The wood boring amphipod

Chelura terebrans

Amaia Green Etxabe

Submitted in partial fulfillment of the requirements for the degree of *Doctor of Philosophy* of the University of Portsmouth

Institute of Marine Sciences, University of Portsmouth

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Amaia Green Etxabe

....

For the Green Etxabe family

Acknowledgements

I would not have started on this endeavour had not been for the encouragement and inspiration of Rocio Barrales. Thank you.

To my Emma, Faye and Lauren, thank you for your patience and making sure that I was still alive when I'd been off the radar too long. Your encouragement over the years and the knowledge that you guys are always there for me is invaluable.

Yasmin Guler, there are no words. I couldn't have done it without you; we shared the painfully long days, late nights, weekends and on rare occasions a sunny sampling session (though they were mainly rainy, muddy and finger numbing sessions). Thank you for losing the plot with me.

Stephen your untiring patience, insightful comments and suggestions (though sometimes long and repeated) have been invaluable. You made long days, nights, and day/nights in the lab bearable, but I thank you most for you seemingly limitless enthusiasm for everything, especially when mine waned. Thank you.

To my family I owe everything, your encouragement of my endeavours kept me going over the years. Especially my mother, Idoia, father, Simon and siblings Chris and Katrin for your belief in the completion of my PHd and for being there, regardless of time of day, and doing anything possible to help.

Abstract

Chelura terebrans is a widely distributed wood boring amphipod belonging to the little studied family Cheluridae. Previous studies have hinted that *C. terebrans* belongs to a small number of animals capable of degrading lignocellulose without the aid of symbiotic gut microorganisms. This study utilises a broad range of techniques to gain a better understanding of *C. terebrans* and their ability to digest wood.

Molecular phylogenetic analyses of two Chelurid species largely agree with the current taxonomic organisation of this family. Examination of *C. terebrans* using scanning electron microscopy has offered a better understanding of their digestive system and revealed, with the exception of robust lateralia, few morphological adaptations to accommodate such an unusual diet. This examination also found no evidence of gut-resident microflora. Furthermore, quantitative real-time PCR confirmed the absence of any substantial resident symbiotic extra- or intracellular bacteria in the digestive tract by revealing very low levels of bacterial 16S gene sequences in comparison to the symbiont-containing isopod *Porcellio scaber*.

Despite finding no evidence for resident symbiont gut-microflora, in-gel and *in vitro* enzymatic assays using extracts isolated from the hepatopancreas suggests that *C. terebrans* possesses a considerable repertoire of endogenous enzymatic capabilities useful for the digestion of wood, including mannosidase, β-glucosidase and β-xylosidase, endo-1, 4-β-glucanase and endo-1, 4-β-xylanase, with extracts also possessing mono- and diphenol oxidase activity. Furthermore, mass spectrometry analysis on gel regions presenting high mono- and diphenol oxidase activity detected several proteins belonging to the glycosyl hydrolase family and haemocyanins.

Two transcriptomic libraries were obtained from the hepatopancreases of *C. terebrans* fed on a diet of either beech (*Fagus sylvatica*) or Scots pine (*Pinus sylvestris*). These data provided sequences and the relative abundances of genes thought to be involved

in lignocellulose digestion. In both cases, a significant number of the total ESTs contributed towards contigs corresponding to genes for glycosyl hydrolases and haemocyanins. Furthermore, overall expression of each glycosyl hydrolase suggested variation according to the substrate on which *C. terebrans* were fed. Comparisons of the relative gene expression seen in the *C. terebrans* transcriptome with those found in both the wood boring isopod *Limnoria quadripunctata* and the non boring amphipod *Echinogammarus marinus* offer insight into the genes important for lignocellulose digestion.

This study represents substantial progress in our understanding of how *C. terebrans* digests wood and has also opened up new avenues of investigation by revealing *C. terebrans* as a potential source of novel lignocellulolytic enzymes.

Table of contents   Page

List of Tables   Page

List of Figures   Page

XVI

List of Abbreviations

°C	degrees Celsius
Acrylamide	acrylamide/bis acrylamide
APS	ammonium persulfate
Benzamidine HCl	benzamedine hydrochloride hydrate
BSA	bovine serum albumin
CCA	copper chromium arsenic
cDNA	complementary-deoxyribonucleic acid
cm	centimeter
Cm	circumferential muscle
CMC	carboxymethyl cellulose
Cs	cardiac stomach
dH$_2$O	distilled water
DMF	dimythylformamide
DNA	deoxyribonucleic acid
DPO	diphenol oxidase
DPX	(Dibutyl phthalate *and* Xylenes) DPX Mountant for Histology, Sigma Aldrich PT. Ltd.
EDTA	ethylenedinitrilo tetraacetic acid
EST	expressed sequence tag
ETS	external transcribed spacer
Fc	Food channel
g	gravitational force
GH	glycosyl hydrolase
Gt	gut
HCL	hydrochloric acid
Histoclear	national diagnostics
HMDS	hexamethyldisilazane

Hp	hepatopancreas
HpL	hepatopancreas lumen
HpT	hepatopancreas tissue
hrs	hours
kDa	killodalton
kV	kilovolts
LB	lysogeny broth
LR white	london Resin white
MBSL	meters below sea level
MBTH	3methyl 2benzothiazolin hydrozon hydrochloride hydrate
MeOH	methanol
min	minute
mL	milliliters
mm	millimeters
mM	millimolar
MPO	monophenol oxidase
mRNA	messenger Ribonucleic acid
NBT	nitro blue tetrozonium
nm	nano meters
Ø	diameter
Oe	oesophagus
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
Pc	posterior caeca
PCR	polymerase chain reaction
Pf	Primary filter
pH	percentage hydrogen
PMSF	phenylmethylsulfonyl fluoride
pNP	para-nitro phenol

Potash Alum	potassium aluminium sulphate
Protogel	30% Acrylamide 0.8% Bisacrylamide (37.5:1) Stabilized Solution
RB	Resolving Buffer (4X) 1.5 M TrisHCl, 0.4% SDS, pH 8. National Diagnostics, ProtoGel
RNA	ribonucleic acid
RTqPCR	real time quantitative polymerase chain reaction
s	seconds
SB	Stacking Buffer (4X) 0.5 M TrisHCl, 0.4% SDS, pH 6. ProtoGel, National Diagnostics
SDS	sodium dodecyl sulphate
SE	Standard Error
SEM	Scanning Electron Microscope
TBE	tris borate ethylenediaminetetraacetic acid
TBST	tris buffered saline tween
TEM	transmission electron microscope
TEMED	NNN'N' Tetramethylethylenediamine
TME	Tris - tris(hydroxymethyl)aminomethane
TRIS	tris(hydroxymethyl)aminomethane
U	unit
UK	United Kingdom
V	volts
v/v	volume per volume
w/v	weight per volume
Wf	Wood fragments
Xylene	Xylenes (o-,m-,p- isomers)and Ethyl Benzene
μL	microliters
μm	micrometers

1 General Introduction

Wood is a versatile and widely used material. Within the marine environment wood is often used to make groynes, pilings and piers and is also a construction material used in aquaculture. However, multiple organisms utilise wood for nourishment and a huge financial loss can be incurred resulting from damage caused by such organisms (Dharmaraj & Nair 1980). Its availability and mechanical structure has led to wood being a preferred natural building material for centuries and it is still a major, current resource. The structure of wood is complex but is key to its varied and desirable properties. The strength to weight ratio, hardness and bending strength are all qualities derived from synergy, the inter relationship between many cells. The chemical composition in wood varies between cells, the part of the tree (branch, stem, root), its age, species and geographical distribution (Pereira, 2007).

The main constituents of wood are cellulose, hemicellulose, lignin and various extractives such as fatty acids and resins. Cellulose is a polymer of β -1, 4 linked D-glucose units (Ioelovich, 2008), with the hydroxyl groups tending to form inter- and intra-molecular hydrogen bonds causing the aggregation of cellulose chains into microfibrils (Barnett & Bonham, 2004). In cell walls, cellulose microfibrils (CMFs) are then covered in hemicelluloses, non-cellulosic polysaccharides, to produce a CMF microlamella. The spaces between the CMF lammellae are filled with the amorphous 3D aromatic polymer lignin, which encases the structure and provides rigidity to the cell wall. Lignin is a highly recalcitrant polymer, which limits the accessibility of the cellulose for reaction with cellulases (Eichhorn et al. 2001). It is thought that organisms which utilise lignocellulose for nourishment, either partly or as the entirety of their diet, must at least partially modify the encrusting phenolic lignin compounds to access the cellulose and hemicellulose (Zhang et al., 2007; King et al., 2010).

1

The digestion of such a heterogeneous and recalcitrant material has been studied in a range of terrestrial and marine animals. For a long time it was assumed that only bacteria, protozoa, fungi and plants were able to synthesise lignocellulases, and that herbivorous metazoans used enzymes derived from these organisms, an idea reinforced by many investigations of animals that exploit symbiotic relationships in their digestive system (e.g Martin, 1987; Tanimura et al., 2012). Symbiotic associations have been found in the Teredinidae (Popham and Dickson, 1973, Waterbury et al., 1983; Distel, 2003; Yang et al., 2009) termites (eg. Watanabe & Tokuda, 2010), beetles (Schloss et al 2009) and woodlice (Zimmer & Topp 1998a, 1998b), these associations are thought to aid in digestion to some degree, however, the full extent of their contribution to digestion is still largely unknown. Endogenous cellulases have since been found in invertebrates, however, these enzymes are not sufficient in isolation as lignocellulose is a difficult substrate to degrade, requiring a more complex suite of enzymes than those required for cellulose alone.

A recent study of the wood borer *Limnoria quadripunctata*, has revealed the presence of endogenous lignocellulasic enzymes. This observation as well being interesting from the perspective of this animal's natural history, it has also revealed a new source of enzymes capable of degrading lignocellulose for industrial use (King et al., 2010). The biofuels industry attempts to exploit the lignocellulosic biomass, the world's most abundant renewable material. However, the biorefining process remains economically unfeasible due to the lack of biocatalysts, a situation that has created incentives for the discovery of novel enzymes (Maki et al., 2009; King et al., 2010). To better understand the functions of these enzymes, a basic understanding of the lignocellulose complex is needed.

1.1 The lignocellulose complex

1.1.1 Cellulose

Cellulose is a glucose polymer and is the most abundant polysaccharide on Earth (Mutwil et al. 2008). Glucose is the product of photosynthesis, used in plants as a source of energy, and is the primary component in the building of plant cell walls. Glucose is a simple monosaccharide commonly found in two forms α-glucose and β-glucose (Figure 1.1 A & B).
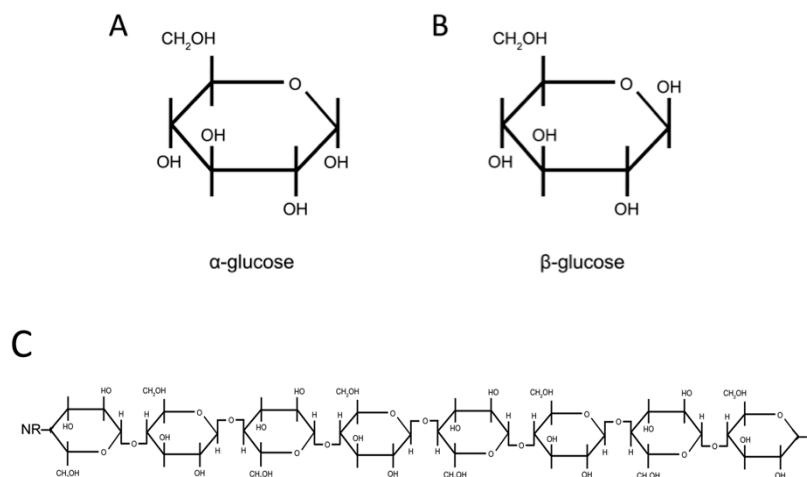
Figure 1.1 Glucose units and polymerisation. A) α-glucose monomer. B) β-glucose monomer. C) β-glucose polymer showing monomers linked by β-(1-4)-Glycosidic bonds (NR- non reducing end, R –reducing end).

α-glucose is usually found as an energy store, in the form of glycogen in animals and as starch in plants whereas β-glucose makes structural material such as cellulose and chitin. In cellulose, the β-glucose units are linked by β-(1-4)-glycosidic bonds that form straight chains of repeating glucose units (Figure 1.1C). The degree of polymerisation is dependent on the location of the cellulose, those found in secondary cell walls are much longer (around 10,000 units) than those found in the primary cell wall (2-6,000). Inter- and intra-molecular hydrogen bonds

form on the hydroxyl group, this strengthens the cellulose chains and promotes the aggregation of the chains into crystalline fibrilar structures, termed microfibrils, that consist of 36 to over 1200 chains and are 3 to 15nm in diameter (Somerville, 2006; Quiroz-Castañeda & Folch-Mallol, 2011; Figure 1.2). Native cellulose (cellulose I) has two phases of crystal cellulose, Iα and Iβ cellulose, the degree of crystallinity and the abundance of each crystal phase differs according to cellulose source. Cellulose Iα can be found in large quantities in algae and bacterial cellulose and a high abundance of cellulose Iβ can be found in wood and cotton.
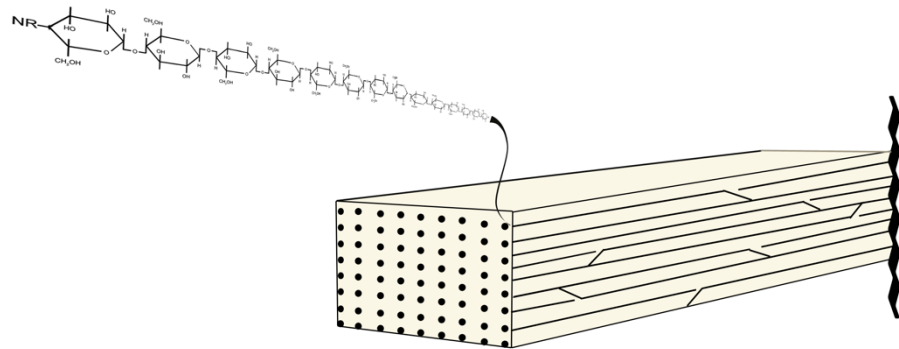
Figure 1.2 Cellulose polymers in a microfibril structure. The straight polymer chains show some areas of dislocation.

The regions of high crystalline structure give great strength and stiffness to wood, with the irregular, amorphous regions (dislocations or slip planes) giving wood its flexibility. However, these dislocations in the regular structure are important for the enzymatic degradation of crystalline structures.

1.1.2 Hemicellulose

Hemicelluloses are heterogeneous polymers consisting of both pentose and hexoses. They are not chemically homogenous like cellulose and the difference in the configuration of these sugars and/or uronic acids (Figure 1.3A & B) makes the formation of hydrogen bonds more difficult and weakens van der Waal forces

between the layers (Walker, 2007). The replacement of the hydroxyl group with others, such as a carboxyl and acetyl groups, makes the formation of a crystalline structure difficult. Hemicelluloses are classified depending on their monomer constituents, the composition of individual hemicelluloses vary between wood species (Figure 1.3C) as does the location in which they are found within the cell. The most abundant hemicelluloses in hard woods are xylans whereas in soft woods hemicellulose is composed mainly of glucomannans (Quiroz-Castañeda & Folch-Mallol, 2011).

Hemicelluloses are branched and there is evidence of bonding with cellulose by hydrogen bonds, and to lignin and pectin by covalent bonds. The role of hemicellulose is complicated, and it has been suggested that it might maintain cellulose integrity in situ, protecting it from cellulases, and it is also thought to be involved in the regulation of cell wall consolidation, playing an important role in determining properties in wood (Beg et al., 2001; Atalla, 2005).
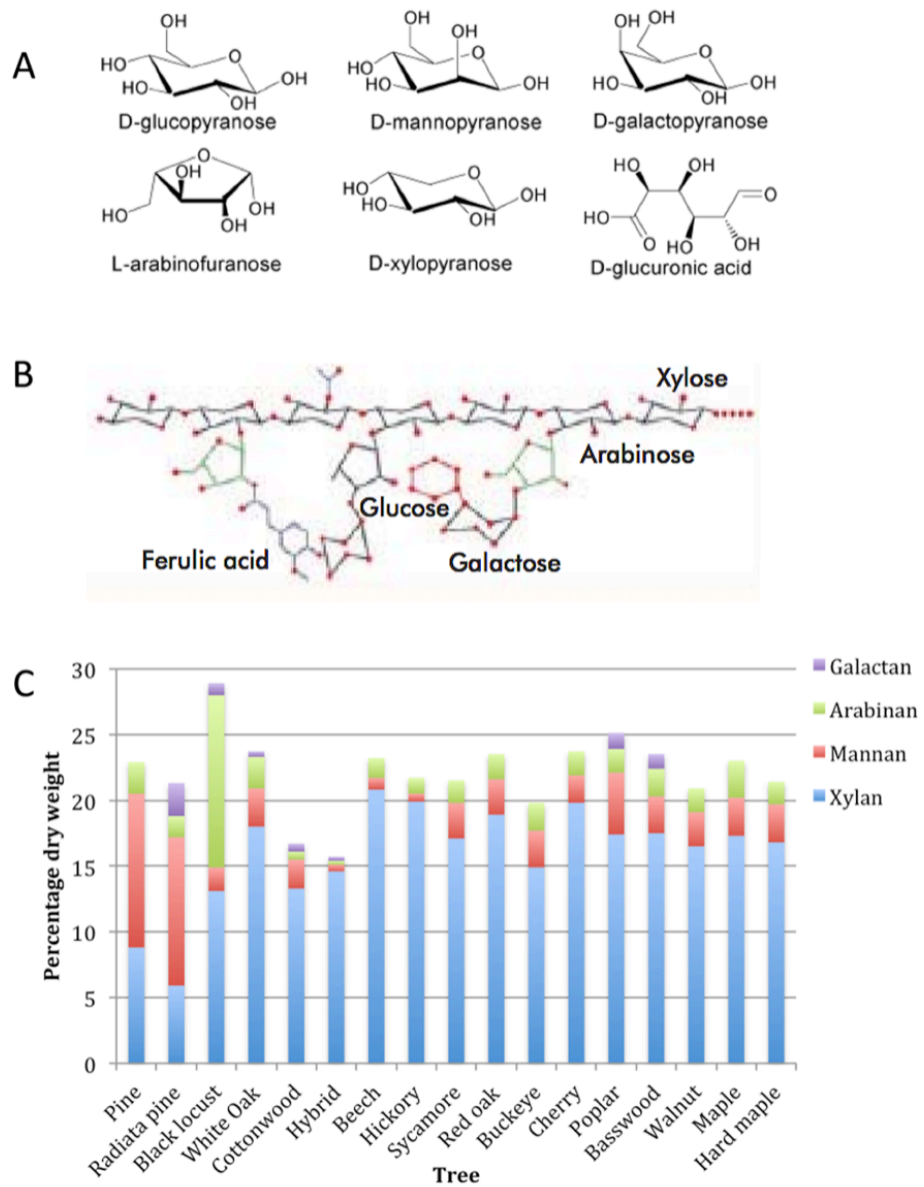
Figure 1.3 Hemicelluloses. A) Important Monomers of hemicellulose (Hansen and Plackett, 2008) B) Example of heterogeneous hemicellulose structure (Quiroz-Castañeda & Folch-Mallol, 2011) C) Percent dry weight compositions of the hemicelluloses of different wood species from North America.

1.1.3 Lignin

Lignins account for 25-35% of the organic matrix in wood. They are highly branched phenolic biopolymer frameworks (Figure 1.4A) with a high molecular mass (600–15000 kDa), which is unique to vascular land plants (Kleinert & Barth, 2008). This polymer is composed of three main monomers, coumaryl-, coniferyl- and sinapyl alcohol (Figure 1.4B), which are found in varying abundance dependent on species and cell types. Softwood species have a high percentage of conifer alcohol (90%), while coniferyl and synapyl alcohols are dominant in hardwoods (Boerjan et al., 2003). Lignin is insoluble and gives strength to cells, is required for plant growth on land, and its hydrophobicity is critical for the movement of water in the conducting vascular elements. Lignin is a resilient polymer that protect the less durable cellulose and hemicellulose from hydrolytic attack (Ruiz-Dueñas & Martínez 2009).
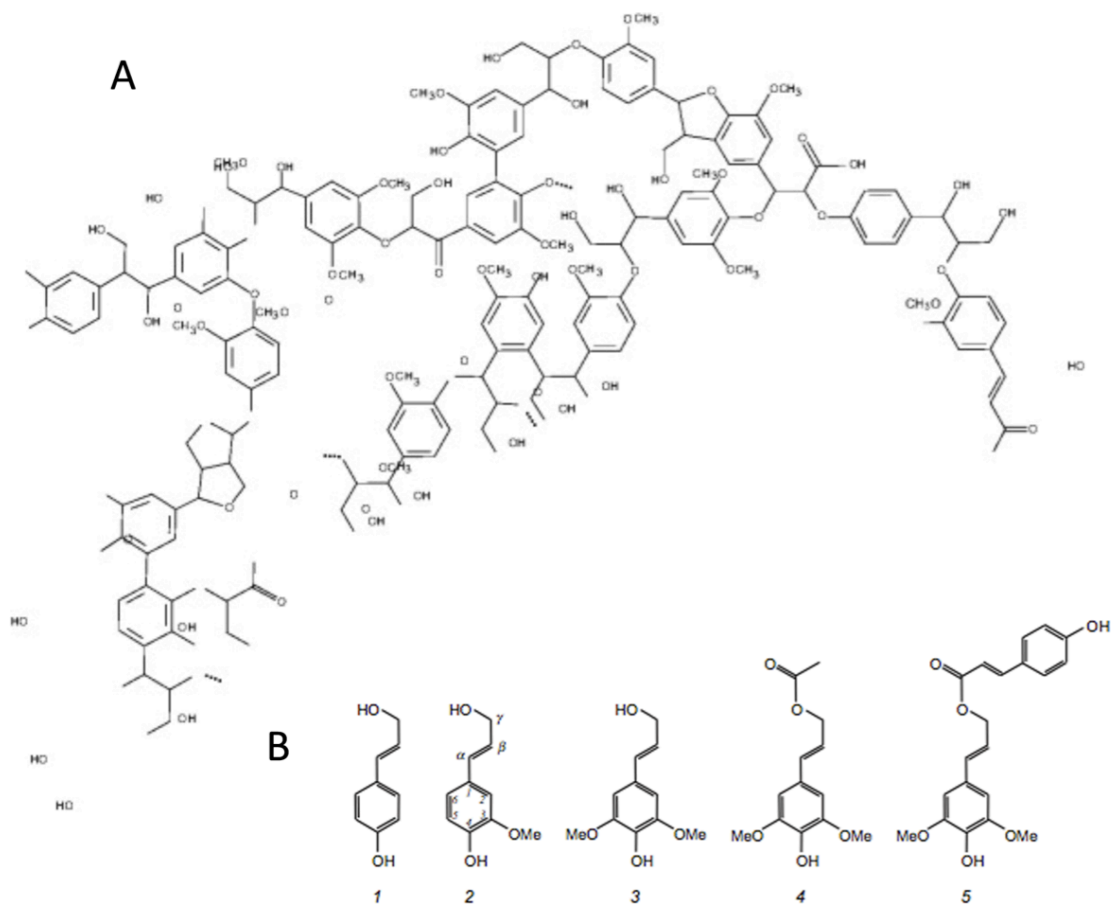
Figure 1.4 Lignins. A) Lignin representation showing branching, heterogeneous monomers http://www.sigmaaldrich.com. B) Three classical and two acylated lignin precursors or monolignols: Classical: 1- p-coumaryl, 2- coniferyl and 3- sinapyl. Acylated: 4- derived from sinapyl alcohol γ-esterified with acetic and 5- p-coumaric acid (Ruiz-Dueñas & Martínez 2009)

1.1.4 Lignocellulose complex

Cellulose and hemicellulose are the major components of wood, representing between 40–60% and 20–40% of dry weight of wood biomass respectively. The remainder is made up of lignin (10–25 % dry weight), extractives and trace amounts of other materials such as pectins and proteins (Lin et al., 2010). It is thought the cellulose microfibrils that are surrounded by hemicelluloses and then

further protected by lignin, which is deposited around them (Panshin and de Zeeuw, 1980). The quantity and distribution of the chemical components in wood is variable within cell type as well as tree species making the complexes difficult to define and study. Furthermore, chemical content and composition is influenced by developmental and environmental conditions in which the tree grows (Pereira, 1988).

1.2 Wood degradation

Wood can last for many centuries under the right conditions, with a range of physical and chemical factors affecting the rate of wood degradation. In the marine environment; organisms are a great contributor and accelerator to this process. Due to the complexity of wood, its biological degradation requires a range of enzymes that work synergistically to break down its various components. In the marine environment, wood degrading organisms such as fungi, bacteria, wood boring molluscs and crustaceans use both physical and chemical mechanisms to degrade wood. One such mechanism employs the use of cellulolytic enzymes.

1.2.1 Wood degradation by enzymatic action

1.2.1.1 *Cellulases*

Cellulase is the general term for enzymes that are capable of degrading cellulose. They are classed as glycosyl hydrolases (GHs) and belong to 20 of the 124 total GH families (www.cazy.org). Cellulases are classified by their characteristics: these include their primary structure, the functional domains present in the enzyme and cleavage sites. Cellulases can be split into 3 groups depending on their enzymatic activity: exoglucanases, endoglucanase and β- glucosidases. Exoglucanases, also known as cellobiohydrolases, liberate cellobiose units from the reducing and non-reducing ends of cellulose chains. Endogluconases hydrolyse β-1-4 linkages at

random, providing more "ends" for the exoglucalases to hydrolyse, while β-glucosidases hydrolyse cellobiose units to glucose monomers (Figure 1.5).

β-glucose   β-glucose   cellobiose   β-glucose

Cellobiohydrolase   Endo β-1,4-glucanase   β-glucosidase   Cutting site   -R   -NR
Exocellobiohydrolase   Endoglucanase   Reducing end   Nonreducing end
Exocellulase
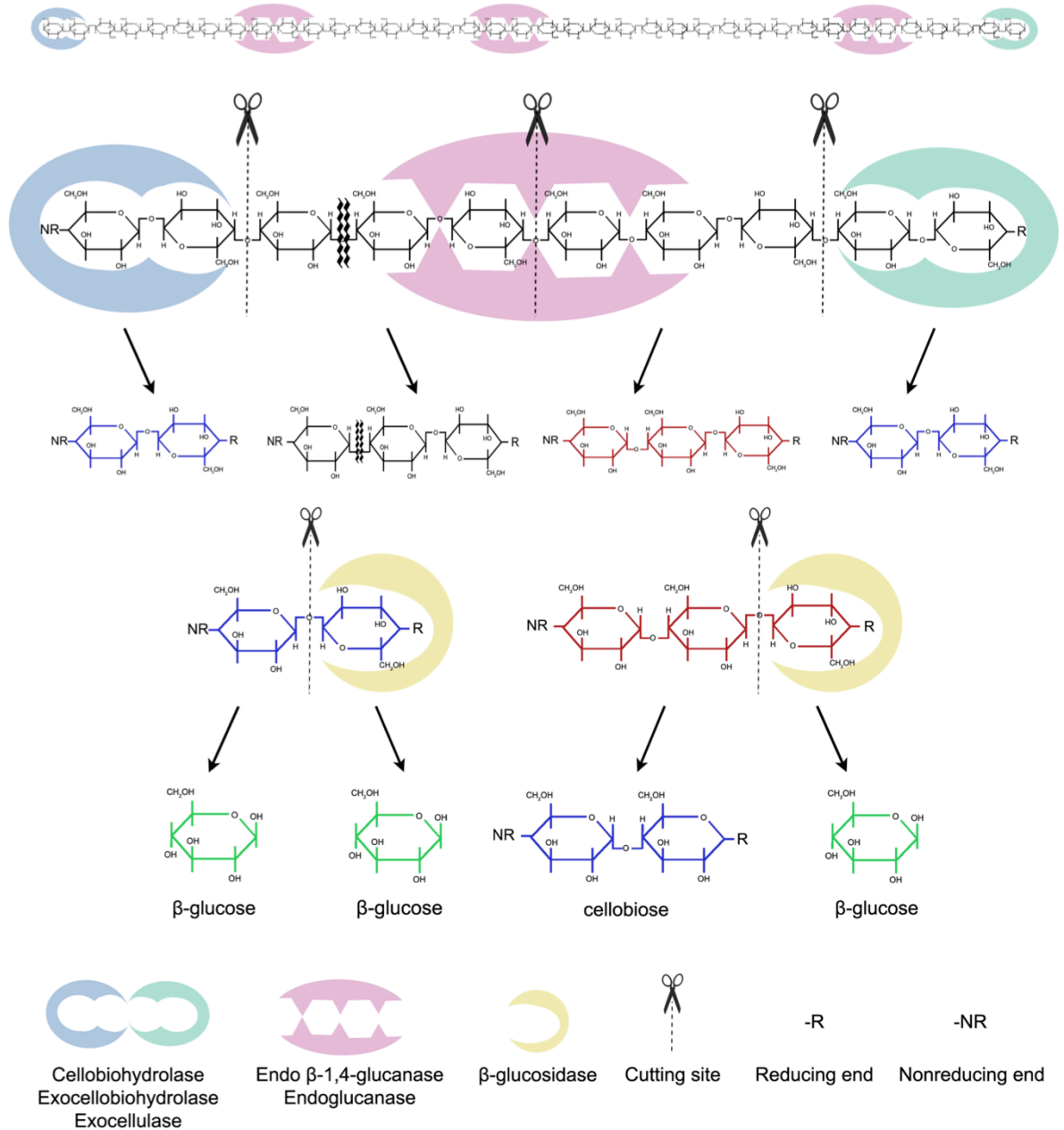
Figure 1.5 Schematic view of the biodegradation of cellulose showing hydrolysis by exo- and endoglucanases and β-glucosidases. Modified from Watanabe (2010).

Although these types of cellulases are produced in their free state in some cellulolytic organisms, others have organised the enzymes in to discrete multi-enzyme complexes, termed cellulosomes (Bayer, 2004), which were first identified in the bacteria *Clostridium thermocellum* (Lamed et al., 1983). The cellulosome is anchored to the S-layer of the bacterial cell wall and attaches to the substrate via a cellulose-binding module (CBM) on the scaffoldin subunit (Figure 1.6). Due to the arrangement of multiple enzymes on a module attached to the substrate itself, it has been suggested that the cellulosome facilitates greater synergism between the enzymes (Fontes & Gilbert, 2010) but it has only, so far, been identified in bacteria**.**
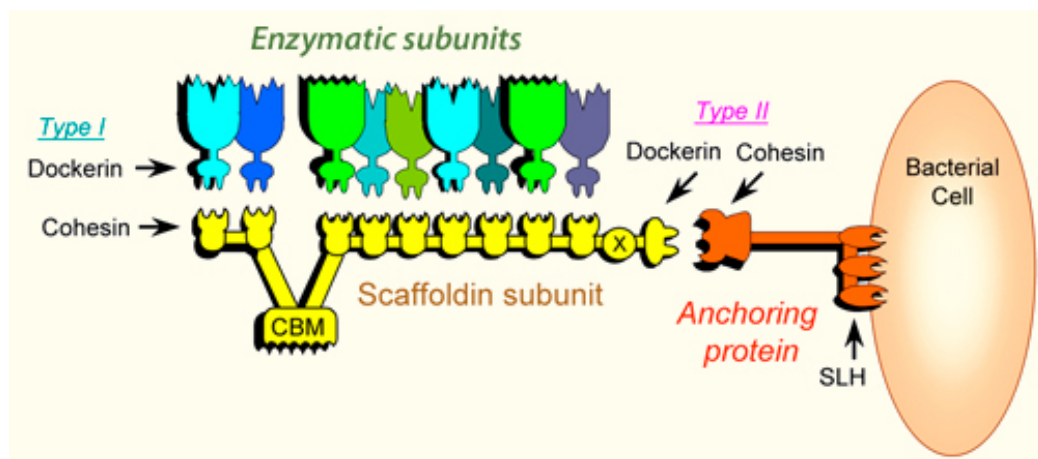
Figure 1.6 Architecture of the *C. thermocellum* cellulosome system from Bayer et al 2009*.* The cellulosome is a muliti-enzyme complex anchored to the bacterial cell, the enzyme subunits and the bacteria attach to the substrate via a cellulose-binding module (CBM) on the scaffolding subunit.

The greater synergism and the processive movement of the cellulosome system is thought to help the enzymes effectively hydrolyse crystalline regions of cellulose (Watanabe, 2010). The large size and diversity of cellulosomes has greatly complicated efforts to probe cellulosome structure and function. The cellulosomes often bind and degrade cellulose, however they contain a wide range of enzymes including hemicellulases.

1.2.1.2 *Hemicellulose and Lignin degradation*

Xylophagus organisms, whether they produce free enzymes or produce cellulosomes, not only contain cellulases but a wide variety of additional enzymes needed to digest more complex heterogeneous polymers found in lignocellulose. The depolymerization of hemicellulose requires carbohydrases and esterases that serve to break the xylan backbone and decouple side-chains that may bind to the lignin components of wood (Figure 1.7, Yang et al., 2009).
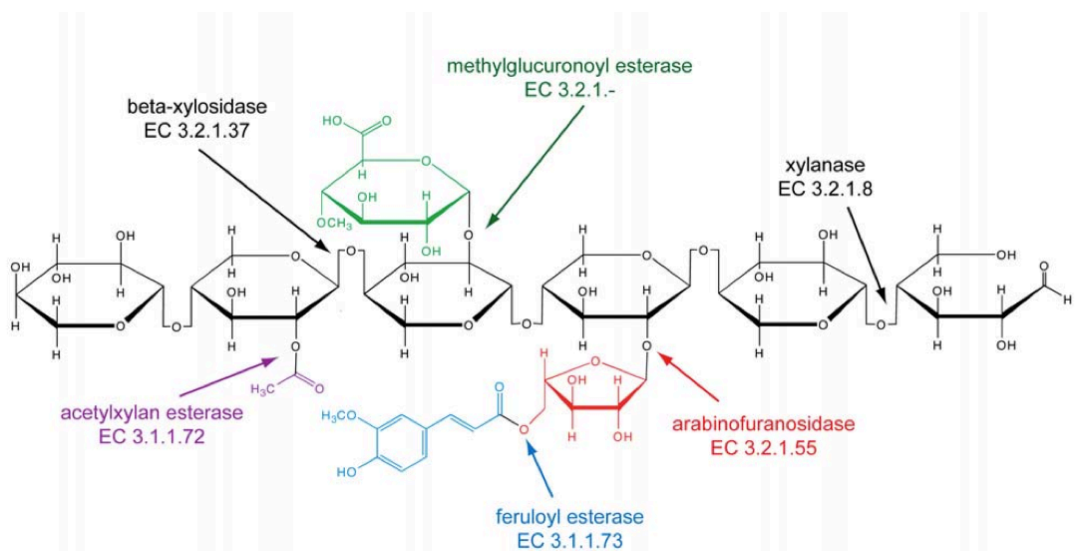
Figure 1.7: Enzymatic components required for the breakdown of a hypothetical hemicellulose with a xylan backbone with corresponding Enzyme commission (EC) number designations (Yang et al., 2009)

Although lignin is recalcitrant to biological and chemical degradation, lignin degradation is a key step in closing the carbon cycle, since its removal enables organisms to utilise other carbohydrates sheltered by the lignin. So far only the white rot fungi, caused by basidiomycota, are able to degrade lignin efficiently by

producing an array of synergistically acting extracellular oxidative enzymes. The major groups of lignolytic enzymes include phenol oxidases (such as laccases) and peroxidases (such as lignin peroxidases, LiPs & manganese peroxidases MnPs). As lignin does not contain hydrolysable bonds, these enzymes must oxidise the aromatic compounds of lignin until the aromatic ring is cleaved, subsequently further degradation is achieved by other enzymes and cofactors. LiPs and MnPs are heme-containing glycoproteins with high redox potentials, which require hydrogen peroxide as an oxidant. LiP degrades non-phenolic lignin units, whereas MnP degrades both phenolic and non-phenolic lignin units (Jensen et al., 1996). Laccases are multicopper proteins possessing mono- and diphenol oxidase activity. The degradation of lignin by laccases is possible in the absence of other ligninases, such as LiP and MnP (Mayer & Staples, 2002). Laccases are capable of oxidizing mediators, which in turn oxidise substrates that would otherwise not be a direct target of the enzyme (i.e compounds with a high molecular mass) therefore extending the substrate range of these enzymes (Torres et al., 2003).

1.3 Microbial wood degradation

Higher filamentous fungi, basidiomycetes, are the dominant wood decomposers and are traditionally categorised depending on the morphological features of the decayed wood and the enzymatic profile. White rot fungi are capable of extensive degradation of lignocellulose whereas brown rot and soft rot fungi preferentially degrade cellulose and hemicellulose (Worrall et al., 1997). Soft rot fungi (such as ascomycetes) are capable of lignin degradation, employing some mechanisms similar to those of white rot fungi (Worrall et al., 1997). However, these fungi have not been studied to the same extent as white-rot fungi so their role in marine environments and mangrove habitats, where ascomycetes predominate, is not

fully understood (Luo et al., 2005). Bacteria capable of lignocellulose degradation are extensively studied and show variety in their modes of attack on wood, ranging from tunnelling, which degrade lignin, to scavenging bacteria that degrade residual materials. Lignocellulasic bacteria are well documented as free-living microbes but also as resident gut flora in animals. For a long time it was assumed that only bacteria, protozoa, fungi and plants were able to synthesise cellulases, and that herbivorous metazoans used the enzymes derived from these organisms, a hypothesis strengthened by many investigations into animals utilising symbiotic systems (e.g Martin, 1987; Tanimura et al., 2012). Although endogenous cellulases have been found in various animals, lignocellulose-degrading animals tend to maintain a close association with lignocellulose degrading microbes (Distel, 2003; Watanabe & Tokuda, 2010; Yang et al., 2009).

1.3.1 Marine wood boring invertebrates

Invertebrates have successfully invaded wood habitats in marine environments such as piers and shipwrecks (Lopez-Anido et al., 2004; Jurgens & Blanchette 2005). Mechanical degradation of wood by borers gives a larger surface area and exposes more native cellulose for microbial activity The combined activities of woodborers and microorganisms can have a damaging effect on marine wood and a methods for controlling them has been a focus of study for some time (eg. Cragg et al., 1999; Schmidt, 2006; Eaton &Hale, 1993; Hochman, 1982; Eaton, 1986). The principal marine wood borers belong to the phyla Mollusca and Arthropoda (Cragg et. al., 1999). They can often be found together using the wood for food and shelter, each leaving distinct marks and causing severe damage.

1.3.1.1 *Mollusca*

Molluscs are capable of attacking both the inner and outer wood layers. Teredinidae and the Pholadidae are bivalve molluscs families and are found to contribute most to wood degradation in the marine environment. Teredinidae, or shipworm, are represented by many genera and species. Shipworms deposit a shell-like material that lines the tunnels into which they bore into using modified shells at their anterior end. These small rasping shells form a pair of abrasive plates that allows the shipworm to burrow into the wood, producing small particles for ingestion and degradation, symbiotic microbes are suggested to aid their digestion (Distel, 2003, Yang et al., 2009). There is much less diversity in the *Pholadidae*, here the main contributors are members from the genera *Martesia* and *Xylophagidae*. Pholads have ridges on their shells that are used for rasping, and are particularly prevalent in tropical waters, although deep-water species are also known to cause extensive damage to wood (Lopez-Anido et al., 2004).

1.3.1.2 *Crustaceans*

It has long been known that wood boring crustaceans are the cause of much marine wood degradation. Major contributors belong to the supraorder *Pericardia*, those include Isopod species from the genus *Limnoria* and *Sphaeroma* as well as species from the Cheluridae belonging to the order Amphipoda. These animals are widespread and are sensitive to environmental factors such as salinity and temperature (Miller, 1926). The *limnoriidae*, also called gribbles, are well known wood borers of around 2-3 mm in length and most are capable of burrowing into wood, however there are algal feeding species. *Sphaeroma* is larger than *Limnoria* and also bores into the wood creating a "honeycomb" of tunnels. They are also hardier and are able to survive extreme salinities and temperatures (Eaton, 1986).

Unlike limnoriids, they are not thought to obtain nutrition from the wood, but are suspension feeders. The amphipod *C. terebrans* is usually found in association with limnoriid spp*.* the reason for their association is unclear.

1.3.1.2.1   Limnoridae

Limnoriids are widespread and can be found in tropical and temperate regions, and it was reported that limnoriid spp*.* ingest wood excavated from their burrows. As many wood boring invertebrates digest wood by maintaining relationships with microorganisms (Martin, 1987), this possibility was also investigated in limnoriids. Boyle & Mitchell (1978) used scanning electron microscopy (SEM) to look at the surfaces of digestive tracts from three wood-boring crustaceans, including *Limnoria quadripunctata* and *L. lignorum*. The failure to observe notable levels of microbiota, in conjunction with a study demonstrating cellulase activity in extracts from *L. lignorum* (Ray & Julian, 1952), lead to the hypothesis that limnoriids might produce endogenous enzymes for the purpose of wood degradation. The digestive gland (hepatopancreas) of *L. quadripunctata* has since been shown to produce transcripts for numerous glycosyl hydrolases, such as those belonging to GH7 and GH9 families, which include cellobiohydrolases and endoglucanases (King et al., 2010)**.**

1.4 Cheluridae

Cheluridae is a family of small marine amphipods that are often found in association with the wood boring isopods *Limnoria* spp. (Coleman & Renz, 2009). The Cheluridae have long been found in marine intertidal wood structures such as those found in harbours. In their review of wood borers Kofoid and Miller (1927) comment on the work of Godfrey Sellius, a distinguished geographical and

historical writer, and his report on the dykes of Holland from 1733. Sellius refers to an organism as 'Springertje', meaning 'small jumper'. He also included two illustrations of the organism in his study, although these were far from detailed (Figure 1.8). From the description of 'Springertje' Kofoid and Miller (1927) were convinced that he was not describing the infamous woodborer *Limnoria* (whose common name in Dutch even today is Springertje). However, they failed to notice its resemblance to *Chelura* unlike Kühne and Becker (1964), who have studied marine borer publications since 1733 and found numerous early references they believe relate to *Chelura,* many reporting on their distinct red colouration and their ability to "jump".

Figure 1.8 Spingertje from Sellius (1733). Above) Animal with red colouration on its back. Below) Animal on wood.

Cheluridae are a small family containing three genera; *Chelura, Tropichelura* and *Nippochelura*, containing total of four species (Figure 1.9). The four species, previously belonging to a single genus, were separated into three genus based on distinctive morphological characteristics by Barnard (1959).

Kingdom	Animalia
Phylum	Arthropoda
Subphylum	Crustacea
Class	Malacostraca
Subclass	Eumalacostraca
Superorder	Peracarida
Order	Amphipoda
Suborder	Gammaridea
Infraorder	Gammarida
Family	Cheluridae

Genus: Chelura   Tropichelura   Nippochelura

Species: *C. terebrans*   *T. insulae*   *T. gomezi*   *N. brevcauda*
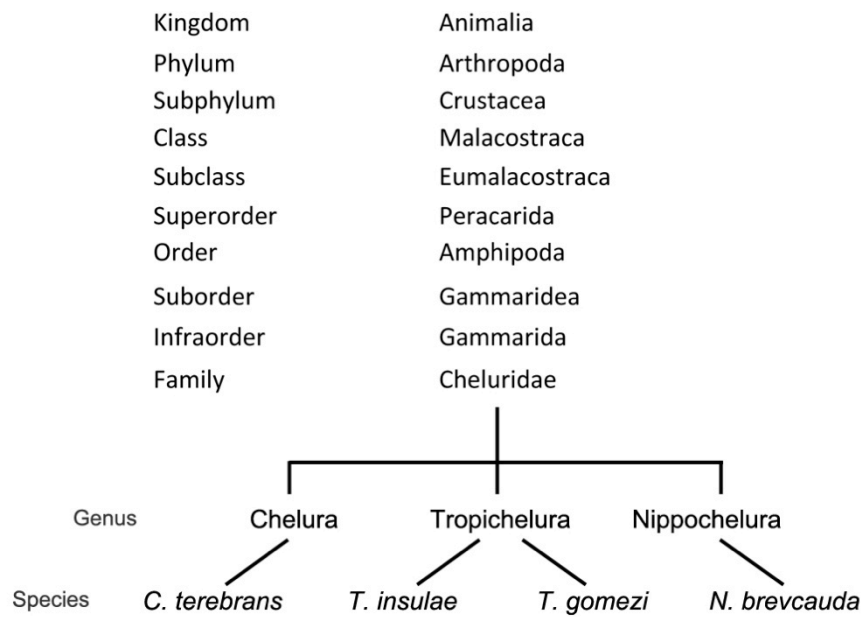
Figure 1.9 Cheluridae taxonomy. The family Cheluridae contains three genera and four species.

They can be found in both temperate and tropical regions (Figure 1.10), and it is thought that the temperature affects the distribution of genera (Barnard, 1959). *C. terebrans* has the most broad distribution and is found in the temperate region (Kühne & Becker, 1964).
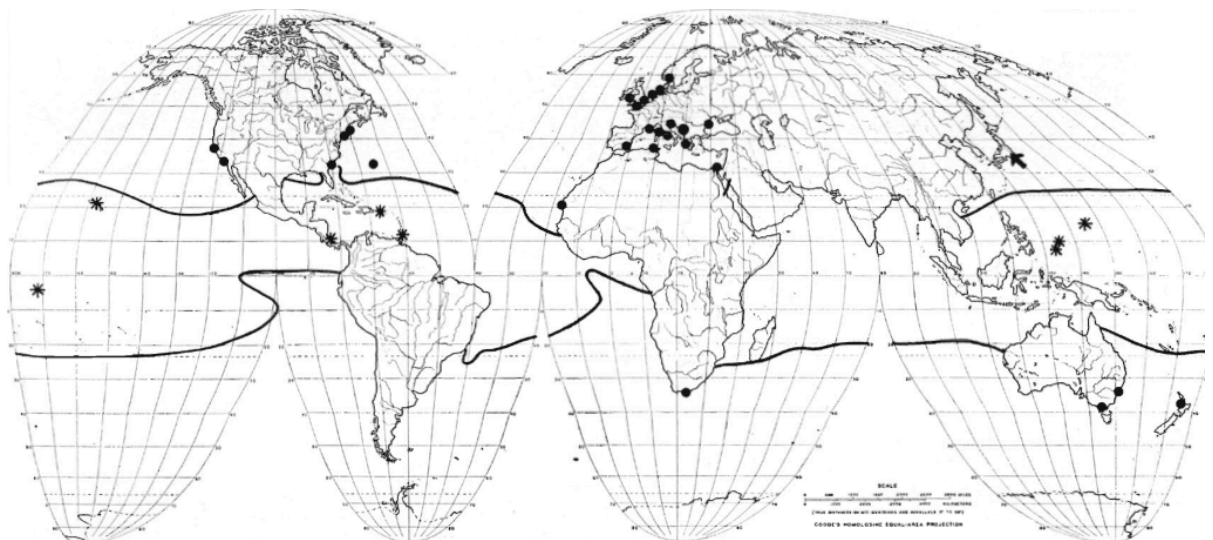
Figure 1.10 World distribution of chelurids in relation to winter isotherms of 22° C. Dots = *Chelura terbrans:* asterisks = *Tropichelura insulae:* arrow = *Nippochelura brevicauda* from Barnard, 1959.

The most extensive study of any Chelurids so far undertaken is that of Kühne and Becker (1964) in a study of *C. terebrans*, which included observations of the main morphological features, feeding habits, reproduction and life cycle. The study also used light microscopy to discern their internal morphology, concentrating on the nervous system, reproductive system and the digestive tract. However, due to the low resolving power of light microscopy, the level of detail provided for many of the structures is limited. Very little work has been performed on the genera *Tropichelura* and *Nippochelura*, and no study on their internal anatomy has been presented.

1.5 *Chelura terebrans*

The external anatomy of is well documented and is largely similar to the general amphipod form with the exception that unlike most amphipods, the body of chelurids is dorsoventrally flattened (Figure 1.11). The males are around 7-9mm (forehead to third uropod) and the females are smaller, measuring 5-6mm. *C. terebrans* are sexually dimorphic, with the morphological differences being more evident in mature individuals. Males posses an extended telson, the second uropods are narrow and long, possessing longer and finer setae along their edges. In contrast females have a shorter telson and second uropod, the uropod is also broader with shortened setae.
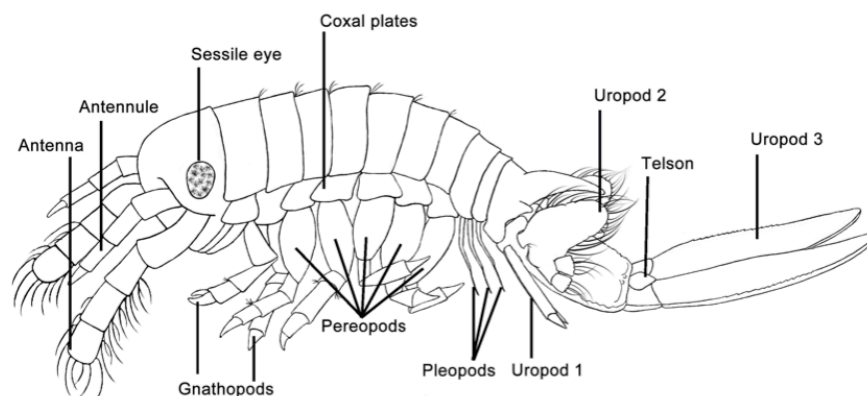
Figure 1.11 Graphical representation of a male *C. terebrans* with narrow long second uropods possessing long setae and extended third uropods.

1.5.1 Wood degradation

Rather than burrowing into wood, like *Limnoria*, *C. terebrans* create troughs in its surfaces, resulting in the creation of large oblique caverns in the wood (Barnard, 1955). The roofs of these caverns often project at the wood surface and are broken off by the action of the sea (Allman, 1847). Settling experiments performed by Kühne and Becker (1964) found that *C. terebrans* preferred soft wood and were unable to excavate a burrow in a short amount of time, confirming the work of Barnard (1955). Yonge (1927) observed that *C. terebrans* enlarges previously excavated limnoriid tunnels but probably would not be able to excavate a tunnel itself unaided. Furthermore, Barnard (Barnard, 1955) proposed that in Nature, *Chelura* would remain exposed on the surface for too long and would risk predation, so would probably require prior activities of limnoriids. However, subsequent studies into the habits of *C. terebrans* have indicated that they are capable of degrading wood without the aid of *Limnoria* (Allman, 1847; Barnard, 1955; Cragg & Daniel, 1992; Kühne & Becker, 1964). Barnard suggests an alternative explanation for finding that limnoriids rapidly enclose themselves in a burrows and *Chelura* do not: 'the wood digestibility of limnoriids may be less efficient than that of chelurids and more wood needs to be consumed by the former' (Barnard, 1955). However the cause and the extent of the dependence of *Chelura* on limnoriids is still open to question

1.5.2 Nutrition

In their experiments Kühne and Becker (1964) found that *C. terebrans* consumes its own faecal pellets as well as that of *Limnoria* spp., and only in part, feeds directly on the wood. Recycling the wood this way, means they are utilising the wood more efficiently than Limnoria (Kühne & Becker, 1964). However, whether

they are consuming the microorganisms with the wood itself or not remains unclear (Cragg & Daniel, 1992). Cragg & Daniel (1992) suggest that the coprophagic tendency of *Chelura* when associated with *Limnoria* indicated that they could benefit from micro-organisms, either through the acquisition of their enzymes or because of the nutritional value of the micororganisms themselves. As a result of their investigations, Kühne and Becker (1964) proposed that no environmental control measures are needed to prevent *C. terebrans* activities in the wild, as they are a "secondary" woodborer 'corresponding to a symbiosis' and their presence in wood is 'not by morphological and physiological adaptation but by means of its special behaviour'.

1.5.3 Symbionts

An interesting observation made following several studies is the apparent lack of micro-organisms in the gut of *C. terebrans* (Boyle & Mitchell, 1978, 1980). Investigations of the digestive tract of *C. terebrans* by light and scanning electron microscopy failed to detect any resident gut flora. These results appear to be in contrast to many other lignocellulose-degrading animals, which rely upon microbial associations to, at least partially, aid in their digestion (Distel, 2003; Watanabe & Tokuda, 2010; Yang et al., 2009). Although it has been demonstrated that *C. terebrans* benefit when fed microbially colonised wood (Kühne & Becker, 1964). However, the ingestion of microbes on the surface of the wood is expected, as the microbes are thought to provide xylophagus animals with a richer source of fixed nitrogen than the wood (Daniel et al., 1991). Indeed *Limnoria spp.* have also been shown to feed on wood with considerable bacterial flora (Daniel et al., 1991).

1.6 Aims of this study

Since the work of Küne and Becker (1964) there have been very few publications on the subject of the Cheluridae. However, information from these relatively few publications has excited questions about the habits and processes of *Chelura terebrans*. The aim of this project is to develop a better understanding of this animal. The following areas will be considered:

1. The genera of the family Cheluridae has been categorised on a solely morphological basis. An investigation using molecular techniques will determine the genetic diversity within Cheluridae and the *C. terebrans* species, and consequently inspect the relationship between morphologically divergent species.

2. Perform a detailed SEM study *C. terebrans* digestive tract morphology. The anatomy of *C. terebrans* has been studied through light microscopy in a study by Kühne & Becker in 1964, however the details of the internal features are not documented. This study aims to provide details of the proventriculus, and make comparisons with the previously detailed internal anatomies of non-wood boring amphipods. This will provide evidence of *C. terebrans* mode of nutrition and reveal any internal anatomical adaptations in a wood boring amphipod.

3. Ascertaining the feeding relationship between *C. terebrans*, *L. quadripunctata* and bacteria. *C. terebrans* is found associated with limnoriids, this study aims to investigate whether *C. terebrans* is reliant on the presence of *L. quadripunctata.* Furthermore, this study will investigate a possible relationship between *C. terebrans* and lignocellulolytic bacteria. As well as giving a greater understanding of the natural history of *C. terebrans*, the discovery of bacterial symbionts, or their lack, will allow an informed assessment of *C. terebrans* as a source of enzymes with potential industrial applications.

4. This study will also aim to investigate the enzymes available to *C. terebrans* and their origin. Assays performed using lysates isolated from specific tissues will indicate the origin and range of enzyme activity in *C. terebrans*.

2   A phylogenetic study of Cheluridae

2.1   Introduction

The family Cheluridae is placed within the amphipod suborder Gammaridea, the infraorder Gammarida, and its four members are split between three genera (World Register of Marine Species (Appeltans et al., 2012). This placement is based on multiple morphological characteristics (Barnard, 1959) as, currently, there are no Cheluridae sequences publicly available from the Cheluidae. Although morphological analysis is still a valuable technique, it has its limitations as morphological differentiation can occur despite genetic similarity. In the same way, genetic diversification can be masked by morphological stasis, making molecular analyses necessary to identify some species (Westram, 2011). For this reason, DNA barcoding was proposed by Hebert et al. (2003) as a complimentary system way of accurately identifying species. This method uses a short genetic sequence from a standard part of the genome, a barcoding region, which allows the identification from any stage of development of known species using a very small amount of tissue (Smith et al., 2012). This also allows for a standardised, rapid and inexpensive method for species identification accessible to non-specialists (Frézal & Leblois, 2008).

For many animal groups, a 648 base pair region of the cytochrome *c* oxidase 1 (CO1) gene is used for species identification, discovery and taxonomy (Bucklin et al., 2011). Whether or not the CO1 barcoding region alone is sufficient for species assignments is controversial. Several problems have been reported suggesting that mitochondrial DNA is an unsuitable marker, these include: symbiont driven changes to diversity (Hurst & Jiggins, 2005), mitochondrial DNA transfer to the nuclear genome, resulting in a non functional copy that gives a false impression of divergence (Buhay, 2009) and inconsistent evolutionary rates among lineages

(Hebert et al., 2003; Bazin et al., 2006; Chu et al., 2009). It has been suggested to be advantageous to use sequences from multiple loci (Neigel et al., 2007). Within arthropods, the CO1 barcoding method is generally used along with other sequences, such as the internal transcribed spacer (ITS) regions, the mitochondrial 16S rRNA gene and the nuclear 18S rRNA gene sequences (Chu et al., 2009; Baird et al., 2011).

This study will aim to acquire sequences useful for the identification of members of the family Cheluridae. These sequences will allow researchers such as molecular taxonomists and ecologists to more conveniently include chelurid species in their studies, giving further insights into the ecology, distribution, diversity of chelurid species. A comparison of these sequences with those of other amphipods will reveal whether the placement of the family Cheluridae within the infraorder Gammarida is reasonable. The analysis will provide insights into the relationships within the Cheluridae, determining whether the level of morphological change is reflected in the sequence divergence. As well as confirming the existing morphological descriptions given by Barnard (1959), this study will investigate the possibility that internal morphological characteristics distinguish members of the family Cheluridae and studies will focus on the foregut structures, especially the lateralia. These are known to be variable even between genera of the Amphipoda (Mekhanikova, 2010) and it has been suggested such structures offer a possible trait for phylogenetic reconstructions (Coleman, 1991).

2.2 Methods

2.2.1 Specimen collection

Specimens were collected from wood in a range of field sites, their location and depth were recorded as coordinates and maximum meters below sea level (max. MBSL) respectively. All collected specimens were placed in 80% ethanol and stored at 4°C until DNA extraction was performed (section 2.2.2.1).

2.2.2 Scanning electron microscopy

Chelurid specimens were dissected and dehydrated in a graded ethanol series before being placed in a 1:1 mixture of ethanol: hexamethyldisilazane, (HMDS, Sigma-Aldrich, UK) for 30 minutes followed by a 100% HMDS solution. This liquid was changed twice and then left overnight to evaporate. The dry samples were mounted on SEM stubs, sputter coated with gold-palladium, and examined using a scanning electron microscope (JEOL JSM-6060LV) in high vacuum mode at an acceleration voltage of 10kV. Secondary electron images were corrected for contrast/brightness and cropped to form montages in an image manipulation program (Adobe Photoshop CS5).

2.2.2.1 *DNA extraction*

Animals were identified under a steromicroscope and blotted dry. The DNA was then extracted from two animals collected from each sampling site using a DNA extraction kit (DNeasy® Blood & Tissue kit, Qiagen) following the 'Purification of Total DNA from Animal Tissues (Spin-Column Protocol)' detailed in the manufacturer's instructions. The quantity of DNA eluted in the 50μl of EB buffer was measured using a spectrophotometer (NanoDrop 1000, Spectrophotometer Thermo scientific).

2.2.2.2 *Primer design*

All the primers were designed using Primer 3 software (Rozen & Skaletsky, 2000) using the default settings, with the exception of the annealing temperature which was limited to 56 - 60°C, and synthesised by Eurofins MWG (www.eurofinsgenomics.eu). The lyophilised primers were diluted in distilled water to give a 100 μM primer stock.

Initially, 'universal' cytochrome c oxidase subunit (CO1) primers described by Folmer et al. (1994) (LCO-1490 5'- GGT CAA CAA ATC ATA AAG ATA TTG G - 3', HCO-2198 5'- TAA ACT TCA GGG TGA CCA AAA AAT CA - 3') were used to amplify the CO1 sequence of *C. terebrans*. However, given the abundant appearance of artifacts (see section Figure 2.6) primers were designed against sequences present in the transcriptomic library of *C. terebrans* (Chapter 6). The CO1 sequence of the amphipod *Gammarus pulex* (GenBank (NCBI) JF965942.1) was used to perform a BLAST search against the *Chelura terebrans* transcriptomic library to identify the equivalent *Chelura* sequence. The retrieved contiguous sequence was then aligned (ClustalX version 2, 2007) with the *G. pulex* sequence in order that primers spanning the equivalent region of the *C. terebrans* CO1 barcoding region could be designed. The resulting primers (ChelCO1 F 5' TCA ACT AAC CAC AAA GAC ATC GG 3' and ChelCO1 R 5' TAG ACT TCG GGG TGG CCA AAA AAT C 3') were used on all specimens. The primers used to amplify a region of the cytochrome b oxidase subunit (Cytb) were designed using a sequence from the *Chelura terebrans* library annotated as cytochrome b oxidase, highest similarity $1.00E^{-150}$ to the cytochrome b from the amphipod *Parhyale hawaiiensis* (AAT69317.1). The resulting primers (ChelCytb – F 5' TTC GGT ACT GTC ACG CAA AC '3 and ChelCytb – R 5' ACA GGG GAT ATG CCA ATT CA '3) were used for all chelurid specimens.

Primers to amplify the chelurid 18S and Internal Transcribed Spacer (ITS) region 1 were designed using ribosomal gene sequences extracted from the *Chelura terebrans* transcriptomic library (Chapter 6). Briefly, the 18S and ITS sequences of the amphipod *Gammarus wilkitzkii* [GenBank (NCBI) JF266608.1 and FJ422963.1 respectively] were used to perform BLAST searches against the *Chelura terebrans* transcriptomic library so to identify the equivalent *Chelura* sequences. The retrieved contiguous sequences were then aligned (ClustalX version 2, 2007) with the *G. wilkitzki* sequences in order that primers spanning the equivalent regions of the *C. terebrans* 18S and ITS 1 regions could be designed. The resulting 18S primers (Chel18SF - 5' AGT AGT CAT ATG CTT GTC TCA AAG ACT 3' and Chel18SR - AAC CGA AGG AGC TTA ACG TG) and ITS primers (ChelITS - F 5' GTA GGT GAA CCT GCG GAA G 3' and ChelITS - R 5' GCG GGT AAT CCC TCC TAA GT 3') were used for all chelurid specimens.

2.2.2.3 *PCR amplification and sequencing*

PCR amplifications were carried out using 10ng of template, 0.27µM primers, 1.25 mM MgCl, 0.2 mM each dNTP, 1x GoTaq PCR buffer, 5U GoTaq® DNA Polymerase (Promega) in volumes of 25µl. The CO1 and 18S amplifications were performed using the following conditions, 5min heated at 95°C followed by 43 cycles 45s at 94°C, 30s at 60°C and 30s at 72°C with a final extension step of 5min at 72°C. The same method was used to amplify the ITS and Cytb sequences using the following amendments: 2 mM MgCl was used, with a 56°C annealing temperature and a 2 minute extension step.

Following amplification, the PCR reactions were cleaned using a PCR reaction clean-up kit (QIAquick PCR Purifiaction Kit, Qiagen) according to the manufacturer's instructions. The purified PCR product was quantified using a spectrophotometer (NanoDrop 1000 Spectrophotometer, Thermo scientific)

before be suitably diluted and sent for sequencing (Source Bioscience) using both the forward and reverse amplifying primers. The sequences were returned in a fasta file format along with the sequence trace in ABI file format permitting the visualization of the sequencing trace and examination of sequence quality scores using the 4Peaks sequence viewer program (version 1.7.1, Griekspoor & Groothuis, 2005).

2.2.3 Sequence alignments

Sequences were aligned using the ClustalW alignment program using the default settings (Larkin et al., 2007) via the EMBL-EBI interface (http://www.ebi.ac.uk). The sequence scores and percentage identities were used to compare levels of sequence similarity. Percentage identities were calculated in LALIGN (Huang & Miller, 1991). The aligned sequences were loaded into Jalview (http://www.compbio.dundee.ac.uk) and annotated according to levels of sequence identity. Tables and figures representing these data were created in an image manipulation program (Adobe Photoshop CS5).

2.2.4 Phylogenetic tree construction

The 18S, ITS, CO1 and Cytb sequences isolated from all samples were aligned using MUSCLE (Multiple Sequence Comparison by Log-Expectation) and a phylogenetic tree was constructed using the maximum likelihood method implemented by the PhyML program (Dereeper et al., 2008). The branching reliability was assessed using the bootstrap method (n = 100).

2.3 Results

2.3.1 Chelurid specimens

Specimens from the genus *Chelura* were collected from two locations around the UK and various locations around the Mediterranean. For comparison samples from the genus *Tropichelura* were collected from the coast of Barbados (Figure 2.1).

Map No.	Species	Country	Location (Closest town)	Depth (Max. MBSL)	Coordinates	
					Latitude	Longitude
1	*C. terebrans*	UK	Portsmouth	5	50.78356	-1.09773
2	*C. terebrans*	Egypt	Sidi Barrani	1145	32.36647	25.70405
3	*C. terebrans*	Croatia	Vis	5	43.00364	16.15952
4	*C. terebrans*	Turkey	Kaş	35	36.20444	29.638981
5	*C. terebrans*	Greece	Kokkari	10	37.77940	26.89007
6	*T. insulae*	Caribbean Islands	Barbados	5	13.15103	-59.43809

Figure 2.1 Locations and information of Cheluridae specimens collected for this study.

2.3.2 External Morphology

The Cheluridae are distinguished by their large urosome and their peculiar uropods (Barnard, 1959). Other characteristics held by this genus include a dorsoventrally depressed body; large urosome with 3 uropod pairs, each pair dissimilar to each other in size and shape. The two species collected in this study are morphologically distinct. *Chelura terebrans* has small gnathopods in comparison to *Tropichelura insulae,* which possess large first gnathopods. Also the third uropods possess small inner rami in *C. terebrans* these are absent in *T. insulae*. The second antennae in *Chelura* are much more setose than that of the other chelurid genus. Species from the genus *Tropichelura* possess a supra-antennal line absent in any of the other chelurid genus. This scanning electron microscopy (SEM) study of the collected of *C. terebrans* and *T. insulae* specimens (Figure 2.2) support the external morphological differences described by Barnard (1959).
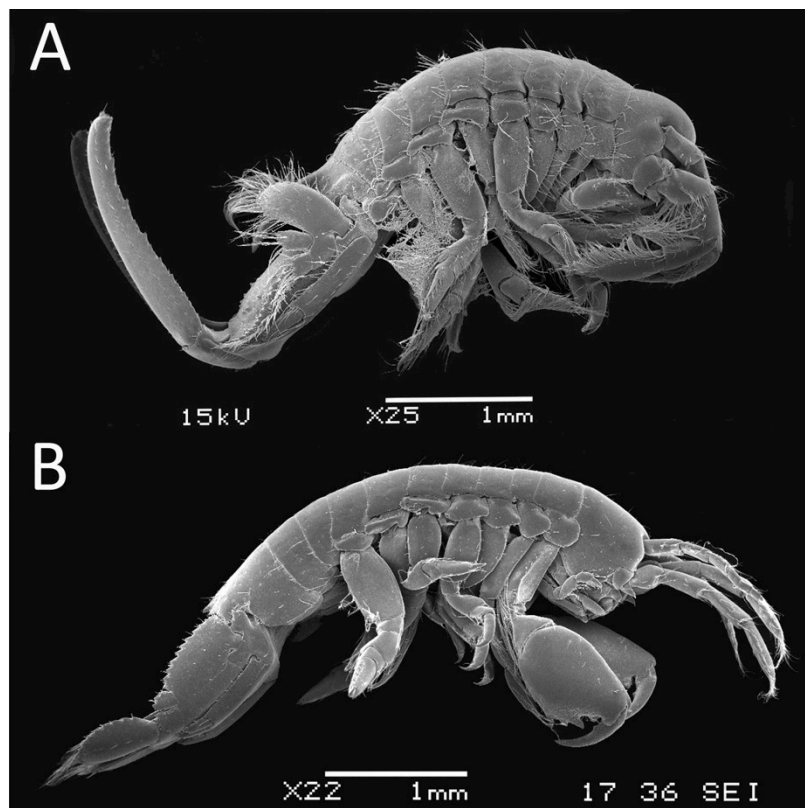
Figure 2.2 External anatomy of *Chelura terebrans* and *Tropichelura insulae* of the family Cheluridae. A) Male *C. terebrans.* B) Male *T. insulae*.

Members of the family Cheluridae are sexually dimorphic. The distinct phenotypes are more evident in the genus *Chelura*, specifically, the first uropods on females are wider, possess shorter setae and the third uropods are shorter in length (Figure 2.3).
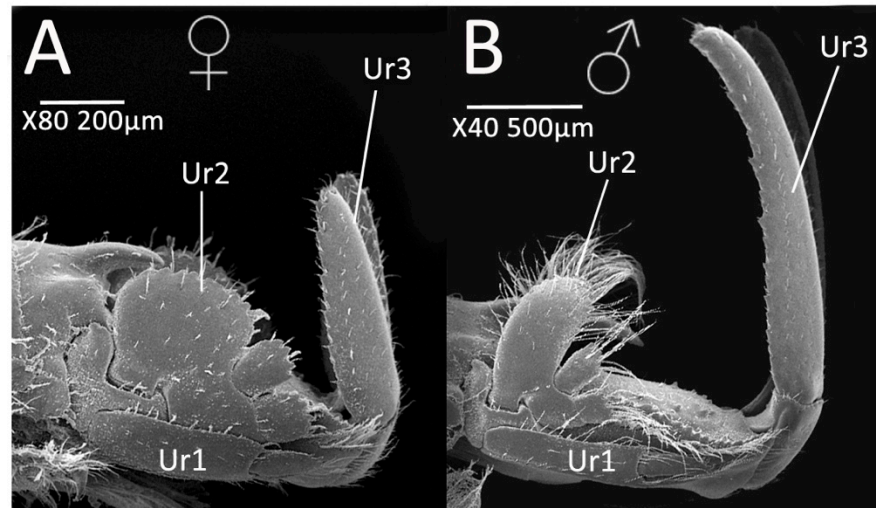
Figure 2.3 Sexual dimorphism in *C. terebrans*. A) Female uropods. B) Male uropods. Ur1 - Uropod1, Ur2 - Uropod2, and Ur3 - Uropod3.

Tropichelura are also sexually dimorphic, although the differences are less striking than *C. terebrans*. In *Tropichelura insulae,* females possess smaller gnathopods than the males, they also posses longer and more profuse setae on the second antennae.
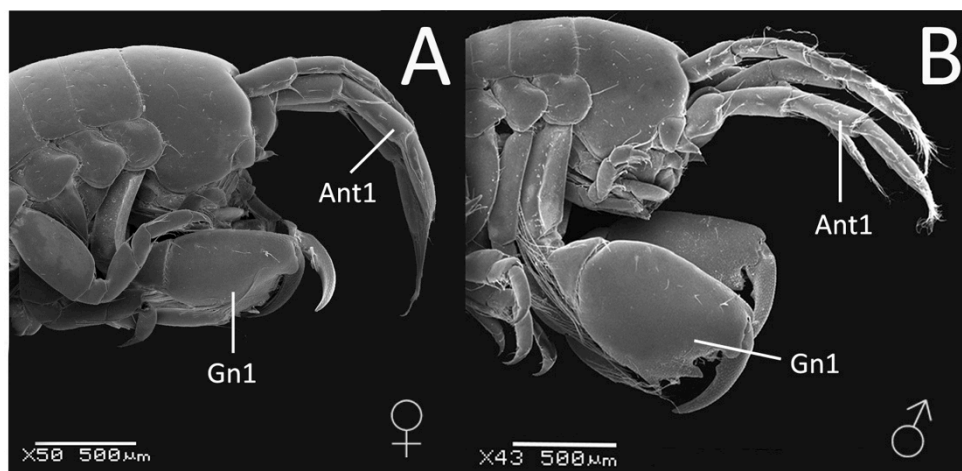
Figure 2.4 Sexual dimorphism in *Tropichelura insulae*. A) Female gnathopods B) Male gnathopods. An1 – antennae Gn1 – gnathopod 1.

31

C. terebrans specimens from each population were examined (Figure 2.1) using SEM, however, no morphological difference could be found between the populations (data not shown).

2.3.3 Lateralia

The lateralia were examined for morphological differences between the species and populations from this study. Although no differences were found between the *C. terebrans* species, the *T. insulae* were found to posses an extra row of plumose setae between the teeth and the setal row seen in *C. terebrans.* Also, a tooth seen on anterior end of the lateralia plates in *C. terebrans* species was not seen in the *Tropichelura* (Figure 2.5). However, it is uncertain whether the tooth is absent in *T. insulae* as it may be covered by the patch of plumose setae present in this area of the lateralia, which is not present in *C. terebrans*.
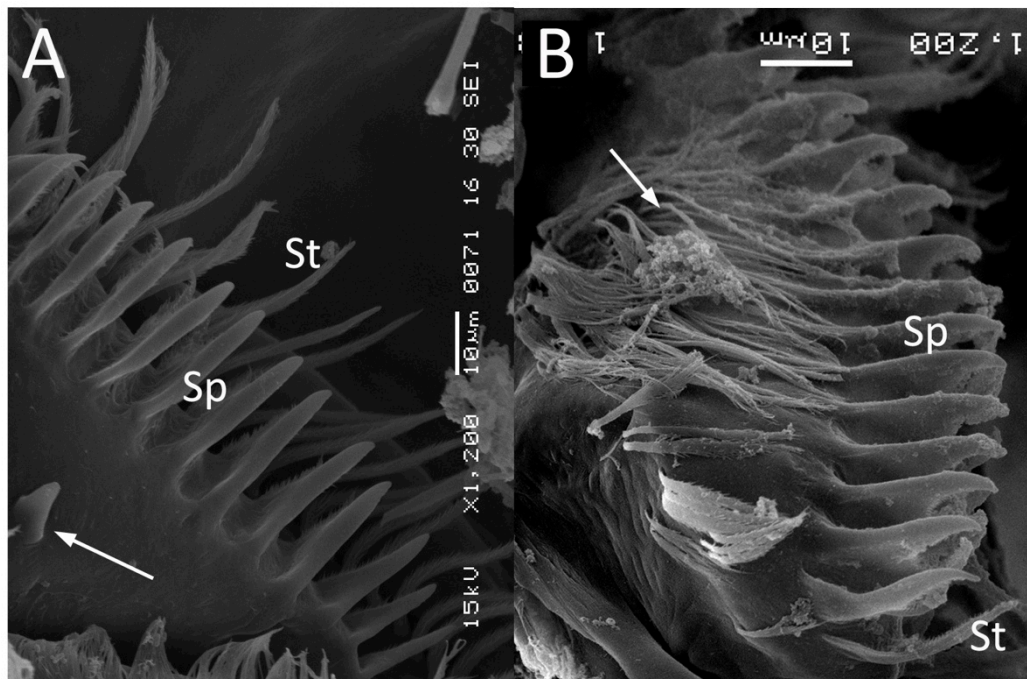
Figure 2.5 Lateralia from *C. terebrans* and *T. insulae*. A) Laterale from *C. terebrans* and B) Laterale from *T. insulae.* Sp – Spines; St – Setea; Arrows indicate area of tooth in *C. terebrans* and eqivalent region in *T. insulae*.

