De novo sastavljanje transkriptoma vrste Procambarus virginalis Lyko, 2017 i ekspresija gena u odgovoru na patogen Aphanomyces astaci Shikora, 1906

Boštjančić, Ljudevit Luka

Sveučilište u Zagrebu

Prirodoslovno-matematički fakultet

Biološki odsjek

Ljudevit Luka Boštjančić

De novo sastavljanje transkriptoma vrste *Procambarus virginalis* Lyko, 2017 i ekspresija gena u odgovoru na patogen *Aphanomyces astaci* Shikora, 1906

Diplomski rad

Zagreb, 2021

This thesis is created in the Institute for Environmental Sciences at the University of Koblenz-Landau, under the supervision of Dr. Kathrin Theissinger and co-supervision of prof. dr. sc. Ivana Maguire. The thesis is submitted for grading to the Department of Biology at the Faculty of Science, University of Zagreb, with the aim of obtaining the Master's degree in Molecular biology.

I would first like to express my deepest gratitude to my supervisor, dr. Kathrin Theissinger, for the warm welcome in her working group, professional guidance and for entrusting me with the subject of this thesis. Your insightful comments and suggestions helped me to organise my thoughts and greatly improve my work.

All of this would not be possible without my co-supervisor prof. Ivana Maguire, who introduced me to the wonderful and exciting world of crayfish. I am thankful for all the guidance, advice and opportunities that were given to me over the past years. Looking back, I realise that You helped me become a better person and hopefully one day a good scientist.

I would like to acknowledge everyone who gave me their insights during writing of my master thesis, especially prof. Odile Lecompte and her working group in Strasbourg.

I would like to express my gratitude to my college Caterina Francesconi, for all her help, patience, and for wonderful time during the stay in Landau. And of course, for cooking recipes!

Through the whole time of writing this thesis, and before that I had amazing luck of being surrounded by my friends: Lena, Marko, Ela, Danijel, Ana, Lucija, Daniel and Leona. Whom have always been there for me through all the ups and downs. I am especially tankful to you Lena, for your companionship, for all the discussions that we had, for the patience, and for all happy moments during our student years!

Lastly, I would like to thank my parents and brothers. You were always the biggest inspiration, motivation, and support in my life!

Happy reading, *"G.N. - Bello ♫"*

BASIC DOCUMENTATION CARD

University of Zagreb

Faculty of Science

Department of Biology Graduation thesis

De novo assembly of _Procambarus virginalis_ Lyko, 2017 transcriptome and gene expression in response to the pathogen _Aphanomyces astaci_ Shikora, 1906

Ljudevit Luka Boštjančić

Department of Molecular Biology, Horvatovac 102a, 10000 Zagreb, Croatia

One of the biggest threats for freshwater crayfish populations is the pathogen *Aphanomyces astaci* Shikora, 1906 from the class Oomycetes that causes a disease lethal for native crayfish species, crayfish plague. The marbled crayfish, *Procambarus virginalis* Lyko, 2017 is an invasive crayfish species, known to be resistant to the crayfish plague disease. In this study the immune response to an *A. astaci* infection of 30 marbled crayfish exposed to two strains of *A. astaci* (PsI- high virulence and As- low virulence) was investigated. Hepatopancreas samples were taken at two time points, three days and 21 days post challenge. From the RNA sequencing reads, high-quality *de novo* transcriptomes of marbled crayfish were assembled containing 60,004 genes. Annotation was available for 41.9% of the transcripts. Differential gene expression analysis revealed in total 102 differentially expressed genes. More differentially expressed genes were observed among the samples from the second time point. Among the differentially expressed genes several known effectors of the crayfish innate immunity were detected. Previously unreported genes with potential role in crayfish innate immunity were also identified. They represent targets for future studies on crayfish innate immunity.

(78 pages, 19 figures, 1 table, 206 references, original in English)

Thesis deposited in the Central Biological Library

Key words: RNA sequencing, marbled crayfish, crayfish plague, innate immunity

Supervisor: Kathrin Theissinger, PhD

Co-supervisor: prof. Ivana Maguire, PhD

Reviewers: Ivana Maguire, PhD

Marin Ježić, PhD

Rosa Karlić, PhD

Substitution: Sandra Hudina, PhD

Thesis accepted: 11th of January 2021.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu

Prirodoslovno-matematički fakultet

Biološki odsjek Diplomski rad

De novo **sastavljanje transkriptoma vrste** ***Procambarus virginalis*** **Lyko, 2017 i ekspresija gena u odgovoru na patogen** ***Aphanomyces astaci*** **Shikora, 1906**

Ljudevit Luka Boštjančić

Zavod za Molekularnu biologiju, Horvatovac 102a, 10000 Zagreb, Croatia

Jedna on najveći prijetnji za populacije nativnih europskih slatkovodnih rakova je patogen *Aphanomyces astaci* Shikora, 1906 iz razreda Oomycetes, koji uzrokuje bolest račju kugu letalnu za nativne vrste slatkovodnih rakova. Mramorni rak, *Procambarus virginalis* Lyko, 2017 invazivna je vrsta i poznati prenositelj račje kuge. U ovom istraživanju proučen je imunosni odgovor 30 jedinki mramornog raka na infekciju s *A. astaci*. Korištena su dva soja patogena; PsI- visoko virulentan i As- nisko virulentan. Uzorkovanje hepatopankreasa rakova iz pokusa provedeno je tri i 21 dan nakon izlaganja patogenu te je provedena izolacija RNA. Iz fragmenata RNA, dobivenih sekvenciranjem, sastavljan je *de novo* transkriptom visoke kvalitete kojeg čini 60,004 gena. Za 41.9% transkripata je pronađena anotacija. Analiza diferencijalne ekspresije otkrila je 102 diferencijalno eksprimirana gena. Većina diferencijalno eksprimiranih gena bila je među uzorcima jedinki uzorkovanih 21 dana nakon infekcije. Među diferencijalno eksprimiranim genima bili su prisutni poznati efektori urođene imunosti rakova: serpin, lektini tipa-C i crustini. Novootkriveni su i geni s potencijalnom ulogom u urođenoj imunosti rakova. Oni predstavljaju predmet za buduća istraživanja urođene imunosti rakova.

(78 stranica, 19 slika, 1 tablica, 206 literaturnih navoda, jezik izvornika: engleski)

Rad je pohranjen u Središnjoj biološkoj knjižnici.

Ključne riječi: RNA sekvenciranje, maramorni rak, račja kuga, urođena imunost

Voditelj: dr. Kathrin Theissinger

Voditelj na matičnom fakultetu: prof. dr. sc. Ivana Maguire

Ocjenitelji: prof. dr. sc. Ivana Maguire

doc. dr. sc. Marin Ježić

doc. dr. sc. Rosa Karlić

Zamjena: doc. dr. sc. Sandra Hudina

Rad prihvaćen: 11. siječnja 2021.

Abbreviations

AaCht1- major chitinase gene 1

ACC- amorphous calcium carbonate

ALF- antilipopolysaccharide factor

ALS- Amyotrophic lateral sclerosis

AMPs- antimicrobial peptides

BBH- bi-directional best hit method

bp- base pair

βGBP- β-1,3-glucan-binding protein

CAP- carbonate apatite

CM- covariance matrix

CTLs- C-type lectins

DE- differentially expressed

Dscam- Down syndrome cell adhesion molecule

GBP- guanylate binding protein

GDP-Guanosine diphosphate

GMP- Guanosine monophosphate

GO- gene onthology

GTP-Guanosine triphosphate

HMM- Hidden Markovnikov models

Ig- immunoglobulin

IFN-γ- Interferon gamma

IMD- immune deficiency homolog

KO- KEGG ontology

LGBP- lipopolysaccharide and β-1,3-glucan-binding protein

LINE- Long interspersed nuclear elements

lncRNA- long non-coding RNA

MMP- maximum mappable prefix

mRNA- messenger RNA

ORF- open reading frame

PCA- principal component analysis

PGBP- peptidoglycan binding protein

PO- phenoloxidase

poly I:C- polyriboinosinic polyribocytidylic acid

ppA- proPO-activating enzyme

proPO- prophenoloxidase zymogen

PRR- pattern recognition receptors

rlog- regularised log transformation

RNAseq- RNA sequencing

rRNA- ribosomal RNA

SPH1- Serine proteinase homolog 1

SPH2- Serine proteinase homolog 2

TE- transposable elements

traf6- tumour necrosis factor receptor associated factor 6

UTR- untranslated region

vst- variance stabilising transformation

WAP- way acidic protein

Table of Contents

1.1. Freshwater crayfish- impact on biodiversity and threats 1

1.2. Invasive crayfish species 1

1.3. *Procambarus virginalis* Lyko, 2017 3

1.3.1. *Procambarus virginalis* as a model organism 5

1.7. Next-generation sequencing studies in freshwater crayfish 5

1.4. *Aphanomcyes astaci* Schikora, 1906 6

1.4.1. Pathogen-host co-evolution and *A. astaci* virulence 7

1.5. Crayfish immune system 9

1.5.1. Crayfish cuticle 10

1.5.2. Regulation and interaction of the crayfish gut microbiome with its immune system . 12

1.5.3. Cellular defence mechanisms 12

1.5.4. Humoral factors 13

1.6. *Procambarus virginalis* infection experiment 18

2. Study aims 20

3. Materials and methods 21

3.1. Quality assessment and filtering 21

3.2. Contamination assessment 21

3.3. *De novo* transcriptome assembly, quality assessment and filtering 23

3.4. Transcriptome annotation 26

3.5. Transcriptome mapping and differential gene expression analysis 27

3.6. Over-representation analysis of KEGG pathways and inspection of gene ontology terms associated with the DE gene subset 29

4. Results 30

4.1. Read pre-processing and contamination assessment 30

4.2. *Procambarus virginalis* transcriptome assembly 35

4.3. Gene expression patterns in hepatopancreatic tissue under *A. astaci* challenge 37

4.4. Annotation of differentially expressed genes in marbled crayfish under *A. astaci* challenge 42

4.4.1. Sampling point 1 42
4.4.2. Sampling point 2 42
4.5. KEGG and GO over-representation analysis of differentially expressed genes 43
5. Discussion 45
5.1 Transcriptome assembly 45
5.2. Transcriptome annotation 47
5.3. Differentially expressed genes and proPO cascade in response to *A. astaci* infection 49
5.3.1. C-type lectin duality and viral response 51
5.3.2. Genes DE in both As and PsI treatment of the late response to the *A. astaci* challenge 53
5.3.3. Specific late immune response to challenge with PsI strain 54
5.4. Broader perspective on the experimental results 55
6. Conclusion 57
7. Literature 58
8. Supplementary I

Introduction

1.1. Freshwater crayfish- impact on biodiversity and threats

Freshwater environments represent only 0.01% of water in the world (Gleick, 1996). Still, more than 100,000 species (6% of all the described species) live in the freshwater environments (Dudgeon et al., 2006). Among them, freshwater crayfish represent a group of organisms important for maintaining a balance in freshwater environments. They have a significant role in freshwater ecosystems, due to their size, long life span, omnivorous feeding, burrowing and bioturbation activities. Therefore, they are considered a keystone species (Füreder et al., 2006). Five European native freshwater crayfish species have been recognised, noble crayfish (*Astacus astacus* (Linnaeus, 1758)), narrow-clawed crayfish (*Pontastacus leptodactylus* (Eschscholtz, 1823)), stone crayfish (*Austropotamobius torrentium* (von Paula Shrank, 1803)), white- clawed crayfish (*Austropotamobius pallipes* (Lereboullet, 1858)) and thick-clawed crayfish (*Pontastacus pachypus* (Rathke, 1837)), all belonging to the family Astacidae. All freshwater crayfish are distributed among four families: Astacidae, Cambaridae, Cambaroididae and Parastacidae (Crandall and De Grave, 2017). The biodiversity of the freshwater ecosystems is increasingly impacted by multiple stressors. In Europe, populations of native freshwater crayfish are threatened by changes in freshwater ecosystems (Kouba et al., 2014, Maguire et al., 2018) caused by global warming and change in climate (Kuoba et al, 2016), pollution and urbanisation. This leads to habitat degradation (Gatti, 2016), regulation of water flow and loss of habitats (Dudgeon et al., 2006). Recently, alien American crayfish were introduced to Europe. Soon afterwards, it turned out that they are invasive and that they outcompete native species. Also, it was shown that they carry the pathogen *Aphanomyces astaci* Schikora, 1906, causative agent of a disease which is lethal to native crayfish species (Holdich et al., 2009, Maguire et al, 2016).

1.2. Invasive crayfish species

The introduction of non-native species to freshwater environments can lead to their establishment as invasive species that negatively influence the freshwater habitats and native species within them (Twardochleb et al., 2013). Invasive freshwater crayfish species represent, along with habitat destruction, the highest threat to European native species (Maguire et al., 2018). They displace native species from their habitats and have several competitive advantages

over native species, e.g. faster growth rate, earlier maturation, higher fecundity, and often higher aggressiveness (Westman et al., 2002, Twardochleb et al., 2013). Non-native invasive freshwater crayfish species were mainly introduced from North America (Jussila et al. 2015). In Croatia, three invasive crayfish species are present: the spiny-cheek crayfish (*Faxonius limosus* (Rafinesque, 1817) present in the Danube and Drava rivers, the signal crayfish (*Pacifastacus leniusculus* (Dana, 1852)) present in the rivers Drava, Mura, Korana, Una and Radonja and the marbled crayfish (*Procambarus virginalis* Lyko, 2017) present in the lake Šoderica near Koprivnica (Maguire et al., 2018). Other non-native crayfish species native to North America have been identified in Europe: *Procambarus clarkii* (Girard, 1852), *Procambarus acutus* (Girard, 1852), *Procambarus alleni* (Faxon, 1884), *Faxonius imminis* (Hagen, 1870), *Faxonius juvenilis* (Hagen, 1870), *Faxonius virilis* (Hagen, 1870), together with *Cherax quadricarinatus* (von Martens, 1868) and *Cherax destructor* Clark, 1936 that originate from Australia (Jussila et al. 2015). In addition to the damage, they cause to native crayfish species, invasive crayfish species can also cause declines in fish populations (Reynolds, 2011), reductions in amphibian populations (Garmardt, 1997) and a general decline in the biodiversity of freshwater ecosystems (McCarthy et al., 2006, Twardochleb et al., 2013). Still, not all non-native freshwater crayfish species introduced to a novel areal have a potential to become invasive. For a non-native species to become invasive, it must successfully undergo an invasion process (**Figure 1**). This process begins with the transport of the species to a novel area, where the species is introduced to a novel ecosystem, overcoming the initial geographical barrier. There, it must establish itself and survive the challenges of a novel environment. If successful, it has to overcome the reproductive barrier to become a naturalised alien species. Finally, its future impact will be determined by the ability to spread outside its new habitat, broadening its ecological and economic impact, becoming a successful invader (Lockwood et al, 2007, Richardson et al., 2000).

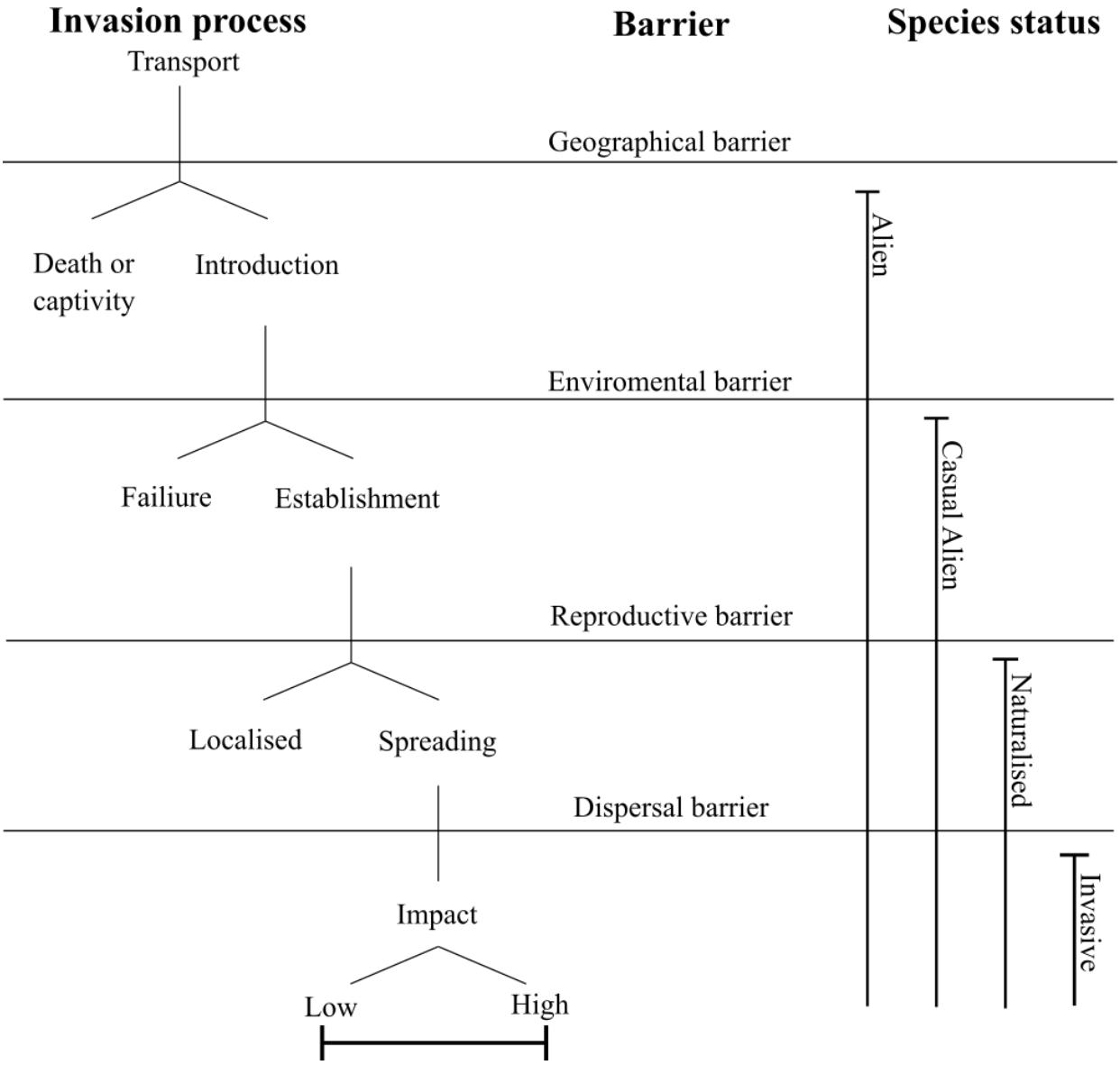

Figure 1. The overview of the invasion process of non-native species. Different barriers that introduced species has to overcome are shown as well as the species status according to Richardson et al., 2000 are shown. The figure was modified according to Lockwood et al., 2007.

1.3. Procambarus virginalis Lyko, 2017

The marbled crayfish (**Figure 2**) belongs to the family Cambaridae, whose members are native to North America (Crandall and De Grave, 2017). It was first recognised among aquarists in the 1990s in Germany (Chucholl and Pfeiffer, 2010) and later became popular because of its pigmented body and fast reproduction in the European and North American pet trade (Faulkes, 2010). In the order Decapoda, it is the only known species that obligatorily reproduces parthenogenetically. This means that one individual of this species (all female) can establish

large populations in a short period of time without the need of mating with a male (all offspring are mother's clones), which makes this species a highly successful invader (Scholtz et al., 2003). The marbled crayfish genome consists of triploid chromosomal set of 276 chromosomes, that is the exact triplicate number of the *Procambarus fallax* (Hagen, 1870) haploid set of chromosomes (Martin et al., 2016). Further studies confirmed that the marbled crayfish triploid genome emerged from the autopolyploid *P. fallax* gamete and mating of two distantly related *P. fallax* individuals (Gutekunst et al, 2018). The marbled crayfish can also be a carrier of the crayfish plague disease agent *A. astaci* (Keller et al. 2014). Breeding, commercial production import and keeping of the marbled crayfish in the European countries is prohibited by European Commission Regulations (EU 2016, Regulation No 1143/2014). Nonetheless, the number of new records of the marbled crayfish from EU countries is still increasing. Established populations of the marbled crayfish outside pet trade have been reported in many European countries: Germany (Chucholl and Pfeiffer, 2010), Croatia (Samardžić et al, 2014), Estonia (Ercoli, 2019), Italy (Vojkovská et al., 2014), Slovakia (Janský and Mutkovič, 2010), Czech Republic (Patoka et al., 2016), Netherlands (Kouba et al., 2014), Sweden (Bohman et al., 2013), Romania (Pârvulescu et al., 2017), Ukraine (Novitsky and Son, 2016) and Hungary (Weiperth et al., 2015).

Figure 2. Adult individual of invasive marbled crayfish (*Procambarus virginalis* Lyko, 2017) (photo by: Christian Huetter).

1.3.1. Procambarus virginalis as a model organism

Model organisms in biological research are animals, plants or microbes that can be used to study a certain biological process. There are three key characteristics that model organisms share: (i) generalisation of discoveries; they are representative for a larger group of organisms beyond themselves (ii) generalisation of the fields of research; they are models for a range of systems and processes and (iii) standardisation and affordability; they can be easily manipulated in the experimental conditions (Leonelli and Ankeny, 2013). The marbled crayfish has a relatively short generation time (~6 months), with each female capable of carrying 50-500 eggs at any time of the year (Vogt et al., 2004). Individuals of marbled crayfish, small in size, are easy to handle in laboratory conditions and show a high tolerance to physical manipulation and adaptability to a wide spectrum of experimental conditions (Vogt, 2008). All individuals of marbled crayfish form a genetically uniform population of clones (Gutekunst et al, 2018). Owning to its triploid genome, the marbled crayfish individuals can buffer the effects of deleterious mutations through the presence of multiple alleles (Gutekunst et al, 2018). Therefore, it is not surprising that since its discovery the marbled crayfish has been recognised as a potentially valuable model organism in genetics, epigenetics, environmental epigenomics, stem cell research, regeneration, evolution, development and cancer biology (Vogt, 2008, Jirikowski et al., 2010, Gutekunst et al, 2018, Hossain et al., 2018).

1.7. Next-generation sequencing studies in freshwater crayfish

High-throughput sequencing has revolutionised the biological field of research. Although, initially restricted to model organisms, reduction in the next-generation sequencing (NGS) costs has led to the large-scale sequencing initiatives of a wide array of species (including non-model organisms) (Eldem et al., 2017). Although species rich, NGS data is available only for a few representatives of the diverse group of freshwater crayfish. Cambarid *P. virginalis* was one of the first sequenced freshwater crayfish with available genomic and transcriptomic assembly (Gutekunst et al, 2018). Currently, together with the marbled crayfish, the genomes of *C. destructor* (GenBank assembly accession: GCA_009830355.1) and *C. quadricatinatus* (Tan et al., 2020) from the family Parastacidae are also available. In addition to the genome assemblies, the transcriptomes of native European freshwater crayfish species, *A. astacus* (Theissinger et al., 2016) and *A. pallipes* (Grandjean et al., 2020), *P. leptodactylus* (Manafrin et al., 2013, Tom et

al., 2013, Mosco et al., unpublished) as well as transcriptomes of non-native crayfish species *P. leniusculus* (Bunkis and Soderhall, unpublished), *C. quadricarinatus* (Galizer et al., 2013, Tan et al., unpublished) and *P. clarkii* (Manafirn et al., 2015, Tom et al., 2014) are also available. Although these assemblies are a step forward in the understanding of the biological processes in the freshwater crayfish, utilisation of these resources is often hindered by the high fragmentation rates, as is the case with the marbled crayfish genomic data is highly fragmented and transcriptome dataset contains only the longest isoforms.

1.4. Aphanomcyes astaci Schikora, 1906

The pathogen *A. astaci* belongs to the class of Oomycetes (order Saprolegiales) (**Figure 3**). Members of the genus *Aphanomyces* show different lifestyles as either plant and animal pathogens or saprotrophs and opportunistic parasitic species (Diéguez-Uribeondo, 2009). *Aphanomyces invadans,* David and Kirk, 1997 closely related to *A. astaci*, causes epizootic ulcerative syndrome in both wild and farmed fish populations (Oidtmann, 2012). Both *A. astaci* and *A. invidans* share a highly specialized parasitic nature with repeated motile zoospore emergence and lack of sexual reproduction (Diéguez-Uribeondo, 2009). Their rapid spread in Europe over large distances is caused by human trade and transport of large fish and shellfish, and introduction of invasive crayfish species (Cerenius et al., 2008, Uneastam and Weiss, 1970). In freshwater crayfish *A. astaci* is a causative agent of the disease, crayfish plague. The first observation of the crayfish plague in Europe was among the crayfish populations in the Po River (Italy), from where it spread through freshwater systems to Germany and France and later to other European countries (Seligo, 1895, Alderman, 1996). Since then, the spread of the disease has caused dramatic alterations in the balance of the European aquatic ecosystems (Souty-Grosset et al., 2006). The reproduction of *A. astaci* through zoospores is asexual. Therefore, the spread of this pathogen is clonal (Cerenius et al., 2008). Zoospores that do not find a host or are in unfavourable conditions can undergo encystment, creating a permanent cyst (lasting for around two weeks). This cyst, once in favourable conditions, can release a new zoospore through the rapture of the cyst wall, a process named repeated zoospore emergence (Cerenius and Soderhall, 1985). *A. astaci* has an ability to go through this cycle several times, allowing it to extend its survival time and chances of finding a host. Finally, *A. astaci* life cycle is closed after a period of hyphal growth within their host and eventual sporulation (Dieguez-Uribeondo et al.,

2006). The timing of spore release varies depending on the host. In noble crayfish it has been established that the release of spores occurs when the crayfish is moribund (Makkonen et al, 2012), while invasive signal crayfish seems to constantly release a low number of spores (Strand et al. 2012).

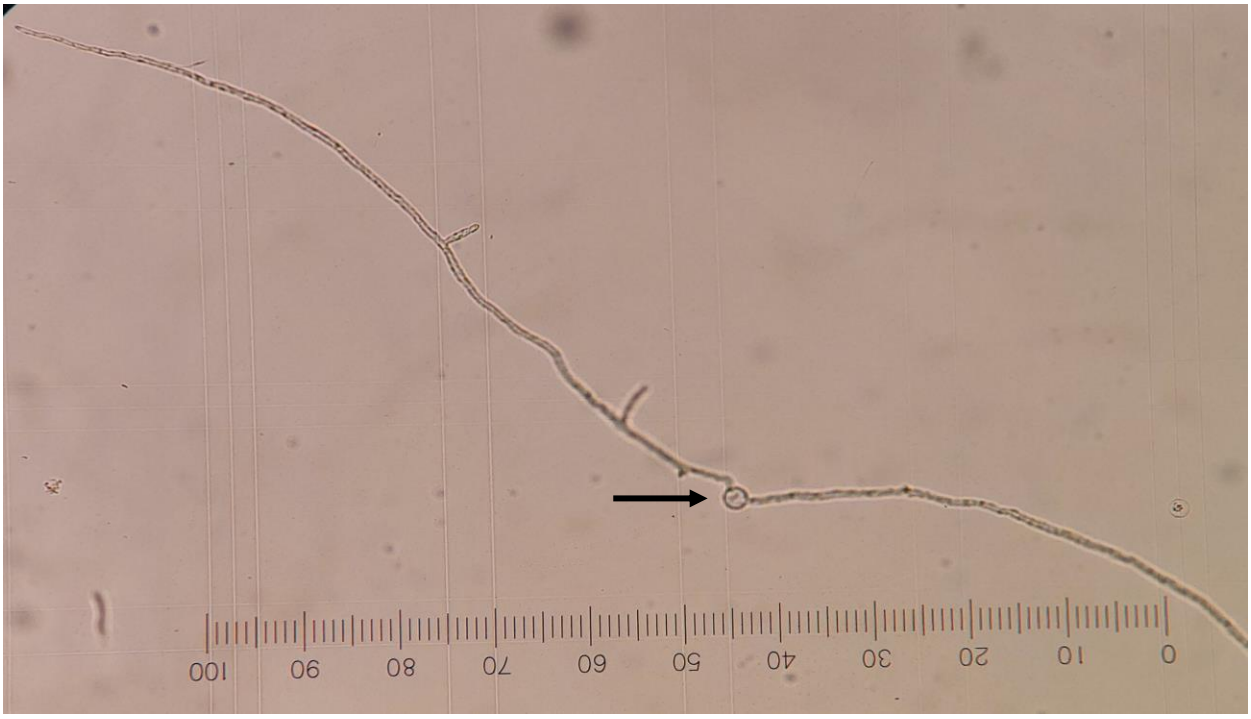

Figure 3. Sporulating spore (arrow) of the crayfish plague pathogen, *Aphanomyces astaci* Schikora, 1906 with the hyphae growing out viewed under the light microscope.

1.4.1. Pathogen-host co-evolution and A. astaci virulence

Today, 150 years after crayfish plague was first recognised in Europe, it is known that foreign non-native freshwater crayfish species are crucial for the pathogen´s transmission (Uneastam and Weiss, 1970). Invasive species originating from North America usually show resistance to the *A. astaci* infection and survive with the pathogen enclosed in their cuticle. They may also carry the pathogen latently until they become immunocompromised or stressed; at that point the onset of the lethal disease happens in the latent carriers (Cerenius et al., 2008). Conversely, native species are highly vulnerable to the pathogen (Souty-Grosset et al., 2006). Since 1970s a drastic decline in native European crayfish populations as well as mass mortalities and range restriction have been observed. This was initially attributed only to the invasive crayfish species. More recently the underlaying infection with *A. astaci* was recognised as a causative agent of the decline in native population numbers (Jussila et al., 2015). Similarly, with

the introduction of invasive crayfish species to Japan, the pathogen *A. astaci* has been shown to affect endemic crayfish species *Cambaroides japonicus* De Haan 1841 (Martín-Torrijos et al., 2018).

Different genotypes of *A. astaci* are present among invasive crayfish species and are often variable in virulence. The oldest strain recognised belongs to the As genotype (low in virulence) and was first isolated from infected noble crayfish, but its invasive host remains unknown (Alderman, 1996). Other strains were linked to the invasive crayfish species: highly virulent PsI and PsII genotypes were isolated from signal crayfish, the Pc genotype is linked to the red swamp crayfish and the Or genotype to the spiny-cheek crayfish (Kozubikova et al., 2011). The co-evolution of different genotypes and their hosts has been largely ignored. It has been hypothesised that the reduced virulence of the As genotype is the consequence of high selective pressure on the pathogen caused by low resistance of the native host to the infection (Jussila et al., 2015). The initial rapid spreading of the strains belonging to the As genotype caused the eradication of a number of native European freshwater crayfish populations. Once host populations were eradicated, the survival of the pathogen itself is compromised (Jussila et al., 2014). Therefore, the selection process towards the lower virulence of *A. astaci* As genotype might have occurred during the last 150 years. A possible mechanism of this reduced virulence could be linked to the inability of the zoospores to attach on or penetrate through the crayfish exoskeleton (Makkonen et al., 2012). On the other side, the highly virulent PsI genotype causes, to these days, almost 100% mortality among the noble crayfish stocks (Makkonen et al., 2014). The PsI genotype is usually introduced to a new water body with its original invasive host, *P. leniusculus*, which acts as a healthy carrier. For this reason, there is no evolutionary pressure for the reduction in the PsI genotype virulence. In fact, unlike the As genotype, once *A. astaci* PsI has caused the eradication of native crayfish populations, it is still able to circulate and survive among *P. leniusculus* populations. Despite the long co-evolution between *A.astaci* and invasive North American crayfish, occurrences of acute crayfish plague infections have been reported among the invasive crayfish populations (Jussila et al., 2014). The selection mechanisms also affect the native crayfish populations challenged by the *A. astaci* infection, leading to increased overall survival rates/time in some populations of noble crayfish (Makkonen et al., 2014), narrow-clawed crayfish (Svoboda et al., 2012), white-clawed crayfish (Manfrin and Pretto 2014, Maguire et al., 2016), stone crayfish (Maguire et al., 2016).

