Molecular ecology of white-clawed crayfish *Austropotamobius pallipes*, in Ireland.

This thesis is submitted by

Daniel J. Brady, B.Sc., M.Sc.

in fulfilment of the requirements of the degree Master of Research

under the supervision of
Dr Luca Mirimin
&
Dr Orla Slattery

at the

Marine & Freshwater
Research Centre,

in

Galway-Mayo
Institute of Technology

Submitted to the Higher Education and Training Awards Council, October 2018

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of M.Sc. is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:

Candidate: Daniel Brady

ID No: G00239178

Date: 10/10/2018

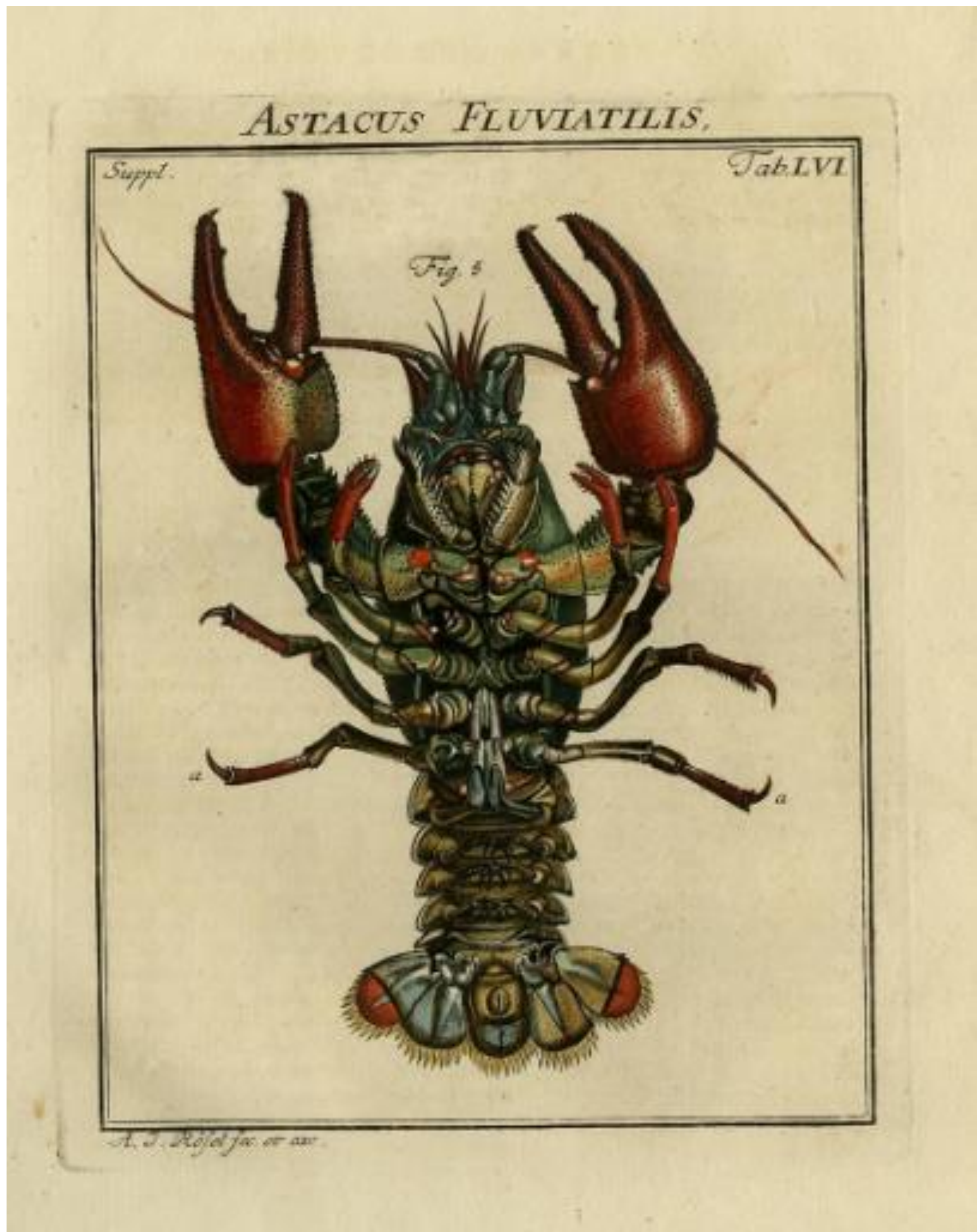
"Common and lowly as most may think the crayfish… it is yet so full of wonders that the greatest naturalist may be puzzled to give a clear account of it"

- Rösel von Rosenhof

Der Insecten Belustigung - 1740

Translated by T. H. Huxley in

The Crayfish: An Introduction to the Study of Zoology



Rösel von Rosenhof

Der Insecten Belustigung – 1740

Austropotamobius pallipes

Contents

Contents ..... **iv**

1.1 Acknowledgements vi

1.2 Table of abbreviations vii

1.3 Framework of the study vii

1.4 Orientation of this thesis vii

Abstract 1

2 Chapter 1: General introduction 2

2.1 Protecting freshwater biodiversity 2

2.2 Crayfish systematics 3

2.3 Biology of *Austropotamobius pallipes* 7

2.4 Major threats to *A. pallipes* 9

2.5 Conservation genetics 10

2.6 Protein biochemistry 12

3 Chapter 2: Genetic population structure and mitochondrial DNA variation of white-clawed crayfish *Austropotamobius pallipes* in Ireland. 13

3.1 Abstract 13

3.2 Introduction 14

3.3 Aims of this study 19

3.4 Materials and methods 20

3.4.1 Field sampling 20

3.4.2 DNA extraction and purification 20

3.4.3 PCR and sequencing 21

3.4.4 Data analysis 22

3.5 Results 24

3.5.1 Genetic diversity 24

3.5.2 Population genetic structure 29

3.6 Discussion 33

3.6.1 Implications for management 36

***4* Chapter 3: Cloning, overexpression and purification of a recombinant serpin KP12 from *Austropotamobius pallipes* 37**

4.1 Abstract 37

4.2 Introduction 38

4.2.1 Enzymes 38

4.2.2 Proteases and protease inhibitors 39

4.2.3 Recombinant protein expression and serpin overexpression 44

4.3 Aims & objectives of this study 48

4.4 Materials & methods 48

4.4.1 RNA extraction, cDNA synthesis & gene amplification 50

4.4.2 Expression plasmid construction 56

4.4.3 Bacterial transformation & construct verification 57
4.4.4 Recombinant KP12 protein over-expression 58
4.4.5 Generating 3-dimensional a structure of KP12 63
4.5 Results 63
4.5.1 General materials 63
4.5.2 RNA extraction, cDNA synthesis & gene amplification 64
4.5.3 Expression plasmid construction 65
4.5.4 Bacterial transformation & construct verification 66
4.5.5 Recombinant protein overexpression 67
4.5.6 MBP_KP12 fusion protein purification 68
4.5.7 Generating 3-dimensional a structure of KP12 73
4.6 Discussion 74
4.6.1 Recombinant plasmid construction 75
4.6.2 Overexpression of the fusion protein 78
4.7 Protein purification 81
4.8 Conclusion and suggestions 83
5 Chapter 4: Implication of this thesis & concluding recommendations 84
6 Appendix 1 87
6.1 Troubleshooting and optimisation of 16S Sanger sequencing 87
6.2 Unedited SDS-PAGE gel 89
7 References 90

1.1 Acknowledgements

My thanks go first and foremost to my siblings; Bláithín, Laura, Shauna, Ciara and Hugh, and of course my parents Helen and Donal. I could not and wouldn't be where I am today without your emotional (and financial) support, the daily chats and banter. You stand above and beyond all the rest.

My surrogate family; Marley, Úna, Elena, Paula, Maria, Conor and Cameron – for keeping me sane.

The office crew; Kristina, Jessica, James, Kerrie, Nóra, Maria and Gráinne, thanks for all the banter and many extended coffee breaks.

Shout out to all the MFRC postgrads! Particularly Adon, we did a serious amount of fieldwork and it was great to have someone that keeps spirits high, the undergrads who were amazing help; Killian, Proinsias and Sibéal, amazing workers. Also, Dr Martin Gammell, it was great being involved on your project.

This project was funded by a Research & Innovation Strategic Endowment Scholarship from the Galway-Mayo Institute of Technology.

Tissue samples were collected during a National Parks and Wildlife Service project (Ref: SPU-C017-2016); Thanks to Dr Heather Lally, Dr Conor Graham and Cóilín Minto for the expert assistance during these projects both round the table and in the field.

Thanks also to Dr Brian Nelson at National Parks and Wildlife Service for his support.

Additional funding was obtained from the Marine Institute's Networking and Travel Grant Scheme.

Express gratitude and acknowledgement for their commitment to this project go to my two supervisors Dr Luca Mirimin and Dr Orla Slattery. I have learned more than I ever could have imagined when we started two years ago, and I'm wholly grateful for all your help. Sincerest thank you, to you both.

1.2 Table of abbreviations

Abbreviation	Definition	Abbreviation	Definition
16S	16S rDNA	nuDNA	nuclear DNA
AFLP	Amplified fragment length polymorphisms	PAMPs	Pattern associated molecular patterns
AIM	Autoinduction Media	PO	Phenoloxidase
cDNA	Complementary DNA	PMSF	Phenylmethanesulfonylfluoride
COI	Cytochrome c oxidase subunit 1	ppA	prophenoloxidase activating enzyme
EBV	Essential biological variables	proPO	Prophenoloxidase
ICS	Indigenous crayfish species	RAPD	Random amplified polymorphic DNA
IPTG	Isopropyl β-D-1-thiogalactopyranoside	qPCR	Real-Time quantitative PCR
MBP	Maltose binding protein	RFLP	Restriction fragment length polymorphisms
mtDNA	mitochondrial DNA	RPM	Revolutions per minute
MCS	Multiple cloning site	SSRs	simple sequence repeats
NBDC	National Biodiversity Data Centre	SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
NICS	Non-indigenous crayfish species	SACs	Special Areas of Conservation

1.3 Framework of the study

This present study is part of broader research ongoing at the Marine and Freshwater Research Centre at the Galway-Mayo Institute of Technology to promote the study and management of the white-clawed crayfish (*A. pallipes*) in Ireland. The overall research areas include; distribution and abundance of *A. pallipes* in Ireland, hatchery rearing methodologies, acoustic monitoring, behavioural studies, molecular ecology (genetics and environmental DNA analyses). This thesis focused on the latter and has a two-fold focus; population genetic analysis of crayfish populations in Ireland (Chapter 2) and characterisation of immune proteins (Chapter 3).

1.4 Orientation of this thesis

This thesis is organised with a general abstract to summarise the overall project, a chapter giving a general introduction to crayfish biology, a chapter detailing the complete population genetic study (abstract -> discussion), a chapter detailing the complete protein biochemistry project (abstract -> discussion), and a concluding chapter including implications of this thesis and concluding remarks.

Abstract

The white-clawed crayfish (*Austropotamobius pallipes*, Lereboullet 1858) is a large freshwater invertebrate that has important functions in maintaining ecosystems. Historically, the species was broadly distributed across western Europe, but populations have been significantly reduced and fragmented due to habitat loss, competition with non-indigenous crayfish species (NICS) and outbreaks of crayfish plague caused by the aquatic mould *Aphanomyces astaci*. Ireland remains free of NICS, but five outbreaks of the plague have been recorded since 2015. Regardless of these recent outbreaks, crayfish populations in Ireland remain the healthiest across Europe. Conservation management plans for the species suggest Irish populations of *A. pallipes* are a suitable restocking source for European locations where the species has been lost but genetic analyses in the early 2000s indicate the stock is genetically homogeneous and lacks diversity. To access this, crayfish were samples from special areas of conservation in Ireland, a region of the 16S rDNA molecular marker was sequenced and genetic diversity was compared within and between populations. Statistically significant pairwise comparisons and molecular analysis of variance values indicate genetic diversity and population structure does exist within Irish white-clawed crayfish populations. Furthermore, to characterise host-pathogen interactions between *A. pallipes* and *A. astaci*, the overexpression of a Kazal type serine protease inhibitor (KP12) was optimised in an *E. coli* host. The fusion protein was purified by affinity chromatography, cleaved and isolated by gel filtration. Appropriate finial concentrations for inhibitory assays were not achieved. However, the optimised overexpression workflow has been detailed and will serve for the future overexpression of KP12 and potentially other Kazal protease inhibitors from the species.

2 Chapter 1:

General introduction

2.1 Protecting freshwater biodiversity

The rivers and lakes that make up freshwater ecosystems account for 0.01% of water on earth covering less than 1% of the earth's surface but support 10% of recorded species (Dudgeon et al. 2006, Strayer and Dudgeon 2010). However, biodiversity loss is occurring in freshwater environments at a substantially greater rate compared to terrestrial environments (Sala et al. 2000), happening at three levels; the ecosystem, the species and genetic levels. With the major threats to freshwater species being; habitat destruction, the introduction of invasive alien species, pollution, and overexploitation of resources (Allan and Flecker 1993, Dudgeon et al. 2006, Strayer and Dudgeon 2010). In an effort to mitigate these threats, Essential Biological Variables (EBV) must be measured within a given environment to report and develop appropriate conservation management strategies, EBV includes species richness, community composition, ecosystem structure and the genetic composition within an ecosystem (Pereira et al. 2013). However, Feld et al. (2009) reported that even well organised environmental projects often omit genetic diversity data, which is the least reported variable when studying ecosystems.

Nevertheless, genetic diversity is a fundamental measure of biodiversity, and genetic data are a vital variable in considerations of managing vulnerable ecosystems and species (Leberg 1990). A reduction of genetic diversity can result in reduced fitness, increase mortality rates, inhibition of population growth and significantly reduce immune responses to infection (Ferguson and Drahushchak 1990, Leberg 1990). Therefore, to appropriately manage vulnerable species, a thorough characterisation of genetic diversity and population structure should be

compiled to inform appropriate management planning for a species and its ecosystem.

2.2 Crayfish systematics

Crayfish (Astacidae) are large nocturnal decapod crustaceans found in a variety of freshwater environments and particularly thrive in calcium-rich rivers, streams and lakes with rocky substrate, vegetation, muddy banks and tree coverage (Matthews and Reynolds 1992). Crayfish are a diverse group with over 650 species ranging across all continents except Antarctica (Crandall and Buhay 2008, Richman et al. 2015). The group is divided into two superfamilies; the Astacoidae with two families, Cambaridae (400+ spp.) and Astacidae (40+ spp.) in the Northern Hemisphere; and the Parastacoidea, which contains a single family, Parastacidae (170+ spp.) in the Southern Hemisphere. The greatest diversity is found across North America with nearly 400 species, followed by Australia and South America with over 150 and 60 species, respectively. In stark contrast, only nine species occur in the African continent, four in East Asia, and five in Western Europe, but 40 in Eurasia as a whole (Crandall and Buhay 2008, Richman et al. 2015). The five Western European species fall into two genera, *Astacus* with three species and *Austropotamobius* with a diverse species complex (containing *A. pallipes,* and *A. italicus* and their subgroups) and *A. torrentium* (Kouba et al. 2014); that are distributed across Western Europe from the Iberian peninsula to Croatia (Figure 1). Naturalists have studied White-clawed crayfish (*A. pallipes*) for centuries (see cover page quote and illustration) , but the species' nomenclature is inconsistent over a wide-ranging period (Table 1). However, consensus on *A. pallipes* occurred around the mid-1970s. Around the same time, interest in white-clawed crayfish ecology began to rise in Ireland, particularly as potential stock to exploit for commercial purposes (Moriarty 1969, Moriarty 1971). Over the following two decades, Irish studies of the species focused primarily on species distribution and ecological functions (McFadden and Fairley 1984b, McFadden and Fairley 1984a, Lucey and

McGarrigle 1987, Holdich and Lowery 1988, Reynolds 1988). However, following the advent of molecular genetics in the 1990s studies emerged that aimed to resolve questions regarding *A. pallipes'* evolutionary origin in Ireland and indicated Irish populations share recent ancestry with western French populations (Grandjean et al. 1997, Souty-Grosset et al. 1997, Grandjean et al. 1998).

Table 1. The current scientific classification and historical nomenclature *of Austropotamobius pallipes*. The left columns show the current taxonomic classification of *A. pallipes* to species level and right columns show historical nomenclature given for the spec

Scientific classification		Historical classifications	Source
Phylum	Arthropoda	*Astacus fluviatilis*	Rutty (1772)
Subphylum	Crustacea		Templeton (1836)
Class	Malacostraca		Scharff (1895)
Order	Decapoda		Fairley (1972)
Family	Astacidae	*Cancer astacus*	Tighe (1802)
Genus	*Austropotamobius*	*Potamobius pallipes*	Scharff (1907)
Species	*A. pallipes*		Kennedy&Fitzmaurice (1971)
		Astacus pallipes	Frost (1948)
			Moriarty (1963)
			Moriarty (1971)
		Potamobius fluviatilis	Karaman (1963)

To date, the majority of *A. pallipes* genetic research has focused on the population structure and evolutionary history and genetic structure of continental European *A. pallipes* populations; in particular, French, Italian and Spanish populations are well characterised. These studies showed that the white-clawed crayfish is characterised by a complex genetic composition, which resulted in a plethora of subspecies being described (Table 2), and the genus being defined as a species complex. This *Austropotamobius* species complex has been well studied using genetic approaches (Table 2), though taxonomic uncertainties have not been fully resolved. Santucci et al. (1997) used allozyme data from restriction fragment length polymorphism (RFLP) analysis to identify three distinct groups within the genus, representing *A. pallipes, A. italicus* and *A. torrentium*. Later, Grandjean et al. (2002b)

used the mitochondrial molecular marker 16S and defined *A. pallipes* and *A. italicus* as two distinct species and further split *A. italicus* into three subspecies. A study by Chiesa et al. (2011) using cytochrome c oxidase subunit 1 (COI) sequencing identified four geographically structured groups among *A. pallipes*, but nuclear DNA analysis by amplified fragment length polymorphisms (AFLP) in the same study did not support the COI geographic structure and displayed only a moderate level of genetic structure. The authors here proposed a single species hypothesis of *A. pallipes,* incorporating *A. italicus* into an *A. pallipes* species complex (Chiesa et al. 2011). Later work by Jelić et al. (2016) contained samples across the species' entire distribution range, using the mitochondrial DNA (mtDNA) molecular markers 16S and COI in combination with several nuclear DNA (nuDNA) markers. Interestingly, the Jelić et al. (2016) study identified mito-nuclear discordance within white-clawed crayfish, that is, white-clawed crayfish show significant inherent variances in patterns of differentiation between mitochondrial and nuclear DNA. As a result, using mtDNA markers wholly in *A. pallipes* has limited abilities in accurately testing phylogenetic hypotheses (Toews and Brelsford 2012), which has been the case for most *A. pallipes* studies. The recent study by Jelić et al. (2016) gives support to Chiesa's 2011 single species hypothesis. However, a consensus has not yet become clear in the literature and at present, uncertainty regarding the overall structure of the *Austropotamobius spp.* complex remains.

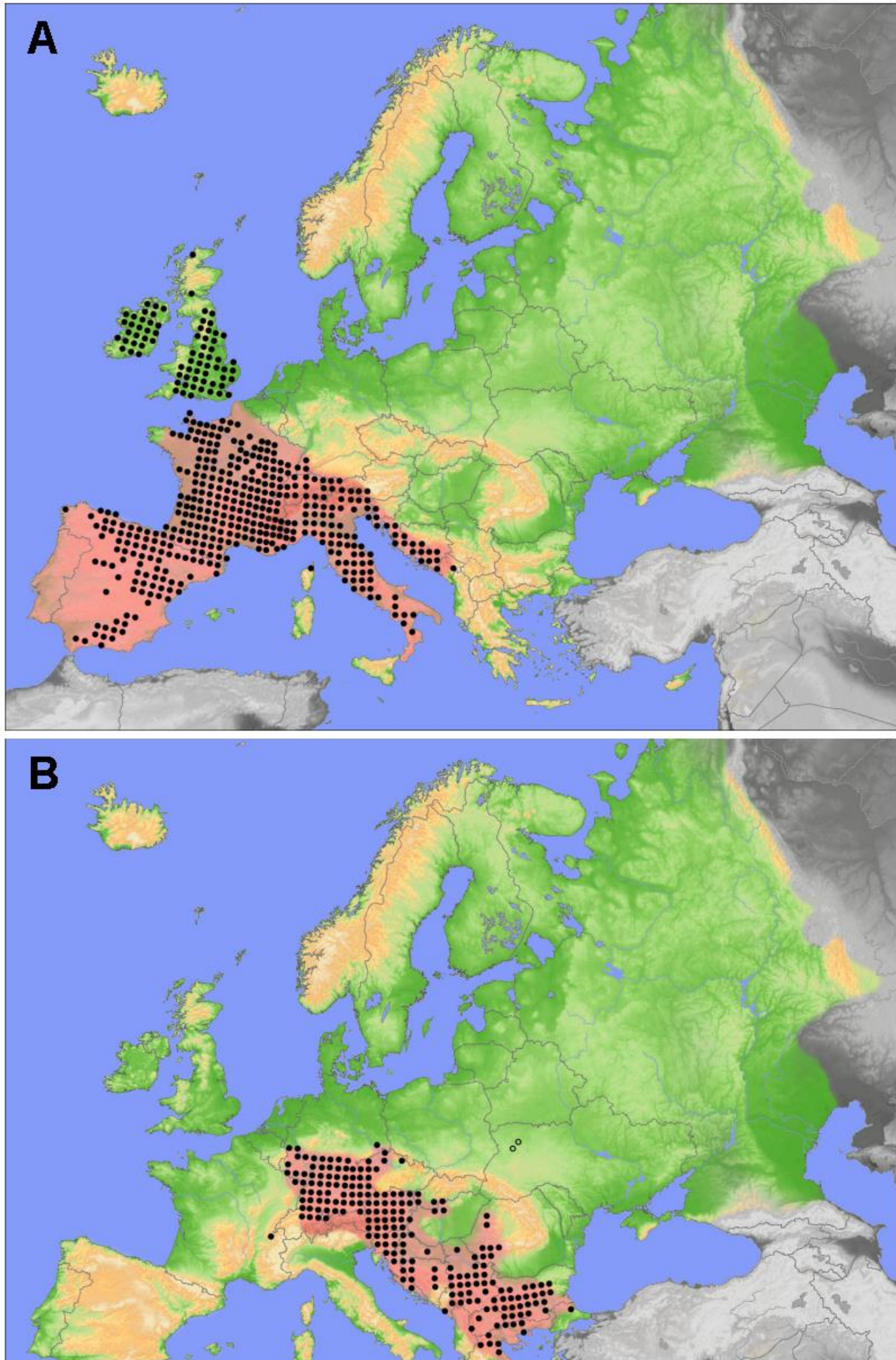


Figure 1. The distribution of *Austropotamobius spp*. across Europe. **1A** shows the distribution range of the *Austropotamobius* species complex of *A. pallipes* and *A. italicus*. **1B** shows the distribution of *A. torrentium*. Native ranges are highlighted in pink. Composite image from Kouba et al. (2014).

Table 2. Systematic hypotheses given for the reclassification of the *Austropotamobius* spp. complex using molecular data and their sources.

Species	Subspecies	Source	Publication
A. pallipes *A. italicus*		Allozyme	Santucci et al. (1997)
A. pallipes	*pallipes* *italicus* *lusitanicus*	mtDNA, RFLP	Grandjean et al. (1998)
A. pallipes *A. italicus* *A. berndhauseri*	*pallipes* *italicus* *carsicus* *carinthiacus*	16S rDNA	Grandjean et al. (2000)
A. pallipes *A. italicus*	*pallipes* *italicus* *carsicus* *carinthiacus*	16S rDNA	Grandjean et al. (2002)
A. pallipes *A. italicus*	*pallipes* *italicus* *carsicus* *carinthiacus* *meridionalis*	16S rDNA	Fratini et al. (2005)
A. pallipes	*pallipes*	COI mtDNA	Trontelj et al. (2005)
A. pallipes	*pallipes* *italicus*	COI, AFLP	Chiesa et al. (2011)

2.3 Biology of *Austropotamobius pallipes*

Like all arthropods, crayfish are invertebrates with an exoskeleton, paired jointed appendages and a segmented body. The hard exoskeleton is composed of chitin and mineralised calcium carbonate. This restricts growth and is periodically moulted to facilitate development (Avise 2017). Juveniles moult several times a

year, lessening to once and twice a year in adult females and males, respectively (Pratten 1980). The body consists of two primary parts: the cephalothorax and abdomen. The former contains the major organ systems beneath the hard carapace and bares four pairs of walking legs with large paired chelipeds (claws). Posterior to the cephalothorax is a six-segmented abdomen composed mostly of muscle tissue with the intestine and ventral abdominal artery spanning the whole length (Lőw et al. 2016). Each abdominal segment contains swimmerets that function in reproduction. The base of the abdomen has adapted into a tail fan consisting of four exterior plates (uropods) and an interior telson where the intestine leads to the anus. Their colouration is highly variable from light to dark brown, olive green, grey or black, with a pale white ventral surface on the cheliped, from which they get their common name. The species can live more than ten years and are relatively slow to reach sexual maturity (3 – 4 years). Breeding occurs in late Autumn (Oct. – Nov) when temperatures drop below 10°C (Matthews and Reynolds 1995), during which the male grasps the female and attaches spermatophores to the ventral surface of her abdomen (McFarlane et al. 2018). The female later manipulates the spermatophores to fertilise her eggs which gestate internally for up to six weeks. The eggs are then deposited *en masse* on the abdomen, attached with glair (sticky mucus) and are aerated by the abdominal swimmerets (Yazicioglu et al. 2016). The females retreat to burrows while brooding their egg mass (4 - 5 months) until fully formed juveniles hatch from their eggs in late Spring. The juveniles remain attached to the female for several days before detaching to fend for themselves.

White-clawed crayfish can be found in a wide range of freshwater bodies, shallow or deep rivers, lakes, streams and canals, and though previously referenced as a bioindicator of good water quality white-clawed crayfish are tolerant of poor-quality water conditions, organic pollutants and eutrophic environments (Füreder and Reynolds 2003). However, they do have a low tolerance to inorganic pollutants and such as pesticides and fertilisers (Füreder and Reynolds 2003).

Behaviourally, the species is primarily nocturnal and retreat to burrows in muddy banks, within macrophyte coverage or beneath rocky substrate to avoid predation or interactions with other crayfish. All asticids are voracious omnivores and *A. pallipes* is no different, feeding on invertebrates including other crayfish, macrophytes and all forms of detritus (Scalici and Gibertini 2007). Juveniles tend to feed more dominantly on arthropods and molluscs and shift to a predominant diet of macrophytes with maturity, though gammarids remain an important food source for all life stages (Reynolds and O'Keeffe 2005).

2.4 Major threats to *A. pallipes*

As with all freshwater species, habitat destruction and modification are a primary concern that can permanently displace native flora and fauna. The most pressing threat regarding white-clawed crayfish in Ireland is the potential introduction of non-indigenous crayfish species (NICS) and the subsequent introduction of the crayfish plague pathogen caused by the aquatic mould *Aphanomyces astaci*. After NICS were introduced from North America into Europe 150 years ago the adverse effects on native populations was immediate (Cornalia 1860). North American crayfish species are tolerant vectors of the plague pathogen, to which the indigenous crayfish species (ICS) of Europe have no or little innate defence. As a result, mass mortalities of ICS occur where infected NICS are introduced. The introduction of NICS and the crayfish plague has devasted populations of white-clawed crayfish across their entire range and has resulted in mostly fragmented and isolated populations across Europe and Britain (Füreder and Reynolds 2003). Ireland has remained free of NICS to date, though the introduction of these alien species remains a persistent threat as NICS are readily available to purchase online from private retailers and breeders within Ireland (Faulkes 2015, Faulkes 2017). Furthermore, studies have confirmed NICS sold in the pet trade can be infected with the plague pathogen (Patoka et al. 2014). The lack of NICS notwithstanding, five recent plague outbreaks have been confirmed in Ireland since 2015: Bruskey

River, River Suir, River Deel, River Barrow and River Lorrha, and are likely the result of contaminated fishing gear or watercrafts.

2.5 Conservation genetics

Biodiversity is measured at a genetic, organismal and ecosystem level to promote the preservation of diverse ecosystems and the natural resources that maintain them. Conservation genetics aims to inform appropriate management, maintenance and restoration of biodiversity by characterising the genetic variability of individuals and populations across ecosystems (Frankham 1995). Characterising the genetic structure of a population is a vital step when planning management strategies for a species and is measured by observing variations in genetic structure that are driven by several forces such as adaptation (more recent events) and speciation (over evolutionary time). Adaptation is a primary evolutionary process that species go through to survive changes in their local environment (Tiffin and Ross-Ibarra 2014); this process is ubiquitous and directly shapes a population's diversity and fitness (Hereford 2009). Adaptations may be structural (i.e. physical features), behavioural (e.g. migration instincts or learned social behaviours) or physiological (i.e. metabolic pathways and organ systems). True adaptations are genetically heritable, as opposed to plasticity in an organism's ability to respond to changes in its environment (Price et al. 2003). Adaptations are acted on by constant environmental pressures over space and time and are precursors to speciation; established populations that become isolated by reproductive barriers (e.g. distance or physical barriers) maintain genetic traits within groups of a given geographical vicinity (Seehausen and Wagner 2014). The modes of speciation can be thought of as a spectrum, from allopatric to sympatric speciation. Allopatry or geographic speciation is merely the isolation of a species into two or more populations that can no longer physically interbreed and can be the result of natural barriers e.g. continental drift or a natural dam forming (Harrison 2012). Sympatric speciation, in contrast to allopatry, is where a new

species arises from an ancestral population while still inhabiting some or all of the same geographic region (Fitzpatrick et al. 2008). Modes of speciation also occur along the spectrum between allopatry and sympatry. Periparty, for example, is a mode of speciation that gives rise to species in small populations that are geographically peripheral to the ancestral group (Lawson et al. 2015). Conservation genetics takes advantage of the genetic variations that arise due to adaptation and speciation events, such as genetic diversity, a measure of the of total genetic variation within a population.

