The effects of a wastewater effluent on population, transcriptomic and metabolomic markers in the freshwater amphipod *Gammarus fossarum*

Domenico Roberto Caputo

The thesis is submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the University of Portsmouth.

May 2020

Declaration

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

Signed:  (Candidate) Date: 14/12/2020

Word Count: 66940

"As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns — the ones we don't know we don't know. And if one looks throughout the history, it is the latter category that tend to be the difficult ones."

Donald Rumsfeld, Pentagon News Briefing (February 2002)

Acknowledgements

This three-year PhD studentship was financially supported by a bursary awarded by the Swiss Federal Institute of Aquatic Science and Technology (EAWAG) (Dübendorf, Switzerland) and the University of Portsmouth (Portsmouth, UK). The metabolomic analysis was financed by the Natural Environmental Research Council (NERC UK) and conducted at University of Birmingham.

I wish to express my deepest gratitude to my supervisors Prof Alex Ford, Dr Inge Werner and Dr Samuel Robson. They gave me constant scientific and moral support throughout my PhD. Many thanks for equipping me with bioinformatics and environmental skills, and more importantly, the right scientific and logic morality. Without their guide, the goals of this project would not have been realised.

I am grateful to Dr Melanie Fischer (aQuaTox - Solutions, Dübendorf, Switzerland), Dr Andrea Schifferli and Dr Thomas Bucher (Swiss Centre for Applied Ecotoxicology, Dübendorf, Switzerland) for their valuable technical support in the sampling and RNA extractions. Thanks for the experimental support provided by Prof Darek Gorecki (University of Portsmouth) and Dr Robin Rumney (University of Portsmouth) in the qPCR analysis. I am also incredibly grateful to Prof Mark Viant and Dr Dorsa Varshavi (University of Birmingham) for providing protocol and reagents for the amphipod metabolic extractions as well as for performing the mass spectrometry and metabolomic data analyses.

I wish to show my gratitude to University of Portsmouth, Swiss Federal Institute of Aquatic Science and Technology (EAWAG) and University of Birmingham for offering office, instruments and scientific facilities for this project. Finally, I want to thank my mother, my father and my brother for their strong support throughout my PhD. They kept me going on through the hardest times, since the first day I moved in the UK for my PhD.

Abstract

In order to prevent further deterioration of river quality in response to anthropogenic contaminants released into freshwaters through wastewater treatment plants (WWTPs), European countries, including Switzerland, financed important river monitoring plans. In addition, measures to improve the sewage treatment technologies were also proposed. Due to a central role in aquatic food web and a marked sensitivity to xenobiotics, the amphipod *Gammarus fossarum* has been defined as an ideal biomarker species for ecotoxicological risk assessment. However, its genome has not yet been completely annotated and attempts to monitor dysfunctions in invertebrates using biomarkers in fish species have produced inconsistent results. The present project was aimed to investigate the impact of a contaminant mixture released by a Swiss sewage effluent on the amphipod species *Gammarus fossarum.* In addition, the transcriptomic and metabolomic differences between male and female amphipods were also explored, in order to increase the knowledge on crustacean sexual biology. Firstly, the concentrations of 55 xenobiotics (pesticides and pharmaceuticals) commonly detected in river waters were measured up- and downstream of the WWTP, in both water samples and amphipods. An evaluation of the differences in the population structure between amphipods sampled above and below the effluent was conducted annually between September 2017 and 2018. A high-throughput sequencing of total RNA from *G. fossarum* was performed employing an Illumina

HiSeq 2500 sequencing platform. The complete transcriptome of *G. fossarum* was assembled and annotated *de novo* and the changes in gene expression between *G. fossarum* sampled up- and downstream of the WWTP were investigated. In addition, the differentially expressed genes between male and female amphipods were also explored. Finally, an Ultra Performance Liquid Chromatography - Mass Spectrometry (UPLC-MS) platform was employed to investigate the metabolomic fingerprints in male and female *G. fossarum* sampled up- and downstream of the effluent. The differential "omics" analyses showed variations in general stress biomarkers, primary metabolism and mitochondrial metabolism in amphipods sampled at the downstream site. However, the toxic pressure did not cause observable abnormalities in amphipod population structures. The comparative analyses of male and female transcriptome and metabolome between animals sampled above and below the effluent showed that the genders may respond differently to anthropogenic pollutants in aquatic environments. Terms related to heart and circulatory processes, muscle system and cell differentiation were found when conducting a gene ontology (GO) analysis on the differentially expressed genes between males and females, suggesting that the sex distinction traits in hormonal system may act on a wide spectrum of molecular networks. Given the lack of molecular information on amphipod species, the data set collected in this project will be useful in future studies to develop new ecotoxicological biomarkers in amphipods.

Table of Contents

List of Figures

List of Tables

Dissemination

The complete transcriptome of *Gammarus fossarum* was deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject accession code PRJNA556212.

The differences in population structure, transcriptome and metabolome between *Gammarus fossarum* amphipods sampled upstream and downstream of the sewage effluent were poster presented at SETAC (Society of Environmental Toxicology and Chemistry) Europe 28th Annual Meeting, Rome, May 13-17th, 2018 and SETAC Europe 29th Annual Meeting, Helsinki, May 26-30th, 2019.

Abbreviations

ACN – Acetonitrile

ANCOVA – Analysis of Covariance

ANOVA – Analysis of Variance

BAF – Bioaccumulation Factor

BLAST – Basic Local Alignment Search Tool

Ct – Threshold Cycle

ddNTP – Dideoxynucleotide Triphosphate

DE – Differential Expression

DGE – Differential Gene Expression

dUTP – Deoxyuracile Triphosphate

EDC – Endocrine Disruptor Chemical

ESI – Electrospray Ionization

FDR – False Discovery Rate

FPKM – Fragments Per Kilobase Million

GC-MS – Gas Chromatography Mass Spectrometry

gDNA – Genomic DNA

GO – Gene Ontology

HSP – Heat Shock Protein

JH – Juvenile Hormone

LC-MS – Liquid Chromatography Mass Spectrometry

LOQ – Limit of Quantitation

NGS – Next-Generation Sequencing

ORF – Open-Reading Frame

PCA – Principal Component Analysis

PCB – Polychlorinated Biphenyls

PCR – Polymerase Chain Reaction

Phred – Phil's Read Editor

PPDB – Pesticide Properties Database

qPCR – Quantitative Polymerase Chain Reaction

SSRI – Selective Serotonin Reuptake Inhibitor

TMS – Trimethylsilyl

TOF – Time of Flight

TU – Toxic Unit

UPLC – Ultra Performance Liquid Chromatography

WWTP – Wastewater Treatment Plant

Chapter 1 – General Introduction

1.1 The importance and status of rivers

Despite their ecological relevance, rivers are among the most impacted ecosystems through water extraction, damming, channel modification, invasive species, pollution and constant climate changes (Gore, 1985; Howarth et al., 2000; Marzin et al., 2012; Gallardo et al., 2016). Rivers are vital for human existence as sources of transportation, irrigation, food, leisure, waste disposal, spiritual inspiration and biodiversity source (Adeloye, 2009). It has been shown that river waters host ~6% of all described species including 33% of all vertebrates and define some of the most biodiverse areas on the planet (Dudgeon et al., 2006). On the other hand, an interesting study by Schinegger et al., (2012) on the European rivers status concluded that 47% of sites are heavily impacted by anthropogenic degradation. In fact, one of the major challenges of the European Union's Water Framework Directive (WFD) over the last decade has been to find common approaches for defining reference conditions and to describe the level of anthropogenic intervention allowed in reference sites (Pardo et al., 2012). Although hydromorphological and physical-chemical parameters are measured to define the conditions of rivers and to set threshold values (King et al., 2011), assessment of water quality in rivers during the 20th century focused on establishing links between pollutants and biota, generally reinforcing the theory that stressors reduced biological diversity (i.e., biodiversity indices) (Hynes, 1994).

Despite the number of the major European water initiatives, to date just ~50% of freshwater bodies are regarded to be in good ecological condition, and in some countries, such as Germany and the Netherlands the value drops to less than 10% (EAA, 2015). River restoration projects are being financed by governments all over Europe and made a legal obligation in several countries (EU WFD, 2000). For instance, in order to prevent further deterioration of river quality in response to an increasing number of anthropogenic chemicals released into freshwaters, Switzerland invested in a massive program for upgrading wastewater treatment plants (WWTPs) (Eggen et al., 2014). With the adaptation of the water protection act in March 2014, the Swiss government decided to monitor the overall condition of the rivers, with the plan of reducing xenobiotic releases starting from densely populated regions, where wastewater from WWTP discharges constitutes an important impact on water quality (Eggen et al., 2014).

1.2 River pollution

Water pollution includes a large number of stressors, such as thermal, biological and chemical contamination (Warren, 1971). However, the chemical stress is particularly concerning in Europe where chemical toxicity coming from anthropogenic activities represents an ecological threat to almost half of all European bodies of water (Malaj et al., 2014). It has been shown that more than 100 000 compounds in commerce are registered in Europe, many of which get transported into water bodies at some stage in their lifecycle (Eggen et al., 2014).

Domestic, industrial and agricultural wastes are among the main sources of pollution in river waters (Voulvoulis et al., 2016). Clearly, the overall chemical composition, the concentrations of the single substances and the potential biological phenomena of compensation and adaptation are essential in evaluating the state of aquatic environments. For this reason, understanding acute and chronical effects of contaminants in freshwater environments is extremely important to clarifying and mitigating the impact of the pollution (Stendera et al., 2012). Unsurprisingly, controlling the chemical wastes released through WWTPs has been one of the main priorities of European water pollution legislation for over twenty years (Kallis et al., 2001) and treating wastewater is considered to be one of the most important forms of water pollution control (Viessman et al., 2009).

1.3 Overview on WWTPs

The sewage treatment is performed by filtering domestic, industrial and agricultural waste waters through the WWTPs (Fig. 1.1&1.2). The research efforts to develop increasingly effective and environmentally balanced wastewater treatment technologies are currently very intense (Guerrero et al., 2011; Eggen et al., 2014; Weber, 2016; Ghernaout et al., 2018). In general, the process to treat wastewater normally involves 3 steps (Defra, 2012; Ahmed et al., 2017):

1) retention of waste in static basins where solids are removed by settlement and scum and lipids are skimmed from the top

2) biological treatment employing processes such as trickle beds or activated sludge. This stage is where most chemical removal occurs
3) treated waters undergo a physical or chemical filtration, such as ozonation but is usually employed where the discharge is going into a “sensitive area” or if direct reuse as a potable supply is intended (Defra, 2012).

The duration and the technology behind each of these treatments define the amount of pollutants that reaches the environment (Defra, 2012).

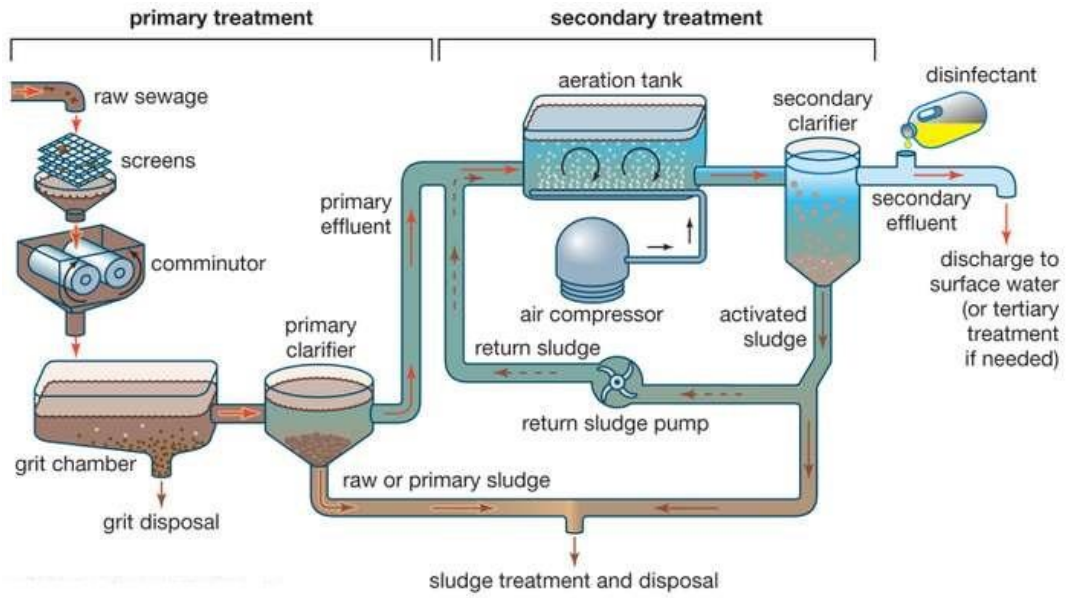

Fig. 1.1: Primary and secondary treatment of sewage, using activated sludge process (EB, 2012).

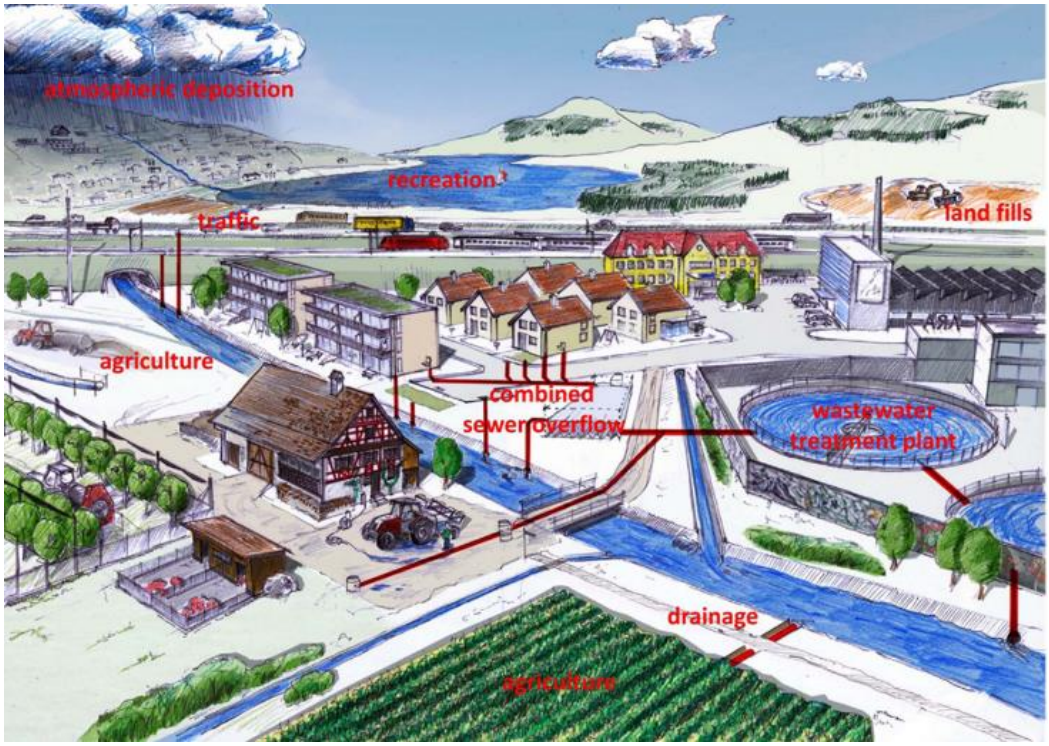

Fig. 1.2: Schematic presentation of diffuse and point source entry paths of pollutants in the environment. *Reproduced with permission from Eggen et al., 2014.*

Although there are differences in the technology used for wastewater treatments and in the level of treatment achieved in different countries, WWTP research programs have common objectives (Eggen et al., 2014) (Fig. 1.3):

- to improve hygienic conditions of receiving waters by functioning as a barrier for faecal bacteria and pathogens
- to improve the water quality removing xenobiotics
- to remove the nutrients nitrogen and phosphorus that are responsible for the eutrophication of aquatic ecosystems.

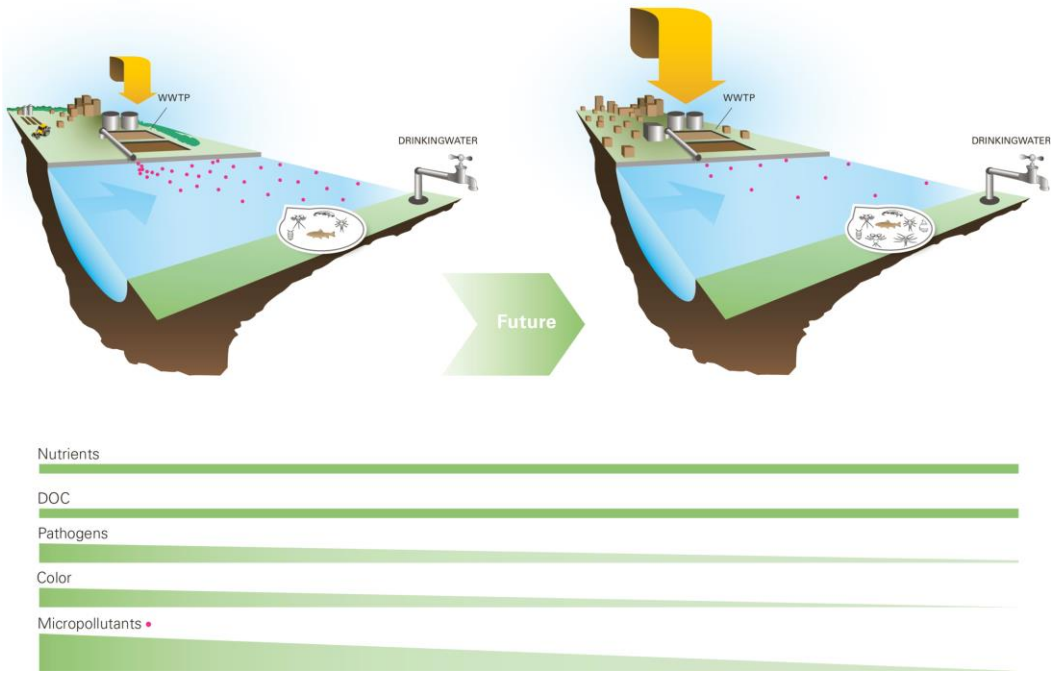

Fig. 1.3: Schematic diagrams of current and future demands and outcomes on wastewater treatment. The left panel shows the current situation in which the loading of degradable organics (DOC), pathogens, nutrients and some micropollutants (MPs) are reduced. The right panel shows

the anticipated future situation with an increased amount of wastewater but with additional treatments. *Reproduced with permission from Eggen et al., 2014.*

1.4 Main effects of wastewater effluents on fish

A large number of studies have focused on individual emerging chemicals and their effects on fish biology considering both physiological (Bonga, 1997; Kakuta et al., 1997; Tetreault et al., 2011; Cazenave et al., 2014) and molecular (Ings et al., 2011; Bahamonde et al., 2014; Al-Salhi et al., 2012; Cavallin et al., 2016) parameters. For instance, the presence of endocrine disruptor compounds (EDCs) in river waters has been of primary concern since the 1970s, when alterations in the population structure of several fish species and the formation of intersex individuals were observed (Crawford et al., 2017). The impact of these substances on aquatic ecosystems has been evaluated using a variety of approaches (Tetreault et al., 2011, Schneider et al., 2015; Trapp et al., 2015). It has been shown that the estrogenic chemical mixtures released by a Canadian WWTP severely impair the production of testosterone and 11-ketotestosterone in greenside darters (*Etheostoma blennioides*) and rainbow darters (*Etheostoma caeruleum*) sampled downstream of the sewage effluent compared to fish sampled upstream (Tetreault et al., 2011). The authors of this study found that these hormonal perturbations lead to variations in the reproductive system of male darters and ultimately to the formation of intersex individuals. Even more severe responses in *Prochilodus lineatus* exposed to an Italian wastewater effluent were found by

Cazenave et al., (2014). In particular, the authors observed an increase in mortality, monocytosis, transaminase activity, antioxidant enzyme activation, lipid oxidative damage in several tissues, and hepatic and muscle glycogen depletion in fish caged at the downstream site of the effluent compared to fish caged upstream, used as reference site. A transcriptomic analysis performed by Ings et al., (2011) on caged rainbow trout (*Oncorhynchus mykiss*) exposed to a whole municipal wastewater effluent highlighted altered expression of the genes encoding the heat shock proteins of 70 and 90 kDa and the enzyme of the cytochrome p450 system *CYP1A1*, indicating a general stress response of the animals as well as an enhanced energy demand in the exposed fish. Using a metabolomic platform, Al-Salhi et al., (2012) found that juvenile rainbow trout living in a British effluent containing high concentrations of domestic wastes accumulated surfactants, naphthols, chlorinated xylenols, phenoxyphenols, chlorophenes, resin acids, mefenamic acid, oxybenzone and steroidal alkaloids in the bile or plasma. As a result of the accumulation of these substances, variations in the plasma concentrations of bile acids and lipids were found, indicating histological perturbations.

In the light of the mentioned examples of toxicological studies on fish, it appears clear that the chemical mixtures released from the sewage effluents can strongly impact the biology of fish species. The effects depend on the composition of the chemical mixtures, the exposure time and the combination of the chemical stress

with others abiotic stressors, such as temperature (Madeira et al., 2013) and parasite infections (Schwaiger, 2011). On top of that, it has been recently demonstrated that the effects of the exposure to the serotonin reuptake inhibitor fluoxetine in adult zebrafish (*Danio rerio)* (F_0) persist for three consecutive generations in the unexposed descendants (F_1 to F_3) (Vera-Chang et al., 2018). Therefore, the possibility that the molecular perturbations following chronical exposure to xenobiotic mixtures could be transferred cross-generationally became of primary concern.

1.5 Amphipods as model species

While the biological effects of single xenobiotics and whole effluents have been mostly focused on fish species (e.g., Ings et al., 2011; Tetreault et al., 2011; Al-Salhi et al., 2012; Cazenave et al., 2014) the use of invertebrates as ecological biomarkers increased over the last few years (e.g., Bossus et al., 2014; Peschke et al., 2014; Lebrun et al., 2017; Gouevia et al., 2018). In general, when investigating the ecotoxicological impacts on the environment, it is essential to use an appropriate organism. It should have the following features:

- sensitive
- representative
- abundant
- subject to typical exposure
- having a critical ecological importance

Ideally, it should be practical to maintain and culture for protracted studies. Generally, amphipods (order of Crustacea) fulfil these requirements very well. They are key members of the aquatic community (Rainbow et al., 2011), perform essential roles in nutrient and energy flow (Graca, 2001), their abundance is an established measure of environmental quality (Gaufin et al., 1956; Malmqvist, 2002; Hutton et al., 2015) and any impact on their populations can have profound implications for the whole ecosystem (Hodkinson et al., 2005).

1.5.1 Taxonomy and habitat

The order Amphipoda is a taxon of malacostracan crustaceans with over 9,900 species described (Balian et al., 2008). They are primarily marine (and occasionally terrestrial), but around 20% live in freshwater (Väinölä et al., 2008). They are an essential component of the aquatic ecosystem (Lowry et al., 2001). Within the order, the most widespread and dominant group (of over 4500 species) is the sub-order Senticaudata (known as Gammaridea), commonly referred to as gammarids (Lowry et al., 2013). Gammarids are particularly important members of aquatic food webs since they are the main link between detritus and “higher” consumers such as fish (Forrow et al., 2000; Kunz et al., 2010). In fact, they have been described as keystone species within chalk streams (Woodward et al., 2008), with a potential to exert strong effects on the structure and processes of the aquatic ecosystem. They are common in fresh and marine environments throughout the world (Schirling et al., 2005; Adam et al., 2010) but rare in the tropics. Typically,

they occur in large numbers (Cold et al., 2004; Ladewig et al., 2006; Dixon et al., 2011), have a relatively short generation time and high reproductive rates (Peschke, 2011), display sexual dimorphism (Felten et al., 2008b), are easily maintained in the laboratory (McCahon et al., 1988a, 1988b) and are widely considered as particularly sensitive to contaminant exposure compared to other crustaceans (Maltby, 1995; Cold et al., 2004; Bloor et al., 2005; Bloor et al., 2006; Felten et al., 2008a; Geffard et al., 2010; Jacobson et al., 2010; Peschke, 2011). Therefore, it is unsurprising that species such as *Gammarus pulex* (Linnaeus 1758), *G. roeseli* (Gervais 1835) and *G. fossarum* (Koch 1835) are recognised as particularly relevant test species when investigating the environmental impact of toxicants in Swiss rivers (Ganser et al., 2018; Kienle et al., 2019). They have been used as a test species in a range of exposures including: nitrogenous compounds (Berenzen et al., 2001), pesticides (Adam et al., 2010), heavy metals (Dedourge-Geffard et al., 2009; Geffard et al., 2010), antibiotics (Bundschuh et al., 2009), herbicides (Bundschuh et al., 2013) and pharmaceuticals (De Lange et al., 2006a; Guler et al., 2010; Bossus et al., 2014) as well as in whole effluent tests (Bundschuh et al., 2011; Schneider et al., 2015; Wigh et al., 2016; Wigh et al., 2017).

1.5.2 Anatomy

G. fossarum amphipods are widespread in Central Europe, especially in the upstream reaches of streams (Westram et al., 2011). The maximum length of males is about 15 mm, while females may reach approximately 10 mm, though

generally they are smaller (Sutcliffe, 1992). The body is curved, laterally compressed and divided into 4 main parts: head, peron, pleon and urosome. The head has two pairs of antenna, complex mouthparts, and a pair of compound-eyes. The peron has 7 pairs of jointed legs classed as pereopods which are used for swimming, crawling and grasping. In mature males, the first two pereopods are enlarged, called gnathopods, and are used to grasp the female. In mature females, attached to pereopods 2-5 are the oosegites: paddle shaped structures that form a brood pouch or marsupium for holding embryos, which are retained by the female until hatching. In both sexes, each segment of the peron also contains a pair of gills. Posterior to the peron is the pleon which contains three pairs of appendages called pleopods, used for circulating water and swimming. The last section, the urosome, has two or three pairs of adapted pleopods called uropods, also used in locomotion (Fig. 1.4). The phenotypical differences between male and female individuals are shown in Fig. 1.5&1.6.

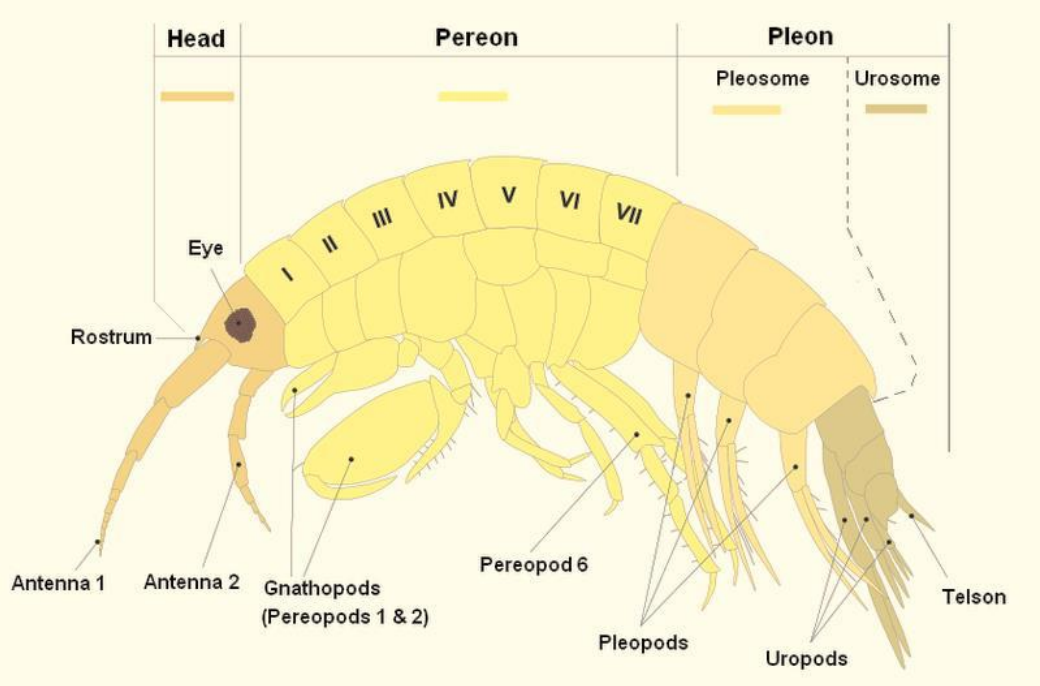

Fig. 1.4: Anatomy of a male amphipod (Lycaon, 2006).

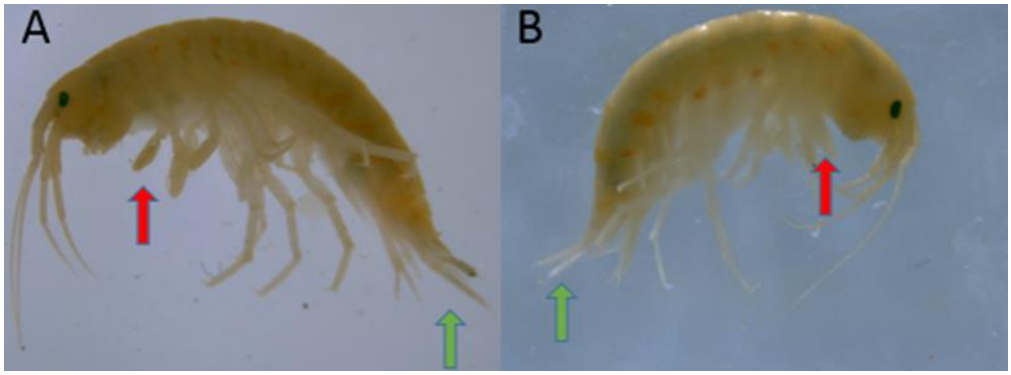

Fig. 1.5: Main differences between *G. fossarum* male (A) and female (B). Female is usually smaller than male and has a less elongated shape. The hand of the first gnathopods is quadrangular and oblong in male. In female, it is thinner (red arrows). Uropods and telson are longer in male, proportionally to the body length (green arrows) (Goedmakers, 1972).

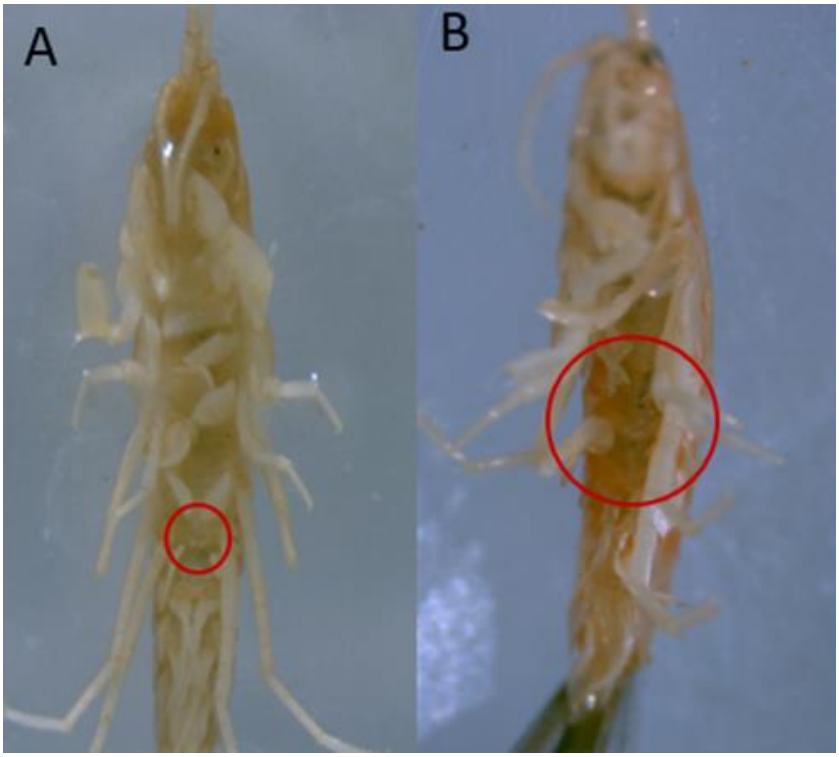

Fig. 1.6: *G. fossarum* sexual appendages. Male genital papillae are red circled in (A). Brood plates in a female amphipod are red circled in (B).

1.5.3 Reproduction

The life span of gammarids is around 1-2 years (Kunz et al., 2010). Sexual maturity (*G. pulex* and *G. fossarum*) is normally reached after having completed about ten moults respectively over 130 days at 13 °C (Mccahon et al., 1988; Pöckl, 1993). At this stage, the genital papillae (penial papillae) of the males are visible and the oostegites (large, flexible plate-like flaps extending from first thoracic segment) of the females are fully developed (Mccahon et al., 1988). Before the female is ready to oviposit, the male grasps the female with his first pair of gnathopods and they remain in precopula (Fig. 1.7) for up to two weeks (Hynes, 1955; Sutcliffe, 1992).

The duration of the precopula phase has been shown to be positively correlated with body size (Hynes, 1955). This ensures the insemination, which is external (Hynes, 1955) as soon as the female moults and releases eggs into the brood pouch (Pascoe et al., 1994). Following this stage, the precopula pair separates and the young hatch after one to three weeks in the female's brood pouch (Kunz et al., 2010). The young remain in the brood pouch four to six weeks until the female's next moult (Kunz et al., 2010). Following leaving of the brood pouch they will mate themselves three to four months later after having reached sexual maturity (Kunz et al., 2010).

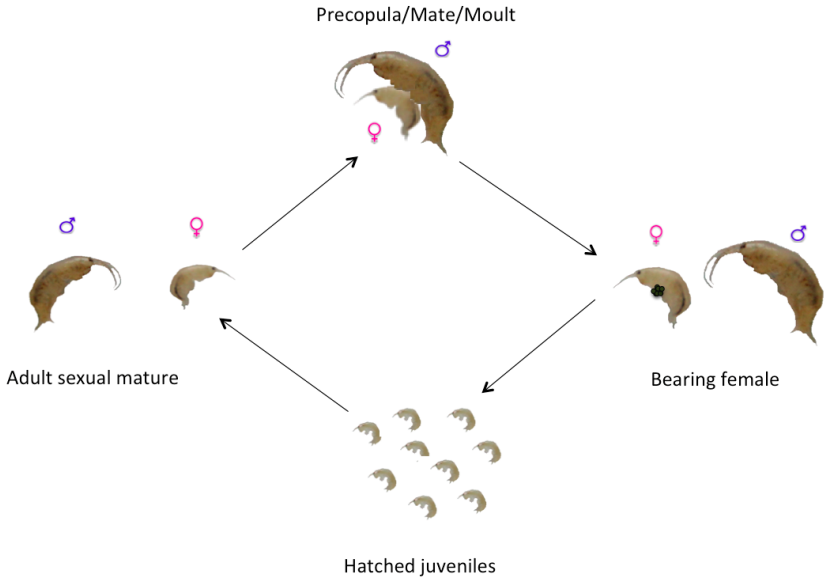

Fig. 1.7: Illustration showing the reproductive cycle of *Gammarus* spp. The cycle includes 4 main steps: precopula pair formation, moulting and subsequent mating; splitting and bearing females until juveniles hatch; juveniles reach sexual maturity and start reproducing.

1.5.4 Moulting

In order for crustaceans to grow, they must periodically shed their exoskeleton, a process known as moulting. In amphipods, moulting occurs concurrently with reproduction and is therefore essential for survival, growth and proliferation. The moult cycle occurs in four main stages based on the changes in the integument and the morphology of setagenesis (Reaka, 1975). The 4 stages are: ecdysis (E), premoult (D), postmoult (A and B) and intermoult (C). A diagram of the main stages of the moult cycle can be found in Fig. 1.8.

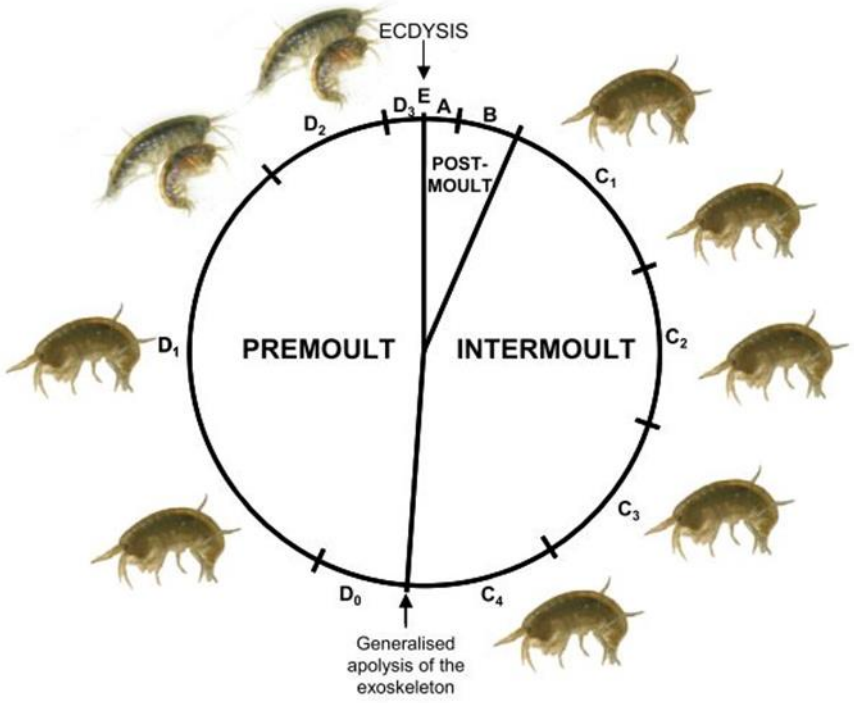

Fig. 1.8: The moult cycle in amphipods. The cycle is divided in post-moult (A and B), intermoult (C) and premoult (D) leading to ecdysis (E), the shedding of the exoskeleton. When males and females pair, the females are approximately nine days before ecdysis, a process which occurs to allow successful reproduction and somatic growth. *Redrawn from Sambles, 2007.*

A breakdown in the membranous layer initiates premoult, causing the epidermis and exoskeleton to separate. Premoult involves the secretion of the epicuticle (outer layer) and exocuticle (middle layer) and the readsorption of the endocuticle. This is quickly followed by ecdysis (Reaka, 1975). Ecdysis is the shedding of the old exoskeleton containing the epicuticle and exocuticle layers; this process is essential for the fertilisation of females and for somatic growth (Sutcliffe, 1992; West, 1997). At this point the animals are particularly vulnerable to predators due to the soft cuticle and impaired movement (West, 1997). After ecdysis the epicuticle hardens (stage A) followed by the mineralisation of the exocuticle. The formation of the endocuticle (stage B) continues into intermoult with the membranous layer formed during late intermoult (stage C4) (Skinner, 1962). At this point in the moult cycle the thickness of the cuticle is at its maximum (Reaka 1975). The moult cycle is under the control of invertebrate specific hormones, such as ecdysteroids, which direct the degradation of old cuticle proteins as well as the biosynthesis of new cuticle proteins (Suzuki et al., 2002). The moult cycle length varies with age, sex and environmental factors such as temperature, photoperiod, food supply and space (Chang, 1995; West, 1997).

1.6 Main effect of wastewater effluents on amphipods

Amphipods as detritivorous species play a fundamental role in the trophic food chain of the aquatic environments (Wigh et al., 2017). These crustaceans have been shown as very sensitive to pollution, particularly from wastewaters (Peschke

et al., 2014; Schirling et al., 2005; Schneider et al., 2015). Similarly to fish species, the consequences of the exposure of amphipods to xenobiotic substances have been studied employing a variety of approaches, including phenotypical/population investigations (Ladewig et al., 2006: Peschke et al., 2014) and molecular studies (Hook et al., 2014; Trapp et al., 2014; Gismondi et al., 2017). The following sub-sections describe studies that employed population, behavioural and molecular approaches to investigate the effects of several anthropogenic contaminants on amphipod species and illustrate the main results.

1.6.1 Population investigations

It is noteworthy that the literature on the population effects of the xenobiotic substances on amphipods (in both targeted studies and analyses of complex chemical mixtures) is variable, depending on a wide range of factors. Indeed, it has been shown that the biological effects of pollutants on the population structure and the parameters currently in use to measure the reproduction activity of these species can be very different depending on the environmental conditions, seasonal variations and chemicals used for the exposures (Ford et al., 2003; Ladewig et al., 2006: Peschke et al., 2014). For instance, Ladewig et al., (2006) observed no alterations in the sex ratio of *G. fossarum* individuals sampled downstream of two German estrogenic effluents and recorded a decrease of the proportion of breeding females and juveniles at downstream site. Gross et al., (2001) found a decrease in the size of *G. pulex* sampled in a sewage effluent

containing endocrine disrupting chemicals. On the other hand, Schneider et al., (2015) recorded an increase in the fecundity index and size of the amphipods exposing the same Gammarus species to a whole estrogenic wastewater effluent in artificial indoor flow-channels under controlled conditions. In addition, the formation of intersex individuals (showing phenotypical features of both sexes) (Fig. 1.9) has been recorded in several amphipod species and attributed to a number of factors, such as parasite infections and exposure to endocrine disruptors (Ford et al., 2008).

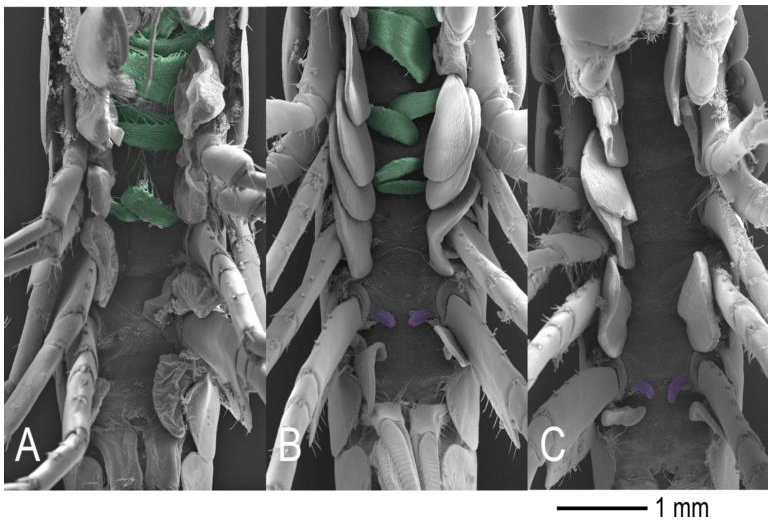

Fig. 1.9: External sexual phenotypes of the amphipod *Dikerogammarus haemobaphes*. Scanning electron microscope pictures showing a normal female with only oostegites (green) (A), an intersex male presenting genital papillae (purple) alongside oostegites (green) with rudimentary setae (B) and a normal male with only genital papillae (purple) (C) (Etxabe et al., 2015).

1.6.2 Behavioural studies

Although measuring population parameters (e.g., sex-ratio, size and fecundity indexes) being fundamental to describe the impact of water pollution on the overall population structure of the amphipods, behavioural analyses investigating the changes in response to the exposure to xenobiotic substances give additional important information and have increased over the last decade (Bossus et al., 2014; Guler et al., 2015; Barros et al., 2017; Lebrun et al., 2017; Neuparth et al., 2019). In particular, associating the altered expression of the molecular markers (genes, proteins or metabolites) in response to the exposure of the amphipods to anthropogenic chemicals in water to behavioural variations (e.g., phototaxis, respiratory activity and swimming velocity) are particularly useful to increase the knowledge on the biological mechanisms triggered by these substances at environmentally relevant concentrations (Bossus et al., 2014). Furthermore, the data obtained from these studies represent a link between physiological and ecological impacts, providing a major endpoint to assess population health and fitness (Craddock et al., 2013).

To date, there is a lack of studies evaluating behavioural variations in amphipods exposed to whole effluents. However, changes in amphipod behavioural parameters in response to a range of substances commonly detected in wastewaters have been investigated. An interesting study by Bossus et al., (2014) reports behavioural data on the responses of the amphipod *Echinogammarus*

marinus to fluoxetine and sertraline, 2 antidepressants commonly detected in river waters. A statistically significant increase in swimming velocity compared to control animals following exposure to 0.1 µg/L of fluoxetine was found. The authors speculated that the lack of significant or reduced effects in higher concentrations of fluoxetine could be due to the inhibition of a finite amount of endogenous serotonin or desensitisation phenomena (Bossus et al., 2014). Other behavioural parameters, such as locomotor and respiratory activities of *Gammarus fossarum* exposed to several metals (Cu, Cd, Ni, Pb and Zn) in a lab study were evaluated by Lebrun et al., (2017). The locomotion activity was evaluated counting the number of animals crossing a radial mark in the middle of a cylindrical beaker 5 times for periods of 30 s, with intervals of 30 s between each counting. The respiratory activity was measured using a O_2-microsensor. The authors found that both locomotor and respiratory activities were significantly affected following mono-metallic exposures. However, the alteration of these parameters was much less pronounced when the animals were exposed to the metals in mixture. Their results highlight the fact that the variations in the behavioural traits are metal-specific and complex additive/inhibitory effects occur when exposing the amphipods to metal mixtures. The effects of crude oil and several refined products on the behaviour of *Gammarus oceanicus* were evaluated by Linden, (1976). A number of sublethal effects appeared during long-term exposure bioassays in lab: the adults showed impaired swimming performance, decreased tendency to pre-copulate, impaired light reaction and decreased

production of larvae. Decreased growth was found among larvae during chronic exposure to crude oil and delayed mortality occurred among adults after a short-term exposure to crude oil with a long recovery period.

1.6.3 Molecular analyses

Looking at the molecular studies focusing on the effects of anthropogenic chemicals on amphipod biology, the literature shows a wide spectrum of variations that can occur when these crustaceans are exposed to xenobiotic substances. Well-established biomarkers have been found in amphipods in response to chemical and abiotic stressors. For instance, an increase of activity of the detoxification/antioxidant enzymes glutathione S-transferase (GST), glutathione peroxidase (GPx), catalase (CAT) and superoxide dismutase (SOD) were found by Turja et al., (2014) following exposure of *Gammarus oceanicus* to nodularin toxins (potent toxins produced by cyanobacteria). Environmental reference values for digestive enzyme activities (amylase, cellulase and trypsin) were provided by Charron et al., (2013). The authors of this study highlighted significant effects of low temperatures (below 7 ^{o}C) on the activities of the digestive enzymes in *Gammarus fossarum*. Reference activity levels of the enzyme acetylcholinesterase (AChE) were analysed by Xuereb et al., (2009). This study is particularly interesting given the massive release of organophosphorous (OPs) and carbamate (CBs) pesticides in aquatic ecosystems through agricultural

wastes, the toxicity of which results in the inhibition of this enzyme (Xuereb et al., 2009).

In addition to the biochemical assays aimed to define the thresholds of physiological enzymatic activities, studies employing high-throughput platforms largely contributed to the knowledge on the effects of aquatic pollution in amphipods. For instance, using a selected reaction monitoring mass spectrometry-based methodology, Gouveia et al., (2018) found altered concentrations of peptides annotated to moult-related proteins and detoxification proteins, such as prophenoloxidase, cythochrome P450 and glutathione S-transferase (GST), in response to field exposure of caged *Gammarus fossarum* to a number of French contaminated effluents. The transcriptomic profiles following whole-sediment exposure of the amphipod *Melita plumulosa* to a series of common environmental contaminants (porewater ammonia, bifenthrin, fipronil, diesel, crude oil, Ni and Zn) were evaluated by Hook et al., (2014) in a lab study. Using a microarray platform, the authors showed changes in transcripts annotated to digestion, growth and moulting, and the cytoskeleton following metal exposure, whereas exposure to petroleum products caused changes in carbohydrate metabolism, xenobiotic metabolism and hormone cycling.

Despite the ecological importance of amphipod species, their genome has not yet been completely annotated. This represents a limiting factor when exploring the

effects of anthropogenic compounds using high-throughput molecular approaches. Nevertheless, the mentioned studies revealed a vast spectrum of phenotypical and molecular variations that can occur in amphipod species in response to exposure to a variety of xenobiotics found in aquatic environments. However, it is important to consider that the biological responses can be very different based on the specific chemical composition of the river waters and that the effects of bioaccumulation (Munz et al., 2018) and additivity with other factors such as temperature (Charron et al., 2013) and even the annual hydrologic features of rivers (Canobbio et al., 2009) make extremely difficult predicting the long-term consequences. It is essential that in the near future ecotoxicological research focuses on discovering new molecular biomarkers to find links between the variations in gene, protein and metabolic pathways and the wide spectrum of phenotypic and behavioural alterations that have been shown to occur in species of ecological relevance in response to anthropogenic chemicals.

1.7 Project aims

Monitoring the impact of an increasing number of xenobiotics in sewage effluents and evaluating the effects of WWTP technologies on aquatic communities are currently of primary concern in European countries. The present project was aimed to investigate the impact of an anthropogenic chemical mixture released by a Swiss WWTP on the amphipod species *Gammarus fossarum.* Ultimately, the main purposes were:

- to enhance understanding of the biological toxicity of domestic, agricultural and industrial wastes on a key species in ecotoxicological monitoring
- to provide an extensive data set including population, transcriptomic and metabolomic parameters useful to develop new biomarkers of exposure to xenobiotics in amphipods

These were achieved through the following analyses:

1) A population analysis on *G. fossarum* amphipods collected in field (**Chapter 2**). The main population parameters, such as sex-ratio, length, weight and number of eggs of the brooding females were recorded on *G. fossarum* populations sampled in September 2017 and 2018, upstream and downstream of a WWTP located at northern Switzerland. The results of this analysis will represent useful information for the scientific community, given the literature gaps on this topic and the discrepancies in the published population data recorded in amphipods. A chemical analysis was also performed on the examined river **(Chapter 2)**. Specifically, the concentrations of several commonly detected compounds in river waters were evaluated upstream and downstream of the WWTP in both water and amphipod samples.
2) The sequencing of the whole *G. fossarum* transcriptome using an RNA-seq platform. This analysis was aimed to generate an extensive transcriptional data set for this important species in ecotoxicological risk assessment. The information on the transcriptome structure and the genetic annotation

data will be an essential resource to develop toxicological biomarkers in amphipods. The strategies used to assemble and annotate the complete transcriptome of *G. fossarum* from the sequencing data generated by an *Illumina HiSeq 2500* platform are described in **Chapter 3**.

3) A differential gene expression (DGE) analysis between *G. fossarum* amphipods collected upstream and downstream of the WWTP as well as between males and females (**Chapter 4**). This investigation will provide the complete set of differentially expressed genes in response to a chronic exposure of *G. fossarum* to the contaminant mixture contained in the river examined. In addition, the differentially expressed genes found between male and female amphipods will represent a solid base to find new sex-specific biomarkers in amphipods. The annotated genes were submitted in the database Panther to explore the molecular pathways impaired in amphipods exposed to the contaminant mixture. A quantitative real-time polymerase chain reaction (RT-qPCR) assay on several significantly changing genes between upstream and downstream amphipod populations detected in the DGE analysis was also performed, in order to experimentally verify the variations in the expression levels.
4) A metabolomic analysis to investigate the metabolic variations between *G. fossarum* amphipods sampled upstream and downstream of a WWTP as well as between males and females, employing a high-throughput untargeted metabolomics approach (**Chapter 5**). The data collected in this

analysis will provide additional molecular information about the metabolic changes in response to exposure of amphipods to xenobiotic compounds in field. The changing metabolites between males and females and the corresponding metabolic pathways will be useful to explore the sex-specific biomarkers in amphipods and will provide a molecular basis for further and more targeted studies to investigate the reproduction biology in crustaceans.

In the last chapter of the thesis (**Chapter 6**), the findings from each approach were compared and related with each other to reach a conclusion on the ecological impact of WWTPs and their components on amphipods. In particular, the phenotypical and population data were related to the molecular data to verify homologies and discrepancies. Moreover, strategies to elucidate in more details the results obtained from untargeted "omics" platforms and to evaluate the biologic effects of anthropogenic chemicals on species of ecological relevance were discussed.

Chapter 2 - Effects of a Swiss wastewater treated effluent on *Gammarus fossarum* population

2.1 Introduction

Water resources provide clean freshwater as ecosystem service and are thus of fundamental importance for anthropogenic activities. The relevance has been highlighted by Hochstrat et al., (2006) by means of a water stress index for various European countries. Freshwater used for domestic and industrial purposes is usually released into aquatic ecosystems as treated wastewater containing complex mixtures of chemicals like pharmaceuticals and personal care products (Daughton et al., 1999). Unfortunately, conventional WWTPs are not, or only partially able, to eliminate micropollutants during treatment processes. Hence, wastewater is one of the major sources of micropollutants in aquatic ecosystems (Seel et al., 1996; Desbrow et al., 1998; Schwarzenbach et al., 2010). The receiving streams and their biological communities may suffer from the chemical and physical (e.g., flow velocity) alterations caused by these point sources (Canobbio et al., 2009; Bundschuh et al., 2011). Although the concentration of the micropollutants can be very low in the aquatic environment (in nanogram range), they are still able to affect the most sensitive species such as fish and macroinvertebrates, potentially leading to strong adverse effects on the aquatic food web (Peschke et al., 2014).

Historically, the impacts of WWTPs on the environment have been investigated with the use of biological methods, namely biotic indices (Wenn, 2008; Morrissey et al., 2013) and chemical sensors (Pejcic et al., 2007). Biological surveys, including biotic indices, provide information on community effects and reflect the overall health of the system. They are an integrated measure of all stressors, including the total toxic effect, and provide additional information on the persistence and bioaccumulation of substances and as such are invaluable in describing the total environmental impact of an effluent (Metcalfe-Smith, 2009). Most of field studies of effluent impacts have been on fish (Triebskorn et al., 2008; Vajda et al., 2008; Al-Bahry et al., 2009; Schultz et al., 2010), though invertebrates are obviously of key relevance (Love, 2017). Invertebrates represent the major part of the animal kingdom by a substantial margin – perhaps 97% (Harley et al., 2015); they occupy critical positions in most ecosystems, facilitating decomposition and the trophic transfer of nutrients and serving as an important food source for fish, bird and amphibian species (Marcarelli et al., 2011); they are often relatively immotile and therefore subject to localised perturbations (Lebrun et al., 2011); they are often present in relatively high numbers and easily collected (Utz et al., 2009). Amphipods have been the subject of many investigations on the impact of WWTPs and their effluents (Chapter 1.6). Amongst all invertebrates in lotic ecosystems, this taxon is arguably one of the most useful as a biological indicator – they are key species in freshwater food web and they have a wide natural distribution (Peschke et al., 2014). Furthermore, the sexes are easily distinguished (Felten et

al., 2008b), and they are notable in being particularly sensitive to pollutants compared to other crustaceans (Felten et al., 2008a; Geffard et al., 2010; Jacobson et al., 2010). In addition, due to their lifestyle as a bottom-dwelling organism, many species may be particularly exposed to higher concentrations of hydrophobic compounds that are common in sewage effluents (Golding et al., 2008).

2.2 Overview on population studies on gammarids

The impact of many chemicals coming from sewage effluents on amphipod behaviour and population structure has already been identified (Ladewig et al., 2006; Bundschuh et al., 2011; Peschke et al., 2014; Schneider et al., 2015; Wigh et al., 2017; Love, 2017; Ganser et al., 2019). However, considering their long and extensive use in ecotoxicology, coupled with their ecological relevance as sentinel species, it is surprising to find that the literature on this aspect appears to be conflicting. The reason for this could be attributed to the rate of variability in the biology and ecology of the gammarids species. In fact, the differences in environmental conditions among different countries, seasonal differences and phenotypic variability in response to natural and anthropogenic perturbations are the main biasing factors. Results on the evaluation of the differences in standard population parameters, such as number of brooding females, sex ratio and size of the amphipods between sites with low and high concentrations of wastewater effluent appear to be discrepant in different studies. This can be seen in Tab. 2.1

which shows the studies that investigated changes in the population structure of amphipods between either up- and downstream of a WWTP or gradually increasing the whole wastewater fraction, in *ex situ* studies. Although these studies used different gammarid species, were performed in different geographic regions and were conducted for different durations, all the wastewaters used for the exposures had high endocrine disrupting potential. Ladewig et al., (2006) observed no consistent pattern showing an influence of two German sewage treatment plants on the sex ratio of *G. fossarum* individuals and recorded a decrease of the proportion of breeding females and juveniles downstream of the effluent. Gross et al., (2001), Peschke et al., (2014) and Schchneider et al., (2015) analysed both population structure and dynamics of *G. pulex* amphipods after the exposure to estrogenic mixtures. Gross et al., (2001) found a decrease in the size of *G. pulex* sampled in a sewage effluent with a high endocrine disrupting potential. On the other hand, Schneider et al., (2015) recorded an increase in the fecundity index and size of the amphipods exposing the same Gammarus species to a whole estrogenic wastewater effluent in artificial indoor flow-channels under controlled conditions.

Gammarids have been shown as very sensitive species to environmental stressors (Chapter 1.6). For instance, changes in water temperature (Pockl et al., 2003; Ladewig et al., 2006), parasite infections (Le Roux 1933; Zohar et al., 1998) and hydrologic features of the streams (Canobbio et al., 2009) can impact their

population structure. However, in addition to the natural seasonal effects, the overall chemical mixture coming from the sewage treatment works represents a highly limiting factor (Bundschuh et al., 2011). Wastewater chemical compositions can be very different, depending on the geographic area and the underlying technology behind the WWTPs (Nelson et al., 2011). Clearly, different compositions and even sewage waters containing the same substances in different concentrations can affect the population parameters differently. Municipal wastewater contains plenty of chemical pollutants that are not fully degraded in sewage treatment plants. These chemicals are discharged into surface waters and to date, it is difficult to predict or to conclude the ecologic impacts of single substances and their mixtures on the environment and invertebrate communities (Schneider et al., 2015).

Amphipod species	**Sex ratio**	**Fecundity Index**	**Size**	***Reference***
G. pulex	No effect	Decrease	Decrease	Gross et al., 2001
G. fossarum	No effect	Decrease	Decrease	Ladewig et al., 2006
G. roeseli and *G. pulex*	Shifted towards females	Decrease	No effect	Peschke et al., 2014
G. pulex	Shifted towards females	Increase	Increase	Schneider et al., 2015

Tab. 2.1: Effects of EDCs containing wastewaters on the main amphipod population parameters.

2.3 *In situ* vs *ex-situ* studies

In addition to a proper evaluation of the variables influencing the results of population studies, it is also fundamental to consider the positive and negative aspects of both in-field investigation and lab studies. Unlike data gathered in the field, laboratory studies offer the possibility of excluding the "noise" of innumerate variables. However, this can be also considered as a limitation. In fact, *ex situ* assays cannot cover all possible interactions between chemicals, interactions with other abiotic and biotic factors or temporal variability of exposure, which field studies can (Piva et al., 2011). Historically, studies directly comparing field and lab exposures have found the latter to be of higher (Sarakinos et al., 1997) or lower (Bloor et al., 2006; Ecetoc, 1997) toxicity. In order to get the full picture there needs to be both - controlled, *ex situ* exposure to effluents and/or their components when studying specific and subtle effects, as well as *in situ* monitoring of natural populations to detect the real impact in the environment.

2.4 Aim and objectives

The aim of the analysis presented in this chapter was to investigate the effects of a Swiss WWTP on natural *Gammarus fossarum* populations. An evaluation of the differences in the population structure between amphipods sampled up- and downstream of the WWTP was conducted in September 2017 and 2018. The following parameters were used to describe the overall population structure at

both up- and downstream sites: sex ratio, number of adults, number of juveniles and number of intersex individuals. Length and weight of the brooding females were also recorded to check whether the amphipod size was correlated to the number of eggs, which was considered as a fecundity parameter. A comparison between the data collected in 2017 and 2018 was also conducted.

2.5 Methods

2.5.1 Collaborations and contributions (Population and Chemical analyses)

Field collection of amphipods was my own work in collaboration with Dr Andrea Schifferli and Dr Thomas Bucher (Swiss Centre for Applied Ecotoxicology, Dübendorf, Switzerland). Population data collection and analysis of the chemical data, including bioaccumulation factors and toxic units calculation, were my own work. Mass spectrometry analysis of water and amphipod samples were performed by Dr Nicole Munz (Institute of Biogeochemistry and Pollutant Dynamics, 8092 Zürich, Switzerland).

2.5.2 Field collection of amphipods

Samples were collected at two locations (above and below a WWTP) along the Eulach river (Elgg, Switzerland) annually between 2017 and 2018 (Fig. 2.1&2.2). This stream receives industrial, agricultural and domestic effluent from a WWTP. A previous report by the Swiss Federal Institute of Aquatic Science and Technology

showed the presence of a high range of contaminants in this stream, including pharmaceuticals, pesticides and personal care products (Fischer et al., 2017). Additionally, the same study reported higher levels of vitellogenin in the hepatic tissue of *Salmo trutta fario* in fish sampled downstream of the WWT compared to fish sampled upstream. An altered expression of the general stress biomarker genes Abcb1 (ATP Binding Cassette Subfamily B Member 1) and PXR (Pregnane X Receptor), the toxic stress genes Cyp1a (Cytochrome P450 family), Cyp3a and GST (Glutathione S-transferase), and the metabolic gene PEPCK (Phosphoenolpyruvate Carboxykinase) were also found in fish sampled below the WWTP (Fischer et al., 2017)

The first sampling was performed in mid-September 2017 and the second one in mid-September 2018. The animals were collected ~50 m and ~100 m upstream, and ~50 m downstream of the ARA wastewater treatment plant (Fig. 2.3) (Tab. 2.2). Two upstream sampling sites were used due to a low number of animals. The amphipods collected from the two upstream sites were united in a single group. *Gammarus fossarum* individuals were collected underneath the stones and leaves at the bottom of the stream, using a standard kick-net method. A net with 1 mm mesh size was used. Approximatively 60 kick-net samplings were performed in each site. The animals were removed from the net using forceps and sorted, separating the species of interest from leaves and other invertebrates. They were placed into 10 L buckets containing stream water and quickly brought back to the

laboratory, in order to place them in controlled conditions. The amphipods were placed in glass tanks with 20 cm depth of constantly aerated stream water and the incubation conditions were 16±2 °C with a 12/12 light-dark cycle, according to Blarer et al., (2016). Gammarids were fed *ad libitum* with alder leaves (*Alnus glutinosa*) collected at the sampling site.

Fig. 2.1: *Gammarus fossarum* sampling location. The banks alongside the fast running streamlined with alder trees, the leaves from which provide food and cover.

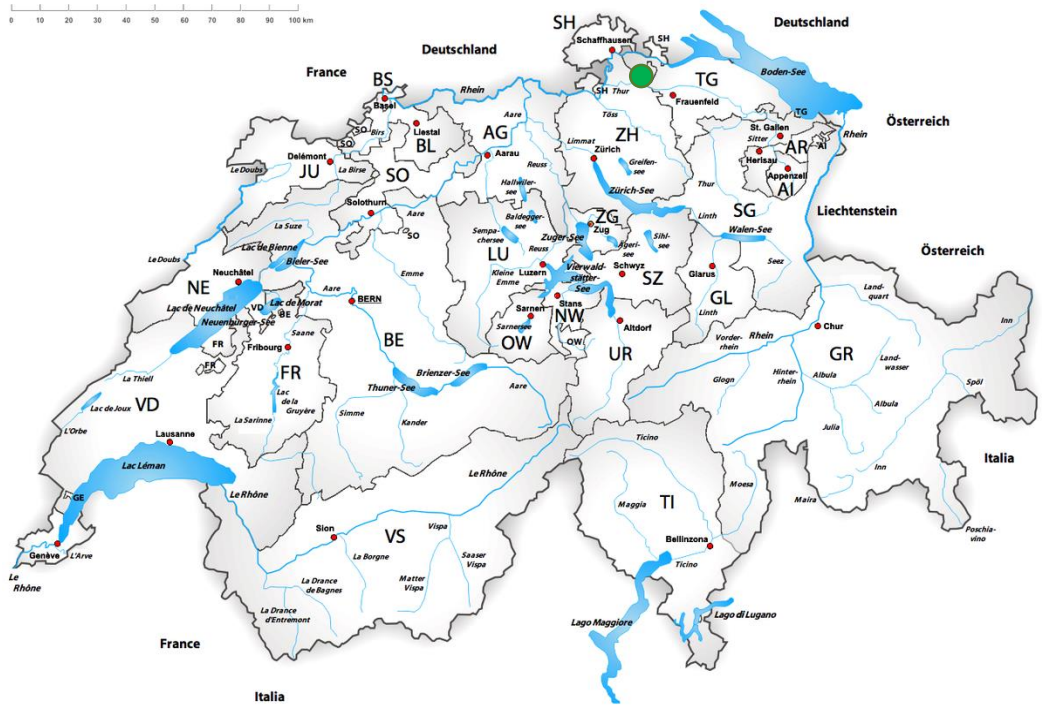

Fig. 2.2: Map of Switzerland lakes. Main lakes and streams located in Switzerland. An approximate location of the sampling site is marked by the green dot.