Due to the possibility of rapid co-evolution with its host (e.g. As strains and native European crayfish), *A. astaci* is an interesting model organism for the studies of the host-parasite interactions (Makkonen et al., 2016). The co-evolution between *A. astaci* and freshwater crayfish host has also been detected on the molecular level. For example, it has been shown that crayfish cuticular and blood proteinase inhibitors are specialised (pathogen specific) for inhibition of *A. astaci* proteolytic enzymes, rather than general for all proteolytic enzymes produced by members of the genus *Aphanomyces* (Dieguez-Uribeondo and Cerenius, 1998). On the other side, in the early stage after *A. astaci* spore germination, lipases and proteinases are released by the pathogen, followed by expression of major chitinase gene *AaCht1* (~10h after spore germination), responsible for chitin degradation. The expression of *AaCht1* happens irrespective of the presence of chitin in the host (Andersson and Cerenius, 2002). Other species within the genus *Aphanomyces* express *AaCht1* chitinase homologs only in the presence of chitin or are not capable of its production (Cerenius et al., 2008). Therefore, it can be concluded that *AaCht1* pre-programed expression originated as a highly specialised trait through the co-evolution of the pathogen with its crayfish host.

1.5. Crayfish immune system

The crayfish immune system relies mainly on the innate immunity (Cerenius and Söderhäll, 2012). Upon contact with the pathogen, three barriers are present: I) physiochemical barrier, which includes melanisation and reaction of proteinases and chitinase inhibitors present in the crayfish cuticle; II) cellular defense mechanisms, cytotoxicity reactions and encapsulation carried out by haemolymph cells; III) humoral factors, from which the most important is activation of the prophenoloxidase cascade (proPO-system), and the release of antibacterial and antifungal peptides together with agglutinins (Rowley, 2016). Although many aspects in the immune response of crayfish to *A. astaci* infection are known, the interplay and expression of individual genes involved in the different immune response pathways at different stages of the infection have not yet been elucidated (**Figure 4.**).

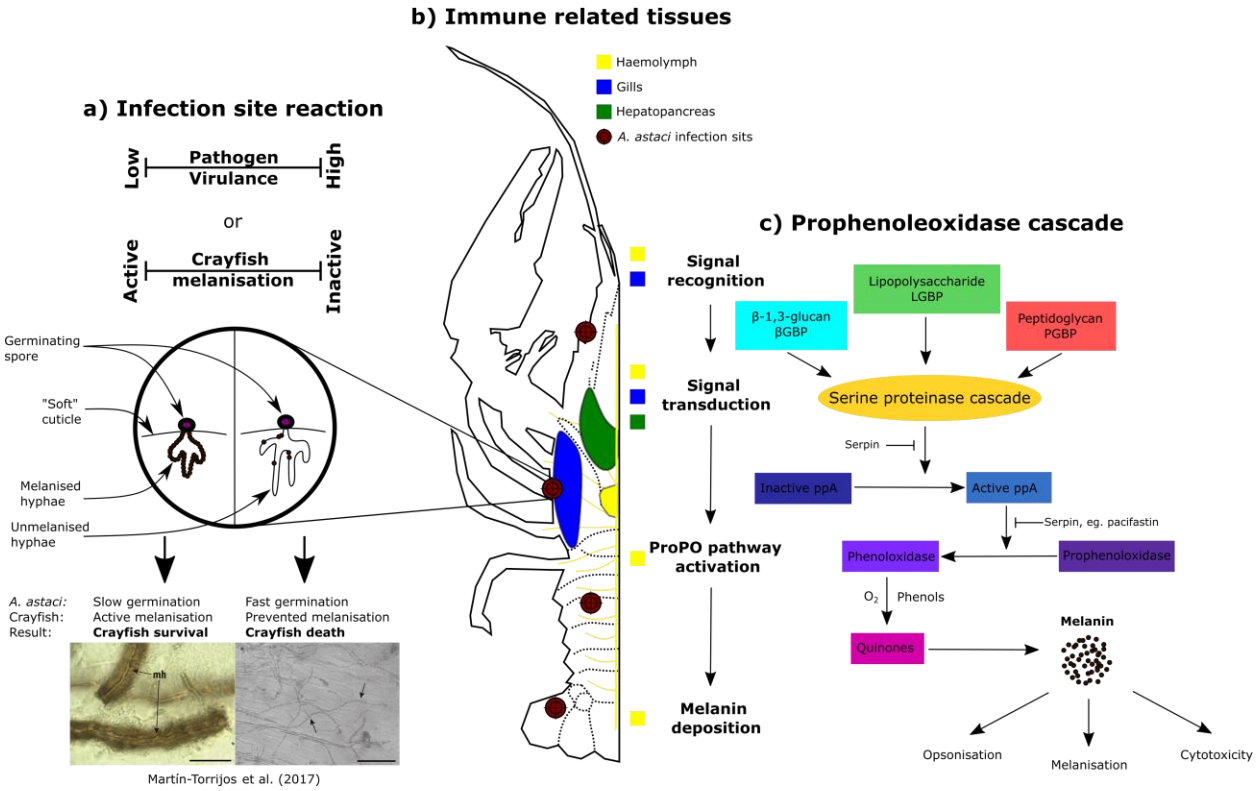

Figure 4. Overview of the crayfish immune response to *A. astaci* infection. **(a)** the reaction in the infection site is shown. Depending on the pathogen virulence and crayfish susceptibility (speed of the immune response), the pathogen can be stopped within the crayfish cuticle via melanisation reaction. Micrographs representing melanised hyphae (left) and unmelanised hyphae (right) in cuticle of *A. pallipes* are adapted from Martín-Torrijos et al., 2017. **(b)** Within the crayfish, three immune related tissues can be recognized: haemolymph with its effectors haemocytes (granular, semi-granular and hyaline cells), gills with nephrocytes and hepatopancreas that carries the metabolic reactions and with phagocytes. **(c)** On the right the canonical proPO cascade is shown. Upon recognition of the fungal (1,3)-β glucan with the (1,3)-β glucan-binding receptors, serine proteinase cascade is activated leading to the cleavage of proPO zymogen and activation of phenoloxidas. In its active from the catalysis of the oxidation of phenols to melanin via quinone intermediates is activated. Sticky melanin is then released by the semi-granular and granular cell in the infection site leading to the pathogen melanisation. The melanisation reaction is also involved in the cellular cytotoxicity and opsonisation reactions (figure adapted from Cerenius and Söderhäll, 2004). βGBP- β-1,3-glucan-binding protein, LGBP- lipopolysaccharide and β-1,3-glucan-binding protein, PGBP- peptidoglycan binding protein, ppA- prophenoloxidase activating enzyme, proPO- prophenoloxidase.

1.5.1. Crayfish cuticle

The cuticle of the crayfish is a chitin rich structure that encloses the animal and represents a formidable barrier against infection. The cuticle consists of three layers, endocuticle, exocuticle and epicuticle (**Figure 5.**). Epidermal cells, which lie on the basal membrane, secrete the building blocks of the endocuticle, chitin fibres and proteins. The endocuticle is the thickest layer of the cuticle. Above the endocuticle, the sclerotized exocuticle is hardened by quinone tanning, calcification and chitin. The outermost layer of the cuticle is its waxy epicuticle. Thousands of secretory helical and ribbon-like pore channels are formed through the crayfish cuticle from the epidermis reaching the epicuticular layer (Roer and Dillaman, 1984). The cuticle

also plays a vital role in the crayfish as a skeleton to which the muscles are attached. During their growth, the crayfish must periodically change its cuticle, through a process called moulting. The crayfish are highly vulnerable to post-ecdysis infections (Rowley, 2016). Furthermore, it has been observed that the highest rates of the *A. astaci* spores release occur during moulting and in moribund crayfish (Svoboda et al., 2013).

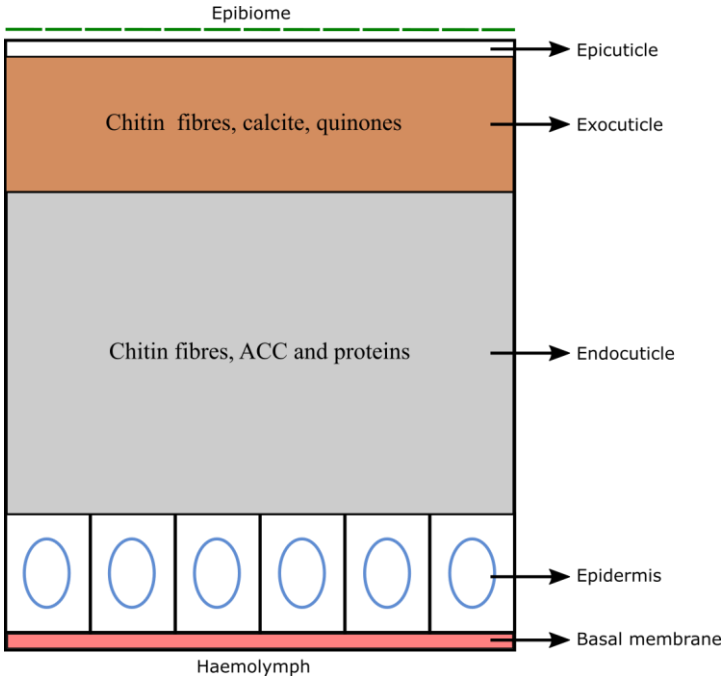

Figure 5. Structure of the crayfish cuticle adapted according to Roer and Dillaman, 1984 and Kunkel et al., 2012.

Early studies of the role of cuticle in the defence against *A. astaci* infection revealed that there is no structural difference between resistant (signal crayfish) and susceptible (noble crayfish) species. After attachment, the zoospore lyses the surface lipid layer and the germ tube that penetrates through the epicuticle forms. But, in invasive signal crayfish, *in vivo* experiments showed that there was neither inward nor outward penetration of intact cuticle pointing to its active resistance (Nyhlén and Unestam, 1975). Alongside chitin, three different minerals rich in Ca^{2+} have been identified in American lobster, *Homarus americanus*, calcite, amorphous calcium carbonate (ACC), and carbonate–apatite (CAP), each in the discrete domains of the cuticle. Calcite forms a continuous, dense layer that provides a physical barrier to the microbial attack. This layer is modulating/increasing the pH of the epicuticular layer by slowly dissolving into it, thus creating and unfavourable environment for microbial growth. Acid resistant CAP is lining the venerable canals that transverse the crayfish cuticle. ACC in the endocuticle is the precursor

to crystalline calcium carbonate forms and upon cuticular injury it can dissolve to form an alkaline antimicrobial shield for the cuticle (Kunkel et al., 2012). Humoral factors including prophenoloxidase cascade (Cerenius, 2010) and antimicrobial peptides (Brisbin et al., 2015) are also present in the cuticle and will be discussed in detail later. Most recently, a study by Orlić et al., 2020 revealed that 33% of the bacteria plate isolated from the crayfish cuticular epibiome of *P. leniusculus* and *P. leptodactylus* are inhibitors of *A. astaci* mycelial growth, indicating the important role of the cuticular epibiome in the resistance to the *A. astaci* infection.

1.5.2. Regulation and interaction of the crayfish gut microbiome with its immune system

Along with the gills, the crayfish digestive canal is unprotected or only partially protected by cuticle and are exposed to the environmental factors. Although, little is known about the microbiome and immune regulation in the intestines of crayfish, in signal crayfish the haemocyte homeostasis-like protein was discovered to be involved in the antimicrobial activities, preventing bacterial pathogenesis (Apitanaysai et al., 2016). Furthermore, a recent study on spiny lobster showed an asymptomatic bacteraemia in its haemolymph (Ooi et al., 2019). This bacteraemia was also identified in other crustaceans, including the freshwater crayfish species *P. clarkii* (Scott and Thune, 1986). The connection between the gut microbiome and the circulatory microbiome has been confirmed in the spiny lobster (Ooi et al, 2017). It is thought that this microbiome community may play a significant role in increasing hosts health, metabolism of protein and non- protein amino acids, modulation of immune response and present a competitors to the invading pathogens (Ooi et al., 2019, Wang and Wang, 2015).

1.5.3. Cellular defence mechanisms

There are four main effector cell types of the crustacean immune system: 1) haemocytes in the circulating haemolymph, 2) their precursors in the hematopoietic tissues of the upper part of the gastrointestinal tract, 3) fixed phagocytes in the hepatopancreas and 4) nephrocytes in gills. Each cell type has its own defined role in the immune system of crayfish. In the cellular defence two response mechanisms are predominant: phagocytosis of small particles and encapsulation of larger ones (Rowley, 2016).

In the circulatory system of crayfish three main types of haemocyte cells could be recognised: hyaline cells, semi-granular cells and granular cells (Bauchau, 1981). The immature

haemocytes, i.e. prohemocytes, are released from the hematopoietic tissues. Both hyaline and semi-granular cells can conduct phagocytosis. Semi-granular cells and granular cells can release/degranulate the products stored in their granules, therefore playing a vital role in the activation of the prophenoloxidase cascade and clotting (Rowley, 2016). It was found that in the signal crayfish, haematopoiesis is controlled by two cytokine-like growth factors, namely christened astakines- 1 and 2. Astakine-1 positively controls the transcription of christened crustacean hematopoietic factor from stem cells in the hematopoietic tissue. Astakine -2 is involved in the differentiation and release of granular haemocytes from the hematopoietic tissue (Lin and Söderhäll, 2011). During the immune response, reactive oxygen species serve as a signal for haematopoiesis activation via regulation of calcium-dependent extracellular transglutaminase that keeps the immature precursors of haemocytes inside hematopoietic tissue (Lin et al., 2008). A defence response called oxidative burst is also regulated by haemocytes and it results in production of free O_2^- radical and H_2O_2 causing oxidative damage to the microorganisms (Sierra et al., 2005). It is worth mentioning that hematopoietic cells are not capable of prophenoloxidase (proPO) production, and mature haemocyte are characterised by the onset of proPO expression (Cerenius and Söderhäll, 2018). In gills, as well as in hepatopancreas and other tissues, haemocytes can aggregate in special structures called nodules. Haemocytes within the nodules show high phagocytic activity and a variety of humoral responses, responsible for stopping bacterial infections (White et al., 1982). Nephrocytes, although not capable of phagocytosis, are capable of pinocytosis of the toxic soluble products of bacterial breakdown in the nodules (White et al., 1985). Another mechanism called encapsulation response is triggered by multicellular parasites. These parasites are unsheathed in the capsules by haemocytes, preventing their spread though the host organism (Rowley, 2016).

1.5.4. Humoral factors

Humoral response is one of the main components of the innate immunity. It involves the synthesis and release of immune proteins into the haemolymph, such as prophenoloxidase cascade effectors, antimicrobial peptides, lectins, and clotting enzymes (Rowley, 2016).

1.5.4.1. Prophenoloxidase cascade

The activation of prophenoloxidase cascade is the most recognised humoral response among crustaceans (**Figure 4**). Phenoloxidase synthesized in its zymogen form

(prophenoloxidase, proPO) is the central enzyme of the pathway. It is cleaved by activating protease (ppA) to the catalytically active PO and 20 kDA N-terminal fragment, ppA-proPO, with a strong agglutination and bacterial killing capacity. Alongside PO, ppA activates the formation of protein peroxinexin involved in pathogen recognition and cell adhesion (Lin et al., 2007). As mentioned above, proPO is released from the granular or semigranular cells in the response to the stimulation by the pathogen (Cerenius et al., 2008). The main products of this response are melanin and antimicrobial toxic quinones. Because of their toxicity this cascade is spatially and temporally finely tuned (Cerenius et al., 2008). Excessive activation of this pathway can cause damage to the host, therefore inhibitory proteins are of outmost importance. Some of the proteins involved in the regulation of this pathway are: pacifastacin, a regulatory inhibitor of ppA (Liang et al., 1997); melanisation inhibition proteins (Söderhäll et al., 2009); caspase 1-like molecule, limiting the proteolysis of proPO and is released concomitantly with the proPO and mannose-binding lectins (Jearaphunt et al., 2014). Lastly, in *D. melanogaster*, several serpin inhibitors of proPO cascade have been identified (De Gregorio et al., 2002). The proPO cascade is also connected with other key immune responses, which include production of opsonin and encapsulation promoting factors (Cerenius et al., 2008).

One of the main constituents of the oomycete cell wall, (1,3)-β-glucan, binds to the specific host membrane (1,3)-β glucan-binding receptors located on the haemocyte cells (Jiravanichpaisal et al., 2006). The (1,3)-β glucan-binding receptors play a role in the activation of proPO cascade by recognition of bacterial peptidoglycans through the complex between lipopolysaccharide- and (1,3)-β -glucan-binding protein and serine protease homologs (SPH1 and SPH2) in signal crayfish (Cerenius and Söderhäll, 2018). Numerous bacterial detection molecules have been found in *P. leniusculus, P. clarkii* and *C. quadricarinatus*, including ficolini-like proteins that can play a role in pattern recognition and bacterial agglutination (Wu et al., 2011), C-type lectins, immunoglobulin- like Dscam molecules that bind bacteria and inhibit bacterial growth (Ng et al., 2014).

Differences between the expression of proPO have been observed in susceptible and resistant crayfish. In the latter, the expression of proPO transcripts is continuously elevated and non-responsive to immune stimuli, while in susceptible crayfish its expression is at lower levels,

although it can be elevated by injection of immune stimulation with (1,3)-β glucan polysaccharide, laminarin (Cerenius et al., 2003).

1.5.4.2. Antimicrobial peptides

Antimicrobial peptides (AMPs) are present in numerous taxa. They are part of the host innate immune response, are potent against microbes, viruses, fungi and other parasites and can play a role in immunomodulation (Strominger, 2009). Other roles have also been attributed to AMPs, such as regulation of haematopoiesis via downregulation of astakine expression (Chang er al., 2013) as well as opsonisation (Liu et al., 2015). Many AMPs are produced and stored in the haemocytes and released upon immune stimulation (Sricharoen et al., 2005). Three types of AMPs were identified in crustaceans: (1) linear α-helical cysteine-free, (2) cationic cysteine-rich peptides containing β sheets, and (3) those with overrepresentation of a particular amino acid (Smith et al., 2010). In red-swamp crayfish it has been shown that trypsins are responsible for the enzymatic processing of newly discovered antibacterial PcnAMP and have a role as an activating enzyme (Zhao et al., 2020).

Some AMPs like cysteine-rich crustins are widely distributed among crustaceans, while others like penaedins (in family Penaeidae) are genera-specific (Smith et al., 2010). Crustins, cationic AMPs, have three main components: signal peptide, multi domain region at N-terminus and whey acidic protein (WAP) domain at the C- terminus. They are classified in five groups based on their structure (type I-V). Recently a new crustin protein was identified in red-swamp crayfish (Pc-crustin 4) which has bacterial growth inhibition activities, both *in vitro* and *in vivo* (Du et al., 2019).

In the signal crayfish, astacidine was identified as the AMP with antibacterial activity, derived from cleavage of hemocyanin (Lee et al., 2003). Astacidin 1 from signal crayfish has an antifungal activity as it causes a pore-formation in the membrane of fungal cells that leads to K^+ ion loss and membrane depolarisation (Choi and Lee, 2014). In red-swamp crayfish astacidins exhibit bacterial cell wall binding and agglutination activities (Shi et al., 2014). In the study conducted by Rončević et al., 2020 four astacidin-like peptides with variable antibacterial activities were identified in red-swamp crayfish (PcAst-1a, PcAst-1b, PcAst-1c and PcAst-2). Their phylogenetic inference suggests that four different families of astacidins exist within infraorder Astacidea (Rončević et al., 2020).

Antilipopolysaccharide factors (ALFs) are small proteins with the hydrophobic N-terminal region. Like AMPs they have a wide range of roles in defence against bacteria, fungi, and viruses. They have been observed in the wide range of crustaceans (Becking et al., 2020). In red-swamp crayfish PcALF1 was identified as highly expressed in haemocytes, hepatopancreas and intestines, where it might play a role in intestinal immunity. This ALF exhibited growth inhibiting activity towards bacterial and fungal microorganisms, as well as opsonic activities, increase the phagocytosis by haemocytes (Sun et al., 2011). Shi et al. (2013) showed that in red-swamp crayfish a 30 kDa protein with lysine motif (LysM) is involved in the regulation of the transcription of crustins and antilipopolysaccharide factors.

1.5.4.3. Toll-like receptors

Toll-like receptors are a part of the invertebrate innate immunity located on the cell membranes. They have been well described in insects where they are critical for expression of AMPs (Paro and Imler, 2016). The first Toll-like receptor in freshwater crayfish, PcToll, was identified by Wang et al. in 2015 in red-swamp crayfish. PcToll has a wide expression across the red-swamp crayfish tissues and it is speculated that it is involved in the basal levels of AMP expression. Since then in red-swamp crayfish, in addition to PcToll, five more Toll like receptors were identified: PcToll2, PcToll3, PcToll4, PcToll5 and PcToll6, with the increasing evidence of their involvement in the antimicrobial, antiviral and antifungal functions in immune related genes expression regulation (Huang et al., 2018). Immune deficiency homolog (IMD) was also identified in the red-swamp crayfish as important in regulation of AMP expression (Lan et al., 2013). Recently, tumour necrosis factor receptor-associated factor 6 (traf6) like gene was described in the red-swamp crayfish. It plays a crucial role as an adaptor and regulator of Toll signalling pathway in crayfish (Li et al., 2020).

1.5.4.4. Lectins

Lectins are proteins capable of binding carbohydrate-binding domains with high specificity. In crustaceans, lectin recognition leads to downstream activation of cellular and humoral responses, agglutination (Jin et al., 2013), endocytosis (Shi et al., 2014), encapsulation and nodule formation (Ling et al., 2006), synthesis of AMPs (Vasta, 2009) and antiviral activities (Zhao et al., 2009). In crayfish they also act as pattern recognition receptors (PRR) activating the proPO cascade (Cerenius et al., 2010). Among PRRs, C-type lectins (CTLs) have a major role in

the innate immunity of freshwater crayfish. Several CTLs were identified in red-swamp crayfish: PcLec1, PcLec3, PcLec4, PcLec5, PcLT participate in the antibacterial response, PcLec2 and Pl-MVL in the prophenoloxidase activating cascade and PcLec6 in the antibacterial as well as in antiviral response (Zhang et al., 2018). In *C. quadricarinatus* serum lectin CqL has the capacity to activate oxidative burst in the crayfish haemocytes (Sánchez-Salgado et al., 2014). At least 14 lectin families have been described, and most recently L-type lectins were recognised in kuruma shrimp involved in the antimicrobial response (Xu et al., 2020).

1.5.4.5. Down syndrome cellular adhesion molecule (Dscam)

The Down syndrome cellular adhesion molecule (Dscam) was first identified from the Down's syndrome critical region of human chromosome 21q22.2-22.3 (Yamakawa et al., 1998). It is a member of the immunoglobulin (Ig) superfamily, with a similar structure in both mammalians and invertebrates. The Dscam molecule consists of three main components, an extracellular region with several Ig and fibronectin type III domains, a transmembrane domain, and a cytoplasmic tail. Unlike its mammalian counterpart, invertebrate Dscam exhibits hypervariability in all domains achieved through a mechanism of alternative splicing during mRNA maturation (Ng et al., 2014). In crustaceans, Dscam can exist in both a membrane bound form and, in a tail-less (secreted) form, encoded on the same genomic locus, similar to the IgM in the vertebrates (Chou et al., 2011). It has been proposed that Dscam, due to its hypervariable Ig domains, is involved in recognition of diverse ligands, epitopes and pathogens. This recognition would be pathogen specific and could include creation of immune "memory" i.e. immune priming reaction in crustaceans (Ng et al., 2014). The main immune response triggered by Dscam after pathogen recognition is an increase in the phagocytosis (Watson et al., 2006). In *C. quadricarinatus,* it has been shown that following the White spot syndrome virus (WSSV) challenge, specific groups of CqDscam variants termed "CqDscam cloud" are elevated and stay elevated up to 14 days post-challenge. Individuals "primed" in that way show a higher resistance in the repeated challenge. Ig2- and Ig3- splice variants were recognised as the most important for pathogen-binding and recognition specificity. Still, it remains unclear if the "correct cloud" of CqDscam is the consequence of isoform selection or maintenance of the hemocyte subpopulation (Ng et al., 2019). Their role in the immune response to fungal infection remains unknown.

1.6. Procambarus virginalis infection experiment

Prior to this study, a controlled infection experiment was conducted with the marbled crayfish collected at Lake Singliser See, Hesse, Germany by Francesconi et al. (submitted). All females were in the similar size group (Carapace length ranging from 35 mm to 47 mm with the mean of 41.07 mm), sexually mature, with 20/30 marbled crayfish produced eggs during the course of the experiment. All crayfish were in the intermoult phase, and moulting was not recorded during the experiment. Marbled crayfish were exposed to two strains of *A. astaci*, strain As (lower virulence) and strain PsI (high virulence). A total of 60 individuals were included in the infection experiment, 20 crayfish exposed to As, 20 to PsI strain, and 20 crayfish as a control group. Three days, 21 days and 45 days after the start of the infection experiment, five crayfish per experimental group were sampled at the first two sampling point and 10 at the last sampling point. The level of *A. astaci* infection in the crayfish was estimated with a qPCR assay (Francesconi et al., submitted). Only one marbled crayfish infected with the high virulence PsI strain showed light symptoms on day 3, but all marbled crayfish survived through the infection experiment. In the PsI-challenged marbled crayfish 14/15 resulted moderately infected in qPCR assay, with ct-values between 30.87 and 42.04. None of the marbled crayfish infected with the less virulent As-strain showed any symptoms and were all negative or very lowly infected with *A. astaci.* Out of 60 crayfish in the infection experiment 30 crayfish (sampled after 3 and 21 days upon infection) were used in this thesis for RNA sequencing analysis according to the sampling scheme in **Figure 6**. Total RNA was isolated from the hepatopancreatic tissue using the NucleoSpin RNA Kit (Macherey Nagel) according to the manufacturer's protocol. The ribosomal RNA (rRNA) was removed from the library using polyA enrichment with magnetic beads. Next-generation sequencing was performed on an Illumina TruSeq platform (Novogene, UK) resulting in ~40 million, 2 x 150 bp unstranded paired-end reads per sample. Each sample represents a biological replicate, as each sample is originating from a different individual. The obtained data is the basis for this thesis. This method was previously successfully employed to study known crayfish pathogens for example: WSSV (Jiao et al., 2019), poly I:C challenge (Dai et al., 2017) and Black May disease in *P. clarkii* (Shen et al., 2020).

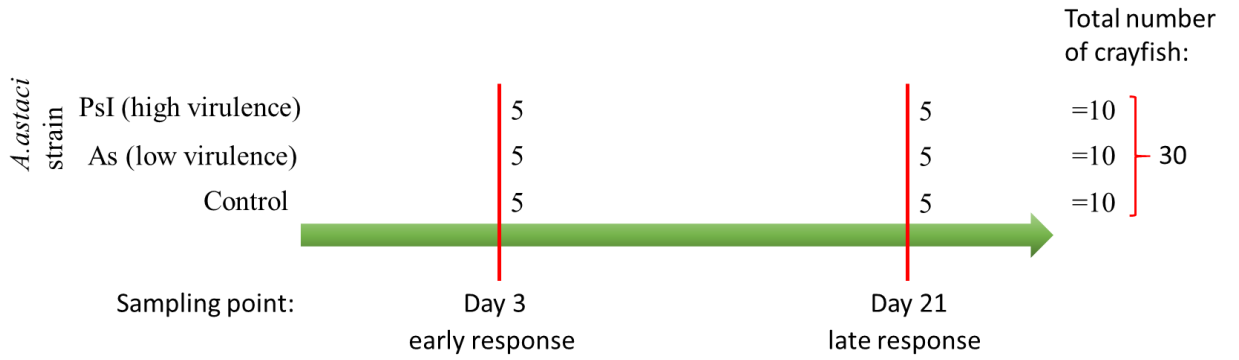

Figure 6. Sampling scheme of the *A. astaci* challenged marbled crayfish for the RNA sequencing. Crayfish were exposed to two strains of *A. astaci* (As and PsI). Two sampling points (Day 3 and Day 21, in the further text sampling point 1 and 2, respectively) are marked with a red line with the number of the sampled crayfish per each treatment. Total number of the crayfish used in the RNAseq experiment is 30**.**

2. Study aims

The general aim of this study is to improve the current knowledge of the immune system of freshwater crayfish *P. virginalis* by studying its immune response to the challenge by a known crayfish pathogen *A. astaci*. *Procambarus virginalis* is an invasive crayfish species, thus likely to show higher resistance to the pathogen challenge.

Specific aims and hypothesis of this study are:

1. *De novo* assembly and annotation of the *P. virginalis* transcriptome

   Hypothesis I: A more complete transcriptome assembly will be obtained as a result of higher sequencing coverage and large sample size.

2. Differential gene expression analysis of the crayfish exposed to the PsI (high virulent) and As (low virulent) genotypes of *A. astaci.*

   Hypothesis II: The genes related to the proPO cascade will not be differentially expressed as they are constantly up-regulated in the invasive crayfish species.

This study will provide valuable sources for the future studies of immune response in this species and other crustaceans and extend the current knowledge of the crayfish plague resistance in invasive freshwater crayfish.

3. Materials and methods

3.1. Quality assessment and filtering

All samples were assessed for quality of the raw reads in the FastQC program (Andrews S., 2010). FastQC allows identification of the problems that originate either from the sequencing or from the library preparation. As an input data for the analysis FastQC uses the reads in the FASTQ format that consist of four lines. The first line always begins with "@" character which is followed by the general information about the read, second line contains the nucleotide sequence, third line starts with a "+" and may contain additional information about the read and the last line contains the quality scores for each base in the read. Quality scores are usually reported in the Phred-33 format and range from 0 (demarked as "!") to 40 (demarked as "I"), with the best score being the latter. Each score represents a probability of the base call to be incorrect and is based on the logarithmic scale. The FastQC analysis outputs several reports for each sample, showing the overall quality of the reads and other parameters important for detection of the possible contamination and assessment of the library complexity. These reports can be summarised using a tool MultiQC (Ewels et al., 2016). MultiQC is a Python package that processes individual analysis outputs of FastQC (and some other programs that can be used in the RNAseq pipeline) and creates a summary report that includes all samples. Raw reads were then processed to remove Illumina sequencing adaptors and to remove low quality sequences in Trimmomatic (Bolger et al. 2014). The following Trimmomatic parameters were used: ILLUMINACLIP:TruSeq3-PE.fa:2:30:1, for removal of Illumina Adaptors, LEADING:15, for removal of the bases from the beginning of the read that have quality scores lower than 15, TRAILING:15, for removal of the bases form the end of the read that have quality scores lower than 15, SLIDINGWINDOW:4:20, for removal of the part of the reads when the average quality of the 4 bases within the read drops below threshold 20, and MINLEN:50 to filter the reads shorter than 50 bp. To access the quality of the reads after processing with Trimmomatic, FastQC was run followed by MultiQC.

