Ageing the unageable: investigating novel methods of ageing crustaceans

Eleanor Adele Fairfield

A thesis submitted for the degree of Doctor of Philosophy
University of East Anglia
School of Biological Sciences

2021

Painting by Sophie Fairfield

Thesis abstract

Crustaceans are extremely difficult to age because of their indeterminate and variable growth, and the moulting of their exoskeleton throughout life. The consequent lack of data on the age structure of populations makes it impossible to reliably predict population dynamics and viability for sustainable fisheries management. The aim of this thesis was to investigate whether crustacean age can be estimated using novel, molecular markers. Three potential markers of age were investigated in laboratory-reared cohorts of red cherry shrimp (*Neocaridina davidi*) and/or the economically-important European lobster (*Homarus gammarus*). First, genome-wide levels of methylation were quantified using a commercially-available, immunoassay-based kit. While the poor reliability afforded by the kit undermined comparisons, there was low variation in global DNA methylation among age groups of lobsters and shrimp. Second, bisulphite sequencing was used to measure site-specific methylation across 5,154 bp of ribosomal DNA sequence in lobsters that ranged in known age from 0–2 years old. A model based on the ten 'best' loci predicted lobster age with high accuracy and precision (ca. 2 months). Finally, the full mitochondrial genomes of lobsters and shrimp were sequenced to high depth using next-generation sequencing. In shrimp, the number of mitochondrial DNA (mtDNA) mutations across 39% of the genome did not differ among seven age groups spanning 54% of the species' lifespan. Sequencing data from the remaining shrimp mitochondrial genome were confounded by probable co-amplification of nuclear DNA of mitochondrial origin (NUMTs). In lobsters, the entire mitochondrial sequencing dataset was confounded by mis-mapping of reads across a duplicated region and likely co-amplification of NUMTs. A range of bioinformatic approaches aimed at removing erroneous sequencing reads was attempted. After thorough filtering of potential false positives, no difference in the number of mtDNA mutations was observed among six lobster age groups. In summary, mtDNA mutation accumulation and genome-wide methylation (at least when measured using the method adopted in this thesis) can be ruled out as viable markers of crustacean age. Conversely, patterns of loci-specific ribosomal DNA methylation show considerable promise as a future tool for crustacean ageing.

Contents

List of figures

List of tables

Acknowledgements

First of all, thank you to my supervisors Martin Taylor and David Richardson. You make a great team and I have learnt a great deal from you. I really appreciate the kindness you have both shown me when I needed it. Thank you to Martin for putting your faith in me to take on this project, and guiding me through what transpired to be some baffling bioinformatics! Despite all the hurdles, I feel incredibly fortunate to have had this experience. Thank you to David for being one of the people that inspired me to work in academia for as long as I have. I feel I have improved a great deal as a scientist while working with you (for nearly nine years!), and I can attribute much of that to your guidance (and always plenty of feedback!).

My research would not have been possible without the help of many other people. Firstly, what is a PhD on crustacean ageing without any crustacean DNA?! A huge thank you to the National Lobster Hatchery, in particular Carly Daniels and Charlie Ellis, and Colin Trundle from the Cornwall IFCA, for proving me with lobster samples. And thank you to Ewen Bell for putting me in touch with Colin and teaching me the lobster lingo. Gathering crustacean DNA from known-age animals proved to be more difficult than Martin and I anticipated. If only there was an established model system for studying crustacean ageing... Let's start our own! And with that, thank you Martin for allowing me to displace some of your cherished catfish for something else cherished, or cherried (the shrimp). It's fair to say I learnt more about plumbing during this PhD than I expected! Thank you to technicians at the Earlham Institute and Zymo Research for transforming my crustacean DNA into a readable sequence of letters. And to the BBSRC Doctoral Training Partnership and the Seafood Innovation Fund for funding this work.

I learnt the hard way that a PhD thesis is not just a product of the group of people directly involved. It's fair to say I found this experience tougher than I imagined. And there are (at least) two people without whom I may have packed up my lobster pencil case (thanks Sarah!) before getting to this point. Thank you to David Murray and Jenny Gill for teaching me the importance of admitting when you need extra help or another perspective.

This PhD has gifted me with something really invaluable, and that's a huge sense of gratification for a fantastic group of friends. Firstly, thank you to my Zeke office comrades past and present. I doubt I'll ever work with such a wonderfully eccentric and funny, and at the same time caring, group of people. I'm gutted that a global pandemic stole six months of time away from the current inhabitants, but so grateful for the hilarious times we did have. My only other regret is not finishing the Lego card collection! To Kat Bebbington, Ellen Bell, David Collins, Jen Donelan, Jessie Gardner, Dan Hewiit, Sarah Marburger, David Murray, Kris Sales and Ram Vasudeva. You have been spectacular friends. Here's to a lifetime of playfulness and laughter. Special thanks to Ellen Bell for being an exceptionally supportive, research group companion, and nearly always being the first to witness, and encourage, office-based mischief. To the wider BIO 01 community, you all play a part in creating what is an amazingly collegial environment to work in. I have never failed to feel inspired, valued or supported by you all.

My family are a constant source of inspiration and comfort. To my mum and dad. This is just one of a long list of things that I wouldn't have attempted without your backing. Thank you for teaching me the value of commitment and encouraging me to simply do something I feel I would enjoy. Weekends with you kept me going. They say that a PhD requires resilience, and I owe all of mine to growing up with two older brothers, haha! Tristan and Alastair, I couldn't begin to count the times that, in a moment of despair, I've thought of you two - how hardworking and brilliant you are - and it's bolstered me back into shape. Thanks to my niece Tilly and nephew Hugo for having no idea what I do at work and it not making the slightest bit of difference, and to Hugo for being the centre of one of my favourite supervisor-based anecdotes.

Finally, to my boyfriend Andrea. I don't want to imagine how this experience would have been without your unshakeable encouragement and thoughtfulness. You know when to ask me about work and when to keep quiet, you know when to leave me alone and when to take me out for an adventure, and you know when I'll appreciate food (always).

Chapter 1

General introduction

1.1 Importance and condition of marine fisheries

Humans have been exploiting marine resources for thousands of years and capture fisheries and aquaculture remain an essential source of employment, food and nutrition worldwide. In 2018, 179 million tonnes of fish (hereafter, "fish" refers to any aquatic/marine animal that is harvested) and shellfish were produced globally with an estimated first-sale value of $401 billion and over 59 million people were employed for harvesting and farming the fish (FAO 2020). Fish exports can be particularly valuable for low- and medium-income countries, accounting for 10% or more of the GDP in some countries (Béné et al. 2016). An estimated 88% of fish produced is used for human consumption (FAO 2020). Over 17% of the global population's intake of animal protein is obtained from fish, which is twice the amount of protein provided by poultry and three times the amount provided by cattle (Béné et al. 2016; FAO 2020). In coastal communities in developing countries, fish provides more than half an individual's intake of animal protein (FAO 2020). Fish are also an important source of other key nutrients including omega-3 fatty acids (FAO 2020). Humans have never consumed as much fish as they do today. Since 1961, the annual increase in fish consumption has been almost twice as high as population growth (FAO 2020).

Fishing has been the fastest growing food industry over the last 40 years and is expected to remain so in the near future (Béné et al. 2016). With the human population expected to exceed 9 billion people by 2050, the fishing industry faces unprecedented challenges to ensure that marine resources can keep pace with demand (Pauly et al. 2002; Béné et al. 2016; FAO 2020). Since 1950, there has been a ten-fold increase in global fishing intensity, driven by rising demand and the development of new fishing technologies (Watson et al. 2012). However, this increased fishing effort has not driven a proportionate increase in catches, which have remained relatively static since the 1980s (FAO 2020). The poor state of our capture fisheries, as a result of overexploitation, climate change, pollution and land conversion, is well documented (e.g. Myers & Worm 2003; Worm et al. 2006; Sumaila & Tai 2020). Indeed, it has been projected that, if the loss of marine species and populations continues its present trend, there will be a total global collapse of all currently fished taxa by 2050 (Worm et al. 2006). However, not all commentators have

painted such a bleak picture; there is evidence that the decline in the status of marine fisheries is "neither inevitable nor irreversible" and that fisheries can be managed successfully to achieve long-term sustainability (Beddington et al. 2007; Hilborn 2007; Daan et al. 2011; Hilborn et al. 2020). Unfortunately, sustainable management tools for fisheries have not been implemented widely and many small stocks remain unmanaged (Beddington et al. 2007; Hilborn et al. 2020). The United Nations Food and Agriculture Organisation (FAO), which has monitored the state of world fisheries since 1950, estimates that 34.2% of commercial fish stocks are fished at biologically unsustainable levels (FAO 2020). The expansion of aquaculture has alleviated some of the pressure on capture fisheries, but may have damaging ecological and environmental impacts (e.g. disease and parasite transmission, pollution and genetic contamination of wild populations). Aquaculture also often depends on capture fisheries to supply feed (Clavelle et al. 2019).

1.2 Fish stock assessments

"A unit stock is an arbitrary collection [of a single species] of fish that is large enough to be essentially self-reproducing…The unit should not be so large as to contain many genetically distinct races of subpopulations within it."

Hilborn and Walters 1992

A sustainable level of fishing is achieved when the total mortality caused by fishing, combined with natural mortality, does not exceed an amount that a population can withstand (via natural recruitment) without significant declines (FAO 2002). To estimate the sustainable level of fishing for a given stock, fisheries managers perform fish stock assessments. The basic premise of a stock assessment is to collate data that can be used to estimate the past and present state of a stock. These data are subsequently used to forecast how stocks will respond to future levels of harvesting using population dynamics models and ultimately inform sustainable management strategies (e.g. catch limits) that minimise the risk of future stock depletion (FAO 1998).

Accurate predictions of population dynamics rely on having a good understanding of the species and stock in question (Beddington et al. 2007). While there have been calls for a more ecosystem-based approach, stock assessments are predominantly performed using single-species population models (Pauly et al. 2002; Quinn & Collie 2005) that comprise solely of biological data from the species of interest. Single-species population dynamics models can be categorised into three main types, which differ in the primary biological data used to inform the model: 1) biomass dynamics models, 2) age-structured models and 3) size-structured models (Punt et al. 2013). These population dynamics models are not, however, mutually exclusive; where more than one type of biological data is available, or in cases where a complete time series for one type of data does not exist, biological data from different sources can be combined and assessed through an 'integrated analysis' (Maunder & Punt 2013; Punt et al. 2013).

1.2.1 Biomass dynamics models

Biomass dynamics models map temporal trends in the number of fish caught combined with a measure of abundance (most commonly catch-per-unit effort) to represent the available biomass in a stock. When a long time-series of biomass data exists for a given stock, future changes in biomass under different fishing scenarios can be predicted (Hilborn 1992). The main strength of biomass dynamics models lies in the simplicity of the data required (Smith & Addison 2003); data on catch and fishing effort can be acquired directly from the fishery, although data from commercial fisheries are commonly combined with standardised fishery-independent surveys to minimise potential bias in fisheries data. However, the simplicity of biomass dynamics models is also their main limitation (Smith & Addison 2003); biomass dynamics models provide a highly simplified representation of population dynamics by ignoring key information about demographic structure (i.e. age or size structure). Consequently, biomass dynamics models rely on a number of assumptions and are only applied to stocks where data beyond measures of catch and abundance are unavailable (Hilborn et al. 1992; FAO 1998; Maunder & Piner 2014).

1.2.2 Age-structured models

Evaluating age structure is a major focus for sustainable fisheries management and provides the foundations for reliably predicting future population dynamics and setting sustainable catch quotas (Campana 2001; Maunder & Piner 2014). Data on age structure are used to estimate parameters such as the proportion of individuals with reproductive potential or at risk of mortality (Campana 2001; Maunder & Piner 2014). Moreover, the age structure of a population is considered one of the most important indicators of a "healthy" fish stock (Brunel & Piet 2013). For example, an overabundance of smaller, younger individuals may be a sign of overexploitation (fishing tends to favour catching larger fish) or environmental stress (Pope et al. 2010). Conversely, populations dominated by larger, older individuals may be suffering from limited recruitment (Pope et al. 2010). For some fish, the process of measuring age is aided by observable morphological characteristics that change predictably with time. Examples of age-informative morphological features include the counting of incremental growth rings in finfish otoliths (Pannella 1971; Campana 2001) and bivalve shells (Kilada et al. 2009). However, accurate age determination is crucial, and there are cases where inaccurate ageing has contributed to the overexploitation of a population (Campana 2001 and references therein). It is therefore not surprising that government laboratories invest heavily in acquiring and analysing age-structure data (Pauly et al. 2002).

1.2.3 Size-structured models

Some fish do not display preserved growth bands for age determination (Jarman et al. 2015). Therefore, growth models based on the size-structure of individuals within populations have been developed (Pauly & Morgan 1987). Size-structured analysis involves identifying modal groups in size-frequency data (usually a standardised measurement of length) and converting the modal groups into different age classes (Maunder & Piner 2014). Compared to age data, size data are inexpensive and easy to obtain. However, converting size data into distinct age classes relies on the presence of discrete size-frequency modes, which rarely exist in natural populations (Phillips 2006). Growth rate is influenced by environmental conditions, such as temperature and food availability (Hartnoll 2001), and so large variation in size can exist among individuals of

the same age (Sheehy et al. 1999; Vogt 2012). Therefore, stock assessments based on size-structured models are typically only applied when the actual age structure of a population cannot be measured.

1.3 Crustacean fisheries

Crustaceans are a group of arthropods comprising almost 67,000 described species (Ahyong et al. 2011). Perhaps the most familiar crustaceans, and certainly the most commercially relevant, are the Decapoda (from the Greek words for "ten" Deca-, and "foot" -pod), including shrimps, lobsters, crayfish and crabs. Hereafter, 'crustaceans' refers exclusively to the Decapoda for simplicity. Crustaceans have a global distribution and dwell in the benthic zone of most aquatic habitats from intertidal to deep water where they feed on detritus and other benthic invertebrates (Boudreau & Worm 2012). Crustaceans themselves are prey for a number of different predators including finfish, birds, invertebrates (mostly other decapods, including conspecifics), marine mammals and, of course, humans (Boudreau & Worm 2012).

Humans have been harvesting crustaceans for hundreds of years but crustaceans have not always been the luxury commodity known today. In the 17th century, lobster was fed to prisoners in America. This practice was deemed so cruel a punishment that some American states had laws to limit the amount of lobster fed to inmates (Townsend 2011). While the global catch of finfish has declined or remained relatively stable since the 1980s (FAO 2020), there has been a rapid and continual expansion in crustacean catch since 1950 (Anderson et al. 2011). In 2018, there were record catches of lobsters and shrimp (FAO 2020). Crustacean catch has the highest value per unit weight of any aquatic animal group; in 2018, crustaceans accounted for 7% of marine capture by weight and 22% by trade value (FAO 2020). In the same year, the UK alone exported nearly 50 thousand tonnes of crustacean with a value of £373 million (MMO 2019), equating to over 20% of the UK export value of all fish caught.

Growth in the crustacean industry has been attributed to overexploitation of finfish stocks, which has forced fishermen to target new, lower trophic species; a global phenomenon known as 'fishing down the food web' (Pauly et al. 1998). Fishermen

have typically switched to targeting smaller pelagic species and benthic invertebrates (Pauly et al. 2002). Concurrently, invertebrate abundance may have increased due to the falling presence of finfish predators (Anderson et al. 2011). A well-known example of 'fishing down the food web' occurred in the Gulf of Maine (United States and Canada). This fishery was once dominated by a plentiful supply of large, predatory groundfish (mainly Atlantic cod (*Gadus morhua*)) but coastal populations of key groundfish species were declared "depleted" in 1949 (Steneck et al. 2011). Many fishers switched to catching lobsters (*Homarus americanus*), which had increased in abundance and expanded their range owing to a lack of predation (Steneck et al. 2011). However, commercial crustacean fisheries are largely unassessed and unregulated owing to a lack of ecological and life-history data (Anderson et al. 2011; Boudreau & Worm 2012). Consequently, increased crustacean fishing effort has caused the overexploitation of some species (Bondad-Reantaso et al. 2012; Boudreau & Worm 2012; Spanier et al. 2015). Even for crustacean populations that have increased in abundance, the long-term effects of increased fishing effort on crustacean stocks are unknown (Boudreau & Worm 2012), not to mention the wider ecological effects on the marine ecosystem (Steneck et al. 2011; Boudreau & Worm 2012).

1.4 Age determination in crustaceans

"*…the determination of age has posed one of the greatest challenges to crustacean biologists*"

Bruce Phillips 2006

A key barrier to the sustainable management of crustacean fisheries is our inability to accurately estimate crustacean age (Smith & Addison 2003; Vogt 2012; Kilada & Driscoll 2017). Crustaceans periodically moult throughout their lives, which involves the complete loss and regeneration of their exoskeleton. As a result, crustaceans lose nearly all (see below) of their hard, calcified structures that might otherwise accumulate visible signatures of age (i.e. growth rings) (Kilada & Driscoll 2017). In the absence of age data, crustacean stock assessments rely on size-structured models based on measurements of carapace length (Vogt 2012; Kilada & Driscoll 2017). However, length is only weakly correlated with age in crustaceans due to

environmental and age-related effects on growth rate (Vogt 2012; Kilada & Driscoll 2017). Age estimates based on size are particularly problematic for long-lived species because fast-growing young and slow-growing old individuals increasingly overlap in size (Vogt 2012; Kilada & Driscoll 2017). Many of the most economically-important crustaceans have (or are predicted to have) long lifespans, including species of clawed lobster (e.g. *H. americanus*: 33 years; *H. gammarus*: 42–72 years), spiny lobster (e.g. *Jasus lalandii*: 40 years) and crab (e.g. *Cancer pagurus*: 21 years) (Vogt 2019). A reliable and accurate ageing method is, therefore, urgently needed for crustacean fisheries management and would have considerable economic and conservation impacts.

While it was not the focus of this thesis, age determination in crustaceans is also of interest to those studying the evolution of age-related deterioration (senescence). Crustaceans exhibit multiple defences against the decline in function with increasing age observed in many other organisms (Vogt 2012). Examples include moulting, tissue regeneration, indeterminate growth, continuous expression of telomerase and detoxification of free radicals (Godwin et al. 2011; Vogt 2012). These characteristics are thought to help protect crustaceans from mechanical and reproductive deterioration over time and may contribute to the longevity seen in some species (Vogt 2012; Vogt 2019).

There is a clear need to develop an ageing tool for crustaceans, both for applied research and to further our understanding of biological ageing. To this end, there have been many attempts to discover alternative approaches for estimating crustacean age. The main methods of crustacean age determination that have been explored so far are summarised and reviewed below.

1.4.1 Captive rearing

Rearing individuals in captivity is the most exact means of measuring age and growth rate. However, captive rearing of the longer-lived, commercially-important crustaceans (see above) is not feasible as this would require potentially several generations of researchers (Vogt 2012). Crustaceans are rarely kept for a long time in captivity from birth and are usually collected from the wild at an already unknown age.

In addition, the optimal conditions of captivity may misrepresent conditions experienced in the wild and provide inaccurate measures of growth rate for use in stock assessments (Vogt 2012).

1.4.2 Tag and recapture

Tag and recapture methods have the benefit of providing a direct means of measuring age in wild crustaceans without the potential environmental biases introduced by captive rearing. Traditional external tags are lost through moulting and can lead to impaired growth and animal mortality (Sharp et al. 2000). Modern internal tags, however, can persist through moulting (reviewed by Hartnoll 2001; Vogt 2012). The most appropriate internal tags for long-lived decapods are passive integrated transponders (PIT) and coded microwire tags (Vogt 2012). Due to their size and weight, PIT tags are only suitable for larger animals (Gibbons & Andrews 2004), which gives them limited applicability across species and age groups. In addition, tags can move within the body, sometimes making them unreadable (Gibbons & Andrews 2004). The main limitation of coded microwire tags is the need to lethally excise the tag for identification (Hartnoll 2001; Vogt 2012). All long-term tag and recapture studies rely on fishers to catch and return tagged individuals, which can introduce bias if fishers fail to notice the tags (especially likely for internal tags) or do not report them (Sharp et al. 2000). Ultimately, the time, cost and potential bias involved in tagging, have not inspired the routine use of tag and recapture programmes for informing crustacean stock assessments.

1.4.3 Lipofuscin accumulation

Lipofuscin is a pigment that accumulates over time in post-mitotic cells (Brunk & Terman 2002). Significant relationships between lipofuscin content and age have been observed in lobsters (*H. Gammarus*; R^2 = 46.8: O'Donovan & Tully 1996), prawns (*Penaeus japonicas*; R^2 = 0.69: Vila et al. 2000), and crabs (*Cherax quadricarinatus*; R^2 = 0.49: Ju et al. 1999), although a more recent study involving *C. quadricarinatus* did not find a significant relationship between lipofuscin content and age (Crowley et al. 2014). The rate of lipofuscin accumulation is affected by environmental effects, such as temperature and diet, which may explain why a large proportion of the variation in age

is not accounted for by lipofuscin concentration alone (Phillips 2006). Furthermore, the methodology for quantifying lipofuscin concentration is labour intensive, highly technical and requires lethal sampling (Hartnoll 2001; Vogt 2012).

1.4.4 Growth bands

It has long been assumed that crustaceans lose all hard, calcified structures during moults. Recently, however, growth bands in the eyestalks and gastric mills of four different crustacean species (*H. americanus*, *Chionoecetes opilio*, *Sclerocrangon boreas* and *Pandalus borealis*) were reported to persist through moulting and correlate with age (Kilada et al. 2012). Eight years after the initial study, this method remains to be properly validated across different species, populations, and a wide range of known ages. Furthermore, it remains contentious to what extent these growth bands are affected by moulting and further research is needed to elucidate the potential environmental effects on the formation and periodicity of growth bands (Sheridan et al. 2016). A study involving four different species (*H. gammarus*, *Nephrops norvegicus*, *Cancer pagurus* and *Necora puber*) suggests that gastric mill growth bands are not retained during moults (Becker et al. 2018). A separate study did not observe a relationship between the number of eyestalk growth bands and age in *Maja squinado* (Sheridan et al. 2020).

1.4.5 Telomere length

Telomeres, the protective caps of non-coding DNA found on the end of chromosomes, shorten throughout life in a range of wild species (reviewed by Dunshea et al. 2011). However, telomere length does not appear to accurately reflect chronological age, likely as a result of the complex interplay of genetic and environmental effects on the rate of telomere loss (reviewed by Dunshea et al. 2011; Jarman et al. 2015). In crustaceans, no relationship between telomere length and age was observed in both long-lived (*Sagmariasus verreauxi* and *Jasus edwardsii*) and short-lived (*Metapenaeus macleayi*) species (Godwin et al. 2011). This is possibly due to the almost ubiquitous expression of telomerase (an enzyme that restores telomere ends) in crustacean tissues (Klapper et al. 1998; Elmore et al. 2008). The activity of telomerase was shown to be independent of lifespan and it has been hypothesised that the high levels of

telomerase expression are related to the ability of lobsters (and other crustaceans) to regenerate injured tissue (Elmore et al. 2008).

1.5 Molecular tools for ageing crustaceans

There is growing interest in the use of molecular tools for estimating age in difficult-to-age species. For example, Jarman and colleagues (2015) reviewed the molecular methods currently available to determine chronological age in wild animals. An appraisal of methylation-based approaches to measuring animal age was published two years later (De Paoli-Iseppi et al. 2017). A molecular marker of age involves measuring a feature (such as abundance or nucleotide sequence) of DNA or RNA, or their associated molecules (e.g. methylation), that changes predictably throughout an animal's lifespan (Jarman et al. 2015). An ideal molecular marker of age would have a perfect linear relationship (r/R^2 = 1 or -1) with chronological age. However, all molecular markers are influenced by environmental effects or inherited variability to some extent (Jarman et al. 2015). Some markers are better predictors of biological age, which describes the progressive decline in function and increased risk of death over time (senescence). Significant variation exists within species, such that individuals of the same chronological age can have different biological ages, owing to genetic, parental or environmental influences on function (Nussey et al. 2013). Despite this, some molecular markers explain a large amount of the variation in chronological age in different species (reviewed by Jarman et al. 2015 and De Paoli-Iseppi et al. 2017).