2.3.4 Analysis of the Cheluridae using the barcoding sequence

Initial amplification using previously designed "universal primers" (Folmer et al., 1994) resulted in the production of multiple PCR products (Figure 2.6, middle lane). This is likely the result of the primers causing the amplification of artifact sequences.
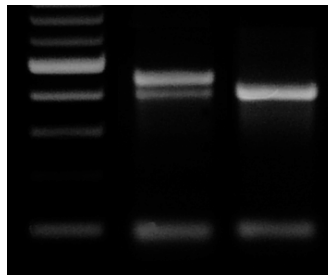
Figure 2.6 Analysis of CO1 barcoding region amplification using agarose gel electrophoresis. Lane 1- 2log DNA ladder; Lane 2 – Products from degenerate primers (Folmer et al., 1994) showing multiple products; Lane 3 – Product amplified using chelurid primers.

To avoid this problem, primers which spanned the equivalent region were designed using sequences available in the *C. terebrans* transcriptome (Chapter 6). These primers produced a single product of the expected size (Figure 2.6, lane 3). CO1 barcoding sequences from Cheluridae can be found in the Appendix.

The CO1 barcoding regions were amplified and sequenced using genomic DNA isolated from all *C. terebrans* and *Tropichelura* specimens. The resulting 658 base pair (bp) sequences were aligned with each other, each pairwise alignment was give a percentage identity (Pi) dependent on the number of base pair changes (Figure 2.7). The alignment of the CO1 sequences was performed in conjunction

with that of *Echinogammarus marinus,* an amphipod from the same suborder Gammaridea.

Number of base pair changes

Percentage identity	C.t UK	C.t Egypt	C.t Croatia	C.t Greece	C.t Turkey	T. insulae	E. marinus
C.t UK		71	70	71	71	2	159
C.t Egypt	89.2		24	23	28	69	172
C.t Croatia	89.4	96.4		15	4	68	169
C.t Greece	89.2	96.5	97.7		19	69	166
C.t Turkey	89.2	95.7	99.4	97.1		69	169
T. insulae	99.7	89.5	89.7	89.5	89.5		159
E. marinus	75.9	73.9	74.3	74.8	74.3	75.8	

Similarity — Highest ——————————→ lowest
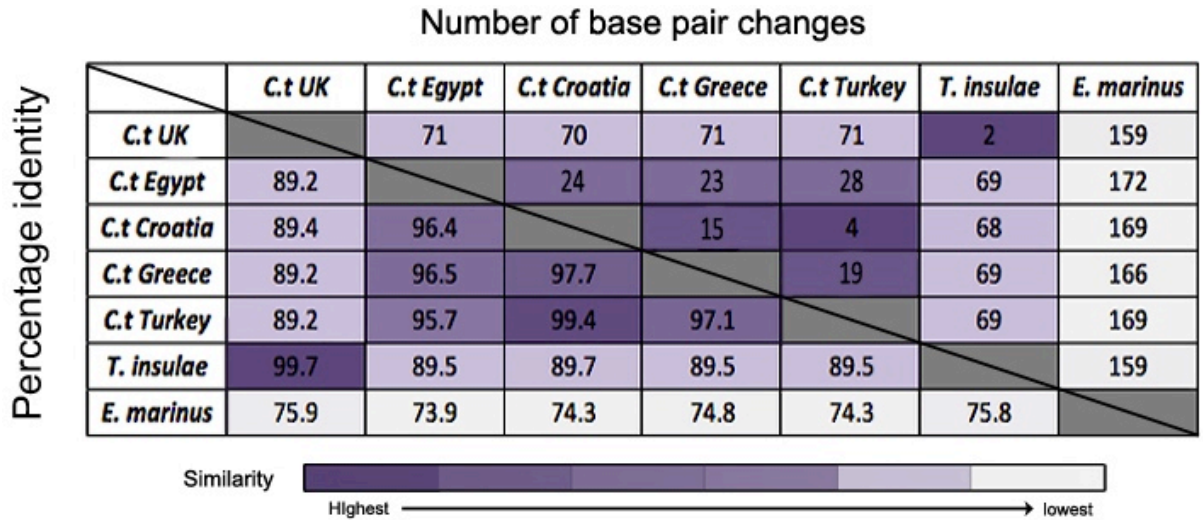
Figure 2.7 CO1 sequence similarity in the family Cheluridae. Numbers below black line give the percentage identity between the two sequences. Numbers above the black line shows the number of base pair changes. The colours highlight the similarity: Dark purple indicates species with the highest similarity and light grey the lowest. *E. marinus* sequence included as an outgroup.

This initial examination shows that the strongest Pi value is between *C. terebrans* UK and *T. insulae* giving a Pi value of 99.7%. *C. terebrans* from Turkey and *C. terebrans* from Croatia present a Pi value of 99.4%. Individual and group sequence alignments are displayed to allow a better visualisation of the sequence divergence (see Figure 2.8, Figure 2.9 & Figure 2.10).

34
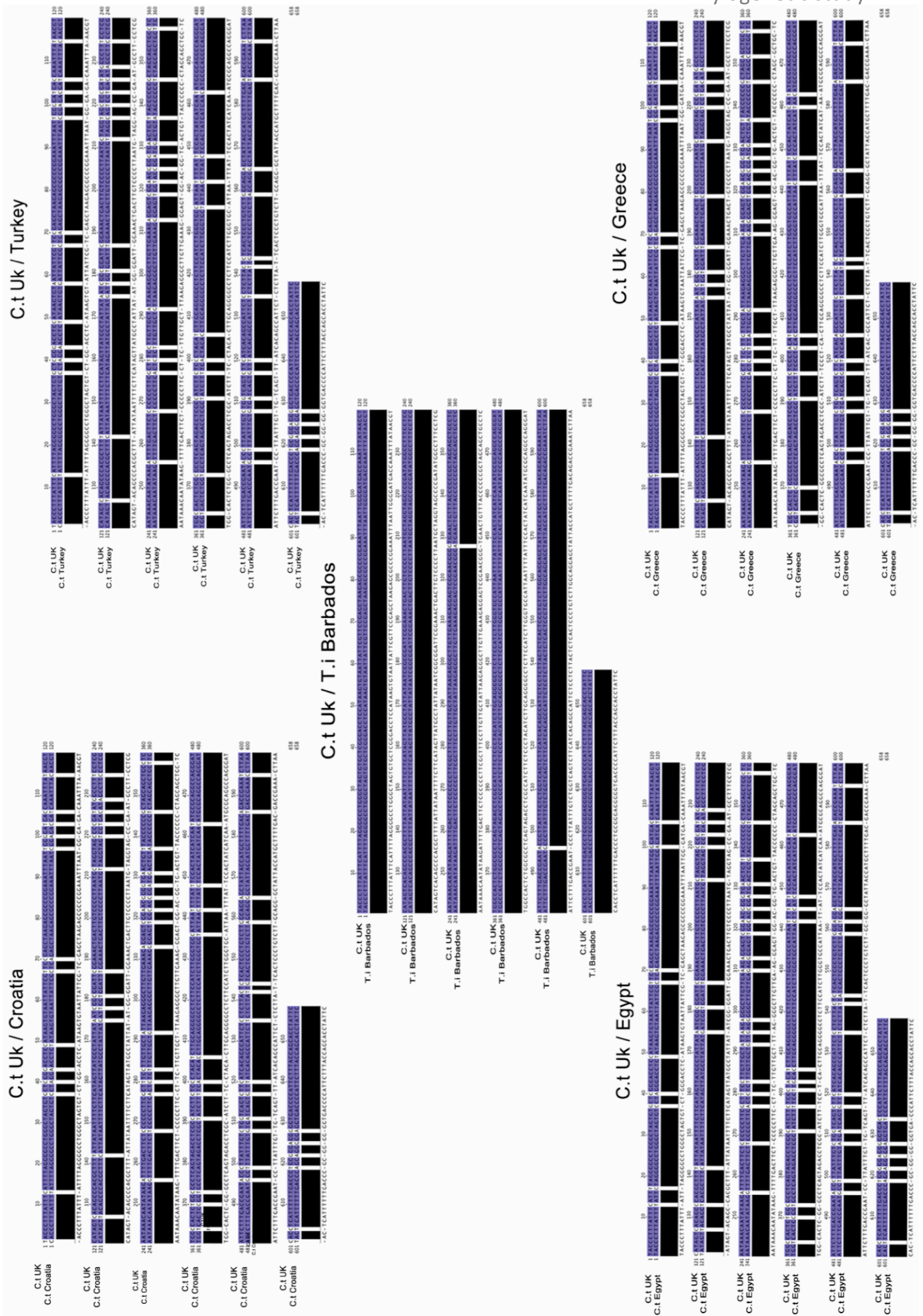
Figure 2.8 Alignments of the CO1 barcoding region of *C. terebrans* from the UK with those of other Chelurid specimens. White lines indicate divergent sequence, black and purple indicate consensus sequence
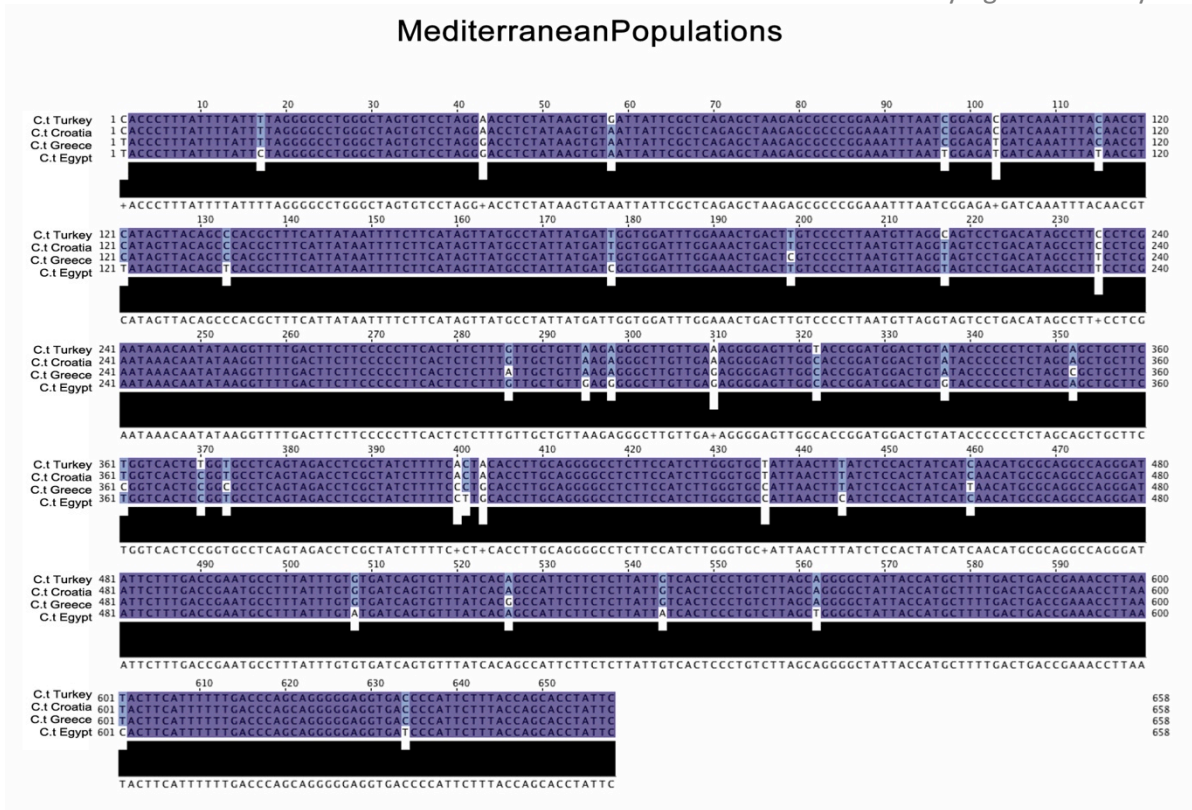
Figure 2.9 Alignments of CO1 barcoding region for specimens taken from Mediterranean populations. White lines indicate divergent sequence, black and purple indicate consensus sequence
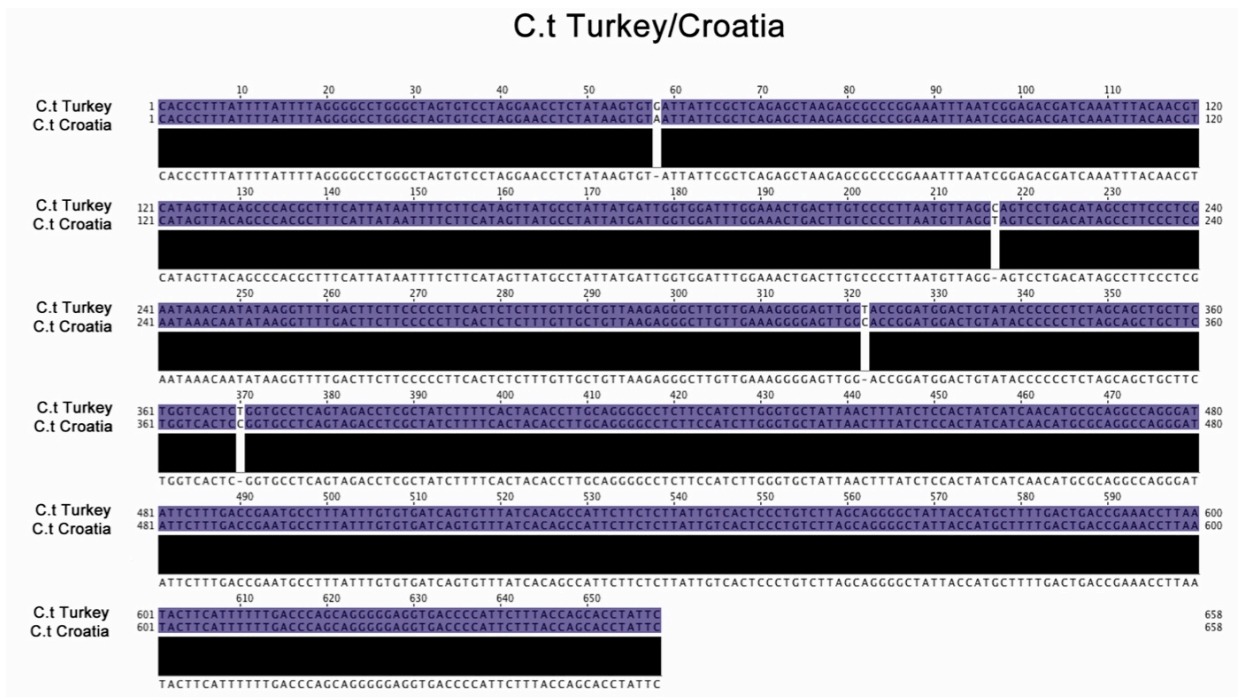
Figure 2.10 Alignments of CO1 barcoding regions of *C. terebrans* specimens from Turkey and Croatia. White lines indicate divergent sequence, black and purple indicate consensus sequence.

There is a discernible difference in CO1 sequences between the *C. terebrans* populations. The CO1 barcoding regions of UK *C. terebrans* samples appear quite different from all the Mediterranean samples, giving a Pi value of around 89%, with 70/71 base pair (bp) differences (Figure 2.8). The Mediterranean samples all appear to share more similar CO1 sequences (Figure 2.9), with the least sequence similarity between samples taken from the Egyptian and Turkish populations (Pi value of 95% and a total of 28 bp changes). The alignments also show that there are only two base pair differences between the UK *C. terebrans* sample and the *T. insulae* sample, giving a Pi value of 99.7%. The divergence of the *C. terebrans* CO1 regions occurs throughout the sequence.

A phylogenetic analysis of the CO1 sequences was performed in conjunction with the CO1 barcoding region of *E. marinus* (GQ341698.1)*.* The analysis of these sequences confirms that the UK *C. terebrans* and *T. insulae* are separated from the other *C. terebrans* samples collected from the Mediterranean. The Mediterranean samples are then split into two groups, Turkey and Croatia samples appearing more closely related to each other than they are to the samples from Greece and Egypt (Figure 2.11).
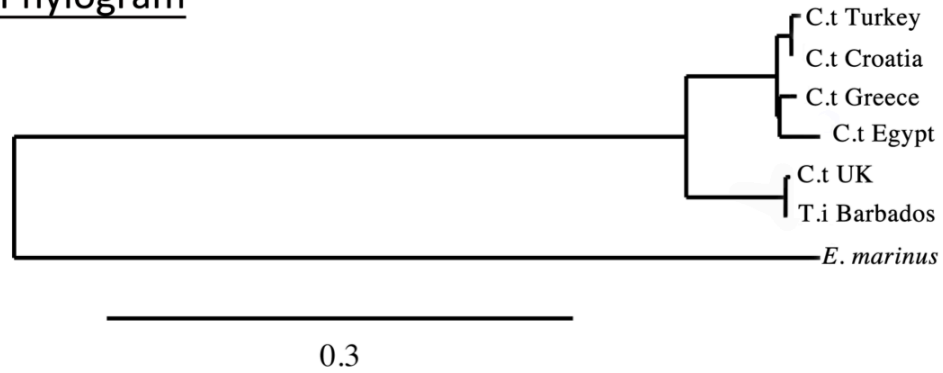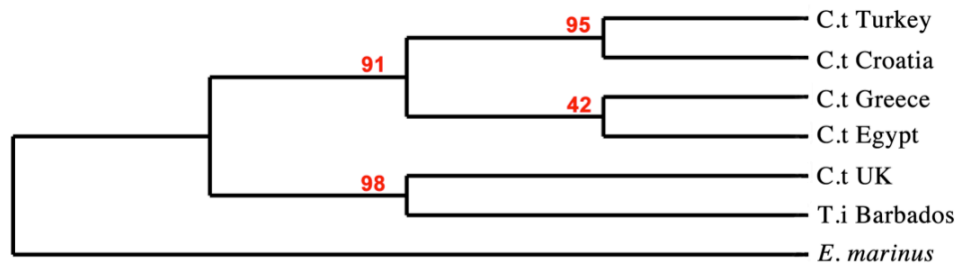
Figure 2.11 CO1 phylogenetic trees created using the barcoding region amplified from Cheluridae samples. CO1 barcoding region from *E. marinus* was used as an out-group. The phylogenetic trees were constructed using the maximum likelihood method implemented by the PhyML program (Dereeper et al., 2008). Phylogram showing branch length proportional to estimated divergence along each branch. Cladogram showing bootstrap values (n=100) for the main branches shown as percentages

2.3.5 Analysis of *the Cheluridae* using Cytochrome b oxidase subunit sequences

To confirm the results of the phylogenetic tree produced using the CO1 sequences, another mitochondrial region, cytochrome b oxidase, was used to perform further analyses (Figure 2.12). The alignment of these sequences was performed in

conjunction with the Cytb sequence of *G. duebeni,* an amphipod from the same suborder Gammaridea.

A

	C.t UK	C.t Egypt	C.t Croatia	C.t Greece	C.t Turkey	T. insulae
C.t Egypt	87.7					
C.t Croatia	87.8	99.8				
C.t Greece	88.1	95.7	96.7			
C.t Turkey	99.9	87.5	87.7	88.0		
T. insulae	92.0	87.5	88.0	88.0	99.4	
G.duebeni	69.3	69.3	69.3	69.0	69.3	69.0

B
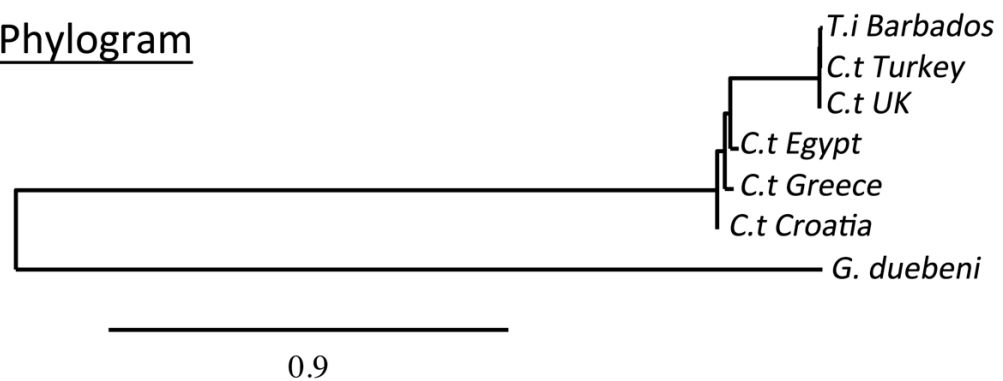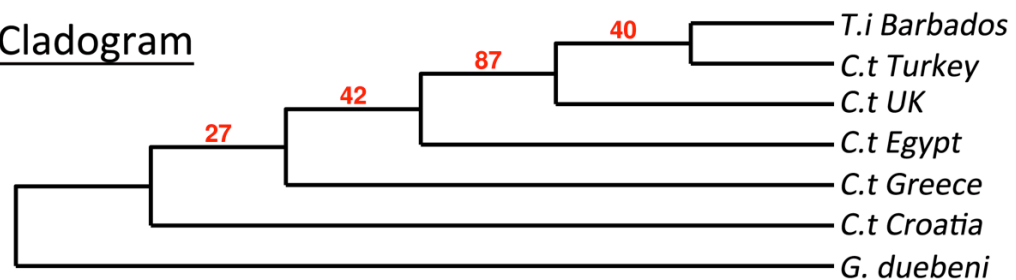
Figure 2.12 Cytochrome b phylogenetic trees created using DNA amplified from Cheluridae samples. A) Sequence similarities in the family Cheluridae*,* dark purple indicates species with the highest similarity and light grey the lowest. B) The phylogenetic trees were constructed using the maximum likelihood method implemented by the PhyML program (Dereeper et al., 2008). Phylogram showing branch length proportional to estimated divergence along each branch. Cladogram showing bootstrap values

39

(n=100) for the main branches shown as percentages. The equivalent Cytb region from *G. duebeni* (JN704067.1) was used as an out-group.

The percentage identities of the Cytb sequences present a different pattern of similarity to those of the CO1 (Figure 2.12). The Cytb sequences suggest that the *C. terebrans* from the UK and Turkey have the highest similarity within the sampled Cheluridae (99.9%). In turn, the UK and Turkey samples show a higher similarity to *T. insulae* than with the other *C. terebrans* sequences. In contrast with the CO1 analysis, the sample taken from Croatia now shows highest similarity to the Egyptian sample, whereas, the CO1 barcoding region suggested that the Croatian sample was most similar to sample from Turkey. The phylogenetic tree created using this mitochondrial gene has positioned *C. terebrans* from Turkey and the UK, along with *T. insulae,* away from the other sequences collected from the Mediterranean (Figure 2.12). This is in contrast with the CO1 sequences, which confidently placed the *C. terebrans* from Turkey with the other Mediterranean samples (Figure 2.11).

2.3.5.1 *18S ribosomal sequence*

Due to the conflicting results suggested by the mitochondrial gene analyses, a 892bp region of the 18S ribosomal RNA gene was amplified and sequenced for all specimens. These were aligned, along with that of *G. wilkitzkii* an amphipod from the same suborder Gammaridea, and a phylogenetic analysis was performed. Initial observations of the 18S sequence scores shows that they appear to be less diverse than those from the mitochondria. Within the *C. terebrans* alignment give sequence scores between 99-100%. In contrast the scores between *T. insulae* and those of the *C. terebrans* species are larger, 91-92% (Figure 2.13).

Figure 2.13 Cheluridae 18S alignments and sequence scores. A) Sequence scores in the family Cheluridae, dark purple indicates species with the highest similarity and light grey the lowest. The equivalent 18S region from *G. wilkitzkii* (JF266608.1) was used as an out-group. B) Alignments of an 18S region from Cheluridae specimens. Alignment scores created in ClustalW using the default settings.

A phylogenetic analysis of the 18S sequences was performed using the equivalent 18S region from *G. wilkitzkii* as an out-group. Consistent with morphological analysis, *T. insulae* is confidently separated from the *C. terebrans* sequences (Figure 2.14).
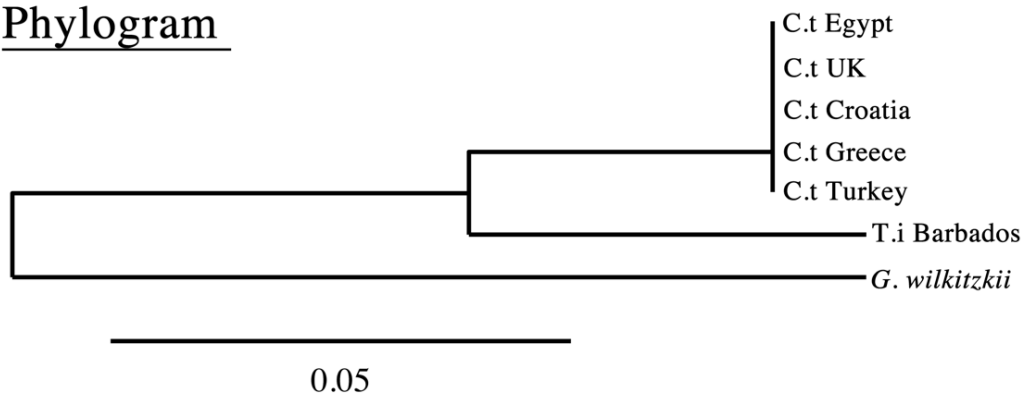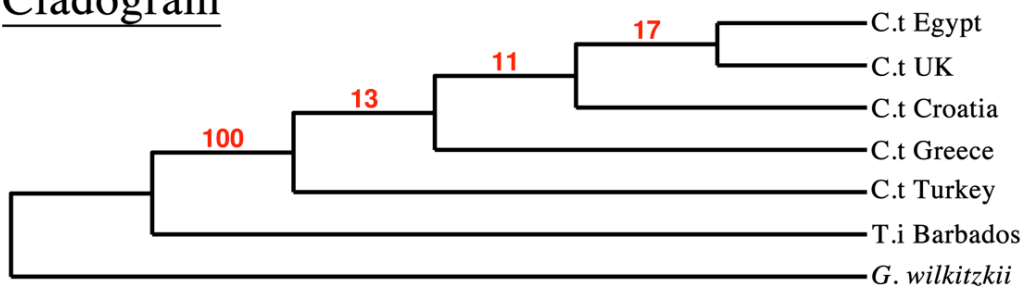
Figure 2.14 18S phylogenetic trees created using DNA amplified from Cheluridae samples. The equivalent 18S region from *G. wilkitzkii* (JF266608.1) was used as an out-group. The phylogenetic trees were constructed using the maximum likelihood method implemented by the PhyML program (Dereeper et al., 2008). Phylogram showing branch length proportional to estimated divergence along each branch. Cladogram showing bootstrap values (n=100) for the main branches shown as percentages.

2.3.6 Analysis of *the Cheluridae* using internal transcribed spacer sequences

To attempt to validate the phylogenetic tree produced using the 18S sequences, another ribosomal region, the internal transcribed spacer region 1 (ITS1), was used to perform further analyses. The alignment of these sequences was performed in conjunction with the ITS1 sequence of *Crangonyx. islandicus,* an amphipod from the same suborder Gammaridea (Figure 2.15).

As seen for the 18S sequences, the *C. terebrans* samples show highest sequence scores with each other, between 87-97. *T. insulae* has a lower similarity score with all *C. terebrans* specimens, between 66-71. The Greek *Chelura terebrans* sample was removed from the ITS analysis due to poor sequence quality, potentially due to the amplification of parasite sequences in the sample.

The phylogenetic analysis of the ITS sequences was performed using the equivalent region from *C. islandicus* as an out-group. Consistent with both the morphological and 18S analysis, *T. insulae* is confidently separated from the *C. terebrans* sequences.

A

	C.t UK	C.t Egypt	C.t Croatia	C.t Turkey	T. insulae
C.t Egypt	97.0				
C.t Croatia	98.0	96.0			
C.t Turkey	89.0	87.0	88.0		
T. insulae	71.0	72.0	72.0	66.0	
C. inslandicus	69.0	68.0	68.0	64.0	66.0

B

Phylogram

0.1
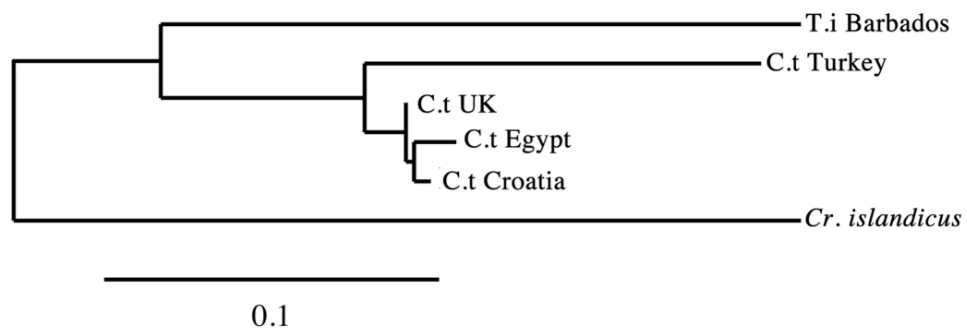
Cladogram
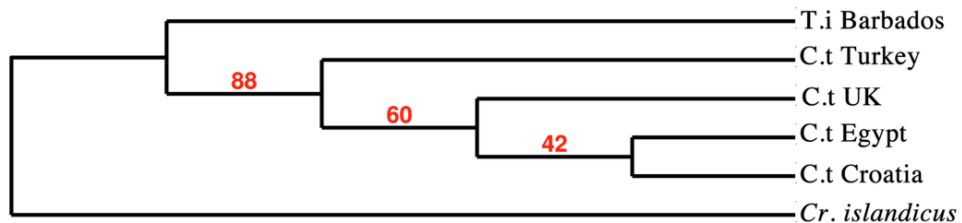
Bootstrap displayed as percentage (N = 10

Figure 2.15 ITS phylogenetic trees created using DNA amplified from Cheluridae samples. A) Sequence similarities in the family Cheluridae, dark purple indicates species with the highest similarity and light grey the lowest. B) The phylogenetic trees were constructed using the maximum likelihood method implemented by the PhyML program (Dereeper et al., 2008). Phyogram showing branch length proportional to estimated divergence along each branch. Cladogram showing bootstrap values (n=100) for the main branches shown as percentages. The equivalent ITS region from *C. islandicus* (JN258064.1) was used as an outgroup.

2.3.7 Cheluridae within the order Amphipoda

The Cheluridae family has been placed within the amphipod infraorder Gammarida, itself placed within the suborder Gammaridea (World Register of Marine Species (Appeltans et al., 2012). As the ribosomal sequences provided consistent phylogenetic trees reflecting the morphology seen in the Cheluridae species, the 18S ribosomal RNA gene sequences were used to investigate whether the Cheluridae placement within the amphipod taxonomic hierarchy is reasonable. A phylogenetic analysis was performed using 18S sequences isolated from the Cheluridae samples and a representative selection from every Gammaridean family publicly available on GenBank (NCBI) (Figure 2.16).
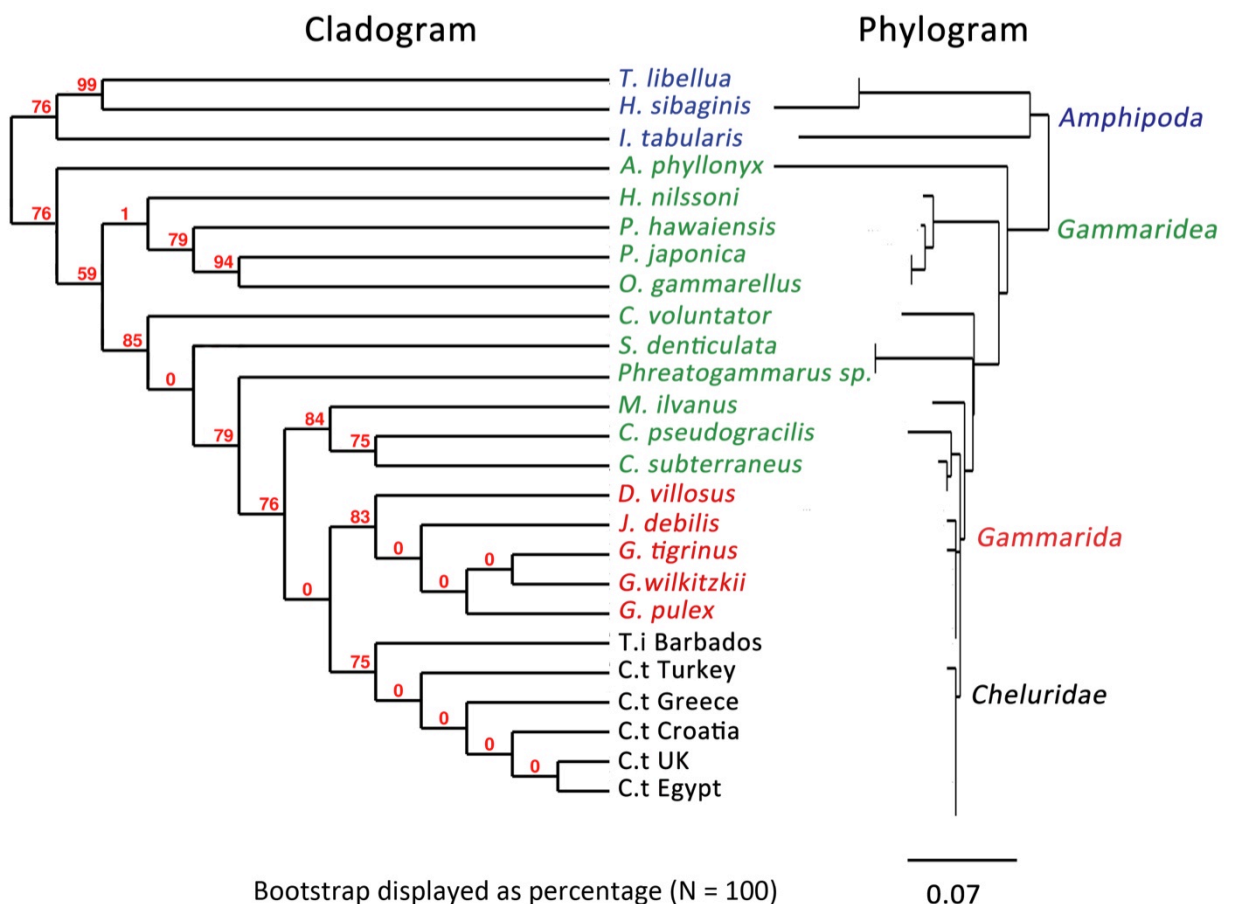
Figure 2.16 Phylogenetic tree created using 18S sequences of amphipod species. Amphipods not in the Gammaridea include *Hyperietta sibaginis* (GU358617) and *Themisto libellula* (JN039368) from the suborder

45

Hyperiidea, and *Ingolfiella tabularis* (DQ378054) from the suborder Ingolfiella. Sequences from the suborder Gammaridea but not in the infraorder Gammarida include *Platorchestia_japonica* (EF582936), *Hyale_nilssoni* (AY826958), *Orchestia gammarellus* (AY826954), *Parhyale hawaiensis* (AY826957) from the infraorder Talitrida, *Corophium volutator* (DQ378027) from the family Corophiidae, *Crangonyx pseudogracilis* (EF582897) from the family Crangonyctidae, *Crangonyx subterraneus* (JQ277470) from the family Crangonyctidae, *Corophium volutator* (DQ378027) from the family Coropgiidae, *Metacrangonyx ilvanus* (HE967296) from the family Metacrangonyctidae, *Arrhis phyllonyx* (AF419235) from the family Oedicerotidae, *Phreatogammarus* sp. (DQ378036.1) from the family Phreatogammaridae and *Salentinella denticulate* (DQ378037.1) from the family Salentinellidae. Sequences from the suborder Gammaridea and in the infraorder Gammarida include *Gammarus wilkitzkii* (FJ422963), *Dikerogammarus villosus* (EF582898), *Gammarus tigrinus* (EF582932), *Gammarus pulex* (EF582923), *Jesogammarus debilis* (EF582934). The analysis also included the 18S sequences isolated from *Cheluridae*. The phylogenetic trees were constructed using the maximum likelihood method implemented by the PhyML program (Dereeper et al., 2008). Phylogram showing branch length proportional to estimated divergence along each branch. Cladogram showing bootstrap values (n=100) for the main branches shown as percentages.

The resulting phylogeny supports the established taxonomic hierarchy of the Amphipoda and the placement of the family Cheluridae within the suborder Gammaridea. Furthermore, the Cheluridae species cannot be reliably separated from species within the infraorder Gammarida suggesting the placement of the Cheluridae family within this infraorder on the basis of morphology is supported by the 18S ribosomal sequence.

2.4 Discussion

Barnard (1959) states that 'the specific contrast in morphological features between each pair of the known [Cheluridae] species' (at the time only one *Tropichelura* species, *T. insulae,* had been identified) 'is of considerable magnitude, greater than found in most genera of other amphipod families'. Indeed this study not only confirmed the numerous external morphological differences between *Tropichelura insulae* and *Chelura terebrans* described by Barnard (1959)*,* but also found a previously unknown internal morphological difference between their lateralia, furthering the differences between the two species. The differences in lateralia between the two chelurid genera is consistant with the extent of differences seen in lateralia between other amphipod genera (Mekhanikova, 2010) and supports the possibility of their use for phylogenetic reconstructions as suggested by Coleman (1991).

The *C. terebrans* samples collected from various locations around the UK and the Mediterranean Sea were also studied using scanning electron microscopy and no external or internal morphological differences were identified between them.

The family Cheluridae have not been incorporated into any published molecular phylogenetic analyses, and no sequences are publicly available for any of its members. This study obtained four sequences typically used for molecular phylogenetic analysis (CO1, Cytb, 18S and ITS1) from *Tropichelura insulae* collected in Barbados and *C. terebrans* species collected from the UK and four locations in the Mediterranean Sea.

Analysis of the sequences obtained in this study found that regions of the 18S and ITS ribosomal gene sequences provided consistent phylogenetic trees reflecting current taxonomy. In contrast, phylogenetic analyses performed using mitochondrial sequences, including the CO1 'barcoding region', did not produce

trees consistent with morphology, the ribosomal sequences or each other. The CO1 sequences isolated from *C. terebrans* collected in the UK and Tropichelura are more similar to each other than they are to the sequences of *C. terebrans* collected in the Mediterranean, a result completely contradicted by the ribosomal analysis.

The Mediterranean CO1 sequences stimulate a range of questions regarding the mechanism responsible for such divergence and are worth further consideration. The transfer of mitochondrial gene sequence to the nuclear genome appears to be a common event in many crustacean groups, including amphipods (Buhay, 2009). As these sequences are no longer coding for functional proteins they are often characterised by mutations resulting in the occurrence of stop codons in the open reading frame, mutations that would not be tolerated in the functioning mitochondrial genome. As non-coding nuclear genes have little or no constraint on their sequence, they can evolve very rapidly, so cannot be used in phylogenetic analyses in conjunction with functioning mitochondrial sequences (Frézal & Leblois, 2008; Buhay, 2009). The translation (using the invertebrate mitochondrial genetic code) of the chelurid CO1 sequences revealed no such stop codons within the sequences. However, the lack of stop codons does not rule out the possibility that these amplified sequences are recently transferred nuclear genes with, thus far, 'minor' mutations. To identify the presence of transferred mitochondrial genes requires the sequencing of a chelurid genome.

It is also worth considering that the chelurid CO1 and Cytb sequences, although variable, are functioning mitochondrial genes. High variability in the CO1 barcoding region has also been found in other arthropods (eg. Astrin, 2006; Whitworth, 2007) showing that the mitochondrial variability in Cheluridae may not be entirely unusual. Indeed, Ballard & Whitlock (2004) and Bazin et al. (2006) have questioned the utility of mitochondrial sequences as they are often under strong selection pressures and evolve under unusual evolutionary rules when compared with other

genomes. This selection could act directly on the mitochondrial DNA itself, but can also arise from maternally transmitted parasites (Hurst & Jiggins, 2005). Inherited parasites are well described in amphipods (Ginsburger - Vogel, 1991; Cordaux, et al., 2001; Terry et al., 2004) however their influence on the evolution of mitochondrial sequences in amphipods has not been explored. Furthermore, it is not known whether parasites infect chelurid species or whether they are responsible for the mitochondrial divergence. Further investigation into the nature and influence of parasites infecting chelurid populations is required.

The collection of multiple *C. terebrans* samples for this study has highlighted the issue of mitochondrial variability in this genus, which a study using a single population might not. Such variability in mitochondrial sequences could cause problems when using them for taxonomy. For this reason, the use of multiple genetic markers, including ribosomal sequences, which appear to give consistent results, should be used for the identification of members from the family Cheluridae until further analysis of the barcoding region in more specimens is completed. Samples from multiple locations are required to establish the extent of mitochondrial gene divergence in this family and whether they correlate to particular influences.

Given the findings in this study it would appear that ribosomal sequences provide a more useful tool for identification of chelurid genera. However, after further study to fully establish the extent of mitochondrial sequence variability in *C. terebrans*, the CO1 gene may prove more effective if attempting to compare *C. terebrans* populations in more localised studies.

3 Functional anatomy of the digestive system in *Chelura terebrans*

3.1 Introduction

The digestive system is well described in amphipods (e.g. Harrison, 1992; Agrawal, 1967; Thiem, 1942; Martin 1964; Icely & Nott 1984; Coleman 1992, 1994), however, the understanding of the more complex proventriculus (stomach) is less detailed (Strus & Storch, 2004). In certain groups of amphipods with specialised diets, the proventriculus is the site of the greatest variation in the digestive system (Coleman, 1994). Strus & Storch (2004) and Coleman (1994) compared the digestive systems of ecologically distinct amphipods and observed that the ultra structure and morphology were very similar. However, differences were found in the lateralia, structures involved in mechanical breakdown of food, and the filter systems found at the anterior of the proventriculus.

Coleman (1992) suggests the armature of the lateralia in certain groups is related to feeding strategies. Although the diet of *C. terebrans* is unclear, they were found to degrade and ingest wood either directly or indirectly through eating faecal pellets (Kühne & Becker, 1964; Cragg & Daniel, 1992). Given the apparent xylphagous diet of *C. terebrans*, an investigation into the anatomy of the proventriculus could reveal specialised structures involved in the processing of food, and could give indications to the extent of specialisation.

The general anatomy of *C. terebrans* was documented by Kühne & Becker (1964) using light microscopy. However, although their analysis of the digestive system was extensive, they were only able to provide a rudimentary description of the proventriculus due to the limitations of the techniques available.

3.1.1 General anatomy of C. terebrans

The external anatomy of *C. terebrans* is well documented and the description of mouth parts and external limbs have been extensive (such as: Kühne & Becker, 1964; Allman, 1847; Bourdillon, 1958). This work describes the basic anatomy of *C. terebrans* as broadly similar to that of other gammaridean amphipods. However, they have some unusual features. These include their body being dorso-ventrally flattened and an extended third pleonite and uropods. *C. terebrans* are described as pale yellow/orange in colour, with brighter red "veins" on the head and along the dorsal median line of the animal, no colour morphs are documented for. *C. terebrans* are sexually dimorphic in a number of anatomical features, including the antennae, which are more setose in males, and particularly in their uropods. The second uropods are narrower, longer and possess more and longer setae in males than females. The third uropod is considerably extended in males and represents the most obvious distinction.
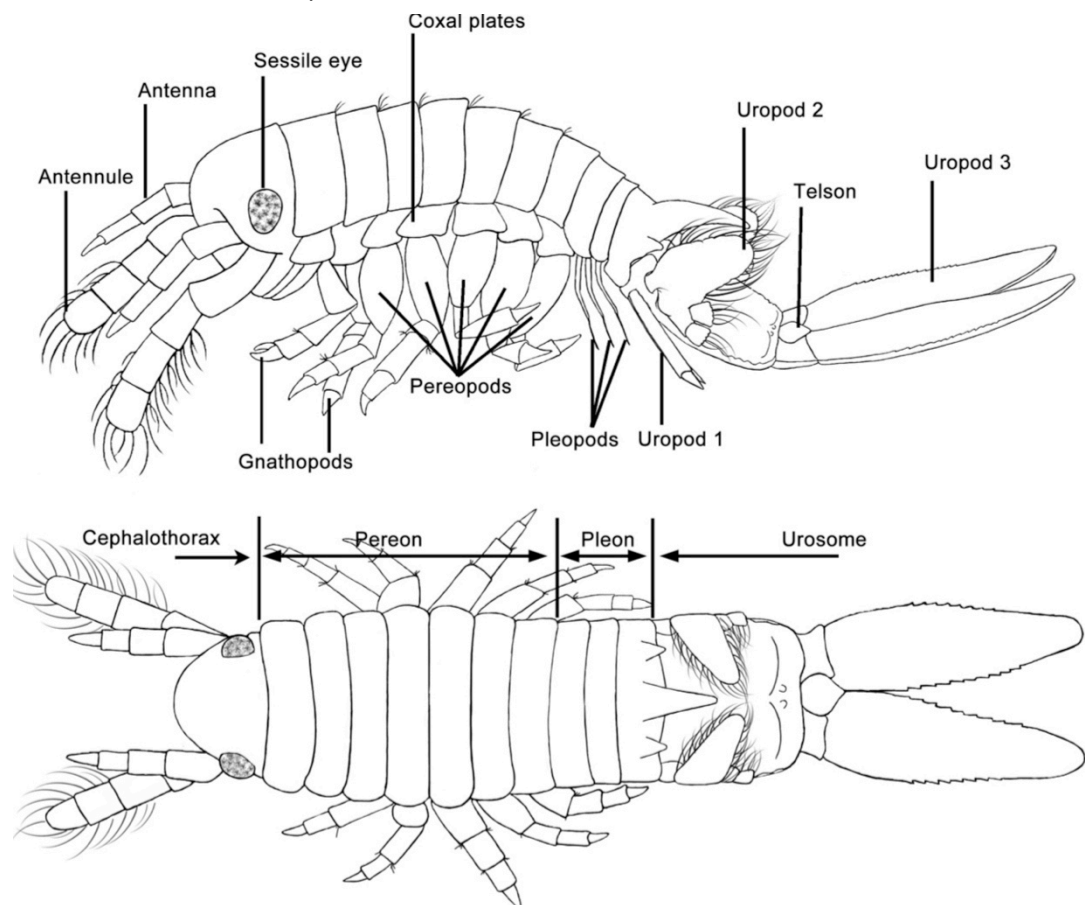
Figure 3.1 The basic external anatomy of a young male *C. terebrans*.

Kühne & Becker (1964) gave a general representation of the digestive system of *C. terebrans* which suggests that, like the external anatomy, in general, the digestive system is much like that of other amphipods. A short oesophagus runs dorsally half way up head where it meets the proventriculus. Here, two rostral paired diverticular (or hepatopancreas) originate on the ventral side of the proventriculus and extend to the urosome. The proventriculus opens into the gut, which extends all the way to the anus, two unpaired posterior caeca extend laterally from the posterior end of the gut (Figure 3.2).
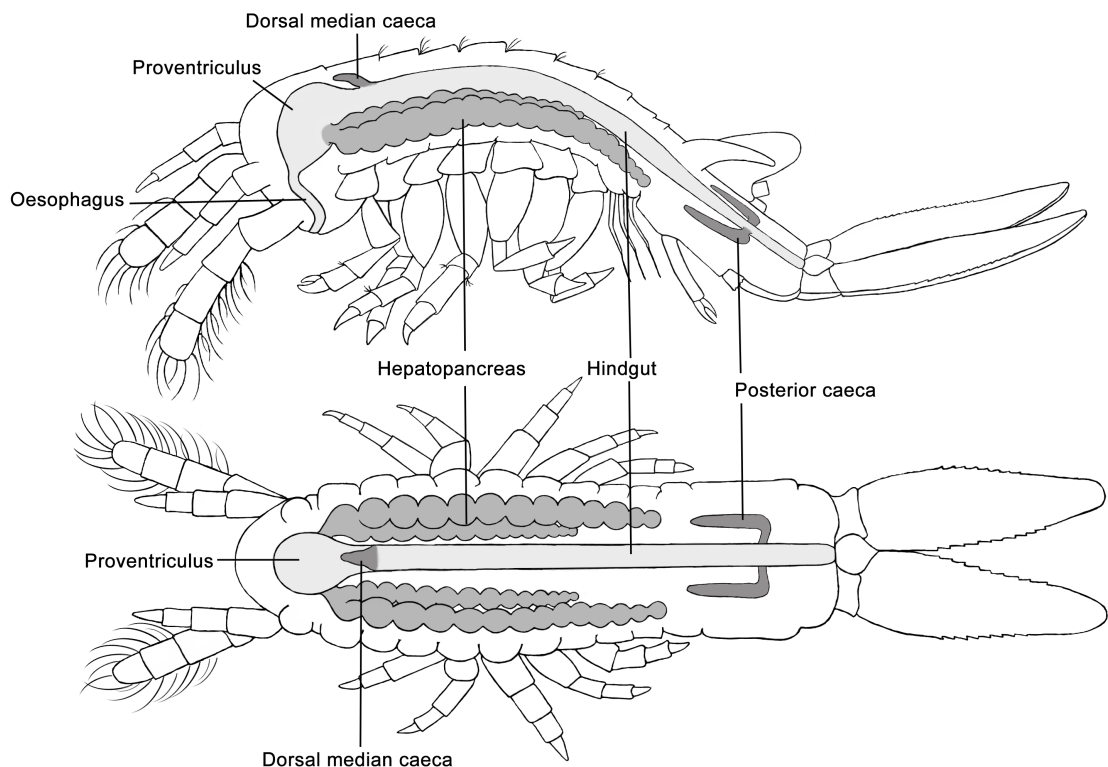
Figure 3.2 The general anatomy of the digestive tract of *C. terebrans*

Only a few studies (such as: Icely and Nott, 1984; Storch, 1987; Strus & Storch 2004; Kobusch, 1998) have used the greater resolution and 3D capacity offered by Scanning Electron Microscopy (SEM) to analyse the peracaridean proventriculus. This chapter aims to document features of the *C. terebrans* digestive system, which have so far not been imaged using SEM. As the site of food processing and the most complex part of the digestive system, the proventriculus will be described in detail in order to ascertain any modifications that have evolved to assist in wood processing.

3.2 Methods

3.2.1 Culturing of *Chelura terebrans*

Cultures of *C. terebrans* were maintained in running seawater at the Institute of Marine Sciences (IMS), University of Portsmouth, Hampshire, UK. The tanks have a natural seasonal regime with light and water temperatures reflecting those of the southern coast of the United Kingdom. However, the tanks have no tidal regime. Animals were fed a variety of wood substrates including *Chlorocardium rodiei* (greenheart) and *Pinus sylvestris* (Scots pine) in the form of blocks or planks (Figure 3.3.B).
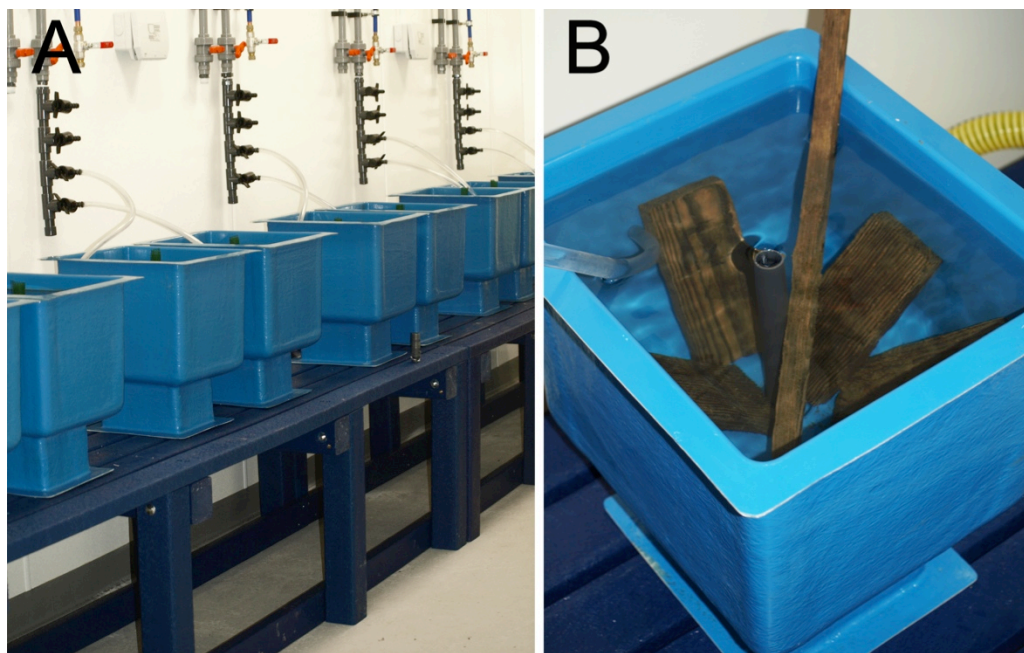
Figure 3.3 Culture tanks at IMS, Portsmouth. A) Tanks are supplied with a steady flow of water from Langstone Harbour. B) Wood in the shallow tanks used to feed woodborers includes scots pine and greenheart.

3.2.2 Collection and extraction techniques

Adult *Chelura terebrans* Philippi (between 5-6mm long) were collected from heavily infested wood (Figure 3.4) a paintbrush. If the animals are not found on the surface, and the wood is sufficiently degraded, layers of wood could be peeled back to expose the animals in their burrows. If animals are hiding inside wood which is still structurally sound or a large number of animals are needed, the wood was removed from the tank, wrapped loosely in paper towel and placed in a tray of seawater for 10min, the animals (both *Limnoria* and *Chelura*) were then collected as they emerged from the wood as a result of reduced oxygen levels within their tunnels.
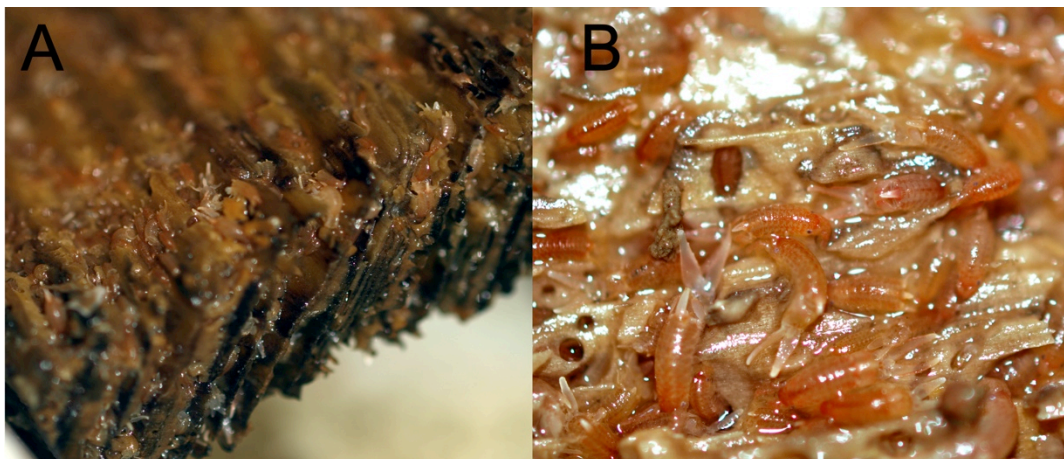
Figure 3.4 *C. terebrans* on the wood. A) They can be found in the grooves in the wood in great numbers. B) In new wood *C. terebrans* populate the surface of the wood creating troughs.

3.2.3 Sample preparation for SEM and LM

Several techniques were used to prepare the samples for light microscopy (LM) and scanning electron microscopy (SEM). The sequences of preparation are displayed in Figure 3.5 and explained further in the section.
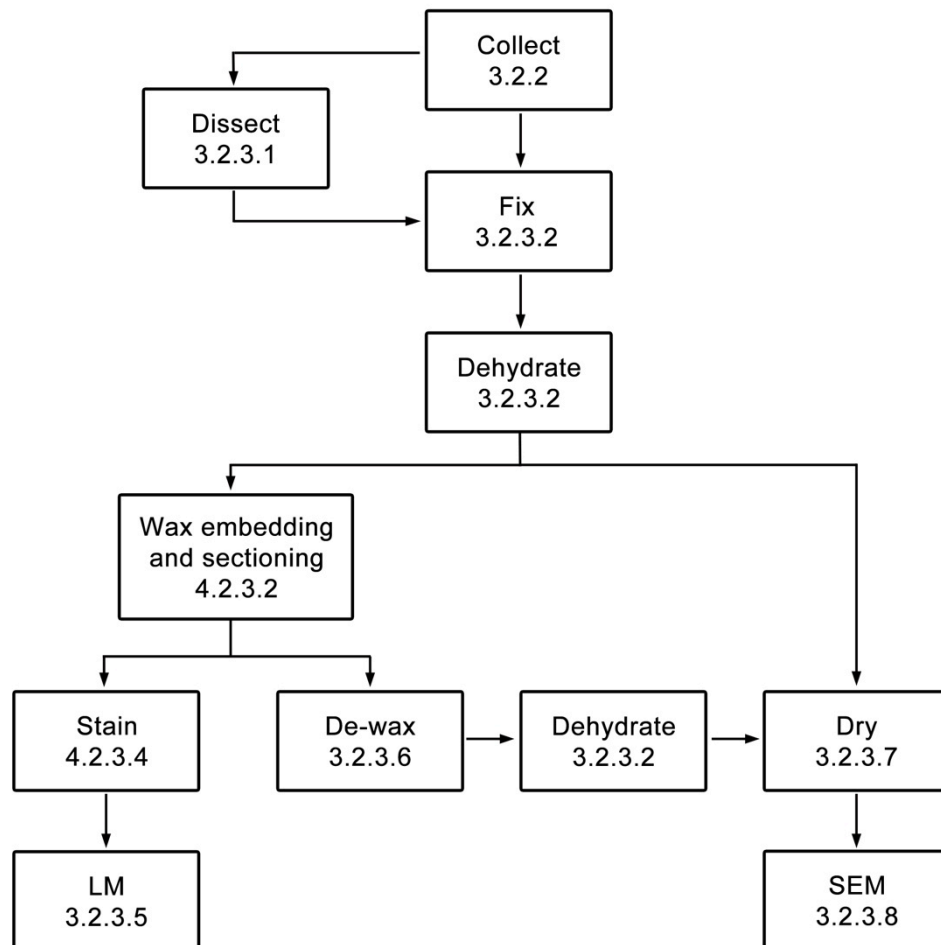
Figure 3.5 A schematic showing sequence of preparative stages for SEM and LM samples.

3.2.3.1 *Dissection*

Some animals were dissected before fixation to obtain only specific parts of the anatomy for inspection. Animals were anaesthetised by immersion in a clove oil

solution (0.4µl mL^{-1} of seawater) until motionless (Rehm, 2009). Once immobilised

using two pairs of forceps (one grasping just behind the head and the other on the

body) the head, hepatopancreas and hindgut were removed from the body. To obtain

images of the isolated proventriculus, it was detached from the head and squeezed

gently until all wood fragments were expelled from its lumen; the proventriculus was

then split in two by inserting the tip of the forceps inside and rubbing the other pair

against the edge.

3.2.3.2 *Fixation and dehydration*

Samples, either whole body or dissected animals, were fixed (3% v/v glutaraldehyde,

0.2 M sodium cacodylate, O.1 M NaCl, 2 mM CaCl$_2$ pH 7.4) for at least 90 minutes at

4°C. They were then rinsed in cacodylate buffer (0.2 M sodium cacodylate, 0.1M NaCl,

2 mM CaCl$_2$ at pH 7.5) for 30 minutes and then dehydrated via a graded series of

ethanol solutions (30, 50, 70, 80, 90 and 100%), all for 30 minutes, and a final 30

minutes in acetone.

3.2.3.3 *Wax embedding and sectioning*

The samples were placed in Histoclear® (National Diagnostics, UK) for 30 minutes and

changed 3 times to displace dehydrating agents. Samples were then transferred to 1:1

Histoclear®: Paraffin wax, and incubated at 60°C for 60 minutes. The Histoclear®: wax

solution was mostly drained off, without allowing samples to be exposed to air, and

replaced by paraffin wax for 30 minutes. This was repeated 3 times. After the third

change the animals were left in molten wax overnight. Samples were set into blocks in

various orientations and 7µm sections were cut using a microtome (Leica Jung Biocut

2035). The sections were placed on to slides in sequence which were then dipped into

warm water to smooth out the sections and temporarily fix the sections to the slide.

3.2.3.4 *Staining for light microscopy*

Sections were de-waxed in xylene for 10 minutes then placed in absolute ethanol for 2 minutes. They were then hydrated through a graded ethanol series to water (90%, 70%, 50% and dH$_2$O, 3 minutes each) and stained using haematoxylin for 10 minutes. Slides were placed in a trough with running water for 5 minutes to rinse out excess dye. Slides were then dipped in acid alcohol, for no longer than 5 seconds, then returned to running water for another 5 minutes. The sections were counter stained in eosin for 5 minutes and again returned to running water for 5 minutes. Water was shaken off gently and the slide placed in 70% ethanol for 3 seconds before being placed in 90% and 100% ethanol, for 1 and 2 minutes respectively. A further 100% ethanol step was performed to ensure complete dehydration of the slides, after which they were placed into xylene twice for 10 minutes. Excess solvent was removed and a cover slip placed over the sections using DPX mountant (Fisher Scientific, UK).

3.2.3.5 *Light Microscopy*

Sections were viewed using a bright-field compound microscope (Leica DM LB2) with a digital camera (JVC KY-F1030U). The images were saved as JPEG files and transferred to an editing program (section 3.2.4).

3.2.3.6 *De-waxing*

Once sectioned, the halves left in the wax blocks were placed in heated xylene for 30 minutes, to ensure complete removal of wax. This was followed by five changes of heated xylene, with the last left overnight at room temperature. The samples were then transferred to 100% ethanol and acetone for ½ hour each before drying (section 3.2.3.7).

3.2.3.7 *Drying*

Samples for SEM preparation were transferred to the low-surface tension solvent hexamethyldisilazane (HMDS, Sigma-Aldrich, UK) and dried by evaporation. After dehydration, samples were transferred to a 1:1 mixture of absolute ethanol:HMDS for 30 minutes then to 100% HMDS. The 100% HMDS was changed twice to ensure the removal of residual ethanol and placed in a fume hood to evaporate overnight.

3.2.3.8 *SEM*

Once dried, individual samples were mounted onto adhesive carbon tabs on SEM stubs. The samples were sputter coated with gold-palladium at 14mA for three minutes in an argon atmosphere (Polaron E6000). The samples where then viewed in a scanning electron microscope (JEOL JSM-6060LV) operated in high vacuum mode at an acceleration voltage of 10kv. Images were saved in TIFF format and then transferred to an editing program (see 3.2.4).

3.2.4 Images

Images were drawn using image manipulation software (Photoshop CS5) in conjunction with a drawing tablet (Wacom, Bamboo CLT-470). SEM images were corrected for contrast/brightness and cropped to form montages in the same program (Photoshop CS5).

3.3   Results

3.3.1 Terminology

3.3.1.1 *Planes of section*

Due to the method used to section *C. terebrans*, which is of small size and has a tendency to curl when placed in fixative, it is difficult to produce specimens with a uniform orientation and with conventional planes of section. To better understand the images shown in this chapter the terms used for sectional planes and anatomical positions are shown in Figure 3.6.
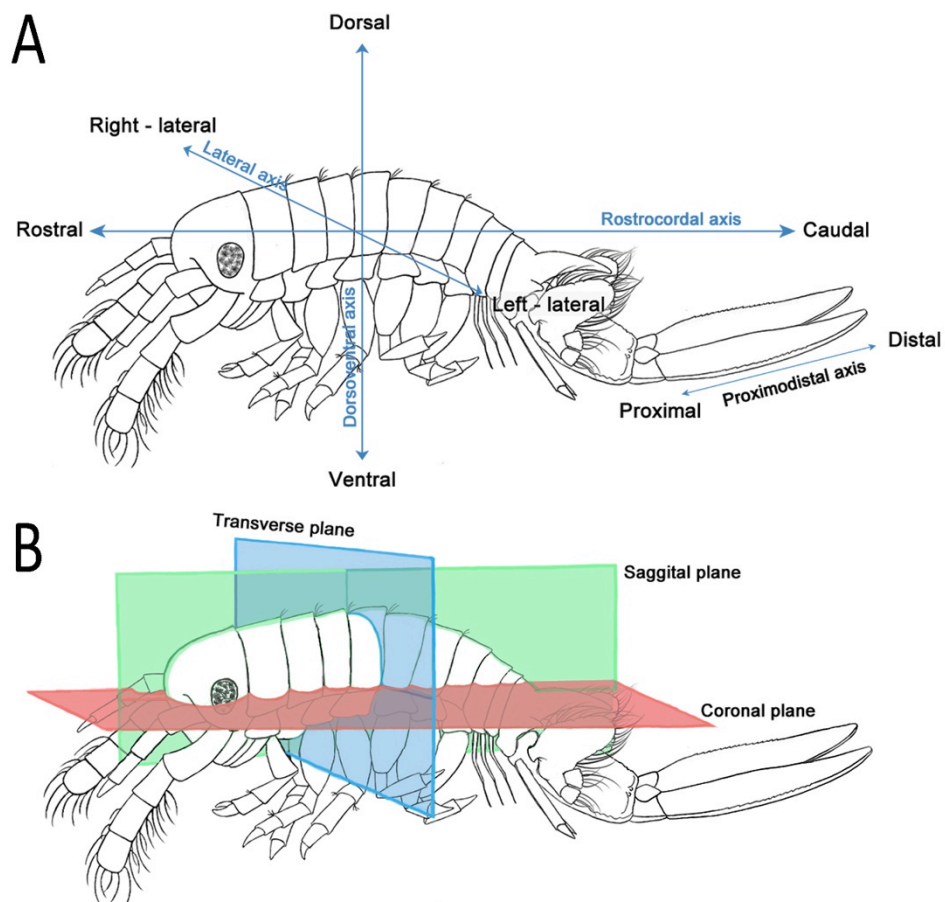
Figure 3.6 Anatomical terms of location and planes of section shown on C. terebrans. A) The anatomical terms of locations shown on *C. terebrans*. B) Sections through *C. terebrans*, the sagittal plane divides the left and right side of the animal, the transverse plane divides the rostral and caudal end of the animal and the coronal plane divides the dorsal and ventral.

3.3.1.2 *Anatomical position*

The term rostral and caudal, coming from the Latin rostrum, "beak" and caudum, "tail", is used instead of the synonymously used anterior and posterior (respectively). In this study, the term anterior will be used to describe the beginning (Latin, ante "before") of an organ. For example, the oesophagus runs dorsoventrally through the head, with the anterior end of the oesophagus being the entrance from the mouth, and the posterior, the join between it and the stomach (Figure 3.7).
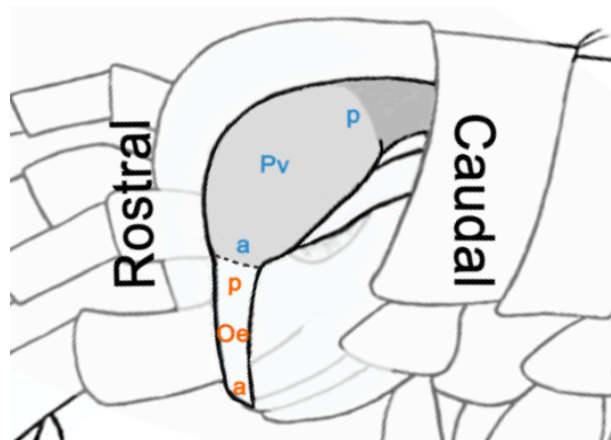
Figure 3.7 The anatomical position of the oesophagus (Oe) and the proventriculus (Pv) showing the anterior (a) and posterior (p) regions of the oesophagus and proventriculus.

3.3.1.3 *Anatomical terminology*

There are many synonyms for anatomical structures. These synonyms in conjunction with difficult to decipher light microscope images seen in early anatomical descriptions lead to difficulties when attempting to search and compare the literature detailing the anatomy of Amphipoda as well as the Pericard in general. The terminologies for anatomical features described and discussed for *C. terebrans* in this chapter are mainly derived from Strus and Storch (2004). These, along with some of the synonyms used by other authors, are displayed in **Table 3.1**.

Table 3.1 Anatomical terminology

Terminology used	Alternative Terminology
Anterior Inferomedianum (Richter & Scholtz, 2001)	Anteromedianum (Strus & Storch, 2004) Inferomedianum anterius (Coleman, 1991) Ventral cardiac ridge (Martin, 1964)
Posterior Inferomedianum (Davolos, 2010)	Inferomedianum (Strus & Storch, 2004) Inferomedianum posterius (Coleman, 1991) Ventral pyloric ridge (Martin, 1964)
Proventriculus (Hassall and Jennings, 1975)	Stomach (Schmitz & Scherrey, 1983)
Lateralia (Strus and Storch, 2004),	Pterocardiac ossicle (Schmitz & Scherrey, 1983) Gastric mill (Schmitz & Scherrey, 1983) Mandril/thorn press of the mouth (Kühne & Becker, 1964 & Thiem, 1942) Lateral ampullae (Hames & Hopkin, 1989) Lateral ridges (Agrawal, 1965) Spinose papilla (Keith, 1974)
Anterior inferolateralia (in the cardiac chamber, new herein)	Inferolateralia anterius (Coleman, 1991) Primary filter (Strus & Storch, 2004)
Posterior inferolateralia (in the pyloric chamber, new herein)	Inferolateralia posterius (Coleman, 1991) Ventral lateral fold (Martin, 1964) Inferolateralia (Strus & Storch, 2004)
Dorsal median caecum (Schmitz & Scherrey, 1983)	Anterior dorsal caecum (Martin, 1964)

Hepatopancreas (Schultz & Kennedy, 1976)	Excretory caeca (Shyamsundari & Rao, 1976) Diverticular (Schultz & Kennedy, 1976) Anterior caeca (Schmitz & Scherrey, 1983) Digestive gland (Keith, 1974)
Posterior caeca (Icely & Nott, 1984)	Rectal caeca (Shyamsundari & Rao, 1976) Caudal, rostral lateral directed diverticula (Kühne & Becker, 1964a)
Circulatory channel (Strus & Storch, 2004)	Dorsal cavity (Coleman, 1991)
Food channel (Strus & Storch, 2004)	Storage cavity (Coleman, 1991)
Primary filter (Strus & Storch, 2004)	Rough filter (Coleman, 1991)
Secondary filter (Strus & Storch, 2004)	Fine filter (Coleman, 1991) Bristle basket (Kühne & Becker, 1964)
Funnel (Strus & Storch, 2004)	Lamina dorsalis (Coleman, 1991)

The terminology regarding the setae is that used by Watling (1989). Where possible the descriptions will be given using terms presented in Figure 3.8.
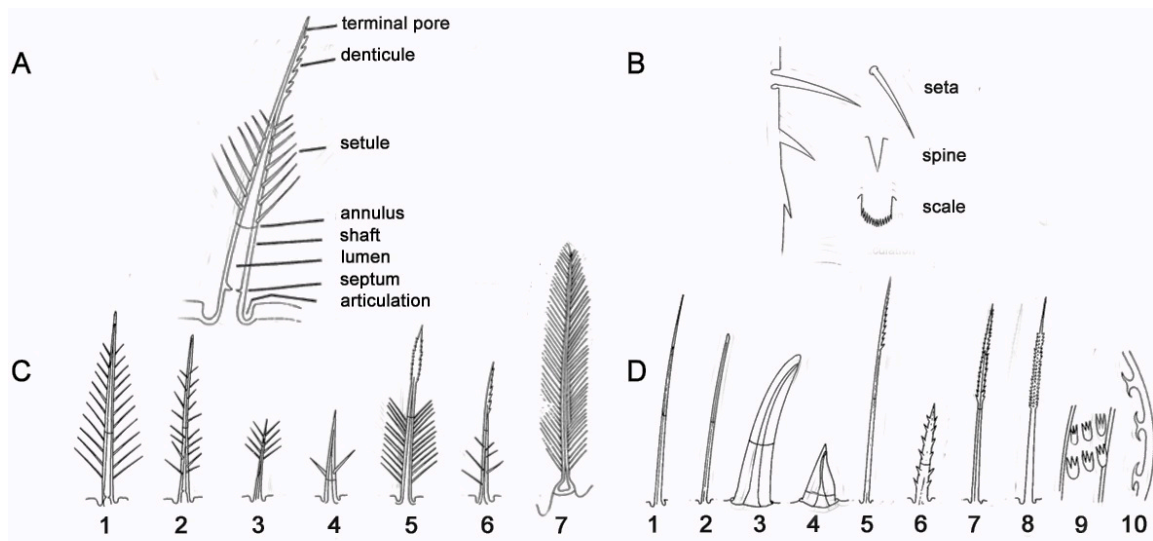
Figure 3.8 Terminology for types of setae found in crustaceans, modified from from Watling (1989). A) The basic morphology of a seta. Annulus, a faint ring circumscribing the shaft. Denticule, a non-articulated extension of the shaft of a seta or spine. Lumen, the hollow canal extending the length of the setal shaft interior. Septum, a basal constriction of the lumen of the setal shaft. Setule, an extension of the shaft of a seta, usually of uniform width from base to tip, and forming an articulated or flexible junction with the shaft. B) The three main types of crustacean setae. Seta, an articulated cuticular extension of virtually any shape or size; may vary from very small (10-20 μm) to very large (> 1 mm in length) and robust, often with a very wide base. Spine, a non-articulated cuticular extension that has a base that is generally not as wide as the structure is long; regardless of its size or shape, a spine has no socket. Scale, a non-articulated cuticular extension of which the base is generally very wide relative to its length; microscopic secondary features may arm the outer margin; this term should not be used to describe very small extensions of setal shafts. C) Setae Type 1, annulate with setules.1 and 3 – Plumose, 2 – pappose, 4 – forked, 5 and 6 – plumodenticulate, 7 – plumose, with supracuticular pocket. D) Setae Type 2, annulate without setules. 1 and 2 – simple, 3 – cuspidate, 4 – conate, 5 to 8 – various types of serrate, 9 – complex denticules of serrate seta, 10 – anvil-shaped denticules.

3.3.2 Abbreviations used for the digestive anatomy

The abbreviations used to describe the digestive anatomy of *C. terebrans* are detailed in Table 3.2.

Table 3.2 Abbreviations used for the digestive anatomy.