3.2. Contamination assessment

After quality assessment and filtering, reads were checked for contamination and potential library complexity issues. The program FastQ Screen (Wingett et al., 2018) allows for identification of external contamination sources in the samples. This program utilises mapping

tools, such as Bowtie 2 (Langmead and Salzberg, 2012) used in this study, to map sample of the reads (100 000 reads) against user-defined indexed database of possible contamination sources. The possible contamination sources chosen in this study included known sequencing contamination sources provided in the default program library: *Homo sapiens* (*ftp://ftp.ensembl.org/pub/current/fasta/homo_sapiens/dna/*) and ribosomal RNA (In house costume database of FastQ Screen, GRCm38_rRNA), as well as user defined sources *P. fallax* mitogenome (NCBI accession number: NC_020021), *A. astacus* transcriptome (NCBI assembly accession: GEDF00000000), and *P. virginalis* genome (NCBI assembly accession: GCA_002838885.1) and *A. astaci* genome (NCBI assembly accession: GCA_003546785). The indexes for the user defined genomes were created using the command "bowtie2-build". Furthermore, the trimmed reads were mapped against *P. virginalis* genome to assess the library complexity and possible DNA contamination of RNAseq reads. Only two samples of RNAseq reads were randomly chosen due to the high fragmentation of the available reference genome which limited the amount of processable samples. This mapping was done using a splice-aware mapping tool STAR- **S**pliced **T**ranscripts **A**lignment to a **R**eference (Dobin et al., 2013). The algorithm of STAR consists of three steps, *ab initio* detection of splice junctions, seed searching and clustering/stitching/scoring. In the seed finding phase read sequence is mapped to the read location in the reference genome sequence if that location in the reference genome represents the longest substring of a read sequence that matches the substring in the in the reference genome (**Figure 7a**). This principle is also called Maximal Mappable Prefix (MMP) and is augmented in STAR algorithm in sequential form (**Figure 7b**). Meaning that the seed from the read will be mapped to a position in a reference genome for example near the donor splice junction and then only the remaining part of the read will be mapped to the acceptor splice site according to the MMP principle. This allows the detection of the splice junctions in a single alignment pass without any a priori knowledge of their loci and properties. In the second phase, the entire read is aligned by stitching together all the seeds aligned to the genome in the first phase using dynamic programming. The alignment of a read is then scored by a local alignment scoring scheme and the stitched combination with the best score is chosen as the best alignment of the read (**Figure 7c**). The mapping results of STAR were summarised using MutiQC tool.

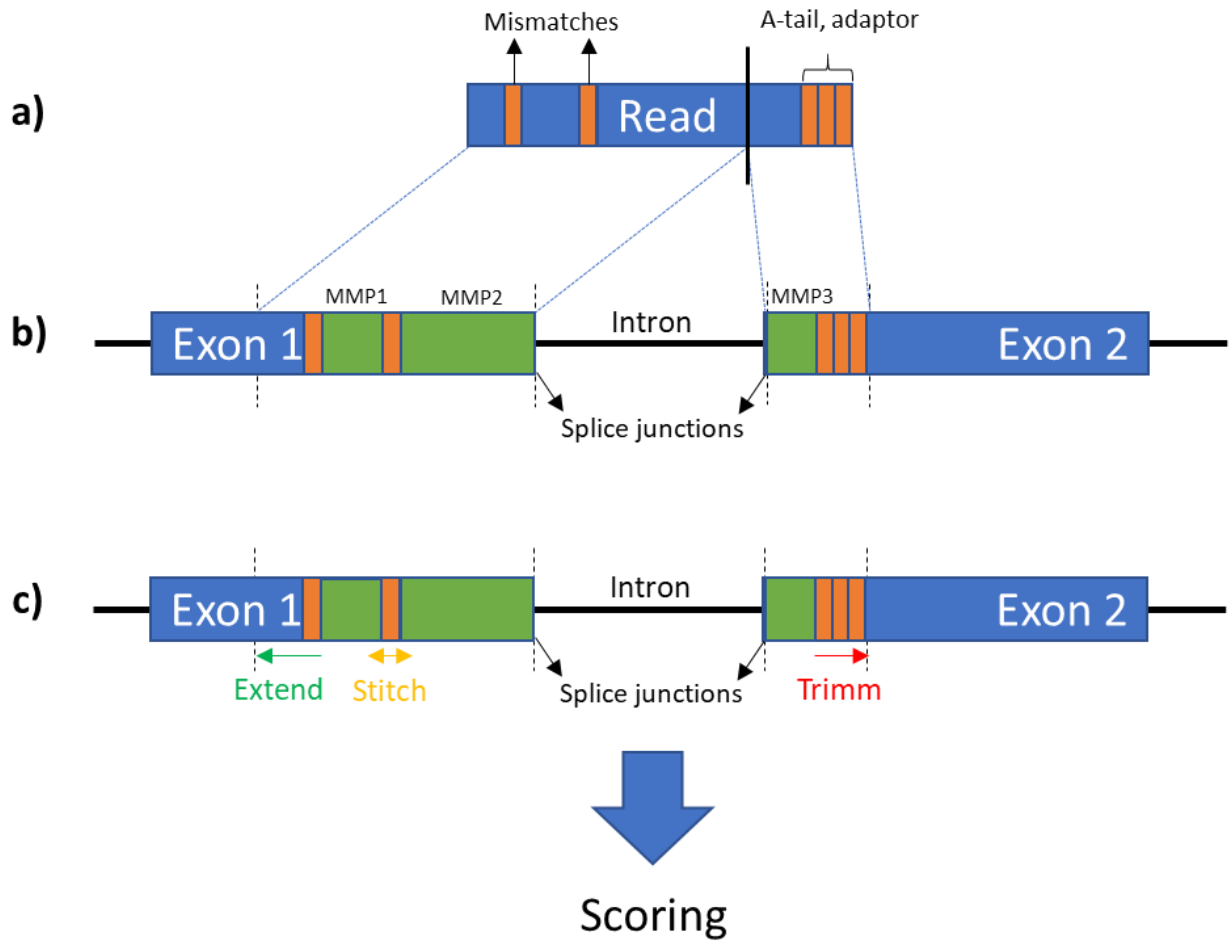

Figure 7. STAR algorithm. Three major steps are represented, **(a)** *ab initio* detection of splice junctions, **(b)** mapping of seeds according to the maximum mappable prefix algorithm **(c)** dynamic programming extension/stitching step and/or trimming followed by scoring of the read alignment. This figure was adapted according to Dobin et al., 2013.

To visualise the mapping of the reads to the genomic loci, the output in SAM format was converted to BAM format using Samtools (Li et al., 2009) and then sorted using Sambamba (Tarasov et al., 2015). After sorting, an index .bai file was created using Sambamba "index" option. These files were used as an input for Integrative Genomics Viewer (IGV) (Robinson et al., 2011) where the mapping was assessed for DNA contamination and potential library complexity issues.

3.3. De novo transcriptome assembly, quality assessment and filtering

Due to high fragmentation of currently available *P. virginalis* genome and transcriptome assembly, the *de novo* transcriptome assembly was conducted using Trinity (Grabherr et al.

2011). As an input for *de novo* assembly three samples were used each representing a different treatment (As, PsI and control), which were likely enriched in transcripts expressed in the different treatments. There are two different approaches to the transcriptome assembly that depend on the available data and resources. One of the approaches relies on the reference genome to which the RNAseq reads are firstly mapped ("mapping first") and then sequences that are overlapping with alignment are merged. The other approach uses the reads to assemble transcripts directly ("assembly first") which can subsequently be mapped to the genome, but this approach is independent of the reference genome. Trinity algorithm belongs to the latter approach, and consists of three software modules: *Inchworm, Chrysalis,* and *Butterfly* that are applied sequentially. In short, *Inchworm* utilises a greedy approach to reconstruct a single transcript representative, all reads are split into k-mers (usually 25bp in length) creating a k-mer dictionary. The most frequent k-mer in the dictionary is used as a seed and then extended in each direction by finding a highest occurring k-mer with a k-1 overlap with the current contig terminus. These k-mers, once used, are sequentially removed from the dictionary, and the whole process is repeated until the entire k-mer dictionary is exhausted (**Figure 8a**). Subsequently, contigs created by *Inchworm* are recursively grouped into connected components (with perfect overlap of k-1 bases between them), and a complete de Bruijn graph (with *k*-1 nodes, and *k* edges) is constructed in *Chrysalis*. The reads are then assigned to the component with which they share the largest number of k-mers (**Figure 8b**). Plausible full-length linear transcripts are reconstructed in *Butterfly* by matching the individual de Bruijn graphs to the original reads and paired end information. In this step the edges that are less supported (with a few reads) are removed, as they likely correspond to the sequencing errors (**Figure 8c**).

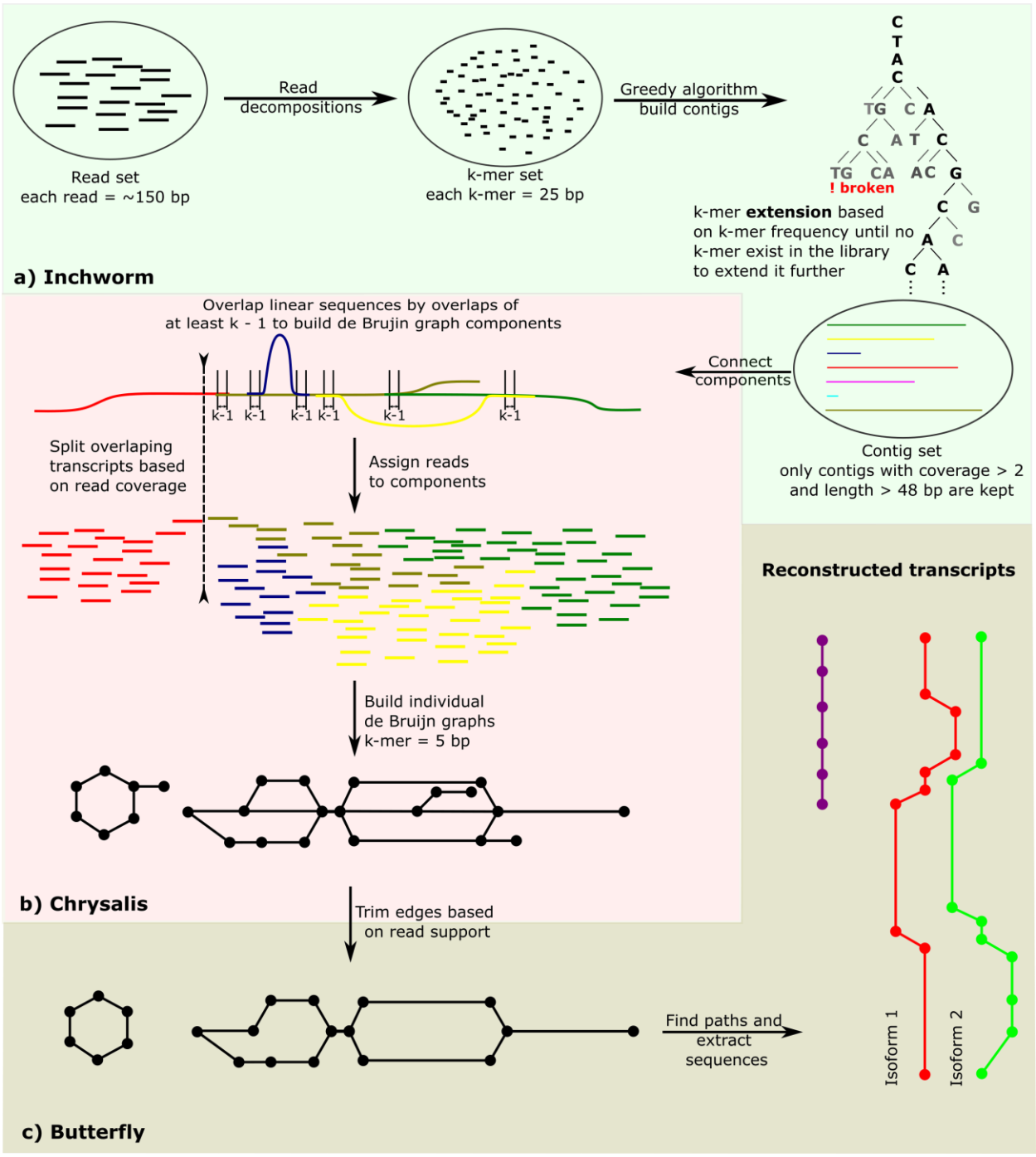

Figure 8. Trinity assembly process algorithm, consisting of three software modules: *Inchworm*, *Chrysalis*, and *Butterfly*. **(a)** In *Inchworm* reads are split into k-mers (usually 25bp in length) creating a k-mer dictionary, followed by the contig reconstruction using greedy approach. **(b)** In *Chrysalis* contigs with perfect overlap of k-1 bases are recursively grouped into connected components. Reads are then assigned to the component with which they share the largest number of k-mer, and individual de Brujin graphs are reconstructed. **(c)** In *Butterfly* edges with less read support are trimmed and graphs are compacted. The best path (based on read coverage and paired-end information) through the graph is found and individual transcripts are extracted.

The quality of assembled transcriptome was estimated by: i) examining the read representation of the assembly, ii) considering the basic assembly statistics as suggested by the Trinity pipeline, and iii) exploring the completeness according to conserved ortholog content. The first quality assessment was conducted by mapping the reads used for *de novo* assembly back to the transcriptome with Bowtie 2 (Langmead and Salzberg, 2012). The assembly statistics was estimated using the TrinityStat script provided with Trinity (Grabherr et al. 2011) and Tansrate software (Smith-Unna et al., 2016). This was followed by the BUSCO ("Benchmarking Universal Single- Copy Orthologs") search against the arthropoda_odb10 database of orthologs (n = 1013). The genes in the BUSCO databases are evolutionary conserved, protein coding genes, under "single-copy control" and are selected from the OrthoDB (Seppey, Manni and Zdobnov, 2019). BUSCO was also performed for the other available transcriptome assemblies for the freshwater crayfish obtained from the NCBI- TSA database (assembly prefix: GAFS01, GAFY01, GBEI01, GADE01, HACK01, HACB02, GBYW01, GARH01, GBEV01, GICG01, GEDF01 and *P. virginalis* assembly form the marmorkrebs.dkfz.de webserver). Filtering of short contigs (<500 bp in length), as they are unlikely to represent a full-length protein coding sequences and reads belonging to the mitochondrial RNA, was conducted in R (R Core Team, 2017). After filtering the assembly quality was re-evaluated and compared to the currently available transcriptome resources for freshwater crayfish. The filtered assembly was used for downstream analysis and annotation.

3.4. Transcriptome annotation

Following the transcriptome assembly, the functional annotation of the transcriptome was conducted using the dammit! annotation pipeline (Scott, 2016). This pipeline utilises gene model building implemented in TransDecoder (https://github.com/TransDecoder/TransDecoder/wiki) to predict protein coding regions of assembled transcripts. These predicted coding regions were matched to four databases: Pfam, OrthoDB, UniRef90 and Rfam. Pfam is a database that contains a collection of protein families, that are represented by multiple sequence alignments and hidden Markovnikov models (HMMs) (El-Gebali et al., 2018). It is searched using *HMMER* which utilises the HMMs (Wheeler and Eddy, 2013). The Pfam database is used both during the TransDecoder's open reading frame (ORF) finding and annotation assignment in dammit! pipeline. OrthoDB contains hierarchical catalogue of orthologs and it attempts to classify

proteins and trace them back to their ancestral ortholog (Krivntseva et al., 2018). UniRef90 is built upon UniRef100 database of the UniProt Reference Clusters, which provide clustered sequence set from the UniProt Knowledgebase. The UniProt database contains a collection of functional information of well annotated proteins. UniRef100 and UniRef90 contain a non-redundant set of sequences, while UniRef90 clusters sequences that have at least 90% sequence identity and an 80% overlap (Suzet et al., 2007). Both UniRef90 and OrthoDB were searched using *LAST* (Kielbasa et al., 2011), which utilises adaptive seeds as an alternative to fixed-length seeds (e.g. implemented in BLAST) to improve sensitivity and suffix arrays to surpass issues caused by repeats (Kielbasa et al., 2011). In addition to protein databases, the predicted coding regions were matched against the Rfam database, which contains RNA families represented by multiple sequence alignments, consensus secondary structures and covariance models (CMs) (Kalvari et al., 2017). These RNA molecules have a conserved secondary structure which can be accurately modelled by covariance models and matched using Infernal ("INFERence of RNA ALignment") (Nawrocki and Eddy, 2013). Annotation features included putative nucleotide and protein matches, 5'- and 3'- untranslated regions (UTRs), exons, mRNA, as well as start and stop codons. As an additional approach for functional annotation, assembled transcripts were mapped to the reference canonical Kyoto Encyclopaedia Genes and Genomes (KEGG), accessed through KEGG automatic annotation server (KAAS) (https://www.genome.jp/kegg/kaas/). Orthologs were assigned using bi-directional best hit (BBH) method with BLAST search (Kanehisa and Sato, 2020.).

3.5. Transcriptome mapping and differential gene expression analysis

All samples were mapped to the newly obtained reference transcriptome using the pseudo-alignment approach implemented in Salmon software (Patro et al., 2017). This approach differs from the other quantification tools that rely on producing the alignment with the transcriptome (un-spliced) or genome (spliced), therefore trading off the sensitivity and specificity for increased speed. Salmon was run in mapping-based mode that consists of two steps: indexing of the reference transcriptome and quantification of the reads mapped to the transcriptome. A number of "flags" were used in the Salmon mapping step to correct the biases that might originate from sequence data. The "–validateMappings" flag within Salmon utilises an improved selective alignment algorithm that mitigates some of the inherited mapping errors that

occur in the lightweight approaches. Salmon increases both the sensitivity and specificity, as it produces the mapping scoring to differentiate between the mapping loci (Srivastava et al., 2020). Salmon is also capable of correcting the sequence-specific biases in the input data that are originating from the random hexamer priming step in Illumina sequencing ("--seqBias" flag) and fragment-level GC biases in the input data (Roberts et al., 2011).

Differential gene expression analysis was conducted according to the DESeq2 protocol (Love et al., 2014) implemented in R environment (**Supplementary code 1**). In short, raw counts from the Salmon output were used as the input. Counts for individual Trinity transcript isoforms were grouped to Trinity genes to reduce the variance (noise) that occurs as the consequence of the mapping step. The count data was then prefiltered to keep the genes that have counts higher/equal to 10 across five samples. The number of samples threshold was chosen to represent the number of samples within each treatment, five and the count threshold was chosen according to the recommended filter in the pipeline. The genes with very low expression are unlikely to carry biologically important information. This step is important, because it lowers the number of multiple statistical tests that need to be performed on the dataset and increases the speed of the analysis. The exploratory analysis of the dataset included two steps, calculation of Poisson distances implemented in PoiClaClu package for heatmap construction in pheatmap package and the analysis of principal components (PCA). Distance calculation was performed after applying the regularized-logarithm transformation (rlog) to reduce the impact of highly expressed genes (i.e. produce a more homoscedastic dataset). Alongside vst transformation the variance stabilizing transformation (vst) and log2 (x+1) transformations were also performed to inspect the data distribution. Following the exploratory analysis, the differential expression (DE) analysis was performed on the filtered raw counts, by applying the DESeq function on the dataset. Following the results of the exploratory analysis a new variable was introduced to account for the reproducing females in the differential gene expression analysis using DESeq2, resulting in the following design for the first sampling point: ~reproduction + groups (treatments), and for the second sampling point: ~groups.

The following comparisons were conducted: Control vs. PsI (time point I), Control vs. PsI (time point II), Control vs. As (time point I), Control vs. As (time point II). Package "EnhancedVolcano" (Blighe et al., 2020) was used for the visualisation of the differentially expressed genes and package "apeglm" for noise removal (Zhu, Ibrahim and Love, 2018). The

list of differentially expressed genes was exported and their counts, log2fold changes and p-values were merged. For each DE gene, dammit! annotation as well as KEGG annotation were assigned and manually inspected. Possible overlaps between the differentially expressed genes at different time points were inspected using Venn diagrams.

3.6. Over-representation analysis of KEGG pathways and inspection of gene ontology terms associated with the DE gene subset

Over-representation analysis of the DE genes with KEGG (KO number) identifier was conducted using the enrichKEGG() function (p-value <0.05) with the option organism: ”ko”, of the ClusterProfiler package (Yu et al., 2012) implemented in R. The resulting over-represented KEGG pathways were illustrated as a barplot graph. KEGG over-representation analysis was also conducted for the whole transcriptome assembly for the set of transcripts that received the KEGG annotation. To obtain the gene ontology (GO) annotation for the DE gene subset, nucleotide sequences of DE genes were imported to the InterPro5 (Jones et al., 2014) with the “-iprlookup” and “-goterms” flags. Gene ontology represents a controlled vocabulary for functional annotation of the nucleotide sequences. Three different aspects are explaining the biological domain of the transcript: molecular function (explains the activity of the gene products at molecular level), cellular component (gives a localisation of the gene product) and biological process (assigns a gene product to a larger biological process context) (Ashburner et al., 2000). Unique differentially expressed transcripts, each representing one Trinity gene that received GO annotation, were imported to WEGO (Web Gene Ontology Annotation Plot) platform (Ye et al., 2018) for the results visualisation. This allowed for comparison in the gene number and percentages representing different GO terms.

4. Results

4.1. Read pre-processing and contamination assessment

The initial assessment of raw reads using FastQC revealed that the average duplication rates of the reads ranged from 71.7% to 81% (average 77.29%), the % GC ranged from 44% to 46% (average 45.18%), the average read length of all samples was 150 bp and the number of reads ranged from 40.6 M to 68.9 M (average 51.64 M) (details in **Supplementary table 1**). The analysis of raw reads revealed the presence of the Illumina sequencing adaptors in the dataset (**Figure 9a**). The adaptors were removed using Trimmomatic and no samples were found with any adapter contamination > 0.1% (not shown). In all samples it was observed that the first 14 bases contain a position specific bias of the nucleotide proportion. This bias was not removed from the samples as it represents the artefact in the sequencing libraries produced by random priming (**Figure 9b**).

After trimming and adaptor removal the average duplication rates of the processed reads ranged from 72.3% to 81.8% (average 77.85%), the % GC ranged from 44% to 46% (average 44.95%), the average read length of all samples was 147.95 bp and the number of reads ranged from 38.0 M to 65.0 M (average 48.78 M) (details in **Supplementary table 1**).

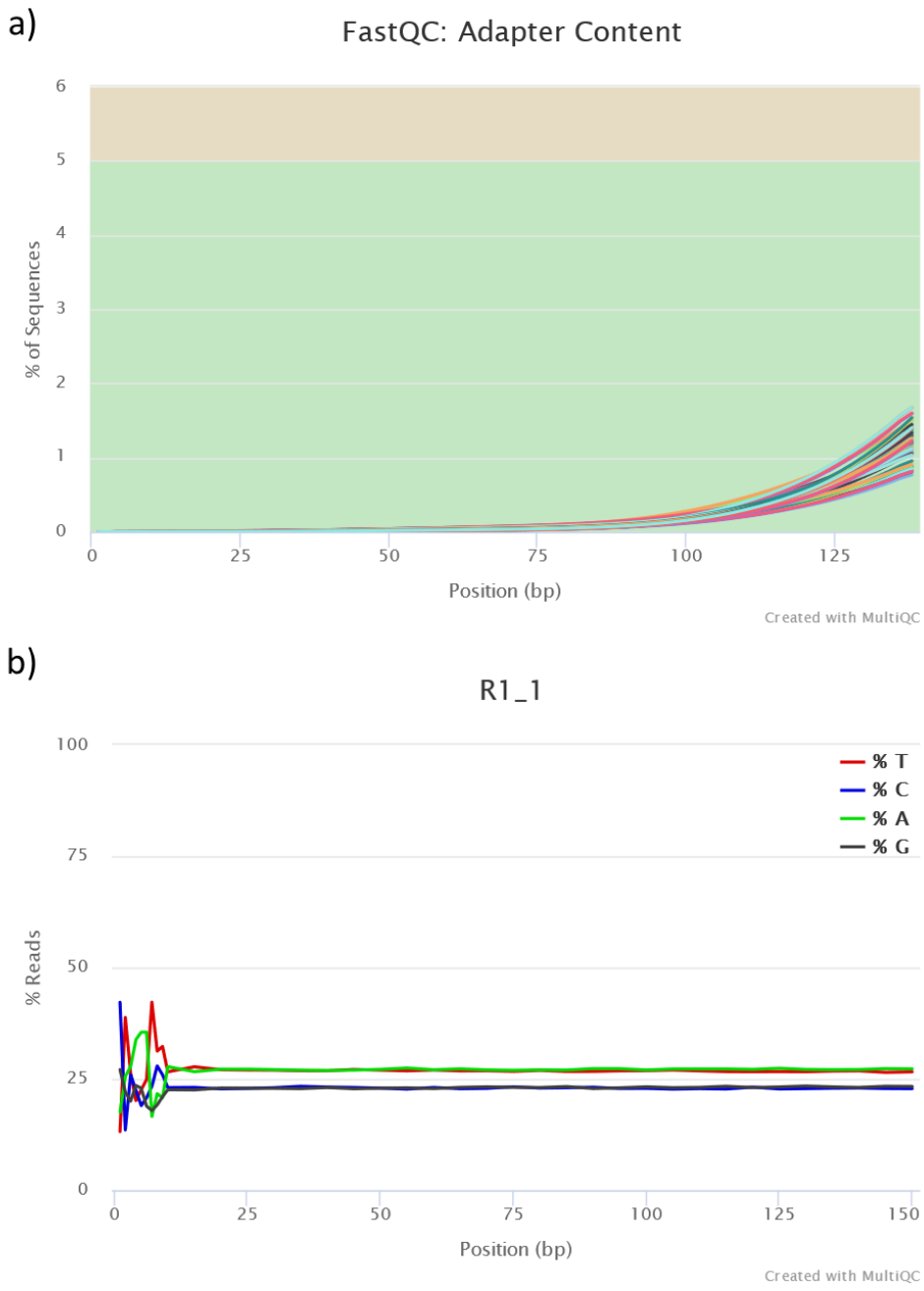

Figure 9. The results of the raw sequence analysis. **(a)** On the left, adaptor content within the reads is shown for each sample **(b)** and on the right the proportion of each base per position in the read is shown for sample R1_1. Plots were produced with the MultiQC software (Ewels et al., 2016).

Analysis of per sequence quality scores and per base mean quality scores revealed the high quality of the sequenced reads (**Figure 10a, b**). The per sequence GC scores showed a bias of sequence to the left (lower GC content) across all samples. The graphs also contained two broad peaks, possibly indicative of contamination with the DNA of other organisms. To reveal possible contamination FastQ Screen analysis (Wingett et al., 2018) was conducted (**Figure 11**).

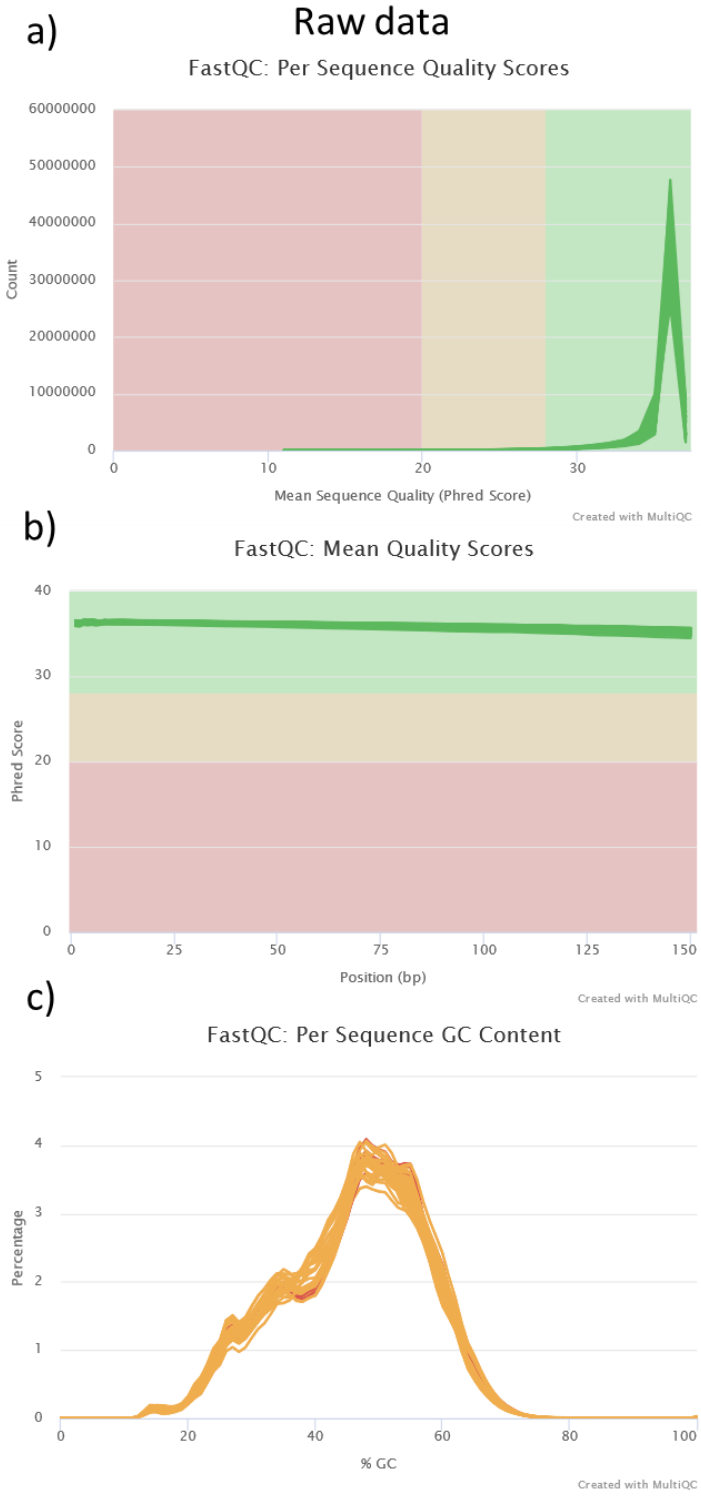

Figure 10. The results of the sample quality assessment of the raw data. **(a)** On the left the per sequence quality scores are shown, **(b)** in the middle mean quality scores are shown **(c)** and on the right per sequence GC content is shown. Plots were produced with the MultiQC software (Ewels et al., 2016).

Although more than 98% of the reads could be assigned to the *P. virginalis* genome, many of them (6% - 10%) could be assigned to the mitochondrial genome of *P. fallax*. Contamination originating from *A. astaci* DNA and rRNA was not detected (or was less than 1%) in all the samples. Around 25% of the contigs were multi mappers between *A. astacus* transcriptome and *P. virginalis* genome. In some samples, low percentage (less than 0.5%) of reads multi mapped between the *H. sapiens* and *P. virginalis* genome. Apart from the *P. virginalis*, transcripts that mapped to only one genome were not detected. It was also observed that some of the transcripts (~ 1.5%) are unique and could not be mapped to the *P. virginalis* genome.