With the exception of a study on telomere length (see section 1.4.5), DNA-based age determination is a poorly-explored avenue for crustacean ageing. With ever-increasing improvements in the cost-effectiveness and ease of collecting and analysing molecular data, there are progressively fewer barriers to implementing genetic tools in government labs. In particular, next-generation sequencing (NGS) has revolutionised our ability to gather vast amounts of DNA sequencing data from non-model organisms at an ever-reducing cost (Ekblom & Galindo 2011). Ongoing improvements to NGS technology are also transforming where DNA can be sequenced and analysed. For example, the recently-developed MinION USB sequencer has been used for genetic species identification in a remote tropical forest (Pomerantz et al. 2018) without the

need for a dedicated lab. In fisheries science this could, for example, allow for port-side sampling of fish for rapid stock assessments. Several changes in DNA or RNA characteristics have been observed to occur in ageing animal tissues that are potentially applicable to crustaceans. These molecular markers of age are summarised below and their potential application to crustaceans is addressed.

1.5.1 Mitochondrial DNA mutation accumulation

Mitochondria are found in nearly all eukaryotic cells where they play an essential role in cellular processes including energy production, metabolism, cell cycling and death. Mitochondria are the only animal organelle to contain their own DNA, which exists in multiple copies (up to several thousand) per cell. Each copy of mitochondrial DNA (mtDNA) within a cell, tissue or individual may be identical (homoplasmy) or exist as a mixture of mutant and wild-type copies (heteroplasmy). Heteroplasmy is attributed to somatic or inherited mutations (Lawless et al. 2020). Somatic mutations are caused by oxidative damage during mitochondrial respiration or polymerase errors during mtDNA replication. Inherited mutations are typically transmitted via heteroplasmic maternal mtDNA in egg cells (Polovina et al. 2020).

Mitochondrial DNA mutations originate in a single copy of mtDNA within a single mitochondrion of a single cell (Stewart & Chinnery 2015). Over time, mtDNA mutations accumulate with age in humans and vertebrate animal models (Larsson 2010 and references therein; Arbeithuber et al. 2020). This proliferation of mtDNA mutations is attributed to accumulating oxidative damage (Fig. 1.1) or clonal expansion of existing mutations during cell division and/or mtDNA replication (Fig. 1.2) (reviewed by Lawless et al. 2020). The age-related increase of mtDNA mutations in humans has been put forward as an aid to estimating age in forensic investigations (Meissner & Ritz-Timme 2010). Intriguingly, the same mtDNA mutations can occur and accumulate in unrelated individuals (Samuels et al. 2013). For example, the prevalence of a common mtDNA deletion of 4,977 bp (Meissner et al. 1999; von Wurmb-Schwark et al. 2002) and a single point mutation (A189G; Thèves et al. 2006) increases with human age. These studies raise the possibility of ageing individuals by quantifying heteroplasmy at one or a few loci, rather than screening whole mitochondrial genomes. Despite this promising

feature, age-related heteroplasmy remains to be investigated in a wide range of animals (Jebb et al. 2018).

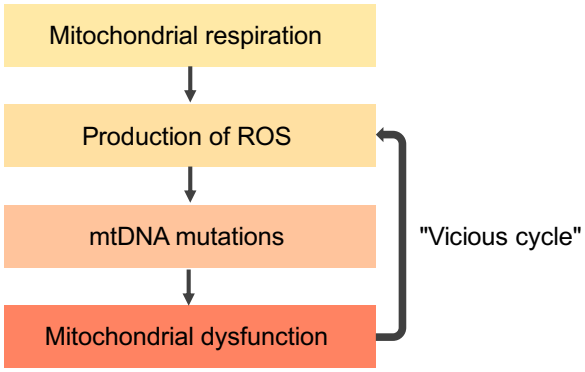


Figure 1.1 | The vicious cycle theory of mtDNA mutation accumulation. This theory posits that mtDNA mutations increase in frequency within a cell as a result of accumulating damage from reactive oxygen species (ROS) produced as a by-product of mitochondrial respiration.

Heteroplasmy has been shown to occur in crustaceans (e.g. Rodríguez-Pena et al. 2020) but nothing is known of age-related changes in heteroplasmy in these animals. NGS is the preferred method for identifying heteroplasmy because high sequencing depth is necessary to identify low-frequency mtDNA mutations (González et al. 2020). A benefit of investigating molecular markers based on mtDNA is that little genetic information from the target species is needed prior to sequencing. Mitochondrial DNA is a short molecule (typically 15–17 kb) and so it is possible to sequence the entire mitochondrial genome without the need to first identify candidate loci (Jarman et al. 2015). Primer design for mtDNA amplification is facilitated by a reference mitochondrial genome, either for the species in question or a close relative. Reference mitochondrial genomes have been published for a large number of crustacean species (Tan et al. 2019).

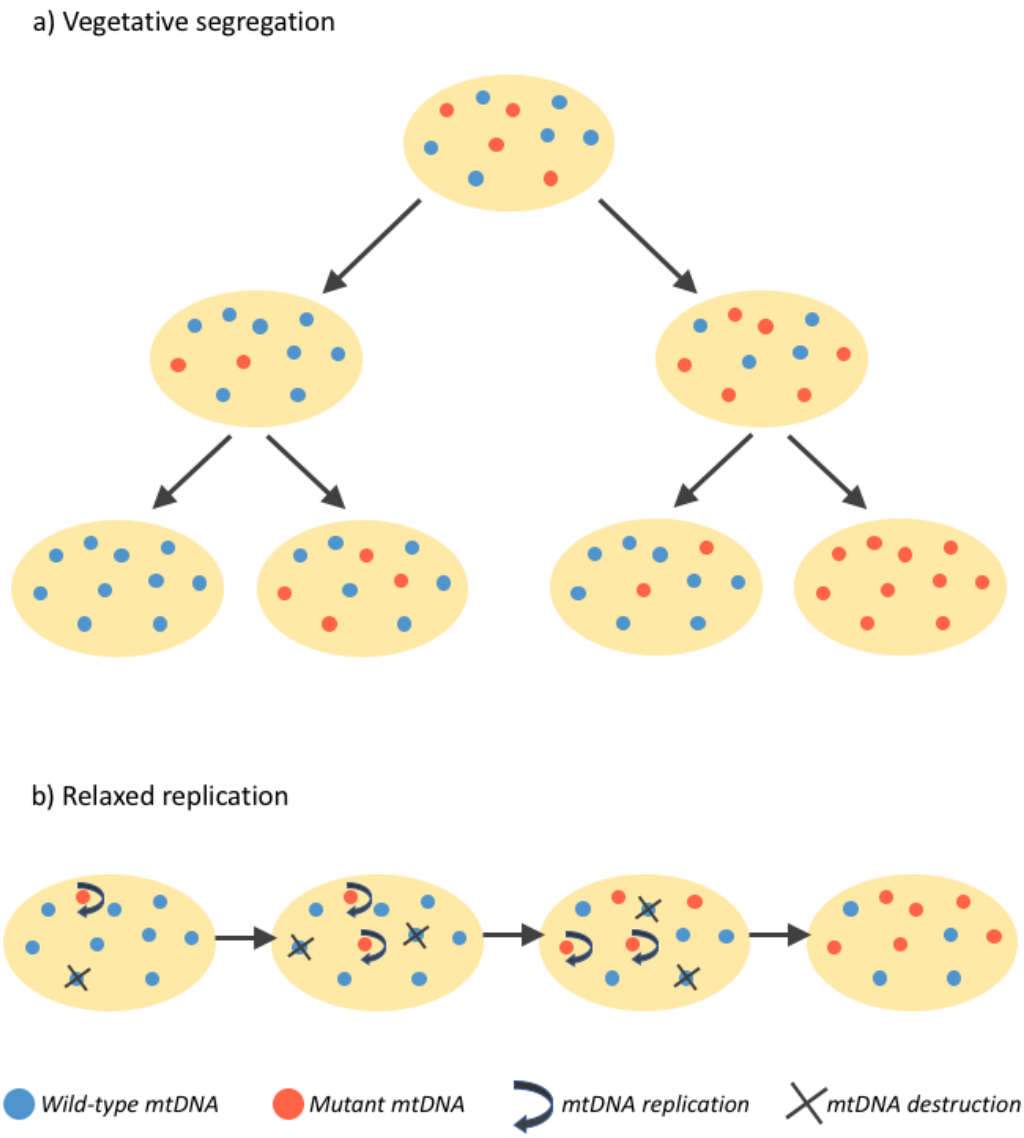


Figure 1.2 | Clonal expansion of mtDNA mutations can occur via a) vegetative segregation in which daughter cells receive different proportions of mutant mtDNA during cell division and b) relaxed replication whereby mutant mtDNA may self-replicate more frequently than wild-type mtDNA within a cell. Adapted from Stewart & Chinnery 2015.

1.5.2 Mitochondrial DNA copy number variation

During mtDNA replication, co-ordination between the degradation and replication of mtDNA helps keep numbers constant to ensure that enough energy is produced to maintain cell function (Montier et al. 2009). Nonetheless, mtDNA copy number can vary with age, although patterns are inconsistent. Some studies have found that mtDNA copy number decreases with age (e.g. human (*Homo sapiens*) skeletal muscle: Wachsmuth et al. 2016; rat (*Rattus norvegicus domestica*) muscle and liver: Barazzoni

et al. 2000; various tissues in fish (*Nothobranchius furzeri*): Hartmann et al. 2011), while other studies found positive associations (human liver samples: Wachsmuth et al. 2016) or no association with age at all (human skeletal and cardiac muscle: Miller et al. 2003; rat heart: Barazonni et al. 2000). These inconsistencies have been attributed to a number of factors including tissue differences and varying methods for quantifying mtDNA copy number (Hartmann et al. 2011; Mengel-From et al. 2014).

With a background of inconsistent results, mtDNA copy number remains to be thoroughly tested as a marker of chronological age in different animals, including crustaceans (Jarman et. al. 2015). Quantifying mtDNA copy number is relatively straightforward by comparing the amount of mtDNA to that of a reference nuclear gene using quantitative real-time PCR (Rooney et al. 2015). Reference genes are necessary to control for experimental variation but there are no perfectly constant reference genes (Jarman et al. 2015). Methods that rely on relative quantification can therefore result in higher measurement error compared to molecular markers involving absolute measurements (e.g. mtDNA heteroplasmy) (Jarman et al. 2015).

1.5.3 Genome-wide and site-specific DNA methylation

Heritable changes to the physical structure of DNA, known as epigenetic modifications, also occur with age. The most well-studied epigenetic change to DNA is methylation, in which methyl groups (CH_3) are added to DNA - a process that is thought to play a role in silencing or modifying gene expression (Schübeler 2015). Methylation usually occurs on position 5 of the nucleotide cytosine and almost exclusively where cytosine precedes guanine (CpG dinucleotides) (Jones et al. 2015).

DNA methylation is not a static modification and varies according to genetic and environmental background, as well as individual age (Jones et al. 2015). A genome-wide loss in methylation with age (epigenetic drift) has been observed in a variety of animals including scallops (*Chlamys farreri*: Lian et al. 2015), humans (Fuke et al. 2004) and chickens (*Gallus gallus domesticus*: Gryzińska et al. 2013). This is thought to occur as a result of a decline in the activity of maintenance methyltransferase enzymes (DNMT1), which are responsible for copying existing epigenetic marks during mitosis

(Jones et al. 2015; Field et al. 2018). However, at individual CpG sites, the degree of methylation has been shown to decrease (hypomethylation) or increase (hypermethylation) with age. Increasing amounts of DNA methylation can arise from the action of *de novo* methyltransferases (e.g. DNMT3b), which methylate previously unmethylated cytosines (De Paoli-Iseppi et al. 2017; Field et al. 2018).

Levels of site-specific DNA methylation have been used to predict age in a growing number of vertebrate taxa, including mammals (*H. sapiens*: Hovrath 2013; *Megaptera novaeangliae*: Polanowski et al. 2014; *Myotis bechsteinii*: Wright et al. 2018), birds (*Ardenna tenuirostris*: De Paoli-Iseppi et al. 2019) and fish (*Dicentrarchus labrax*: Anastasiadi & Piferrer 2020). For example, humpback whale (*M. novaeangliae*) age can be predicted with high accuracy ($R^2 = 0.79$; $p < 0.001$) and precision (standard deviation = 3 years) using percentage methylation at just three CpG sites (Polanowski et al. 2014). In addition, there is evidence to suggest that DNA methylation may be less influenced by environmental factors than other molecular markers. For example, DNA methylation of genes expressed in human skin changes independently of factors such as sun exposure (Grönniger et al. 2010). DNA methylation has been described as the most promising marker of animal age (Jarman et al. 2015; De Paoli-Iseppi et al. 2017) and studies that look to apply methylation-based markers to lobsters specifically have been encouraged (Silva et al. 2019).

Aside from a preliminary study that compared genome-wide methylation levels between six juveniles and six adults belonging to two crayfish species (*Procambarus fallax* and *P. virginalis*; Vogt et al. 2015), little is known about DNA methylation changes with crustacean age. Genome-wide methylation can be quantified in non-model species using a range of approaches that do not require species-specific primers or probes. These methods range from high resolution but complex and costly whole genome bisulphite sequencing, to lower resolution but easy and affordable immunoassay-based methods (Kurdyukov & Bullock 2016). Quantifying site-specific methylation can be costly and is best achieved by targeting a small region of the genome (Kurdyukov & Bullock 2016). Candidate loci that show age-related methylation changes have been characterised in some vertebrate species (e.g. Polanowski et al. 2014; Anastasiadi & Piferrer 2020) but the same loci may not be present or identifiable

in crustaceans. However, site-specific methylation has recently been shown to vary with age in ribosomal DNA (rDNA) loci in mice (Wang & Lemos 2019). This region of the genome is evolutionarily-conserved and may allow for age estimation across the animal kingdom, although this remains to be tested (Wang & Lemos 2019).

1.5.4 MicroRNA and messenger RNA expression

MicroRNAs (miRNAs) are short (between 18–25 nucleotides in length), non-coding RNAs that bind to messenger RNA molecules (mRNA) and downregulate expression of protein-coding genes (Kinser & Pincus 2020). Patterns of up- and down-regulation of miRNA expression with age have been observed in a range of vertebrate and invertebrate organisms (reviewed by Kinser & Pincus 2020), including humans (Huan et al. 2018) and nematodes (*Caenorhabditis elegans*: Pincus et al. 2011). Changes to the abundance of miRNA with age are associated with factors such as declines in physical function and age-related diseases (reviewed by Kinser & Pincus 2020). These changes have been used to create regression-based age prediction models. For example, in humans, age predicted from the expression levels of 80 miRNAs explained 49% of the variation in known age (R^2 = 0.49; p < 0.001: Huan et al. 2018). Age-related changes in expression also occur at the mRNA level. For example, mRNA levels at two genes proved useful for age determination in field-caught mosquitoes (*Anopheles gambiae*; R^2 = 0.43; p < 0.001: Wang et al. 2013).

Individual micro or messenger RNA levels have not yet been used for age determination in wild crustaceans, although miRNA expression was found to differ across different developmental stages of *Daphnia pulex* (Chen et al. 2014). Changes in miRNA expression can be assayed by qPCR, and since miRNAs have now been identified and validated in *D. pulex* (Chen et al. 2014), this should reduce the development costs of applying this method to other crustaceans, since pre-existing assays often function in close relatives (Jarman et al. 2015). However, measurement precision using qPCR can be poor owing to the need to quantify gene expression relative to a reference gene (Jarman et al. 2015; see section 1.5.2). Low measurement precision may help to explain why a large amount of variation in age (≥ 51%) is

unexplained by miRNA/mRNA-predicted age in humans (Huan et al. 2018) and mosquitoes (Wang et al. 2013).

1.6 Study system

1.6.1 European lobster (*Homarus gammarus*)

The European lobster is one of the most economically-valuable crustaceans targeted by fishers throughout Europe. The European lobster is harvested across its range from the shallow coastal areas of the northeast Atlantic Ocean (Fig. 1.3). There is pronounced population genetic structure between lobsters from the Atlantic and Mediterranean basins, but little genetic structure around the UK and Ireland (Jenkins et al. 2019). The amount of European lobster caught has been increasing since the 1980s (Fig. 1.4). Most European lobster is landed in the UK (Agnalt et al. 2009) where lobster catch is valued at more than £44 million per annum (MMO 2019), making lobsters one of the most valuable species caught in the UK by weight.

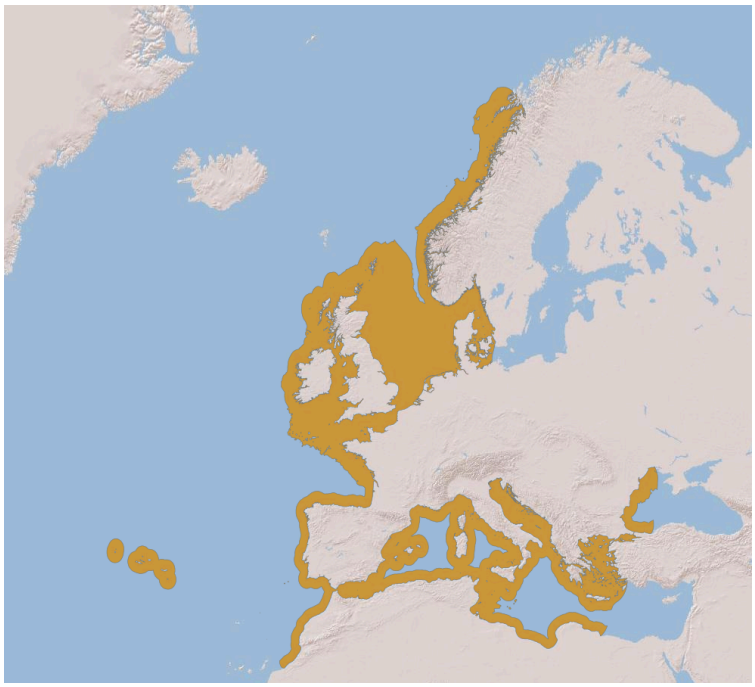


Figure 1.3 | Map of western Europe and northern Africa with the spatial distribution of the European lobster (*Homarus gammarus*) in orange. Downloaded from The IUCN Red List of Threatened Species (www.iucnredlist.org) on 11 June 2020.

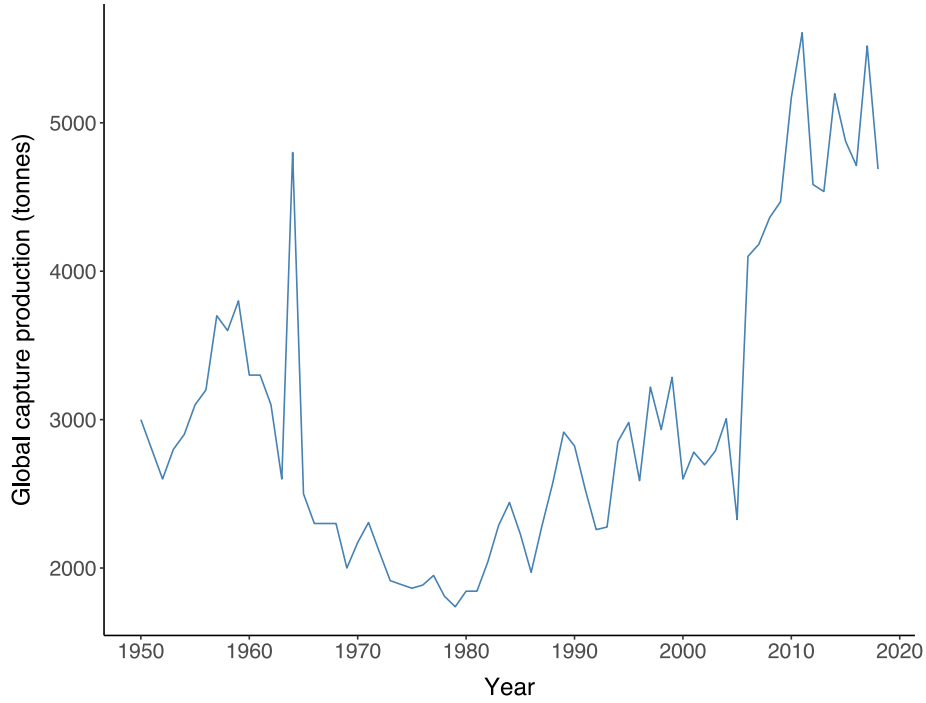


Figure 1.4 | Global capture production of the European lobster (*Homarus gammarus*) from 1950 to 2018. Data gathered by FAO Fisheries & Aquaculture.

European lobsters are primarily trapped using baited pots and are targeted by both commercial and recreational fishers. Unlike many other commercial fish stocks, European lobsters are predominantly harvested at a local scale and are therefore crucial to supporting local fishing communities and regional economies (Ellis et al. 2015). However, European lobsters are vulnerable to overexploitation and mismanagement. Historically, Norway was one of the main suppliers of European lobster to the European market but between 1930 and 1970 the stock declined by 92% and stock size and recruitment remain low (Agnalt et al. 2007; 2009). Stock declines have also occurred in Mediterranean populations (Spanier et al. 2015). Evidence that European lobsters are susceptible to overharvesting has set a precedent for ensuring that stocks are managed sustainably across their range.

Management of European lobster stocks includes limits on minimum landing size (87–90 mm carapace length depending on location). Tag-and-recapture studies suggest that lobsters reach this size after approximately four years (Bannister & Addison 1998; Uglem et al. 2005; Schmalenbach et al. 2011). These size limits are in place to prevent lobsters being removed from the fishery before they have contributed offspring. However, the size at which European lobsters reach sexual maturity is highly variable,

at least for female lobsters (Tully et al. 2011). Further measures for enhancing lobster stocks include the use of lobster hatcheries, in which lobsters are reared through their vulnerable larval stages and then released into the wild as juveniles where they contribute to fishery recruitment (Jenkins et al. 2020).

1.6.2 Red cherry shrimp (*Neocaridina davidi*)

Red cherry shrimp are freshwater shrimp native to Asia. Since their introduction to the aquarium trade in 2003, red cherry shrimp have gained popularity among aquarists due to their bright colouration, range of colour morphs and the ease at which they can be acquired commercially and reared in captivity (Nur & Christianus 2013). Red cherry shrimp are short lived (ca. 13 months in captivity: Schoolmann & Arndt 2018) with a well-characterised life cycle (Nur & Christianus 2013). They also breed year-round, can tolerate a wide range of water pH and temperature, and are easy to rear in large numbers (Mykles & Hui 2015). Consequently, red cherry shrimp are an excellent model species for studying crustacean ageing. Following an assessment of different decapod crustaceans, red cherry shrimp were advocated as an ideal model species for investigating crustacean genomics (Mykles & Hui 2015). Draft nuclear and mitochondrial genomes for red cherry shrimp have been assembled and both genomes appear to be free of extensive duplication and/or loss of genes or rearrangement (Kenny et al. 2014). This suggests that the red cherry shrimp genome can represent the genomes of other decapod species (Mykles & Hui 2015).

1.7 Aims of this thesis

The overall aim of this thesis was to test whether various genetic markers have potential for estimating age in crustaceans. The main motivation was to identify a novel, molecular method that could ultimately be used for assessing population age structure in stock assessments. To this end, potential methods must not only be strong predictors of chronological age, but also have options for scalability and be relatively easy to use and interpret. Three molecular markers of age were chosen for investigation: 1) Genome-wide methylation 2) Site-specific DNA methylation and 3) mtDNA heteroplasmy. Markers based on DNA methylation were selected because they

have so far shown to be the strongest predictors of chronological age in a range of species (e.g. De Paoli-Iseppi et al. 2017 and references therein). Heteroplasmy was chosen because surprisingly little is known about age-related changes to mtDNA mutations outside of model organisms (Jebb et al. 2018), despite the highlighted potential for heteroplasmy-based markers for predicting human age (Meissner & Ritz-Timme 2010). It is also relatively easy to screen for heteroplasmy without prior information for the target species (Jarman et al. 2015).