The elucidation of genetic diversity and population structure within and between populations informs appropriate management planning to maintain and promote biodiversity and is a primary focus of conservation genetics (Queller and Goodnight 1989, Frankham 2003, Burton 2009). However, a number of different molecular markers can be used to answer a variety of biological questions, and the choice of marker in a population genetics study depends on the question posed (Wan et al. 2004). When studying population genetics in animal species, the sources of DNA used are either mitochondrial or nuclear. mtDNA is a small circular chromosome that is maternally inherited. The genes coded by the mitochondria are vital to cellular physiology and are therefore highly conserved among all animals. In *A. pallipes* the mitochondrial genome is composed of 13 protein-coding genes, 22 transfer RNAs, two ribosomal subunits and a noncoding control region with a function in transcription and replication (Grandjean et al. 2015). Sequences from mtDNA are such as 16S and COI are two of the most commonly used molecular markers in population genetics studies and are often combined or used in conjunction with additional markers (from nuclear DNA) to answer complex questions regarding phylogenetics and species dispersal (Wan et al. 2004). Other commonly employed markers are microsatellites found in nuclear DNA (nuDNA). These simple sequence repeats (SSRs) are characterised by motifs of 1 – 6 bases as repeating from tens to hundreds of times (Kim and Sappington 2013). Though initially thought to be randomly distributed among the genome, there is evidence

to support structure to their genomic position, with a potential effect on gene expression profiles of different cell types (Lawson and Zhang 2008). Regardless, microsatellites do occur in the nuDNA and often in non-coding regions, and these loci mutate readily and can be hypervariable within populations (Selkoe and Toonen 2006). This characteristic of microsatellites has made them a popular choice in population genetic studies and they are commonly used in agriculture and aquaculture to characterise and promote genetic diversity within breeding stocks and natural resources (MacHugh et al. 1998, Mirimin et al. 2016) and to monitor traceability of food sources (Dalvit et al. 2007).

2.6 Protein biochemistry

As some crayfish species are of commercial value in continental Europe a great deal of research has been completed regarding host-pathogen interactions between crayfish and *A. astaci*. Not surprisingly, the focus has mostly been mostly on those species of commercial interest such as signal crayfish (*Pacifastacus leniusculus*), a NICS introduced from North America for aquaculture. These studies identified many interesting proteins from *A. astaci* using biochemical techniques and methodologies, and several key virulence factors have been characterised specific to host-pathogen interactions between crayfish and the mould (Söderhäll and Unestam 1979, Häll and Söderhäll 1982, Aspán et al. 1990). These techniques required the sacrifice of tens to hundreds of crayfish to isolate single proteins. In the past few decades, recombinant DNA technology has enabled protein biochemists to clone genes of interest into expression systems (e.g. bacteria, and yeast) and overexpress large quantities of a *protein of interest* to characterise and assay (Somprasong et al. 2006, Donpudsa et al. 2010). However, no such work has been completed regarding *A. pallipes*, a crayfish highly susceptible to the mould pathogen, and so many questions regarding the species vulnerability to the pathogen remain.

3 Chapter 2: Genetic population structure and mitochondrial DNA variation of white-clawed crayfish *Austropotamobius pallipes* in Ireland.

3.1 Abstract

The white-clawed crayfish (*Austropotamobius pallipes*) is a large freshwater invertebrate with a high ecological value, but of little commercial significance in Ireland. The species is native to western Europe and naturalised in the British Isles, though native populations are fragmented due to competition with invasive species and the virulent pathogens. In Ireland, *A. pallipes* is well distributed throughout the country with the healthiest remaining populations across the species' range. However, current conservation management plans for the species assume that Irish populations lack genetic diversity and structure. To investigate such assumption, tissue samples were collected from three sites across Ireland and processed to obtain DNA sequence data of the mitochondrial molecular marker 16S rDNA. These data identified four haplotypes present within the tested Irish crayfish populations. Standard genetic diversity statistics were comparable to several native Spanish and Italian populations. Statistically significant fixation indices (F_{ST} = 0.171 - 0.23) were generated for all pairwise comparisons between Irish populations. Furthermore, AMOVA (Φ_{ST}) identified 30% of haplotype variation was distributed between populations, indicating population genetic structure. Comparison with four European populations identified a shared haplotype between Irish and French populations. These data indicate population genetic structure within Irish *A. pallipes* populations and support a recent common ancestry with French populations of *A. pallipes*.

3.2 Introduction

The white-clawed crayfish (*Austropotamobius pallipes,* Lereboullet 1858) has a widespread distribution across Europe ranging from the British Isles (Ireland and Great Britain) across the Iberian peninsula, western and central Europe to Montenegro (Holdich 2002). Though once abundant across this range, white-clawed crayfish is listed as endangered on the Red List of threatened species requiring special conservation measures under Annexes II and IV of the EU Habitats Directive (92/43/EEC and 94/62/EU)(Holdich 2003, Souty-Grosset et al. 2003, Diéguez-Uribeondo et al. 2008, Reynolds et al. 2010a, Jelić et al. 2016). In Ireland, *A. pallipes* is a naturalised species with two hypotheses regarding its introduction. Reynolds (1997) suggests crayfish were translocated from France around the 12th century, and Lucey (1999) hypothesised Irish populations originated from Britain around the 17th century. Since its introduction, white-clawed crayfish have colonised almost every region of Ireland, establishing itself as an ecosystem engineer. In fact, thanks to its role on habitat maintenance, predation of benthic fauna, and grazing of macrophytes, the white-clawed crayfish influences primary production and ecosystem structure (Matthews and Reynolds 1992, Matthews et al. 1993, Reynolds 1998). Furthermore, crayfish are a valuable food source for a wide range of fish, particularly salmonids and European eel (Moriarty 1963), birds and mammal species such as heron and otters (McFadden and Fairley 1984b, Holdich and Lowery 1988).

To monitor the species, a distribution map of white-clawed crayfish was assembled by Lucey and McGarrigel (1987) and was composed of over 300 records collected between 1976 and 1987. White-clawed crayfish were positively identified at 13.5% of regularly sampled sites (n = 1,800), and these data were compiled to form a baseline from which future changes in distribution could be made. These data indicated healthy populations were well distributed across the country and noted that the species particularly thrives in limestone-rich midlands. So much so that these populations were identified as a potential restocking source for other

European regions (Lucey and McGarrigle 1987, Reynolds 1988, Lucey 1999, Reynolds et al. 2002a, Andréanne and Reynolds 2003, Demers et al. 2005, Reynolds et al. 2010a, Reynolds et al. 2010b, Reynolds et al. 2013). An updated distribution map was compiled by Dimears et al. (2005), with data from 1990 to 2003 where crayfish were positively identified at 17.6% of regularly sampled sites (*n* = 3,600). These data indicate crayfish distribution has remained widespread across Ireland in the sampled timeframe and recommendations on managing and maintaining such populations have been developed by Reynolds (1988), Matthews and Reynolds (1992), Reynolds (1997), Demers and Reynolds (2002), Andránne and Reynolds (2003) Demers et al. (2005) and Reynolds (2009).

The abundance of crayfish in Ireland notwithstanding, white-clawed crayfish and indeed all indigenous crayfish species (ICS) in Europe face the most significant number of threats of crayfish globally (Crandall and Buhay 2008), mainly; habitat destruction, competition with Non-Indigenous Crayfish Species (NICS), and recurrent outbreaks of the crayfish plague caused by *Aphanomyces astaci* (Holdich and Lowery 1988, Holdich 2003, Richman et al. 2015). Crayfish plague outbreaks are of particular concern regarding crayfish conservation as 100% mortality rates can occur in susceptible populations in as little as seven to 21 days (Makkonen et al. 2012). Consequently, *A. pallipes* has declined between 50 – 80% across their European range (Füreder et al. 2010).

In Ireland, little change has been recorded in the species distribution since the 1970s due in large fact to the absence of NICS, hence reinforcing the potential ecological value of the Irish populations (Demers et al. 2005). However, in 2015 the first crayfish plague outbreak was confirmed in Ireland by genetic analysis occurring in the Bruskey River, this was followed by confirmed outbreaks in the River Suir, River Barrow, River Deel and the River Lorrha in 2017 (NBDC 2017). Moreover, these genetic analyses identified three separate strains of *A. astaci* between these sites, indicating the plague pathogen was introduced to Ireland on three separate occasions in the two-year period (NBDC 2017). While only recently occurring in

Ireland, the crayfish plague has had devastating effects on crayfish populations in Europe for over a century and coupled with habitat degradation and competition with NICS, in many regions white-clawed crayfish has been reduced to fragmented populations in headwaters and brooks (Füreder and Reynolds 2003). Such population fragmentation could result in small genetically homogeneous and isolated populations that, due to restricted gene flow could be at risk of inbreeding and extinction (Wright 1931, Nei et al. 1975, Shaffer 1981, Frankham 1999). In theory, appropriate and informed conservation management of endangered species should always incorporate factors such as population size and genetic diversity and structure (Rai, Boyce 1992, Neaves et al. 2015). However, these data can be time-consuming and expensive to collect, process and analyse.

Moreover, if conservation management planning for a vulnerable species take place after a crisis event occurs (e.g. plague outbreak), then best practice regarding the retention of genetic diversity may be overlooked to facilitate immediate action, such as restocking (Soulé 1980, Frankham 2003, Frankham et al. 2014). It is therefore sensible to obtain and incorporate these genetic data into management plans before further plague outbreaks occur. In Ireland, white-clawed crayfish are listed as a *qualifying interest* in 15 special areas of conservation (SACs) (Figure 2). SACs are sites of conservation importance at a European level, and conservation management plans were developed for each SAC including objectives specifically designed to monitor the qualifying interests within them. Specific habitats and species must be protected within SACs, such as; raised bogs, sand dunes and heaths, otter, salmon and white-clawed crayfish. SACs must be monitored periodically to record the distribution of qualifying interests within them.

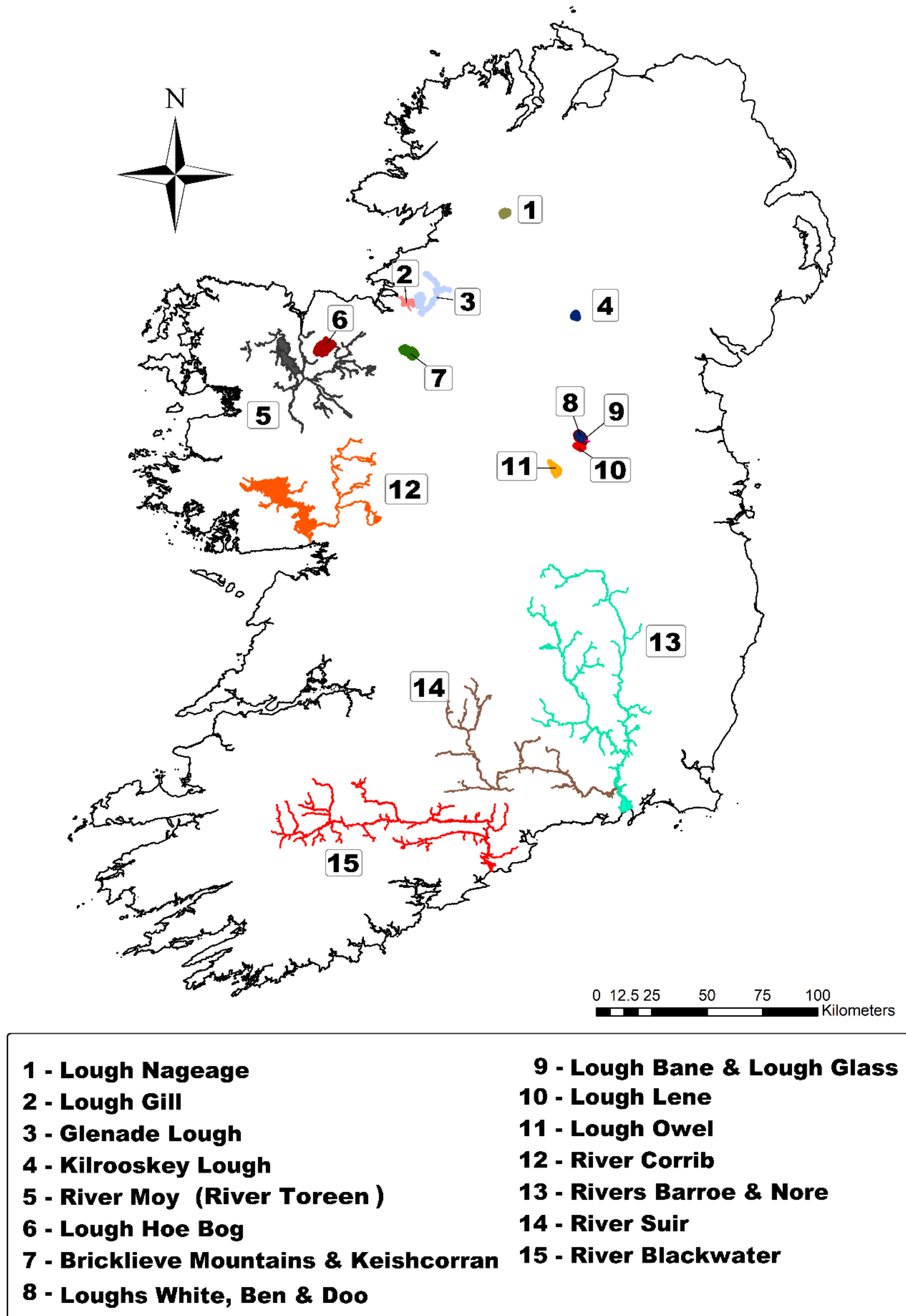


Figure 2. The 15 special areas of conservation in Ireland where *Austropotamobius pallipes* is listed as a qualifying interest (listed on Annex I and Annex II of the Eu habitats directive). N.B. – lakes and rivers illustrate their position but are not to scale.

Many genetic analyses have been undertaken on European and Irish crayfish populations (Souty-Grosset et al. 1997, Grandjean et al. 2002a, Diéguez-Uribeondo

et al. 2008). However, the studies including Irish populations focus mainly on the origin of *A. pallipes* into Ireland. The earliest analytical attempts (protein electrophoresis) to identify species by Albrecht & von Hagen (1981) failed to distinguish between related crayfish genera. Later work by Attard and Pasteaur (1984) identified loci to differentiate between *Astacus* sp. and *Austropotamobius* spp., and further research by Attard and Vianet (1985) identified a low level of heterozygosity between Irish and French populations. In the late 90's the advancement of molecular genetics shed more light on the origin of *A. pallipes*. Using restriction fragment length polymorphisms (RFLPs) Souty-Grosset et al. (1997) identified low levels of genetic variation between Welsh, English and French populations. Grandjean et al. (1997, 2000) confirmed high similarities between French, English and northern Italian crayfish populations, separating them as distinct conservation units from Spanish, Austrian and Slovenian populations. Gouin et al. (2001) also confirmed these findings utilising random amplified polymorphic DNA (RAPD). Later, using RFLP Gouin et al. (2003) showed a single mitochondrial DNA (mtDNA) haplotype within Irish crayfish populations, this haplotype was also identified but rare among French populations. Additionally, RAPD analysis identified a clinal reduction in genetic variability from the south to the north of Ireland (Gouin et al. 2003) and the authors suggest *A. pallipes* was translocated to southern Ireland from France and expanded northward, colonising the country.

Regarding European populations, a much broader range of molecular analyses have been conducted. The phylogeography and genetic structure of populations have been well defined in Austria (Baric et al. 2005) Italy (Zaccara et al. 2004, Baric et al. 2005, Fratini et al. 2005), Spain (Diéguez-Uribeondo et al. 2008, Pedraza-Lara et al. 2010, Matallanas et al. 2016) and France (Souty-Grosset et al. 1997, Gouin et al. 2002, Gouin et al. 2006). A recent study by Jelić et al. (2016) used species delamination methods on mtDNA (16S and COI) from across the species range and proposed a four primary species hypothesis. The study also compared nuDNA

regions, internal transcribed spacers (ITS), and found no statistically significant variation in ITS markers. The study further reconstructed ancestral ranges and divergence time estimates and hypothesise the recent common ancestor of mitochondrial groups originated from Dalmatia (Jelić et al. 2016). These questions, regarding phylogenetics and population structure, are distinct and should be approached using different molecular approaches. To resolve taxonomic uncertainties, molecular markers from mitochondrial DNA (mtDNA) are particularly useful as mtDNA is conserved among animals, is haploid and maternally inherited without recombination, mtDNA also lacks intergenic regions (Avise et al. 1987). Sequence conservation varies among the mtDNA molecule, the 12S ribosomal subunit is highly conserved and can be applied to determine phylogenetic relationships (Wan et al. 2004), in comparison the control region mutates much more rapidly and is used in population studies (Brown et al. 1986, Moritz et al. 1987). Noncoding nuclear regions of the genome mutate up to 1,000 times faster and are suitable for resolving recent changes in genetic diversity (Wan et al. 2004).

As Irish populations of the white-clawed crayfish are often referred to as a suitable restocking source (Reynolds 1988, Reynolds et al. 2002a, Demers et al. 2005), a systematic genetic analysis of the species structure and phylogeography should be completed. These data can provide a more in-depth characterisation of Irish crayfish populations that will enable appropriate management planning for this valuable resource before future outbreaks of the plague occur.

3.3 Aims of this study

This study aimed to increase the sample coverage area for genetic analysis of *A. pallipes* across Ireland and identify whether population genetic structure exists within Irish crayfish populations. To meet this aim, tissue samples were collected from across Ireland; the molecular marker 16S was PCR amplified and sequenced. These data were then analysed to identify whether any population genetic

structure is present in Irish crayfish populations. The data were also statistically compared with native populations of *A. pallipes* in Europe.

3.4 Materials and methods

3.4.1 Field sampling

Sampling was conducted under licences issued by National Parks and Wildlife Service, Department of Arts, Heritage, Regional, Rural and Gaeltacht Affairs (No. C73/2017; No.C99/2017). During the summer period of 2017 crayfish were sampled from three Special Areas of Conservation (SAC) across Ireland (Figure 3); Kilrooskey Lough SAC 54.192, -7.245; Lough Owel SAC 53.556, -7.367273 and River Toreen in the Toreen SAC 54.111, -9.324 (Figure 3). Crayfish were sampled by a standardised hand searching method searching under 100 refuges (rocks and debris) in 100m stretches, or by crayfish pots, four pots were evenly spread along a ten-meter rope and baited with mackerel and left overnight. Tissue samples were collected from captured individuals as V-notch clippings from the uropod and stored in 70% ethanol. Crayfish were immediately returned to the site of capture.

3.4.2 DNA extraction and purification

V- notch tissue samples were blotted dry, and a scalpel was used to separate the dorsal/ventral section of the uropod. The samples were then washed and blotted dry three times with distilled water. DNA was extracted and purified with the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol, except digesting in proteinase K overnight and eluting the DNA with 50µL of warmed (55ºC) distilled H_20. DNA extracts were stored at -20°C prior to further use.

3.4.3 PCR and sequencing

PCR amplifications were performed in 25 µL reactions with Promega GoTaq (1x GoTaq® Green Reaction Buffer, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 1.25 u GoTaq® DNA Polymerase, 1.0 µM forward and reverse primers, 20 – 100 ng DNA template). Universal 16S rDNA primers designed by Simon et al. (1994) are widely used in invertebrate studies as a molecular marker and have been used in previous *A. pallipes* studies (Jelić et al. 2016). Thus, this marker was chosen to enable comparison to previously published DNA sequence data (Table 1). An approximate 540 bp section of the mitochondrial 16S rDNA gene was amplified with the universal 16Sar (5′ CGCCTGTTTATCAAAAACAT 3′) and 16Sbr (5′CCGGTCTGAACTCAGATCACGT 3′) primers (Simon et al. 1994), under the following cycling conditions: initial denaturation at 95°C for 3 minutes, 35 cycles including 95°C for 30 seconds, 59°C for 30 seconds, 72°C for 30 second; and a final extension step at 72°C for 5 minutes. The resulting PCR products were visualised on 1.2% agarose gel composed of 1x TAE buffer and SYBR™ Safe DNA gel stain (ThermoFisher Scientific). PCR products were purified with a QIAquick PCR Purification Kit (QIAGEN), and Sanger sequencing was completed using one or both primers on an ABI 3730XL sequencer (GATC, Konstanz, Germany). Sequencing required considerable troubleshooting, which is reported in Appendix 1 section 6.1.

Table 3. Sampling location, sample sizes (*n*), coordinates, GenBank accession numbers and data sources.

Location	Waterbody	*n*	Coordinates	GenBank accession	Source
Ireland	Kilrooskey Lough	27	54.192N, -7.245E		This study
Ireland	Lough Owel	27	53.556N, -7.367E		This study
Ireland	River Toreen	29	54.111N, -9.324E		This study
Bosnia & Herzegovina	Žukovica River	3	43.544N, 17.226E	KX370155 - KX370157	Jelić et al. (2016)
Bosnia & Herzegovina	Vriošica River	3	43.238N, 17.486E	KX370158 - KX370160	Jelić et al. (2016)
Bosnia & Herzegovina	Tihaljina River	3	43.336N, 17.324E	KX370161 - KX370163	Jelić et al. (2016)
Bosnia & Herzegovina	Tribistovo	3	43.534N, 17.344E	KX370192 - KX370194	Jelić et al. (2016)
Bosnia & Herzegovina	Jelica	2	43.550N, 17.402E	KX370206 - KX370207	Jelić et al. (2016)
Bosnia & Herzegovina	Mrtvica River	2	43.951N, 17.226E	KX370212 - KX370213	Jelić et al. (2016)
Bosnia & Herzegovina	Bilila Mokrašnica	2	43.395N, 17.596E	KX370195 - KX370196	Jelić et al. (2016)
Bosnia & Herzegovina	Zorića Gaj	4	43.375N, 17.332E	KX370197 - KX370200	Jelić et al. (2016)
Bosnia & Herzegovina	Gaj	2	43.375N, 17.332E	KX370208 - KX370209	Jelić et al. (2016)
Austria	Plansee Lake	2	47.475N, 10.816E	AJ242700, AF237607	Grandjean et al. (2000)
Austria	Plansee Lake	1	47.475N, 10.816E	AJ242706	Largiadèr et al. (2000)
Austria	Plansee Lake	21	47.475N, 10.816E	KX370465 - KX370485	Jelić et al. (2016)
France	Val Renard	1	48.876N, -0.412E	AF237595	Grandjean et al. (2000)
France	Val Renard	7	48.876N, -0.412E	KX370244 - KX370250	Jelić et al. (2016)
France	Gatineau	8	46.578N, 0.339E	KX370409 - KX370416	Jelić et al. (2016)
France	Haute Saone regio	13	47.757N, 6.156E	KX370430 - KX370442	Jelić et al. (2016)
France	Puit d Enfer	9	46.449N, -0.207E	KX370486 - KX370494	Jelić et al. (2016)
France	Vonne	7	46.612N, -0.433E	KX370498 - KX370504	Jelić et al. (2016)
France	Crochatiere	11	46.287N, 0.676E	KX370365 - KX370375	Jelić et al. (2016)
France	Roya River	14	43.925N, 7.521E	KX370218 - KX370231	Jelić et al. (2016)
Italy	Pozzatoio	14	43.401N, 12.977E	KX370512 - KX370525	Jelić et al. (2016)
Italy	Amandole	11	43.428N, 13.003E	KX370526 - KX370537	Jelić et al. (2016)

3.4.4 Data analysis

The 16s rDNA sequences were viewed and aligned using Muscle in MEGA7 (Kumar et al. 2016), the sequences were trimmed, and polymorphisms were manually inspected. Diversity estimates were carried out for each of the three Irish sampling locations as well as for the combined Irish dataset. To compare Irish populations with native European populations, 16S rDNA sequences of 143 *A. pallipes* were

retrieved from biogeographic regions defined for the species by Jelić et al. (2016) from GenBank, including; Austria (n = 24), Bosnia and Herzegovina (n = 27), Italy (n = 26) and France (n = 56) (Table 1). Mean nucleotide composition was calculated in MEGA7. Number of haplotypes (h), parsimony informative sites, nucleotide diversity (π), haplotype diversity (Hd), number of polymorphic (segregating) sites (S), and the variance of haplotype diversity (with standard deviation) were calculated using Arlequin v.3.5.2 (Excoffier et al. 2005). To test for selective neutrality, Tajima's D (Tajima 1989) and Fu's Fs (Fu 1997) statistics were run on Arlequin with 20,0000 simulated samples to test statistical significance. To investigate population differentiation within and between sampled areas, the Φ_{ST} fixation index was calculated by analysis of molecular variance (AMOVA) using Arlequin, with 40,000 permutations at a significance level of 0.05 (Excoffier et al. 1992). The pairwise F_{ST} fixation index based on haplotype frequencies between locations was calculated by projected distance matrix with 20,000 permutations, with significance levels of = 0.05, 0.01 and 0.001. The F_{ST} fixation index ranges 0 to 1; zero indicates the two groups are freely interbreeding, and nearer one signifies the two groups do not interbreed, indicating population structure. The TN93+G mutation model was calculated in MEGA7 to have the best substitution parameters for the dataset using the Model Selection function, and maximum parsimony (MP) method with 1,000 bootstrap replications was used to generate a dendrogram representing the evolutionary relationship among haplotypes in MEGA7. A 16S rDNA sequence of a related species (*Austropotamobius torrentium*) was obtained from GenBank (JF293381) and included in the analysis as an outgroup. The MP tree was constructed using the Subtree-Pruning-Regrafting algorithm (Nei and Kumar 2000). All positions containing missing data were not taken into consideration. The Templeton, Crandall and Sing (TCS) method (1992) was used to calculate the genealogical relationships between haplotypes to generate a TCS haplotype network in PopART (Clement et al. 2002) and edited in Microsoft PowerPoint to include haplotype distribution between populations.

3.5 Results

3.5.1 Genetic diversity

Genetic diversity analyses were calculated using a 16S molecular marker from samples collected at Kilrooskey Lough (n = 27), Lough Owel (n = 27), and the River Toreen sites (n = 29). The dataset included a 398bp fragment from the 83 Irish individuals (Table 3). These data included six polymorphic sites and one parsimoniously informative site. The mean nucleotide composition was: C 9.5%, T 35.49%, A 34.60%, G 20.41%, with an A/T composition of 70.1% and C/G composition of 29.9%. Four haplotypes were identified among the Irish samples, one of which matched a previously published haplotype (Table 2). Three haplotypes were identified in the Kilrooskey Lough samples, two haplotypes were identified in the Lough Owel samples and one haplotype among the River Toreen samples (Table 1). Hap_01 was the most abundant haplotype observed in 63 individuals (75%) in the Kilrooskey Lough, Lough Owel and River Toreen samples, and Hap_01 was the only haplotype identified in the River Toreen samples. Hap_02 and Hap_03 were only identified in Kilrooskey Lough samples, and Hap_04 was only identified in the Lough Owel samples (Table 1).

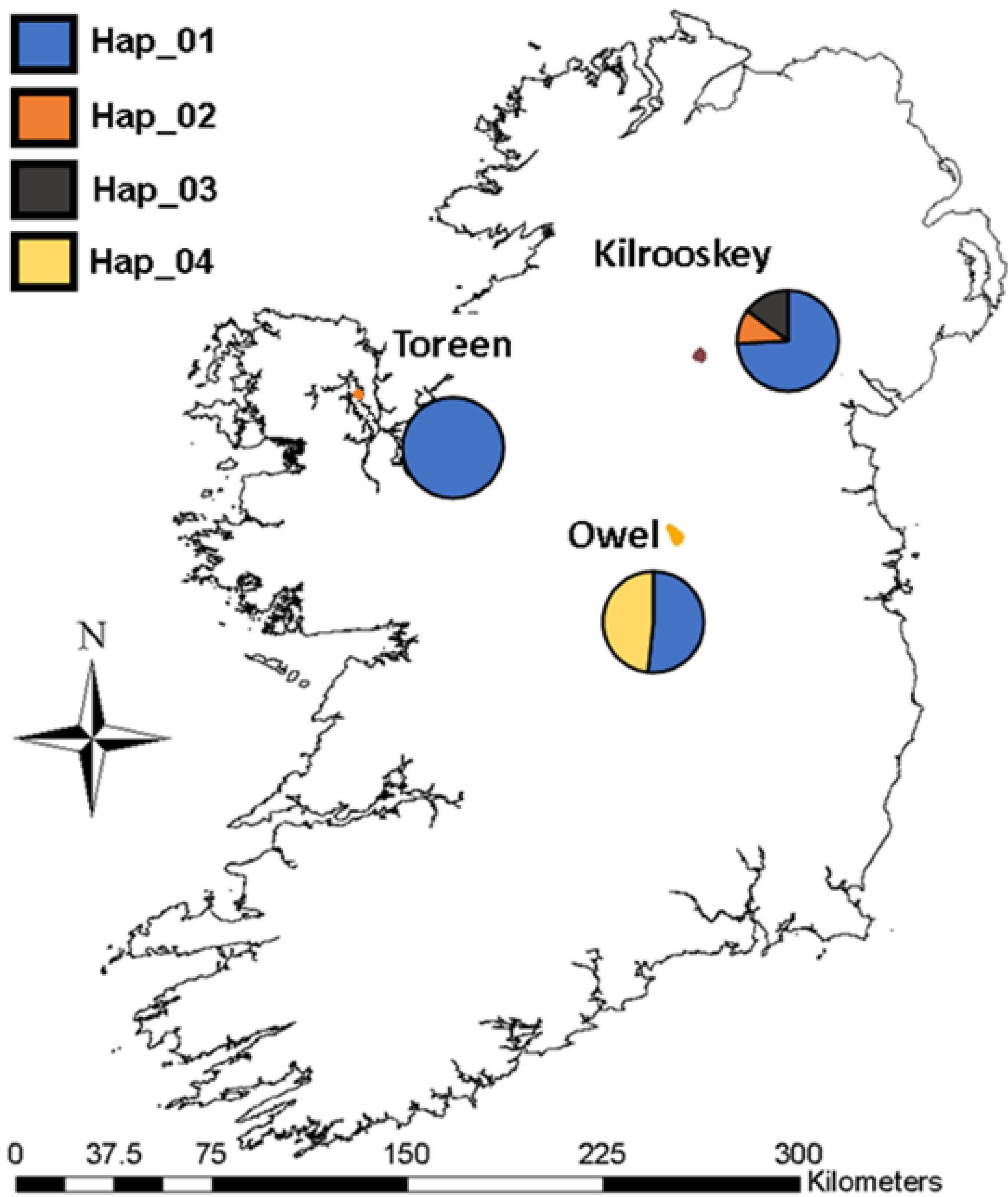


Figure 3. Three sampling sites with haplotype frequencies at each site. Kilrooskey Lough 54.192, -7.245; Lough Owel 53.556, -7.367273 and River Toreen 54.111, -9.324. At Kilrooskey Hap_01 = 74 %, Hap_02 = 11 % and Hap_03 = 15 %. At Lough Owel Hap_01 = 52 % and Hap_04 = 48 %. At River Toreen Hap_01 = 100 %.