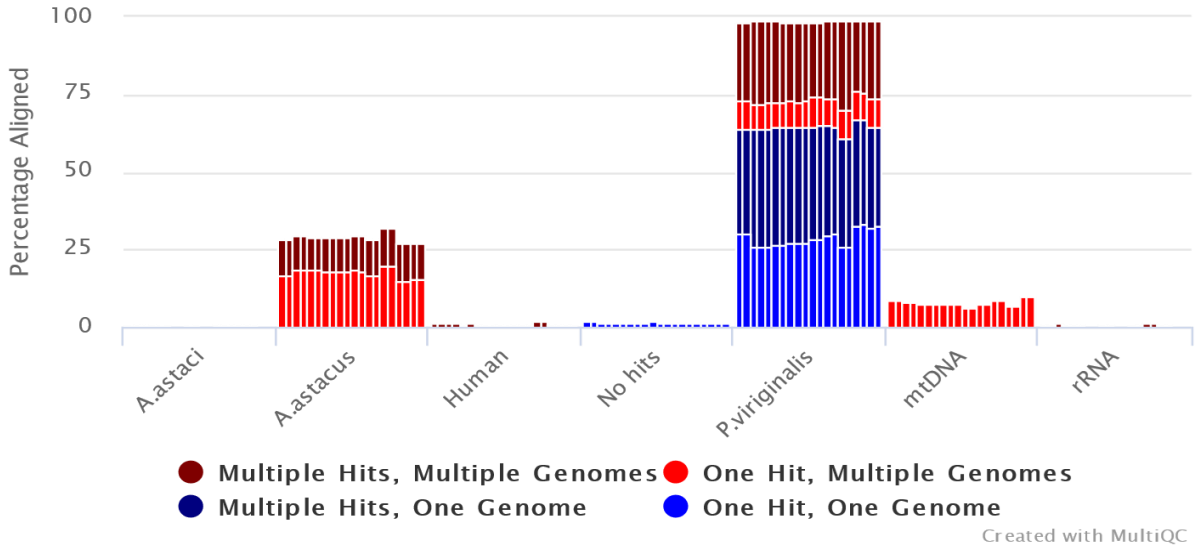

Figure 11. FastQ Screen (Wingett et al., 2018) results, on the dataset of processed reads belonging to the As treatment. Reads were screened against the genome of *A. astaci*, the transcriptome of *A. astacus*, the genome of *Homo sapiens* (human), the genome of *P. virginalis*, the *P. fallax* mitochondrial genome (mtDNA) and the internal rRNA database. Transcripts unique for one genome are coloured in blue, and transcripts shared across multiple genomes are coloured red. The number of mapped transcripts is shown in percentages (%) of the reads that mapped to the database. This plot was produced with the MultiQC software (Ewels et al., 2016).

The mapping results were congruent with the mapping results obtained using FastQScreen, as 74.1% and 77.4% of the reads mapped uniquely and 15.6% and 12.2% mapped to multiple loci against the genome of the *P. virginalis* (**Figure 12a**). Reads mapped to too many loci represented 0.2% for both samples and unmapped reads (too short) represented 10.00% and 10.1%, while unmapped (other) represented 0.2% and 0.1%, respectively. The visual inspection

of the mapped reads neither showed issues with the library complexity (i.e., columns of the some read sequenced multiple times) nor with DNA contamination (**Figure 12b**).

Figure 12. (a) Results of the read mapping to *P. virginalis* genome with STAR (Dobin et al., 2013) for the control samples R21 and R22. This plot was produced by the MultiQC software (Ewels et al., 2016). **(b)** The visual inspection of the mapping with IGV (Robinson et al., 2011). The reads are represented as grey arrows, across the 5550 bp region of contig SEQ1:57450:63000. Mismatched bases in individual reads are coloured, the intron spanning reads relate to a light blue line.

4.2. Procambarus virginalis transcriptome assembly

The results of transcriptome assembly with Trinity are shown in **Table 1**. The final assembly was constructed from 142.1 million of the paired-end reads and contained 111,770 contigs representing 60,004 genes. The minimal contig length was 501 bp and the maximal contig length was 32579 bp, with N50 value of 3039 bp. The average mapping rate with Salmon (Patro et al., 2017) was 86,21%. BUSCO analyses showed 93.1% completeness for the final (processed) assembly, with 1.3% fragmented BUSCOs and 5.6% missing BUSCOs. The BUSCO assembly statistics was compared to 12 currently available transcriptome assemblies of the freshwater crayfish of infraorder Astacidea. The completeness of these assemblies was as follows: 82.8% for *P. virginalis* (Gutekunst et al., 2018), 82.0 % for *A. pallipes* (Grandjean et al., 2020), 83.3 % for *A. astacus* (Theissinger et al., 2016), 83.8 %, 61.7 % and 84.1 % for *P. leptodactylus* (Tom et al., 2013, Manafrin et al., 2013, Mosco et al., unpublished, respectively), 14.2 %, 90.9 % and 88.9 % for *C. quadricarinatus* (Glazer et al., 2013, Tan et al., unpublished 1 and 2, respectively), 13.7% for *P. leniusculus* (Bunikis and Soderhall, unpublished), 90.9% and 84.3% for *P. clarkii* (Manfrin et al., 2015, Tom et al., 2014) (**Figure 13**).

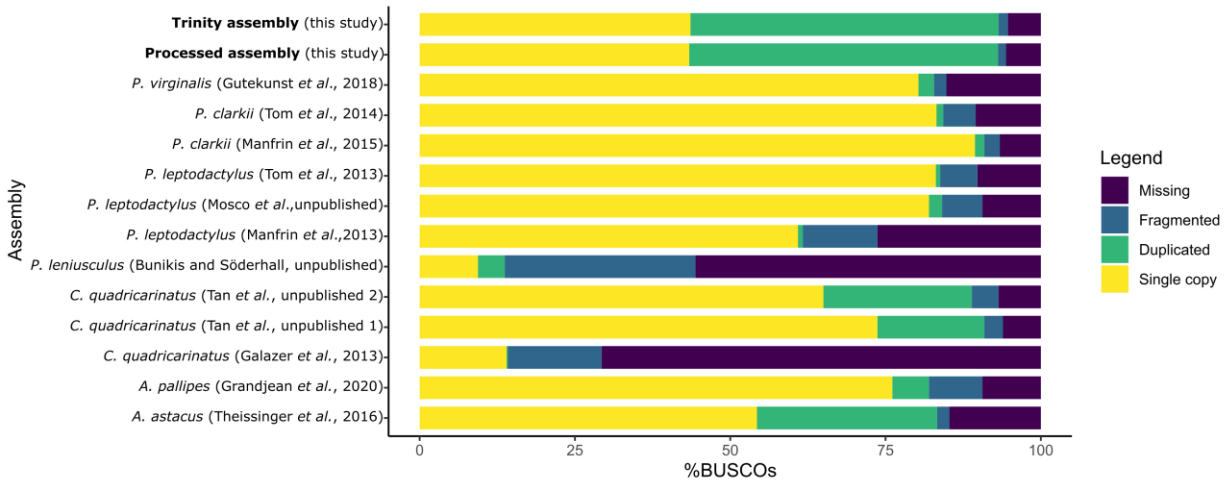

Figure 13. Benchmarking Universal Single-Copy Orthologs (BUSCO) results. Comparison of the transcriptome assemblies obtained in this study (Trinity assembly and Processed assembly) with the publicly available transcriptome assemblies for the 12 freshwater crayfish of infraorder Astacidea. For each species N=1013 single copy arthropod orthologs were searched. Percentage in yellow (single copy BUSCOs) and green (duplicated BUSCOs) indicate complete orthologs detected, and percentage in blue and purple indicate fragmented and missing BUSCOs, respectively.

Table 1. Assembly statistics calculated by TrinityStat script and TransRate program. The results for the Trinity assembly and Processed assembly are shown.

	Trinity assembly	**Processed assembly**
Total Trinity genes	194279	60004
Total Trinity transcripts	284417	111770
G+C content (%)	41.42	40.94
Contig min (bp)	169	501
Contig max (bp)	32579	32579
Mean contig length (bp)	946.97	1929.98
Contigs >1k	62257	62252
Contigs >10k	735	735
Contigs N90	321	766
Contigs N70	863	1783
Contigs N50	2251	3039
Contigs N30	4022	4697
Contigs N10	7178	7803
Mean mapping rate (%)	95.61	86.21

Gene model building using Transdecoder predicted 74,316 (66.49% of the total number of transcripts) coding regions *P. virginalis*. Predicted coding regions were matched to the protein family databases Pfam, OrthoDB and UniRef90 and non-coding RNA database (Rfam) databases. In total, 46,834 (41.9%, Pfam = 29,881, OrthoDB = 44,722, UniRef90 = 6538 and Rfam = 636) of the initial (processed assembly) transcripts were annotated when combining results of all searches for *P. virginalis* (**Figure 14a**). Annotation features include putative nucleotide and protein matches, 5'- and 3'- UTRs, exons, mRNA, as well as start and stop codons. As an additional approach for functional annotation and categorisation assembled transcripts were mapped to the reference canonical KEGG database and 16,207 transcripts were annotated across 401 KEGG pathways. Among the represented pathways, metabolic pathways (829) and pathways of biosynthesis of secondary metabolites were the most represented (**Figure 14b**).

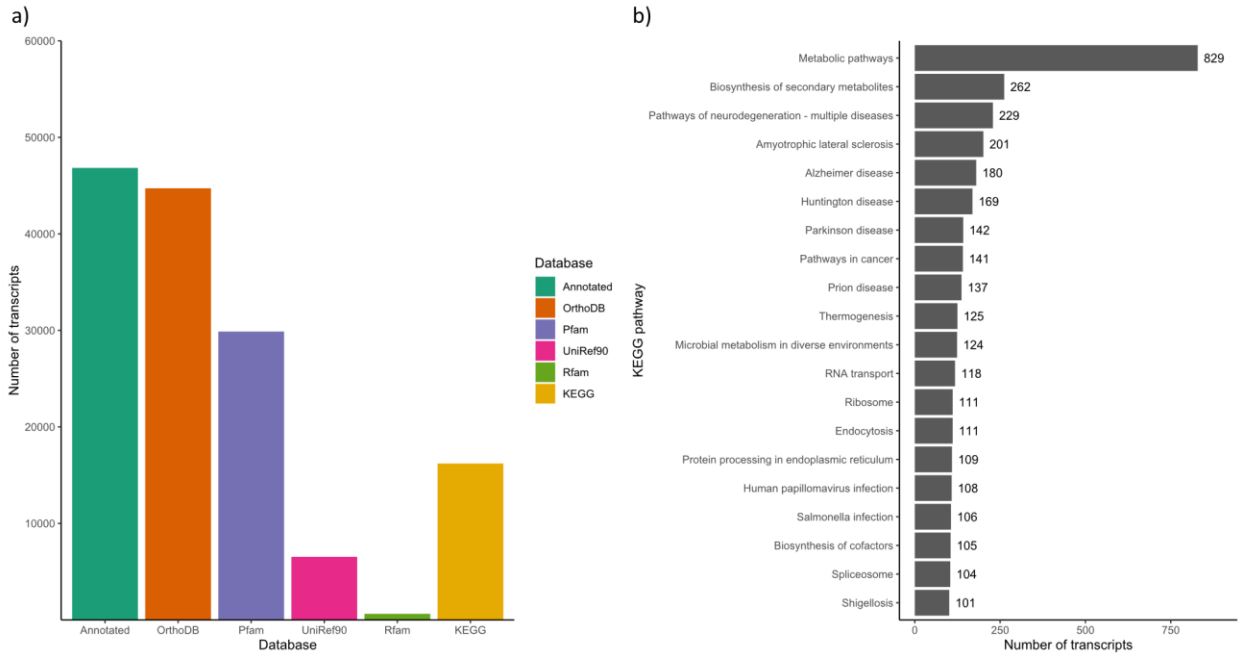

Figure 14. Annotation results for *P. virginalis* processed transcriptome. **(a)** Number of annotated coding transcripts (predicted by Transdecoder) with dammit! pipeline per database OrthoDB, Pfam, UniRef90 and KEGG. **(b)** KEGG Classification of transcripts genes. The top 20 most abundant KEGG pathways in the *P. virginals* transcriptome assembly are shown.

4.3. Gene expression patterns in hepatopancreatic tissue under A. astaci challenge

Transcript counts were obtained by mapping the processed reads to the reference transcriptome with Salmon. Individual counts for Trinity transcripts were grouped to Trinity genes. The count table included all 60,004 Trinity genes, with 36,603 genes passing the initial expression filtering step (counts higher or equal than 10 across 5 samples) applied on the whole dataset. The initial exploratory analysis of samples was conducted on the rlog transformed dataset. Regularized-logarithm transformation was chosen because it showed better performance in controlling between sample variances than vst and log2(x+1) transformation (**Figure 15a**). Exploratory analysis of Euclidean distances between samples revealed undelaying structure in the samples (**Figure 15b**). This structure was confirmed in the analysis of principal components (**Figure 15c**). The data indicates two groups, which are not following pattern caused by *A. astaci* challenge or potential differences between the sampling points. Seven samples (sample: 9, 48, 47, 8, 46, 26, 49) grouped in the bottom right corner of the heat map/top right corner of the principal component analysis. Thorough investigation of the research notes revealed that these samples are connected to the parthenogenic marble crayfish that did not undergo the process of reproduction during the infection experiment.

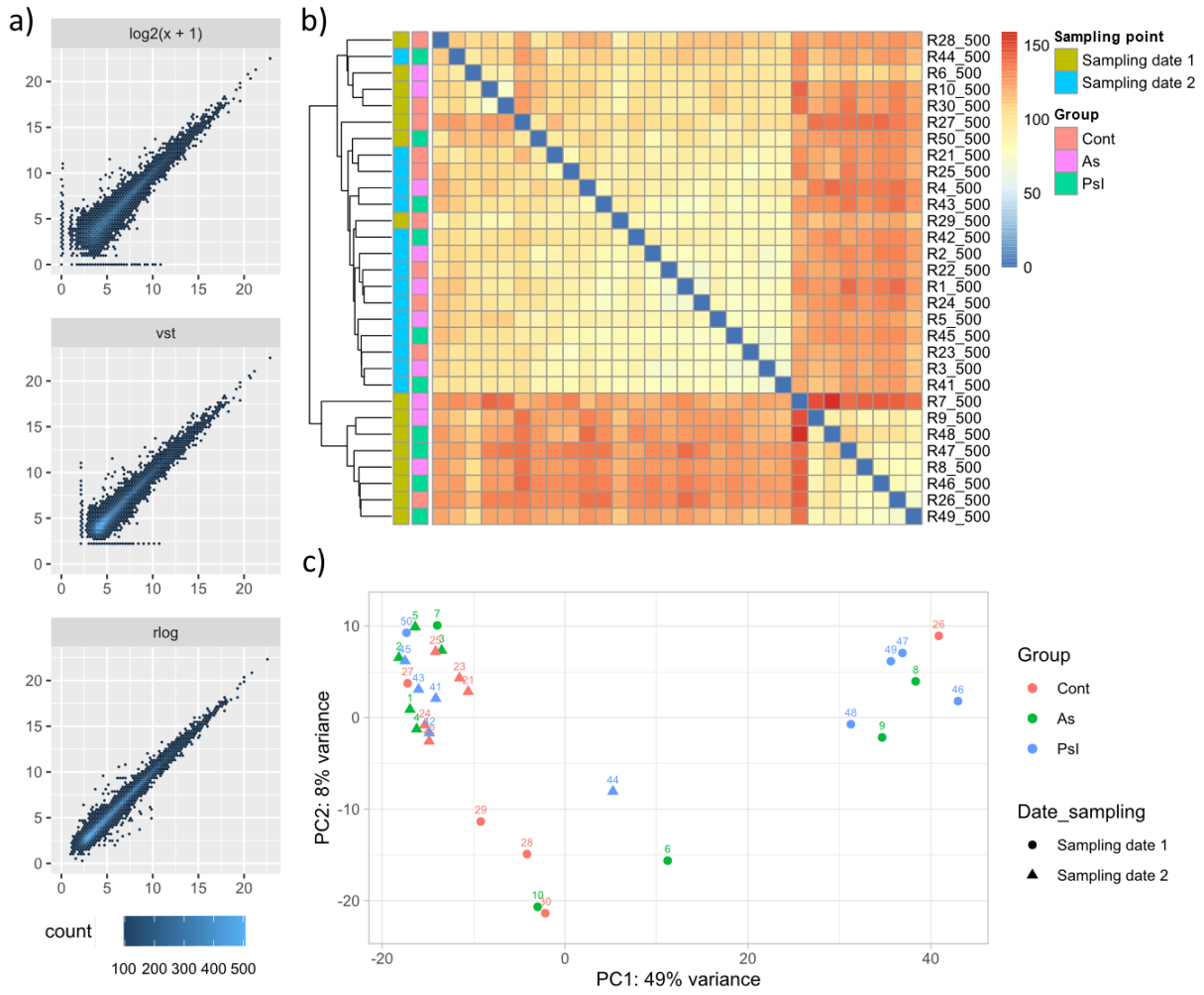

Figure 15. Results of the exploratory analysis. The scatterplots of differences between the transformed counts form sample 1 (vertical axis) and sample 2 (horizontal axis). **(a)** Three transformations were used: log2(x+1) transformation of normalized counts (top), variance stabilising transformation (vst) (middle), and regularised log (rlog) (bottom). **(b)** Heatmap produced from the rlog transformed data. The colour scale represents Euclidean distances between individual samples, clustering dendrogram is shown on the left. **(c)** Graph representing results of the principal component analysis (PCA) on the rlog transformed dataset.

Additional adjustment of the DESeq2 pipeline was needed to remove the reproduction batch effect that occurred in samples belonging to the first sampling point. Exploratory analysis was repeated, after applying removeBatchEffect() function implemented in limma R package (Ritchie et al., 2015) on the rlog and vst transformed dataset (**Figure 16.**) Thus, PCA plot and Euclidean distance heatmap deprived of visible batch effects were obtained. Samples 7 and 8 of the first sampling point dataset (both samples belonging to the As treatment) showed the highest variance from the rest of the samples (**Figure 16b, c**). All samples of the second sampling point grouped together in the lower left corner of the PCA plot, apart from the sample 40.

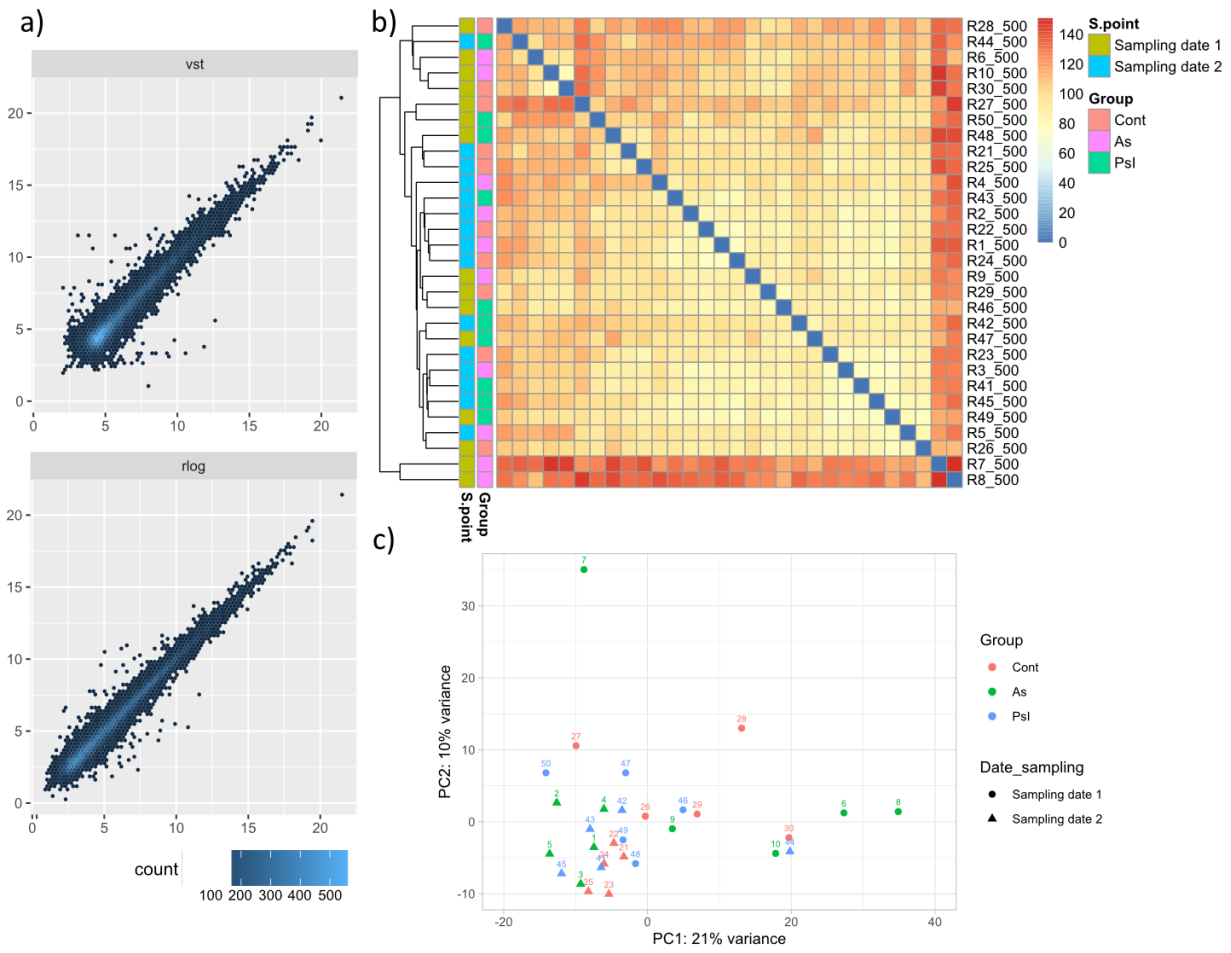

Figure 16. Results of the exploratory analysis on the datasets with removed batch effects. The variance of caused by the reproducing marbled crayfish females was controlled using removeBatchEffect() function implemented in limma R package (Ritchie et al., 2015) . **(a)** The scatterplots of differences between the transformed counts form sample 1 (vertical axis) and sample 2 (horizontal axis). Two transformations were used the variance stabilising transformation (vst) (top) and regularised log (rlog) (bottom). **(b)** Heatmap produced from the rlog transformed data with batch effect removal. The colour scale represents Euclidean distances between individual samples, clustering dendrogram is shown on the left. **(c)** Graph representing results of the principal component analysis (PCA) on the rlog transformed dataset with batch effect removal.

Results of DE analysis conducted in DESeq2 (Love et al., 2014) are shown on the **Figure 17**. For the first sampling point 35,092 genes (above the expression threshold of 10 in at least 5 samples) were analysed, with four DE genes in the As treatment and nine DE genes in the PsI treatment at the p-value= 0.05. Among them only two genes in the As treatment and four genes in the PsI treatment (for the sampling point 1) exceeded the fold change threshold of 2 (meaning that their expression is two times higher/lower than expression in control samples). For the second sampling point 34486 genes were analysed, with 33 DE genes in the As treatment and 71 DE genes in the PsI treatment. Mapping of the Trinity genes back to Trinity transcripts resulted in 20

(4 As and 16 PsI) DE trinity transcripts for sampling point 1 and 303 (95 As and 208 PsI) DE trinity transcripts for sampling point 2.

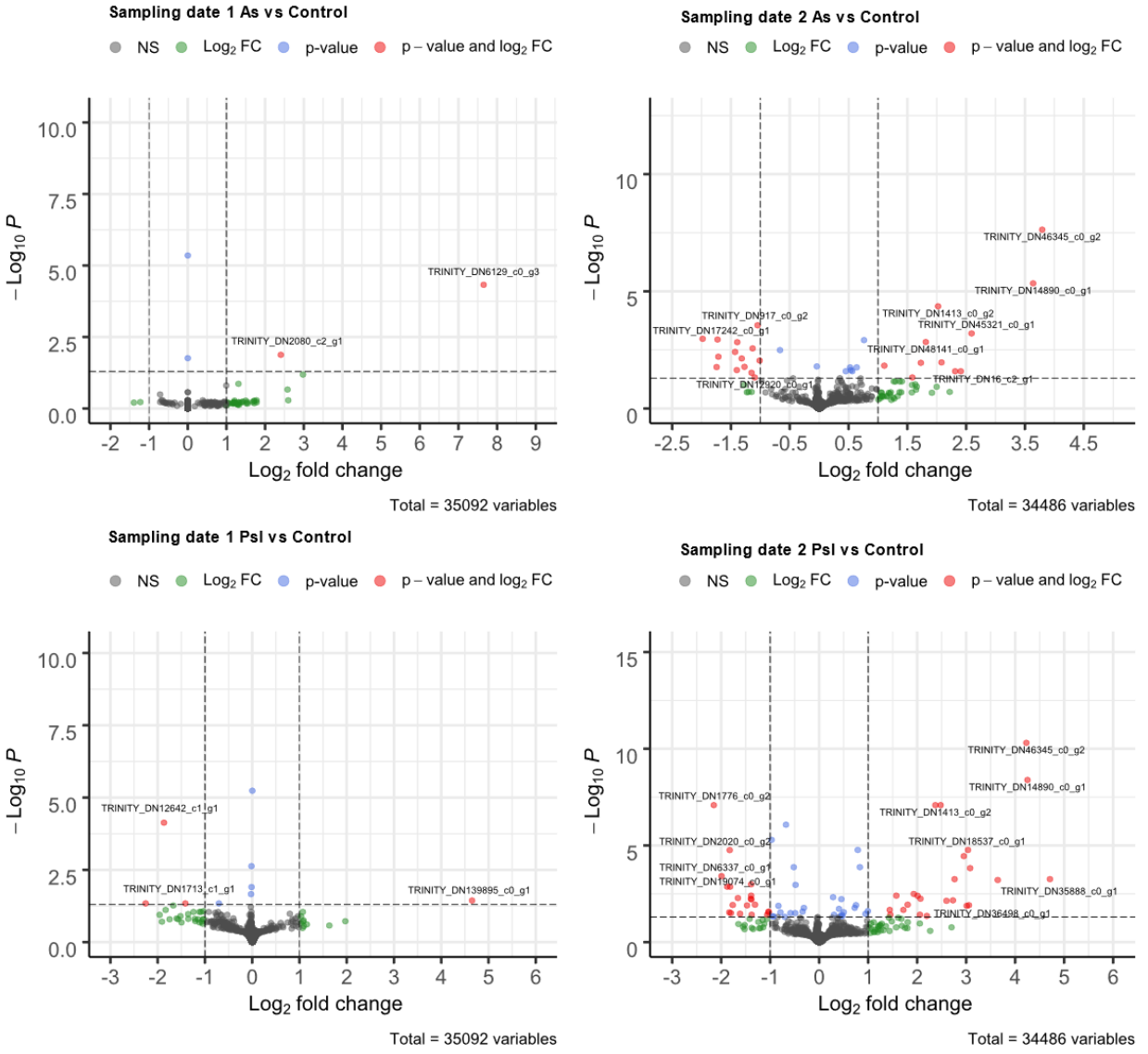

Figure 17. Results of the differential gene expression analysis represented as volcano plots. Volcano plots show the distribution of the genes based on the -$\log_{10}$P (y- axis) and $\log_2$fold (x- axis). The threshold values are represented as lines dashed lines (p-value=0.05, Fold change=2). Genes that are above both fold change and p-value threshold are coloured red. The figures are produced with EnhancedVulcano R package (Blighe et al., 2020).

In total 102 DE genes were discovered. One DE gene was expressed in both As and PsI challenged samples of the first sampling point, and 14 DE genes were observed expressed in both As and PsI challenged samples of the second sampling point. DE genes shared among sampling point have not been observed (**Figure 18**).

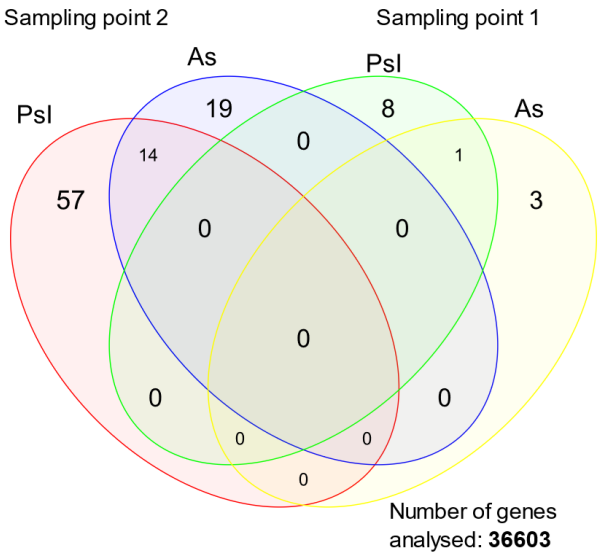

Figure 18. Venn diagram representing differentially expressed (DE) genes in all treatments. On the right DE genes of the first sampling point (high virulent-PsI; green, low virulent-As: yellow) are shown and on the left side DE expressed genes of the second sampling point (high virulent-PsI; red, low virulent-As; blue) are shown. In total across two sampling points 36,603 genes were tested for differential expression.

4.4. Annotation of differentially expressed genes in marbled crayfish under A. astaci challenge

Differentially expressed genes for each sampling point are presented in Supplementary table 2.

4.4.1. Sampling point 1

In the first sampling point in both As and PsI treatment groups, there were no transcripts with annotation related to the crayfish immune response (**Supplementary table 2**). There was no available annotation for transcripts in the As treatment of the first sampling point. For the PsI treatment three DE transcripts had no available annotation and one transcript was described as uncharacterised protein. Among the annotated transcripts of the PSI group viral RNA-dependent RNA polymerase was detected as most up-regulated and organic anion-transporting polypeptide and Pleckstrin homology domain containing protein as most down-regulated. Immune related C-type lectin was also detected among the down-regulated DE genes of PsI treatment.

4.4.2. Sampling point 2

Among the DE genes of the As treatment 12 DE genes were not annotated through the dammit! pipeline. Among the annotated genes seven genes are part of the ribosomal subunits, six among the most up-regulated and one down-regulated. Up-regulated genes with a known role in the crayfish immune system were annotated as the Cytochrome P450 and regulation of nitric-oxide synthase activity, and one gene (colony stimulating factor 3 receptor (granulocyte)) was down-regulated. Genes associated with the cell cytoskeleton (Titin isoform X7, Mucin like domain containing protein and Microtubule associated protein 1b) were also detected as down-regulated. The 22 DE genes in the PsI treatment were not annotated through the dammit! pipeline. Among the annotated genes eight genes are part of the ribosomal subunits and up-regulated. Up-regulated genes with known role in the crayfish immune system were annotated as the Serine protease inhibitor, crustin, lectin C as up-regulated and uncharacterised protein with Immunoglobulin domains and colony stimulating factor 3 receptor (granulocyte) as down-regulated. Genes associated with the cell cytoskeleton (Intermediate filament protein, Titin-like protein, Invertebrate connectin, Mucin like domain containing protein and Microtubule associated protein 1b) were detected as down-regulated. Several transcripts with L1 retrotransposon domains were detected among the most up-regulated DE genes. DE involved in

the guanylate metabolism (guanylate synthetase and guanylate binding protein) with GDP as final product were also up-regulated.

4.5. KEGG and GO over-representation analysis of differentially expressed genes

For the first sampling point one DE gene (PsI treatment) and for the second sampling point 16 DE genes (five As and 11 PsI treatment) were annotated through KEGG database. GO annotation was assigned to three DE of the first sampling point (PsI treatment) and 21 DE genes of the second sampling point (14 PsI and 7 As). Binding and catalytic activity were the most represented molecular fucntion GO terms among the samples of the second sampling point. The most represented biological pathways were metabolic and cellular process (**Figure 19b, c**). Binding, catalitic activity as well as metabolic process were also represented in the DE genes of the As sample of the first sampling point. Analysis of over-represented KEGG terms on the whole processed transcriptome assembly revealed neurodegenerative pathways and conditions (Amyotrophic lateral sclerosis, Huntington disease, Pathways of neurodegeneration, Parkinson disease, Alzheimer disease) as the most over-represented in the assembly. Pathways and conditions important for the hepatopancreas tissue (Protein processing in ER, Thermogenesis, Endocytosis, Autophagy, Lysosome, Peroxisome) were also revealed as over-represented (**Figure 19d**). Over-representation analysis of DE genes of the PsI treatment (second sampling point) showed similar pattern of overrepresented terms related to the neurodegenerative conditions as the analysis of whole assembly (**Figure 19f**). Pathways over-represented among the DE genes of the As treatment (second sampling point) were enriched by 1 gene (**Figure 19e**).