The specific aims of this thesis were as follows:

Chapter 2: *Are lobster and shrimp genomes methylated and, if so, are there differences in genome-wide levels with age?* Genome-wide methylation was quantified using a commercially-available, immunoassay kit in different age groups of lobsters and shrimp.

Chapter 3: *Can site-specific levels of methylation in ribosomal DNA be used to predict chronological age in European lobsters?* Site-specific DNA methylation was quantified in lobster ribosomal DNA using bisulphite sequencing and the ability of a small subset of the screened loci to predict lobster age was assessed.

Chapter 4: *Can mitochondrial heteroplasmy be used to predict chronological age in red cherry shrimp?* The mitochondrial genomes of multiple shrimp were amplified and sequenced using NGS. After filtering of likely false positives, the number of mtDNA mutations in 39% of the genome was compared across seven age cohorts.

Chapter 5: *Can mitochondrial heteroplasmy be used to predict chronological age in European lobsters?* The mitochondrial genomes of multiple lobsters were amplified and sequenced using NGS. Most of the dataset was affected by spurious sequencing reads and a range of bioinformatic approaches aimed at removing false positive mtDNA mutations was attempted. After thorough filtering, the number of mtDNA mutations across 48% of the genome was compared across six age cohorts.

A synthesis of the findings presented in this thesis is provided in **chapter 6**. This chapter also outlines avenues for future research and highlights the areas that should be prioritised for further development.

Chapter 2

Occurrence and age-related patterns of global DNA methylation in two decapod crustaceans (*Homarus gammarus* and *Neocaridina davidi*)

2.0 Abstract

Estimating age in crustaceans is complicated but is crucial for predicting population dynamics for the sustainable management of economically-important fisheries. Age-related changes in DNA methylation occur in many vertebrates, but very little is known about this phenomenon in crustaceans. To be practical for estimating age in stock assessments, molecular ageing methods must be relatively cheap, simple and preferably rapid. Here, genome-wide (global) levels of methylation were measured using a commercially-available ELISA kit in different age groups of the economically-important European lobster (*Homarus gammarus*) and laboratory-reared cohorts of red cherry shrimp (*Neocaridina davidi*). Both lobsters and shrimp were found to have methylated genomes. Global DNA methylation did not differ between larval (1 day old) and adult (≥ 4 years old) lobsters. Age had a significant effect on global DNA methylation across shrimp life-cycle stages (eggs, larvae, juveniles and adults) but this effect was driven by the larval cohort, which had significantly higher methylation levels than the juvenile and adult age groups. However, the shrimp data were confounded by high variability in the methylated standards supplied in the kit for quantifying DNA methylation. These results suggest that global DNA methylation, measured using an ELISA kit, is not an informative and/or reliable marker of chronological age in crustaceans.

2.1 Introduction

Global catches of crustaceans (mainly lobsters, shrimp and crabs) have been increasing since the 1950s (Anderson et al. 2011) and account for 7% of marine capture by weight and 22% of fish trade value (FAO 2020). Nevertheless, concerns have been raised over the long-term sustainability of crustacean fisheries, which are largely unassessed and unregulated owing to a lack of ecological and life-history data (Anderson et al. 2011; Boudreau & Worm 2012). Specifically, information on population age structure is crucial for predicting population dynamics and managing sustainable fisheries (Campana 2001). However, crustaceans are difficult to age because they moult throughout their lives and there is extensive individual variation in size-at-age (Vogt 2012; Kilada & Driscoll 2017).

Molecular makers of chronological age are an increasingly-attractive tool in molecular ecology and have shown great promise for estimating age in a number of wild animal systems (Jarman et al. 2015; De Paoli-Iseppi et al. 2017). These involve measuring a feature of DNA or RNA, or their associated molecules, that changes consistently with age among individuals. The most widely-tested molecular markers of animal age include change in average telomere length, the frequency of mitochondrial DNA mutations, and epigenetic alterations such as DNA methylation (Jarman et al. 2015; De Paoli-Iseppi et al. 2017). For example, average telomere length decreases significantly with age in some bird species, e.g. Seychelles warblers (*Acrocephalus sechellensis*: Barrett et al. 2013), and percentage DNA methylation at just three loci predicts the age of Humpback whales (*Megaptera novaeangliae*) with an average accuracy of 3 years (Polanowski et al. 2014). Despite this progress, molecular-based age determination has not yet been successfully applied to crustaceans. One study that did investigate telomere length in crustaceans, found no relationship with age in either long-lived (*Sagmariasus verreauxi* and *Jasus edwardsii*) or short-lived (*Metapenaeus macleayi*) species (Godwin et al. 2011).

DNA methylation is the addition of methyl groups (CH_3) to DNA, which is thought to play an important role in silencing gene expression (Schübeler 2015). DNA methylation

typically occurs on position 5 of the nucleotide cytosine, resulting in 5-methylcytosine (5-mC), and almost exclusively where cytosine precedes guanine (CpG dinucleotides) (Jones et al. 2015). It is a dynamic process that can vary among species and tissues, according to genetic and environmental variation, and with individual age (Jones et al. 2015). In vertebrates, most changes appear to occur during early-life when methylation patterns are established by methyltransferase enzymes (Jones et al. 2015). Vertebrate methylation is at its lowest level in early embryo development before increasing in later developmental stages (Mhanni & McGowan 2004; Zeng & Chen 2019) and the first few years of life (Herbstman et al. 2013). Thereafter, methylation patterns are not consistently maintained resulting in a genome-wide (global) loss over time (Jones et al. 2015). Global DNA methylation declines with advancing age in a range of animals, including humpback salmon (*Oncorhynchus gorbuscha*: Berdyshev et al. 1967), Zhikong scallops (*Chlamys farreri*: Lian et al. 2015), humans (*Homo sapiens*: Fuke et al. 2004), mice (*Mus musculus*: Wilson et al. 1987) and chickens (*Gallus gallus domesticus*: Gryzińska et al. 2013). These studies suggest that global DNA methylation may be a useful proxy marker of animal age for a variety of species.

Any method for estimating age in fisheries management needs to be cheap, quick and simple enough to be applied to large numbers of individuals, and across diverse species and laboratory settings. To this end, a marker of age based on levels of global DNA methylation appears to be a favourable option; methods exist for measuring global DNA methylation that are quick, accessible and affordable, and do not require prior genetic information for the target species. Whole genome bisulphite sequencing is the most comprehensive and sensitive approach to quantifying global DNA methylation but costs several thousand pounds per sample (Kurdyukov & Bullock 2016). HPLC-UV (high performance liquid chromatography-ultraviolet) also provides high resolution but requires specialist equipment, technical expertise and large amounts of DNA (Kurdyukov & Bullock 2016). Alternatively, enzyme-linked immunosorbent assays (ELISAs) simply require a microplate reader and one of the several commercially-available kits (Kurdyukov & Bullock 2016). Therefore, ELISAs are

an attractive choice for ecological studies for which funding and access to well-equipped laboratories may be limited.

Patterns of global DNA methylation have been well-characterised in vertebrates but there is a paucity of data for invertebrates (Lian et al. 2015). Vertebrate genomes are almost always highly methylated, whereas invertebrate genomes appear to be more variable in their methylation state (Lian et al. 2015). DNA methylation has been studied in a handful of crustaceans, with global levels ranging from no detectable methylation in brine shrimp (*Artemia* sp.: Warner & Bagshaw 1984), 0.05% in prawns, (*Macrobrachium rosenbergii*: Feng et al. 2007) to 3% in crayfish (*Procambarus fallax*: Vogt et al. 2015). Almost nothing is known about age-related patterns of DNA methylation in crustaceans; global DNA methylation did not vary with age in two crayfish species (*P. fallax* and *P. virginalis*) but this was based on two age groups (juveniles vs. adults) with very small sample sizes (n = 1–3) per group (Vogt et al. 2015).

This study aims to investigate whether global DNA methylation can be used as a proxy marker of age in European lobsters (*Homarus gammarus*) and red cherry shrimp (*Neocaridina davidi*). The European lobster is a commercially-important species, harvested in shallow coastal areas across the northeast Atlantic. The fishery is valued at more than £44 million per annum to the UK alone (MMO 2019). This species has an estimated lifespan of 42–72 years (Sheehy et al. 1999) and is rarely (if ever) maintained in captivity across its lifespan. This makes obtaining DNA samples across a wide-range of known age individuals extremely difficult. Therefore, this study also used captive-reared cohorts of red cherry shrimp, which are short-lived, breed year-round and are easy to rear in large numbers, making them an excellent model system for studying crustacean ageing. Global DNA methylation was quantified using a commercially-available ELISA kit and levels were compared between larval (1 day old) and adult (≥ 4 years) European lobsters, and among all life cycle stages (eggs, larvae, juveniles, and adults; Nur & Christianus 2013) of red cherry shrimp.

2.2 Methods

2.2.1 Study species and sampling

Whole claws were sampled from larval lobsters (*n* = 10) reared at the National Lobster Hatchery in Cornwall (UK) one day post-hatching. Adult lobsters (*n* = 10) caught off the coast of Cornwall were sampled by clipping the terminal end of an antenna. All wild-caught lobsters were above minimum landing size (88–137 mm carapace length (CL)) and therefore estimated to be mature individuals ≥ 4 years of age. This estimated minimum age is based on size-at-age data from previous mark-recapture studies (Bannister & Addison 1998; Uglem et al. 2005; Schmalenbach et al. 2011).

Red cherry shrimp were sourced from Neil Hardy Aquatica Ltd and reared at the University of East Anglia, UK. Known-age cohorts were established by placing 5–7 berried (egg-carrying) females from broodstock aquaria into separate cohort aquaria. Females fertilised within eight days of each other were placed in the same cohort tank. Mothers were returned to the broodstock following hatching of the eggs. Cohort age was calculated as days post-fertilisation of the first female added to that aquarium and therefore represents maximum age. To maintain the known age of each cohort, shrimp that subsequently became pregnant within cohorts were temporarily placed in holding tanks until egg hatching and then returned to their respective aquarium. Whole legs were sampled from larval, juvenile and adult shrimp 40 (± 5), 75 (± 8) and 120 (± 8) days post-fertilisation, respectively, spanning ca. 30% of the expected lifespan (Schoolmann & Arndt 2018). For each age category, a random subset of 25 shrimp was sampled from one of the cohort tanks. Whole eggs (*n* = 25) were sampled from a single berried female 7 days post-fertilisation. All aquaria contained 50% de-chlorinated tap water and 50% RO water maintained at 22–24^{o}C. Shrimp were fed *ad libitum* with commercial fish wafers (Tetra, UK). All lobster and shrimp tissue samples were stored in separate vials of absolute ethanol at 4^{o}C.

2.2.2 DNA extraction and global DNA methylation quantification

Genomic DNA was extracted from ca. 2 mm^3 of tissue excised from within the lobster claws or antennae, or from whole shrimp legs or eggs, using a salt-precipitation protocol (modified from Aljanabi & Martinez 1997) and resuspended in H_2O. DNA concentration and purity were verified using a NanoDrop 8000 Spectrophotometer (Thermo Scientific) and extracts were diluted to 25 ng/µl with ddH_2O.

Genome-wide methylation was determined using the MethylFlash™ Global DNA Methylation (5-mC) ELISA Easy Kit (Epigentek) according to the manufacturer's protocol. The ELISA kit utilises a 5-mC-specific antibody, which was quantified colourimetrically by reading absorbance (OD) at 450 nm using a microplate reader (BMG Labtech Ltd, UK) (Fig. 2.1). Absorbance at 450 nm is directly proportional to the percentage of methylated DNA.

A standard curve was generated for each plate by plotting the mean OD for a dilution series (0.1, 0.2, 0.5, 1.0, 2.0 and 5.0%) made from a methylated DNA standard supplied in the kit. A negative control (0% methylation standard supplied in the kit) was also included. All standards and lobster/shrimp DNA were measured in duplicate and analyses performed on the mean. A polynomial second order regression was fitted to each standard curve in accordance with the manufacturer's protocol. Global DNA methylation was subsequently calculated using equation (1), where *b* is slope 2 and *a* is slope 1 of the standard curve, *Y* = delta OD (standard OD - negative control OD), and *S* is the amount of input DNA in ng.

$$Percentage\ methylation = \frac{\left(b^2 + 4aY\right)^{0.5} - b}{2a} \div S\ x\ 100 \qquad (1)$$

2.2.3 Assessing ELISA kit performance

To evaluate intra-plate measurement precision, the coefficient of variation (CV) in delta OD was calculated for each technical replicate. Repeatability between technical replicates was further assessed through Pearson's correlation using 'cor.test{stats}' (R Core Team 2017a).

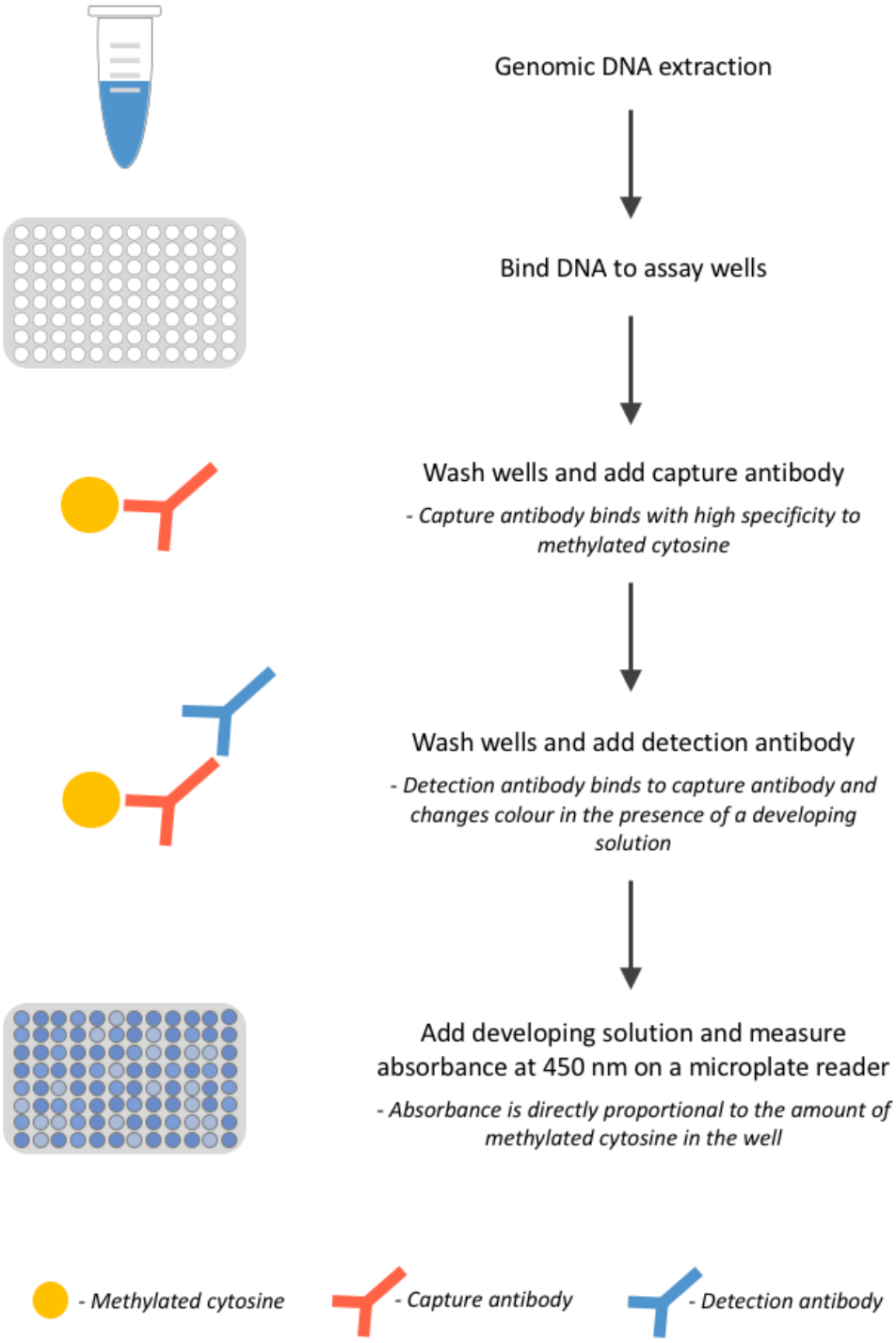


Figure 2.1 | Overview of protocol for quantifying global DNA methylation using the MethylFlash™ Global DNA Methylation (5-mC) ELISA Easy Kit by Epigentek.

It is not possible to extrapolate percentage DNA methylation from the standard curve using polynomial regression for samples that fall outside of the range of the standards. Therefore, any DNA sample with a delta OD reading higher than the highest methylated standard (5%) was removed from the dataset. Three plates (kits) were used in total, belonging to two kit manufacturing batches (Plate 1 = batch no. 8091208; Plates 2 and 3 = batch no. 8110003). To assess inter-plate variability, a 'golden sample' (DNA from the same lobster) was included on each plate to calculate the average CV across plates. All lobster DNA was measured on plate 1 and delta OD for all individuals (n = 20) fell within the range of the standard curve and could be retained for downstream analysis. Shrimp DNA was measured across plates 1–3 with DNA from shrimp of different ages randomly distributed across plates to minimise plate effects. For plate 1, delta OD for most shrimp DNA (n = 16/20; 80%) was lower than the highest methylated standard. However, only a small number of shrimp had a delta OD reading within the standard curve range for plates 2 (11/40; 27.5%) and 3 (3/40; 7.5%).

2.2.4 Interrogating issues with shrimp DNA falling beyond the standard curve range

To investigate why the delta ODs for the majority of shrimp fell beyond the standard curve range for the second batch of plates (plates 2 and 3) and not the first (plate 1), delta OD was compared across plates for the standards and shrimp DNA separately. There are two possible reasons for the inter-plate differences:

1) Delta OD for the supplied standards was lower for plates 2 and 3 compared to plate 1, but raw absorbance was consistent across plates for shrimp DNA (i.e. the standards were different in each kit, but the raw ODs for the shrimp samples were consistent)
2) Delta OD for the standards was consistent across plates, but higher for the shrimp DNA on plates 2 and 3 compared to plate 1. This would suggest that the shrimp on plates 2 and 3 contain levels of DNA methylation larger than the highest standard (5% methylation)

Generalized linear mixed-effects models (GLMMs) were fitted using 'glmer{lme4}' (Bates et al. 2015) with a Gaussian error distribution and an 'identity' link function with

plate number as a fixed predictor and delta OD as the response. Percentage DNA methylation of the standard or shrimp age were included as nested random effects when comparing standards and shrimp DNA, respectively. For both models, inclusion of a random effect resulted in a significantly lower model Akaike's Information Criterion ($p < 0.01$).

2.2.5 Re-calibrating standard curves

Delta OD was consistent among plates for shrimp DNA but not for the standards (see results section 2.3.1; Fig. 2.2). The inconsistency in standards among plates appeared to be linked to different kit batches. Delta ODs for plate 2 and 3 standards (same batch) were lower than plate 1 standards (different batch). In an attempt to recover some data, the standards on plates 2 and 3 were re-calibrated using the 5% methylated standard on plate 1. This increased the absorbance values of plate 2 and 3 standards to minimise the number of DNA samples falling beyond the standard curve range. Specifically, re-calibrated delta OD of the standards was calculated using equation (2) where 1.2055 is delta OD for the 5% methylated standard on plate 1.

$$Re\text{-}calibrated\ standard\ delta\ OD = \frac{standard\ delta\ OD}{(delta\ OD\ of\ 5\%\ standard \div 1.2055)} \qquad (2)$$

2.2.6 Global DNA methylation and age

Independent samples *t*-tests were performed using 't.test{stats}' to compare global DNA methylation between lobster age groups, and between small (88–97 mm CL) and large (104–137 mm CL) wild lobsters. There was large size variation among the wild lobster cohort and these individuals may differ markedly in age (Sheehy et al. 1999).

To compare global DNA methylation among the four age groups of shrimp, a GLMM was fitted using 'glmer{lme4}' with a Gamma error distribution and a 'log' link function. Age was included as a fixed predictor. Plate ID was included as a nested random effect to account for inter-plate variability. Inclusion of Plate ID as a random effect resulted in a significantly lower model Akaike's Information Criterion ($p < 0.001$). Analyses were performed with/without eggs to control for tissue differences.

To test the effect of intra-plate variation, analyses of age/size were repeated for lobster and shrimp DNA samples with a CV in delta OD between technical replicates ≤ 10% (a commonly accepted threshold).

2.2.7 Statistics

Statistical analyses were performed in R (version 3.3.2; R Core Team 2020) using RStudio (version 1.0.136; RStudio Team 2016). For independent samples *t*-tests, normality and variance were assessed using 'shapiro.test{stats}' and 'leveneTest{car}' (Fox & Weisberg 2019). For GLMMs, error distributions and link functions were selected by examining residual and normality plots using 'Plot{graphics}' (R Core Team 2017b). Model AICs with and without random effects were compared using 'anova{stats}'. Model main effects were assessed through 'Anova{car}'. Post-hoc pairwise Tukey comparisons were performed using 'lsmeans{lsmeans}' (Lenth et al. 2016). Plots were generated using 'ggplot{ggplot2}' (Wickham 2016).

2.3 Results

2.3.1 ELISA kit performance

Delta OD for the standards differed among plates (Gaussian GLMM: $\chi^2_{(2,18)} = 22.297$; $p < 0.001$; Fig. 2.2a). Delta OD of the standards was significantly lower on plate 2 (Tukey's test: $t = 2.762$; $p = 0.048$) and plate 3 (Tukey's test: $t = 4.698$; $p = 0.002$) compared to plate 1, but did not differ between plates 2 and 3 (Tukey's test: $t = 1.935$; $p = 0.179$). Re-calibrating the standard curves removed the variation in standard delta OD across plates (Gaussian GLMM: $\chi^2_{(2,18)} = 4.873$; $p = 0.087$; Fig. 2.2b). Delta OD was consistent among plates for shrimp DNA (Gaussian GLMM: $\chi^2_{(2,100)} = 0.247$; $p = 0.884$; Fig. 2.2c).

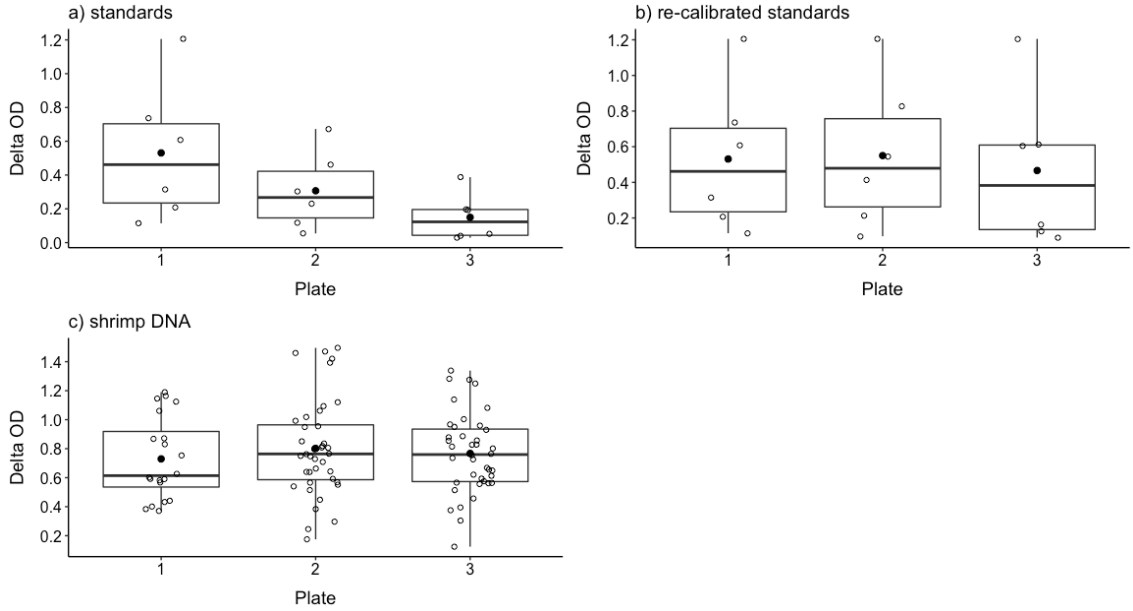


Figure 2.2 | Delta OD (absorbance at 450 nm) for a dilution series of methylated standards (a and b) and red cherry shrimp (*Neocaridina davidi*) DNA (c) across three ELISA plates prepared using the MethylFlash™ Global DNA Methylation (5-mC) ELISA Easy Kit (Epigentek). Plates 2 and 3 were from the same kit batch. Standards on plates 2 and 3 were re-calibrated (b) using the 5% methylated standard on plate 1 following equation (2).