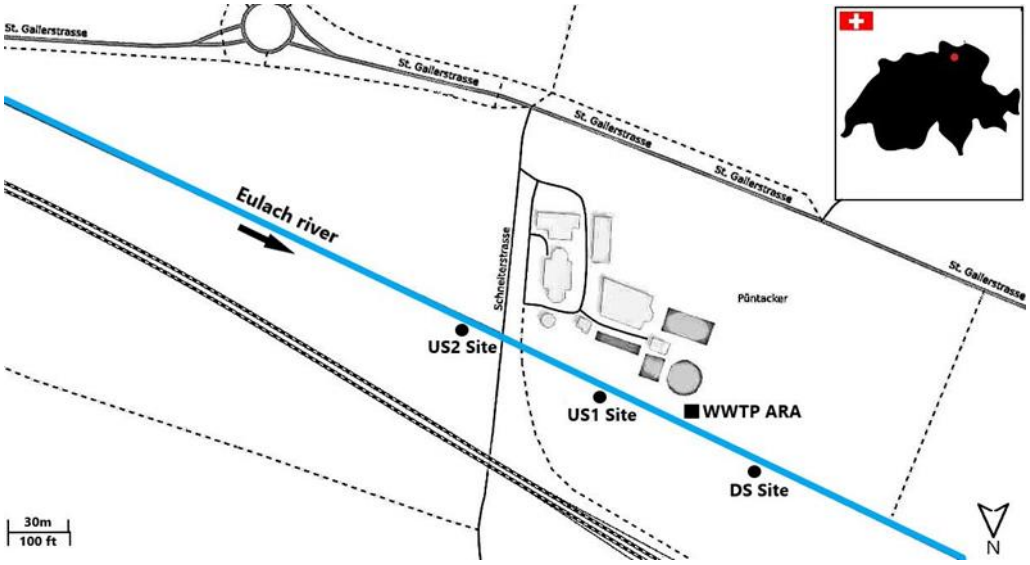

Fig. 2.3: Sampling sites at the Eulach river. Gammarids were collected at DS Site: downstream site, US1 Site: upstream site one, US2 Site: upstream site two. The ARA wastewater treatment plant is

represented by the square on the map. This map is based on OpenStreetMap© (https://www.openstreetmap.org).

Placemark	Sampling point	Latitude	Longitude	Latitude-dec	Longitude-dec
DS site	Downstream	47°30'04.43"N	8°51'05.65"E	47.501230556	8.851569444
WWTP ARA	Wastewater treatment plant	47°30'03.71"N	8°51'07.52"E	47.501030556	8.852088889
US1 site	Upstream site 1	47°30'04.23"N	8°51'09.40"E	47.500897222	8.852611111
US2 site	Upstream site 2	47°30'04.73"N	8°51'12.96"E	47.500480556	8.850822222

Tab. 2.2: Geographic coordinates where *Gammarus fossarum* amphipods were sampled. DS: downstream site of the WWTP; WWTP ARA: ARA wastewater treatment plant; US1: first upstream site of the WWTP; US2: second upstream site of the WWTP.

2.5.3 Population data collection

A total of 609 (326 upstream and 283 downstream) and 193 (103 upstream and 90 downstream) amphipods were used for population data collection on the animals from the 2017 and the 2018 sampling, respectively. Gammarids collected at both upstream and downstream sites were placed in 15 mL petri dishes and classified in two general size categories: juveniles and adults, according to the classification found in literature: young/adults (>5 mm) and juvenile/immature specimens (< 5 mm) (Adam et al., 2010). An examination under stereo microscope

(Leica S8AP0, magnification up to 80X) was carried out for both sex determination (Fig. 1.5&1.6) and evaluation of intersex individuals (Fig. 1.9). Juveniles were excluded from the sex determination analysis since genital papillae or brood plates may not be completely developed in immature individuals (Chapter 1.5.3). Females without eggs were separated from females with eggs and a count of the eggs was carried out manually, using stainless steel forceps to remove them from the brood plates. After sex determination and eggs removal from the brooding females, the amphipods were dried on paper towels, in order to remove any residual liquid. Brooding females were straightened along a ruler and lengths were measured to the nearest millimetre (mm). Lengths were recorded from the base of the first antennae to the base of the third pair of uropods, following the outline of the gammarids (Fig. 2.4) (Mayer et al., 2012). Weights were estimated in mg and measured on dried animals, using an analytic balance (Fisherbrand PS-100, Max 100g - d=0.1mg).

Fig. 2.4: *Gammarus fossarum* length measurement. Length measurement was taken considering the line between the connection point of the first antennae and the base of the third pair of uropods. Red crosses mark the two extremes.

2.5.4 Chemical analysis

The concentrations of some of the most common pesticides and pharmaceuticals found in river waters were measured in both surface water and amphipods in September 2017. An online solid phase extraction (SPE) followed by liquid chromatography coupled to high-resolution tandem mass spectrometry (LC-HRMS/MS) was employed. Temperature, pH and conductivity were recorded at up- and downstream sites annually, in September 2017 and 2018 (Tab. 2.3).

2.5.4.1 Samples for the chemical analysis

Concurrently with the amphipod sampling for the population data, two water samples of 1 L were collected from upstream and downstream sites of the Eulach

river and stored on ice for the transport. Water samples were quickly brought back to the laboratory where they were stored at -20 °C, until chemical analysis. An approximate number of 100 gammarids were also collected, in order to perform a chemical analysis of the amphipods. The animals were stored in falcon tubes on dry ice for the transport and at -80 °C in the laboratory.

2.5.4.2 Preparation of water samples

Both water samples preparation and chemical analysis were conducted according to the methodology described in Munz et al., (2017). Briefly, samples were thawed at room temperature and the pH was adjusted to 6.5-6.7. All samples were filtered (GF/F, 0.7 µm, ∅ 47 mm, Whatman, UK) and spiked with internal standards prior to enrichment. The cartridges were manually filled with 200mg EnviCarb (only cartridges for offline measurement), 350 mg or 9 mg of Strata X-AW (33 µm), Strata X-CW (25 µm, both Phenomex, Brechbühler AG, Switzerland) and Isolute ENV+ (70 µm, Biotage, Sweden) in a ratio of 1:1:1.5 and 200 mg or 9 mg OasisHLB (15 µm, Waters) for offline or online cartridges, respectively. Offline cartridges were conditioned with methanol and nanopure water and eluted in opposite flow direction with 6 mL of ethylacetate and methanol (50:50) containing 0.5% ammonia, 3 mL of ethylacetate and methanol (50:50) containing 1.7% formic acid and 2 mL of methanol. The combined neutral extracts were evaporated at 40°C under a stream of nitrogen to 100 µL and diluted with nanopure water to a final volume of 1 mL. For the online setup, the pH of 20 mL aliquots was automatically

adjusted to pH 7 with 80 µL of 0.5M citrate buffer prior to enrichment. Elution was achieved in back-flush mode with methanol containing 0.1% formic acid.

2.5.4.3 Chemical analysis of water samples

Chromatographic separation was performed using a XBridge C18 column (3.5 µm, 2.1 x 50 mm, Waters; for offline samples) with pre-column (2.1 x 10 mm) or an Atlantis T3 C18 column (5 µm, 150 mm, Waters; for online samples) with methanol acidified with 0.1% formic acid and nanopure water acidified with 0.1% formic acid as eluents. To prevent carry-over in the online setup the loop and extraction cartridge were flushed with ACN between samples. The HPLC was connected to an electrospray ionization source of a QExactive plus mass spectrometer (Thermo Fisher Scientific) which was operated in data-dependent acquisition mode to trigger MS/MS spectra with an inclusion list of all target compounds for the offline samples and in data-independent acquisition mode for the online samples. Quantification was conducted with TraceFinder v3.2 (Thermo Fisher Scientific) using internal standards and an external calibration curve in nanopure water. For substances without own internal standard the closest-matching internal standard according to retention time and structure was used. These substances were corrected for relative recovery using spiked control samples. Controls of nanopure water (with internal standard) were used to prevent and quantify potential carry-

over. Limit of quantifications (LOQ) based on the external calibration curve were corrected for matrix effects and carry-over, if detected.

2.5.4.4 Preparation of amphipod samples

Thawed gammarids were quickly rinsed with nanopure water, blotted dry with tissue, and weighed into a 2 mL microcentrifuge tube to a final weight of approximately 500 mg per sampling site (~50 organisms, depending on size). After the addition of 80 µL internal standard (1 mg/L) they were stored overnight at 4 °C. The remaining solvent was shortly evaporated with a gentle stream of nitrogen, then 500 mg of 1 mm zirconia/silica beads (BioSpec Products, Inc., U.S.A.), 500 µL of acetonitrile (ACN), and 500 µL of nanopure water were added. Extraction and homogenization were carried out using a Fast Prep bead beater (MP Biomedicals, Switzerland) in two cycles of 15 s at 6 m/s with cooling on ice in between. Afterward, samples were centrifuged (6 min, 10 000 rpm, 20 °C) and 800 µL of the supernatant were transferred into a tube containing 500 mg QuEChERS salts (4:1, MgSO4:NaCl, Agilent Technologies), vortexed, centrifuged again, and the superantant was transferred to a new tube. ACN (500 µL) was added to the first homogenate with the already used QuEChERS salts and all the steps were repeated to increase recoveries. For a further clean-up, especially for the elimination of lipids, heptane (500 µL) was added to the combined supernatant (800 µL). After vortexing and centrifuging (6 min, 10 000 rpm, 20 °C), heptane (400

µL) was removed and a second heptane extraction (500 µL) was carried out. Finally, 700 µL of the ACN phase (bottom layer) was transferred to a clean HPLC glass vial and was filled with methanol to a final volume of 2 mL. The extracts were stored at 4 °C until analysis.

2.5.4.5 Chemical analysis of amphipod samples

Similarly to the chemical analysis conducted on water samples, all gammarid extracts were analysed using an online SPE LC-HRMS/MS platform. For the gammarid extracts, 200 µL of the extract was spiked into an online vial filled with 20 mL nanopure water. Quantification of the 54 target substances was performed with internal standard calibration using the software TraceFinder v3.2/v4.1 (Thermo Scientific). After filtering out the background, only targets which showed an acceptable signal (i.e., peak shape and retention time) in the calibration curve and in at least one of the selected samples were considered for further quantification in TraceFinder. Unfortunately, there was not enough material to perform the analysis in replicate. However, previous studies applying the same mass spectrometry platform on *G. fossarum* amphipods reported no significant differences between the replicates (Munz et al., 2017; Munz et al., 2018).

2.5.5 Bioaccumulation factors (BAFs)

Bioaccumulation factors (BAFs) for the analysed compounds were calculated, where possible. Because of a slower uptake through cell membranes, polar substances tend to have lower bioaccumulation rates compared to hydrophobic compounds (Barron, 1990). The calculation of the BAFs did not include hydrophobicity and pH values, hence they were defined as "apparent" BAFs. This calculation was useful to determine whether one or more substances showed significant accumulation rates. Apparent BAFs [L/kg] were calculated as the ratio of the internal concentration in amphipods (C_{int}) [ng/kg wet weight (w.w.)] and the exposure concentration (C_w) [ng/L] in water (eq. 2.1) (Munz et al., 2018).

Equation 2.1

$$BAF = \frac{C_{int}}{C_w}$$

This calculation was possible for only 1 substance at upstream site and 8 out of 55 substances at downstream site, since the remaining substances showed an internal concentration in gammarids below the limit of quantitation (LOQ) (File S2.1 – Appendix B).

2.5.6 Toxic Units (TUs)

To translate chemical concentrations into ecotoxicologically relevant and comparable values, the inherent toxicity, expressed in toxic units (TUs) (Liess et

al., 2005; Munz et al., 2018), was determined for all the compounds for which EC_{50} values were available either in literature or in online databases. Pesticide Properties DataBase (PPDB) was consulted for the EC_{50} values of chemicals classified as pesticides, whilst EC_{50} values for pharmaceuticals and other classes of micropollutants were taken from Könemann et al., (2019). Calculations of TUs were performed by dividing the concentrations of the single compounds measured in water by the acute EC_{50} (48 h) for either *G. pulex*, or, if no effect data was available, *Daphnia magna.* In order to determine the overall toxicity of the compounds detected in the gammarids for which it was possible to calculate TU values, the single TUs were summed up to sumTU (eq. 2.2), which is based on the assumption of toxicity additivity (Warne et al., 1995). If the threshold value of −3.0 was exceeded by the sumTU, chronic effects cannot be excluded (Liess et al., 2008) (File S2.1 – Appendix B).

Equation 2.2 $$sumTU = \log \sum \left(\frac{C_i^w}{EC_{50.i}} \right)$$

2.6 Results

2.6.1 Population structures (September 2017)

A total of 609 (326 upstream and 283 downstream) individuals were used to collect the population data on the amphipods of the first sampling. The overall population structure of both up- and downstream populations can be seen in Fig.

2.5. The following sex ratios were calculated for the populations collected at upstream and downstream sites, respectively: 48.7% ♂ - 51.3% ♀ and 46% ♂ - 53.9% ♀. A chi-squared test indicated no statistically significant difference between the sex ratios of up- and downstream populations ($\chi 2$= 0.386, df=1, p-value= 0.534, α = 0.05). An observational analysis did not show intersex phenotypes in amphipods sampled both upstream and downstream of the WWTP.

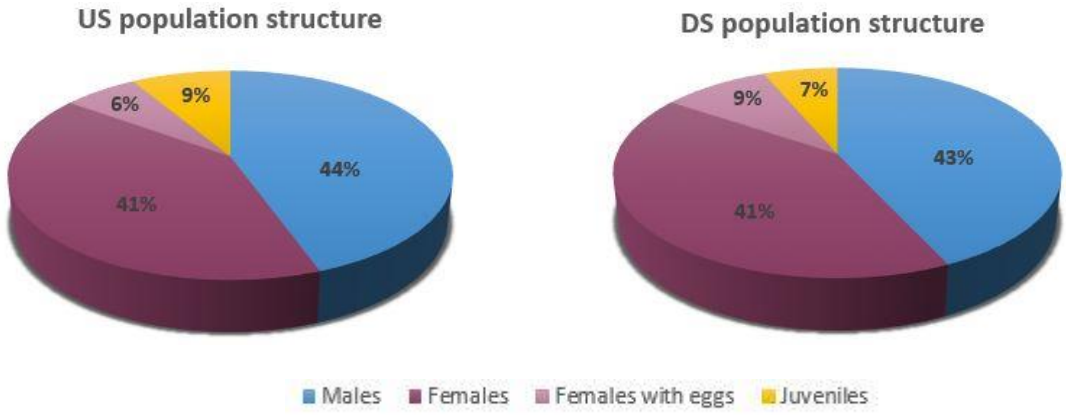

Fig.2.5: Upstream (left) and downstream (right) population structures (Sep 2017). Number of males, females, females with eggs and juveniles are expressed in percentage of the overall sample. Upstream (left) and downstream (right) population structures (US; n=326), (DS; n=283).

2.5.1.1 Number of eggs

A total of 47 brooding females (21 upstream and 26 downstream) were found in the first sampling. Length, weight and number of eggs of these females with eggs were recorded and plotted (Fig. 2.6&2.7). A positive correlation between the number of eggs and both the lengths and weights of the animals was found. This

was true for brooding females sampled at both up- and downstream sites (Fig. 2.6&2.7). An ANCOVA test comparing the number of eggs against the sampling site (including length and weight as covariates) showed no statistically significant difference between up- and downstream populations (F= 1.801, p-value= 0.187, α = 0.05). In addition, an ANOVA test did not reveal statistically significant differences when comparing the size of the brooding females (length (mm); weight (mg)) between up- and downstream populations (Length: F=1.110, df= 46, p-value=0.298, α = 0.05; Weight: F= 0.009, df= 46, p-value= 0. 926, α = 0.05).

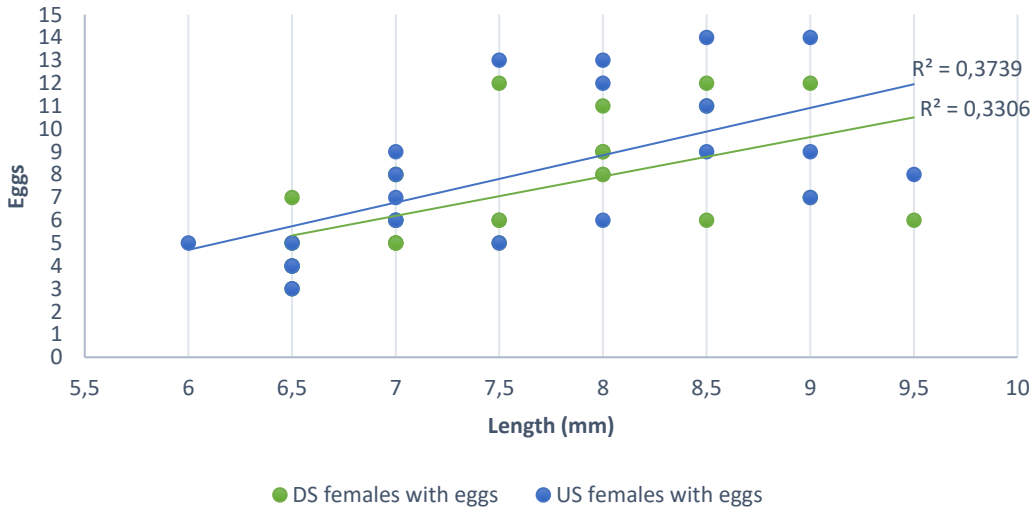

Fig. 2.6: Length of the brooding females plotted against the number of eggs (Sep 2017 sampling). Length and number of eggs measured on the brooding females collected at upstream (US; n=21) and downstream sites (DS; n=26). R^2 values are shown on the top right of the trend lines.

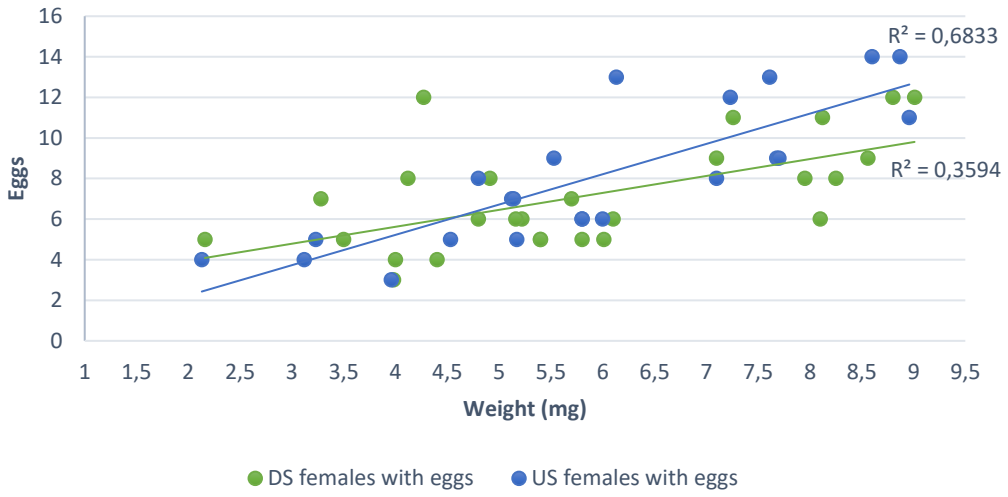

Fig. 2.7: Weight of the brooding females plotted against the number of eggs (Sep 2017 sampling). Weight and number of eggs measured on the brooding females at upstream (US; n=21) and downstream sites (DS; n=26). R^2 values are shown on the top right of the trend lines.

2.6.2 Population structures (September 2018)

The number of amphipods collected in September 2018 was lower compared to the previous year. The overall population structure was recorded on a total of 193 individuals (Fig. 2.8). The same measurements as the first sampling were performed to investigate potential differences between up- and downstream populations. The sex ratios for the upstream and downstream populations were 42.3% ♂ - 57.7% ♀ and 37.6% ♂ - 62.3% ♀, respectively.

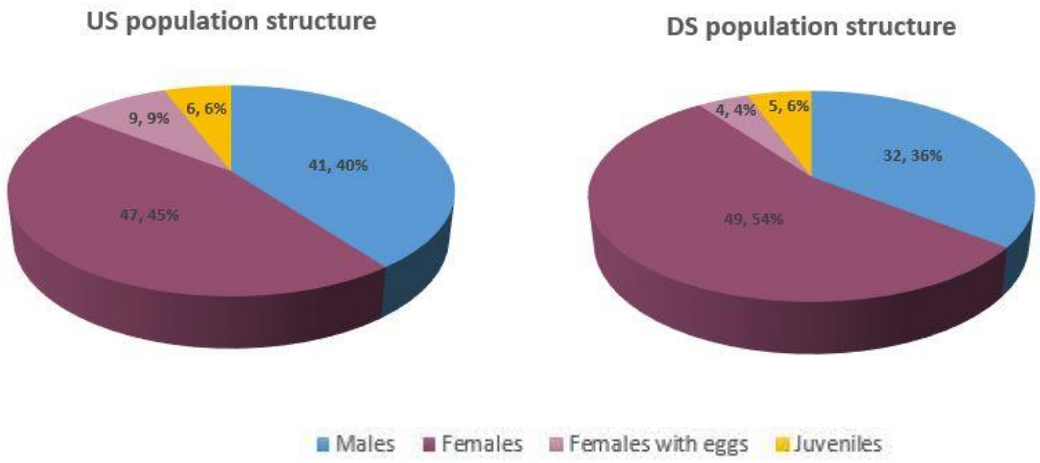

Fig. 2.8: Upstream (left) and downstream (right) population structures (Sep 2018). Number of males, females, females with eggs and juveniles are expressed in percentage of the overall sample (US; n=103), (DS; n=90).

Despite a slight female biased sex ratio observed in 2018, no significant differences were observed in the proportions of males and females between upstream and downstream populations ($\chi 2 = 0.403$, df=1, p-value= 0.526, $\alpha = 0.05$). An evaluation of the sexual phenotype (Chapter 1.5.2) did not show any intersex individual at both upstream and downstream sampling sites.

2.6.2.1 Number of eggs

As for the lower total number of animals compared to the sampling of the previous year, the number of brooding females found in the second sampling was lower too: a total of 13 brooding females (9 upstream and 4 downstream) were found in the second sampling. Plots of the number of eggs against length and weight of the

animals could not show a consistent picture of the relationship between the variables. For this reason, the correlation between the number of eggs and length and weight of the brooding females was not checked for the second sampling. Average values of length, weight and number of eggs of the brooding females along with the corresponding standard deviations (σ) were calculated (Tab. 2.3).

	Avg.length (mm)	σ	Avg.weight (mg)	σ	Avg.number of eggs	σ
US	9.05	1.99	14.33	4.94	8.78	2.54
DS	9.17	1.94	11.3	3.15	8.5	2.65

Tab. 2.3: Average length, weight and number of eggs of the brooding females collected upstream (US; n=9) and downstream (DS; n=4) of the WWTP in September 2018.

2.6.3 Comparison between population data recorded in 2017 and 2018

A chi-squared test revealed no statistically significant differences when comparing the proportions of males and females between 2017 and 2018 samplings, in both upstream ($\chi2$= 1,199, df=1, p-value = 0.273, α = 0.05) and downstream sites ($\chi2$=, df=1, p-value = 0,175, α = 0.05). Because of a low number of brooding females sampled in 2018 (upstream, n=9; downstream, n=4), a statistical comparison of size (length (mm); weight (mg)) and number of eggs between Sep 2017 and Sep

2018 was not conducted. Average values of these parameters along with the corresponding standard errors were plotted (Fig. 2.9).

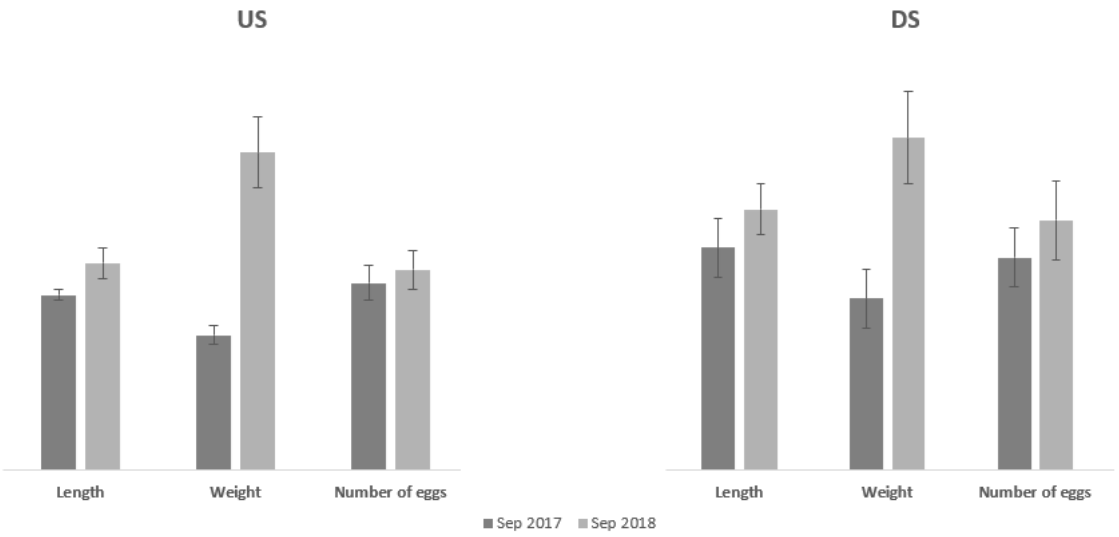

Fig. 2.9: Length (mm), weight (mg) and number of eggs of the brooding females sampled in September 2017 and 2018 upstream (US) and downstream (DS) of the sewage effluent plotted as mean ± standard error (σ/√n).

2.6.4 Detected substances and concentration patterns

When looking at the concentrations of the analysed chemicals in surface water, 34 out 55 compounds showed a concentration below the LOQ at upstream site (File S2.1 – Appendix B). On the other hand, only 18 compounds showed a concentration below the limit of detection at downstream site (File S2.1 – Appendix B). As expected, higher concentrations were found in water samples collected at downstream site compared to upstream (Fig. 2.11), for most of the chemical classes analysed (Fig. 2.10).

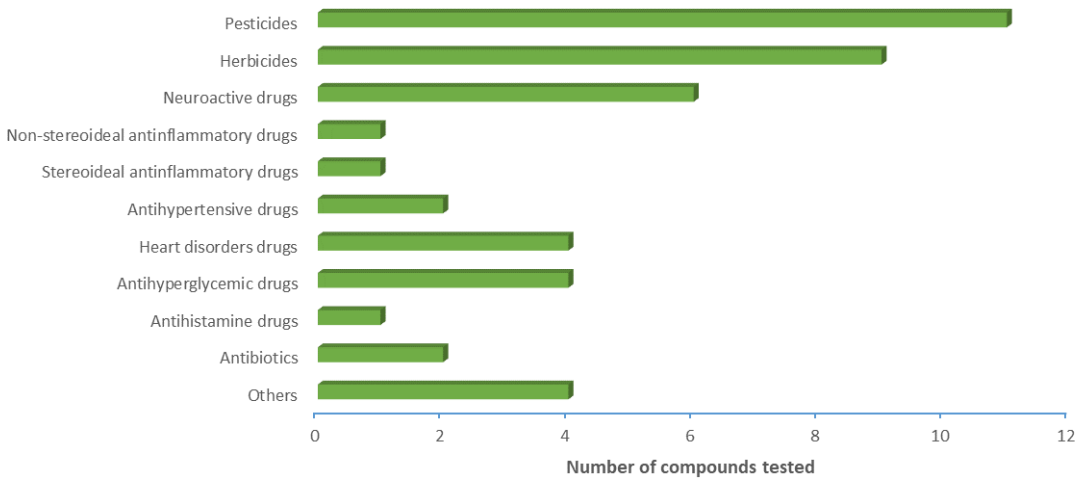

Fig. 2.10: Number of compounds tested (n=55) split into chemical classes (n=11).

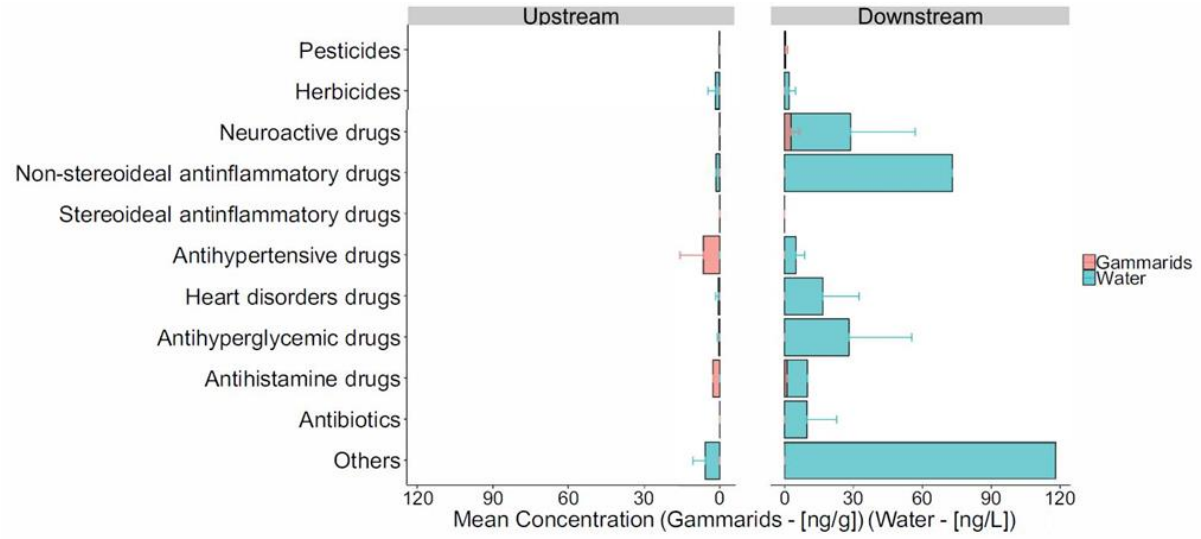

Fig. 2.11: Concentration of the 11 analysed chemical classes in both surface water (blue bars) and gammarids (pink bars). The average concentration of the substances belonging to each chemical class were used. Error bars represent the standard error.

51 out of 55 analysed substances in gammarids were found below the LOQ at upstream site and 34 out 55 at downstream site. For the substances for which the concentrations were detectable in gammarids, the concentration trend at upstream and downstream sites turned out to be different compared to water samples. For example, only 2 out of 11 chemical classes (i.e., pesticides and neuroactive drugs) showed a higher average concentration at downstream compared to upstream in amphipod samples (Fig. 2.11).

2.6.5 Physicochemical water parameters

Three main physicochemical parameters were recorded at both up- and downstream sites annually, in September 2017 and 2018 (Tab. 2.4). The monitored sites had similar temperature and pH profiles but differed by their electric conductivity with an average of 543.5 µS/cm for the upstream site and 760 µS/cm for the downstream site, between September 2017 and 2018 (Tab.2.4).

	Temperature (Sep 2017)	**Temperature (Sep 2018)**	**pH (Sep 2017)**	**pH (Sep 2018)**	**Conductivity (Sep 2017)**	**Conductivity (Sep 2018)**
Upstream	16.6 °C	17.6 °C	8.4	8.5	557 µS/cm	530 µS/cm
Downstream	16.9 °C	17.8 °C	8.3	8.2	645 µS/cm	875 µS/cm

Tab. 2.4: Physicochemical parameters measured at the Eulach river in September 2017 and 2018. The measurements were conducted above (Upstream) and below (Downstream) the ARA wastewater treatment plant.

2.7 Discussion

2.7.1 Population analysis

The aim of this chapter was to compare the populations of amphipods sampled up- and downstream of a Swiss WWTP. It is recognized that exposure to very low levels of pollution can take a long time to have an effect (Thorpe et al., 2008). Furthermore, the life-span of Gammarid species in northern latitudes is relatively long at 2 years (Sutcliffe, 1993), thus changes in population structure are likewise protracted (Paganelli et al., 2016). Also, the tolerance to pollution itself may vary temporally (Dehedin et al., 2013). Therefore, two samplings in two consecutive years (September 2017 and 2018) were performed. The following parameters were used to describe the status of both up- and downstream populations: sex ratio, number of adults, number of juveniles, number of intersex individuals and number of eggs of the brooding females, as a fecundity parameter. In order to check whether the size of the females was correlated to the number of eggs, length and weight of the brooding females were also measured. The number of amphipods found in precopula stage was very low in both Sep 2017 and Sep 2018 samplings (2017: US, n=8; DS, n=5. 2018: US, n=3; DS, n=2). This may be explained by the reproduction period of gammarid species. In fact, the reproduction rate is much reduced in autumn (October-November) (Pöckl et al., 2003) and it is possible that a breeding resting stage was starting in the sampling period. Furthermore, the amphipods were collected using a standard kick-net method, thus the separation of animals in precopula could not be ruled out during the sampling

procedure. In order to avoid an underestimation of the proportion of precopula pairs in the population, the number of amphipods in precopula was not considered as a reproduction parameter in any of the two population analyses.

2.7.2 Population structures and sex ratios

No statistically significant differences were found in the overall population structures of the animals from the 2017 sampling, in both up- and downstream sites. The sex ratios were 48.7% ♂ - 51.3% ♀ and 46% ♂ - 53.9% ♀ at up- and downstream sites, respectively. Population structures shifted towards one of the genders in apparently uncontaminated sites have been recorded in several Gammarus species including *G. duebeni* (Jones et al., 1992) and *G. pulex* (Duran, 2007). However, a percentage of ~50-50% of males and females recorded in both up- and downstream sites suggests no effects of the sewage effluent on sex ratios. No intersex individuals were found in either sampling site, indicating the absence of sex biasing factors on amphipods, such as parasite infections (Bulnheim, 1965; Engelstädter et al., 2009) or high concentrations of endocrine disrupting chemicals (Ford et al., 2004, Ford et al., 2012). A low number of juveniles compared to the overall population size, for both up- and downstream sites (9% upstream and 7% downstream) was observed and can be explained by a decrease in the reproduction rate during the sampling period (Pöckl, et al., 2003).

In order to have biologically comparable results in both the population and the molecular analyses (Chapters 3–5) all samplings were performed in mid-September. In 2018, a particularly dry summer as well as a lack of precipitation may have caused the sampling of a lower number of animals than the previous year. In fact, the number of amphipods was substantially lower compared to Sep 2017, in both up- and downstream sites. However, a total of 193 amphipods allowed to get a picture of the overall population structure in both up- and downstream sites. A statistically similar pattern as the 2017 sampling was found in 2018 with no significant differences in the overall population structure between up- and downstream sites. Interestingly, slightly female biased sex ratios were found in amphipod populations sampled in 2018, in both sampling sites. However, a low number of amphipods may have represented a biasing factor in the evaluation of the sex ratios in September 2018.

Because of their ability to interfere with the normal function of the crustacean endocrine system, for example by binding to receptors for sex hormones (Schneider et al., 2015), the presence in the water of endocrine disrupting compounds has been described as a sex biasing factor in crustaceans (Ford et al., 2012). In this study, the chemical analysis conducted on the Eulach river was not focused on evaluating the concentrations of endocrine disruptors. Therefore, it is not possible to exclude that an increase in concentration of EDCs may have occurred in 2018. However, no intersex individuals were found in in the amphipod

populations collected in both 2017 and 2018. Furthermore, no statistically significant differences were found when comparing the proportions of males and females of 2017 and 2018 samplings (section 2.6.3), suggesting that no deterioration triggering changes in the amphipod population structure or sexual phenotype occurred in 2018.

2.7.3 Fecundity and size of brooding females

Length, weight and number of eggs of the brooding females were recorded in order to evaluate any potential differences between up- and downstream populations, in both samplings. Although the biological stage of the eggs was not investigated, since the cytological aspects were not considered in the population analysis, the literature shows that the number of eggs is an established parameter of fecundity in amphipods (Pöckl, 1990; Ladewig et al., 2006; Peschke et al., 2014). In accordance with previous studies (Franke, 1977; Ford et al., 2003; Franken, 2005) the brooding females sampled in September 2017 showed a correlation between size and number of eggs (Fig. 2.6&2.7). Despite the coefficients of determination (R^2) of the scatter plots in Fig.2.6&2.7 were less than 50% for both "length - number of eggs" and "weight - number of eggs" plots, the trend lines showed a similar slope. Low R^2 coefficients were probably due to a statistically low number of brooding females collected.

An ANCOVA test comparing the number of eggs of the brooding females collected in 2017 against the sampling site (including length and weight as covariates) found no statistically significant difference between up- and downstream populations. Furthermore, no significant difference was found comparing amphipods size between up- and downstream populations. Although it is not possible to exclude sub-lethal or long-term effects, these data indicate no significant effects caused by the effluent on amphipod fecundity. The correlation between the number of eggs and length and weight of the brooding females was not checked for the second sampling, since statistical analyses conducted using a low number of brooding females sampled in 2018 (upstream, n=9; downstream, n=4) could not provide a picture representative of the whole population. When comparing length and number of eggs of the brooding females sampled in 2017 and 2018, no evident differences were observed (Fig. 2.9). On the other hand, females carrying eggs sampled in 2018 appeared heavier than 2017, in both up- and downstream sites (Fig. 2.9). However, this difference is probably due to a biasing effect caused by a low number of brooding females sampled in 2018 and cannot be considered as statistically reliable.

2.7.4 Chemical analysis

The concentrations of a total of 55 compounds, including some of the most common compounds detected in river waters, was measured for both surface water and amphipods (File S2.1 -Appendix B). This analysis included several classes

of pharmaceuticals, pesticides and other classes of micropollutants that could potentially have effects on aquatic species. The initial hypothesis was that the overall concentration of the chemicals at upstream site was lower compared to downstream, since the natural flow of the river carried to downstream the substances not retained by the WWTP. Overall, this hypothesis was confirmed by the fact that only ~37% of the analysed substances was detectable in surface water at the upstream site, in contrast to a nearly double percentage (~67%) of detectable compounds at the downstream site. A lower concentration of chemicals upstream of the discharge was consistent with the attenuation of conductivity measurements at this site compared to the downstream site (Tab. 2.4). The concentrations of the detected chemical classes, obtained calculating a mean of the concentration values of the single compounds in water samples, were found higher at downstream site for 9 out of 11 chemical classes analysed (Fig. 2.11). For some substances, a similar or higher concentration was found at upstream compared to downstream. While these detections were mainly observed for pesticides (clothianidin, propamocarb, thiamethoxam) and herbicides (atrazine, isoproturon, simazine, terbuthylazine), caffeine was also found at higher concentration at upstream site. This phenomenon is not uncommon and can be due to a number of factors. In particular, additional releases through combined sewer overflows or man-made hydraulic shortcuts (such as road storm drains or manholes of drainage systems) during rain events occurring before the sampling day, pipe leakages or wrong sewer connections

could lead to detections of certain substances upstream of WWTPs (Joss et al., 2008; Bradley et al., 2016). In addition, pesticide applications are mainly performed in spring and summer, thus concentration peaks for these substances in aquatic environments can be detected in late summer (August - September).

When comparing the average concentrations of the single compounds for each chemical class in gammarids between up- and downstream, the concentration patterns were different compared to water samples, with only 2 out of 11 analysed chemical classes having higher concentration values at downstream (pesticides and neuroactive drugs).

2.7.5 Bioaccumulation and toxic units

The chemical analysis of the internal concentrations in gammarids was performed to detect potential bioaccumulation, comparing the freely dissolved fraction with the internal fraction in gammarids, for each compound. Particularly in Swiss rivers, gammarids have been shown to bioaccumulate a wide range of chemicals, both polar and non-polar (Munz et al., 2018). A regulation of the European Parliament (EC - No1907/2006) set a threshold value of 2000 L/Kg for the BAF values of the bioaccumulative compounds. A value >2000 L/Kg represents a significant bioaccumulation rate. Therefore, apparent BAFs were calculated for the compounds analysed in this study. However, BAF calculations were possible for only 1 substance at upstream site and 8 out of 55 analysed substances at

downstream site, since the remaining substances showed a concentration in gammarids below the limit of detection (File S2.1 – Appendix B). Only one compound (the neonicotinoid pesticide imidacloprid) was shown having a BAF above 2000 L/Kg at downstream site, whilst the BAF of the remaining compounds being well below the threshold value.

Only ~5% and ~18% of the tested chemicals was detectable in gammarids at upstream and downstream sites, respectively. In addition, it is noteworthy that 8 out of 11 compounds detectable in gammarids for at least 1 of the 2 sampling sites taken into account showed a higher concentration at downstream and they belonged to only 2 of the 11 chemical classes analysed: neuroactive drugs (amisulpride, citalopram, venlafaxine, carbamazepine) and pesticides (climbazol, imidacloprid, iprovalicarb, thiacloprid). These results are supported by previous studies showing that the fraction of neuroactive drugs and neonicotinoid pesticides (e.g., imidacloprid) in river waters is increasing significantly in recent years and substantially influence the total toxic pressure in water (Brandão et al., 2013; Munz et al., 2018). In fact, in this study imidacloprid was the only compound having a BAF>2000 L/Kg (3153.846 L/Kg) and one of the chemicals with the closest value to the threshold value of toxic pressure (eq. 2.2) of -3 (-4.2) at the downstream site, across all the analysed substances.

Because of a lack in EC_{50} values in gammarids for many ecologically relevant substances in literature, the calculations of the toxic units were possible for only 12 and 19 compounds out of 55 analysed chemicals upstream and downstream of the WWTP, respectively. Based on the assumption of toxicity additivity (Warne et al., 1995), the single logTU values were summed up to sumTU (eq. 2.2). In general, if the threshold value of −3.0 is exceeded by the sumTU, chronic effects cannot be excluded (Liess et al., 2008). None of the individual compounds showed a logTU value above the threshold value of -3, neither at upstream nor downstream. However, the sum of the logTUs values was -5.2 at the upstream site and -3.67 at downstream (File S2.1 – Appendix B). Despite these sumTU values being still below the threshold value of -3 at both sampling sites, they were calculated including only ~22% of the compounds at upstream and ~35% at downstream. Therefore, it is possible that a sumTU calculated including all the compounds belonging to this chemical mixture would give a TU value above -3 at both sites.

These findings reveal that, although no clear bioaccumulation effects being shown looking at both BAF and TU values and no statistically significant differences between up- and downstream being found evaluating the population parameters in both 2017 and 2018, chronic or long-term toxic effects on the amphipod populations cannot be excluded.

Chapter 3 - Building a *Gammarus fossarum* transcriptome

3.1 Introduction to "omics" sciences in aquatic ecotoxicology

According to the Web of Science (Clarivate Analytics 2017), the term "omics" first appeared in 1999. "Omics" can be defined as a discipline of science and engineering for analysing the interactions and functions of biological information entities in various –*ome* layers or clusters of life. Although "omics" sciences have been split into many sub-disciplines (e.g., genomics, transcriptomics, proteomics, metabolomics, lipidomics, epigenomics, functionomics, immunogenomics, immunoproteomics, interactomics and pathomics), they all fall within the field of systems biology (Leung, 2018). Systems biology aims to integrate global responses within an organism from genotype to phenotype and is referred to as the integrated study of "omics" disciplines. On the other hand, the term "system toxicology" has been used to describe the integration of systems biology approaches with traditional toxicology. Although the concept is still developing within the field of aquatic ecotoxicology, the intention is to provide an integrated perspective among transcriptomic, proteomic, metabolomic and whole-organism (or even population-level) responses to specific physiological changes, which may have resulted from an environmental exposure (Sturla et al., 2014). With the advent of next-generation sequencing (NGS) methods, as well as the applications of mass spectrometry to biological systems, researchers can now look at the whole picture of the system, as opposed to looking at individual genes, proteins or

metabolites. In fact, multi-omics sciences have revolutionized scientific research, since they are able to simultaneously investigate hundreds of thousands of biomolecules at the same time, facilitating a more holistic understanding of the organism physiological status (Simmons et al., 2015).

The long-term vision of integrating “omics” approaches in the environmental monitoring programs and risk assessment has been recognized for some time, perhaps as early as some of the first cDNA-based microarray applications in environmental science at the turn of the millennium (Hogstrand et al., 2002; Larkin et al., 2003; Neumann et al., 2002; Miracle et al., 2003). For instance, over the last few years, estrogenic pharmaceuticals in the environment have been a primary concern and efforts have been directed towards identifying estrogen-responsive genes in fish using this technology (Larkin et al., 2003; Larkin et al., 2002). The goal was to develop molecular biomarkers for estrogens and endocrine disrupting compounds. There now exists more than 100 peer-reviewed studies that report on transcriptional profiles in fish and aquatic invertebrates following estrogenic treatments, an impressive dataset that can be leveraged with other databases to identify estrogen-responsive gene networks (Feswick et al., 2017). The use of “omics” sciences in the ecological field has also extended to the research for molecular biomarkers, the study of the impaired biological pathways in response to the exposure to chemical substances and in general, to an increasingly accurate characterization of large-scale molecular variations in organisms of ecological

interest (Martyniuk, 2018). In consideration of these premises, the main objective for "omics" in environmental sciences is ultimately to identify biologically meaningful molecular clusters that predict adverse outcomes which lead to negative impacts on individual fitness (Fig. 3.1).

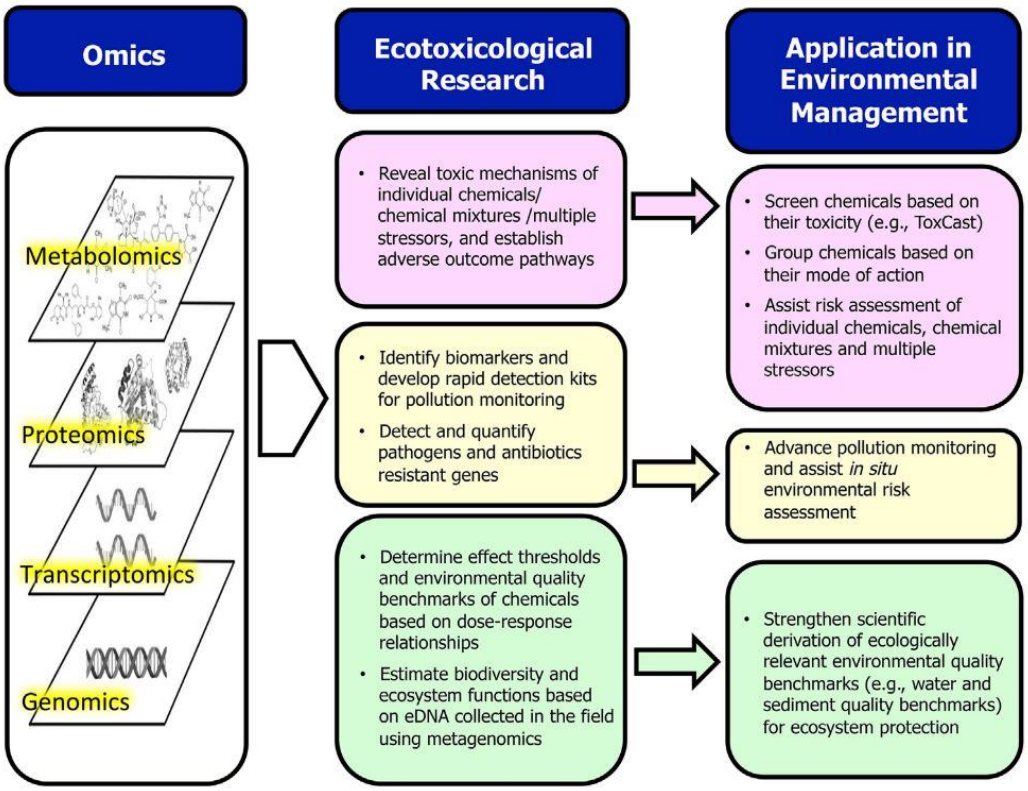

Fig. 3.1: Schematic diagram illustrating the main applications of "omics" platforms in environmental sciences. eDNA: environmental DNA; ToxCast: Toxicity Forecaster. *Reproduced with permission from Leung, 2018.*

3.2 Omics vs targeted approaches

Traditionally, in toxicity-markers research, measures of the cellular levels provide valuable information on the mode of action of uncharacterized chemicals and, the health status of exposed organisms, can act as a means to extrapolate beyond model organisms, and can be integrated into predictive risk models (Poynton et al., 2018). However, strategies focused to elucidate the toxicity of particular compounds on model organisms have strong limitations. Targeted assays alone cannot cover all possible interactions between chemicals, investigate all the molecular pathways affected and evaluate all possible changes in biomolecule interactions. Multi-omics studies, on the other hand, have the capability to provide simultaneous multi-faceted analyses of these complex interactions (Bahamonde et al., 2016; Brockmeier et al., 2017). Although interpreting and collating the significant amount of data provided by “omics” studies requires a range of statistical analyses and bioinformatics software, multi-omics sciences can offer a more thorough description of the biological system under examination. However, with particular regard to environmental/toxicology studies, one of the most significant challenges is that “omics” data usually include natural and experimental variability from multiple endpoints. That variability must be characterized and attributed in order to correctly interpret biological responses to environmental stressors. There are many potential sources of biological variation in a conventional ecotoxicology study. For example, seasonal and temporal variation may affect reproductive and metabolic endpoints (Watanabe et al.,

2007). Complex mixtures, such as effluent discharges, and differences in abiotic factors between sampling locations also add variability (Kovacs et al., 2013; Martinović-Weigelt et al., 2014). However, ubiquitous sources of variation (such as genetic variation among individuals, life history, and trophic interactions), which are often apparent when measuring conventional endpoints, can become increasingly problematic and can confound "omics" datasets. Therefore, the experimental design and the strategy for the statistical analyses become critically important in systems toxicology studies. Fig. 3.2 provides an illustrative diagram of the main sources of variability in ecotoxicology studies.

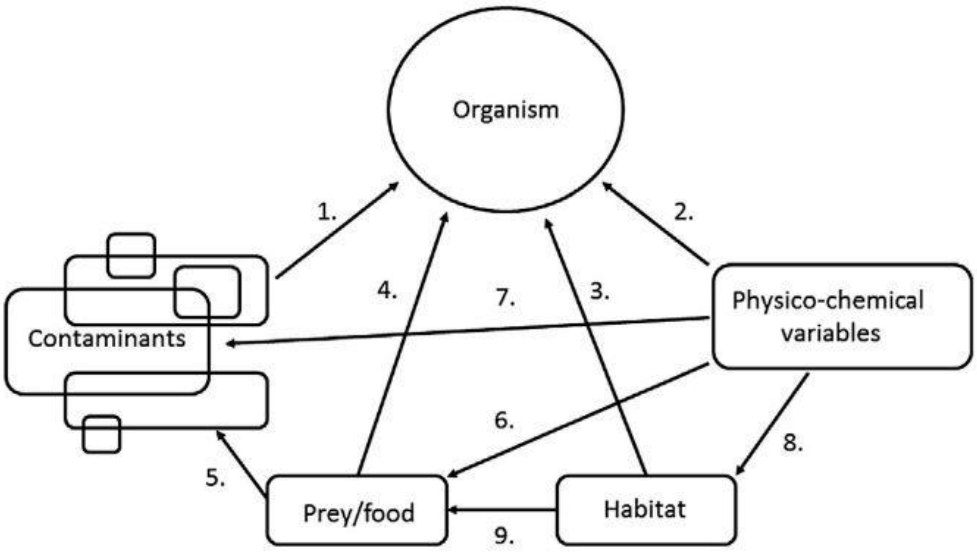

Fig. 3.2: Multiple factors which may affect the organism as stressors. 1: Exposure and effect of contaminants (possible outcomes being additivity/synergism/antagonism). 2: Physicochemical variables (e.g., climatic conditions). 3: Habitat changes. 4: Availability, type and nutritional value of food. 5: The type of food influence type and magnitude of contaminant exposure. 6: Physical variables influence availability of food (e.g., abundance of prey species). 7: Changes in environmental variables influence contaminant bioavailability (e.g., by transport/advection,

diffusion, adsorption etc.). 8: Physico-chemical variables also affect the habitat of the organism. 9: The habitat of the organism is also the habitat of its prey organism, thus influencing on type and availability of food. *Reproduced with permission from Beyer et al., 2014.*

Despite particular attention must be paid to normalize biological variation sources and analyse an amount of data, which increase with the technological advancement of "omics" platforms, it is clear that the perspective to find reliable biomarkers of water pollution is arousing more and more interest in aquatic toxicology field. In particular, the possibility to get a whole molecular fingerprint in a given organism investigating the entire set of interactions between different types of biomolecules in response to either *in-situ* or *ex-situ* chemical exposures, doubtless make the use of "omics" platforms the new frontier in aquatic ecotoxicology field.

3.3 Review on nucleic acids sequencing strategies

DNA sequencing technologies provide information on nucleotide alignment of nucleic acid sequences, such as genomic or complimentary DNA, and largely facilitate the biological studies by allowing researchers to decode the genome of living organisms (Pop et al., 2008). In the 1970s, several sequencing strategies were reported, for instance the specific chemical degradation-based DNA sequencing approach described by Maxam and Gilbert (1977). Since the chain-termination sequencing method was firstly introduced by Sanger and his

colleagues in 1977, this sequencing technique has been widely applied (Sanger et al., 1977). The chain termination sequencing method, also named as the Sanger sequencing method, or the capillary sequencing when the fluorescence dye and capillary electrophoresis were introduced, employs 2',3'-dideoxynucleotides (ddNTPs) to terminate the extension of DNA synthesis. Random incorporation of ddNTPs in the synthesis procedure results in a selection of sequences with various lengths, which are then used to determine the ordering of nucleotide on the DNA sequence by electrophoresis (Sanger et al., 1977; Metzker, 2005). In recent years, there emerged a variety of NGS technologies, such as GS FLX (454) sequencing (Roche), Illumina, SOLiD technology (Applied Biosystems) and Nanopore sequencing (Oxford Nanopore Technologies), which has been defined as 3rd - generation sequencing platform (Schadt et al., 2010). Although each of these revolutionary sequencing technologies represents a different chemistry, and each has their own pros and cons (Tab. 3.1), they all share the common feature of being able to process millions of sequencing reactions in parallel therefore dramatically increase the speed of nucleic acids sequencing (Mardis, 2008).

Platform	Library/template preparation	Chemistry	Reads length	Run time (days)	Gb per run	Pros	Cons
GS FLX (454) **(Roche)**	Frag/MP emPCR	PS	330 bp	0.35	0.45	Long reads improve mapping in repetitive regions	High reagent cost; high error rate in homo-polymers
Illumina	Frag/MP solid-phase	RTs	100-150 bp	4-9	10-1000	Currently the most widely used platform	Short reads
SOLiD technology **(Applied Biosystems)**	Frag/MP emPCR	Cleavable probes SBL	50 bp	7-14	30-50	Low error rate	Long run times
Nanopore **(Oxford Nanopore Technologies)**	Frag/SS	Transmembrane channels	10-100 Kb	No fixed time sequencing	50 Gb-5 Tb depending on the instrument	Long reads, high portability of the platform	Accuracy depends on sample parameters

Tab. 3.1: Performance data of the most commonly used NGS sequencing platforms. Frag: fragment; GA: Genome Analyser; GS: Genome Sequencer; MP: mate-pair; NGS: next-generation sequencing; PS: pyrosequencing; RT: reversible terminator; SBL: sequencing by ligation; SOLiD: support oligonucleotide ligation detection; SS: single strand. (Metzker et al., 2010; Loose et al., 2016).

3.4 RNA-seq: NGS sequencing for gene expression studies

NGS technologies have been widely used to obtain sequence information on the genomes of a large number of organisms. Their use has been fundamental for the creation of genome databases, studies on gene structure, evolutionary studies and a range of other applications in the whole biology field (McGettigan, 2013). However, as well as genomic sequencing, high-throughput sequencing technologies have become a valuable resource for the study of the abundance of RNA molecules present in a cell at a given moment under certain conditions – the so-called transcriptome. RNA sequencing (RNA-seq) has been used to analyse unknown transcript sequences, estimate gene expression levels and study single nucleotide polymorphisms (Wang et al., 2010). It has been shown that RNA-seq provides many advantages over microarray technology in gene expression studies, although more complex tools for the data analysis are required (Mortazavi et al., 2008). One of the main tasks in the RNA-seq data analysis is the identification of a set of differentially expressed genes or transcripts. Acquiring data from a differential expression (DE) analysis of individual transcripts is essential to shed light on a wide range of problems, such as identifying differences between tissues (Mortazavi et al., 2008), understanding developmental changes (Graveley et al., 2011) and microRNA target prediction (Xu et al., 2010). To perform an effective DE analysis, it is important to obtain accurate estimates of expression for each sample, but it is equally important to properly account for all sources of variation, technical and biological, in order to avoid spurious DE calls (Robinson et al., 2007;

Anders et al., 2010). For the most used RNA-seq procedures (e.g., sequence by synthesis platforms, such as GS FLX 454 and Illumina) a studied specimen of transcriptome is synthesized into cDNA, amplified, fragmented and then sequenced using a high-throughput sequencing device. This process results in a dataset consisting of up to hundreds of millions of short sequences, or reads, encoding observed nucleotide sequences. The length of the reads depends on the sequencing platform and typically ranges from 50-300 base pairs (short reads) to more than 1kb (long reads) (Tab. 3.1). Reads have to be either aligned to a reference genome by an alignment tool to determine the sequence from which they originate, or else must be assembled *de novo* into contiguous transcript sequences if no reference sequence exists. With proper sample preparation, the number of reads aligning to a certain gene can be thought of as being approximately proportional to the abundance of fragments of transcripts for that gene within the sample (Mortazavi et al., 2008) allowing the study of gene expression (Cloonan et al., 2008). However, during the process of transcription, most eukaryotic genes can actually form multiple transcript isoforms, each sharing coding parts of their sequence, in a process called “alternative splicing” (Kim et al., 2008). Since the splicing variants of a particular gene can share many reads, problems of proportionality between the number of reads and actual gene expression may come out. Therefore, an estimation of transcript expression levels needs to be conducted in a probabilistic manner. Alternatively, it is possible to only use reads that map with no ambiguity to a particular transcript.

A range of platforms are available for transcriptome sequencing and new tools for an increasingly accurate data analysis are continuously developed. In general, sequencing technologies include a number of methods that are grouped broadly as template preparation, sequencing and imaging, and data analysis. The unique combination of specific protocols distinguishes one technology from another and determines the type of data produced from each platform. These differences in data output present challenges when comparing platforms based on data quality and cost (Metzker, 2010). The following section will give a brief overview on one of the most widely used sequencing platform for RNA sequencing experiments: Illumina. This platform was used in the present study to obtain a complete transcriptome profile of the amphipod *G. fossarum*.

3.5 Illumina sequencing

Due to an exceptional sequencing speed and well-established protocols for data analysis, Illumina sequencing is likely the most used technology for RNA-seq studies to date. For Illumina sequencing, RNA extraction protocol, cDNA library preparation and fragmentation strategies are very similar to other NGS technologies. However, the solid-phase amplification is to be considered as a peculiar template preparation method of this sequencing technology (Fig. 3.3). The process involves two basic steps: initial priming and extending of the single-stranded, single-molecule template, and bridge amplification of the immobilized template with immediately adjacent primers to form clusters (Fig. 3.3).

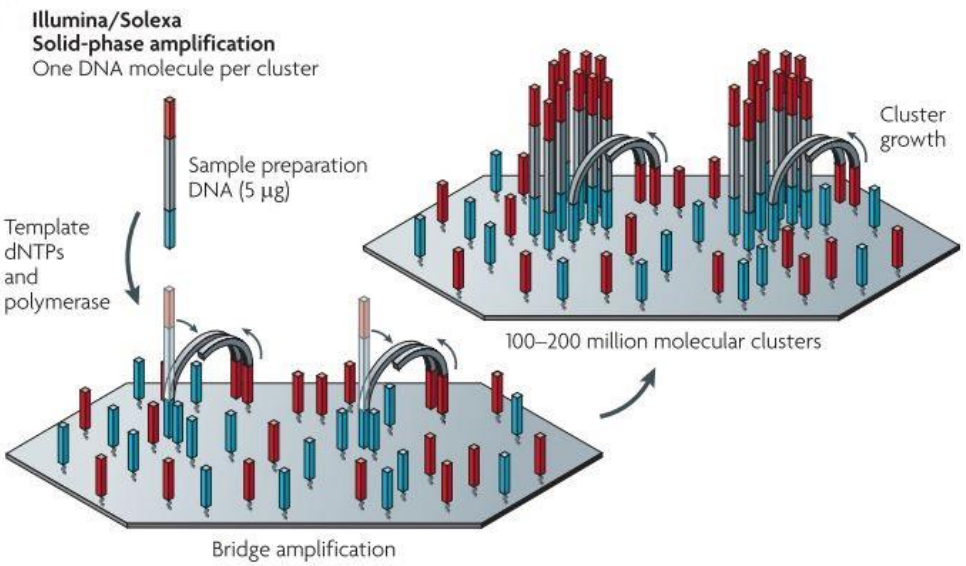

Fig. 3.3: Illumina solid-phase amplification scheme. Sequencing templates are immobilized on a proprietary flow cell surface designed to present the DNA in a manner that facilitates access to enzymes while ensuring high stability of surface bound template and low non-specific binding of fluorescently labelled nucleotides. Solid-phase amplification creates many thousands of identical copies of each single template molecule in close proximity (diameter of 1 µm or less). Because this process does not involve photolithography, mechanical spotting or positioning of beads into wells, densities on the order of ten million single-molecule clusters per cm^2 are achieved (Metzker et al., 2010).

Each cluster on the plate will contain hundreds of thousands of DNA molecules with the same sequence and this will be fundamental for the subsequent imaging analysis. In fact, during the PCR reactions needed for the sequencing-by-synthesis step, dye-labelled 3' blocked reversible terminator nucleotides are added to the reaction mix and compete for addition to the growing chain. In each extension cycle, only one nucleotide will be incorporated, based on the sequence of the

template. After the addition of the nucleotide, the clusters are excited by a light source and the characteristic fluorescence signal is detected through high-resolution photography. The software of the sequencer is able to elaborate billions of fluorescence signals coming from the clusters, recording the nucleotide sequences of the templates contained in each single cluster. Although the detection process being accurate and well-established, ambiguous signals readings can occur (e.g., for overlapping clusters). Hence, a quality score (base calling quality score) is also recorded for each nucleotide sequenced, using the Phred (Phil's Read Editor) algorithm (Prosdocimi et al., 2003). This score (*Q* in eq. 3.1) is defined as a property which is logarithmically related to the base calling error probabilities (*P* in eq. 3.1) (Prosdocimi et al., 2003).

Equation 3.1 $$Q = -10 \log_{10} P$$

The fragments of known sequence are called “reads”. After reads have been generated, they are aligned to a known reference sequence or assembled *de novo.* The decision to use either strategy is dependent on whether a pre-existing reference genome is available. For example, identifying and cataloguing genetic variation in multiple strains of highly related genomes can be accomplished by aligning NGS reads to their reference genomes. On the other hand, *de novo* assembly is the best choice for poorly annotated species with no reference genome to map the sequencing data to (Shingal, 2013).

