per base read depths (the total read base count) for each cORF is also provided in Table 1.

Both the ML and BI phylogenetic analyses generated similar topologies (Figure 1). The putative enORFs were all placed into the monophyletic Malacostraca endoglucanase clade (support values: bootstrap (bp)=99, posterior probability (pp)=0.99; Figure 1). The majority of enORFs were placed into a monophyletic oniscidean isopod endoglucanase clade (bp=97, pp=1), forming two major subclades (I and II) with all canonical enORFs clustered into Subclade_I (bp=100, pp=1). The phylogenetic placements of enORFs within the oniscidean clade with high support values indicate that the endogenous proteins are most likely encoded by the isopod genomes.

3.2 | Evolutionary selective pressures

To detect possible positive selection associated with amino acid codon sites over the whole phylogeny, multiple tests of selection

pressures were carried out. Evidence for positive selection over the whole phylogeny was not detected using the site models (Model comparisons: M1a–M2a and M7–M8) using a likelihood ratio test (LRT) at 95% confidence (Table S3); however, the FEL model showed a single codon under diversifying positive selection (codon 77: LRT=4.194, p -value=0.041 $< .05$). The comparison of M3–M0 models provided strong evidence of heterogeneous selective pressures across the codon sites over the phylogeny (LRT: 31.064, $df=4$, χ^2 p -value=.000).

To assess whether there was evidence of an adaptive shift in selection pressure or positive selection acting on the endoglucanase genes of the subterranean species, branch models were used. For the branch models (Table 2), the H_1 hypothesis (H_0 as null) was found to be highly significant (H_0 – H_1 , $df=1$, χ^2 p -value=.000), while the H_2 model (H_0 as null) was not supported (H_0 – H_2 , $df=1$, χ^2 p -value=.142). In addition, the H_3 hypothesis was not favored over H_1 (H_1 as null; H_1 – H_3 : $df=1$, χ^2 p -value=.584), with H_4 also not being supported (H_3 as null) by the data (H_3 – H_4 , $df=1$, χ^2 p -value=.326). Based on the LRT analyses and AIC criteria, the H_1 hypothesis (Table 2) was the best-fit model of ω variations along the branches, suggesting there was an adaptive shift in selective pressure acting on the endoglucanase genes of the troglobite and troglophile isopod species. The branch-site model of positive selection (Model A) was found to be significant at 95% confidence (null model fixes $\omega_2=1$; LRT=4.008, $df=1$, χ^2 p -value=.045). As a result, NEB and BEB detected site classes of positive selection with high posterior probabilities, on a few codon sites, along the subterranean branches (Table 3). Moreover, aBSREL found evidence of episodic diversifying selection for the troglobite branch on the phylogeny ($\omega_1=0.00$ (95%), $\omega_2=5.09$ (5.1%), LRT=4.389, p -value=.040; 95% confidence). BUSTED also found evidence

TABLE 1 Transcripts associated with endoglucanase and exoglucanase (*) genes in surface (Ω), and subterranean troglobite (\ddagger) and troglophile (\dagger) species, enzyme family (EF), Enzyme Commission (E.C.) number, transcript length (TL), total read base count (i.e., the sum of per base read depths (Bedcov)), ORF length (aa), the estimated molecular weight of the predicted peptides, and accession numbers. Canonical and partial (<150 aa) cORFs are shown with C and P, respectively.

Species	Cellulase_ Transcripts	EF	E.C.	TL (bp)	Bedcov	ORF_length (aa)	C/P	MW (KDa)	Accessions
<i>Porcellionides pruinosus</i> ^{Ω}	NODE_3571	GH9	3.2.1.4	1306	12,297	412	C	46.00	OR31528
<i>Paraplatyarthrus subterraneus</i> ^{\ddagger}	NODE_3463	GH9	3.2.1.4	1494	12,759	481	C	53.36	OR31527
<i>Paraplatyarthrus</i> sp. ^{\dagger}	NODE_5007	GH9	3.2.1.4	1265	23,509	426	C	47.27	OR31529
<i>Paraplatyarthrus</i> sp. ^{\dagger}	NODE_6006	GH9	3.2.1.4	1123	13,461	356	C	38.67	OR31530
<i>Paraplatyarthrus</i> sp. ^{\dagger}	NODE_20685	-	-	361	2011	119	P	12.73	OR31531
<i>Paraplatyarthrus</i> sp. ^{\dagger}	NODE_24755	-	-	308	392	102	P	11.29	OR31532
<i>Paraplatyarthrus</i> sp. ^{\dagger*}	NODE_18582	-	-	401	1246	133	P	14.12	OR31533
<i>Porcellionides pruinosus</i> ^{Ω}	NODE_23006	-	-	300	-	100	P	11.29	OR31534