Polynomial regressions fitted through the re-calibrated standard curves had an average R^2 across plates of 0.96 (Fig. 2.3). Average intra-plate CV was 13.3%. Absorbance between technical replicates was significantly correlated (Pearson's r = 0.73, $p < 0.001$). Average inter-plate CV was 61.7% and 19.7% for original and re-calibrated standards, respectively (Table 2.1) and 19.6% for the golden sample.

Table 2.1 | Delta OD (absorbance at 450 nm) and inter-plate coefficient of variation (CV) for a dilution series of methylated standards across three ELISA plates prepared using the MethylFlash™ Global DNA Methylation (5-mC) ELISA Easy Kit (Epigentek). Plates 2 and 3 were from the same kit batch. Standards on plates 2 and 3 were re-calibrated using the 5% methylated standard on plate 1 following equation (2).

	Original standard curve				Re-calibrated standard curve			
Percentage methylation	Delta OD				Delta OD			
	Plate 1	Plate 2	Plate 3	CV (%)	Plate 1	Plate 2	Plate 3	CV (%)
0.1	0.115	0.055	0.041	56.3	0.115	0.098	0.126	12.6
0.2	0.208	0.118	0.029	75.5	0.208	0.212	0.090	40.6
0.5	0.316	0.231	0.052	67.4	0.316	0.414	0.162	42.8
1.0	0.608	0.304	0.197	57.8	0.608	0.544	0.611	6.4
2.0	0.736	0.462	0.194	58.4	0.736	0.828	0.604	15.6
5.0	1.206	0.673	0.388	55.0	1.206	1.206	1.206	0.0
			Average =	61.7			Average =	19.7

2.3.2 Global DNA methylation and age

Global DNA methylation did not differ between young and old lobsters (independent samples t-test: t = 0.314; p = 0.757; Fig. 2.4), or between small and large wild lobsters (independent samples t-test: t = 0.333; p = 0.748; Fig. 2.4).

Global DNA methylation was not compared among shrimp cohorts using the raw data because only a small number of individuals (30/100) had a delta OD within the standard curve range (Fig. 2.5a): 11 eggs, 1 larva, 6 juveniles and 12 adults. After re-calibrating the standard curves, 86 individuals remained: 23 eggs, 18 larvae, 25 juveniles, and 20 adults. Using the re-calibrated standard curves, age had a significant effect on global DNA methylation in shrimp (Gamma GLMM: $\chi^2_{(3,86)}$ = 13.656; p = 0.003; Fig. 2.5b). Shrimp larvae had significantly higher levels of global DNA methylation

compared to adults (Tukey's test: $z = 3.628$; $p = 0.002$). No other pairwise comparisons were significant (Tukey's test: $z \leq 2.550$; $p \geq 0.053$). When excluding shrimp eggs, global DNA methylation still differed among the remaining shrimp age groups (Gamma GLMM: $\chi^2_{(2,75)} = 13.197$; $p = 0.001$). This effect was again driven by larval shrimps, which had higher levels of global DNA methylation compared to juveniles (Tukey's test: $z = 2.378$; $p = 0.046$) and adults (Tukey's test: $z = 3.602$; $p = 0.001$). Global DNA methylation did not differ between juveniles and adults (Tukey's test: $z = 1.486$; $p = 0.297$). The main results were not affected after removing lobster/shrimp DNA samples with a CV between technical replicates > 10% (Supplementary Table S2.1).

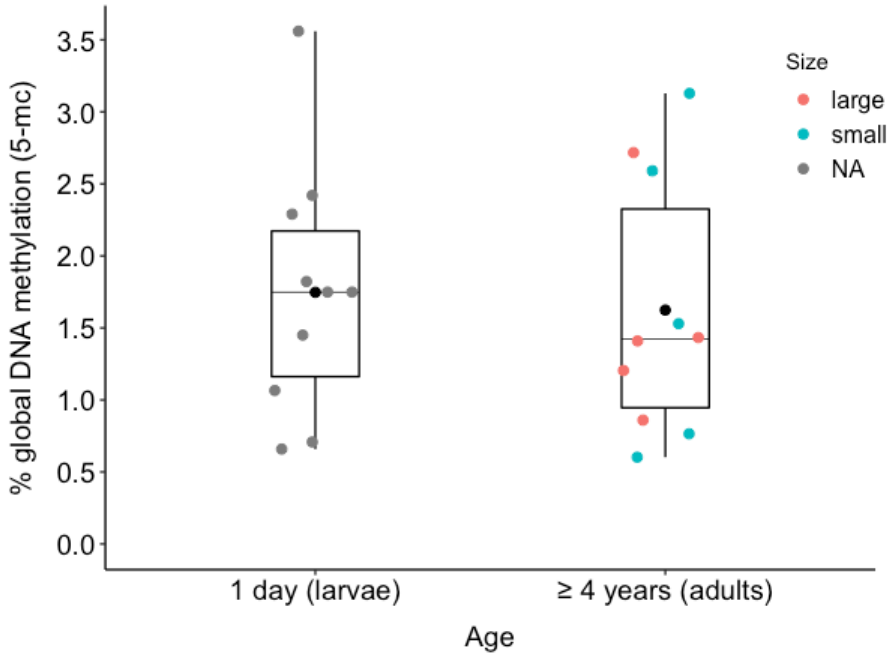


Figure 2.4 | Percentage global DNA methylation in different age groups of the European lobster (*Homarus gammarus*). Size according to carapace length: large = 104–137 mm, small = 88–97 mm, NA = data unavailable. Boxplots display a mean dot (solid black point), median line, inter-quartile range (IQR) boxes, 1.5*IQR whiskers, and data points.

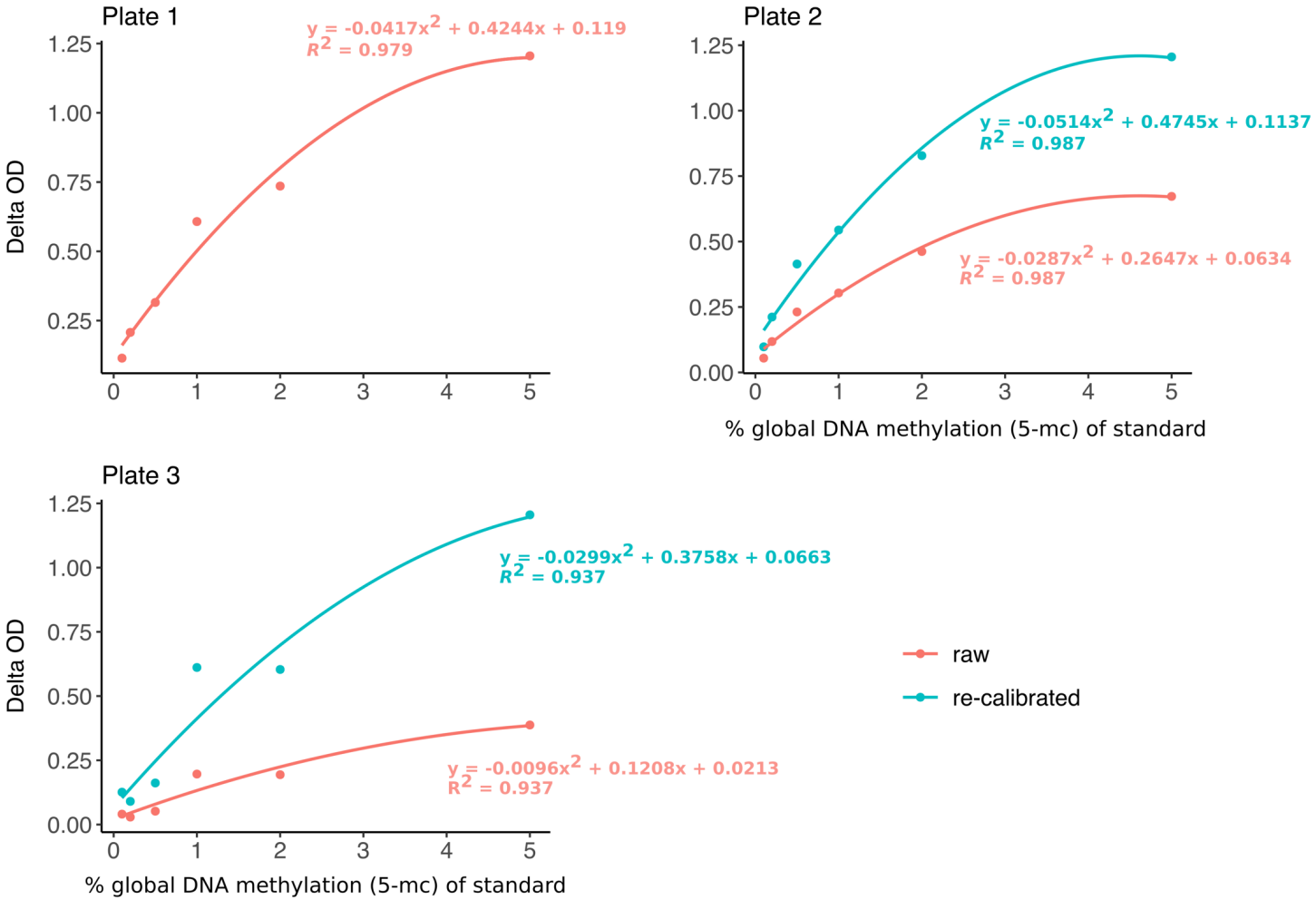


Figure 2.3 | Standard curves generated by plotting the mean delta absorbance (OD) at 450 nm for a dilution series made from a methylated DNA standard. Fitted lines are polynomial second order regressions for raw (red) and re-calibrated (blue) delta OD.

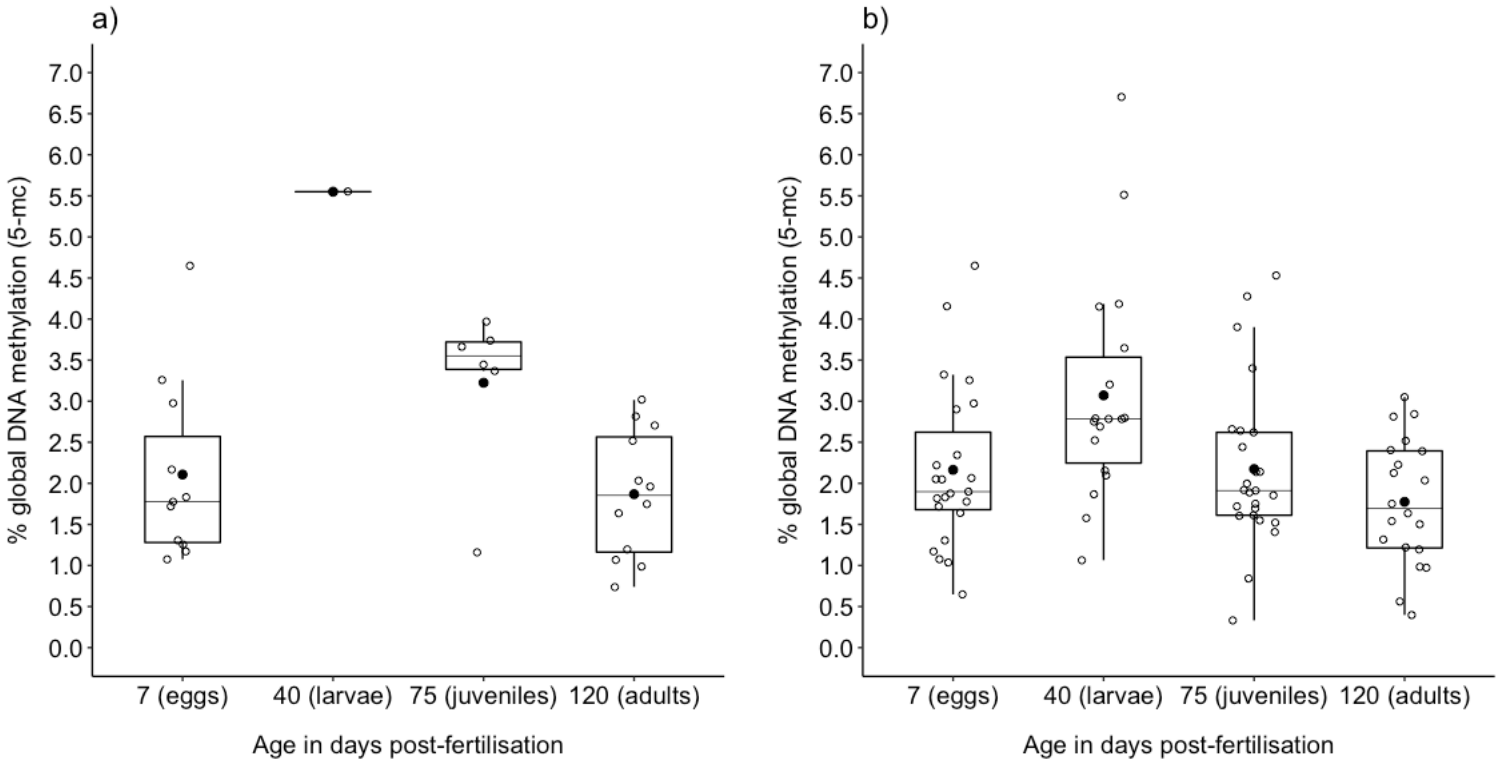


Figure 2.5 | Percentage global DNA methylation in different age groups of red cherry shrimp (*Neocaridina davidi*) with raw (a) and re-calibrated (b) standard curves. Boxplots display a mean dot (solid black point), median line, inter-quartile range (IQR) boxes, 1.5*IQR whiskers, and data points.

2.4 Discussion

This study investigated whether global DNA methylation, measured using a rapid and cost-effective ELISA kit, could be used to differentiate age groups of European lobsters and red cherry shrimp. Both species were confirmed to have methylated genomes, adding to our limited understanding of DNA methylation in invertebrates. Global DNA methylation did not differ between hatchery-reared larvae and wild-caught adults of the European lobster. Comparing shrimp age groups was compromised by high inter-plate variation in the standard curves. After re-calibrating the standard curves, global DNA methylation was found to differ among four known-age shrimp cohorts (sampled from eggs to adults) reared in aquaria. However, this effect was driven by the larval shrimp cohort, which had significantly higher levels of global DNA methylation compared to juveniles and adults. Global DNA methylation did not differ between juvenile and adult shrimp. These results, combined with kit reliability issues, suggest global DNA methylation measured using a commercially-available ELISA kit is not a viable ageing tool for crustaceans.

The absence of age-related variation in global DNA methylation in European lobsters is unusual among animals; age-related declines in global DNA methylation have been observed in humpback salmon (Berdyshev et al. 1967), humans (Fuke et al. 2004), mice (Wilson et al. 1987), Zhikong scallops (Lian et al. 2015) and chickens (Gryzińska et al. 2013). However, the current study is in-line with previous crustacean research, which found no difference in global DNA methylation between two age groups (juvenile vs. adults) of two crayfish species (*P. fallax* and *P. virginalis*: Vogt et al. 2015). In addition, a recent study on birds did not observe significant age-related differences in global DNA methylation (*Ardenna tenuirostris*: De Paoli-Iseppi et al. 2019). This suggests that relationships between global DNA methylation and age are not universally found across taxa.

The lack of signal identified in the current study could be related to the tissue used for study. Due to sampling constraints on adult tissue, European lobster DNA was extracted from larval claws but antennae had to be sampled from adults. Some tissue

dependency is thought to exist in age-related DNA methylation, although large-scale tissue comparisons are scarce (Slieker et al. 2018). A multi-tissue comparison of global DNA methylation was performed in different age groups of Zhikong scallop (Lian et al. 2015). Declines in global DNA methylation were observed in all scallop tissues investigated but this relationship was only significant in soft tissue, mantle, kidney and hepatopancreas, and not gill, gonad or adductor muscle. For this study, antennae were provided from wild-caught, adult lobsters because antennae can be non-destructively sampled without commercial loss to the catch, but antennae are not yet developed in early larval stages of European lobsters (Rötzer et al. 2015). It is possible that age-related changes occur in other crustacean tissues but no differences in global DNA methylation were observed across three diverse tissue types (muscle, hepatopancreas, and ovary) in *P. fallax* and *P. virginalis* crayfish (Vogt et al. 2015).

While it may not be possible to definitively compare larval and adult European lobsters due to tissue differences, the results from this study suggest that any age-related changes in global DNA methylation in European lobsters are minimal, at least in adults. The adult lobsters sampled in this study were estimated to be at least 4 years old based on CL (Bannister & Addison 1998; Uglem et al. 2005; Schmalenbach et al. 2011). Due to the large variation in CL among the wild-caught individuals sampled in this study (88–137 mm), the adult lobsters are likely to differ substantially in age (Sheehy et al. 1999). For example, using a calibration dataset of lipofuscin content in a set of known age European lobsters, a CL between 85–120 mm was estimated to equate to lobsters between 4 and 20 years old. Variation in size-at-age increased with age and became particularly extreme for individuals with a CL > 130 mm (Sheehy et al. 1999). In this study, global DNA methylation did not differ between small (88–97 mm CL) or large lobsters (104–137 mm CL) despite the potentially large difference in age. Therefore, it is unlikely that global DNA methylation will be a useful ageing tool in European lobsters.

The laboratory-reared red cherry shrimp made it possible to investigate global DNA methylation changes across ca. 30% of the species' lifespan, encompassing every life cycle stage, and under controlled environmental conditions. With the exception of the

shrimp larvae, no significant differences in global DNA methylation were observed among age groups. Larval shrimp had significantly higher levels of global DNA methylation compared to juveniles and adults. Early-life increases in genome-wide methylation have been observed in several vertebrates (Mhanni & McGowan 2004; Herbstman et al. 2013; Zeng & Chen 2019) but little, if anything, is known about this phenomenon in invertebrates. Adult shrimp had lower levels of global DNA methylation compared to juveniles but this difference was not significant. However, quantifying global DNA methylation in shrimp was complicated by having measured individuals across three plates with high inter-plate variability in the standard curves (mean inter-plate CV in standards = 61.7% before re-calibration). It is possible that a more sensitive method for quantifying global DNA methylation would detect a significant difference between juvenile and adult shrimp.

ELISA-based methods for quantifying global DNA methylation are quick, affordable, and can be applied to large sample sizes, making them an attractive choice for use in fisheries laboratories and ecological studies. On the other hand, intra- and inter-plate variability can lead to low measurement precision and poor resolution (Kurdyukov & Bullock 2016). However, subtle differences are of less importance when developing an ageing assay, which depends on being able to detect pronounced differences that give rise to distinct, ecologically-informative age classes. Nevertheless, in this study, both intra- (13.3%) and inter-plate CV (19.6%) were high. However, exclusion of samples with a CV > 10% between technical replicates did not affect the main findings (Supplementary Table S2.1). It is difficult to gauge how the reliability of these data compares to other studies; intra-plate CV is often not reported (e.g. Gomes et al. 2012; Fang et al. 2013; Lian et al. 2015; Gryzińska et al. 2016; Guarasci et al. 2018) but high intra-plate CV has been observed (in some cases ≥ 40%: Gryzińska et al. 2013; Agrelius et al. 2018). In addition, despite the importance of controlling for plate-to-plate variability, no mention of inter-plate CV was found in equivalent studies. However, concerns over the reliability of the ELISA kit used in this study have been raised by other researchers on ResearchGate and within doctoral theses (Price 2009; Lisanti 2013), suggesting that the problems encountered in this study are not uncommon.

The main conclusions from this study are that global DNA methylation did not differ between European lobster age groups, showed minimal variation in European lobster adults and was similar across each life cycle stage of red cherry shrimp except larvae. Furthermore, this study raises doubts about the reliability of ELISA-based methods for quantifying global DNA methylation. Any absence of age-related patterns of global DNA methylation does not rule out the potential for age-dependent changes in site-specific DNA methylation in these species. Given the promising results regarding epigenetic ageing 'clocks' based on quantifying methylation at a few CpG loci in other animals (De Paoli-Iseppi et al. 2017), site-specific DNA methylation now provides arguably the best way forward for further investigation in crustaceans.

2.5 Supplementary information

Supplementary Table S2.1 | Relationship between age or size and global DNA methylation in European lobsters (*Homarus gammarus*) and red cherry shrimp (*Neocaridina davidi*) when excluding samples that had a coefficient of variation in delta absorbance between technical replicates > 10%. Lobster age and size groups were compared using an independent samples *t*-test. A generalized linear-mixed effect model was fitted to investigate the effect of age on global DNA methylation in shrimp (Gamma GLMM: $\chi^2_{(2,39)} = 7.483$; $p = 0.024$) with plate ID as a nested random effect. Post-hoc model outputs shown below. Significant contrasts in bold and underlined.

Species	Contrasts	Estimate	SE	*t/z* ratio	*p*
H. gammarus	Larvae vs Adults	-	-	-0.572	0.581
H. gammarus	Small vs Large wild lobsters	-	-	0.944	0.399
N. davidi	Larvae vs Juveniles	0.465	0.215	2.163	0.078
	Larvae vs Adults	**0.583**	**0.221**	**2.637**	**0.023**
	Juveniles vs Adults	0.118	0.195	0.604	0.818

Chapter 3

Ageing European lobsters (*Homarus gammarus*) using DNA methylation of evolutionarily-conserved ribosomal DNA

3.0 Abstract

Crustaceans are extremely difficult to age because of their indeterminate growth and the moulting of their exoskeleton throughout life. The consequent lack of information on population age structure in crustaceans prevents accurate assessment of population dynamics and, therefore, seriously hampers sustainable fisheries management. A recent genomic study showed that DNA methylation of the evolutionarily-conserved ribosomal DNA (rDNA) may allow for age prediction across diverse species. However, the rDNA epigenetic clock remains to be tested in wild animals, despite its potential to inform ecological and evolutionary understanding, and conservation and management practices. Here, patterns of methylation with age were measured across 5,154 bp of rDNA sequence in the economically-important European lobster. There was a significant linear relationship between percentage rDNA methylation in claw tissue and age at 67% of quality-filtered loci (n = 237) across 0–24.8-months-old lobsters (n = 133). A multiple linear regression based on just ten loci allowed for the accurate and precise age estimation of individuals (R^2 = 0.911, standard deviation = 2.024 months). Applying this ageing model to antennal DNA from wild lobsters of unknown age (n = 38) resulted in estimated ages that were lower than expected based on size-at-age data. This may be caused by tissue-specificity of rDNA methylation, overfitting of the model to known-age lobsters, or non-linear relationships between percentage rDNA methylation and age in older individuals. Further validation of this ageing tool is required across a wider range of known-age individuals and multiple tissues.

3.1 Introduction

Information on the age structure of animal populations is fundamental to understanding their ecology, evolution and conservation (Jarman et al. 2015; De Paoli-Iseppi et al. 2017). Animal age can be used to predict mortality risk, reproductive potential and susceptibility to parasites (De Paoli-Iseppi et al. 2019). In fisheries management, age structure is a key predictor of population dynamics and is therefore crucial for managing sustainable fisheries (Campana 2001). However, animal age is often very difficult to measure. Some animals exhibit physical features that are correlated with age, such as growth rings in fish otoliths (Panella 1971) and bivalve shells (Kilada et al. 2009), and tooth length in deer (Pérez-Barbería et al. 2014), but many, if not most, animals lack such characteristics. Accurate age estimates are often only attainable through expensive tracking or marking studies (De Paoli-Iseppi et al. 2019) but these approaches are not practical for many species, especially those that are long lived or inhabit environments that are difficult to access.