Abbreviation	Anatomy	Abbreviation	Anatomy
Ail	Anterior inferolateral	Hg	Gut
Ant	Antennae	Hp	Hepatopancreas
Bl	Bolus	Il	Posterior inferolateralia.
Cc	Circulatory channel	Lbr	Labrum
Cm	Circular muscles	lSt	Long setae
Cu	Cuticle	Lt	Lateralia
DMC	Dorsal median caeca	lu	Lumen
Dty	Dactylos	Md	Mandible
Ey	Eye	Mdp	Mandibular palp
Fc	Food channel	Ms	Muscle
Fic	Filter channel	Mx	Maxillae
Fm	Food mass	Mx 1	First maxillae
Fn	Funnel	Mx 2	Second maxillae
Gdp	Gnathopod		

3.3.3 Mouthparts

The oesophagus is situated rostro-ventrally on the head and the anterior opening is surrounded by complex mouthparts. The oesophagus is surrounded by the labrum (upper lip) on the rostral side and laterally by the mandibles which form a transverse biting mechanism. Two pairs of maxillae cover the mandibles, and these maxillae are covered by the first maxilipeds that are fused at their basal joints and act like a bottom lip, all of which are rostral to the level of the oesophageal opening (Figure 3.9).
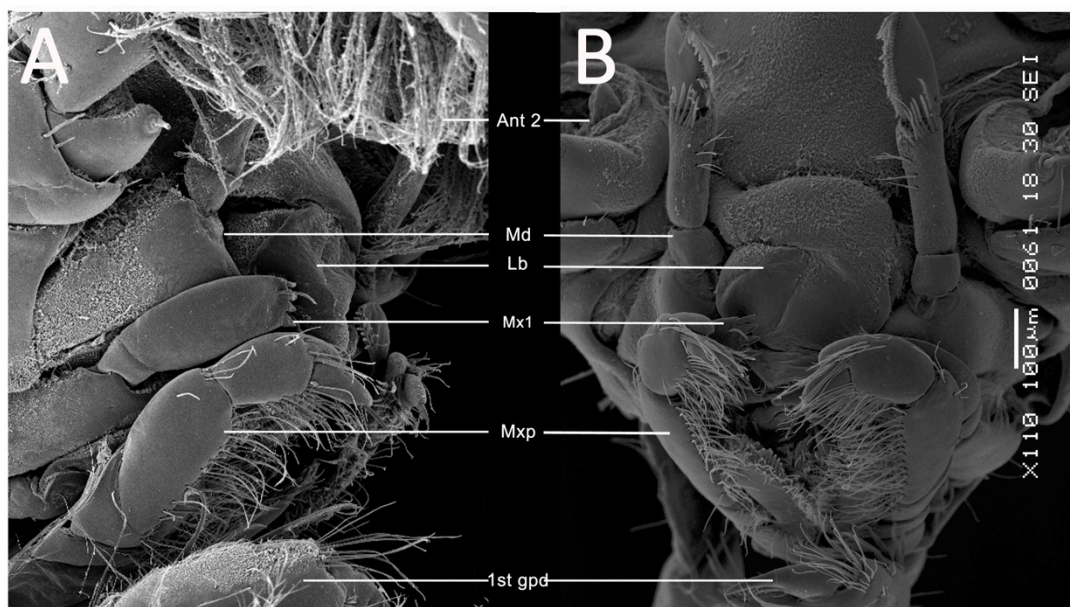
Figure 3.9 *In situ* mouthparts of C. terebrans. A) Right side lateral view of the head of *C. terebrans* showing mouthparts. Rostral end is on the left. B) Showing mouthparts from a frontal "head on" view. Ant 2 – antennae 2; Gdp - gnathopod; Lb – labrum; Md – mandible; Mx – maxillae; Mxp – maxillipeds.

The mandibles of *C. terebrans* are asymmetrical, as the right is shorter than the left. They possess a mandibular palp which is directed dorsally at rest, the final segment possesses plumose setae along the rostral margin which are approximately 50 μm in length and extend at right angles (Figure 3.10).

65
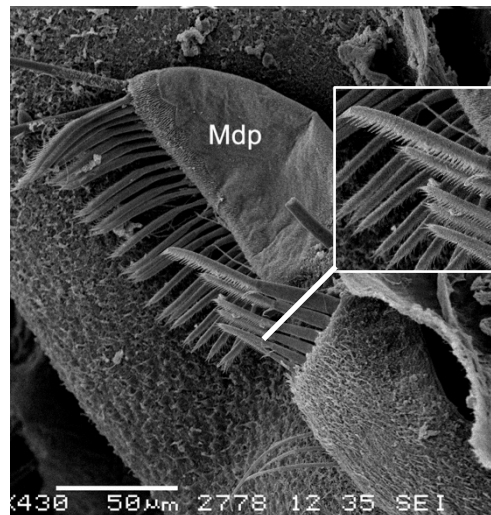
Figure 3.10 Detail of the last segment of the manibular palp (Mdp) with plumose setae; Insert - magnification of the plumose setae.

The first maxillae have an inner plate, outer plate and palp, all of which reach over the labrum (Figure 3.11). The first maxillae are also asymmetrical, the setae on the distal margin on the outer plate are their most distinguishing feature. These setae differ in number and shape (Figure 3.12), the right maxilla possesses 4-5 shorter stouter setae on its margin and the left 6-7 longer, narrower setae (Figure 3.12).
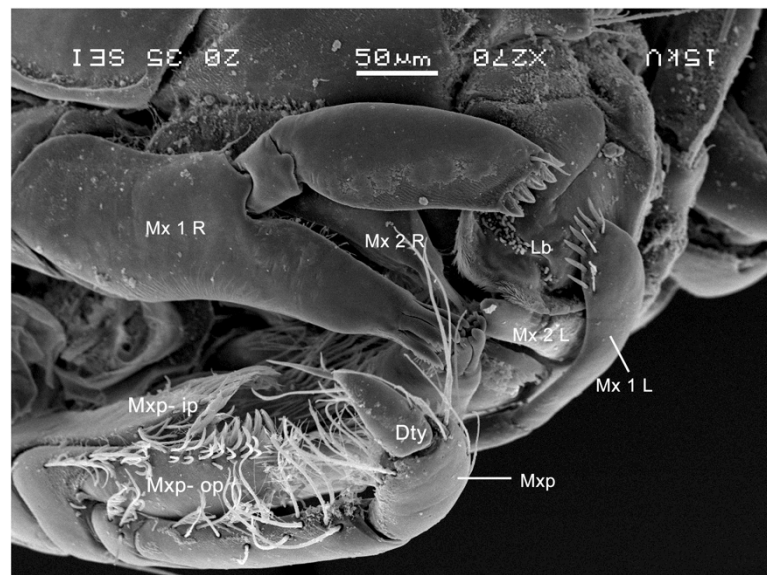
Figure 3.11 Mouthparts of *C. terebrans*. A) lateral view of the mouthparts right side of maxilliped is removed exposing the first maxilla. Lb- labrum; Mx – maxillae; Dty – dactylus; Mxp ip – maxilliped inner plate; Mxp op - maxilliped outer plate; Mx 1 – first maxillae; Mx 2 ; Second maxillae R- right; L – left.
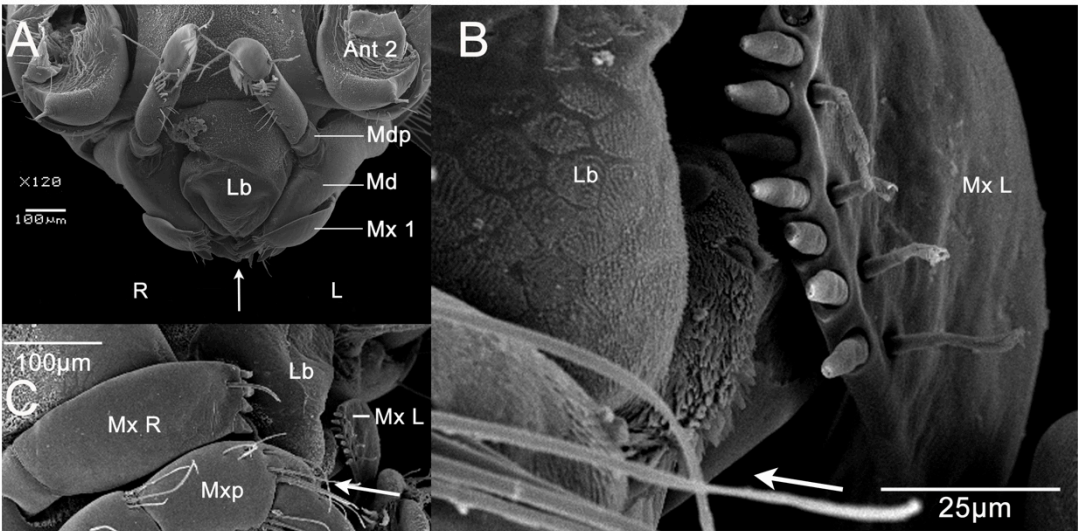
Figure 3.12 The maxillae of *C. terebrans*. A) The maxillae (Mx) and mandible (Md) exposed by removal of maxillipeds. B) The first maxilla on the right side (Mx R) with 4 setae on its margin. C) The first maxilla on the left side (Mx L) with 8 longer setae. The labrum (Lb) is visible with small setae on its surface, which become denser on its ventral side. All arrows indicate the direction of the oesophagus. Ant – 2 antenae; Mdp – madibular palp.

The many muscles needed for such complex mouthparts extend dorsally around the proventriculus and insert into the exoskeleton on the dorsal side of the head (Figure 3.13).
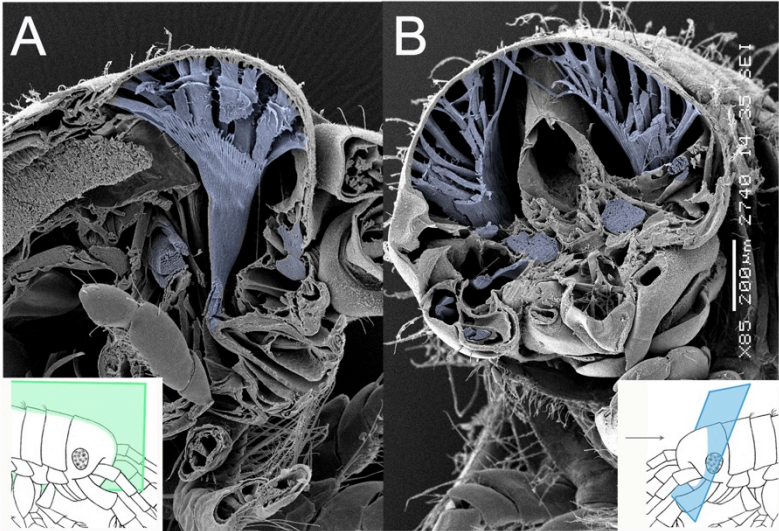
Figure 3.13 Muscles (blue) associated with the mouthparts of *C. terebrans*. A) The sagittal section, giving a lateral view of muscles from the mandible. B) Transverse section of head. Inserts indicate orientation of section and direction of view (arrow).

3.3.4 Overview of the digestive tract

The oesophagus is simple and leads to the more complex proventriculus. *C. terebrans* possess two pairs of hepatopancreas lobes (Figure 3.14), a feature which varies in number between amphipod species, and a simple gut with a dorsal median caecum and unpaired posterior caeca (Figure 3.14). The hepatopancreas appears posterior to the proventriculus and are connected on the ventral side of the proventriculus at its posterior end.
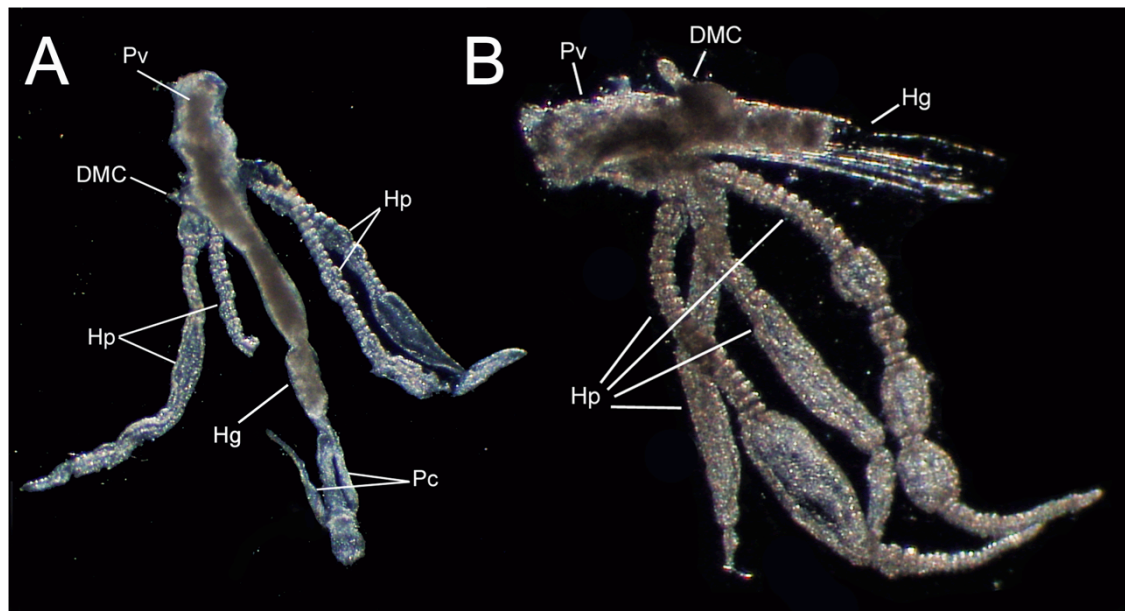
Figure 3.14 Digestive tract of *C. terebrans* visualised using the light microscope. A) Digestive tract viewed from the dorsal surface. B) Digestive tract visualised from the left lateral surface, the hindgut (Hg) on this specimen has broken away. DMC – dorsal median caeca; Hp – hepatopancreas; Pc – posterior caeca; Pv – proventriculus.

The proventriculus is the most structurally complex region of the digestive tract. The plates of its cuticular lining produce folds that protrude into the lumen creating distinct chambers and channels. The proventriculus is the site where food fragments are compressed into a food mass and through which the digestive fluids are circulated and

filtered. The posterior end of the proventriculus joins the gut. Here a single dorsal median caeca protrudes rostrally and rests on the dorsal side of the proventriculus (Figure 3.15). The gut extends the whole length of the body. At its posterior end, within the pleon region, two lateral, rostrally-directed posterior caeca extend laterally from the gut before reaching the anus located under the telson between the third uropods.
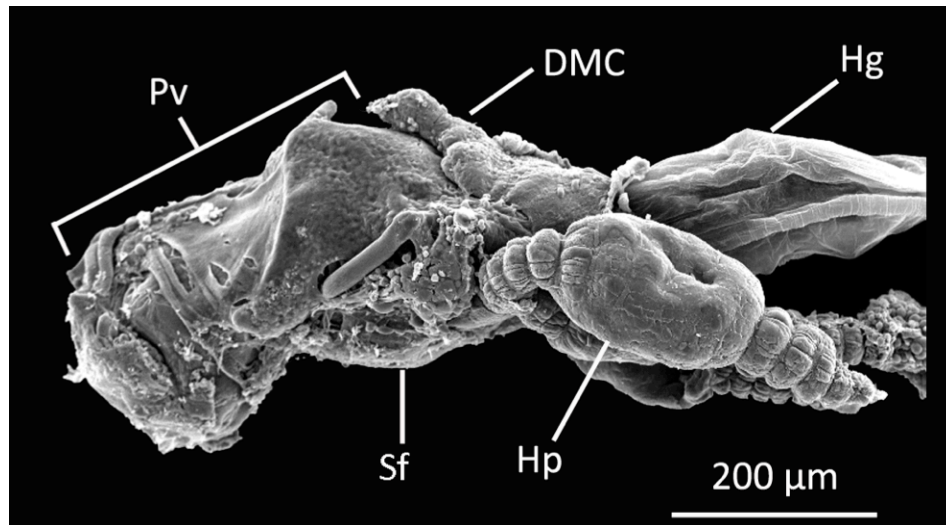
Figure 3.15 The left lateral view of the proventriculus (Pv) of _C. terebrans_ joining the hepatopancreas (Hp) and gut (Hg). The dorsal median caeca (DMC) can be seen resting over the dorsal side of the proventriculus at its rostral side. Sf – secondary filter.

3.3.5 Oesophagus

The oesophagus is dorso-ventrally orientated in an elongated S-shape. It is essentially a trapezoid shaped tube with the large base on its rostral side (see Figure 3.16C). It is approximately 300µm in length with a long pappose spike (articulation or pore not visible) projecting into the lumen from the rostral side at its anterior end (Figure 3.16B).
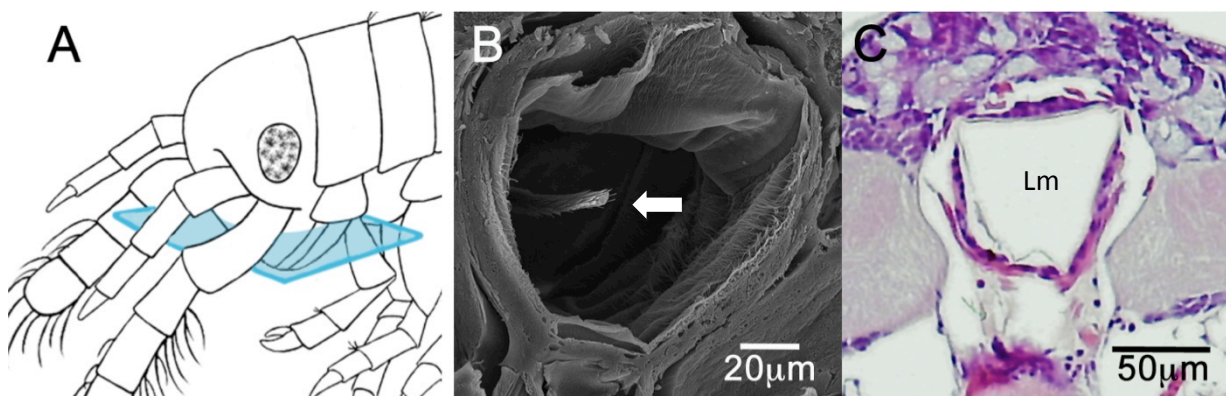
Figure 3.16 The oesophagus of *C. terbrans*. A) Shows the plane of section of the images. B) SEM image of oesophagus in cross-section, the arrow indicating the spike extending into the lumen (Lm) from the rostral wall at the top of the image. C) Light microscope image of the oesophagus stained with H&E, showing the lumen of the rounded triangular tube of the oesophagus (the rostral end of the animal at the top of the image).

The pappose spike extends into the lumen and bends posteriorly near its base. It possesses short, lenguiform, posteriorly-directed setules that appear to each have a supracuticular pocket (Figure 3.17B). A few setules on the caudal side of the spike appear to have similar features, however, they are very short and are only as long as they are broad (Figure 3.17A). The oesophagus was observed to have two distinct surfaces; the anterior of the caudal wall possesses intermittently spaced and anteriorly

directed short spikes (articulation not visible, Figure 3.17B) of approximately 1 - 2μm in length. The surface then changes after the first S-bend and along with the caudal and lateral walls becomes covered in long, simple, posteriorly-directed fine setae (Figure 3.17C), which are between 0.5 - 2μm in diameter at the base and more than 5μm in length, giving a fairly dense covering of fine setae along these walls. No exogeonous material was found within the oesophagus of any of the samples.
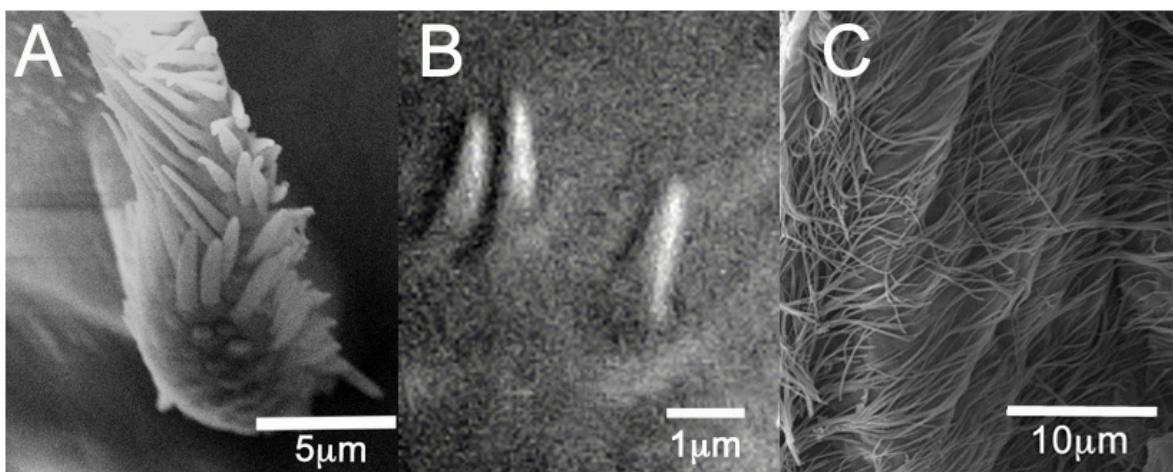
Figure 3.17 Oesophageal features of *C. terebrans*. A) Anteroposterior oesophageal spine projecting from the rostral wall with lenguiform setules. B) The anteriorly directed spines of the anterior rostral surface. C) The fine simple seta of the posterior surfaces.

3.3.6 Proventriculus overview

The proventriculus is located rostro-medially in the head region (Figure 3.18) suspended by extrinsic muscles attached to the exoskeleton. In an adult the proventriculus is approximately 500μm along on its rostrocaudal axis, extending slightly into the second pereon segment and approximately 200μm in length dorsoventrally. The proventriculus is widest at its rostral end, approximately 200μm across, and gradually tapering posteriorly into the funnel that leads into the gut.
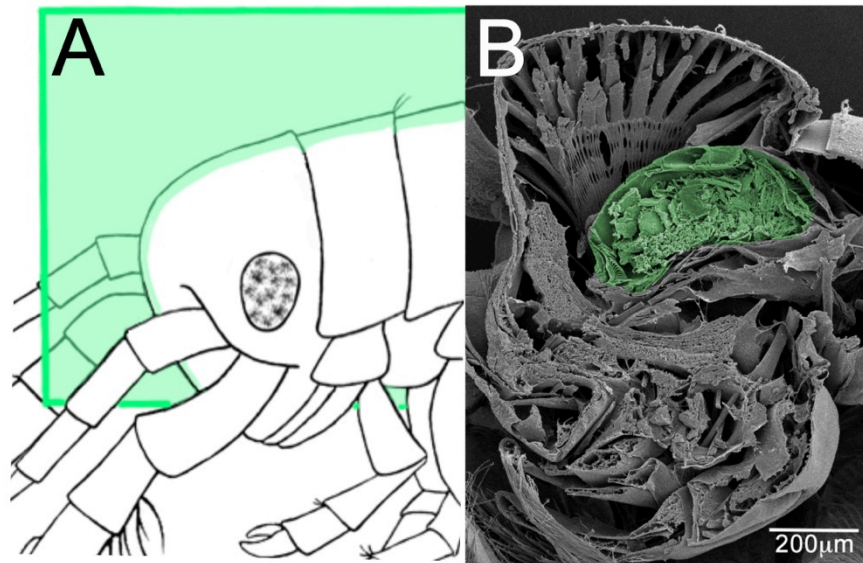
Figure 3.18 The position of the proventriculus in the head of *C. terebrans*. A) The coronal plane of section and orientation of the *C. terebrans* specimen. B) Corresponding SEM image of the section shows the position of the proventriculus (green) within the head.

The proventriculus is the most complex part of the digestive system the digestive tract with its intricate cuticular folds creating chambers and channels. The proventriculus is divided into a cardiac and pyloric chamber by two pairs of lateral cuticular plates (Figure 3.19). The cardiac chamber is at the anterior of the proventriculus and contains the lateral cuticular ossicles, or lateralia, that guard the oesophageal opening into the proventriculus (Figure 3.20). These ossicles are armed with a row of tooth-like structures followed rostrally by a row of long setae. The cardiac chamber also possesses the primary filter at its ventral region, which leads into the anterior ventral filter channel. The pyloric chamber is characterised by a deep dorso-ventral fold, the inferomedianum. This area is separated off from the food mass by large lateral folds, the inferolateralia, and the secondary filter (Figure 3.20).
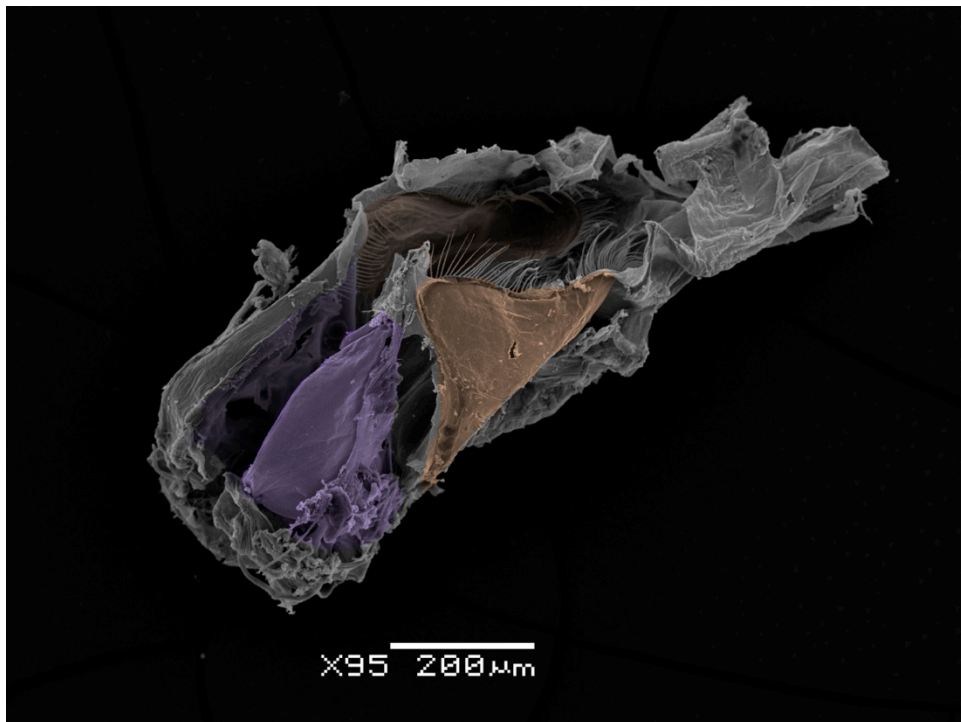
Figure 3.19 Lateral cuticular plates in the proventriculus of *C. terebrans*. The left lateral view of a dissected proventriculus to show the cuticular plates. Anterior lateral plates of the cardiac chamber (purple). The posterior lateral plates of the pyloric chamber (orange).
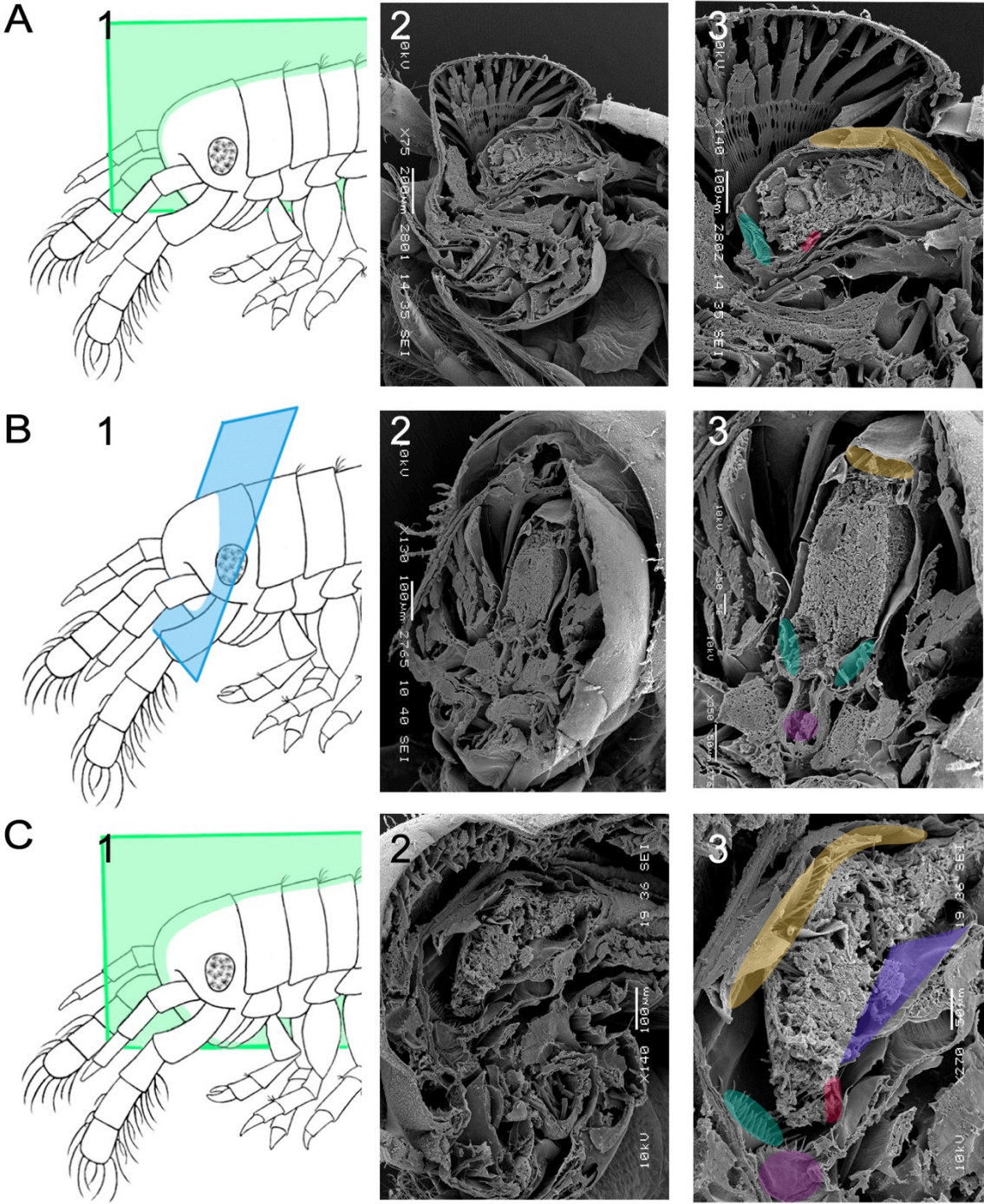
Figure 3.20 Features of the *C. terebrans* proventriculus. Column 1) The plane of section in their respective rows. Column 2) SEM image of the proventriculus at this section. Column 3) A closer view of the proventriculus with the features highlighted. Circulation channel (yellow), lateralia (green), oesophagus (pink), posterior inferolateralia (purple), primary filter (red). A) Sagittal section through the head at a slightly oblique angle. B) An oblique, transverse section through the head, cutting through the middle of the stomach on its dorsal side and cordially through the lateralia on the ventral side. C) An oblique sagittal section, slightly to the left of the medial line.

The lumen of the proventriculus is divided longitudinally into three incomplete channels that extend through both the cardiac and pyloric chambers, with the openings of the chambers guarded by setae. A shallow dorsal channel, the circulatory channel, is thought to aid the circulation of digestive fluids (Figure 3.21B). The main food channel at the centre is separated from a ventral filtering channel by the anterior and posterior inferolateralia (Figure 3.21).
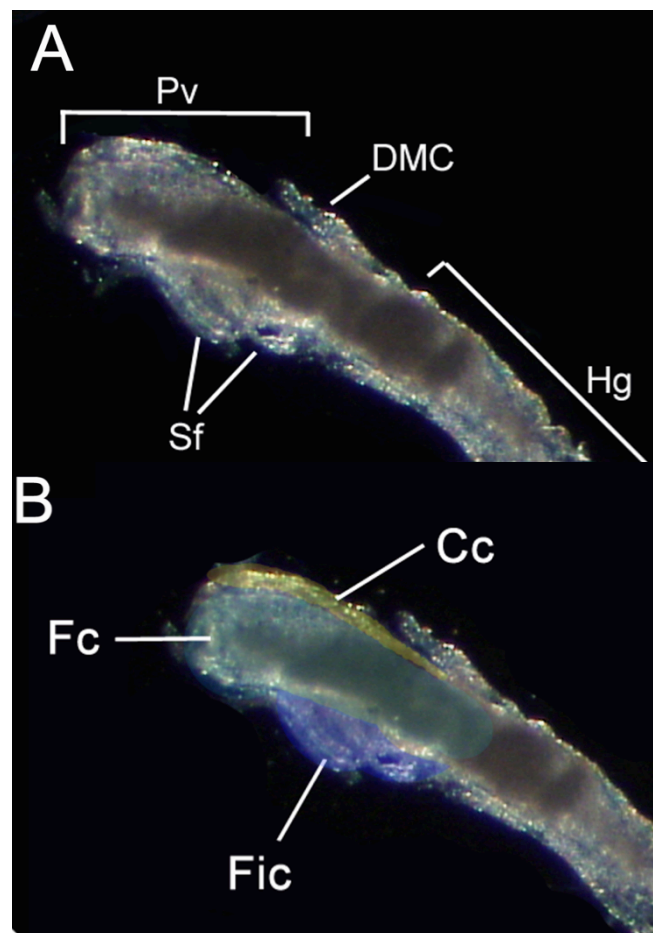
Figure 3.21 The proventriculus as part of the *C. terebrans* digestive tract. A) The proventriculus (Pv) is followed by the dorsal median caeca (DMC) and the gut (Hg). B) The proventriculus can be split into three main sections: the food channel (Fc, green) is separated from the filter channels (Fic, blue) by the inferolateralia, and the circulatory channel (Cc, yellow) by long setae. Sf - secondary filter.

The muscles surrounding the proventriculus act to contract and expand the proventriculus. The bands of intrinsic circular muscles under the filtering system create semi circles of muscles capable of pulling the ventral side of the proventriculus dorsally (Figure 3.22). Extrinsic muscles are also attached to the dorsal and ventral plates of the proventriculus and extend to the exoskeleton.
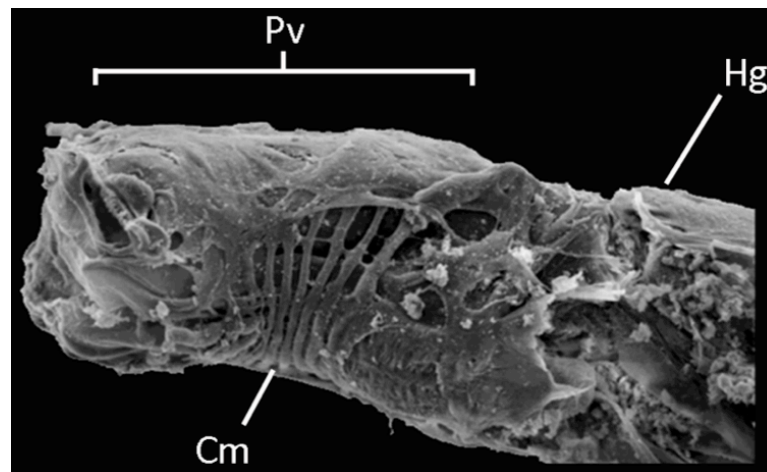
Figure 3.22 The intrinsic musculature surrounding the proventriculus of *C. terebrans*. A) Circular muscles (Cm) surrounding the proventriculus (Pv). Hg – gut.

3.3.7 The Cardiac Chamber

The cardiac chamber is an expanded region found at the rostral end of the proventriculus. The cardiac stomach is the site where the food enters the proventriculus from the oesophagus. Here the lateralia can be found guarding either side of the oesophageal opening (Figure 3.23D). The cardiac chamber is 140μm on its dorsal side and is characterised by the lateralia and two main lateral folds on the dorsal and ventral side of the chamber. The dorsolateral folds, the superolateralia, create the anterior portion of the circulation channel (Cc, Figure 3.23B). The folds are bordered by posteriorly directed setae that interlock with those of the opposite plate and section off a shallow chamber. The anterior lateral folds are similarly bordered and create the

primary filter channels (Ail, Figure 3.23B), with the food channel in the centre portion of the cardiac chamber forming the largest channel. The cardiac chamber is delimited by the edge of the cuticular palate, which is fringed with posteriorly directed setae, preventing the reflux of food once it is passed from the cardiac chamber.
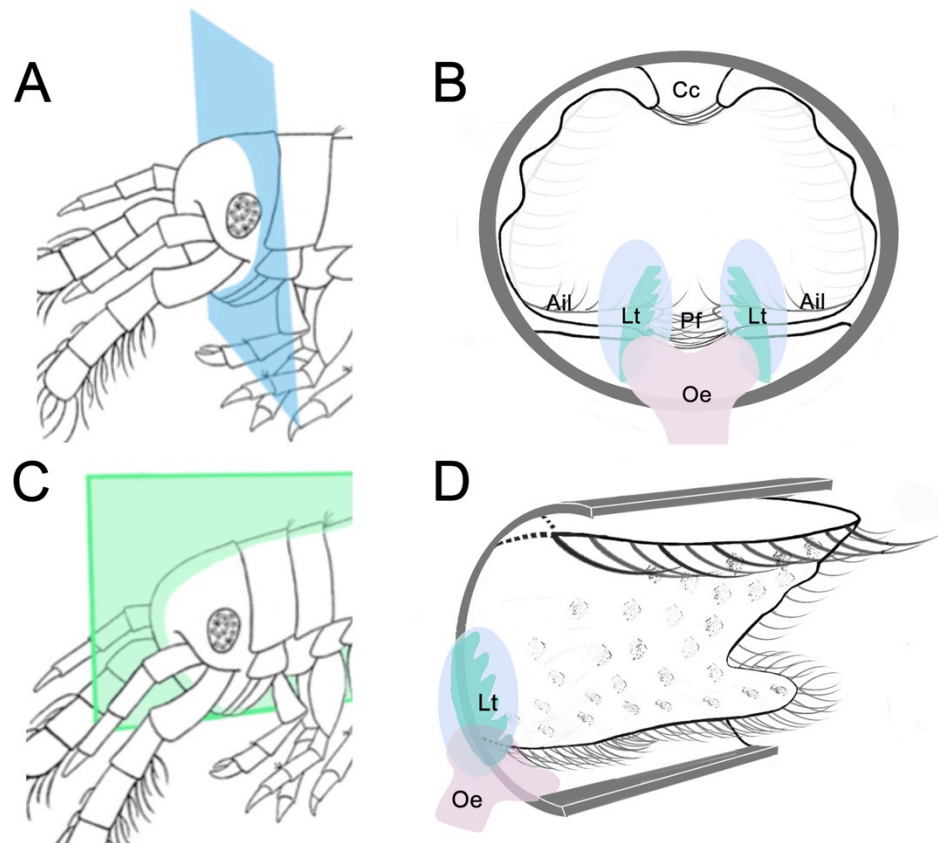
Figure 3.23 Schematic of the *C. terebrans* cardiac stomach based on SEM images. A) Orientation of section in B. B) The cardiac stomach from the rostral wall into the lumen of the cardiac chamber. The lateralia (green) at the anterior of the chamber guarding the entrance to the oesophagus (pink). The anterior inferolateralia (Ail) protrude into the lumen on the ventral side to create the primary filter (Pf). The dorsal lateral folds create the circulatory channel (Cc). C) Orientation of section in D. D) Oblique sagittal angled section showing one of the lateral plates with the lateralia (Lt) and the oesophageal opening (Oe). Setae surround the fringe of the plate; those on the anterior lateral fold are shorter and form the primary filter. The longer setae on the dorsal fold form the circulatory channel.

The setae that create the circulatory channel are approximately 100μm long and simple. They are broader at the base, oval in cross section and possess lumen (Figure 3.24A & B). The surface of the cardiac chamber is predominantly smooth with small posteriorly-directed setae. These setae are acutely-angled close to the surface and are grouped in rows that are themselves, in most areas, grouped into scale shapes (Figure 3.24C & D).
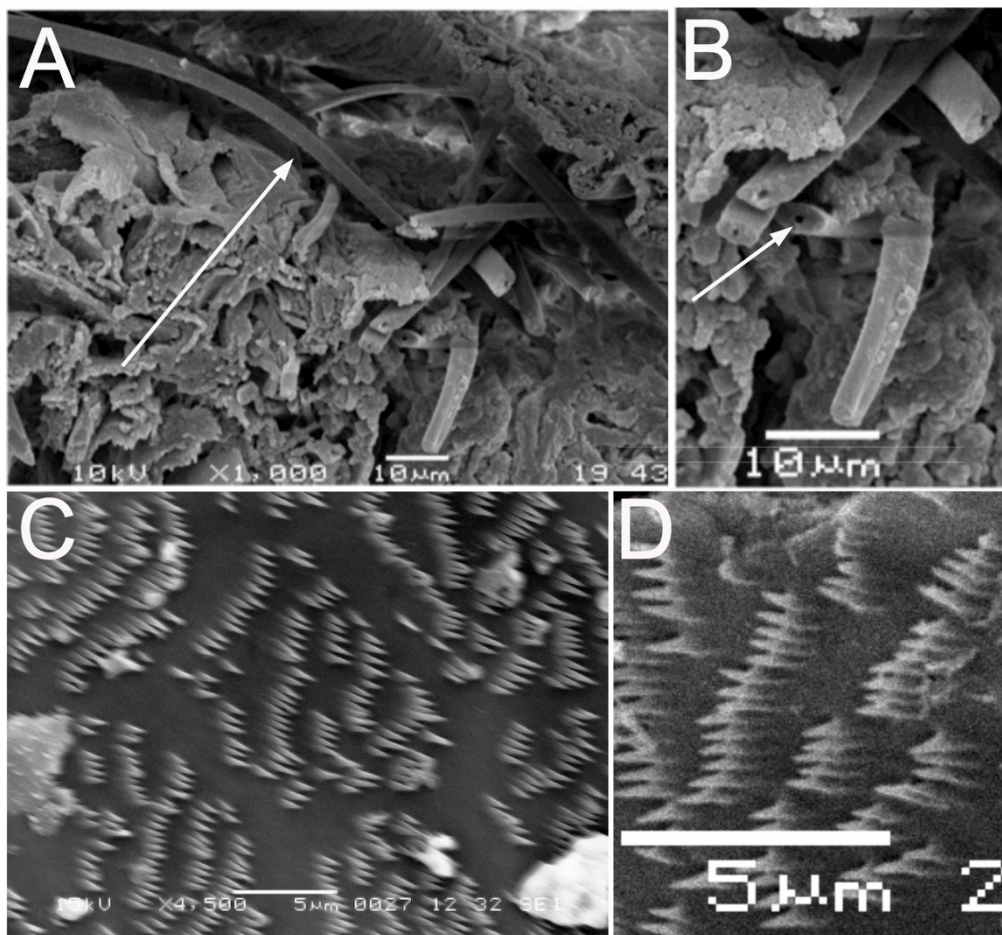
Figure 3.24 Setal arrangement in the cardiac stomach of *C. terebrans*. A) Long simple posteriorly directed setae (arrow) from the dorsal side of the cuticular plate. B) Magnified dorsal setae showing oval cross section and lumen (arrow). C) The setal arrangement on the ventral surface of the plates, clusters of setae are arranged in scale shaped groups. D) Magnification of the setae making up the setal scale clusters.

3.3.7.1.1   Lateralia

The junction between the oesophagus and the proventriculus is guarded by two lateral ossicles, the lateralia. The lateralia are approximately 130μm dorsoventrally in length and 40μm wide in adults and are found at the anterior of the stomach (Figure 3.25B). They extend further dorsally and ventrally than the oesophageal opening, (Figure 3.25C). The lateralia consist of a row of tricuspid spines, approximately 15μm in length, and caudally, a row of long serrate setae, in which both rows are directed posteriorly. The tricuspid spines are approximately 30μm long and are shaped like a chelicerate claw. The longest spike is on the rostral side of the spine and is hooked posteriorly. The caudal spike is slightly smaller than the rostral spike and is hooked more sharply in the anterior direction, with the central spike being much shorter and straight. The spines have many small sharp spinules, approximately 0.5μm in length, on their caudal facing edge (Figure 3.25D). The laterialia are similar in on either side, however, it is unclear whether they are interlacing when in movement. The SEM examination suggests that this could be the case, the preparative techniques which allow visualisaton of the proventriculus may result in the misrepresentation of their natural positioning. Extrinsic muscles extend from the proventriculus in the region of the lateralia and insert into the exoskelleton on the lateral sides of the head.
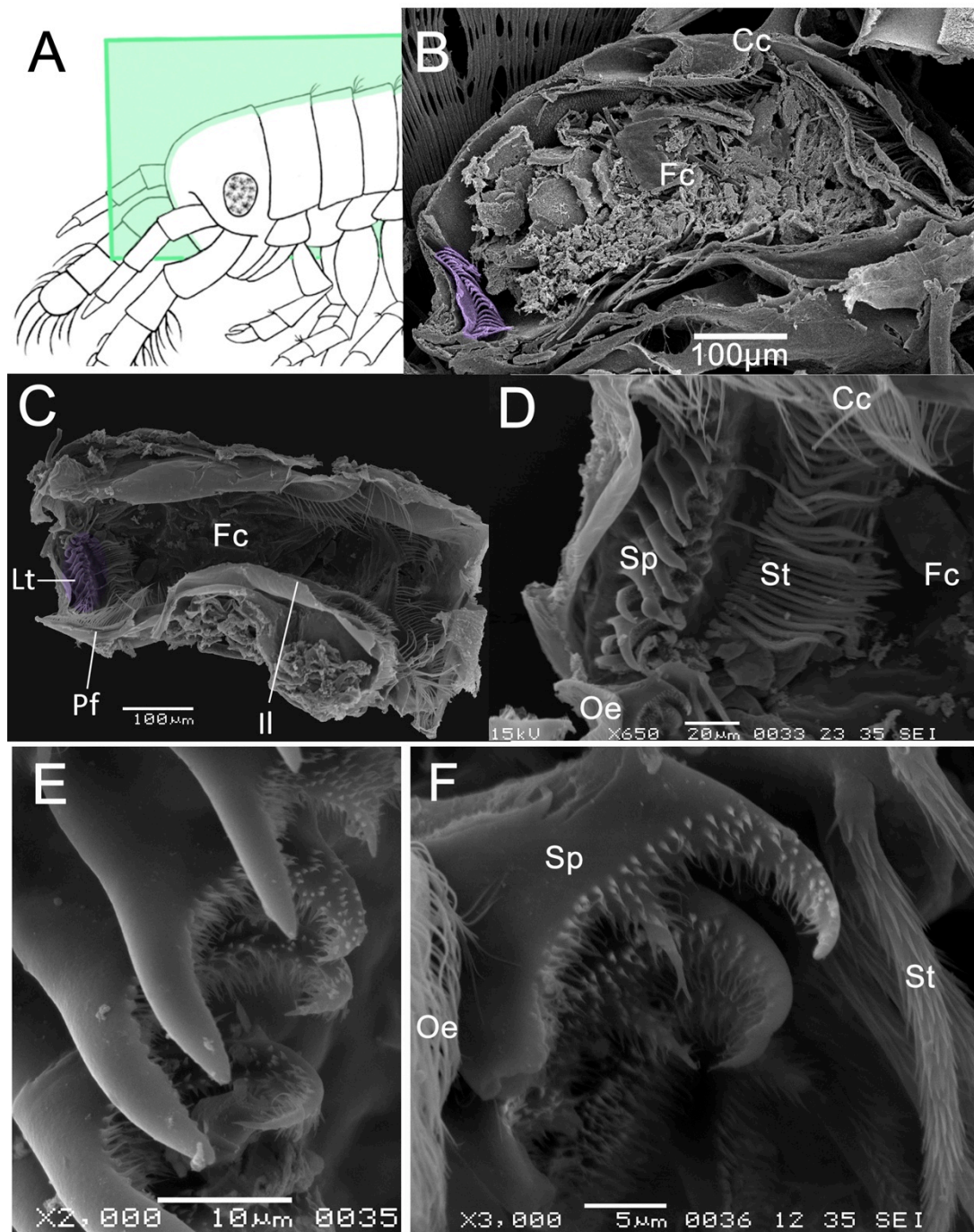
Figure 3.25 Lateralia of *C. terebrans.* A) All images in this figure are taken from animals sectioned in the coronal plane. B) Position of the lateralia (purple) in the proventriculus. C & D) Lateralia (Lt) positioned next to the opening from the oesophagus (Oe) showing the cuspid spines (Sp) and setae (St) E) The posteriorly directed cuspid spines with sharp spinules on their caudal surface F) Cuspid spine (Sp) with rostral spike broken, and posteriorly directed setae (St) of the lateralia. Cc – circulatory channel; Fc – food channel; Pf – primary filter. Il - posterior inferolateralia.

3.3.7.2 *Primary filters*

The primary filter is constructed from setae protruding from the anterior inferolateralia into the lumen. The plumose setae from either side interlock along the median line of the stomach creating a mesh. The setae are approximately 2 μm apart with setules approximately 1μm apart (Figure 3.26). The anterior of the chamber has a median medial ridge which does not protrude above the anterior inferomedianum, creating two posteriorly directed channels.
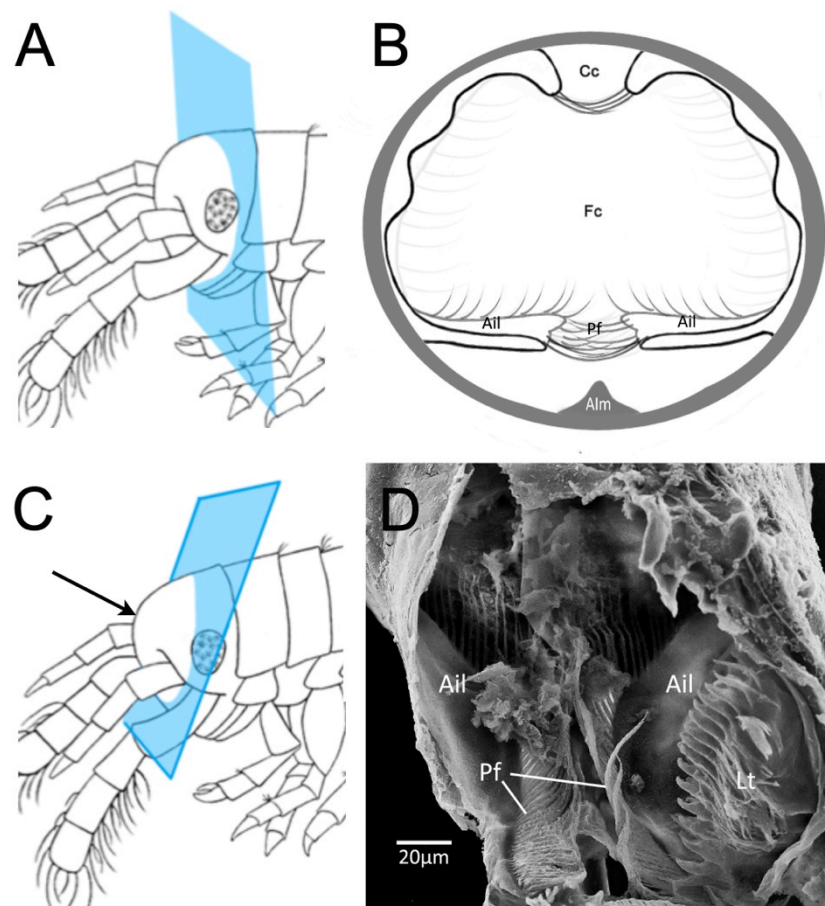
Figure 3.26 Primary filter of *C. terebrans*. A) The orientation of the section in B. B) A section through the cardiac stomach just posterior to the lateralia. This shows the setae from the anterior inferolateralia (Ail) closing off the anterior filter channel from the food channel (Fc). The medial ridge (Alm) separates the filter channel in two. C) Orientation of section in D. D) The anterior inferolateralia (Ail) possessing marginal setae, creating the primary filter (Pf). Lt – lateralia.

3.3.8 Pyloric Chamber

The pyloric chamber of the proventriculus is found at the posterior of the proventriculus and, at around 400μm, it is larger than the cardiac chamber. The posterior cuticular plates of the pyloric chamber also possess two large outfoldings. The dorsolateral folds and the ventral lateral folds of the pyloric region are a functional continuation of those found in the cardiac chamber, and result in the circulation channel and the filter channel. As in the cardiac region, both folds are fringed with setae. In the pyloric chamber, the plates narrow and taper downward at their posterior ends and form a funnel leading into the gut (Figure 3.27). The tapering of the plates means the filter channel spirals around the proventriculus from the dorsal side to the ventral. The proventriculus is at its largest in the pyloric chamber, as the filter channel extends in a ventral direction. A large ventromedial ridge, the Inferomedianum, divides the dorsal surface of the filter chamber, the secondary filter is found on the lateral walls of the Inferomedianum. The filter chamber is separated from the main food chamber by large lateral out foldings, the posterior Inferolateralia.
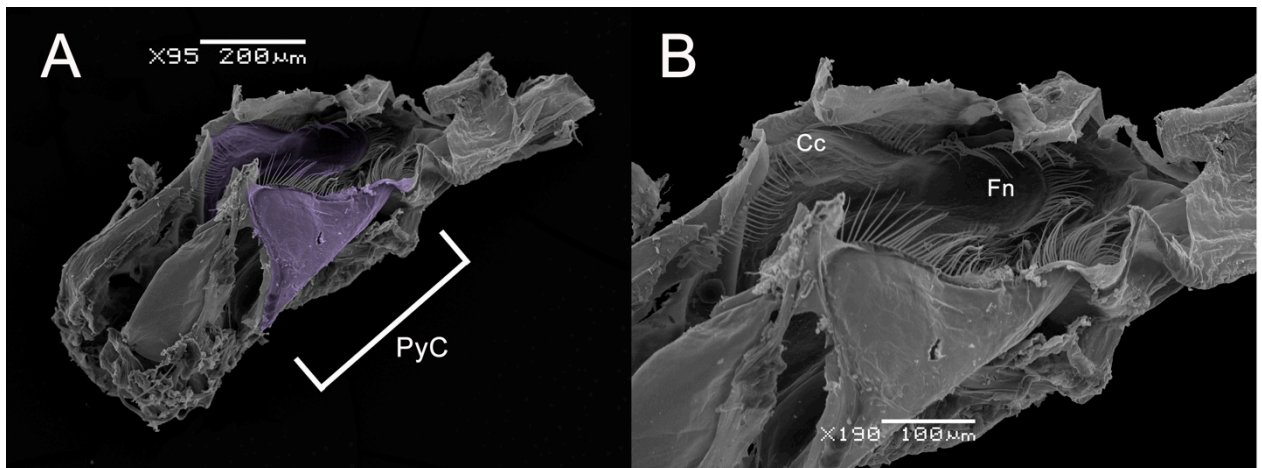
Figure 3.27 The pyloric chamber of *C. terebrans* from the left lateral side. A) The proventriculus showing the location of the pyloric chamber (PyC) with cuticular plates (purple) B) Magnification of the pyloric chamber showing the circulatory chamber (Cc) and the tapered posterior region of the plate creating a funnel (Fn).

82

3.3.8.1 *Posterior inferolateralia*

The inferolateralia are lateral folds that extend into the lumen and span the entire length of the pyloric stomach. They are positioned above the posterior inferomedianum and meet in the middle (sometimes overlapping) separating the food channel from the posterior filter chamber (Figure 3.28). The inferolateralia are deflected dorsally as they approach the median line of the proventriculus, creating gutters along the lateral walls and a ridge along the median line (indicated by arrow Figure 3.28 A2 & A3).
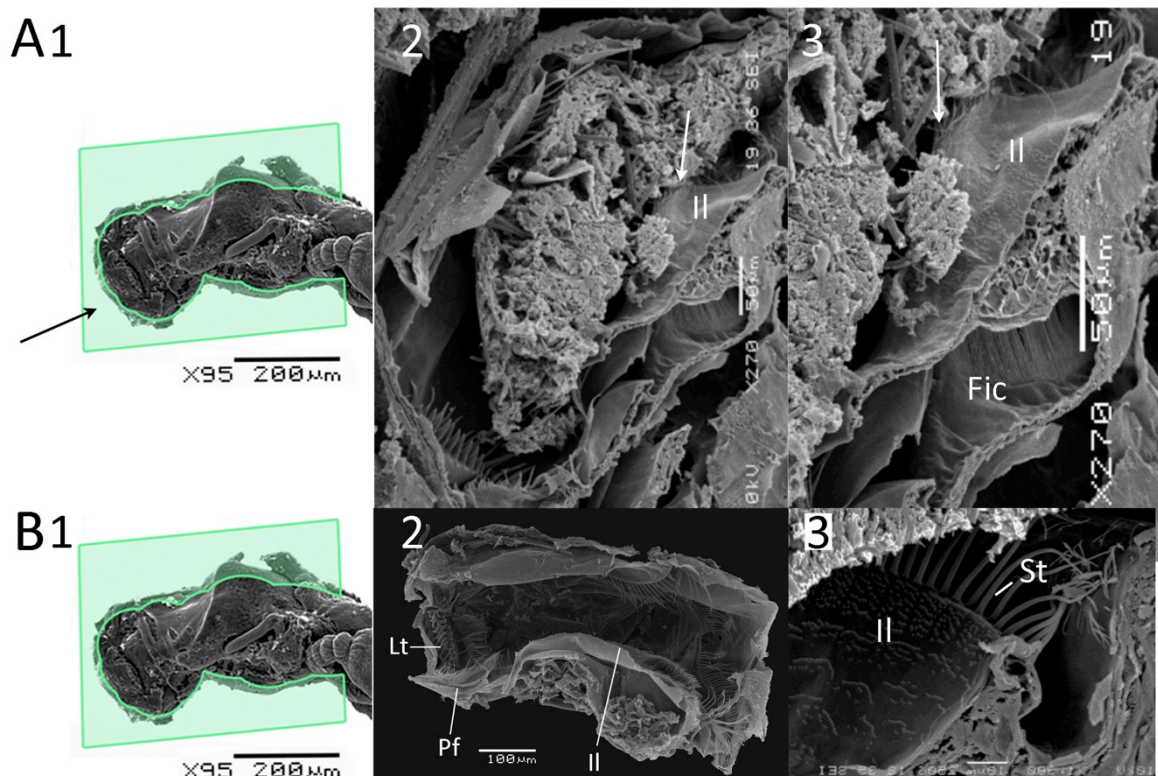
Figure 3.28: Inferolateralia of *C. terebrans*. A1) Position of sagittal section seen in images A2 & A3, arrow indicates direction of view. A2) Proventriculus with wood particles in the food channel, arrow indicates the ridge caused by the meeting of the two posterior inferolateralia (Il) shown slightly to the right of the arrow. A3) A magnified view of the inferolateralia (Il) showing the surface of the lateral fold preventing the penetration of food particles into the lower filter channel (Fic). B1) Position of sagittal section seen in images B2 & B3. B2) lateral view showing the length of the posterior inferlateralia in the cardiac stomach. B3) The fine posteriorly directed setae on the surface of the inferolateralia (Il)

and the longer setae (St) along the posterior margin. Pf – primay filter; Lt – lateralia.

The medial margin of each fold, where the two inferolateralia meet, are thicker and possess a fringe of dense short (approximately 7µm) setae. The caudal margin is fringed with fewer long setae, approximately 60µm in length, which traverse the channel toward the inferomedianum and secondary filter (Figure 3.28B3 & Figure 3.29). The dorsal facing surface of the inferolateralia is similar to that of the cardiac chamber, possessing small posteriorly directed rows of setae grouped into scale shapes (Figure 3.29).
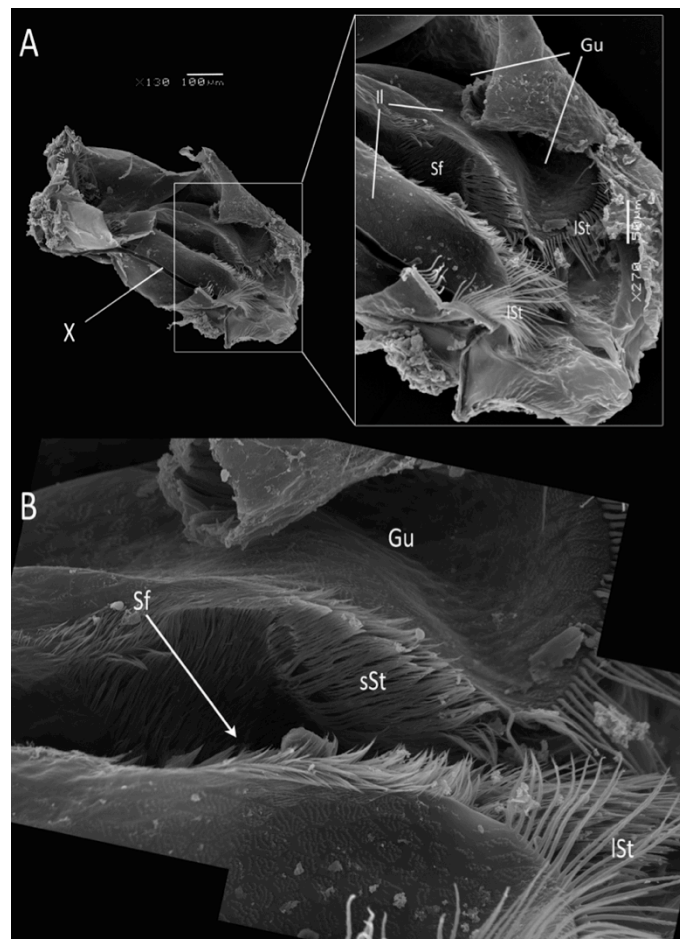
Figure 3.29 Posterior Inferolateralia of *C. terebrans*. A) A dissected proventriculus, the rostral side is to the left. The posterior inferolateralia have been separated slightly and left inferolateralia was broken along its length during preparation (X). Magnification on the right shows inferolateralia (Il) on either side of the proventriculus and the secondary filter (Sf) between the two folds. B) The ridge of the posterior inferolateralia showing thickened medial

edge with short setal margin and ventromedial surface (sSt). Gutter (Gu) created by the median rise of the inferolateralia. Long setae (lSt) at the posterior of the inferolateralia.

3.3.9 Secondary filters

The secondary filters are positioned below the posterior inferolateralia (Figure 3.30). The filter channel is divided medially by the posterior inferomedianum, here two tiers of plumose setae are inserted in rows on its lateral sides, forming the secondary filter (Figure 3.31).
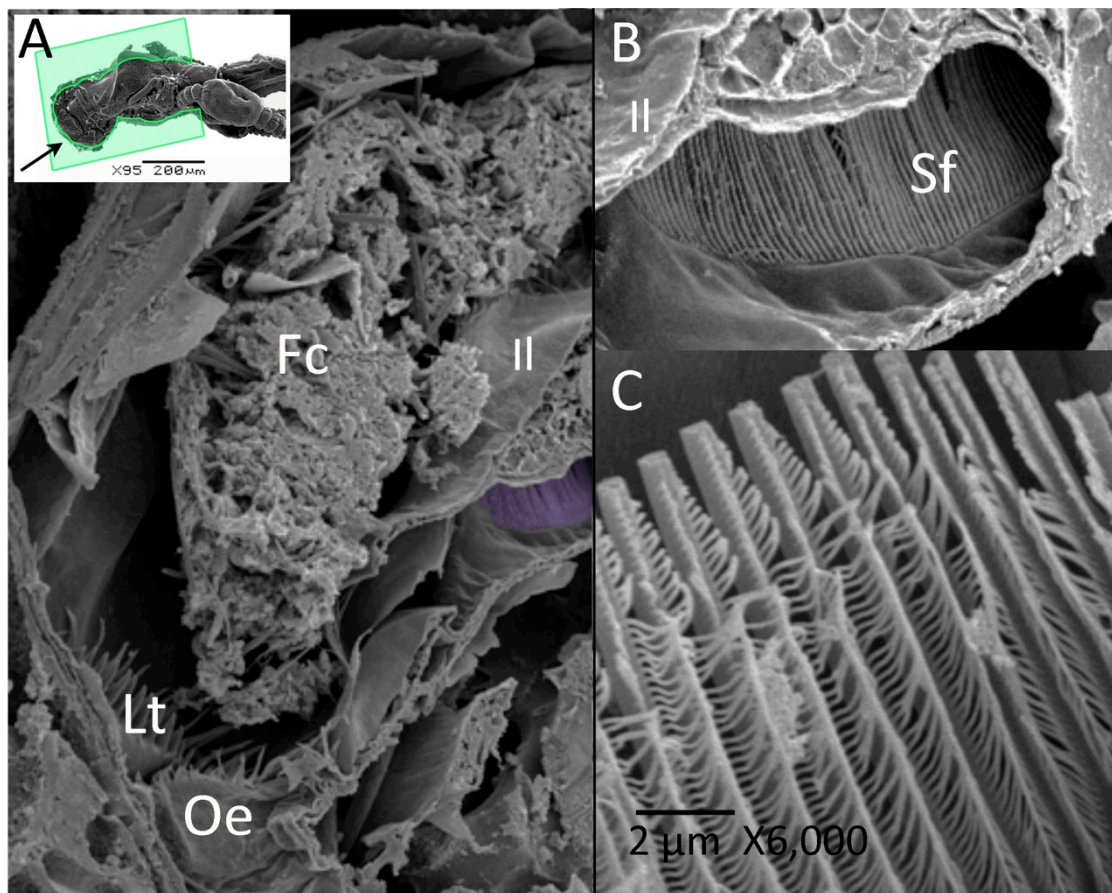
Figure 3.30 Location of the *C. terebrans* secondary filter. A) View through proventriculus from the rostral end at the oesophagus (Oe) and the lateralia (Lt) towards the posterior end. The posterior inferolateralia (Il) separates the wood fragments in the food channel (Fc) from the secondary filter (purple). B) The secondary filter (Sf) viewed from the left lateral wall of the filter channel. C) The concave plumose setae of the secondary filter with joined setules.

The lower row of setae overlaps the bases of the upper row. The bases of each row protrude from the inferomedianum to create two posteriorly directed channels (Figure 3.31). The setae in these rows are uniformly distributed and close enough for the setules to link, approximately 1μm distance from each other (Figure 3.31). The setae are concave in cross section, with the setules on their fringe directed toward the inferomedianum. The setules are also uniformly spaced (0.1μm apart) and are approximately 1μm long and 0.1μm wide at their base.
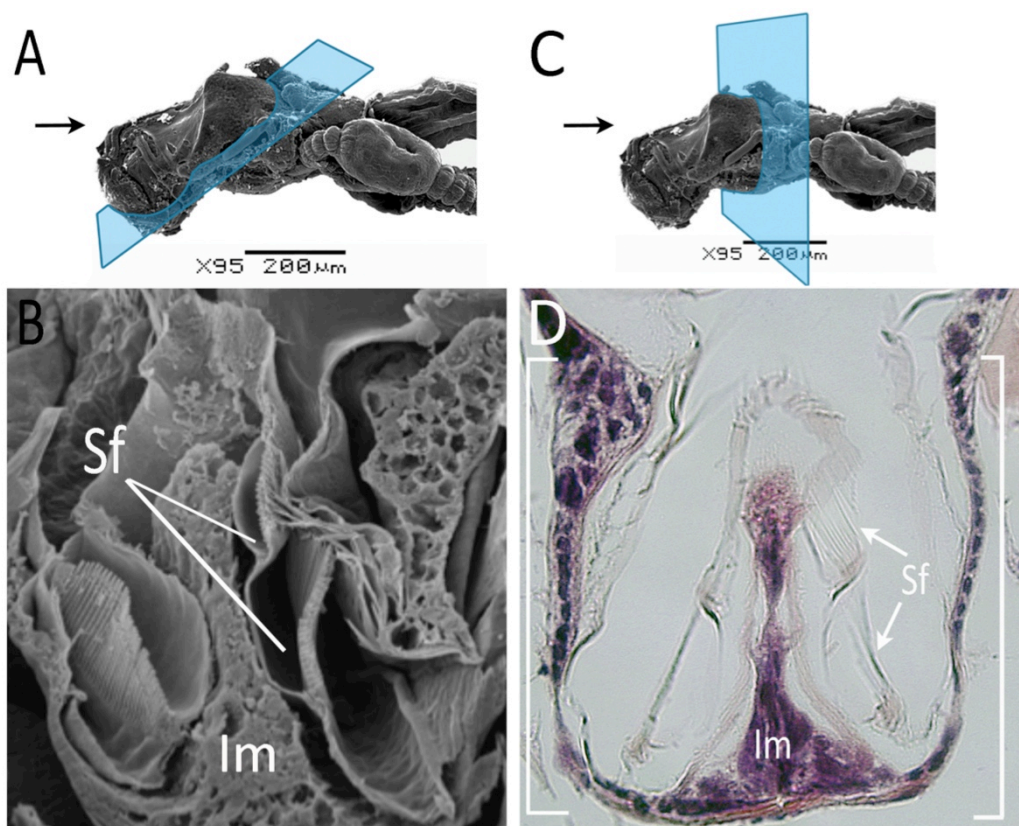
Figure 3.31 The secondary filter of *C. terebrans* in reference to the inferomedianum. A) Indicating the angle of transverse section in image B, direction of view shown by arrow. B) The filter channel showing the posterior inferomedianum (Im) and two tiers of setae forming part of the secondary filter (Sf) and posterior dorsolateral channels created by the setal rows. C) Location of transverse section of image in D, direction of view shown by arrow D) A light microscopy image of filter channel (boxed) with the inferomedianum (Im) dividing the channel in two, with two rows of the secondary filter setae (Sf).

The posterior inferomedianum is shorter and narrower anteriorly but then broadens and rises dorsally at the level of the secondary filter setae. Here, it rises and enters the food channel to the level of the inferolateralia (Figure 3.32), creating a tongue like protrusion that is covered by posteriorly directed setae, which appear iridescent under a light microscope.
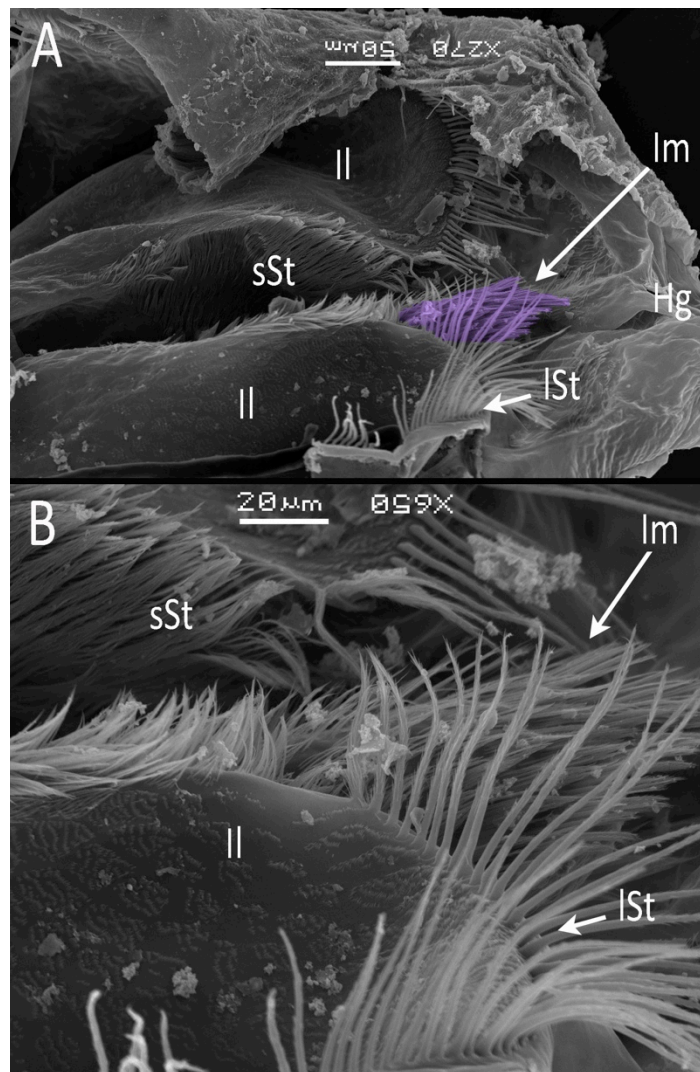
Figure 3.32 Inferomedianum of *C. terebrans*. A) The posterior of the pyloric chamber viewed from the left side. The inferomedianum (Im, purple) protruding into the food chamber between, and only to the level of, the inferolateralia (Il). B) Magnification of the the inferomedianum between the long posterior plumose setae (lSt) of the inferolateralia (Il). Hg – Gut; sSt – short setae.

At the posterior margin of the posterior inferolateralia, the posterior wall of the proventriculus forms a channel underneath the posterior lip of the inferolateralia

(Figure 3.33 & Figure 3.34). The posterior long setae of the two inferolateralia traverse the channel. The inferomedianum fills the gap between the two at the median line. The channel is gutter-shaped and leads to the inferomedianum.
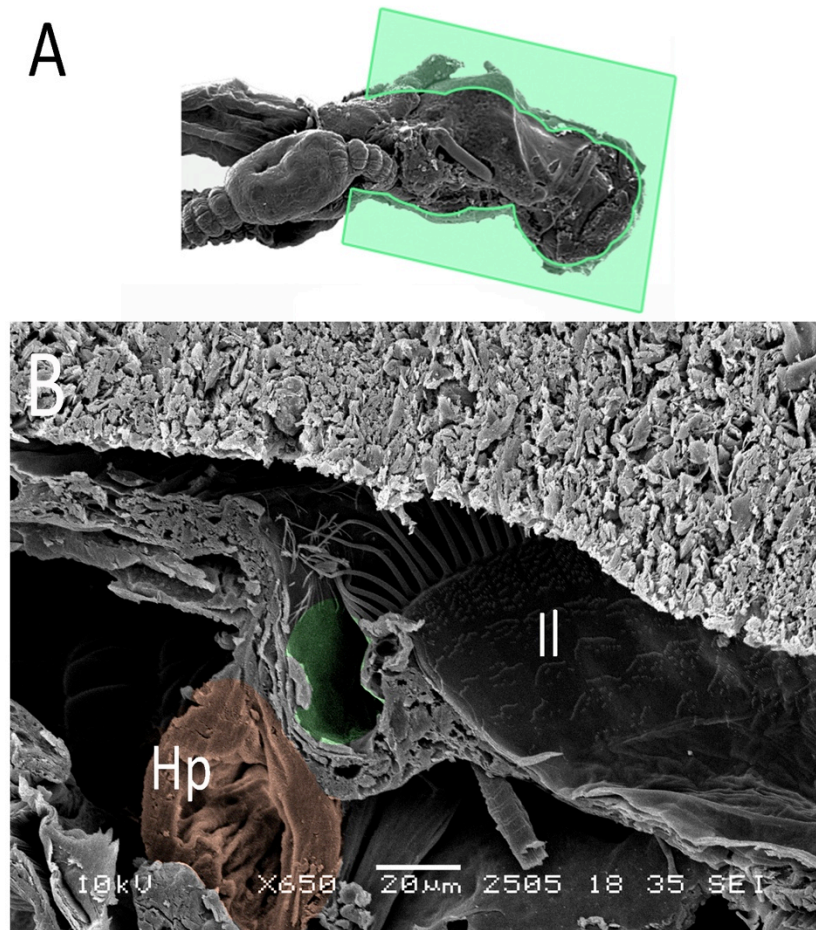
Figure 3.33 The channels at the posterior end of the proventriculus leading to the inferomedianum and the secondary filter in *C. terebrans.* A) Orientation of section through the proventriculus in B. B) View from the right hand lateral side of the posterior wall of the proventriculus, forming a channel (green) towards the inferomedianum. The posterior setae of the inferolateralia (II) traversing the channel toward the inferomedianum (green) and, below this, the channel that leads from the secondary filter to the hepatopancreas (Hp, orange).