TABLE 2 Parameter estimates under variable ω ratios among branches, log-likelihood (L) values under different models following ML estimations; the number of free parameters (np), Akaike Information Criterion (AIC) index, and LRTs ($2\Delta L$) of branch model comparisons using χ^2 distribution (df = 1, only nested models were compared).

Branch models and their comparisons	ω_{E_0}	ω_{E_1}	ω_{E_2}	ω_{E_3}	L	np	AIC	LRT ($2\Delta L$)
$H_0: \omega_{E_3} = \omega_{E_0} = \omega_{E_1} = \omega_{E_2}$	0.050	0.050	0.050	0.050	-1866.285	7	3746.57	-
$H_1: \omega_{E_3} = \omega_{E_0} \neq \omega_{E_1} = \omega_{E_2}$	0.033	0.165	0.165	0.033	-1857.967	8	3731.934	-
$H_2: \omega_{E_3} \neq \omega_{E_0} = \omega_{E_1} = \omega_{E_2}$	0.063	0.063	0.063	0.036	-1865.211	8	3746.422	-
$H_3: \omega_{E_3} = \omega_{E_0} \neq \omega_{E_1} \neq \omega_{E_2}$	0.032	0.127	1.056	0.032	-1857.534	9	3733.068	-
$H_4: \omega_{E_3} \neq \omega_{E_0} \neq \omega_{E_1} \neq \omega_{E_2}$	0.028	0.125	1.827	0.037	-1857.335	10	3734.67	-
H_0 vs H_1	-	-	-	-	-	-	-	16.636**
H_0 vs H_2	-	-	-	-	-	-	-	2.149
H_1 vs H_3	-	-	-	-	-	-	-	0.351
H_3 vs H_4	-	-	-	-	-	-	-	0.398

Note: Two asterisks indicate a very significant LRT ($p < .01$, χ^2 critical value = 6.63).

TABLE 3 Parameters in the branch-site Model A including the proportion of site classes, ω for background (surface species) and foreground (subterranean species) lineages, and positively selected sites (PSS, $\omega > 1$) using BEB and NEB methods along the foreground branches.

Site class	Proportion	Background ω	Foreground ω	PSS using BEB	PSS using NEB
0 ($0 < \omega_0 < 1$)	0.913	0.018	0.018	K* (15), W* (85), W* (91), G (189)	K (15), W* (85), W* (91), Q (53), A (72), S (77), L (145), G (189), T (202), T (220), D (248)
1 ($\omega_1 = 1$)	0.036	1.000	1.000		
2a	0.049	0.018	4.223		
Background ($0 < \omega_0 < 1$)					
Foreground ($\omega_2 \geq 1$)					
2b	0.002	1.000	4.223		
Background ($\omega_2 = 1$)					
Foreground ($\omega_2 \geq 1$)					

Note: Amino acids with bold font indicate positive sites with posterior probability (p) > .99, asterisks show positive sites with $p > .95$, and plain font denotes positive sites with $p > .5$. Site numbers are shown in parentheses.

of diversifying positive selection for at least one site on at least a test branch for the subterranean branches (lnL = -1825.4, LRT, p -value = .021; Table S4). The tests of relaxed selection were not significant ($p > .05$; Table S5).

4 | DISCUSSION

This study provides the first evidence for the expression of cellulase genes in subterranean isopods and shows their phylogenetic

relationships to other known endoglucanase genes in invertebrates. The monophyly of the putative enORFs with the isopods' endogenous endoglucanase orthologs, including the experimentally verified endoglucanase in *Porcellio scaber* (Kostanjsek et al., 2010; Figure 1, Subclade_1), and *Armadillidium vulgare* (Bredon et al., 2018, 2019), supports the endogenous origin of the enORFs in both subterranean and surface isopods analyzed here.