Recently, molecular markers of age have generated interest among those looking to develop affordable, accurate, non-lethal, minimally-invasive methods for estimating animal age (Jarman et al. 2015; De Paoli-Iseppi et al. 2017). These involve measuring a feature of DNA or RNA, or their associated molecules, that changes consistently over time among individuals. Telomere length, which declines throughout life in many species (reviewed by Dunshea et al. 2011; Barrett et al. 2013), was the first genetic marker of age to receive widespread attention among molecular ecologists. However, despite its initial promise, telomere length does not accurately predict chronological age in many animals, likely as a result of the complex interplay of genetic and environmental effects (Monaghan & Haussman 2006; Bize et al. 2009; Dunshea et al. 2011; Jarman et al. 2015). Other molecular methods of age determination have been suggested, based on changes to DNA damage or abundance throughout life (e.g. mitochondrial DNA heteroplasmy or copy number). However, to date, none have successfully been applied as molecular markers of age in a wild animal (Jarman et al. 2015; Jebb et al. 2018).

A more recently-explored and promising avenue for developing a molecular ageing assay is DNA methylation. DNA methylation is an epigenetic change in which methyl groups are added to DNA, almost exclusively where cytosine precedes guanine (CpGs) (Jones et al. 2015). CpG methylation plays an important role in controlling gene expression (Schübeler 2015). A gradual decline in genome-wide (global) methylation with increasing age has been observed in many taxa, including fish (*Oncorhynchus gorbuscha*: Berdyshev et al. 1967), mammals (*Homo sapiens*: Fuke et al. 2004; *Mus musculus*: Wilson et al. 1987), invertebrates (*Chlamys farreri*: Lian et al. 2015) and birds (*Gallus gallus*: Gryzińska et al. 2013). This age-related decline - combined with increasing variance among individuals - in global DNA methylation, has been referred to as 'epigenetic drift' and is thought to occur due to the imperfect and weakening maintenance of DNA methylation over time (Jones et al. 2015; Field et al. 2018). At individual CpGs, the amount of methylation undergoes a positive (hypermethylation) or negative (hypomethylation) linear relationship with chronological age in a range of animals (Table 3.1). Such relationships may remain linear irrespective of how long an animal lives; i.e. the 'tick rate' of CpG methylation differs according to lifespan (Field et al. 2018; Lowe et al. 2018). The underlying mechanisms affecting individual loci are unknown (De Paoli-Iseppi et al. 2017) but site-specific CpG-predictors of age (termed 'epigenetic clocks') are highly correlated with age and display low margins of error in every species studied to date (Table 3.1). Being able to estimate chronological age by measuring CpG methylation at just a handful of loci has made it possible to implement epigenetic ageing tools in a diverse array of areas including human forensics (Shabani et al. 2018) and marine science (Polanowski et al. 2014; Beal et al. 2019).

Until recently, a key barrier to developing an ageing tool in non-model organisms based on epigenetic clocks was the need to have high quality, full-length genomic data for either the species of interest, or a closely-related species. There are next-generation sequencing (NGS) approaches that make it possible to attain site-level methylation data across entire genomes without prior genomic information, but these methods are costly (Kurdyukov & Bullock 2016). Developing a more cost-effective, targeted ageing assay based on a limited number of CpGs relies on being able to identify differentially-methylated regions using existing genetic data (De Paoli-Iseppi et

al. 2019). Extensive gene-level information on age-related DNA methylation exists for humans (e.g. Horvath 2013), so many studies on the existence of epigenetic clocks in non-model animals have targeted orthologous sequences of the age-related genes identified in humans (e.g. Polanowski et al. 2014; Beal et al. 2019). However, this approach is only feasible for closely-related species, which likely explains the bias towards mammals in epigenetic clock studies (Table 3.1). More recently, an epigenetic clock has been described that may be applicable across the animal kingdom (Wang & Lemos 2019). This new ageing tool is based on methylated cytosines in ribosomal DNA (rDNA), which is the most evolutionarily-conserved region of the genome. Across ca. 13,000 bp of rDNA sequence in mice, 620 age-informative CpGs (66.8%) were discovered. Many CpGs were found to occur in distantly-related taxa; for example, more than 70% of human CpGs in the 18S and 5.8S genes of rDNA are found in species as divergent as zebrafish (*Danio rerio*). It remains to be known whether the rDNA ageing clock can be applied to wild animals.

Crustacean fisheries are a major industry with substantial benefits for human livelihoods and food security worldwide. Crustacean catch has the highest export value per live weight of any aquatic animal group, with a 22% global share by trade value (FAO 2020). However, concerns have been raised over the long-term sustainability of crustacean fisheries. Crustacean stocks are largely unassessed and unregulated owing to the difficulties of accurately estimating their chronological age and therefore making reliable predictions about population dynamics (Anderson et al. 2011; Boudreau & Worm 2012). Crustaceans are difficult to age because they moult throughout their lives and show indeterminate growth but with extensive individual variation in size-at-age (Vogt 2012; Kilada & Driscoll 2017). Several alternative methods for estimating crustacean age have been investigated (reviewed by Vogt 2012; Kilada & Driscoll 2017) but none have been adopted for routine use due to technical limitations (tag and recapture, lipofuscin content), limited or no association with chronological age (lipofuscin content, telomere length), or because they are affected by moulting (growth bands; Becker et al. 2018). Therefore, a reliable and accurate ageing method is urgently needed for crustacean fisheries management and would have considerable economic and conservation impacts. In a recent review of

future genetic tools for lobster management, DNA methylation-based markers were highlighted as a possible solution to age estimation (Silva et al. 2019).

The European lobster (*Homarus gammarus*) is an economically-important species harvested across its range in the shallow, coastal areas of the northeast Atlantic Ocean. Lobster landings are valued at more than £44 million per annum to the UK alone (MMO 2019). Stock assessments are currently based on tracking the change in length-frequencies across years at the population level to estimate future population resilience to fishing (CEFAS 2018). The European lobster has an estimated lifespan of 42–72 years (Sheehy et al. 1999), and length is not an accurate predictor of age (and therefore population dynamics) in such long-lived species because fast-growing, young individuals increasingly overlap in size with slow-growing, old individuals (Vogt 2012).

Here, the use of the ribosomal epigenetic clock of chronological age was tested in European lobsters. Specifically, percentage DNA methylation was quantified at individual CpGs across rDNA in known-age, hatchery-reared and unknown-age, wild lobsters using targeted bisulphite sequencing. Lasso regression was used to select the 'best' loci for predicting chronological age in known-age lobsters and these loci were used to create a multiple linear regression model for age prediction. The regression model was subsequently used to predict age in wild lobsters estimated to be ≥ 4 years old. This study was the first to investigate the applicability of site-specific, DNA methylation-based markers for age estimation in crustaceans, and the first to test the rDNA epigenetic clock of ageing in a wild animal.

Table 3.1 | Existing 'epigenetic clock' studies where the amount of methylation at a varying number of CpGs has successfully been used to estimate chronological age. For each study, data from the 'best' combination of CpGs are shown - i.e. those with the highest correlation coefficient and precision using multiple linear regression. Reported lifespan is average unless otherwise stated. Not all data were available for every study. Precision is the difference (mean, median or standard deviation) between known and estimated ages following model cross-validation. Lifespan, age range and precision are reported in years.

Species	Lifespan	Age range studied	Tissue	*n* genes	*n* CpGs	R^2/r	*p*	Precision	Reference
Birds									
Short-tailed shearwaters (*Ardenna tenuirostris*)	Max = 39	5–21	Blood	7	7	0.59	< 0.001	2.81	De Paoli-Iseppi et al. 2019
Fish									
European sea bass (*Dicentrarchus labrax*)	15	0.6–10.5	Muscle	4	28	0.82	< 0.001	2.15	Anastasiadi & Piferrer 2020
Mammals									
Bechstein's bat (*Myotis bechsteinii*)	Max = 21	0–14	Wing	3	7	0.58	< 0.001	1.52	Wright et al. 2018
Bottlenose dolphins (*Tursiops truncates*)	40–50	2–36	Skin	2	2	0.78	< 0.001	5.14	Beal et al. 2019
Chimpanzees (*Pan troglodytes*)	> 33 (captive)	1–39	Blood	2	5	0.74	0.03	5.41	Ito et al. 2018
		1–58.9	Blood	NA	80	0.96	< 0.001	2.4	Guevara et al. 2020

Table 3.1 (continued)

Species	Lifespan	Age range studied	Tissue	*n* genes	*n* CpGs	R^2/*r*	*p*	Precision	Reference
Domestic dogs (*Canis familiaris*)	12	0–14	Blood	NA	41	1.0	< 0.001	0.05	Thompson et al. 2017
Gray wolves (*Canis lupus*)	6–8	0–8	Blood	NA	67	0.96	< 0.001	0.04	Thompson et al. 2017
Humans (*Homo sapiens*)	79	18–70	Saliva	3	3	73	NA	5.2	Bocklandt et al. 2011
	79	0–100	Multi-tissue	NA	353	0.96	< 0.001	3.6	Horvath 2013
Humpback whales (*Megaptera novaeangliae*)	Max = 95	0–30	Skin	3	3	0.79	< 0.001	2.99	Polanowski et al. 2014
Mice (*Mus musculus*)	> 1.9	0.2–2.3	Multi-tissue	3	3	0.95	NA	0.09	Han et al. 2018

3.2 Methods

3.2.1 Study species and sampling

Tissue samples (claws) were obtained from European lobsters of different known ages (0.0–24.8 months) reared at the National Lobster Hatchery (NLH) in Cornwall, UK (Table 3.2). Unknown-age, wild lobsters caught off the coast of Cornwall were sampled by clipping the terminal end of an antenna (Table 3.2). All wild-caught lobsters were above minimum landing size (MLS) (88–137 mm carapace length (CL)) and therefore estimated to be ≥ 4 years old. This estimated minimum age is based on size-at-age data from previous mark-recapture studies (Bannister & Addison 1998; Uglem et al. 2005; Schmalenbach et al. 2011). The large size range of the wild lobsters suggests they may differ substantially in age (Sheehy et al. 1999).

Table 3.2 | Sample demography of the 171 European lobsters (*Homarus gammarus*) sequenced in this study. Ages are time post-hatching. Error in age estimates arises for individuals that were graded by moult stage rather than hatch date. Known-age lobsters were reared at the National Lobster Hatchery (NLH). NLH lobsters ≥ 7.3-months-old were deployed into sea-based containers (Daniels et al. 2015) off the coast of Cornwall (UK) ca. 1-month post-hatching. Wild-caught lobsters also caught off the coast of Cornwall. CL = carapace length. All lobsters sampled in 2018.

Source	Age at sampling (months)	Age uncertainty (months)	Tissue	Mean CL (mm) (± SD)	*n*
NLH	0.0	0	Claw	NA	27
NLH	1.8	0	Claw	NA	29
NLH	7.3	± 0.5 months	Claw	11.2 (2.0)	26
NLH	12.5	± 14 days	Claw	16.0 (2.3)	19
NLH	24.8	± 14 days	Claw	35.6 (3.1)	32
Wild caught	≥ 48	Unknown	Antenna	103.9 (15.1)	38

3.2.2 DNA extraction and rDNA reference Sanger sequencing

Genomic DNA was extracted from ca. 2 mm^3 of tissue excised from within the appendages (claws or antennae) using a salt-precipitation protocol (modified from

Aljanabi & Martinez 1997) and resuspended in H_2O. DNA concentration and purity were verified using a NanoDrop 8000 Spectrophotometer (Thermo Scientific).

Ribosomal DNA occurs in tandemly-repeated clusters separated by non-transcribed intergenic spacers (Dyomin et al. 2016). Each rDNA cluster comprises three genes essential for ribosome functions (18S, 5.8S and 28S rRNAs), internal transcribed spacers (ITS1 and ITS2) and external transcribed spacers (5'ETS and 3'ETS) (Fig. 3.1). Animal rDNA clusters range in length between 8–14 kb (Dyomin et al. 2016). Partial sequences for 18S and 28S in *H. gammarus* were available in GenBank (Accession numbers: DQ079749 and DQ079789, respectively). To recover additional reference rDNA sequences for *H. gammarus*, a combination of published primers and new primers were tested and designed (Table 3.3). New primers were manually designed using cross-species alignments of all publicly-available rDNA sequences for the European lobster, American lobster (*H. americanus*) and Norway lobster (*Nephrops norvegicus*) viewed in AliView (Larsson 2014). Primer3 software (Rozen & Skaletsky 2000) was used to ensure compatible annealing temperatures, appropriate GC content (40–60%), and to minimise secondary structures (hairpins) and primer dimer formation.

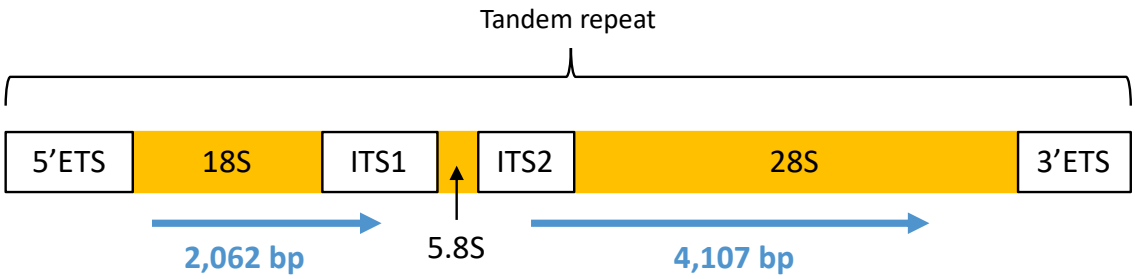


Figure 3.1 | Structure of ribosomal DNA clusters after (Dyomin et al. 2016). Blue arrows represent approximate locations of the regions sequenced in this study (total length = 6,169 bp) in European lobsters (*Homarus gammarus*).

Polymerase chain reactions (PCRs) were performed in 10 µl, consisting of 5 µl TopTaq Master Mix (Qiagen), 0.2 µl (10 µM) each primer, 4.1 µl ddH_2O and 0.5 µl DNA. Thermal cycle conditions were initial denaturation at 94°C for 3 minutes, followed by 25–30 cycles of denaturation (94°C, 30 seconds), annealing (30 seconds) and extension (72°C), with a final extension step at 72°C for 10 minutes. For each successful primer

pair, optimised cycle numbers, annealing temperatures and initial extension times are provided in Table 3.3.

Successful amplification was verified on a 1.5% agarose gel. Amplicons (5 µl) were cleaned with 0.1 µl of Exo1 (Thermo Scientific), 0.2 µl FastAP (Thermo Scientific) and 4.7 µl ddH_2O at 37°C for 15 minutes and 85°C for 15 minutes, then sequenced with Sanger sequencing (Eurofins). Sequence chromatograms were viewed in FinchTV (www.digitalworldbiology.com/FinchTV) and poor-quality regions removed. All available *H. gammarus* sequences were subsequently merged in AliView where possible to produce continuous sequence. The resulting continuous sequences (*n* = 2, see results section 3.3.1) were used as the regions of interest for targeted bisulphite sequencing.

3.2.3 Targeted bisulphite sequencing

Targeted bisulphite sequencing was conducted by Zymo Research (Irvine, California) including primer design, validation and bioinformatics. An overview of the protocol is provided in Fig. 3.2. Primers were designed to target CpGs across the specified regions of interest with Rosefinch - Zymo's proprietary primer design tool. Primers were designed such that PCR amplicons were between 100–300 bp and to avoid annealing to CpGs. Where this was not possible, primers were synthesised with a pyrimidine (C or T) at the CpG cytosine in the forward primer, or a purine (A or G) in the reverse primer, to minimise amplification bias to either the methylated or unmethylated allele. Primers were tested using real-time PCR (RT-PCR) with 1 ng of bisulphite-converted control DNA (from an individual lobster; L312) measured in duplicate. The presence of a single, specific PCR product was confirmed by analysing the RT-PCR melt curves. RT-PCRs were deemed successful if the following criteria were met: average crossing point (Cp) values were less than 40, duplicate Cp values did not differ by more than one, the plateau phase was reached before the run ended at cycle 45, melting curves were in the expected range for PCR products, and duplicates had calculated primer melting temperatures within 10% of the coefficient of variation.

Table 3.3 | Primers used in this study to PCR amplify rDNA in European lobsters (*Homarus gammarus*). All primers were successfully used for subsequent Sanger sequencing except ITS2_rev_SP-1-3, which failed to produce clean sequence despite several attempts at optimisation.

Primer name	Primer sequence (5' → 3')	Target	Annealing (°C)	Extension (minutes)	*n* cycles	Reference
ITS1_fwd_SP-1-5	CAC ACC GCC CGT CGC TAC TA	18S:ITS1	60	1	30	Chu et al. 2001
ITS2_rev_SP-1-3	ATT TAG CTG CGG TCT TCA TC					Chu et al. 2001
ITS1_fwd_SP-1-5	CAC ACC GCC CGT CGC TAC TA	ITS2:28S	60	2	25	Chu et al. 2001
28S_rev_PT3	TTC AGT CGC CCT TAC TAA GGG AAT CC					Tang et al. 2003
28S_gam_1F	GGT TAT CCC AGG CAG CAT TG	28S	55	1	25	This study
28S_gam_2R	AGT CAT AGT TAC TCC CGC CG					This study
28S_gam_3F	CGG CGG GAG TAA CTA TGA CT	28S	60	1	25	This study
28S_gam_4R	ATC GAT AGG CCT TGC TTT CG					This study

Following primer validation, lobster DNA was bisulphite converted using the EZ DNA Methylation-Lightning™ Kit (Zymo Research). Multiplex amplification of all DNA samples was performed using the Fluidigm Access Array™ System. The resulting amplicons were pooled and barcoded according to the Fluidigm instrument's guidelines. After barcoding, pooled amplicons were purified (ZR-96 DNA Clean & Concentrator™) and then prepared for sequencing using a MiSeq V2 300bp Reagent Kit (Illumina) and paired-end sequencing protocol according to the manufacturer's guidelines.

3.2.4 Quantifying percentage CpG methylation

NGS reads were identified using standard Illumina base-calling software and then analysed using a Zymo Research proprietary analysis pipeline. Low quality nucleotides and adapter sequences were removed. Paired-end sequence reads were aligned back to the reference sequences using Bismark, an aligner optimised for bisulphite sequencing data and DNA methylation calling (Krueger & Andrews 2011). Primer binding regions were removed from amplicons during rDNA methylation calling. Percentage methylation of each CpG was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T, multiplied by 100. Methylated sites with less than ten reads, or with missing data across individuals, were excluded.

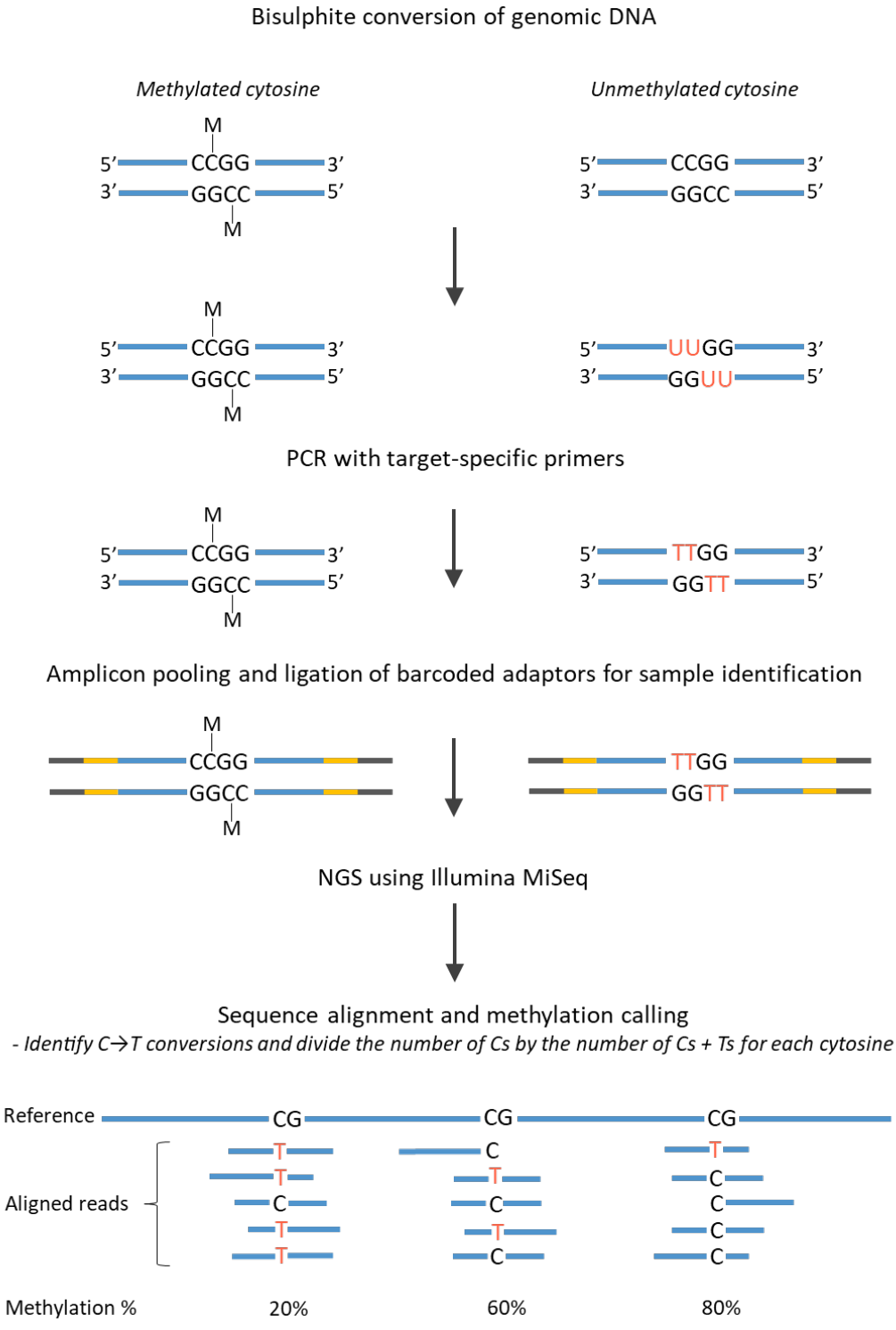


Figure 3.2 | Overview of targeted bisulphite sequencing protocol.

3.2.5 Developing an ageing tool for known-age lobsters

For lobsters of known age (n = 133), the relationship between percentage methylation and age was initially assessed for each quality-filtered CpG using simple linear regression ('lm{stats}': R Core Team 2017a). A Bonferroni–Holm correction (Holland & Copenhaver 1987) was applied to control for multiple comparisons using 'p.adjust{stats}'. Percentage methylation displayed a significant relationship with age at the majority of CpGs ($p < 0.05$; see results section 3.3.3). Therefore, penalised lasso regressions were fitted using 'cv.glmnet{glmnet}' (Friedman et al. 2010) with alpha = 1 (lasso). The lasso (least absolute shrinkage and selection operator) algorithm places a constraint on the absolute sum of the regression coefficients and reduces some of them to zero; non-zero coefficients are retained in the model, whereas coefficients reduced to zero are removed. This method is useful for selecting predictors, and reduces overfitting and multicollinearity (Tibshirani 1996). The size of the penalty was determined by the mean value of lambda 1 standard error (λ1se), which was calculated after 100 iterations of the default 10-fold cross validation in 'cv.glmnet{glmnet}'. This function randomly subdivides the data into ten groups, with nine parts used for training the model (the relationship between percentage rDNA methylation and age) and the remaining group for model testing. Methylated sites that passed the λ1se threshold, and had non-zero regression coefficients, were further filtered so that only loci with an $R^2 > 0.5$ according to simple linear regression were retained (n = 10). The ten remaining CpGs were incorporated into a multiple linear regression model to create an ageing assay for known-age lobsters with predicted age plotted against known age. Finally, stepwise backwards regression was performed to assess whether the number of CpGs in the ageing model could be further reduced for improved cost-efficiency without significantly altering the model fit according to the Akaike Information Criterion (AIC) using 'stepAIC{MASS}' (Venables & Ripley 2002).