The channels from the secondary filter continue posteriorly until they are beneath the posterior lip of the posterior inferolateralia. Here, they converge into a chamber at the posterior of the proventriculus, the hepatopancreal chamber (Figure 3.34 & Figure 3.35).
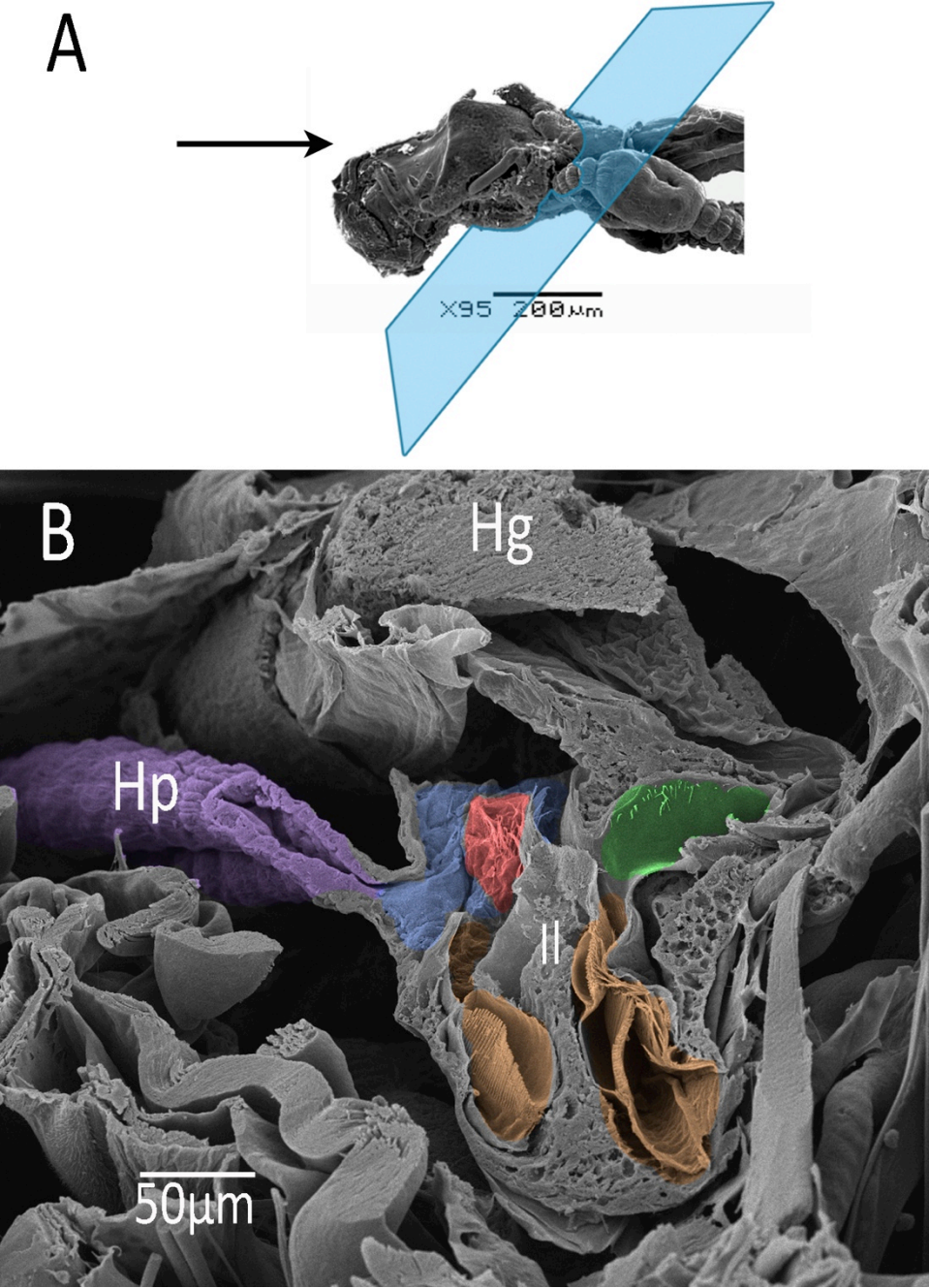
Figure 3.34 The secondary filter channels in *C. terebrans* leading to the hepatopancreal chamber. A) Orientation of specimen section shown in B, arrow indicates direction of view. B) The secondary filter and the filter channel (orange) leading to the hepatopancreal chamber (blue), which in turn leads to the hepatopancreas (Hp, purple). The post inferomedianum at the centre (red) and the channel at the posterior of the pyloric stomach (green) leading to the inferomedianum (Im).
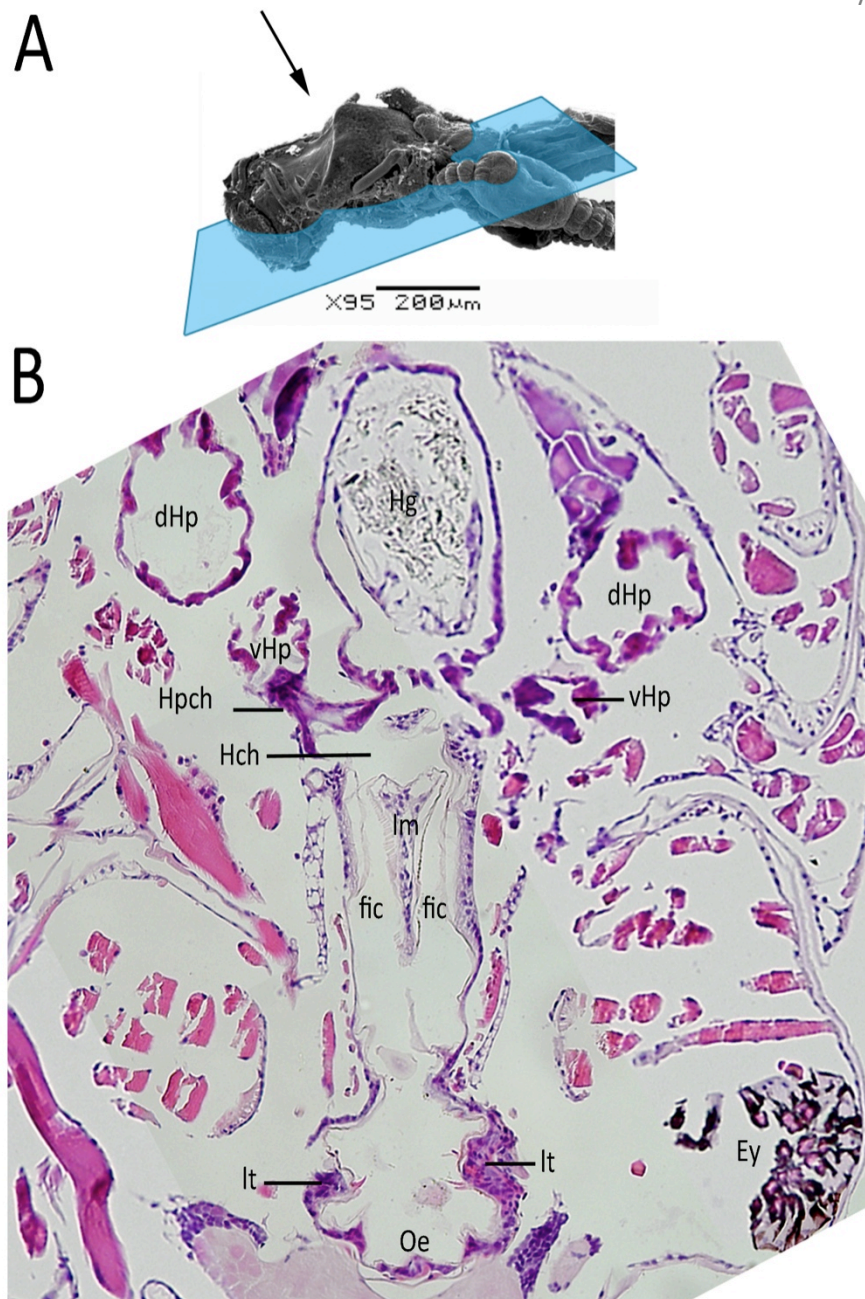
Figure 3.35 Section from the ventral portion of the proventriculus in *C. terebrans*. A) Orientation of histological section in B. B) Oblique section of the ventral side of the proventriculus (Note: the left lateral side is cut lower than the right making the eye (Ey) visible on the left side). The rostral side is at the bottom and the caudal side at the top of the image. The opening of the oesophagus (Oe) entering the cardiac region with its lateralia (lt). Posterior to this, the secondary filter channels (sFc) on either side of the inferomedianum (Im) opens into the hepatopancreal chamber (Hch). The hepatopancreal chamber leads into the hpatopancreal channels (Hpch) and then to the dorsal and ventral hepatopancreas (dHp & vHp respectively). Food can be seen in the gut (Hg) posterior to the proventriculus.

90

3.3.10   Hepatopancreas

Chelura terebrans has two pairs of blind ending sacs, hepatopancreatic lobes. The two lobes are joined at their proximal end where they connect to the hepatopancreatic chamber on the ventral side of the proventriculus (Figure 3.36). The paired lobes are orientated longitudinally in the body running parallel to the hindgut. In transverse section the lobes are arranged on top of one another and surround the ventral side of the hindgut (Figure 3.37). The dorsal lobes are longest and extend into the pleon, whereas, the ventral lobes are approximately two-thirds that length (Figure 3.36B).
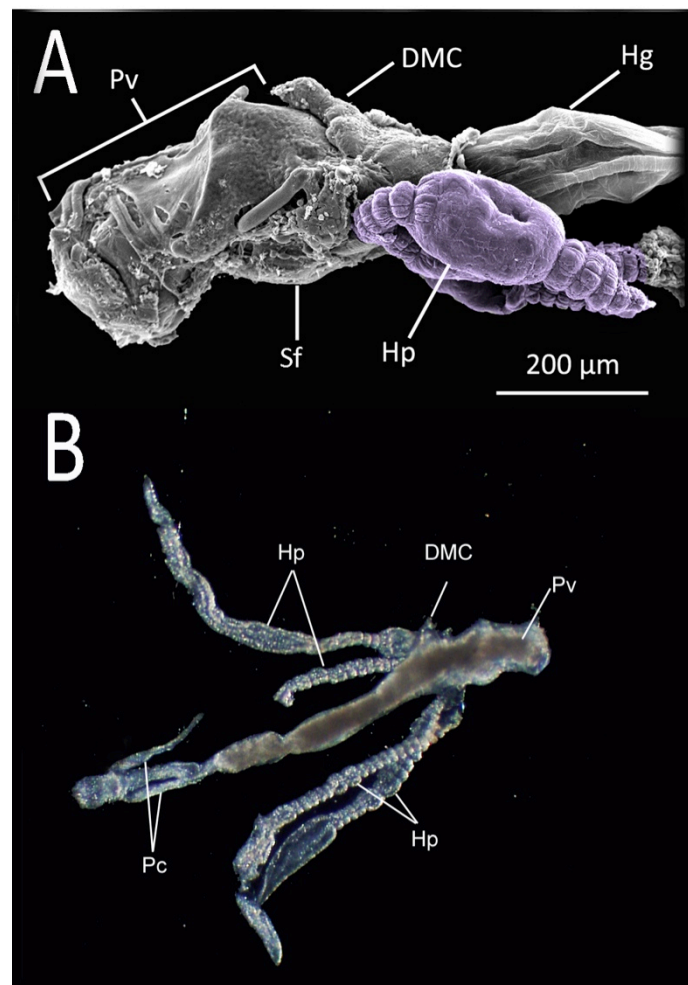
Figure 3.36 The hepatatopancreas in relation to the digestive system in *C. terebrans.* A) The proventriculus (Pv) showing the hepatopancreas (Hp) emerging from the ventral side at its posterior. B) A disected digestive system,. The proventruiculus (Pv) to the right leading onto the posterior digestive system. the hepatopancreas (Hp) lobes differ in length. DMC – dorsal median caeca; Hg – Gut; Pc – posterior caeca; Sf – secondary filter.
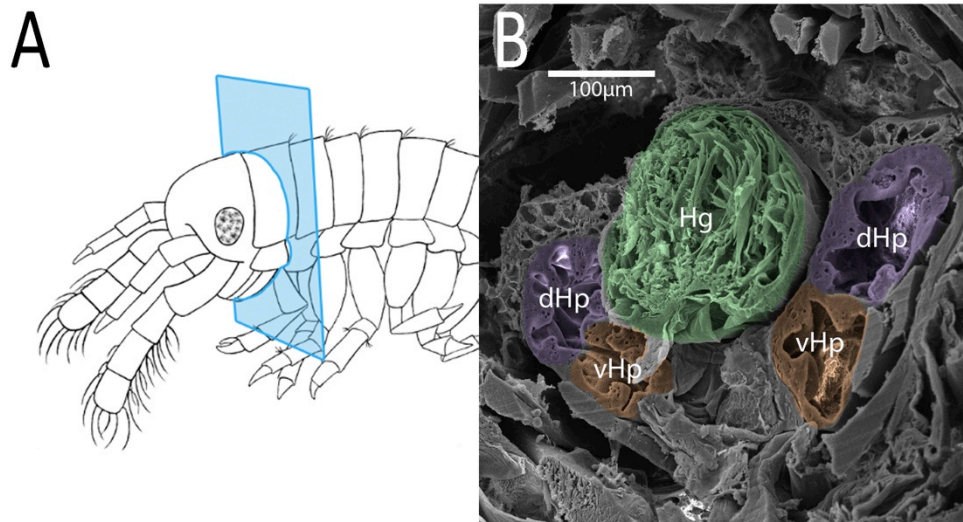
Figure 3.37 Transverse section of the hepatopancreas in *C. terebrans*. A) Orientation of section in B. B) The hepatopancreal lobes around the gut (Hg) each pair possess a ventral (vHp) and dorsal lobe (dHp) lateral to the gut.

The channels from the hepatopancreas pairs open ventrolaterally into the posterior of the hepatopancreas chamber, a narrow chamber at the posterior of the proventriculus. In this chamber a ventromedial ridge, the post inferomedianum, rises in its centre into the lumen of the chamber (Figure 3.38 & Figure 3.39). It has a two-pronged anterior each directed toward the filter channel on either lateral side of the inferomedianum, at the level of the hepatopancreal channels the prongs merge and the post inferomedianum widens at its posterior to form a concave lateral wall between the hepatopancreal channels and the filter channels.
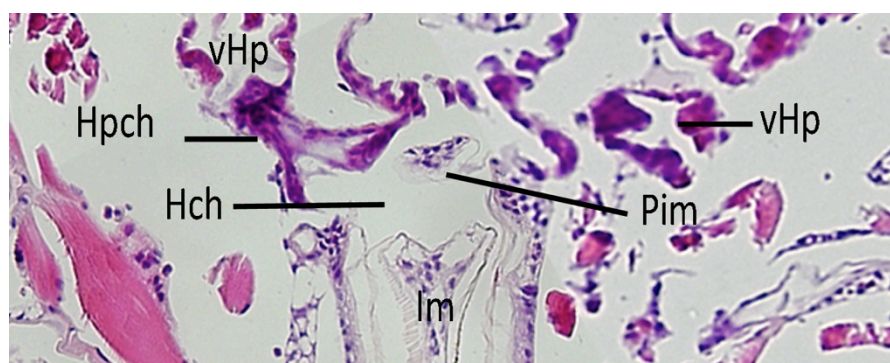
Figure 3.38 Post inferomedianum at the centre of the hepatopancreal chamber. Channels (Hpch) from the hepatopancreas (vHp) to the hepatompancreal chamber (Hch) with the posterior of the post inferomedianum (Pim) at its centre. Im - inferomedianum
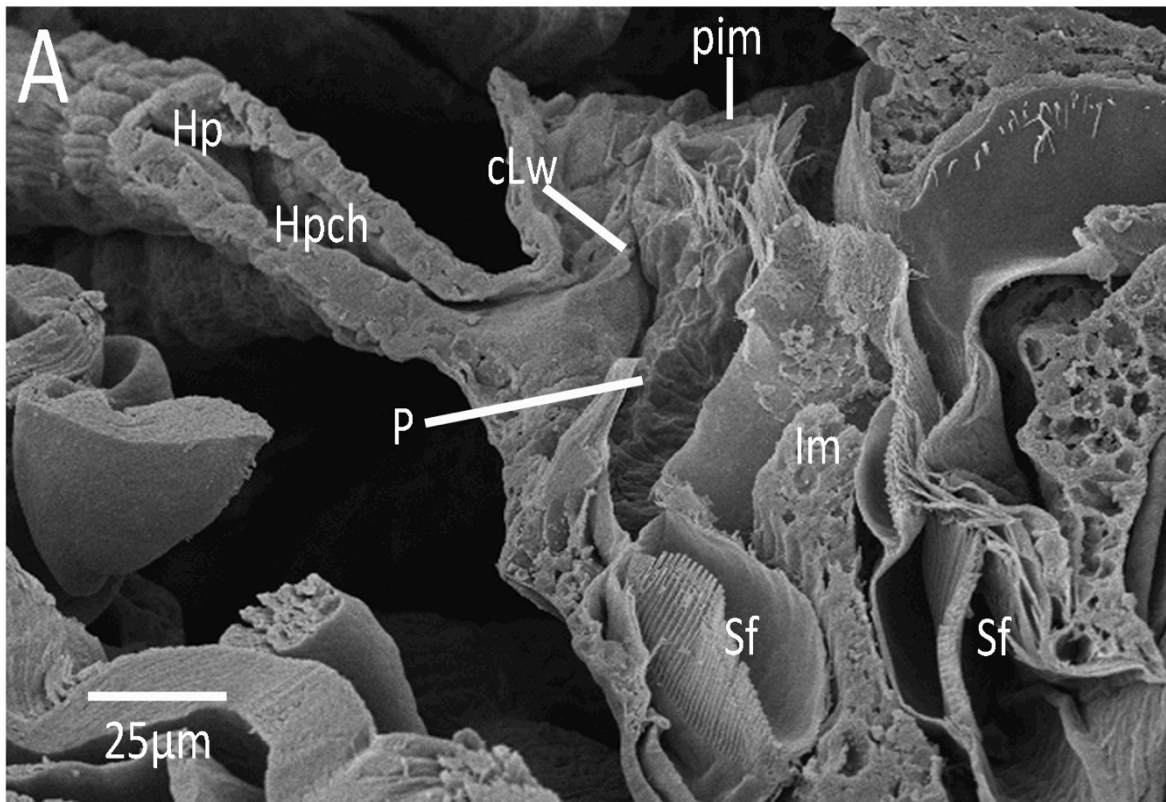
Figure 3.39 Hepatopancreal chamber of *C. terebrans*. Magnification of the post inferomedianum (Pim) showing the right lateral side. The right prong (p) making the concave lateral wall (cLw) between the hepatopancreal channel (Hpch) and the secondary filter (Sf).

The hepatopancreatic lobes have a meshwork of longitudinal and circular muscles, which are visible on the basal surface of the glandular epithelum under SEM, giving the lobes a segmented appearance (Figure 3.40A & B). The distal end of the lobes has a different surface appearance, here it appears wrinkled and the musculature is less defined (Figure 3.40B & C), this could be an artifact. Under the light microscope, the cells on the distal end also appear smaller than those in the more proximal regions, where many cells have a large vacuole (Figure 3.40D).
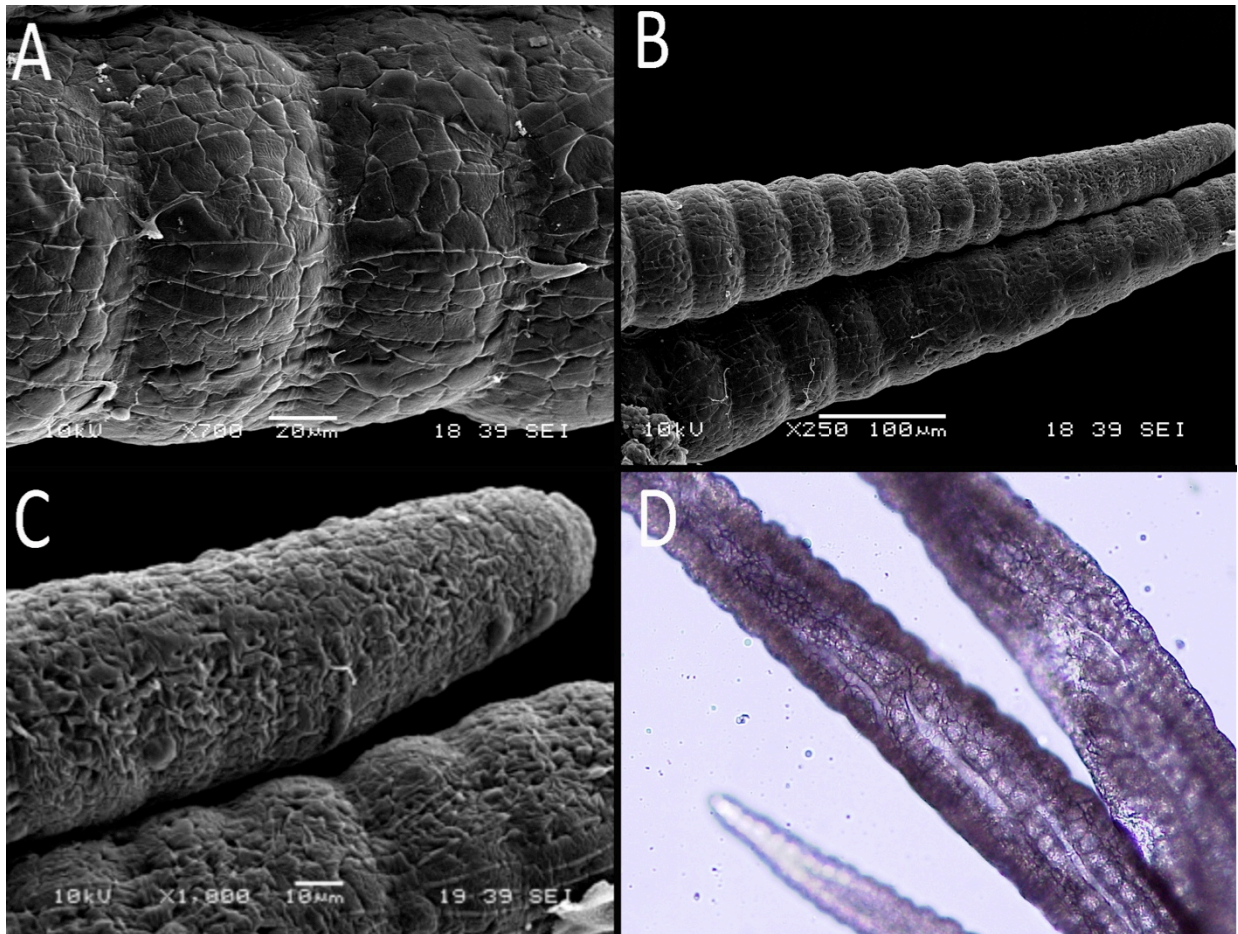
Figure 3.40 Hepatopancreal lobes of *C. terebrans*. A) Outer surface of the hepatopancreal lobes showing the musculature. B) Evident external surface transition from mid to distal region of hepatopancreas C) The distal end of the hepatopancreal lobe. D) Hepatopancreas viewed under a light microscope revealing the mid section cells with large vacuoles.

On the inner surface of the hepatopancreas, cell borders are visible and cells can be seen to bulge into the luminal space. The cells that line the inner surface have microvilli protruding into the lumen and are approximately 1.5µm in length (Figure 3.41).
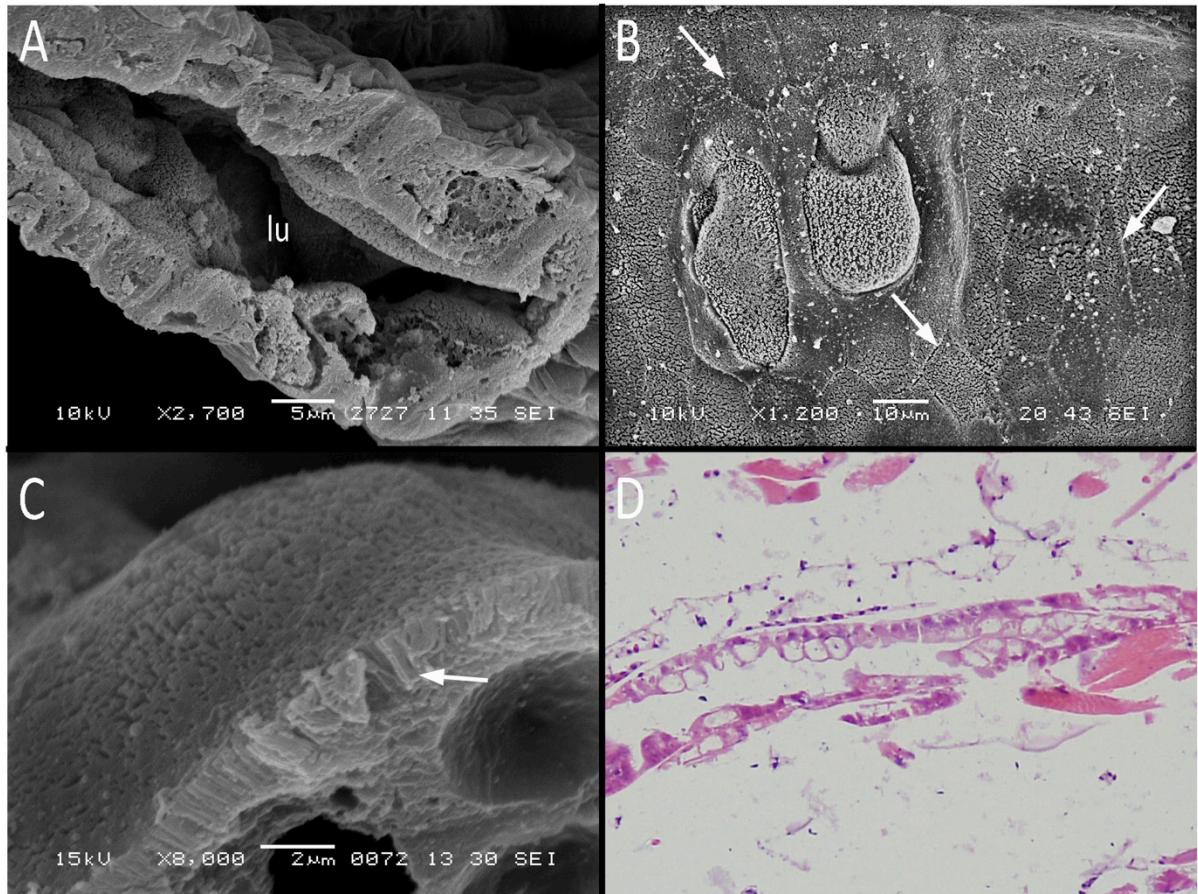
Figure 3.41 Inner surface cells lining the hepatopancreas of *C. terebrans*. A) Sagittal section through the hepatopancreas. Cells bulging on the inner surface into the lumen (lu) can be seen. B) The inner surface of the hepatopancreas when muscles are relaxed, showing a flat surface for the majority of this section and two larger bulging cells. Examples of cell borders are highlighted with arrows. C) Transverse section through the hepatopancreas showing the microvilli (arrow). D) A sagittal section through the middle of the mid region of the hepatopancreas showing cells with large vacuoles bulging into the lumen.

Transverse sections of the hepatopancreas examined using light microscopy revealed different cell types. Toward the distal end, cells were found to be columnar and narrow in shape, resembling the fibrillous or reserve cells (Figure 3.42A & B). Further toward the proximal regions some cells have large vacuoles. These cells were found to have protruded most into the luminal space, with the vacuole in these cells closer to the luminal surface. These cells resemble blast cells (Figure 3.42C – F).
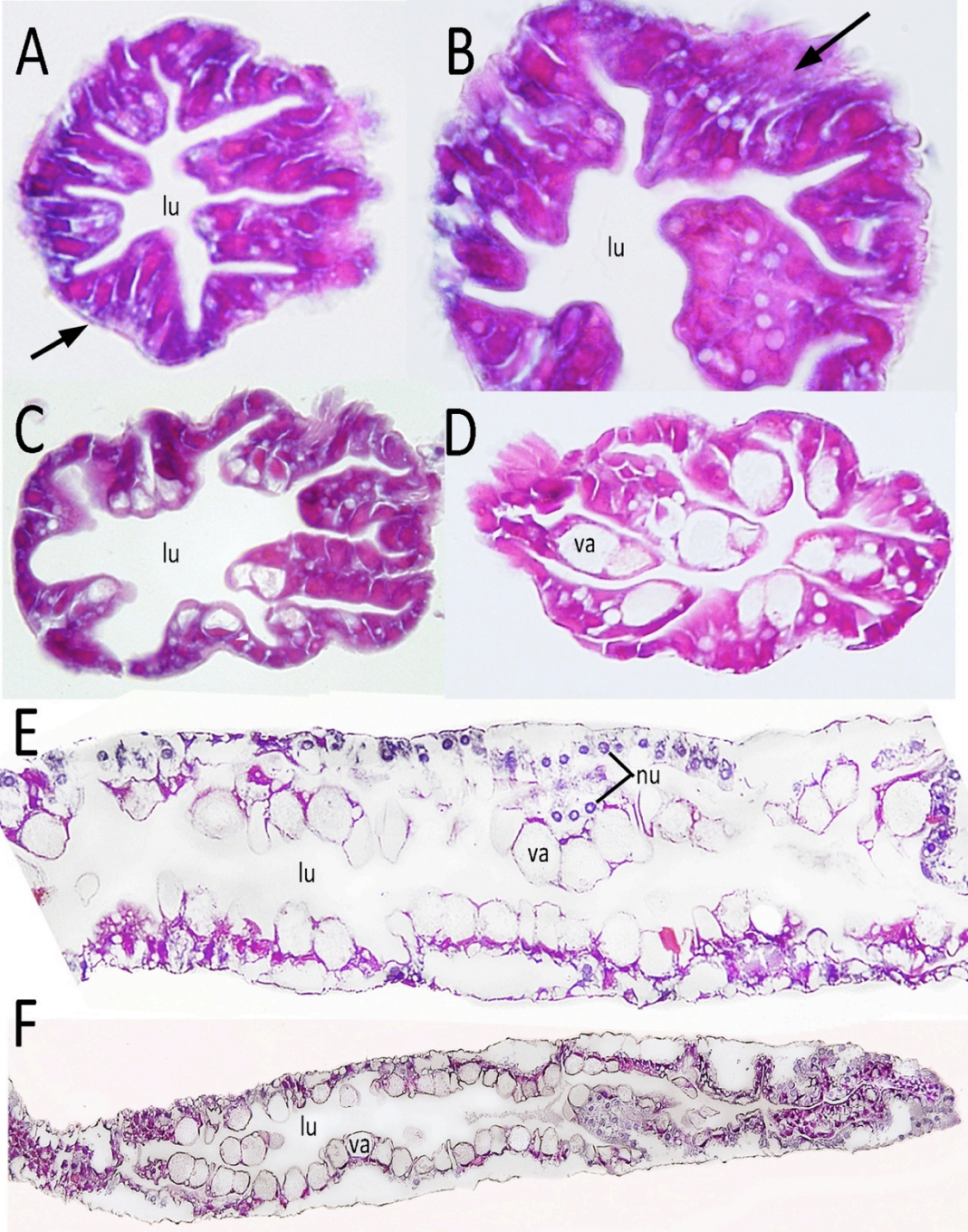
Figure 3.42 Hepatopancreas cells in *C. terebrans*. A) Transverse section of distal region of the hepatopancreas showing cells which are columnar in shape (arrow) surrounding the lumen (lu). B) Magnification of columnar cells. C & D) Sagittal section through the mid region of the hepatopancreas showing cells with larger vacuoles (va). E) Longitudinal section through the midsection of the hepatopancreas showing cells which are narrow at the basal edge, containing a clear nucleus (nu) and possessing large vacuoles (va) at the edge of the lumen (lu). F) hepatopancreal channel on the left side showing a large abundance of vacuolar cells at the proximal edge.

3.3.11   Dorsal median caecum

The dorsal median caeca is located at the posterior of the proventriculus and the anterior to the hepatopancreal junction. It begins in the first peraeoneal segment, on the lateral sides of the gut and extends rostrally over the dorsal surface of the pyloric chamber of the proventriculus (Figure 3.43A). Externally it is broad at its base on the lateral sides and narrows to a point on the dorsal side.
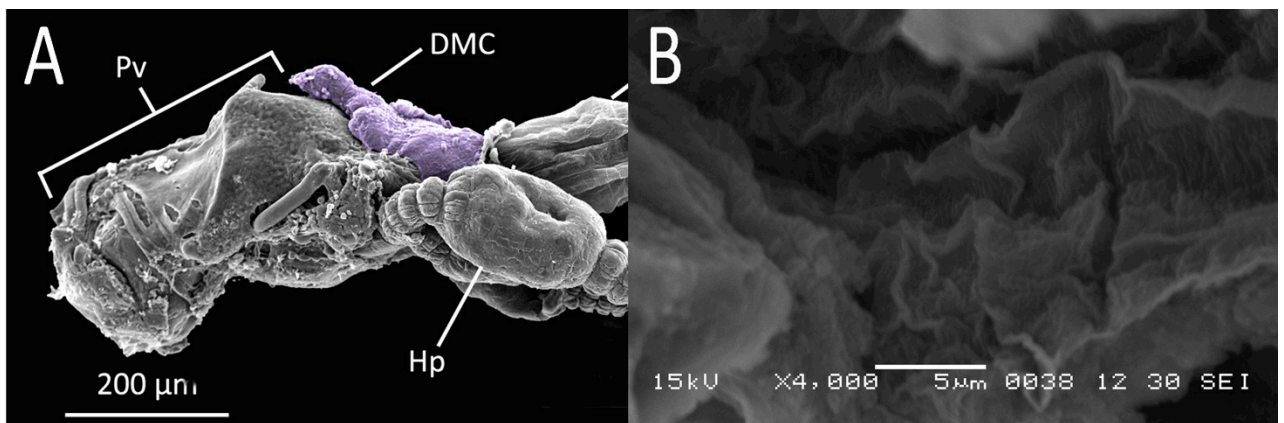
Figure 3.43 The dorsal median caeca in relation to the proventriculus of *C. terebrans*. A) The dorsal median caeca (DMC, purple) is found posterior to the proventriculus (Pv) and the hepatopancreas (B) The luminal surface of the caecum Hp). Hg – gut.

Sections stained using eosin and haematoxylin show the dorsal median caeca deeply stained (Figure 3.44). However it is difficult to detect any cellular structure. The internal surface at its tip is wrinkled and no microvilli are visible (Figure 3.43B), there is no external difference between the proximal or distal end. The dorsal median caeca does not have the same defined musculature as the hepatopancreatic lobes.
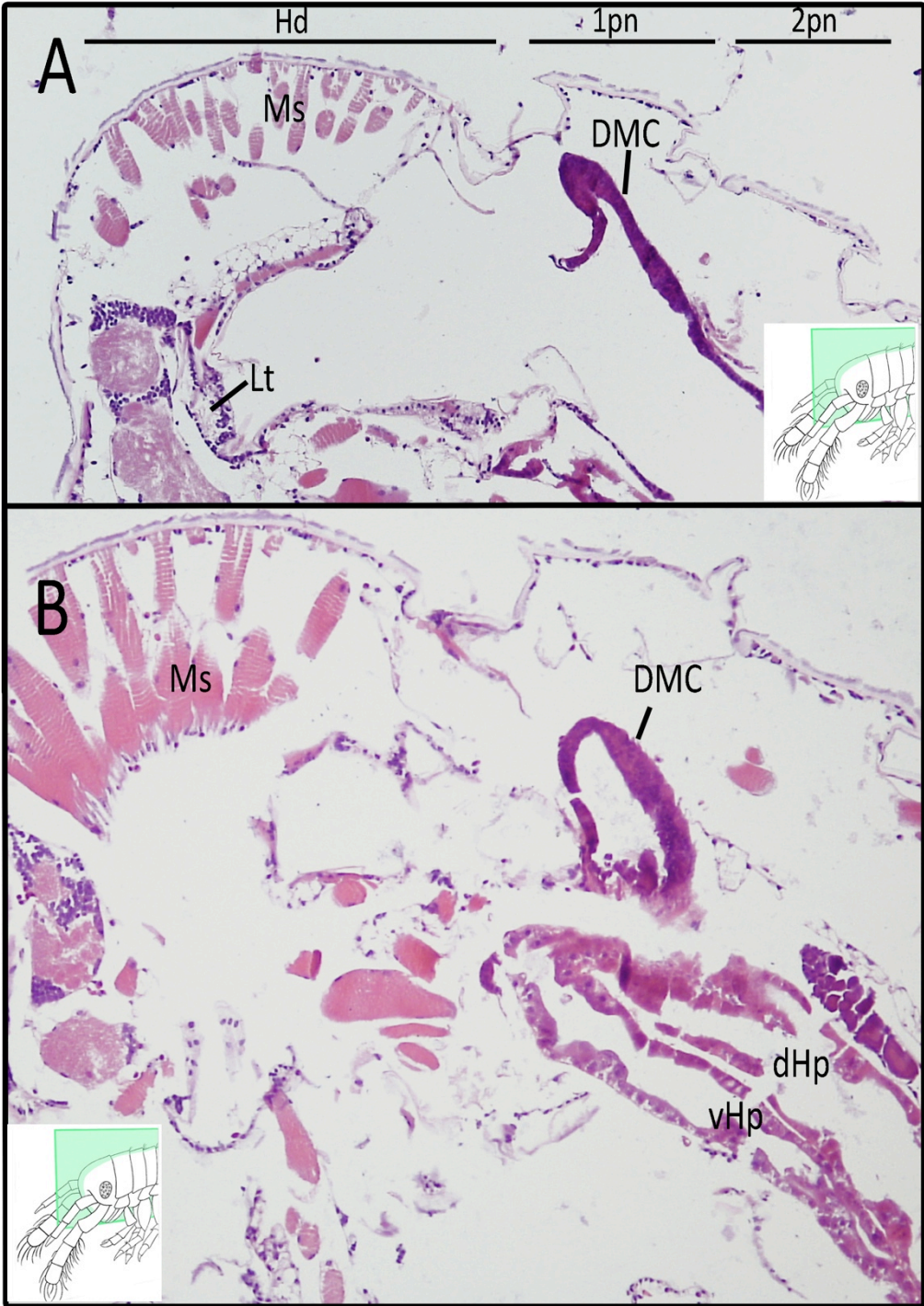
Figure 3.44 Sections of the *C. terebrans* dorsal median caeca viewed under a light microscope. Inset images show orientation of section. A) Saggital section of the head (Hd) and first peroneal segments (1 & 2Pn) close to the median, showing the tip of the dorsal median caeca (DMC). B) Magnification of a section to the right of the median line. The dorsal median caeca (DMC) is deeply stained and shows little cellular structure. dHp – dorsal hepatopanceas; vHp – ventral hepatopancreas; Ms – Muscle; Lt – lateralia.

3.3.12   Gut

The gut was found as described in Kühne & Becker (1964), a straight cylindrical tube for most of the length and begins at the posterior end of the proventriculus and extends to the anus, which is located under the telson (Figure 3.45A & B). The external surface of the gut has longitudinal and circular muscles. These muscles contract creating a wrinkled appearance (Figure 3.45C) when there is little food in the gut and appear to contract sharply behind a bolus (Figure 3.45D). These contractions were seen at various positions throughout the gut and are consistent with light microscopy studies of live animals where several bolus can be seen in the gut at any time.
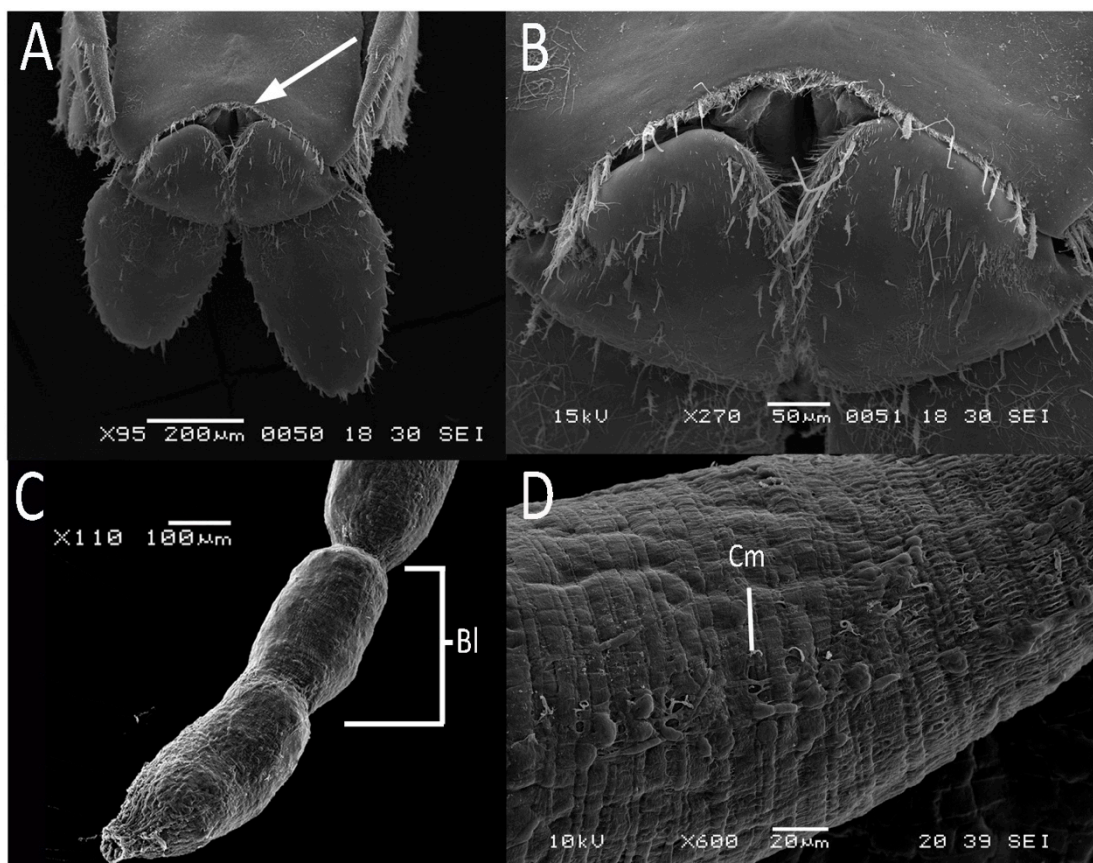
Figure 3.45 The anus and exterior surface of the gut in *C. terebrans*. A) Ventral view of the urosome of a female *C. terebrans* showing anus (arrow) B) Magnification of the anus in A, located under the third uropods. C) Circular muscle (Cm) of the gut, contracted on the right of the image. D) Sharp contraction of circular muscles anterior and posterior to the bolus (Bl).

99

It was not possible to define the distinction between the mid and hindgut in this study. A chitinous intima was found to line the gut from the second pereoneal segment to the posterior regions. The intima was found wrinkled where the muscles were contracted. The posterior end of the gut is lined by singular spines, approximately 1μm in length, (Figure 3.46).
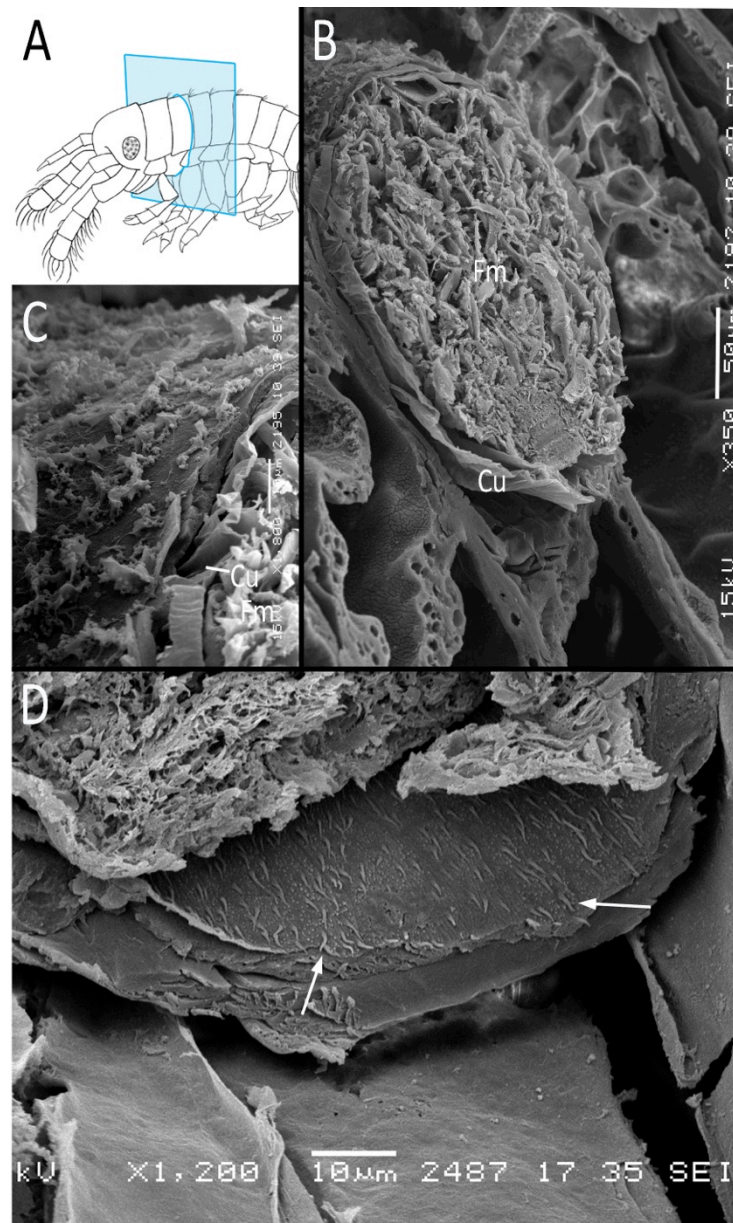
Figure 3.46 Cuticular lining and spikes at the posterior of the gut of *C. terebrans*. A) Orientation of sectioning in B and C. B) A transverse section through the gut in the second peraeonal segment, showing layers of the gut including the cuticle lining (Cu). C) Magnification of the gut wall at this location. D) The posterior lining with spines (arrow). Fm – food mass.

3.3.12.1 *Peritrophic membrane*

Tightly packed wood fragments were found throughout the gut lumen. From the second-third peraeoneal segment, the food mass is enclosed in a peritrophic membrane (Figure 3.47), however the exact location of the secretion of this membrane has not been identified. The membrane is thin and coats the food mass entirely. However, in some places there were small holes, though this could be an artefact of preparation techniques. The membrane appears to be attached to the food mass itself, as there was always a gap between the peritrophic membrane and the epithelium (Figure 3.47C). However, this could also be an artefact of preparation.
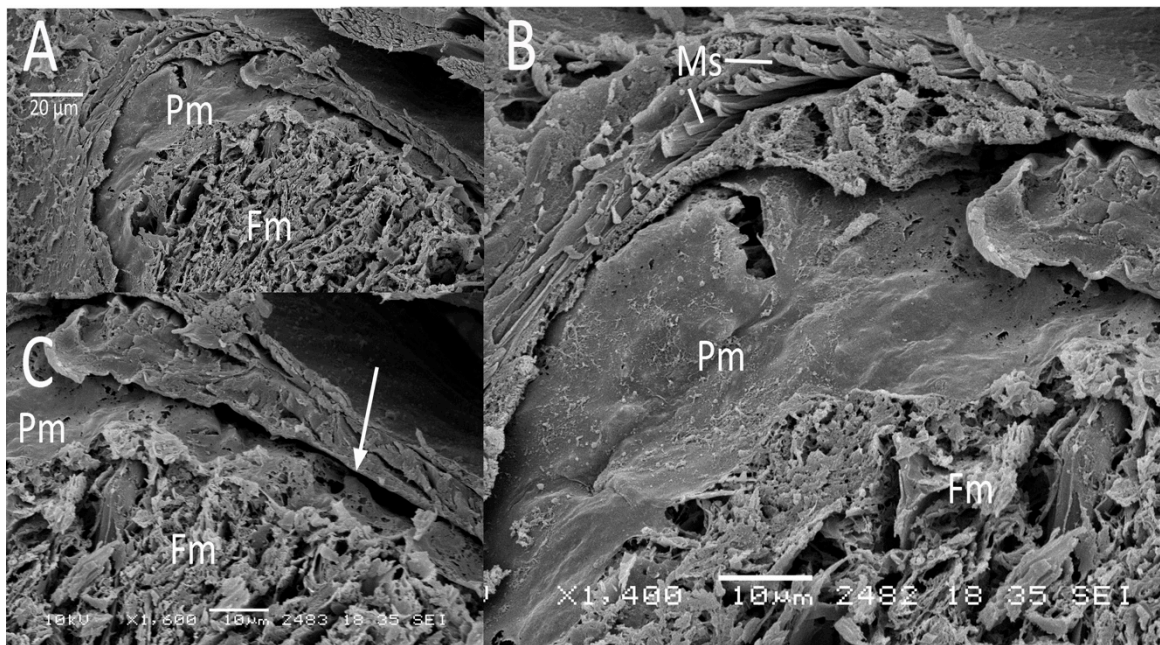
Figure 3.47 Peritrophic membrane in the gut of *C. terebrans.* A) Oblique cut through the gut showing the food mass (Fm) surrounded by the peritrophic membrane (Pm) enclosed in the gut wall (Hg). B) Magnification of A. C) Gap (arrow) between the peritrophic membrane. Ms – muscle.

3.4 Discussion

Generally, the superficial features of the digestive system of *C. terebrans* were found to correspond to the descriptions by Kühne & Becker (1964). Although our understanding of the proventriculus is not as detailed as other components of the amphipod digestive system, on the basis of the existing literature. It appears that the proventriculus of *C. terebrans* has the same fundamental organisation as that of other amphipods (Such as: Martin 1964; Coleman 1994, 1992; Strus & Storch, 2004; Davolos et al., 2010).

Using light microscopy, Kühne & Becker (1964) described the lateralia as posteriorly directed 'thorns', similar to the spines on the lateralia of many other amphipod species (such as: Davolos et al., 2010; Mekhanikova, 2010; Martin 1964; Coleman 1994, 1992; Strus & Storch, 2004). In this study, however, the SEM analysis revealed the lateralia were not simple "thorn"-shaped spines but more elaborate spines with multiple points and secondary spines on the caudal facing edges. The discrepancy between the findings in this study and those given by Kühne & Becker (1964) is likely to be the result the increased detail and dimensionality offered by SEM. The lateralia of *C. terebrans* appear to have evolved more elaborate spines, when compared to those described using SEM in the literature (Strus & Storch, 2004). Although the modification to the functioning of these structures is unknown, it is plausible that they could be specialised in *C. terebrans*, such that they aid in the mechanical degradation of wood. It tempting to speculate that the posteriorly directed spines cause larger wood fibres, caught between the 'claws' of opposing spines, to be splintered into fractions as a result of a scissor-like action.

In their studies Kühne & Becker (1964) were unsure whether the post inferomedianum structure mentioned in previous descriptions of amphipod anatomy (Martin, 1964) was present in the hepatopancreal chamber anterior to the hepatopancreas in *C. terebrans*. This structure was easily identified using the SEM techniques employed in this study. Moreover, with the advantage of the three dimensional view offered by SEM, it would

appear this structure acts to help direct digestive fluids anteriorly from the hepatopancreas into the dorsal channels of the proventriculus.

It is worth noting that the proventriculus and the gut was always filled with wood fragments, regardless of whether specimens were subjected to a week long starvation period before fixation. This provides more evidence for a truly xylophagus diet for *C. terebrans,* as first suggested by Allman (1847), as the level of wood fragments found were highly unlikely to be a by-product of foraging for bacteria, as suggested by Barnard (1955).

C. terebrans possesses both primary and secondary filters constructed of setae, which separate the food channel from the hepatopancreal chamber. The primary filter found at the anterior of the stomach is coarse but would be sufficient to stop larger food particles (>2μm) from entering the shallow filter channel. The secondary filter is found posterior to the primary filter on the lateral walls of the inferomedianum. This filter is much finer, the setae over lap and the setules on the setae are approximately 0.1μm apart. It is uncertain whether food particles are able to pass through this filter, though none were found in the hepatopancreal chamber or lobes. The size of the secondary filter is also likely to be sufficient to prevent bacteria from entering the hepatopancreal lobes.

Due to the size and colouration of *C. terebrans* it is difficult to observe the proventriculus working in live animals. However, using the observations seen in this chapter and descriptions of similar structures operational in other amphipods (Martin 1964; Coleman 1994, 1992; Strus & Storch, 2004, Davolos et al., 2010,) a prediction of how food processes through the proventriculus can be made. Food enters the oesophagus and, with the aid of posteriorly directed spines, makes its way to the proventriculus. The entrance to the cardiac chamber is guarded by lateralia possessing curved teeth; the food is triturated further as the lateralia rotate towards each other during cardiac region contractions (Martin, 1964). The food particles pass into the large

cardiac chamber where they are confined to the main channel, unable to pass into dorsal or ventral channels due to setae guarding their small openings. The contractions of the proventriculus, with the aid of many caudally directed setae on the surface and margins of outfoldings, cause the particles to move posteriorly into the main chamber of the pyloric region. The hepatopancreal lobes covered in many circular muscles contract causing digestive fluids to be ejected into the hepatopancreal chamber where the posterior inferomedianum directs the fluid either anteriorly along the dorsal filter chamber of the proventriculus or posteriorly into the gut. The fluids passing anteriorly through the filter chamber pass into the anterior of the cardiac chamber where they are then deflected posteriorly through the empty dorsal channel. This circulation of digestive fluids would aid the saturation of the food mass. The contraction of the circular muscles on the ventral side of the proventriculus causes the food mass to be pressed by the inferolateralia below. When these muscles are then relaxed the inferolateralia return to their original position and the proventriculus regains normal capacity. With this action, digestive fluids are sucked down into the filter chamber. This can be achieved either through the primary filter or passing through the setal margins of the posterior inferomedianum. Channels direct fluid to the secondary filter before returning to the hepatopancreas, this can be achieved either ventrally through the filter channels or dorsally from a channel created by the posterior wall of the proventriculus.

The morphological analysis of the *C. terebrans* digestive system carried out in this study suggests that the overall anatomical architecture closely resembles the digestive systems of previously examined amphipod species. This conclusion was also reached following examinations of *Macarorchestia remyi* (Davolos, 2012), an amphipod that degrades wood with the aid of bacterial symbionts. However, there were modifications to the lateralia seen in *C. terebrans* that were not seen in *M. remyi* (Davolos, 2012). Such modifications, although relatively subtle, occur at the location important for the processing of food once it has entered the digestive tract and may be important

adaptations for a truly xylophagus amphipod. A further understanding of the digestive system in *C. terebrans* could be inferred by the use of transmission electron microscopy to explore its cellular structure (Sousa et al., 2005; Strus & Storch, 2004).

4 Primary nutrition and symbiotic relationships of *Chelura terebrans*

4.1 Introduction

In the admittedly limited literature, there are no reports of *C. terebrans* occurring in the wild in the absence of limnoriids. Indeed, much of the literature discusses their co-habiting nature and some of the earlier work seems to confuse the two organisms (Kühne & Becker, 1964). The extent of their association has generated speculation on the possible advantages that *C. terebrans* receives from the relationship, ranging from shelter and protection to dietary benefits, and it has even been suggested that they cannot survive without limnoriids (Kühne & Becker, 1964). However, investigation into the tunnelling capabilities and feeding habits of *C. terebrans* and has shown that they are able to degrade wood without the aid of limnoriids spp*.* (Barnard, 1955; Cragg & Daniel, 1992; Kühne & Becker, 1964). Barnard (1955) explained their relationship by suggesting that in Nature, due to the mode in which they attack wood, *C. terebrans* would remain exposed on the surface for too long and would risk predation, so would probably require limnoriids for their prior tunneling activities. Observation of the association between *C. terebrans* and species from the limnoriidae*,* have led to the hypothesis that *C. terebrans* consumes limnoriid faecal pellets. Kühne and Becker (1964) performed laboratory investigations on groups of *C. terebrans* and found that faecal pellets from *Limnoria tripunctata* and from the house longhorn beetle (*Hylotrupes bajulus*) provided sufficient nutritional value for survival, but not for rearing, this was also found to be the case with sapwood. They concluded that even though *C. terebrans* are capable of degrading the wood by themselves in laboratory conditions, they were unable to live more than several generations without the presence of limnoriids.

Regardless as to the origin of ingested wood fragments, either from limnoriid faecal pellet matter or degradation of the wood itself, *C. terebrans* still requires a capacity for the digestion of wood for nutrition. The digestion of lignocellulosic material has been studied in a variety of terrestrial and marine species, where it was consistently observed that these animals required microbial associations, to at least, partially aid in its digestion (Distel, 2003; Watanabe & Tokuda, 2010; Yang et al., 2009). An interesting observation made in several studies (Boyle & Mitchell, 1978, Boyle & Mitchell., 1980) is the apparent lack of microorganisms in the gut of *C. terebrans*. The study of their digestive tract by light and scanning electron microscopy failed to detect any resident gut flora. Boyle & Mitchell (1978) used scanning electron microscopy (SEM) to look at the surfaces of digestive tracts from three wood-boring peracarid crustaceans sustained by a diet entirely, or almost entirely, of lignocellulose. They observed an apparent absence of microorganisms in the digestive tract of the marine amphipod *Chelura terebrans* as well as isopods from the family Limnoriidae.

The failure to observe significant levels of microbiota, in conjunction with a study demonstrating cellulase activity in extracts from *Limnoria lignorum* (Ray & Julian, 1952), has led to the hypothesis that *C. terebrans* and limnoriids produce endogenous enzymes capable of wood degradation. Furthermore, *L. quadripunctata* has since been shown to produce transcripts for numerous glycosyl hydrolases (GHs) such as those belonging to GH7 and GH9 families, which include cellobiohydrolases (King et al., 2010). Although these findings are suggestive of truly endogenous enzymatic capabilities, cellulolytic symbiotic bacteria are known to reside intracellularly in some wood-boring invertebrates, for example, isopods (Kostanjsek et al., 2004a, 2004b) and bivalves (Distel et al., 2002; Distel & Roberts, 1997). However, despite the fact that intercellular bacteria would have eluded SEM investigations, the possibility of bacterial symbionts in *C. terebrans* and limnoriids has never been studied using sensitive

molecular techniques, leaving the possibility that such bacteria could contribute to the digestion of lignocellulose.

In order to investigate whether *C. terebrans* is reliant on the presence of *L. quadripunctata,* this study observed the feeding habits of *C. terebrans* and *L. quadripunctata,* alone and together, in order to determine how the cohabitation affects the rate of faecal pellet production and reveal any dietary preference towards coprophagy. To investigate a possible relationship with lignocellulolytic bacteria, this study will also compare the bacterial burden of the *C. terebrans* hepatopancreas and gut with those of other peracarid crustaceans. Quantitative PCR techniques were used to estimate the ratio of bacterial sequences to those of the animal itself. This ratio was then compared to that of the wood borer *Limnoria quadripunctata,* the terrestrial detritivore *Porcellio scaber* (well known for its hepatopancreatic microbial community) and the algal-feeding amphipod *Echinogammarus marinus.*

4.2 Methods

4.2.1 *Chelura terebrans* culturing

Cultures of *C. terebrans* were kept in tanks at the Institute of Marine Sciences. These tanks only contained *C. terebrans* populations which were fed on a variety of wood including Scots pine and greenheart in the form of blocks or planks (see 3.2.1). The tanks have a natural seasonal regime with light and water temperatures reflecting those of the southern coast of the United Kingdom. However, the tanks have no tidal regime.

4.2.2 Study of faecal pellets and food mass

Chelura terebrans Philippi, specimens from laboratory cultures (see section 3.2.1) and faecal pellets were fixed (3 % glutaraldehyde buffered in 0.2 M cacodylate pH 7.4) and dehydrated in a graded ethanol series before being embedded in paraffin. Specimens were sectioned (Leica Jung Biocut 2035), de-waxed in xylene, dehydrated through a graded ethanol series and taken through transitional steps (100 % EtOH to 100 % hexamethyldisilazane, HMDS) then evaporation dried. The dry samples were mounted on SEM stubs, sputter coated with gold-palladium, and examined for the presence of microflora in the digestive tract and faecal pellets using a scanning electron microscope (JEOL JSM-6060LV) in secondary electron and high vacuum mode at an acceleration voltage of 10kv.

Faecal pellets were observed with a bright-field compound microscope (Leica DM LB2) and images captured using with a digital camera (JVC KY-F1030U). Other faecal pellet specimens were placed between polarising filters, one attached to the lens and the other below the specimen on top of the stage plate. The filters were rotated so that they allowed light in at right angles to each other, effectively blocking the light coming through to the lens.

4.2.3 Feeding experiment

4.2.3.1 *Substrate preparation and animal collection*

Scots pine sapwood (SPS) chips (machined using a fine saw into chips of 0.5 x 0.3 x 2cm) were pre-leached in filtered and autoclaved seawater in a vacuum desiccator. This water was replaced twice prior to the experiment to remove the most soluble extractives before the experiment. *C. terebrans* and *L. quadripunctata* were collected from mixed cultures kept in a tank at the IMS (Portsmouth see section 3.2.1) and were identified using a stereomicroscope.

4.2.3.2 *Test groups, pellet collection and wood weight*

Animals were separated into groups and placed in multi-well culture trays (Table 4.1). Animals were placed in cylindrical wells (22mm Ø) filled with seawater that had been autoclaved and filtered (0.2 μm) and a single SPS wood chip. All the groups contained 12 replicates and had their faecal pellets collected daily for three weeks. The groups and their replicates were duplicated and these had their faecal pellets collected at the end of three weeks.

Table 4.1 The number and combination of animals in each group for feeding study

	Group				
	A	B	C	D	E
C. terebrans	1	1	0	2	0
L. quadripunctata	1	0	1	0	2

The samples were kept out of direct light at 20 ± 1°C for three weeks. For those groups that had their faecal pellets removed, this was completed at the same time each day using a Pasteur pipette to cause minimal disturbance to the animal. The faecal pellets of the two borers are distinctive in cross section (Figure 4.1), which allowed them to be individually counted under a stereomicroscope.

Figure 4.1 Faecal pellets of *L. quadripunctata* (left) and *C. terebrans* (right) in cross section

After the 3 weeks the SPS wood chips used and control SPS chips, which were left in autoclaved filtered seawater without animals were saturated with distilled water (with 5 water changes) and then dried at 60°C for 48hrs and weighed. Results were analysed [Minitab, Inc. (2005). MINITAB statistical software] and significant differences calculated using Tukey pairwise comparison.

4.2.4 PCR Assay to detect symbiotic bacteria

4.2.4.1 *Animal and substrate collection*

P. scaber were collected from Fort Cumberland (Grid reference: SZ 67992 98979) and *E. marinus* from Langstone Harbour (Grid reference: SZ 67516 99350), both *C. terebrans* and *L. quadripunctata* were collected from culture tanks at Portsmouth University. All were identified using a stereomicroscope (Leica DM LB2). Food substrate was collected simultaneously for *P. scaber* (leaves) and *E. marinus* (*Ascophyllum nodosum*). SPS wood wafers were left to soak in the culture tank for two weeks prior to the experiment,

allowing bacteria but not animals to colonise their surfaces, and then used to feed the *C. terebrans* and *L. quadripunctata*.

4.2.4.2 *Animal preparation*

The collected animals were kept in isolated tanks for one week along with their food substrate. Tanks for *Chelura* and *Limnoria* samples had a constant flow of running seawater from Langstone harbour (see 3.2.1). *E. marinus* were kept in a tank of shallow (less than 5cm deep) running seawater also from Langstone harbour, the tank contained *Ascophyllum nodosum* , rocks and pebbles for hiding places to try to prevent cannibalism. *P. scaber* were kept in a tank with leaf substrate dampened with dH_2O.

4.2.4.3 *Dissection and sample preparation*

All animals were individually dipped in 70% ethanol and then in either seawater that had been filtered (0.2 μm) and autoclaved (for marine species) or dH_2O (for terrestrial species); this was repeated three times before blotting dry and dissecting the animal under a stereomicroscope. The hepatopancreas, gut and muscle tissue were all placed in separate mircocentrifuge tubes containing 180μl buffer ATL and 25μl proteinase K (both Qiagen). The average size of each species dictated the number of animals needed to extract enough DNA to make a sample, however in all cases, tissues from at least 5 animals were combined (*P. scaber* n=5, *L. quadripunctata* n=100, *C. terebrans* n=30, *E. marinus* n=10) Three replicate groups were made for each group except for *L. quadripunctata* that had two replicates due to the large amount of animals required to obtain a viable sample.

4.2.4.4 *DNA extraction*

DNA was extracted from the feeding substrate as well as the dissected hepatopancreas and gut from each animal using a DNeasy blood/tissue extraction kit (Qiagen). DNA isolation protocols for animal tissue were combined with those for gram-negative and

positive bacteria (Sarma-Rupavtarm et al., 2004) shown in brief in Figure 4.2. Samples containing tissue/substrate, ATL buffer and proteinase K were ground using a pestle and mortar, then and vortexed for 15 seconds before incubating at 52°C overnight (at least 12hours) with another two 15 second vortexing in between. After overnight incubation the samples were centrifuged and the pellet and the supernatant were separated. The pellet was resuspended in lysis buffer (20 mM Tris·Cl pH 8. plus 2 mM sodium EDTA and 1.2% Triton® X-100) and the proteinase K inactivated by incubation at 80°C for 10 minutes. Lysozyme was then added to 20 mg mL^{-1} and incubated for 45 min at 37°C. RNase A (400µg) was added and left for 2 minutes at room temperature before more proteinase K and AL buffer, 25µl and 200µl respectively, was added and then incubated for 30 minutes at 70°C. After incubation, ethanol was added to give a final concentration of 33% v/v. To the supernatant, 400µg of RNase A was added to the lysate solution and this was then incubated at room temperature for 2 minutes followed by the addition of ethanol to a final concentration of 33%. The supernatant lysate and the pellet lysate were applied together into the Qiagen spin-column. This was then washed with AW1 and AW2 buffer and then eluted with buffer AE following manufacturers instructions.
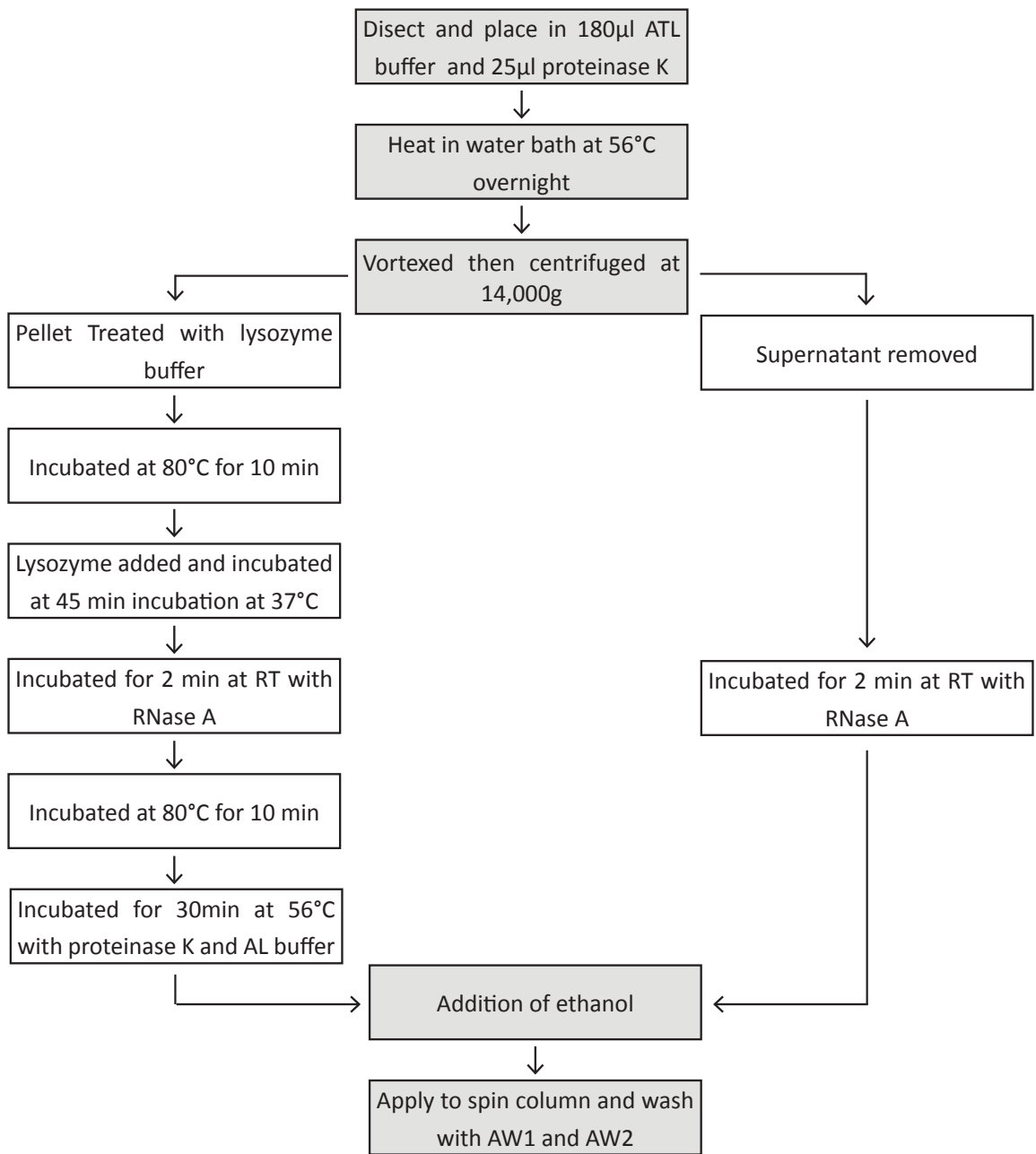
Figure 4.2 Schematic illustrating the steps for the DNA isolation protocol combining that for animal tissue with those for gram-negative and -positive bacteria (Sarma-Rupavtarm et al., 2004).

4.2.4.5 *Wolbachia screen*

DNA samples from *P. scaber* individuals were screened for the presence of *Wolbachia* a bacterial endosybiont of woodlice, using primers specific to the *Wolbachia* 16S ribosomal RNA gene (Werren & Windsor, 2000). Only the DNA of individuals clear of *Wolbachia* infection was used in the bacterial amplification (see 4.2.4.8).

4.2.4.6 *Haemocyanin Primers*

Primers were designed using Primer 3 software (Rozen & Skaletsky, 2000) for each animal corresponding to a 120-129bp fragment of a haemocyanin gene. Sequences from GenBank were used for *P. scaber* (ACS44711.1), *L. quadripunctata* (GU166295.1) and *E. marinus* (personal communication from Drs Yang & Short see section 6.7). Primers for *C. terebrans* were designed using sequences from the transcriptome (see Chapter 6)

4.2.4.7 *Bacterial primers*

Bacterial primers used in this experiment needed to amplify a sequence suitable for qPCR which were also capable of amplifying a wide range of bacterial groups.

Multiple 16S gene primer sequences previously described as 'universal' (Lane et al., 1991; Turner et al., 1999) were tested in various combinations to establish the primer set producing least primer-dimer artefact. These primer sequences were then compared to bacterial 16S ribosomal RNA gene sequences stored in GenBank using a BLAST analysis. These comparisons led to the prediction that the chosen 'universal' primers (1237F and 1391R, Turner et al 1999) would indeed amplify the 16S ribosomal gene sequences of bacteria belonging to the majority of bacterial phyla. The potential exceptions being Cyanobacteria and Chlorobi due to a single base pair mismatch in the

forward primer. In addition, the phylum Planctomyces, some classes of Actinobacteria, and the class β-proteobacteria may not be detected due to a two base pair mismatch in the forward primer. Furthermore, there is a single base pair mismatch between the reverse primer and bacteria from the class ε-proteobacteria. However, it should be noted that the primers might still amplify sequences from these groups as none of the mismatches occur in the three 3' nucleotides most likely to hinder or prevent primer extension. Overall, amplification of the bacterial 16S ribosomal RNA gene using the 1237F and 1391R primers offered the best combination of sensitivity, qPCR suitability and universality. Further consideration of primer choice and their limitations are discussed later (section 4.4).

Table 4.2 Primers for host genomic sequences and bacterial 16S sequences used for bacterial quantification.