Based on the branch models, there is evidence that the subterranean species were subjected to a homogenous adaptive shift in selective pressure (a 5-fold change in selection pressure; H_1 hypothesis, Table 2) compared with the ancestral lineage and surface species. This shift in selective pressure could be linked to the evolutionary transition from surface to subterranean environments. It is estimated that this transition by *Paraplatyarthus* species in WA took place between the mid-Miocene and the Pleistocene, coinciding with the onset of aridification in Australia (Javidkar et al., 2018). Furthermore, branch-site models of evolution (Model A, aBSREL, and BUSTED) provided evidence of positive selection for the subterranean species that may indicate novel adaptations to the subterranean environment at the associated amino acid sites. These findings could also suggest that the acquisition of adaptive features and, as a result, the retention of endoglucanase genes in the subterranean isopods, which have persisted over long evolutionary times in calcrete aquifers, may be a key factor for their long-term survival in a potentially low-energy habitat. Although our study detected positive selective signals on the subterranean branches, such signals were not identified for GH9 genes (endo- β -1,4-glucanases (E.C. 3.2.1.4)) in terrestrial isopods (Bredon et al., 2019); the latter found just a single site under positive selection using the program FUBAR.

In contrast to other studies on the expression of cellulase genes in isopods using whole animal bodies, digestive tract, or hepatopancreas, this study utilized head samples to elucidate the likely role of cephalic tissues (e.g., mouth tissues, salivary glands) in the secretion of endogenous cellulolytic enzymes in isopods. The expression of endogenous endoglucanase from salivary glands has also been confirmed in several termite species (e.g., Nakashima et al., 2002). However, to better understand the evolutionary history of the complete set of endoglucanase genes in subterranean isopod species and improve recognition of the role and contribution of different tissues in the expression of cellulase genes, digestive tract, and hepatopancreas tissues also need to be utilized.

The occurrence of multiple enORFs per transcriptome, found for the troglophile (four enORFs) and surface (two enORFs) species in this study, is common in terrestrial isopods (Bredon et al., 2018, 2019). According to Bredon et al. (2019), the number of endogenous cellulase genes (GH9) identified in the digestive tracts of terrestrial isopods ranged from one to 12, reflecting different evolutionary histories. In this study, although the difference in the number of predicted cORFs among the species could be due to the difference in sequencing reads between libraries, based on the BUSCO analyses, the troglobite species, with just a single enORF identified, showed

the same or even a higher BUSCO score compared to those of other assemblies (C+F: 79.6%; Table S1). Also, the medium-level BUSCO scores recovered here may indicate that some cORFs are missing from our prediction analyses. Moreover, considering the different tissue used in this study, a direct quantitative comparison with data from published research is not plausible.

The selection analyses, and previous biogeographic analyses of the subterranean isopods (Javidkar et al., 2018), indirectly suggest that the calcrete aquifers have been fed with lignocellulose hydrocarbons for millions of years. This lignocellulose is likely to come from root systems that penetrate into the calcrete and down to the water table, where they also provide a direct (via root grazing) or indirect (via exudates) source of carbon for other crustaceans, such as Amphipoda and Copepoda, in the groundwater ecosystem; It has been suggested that exudates may be derived from termites in the terrestrial environment (Saccò et al., 2022), but it is also likely that subterranean isopods, by directly feeding on root material, also contribute to energy flows into the groundwater.

AUTHOR CONTRIBUTIONS

Mohammad Javidkar: Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); resources (lead); software (equal); supervision (equal); validation (equal); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Steven J. B. Cooper:** Conceptualization (equal); funding acquisition (lead); investigation (supporting); methodology (supporting); project administration (lead); software (supporting); supervision (equal); writing – original draft (supporting); writing – review and editing (equal). **Nahid Shokri Bousjein:** Data curation (supporting); formal analysis (equal); investigation (equal); methodology (equal); software (equal); validation (equal); visualization (equal); writing – review and editing (supporting). **William F. Humphreys:** Conceptualization (supporting); funding acquisition (lead); investigation (supporting); project administration (lead); resources (equal); supervision (supporting); writing – review and editing (supporting). **Rachael A. King:** Conceptualization (supporting); investigation (supporting); supervision (equal); writing – review and editing (supporting). **Andrew D. Austin:** Conceptualization (equal); funding acquisition (lead); investigation (supporting); project administration (lead); supervision (equal); writing – original draft (supporting); writing – review and editing (supporting).