3.2.6 Assessing the precision of the ageing tool

To assess the precision of the ageing tool for known-age lobsters, the multiple regression analysis was re-run using leave-one-out cross-validation (LOOCV) (Picard & Dennis Cook 1984). This method involves applying the multiple linear regression model

to all but one individual at a time until all individuals have been left out. Precision was subsequently quantified as the standard deviation (SD) of the mean difference between known and estimated ages. Model precision was compared among age groups using 'Anova{car}' (Fox & Weisberg 2019).

3.2.7 Effect of sex on age prediction in known-age lobsters

The sex was known for all 24.8-month-old lobsters (*n* = 32). Independent samples *t*-tests ('t.test{stats}') were used to compare: 1) percentage methylation for each of the ten CpGs used in the ageing assay for all known-age lobsters and 2) estimated age from a multiple linear regression of the same ten loci, between males and females. When comparing percentage methylation for individual CpGs (1), a Bonferroni–Holm correction (Holland & Copenhaver 1987) was applied using 'p.adjust{stats}'.

3.2.8 Predicting age in unknown-age, wild lobsters

Ages for the unknown-age, wild lobsters (*n* = 38) were estimated using two different methods for selecting CpGs. For method 1, the wild lobsters were aged by applying the ageing tool developed for known-age lobsters (ten CpGs). For method 2, CpGs were selected from the list of 237 loci that had a significant relationship with known age (see results section 3.3.3) according to four criteria below (Fig. 3.3), with justification thereafter:

1) the linear relationship between percentage rDNA methylation and age had an R^2 greater than 0.2 across known-age lobsters
2) for sites that were *hyper-methylated* with age in known-age lobsters: a) the *maximum* percentage rDNA methylation observed in known-age lobsters *did not exceed 50%* and, b) the *mean* percentage rDNA methylation was *higher* in wild lobsters compared to known-age lobsters
3) for sites that were *hypo-methylated* with age in known-age lobsters: a) the *minimum* percentage rDNA methylation observed in known-age lobsters *exceeded 50%* and b) the *mean* percentage rDNA methylation was *lower* in wild lobsters compared to known-age lobsters

4) the coefficient of variation (CV) in percentage rDNA methylation was higher within the wild lobster cohort compared to across known-age individuals

The rationale for these criteria was to select loci that had a linear relationship among known-age lobsters (1), were not likely to reach saturation (i.e. become completely unmethylated or fully methylated) at a young age (2a and 3a), for which the linear phase may extend into older age (2b and 3b), and where variability in rDNA methylation was higher among wild lobsters compared to across known-age lobsters to reflect the likely wider range of ages in the wild lobster cohort (4). The filtered CpGs were incorporated into a multiple linear regression for the known-age individuals, and this model subsequently used to predict age in the wild lobsters. The relationship in estimated age between the two methods for selecting loci was assessed using Pearson's correlation ('cor.test{stats]') following log transformation of the data. The age at which loci were predicted to reach saturation (x) was calculated using the linear equation for percentage methylation versus known age for each CpG using equation (1), where m = *y*-intercept and *c* = slope and *y* is set to 0 or 1 for hypo- and hyper-methylated loci, respectively.

$$x = \frac{y - m}{c} \qquad (1)$$

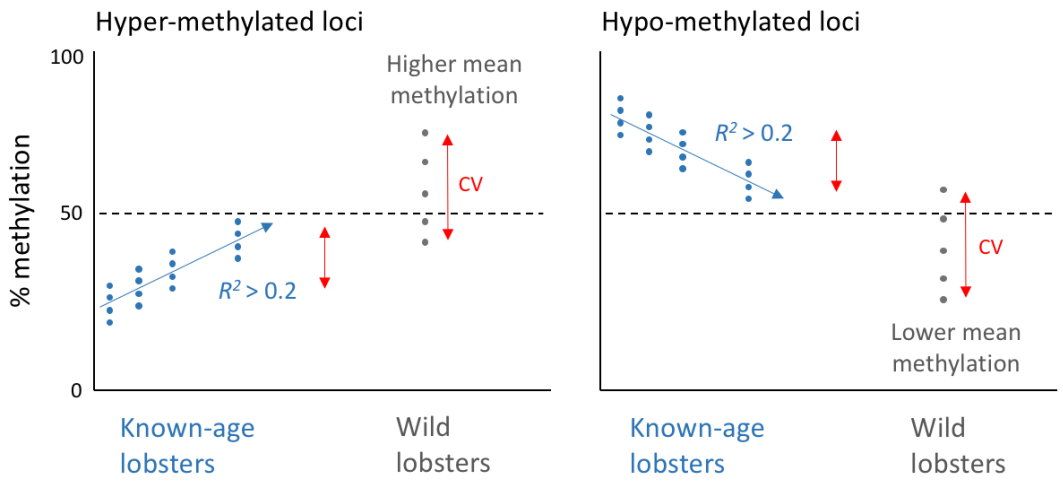


Figure 3.3 | Schematic of the process for selecting loci to estimate age in wild European lobsters (*Homarus gammarus*) (grey) based on features of the relationship between percentage rDNA methylation and age in known-age lobsters (blue), and differences in the mean and coefficient of variation (CV) in percentage rDNA methylation between known-age and wild lobsters.

3.2.9 Relationship between size and (estimated) age

Size (CL) data were available for all hatchery-reared lobsters ≥ 7.3 months old. For these individuals, the relationship between known age and CL was assessed with a Pearson's correlation using 'cor.test{stats}'. Pearson's correlation was also performed for the relationship between CL and estimated age in wild lobsters.

3.2.10 Statistics

Statistical analyses were performed in R version 3.6.3 (R Core Team 2020) using RStudio version 1.2.5003 (RStudio Team 2016). Plots were produced using 'ggplot{ggplot2}' (Wickham 2016) or 'Plot{graphics}' (R Core Team 2017b). For independent samples *t*-tests, normality and variance were assessed using 'shapiro.test{stats}' and 'leveneTest{car}', respectively. 'shapiro.test{stats}' was also used to check normality for Pearson's correlations and log transformation of variables performed when necessary. 'Plot{graphics}' was used to visualise residual and normality plots to test the assumptions of linear models.

3.3 Results

3.3.1 rDNA Sanger sequencing

Two continuous reference sequences spanning partial 18S through to the start of ITS1 (2,062 bp) and the end of ITS2 through to partial 28S (4,107 bp) were generated (Fig. 3.1). These two sequences were used as the regions of interest for bisulphite sequencing. Sequences for the gap between the two regions (end of ITS1 to beginning of ITS2) were not possible to obtain despite testing the primers published in Chu et al. 2001, ITS2 primers described in Harris & Crandall (2000), and primers designed in this study (Gam_ITS1_F: 5'-AGTCGTAACAAGGTTTCCGT-3' and Gam_ITS2_R: 5'-TCTTCACCACCGACATTACCA-3'), possibly because of intra-individual variation in ITS sequences (e.g. Harris & Crandall 2000; Bower et al. 2009).

3.3.2 Bisulphite sequencing quality control

Bisulphite conversion rates were greater than 99% for each DNA sample. Amplicons were successfully generated from bisulphite-converted DNA for 5,154 bp across the two regions (84% of combined length) (Supplementary Fig. S3.1). A total of 436 CpGs were sequenced and 355 were retained for downstream analyses following removal of sites that were not successfully sequenced across all individuals or had fewer than ten reads. Average coverage across the filtered CpGs was 7,605x.

3.3.3 Age prediction using CpG methylation in known-age lobsters

Simple linear regressions showed that percentage methylation had a significant relationship with age at 67% of filtered CpGs (n = 237/355) (Bonferroni–Holm corrected $p < 0.05$). Forty of the 355 CpGs passed the λ1se cut-off using lasso regression (Table 3.4; Supplementary Fig. S3.2). These forty markers were limited to ten CpGs when requiring the R^2 from simple linear regression to exceed 0.5. Among these ten loci, seven showed a negative relationship with age (hypomethylated) and three demonstrated a positive relationship (hypermethylated) (Fig. 3.4). The age-related CpGs were evenly-distributed along the rDNA sequence (Fig. 3.5).

Estimated age according to percentage methylation across the ten chosen loci had a highly significant relationship with chronological age ($p < 0.001$), explaining 91% of the variation in chronological age ($R^2 = 0.911$; Fig. 3.6a). The significant *y*-intercept indicates that the age of young lobsters may be slightly overestimated (*y*-intercept = 0.851; $p < 0.05$). Based on the slope (0.911), the age of older lobsters may be slightly underestimated. Stepwise backwards regression further reduced the number of CpGs for predicting age to five loci (ITS1_1874, 18S_366, ITS2_275, 28S_2770, 28S_3932) without a significant effect on the model AIC. A multiple linear regression model incorporating these five CpGs also explained 91% of the variation in chronological age ($R^2 = 0.909$; $p < 0.001$).

Table 3.4 | The forty CpGs with non-zero coefficients in the lasso regression model, which assessed the relationship between percentage methylation and age in European lobsters (*Homarus gammarus*) at 355 CpGs. R^2 and *p*-values from simple linear regression of percentage methylation with age for each CpG. A Bonferroni–Holm correction was applied to all *p*-values to account for multiple comparisons. Highlighted in bold and underlined are the ten CpGs used to estimated age in known-age lobsters. Positions according to the reference sequences compiled in this study.

Gene	Position	Lasso coefficient	R^2	Adjusted *p*
18S	242	2.142	0.068	0.148
18S	340	-38.869	0.411	<0.001
18S	**366**	**-34.325**	**0.666**	**<0.001**
18S	459	8.531	-0.008	1.000
18S	627	6.647	0.204	<0.001
18S	631	32.567	0.369	<0.001
18S	659	20.065	0.395	<0.001
18S	914	-3.925	0.036	1.000
18S	1026	32.553	0.286	<0.001
18S	**1304**	**-0.706**	**0.615**	**<0.001**
18S	1667	-4.005	0.172	<0.001
ITS1	1793	40.592	0.001	1.000
ITS1	**1874**	**-46.979**	**0.593**	**<0.001**
ITS2	**249**	**-37.095**	**0.528**	**<0.001**

Table 3.4 (continued)

Gene	Position	Lasso coefficient	R^2	Adjusted *p*
ITS2	**273**	**-0.037**	**0.504**	**<0.001**
ITS2	**275**	**-13.954**	**0.561**	**<0.001**
28S	587	-4.142	0.462	<0.001
28S	647	-0.128	0.331	<0.001
28S	992	27.823	0.120	0.004
28S	**1057**	**-59.228**	**0.621**	**<0.001**
28S	1116	7.712	0.246	<0.001
28S	1167	-2.650	0.155	<0.001
28S	1202	-4.092	0.055	0.354
28S	1303	-4.399	0.377	<0.001
28S	1307	-4.001	0.027	1.000
28S	1358	6.281	0.069	0.139
28S	1413	-0.410	0.260	<0.001
28S	1710	-4.115	0.120	0.004
28S	1843	-0.525	0.337	<0.001
28S	2165	9.587	0.004	1.000
28S	2590	-9.452	0.404	<0.001
28S	2616	-2.202	0.031	1.000
28S	**2761**	**4.871**	**0.699**	**<0.001**
28S	**2770**	**6.823**	**0.520**	**<0.001**
28S	2788	7.281	0.222	<0.001
28S	3048	0.255	-0.001	1.000
28S	3231	-1.722	0.191	<0.001
28S	3538	2.087	0.002	1.000
28S	3800	0.226	0.058	0.300
28S	**3932**	**14.824**	**0.635**	**<0.001**

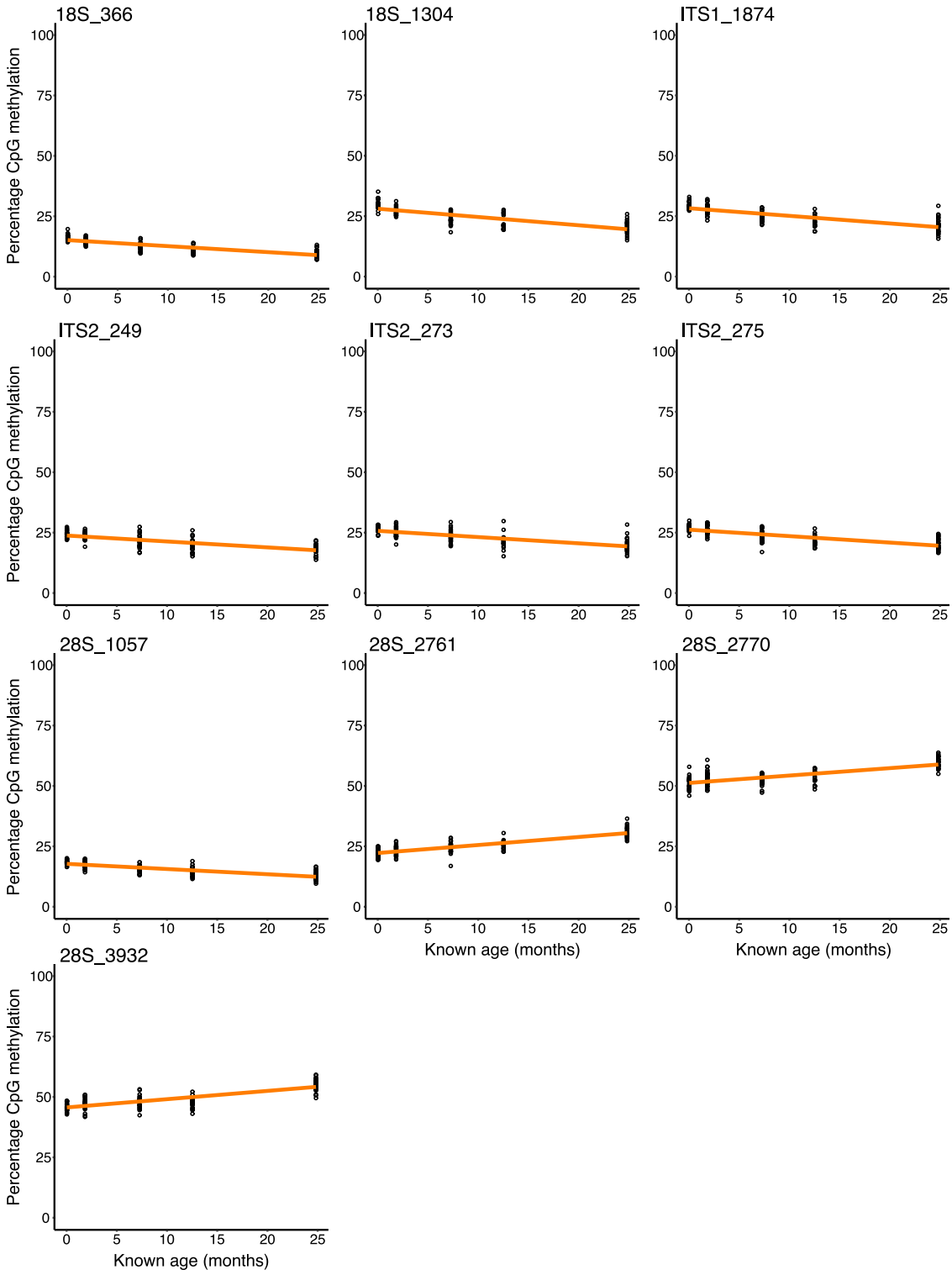


Figure 3.4 | Linear relationships for percentage methylation with age for the ten CpGs incorporated in the multiple regression model to estimate age in European lobsters (*Homarus gammarus*). Regression coefficients and *p*-values for the (orange) regression lines provided in Table 3.4.

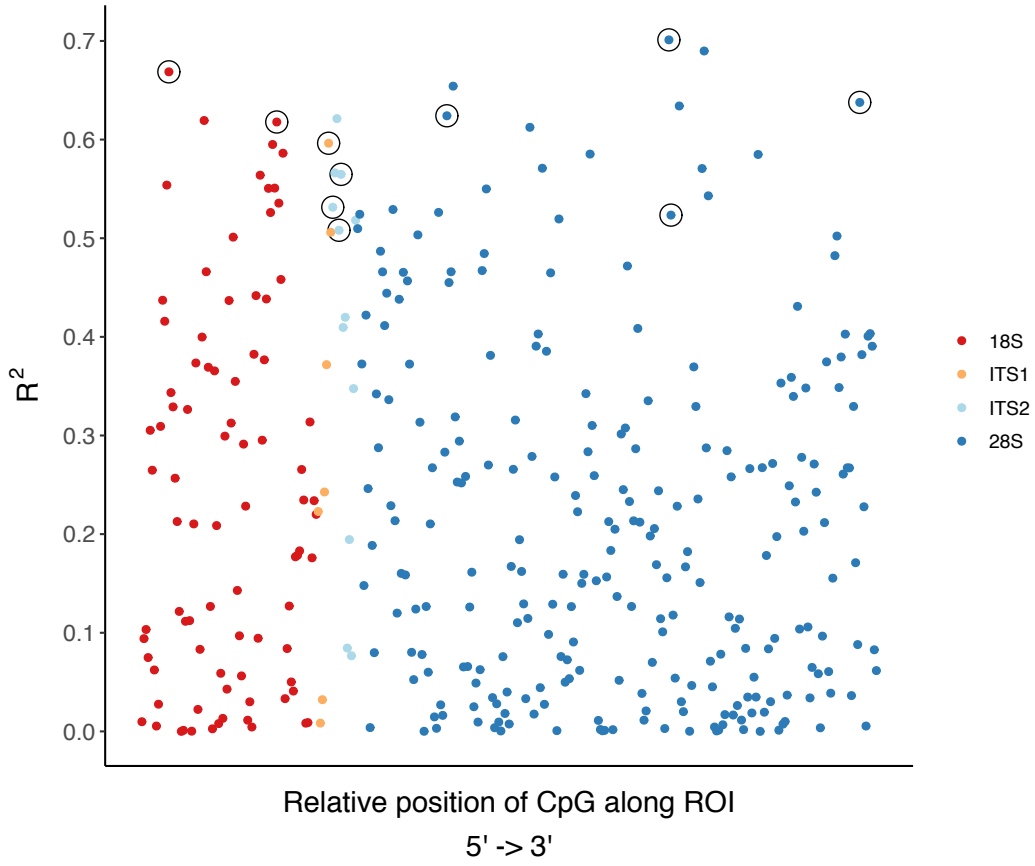


Figure 3.5 | Simple linear regression coefficients for the CpGs (*n* = 237) that showed significant age-related methylation changes in European lobsters (*Homarus gammarus*) (Bonferroni–Holm corrected $p < 0.05$). Sites ordered according to their relative position along the rDNA region of interest (ROI) and coloured by gene region. Encircled are the ten CpGs used in the multiple linear regression for predicting age in known-age lobsters.

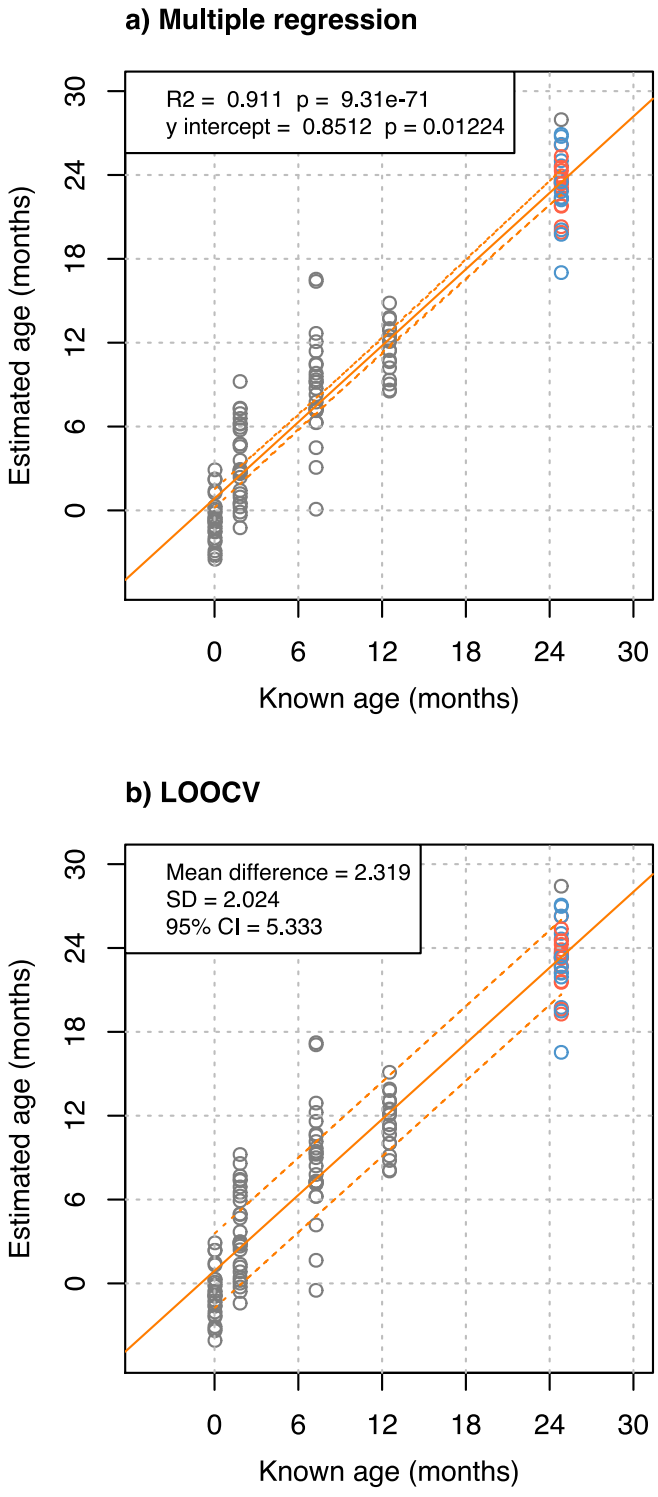


Figure 3.6 | a) Multiple linear regression for estimated age based on percentage methylation at ten CpG loci in European lobsters (*Homarus gammarus*) (*n* = 133) and b) the precision of the model as determined using a leave-one-out cross-validation analysis (LOOCV). Orange dotted lines are 95% confidence intervals for the regression line (solid orange lines). Unknown sex in grey, females in red and males in blue.

3.3.4 Ageing model precision

The LOOCV analysis estimated the precision of the ageing model to be 2.024 months - the SD of the mean difference between known and estimated ages (Fig. 3.6b). The precision of age estimates did not differ significantly among age groups (ANOVA: F = 2.341; p = 0.059; Fig. 3.7). Reducing the model to five CpGs marginally improved the precision (SD = 1.923 months).