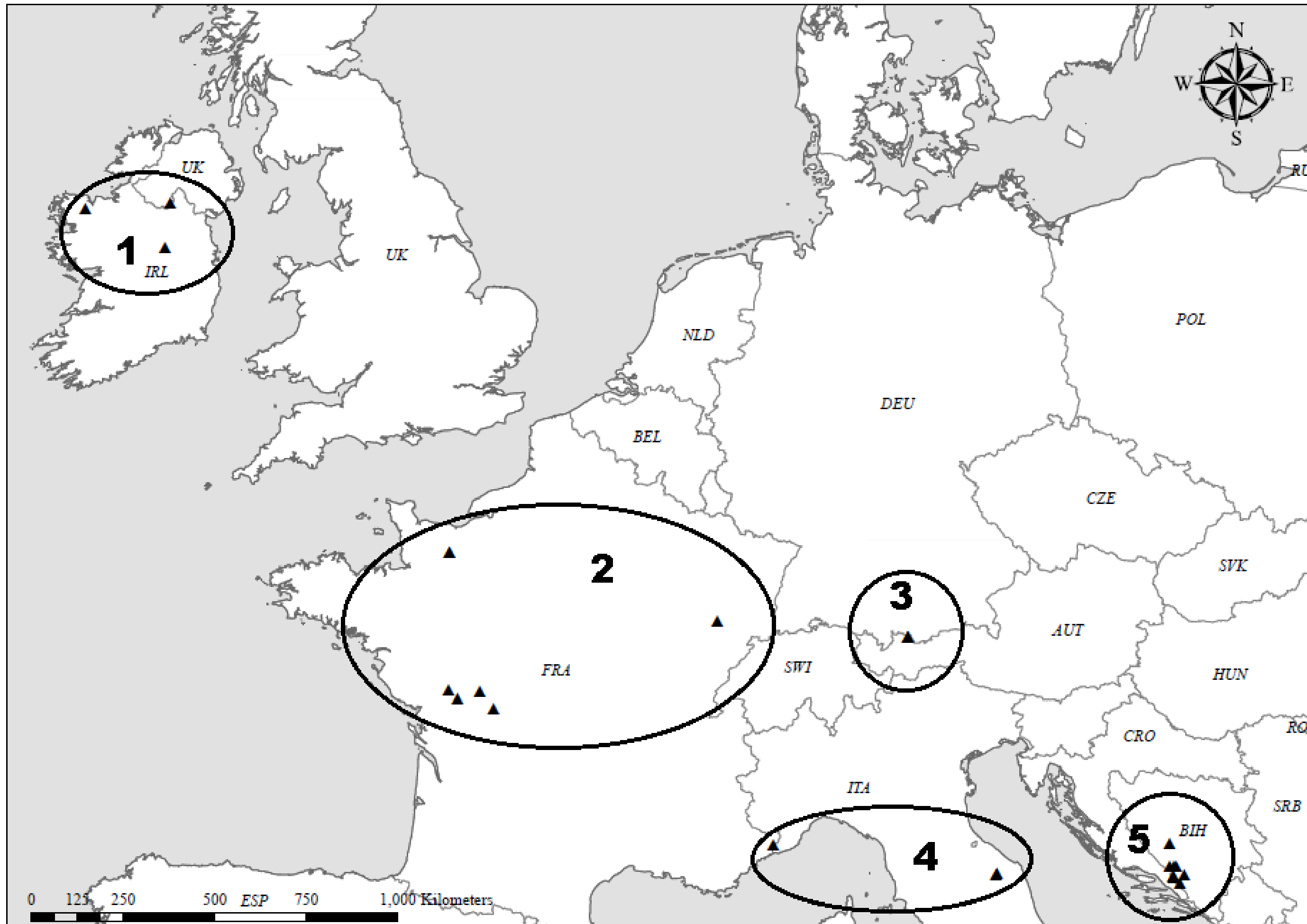


Figure 4. Locations of *Austropotamobius pallipes* samples collected in this study and obtained from GenBank. **1** Ireland, **2** France, **3** Austria, **4** Italy, **5** Bosnia and Herzegovina.

Table 4. Haplotypes frequencies and relative frequencies of the 16S rDNA mtDNA molecular marker of *Austropotamobius pallipes* from three Irish and four European locations.

Site locations	Ireland			Continental Europe			
Haplotype	**Kilrooskey**	**Owel**	**Toreen**	**Austria**	**Bosnia & Herzegovina**	**France**	**Italy**
Hap_01	20 (0.741)	14 (0.519)	29 (1)			32 (0.571)	
Hap_02	3 (0.111)						
Hap_03	4 (0.148)						
Hap_04		13 (0.481)					
Hap_05				23 (1)			
Hap_06					23 (0.852)		
Hap_07					2 (0.0741)		
Hap_08					2 (0.0741)		
Hap_09						10 (0.179)	
Hap_10						14 (0.250)	18 (0.692)
Hap_11							8 (0.308)

Haplotype diversity was greatest among the Lough Owel samples (Hd = 0.519) followed by Kilrooskey Lough (Hd = 0.433), though Kilrooskey Lough had a greater nucleotide diversity (π = 0.661) compared to Lough Owel (π = 0.519; Table 5). A lack of haplotype variation was found within the River Toreen samples, where Hap_01 dominated. The pooled Irish samples give similar values to those from Lough Owel and Kilrooskey sites (Table 5).

For the European samples, three haplotypes were identified in the French samples, of those, Hap_01 was the most abundant and is shared in each of the three Irish sites, of the remaining two haplotypes, Hap_09 is unique to its location and Hap_10 is shared with an Italian site (Table 4). Samples from Bosnia & Herzegovina contained three unique haplotypes, Italian samples contained one additional unique haplotype, and Austrian samples contained one unique haplotype. Between the continental European locations, the French samples showed the greatest haplotype diversity (Hd = 0.590), nucleotide diversity (π = 0.01229) and had the greatest mean number of nucleotide differences (k = 4.881; Table 5), the Bosnia & Herzegovina samples showed lower variation in haplotype diversity (Hd = 0.274),

nucleotide diversity (π = 0.00072) and mean number of nucleotide differences (k = 0.285). The Italian diversity statistics were intermediate (Hd = 0.443, π = 0.443, k = 3.102). No variation was observed within the Austrian samples.

Table 5. Genetic diversity statistics for 16S rDNA sequences of *A. pallipes* from three Irish population and four crayfish populations from Austria, Bosnia & Herzegovina, Italy and France.

Population	*n*	k	S	π	h	Hd	F	D
Kilrooskey	27	0.661	2	0.00169	3	0.433	0.559	0.57946
Owel	27	0.519	1	0.00131	2	0.519	1.648	1.61222
Toreen	29	-	-	-	1	-	-	-
All Ireland	83	0.517	3	0.00132	4	0.401	-0.428	-0.26141
Austria	23	-	-	-	1	-	-	-
Bos. & Herz.	27	0.285	2	0.00072	3	0.274	-1.05158	-0.95426
Italy	26	3.102	7	0.00781	2	0.443	8.30547	2.11445*
France	56	4.881	13	0.01229	3	0.590	12.07215	2.13151*

Number of sequences, n; mean no. of nucleotide differences, k; no. of polymorphic (segregating) sites, S; nucleotide diversity, π; number of haplotypes, h; haplotype diversity, Hd; Fu's F statistic, F; Tajima's D, D. Statistical significance for Tajima's D: *, $P < 0.05$; *.

Tajima's D values were positive at both sites but greater for the Lough Owel samples (D = 1.61222) compared to Kilrooskey Lough (D = 0.57946), when all three Irish sites were pooled the value became negative (D = -0.26141) (Table 5). Fu's Fs showed the same trend being greater in the Lough Owel samples (Fs = 1.648) compared to Kilrooskey Lough (Fs = 0.559), the pooled samples also gave a negative value (Fs = -0.428). Neither Tajima's D nor Fu's Fs statistics were statistically significant for the Irish locations, nor was the pooled Irish data ($P < 0.05$; Table 5).

3.5.2 Population genetic structure

The AMOVA based on haplotype frequencies showed that for Irish populations 30 % of variation falls among groups and 70 % falls within groups with a fixation index Φ_{ST} = 0.3044 (Table 6). The AMOVA conducted for the three Irish and four European locations found that 20 % of overall variation falls within groups (Table 6), and 80 % of variation occurring among the groups, with a Φ_{ST} = 0.079574.

Table 6. Analysis of molecular variance of the 16S molecular marker from three Irish and four European populations of *Austropotamobius pallipes*.

	Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Irish populations	Among populations	2	4.051	0.06767 Va	30.44
	Within populations	80	13.370	0.15463 Vb	69.56
	Total	82	16.422	0.22230	
	ΦST 0.3044				
EU populations	Among populations	6	650.839	3.57082	79.57
	Within populations	208	190.659	0.91663	20.43
	Total	214	841.498	4.48745	
	Φ_{ST} = 0.79574				

Analysis of molecular variance of the 16S molecular marker and haplotypes from individuals within three populations of *Austropotamobius pallipes* in Ireland, and between Irish and four European populations.

All pairwise comparisons were statistically significant between sites (Table 7). The pairwise comparison calculated between the Kilrooskey Lough/Lough Owel (F_{ST} =

0. 22751) was statistically significant at P < 0.001, the pairwise comparison between Kilrooskey Lough and River Toreen (F_{ST} = 0.17195) was statistically significant P < 0.01, respectively (Table 7). The Lough Owel/River Toreen pairwise comparison gave a similar F_{ST} value (0.22505) and was also statistically significant at P < 0.001. The F_{ST} comparison between Irish and European populations was higher in nearly all cases except for those between Irish and France and locations; the Kilrooskey Lough F_{ST} = 0.10588 and was statistically significant P < 0.01 = 0.00293 ± 0.0016; Lough Owel comparison F_{ST} = 0.2811, and was statistically significant at P < 0.001, and River Toreen comparison F_{ST} = 0.24794, and again was statistically significant at P < 0.001 (Table 7). All other pairwise comparisons between continental European sites were statistically significant at P < 0.001.

Table 7. Pairwise fixation indices (F_{ST}) for 16S rDNA mtDNA haplotypes for three Irish and four European *Austropotamobius pallipes* sampled locations.

Population	Kilrooskey	Owel	Toreen	Bosnia & Herzegovina	Austria	Italy	France
Kilrooskey	0	0.00098 ± 0.0010	0.00684 ± 0.0027				0.00293 ± 0.0016
Owel	**0.22751**	0					
Toreen	0.17195**	**0.22505**	0				
Bosnia & Herzegovina	**0.64672**	**0.60399**	**0.86769**	0			
Austria	**0.76927**	**0.72462**	**1.00000**	**0.85339**	0		
Italy	**0.56200**	**0.51879**	**0.78854**	**0.64282**	**0.76735**	0	
France	0.10588**	**0.20811**	**0.24794**	**0.53665**	**0.62353**	**0.36237**	0

Conventional $F_{ST'}$ fixation indices calculated from haplotype frequencies are included below the diagonal, F_{ST} P-values greater than 0.0 are included above the diagonal. Significant values P < 0.001 are in bold, P < 0.01 with **, and P < 0.05 are indicated with an *.

The MP tree showed that Hap_01 is shared between Irish and French populations of *A. pallipes* (*Figure 5*), Hap_01 is the dominant haplotype in all three Irish locations

and France. Hap_9 is also clustered with Hap_01 but is only identified in French samples. Of the remaining European haplotypes; Haplotypes 06, 07 and 08 were found within the Bosnia & Herzegovina samples and Hap_10 and Hap_11 were found within the Italian samples. However, Hap_10 was also found within the French samples. Hap_05 was the only Austrian haplotype identified and was equidistant from all other haplotypes.

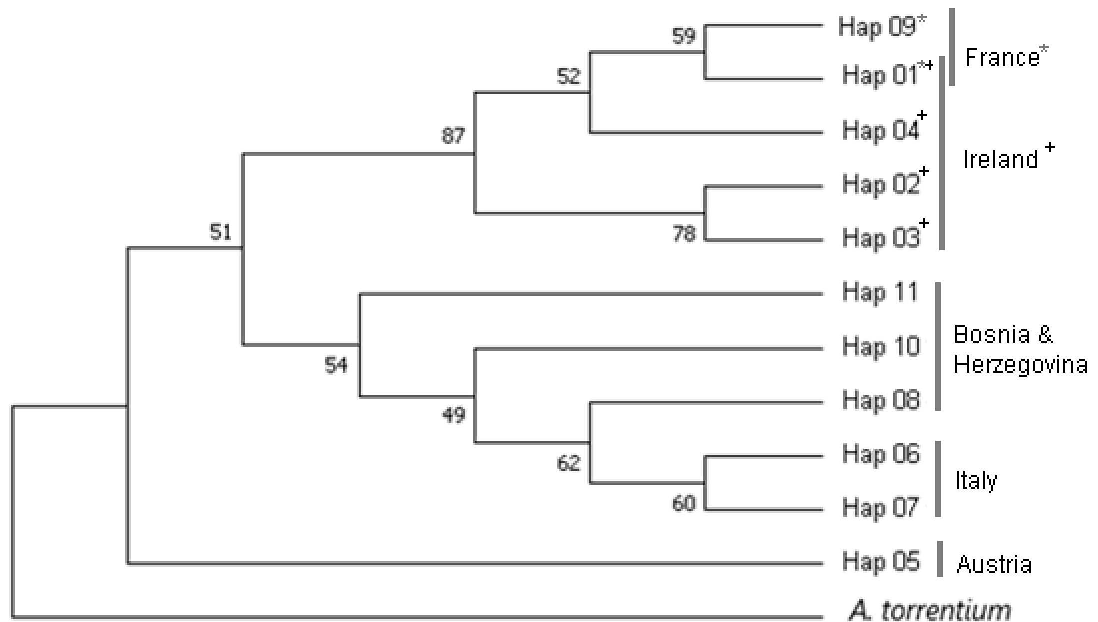


Figure 5. Maximum parsimony analysis of eleven 16S rDNA haplotypes from *Austropotamobius pallipes* found in three sampled Irish locations (Kilrooskey Lough, Lough Owel and River Toreen) and four continental European locations (Austria, Bosnia & Herzegovina, Italy and France) with *A. torrentium* as an outgroup. The bootstrap consensus tree was inferred from 1000 replicates to represent the evolutionary history of the haplotypes (Felsenstein 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed, analyses were conducted in MEGA7 (Kumar et al. 2016).

The TCS haplotype network (Figure 6) shows the most abundant haplotype (Hap_01) was identified in all Irish samples and French samples, being separated by a single mutation from Hap_02 and Hap_04 which were only found among Irish samples. Hap_03 found only in Irish samples was two mutations from Hap_01. Hap_09 found only in French samples was a single mutation from Hap_01 (Figure 6, Table 4). Hap_02 contained ten mutations from the next closest continental European haplotype, Hap_10, found in French samples. Hap_05 (Austria) and Hap_11 (Italy) were most distantly related with four and five mutations from the nearest inferred haplotype (Figure 6).

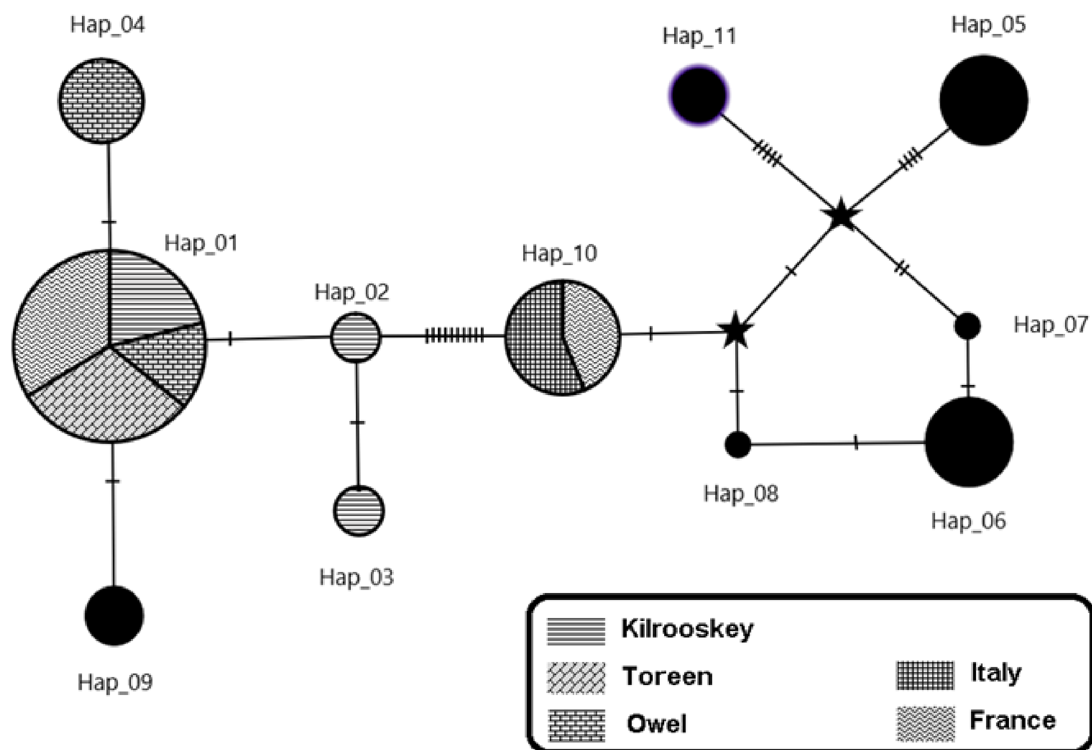


Figure 6. TCS haplotype network of 16S rDNA mtDNA for *Austropotamobius pallipes* from three Irish locations (Kilrooskey Lough, Lough Owel and River Toreen) and four continental locations (Austria, Bosnia & Herzegovina, Italy and France). Haplotype distribution within populations is indicated where shared haplotypes occur and can be referenced in table 4. Black circles indicate the haplotype was not shared between populations of *A. pallipes*. Hatch marks indicate single mutational changes between haplotypes and stars represent inferred or unsampled haplotypes.

3.6 Discussion

In this study, the 16S rDNA molecular marker was used to access genetic diversity and population structure among Irish crayfish *Austropotamobius pallipes* from three Special Areas of Conservation (SAC) in Ireland. These data indicate a prevalent haplotype in the sampled Irish populations of *A. pallipes* (Hap_01), having an occurrence of 100% of individuals in the River Toreen, 74% in Kilrooskey Lough and 52% in Lough Owel (Figure 6; Table 4). The same haplotype was also shared in 57% of French *A. pallipes* samples. Jelić et al. (2016) reports that within defined biogeographic regions across Europe; eight 16S haplotypes are present in Central and Western Europe, *n* = 128 (French samples), 18 haplotypes are present in Italy, *n* = 144 (Italian samples), 19 haplotypes are present in Dalmatia, *n* = 153 (Bosnia & Herzegovinian samples) and one haplotype in the upper Danube (Austrian samples) *n* = 24. The low number of haplotypes found in the Irish sites is comparable to those of Spain and France and substantially fewer compared to the populations within Italy and Bosnia & Herzegovina (Jelić et al. 2016). This may be a consequence of *A. pallipes'* recent introduction to Ireland. If founder populations were introduced into the southern parts of Ireland, as proposed by Gouin et al. (2003), and these populations had low genetic diversity then the dominant haplotypes in the founder groups would likely predominate today (i.e. a founder event). Moreover, mutations that may have occurred since their introduction would be unique to extant white-clawed crayfish populations in Ireland. Detailed reports of 16S Hd are limited in the literature. However, Hd values in Irish *A. pallipes* populations (Table 5) are comparable with those reported in Spain (Matallanas et al. (2016). Moreover, Irish populations have a similar Hd range compared to four of the ten Italian populations reported by Berniniet et al. (2016), though additional six Italian populations had greater Hd. However, this remains an interesting finding as Italian populations are some of the most genetically diverse across Europe (Zaccara et al. 2004, Baric et al. 2005, Fratini et al. 2005, Bernini et al. 2016).

The Irish populations were shown to not deviate from neutrality as no statistical significance was seen for Tajima's D or Fu's Fs statistics (Table 5). The relatively larger values for Lough Owel (D = 1.612 and Fs = 1.648) could be indicative of a small sample size. However, Tajimas's D and Fu's Fs are calculated using haplotype frequency data it is therefore likely the high values given at Lough Owel are caused by the low number of haplotypes (*n* = 2), with near equal haplotype frequencies (52 and 48 %). Moreover, when the data for all Irish sites were pooled these values were reduced to near zero. These results coupled with a lack of statistical significance imply that the genetic variation in haplotype frequencies within Irish crayfish population is derived from neutral mutations (Kimura 1991).

AMOVA was used to tests for patterns of subpopulations structure within and between groups, indicating Irish population are not genetically homogeneous. Over 70 %, of variation in the data observed within populations and 30 % among or between the three populations (statistically significant, P = < 0.001). When calculated for the European groups, AMOVA indicated 80% of variation falls between groups and 20 % among groups (statistically significant, P = < 0.001). In contrast, Largiadèr et al. (2000) analysed seven *A. pallipes* populations in Switzerland and northern Italy and found 75.5 % of genetic variation fell between groups, 16 % within populations, with the remaining variation found within subpopulations. Moreover, in the 25 Spanish populations studied by Matallanas et al. (2016), 85 % of variation fell between and 15 % within populations. These latter two studies on native European populations show considerably greater genetic structure within their native crayfish populations compared to the Irish populations, which is not unexpected given the species' recent introduction to Ireland. However, 30% variation between Irish populations is significant and indicates genetic structure within Irish crayfish populations.

The genetic structure between populations was carried out using the F_{ST} fixation index. The F_{ST} indices were unsurprisingly highest for the comparisons between European and Irish populations, and greatest F_{ST} was between the Toreen/Austria

populations. This is not-unexpected as each group contain one unique haplotype. The French/ Irish comparisons had the lowest F_{ST} indices in this regard and were similar to the pairwise comparisons values given for the Irish populations (Table 7). The data generated in this study do identify statistically significant structure within Irish crayfish populations, and furthermore support a recent divergence from an ancestorial population with the shared haplotype Hap_01 (Ireland and France). The implementation of an additional molecular marker would significantly reduce uncertainty when interpreting these results. A study of *A. pallipes* by Grandjean (2002a) using only 16S reported very low diversity in crayfish populations from the Iberian peninsula. However, later work by Pedraza-Lara (2010) utilising a number of markers on the same populations showed that there was indeed greater diversity within these same populations. It would, therefore, be appropriate to incorporate further molecular markers from a range of locations for Irish populations.

The Maximum Parsimony analysis provided further support to an association between Irish and French crayfish clustering the haplotypes distributed between these populations (Figure 5), which is reflective of the F_{ST} indices. This cluster generates a distinct clade from the remaining European haplotypes. The MP analysis identified the Austrian haplotype to be unique in the dataset. The generated TCS network sorted the sample groups similarly, with an Irish and French cluster separated from the rest of the European groups (Figure 6).

The data generated in this present study using a 16S molecular marker indicates a population genetic structure between *A. pallipes* populations in Ireland and the 16S marker has higher resolution compared to the allozyme data generated in earlier molecular genetics analyses of Irish crayfish populations (Crandall and Fitzpatrick Jr 1996). These past studies all indicate Irish populations shared one single haplotype, lack genetic structure and shared recent ancestry with French *A. pallipes* populations (Grandjean et al. 1997, Souty-Grosset et al. 1997, Grandjean and Souty-Grosset 2000, Gouin et al. 2001, Gouin et al. 2003). However, this present study provides a dataset that does identify population genetic structure between

Irish crayfish populations but adds further support to earlier findings that highlight the interesting association between Irish and French *A. pallipes* populations. This relationship should certainly be studied in greater detail, and with the addition of further molecular markers from mtDNA and nuDNA, a more detailed characterisation of *A. pallipes* population genetic structure for Irish populations can be achieved. Likewise, a higher resolution overview of Irish *A. pallipes* ancestry could also be provided utilising these data.

3.6.1 Implications for management

Appropriate management planning for vulnerable species should incorporate population genetic data to maintain and promote genetic population structure and diversity. However, these data are often overlooked. In Ireland, past characterisations of crayfish populations have not identified any variability in the genetic structure of crayfish populations, but by implementing a higher resolution molecular marker, this study shows that there is indeed genetic structuring between populations. This population structure should be considered when designing and applying management strategies, such as restocking.

Conclusions

- Population genetic structure has been identified in Irish crayfish populations using the molecular marker 16S.
- The association between Irish and French populations is further supported.
- Appropriate conservation management plans for *A. pallipes* in Ireland should be reviewed to reflect the emerging population genetic structure, particularly in special areas of conservation.

4 Chapter 3: Cloning, overexpression and purification of a recombinant serpin KP12 from *Austropotamobius pallipes*

4.1 Abstract

Serpins are protease inhibitors found in plants, animals and microorganisms that have a function to regulate proteolytic enzymes and are an essential component of the innate immune response of crayfish. The Kazal domain containing serpin KP12 was amplified from white-clawed crayfish (*Austropotamobius pallipes*) cDNA, cloned into a pMAL-p4x expression vector and the serpin was overexpressed as a maltose binding protein-KP12 fusion. Overexpression was optimised in autoinduction media within an *Escherichia coli* BL21 (DE3) host. Expression of the MPB-KP12 fusion protein was confirmed by SDS-PAGE and western blot. The fusion protein was recovered using cell lysis buffer, and sonication and purified by affinity chromatography using an Amylose Resin column. The protein was concentrated using centrifugation spin columns, and cleavage of the MBP from the KP12 was optimised with Factor Xa. The cleaved proteins were then separated by gel filtration (Superdex 200 column). Initial attempts to concentrate the purified KP12 resulted in significant loss of the serpin, this step requires optimisation however the work presented here outlines a successful method for the overexpression of this serpin in *E. coli*.

4.2 Introduction

4.2.1 Enzymes

Enzymes are macromolecules that act as catalysts to speed up biological reactions and function by converting substrates (the molecule an enzyme acts upon) into products (the molecules produced after the enzymatic reaction). Enzymes are highly specific molecules that are neither consumed during a reaction nor become the product of the reaction. Enzymatic catalysis can increase biochemical reaction times by orders of magnitude (Radzicka and Wolfenden 1995), with a substrate's half-life changing from millions of years to a fraction of a second (Wolfenden and Snider 2001). Enzymes are (mostly) proteins, and as proteins, their function is dependent on their structure (Kraut 1977), which is arranged at four levels. The primary structure is simply a linear sequence of amino acids linked by peptide bonds. Secondary structure refers to how the linear sequence of amino acids fold upon itself and is determined by backbone interactions that are maintained mostly by hydrogen bonds. Tertiary structure relates to a higher order of folding stabilised by hydrogen bonds, van der Walls interactions, hydrophobic packing and the formation of disulphide bridges between peptide chains. The quaternary structure is where two or more peptide chains are bound to form a single operational unit. To function correctly, an enzyme must have proper confirmation of folding at all four levels of structure. Furthermore, to catalyse a biochemical reaction an enzyme must bind to its specific substrate and this takes place in the enzyme's active site, a small portion (10 – 20%) of the overall structure (Alberts et al. 2013).

Enzymes are grouped into six classes that are defined by the specific mechanism they employ to catalyse a biochemical reaction. These six classes include; *Oxidoreductases* that catalyse the oxidation or reduction of a substrate by electron transfer. *Transferases* that catalyse functional group transfer from a donor to an acceptor molecule. *Hydrolases* that catalyse the hydrolysis (addition of H_2O) to a substrate. *Lyases* that catalyse the breaking or formation of double bonds in

molecules by the addition or removal of functional groups. *Isomerases* catalyse the rearrangement of functional groups within an existing molecule (producing isoforms). *Ligases* catalyse the formation of bonds to join (ligate) two molecules. Each class of enzyme contains several sub-classes (Webb 1992, Barrett 1994).

4.2.2 Proteases and protease inhibitors

4.2.2.1 Proteases

Proteases, also called *peptide hydrolases* or *peptidases* are enzymes found in all organisms and have diverse roles in a wide range of physiological functions, including digestion (trypsin), apoptosis (tryptase), inflammation (chymase) and immune responses. Proteases are hydrolases and function by cleaving proteins and polypeptides, cutting and replacing peptide bonds between amino acids with H_2O molecules (hydrolysis).

Proteases are classified by three criteria; 1) the reaction the enzyme catalyses, 2) the chemical composition of the catalytic triad found at the enzyme's active site, and 3) the conserved evolutionary homology of the molecule sequence (Barrett 1994). Proteases are first divided into two subclasses; endopeptidases that cleave site-specifically within the target molecule, and exopeptidases that cleave at either terminus of the polypeptide. Proteases are then grouped into families based on the catalytic moiety at their active site and include; serine-, cysteine-, aspartic-, metallo-, glutamic-, threonine- and asparagine-proteases (Oda 2011). These families include over 60 evolutionary lineages that show high levels of convergent evolution with many performing the same biochemical reactions but by different catalytic mechanisms (Rawlings and Barrett 1993). However, most proteases are then grouped into six clans based on their phenotype (Rawlings and Barrett 1994).

Abnormal proteolytic activity can result in disease states (e.g. self-digestion in pancreatitis, and inflammation), and several biological mechanisms have been

adapted to reduce non-specific proteolytic activity. Firstly, proteases are expressed as inactive zymogens to prohibit direct proteolytic activity upon expression. Protease inhibitors act to regulate proteolytic activity and are divided into two general classes: *active site inhibitors* that function but bind to the active site of a protease inhibiting further proteolytic activity, and α2-macroglobulins that act as opsonins by binding to a target molecule and marking the subsequent complex for degradation.

4.2.2.2 Serine proteases and serine protease inhibitors

Nearly a third of all proteases are serine proteases that are divided into over 20 families (Rawlings and Barrett 1993). These enzymes contain an *Asp-His-Ser* catalytic triad, that is a base histidine residue, an electrophile aspartate residue, and a nucleophile serine residue, the latter giving the group their name (Barrett 1994, Rawlings and Barrett 1994, Hedstrom 2002). Subclasses of serine proteases are distinguished by non-homologous secondary and tertiary motifs. However, the subclasses maintain an identical catalytic triad composition (Perona and Craik 1995, Dodson and Wlodawer 1998). Chymotrypsin, for example, is that largest clan of serine proteases, containing ten families of predominantly endopeptidases. Subtilisin is the second largest family, and they are also endopeptidases (Rawlings et al. 2004). The two families share the *Asp-His-Ser* catalytic triad, though the orientations of the sequences are different between the families (*His-Asp-Ser* in chymotrypsin*)*. Furthermore, their 3D structure is non-homologous (Rawlings and Barrett 1994), indicating the two groups are not evolutionarily related.

<u>Ser</u>ine <u>p</u>rotease <u>in</u>hibitors (serpins) are a well-studied large class of protease inhibitors that function to inhibit the proteolytic activity of proteases and are ubiquitous in plants, animals and microorganisms (Laskowski Jr and Kato 1980, Ryan 1990). Protease inhibitors function by forming very stable complexes with

proteolytic enzymes and are generally small molecules (5 – 80 kDa) with numerous disulphide bridges (Laskowski Jr and Kato 1980, Singh and Rao 2002).