Species	Primer	Forward (5'- 3')	Product size
C. terebrans haemocyanin	CtHc1F	CGG TGT CCA TGG AGA GAA GT	129bp
	CtHc1R	GAG AAT GTA ACT TTC ACG ATT TGG	
L. quadripunctata haemocyanin	LqHc1F	TGA AAC CTA CCC AGA CAA ACG	121bp
	LqHc1R	TTC TCT GTG CTT GAT GGT AAC AA	
P. scaber haemocyanin	PsHc1F	CCC AGA CAA ACG TCC TCT TG	120bp
	PsHc1R	TCA TGA TGG TCA TCA AAG TGA A	
E. marinus haemocyanin	EmHc3F	TTC TCT GTG CTT GAT GGT AAC AA	123bp
	EmHc3R	GGA CAG AAA TAC CCC GAC AA	
Bacteria 'Universal' 16S	1237F	GGG CTA CAC ACG YGC WAC	154bp (*E. coli*)
	1391R	GAC GGG CGG TGT GTR CA	

4.2.4.8 *Amplification*

All reactions were performed in triplicate in a 15µl reaction mixture volume containing 10ng of template, 0.27 µM of each primer and qPCR mix containing SYBR® Green (QuantiTec® PCR kit, Qiagen). The PCR was performed using the following conditions, 15min, 95°C heated step followed by 47 cycles of 15s at 95°C, 30s at 60°C and 15s at 72°C (Eco Illumina).

4.2.4.9 *Bacterial identification*

Bacterial PCR products of particular interest were then cleaned using a QIAquick-spin PCR purification kit (Qiagen), sequenced (Sanger Sequencing, Source Biosciences UK) before blastn was used to make comparisons against sequences from Genbank (BLAST, NCBI).

An ~1.3kb fragment of the 16S ribosomal DNA gene from the bacteria found in the *L. quadripunctata*. This was amplified from the original hepatopancreas sample using the primers 27F, 1459sR (Lane, 1991) using the PCR conditions 94°C for 45s, 58°C for 30s and 72°C for 30s for a total of 40 cycles (G Storm GS1) before being sequenced and analysed as described above.

4.2.4.10 *Screen for hepatopancreatic bacteria*

Primers were designed for the bacteria found in the *L. quadripunctata* (LimDLSF – TGG CAG TGA CAA AG AGT TGC, LimDLSR – CAA CAT GCT GAT TTG CGA TT) and PCR performed on all samples (G Storm GS1) using the PCR conditions 94°C for 45s, 58°C for 30s and 72°C for 30s for a total of 40 cycles.

4.3 Results

4.3.1 Chelura terebrans culturing

Several cultures of isolated *C. terebrans* populations have been well established for just under 4 years. These populations have flourished from a start population of approximately 500 animals. They have been fed on a variety of wood, degrading it to the extent of destruction (Figure 4.3.).
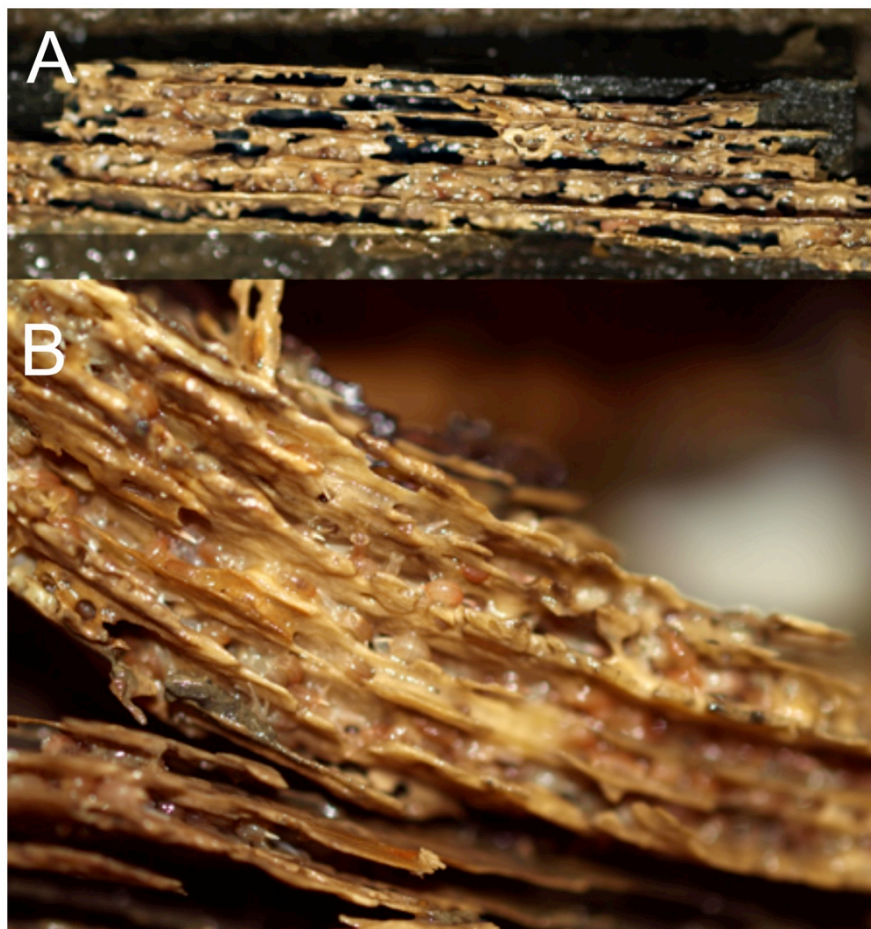
Figure 4.3 Wood destruction by *C. terebrans* in laboratory conditions.

The cut of the wood also appears to influence the mode of attack employed by *C. terebrans* (Figure 4.4). Sometimes the wood appears to be degraded at the softer earlywood before the denser latewood, creating their characteristic troughing (Figure 4.4A), though the latewood is also degraded when little earlywood is left. However, on

other woodcuttings the attack appears to be less discriminating (Figure 4.4B), and *C. terebrans* work from the uppermost surface inwards creating layers, which are eventually, weaken to the point of falling away if moved.
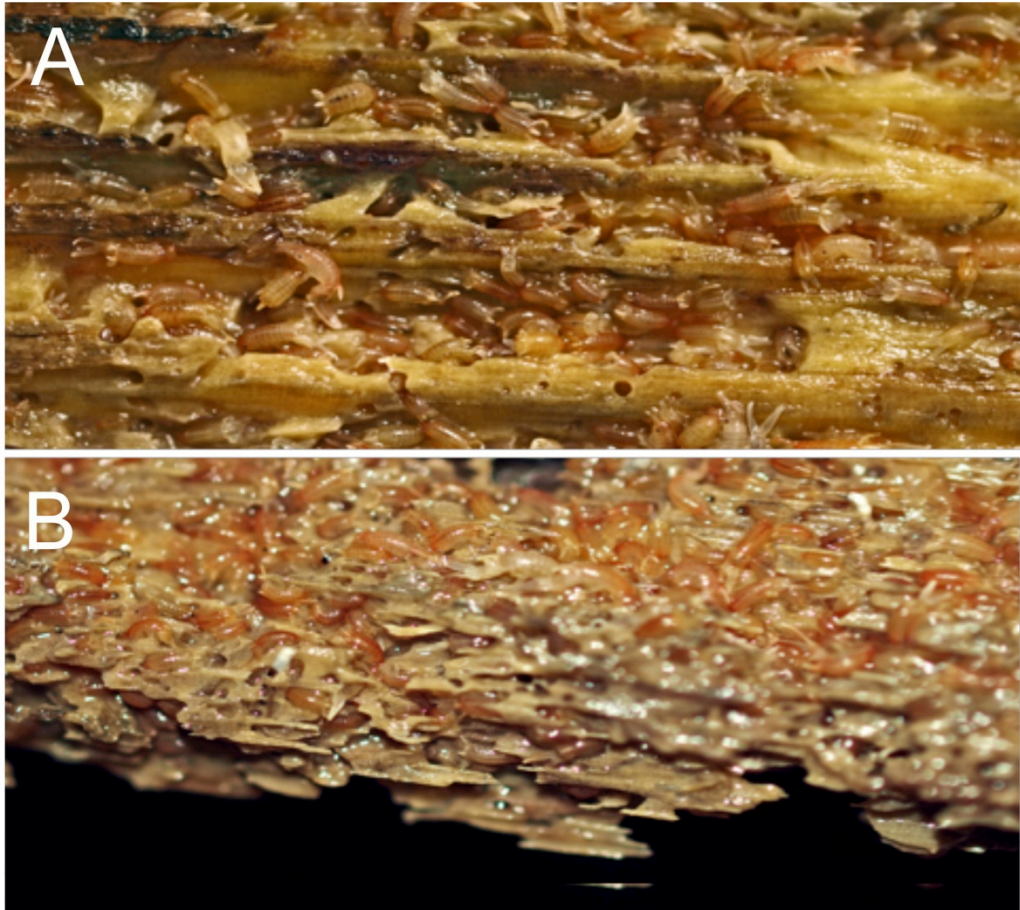
Figure 4.4 *C. terebrans* wood degradation strategies. A) Troughs in the early wood created by *C. terebrans*. B) Less discriminating wood degradation by *C. terebrans.*

After three years in the aquarium tanks (see 3.2.1) *C. terebrans* was found to dominate cultures originally consisting of various wood borers. The *C. terebrans* populations are visible in large groups on the surface with very little *L. quadripunctata* left on or in the wood. Furthermore, *C. terebrans* were found to be the only remaining species on older cut wood that was originally inhabited by many other species (Figure 4.5). On this piece of wood, formally a whole cross section, not only were *C. terebrans* the last inhabitants

of this dense wood, they were flourishing. Here we can see that both the heartwood (Figure 4.5A) and the sapwood have both been degraded, however, the knots in the wood appear to remain more or less intact (Figure 4.5B).
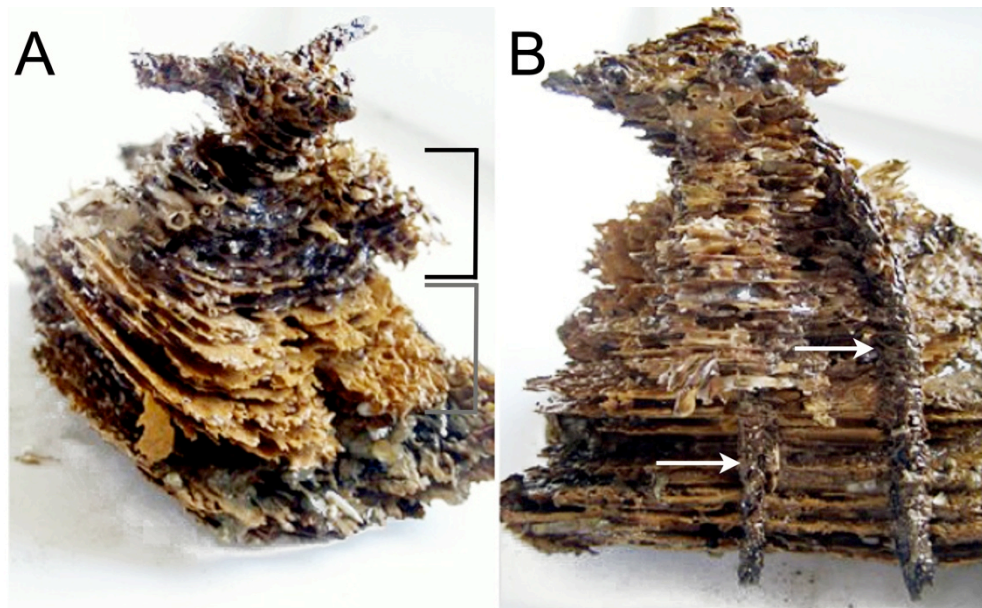
Figure 4.5 Remains of log degraded by *C. terebrans* A) Sapwood (light coloured and heartwood (dark coloured) degraded by *C. terebrans*. B) Relatively intact wood knots, indicated by arrows.

4.3.2 Study of faecal pellets and food mass

The scanning electron microscopy did not reveal microbes on the luminal surfaces of the hepatopancreas, gut or dorsal caecum of *C. terebrans* (Figure 4.6A & B). Densely-packed wood flakes were found in the food mass of the digestive tract and faecal pellets. Wood fragments in the digestive tract were found in a size gradient, fragments found on the dorsal side were generally larger than the fragments found at the ventral side which were much finer and much more densely packed (Figure 4.6C). Some of the wood flakes in the food mass showed evidence of bacterial degradation (Figure 4.6D) similar to that shown by erosion bacteria (Björdal, 2012).
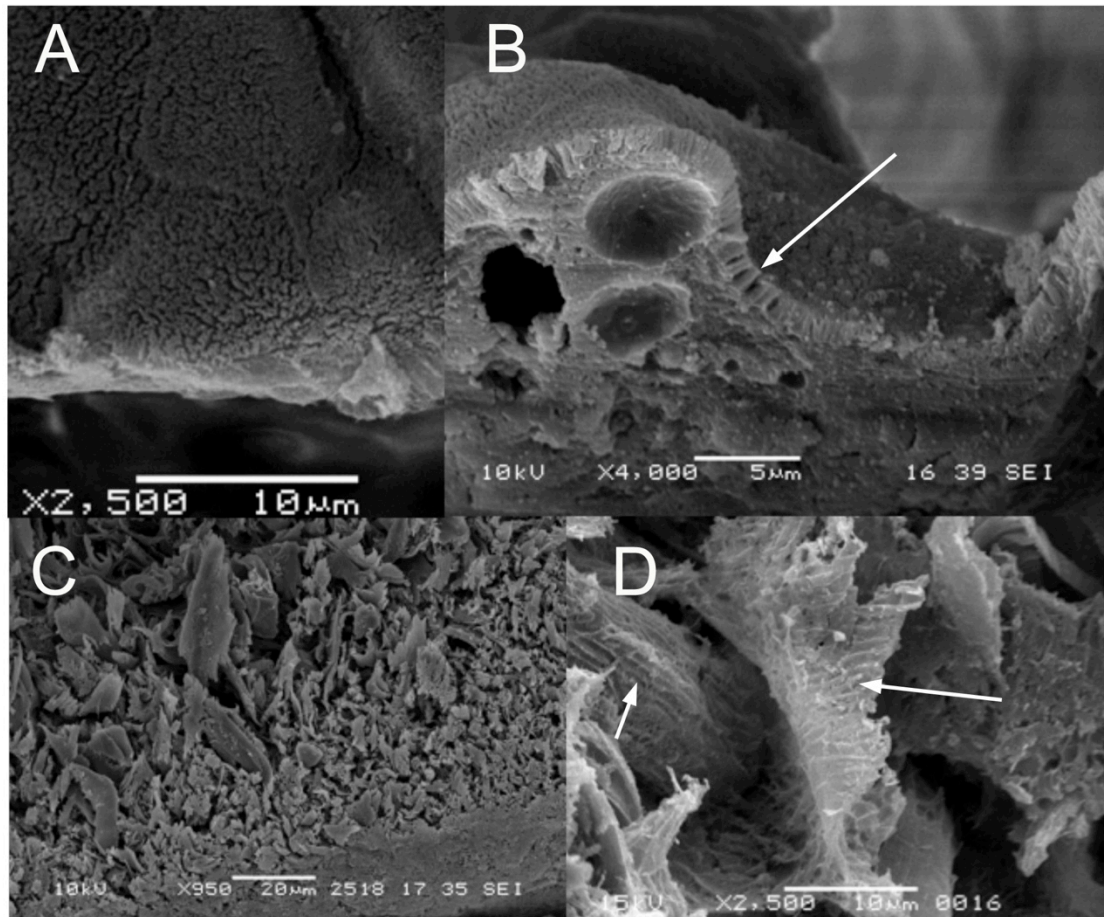
Figure 4.6. SEM analysis of *C. terebrans* digestive system. A) Luminal surface of the hepatopancreas. B) Transverse section of the hepatopancreas showing villi (arrow). C) Food mass found in the digestive tract showing wood flakes. Ventral surface top left corner, dorsal surface bottom right corner. D) Bacterial activity shown on a wood flake from the digestive tract indicated by arrows.

Faecal pellets were found also to contain mainly densely packed wood fragments, however the size gradient of wood fragments were not found to identically follow that of the digestive tract but were found to have the more densely packed finer wood fragments just below the lateral line (Figure 4.7A). A few rod bacteria were found on the surface of the faecal pellets, however none were found in the centre of the faecal pellets (Figure 4.7B).
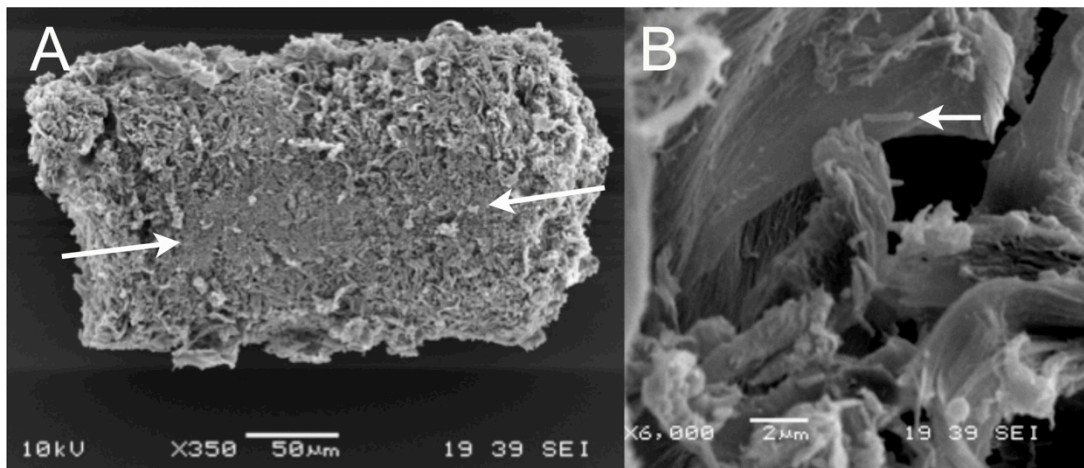
Figure 4.7 SEM study of *C. terebrans* faecal pellets. A) Faecal pellet in the sagittal plane. Smallest wood fragments in the lateral line indicated by arrow B) Rod bacteria found on the surface fragments of a faecal pellet, indicated by arrow.

Placing an object between the two polarising filters gives an indication of its structure. An amorphous structure refracts minimal light while a more uniform structure will refract a greater amount of light.
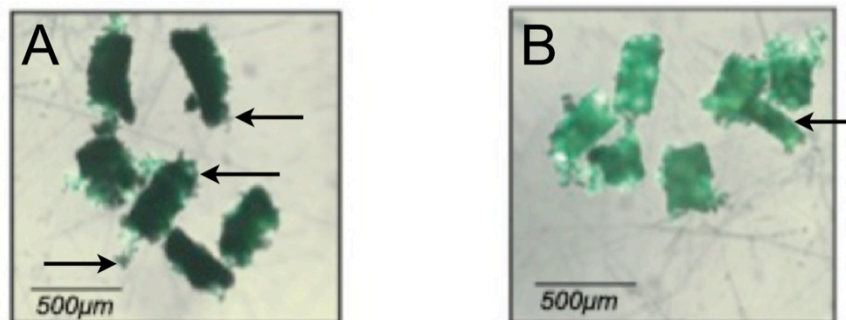
Figure 4.8 Faecal pellets photographed between to polarising filters. A) *C. terebrans* B) *L. quadripunctata* pellets.

Faecal pellets were viewed using a stereo-microscope to see if the crystalline cellulose could be picked up. The faecal pellets of *C. terebrans* appear to not refract a large amount of light (Figure 4.8A). Although there are some areas of birefringence, even the

edges containing relatively little material remain quite dark. *L. quadripunctata* faecal pellets of similar size were used for comparison against *C. terebrans.* However, the faecal pellets of *L. quadripunctata* appear to glow when placed between the two filters (Figure 4.8B). This could be due to birefringence caused by the crystalline structure of the wood in the faecal pellets. Even when the faecal pellet is turned on its side (indicated by the arrow Figure 4.8), so that the light has to travel through more material, the light is still refracted. The difference in light refraction between the faecal pellets of both species suggests that those of *C. terebrans* could be more amorphous, that is degraded, than that of *L. quadripunctata*.

4.3.3 Interactions with Limnoria

The wood chips recovered from the co-habitation experiment were measured to determine wood weight loss. The wood weight loss was found to be higher in all *Chelura* only groups than for that of *L. quadripunctata* containing groups, though not significant in all cases (Figure 4.9). The presence of another individual, either *L. quadripunctata* or *C. terebrans,* has no significant effect on the wood weight loss for a *C. terebrans* individual. A slight decrease in wood loss observed in the mixed species group when faecal pellets were left in the wells during water changes may support the theory that one or the other may be coprophagic*.* Furthermore, the lowest wood weight loss is observed when faecal pellets were not removed from wells containing single *L. quadripunctata* suggesting that *L. quadripunctata* could be consuming their own faecal pellets. However, as stated, there is no significant difference in the wood weight loss if an individual is alone or accompanied. This is not the result we would have expected if either species were truly coprophagic*.* In this case, we would expect to observe a reduction in wood weight loss, as one or both individuals consume less wood in favour of faecal pellets.
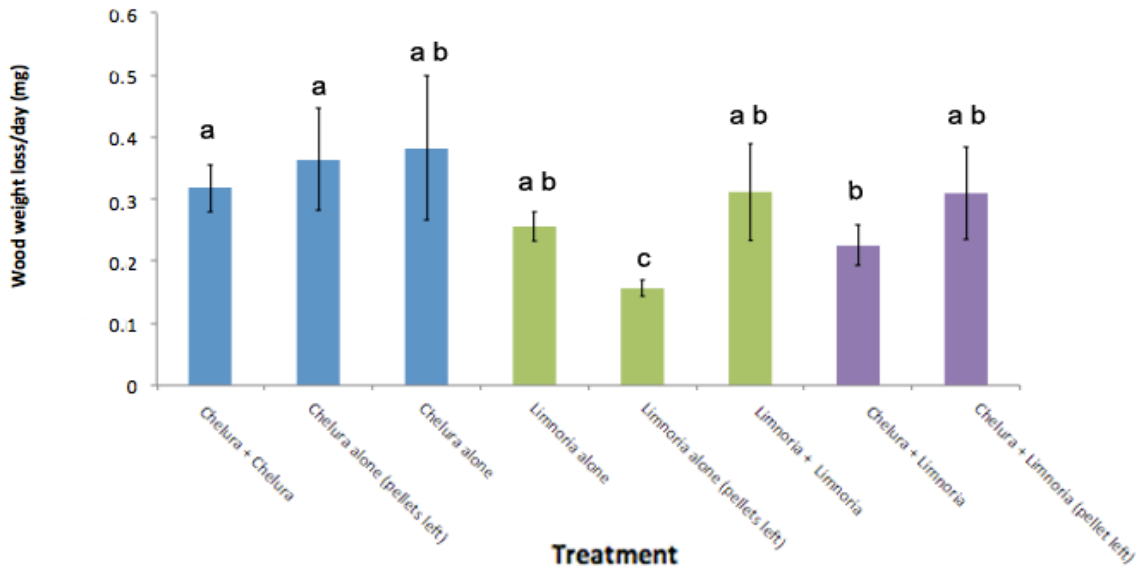
Figure 4.9. Wood weight loss from each co-habitation experimental group. Blue – *Chelura*, Green – *Limnoria*, Purple – Both wood borers together. Letters indicate significant differences using Tukey comparison (Minitab).

Chelura create fine wood debris as they browse the surface, and this debris could account for some of the observed wood loss. The faecal pellets of *C. terebrans* and *L. quadripunctata* are distinguishable in cross-section (Figure 4.1), by counting the number of faecal pellets produced, a better calculation of the amount of wood consumed by an individual animal could be made. These data show that there is a large difference in the amount of pellets produced, from each species; *C. terebrans* produce fewer faecal pellets than *L. quadripunctata*, which is surprising given the wood weight loss is very similar. However, observations regarding the production of wood debris during browsing could account for this apparent discrepancy.

The pellets produced by *C. terebrans* are larger than those of *L. quadripunctata*, however *L. quadripunctata* produce many more than *C. terebrans*. Aside from the overall quantity of pellets, the experimental groups containing *C. terebrans* or *L. quadripunctata* show a similar trend in pellet production rate (Figure 4.10), suggesting that both species react to each other's presence in a similar manner. These results are

also consistent with *C. terebrans* being facultative coprophagic, given the slight dip in *L. quadripunctata* pellet numbers when they co-habit, although this dip could be due to *L. quadripunctata* feeding less in the presence of *C. terebrans*. However, given the similar pattern of pellet production, the same could be said of *L. quadripunctata. Chelura* produce more faecal pellets when they are alone or with their own species than when they are with *L. quadripunctata*, suggesting that *C. terebrans* do not require the aid of *L. quadripunctata* in order to digest wood.
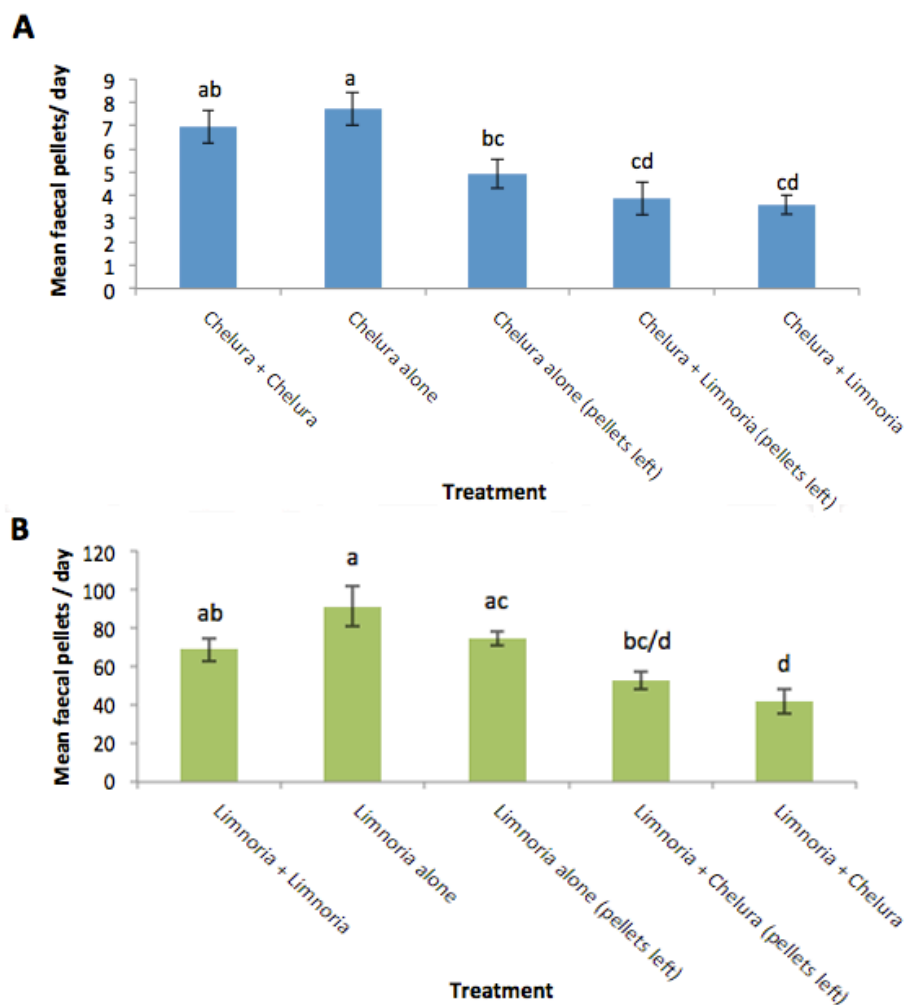
Figure 4.10 Faecal pellet production of *C. terebrans* and *L. quadripunctata* in each co-habitation experimental group. A) Faecal pellets produced by *C. terebrans* in each experimental group B) Faecal pellets produced by *L. quadripunctata* in each co-habitation experimental group**.**

4.3.4 Microbial study

To investigate bacterial levels in *C. terebrans* and make comparisons with other pericarid crustaceans, the ratio of bacterial to host genes in the hepatopancreas and gut was estimated by the amplification of bacterial 16S ribosomal DNA and a host haemocyanin gene. These ratios were then referenced to bacterial levels seen in muscle tissue from *C. terebrans* to control for contamination. Muscle tissue was more difficult to extract than any of the hepatopancreas or gut samples for any species and therefore, assuming the muscle tissue was effectively sterile to begin with, allowed the samples to be referenced against a more stringent control for contamination. The suitability of the using DNA isolated from muscle tissue as a reference sample was shown by the finding that although bacterial sequences were amplified in the muscle sample, the PCR product reached its exponential phase of amplification closest to that of the bacterial no template controls (data not shown).
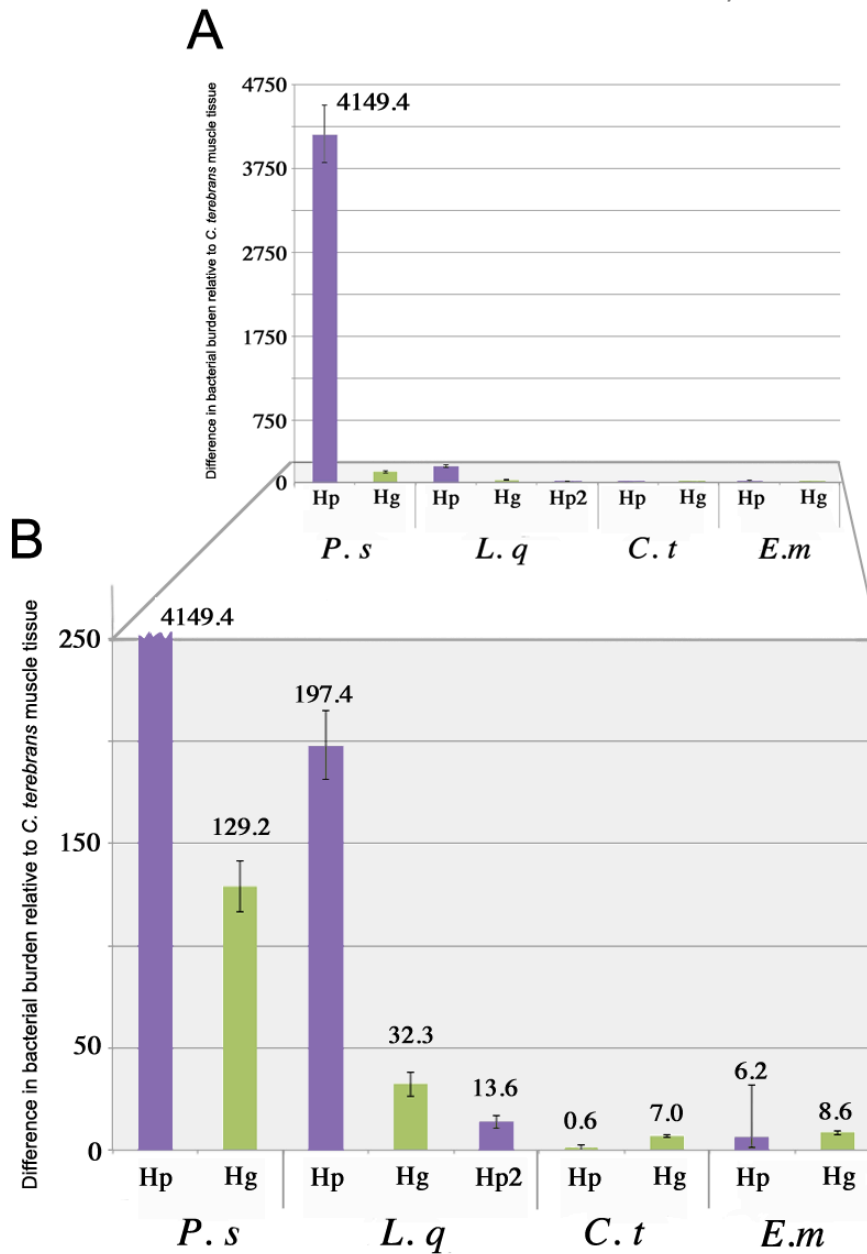
Figure 4.11. Bacterial sequences in the hepatopancreas and gut compared to the host DNA sequences. A) Full graph showing the difference between *P. scaber* ratio and those of the other animals. B) Graph showing a magnified portion from graph A, from 0-250 on the X-axis. Values are displayed at the top of each data bar, purple - hepatopancreas (Hp), green - gut (Hg). Error bars relate to technical repeats.

The *P. scaber* samples show a much greater number of bacterial sequences in comparison to the muscle reference (Figure 4.11A), the hepatopancreas and gut display 4149-fold and 129-fold more bacterial sequences than the muscle respectively. In contrast, *C. terebrans* displays the smallest number of bacterial sequences, with the

hepatopancreas presenting 0.6 times that seen muscle, almost 6916 times less than that in the hepatopancreas of *P. scaber* (Figure 4.11B). The qPCR also shows very few bacterial sequences in *E. marinus*. All species tested, except *P. scaber* and *L.q* Hp (discussed below)*,* display a lower bacterial sequence number in the hepatopancreas than the gut.

4.3.5 Investigation of bacterial species found in the hepatopancreas of *L. quadripunctata*

The first hepatopancreas sample of *L. quadripunctata* (*L.q* Hp, Figure 4.11B) was found to contain a large amount of bacteria in comparison to its gut sample (*L.q* Hg) and all samples from *C. terebrans* and *E. marinus*. Another *L. quadripunctata* hepatopancreas sample was prepared (*L.q* Hp2) and added to the experiment. This showed a much lower number of bacterial sequences than the first sample (*L.q* Hp), slightly less than that of the gut (*L.q* Hg) and only slightly higher than those found in *C. terebrans* and *E. marinus* samples. The PCR product from the *L. quadripunctata* hepatopancreas sample (*L.q* Hp) showed a very sharp peak in the melt curve (data not shown) suggesting the presence of a single dominant amplicon. A 1.3kb 16S ribosomal region was amplified and sequenced from the original sample (*L.q* Hp). A single sequence trace was obtained and comparison with sequences in GenBank (NCBI) using BLAST showed the closest related species to be *Candidatus Hepatincola* porcellionum (92.7% across sequenced region). Primers were then designed to amplify sequences of this new bacterium and were used to screen all samples from the experiment in addition to DNA extracted from the wood on which the limnoriids where fed. However, while the *L.q* Hp sample produced a single PCR product of the expected size, no product was amplified in any other sample (data not shown).

4.4 Discussion

This chapter has investigated the reliance of *C. terebrans* on *L. quadripunctata,* their diet as well as their potential relationship with bacterial endosymbionts. The feed tests in combination with the effective culturing showed that the *C. terebrans* were capable of not only thriving on Scots pine wood chips without faecal pellets of *L. quadripunctata* to supplement their diet but also reproducing at a successful rate to populate the wood. This contradicts some of the reports of Kühne & Becker (1964) regarding their reliance for *L. quadripunctata* for survival beyond several generations. The failure of *C. terebrans* cultures reported in this study could have been the result of any number of alterations in the environment in which they were kept, as *C. terebrans* have been found to be more sensitive to oxygen levels and other environmental factors than *L. quadripunctata* (Kühne & Becker, 1964). Although *C. terebrans* were found not to rely on faecal pellets, the feeding experiment did not rule out a coprophagic nature in either *C. terebrans* or their close associate *L. quadripunctata.*

The SEM examination of the digestive tract and faecal pellets in *C. terebrans* found only densely packed wood fragments of differing sizes, suggesting that the animals are utilising the wood as their primary food source. The fact that some of these animals were sourced from an established aquarium monoculture of *C. terebrans* fed on a variety of wood spp. also supports the hypothesis that *C. terebrans* maintains a proclivity for wood in the absence of *L. quadripunctata.* The amount of wood found in the gut was clearly not a by-product of foraging for bacteria as suggested by Barnard (1955). The polarised imaging of the faecal pellets from both animals suggests that the particles/pellets excreted by *C. terebrans* has lower crystallinity than in faecal pellets of *L. quadripunctata*, the substrates fed to the animals were from the same source removing the possibility that the result is due to *C. terebrans* feeding on softer wood. If the images do indicate a more crystalline wood structure in the *L. quadripunctata* faecal pellets, this would contradict the assumption that the wood fragments

129

remaining in the faecal pellets are easier to digest, as the crystalline structure requires many more enzymes or non-enzymatic processes to digest. The difference in faecal pellet refraction is consistent with the difference in gut passage time between the two species. The very little refraction seen in *C. terebrans* faecal pellets indicated that the wood passing through has undergone greater levels of degradation than that of *L. quadripunctata.* This better degradation of the wood could be linked to the much longer in passage time seen in *C. terebrans* and supports a theory that *C. terebrans* utilise the wood more efficiently (Kühne & Becker, 1964).

Although there was no visible evidence of ingested microbes in the digestive tract of *C. terebrans* examined for this study, their presence on ingested wood has been previously shown (Daniel et al., 1991) and it has been demonstrated that *C. terebrans* benefit when fed wood containing microbial flora (Kühne & Becker, 1964). Indeed, although no bacteria were found, wood fragments in the gut did show evidence of previous bacterial degradation. However, the ingestion of microbes on the surface of the wood is inevitable in field conditions, these microbes are thought to provide xylophagus animals with a richer source of nitrogen than the wood (Daniel et al., 1991). Indeed *Limnoria spp.* have also been shown to feed on wood with considerable bacterial flora (Daniel et al., 1991). The SEM study confirmed Boyle and Mitchell's result in finding no lumen-resident bacteria in the hepatopancreas or gut of *C. terebrans* (Boyle & Mitchell, 1978).

4.4.1.1 *Experimental design of bacterial assay*

As the SEM analysis can only reveal evidence for extracellular bacteria, a molecular approach was used to test for any bacterial DNA sequences (extra- or intracellular) in the hepatopancreas and gut. This study has attempted to compare the overall bacterial levels in the digestive system of various perecarid crustaceans. In theory, a qPCR assay utilising truly universal primers should be both highly sensitive and capable of detecting

the DNA sequences of all bacteria. However, even though a perfect bacterial assay is not achievable in practice, steps were taken to maximise the sensitivity and universality of the assay. The primers employed needed to amplify sequences from a wide range of bacterial groups, in other words, they had to be as 'universal' as possible, for this reason careful consideration was given to the choice of 16S primers (4.2.4.7). However, other viable universal primers were considered. The small subunit (16S) ribosomal RNA gene and the beta subunit of DNA polymerase (*rpoB*) represent plausible PCR amplification targets for measuring overall bacterial levels (Case et al., 2007, Lane, 1991, Vos et al., 2012, Ward et al., 1990). The *rpoB* gene is invaluable for studying the diversity of bacterial communities (Case et al., 2007; Dahllöf et al., 2000; Peixoto et al., 2002; Rantsiou et al., 2004; Santos & Ochman, 2004; Vos et al., 2012) however, it is not conserved enough to be used as a universal bacterial marker due to saturation in third codon positions over long timescales, an issue associated with the use of any protein-coding gene (Case et al., 2007; Silkie & Nelson, 2009; Vos et al., 2012). The 16S ribosomal RNA gene sequence is more conserved and therefore presents a better candidate. In addition, due to multiple 16S ribosomal gene copies in many bacterial genomes, the 16S gene also offers a lower bacterial detection limit than *rpoB* (Case et al., 2007).

Although the 16S gene clearly has advantages, there are also drawbacks that require consideration. While the occurrence of multiple 16S gene copies is advantageous in regards to detection limits, it presents issues when attempting to compare bacterial levels. If a bacterial community present in the digestive system of an organism is dominated by species containing notably above or below average copy numbers of the 16S gene, a false impression of the overall bacterial levels would be obtained. The variation in 16S gene copy number is relatively small between most bacterial groups, between 1-6 for majority of groups and 1-13 in extreme cases (Case et al., 2007), and so such variations should only confound findings if differences in the bacterial levels fall

131

within these ranges. However, it does mean that comparative bacterial levels obtained using a comparison of 16S sequences can only ever be used as guide to overall bacterial levels rather than a definitive measurement.

Quantitative PCR revealed that the digestive system in *C. terebrans* possesses almost 7000-fold less bacterial sequences than that of the symbiont-containing *P. scaber*. Furthermore, this low bacterial burden is comparable to that seen in *E. marinus*, suggesting that low bacterial burdens are not exceptional in amphipods. The extent of variation between bacterial levels in *P. scaber* and those of the other animals (~300-6900 fold difference when using the second hepatopancreas sample as being representative of *L. quadripunctata*) is not likely the result of bias toward bacterial species with high 16S gene copy numbers, as a variation of 1-13 copies is the maximum variation reasonably expected on the basis of comparisons between divergent bacteria (Case et al., 2007). However, removing the comparisons with *P. scaber*, the differences observed between the other perecarid crustaceans could very reasonably be accounted for by 16S gene copy number variation. It is well known that *P. scaber* possesses cellulolytic hepatopancreal endosymbionts that enhance its enzymatic capabilities (Kostanjšek et al., 2010, Zimmer & Topp, 1998a). However, the relatively low bacterial burden observed for the other species examined in this study suggests that they do not exploit such a relationship and is perhaps, indicative of a difference in their digestive mechanism. *C. terebrans* samples in this experiment have shown very low bacterial burdens, indeed the hepatopancreas appears to have a marginally lower level of bacteria than the muscle tissue. Assuming the muscle tissue is effectively sterile, these low levels, especially that of the *C. terebrans* hepatopancreas, could be accounted for by unavoidable contamination resulting from the dissection process. Furthermore, contamination of commercial reagents with bacterial sequences is well documented (Mühl et al., 2010; Böttger, 1990; Goto et al., 2006; Hughes, 1994) and these contaminants, however slight, rule out the possibility of an absolute negative

result. Therefore the bacterial levels observed in all species but *P. scaber* are so low they could be considered contamination. However, the observation of consistent differences in bacterial sequence numbers between the hepatopancreas and gut of each species, and in all replicates, indicates the detection of genuine differences in levels of 16S bacterial sequences. Although, due to variation in 16S gene copy numbers, these differences may not reveal genuine differences in bacterial levels.

Any assay that attempts to use bacterial 16S DNA sequences as a proxy for determining bacterial levels in a community has inevitable limitations. These limitations result from the usual assumptions associated with any qPCR experiment, as well as the level of primer universality and the 16S gene copy number variation in found in bacteria. However, the conservation of both primer binding sites across a wide range of bacterial phyla, in addition to the extreme variation between bacterial sequence levels observed not being explicable by 16S gene copy number differences alone, suggest that while the assay cannot be considered a definitive measurement of bacterial levels, it is still nonetheless telling of a meaningful variation. Despite the limitations, such an experiment is necessary as a result of claims regarding the low bacterial levels, or even sterility, in the digestive systems of lignocellulose digesting perecarids (Boyle & Mitchell, 1978; King et al., 2010; Kobayashi et al., 2012), and the implications of such a claim on the potential utilisation of these organisms for the development of future biofuel technologies (King et al., 2010).

The qPCR assay indicated a high level of bacteria in the first hepatopancreas sample of *L. quadripunctata*, a species with the digestive system apparently devoid of bacteria (King et al., 2010). A single trace was obtained after sequencing of the PCR product and comparison with sequences in GenBank (NCBI) using BLAST revealed the closest related species to be *Candidatus Hepatincola* porcellionum (92.7% across sequenced region), a bacterium found intimately associated with the endothelium of the hepatopancreas in *P. scaber* (Wang et al., 2004).

The bacterium found was unlikely to be the result of contamination for several reasons. Firstly, it was the only non-*P. scaber* sample with such high bacterial quantities, the melt curve of the qPCR product for this sample displayed a sharp peak, indicative of a single dominant sequence. This is in comparison to all other samples, where a broader melt curve containing multiple small peaks was observed. Furthermore, the chromatogram obtained following sequencing of the cleaned PCR product produced a clear sequence trace, again indicating the sample was dominated by a single bacterial species. In addition, subsequent attempts to detect this bacterium in all other samples used in this study, including DNA extracted from the wood on which the limnoriids were fed and the second *L. quadripunctata* hepatopancreas sample, resulted in negative findings. The bacterium found in the *L. quadripunctata* hepatopancreas was found to be a relative of *Candidatus Hepatincola porcellionum*, with a sequence divergence above 3% (7.3% across the 1.3kb 16S region) suggesting that it should be considered a different species (Cohan, 2002). *Candidatus Hepatincola porcellionum* belongs to the Rickettsiales, these are known to be intracellular symbionts or pathogens of many animals (Darby et al., 2007; Wang et al., 2004). However, in this case, the absence of this bacterium in the other *L. quadripunctata* sample is perhaps indicative of an infection. Due to the requirement of a large number of *L. quadripunctata* to attain sufficient amounts of DNA in all samples, we are unable to discern the number of infected individuals contributing to the observed bacterial levels in this sample.

This study supports the idea that *C. terebrans* does indeed utilise wood as their primary food source. The study also suggests they are capable of digesting wood independently of *L. quadripunctata* and, relative to *P. scaber*, without large quantities of bacteria in their digestive system. This raises questions as to the origin of the enzymes utilised for the digestion of such a recalcitrant food source.

5 Lignocellulolytic enzymes in the digestive tract of *Chelura terebrans*

5.1 Introduction

Lignocellulose is synthesized by all higher plants and is composed of complex polymers, consisting of cellulose and hemicelluloses encrusted by lignins. Cellulose is the most abundant constituent of lignocellulose and is exploited by a wide range of invertebrates as a food source *e.g.*, insects (e.g. Martin 1982; Treves & Martin 1994; Wanatabe, 1998), molluscs (e.g. Sakamoto et al., 2007), and crustaceans (e.g. Zimmer & Topp, 1998; King, 2010). However, lignin is a highly recalcitrant polymer, which limits the accessibility of the cellulose for degradation by cellulases. It is thought that organisms which utilise lignocellulose as a food source, either partly or as a whole, must at least partially modify the encrusting phenolic lignin compounds to access the cellulose and hemicellulose (Zhang et al., 2007; King et al., 2010).

Cellulase is the general term for glycosyl hydrolase enzymes that are capable of degrading cellulose, and are classified by their characteristics: these include the cleavage site location, domains and their primary structure. The primary structure is used to group the glycosyl hydrolases (GHs) into family groups, of which cellulases are present in GH families 1, 3, 5, 6, 7, 8, 9, 10, 12, 19, 26, 30, 44, 45, 48, 51, 61, 74, 116, and 124 (www.cazy.org). For a long time it was assumed that only bacteria, protozoa, fungi and plants were able to synthesise cellulases, and that herbivorous metazoans used the cellulases derived from these organisms, an idea reinforced by many investigations of animals with symbiotic systems (e.g. Martin, 1987; Tanimura et al., 2012). However, a gene encoding a β-1,4- endoglucanase was identified in the termite *Reticulitermes speratus* (Watanabe, 1998). Subsequently, endogenous cellulases were found in a number of invertebrates, with the first crustacean identified with an

endogenous cellulase (β-1,4-endoglucanase) being the freshwater crayfish *Cherax quadricarinatus* (Byrne et al., 1999).

Although cellulases are essential for the final stages of wood degradation, lignocellulose is a more difficult substrate to degrade, requiring a more complex suite of enzymes than those required for cellulose alone. In 1973, wood-boring shipworm from the family Teredinidae were found to have symbiotic bacteria in the gills (gland of Deshayes) (Popham and Dickson, 1973). It was later discovered that these bacteria are capable of fixing atmospheric nitrogen (Waterbury et al., 1983) and 'encode genes associated with the destruction of terrestrial woody material' (Yang, 2009). Since then, other bivalves using wood for nutrition, such as the deep sea bivalve *Xylophaga washingtona*, have also been found to house symbiotic bacteria. However, it remains unclear the extent to which these symbionts contribute to the digestion of wood. Endogenous enzymes useful for the digestion of hemicelluloses (β-1,4-glucanases and β-glucosidase) have been found in the amphipod *Hirondellea gigas* (Kobayashi et al., 2012) and the brackish water bivalve *Corbicula japonica* (Sakamoto et al., 2007*).* Although the importance of plant material to the diet of these animals remains unclear, this finding does show it may not be altogether unusual for invertebrates to possess a lignocellulolytic capacity. Until 2010 no report had revealed evidence of an endogenous endoglucanase in an invertebrate, the possession of which was believed to be an important component enabling the complete digestion of native cellulose. However, an analysis of expressed sequence tags (ESTs) obtained from the hepatopancreas of the wood-boring isopod *Limnoria quadripunctata* revealed sequences believed to be endoglucanase-like (King et al., 2010).

Previous analysis suggests that *C. terebrans* utilises wood as its primary food source and maintains a proclivity for wood in the absence of *L. quadripunctata* (Chapter 4)*.* To be able to access the energy from such a complex food source, *C. terebrans* must have access to a full complement of enzymes capable of hydrolysing a wide range of bonds.

The low levels of microbial 16S sequences in the digestive system of *C. terebrans* has made it unlikely that an internal population of cellulose-digesting symbionts is supplying the enzymes (Chapter 4). Therefore, if *C. terebrans* is indeed processing the wood for nutrition, the enzymes used for its digestion could be acquired from fungal or bacterial microbes on the wood that remain active after ingestion by *C. terebrans*, such as that seen in the cerambycid beetle larvae (Martin, 1992), or could be produced endogenously.

In this chapter the lignocellulose-degrading enzymes available to *C. terebrans* will be investigated by testing the capacity of hepatopancreatic and gut extracts to break specific bonds found in the hemicellulose complex. This study also investigates the presence of phenoloxidase activity, suggested to be important in the degradation of lignicellulose (Zimmer, 2002; Kirby, 2007; Neuhauser & Hartenstein, 1976).

5.2 Methods

5.2.1 Buffer, media and stains

5.2.1.1 *Sample preparation*

Reaction Buffer: 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM phenylmethylsulfonyl chloride (PMSF), 1 mM benzamidine hydrochloride

10mL of solution was made as follows:

- 0.5 mL 1M Tris-HCl pH 7.5
- 125 µL PMS
- 0.002 g Benzamidine hydrochloride hydrate
- 300 µL 5 M NaCl
- 9.375 mL dH$_2$O

This solution was then filtered (0.2 µm).

5.2.1.2 *Gel electrophoresis and plate media*

10 % SDS Polyacrylamide Resolving gel:

20 mL of solution was made as follows:

- 6.7 mL 30 % Protogel (National Diagnostics)
- 8.08 mL dH$_2$O
- 5 mL 4 × RB [1.5 M Tris-HCl, 0.4 % SDS, pH 8 (ProtoGel, National Diagnostics)]
- 200 µL APS (10 % w/v)
- 20 µL TEMED

4 % SDS Polyacrylamide Stacking Gel:

10 mL of solution was made as follows:

- 1.3 mL 30 % Protogel (National Diagnostics)
- 2.5 mL 4 × SB (0.5 M TrisHCl, 0.4 % SDS, pH 6 ProtoGel, National Diagnostics)
- 6.1 mL dH$_2$O
- 50 µL APS (10 % w/v)
- 10 µL TEMED

Native gel: 8 % Acrylamide, 1 × TBE (89 mM Tris-Borate pH 8.3, 2 mM Na$_2$EDTA) 0.1% APS

21 mL of solution was made as follows:

- 5.6 mL of 30 % Acrylamide bis-acrylamide
- 2.1 mL of 10 × TBE (National Diagnostics)
- 13.069 mL dH$_2$O
- 210 μL APS (10 % w/v)
- 21 μL TEMED

Running Buffer: 0.089 M Tris base, 0.089 M boric acid (pH 8.3) and 2 mM Na$_2$EDTA

1 L of solution was made as follows:

- 50 mL 1 × TBE (National Diagnostics)

The solution was adjusted to 1 L with dH$_2$O.

Loading Buffer: 50 mM Tris-HCl pH 7.5, 20 % w/v Ficoll

10mL of solution was made as follows:

- 2 g Ficoll
- 0.5 mL 1M Tris-HCl pH 7.5
- 9.5 mL dH$_2$O

CMC Plates: 1.7 % w/v Agar **,** 0.5 % w/v CMC

400 mL of solution was made as follows:

- 8.5 g Agar (Fisher Scientific)
- 2 g CMC (Sigma-Aldrich)

The solution was made up to 400 mL with dH$_2$O and autoclaved at 121 °C for 20 minutes.

5.2.1.3 *Stains*

2 % Congo red Solution

500 mL of solution was made as follows:

- 10 g Congo red (Sigma-Aldrich)
- 500 mL dH$_2$O

Monophenol Oxidase stain (Nellaiappan & Banu, 1991): 10 mM TME, 10 mg PMS, 20 mg NBT, 50 mM Tris-HCl pH 7.5

20mL of solution was made as follows:

- 0.039 g TME (Sigma-Aldrich)
- 0.01 g PMS
- 2 × 10 mg NBT tablets (Sigma-Aldrich)
- 1 mL Tris-HCl pH 7.5
- 18 mL dH$_2$O

Solution is light sensitive and was kept in the dark before and during the staining process.

Diphenol Oxidase stain (Gasparic et al., 1977): 10 mM Catechol, 50 mM Tris-HCl pH 7.5, 0.3 % w/v 3-methyl-2-benzothiazolinone hydrazone hydrochloride - MBTH

20 mL of solution was made as follows:

Solution 1)   - 0.024 g MBTH (3-methyl 2-benzothiazolin-hydrozone hydrochloride hydrate)

  - 8mL DMF (di-methylformamide)

Solution 2)   - 0.176g pyrocatechol

  - 1.6mL 1M Tris-HCl pH7.5

  - 30.4mL dH$_2$0

Solutions 1 and 2 were mixed together kept in the dark before and during the staining process.

Laccase stain: 2 mM syringaldazine, 50 mM Tris-HCl pH 7.5 50 % v/v EtoH

10 mL of solution was made as follows:

- 0.014 g syringaldazine
- 1 mL 1 M Tris-HCl pH 7.5
- 10 mL 100% v/v EtOH
- 9 mL dH$_2$O

5.2.1.4 *In vitro enzyme assay*

5 × Enzyme reaction buffer: 250 mM Tris-HCl pH 7.5, 750 mM NaCl, 5 mM PMSF, 5 mM benzamidine hydrochloride

- 2.2 mL dH$_2$O
- 1 mL 1 M Tris-HCl pH 7.5
- 600 μL 5 M NaCl
- 200 μL PMSF
- 0.003 g benzamidine hydrochloride

5.2.1.5 *4-NP specific bond breakage assays*

5.2.1.5.1   Substrate solutions

The following substrates were used for each assay and were made up to 20 mM solutions:

- Cellobiohydrolase assay - 4-Nitrophenyl β-D-cellobioside (Sigma-Aldrich)
- Galactosidase assay - 4-Nitrophenyl β-D-galactopyranoside (Sigma-Aldrich)
- Mannosidase assay - 4-Nitrophenyl β-D-Mannopyranoside (Sigma-Aldrich)
- Xylanase assay - 4-Nitrophenyl β-D-glucopyranoside (Sigma-Aldrich)
- Glucosidase assay - 4-Nitrophenyl β-D-xylopyranoside (Sigma-Aldrich)

All were made to 20 mM solutions

5.2.1.5.2   Positive control enzyme solutions

The following enzymes were purchased to act as positive controls:

- Cellobiohydrolase assay - β- glucosidase (CAZyme ^{TM}, Lucigen)
- Galactosidase assay - β-Galactosidase (Sigma-Aldrich)
- Mannosidase assay - β-Mannosidase (Sigma-Aldrich)
- Xylanase and Glucosidase assay - Enzyme cocktail (Both Novozymes, Denmark) made up as follows:
    - 4 parts Celluclast 1.5L (CCN03108)
    - 1 part Novozyme 188 (DCN00211)

    The cocktail was purified with a GE Healthcare HiTrap desalting column and

    eluted in 25mM sodium acetate pH 4.5.

The positive control enzyme solutions were made up as described in Table 5.1.

Table 5.1 Positive control solutions for enzymatic assays.

Assay	Enzyme	5× Enzyme reaction buffer	dH$_2$0
Cellobiohydrolase	8.03μl β- glucosidase	20μl	66.97μl
Galactosidase	0.3μl galactosidase	20μl	74.7μl
Mannosidase	10μl mannosidase	20μl	65μl
Xylanase	2.5μl Enzyme cocktail	20μl	72.5μl
Glucosidase	2.5μl Enzyme cocktail	20μl	72.5μl

5.2.1.6 *Controls for CMC plate, MpO, DpO and Laccase assay.*

The following enzymes were purchased to act as positive controls:

- CMC, MpO and DpO assay- Agaricus bisporus (mushroom) tyrosinase (Sigma-Aldrich)

- Laccase assay- *Rhus vernicifera* (Sigma)
    - 83.3 mg
    - 999.6 μL 50mM Tris-HCl pH 7.5

Centrifuged at 12,000 g for 1 minute and the supernatant used.

5.2.2 Sample preparation

Animals obtained from laboratory tanks (see section 3.2.1) were anaesthetised on ice, and dissected in reaction buffer [50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM phenylmethylsulfonyl chloride (PMSF), 1 mM benzamidine hydrochloride]. The guts (Gt) and hepatopancreases (Hp) removed from each animal were separately pooled and placed into microcentrifuge tubes containing 20 µL of reaction buffer. Samples were spun at 6000 × g for 60 seconds at room temperature before the supernatant was removed and the volume of gut luminal fluid and hepatopancreas luminal fluid were recorded respectively. Subsequently, 20 µL of reaction buffer was added to the remaining tissue samples, which were thoroughly crushed using a pestle, and centrifuged at 14,000 × g. The supernatant was removed and volume recorded.

5.2.3 Carboxymethylcellulose (CMC) agar plate assay

The hepatopancreas and gut of fifty specimens of *C. terebrans* were extracted by dissection (dissection buffer: 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM phenylmethylsulfonyl chloride (PMSF), 1 mM benzamidine hydrochloride) and the organs were pooled into two microcentrifuge tubes. These were then centrifuged at 14000 x g for 60 seconds. Aliquots (100 µL) of the extract (denatured, 80 °C for 15 seconds, and non-denatured), positive control [enzyme cocktail of Celluclast 1.5L and Novozyme 188 in a 4:1 ratio (Novozymes, Denmark)], and negative control (buffer only and leg muscle tissue of *Echinogammarus marinus*, a non-wood boring amphipod) were prepared. The samples were dispensed into 8 mm diameter wells bored into 1.7 % w/v agar plates containing 0.5 % w/v sodium carboxymethyl cellulose (sodium-CMC), and incubated overnight at 25 °C. The wells were rinsed with distilled water and the

plates were then flooded with 1 % w/v Congo red and incubated at room temperature for 15 minutes before the excess was poured off and the plates rinsed using 1 M NaCl.

5.2.4 In vitro Azo-dye assay for endo- 1, 4 - β-glucosidase and endo- 1, 4 - β- Xylanase

Endo-1, 4-β-glucanase and endo-1, 4-β-xylanase activities were assayed using azo-dye derivatives of CMC and wheat arabinoxylan respectively (Megazyme, Ireland). An enzyme cocktail was used as the positive control [Celluclast 1.5L and Novozyme 188 in a 4:1 ratio (Novozymes, Denmark)]. The hepatopancreases and hindguts of 100 dissected *C. terebrans* specimens (2 × dissection buffer) were pooled into two tubes as before. Negative controls (as described, as for section 5.2.2, were prepared as per the manufacturer's instructions. Test samples and controls were incubated at 25 °C for 90 minutes. Denatured samples were incubated at 80 °C for the same period of time. Release of soluble azo-dye was measured in triplicate at 590 nm using a Nanodrop 1000 spectrophotometer (ThermoScientific, UK).

5.2.5 In vitro 4-Nitrophenol enzyme assays for β-Cellobiohydrolase, β-Galactosidase, β-Glucosidase, β-Mannosidase and β-Xylanase

Feeding animals were collected from the isolated laboratory culture (see 3.2.1). The hepatopancreas and gut of 250 animals were dissected in reaction buffer and placed into separate microcentrifuge tubes containing 100 μL of reaction buffer. The total amount of sample in each microcentrifuge tube was made up to 400 μL using reaction

buffer and centrifuged at 6000 x g for 5 minutes. The supernatant was removed from both, placed into separate microcentrifuge tubes, labelled Hp and Gt and kept on ice while the next samples were prepared. The negative control was made using an equivalent amount of *Echinogammarus marinus* muscle tissue as the original dissected tissue samples (approximately 2 mg). The muscle was crushed, spun and the supernatant removed as for the tissue samples of *C. terebrans*. All the supernatant samples were made up to 475 µL using the reaction buffer and two aliquots of 47.5 µL were taken from each sample, labelled and kept on ice. Positive controls were made using 5 × running buffer as described in Table 5.1.

The positive controls were split into two lots of 47.5 µL, and five blanks made up using 20 µL 5 × Buffer + 75 µL dH$_2$O all were kept on ice. Each set of enzyme tests consists of: a blank, positive control (× 2), negative control (× 2), Hp Lumen (× 2), Hp Tissue (× 2), Gt Lumen (× 2), Gt Tissue (× 2). One of each replicate from all enzyme test (i.e. positive cont., HpL, HpT, GtL, GtT), except cellobiohydrolase positive control, were placed into 80 °C for denaturing for 90 minutes. All other samples were on ice. One of the replicates for cellobiohydrolase positive control was denatured by incubation at 90 °C of 90 minutes. After denaturation, the samples were centrifuged briefly and 2.5 µL of the corresponding pNP substrate (pNP-β-Cellobioside, pNP-β-Glucopyranoside, pNP-β-Galactopyranosid, pNP-β-Mannopyranoside and pNP-β-Xylopyranoside, all Sigma-Aldrich) was added to each enzyme set. All samples from all enzyme tests, except the positive controls for cellobiohydrolase, were incubated at 25 °C for 90 minutes. The positive controls for cellobiohydrolase were incubated at 70 °C for 90 minutes. After incubation, 50 µL of 1 M Na$_2$CO$_3$ was added to each set of samples, and then the resulting mixture was vortexed for 15 seconds. The samples were centrifuged for 20 seconds at 600 x g and the absorbance measured at 405 nm and compared to a calibration curve made using various concentrations of pNP (1 M, 500 mM, 250 mM, 125 mM 62.5 mM and 32.25 mM).

145

5.2.6 Assay for monophenol and diphenol oxidase

5.2.6.1 *In gel assay*

Native gels (8 % acrylamide bis-acrylamide, 1 × TBE 89 mM Tris-borate pH 8.3, 2 mM Na$_2$-EDTA, 0.1 % APS) were pre-run at 100 V at room temperature for 1 hour with 1 × TBE running buffer. Samples were prepared as in section 5.2.2, before loading buffer was added to each sample at a ratio of 4:1 sample to buffer. These were centrifuged again at 14000 × g for 60 seconds and 12.5 µL of each sample were loaded into the gel. 12.5 µL of 0.01 U/µL mushroom tyrosinase (Sigma-Aldrich) was also loaded onto the gel to act as a positive control. Gels were run in duplicate at 100 V for 75 minutes at room temperature, one stained using Safestain (Invitrogen) and the other stained for either monophenol oxidase activity (10 mM TME, 10 mg PMS, 20 mg NBT, 50 mM Tris-HCl pH 7.5) or diphenol oxidase activity (0.3 % w/v MBTH, 10 mM catechol, 50 mM 1 M Tris-HCl pH 7.5), at room temperature in the dark.

5.2.7 Assay for laccase

Native gels [8 % acrylamide bis-acrylamide, 1 × TBE 89 mM Tris-borate pH 8.3, 2 mM Na$_2$-EDTA) 0.1 % APS] were pre-run at 100 V at room temperature for 1 hour with 1 × TBE running buffer. Samples were prepared as in section 5.2.2 and run in duplicate at 100 V for 75 minutes at room temperature before being stained with either Safestain (Invitrogen) or 50 mM Tris-HCl pH 7.5, 2 mM syringaldazine, 23.75 % v/v MeOH.

5.2.8 Mass spectrometry on bands from mono- and diphenol oxidase in gel assay

Bands from the mono- and di-phenol oxidase in-gel assays (5.2.6.1) displaying phenoloxidase activity were excised. These bands were sent to Mr Will Eborall at the University of York to be analysed using MALDI-TOF/TOF mass spectrometry following a trypsin digest. Methods for this section were performed and provided by Mr Will Eborall, (University of York).

Positive-ion MALDI mass spectra were acquired over a mass range of m/z 800-4000 obtained using a Bruker Ultraflex III in reflection mode, equipped with an Nd:YAG smart beam laser. Final mass spectra were externally calibrated against an adjacent spot containing six peptides (des-Arg1-Bradykinin, 904.681; Angiotensin I, 1296.685; Glu1-Fibrinopeptide B, 1750.677; ACTH (1-17 clip), 2093.086; ACTH (18-39 clip), 2465.198; ACTH (7-38 clip), 3657.929.). Monoisotopic masses were obtained using a SNAP averaging algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) and an signal-to-noise (S/N) threshold of 2.

The ten strongest peaks of interest (with a signal/noise ratio greater than 30) for each spot were selected for MS/MS fragmentation. Fragmentation was performed in LIFT mode without the introduction of a collision gas. The default calibration was used for MS/MS spectra, which were baseline-subtracted and smoothed (Savitsky-Golay, width 0.15 m/z, cycles 4); monoisotopic peak detection used a SNAP averaging algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) with a minimum S/N of 6. Bruker flexAnalysis software (version 3.3) was used to perform the spectral processing and peak list generation for both the MS and MS/MS spectra. Tandem mass spectral data were submitted to database searching using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.2.0), through the Bruker ProteinScape interface (version 2.1). Search criteria included: Enzyme, Trypsin; Fixed modifications, Carbamidomethyl (C); Variable modifications, Oxidation (M); Peptide tolerance, 250

ppm; MS/MS tolerance, 0.5 Da; Instrument, MALDI-TOF-TOF; Database, non-redundant peptide database, GenBank (NCBI).

Methods provided by Will Eborall (University of York)

5.3 Results

5.3.1 Cellulolytic activity

A CMC agar plate assay was performed as a preliminary test for cellulolytic activity. The native *C. terebrans* hepatopancreas sample displays the largest area of clearance, indicative of CMC digestion (right-hand well, Figure 5.1C), with the gut presenting a reduced but still obvious area of clearance (right-hand well, Figure 5.1D). In contrast, no activity was seen in the denatured controls (left-hand wells, Figure 5.1C & D) suggesting the reactions observed in the native samples are the result of enzymatic activity. The same patterns of activity were observed in all five experimental replicates (data not shown). The *E. marinus* muscle sample showed no activity in either the denatured or native sample (left- and right-hand well respectively, Figure 5.1A) with the mushroom tyrosinase positive control presenting a clear activity (right-hand well in Figure 5.1B). The positive control appears to have some thermostability as, although reduced, the denatured control still has a modicum of activity (left-hand well, Figure 5.1B).
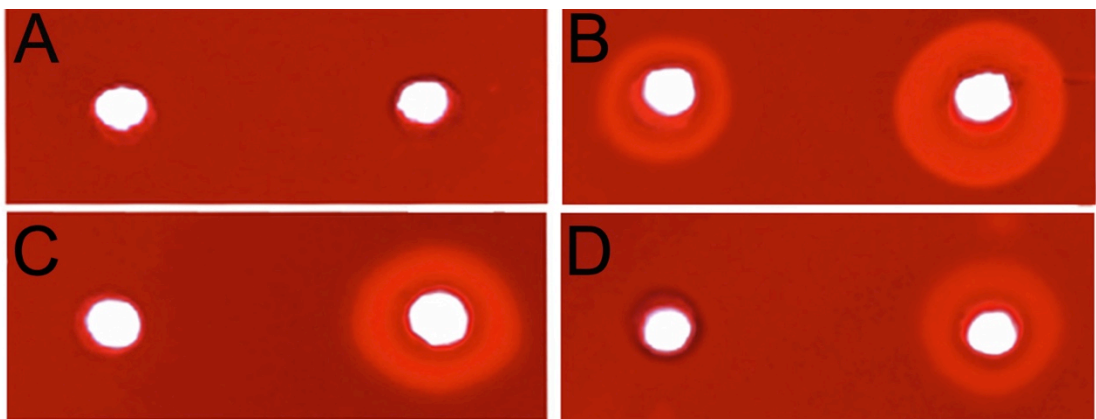
Figure 5.1 Cellulose plate assay. Discolouration represents activity, left hole on each figure shows heat denatured sample and the right hole the native sample. A) Negative control, *E. marinus* muscle showing no activity. B) Positive control, mushroom tyrosinase, showing markedly more activity in the native sample than the heat denatured. C) Extract from the lumen of the hepatopancreas of *C. terebrans* showing activity in the native sample and none

in the denatured. D) Extract from the gut of *C. terebrans* showing less activity in the native sample than the hepatopancreas sample and none in the denatured.

5.3.2 Cellulolytic enzyme assays

5.3.2.1 *Azo-dye derivatives*
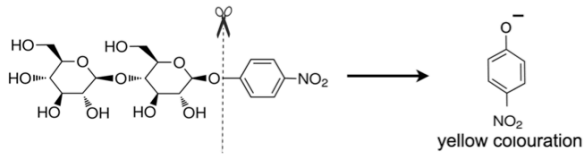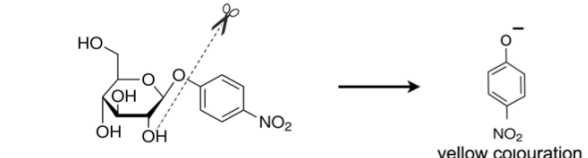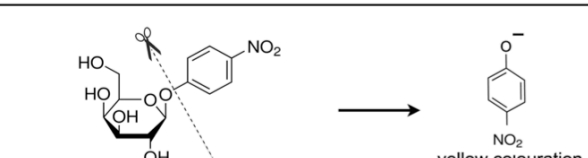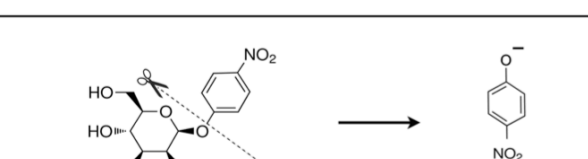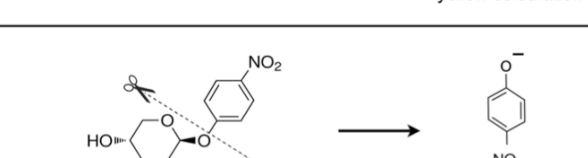
The hepatopancreas and the gut samples gave positive results in Azo-dye assays designed to detect the presence of endo-1, 4-β-glucanase and endo-1, 4-β-xylanase activity. The absorbances were read at 590 nm and presented in Table 5.2. Absorbance was observed in all assays (each performed in triplicate) using native hepatopancreas and gut samples (Table 5.2), while little activity was detected in heat-denatured samples and blank controls. The endo-1, 4-β-glucanase displays greater activity than the endo-1, 4-β-xylanase, while in all assays the hepatopancreas gave higher absorbance readings than the gut (Table 5.2).

Table 5.2 Azo-dye assay for endo-1, 4-β-glucanase and endo-1, 4-β-xylanase activity.

Sample	Average absorbance (y) at 590nm	
	endo-1, 4-β-glucanase	endo-1, 4-β-xylanase
Positive	0.193 ± 0.002	0.192 ± 0.013
Positive denatured	0.032 ± 0.002	0.042 ± 0.005
Hepatopancreas	0.111 ± 0.002	0.120 ± 0.003
Hepatopancreas denatured	0.004 ± 0.001	0.000 ± 0.001
Hindgut	0.128 ± 0.006	0.072 ± 0.018
Hindgut denatured	0.004 ± 0.002	0.003 ± 0.001
Negative	0.002 ± 0.001	0.002 ± 0.001

5.3.2.2 *In vitro 4-Nitrophenol enzyme assays for β-Cellobiohydrolase, β-Galactosidase, β-Glucosidase, β-Mannosidase and β-Xylanase*

Para Nitrophenol (4-Np) cleavage assays were undertaken to detect cellobiohydrolase, glucosidase, mannosidase, galactosidase, and xylosidase activities in the digestive system of *C. terebrans*. The cleaved chromophore, para Nitrophenol (4-Np), from the substrate becomes detectable when the addition Na_2CO_3 causes it to take its ionic form (Figure 5.2).

Figure 5.2 Enzyme cleavage sites in Paranitrophenol (4-NP) assays. Specific enzyme activities cleave the 4-NP molecule and the addition of $NaCO_3$ results in the 4-NP taking its anionic form giving a yellow colouration.

151

The cleavage of pNP was observed in all assays (each performed in triplicate) using native hepatopancreas and gut samples (Table 5.3, Figure 5.3) while no activity was detected in heat-denatured samples and blank controls. In all assays, the hepatopancreas fluid samples cleaved more pNP than fluid from the gut and the greatest level of activity was observed in the glucosidase assay, with 1.28 and 0.63 nmol of pNP produced by 50 animals/hour in the hepatopancreas and gut respectively (Table 5.3, Figure 5.3)

Table 5.3 4-Nitrophenol (4-NP) assays for specific enzymatic bond breakages, showing the amount of 4NP cleaved by 50 animals / hour.

Activity	Useful for the degradation of:	Tissue sample	Average absorbance (y) at 405nm	Activity in amount of cleaved 4NP [x=(y/0.0008)-0.0117] mM
Cellobiohydrolase	Cellulose	Hp	0.2617	327.1
		Hindgut	0.0370	46.2
β-glucosidase	Cellulose	Hp	0.3077	384.6
		Hindgut	0.1507	188.3
β-mannosidase	Hemicellulose	Hp	0.1313	164.2
		Hindgut	0.0613	76.7
β-galactosidase	Cellulose	Hp	0.0547	68.3
		Hindgut	0.0490	61.2
β-xylanase	Hemicellulose	Hp	0.2173	271.7
		Hindgut	0.0580	72.5
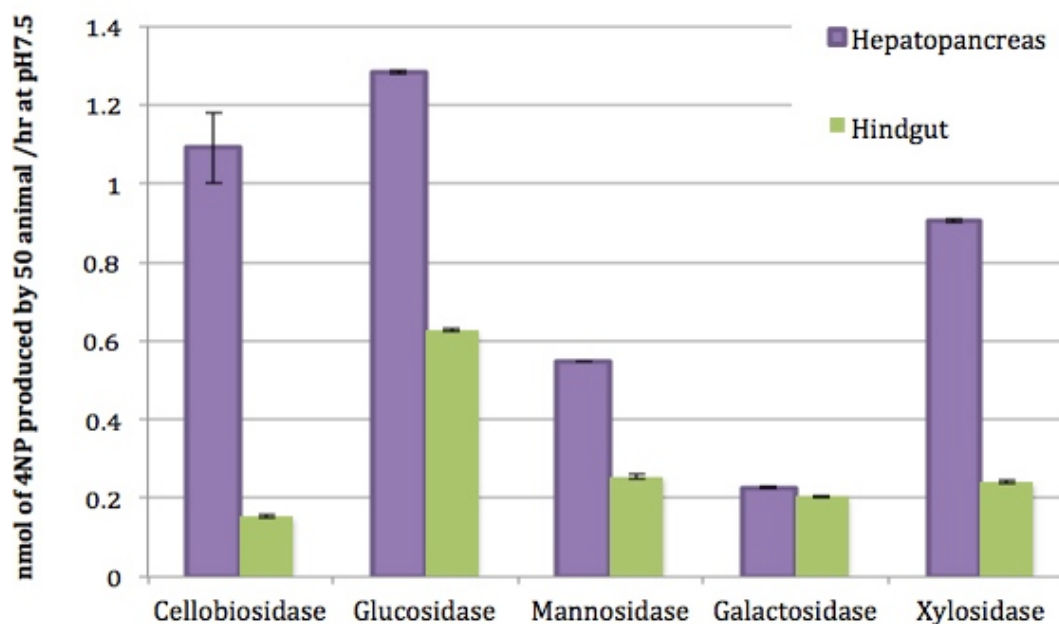
Figure 5.3 4Nitrophenol enzyme activity assay All assays show activity in both the hepatopancreas and the hindgut. The amount of 4-NP cleaved was measured in triplicate at 405 nm using a spectrophotometer (Nanodrop 1000, ThermoScientific UK) and the concentration estimated from a calibration curve [Conc. Of 4NP produced [x=(y/0.0008)-0.0117].