ACKNOWLEDGMENTS

Special thanks to Ms. Kathy Saint for her help in the laboratory. This project was supported by Australian Research Council Linkage grants (LP100200494 and LP140100555) to ADA, SJBC, and WFH, which included the following partners: The University of Adelaide, Flinders University, Curtin University of Technology, South Australian Museum, Western Australian Museum, Department of Biodiversity, Conservation and Attractions, Bennelongia Pty Ltd, Biota Environmental Sciences Pty Ltd, and Minara Resources Limited.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

All enROF and cORF sequences will be deposited in GenBank upon acceptance (Table 1). The datasets for the phylogenetic inference and selection pressure analyses are provided as Appendix S1.

ORCID

Mohammad Javidkar  <https://orcid.org/0000-0002-0781-8687>

Steven J. B. Cooper  <https://orcid.org/0000-0002-7843-8438>

Nahid Shokri Bousjeijn  <https://orcid.org/0000-0002-2516-8657>

Rachael A. King  <https://orcid.org/0000-0001-8089-7599>

Andrew D. Austin  <https://orcid.org/0000-0002-9602-2276>

REFERENCES

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., & Pyskin, A. V. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*, 455–477.
- Bredon, M., Dittmer, J., Noël, C., Moumen, B., & Bouchon, D. (2018). Lignocellulose degradation at the holobiont level: Teamwork in a keystone soil invertebrate. *Microbiome*, *6*, 1–19.
- Bredon, M., Herran, B., Lheraud, B., Bertaux, J., Grève, P., Moumen, B., & Bouchon, D. (2019). Lignocellulose degradation in isopods: New insights into the adaptation to terrestrial life. *BMC Genomics*, *20*, 1–14.
- Cooper, S. J. B., Bradbury, J. H., Saint, K. M., Leys, R., Austin, A. D., & Humphreys, W. F. (2007). Subterranean archipelago in the Australian arid zone: Mitochondrial DNA phylogeography of amphipods from Central Western Australia. *Molecular Ecology*, *16*, 1533–1544.
- Cooper, S. J. B., Hinze, S., Leys, R., Watts, C. H. S., & Humphreys, W. F. (2002). Islands under the desert: Molecular systematics and evolutionary origins of stygobitic water beetles (coleoptera: Dytiscidae) from Central Western Australia. *Invertebrate Systematics*, *16*, 589–598.
- Cooper, S. J. B., Saint, K. M., Taiti, S., Austin, A. D., & Humphreys, W. F. (2008). Subterranean archipelago: Mitochondrial DNA phylogeography of stygobitic isopods (Oniscidea: *Haloniscus*) from the Yilgarn region of Western Australia. *Invertebrate Systematics*, *22*, 195–203.
- Culver, D. C., & Pipan, T. (2009). *Biology of caves and other subterranean habitats*. Oxford University Press.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008.
- Davison, A., & Blaxter, M. (2005). Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Molecular Biology and Evolution*, *22*, 1273–1284.
- Guzik, M. T., Abrams, K. M., Cooper, S. J. B., Humphreys, W. F., Cho, J.-L., & Austin, A. D. (2008). Phylogeography of the ancient Parabathynellidae (crustacea: Bathynellacea) from the Yilgarn region of Western Australia. *Invertebrate Systematics*, *22*, 205–216.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., & MacManes, M. D. (2013). De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protocols*, *8*, 1494–1512.
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*, 518–522.
- Huelsenbeck, J. P., & Ronquist, F. (2005). Bayesian analyses of molecular evolution using MrBayes. In R. Nielsen (Ed.), *Statistical methods in molecular evolution* (pp. 183–226). Springer.
- Humphreys, W. F. (2006). Aquifers: The ultimate groundwater-dependent ecosystems. *Australian Journal of Botany*, *54*, 115–132. <https://doi.org/10.1071/BTO4151>