3.6 Omics sciences applied to gammarids

Because of their ecological representativeness, invertebrates (particularly amphipods), are commonly employed as test organisms in ecotoxicological assessment. Among amphipods, the genus *Gammarus* represents the greatest number of epigean freshwater species distributed throughout the Northern Hemisphere. They are commonly used as sentinel species in freshwater risk assessment, for several reasons. First, they are widespread and found throughout a large habitat range, where they often occur at high densities. Second, they occupy a large trophic repertoire: herbivores, predators, and detritivores playing a major role in leaf-litter breakdown processes (Felten, 2003). They also constitute a food reserve for macroinvertebrates and fish. Finally, gammarids can be easily maintained in the laboratory or used for *in situ* bioassays (Kunz et al., 2010), in which one can assess the impact of pollutants by measuring molecular markers related to diverse modes of action, such as neurotoxicity (Xuereb et al., 2009), as well as by using life-history-trait reproductive features (Geffard et al., 2010). Alterations of sexual phenotype (intersexuality) have also been reported *in situ* (Jungmann et al., 2004), as well as alterations by xenobiotics of various physiological parameters related to reproductive success (i.e., gametogenesis, embryogenesis, fecundity, or moult) (Geffard et al., 2010). Surprisingly, despite all the features that made amphipods ideal for aquatic ecotoxicology research, they are still considered as poorly annotated species compared to other model organisms used in ecotoxicology, such as *Daphnia* or fish species. For this reason,

most biomarkers employed for amphipods were the result of a direct transposition from vertebrates, despite a deep evolutionary divergence (Trapp et al., 2014). The major drawback of this approach leads to many false-positives or a scarcity of results as soon as a non-model organism, distantly related to a sequenced organism, is analysed. In fact, only highly conserved and ubiquitous genes and proteins will be correctly identified. It is noteworthy that lineage-specific genes are more likely to be linked to the organism's unique biology, as demonstrated by the characterization of *Daphnia pulex* genome, a freshwater microcrustacean whose orphan genes (i.e., genes specific to a particular taxonomic group with no detectable homologs in genomes of other lineages) have been shown to be among the most ecoresponsive (Trapp et al., 2014). For instance, Colbourne et al., (2011) reported an altered expression of many unannotated gene sequences in *D. pulex* genome, in response to a variety of stress sources (e.g., hypoxia, exposure to metals, high salinity, food deprivation). Although the acquisition of more detailed genomic information in crustaceans will provide a huge support to molecular ecotoxicology research, gene products from non-model organisms responding to environmental challenges are currently overlooked.

So far, several studies have investigated the effects of water pollution on gammarids, using multi-omics platforms. For example, Leroy et al., (2010) investigated the proteomic profile of *Gammarus pulex* after exposure to polychlorinated biphenyls (PCBs), showing that pentose phosphate, cytoskeleton

organization and energetic metabolism are among the main impaired pathways. Short et al., (2014) used two NGS platforms to analyse gene expression profiles in intersex *Echinogammarus marinus* amphipods to find feminization biomarkers. Trapp et al., (2015) performed a shotgun tandem mass spectrometry analysis to investigate changes in proteomic profile of *Gammarus fossarum* in response to endocrine disruptor chemicals (EDCs). In this study, the authors found out a dose-dependent relationship between male spermatozoon production and concentration of several xenobiotics (i.e., cadmium, methoxyfenozide, and pyriproxyfen) while no induction of female-specific proteins was noted.

Although our knowledge of the biomolecular workings of the amount of molecular information on amphipods remains inferior to that of other model organisms, the availability of molecular data for these species have substantially increased over the last few years (e.g., Trapp et al., 2015; Cogne et al., 2019; Caputo et al., 2020). Availability of these data allows for the comparison of amphipods with data from other species available from online data resources. By comparing with molecular pathways known to be affected by exposure to a variety of substances from previous studies, inferences may be made as to the effects of ecotoxicity on amphipods. New strategies are being promoted to address the magnitude and wide range of effects elicited by chemicals and deficiencies in current toxicity testing approaches (Major et al., 2018). These strategies include developing adverse outcome pathway models that connect “key events” that are predictive

of harmful results, from molecular perturbations to ecologically relevant effects. In addition, comparative toxicogenomic approaches to identify evolutionarily conserved toxicological pathways and target sites enable cross-species predictions of adverse effects (Poynton et al., 2018).

3.7 Aim and objectives

In the present study, a high-throughput sequencing of total RNA from the amphipod species *G. fossarum* was performed employing an Illumina *HiSeq 2500* sequencing device. The first aim of the analysis was to use a *de novo* assembly strategy to build a complete *G. fossarum* transcriptome, in order to provide an extensive transcriptional resource for this important species in ecotoxicological risk assessment. Secondly, the transcripts set was annotated against a range of databases, both at nucleotide and protein level, to increase the molecular information on this amphipod species. An overview of the tools and strategies used to assemble and annotate the *G. fossarum* transcriptome can be seen in Fig. 3.6.

3.8 Methods

3.8.1 Collaborations and contributions (*G. fossarum* transcriptome assembly)

A subsample of amphipods collected for the population and chemical analyses were used for the RNA sequencing experiment and were my own work.

Dissections, RNA extractions and assessment of RNA quantity and quality were my own work in collaboration with Dr Melanie Fischer (aQuaTox - Solutions, Dübendorf, Switzerland). Library preparation and RNA sequencing analysis were conducted by GATC Biotech (Konstanz, Germany). *Gammarus fossarum* sub-species assignment was my own work. Quality control, assembly and annotation of the transcriptome were performed by Dr Samuel Robson (University of Portsmouth, St Michael's Building, Portsmouth, UK).

3.8.2 Amphipod sampling

A subsample of 100 amphipods from the sampling of September 2017 (Chapter 2.5.2) was used for the transcriptomics analysis. A total of 100 amphipod dissections were conducted, but in order to increase the RNA yield, 5 independent amphipods were randomly pooled per replicate, resulting in a total of 20 distinct pools. Each pool was considered an independent biological replicate for downstream analyses. Sampling was conducted independently for both males and females (5 male pools upstream, 5 male pools downstream - 5 female pools upstream, 5 female pools downstream).

3.8.3 Dissections and total RNA extractions

Total RNA was extracted from *G. fossarum* total internal tissues using RNeasy® Mini Kit (Qiagen, Hombrechtikon, Switzerland), following manufacturer's instructions. Fresh amphipods were anaesthetised for 10-15 min in a 5% (v/v)

clove oil solution prepared in water and washed in DEPC water (Sigma-Aldrich, Shnelldorf, Germany) prior to dissection, to remove any residual debris. Twenty nuclease-free Eppendorf tubes were labelled using thermal labels and placed on ice for 20 min, in order to pre-cool them until the dissection step. Dissections were performed under a stereo-binocular (x3-4 magnification; SZ2 –ILST, Olympus), using stainless steel forceps. Heads were removed from the body, allowing an easier removal of the internal tissues. Internal tissues were washed in DEPC water and placed in the 1.5 mL Eppenderof tubes previously cooled. Tubes were quickly snap-frozen in liquid nitrogen and placed at -80 °C until RNA extractions. Tubes were placed in a – 20 °C pre-cooled rack before the extractions, preventing the tissues from thawing before adding the lysis buffer. For each tube, one pre-treated stainless-steel bead (Qiagen, Hilden, Germany) and 350 µL of lysis buffer plus 10µL β-mercaptoethanol (Sigma-Aldrich, Buchs, Switzerland) were added and the samples were immediately placed into the adaptors of a Tissue Lyser II® machine (Qiagen, Hilden, Germany). Stainless-steel beads were previously subject to two treatments of 15% (v/v) H_2O_2 and 70% (v/v) ethanol washings followed by 20 min of UV irradiation, to remove any potential chemical/biological contamination. Three 20 sec mechanical stirring cycles at 30 Hz speed were performed, in order to disrupt the tissues and homogenize the cell suspension. The lysed tissue samples were centrifuged at full speed at room temperature for 3 mins to separate the cell debris from the supernatant and the supernatants were transferred to fresh tubes.

3.8.4 Assessment of RNA quantity and purity

To quantify the RNA, 1,5 µL of each sample was analysed using a NanoDrop™ ND 1000 spectrophotometer (Witec, Littau, Switzerland). The RNA samples concentration expressed in ng/µL as well as the absorbance at 230 nm, 260 nm and 280 nm were recorded. 260/280 and 260/230 ratios were also recorded, to verify the RNA purity (Tab. 8.1 – Appendix A). After ascertaining that both 260/280 and 260/230 ratios were close to 2, an RNA quality assessment by Agilent 2100 Bioanalyzer® (Agilent Technologies, Wahausel, Germany) was carried out following manufacturer's instructions.

3.8.5 RNA quality assay using Agilent 2100 Bioanalyzer®

An RNA quality assay was performed using an Agilent 2100 Bioanalyzer® device (Agilent Technologies). The Bioanalyzer is a chip-based capillary electrophoresis which allows to obtain quantitative data from protein and DNA samples as well as information about the integrity of RNA samples. The electrophoretic assay on the Agilent 2100 Bioanalyzer® is based on traditional gel electrophoresis principles that have been transferred to a chip format. The chip format significantly reduces separation time and sample consumption, compared to the traditional electrophoresis. Charged biomolecules, such as DNA or RNA are electrophoretically driven by a voltage gradient similarly to slab gel electrophoresis. Because of a constant mass-to-charge ratio and the presence of a sieving polymer matrix, the molecules are separated by size, with smaller

fragments migrating faster than larger ones. Dye molecules intercalate into DNA or RNA strands. These complexes are detected by laser-induced fluorescence. Data is translated into gel-like images (bands) and electropherograms (peaks). With the help of a ladder containing fragments of known sizes and concentrations, a standard curve of migration time versus fragments size is plotted. From the migration times measured for each fragment in the sample, the size is calculated. Two marker fragments (for RNA only one marker fragment) are run with each of the samples bracketing the overall sizing range. The “lower” and “upper” markers are internal standards used to align the ladder data with data from the sample wells. This is necessary to compensate for drift effects that may occur during the course of a chip run. For DNA and protein assays, quantitation is done with the help of the upper marker. The area under the upper marker peak is compared with the sample peak areas. Because the concentration of the upper marker is known, the concentration for each sample can be calculated. Besides this relative quantitation, an absolute quantitation is available for protein assays, using external standard proteins. For RNA assays, quantitation can be conducted comparing the sum of the sample peak areas with the ladder area. More importantly for downstream analyses (e.g., Microarray or RNA-seq), the software also calculates an RNA integrity number (RIN), which is a measure of the integrity status of the RNA sample (Agilent Technologies, 2018).

3.8.5.1 RIN value

RNA Integrity Number (RIN) is usually calculated by an algorithm for assigning integrity values to RNA measurements, at the end of Bioanalyzer run. In fact, the integrity of RNA is a major concern for gene expression studies and traditionally has been evaluated using the 28S to 18S rRNA ratio, which should be >2 for good quality RNA (Schroeder et al., 2006). RIN for a sample is computed using several characteristics of an RNA electropherogram trace, including 18S and 28S rRNA peaks, the area under the ladder and the total area under the graph (Schroeder et al., 2006). The algorithm assigns an electropherogram a value of 1 to 10, with 10 representing no significant degradation. Although the RNA profiles from eukaryotes are dominated by the presence of conserved 18S and 28S rRNA species, the total RNA profiles of some arthropods (e.g., insects and several species of crustaceans, such as amphipods) differ substantially from other profiles (Fabrick et al., 2017) (Fig. 3.4). Therefore, Agilent 2100 Bioanalyzer® is not able to calculate a discrete RIN value for these species. In fact, the rRNA of several arthropods breaks in two very close 28S rRNA picks during the electrophoretic run. The two fragments run roughly to the same spot as the 18S, creating a large fuzzy band (~1kb) of RNA, resulting in a poor peak separation (Fig. 3.4). However, an Agilent 2100 Bioanalyzer® analysis was still important to get a detailed gel picture showing the overall quality of the RNA samples.

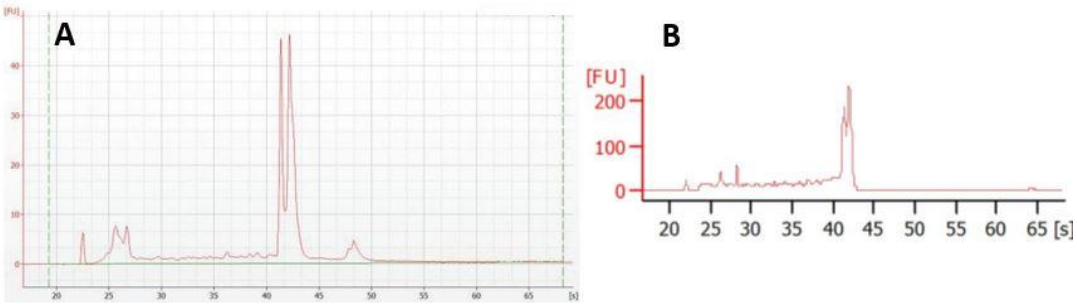

Fig. 3.4: Electropherograms obtained loading total RNA samples extracted from *Drosophila melanogaster* (A) (Fabrick et al., 2017) and *Gammarus fossarum* (B) on an Agilent 2100 Bioanalyzer®.

3.8.6 Chip loading

The samples were diluted 1:100 – 1:120 to get a final concentration of 500-5000 pg/µL, in order to make them suitable for a Pico-Chip (Agilent Technologies, Waldbronn, Germany) (Fig. 3.5) detection, subsequently loaded in the Bioanalyzer machine. After loading 9 µL of gel matrix in the wells marked "G", 9 µL of conditioning solution were added in the "CS" well and 5 µL of RNA Pico marker were placed in the "stair" well. 1 µL of each RNA sample was loaded in the remaining wells. A maximum of 11 samples can be simultaneously analysed in an Agilent 2100 Bioanalyzer® chip, thus 2 chips were used to analyse all the RNA samples.

Fig. 3.5: RNA Pico-Chip. The chip accommodates sample wells, gel wells and a well for an external standard (ladder). Micro-channels are fabricated in glass to create interconnected networks among these wells. During the chip preparation, the micro-channels are filled with a sieving polymer and fluorescence dye. Once the wells and channels are filled, the chip becomes an integrated electrical circuit. The 16-pin electrodes of the cartridge are arranged so that they fit into the wells of the chip. Each electrode is connected to an independent power supply that provides control and flexibility (Agilent Technologies, 2008).

3.8.7 RNA sequencing

An Illumina TruSeq Stranded mRNA library kit (Illumina – TrueSeq Stranded mRNA Reference Guide) was used to generate cDNA libraries for each of the 20 RNA samples. “Stranded” library preparation protocols are usually preferred to “non-stranded” methods in RNA sequencing assays for the following reasons (Hou et al., 2015):

- provide information about which of the two DNA strands a given transcript was derived
- enable the detection of antisense expression

- allow to map reads that could otherwise not be uniquely mapped to a reference genome
- provide a more accurate and complete picture of the transcriptome

Because the transcriptomic analysis was focused on obtaining information about the protein-coding transcripts, poly-A containing mRNA molecules were purified using poly-T oligo attached magnetic beads. Following purification, mRNA molecules were fragmented treating the samples with Mg^{2+} cations at 94 ^{o}C. Cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random nucleotide hexamers as primers. Actinomycin D (a molecule which binds double stranded DNA) was added to the reaction mixes to prevent spurious DNA-dependent synthesis. In order to hydrolyse RNA strands in RNA/DNA hybrid, the enzyme RNase H was also added. The synthesis of the second strand cDNA was performed using DNA Polymerase I. Strand specificity was achieved by replacing deoxythymidine triphosphate nucleotides (dTTPs) with deoxyuracile triphosphate nucleotides (dUT) in the second strand reaction mixes. In fact, the polymerase preferentially works using DNA strands containing thymidine, thus the incorporation of dUTPs ensures an over-representation of sense strands during the sequencing process. In preparation of the hybridization of cDNA on the Illumina flow cell, adapters of known sequence were bound to the molecules. Library preparation and sequencing were performed by GATC Biotech (Konstanz, Germany). The sequencing was conducted on an Illumina *HiSeq 2500* platform to generate paired-end 150 bp reads.

3.8.8.1 Quality Control

Quality control, assembly and annotation of the transcriptome were performed by Dr Samuel Robson (University of Portsmouth, St Michael's Building, Portsmouth, UK). Quality control of the raw reads was performed using FastQC v0.11.7 (Andrews, 2010). A Phred score analysis was carried out to check the overall sequencing quality of both forward and reverse reads (File S3.1 – Appendix B). Species-specific sequence purity was assessed by using a multiple genome alignment approach, by mapping reads against a database of nearly 50 different model species taken from the Ensembl database (Aken et al., 2017) using MGA v1.4 (Hadfield et al., 2014). In addition, a draft transcriptome for *Gammarus chevreuxi* (Truebano et al., 2016) and a rough transcriptome for *Gammarus fossarum* generated by pulling out the 383 sequences in RefSeq matching the following search parameters - "*Gammarus fossarum*"[porgn] were included. Read trimming was performed using Trim Galore v0.4.4 (Krueger, 2012) using the following parameters "--illumina -q 20 --stringency 5 -e 0.1 --length 20 --trim-n". This process was aimed to remove Illumina adapters and low-quality sequences, while preserving the longest high-quality part of the reads. Reads with a length < 20bp were removed.

3.8.6.2 Assembly and annotation

Reads were combined across the data set and used to generate a putative transcriptome assembly using Trinity v2.5.1 (Grabherr et al., 2011) with

parameters "--seqType fq --max_memory 100G --CPU 24 --min_contig_length 200 --min_kmer_cov 1 --SS_lib_type RF --verbose --full_cleanup". Unique transcript sequences were clustered into potential alternatively spliced isoform groups and paralogous "genes". A TransDecoder v5.0.2 (Haas et al., 2015) analysis was run using default parameters to identify open reading frames (ORF) of 100 amino acids or more within transcripts and putative protein amino acid sequences were produced. Transcripts were annotated against the Universal Protein Knowledge Base (UniProtKB) SwissProt database (The UniProt Consortium, 2017) using BLAST (Altschul, 1990), either at the protein level by taking the TransDecoder derived peptide sequence (using "blastp") or from the translated nucleotide sequence directly if no ORF was identified (using blastx). Additional annotation was performed against the Protein family (Pfam) database (Finn et al., 2013) using HMMER (Finn et al., 2011), Clusters of Orthologous Groups of proteins (eggNOG) database (Huerta-Cepas et al., 2015), the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa, 2002), and Gene Ontology (GO) database (Ashburner et al., 2000). Results were collated into a single output table using Trinotate v3.02 (http://trinotate.github.io/). Transcriptome completeness was assessed by comparing the assembly against a database of metazoan universal single copy orthologs using BUSCO v2.0 (Simão et al., 2015) (Fig. 3.6).

3.8.6.3 Coverage

In order to estimate the sequencing depth, a coverage value was calculated. The sequencing coverage is formally defined by the number of unique reads that include a given nucleotide in the reconstructed sequence (Illumina – Estimating Sequencing Coverage). The coverage metric can be expressed as a percentage providing a measure of the number of times a reference genome (or an assembled transcriptome) has been covered by the sequenced reads (eq. 3.2). It is noteworthy that the coverage requirements of a sequencing analysis strongly depend on the downstream applications of the data set. In particular, the required values can be very variable, from a 10X for the detection of mutations, single nucleotide polymorphisms (SNPs) or rearrangements in human genomes to around 100X for chromatin immunoprecipitation sequencing experiments (ChIP-Seq) (Illumina, 2014).

Equation 3.2

$$C = \frac{N(2L)}{G}$$

C is the coverage value, *N* represents the total number of reads, *L* is the read length (multiplied by 2 when a paired-end sequencing strategy is used) and *G* is the total length of the assembled transcriptome.

3.8.8 *Gammarus fossarum* sub-species assignment

A sequence alignment of the whole transcriptome against the complete *G. fossarum* mitochondrial genome (Macher et al., 2017) was conducted using Blast2GO 5 Basic software (Conesa et al., 2005), in order to identify transcripts corresponding to mitochondrial genes that may allow for taxonomic assignment (i.e., rRNA 16S and CO1) (Müller, 2000; Weiss et al., 2014). The following parameters were used for the analysis:

- E-value: $1.0E^{-3}$
- Number of blast hits: 20
- Word size: 11
- HSP length cutoff: 33

BLAST was then used to align putative mitochondrial transcripts against the NCBI non-redundant database (Altschul et al., 1990) for "amphipods – taxid:6821", to assign both *G. fossarum* sub-species and CO1 type.

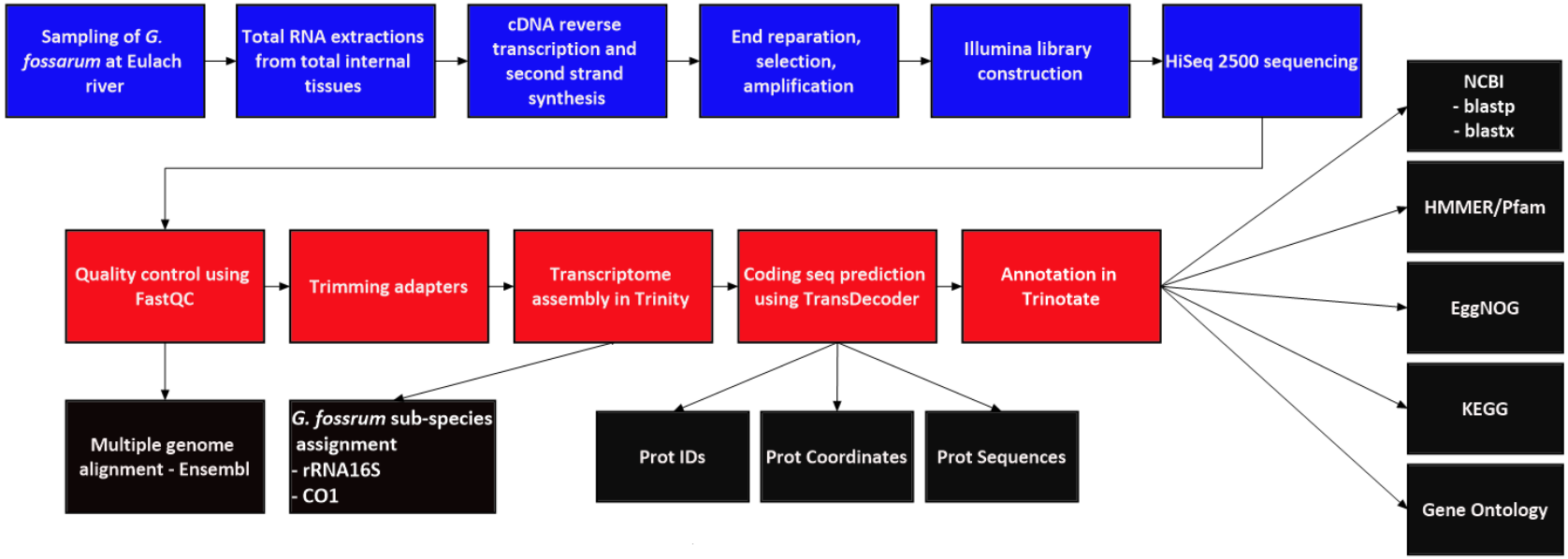

Fig. 3.6: Overview on the workflow used for generation and annotation of the *Gammarus fossarum* transcriptome. Blue steps are wet-lab procedures, red steps are *in silico* and black steps represent the outputs.

3.9 Results

3.9.1 Assessment of RNA quantity and purity

A NanoDrop™ ND 1000 spectrophotometer was used to measure the concentrations of the RNA samples and the absorbance at 230 nm, 260 nm and 280 nm. Nucleic acids absorb at the wavelength of 260 nm, proteins absorb at 280 nm, and phenol and carbohydrates absorb at 230 nm (Thermo Fisher Scientific, 2008). A 260/280 ratio of ~2 is generally accepted as "pure" RNA, while 260/230 ratio is used as a secondary measure of nucleic acid purity, and should be ~2 (Thermo Fisher Scientific, 2008). All samples showed 260/280 ratios close to 2, indicating that the extraction procedure generated good purity RNA samples (Tab. 8.1 – Appendix A).

3.9.2 Agilent 2100 Bioanalyzer® assay

After diluting the RNA extracts to reach the proper concentration for the Agilent Pico-chips (500-5000 pg/µL), 1 µL of each RNA sample was used for the chip loading. In order to perform an RNA quality assay, the chips were loaded in an Agilent 2100 Bioanalyzer® device and a capillary electrophoresis was run following manufacturer's instructions. Fig. 3.7 shows the electrophoretic traces of the RNA samples. The bands at ~1 kb in each lane correspond to intact ribosomal RNA, indicating that a significant portion of intact RNA was still present within the samples.

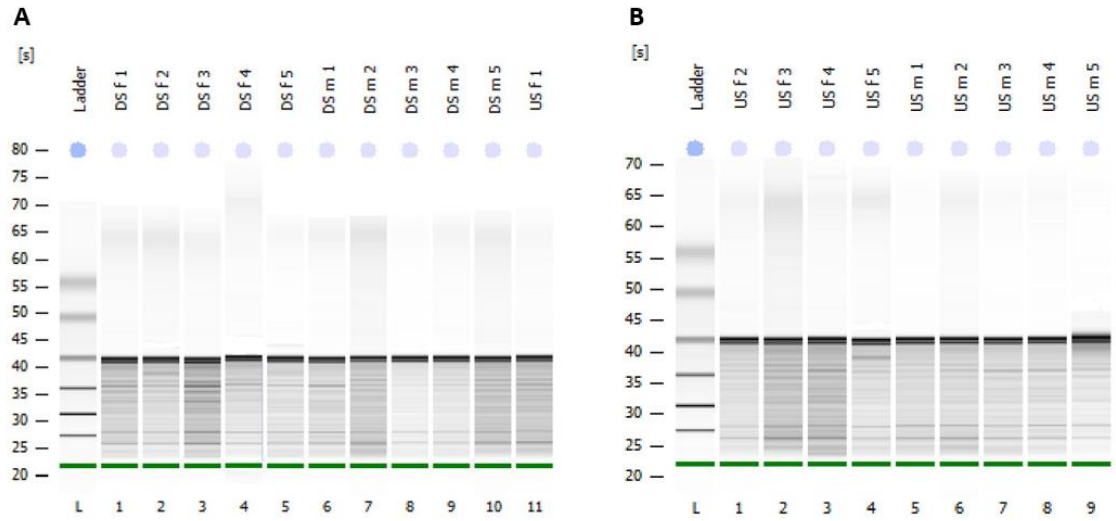

Fig. 3.7: Agilent 2100 Bioanalyzer® gel pictures. Pictures obtained following a capillary electrophoresis run on the total RNA extracts (A: first group of samples; B: second group of samples). The lanes marked with [s] show the migration times expressed in seconds. The lanes marked with "Ladder" contain a 4000-25 bp DNA ladder. The intense bands corresponding to ~40 sec of migration time (~1 kb) in each lane correspond to intact ribosomal RNA.

3.9.3 *Gammarus fossarum de novo* transcriptome assembly

The data set generated in this study consists of a total of 325,393,762 paired-ends 150 bp reads across the 20 samples (Tab. 3.3). From these, a transcriptome 427,679,404 bp in length was assembled. The coverage for this assembly was 228.25X. Trimming of reads resulted in the loss of only 0.1% of reads. The Phred encoded base calling quality score plots for both forward and reverse reads obtained from the sample DS♀1 can be seen in Fig. 3.9 and are representative of all samples (File S3.1 – Appendix B). In total, 324,958,898 quality-trimmed reads were used for *de novo* transcriptome assembly. The final assembly consisted of

680,840 transcripts, clustered into 407,060 genes by the Butterfly portion of the Trinity algorithm (Grabherr et al., 2011). In this process, the algorithm grouped transcripts into clusters based on shared sequence content (e.g., isoforms with shared exons). The transcriptome consisted of a total of 427,679,404 bp and showed a GC content of 44.68 %. Transcripts ranged in size from 100 bp to 31,653 bp, with a median transcript size of 628.16 bp and an N50 of 1,026 bp (Tab. 3.4). N50 is a parameter describing the quality of assembled genomes/transcriptomes that are fragmented in contigs of different length. Among the longest contigs, N50 represents the length where all contigs greater or equal to that length, make up at least 50% of the assembly length (Baker, 2012) (Fig. 3.8).

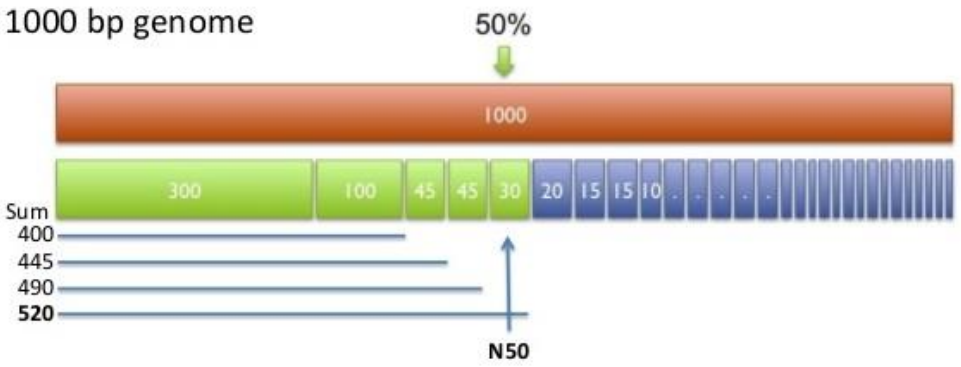

Fig. 3.8: Schematic of N50 assembly parameter. The set of assembled contigs is sorted from the longest to the shortest contig. Contig sizes are summed until half of total assembly size (i.e., sum of all contig sizes in the assembly) is reached. The last contig size added in the calculation is N50.

In order to exclude low abundance transcripts from the N50 calculation, N50 values for the sequences expressed within the data set (Ex%N50) were also calculated. Briefly, the transcripts were ordered based on their abundance; transcripts for which the number of reads mapping to them accounted for 50% of the total number of reads were used in the calculation (Tab. 3.2).

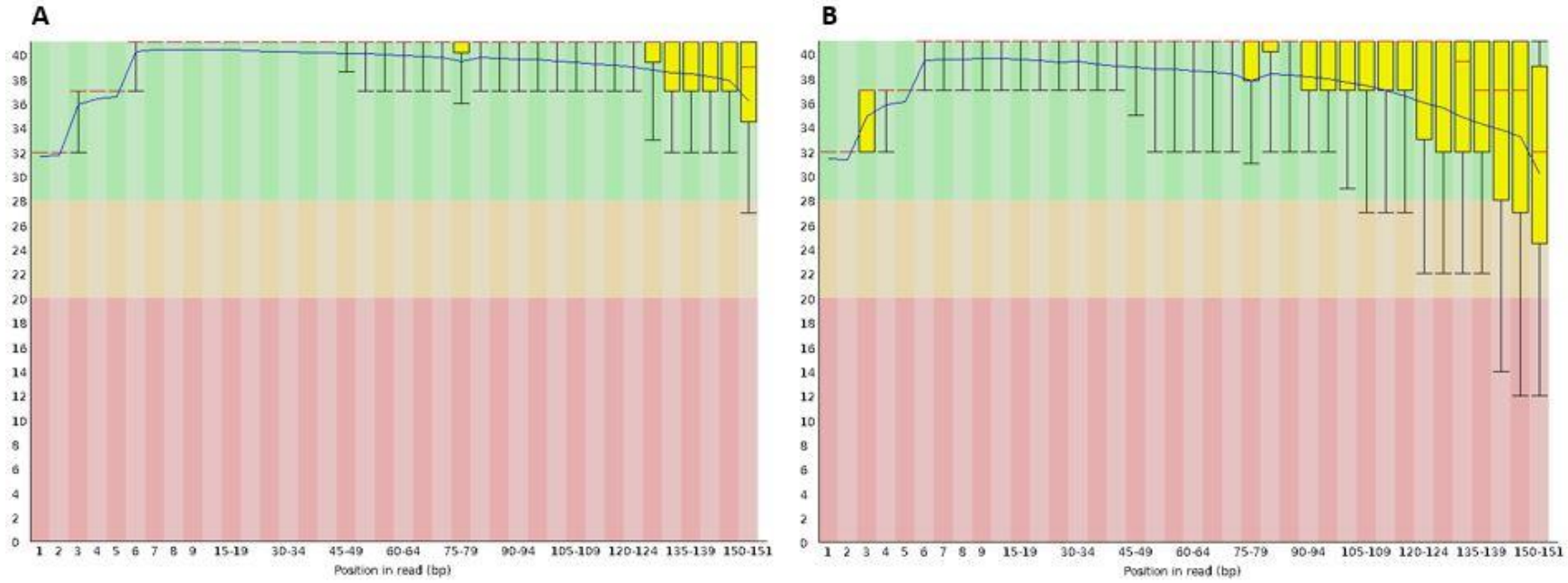

Fig. 3.9: Base calling quality score panels for both forward (A) and reverse (B) reads obtained from the sample DS♀1. The positions on the reads in 5'-3' direction are shown on the x-axis and the base calling quality scores on the y-axis. Scores in the yellow and red zones are considered as less reliable.

E	ExN50	num_transcripts
80	1762	9248
81	1784	10670
82	1796	12354
83	1806	14346
84	1838	16720
85	1869	19579
86	1885	23040
87	1891	27271
88	1925	32513
89	1924	39108
90	1922	47563
91	1908	58675

92	1890	73814
93	1854	95037
94	1782	125104
95	1675	167085
96	1533	224515
97	1352	301980
98	1152	410271
99	1080	624084
100	1005	695754

Tab. 3.2: N50 values for expressed reads in the data set (ExN50). "E": percentile for the expression in the data. "ExN50": N50 score when looking at a subset of transcripts. "num_transcripts": number of transcripts.

The Ex90N50 was 1,922 bp and is accounted for 47,563 (6.8%) of the transcripts (Tab. 3.2). Mapping of the reads back to this assembly showed an overall mapping rate of 65.7%.

Sample	Read Count	Trimmed reads	Mapped reads	Mapping %
Downstream Female Rep1	22650699	22623094	14726604	65.1 %
Downstream Female Rep2	13517269	13497408	9001357	66.7 %
Downstream Female Rep3	17514244	17494057	11642836	66.6 %
Downstream Female Rep4	14999538	14973541	10426769	69.6 %
Downstream Female Rep5	12163468	12142453	7967433	65.6 %
Downstream Male Rep1	20428324	20405964	12447287	61.0 %
Downstream Male Rep2	14631291	14607452	9925433	67.9 %
Downstream Male Rep3	16152782	16135765	10708075	66.4 %
Downstream Male Rep4	18340712	18320979	12263090	66.9 %
Downstream Male Rep5	16527724	16503722	9593208	58.1 %
Upstream Female Rep1	12332374	12319711	9112684	74.0 %
Upstream Female Rep2	11168506	11147888	7621486	68.4 %
Upstream Female Rep3	19176265	19154074	13389168	69.9 %
Upstream Female Rep4	15397710	15379802	9375903	61.0 %
Upstream Female Rep5	22736072	22697108	14779999	65.1 %
Upstream Male Rep1	12404555	12386567	8213932	66.3 %
Upstream Male Rep2	22943358	22913556	14657959	64.0 %
Upstream Male Rep3	12283566	12274253	8262388	67.3 %
Upstream Male Rep4	17513347	17490469	10802000	61.8 %
Upstream Male Rep5	12511958	12491035	8482072	67.9 %

Tab. 3.3: Read parameters for each sample. Read count: total number of reads obtained from each sample; Trimmed reads: number of reads following the trimming process; Mapped reads: number of reads that mapped back to the assembly; Mapping %: mapping rate of the reads of each sample to the assembly.

Transcriptome completeness was assessed by comparing against a database of 978 metazoan universal single-copy orthologs using BUSCO (Simão et al., 2015). In this assembly, 942 (96.4%) universal single-copy orthologs were present in a complete form, with 335 (34.3%) showing a single copy and 607 (62.1 %) showing 2 or more copies. These genes may correspond to multiple isoforms in the

transcriptome. 19 (1.9 %) were found in a fragmented form whilst only 17 (1.7%) were missing. These data suggest that this assembly represents a very complete transcriptome for *G. fossarum*. Annotation against the UniProt/Swissprot database identified candidate hits for only 80,476 (11.8%) of the transcripts. However, 53.5% of transcripts with an identified open-reading frame (ORF) (representing coding genes) showed a hit, whilst only 5.6% of transcripts with no ORF showed a hit. The vast majority of non-ORF transcripts (78.7 %) were shorter in length than 500 bp (Fig. 3.10). This suggests that the majority of non-annotated non-coding transcripts is likely a result of fragmented RNA.

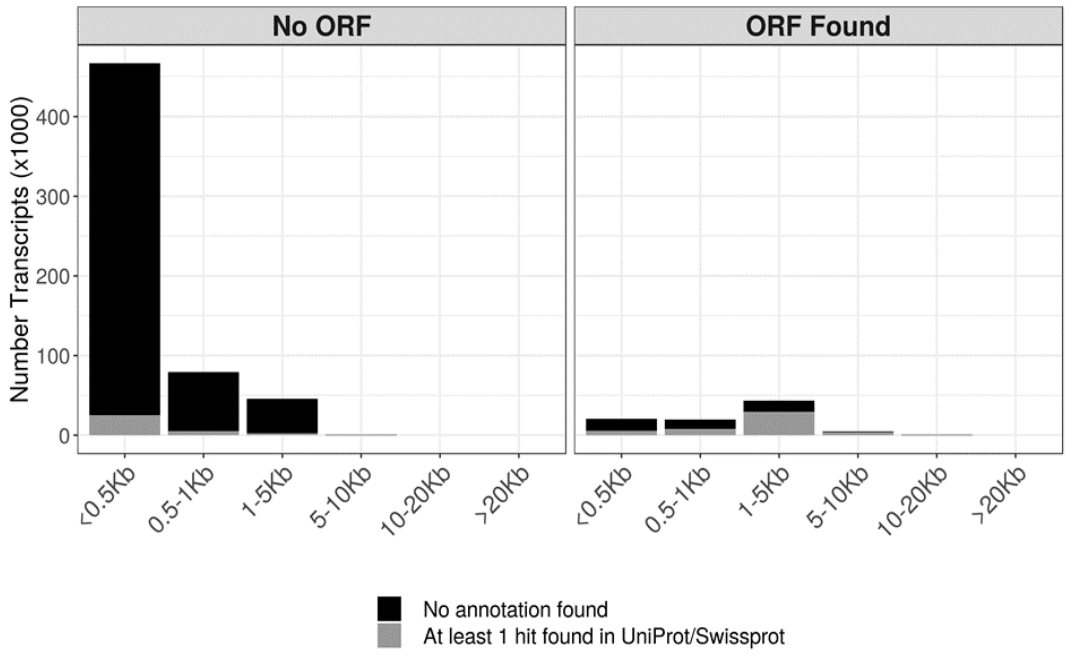

Fig. 3.10: Annotation rates for both the transcripts with an ORF and with no ORF found. The plot shows the number of transcripts with no ORF (left) and with ORF (right) found that show at least a hit against the UniProt/Swissprot database. The transcripts were ordered by their length (x-axis).

Transcripts were filtered to remove transcripts that were assigned to genes from archea, bacteria or viruses, transcripts with no ORF and transcripts less than 500 bp. In addition, for multiple transcripts in the Trinity output, only the longest transcript was retained for future analyses. This produced a filtered data set of 20,836 distinct genes.

Total Trinity 'genes'	407,060
Total Trinity transcripts	680,840
Percent GC	44.68%
Longest contig	31,653 bp
Shortest contig	100 bp
Greater than 10 Kb	352
Greater than 5 Kb	5,389
Greater than 2 Kb	94,819
Transcript contig N10	4,671 bp
Transcript contig N20	3,075 bp
Transcript contig N30	2,184 bp
Transcript contig N40	1,531 bp
Transcript contig N50	1,026 bp

Median contig length	324 bp
Average contig	628.1 bp
Total assembled bases	427,679,404

Tab. 3.4: Summary statistics of *Gammarus fossarum* transcriptome assembly. The table shows the main parameters describing the overall structure and quality of the assembly.

3.9.4 Annotation

The overall rate of annotation against Uniprot/Swissprot was 80,476 annotated transcripts out of 680,840 total transcripts (11.8%). However, the rate of annotation becomes more significant when focusing on transcripts that are more likely to represent genes. In fact, 53.5 % of transcripts with an identified open-reading frame (ORF) showed a hit, whilst only 5.6% of transcripts with no ORF showed a hit (Fig. 3.10). In order to filter out most of non-coding transcripts and false positives, only the transcripts with the following features were retained:

- Transcripts that were not assigned to genes from archea, bacteria or viruses
- Transcripts with an identified complete ORF
- Transcripts greater than 500 bp in length
- Fragments per Kilobase per Million (FPKM) > 1

20,836 out of 38,493 (54.13%) annotated transcripts were found to pass the filters above. Fig. 3.11 shows the number of annotated transcripts that gave hits from either bacteria, archaea, viruses, eukaryotes or unknown. Interestingly, when looking specifically at the eukaryotic genes, human, mouse and fly were the species that gave more hits. This might be due to the presence of highly conserved genes and the fact that these species are amongst the most well-categorised species within the UniProt database. The number of transcripts annotated against Eukaryotes, using the filters mentioned above was 11,203 out of 20,836 (53.8%). In addition, 7,449 transcripts out of 11,203 (66.49%) were annotated against unique genes from UniProt. In order to obtain a biological overview on the whole set of annotated transcripts, the transcripts were split into different GO-slim terms. This classification was conducted in all three GO categories: Cellular component, Biological Process and Molecular function (Fig. 3.13). An additional annotation specifically against *Drosophila melanogaster* genome was performed (File S3.2 – Appendix B). The analysis identified 105,843 unique *Drosophila* genes (15.5% of the total number of transcripts) using the blastp tool. An annotation of the whole set of transcripts was also performed against the Clusters of Orthologous Groups of proteins (eggNOG) database (Fig. 3.12).

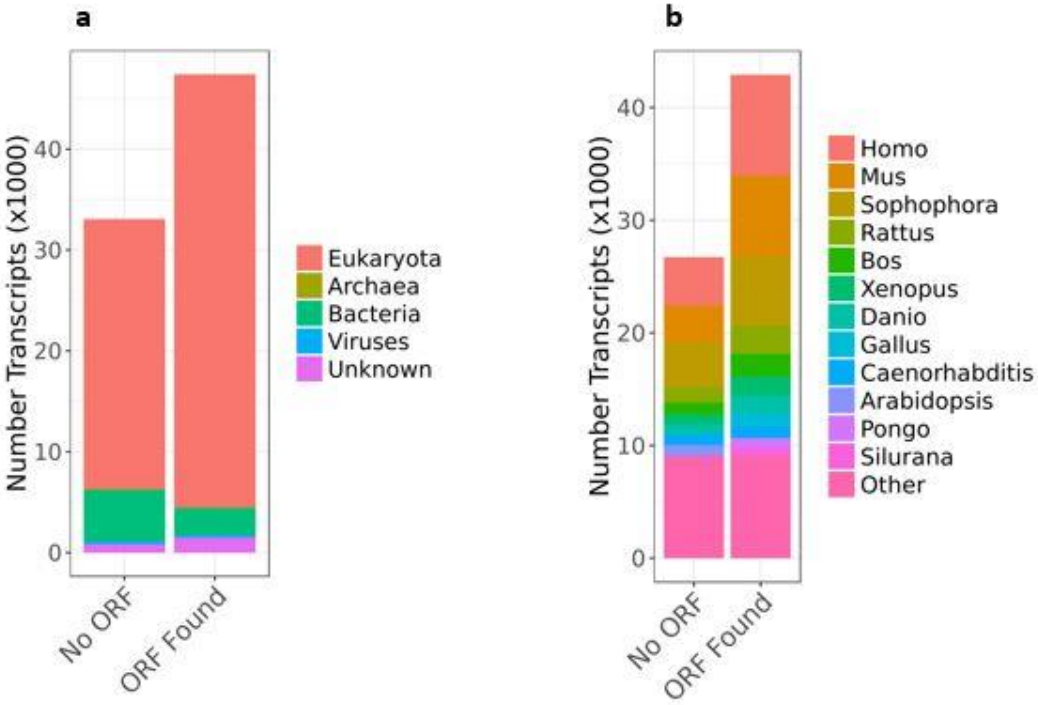

Fig. 3.11: Number of transcripts annotated against different biological domains (a) and genera (b) in UniProt database.

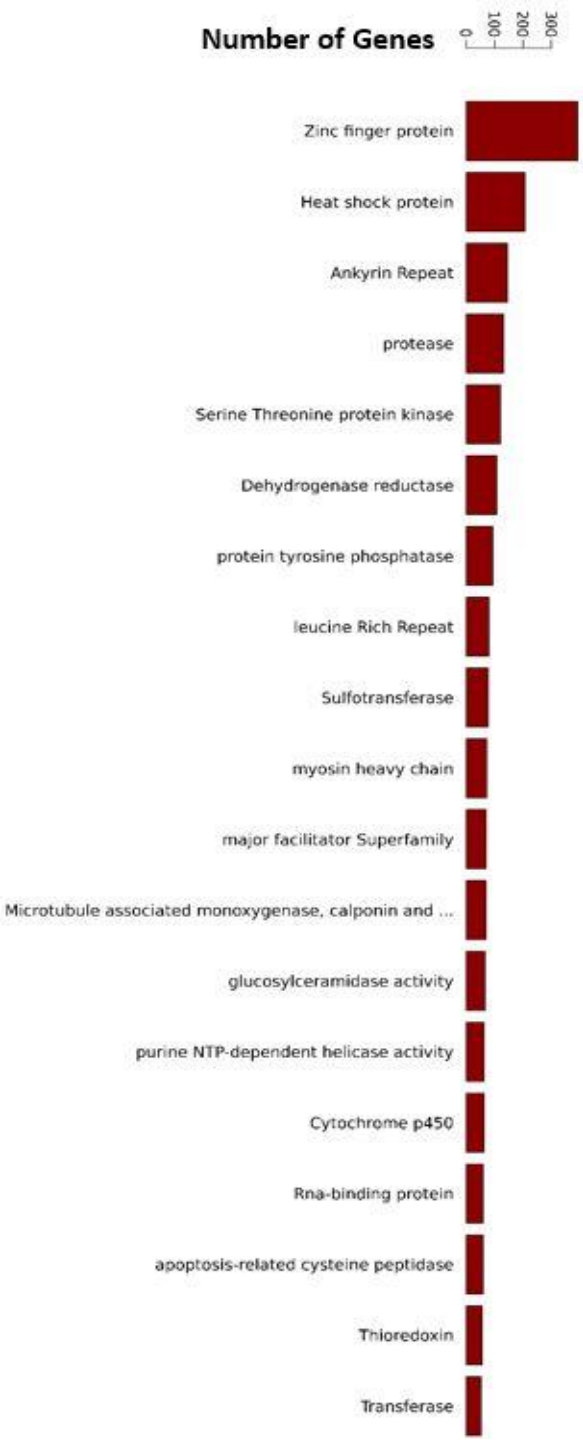

Fig. 3.12: eggNOG annotation plot. Number of transcripts annotated in different gene functions against eggNOG database. Gene classes that showed a hit with less than 10 transcripts in the transcriptome are not shown in the plot.

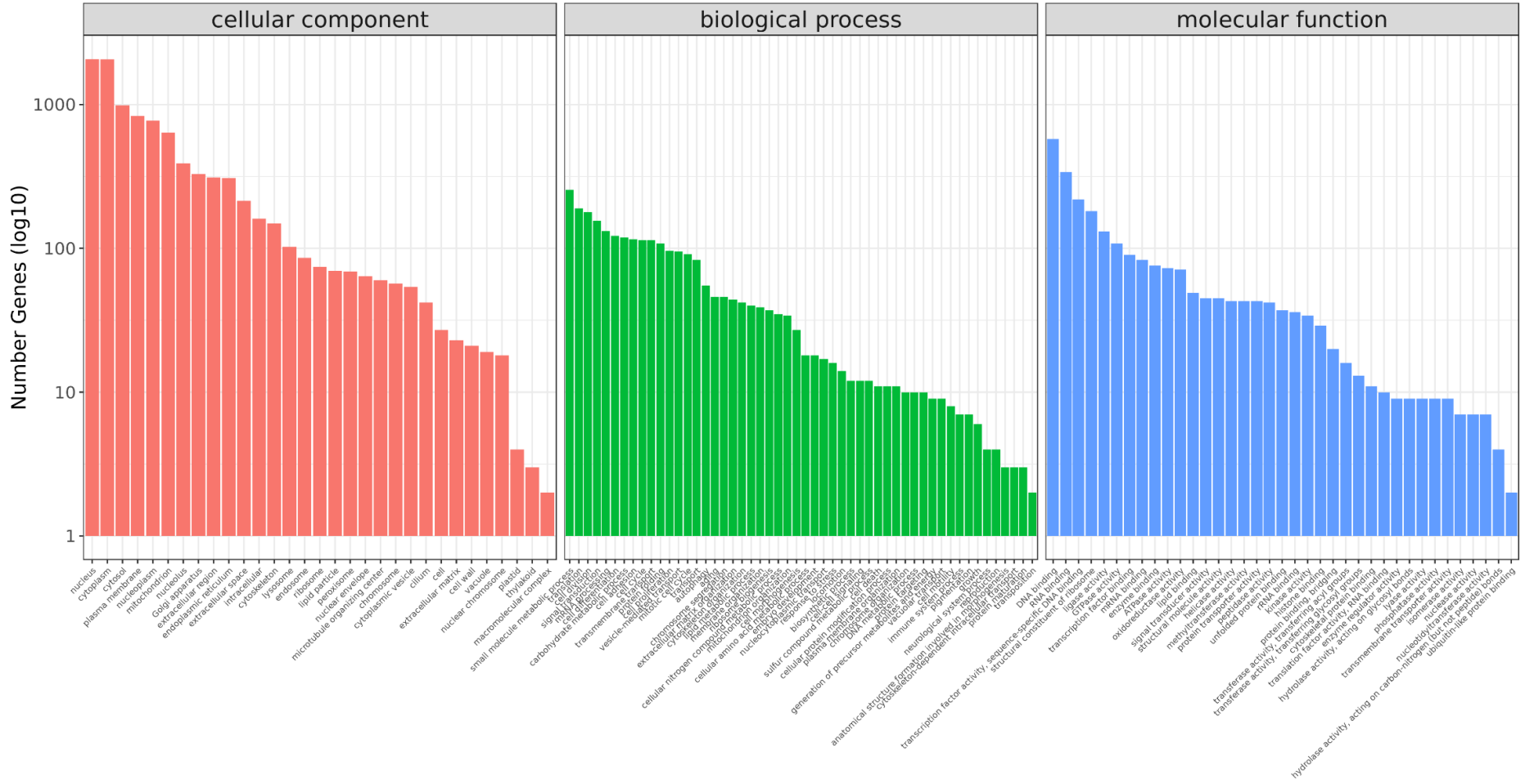

Fig. 3.13: GO terms obtained from a GO-slim analysis conducted on the genes annotated against the UniProt database.

3.9.5 *Gammarus fossarum* sub-species assignment

Previous studies have reported strong genetic differences among populations of *G. fossarum* (Siegismund, 1988; Siegismund et al., 1991, Müller 2000). Although this amphipod has been a subject of controversial debate in taxonomy for decades (e.g., Roux, 1970; Goedmakers, 1972; Jażdżewski, 1977; Goedmakers, 1980; Scheepmaker et al., 1989; Müller, 1998; Westram et al., 2011), to date at least 3 main sub-species have been identified based on nucleotide differences within the mitochondrial gene encoding rRNA 16S (Müller, 2000). The 16S rRNA gene is often used for phylogenetic studies (Weisburg et al., 1991). Whilst highly conserved across all species for the regions responsible for the secondary structure, it also contains a number of hypervariable regions that differ between species. These hypervariable regions act as useful markers for phylogenetic analysis, since more closely related species will have less divergent sequences in these regions (Weisburg et al., 1991). 16S species-specific signatures have been particularly useful for amphipod taxonomy (Müller, 2000; Weiss et al., 2014). An additional taxonomic system in *G. fossarum* based on the differences in the gene encoding cytochrome oxidase 1 (CO1) was proposed by Weiss et al., (2014). In this study, a BLAST analysis of the transcriptome assembly was conducted against the complete *G. fossarum* mitochondrial genome (Macher et al., 2017). Putative mitochondrial transcript sequences from this assembly were aligned against the amphipod NCBI nucleotide database using BLAST in order to identify transcripts corresponding to rRNA 16S and CO1. Among the 20 putative mitochondrial

transcripts, one (TRINITY_DN95725_c7_g2_i2) showed a hit against "*Gammarus fossarum* type A mitochondrial partial 16S rRNA gene, haplotype A14" suggesting that these data are specific to sub-species assignment (*G. fossarum* A) (Müller, 2000) (Tab. 3.5). Additionally, another transcript (TRINITY_DN93353_c3_g3_i1) showed a hit against "*Gammarus fossarum* isolate Gfos_47 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial" and "*Gammarus fossarum* isolate Gfos_45 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial" (Tab.3.5). One other transcript (TRINITY_DN93353_c3_g1_i3) also showed hits against CO1, but from the amphipod taxa *Chiltoniidae*.

Transcript ID	Max score	Total score	Query cover	E-value	% identity	Best hit description	Accession
TRINITY_DN95725_c7_g2_i2 *	725	725	17%	0.0	100%	*G. fossarum* type A mitochondrial partial 16S rRNA gene, haplotype A14	AJ269600.1
TRINITY_DN93353_c3_g3_i1	1110	1110	73%	0.0	100%	*Gammarus fossarum* isolate Gfos_47 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	MG986797.1
	1110	1110	73%	0.0	100%	*Gammarus fossarum* isolate Gfos_45 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	MG986795.1

Tab. 3.5: *Gammarus fossarum* sub-type assignment BLAST parameters. A BLAST analysis of the complete *G. fossarum* mitochondrial genome was conducted against the total assembly, in order to identify the transcripts corresponding to rRNA 16S and CO1 genes. The table shows the BLAST parameters of the best hits obtained when repeating the BLAST analysis for the transcripts coding rRNA 16S and CO1 genes in the NCBI database.

*: Top 2 hits for TRINITY_DN95725_c7_g2_i2 were “Gammarus fossarum mitochondrion, complete genome” and “Gammarus fossarum isolate NEU07 16S ribosomal RNA gene, partial sequence; mitochondrial”. The table shows the third hit, which was useful for the sub-species assignment (G. fossarum A), according to Müller, (2000) classification.

3.10 Discussion

Much has changed in recent years on the awareness of the importance of obtaining detailed molecular information in species of ecotoxicological interest (Simmons et al., 2015). The search for new molecular biomarkers for the evaluation of the status of natural habitats as well as the study of specific pathways affected as a result of anthropogenic activities, has increased exponentially (Pascoe et al., 2003; Atli et al., 2007; Leroy et al., 2010; Sanchez et al., 2011; Brandão et al., 2013; Bahamonde et al., 2014; Schneider et al., 2015; Gismondi et al., 2017; Lebrun et al., 2017; Gouveia et al., 2018). On the other hand, the use of "omics" platforms has greatly increased the depth of the molecular analyses, allowing the acquisition of data on hundreds of thousands of molecules simultaneously. This provides a valuable resource for ecotoxicological research (Simmons et al., 2015). Particular attention has been given to a sub-group of Crustacea, amphipods, due to their sensitivity to aquatic pollutants (Trapp et al., 2015; Wigh et al., 2017) and their central role in the freshwater food web (Dangles et al., 2001). However, despite their importance in ecotoxicology field, the lack of molecular information on these species still represents a limiting factor and a more detailed genomic annotation is fundamental to highlighting homologies and compared to other model organisms. In this study, an RNA sequencing analysis was performed on total RNA extracted from the amphipod *G. fossarum*, using 150 bp paired-end reads sequenced with Illumina NGS technology. Subsequently, a *de novo* assembly strategy was applied, in order to

generate a putative transcriptome assembly using the Trinity software. The assembly produced a total of 680,840 transcripts, clustered into 407,060 genes.

3.10.1 RNA quality

The quality of the RNA samples was evaluated using an Agilent 2100 Bioanalyzer® machine. The electrophoretic gels showed smears below the bands corresponding to 40 sec of migration time, indicating a partial degradation of the extracted RNA (Fig. 3.7). However, the bands at 40 sec migration time appear intense compared to the background and represent a significant portion of intact ribosomal RNA (Fig. 3.7). It is noteworthy that whilst degradation of RNA may bias the estimated relative expression levels, the data obtained using RNA-seq devices based on short-reads sequencing (e.g., Illumina) have been shown to tolerate RNA degradation phenomena in differential gene expression analyses (Romero et al., 2014).

3.10.2 Quality of the transcriptome

Base calling quality was excellent across the entire data set, with the vast majority of reads showing average quality scores greater than 30 (Fig. 3.9) (File S3.1 – Appendix B). The Phred score plots do not indicate any abnormalities concerning the quality of the sequenced reads for all samples (File S3.1 – Appendix B). The Phred scores at the 5'-ends of the reads tended to be lower compared to the central portion, since the sequencing device performs several calibration cycles

during the first sequencing reactions. As expected, quality also drops off towards the 3'-end of the reads, due to incomplete washing between cycles, and is worse for reverse reads. This is a result of the fact that sequencing data for reverse strands are generated after the forward strands, allowing more background signal to accumulate and decreasing the overall base calling quality.

An N50 of 1026 bp was calculated for the assembly performed in this study (Tab. 3.4). This value is actually quite low compared to the values of N50 that can be obtained from a *de novo* genome assembly using the Illumina technology (8-9 x 10^4 bp) (Illumina, 2010), indicating that a large proportion of the transcripts was composed by short fragments that could not be assembled into contigs. On the other hand, a coverage value of 228.25X was obtained. Sequencing generated a total of 325,393,762 paired end reads across the 20 samples. Mapping of reads against a range of different target genomes identified similar mapping rates for each of the 20 samples, with around 50% of reads mapping to either *G. fossarum*, *G. pulex* or *E. marinus*, whilst the remaining reads did not map to any other model organism, indicating no sign of cross-contamination.

Overall, the quality metrics calculated on this assembly (i.e., %GC, longest contig, average contig, number of contigs longer than 10 Kb, N10, N50, number of genes with an identified ORF) were in agreement with a recent study by Cogne et al., (2019) who performed a *de novo* assembly of 7 gammarid taxonomic groups,

including *G. fossarum*. Furthermore, the results of the single-copy orthologs analysis performed on the transcriptome were comparable to the previously mentioned study by Cogne et al., (2019), with 96.4% universal single-copy orthologs identified in a complete form. These data taken together provide evidence of the high quality and completeness of the assembly produced in this study.

3.10.3 Annotation

Annotation against the UniProt/Swissprot database identified candidate hits for only 80,476 (11.8%) of the transcripts. 53.5% of transcripts with an identified open-reading frame (ORF) showed a hit, whilst only 5.6% of transcripts with no ORF showed a hit. The vast majority of non-ORF transcripts (78.7%) were shorter than 500 bp (Fig. 3.10). These data suggest that the majority of non-annotated non-coding transcripts are likely a result of fragmented RNA. Given the lack of molecular information on amphipod species, it is unsurprising to find a total annotation rate below 50% against other amphipod genomes. An analysis of gene ontology terms performed on the total transcriptome showed a wide range of terms, in all three GO categories (Cellular component, Biological Process and Molecular function) (Fig. 3.13). This indicates that the genes identified are involved in a multitude of different functions. No annotation in the UniProt database was identified for almost 90% of the identified transcripts. However, when focusing on transcripts with a complete ORF identified (more likely representing coding

genes), more than 50% of the transcripts showed candidate hits in the UniProt annotation database. In addition, a comparison of transcripts with the eggNOG database identified a wide range of protein classes potentially useful in ecotoxicological research on amphipod species, for example heat shock proteins and enzymes belonging to the cytochrome p450 system (Fig. 3.12). In particular, HSPs have been demonstrated to be involved in crustacean stress responses to a wide range of both biotic and abiotic stress sources, such as pollution exposure (De Pomerai, 1996), thermic stress and microsporidian infections (Grabner et al., 2014). The eggNOG annotation also identified transcripts coding enzymes belonging to cytochrome p450 enzymatic system, which are universally known to be involved in the detoxification effort in a wide range of organisms, including human (Hernández et al., 2013) and ecotoxicologically relevant species, such as fish (Roberts et al., 2005; Ings et al., 2011) and crustaceans (David et al., 2003; Del Brio et al., 2019). Despite a large portion of unannotated transcripts, this dataset will provide a useful resource of genomic information in this poorly annotated species and will represent a reference source for further and more focused molecular analyses on *Gammarus fossarum* as well as other amphipod species. The complete transcriptome dataset was deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession code PRJNA556212.

3.10.4 *Gammarus fossarum* sub-species assignment

A correct sub-type assignment of the amphipod *G. fossarum* is essential. In fact, the genetic differences between the sub-types are considered strong enough to prevent crossbreeding in a natural setting (Müller, 1998, Weiss et al., 2013). Although Pinkster & Scheepmaker (1994) were able to obtain *G. fossarum* F1 juveniles from ex-situ crossbreeding experiments, Müller, the author who differentiated *G. fossarum* into 3 distinct cryptic species (A, B, C) (Müller, 1998; Müller, 2000), found out that sub-types A and B are to be considered reproductively isolate in a natural setting (Müller, 1998). An additional taxonomic system in *G. fossarum*, based on the differences in the faster evolving gene encoding cytochrome oxidase 1 (CO1) was proposed by Weiss et al., (2014). In this study, the sub-type A, haplotype 14 was attributed to *G. fossarum* using a BLAST analysis of the assembly against the complete *G. fossarum* mitochondrial genome (Macher et al., 2017). Although the BLAST analysis did not allow an unambiguous assignment of the CO1 type, it revealed CO1-45 and CO1-47 as most probable CO1 types.

Chapter 4 - Differential Gene Expression Analysis (DGE)

4.1 Introduction

Amphipods represent important model organisms for understanding the effects of exposure to complex chemical combinations in the field. Despite this, there is a clear lack of published genomic information on the subject allowing for a more mechanistic and deterministic approach to environmental toxicology. One of the main aims of the present project was to investigate the gene expression profiles of *Gammarus fossarum* amphipods sampled up- and downstream of a Swiss WWTP, identifying changes in the molecular machinery brought about in response to chemical waste in the water system. Such differentially expressed genes could potentially be useful as molecular biomarkers of xenobiotic exposure in amphipods.

Recent advances in “omics” and high-throughput methodologies have been successfully applied to aquatic species to determine new molecular biomarkers (genes, proteins and metabolites) altered in response to exposure to anthropogenic pollutants in their natural environment (Mortensen et al., 2007, Ings et al., 2011; Bahamonde et al., 2014; Martinović-Weigelt et al., 2014; Poynton et al., 2018). For example, using the expression data obtained from RNA-seq analyses researchers can look at the whole gene expression profile of an organism, as opposed to investigating individual gene pathways (Chapter 3.4). A wide range of studies have focused on evaluating the changes in the gene expression profiles of ecologically relevant species following exposure to single contaminants including endocrine disruptors (Mortensen et al., 2007; Schneider et al., 2015), metals (Hook et al., 2014; Poynton et al., 2018) and pharmaceuticals (Hampel et al., 2010; Mezzelani et al., 2018). The literature also shows various studies that investigated the changes in the gene expression profiles of fish species sampled up- and downstream of sewage effluents (Ings et al., 2011; Bahamonde et al., 2014; Martinović-Weigelt et al., 2014). For instance, it has been shown that endocrine disruptor compounds, such as 17α-ethynylestradiol induce variations in hepatic biotransformation and hormonal response pathways in Atlantic salmon (*Salmo salar*) through alterations in the expression of hormone responsive genes (*Vtg* genes and *Zr-proteins* genes) and detoxification genes (*CYP1A1, CYP3A, GST*) (Mortensen et al., 2007). Transcripts annotated to digestion, growth, moulting and

cytoskeleton components were found altered by Hook et al., (2014) when exposing the amphipod *Melita Plumulosa* to Ni and Zn. Gene expression responses become more complex when exposing aquatic species to whole effluents. Ings et al., (2011) showed altered expression of the genes encoding the heat shock proteins of 70 and 90 kDa and the enzyme of the cytochrome p450 system *CYP1A1* in rainbow trout (*Oncorhynchus mykiss*) caged downstream of a municipal effluent compared to animals caged upstream, used as reference site. The authors also found significant alterations in immune related genes, stress related genes and genes coding hormone receptors, highlighting a wide range of stress responses in exposed fish. The mentioned studies contributed to develop molecular biomarkers of exposure to commonly detected xenobiotics in aquatic environments. Furthermore, gene expression studies conducted on ecologically relevant species highlight that complex molecular interactions occur when animals are exposed to anthropogenic contaminants in mixture (Ings et al., 2011; Martinović-Weigelt et al., 2014). On top of that, concentrations of individual analytes in sewage effluents can vary substantially over time (Nelson et al., 2011), making the biological alterations on the local fauna hard to predict.