Early work to identify and characterise serpins was completed by biochemical methods in a range of invertebrates, and the number of serpin genes within a given species varies widely including between closely related taxa. Some of the first serpins confirmed in invertebrates were from the silkworm (*Bombyx mori*) which now has 34 identified serpins (Zou et al. 2009); the red flour beetle (*Tribolium castaneum)* has 31 sequenced serpins (Zou et al. 2007), and the fruit fly (*Drosophila melanogaster*) has 29 recorded serpins (Garrett et al. 2009). Other invertebrates have considerably fewer, with the honeybee (*Apis mellifera*) and tsetse fly (*Glossina morsitaans*) containing seven and ten serpins, respectively (Evans et al. 2006, Mwangi et al. 2011). Other groups vary widely, for example, in clade Acari (mites and ticks) ten serpins are reported in the scabies mite (*Sarcoptes scabiei*) (Rider et al. 2015), while the blacklegged tick (*Ixodes scapularis*) contains 45 (Mulenga et al. 2009). In the freshwater crayfish, *Pacifastacus leniusculus* eleven serpins are reported (Wu 2011) . Serine proteinase inhibitors are grouped into several families including Kazal type, Kunitz, Bowman-Birk and pacifastins (Laskowski Jr and Kato 1980). The Kazal domain containing inhibitors contain three intra-domain disulphide bonds formed by six conserved cysteine residues (Bode and Huber 1992) and are believed to be involved in innate immune responses (Polanowski and Wilusz, Kanost 1999). Early work in crayfish by Johansson et al. (1994) cloned a Kazal type serpin of 209 amino acids with a molecular mass of 22.7 kDa that contains four domains, but though the serpin was shown to inhibit chymotrypsin and subtilisin, its biological function is unknown. Li et al. (2009) report a three domain Kazal-type serine proteinase inhibitor (hcPcSPI1) in the haemocytes, heart, gill and intestine of the crayfish *Procambarus clarkii*. The domains and proteins of hcPcSPI1 were overexpressed, and each was shown to inhibit subtilisin A and proteinase K. Furthermore, hcPcSPI1 was shown by Real-Time quantitative PCR (qPCR) to be upregulated in

the presence of the bacteria *Escherichia coli*, indicating the Kazal type inhibitor has some immune function (Li et al. 2009).

4.2.2.3 The invertebrate immune response

All invertebrates lack an adaptive immune response and rely on innate immune systems to protect against pathogenic infections (Söderhäll 2011). These responses are grouped into *cellular* responses (phagocytosis by haemocyte entrapment) and *humoral* responses. Humoral responses include Toll pathway activation that results in the production of antimicrobial peptides, and activation of proteolytic cascades (e.g. the prophenoloxidase activating system). Cellular and humoral responses work in conjunction to defend against invading pathogens as cellular responses aid in the initiation of humoral responses that in turn affect haemocyte activation (Jiravanichpaisal et al. 2006, Cerenius et al. 2008). In crustaceans, a critical humoral immune response is the prophenoloxidase (proPO)-activating system, a proteolytic cascade that neutralises pathogens upon recognition of pattern associated molecular patterns (PAMPs) by the host's pattern recognition receptors (Cerenius and Söderhäll 2004). PAMPs such as lipopolysaccharides, B-1, 3 glucan or extracellular glycoproteins from pathogens initiate the encapsulation of the pathogen in melanin (Söderhäll and Unestam 1979, Cerenius et al. 2008, Söderhäll 2011). Moreover, serpins generate complexes with PAMPs to direct melanisation to the site of infection, curtailing uncontrolled and nonspecific melanisation (Cerenius and Söderhäll 2004).

The melanisation process is initiated by phenoloxidase (PO), a serine proteinase that catalyses the oxidation of phenols to quinones which in turn polymerise into melanin (Söderhäll and Cerenius 1998). Though melanisation is a fast and effective immune response to infection, the products of the proPO activating system (phenols and quinones) are cytotoxic to the host, and their production must be highly regulated to avoid damage to the host (Wang et al. 2001). In

healthy crayfish, prophenoloxidase is expressed and maintained as an inactive zymogen, which in the presence of PAMPs, initiates a serine protease termed *prophenoloxidase activating enzyme* (ppA) that activates the proPO zymogen into active PO, which in turn, initiates the melanisation process (Cerenius et al. 2010).

Many of the proteins involved in the prophenoloxidase-activating system have been studied and characterised. Aspán and Söderhäll (1991) identified a 36 kDa serine proteinase purified from *P. leniusculus* haemocytes, showing the enzyme converts zymogenic proPO into an active 62 kDa phenoloxidase by cleaving chromogenic peptide substrates (ppA). Regulatory enzymes that inhibit nonspecific activation of PO have also been identified and characterised. Two protease inhibitors were isolated from *Pacifastacus leniusculus* plasma by Hergenhahn et al. (1987, 1988): a 155 kDa trypsin inhibitor (pacifastin) and a 190 kDa α_2-macroglobulin dimer. Later studies by Aspán et al. (1990, 1991) showed the trypsin inhibitor to have high inhibition towards ppA and the α_2-macroglobulin to have low inhibition towards ppA. A third subtilisin inhibitor was also identified. However, this did not affect ppA. Häll and Söderhäll (1982) isolated a proteinase inhibitor from haemocytes of *Astacus astacus* by differential centrifugation with a molecular mass of 23 kDa. The molecule was found to inhibit subtilisin and a commercial protease (Pronase) that contained several serine proteases, but neither trypsin nor chymotrypsin were inhibited. Häll and Söderhäll (1982) assayed the 23 kDa enzymes inhibitory activity against three *Aphanomyces spp.;* the enzyme only inhibited a protease from *Aphanomyces astaci*, an aquatic mould that infects freshwater crayfish. Li et al. (2009) identified three Kazal proteinases from *Procambarus clarkii* (hcPcSPI2, hpPcSPI3, and hpPcSPI4) and these inhibitors were all upregulated after challenge with *Vibrio angullarum,* though the inhibitory activity was only assayed for hcPcSPI2, which was shown to inhibit subtilisin and trypsin weakly. Donpudsa et al. (2010a) identified two further serine protease inhibitors (KP12 and KP18) from the crayfish *P. leniusculus*, showing that KP12 has strong inhibition against subtilisin but weak inhibition

against trypsin, whereas KP18 only inhibited subtilisin. Furthermore, KP12 was shown to inhibit extracellular proteinases from *A. astaci* (Donpudsa et al. 2010a).

These past studies have characterised several proteases and serpins in the general astacid (crayfish) proPO immune response, and interestingly, the *P. lenisculus* KP12 serpin identified and characterised by Donpudsa et al. (2010a) was shown to inhibit proteolysis of chromogenic peptide substrates caused by the pathogen *A. astaci's* extracellular proteinases. Therefore, it was decided for this study to focus on characterisation of a homologous KP12 serpin from *A. pallipes* to further investigating more specific interactions between *A. pallipes* serpins and *A. astaci* serine proteases.

4.2.3 Recombinant protein expression and serpin overexpression

Protein biochemistry requires milligram quantities of purified and functional protein to complete functional and characterisation assays. To this end, it is often infeasible for ethical and logistical reasons to purify native proteins from their host organism. This is particularly important when working with endangered or rare species as the sacrifice or extraction of protein may not be an option from a number of individuals; for example, early studies on the characterisation of single proteins in the crayfish *P. leniusculus* required the sacrifice of 170 specimens to isolate and purify a target protein (Johansson et al. 1994). Furthermore, if the protein of interest has a low expression level then a large quantity of tissue will be required; additionally, if the expression of the protein must be induced, such as with some immune genes, then ethical implications regarding animal welfare arise.

Recombinant protein expression overcomes these issues by cloning a gene of interest into a plasmid vector that can be cheaply and quickly induced and overexpressed in a host organism *in vivo*. The general workflow for recombinant protein overexpression is straightforward (Figure 7). However, in practice many

factors must be taken into account such as the choice of expression vector, choice of host (most commonly an *E. coli* strain), whether the protein of interest affects host growth, ability of the host to express sufficient quantities of correctly folded, active target protein (Bird et al. 2004, Pearce and Cabrita 2011, Rosano and Ceccarelli 2014). There are many advantages to using prokaryotic hosts such as *E. coli*, namely: they are cheap and easy to culture to high densities (Sezonov et al. 2007), and once made chemically competent, *E. coli* readily take up transposable elements such as plasmids (Pope and Kent 1996). The choice of host species is of great importance, particularly for complex eukaryotic proteins that require post-translational modifications such as glycosylation, which is not performed by most prokaryotic systems (Rosano and Ceccarelli 2014). Serpins typically do not require complex posttranslational modifications and are generally successfully expressed in appropriate *E. coli* host strains (Pearce and Cabrita 2011).

The plasmid vector should also be carefully chosen, as the promoter of a given vector system dramatically influences soluble and insoluble protein distribution (Pearce and Cabrita 2011). One of the most commonly used promoters is the *lac* promoter derived from the *lac* operon of *E. coli*. This promoter induces the expression of downstream transcripts in the presence of lactose. However, if other sugars are present in the culture media, such as glucose, the promoter is not fully induced until the glucose has been metabolised (Rosano and Ceccarelli 2014). This is the principal behind autoinduction media (AIM) which contains lactose, glucose and glycerol. Using AIM, expression remains at a low level for the first several hours when glucose is the main energy source, giving the host time to replicate exponentially. Once all the glucose in the media is metabolised, the cells being to utilise lactose, which initiates transcription and over-expression of the gene of interest under control of the lac promotor. In the later stages of the culture, when the lactose has been metabolised, glycerol can be utilised to maintain the culture. AIM is an effective and efficient culture method that promotes greater cell densities and a higher yield of recombinant protein over

traditional induction with Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Studier, 2014). The pUC vector system utilises the lac promoter. A modification of the *lac* promoter is *tac*, a combination of tryptophan (*trp*) and the *lac* UV5 promoter; the *trp* promoter follows the same principal as *lac* but provides considerably greater expression than *lac* (De Boer et al. 1983). The *trp* promoter is utilised by the pMAL vector system. The commonly used pET system incorperating the T7 promoter (Rosano and Ceccarelli 2014).

A further consideration when selecting the vector is codon bias, that is, an abundance of codons in the gene of interest that are uncommon in the host genome. Arginine, proline, leucine and isoleucine are low abundance codons in the *E. coli* genome. Therefore, the host has a low abundance of tRNA corresponding to these codons, which results in reduced expression of the protein. Several Rosetta *E. coli* strains (Novagen) are commercially available that carry plasmids for these low abundance tRNAs (Bird et al. 2004).

Several serpins from freshwater crayfish have been cloned and overexpressed and many from decapods and invertebrates as a whole have been produced recombinantly. Donpudsa et al. (2010a) detail the recombinant expression of two serpins (KP12 and KP18) from *P. leniusculus*, and subsequent functional studies. In their research the authors cloned the genes and also domains of the serpins separately, ligating the gene fragments into the pVR500 vector, derived by deleting the 5′ His·Tag and S·Tag from the pET-32a (+) vector. This resulted in a single His·Tag downstream (C-terminal end) of the reading frame to generate a fusion protein for protein purification and a Trx·Tag (thioredoxin) at the N-terminal portion. The fusion protein was expressed in Rosetta gami B (DE3) *E. coli,* and each of the serpins and their domains were expressed as soluble proteins (Donpudsa et al. 2010a). After induction (with IPTG) the cells were lysed with a phosphate-buffered saline pH 7.4. The soluble recombinant fusion protein was purified with a Ni-NTA agarose column, and the purified fusion protein was cleaved with enterokinase to remove the Trx·tag, then purified again with the Ni-

NTA column. The authors found that the expressed proteins each had a greater molecular mass that was predicted, and found KP12 inhibits strongly against subtilisin, and KP18 inhibits strongly against trypsin. This inhibition was reflected in the recombinant KP12_domain and KP18_domain.

Li et al. (2009) report the overexpression and characterisation of a three-domain Kazal type serpin (hcPcSPI1) as well as separate characterisation of each of the three individual domains. The protein was found predominantly in the haemocytes of *Procambarus clarkii*. cDNA was synthesised from the hepatopancreas and gill after immune stimulation with white spot syndrome virus. The gene and each of the three domains were amplified from cDNA and ligated into a pET30a (+) vector. hcPcSPI1was transformed into *E. coli* BL21 (DE3), and the domains were transformed into a Rosetta (DE3) *E. coli* with IPTG induced overexpression. The cells were harvested resuspended in lysis buffer (phosphate-buffered saline, pH 8 with 1% TritonX 100) and sonicated. The cell lysate was purified using a His-Bind (Novagen) column. hcPcSPI1 is composed of 168 amino acids with a predicted molecular mass of 18.2 kDa. hcPcSPI1 and its three domains were shown to inhibit proteinase K and subtilisin strongly.

Somprasong (2006) report the overexpression of a five-domain Kazal type 29.1 kDa serpin (SPI*Pm*2) from the shrimp *Penaeus monodon*. The five-domain Kazal nucleotide sequence (excluding the single peptide) was amplified by PCR and ligated into a pET22b (-) vector containing a His-tag. Overexpression was induced with IPTG in a Rosetta (DE3) host. The cells were harvested, and the pellet was frozen, thawed at room temperature and resuspended in lysis buffer (50 mM Tris-HCl, pH8, 5% glycerol and 50 mM NaCl), followed by sonication. The protein was purified from inclusion bodies and solubilised with 50 mM sodium carbonate, pH 10 overnight. The SPI*Pm*2 recombinant protein was purified by Sephadex G-100 gel filtration. SPI*Pm2* was shown to be strongly inhibitory against subtilisin and elastase but weakly against trypsin. Chymotrypsin was not inhibited.

Obtaining the milligram quantities of protein required to complete characterisation assays from a host can demand large quantities of tissue from the species of interest, which is a severe limiting factor when studying protected species such as *A. pallipes*. However, recombinant protein overexpression is a well-established method for the production of large quantities of a protein of interest, and the technique has been applied to successfully synthase and study serpins from all kingdoms of life, including astacids (Bird et al. 2004, Pearce and Cabrita 2011). Therefore, for this work, recombinant protein overexpression was deemed the most appropriate method to obtain large quantities of KP12.

4.3 Aims & objectives of this study

This study aimed to characterise a serpin (KP12) from the freshwater crayfish *A. pallipes*. To meet this aim, cDNA corresponding to the candidate serpin gene was amplified from *A. pallipes* RNA, and the gene was cloned and overexpressed as a fusion protein in combination with maltose binding protein (MPB) in an *E. coli* host using a pMAL vector system. The recombinant KP12 protein was purified by affinity chromatography and size exclusion chromatography.

4.4 Materials & methods

A flowchart is presented (Figure 7) to show the workflow in this study for the overexpression and purification of the recombinant serpin KP12 in this study.

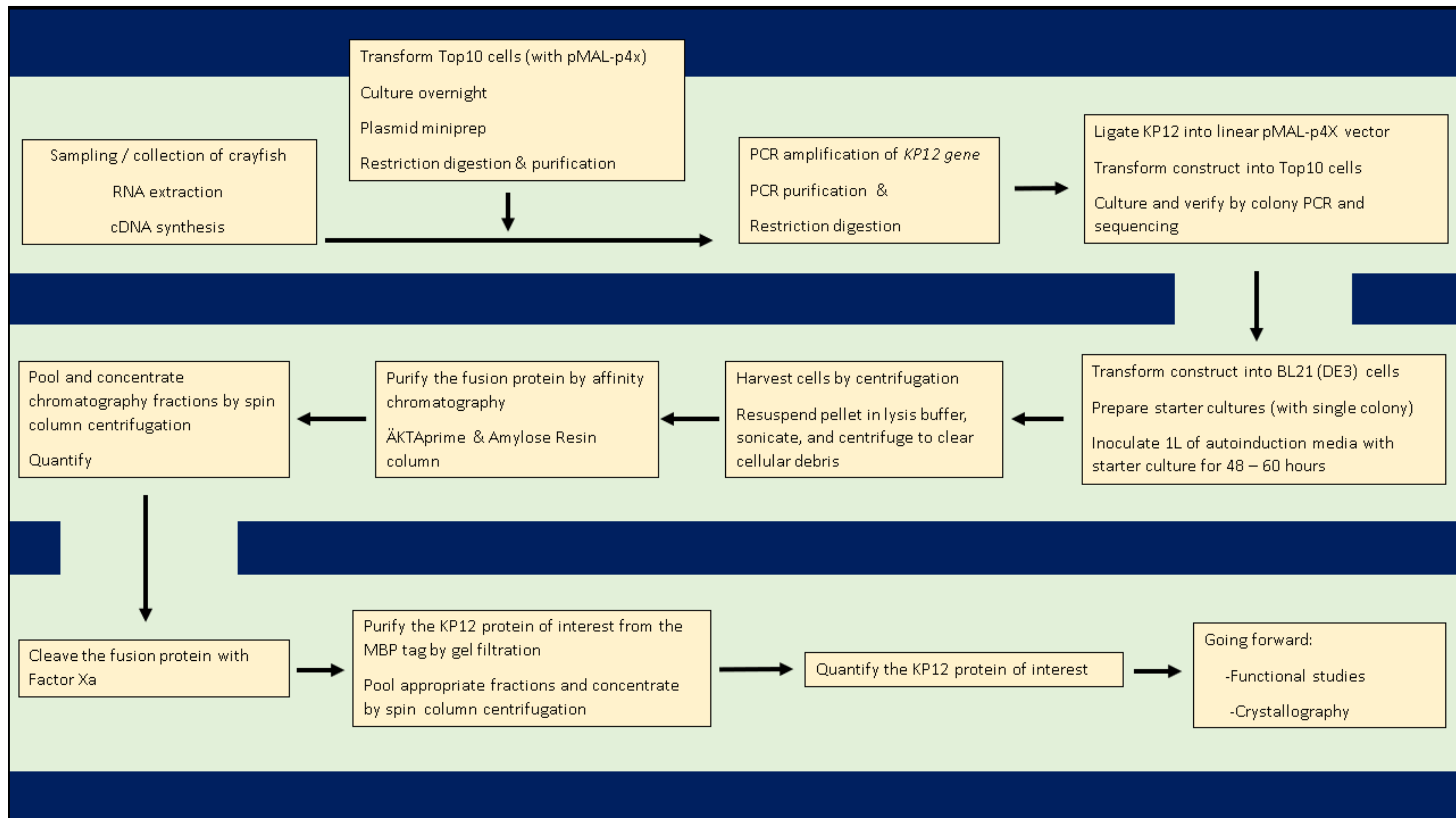


Figure 7. The workflow for recombinant protein overexpression and purification used in this study.

4.4.1 RNA extraction, cDNA synthesis & gene amplification

Live *A. pallipes* were obtained from the in-house crayfish behaviour facility at Galway-Mayo Institute of Technology. Haemolymph was extracted from the ventral sinus of live *A. pallipes* by penetrating through the leg joints using a 19-gauge hypodermic needle and extracting directly into 1 mL of TRI Reagent® (Sigma-Aldrich). To obtain tissue samples, the specimen was sacrificed by destroying the nerve ganglion with a scalpel and placing the dissected tissue into TRI Reagent®. RNA extraction was performed immediately. RNA quantity and sample quality ($A_{260/280}$) was checked on an Epoch Spectrophotometer using a Take3 Micro-Volume Microplate. RNA integrity was assessed by agarose gel electrophoresis (1% gel with 18 Ω H_2O for TAE, as per 4.4.1.1), nondegraded RNA shows two intense bands representing the 28S and 18S ribosomal sub-units. No bands and/or a smear in the lane indicated the RNA has degraded in the sample.

The RNA samples were DNase treated for contaminating DNA using the Precision™ DNase kit (Primerdesign Ltd.). cDNA was synthesised using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems™) and stored at -80°C.

Primers designed by Donpudsa et al. (2010a) for the serine protease inhibitor KP12 in *Pacifastacus leniusculus* (Table 8) were used to cross amplify *kp12* in *A. pallipes*. Standard primers (Table 8) were used to initially amplify the genes in a PCR reaction of: Phusion DNA polymerase (Thermo Scientific), 1x Phusion Green HF buffer, 200 µM each dNTP, 0.5 µM forward and reverse primer, 3 µL of cDNA template, 0.02 U/µL of Phusion DNA polymerase. The PCR cycling conditions were: initial denaturation at 98 °C for 30 seconds, (98 °C for 15 seconds, 59 °C for 30 seconds and 72 °C for 15 seconds) x 30, and a final extension at 72 °C for 10 minutes on a Whatman Thermocycler (Biometra). The PCR product was checked on a 1.2% agarose gel and purified using the QIAquick PCR purification kit. The *A. pallipes kp12* PCR product was then used as DNA template to amplify the *kp12*

gene using the cloning primers KP12_MBP_F and KP12_MBP_R (Table 8) using the same PCR cycling conditions. The PCR products were purified (QIAquick Kit) following manufacturer guidelines.

Table 8. List of oligonucleotides used to amplify serpin KP12.

Primer nam	Plasmid vector	Restriction enzyme	Oligonucleotide	GenBank	Reference
KP12_std_F			GCTCCTGATATAGCTTTCCCC	EU433325	Donpudsa et al (2010a)
KP12_std_R			ATGCTAGAGGCGGATAAACAG		
KP12 MBP_F	pMALp4X	EcoRI	CGC **G*AATTC**ACAAGCCCTACGCG		This study
KP12 MBP_R		BamHI	CCG **G*GATCC** TTAACCTGTGCATTCAC		

The restriction sites are in bold, an * indicates the specific digestion sites, the 5' flanker sequence is in underlined.

4.4.1.1 Agarose gel electrophoresis

Agarose gel electrophoresis is a technique used to separate macromolecules by molecular weight and is the standard method when visualising nucleic acids such as DNA and RNA. Gel density can be increased (1.5 – 2% agarose w/v) to enable the separation of smaller molecular weights nucleic acids and decreased for larger molecules (0.6 – 0.9 % agarose w/v) such as plasmids and kb length PCR amplicons. Agarose gels are composed of agarose powder, buffer and nucleic acid stain. In this study all agarose gels were composed of agarose powder, 1x TAE buffer (50x stock – 242 g Tris base, 57.1 mL glacial acetic acid and 100 mL of 500 mM EDTA, pH 8) and SYBR™ Safe DNA gel stain, (ThermoFisher Scientific). Agarose powder was melted in TAE using a microwave, the agarose was observed and swirled several times. The molten agar was then allowed to cool for several minutes, and SYBR™ Safe was added as a 1:20,000 dilution. The agar was then poured into the casting tray and left to set or set in a refrigerator at 4 °C. The cast gel is submerged in 1x TAE, and the samples are loaded with a molecular weight size ladder as a standard. For this study, HyperLadder™ (Bioline) standards were

used. Gels were run at 80 volts for smaller molecules to 150 volts for larger molecules. The duration of the run depended on the density of the gel and the casting tray used (30 – 60 minutes). Agarose gels were imaged on a Gel Doc EZ (Bio-Rad).

4.4.1.2 Plasmid vector - pMAL-p4X

The pMAL-p4X plasmid contains a sequence coding for MBP (Figure 8), which translates a fusion protein of *MBP + protein of interest* that can be isolated and purified by amylose affinity chromatography. MBP can provide higher purity products compared to more commonly used histidine purification tags. The pMAL-p4X plasmid vector (Addgene plasmid #75289) was a gift from Paul Riggs (Walker et al. 2010) and annotated in figure 2. The plasmid was transformed and maintained in Top10 *E. coli* glycerol stocks at -80°C. Large quantities (micrograms) of plasmids are required to manipulate nucleic acids and construct the recombinant expression plasmid. To obtain these quantities, streak plates (LB agar with Amp100 [final conc. ampicillin 100 µg/mL, hereafter amp100]) were prepared from Top10 *E. coli* glycerol stocks containing the plasmid and were incubated overnight at 37°C. From the streak plate, single colonies were selected, and 5 mL aliquots of Lysogeny broth (LB broth) (with amp100) were inoculated and incubated with vigorous shaking (250 rpm; Forma Orbital Shaker, Thermo Scientific) at 37°C for 8 to 16 hours. Plasmids were purified with the QIAprep Spin Miniprep Kit (Qiagen) and eluted with 100 µL of warmed (65°C) distilled H_2O. DNA yields were quantified with a Qubit 3.0 Fluorometer (ThermoFisher), and quality ($A_{260/280}$) was checked with an Epoch Spectrophotometer and a Take3 micro-volume Microplate (BioTek, Mason Technology).

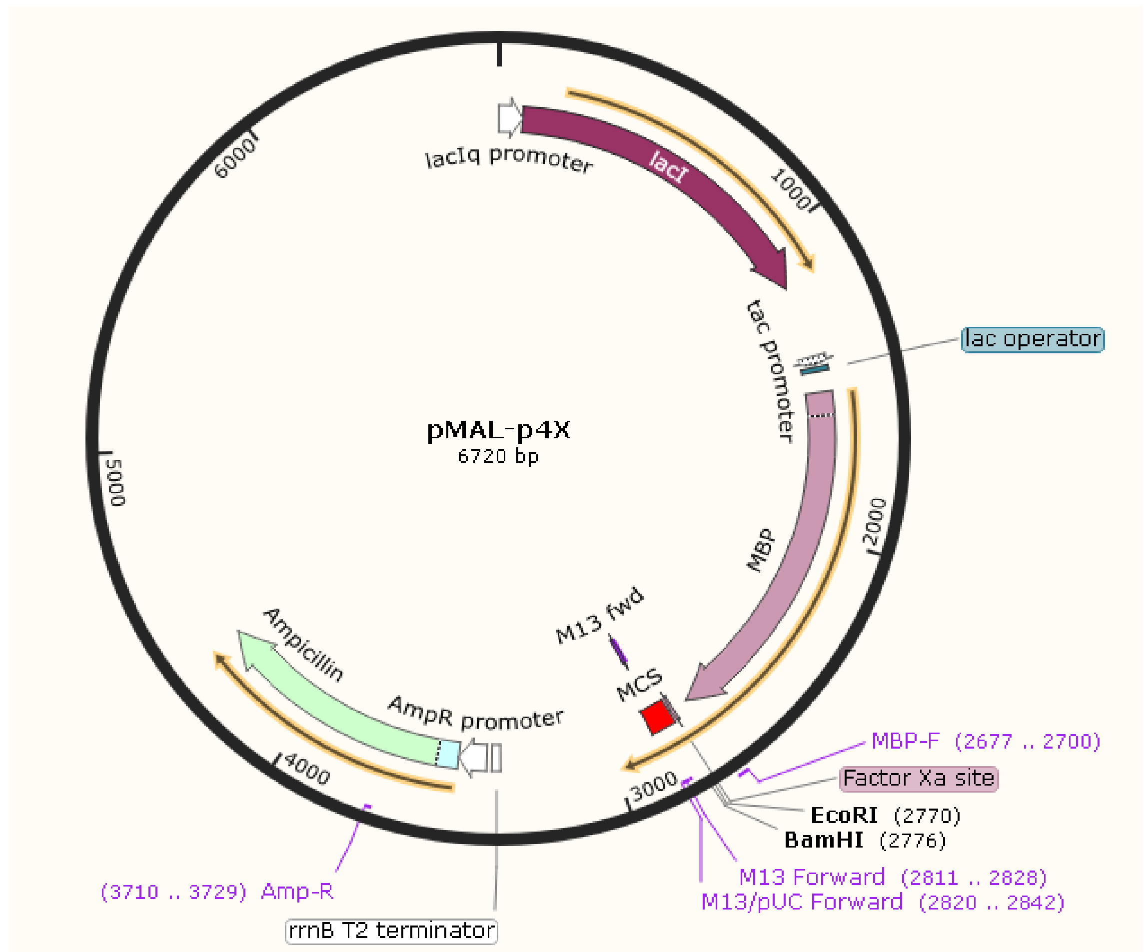


Figure 8. A vector map of pMAL-p4X showing the maltose binding protein sequence upstream of the multiple cloning site (MCS) where the EcoRI and BamHI restriction sites positioned used for this study are positioned within the MCS. The ampicillin resistance sequence is also annotated. Generated in SnapGene viewer v 4.2.5

4.4.1.3 Preparing chemically competent cells

Chemically competent strains of Top10 and BL21 (DE3) *E. coli* were prepared for the cloning of plasmid constructs. The protocol for preparing both strains was the same. A streak plate of the cell strain was prepared from in-house glycerol *E. coli* stocks on LB agar plates and incubated overnight at 37ºC. A single colony was used to inoculate 5 mL LB broth and incubated at 37ºC overnight. From the overnight culture, 0.5 mL was used to inoculate 50 mL of LB broth and incubated at 37 ºC with vigorous shaking (250 RPM) to an optical density (600 nm) of 0.5 –

0.7, then incubated on ice for ten minutes. At 4 °C the cells were centrifuged at 4,000 g (Sorvall Legend Micro 21R centrifuge, Thermo Scientific) for five minutes and gently resuspended by pipetting in sterile $CaCl_2$ and incubated on ice for 20 minutes to several hours. The cells were then pelleted at 4,000 g and resuspended in fresh $CaCl_2$, repeated twice more, without an extended incubation on ice for the latter two washes. The final pellet was resuspended in 0.5 mL of ice-cold 0.1 M $CaCl_2$, 15% glycerol. 50 µL aliquots of the chemically competent cells were prepared in prechilled 1.5 mL microcentrifuge tubes and stored at -80 °C until required.

4.4.1.4 SDS-PAGE and protein verification

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) is used to separate macromolecules, particularly proteins based on their molecular mass. Samples were applied to a hand cast 4% stacking gel (1.98 mL 30% Acrylamide/bis, 3.78 mL 0.5M Tris-HCl pH 6.8, 150 µL 10% SDS, 9 mL dH_2O, 15 µL TEMED (Sigma Aldrich), 15 µL APS) and 12% resolving gel (6.0 mL 30% Acrylamide/bis, 3.75 mL 1.5M Tris-HCl pH 8.8, 150 µL 10% SDS, 5.03 mL dH2O, 15 µL TMED (Sigma Aldrich) and 150 µL 10% APS) polyacrylamide gels and electrophoresis was allowed to proceed at 100 V for 25 minutes followed by 200 V for 60 minutes.

4.4.1.4.1 Coomassie staining

Gels are washed with distilled water (dH_2O) three times for five minutes each and stained with Coomassie blue (Bio-Rad) for several hours to overnight with slow agitation on a mini see-saw rocker (SSM4, Stuart). The Coomassie was decanted, and the gel was rinsed three times with dH_2O water and placed in de-stain (30% methanol, 10% acetic acid and 60% H_2O) for several hours with slow agitation. The gel was imaged on a Gel Doc EZ (Bio-Rad).