- Hüppop, K. (2012). Adaptation to low food. In W. B. White & D. C. Culver (Eds.), *Encyclopedia of caves* (2nd ed., pp. 1–9). Elsevier Academic Press.
- Javidkar, M., Cooper, S. J. B., Humphreys, W., King, R. A., Judd, S., & Austin, A. D. (2018). Biogeographic history of subterranean isopods from groundwater calcrete islands in Western Australia. *Zoologica Scripta*, *47*, 206–220.
- Javidkar, M., Cooper, S. J. B., King, R. A., Humphreys, W., Bertozzi, T., Stevens, M. I., & Austin, A. D. (2016). Molecular systematics and biodiversity of oniscidean isopods in the groundwater calcretes of Central Western Australia. *Molecular Phylogenetics and Evolution*, *104*, 83–98.
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, *36*, W5–W9.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*, 587–589.
- Kostanjsek, R., Milatovic, M., & Srus, J. (2010). Endogenous origin of endo-beta-1,4-glucanase in common woodlouse *Porcellio scaber* (crustacea, isopoda). *Journal of Comparative Physiology B*, *180*, 1143–1153.
- Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., & Usadel, B. (2012). R obi NA: A user-friendly, integrated software solution for RNA-seq-based transcriptomics. *Nucleic Acids Research*, *40*(W1), W622–W627.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*(D1), D490–D495.
- Nakashima, K., Watanabe, H., Saitoh, H., Tokuda, G., & Azuma, J. I. (2002). Dual cellulose-digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki. *Insect Biochemistry and Molecular Biology*, *32*, 777–784.
- Saccò, M., Blyth, A. J., Humphreys, W. F., Middleton, J. A., White, N. E., Campbell, M., Mousavi-Derazmahalleh, M., Laini, A., Hua, Q., Meredith, K., & Cooper, S. J. (2020). Tracking down carbon inputs underground from an arid zone Australian calcrete. *PLoS One*, *15*(8), e0237730.
- Saccò, M., Campbell, M. A., Nevill, P., Humphreys, W. F., Blyth, A. J., Grierson, P. F., & White, N. E. (2022). Getting to the root of organic inputs in groundwaters: Stygofaunal plant consumption in a Calcrete aquifer. *Frontiers in Ecology and Evolution*, *10*, 202.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210–3212.
- Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., & Kelly, S. (2016). TransRate: Reference-free quality assessment of de novo transcriptome assemblies. *Genome Research*, *26*(8), 1134–1144.
- Thurman, E. M., & Malcolm, R. L. (1981). Preparative isolation of aquatic humic substances. *Environmental Science & Technology*, *15*(4), 463–466.
- Watanabe, H., Nakamura, M., Tokuda, G., Yamaoka, I., Scrivener, A. M., & Noda, H. (1997). Site of secretion and properties of endogenous endo- β -1, 4-glucanase components from *Reticulitermes speratus* (Kolbe), a Japanese subterranean termite. *Insect Biochemistry and Molecular Biology*, *27*, 305–313.
- Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution*, *35*, 773–777.
- Yang, Z. (2007). PAML 4: A program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591.
- Zhang, D., Lax, A. R., Bland, J. M., & Allen, A. B. (2011). Characterization of a new endogenous endo- β -1, 4-glucanase of Formosan subterranean

termite (*Coptotermes formosanus*). *Insect Biochemistry and Molecular Biology*, 41, 211–218.

Zheng, J., Ge, Q., Yan, Y., Zhang, X., Huang, L., & Yin, Y. (2023). dbCAN3: Automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Research*, 328, W115–W121.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Javidkar, M., Cooper, S. J. B., Shokri Bousjein, N., Humphreys, W. F., King, R. A., & Austin, A. D. (2023). Evolution of endoglucanase genes in subterranean and surface isopod crustaceans from Central Western Australia. *Ecology and Evolution*, 13, e10552. <https://doi.org/10.1002/ece3.10552>