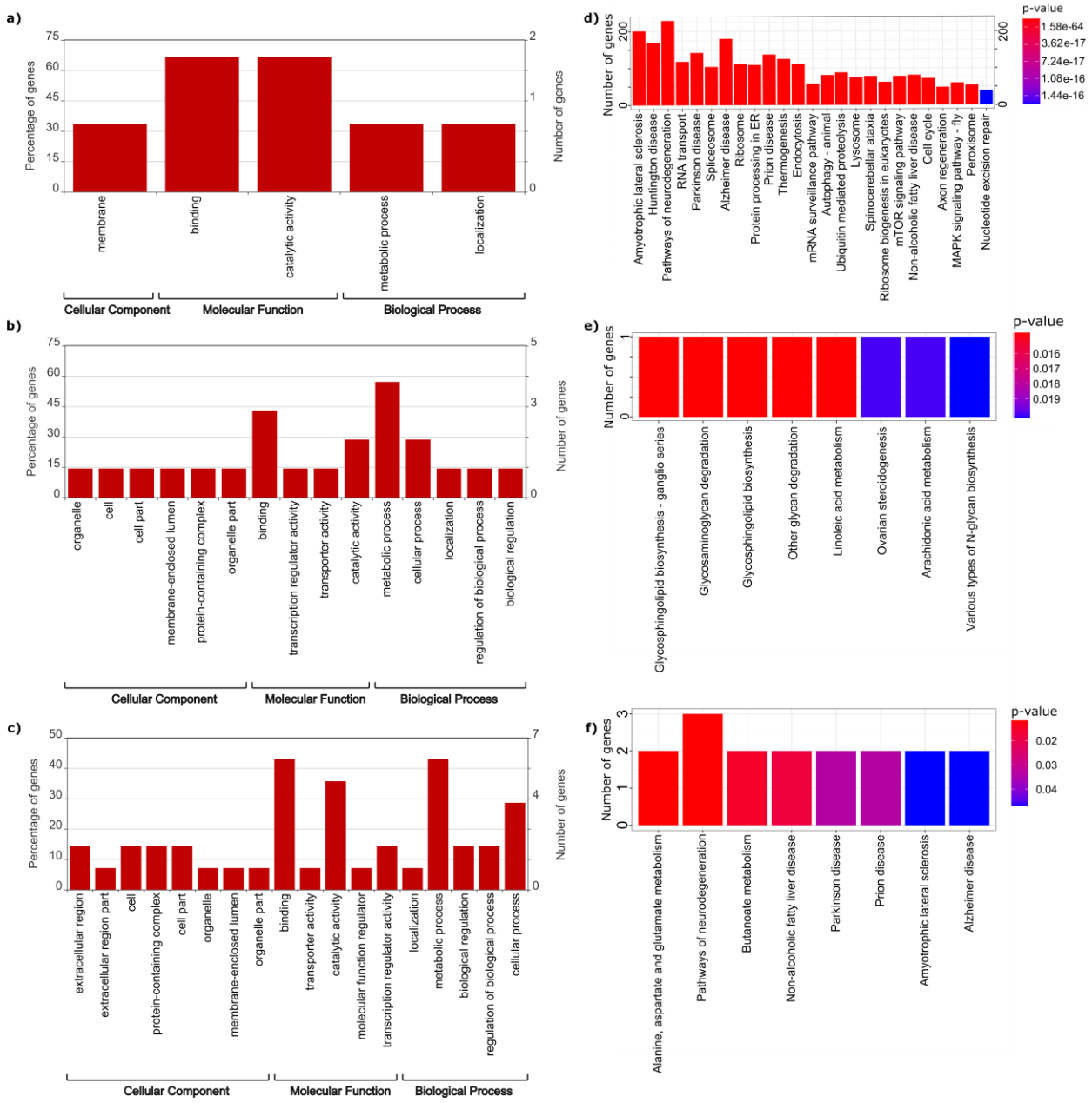

Figure 19. Representation of Gene onthology (GO) terms in the DE genes, conducted using WEGO (Ye et al., 2018) and results of the KEGG over-representation analysis conducted in R. **(a)** GO representation for the PsI treatment- sampling point 1, **(b)** GO representation for the As treatment- sampling point 2, **(c)** GO representation for the PSI treatment sampling point 2, **(d)** top 20 KEGG over-represented pathways (p-value < 0.05) in thewhole transcriptome assembly of hepatopancreas tissue, **(e)** KEGG over represented pathways in the DE genes of As treatment- sampling point 2, **(f)** KEGG over represented pathways in the DE genes of PsI treatment at sampling point 2. 11

5. Discussion

The overarching goal of this study was to uncover the molecular effects of an *A. astaci* infection on the invasive marbled crayfish. The marbled crayfish had been exposed to two strains of *A. astaci*, the low virulent As strain and the high virulent PsI strain. Initial results of the infection experiment showed that the marbled crayfish can successfully cope with the active *A. astaci* infection (Francesconi et al., submitted), as was previously observed in wild *P. virginalis* populations (Keller et al., 2014). To gain a better understanding of the disease resistance mechanisms, samples obtained from the hepatopancreatic tissue were RNA-sequenced for subsequent gene expression analyses. With the *de novo* assembly approach implemented in Trinity a high-quality transcriptome of marbled crayfish was recovered. For the first time gene expression patterns between *A. astaci* infected and uninfected invasive marbled crayfish were compared, revealing a total of 102 DE genes across 4 conditions. Among them, genes related to the proPO cascade activation were not up- regulated confirming the initial hypothesis of the study.

5.1 Transcriptome assembly

Overall quality assessment of the reads was the first step in this study. Initial results of quality assessment revealed the high quality of sequenced reads. Raw read evaluation revealed the sequence bias in the first 14 bases of reads (**Figure 9b**). This is the consequence of the random hexamer priming. Random hexamers should all be presented with uniform base frequency in the priming mix and have an equal efficiency, but this is often not the case. Removal of the first 14 bases would not have fixed this issue as the biased bases within priming hexamers are present within the reads themselves, so their removal across all reads would lead to loss of almost 10% of the data. Furthermore, quantification programs like Salmon are capable of accounting for this bias (Patro et al., 2017). High levels of duplicate reads could indicate issues with the library complexity. They are also expected in the RNAseq libraries, as multiple copies of the same transcript are present at the same time. No contamination of the reads was observed. Observation of the multi mapping reads between *A. astacus* transcriptome and *P. virginalis* genome (~25%) is expected as they are closely related species. Unique transcripts not mapped to

the *P. virginalis* genome were also observed and they could represent novel, previously unidentified transcripts.

One of the aims of this study was to obtain a more complete transcriptome assembly. This was achieved due to the high overall quality of sequenced reads (**Figure 10**) and sufficient sequencing coverage (~30 million reads per sample). The transcriptome assembly process with Trinity resulted in the high-quality transcriptome assembly (**Table 1**). Comparative analysis of the transcriptome completeness based on the BUSCO scores placed the marbled crayfish transcriptome assembly presented in this study as the most complete among the available freshwater crayfish assemblies, with 93,1% BUCSO completeness. In the final assembly 60,004 (initial assembly 194,279) Trinity genes were present with N50 value of 3,039 bp (initial assembly 2,251 bp), outperforming previous marbled crayfish assembly with 22,338 transcripts and N50 value of 2,651 (Gutekunst et al., 2018).

Although N50 value and gene counts are often taken as the measure of assembly quality, their interpretation should be taken with caution, as those values can often be misleading (Ayling et al., 2020). In the case of the assembly presented in this study, both N50 value and Trinity gene counts were improved after filtering by length and removal of marbled crayfish mitogenome sequences. In the study by Gutekunst et al., 2018, transcriptome assembly was first clustered based on the sequence similarity (97%) with CD-HIT-EST (Fu et al., 2012) and the longest isoform for every Trinity gene was selected followed by transcript merging with CAP3 (Huang and Madan., 1999). From their assembly, transcripts containing more than 10% of repetitive segments identified with RepeatMasker were removed. Although this process might seem as "biologically correct" it drastically alters the final assembly statistics. In the initial attempt to map the reads to the Gutekunst et al. (2018) transcriptome assembly only 52.40% - 61.30% (results not shown) of the reads were mapped. While the mapping rates of this study's assemblywere 94.4% - 96.3% (before filtering) i.e. 83.40% – 88.50% after filtering.

Many approaches are available for the transcriptome assembly clean-up, each with its own drawback. Algorithms that utilise **filtering by similarity** (like CD-HIT-EST), are efficient due to their high speed, but often biased (as the similarity threshold is user defined) and can lead to the elimination of the alternatively spliced transcript isoforms that are indeed present in the species transcriptome (Kerkvliet et al., 2019). **Filtering by expression** has also been utilised as a

method of improving the transcriptome assembly quality, especially for differential gene expression analysis purposes. This step is often unnecessary as the lowly expressed transcripts in one tissue could be enriched in another tissue, therefore this could lead to failure of recognising tissue specific transcripts and isoforms in future studies. It should be noted that differential gene expression analysis tools offer the user a possibility to eliminate lowly expressed transcripts from the dataset post mapping, which is important for the background noise reduction and the number of false positives control (Love et al., 2014). **Filtering by the presence of open reading frames (ORF)** e.g. ones predicted by Transrate, as well as removal of transcripts with repetitive segments identified with RepeatMasker, is suitable for retaining protein coding transcripts. On the other hand, expressed non-coding RNAs and transposable elements rich in repetitive segments could also be eliminated in this step. **Filtering by length** was applied in this study. It eliminates short transcripts from the assembly (<500 bp) as they are considered to represent only fragments of the full-length transcripts. This was confirmed as the reduction in the mapping rate between initial and processed Trinity assembly can mostly be attributed to the elimination of contigs belonging to the marbled crayfish mitogenome (**Figure 11**) which makes up ~10% of all reads. The BUSCO statistics between raw and processed assembly remained unchanged (**Figure 13**). Finally, various filtering options have been shown to correct for assembly errors e.g. eliminating chimeric contigs, spurious insertions in contigs or local mis-assembly of contigs (Cabau et al., 2017; Kerkvilet et al., 2019; Smith-Unna et al., 2016). It is important to consider the trade-off between the assembly error elimination (redundancy removal) and the level of assembled transcripts conservation in the light of the biological question at hand.

5.2. Transcriptome annotation

The goal of the transcriptome annotation process is to connect different transcripts with the biological information available to date. This process is usually starting with the basic sequence annotation based on the homology to the known sequences. The initial annotation step for the marbled crayfish transcriptome assembly was conducted using the dammit! annotation pipeline (Scott, 2016), that involves four different resource databases: protein domain database, Pfam (El-Gebali et al., 2018), ortholog database, OrthoDB (Krivntseva et al., 2018), non-redundant cluster of proteins database, UniRef90 (Suzet et al., 2007) and non-coding RNA database Rfam (Kalvari et al., 2017). Combining the information and resources of these

databases, 41.9% of transcripts in the final assembly received annotation (**Figure 14a**). This is comparable to the percentage of transcripts annotated in the Caribbean spiny lobster, where 50.33% of the transcripts where annotated with the same pipeline. This annotation rate (~ 50%) is typical for non-model organisms (Baeza and MacManes, 2020). Still, the annotation process in the non-model organisms' transcriptomes is lagging the investigations in model organisms. This problem has been particularly well explained in the study by Clark et al. (2016), where they focused on the annotation of the crustacean immune related genes in the American lobster. Their approach utilised the availability of Insect Innate Immunity Database (IIID; Brucker et al. 2012) to overcome the limitations of the transcriptome annotation with the traditional methods. Unfortunately, this resource is no longer available at the original publication website. It is important to note that today, out of 1,076 sequenced genomes in the Pancrustacea, 1,016 belong to the Hexapods, and only 60 to the Crustaceans (Genomes-NCBI Datasets, accessed: November 2020).

In the functional classification the marbled crayfish transcripts were annotated across 401 KEGG pathways, which is comparable to the KEGG classification in the Caribbean spiny lobster and the red-swamp crayfish, where 390 and 311 KEGG pathways were represented, respectively (Baeza and MacManes, 2020, Shen et al., 2014). Comparable results were also achieved in respect of the top 20 most abundant KEGG pathways (**Figure 14b**) in marbled crayfish, with the highest number of transcripts assigned to the metabolic process and biosynthesis of secondary metabolites as in the red-swamp crayfish (Shen et al., 2014, Ali et al., 2015). Among the top 20 most abundant pathways KEGG terms were assigned to the pathways of neurodegeneration, as well as several other disease related conditions, including pathways in ALS, Alzheimer, Huntington, Parkinson, cancer, prion, viral and bacterial diseases. As expected, these disease related KEGG pathways were also among the over-represented KEGG pathways on the whole transcriptome assembly level (**Figure 19d**). Although a variety of KEGG pathways were represented, KO terms were assigned to only 14.5% of the final transcriptome assembly. Most of the transcripts in the assembly did not receive any annotation (58.9%). The problem of unannotated genes was also observed among the DE genes (**Supplementary table 2**), where many DE genes remained without annotation or were annotated as uncharacterised proteins within the dammit! pipeline. An even lower amount of DE genes received KEGG ID assignment. For the samples of the second sampling point that received KEGG annotation, over-

representation analysis revealed that the pathways over-represented in the whole transcriptome were also over-represented in the DE gene PsI subset (**Figure 19f**), while the over-represented DE genes in the As subset contained only one gene per subset. Gene ontology assignment with InterPro5 (Jones et al., 2014) showed similar performance to the assignment of KEGG terms among the DE genes. Analysis of GO terms with WEGO across DE gene subset revealed metabolic and cellular process as the most represented biological process and catalytic activity and binding as the most represented molecular function (**Figure 19a-c**). These terms are usually well represented across the whole transcriptome assemblies (Jiao et al., 2019, Shen et al., 2014, Shen et al., 2020). This could be resolved with future efforts in the field of Crustacean genomics as well as further breakthrough in studying the molecular processes in Crustaceans. Lastly, recent advances in the field of transcriptomics revealed that only 50% of the human transcriptome assembly consists of protein-coding genes (Eldem et al., 2017). Long non-coding RNAs (lncRNAs), small non-coding RNAs, as well as mobile elements are also expressed in the eukaryotic genomes. It is important to recognise these elements in the future studies by considering them in the transcriptome assembly and annotation process.

5.3. Differentially expressed genes and proPO cascade in response to A. astaci infection

The hepatopancreas represents an integrated organ of the crayfish immunity and metabolism. It is connected to the crustacean digestive tract and responsible for secretion of digestive enzymes, detoxification of toxic materials, regulation of metabolic processes and biogenesis, as well as nutrient absorption (Rőszer et al., 2014). Fixed phagocytes in the exterior surfaces of arterioles in haemal spaces, responsible for the substance clearance are also shown to be present in the hepatopancreas (Johnson, 1987). Crustacean hepatopancreas, particularly susceptible to the viral infections, can be used for diagnostics of many diseases based on its pathology or its physiological state (Shen et al., 2020). Recent studies in *P. clarikii* have shown hepatopancreas active role in the immune response to heavy metal detoxification (Meng et al., 2019, Zhang et al., 2019), as well as viral disease response (Dai et al., 2017, Jiao et al., 2019, Shen et al., 2020), uncovering numerous DE genes. In this study, the number of DE genes ranged from four in the As-treated crayfish of the first sampling point to 71 in the PsI-treated crayfish of the second sampling point (**Figure 18**). Based on the sole number of DE genes detected in the hepatopancreas of marbled crayfish it can be presumed that the reaction to the pathogen infection

was low to mild. Similarly, in the response to the Black May disease in *P. clarkii*, it has been observed that hepatopancreas (476 DE genes) had almost two times lower number of DE genes than muscle tissue (964 DE genes) and three times lower amount of DE genes than the crayfish gills (1703 DE genes), although DE genes in hepatopancreas showed relation to immune response and metabolic pathways (Shen et al., 2020). Therefore, a higher amount of DE genes could be expected in the tissues directly exposed to the environmental factors (pathogen), such as gills, intestines and haemolymph.

The most important reaction of the crustacean immune response is the activation of the proPO cascade. The synthesis of proPO is localised in the haemocytes, mainly in semi granular and granular cells. Due to the high toxicity of the proPO cascade intermediate products, this reaction is spatially and temporally finely tuned (Cerenius et al., 2008). Hemocyanin, which can exhibit phenoloxidase activity, is synthesised in the hepatopancreas (Soderhall and Cerenius, 1998). Therefore, an increase of the proPO expression would not be expected in the hepatopancreas, at least because the tissue is not affected by the oomycete growth. Furthermore, in *P. leniusculus* the resistance to an *A. astaci* infection was attributed to the constant up-regulation of the proPO (and peroxinexin) with no further increase in the proPO levels, whereas the native crayfish species showed increased proPO activation after imunostimulation (Cerenius et al., 2003). Unfortunately, in the study be Cerenius et al. (2003) these differences were only observed during the first 12 h post infection (a time-point not sampled in this study), and long-term effect has not been studied, making a direct comparison to this study difficult.

Studies of the *A. astaci* infection have shown that invasive crayfish can limit pathogen growth in their cuticle and act as a latent carrier of the disease. Some native European crayfish populations can also resist the infection with *A. astaci*, even when infected with a high virulent Pc genotype. This resistance is caused by encapsulation and strong melanisation of the penetrating hyphae in the cuticle, thus exhibiting a strong immune reaction like the North American invasive crayfish species (Gruber et al., 2014, Martín-Torrijos et al., 2017). Similarly, marbled crayfish populations latently infected with the Pc and possibly Or genotype have been detected (Keller et al., 2014). As mentioned above, latently infected crayfish constantly release spores in their environment (Strand et al., 2012). Under stressful conditions (e.g. laboratory conditions in restricted space can lead to the increased spore density) the latently infected

invasive crayfish species show a higher *A. astaci* load and can even succumb to the infection leading to their death (Söderhäll and Cerenius, 1992, Keller et al., 2014). This means that a late immune response, and possible proPO cascade activation, would be expected in the invasive marbled crayfish. The differential expression of proPO was not observed in this study, although proPO cascade related serine proteinase inhibitor (serpin, *DN12252 c0 g1*) was up-regulated in the PsI treated crayfish of the second sampling point (21 days after infection). This was in accordance with the second hypothesis of this study and expectation of a constantly up-regulated proPO cascade in the invasive crayfish. Serpins are regulatory inhibitors of the proPO cascade, important for prevention of superfluous activation and production of proPO cascade compounds that could be toxic to the animal (Cerenius et al., 2010). Specifically, in *D. melanogaster* Serpin-27A has been identified as the inhibitor of the terminal protease prophenoloxidase activating enzyme (Gergorio et al., 2002). In this study, serpin up-regulation in the hepatopancreas, 21 days post challenge with the high virulent PsI strain, could be triggered as a response to the up-regulated proPO system combating a more active *A. astaci* infection occurring in the crayfish haemolymph. Since proPO is mainly produced in the crayfish hemocytes and degranulated in the extracellular space, higher expression of the serpin could have protective function in preventing the potential damage caused by the cleavage of proPO zymogen within the hepatopancreatic tissue.

5.3.1. C-type lectin duality and viral response

The two sample groups (As and PsI) belonging to the first sampling point (three days post challenge) revealed only 12 DE genes. Based on the available annotation, it is hard to interpret their potential role in the immune system. All DE genes in the As treated crayfish remained un-annotated. Among the DE genes in the PsI treatment up-regulated Viral RNA-dependent RNA polymerase (*DN139895 c0 g1*) and down-regulated C-type lectin-like (*DN6324 c0 g1*) are the most interesting in the context of this study. During GO assignment with InterPro5, transcripts belonging to the gene *DN139895 c0 g1* showed Caliciviridae signatures (domain annotations). These viruses belong to the positive sense single-stranded RNA viruses, and have been associated with a number of diseases in vertebrates, while until today they have not been recognised among invertebrates. Their genome size is in the range of 7.4 to 8.3 kilobases, which corresponds to the *DN139895 c0 g1* transcript length of 7,722 bp (Asanaka et

al., 2005; Yu et al., 2013). This finding could also explain the shift in the per sequence GC plot (**Figure 10**), as GC distribution could be altered by the viral contamination. It might also be possible that the viral spread was triggered by the infection challenge with the high virulent PsI strain, causing an opportunistic infection, as it is known that the hepatopancreas microbiome composition can be altered during infection (Cornejo-Granados et al., 2017).

In this study the microbial community composition of the hepatopancreas has not been considered, although transcriptomic studies on this scale should also be viewed from a metagenomic perspective. Therefore, future studies should focus on identification of the microbial communities in the marbled crayfish transcriptome assembly as well as on changes in their composition during infection with *A. astaci*. C-type lectins (CTLs) are a known effector of the crustacean innate immunity, with their carbohydrate recognition domain. They can also exhibit antiviral role activities (Zhao et al., 2009). Therefore, it would be expected that they are up-regulated in the crayfish potentially infected with the member of Caliciviridae. However, in the kuruma shrimp the CTL, which is critical for binding to calreticulin on cell membrane, can be hijacked by White Spot Syndrome Virus (WSSV) during infection to assist the viral particle entry in the host cell (Wang et al, 2014). Similar observation has been made in the *Aedes aegypti* that can collaborate the cell entry of the West Nile virus (Cheng et al., 2010). In the closely related red-swamp crayfish down-regulation of CTL 96 h post WSSV infection has been observed (Shi et al., 2010). In the light of these findings, the reduction of CTL (*DN6324 c0 g1*) expression could be explained as the host defence mechanism, preventing the viral entry.

Differential expression of different CTL (*DN7962 c0 g1)* was present in the PsI-treated crayfish of the second sampling point. In this case, CTL was up-regulated. Many different subtypes of CTLs have also been observed in the red-swamp crayfish, where they have been linked to the ProPO cascade, antiviral as well as antibacterial response (Zhang et al., 2018). It can be speculated that this up-regulation is triggered as a late immune response to the infection with a high virulent PsI strain. Duality of the CTLs roles in the crayfish immune system is linked to their common characteristics of a high binding affinity to their substrates (Wang et al., 2014). Further characterisation of the CTLs subtypes in the marbled crayfish is needed to establish a firm link between their expression and molecular functions.

5.3.2. Genes DE in both As and PsI treatment of the late response to the A. astaci challenge

Although most of the DE genes were treatment specific, there were some genes that were DE across all samples (n = 14) of the second sampling point (21 days after infection) (**Figure 18**). Among them, several transcripts related to the large ribosomal subunit (LSU) and small ribosomal subunit (SSU) were up-regulated (Supplementary table 2). Ribosomal RNA, that constitutes a significant proportion of the eukaryotic RNA, is usually removed prior to the RNA sequencing by ribosomal depletion or poly-A enrichment techniques. In this study a poly-A enrichment with magnetic beads was performed during sequencing library preparation. The analysis with FastQ Screen (Wingett et al., 2018) confirmed that the rRNA was not over-represented in the read data, with slight variance in rRNA abundance (less than ~ 1% of the reads) (Figure 10). Most of the DE genes associated with the ribosomal RNA had a low base mean expression of (~ 20.69) and only one transcript had a higher base mean expression of 337.57 (*DN1413 c0 g2*). Differentially expressed ribosomal genes were also observed in the response to the Black May disease (Shen et al., 2020) and lipopolysachariide challenge (Jiao et al., 2019). The up-regulation in rRNA genes, should be interpreted with caution. Prior sequencing a poly-A enrichment is conducted, therefore the variances in the rRNA expression can be attributed to the successfulness of the enrichment process and should not be interpreted in the disease related context. These slight variances in the rRNA content were observed in the contamination screening with FastQ Screen (**Figure 11**). Two other shared up-regulated genes were *DN645 c2 g1* and *DN863 c0 g1*. The first (*DN645 c2 g1*) was not labelled with any annotation and the second (*DN863 c0 g1*) contained an MFS-1 domain. The major facilitator super family (MFS) is responsible for the transmembrane transportation of the small solutes in the response to chemiosmotic ion gradient (Pao et al., 1998).

Four shared down-regulated genes were also detected: Mucin-like domain containing protein (*DN1776 c0 g2*), Invertebrate connectin (*DN18278 c0 g1*), Microtubule-associated protein 1B (*DN19074 c0 g1*), Colony stimulating factor 3 receptor -granulocyte (*DN51753 c0 g1*) and uncharacterised metal ion binding protein (*DN6337 c0 g1*). Mucins and mucin-like proteins represent a diverse and heterogeneous group of O-glycosylated proteins that are main mucosal components. They act as a barrier for protection and interaction with cell receptors, and are capable of creating gel-like layer in the extracellular space of organs (Carraway and Hall, 1991). They have been attributed with various immune related roles, cellular regeneration,

differentiation, signalling, adhesion and are aslo associated with granulocitin (mucin-binding) lectins (Iwanaga et al., 2005). Mucins are high expressed in the epithelial and endothelial cells, as well as speciallised cells of the gastrointestinal, reproductive and respiratory tract in humans. Alteration in their expression is associated with many human disease conditions, especially cancer (Pinzón et al., 2019). Therefore, the down-regulation of their expression could be connected to the inhibition of the host immune response, caused by the crayfish plague pathogen. This would render the hepatopancreas tissue more susceptible to the pathogen invasion. Different transcripts related to the cytoskeletal filaments and cytoskeletal proteins were also down-regulated in the As and PsI samples of the second sampling point. The colony stimulating factor 3 receptor (granulocyte) (*DN51753 c0 g1*) is a cell surface receptor for granulocyte colony stimulating factor, which initiates cell proliferation and differentiation of granulocytes and macrophages (Fukungana et al., 1991). Its involvement in modulation of the crustacean innate immune response is unknown.

5.3.3. Specific late immune response to challenge with PsI strain

Among the differentially expressed genes of the late immune response to the PsI challenge, crustin (*DN83351 c0 g1*, 615 bp) was strongly up-regulated (Supplementary table 2). Crustins are antimicrobial peptides present among the large group of crustaceans. Different crustin isoforms have been identified in the red-swamp crayfish (Du et al., 2019), where they have been linked to the immune response of bacterial as well as viral infection (Wang et al., 2015, Shi et al., 2010). All crustins, including crustin encoded by *DN83351 c0 g1* transcript, contain signal peptide domain (results not shown), and a whey acidic protein (WAP) domain (Tassanakajon et al., 2015). Little is known about the involvement of crustins in the anti-fungal response in crustaceans. Amino acid content of the inter signal peptide-WAP domain revealed a high proportion of proline residues characteristic for type III crustins. In Chinese white shrimp, rSWDFc crustin, among others, has been shown to possess anti-fungal activity (Jia et al., 2008). Role of this crustin in the immune response of marbled crayfish might be better understood in the future, as it can represent a potential target for knock-out and over-expression studies.

As a novel transcript involved in the crustacean immune system up-regulation in several L1 retrotransposon domain-containing transcripts (*DN23859 c0 g1*, *DN18537 c0 g1*, *DN36498 c0 g1*) in the late response of the PsI treated marbled crayfish was uncovered. L1

retrotransposons belong to a class of the autonomous long interspersed nuclear element (LINE) which in marbled crayfish make up for ~10% of genomic proportion (Gutekunst et al., 2018), similar to the genomic proportion in red claw crayfish ~15% (Tan et al., 2020). The knowledge about the expression of mobile elements and their characteristic in crustaceans is limited, although they represent an interesting model for their study due to the great variation in genome sizes as well as a high species and lifestyle diversity (Piednoël et al., 2013). Mobile element activation can be achieved during stressful events, environmental changes, and can be associated with the emergence of new lineages, species, populations, or subpopulations (Oliver and Greene, 2009). In humans, over-expression of LINE-1 has been linked to the apoptosis, DNA damage and repair, cellular plasticity, as well as stress responses and promotions of tumour progression (Zhang et al., 2020).

Co-evolution of the transposable elements with the initiation of the innate immune response, which is linked to the interferon-γ response, has been documented in mammals (Chuong et al., 2016). One of the mediators of the response to the interferon-γ (IFN- γ) and other cytokines are the guanylate binding proteins. One of them is also up-regulated (*DN3059 c1 g1*)) in the PsI treated crayfish of the second sampling point. Both IFN-γ and GBP are responsible for the antiviral and anti-microbial response. While GBP clave GTP to GDP and GMP in varying ratios, the physiological function of this activity is not understood (Praefcke, 2018). In this context, up-regulated Guanylate kinase (*DN17789 c0 g1)* might be important for the maintenance of cellular GDP levels, important for the GBP activity and oligomerisation (Honkala et al., 2020). Further studies are needed to unravel the molecular interactions between these enzymes as well as their role in the Crustacean immunity.

5.4. Broader perspective on the experimental results

As mentioned in the introduction, *A. invadans* is a pathogen that affects freshwater fish populations causing the disease epizootic ulcerative syndrome (EUS). Over 160 fish species are susceptible to this pathogen, while some show resistance (Oidtmann, 2012). In the *A. invadans* resistant common carp, *Cyprinus curpio* Linnaeus, 1758, it has been shown that, during time course infection experiment multiple immune related gene pathways are up-regulated (Verma et al., 2020). Direct comparison between the immune response of the crayfish and fishes is not possible, because fish, like all vertebrates, have also a more complex adaptive immune response,

unlike crayfish which rely only on the innate immunity. Nonetheless, several correlations between these two models exist. Firstly, both *A. invadans* resistant fish and *A. astaci* resistant crayfish rely on the isolation of the pathogen to prevent its spread. In crayfish, this occurs in the exposed cuticle through the melanisation reaction triggered by the proPO cascade (Cerenius et al., 2008). On the other hand, in common carp the hyphae are surrounded by well-developed granulomas that prevent the spread of hyphae into the soft tissue. Secondly, in the resistant common carp histopathological alterations were not observed outside of the infection site (Verma et al., 2020), whereby spread of the infection into the adjoining tissues was reported in susceptible fish species (Vishwanath et al., 1998). Likewise, resistant latently infected crayfish species do not show alterations in their organs, while in susceptible crayfish species, after progradation through the basal lamina, *A. astaci* spreads throughout the crayfish body (Cammà et al., 2010, Hsieh et al., 2016). Lastly, two recent studies have shown that innate immune response is greatly involved in defence against *A. invadans* in both resistant (Varma et al., 2020) and susceptible fish species (Pradhan et al., 2020). This approach also revealed that *A. invadans* can modulate the expression of the host immune related genes, thus evading its defence mechanisms in the susceptible rohu fish *Labeo rohita* Hamilton, 1822 (Pradhan et al., 2020). One of the particularly interesting DE pathways is up-regulated GBP-associated pathway in the resistant common carp, which was correlated with the recognition of pathogen-associated molecular patterns (Verma et al., 2020) and, as previously mentioned, also up- regulated in the marbled crayfish. Comparative approach to the crayfish immune response to the *A. astaci* infection, between resistant and susceptible crayfish, could yield a better understanding of the disease progression and resistance mechanisms.