Although a DGE analysis conducted on RNA sequencing data allows to investigate the changes in the total gene expression profiles between two or more conditions, the expression data need to be validated using targeted techniques. For instance, measuring the changes in expression of properly selected genes in response to

external stimuli through quantitative polymerase chain reaction (qPCR) remains one of the most reliable approaches (Mehennaoui et al., 2018). qPCR is currently described as one of the most reliable techniques to assess these changes due to its effectiveness, sensitivity and reproducibility (Thornton et al., 2011). qPCR is used to determine the gene expression levels by monitoring the amplification of targeted cDNA molecules during a polymerase chain reaction (PCR), in real time. Starting with RNA samples, a reverse transcription (RT) step to generate cDNA for the subsequent qPCR reaction is performed. The gene expression is evaluated measuring the increase in fluorescence of specific probes or fluorescent compounds added to the reaction mix, during a standard PCR reaction. Although a number of oligonucleotide fluorescent probes are available (e.g., Taq Man©, Beacons, Scorpions, Amplifluor), one of the most widely used compound for the detection of cDNA molecules is the SYBR™ green. This fluorescent probe represents a cost-saving and a well-established option to evaluate gene expression through qPCR. Briefly, the qPCR device detects the fluorescence produced during the amplification process by adding a DNA intercalating dye (SYBR™ green) that fluoresces upon binding to double-stranded DNA. Following the synthesis and binding of the fluorescent compound to DNA synthesized during qPCR, the quantity of amplified DNA and the melting point of the resulting amplicon can be measured (Fig. 4.1). For each evaluated gene, the final output of qPCR is a Ct (Threshold Cycle) value. This value represents how many cycles the fluorescence detector of the qPCR device takes to detect a fluorescence signal significantly

above the background (Fig. 4.2). Therefore, the Ct cycle for a particular gene is inversely related to its expression level within the sample.

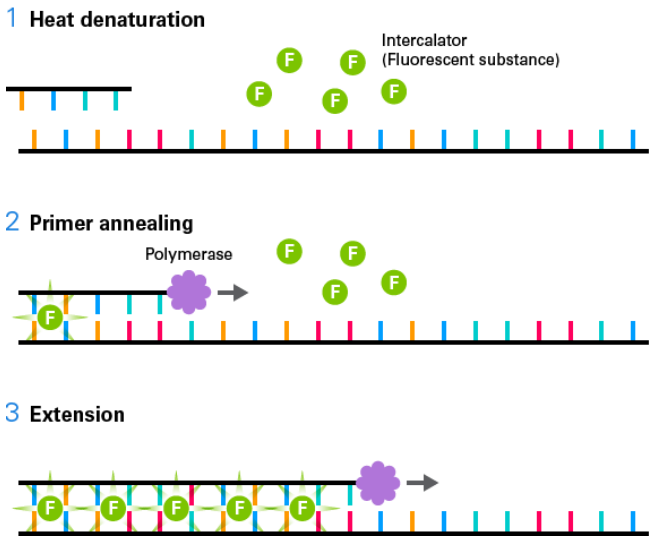

Fig. 4.1: Schematic representation of the SYBRTM green based qPCR method.

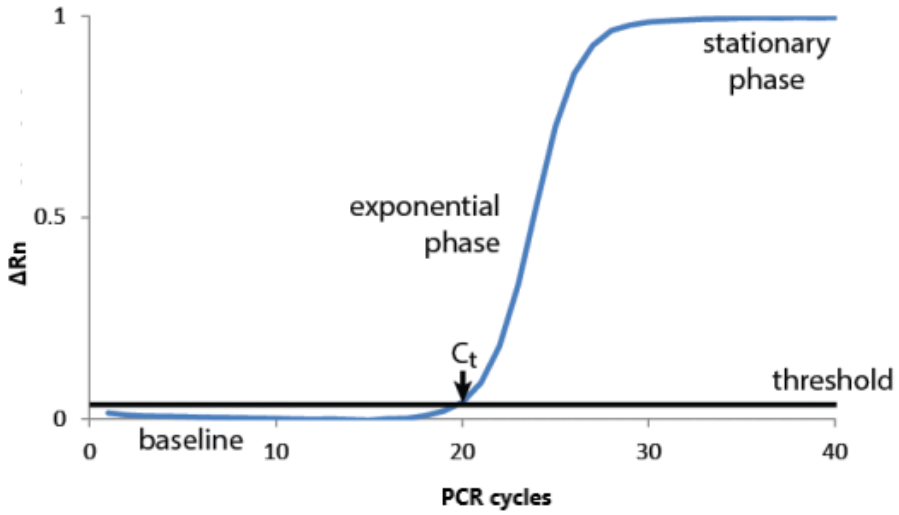

Fig. 4.2: Graphic representation of the Ct value in a qPCR amplification plot. Rn represents the fluorescent signal from SYBR™ green normalized to the signal of the passive reference dye for a given reaction. The ΔRn is the Rn value of an experimental reaction minus the Rn value of the baseline signal generated by the instrument.

4.2 Aim and objectives

Following the phenotypical observations and the evaluation of the amphipod population structures (Chapter 2), the analysis described in this chapter was aimed to evaluate the impact of the toxic pressure downstream of a WWTP on amphipod gene expression profiles. The changes in gene expression between *G. fossarum* sampled downstream of the WWTP and upstream, used as reference site, were explored. In particular, the complete transcriptome of *G. fossarum* was assembled and annotated (Chapter 3) and the read counts were loaded in the programming language R for the differential analyses. A first differential analysis between up- and downstream samples was conducted on both male and female amphipods to evaluate potential differences between the genders in the response to the toxic

pressure in the stream. In addition, in order to generate a data set useful to find new sex-specific biomarkers in crustaceans, a second DGE analysis was performed between male and female amphipods. To experimentally validate the results obtained from the DGE analysis, a qPCR experiment was set up. Based on bioinformatics parameters, 5 transcripts were chosen from the list of changing transcripts between up- and downstream in *G. fossarum* males and 5 transcripts were chosen from the female list. To conduct a relative quantification of the chosen transcripts between upstream and downstream populations, 2 reference genes were also selected.

4.3 Methods

4.3.1 Collaborations and contributions (DGE analysis)

In field amphipod sampling was my own work in collaboration with Dr Andrea Schifferli and Dr Thomas Bucher (Swiss Centre for Applied Ecotoxicology, Dübendorf, Switzerland). Transcript filtering, mapping of the reads and differential gene expression analysis were conducted by Dr Samuel Robson (University of Portsmouth, St Michael's Building, Portsmouth, UK). Panther functional investigation and Gene Ontology analysis of the differentially expressed genes were my own work. RNA extractions, assessment of RNA quantity and quality,

primer design and qPCR data acquisition were my own work in collaboration with Dr Robin Rumney and (University of Portsmouth).

4.3.2 Amphipod sampling

Sampling procedures of the amphipods used for the transcriptomic analysis are described in Chapter 2.5.2. Due to a lack of RNA remaining from the initial transcriptome sampling, additional sampling was required to generate RNA for qPCR validation. To ensure that the results of the qPCR analysis were biologically comparable with the results obtained from the DGE analysis, the animals used for this experiment were sampled in the same season and at the same geographical coordinates (Tab. 2.2) used for the sequencing experiment, albeit 2 years apart (mid-September 2019). 10 animals were sampled below the WWTP and 10 animals were sampled above the WWTP, for a total of 20 amphipods (5 male and 5 female biological replicates per sampling site). The sex determination was conducted evaluating the presence of genital papillae in males and brood plates in females (Chapter 1.5.2). Sampling and animal handling procedures prior to RNA extractions for the qPCR experiment were analogous to the RNA-seq experiment (Chapter 2.5.2). Each amphipod was cut in 3-4 pieces using a scalpel. In preparation for the RNA extractions, the sections belonging to each animal were placed in 500 µL of a TRIzol™ solution (Thermo Fisher Scientific, Bremen, Germany). The tubes were placed on dry ice and sent to the Institute of Marine Sciences (Portsmouth, UK).

4.3.3 Transcript filtering

A filtering process was applied to the complete transcript set (File S4.1 – Appendix B) prior to DGE analysis, in order to filter out low-abundance transcripts, transcripts that showed evidence of representing contamination from non-eukaryotic sources, and transcripts unlikely to represent coding genes of interest (File S4.2 – Appendix B). Transcripts having the following features were excluded from the differential gene lists:

- transcripts annotated against genes belonging to bacteria, archaea or viruses
- transcripts with no complete open reading frame (ORF) detected
- transcripts with a maximum FPKM (Fragments Per Kilobase per Million) value across all samples less than 1
- transcripts less than 500 bp in length

Given the nature of transcriptome assembly, there may be redundancy present in the assembly, with multiple distinct transcripts representing the same gene, due to the presence of multiple isoforms or of fragmented RNAs. However, for the sake of brevity, the term “gene” will be used from here on to describe the transcripts present in the filtered assembly.

4.3.4 Differential Gene Expression Analysis

A differential gene expression analysis was conducted using the DESeq2 package in R, using the group (Male Upstream, Male Downstream, Female Upstream and

Female Downstream) as the independent variable. P-values were corrected for multiple testing by using the Benjamini-Hochberg FDR correction. Differentially expressed transcripts were identified between all pairs of groups by specifying the contrasts. Transcripts were identified as differentially expressed if they showed a fold-change greater than 2-fold (either up or down) between groups with an adjusted p-value ≤ 0.05. In addition, transcripts with a mean FPKM value across all samples below 1 in both groups of the evaluated comparisons ("Downstream vs Upstream" and "Females vs Males") were not included in the differential expression analysis, in order to avoid inflated fold-changes between low-abundance transcripts.

4.3.5 Functional analysis

4.3.5.1 Overview on Gene Ontology

Given the large number of genes detected by DGE analyses performed on RNA-seq data, exploring their function studying the pathways corresponding to single genes would be a very long task. An effective strategy to conduct functional analyses on large sets of genes is represented by the gene ontology (GO) functional enrichment. The Gene Ontology (GO) is a major bioinformatics initiative to unify the representation of gene and gene product attributes across all species (GO Consortium, 2008). The project was originally conceived in 1988 with the following aims (Ashburner et al., 2000; Dessimoz et al., 2017):

- maintain and develop a constantly updated vocabulary of gene and gene product attributes
- annotate genes and gene products, and assimilate and disseminate annotation data
- provide tools for the functional interpretation of experimental data using the GO vocabulary (e.g., enrichment analysis).

In general, two types of functional analyses can be conducted exploring the gene ontologies: GO-complete and GO-slim analyses. GO-complete terms refer to the original full set of annotations maintained by GO consortium, in a branched system with a certain degree of redundancy (i.e., the same gene can be associated with different terms or sub-terms). On the other hand, GO-slim terms are more generic and represent an ideal choice to conduct a preliminary functional analysis on a given set of genes, minimising the redundancy.

4.3.5.2 Enrichment analysis

In order to perform functional enrichment analyses on large sets of genes it is possible to submit the gene names in the GO website (http://geneontology.org/) or in the Panther database (Mi et al., 2005), which includes functional tools relying on the official gene ontologies. Following the submission of the gene names, the user choses the GO category to explore (Biological process, Molecular function or Cellular component) and selects the gene ontology of a given species (e.g., Human, *Drosophila melanogaster*, *Saccharomyces cerevisiae*). The output of the analysis is

a set of biological processes, molecular functions or cellular components (i.e., GO-terms) shared by the submitted genes. The rate and the statistical significance of the enrichments are described by fold-enrichment and FDR values, respectively. The fold-enrichment represents the "magnitude" of the enrichment and can be described as ratio of the percentage of genes enriched in a particular GO-term within the submitted gene set over the number of genes associated with that GO-term in the genome of the species explored. In order to calculate the significance of the enrichments, p-values are also calculated. The p-value represents the probability of having an over-representation of the submitted genes in a specific GO-term compared to the probability of an enrichment using a random set of genes. For instance, if a significance level of 0.05 is set, there is a 5% probability that the enrichments with a p-value<0.05 are false-positives. Because multiple enrichment tests are performed in a GO analysis, p-values need to be corrected for multiple testing (Benjamini-Hochberg FDR correction), thus FDR values for each GO-term are calculated and represent the statistical significance of the enrichments.

4.3.6 Gene Ontology analysis

In order to functionally classify the differential transcripts detected in both "Downstream vs Upstream" and "Females vs Males" comparisons, gene ontology (GO) analyses were conducted using the database Panther (Mi et al., 2005). To avoid redundancy of GO-terms, these preliminary analyses were focused on the

GO-slim sets of ontologies, exploring “Biological process”, “Molecular function” and “Cellular component” categories. Gene names annotated against UniProt were submitted and, consistently with the functional analysis performed on the differential metabolites (Chapter 5.6.3), *Drosophila melanogaster* gene database was interrogated. In addition, Panther was also employed to conduct GO-complete analyses on the differential transcripts annotated against UniProt for both comparisons. The “Statistical over-representation” tool of Panther database was used, selecting Fisher’s exact test to calculate the statistical significance of the enrichments. GO terms enrichments with an FDR<0.05 were considered as statistically significant and are shown in the Files S4.3-4.11 (Appendix B). An additional GO-complete “Biological process” analysis was conducted on the overlapping genes between “Downstream female vs Downstream male” and “Upstream female vs Upstream male” comparisons (File S4.12&4.13 – Appendix B). This subset of genes was of interest because it contained general sex biomarkers, excluding the influence of the sampling site.

4.3.7 RNA extraction

For the RNA-seq analysis, fresh amphipods were dissected and the total RNA was extracted from the internal tissues (Chapter 3.8.3). On the other hand, the animals for the qPCR validation were sent from Switzerland on dry ice. Due to a technical difficulty in dissecting thawed animals, a different protocol for tissue disruption/homogenization was used for the samples used for the qPCR

experiment. The amphipods were subjected to a 2 min exposure to ultrasounds in a VWr water bath (VWr – USC 300 T, Lutterworthfor, UK) for tissue lysis. The samples were centrifuged at 15000 rpm for 15 min at 4 ^{o}C, in order to precipitate both cell and exoskeleton debris. The supernatants were separated from the precipitated biological debris and the total RNA was extracted using an RNeasy® Plus Universal Mini Kit (Qiagen, Manchester, UK), following manufacturer's instructions.

4.3.8 Assessment of RNA quantity and purity

1 µL of each RNA sample was loaded on a NanoDrop™ ND 1000 spectrophotometer (Thermo Fisher Scientific, Altrincham, UK) to quantify the RNA. The RNA samples concentration expressed in ng/µL as well as the absorbance at 230 nm, 260 nm and 280 nm were recorded (Tab. 8.2 – Appendix A). 260/280 and 260/230 ratios were also recorded, to verify the RNA purity.

4.3.9 RNA quality assessment

5 µL of each RNA sample, corresponding to 500 ng of RNA, were used for an RNA quality assessment through electrophoresis on agarose gel. A 1:100 (w/v) dilution of agarose (Sigma, Gillingham, UK) in 1X TAE (Tris-acetate-EDTA) buffer was prepared. An aliquot of a 10mg/mL ethidium bromide solution (Invitrogen, Carlsbad, US) was added to the agarose solution in a 1:20000 (v/v) ratio for gel

staining. The RNA samples and 10 µL of a 100 bp DNA ladder were loaded on the gel. The voltage for the electrophoretic run was set to 25 V.

4.3.10 Mapping and quality Control

Transcripts were generated by combining the read data from all samples and using a *de novo* assembly approach (Chapter 3). Following transcriptome assembly, transcript abundance of male and female samples collected up- and downstream of the WWTP was calculated for individual replicates using the software Kallisto (Bray et al., 2016) and expressed in FPKM (Fragments Per Kilobase per Million) (File S4.14 – Appendix B). This value represents an estimation of gene expression normalised for the gene length and library size (i.e., sequencing depth) of the sample (eq. 4.1).

Equation 4.1

$$FPKM = \frac{N}{\left(\frac{L}{10^3}\right)\left(\frac{T}{10^6}\right)}$$

N is the number of reads mapped to a gene, *L* represents the gene length and *T* is total number of mapped reads of the sample. Because the RNA-seq analysis performed in this project was a paired-end sequencing (Chapter 3.8.7), the algorithm takes into account that two reads can map to one fragment, thus it will count that fragment once. The FPKM values corresponding to all assembled transcripts for each of the 20 RNA samples in the data set (File S4.14 – Appendix

B) were loaded into the statistical programming language R (RC Team, 2015) for the analysis. The package DESeq2 (Love et al., 2014) was used to prepare the read counts for differential expression analysis, by generating variance stabilised log-scaled read values. These were used for data visualisation to avoid the effects of low abundance transcripts on the clustering. Cluster plots were built based on Euclidean distance (eq. 4.2) between the transformed read values for all 680,840 transcripts across the 20 samples and for the top 500 transcripts with the highest variance across the samples.

Equation 4.2

$$Dab = \sqrt{\sum_{c} (eas - ebs)^2}$$

D_{ab} is the distance between expression patterns for genes *a* and *b*, e_{as} is the expression level of gene *a* in sample *s* and e_{bs} is the expression level of gene *b* in the same sample. The equation is expressed for the example genes *a* and *b*, but includes the sum of the Euclidean distances of all the genes obtained in the assembly, for each sample. A principal component analysis (PCA) was also performed using the top 500 changing transcripts (based on the variance across samples).

4.3.11 Genes for qPCR validation

A total of 10 significantly changing genes between up- and downstream populations (5 from the male list and 5 from the female list) (Tab. 4.1&4.2) detected in the DGE analysis were chosen for a qPCR validation. Genes of interest were prioritised for validation based on the fold-change values calculated in the DGE analysis. Overall, 10 annotated genes with an absolute log_2-transformed fold-change greater than 20 between upstream and downstream populations, annotated against the UniProt database were selected.

Transcript ID	Gene Name	Gene Description	log_2(FC)
TRINITY_DN93483_c0_g1_i7	*CP2L1*	Cytochrome P450 2L1	-25.99
TRINITY_DN102258_c0_g1_i5	*LDAH*	Lipid droplet-associated hydrolase	37.27
TRINITY_DN83373_c0_g1_i1	*H90A1*	Heat shock protein HSP 90-alpha 1	31.10
TRINITY_DN108005_c11_g5_i2	*MYP2*	Myelin P2 protein	-34.96
TRINITY_DN107108_c0_g2_i4	*DHSD*	Succinate dehydrogenase [ubiquinone] cytochrome b small subunit, mitochondrial	22.28

Tab. 4.1: Details about the changing genes in the "Downstream vs Upstream" comparison selected from the male list for a qPCR validation. "Gene Name" and "Gene Description" refer to the annotation conducted against UniProt database.

Transcript ID	Gene Name	Gene Description	log_2(FC)
TRINITY_DN114991_c1_g1_i2	*ODO1*	2-oxoglutarate dehydrogenase, mitochondrial	22.56
TRINITY_DN112167_c0_g1_i11	*AMPN*	Aminopeptidase N	24.64
TRINITY_DN112074_c0_g3_i2	*SLO*	Calcium-activated potassium channel slowpoke	-22.37
TRINITY_DN103612_c0_g1_i1	*NPAB*	Neuroparsin-A	-20.60
TRINITY_DN103329_c0_g3_i7	*ZC3HE*	Zinc finger CCCH domain-containing protein 14	21.81

Tab. 4.2: Details about the changing genes in the "Downstream vs Upstream" comparison selected from the female list for a qPCR validation. "Gene Name" and "Gene Description" refer to the annotation conducted against UniProt database.

2 housekeeping genes were also selected to act as a qPCR data normalization (Tab. 4.3). Based on the RNA sequencing data, *CYCG* was among the genes with the lowest variance across the samples, therefore potentially suitable to work as reference gene in a qPCR experiment. In order to test one of the most stable reference genes to be used in a qPCR experiment using *G. fossarum* according to Mehennaoui et al., (2018), an additional housekeeping gene, such as *GPADH* was also selected (Tab. 4.3).

Transcript ID	Gene Name	Gene Description	Average FPKM across all samples	σ
TRINITY_DN109602_c0_g1_i3	*CYCG*	Cyclin G	8.66	2.77
TRINITY_DN101022_c0_g1_i4	*GAPDH*	Glyceraldehyde-3-P dehydrogenase	509.6	172.23

Tab. 4.3: Details about the reference genes chosen for the qPCR experiment. "Gene Name" and "Gene Description" refer to the annotation conducted against UniProt database.

4.3.11.1 Primer Design

Primer3Plus online tool (Untergasser et al., 2012) was employed for the design of the primers for both the genes to validate and the housekeeping genes. Considering that *G. fossarum* is still a poorly annotated species, a study on the structures of the chosen genes aimed to design exon-exon primers was not possible. Therefore, the whole transcript nucleotide sequences corresponding to the chosen genes were loaded in the online software and the following parameters were set:

- PCR product size: 150-220 nucleotides
- Primer T_m (Melting Temperature): 59-61 °C with an optimum of 60 °C
- Primer GC content: 50-60%
- Primer size: 19-21 nucleotides with an optimum of 20 nucleotides

One primer set for each gene was selected. The oligonucleotides were synthetized by Eurofins Genomics (Abingdon, UK). Tables 4.4, 4.5 and 4.6 show the primer sequences and the details for each oligonucleotide.

Gene	Orientation	Sequence (5' –> 3')	Size	T_m (°C)	Product size (bp)	GC – content (%)
CP2L1	F	GAGACTTCATCGACGCCTTC	20	59.4	191	55
CP2L1	R	CTGTATTTTGGCCTGCACCT	20	57.3		50
LDAH	F	CAGGAAGTCAGTGTGGAGCA	20	59.4	191	55
LDAH	R	GGAGGTAGCAGCTGATGGAG	20	61.4		60
H90A1	F	ACATCTAGAGGAGCGCCGTA	20	59.4	158	50
H90A1	R	TGGGTTTATCTTCGGACTCG	20	57.3		55
MYP2	F	GGTGCAGAAGGCTAGCAAAG	20	59.4	150	55
MYP2	R	CTGAAGGGACCATGAAAGGA	20	57.3		50
DHSD	F	CTTCTGGCCTTATCGCTCAC	20	59.4	209	55
DHSD	R	AGCGTCCAGAGCATTGAGAT	20	57.3		50

Tab. 4.4: Details about the primers designed for the amplification of the male genes.

Gene	Orientation	Sequence (5' –> 3')	Size	T_m (°C)	Product size (bp)	GC – content (%)
ODO1	F	GTGACCCACGGCTAAGGATA	20	59.4	159	55
ODO1	R	AGATGTCCAGCCAGAGAGGA	20	59.4		55

AMPN	F	ATCAGAGAAGGCGGTGAGAA	20	57.3	177	50
AMPN	R	GTCCTGCTTCCTCACTCCAG	20	61.4		60
SLO	F	GTTGTCGAGGAGGATGTGGT	20	59.4	182	55
SLO	R	ATCCTGATTGTCCCAACGTC	20	57.3		50
NPAB	F	GCACTCACCCAATCACTCCT	20	59.4	172	55
NPAB	R	CTGGTTCCGGCAGAATATGT	20	57.3		50
ZC3HE	F	GCAAGTGATGAGTTGGAGCA	20	57.3	205	50
ZC3HE	R	AGTGTTGTGGAGGACCAAGG	20	59.4		55

Tab. 4.5: Details about the primers designed for the amplification of the female genes.

Gene	Orientation	Sequence (5' –> 3')	Size	T_m (°C)	Product size (bp)	GC – content (%)
CYCG	F	GATGATGCTGGTGGATGATG	20	57.3	176	50
CYCG	R	AGATAGCGTTGGAGCCTGAA	20	57.3		50
GAPDH	F	ACCAGCACCCCTTTTTCTCT	20	57.3	178	50
GAPDH	R	CTGTGCAGGTCAAATCGAGA	20	57.3		50

Tab. 4.6: Details about the primers designed for the amplification of the housekeeping genes.

Each lyophilized primer sample was reconstituted in the appropriate volume of Nuclease-free water (Qiagen, Manchester, UK) to obtain a concentration of 100pmol/µL. In order to make the primers suitable for the cDNA synthesis, PCR

and qPCR protocols, a 1:10 dilution in Nuclease-free water was performed for each primer.

4.3.11.2 DNase I treatment

To remove any genomic DNA (gDNA) contamination, a DNase I protocol was performed on the RNA extracts. A Precision™ DNase kit (Primer Design, Chandler's Ford, UK) was used following manufacturer's instructions. 1,5 µg of RNA for each sample were used for the protocol, except for the samples US♀2 and US♀5 for which the nucleic acid yields were not sufficient to use 1,5 µg of RNA. For these samples, 1 µg was used instead. The sample DS♀3 was excluded because of a low RNA yield (Tab. 8.2 – Appendix A).

4.3.11.3 cDNA synthesis

Following the DNase treatment, an aliquot of 500 ng of RNA for each sample was used for the cDNA synthesis. A SuperScript VILO cDNA Synthesis kit (Invitrogen, Renfrew, UK) was employed to synthetize first-strand cDNA in 20 µL total reaction volume for each sample, following manufacturer's instructions. For each sample, 2 PCR reaction mixes were prepared:

1) a standard PCR reaction mix

2) a negative control lacking the reverse transcriptase enzyme (-RT)

4.3.11.4 Primer specificity test

One cDNA male sample (US♂1) and one cDNA female sample (US♀1) were used as templates to test the specificity of each primer set. A PCR experiment using 1 µL of cDNA for each sample and 1 µL of each primer in a 25 µL of total reaction volume was set up. A GoTaq Green Master Mix (Promega, Madison, US) was employed to perform the PCR amplifications. After an initial denaturation step at 95 °C for 2 min, 35 amplification cycles were performed with the following parameters:

- Denaturation: 95 °C – 1 min
- Annealing: 60 °C – 1 min
- Extension: 72 °C – 1 min

A final extension step at 72 °C for 5 min was set. To visualise the amplification products, an electrophoresis on a 1% agarose gel was carried out. The gel was prepared using the same reagents and the same concentrations as the gel employed for the RNA quality assessment (section 4.3.8). 10 µL of each PCR reaction were loaded on the gel and 10 µL of a 100 bp DNA ladder were also loaded to check the seizes of the PCR products. A voltage of 100 V was set for the electrophoresis run and a gel picture was taken by exposing the gel to a UV light source (Fig. 4.19).

4.3.11.5 cDNA quality control

To verify the functioning of the DNase I protocol previously performed on the RNA samples, a PCR experiment using the primer set for the reference gene *CYCG* and the cDNA samples as templates was set up. The same PCR methods as the primer specificity test were employed (section 4.3.10.5). For the visualization of the PCR products, 10 µL of each sample were loaded on a 1% agarose gel.

4.3.11.6 Plate preparation and qPCR

Two 96-well plates were used for a qPCR experiment (the first one for male samples and the second one for female samples) employing a Viia 7 Real-Time PCR instrument (Thermo Fisher Scientific, Altrincham, UK). Tab. 4.7 shows the components loaded in each well. The PCR method is described in Fig. 4.3.

Component	Volume
Precision™ 2X qPCR Mastermix (Primer design, Chandlers Ford, UK) premixed with SYBR green	9 µL
Forward primer	0.5 µL
Reverse primer	0.5 µL
Template	1 µL
Nuclease-free water	7 µL
Final volume	**18 µL**

Tab. 4.7: Components of the qPCR reaction mixes. 25 ng of cDNA (template) for each sample were loaded in each well. The primers were reconstituted in a volume of Nuclease-free water to obtain a concentration of 100pmol/µL and then diluted 1:10 before adding them to the reaction mixes.

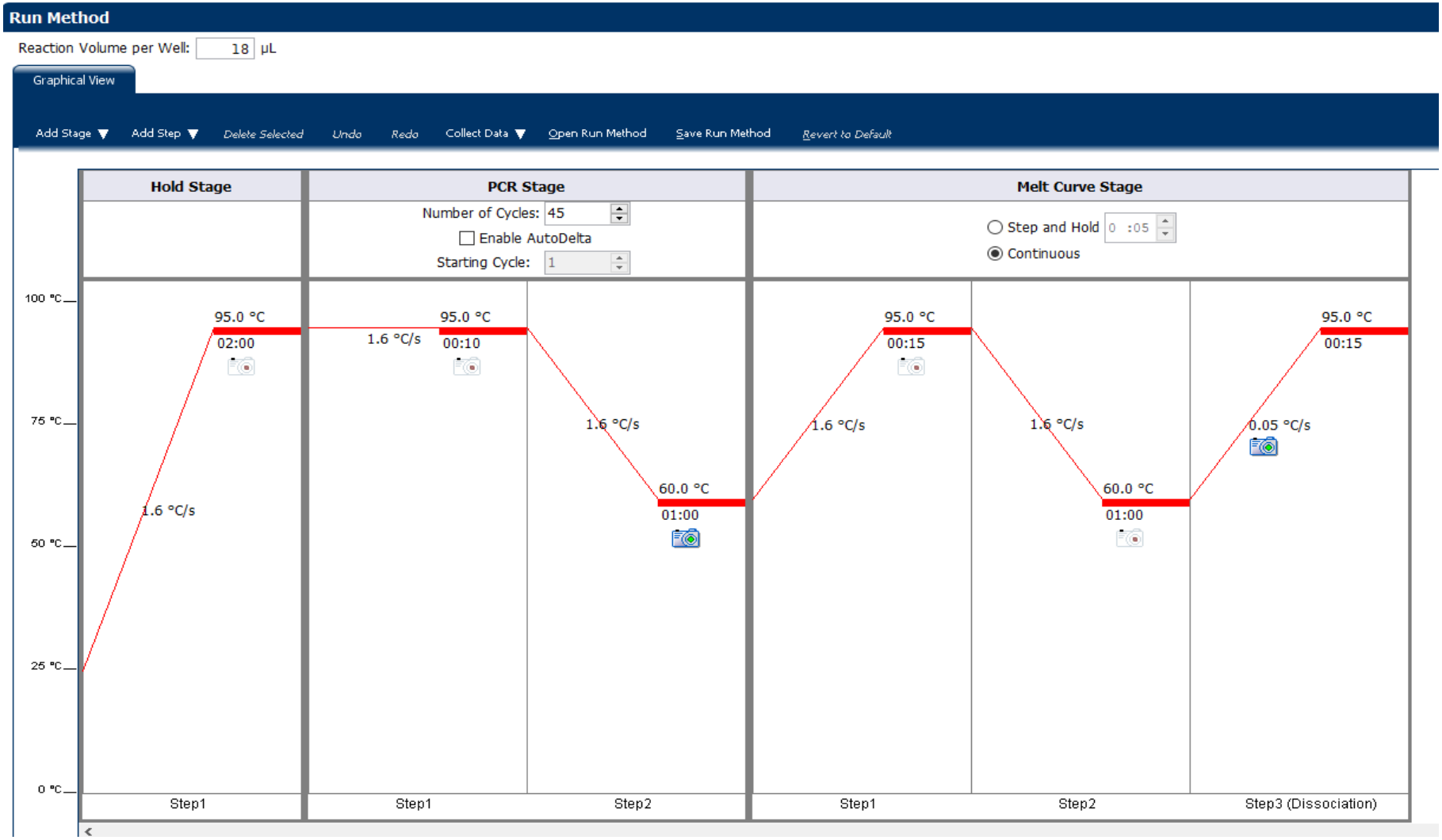

Fig. 4.3: qPCR method. The PCR method used for the qPCR experiment was recommended by Primer Design when employing SYBR™ green as fluorescent probe.

An evaluation of the melt curves and the amplification plots were conducted to confirm the specificity of the PCR products in each well. Nuclease-free water was used in place of template in the negative control reactions. Furthermore, -RT controls for each sample were loaded on the plates to check the presence of residual gDNA.

4.3.11.7 qPCR data analysis

Because qPCR is a much more sensitive technique for the detection of nucleic acid compared to electrophoresis, the qPCR data were used to identify samples that contained residual gDNA. In fact, the presence of bands in lanes corresponding to -RT controls in the cDNA gels may be due to spill over/diffusion phenomena occurred during gel loading. Biological replicates that showed Ct values in the corresponding -RT controls less than 30 for at least one of the evaluated genes were excluded from the data analysis. A relative quantification for each gene of interest was conducted between upstream (reference) and downstream (test) samples. In each replicate, the Ct values for the genes of interest were normalised with the Ct values of the housekeeping gene (*CYCG*), calculating ΔCt values. The normality of the ΔCt values for each group was tested using a Shapiro-Wilk test. Mann-Whitney U-tests were performed to evaluate the statistical significance of the differences between the ΔCt values of upstream and downstream samples, using a significance level of $p<0.05$. The results are expressed as the mean ±

standard error ($\sigma/\sqrt{n}$). Relative expression values for each gene of interest were calculated using the ΔΔCt method (Schmittgen et al., 2008). Specifically, overall ΔΔCt values for each examined gene were calculated as the difference of the average ΔCt values of the downstream replicates (test group) minus the average ΔCt values of the downstream replicates (control group). Overall fold-changes were expressed as $2^{-\Delta\Delta Ct}$.

4.4 Results

4.4.1 Assessment of RNA quantity and purity

A NanoDrop™ ND 1000 spectrophotometer was used to measure the concentrations of the RNA samples and the absorbance at 230 nm, 260 nm and 280 nm. Several samples showed a low 260/230 ratio, indicating residual phenol or carbohydrates (Thermo Fisher Scientific, 2008). However, all samples showed 260/280 ratios close to 2, indicating that the extraction protocol successfully filtered out most of protein component, with an enrichment in nucleic acid (Thermo Fisher Scientific, 2008) (Tab. 8.2 – Appendix A).

4.4.2 Clustering

Two plots were built for the visualization of the gene expression trends across the samples using the pheatmap package in R (Kolde, 2019). A first heat map was built including the entire set of assembled transcripts (680,840) (Fig. 4.4) and a second heat map was built considering the 500 transcripts with the highest variance across the samples in the data set (Fig. 4.5). A darker colour represents a lower Euclidean distance (eq. 4.2) between the samples, indicating that the samples are more similar in their overall transcript expression profiles.

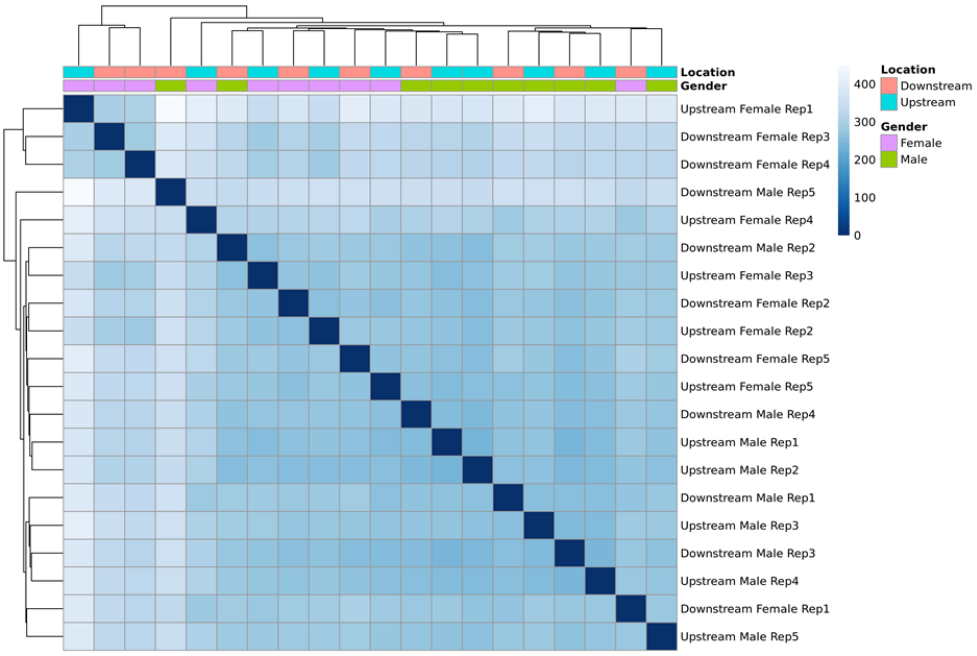

Fig. 4.4: Heat map built including the entire set of transcripts. The overall similarity in expression across all transcripts between the samples is represented by a scale from blue (highest) to white (lowest). Samples are clustered such that more similar samples are closer together, allowing the visualization of similar sample groups. Sampling site (Upstream or Downstream) and Gender (Male or Female) are annotated above the heatmap.

Overall, the samples did not show a distinct clustering within their specific groups – neither up- and downstream nor male and female groups. An exception is the clustering of three female samples (“Upstream Female Rep1”, “Downstream Female Rep3” and “Downstream Female Rep4”) which appear distinct from all other samples.

The second plot was built including 500 transcripts with the highest variance in the data set (Fig. 4.5). Similarly to the first cluster plot (Fig. 4.4), the samples did not show a clear clustering based on their grouping. A distinction between a small subset of female samples distinct from the remaining samples (Fig. 4.5 - top left), and a cluster between two female samples (“Downstream Female Rep1” and “Upstream Female Rep4”) and the male sample “Downstream male Rep5” are evident. These three samples may represent outliers of the analysis. In fact, no anomalies neither in the read quality control nor in the RNA quality control were found for these samples. Although a large male cluster in the middle can be observed, it also contains the female samples “Upstream female Rep5” and “Downstream female Rep5”. Therefore, similarly to the heat map built including the entire set of transcripts, a discrete gene expression clustering based on the gender or the sampling site was not observed.

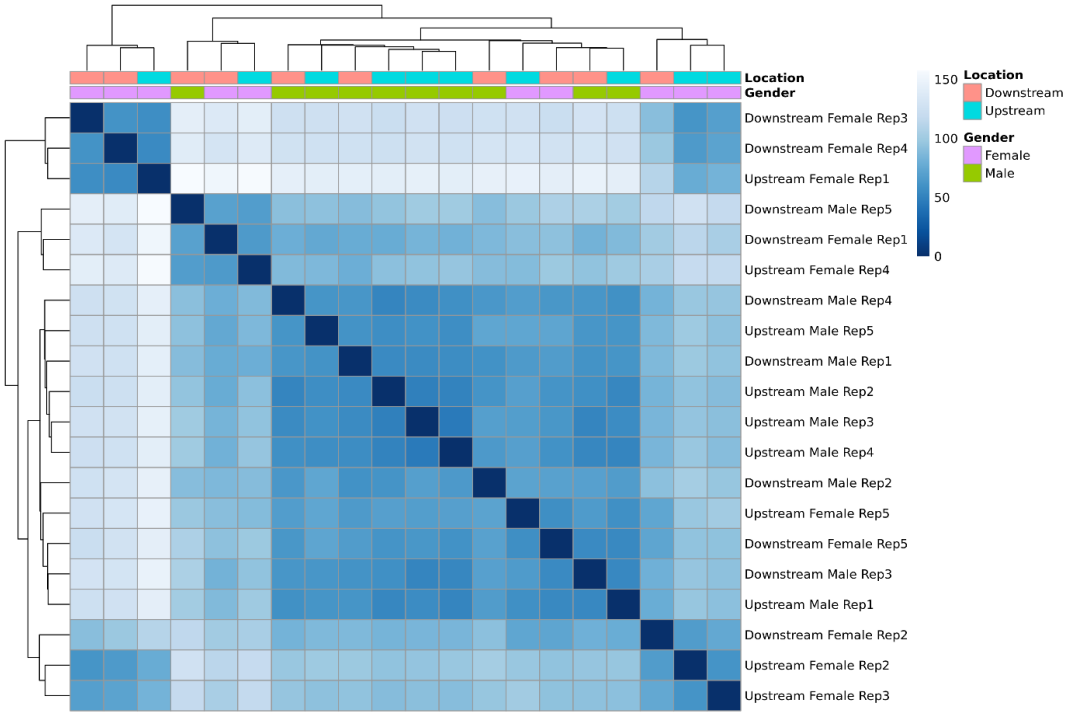

Fig. 4.5: Heat map built including the top 500 changing transcripts. The overall similarity in expression across all transcripts between the samples is represented by a scale from blue (highest) to white (lowest). Samples are clustered such that more similar samples are closer together, allowing the identification of similar sample groups. Sampling site (Upstream or Downstream) and Gender (Male or Female) are annotated above the heatmap.

4.4.3 PCA

A PCA analysis including the first (PC1) and the second (PC2) greatest sources of variation over the top 500 transcripts with the highest variance across the samples in the data set was performed (Fig. 4.6). A cluster of male samples (Fig. 4.6 – top left) is clear. On the other hand, female samples appear much more disperse with the greatest source of variation in the data (PC1 = 48.32%) separating the males from a subset of female samples. The second greatest source of variation in the

data (PC2 = 15.01%) separates a small group of 3 samples (including 2 female samples and a male outlier) (Fig. 4.6 – bottom left) from the remaining samples. However, there does not appear to be a great amount of variation separating male upstream and downstream samples.

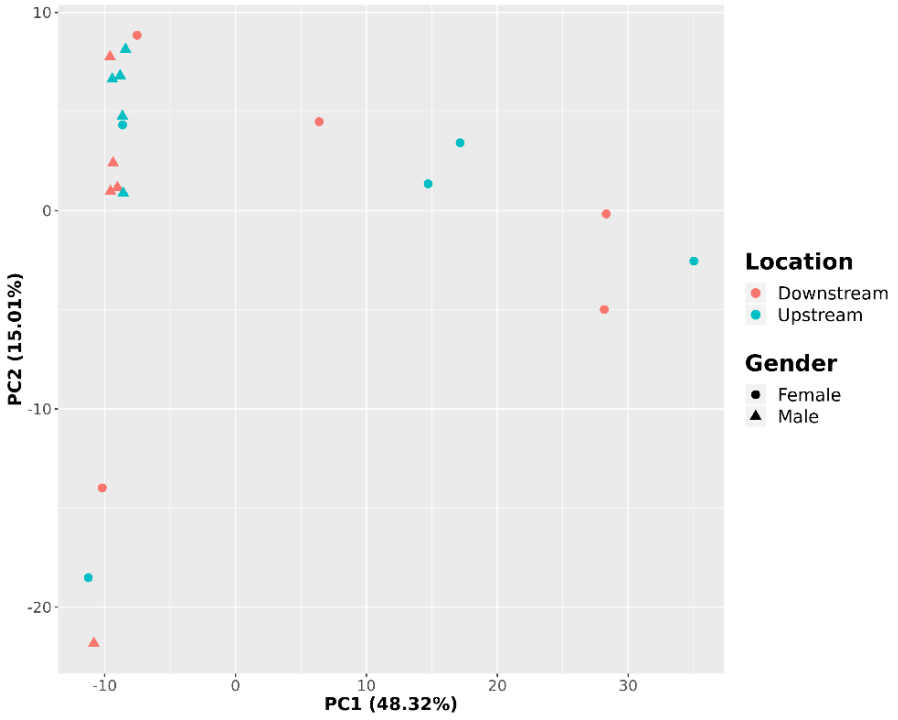

Fig. 4.6: PCA plot built using the top 500 changing genes – PC1 vs PC2.

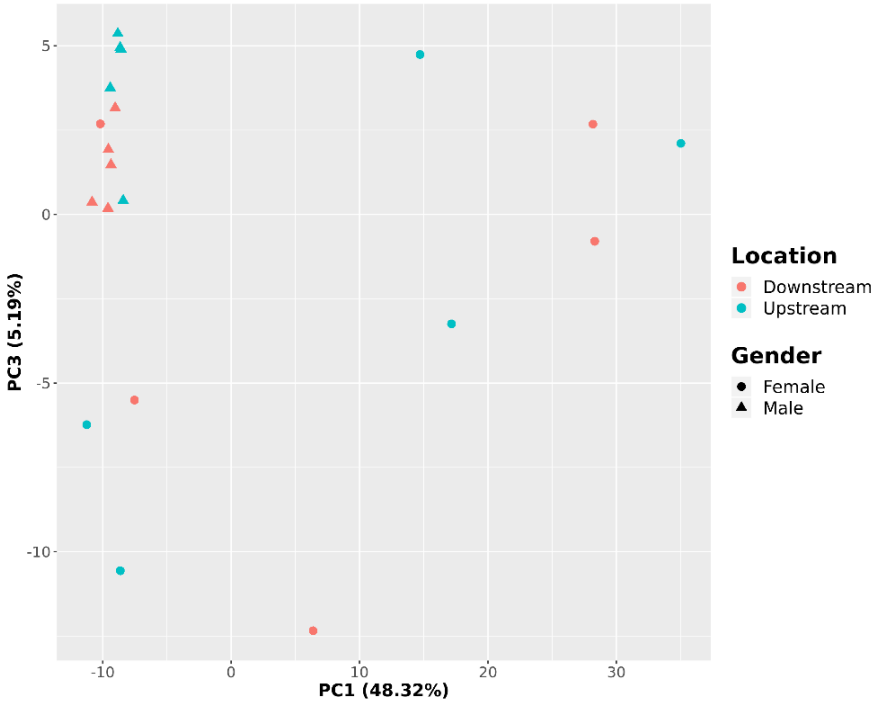

Fig. 4.7: PCA plot built using the top 500 changing genes – PC1 vs PC3.

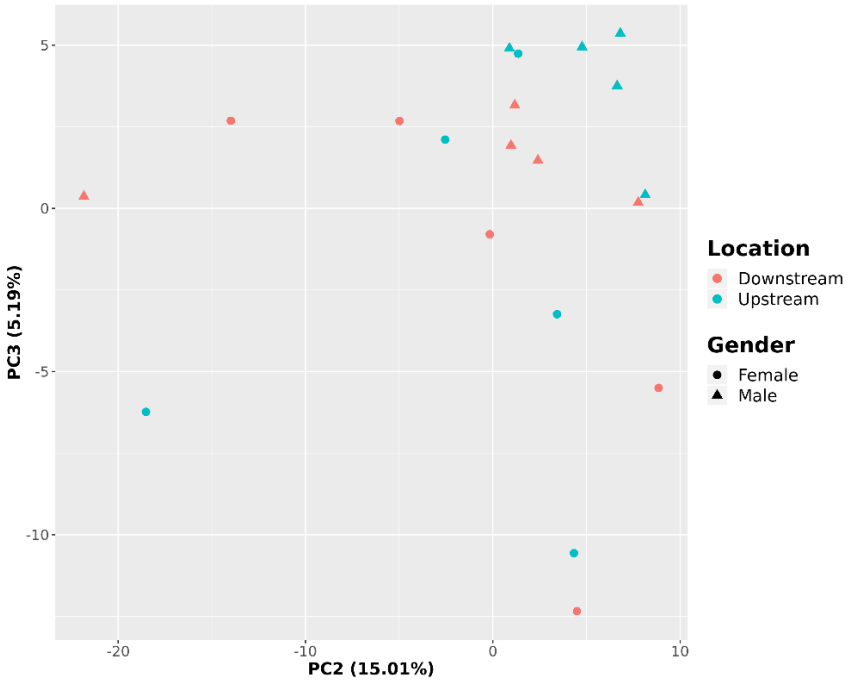

Fig. 4.8: PCA plot built using the top 500 changing genes – PC2 vs PC3.

Although the first 2 PCs explain the majority of the variation in the data (~63%) (Fig. 4.6), the other portions of variation, such as PC1-PC3 and PC2-PC3 are shown in Fig. 4.7 and 4.8, respectively. However, none of the PC plots built considering the first 3 PCs was able to identify clustering that systematically separates the upstream from the downstream samples.

4.4.4 Number of differentially expressed genes

In order to visualise the number of differentially expressed genes detected in the DGE analysis, a Venn diagram including the number of significantly changing genes for all analysed comparisons was built (Fig. 4.9).

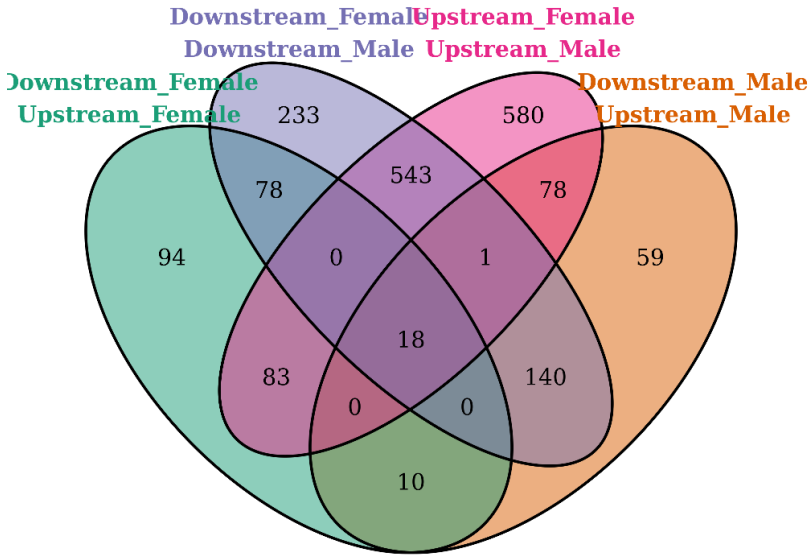

Fig. 4.9: Venn diagram showing the number of significantly changing genes in all comparisons.

A total of 306 and 283 differential genes were found when comparing the downstream and upstream populations for both male and female amphipods, respectively. Only 28 overlapping genes between the “Downstream female vs Upstream female” and “Downstream male vs Upstream male” comparisons were found (4.75%), supporting the hypothesis that male and female amphipods may respond differently to aquatic pollution exposure. On the other hand, the number of differentially expressed genes between males and females was higher compared to the number of differential genes between up- and downstream populations. A total of 1303 and 1013 differential genes were found when comparing male and female amphipods for both up- and downstream populations, respectively. 561 genes were found in common between “Upstream female vs Upstream male” and “Downstream female vs Downstream male” comparisons (24.22%).

4.4.5 Number of the differential genes: “Downstream vs Upstream”

The overall number of differentially expressed genes between up- and downstream populations, for both males and females, considering the filtered genes (section 4.3.4) is shown in Tab. 4.7. The genes were split into upregulated and downregulated between the two groups. Files S4.15&4.16 (Appendix B) show the complete list of differential genes between up- and downstream samples, including the UniProt gene descriptions, $\log_2$-transformed fold-changes and FPKM values for each gene.

	♂	♀
Upregulated genes	236	93
Downregulated genes	70	190

Tab. 4.7: Number of statistically significant differentially expressed genes between up- and downstream populations, for both males and females.

The differential genes between up- and downstream populations appeared to be mostly upregulated in male amphipods sampled at the downstream site, whilst the majority of differential genes detected in female amphipods were found downregulated (Tab. 4.7). This supports the hypothesis that substantial differences between the genders in the response to xenobiotic mixtures in water may be found in amphipods.

Volcano plots in Fig. 4.10 – 4.13 show the statistical significance (p-value) of the gene expression plotted against the magnitude of the change (fold-change). These scatter plots enable a visual identification of the transcripts that display large magnitude changes between two compared conditions (Li, 2012). Significantly downregulated and upregulated genes are represented by blue and red dots, respectively. The plots also show significantly changing genes (fold-change>2 and p-value<0.05) with a mean FPKM value across all samples less than 1 in both

groups of the evaluated comparisons (black dots). These genes were not included in the DGE analysis (section 4.3.4).

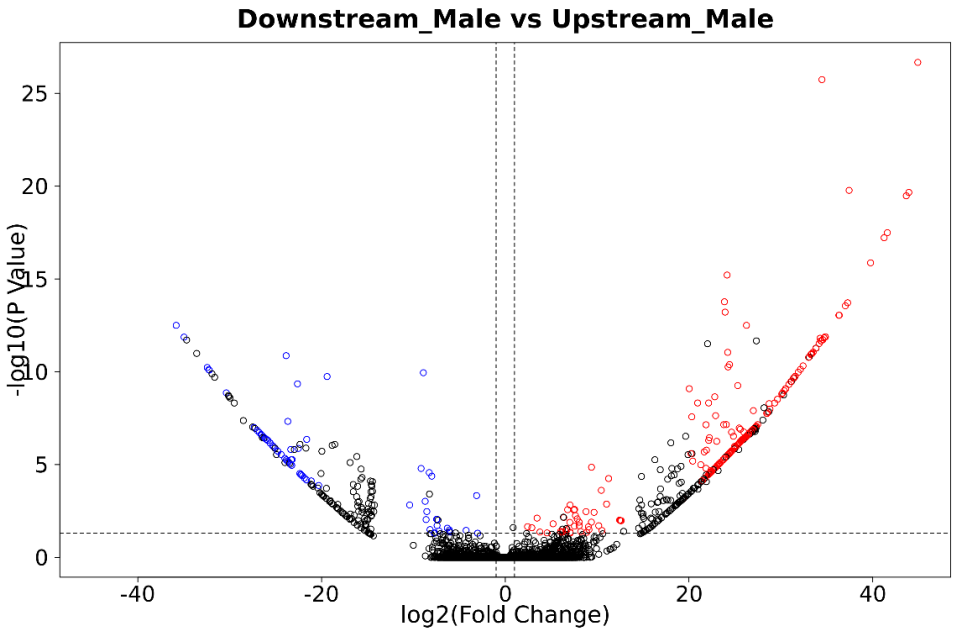

Fig. 4.10: Volcano plot built on the male differential genes between up- and downstream populations. $\log_2$ fold-changes are shown on the x-axis and the $-\log_{10}$ of the p-values on the y-axis. Downregulated genes are highlighted in blue, whilst upregulated genes are highlighted in red. Black dots represent genes with a mean FPKM value across all samples less than 1 in both groups of the evaluated comparisons.

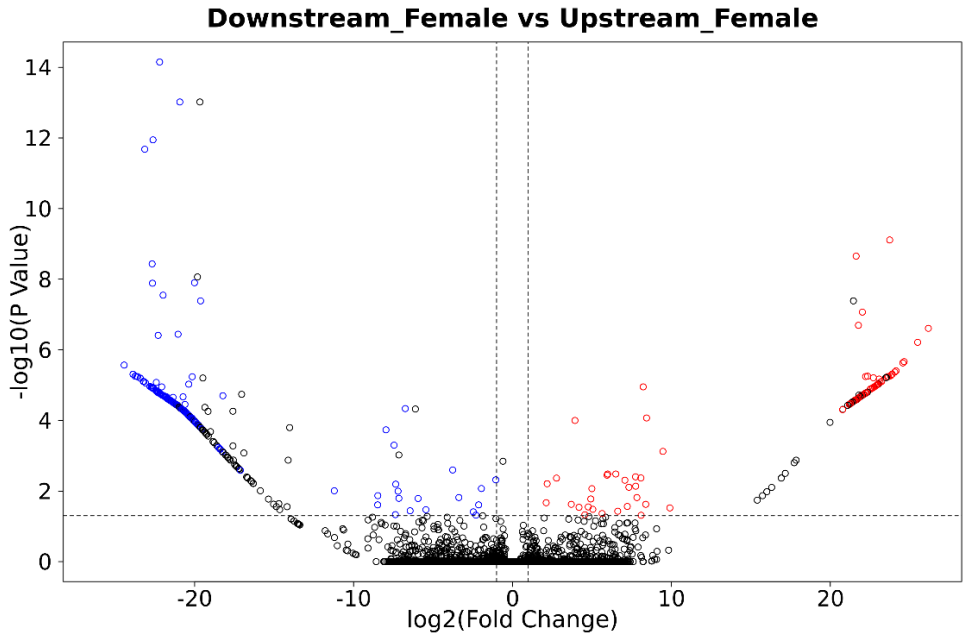

Fig. 4.11: Volcano plot built on the female differential genes between up- and downstream populations. $\log_2$ fold-changes are shown on the x-axis and the $-\log_{10}$ of the p-values on the y-axis. Downregulated genes are highlighted in blue, whilst upregulated genes are highlighted in red. Black dots represent genes with a mean FPKM value across all samples less than 1 in both groups of the evaluated comparisons.

4.4.6 Number of the differential genes: "Females vs Males"

Tab. 4.8 shows the overall number of differentially expressed genes between male and female samples, for both up- and downstream sites, considering the filtered genes (section 4.3.3). Similarly to Tab. 4.7, the genes were split into upregulated and downregulated between the two groups. Files S4.17&4.18 (Appendix B) show the complete list of differential genes between males and females, including the UniProt gene descriptions, $\log_2$ fold-changes and FPKM values for each gene.

	Upstream	Downstream
Upregulated genes	1210	776
Downregulated genes	93	237

Tab. 4.8: Number of statistically significant differentially expressed genes between males and females, for both up- and downstream populations.

The number of differential genes between the genders detected upstream of the WWTP was higher compared to the number of differential genes detected at the downstream site. It is possible that the variations in gene expression profiles triggered by the exposure to contaminants in water may have biased the detection of differences in gene expression between male and female amphipods, increasing the background noise in the analysis.

In order to visualise the statistical significance and the fold-changes of the differential genes between males and females, the same type of volcano plots built for the differential genes between up- and downstream populations (Fig. 4.10&4.11) were built for the differential genes between male and female amphipods sampled up- (Fig. 4.11) and downstream (Fig. 4.12) of the sewage effluent.

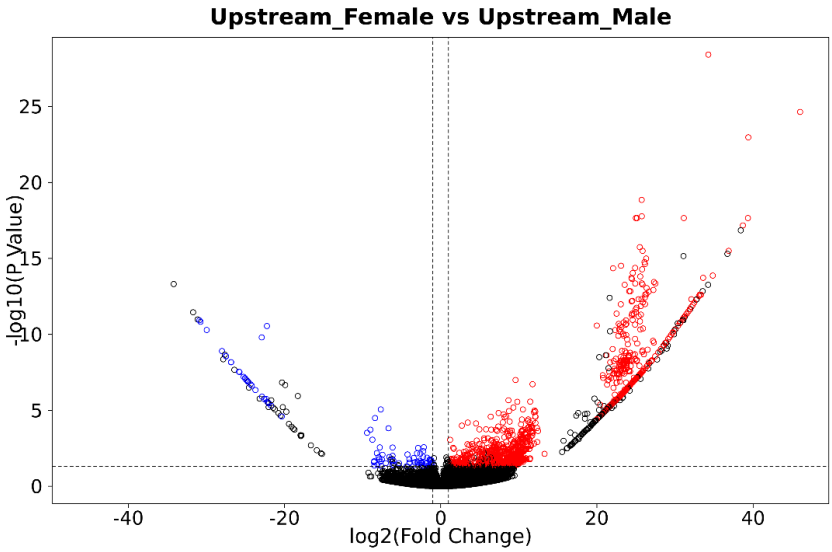

Fig. 4.12: Volcano plot built on the differential transcripts between male and female samples, collected upstream of the WWTP. $\log_2$ fold-changes are shown on the x-axis and the $-\log_{10}$ of the p-values on the y-axis. Downregulated genes are highlighted in blue, whilst upregulated genes are highlighted in red. Black dots represent genes with a mean FPKM value across all samples less than 1 in both groups of the evaluated comparisons.

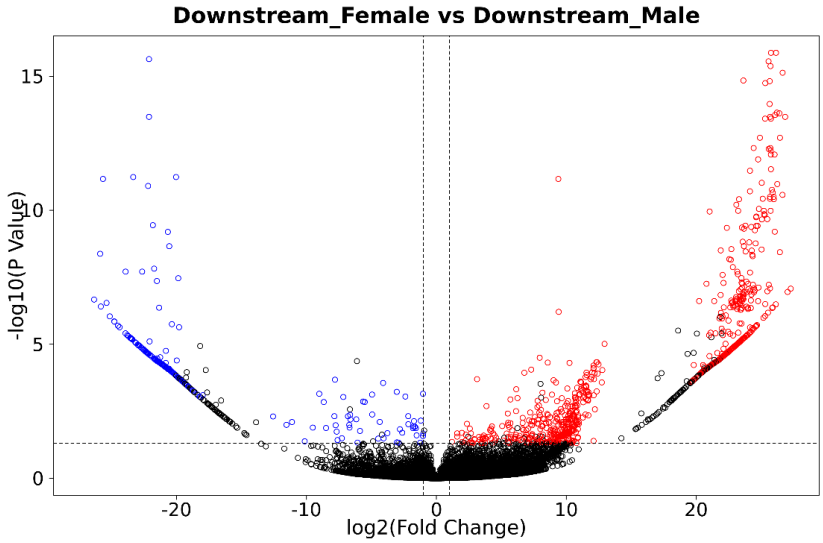

Fig. 4.13: Volcano plot built on the differential transcripts between male and female samples, collected downstream of the WWTP. $\log_2$ fold-changes are shown on the x-axis and the $-\log_{10}$ of the p-values on the y-axis. Downregulated genes are highlighted in blue, whilst upregulated genes

are highlighted in red. Black dots represent genes with a mean FPKM value across all samples less than 1 in both groups of the evaluated comparisons.

4.4.7 Functional overview

A functional classification of the differentially expressed genes annotated against UniProt database was focused on the GO-slim ontologies and conducted for both "Downstream vs Upstream" and "Females vs Males" comparisons. The pie charts in Fig. 4.14 and 4.15 show the proportions of the biological processes, molecular functions and cellular components where the differentially expressed genes were detected interrogating the *Drosophila melanogaster* whole-genome database.

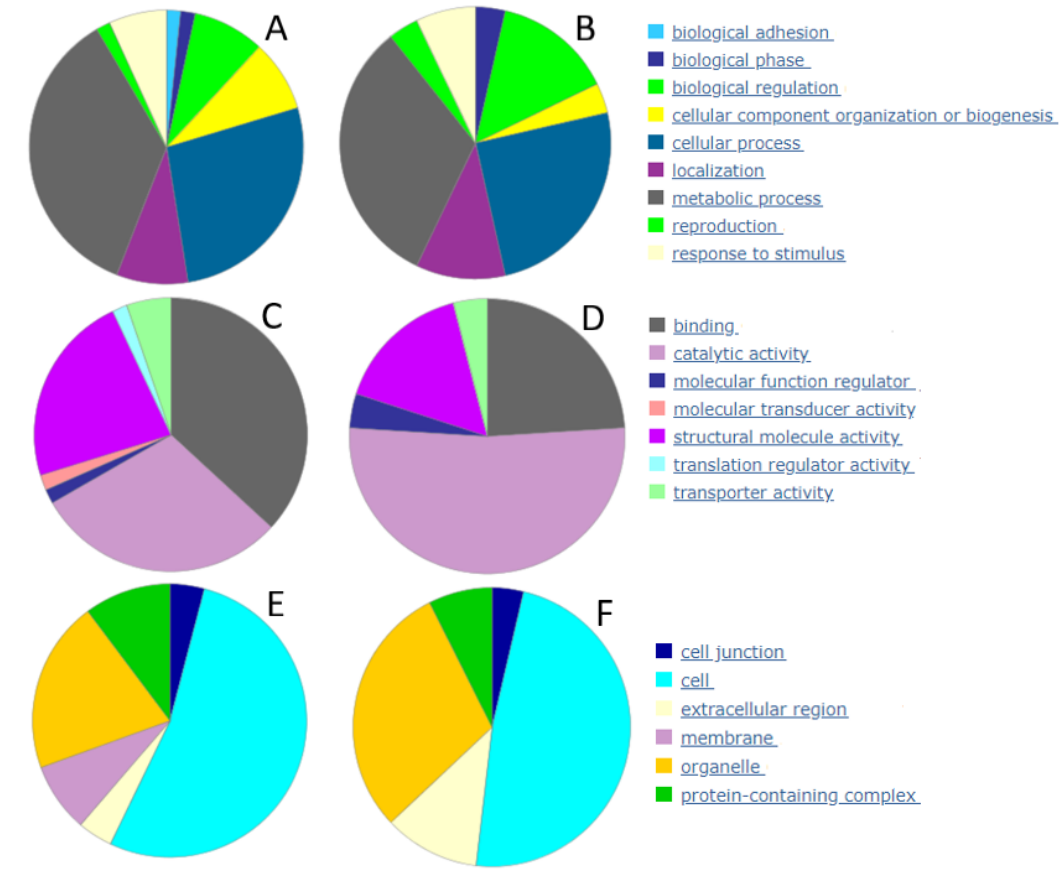

Fig. 4.14: GO-slim analysis conducted on the differential genes between upstream and downstream populations. The pie-charts were built submitting the differential genes annotated against UniProt database between up- and downstream populations for both males (A, C, E) and females (B, D, F) in the Panther database. Biological Process (A, B), Molecular Function (C, D) and Cellular Component (E, F) GO-categories were explored.