5.3.3 Monophenol and Diphenol oxidase activity

Assays were performed to detect the presence of both monophenol oxidase (MpO) and diphenol oxidase (DpO) activity in *C. terebrans*. All samples show activity in the MpO assay, as evident by the clear and violet colouration throughout the length of the sample lanes (Figure 5.4). The cause of the colouration on the gel is described by (Sidjanski et al., 1997) as a transfer of electrons from a mono- phenolic substrate in the sample to an electron acceptor, in this case the phenazine methosulfate (PMS), and subsequently the reduction of nitro blue tetrazolium (NBT) into a violet-coloured formazan visible in the gel. Indeed, there are areas where the *C. terebrans* samples

appear more enzymatically active than the mushroom tyrosinase sample used as a positive control (indicated with arrows on the right hand side of Figure 5.4.
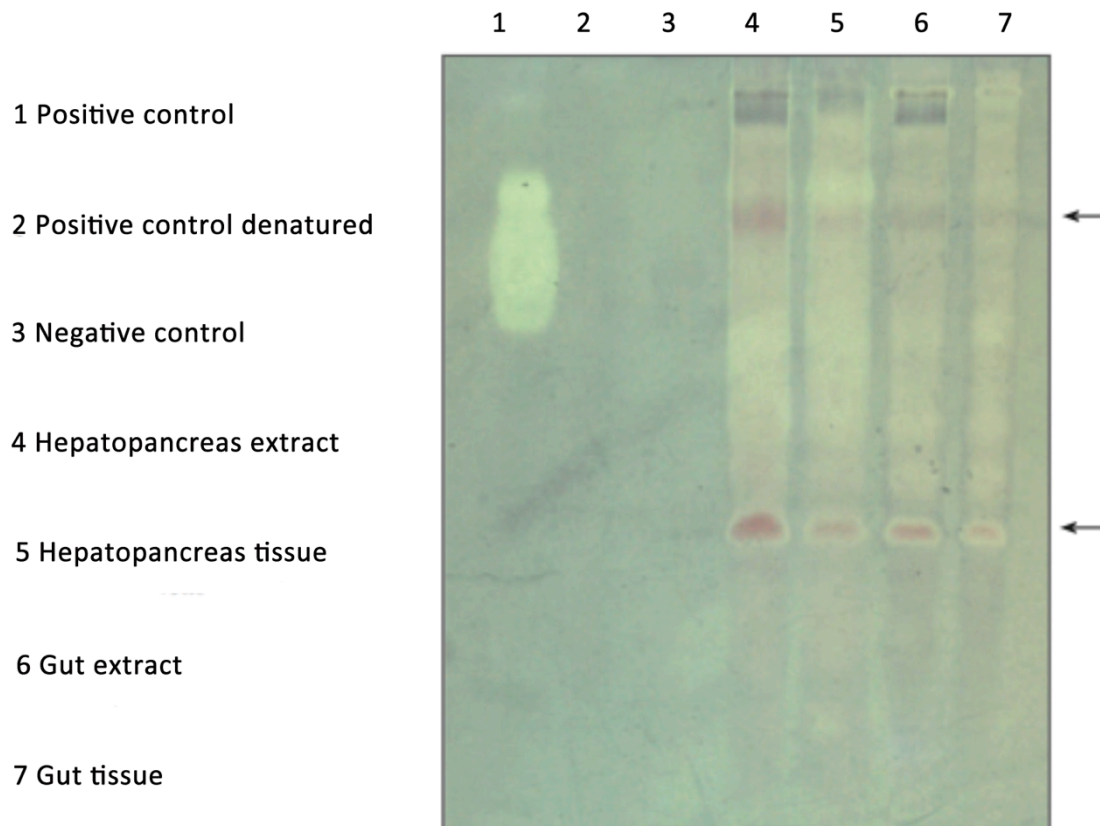
Figure 5.4 Monophenol oxidase Assay (MpO). Positive control in lane 1 shows a clearing of colouration indicating a positive MPO activity. The denatured sample (lane 2) and the negative control (lane 3) showed no activity. Arrows on the right indicate the greatest reaction bands in the *C. terebrans* samples.

The diphenol reaction occurs by the oxidation of catechol to its quinnone product this then condenses with the MBTH to form a red stain. Dark red banding indicates positive DpO activity with a large amount of activity shown by the positive control (lane 1, Figure 5.5) Activity can be seen also bee seen in all the test samples, with the strongest bands indicated by arrows (Figure 5.5). The lumen extracts of both the hepatopancreas and gut samples show the greatest activity. All heat-denatured and negative control samples appeared blank, indicating the reactions present in the native samples are a

result of enzymatic activity. Experimental repeats consistently demonstrated a greater reaction in the hepatopancreas than the gut for both MpO and DpO assays, with tissue samples consistently presenting lower activity than those from the lumen (compare lanes 4 & 5 and 6 & 7 in Figure 5.4 and Figure 5.5).
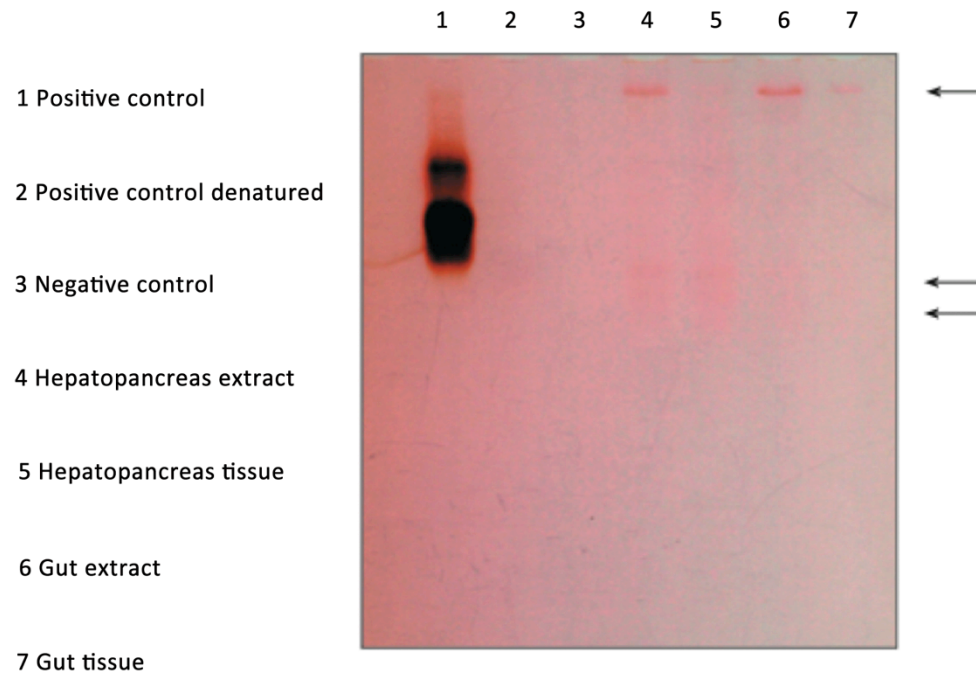
Figure 5.5 Diphenol oxidase Assay (DpO). Positive control in lane 1 shows dark colouration indicating a positive DPO activity. The denatured sample (lane 2) and the negative control (lane 3) showed no activity. Arrows on the right indicate the greatest reaction bands in the *C. terebrans* samples. The activity seen in the extract sample is greater than any tissues sample.

5.3.4 Laccase activity

Presence of laccase activity was probed for, as with MPO and DPO activity, except using a laccase-specific stain. The laccase positive control from *Rhus vernicifera* (Sigma-Aldrich) shows a bright Pink colouration in lane 1 (Figure 5.6), indicating a positive activity. There was no activity was evident in any of the test samples or denatured and negative controls.

1 Positive control

2 Positive control denatured

3 Negative control

4 Hepatopancreas extract

5 Hepatopancreas tissue

6 Gut extract

7 Gut tissue

Figure 5.6 Laccase activity. The only activity evident in the gel was seen in the positive control, indicated by the bright pink area. All repeats of this study showed the same result.

5.3.5 Mass spectrometry on bands from mono and di-phenol oxidase in gel assay

Mass spectrometry analysis (MALDI-TOF TOF) was able to identify proteins in all bands tested exhibiting phenol oxidase activity. Equivalent bands from the hepatopancreas and gut assays were tested and in all cases the same proteins were identified (Figure 5.7). Monophenol oxidase bands presented proteins identifiable as haemocyanin as

well as glycosyl hydrolases (GH). The GH family proteins include; GH2, similar to a β-mannosidase, GH7, a cellobiohydrolase, GH9, a β-1,4-endoglucanase and GH16, similar to a β-1,3-glucan binding protein. The bands showing diphenol oxidase activity were annotated as haemocyanins and GH proteins from the families GH2 and 9. Several bands were tested where there appeared to be no activity and no proteins could be identified from these bands.
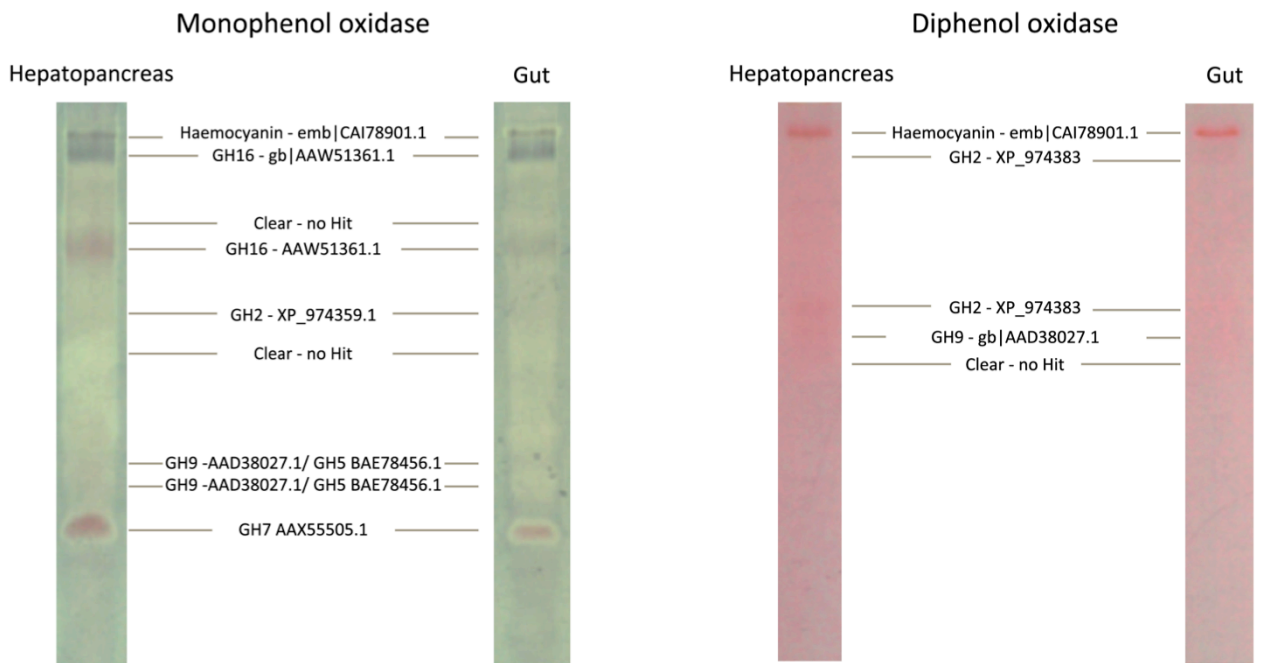
Figure 5.7 Phenoloxidase assay gels illustrating the location of proteins identified from the *C. terebrans* transcript libraries by mass spectrometery (MALDI-TOF-TOF).

157

5.4 Discussion

All enzymatic activities assayed for, with the exception of laccase, were identified in *C. terebrans* extracts, with the lack of activity in any of the denatured samples indicating the observed results are due to enzymatic action. The presence of enzymatic capabilities that could be useful for the digestion of wood, including mannosidase, β-glucosidase and β-xylosidase, endo-1, 4-β-glucanase and endo-1, 4-β-xylanase, with extracts also possessing mono- and diphenol oxidase activity in the digestive tract of *C. terebrans*. These findings further support the hypothesis that *C. terebrans* utilises lignocellulose as a food source.

The cellulases present appear to be both endo-and exocellulases that could work synergistically to digest the amorphous and crystalline cellulose (Watanabe, 2010). Xylan is the most abundant of the hemicelluloses and although all the activities demonstrated would be useful in the degradation of hemicellulose, *C. terebrans* does demonstrate more β-D-xylopyranosidase activity than that observed for both β-mannosidase and β-galactosidase. Cellulase, mannanase, xylanase, and α-glycosidase have been found in an omnivorous amphipod *Hirondellea gigas* (Kobayashi et al., 2012). However, the importance of plant material to the diet of *H. gigas* is open to question and, unlike *C. terebrans,* it does not appear to rely primarily on lignocellulose digestion. In keeping with this, no lignin modifying/degradative activity has been identified in *H. gigas*'. However, *C. terebrans* does display mono- and diphenol oxidase activity consistent with a tyrosinase–like activity, potentially important as a tyrosinase found in the bacterium *Streptomyces spp.* is thought to be involved in lignin degradation (Kirby, 2007). Consistent with this, extracellular tyrosinase from the fungus *Aspergillus* was found to be rapidly induced by the presence of lignin (Gukasyan, 1999). Phenoloxidase activity has been observed in the digestive tracts of some crustaceans, such as *Gammarus pulex and Porcellio scaber* (Zimmer & Bartholme, 2003)*, that feed at least partly on lignocellulose. It has been shown that haemocyanins, besides being

oxygen carriers, can function as phenoloxidases (tyrosinase/catecholoxidase) (Decker 2007), and it has been hypothesised as the mechanism employed by *L. quadripunctata* for lignin modification, as no prophenol oxidases, laccases or peroxidases were identified in its transcriptome sequence (King et al., 2010). The identification in *C. terebrans* of haemocyanins in the gel regions showing high phenoloxidase activity is consistent with this hypothesis. However, the identification of GH family proteins is perhaps more difficult to interpret. It is possible that the appearance of haemocyanins and GH family proteins in the examined gel regions is simply a reflection of their high abundance. Whether the phenoloxidase activity found in *C. terebrans* results from the activity of haemocyanins, as indicated by the mass spectrometry, needs to be further investigated along with any possible effect it may have on lignin.

Zimmer and Bartholmé (2003) suggest that enzymes must be present in both gut and hepatopancreas to be considered endogenous or functionally endogenous [if microbes are present in the hepatopancreas (Zimmer & Topp, 1998)], as activities present in the gut but not in the hepatopancreas would indicate that the enzymes are acquired through feeding. In this study, all enzymatic activities in *C. terebrans* were not only shown to be present in both the gut and hepatopancreas but also found to be more intense in the latter. Furthermore, the microbial experiment (see section 4.3.4) consistently showed fewer bacterial 16S ribosomal sequences in the hepatopancreas than the gut. This, in combination with the higher levels of activity in the hepatopancreas, strongly suggest that the enzymes are truly endogenous to *C. terebrans*. These results appear to be in contrast to many other lignocellulose-degrading animals, which rely upon microbial associations to at least partially aid in their digestion of lignocellulosic materials (Distel, 2003; Watanabe & Tokuda, 2010; Yang et al., 2009).

The investigation into the enzymatic activities available to *C. terebrans* has indicated that the enzyme activities found are of endogenous origin. To confirm their truly

endogenous origin and identify the genes involved, the transcriptome of *C. terebrans* must be inspected for ESTs corresponding to lignocellulolytic enzymes.

6 The hepatopancreatic transcriptome of *Chelura terebrans*

6.1 Introduction

The work presented in this study so far indicates that *Chelura terebrans* consumes wood and is likely to be truly xylophagous (Chapters 3 & 4). *C. terebrans* does not appear to possess high levels of bacterial 16S ribosomal sequences in its digestive system, at least compared to a wood eating isopod with well-established relationships with cellulolytic symbiotic bacteria (Chapter 4). This suggests that *C. terebrans* does not utilize gut-resident bacteria and that it probably obtains nutrition from the wood via the production of endogenous enzymes. Indeed, *in vitro* enzymatic assays using extracts isolated from the hepatopancreas suggest that *C. terebrans* possess a repertoire of endogenous enzymatic capabilities useful for the digestion of wood, including mannosidase, β-glucosidase and β-xylosidase, endo-1, 4-β-glucanase and endo-1, 4-β-xylanase activities. In-gel assays revealed that extracts also possess mono- and di-phenoloxidase activity for potential lignin modifying capabilities (Chapter 5, Figure 5.4 & Figure 5.5). Furthermore, mass spectrometry analysis on excised gel fragments testing positive for mono- and di-phenoloxidase activity implicated several proteins belonging to the glycosyl hydrolase (GH) families and haemocyanins (Fig 5.6). The lack of gut-resident bacterial symbionts and high levels of enzymatic activity in the hepatopancreatic extracts combined with its apparent endogenous origin, increases the potential value of transcript sequences present in the *C. terebrans* hepatopancreas. It is possible that the suite of enzymes present in *C. terebrans* is sufficient for the effective modification of lignin and degradation of hemicellulose. Therefore an analysis of the mRNA transcripts expressed by the hepatopancreas of *C. terebrans* would greatly advance our understanding of this process.

The 454 sequencing platform (454-life sciences, www.454.com) uses a pyrosequencing method (Ronaghi et al., 1998) in combination with 'massive parallelisation' via emulsion PCR (Margulies et al., 2005). A schematic of the methodology taken from Ellegren (2008) is illustrated in (Figure 6.1) . Applying this technology to *C. terebrans* offers the possibility of characterising its hepatopancreatic transcriptome, obtaining both sequence data and relative abundances of transcripts present. Such analysis has been applied to increase our genomic understanding of non-model organisms (Ellegren, 2008), and has recently been utilised to characterise the hepatopancreatic transcriptome of the wood boring isopod *Limnoria quadripunctata* (King et al, 2010). An analysis that suggested enzymes from the GH7 family are important for efficient wood degradation.

This chapter will examine the *C. terebrans* hepatopancreatic transcriptome using the 454 GS FLX Titanium sequencing platform. In doing so it will help confirm the truly endogenous origin of enzymatic capabilities described in Chapter 5 and, by comparison with the transcriptome of the non-boring amphipod *Echinogammarus marinus* (personal communication) and the library already published for *L. quadripunctata* (King et al., 2010), identify the enzymes involved in lignocellulosic degradation. Moreover, it will compare the transcriptome of *C. terebrans* fed on hardwood (Beech - *Fagus sylvatica)* and soft wood (Scots pine sapwood - *Pinus sylvestris)* to ascertain whether the trancriptomic output of *C. terebrans* is tailored to particular diets.
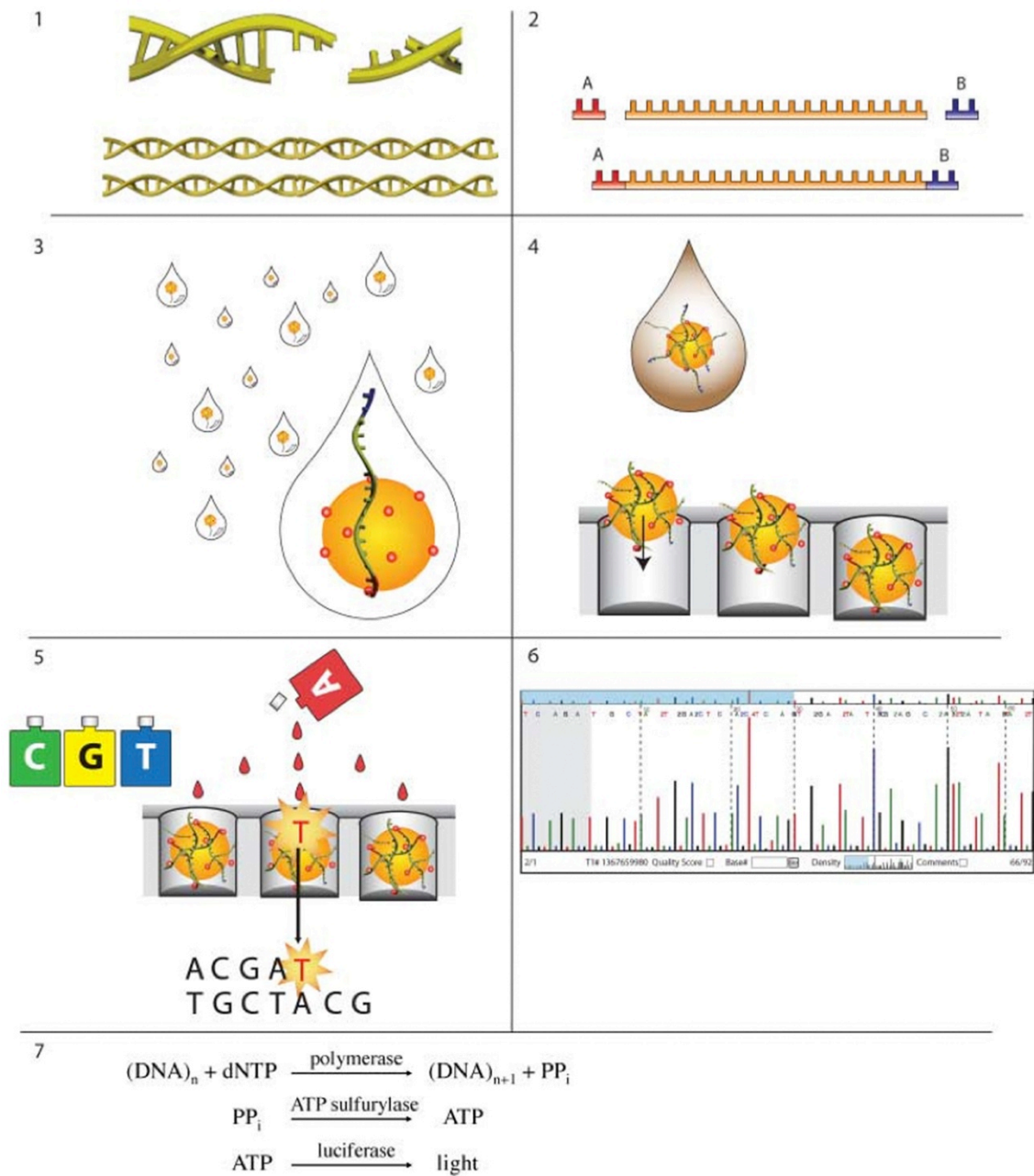
Figure 6.1 Schematic illustration of 454 sequencing from Ellegren, 2008. A schematic illustration of the different steps in 454-sequencing. (1) Short amplicons, fragmented DNA or cDNA is used as starting template for further processes. (2) 5'- and 3'-end specific adapters are ligated to single-strand fragments, creating a library for PCR amplification. (3) Fragments are immobilized through a biotin tag on one of the adaptors that binds to streptavidin-coated beads. Each bead will come to carry just one fragment. Beads are then emulsified in a water-in-oil mixture. (4) Each drop of oil contains the necessary ingredients for PCR and thereby forms a microreactor

for amplification. Massively parallel amplification is carried out in the emulsion. Beads, with amplified fragments bound to them, are released from oil and are loaded onto a fibre optic chip, a picotiter plate, for sequencing. Only one bead will fit in each ≈ 44 µm well. (5) Pyrosequencing takes place by a sequential flow of sequencing reagents across the plate. When a complementary nucleotide is added to a particular template in an extension reaction, a light signal is generated. (6) The final result is a pyrogram in which the height of each signal is proportional to the number of adjacent nucleotides that are identical. In each cycle of the sequential addition of the four different nucleotides, no signal is seen when noncomplementary nucleotides are added. (7) Description of the chemical reactions leading to a light signal. Illustration: Ola Lundström. The pyrogram was kindly provided by Anders Götherström.

6.2   Methods

6.2.1 Animal collection and feeding

Animals above 5 mm in length (from rostrum to telson) were collected from the mixed laboratory culture (section 3.2.1) and placed in individual wells of 12-well plates containing seawater. These were fed on either chips of Scots pine sapwood (*Pinus sylvestris)* or beech (*Fagus sylvatica)*; the seawater was changed four times a week for two weeks. Animals that had been feeding, identified from regular faecal pellet production and a full gut, were used for RNA extraction.

6.2.2 RNA extraction

Twenty animals were dissected from each food group (Scots pine or beech) and their hepatopancreases were removed and placed in 800 μL of Trizol (Invitrogen). Samples were then crushed using pestles, vortexed for 10 seconds and left to incubate at room temperature for 5 minutes to allow nucleoprotein complexes to dissociate. To aid precipitation, 200 μg glycogen was added to each sample to act as a carrier and samples were then centrifuged at 14,000 g at room temperature for 5 minutes. The top phase was transferred to a fresh microcentrifuge tube and 400 μL of isopropanol was added before being mixed by inverting the tube. The samples were incubated at room temperature for 15 minutes before centrifugation at 14,000 g at 4°C for 10 minutes. The supernatant from around the resulting pellet was replaced by 1 mL of cold 70 % ethanol, the samples were then centrifuged at 14,000 g at 4°C for 5 minutes. The supernatant was removed and the microcentrifuge tubes were inverted and left to dry for 10 minutes before the pellet was resuspended by the addition of 102μl of distilled RNAse free water. The samples were sent to Dr Marcelo Kern at the University of York where the integrity and concentration of the RNA was established using a Bioanalyser (Agilent 2100) (Figure 6.2).
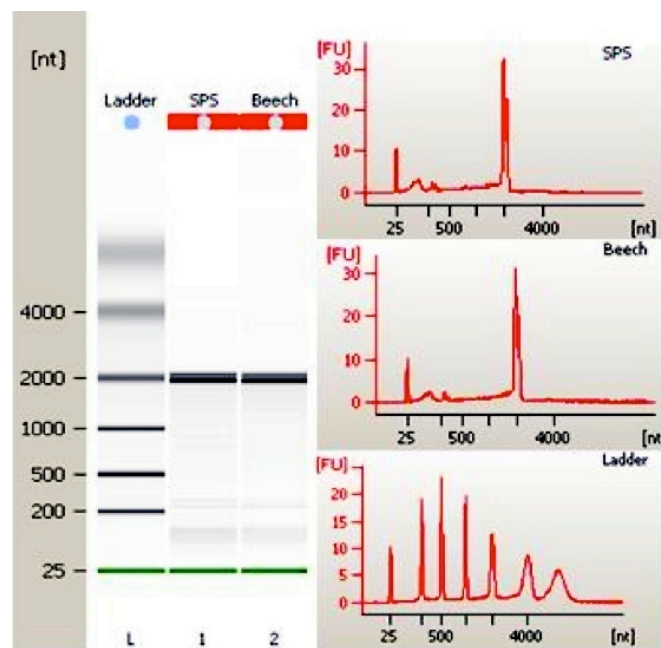
Figure 6.2. Analysis of hepatopancreatic RNA extractions from *C. terebrans* fed on SPS and Beech. Screen shot of Bioanalyser (Agilent 2100) results demonstrates the RNA has discrete bands, sharp peaks, flat baselines and no shoulders. This indicates intact RNA presenting the expected profile of an arthropod.

6.2.3 RNA clean up and cDNA synthesis

Methods in section 6.2.3 are courtesy of Dr. Marcelo Kern, University of York.

6.2.3.1 *RNA clean up and concentration*

The RNA was cleaned and concentrated using a commercially available kit (RNeasy Clean-Up and concentration kit, Qiagen), which removes gDNA, polysaccharides and impurities. RNA was eluted in a final volume of 12 μL of RNAse-free water.

6.2.3.2 *First strand cDNA synthesis*

First strand cDNA was synthesized using RNA isolated from *C. terebrans* fed on SPS and Beech using a commercially available kit (SMART cDNA library construction kit, Clonetech). For each library, five first strand reactions were set up containing 1 μL of the SMartIV oligonucleotide (Switching Mechanism At 5' end of RNA Transcript oligonucleotide 5'-AAG CAG TGG TAT CAA CGC AGT ACT CCR ACr GrG rG- 3'), 1 μL CDS/III 3' (CDS/III 3' PCR 5 ' ATT CTA GAT CCR ACA TGT TTT TTT TTT TTT TTT TTT TVN - 3') and 200 ng RNA. Each reaction was incubated at 72°C for seven minutes then chilled on ice for a further two minutes before the addition of 2 μL 5x First strand buffer, 1 μL 20 mM DTT, 1 μL 10 mM dNTP and 1 μL Superscript IV reverse transcriptase to each 1st strand reaction and incubation for two hours at 42°C. (rG = guanosine ribonucleotide; R = A or G; V = A or C or G; N = A, or C or G or T)

6.2.3.3 *Second strand cDNA synthesis*

The first strand cDNA samples were then combined with reagents in Table 6.1 to create eight 100 μL reactions for each library and PCR amplified (95°C for 1 min, followed by 21 cycles of 95°C for 15 s, 68°C for 6 min).

Table 6.1 The reagents and associated volume used for the second strand cDNA synthesis. * 5' PCR primer sequence 5-AAG CAG TGG TAT CAA CGC AGT AC-3'

Regents	Volume in 1x buffer
H2O	77 μL
10 x advantage PCR	10 μL
10 mM dNTP	2 μL
10 μM 5'PCR*	2 μL
10 μM CDS III 3'	2 μL
1st Strand cDNA	5 μL
50 x Adv 2 Polymerase	2 μL
Total volume	100 μL

The eight 100 μL reactions for each library were then pooled and applied to a 30 kDa filter column (Amicon ultra 0. 5 mL MWCO 30 K, UFC503024) and centrifuged at 9000 g at room temperature for 10 minutes to concentrate both double stranded cDNA libraries into a volume of 80 μL.

Following concentration, a MmeI restriction enzyme digest was performed to remove primer sequences from the 5' and 3' ends of the cDNA. Next, 76 μL double stranded cDNA, 10 μL NEB buffer 4, 5 μL MmeI restriction enzyme (NEB) and 9 μL S-adenosylmethionine (SAM) were combined to produce a 100 μL restriction digest reaction for each library. The reactions were incubated for 4 hours at 37°C before the products were cleaned using a PCR purification kit (QIAquick, Qiagen) following the manufacturer's protocol. The Beech library yielded 244 ng/μL double stranded cDNA and the SPS library yielded 290 ng/μL.

6.3 Libraries

The amplified cDNA was sequenced at the University of York using a 454 GS-FLX titanium pyrosequencing platform.

6.3.1 Contig assembly and annotation of expressed sequence tag libraries

The Expressed Sequenced Tags (ESTs) libraries were assembled into contiguous sequences (contigs) and annotated by Dr Yi Li at the Centre for Novel Agricultural Products (CNAP) at the University of York.

Following sequencing, the initial sets of ESTs were subjected to a quality control procedure. This involves excising sequence stretches that have significant matches to the primers, and choosing the longest possible stretch (minimum 40 base pairs) containing no more than 3% unknown (N) residues. ESTs were assembled into unigene contigs using the CAP3 DNA Sequence Assembly Program (Huang & Madan, 1999). These contigs were then annotated using the WU-BLAST2.0 program (Altschul et al., 1997) and the BLASTx algorithm was utilised to search against the non-redundant peptide database available from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov).

6.4 Contig Library analysis

Gene ontology identities were determined for annotatable genes in the libraries using Blast2GO (www.Blast2GO.com). The GO ontology is structured as a directed acyclic graph, and each term has defined relationships to one or more other terms in the same domain, and sometimes to other domains. Additional bioinformatic analyses were performed with programs available on the NCBI (www.ncbi.nlm.nih.gov), EMBL-EBI (http://www.ebi.ac.uk/) and ExPASy (http://expasy.org) servers.

6.5 Phylogenetic Trees

The translated sequences were aligned using MUSCLE (v3.7, www.phylogeny.fr) and a phylogenetic tree was constructed using the maximum likelihood method implemented by the PhyML (v3.0, www.phylogeny.fr) program, with the reliability of the branching being assessed using the bootstrap method (n = 100) (Dereeper et al., 2008).

6.6 Structural predictions of *C. terebrans* GH7 proteins

The structural modeling was performed by Dr John McGeehan at the University of Portsmouth. The atomic coordinates of the *Limnoria quadripunctata* GH7B (PDB ID: 4GWA) together with the *C. terebrans* GH7B amino acid sequence were used as a template for generating the structural model with the SWISS-MODEL (http://swissmodel.expasy.org)

6.7 *Echinogammarus marinus* sequences

Comparisons using sequences from a transcriptomic library derived from the hepatopancreas of *Echinogammarus marinus* were provided by Drs Gongda Yang and Stephen Short through personal communication [for information on the library construction see Yang (2013)].

6.8 Results

6.8.1 Basic analysis of C. terebrans EST libraries

The 454-sequencing produced a combined total of 123493 ESTs from the hepatopancreas of *C. terebrans*. The EST libraries constructed using RNA isolated from animals fed on SPS produced a library containing 66681 ESTs; the library constructed from animals fed on beech produced 56812 ESTs. The properties of the EST libraries are detailed in Table 6.2.

Table 6.2 Properties of EST and Contig libraries produced using the 454 sequencing and assembled using the CAP3 DNA assembly program.

Property of library	SPS Library	Beech Library
Number of nucleotides	22382928	18311176
Total number of ESTs	66681	56812
Average length of ESTs (bp)	247	237
No. of contigs produced	2612	2401
Average length of contigs (bp)	418	401
No. of singlets	9976	6699
No. of ESTs contributing to 35 most expressed contigs	28607	32187

Both libraries show a similar percentage of EST singlets (ESTs not incorporated in a contig) and broadly similar percentage of ESTs annotating to sequences in the non-redundant peptide database (E value $\leq 1e^{-5}$). However, the percentage of ESTs contributing to the most abundantly represented 35 contigs is larger in the beech-fed library than the SPS (Figure 6.3).

SPS Library

Beech Library
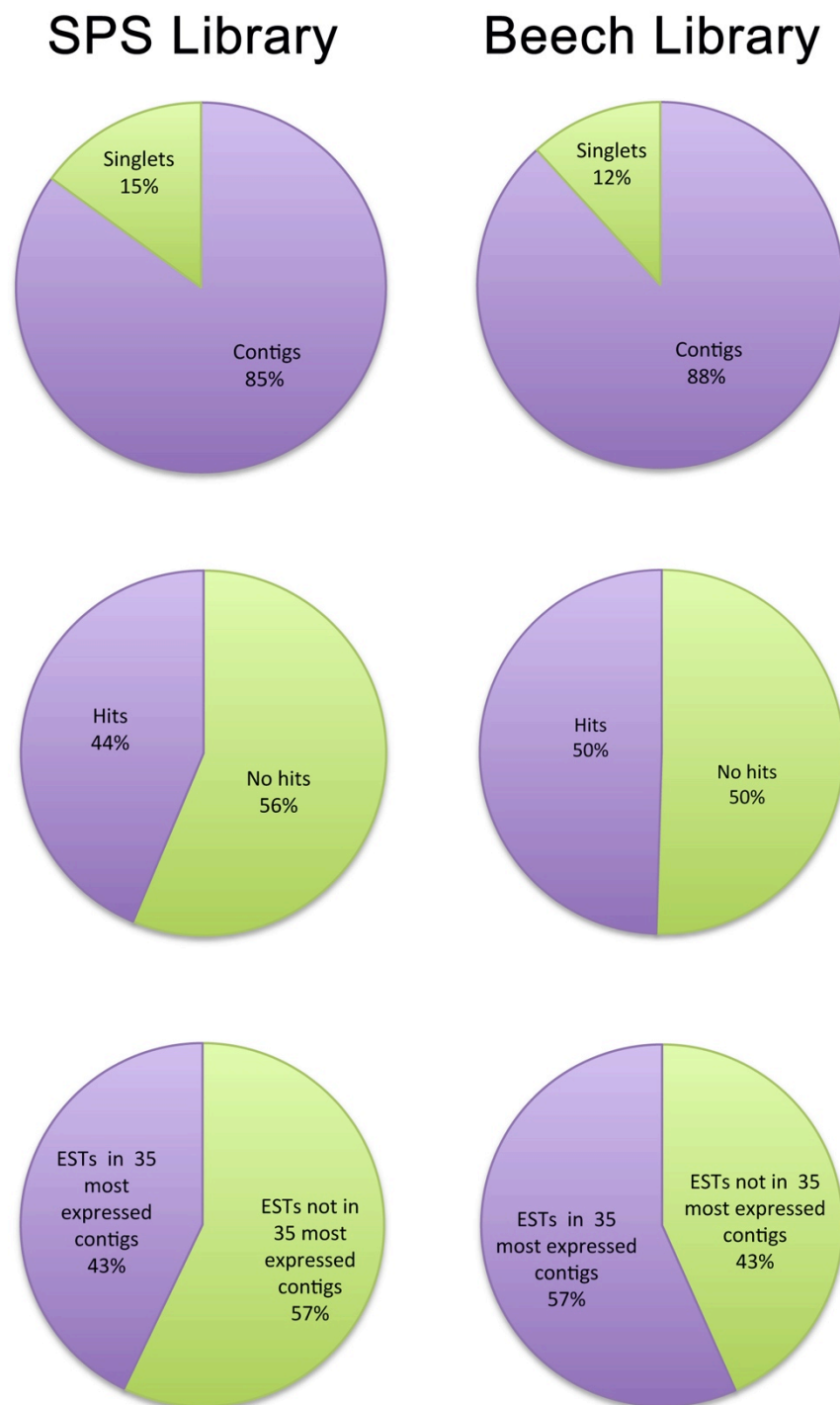
Figure 6.3 Analysis of EST and Contig libraries. A) Percentage of ESTs singlets and ESTs mapping to contigs. B) Percentage of ESTs annotating to sequences in the non-redundant peptide database (NCBI) using the WU-BLAST2.0 program (Altschul et al., 1997) and the BLASTx algorithm E value ≤ 1e^{-5}. C) Percentage of ESTs contributing to the most highly expressed 35 contigs.

6.8.2 Functional annotation of ESTs and Contig libraries

Due to the high number of ESTs contributing to the 35 most abundantly represented contigs (Figure 6.3), these will be the focus of the following section. This does not make the assumption that weakly expressed genes do not play functionally important roles; rather, it allows a more manageable data set for initial analysis.

Of the 35 most abundantly expressed contigs, 23 could be annotated in the SPS library and 22 in the beech (E value $\leq 1^{-5}$). The annotated functions of these contigs are presented in Figure 6.4. Both libraries show many ESTs representing haemocyanins and GH proteins as well as Cadherin D.
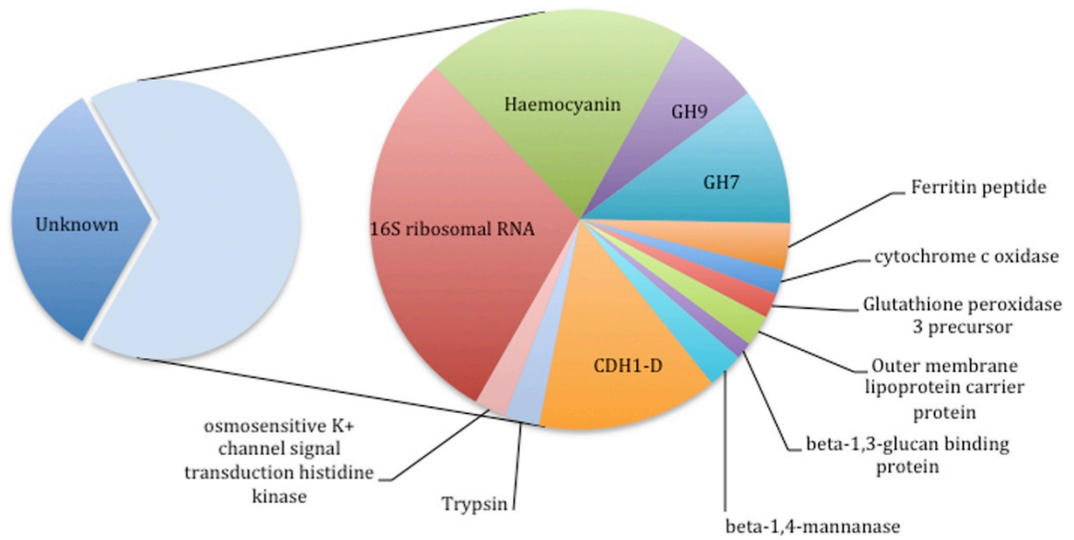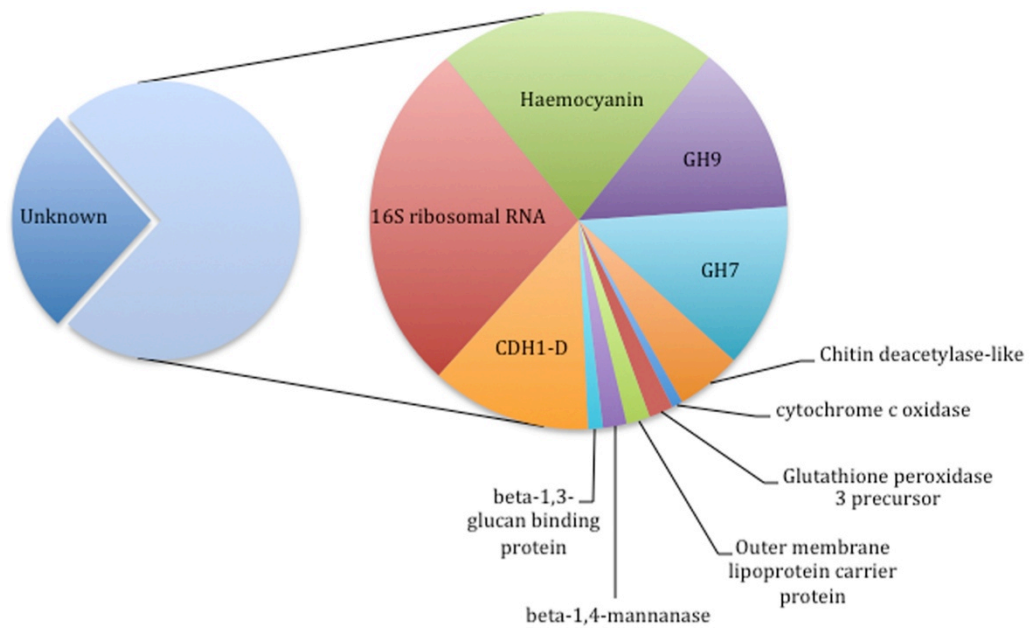
Figure 6.4 Annotations of the 35 most abundantly expressed contigs in SPS and beech libraries. A) Functions of the 23 annotated contigs in SPS. B) Functions of the 22 annotated contigs in beech. Pie charts represent the proportions of ESTs contributing to each annotation.

An analysis of the Gene Ontology (GO) terms associated with the annotatable genes in the 35 most abundantly represented contigs was performed using Blast2GO software (www.blast2go.com). Gene Ontology terms have defined relationships to one or more other terms in the same level, and sometimes other levels. For this study, levels 2 and 6 were found to be the most informative and are presented in Figure 6.5 & Figure 6.6, however all GO terms and their parent terms are shown in the appendix. Level 2 GO terms give a more generalised functional categorisation, whereas level 6 terms show more specialised functions. The biological functions associated with the most abundantly represented contigs in each library are presented in Figure 6.5 and Figure 6.6. A large proportion of the functions assigned are related to biological processes concerning metabolism and localisation (the establishment and maintenance of substrate location). More specifically, some of these metabolic processes are involved in oxygen transport and primary metabolic processes. Both libraries have ESTs mapping to similar functions, including oxygen transport and hemicellulose metabolic processes. The beech library displays less functional categories at the 6th level, indicating less diversity in the 35 most represented contigs.
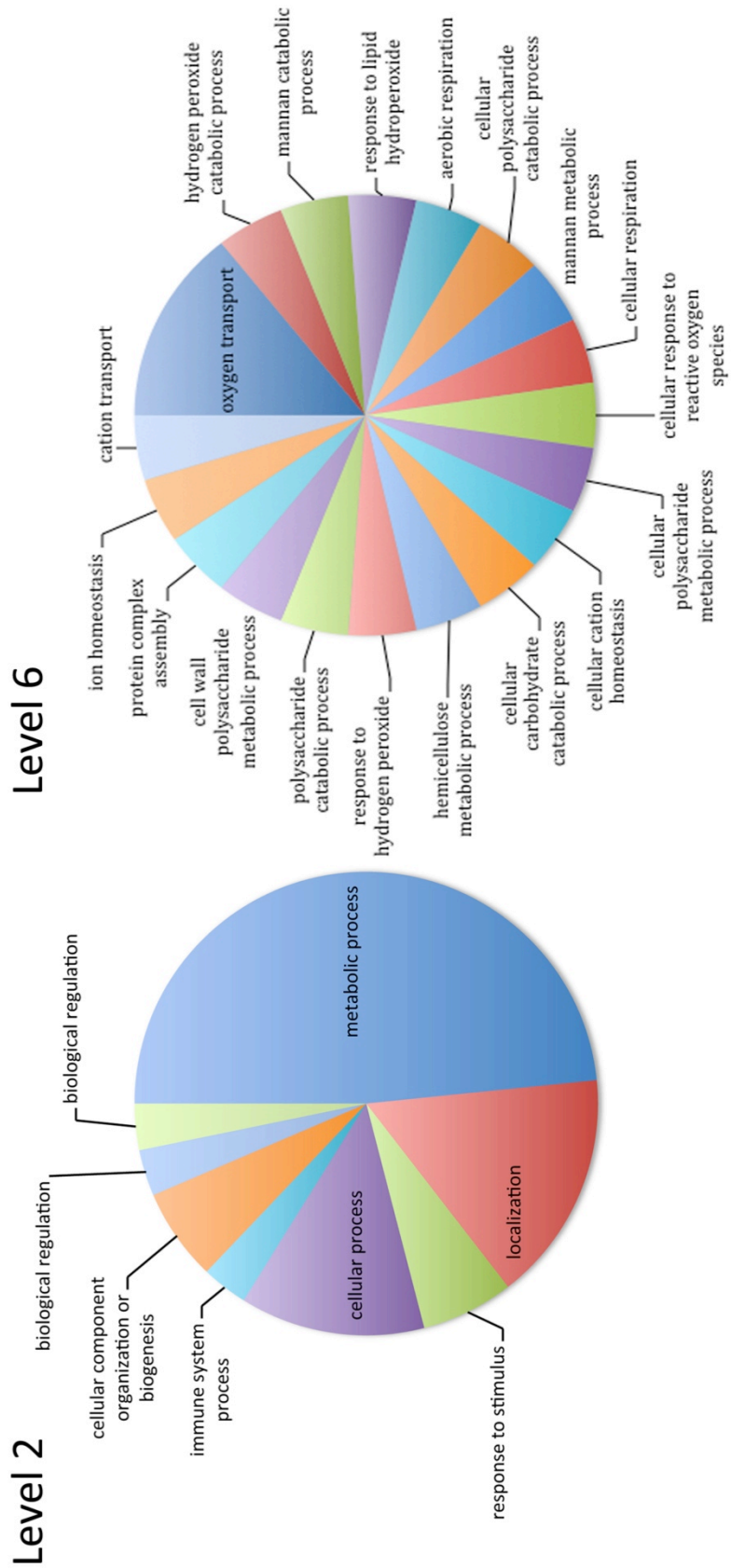
Figure 6.5. The GO terms associated with the annotated contigs present in the 35 most abundantly represented contigs of the SPS library. A) GO terms at level 2. B) GO terms at level 6.
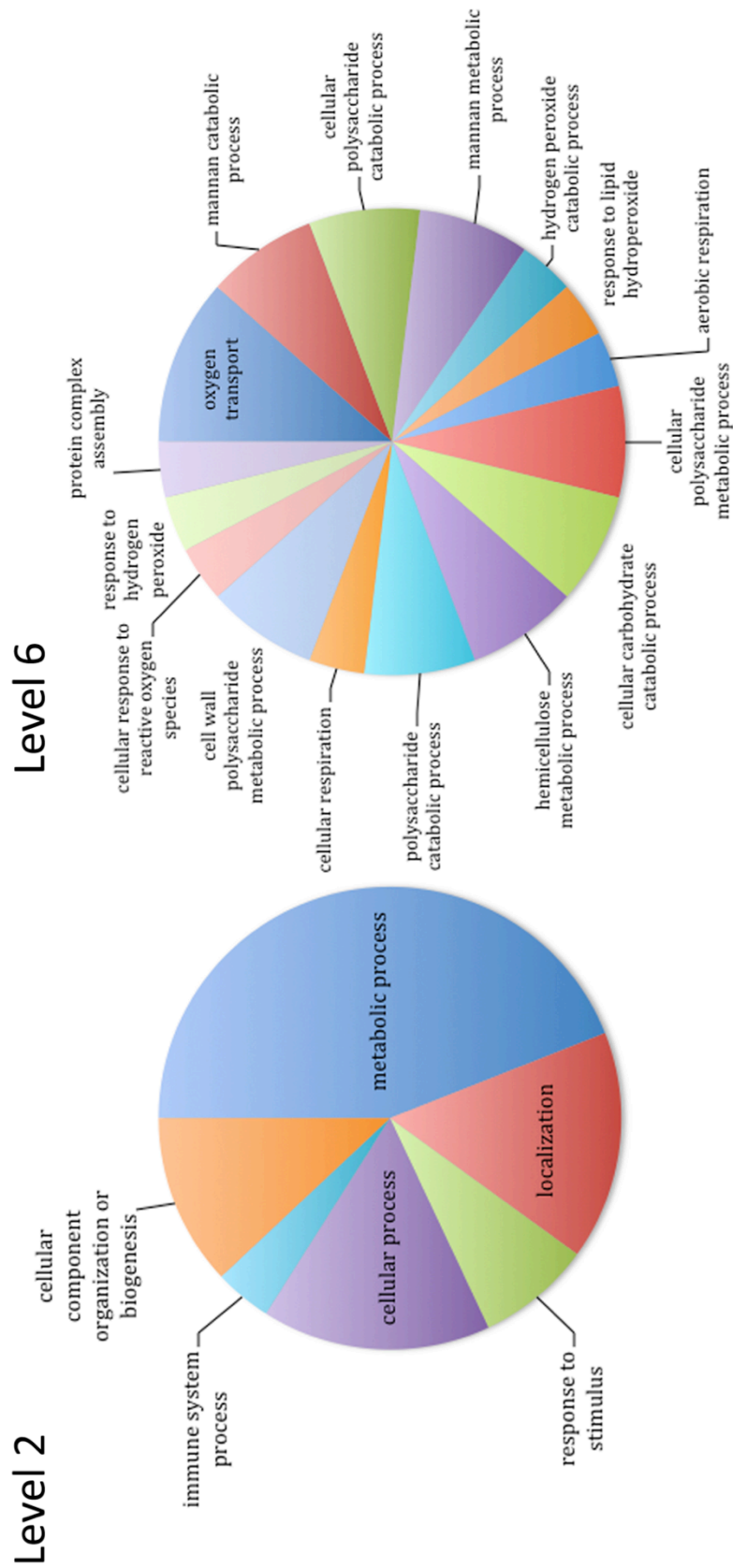
Figure 6.6. The GO terms associated with the annotated contigs present in the 35 most abundantly represented contigs of the beech library. A) GO terms at level 2. B) GO terms at level 6.

6.8.3 Overview of the enzymatic groups

An overview of the transcriptomic library for the hepatopancreas of *L. quadripunctata* has already been published (King et al., 2010). To allow comparison, similar analyses were performed using the *C. terebrans* libraries, examining the EST representation corresponding to the same sequence categories (Figure 6.7). The transcriptome of *C. terebrans* is dominated by a relatively small number of sequence types (Figure 6.7A & B). The most abundant transcripts correspond to GH family proteins and haemocyanins, representing 12.2 % and 8.7 % of the SPS-fed library respectively, and 8.1 % and 16.6 % of the beech library respectively. The *C. terebrans* and *L. quadripunctata* libraries show a broadly similar distribution of EST annotation, although the *C. terebrans* libraries show a smaller percentage of ESTs contributing to the selected categories. The most represented types of ESTs in the *L. quadripunctata* library are the GH family proteins and the haemocyanins; the same can also be seen in the *C. terebrans* libraries. The ESTs contributing to the GH family proteins and haemocyanins are proportionally larger in the *L. quadripunctata* library than in either of the *C. terebrans* libraries. The ESTs contributing to proteases and ferritins are larger in the SPS-fed library.
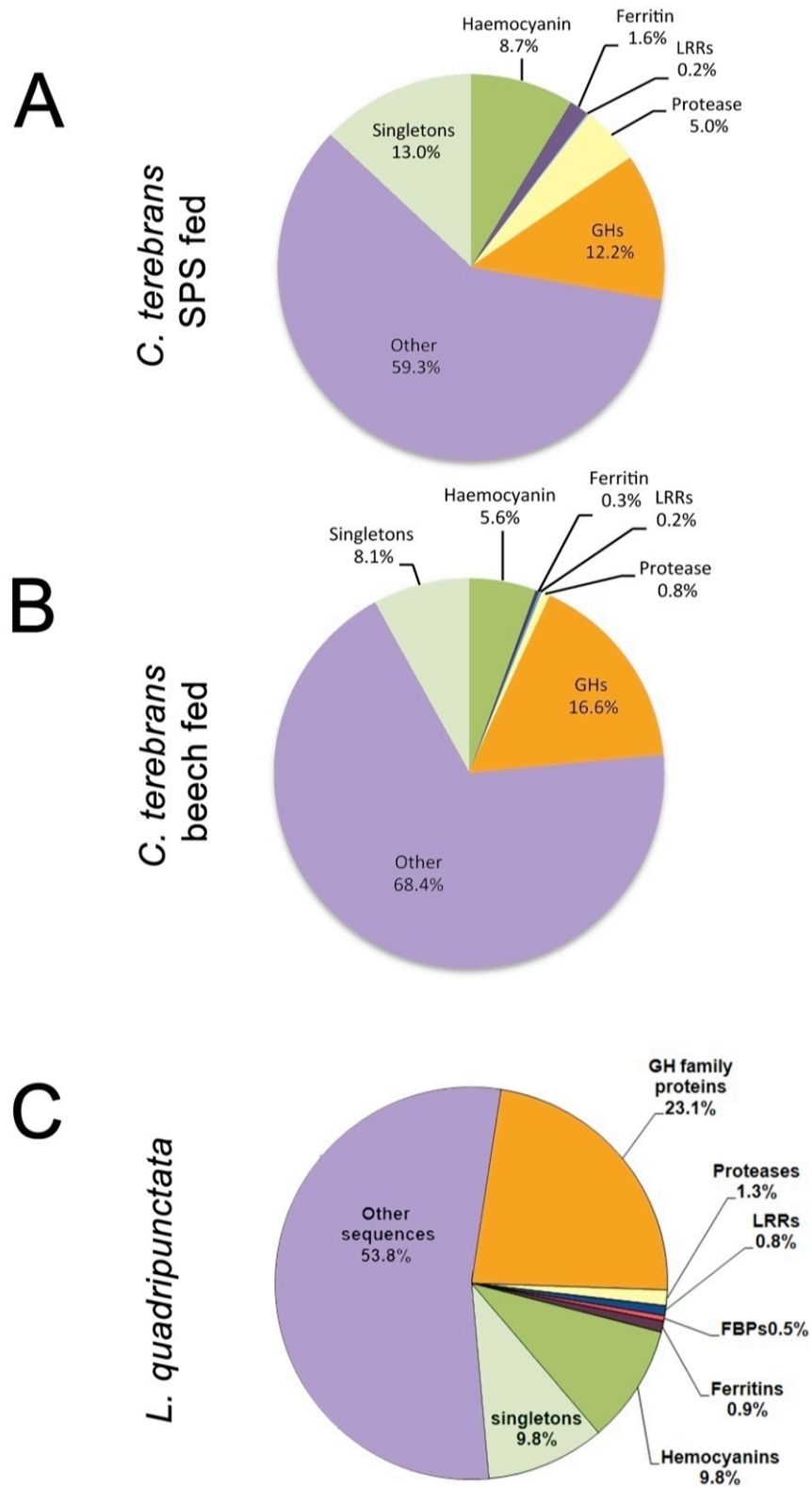
Figure 6.7 Overview of the most represented types of ESTs in the libraries.

A) Overview of the library produced from *C. terebrans* fed on SPS. B) Overview of the library produced from *C. terebrans* fed on beech. C) Overview from a library produced from the hepatopancreas of *Limnoria quadripunctata* taken from King et al, 2010).

179

The comparison with *L. quadripunctata* highlighted similarities in the most represented sequence types with those of *C. terebrans*. A table published in King et al. (2010) presented the most abundantly represented annotated ESTs; for comparison, similar tables were prepared for libraries of both *C. terebrans* (Table 6.3) and for the non-boring amphipod *E. marinus* (Table 6.4).

SPS

	Annotation	Organism	Function	E- value	GenBank	ESTs	%
1	CDH1-D [Gallus gallus]	*Gallus gallus*	cell-cell adhesion	8.00E-25	gb|AAL31950.1|	2651	4.0
2	hemocyanin subunit 1	*Gammarus roeseli*	Oxygen transporter	0	emb|CAI78901.1|	1969	3.0
3	cellobiohydrolase	*Schizophyllum commune*	Cellulose digestion (GH7 family)	1.00E-148	gb|AAX55505.1|	1127	1.7
4	hemocyanin	*Pacifastacus leniusculus*	Oxygen transporter	0	gb|AAM81357.1|AF522504_1	976	1.5
5	hemocyanin subunit 1	*Gammarus roeseli*	Oxygen transporter	0	emb|CAI78901.1|	910	1.4
6	ferritin peptide	*Fenneropenaeus chinensis*	Iron storage protein	3.00E-71	gb|ABB05537.1|	688	1.0
7	beta 1,4-endoglucanase	*Cherax quadricarinatus*	Cellulose digestion (GH9 family)	1.00E-121	gb|AAD38027.1|AF148497_1	611	0.9
8	serine collagenase 1 precursor	Celuca pugilator	Protease	2.00E-72	gb|AAC47030.1|	560	0.8
9	beta-1,4-mannanase	*Haliotis discus hannai*	Mannan digestion (GH5)	5.00E-72	dbj|BAE78456.1|	532	0.8
10	Trypsin	Paralithodes camtschaticus	Protease	2.00E-95	gb|AAL67442.1|AF461036_1	509	0.8
11	cellobiohydrolase	*Schizophyllum commune*	Cellulose digestion (GH7 family)	1.00E-146	gb|AAX55505.1|	391	0.6
12	metalloproteinase, putative	*Aedes aegypti*	Putitative Protease	1.00E-53	gb|EAT44806.1|	388	0.6
13	Glutathione peroxidase 3 precursor	*Hylobates lar*	Oxidoreductase/Peroxidase	9.00E-37	sp|Q4AEH5|GPX3_HYLLA	365	0.5
14	cytochrome c oxidase subunit 1	*Parhyale hawaiiensis*	Oxidoreductase	0	gb|AAT69307.1|	359	0.5
15	beta 1,4-endoglucanase	*Cherax quadricarinatus*	Cellulose digestion (GH9 family)	1.00E-150	gb|AAD38027.1|AF148497_1	337	0.5
16	beta 1,4-endoglucanase	*Cherax quadricarinatus*	Cellulose digestion (GH9 family)	1.00E-161	gb|AAD38027.1|AF148497_1	302	0.5
17	cellobiohydrolase	*Schizophyllum commune*	Cellulose digestion (GH7 family)	1.00E-153	gb|AAX55505.1|	249	0.4
18	beta-1,3-glucan binding protein	*Litopenaeus vannamei*	Glucan binding (GH16 family)	1.00E-125	gb|AAW51361.1|	249	0.4
19	cellobiohydrolase	*Schizophyllum commune*	Cellulose digestion (GH7 family)	1.00E-39	gb|AAX55505.1|	241	0.4
20	beta-1,4-mannanase	*Haliotis discus hannai*	Mannan digestion (GH5)	1.00E-56	dbj|BAE78456.1|	236	0.4

Beech

	Annotation	Organism	Function	E- value	GenBank	ESTs	%
1	CDH1-D	*Gallus gallus*	cell-cell adhesion	3.00E-21	gb|AAL31950.1|	2926	5.2
2	beta 1,4-endoglucanase	*Cherax quadricarinatus*	Cellulose digestion (GH9 family)	1.00E-121	gb|AAD38027.1|AF148497_1	1938	3.4
3	hemocyanin subunit 1	*Gammarus roeseli*	Oxygen transporter	0	emb|CAI78901.1|	1679	3.0
4	cellobiohydrolase	*Schizophyllum commune*	Cellulose digestion (GH7 family)	1.00E-146	gb|AAX55505.1|	1550	2.7
5	Chitin deacetylase-like, isoform C	*Drosophila melanogaster*	chitin binding	1.00E-70	ref|NP_001036326.1|	1195	2.1
6	cellobiohydrolase	*Schizophyllum commune*	Cellulose digestion (GH7 family)	7.00E-42	gb|AAX55505.1|	1011	1.8
7	beta 1,4-endoglucanase	*Cherax quadricarinatus*	Cellulose digestion (GH9 family)	1.00E-161	gb|AAD38027.1|AF148497_1	983	1.7
8	hemocyanin	*Pacifastacus leniusculus*	Oxygen transporter	1.00E-179	gb|AAM81357.1|AF522504_1	819	1.4
9	hemocyanin subunit 1	*Gammarus roeseli*	Oxygen transporter	0	emb|CAI78901.1|	712	1.3
10	Uncharacterized protein	*Apis mellifera*	Similar to (GH 38/57 family)	7.00E-07	ref|XP_001120478.1|	532	0.9
11	cellobiohydrolase	*Schizophyllum commune*	Cellulose digestion (GH7 family)	5.00E-33	gb|AAX55505.1|	449	0.8
12	Glutathione peroxidase 3 precursor	*Hylobates lar*	Oxidoreductase/Peroxidase	3.00E-37	sp|Q4AEH5|GPX3_HYLLA	444	0.8
13	beta-1,3-glucan binding protein	*Litopenaeus vannamei*	Glucan binding (GH16 family)	1.00E-125	gb|AAW51361.1|	272	0.5
14	beta-1,4-mannanase	*Haliotis discus hannai*	Mannan digestion (GH5)	3.00E-56	dbj|BAE78456.1|	222	0.4
15	beta-1,4-mannanase	*Haliotis discus hannai*	Mannan digestion (GH5)	5.00E-72	dbj|BAE78456.1|	210	0.4
16	chitin binding	*Drosophila melanogaster*	chitin binding/ chitin metabolic process	5.00E-12	gb|AAL49119.1|	209	0.4
17	beta 1,4-endoglucanase	*Cherax quadricarinatus*	Cellulose digestion (GH9 family)	7.00E-75	gb|AAD38027.1|AF148497_1	199	0.4
18	cytochrome c oxidase subunit I	*Pagurus longicarpus*	Oxidoreductase	1.00E-177	ref|NP_150616.2|	198	0.3
19	ferritin peptide	*Fenneropenaeus chinensis*	Iron storage protein	3.00E-71	gb|ABB05537.1|	176	0.3
20	1,4-beta-cellobiohydrolase	*Cochliobolus carbonum*	Cellulose digestion (GH7 family)	7.00E-26	sp|Q00328|GUX1_COCCA	172	0.3

Table 6.3 Most abundantly represented annotated ESTs from the libraries of *C. terebrans*. Blue – GH7 family proteins; Green – GH9 family proteins; Yellow – GH5 family proteins.

L. quadripunctata

	Annotation	Organism	Function	E- value	GenBank	ESTs	%
1	Cellulase	Pseudotrichonympha grassii	Cellulose digestion (GH7 family)	1.00E-145	BAB69425	23,668	8.4
2	Putative glycosyl hydrolase family 7	Protist from H. sjoestedti	Cellulose digestion (GH7 family)	2.00E-149	BAB69425	16,632	5.9
3	β-1,4-endoglucanase	Cherax quadricarinatus	Cellulose digestion (GH9 family)	6.00E-127	AAD38027	15,755	5.6
4	Hemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CAI78901	13,323	4.7
5	Hemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CAI78902	8,473	3.0
6	Hemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CAI78903	7,245	2.6
7	Hemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CAI78904	6,272	2.2
8	Hemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CAI78905	5,001	1.8
9	Hemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CAI78906	3,766	1.3
10	β-1,4-endoglucanase	Cherax quadricarinatus	Cellulose digestion (GH9 family)	1.00E-130	AAD38027	3,549	1.3
11	Fatty acid-binding protein	Bufo arenarum	Unknown	5.00E-08	P83409	3,082	1.1
12	Hemocyanin	Pacifastacus leniusculus	Oxygen transporter	0	AAM81357	2,916	1.0
13	Ferritin peptide	Fenneropenaeus chinensis	Iron storage protein	1.00E-57	ABB05537	2,455	0.9
14	Hypothetical protein	Branchiostoma floridae	Unknown (Leucine-rich repeat)	1.00E-24	XP_002223116	2,330	0.8
15	Lysosomal β-galactosidase	Canis lupus familiaris	β-galactosidase (GH35 family)	2.00E-132	ABA43388	2,293	0.8
16	Chymotrypsin BII	Litopenaeus vannamei	Protease	9.00E-67	P36178	2,094	0.7
17	β-1,4-mannanase precursor	Cryptopygus antarcticus	Mannan digestion	1.00E-75	ABV68808	1,720	0.6
18	β-1,4-endoglucanase	Cherax quadricarinatus	Cellulose digestion (GH9 family)	7.00E-135	AAD38027	1,660	0.6
19	Trypsin	Litopenaeus vannamei	Protease	2.00E-65	CAA60129	1,546	0.6
20	Hypothetical protein	Branchiostoma floridae	Unknown (Leucine-rich repeat)	1.00E-24	XP_002223116	1,455	0.5

E. marinus

	Annotation	Organism	Function	E-value	GenBank	ESTs	%
1	ATP synthase subunit a	Gammarus duebeni	ATP production	4E-33	YP_006234446	4975	5.0
2	Haemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CA178901	4637	4.6
3	Haemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CA178901	3429	3.4
4	Haemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CA178901	2319	2.3
5	Haemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CA178901	1497	1.5
6	Glycosyl hydrolase family 7	Limnoria quadripunctata	Cellulose digestion (GH7)	0	ADB85438	1415	1.4
7	Cytochrome c oxidase subunit III	Gammarus duebeni	Oxidoreductase	2E-40	YP_006234447	1203	1.2
8	Cytochrome c oxidase subunit I	Gammarus duebeni	Oxidoreductase	1E-80	YP_006234443	1099	1.1
9	Haemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	1E-69	CA178901	894	0.9
10	Monooxygenase non-catalytic subunit (LLR)	Oplophorus gracilirostris	Stabilises protein conformation	5E-30	Q9GV46	862	0.8
11	Trypsin 1a	Panulirus argus	Hydrolase/Protease	3E-113	ADB66711	807	0.8
12	Trypsin 1a	Panulirus argus	Hydrolase/Protase	9E-53	ADB66711	788	0.8
13	Destabilase I (Invertebrate-type lysozyme)	Litopenaeus vannamei	Hydrolase/Glycosidase (GH22)	4E-42	ABD65298	712	0.7
14	Ferritin peptide	Fenneropenaeus chinensis	Iron storage protein	2E-91	ABB05537	626	0.6
15	Lectin 1	Macrobrachium rosenbergii	Cell surface receptor	1E-37	AFN20597	602	0.6
16	Lectin 1	Macrobrachium rosenbergii	Cell surface receptor	5E-35	AFN20597	530	0.5
17	Beta-1-3-gulcan binding protein	Homarus gammarus	Carbohydrate metabolism (GH16)	2E-162	CAE47485	491	0.5
18	Glycosyl hydrolase family 7	Limnoria quadripunctata	Cellulose digestion (GH7)	9E-141	ADB85437	457	0.5
19	cytochrome c oxidase subunit I	Gammarus duebeni	Oxidoreductase	9E-160	YP_006234443	427	0.4
20	Haemocyanin subunit 1	Gammarus roeseli	Oxygen transporter	0	CA178901	418	0.4

Table 6.4 Most abundantly represented annotated ESTs of *L. quadripunctata* library (King et al, 2010) and the non-boring amphipod *E. marinus*. Blue - GH7 family proteins, Green – GH9 family proteins.

The twenty most abundantly represented annotated contigs in all the species shows a difference in the representation of GH family proteins. ESTs corresponding to GH families 5,7 and 9, are relatively well represented in the wood borers. However, for *E. marinus,* the glycosl hydrolases belonging to GH families 5 and 9 are absent in the most abundantly represented contigs. GH family 7 proteins are represented in the twenty most abundantly represented contigs; however, the number of ESTs contributing to these contigs represents just 0.9 % of the total library. In contrast, the glycosyl hydrolases in the *C. terebrans* and the *L. quadripunctata* libraries have considerably higher representation. The glycosyl hydrolase family 16 is represented in the 20 most abundantly expressed contigs of *C. terebrans* and *E. marinus*;

however, it is absent from the equivalent list for *L. quadripunctata*. A large abundance of ESTs contribute to an unknown protein which annotates as a protein similar to a GH38 and GH57 in the beech library. This is not seen in the most abundantly represented contigs of any other library.