6. Conclusion

In this study, the gene expression in hepatopancreas transcriptome of marbled crayfish was analysed after an infection challenge with two strains of *A. astaci*, As (low virulent) and PsI (high virulent). High-quality *de novo* transcriptome of marbled crayfish was assembled containing 60,004 genes, with 41.9% of transcripts receiving annotation. Differential gene expression analysis revealed in total 102 differentially expressed genes. In the early response, three days post challenge, only few (N = 12) genes were differentially expressed. Among them the presence of the Caliciviridae and C-type lectin down-regulation was observed in the PsI treated crayfish. But, substantial initial response to the *A. astaci* infection could not be observed in marbled crayfish hepatopancreas. A higher number of differentially expressed genes (N = 90) was observed in the samples 21 days post infection, suggesting a later response to the infection, possibly due to the prolonged time in the experimental conditions, which can cause stress to the crayfish. Effector genes of the canonical proPO pathway were not up-regulated in the hepatopancreas tissue of the challenged crayfish. However, their inhibitor, serpin, was up-regulated in the PsI treated crayfish (sampling point II), suggesting that suppression of the proPO cascade occurs in hepatopancreas. Previously unreported genes with a potential role in crayfish innate immunity were also detected. Some of them, annotated as retrotransposable elements (L1), as well as GBP, have only recently been recognised as potential mediators of the immune response in vertebrates. This study represents a significant contribution to the emerging crayfish genomics field, providing a dataset for future comparative transcriptomic analysis and gene identification. In conclusion, this study provided new insights into the crayfish immune response to an *A. astaci* challenge, as well as identified potential new challenges for future studies of crayfish innate immunity.

7. Literature

Alderman, D. J. 1996. Geographical spread of bacterial and fungal diseases of crustaceans. Revue scientifique et technique (International Office of Epizootics), **15(2)**, 603-632.

Ali, M., Pavasovic, A., Amin, S., Mather, P. and Prentis, P., 2015. Comparative analysis of gill transcriptomes of two freshwater crayfish, *Cherax cainii* and *C. destructor*. Marine Genomics, **22**, 11-13.

Andersson MG, Cerenius L 2002. Analysis of chitinase expression in the crayfish plague fungus *Aphanomyces astaci*. Diseases of Aquatic Organisms, **51**, 139-47.

Andrews, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data, Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Apitanyasai, K., Noonin, C., Tassanakajon, A., Söderhäll, I., Söderhäll, K., 2016. Characterization of a hemocyte homeostasis-associated-like protein (HHAP) in the freshwater crayfish *Pacifastacus leniusculus*. Fish and Shellfish Immunology, **58**, 429-435.

Asanaka M, Atmar RL, Ruvolo V, Crawford SE, Neill FH, Estes MK., 2005. Replication and packaging of Norwalk virus RNA in cultured mammalian cells. Proceedings of the National Academy of Sciences of the United States of America, **102(29)**, 10327-32

Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., Davis, A., Dolinski, K., Dwight, S., Eppig, J., Harris, M., Hill, D., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J., Richardson, J., Ringwald, M., Rubin, G. and Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nature Genetics, **25(1)**, 25-29.

Ayling M, Clark MD, Leggett RM. 2020. New approaches for metagenome assembly with short reads. Briefngs in Bioinformatics, **21(2)**, 584-594.

Baeza, J. A., MacManes, M., 2020. De novo assembly and functional annotation of the heart+hemolymph transcriptome in the Caribbean spiny lobster Panulirus argus, Marine Genomics, **54**, 100783.

Bauchau, A.G., 1981. Crustaceans. In: Ratcliffe, N.A., Rowley, A.F. (Eds.), Invertebrate Blood Cells, vol. 2. Academic Press, London, 385-420.

Becking, T., Delaunay, C., Cordaux, R., Berjeaud, J., Braquart-Varnier, C., Verdon, J. 2020. Shedding Light on the Antimicrobial Peptide Arsenal of Terrestrial Isopods: Focus on Armadillidins, a New Crustacean AMP Family. Genes, **11(1)**, 93.

Blighe, K., Rana, S., Lewis, M., 2020. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.8.0.

Bohman P, Edsman L, Martin P, Scholtz G 2013. The first Marmorkrebs (Decapoda: Astacida: Cambaridae) in Scandinavia. BioInvasions Records **2**, 227-232.

Bolger, A. M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, **30(15)**, 2114-2120.

Brisbin, M.M., McElroy, A.E., Espinosa, E.P., Allam, B., 2015. Antimicrobial activity of the cuticle of the American lobster, *Homarus americanus*. Fish and Shellfish Immunology, **44**, 542–546.

Brucker, R., Funkhouser, L., Setia, S., Pauly, R., Bordenstein, S., 2012. Insect Innate Immunity Database (IIID): An Annotation Tool for Identifying Immune Genes in Insect Genomes. Plos ONE, **7(9)**, e45125.

Cammà, C., Ferri, N., Zezza, D., Marcacci, M., Paolini, A., Ricchiuti, L., Lelli, R. 2010. Confirmation of crayfish plague in Italy: detection of *Aphanomyces astac*i in white clawed crayfish. Diseases Of Aquatic Organisms, **89**, 265-268.

Carraway, K., Hull, S. 1991. Cell surface mucin-type glycoproteins and mucin-like domains. Glycobiology, **1(2)**, 131-138.

Cazzolla Gatti, R. 2016. Freshwater biodiversity: a review of local and global threats. International Journal of Environmental Studies, **73(6)**, 887-904.

Cerenius, L., Andersson, M., Söderhäll, K., 2008. *Aphanomyces astaci* and Crustaceans. Oomycete Genetics and Genomics, **21**, 425-433.

Cerenius L, Bangyeekhun E, Keyser P, Söderhäll I, Söderhäll K 2003. Host prophenoloxidase expression in freshwater crayfish is linked to increased resistance to the crayfish plague fungus, *Aphanomyces astaci*. Cellular Microbiology, **5**, 353-357.

Cerenius, L., Kawabata, S., Lee, B., Nonaka, M., Söderhäll, K., 2010. Proteolytic cascades and their involvement in invertebrate immunity. Trends in Biochemical Sciences, **35(10)**, 575-583.

Cerenius, L., Lee, B. L., Söderhäll, K., 2008. The proPO-system: pros and cons for its role in invertebrate immunity. Trends in Immunology, **29(6)**, 263–271.

Cerenius L, Söderhäll, K., 1985. Repeated zoospore emergence as a possible adaptation to parasitism in *Aphanomyces*. Experimental Mycology, **9**, 259-263.

Cerenius L, Söderhäll, K., 2004. The prophenoloxidase-activating system in invertebrates. Immunological Reviews, **198**, 1016-126.

Cerenius, L., Söderhäll, K., 2012. Crustacean immune responses and their implications for disease control, Infectious Disease in Aquaculture, 69-87.

Cerenius, L., Söderhäll, K. 2018. Crayfish immunity – Recent findings., Developmental and Comparative Immunology,**80,** 94-98.

Chang, Y., Lin, C., Tsai, C., Siva, V., Chu, C., Tsai, H., Song, Y. (2013). The New Face of the Old Molecules: Crustin Pm4 and Transglutaminase Type I Serving as RNPs Down-Regulate Astakine-Mediated Hematopoiesis., Plos ONE,**8**.

Cheng, G., J. Cox, P. Wang, M. N. Krishnan, J. Dai, F. Qian, J. F. Anderson, and E. Fikrig. 2010. A C-type lectin collaborates with a CD45 phosphatase homolog to facilitate West Nile virus infection of mosquitoes., Cell **142**, 714-725.

Choi, H., Lee, D.G., 2014. Antifungal activity and pore-forming mechanism of astacidin 1 against *Candida albicans*., Biochemie, **105**, 58-63.

Chou, P.H., Chang, H.S., Chen, I.T., Lee, C.W., Hung, H.Y., Wang, H.C., 2011. *Penaeus monodon* Dscam (PmDscam) has a highly diverse cytoplasmic tail and is the first membrane-bound shrimp Dscam to be reported., Fish and Shellfish Immunology., **30**, 1109-1123.

Chucholl C. and Pfeiffer M., 2010. First evidence for an established Marmorkrebs (Decapoda, Astacida, Cambaridae) population in Southwestern Germany, in syntopic occurrence with *Orconectes limosus* (Rafinesque, 1817). Aquatic Invasions, **5**, 405-412.

Chuong, E., Elde, N., & Feschotte, C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science, **351(6277)**, 1083-1087.

Clark, K. F., Greenwood, S. J., 2016, Next-generation sequencing and the crustacean immune system: The need for alternatives in immune gene annotation, Integrative and Comparative Biology, **56 (6)**, 1113-1130.

Cornejo-Granados, F., Lopez-Zavala, A.A., Gallardo-Becerra, L., Mendoza-Vargas, A., Sánchez, F., Vichido, R., Brieba, L.G., Viana, M.T., Sotelo-Mundo, R.R., Ochoa-Leyva, A., 2017, Microbiome of Pacific Whiteleg shrimp reveals differential bacterial community composition between Wild, Aquacultured and AHPND/EMS outbreak conditions. Scientific Reports, **7**, 11783.

Crandall, K.A., and De Grave, S. 2017. An updated classification of the freshwater crayfishes (Decapoda: Astacidea) of the world, with a complete species list., Journal of Crustacean Biology., **37**, 615-653.

Dai, L., Abbas, M., Kausar, S., Zhou, Y., 2017. Transcriptome analysis of hepatopancraes of Procambarus clarkii challenged with polyriboinosinic polyribocytidylic acid (poly I:C). Fish and Shellfish Immunology, **71**, 144-150.

De Gregorio, E., Han, S., Lee, W., Baek, M., Osaki, T., Kawabata, S., Lee, B., Iwanaga, S., Lemaitre, B. and Brey, P., 2002. An Immune-Responsive Serpin Regulates the Melanization Cascade in Drosophila. Developmental Cell, **3(4)**, 581-592.

Dieguez-Uribeondo, J., Cerenius, L. 1998. The inhibition of extracellular proteinases from *Aphanomyces spp*. by three different proteinase inhibitors from crayfish blood. Mycological Research, **102(7)**, 820-824.

Diéguez-Uribeondo J., García M.A., Cerenius L., Kozubíková E., Ballesteros I., Windels C., Weiland J., Kator H., Söderhäll K., Martín MP. 2009. Phylogenetic relationships among plant and animal parasites, and saprotrophs in *Aphanomyces* (Oomycetes)., Fungal Genetics and Biology, **46(5)**, 365-76.

Diéguez-Uribeondo J., Huang T.S., Cerenius L., Söderhäll K. 1995. Physiological adaptation of an *Aphanomyces astaci* strain isolated from the freshwater crayfish *Procambarus clarkii*. Mycological Research., **99**, 574-578.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T. R., 2013. STAR: ultrafast universal RNA-seq aligner, Bioinformatics, **29(1)**, 15-21.

Du, Z., Li, B., Shen, X., Wang, K., Du, J., Yu, X., Yuan, J. 2019. A new antimicrobial peptide isoform, Pc-crustin 4 involved in antibacterial innate immune response in freshwater crayfish, *Procambarus clarkii*., Fish and Shellfish Immunology, **94**, 861-870.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.I., Knowler, D.J., Lévêque, C. and Naiman, R.J., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews, **81(2)**, 163-182.

Eldem, V., Zararsiz, G., Taşçi, T., Duru, I. P., Bakir, Y., Erkan, M. 2017. Transcriptome analysis for non- model organism: current status and best-practices. In: Applications of RNA-Seq and Omics Strategies-from Microorganisms to Human Health. IntechOpen, Rijeka, 55–71

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S., Luciani, A., Potter, S., Qureshi, M., Richardson, L., Salazar, G., Smart, A., Sonnhammer, E., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. and Finn, R., 2018. The Pfam protein families database in 2019. Nucleic Acids Research, **47(D1)**, D427-D432.

Ercoli, F. 2019. First record of an established marbled crayfish *Procambarus virginalis* (Lyko, 2017) population in Estonia., Bioinvasions Records, **8(3)**, 675-683.

EU (2016) Commission Implementing Regulation 2016/1141 of 13 July 2016. Adopting a list of alien species of Union concern pursuant to Regulation (EU) No. 1143/2014 of the European Parliament and of the Council. Official Journal of the European Union L., **189**, 4–8.

Ewels, P., Magnusson, M., Lundin, S., Käller, M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, **32(19)**, 3047-3048.

Faulkes Z., 2010. The spread of the parthenogenetic marbled crayfish, Marmorkrebs (*Procambarus sp*.), in the North American pet trade. Aquatic Invasions, **5**, 447-450.

Francesconi, C., Makkonen, J., Schrimpf, A., Jussila, J., Kokko, H., Theissinger K., Elevated resistance and latent infections caused by reduced virulence: Marbled crayfish and noble crayfish challenged with the crayfish plague disease agent, submitted manuscript.

Fu, L., Niu, B., Zhu, T., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics, **28 (23)**, 3150-3152.

Fukunaga, R., Ishizaka-Ikeda, E., Pan, C. X., Seto, Y., Nagata, S. 1991. Functional domains of the granulocyte colony-stimulating factor receptor. The EMBO journal, **10(10)**, 2855–2865.

Füreder, L. 2006. Indigenous crayfish habitat and threats. In: C. Souty-Grosset, D. M. Holdich, P. Y. Noël, J. D. Reynolds, and P. Haffner (eds.), Atlas of Crayfish in Europe (Paris: Muséum national d'Histoire naturelle), 26-43.

Glazer L., Tom M., Weil S., Roth Z., Khalaila I., Mittelman B., Sagi A., 2013. Hemocyanin with phenoloxidase activity in the chitin matrix of the crayfish gastrolith. Journal of Experimental Biology, **216**, 1898-904.

Gamradt, S. C., Kats L. B., Aanzalone C. B. 1997. Aggression by non-native crayfish deters breeding in California newts. Conservation Biology **11**, 793-796.

Gleick, P.H., 1996. Water resources. In: S.H. Schneider (Ed.) Encyclopedia of Climate and Weather (New York: Oxford University Press), 817–823.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology, **29(7)**, 644-652.

Grandjean, F., Gan, H., Moumen, B., Giraud, I., Hatira, S., Cordaux, R., Austin, C. 2020. Dataset for sequencing and de novo assembly of the European endangered, white-clawed crayfish (*Austropotamobius pallipes*) abdominal muscle transcriptome., Data In Brief, **29**, 105166.

Gruber, C., Kortet, R., Vainikka, A., Hyvarinen, P., Rantala, M.J., Pikkarainen, A., Jussila J., Makkonen J., Kokko H., Hirvonen H., 2014. Variation in Resistance to the Invasive Crayfish Plague and Immune Defence in the Native Noble Crayfish. Annales Zoologici Fennici. 2014, **51(4)**, 371-89.

Gutekunst, J., Andriantsoa, R., Falckenhayn, C., Hanna, K., Stein, W., Rasamy, J., Lyko, F. 2018. Clonal genome evolution and rapid invasive spread of the marbled crayfish. Nature Ecology and Evolution, **2(3)**, 567-573.

Hsieh, C., Huang, C., Pan, Y. 2016. Crayfish plague *Aphanomyces astaci* detected in redclaw crayfish, Cherax quadricarinatus in Taiwan. Journal of Invertebrate Pathology, **136,** 117-123.

Hossain, M., Patoka, J., Kouba, A., Buřič, M. 2018. Clonal crayfish as biological model: a review on marbled crayfish. Biologia, **73(9)**, 841-855.

Holdich, D. M., Reynolds, J. D., Souty-Grosset, C. Sibley, P. J. 2009. A review of the ever increasing threat to European crayfish from non-indigenous crayfish species. Knowleadge and Management of Aquatic Ecosystems., **11**, 394-395.

Honkala, A., Tailor, D., Malhotra, S. 2020. Guanylate-Binding Protein 1: An Emerging Target in Inflammation and Cancer., Frontiers In Immunology, **10**.

Huang, Y., Chen, Y., Hui, K., Ren, Q. 2018. Cloning and Characterization of Two Toll Receptors (PcToll5 and PcToll6) in Response to White Spot Syndrome Virus in the Red Swamp Crayfish *Procambarus clarkii*., Frontiers In Physiology, **9**.

Huang X, Madan A., 1999. CAP3: A DNA sequence assembly program. Genome Res., **9(9)**, 868-77.

Iwanaga, S., Lee, B., 2005, Recent Advances in the Innate Immunity of Invertebrate Animals. BMB Reports, **38(2)**, 128-150.

Janský V., Mutkovič A. 2010. Marbled crayfish—*Procambarus sp*. (Crustacea: Decapoda: Cambaridae)—first find in Slovakia. Acta Rerum Naturalium Musei Natuionalis Slovenici **56**. 64-67.

Jearaphunt M, Noonin C, Jiravanichpaisal P, Nakamura S, Tassanakajon A, Söderhäll I, Söderhäll K. 2014. Caspase-1-like regu- lation of the proPO-system and role of ppA and Caspase-1-like cleaved peptides from proPO in innate immunity. PLoS Pathog. **10**, e1004059.

Jia, Y.P., Sun, Y.D., Wang, Z.H., Wang, Q., Wang, Z.W., Zhao, Z.F., Wang, J.X., 2008. A single whey acidic protein domain (SWD)-containing peptide from fleshy prawn with antimicrobial and proteinase inhibitory activities. Aquaculture **284**, 246–259.

Jiao, T., Yang, T., Wang, D., Gao, Z., Wang, J., Tang, B. Liu, Q., Zhang, D. And Dai, L. 2019. Characterization and expression analysis of immune-related genes in the red swamp crayfish, *Procambarus clarkii* in response to lipopolysaccharide challenge. Fish and Shellfish Immunology, **95**, 140-150.

Jin X.K., Li S., Guo X.N., Cheng L., Wu M.H., Tan S.J., Zhu Y.T., Yu A.Q., Li W.W., Wang Q., 2013. Two antibacterial C-type lectins from crustacean, Eriocheir sinensis, stimulated cellular encapsulation in vitro. Developmental and Comparative Immunology. **41**, 544-552.

Jiravanichpaisal P., Lee B.L., Söderhäll K. 2006. Cell-mediated immunity in arthropods: Hematopoiesis, coagulation, melanization and opsonization. Immunobiology., **211**, 213-236.

Jirikowski, G., Kreissl, S., Richter, S., Wolff, C. 2010. Muscle development in the marbled crayfish—insights from an emerging model organism (Crustacea, Malacostraca, Decapoda)., Development Genes And Evolution, **220(3-4)**, 89-105.

Johnson, P.T., 1987., A review of fixed phagocytic and pinocytotic cells of decapod crustaceans, with remarks on hemocytes. Developmental and Comparative Immunology, **11**, 679-704.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., Hunter, S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics (Oxford, England), **30(9)**, 1236-1240.

Joseph G. Kunkel, Wolfram Nagel, Michael J. Jercinovic 2012. Mineral Fine Structure of the American Lobster Cuticle, Journal of Shellfish Research **31(2)**, 515-526.

Jussila J., Makkonen J., Vainikka A., Kortet R., Kokko H. 2014. Crayfish plague dilemma: how to be a courteous killer? Boreal Environment Research, **19**, 235-244.

Jussila, J., Vrezec, A., Makkonen, J., Kortet, R., Kokko, H. 2015. Invasive Crayfish and Their Invasive Diseases in Europe with the Focus on the Virulence Evolution of the Crayfish Plague. Biological Invasions in Changing Ecosystems: Vectors, Ecological Impacts, Management and Predictions, **2**, 183-211.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E., Rivas, E., Eddy, S., Bateman, A., Finn, R. and Petrov, A., 2017. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Research, **46(D1)**, D335-D342.

Kanehisa, M. and Sato, Y. 2020. KEGG Mapper for inferring cellular functions from protein sequences. Protein Science. **29**, 28-35.

Keller, N., Pfeiffer, M., Roessink, I., Schulz, R., Schrimpf, A. 2014. First evidence of crayfish plague agent in populations of the marbled crayfish (*Procambarus fallax* forma *virginalis*). Knowledge and Management Of Aquatic Ecosystems, **414**, 15.

Kerkvliet, J., Fouchier, A., Wijk, M., Groot, A., 2019. The Bellerophon pipeline, improving de novo transcriptomes and removing chimeras. Ecology and Evolution, **9(18)**, 10513-10521.

Kielbasa, S., Wan, R., Sato, K., Horton, P. and Frith, M., 2011. Adaptive seeds tame genomic sequence comparison. Genome Research, **21(3)**, pp.487-493.

Kouba, A., Petrušek, A. and Kozák, P. 2014. Continental-wide distribution of crayfish species in Europe: update and maps. Knowledge and Management of Aquatic Ecosystems, **413**, 5.

Kouba, A., Tikal, J., Cisar, P., Vesely, L., Fort, M., Priborsky, J., Patoka J., Buřič M.. 2016. The significance of droughts for hyporheic dwellers: Evidence from freshwater crayfish. Scientific Reports, **6**, 1-7.

Kozubíková E., Viljamaa-Dirks S., Heinikainen S., Petrusek A. 2011. Spiny-cheek crayfish *Orconectes limosus* carry a novel genotype of the crayfish plague pathogen *Aphanomyces astaci*. Journal of Invertebrate Pathology, **108**, 214-216

Kriventseva, E., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. and Zdobnov, E., 2018. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Research, **47(D1)**, D807-D811.

Lan, J., Zhou, J., Zhang, X., Wang, Z., Zhao, X., Ren, Q., Wang, J. 2013. Characterization of an immune deficiency homolog (IMD) in shrimp (*Fenneropenaeus chinensis*) and crayfish (*Procambarus clarkii*). Developmental and Comparative Immunology, **41(4)**, 608-617.

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods., **9**, 357-359.

Lee, S.Y., Lee, B.L., Söderhäll, K., 2003. Processing of an antibacterial peptide from hemocyanin of the freshwater crayfish *Pacifastacus leniusculus*. Journal of Biological Chemistry. **278**, 7927-7933

Leonelli, S., Ankeny, R. 2013. What makes a model organism?. Endeavour, **37(4)**, 209-212.

Li, B., Yang, B., Shen, X., Wang, K., Wei, Z., Du, Z. 2020. Molecular characterization and expression analysis of tumor necrosis factor receptor-associated factor 6 (traf6) like gene involved in antibacterial innate immune of freshwater crayfish, *Procambarus clarkii*. Fish and Shellfish Immunology, **104,** 517-526.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup, 2009. The Sequence alignment/map (SAM) format and SAMtools, Bioinformatics, **25(16),** 2078-9.

Liang, Z.C., Sottrup-Jensen, L., Aspan, A., Hall, M., Söderhäll, K., 1997. Pacifastin, a novel 155-kDa heterodimeric proeinase inhibitor containing a unique transferrin chain. Proceedings of the National Academy of Sciences of the United States of America, **94**, 6682-6687.

Lin, X., Söderhäll, I., 2011. Crustacean hematopoiesis and the astakine cytokines. Blood **117**, 6417-6424.

Lin, X., Söderhäll, K., Söderhäll, I., 2008. Transglutaminase activity in the hemato- poietic tissue of a crustacean, Pacifastacus leniusculus, importance in hemocyte homeostasis. BMC Immunology, **9**, 58.

Ling E, Yu X. 2006. Cellular encapsulation and melanization are enhanced by immulectins, pattern recognition receptors from the tobacco hornworm *Manduca sexta*. Developmental and Comparative Immunology, **30**, 289e99

Liu, N., Lan, J., Sun, J., Jia, W., Zhao, X., Wang, J. 2015. A novel crustin from *Marsupenaeus japonicus* promotes hemocyte phagocytosis. Developmental and Comparative Immunology, **49(2)**, 313-322.

Lockwood L., Hoopes J., Martha F., Marchetti P. 2010. Invasion ecology. Blackwell publishing, Oxford, United Kingdom.

Love, M.I., Huber, W., Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology **15**, 550.

Maguire, I., Jelić, M., Klobučar, G., Delpy, M., Delaunay, C. and Grandjean, F. 2016. Prevalence of the pathogen *Aphanomyces astaci* in freshwater crayfish populations in Croatia. Diseases of Aquatic Organisms, **118**, 45-53.

Maguire, I., Klobučar, G., Žganec, K., Jelić, M., Lucić, A. and Hudina, S. 2018. Recent changes in distribution pattern of freshwater crayfish in Croatia − threats and perspectives. Knowledge and Management of Aquatic Ecosystems, **419(2)**, 62.

Makkonen J., Jussila J., Kortet R., Vainikka A. Kokko H. 2012. Differing virulence of *Aphanomyces astaci* isolates and elevated resistance of noble crayfish *Astacus astacus* against crayfish plague. Diseases of Aquatic Organisms. **102**, 129-136.

Makkonen J., Kokko H., Vainikka A., Kortet R. Jussila J. 2014. Dose-dependent mortality of the noble crayfish (*Astacus astacus*) to different strains of the crayfish plague (*Aphanomyces astaci*). Journal of Invertebrate Pathololgy. **115**, 86-91.

Makkonen, J., Vesterbacka, A., Martin, F., Jussila, J., Diéguez-Uribeondo, J., Kortet, R., Kokko, H. 2016. Mitochondrial genomes and comparative genomics of *Aphanomyces astaci* and *Aphanomyces invadans*. Scientific Reports, **6(1)**.

Manfrin A., Pretto T. 2014. Aspects of health and disease prevention. In: RARITY. Eradicate invasive Louisiana red swamp and preserve native white clawed crayfish in Friuli Venezia Giulia, RARITY project LIFE10 NAT/IT/000239, pp. 123–125.

Manfrin C., Tom M., De Moro G., Gerdol M., Giulianini P.G., Pallavicini A., 2014, The eyestalk transcriptome of red swamp crayfish *Procambarus clarkii*. Gene., **557(1),** 28-34.

Manfrin C., Tom M., De Moro G., Gerdol M., Guarnaccia C., Mosco A., Pallavicini A., Giulianini P.G., 2013, Application of D-Crustacean Hyperglycemic Hormone Induces Peptidases Transcription and Suppresses Glycolysis-Related Transcripts in the Hepatopancreas of the Crayfish *Pontastacus leptodactylus* - Results of a Transcriptomic Study. Plos ONE., **8(6)**, e65176.

Martin, P., Thonagel, S. Scholtz, G. 2016. The parthenogenetic Marmorkrebs (Malacostraca: Decapoda: Cambaridae) is a triploid organism. Journal of Zoological Systematics and Evolutionary Research., **54**, 13-21

Martín-Torrijos, L., Campos Llach, M., Pou-Rovira, Q., Diéguez-Uribeondo, J. 2017. Resistance to the crayfish plague, *Aphanomyces astaci* (Oomycota) in the endangered freshwater crayfish species, *Austropotamobius pallipes*. PLOS ONE, **12(7)**, e0181226.

Martín-Torrijos, L., Kawai, T., Makkonen, J., Jussila, J., Kokko, H., Diéguez-Uribeondo, J. 2018. Crayfish plague in Japan: A real threat to the endemic *Cambaroides japonicus*. PLOS ONE, **13(4)**, e0195353.

McCarthy, J. M., Hein, C. L., Olden, J. D., Vander Zanden. M. J., 2006. Coupling long-term studies with meta- analysis to investigate impacts of non-native crayfish on zoobenthic communities. Freshwater Biology, **51**, 224-235

Meng, X., Hong, L., Yang, T., Liu, Y., Jiao, T., Chu, X., Zhang, D., Wang, J., Tang, B., Liu, Q., Zhang, W. And He, W., 2019. Transcriptome-wide identification of differentially expressed genes in *Procambarus clarkii* in response to chromium challenge. Fish and Shellfish Immunology, **87**, 43-50.

Nawrocki, E. and Eddy, S., 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics, **29(22)**, 2933-2935.

Ng, T., Chiang, Y., Yeh, Y., Wang, H. 2014. Review of Dscam-mediated immunity in shrimp and other arthropods. Developmental and Comparative Immunology, **46(2)**, 129-138.

Novitsky RA, Son MO 2016. The first records of Marmorkrebs (*Procambarus fallax* (Hagen, 1870) f. *virginalis*) (Crustacea, Decapoda, Cambaridae) in Ukraine. Ecologica Montenegrina **5**, 44-46.

Nyhlén, L., Unestam, T. 1975. Ultrastructure of the penetration of the crayfish integument by the fungal parasite, *Aphanomyces astaci*, Oomycetes. Journal of Invertebrate Pathology, **26(3)**, 353-366.

Oidtmann, B. 2012. Review of biological factors relevant to import risk assessments for Epizootic Ulcerative Syndrome (*Aphanomyces invadans*). Transboundary and Emerging Diseases, **59**, 26-39.

Oliver, K.R., Greene, W.K., 2009. Transposable elements: powerful facilitators of evolution. Bioessays **31**, 703-714.

Ooi, M.C., Goulden, E.F., Smith, G.G., Bridle, A. R., 2019 Haemolymph microbiome of the cultured spiny lobster *Panulirus ornatus* at different temperatures. Scientific reports, **9**, 1677.

Ooi, M. C., Goulden, E. F., Smith, G.G., Nowak, B. F., Bridle, A. R. 2017 Developmental and gut-related changes to the microbiomes of the cultured juvenile spiny lobster *Panulirus ornatus*. FEMS Microbiol. Ecol., **93(12)**.

Orlić, K., Šver, L., Burić, L., Kazazić, S., Grbin, D., Maguire, I., Pavić, D., Hrašćan, R., Vladušić, T., Hudina, S. Bielen, A., 2020. Cuticle-associated bacteria can inhibit crayfish pathogen *Aphanomyces astaci*: Opening the perspective of biocontrol in astaciculture. Aquaculture, p.736112.

Pao, S., Paulsen, I., Saier, M., 1998. Major Facilitator Superfamily. Microbiology and Molecular Biology Reviews, **62(1)**, 1-34.

Paro, S., Imler, J.-L., 2016. Immunity in insects. In: Encyclopedia of Immunology Elsevier Science, **68**, 383-398.

Pârvulescu, L., Togor, A., Lele, S.-F., Scheu, S., Șinca, D., Panteleit, J., 2017 First established population of marbled crayfish *Procambarus fallax* (Hagen, 1870) f. *virginalis* (Decapoda, Cambaridae) in Romania. BioInvasions Records **6,** 357-362.

Patoka J., Buřič M., Kolář V., Bláha M., Petrtýl M., Franta P., Tropek R., Kalous L., Petrusek A., Kouba A., 2016. Predictions of marbled crayfish establishment in conurbations fulfilled: Evidences from the Czech Republic. Biologia, **71**, 1380-1385.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., Kingsford, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nature methods, **14(4)**, 417-419.

Piednoël, M., Donnart, T., Esnault, C., Graça, P., Higuet, D., Bonnivard, E., 2013. LTR-Retrotransposons in R. exoculata and Other Crustaceans: The Outstanding Success of GalEa-Like Copia Elements. Plos ONE, **8(3)**, e57675.

Pinzón Martín, S., Seeberger, P., Varón Silva, D. 2019. Mucins and Pathogenic Mucin-Like Molecules Are Immunomodulators During Infection and Targets for Diagnostics and Vaccines. Frontiers in Chemistry, **7**.

Pradhan, P., Verma, D., Peruzza, L., Gupta, S., Haq, S., Shubin, S. V., Morgan, K. L., Trusch, F., Mohindra, V., Hauton, C., Van West, P., Sood, N., 2020. Molecular insights into the mechanisms of susceptibility of *Labeo rohita* against oomycete *Aphanomyces invadans*. Scientific Reports, **10(1)**.

Praefcke, G. (2018). Regulation of innate immune functions by guanylate-binding proteins. International Journal of Medical Microbiology, **308(1)**, 237-245.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reynolds, J. D., 2011. A review of ecological interactions between crayfish and fish, indigenous and introducted. Knowledge and Management of Aquatic Ecosystems, **401**.