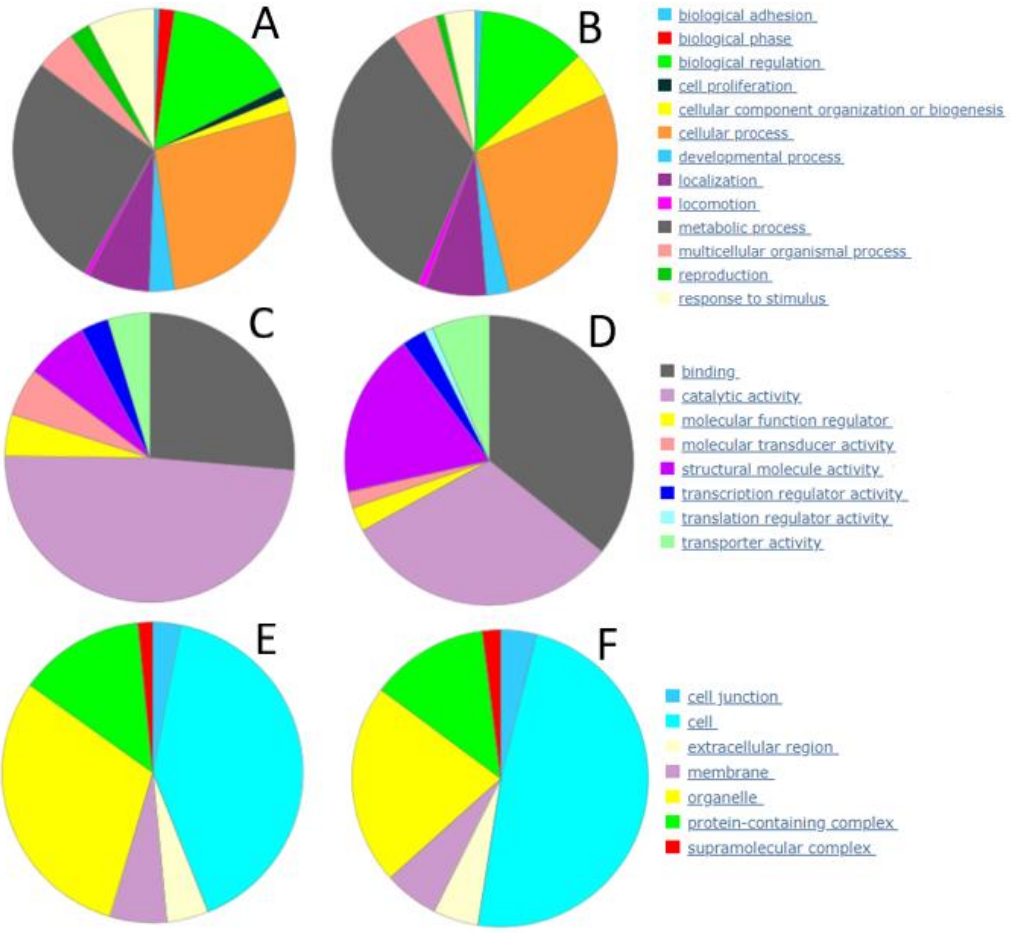

Figure 2Fig. 4.15: GO-slim analysis conducted on the differential genes between male and female amphipods. The pie-charts were built submitting the differential genes annotated against UniProt database between males and females for both upstream (A, C, E) and downstream (B, D, F) populations in the Panther database. Biological Process (A, B), Molecular Function (C, D) and Cellular Component (E, F) GO-categories were explored.

The results of the GO-complete analyses are shown in Files S4.3-4.13 (Appendix B). Fig. 4.16 shows the statistically enriched GO terms submitting a total of 170 (25.11% of the total of upstream and downstream genes annotated against UniProt) overlapping genes between “Downstream female vs Downstream male”

and “Upstream female Upstream male” lists of differentially expressed genes (Fig. 4.9).

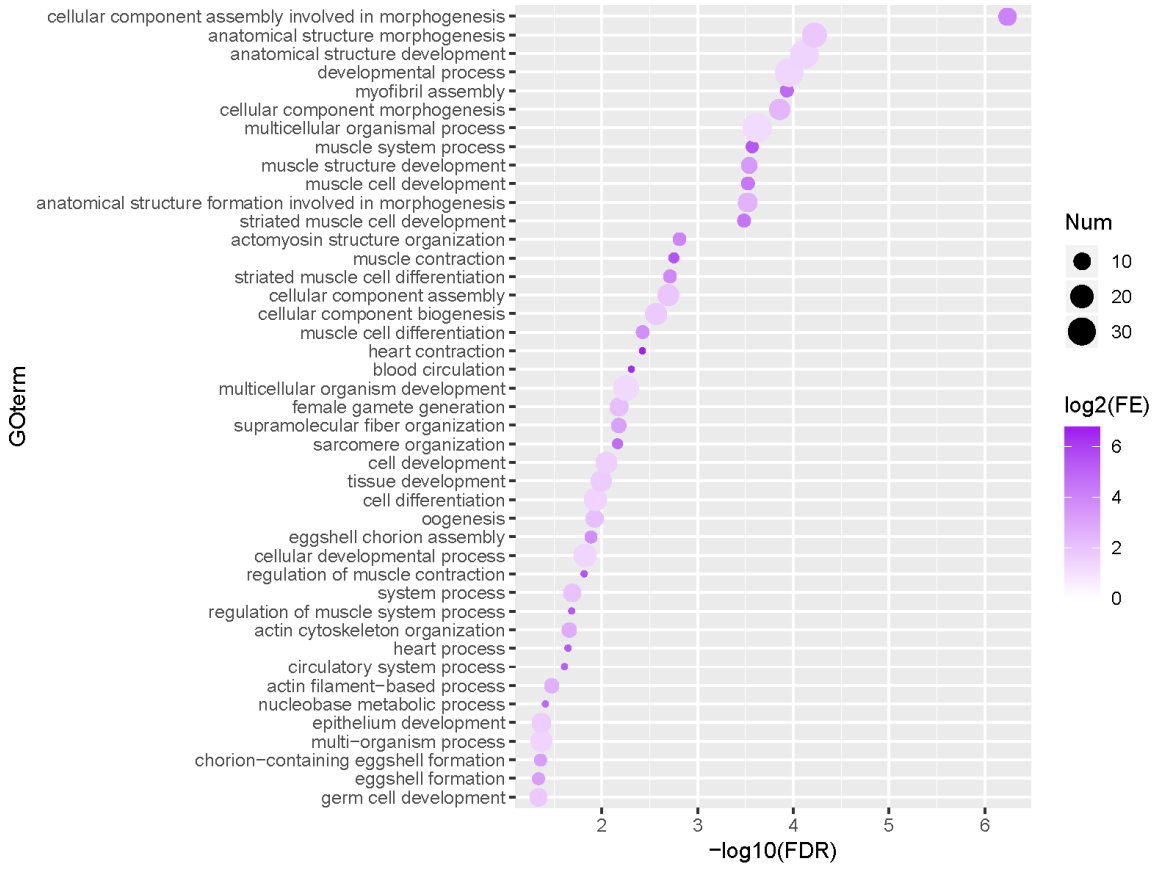

Fig. 4.16: Results obtained from a GO-complete over-representation analysis conducted on the genes annotated against UniProt that showed differential expression between males and females, both upstream and downstream of the WWTP (n=170). The GO category “Biological process” was explored and *Drosophila melanogaster* genome database was interrogated. -log10 transformed FDRs are shown on the x-axis and GO-terms on the y-axis. Bubble size and colour scale show the number of genes found in different biological processes and the log2 (fold-enrichment), respectively.

4.4.8 RNA quality assessment

Following the extraction of total RNA from the amphipods, DNase I was used to remove any residual gDNA from the RNA samples. 500 ng of RNA were used for an RNA quality assessment through electrophoresis on 1% agarose gel (Fig. 4.17&4.18). The RNA quality was assessed on 17 samples out of 20 total RNA extracts, since samples US♀2, US♀5 and DS♀3 showed low nucleic acid yields and did not contain enough material for both the cDNA synthesis and the RNA gel (Tab. 8.2 – Appendix A).

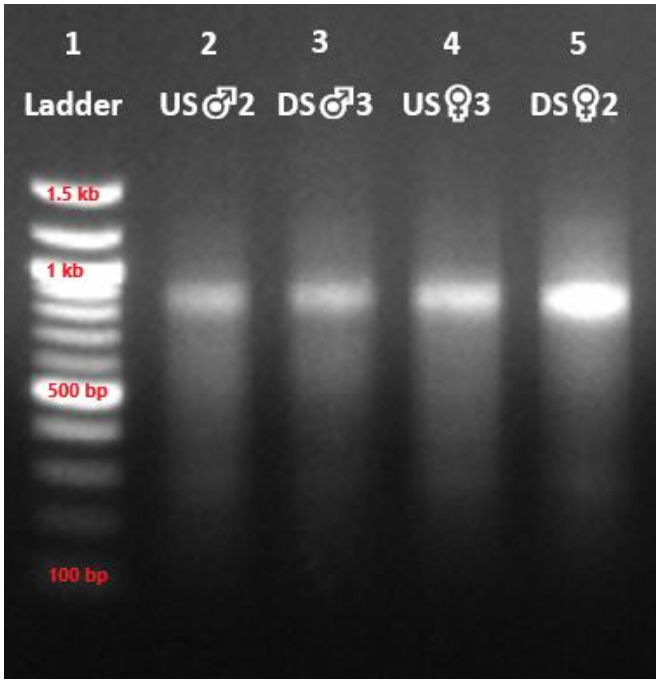

Fig. 4.17: Electrophoretic gel performed on the total RNA extracted from samples US♂2, DS♂3, US♀3 and DS♀2 (RNA gel 1). 2 male samples (lanes 2-3) and 2 female samples (lanes 4-5) were loaded on a 1% agarose gel.

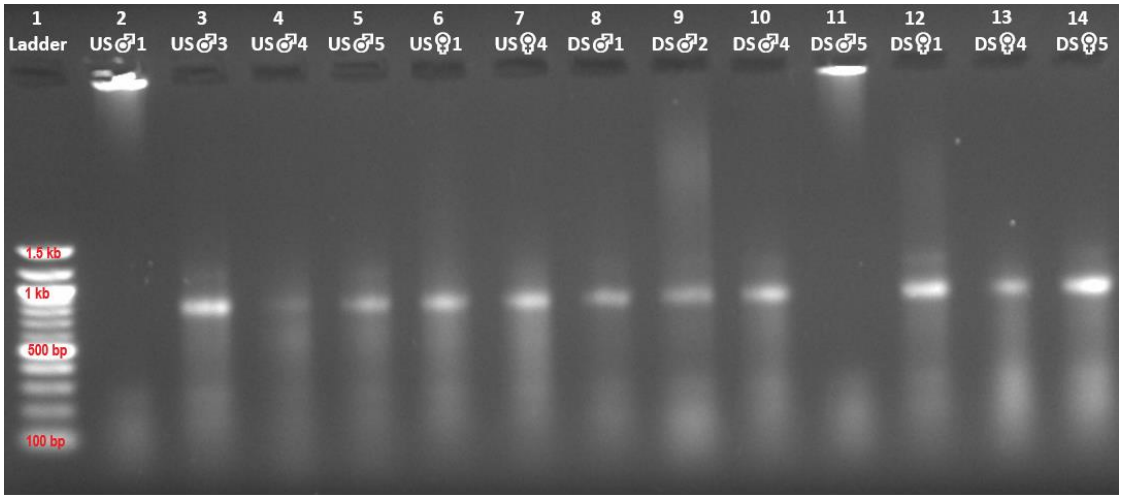

Fig. 4.18: Electrophoretic gel performed on the total RNA extracted from samples US♂1, US♂3, US♂4, US♂5, US♀1, US♀4, DS♂1, DS♂2, DS♂4, DS♂5, DS♀1, DS♀4, DS♀5 (RNA gel 2). 8 male samples (lanes 2-5, 8-11) and 5 female samples (lanes 6-7, 12-14) were loaded on a 1% agarose gel.

The DNA ladder in lane 1 of both gel 1 (Fig. 4.17) and gel 2 (Fig. 4.18) covers a 1500-100 bp size range. The gels show slight smears below the 1 kb ladder for most of samples, indicating a partial degradation of the extracted RNA. However, the bands of ~1Kb appear intense compared to the background and correspond to intact ribosomal RNA (Fig. 4.17&4.18) (see Chapter 3.8.5). The bands on top of lanes 2 (US♂1) and 11 (DS♂5) and the smears on top of lanes 9 (DS♂2) and 12 (DS♀2) indicate residual gDNA within the corresponding samples (Fig. 4.18).

4.4.9 Primer specificity test

Because of a lack of information in literature about the genes chosen for the qPCR validation, a primer set for each gene of interest and the housekeeping genes were designed *de novo*. The specificity of each primer set was assessed performing a PCR experiment. A cDNA male replicate (US♂1) and a cDNA female replicate (US♀1) were used as templates. An electrophoresis on a 1% agarose gel was set up to visualise the PCR products (Fig. 4.19).

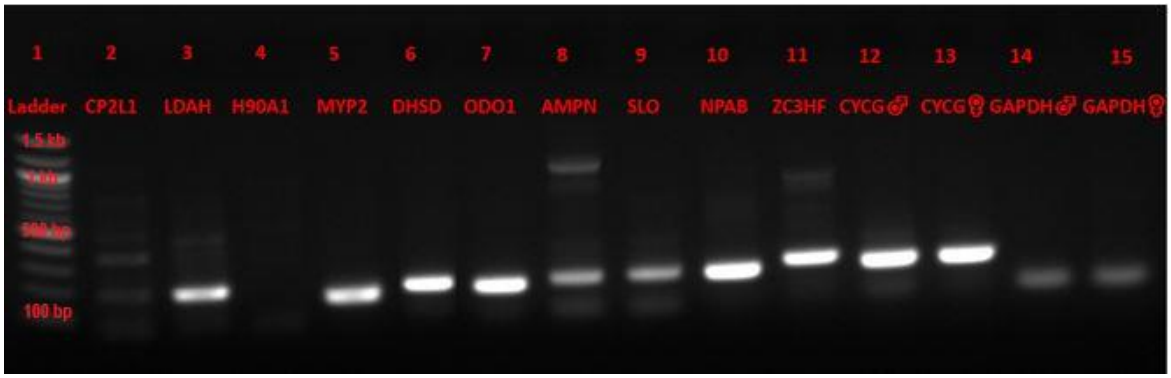

Fig. 4.19: Electrophoretic gel performed to test the specificity of the primers designed for the qPCR experiment. Gel picture showing the PCR amplification products using the primer sets designed for both male (lanes 2-6) and female (lanes 7-11) genes. Lanes 12-15 show the PCR amplification products using the primers designed for the housekeeping genes, using both male (lanes 12,14) and female (13,15) templates (US♂1 and US♀1).

The bands at ~200 bp in lanes 3,5,6,7 and lanes 9-13 represent specific amplification products. The primer set for the male gene *CP2L1* (lane 1) gave nonspecific amplification products. The primers designed for the male gene *H90A1* (lane 4) did not give amplification products. The primer sets for the female genes

AMPN (lane 8) and *ZC3HF* (lane 11) showed non-specific amplifications. The primers used for the housekeeper *CYCG* showed specific amplification products in both male (lane 12) and female (lane 13) samples. The primers designed for the housekeeper *GAPDH* did not give amplification products neither using a male sample (lane 14) nor a female sample (lane 15) as templates. The bottom bands (< 100 bp) in each lane correspond to the primers. The genes for which the designed primer sets did not give amplification products (*H90A1*, *GAPDH*) or showed non-specific amplification products (*AMPN*, *CP2L1, ZC3HF)* were not evaluated in the qPCR experiment.

4.4.10 cDNA control

To verify the overall functioning of the DNase I protocol, both the RNA extracts and cDNA samples were loaded on a 1% agarose gel. A PCR experiment using the cDNA samples as templates and the primers for the housekeeping gene *CYCG* was performed. For each sample (excluding DS♀3 due to a low RNA yield), both standard cDNA and -RT controls were loaded on 1% agarose gels (Fig. 4.20 – 4.22).

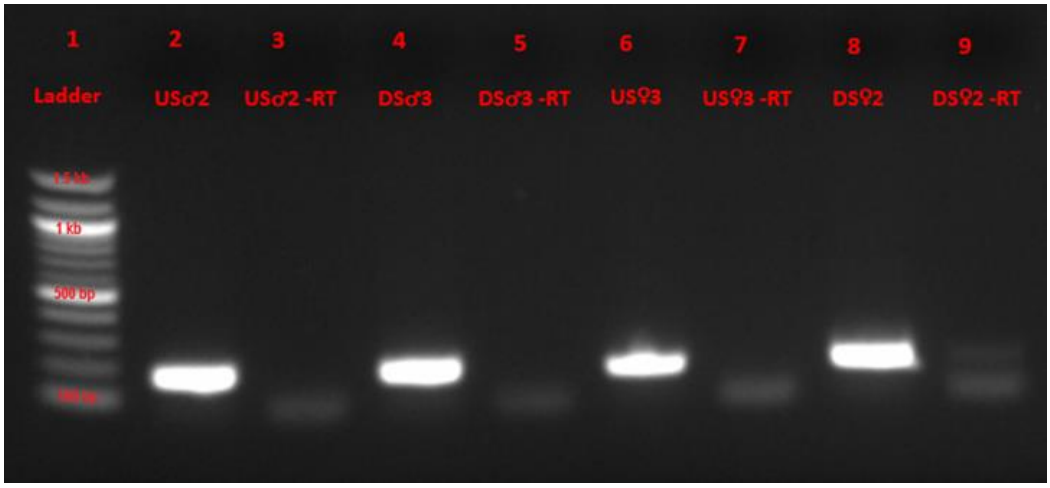

Fig. 4.20: Electrophoretic gel performed on the cDNA corresponding to samples US♂2, DS♂3, US♀3, DS♀2 (cDNA gel 1). Gel picture showing the PCR amplification products of the cDNA samples using the primer set for the housekeeping gene *CYCG*. Lanes 2,4,6,8 show standard PCR products while cDNA synthesis reactions lacking reverse transcriptase (-RT) were used as negative controls and can be seen in lanes 3,5,7,9.

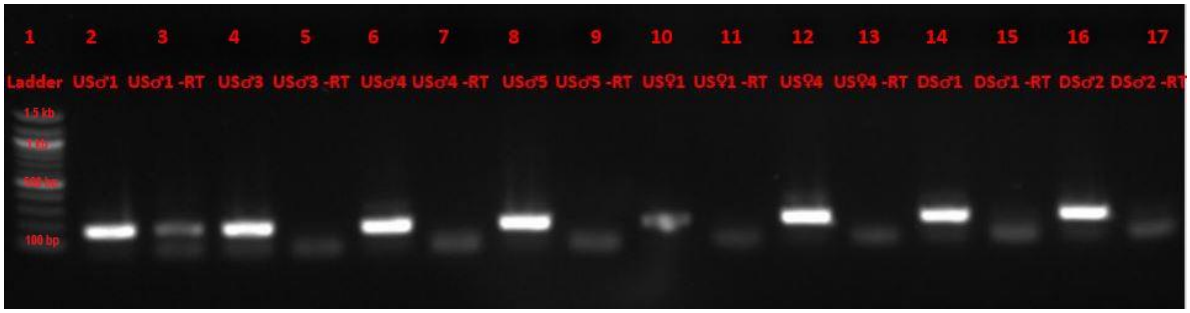

Fig. 4.21: Electrophoretic gel performed on the cDNA corresponding to samples US♂1, US♂3, US♂4, US♂5, US♀1, US♀4, DS♂1, DS♂2 (cDNA gel 2). Gel picture showing the PCR amplification products of the cDNA samples using the primer set for the housekeeping gene *CYCG*. Similarly to the cDNA gel 1 (Fig. 4.20), standard cDNA synthesis reactions and -RT controls were loaded alternately.

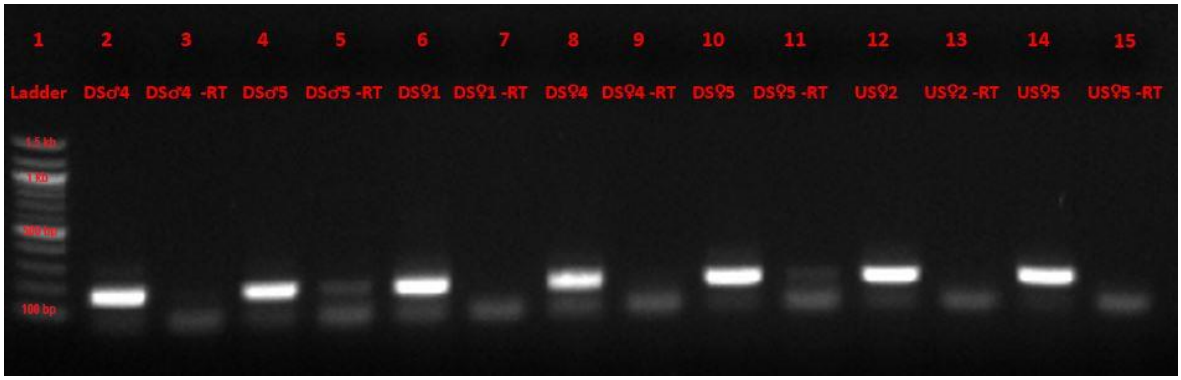

Fig. 4.22: Electrophoretic gel performed on the cDNA corresponding to samples DS♂4, DS♂5, DS♀1, DS♀4, DS♀5, US♀2, US♀5 (cDNA gel 3). Gel picture showing the PCR amplification products of the cDNA samples using the primer set for the housekeeping gene *CYCG*. Similarly to the cDNA gels 1 and 2 (Fig. 4.20&4.21), standard cDNA synthesis reactions and -RT controls were loaded alternately.

Consistently with the RNA gels (Fig. 4.17&4.18), the samples US♂1 (lanes 2,3 – Fig. 4.21) and DS♂5 (lanes 4,5 – Fig. 4.22) showed a band in the no-RT controls, indicating a potential presence of residual gDNA in the cDNA samples. Although the RNA sample DS♀5 did not show signs of cDNA contamination when loaded in a 1% agarose gel (lane 14 – Fig. 4.18), it did give an amplification product in the -RT control sample (lane 11 - Fig. 4.22). In fact, the RNA gel could have hidden the presence of residual gDNA in this sample that a PCR amplification reaction was able to highlight. The lanes corresponding to -RT controls of all the other samples were clean, indicating that the DNase I treatment successfully removed the genomic DNA.

4.4.11 qPCR results

-RT controls corresponding to the sample US♂1 for the genes *CYCG* and *MYP2* showed Ct values less than 30. Therefore, this replicate was excluded from the data analysis. Although the primers designed for the gene *SLO* showed a single amplification product in the primer specificity gel (Fig. 4.19), multiple amplification curves for this gene were generated in the qPCR experiment. Therefore, *SLO* was not included in the data analysis. Shapiro-Wilk tests conducted on the ΔCt values corresponding to each gene of interest showed that the values were not normally distributed in any of the evaluated groups ($p<0.05$). Hence, non-parametric tests (i.e., Mann-Whitney U-tests) were performed to calculate the statistical significance of the variations between the normalised gene expression values of upstream and downstream samples, using a significance level of $p<0.05$ (Tab. 4.9). Fig. 4.23 show the differences in normalised gene expression values between up- and downstream groups, for each evaluated gene. Tab. 4.10 shows the average relative expression values across the replicates in each group and the fold changes, for all evaluated genes.

♂	
Gene	**p-value (Mann-Whitney)**
LDAH	0.401
MYP2	0.022
DHSD	0.868
♀	
ODO1	0.044
NPAB	0.567

Tab. 4.9: P-values obtained from the Mann-Whitney U-tests performed to calculate the statistical significance of the variations between the normalised gene expression values (ΔCt values) of upstream and downstream samples, for the genes evaluated in both males and females.

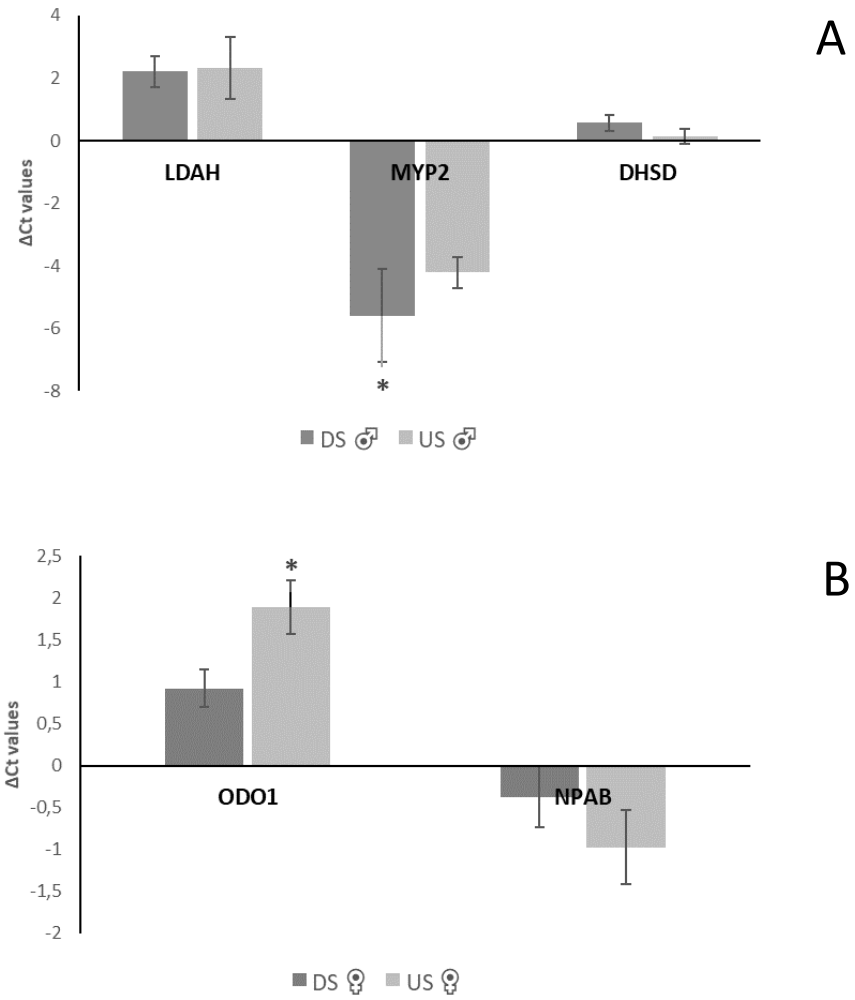

Fig. 4.23: Normalised expression levels for both male (A) and female (B) genes tested though qPCR. The mean of the expression values of each gene normalised with the expression levels of the reference gene *CYCG* (ΔCt values), including the biological replicates belonging to both upstream (US) and downstream (DS) groups were plotted. Error bars represent the standard error. Genes that showed statistically changing ΔCt values between up- and downstream replicates in a Mann-Whitney U-test ($p<0.05$) were marked with a "*".

A

Gene name	Average ΔΔCt	Fold-change ($2^{-\Delta\Delta Ct}$)
LDAH	-0,111	1,080
MYP2	-1,389	2,619
DHSD	0,442	0,737

B

ODO1	-0.966	1,9541
NPAB	0.602	0,6590

Tab. 4.10: Average relative expression and fold-change values in "Downstream vs Upstream" comparison for the genes evaluated in both male (A) and female (B) amphipods.

4.5 Discussion

The degree to which ecologically relevant species, such as amphipods are affected by an increasing number of anthropogenic pollutants released in aquatic environment through sewage effluents is a matter of continued study. A wide range of approaches have been used to investigate the effects of these compounds on amphipod biology, such as population studies, behavioural studies, and molecular studies (Chapter 1.6). The literature shows that a considerable part of these studies has focused on exploring the biological responses of amphipods to single xenobiotic compounds, whilst a lack of analyses investigating the effects of anthropogenic chemicals in mixture is evident. Although the evaluation of physiological and molecular responses to the most commonly detected compounds in sewage effluents being of crucial importance to understanding the mechanisms of toxicity, it has been shown that complex and hardly predictable effects occur in aquatic species when chronically exposed to anthropogenic chemical mixtures in their natural environment (David et al., 2017; Wigh et al., 2017). In the present project, phenotypical (Chapter 2), transcriptomic and metabolomic (Chapter 5) alterations in *G. fossarum* amphipods chronically exposed to a contaminant mixture released by a Swiss wastewater discharge were investigated. For the transcriptomic analysis, an RNA-seq approach was used to obtain the complete transcriptome of *G. fossarum* (Chapter 3). The transcriptome was assembled and annotated, and a DGE analysis was conducted in order to find the differentially expressed genes between amphipods sampled downstream

(contaminated site) and upstream (reference site) of the WWTP. Differences in gene expression profiles between male and female amphipods were also explored. To experimentally verify the variations in gene expression, a qPCR assay on several significantly changing transcripts between upstream and downstream amphipod populations detected in the DGE analysis was also performed.

4.5.1 Preliminary clustering

In order to check whether discrete transcript expression clusters could be observed according to the samples grouping (Upstream male, Upstream female, Downstream male, Downstream female) 2 heat maps were built: 1) a heat map including the whole set of transcripts (Fig. 4.4); 2) a heat map including the top 500 transcripts with the highest variance across the samples (Fig. 4.5). A PCA analysis using the top 500 changing transcripts was conducted to separate the samples based on the 3 greatest sources of variations (PC1, PC2 and PC3) (Fig 4.6 – 4.8). Overall, distinct expression clusters based on the gender or the sampling site were not observed in the heat maps. Although the PCA plots did not show a clear distinction between upstream and downstream samples, females appeared more disperse with the greatest source of variation compared to males (Fig. 4.6&4.7). A higher rate of separation of female samples was unsurprising, since changes in physiological parameters have been shown for female amphipods in different development/reproductive stages (Hyne et al., 2011). In fact, while male amphipods are available for mating during most of their moult (Sutcliffe, 2010),

females are sexually receptive for only a brief period and are constantly subject to complex morphological and hormonal changes during different ovarian and moulting cycles (Hyne, 2011). For instance, ecdysteroideal hormones co-regulate an increased synthesis of lipids and proteins during the formation of the new cuticle with the development of the brood plates (Hyne, 2011). Therefore, a higher variability in gene expression in female samples was expected, since the amphipods were not collected in the same development/reproduction stage. It is possible that the background noise caused by the biological variability could have hidden a clearer grouping of the samples based on the sampling site and the gender.

4.5.2 Differentially expressed genes: "Downstream vs Upstream"

4.5.2.1 Functional overview

The DGE analysis revealed a total of 306 and 283 differential genes between up- and downstream populations for the male and female lists, respectively (Tab. 4.7&4.8). Despite the two lists of genes showing little overlap (Fig. 4.9), a functional classification conducted in Panther revealed very similar proportions of genes corresponding to different GO terms. This was true in all three GO categories explored, such as Biological process, Cellular component and Molecular function (Fig. 4.14). Looking at the "Biological process" category, the most abundant portions of genes were found in "metabolic process" (GO:0008152) and "cellular process" (GO:0009987), for both male and female lists (Fig. 4.14).

Interestingly, *HSP90* (Heat shock protein 90 kDa) and *HSP70* (Heat shock protein 70 kDa) were found in the "response to stimulus" class (GO:0050896), in both male and female lists of differential genes between up- and downstream samples (File S4.15&4.16 – Appendix B). *HSP* genes encode evolutionary conserved molecular chaperones involved in protein folding and are considered as biomarkers of general stress responses in both crustacean and fish species (Moreira-de-sousa et al., 2018). Historically, the synthesis of these proteins was associated with exposure of *D. melanogaster* to high temperatures (Morimoto et al., 1984; Meyer et al., 1999). However, a number of studies showed altered expression of the genes encoding Hsp proteins in ecologically relevant species (e.g., mollusks, fish and crustaceans) in response to the exposure to a wide range of stress sources, including xenobiotics (Ekambaram et al., 2017), metals (Söyüt et al., 2012) and variations in the physico-chemical parameters of the water (Mi'covi'c et al., 2009). In addition, the expression of the genes encoding Hsp proteins have been shown to be upregulated or downregulated, depending on the organism studied and the stress source (Moreira-de-sousa et al., 2018). In this study, 2 transcripts were annotated as *HSP90* (TRINITY_DN102744_c2_g4_i1 and TRINITY_DN77657_c0_g1_i1) in the annotation against UniProt and both were found downregulated in female amphipods sampled downstream of the WWTP (File S4.16 – Appendix B). The transcript TRINITY_DN83373_c0_g1_i1 was annotated as *HSP90A1* in the male list of differentially expressed genes between up- and downstream samples and was found upregulated in amphipods sampled

downstream of the effluent (File S4.15 – Appendix B). The transcripts TRINITY_DN91778_c1_g1_i3 and TRINITY_DN101719_c0_g1_i1 were annotated as *HSP70* in the female and male lists, respectively (File S4.15&4.16 – Appendix B). Both transcripts were downregulated in amphipods sampled downstream of the WWTP. Whilst the transcripts annotated as *HSP70* were downregulated in amphipods sampled downstream of the effluent, *HSP90* was found downregulated in females and upregulated in males (File S4.15&4.16 – Appendix B). Strong discrepancies in the response to environmental stressors between male and female amphipods have been reported by a number of studies (Gismondi et al., 2012; Gismondi et al., 2013; ; Foucreau et al., 2014; Barros et al., 2017; Bedulina et al., 2017). For instance, an interesting study by Bedulina et al., (2017) showed a different response to thermal stress of the amphipods *Eulimnogammarus verrucosus* and *E. cyaneus* between the genders. The authors used a differential proteomics approach to compare the proteomic profiles of control amphipods kept at 6-7 ^{o}C and amphipods exposed to a 1h heat shock (24.5-25.5 ^{o}C). Significantly lower levels of Hsp70 were found in females of *E. verrucosus* after the heat shock compared to males, although no differences between the genders were found in *E. cyaneus.* Among the proteins with different expression between males and females of *E. verrucosus*, other heat shock proteins, such as Hsp60 and Hsp90, were identified using mass spectrometry. Their data highlight that male and female amphipods can show very different expression trends of heat shock proteins and these differences depend on the amphipod species

examined. In this project, the presence of differentially expressed genes between up- and downstream populations coding heat shock proteins was expected, since alterations of expression of both the 70 and 90 kDa isoforms have been observed in aquatic species in response to a wide range of stress sources, including thermic stress (Madeira et al., 2013), chemical exposure (Zhao et al., 2012) and parasite infections (Grabner et al., 2014). Further studies focusing on evaluating the expression levels of the genes encoding Hsp proteins in *G. fossarum* amphipods exposed to xenobiotic mixtures will clarify what substances trigger the variations in the expression of these important stress biomarkers, and the differences between the genders.

The gene annotated as tuberculosis sclerosis complex 1 (*TSC1*) (transcript code: TRINITY_DN111480_c2_g3_i2) was found downregulated in female amphipods sampled downstream of the WWTP compared to the upstream site (File S4.16 – Appendix B). Similarly to *HSP90* and *HSP70*, this gene was found in the "response to stimulus" class when functionally classifying the differentially expressed genes between up- and downstream female amphipods. Although a literature search having highlighted a lack of studies focusing on the functions of this gene in crustaceans, *TSC1* has been described as a tumour suppressor in *Drosophila melanogaster* (Sun et al., 2010). Specifically, *TSC1* has been shown to form a functional complex with *TSC2* that negatively regulates target of rapamycin (TOR), an evolutionarily conserved kinase that plays a central role in cell growth and

biosynthetic processes (Sun et al., 2010). Inactivating mutations of *TSC1* have been associated to an increase in cell number and organ size in *D. melanogaster* (Potter et al., 2001). On the other hand, the co-overexpression of *TSC1* and *TSC2* has been demonstrated to cause a decrease in cell size, cell number, and organ size, suggesting that *TSC2* may act as an epistatic regulator on *TSC1* (Potter et al., 2001). The fact that the gene *TSC1* was found differentially expressed only between female amphipods sampled up- and downstream of the sewage effluent highlights that males and females might respond differently to environmental pollutants. However, no other genes belonging to the *TSC1* pathway were found differentially expressed between up- and downstream female samples. It is currently unclear whether the background noise in the sequencing data caused by the biological variability may have hidden important changes in gene expression. Further and more targeted studies will be needed to investigate the functions of *TSC1* in crustaceans. In the present project, the histological variations between amphipods sampled above and below the sewage effluent were not evaluated, hence the long-term effects of aquatic contaminants on the TSC1-TOR pathway in amphipods will need to be investigated in future studies. Based on an evident increase in size of *D. melanogaster* organs containing a majority of $Tsc1^{-}$ mutant cells (Potter et al., 2001), potential variations in cell proliferation processes and organ size will need to be evaluated employing immunohistochemical and electron microscopy assays.

4.5.2.2 Molecular function and Cellular component

Exploring the "Molecular function" category, most of genes belonged to "catalytic activity" (GO:0003824), "binding" (GO:0005488) and "structural molecule activity" (GO:0005198) (Fig. 4.14 C,D). On the other hand, the majority of genes within the "Cellular component" category was associated to "cell" (GO:0005623) and "organelle" (GO:0043226), for both male and female lists (Fig. 4.14 E,F). The *D. melanogaster* orthologs that corresponded to these GO terms included *HSP* genes, genes associated to RNA transcription and maturation (e.g., the RNA splicing factors *CPSF5* and *U2AF1,* and the zinc-finger transcription factor *SUS*), genes encoding ribosomal proteins (*RL14, RL3* and *RSSA*) as well as genes encoding metabolic components, such as the subunit D of the complex of the respiratory chain succinate dehydrogenase (*DHSD*) and the lipid droplet-associated hydrolase *LDAH*. This functional analysis showed that the differentially expressed genes between up- and downstream samples were involved in a range of cellular processes, suggesting that the toxic pressure downstream of the sewage effluent could lead to perturbations in a wide spectrum of important transcriptional, translational and metabolic processes. In addition, the presence of *HSP* genes in both lists indicates that a general stress response may be triggered by the exposure to the contaminants in water. This stress response may be sub-lethal, in fact no evident differences between the population structures of amphipods sampled above and below the sewage effluent were observed (Chapter 2.6). Further phenotypical and molecular analyses as well as a constant monitoring of

the chemical composition of the stream will validate potential long-term toxic effects on amphipod populations.

4.5.2.3 GO-complete analysis

Following a preliminary functional classification, a GO-complete analysis was conducted on the differential genes between up- and downstream populations, in order to have a complete picture of the over-represented pathways within the *D. melanogaster* genome. Although no statistically enriched pathways were found in "Biological process" GO category when comparing the list of differentially expressed genes between up- and downstream populations in females, an enrichment in "metabolic process" (GO:0008152) class was observed submitting the male list (File S4.3 – Appendix B). In particular, "protein metabolic process", "primary metabolic process" and "nitrogen compound metabolic process" were found among the significantly over-represented pathways (File S4.3 – Appendix B). Interestingly, a pathway analysis conducted on the differential metabolites in males between amphipods sampled up- and downstream the WWTP also revealed several *D. melanogaster* metabolic pathways, including "One carbon pool by folate", which was statistically enriched ($p<0.05$) (Chapter 5.7.3.1). Several metabolic networks may be impaired at sub-lethal level in males exposed to the effluent, although long-term studies will be needed to clarify the actual ecological impact of these alterations. In fact, the molecular variations detected by using "omics" platforms were not evident in the population data (Chapter 2.6). On the

other hand, alterations in energetic metabolism were also speculated. “One carbon pool by folate” pathway partially takes place at mitochondrial level (Zong et al., 2016) and a number of genes with mitochondrial expression were found within the list of changing genes between up- and downstream male populations (e.g., *MTCH2, PCCA, DHE3, SL9B2, RM50, ODO1, DHSD, TIM8, COX16*) (File S4.15 – Appendix B). Due to the presence of a single significantly changing metabolite between up- and downstream female samples, a pathway investigation was not possible for females also in the metabolomic data analysis (Chapter 5.7.3.1). It has been shown that female amphipods are subject to complex molecular changes during their lifecycle (Hyne, 2011). In this project, females were not sampled in the same development/reproduction stage, thus the fact that no significantly over-represented pathways came out in “Biological process” class when submitting the female differential genes between up- and downstream populations was likely due to a strong biological variability in female amphipods.

4.5.3 Differentially expressed genes: “Females vs Males”

4.5.3.1 Functional overview

A total of 1303 and 1013 genes were found differentially expressed between males and females in up- and downstream lists, respectively (Tab. 4.8). Running a preliminary functional classification submitting the 2 lists of genes in Panther, similar proportions of different GO terms were observed for up- and downstream samples for all 3 GO categories explored (Fig. 4.15), although a percentage less

than 50% of genes was found in common between the 2 lists (25.11% of the total of differential genes annotated against UniProt of both “Upstream female vs Upstream male” and “Downstream female vs Downstream male” lists) (File S4.17&4.18 – Appendix B).

4.5.3.2 GO-complete analysis

Similarly to the approach used for the differential genes between up- and downstream populations, GO-complete enrichment analyses were performed submitting the differential genes between males and females in Panther, for both up- (File S4.9-4.11 – Appendix B) and downstream (File S4.7&4.8 – Appendix B) sampling sites. Unsurprisingly, submitting the list of changing genes between the genders at upstream site investigating the “Biological process” class, the analysis showed a significant enrichment in a wide range of GO-terms involved in reproduction, for example “oogenesis” (GO:0048477), “gamete generation” (GO:0007276) and “sexual reproduction” (GO:0019953) (File S4.9 – Appendix B). Intriguingly, this analysis also showed GO-terms related to stress response, such as “response to external stimulus” (GO:0009605) and “response to stress” (GO:0006950) (File S4.9 – Appendix B). Although the upstream site was considered as a reference - theoretically uncontaminated - site in the design of this study, the chemical analysis did reveal the presence of xenobiotics above the sewage discharge, in both water and amphipod samples (Chapter 2.6.4). This was also observed by Munz et al., (2018) who analysed the internal concentrations of a

range of substances in *G. fossarum* amphipods up- and downstream a WWTP and shown that for some compounds, similar or higher concentrations were detected upstream compared to downstream. Therefore, the detection of stress response pathways in the upstream population was not totally unexpected.

In order to find general sex biomarkers, excluding the influence of the sampling site, a GO-complete "Biological process" analysis conducted on the changing genes between males and females shared between up- and downstream samples was performed. The analysis revealed GO terms related to a wide range of processes, including gamete formation (meiotic break formation, eggshell formation), morphogenesis (cytoskeleton formation, angiogenesis, signal transduction), circulatory system (vascular contractility, nitric oxide synthesis) and muscle development (myofibril assembly) (Fig. 4.16). The myocyte expression of a portion of differentially expressed genes between males and females was also observed in a GO-complete "Cellular component" analysis, which revealed a range of GO terms related to muscle developing and structure (File S4.8&4.11 – Appendix B). It is well known that the vast majority of differences between male and female crustaceans are related to their sexual organs and gamogenesis (Pamuru, 2019), which are finely regulated by their endocrine system (Hyne et al., 2011). However, given the wide spectrum of molecular and physiological alterations observed in response to the exposure to endocrine disruptor compounds (e.g., formation of intersex individuals) (Ford et al., 2008; Schneider et al., 2015) the research on

sexual biology in crustaceans is still intense. Although further analyses will be necessary to investigate in detail the whole range of differences between the genders, the need for molecular sex biomarkers is evident. In fact, the results obtained from both the transcriptomic and metabolomic analyses suggest that males and females may respond differently to chemical stressors, even at a sub-lethal level, when no clear effects on the overall population structure are observed. It is noteworthy though, that the presence of terms related to heart and circulatory processes, muscle system as well as cell differentiation and development among the over-represented GO-terms when submitting the overlapping genes between "Upstream females vs Upstream males" and "Downstream females vs Downstream males" comparisons (Fig. 4.16) will need to be investigated in more detail. The differences between the genders may not be limited to reproduction-related processes and structures. More likely, the sex distinction traits in hormonal system may act on a wider spectrum of molecular networks. A deeper understanding of these interactions could potentially clarify the differences between the genders in the response to anthropogenic pollutants in water, ultimately helping to elucidate the ecological impact.

4.5.4 qPCR experimental design

Based on bioinformatics parameters obtained in the RNA-seq data analysis, 10 transcripts were selected from the list of statistically changing transcripts between amphipods sampled up- and downstream of a WWTP, for an experimental

validation through qPCR. In particular, 10 genes with interesting biological functions were selected from the annotated genes against UniProt database with a high fold-change between up- and downstream samples (log_2(fold-change)>|20|). Having the DGE analysis detected different lists of differentially expressed genes between up- and downstream populations in males and females, 5 genes were chosen from the female changing transcripts and 5 different genes were chosen from the male list. Mehennaoui et al., (2018) recommend using 2 reference genes in qPCR analyses using *G. fossarum.* Therefore, 2 housekeeping genes (i.e., *CYCG,* involved in the G2/M phase of cell cycle and *GAPDH*, coding the glyceraldehyde-3-P dehydrogenase enzyme) were selected from the transcriptomic dataset. Although *CYCG* was not highly expressed, it was among the genes with the lowest variance across the samples (Tab. 4.3). Mehennaoui et al., (2018) identified 3 particularly stable genes in *G. fossarum.* Using a number of algorithms for the calculations of the stability coefficients, the authors identified *Clathrin*, *SDH* and *GAPDH* (in ascending order of stability coefficients) as the most stable reference genes for qPCR data normalization on *G. fossarum* sp. In the transcriptomic dataset, 2 potential transcripts were annotated as *Clathrin* (*CLH1*) in the annotation against UniProt. However, these transcripts showed a low FPKM value (TRINITY_DN96217_c3_g2_i2 - FPKM=64 and TRINITY_DN116446_c15_g3_i4 - FPKM=33). Similarly to *Clathrin*, a low FPKM value (FPKM=7.5) was detected for the transcript annotated as *SDH* (TRINITY_DN110356_c0_g5_i1). Although a high variance across the samples was

detected for *GAPDH* (*G3P* in this dataset) (Tab 4.3), a t-test showed no significant difference (t-test: df=19; p-value=0.171; α=0.05) between the groups, suggesting that the observed fluctuations were consistent between males/females and upstream/downstream samples. Furthermore, *GAPDH* was highly expressed in the data set (Tab. 4.3). For these reasons, it was chosen as a second housekeeper. In order to validate the variations in expression specifically detected on the transcript sequences of our dataset, the primer sets for both the genes to validate and the housekeeping genes were designed *de novo*, using the corresponding transcript sequences as templates. Unfortunately, the primer set used for *GPADH* gave no amplification products when running a PCR on both male and female cDNA samples, suggesting that the designed oligonucleotides did not anneal to the sequence detected in the RNA-seq experiment and annotated as *GPADH.* However, a PCR performed using the primers for *CYCG* and both male and female cDNA samples as templates showed a single amplification product (Fig. 4.19). RNA-seq data showed a low variance for the FPKM values of *CYCG* across all samples (Tab. 4.3), indicating very similar expression levels regardless of the sampling site and gender. Comparable Ct values of this gene were also detected in the qPCR experiment between up- and downstream samples for both males (Mann-Whitney U-test: p-value=0.101; α=0.05) and females (Mann-Whitney U-test: p-value=0.206; α=0.05). Therefore, *CYCG* was used as reference gene in the qPCR validation experiment. This gene was not analysed in the study conducted by Mehennaoui et al., (2018), since the authors focused on the most commonly used

reference genes in qPCR experiments on amphipods. Furthermore, it is noteworthy that the expression of reference genes can be influenced by biotic or abiotic stress as well as developmental stage and tissue type (Mehennaoui et al., 2018). This is even more true in particularly sensitive species, such as *G. fossarum* (Wigh et al., 2017). The experiments of the present project were conducted with the aim of measuring the molecular variations between amphipods sampled above and below a sewage treatment plant (chronically exposed to a complex contaminant mixture), minimising laboratory treatments. It is therefore possible that the expression of *CYCG* may variate in response to different treatments and further investigations will be needed to confirm the general suitability of *CYCG* as reference gene in *G. fossarum*. However, both the qPCR data discussed in this chapter and RNA-seq data showed that the expression levels of this gene were very similar in all samples, making it ideal as reference gene in this study.

Because the primer sets for the genes *H90A1*, *GAPDH* did not give any amplification product when running a standard PCR on cDNA samples (Fig. 4.19), the expression of the corresponding genes was not evaluated in the qPCR experiment. The genes *AMPN*, *CP2L1, ZC3HF* were also excluded from the analysis, since the corresponding primers gave multiple amplification products (Fig. 4.19). It is possible that these genes were present in multiple isoforms and the chosen primers annealed to the shared nucleotide sequences. The female gene *SLO* showed multiple curves in the qPCR amplification plots, indicating that the

corresponding primers were not specific for the selected template. It was therefore not possible to include this gene in the data analysis. Following the mentioned exclusions, the number of genes analysed through qPCR was reduced to 3 (*LDAH, MYP2, DHSD*) for males and 2 (*ODO1, NPAB*) for females.

4.5.4.1 Significantly changing genes in males

Among the evaluated genes in males, *MYP2* was found statistically downregulated in downstream samples compared to the controls (Fig. 4.23 A) (Tab. 4.10 A). RNA-seq also showed a significant down-regulation of this gene in male amphipods sampled downstream of the WWTP (Tab. 4.1). In this transcriptomic data set, *MYP2* sequence was annotated as "Myelin P2 protein" against UniProt database. In general, the protein coded by this gene constitutes a fraction of the myelin complex in peripheral nervous system (https://www.uniprot.org/). Although a literature search having highlighted a lack of studies focusing on the effects of whole sewage effluents on peripheral nervous system in crustaceans, cytological and histological alterations in peripheral nervous system were found in fish exposed to contaminated waters (Burkhardt-Holm et al., 1997). Particularly, decompaction of myelin sheaths was observed in brown trout (*Salmo trutta*) exposed to an undiluted treat wastewater effluent, indicating peripheral nerve degeneration (demyelination) (Burkhardt-Holm et al., 1997). Functional analyses conducted on the changing transcripts and metabolites between male amphipods collected upstream and downstream the WWTP did not show enrichments in

neuronal pathways (Chapters 4&5). However, despite the high rate of biological variation detected in both RNA-seq and qPCR analyses, the expression levels of *MYP2* were found significantly lower downstream of the effluent compared to upstream in male samples, in both experiments. Therefore, the validated altered expression of a gene coding an important constituent of myelin sheaths in male amphipods collected below the WWTP may represent a consequence of the exposure to the contaminant mixture.

4.5.4.2 Significantly changing genes in females

The data analysis showed a statistically significant upregulation of the gene *ODO1* in amphipods collected downstream of the WWTP compared to the reference site (Fig. 4.23 B) (Tab. 4.10 B). This was in agreement with RNA-seq results that showed an upregulation of this gene in downstream samples (Tab. 4.2). *ODO1*, also known as *OGDH,* is a gene encoding one subunit of the 2-oxoglutarate dehydrogenase complex. This complex catalyses the conversion of 2-oxoglutarate (α-ketoglutarate) to succinyl-CoA and CO_2 during the Krebs cycle at mitochondrial level (https://www.ncbi.nlm.nih.gov/). Although further investigations would be needed to confirm a role of this gene in amphipod stress response, other genes at mitochondrial localization were detected among the differentially expressed genes between females collected up- and downstream of the WWTP in the DGE analysis, such as *HCD2* (3-hydroxyacyl-CoA dehydrogenase type-2) and *FXRD1* (FAD-dependent oxidoreductase domain-containing protein 1) (File S4.16 –

Appendix B). Furthermore, the metabolomic analysis revealed that "1C-metabolism", a pathway partially taking place at the mitochondrial level (Zong et al., 2016), was among the statistically enriched pathways when submitting the differential metabolites between male amphipods sampled up- and downstream of the WWTP in the MetaboAnalyst database (Chapter 5.7.3.1). The analyses conducted in this project were not specifically focused on the evaluation of mitochondrial stress in amphipods. However, both transcriptomic and metabolomic data showed alterations in mitochondrial components, suggesting that mitochondrial processes may be affected by the exposure to the chemical mixture contained in the effluent. These alterations could be present in both male and female amphipods and further analyses will be needed to investigate the differences in mitochondrial stress responses between the genders.

4.5.4.3 Not validated genes

Despite the high fold-fold changes detected in DGE analysis for the male genes *LDAH* and *DHSD* (Tab. 4.1), and for the female gene *NPAB* (Tab. 4.2), no statistically significant variations in gene expression were found in the qPCR analysis when comparing amphipods sampled up- and downstream of the WWTP (Tab. 4.9). Although a wide range of ecotoxicological studies having confirmed the results obtained from the application of transcriptomic platforms using qPCR assays (e.g., Ings et al., 2011; Hook et al., 2014; Short et al., 2014), both technical and biological variation sources can be strong limiting factors in validation experiments

(Rajkumar et al., 2015). Whilst pooling samples represents a cost-saving strategy for RNA-seq experiments and allows to increase the RNA yield, qPCR is a targeted approach and is usually employed to evaluate the changes in gene expression using independent replicates. This difference in the experimental design can potentially lead to a different estimation of the biological variability during the data analysis. Furthermore, if the samples used for the transcriptomic experiment and the qPCR validation are not collected in the same sampling, changes in gene expression due to seasonal variations and fluctuations in chemical composition of the river may bring additional sources of variation (Davie et al., 2009; Nelson et al., 2011; Guler et al., 2015; Munz et al., 2018). In this project, due to logistic reasons, the amphipods for the sequencing and qPCR experiments were sampled 2 years apart. Therefore, the discrepancies in the results of the 2 analyses could be due to the previously mentioned sources of variations. In addition, a false positive rate of 0.05 was set in the DGE analysis (section 4.3.4) to detect the significantly changing genes between the groups, thus the presence of false positives within the lists of changing genes cannot be excluded. Nevertheless, the qPCR data analysis revealed relative expression levels between up- and downstream samples consistent with the DGE analysis for the genes *MYP2* and *ODO1*. Moreover, the data set obtained from the sequencing of the complete transcriptome of *G. fossarum* (deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession code PRJNA556212) and the results of the DGE analysis will represent a valuable resource for future molecular studies on poorly

annotated crustacean species, such as amphipods. Further long-term studies investigating the fluctuations in chemical composition of the rivers and their effects on amphipod gene expression profiles will be essential.

Chapter 5 - Metabolomics analysis

5.1 Introduction

Environmental metabolomics is the application of metabolomics to characterise the interactions of organisms with their environment. This approach has many advantages for studying organism-environment interactions and for assessing organism function and health at the molecular level (Bundy et al., 2009). Like genomics, transcriptomics and proteomics, the use of metabolomic platforms have increased significantly in environmental molecular investigations during the last decade (Gómez-Canela et al., 2016; Kovacevic et al., 2016; Chiu et al., 2017; David et al., 2017). In fact, this relatively new "omic" science is increasingly being applied in post-genomic sciences to study a wide range of biological systems including microorganisms (Smedsgaard et al., 2005; MacKenzie et al., 2008), plants (Hall, 2006; William et al., 2006), mammals (Kell et al., 2006; Dunn et al., 2007) and other ecologically relevant organisms (Chiu et al., 2017; David et al., 2017; Gómez-Canela et al., 2016; Kovacevic et al., 2016). In particular, metabolomics is finding an increasing number of applications in the environmental sciences, ranging from understanding organismal responses to abiotic pressures, to investigating the responses of organisms to other biota. One of the main advantages is that these interactions can be studied from individuals to populations, which can be related to both the traditional field of ecophysiology and from instantaneous effects to those over evolutionary time scales, the latter enabling studies of genetic adaptation (Bundy et al., 2009). It is also important to

emphasise that metabolomic investigations clarify the actual functional status of the organism (cell, tissue or biofluid), which is strictly related to organism phenotype.

In general, metabolomic studies can use two different approaches: targeted and untargeted. The scope of these two types of approaches is different and they both have advantages and disadvantages. Targeted metabolomics approaches have a low detection limit and enable the absolute quantification of the sample. However, when using a targeted strategy, the instruments are set to only detect one or few classes of biological molecules (e.g., total carbohydrates or total phenolics) (Davey et al. 2007). Therefore, targeted approaches will not allow the discovery of unknown compounds. On the other hand, untargeted metabolomics approaches provide a global view of a sample. In this case, the disadvantages include the complex informatics required to interpret the results. In addition, the semi-quantitative nature of the methods and the need to validate any identified compounds are also among the main disadvantages (Menni et al., 2017).

5.2 Metabolomics in environmental toxicology

The literature shows many examples of applications of metabolomic platforms in environmental toxicology. Studies that employed metabolomic platforms to investigate the effects of single contaminants and whole effluents on fish increased the molecular knowledge on the biological responses of these species

to contaminants as well as on the pharmacokinetics of xenobiotics in animal tissues. For instance, it has been shown that the exposure of the brown bullhead (*Ameiurus nebulosus*) to Aroclor (a polychlorinated biophenyl mixture) leads to a strong immunosuprression, ultimately causing an increased susceptibility to opportunistic parasite infections (Iwanowicz et al., 2009). Using a UPLC-TOF/MS (Ultra Performance Liquid Chromatography - Time of Flight Mass Spectrometry) platform, Al-Salhi et al., (2012) showed a clear metabolic separation between control rainbow trout (*Oncorhynchus mykiss*) sampled upstream of a British WWTP and animals exposed to undiluted effluent for 10 days. The authors started from the hypothesis that fish tend to bioaccumulate xenobiotics when exposed to contaminated effluent waters and applied an untargeted metabolomics approach to identify both xenobiotics and their metabolization products in bile and plasma of the animals. Many of the contaminants detected in trout bile and plasma derived from commonly used surfactants. A variety of phenolics (e.g., dichlorophenol, trichlorophenol, chloroxylenol, diclosan and triclosan) were also identified in bile from effluent-exposed trout, and these were predominantly conjugated to glucuronic acid. Interestingly, bile acids were found in trout plasma as well as in hepatic fluids. It is therefore possible that a chronical exposure to complex xenobiotic mixtures may lead to histological injuries as well as molecular changes (Al-Salhi et al., 2012). In fact, the majority of the compounds are metabolised in the liver prior to be transported to the bile, thus a liver injury can cause an increased release of bile acids into the plasma (Rosen et al., 2001). An

interesting study by David et al., (2017) shows that roach (*Rutilus rutilus*) exposed to a wastewater effluent rich in pharmaceutical wastes (e.g., nonsteroidal anti-inflammatory drugs and SSRIs) leads to a prostaglandin reduction, a tryptophan/serotonin pathway impairment and changes in lipid metabolism. The authors employed a nanoflow-nanospray mass spectrometry untargeted profiling technique to identify the changes in plasma metabolome between control fish exposed to clean water and fish exposed to contaminated water for 2 weeks. Surprisingly, metabolite disruptions were not explained by altered expression of genes encoding enzymes for which related metabolites were found to change in the prostaglandin and serotonin pathways. This highlights that metabolic disruptions may not be due to changes in expression of biosynthetic genes, but rather may arise from direct inhibition of enzyme activity (David et al., 2017).

5.3 Applications of metabolomics in crustaceans: perspectives and limits

High-throughput studies on ecologically important crustaceans, such as isopods, amphipods and euphausids are increasing more and more in recent years. The interest is focused on how crustaceans respond to environmental stress sources, including temperature, viral and bacterial infection, metal and organic toxicants (Stillman et al., 2015). In general, “omics” platforms allow to take a picture of the whole molecular phenotype of an organism. This becomes crucial when studying crustacean species in environmental studies. In fact, crustaceans are constantly exposed to a wide range of anthropogenic stressors in their natural environment,

hence the application of techniques capable of detecting the whole set of molecular variations caused by the exposure to contaminants became fundamental (Schock et al., 2010). To date, the literature shows a wide range of studies that applied metabolomic platforms to crustacean species, in order to elucidate their response to both chemical and physical environmental stressors (Samuelsson et al., 2011; Nagato et al., 2013; Nagato et al., 2016; Kovacevic et al., 2016). Starting from one of the most used crustacean sentinel species in the ecotoxicology field, such as *Daphnia magna*, a number of studies have been performed to evaluate the responses of this organism to a range of chemicals. For instance, the molecular responses of *D. magna* following exposure to several commonly detected anthropogenic chemicals in freshwater ecosystems, such as organophosphates pesticides, bisphenol A (Nagato et al., 2016), metals (Nagato et al., 2013) and pharmaceuticals (Kovacevic et al., 2016) have been evaluated using metabolomic platforms. Employing a ^{1}H NMR platform, Nagato et al., (2016) demonstrated that energetic molecules (e.g., glucose and lactate) and amino acids are in an inverse proportionality relationship after exposure of *D. magna* to diazinon and malathion, 2 organophosphates pesticides. Examining the metabolomic changes after exposure to bisphenol A, the authors also showed that the observed responses were not linearly concentration dependent, due to the complexity of the underlying biochemical processes (Graney et al., 2011). In fact, performing toxicity tests only using mortality as endpoint can be often unreliable, as they do not provide information about the changes that are occurring at sub-

lethal levels (Nagato et al., 2016; Bownik, 2019). To support the complexity of predicting adverse responses after exposure to chemicals using high-throughput metabolomic platforms (e.g., ^{1}H NMR), Kovacevic et al., (2016) showed that an aminoacidic depletion occurred when *D. magna* was exposed to sub-lethal concentrations of common pharmaceuticals, such as carbamazepine, ibuprofen and triclosan. However, additional effects (e.g., misregulation of several intermediates of Krebs cycle) were only visible after exposure to higher concentrations of carbamazepine, suggesting that even more complex responses may occur when crustaceans are chronically exposed to contaminant mixtures containing hundreds of different compounds.

Studies applying metabolomic platforms on amphipods have also been performed. In particular, NMR and GC mass spectrometry platforms have been used to identify the metabolic fingerprints of several amphipod species in response to chronical exposure to complex chemical mixtures released by sewage effluents (Chiu et al., 2017) or single contaminants, in lab studies (Ralston-Hooper et al., 2011; Gómez-Canela et al., 2016). Metabolites involved in β-oxidation and lipid metabolism were found when exposing the amphipod *Hyalella azteca* to atrazine, one of the most detected herbicides in the U.S., suggesting possible disruption in energy metabolism (Ralston-Hooper et al., 2011). Significant changes in metabolites involved in oxidative stress, protein synthesis and a broad range of signaling cascades were found when comparing control *G. pulex* amphipods with

amphipods exposed to 3 commonly detected pharmaceuticals in wastewaters (triclosan, nimesulide and propranolol) (Gómez-Canela et al., 2016). Chiu et al., (2017) employed an NMR platform to explore the metabolic changes between *Hyalella azteca* amphipods sampled in a reference stream and a number of Taiwan contaminated rivers. Amino acid metabolism, Krebs metabolism, glycolysis and gluconeogenesis networks were found among the enriched pathways when submitting the differential metabolites in MetaboAnalyst database using *D. melanogaster* data set. Their results suggest a significant impairment in primary metabolic processes in amphipods exposed to whole effluents.