4.4.1.5 Western Blot

To verify protein over-expression, SDS-PAGE was completed as in the Coomassie protocol without Coomassie staining. The unstained gel and six sheets of absorbent filter paper were cut to gel size and placed in transfer buffer (25 mM Tris-HCl pH 7.6, 192 mM glycine, 20% methanol, 0.03% sodium dodecyl sulfate (SDS). One sheet of PVDF membrane (Immobilon®, Sigma-Aldrich) was cut to the same size and soaked in methanol for three minutes to activate the membrane. The gel, filter paper and membrane were further soaked in transfer buffer for ten minutes. The materials were assembled into a transfer sandwich on a semi-dry blotter unit (V20-SDB, SCIE-PLAS) as depicted in figure 9. The sandwich was placed on the base of the semidry blotter, which is positively charged, and the top (upper negative electrode) was attached. The transfer was run for 55 minutes at 400 mAmps and 25 volts. After the transfer, the PVDF membrane was stained with Ponceau stain to verify the protein was transferred from the SDS-PAGE gel to the PVDF membrane. The membrane was then washed in TBS buffer (10x stock: 24g Tris-HCl, 5.6g Tris base, 88 g NaCl, to 1 L with dH_2O) and incubated in blocking buffer (4% w/v powder milk in TBST [stock: 100 mL TBS 10x, 899 mL dH_2O, 1 mL Tween 20]) for one hour with gentle shaking. The membrane was then washed three times with TBST for five minutes. The PVDF membrane was incubated with a 1:2,000 dilution of Anti-MBP Monoclonal Antibody (New England BioLabs Inc.), diluted 1% w/v BSA in TBST for one hour with gentle shaking. The antibody was decanted, and the membrane was washed with TBST three times for five minutes each. The final signal was developed with 2 mL of Pierce™ CN/DAB substrate solution (200 µL) and stable peroxide substrate buffer (1,800 µL) (Thermo Scientific). Signal development was stopped when the bands were visibly clear by washing the membrane with dH_2O.

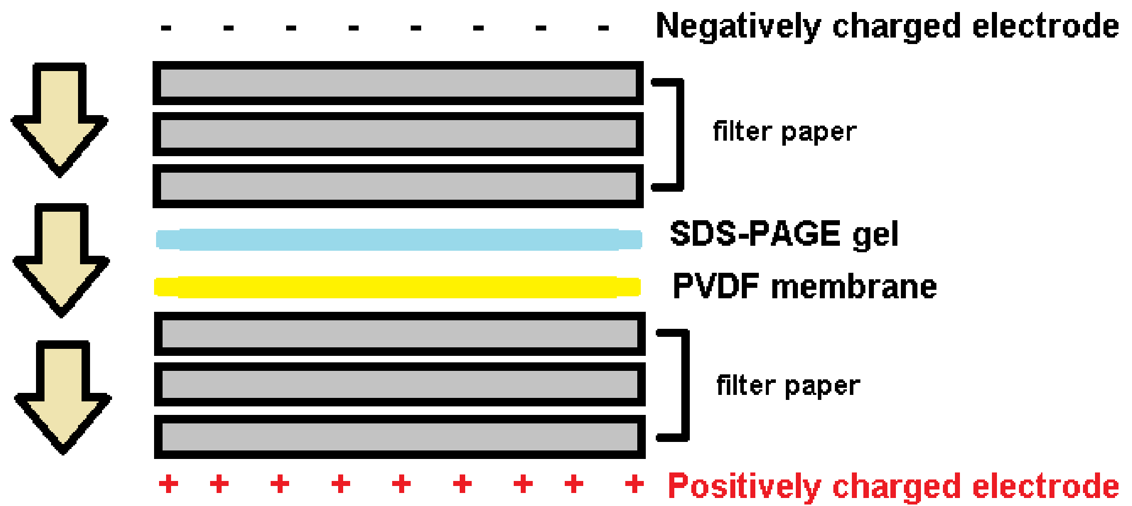
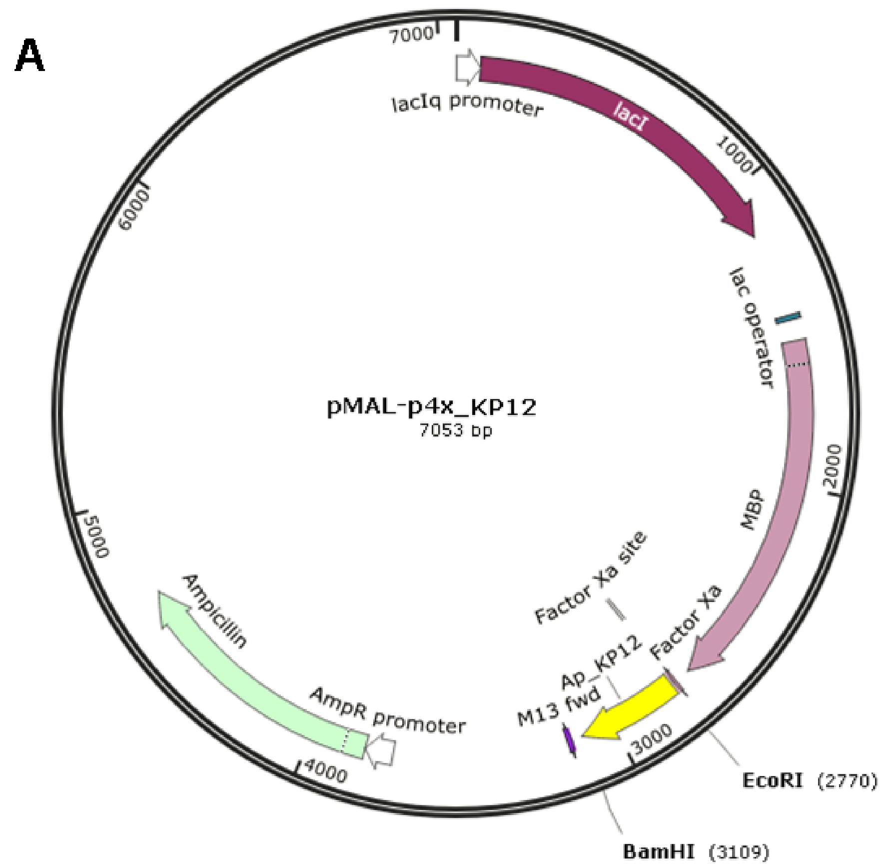


Figure 9. Cross-section view of a Western Blot transfer stack. The current travels in the direction of the arrows from the negatively charged electrode to the positively charged electrode causing the migration of proteins in the SDS-PAGE onto the PVDF-membrane.

4.4.2 Expression plasmid construction

To generate the pMAL-p4X_KP12 construct Figure 10a) 1 µg of purified pMAL-p4X plasmid and 1 µg of *kp12* purified PCR product was restriction digested with BamHI and EcoRI (with 1x Bam Buffer) at 37°C for 90 minutes. The digested plasmid and *kp12* insert were gel purified (1.2% low melting point agarose gel) and the DNA was purified with the QIAquick Gel Extraction Kit (Qiagen) and quantified with the Qubit 3.0 Fluorometer. The linearized vector and insert were ligated with Anza™ T4 DNA Ligase Master Mix (insert:vector ratio of 3:1) in a final volume of 20µL with incubation at room temperature for 30 minutes. Five microliters of the ligation reaction was immediately transformed into chemically competent Top10 *E. coli* (as per 4.4.1.3), and the remainder was frozen at -80°C.



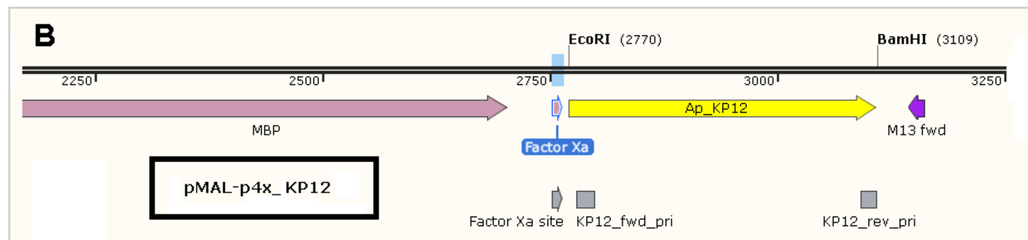


Figure 10**A**. A vector map of pMAL-p4X_KP12 construct showing the KP12 gene inset downstream of the maltose binding protein tag sequence. Figure 10**B** shows the linearised construct map with the Factor Xa cleavage site highlighted. Generated in SnapGene viewer v 4.2.5.

4.4.3 Bacterial transformation & construct verification

Chemically competent Top10 cell cells were thawed on ice for 30 minutes, 5µL of the ligation reaction was added, and the mixture was incubated on ice for 30 minutes. The cells were then heat shocked at 42 °C for 45 seconds and immediately incubated on ice for 3 minutes, 250 µL of LB broth was added, and the reaction was transferred to a sterile 15 mL falcon tube and incubated at 37 °C with vigorous shaking for 60 minutes. The cells were then spread on LB agar (with

Amp100) and incubated overnight at 37 ºC. Colony PCR was performed in 25μL reactions with Promega GoTaq (1x GoTaq® Green Reaction Buffer, 1.5 mM $MgCl_2$, 0.2 mM each dNTP), 1.25 U GoTaq® DNA Polymerase, 1.0 μM forward and reverse KP12 primer. A single colony was smeared on the bottom of a PCR tube to act as DNA template. The PCR cycling conditions were: initial denaturation at 98ºC for 10 minutes; 35 cycles of 95ºC for 30 seconds, 59ºC for 30 seconds, 72ºC for 30 seconds; and a final extension step at 72ºC for 5 minutes. The PCR products were visualised on 1.2% agarose gel. Colonies positive for the KP12 insert were used to inoculate 10 mL of LB broth (with Amp100) and incubated with vigorous shaking for 8 hours. The putative plasmid construct was purified with the QIAprep Spin Miniprep Kit (Qiagen), and Sanger sequencing was completed to verify the KP12 sequence on an ABI 3730XL sequencer (GATC, Konstanz, Germany) using the standard M13 forward primer (5´-GTAAAACGACGGCCAG-3´) supplied by GATC.

4.4.4 Recombinant KP12 protein over-expression

Starter cultures for protein over-expression were prepared by inoculating 10 mL of sterile LB broth (with Amp100, hereafter amp100) with a single transformant colony and incubated overnight with vigorous shaking (250 rpm, Forma Orbital Shaker). The media chosen for protein over-expression was a simplified autoinduction media (AIM) prepared as per Studier (2014). Each litre of media contained 6g Na2HPO4 and 3g KH2PO4, 20g tryptone, 5g yeast extract, 5 g NaCl, 60% v/v glycerol, 10% w/v glucose, 8% w/v lactose and amp100. Expression cultures were inoculated 1:100 will starter cultures and incubated with vigorous shaking (275 rpm) for 60 hours. A no plasmid control culture of BL21 (DE3) was concurrently incubated, excluding ampicillin in the medium.

100 µL aliquots were taken from the no plasmid control and induced culture then mixed with an equivalent volume of 2x SDS-dye (50 mM Tris-HCl, pH6.8, 2% w/v

SDS, 0.1% w/v bromophenol blue, 10% v/v glycerol), placed in a boiling water bath for 7-10 minutes and frozen at -80°C.

4.4.4.1 Optimising protein overexpression with ethanol

Irving et al. (2003) optimised the overexpression of a serpin in *E. coli* (strain SG13009) with the addition of ethanol to LB media, and Chhetri et al. (2015) describe a method to enhance the overexpression of recombinant proteins by the addition of analytical grade ethanol to LB media during incubation. To test whether ethanol improves the yield of the recombinant pMALp4X_KP12 fusion protein the pMALp4X_KP12 plasmid was transformed into chemically competent BL21 *E. coli* and incubated at 37°C overnight with shaking at 250 rpm. The following morning single colonies were used to inoculate seven 15 mL cultures: 1) BL21 in LB broth, 2) BL21 in AIM, 3) pMALp4X_KP12 in AIM, 4) pMALp4X_KP12 in AIM with 1% EtOH, 5) pMALp4X_KP12 in AIM with 2% EtOH, 6) pMALp4X_KP12 in AIM with 3% EtOH, 7) pMALp4X_KP12 in AIM with 4% EtOH. At three time points (18 hours, 42 hours and 66 hours post inoculation) 30 µL aliquots were taken from each culture and mixed with 30 µL of 2x SDS dye then boiled for 7 minutes. Samples were run on an SDS-PAGE gel (4% stacking/12%resolving gel, as per 4.4.1.4) and stained with Coomassie Brilliant Blue (R-250, BioRad) to visualise overexpression of the pMALp4x_KP12 recombinant protein. SDS-PAGE gels were imaged on a Bio-Rad Gel Doc™ EZ System.

4.4.4.2 Overexpression validation and cell lysis

Cell cultures were centrifuged at 4,000 g for 10 minutes at 4°C (Sorvall™ RC 6 Plus, Thermo Scientific) and pellets were resuspended in 3x the pellet weight (in mL) in chilled cell lysis buffer (10 mM Tris-HCl, pH 8.0, 100 mM NaCl). 10 µL of phenylmethanesulfonyl fluoride (PMSF) (Sigma Aldrich) was added per mL of cell

suspension, and the suspension was frozen at -20ºC for a minimum of 24 hours. After thawing, lysozyme was added to a final concentration of 300 µg/mL and the suspension was agitated on ice for 60 minutes. To complete the cell lysis, the suspension was then sonicated on ice for two minutes *on*, at 50% amplitude in 20/40 second *on/off* cycles and centrifuged at 4ºC for 60 minutes at 18,000 g (Sorvall™ RC 6 Plus, Thermo Scientific). The supernatant containing the recombinant fusion protein was stored on ice and refrigerated at 4ºC or immediately used for protein purification (as per 4.4.4.1). Protein overexpression was verified by SDS-PAGE (as per 4.4.1.4), Coomassie staining (as per 4.4.1.4.1) and western blot (as per 4.4.1.5).

4.4.4.3 KP12 Protein purification

After recombinant protein overexpression is verified, the protein of interest must be separated from native cellular proteins, cellular debris and finally the fusion protein tag. The pMAL-p4x expression vector system generates a fusion protein containing MBP fused to the protein of interest. Using affinity chromatography, the maltose binding protein tag binds to an Amylose Resin column and is washed to clear cellular debris. The fusion protein is eluted, cleaved and separated by size exclusion chromatography (gel filtration chromatography), which separates molecules in a sample by molecular size and should result in a purified protein of interest.

4.4.4.1 Fusion protein purification

The overexpressed fusion protein was purified by affinity chromatography with the ÄKTAprime (General Electric) chromatography system. The affinity column used was a 1 x 5 cm glass Econo Column (BioRad) packed with a 3 mL final bed volume of Amylose Resin High Flow (New England Bioscience). The column was

equilibrated with 5 column volumes of Equilibration Buffer (20mM Tris, 200mM NaCl, 1mM EDTA). To bind the MBP-KP12 overexpressed protein to the column, cell lysate (prepared as per 4.4.4.1) was loaded at a flow rate of 1 mL/min onto the column. Unbound contaminant proteins were washed from the column with 10 column volumes of Equilibration Buffer. The MBP-KP12 fusion protein was eluted in 5 mL of Elution Buffer (50mM Tris-HCl pH 8.0, 20mM NaCl, 1mM EDTA, 10mM maltose). MBP-KP12 peak fractions were pooled and concentrated. The methods attempted to concentrate the fusion protein were: SnakeSkin™ Dialysis Tubing (ThermoFisher) incubated with PEG-100 at 4 °C for approximately 4 hours. Furthermore, 10 kDa cut-off Amicon® Ultra Centrifuge Filters (Sigma-Aldrich) were loaded with 2 mL of the pure MBP-KP12 fusion protein and centrifuged at 4,400 rpm at 4 °C. The concentrated fusion protein was then quantified using the BCA method (Pierce BCA protein assay kit, ThermoFisher).

4.4.4.2 Optimisation of fusion protein cleavage with Factor Xa

Once the fusion protein was verified by SDS-PAGE and western blot, the affinity tag (MBP) was cleaved and separated from the target protein (KP12). Cleavage was completed by Factor Xa, a peptidase that is composed of two polypeptides linked by a disulfide bond. It is a site-specific endopeptidase (Ile-Glu-Gly-Arg) that cleaves the C-terminal peptide bond of the recognition sequence, and ideal cleavage conditions must be optimised for each protein. Four 40 µl test solutions were prepared to contain; 20 µg of the fusion protein, 1x reaction buffer (20 mM Tris-HCl, pH 6.5; 50 mM NaCl; 1 mM $CaCl_2$) and varying Factor Xa concentrations (Table 9).

Table 9. The composition of reaction solutions for optimisation experiment of Factor Xa cleavage of pMAL-p4x_ KP12.

Reagent	Volume *(µL)*	Final conc.
MBP_KP12	14	0.5 µg/mL
Factor Xa[a]	1	1.0x[b], 0.2x, 0.05x
Reaction buffer	25	1x[c]

a = 1ug of Factor Xa cleaves 50 µg of substrate in 6 hours;

b = 1.0 X Factor Xa initial concentration is 1 mg/mL;

c = 1X Reaction buffer (20 mM Tris-Cl, pH 6.5; 50 mM NaCl; 1 mM CaCl2)

The 40 µl reactions were incubated at 22°C and at four-time points (3hrs, 7hrs, 10hrs, 27hrs) 8 µL was taken from each reaction, mixed thoroughly with 2µL of 6x SDS-PAGE dye, boiled for 7.5 – 10 minutes and frozen. Cleavage efficiency was determined by SDS-PAGE (as per 4.4.1.4).

4.4.4.3 KP12 purification by gel filtration

Following the optimisation of Factor Xa cleavage, the optimal reaction conditions were applied to an up-scaled volume of the MBP-KP12 fusion protein. After overexpression and affinity chromatography, 5.4 mg of the MBP-KP12 fusion protein was digested with 0.2x Factor Xa in reaction buffer to a final volume of 8 mL, incubating at 23 °C overnight. The Factor Xa cleavage reaction was concentrated by spin centrifugation (3 kDa cut off) to a final volume of 2 mL. The MBP and KP12 proteins were purified by size exclusion chromatography (gel filtration) using a Superdex 200 column (GE Healthcare). The column was equilibrated with 10 column volumes of Equilibration Buffer (20mM Tris, 200mM NaCl, 1mM EDTA) before the 2 mL cleavage reaction mixture was injected into the ÄKTAprime chromatography system, the system was loaded with equilibration buffer at a flow rate of 0.2 mL/min to separate the proteins and monitored at 280 nm.

4.4.5 Generating 3-dimensional a structure of KP12

The *kp12* nucleic acid sequences from *A. pallipes* was aligned in UCSF Chimera (Pettersen et al. 2004) and predicted secondary structures were annotated. The amino acid sequences were individually uploaded to Phyre2 Protein Homology/analogy Recognition Engine V 2.0) and the *Intensive* modelling mode was used to predict the three-dimensional structure of the KP12 serpin. Hidden Markov models were used to detect and align the sequences (Kelley et al. 2015) against known protein structures in the protein databank (www.rcsb.org). The file generated by Phyre2 was imported to UCSF Chimera, and a 3D ribbon diagram was generated (coloured by rainbow N → C terminus). Condon that contribute to codon bias were highlighted (arginine, proline, leucine and isoleucine).

4.5 Results

4.5.1 General materials

4.5.1.1 Oligonucleotides for serpin *k12*

There are no serine protease inhibitor gene sequences available for *A. pallipes* on GenBank. Therefore, amplification was completed by cross amplifying KP12 using oligonucleotides designed for the related species *Pacifastacus leniusculus* (KP12_std_F and KP12_std_R) by Donpudsa et al. (2010a), listed in table 8.

4.5.2 RNA extraction, cDNA synthesis & gene amplification

Total RNA was extracted from haemolymph, gonad tissue and gill tissue of *A. pallipes*, and quantified on a BioTek spectrophotometer with two replicates per sample, the mean quantities (ng/µL), $A_{260/280}$ values between 1.8 and 2.0 indicate no contaminants are resent in the sample (Table 10). All the RNA samples yielded good concentrations of RNA varying from 569 – 796 ng/µL. RNA quality determined by $A_{260/280}$ vary from 2.05 to 2.08 for samples 2 – 4 which are all excellent quality without contamination. Sample 1 has a very low $A_{260/280}$ of 1.10 indicating heavy protein contamination (Table 10).

RNA integrity was checked on a 1% agarose gel and showing strong, distinct bands of the 18S and 28S ribosomal subunits (Figure 11) indicating the RNA was not degraded and of sufficient quality to complete RT-PCR (RT-PCR). The synthesised cDNA was used as DNA template to amplify a 324 bp KP12 gene; this PCR product was purified and used as DNA template to amplify KP12 using KP12_MBP_F and KP12_MBP_R cloning primers. This product was purified and double digested with EcoRI and BamHI, then visualised on a 1.2% agarose gel with HyperLadder 50bp as per (Figure 12). A clear single band of approximately 300 bp present (Figure 12), indicating the KP12 gene was amplified.

Table 10. Total RNA extraction, quantity and quality from different tissues in *Austropotamobius pallipes*.

RNA sample	Tissue source	Quantity (ng/µL)	$A_{260/280}$
1	Haemolymph	691	1.10
2	Haemolymph	796	2.05
3	Gonad	544	2.08
4	Gill	569	2.07

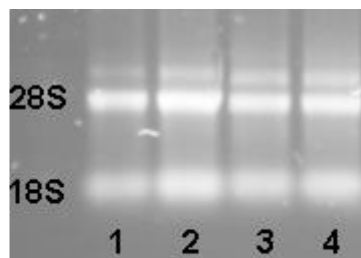


Figure 11. RNA integrity gel for *Austropotamobius pallipes* RNA extracts (1% agarose gel with SYBR Safe).

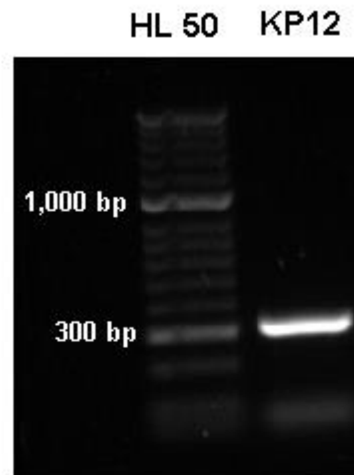


Figure 12. EcoRI and BamHI double digested KP12 PCR product from *Austropotamobius pallipes*, with HyperLadder 50bp (HL50) (1.2% agarose gel with SYBR Safe).

4.5.3 Expression plasmid construction

The EcoRI/BamHI double digested pMAL-p4X plasmid should equate to approximately 6.7 kb in length and the digestion product showed a band between the 6 kb and 8 kb bands on HyperLadder 1 kb (Figure 13), indicating successful double digestion. The digested KP12 insert was ligated into the linearised pMAL-p4X vector. The putative pMAL-p4x_KP12 construct was transformed into chemically competent Top10 *E. coli* and cultured on an LB agar plate with amp100 at 37 ^{o}C overnight. The following morning, multiple single colonies were visible on the plates except for the control (no plasmid) plate.

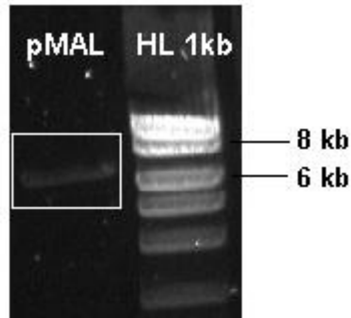


Figure 13. The EcoRI/BamHI double digested pMAL-p4X (white box) run with HyperLadder 1kb (HL 1 kb) (0.8% agarose gel with SYBR Safe).

4.5.4 Bacterial transformation & construct verification

To confirm whether single colonies on the overnight LB agar plate contained the pMAL-p4X_*kp12* construct, a colony PCR was performed on three colonies and a no template control (dH_2O) (Figure 14). Following plasmid purification (miniprep), samples 1 and 2 were sent for Sanger sequencing. Sample 1 did not contain the *kp12* insert; sample 2 did contain the *kp12* insert.

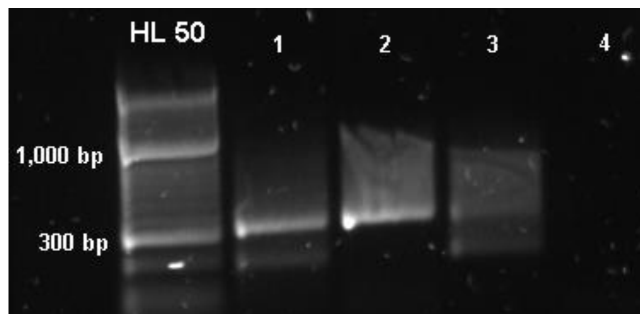


Figure 14. Colony PCR of three Top10 *E. coli* colonies grown overnight on LB agar with ampicillin at a final concentration of 100 µg/mL following transformation with pMAL-p4X_ *kp12* ligation reaction mixture (lanes 1 – 3) and a no template control (lane 4). Run with HyperLadder 50 bp (far left) on a 1.2% agarose gel stained with SYBR Safe.

4.5.5 Recombinant protein overexpression

4.5.5.1 Optimising protein overexpression with ethanol

To observe whether fusion protein expression could be enhanced by the addition of EtOH to the culture broth, molecular grade ethanol was added to cultures at varying concentrations. Samples were taken at three-time points, mixed with SDS loading dye and over-expression was observed by SDS-PAGE (4% stacking/12% resolving gel, as per 4.4.1.4), MBP is 42.5 kDa, and KP12 is approximately 11.5 kDa. Therefore the recombinant protein is approximately 54 kDa. Overexpression appears greatest in AIM with no EtOH added and at 48 hours incubation (Figure 15). The overexpression observed from the SDS-PAGE (Figure 15) consistently showed that protein yield appears greatest where no EtOH was added to the culture media and also at timepoint 2.

4.5.5.2 Protein overexpression validation and cell lysis

After a 60-hour incubation at 37 ºC with vigorous shaking, 3 litres of AIM media yielded a pellet weighing 32.45 g. Following cell lysis, sonication and centrifugation (as per 4.4.4.2), 59 mL of cell lysate was prepared for affinity chromatography.

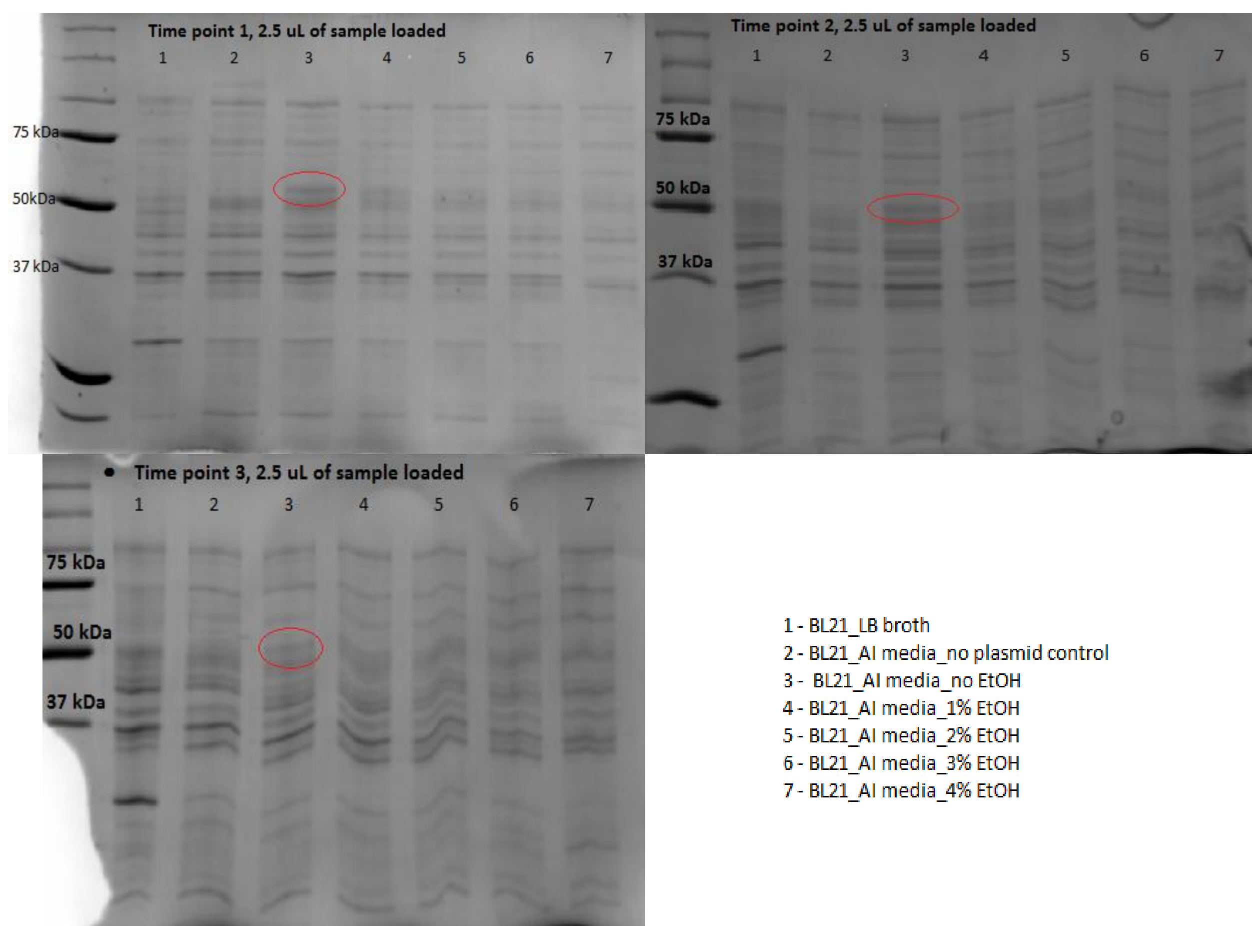


Figure 15. The overexpressed MBP_KP12 in BL21 (DE) *E. coli* with 0 to 4% analytical grade ethanol, a no plasmid control in autoinduction media and host culture control in LB broth. SDS-PAGE gel of 4% stacking-12% resolving gel.

4.5.6 MBP_KP12 fusion protein purification

The MBP_ KP12 fusion protein was purified using affinity chromatography loading onto an Amylose Resin column using an ÄKTAprime HPLC system. The fusion protein was eluted, and the five 1 mL collected fractions were pooled, yielding 11.4 mg total fusion protein. The fusion protein (approximately 54 kDa) can be seen in lane 8 of figure 16; there is also a large quantity of MBP (42.5 kDa) and a potential aggregate mass higher in the lane that is not present in the other samples. The fraction 4 – 7 were collected in figure 16 can be compared to the chromatogram generated during affinity chromatography (Figure 17) that shows

cleared native cellular proteins (10 min to 140 min) followed by the eluted fusion protein (180 to 200 min). The western blot confirmed the presence of MBP-KP12 fusion protein in the pooled purified fractions using anti-MBP antibody immunoblotting (Figure 18).