Richardson, D., Pysek, P., Rejmanek, M., Barbour, M., Panetta, F., West, C. 2000. Naturalization and invasion of alien plants: concepts and definitions. Diversity Distributions, **6(2)**, 93-107.

Ritchie M.E., Phipson B., Wu D., Hu Y., Law C.W., Shi W., Smyth G.K., 2015. "limma powers differential expression analyses for RNA-sequencing and microarray studies." Nucleic Acids Research, **43(7)**, e47

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J., Pachter, L. 2011. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biology*, ***12*****(3)**, R22.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., Mesirov. J.P., 2011. Integrative Genomics Viewer. Nature Biotechnology **29**, 24-26.

Roer, R., Dillaman, R. 1984. The Structure and Calcification of the Crustacean Cuticle. *American Zoologist*, ***24*****(4)**, 893-909.

Rončević, T., Čikeš-Čulić, V., Maravić, A., Capanni, F., Gerdol, M., Pacor, S., Tossi, A., Giulianini, P., Pallavicini, A. and Manfrin, C., 2020. Identification and functional characterization of the astacidin family of proline-rich host defence peptides (PcAst) from the red swamp crayfish

(*Procambarus clarkii*, Girard 1852). Developmental and Comparative Immunology, **105**, 103574.

Rowley, A. 2016. The Immune System of Crustaceans. Encyclopedia of Immunobiology, 437-453.

Rőszer, T., 2014, The invertebrate midintestinal gland ('hepatopancreas') is an evolutionary forerunner in the integration of immunity and metabolism. Cell and Tissue Research.,**358**, 685–695.

Samardžić M., Lucić A., Maguire I., Hudina S., 2014. The first record of the marbled crayfish (*Procambarus fallax* (Hagen, 1870) f. *virginalis*) in Croatia. Crayfish News, **36 (4).**

Sánchez-Salgado, J., Pereyra, M., Vivanco-Rojas, O., Sierra-Castillo, C., Alpuche-Osorno, J., Zenteno, E., Agundis, C. 2014. Characterization of a lectin from the craysfish *Cherax quadricarinatus* hemolymph and its effect on hemocytes. Fish and Shellfish Immunology, **39(2)**, 450-457.

Scholtz, G., Braband, A., Tolley, L., Reimann, A., Mittmann, B., Lukhaup, C., Steuerwald, F., Vogt, G., 2003. Parthenogenesis in an outsider crayfish. Nature, **421(6925)**, 806-806.

Scott C. 2016 dammit: an open and accessible de novo transcrip- tome annotator, https://github.com/dib-lab/dammit, accessed 30th of October 2020.

Scott, J. R., Thune, R. L., 1986. Bacterial flora of hemolymph from red swamp crawfish, *Procambarus clarkii* (Girard), from commercial ponds. Aquaculture, **58**, 161-165.

Seligo, A. 1895. Bemerkungen iiber Krebspest, Wasserpest, Lebensverhaltnisse des Krebses, Zeitschrift fiir Fischerei und deren Hilfswissenschaften, **3**, 247.

Seppey, M., Manni, M., Zdobnov, E. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods in Molecular Biology, 227-245.

Shen, G., Zhang, X., Gong, J., Wang, Y., Huang, P., & Shui, Y., Xu, Z. And Shen, H., 2020. Transcriptomic analysis of Procambarus clarkii affected by "Black May" disease. Scientific Reports, **10(1)**.

Shen, H., Hu, Y., Ma, Y., Zhou, X., Xu, Z., Shui, Y., Li, C., Xu, P., Sun, X., 2014. In-Depth Transcriptome Analysis of the Red Swamp Crayfish *Procambarus clarkii*. Plos ONE, ***9*(10)**, e110548.

Shi, X., Li, X., Wang, S., Zhao, X., Wang, J. 2010. Transcriptome analysis of hemocytes and hepatopancreas in red swamp crayfish, *Procambarus clarkii*, challenged with white spot syndrome virus. Invertebrate Survival Journal, **7(1)**, 119-131.

Shi, X.Z., Zhao, X.F., Wang, J.X., 2014. A new type of antimicrobial peptide astacidin functions in antibacterial immune response in red swamp crayfish *Procambarus clarkii*, Developmental and Comparative Immunology, **43**, 121e128.

Shi, X.-L., Wang, L., Xu, S., Zhan, X.-W., Zhao, X.-F., Vasta, G. R., Wang, J.-W., 2014. A galectin from the Kuruma shrimp (*Marsupenaeus japonicus*) functions as an opsonin and promotes bacterial clearance from hemolymph. Plos ONE, **9**, e91794.

Sierra C., Lascurain R., Pereyra A., Guevara J., Martınez G., Agundis C., Zenteno, E., Vázquez, L., 2005. Participation of serum and membrane lectins on the oxidative burst regulation in *Macrobrachium rosenbergii* hemocytes. Developmental and Comparative Immunology, **29**, 113e21.

Smith, V.J., Desbois, A.P., Dyrynda, E.A., 2010. Conventional and unconventional antimicrobials from fish, marine invertebrates and micro-algae. Marine Drugs **8**, 1213-1262.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. and Kelly, S., 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Research, **26(8)**, 1134-1144.

Souty-Grosset, C., Holdich, D. M., Noël, P. Y., Reynolds, J. D., Haffner P. 2006. Atlas of Crayfish in Europe, Paris: Muséum national d'Histoire naturelle.

Söderhäll, K., Cerenius, L., 1992., Crustacean immunity. Annual Review of Fish Diseases., **2**, 3-23.

Söderhäll, K., Cerenius, L., 1998. Role of the prophenoloxidase-activating system in invertebrate immunity. Current Opinion in Immunology **10**, 23–38

Söderhäll, I., Wu, C., Novotny, M., Lee, B.L., Söderhäll, K., 2009. A novel protein acts as a negative regulator of proPO activation and melanisation in the freshwater crayfish, *Pacifastacus leniusculus*. Journal of Biological Chemistry. **284**, 6301–6310

Sricharoen, S., Kim, J.J., Tunkijjaukij, S., Söderhäll, I., 2005. Exocytosis and proteomic analysis of the vesicle content of granular hemocytes from a crayfish. Developmental and Comparative Immunology, **29**, 1017e1031

Srivastava, A., Malik, L., Sarkar, H., Zakeri, M., Almodaresi, F., Soneson, C. Love, M. I., Kingsford, C., Patro, R. 2020. Alignment and mapping methodology influence transcript abundance estimation. Genome Biology, **21(1)**.

Strand D.A., Jussila J., Viljamaa-Dirks S., Kokko H., Mak- konen J., Holst-Jensen A., Viljugrein H. Vrålstad T. 2012. Monitoring the spore dynamics of *Aphanomyces astaci* in the ambient water of latent carrier crayfish. Veterinary Microbiology. **160**, 99-10.

Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M., Paulovich, A., Pomeroy, S., Golub, T., Lander, E. and Mesirov, J., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, **102(43)**, 15545-15550.

Storminger, J.L., 2009. Animal antimicrobial peptides: ancient players in innate immunity. Journal of Immunology, **182**, 6633-6634.

Sun, C., Xu, W., Zhang, H., Dong, L., Zhang, T., Zhao, X., Wang, J. 2011. An anti-lipopolysaccharide factor from red swamp crayfish, *Procambarus clarkii*, exhibited antimicrobial activities *in vitro* and *in vivo*. Fish and Shellfish Immunology, **30(1)**, 295-303.

Suzek, B., Huang, H., McGarvey, P., Mazumder, R. and Wu, C., 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics, **23(10),** 1282-1288.

Svoboda, J., Kozubíková-Balcarová, E., Kouba, A., Buřič, M., Kozák, P., Diéguez-Uribeondo, J., Petrusek, A., 2013. Temporal dynamics of spore release of the crayfish plague pathogen from its natural host, American spiny-cheek crayfish (*Orconectes limosus*), evaluated by transmission experiments. Parasitology, **140(6)**, 792-801.

Svoboda J., Kozubíková E., Kozák P., Kouba A., Bahadir Koca S., Diler Ö., Diler I., Policar T., Petrusek A. 2012. PCR detection of the crayfish plague pathogen in narrow-clawed crayfish inhabiting Lake Eğirdir in Turkey. Diseases of Aquatic Organisms, **98**, 255–259.

Tan, M., Gan, H., Lee, Y., Grandjean, F., Croft, L., Austin, C. 2020. A Giant Genome for a Giant Crayfish (*Cherax quadricarinatus*) With Insights Into cox1 Pseudogenes in Decapod Genomes. Frontiers in Genetics, **11**.

Tarasov, A., Vilella, A., Cuppen, E., Nijman, I. and Prins, P., 2015. Sambamba: fast processing of NGS alignment formats. Bioinformatics, **31(12)**, 2032-2034.

Tassanakajon, A., Somboonwiwat, K., Amparyup, P. 2015. Sequence diversity and evolution of antimicrobial peptides in invertebrates. Developmental and Comparative Immunology, **48(2)**, 324-341.

Theissinger K, Falckenhayn C, Blande D, Toljamo A, Gutekunst J, Makkonen J, Jussila J, Lyko F, Schrimpf A, Schulz R, Kokko H. 2016. *De Novo* assembly and annotation of the freshwater crayfish *Astacus astacus* transcriptome. Marine Genomics. **28**,7-10.

Tom M., Manfrin C., Chung S.J., Sagi A., Gerdol M., De Moro G., Pallavicini A., Giulianini P.G., 2014, Expression of cytoskeletal and molt-related genes is temporally scheduled in the hypodermis of the crayfish *Procambarus clarkii* during premolt. Journal of Experimental Biology, **217**, 4193-202.

Tom M, Manfrin C, Giulianini PG, Pallavicini A., 2013, Crustacean oxi-reductases protein sequences derived from a functional genomic project potentially involved in ecdysteroid hormones metabolism - a starting point for function examination, General and Comparative Endocrinology, **194**, 71-80

Tom, M., Manfrin, C., Mosco, A., Gerdol, M., De Moro, G., Pallavicini, A. and Giulianini, P., 2014. Different transcription regulation routes are exerted by L- and D-amino acid enantiomers of peptide hormones. Journal of Experimental Biology, **217(24)**, 4337-4346.

Twardochleb LA, Olden JD, Larson ER. 2013. A global meta-analysis of the ecological impacts of nonnative crayfish. Freshwater Science, **32,** 1367-1382.

Unestam, T., Weiss, D. 1970. The Host-Parasite Relationship between Freshwater Crayfish and the Crayfish Disease Fungus *Aphanomyces astaci*: Responses to Infection by a Susceptible and a Resistant Species. Journal of General Microbiology, **60(1)**, 77-90.

Vasta GR. 2009. Roles of galectins in infection. Nature Reviews Microbiology, **7**, 424e38.

Verma, D., Peruzza, L., Trusch, F., Yadav, M., Ravindra, Shubin, S. V., Morgan, K. L., Mohindra, V., Hauton, C., Van West, P., Pradhan, P., Sood, N. 2020, Transcriptome analysis reveals immune pathways underlying resistance in the common carp *Cyprinus carpio* against the oomycete *Aphanomyces invadans*. Genomics (in press).

Vishwanath, T.S., Mohan, C.V., Shankar, K.M., 1998, Epizootic Ulcerative Syndrome (EUS), associated with a fungal pathogen, in Indian fishes: histopathology-`a cause for invasiveness', Aquaculture **165(1-2)**, 1-8.

Vogt, G., Tolley, L. Scholtz, G. 2004. Life stages and reproductive components of the Marmorkrebs (marbled crayfish), the first parthenogenetic decapod crustacean. Journal of Morphology, **261**, 286-311

Vogt G., 2008. The marbled crayfish: a newmodel organism for research on development, epigenetics and evolutionary biology. Journal of Zoology, **276**, 1-13.

Vojkovská R., Horká I., Tricarico E., Ďuriš Z. 2014. New record of the parthenogenetic marbled crayfish *Procambarus fallax* f. *virginalis* from Italy. Crustaceana, **87**, 1386-1392.

Wang, X.-W., Wang, J.-X. 2015. Crustacean hemolymph microbiota: Endemic, tightly controlled, and utilization expectable. Molecular Immunology, **68**, 404–411.

Wang X.W., Xu Y.H., Xu J.D., Zhao X.F., Wang J.X., 2014, Collaboration between a soluble C-type lectin and calreticulin facilitates white spot syndrome virus infection in shrimp. Journal of Immunology. **193(5)**, 2106-17.

Wang, Z., Chen, Y., Dai, Y., Tan, J., Huang, Y., Lan, J., Ren, Q. 2015. A novel vertebrates Toll-like receptor counterpart regulating the anti-microbial peptides expression in the freshwater crayfish, *Procambarus clarkii*. Fish and Shellfish Immunology, **43(1)**, 219-229.

Watson, F.L., Püttmann-Holgado, R., Thomas, F. Lamar, D.L., Hughes, M., Kondo, M., Rebel, V.I., Schmucker, D., 2006. Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science **309**, 1874–1878.

Westman K., Savolainen R., Julkunen M. 2002. Replacement of the native crayfish *Astacus astacus* by the introduced species *Pacifastacus leniusculus* in a small enclosed Finnish lake: a 30 year study. Ecography, **25**, 53-73.

Weiperth A., Csányi B., Gál B., György Á.I., Szalóky Z., Szekeres J., Tóth B., Puky M. 2015. Egzotikus rákhalés kétéltűfajok a Budapest környéki víztestekben [Exotic crayfish, fish and amphibian species in various water bodies in the region of Budapest]. Pisces Hungarici, **9**, 65-70.

Wheeler, T. and Eddy, S., 2013. nhmmer: DNA homology search with profile HMMs. Bioinformatics, **29(19)**, 2487-2489.

White, K.N., Ratcliffe, N.A., 1982. The segregation and elimination of radio and fluorescent-labelled marine bacteria from the haemolymph of the shore crab, *Carcinus maenas*. Journal of the Marine Biological Association UK, **62**, 819-833

White, K.N., Ratcliffe, N.A., Rossa, M., 1985. The antibacterial activity of haemocyte clumps in the gills of the shore crab, *Carcinus maenas*. Journal of the Marine Biological Association UK, **65**, 857-870.

Wingett, S. W., Andrews, S. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research, **7**, 1338.

Wu, C., Söderhäll, K., Söderhäll, I., 2011. Two novel ficolin-like proteins (FLPs) act as pattern recognition receptors for invading pathogens in the freshwater crayfish *Pacifastacus leniusculus*. Proteomics, **11**, 2249e2264.

Xu, S., Jing, M., Liu, W. Y., Dong, H., Kong, D. M., Wang, Y. R., Zhang, H. H., Yue, Z., Li, Y. J., Jiao, F., Xie, S. Y. 2020. Identification and characterization of a novel L-type lectin (MjLTL2) from kuruma shrimp (*Marsupenaeus japonicus*). Fish and shellfish immunology, **98**, 354–363.

Yamakawa, K., Huot, Y. K., Haendelt, M. A., Hubert, R., Chen, X. N., Lyons, G. E., Korenberg, J. R. 1998. DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. Human molecular genetics, **7(2)**, 227-237.

Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., Xu, H., Huang, X., Li, S., Zhou, A., Zhang, X., Bolund, L., Chen, Q., Wang, J., Yang, H., Fang, L. and Shi, C., 2018. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. Nucleic Acids Research, **46(W1)**, W71-W75.

Yu, G., Wang, L., Han, Y., He, Q., 2012. “clusterProfiler: an R package for comparing biological themes among gene clusters.” OMICS: A Journal of Integrative Biology, **16(5)**, 284-287.

Yu G, Zhang D, Guo F, Tan M, Jiang X, Jiang, W., 2013, Cryo-EM Structure of a Novel Calicivirus, Tulane Virus. Plos ONE, **8(3)**, e59817.

Zhang, Y., Li, Z., Kholodkevich, S., Sharov, A., Feng, Y., Ren, N., Sun, K. 2019. Cadmium-induced oxidative stress, histopathology, and transcriptome changes in the hepatopancreas of freshwater crayfish (*Procambarus clarkii*). Science of The Total Environment, **666**, 944-955.

Zhang, X., Man, X., Huang, X., Wang, Y., Song, Q., Hui, K., Zhang, H. 2018. Identification of a C-type lectin possessing both antibacterial and antiviral activities from red swamp crayfish. Fish and Shellfish Immunology, **77**, 22-30.

Zhang, X., Zhang, R., & Yu, J. 2020. New Understanding of the Relevant Role of LINE-1 Retrotransposition in Human Disease and Immune Modulation. Frontiers in Cell And Developmental Biology, **8**.

Zhao, Z.Y., Yin, Z.X., Xu, X.P., Weng, S.P., Rao, X.Y., Dai, Z.X., Luo, Y.W., Yang, G., Li, Z.S., Guan, H.J., Li, S.D., Chan, S.M., Yu, X.Q., He, J.G., 2009. A novel C-type lectin from the shrimp *Litopenaeus vannamei* possesses anti-white spot syndrome virus activity, Journal of Virology, **83(1)**, 347-56

Zhu, A., Ibrahim, J. and Love, M., 2018. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics, **35(12)**, 2084-2092.

8. Supplementary

Supplementary Table 1. Reads statistics for raw and processed samples.

		Raw reads				*Processed reads*			
Treatment	***Sample Name***	**Duplicated (%)**	**% GC**	**Length**	**M Seqs**	**Duplicated (%)**	**% GC**	**Length**	**M Seqs**
As	***R10_forward_paired***	80.80	44	150	63.8	81.10	44	148	60.3
As	***R10_reverse_paired***	81.00	44	150	63.8	81.80	44	148	60.3
As	***R1_forward_paired***	77.90	45	150	47.7	78.20	45	148	45.2
As	***R1_reverse_paired***	77.30	46	150	47.7	78.30	45	148	45.2
As	***R2_forward_paired***	76.80	46	150	47.6	77.10	46	148	44.8
As	***R2_rreverse_paired***	76.50	46	150	47.6	77.40	46	148	44.8
As	***R3_forward_paired***	78.80	45	150	61.9	79.10	45	148	58.6
As	***R3_rreverse_paired***	78.50	45	150	61.9	79.50	45	148	58.6
As	***R4_forward_paired***	79.70	46	150	52.9	80.00	46	148	49.9
As	***R4_rreverse_paired***	79.10	46	150	52.9	80.10	46	148	49.9
As	***R5_forward_paired***	76.50	46	150	47.2	76.80	46	148	44.6
As	***R5_rreverse_paired***	76.10	46	150	47.2	76.90	46	148	44.6
As	***R6_forward_paired***	77.10	45	150	52.3	77.30	45	148	49.7
As	***R6_rreverse_paired***	76.90	45	150	52.3	77.70	45	148	49.7
As	***R7_forward_paired***	80.60	46	150	47	80.90	45	148	44.6
As	***R7_rreverse_paired***	80.40	46	150	47	81.30	46	148	44.6
As	***R8_forward_paired***	79.00	46	150	41.2	79.30	45	148	39.1
As	***R8_rreverse_paired***	78.70	46	150	41.2	79.50	45	148	39.1
As	***R9_forward_paired***	77.80	44	150	51.4	78.10	44	148	48.7
As	***R9_rreverse_paired***	77.60	44	150	51.4	78.50	44	148	48.7
Control	***R21_forward_paired***	77.00	45	150	51.2	77.20	44	148	47.5
Control	***R21_reverse_paired***	75.30	45	150	51.2	76.40	44	147	47.5
Control	***R22_forward_paired***	77.20	46	150	48.3	77.20	46	148	45.4
Control	***R22_reverse_paired***	76.60	46	150	48.3	77.30	46	148	45.4
Control	***R23_forward_paired***	77.50	45	150	50.6	77.60	45	148	48.2

Control	***R23_reverse_paired***	76.80	45	150	50.6	77.60	45	148	48.2
Control	***R24_forward_paired***	79.30	46	150	62.6	79.60	46	148	59.5
Control	***R24_reverse_paired***	78.60	46	150	62.6	79.70	46	148	59.5
Control	***R25_forward_paired***	79.10	44	150	60	79.50	44	148	56.8
Control	***R25_reverse_paired***	79.00	45	150	60	79.90	44	148	56.8
Control	***R26_forward_paired***	79.20	45	150	61.1	79.60	45	148	57.8
Control	***R26_reverse_paired***	78.90	45	150	61.1	80.00	45	148	57.8
Control	***R27_forward_paired***	79.90	45	150	52.9	80.30	45	148	50.2
Control	***R27_reverse_paired***	79.60	45	150	52.9	80.60	45	148	50.2
Control	***R28_forward_paired***	75.20	46	150	54.8	75.50	46	148	51.8
Control	***R28_reverse_paired***	75.40	46	150	54.8	76.10	46	148	51.8
Control	***R29_forward_paired***	78.70	46	150	68.9	79.20	45	148	65
Control	***R29_reverse_paired***	78.30	46	150	68.9	79.40	45	148	65
Control	***R30_forward_paired***	72.10	45	150	47.1	72.30	45	148	44.6
Control	***R30_reverse_paired***	71.70	45	150	47.1	72.50	45	148	44.6
PsI	***R41_forward_paired***	78.00	45	150	52.2	78.10	45	148	49.4
PsI	***R41_reverse_paired***	76.90	45	150	52.2	77.80	45	148	49.4
PsI	***R42_forward_paired***	76.90	44	150	47	76.90	44	148	44.4
PsI	***R42_reverse_paired***	76.00	44	150	47	76.70	44	148	44.4
PsI	***R43_forward_paired***	80.40	46	150	47.6	80.40	45	148	44.9
PsI	***R43_reverse_paired***	78.90	46	150	47.6	79.90	45	148	44.9
PsI	***R44_forward_paired***	76.00	44	150	47.4	76.00	44	148	44.6
PsI	***R44_reverse_paired***	75.00	44	150	47.4	75.70	44	148	44.6
PsI	***R45_forward_paired***	79.20	45	150	51.3	79.40	45	148	48.5
PsI	***R45_reverse_paired***	78.00	45	150	51.3	79.00	45	148	48.5
PsI	***R46_forward_paired***	76.40	45	150	44.6	76.40	45	148	42
PsI	***R46_reverse_paired***	75.40	45	150	44.6	76.30	45	148	42
PsI	***R47_forward_paired***	74.70	45	150	44.7	74.70	45	148	42.4
PsI	***R47_reverse_paired***	73.70	45	150	44.7	74.60	45	148	42.4
PsI	***R48_forward_paired***	74.70	45	150	47	74.80	44	148	44.3
PsI	***R48_reverse_paired***	73.10	45	150	47	74.10	44	148	44.3

PsI	***R49_forward_paired***	74.70	44	150	40.6	74.80	44	148	38
PsI	***R49_reverse_paired***	73.60	44	150	40.6	74.50	44	147	38
PsI	***R50_forward_paired***	77.30	45	150	56.2	77.30	45	148	52.6
PsI	***R50_reverse_paired***	76.20	46	150	56.2	77.00	45	147	52.6
	Mean	77.29	45.18	150.00	51.64	77.85	44.95	147.95	48.78
	Median	77.23	45.20	150.00	51.43	77.79	44.97	147.95	48.59
	Min	71.70	44.00	150.00	40.60	72.30	44.00	147.00	38.00
	Max	81.00	46.00	150.00	68.90	81.80	46.00	148.00	65.00

Supplementary table 2. List of differentially expressed genes and their annotations for marbled crayfish challenged with two strains of *A. astaci* (low virulence As strain and high virulence PsI strain) sampled at two sampling points. Sampling point I- 3 days after treatment and Sampling point II- 21 days after treatment.

Trinity Gene ID	**Isoform count**	**Base Mean expression**	**log_2 fold change**	**p-value**	**Annotation database**	**Annotation**	**Description**	**KEGG**
a) Sampling point I- As treatment								
DN3782 c0 g4	1	81.17	23.05	1.28E-10	NA	NA	NA	NA
DN6129 c0 g3	1	10.72	6.58	2.71E-09	NA	NA	NA	NA
DN2080 c2 g1	1	40.2	2.65	1.14E-06	NA	NA	NA	NA
DN1 c6 g1	1	42.64	-1.2	1.99E-06	NA	NA	NA	NA
b) Sampling point I- PsI treatment								
DN3782 c0 g4	1	81.17	22.45	3.79E-10	NA	NA	NA	NA
DN139895 c0 g1	1	1910.21	5.28	1.43E-05	Pfam	RNA helicase, RNAdRNApoly	Viral RNA-dependent RNA polymerase	NA
DN19756 c0 g1	2	471.66	-0.81	2.46E-05	NA	NA	NA	NA
DN1713 c1 g1	1	333.38	-1.63	2.69E-05	NA	NA	NA	NA
DN12642 c1 g1	5	224.56	-2	9.67E-09	Pfam	DUF4499	Uncharacterised protein	NA
DN6324 c0 g1	1	434.93	-2.59	2.16E-05	InterPro5	C-type lectin-like	C-type lectin	K14356
DN2571 c1 g3	1	1640.45	-3.63	7.15E-06	Pfam	PEPCK	Phosphoenolpyruvate carboxykinase	NA
DN1271 c1 g1	1	638.09	-5.12	3.26E-06	OrthoDB	ENSTRUP00000002137	Pleckstrin homology domain	NA

DN2491 c0 g2	3	1195.46	-5.3	4.62E-07	Pfam	OATP	Organic anion-transporting polypeptide	NA
c) Sampling point II- As treatment								
DN46345 c0 g2	2	23.1	3.95	7.18E-13	Rfam	SSU rRNA eukarya	Small ribosomal subunit	NA
DN14890 c0 g1	4	17.4	3.84	2.81E-10	Rfam	LSU rRNA eukarya	Large ribosomal subunit	NA
DN45321 c0 g1	1	25.04	2.8	9.65E-08	NA	NA	NA	NA
DN16 c2 g1	1	387.74	2.76	2.19E-05	Pfam	p450	Cytochrome P450 2J3-like	K07418
DN28222 c0 g2	1	10.94	2.64	2.30E-05	Rfam	LSU rRNA eukarya	Large ribosomal subunit	NA
DN439 c0 g4	3	127.95	2.33	5.64E-06	NA	NA	NA	NA
DN1413 c0 g2	2	337.57	2.15	4.00E-09	Rfam	LSU rRNA eukarya	Large ribosomal subunit	NA
DN48141 c0 g1	2	22.77	1.98	4.04E-07	Rfam	SSU rRNA eukarya	Small ribosomal subunit	NA
DN10004 c0 g1	1	21.62	1.94	6.21E-06	NA	NA	NA	NA
DN18944 c0 g1	4	27.59	1.85	4.75E-05	Rfam	LSU rRNA eukarya	Large ribosomal subunit	NA
DN20347 c2 g2	3	534.67	1.25	8.69E-06	Pfam	FCH, SH3 domain	Regulation of nitric-oxide synthase activity	K20126
DN7783 c0 g1	2	312.93	0.83	2.98E-07	Pfam	OHCU decarbox	OHCU decarboxylase	K13485
DN645 c2 g1	5	118.77	0.72	1.25E-05	NA	NA	NA	NA
DN4193 c0 g1	1	861.93	0.64	2.38E-05	OrthoDB	ENSDARP00000106918	Uncharacterised protein	NA
DN863 c0 g1	7	1605.92	0.61	1.84E-05	Pfam	MFS 1	Membrane transport	NA
DN140686 c0 g1	1	203.48	0.6	1.30E-05	Pfam	Med6	Activation of many RNA polymerase II promoters	K15128
DN631 c0 g1	7	7607.96	0.51	2.35E-05	Pfam	Glycohydro 20b2	Glycoside hydrolase family	K12373
DN1413 c0 g1	8	4805.89	-0.73	1.19E-06	Rfam	LSU rRNA eukarya	Large ribosomal subunit	NA
DN917 c0 g2	1	140	-1.13	3.47E-08	NA	NA	NA	NA
DN14571 c0 g1	4	38.68	-1.13	4.42E-06	NA	NA	NA	NA
DN15920 c0 g1	3	48.07	-1.25	9.19E-07	NA	NA	NA	NA
DN12920 c0 g1	2	38.95	-1.28	4.80E-05	OrthoDB	XP 004209983.1	RNA-directed DNA polymerase from mobile element jockey-like	NA
DN1814 c0 g1	3	537.77	-1.33	2.85E-05	NA	NA	NA	NA
DN18278 c0 g1	5	66.9	-1.44	1.08E-05	UniRef90	UPI000DBEE85B	Invertebrate connectin	NA
DN6767 c1 g1	1	29.41	-1.47	3.38E-06	NA	NA	NA	NA
DN24 c2 g2	1	36.26	-1.53	4.54E-07	NA	NA	NA	NA
DN1776 c0 g2	3	294.48	-1.58	1.55E-06	Pfam	Mucin-like	Mucin-like domain-containing protein	NA
DN20079 c1 g1	4	21.72	-1.6	1.80E-05	NA	NA	NA	NA

DN19074 c0 g1	1	22.92	-1.83	9.83E-06	OrthoDB	ENSETEP00000013416	Microtubule-associated protein 1B	NA
DN17242 c0 g1	2	15.25	-1.88	2.45E-07	OrthoDB	Sakowv30043123m	EGF-like calcium-binding domain	NA
DN51753 c0 g1	8	349.58	-1.91	2.66E-06	OrthoDB	ENSSTOP00000000559	colony stimulating factor 3 receptor (granulocyte)	NA
DN17169 c0 g1	1	28.12	-1.98	1.15E-05	NA	NA	NA	NA
DN6337 c0 g1	1	48.53	-2.16	1.96E-07	OrthoDB	FBpp0261707	Metal ion binding, uncharacterised protein	NA
d) Sampling point II- PsI treatment								
DN35888 c0 g1	7	2866.73	5.12	2.55E-07	NA	NA	NA	NA
DN14890 c0 g1	4	17.4	4.42	2.37E-13	Rfam	LSU_rRNA_eukarya	Large ribosomal subunit	NA
DN46345 c0 g2	2	23.1	4.37	1.43E-15	Rfam	SSU_rRNA_eukarya	Small ribosomal subunit	NA
DN57682 c1 g2	4	17.06	3.98	3.18E-07	NA	NA	NA	NA
DN12252 c0 g1	3	23.17	3.47	1.39E-05	Pfam	Serpin	Serine protease inhibitor	NA
DN36498 c0 g1	2	11.04	3.43	1.61E-05	Pfam	Transposase_22_RBD-like domain, Tnp_22_trimer	L1 retrotransposon	NA
DN28222 c0 g2	1	10.94	3.32	6.04E-08	Rfam	LSU_rRNA_eukarya	Large ribosomal subunit	NA
DN18537 c0 g1	6	52.48	3.24	4.32E-09	OrthoDB	ENSMMUP00000040168	L1 retrotransposon	NA
DN5965 c0 g1	3	58.51	3.16	1.14E-08	NA	NA	NA	NA
DN23859 c0 g1	1	19.2	3.07	6.61E-06	Pfam	L1 retrotransposon	L1 retrotransposon	NA
DN5306 c0 g1	2	26.99	3.01	2.72E-07	Rfam	SSU_rRNA_eukarya	Small ribosomal subunit	NA
DN14628 c0 g2	1	20.55	2.92	7.16E-06	OrthoDB	ENSNLEP00000023933	Endonuclease/exonuclease/phosphatase superfamily	NA
DN48141 c0 g1	2	22.77	2.6	1.19E-11	Rfam	SSU_rRNA_eukarya	Small ribosomal subunit	NA
DN20205 c0 g1	2	15.14	2.59	8.43E-05	OrthoDB	OVOC12968	Cell adhesion, predicted protein	NA
DN1413 c0 g2	2	337.57	2.48	1.15E-11	Rfam	LSU_rRNA_eukarya	Large ribosomal subunit	NA
DN55362 c0 g1	2	40.09	2.41	7.17E-05	Pfam	AMP-binding, AMP-binding C	Acetyl-CoA synthetase-like, uncharacterised protein	NA
DN10823 c0 g1	1	9.57	2.31	4.75E-06	NA	NA	NA	NA
DN713 c0 g1	10	2391.17	2.24	3.02E-06	Pfam	Sec1	Vesicle transport processes	K23281
DN18944 c0 g1	4	27.59	2.14	2.14E-06	Rfam	LSU_rRNA_eukarya	Large ribosomal subunit	NA
DN12870 c0 g1	2	16.04	2.05	1.20E-05	Rfam	LSU_rRNA_eukarya	Large ribosomal subunit	NA
DN83351 c0 g1	1	574.02	1.98	2.90E-05	Pfam	WAP domain	Crustin	K23638
DN20109 c0 g1	2	18.67	1.75	2.84E-06	NA	NA	NA	NA
DN1069 c0 g1	2	946.98	1.71	7.22E-05	OrthoDB	LDEC000748-RA	Uncharacterised protein	NA