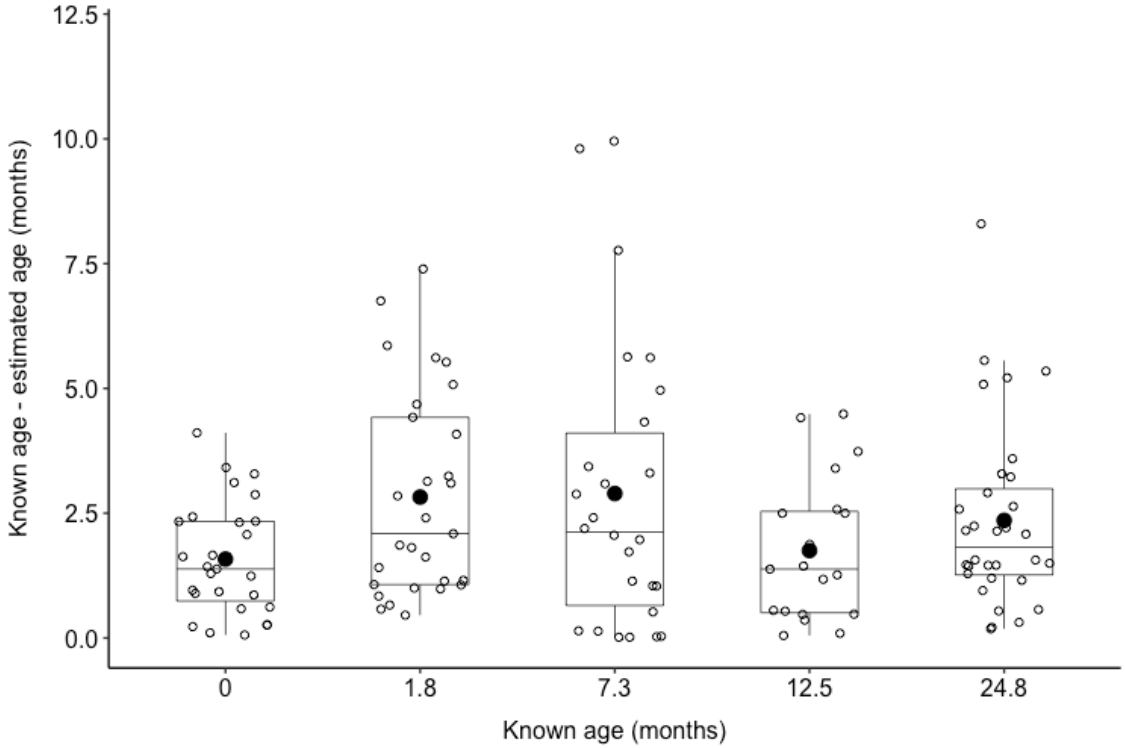


Figure 3.7 | Differences between known and estimated ages across five age groups of the European lobster (*Homarus gammarus*). Boxplots display a mean dot (solid black point), median line, inter-quartile range (IQR) boxes, 1.5*IQR whiskers, and data points.

3.3.5 Effect of sex on age prediction in known-age lobsters

In lobsters that were 24.8 months old (the only cohort in which individuals could be sexed), females had higher mean levels of methylation at each of the ten CpGs in the ageing model, but these differences were not significant (Bonferroni–Holm corrected p > 0.05; Fig. 8a; Table 3.5). Males and females did not differ in estimated age using a

multiple linear regression of the same ten loci (independent samples *t*-test: $t = 0.090$; $p = 0.929$; Fig. 8b).

Table 3.5 | Summary of independent samples *t*-tests comparing percentage methylation at ten CpGs in male ($n = 16$) and female ($n = 15$) European lobsters (*Homarus gammarus*) (df = 29). Adjusted *p* following Bonferroni–Holm correction.

Position	Mean difference (females – males)	*t*	*p*	Adjusted *p*
18S_366	0.008	1.579	0.125	0.382
18S_1304	0.017	2.092	0.045	0.363
ITS1_1874	0.021	2.355	0.025	0.229
ITS2_249	0.014	2.060	0.048	0.363
ITS2_273	0.016	1.723	0.095	0.382
ITS2_275	0.012	1.462	0.155	0.382
28S_1057	0.011	1.841	0.076	0.380
28S_2761	0.006	0.768	0.449	0.449
28S_2770	0.015	1.980	0.057	0.363
28S_3932	0.021	2.583	0.015	0.151

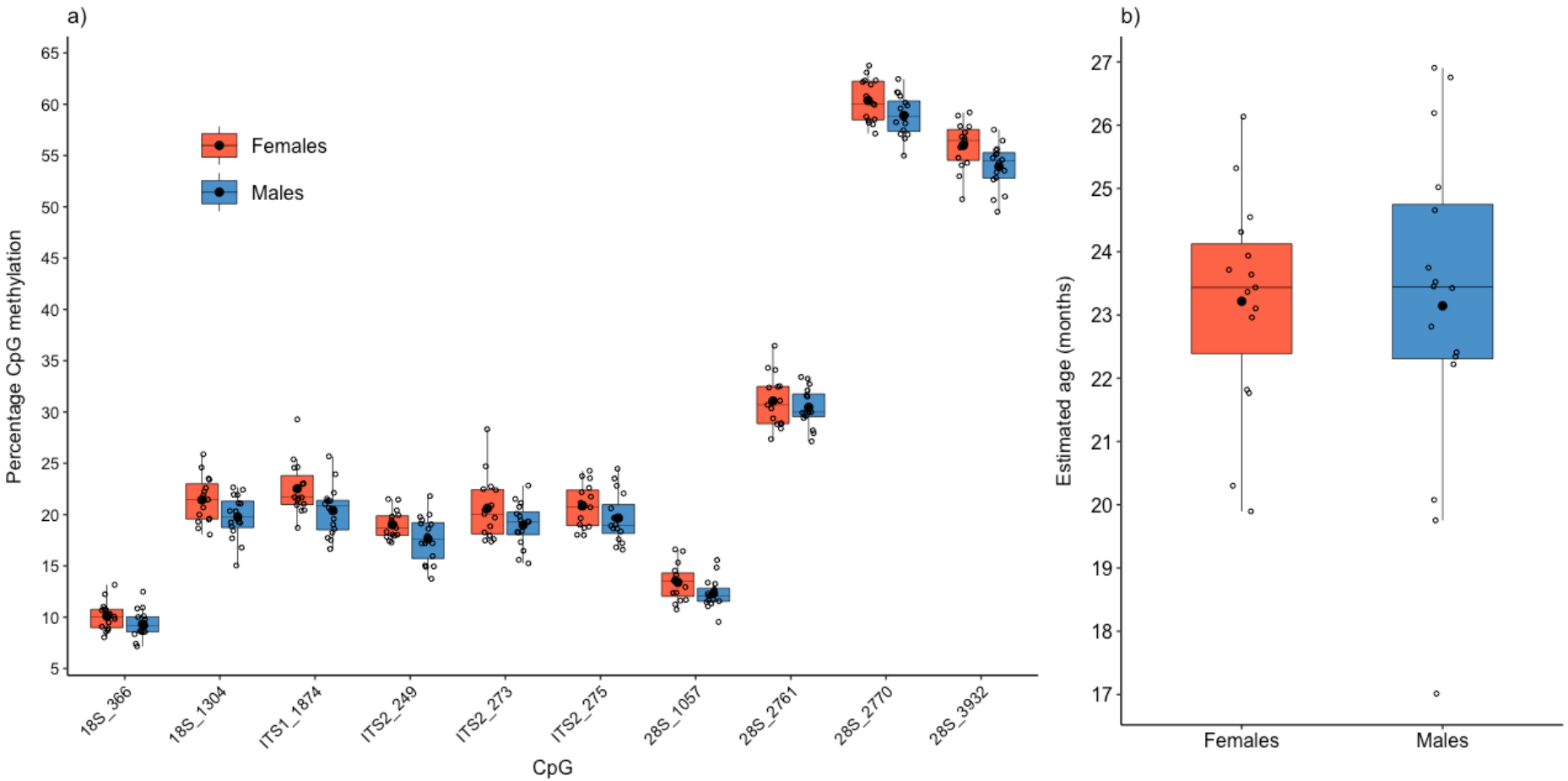


Figure 3.8 | Independent samples *t*-tests revealed no significant differences between male and female European lobsters (*Homarus gammarus*) at 24.8 months of age in a) percentage methylation across ten CpGs (Bonferroni–Holm corrected $p > 0.05$) and b) estimated age from a multiple linear regression model of the same ten loci ($p > 0.05$). Boxplots display a mean dot (solid black point), median line, inter-quartile range (IQR) boxes, 1.5*IQR whiskers, and data points.

3.3.6 Predicting age in unknown-age, wild lobsters

The ages of unknown-age, wild lobsters were predicted using two methods. When applying the ageing model developed for known-age lobsters (method 1), wild lobsters were estimated to be 25.0 months old on average (range = 15.0–41.5 months). The ten loci in this model were estimated to become saturated (fully methylated or completely unmethylated) at 134.2 months (ca. 11 years) on average (range = 60.0–284.6 months). In an attempt to improve age estimation in the wild lobsters, a different set of loci were chosen based on a set of criteria aimed at selecting CpGs with a higher probability of having a linear relationship into later life (method 2; Fig. 3.3). Seven loci were selected (18S_861, 18S_1026, 28S_1116, 28S_2154, 28S_2653, 28S_2761, 28S_3174), only one of which was included in the original model. An ageing assay created for known-age lobsters using percentage methylation at the seven CpGs had an R^2 of 0.813 ($p < 0.001$) and, according to a LOOCV, predicted age in known-age lobsters with a standard deviation of 2.610 months (Supplementary Fig. S3.3). These seven loci were predicted to become saturated at 430.8 months on average, equivalent to ca. 36 years (range = 225.8–800.0 months). Despite the fact these seven loci were predicted to reach saturation much later in life, method 2 resulted in only marginally higher age estimates for wild lobsters compared to method 1 (mean = 26.1 months; range = 13.2–54.3 months). Estimated ages were highly correlated between the two methods (Pearson's $r = 0.874$; $p < 0.001$).

3.3.7 Relationship between size and (estimated) age

Size (CL) was positively correlated with age in known-age lobsters 7.3–24.8 months old ($r = 0.971$; $p < 0.001$; Supplementary Fig. S3.4a). There was no relationship between CL and estimated age (using method 1) in wild lobsters ($r = 0.222$; $p = 0.180$; Supplementary Fig. S3.4b).

3.4 Discussion

This study sought to investigate whether patterns of CpG methylation in rDNA could be used to estimate age in difficult-to-age European lobsters. Percentage methylation had a significant relationship with chronological age at a large number of the inspected loci (n = 237; 67% of CpGs that passed quality control) in European lobsters ranging in age from 0.0–24.8 months. Ten loci were chosen by lasso regression to construct a multiple linear regression model. The aging model predicted lobster age with high accuracy (R^2 = 0.911; p < 0.001) and precision (SD = 2.024 months). Stepwise regression further revealed that the ageing assay could be reduced to five CpGs without a significant effect on the explanatory power of the model while maintaining high accuracy and precision (R^2 = 0.909; p < 0.001; SD = 1.923 months). The accuracy of this model is among the highest reported in any arnimal, with the average r/R^2 for epigenetic clocks developed in other animals, based on alternative regions of the genome with a comparably low number of CpGs (n = 2–7), being 0.74 (range = 0.58–0.95) (Table 3.1). These results suggest that the measurement of rDNA methylation holds considerable promise as a cost-effective marker of age in European lobsters and helps to confirm the hypothesis that rDNA may harbour an evolutionarily-conserved clock of animal age (Wang & Lemos 2019). Further work is required to test the assay on older, known-age European lobsters and across different tissues.

The age range of individuals used to calibrate the ageing assay represents a small proportion of the estimated lifespan for European lobsters (42–72 years; Sheehy et al. 1999). Here, the oldest, known-age lobsters were just over 24 months old. Obtaining tissue samples across a broad range of known ages is extremely difficult for this long-lived species, largely because it is harvested from the wild (not farmed) and cannot easily be individually tagged for recapture studies (due to moulting). This limitation (the lack of known age individuals) will apply to most studies involving economically-important crustaceans, which frequently have long lifespans (Vogt 2019). However, obtaining older, known-age European lobster DNA should be possible in the future with further sampling of lobster hatchlings reared by the National Lobster Hatchery since 2016, which are now at least 4 years old. Re-calibrating the ageing assay for

known 4-year-old lobsters (or older) would be of particular interest to fisheries management, as this is the minimum estimated age for exploited lobsters (see below).

In an attempt to test the ageing model's applicability to older lobsters, without having exact, known-age individuals, the assay created for known-age lobsters was used to predict the ages of unknown age, wild lobsters. This resulted in an average estimated age of 25.0 months (range = 15.0–41.5 months). Based on the size of the wild lobsters, this is likely an underestimate of their age. The most reliable information on size-at-age for European lobsters in the UK comes from a re-capture study in which thousands of hatchery-reared juvenile (stage VII) lobsters were microwire tagged, released from Bridlington on the east coast of the UK, and subsequently re-captured (Bannister & Addison 1998). Unfortunately, tissue samples were never taken as part of this study, however, tagged lobsters reached MLS (88 mm) 4–9 years after release, suggesting that the wild lobsters in this study are at least this old based on their size (88–137 mm CL). Similar patterns of growth were observed in recapture studies in other sites around the UK (numbers unreported: Bannister & Addison 1998), Heligoland Archipelago off the north-west coast of Germany (MLS = 85 mm CL at 4–7 years: Schmalenbach et al. 2011), and Norway (MLS = 88 mm CL at 4–8 years: Uglem et al. 2005), suggesting the minimum age at which lobsters reach legal size shows little population dependency.

Epigenetic clocks may fail to provide accurate estimates of age in older individuals if the CpGs of interest reach saturation before old age. In other words, if loci become fully methylated or completely unmethylated early in life. Based on the linear regression equation for the known-age lobsters, the original ten loci in the rDNA ageing model were predicted to reach saturation at 134.2 months (ca. 11 years) on average (range = 60.0–284.6 months), considerably above the minimum predicted age of the wild lobsters (4 years). The age predictions for wild lobsters were not increased by deliberately choosing rDNA loci (n = 7) with a higher likelihood of being applicable to older lobsters. This higher likelihood was based on criteria related to the strength of the relationship between percentage rDNA methylation and age in known-age lobsters, variability in rDNA methylation between known-age and wild lobsters, and

the estimated age at saturation (see methods section 3.2.8; Fig. 3.3). Methylation at these seven loci still provided accurate and precise estimates of age in known-age lobsters (R^2 = 0.813; SD = 2.610 months) and they were not predicted to reach saturation until 430.8 months on average, equivalent to ca. 36 years (range = 225.8–800.0 months). However, wild lobster ages were still underestimated (mean = 26.1 months; range = 13.2–54.3 months). This analysis suggests that saturation is not the cause of the apparent underestimation of age in wild lobsters. However, it cannot be ruled out that the regressions have been over-calibrated for the known-age lobsters and that the age of saturation is earlier than predicted. Further testing on older known-age individuals is needed to assess these possibilities.

Another possible cause of the apparent under-estimation of wild lobster age is that rDNA methylation changes are non-linear with age and reach a plateau phase (irrespective of saturation). Previous studies on epigenetic changes have primarily shown linear trends across wide age ranges in a number of different species, tissues and genomic regions (e.g. Bocklandt et al. 2011; Polanowski et al. 2014; Ito et a. 2018; Anastasiadi & Piferrer 2020). Data from human studies, however, suggest non-linear epigenetic changes do exist but that such trends are restricted to early life (Horvath 2013; Snir et al. 2019). Specifically, human DNA methylation has been shown to change at an accelerated rate in young individuals (birth to adolescents) and thereafter varies in a decelerated, linear fashion from early adulthood through to old age (Horvath 2013; Snir et al. 2019). During the adulthood phase of the Horvath epigenetic clock, predicted ages increased at the same rate as chronological age on average (Horvath 2013). Such trends of early acceleration followed by deceleration are best described by a logarithmic function with age (Snir et al. 2019). If a similar pattern occurs in European lobsters, a linear regression equation would result in an over-estimation of age in older individuals, which does not fit the observation in this study.

An alternative explanation is that the underestimation of age in wild lobsters may have occurred because epigenetic changes with age are tissue dependent. Claws and antennae were sampled from known-age and wild lobsters, respectively. If rDNA methylation levels change at different rates across tissues, ageing models need to be

re-trained and re-tested for different tissue types. Tissue-dependent patterns of age-related CpG methylation have been observed in humans (Christensen et al. 2009; Horvath 2013), mice (Maegawa et al. 2010; Spiers et al. 2016) and fish (Anastasiadi & Piferrer 2020). For example, in European sea bass, an epigenetic predictor of age created using muscle DNA was found to perform well in the testis but failed to accurately predict age from ovary tissue (Anastasiadi & Piferrer 2020). On the other hand, methylation status at some loci allows for multi-tissue predictors of age. 'Horvath's clock', for example, can accurately estimate human age from any of 51 different tissue and cell types based on the weighted average of 353 CpGs (Horvath 2013). Multi-tissue predictors tend to come at the cost of requiring more CpGs to capture the variation across tissues (Horvath 2013; Stubbs et al. 2017). Antennae were collected from wild lobsters as these can be sampled non-destructively, and without compromising the commercial value of the catch by removing a claw. Other non-destructive tissue samples include legs and pleopods, which lobsters can autotomize and regrow (Butler 2017). A useful and straightforward first step to exploring whether rDNA methylation changes in European lobsters are tissue-specific would be to acquire claw samples from wild lobsters around Cornwall, of similar size to those used in this study, and see how the range of estimated ages compares to the range reported here.

Male and female lobsters did not differ in percentage methylation at any of the ten individual loci included in the ageing model or in estimated age overall, but data were only available for 24.8-month-old individuals. Moreover, there were indications that if larger sample sizes had been used, there may have been differences: females had consistently higher methylation levels across all loci (although not statistically significant). Sex-related differences in methylation at age-related CpGs have been reported from bottlenose dolphins (Beal et al. 2019) and short-tailed shearwaters (De Paoli-Iseppi et al. 2019) but only for a portion of the CpGs investigated and these differences did not affect the multiple regression ageing models. None of the age-related CpGs investigated in humpback whales displayed sex-specific regressions (Polanowski et al. 2014). These results suggest that sex-dependent DNA methylation is context specific. Future work should investigate whether sex-specific differences in rDNA methylation occur across European lobsters of different known ages. If

differences exist, and sex accounts for unexplained variation in age in the ageing model, sex-specific regressions could improve the accuracy of epigenetic age estimation in lobsters.

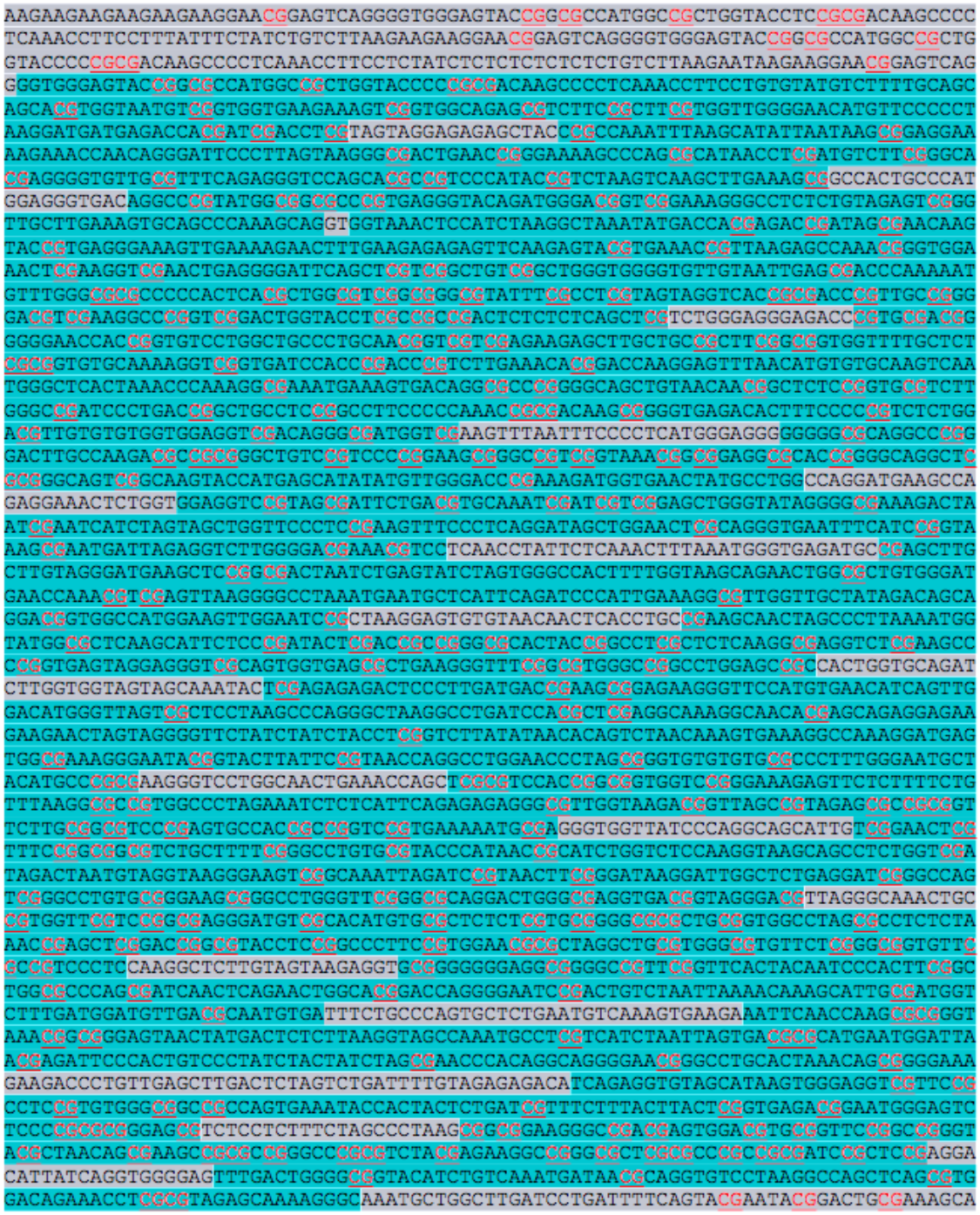
In conclusion, this study is the first to investigate the rDNA epigenetic clock of ageing in a wild animal and suggests that this method holds considerable promise as an ageing tool for European lobsters. Further development and validation of this work is needed before the method can be applied for use in fisheries science. As an area of priority, the rDNA ageing assay requires testing across a wider age range of European lobsters, and with a comparison of the effect of tissue types (including claw if comparisons are to be made directly with the present study). Such information will help to shed light on the main unanswered questions presented here: are age-related patterns of rDNA methylation a) linear into adulthood and/or b) tissue-specific? Information on whether the rDNA epigenetic clock is sex or population dependent will also be of value before such a tool can be widely adopted.