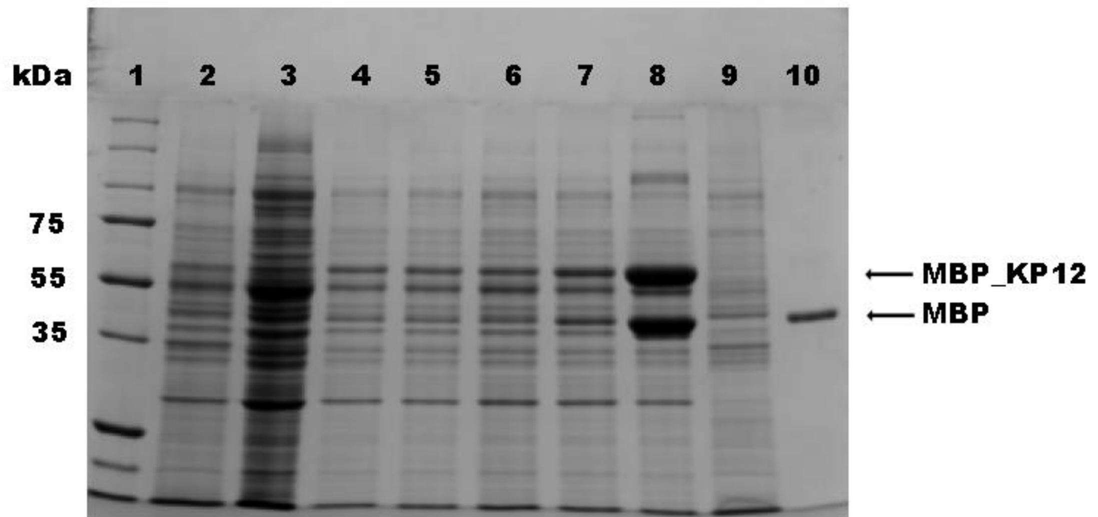


Figure 16. An SDS-PAGE gel of the stages of MBP_KP12 fusion protein purification by affinity chromatography using an Amylose Resin column. The samples present in each lane are as follows: 1 = unstained ladder; 2 = induced cells at 24 hours; 3 = induced cells at 6

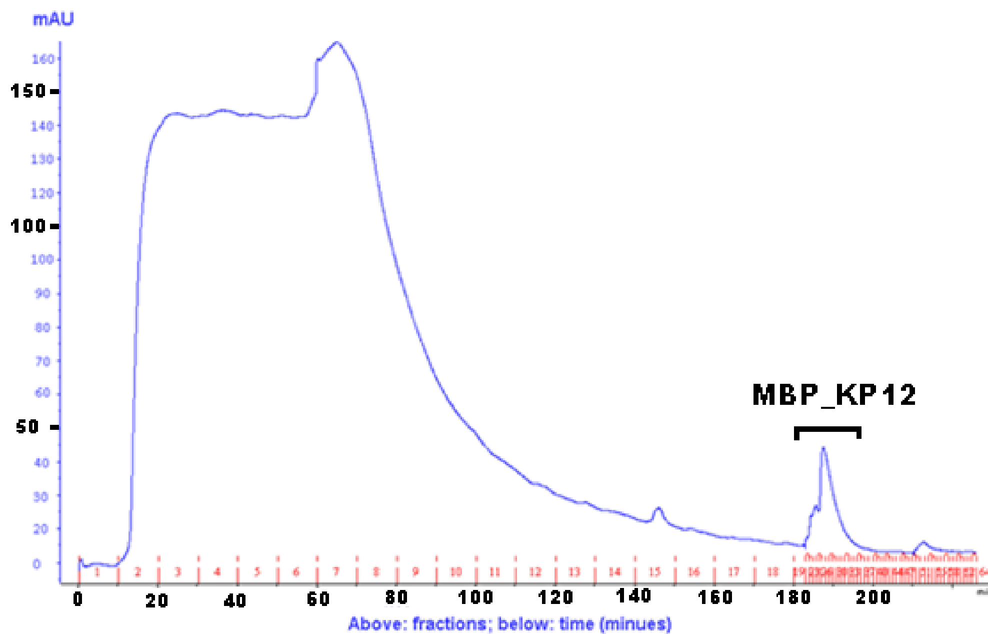


Figure 17. A chromatogram of the purification of MBP_KP12 from whole cell protein extract on an ÄKTAprime FPLC system with an Amylose Resin affinity column. The absorbance (at 280 nm) shows the presence of protein. The first peak (fractions 2 to 10) represents the elution of unbound contaminant proteins. The second peak is the MBP-KP12 fusion protein eluted with 10 mM maltose at approx. 180 minutes.

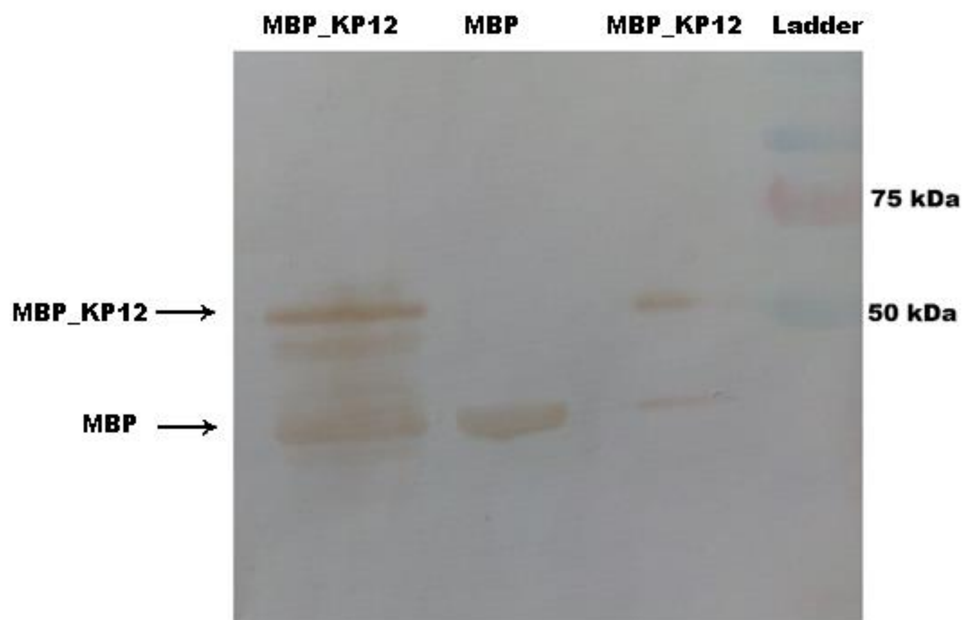


Figure 18. Western blot immunoassay of MBP-KP12 fusion protein with MBP control. The western blot shows the presence of MBP_KP12 and MBP by anti-MBP monoclonal antibody staining. The first lane shows fusion protein at the appropriate molecular weight for MBP_KP12 (54 kDa) with degraded fusion protein. The second lane is a pure MBP control. The third lane is a smaller quantity of MBP-KP12 and MBP can be seen in this sample also. The size ladder confirms the molecular weight of the transferred protein.

An attempt to concentrate the fusion protein by dialysis did result in a slight increase in concentration from 0.78 mg/mL to 0.81 mg/mL. However, the sample became cloudy which can indicate aggregation and precipitation of the protein. Spin columns enabled a higher concentration to be achieved. Therefore this approach was adopted.

4.5.6.1 Optimisation of fusion protein cleavage with Factor Xa

The SDS-PAGE showed complete digestion of MBP-KP12 digested with 1x Factor Xa (Figure 19) across all time points; samples digested with the 0.2x Factor Xa concentration showed an increase in cleavage of the fusion protein over time, with complete cleavage at time point 4, though the KP12 appears to have a higher yield at time point 3 for the 0.2x concentration. Limited cleavage is seen

with a Factor Xa concentration of 0.05x over the first three time points, with significant cleavage at time point 4; the negative control shows no cleavage and no KP12 protein Figure 19.

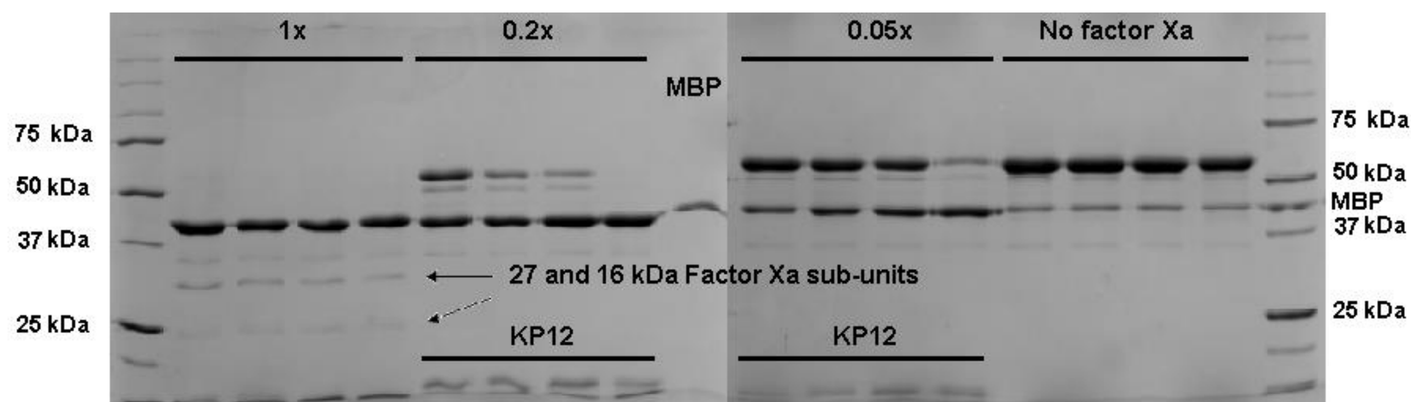


Figure 19. A composite image showing the optimisation analysis of Factor Xa cleavage efficiency on MPB_KP12 fusion protein. Extracts were prepared from chemically competent BL21 (DE3) cells transformed with pMALp4X_*kp12* (as per 4.4.1.3). Recombinant protein overexpression was carried out in autoinduction media for 60 hours, and the expressed MBP_KP12 fusion protein was purified on an ÄKTA Prime with an Amylose Resin affinity column. The recombinant MBP_KP12 fusion protein was then digested with Factor Xa at varying concentrations (1x, 0.2x, 0.05x and a no Factor Xa control) and sampled across four time points (T1 = 3 hrs, T2 = 7 hrs, T3 = 10 hrs and T4 = 27 hrs), then analysed on a 4% stacking-12% resolving SDS-PAGE gel (as per 4.4.1.4) to visualise optimal cleavage conditions.

4.5.6.1.1 Upscaling Factor Xa cleavage

The Factor Xa cleavage incubation was upscaled using a final 0.2x concentration of Factor Xa and 5.4 mg of the fusion protein. After a 10-hour incubation at 22 °C and concentration by spin-column centrifugation, the total quantity of protein was 3 mg.

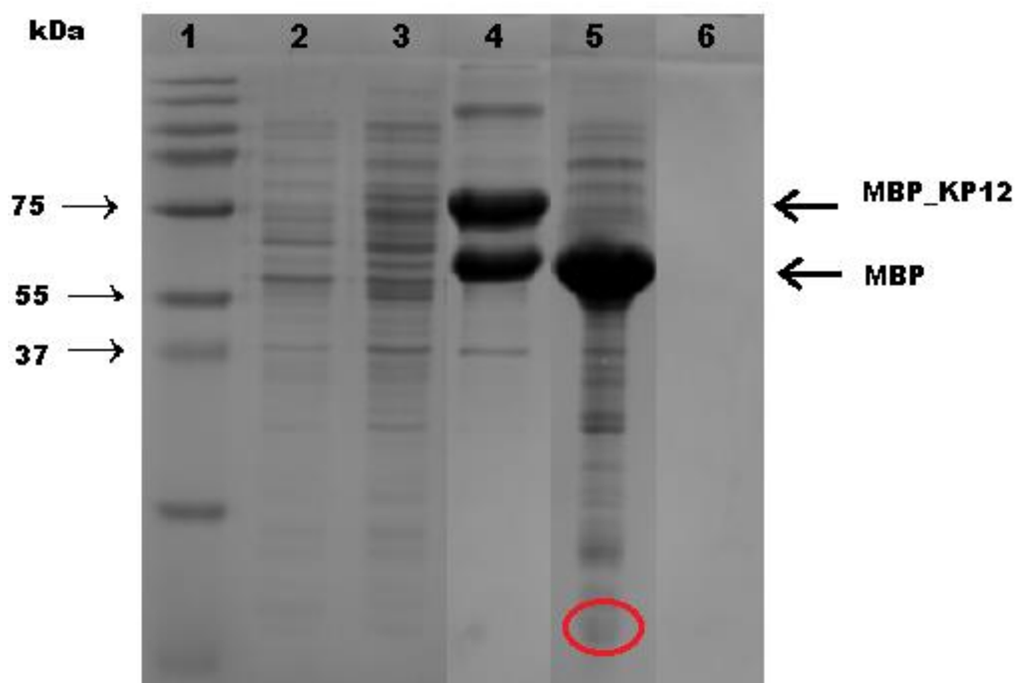


Figure 20. A summary SDS-PAGE gel (composite image) of MBP-KP12 over-expression, purification by affinity chromatography and separation by the target protein separation by gel filtration. Lane 1 is pre-stained protein ladder, lane 2 is no pMAL-p4X_KP12 plasmid control BL21 (DE3) cells in LB broth, lane 3 is induced pMAL-p4X_KP12 in BL21 (DE3) cells in autoinduction media, lane 4 is post MBP affinity chromatography, lane 5 is post Factor Xa cleavage, and lane 6 is KP12 post gel filtration and concentration (no band). The unedited image can be viewed in Appendix 1.

4.5.6.2 KP12 purification by gel filtration

After cleavage of the fusion protein and concentration of the cleavage reaction mixture, the MBP and KP12 protein of interest were separated and purified by gel filtration. Gel filtration separates molecules by size, with larger molecules passing through the column more quickly. The chromatogram (Figure 21) shows two clear peaks in the flow through; the first peak (MBP) eluted after approximately 90 minutes and was collected in fractions 10 and 11. The second protein (KP12) eluted after approximately 130 minutes and was collected in fractions 22 and 23. Fractions 20 and 21 skipped during the run. However, approximately 0.2 mL was retrieved from these factions. In total, three milligrams of cleaved MBP-KP12 were purified by gel filtration, and 0.42 mg/mL of purified KP12 were obtained in approximately

2.2 mL. This was concentrated to 0.2 mL by spin column centrifugation (3 kDa cut-off concentrator), but the final KP12 yield was only 0.07 mg.

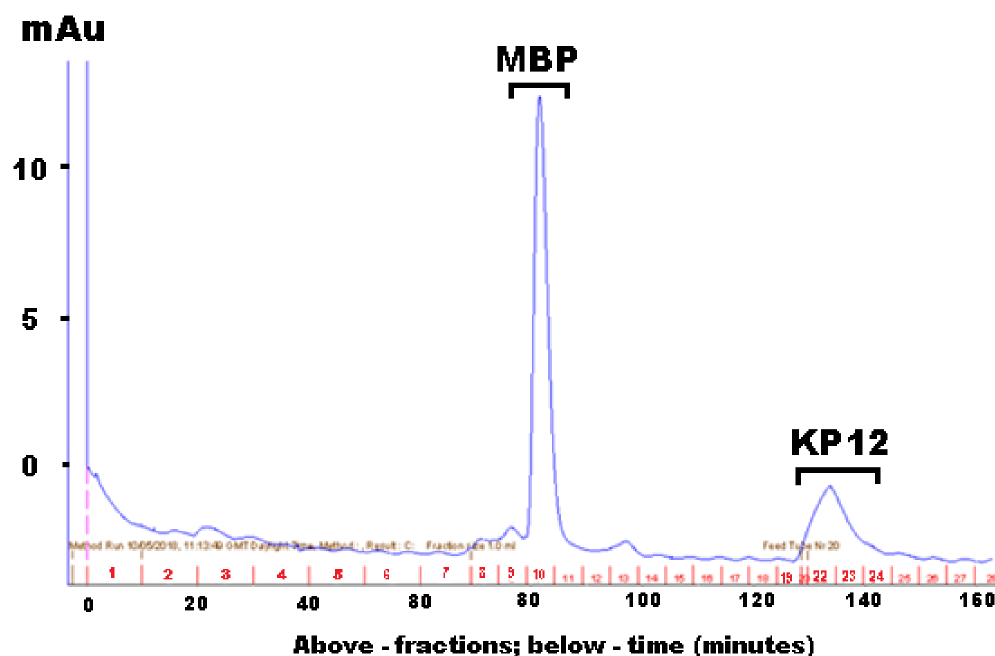


Figure 21. A chromatogram of the separation of maltose binding protein from KP12 after cleavage with Factor Xa. The first peak at 90 minutes is purified maltose binding protein and the second peak at approximately 125 minutes is purified KP12.

4.5.7 Generating 3-dimensional a structure of KP12

The molecular mass of KP12 is predicted to be 11.4 kDa, and a pI = 7.79. The *Intensive* model run by Phyre2 predicted 92% of KP12 residues (with 100 % confidence) that the match is truly homologous to a 2 domain Kazal serpin rhodniin (c1tbqS) from the assassin bug (*Rhodnius prolixus*). The 3D image is of KP12 is available on page 80, figure 22.

4.6 Discussion

Theoretically, recombinant protein synthesis is straightforward. A gene of interest is amplified, purified and cloned into a plasmid vector. The plasmid vector is transformed into an expression host and expression of the protein is induced. The cells are harvested, the fusion protein is purified from the cell lysate, and the protein of interest is cleaved from its fusion tag before being isolated in a pure and functional state to complete protein characterisation and inhibitory assays. In practice, the dynamic form and function proteins means many of the individual steps of the process must be optimised for each unique protein. The vector system (i.e. plasmid), expression host, culture media, method of overexpression induction, method of cell lysis, the protein purification system and the method of concentrating the protein each require considered forethought or alternatively must be optimised. These factors will be further discussed in relation to the workflow and results of this project while identifying a combination of methods described in the literature. Finally, recommendations will be made to maximise KP12 overexpression and yield for future studies.

In this study, a gene for the serpin KP12 was amplified from white-clawed crayfish *Austropotamobius pallipes*, using cDNA, the *kp12* gene was successfully cloned into the expression vector pMAL-p4X; containing a tac promoter, maltose binding protein fusion tag and ampicillin resistance gene. The plasmid construct was then transformed and overexpressed in an *E. coli* BL21 (DE3) host, cultured in autoinduction media. After cell lysis and affinity chromatography (cross-linked Amylose Resin column), this workflow yielded 11.4 mg of the fusion protein. Next, 5.4 mg of fusion protein was cleaved with Factor Xa, and the overnight incubation reaction was concentrated by spin column centrifugation. The cleaved MBP/KP12 reaction was purified by size exclusion chromatography and eluted KP12 fractions were then concentrated by spin column centrifugation. However, the final yield of KP12 protein required to complete functional studies was not achieved (final yield = 0.07 mg KP12). This loss of KP12 during the final concentration step was a

consistent issue that was not resolved during the project. The search for the lost protein included pooling and concentrating five fractions pre and post the KP12 chromatogram peak (Figure 21), but no increase in KP12 was observed. A further effort was to dislodge any protein bound to the concentrator; the membrane was rinsed thoroughly by pipetting but with no increase in KP12. Another attempt was made to clear the membrane using a vortex, again with no change in final KP12 quantity. The following discussion will compare the methods used in this study in with past serpin overexpression studied and recommendations will be proposed to change the present workflow which will enhance future overexpression studied with KP12.

4.6.1 Recombinant plasmid construction

To study the KP12 serpin from *A. pallipes*, the gene of interest had first to be identified in the genome. However, though *A. pallipes* has been studied broadly at an ecological level, the details of their phylogeny and biogeography have been constructed recently (Jelić et al. 2016). These data required for population genetic studies are generally limited to microsatellites which are non-coding nuclear markers, and fragments of the universal mitochondrial molecular markers COI and 16S rRNA, and therefore limited for functional protein studies. Moreover, only 49 protein sequences for *A. pallipes* are available on the UniProt protein repository (20/09/18), of these 48 are for mitochondrial genes, mostly COI or its subunits, and one is for Histone H3. This is a serious limiting factor when aiming to work on genes that have not yet been confirmed in the species. One might next turn to standard biochemistry methods to confirm the protein is indeed expressed in *A. pallipes*. However, these techniques require large quantities of crayfish haemolymph; for example, Johansson et al. (1994) sacrificed 170 *P. leniusculus* specimens for the purification of Kazal type protease inhibitor, and *P. leniusculus* are themselves much larger than *A. pallipes*. As a protected species, this method is not an option.

Newer approaches such as transcriptomics and proteomics can identify and sequence the gene transcripts and proteins present in a sample, *de novo*. However, though the price of these techniques is continually decreasing, they remain prohibitively expensive. The cross-amplification of genes using primers designed from closely related species has been shown to be an effective method to amplify speculative gene targets; Migdał et al. (2016) cross-amplified coat colour genes between several species of rodent, and Jaramillo-Correa et al. (2003) cross-amplified mtDNA between conifer species. Although effective, this method does increase uncertainty into the study of the subsequent gene and protein. As the primer sequence is incorporated into the new amplicon, any polymorphisms in the primer region of the gene of interest will be replaced with the cross-amplification primer sequence, which may potentially change the structure of the protein.

Notwithstanding, the method is cost-effective and accessible, and for this study the Pl_KP12 forward and reverse primers designed by Donpudsa et al. (2010a) in the signal crayfish *P. leniusculus* successfully amplified the homologous gene *kp12* from *A. pallipes*. Additional genes were also amplified using this technique including Ap_KP18 a serpin designed alongside Pl_KP12, and an antimicrobial peptide (Ap_Crustin1) with primers again designed for *P. leniusculus* (Donpudsa et al. 2010a). The success in cross-amplifying these genes suggests this method may be applicable to a whole range of conserved genes between the two species.

With the target identified in the species of interest, the next consideration was the selection of an appropriate vector plasmid. The choice of a pMAL-p4X expression vector system had several advantages for the overexpression of KP12. The vector has a replicon resulting in a low copy number of plasmids within the host, and this is beneficial as a greater plasmid count can cause stress to the host and inhibit cell growth (Birnbaum and Bailey 1991). The *tac* promoter is also of benefit as it is considerably stronger than its *lac* derivative and induces stable transcription of *kp12* when the cell culture has reached a high density (De Boer et al. 1983). Moreover, the pMAL system has been successfully utilised for the overexpression

of serpins in previous studies, such as a chymotrypsin-like protease inhibiting serpin (Schick et al. 1997).

A further important consideration regarding the vector system is the drug resistance gene. An ampicillin resistance marker is the standard antibiotic resistance used in overexpression studies and is utilised in pMAL-p4X vector system. A significant drawback to ampicillin is that its function depletes overtime; a periplasmic enzyme produced in *E. coli* (β-lactamase) inactivates β-lactam antibiotics such as ampicillin. Hence, in the initial culture ampicillin was in a concentration sufficient to ensure only target host cells could grow. However, β-lactamase is continually expressed in *E. coli* and after several hours the antibiotic resistance inferred from ampicillin is functionally diminished (Korpimäki et al. 2003). Though overexpression seemed to yield a substantial quantity of fusion protein (Figure 9, Table 4), amp100 was only added at the start of the 60-hour incubation, and there is potential for contamination during the prolonged shaking incubation. Furthermore, plasmids are metabolically expensive for a host cell to maintain and they can be lost over time without evolutionary pressure to preserve them, such as ampicillin, (Smith and Bidochka 1998, Baneyx 1999). To overcome this, Sivashanmugam (2009) suggest inoculation of the culture media with amp100 at regular intervals, and this is a factor that should be addressed following on from this study.

The inclusion of MBP as the fusion tag has several benefits; the fusion tag is detectable by immunoblotting to verify expression, the tag enhances the solubility of the fusion protein (Kapust and Waugh 1999), and has intrinsic chaperone activity that can improve protein folding (Raran-Kurussi and Waugh 2012). A potential issue with the pMAL-p4x system is that the MBP fusion protein is cleaved at a Factor Xa cleavage site. Factor Xa is generally an efficient protease but can cause non-specific or promiscuous cleavage, and it is recommended the protease be optimised to identify the lowest concentration of protease required for maximum cleavage of a given fusion protein (Block et al. 2015). A further consideration in

using Factor Xa is the expense of the protease. The optimal cleavage conditions for MBP_KP12 using Factor Xa were found to be 0.2x protease concentration incubated at 23°C for >10hrs (Figure 19) and non-specific cleavage was not observed. If an expense or promiscuous cleavage was observed, then an alternative cleavage protease should be utilised. The protease TEV, for example, has considerably greater cleavage specificity, is more cost-effective and is relatively simple to overexpress and purify in one's own lab (Van Den Berg et al. 2006). However, TEV is used less commonly in commercial vector systems, and pMAL/TEV plasmids have been discontinued by some vendors such as NEB (Van Den Berg et al. 2006). Alternatively, one could modify the Factor Xa cleavage site in pMAL-p4X to a TEV site by site-directed mutagenesis (Hallak et al. 2017, Zeng et al. 2018).

4.6.2 Overexpression of the fusion protein

Consideration should also be given to the host system. *E. coli* has been dubbed the workhorse of protein overexpression, with countless strains of *E. coli* now available, each with advantages and drawbacks (Bird et al. 2004). *E. coli* is convenient and economical to culture to high densities, and its metabolic processes have been studied in detail. A vast range of recombinant serpins have been overexpressed and purified using the host system such as; plasminogen inhibitors (Zhou et al. 1997), α1-antitrypsins (Hopkins et al. 1993, Zhou et al. 2001) and chimeric serpins (Bottomley and Stone 1998), to name a few.

Additionally, there is no report of serpin induced toxicity in the host, though the system is not without its limitations. Considering the success of the host in expressing serpins in past studies, *E. coli* BL21 (DE3) was chosen as the host for this project. Albeit a proven host in overexpression studies, there are potential limiting factor regarding this host system and consideration should be given to codon bias. As the *E. coli* proteome is low in arginine, proline, leucine and isoleucine, the host only generates a low number of tRNAs coding for their associated amino acids

(Bird et al. 2004), this can result in low expression of proteins rich in these residues. KP12 does contain a number of these amino acids throughout the mature peptide (Figure 22). However, whether codon bias affects KP12 synthesis in BL21 (DE3) would require comparative expression in another host. Strains such as Rosetta gami B (DE3) contain tRNA encoded plasmids to provide a consistent supply of these amino acids during peptide synthesis. The work by Donpudsa et al. (2010a) expressed soluble KP12 and KP18 from *P. leniusculus* in Rosetta gami B (DE3) and purified soluble and functional protein to complete inhibition and kinetics assays. Donpudsa et al. (2010b) again used the same host to overexpress two functional antimicrobial peptides, Crustin1 and Crustin2. Moreover, a range of MBP fusion serpins have been successfully overexpressed in Rosetta gami B (DE3) from a number of species, including; bovine cathepsin L (Hwang et al. 2005), Protein C Inhibitor from humans (Li et al. 2007) and Aeropin from *Pyrobaculum aerophilum* (Cabrita et al. 2007). Therefore, a change from BL21 (DE3) to Rosetta gami B (DE3) could be a simple and effective way to enhance protein synthesis. A further reported protocol that can enhance fusion protein synthesis is the addition of ethanol to the host culture, and a small-scale experiment was carried out. The results showed ethanol initially appears to inhibit cell growth (Figure 9, Figure 14). However, a direct comparison between the timescales cannot be inferred as the samples were not standardised to a uniform optical density (e.g. O.D. 1). This experiment should be repeated, and the samples should be normalised to a uniform O.D.

In this study, the yield of purified MBP_KP12 from BL21 (DE3) equated to 3.8 mg/L. In comparison, Bottomley and Stone (1998) yielded 1 mg/L of a chimeric serpin fusion protein from BL21 (DE3), Schick et al. (1997) yielded 3 mg/L of the serpin Antigen 2 fusion protein from BL21 (DE3) and Blacque et al. (2002) purified 5 mg/L of a tumour suppressor serpin fusion protein in the same host. Hence, the yield for this project appears relatively good. However, a direct comparison between these studies should be interpreted cautiously as different vector systems were used in

each study, and the effects of their promoters, fusion peptides, culture conditions and subsequent purifications will have affected final yield.

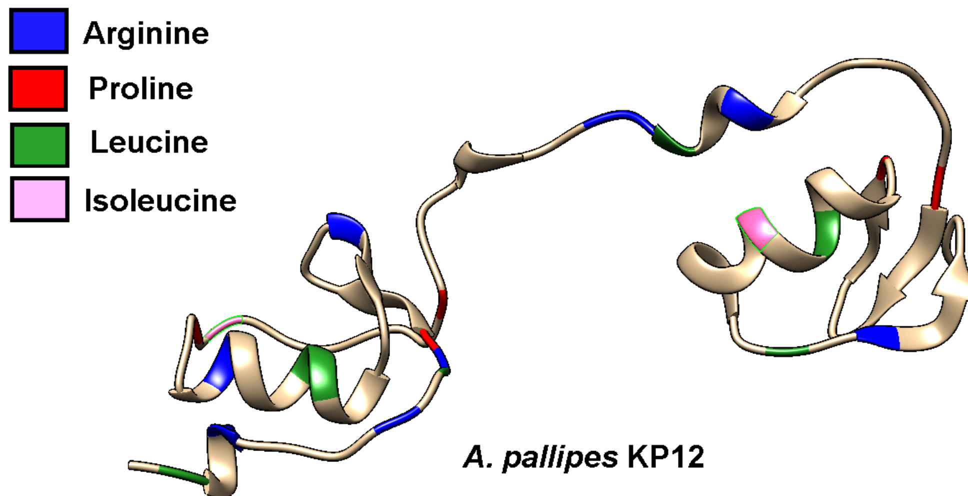


Figure 22. Distribution of arginine, proline, leucine and isoleucine in the serpin KP12 from *Austropotamobius pallipes*, created as per 4.4.5.

The next consideration is overexpression induction. The dominant method referenced in the literature is the addition of IPTG to the culture at mid-log growth phase (optical density 600 nm = 0.5 – 0.8). IPTG shares structural homology with lactose which is the molecule that natively binds to the *lac* promoter of the plasmid vector; this initiates transcription of the fusion gene and ampicillin mRNA transcripts. The primary limiting factor when using IPTG is the cost (€75/g, Sigma-Aldrich 22/09/18), and alternatives methods, such as autoinduction media (AIM), do not require the addition of reagents to the culture (Studier 2014). The glucose and glycerol contained in AIM more readily bind to the vectors *lac* promoter than lactose which inhibits transcription initiation (Grossman et al. 1998). The *E. coli* host metabolises these sugars before lactose, and after their depletion lactose will bind to the *lac* promoter and promote transcription of the plasmid. This delay enables the culture to reach a mid-log phase before transcription is initiated. AIM is a cost-effective method and was shown in this study to be a successful method of overexpression of the serpin KP12. However, AIM is also not without its limitations, the culture incubation is considerably prolonged (24 – 60 hours) and should be

optimised. This extended incubation is particularly important when considering antibiotic resistance, as discussed previously the culture should be inoculated regularly with ampicillin.