Despite the huge amount of data generated by using metabolomics strategies to a wide range of species, in both targeted and untargeted studies, it is noteworthy that currently the metabolomic databases (e.g., Human Metabolome Database, MetaboAnalyst, BiGG) contain metabolic data for *Homo sapiens* and a limited number of model organisms, for example *Drosophila melanogaster*, *Danio rerio* and *Caenorhabditis elegans*. In fact, unlike functional analyses conducted using genomic or proteomic data sets, where in most cases it is possible to interrogate databases specific for the species being studied, the lack of species-related metabolomic databases still represents a strong limiting factor in terms of data interpretation. This is even more true in untargeted studies where the whole metabolic fingerprint of an organism is evaluated. In most cases, researchers are limited to use metabolic data sets of evolutionary distant species to conduct

functional analyses. On top of that, it has been shown that both size and composition of the metabolome vary greatly, depending on the organism studied. For example, *Saccharomyces cerevisiae* contains approximately 600 metabolites (Famili et al., 2003), the plant kingdom has an estimated 200,000 primary and secondary metabolites (Fiehn, 2002) and the human metabolome is even larger in size and more complex in composition (Dunn et al., 2005). These strong differences in the metabolomes of different species represent one of the toughest challenges in annotating metabolomics data. Although strategies aimed to using different species to conduct metabolic pathway analyses being still useful to conduct a preliminary functional classification of large metabolomic data sets, further studies are needed to collect detailed metabolic data from ecologically relevant species. This will allow the environmentally important organisms to be adequately represented in terms of metabolomic data.

5.4 Overview on metabolomic platforms

Metabolomic studies can be technically defined as those analyses based on the simultaneous measurement of multiple metabolites, using inherently parallel analytical techniques such as NMR (Nuclear Magnetic Resonance) spectroscopy or MS (Mass Spectrometry), followed by appropriate statistical analysis that typically employs multivariate or else repeated univariate tests (Bundy et al., 2009). The appropriate platform to be used depends on the purpose of the analysis. In fact, whilst mass spectrometry provides an excellent approach that can offer a

combined sensitivity and selectivity for metabolomics research, NMR also provides information on the molecular structures of metabolites and does not require extra steps for sample preparation, such as separation or chemical derivatization (Emwas et al., 2015) (Fig. 5.1).

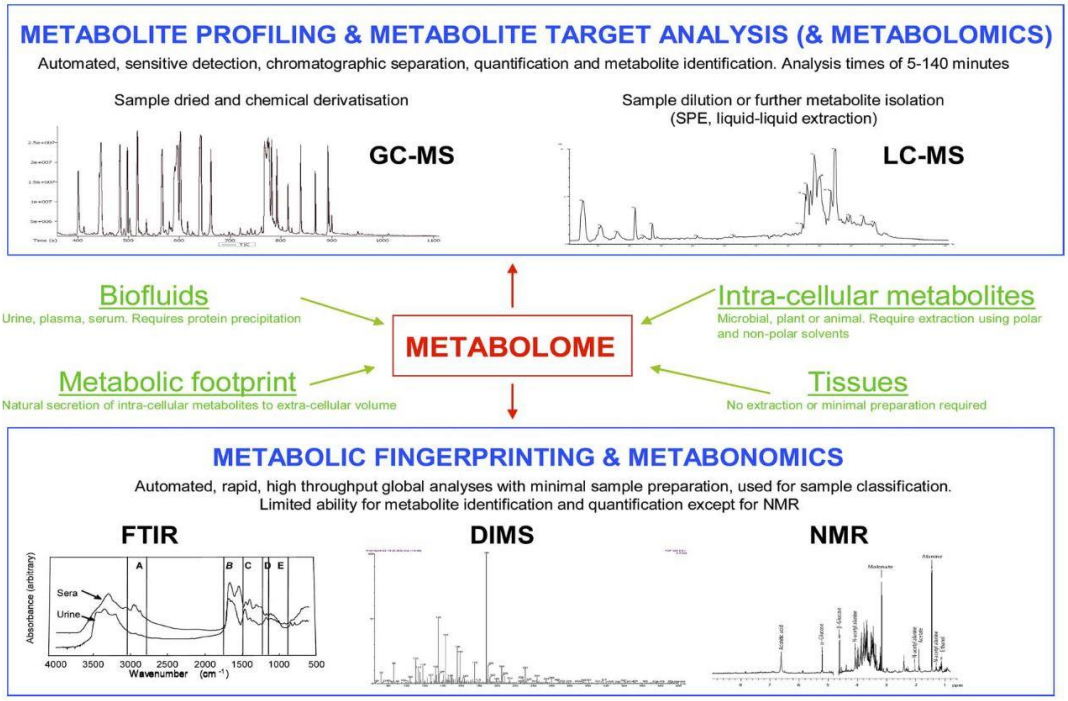

Fig. 5.1: Platforms used in metabolomics. Different metabolomics-based strategies for sample preparation and sample analysis. *Reproduced with permission from Dunn et al., 2005.*

5.4.1 NMR spectroscopy

NMR spectroscopy allows to perform rapid, non-destructive, high-throughput metabolic investigations that require a minimal sample preparation (Reo et al., 2002; Lindon et al., 2003). NMR platforms function by the application of strong

magnetic fields and radio frequency (RF) pulses to the nuclei of atoms. For atoms with either an odd atomic number (e.g., ^{1}H) or odd mass number (e.g., ^{13}C), the presence of a magnetic field will cause the nucleus to possess spin, termed nuclear spin. Absorption of RF energy will then allow the nuclei to be promoted from low-energy to high-energy spin states, and the subsequent emission of radiation during the relaxation process is detected. Ultimately, this process allows to obtain very accurate information on the molecular structure of the metabolites (Dunn et al., 2005). The output of an NMR analysis is an NMR spectrum, which is related to a measure called “chemical shift”. The chemical shift depends on the effect of shielding by electrons orbiting the nucleus. The chemical shift for ^{1}H NMR is calculated as the difference (in ppm) between the resonance frequency of the observed proton and that of a reference proton present in a reference compound (e.g., tetramethylsilane). The measured chemical shifts vary based on the different atoms: 0–10 ppm for ^{1}H; 0–250 ppm for ^{13}C. The signal intensity depends on the number of identical nuclei, and the presence of complex samples does not interfere with measured intensity as ionisation suppression does with electrospray ionisation. This allows to obtain also quantitative information (Dunn et al., 2005).

5.4.2 Mass spectrometry

Although the sensitivity of NMR spectroscopy has increased enormously and improvements continue to emerge, it still represents a weak point of NMR compared with mass spectrometry (MS). MS-based metabolomics provides an

excellent approach that can offer a combined sensitivity and selectivity platform for metabolomics research (Emwas et al., 2015). To date, MS is the most widely applied technology in metabolomics, as it provides a blend of rapid, sensitive, selective, qualitative and quantitative analyses with the ability to identify metabolites and to discover new compounds. MS principle is based on ion formation and separation according to their mass-to-charge (*m/z*) ratio and detection of the separated ions. These steps are shared by all the MS platforms, although ion formation and separation strategies, sample preparation methods and data analysis can be very different depending on the platform used. The following sections summarise technical principles and applications of the two most used MS platforms in both clinical and environmental fields: GC (Gas Chromatography)-MS and LC (Liquid Chromatography)-MS.

5.4.3 GC-MS

In a GC-MS analysis, volatile and thermally stable compounds are first separated by GC and then eluting compounds are detected by electron-impact mass spectrometers. In this technique, a proper chemical derivatization of the sample at room or elevated temperatures is essential to provide volatility and thermal stability prior to analysis. In fact, the difficulty to analyse non-volatile and high-molecular weight compounds represents the main disadvantage of this platform (Dunn et al., 2005). Because of the vast range of chemical classes of metabolites, usually two stages of derivatization are employed. First, functional groups are

converted to oximes with O-alkylhydroxylamine solutions, followed by formation of trimethylsilyl (TMS) esters with silylating reagents (typically N-methyl-N-(trimethylsilyl)-trifluoroacetamide), to replace exchangeable protons with TMS groups. Oxime formation is required to eliminate undesirable slow and reversible silylation reactions with carbonyl groups, whose products can be thermally labile. Some metabolites contain exchangeable protons and hence a range of derivatisation products are formed. For example, amino acids and carbohydrates will form multiple derivatisation products, whereas organic acids often react to create only one detected product. Chromatograms in output are often complex, containing hundreds of metabolite peaks and are further complicated by multiple derivatisation products. Therefore, either long run times (greater than 60 min) (Roessner et al., 2002) or a combination of fast acquisition rate TOF (Time of Flight) instruments coupled with deconvolution software (Dunn et al., 2005) are used.

5.4.4 Electrospray LC-MS

LC-MS provides metabolite separation by liquid chromatography and ion source, such as an electrospray ionization (ESI) system (Fenn et al., 1989). ESI uses electrical energy to assist the transfer of ions from solution into the gaseous phase before they are subjected to mass spectrometric analysis. Ionic species in solution can be analysed by ESI-MS with increased sensitivity. Neutral compounds can also be converted to ionic form in solution or in gaseous phase by protonation or cationization and hence can be studied by ESI-MS (Ho et al., 2003). Using ESI as ion

source, a single metabolite can be detected as multiple features in either positive or negative ion modes. For example, a metabolite may be detected as protonated and sodiated ions in positive ion mode and as deprotonated ion in negative ion mode (Brown et al., 2009). This can lead to an overestimation of the number of detected metabolites and makes difficult to determine the molecular formula, as the type of ions formed is often unknown. The transfer of ionic species from solution into the gas phase by ESI involves three steps: dispersal of a fine spray of charge droplets, solvent evaporation and ion ejection from highly charged droplets (Fig. 5.2). LC-MS systems involve easy sample preparation protocols compared to GC-MS. Chemical derivatization steps are usually not required and the sample preparation only involves proper metabolic extraction protocol using different organic solvents (e.g., methanol-chloroform, acetonitrile/methanol, isopropanol/methanol), depending on the application and the matrix studied (Liebeke et al., 2012).

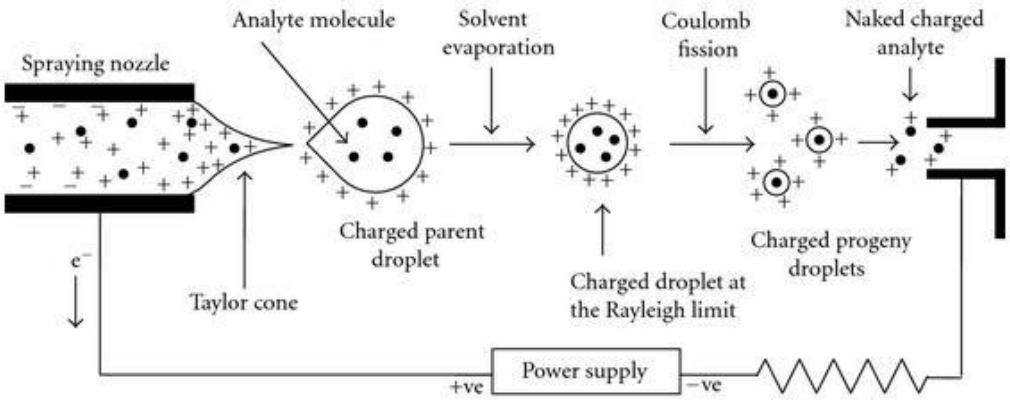

Fig. 5.2: Schematic representation of the electrospray ionization process. A mist of highly charged droplets with the same polarity as the capillary voltage is generated from the spraying nozzle. The application of a nebulising gas (e.g., N_2) which shears around the eluted sample solution and a high temperature within the chamber (> 200 ^oC) enhance a higher flow rate. The charged droplets, generated at the exit of the electrospray tip, pass down a pressure gradient and potential gradient toward the analyser of the mass spectrometer (Banerjee et al., 2012).

5.4.5 Mass analyser

Although the ion source being an essential part of a mass spectrometer, the mass analyser represents the heart of the instrument. In the mass analyser, different types of ions (*m/z*) of an ion beam are separated, and then they are passed to the detector. Many types of mass analysers are available (Hoffmann et al., 2001; Gross et al., 2004). For example, magnetic/electric sector mass analyser, linear quadrupole ion trap (LIT), three-dimensional quadrupole ion trap (QIT), orbitrap, TOF mass analyser, and ion cyclotron resonance mass analyser (ICR), all of these using static or dynamic magnetic/electric fields (Banerjee et al., 2012). Proper

selection of the mass analyser depends on the resolution, mass range, scan rate, and detection limit required for the application (Banerjee et al., 2012).

5.4.6 MS/MS analysis

The precursor ions of interest can be mass selected and further fragmented in a collision cell. This process is defined as “tandem mass spectrometry” or MS/MS (McLafferty, 1983; Todd, 1991). In an MS/MS experiment, a precursor ion is mass selected by a mass analyser Q1 and then focused into a reaction cell q2 (collision cell) where it undergoes a gas-phase chemical reaction. The collision of the ions with neutral gas molecules (e.g., N_2, He_2, Ar) gives different product ions with different masses, which are then passed to a third mass analyser. This last mass analyser scans the masses of the product ions and generates an ion spectrum. The mass analysers are set up in series either in space (sector, triple quadrupole, and hybrid instruments) or in time (trapping instruments) (Banerjee et al., 2012). This process is widely used in both proteomics and metabolomics, since it allows to obtain detailed structural information on biomolecules. MS/MS spectra can be uploaded in online databases to identify the metabolite mixtures in a process known as molecular annotation.

5.5 Aim and objectives

Having recorded the differences in gene expression between amphipods sampled above and below a Swiss WWTP (Chapter 4), a UPLC-MS/MS platform was

employed to investigate the metabolic fingerprints in *G. fossarum* populations sampled at the same sites used for the transcriptomic analysis. The hypothesis was that the genes found differentially expressed between amphipods sampled up- and downstream of the effluent were biochemically related to the changing metabolites detected in the metabolomics analysis. The differential metabolites could be used as a base to further studies focusing on biological pathways affected by exposure to contaminant mixtures released through sewage effluents, in amphipods. In addition, a pathway analysis was conducted using the differential metabolites between male and female *G. fossarum* amphipods to explore the metabolic differences between the genders. The enriched sex-specific metabolic pathways, in conjunction with the transcriptomic data, will provide an extensive data set useful to develop new sex-specific markers in crustaceans.

5.6 Methods

5.6.1 Collaborations and contributions (Metabolomics analysis)

Amphipods for the metabolomics analysis were collected by Dr Andrea Schifferli and Dr Thomas Bucher (Swiss Centre for Applied Ecotoxicology, Dübendorf, Switzerland). Metabolic extractions were my own work in collaboration with Dr Dorsa Varshavi (University of Birmingham). Mass spectrometry, statistical analyses and metabolic annotation were conducted by Dorsa Varshavi and Prof Mark Viant (University of Birmingham). Pathway analyses in MetaboAnalyst and functional investigation of the differential metabolites were my own work.

5.6.2 Metabolic extractions

Sampling of *G. fossarum* for the metabolomic analysis was conducted in mid-September 2017 at the same site used for both population and transcriptomic analyses (Chapter 2.5.2). A total of 25 males (13 upstream and 12 downstream) and 16 females (7 upstream and 9 downstream) were used. An endometabolome extraction was performed using a standard methanol-chloroform protocol (Liebeke et al., 2012). Briefly, the amphipods were weighted and individually placed in 2 mL Precellys tubes (LifeScience Products, Cheltenham, UK) kept on dry ice. 320 µl of methanol and 128 µl of dH_20 were added to each Precellys tube. The tubes were placed in a Precellys 24 homogeniser machine (LifeScience Products, Cheltenham, UK) and 2 homogenization cycles of 10s at 6400 rpm were set. Homogenised mixtures were transferred into 1.8 mL glass vials with aluminium lined caps (Fisher Scientific, Leicestershire, UK). The vials were placed on ice and 320 µl of chloroform plus 160 µl of dH_20 were added to each vial. Vials were vortexed at full power for 30s each and left on ice for 10 mins. Samples were then centrifuged at 3000 rpm at 4 °C for 10 min and stabilised on bench for 5 min at room temperature. Both upper (polar) and lower (non-polar) phases were separately placed in new tubes (1,5 mL tubes for polar phases and 1,8 mL glass vials for non-polar phases). Finally, samples were dried using a Speed Vac Concentrator (Thermo Fisher, Warrington, UK) for polar phases, and a nitrogen stream for non-polar phases. Samples were stored at -80 °C until mass spectrometry analysis.

5.6.3 Ultra-Performance Liquid Chromatography Mass Spectrometry analysis

MS analysis, data processing and metabolite annotation were kindly performed by Prof Mark Viant and his team (University of Birmingham, School of Biosciences, Birmingham, UK). Samples were analysed applying two Ultra Performance Liquid Chromatography - Mass Spectrometry (UPLC-MS) methods, using a Dionex UltiMate 3000 Rapid Separation LC system (Thermo Fisher Scientific, MA, USA) coupled with a heated electrospray Q Exactive Focus mass spectrometer (Thermo Fisher Scientific, MA, USA). Polar extracts were reconstituted in acetonitrile/water (75:25) and analysed on an Accucore-150-Amide-HILIC column (100 x 2.1 mm, 2.6 µm, Thermo Fisher Scientific, MA, USA). Mobile phase A consisted of 10 mM ammonium formate and 0.1% formic acid in 95% acetonitrile/water and mobile phase B consisted of 10 mM ammonium formate and 0.1% formic acid in 50% acetonitrile/water. Flow rate was set for 0.50 $mL.min^{-1}$ with the following gradient: t=0.0, 1% B; t=1.0, 1% B; t=3.0, 15% B; t=6.0, 50% B; t=9.0, 95% B; t=10.0, 95% B; t=10.5, 1% B; t=14.0, 1% B, all changes were linear with curve = 5. The column temperature was set to 35 °C and the injection volume was 2 µL. Data were acquired in positive and negative ionisation modes separately within the mass range of 70 – 1050 *m/z* at resolution 70,000 (FWHM at *m/z* 200). Ion source parameters were set as follows: Sheath gas = 53 arbitrary units, Aux gas = 14 arbitrary units, Sweep gas = 3 arbitrary units, Spray Voltage = 3.5kV, Capillary temp. = 269 °C, Aux gas heater temp. = 438°C. Non-polar extracts were reconstituted in isopropanol/water (75:25) and analysed on a Hypersil GOLD

column (100 x 2.1mm, 1.9 µm; Thermo Fisher Scientific, MA, USA). Mobile phase A consisted of 10 mM ammonium formate and 0.1% formic acid in 60% acetonitrile/water and mobile phase B consisted of 10 mM ammonium formate and 0.1% formic acid in 90% propan-2-ol/water. Flow rate was set for 0.40 $mL.min^{-1}$ with the following gradient: t=0.0, 20% B; t=0.5, 20% B, t=8.5, 100% B; t=9.5, 100% B; t=11.5, 20% B; t=14.0, 20% B, all changes were linear with curve = 5. The column temperature was set to 55 °C and the injection volume was 2µL. Data were acquired in positive and negative ionisation mode separately within the mass range of 150 – 2000 *m/z* at resolution 70,000 (FWHM at *m/z* 200). Ion source parameters were set as follows: Sheath gas = 50 arbitrary units, Aux gas = 13 arbitrary units, Sweep gas = 3 arbitrary units, Spray Voltage = 3.5kV, Capillary temp. = 263 °C, Aux gas heater temp. = 425 °C. A Thermo Exactive Tune 2.8 SP1 build 2806 was used as instrument control software in both cases and data were acquired in profile mode. Quality control (QC) samples were analysed as the first ten injections and then every seventh injection with two QC samples at the end of the analytical batch. Two blank samples were analysed, the first as the sixth injection and then the second at the end of each batch.

5.6.3.1 Raw data processing

Raw data acquired in each analytical batch were converted from the instrument-specific format to the mzML file format applying the open access ProteoWizard software (Kessner et al., 2008) for both HILIC (Hydrophilic Interaction Liquid

Chromatography) and lipids. Deconvolution was performed with XCMS software (Smith et al., 2006) according to the following settings: Min peak width (4 for HILIC and 6 for lipids); max peak width (30); ppm (12 for polar phases and 14 for lipids); mzdiff (0.001); gapInit (0.5 for HILIC and 0.4 for lipids); gapExtend (2.4); bw (0.25); mzwid (0.01). A data matrix of metabolite features (*m/z*-retention time pairs) vs samples was built with peak areas provided where the metabolite feature was detected for each sample.

5.6.3.2 Metabolite Annotation

Putative annotation of metabolites or metabolite groups was performed by applying the PUTMEDID-LCMS workflows operating in the Taverna workflow environment (Brown et al., 2011). 12 ppm mass error for HILIC and 14 ppm mass error for Lipid data and a retention time range of 2s in feature grouping and molecular formula and metabolite matching were applied. As different metabolites can be detected with the same accurate *m/z* (e.g., isomers with the same molecular formula), multiple annotations can be observed for a single detected metabolite feature. Also, a single metabolite could be detected as multiple molecules, particularly as a different type of ion (e.g., protonated or sodiated ions). All molecules were annotated according to guidelines for reporting of chemical analysis results, specifically to Metabolomics Standards Initiative level 2 (Fiehn et al., 2007).

5.6.3.3 Quality Control and Quality Assessment

A quality assurance and quality control (QA/QC) assessment were performed to measure drift across retention time, *m/z* and signal intensity and identify potential outliers. The first five QCs were used to equilibrate the analytical system and therefore subsequently removed from the data prior to data analysis. Principal Components Analysis (PCA) was performed to assess the technical variability (measured by the replicate analysis of a pooled QC sample) and biological variability as part of the quality control process. Prior to PCA missing values in the data were replaced by applying k nearest neighbour (kNN) missing value imputation (k = 5) followed by probabilistic quotient normalisation (PQN) (Dieterle et al., 2006), and glog transformation (Motakis et al., 2006) prior to data analysis.

5.6.3.4 Peak Matrix Filtering

The data from the pooled QC samples were applied to perform QC filtering. For each metabolite feature detected, QC samples 1-5 were removed and the relative standard deviation and percentage detection rate were calculated using the remaining QC samples. Blank samples at the start and end of a run were used to remove features from non-biological origins. Any feature with an average QC intensity less than 20 times the average intensity of the blanks was removed. Any sample with >50% missing values was excluded from further analysis. Metabolite features with an RSD (spectrum-wide Relative Standard Deviations) >30% and

present in less than 90% of the QC samples were deleted from the dataset. Features with a <50% detection rate over all samples were also removed.

5.6.3.5 Statistical Analysis

5.6.3.5.1 Univariate Analysis

All Univariate statistics were performed in the R environment using the Univariate statistics function made available by Workflow4Metabolomics. Probabilistic quotient normalisation (PQN, mean QC applied) of the data was performed prior to differential analysis and t-tests ($p<0.05$) were used to identify features showing a significant difference in intensities between groups. This required the application of multiple tests (one for each metabolite feature), so a False Discovery Rate (FDR) correction was also performed.

5.6.3.5.2 Pathway analysis

In order to investigate the biological pathways where the differential metabolites between males and females and between amphipods sampled up- and downstream of the WWTP were involved, the "Pathway Analysis" tool of the online database MetaboAnalyst (Xia et al., 2015) was employed (Fig. 5.3). Since only one significantly changing metabolite between upstream and downstream was found in female amphipods, the pathway analysis for the differential metabolites between up- and downstream was only conducted using the male compound list. Differential metabolite features detected in all ionization modes

with an FDR<0.05 in univariate analyses were submitted in the database. When multiple metabolites were annotated to a metabolite feature (e.g., multiple molecular formulae annotated to a single *m/z* signal) they all were submitted. Since the available species for both enrichment and pathway analyses were limited to *Homo sapiens* and the most used model organisms, *Drosophila melanogaster* metabolic database was interrogated. Hypergeometric test and relative-betweeness centrality were set for the overrepresentation and pathway topology analyses, respectively (Xia et al., 2011). Similarly to an enrichment analysis conducted on a set of genes (Chapter 4.3.5.2), hypergeometric test is used to evaluate whether a metabolite set is represented more than expected by chance within the interrogated compound database (Xia et al., 2011). The p-value in output represents the statistical probability of the enrichment. On the other hand, the pathway topology analysis is aimed to give a measure of importance of the matched metabolites within the enriched metabolic network. The output is a numeric value, which is based on the number of shortest paths going through the matched metabolites (nodes) within the metabolic network (Aittokallio et al., 2006).

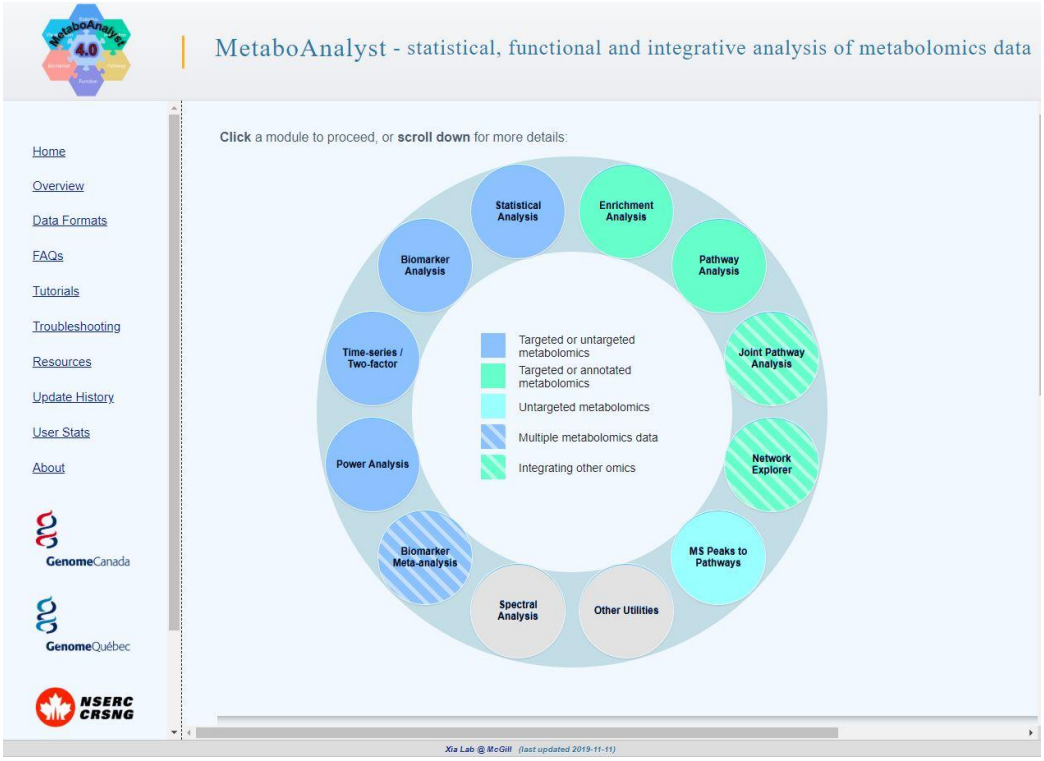

Fig. 5.3: MetaboAnalyst database homepage (https://www.metaboanalyst.ca/). The tools provided by the database can be used for statistical, functional and integrative analyses of high-throughput metabolomic data. Although the available species for enrichment and pathway analyses being limited to human and the most common model organisms, this database is a valuable resource for both targeted and untargeted metabolomic functional analyses.

5.7 Results

5.7.1 Annotation and changing metabolites

m/z values in output from the mass spectrometry analysis were annotated using PUTMEDID-LCMS software (Brown et al., 2011). The overall annotation rates can be seen in Tab. 5.1.

US - DS Males **A**

	Tot Features	Unidentified Features	% Annotation	Outliers
HILIC-NEG	5232	608	88.38	0
HILIC-POS	6439	825	87.19	1
LIPID-NEG	2976	372	87.62	0
LIPID-POS	7271	1017	86.01	1

US - DS Females **B**

	Tot Features	Unidentified Features	% Annotation	Outliers
HILIC-NEG	5206	607	88.34	1
HILIC-POS	6380	812	87.27	1
LIPID-NEG	2962	370	87.51	1
LIPID-POS	7203	1004	86.06	0

Males - Females				C
	Tot Features	**Unidentified Features**	**% Annotation**	**Outliers**
HILIC-NEG	5242	610	88.36	3
HILIC-POS	6450	827	87.18	1
LIPID-NEG	2978	372	87.51	2
LIPID-POS	7294	1020	86.01	2

Tab. 5.1: Annotation rates calculated for "Upstream Male – Downstream Male" (A), "Upstream Female – Downstream Female" (B) and "Total Male – Total Female" (C) comparisons. The data are shown for each ionisation mode used. "Tot Features": Total number of detected *m/z* signals; "Unidentified Features": *m/z* signals not identified in PUTMEDID, including both unidentified products and unmatched ion adducts; "% Annotation": overall percentage of annotation; "Outliers": number of samples outside of the 95% confidence interval based on the first two principal components in the PCA analysis. HILIC-NEG (Hydrophilic Negative), HILIC-POS (Hydrophilic Positive), LIPID-NEG (Lipids Negative), LIPID-POS (Lipids Positive).

A total of 141 and 1 features were found statistically changing (t-test, $p<0.05$) when comparing up- and downstream populations for males and females, respectively (Tab. 5.2) (File S5.1-5.8 – Appendix B). Looking at the changing metabolites between males and females, a total of 170 statistically changing

features was found (t-test, $p<0.05$) using hydrophilic positive ionisation mode (Tab 5.3) (File S5.10 -Appendix B). No statistically changing metabolites between males and females were found using hydrophilic negative (File S5.9 – Appendix B), lipids negative (File S5.11 – Appendix B) and lipids positive (File S5.11 – Appendix B) ionisation modes.

US - DS (Male)	Differential features
HILIC-NEG	28
HILIC-POS	69
LIPID-NEG	0
LIPID-POS	44

US - DS (Female)	Differential features
HILIC-NEG	0
HILIC-POS	0
LIPID-NEG	0
LIPID-POS	1

Tab. 5.2: Number of statistically changing metabolite features found in upstream-downstream (US-DS) comparison in each ionisation mode used, in males (left) and females (right). HILIC-NEG (Hydrophilic Negative), HILIC-POS (Hydrophilic Positive), LIPID-NEG (Lipids Negative), LIPID-POS (Lipids Positive).

Total Males - Total Females	Differential features
HILIC-NEG	0
HILIC-POS	170
LIPID-NEG	0
LIPID-POS	0

Tab. 5.3: Number of statistically changing metabolite features found in males-females comparison in each ionisation mode used. HILIC-NEG (Hydrophilic Negative), HILIC-POS (Hydrophilic Positive), LIPID-NEG (Lipids Negative), LIPID-POS (Lipids Positive).

5.7.2 Principal component analysis

PCA plots were built to check whether a distinct clustering of upstream and downstream samples (Fig. 5.4&5.5) as well as between males and females (Fig. 5.6) could be observed. Samples that were outside of the 95% confidence interval (dashed grey line in Fig. 5.4 – 5.6) based on the first two principal components were classified as outliers and excluded from the analysis. No discrete clustering was observed when plotting the first two components of upstream-downstream comparison, for the metabolites detected in all the ionisation modes (Fig 5.4&5.5).

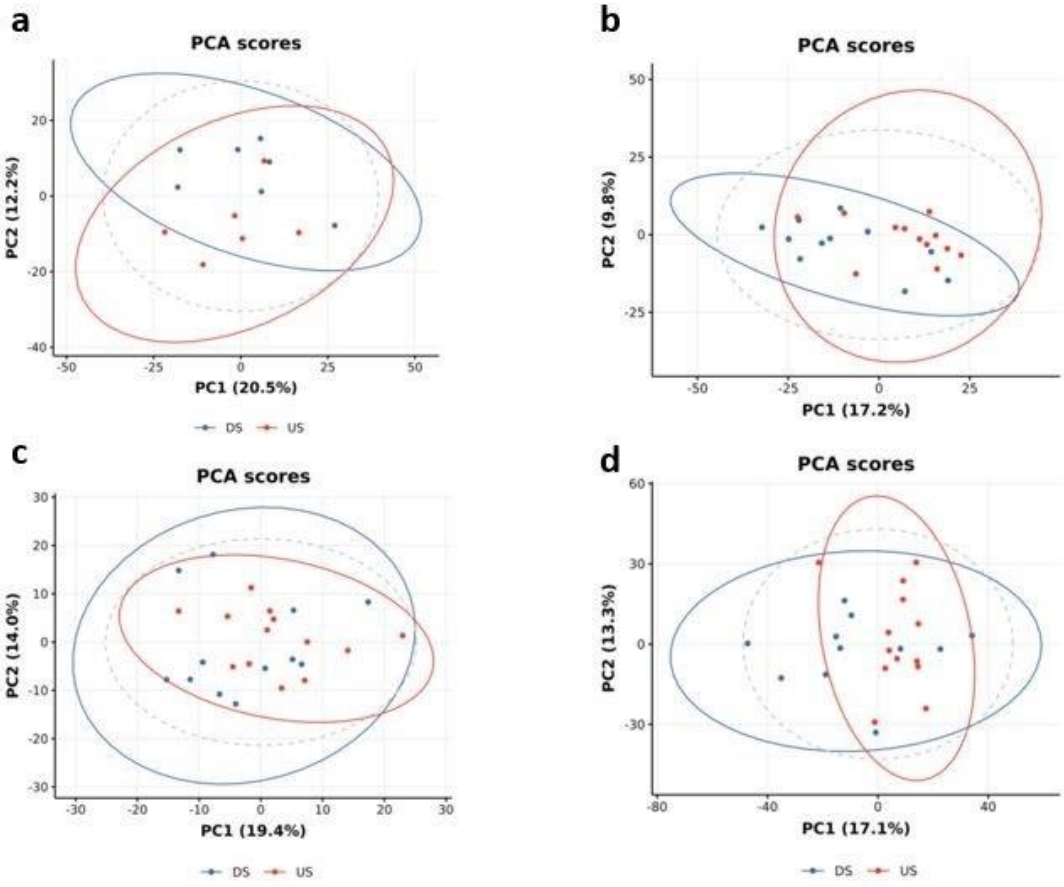

Fig. 5.4: PCA plots for components 1 and 2 of the upstream (US) – downstream (DS) metabolomic comparison in males. The analysis included metabolites detected in each ionisation mode used: a) hydrophilic negative; b) hydrophilic positive; c) lipids negative; d) lipids positive. Red and blue ellipses represent the 95% confidence interval for upstream and downstream groups, respectively. The grey dashed ellipse represents the 95% interval for all samples (ignoring group).

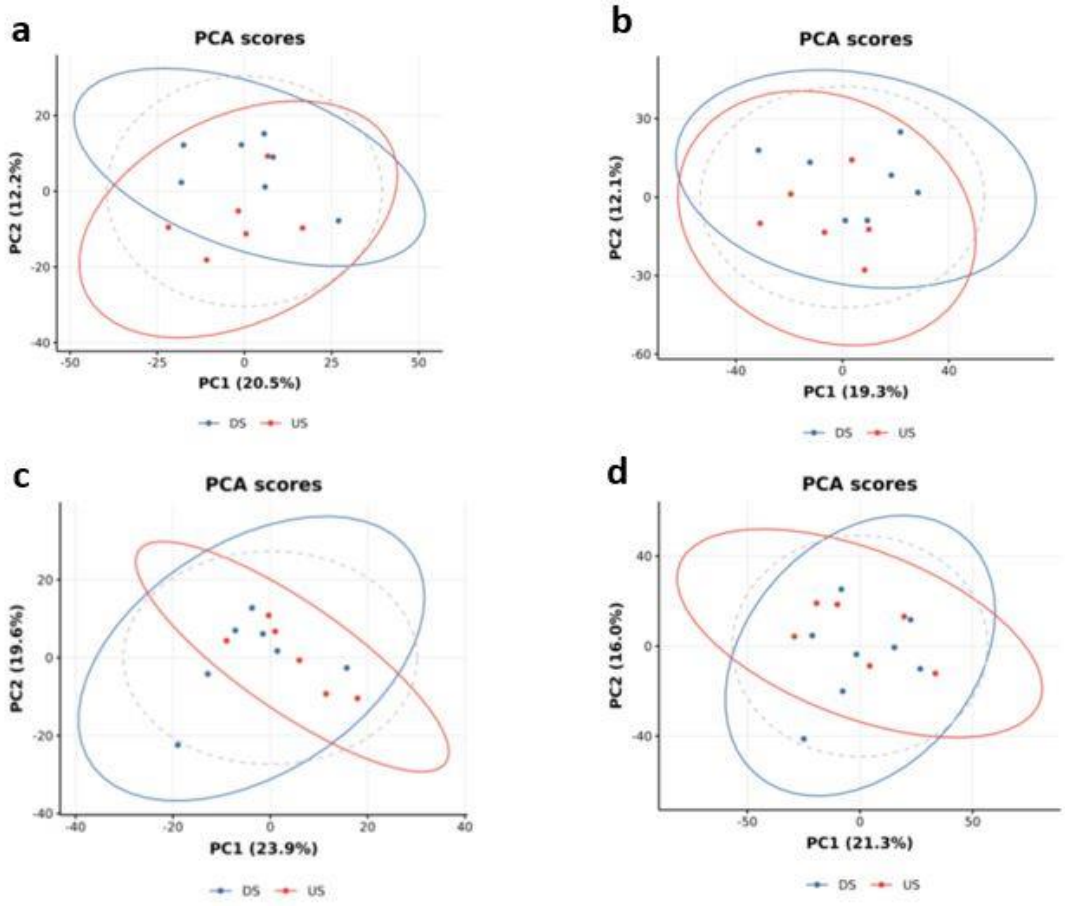

Fig. 5.5: PCA plots for components 1 and 2 of the upstream (US) – downstream (DS) metabolomic comparison in females. The analysis included metabolites detected in each ionisation mode used: a) hydrophilic negative; b) hydrophilic positive; c) lipids negative; d) lipids positive. Red and blue ellipses represent the 95% confidence interval for upstream and downstream groups, respectively. The grey dashed ellipse represents the 95% interval for all samples (ignoring group).

No evident cluster for male and female samples was found when plotting the first two components of males-females comparison, including the metabolites detected in all ionisation modes (Fig. 5.6).

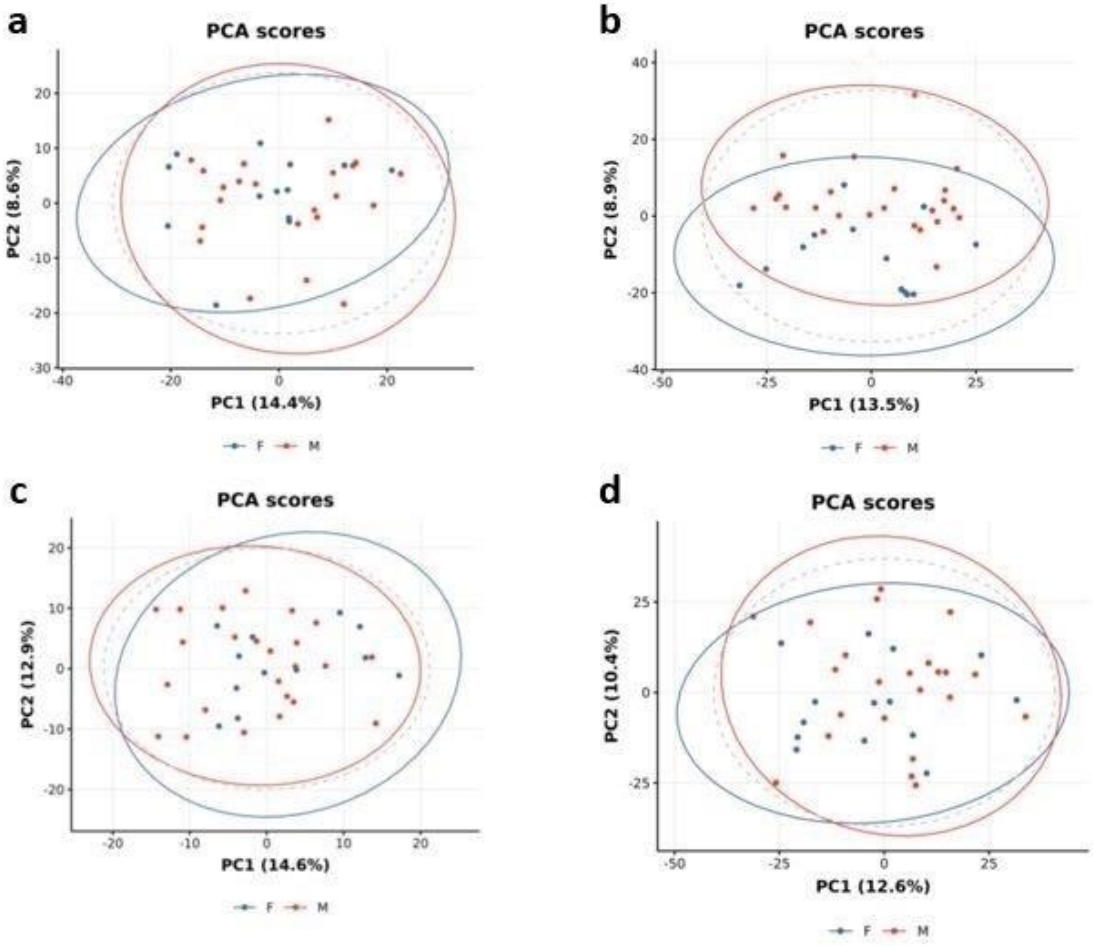

Fig. 5.6: PCA plots for components 1 and 2 of the male (M) – female (F) metabolomic comparison. The analysis included metabolites detected in both upstream and downstream samples, in each ionisation mode used: a) hydrophilic negative; b) hydrophilic positive; c) lipids negative; d) lipids positive. Red and blue ellipses represent the 95% confidence interval for upstream and downstream groups, respectively. The grey dashed ellipse represents the 95% interval for all samples (ignoring group).

5.7.3 Pathway analyses

5.7.3.1 Pathway analysis: upstream-downstream

Submitting the male differential metabolite compound name list in MetaboAnalyst, the database was not able to recognise 309 compounds out of 860 total metabolite features (35.93%). The submitted compound names matched within 22 *D. melanogaster* metabolic pathways (Tab. 5.4) (Fig. 5.7). “One carbon pool by folate” was the only metabolic pathway with a statistically significant enrichment in the dataset ($p<0.05$). The components within the “One carbon pool by folate” pathway in *D. melanogaster* were explored consulting the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Fig. 5.8).

Pathway name	Match status	p	-log(p)	Impact
One carbon pool by folate	3/8	0.0262	3.642	0.36061
Ascorbate and aldarate metabolism	1/6	0.42313	0.86007	0
Taurine and hypotaurine metabolism	1/7	0.47384	0.74689	0.2
Valine, leucine and isoleucine biosynthesis	1/8	0.52013	0.65367	0
α-linoleic acid metabolism	1/8	0.52013	0.65367	0
Retinol metabolism	1/8	0.52013	0.65367	0
Sulfur metabolism	1/8	0.52013	0.65367	0.31915
Caffeine metabolism	1/10	0.60097	0.50921	0
Arachidonic acid metabolism	1/13	0.69763	0.36006	0
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	1/13	0.69763	0.36006	0.00476
Starch and sucrose metabolism	1/14	0.72439	0.32243	0.03918
Tryptophan metabolism	2/30	0.75516	0.28082	0.12451
Glycerophospholipid metabolism	2/32	0.7868	0.23979	0.12758
Sphingolipid metabolism	1/18	0.80992	0.21082	0.1875
Glutathione metabolism	1/26	0.90999	0.094323	0

Galactose metabolism	1/27	0.91805	0.085499	0.02976
Folate biosynthesis	1/28	0.9254	0.077526	0.00687
Aminoacyl-tRNA biosynthesis	2/48	0.9349	0.067312	0.08334
Glycine, serine and threonine metabolism	1/30	0.9382	0.063792	0.01087
Valine, leucine and isoleucine degradation	1/38	0.971	0.029427	0.01038
Pyrimidine metabolism	1/40	0.97602	0.024269	0.05771
Purine metabolism	1/63	0.99738	0.0026207	0.03528

Tab 5.4: *Drosophila melanogaster* metabolic pathways obtained submitting the male metabolites differentially expressed between up- and downstream samples in MetaboAnalyst. Pathway name: name of the metabolic pathway; Match status: number of metabolites matching to a particular pathway in ratio with the total number of metabolites of that pathway; p: statistical significance of the enrichment expressed as a p-value; -log(p): -log transformed p-value; Impact: measure of importance of the matched metabolites within the metabolic pathway.

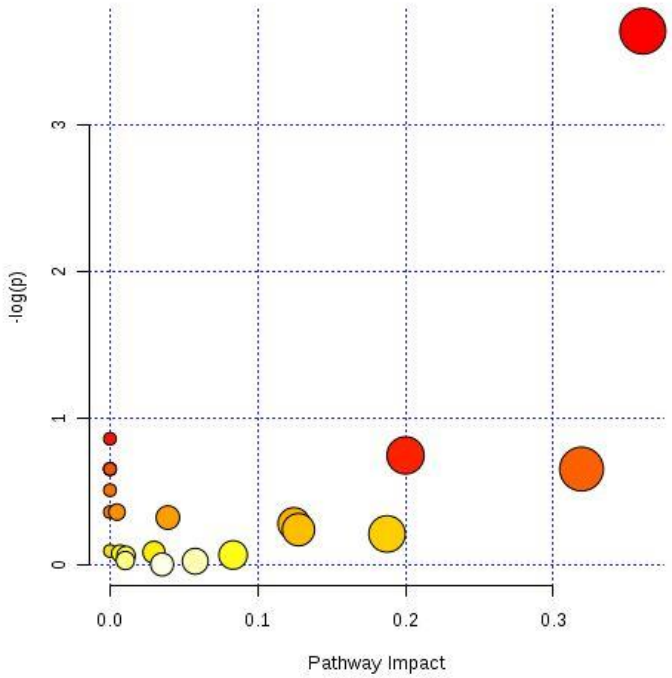

Fig. 5.7: Bubble chart showing the results of the pathway analysis conducted on the differential metabolites between male amphipods sampled upstream and downstream of the WWTP. Pathway impact of the matched metabolites is shown on the x-axis and -log(p-values) on the y-axis. The bubble colour represents the p-values for each pathway, in a scale from red (lowest) to white (highest). The bubble on the top right corresponds to “One carbon pool by folate” pathway.

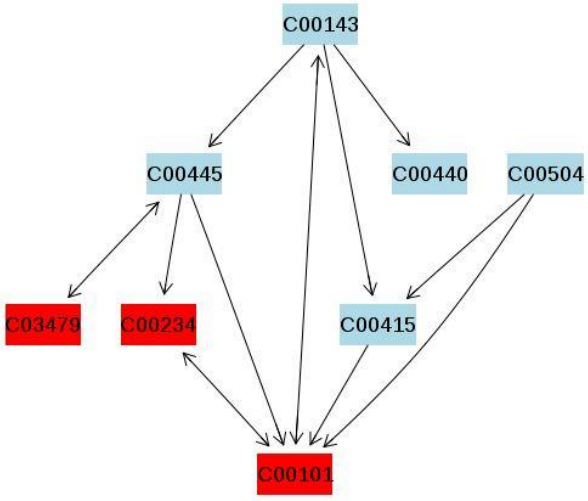

Fig. 5.8: "One carbon pool by folate" KEGG metabolic network in *Drosophila melanogaster*. The boxes contain the KEGG codes of the metabolites involved in this pathway. The metabolites of the male differential metabolites (upstream vs downstream) that matched in this pathway are highlighted in red (C03479: Folinic acid; C00234: 10-Formyltetrahydrofolate; C00101: Tetrahydropholate).

5.7.3.2 Pathway analysis: males-females

Similarly to the strategy used to investigate the metabolic pathways where the differential metabolites between amphipods sampled up- and downstream were involved, the differential compounds between males and females were submitted in MetaboAnalyst, using the "Pathway Analysis" tool. Also in this case, *D. melanogaster* metabolic database was interrogated. Although the database did not recognise 521 compounds out of a total of 1156 metabolite features (45.07%) (including metabolites detected in all ionisation modes), a statistically significant

enrichment in “Insect hormone biosynthesis” pathway was found (Tab. 5.5) (Fig. 5.9&5.10).

Pathway name	Match status	p	-log(p)	Impact
Insect hormone biosynthesis	5/21	0.020184	3.9029	0.29932
Arachidonic acid metabolism	2/13	0.274	1.2946	0
Sphingolipid metabolism	2/18	0.42236	0.86189	0.07917
Glycerophospholipid metabolism	3/32	0.42236	0.75451	0.1712
α-Linolenic acid metabolism	1/8	0.48324	0.72725	0
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	1/13	0.65885	0.41726	0.00476
Pentose and glucuronate interconversions	1/16	0.73435	0.30877	0.14062
Fatty acid degradation	2/38	0.81822	0.20063	0
Tyrosine metabolism	1/33	0.93654	0.065565	0.08677
Fatty acid elongation	1/37	0.95485	0.0462	0
Pyrimidine metabolism	1/40	0.96506	0.035568	0.00761
Fatty acid metabolism	1/43	0.97298	0.027395	0.01681

Tab. 5.5: *Drosophila melanogaster* metabolic pathways obtained submitting the differential metabolites between male and female amphipods in MetaboAnalyst. “Pathway name”: name of the metabolic pathway; “Match status”: number of metabolites matching to a particular pathway

in ratio with the total number of metabolites of that pathway; “p”: p-value; “-log(p)”: -log transformed p-value; “Impact”: measure of importance of the matched metabolites within the metabolic pathway.

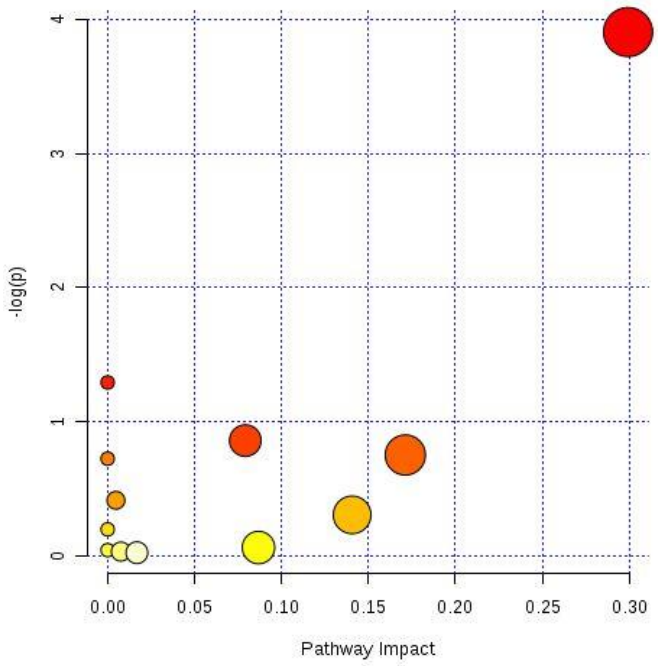

Fig. 5.9: Bubble chart showing the results of the pathway analysis conducted on the differential metabolites between male and female amphipods. Pathway impact of the matched metabolites is shown on the x-axis and -log(p-values) on the y-axis. The bubble colour represents the p-values for each pathway, in a scale from red (lowest) to white (highest). The bubble on the top right corresponds to “Insect hormone biosynthesis” pathway.

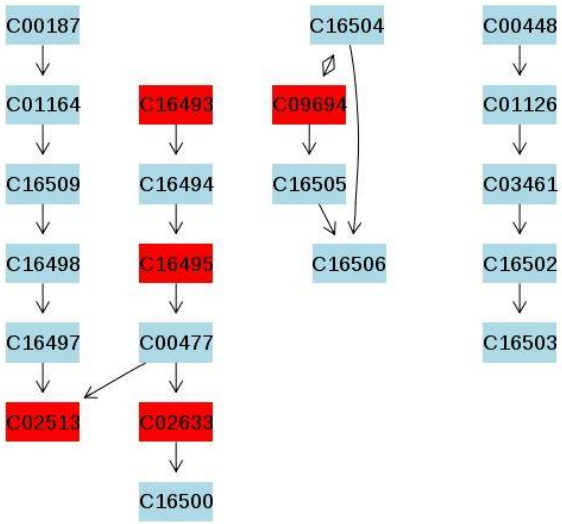

Fig. 5.10: "Insect hormone biosynthesis" KEGG metabolic network in *Drosophila melanogaster*. The boxes contain the KEGG codes of the metabolites involved in this pathway. The differential metabolites between male and female amphipods that matched in this pathway are highlighted in red (C16493: 3β,5β-ketodiol; C09694: Juvenile hormone III; C16495: 2-Deoxyecdysone; C02513: 3-Dehydroecdysone; C02633: 20-Hydroxyecdysone).

5.8 Discussion

Within the field of chemical safety and risk assessment, there is a need to assess toxicity of a continuously growing number of chemicals using finite resources while addressing the ethical concerns surrounding the use of reliable animal alternatives (Krewski, et al., 2010; SCENIHR, 2012). “Omics” technologies enable researchers to assess the responses of tens of thousands of genes and their products in a single sample (Aardema et al., 2002). In combination with advances in statistical data analysis and an exponential increase of databases specifically designed for data obtained from different platforms, “omics” datasets are used to “learn” the structure of biological pathways from observational data (Mitra et al., 2013). In addition, the molecular picture as well as the set of biological responses become even more detailed when using multiple “omics” platforms in parallel. For example, Trapp et al., (2016) used a proteogenomics strategy (genomic and proteomic platforms) to generate a molecular report on the reproductive tissues of *G. fossarum* females. Beale et al., (2017) applied both metagenomics and metabolomics approaches to an Australian urban river system in order to investigate surface water quality and characterise the bacterial population changes in response to water contaminants. In the present project, two “omics” platforms (transcriptomics and metabolomics) were used to investigate the potential molecular differences between *G. fossarum* amphipods sampled up- and downstream of a sewage effluent located at northern Switzerland. The molecular fingerprints of both male and female amphipods were also acquired, in order to

explore the metabolic differences between the genders in this ecologically important species. A mass spectrometry analysis using a UPLC-MS/MS platform was performed on the total metabolite mixtures extracted from male and female amphipods, sampled both upstream and downstream of a WWTP. The metabolic annotation was performed by applying the PUTMEDID-LCMS workflow (Brown et al., 2011) and a set of potential metabolites were assigned to each metabolite feature. Since an untargeted metabolomics approach was used (section 5.1), there was no way to assign a single metabolite to each metabolic feature detected. Therefore, the whole set of metabolites corresponding to each differential metabolite feature detected in "Upstream vs Downstream" and "Males vs Females" comparisons were submitted in the MetaboAnalyst database. To date, no amphipod specific metabolic databases are available, hence the fruit fly (*Drosophila melanogaster)* metabolic database was interrogated. This species was chosen, since it is a widely used arthropod for molecular annotation in a vast range of studies on amphipods employing "omics" platforms (Bossus et al. 2014; Trapp et al., 2014; Poynton et al., 2018; Caputo et al., 2020).

5.8.1 Differential metabolites between upstream and downstream

With an average annotation rate of 87.29%, calculated including the annotation rates of all the ionisation mode used, an excellent coverage of the metabolic fingerprint in females was obtained for the "Upstream-Downstream" comparison (Tab. 5.1 A,B). However, no significant metabolic changes between upstream and

downstream were detected and no distinct clustering could be observed in the PCA plots for female amphipods (Fig. 5.5). In fact, only one significantly changing metabolite (ω-6 long-chain polyunsaturated fatty acid: heneicosadienoic acid) was found in lipid positive ionization mode in females, when comparing upstream and downstream samples (File S5.8 – Appendix B). The literature shows several studies investigating the biological functions of polyunsaturated long-chain fatty acids in crustaceans. In particular, they appear to be mainly involved in growth (Kolanowski et al., 2007) and development (Romano et al., 2016). However, their content in freshwater gammarids was found to be the lowest compared to other gammarids species (Baeza-Rojano et al., 2014). Several studies conducted in fish demonstrate significant differences in lipid composition between control animals and animals exposed to whole sewage effluents (Samuelsson et al., 2011; Al-Salhi et al., 2012; David et al., 2017). Although both fish species and chemical compositions of the tested effluents were different, the mentioned studies brought to light that evident and biologically relevant changes in the lipidome of animals exposed to complex contaminant mixtures occur. For instance, a reduction in prostaglandins (molecules formed from arachidonic acid, a membrane phospholipid) was observed in roach (*Rutilus rutilus*) exposed to a wastewater effluent containing pharmaceuticals (David et al., 2017). An impairment of the physiological ratios between different cholesterol-related lipoproteins (i.e., HDL, LDL, VLDL) was recorded when exposing rainbow trout (*Oncorhynchus mykiss*) to whole sewage effluents (Samuelsson et al., 2011).

Perturbations in plasma concentrations of lysophospholipids and sphingosine (key components of cell membranes) were found when rainbow trout (*Oncorhynchus mykiss*) were exposed to wastewaters containing a variety of xenobiotics, including surfactants, phenoxyphenols and steroidal alkaloids (Al-Salhi et al., 2012). However, in the present project, no differential lipid compounds were detected other than heneicosadienoic acid, when comparing female amphipods between up- and downstream sites. In fact, it was not possible to perform a pathway enrichment analysis using a single compound. Thus, different signal intensities corresponding to heneicosadienoic acid, detected between upstream and downstream samples, cannot be considered as biologically relevant. This mirrored the results of the functional analysis performed on the transcriptomic data, whereby no significantly enriched pathways were found when running a GO-complete "Biological process" analysis on the differentially expressed genes between up- and downstream female amphipods (Chapter 4.5.2.3).

A good average annotation rate (87.30%) was found when submitting the metabolic features detected in "Male Upstream - Male Downstream" comparison in the PUTMEDID database. Similar plots for male amphipods were found when comparing upstream and downstream samples, with no distinct clustering in the PCA plots (Fig. 5.4). However, hints of metabolic changes were detected in males in all ionisation modes except lipid negative, in the comparative analysis. In total, 141 metabolic features (accounting for a total of 860 potential metabolites) were

found statistically different between male amphipods sampled at up- and downstream sites (t-test, $p<0.05$) (Tab. 5.4). This was a sufficient number of compounds to conduct a pathway analysis, to check potential enrichments in the *D. melanogaster* metabolic network. “One carbon pool by folate” turned out to be the only statistically enriched pathway ($p<0.05$) (Tab. 5.4). One-carbon (1C) metabolism, mediated by the folate cofactor, is a universal metabolic process that activates and transfers 1C units for biosynthetic processes including purine and thymidine synthesis and homocysteine remethylation (Ducker et al., 2017). It has been shown to support a broad set of transformations known as one-carbon (1C) metabolism. Whereas most bacteria, yeasts and plants can synthesize folate, animals require dietary folate intake (Ducker et al., 2017). Although there are no studies specifically investigating this important metabolic pathway in crustaceans, an interesting study by Shiau et al., (2001) demonstrates the relevance of folate metabolism in the Asian tiger prawn *Penaeus monodon.* The authors conducted a feeding trial to determine the dietary folic acid requirement of juveniles. When juvenile *P. monodon* were fed a folic acid-free basal diet, there was a clear need for dietary folic acid supplementation. In fact, the weight gain was evident when gradually increasing the amount of folic acid in their dietary intake. Of greater interest for the present work, a study conducted by Wang et al., (2017) highlights that 1C metabolism may be directly involved in the stress response of the banana shrimp *Fenneropenaeus merguiensis* following exposure to ammonia in water. The authors performed an RNA sequencing experiment followed by a DGE analysis

between control animals and animals exposed to environmentally relevant concentrations of ammonia nitrogen. Subsequently, they used the differentially expressed genes to perform a pathway analysis in the KEGG database and “one carbon pool by folate” was found among the enriched pathways. Despite a low percentage of genes (0.23%) being found in the folate metabolic network, the fact that there were differentially expressed genes involved in this pathway along with genes involved with the response to the toxic pressure (e.g., cytochrome p450) (David et al., 2003; Del Brio et al., 2019), supports a potential involvement of 1C metabolism in the responses to xenobiotic exposure. In the present study, 1C metabolism was the only significantly enriched metabolic pathway when submitting the differential metabolites between up- and downstream samples in MetaboAnalyst. Therefore, targeted investigations to elucidate in detail the roles of this metabolic pathway in amphipod stress response will be fundamental. Moreover, although the database did not recognise ~35% of the differential compounds between up- and downstream samples, other important metabolic pathways came out from the enrichment analysis (e.g., caffeine metabolism, arachidonic acid metabolism, tryptophan metabolism) (Tab. 5.4). An enrichment in these pathways may indicate different types of stress responses of the amphipods, such as changes in lipid metabolism and neurological processes. However, the p-values for these pathways were all above the threshold of statistical significance. This might be due to the low number of matched metabolites in the database rather than the presence of false positives. Although

differentially expressed male genes between up- and downstream closely associated to 1C-metabolism were not found in the transcriptomic analysis, a number of genes encoding proteins at mitochondrial localization (e.g., *DHSD, MTCH2, PCCA, DHE3*) were found (File S4.15 – Appendix B). Considering that part of the 1C-metabolism takes place a mitochondrial level (Zong et al., 2016), metabolomic and transcriptomic data taken together may indicate a mitochondrial stress caused by chronic exposure to a complex chemical mixture.

5.8.2 Differential metabolites between males and females

Although no distinct clustering of male and female samples was observed in the PCA plots, a total of 170 metabolic features (accounting for a total of 1156 potential metabolites) were differentially detected in the comparative analysis in hydrophilic positive ionisation mode (t-test, $p<0.05$) (File S5.10 – Appendix B). The whole set of metabolites was submitted in MetaboAnalyst to run a pathway analysis using *D. melanogaster* metabolic database. "Insect hormone biosynthesis" was found statistically enriched with 5 out of a total of 21 metabolites matching within this pathway ($p<0.05$) (Tab. 5.5). Interestingly, Juvenile hormone (JH) III was among the matching metabolites (Fig. 5.10). JH has central roles in the regulation of insect development and reproduction (Laufer et l., 1987). The physiology of its homolog in crustacean (i.e., methyl farnesoate) has been described (Schneiderman et al., 1958; Homola et al., 1977; Laufer et al., 1987). In particular, the sesquiterpenoid methyl farnesoate (MF) is synthesized by

crustacean mandibular organs and is present in the haemolymph (Homola et al., 1977). MF has been shown to be involved in a number of processes in crustaceans, such as stimulation of general protein synthesis, promotion of the moult cycle and reproduction (Homola et al., 1977). The other metabolites that matched in the *D. melanogaster* hormone biosynthesis network were 3β,5β-ketodiol, deoxyecdysone, 3-dehydroecdysone and 20-hydroxyecdysone (Fig. 5.10). These compounds are known as ecdysteroids and in insects, they have been shown to be involved in the timing of moulting and metamorphosis (Koolman, 1990). Fine coordination mechanisms of moulting have been described in female amphipods, especially during the reproduction period (Hyne et al., 2011). In fact, coordination of moulting of the rigid exoskeleton with the ovarian cycle facilitates the movement of the oocytes through the oviduct into the marsupium and the pairing process with the male (Scheader, 1983). On the other hand, males are considered available for mating during most of their moult cycle (Sutcliffe, 1992). The presence of ecdysteroids among the differential metabolites between male and female amphipods is therefore unsurprising and highlights the molecular differences between the genders, particularly in moulting and reproduction processes (Hyne et al., 2011). This was confirmed by a DGE analysis conducted on the changing genes between males and females, where genes encoding proteins with GO terms associated with reproduction and gametogenesis in *D. melanogaster* were found (Chapter 4.4.7). Future targeted metabolomics analyses investigating moulting and reproduction metabolites in amphipods may provide a

deeper understanding on the structure of these pathways and highlight homologies and discrepancies with other arthropods. The metabolic extracts for the targeted analyses should come from amphipods in the same development/reproduction stage, in order to study the characteristic metabolic fingerprint of each stage.