DN7962 c0 g1	1	33.01	1.66	2.94E-05	Pfam	Lectin_C	Lectin C	NA
DN17789 c0 g1	5	174.46	1.15	3.45E-05	Pfam	Guanylate kinase/PDZ/SH3_1/2	Guanylate kinase, Uncharacterised protein	NA
DN52580 c0 g1	1	129.23	1.1	5.88E-05	OrthoDB	R9X8P5_ASHAC	Uncharacterised protein	NA
DN645 c2 g1	5	118.77	0.89	4.99E-08	NA	NA	NA	NA
DN5470 c0 g1	2	110.77	0.85	1.34E-05	KEGG	Apolipoprotein D and lipocalin family protein	NA	K03098
DN863 c0 g1	7	1605.92	0.84	4.47E-09	Pfam	MFS_1	Membrane transport	NA
DN1748 c1 g1	2	2929.06	0.83	2.10E-05	Pfam	Aminotran_3	Aminotransferase class-III	K13524
DN3059 c1 g1	1	658.2	0.58	4.26E-05	Pfam	GBP, FYVE x2	Guanylate-binding protein	K17603
DN1380 c0 g1	5	419.95	0.55	8.85E-05	NA	NA	NA	NA
DN140686 c0 g1	1	203.48	0.54	7.19E-05	Pfam	Med6	Activation of many RNA polymerase II promoters	K15128
DN789 c0 g1	7	2422.42	0.51	5.37E-06	Pfam	Aldedh	Oxidation of aldehydes to carboxylic acids	K23234
DN32562 c1 g2	3	1271.85	0.5	8.83E-05	OrthoDB	3944354.p	Oxidoreductase domain, uncharacterised protein	NA
DN1113 c1 g1	2	3797.73	0.46	2.48E-05	Pfam	Aldedh	Succinate-semialdehyde dehydrogenase	K00139
DN4093 c0 g1	1	28399.91	0.34	7.13E-05	OrthoDB	LFUL008294-RA	Basic-leucine zipper (bZIP) domain transcription factor	K04374
DN11941 c0 g2	2	1878.06	0.31	3.71E-06	Pfam	Fer2_3	Succinate dehydrogenase/fumarate reductase	K00235
DN454 c1 g1	5	2465.73	-0.35	2.16E-05	NA	NA	NA	NA
DN246 c3 g1	5	783.76	-0.39	3.81E-05	OrthoDB	FOCC010058-RA	Uncharacterised protein	NA
DN1936 c2 g2	2	1371.02	-0.53	6.32E-07	NA	NA	NA	NA
DN634 c0 g1	2	1987.27	-0.54	4.89E-08	NA	NA	NA	NA
DN1284 c0 g1	6	374.04	-0.57	5.08E-05	NA	NA	NA	NA
DN651 c0 g1	4	194.65	-0.68	5.17E-05	Pfam	RAM	Maintainence of mRNA expression levels	NA
DN285 c1 g2	1	720.61	-0.71	1.47E-10	Pfam	RRM_1	RNA binding protein	K13107
DN15225 c0 g1	1	50.69	-0.83	0.000102	NA	NA	NA	NA
DN23876 c0 g1	7	99.6	-0.92	4.47E-05	OrthoDB	ENSTRUP00000008013	Wilm's tumour protein	NA
DN36223 c0 g1	2	34.8	-0.95	1.57E-05	OrthoDB	BGLTMP000799-PA	Intermediate filament protein	NA
DN1620 c0 g1	8	364.87	-1.03	1.04E-09	Pfam	RRM_6	RNA recognition motif	NA
DN9141 c0 g1	4	50.04	-1.1	9.51E-05	Pfam	Ig_3 x 3, I-set, C2-set_2 domain	Uncharacterised protein	NA
DN3139 c0 g1	2	72.32	-1.14	8.99E-05	NA	NA	NA	NA

DN8576 c0 g1	2	57.65	-1.19	3.73E-05	NA	NA	NA	NA
DN1610 c3 g1	1	106.51	-1.2	4.24E-05	NA	NA	NA	NA
DN4081 c0 g2	1	73.8	-1.25	7.09E-05	OrthoDB	BGLTMP004142-PA	Uncharacterised protein	NA
DN7737 c0 g1	3	40.83	-1.49	1.20E-05	UniRef90	UPI0007718C59	Uncharacterised protein	NA
DN3808 c3 g1	1	56.34	-1.52	5.62E-07	NA	NA	NA	NA
DN18278 c0 g1	5	66.9	-1.54	2.72E-06	UniRef90	Q95YM2	Invertebrate connectin	NA
DN7979 c0 g1	3	978.02	-1.56	5.72E-06	NA	NA	NA	NA
DN1981 c4 g2	1	29.61	-1.56	4.98E-06	OrthoDB	Bm2572b	Cyclophilin-type peptidyl-prolyl cis-trans isomerase,chaperone-like functions	NA
DN3567 c0 g1	2	288.95	-1.62	7.02E-05	NA	NA	NA	NA
DN4145 c0 g1	3	29.64	-1.68	1.46E-05	OrthoDB	HSAL11629-PA	Polyphosphate kinase 2 family	NA
DN5846 c0 g3	2	23.98	-1.85	4.27E-06	NA	NA	NA	NA
DN40700 c0 g1	3	11.63	-1.88	5.70E-05	OrthoDB	ADAC002598-PA	Ribulose-phosphate 3-epimerase	NA
DN2020 c0 g2	6	946.78	-1.95	5.04E-09	OrthoDB	BGLTMP003164-PA	Transient receptor potential cation channel	NA
DN51753 c0 g1	8	349.58	-2	8.64E-07	OrthoDB	ENSSTOP00000000559	Colony stimulating factor 3 receptor (granulocyte)	NA
DN9256 c0 g1	1	20.24	-2	1.28E-05	NA	NA	NA	NA
DN19074 c0 g1	1	22.92	-2.06	8.35E-07	OrthoDB	ENSETEP00000013416	Microtubule-associated protein 1B	NA
DN131436 c0 g1	1	8.66	-2.09	5.07E-05	UniRef90	UPI00094E451C	Titin-like protein	NA
DN69895 c0 g1	1	9.14	-2.12	4.52E-05	NA	NA	NA	NA
DN6337 c0 g1	1	48.53	-2.16	1.67E-07	OrthoDB	FBpp0261707	Metal ion binding, Uncharacterised protein	NA
DN1776 c0 g2	3	294.48	-2.25	1.06E-11	Pfam	Mucin-like	Mucin-like domain-containing protein	NA

Supplementary code 1. RScript used for the Differential gene expression analysis and KEGG over-representation analysis

```
---
title: "Master_data_analysis"
output: html_notebook
---
Data import
```{r}
#Libraries used in the analysis
library(tximport)
library(stringr)
library(DESeq2)
library(ggplot2)
library(hexbin)
library(magrittr)
library(vsn)
library(PoiClaClu)
library(pheatmap)
library(RColorBrewer)
library(FactoMineR)
library(factoextra)
library(dplyr)
library(tibble)
library(limma)
library(apeglm)
library(EnhancedVolcano)
library(readr)
library(data.table)
library(clusterProfiler)
library(Biostrings)
library(enrichplot)

#Samples is an excel spreadsheet that contains information about the data
(biological conditions, sampling)
Samples <- read_excel("Samples.xlsx")

#This code leads to the directories that contain the sample counts produced
by Salmon and imports them to R
dir<- file.path("F:", "Salmon", "Trinity", "Salmon_trinity_results")
files<- file.path(dir, "500", Samples$Sample_name,"quant.sf")
names(files)<-paste("samples",1:30)
all(file.exists(files))
txi<-tximport(files, type="salmon", txOut = T)

#Obtain Trinity gene names from Trinity transcripts
names<-rownames(txi$counts)
names2<-str_extract(names, ".*_.*_.*_.*[1,2,3,4,5,6,7,8,9,0]_")
names3<-str_extract(names2, ".*_.*_.*_.*[1,2,3,4,5,6,7,8,9,0]")
tx2gene<-data.frame(names,names3)
txi2<-tximport(files, type="salmon", tx2gene = tx2gene)
nrow(txi2$counts)
```

#Filtering the count table by expression
```{r}
```
```

```
#Preparation of DESeq2 object
Samples<-as.data.frame(Samples)
rownames(Samples)<-colnames(txi2$counts)
Samples$Group<-factor(Samples$Group, levels = c("Cont","As","PsI"))
Samples$Date_sampling<-as.factor(Samples$Date_sampling)
dds<-DESeqDataSetFromTximport(txi2, Samples, ~Reproduction + Group)

#Applaying expression filter, count higher or equal to 10 in more then 5
samples
keep <- rowSums(counts(dds)>=10)>=5
dds<-dds[keep,]
nrow(dds)
```

#Data transformations figure 15a
```{r}

vsd <- vst(dds, blind = FALSE)
rld <- rlog(dds, blind = FALSE)
dds2 <- estimateSizeFactors(dds)
dfvsd<-as_data_frame(assay(vsd)[, 1:2]) %>% mutate(transformation = "vst")
dfrld<- as_data_frame(assay(rld)[, 1:2]) %>% mutate(transformation = "rlog")
dflog2<-as_data_frame(log2(counts(dds2, normalized=TRUE)[, 1:2]+1))
%>%mutate(transformation = "log2(x + 1)")
colnames(dflog2)[1:2] <- c("samples 1", "samples 2")
df<-rbind(dfvsd,dfrld,dflog2)
lvls <- c("log2(x + 1)", "vst", "rlog")
colnames(df)[1:2] <- c("x", "y")
df$transformation <- factor(df$transformation, levels=lvls)
levels(df$transformation)
#Creating and saving the plot
svg("Count_transformations_09112020.svg")
ggplot(df, aes(x = x, y = y)) + geom_hex(bins = 80) + coord_fixed() +
facet_grid( . ~ transformation)
dev.off()
```
Exploratory analysis of the data
```{r}
#Calculation of between sample Euclidean distance Figure 15b

Edist <- dist(t(assay(rld)))
Matrix <- as.matrix( Edist )
rownames(Matrix) <- paste(rld$Sample_name, sep="-" )
colnames(Matrix) <- NULL

colors <- colorRampPalette(rev(brewer.pal(9, "Blues")))(255)
rowanno<-data.frame("Group"=rld$Group,"S.point"=rld$Date_sampling)
rownames(rowanno)<-rownames(Matrix)
svg("Heatmap.svg", height = 5, width = 8)
pheatmap(Matrix,clustering_distance_rows =
"euclidean",clustering_distance_cols = "euclidean",treeheight_col=0,
annotation_row=rowanno)
dev.off()


```
#Principal component analysis Figure 15c

```{r}
#With batch effects Figure 15a
pcaData <- plotPCA(rld, intgroup = c( "Date_sampling", "Group"), returnData =
TRUE)
percentVar <- round(100 * attr(pcaData, "percentVar"))

pdf("PCA.pdf", height = 4, width=7)
row.names(pcaData)<-c(1:10,21:30, 41:50)
ggplot(pcaData, aes(x = PC1, y = PC2, color = Group, shape = Date_sampling,
label=row.names(pcaData))) +
  geom_point(size =2) +
  xlab(paste0("PC1: ", percentVar[1], "% variance")) +
  ylab(paste0("PC2: ", percentVar[2], "% variance")) +
  coord_fixed()+theme_light()+geom_text(size=2.2, position
=position_dodge(width = 0.5), vjust = -1 )+ expand_limits(y = c(-23, 12))
dev.off()

```
#Accounting for the reproduction batch effect in the dataset
```{r}
#Removing batch effects
mat1 <- assay(rld)
mat1 <- limma::removeBatchEffect(mat1, rld$Reproduction)
assay(rld) <- mat1
mat2 <- assay(vsd)
mat2 <- limma::removeBatchEffect(mat2, vsd$Reproduction)
assay(vsd) <- mat2
#Data transformations figure 16a
dfvsd<-as_data_frame(assay(vsd)[, 1:2]) %>% mutate(transformation = "vst")
dfrld<- as_data_frame(assay(rld)[, 1:2]) %>% mutate(transformation = "rlog")
df<-rbind(dfvsd,dfrld)
lvls <- c("vst", "rlog")
colnames(df)[1:2] <- c("x", "y")
df$transformation <- factor(df$transformation, levels=lvls)
levels(df$transformation)
#Creating and saving the plot
svg("Count_transformations_21112020_nobatch.svg")
ggplot(df, aes(x = x, y = y)) + geom_hex(bins = 80) + coord_fixed() +
facet_grid( . ~ transformation)
dev.off()

#Calculation of between sample Euclidean distance Figure 16b
Edist <- dist(t(assay(rld)))
Matrix <- as.matrix( Edist )
rownames(Matrix) <- paste(rld$Sample_name, sep="-" )
colnames(Matrix) <- NULL

colors <- colorRampPalette(rev(brewer.pal(9, "Blues")))(255)
rowanno<-data.frame("Group"=rld$Group,"S.point"=rld$Date_sampling)
rownames(rowanno)<-rownames(Matrix)
svg("Heatmap2.svg", height = 5, width = 8)
pheatmap(Matrix,clustering_distance_rows =
"euclidean",clustering_distance_cols = "euclidean",treeheight_col=0,
annotation_row=rowanno)
dev.off()
#Principal component analysis Figure 16c
```

```
pcaData <- plotPCA(rld, intgroup = c( "Date_sampling", "Group"), returnData =
TRUE)
percentVar <- round(100 * attr(pcaData, "percentVar"))
svg("PCA2.svg", height = 4, width=9)
row.names(pcaData)<-c(1:10,21:30, 41:50)
ggplot(pcaData, aes(x = PC1, y = PC2, color = Group, shape = Date_sampling,
label=row.names(pcaData))) +
  geom_point(size =2) +
  xlab(paste0("PC1: ", percentVar[1], "% variance")) +
  ylab(paste0("PC2: ", percentVar[2], "% variance")) +
  coord_fixed()+theme_light()+geom_text(size=2.2, position
=position_dodge(width = 0.5), vjust = -1 )+ expand_limits(x = c(-20, 40))
dev.off()
```
#Differential gene expression analysis- sampling date 1
```{r}
#Importing data for the analysis
Sampling_date1500 <- read_excel("Sampling_date1500.xlsx")
filesSD1<- file.path(dir, "500", Sampling_date1500$Sample_name,"quant.sf")
names(filesSD1)<-paste("samples",1:15)
all(file.exists(filesSD1))
txiSD1<-tximport(filesSD1, type="salmon", txOut = T)
#Convert Trinity transcripts to Trinity genes
names<-rownames(txiSD1$counts)
names2<-str_extract(names, ".*_.*_.*_.*[1,2,3,4,5,6,7,8,9,0]_")
names3<-str_extract(names2, ".*_.*_.*_.*[1,2,3,4,5,6,7,8,9,0]")
tx2gene<-data.frame(names,names3)
txiSD1genes<-tximport(filesSD1, type="salmon", tx2gene = tx2gene)
nrow(txiSD1genes$counts)

#Filtring by expression
#preparation of DESeq2 object
SamplesSD1<-as.data.frame(Sampling_date1500)
rownames(SamplesSD1)<-colnames(txiSD1genes$counts)
SamplesSD1$Group<-factor(SamplesSD1$Group, levels = c("Cont","As","PsI"))
ddsSD1<-DESeqDataSetFromTximport(txiSD1genes, SamplesSD1, ~Reproduction +
Group)
#Applaying expression filter
keepSD1 <- rowSums(counts(ddsSD1)>=10)>=5
ddsSD1<-ddsSD1[keepSD1,]
nrow(ddsSD1)
```
#Differential gene expression analysis
```{r}
ddsSD1<-DESeq(ddsSD1)
resSD1<-results(ddsSD1)
resultsNames(ddsSD1)
resultsSD1PsI <- as.data.frame(results(ddsSD1, name="Group_PsI_vs_Cont")) %>%
  rownames_to_column("GeneID")
resultsSD1As <- as.data.frame(results(ddsSD1, name="Group_As_vs_Cont")) %>%
  rownames_to_column("GeneID")
#exporting the list of DE genes for further analysis
arrange(resultsSD1PsI, padj)%>% filter(padj <= 0.05)-> PSISD1 %>%
  write.csv("PsIvsCont_DESeqSD1_DE2.csv")
arrange(resultsSD1As, padj)%>% filter(padj <= 0.05)%>%
  write.csv("AsvsCont_DESeqSD1_DE2.csv")
```

```
#Enhanced vulcano plots
resultsNames(ddsSD1)
resSHSD1As <- lfcShrink(ddsSD1, coef="Group_As_vs_Cont", type="apeglm")
resSHSD1PsI <- lfcShrink(ddsSD1, coef="Group_PsI_vs_Cont", type="apeglm")
svg("EnhancedVolcanoSD1As2.svg")
print(EnhancedVolcano(resSHSD1As,lab=rownames(resSHSD1As),x="log2FoldChange",
y="padj",pCutoff = 0.05)+
    ggplot2::coord_cartesian(xlim=c(-2, 9)) +
    ggplot2::scale_x_continuous(
      breaks=seq(-2,9, 1)))
dev.off()
svg("EnhancedVolcanoSD1PsI2.svg")
print(EnhancedVolcano(resSHSD1PsI,lab=rownames(resSHSD1PsI),x="log2FoldChange
",y="padj",pCutoff = 0.05)+
    ggplot2::coord_cartesian(xlim=c(-3, 6)) +
    ggplot2::scale_x_continuous(
      breaks=seq(-3,6, 1)))
dev.off()
```

#Differential gene expression analysis- sampling date 2

```{r}
#importing data for the analysis
Sampling_date2500 <- read_excel("Sampling_date2500.xlsx")
filesSD2<- file.path(dir, "500", Sampling_date2500$Sample_name,"quant.sf")
names(filesSD2)<-paste("samples",1:15)
all(file.exists(filesSD2))
txiSD2<-tximport(filesSD2, type="salmon", txOut = T)
#Convert Trinity transcripts to Trinity genes
names<-rownames(txiSD2$counts)
names2<-str_extract(names, ".*_.*_.*_.*[1,2,3,4,5,6,7,8,9,0]_")
names3<-str_extract(names2, ".*_.*_.*_.*[1,2,3,4,5,6,7,8,9,0]")
tx2gene<-data.frame(names,names3)
txiSD2genes<-tximport(filesSD2, type="salmon", tx2gene = tx2gene)
nrow(txiSD2genes$counts)
#Filtring by expression
#preparation of DESeq2 object
SamplesSD2<-as.data.frame(Sampling_date2500)
rownames(SamplesSD2)<-colnames(txiSD2genes$counts)
SamplesSD2$Group<-factor(SamplesSD2$Group, levels = c("Cont","As","PsI"))
ddsSD2<-DESeqDataSetFromTximport(txiSD2genes, SamplesSD2, ~Group)
#applaying expression filter
keepSD2 <- rowSums(counts(ddsSD2)>=10)>=5
ddsSD2<-ddsSD2[keepSD2,]
nrow(ddsSD2)
```

#Differential gene expression

```{r}
ddsSD2<-DESeq(ddsSD2)
resSD2<-results(ddsSD2)
resultsSD2PsI <- as.data.frame(results(ddsSD2, name="Group_PsI_vs_Cont")) %>%
  rownames_to_column("GeneID")
resultsSD2As <- as.data.frame(results(ddsSD2, name="Group_As_vs_Cont")) %>%
  rownames_to_column("GeneID")
#exporting the list of DE genes for further analysis
arrange(resultsSD2PsI, padj)%>% filter(padj <= 0.05)%>%
  write.csv("PsIvsCont_DESeqSD2_DE.csv")
```

```
arrange(resultsSD2As, padj)%>% filter(padj <= 0.05)%>%
  write.csv("AsvsCont_DESeqSD2_DE.csv")
#Enhanced vulcano plots
resultsNames(ddsSD2)
resSHSD2As <- lfcShrink(ddsSD2, coef="Group_As_vs_Cont", type="apeglm")
resSHSD2PsI <- lfcShrink(ddsSD2, coef="Group_PsI_vs_Cont", type="apeglm")

svg("EnhancedVolcanoSD2As.svg")
print(EnhancedVolcano(resSHSD2As,lab=rownames(resSHSD2As),x="log2FoldChange",
y="padj",pCutoff = 0.05)+
    ggplot2::coord_cartesian(xlim=c(-2.5, 5)) +
    ggplot2::scale_x_continuous(
      breaks=seq(-2.5,5, 1)))
dev.off()

svg("EnhancedVolcanoSD2PsI.svg")
print(EnhancedVolcano(resSHSD2PsI,lab=rownames(resSHSD2PsI),x="log2FoldChange
",y="padj",pCutoff = 0.05)+
    ggplot2::coord_cartesian(xlim=c(-3, 6)) +
    ggplot2::scale_x_continuous(
      breaks=seq(-3,6, 1)))
dev.off()
```

Venn diagrams

```{r}
#Venn diagram SD 1
venn_data1 <- data.frame( PsI1= resultsSD1PsI$padj<=0.05,
                          As1 = resultsSD1As$padj <= 0.05)
venn_data1[is.na(venn_data1)]<-FALSE
svg("VennDiagramSD1.svg")
VDSD1<-vennDiagram(venn_data1,circle.col=c("red", "blue"))

print(VDSD1)
dev.off()

#Venn diagram SD2
venn_data2 <- data.frame( PsI2= resultsSD2PsI$padj<=0.05,
                          As2 = resultsSD2As$padj <= 0.05)
venn_data2[is.na(venn_data2)]<-FALSE
svg("VennDiagramSD2.svg")
VDSD2<-vennDiagram(venn_data2,circle.col=c("red", "blue"))
print(VDSD2)
dev.off()

#Venn diagram_overlap all
venn_data3 <- cbind(venn_data2,T_ID=resultsSD2As$GeneID)
venn_data4 <- cbind(venn_data1,T_ID=resultsSD1As$GeneID)
venn_data5 <- merge(venn_data3,venn_data4, all=T)
venn_data5[is.na(venn_data5)] <- FALSE
VD5<- venn_data5[,2:5]
svg("VenndiagramSD1SD2-5.svg")
VDSD3<-vennDiagram(VD5,circle.col=c("red", "blue", "green","yellow"))
print(VDSD3)
dev.off()
```

```
#Parse DE genes with their annotation
```{r}
Transcript_annotation <-
read_csv("F:/Marble/Pvir_annotation/NO_mtDNA_500.fasta.dammit/Transcript_anno
tation.csv")
#Name_map <-
read_csv("F:/Marble/Pvir_annotation/NO_mtDNA_500.fasta.dammit/NO_mtDNA_500.fa
sta.dammit.namemap.csv")
Name_map$seqid<-Name_map$q_name
Name_map$q_name<-NULL
Name_map
Annotation_Procambarus<-
merge(Name_map,Transcript_annotation,"seqid",all=TRUE)
names<-Annotation_Procambarus$original
names2<-str_extract(names, ".*_.*_.*_.*_.*[1,2,3,4,5,6,7,8,9,0] 1")
names3<-str_extract(names2, ".*_.*_.*_.*_.*[1,2,3,4,5,6,7,8,9,0]")
Annotation_Procambarus$Transcript_name<-names3
names2<-str_extract(names, ".*_.*_.*_.*[1,2,3,4,5,6,7,8,9,0]_")
names3<-str_extract(names2, ".*_.*_.*_.*[1,2,3,4,5,6,7,8,9,0]")
Annotation_Procambarus$Gene_name<-names3
write.csv(Annotation_Procambarus,"Annotation_procambarus.csv")
Annotation_Procambarus$GeneID<-Annotation_Procambarus$Gene_name
#create DE gene objects
arrange(resultsSD1PsI, padj)%>% filter(padj <= 0.05)-> PsISD1
arrange(resultsSD2PsI, padj)%>% filter(padj <= 0.05)-> PsISD2
arrange(resultsSD1As, padj)%>% filter(padj <= 0.05)-> AsSD1
arrange(resultsSD2As, padj)%>% filter(padj <= 0.05)-> AsSD2

#merge annotation with DE gene files
DEPsISD1Annotated<-merge(PsISD1,Annotation_Procambarus,"GeneID")
write.csv(DEPsISD1Annotated,"DEPsISD1Annotated.csv")
PsISD1$TranscriptID
DEAsSD1Annotated<-merge(AsSD1,Annotation_Procambarus,"GeneID")
write.csv(DEAsSD1Annotated,"DEAsSD1Annotated.csv")
DEPsISD2Annotated<-merge(PsISD2,Annotation_Procambarus,"GeneID")
write.csv(DEPsISD2Annotated,"DEPsISD2Annotated.csv")
PsISD1$TranscriptID
DEAsSD2Annotated<-merge(AsSD2,Annotation_Procambarus,"GeneID")
write.csv(DEAsSD2Annotated,"DEAsSD2Annotated.csv")
```
#Over- representation analysis KEGG
```{r}
#importing cleaned dataset from KEGG mapper
KEGGmapping_cleaned <-
read_delim("F:/Marble/Pvir_annotation/KEGG/KEGGmapping_cleaned.txt",
+     "\t", escape_double = FALSE, trim_ws = TRUE)

Top20<-KEGGmapping_cleaned[1:20,]
svg("KEGG_representatin.svg")
ggplot(Top20,aes(x=reorder(Pathway,Count), y=Count))+geom_bar(stat =
"identity")+ coord_flip(ylim=c(0, 900))+theme_classic()+
  geom_text(aes(label=Count), hjust=-0.3, size=3.5)+ylab("Number of
transcripts")+xlab("KEGG pathway")
dev.off()
```
```

```
#Over- representation analysis
```{r}
#Kegg annotation_K_identifiers
KEGG_anno <-
read_csv("F:/Marble/Pvir_annotation/KEGG/Procambarus_virginalis_KEGG_forR.csv
")
K_DEAsSD1<- merge(KEGG_anno,DEAsSD1Annotated,"seqid",all=F)
#no gene annotated with KEGG
K_DEAsSD2<- merge(KEGG_anno,DEAsSD2Annotated,"seqid",all=F)
write.csv(K_DEAsSD2,"KEGG_DEAsSD2.csv")
K_DEPSISD1<- merge(KEGG_anno,DEPsISD1Annotated,"seqid",all=F)
write.csv(K_DEPSISD1,"KEGG_DEPSISD1.csv")
K_DEPSISD2<- merge(KEGG_anno,DEPsISD2Annotated,"seqid",all=F)
write.csv(K_DEPSISD2,"KEGG_DEPSISD2.csv")

geneListSD2Psi<-K_DEPSISD2$lfcSE
names(geneListSD2Psi)<-as.character(K_DEPSISD2$KEGGid)
geneListSD2Psi<-sort(geneListSD2Psi,decreasing = T)

kkSD2_PSI<-enrichKEGG(gene=unique(KEGG_DEPSISD2$KEGGid),
organism="ko",pvalueCutoff = 0.05)
head(kkSD2_PSI, n=100)
svg("Kegg_enriched pathwaysSD2_PSI.svg")
barplot(kkSD2_PSI)
dev.off()
svg("Kegg_enriched pathwaysSD2_AS.svg")
barplot(kkSD2_As)
dev.off()
kkSD2_As<-enrichKEGG(gene=unique(KEGG_DEAsSD2$KEGGid),
organism="ko",pvalueCutoff = 0.05)
head(kkSD2_As, n=100)
kkSD2_PSISD1<-enrichKEGG(gene=unique(KEGG_DEPSISD1$KEGGid),
organism="ko",pvalueCutoff = 0.05)


```
Extract FASTA sequences of DE genes
```{r}
Fasta<-readDNAStringSet("NO_mtDNA_500.fasta")
names(Fasta)
Fasta[DEPsISD1Annotated$seqid] %>% writeXStringSet("DEPsISD1.fasta")
Fasta[DEPsISD2Annotated$seqid]%>% writeXStringSet("DEPsISD2.fasta")
Fasta[DEAsSD1Annotated$seqid]%>% writeXStringSet("DEAsSD1.fasta")
Fasta[DEAsSD2Annotated$seqid]%>% writeXStringSet("DEAsSD2.fasta")
```
```

CURRICULUM VITAE

Ljudevit Luka Boštjančić ljudevit.luka@gmail.com

Born: 1st of October 1996 in Zagreb, Croatia

Education

2015.-2018. Bachelor of Science in Molecular Biology, Faculty of Science, Department of Biology, Zagreb

2018.-2021. University Graduate Programme in Molecular Biology, Faculty of Science, University of Zagreb (Croatia)

Work and research experience

2017.-2021. Intern at the Laboratory for molecular analyses, Division of Zoology, Department of Biology, Faculty of Science, University of Zagreb, under the supervision of prof. Ivana Maguire, PhD

2020. Erasmus+ student internship at the Institute for Environmental Sciences, University of Koblenz-Landau, under the supervision of Kathrin Theissinger, PhD

Conferences

2018. 13th Croatian biological congress with international participation in Poreč, Croatia, poster presentation

2019. 5th Students' Symposium in Biology and Life Sciences - SISB5, Zagreb, oral presentation

2019. Regional European IAA meeting “IAA Gotland 2019 – Crayfish and Aquaculture”- poster presentation

2019. International medical conference ZIMS, Zagreb, Croatia Workshop leader: Mutations and disease, a bioinformatics approach Tracking down Alzheimer’s

Awards

2019. Rector's award for individual scientific work (one/two authors) in natural sciences

2019.-2020. City of Zagreb Scholarship for excellence

Projects

2019.-2020. Project leader of the student project founded by the University of Zagreb Student union grant, “Development of new karyotypization techniques of freshwater crayfish (family: Astacidae)”, KarioInAstacidae.

Publications

Lovrenčić, L., Bonassin, L., **Boštjančić, L.L.**, Podnar, M., Jelić, M., Klobučar, G., Jaklič, M., Slavevska-Stamenković, V., Hinić, J., Maguire., I., 2020 New insights into the genetic diversity of the stone crayfish: taxonomic and conservation implications. BMC Evol. Biol. **20**, 146

Boštjančić, L.L., Bonassin, L., Anušić L., Lovrenčić, L., Besendorfer, V., Maguire, I., Grandjean, F., Austin, C.M., Greve, C., Hamadou A.B., Mlinarec, J., 2020, The *Pontastacus leptodactylus* (Astacidae) Repeatome Provides Insight Into Genome Evolution and Reveals a Remarkable Diversity of Satellite DNA. Front. Genet. **11**, 611745.

Skills

Languages Croatian (native proficiency), English (full professional proficiency)

Software R, Linux