4.7 Protein purification

Once the fusion/protein of interest was appropriately overexpressed in a host system, the fusion protein was purified from native cellular proteins. This purification process may also require optimisation at several stages, including cell lysis, fusion protein purification, fusion protein cleavage and concentration of the isolated target protein. For this study, cell lysis followed a standard workflow. Harvested cells were resuspended in lysis buffer with lysozyme to inhibit native *E. coli* proteases from degrading the fusion protein, followed by sonication and centrifugation (Pearce and Cabrita 2011). Following lysis, soluble protein remains in the lysate, and insoluble protein is found in the pellet. The MBP fusion tag used in this study enhances the solubility of the fusion protein (Kapust and Waugh 1999, Esposito and Chatterjee 2006), and Donpudsa et al. (2010a) found that Pl_KP12 and Pl_KP18 were readily soluble molecules, purifying the fusion protein from lysate only. If low yields of soluble protein are observed in the lysate, then purification from the lysed pellet could be considered, though this process requires extensive recovery, denaturation and refolding steps that do not guarantee efficient recovery of active protein (Bird et al. 2004). However, purifying the fusion protein by fusion protein by affinity chromatography did yield a significant quantity of material, a total of 11.4 mg of MBP-KP12 was collected in the elution fractions.

In the upscaled Factor Xa cleavage reaction 5.4 milligrams the fusion protein was digested and incubated overnight, the step to inactivate Factor Xa with Dansyl-glu-gly-arg-chloromethyl ketone was omitted from this study, and the reaction mixture was concentrated by spin column centrifugation to 2 mL final volume. The concentrated reaction mixture was then loaded onto the size exclusion column, and the eluted KP12 fractions were pooled and concentrated by spin column

centrifugation. However, significantly reduced the quantity of KP12 was quantified post gel filtration (0.077 mg) and this significant loss of KP12 was not resolved during this project. Some hypotheses addressing this were: quantifying the pre and post-elution fraction to see whether the ÄKTAprime and PC readout were synchronised, but no protein was found in the surround fractions. Nor was there protein present in the spin column filtrate. Another thought was that protein (11.5 kDa) may be bound to the 10 kDa cut-off filter membrane but attempts to vigorously dislodge any bound KP12 did not increase yield. Concentration was also completed with 3 kDa cut-off filters with the same result. As size exclusion chromatography is a standard and reliable technique (Kågedal et al. 1991), the issue here likely occurs during the upscaled Factor Xa cleavage or during the concentration of the reaction mixture. Some change to this workflow going forward that may improve the final KP12 yield would be to inactivate the Factor Xa protease. As the 15 mL cleave reaction is concentrated to 2 mL, the protease concentration increases considerably and may begin non-specifically digesting the molecules. An additional change to this workflow would be to change the post-cleavage concentration method. The concentrator membranes used had a 10 or 3 kDa cut-off size, and no protein was quantifiable in the flow through. A further thought was that the 11.5 kDa KP12 was bound to the spin column membrane, but several attempts to rinse the membrane by pipetting and vortex were unsuccessful. An alternative method of concentration at this final stage could be such as dialysis, though the Snake Skin dialysis method did not provide good results when concentrating the fusion protein other devices such as Tube-O-DIALYZER™ could be considered. Given the large volume of the cleavage reaction, a method such as protein precipitation by salting out could also be considered. However, this technique can alter the structure and activity of the target protein.

4.8 Conclusion and suggestions

Milligram quantities of soluble functional protein are required to complete characterisation and functional studies, and overexpression is an invaluable technique of generating these large quantities recombinant proteins. This study has shown that *E. coli* is a suitable host for the overexpression of the serpin KP12. Taking the results obtained in this study into account, the suggested workflow for overexpressing KP12 is as follows (alterations are italicised):

- Cross amplify KP12
- Ligate into the vector system *pET His6 MBP TEV*
- Transform into host *Rosetta gami (DE3)*
- Overexpress in autoinduction media, inoculate with ampicillin (*periodically*)
- Harvest cell (up to 60 hours)
- Follow standard cell lysis protocol (buffer, lysozyme, sonication)
- Purify the fusion protein from cell lysate by affinity chromatography
- Concentrate the fusion protein with 10 kDa spin columns
- Digest the fusion protein with TEV protease (degrades after 20 hours)
- Isolate the protein of interest by size exclusion chromatography
- Concentrate the purified protein of interest by 3 kDa spin column
    - Alternatively, *dialyse the sample*
    - Alternatively, precipitate the protein by *salting out*

5 Chapter 4: Implication of this thesis & concluding recommendations

Chapter 2: By implanting a higher resolution molecular marker than previous studies, this project has identified a level of population genetic structure within Irish *A. pallipes* populations, which have until now been characterised as a genetically homogeneous populations (Gouin et al. 2001, Reynolds et al. 2002b). The current methodological approaches suggested by Souty-Grosset and Reynolds (2009) for conservation and restocking procedures regarding *A. pallipes* explicitly refers to Irish crayfish stocks as very uniform. However, the authors also described how analysis of the 16S gene has been appropriately implemented to identify genetic diversity in Italian populations (Bertocchi et al. 2008). It is therefore recommended that findings from this study should be taken into account in future management strategies to protect the species in Ireland, while additional molecular markers should be screened to provide a higher resolution characterisation of crayfish genetic structure in Ireland. Several conservation efforts in Europe have taken genetic diversity into account, such as the identification of several suitable locations for restocking in Granada, Spain. The study describes the genetic diversity of *A. pallipes* within the region and the importance to maintain the genetic structure when restocking the identified sites (Gil-Sanchez et al., 2002).

In Ireland, very few restocking events have been recorded, which has been attributed to crayfish receiving protected status (Reynolds, 1997). However, management plans have been developed (Reynolds et al. 2002) and do incorporate the necessity of maintaining genetic diversity when restocking. However, the plans do state Irish populations of crayfish are rather homogenous and suggest using RAPD as a method to determine genetic diversity. Moreover, Reynolds et al. (2000) describes the restocking procedure of White Lough in county Westmeath, where crayfish were wiped out in the 1980s. The project thoroughly incorporated genetic concerns for the restocking plan. However, this project used the genetic analyses for Souty-Grosset et al. (1997) and Gouin et al. (2003) which did not identify any

significant genetic diversity within Irish crayfish population (using RAPD markers). The restocking project was successful, and an abundant stock has since been recorded in the lake (Gammell et al., unpublished). Considering these results, future studies should include White Lough as a sampling location regarding genetic analyses, as a successfully restocked lake it would generate an interesting and unique dataset for the species at a European level.

Chapter 3: This study has successfully overexpressed a recombinant serpin KP12 from *A. pallipes,* that appears to be the first protein recombinantly overexpressed from the species. To date, host-pathogen interaction studies involving A. astaci and astacid species have been focused on commercially valuable and invasive species such as *P. leniusculus* and *P. clarkia.* The optimisation of the overexpression protocol of KP12, this project represents an important first step for future continued work to elucidate host-pathogens in this vulnerable species. Furthermore, this is a good method for production of KP12 but could be applied to additional different protease inhibitors such as KP18, the sequence for which was also amplified during this project.

Once the concentration of purified KP12 has been optimised, inhibitory assays can be completed against specific proteases isolated from the plague pathogen *A. astaci*. Donpudsa et al. (2010a) did complete inhibitory assays with KP12 and KP18 against extracellular serine proteinases from *A. astaci*. However, these proteinases were filtered from culture media and the specific proteases included in the assay were not identified. Two serine proteinases from *A. astaci* were identified and overexpressed by Bangyeekhun et al. (2001) but no specific inhibitory assays between the pathogens proteinases and crayfish inhibitors have been completed but would be an excellent future direction for research following on from this study.

This project represents an important contribution to *A. pallipes* research in Ireland and Europe. Genetic research has been lacking on the species since the work of Gouin et al. (2003), which characterised Irish populations as having a homogeneous

genetic structure. Though this characterisation was appropriate for the data generated from that study, that conclusion may have contributed to the lack of research interest, e.g. implementing further molecular markers across the species range. The species is still well distributed across Ireland but considering the recent mass mortality events caused by the crayfish plague, a thorough characterisation of *A. pallipes*' genetic structure should be completed before further populations are permanently lost to the virulent pathogen *A. astaci*.

6 Appendix 1

6.1 Troubleshooting and optimisation of 16S Sanger sequencing

While completing these studies both the population genetics and protein overexpression lab work had significant troubleshooting and optimisation stages. Regarding protein overexpression, optimisation is required at potentially every step, and this is discussed in detail in chapter 3. Regarding the population genetics work, preliminary analyses of three molecular markers was completed (16S, COI and microsatellites). These data indicated COI and microsatellites lacked much variation between populations and 16S was most variable between populations, and so it was decided to focus the project's resources on one marker that would access preliminary population structure across Ireland. This choice could be criticised for the fact that a single marker provides a lower resolution picture of the overall genetic structure. However, this masters project is part of a greater research group, and by focusing on 16S, this thesis can now state preliminary findings suggesting population structure does exist between Irish crayfish populations. Such a result warrants further study, and even though COI and microsatellites showed initially low levels of variation, these markers should be further investigated over a wider-ranging sample distribution. This would bring Irish crayfish population genetics research up to the level found in most European countries.

The greatest troubleshooting issue regarding obtaining good quality sequence data (i.e. signal drop-off) was unavoidable and arose during the sequencing step. Signal drop-off can occur during Sanger sequencing for several reasons, such as a low primer or amplicon concentration in the reaction. A further cause of signal loss during sequencing is repetitive stretches of single nucleotides (mononucleotide stretches). When present in a sequence these regions can cause the signal to slip, causing frequent double peaks or signal drop-off. *A. pallipes* has two mononucleotide stretches 50 bp apart (Figure 23). With the sequences dropping

off at around 250/300 bp these sequences were too short to be included in the data analyses.

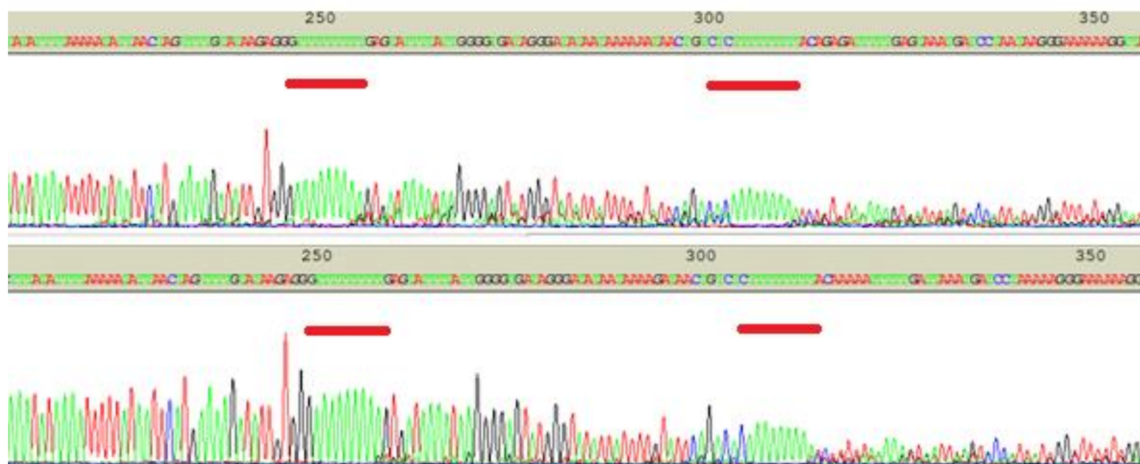


Figure 23. Chromatogram showing Sanger sequencing drop-off and double peaks in 16S after mononucleotide stretches in *Austropotamobius pallipes*.

GATC (the sequencing company) state sequence slippage is a known limitation of Sanger sequencing and chemistries are available to try and overcome the issue (pers. comm with GATC customer service). However, these chemistries are considerably more expensive than standard sequencing, and the company declined to provide their specific details. Further solutions were sought in the literature, but no specific recommendations were found regarding mononucleotide induced slippage. In searching science forums, a suggestion to include DMSO to 1% of the final reaction volume was found, this seemed to improve sequencing results. However, at this stage of the project too few sequencing reactions remained to conclude whether DMSO did improve the results. For several samples, sequencing was also completed using both the forward and reverse primer, which did provide some additional sequence data. However, where slippage occurred on the first mononucleotide stretch for both sequencing reactions using forward and reverse primer, then the 50 bp region between the mononucleotide stretches was still missing, and as a result, two sequencing reactions were lost.

A considerable quantity of resources were invested in the 16S marker, for a low return rate. However, the results provide exciting findings. Moreover, as these data are particularly difficult to obtain, it makes the dataset inherently more valuable. Going forward, the best course of action would be to purchase the more expensive sequencing service. However, many samples are required for thorough population genetics analyses, and this additional expense is likely prohibitive. A good option would investigate the effect of DMSO on sequencing in greater detail.

6.2 Unedited SDS-PAGE gel

The unedited gel image used to generate a composite summary gel for section 4.5.6.1.1*Upscaling Factor Xa cleavage*

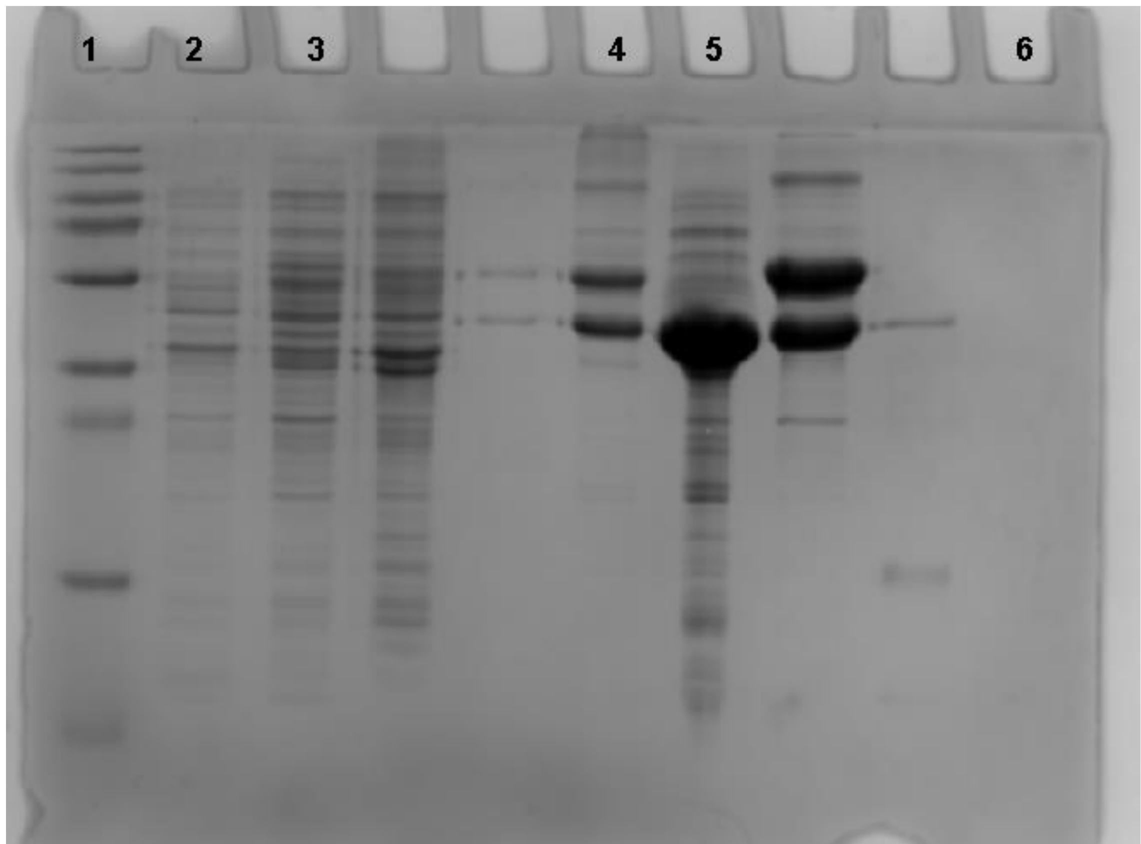


Figure 24. The unedited gel image used to generate a composite summary gel for figure 20. Upscaling Factor Xa cleavage. The lanes used to generate the composite image are numbered above.

7 References

Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P., 2013. *Essential Cell Biology.* Garland Science.

Albrecht, H. & Von Hagen, H.O., 1981. 'Differential weighting of electrophoretic data in crayfish and fiddler crabs (decapoda: Astacidae and ocypopidae)'. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry,* 70 (3), pp. 393-399.

Allan, J.D. & Flecker, A.S., 1993. 'Biodiversity conservation in running waters'. *BioScience,* 43 (1), pp. 32-43.

Andréanne, D. & Reynolds, J.D., 2003. 'The distribution of the white-clawed crayfish, *Austropotamobius pallipes*, in eight catchments in Ireland in relation to water quality'. *Management & Conservation of Crayfish*, p. 94.

Aspán, A., Hall, M. & Söderhäll, K., 1990. 'The effect of endogeneous proteinase inhibitors on the prophenoloxidase activating enzyme, a serine proteinase from crayfish haemocytes'. *Insect Biochemistry,* 20 (5), pp. 485-492.

Aspán, A. & Söderhäll, K., 1991. 'Purification of prophenoloxidase from crayfish blood cells, and its activation by an endogenous serine proteinase'. *Insect biochemistry,* 21 (4), pp. 363-373.

Attard, J. & Pasteur, N., 1984. 'Variability and genetic differentiation in 5 species of the crayfish Astacidae'. *Biochemical Systematics and Ecology,* 12 (1), p. 109.

Attard, J. & Vianet, R., 1985. 'Variabilité génétique et morphologique de cinq populations de l'écrevisse Européenne *Austropotamobius pallipes* (Lereboullet 1858) (Crustacea, Decapoda)'. *Canadian Journal of Zoology,* 63 (12), pp. 2933-2939.

Avise, J.C., 2017. 'Invertebrate zoology (other than insects)'. In*: From aardvarks to zooxanthellae.* Springer, pp. 81-90.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. & Saunders, N.C., 1987. 'Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics'. *Annual review of ecology and systematics,* 18 (1), pp. 489-522.

Baneyx, F., 1999. 'Recombinant protein expression in *Escherichia coli*'. *Current opinion in biotechnology,* 10 (5), pp. 411-421.

Bangyeekhun, Eakaphun, Lage Cerenius, and Kenneth Söderhäll. "Molecular cloning and characterization of two serine proteinase genes from the crayfish plague fungus, Aphanomyces astaci." *Journal of invertebrate pathology* 77.3 (2001): 206-216.

Baric, S., Höllrigl, A., Füreder, L. & Dalla Via, J., 2005. 'Mitochondrial and microsatellite DNA analyses of *Austropotamobius pallipes* populations in south tyrol (italy) and tyrol (austria)'. *Bulletin Français de la Pêche et de la Pisciculture,* (376-377), pp. 599-612.

Barrett, A.J., 1994. '[1] Classification of peptidases'. In*: Methods in enzymology.* Elsevier, pp. 1-15.

Bernini, G., Bellati, A., Pellegrino, I., Negri, A., Ghia, D., Fea, G., Sacchi, R., Nardi, P.A., Fasola, M. & Galeotti, P., 2016. 'Complexity of biogeographic pattern in the endangered crayfish *Austropotamobius italicus* in Northern Italy: Molecular insights of conservation concern'. *Conservation genetics,* 17 (1), pp. 141-154.

Bertocchi, S., Brusconi, S., Gherardi, F., Grandjean, F. & Souty-Grosset, C., 2008. 'Genetic variability of the threatened crayfish *Austropotamobius italicus* in tuscany (italy): Implications for its management'. *Fundamental and Applied Limnology/Archiv für Hydrobiologie,* 173 (2), pp. 153-164.

Bird, P.I., Pak, S.C., Worrall, D.M. & Bottomley, S.P., 2004. 'Production of recombinant serpins in *Escherichia coli*'. *Methods,* 32 (2), pp. 169-176.

Birnbaum, S. & Bailey, J.E., 1991. 'Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*'. *Biotechnology and bioengineering,* 37 (8), pp. 736-745.

Blacque, O.E. & Worrall, D.M., 2002. 'Evidence for a direct interaction between the tumor suppressor serpin, maspin, and types i and iii collagen'. *Journal of Biological Chemistry,* 277 (13), pp. 10783-10788.

Block, H., Maertens, B., Spriestersbach, A., Kubicek, J. & Schäfer, F., 2015. 'Proteolytic affinity tag cleavage'. In*: Methods in enzymology.* Elsevier, pp. 71-97.

Bode, W. & Huber, R., 1992. 'Natural protein proteinase inhibitors and their interaction with proteinases'. *European Journal of Biochemistry,* 204 (2), pp. 433-451.

Bottomley, S.P. & Stone, S.R., 1998. 'Protein engineering of chimeric serpins: An investigation into effects of the serpin scaffold and reactive centre loop length'. *Protein engineering,* 11 (12), pp. 1243-1247.

Boyce, M.S., 1992. 'Population viability analysis'. *Annual review of Ecology and Systematics,* 23 (1), pp. 481-497.

Brown, G.G., Gadaleta, G., Pepe, G., Saccone, C. & SbisA, E., 1986. 'Structural conservation and variation in the d-loop-containing region of vertebrate mitochondrial DNA. *Journal of molecular biology,* 192 (3), pp. 503-511.

Burton, R.S., 2009. 'Molecular markers, natural history, and conservation of marine animals'. *BioScience,* 59 (10), pp. 831-840.

Cabrita, L.D., Irving, J.A., Pearce, M.C., Whisstock, J.C. & Bottomley, S.P., 2007. 'Aeropin from the extremophile *Pyrobaculum aerophilum* bypasses the serpin misfolding trap'. *Journal of Biological Chemistry,* 282 (37), pp. 26802-26809.

Cerenius, L., Jiravanichpaisal, P., Liu, H.-p. & Soderhall, I., 2010. 'Crustacean immunity'. In*: Invertebrate immunity.* Springer, pp. 239-259.

Cerenius, L., Lee, B.L. & Söderhäll, K., 2008. 'The proPO-system: Pros and cons for its role in invertebrate immunity'. *Trends in immunology,* 29 (6), pp. 263-271.

Cerenius, L. & Söderhäll, K., 2004. 'The prophenoloxidase-activating system in invertebrates'. *Immunological reviews,* 198 (1), pp. 116-126.

Chhetri, G., Kalita, P. & Tripathi, T., 2015. 'An efficient protocol to enhance recombinant protein expression using ethanol in *Escherichia coli*'. *MethodsX,* 2, pp. 385-391.

Chiesa, S., Scalici, M., Negrini, R., Gibertini, G. & Marzano, F.N., 2011. 'Fine-scale genetic structure, phylogeny and systematics of threatened crayfish species complex'. *Molecular phylogenetics and evolution,* 61 (1), pp. 1-11.

Clement, M., Snell, Q., Walker, P., Posada, D. & Crandall, K. TCS: Estimating gene genealogies'. [Online]. 2002. IEEE.

Cornalia, E., 1860. 'Sulla malattia dei gamberi'. *Atti Soc Ital Sci Nat,* 2, pp. 334-336.

Crandall, K.A. & Buhay, J.E., 2008. 'Global diversity of crayfish (Astacidae, Cambaridae, and Parastacidae – Decapoda) in freshwater'. *Hydrobiologia,* 595 (1), pp. 295-301.

Crandall, K.A. & Fitzpatrick Jr, J.F., 1996. 'Crayfish molecular systematics: Using a combination of procedures to estimate phylogeny'. *Systematic Biology,* 45 (1), pp. 1-26.

Dalvit, C., De Marchi, M. & Cassandro, M., 2007. 'Genetic traceability of livestock products: A review'. *Meat Science,* 77 (4), pp. 437-449.

De Boer, H.A., Comstock, L.J. & Vasser, M., 1983. 'The tac promoter: A functional hybrid derived from the *trp* and *lac* promoters'. *Proceedings of the National Academy of Sciences,* 80 (1), pp. 21-25.

Demers, A., Lucey, J., McGarrigle, M.L. & Reynolds, J.D. 'The distribution of the white-clawed crayfish, *Austropotamobius pallipes*, in Ireland'. *Biology and Environment: Proceedings of the Royal Irish Academy*. JSTOR.

Demers, A. & Reynolds, J., 2002. 'A survey of the white-clawed crayfish, *Austropotamobius pallipes* (lereboullet), and of water quality in two catchments of Eastern Ireland'. *Bulletin Français de la Pêche et de la Pisciculture,* (367), pp. 729-740.

Diéguez-Uribeondo, J., Royo, F., Souty-Grosset, C., Ropiquet, A. & Grandjean, F., 2008. 'Low genetic variability of the white-clawed crayfish in the Iberian peninsula: Its origin and management implications'. *Aquatic Conservation: Marine and Freshwater Ecosystems,* 18 (1), pp. 19-31.

Dodson, G. & Wlodawer, A., 1998. 'Catalytic triads and their relatives'. *Trends in biochemical sciences,* 23 (9), pp. 347-352.

Donpudsa, S., Söderhäll, I., Rimphanitchayakit, V., Cerenius, L., Tassanakajon, A. & Söderhäll, K., 2010a. 'Proteinase inhibitory activities of two two-domain kazal proteinase inhibitors from the freshwater crayfish *Pacifastacus leniusculus* and the importance of the p2 position in proteinase inhibitory activity'. *Fish & shellfish immunology,* 29 (5), pp. 716-723.

Donpudsa, S., Rimphanitchayakit, V., Tassanakajon, A., Söderhäll, I. & Söderhäll, K., 2010b. 'Characterization of two crustin antimicrobial peptides from the freshwater crayfish *Pacifastacus leniusculus*'. *Journal of invertebrate pathology,* 104 (3), pp. 234-238.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.-I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.-H., Soto, D. & Stiassny, M.L.J., 2006. 'Freshwater biodiversity: Importance, threats, status and conservation challenges'. *Biological reviews,* 81 (2), pp. 163-182.

Esposito, D. & Chatterjee, D.K., 2006. 'Enhancement of soluble protein expression through the use of fusion tags'. *Current opinion in biotechnology,* 17 (4), pp. 353-358.

Evans, J.D., Aronstein, K., Chen, Y.P., Hetru, C., Imler, J.L., Jiang, H., Kanost, M., Thompson, G.J., Zou, Z. & Hultmark, D., 2006. 'Immune pathways and defence mechanisms in honey bees *Apis mellifera*'. *Insect molecular biology,* 15 (5), pp. 645-656.

Excoffier, L., Smouse, P.E. & Quattro, J.M., 1992. 'Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data'. *Genetics,* 131 (2), pp. 479-491.

Fairley, J.S., 1972. 'Food of otters (*Lutra lutra*) from Co. Galway, Ireland, and notes on other aspects of their biology'. *Journal of Zoology,* 66 (APR), p. 469.

Faulkes, Z., 2015. 'A bomb set to drop: Parthenogenetic marmorkrebs for sale in Ireland, a european location without non-indigenous crayfish'. *Management of Biological Invasions*.

Faulkes, Z., 2017. 'Slipping past the barricades: The illegal trade of pet crayfish in Ireland'. *Biology and Environment: Proceedings of the Royal Irish Academy,* 117B, No. 1, pp., pp. 15-23.

Feld, C.K., Martins da Silva, P., Paulo Sousa, J., De Bello, F., Bugter, R., Grandin, U., Hering, D., Lavorel, S., Mountford, O. & Pardo, I., 2009. 'Indicators of biodiversity and ecosystem services: A synthesis across ecosystems and spatial scales'. *Oikos,* 118 (12), pp. 1862-1871.

Felsenstein, J., 1985. 'Confidence limits on phylogenies: An approach using the bootstrap'. *Evolution,* 39 (4), pp. 783-791.

Ferguson, M.M. & Drahushchak, L.R., 1990. 'Disease resistance and enzyme heterozygosity in rainbow trout'. *Heredity,* 64 (3), pp. 413-417.

Fitzpatrick, B.M., Fordyce, J.A. & Gavrilets, S., 2008. 'What, if anything, is sympatric speciation?'. *Journal of evolutionary biology,* 21 (6), pp. 1452-1459.

Frankham, R., 1995. 'Conservation genetics'. *Annual review of genetics,* 29 (1), pp. 305-327.

Frankham, R., 1999. 'Quantitative genetics in conservation biology'. *Genetics Research,* 74 (3), pp. 237-244.

Frankham, R., 2003. 'Genetics and conservation biology'. *Comptes Rendus Biologies,* 326, pp. 22-29.

Frankham, R., Bradshaw, C.J.A. & Brook, B.W., 2014. 'Genetics in conservation management: Revised recommendations for the 50/500 rules, red list criteria and population viability analyses'. *Biological Conservation,* 170, pp. 56-63.

Fratini, S., Zaccara, S., Barbaresi, S., Grandjean, F., Souty-Grosset, C., Crosa, G. & Gherardi, F., 2005. 'Phylogeography of the threatened crayfish (genus *Austropotamobius*) in Italy: Implications for its taxonomy and conservation'. *Heredity,* 94 (1), p. 108.

Frost, W.E. 'R. Liffey survey iv. The fauna of the submerged" Mosses" In an acid and an alkaline water'. [Online]. 1941. JSTOR.

Fu, Y.-X., 1997. 'Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection'. *Genetics,* 147 (2), pp. 915-925.

Füreder, L., Gherardi, F., Holdich, D., Reynolds, J., Sibley, P. & Souty-Grosset, C., 2010. 'Austropotamobius pallipes'. *IUCN 2010: IUCN Red List of Threatened Species. Version 2010.4*.

Füreder, L. & Reynolds, J.D., 2003. 'Is *Austropotamobius pallipes* a good bioindicator?'. *Bulletin Français de la Pêche et de la Pisciculture,* (370-371), pp. 157-163.

Garrett, M., Fullaondo, A., Troxler, L., Micklem, G. & Gubb, D., 2009. 'Identification and analysis of serpin-family genes by homology and synteny across the 12 sequenced drosophilid genomes'. *BMC genomics,* 10 (1), p. 489.

Gil-Sanchez, Jose M., and Javier Alba-Tercedor. "Ecology of the native and introduced crayfishes *Austropotamobius pallipes* and *Procambarus clarkii* in southern Spain and

implications for conservation of the native species." Biological Conservation 105.1 (2002): 75-80.

Gouin, N., Grandjean, F., Bouchon, D., Reynolds, J.D. & Souty-Grosset, C., 2001. 'Population genetic structure of the endangered freshwater crayfish *Austropotamobius pallipes*, assessed using rapd markers'. *Heredity,* 87 (1), pp. 80-87.

Gouin, N., Grandjean, F., Pain, S., Souty-Grosset, C. & Reynolds, J., 2003. 'Origin and colonization history of the white-clawed crayfish, *Austropotamobius pallipes*, in ireland'. *Heredity,* 91 (1), pp. 70-77.

Gouin, N., Grandjean, F. & Souty-Grosset, C., 2006. 'Population genetic structure of the endangered crayfish *Austropotamobius pallipes* in France based on microsatellite variation: Biogeographical inferences and conservation implications'. *Freshwater Biology,* 51 (7), pp. 1369-1387.