6.8.4 GH family proteins

Sequences belonging to glycosyl hydrolase families have been shown to be one of the most abundantly expressed groups in *C. terebrans* and *L. quadripunctata* (Table 6.3 & Table 6.4). In the *C. terebrans* libraries, ESTs annotated as GH family proteins amounted to 12.2 % and 16.6 % of ESTs in the SPS and beech library respectively, and 26.9 % of all ESTs in the library of *L. quadripunctata* (King et al., 2010; Figure 6.7). ESTs representing particular GH families in both *C. terebrans* libraries were compared to those found in the *L.* quadripunctata library. Amongst the GH families found, ESTs representing GH7 and 9 dominate both the *C. terebrans* libraries (Figure 6.8). The proportion of ESTs representing GH9 and GH38 is greater in the beech library than that observed in the SPS library, which is dominated by ESTs representing GH7 families. The percentage of ESTs representing the GH5 family appears greater in the SPS library than that observed in the beech library. The equivalent analysis for the *L. quadripunctata* library (King et al., 2010) is presented for comparison. Here levels of ESTs representing the GH7 and 9 families dominate the library to a greater extent than that seen in either *C. terebrans* libraries, where ESTs representing GH families other than 7 and 9 make up a larger proportion (Figure 6.8). The unknown protein annotated as 38/57 in the beech library is also seen in the SPS library at a lower overall percentage, representing 2 % of GH transcripts. Although this is not highlighted in the *L. quadripunctata* library for comparison we can assume that if GH38/57 families are present, they cannot

represent more than 1.3 % of transcripts annotated as GH family proteins in the *L. quadripunctata* library.
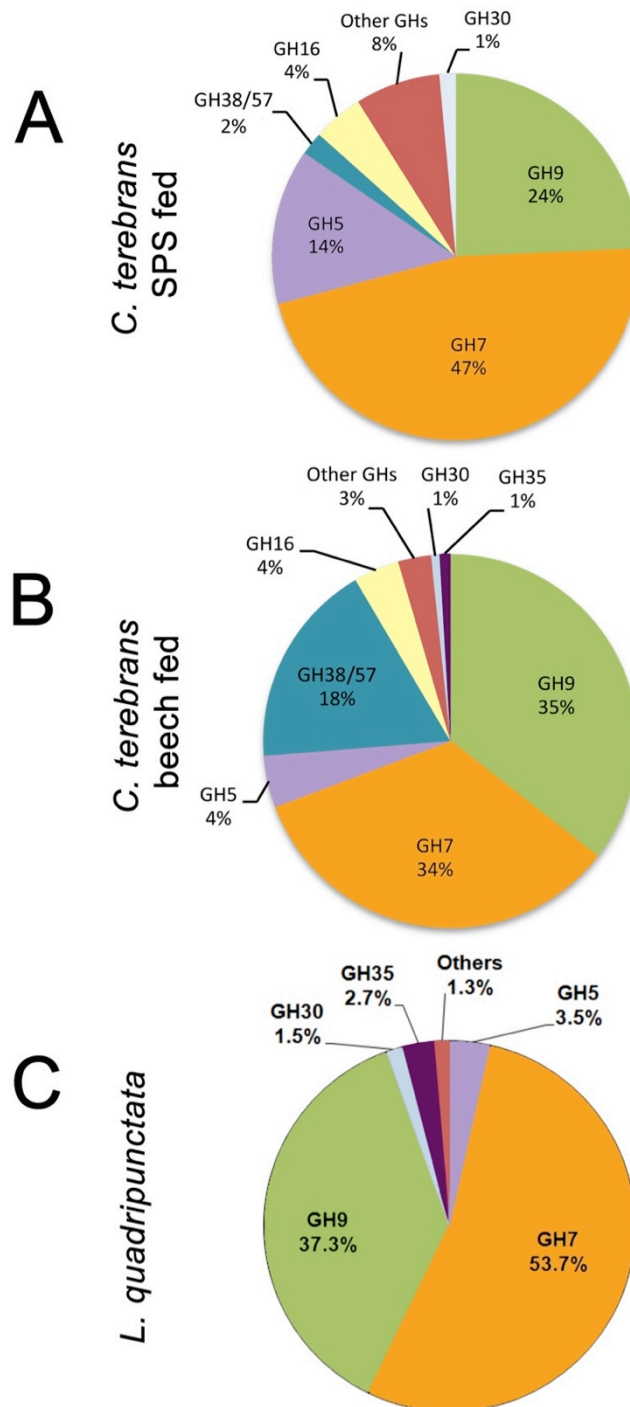
Figure 6.8 Overview of ESTs contributing to GH family proteins in the libraries. A) Overview of the library produced from *C. terebrans* fed on SPS*.* B) Overview of the library produced from *C. terebrans* fed on beech. C) Overview form a library produced from the hepatopancreas of *Limnoria quadripunctata.*

ESTs belonging to GH families following CAZy nomenclature for carbohydrate active enzymes (Cantarel et al, 2009) were identified in both the *C. terebrans* libraries, as well as those of *L. quadripunctata* and *E. marinus*. The most dominant glycosyl hydrolase proteins in both *C. terebrans* libraries and that of *L. quadripunctata* belong to the GH 7 and 9 families. Examination of the contigs annotated as glycosyl hydrolase family proteins indicated the existence of at least four distinct GH7 and four GH9 proteins (Sequence 6.1 & Sequence 6.2)

C. terebrans GH7A
MKLAAAVFLGLACVVYGQQAGTQTEEVHLSLPIQNCESGSCSSESTSIVLDSNWRWAHA
VDDYTNCYDGNDWVDEYCPDAATCTENCAIDGVDEASWTSTYGISASGDGVTLTFVTEG
TYSTNIGSRVYLLASDTAYRMFYLLNREFTVDIDSSNLPCGLNGALYFVEMEEDGGMGS
YSTNTAGAEYGTGYCDAQCPHDMKFIAGQANSDGWEPSDDDQNAGTGQYGICCFEMDIW
EANSQAQSFTTHSCSVSGYYPCQGTDCGDNGSDRYSGVCDKDGCDWAAYRNNQLDFYGP
GFTVDSGSTITVITQFLTSDGSDNGQLSEVRRIYVQGGQEIQNTVVNWEGVAEYDSITS
EYCDDIKDFFGDEPDFQAKGGLSAMGDSLSRGHVLVMSLWDDHYAHMLWLDSSYPTEAD
PSTPGIARGPCPSDGGDPAVIEQENPGATVTFSNIQIGSIGSYKSRLAPSK

C. terebrans GH7B
MKLALVVFLGLACVVYGQQAGTQTEEVHLSMPIQNCESGSCSSESTSIVLDSNWRWAHA
IDDYTNCYDGNEWIEEYCPDAATCTENCAIDGVDDASWSSTYGISSSGDGITLTFVTEG
TYSTNIGSRVYLLASDTAYRMFYLLNREFTVDIDSSNLPCGLNGALYFVEMEEDGGMGS
YSTNTAGAEYGTGYCDAQCPHDMKFIAGQANSDDWVPSEDDQNAGTGKYGICCFEMDIW
EANSMAQSFTTHSCSVSGYYPCEGTDCGDNGSDRYSGVCDKDGCDWAAYRLNQLDFYGS
GMTVDSSQTITVVTQFITSGGDNGQLSEVRRIYVQGGQEIQNTVVNWDGVTEYDSITSE
YCDEIKDFFGDEPDFQAKGGLQAMGESLSRGHVLVMSLWDDHYAHMLWLDSSYPTEDDP
STPGVARGPCPTDGGDPAVIEQENPGATVTFSNVQIGPIGSYKSRLAPSK

C. terebrans GH7D
MKLAVVVFLGLACVVYGQQAGTQTEEVHLSMPIQNCESGSCSSESTSIVLDSNWRWAHA
VDDYTNCYDGNDWVEEYCPDAATCTENCAIDGVDDASWSSTYGISSSGDGITLTFVTEG
TYSTNIGSRVYLLASDTAYYMFYLLNREFTVDIDSSNLPCGLNGALYFVEMEEDGGMAS
YSTNTAGAEYGTGYCDAQCPHDMKFIAGQANSDGWEPSEDDQNAGTGQYGICCFEMDIW
EANSQAQSFTTHSCSVSGYYPCQGTDCGDNGSDRYSGVCDKDGCDWAAYRLNQLDFYGP
GMTVDSGSTITVVTQFITSDGSDNGQLSEVRRIYVQGGQEIQNTVVNWDGVTEYDSITS
EYCDEIKDFFGDEPDFQAKGGMQAMGESLSRGHVLVMSLWDDHYAHMLWLDSSYPTEAD
PSTPGIARGPCPSDGGDPAVIEQENPGATVTFSNVQIGSIGSYKSRLAPAK

C. terebrans GH7E
MKGALFLLVWSLICWACHGQQGGTATEEYHLPFPMQSCDSSGCQEEATTIVIDADVRQM
YSAADSSVSCQTGQGWDSELCPDGITCAENCALDGIDEAGYQSIYGVTSDGETIDFKYL
TESSYGTNIGGRLYLLASENEYRLFKMKNREFTIDVDTSQLPCGLNGAVFFVEMEGDGG
VSSYPSNSAGAKYGTGYCDAQCPTYLRVIAGEVNIGQSNYGICCGELDIWEANSAAQTY
TFHTCTSEGYYPCQGVDCGSDATGDHYNGVCDKDGCEFGAYRMNQHEYYGPDSSFDVDS
SQPITVVTQFLTEDGTDEGDLVEMRRLYVQNNQVIRNTRVNFEGIPDYDSITDEFCNDY
KELFGEVPDFAAKGGLQSMGEALDRGMVLVLCLWEDYSNHMQWLDGISPADGDPSDPGV
LRGPCPADSGRPGQMHQNYPDAYVRYSNVKFGTINSTFTL

Sequence 6.1 Glycosyl hydrolase family 7 sequences identified in *C. terebrans*.

C. terebrans GH9A

MWTVAVVAVLLVAAGGLTNAQSNPCSGSGQGPYDYQQALCMSILFYEAQRSGPLPGNMR
IDWRGDSAMNDGSDNGVDLTGGYYDAGDHVKFGFPMAYSVTVLAWSLLSYRSGYTTAGQ
VGYAEDAVKWGTDYFLKAHSGDNVLYGQVGDGNSDHAYWGRPEDMTMNRPSWAITNSAP
GSDLAGETSAALAAASIVFQSDSSYSSQCLAAARNLYDFADAHRELYHNSITAAADFYK
SWSGYGDELAWAAVWLYRATGEASYESAAKGLWTEFDIGLGSDRFDWDNKYAGVQVLFA
EEFGDSTYTSAVQSFMSGVRRTQTTPGGMVYIDQWGSLRHAVNVAFIGFKAADLGLDAS
ANRNFASDQVNYALGSVGHSFMVGFGNNPPTQPHHRGSSCPDPPDSCTDGWAQSQSGPN
PQTLWGAMVGGPDENDGYTDDRNDYVHNEVACDYNAALTGALAAMVTLNL

C. terebrans GH9B

MRIDWRGDSALGDGSDVGLDLTGGYYDAGDHVKFGFPMAYSVTVLAWGLLSYRAGYQTA
GQVGFAEDAVRWGAEYFLKAHSGNNVLYGQVGDGDIDHSYWGRPEDMTMSRPAWDINAN
APGTDLAGETSAALAAASIVFESSDSAFSAECLAASRNLFDFADQYRATYTSSIPNAAS
FYESYSGYGDELAWSAIWLYKATGESTYETKGKAFYTEYEIQYSGYGFGWDNKYSGVQI
LFAEEFGGTYVDAVQGFMEDTRAFPHTPGGMVYIDQWGSLRNAINVAFIGFKAADLGVD
ANTNRNWAAQQVDYALGSVGHSYVVAFGVNPPERPHHRGSSCPDPPDSCTDGWAQNQDG
PNPQILYGAMVGGPDENDGYVDDRNDYVHNEVACDYNAALTGSLAAMVVNNM

C. terebrans GH9D

MFFSVALVAGLLIAGSGLTDAVGYCGAIGGSPYDYGELLCHAIVFYEANRSGRLPSNQR
VEWRGDSALDDGSDNGVDLEGGYYDAGDLVKFGFPLAFTVSFLSWGILSYGDGYTQAGQ
LDNAHAAIRWATDFILKAQTGPTTLWGQVGDGDLDHSLWGRPEDMTMNRPSFKIDASKP
GSELAGESAAGLAAASIVFKNSGDNSYANTLLSAARELFDFANDHRANYDVSMPEASDF
YHSYSGYGDELTWAAIWMYKATGEASYRSTAESFYTEFEIQYGASFSWENKKGGCIALF
AEELGGNYVNILKSHAQELRAISTTPGGLSYFQHWGSLRYALNHAFIAFKAADLGQDVS
QNNQFGERQVNYVLGDNPRSSSYVVGVLNNPPLRCHHRPASCPDLGTPCGWDYGNSPDP
NPHILYGALVGGPDDSDGYTDERQNWEQNEVGCDYNAALSSVLARMVSNQIKNGEFSQT
NQ

C. terebrans GH9E

MVFQDTDAAYAAECLEVARDLFDFADQYRLMYHESIIDAANFYKSFNGFGDELGWGSMW
LFKATGEQNYADKAMEYWDEFDIHYNGYGFSWDNKHSGAQILYAEEFGVSPYTDAVELF
MSNMRGREHTPGGMFYLDPWGSLRHAVNVAFIGFKAADLGIDETVNRDWAIQQVQYALG
SVGHSYIVGFGVNPPERPHHRASSCPDLPDSCTDGWAEEQPGPNPQVLYGAMVGGPDQD
DGYTDERMDYQHNEVACDYNAALTGALAQMVILNS

Sequence 6.2 Glycosyl hydrolase family 9 sequences identified in *C. terebrans.*

The GH5 proteins are also a highly expressed family group in *C. terebrans*, with at least two distinct sequences identified in this library.

```
C.terebrans GH5A
RPVDTQYSDGGCRASTLVLAMLKLLVVLALVAVSQTQAARLSVGGSSLNYGGQQVYLNGV
NAAWNSYGYDFGNGGYDGSLEGWMREINQAGGNSFRVWVHVEGYSSPNFDGQGMVTACDS
SGQFIGDVQTFLDSAQSSNILVTLTLWNGAVLSNQPYHDLIMDDNKLQSYLDNCLAPLAQ
AVAGHPALGSWEPINEPEGSVQVASDSNPCYDTTLIGESGAGWTGANIPMQRFLRFIGKM
NQVIRSNDPGGIISVGSWGQFPQSDAFSNTHNHYTDSCLNGAAGGSNAGVDFAQMHTYDW
GDSWSPNAPFTVSASAYGLGMPIVIGEFSSDCSLGTPLSELHENGYTNGYSGTWTWHWAA
TGECSDSRDVQRQGLGALAGRSDNGVVDINVG

C.terebrans GH5B
YSDGGPQKVPSGNTVNMTRIVLLVLLAGLGVTQAARLSVSGKDLMYDGEKVYLSGVNAAW
NSYGYDFGNGNYDGTLENWLQDIGSSGGRALRMWVHVEGFSTPNYNNQGMVTACDSTGDF
IDDIISLLDAAQQSDVLVTLTLWNGAVLSNQVYSDMILEDAKLESYLDNCLAPLAQAVAG
HPALGSWEAINEGEGSIQVASDSNWCYDTTIIGREGAGWSGANIPMERMLHFIGRINQVV
RANDPTGMITLGSGHNEQSDAFSNTVYHWSDECLNGAAGGSGAGLDFNQMHTYEWQGEWT
TNAPFTVDADDYGLTKPIVIGEFSSDCATGIPMNDLLEYAYTRGYSGTWTWHWGGTGDCS
DTRAAQRTAFGYLIDRTENGLVDIDVN
```

Sequence 6.3 Glycosyl hydrolase family 5 sequences identified in *C. terebrans.*

The distributions of these GH5, 7 and 9 family proteins within those so far identified and publicly available on GenBank (NCBI), may indicate whether the genes encoding these proteins are of endogenous origin. A phylogenetic tree was made containing the GH9 protein sequences identified in *C. terebrans* and *E. marinus* with other GH9s already identified, including those of *L. quadripunctata*. Another tree was made in this way for the GH7 proteins of *C. terebrans* and *E. marinus*, The proteins from *C. terebrans* corresponding to the GH9 family clearly group with other crustacean sequences (Figure 6.9) separate from, plants, fungi and eubacteria.
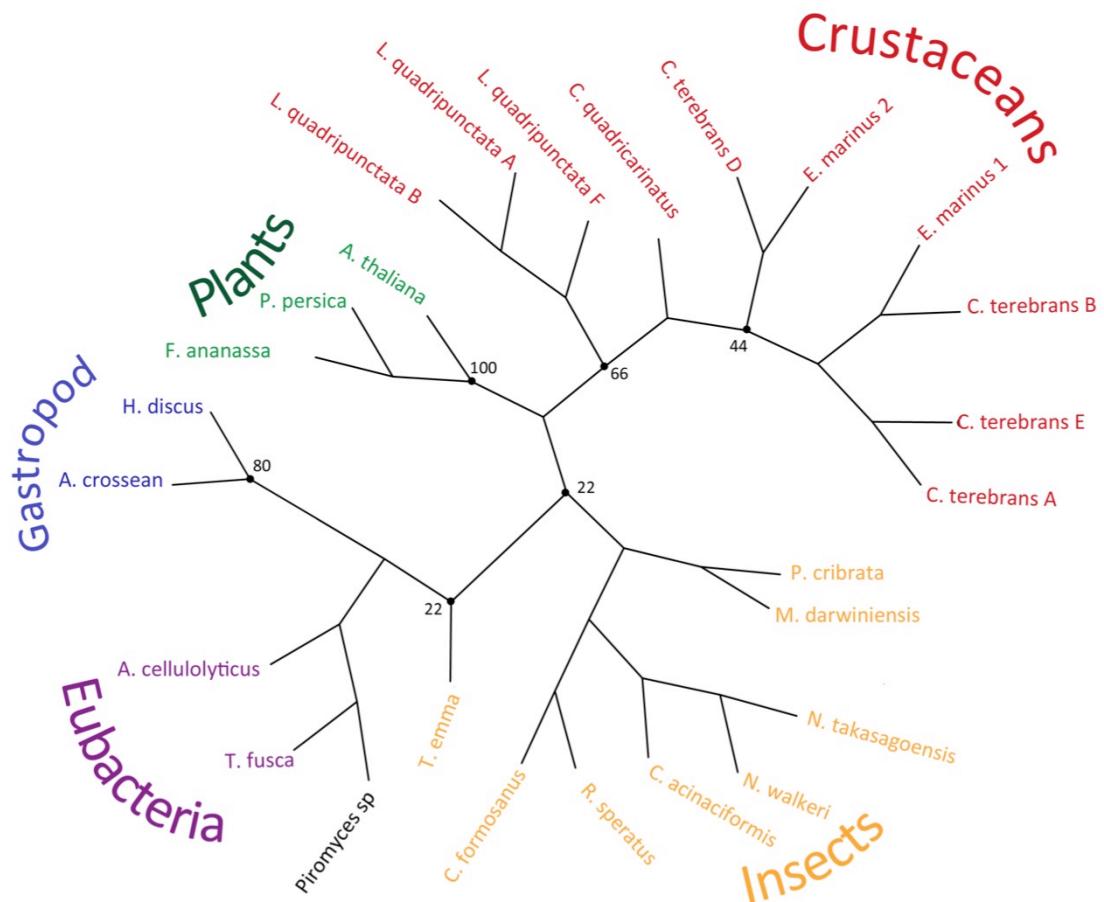
Figure 6.9 GH9 family proteins of *C. terebrans*, other arthropods, gastropods, eubacteria, plants and a fungus shown on an unrooted phylogenetic tree. Arthropod sequences include *L. quadripunctata* (FJ940759.1, FJ940760.1, FJ940761.1), *Cherax quadricarinatus* (AAD38027.1), *Teleogryllus emma* (ABV32557.1), *Panesthia cribrata* (AAF80584.1), *Mastotermes darwiniensis* (CAD54730), *Nasutitermes takasagoensis* (BAA33708), *Nasutitermes walkeri* (BAA33709*) Coptotermes acinaciformis* (AAK12339), *Reticulitermes speratus* (BAA31326), *Coptotermes formosanus* (BAB40696) and *E. marinus* (personal communication). Gastropod sequences include *Ampullaria crossean* (ABD24280) *Haliotis discus* (ABO26609). Eubacterial sequences include *Acidothermus cellulolyticus* (YP_873459) and *Thermobifida fusca* (YP_290232). Plant sequences include Arabidopsis thaliana (NP_192843), *Fragaria x ananassa* (AAC78298)*, Prunus persica* (CAI68019). *Chytridiomycote fugus* sequence from Piromyces sp. (AAM81967). Branch length proportional to estimated divergence along each branch. Bootstrap values (n=100) for the main branches are shown as percentages.

The GH family 7 proteins from *C. terebrans* also group with other crustacean GH7 proteins, separate from protists and fungi (Figure 6.10).
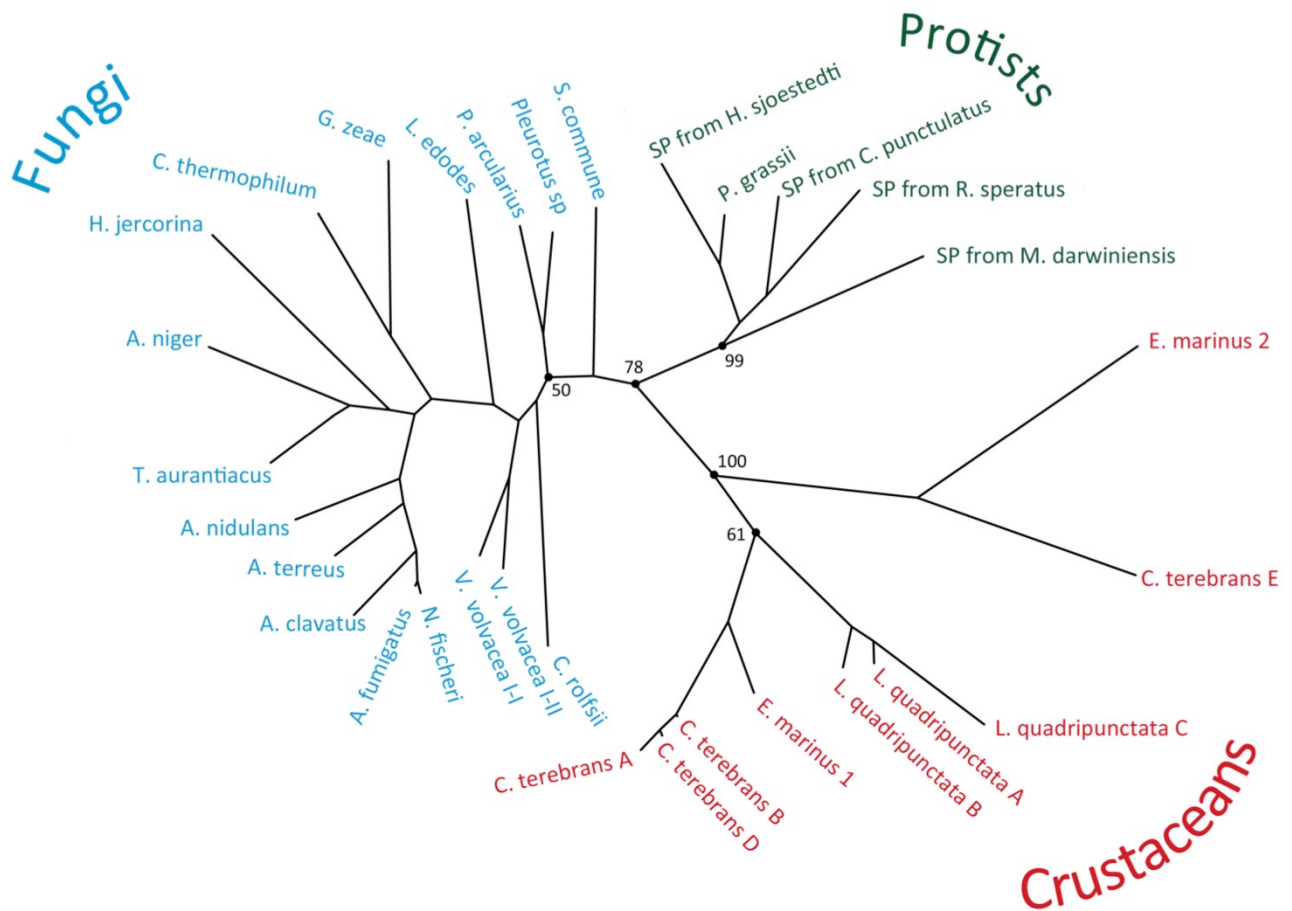
Figure 6.10 GH7 family proteins of *C. terebrans* and other crustaceans with those of ascomycete and basidiomycete fungi and protists shown on an unrooted phylogenetic tree. Ascomycete sequences include *Aspergillus clavatus* (GenBank Accession XP_001272622), *Aspergillus fumigatus* (XP_750600), *Aspergillus nidulans* (AAM54069), *Aspergillus niger* (AAF04492), Aspergillus terreus (XP_001212905), *Chaetomium thermophilum* (CAM98448), *Gibberella zeae* (AAR02398), *Hypocrea jercorina* Cel7A (P62694), *Neosartorya fischeri* (XP_001257823), and *Thermoascus aurantiacus* (CAM98447). Basidiomycete sequences include *Corticium rolfsii* (JC7979), *Lentinula edodes* (AAK95563), *Pleurotus* sp. "Florida" (CAK18913), *Polyporus arcularius* (BAF80326), *Schizophyllum commune* (AAX55505), *Volvariella volvacea* cellobiohydrolase I-I (AAT64006), and cellobiohydrolase I-II (AAT64007). Crustacean sequences include *L. quadripunctata* GH7A (FJ940756), GH7B (FJ940757), and GH7C (FJ940758) and those from *E.*

marinus (personal communication). Protist sequences include *Pseudotrichonympha grassii* (BAB69426) and a selection of uncultured symbiont protists isolated from the guts of the wood roach *Cryptocercus punctulatus* (USPCs1; BAF57469), *Hodotermopsis sjoestedti*, (USPHs1; BAF57344), *Mastotermes darwiniensis* (USPMd; BAF57427), and *Reticulitermes speratus* (USPRs1; BAF57302). Unrooted tree with branch length proportional to estimated divergence along each branch. Bootstrap values (n=100) for the main branches are shown as percentages.

The GH family 5 proteins from *C. terebrans* also group with other crustacean GH5 proteins, separate from protists, plants and fungi (Figure 6.11).
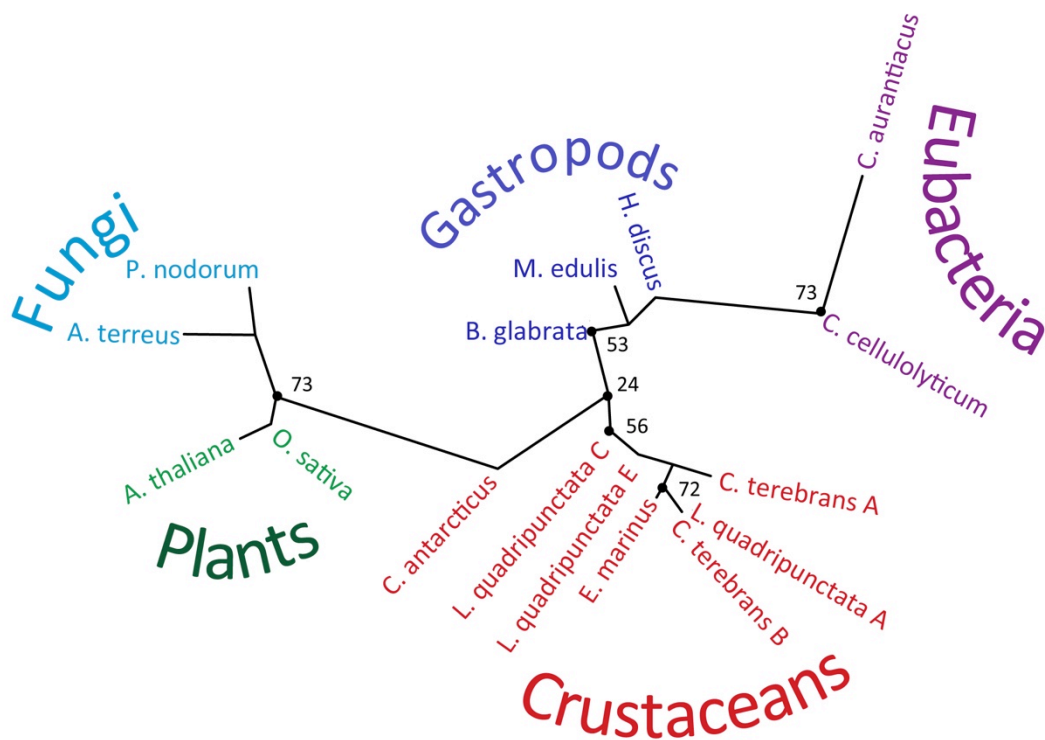
Figure 6.11 GH5 proteins of *C. terebrans*, with other arthropods, gastropods, plants, fungi, and eubacteria shown on an unrooted phylogenetic tree. Crustacean sequences include *L. quadripunctata* GH5A (GU066826-8), GH5C (GU066827), and GH5E (GU066828) and *E. marinus* (personal communication) and an additional arthropod GH5 from *Cryptopygus antarcticus* (ABV68808), *Biomphalaria glabrata* (AAV91523), *Mytilus edulis*

(Q8WPJ2), *Haliotis discus discus* (ACJ12612), *Haliotis discus hannai* (BAE78456), and *Branchiostoma floridae* (1; XP_002227661, 2; XP_002245006). Eubacterial sequences include *Chloroflexus aurantiacus* (YP_001635252) and *Clostridium cellulolyticum* (AAG45159). Plant proteins include *Arabidopsis thaliana* (NP_171733) and *Oryza sativa* (Q6Z310). Fungal sequences include *Phaeosphaeria nodorum* (XP_001799860) and *Aspergillus terreus* (EAU29440). Unrooted tree with branch length proportional to estimated divergence along each branch. Bootstrap values (n=100) for the main branches are shown as percentages.

6.8.5 Structural modelling of *Chelura terebrans* GH7 proteins

Alignment of the *C. terebrans* GH7 protein sequences reveals that GH7E is lacking a highly conserved loop region found in GH7A, B and D between residues 209-220 (Figure 6.12). The structural predictions of *C. terebrans* GH7B and E reveal the potential consequence of the sequence absent from GH7E. The deletion of this region is predicted to remove a loop (Figure 6.13), changing the normally obstructed catalytic substrate tunnel (Figure 6.13B) into a more open conformation in GH7E (Figure 6.13C) (alignments and structural predictions of *C. terebrans* GH7 proteins were carried out by Dr. John McGeehan at the University of Portsmouth).
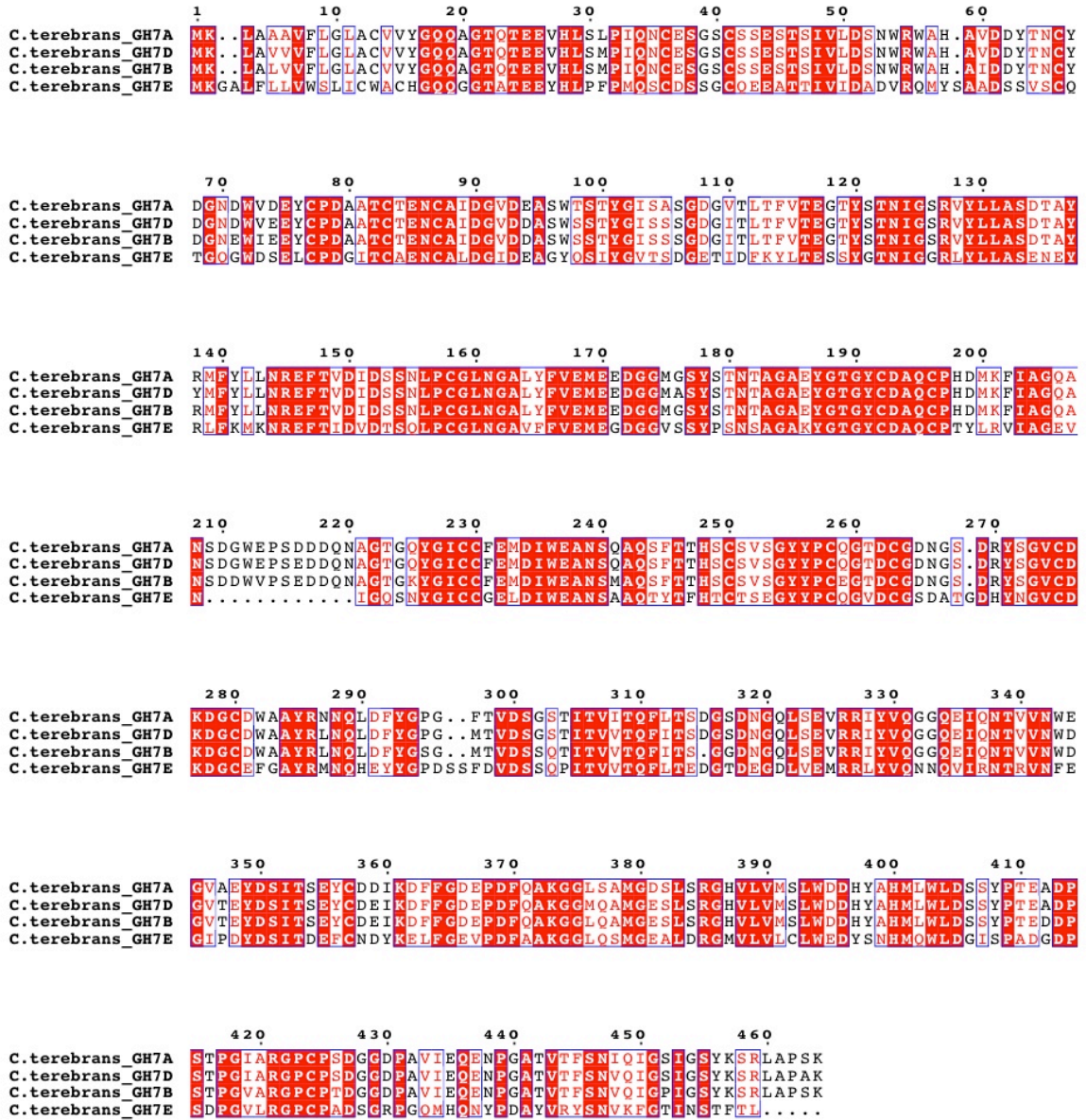
Figure 6.12 Primary sequence alignments of *Chelura terebrans* GH7 proteins. Alignment of *C. terebrans* GH7A, B, D and E. The alignment was performed in ClustalW (Larkin et al., 2007) using the default settings via the EMBL-EBI interface (http://www.ebi.ac.uk) and enhanced using ESPript2.2 (http://espript.ibcp.fr) (Gouet et al., 1999). Red boxes – strict identity; red character – similarity within a group.
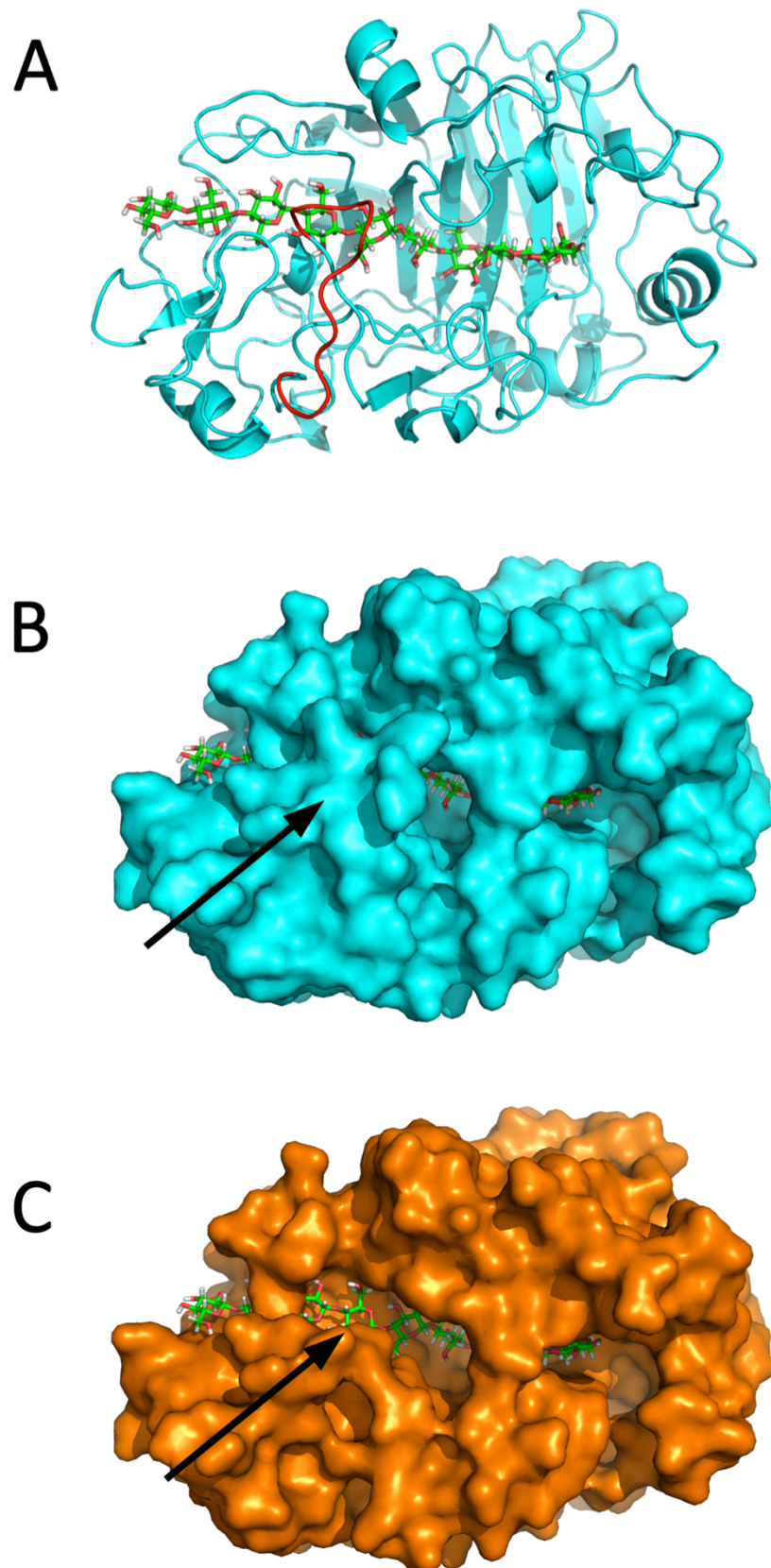
Figure 6.13 Structural modelling of *Chelura terebrans* GH7 proteins. A) Cartoon structural representation of *C. terebrans* GH7B showing secondary structure elements. A cellulose molecule was docked into the catalytic

substrate tunnel and the loop present in GH7A, B and D but missing in GH7E is shown in red. B) Surface model prediction of *C. terebrans* GH7B showing a cellulose molecule enclosed within the catalytic substrate tunnel. The atomic coordinates of the *Limnoria quadripunctata* GH7B (PDB ID: 4GWA) together with the *C.terebrans* GH7B amino acid sequence were used as a template for generating the structural model with the SWISS-MODEL (http://swissmodel.expasy.org). The loop obstructing the catalytic substrate tunnel in GH7a, B, and D is indicated by the arrow. C) Equivalent surface model prediction of *C. terebrans* GH7E showing a cellulose molecule within the catalytic substrate tunnel and the relatively more open region caused by the missing loop indicated by the arrow. Alignments and structural predictions of *C. terebrans* GH7 proteins were carried out by Dr. John McGeehan at the University of Portsmouth.

6.8.6 Haemocyanins

BLAST analysis of both *C. terebrans* libraries found no prophenoloxidases in the hepatopancreas transcriptome. However, this study has already shown a correlation between *C. terebrans* haemocyanin and phenoloxidase activity in the hepatopancreas extract (Chapter 5). The number of ESTs representing haemocyanins in both libraries indicates that this is a very abundantly expressed group in the hepatopancreas. In the *C. terebrans* libraries, ESTs annotated as haemocyanins amounted to 8.7 % and 5.6 % of ESTs in the SPS and beech library respectively, and 9.8 % of all ESTs in the library of *L. quadripunctata* (King et al., 2010; Figure 6.7). Examination of the contigs annotated as haemocyanins indicated the existence of at least three distinct haemocyanin subunits of around 670 amino acids in length (Sequence 6.4).

C. terebrans haemocyanin 1
MRALLVLCLFVTGSLAWPSFLNDAPVVSLAKKQQDINRLVFRINEPLRFEDLKTIASTFD
PVADLSLYTDGGVAVQALVTELTDGRLLEQKHWFSLFNPRQREEALLLFNVLMTASTWET
AVNNAAYFRERVNEGEFVYALYAAVIHAEAGAGIVLPPLYEVTPHLFTNSEVIQKAYTAQ
MTQIPGKFEMAFTGSQRNPEQRVAYFGEDIGLNVHHVTWHLDFPFWWQDSYGYNLDRKGE
LFFWAHHQLTVRFDSERLSNYLDLVDELYWDRPIVEGFAPHTTYRFGGEFPARPDNVQFE
DVDGVIRVRDMIIHEDRIRDAIAHGYITAADGTKLNINDAGGINLLGDIVESSVYSPNQA
YYGALHNEAHILLGRQGDPHGKFKLPPGVMEHFETATRDPAFFRLHKYMDNLFKEHKDTL
PAYTKEELEFVGIELESVTVNGPLETFFEDFDVDLVNAVDTSPEVDDVAIGATVKRLNHK
PFSVAFEINNNNGVEKHGVVRTFLCPRRDVNGVIYSFEEGRWNCIELDKFWTKLAPGSNT
VTRDSSDAAVTVPDVPSFQTLIDTADAAVASGSPLGLEKFDRSCGIPNRLLLPKGTVEGL
EFALAVAVTDGETDAQHDVLESAGAHSHAQCGVHGEKYPDHQPMGFPLDRRIEDERIILA
VPNIQYQIVKVTFSG

C. terebrans haemocyanin 2
MKLLILCVVVTSALAWPSSYVRDDYAVNDAAPSLAKRQQDVNRILYRVTEPLQFEELVTA
AANFDPVADTSIYTDGGKAAQTLVDEITDGRVLERHHWFSLFNTRQRTEALLLFDVFAHV
KTWETAINAAAYFHDKINEGEFVYAVYAAVIHSPLGAGIVLPPLYEVTPHLFTNSEVIQK
AYTAQMTQTPGKFKMEFTGSQKNPEQHVAYFGEDIGMNVHHVTWHLDFPFWWDDAYGYHL
DRKGELFFWAHHQLTVRFDAERLSNHLDLVDELYWDRPIVEGFAPHTTYRYGGEFPARPD
NVQFEDVDGVIRVRDMIIHESRIRDAIAHGYITAKDGSQIDIRDEHGIDHLGDIIESSVY
SPNIQYYGALHNEAHIILGRQGDPHGKFNLPPSVMEHFETATRDPAFFRLHKYMDGIFKE
HKDSLPPYTKEQIGFTGVHLTGVSVEGELETFFEDFEFDLKQAVDLSEAVEEVDLSAYVS
RLNHKEFAYNFDITSDTAEAHAVIRVFLCPRRDNNGVIFTFEEGRYNCIEMDKFWTKLSS
GANSIKRKSSESSVTVPDIPSFSSLVAQADEAVASGSELHLEEFDRSCGIPNRMLLPKGT
KDGMEFALVVAVTDGSTDAQHEALESAAAHSHAQCGVHGEKYPDHQPMGFPLDRRIPDDR
VFLTSDNVAYTIVKVFYKE

C. terebrans haemocyanin 3
MKLILVLAAVAAVCVADDLAVKQQTINRLLLKVTEPIRSYFTDLSALSKTYDPRSSGIAS
AAALLDEIEAGRVLPQKQIFSLFNDRQREEALLLIDTLLAATDFDTFKGNAAYFREKVNE
GEFVYALYVAVTHSDLTQDVVLPPLYEVTPHLFTNSEVINKAYSAKMTQTPGRFKMEFTG
SQRNPEQRVAYFGEDVGMNSHHVNWHMDFPFWWKGYSIDRKGELFFWVHHQLTARFDAER
LSNYLSVVDELYWDRPIYEGFAPHTTYRYGGEFPSRPDNKYFEDVNGVARIRDMKIIESR
IHDAIDHGYIIDAEGTQIPLDAENGIDILGAVIESSTSSPNVQYYGALHNTAHKMLGRQA
DPQGKFKLPPGVMEHFETATRDPSFFRLHKYMNNIFKDFKNTLPPYTAEELGYANAEITS
LSIDGALETYFEDYEFSLLNAIDDTESIDDVAISTYVSRLNHKDFAYNIGVKANADEVAT
VRIFLCPKYDSNGVEYTLDEARWGCIQVDKFWTQLTAGSNTIVRQSSDSSVTIPDRVPFA
TLIEEADAAVAAGSDLPVYNPRGCGQPQRLLLPKGNKEGLDLELFVSITSGEDATIDDLT
TNDYGGSYSYCGIKGQKYPDSRAMGYPVDRHIEDDRLFRIPNIKWTTVKVFFTEQ

Sequence 6.4 Haemocyanin subunits identified in *C. terebrans*

The proportions of ESTs annotated as haemocyanins representing each subunit found in *C. terebrans* (Sequence 6.4) were compared in both libraries (Figure 6.14). The proportions of these ESTs were very similar for each library. The number of ESTs representing haemocyanin 1 in *C. terebrans* dominates the ESTs annotated as haemocyanins. The proportions of ESTs representing haemocyanins 2 and 3 are almost equal in both libraries
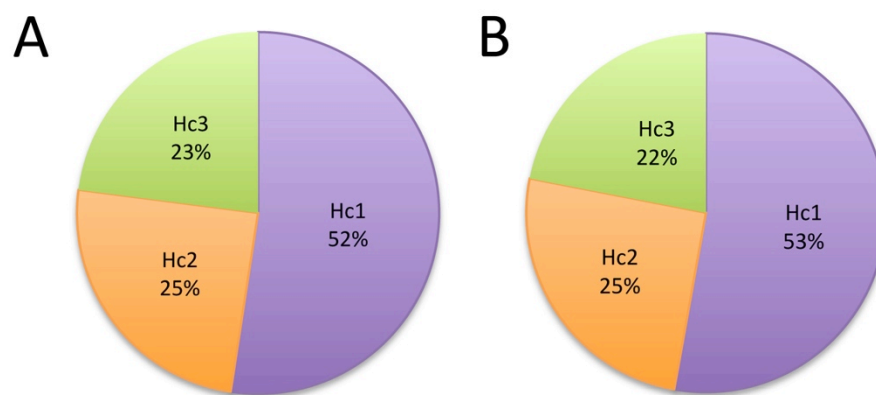
Figure 6.14 Proportions of ESTs annotated as haemocyanins representing each subunit found in *C. terebrans*. A) SPS fed library B) Beech fed library. Hc – haemocyanin

Four haemocyanin subunits were reported in *L. quadripunctata* (GU166295-98, King et al., 2010) and appear very closely related. Indeed, in a phylogenetic analysis, these four haemocyanin subunits formed their own clade within the isopod group (King et al., 2010). As in *C. terebrans*, three haemocyanin subunits were also found in the transcriptome of *E. marinus* (pers. comm.). To better understand the relationship between the haemocyanins found in *C. terebrans*, a phylogenetic tree was produced using these sequences and those of several other crustaceans available on GenBank (NCBI) (Figure 6.15).
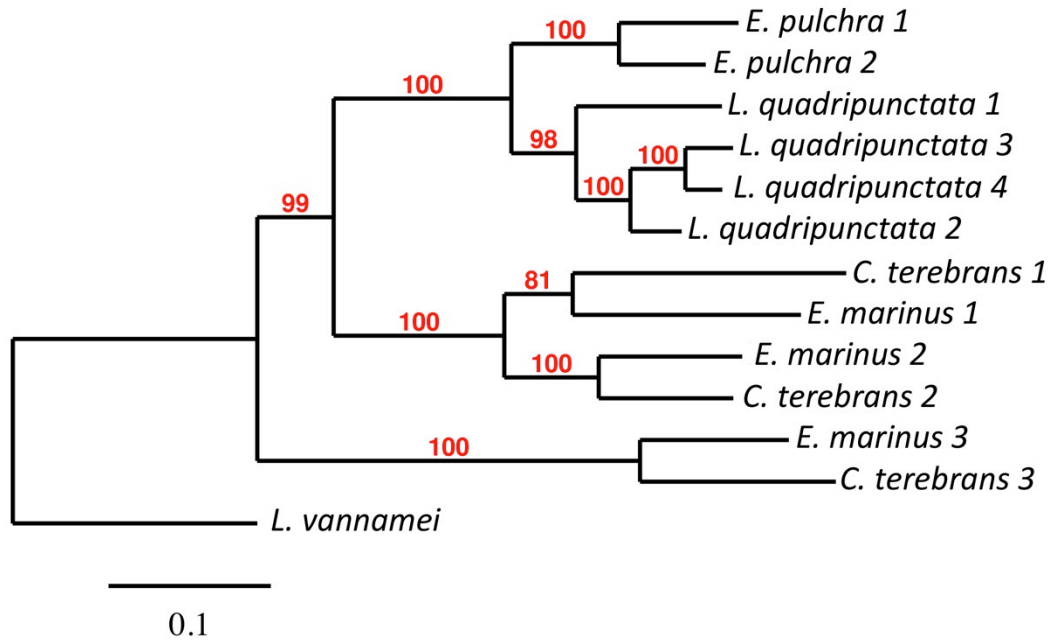
Figure 6.15 Phylogenetic analysis of haemocyanin subunit*s.*

Phylogenetic tree showing the relationship between heamocyanin proteins of *C. terebrans* and other crustaceans. Haemocycanin proteins from the decapod *Litopenaeus vannamei* (CAB85965.1) the isopods *Limnoria quadripunctata* (GU166295-98) and *Eurydice pulchra* (1 – ACS44712 & 2 – ACS 44713) and the amphipods *Echinogammarus marinus* (personal communication) and *Chelura terebrans.* All branches are drawn to scale as indicated by the scale bar. Boot strap values (n=100) for the branches are shown as percentages.

As seen in the King et al. study (2010), the haemocyanin subunits from *L. quadripunctata* form their own clade (Figure 6.15). However, unlike *L. quadripunctata*, each haemocyanin subunit in *C. terebrans* pairs with an orthologous sequence found in *E. marinus*. The tree also suggests that the haemocyanin subunits 1 and 2 of the amphipods are more closely related to those of the isopods to their own haemocyanin subunit 3.

Given the *C. terebrans* haemocyanin subunits posses orthologues in *E. marinus*, it may be informative to compare the relative numbers of ESTs contributing to each subunit between the two species (Figure 6.16).
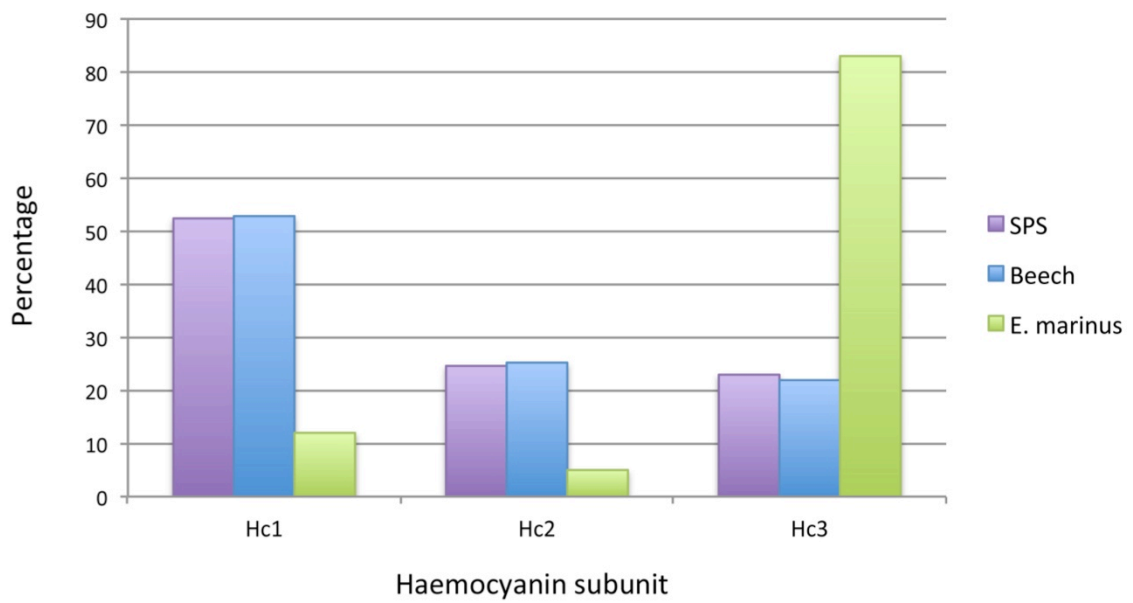
Figure 6.16 Comparison of ESTs contributing to each haemocyanin subunit found in *C. terebrans* and *E. marinus*. Graph shows the percentage of ESTS annotated as haemocyanins representing each subunit. Hc – Haemocyanin orthologue.

The EST distributions suggest that there is a difference in the haemocyanin subunit expression between the wood boring and non-boring amphipods. In the *C. terebrans* libraries haemocyanin subunit 1 is the most abundantly represented, approximately half of all ESTs representing haemocyanins. Haemocyanin transcripts for subunits 2 and 3 in *C. terebrans* were found in similar numbers, each approximately representing a quarter of all ESTs annotated as haemocyanins. In contrast, the *E. marinus* ESTs annotated as haemocyanins are dominated by those representing haemocyanin subunit 3.

6.9 Discussion

Initial analysis of the *C. terebrans* hepatopancreas libraries revealed that approximately 80% of the ESTs in each library assembled to create contigs, leaving relatively few singlets/singletons. The five most abundantly represented contigs are contributed to by approximately 25% of the total ESTs in the libraries, with approximately 50% of all ESTs in the libraries contributing to the 35 most abundantly represented contigs. This is in keeping with findings in other crustaceans that suggest the hepatopancreas produces large quantities of proteins from relatively few genes (Jiang et al., 2009; King et al., 2010).

The EST library derived from *C. terebrans* fed on beech (from now on referred to as the beech library) produced a lower number of ESTs than the library derived from *C. terebrans* fed on SPS (from now on referred to as the SPS library). This could have been due to any number of variations during the preparatory process or during the 454-sequencing run. However, the overall EST distribution appears to be relatively similar between the two *C. terebrans* libraries; certainly they are more similar to each other than to the transcriptome of *L. quadripunctata*.

The most abundant sequences correspond to groups annotated as GH family proteins (discussed in sections 1.2.1 & 5.1), which represent 12 % and 17 % of all ESTs in the SPS and beech libraries respectively, and haemocyanins, which represent 9 % and 6 % of all ESTs in the SPS and beech libraries respectively.

The abundance of ESTs representing haemocyanin contigs is not surprising given its role as an oxygen transporter and its production is one of the key roles of the hepatopancreas (Kusche & Burmester, 2001). *C. terebrans* appears to have the same number of haemocyanin subunits present in the hepatopancreas library derived from the non-boring amphipod *E. marinus* (from now on referred to as the *E. marinus* library). The *E. marinus* library also has a similarly high number of ESTs representing haemocyanins. Furthermore, phylogenetic analysis reveals that

198

haemocyanins in *C. terebrans* are closely related to those of other peracarids. Haemocyanins have been shown to possess both phenoloxidase activity once a conformational change has been induced through the action of detergents and/or salts (Zlateva et al., 1996; Decker et al., 2001), and antimicrobial activity (Nagai et al., 2001). As no prophenoloxidases, laccases or peroxidases were found in the transcriptome sequence of *L. quadripunctata*, King et al. (2010) suggested that the *L. quadripunctata* haemocyanin could be proteolytically converted into active phenoloxidases and hypothesized this as the mechanism employed by *L. quadripunctata* for lignin modification. BLAST analysis of both *C. terebrans* libraries also found no prophenoloxidases, laccases or peroxidases in the hepatopancreas transcriptome. However, this study has shown a correlation between *C. terebrans* haemocyanin and phenoloxidase activity in the hepatopancreas extract (Chapter 5) and haemocyanins could be responsible for lignin modification and affect cellulose crystallinity, but this requires further investigation.

Annotation of the transcriptomic library suggested the presence of genes belonging to twelve GH families (following CAZy nomenclature). The majority of these sequences are annotated as belonging to the GH5, 7 and 9 families, with a large number of ESTs characterised tentatively as belonging to the family GH38/57 in the beech library. Although the proportion of ESTs contributing to GH sequences is broadly similar, the beech and SPS libraries show a difference in the distribution of the transcripts representing particular GH families. In the beech library the GH7 and 9 families appear to be represented by a very similar number of ESTs; however, in the SPS library the GH7 family dominates. Given the strikingly similar levels of ESTs representing specific haemocyanin subunits across both libraries, the altered distribution of ESTs contributing to particular GH families could indicate a difference in the expression of genes involved with cellulose or hemicellulose degradation in response to a different diet. The beech library also shows a higher percentage of overall ESTs contributing to GH family sequences: this could indicate that *C.*

terebrans requires more of the major GH enzymes to degrade a hardwood. To confirm the differences in the expression of GH families observed in the SPS and beech libraries, a comparison of GH expression in animals fed on a variety of diets using qPCR or repeat high-throughput sequencing runs would need to be undertaken, allowing robust statistical analysis to confidently determine expression variation.

Alignment of the GH7 proteins from *C. terebrans* reveals that a highly conserved loop region is absent from GH7E. The subsequent structural predictions of *C. terebrans* GH7B revealed that this loop, in part, encloses the catalytic substrate tunnel and appears to correspond closely to other GH7 exo-like enzymes, including those from *L. quadripunctata* (unpublished observation). The consequence of the deletion of this loop is that the catalytic substrate tunnel is endowed with a more open conformation. It is intriguing that *C. terebrans* appears to possess an additional GH7 protein, which in the absence of this loop, results in a conformation more similar to an endo-like enzyme such as the endoglucanase I from *T. reesei* (PDB ID: 1EG1). The importance of this divergent GH7 for the processing of lignocellulose is not known and requires further investigation, however, it is tempting to speculate that the organism may recruit this additional enzyme when cellulosic substrates of a very closed nature (i.e. without the free-ends that are required by the standard exo-like enzymes) are present. It will be interesting to investigate the relative expression of this cohort of GH7s under the influence of different substrates and feeding regimes.

The sequencing of the hepatopancreatic transcripts of *C. terebrans* and other crustaceans allows for a consideration of the origins and mechanisms that have enabled *C. terebrans* to degrade lignocellulose. The origin of GH family enzymes in *C. terebrans* could be explained by the horizontal transfer of genes from symbiotic microorganisms or vertical transfer of genes from their ancestors. At least two

distinct sequences predicted to encode GH5 proteins were identified which possessed a high similarity to a β-1,4-mananase from the gastropod *Haliotis discus hannai* (Ootsuka et al., 2006). Four sequences were predicted to encode distinct GH7 family proteins, with the BLAST analysis revealing these sequences show a high level of similarity to cellobiohydrolases in the fungi *Cochliobolus carbonum* (Sposato et al., 1995) and *Schizophyllum commune* (Raudaskoski & Uuskallio, 2005). In addition, at least four distinct sequences annotating to the GH9 family were found in the libraries of *C. terebrans*; BLAST analysis of these sequences show highest similarity to β-1,4-endoglucanases from the crayfish *Cherax quadricarinatus* (Byrne et al., 1999). Phylogenetic analysis of GH family genes reveals that *C. terebrans* sequences consistently group with those found in other crustaceans, including *L. quadripunctata,* suggesting vertical transmission from a common ancestor. This does not mean that the genes could not have originated in a microorganism and been horizontally transferred, just that it would not have been a recent transfer event, but rather an event that occurred in the common ancestor of the crustaceans used for this analysis.

The apparent lack of resident bacteria or fungi in the hepatopancreas of *C. terebrans* (Chapter 4) and the phylogenetic analyses suggesting no recent horizontal transfer of cellulolytic genes from microorganisms fails to answer how *C. terebrans* has evolved a cellulolytic capacity. There are many mechanisms that may facilitate the evolution of this capacity, such as point mutations. This could result in amino acid substitutions that may alter the protein's tertiary structure and thereby modify its function. The phylogenetic analysis reveals some clearly orthologous GH sequences have diverged in sequence. The implications of these sequence changes need to be investigated through structural analysis and functional assays.

A further mechanism which may have enabled the evolution of *C. terebrans* capacity for lignocellulolytic degradation is the modification of gene regulation (Romero et

al., 2012), such that the genes involved in lignocellulolytic degradation are up-regulated. The apparent difference in the distribution of ESTs corresponding to GH families in *C. terebrans* and *L. quadripunctata* (King et al., 2010) in comparison to that of the *E. marinus* library suggests that this mechanism may have been a contributing factor. Although *E. marinus* clearly has genes representing GH9 families, no GH9 contig has enough contributing ESTs to place it in the list of the 20 most abundantly expressed genes, however, it should be noted that, as yet, there is no information on the activity levels of these respective enzymes. To confirm this finding comparative qPCR analysis or further high-throughput sequencing experiments would need to be performed, incorporating a greater range of amphipod species and other species from the Cheluridae family.

Gene duplication followed by sequence divergence is another potential mechanism facilitating the evolution of modified function (Taylor & Raes, 2004). If this mechanism contributed to the evolution of cellulolytic capacity in *C. terebrans*, we would expect an expansion of particular gene families. Although several GH gene sequences found in *C. terebrans* possessed a clear orthologue in the *E. marinus* library, some *C. terebrans* sequences, such as, GH5A, GH7A, B and D and GH9A and E, appear to lack a clearly identifiable corresponding *E. marinus* orthologue. The lack of an obvious orthologue may be the result of its absence from the *E. marinus* genome, or of the gene having such a low level of expression in *E. marinus* that it does not appear in the assembled transcriptome. To truly distinguish between massive down-regulation or complete absence of orthologous genes requires the full genome sequencing and comparison of closely related amphipod genomes.

The extent to which all these mechanisms have contributed to the lignocellulolytic capacity in *C. terebrans* remains unclear. Indeed, additional post-transcriptional regulatory mechanisms, such as alternative splicing (Keren et al., 2010), translational regulations by microRNAs (Niwa & Slack, 2007) and post-translational protein

modifications (Seo & Lee, 2004) could also have been important for the evolution of modified function and require further study.

The transcriptome of the *C. terebrans* hepatopancreas has provided information on the types of genes present and their relative abundances. The analysis described in this study is preliminary and further comparisons with fellow wood boring crustaceans and closely related non-boring amphipods promises an even more confident identification of the genes critical for lignocellulosic degradation. Further study would involve the cloning and functional characterisation of critical proteins and experiments involving probing protein-protein interactions *e.g.*, protein immunoprecipitations, analytical ultracentrifugation, EMSA, *etc.* to establish whether any complexes form between the lignocellulolytic enzymes, analogous to the cellulosomes documented in bacteria (Bayer, 2004*;* Lamed et al., 1983).

7 General discussion

Initial morphological analysis of members of the family Cheluridae placed all known species into a single genus (Allman, 1847). Since then, on the basis of further analysis of morphology, the four species have been separated in to three morphologically distinct genera (Barnard, 1959). However, since *C. terebrans* has not been incorporated into any molecular phylogenetic analyses, it was not known whether a phylogeny based on gene sequence comparisons would support the taxonomy developed using morphological characteristics. This study found regions of the 18S and ITS ribosomal genes sequences provided consistent phylogenetic trees reflecting current taxonomy. In contrast, phylogenetic analyses performed using mitochondrial sequences, including the CO1 'barcoding region', did not produce trees consistent with morphology, the ribosomal sequences or each other. Such findings suggest the mitochondrial genes in some *C. terebrans* populations are undergoing, or have undergone, unusually fast rates of evolution. These findings support other studies that have highlighted difficulties using the CO1 gene as a phylogenetic marker (e.g. Astrin, 2006). Further analysis of chelurid samples from multiple locations is required to establish the extent of mitochondrial gene divergence in this family and whether they correlate to particular influences. Given the findings in this study, it would appear that ribosomal sequences provide a more useful tool for understanding of the relationships between chelurid populations over large geographical distance and for species identification. However, the CO1 sequence may be more effective when comparing *C. terebrans* populations in a more localised study.

Allman (1847) described *C. terebrans* as truly xylophagous. However, due to its slow attack of wood (Kühne & Becker, 1964) and apparent lack of gut resident symbionts (Boyle & Mitchell, 1978), so often present in wood degrading animals, doubts were cast on the true diet of *C. terebrans*. The wood degrading ability of *L. quadripunctata* was undisputed due to the voraciousness with which it attacks wood, despite it also

apparently lacking any resident symbionts. This study has shown that *C. terebrans* do indeed ingest wood and, although not thought possible by some (Kühne & Becker, 1964), are capable of surviving for more than a few generations in the absence of *L. quadripunctata*. Scanning electron microscopy was used to examine the digestive tract of *C. terebrans* and found no evidence of bacterial cells in the hepatopancreas or gut. Furthermore, quantitative real-time PCR confirmed the absence of any substantial symbiotic extra- or intracellular bacteria resident in the digestive tract by revealing very low levels of bacterial 16S gene sequences in comparison to the symbiont-containing isopod *Porcellio scaber*.

Barnard (1955) suggested that the 'wood digestibility of limnoriids may be less efficient than that of chelurids and more wood needs to be consumed by the former'. Indeed, this study supports findings that the transition time for food in *C. terebrans* is very slow, less than 10 faecal pellets a day, in comparison to *L. quadripunctata,* which produces around 80-100 faecal pellets a day (Borges, 2009). Further study is needed to determine the extent of wood degradation as it passes through the gut of *C. terebrans* and could involve sugar component analysis of food mass in the proventriculus, hindgut and faecal pellets. FT-IR spectroscopy and X-ray diffraction studies could also aid in the measurement of changes to wood chemistry and structure after digestion. The SEM examination of the gross digestive system and fine structure of the proventriculus found, with the exception of robust lateralia, few morphological adaptations to accommodate this unusual diet. Transmission electron microscopy examination will provide greater detail of the digestive tract, offering possible functional insights of individual cells in the hepatopancreas.

The transcriptomic libraries and enzymatic assays suggest that *C. terebrans* possess a considerable repertoire of endogenous enzymatic capabilities potentially useful for the digestion of wood. Mass spectrometry analysis on gel regions presenting high mono- and diphenol oxidase activity identified several proteins belonging to the Glycosyl

Hydrolase (GH) families and haemocyanin. The identification of haemocyanins in these gel regions is consistent with phenoloxidase activity presented by haemocyanins of other crustaceans (Decker et al., 2001, 2007; Nagai et al., 2001). The identification of GH family proteins is more difficult to interpret, however, it would be interesting to speculate that the presence of GH proteins in these regions is due to the formation of protein complexes, analogous to the cellulosomes documented in bacteria (Lamed et al., 1983; Bayer, 2004). It is also possible that the appearance of the haemocyanins and GH family proteins in the examined gel regions following the phenol oxidase assay is simply a reflection of their high abundance. Nevertheless, the possibility of complexes as a mechanism for efficient wood degradation in *C. terebrans* requires further investigation.

The degradation of wood also requires the modification of lignin. A lignin-modification mechanism suggested for *L. quadripunctata* involves the cleavage of haemocyanins to induce phenoloxidase activity, a capacity seen in the haemocyanins of other crustaceans (e.g. Decker & Tuczek, 2000). The potential involvement of the haemocyanins in a range of activities in *C. terebrans* requires further investigation. This includes the inducement of potential lignin degrading and antimicrobial properties resulting from C- and N-terminal cleavages respectively (e.g. Decker & Tuczek, 2000; Nagai et al., 2001). The presence of highly expressed contigs with associated functions involving ferritin, hydrogen peroxide and reactive oxygen species in the transcriptomic library highlights Fenton reactions as another possible lignin degrading mechanism in *C. terebrans*. Fenton reactions, utilised by white-rot fungi to degrade lignin, involve the production of reactive oxygen species using hydrogen peroxide and ferrous iron (H_2O_2 + $Fe^{2+} \rightarrow {}^\bullet OH + OH^- + Fe^{3+}$) (Gomez-Toribio *et al.*, 2009). Indications of the presence of components critical for the Fenton reaction, which can also mediate changes to cellulose crystallinity (Arantes et al., 2011), in the transcriptome are intriguing and

could be investigated by the injection of dyes in to the digestive tract of *C. terebrans* which fluoresce once oxidized by reactive oxygen species.

The *C. terebrans* transcriptomic libraries have helped highlight similarities and differences between wood boring and non-boring crustaceans, opening new avenues of investigation into the genes important for the digestion of lignocellulose. These libraries will also improve our understanding of the evolution of the lignocellulosic capacity when combined with further analyses and comparison with related species. This study of the transcriptomic library of *C. terebrans* is preliminary, and further work is needed to gain a full understanding of the genes important for digestion in *C. terebrans*. Initial future work would include the cloning of genes encoding enzymes of interest and elucidation of their specific action and level of activity. In situ hybridisation could be utilised to investigate the expression of target genes and Western blot experiments could verify protein expression. Furthermore, immunoprecipitation experiments could be used to identify whether the enzymes involved in lignocellulose digestion form complexes.

There is now considerable interest in lignocellulose degrading organisms as a novel source of enzymes that can be commercially produced by biotechnological means for the production of sustainable liquid biofuels from plant biomass (Carroll & Someville, 2009; Wong, 2009; Horn et al., 2012). The search for enzyme systems capable of obtaining sugars for fermentation has so far overlooked the wood boring amphipod *C. terebrans*. This study has provided new insights into our understanding of wood digestion in *C. terebrans* and highlighted this wood borer as a potential source for lignocellulose degrading enzymes, making *C. terebrans* relevant to current scientific interest.

8 References

Agrawal, B. B. L., & Goldstein, I. J. (1967). Protein-carbohydrate interaction: VI. Isolation of concanavalin a by specific adsorption on cross-linked dextran gels. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *147*(2), 262-271.

Agrawal, **V. P**. (1965). Feeding appendages and the digestive system of Gammarus pulex. *Acta Zoologica, 46*(1-2), 67-81.

Allman, **G. J.** (1847). On Chelura terebrans, Philippi, an amphipodous crustacean destructive to submarine timberworks. *Annals and Magazine of Natural History, Ser, 1*(19), 361-370.

Altschul, S. F., Warren Gish, Webb Miller, Eugene W. Myers, David J. Lipman. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology, 215*, 403-410.

Appeltans, W; Bouchet, P; Boxshall, GA; De Broyer, C; de Voogd, NJ; Gordon, DP; Hoeksema, BW; Horton, T; Kennedy, M; Mees, J; Poore, GCB; Read, G; Stöhr, S; Walter, TC; Costello, MJ (eds) (2013). WoRMS - World Register of Marine Species. *Accessed at* http://www.marinespecies.org *on 2013-01-27.*

Arantes, V., Milagres, A. M., Filley, T. R., & Goodell, B. (2011). Lignocellulosic polysaccharides and lignin degradation by wood decay fungi: the relevance of nonenzymatic Fenton-based reactions. *Journal of industrial microbiology & biotechnology*, *38*(4), 541-555.

Astrin, J. J., Huber, B. A., Misof, B., & Klütsch, C. F. C. (2006). Molecular taxonomy in pholcid spiders (Pholcidae, Araneae): evaluation of species identification methods using CO1 and 16S rRNA. *Zoologica Scripta, 35*(5), 441-457.

Atalla, R. H. (2005). The role of the hemicelluloses in the nanobiology of wood cell walls: a systems theoretic perspective. In *Proceedings of the hemicelluloses workshop* (pp. 37-57).

Baird, H. P., Miller, K. J., & Stark, J. S. (2011). Evidence of hidden biodiversity, ongoing speciation and diverse patterns of genetic structure in giant Antarctic amphipods. *Molecular Ecology, 20*(16), 3439-3454.

Ballard, J. W. O., & Whitlock, M. C. (2003). The incomplete natural history of mitochondria. *Molecular Ecology, 13*(4), 729-744.

Barnard, J. L. (1959). Generic partition in the amphipod family Cheluridae, marine wood borers AND Liljeborgiid amphipods of southern Californian coastal bottoms, with a revision of the family. *Pacific Naturalist, 1*(3-4), 1-28.

Barnard, L. J. (1955). The occurrence of Chelura terebrans Philippi in Los Angeles harbor *Essays in the Natural History in Honour of Captain Allan Hancock* (pp. 87-92).

Barnett, J. R., & Bonham, V. A. (2004). Cellulose microfibril angle in the cell wall of wood fibres. *Biological Reviews, 79*(02), 461-472.

Bayer, E. A., Belaich, J.-P., Shoham, Y., & Lamed, R. (2004). The Cellulosomes: Multienzyme Machines for Degradation of Plant Cell Wall Polysaccharides. *Annual Review of Microbiology, 58*(1), 521-554.

Bazin, E., Glémin, S., & Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science, 312*(5773), 570-572.

Beg, Q. K., Kapoor, M., Mahajan, L., & Hoondal, G. S. (2001). Microbial xylanases and their industrial applications: a review. *Applied Microbiology and Biotechnology, 56*(3), 326-338.

Björdal, C. G. (2012). Evaluation of microbial degradation of shipwrecks in the Baltic Sea. *International Biodeterioration & Biodegradation*, *70*, 126-140.

Boerjan, W., Ralph, J., & Baucher, M. (2003). Lignin biosynthesis. *Annual Review of Plant Biology, 54*(1), 519-546.

Bourdillon, A. (1958). La dissemination des Crustaces xylophage *Limnoria tripunctata* Menzies et *Cheluraterebrans* Philippi. . *Annals of Biology, 34*, 437 - 463.

Boyle, P. J., & Mitchell, R. (1978). Absence of microorganisms in crustacean digestive tracts. *Science, 200*(4346), 1157-1159.

Boyle, P. J., & Mitchell, R. (1980). Interaction between microorganisms and wood boring crustaceans. In T. S. Oxley, D. Allsopp & G. Becker (Eds.), *Biodeterioration* (pp. 179-186). London: Pitman.

Bucklin, A., Steinke, D., & Blanco-Bercial, L. (2011). DNA barcoding of marine metazoa. *Annual Review of Marine Science, 3*, 471-508.

Buhay, J. E. (2009). "COI-Like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. *Journal of Crustacean Biology, 29*(1), 96-110.

Byrne, K. A., Lehnert, S. A., Johnson, S. E., & Moore, S. S. (1999). Isolation of a cDNA encoding a putative cellulase in the red claw crayfish Cherax quadricarinatus. [doi: DOI: 10.1016/S0378-1119(99)00396-0]. *Gene, 239*(2), 317-324.

Böttger, E. C. (1990). Frequent contamination of Taq polymerase with DNA. *Clin Chem, 36*(6), 1258-1259.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. [Article]. *Nucleic Acids Research, 37*, D233-D238.

Carroll, A., & Somerville, C. (2009). Cellulosic biofuels. *Annual review of plant biology, 60*, 165-182.

Case, R. J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W. F., & Kjelleberg, S. (2007). Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol, 73*(1), 278-288.

Chu, K., Xu, M., & Li, C. (2009). Rapid DNA barcoding analysis of large datasets using the composition vector method. *BMC bioinformatics, 10* (Suppl 14), S8.

Cohan, F. (2002). What are bacterial species? [Review]. *Annual Review of Microbiology, 56*, 457-487.

Coleman, C. O. (1991). Comparative fore-gut morphology of antarctic amphipoda (crustacea) adapted to different food sources. *Hydrobiologia, 223*, 1-9.

Coleman, C. O. (1992). Foregut morphology of Amphipoda (Crustacea). An example of its relevance for systematics. *Ophelia, 36*(2), 135-150.

Coleman, C. O. (1994). Comparative anatomy of the alimentary canal of hyperiid amphipods. *Journal of Crustacean biology*, 346-370.

Cordaux, R., Michel-Salzat, A., & Bouchon, D. (2001). Wolbachia infection in crustaceans: novel hosts and potential routes for horizontal transmission. *Journal of Evolutionary Biology, 14*(2), 237-243.

Cragg, S. M., & Daniel, G. (1992). Chelura terebrans (Crustacea: Amphipoda) is capable of degrading wood independently of its associate Limnoria (Doc. No. IRG/WP/4180-92 ed., pp. 1-10). *International Research Group on Wood Preservation*: *International Research Group on Wood Preservation*.

Cragg, S. M., Pitman, A. J., & Henderson, S. M. (1999). Developments in the understanding of the biology of marine wood boring crustaceans and in methods of controlling them. *International Biodeterioration & Biodegradation, 43*(4), 197-205.

Dahllöf, I., Baillie, H., & Kjelleberg, S. (2000). rpoB-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. *Appl Environ Microbiol, 66*(8), 3376-3380.

Daniel, G., Nilsson, T., & Cragg, S. (1991). Limnoria lignorum ingest bacterial and fungal degraded wood. *Holz als Roh- und Werkstoff, 49*(12), 488-490.

Darby, A. C., Cho, N.-H., Fuxelius, H.-H., Westberg, J., & Andersson, S. G. E. (2007). Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *23*(10), 511-520.

Davolos, D., Pavesi, L., Accordi, F., & De Matthaeis, E. (2010). Morphology and histology of the mouthparts and gut system of Macarorchestia remyi (Schellenberg, 1950)(Amphipoda, Talitridae). *Zool. baetica, 21*, 151-178.

Decker, H., Hellmann, N., Jaenicke, E., Lieb, B., Meissner, U., & Markl, J. (2007). Minireview: Recent progress in hemocyanin research. *Integrative and comparative biology, 47*(4), 631-644.

Decker, H., Ryan, M., Jaenicke, E., & Terwilliger, N. (2001). SDS-induced Phenoloxidase Activity of Hemocyanins fromLimulus polyphemus, Eurypelma californicum, andCancer magister.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research, 36*(suppl 2), W465-W469.

Dharmaraj, K., & Nair, N. B. (1980). Wood-boring organisms in relation to aquaculture along the coasts of India. Paper presented at the *Symposium on Coastal Aquaculture*, Cochin India, 12-18. January 1980.

Distel, D. L. (2003). The biology of marine wood boring bivalves and their bacterial endosymbionts *Wood Deterioration and Preservation* (Vol. 845, pp. 253-271).

Distel, D. L., Morrill, W., MacLaren-Toussaint, N., Franks, D., & Waterbury, J. (2002). Teredinibacter turnerae gen. nov., sp. nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-proteobacterium isolated from the gills of wood-boring molluscs (Bivalvia: Teredinidae). *Int J Syst Evol Microbiol, 52*(6), 2261-2269.