Chapter 6 - General Discussion

The regulations that aim to limit the impact of pollution on the environment have evolved in response to a massive increase in the use of substances, such as pesticides and pharmaceuticals, all over the world. For instance, in the European Union, 327642 tons of pesticides were sold in 2001 and 2300 tons were sold just in the Swiss territory in 2008 (Wenger et al., 2012). Most toxic among this large group of chemicals are the synthetic organic molecules, which share the common property of damaging the nervous system of animals (Werner et al., 2012). Long before the onset of environmental regulation, biological tools based on indicator species were used to detect environmental hazards, such as the "canary in the coal mine" used to warn miners of dangerous levels of carbon monoxide and methane. Standardized biological methods to measure water quality developed quickly after the US EPA initiated a national policy in 1984 to control toxic substances based on a water quality approach. The issuance of permits for effluent discharges into surface waters was subsequently tied to whole effluent testing using standardized toxicity tests. Such tests enable the direct measurement of toxicity independent of the number of causative chemicals or mixture effects. In addition, aquatic community indices such as the saprobic (Kolkwitz et al., 1909) or SPEAR (SPEcies At Risk) indices (Liess et al., 2008) integrate the effects of all chemical, physical, and biological stressors acting in a system. Furthermore, chronic effects in the form of sub-lethal damage to organisms can be observed at concentrations found

in the environment. These include impairment of the reproductive (Sumpter, 2005) or immune system (Arkoosh et al., 2001), and genetic (Shugart, 1995), developmental, and behavioural changes (Weis et al., 1995, Sandahl et al., 2007). They can severely reduce ecological fitness and ultimately survival, since the individual must be able to successfully compete with others for food, avoid predation, reproduce, and cope with pathogens and other environmental stressors. Such effects are not easily detected and can act for long periods of time before being recognized.

In general, because small amounts of anthropogenic compounds are gradually bioaccumulated by aquatic species and only slowly excreted, threat of chronic poisoning and degenerative changes has become of primary concern over the last 50 years (Werner et al., 2012). In order to prevent further deterioration of river quality in response to an increasing number of contaminants released into freshwaters through WWTPs, European countries, including Switzerland, financed important river monitoring plans. In addition, measures to improve the sewage treatment technologies were also proposed (EU WFD, 2000; Eggen et al., 2014; Busch et al., 2016).

Due to a central role in aquatic food web and a marked sensitivity to xenobiotics, the amphipod *Gammarus fossarum* has been defined as an ideal biomarker species for ecotoxicological risk assessment in freshwater systems (Adam et al.,

2010; Kunz et al., 2010; Trapp et al., 2015; Wigh et al., 2017). Previous studies have shown a wide range of biological perturbations when *G. fossarum* amphipods were exposed to single xenobiotics (Adam et al., 2010; Besse et al., 2013; Bossus et al., 2014; Trapp et al., 2015) or chemical mixtures (Gouveia et al., 2017; Wigh et al., 2017). However, despite the importance of this species in monitoring aquatic ecosystems, its genome has not yet been completely annotated. Although some progress has been made in obtaining molecular information on amphipods using high-throughput "omics" platforms (Gismondi et al., 2016; Gómez-Canela et al., 2016; Poynton et al., 2018; Cogne et al., 2019; Caputo et al., 2020), further studies are needed to make these ecologically important species adequately represented in terms of "omics" data sets. This will facilitate the study of the pathways affected by exposure to the contaminants released in aquatic environments, ultimately helping to develop new ecotoxicological biomarkers.

In the present study, the complete transcriptome of *G. fossarum* was assembled and annotated *de novo*. In addition, the differences in population structure, gene expression and in metabolome between amphipods sampled upstream and downstream of a Swiss WWTP were explored. The main aims were:

- to generate a data set (transcriptome) that will be useful in future studies to develop new ecotoxicological biomarkers in amphipods

- to evaluate the biological impact of a chronic exposure to an anthropogenic chemical mixture on amphipod gene expression and metabolomic profiles.

Gene expression and metabolomic differences between male and female amphipods were also investigated, in order to provide a molecular data set useful for developing new sex-specific makers in crustaceans.

In this chapter, key findings of previous chapters were reviewed and the main results of population, transcriptomic and metabolomic analyses are discussed. The data collected in the population analyses were compared with the outputs of the "omics" platforms, highlighting homologies and discrepancies. Based on the limitations of the methodologies presented, future research strategies to investigate the impact of xenobiotics on amphipods were also proposed.

6.1 Chemical analysis

One objective of the present project was to evaluate the overall toxic stress in the stream studied. To achieve this, the concentrations of 55 compounds commonly detected in sewage effluents, including several classes of pharmaceuticals and pesticides, were measured in both surface water and amphipods sampled upstream and downstream of the WWTP in September 2017 (Chapter 2.5.3).

Overall, the concentrations detected in water and amphipods were in a similar range as values reported in other studies investigating Swiss rivers (Munz et al., 2017; Munz et al., 2018). Unsurprisingly, ~37% of the number of analysed substances at the upstream site was detectable in water samples, in contrast to ~67% of detectable compounds at the downstream site. The analysed substances were split into chemical classes (Fig. 2.10) and a mean of the concentration values of the single compounds belonging to each class was calculated (Fig. 2.11). The average concentrations of 9 out of 11 chemical classes analysed were higher at the downstream site in water samples (Fig. 2.11). On the other hand, only ~5% and ~18% of the evaluated substances were above the limit of detection in gammarids collected from upstream and downstream sites, respectively (Fig. 2.11). Only 2 out of 11 chemical classes analysed (i.e., pesticides and neuroactive drugs) showed a higher average concentration at the downstream site in amphipod tissues (Tab. S2.1 – Appendix B) (Fig. 2.11).

In order to check whether bioaccumulation was occurring in amphipods, apparent BAFs were calculated for the substances found above the detection limit in both water samples and gammarids (Tab. S2.1 – Appendix B). Surprisingly, imidacloprid (a neonicotinoid insecticide) was the only compound with a BAF value above the threshold of potential bioaccumulation (2000 L/Kg) (EC - No1907/2006) at downstream site (Tab. S2.1 – Appendix B). Neonicotinoids are very polar substances (log K_{ow} = 0.6-1.3), hence the bioaccumulation rate in organisms is

supposed to be lower compared to non-polar compounds, which more easily pass through the phospholipidic layer of cell membranes (Munz et al., 2017; Munz et al., 2018). However, despite a relatively low rate of bioaccumulation and fast elimination of neonicotinoid pesticides described in gammarids (Ashauer et al., 2017), these compounds are known to be highly toxic toward aquatic invertebrates (Beketov et al., 2008; Munz et al., 2018). In a previous study investigating the concentrations of a wide range of xenobiotics in *G. fossarum* amphipods collected in several Swiss rivers, citalopram (a polar neuroactive drug) and thiacloprid (a neonicotinoid pesticide) were found among the substances with highest detection frequencies (Munz et al., 2018). Interestingly, Englert et al., (2017) investigated the accumulation of neonicotinoids in gammarids and revealed a high impact of dietary exposure on the accumulation process. The authors showed higher bioaccumulation rates when gammarids were exposed to contaminated water and fed on contaminated leaves compared to water exposure only. In addition, although highly hydrophilic compounds tend to have lower rates of bioaccumulation, their concentrations can substantially vary between seasons (Spycher et al., 2018). For instance, pesticide applications are mainly performed in spring and summer and concentration peaks for these substances can be detected in late summer (August - September). This may explain the detection of imidacloprid in gammarids in this study.

Since benthic organisms such as gammarids continuously take up substances from water and sediments, they can act as a passive sampler for bioavailable substances (Vrana et al., 2005), taking into account that some substances are metabolised more quickly than others. Thus, they can be used as a tool in the evaluation of the toxic pressure by analysing the substances that are accumulated in the tissues. Previous studies have applied the calculation of the toxic units (TUs) to translate chemical concentrations in water into ecotoxicologically relevant and comparable values (Munz et al., 2018; Könemann et al., 2019). TU values are calculated as the ratio between the measured concentration of the contaminants in surface water over known toxicity endpoints (e.g., acute EC_{50} - 48h). Based on the assumption of toxicity additivity (Warne et al., 1995), the TU values can be summed up providing a measure of the toxic pressure of contaminant mixtures on the organisms studied. In general, if the threshold value of −3.0 is exceeded by the logarithm of the sum of TU values, chronic effects cannot be excluded (Liess et al., 2008). In this study, TUs were determined for all the detected compounds for which EC_{50} values were available either in the literature or in online databases (12 at upstream site and 19 at downstream) to quantify the toxic pressure in the stream. TU values were calculated as the ratio between the measured concentration of the contaminants in surface water over acute EC_{50} - 48h values for either *G. pulex*, or, if no effect data was available, *Daphnia magna* (Tab. S2.1 – Appendix B). The single TUs were summed up to sumTU (eq. 2.2) and, in accordance with the literature

(Liess et al., 2008; Könemann et al., 2019), a threshold of −3.0 was set to predict potential chronic effects.

None of the individual compounds showed a TU value above the threshold value of -3, neither at upstream nor downstream site. SumTU values were -5.2 and -3.67 at up- and downstream sites, respectively (Tab. S2.1 – Appendix B). Despite these values being below the threshold value of -3 at both sampling sites, they were calculated including only ~22% of the compounds at upstream and ~35% at downstream. In fact, the calculation of the TUs was not possible for most of the evaluated substances either for a lack of toxicity data in literature or because the concentration in water was found below the limit of detection (Tab. S2.1 – Appendix B). Therefore, it is possible that a sumTU calculated including all the compounds belonging to the chemical mixture would reach or exceed the threshold of -3 at both sites. According to Liess et al., (2008), this may indicate a long-term risk to macroinvertebrates populations.

6.2 Population analysis

Previous studies reported alterations in the population structure of amphipod species in response to the exposure to commonly detected contaminants in aquatic environments (Ladewig et al., 2006; Bundschuh et al., 2011; Peschke et al., 2014; Schneider et al., 2015; Love, 2017; Wigh et al., 2017; Ganser et al., 2019). Variations in the proportions of males and females, size and fecundity rate have

been shown in amphipods after lab exposure to xenobiotics (e.g., EDCs) (Schneider et al., 2015) and whole effluents (Gross et al., 2001; Ladewig et al., 2006; Peschke et al., 2014). In this study, amphipods were collected upstream and downstream of a WWTP annually between Sep 2017 and Sep 2018. Sex ratio, number of adults, number of juveniles and number of intersex individuals were recorded to describe the population structures of amphipods sampled above and below the sewage effluent. The number of eggs of the brooding females was used as fecundity parameter. Length and weight of the brooding females were also recorded, in order to evaluate a potential correlation between size and fecundity.

Although studies conducted on a higher number of amphipods would provide a more statistically consistent picture of the overall population structure, no significant differences were found when comparing upstream and downstream populations in September 2017 (Chapter 2.6.1) and 2018 (Chapter 2.6.2). This was true for all parameters evaluated. In this project, both chemical and molecular analyses were conducted on amphipods sampled in 2017. However, the population data suggest that no deterioration triggering changes in the amphipod population structure occurred in 2018 compared to the previous year. In fact, no significant differences were found when comparing the overall population structures of 2017and 2018 samplings (Chapter 2.6.3). In accordance with the literature (Franke, 1977; Ford et al., 2003; Franken, 2005), the brooding females sampled in September 2017 showed a positive correlation between size (length

(mm); weight (mg)) and number of eggs, in both up- and downstream sites (Fig. 2.6&2.7). The relationship between size and fecundity of the brooding females was not checked for the second sampling, since statistical analyses conducted using a low number of females with eggs sampled in 2018 (upstream, n=9; downstream, n=4) could not be representative of the population. When comparing length and number of eggs of the brooding females sampled in 2017 and 2018, no evident differences were observed (Fig. 2.9). On the other hand, females carrying eggs sampled in 2018 appeared heavier than 2017, in both up- and downstream sites (Fig. 2.9). This difference in weight needs to be interpreted cautiously though, since it was observed on a sample of only 9 and 4 females sampled in 2018 at up- and downstream sites, respectively.

The proportion of intersex individuals within *Gammarus fossarum* populations has been shown to range from 0 to 24% (Nagel et al., 2002; Jungmann et al., 2004; Ladewig et al., 2006). It has been demonstrated that the presence or the detection frequency of intersex amphipods depends on the sampling site (e.g., up- or downstream of a sewage discharge) as well as the chemical compositions of the stream (Nagel et al., 2002; Jungmann et al., 2004; Ladewig et al., 2006). In this study, an observational analysis of the sexual phenotype did not reveal intersex individuals neither in 2017 (Chapter 2.6.1) nor in 2018 (Chapter 2.6.2) samplings. The literature shows that the exposure to several stress sources (e.g., exposure to EDCs or parasite infections) can lead to the formation of intersex individuals in

Gammarus species (Bulnheim, 1965; Ladewig et al., 2006; Engelstädter et al., 2009; Peschke et al., 2014; Schneider et al., 2015). Although the chemical analysis performed in this study was not aimed to the detection of EDCs in water, the absence of intersex individuals suggests that the toxic pressure downstream of the effluent did not cause observable changes in amphipod sexual phenotype.

Further analyses will be needed to provide more detailed toxicological information and evaluate the long-term biological impact. However, the data collected in this study showed that the overall fraction of xenobiotics in amphipods did not cause evident changes in the population structure. In fact, no significant differences were found comparing amphipods sampled up- and downstream of the sewage effluent in the population analyses.

6.3 *Gammarus fossarum* complete transcriptome

Amphipods are still considered as poorly annotated species and further studies will be needed to collect more detailed molecular data on these ecologically important species. Previous studies have increased the molecular knowledge on amphipods applying "omics" platforms. For instance, a recent study by Cogne et al., (2019) was aimed to explore the genetic diversity between different amphipod species, including the three main sub-types of *G. fossarum* (Müller types A, B and C) (Müller, 2000), *G. pulex*, *Echinogammarus marinus* and *E. berilloni.* Using a proteogenomic approach, Trapp et al., (2015) analysed the diversity in the reproductive system of several amphipod species, such as *G. fossarum*, *G. pulex*,

G. roeseli and *Hyalella azteca* identifying their ovarian proteome. In this project, a high-throughput sequencing of *G. fossarum* (Müller type A) (Müller, 2000) total RNA was performed employing an Illumina *HiSeq 2500* sequencing device. The complete transcriptome was assembled *de novo* and annotated against a range of databases to increase the molecular information on this species. The annotated transcriptome produced as part of this project (Chapter 3) will significantly add to the gene discovery work on amphipods.

Base calling quality was high across the entire read dataset and an overall coverage of 228.25X was obtained in the sequencing. The assembly produced a total of 680,840 transcripts, clustered into 407,060 genes. Mapping of reads against a range of genomes identified ~50% of reads mapping to either *G. fossarum*, *G. pulex* or *E. marinus*, whilst the remaining reads did not map to any other model organism, indicating the absence of cross-species contamination within the samples.

Based on the evaluation of the content of evolutionary conserved orthologs in the transcriptome, a database of 978 metazoan universal single-copy orthologous genes was interrogated using the software BUSCO (Simão et al., 2015). The analysis identified 942 (96.4%) universal single-copy orthologs in a complete form, with 335 (34.3%) showing a single copy and 607 (62.1 %) showing 2 or more copies. Single-copy orthologs found in multiple copy in the assembly may correspond to

different gene isoforms. Notably, only 19 (1.9 %) were found in a fragmented form and 17 (1.7%) were not identified, confirming the completeness of the assembled transcriptome. The results of the single-copy orthologs analysis were in agreement with a recent study by Cogne et al., (2019) who performed a *de novo* assembly of 7 gammarid taxonomic groups.

Probably because of the presence in the transcriptome of erroneous transcripts (e.g., transcripts composed by short incomplete fragments or transcripts assembled using reads belonging to different splicing variants), the annotation against UniProt database was not able to identify ~90% of the transcripts in the assembly (Chapter 3.9.4). On the other hand, when focusing on transcripts with a complete ORF identified, more than 50% of the transcripts showed candidate hits in the UniProt annotation database (Chapter 3.9.4). This indicates that, excluding a large portion of artefacts of the assembly process, more than half of the transcripts that more likely represent genes was annotated.

A functional annotation against the eggNOG database identified a wide range of gene classes potentially useful in ecotoxicological research on amphipod species, for example HSPs and cytochrome p450 system (Fig. 3.12). HSPs have been demonstrated to be involved in crustacean stress responses to a wide range of both biotic and abiotic stress sources, such as exposure to aquatic pollutants (De Pomerai, 1996), thermic stress and microsporidian infections (Grabner et al.,

2014). Cytochrome p450 enzymes have been shown to be involved in detoxification processes of ecotoxicologically relevant species, such as fish (Roberts et al., 2005; Ings et al., 2011) and crustaceans (David et al., 2003; Del Brio et al., 2019).

6.4 Differential analyses using "omics" platforms

In order to explore the changes in the transcriptomic profiles between amphipods sampled above and below the WWTP, a DGE analysis was conducted using the transcript expression data obtained in the RNA-seq data analysis (Chapter 4). In addition, using a UPLC-MS/MS platform, the metabolic fingerprints in amphipod populations sampled at both sites were acquired (Chapter 5). Transcriptomic and metabolomic differences between male and female amphipods were also explored, in order to generate a data set that will be useful in the research on crustacean sexual biology. The differentially expressed genes between up- and downstream populations as well as between males and females were functionally classified using the Panther database, interrogating the *Drosophila melanogaster* GO-slim ontology (Chapter 4.4.7). Panther was also employed to conduct a functional analysis on the differential genes annotated against UniProt interrogating the GO-complete *D. melanogaster* ontology (Files S4.3-4.6 – Appendix B).

In order to investigate the metabolic pathways where the differential metabolites detected in “Upstream vs Downstream” and “Males vs Females” comparisons were involved, pathway analyses were performed. Consistently with the functional analysis conducted on the differentially expressed genes, the *D. Melanogaster* metabolomic data set was interrogated (Chapter 5.6.3).

6.4.1 Comparison between upstream and downstream amphipods

Previous studies have reported distinct clusters based on the gender when performing a principal component analysis on amphipod gene expression profiles (Ford et al., 2008; Short et al., 2012). Considering the documented differences between male and female crustaceans (Hyne et al., 2011; Pamuru, 2019) and between amphipods sampled above and below sewage effluents (Schirling et al., 2004; Ladewig et al., 2006; Besse et al., 2012; Peschke et al., 2014), a clear grouping of the samples based on the gene expression and metabolomic profiles was expected. Surprisingly, tests aimed to grouping the samples based on their transcriptomic (Fig. 4.6) and metabolomic (Fig. 5.4&5.5) profiles did not show distinct clusters between *G. fossarum* sampled up- and downstream of the effluent. It is currently unclear whether the background noise caused by the biological variability in the “omics” data could have hidden a clustering of the samples based on their gene expression and metabolic profiles.

Despite male and female lists of differential genes between up- and downstream populations showing little overlap (Fig. 4.9), a functional classification conducted in Panther revealed similar proportions of genes corresponding to a wide range of GO terms, including “metabolic process” and “response to stimulus”. This was true in all three GO categories explored, such as Biological process, Cellular component and Molecular function (Fig. 4.14), suggesting that the toxic pressure in the stream may trigger a molecular response in both male and female amphipods, even though the genders may respond differently.

Genes encoding the heat shock proteins HSP90 and HSP70 were detected among the differentially expressed genes between amphipods sampled up- and downstream of the WWTP, in both males and females (Files S4.15&4.16 – Appendix B). Changes in the expression of these stress biomarkers have been shown in ecologically relevant species (e.g., fish and invertebrates) in response to a wide range of stress sources, including thermic stress (Madeira et al., 2013), chemical exposure (Zhao et al., 2012) and parasite infections (Grabner et al., 2014). A sub-lethal stress response triggered by the toxic pressure in the stream appears probable, although the molecular variations were not mirrored by the population analysis, where no significant differences between up- and downstream sites were observed.

The gene *TSC1* was found downregulated in female amphipods sampled downstream of the WWTP compared to the upstream site (File S4.16 – Appendix B). Although the functions of this gene in crustaceans have not yet been described, *TSC1* has been shown as a tumour suppressor in *Drosophila melanogaster* (Sun et al., 2010). The gene product of *TSC1* has been demonstrated to form a functional complex with the protein TSC2 that negatively regulates target of rapamycin (TOR), an evolutionarily conserved kinase involved in cell growth and metabolic processes (Sun et al., 2010). Inactivating mutations of *TSC1* have been associated to an increase in cell number and organ size in *D. melanogaster* (Potter et al., 2001). On the other hand, the co-overexpression of *TSC1* and *TSC2* has been demonstrated to cause a decrease in cell size, cell number, and organ size, suggesting that *TSC2* may act as an epistatic regulator on *TSC1* (Potter et al., 2001). The gene *TSC1* was found differentially expressed only between female amphipods sampled up- and downstream of the sewage effluent, highlighting that males and females may respond differently to environmental pollutants. A histological comparison between amphipods sampled up- and downstream the effluent discharge was not performed in this study. Therefore, further analyses will be needed to verify whether variations in the expression of the genes within the TSC-TOR pathway could lead to organ malformations, similarly to what has been observed in *D. melanogaster* (Potter et al., 2001).

Genes associated to a broad spectrum of GO terms were detected when comparing the differential genes between amphipods sampled up- and downstream of the WWTP (Files S4.3-4.6 – Appendix B). Although no statistically enriched pathways were found in the "Biological process" category when running a GO-complete analysis on the list of differential genes in females, an enrichment in "metabolic process" class was observed submitting the male list (File S4.3 – Appendix B). Particularly, "protein metabolic process", "primary metabolic process" and "nitrogen compound metabolic process" were found among the significantly over-represented pathways (File S4.3 – Appendix B).

Hints of metabolic changes between up- and downstream samples were also detected in males in the metabolomic analysis (File S5.1-5.4 – Appendix B), allowing to run a pathway enrichment analysis on the differential metabolites. On the other hand, only one significantly changing metabolite (ω-6 long-chain polyunsaturated fatty acid: heneicosadienoic acid) was found when comparing *G. fossarum* females sampled up- and downstream of the effluent (File S5.6 – Appendix B). Since the results of the differential metabolomics analysis were not validated by targeted analyses, it is currently unclear whether the detection of heneicosadienoic acid among the differential metabolites between up- and downstream female amphipods can be biologically relevant. Targeted metabolomics studies will be needed to confirm and elucidate a potential role of heneicosadienoic acid in amphipod stress response.

Female total gene expression profiles appeared to be more disperse than males with the two greatest sources of variation (PC1 and PC2) (Fig. 4.6). A high biological variability in female amphipods sampled in different moulting/reproduction stages may have caused a background noise in the “omics” analyses. It is possible that this background noise may have hidden a portion of molecular changes between female amphipods sampled up- and downstream of the effluent, acting as a biasing factor in the pathway enrichment analyses.

A wide range of metabolic pathways were detected when running and enrichment analysis on male differential metabolites between up- and downstream samples. Among these, “One carbon (1C) pool by folate” was found statistically enriched and a number of other interesting metabolic networks (e.g., caffeine metabolism, arachidonic acid metabolism, tryptophan metabolism) were also detected (Tab. 5.4). 1C-metabolism is a universal metabolic process involved in the activation and transfer of 1C units for biosynthetic processes (Ducker et al., 2017) and partially takes place at mitochondrial level (Zong et al., 2016). Several genes with mitochondrial expression were also found within the list of differential genes between up- and downstream populations in both males and females (e.g., *MTCH2, PCCA, DHE3, SL9B2, RM50, ODO1, DHSD, TIM8, COX16*) (Files S4.15&4.16 – Appendix B), suggesting alterations in energetic metabolism. A literature search highlighted a lack of studies investigating the roles of 1C-metabolism in amphipod

stress response. Targeted metabolomics and proteomics analyses will be needed to investigate in detail the perturbations in this metabolic network in response to the exposure to single contaminants as well as xenobiotic mixtures in field studies.

Overall, the results of both metabolomic and transcriptomic pathway analyses suggest that the amphipod responses to toxic stress were subtle and may be very different between the genders. The importance of considering the gender in ecotoxicological studies on gammarids was already highlighted by Gismondi et al., (2013). The authors of this study investigated the influence of gender on the detoxification response of *G. roeseli* to Cd exposure. In particular, they measured glutathione, metallothionein, γ-glutamylcystein ligase, carotenoid, protein, lipid and glycogen levels in amphipods exposed to Cd. The levels of the biomarker of toxic effect malondialdehyde (Gismondi et al., 2013) were also measured in both males and females. Interestingly, lower malondialdehyde levels were found in females after exposure to Cd compared to males, and glycogen contents decreased only in females. The authors speculated that females might have more effective detoxification processes in response to Cd exposure. Another study by Bedulina et al., (2017) showed a different response to thermal stress of the amphipods *Eulimnogammarus verrucosus* and *E. cyaneus* between the genders. In particular, the authors used a differential proteomics approach to compare the proteomic profiles of control amphipods kept at 6-7 ^{o}C and amphipods exposed to a 1h heat shock (24.5-25.5 ^{o}C). Lower levels of Hsp70 were found in females of *E.*

verrucosus after the heat shock compared to males, although no differences between the genders were found in *E. cyaneus*. Among the proteins with different expression between males and females of *E. verrucosus*, other heat shock proteins, such as Hsp60 and Hsp90, were identified. Their data highlight that male and female amphipods can show very different HSP expression patterns in response to thermal stress, and these differences depend on the amphipod species examined. Other studies indicate gender-specific responses in amphipods exposed to other environmental stressors, such as hypoxia, salinity (Hoback et al., 1996; Sorom et al., 2010) and xenobiotic stress (Gismondi et al., 2012, Barros et al., 2017; Foucreau et al., 2017). Even more complex biological effects are expected when amphipods are exposed in the field to complex contaminant mixtures containing hundreds of different compounds and multiple stressors. There is clearly a need for further studies investigating these responses. Furthermore, understanding the differences in the impact of the toxic stress between genders becomes fundamental in ecotoxicological risk assessment, especially when the overall ecological status is evaluated, and long-term effects are predicted.

6.4.2 Sex-specific pathways

A PCA conducted on the gene expression profiles of all the analysed samples revealed a higher rate of separation with the greatest source of variation in females (Fig. 4.6). This was expected since amphipods collected were not at the

same development/reproduction stage. In fact, gene expression profiles, and in turn, the hormonal parameters, have been shown to be highly variable during different moulting and reproduction stages, especially in female amphipods (Sutcliffe, 2010; Hyne et al., 2011, Xuereb et al., 2011). On the other hand, metabolic profiles did not show evident differences between males and females in PCA plots (Fig. 5.6). In fact, when plotting male and female samples against the first and the second greatest sources of variations in the metabolomic profiles, no distinct clusters based on gender were observed. This was true in all ionization modes used (Fig. 5.6). In general, analysing gene expression and acquiring metabolic fingerprints can have very different outputs. Whilst the transcriptome reflects the genes that are actively expressed at a given moment, the metabolome is the final downstream product of gene transcription and its strictly related to the phenotype (Horgan et al., 2011). In addition, the proteome – which directly influences the metabolome – is constantly subject to a large number of regulation mechanisms (e.g., protein synthesis regulation and a wide range of post-translational modifications) (Merrick, 1992; Theodorescu et al., 2007), adding variability to the system. Acquiring the proteomic profiles of amphipods sampled in the stream analysed in this study and identifying the protein pathways of up- and downstream populations could fill the gap between the transcriptomic and metabolomic data. The presence of proteins annotated as components of primary metabolic processes could further confirm the data obtained in the transcriptomic and metabolomic analyses. Furthermore, a differential proteomics strategy (e.g.,

two-dimensional electrophoresis) (Monteoliva et al., 2004) comparing male and female amphipods could confirm the differences between the genders in the response to contaminants in water observed in this study. Ideally, future studies conducted using amphipods sampled in a synchronised moulting/reproduction stage (i.e., in precopula pair) (Chapter 1.5.4) may minimise the biological variability in females, allowing a clearer data interpretation.

In order to explore the general sex-specific pathways excluding the influence of the sampling site, a GO-complete "Biological process" analysis was conducted on the differential genes between males and females shared between up- and downstream samples. This analysis revealed GO terms related to a wide range of processes, including gamete formation, morphogenesis, circulatory system and muscle development (Fig. 4.16). GO terms related to gamete generation and development were expected, since the main differences between the genders in crustaceans are related to the reproduction system (Hyne et al., 2011; Pamuru, 2019). However, the presence of GO terms associated to heart and circulatory processes, muscle system, and cell differentiation and development among the over-represented processes when submitting the differential genes between males and females (Fig. 4.16) should be investigated in greater detail. It is possible that the amphipod endocrine system, which is different between males and females (Hyne et al., 2011), may influence a much wider spectrum of molecular networks than solely reproduction processes. Considering the ecotoxicological

importance of gammarid species and the gender-specific differences in their response to environmental stressors (Gismondi et al., 2012; Gismondi et al., 2013; Barros et al., 2017; Bedulina et al., 2017; Foucreau et al., 2017), the need for new sex-specific biomarkers appears clear. Future studies aimed to investigate the molecular differences between males and females in all *–ome* layers and their influence on the response to contaminants and multiple stressors will be fundamental.

Since a single *m/z* signal can be annotated to multiple molecular formula in an untargeted metabolomic analysis (e.g., different ion adducts or structural isomers), a pathway analysis using the differential metabolites between male and female amphipods shared between up- and downstream populations could not give accurate results. Therefore, the whole set of both upstream and downstream differential metabolites between the genders, including metabolites detected in all the ionization modes, were used for a pathway analysis in MetaboAnalyst (Chapter 5.7.3.2). Consistently with the functional analyses conducted on the differentially expressed genes, the *D. melanogaster* metabolic database was interrogated. The database was not able to recognise ~45% of the submitted compounds. Probably because of a low number of metabolites identified by the database, “Insect hormone biosynthesis” turned out to be the only significantly enriched pathway. In particular, JH and several ecdysteroids were found to match within this network. The crustacean homolog of JH is methyl farnesoate, which has

been shown to have a wide range of biological functions, such as stimulation of general protein synthesis, promotion of the moult cycle and reproduction (Homola et al., 1977). On the other hand, ecdysteroids have been shown to be involved in the timing of moulting and metamorphosis in arthropods (Koolman, 1990; Chang 1993). Given the documented differences between the genders in moulting and reproduction processes in amphipods (Hyne et al., 2011), the presence of ecdysteroids was therefore unsurprising. Although several other pathways (e.g., arachidonic acid metabolism, sphingolipid metabolism, glycerophospholipid metabolism, tyrosine metabolism) were found in the functional analysis, the enrichments were not statistically significant (Tab. 5.5). However, this might be due to a low number of metabolites that gave a match in the database rather than the presence of false positives. In addition, whilst the current metabolomic databases contain detailed metabolic data for *Homo sapiens* and a limited number of model organisms, it is possible that several compound names corresponding to crustacean metabolites were not recognised when interrogating the data set of a different species. Further studies aimed to collect and annotate metabolomic data from amphipod species will allow these environmentally important organisms to be adequately represented in terms of metabolomic data sets.

6.5 Limitations of the project

In the present project, a transcriptomic and a metabolomic analysis were conducted in parallel to:

- provide a transcriptomic data set that will be useful in future studies to develop new ecotoxicological biomarkers in amphipods
- evaluate the biological impact of a complex contaminant mixture released through a Swiss WWTP on *Gammarus fossarum* amphipods sampled in filed
- explore the differences in the transcriptome and the metabolome between male and female amphipods

To achieve these goals, the complete transcriptome of *G. fossarum* was assembled and annotated. In addition, differences in the population structure and the gene expression profiles as well as the changes in the metabolome were compared between amphipods sampled up- and downstream of the effluent and between male and female amphipods. The data set generated in this study could represent a valuable resource to find new ecotoxicological biomarkers in amphipods and to obtain new information on crustacean sexual biology.

Although no significant differences in the population structure were observed comparing amphipods sampled above and below the sewage effluent (Chapter 2.6.1&2.6.2), the low number of animals sampled represented a limiting factor in terms of statistical significance of the results. In fact, whilst the population parameters were recorded on a total of 609 amphipods in September 2017, only 193 amphipods were sampled in September 2018. A particularly dry summer as well as a lack of precipitation may have caused the sampling of a lower number of

animals in 2018 compared to the previous year. Studies conducted on a larger number of amphipods would provide a more statistically consistent picture of the population structures. Additionally, given that the composition of the contaminant mixtures in water systems can substantially vary between seasons (Nelson et al., 2011; Munz et al., 2018), the ecological monitoring should be conducted on a longer-term basis. This would ensure that the overall status of the system can continuously be evaluated and that the anthropogenic impact on the local fauna can be more accurately predicted.

A low number of brooding females and juveniles compared to the overall population size was observed in both 2017 and 2018 samplings (Chapter 2.6.1&2.6.2). Although the seasonal life-cycles in amphipods may vary in different geographic areas, their reproduction rate in European regions has been described to be much reduced in autumn (October-November) (Pöckl et al., 2003). Therefore, it is possible that a breeding resting-stage was starting during the sampling period. For logistic reasons, all the samplings in this project were performed in mid-September. Further studies aimed to monitor both the amphipod population structure and fecundity rate should be conducted in early summer (June-July), when the breeding activity is pronounced (Pöckl et al., 2003).

Because there was no way to avoid the separation of the animals in precopulatory mate guarding during the kick-net sampling procedure, the amphipods used in this

project were not collected in the same moulting/reproduction stage. This represented a limitation in the interpretation of the "omics" data and added biological variability to the molecular analyses. It is possible that the background noise caused by the biological variability having prevented a clearer grouping of the samples based on their gene expression and metabolic profiles. Future studies on amphipod response to whole effluents should be conducted on animals in precopula pair, in order to minimise the variability caused by different gene expression profiles and hormonal composition in different moulting stages. In field caged studies or lab exposure studies using different fractions of whole effluent could represent viable options to avoid the separation of animals in precopula. Nevertheless, important biological processes (e.g., general stress response, primary metabolism, mitochondrial energetic metabolism) were observed in the functional analyses conducted on the differential genes (Chapter 4.4.7) and metabolites (Chapter 5.7.3) between amphipods sampled upstream and downstream of the effluent discharge. However, these alterations were not mirrored by the population data where no changes in the main population parameters were observed and no animals showing an intersex phenotype were found. Therefore, the molecular alterations are to be considered sub-lethal and need to be confirmed by long-term analyses. In addition, targeted analyses will need to be performed to clarify the roles of the genes and metabolites found in this project in amphipod stress response.

A total of 10 genes were selected from the lists of differentially expressed genes between up- and downstream populations detected in the DGE analysis to perform an experimental validation of the differences in expression through qPCR (Chapter 4.3.11). Because of a lack of information in the literature about the genes chosen for the validation experiment, a study on the structures of the chosen genes aimed to design exon-exon junction spanning primers was not possible, thus the primer sets were designed *de novo*. The primer sets designed for 5 out of 10 genes chosen for the validation experiment showed multiple amplification products or no amplification in the PCR tests (Fig. 4.19). Whilst multiple PCR products may be due to the amplification of different gene isoforms or non-target sequences, the absence of amplification products may indicate the formation of secondary structures in the primers. For time reasons, ordering new primer sets and sequencing the amplification products corresponding to each gene chosen for the qPCR validation experiment was not possible, therefore only the genes for which the corresponding primers gave a single amplification product were included in the qPCR data analysis.

Despite high-fold changes being detected in the DGE analysis for all the differential genes between up- and downstream populations chosen for the validation experiment (Tab. 4.1&4.2), only 2 (*MYP2* and *ODO1*) out of 5 genes evaluated through qPCR (Male list: *LDAH, MYP2, DHSD*; Female list: *ODO1, NPAB*) were experimentally validated (Chapter 4.4.11). The absence of correlation between

RNA-seq and qPCR expression data can be due to a number of factors. Firstly, due to logistic reasons, the RNA-seq and qPCR experiments were performed 2 years apart, hence variations in the chemical composition of the effluent cannot be excluded and may have represented a source of variability between the 2 experiments. Secondly, the RNA extracted from 5 animals per replicate were pooled for each sample in the RNA-seq analysis, in order to increase the RNA yield. On the other hand, the RNA extracted from single amphipods were used for the qPCR assay. This difference in the experimental design may have caused a different estimation of the biological variability during the data analysis. Finally, a false positive rate of 0.05 was set in the DGE analysis (Chapter 4.3.4) to detect the significantly changing genes between the groups, hence the presence of false positives within the lists of changing transcripts cannot be excluded.

6.6 Novelty and main findings

The present project was the first to date to use a transcriptomics and a metabolomics platform in parallel to explore the molecular changes in filed amphipods exposed to a whole effluent.

A number of limitations about applying a non-targeted multi-omics approach to evaluate the impact of a complex contaminant mixture on a biomarker amphipod species have been brought to light in this study. Nevertheless, the data set generated will represent a valuable resource to develop new ecotoxicological biomarkers in crustaceans.

A complete *Gammarus fossarum* transcriptome was generated and made available to the scientific community. The transcriptome was deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession code PRJNA556212.

Overall, this study resulted in the following findings:

- the toxic pressure at the Eulach river did not cause observable abnormalities in *Gammarus fossarum* population structure

- the xenobiotic mixture in the water did not lead to the formation of intersex individuals within *G. fossarum* populations

- the differential "omics" analyses highlighted that variations in general stress biomarkers, primary metabolism, RNA transcription and maturation and mitochondrial energetic metabolism may be triggered by amphipod exposure to aquatic contaminants present in the wastewater effluent

- the comparative analyses of male and female transcriptome and metabolome between animals sampled above and below the WWTP showed that the genders may respond differently to anthropogenic pollutants in aquatic environments

- the comparison between male and female metabolome and transcriptome showed that the differences between amphipod genders may not be limited to reproduction-related processes. The sex distinction traits in hormonal system may act on a wider spectrum of molecular networks, such as heart and circulatory processes, muscle system and cell differentiation and development.

6.7 Future perspectives

The ecological implications of the exposure of aquatic organisms to anthropogenic contaminants can vary substantially depending on the geographical area, water chemical composition and concentrations of the single pollutants in the aquatic environment (Snape et al., 2004). Furthermore, the exposure to very low levels of contaminants can take a long time to have biological effects (Thorpe et al., 2008), thus observable changes within populations are likewise protracted (Paganelli et al., 2016). Also, the tolerance to pollution itself can vary temporally (Dehedin et al., 2013). For these reasons, a constant monitoring of the rivers will be fundamental to accurately predict the effects of the anthropogenic xenobiotics on

aquatic communities. For instance, chemical, physiological and molecular analyses on amphipod species sampled in all Swiss rivers, conducted on a long-term basis, may provide a consistent picture of the overall ecological status of freshwater environments. Comparative studies aimed to evaluate the biological impact of aquatic contaminants after the application of improvement measures of the WWTPs will elucidate the benefits of new wastewater treatment technologies on the aquatic environments.

Sampling a statistically consistent number of amphipods to conduct population and molecular analyses can represent a significant challenge. This is even more true in late summer or early autumn (September-November), when gammarid reproduction rate is much reduced (Pöckl et al., 2003). Future population and molecular analyses on amphipod species should be conducted in early summer (June-July), when the breeding activity is pronounced (Pöckl et al., 2003). This will ensure the sampling of a consistent number of animals to carry out the analyses and will avoid an underestimation of the number of juveniles and precopula pairs due to the breeding stage.

When performing high-throughput “omics” analyses, a high rate of biological variability within the samples can represent a strong limiting factor in terms of data interpretation. In this project, the amphipods were not sampled in a synchronised moulting/reproduction stage. It is probable that the resulting

background noise in the molecular analyses may have hidden important changes between up- and downstream populations, especially in female amphipods, where complex hormonal fluctuations take place during different breeding stages (Sutcliffe, 2010; Hyne et al., 2011, Xuereb et al., 2011). Future studies should be conducted on amphipods in precopula mate guarding, reducing considerably the biological variability.

Because of the complex interactions among the *-ome* layers (genome, transcriptome, proteome and metabolome) in organisms, the results obtained applying a "omics" platform may not (or only partially) match with the data recorded applying a different platform. For example, the differential proteins between a control sample and a treated sample may not match with the corresponding genes detected in a differential gene expression analysis. Although these discrepancies in the results can be due to many technical and/or biological factors, it is crucial that all *-ome* layers are explored when comparing two biological conditions. Looking at the whole spectrum of changing biomolecules in a given organism compared to control samples, more accurate inferences can be made and the potential discrepancies in the results can be explained more accurately. Further studies on the evaluation of the amphipod molecular responses to aquatic contaminants should record the gene expression, proteomic and metabolomic profiles simultaneously. Using this approach, a more holistic understanding of the amphipod physiological status can be achieved.

Chapter 7 - References

Aardema, M. J., and MacGregor, J. T., (2002). Toxicology and genetic toxicology in the new era of “toxicogenomics”: Impact of “omics” technologies. Mut. Res. 499, 13–25.

Adam, O., Degiorgi, F., Crini, G. and Badot, P.M., (2010). High sensitivity of Gammarus sp. juveniles to deltamethrin: outcomes for risk assessment. Ecotoxicology and environmental safety, 73(6), 1402-1407.

Adeloye, A. J. (2009). Rivers and human development. Fresh surface water, 3, 117.

Agilent Technologies, (2008). Agilent RNA 6000 Pico Kit Quick Start Guide.

Agilent Technologies, (2018). 2100 Expert Software User’s Guide. Agilent 2100 Bioanalyzer System Manual.

Aittokallio, T., & Schwikowski, B., (2006). Graph-based methods for analysing networks in cell biology. Briefings in bioinformatics, 7(3), 243-255.

Aken, B.L., Achuthan, P., Akanni, W., Amode, M.R., Bernsdorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P. and Gil, L., (2017). Ensembl 2017. Nucleic acids research, 45(D1), D635-D642.

Al-Bahry, S., Mahmoud, I., Al-Belushi, K., Elshafie, A., Al-Harthy, A., & Bakheit, C., (2009). Coastal sewage discharge and its impact on fish with reference to antibiotic resistant enteric bacteria and enteric pathogens as bio-indicators of pollution. Chemosphere, 77(11), 1534-1539.

Al-Salhi, R., Abdul-Sada, A., Lange, A., Tyler, C. R., & Hill, E. M. (2012). The xenometabolome and novel contaminant markers in fish exposed to a wastewater treatment works effluent. Environmental science & technology, 46(16), 9080-9088.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Anders, S. and Huber, W., (2010). Differential expression analysis for sequence count data. Genome Biol., 11, R106.

Andrews, S., (2010). FastQC: a quality control tool for high throughput sequence data.BabrhamBioinformatics,URL:https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arkoosh, M. R., Clemons, E., Huffman, P., Kagley, A. N., Casillas, E., Adams, N., and Stein, J. E., (2001). Increased susceptibility of juvenile chinook salmon to vibriosis after exposure to chlorinated and aromatic compounds found in contaminated urban estuaries. Journal of Aquatic Animal Health, 13(3), 257-268.

Ashauer R, Kuhl R, Zimmer E, Junghans M, (2017). Effect Modelling to Evaluate Water Quality Monitoring Strategies. University of York, Heslington, York, UK.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., & Harris, M. A., (2000). Gene ontology: tool for the unification of biology. Nature genetics, 25(1), 25-29.

Atli, G. & Canli, M., (2007). Enzymatic responses to metal exposures in a freshwater fish Oreochromis niloticus. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 145(2), 282-287.

Baeza-Rojano, E., Hachero-Cruzado, I., & Guerra-García, J. M., (2014). Nutritional analysis of freshwater and marine amphipods from the Strait of Gibraltar and potential aquaculture applications. Journal of sea research, 85, 29-36.

Bahamonde, P. A., Feswick, A., Isaacs, M. A., Munkittrick, K. R., & Martyniuk, C. J., (2016). Defining the role of omics in assessing ecosystem health: Perspectives from the Canadian environmental monitoring program. Environmental toxicology and chemistry, 35(1), 20-35.

Bahamonde, P.A., Tetreault, G.R., McMaster, M.E., Servos, M.R., Martyniuk, C.J. and Munkittrick, K.R., (2014). Molecular signatures in rainbow darter (Etheostoma caeruleum) inhabiting an urbanized river reach receiving wastewater effluents. Aquatic Toxicology, 148, 211-220.

Baker, M., (2012). De novo genome assembly: what every biologist should know. Nature methods, 9(4), 333-337.

Balian, E., Segers, H., Lévêque, C., & Martens, K., (2008). The freshwater animal diversity assessment: an overview of the results. Hydrobiologia, 595(1), 627-637.

Banerjee, S., & Mazumdar, S., (2012). Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. International journal of analytical chemistry, 2012.

Barron, M.G. Bioconcentration, (1990). Will water-borne organic chemicals accumulate in aquatic organisms? Environ. Sci. Technol. 24 (11), 1612−1618.

Barros, S., Montes, R., Quintana, J.B., Rodil, R., Oliveira, J.M., Santos, M.M. & Neuparth, T., (2017). Chronic effects of triclocarban in the amphipod Gammarus locusta: Behavioural and biochemical impairment. Ecotoxicology and environmental safety, 135, 276-283.

Beale, D. J., Karpe, A. V., Ahmed, W., Cook, S., Morrison, P. D., Staley, C., & Palombo, E. A., (2017). A community multi-omics approach towards the assessment of surface water quality in an urban river system. International journal of environmental research and public health, 14(3), 303.

Bedulina, D., Meyer, M.F., Gurkov, A., Kondratjeva, E., Baduev, B., Gusdorf, R. and Timofeyev, M.A., (2017). Intersexual differences of heat shock response between two amphipods (Eulimnogammarus verrucosus and Eulimnogammarus cyaneus) in Lake Baikal. PeerJ, 5, e2864.

Beketov, M.A. and Liess, M., (2008). Acute and delayed effects of the neonicotinoid insecticide thiacloprid on seven freshwater arthropods. Environmental Toxicology and Chemistry: An International Journal, 27(2), 461-470.

Beracko, P., Sýkorová, A., & Štangler, A., (2012). Life history, secondary production and population dynamics of Gammarus fossarum (Koch, 1836) in a constant temperature stream. Biologia, 67(1), 164-171.

Berenzen, N., Schulz, R., & Liess, M., (2001). Effects of Chronic Ammonium And Nitrite Contamination On The Macroinvertebrate Community In Running Water Microcosms. Wat. Res., Vol. 35(No.14), 3478–3482.

Besse, J.P., Coquery, M., Lopes, C., Chaumot, A., Budzinski, H., Labadie, P. and Geffard, O., (2013). Caged Gammarus fossarum (Crustacea) as a robust tool for the characterization of bioavailable contamination levels in continental waters: towards the determination of threshold values. Water research, 47(2), 650-660.

Beyer, J., Petersen, K., Song, Y., Ruus, A., Grung, M., Bakke, T., & Tollefsen, K. E., (2014). Environmental risk assessment of combined effects in aquatic ecotoxicology: a discussion paper. Marine environmental research, 96, 81-91.

Blarer, P. and Burkhardt-Holm, P., (2016). Microplastics affect assimilation efficiency in the freshwater amphipod Gammarus fossarum. Environmental Science and Pollution Research, 23(23), 23522-23532.

Bloor, M. C., & Banks, C. J., (2006). An evaluation of mixed species in-situ and ex-situ feeding assays: The altered response of Asellus aquaticus and Gammarus pulex. Environ Int, 32(1), 22-27.

Bloor, M. C., Banks, C. J., & Krivtsov, V., (2005). Acute and sublethal toxicity tests to monitor the impact of leachate on an aquatic environment. Environ Int, 31(2), 269-273.

Bonga, S. E., (1997). The stress response in fish. Physiological reviews, 77(3), 591-625.

Bossus, M. C., Guler, Y. Z., Short, S. J., Morrison, E. R., & Ford, A. T., (2014). Behavioural and transcriptional changes in the amphipod Echinogammarus marinus exposed to two antidepressants, fluoxetine and sertraline. Aquatic Toxicology, 151, 46-56.

Bownik, A., (2019). Physiological endpoints in daphnid acute toxicity tests. Science of The Total Environment, 134400.

Brandão, F.P., Rodrigues, S., Castro, B.B., Goncalves, F., Antunes, S.C. and Nunes, B., (2013). Short-term effects of neuroactive pharmaceutical drugs on a fish species: biochemical and behavioural effects. Aquatic toxicology, 144, 218-229.

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L., (2016). Near-optimal probabilistic RNA-seq quantification. Nature biotechnology, 34(5), 525-527.

Brockmeier, E. K., Hodges, G., Hutchinson, T. H., Butler, E., Hecker, M., Tollefsen, K. E., & Colbourne, J., (2017). The role of omics in the application of adverse outcome pathways for chemical risk assessment. Toxicological Sciences, 158(2), 252-262.

Brown, M., Dunn, W. B., Dobson, P., Patel, Y., Winder, C. L., Francis-McIntyre, S., & Swainston, N., (2009). Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. Analyst, 134(7), 1322-1332.

Brown, M., Wedge, D.C., Goodacre, R., Kell, D.B., Baker, P.N., Kenny, L.C., Mamas, M.A., Neyses, L. and Dunn, W.B., (2011). Automated workflows for accurate mass-

based putative metabolite identification in LC/MS-derived metabolomic datasets. Bioinformatics, 27(8), 1108-1112.

Bulnheim H-P, (1965) Untersuchungen über Intersexualität bei Gammarus duebeni (Crustacea, Amphipoda). Helgol Wiss Meersunters 12:349–394.

Bundschuh, M., Hahn, T., Gessner, M. O., & Schulz, R., (2009). Antibiotics as a chemical stressor affecting an aquatic decomposer–detritivore system. Environmental Toxicology and Chemistry, 28(1), 197-203.

Bundschuh, M., Zubrod, J. P., Klemm, P., Elsaesser, D., Stang, C., & Schulz, R., (2013). Effects of peak exposure scenarios on Gammarus fossarum using field relevant pesticide mixtures. Ecotoxicology and Environmental Safety, 95, 137-143.

Bundschuh, M., Zubrod, J.P. and Schulz, R., (2011). The functional and physiological status of Gammarus fossarum (Crustacea; Amphipoda) exposed to secondary treated wastewater. Environmental Pollution, 159(1), 244-249.

Bundy, J. G., Davey, M. P., & Viant, M. R. (2009). Environmental metabolomics: a critical review and future perspectives. Metabolomics, 5(1), 3.

Burkhardt-Holm, P., Escher, M., & Meier, W., (1997). Waste-water management plant effluents cause cellular alterations in the skin of brown trout. Journal of Fish Biology, 50(4), 744-758.

Busch, W., Schmidt, S., Kühne, R., Schulze, T., Krauss, M., & Altenburger, R., (2016). Micropollutants in European rivers: A mode of action survey to support the development of effect-based tools for water monitoring. Environmental toxicology and chemistry, 35(8), 1887-1899.

Canobbio, S., Mezzanotte, V., Sanfilippo, U. and Benvenuto, F., (2009). Effect of multiple stressors on water quality and macroinvertebrate assemblages in an effluent-dominated stream. Water, air, and soil pollution, 198(1-4), 359-371.

Caputo, D.R., Robson, S.C., Werner, I. and Ford, A.T., (2020). Complete transcriptome assembly and annotation of a critically important amphipod species in freshwater ecotoxicological risk assessment: Gammarus fossarum. Environment International, 137, 105319.

Cavallin, J. E., Jensen, K. M., Kahl, M. D., Villeneuve, D. L., Lee, K. E., Schroeder, A. L., & Blackwell, B. R., (2016). Pathway-based approaches for assessment of real-time exposure to an estrogenic wastewater treatment plant effluent on fathead minnow reproduction. Environmental toxicology and chemistry, 35(3), 702-716.

Cazenave, J., Bacchetta, C., Rossi, A., Ale, A., Campana, M., & Parma, M. J., (2014). Deleterious effects of wastewater on the health status of fish: a field caging study. Ecological indicators, 38, 104-112.

Chang, E. S., (1993). Comparative endocrinology of molting and reproduction: insects and crustaceans. Annual review of entomology, 38(1), 161-180.

Chang, E. S., (1995). Physiological and biochemical changes during the molt cycle in decapod crustaceans: an overview. Journal of Experimental Marine Biology and Ecology, 193(1-2), 1-14.

Charron, L., Geffard, O., Chaumot, A., Coulaud, R., Queau, H., Geffard, A. and Dedourge-Geffard, O., (2013). Effect of water quality and confounding factors on digestive enzyme activities in Gammarus fossarum. Environmental Science and Pollution Research, 20(12), 9044-9056.

Chiu, K. H., Dong, C. D., Chen, C. F., Tsai, M. L., Ju, Y. R., Chen, T. M., & Chen, C. W. (2017). NMR-based metabolomics for the environmental assessment of Kaohsiung Harbor sediments exemplified by a marine amphipod (Hyalella azteca). Marine pollution bulletin, 124(2), 714-724.

Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., & Robertson, A. J., (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nature methods, 5(7), 613.

Cogne, Y., Degli-Esposti, D., Pible, O., Gouveia, D., François, A., Bouchez, O., & Chaumot, A., (2019). De novo transcriptomes of 14 gammarid individuals for proteogenomic analysis of seven taxonomic groups. Scientific data, 6(1), 1-7.

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., & Bauer, D. J. (2011). The ecoresponsive genome of Daphnia pulex. Science, 331(6017), 555-561.

Cold, A., & Forbes, V. E., (2004). Consequences of a short pulse of pesticide exposure for survival and reproduction of Gammarus pulex. Aquat Toxicol, 67(3), 287-299.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M., (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21(18), 3674-3676.

Craddock, N., & Sklar, P., (2013). Genetics of bipolar disorder. The Lancet, 381(9878), 1654-1662.

Crawford, G., Hurrel, P., Paroschy, K., & Pereira, C., (2017). Pharmaceuticals and other endocrine disrupting compounds in natural water systems. Muskoka Watershed Council, Bracebridge.

Dangles, O., and Guerold, F., (2001) Linking shredders and leaf litter pro- cessing: insights from an acidic stream study. Int. Rev. Hydrobiol. 86, 395–406.

Daughton, C.G., Ternes, T.A., (1999). Pharmaceuticals and personal care products in the environment: agents of subtle change.Environmental Health Perspectives 107, 907e938.

Davey, M. P., Harmens, H., Ashenden, T. W., Edwards, R.,&Baxter, R. (2007). Species-specific effects of elevated CO2 on resource allocation in Plantago maritima and Armeria maritima. Biochem- ical Systematics and Ecology, 35, 121–129.

David, A., Lange, A., Abdul-Sada, A., Tyler, C. R., & Hill, E. M. (2017). Disruption of the prostaglandin metabolome and characterization of the pharmaceutical exposome in fish exposed to wastewater treatment works effluent as revealed by nanoflow-nanospray mass spectrometry-based metabolomics. Environmental science & technology, 51(1), 616-624.

David, P., Dauphin-Villemant, C., Mesneau, A. and Meyran, J.C., (2003). Molecular approach to aquatic environmental bioreporting: differential response to environmental inducers of cytochrome P450 monooxygenase genes in the detritivorous subalpine planktonic Crustacea, Daphnia pulex. Molecular ecology, 12(9), 2473-2481.

Davie, A., Minghetti, M., & Migaud, H., (2009). Seasonal variations in clock-gene expression in Atlantic salmon (Salmo salar). Chronobiology international, 26(3), 379-395.

De Lange, H. J., Noordoven, W., Murk, A. J., Lurling, M., & Peeters, E. T., (2006a). Behavioural responses of Gammarus pulex (Crustacea, Amphipoda) to low concentrations of pharmaceuticals. Aquat Toxicol, 78(3), 209-216.

De Pomerai, D.I., (1996). Heat-shock proteins as biomarkers of pollution. Human & experimental toxicology, 15(4), 279-285.

Dedourge-Geffard, O., Palais, F., Biagianti-Risbourg, S., Geffard, O., & Geffard, A., (2009). Effects of metals on feeding rate and digestive enzymes in Gammarus fossarum: an in situ experiment. Chemosphere, 77(11), 1569-1576.

Dehedin, A., Piscart, C., & Marmonier, P., (2013). Seasonal variations of the effect of temperature on lethal and sublethal toxicities of ammonia for three common freshwater shredders. Chemosphere, 90(3), 1016-1022.

Del Brio, J., Lares, B.A., Parra-Morales, L.B., Sanchez, V.G., Montagna, C.M. and Venturino, A., (2019). Differential detoxifying responses to crude oil water-accommodated fraction in Hyallela curvispina individuals from unpolluted and contaminated sites. Environmental toxicology and pharmacology, 70, 103-191.

Desbrow, C., et al., (1998). Identification of estronic chemicals in STW Effluent. 1. Chemical fractionation and in vitro biological screening.Environ. Sci. Technol. 32, 1549–1558.

Dessimoz, C., & Škunca, N., (2017). The Gene Ontology Handbook (Vol. 1446, pp. 15-24). New York, NY, USA, Humana Press.

Dieterle, F., Ross, A., Schlotterbeck, G. and Senn, H., (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical chemistry, 78(13), 4281-4290.

Dixon, M. J., & Shaw, P. J., (2011). Watercress and Water Quality: The Effect of Phenethyl Isothiocyanate on the Mating Behaviour of Gammarus pulex. International Journal of Zoology, 2011.

Ducker, G. S., & Rabinowitz, J. D., (2017). One-carbon metabolism in health and disease. Cell metabolism, 25(1), 27-42.

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

Dunn, W. B., & Ellis, D. I., (2005). Metabolomics: current analytical platforms and methodologies. TrAC Trends in Analytical Chemistry, 24(4), 285-294.

Dunn, W. B., Broadhurst, D. I., Deepak, S. M., Buch, M. H., McDowell, G., Spasic, I., & Neyses, L. (2007). Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate. Metabolomics, 3(4), 413-426.

Duran, M., (2007). Life Cycle of Gammarus pulex (L.) in the River Yeflil›rmak. Turk J Zool, 31, 389-394.

EAA., (2015). The European environment — state and outlook 2015. URL: http://www.eea.europa.eu/soer.

EB, Web Encyclopedia Britannica, M.W., (2012), May.

EC, (2006). No1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) as regards Annex XII; The European Parliament and the Council of the European Union; Commission, ed. Off. J. Eur. Union.

Ecetoc., (1997). The value of aquatic model ecosystem studies in ecotoxicology. Retrieved from Brussels, Belgium.

Eggen, R. I., Hollender, J., Joss, A., Schärer, M., & Stamm, C., (2014). Reducing the discharge of micropollutants in the aquatic environment: the benefits of upgrading wastewater treatment plants.

Ekambaram, P., &Narayanan, M., (2017). Upregulation ofHSP70 extends Cytoprotection to fish brain under xenobiotic stress. Journal ofFisheriesSciences.com, 11(1), 11–20.

Emwas, A.H.M., (2015). The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. In *Metabonomics* (pp. 161-193). Humana Press, New York, NY.

Engelstädter, J.; Hurst, G. D. D., (2009). The ecology and evolution of microbes that manipulate host reproduction. Ann. Rev. Ecol., Evol., Syst., 40 (1), 127−149.

Englert, D., Zubrod, J.P., Link, M., Mertins, S., Schulz, R. and Bundschuh, M., (2017). Does waterborne exposure explain effects caused by neonicotinoid-contaminated plant material in aquatic systems?. Environmental science & technology, 51(10), 5793-5802.

Etxabe, A. G., Short, S., Flood, T., Johns, T., & Ford, A. T., (2015). Pronounced and prevalent intersexuality does not impede the 'Demon Shrimp'invasion. PeerJ, 3, e757.

Fabrick J. A., Hull J., (2017). Assessing Integrity of Insect RNA. Agilent Technologies Manual.

Famili, I., Förster, J., Nielsen, J., & Palsson, B. O., (2003). Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proceedings of the National Academy of Sciences, 100(23), 13134-13139.

Felten, V., (2003). Effects of the acidification of Vosges streams on the biology, ecology and ecophysiology of Gammarus fossarum Koch, 1835 (Crustacea Amphipoda): Integrated approach at different levels of organization. University of Metz, 340.

Felten, V., Charmantier, G., Charmantier-Daures, M., Aujoulat, F., Garric, J., & Geffard, O., (2008a). Physiological and behavioural responses of Gammarus pulex exposed to acid stress.Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 147(2), 189-197.

Felten, V., Charmantier, G., Mons, R., Geffard, A., Rousselle, P., Coquery, M.et al., (2008b). Physiological and behavioural responses of Gammarus pulex (Crustacea: Amphipoda) exposed to cadmium. Aquat Toxicol, 86(3), 413-425.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M., (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science, 246(4926), 64-71.

Feswick, A., Munkittrick, K.R., Martyniuk, C.J., (2017). Estrogen-responsive gene networks in the teleost liver: What are the key molecular indicators? Environ. Toxicol. Pharmacol. 56, 366–374.

Fiehn, O., (2002). Metabolomics—the link between genotypes and phenotypes. In Functional genomics. Springer, Dordrecht, 155-171.

Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., & Fostel, J., (2007). The metabolomics standards initiative (MSI). Metabolomics, 3(3), 175-178.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J. & Sonnhammer, E.L., (2013). Pfam: the protein families database. Nucleic acids research, 42(D1), D222-D230.

Finn, R.D., Clements, J. and Eddy, S.R., (2011). HMMER web server: interactive sequence similarity searching. Nucleic acids research, 39(suppl_2), W29-W37.

Fischer, S., Fischer, M., Kristin, S. & Werner I., (2017). Wirkungsorientierte Gewässerüberwachung: Biomonitoring mit Forellen - Studie im Auftrag des Bundesamtes für Umwelt (BAFU). URL:

https://www.oekotoxzentrum.ch/projekte/aquatischeoekotoxikologie/biomarker-in-der-bachforelle/.

Ford, A. T., & Glazier, D. S., (2008). Persistently high levels of intersexuality in male-biased amphipod populations. Zoology, 111(5), 401-409.

Ford, A.T., (2012). Intersexuality in Crustacea: an environmental issue? Aquatic toxicology, 108, 125-129.

Ford, A.T., Fernandes, T.F., Rider, S.A., Read, P.A., Robinson, C.D. and Davies, I.M., (2003). Measuring sublethal impacts of pollution on reproductive output of marine Crustacea. Marine Ecology Progress Series, 265, 303-309.

Ford, A.T., Fernandes, T.F., Rider, S.A., Read, P.A., Robinson, C.D. and Davies, I.M., (2004). Endocrine disruption in a marine amphipod? Field observations of intersexuality and de-masculinisation. Marine environmental research, 58(2-5), 169-173.

Forrow, D. M., & Maltby, L., (2000). Toward a mechanistic understanding of contaminant-induced changes in detritus processing in streams: Direct and indirect effects on detritivore feeding. Environmental Toxicology and Chemistry, 19(8), 2100-2106.

Foucreau N, Cottin D, Piscart C, Hervant F., (2014). Physiological and metabolic responses to rising temperature in Gammarus pulex (Crustacea) populations living under continental or Mediterranean climates. Comparative Biochemistry and Physiology Part A: Molecular & integrative Physiology 168:69–75.

Franke U., (1977). Experimentelle Untersuchungen zur Respiration von Gammarus fossarum KOCH 1835 (Crustacea-Amphipoda) in AbhUngigkeit von Temperatur, Sauerstoffkonzentration und Wasserbewegung. Arch Hydrobiol 48(suppl):369–411.

Franken RJM, Waluto B, Peeters ETHM, Gardeniers JJP, Bejer JAJ, Scheffer M., (2005) Growth of shredders on leaf litter biofilms: The effect of light intensity. Freshw Biol 50:459–466.

Gallardo, B., Clavero, M., Sánchez, M.I. and Vilà, M., (2016). Global ecological impacts of invasive species in aquatic ecosystems. Global change biology, 22(1), 151-163.

Ganser, B., Bundschuh, M., Werner, I., Homazava, N., Vermeirssen, E. L., Moschet, C., & Kienle, C., (2019). Wastewater alters feeding rate but not vitellogenin level of Gammarus fossarum (Amphipoda). Science of The Total Environment, 657, 1246-1252.

Gaufin, A. R., & Tarzwell, C. M., (1956). Aquatic macro-invertebrate communities as indicators of organic pollution in Lytle Creek. Sewage and Industrial Wastes, 28(7), 906-924.

Geffard, O., Xuereb, B., Chaumot, A., Geffard, A., Biagianti, S., Noel, C.et al. (2010)., Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment. Environ Toxicol Chem, 29(10), 2249-2259.