Gouin, N., Souty-Grosset, C., Ropiquet, A. & Grandjean, F., 2002. 'High dispersal ability of *Austropotamobius pallipes* revealed by microsatellite markers in a french brook'. *Bulletin Français de la Pêche et de la Pisciculture,* (367), pp. 681-689.

Grandjean, F., Bouchon, D. & Souty-Grosset, C., 2002a. 'Systematics of the european endangered crayfish species *Austropotamobius pallipes* (decapoda: Astacidae) with a re-examination of the status of austropotamobius berndhauseri'. *Journal of Crustacean Biology,* 22 (3), pp. 677-681.

Grandjean, F., Frelon-Raimond, M. & Souty-Grosset, C., 2002b. 'Compilation of molecular data for the phylogeny of the genus *Austropotamobius*: One species or several?'. *Bulletin Francais de la Peche et de la Pisciculture,* (367), pp. 671-680.

Grandjean, F., Gouin, N., Frelon, M. & Souty-Grosset, C., 1998. 'Genetic and morphological systematic studies on the crayfish *Austropotamobius pallipes* (Decapoda: Astacidae)'. *Journal of Crustacean Biology,* 18 (3), pp. 549-555.

Grandjean, F. & Souty-Grosset, C., 2000. 'Mitochondrial DNA variation and population genetic structure of the white-clawed crayfish, A*ustropotamobius pallipes pallipes*'. *Conservation genetics,* 1 (4), pp. 309-319.

Grandjean, F., Souty-Grosset, C., Raimond, R. & Holdich, D., 1997. 'Geographical variation of mitochondrial DNA between populations of the white-clawed crayfish *Austropotamobius pallipes*. *Freshwater Biology,* 37 (2), pp. 493-501.

Grandjean, F., Tan, M.H., Gan, H.Y., Gan, H.M. & Austin, C.M., 2015. 'The complete mitogenome of the endangered white-clawed freshwater crayfish *Austropotamobius pallipes* (lereboullet, 1858)(crustacea: Decapoda: Astacidae)'. *Mitochondrial DNA*, pp. 1-2.

Grossman, T.H., Kawasaki, E.S., Punreddy, S.R. & Osburne, M.S., 1998. 'Spontaneous camp-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability'. *Gene,* 209 (1), pp. 95-103.

Hallak, L.K., Berger, K., Kaspar, R., Kwilas, A.R., Montanaro, F. & Peeples, M.E., 2017. 'Efficient method for site-directed mutagenesis in large plasmids without subcloning'. *PloS one,* 12 (6), p. e0177788.

Harrison, R.G., 2012. 'The language of speciation'. *Evolution: International Journal of Organic Evolution,* 66 (12), pp. 3643-3657.

Hedstrom, L., 2002. 'Serine protease mechanism and specificity'. *Chemical reviews,* 102 (12), pp. 4501-4524.

Hereford, J., 2009. 'A quantitative survey of local adaptation and fitness trade-offs'. *The American Naturalist,* 173 (5), pp. 579-588.

Holdich, D., 2002. 'Distribution of crayfish in europe and some adjoining countries'. *Bulletin Français de la Pêche et de la Pisciculture,* (367), pp. 611-650.

Holdich, D.M., 2003. 'Crayfish in europe–an overview of taxonomy, legislation, distribution, and crayfish plague outbreaks'. *Management & Conservation of Crayfish*, pp. 15-34.

Holdich, D.M. & Lowery, R.S., 1988. 'Freshwater crayfish: Biology, management and exploitation'.

Hopkins, P.C.R., Carrell, R.W. & Stone, S.R., 1993. 'Effects of mutations in the hinge region of serpins'. *Biochemistry,* 32 (30), pp. 7650-7657.

Hwang, S.-R., Stoka, V., Turk, V. & Hook, V.Y.H., 2005. 'The novel bovine serpin endopin 2c demonstrates selective inhibition of the cysteine protease cathepsin l compared to the serine protease elastase, in cross-class inhibition'. *Biochemistry,* 44 (21), pp. 7757-7767.

Häll, L. & Söderhäll, K., 1982. 'Purification and properties of a protease inhibitor from crayfish hemolymph'. *Journal of Invertebrate Pathology,* 39 (1), pp. 29-37.

Irving, J.A., Cabrita, L.D., Rossjohn, J., Pike, R.N., Bottomley, S.P. & Whisstock, J.C., 2003. 'The 1.5 å crystal structure of a prokaryote serpin: Controlling conformational change in a heated environment'. *Structure,* 11 (4), pp. 387-397.

Jaramillo-Correa, J.P., Bousquet, J., Beaulieu, J., Isabel, N., Perron, M. & Bouillé, M., 2003. 'Cross-species amplification of mitochondrial DNA sequence-tagged-site markers in conifers: The nature of polymorphism and variation within and among species in picea'. *Theoretical and Applied Genetics,* 106 (8), pp. 1353-1367.

Jelić, M., Klobučar, G.I.V., Grandjean, F., Puillandre, N., Franjević, D., Futo, M., Amouret, J. & Maguire, I., 2016. 'Insights into the molecular phylogeny and historical biogeography of the white-clawed crayfish (Decapoda, Astacidae)'. *Molecular phylogenetics and evolution,* 103, pp. 26-40.

Jiravanichpaisal, P., Lee, B.L. & Söderhäll, K., 2006. 'Cell-mediated immunity in arthropods: Hematopoiesis, coagulation, melanization and opsonization'. *Immunobiology,* 211 (4), pp. 213-236.

Johansson, M.W., Keyser, P. & SÖDerhÄLl, K., 1994. 'Purification and cDNA cloning of a four-domain kazal proteinase inhibitor from crayfish blood cells'. *European Journal of Biochemistry,* 223 (2), pp. 389-394.

Kanost, M.R., 1999. 'Serine proteinase inhibitors in arthropod immunity'. *Developmental & Comparative Immunology,* 23 (4-5), pp. 291-301.

Kapust, R.B. & Waugh, D.S., 1999. '*Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused'. *Protein Science,* 8 (8), pp. 1668-1674.

Karaman, M.S., 1963. 'Studie der astacidae (Crustacea, Decapoda) ii. Teil'. *Hydrobiologia,* 22 (1-2), pp. 111-132.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E., 2015. 'The phyre2 web portal for protein modeling, prediction and analysis'. *Nature protocols,* 10 (6), p. 845.

Kennedy, M. & Fitzmaurice, P. 'Growth and food of brown trout *Salmo trutta* (L.) in Irish waters'. [Online]. 1971. JSTOR.

Kim, K.S. & Sappington, T.W., 2013. 'Microsatellite data analysis for population genetics'. In: *Microsatellites.* Springer, pp. 271-295.

Kimura, M., 1991. 'The neutral theory of molecular evolution: A review of recent evidence'. *The Japanese Journal of Genetics,* 66 (4), pp. 367-386.

Korpimäki, T., Kurittu, J. & Karp, M., 2003. 'Surprisingly fast disappearance of β-lactam selection pressure in cultivation as detected with novel biosensing approaches'. *Journal of microbiological methods,* 53 (1), pp. 37-42.

Kouba, A., Petrusek, A. & Kozák, P., 2014. 'Continental-wide distribution of crayfish species in europe: Update and maps'. *Knowledge and Management of Aquatic Ecosystems,* (413).

Kraut, J., 1977. 'Serine proteases: Structure and mechanism of catalysis'. *Annual review of biochemistry,* 46 (1), pp. 331-358.

Kumar, S., Stecher, G. & Tamura, K., 2016. 'Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets'. *Molecular biology and evolution,* 33 (7), pp. 1870-1874.

Kågedal, L., Engström, B., Elelgren, H., Lieber, A.-K., Lundström, H., Sköld, A. & Schenning, M., 1991. 'Chemical, physical and chromatographic properties of superdex 75 prep grade and superdex 200 prep grade gel filtration media'. *Journal of Chromatography A,* 537, pp. 17-32.

Largiadèr, C.R., Herger, F., Lörtscher, M. & Scholl, A., 2000. 'Assessment of natural and artificial propagation of the white-clawed crayfish (*Austropotamobius pallipes* species complex) in the alpine region with nuclear and mitochondrial markers'. *Molecular Ecology,* 9 (1), pp. 25-37.

Laskowski Jr, M. & Kato, I., 1980. 'Protein inhibitors of proteinases'. *Annual review of biochemistry,* 49 (1), pp. 593-626.

Lawson, L.P., Bates, J.M., Menegon, M. & Loader, S.P., 2015. 'Divergence at the edges: Peripatric isolation in the montane spiny throated reed frog complex'. *BMC evolutionary biology,* 15 (1), p. 128.

Lawson, M.J. & Zhang, L., 2008. 'Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5′-utr region'. *Gene,* 407 (1), pp. 54-62.

Leberg, P.L., 1990. 'Influence of genetic variability on population growth: Implications for conservation'. *Journal of Fish Biology,* 37 (sA), pp. 193-195.

Li, W., Adams, T.E., Kjellberg, M., Stenflo, J. & Huntington, J.A., 2007. 'Structure of native protein c inhibitor provides insight into its multiple functions'. *Journal of Biological Chemistry,* 282 (18), pp. 13759-13768.

Li, X.-C., Wang, X.-W., Wang, Z.-H., Zhao, X.-F. & Wang, J.-X., 2009. 'A three-domain kazal-type serine proteinase inhibitor exhibiting domain inhibitory and bacteriostatic activities from freshwater crayfish *Procambarus clark*ii'. *Developmental & Comparative Immunology,* 33 (12), pp. 1229-1238.

Lucey, J., 1999. 'A chronological account of the crayfish *Austropotamobius pallipes* (Lereboullet) in Ireland'. *Bulletin-Irish Biogeographical Society,* 23, pp. 143-161.

Lucey, J. & McGarrigle, M.L., 1987. 'The distribution of the crayfish *Austropotamobius pallipes* (Lereboullet) in Ireland'.

Lőw, P., Molnár, K. & Kriska, G., 2016. 'Dissection of a crayfish (*Astacus astacus*)'. In*: Atlas of animal anatomy and histology.* Springer, pp. 101-137.

MacHugh, D.E., Loftus, R.T., Cunningham, P. & Bradley, D.G., 1998. 'Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers'. *Animal Genetics,* 29 (5), pp. 333-340.

Makkonen, J., Jussila, J., Kortet, R., Vainikka, A. & Kokko, H., 2012. 'Differing virulence of *Aphanomyces astaci* isolates and elevated resistance of noble crayfish *Astacus astacus* against crayfish plague'. *Diseases of Aquatic Organisms,* 102 (2), pp. 129-136.

Matallanas, B., Ochando, M.D., Alonso, F. & Callejas, C., 2016. 'Update of genetic information for the white-clawed crayfish in spain, with new insights into its population genetics and origin'. *Organisms Diversity & Evolution,* 16 (3), pp. 533-547.

Matthews, M. & Reynolds, J.D., 1992. 'Ecological impact of crayfish plague in Ireland'. *Hydrobiologia,* 234 (1), pp. 1-6.

Matthews, M.A., Reynolds, J.D. & Keatinge, M.J., 1993. 'Macrophyte reduction and benthic community alteration by the crayfish *Austropotamobius pallipes* (Lereboullet)'. *Freshwater Crayfish,* 9, pp. 289-299.

McFadden, Y.M.T. & Fairley, J.S., 1984a. 'Fish predation as a possible influence on the sizes of crayfish eaten by otters'. *The Irish Naturalists' Journal,* 21 (8), pp. 364-364.

McFadden, Y.M.T. & Fairley, J.S., 1984b. 'Food of otters Lutra lutra (L.) in an Iirish limestone river system witn special reference to the crayfish *Austropotamobius pallipes* (Lereboullet)'. *Journal of Life Sciences Royal Dublin Society,* 5 (1), pp. 65-76.

McFarlane, A., O'Brien, J. & Gammel, M., 2018. 'Observations on breeding of native Irish white-clawed crayfish (*Austropotamobius pallipes*) in captivity.'. *Irish Naturalists Journal,* 36, p. 18.

Migdał, Ł., Ząbek, T., Migdał, A., Łapiński, S., Bieniek, J. & Bugno-Poniewierska, M., 2016. 'Cross species amplification of coat colour genes in nutria (*Myocastor coypus mol)*'. *Folia biologica,* 64 (2), pp. 105-111.

Mirimin, L., Kerwath, S., Macey, B., Lamberth, S.J., Cowley, P.D., Bester-van der Merwe, A., Bloomer, P. & Roodt-Wilding, R., 2016. 'Genetic analyses of overfished silver kob *Argyrosomus inodorus* (Scieanidae) stocks along the southern African coast'. *Fisheries research,* 176, pp. 100-106.

Moriarty, C., 1963. 'Food of perch (*Perca fluviatilis*, L.) and trout (*Salmo trutta*, L.) in an Irish reservoir'. pp. 1-31.

Moriarty, C., 1969. 'Freshwater crayfish 1968'. Department of Agriculture and Fisheries (Fisheries Division). [online: Marine Institute Reposatory]

Moriarty, C., 1971. 'The crayfish astacus pallipes of an irish lake'. Department of Agriculture and Fisheries (Fisheries Division).). [online: Marine Institute Reposatory]

Moritz, C., Dowling, T.E. & Brown, W.M., 1987. 'Evolution of animal mitochondrial DNA: Relevance for population biology and systematics'. *Annual Review of Ecology and systematics,* 18 (1), pp. 269-292.

Mulenga, A., Khumthong, R. & Chalaire, K.C., 2009. '*Ixodes scapularis* tick serine proteinase in hibitor (serpin) gene family; annotation and transcriptional analysis'. *BMC genomics,* 10 (1), p. 217.

Mwangi, S., Murungi, E., Jonas, M. & Christoffels, A., 2011. 'Evolutionary genomics of *Glossina morsitans* immune-related clip domain serine proteases and serine protease inhibitors'. *Infection, Genetics and Evolution,* 11 (4), pp. 740-745.

NBDC, 2017. *Crayfish plague: Additional resources, notes update - october 2017* [Online: https://tinyurl.com/y8uug9eg]. National Biodiversity Data Centre. [Viewed 15/08/2018].

Neaves, L.E., Eales, J., Whitlock, R., Hollingsworth, P.M., Burke, T. & Pullin, A.S., 2015. 'The fitness consequences of inbreeding in natural populations and their implications for species conservation – a systematic map'. *Environmental Evidence,* 4 (1), p. 5.

Nei, M. & Kumar, S., 2000. *Molecular evolution and phylogenetics.* Oxford university press.

Nei, M., Maruyama, T. & Chakraborty, R., 1975. 'The bottleneck effect and genetic variability in populations'. *Evolution,* 29 (1), pp. 1-10.

Oda, K., 2011. 'New families of carboxyl peptidases: Serine-carboxyl peptidases and glutamic peptidases'. *The Journal of Biochemistry,* 151 (1), pp. 13-25.

Patoka, J., Kalous, L. & Kopecký, O., 2014. 'Risk assessment of the crayfish pet trade based on data from the Czech Republic'. *Biological Invasions,* 16 (12), pp. 2489-2494.

Pearce, M.C. & Cabrita, L.D., 2011. 'Production of recombinant serpins in *Escherichia coli*'. In*: Methods in enzymology.* Elsevier, pp. 13-28.

Pedraza-Lara, C., Alda, F., Carranza, S. & Doadrio, I., 2010. 'Mitochondrial DNA structure of the Iberian populations of the white-clawed crayfish, *Austropotamobius italicus italicus* (faxon, 1914)'. *Molecular Phylogenetics and Evolution,* 57 (1), pp. 327-342.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M. & Cardoso, A.C., 2013. 'Essential biodiversity variables'. *Science,* 339 (6117), pp. 277-278.

Perona, J.J. & Craik, C.S., 1995. 'Structural basis of substrate specificity in the serine proteases'. *Protein Science,* 4 (3), pp. 337-360.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. & Ferrin, T.E., 2004. UCSF Chimera - a visualization system for exploratory research and analysis'. *Journal of computational chemistry,* 25 (13), pp. 1605-1612.

Polanowski, Antoni, and Tadeusz Wilusz. "Serine proteinase inhibitors from insect hemolymph." *ACTA BIOCHIMICA POLONICA-ENGLISH EDITION-* 43 (1996): 445-454.

Pope, B. & Kent, H.M., 1996. 'High efficiency 5 min transformation of *Escherichia coli*'. *Nucleic acids research,* 24 (3), pp. 536-537.

Pratten, D.J., 1980. 'Growth in the crayfish *Austropotamobius pallipes* (Crustacea: Astacidae)'. *Freshwater Biology,* 10 (5), pp. 401-402.

Price, T.D., Qvarnström, A. & Irwin, D.E., 2003. 'The role of phenotypic plasticity in driving genetic evolution'. *Proceedings of the Royal Society of London B: Biological Sciences,* 270 (1523), pp. 1433-1440.

Queller, D.C. & Goodnight, K.F., 1989. 'Estimating relatedness using genetic markers'. *Evolution,* 43 (2), pp. 258-275.

Radzicka, A. & Wolfenden, R., 1995. 'A proficient enzyme'. *Science,* 267 (5194), pp. 90-93.

Rai, U.K., 'Minimum sizes for viable population and conservation biology'. *Our Nature,* 1 (1), pp. 3-9.

Raran-Kurussi, S. & Waugh, D.S., 2012. 'The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated'. *PloS one,* 7 (11), p. e49589.

Rawlings, N.D. & Barrett, A.J., 1993. 'Evolutionary families of peptidases'. *Biochemical Journal,* 290 (1), pp. 205-218.

Rawlings, N.D. & Barrett, A.J., 1994. '[2] Families of serine peptidases'. In*: Methods in enzymology.* Elsevier, pp. 19-61.

Rawlings, N.D., Tolle, D.P. & Barrett, A.J., 2004. 'Evolutionary families of peptidase inhibitors'. *Biochemical Journal,* 378 (3), pp. 705-716.

Reynolds, J., 1997. 'The present statuts of freshwater crayfish in Ireland'. *Bulletin Français de la Pêche et de la Pisciculture,* (347), pp. 693-700.

Reynolds, J. D., Catherine Souty-Grosset, Nicolas Gouin, Sandra Devaney, and Frédéric Grandjean. 2000."Experimental restocking of native crayfish in White lake, Co. Westmeath, Ireland." In *Crayfish Conference Leeds*, vol. 26, p. 110. Environment agency Leeds.

Reynolds, J., Lynn, D., O'keeffe, C., Lucey, J., Clabbey, K., Mc-Garrigle, M. & King, J., 2010a. 'Conservation assessment and current status of protected white-clawed crayfish, A*ustropotamobius pallipes* (Lereboullet)'. *Ireland. Freshwater Crayfish,* 17, pp. 123-127.

Reynolds, J., Souty-Grosset, C. & Richardson, A., 2013. 'Ecological roles of crayfish in freshwater and terrestrial habitats'. *Freshwater Crayfish,* 19 (2), pp. 197-218.

Reynolds, J.D., 1988. 'Crayfish extinctions and crayfish plague in central Ireland'. *Biological Conservation,* 45 (4), pp. 279-285.

Reynolds, J.D., 1998. 'Conservation management of the white-clawed crayfish *Austropotamobius pallipes* part 1'.

Reynolds, J.D., 2009. 'The current status of white-clawed crayfish in Ireland'. *Crayfish Conservation in the British Isles*, p. 35.

Reynolds, J.D., Demers, A. & Marnell, F., 2002a. 'Managing an abundant crayfish resource for conservation - *A. pallipes* in Ireland'. *Bulletin Français de la Pêche et de la Pisciculture,* (367), pp. 823-832.

Reynolds, J.D., Gouin, N., Pain, S., Grandjean, F., Demers, A. & Souty-Grosset, C., 2002b. 'Irish crayfish populations: Ecological survey and preliminary genetic findings'. *Freshwater crayfish,* 13, pp. 584-594.

Reynolds, J.D., O'Connor, W., O'Keeffe, C. & Lynn, D., 2010b. *A technical manual for monitoring white clawed crayfish Austropotamobius pallipes in Irish lakes*, Dublin: National Parks and Wildlife Service, D.O.T.E., Heritage and Local Government.

Reynolds, J.D. & O'keeffe, C., 2005. 'Dietary patterns in stream-and lake-dwelling populations of *Austropotamobius pallipes*'. *Bulletin Français de la Pêche et de la Pisciculture,* (376-377), pp. 715-730.

Richman, N.I., Böhm, M., Adams, S.B., Alvarez, F., Bergey, E.A., Bunn, J.J., Burnham, Q., Cordeiro, J., Coughran, J. & Crandall, K.A., 2015. 'Multiple drivers of decline in the global status of freshwater crayfish (Decapoda: Astacidea)'. *Phil. Trans. R. Soc. B,* 370 (1662), p. 20140060.

Rider, S.D., Morgan, M.S. & Arlian, L.G., 2015. 'Draft genome of the scabies mite'. *Parasites & vectors,* 8 (1), p. 585.

Rosano, G.L. & Ceccarelli, E.A., 2014. 'Recombinant protein expression in *Escherichia coli*: Advances and challenges'. *Frontiers in microbiology,* 5**,** p. 172.

Rutty, John. *An Essay Towards a Natural History of the County of Dublin: Accomodated to the Noble Designs of the Dublin Society.* Vol. 2. W. Sleater, 1772. Ryan, C.A., 1990. 'Protease inhibitors in plants: Genes for improving defenses against insects and pathogens'. *Annual review of phytopathology,* 28 (1)**,** pp. 425-449.

Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B. & Kinzig, A., 2000. 'Global biodiversity scenarios for the year 2100'. S*cience,* 287 (5459)**,** pp. 1770-1774.

Santucci, F., Iaconelli, M., Andreani, P., Cianchi, R., Nascetti, G. & Bullini, L., 1997. 'Allozyme diversity of European freshwater crayfish of the genius *Austropotamobius*'. *Bulletin Français de la Pêche et de la Pisciculture,* (347)**,** pp. 663-676.

Scalici, M. & Gibertini, G., 2007. 'Feeding habits of the crayfish *Austropotamobius pallipes* (Decapoda, Astacidae) in a brook in Latium (central Italy)'. *Italian Journal of Zoology,* 74 (2)**,** pp. 157-168.

Scharff, R.F., 1895. 'The freshwater crayfish (*Astacus fluviatilis*) in Co. Dublin'. *The Irish Naturalist,* 4 (2)**,** pp. 50-50.

Scharff, R.F., 1907. *European animals: Their geological history and geographical distribution.* A. Constable & Company Limited.

Schick, C., Kamachi, Y., Bartuski, A.J., Çataltepe, S., Schechter, N.M., Pemberton, P.A. & Silverman, G.A., 1997. 'Squamous cell carcinoma antigen 2 is a novel serpin that inhibits the chymotrypsin-like proteinases cathepsin g and mast cell chymase'. *Journal of Biological Chemistry,* 272 (3)**,** pp. 1849-1855.

Seehausen, O. & Wagner, C.E., 2014. 'Speciation in freshwater fishes'. *Annual review of ecology, evolution, and systematics,* 45**,** pp. 621-651.

Selkoe, K.A. & Toonen, R.J., 2006. 'Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers'. *Ecology letters,* 9 (5)**,** pp. 615-629.

Sezonov, G., Joseleau-Petit, D. & D'Ari, R., 2007. '*Escherichia coli* physiology in luria-bertani broth'. *Journal of bacteriology,* 189 (23)**,** pp. 8746-8749.

Shaffer, M.L., 1981. 'Minimum population sizes for species conservation'. *BioScience,* 31 (2), pp. 131-134.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P., 1994. 'Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers'. *Annals of the entomological Society of America,* 87 (6), pp. 651-701.

Singh, R.R. & Rao, A.G.A., 2002. 'Reductive unfolding and oxidative refolding of a bowman–birk inhibitor from horsegram seeds (*Dolichos biflorus*): Evidence for 'hyperreactive'disulfide bonds and rate-limiting nature of disulfide isomerization in folding'. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology,* 1597 (2), pp. 280-291.

Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J. & Li, Q., 2009. 'Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*'. *Protein Science,* 18 (5), pp. 936-948.

Smith, M.A. & Bidochka, M.J., 1998. 'Bacterial fitness and plasmid loss: The importance of culture conditions and plasmid size'. *Canadian journal of microbiology,* 44 (4), pp. 351-355.

Somprasong, N., Rimphanitchayakit, V. & Tassanakajon, A., 2006. 'A five-domain kazal-type serine proteinase inhibitor from black tiger shrimp *Penaeus monodon* and its inhibitory activities'. *Developmental & Comparative Immunology,* 30 (11), pp. 998-1008.

Soulé, M., 1980. 'Thresholds for survival: Maintaining fitness and evolutionary potential'. *Comservation biology: an evolutionary-ecological perspective*, pp. 151-169.

Souty-Grosset, C., Grandjean, F. & Gouin, N., 2003. 'Keynote involvement of genetics in knowledge, stock management and conservation of *Austropotamobius pallipes* in Europe'. *Bulletin Français de la Pêche et de la Pisciculture,* (370-371), pp. 165-179.

Souty-Grosset, C., Grandjean, F., Raimond, R., Frelon, M., Debenest, C. & Bramard, M., 1997. 'Conservation genetics of the white-clawed crayfish *Austropotamobius pallipes*: The usefulness of the mitochondrial DNA marker'. *Bulletin Français de la Pêche et de la Pisciculture,* (347), pp. 677-692.

Souty-Grosset, C. & Reynolds, J., 2009. 'Current ideas on methodological approaches in European crayfish conservation and restocking procedures'. *Knowledge and Management of Aquatic Ecosystems,* (394-395), p. 01.

Strayer, D.L. & Dudgeon, D., 2010. 'Freshwater biodiversity conservation: Recent progress and future challenges'. *Journal of the North American Benthological Society,* 29 (1), pp. 344-358.

Studier, F.W., 2014. 'Stable expression clones and auto-induction for protein production in *E. coli*'. In: *Structural genomics.* Springer, pp. 17-32.

Söderhäll, K., 2011. *Invertebrate immunity.* Springer Science & Business Media.

Söderhäll, K. & Cerenius, L., 1998. 'Role of the prophenoloxidase-activating system in invertebrate immunity'. *Current opinion in immunology,* 10 (1), pp. 23-28.

Söderhäll, K. & Unestam, T., 1979. 'Activation of serum prophenoloxidase in arthropod immunity. The specificity of cell wall glucan activation and activation by purified fungal glycoproteins of crayfish phenoloxidase'. *Canadian Journal of Microbiology,* 25 (3), pp. 406-414.

Tajima, F., 1989. 'Statistical method for testing the neutral mutation hypothesis by DNA polymorphism'. *Genetics,* 123 (3), pp. 585-595.

Templeton, A.R., Crandall, K.A. & Sing, C.F., 1992. 'A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. Iii. Cladogram estimation'. *Genetics,* 132 (2), pp. 619-633.

Templeton, R., 1836. 'Catalogue of Irish Crustacea, Myriapoda and Arachnoida, selected from the papers of the late john templeton, esq'. *Annals and Magazine of Natural History (Series 1) Vol,* 9, pp. 9-14.

Tiffin, P. & Ross-Ibarra, J., 2014. 'Advances and limits of using population genetics to understand local adaptation'. *Trends in ecology & evolution,* 29 (12), pp. 673-680.

Tighe, W., 1802. *Statistical observations relative to the county of Kilkenny: Made in the years 1800 & 1801.* Graisberry and Campbell.

Toews, D.P.L. & Brelsford, A., 2012. 'The biogeography of mitochondrial and nuclear discordance in animals'. *Molecular Ecology,* 21 (16), pp. 3907-3930.

Van Den Berg, S., Löfdahl, P.-Å., Härd, T. & Berglund, H., 2006. 'Improved solubility of TEV protease by directed evolution'. *Journal of biotechnology,* 121 (3), pp. 291-298.

Walker, I.H., Hsieh, P.-c. & Riggs, P.D., 2010. 'Mutations in maltose-binding protein that alter affinity and solubility properties'. *Applied microbiology and biotechnology,* 88 (1), pp. 187-197.

Wan, Q.H., Wu, H., Fujihara, T. & Fang, S.G., 2004. 'Which genetic marker for which conservation genetics issue?'. *Electrophoresis,* 25 (14), pp. 2165-2176.

Wang, R., Lee, S.Y., Cerenius, L. & Söderhäll, K., 2001. 'Properties of the prophenoloxidase activating enzyme of the freshwater crayfish, *Pacifastacus leniusculus*'. *European Journal of Biochemistry,* 268 (4), pp. 895-902.

Webb, E.C., 1992. *Enzyme nomenclature 1992. Recommendations of the nomenclature committee of the international union of biochemistry and molecular biology on the nomenclature and classification of enzymes.* Academic Press.

Wolfenden, R. & Snider, M.J., 2001. 'The depth of chemical time and the power of enzymes as catalysts'. *Accounts of chemical research,* 34 (12), pp. 938-945.

Wright, S., 1931. 'Evolution in mendelian populations'. *Genetics,* 16 (2), p. 97.

Wu, C., 2011. 'Innate immune proteins in a crustacean *Pacifastacus leniusculus*'. *Acta Universitatis Upsaliensis*.

Yazicioglu, B., Reynolds, J. & Kozák, P., 2016. 'Different aspects of reproduction strategies in crayfish: A review'. *Knowledge and Management of Aquatic Ecosystems,* (417), p. 33.

Zaccara, S., Stefani, F., Galli, P., Nardi, P.A. & Crosa, G., 2004. 'Taxonomic implications in conservation management of white-clawed crayfish (*Austropotamobius pallipes*)(Decapoda, Astacidae) in northern Italy'. *Biological Conservation,* 120 (1), pp. 1-10.

Zeng, F., Zhang, S., Hao, Z., Duan, S., Meng, Y., Li, P., Dong, J. & Lin, Y., 2018. 'Efficient strategy for introducing large and multiple changes in plasmid DNA. *Scientific reports,* 8 (1), p. 1714.

Zhou, A., Carrell, R.W. & Huntington, J.A., 2001. 'The serpin inhibitory mechanism is critically dependent on the length of the reactive center loop'. *Journal of Biological Chemistry*.

Zhou, A., Jiang, X., Dou, F., Zhu, D. & Xu, X., 1997. 'Renaturation, purification, and characterization of human plasminogen activator inhibitor type 2 (pai-2) accumulated at high level in *Escherichia coli*'. *The Journal of Biochemistry,* 121 (5), pp. 930-934.

Zou, Z., Evans, J.D., Lu, Z., Zhao, P., Williams, M., Sumathipala, N., Hetru, C., Hultmark, D. & Jiang, H., 2007. 'Comparative genomic analysis of the Tribolium immune system'. *Genome biology,* 8 (8), p. R177.

Zou, Z., Picheng, Z., Weng, H., Mita, K. & Jiang, H., 2009. 'A comparative analysis of serpin genes in the silkworm genome'. *Genomics,* 93 (4), pp. 367-375.