Distel, D. L., & Roberts, S. J. (1997). Bacterial Endosymbionts in the Gills of the Deep-Sea Wood-Boring Bivalves Xylophaga atlantica and Xylophaga washingtona. *Biol Bull, 192*(2), 253-261.

Eaton, R. A. (1986). Preservation of wood in the sea. *The Biology of Marine Fungi*, 30, 355.

Eaton, R. A., & Hale, M. D. C. (1993). *Wood: decay, pests and protection*: Chapman and Hall Ltd.

Eichhorn, S. J., Sirichaisit, J., & Young, R. J. (2001). Deformation mechanisms in cellulose fibres, paper and wood. *Journal of materials science, 36*(13), 3129-3135.

Ellegren, H. (2008). Sequencing goes 454 and takes large-scale genomics into the wild. *Molecular Ecology, 17*(7), 1629-1631.

Folmer, O., Hoeh, W. R., Black, M. B., & Vrijenhoek, R. C. (1994). Conserved primers for PCR amplification of mitochondrial DNA from different invertebrate phyla. *Molecular Marine Biology and Biotechnology, 3*, 294-299.

Frézal, L., & Leblois, R. (2008). Four years of DNA barcoding: current advances and prospects. *Infection, Genetics and Evolution, 8*(5), 727-736.

Gasparic, J., Svobodova, D., & Pospisilova, M. (1977). Identification of organic compounds. Part LXXXVI. Investigation of the color reaction of phenols with the MBTH (3-methyl-2-benzothiazolinone hydrazone hydrochloride) reagent. *Mikrochim Acta*, *1*, 241-250.

Ginsburger-Vogel, T. (1991). Intersexuality in Orchestia mediterranea Costa, 1853, and Orchestia aestuarensis Wildish, 1987 (Amphipoda): A Consequence of Hybridization or Parasitic Infestation? *Journal of Crustacean Biology*, 530-539.

Gomez-Toribio, V., Garcia-Martin, A. B., Martinez, M. J., Martinez, A. T., & Guillen, F. (2009). Induction of Extracellular Hydroxyl Radical Production by White-Rot Fungi through Quinone Redox Cycling. *Appl. Environ. Microbiol., 75*(12), 3944-3953.

Goto, M., Ando, S., Hachisuka, Y., & Yoneyama, T. (2006). Contamination of diverse nifH and nifH-like DNA into commercial PCR primers. *FEMS microbiology letters, 246*(1), 33-38.

Gouet, P., Courcelle, E., & Stuart, D. I. (1999). ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics, 15*(4), 305-308.

Griekspoor, A., & Groothuis, T. (2005). 4Peaks: a program that helps molecular biologists to visualize and edit their DNA sequence files v1. 7.

Gukasyan, G. S. (1999). Effect of lignin on growth and tyrosinase activity of fungi from the genus Aspergillus. *Biochemistry. Biokhimiia, 64*(2), 223.

Hames, C. A. C., & Hopkin, S. P. (1989). The structure and function of the digestive system of terrestrial isopods. *Journal of Zoology, 217*(4), 599-627.

Hansen, N. M. L., & Plackett, D. (2008). Sustainable films and coatings from hemicelluloses: A review. *Biomacromolecules, 9*(6), 1493-1505.

Harrison, F. W. (Ed.). (1992). *Microscopic Anatomy of Invertebrates* (Vol. 9).

Hassall, M., & Jennings, J. B. (1975). Adaptive features of gut structure and digestive physiology in the terrestrial isopod philoscia muscorum (scopoli) 1763. *Biol Bull, 149*(2), 348-364.

Hebert, P. D. N., Cywinska, A., & Ball, S. L. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270*(1512), 313-321.

Hill, C. L. R., & Kofoid, C. A. (1927). *Marine borers and their relation to marine construction on the Pacific coast: being the final report of the San Francisco Bay Marine Piling Committee*: The Committee.

Hochman, H. (1982). 7. Degradation and Protection of Wood from Marine Organisms. *Wood Deterioration and Its Prevention by Preservative Treatments: Degradation and Protection of Wood, 1*, 247.

Horn, S. J., Vaaje-Kolstad, G., Westereng, B., & Eijsink, V. G. H. (2012). Novel enzymes for the degradation of cellulose. *Biotechnology for Biofuels, 5*(45).

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome research, 9*(9), 868-877.

Huang, X., & Miller, W. (1991). A time-efficient linear-space local similarity algorithm. *Advances in Applied Mathematics, 12*(3), 337-357.

Hughes, M. S., Beck, L. A., & Skuce, R. A. (1994). Identification and elimination of DNA sequences in Taq DNA polymerase. *Journal of clinical microbiology, 32*(8), 2007-2008.

Hurst, G. D. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences, 272*(1572), 1525-1534.

Icely, J. D., & Nott, J. A. (1984). On the morphology and fine structure of the alimentary canal of Corophium volutator (Pallas)(Crustacea: Amphipoda). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 49-78.

Ioelovich, M. (2008). Cellulose as a nanostructured polymer: a short review. *BioResources, 3*(4), 1403-1418.

Jensen, K. A., Bao, W., Kawai, S., Srebotnik, E., & Hammel, K. E. (1996). Manganese-Dependent Cleavage of Nonphenolic Lignin Structures by Ceriporiopsis subvermispora in the Absence of Lignin Peroxidase. *Applied and environmental microbiology, 62*(10), 3679-3686.

Jiang, H., Cai, Y. M., Chen, L. Q., Zhang, X. W., Hu, S. N., & Wang, Q. (2009). Functional Annotation and Analysis of Expressed Sequence Tags from the Hepatopancreas of Mitten Crab (Eriocheir sinensis). *Marine Biotechnology, 11*(3), 317-326.

Jurgens, J. A., & Blanchette, R. A. (2005). Characterization of wood destroying microorganisms in archaeological woods from marine environments.

Keith, D. E. (1974). A comparative study of the digestive tracts of Caprella equilibra Say and Cyamus boopis Lütken (Amphipoda, Caprellidea). *Crustaceana*, 127-132.

Keren, H., Lev-Maor, G., & Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics*, *11*(5), 345-355.

King, A. J., Cragg, S. M., Li, Y., Dymond, J., Guille, M. J., Bowles, D. J., et al. (2010). Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proceedings of the National Academy of Sciences of the United States of America, 107*(12), 5345-5350.

Kirby, R. (2005). Actinomycetes and lignin degradation. *Advances in applied microbiology, 58*, 125-168.

Kleinert, M., & Barth, T. T. (2008). Towards a Lignincellulosic Biorefinery: Direct One-Step Conversion of Lignin to Hydrogen-Enriched Biofuel. *Energy & Fuels*, 22, 1371–1379

Kobayashi, H., Hatada, Y., Tsubouchi, T., Nagahama, T., & Takami, H. (2012). The Hadal Amphipod Hirondellea gigas possessing a unique cellulase for digesting wooden debris buried in the deepest seafloor. *PLoS One, 7*(8), e42727.

Kobusch, W. (1998). The foregut of the Mysida (Crustacea, Peracarida) and its phylogenetic relevance. *Philosophical Transactions of the Royal Society B: Biological Sciences, 353*(1368), 559.

Kofoid, C. A., & Miller, R. C. (1927). Biological section. In C. L. Hill & C. A. Kofoid (Eds.), *Marine borers and their relation to marine construction on the Pacific Coast. Final Report of the Sand Francisco Bay marine piling comittee.* (pp. 188-343).

Kostanjsek, R., Lapanje, A., Rupnik, M., Strus, J., Drobne, D., & Avgustin, G. (2004). Anaerobic bacteria in the gut of terrestrial isopod Crustacean Porcellio scaber. *Folia Microbiol (Praha), 49*(2), 179-182.

Kostanjsek, R., Strus, J., Drobne, D., & Avgustin, G. (2004). 'Candidatus Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod Porcellio scaber (Crustacea: Isopoda). *Int J Syst Evol Microbiol, 54*(Pt 2), 543-549.

Kostanjšek, R., Milatovič, M., & Štrus, J. (2010). Endogenous origin of endo-β-1,4-glucanase in common woodlouse <i>Porcellio scaber</i> (Crustacea, Isopoda). *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology, 180*(8), 1143-1153.

Kusche, K., & Burmester, T. (2001). Molecular cloning and evolution of lobster hemocyanin. *Biochemical and biophysical research communications, 282*(4), 887-892.

Kühne, H., & Becker, G. (1964). Der Holz-Flohkrebs Chelura terebrans Philippi (Amphipoda, Cheluridae). *Zeitschrift für angewante Zoologie, Beiheft 1*, 1-141.

Lamed, R., Setter, E., Kenig, R., & Bayer, E. A. (1983). *Cellulosome: a discrete cell surface organelle of Clostridium thermocellum which exhibits separate antigenic, cellulose-binding and various cellulolytic activities.* Paper presented at the Biotechnol. Bioeng. Symp.;(United States).

Lane, D. J. (1991). 16S/23S rRNA sequencing. In E. Stakebrandt & M. Goodfellow (Eds.), *Nucleic Acid Techniques in Bacterial Systematics.* (pp. 133). Chichester, New York.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*(21), 2947-2948.

Lin, L., Yan, R., Liu, Y., & Jiang, W. (2010). In-depth investigation of enzymatic hydrolysis of biomass wastes based on three major components: Cellulose, hemicellulose and lignin. *Bioresource technology, 101*(21), 8217-8223.

Lopez-Anido, R., Michael, A. P., Goodell, B., & Sandford, T. C. (2004). Assessment of Wood Pile Deterioration due to Marine Organisms. *Journal of Waterway, Port, Coastal, and Ocean Engineering, 130*(2), 70-76.

Luo, W., Vrijmoed, L. L. P., & Jones, E. B. (2005). Screening of marine fungi for lignocellulose-degrading enzyme activities. *Botanica Marina, 48*(5), 379-386.

Maki, M., Leung, K. T., & Qin, W. (2009). The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *International journal of biological sciences, 5*(5), 500.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature, 437*(7057), 376-380.

Martin, A. L. (1964). The alimentary canal of marinogammarus obtusatus (Crustacea, Amphipoda). *Proceedings of the Zoological Society of London, 143*(4), 525-544.

Martin, M. M. (1983). Cellulose digestion in insects. *Comparative Biochemistry and Physiology Part A: Physiology, 75*(3), 313-324.

Martin, M. M. (1987). *Invertebrate-microbial interactions. Ingested fungal enzymes in arthropod biology*: Cornell University Press.

Martin, M. M. (1992). The evolution of insect-fungus associations - from contact to stable symbiosis. *American Zoologist, 32*(4), 593-605.

Mayer, A. M., & Staples, R. C. (2002). Laccase: new functions for an old enzyme. [doi: DOI: 10.1016/S0031-9422(02)00171-1]. *Phytochemistry, 60*(6), 551-565.

Mekhanikova, I. V. (2010). Morphology of mandible and lateralia in six endemic amphipods (Amphipoda, Gammaridea) from Lake Baikal, in relation to feeding. *Crustaceana, 83*(7), 865-887.

Miller, R. C. (1926). Ecological relations of marine wood-boring organisms in San Francisco Bay. *Ecology, 7*(3), 247-254.

Mutwil, M., Debolt, S., & Persson, S. (2008). Cellulose synthesis: a complex complex. *Current opinion in plant biology, 11*(3), 252-257.

Mühl, H., Kochem, A. J., Disqué, C., & Sakka, S. G. (2010). Activity and DNA contamination of commercial polymerase chain reaction reagents for the universal 16S rDNA real-time polymerase chain reaction detection of bacterial pathogens in blood. *Diagn Microbiol Infect Dis, 66*(1), 41-49.

Nagai, T., Osaki, T., & Kawabata, S.-i. (2001). Functional Conversion of Hemocyanin to Phenoloxidase by Horseshoe Crab Antimicrobial Peptides. *Journal of Biological Chemistry, 276*(29), 27166-27170.

Neigel, J., Domingo, A., & Stake, J. (2007). DNA barcoding as a tool for coral reef conservation. *Coral Reefs, 26*(3), 487-499.

Nellaiappan, K., & Banu, M. J. (1991). Demonstration of monophenoloxidase activity of tyrosinase after electrophoresis. *Biotechnic & histochemistry*, *66*(3), 125-130.

Neuhauser, E., & Hartenstein, R. (1976). On the presence of *O*-demethylase activity in invertebrates. *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology, 53*(1), 37-39.

Niwa, R., & Slack, F. J. (2007). The evolution of animal microRNA function. *Current opinion in genetics & development*, *17*(2), 145-150.

Ootsuka, S., Saga, N., Suzuki, K., Inoue, A., & Ojima, T. (2006). Isolation and cloning of an endo-β-1, 4-mannanase from Pacific abalone *Haliotis discus hannai*. *Journal of biotechnology, 125*(2), 269-280.

Panshin, A. J., & Zeeuw, C. (1980). *Textbook of wood technology*: McGraw-Hill Book Co.

Peixoto, R. S., da Costa Coutinho, H. L., Rumjanek, N. G., Macrae, A., & Rosado, A. S. (2002). Use of rpoB and 16S rRNA genes to analyse bacterial diversity of a tropical soil using PCR and DGGE. *Lett Appl Microbiol, 35*(4), 316-320.

Pereira, H. (1988). Variability in the chemical composition of plantation eucalypts (Eucalyptus globulus Labill.). *Wood and fiber science, 20*(1), 82-90.

Popham, J. D., & Dickson, M. R. (1973). Bacterial associations in the teredo Bankia australis (Lamellibranchia: Mollusca). *Marine Biology, 19*(4), 338-340.

Quiroz-Castañeda, R. E., & Folch-Mallol, J. L. (2011). Proteínas que remodelan y degradan la pared celular vegetal: perspectivas actuales. *Biotecnología Aplicada, 28*(4), 205-215.

Rantsiou, K., Comi, G., & Cocolin, L. (2004). The rpoB gene as a target for PCR-DGGE analysis to follow lactic acid bacterial population dynamics during food fermentations. [Article]. *Food Microbiology, 21*(4), 481-487.

Ray, D. L., & Julian, J. R. (1952). Occurrence of Cellulase in Limnoria. [10.1038/169032a0]. *Nature, 169*(4288), 32-33.

Rehm, E. J., Hannibal, R. L., Chaw, R. C., Vargas-Vila, M. A., & Patel, N. H. (2009). Injection of Parhyale hawaiensis blastomeres with fluorescently labeled tracers. *Cold Spring Harbor Protocols, 2009*(1), pdb. prot5128.

Richter, S., & Scholtz, G. (2001). Phylogenetic analysis of the Malacostraca (Crustacea). *Journal of Zoological Systematics and Evolutionary Research, 39*(3), 113-136.

Romero, I. G., Ruvinsky, I., & Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics, 13*(7), 505-516.

Ronaghi, M. (1998). *Pyrosequencing: a tool for sequence-based DNA analysis*: Tekniska högsk.

Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol, 132*(3), 365-386.

Rubinoff, D., Cameron, S., & Will, K. (2006). A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *Journal of Heredity, 97*(6), 581-594.

Ruiz-Dueñas, F. J., & Martínez, Á. T. (2009). Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microbial Biotechnology, 2*(2), 164-177.

Sakamoto, K., Touhata, K., Yamashita, M., Kasai, A., & Toyohara, H. (2007). Cellulose digestion by common Japanese freshwater clam Corbicula japonica. *Fisheries Science, 73*(3), 675-683.

Santos, S. R., & Ochman, H. (2004). Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environmental Microbiology, 6*(7), 754-759.

Sarma-Rupavtarm, R., Ge, Z., Schauer, D., Fox, J., & Polz, M. (2004). Spatial distribution and stability of the eight microbial species of the altered Schaedler flora in the mouse gastrointestinal tract. [Article]. *Applied and Environmental Microbiology, 70*(5), 2791-2800.

Schmidt, O. (2006). *Wood and tree fungi: biology, damage, protection, and use*: Springer.

Schmitz, E. H., & Scherrey, P. M. (1983). Digestive anatomy of hyalella-azteca (*Crustacea,* Amphipoda). *Journal of Morphology, 175*(1), 91-100.

Schultz, T. W., & Kennedy, J. R. (1976). The fine structure of the digestive system of Daphnia pulex (Crustacea: Cladocera). *8*(3), 479-490.

Seo, J., & Lee, K. J. (2004). Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *Journal of biochemistry and molecular biology*, *37*(1), 35-44.

Shyamsundari, K., & Rao, K. H. (1976). Studies on the Alimentary Canal of Amphipods. Excretory Caeca. *Crustaceana, 31*(2), 190-192.

Sidjanski, S., Mathews, G. V., & Vanderberg, J. P. (1997). Electrophoretic Separation and Identification of Phenoloxidases in Hemolymph and Midgut of Adult Anopheles stephensi Mosquitoes. *The Journal of Parasitology, 83*(4), 686-691.

Silkie, S. S., & Nelson, K. L. (2009). Concentrations of host-specific and generic fecal markers measured by quantitative PCR in raw sewage and fresh animal feces. *Water Res, 43*(19), 4860-4871.

Smith, M. A., Bertrand, C., Crosby, K., Eveleigh, E. S., Fernandez-Triana, J., Fisher, B. L., et al. (2012). Wolbachia and DNA Barcoding Insects: Patterns, Potential, and Problems. *PloS one, 7*(5), e36514.

Somerville, C. (2006). Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol., 22*, 53-78.

Sousa, L. G., Cuartas, E. I., & Petriella, A. M. (2005). Fine structural analysis of the epithelial cells in the hepatopancreas of Palaemonetes argentinus (Crustacea, Decapoda, Caridea) in intermoult. *Biocell, 29*(1), 25-31.

Sposato, P., Ahn, J. H., & Walton, J. D. (1995). Characterization and disruption of a gene in the maize pathogen Cochliobolus carbonum encoding a cellulase lacking a cellulose binding domain and hinge region. *MPMI-Molecular Plant Microbe Interactions, 8*(4), 602-609.

Storch, V. (1987). Microscopic anatomy and ultrastructure of the stomach of porcellio-scaber (crustacea, isopoda). *Zoomorphology, 106*(5), 301-311.

Strus, J. (1987). The effects of starvation on the structure and function of the hepatopancreas in the isopod Ligia italica. *Investigacion Pesquera, 51*.

Strus, J., & Storch, V. (2004). Comparative electron microscopic study of the stomach of Orchestia cavimana and Arcitalitrus sylvaticus (crustacea: Amphipoda). *J Morphol, 259*(3), 340-346.

Tanimura, A., Liu, W., Yamada, K., Kishida, T., & Toyohara, H. (2012). Animal cellulases with a focus on aquatic invertebrates. *Fisheries Science*, 1-13.

Taylor, J. S., & Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet., 38*, 615-643.

Terry, R. S., Smith, J. E., Sharpe, R. G., Rigaud, T., Littlewood, D. T. J., Ironside, J. E., et al. (2004). Widespread vertical transmission and associated host sex–ratio distortion within the eukaryotic phylum Microspora. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 271*(1550), 1783-1789.

Thiem, E. (1942). Untersuchungen über den Darmkanal und die Nahrungsaufnahme von *Synurella ambulans* (Fr. Müller) (Crust. Amph.).-*Zeitschrift für Morphologie und Ökologie der Tiere 38*, 63- 79.

Torres, E., Bustos-Jaimes, I., & Le Borgne, S. (2003). Potential use of oxidative enzymes for the detoxification of organic pollutants. *Applied Catalysis B: Environmental, 46*(1), 1-15.

Treves, D. S., & Martin, M. M. (1994). Cellulose digestion in primitive hexapods: effect of ingested antibiotics on gut microbial populations and gut cellulase levels in the firebrat, Thermobia domestica (Zygentoma, Lepismatidae). *Journal of chemical ecology, 20*(8), 2003-2020.

Turner, S., Pryer, K. M., Miao, V. P., & Palmer, J. D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol, 46*(4), 327-338.

Vos, M., Quince, C., Pijl, A. S., de Hollander, M., & Kowalchuk, G. A. (2012). A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. *PLoS One, 7*(2), e30600.

Walker, J. C. F. (2007). *Primary wood processing: principles and practice*: Springer.

Wang, Y., Stingl, U., Anton-Erxleben, F., Geisler, S., Brune, A., & Zimmer, M. (2004). "Candidatus hepatoplasma crinochetorum," a new, stalk-forming lineage of Mollicutes colonizing the midgut glands of a terrestrial isopod. *Appl Environ Microbiol, 70*(10), 6166-6172.

Ward, D. M., Weller, R., & Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature, 345*(6270), 63-65.

Watanabe, H., Noda, H., Tokuda, G., & Lo, N. (1998). A cellulase gene of termite origin. *Nature, 394*(6691), 330-331.

Watanabe, H., & Tokuda, G. (2010). Cellulolytic Systems in Insects. [Review]. *Annual Review of Entomology, 55*, 609-632.

Waterbury, J. B., Turner, R. D., & Calloway, C. B. (1983). A cellulolytic nitrogen-fixing bacterium cultured from the gland of Deshayes in shipworms(Bivalvia: Teredinidae). *Science(Washington), 221*(4618), 1401-1403.

Watling, L. (1989). A classification system for crustacean setae based on the homology concept. *Crustacean issues, 6*, 15-26.

Weber, M., Salo, V., Uuskallio, M., & Raudaskoski, M. (2005). Ectopic expression of a constitutively active Cdc42 small GTPase alters the morphology of haploid and dikaryotic hyphae in the filamentous homobasidiomycete *Schizophyllum commune*. *Fungal Genetics and Biology, 42*(7), 624-637.

Werren, J. H., & Windsor, D. M. (2000). Wolbachia infection frequencies in insects: evidence of a global equilibrium? *Proceedings of the Royal Society of London. Series B: Biological Sciences, 267*(1450), 1277-1285.

Westram, A. M., Jokela, J., Baumgartner, C., & Keller, I. (2011). Spatial distribution of cryptic species diversity in European freshwater amphipods (Gammarus fossarum) as revealed by pyrosequencing. *PloS one, 6*(8), e23879.

Whitworth, T. L., Dawson, R. D., Magalon, H., & Baudry, E. (2007). DNA barcoding cannot reliably identify species of the blowfly genus Protocalliphora (Diptera:

Calliphoridae). *Proceedings of the Royal Society B: Biological Sciences, 274*(1619), 1731-1739.

Wong, D. W. S. (2009). Structure and Action Mechanism of Ligninolytic Enzymes. [Review]. *Applied Biochemistry and Biotechnology, 157*(2), 174-209.

Worrall, J. J., Anagnost, S. E., & Zabel, R. A. (1997). Comparison of wood decay among diverse lignicolous fungi. *Mycologia*, 199-219.

Yang, G. (2013). *Intersexuality and endocrine disruption in the amphipod Echinogammarus marinus- From genes to physiology.* University of Aberdeen, Unpublished Thesis.

Yang, J. C., Madupu, R., Durkin, A. S., Ekborg, N. A., Pedamallu, C. S., Hostetler, J. B., et al. (2009). The complete genome of Teredinibacter turnerae T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *Plos One, 4*(7), e6085.

Yonge, C. M. (1927). The absence of cellulase in Limnoria. *Nature, 119*, 855.

Zhang, Y. H. P., Ding, S. Y., Mielenz, J. R., Cui, J. B., Elander, R. T., Laser, M., et al. (2007). Fractionating recalcitrant lignocellulose at modest reaction conditions. *Biotechnology and Bioengineering, 97*(2), 214-223.

Zimmer, M., & Bartholme, S. (2003). Bacterial endosymbionts in Asellus aquaticus (Isopoda) and Gammarus pulex (Amphipoda) and their contribution to digestion. *Limnology and Oceanography, 48*(6), 2208-2213.

Zimmer, M., Danko, J. P., Pennings, S. C., Danford, A. R., Carefoot, T. H., Ziegler, A., et al. (2002). Cellulose digestion and phenol oxidation in coastal isopods (Crustacea : Isopoda). *Marine Biology, 140*(6), 1207-1213.

Zimmer, M., & Topp, W. (1998a). Do woodlice (Isopoda : Oniscidea) produce endogenous cellulases? *Biology and Fertility of Soils, 26*(2), 155-156.

Zimmer, M., & Topp, W. (1998b). Microorganisms and cellulose digestion in the gut of the woodlouse Porcellio scaber. *Journal of Chemical Ecology, 24*(8), 1397-1408.

Zlateva, T., Di Muro, P., Salvato, B., & Beltramini, M. (1996). The o-diphenol oxidase activity of arthropod hemocyanin. [doi: DOI: 10.1016/0014-5793(96)00326-2]. *FEBS Letters, 384*(3), 251-254.

9 Appendix

9.1 CO1 sequences from Cheluridae populations

>UK *C. terebrans*
TACCCTTTATTTCATTTTAGGGGCCTGGGCTAGTGTGCTCGGGACCTCCATAAGTGTAATTATTCGTTCCGAGCTAAGAGCGCCC
GGAAATTTAATTGGGGATGACCAAATTTATAACGTCATAGTCACAGCCCACGCTTTTATTATAATTTTCTTCATAGTTATGCCTA
TTATAATCGGCGGATTCGGAAACTGACTTGTCCCCTTAATGCTAGGTAGCCCCGATATGGCCTTTCCTCGAATAAACAATATAAG
ATTTTGACTTCTCCCCCCTTCGCTTTCCTTGTTGCTATTAAGAGGGCTTGTTGAAAGAGGAGTCGGAACGGGGTGAACTGTTTAC
CCCCCGCTAGCAGCTGCCTCTGGCCACTCTGGCGCCTCAGTAGACCTCGCCATCTTCTCCCTACATCTTGCAGGGGCCTCTTCCA
TCTTGGGTGCCATTAATTTTATTTCCACTATCATCAATATGCGCAGGCCAGGGATATTCTTTGACCGAATACCCTTATTTGTCTG
GTCAGTCTTCATCACAGCCATTCTCCTCTTACTCTCACTCCCTGTCTTGGCAGGAGCTATTACCATGCTTTTGACAGACCGAAAT
CTTAACACCTCATTTTTTGACCCTGCGGGAGGGGGTGACCCCATTCTTTACCAGCACCTATTC

>Egypt *C. terebrans*
TACCCTTTATTTTATTCTAGGGGCCTGGGCTAGTGTCCTAGGGACCTCTATAAGTGTAATTATTCGCTCAGAGCTAAGAGCGCCC
GGAAATTTAATTGGAGATGATCAAATTTATAACGTTATAGTTACAGCTCACGCTTTCATTATAATTTTCTTCATAGTTATGCCTA
TTATGATCGGTGGATTTGGAAACTGACTTGTCCCCTTAATGTTAGGTAGTCCTGACATAGCCTTTCCTCGAATAAACAATATAAG
GTTTTGACTTCTTCCCCCTTCACTCTCTTTGTTGCTGTTGAGGGGCTTGTTGAGAGGGGAGTTGGCACCGGATGGACTGTGTAC
CCCCCTCTAGCAGCTGCTTCTGGTCACTCCGGTGCCTCAGTAGACCTCGCTATCTTTTCCTTGCACCTTGCAGGGGCCTCTTCCA
TCTTGGGTGCCATTAACTTCATCTCCACTATCATCAACATGCGCAGGCCAGGGATATTCTTTGACCGAATGCCTTTATTTGTATG
ATCAGTGTTTATCACAGCCATTCTTCTCTTATTATCACTCCCTGTCTTAGCTGGGGCTATTACCATGCTTTTGACTGACCGAAAC
CTTAACACTTCATTTTTTGACCCAGCAGGGGGAGGTGATCCCATTCTTTACCAGCACCTATTC

>Croatia *C. terebrans*
CACCCTTTATTTTATTTTAGGGGCCTGGGCTAGTGTCCTAGGAACCTCTATAAGTGTAATTATTCGCTCAGAGCTAAGAGCGCCC
GGAAATTTAATCGGAGACGATCAAATTTACAACGTCATAGTTACAGCCCACGCTTTCATTATAATTTTCTTCATAGTTATGCCTA
TTATGATTGGTGGATTTGGAAACTGACTTGTCCCCTTAATGTTAGGTAGTCCTGACATAGCCTTCCCTCGAATAAACAATATAAG
GTTTTGACTTCTTCCCCCTTCACTCTCTTTGTTGCTGTTAAGAGGGCTTGTTGAAAGGGGAGTTGGCACCGGATGGACTGTATAC
CCCCCTCTAGCAGCTGCTTCTGGTCACTCCGGTGCCTCAGTAGACCTCGCTATCTTTTCACTACACCTTGCAGGGGCCTCTTCCA
TCTTGGGTGCTATTAACTTTATCTCCACTATCATCAACATGCGCAGGCCAGGGATATTCTTTGACCGAATGCCTTTATTTGTGTG
ATCAGTGTTTATCACAGCCATTCTTCTCTTATTGTCACTCCCTGTCTTAGCAGGGGCTATTACCATGCTTTTGACTGACCGAAAC
CTTAATACTTCATTTTTTGACCCAGCAGGGGGAGGTGACCCCATTCTTTACCAGCACCTATTC

>Greece *C. terebrans*
TACCCTTTATTTTATTTTAGGGGCCTGGGCTAGTGTCCTAGGGACCTCTATAAGTGTAATTATTCGCTCAGAGCTAAGAGCGCCC
GGAAATTTAATCGGAGATGATCAAATTTACAACGTCATAGTTACAGCCCACGCTTTCATTATAATTTTCTTCATAGTTATGCCTA
TTATGATTGGTGGATTTGGAAACTGACTCGTCCCCTTAATGTTAGGTAGTCCTGACATAGCCTTTCCTCGAATAAACAATATAAG
GTTTTGACTTCTTCCCCCTTCACTCTCTTTATTGCTGTTAAGAGGGCTTGTTGAGAGGGGAGTTGGCACCGGATGGACTGTATAC
CCCCCTCTAGCCGCTGCTTCCGGTCACTCCGGCGCCTCAGTAGACCTCGCTATCTTTTCCCTGCACCTTGCAGGGGCCTCTTCCA
TCTTGGGTGCCATTAACTTTATCTCCACTATCATTAACATGCGCAGGCCAGGGATATTCTTTGACCGAATGCCTTTATTTGTGTG
ATCAGTGTTTATCACGGCCATTCTTCTCTTATTGTCACTCCCTGTCTTAGCAGGGGCTATTACCATGCTTTTGACTGACCGAAAC
CTTAATACTTCATTTTTTGACCCAGCAGGGGGAGGTGACCCCATTCTTTACCAGCACCTATTC

>Turkey *C. terebrans*
CACCCTTTATTTTATTTTAGGGGCCTGGGCTAGTGTCCTAGGAACCTCTATAAGTGTGATTATTCGCTCAGAGCTAAGAGCGCCC
GGAAATTTAATCGGAGACGATCAAATTTACAACGTCATAGTTACAGCCCACGCTTTCATTATAATTTTCTTCATAGTTATGCCTA
TTATGATTGGTGGATTTGGAAACTGACTTGTCCCCTTAATGTTCCCTCGAATAAACAATATAAG
GTTTTGACTTCTTCCCCCTTCACTCTCTTTGTTGCTGTTAAGAGGGCTTGTTGAAAGGGGAGTTGGTACCGGATGGACTGTATAC
CCCCCTCTAGCAGCTGCTTCTGGTCACTCTGGTGCCTCAGTAGACCTCGCTATCTTTTCACTACACCTTGCAGGGGCCTCTTCCA
TCTTGGGTGCTATTAACTTTATCTCCACTATCATCAACATGCGCAGGCCAGGGATATTCTTTGACCGAATGCCTTTATTTGTGTG
ATCAGTGTTTATCACAGCCATTCTTCTCTTATTGTCACTCCCTGTCTTAGCAGGGGCTATTACCATGCTTTTGACTGACCGAAAC
CTTAATACTTCATTTTTTGACCCAGCAGGGGGAGGTGACCCCATTCTTTACCAGCACCTATTC

>Tropchelura insulae
TACCCTTTATTTCATTTTAGGGGCCTGGGCTAGTGTGCTCGGGACCTCCATAAGTGTAATTATTCGTTCCGAGCTAAGAGCGCCC
GGAAATTTAATTGGGGATGACCAAATTTATAACGTCATAGTCACAGCCCACGCTTTTATTATAATTTTCTTCATAGTTATGCCTA
TTATAATCGGCGGATTCGGAAACTGACTTGTCCCCTTAATGCTAGGTAGCCCCGATATGGCCTTTCCTCGAATAAACAATATAAG
ATTTTGACTTCTCCCCCCTTCGCTTTCCTTGTTGCTATTAAGAGGGCTTGTTGAAAGAGGAGTCGGAACGGGGTGAACTGTTTAC
CCCCCGCTAGCAGCTGCCTCTGGCCACTCTGGCGCCTCAGTAGACCTCGCCATCTTCTCCCTACATCTTGCAGGGGCCTCTTCCA
TCTTGGGTGCCATTAATTTTATTTCCACTATCATCAATATGCGCAGGCCAGGGATATTCTTTGACCGAATGCCCTTATTTGTCTG
GTCAGTCTTCATCACAGCCATTCTCCTCTTACTCTCACTCCCTGTCTTGGCAGGAGCTATTACCATGCTTTTGACAGACCGAAAT
CTTAACACCTCATTTTTTGACCCTGCGGGAGGGGGTGACCCCATTCTTTACCAGCACCTATTC

9.2 ITS sequences from Cheluridae populations

>ITS UK *C. terebrans*
CGTGGTCGAGGCTCACGCGTGGGCCTCGGACGATGCAAAATAAACAAACTCAGACACTCTCCTTGCTTTGCTTTGAGCTGAACAA
AATATAACGTTCGGCGGTACGAAACGAGCAGTTTCTAAACCAAAATACAACTCTTAGCGGTGGATCACTTGGCCTGTGGAAGCTA
TGAAGACCGTAGCTAAGTGTGAGAATGGCAGCGAGCCGTCGCTATATGCGCTGTCCCCTTCTCTCATATGTCGAATGCACATTGC
GCCCCACCCTAGTGGTTGGCATTCACGGACAAAAACCGGTGAGCTGGATCAATCGCAAATGTGTGCTCCCGAGCAACGGCGGGAG
GTCAAACGCGCACGAGTGTGCGGTCCAGTTCCACGTTAGTAGTTCGGCCGACCCGATGGTGGGGCACTCCTAGTCGAGTGTGGTG
GACACACGCAAACAACTCGTTCGAAGTTGGCGTGATGATCTGTGACGATTTGGGTCAACATCAGTTGAACCCGCATGCATATATA
ACTGCCTTGGCTCGTACCTCTGGGCGCGGTCTCACCGAGTCGCGTCCATTGTA

>ITS Tropichelura insulae
CGTGGTCGCGACCGATCGTTCGGTCGCGGACGATGCAAAACAAACAAACATCANGCCCCTTTTTGTGTTTTATCTTTGATAACCG
ACCCTTTGCGCTCGGCGGTACGGGACTCGAGCGTCCCCAAACCACCATACAACTCTGAGCGGTGGATCACTAGGCCTGTGGAAGC
TATGAACACCATACCCAAGTGTGAGAATTGCAGCGAGCCGTCGCTACATGCGCTGTTATTTTCTCTCACATGTCAAATGCACATT
GCCCCCCACCCTAGCGGTTGGCATATCTTGTGGCTCTCGCACGAAAGCTGGTGAGCTGTAAAGAATTTTCTTGTTTGTGCGCGGC
TGCGCGCACACGCACAGGGAGCTCCGTGTTAGTAGTCCGACCGACCCGATGGTGGGGCACTCCTACTCGAGTGTGGGGCCCGCG
GCGCTTCTCTCACCCCTGTGCGCTGACTCAGACTCTGTCCGTTTGATTTCTGGTGAAATATTTGGGACAACGCCAGTTGCTCCCG
ACAGTCTCATCTGCCTTGGCTCGAACCTCTGGGCGCGTGTGAGAGCGCGCGCGTCCATTGCA

>ITS Egypt *C. terebrans*
CGTGGTCGAGGCTCACGCGTGGGCCTCGGACGATGCAAAATAAACAAACTCACACACTCTCCTTGCTTTGCTTTGAGCTGAACAA
AATATAACGTTCGGCGGTACGAAACGAGCAGTTTCTAAACCAAAATACAACTCTTAGCGGTGGATCACTTGCCCTGTGGAAGCTA
TGAAGACCGTAACTAAGTGTGAGAATGGCAGCGAGCCCTCGCTATATGCGCTGTCCCCTTCTCTCATATGTCGAATGCACATTGC
GCCCCACCCTAGTGGTTGGGATTCACGGACAAAAACCGGTGAGCTGGATCAATCGCAAATGTGTGCTCCCGAGCAACGGCGGGAG
GTCAAACGCGCACGAGTGTGCGGGCCAGTTCCACGTTAGTATTTCGGCCGACCCGATGGTGGGGCACTCCTACTCGAGTGTGGTG
GACACGCACACACAACTCGTTCNAANTTGGCGTGATGATCTGTGACTATTTGGGTCAACATCANTTGAACCCGCATGCATATATA
ACTGCCTTGGCTCGTACCTCTGGGCGCGGTCTCACCGACTCGCGTCCATTGTA

>ITS Croatia *C. terebrans*
CGTGGTCGAGGCTCACGCGTGGGCCTCGGACGATGCAAAATAAACAAACTCAAACACTCTCCTTGCTTTGCTTTGAGCTGAACAA
AATATAACGTTCGGCGGTACGAAACGAGCAGTTTCTAAACCAAAATACAACTCTTAGCGGTGGATCACTTGGCCTGTGGAAGCTA
TGAATACCGTAACTAAATGTGAGAATGGCAGCGAGCCGTCGCTATATGCGCTGTCCCCTTCTCTCATATGTCGAATGCACATTGC
GCCCCACCCTAGTGGTTGGCATTCACGGACAAAAACCGGTGAGCTGGATCAATCGCAAATGTGTGCTCCCGAGCAACGGCGGGAG
GTCAAACGCGCACGAGTGTGCGGTCCAGTTCCACGTTAGTAGTTCGGCCGACCCGATGGTGGGGCACTCCTAGTCGAGTGTGGTG
GACACACACAAACAACTCGTTCGAAGTTGGCGTGATGATCTGTGACGATTTGGGTCAACATCAGTTGAACCCGCATGCATATATA
ACTGCCTTGGCTCGTACCTCTGGGCGCGGTCTCACCGAGTCNCGTCCATTGTA

>ITS Turkey *C. terebrans*
CGTGGTCGAGGCTCACGCGTGGGCCTCGGACGATGCAAAATAAACAAACTCAGACACTATCCTTGCTTTGCTTTGAGCTGAACAA
AATATGAAGTTCGGCGGTACGAAACAAGAAGTTTCTAAACCAAAATACAACTCTTATCGGTGGATCACTTGTCCTGTGGAAACTA
TGAAAACCGTAGCTAAATGTGAGAATGGGAGCGAGCCGCCGCTATATGCGCTGTCCCCTTCTCTCTTATGTCGAATGCGCATTGC
GCCCCCCCCCATTGGTTGGGGTTCACGGACAAAAACCGGTGAGCTGGATCAAACACAAATGTGTGCTCTCGAGCAACGGCGGGAG
GTCAAACGCGCACGAGTGTGCGGTCCACTTCCACGTTATTAGTTCGGCCGACCCCATGGTGGGGCACTCCCATTCTAGTGTGGTG
GACACACGCAAACAACTCGTTCTAAATTGGCGTGATGTGTTGTGATAAGTTTGGTCATCATCTGATGACACCCCACACCCATATA
ACTGTGTTGGGTCTCACCTCTGGGGGCGGGCTCACCTCGCCTCGATCCTTGTA

9.3 Cytochrome b oxidase subunit sequences from Cheluridae populations

> Cytb Egypt *C. terebrans*
TGTCTTTACCTTCATATAGGCCGGGGTATCTACTACTCCTCCTTTACATTAACGCACACCTGAAATATCGGTGTCACTATCCTAA
TCCTCACTATAGCCACAGCTTTTATAGGGTACGTTCTCCCTGTCAATCAAATATCCTTTTGGGGGGCATCAGTTATTACCAATCT
TTTCTCTGAAGTCCCCTACATTGGGCCGGATATTGTCCGTTTAATGTGGGGGGGTGTGTCCATCGATAACCCTACTATTGTACGC
TTCTTTACCTTCCACTTCATCCTCCCCTTTGTCATCCTAGCTATAGTTGTAGTCCATATCACCCTTCTTCACCAGACGGGCTCCA
GAAACCCCCTGGGTATCCCCTCTAGTCTAGATAAAACGCCCTTTCACTCACACTTCTCTTCAAAGGATTTACTAGGCGTTATTAT
TGTCCTTCTCCTATTCACAACTTTATGCCTCTACTACCCAATGATTTTAGGGGACGACGAAAATTTTAACAGAGCTGACCCCGCC
GTCACTCCTCACCATATCCAACCTGAATGGTACTTTCTCTTTGCCTACGCCATCTTACGCTCTATTCCAAACAAATTAGGGGGTG
TTATCGCCCTAGCACTCTCAGTCCTCATTTTCTATGCCATACCTTTCACTTTCCTCGGACGGGTAAAGAGCTCCTCCTTTTACCC
TCTCAATAAATTTCTCTT

> Cytb Croatia *C. terebrans*
TGTCTTTACCTGCATATAGGCCGGGGTATCTACTACTCCTCCTTTACATTAACGCACACCTGAAATATCGGTGTTACCATCCTAA
TCCTTACCATAGCCACAGCTTTTATAGGATACGTTCTGCCTGTTAATCAAATATCCTTCTGGGGGGCATCAGTTATCACCAATCT
TTTCTCTGAAGTCCCCTACATTGGGCCGGATATTGTCCGTTTAATGTGGGGGGGTGTATCCATCGATAACCCTACTATCGTGCGC
TTTTTTACCTTCCACTTCATCCTCCCCTTTGTTATCCTAGCTATAGTTGTGGTCCATATTACCCTTCTTCACCAAACGGGCTCCA
GAAACCCTCTAGGTGTCCCCTCGAGTCTAGATAAAACGCCCTTTCACTCACACTTCTCTTCAAAGGATTTACTAGGCGTTATTAT
TGTCCTTCTCCTATTCACAACTTTGTGCCTTTACTACCCAATGATTTTAGGGGACGACGAAAATTTTAACAGAGCTGACCCCGCC
GTTACTCCTCACCACATCCAACCTGAATGGTACTTCCTCTTTGCCTACGCCATCTTACGCTCTATTCCAAACAAATTAGGGGGTG
TTATCGCCCTAGCACTCTCAGTCCTCATTTTCTATACCATACCTTTCACTTTCCTCGGACGGGTAAAAAGCTCCTCCTTTTACCC
TCTCAATAAATTTCTCTT

> Greece *C. terebrans*
TGTCTTTACCTGCATATAGGCCGGGGTATCTACTACTCCTCCTTTACATTAACGCACACCTGAAATATCGGTGTCACTATCCTAA
TCCTCACCATAGCCACAGCTTTTATAGGATACGTCCTCCCTGTTAATCAAATATCCTTTTGGGGAGCATCAGTTATCACCAATCT
TTTCTCTGAAGTCCCCTATATTGGGCCGGATATTGTCCGTTTAATATGGGGAGGTGTATCCATCGATAACCCTACTATTGTACGC
TTTTTTACCTTCCACTTCATCCTCCCCTTTGTTATCCTGGCTATAGTTGTGGTCCATATTACCCTTCTTCACCAAACGGGCTCCA
GAAACCCTCTGGGTGTCCCCTCGAATCTAGATAAAACGCCCTTTCACTCACACTTCTCTTCAAAGGATTTACTAGGTGTTATTAT
TGTCCTTCTCCTCTTCACAACTTTGTGCCTCTACTACCCAATGATTTTAGGGGATGACGAAAATTTTAATAGAGCTGACCCCGCC
GTCACTCCTCACCATATCCAACCTGAATGGTACTTCCTCTTTGCCTACGCCATCTTACGCTCTATTCCAAACAAATTAGGGGGTG
TTATCGCCCTAGCACTCTCAGTCCTCATCTTCTATACCATACCTTTCACTTTCCTCGGACGGGTAAAAAGCTCCTCCTTTTACCC
TCTCAATAAATTTCTCTT

> Cytb UK *C. terebrans*
TGTCTTTACCTGCATATAGGCCGTGGCATTTACTACTCCTCCTTTACATTGACACACACCTGAAATATCGGTGTCACTATCCTAA
TCCTGACTATAGCCGCAGCTTTCATAGGATATGTCCTCCCTGTTAATCAAATATCCTTTTGAGGAGCATCAGTTATTACCAATCT
TTTCTCTGAGGTCCCTTACGTAGGGCCAGATATTGTACGCCTCATATGGGGGGGAGTATCCATCGACAACCCTACTATCGTCCGG
TTCTTTACCTTTCATTTCATCCTTCCCTTTGTTATCCTAGCTATAGTTGTGGTTCACATCACCTTACTCCACCAGACGGGTTCCA
GAAACCCTTTAGGGGTCCCCTCCGGTCTAGACAAAACGCCCTTTCACCCACATTTCTCATCAAAAGATATACTAGGTGTTATCGT
TGTCCTACTCCTATTCACAACTCTATGCCTCTACTATCCAATGATTTTAGGGGACGATGAAAATTTTAATAGAGCTGACCCCGCT
GTGACTCCCCACCACATCCAACCAGAGTGGTACTTCCTCTTTGCCTACGCCATCCTACGCTCTATCCCAAACAAGTTGGGGGGGG
TTATCGCCCTAGCGCTTTCAGTCCTCATCTTCTATGCCCTCCCTTTTACTTTCCTTGGGCGGGTAAAAAGCTCCTCCTTTTATCC
CCTCAATAAATTTCTCTT

> Cytb Tropichelura insulae
TGTCTTTACCTGCATATAGGCCGTGGCATTTACTACTCCTCCTTTACATTGACACACACCTGAAATATCGGCGTCACCATCCTAA
TCCTGACTATAGCCACAGCTTTCATAGGATATGTCCTCCCTGTTAATCAAATATCCTTTGAGGAGCATCAGTTATTACCAATCT
TTTCTCTGAGGTCCCTTACGTAGGGCCAGATATTGTACGCCTCATATGGGGGGGAGTATCCATCGACAACCCTACTATCGTCCGG
TTCTTTACCTTTCATTTCATCCTTCCCTTTGTTATCCTAGCTATAGTTGTGGTTCACATCACCTTACTCCACCAGACGGGTTCCA
GAAACCCTTTAGGGGTCCCCTCCGGTCTAGACAAAACGCCCTTTCACCCACATTTCTCATCAAAAGATATACTAGGTGTTATCGT
TGTCCTACTCCTATTCACAACTCTATGCCTCTGCTATCCAATGATTTTAGGGGACGATGAAAATTTTAATAGAGCTGACCCCGCT
GTGACTCCCCACCACATCCAACCAGAGTGGTACTTCCTCTTTGCCTACGCCATCCTACGCTCTATCCCAAACAAGTTGGGGGGGG
TTATCGCCCTAGCGCTCTCAGTCCTCATCTTCTATGCCCTCCCTTTTACTTTCCTTGGGCGGGTAAAAAGCTCCTCCTTTTATCC
CCTCAATAAATTTCTCTT

> Cytb Turkey *C. terebrans*
TGTCTTTACCTGCATATAGGCCGTGGCATTTACTACTCCTCCTTTACATTGACACACACCTGAAATATCGGCGTCACTATCCTAA
TCCTGACTATAGCCGCAGCTTTCATAGGATATGTCCTCCCTGTTAATCAAATATCCTTTTGAGGAGCATCAGTTATTACCAATCT
TTTCTCTGAGGTCCCTTACGTAGGGCCAGATATTGTACGCCTCATATGGGGGGGAGTATCCATCGACAACCCTACTATCGTCCGG
TTCTTTACCTTTCATTTCATCCTTCCCTTTGTTATCCTAGCTATAGTTGTGGTTCACATCACCTTACTCCACCAGACGGGTTCCA
GAAACCCTTTAGGGGTCCCCTCCGGTCTAGACAAAACGCCCTTTCACCCACATTTCTCATCAAAAGATATACTAGGTGTTATCGT
TGTCCTACTCCTATTCACAACTCTATGCCTCTACTATCCAATGATTTTAGGGGACGATGAAAATTTTAATAGAGCTGACCCCGCT
GTGACTCCCCACCACATCCAACCAGAGTGGTACTTCCTCTTTGCCTACGCCATCCTACGCTCTATCCCAAACAAGTTGGGGGGGG
TTATCGCCCTAGCGCTTTCAGTCCTCATCTTCTATGCCCTCCCTTTTACTTTCCTTGGGCGGGTAAAAAGCTCCTCCTTTTATCC
CCTCAATAAATTTCTCTT

9.4 18S sequences from Cheluridae populations

>18S UK *C. terebrans*
TGCATGTCTAAGTCCAAGCTGTGTCTGACACGGCGAGACCGCGGACGGCTCATTAAATCAGTCGTGGTCCAAAT
GGGCCAGCGTTAACTCTACTTGGATAACTGTGGTAATTCCAGAGCTAATACATGCAACTCTTGTCCCGATGGCGCTCGTC
TGATCTCTTGCAGGTCGCGGACGTTCTGACGGATGCTTTTATTAGACCAAACCGCTGAGGGCGCTGTCACCTGCGTGGCA
GCGTTTGACTCGTGTTATGGTGACTCTGGATAACCTCTTTTGACAAGCGCACGTACCCCTCTTTTGAGCGGGACGGCGCT
GCCACTTTCGAGTGTCTGCCTTATCAGCTCTCAACCGTTCGTTATGTGCGACCGATGGCTTTGACGGGTAACGGGGAATC
AGGGTTCGATTCCGGAGAGGCAGCCTGAGAGACGGCTACCACGTCCAAGGACGGCAGCAGGCACGCAAATTACCCAATCC
CAGCAACTGGGGGAGGTAGTGACGAAATCTAACGATGCGGGCGCCCTTTCTGGGGCCCCGCAATCGGAATGAGCACTTTC
TAAACAACCTGTTGAGAACCTACTGAAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTTCAGCAGCATCTATT
AAAGTTGCTGCGGTTAAAAGGCTCGCAGTTGAATCTCAGTATCGAGCGCAGGCAGGTGGACGCTGGGGCGTGGGATCACA
AGGCGCGGCACGGACGCTTCTCTCTGGTGTTTCCTCTCACGGATCGCCGGGTTTTGTCGCACCGACCGCGTCTCTCACAA
ACCCCCGTCCGTGACAAGCCTCCTCGCTTGCGTACATATCGGAATCTGTCAAACGGAGTGTGCGCTGCTCTCTCGCTTTG
CTCCCTGCATTCTTTGTT

>18S Turkey *C. terebrans*
TGCATGTCTAAGTCCAAGCTGTGTCTGACACGGCGAGACCGCGGACGGCTCATTAAATCAGTCGTGGTCCAAATGGGCCAGCGTT
AACTCTACTTGGATAACTGTGGTAATTCCAGAGCTAATACATGCAACTCTTGTCCCGATGGCGCTCGTCTGATCTCTTGCAGGTC
GCGGACGTTCTGACGGATGCTTTTATTAGACCAAACCGCTGAGGGCGCTGTCACCTGCGTGGCAGCGTTTGACTCGTGTTATGGT
GACTCTGGATAACCTCTTTTGACAAGCGCACGTACCCCTCTTTTGAGCGGGACGGCGCTGCCACTTTCGAGTGTCTGCCTTATCA
GCTCTCAACCGTTCGTTATGTGCGACCGATGGCTTTGACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGCAGCCTGAGAG
ACGGCTACCACGTCCAAGGACGGCAGCAGGCACGCAAATTACCCAATCCCAGCAACTGGGGGAGGTAGTGACGAAATCTAACGAT
GCGGGCGCCCTTTCTGGGGCCCCGCAATCGGAATGAGCACTTTCTAAACAACCTGTTGAGAACCTACTGAAGGGCAAGTCTGGTG
CCAGCAGCCGCGGTAATTCCAGCTTCAGCAGCATCTATTAAAGTTGCTGCGGTTAAAAGGCTCGCAGTTGAATCTCAGTATCGAG
CGCAGGCAGGTGGACGCTGGGGCGTGGGATCACAAGGCGCGGCACGGACGCTTCTCTCTGGTGTTTCCTCTCACGGATCGCCGGG
TTTTGTCGCACCGACCGCGTCTCTCACAAACCCCCGTCCGTGACAAGCCTCCTCGCTTGCGTACATATCGGAATCTGTCAAACGG
AGTGTGCGGTGCTCTCTCGCTTTGCTCCCTGCATTCTTTGTT

>18S Croatia *C. terebrans*
TGCATGTCTAAGTCCAAGCTGTGTCTGACACGGCGAGACCGCGGACGGCTCATTAAATCAGTCGTGGTCCAAATGGGCCAGCGTT
AACTCTACTTGGATAACTGTGGTAATTCCAGAGCTAATACATGCAACTCTTGTCCCGATGGCGCTCGTCTGATCTCTTGCAGGTC
GCGGACGTTCTGACGGATGCTTTTATTAGACCAAACCGCTGAGGGCGCTGTCACCTGCGTGGCAGCGTTTGACTCGTGTTATGGT
GACTCTGGATAACCTCTTTTGACAAGCGCACGTACCCCTCTTTTGAGCGGGACGGCGCTGCCACTTTCGAGTGTCTGCCTTATCA
GCTCTCAACCGTTCGTTATGTGCGACCGATGGCTTTGACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGCAGCCTGAGAG
ACGGCTACCACGTCCAAGGACGGCAGCAGGCACGCAAATTACCCAATCCCAGCAACTGGGGGAGGTAGTGACGAAATCTAACGAT
GCGGGCGCCCTTTCTGGGGCCCCGCAATCGGAATGAGCACTTTCTAAACAACCTGTTGAGAACCTACTGAAGGGCAAGTCTGGTG
CCAGCAGCCGCGGTAATTCCAGCTTCAGCAGCATCTATTAAAGTTGCTGCGGTTAAAAGGCTCGCAGTTGAATCTCAGTATCGAG
CGCAGGCAGGTGGACGCTGGGGCGTGGGATCACAAGGCGCGGCACGGACGCTTCTCTCTGGTGTTTCCTCTCACGGATCGCCGGG
TTTTGTCGCACCGACCGCGTCTCTCACAAACCCCCGTCCGTGACAAGCCTCCTCGCTTGCGTACATATCGGAATCTGTCAAACGG
AGTGTGCGCTGCTCTCTCGCTTTGCTCCCTGCATTCTTTGTT

>18S Greece *C. terebrans*
TGCATGTCTAAGTCCAAGCTGTGTCTGACACGGCGAGACCGCGGACGGCTCATTAAATCAGTCGTGGTCCAAATGGGCCAGCGTT
AACTCTACTTGGATAACTGTGGTAATTCCAGAGCTAATACATGCAACTCTTGTCCCGATGGCGCTCGTCTGATCTCTTGCAGGTC
GCGGACGTTCTGACGGATGCTTTTATTAGACCAAACCGCTGAGGGCGCTGTCACCTGCGTGGCAGCGTTTGACTCGTGTTATGGT
GACTCTGGATAACCTCTTTTGACAAGCGCACGTACCCCTCTTTTGAGCGGGACGGCGCTGCCACTTTCGAGTGTCTGCCTTATCA
GCTCTCAACCGTTCGTTATGTGCGACCGATGGCTTTGACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGCAGCCTGAGAG
ACGGCTACCACGTCCAAGGACGGCAGCAGGCACGCAAATTACCCAATCCCAGCAACTGGGGGAGGTAGTGACGAAATCTAACGAT
GCGGGCGCCCTTTCTGGGGCCCCGCAATCGGAATGAGCACTTTCTAAACAACCTGTTGAGAACCTACTGAAGGGCAAGTCTGGTG
CCAGCAGCCGCGGTAATTCCAGCTTCAGCAGCATCTATTAAAGTTGCTGCGGTTAAAAGGCTCGCAGTTGAATCTCAGTATCGAG
CGCAGGCAGGTGGACGCTGGGGCGTGGGATCACAAGGCGCGGCACGGACGCTTCTCTCTGGTGTTTCCTCTCACGGATCGCCGGG
TTTTGTCGCACCGACCGCGTCTCTCACAAACCCCCGTCCGTGACAAGCCTCCTCGCTTGCGTACATATCGGAATCTGTCAAACGG
AGTGTGCGGTGCTCTCTCGCTTTGCTCCCTGCATTCTTTGTT

>18S Tropichelura insulae
TGCATGTCTAAGTACAAGCCGTGTCTGACACGGCGAGACCGCGGACGGCTCATTAAATCAGTCGTGGTCCGAATGGGCCAGTGTT
TTGAACTTACTTGGATAACTGTGGTAATTCCAGAGCTAATACATGCAACTGATATCCCGATGGCGTGCCGCTGATCTCTTGCAGA
TCGGGTTCGTTCTGACGGATGCTTTTATTAGACCAAACCGCTGGGGGCGCGCGCCTCGCCTCGCGCGTGGCCGCGCTTGACTCGT
GCTATGGTGACTCTGGATAACCTTTTCTGACAAGCGCACGTACCCCTCTTTTGAGCGGGACGGCGCTGCCACTTTCGAGTGTCTG
CCTTATCAGCTTTCAACCGCTCGTTATGTGCGACCGATGGCTTTGACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGCAG
CCTGAGAGACGGCTACCACGTCCAAGGACGGCAGCAGGCACGCAAATTACCCAATCCCAGCAACTGGGGGAGGTAGTGACGAAAT
CTAACGATGCGGGCGCCCTTTCTGGGGCCCCGCAATCGGAATGAGCACTTTCTAAACAACCTGTTGAGAACCTACTGAAGGGCAA
GTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTTCAGCAGCATCTATTAAAGTTGCTGCGGTTAAAAGGCTCGCAGTTGAATCTCA
GTATCGAGCGCAGGCAGGTGGACGCTGGGGCGTGGGATCACAAGGCGAGGCACGGGCGCGCGCGCGGTTCGCTTCTACTGCGG
CCGTGCGTCGCCGCACCGACCTCGTCTCTCACAAGCCCCCGTCCGTAACAAGCCTCCTCGCTTGCGTACATATCGGAATCTGTCA
CACGGAGTGTGGGCGCTCGCTCGTCCGTCTCGTCGTGCTCTCCTCCCC

>18S Egypt *C. terebrans*
TGCATGTCTAAGTCCAAGCTGTGTCTGACACGGCGAGACCGCGGACGGCTCATTAAATCAGTCGTGGTCCAAATGGGCCAGCGTT
AACTCTACTTGGATAACTGTGGTAATTCCAGAGCTAATACATGCAACTCTTGTCCCGATGGCGCTCGTCTGATCTCTTGCAGGTC
GCGGACGTTCTGACGGATGCTTTTATTAGACCAAACCGCTGAGGGCGCTGTCACCTGCGTGGCAGCGTTTGACTCGTGTTATGGT
GACTCTGGATAACCTCTTTTGACAAGCGCACGTACCCCTCTTTTGAGCGGGACGGCGCTGCCACTTTCGAGTGTCTGCCTTATCA

```
GCTCTCAACCGTTCGTTATGTGCGACCGATGGCTTTGACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGCAGCCTGAGAG
ACGGCTACCACGTCCAAGGACGGCAGCAGGCACGCAAATTACCCAATCCCAGCAACTGGGGGAGGTAGTGACGAAATCTAACGAT
GCGGGCGCCCTTTCTGGGGCCCCGCAATCGGAATGAGCACTTTCTAAACAACCTGTTGAGAACCTACTGAAGGGCAAGTCTGGTG
CCAGCAGCCGCGGTAATTCCAGCTTCAGCAGCATCTATTAAAGTTGCTGCGGTTAAAAGGCTCGCAGTTGAATCTCAGTATCGAG
CGCAGGCAGGTGGACGCTGGGGCGTGGGATCACAAGGCGCGGCACGGACGCTTCTCTCTGGTGTTTCCTCTCACGGATCGCCGGG
TTTTGTCGCACCGACCGCGTCTCTCACAAACCCCCGTCCGTGACAAGCCTCCTCGCTTGCGTACATATCGGAATCTGTCAAACGG
AGTGTGCGCTGCTCTCTCGCTTTGCTCCCTGCATTCTTTGTT
```

227

Sequence Distribution in SPS library

Graph Level	GO Term	#Seq	Score	Parents
2	metabolic process	15	10.00	biological_process
1	biological_process	16	6.69	
3	oxidation-reduction process	5	4.82	metabolic process
6	oxygen transport	3	3.00	gas transport
4	transport	5	2.16	establishment of localization
5	gas transport	3	1.80	transport
3	macromolecule metabolic process	3	1.49	metabolic process
3	establishment of localization	5	1.29	localization, biological_process
6	hydrogen peroxide catabolic process	1	1.00	cellular response to hydrogen peroxide, cellular catabolic process, hydrogen peroxide metabolic process
4	electron transport chain	1	1.00	oxidation-reduction process, generation of precursor metabolites and energy
4	innate immune response	1	1.00	immune response, defense response
8	cellular iron ion homeostasis	1	1.00	iron ion homeostasis, cellular metal ion homeostasis
6	mannan catabolic process	1	1.00	cellular polysaccharide catabolic process, mannan metabolic process, cellular cell wall macromolecule catabolic process
5	proteolysis	1	1.00	protein metabolic process
6	response to lipid hydroperoxide	1	1.00	response to hydroperoxide
6	aerobic respiration	1	1.00	cellular respiration
9	iron ion transport	1	1.00	transition metal ion transport
9	protein homotetramerization	1	1.00	protein homooligomerization, protein tetramerization
3	cellular metabolic process	3	0.98	metabolic process, cellular process
4	cellular catabolic process	2	0.96	cellular metabolic process, catabolic process
4	generation of precursor metabolites and energy	1	0.82	cellular metabolic process
2	localization	5	0.78	biological_process
3	response to stress	2	0.65	response to stimulus
4	defense response	1	0.60	response to stress
5	response to hydroperoxide	1	0.60	response to oxidative stress
7	cellular response to hydrogen peroxide	1	0.60	response to hydrogen peroxide, cellular response to reactive oxygen species
6	cellular polysaccharide catabolic process	1	0.60	cellular polysaccharide metabolic process, polysaccharide catabolic process, cellular carbohydrate catabolic process
6	mannan metabolic process	1	0.60	cellular cell wall macromolecule metabolic process, cellular polysaccharide metabolic process, hemicellulose metabolic process
8	protein homooligomerization	1	0.60	protein oligomerization
8	transition metal ion transport	1	0.60	metal ion transport
6	cellular respiration	1	0.60	energy derivation by oxidation of organic compounds
3	immune response	1	0.60	response to stimulus, immune system process
7	cellular metal ion homeostasis	1	0.60	metal ion homeostasis, cellular cation homeostasis
4	protein metabolic process	1	0.60	primary metabolic process, macromolecule metabolic process
8	iron ion homeostasis	1	0.60	cation homeostasis
8	protein tetramerization	1	0.60	protein oligomerization
5	cellular cell wall macromolecule catabolic process	1	0.60	cell wall macromolecule catabolic process, cellular cell wall macromolecule metabolic process, cellular catabolic process
5	hydrogen peroxide metabolic process	1	0.60	reactive oxygen species metabolic process

2	response to stimulus	2	0.54	biological_process
4	response to oxidative stress	1	0.49	response to chemical stimulus, response to stress
3	catabolic process	2	0.49	metabolic process
2	cellular process	4	0.46	biological_process
3	primary metabolic process	2	0.44	metabolic process
6	cellular response to reactive oxygen species	1	0.36	cellular response to oxidative stress, response to reactive oxygen species
6	cellular polysaccharide metabolic process	1	0.36	polysaccharide metabolic process, cellular carbohydrate metabolic process, cellular macromolecule metabolic process
7	protein oligomerization	1	0.36	protein complex assembly
5	cell wall macromolecule catabolic process	1	0.36	cell wall macromolecule metabolic process, catabolic process
8	metal ion homeostasis	1	0.36	cation homeostasis
7	cation homeostasis	1	0.36	ion homeostasis
4	reactive oxygen species metabolic process	1	0.36	cellular metabolic process
5	cellular cell wall macromolecule metabolic process	1	0.36	cellular macromolecule metabolic process, cellular cell wall organization or biogenesis, cell wall macromolecule metabolic process
2	immune system process	1	0.36	biological_process
4	energy derivation by oxidation of organic compounds	1	0.36	oxidation-reduction process, generation of precursor metabolites and energy
6	cellular cation homeostasis	1	0.36	cation homeostasis, cellular ion homeostasis
6	cellular carbohydrate catabolic process	1	0.36	cellular carbohydrate metabolic process, carbohydrate catabolic process
6	hemicellulose metabolic process	1	0.36	cell wall polysaccharide metabolic process
7	metal ion transport	1	0.36	cation transport
6	response to hydrogen peroxide	1	0.36	response to reactive oxygen species
6	polysaccharide catabolic process	1	0.36	polysaccharide metabolic process, macromolecule catabolic process, carbohydrate catabolic process
3	response to chemical stimulus	1	0.29	response to stimulus
5	response to reactive oxygen species	1	0.22	response to inorganic substance, response to oxidative stress
4	cellular macromolecule metabolic process	1	0.22	cellular metabolic process, macromolecule metabolic process
6	cell wall polysaccharide metabolic process	1	0.22	polysaccharide metabolic process, cell wall macromolecule metabolic process
5	cellular response to oxidative stress	1	0.22	cellular response to stress, cellular response to chemical stimulus, response to oxidative stress
6	protein complex assembly	1	0.22	macromolecular complex assembly, protein complex biogenesis, protein complex subunit organization
4	cellular carbohydrate metabolic process	1	0.22	cellular metabolic process, carbohydrate metabolic process
4	cellular cell wall organization or biogenesis	1	0.22	cellular component organization or biogenesis at cellular level, cell wall organization or biogenesis
5	cellular ion homeostasis	1	0.22	ion homeostasis, cellular chemical homeostasis
6	ion homeostasis	1	0.22	chemical homeostasis
4	carbohydrate catabolic process	1	0.22	carbohydrate metabolic process, catabolic process
4	macromolecule catabolic process	1	0.22	macromolecule metabolic process, catabolic process
4	polysaccharide metabolic process	1	0.22	macromolecule metabolic process, carbohydrate metabolic process
6	cation transport	1	0.22	ion transport
4	cell wall macromolecule metabolic process	1	0.22	macromolecule metabolic process, cell wall organization or biogenesis
4	cellular response to chemical stimulus	1	0.13	response to chemical stimulus, cellular response to stimulus
4	response to inorganic substance	1	0.13	response to chemical stimulus

4	cellular chemical homeostasis	1	0.13	chemical homeostasis, cellular homeostasis
4	cellular response to stress	1	0.13	response to stress, cellular response to stimulus
5	chemical homeostasis	1	0.13	homeostatic process
4	protein complex biogenesis	1	0.13	cellular component biogenesis
3	cellular component organization or biogenesis at cellular level	1	0.13	cellular component organization or biogenesis, cellular process
3	cell wall organization or biogenesis	1	0.13	cellular component organization or biogenesis
4	carbohydrate metabolic process	1	0.13	primary metabolic process
5	ion transport	1	0.13	transport
5	protein complex subunit organization	1	0.13	macromolecular complex subunit organization
5	macromolecular complex assembly	1	0.13	macromolecular complex subunit organization, cellular component assembly
2	cellular component organization or biogenesis	2	0.11	biological_process
4	cellular component assembly	1	0.08	cellular component biogenesis, cellular component organization
3	cellular homeostasis	1	0.08	homeostatic process, cellular process
4	homeostatic process	1	0.08	regulation of biological quality
4	macromolecular complex subunit organization	1	0.08	cellular component organization
3	cellular response to stimulus	1	0.08	response to stimulus, cellular process
3	regulation of biological quality	1	0.05	biological regulation
3	cellular component organization	1	0.05	cellular component organization or biogenesis
3	cellular component biogenesis	1	0.05	cellular component organization or biogenesis
2	biological regulation	1	0.03	biological_process

Sequence Distribution in beech library

Graph Level	GO Term	#Seq	Score	Parents
2	metabolic process	11	5.38	biological_process
1	biological_process	12	3.89	
3	oxidation-reduction process	4	3.82	metabolic process
6	oxygen transport	3	3.00	gas transport
4	carbohydrate metabolic process	4	2.26	primary metabolic process
4	transport	4	2.08	establishment of localization
6	mannan catabolic process	2	2.00	cellular polysaccharide catabolic process, mannan metabolic process, cellular cell wall macromolecule catabolic process
5	gas transport	3	1.80	transport
3	primary metabolic process	4	1.36	metabolic process
4	cellular catabolic process	3	1.32	cellular metabolic process, catabolic process
3	establishment of localization	4	1.25	localization, biological_process
6	cellular polysaccharide catabolic process	2	1.20	cellular polysaccharide metabolic process, polysaccharide catabolic process, cellular carbohydrate catabolic process
6	mannan metabolic process	2	1.20	cellular cell wall macromolecule metabolic process, cellular polysaccharide metabolic process, hemicellulose metabolic process
5	cellular cell wall macromolecule catabolic process	2	1.20	cell wall macromolecule catabolic process, cellular cell wall macromolecule metabolic process, cellular catabolic process
3	cellular metabolic process	4	1.11	metabolic process, cellular process
6	hydrogen peroxide catabolic process	1	1.00	cellular response to hydrogen peroxide, cellular catabolic process, hydrogen peroxide metabolic process
4	electron transport chain	1	1.00	oxidation-reduction process, generation of precursor metabolites and energy
4	innate immune response	1	1.00	immune response, defense response
6	response to lipid hydroperoxide	1	1.00	response to hydroperoxide
6	aerobic respiration	1	1.00	cellular respiration
9	protein homotetramerization	1	1.00	protein homooligomerization, protein tetramerization
4	generation of precursor metabolites and energy	1	0.82	cellular metabolic process
2	localization	4	0.75	biological_process
6	cellular polysaccharide metabolic process	2	0.72	polysaccharide metabolic process, cellular carbohydrate metabolic process, cellular macromolecule metabolic process
5	cell wall macromolecule catabolic process	2	0.72	cell wall macromolecule metabolic process, catabolic process
5	cellular cell wall macromolecule metabolic process	2	0.72	cellular macromolecule metabolic process, cellular cell wall organization or biogenesis, cell wall macromolecule metabolic process
6	cellular carbohydrate catabolic process	2	0.72	cellular carbohydrate metabolic process, carbohydrate catabolic process
6	hemicellulose metabolic process	2	0.72	cell wall polysaccharide metabolic process
6	polysaccharide catabolic process	2	0.72	polysaccharide metabolic process, macromolecule catabolic process, carbohydrate catabolic process
3	response to stress	2	0.65	response to stimulus
3	catabolic process	3	0.62	metabolic process
4	defense response	1	0.60	response to stress
5	response to hydroperoxide	1	0.60	response to oxidative stress
7	cellular response to hydrogen peroxide	1	0.60	response to hydrogen peroxide, cellular response to reactive oxygen species
8	protein homooligomerization	1	0.60	protein oligomerization
6	cellular respiration	1	0.60	energy derivation by oxidation of organic compounds
3	immune response	1	0.60	response to stimulus, immune system process
8	protein tetramerization	1	0.60	protein oligomerization

5	hydrogen peroxide metabolic process	1	0.60	reactive oxygen species metabolic process
2	response to stimulus	2	0.54	biological_process
2	cellular process	4	0.50	biological_process
4	response to oxidative stress	1	0.49	response to chemical stimulus, response to stress
4	cellular macromolecule metabolic process	2	0.43	cellular metabolic process, macromolecule metabolic process
6	cell wall polysaccharide metabolic process	2	0.43	polysaccharide metabolic process, cell wall macromolecule metabolic process
4	cellular carbohydrate metabolic process	2	0.43	cellular metabolic process, carbohydrate metabolic process
4	cellular cell wall organization or biogenesis	2	0.43	cellular component organization or biogenesis at cellular level, cell wall organization or biogenesis
4	carbohydrate catabolic process	2	0.43	carbohydrate metabolic process, catabolic process
4	macromolecule catabolic process	2	0.43	macromolecule metabolic process, catabolic process
4	polysaccharide metabolic process	2	0.43	macromolecule metabolic process, carbohydrate metabolic process
4	cell wall macromolecule metabolic process	2	0.43	macromolecule metabolic process, cell wall organization or biogenesis
6	cellular response to reactive oxygen species	1	0.36	cellular response to oxidative stress, response to reactive oxygen species
7	protein oligomerization	1	0.36	protein complex assembly
4	reactive oxygen species metabolic process	1	0.36	cellular metabolic process
2	immune system process	1	0.36	biological_process
4	energy derivation by oxidation of organic compounds	1	0.36	oxidation-reduction process, generation of precursor metabolites and energy
6	response to hydrogen peroxide	1	0.36	response to reactive oxygen species
3	response to chemical stimulus	1	0.29	response to stimulus
3	cellular component organization or biogenesis at cellular level	2	0.26	cellular component organization or biogenesis, cellular process
3	cell wall organization or biogenesis	2	0.26	cellular component organization or biogenesis
3	macromolecule metabolic process	2	0.26	metabolic process
5	response to reactive oxygen species	1	0.22	response to inorganic substance, response to oxidative stress
5	cellular response to oxidative stress	1	0.22	cellular response to stress, cellular response to chemical stimulus, response to oxidative stress
6	protein complex assembly	1	0.22	macromolecular complex assembly, protein complex biogenesis, protein complex subunit organization
2	cellular component organization or biogenesis	3	0.18	biological_process
4	cellular response to chemical stimulus	1	0.13	response to chemical stimulus, cellular response to stimulus
4	response to inorganic substance	1	0.13	response to chemical stimulus
4	cellular response to stress	1	0.13	response to stress, cellular response to stimulus
4	protein complex biogenesis	1	0.13	cellular component biogenesis
5	protein complex subunit organization	1	0.13	macromolecular complex subunit organization
5	macromolecular complex assembly	1	0.13	macromolecular complex subunit organization, cellular component assembly
4	cellular component assembly	1	0.08	cellular component biogenesis, cellular component organization
4	macromolecular complex subunit organization	1	0.08	cellular component organization
3	cellular response to stimulus	1	0.08	response to stimulus, cellular process
3	cellular component organization	1	0.05	cellular component organization or biogenesis
3	cellular component biogenesis	1	0.05	cellular component organization or biogenesis