Ghernaout, D., Alshammari, Y., & Alghamdi, A., (2018). Improving energetically operational procedures in wastewater treatment plants. International Journal of Advanced and Applied Sciences, 5, 64-72.

Gismondi E, Beisel JN, Cossu-Leguille C., (2012). Influence of gender and season on reduced glutathione concentration and energy reserves of Gammarus roeseli. Environmental Research 118:47–52.

Gismondi, E., Cossu-Leguille, C. and Beisel, J.N., (2013). Do male and female gammarids defend themselves differently during chemical stress?. Aquatic toxicology, 140, 432-438.

Gismondi, E., Thomé, J. P., Urien, N., Uher, E., Baiwir, D., Mazzucchelli, G., & Lebrun, J. D., (2017). Ecotoxicoproteomic assessment of the functional alterations caused by chronic metallic exposures in gammarids. Environmental Pollution, 225, 428-438.

GO Consortium, (2008). The gene ontology project in 2008. Nucleic acids research, 36(suppl_1), D440-D444.

Goedmakers, A., (1972). Gammarus fossarum Koch, 1835: Redescription based on neotype material and notes on its local variation (Crustacea, Amphipoda). Bijdragen tot de Dierkunde, 42(2), 124-138.

Goedmakers, A., (1980). Microgeographic races of Gammarus fossarum Koch, 1836. Crustaceana. Supplement, 216-224.

Golding, C. J., Gobas, F. A., & Birch, G. F., (2008). A fugacity approach for assessing the biooaccumulation of hydrophobic organic compounds from estuarine sediment. Environmental Toxicology and Chemistry, 27(5), 1047-1054.

Gómez-Canela, C., Miller, T. H., Bury, N. R., Tauler, R., & Barron, L. P. (2016). Targeted metabolomics of Gammarus pulex following controlled exposures to selected pharmaceuticals in water. Science of the Total Environment, 562, 777-788.

Gore, J.A., (1985). Restoration of rivers and streams. Butterworth Publishers, United States.

Gotz, C.W., Kase, R., & Hollender, J. (2010). Mikroverunreinigungen – Beurteilungskonzept für orga- nische Spurenstoffe aus kommunalem Abwasser. Studie im Auftrag des BAFU. Eawag, Dübendorf.

Gouveia, D., Bonneton, F., Almunia, C., Armengaud, J., Quéau, H., Degli-Esposti, D., Geffard, O. & Chaumot, A., (2018). Identification, expression, and endocrine-disruption of three ecdysone-responsive genes in the sentinel species Gammarus fossarum. Scientific reports, 8(1), 1-13.

Gouveia, D., Chaumot, A., Charnot, A., Almunia, C., François, A., Navarro, L., Armengaud, J., Salvador, A. and Geffard, O., (2017). Ecotoxico-proteomics for aquatic environmental monitoring: first in situ application of a new proteomics-based multibiomarker assay using caged amphipods. Environmental science & technology, 51(22), 13417-13426.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. and Chen, Z., (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology, 29(7), 644.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. and Chen, Z., (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology, 29(7), 644.

Grabner D. S., Schertzinger G., & Sures B., (2014). Effect of multiple microsporidian infections and temperature stress on the heat shock protein 70 (hsp70) response of the amphipod Gammarus pulex. Parasites & vectors, 7(1), 170.

Graca, M., Maltby, L., & Calow, P., (1994). Comparative ecology of Gammarus pulex (L.) and Asellus aquaticus (L.) I: population dynamics and microdistribution. Hydrobiologia, 281(3), 155-162.

Graney, R.L., Giesy, J.P., (1986). Effects of long-term exposure to pentachlorophenol on the free amino-acid pool and energy reserves of the fresh-water amphipod Gammarus pseudolimnaeus Bousfield (Crustacea, Amphipoda). Ecotoxicol. Environ. Saf. 12, 233–251.

Graveley, B.R. et al., (2011). The developmental transcriptome of Drosophila melanogaster. Nature, 471, 473–479.

Gross J. H., (2004). Mass Spectrometry, Springer, Heidelberg, Germany.

Gross, M.Y., Maycock, D.S., Thorndyke, M.C., Morritt, D. and Crane, M., (2001). Abnormalities in sexual development of the amphipod Gammarus pulex (L.) found below sewage treatment works. Environmental Toxicology and Chemistry: An International Journal, 20(8), 1792-1797.

Guler, Y., & Ford, A. T., (2010). Anti-depressants make amphipods see the light. Aquatic Toxicology, 99(3), 397-404.

Guler, Y., Short, S., Etxabe, A.G., Sherhod, C.M., Kille, P. and Ford, A.T., (2015). Impacts of a newly identified behaviour-altering trematode on its host amphipod: from the level of gene expression to population. Parasitology, 142(12), 1469-1480.

Haas, B. and Papanicolaou, A., (2015). TransDecoder (find coding regions within transcripts). URL: https://github.com/TransDecoder/TransDecoder.

Hadfield, J. and Eldridge, M.D., (2014). Multi-genome alignment for quality control and contamination screening of next-generation sequencing data. Frontiers in genetics, 5, 31.

Hall, R. D. (2006). Plant metabolomics: from holistic hope, to hype, to hot topic. New phytologist, 169(3), 453-468.

Hampel, M., Alonso, E., Aparicio, I., Bron, J. E., Santos, J. L., Taggart, J. B., & Leaver, M. J., (2010). Potential physiological effects of pharmaceutical compounds in Atlantic salmon (Salmo salar) implied by transcriptomic analysis. Environmental Science and Pollution Research, 17(4), 917-933.

Harley, J., & Miller, S., (2015). Zoology (10th Edition): McGraw-Hill Education.

Hernández, A.F., Parrón, T., Tsatsakis, A.M., Requena, M., Alarcón, R. and López-Guarnido, O., (2013). Toxic effects of pesticide mixtures at a molecular level: their relevance to human health. Toxicology, 307, 136-14.

Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., & Tai, H. L., (2003). Electrospray ionisation mass spectrometry: principles and clinical applications. The Clinical Biochemist Reviews, 24(1), 3.

Hochstrat, R., Wintgens, T., Melin, T., Jeffrey, P., (2006). Assessing the European wastewater reclamation and reuse potential e a scenario analysis.Desalination 188 (1e3), 1e8.

Hoback WW, Barnhart MC., (1996). Lethal limits and sublethal effects of hypoxia on the amphipod Gammarus pseudolimnaeus. Journal ofthe North American Benthological Society 15(1):117–126.

Hodkinson, I. D., & Jackson, J. K., (2005). Terrestrial and aquatic invertebrates as bioindicators for environmental monitoring, with particular reference to mountain ecosystems. Environmental management, 35(5), 649-666.

Hoffmann E. D. and V. Stroobant, (2001). Mass Spectrometry Principles and Applications, John Wiley & Sons, Chichester, UK.

Hogstrand, C., Balesaria, S., Glover, C.N., (2002). Application of genomics and proteomics for study of the integrated response to zinc exposure in a non-model fish species, the rainbow trout. Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 133, 523–535.

Homola, E., & Chang, E. S., (1997). Methyl farnesoate: crustacean juvenile hormone in search of functions. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 117(3), 347-356.

Hook, S. E., Osborn, H. L., Spadaro, D. A., & Simpson, S. L., (2014). Assessing mechanisms of toxicant response in the amphipod Melita plumulosa through transcriptomic profiling. Aquatic Toxicology, 146, 247-257.

Horgan, R. P., & Kenny, L. C., (2011). 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. The Obstetrician & Gynaecologist, 13(3), 189-195.

Hou, Z., Jiang, P., Swanson, S.A., Elwell, A.L., Nguyen, B.K.S., Bolin, J.M., Stewart, R. and Thomson, J.A., (2015). A cost-effective RNA sequencing protocol for large-scale gene expression studies. Scientific reports, 5, 9570.

Howarth, R.W., Anderson, D.B., Cloern, J.E., Elfring, C., Hopkinson, C.S., Lapointe, B., Maloney, T.J., Marcus, N., McGlathery, K., Sharpley, A.N. and Walker, D., (2000). Issues in ecology: Nutrient pollution of coastal rivers, bays, and seas.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M. and Jensen, L.J., (2015). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic acids research, 44(D1), D286-D293.

Hutton, M., Venturini, N., García-Rodríguez, F., Brugnoli, E., & Muniz, P., (2015). Assessing the ecological quality status of a temperate urban estuary by means of benthic biotic indices. Marine Pollution Bulletin, 91(2), 441-453.

Hyne, R. V., (2011). Review of the reproductive biology of amphipods and their endocrine regulation: identification of mechanistic pathways for reproductive toxicants. Environmental toxicology and chemistry, 30(12), 2647-2657.

Hynes HBN. Historical perspective and future direction of biological monitoring of aquatic systems. In: Loeb SL, Spacie A, editors. Biological monitoring of aquatic systems. Boca Raton: Lewis Publishers., (1994). 11–22.

Hynes, H. B. N., (1955). The Reproductive Cycle of Some British Freshwater Gammaridae. Journal of Animal Ecology, 24(2), 352-387.

Illumina, (2010). De Novo assembly using Illumina Reads. Technical Note: Sequencing.

Illumina, (2014). Estimating Sequencing Coverage. Technical Note: Sequencing.

Ings, J. S., Servos, M. R., & Vijayan, M. M., (2011). Hepatic transcriptomics and protein expression in rainbow trout exposed to municipal wastewater effluent. Environmental science & technology, 45(6), 2368-2376.

Iwanowicz, L. R.; Blazer, V. S.; McCormick, S. D.; VanVeld, P. A.; Ottinger, C. A. Aroclor 1248 exposure leads to immunomodulation, decreased disease resistance and endocrine disruption in the brown bullhead, Ameiurus nebulosus. Aquat. Toxicol, (2009), 93 (1), 70−82.

Jacobson, T., Holmström, K., Yang, G., Ford, A. T., Berger, U., & Sundelin, B., (2010). Perfluorooctane sulfonate accumulation and parasite infestation in a field population of the amphipod Monoporeia affinis after microcosm exposure. Aquatic Toxicology, 98(1), 99-106.

Jażdżewski, K., (1977). Remarks on the morphology of Gammarus fossarum Koch, 1835, and Gammarus kischineffensis Schellenberg, 1937. Crustaceana. Supplement, 201-211.

Jones, M., & Johnson, I., (1992). Responses of the brackish-water amphipod Gammarus duebeni (Crustacea) to saline sewage. Netherlands journal of sea research, 30, 141-147.

Jungmann, D., Ladewig, V., Ludwichowski, K. U., Petzsch, P., & Nagel, R., (2004). Intersexuality in Gammarus fossarum KOCH-a common inducible phenomenon? Archiv für Hydrobiologie, 159(4), 511-529.

Kakuta, I., & Murachi, S., (1997). Physiological response of carp, Cyprinus carpio, exposed to raw sewage containing fish processing wastewater. Environmental Toxicology and Water Quality: An International Journal, 12(1), 1-9.

Kallis, G., & Butler, D., (2001). The EU water framework directive: measures and implications. Water policy, 3(2), 125-142.

Kanehisa, M., (2002). The KEGG database. In Novartis Foundation Symposium. Chichester; New York, 91-100.

Kell, D. B. (2006). Systems biology, metabolic modelling and metabolomics in drug discovery and development. Drug discovery today, 11(23-24), 1085-1092.

Kessner, D., Chambers, M., Burke, R., Agus, D. and Mallick, P., (2008). ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics, 24(21), 2534-2536.

Kienle, C., Vermeirssen, E. L., Schifferli, A., Singer, H., Stamm, C., & Werner, I., (2019). Effects of treated wastewater on the ecotoxicity of small streams–Unravelling the contribution of chemicals causing effects. PloS one, 14(12).

Kim, E., Goren, A., & Ast, G., (2008). Alternative splicing: current perspectives. Bioassays, 30(1), 38-47.

King, R.S. and Baker, M.E., (2011). An alternative view of ecological community thresholds and appropriate analyses for their detection: comment. Ecological Applications, 21(7), 2833-2839.

Kolanowski, W., Stolyhwo, A. and Grabowski, M., (2007). Fatty acid composition of selected fresh water gammarids (Amphipoda, Crustacea): a potentially innovative source of omega-3 LC PUFA. Journal of the American Oil Chemists' Society, 84(9), 827-833.

Kolde, R., (2019). Pheatmap: pretty heatmaps. R package version, 61: 617.

Könemann, S., Müller, Y., Tschentscher, D., Krauss, M., Inostroza, P. A., Brückner, I., & Hollert, H., (2019). Combination of In Situ Feeding Rate Experiments and Chemical Body Burden Analysis to Assess the Influence of Micropollutants in Wastewater on Gammarus pulex. International journal of environmental research and public health, 16(5), 883.

Koolman, J., (1990). Ecdysteroids. Zoological science, 7(4), 563-580.

Kolkwitz, R., & Marson, M. (1909). Ökologie der tierischen Saprobien Internationale Revue der gesamten. Hydrobiologie, 2, 126-152.

Kovacevic, V., Simpson, A. J., & Simpson, M. J. (2016). 1H NMR-based metabolomics of Daphnia magna responses after sub-lethal exposure to triclosan, carbamazepine and ibuprofen. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 19, 199-210.

Kovacs, T. G., Martel, P. H., O'Connor, B. I., Hewitt, L. M., Parrott, J. L., McMaster, M. E., & Van Den Heuvel, M. R., (2013). A survey of Canadian mechanical pulp and paper mill effluents: Insights concerning the potential to affect fish reproduction. Journal of Environmental Science and Health, Part A, 48(10), 1178-1189.

Krewski, D., Acosta, D., Jr., Andersen, M., Anderson, H., Bailar, J. C., 3rd, Boekelheide, K., Brent, R., Charnley, G., Cheung, V. G., Green, S., (2010). Toxicity testing in the 21st century: A vision and a strategy. Journal of Toxicol. Environ. Health B Crit. Rev. 13, 51–138.

Krueger, F., (2012). Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq).

Kunz, P.Y., Kienle, C. & Gerhardt, A., (2010). Gammarus spp. in aquatic ecotoxicology and water quality assessment: toward integrated multilevel tests. In Reviews of Environmental Contamination and Toxicology Volume 205, 1-76.

Ladewig, V., Jungmann, D., Köhler, H.R., Schirling, M., Triebskorn, R. and Nagel, R., (2006). Population structure and dynamics of Gammarus fossarum (Amphipoda) upstream and downstream from effluents of sewage treatment plants. Archives of environmental contamination and toxicology, 50(3), 370-383.

Larkin, P., Folmar, L.C., Hemmer, M.J., Poston, A.J., Denslow, N.D., (2003). Expression profiling of estrogenic compounds using a sheepshead minnow cDNA macroarray. EHP Toxicogenomics 111, 29–36.

Larkin, P., Sabo-Attwood, T., Kelso, J., Denslow, N.D., (2002). Gene expression analysis of largemouth bass exposed to estradiol, nonylphenol, and p,p'-DDE. Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 133, 543–557.

Laufer, H., Borst, D., Baker, F. C., Reuter, C. C., Tsai, L. W., Schooley, D. A., & Sinkus, M., (1987). Identification of a juvenile hormone-like compound in a crustacean. Science, 235(4785), 202-205.

Le Roux M-L., (1933) Recherches sur la sexualite' des gammariens. Bull Biol 16(suppl): 1–139.

Lebrun, J. D., Perret, M., Uher, E., Tusseau-Vuillemin, M.-H., & Gourlay-Francé, C., (2011). Waterborne nickel bioaccumulation in Gammarus pulex: Comparison of mechanistic models and influence of water cationic composition. Aquatic Toxicology, 104(3), 161-167.

Lebrun, J.D., Uher, E. and Fechner, L.C., (2017). Behavioural and biochemical responses to metals tested alone or in mixture (Cd-Cu-Ni-Pb-Zn) in Gammarus fossarum: From a multi-biomarker approach to modelling metal mixture toxicity. Aquatic Toxicology, 193, 160-167.

Leroy, D., Haubruge, E., De Pauw, E., Thomé, J. P., & Francis, F., (2010). Development of ecotoxicoproteomics on the freshwater amphipod Gammarus pulex: identification of PCB biomarkers in glycolysis and glutamate pathways. Ecotoxicology and environmental safety, 73(3), 343-352.

Leung, K. M., (2018). Joining the dots between omics and environmental management. Integrated environmental assessment and management, 14(2), 169-173.

Li, W., (2012). Volcano plots in analyzing differential expressions with mRNA microarrays. Journal of bioinformatics and computational biology, 10(06), 1231003.

Liebeke, M., & Bundy, J. G., (2012). Tissue disruption and extraction methods for metabolic profiling of an invertebrate sentinel species. Metabolomics, 8(5), 819-830.

Liess, M., Ohe, P.C.V.D., (2005). Analyzing effects of pesticides on invertebrate communities in streams. Environ. Toxicol. Chem. 24 (4), 954–965.

Liess, M.; Schäfer, R.B.; Schriever, C.A., (2008). The footprint of pesticide stress in communities -Species traits reveal community effects of toxicants. Sci. Total Environ., 406, 484–490.

Lindén, O., (1976). Effects of oil on the amphipod Gammarus oceanicus. Environmental Pollution (1970), 10(4), 239-250.

Lindon, J. C., Holmes, E., & Nicholson, J. K., (2003). Peer Reviewed: So What's the Deal with Metabolomics?. ACS Publications, 384A.

Loose, M., Malla, S., & Stout, M., (2016). Real-time selective sequencing using nanopore technology. Nature methods, 13(9), 751.

Love, A.C., (2017). The sub-lethal impacts of wastewater effluent and its components on Gammarus pulex (Doctoral dissertation, University of Portsmouth).

Love, M. I., Huber, W., & Anders, S., (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology, 15(12), 550.

Lowry, J. K., & Myers, A. A., (2013). A Phylogeny and Classification of the Senticaudata subord. nov. Crustacea: Amphipoda. Zootaxa, 3610(1), 1-80.

Lowry, J., & Springthorpe, R., (2001). Amphipoda: families. Version, 1, 2.

Macher, J.N., Leese, F., Weigand, A.M. and Rozenberg, A., (2017). The complete mitochondrial genome of a cryptic amphipod species from the Gammarus fossarum complex. Mitochondrial DNA Part B, 2(1), 17-18.

MacKenzie, D. A., Defernez, M., Dunn, W. B., Brown, M., Fuller, L. J., de Herrera, S. R. S., & Goodacre, R. (2008). Relatedness of medically important strains of Saccharomyces cerevisiae as revealed by phylogenetics and metabolomics. Yeast, 25(7), 501-512.

Madeira, D., Narciso, L., Cabral, H. N., Vinagre, C., & Diniz, M. S., (2013). Influence of temperature in thermal and oxidative stress responses in estuarine fish. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 166(2), 237-243.

Major, K. M., Weston, D. P., Lydy, M. J., Wellborn, G. A., & Poynton, H. C., (2018). Unintentional exposure to terrestrial pesticides drives widespread and predictable

evolution of resistance in freshwater crustaceans. Evolutionary applications, 11(5), 748-761.

Malaj, E., Peter, C., Grote, M., Kühne, R., Mondy, C. P., Usseglio-Polatera, P.et al., (2014). Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. Proceedings of the National Academy of Sciences, 111(26), 9549-9554.

Malmqvist, B., (2002). Aquatic invertebrates in riverine landscapes. Freshwater Biology, 47(4), 679-694.

Maltby, L., (1995). Sensitivity of the crustaceans Gammarus pulex (l.) and Asellus aquaticus to short-term exposure to hypoxia and unionized ammonia: observations and possible mechanisms. Wat. Res., 29(3), 781-787.

Marcarelli, A. M., Baxter, C. V., Mineau, M. M., & Hall, R. O., (2011). Quantity and quality: unifying food web and ecosystem perspectives on the role of resource subsidies in freshwaters. Ecology, 92(6), 1215-1225.

Mardis, E. R., (2008). Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet., 9, 387-402.

Martinović-Weigelt, D., Mehinto, A. C., Ankley, G. T., Denslow, N. D., Barber, L. B., Lee, K. E., & Villeneuve, D. L., (2014). Transcriptomic effects-based monitoring for endocrine active chemicals: Assessing relative contribution of treated wastewater to downstream pollution. Environmental science & technology, 48(4), 2385-2394.

Martyniuk, C. J., (2018). Are we closer to the vision? A proposed framework for incorporating omics into environmental assessments. Environmental toxicology and pharmacology, 59, 87-93.

Marzin, A., Archaimbault, V., Belliard, J., Chauvin, C., Delmas, F., & Pont, D., (2012). Ecological assessment of running waters: do macrophytes, macroinvertebrates, diatoms and fish show similar responses to human pressures? Ecological Indicators, 23, 56-65.

Maxam, A. M., & Gilbert, W., (1977). A new method for sequencing DNA. Proceedings of the National Academy of Sciences, 74(2), 560-564.

Mayer G, Maas A, Waloszek D., (2012). Mouthpart Morphology of Three Sympatric Native and Nonnative Gammaridean Species: Gammarus pulex, G. fossarum, and Echinogammarus berilloni (Crustacea: Amphipoda). International Journal of Zoology 2012:1-23.

McCahon, C. P., & Pascoe, D., (1988a). Culture techniques for three freshwater macroinvertebrate species and their use in toxicity tests. Chemosphere, 17(12), 2471-2480.

McCahon, C. P., & Pascoe, D., (1988b). Use of Gammarus pulex (L.) in safety evaluation tests: Culture and selection of a sensitive life stage. Ecotoxicology and Environmental Safety, 15(3), 245-252.

McGettigan, P. A., (2013). Transcriptomics in the RNA-seq era. Current opinion in chemical biology, 17(1), 4-11.

McLafferty F. W., (1983). Tandem Mass Spectrometry, John Wiley & Sons, New York, NY, USA.

Mehennaoui, K., Legay, S., Serchi, T., Guérold, F., Giamberini, L., Gutleb, A. C., & Cambier, S., (2018). Identification of reference genes for RT-qPCR data normalization in Gammarus fossarum (Crustacea Amphipoda). Scientific reports, 8(1), 1-8.

Menni, C., Zierer, J., Valdes, A.M. and Spector, T.D., (2017). Mixing omics: combining genetics and metabolomics to study rheumatic diseases. Nature Reviews Rheumatology, 13(3), 174.

Metcalfe-Smith, J., (2009). Biological Water-Quality Assessment of Rivers: Use of Macroinvertebrate Communities. The Rivers Handbook: Hydrological and Ecological Principles, Volume Two, 144-170.

Metzker, M. L., (2005). Emerging technologies in DNA sequencing. Genome research, 15(12), 1767-1776.

Metzker, M. L., (2010). Sequencing technologies—the next generation. Nature reviews genetics, 11(1), 31-46.

Meyer, T. N., & Silva, A. L., (1999). Resposta celular ao estresse. Revista da Associação Médica Brasileira, 45(2), 181–188.

Mezzelani, M., Gorbi, S., & Regoli, F., (2018). Pharmaceuticals in the aquatic environments: evidence of emerged threat and future challenges for marine organisms. Marine environmental research, 140, 41-60.

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., ... & Kitano, H., (2005). The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic acids research, 33(suppl_1), D284-D288.

Mi'covi'c, V., Bulog, A., Kuˇci ´c, N., Jakovac, H., & Radoˇsevi'c-Staˇsi´, B., (2009). Metallothioneins and heat shock proteins 70 in marine mussels as sensors of environ- mental pollution in Northern Adriatic Sea. Environmental Toxicology and Pharmacology, 28,439–447.

Miracle, A.L., Toth, G.P., Lattier, D.L., (2003). The path from molecular indicators of exposure to describing dynamic biological systems in an aquatic organism: microarrays and the fathead minnow. Ecotoxicology 12, 457–462.

Mitra, K., Carvunis, A.-R., Ramesh, S. K., and Ideker, T., (2013). Integrative approaches for finding modular structure in bio- logical networks. Nat. Rev. Genet. 14, 719–732.

Monteoliva, L., & Albar, J. P., (2004). Differential proteomics: an overview of gel and non-gel based approaches. Briefings in Functional Genomics, 3(3), 220-239.

Moreira-de-Sousa, C., de Souza, R. B., & Fontanetti, C. S., (2018). HSP70 as a biomarker: an excellent tool in environmental contamination analysis—a review. Water, Air, & Soil Pollution, 229(8), 264.

Morimoto, R., & Fodor, E., (1984). Cell-specific expression of heath shock proteins in chicken reticulocytes and lympho- cytes. The Journal ofCell Biology, 99,1316–1323.

Morrissey, C. A., Boldt, A., Mapstone, A., Newton, J., & Ormerod, S. J., (2013). Stable isotopes as indicators of wastewater effects on the macroinvertebrates of urban rivers. Hydrobiologia,700(1), 231-244.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B., (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods, 5(7), 621.

Mortensen, A. S., & Arukwe, A., (2007). Effects of 17α-ethynylestradiol on hormonal responses and xenobiotic biotransformation system of Atlantic salmon (Salmo salar). Aquatic toxicology, 85(2), 113-123.

Motakis, E.S., Nason, G.P., Fryzlewicz, P. and Rutter, G.A., (2006). Variance stabilization and normalization for one-color microarray data using a data-driven multiscale approach. Bioinformatics, 22(20), 2547-2553.

Müller, J., (1998). Genetic population structure of two cryptic Gammarus fossarum types across a contact zone. Journal of Evolutionary Biolog, 11, 79–101.

Müller, J., (2000). Mitochondrial DNA variation and the evolutionary history of cryptic Gammarus fossarum types. Molecular Phylogenetics and Evolution, 15(2), 260-268.

Nagato, E. G., Jessica, C., Lankadurai, B. P., Poirier, D. G., Reiner, E. J., Simpson, A. J., & Simpson, M. J., (2013). 1H NMR-based metabolomics investigation of Daphnia magna responses to sub-lethal exposure to arsenic, copper and lithium. Chemosphere, 93(2), 331-337.

Nagato, E. G., Simpson, A. J., & Simpson, M. J., (2016). Metabolomics reveals energetic impairments in Daphnia magna exposed to diazinon, malathion and bisphenol-A. Aquatic Toxicology, 170, 175-186.

Nagel, R., Jungmann, D., Schirling, M., Ladewig, V., Koehler, A., & Triebskorn, R., (2002). Intersexuality in Gammarus fossarum Koch, 1835 (Amphipoda). Crustaceana, 75(11), 1289-1299.

Nelson, E. D., Do, H., Lewis, R. S., & Carr, S. A., (2011). Diurnal variability of pharmaceutical, personal care product, estrogen and alkylphenol concentrations in effluent from a tertiary wastewater treatment facility. Environmental science & technology, 45(4), 1228-1234.

Neumann, N.F., Galvez, F., (2002). DNA microarrays and toxicogenomics: applications for ecotoxicology? Biotechnol. Adv. 20, 391–419.

Neuparth, T., Lopes, A.I., Alves, N., Oliveira, J.M. and Santos, M.M., (2019). Does the antidepressant sertraline show chronic effects on aquatic invertebrates at environmentally relevant concentrations? A case study with the keystone amphipod, Gammarus locusta. Ecotoxicology and environmental safety, 183, 109486.

Paganelli, D., Caronni, S., Marchini, A., Gazzola, A., & Sconfietti, R., (2016). Dynamics and population structure of native Echinogammarus stammeri (Karaman, 1931) (Crustacea: Amphipoda) and non-native Gammarus roeselii Gervais, 1835 (Crustacea: Amphipoda) occurring in sympatry in Northern Italy. Italian Journal of Zoology, 1-8.

Pamuru, R. R., (2019). Endocrinology of Reproduction in Crustaceans. In Comparative Endocrinology of Animals. IntechOpen.

Pardo, I., Gómez-Rodríguez, C., Wasson, J. G., Owen, R., van de Bund, W., Kelly, M., & Mengin, N., (2012). The European reference condition concept: a scientific and technical approach to identify minimally-impacted river ecosystems. Science of the Total Environment, 420, 33-42.

Pascoe, D., Karntanut, W. and Müller, C.T., (2003). Do pharmaceuticals affect freshwater invertebrates? A study with the cnidarian Hydra vulgaris. Chemosphere, 51(6), 521-528.

Pascoe, D., Kidwards, T. J., Maund, S. J., Muthi, E., & Taylor, E. J., (1994). Laboratory and field evaluation of a behavioural bioassay—The Gammarus pulex (L.) precopula separation (GaPPS) test. Water Research, 28(2), 369-372.

Pejcic, B., Eadington, P., & Ross, A., (2007). Environmental monitoring of hydrocarbons: a chemical sensor perspective. Environmental Science & Technology, 41(18), 6333-6342.

Peschke, K., Geburzi, J., Ko, H.R., Wurm, K. and Triebskorn, R., (2014). Invertebrates as indicators for chemical stress in sewage-influenced stream systems: toxic and endocrine effects in gammarids and reactions at the community level in two tributaries of Lake Constance, Schussen and Argen. Ecotoxicology and environmental safety, 106, 115-125.

Pinkster, S. and Scheepmaker, M., (1994). Hybridization experiments and the taxonomy of Gammarus (Amphipoda): A contribution to the understanding of controversial results. Crustaceana, 129-143.

Piva, F., Ciaprini, F., Onorati, F., Benedetti, M., Fattorini, D., Ausili, A., (2011). Assessing sediment hazard through a weight of evidence approach with bioindicator organisms: a practical model to elaborate data from sediment chemistry, bioavailability, biomarkers and ecotoxicological bioassays. Chemosphere, 83(4), 475-485.

Pöckl, M., (1993). Reproductive potential and lifetime potential fecundity of the freshwater amphipods Gammarus fossarunt and G. roeseli in Austrian streams and rivers. Freshwater Biology, 30(1), 73-91.

Pöckl, M., Webb, B.W. & Sutcliffe, D.W., (2003). Life history and reproductive capacity of Gammarus fossarum and G. roeseli (Crustacea: Amphipoda) under naturally fluctuating water temperatures: a simulation study. Freshwater biology, 48(1), 53-66.

Pop, M., & Salzberg, S. L., (2008). Bioinformatics challenges of new sequencing technology. Trends in genetics, 24(3), 142-149.

Potter, C. J., Huang, H., & Xu, T., (2001). Drosophila Tsc1 functions with Tsc2 to antagonize insulin signaling in regulating cell growth, cell proliferation, and organ size. Cell, 105(3), 357-368.

Poynton, H. C., Hasenbein, S., Benoit, J. B., Sepulveda, M. S., Poelchau, M. F., Hughes, D. S., & Werren, J. H., (2018). The toxicogenome of Hyalella azteca: a model for sediment ecotoxicology and evolutionary toxicology. Environmental science & technology, 52(10), 6009-6022.

Prosdocimi, F., Peixoto, F. C., & Ortega, J. M., (2003). DNA sequences base calling by PHRED: Error pattern analysis.

Rainbow, P. S., & Luoma, S. N., (2011). Trace metals in aquatic invertebrates: Taylor and Francis Books: Boca Raton, FL.

Rajkumar, A. P., Qvist, P., Lazarus, R., Lescai, F., Ju, J., Nyegaard, M., & Christensen, J. H., (2015). Experimental validation of methods for differential gene expression analysis and sample pooling in RNA-seq. BMC genomics, 16(1), 548.

Ralston-Hooper, K. J., Adamec, J., Jannash, A., Mollenhauer, R., Ochoa-Acuña, H., & Sepúlveda, M. S., (2011). Use of GC× GC/TOF-MS and LC/TOF-MS for metabolomic analysis of Hyalella azteca chronically exposed to atrazine and its primary metabolite, desethylatrazine. Journal of applied toxicology, 31(5), 399-410.

RC Team, (2015). R: a language and environment for statistical computing. R Foundation for Statistical Computing.

Reaka, M. L., (1975). Molting in stomatopod crustaceans. I. Stages of the molt cycle, setagenesis, and morphology. Journal of Morphology, 146(1), 55-80.

Reo, N. V., (2002). NMR-based metabolomics. Drug and chemical toxicology, 25(4), 375-382.

Ridder, L. and Wagener, M., (2008). SyGMa: combining expert knowledge and empirical scoring in the prediction of metabolites. ChemMedChem: Chemistry Enabling Drug Discovery, 3(5), 821-832.

Roberts, A.P., Oris, J.T., Burton Jr, G.A. and Clements, W.H., (2005). Gene expression in caged fish as a first-tier indicator of contaminant exposure in

streams. Environmental Toxicology and Chemistry: An International Journal, 24(12), 3092-3098.

Robinson, M.D. and Smyth, G.K., (2007). Moderated statistical tests for assessing differences in tag abundance. Bioinformatics, 23, 2881–2887.

Roessner, U., Willmitzer, L., & Fernie, A., (2002). Metabolic profiling and biochemical phenotyping of plant systems. Plant Cell Reports, 21(3), 189-196.

Romano, N., Safee, M. A., Ebrahimi, M., & Arshad, A., (2016). Fatty acid compositional changes during the embryonic development of the swimming crab, Portunus pelagicus (Portunidae: Decapoda). Invertebrate Reproduction & Development, 60(2), 112-117.

Romero, I.G., Pai, A.A., Tung, J. and Gilad, Y., (2014). RNA-seq: impact of RNA degradation on transcript quantification. BMC biology, 12(1), 42.

Rosen, M.J., Li, F., Morrall, S.W. and Versteeg, D.J., (2001). The relationship between the interfacial properties of surfactants and their toxicity to aquatic organisms. Environmental science & technology, 35(5), 954-959.

Roux, A. L., (1971). Les Gammares du groupe pulex. Essai de systématique biologique. II. Quelques caractéristiques écologiques et physiologiques. Arch. Zool. exp. gén, 112, 471-503.

Sambles, C., (2007). Derivation of the molecular basis of endocrine disruption in aquatic invertebrates. Doctoral dissertation, Cardiff University, UK.

Samuelsson, L. M., Björlenius, B., Förlin, L., & Larsson, D. J., (2011). Reproducible 1H NMR-based metabolomic responses in fish exposed to different sewage effluents in two separate studies. Environmental science & technology, 45(4), 1703-1710.

Sanchez, W., Sremski, W., Piccini, B., Palluel, O., Maillot-Marechal, E., Betoulle, S., Jaffal, A., Ait-Aissa, S., Brion, F., Thybaud, E. & Hinfray, N., (2011). Adverse effects in wild fish living downstream from pharmaceutical manufacture discharges. Environment International, 37(8), 1342-134.

Sandahl, J. F., Baldwin, D. H., Jenkins, J. J., & Scholz, N. L., (2007). A sensory system at the interface between urban stormwater runoff and salmon survival. Environmental Science & Technology, 41(8), 2998-3004.

Sanger, F., Nicklen, S., & Coulson, A. R., (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences, 74(12), 5463-5467.

Sarakinos, H. C., & Rasmussen, J. B., (1997). Use of bioassay-based whole effluent toxicity (WET) tests to predict benthic community response to a complex industrial effluent. Journal of Aquatic Ecosystem Stress and Recovery, 6(2), 141-157.

SCENIHR (Scientific Committee on Emerging and Newly Identified Health Risks)., (2012). Addressing the New Challenges for Risk Assessment: Discussion paper approved for public consultation in view of receiving feedback from stakeholders for its further development, October 8, 2012.

Schadt, E. E., Turner, S., & Kasarskis, A., (2010). A window into third-generation sequencing. Human molecular genetics, 19(R2), R227-R240.

Scheepmaker, M. & J. van Dalfsen, (1989). Genetic differentia- tion in Gammarus fossarum and G. caparti (Crustacea, Amphipoda) with reference to G. pulex pulex in north-western Europe. Bijdragen Tot De Dierkunde, 59, 127–139.

Schill R.O. and Köhler H.R., (2004). Does the environment or the source of the population define stress status and energy supply in the freshwater amphipod, Gammarus fossarum? Ecotoxicology, 13(7), 683-695.

Schirling, M., Jungmann, D., Ladewig, V., Nagel, R., Triebskorn, R., & Köhler, H. R., (2005). Endocrine effects in Gammarus fossarum (Amphipoda): influence of wastewater effluents, temporal variability, and spatial aspects on natural populations. Archives of environmental contamination and toxicology, 49(1), 53-61.

Schmittgen, T.D. and Livak, K.J., (2008). Analyzing real-time PCR data by the comparative Ct method. Nature protocols, 3(6), 1101.

Schneider, I., Oehlmann, J., & Oetken, M., (2015). Impact of an estrogenic sewage treatment plant effluent on life-history traits of the freshwater amphipod Gammarus pulex. Journal of Environmental Science and Health, Part A, 50(3), 272-281.

Schneiderman, H.A. and Gilbert, L.I., (1958). Substances with juvenile hormone activity in Crustacea and other invertebrates. The Biological Bulletin, 115(3), 530-535.

Schock, T. B., Stancyk, D. A., Thibodeaux, L., Burnett, K. G., Burnett, L. E., Boroujerdi, A. F., & Bearden, D. W., (2010). Metabolomic analysis of Atlantic blue

crab, Callinectes sapidus, hemolymph following oxidative stress. Metabolomics, 6(2), 250-262.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T., (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC molecular biology, 7(1), 3.

Schultz, M. M., Furlong, E. T., Kolpin, D. W., Werner, S. L., Schoenfuss, H. L., Barber, L. B.et al., (2010). Antidepressant pharmaceuticals in two US effluent-impacted streams: occurrence and fate in water and sediment, and selective uptake in fish neural tissue. Environmental Science & Technology, 44(6), 1918-1925.

Schwaiger, J., (2001). Histopathological alterations and parasite infection in fish: indicators of multiple stress factors. Journal of Aquatic Ecosystem Stress and Recovery, 8(3-4), 231-240.

Schwarzenbach, R. P., Egli, T., Hofstetter, T. B., von Gunten, U., & Wehrli, B., (2010). Global water pollution and human health. Annual Review of Environment and Resources, 35, 109-136.

Schwarzenbach, R.P., et al., (2006). The challenge of micropollutants in aquatic systems. Science 313, 1072–1077.

Seel, P., et al., (1996). Kläranlagen als Haupteintragspfad für Pflanzenschutzmittel in ein Fließgewässer: Bilanzierung der Einträge. Vom Wasser 86, 247–262.

Sheader M., (1983). The reproductive biology and ecology of Gammarus duebeni (Crustacea: Amphipoda) in southern England. J Mar Biol Ass U K 63:517–540.

Shiau, S. Y., & Huang, S. Y., (2001). Dietary folic acid requirement determined for grass shrimp, Penaeus monodon. Aquaculture, 200(3-4), 339-347.

Short, S., Yang, G., Guler, Y., Green Etxabe, A., Kille, P., & Ford, A. T., (2014). Crustacean intersexuality is feminization without demasculinization: implications for environmental toxicology. Environmental science & technology, 48(22), 13520-13529.

Short, S., Yang, G., Kille, P., & Ford, A. T., (2014). Vitellogenin is not an appropriate biomarker of feminisation in a crustacean. Aquatic toxicology, 153, 89-97.

Shugart, L. R., (1995). Environmental genotoxicology. In “Fundamentals of Aquatic Toxicology”, Rand GM.

Siegismund, H. R., & Müller, J., (1991). Genetic structure of Gammarus fossarum populations. Heredity, 66(3), 419.

Siegismund, H. R., (1988). Genetic differentiation in populations of the freshwater amphipods Gammarus roeseli and Gammarus fossarum. Hereditas, 109(2), 269-276.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M., (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31(19), 3210-3212.

Simmons, D. B., Benskin, J. P., Cosgrove, J. R., Duncker, B. P., Ekman, D. R., Martyniuk, C. J., & Sherry, J. P., (2015). Omics for aquatic ecotoxicology: Control

of extraneous variability to enhance the analysis of environmental effects. Environmental toxicology and chemistry, 34(8), 1693-1704.

Singhal, S., (2013). De novo transcriptomic analyses for non-model organisms: An evaluation of methods across a multi-species data set. Molecular Ecology Resources, 13(3), 403-416.

Skinner, D. M., (1962). The structure and metabolism of a crustacean integumentary tissue during a molt cycle. The Biological Bulletin, 123(3), 635-647.

Smedsgaard, J., & Nielsen, J. (2005). Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. Journal of experimental botany, 56(410), 273-286.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. and Siuzdak, G., (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Analytical chemistry, 78(3), 779-787.

Snape, J. R., Maund, S. J., Pickford, D. B., & Hutchinson, T. H., (2004). Ecotoxicogenomics: the challenge of integrating genomics into aquatic and terrestrial ecotoxicology. Aquatic toxicology, 67(2), 143-154.

Söyüt, H., Beydemir, S., Ceyhun, S. B., Erdogan, O., & Kaya, E. D., (2012). Changes in carbonic anhydrase activity and gene expression ofHsp70 in rainbow trout (Oncorhynchus mykiss) muscle aft er exposure to some metals. Turkish Journal of Veterinary and Animal Science, 36(5), 499–508.

Sornom P, Felten V, Médoc V, Sroda S, Rousselle P, Beisel JN., (2010). Effect of gender on physiological and behavioural responses of Gammarus roeseli

(Crustacea Amphipoda) to salinity and temperature. Environmental pollution 158(5):1288–1295.

Spycher, S., Mangold, S., Doppler, T., Junghans, M., Wittmer, I., Stamm, C. and Singer, H., (2018). Pesticide risks in small streams—How to get as close as possible to the stress imposed on aquatic organisms. Environmental science & technology, 52(8), 4526-4535.

Stendera, S., Adrian, R., Bonada, N., Cañedo-Argüelles, M., Hugueny, B., Januschke, K.et al., (2012). Drivers and stressors of freshwater biodiversity patterns across different ecosystems and scales: a review. Hydrobiologia, 696(1), 1-28.

Stillman, J.H. and Hurt, D.A., (2015). Crustacean genomics and functional genomic responses to environmental stress and infection. Physiology, 4, 420.

Sturla, S. J., Boobis, A. R., FitzGerald, R. E., Hoeng, J., Kavlock, R. J., Schirmer, K., & Peitsch, M. C., (2014). Systems toxicology: from basic research to risk assessment. Chemical research in toxicology, 27(3), 314-329.

Sumpter, J. P., (2005). Endocrine disrupters in the aquatic environment: an overview. Acta hydrochimica et hydrobiologica, 33(1), 9-16.

Sun, P., Quan, Z., Zhang, B., Wu, T., & Xi, R., (2010). TSC1/2 tumour suppressor complex maintains Drosophila germline stem cells by preventing differentiation. Development, 137(15), 2461-2469.

Sutcliffe DW., (1992). Reproduction in Gammarus (Crustacea: Amphipoda): Male strategies. Freshwater Forum 3, 97–109.

Sutcliffe, D. W., (1993). Reproduction in Gammarus (Crustacea, Amphipoda): female strategies. Freshwater Forum, 3(1), 26-64.

Sutcliffe, D., (1992). Reproduction in Gammarus (Crustacea, Amphipoda): basic processes. Freshwater Forum, 2(2), 102-128.

Sutcliffe, D.W., (2010). Reproduction in Gammarus (Crustacea, Amphipoda) female strategies. In Freshwater Forum. Vol. 3, No. 1.

Suzuki, Y., Matsuoka, T., Iimura, Y., & Fujiwara, H., (2002). Ecdysteroid-dependent expression of a novel cuticle protein gene BMCPG1 in the silkworm, Bombyx mori. Insect biochemistry and molecular biology, 32(6), 599-607.

Tetreault, G. R., Bennett, C. J., Shires, K., Knight, B., Servos, M. R., & McMaster, M. E., (2011). Intersex and reproductive impairment of wild fish exposed to multiple municipal wastewater discharges. Aquatic Toxicology, 104(3-4), 278-290.

The UniProt Consortium, (2017). On expert curation and scalability: UniProtKB/Swiss-Prot as a case study. Bioinformatics, 33(21), 3454-3460.

Theodorescu D, Mischak H, (2007). Mass spectrometry-based proteomics in urine biomarker discovery. Wld J Urol, 25:435–43.

Thermo Fisher Scientific, (2008). 260/280 and 260/230 Ratios. T009-Technical Bulletin, 1-2.

Thornton, B., & Basu, C., (2011). Real-time PCR (qPCR) primer design using free online software. Biochemistry and Molecular Biology Education, 39(2), 145-154.

Thorpe, K. L., Benstead, R., Eccles, P., Maack, G., Williams, T., & Tyler, C. R., (2008). A practicable laboratory flow-through exposure system for assessing the health effects of effluents in fish. Aquatic Toxicology, 88(3), 164-172.

Todd, J. F., (1991). Recommendations for nomenclature and symbolism for mass spectroscopy (including an appendix of terms used in vacuum technology). Pure and applied chemistry, 63(10), 1541-1566.

Trapp, J., Almunia, C., Gaillard, J. C., Pible, O., Chaumot, A., Geffard, O., & Armengaud, J., (2016). Proteogenomic insights into the core-proteome of female reproductive tissues from crustacean amphipods. Journal of proteomics, 135, 51-61.

Trapp, J., Armengaud, J., Pible, O., Gaillard, J. C., Abbaci, K., Habtoul, Y. & Geffard, O., (2015). Proteomic investigation of male Gammarus fossarum, a freshwater crustacean, in response to endocrine disruptors. Journal of proteome research, 14(1), 292-303.

Trapp, J., Armengaud, J., Salvador, A., Chaumot, A., & Geffard, O., (2014). Next-generation proteomics: toward customized biomarkers for environmental biomonitoring. Environmental science & technology, 48(23), 13560-13572.

Triebskorn, R., Telcean, I., Casper, H., Farkas, A., Sandu, C., Stan, G.et al., (2008). Monitoring pollution in River Mureş, Romania, part II: metal accumulation and histopathology in fish. Environmental Monitoring and Assessment, 141(1-3), 177-188.

Truebano, M., Tills, O. and Spicer, J.I., (2016). Embryonic transcriptome of the brackishwater amphipod Gammarus chevreuxi. Marine genomics, 28, 5-6.

Turja, R., Guimarães, L., Nevala, A., Kankaanpää, H., Korpinen, S. and Lehtonen, K.K., (2014). Cumulative effects of exposure to cyanobacteria bloom extracts and benzo [a] pyrene on antioxidant defence biomarkers in Gammarus oceanicus (Crustacea: Amphipoda). Toxicon, 78, 68-77.

UNEP, (2012). Global Chemical Outlook; Towards Sound Management of Chemicals; United Nations Environment Programme.

Union European (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy ("Water Framework Directive"). Off J Eur Commun 327(1):1–72.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G., (2012). Primer3—new capabilities and interfaces. Nucleic acids research, 40(15), e115-e115.

Utz, R. M., Hilderbrand, R. H., & Boward, D. M., (2009). Identifying regional differences in threshold responses of aquatic invertebrates to land cover gradients. Ecological Indicators, 9(3), 556- 567.

Väinölä, R., Witt, J., Grabowski, M., Bradbury, J. H., Jazdzewski, K., & Sket, B., (2008). Global diversity of amphipods (Amphipoda; Crustacea) in freshwater. Hydrobiologia, 595(1), 241-255.

Vajda, A. M., Barber, L. B., Gray, J. L., Lopez, E. M., Woodling, J. D., & Norris, D. O., (2008). Reproductive disruption in fish downstream from an estrogenic wastewater effluent. Environmental Science & Technology, 42(9), 3407-3414.

Vera-Chang, M. N., St-Jacques, A. D., Gagné, R., Martyniuk, C. J., Yauk, C. L., Moon, T. W., & Trudeau, V. L., (2018). Transgenerational hypocortisolism and behavioral disruption are induced by the antidepressant fluoxetine in male zebrafish Danio rerio. Proceedings of the National Academy of Sciences, 115(52), E12435-E12442.

Viessman, W., Hammer, M. J., Perez, E. M., & Chadik, P., (2009). Water supply and pollution control: Pearson Prentice Hall New Jersey.

Voulvoulis, N., & Georges, K., (2016). Industrial and Agricultural Sources and Pathways of Aquatic Pollution Impact of Water Pollution on Human Health and Environmental Sustainability (pp. 29-54): IGI Global.

Vrana, B.; Allan, I.J.; Greenwood, R.; Mills, G.A.; Dominiak, E.; Svensson, K.; Knutsson, J.; Morrison, G., (2005). Passive sampling techniques for monitoring pollutants in water. TrAC Trends Anal. Chem., 24, 845–868.

Wang, W., Yang, S., Wang, C., Shi, L., Guo, H., & Chan, S., (2017). Gill transcriptomes reveal involvement of cytoskeleton remodeling and immune defense in ammonia stress response in the banana shrimp Fenneropenaeus merguiensis. Fish & shellfish immunology, 71, 319-328.

Wang, X et al., (2010). Isoform abundance inference provides a more accurate estimation of gene expression levels in RNA-seq. J. Bioinform. Comput. Biol., 8, 177–192.

Warne, M.S.J.; Hawker, D.W., (1995). The number of components in a mixture determines whether synergistic and antagonistic or additive toxicity predominate: The funnel hypothesis. Ecotoxicol. Environ. Saf., 31, 23–28.

Warren, C.E., (1971). Biology and water pollution control. W. B. Saunders Company, West Washington Square, Philadelphia, United States.

Watanabe, K. H., Jensen, K. M., Orlando, E. F., & Ankley, G. T., (2007). What is normal? A characterization of the values and variability in reproductive endpoints of the fathead minnow, Pimephales promelas. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 146(3), 348-356.

Weber, R., (2016). Updated Water and Wastewater Treatment Plants in South Carolina and Reno.

Weisburg, W.G., Barns, S.M., Pelletier, D.A. & Lane, D.J., (1991). 16S ribosomal DNA amplification for phylogenetic study. Journal of bacteriology, 173(2), 697-703.

Weiss, M., Macher, J.N., Seefeldt, M.A. and Leese, F., (2014). Molecular evidence for further overlooked species within the Gammarus fossarum complex (Crustacea: Amphipoda). Hydrobiologia, 721(1), 165-184.

Weis, J. S., & Weis, P., (1995). Effects of embryonic exposure to methylmercury on larval prey-capture ability in the mummichog, fundulus heteroclitus. Environmental Toxicology and Chemistry: An International Journal, 14(1), 153-156.

Werner, I., & Hitzfeld, B. (2012). 50 Years of Ecotoxicology since Silent Spring–A Review. GAIA-Ecological Perspectives for Science and Society, 21(3), 217-224.

Wenn, C. L., (2008). Do freshwater macroinvertebrates reflect water quality improvements following the removal of point source pollution from Spen Beck, West Yorkshire. Earth and Environment, 3, 369-406.

West, J. M., (1997). Ultrastructural and contractile activation properties of crustacean muscle fibres over the moult cycle. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 117(3), 333-345.

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

Wigh, A., Devaux, A., Brosselin, V., Gonzalez-Ospina, A., Domenjoud, B., Aït-Aïssa, S.et al., (2016). Proposal to optimize ecotoxicological evaluation of wastewater treated by conventional biological and ozonation processes. Environmental Science and Pollution Research, 23(4), 3008-3017.

Wigh, A., Geffard, O., Abbaci, K., Francois, A., Noury, P., Bergé, A., Vulliet, E., Domenjoud, B., Gonzalez-Ospina, A., Bony, S. & Devaux, A., (2017). Gammarus fossarum as a sensitive tool to reveal residual toxicity of treated wastewater effluents. Science of the Total Environment, 584, 1012-1021.

William Allwood, J., Ellis, D. I., Heald, J. K., Goodacre, R., & Mur, L. A. (2006). Metabolomic approaches reveal that phosphatidic and phosphatidyl glycerol phospholipids are major discriminatory non-polar metabolites in responses by Brachypodium distachyon to challenge by Magnaporthe grisea. The Plant Journal, 46(3), 351-368.

Woodward, G., Papantoniou, G., Edwards, F., & Lauridsen, R. B., (2008). Trophic trickles and cascades in a complex food web: impacts of a keystone predator on stream community structure and ecosystem processes. Oikos, 117(5), 683-692.

Xia, J., & Wishart, D. S., (2011). Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. Nature protocols, 6(6), 743.

Xia, J., Sinelnikov, I. V., Han, B., & Wishart, D. S., (2015). MetaboAnalyst 3.0—making metabolomics more meaningful. Nucleic acids research, 43(W1), W251-W257.

Xu, G. et al., (2010). Transcriptome and targetome analysis in MIR155 expressing cells using RNA-seq. RNA, 1610–1622.

Xuereb, B., Bezin, L., Chaumot, A., Budzinski, H., Augagneur, S., Tutundjian, R., & Geffard, O., (2011). Vitellogenin-like gene expression in freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization in females and potential for use as an endocrine disruption biomarker in males. Ecotoxicology, 20(6), 1286-1299.

Xuereb, B., Chaumot, A., Mons, R., Garric, J. and Geffard, O., (2009). Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda): intrinsic variability, reference levels, and a reliable tool for field surveys. Aquatic toxicology, 93(4), 225-233.

Zhao, L. and Jones, W.A., (2012). Expression of heat shock protein genes in insect stress responses. Invertebrate Survival Journal, 9(1), 93-101.

Zohar S, Holmes JC., (1998). Pairing success of male Gammarus lacustris infected by two acanthocephalans: A comparative study. Behav Ecol 9:206–211.

Zong, W. X., Rabinowitz, J. D., & White, E., (2016). Mitochondria and cancer. Molecular cell, 61(5), 667-676.

Chapter 8 - Appendices

8.1 Appendix A

Sample	Concentration (ng/µL)	260/280	260/230
US♂1	517.6	2.11	2.3
US♂2	168.6	2.1	1.65
US♂3	263.4	2.1	2.28
US♂4	289.1	2.09	2.22
US♂5	142	2.07	1.82
DS♂1	302.7	2.09	2.3
DS♂2	140.1	2.09	1.8
DS♂3	446.1	2.09	2.3
DS♂4	271.4	2.12	1.71
DS♂5	156.9	2.08	1.72
US♀1	215.2	2.15	1.6
US♀2	193	2.1	1.2
US♀3	164.1	2.1	1.91
US♀4	388.2	2.12	2.16
US♀5	249.1	2.09	2.19
DS♀1	255.8	2.13	1.26
DS♀2	210.5	2.11	1.83
DS♀3	298.1	2.13	1.72
DS♀4	180.8	2.12	2.13
DS♀5	388.7	2.15	1.37

Tab. 8.1: Results of the RNA purity assay performed on the samples used for the RNA-seq experiment (Chapter 3.8.4). US: Upstream; DS: Downstream.

Sample	Concentration (ng/µL)	260/280	260/230
US♂1	285.93	2.06	1.08
US♂2	378.83	2.02	0.85
US♂3	195.55	2.06	1.12
US♂4	569.32	2.06	1.34
US♂5	300.75	2.05	1.04
DS♂1	162.07	2.04	0.46
DS♂2	200.59	2.05	0.56
DS♂3	355.41	2.01	1.12
DS♂4	224.21	2.02	0.75
DS♂5	205.6	2.02	0.95
US♀1	292.11	2.06	1.85
US♀2	163.91	2.03	1.99
US♀3	411.06	2.06	2.08
US♀4	444.18	2.06	1.38
US♀5	144.53	2.1	1.81
DS♀1	245.43	2.06	1.42
DS♀2	323.57	2.08	1.31
DS♀3	77.76	1.95	0.51
DS♀4	158.05	2.03	1.83
DS♀5	228.45	2.06	1.38

Tab. 8.2: Results of the RNA purity assay performed on the samples used for the qPCR experiment (Chapter 4.4.8). US: Upstream; DS: Downstream.

8.2 Appendix B

List of tables not embedded in the thesis but provided as supplementary files:

File S2.1: Results of the chemical analysis (Chapter 2.5.3). The table includes the concentration values of 55 contaminants measured upstream and downstream of the ARA wastewater treatment plant (WWTP) in both water (ng/L) and amphipod (ng/g) samples. Apparent bioaccumulation factors (BAFs) and toxic units (TUs) are also reported.

File S3.1: Results of the quality control of the raw reads generated in the RNA-seq experiment, for all 20 samples (Chapter 3.8.7). FastQC v0.11.7 was used to perform the quality control of the reads. The reports include plots showing the base calling quality scores (Phred scores) reads, per tile sequence quality, per sequence quality scores, per base sequence content, per sequence GC, per base N content, sequence length distribution, sequence duplication level, overrepresented sequences and content of adapter sequences, for both forward (R1) and reverse (R2) reads. DS: downstream; US: upstream; Rep: replicate.

File S3.2: Annotation report of the whole set of transcripts assembled in the RNA-seq experiment against *Drosophila melanogaster* protein database (Chapter 3.9.4). The annotation was performed against *D. melanogaster* NCBI Reference

Sequence (Refseq) database using the blastp tool for the sequence alignments (Chapter 3.9.4).

File S4.1: FASTA file containing the transcript codes, length and the nucleotide sequences for the whole set of transcripts (Chapter 4.3.4).

File S4.2: FASTA file containing the transcript codes, length and the nucleotide sequences for the transcripts > 500 bp in length that were not annotated against genes belonging to bacteria, archaea or viruses, with a complete open reading frame (ORF) detected and with a maximum FPKM value across all samples > 1. These filtered transcripts were included in the differential gene expression analysis (Chapter 4.3.4).

File S4.3: Results of the gene ontology (GO) analysis conducted submitting the differentially expressed genes between male amphipods sampled up- and downstream of the WWTP in the database Panther. The Biological process GO category was explored interrogating the *D. melanogaster* GO-complete ontology (Chapter 4.4.7).

File S4.4: Results of the gene ontology (GO) analysis conducted submitting the differentially expressed genes between male amphipods sampled up- and

downstream of the WWTP in the database Panther. The Molecular function GO category was explored interrogating the *D. melanogaster* GO-complete ontology (Chapter 4.4.7).

File S4.5: Results of the gene ontology (GO) analysis conducted submitting the differentially expressed genes between male amphipods sampled up- and downstream of the WWTP in the database Panther. The Cellular component GO category was explored interrogating the D. melanogaster GO-complete ontology (Chapter 4.4.7).

File S4.6: Results of the gene ontology (GO) analysis conducted submitting the differentially expressed genes between female amphipods sampled up- and downstream of the WWTP in the database Panther. The Cellular component GO category was explored interrogating the *D. melanogaster* GO-complete ontology (Chapter 4.4.7).

File S4.7: Results of the gene ontology (GO) analysis conducted submitting the differentially expressed genes between male and female amphipods sampled downstream of the WWTP in the database Panther. The Molecular function GO category was explored interrogating the *D. melanogaster* GO-complete ontology (Chapter 4.4.7).

File S4.8: Results of the gene ontology (GO) analysis conducted submitting the differentially expressed genes between male and female amphipods sampled downstream of the WWTP in the database Panther. The Cellular component GO category was explored interrogating the *D. melanogaster* GO-complete ontology (Chapter 4.4.7).

File S4.9: Results of the gene ontology (GO) analysis conducted submitting the differentially expressed genes between male and female amphipods sampled upstream of the WWTP in the database Panther. The Biological process GO category was explored interrogating the *D. melanogaster* GO-complete ontology (Chapter 4.4.7).

File S4.10: Results of the gene ontology (GO) analysis conducted submitting the differentially expressed genes between male and female amphipods sampled upstream of the WWTP in the database Panther. The Molecular function GO category was explored interrogating the *D. melanogaster* GO-complete ontology (Chapter 4.4.7).

File S4.11: Results of the gene ontology (GO) analysis conducted submitting the differentially expressed genes between male and female amphipods sampled upstream of the WWTP in the database Panther. The Cellular component GO

category was explored interrogating the *D. melanogaster* GO-complete ontology (Chapter 4.4.7).

File S4.12: Results of the gene ontology (GO) analysis conducted submitting the overlapping genes between "Upstream females vs Upstream males" and "Downstream females vs Downstream males" lists of differentially expressed genes in the database Panther. The Biological process GO category was explored interrogating the *D. melanogaster* GO-complete ontology (Chapter 4.4.7).

File S4.13: Results of the gene ontology (GO) analysis conducted submitting the overlapping genes between "Upstream females vs Upstream males" and "Downstream females vs Downstream males" lists of differentially expressed genes in the database Panther. The Cellular component GO category was explored interrogating the *D. melanogaster* GO-complete ontology (Chapter 4.4.7).

File S4.14: Read counts expressed in Fragments Per Kilobase Million (FPKM) for all the transcripts across the 20 samples (Chapter 4.3.4).

File S4.15: List of differentially expressed genes between male amphipods sampled up- and downstream of the WWTP. The table includes the UniProt gene

descriptions, log_2-transformed fold-changes and FPKM values for each gene (Chapter 4.4.5).

File S4.16: List of differentially expressed genes between female amphipods sampled up- and downstream of the WWTP. The table includes the UniProt gene descriptions, log_2-transformed fold-changes and FPKM values for each gene (Chapter 4.4.5).

File S4.17: List of differentially expressed genes between male and female amphipods sampled upstream of the WWTP. The table includes the UniProt gene descriptions, log_2-transformed fold-changes and FPKM values for each gene (Chapter 4.4.6).

File S4.18: List of differentially expressed genes between male and female amphipods sampled downstream of the WWTP. The table includes the UniProt gene descriptions, log_2-transformed fold-changes and FPKM values for each gene (Chapter 4.4.6).

File S5.1: List of differential metabolite features (MFF) between male amphipods sampled up- and downstream of the WWTP detected using the hydrophilic negative (HILIC-NEG) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion

adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.2: List of differential metabolite features (MFF) between male amphipods sampled up- and downstream of the WWTP detected using the hydrophilic positive (HILIC-POS) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.3: List of differential metabolite features (MFF) between male amphipods sampled up- and downstream of the WWTP detected using the lipids negative (LIPIDS-NEG) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.4: List of differential metabolite features (MFF) between male amphipods sampled up- and downstream of the WWTP detected using the lipids positive (LIPIDS-POS) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.5: List of differential metabolite features (MFF) between female amphipods sampled up- and downstream of the WWTP detected using the hydrophilic negative (HILIC-NEG) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.6: List of differential metabolite features (MFF) between female amphipods sampled up- and downstream of the WWTP detected using the hydrophilic positive (HILIC-POS) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.7: List of differential metabolite features (MFF) between female amphipods sampled up- and downstream of the WWTP detected using the lipids negative (LIPIDS-NEG) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.8: List of differential metabolite features (MFF) between female amphipods sampled up- and downstream of the WWTP detected using the lipids positive

(LIPIDS-POS) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.9: List of differential metabolite features (MFF) between male and female amphipods sampled both up- and downstream of the WWTP detected using the hydrophilic negative (HILIC-NEG) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.10: List of differential metabolite features (MFF) between male and female amphipods sampled both up- and downstream of the WWTP detected using the hydrophilic positive (HILIC-POS) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.11: List of differential metabolite features (MFF) between male and female amphipods sampled both up- and downstream of the WWTP detected using the lipids negative (LIPIDS-NEG) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion

adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

File S5.12: List of differential metabolite features (MFF) between male and female amphipods sampled both up- and downstream of the WWTP detected using the lipids positive (LIPIDS-POS) ionisation mode. The file includes the *m/z* values, chromatographic retention times (RT), putative metabolite annotations, ion adducts formed for all the detected features. The results of the t-tests are also reported (Chapter 5.7).

FORM UPR16

Research Ethics Review Checklist

Please include this completed form as an appendix to your thesis (see the Postgraduate Research Student Handbook for more information

Postgraduate Research Student (PGRS) Information			**Student ID:**	843898
Candidate Name:	Domenico Roberto Caputo			
Department:	School of Biological Sciences	**First Supervisor:**	Alex Ford	
Start Date: (or progression date for Prof Doc students)	01/02/2017			

Study Mode and Route:	Part-time ❑	MPhil ❑	Integrated Doctorate (NewRoute) ❑
	Full-time ☒	MD ❑	Prof Doc (PD) ❑
		PhD ☒	

Title of Thesis:	The effects of a wastewater effluent on population, transcriptomic and metabolomic markers in the freshwater amphipod Gammarus fossarum
Thesis Word Count: (excluding ancillary data)	65717

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:
(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: http://www.ukrio.org/what-we-do/code-of-practice-for-research/)

a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	**YES**
b) Have all contributions to knowledge been acknowledged?	**YES**
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	**YES**
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	**YES**
e) Does your research comply with all legal, ethical, and contractual requirements?	**YES**

*Delete as appropriate