3.5 Supplementary information

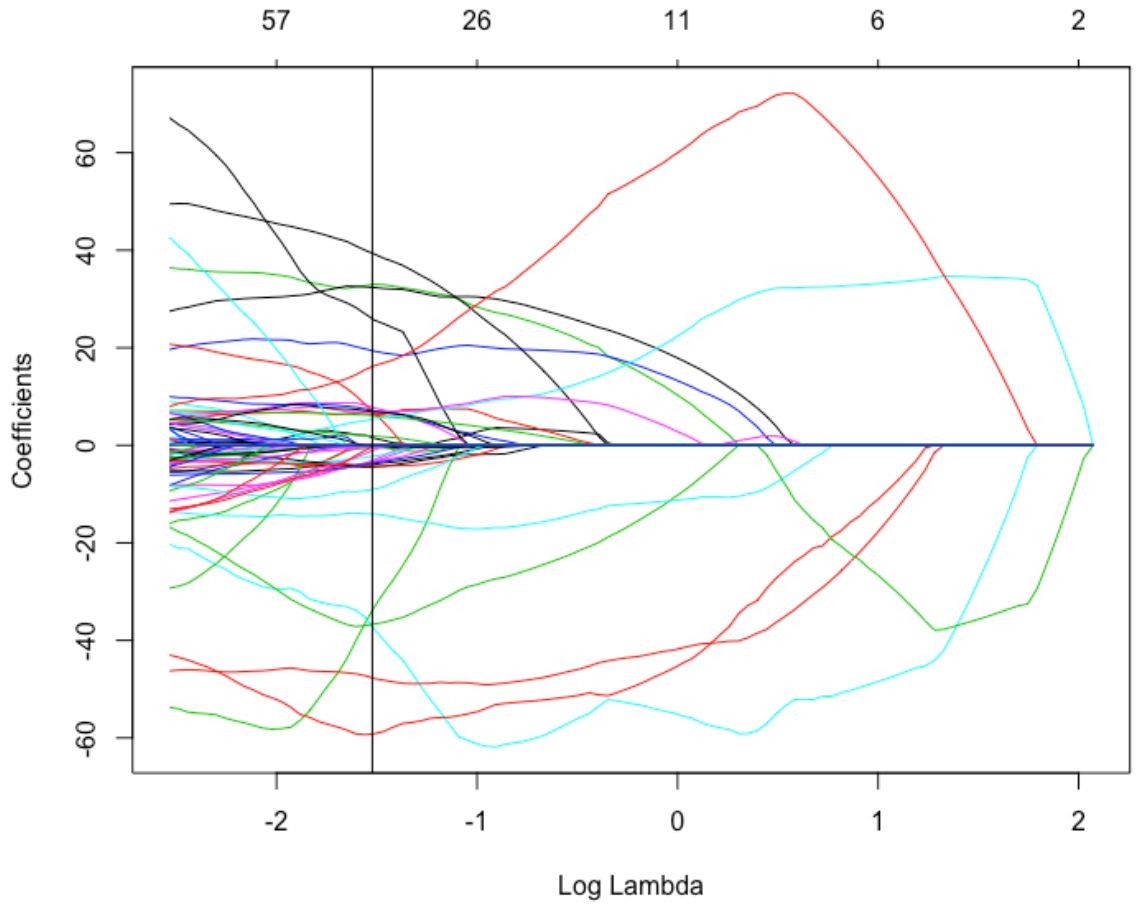
a) 18S–ITS1

CATGCATGTGTAAGTACAAGCCGATTTAAGGCGAAACCGCGAATGGCTCATTAAATCAGCTATGTTTCATTGGAT
CTGTAAACCCACTTACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCAATTGGTCTCTGACCGCAAGGGA
AGAGCGCTTTTATTAGTTCAAAACTGGTCGGGCCCCGGTCCGTAACCCACCTGTGGTGAATCTGAATAACTTCGG
GCTGAGCGCACGGTCTCCGCACCGGCGCCGCTTCTTTCAAGTGTCTGCCTTATCAGCTTTCGATTGTAGGTTATG
CGCCTACAATGGCTATAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCA
CATCTAAGGAAGGCAGCAGGCACGCAAATTACCCACTCCCGGCACGGGGAGGTAGTGACGAAAAATAACGATGTG
AGTCTCATCCGAGGCCTCGCAATCGGAATGAGTACACTTTAAATCCTTTAACGAGTATCCATTGGAGGGCAAGTC
TGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAAAGTTGTTGCGGTTAAAAAGCTCGTAGTTG
GATCTCAGTTCCGGACTGACGGTGCACCGCCCGGTGTTTACTGTCACGCTCCGAACAGCCGCCCCGCCGGCTCGC
ACGGGATGCTCTTTGTCGAGTGTCCCGAGTGGCCGGAGAGTTTACTTTGAAAAAATTAGAGTGCTCAGAGCAGGC
TATTTGAATGGCCCGAATGGTGATGCATGGAATAATGGAATAGGACCTCGGTTCTATTTTGTTGGTTTTCGGAAC
CAGAGGTAATGACTAATCGGAACAGGCGGGGGCATTCGTATTGCGACGCTAGAGGTGAAATTCTTGGACCGTCGC
AAGACGAACTACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGAGGTTCGAAGGCG
ATCAGATACCGCCCTAGTTCTAACCATAAACGATGCCAACTAGCGATCCGCCGGCGTTATTCCCATGACCCGGCG
GGCAGCTTCCGGGAAACCAAAGTCTTTGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGA
CGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAACCTCACCAGGCCCAGACAC
CGGAAGGATTGACAGATTGAGAGCTCTTTCTCGATTCCGTGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAG
CGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTGGCCTACTAAGTAGTCGACGGATCTCCAGAAAATGGT
GTCCAGTTCGCAACTTCTTCTTAGAGGGATAAGCGGCAATTCTAGCCGCACGAGATTGAGCAATAACAGGTCTGT
GATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGAAGGGATCAACGAGTTTTCCCCCTCCGAGAGGAGC
GGGTAACCCGTTCAAAGCCTTTCTTGATAGGGATTGGGGCTTGCAATTGTTTCCCATGAACGAGGAATTCCCAGT
AAGTGCAAGTCATCAGCTTGCGCTGACTACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTACTACCGATTGAA
TGATTTAGTGAGGCCTTCGGACTGGCGCTCTTGGATTGTCGGCCCCGTTGCCCGCGAGGGTTTCTGGGGTCGTC
GCCTCGAGCTGACGGAAAGATGTCCAAACTTGATCATTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGGT
GAACCTGCGGAAGGATCATTAACAGAGAGTAAGAGAACGGGGCGCTCTTAAAACTACTACTCTACTGATTTGGGC
GGTCTCGGGCTCTGCTCTGGCAACTCCAGCAGCGAATAGAGGACAATCCCGGCCCCCGCACCAAAAATAAATAAA
TACCTCGTTGTAGGATGAACCACACTCGGGCGAGAGCGCGAAGACCGGAGGAGGACGGACGGGGCTTTGGACGGC
TGCCCGTCTTTTCCTTACTACCGGCGAATAAGCCTCTTGACCCACAAAGAGTGAAATTCCTAAGCTTAACAACAA
CAAC

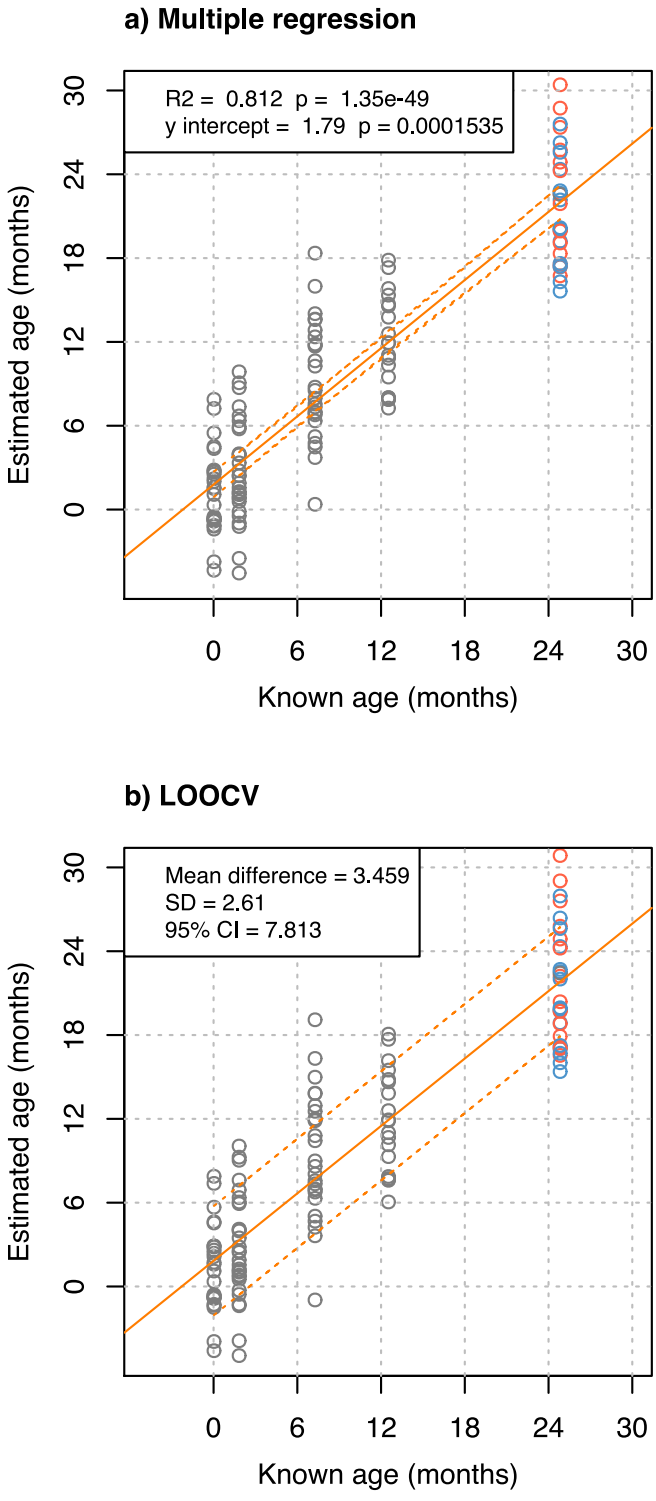
b) ITS2–28S



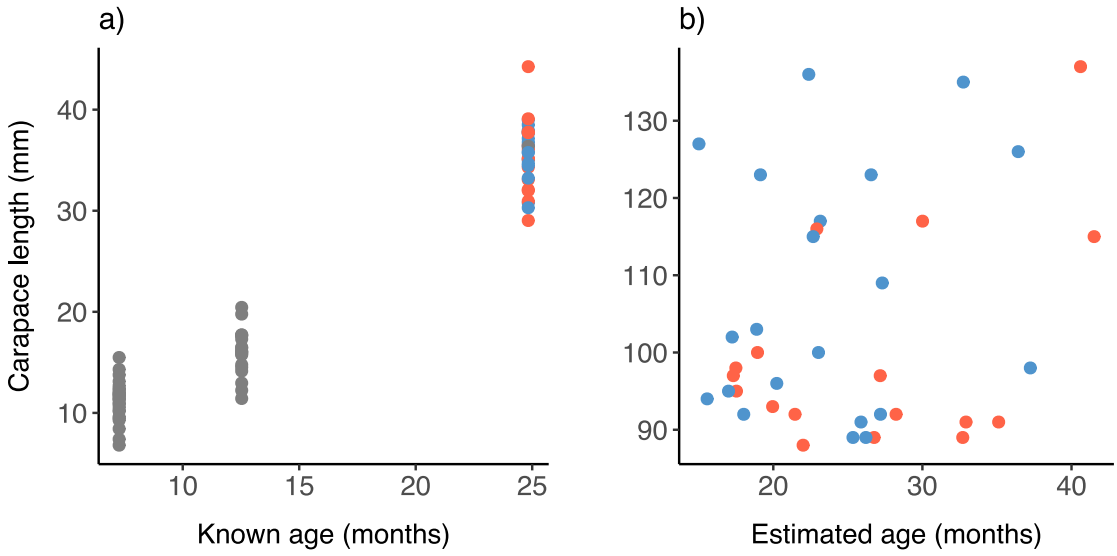
Supplementary Figure S3.1 | Sequences of the two ribosomal DNA regions of interest (a and b) for European lobsters (*Homarus Gammarus*). Regions successfully covered by amplicons for bisulphite sequencing in blue and regions not tested in grey. CpG dinucleotides in red.



Supplementary Figure S3.2 | Lasso regression coefficient pathways for the relationship between percentage methylation at individual CpGs (*n* = 355) and European lobster (*Homarus gammarus*) age at different levels of log lambda. The top axis represents the number of predictor variables with non-zero coefficients. CpGs that had non-zero coefficients above the lambda 1 standard error cut-off (vertical line) were selected for downstream analysis (*n* = 40).



Supplementary Figure S3.3 | a) Multiple linear regression for estimated age based on percentage methylation at seven CpG loci in European lobsters (*Homarus gammarus*) (*n* = 133) and b) the precision of the model as determined using a leave-one-out cross-validation analysis (LOOCV). Orange dotted lines are 95% confidence intervals for the regression line (solid orange lines). Unknown sex in grey, females in red and males in blue.



Supplementary Figure S3.4 | Relationship between (estimated) age and carapace length in a) known-age and b) wild European lobsters (*Homarus gammarus*). Ages of wild lobsters were estimated using a multiple linear regression of percentage methylation at ten CpG loci created using known-age lobsters. Unknown sex in grey, females in red and males in blue.

Chapter 4

The number of mitochondrial DNA mutations with age in laboratory-reared red cherry shrimp (*Neocaridina davidi*)

4.0 Abstract

Genetic mutations accumulate with age in the mitochondrial genome of humans and some vertebrate animal models, creating a mixture of wild-type and mutant mitochondrial DNA within individuals known as heteroplasmy. Surprisingly little is known about whether heteroplasmy accumulates with age in different species, despite its potential usefulness as a marker of age in difficult-to-age animals, such as crustaceans. In this study, whole mitochondrial genomes of red cherry shrimp (*Neocaridina davidi*) were amplified and sequenced at very high depth and point heteroplasmy compared across seven known-age cohorts sampled 7–210 days post-fertilisation (spanning ca. 54% of the species' typical lifespan in captivity). Eleven unique heteroplasmies were detected across ca. 6 kb of mitochondrial sequence in 127 shrimp. The number of heteroplasmies did not differ among age groups; shrimp from each age cohort had between 0–3 heteroplasmies. Individuals contained multiple haplotypes across the remaining 9.5 kb of the mitochondrial genome, but patterns of haplotype and nucleotide diversity points towards contamination from nuclear mitochondrial DNA copies (NUMTs). These results suggest that heteroplasmy is not a viable marker of chronological age in red cherry shrimp. However, additional data are needed on the distribution and frequency of heteroplasmy across the entire mitochondrial genome and lifespan, and in different tissues, to conclusively rule-out mitochondrial DNA mutation accumulation in this species.

4.1 Introduction

Animal age determines many important ecological parameters, such as reproduction potential, mortality risk and susceptibility to parasites (Jarman et al. 2015). The estimation of age in wild populations is therefore of great importance in conservation biology. In fisheries management, information on population age structure underpins population dynamic models that inform sustainable management practices (Campana 2001). Some animals can be aged using morphological features that change over time, such as growth rings in fish otoliths (Panella 1971) and bivalve shells (Kilada et al. 2009), and tooth length in deer (Pérez-Barbería et al. 2014). However, many, if not most, animals lack such easily-measurable characteristics.

Molecular methods of age determination may provide the solution for estimating age in otherwise difficult-to-age species (Jarman et al. 2015). These involve measuring a feature of the abundance or sequence of DNA or RNA, or associated molecules, that changes throughout an animal's lifespan. DNA can often be non-lethally sampled and increasing improvements to the cost-effectiveness and accessibility of molecular technology has made the use of molecular tools feasible for ecology projects around the world (Jarman et al. 2015). A number of molecular markers of age have been investigated in animals (reviewed by Jarman et al. 2015; De Paoli-Iseppi 2017). Examples include the loss of telomere length (e.g. Haussmann & Vleck 2002), varying levels of DNA methylation (e.g. De Paoli-Iseppi et al. 2019) and the accumulation of mitochondrial DNA (mtDNA) mutations over time (e.g. Thèves et al. 2006). Molecular markers based on changes in telomere length and DNA methylation have been the focus of a large number of studies (reviewed by Dunshea et al. 2011 and De Paoli-Iseppi 2017, respectively) but little attention has been given to mutation accumulation in mtDNA (Jebb et al. 2018).

Mitochondria are found in nearly all animal cells where they play an essential role in cellular processes including energy production, metabolism, cell cycling and death. They are the only component of animal cells that contain their own genetic material outside of the nucleus. In most animals, the mitochondrial genome (mitogenome) is a

small (15–17 kb), double-stranded, circular molecule, which contains 13 protein-coding genes essential for the electron transport chain, 22 translation RNAs (tRNAs) and two ribosomal RNAs (rRNAs) necessary for DNA translation, and a non-coding control region. Mitochondrial DNA is abundant relative to nuclear DNA; there are typically just two copies of nuclear DNA per eukaryotic cell but up to several thousand of copies of mtDNA (Stewart & Chinnery 2015). Within an individual cell, tissue or organism, different copies of mtDNA may be identical (homoplasmy) or exist as a mixture of mutant and wild-type mtDNA (heteroplasmy) that differs at individual nucleotides (point heteroplasmy) or in length (length heteroplasmy).

Heteroplasmy can arise through the inheritance of mutant mtDNA, or as a result of *de novo* somatic mutations that are acquired throughout life (Lawless et al. 2020). Inheritance of mtDNA mutations usually occurs via maternal transmission of heteroplasmic mtDNA in egg cells. Paternal leakage of sperm mtDNA is rare; many mechanisms exist to destroy sperm mtDNA, both before and after fertilisation, to reduce the spread of potentially harmful mutations (Polovina et al. 2020). During mother-to-offspring transmission, a very small proportion of mtDNA molecules repopulate the oocyte (Barrett et al. 2019). Due to random subsampling and/or purifying selection, this bottleneck leads to a shift in the level of heteroplasmy and explains how mutant molecules can increase in frequency or be purged from one generation to the next (Stewart & Larsson 2014). Somatic mutations may result from oxidative damage or polymerase errors during mtDNA replication (Lawless et al. 2020).

All mtDNA mutations originate in a single copy of mtDNA within a single mitochondrion of a single cell (Stewart & Chinnery 2015). However, over time, mtDNA deletions and/or point mutations have been shown to increase in number and frequency in humans (*Homo sapiens*) and other vertebrates, including monkeys (*Macaca mulatta*) and mice (*Mus musculus*) (Larsson 2010 and references therein; Li et al. 2015; Arbeithuber et al. 2020). One explanation for this age-related increase of mtDNA mutations is accumulating damage from reactive oxygen species (ROS). Mitochondria produce large quantities of ROS, as a by-product of cellular respiration, leaving the mtDNA exposed to oxidative damage (Trifunovic 2006). ROS-induced

damage has been postulated to lead to a 'vicious cycle' of reduced respiratory efficiency, causing enhanced ROS production and therefore further accumulation of mtDNA mutations (Harman 1972; Payne & Chinnery 2015). Alternatively, recent evidence suggests that clonal expansion of existing mtDNA mutations plays an important role in driving mtDNA mutation accumulation (Stewart & Chinnery 2015). In dividing cells, there is unequal, random partitioning of mtDNA molecules to the daughter cells, via 'vegetative segregation'. Daughter cells therefore receive different proportions of mutant and wild-type mtDNA genomes, which leads to shifts in the level of heteroplasmy over cell generations (Stewart & Chinnery 2015). In non-dividing cells, mtDNA is continuously destroyed and replicated in a 'relaxed' manner, with no mechanism in place to ensure that each molecule is replicated only once during a cycle (Stewart & Chinnery 2015). Either by random chance and/or selective mechanisms, a mutant mtDNA molecule may replicate more frequently than the wild-type molecule (Lawless et al. 2020).

Once a mtDNA mutation exceeds a frequency threshold within a cell (60–90%; Rossignol et al. 2003), it can cause mitochondrial respiration dysfunction. Therefore, the accumulation of mtDNA mutations has received a great deal of interest in the context of understanding the impacts on human health and disease (e.g. Taylor & Turnbull 2005; Stewart & Chinnery 2015). A less-well-studied aspect of heteroplasmy is whether age-related changes can be used to estimate chronological age. In humans, the age-related increase of specific mtDNA mutations has been put forward as an aid to estimating age in forensic investigations (Meissner & Ritz-Timme 2010). Surprisingly, just one study to date has explored the link between the number of mtDNA mutations and age in a wild animal; no significant association between heteroplasmy and age was observed across the mitogenome of long-lived, greater mouse-eared bats (*Myotis myotis*: Jebb et al. 2018).

mtDNA is particularly useful for chronological age determination; the high copy number means it is possible to extract intact mtDNA from aged specimens and in tissues that may otherwise have little or degraded nuclear DNA, such as hair or fingernails (Just et al. 2015; Barrett et al. 2019). In addition, the short length of mtDNA

allows heteroplasmy to be screened throughout the entire mitogenome without prior genetic information of the target species (Jarman et al. 2015; Barrett et al. 2019). The lack of progress in investigating heteroplasmy as a marker of animal age may be due to the technical limitations of traditional approaches for measuring heteroplasmy (Sanger sequencing) and the analytical challenges of current methods (next-generation sequencing (NGS)). Sanger sequencing can be labour-intensive, costly and exhibit low sensitivity. Therefore, early heteroplasmy studies were typically limited to a few individuals and/or a small portion of the mitogenome, and were only capable of detecting high levels of heteroplasmy (ca. 15% or higher) (Duan et al. 2018). The higher coverage afforded by NGS has allowed for the detection of low-level heteroplasmy, in some cases less than 1% (Duan et al. 2018). High resolution data are crucial for characterising heteroplasmy because most mtDNA variants exist at low frequencies (e.g. Jebb et al. 2018). However, bioinformatic analysis of NGS data is complicated by the need to consider sequencing error and possible contamination from nuclear copies of mtDNA sequences (NUMTs) (Santibanez-Koref et al. 2019; Dierckxsens et al. 2020), both of which can give the appearance of real mitochondrial variants (pseudo heteroplasmy).

In this study, age-related heteroplasmy was investigated in laboratory-reared cohorts of red cherry shrimp (*Neocaridina davidi*). Crustaceans are notoriously difficult to age (reviewed by Vogt 2012; Kilada & Driscoll 2017), which limits our ability to predict population dynamics and inform sustainable management of these valuable fisheries. An accurate ageing method is urgently needed for crustacean fisheries management and would have considerable economic and conservation impacts. However, many of the most economically-important crustaceans are long lived (Vogt 2019), which hinders efforts to investigate alternative methods of ageing across a wide range of known ages. Red cherry shrimp are short lived, with most individuals living for ca. 13 months in captivity (Schoolmann & Arndt 2018). They also breed year-round and are easy to rear in large numbers, making them an ideal model system for studying crustacean ageing.

Point heteroplasmy was characterised in red cherry shrimp using deep NGS. Some individuals contained linked single-nucleotide polymorphisms (SNPs), whereby nearby SNPs occur at similar frequency, across a portion (9.5 kb) of the mitogenome. Linked SNPs were phased into multiple, within-individual haplotypes. Such patterns are suggestive of genuine heteroplasmy or pseudo heteroplasmy from contamination by NUMTs or other sources. Haplotype and nucleotide diversity were assessed in this region to deduce the most likely cause. For the remaining region (6 kb), the number of mtDNA point mutations was compared among seven known-age shrimp cohorts spanning ca. 54% of the species' predicted lifespan. This study is one of the first to investigate age-dependent heteroplasmy outside model organisms (Jebb et al. 2018) and the first to explore the potential of heteroplasmy as a maker of chronological age in crustaceans.

4.2 Methods

4.2.1 Study species and sampling

Red cherry shrimp were sourced from Neil Hardy Aquatica Ltd and reared at the University of East Anglia, UK. Known-age cohorts were established by placing 5–7 berried (egg-carrying) females from broodstock aquaria into separate cohort aquaria. Females fertilised within eight days of each other were placed in the same cohort tank. Mothers were returned to the broodstock following hatching of the eggs. Cohort age was calculated as days post-fertilisation of the first female added to that aquarium and therefore represents maximum age. To maintain the known age of each cohort, shrimp that subsequently became pregnant within cohorts were temporarily placed in holding tanks until egg hatching and then returned to their respective aquarium. Tissue samples (walking legs) were taken from a random subset of shrimp in a cohort tank between 40- and 210-days post-fertilisation of the broodstock (Table 4.1). Shrimp eggs were collected from a single berried female. Tissue samples were stored in absolute ethanol at 4°C. All aquaria contained 50% de-chlorinated tap water and 50% RO water maintained at 22–24°C. Shrimp were fed *ad libitum* with commercial fish wafers (Tetra, UK).

Table 4.1 | Sample demography of the 151 red cherry shrimp (*Neocaridina davidi*) sequenced in this study (not including technical replicates). Sample sizes for the 128 shrimp that passed sequencing and mapping quality control in parentheses. Age estimates are time post-fertilisation of the broodstock. Error in age estimates represents the maximum number of days between fertilisation of the broodstock for that cohort. Life-cycle stage calculated from Nur & Christianus 2013.

Age in days	Life-cycle stage	*n*	Tissue
7	Egg	25 (25)	Whole eggs
40 (± 5)	Larva	17 (16)	Walking leg
60 (± 8)	Larva	25 (22)	Walking leg
75 (± 8)	Juvenile	25 (16)	Walking leg
90 (± 8)	Adult	26 (26)	Walking leg
120 (± 8)	Adult	24 (18)	Walking leg
210 (± 8)	Adult	9 (5)	Walking leg

4.2.2 DNA extraction and long-range mtDNA amplification

Genomic DNA was extracted from ca. 2 mm^3 of leg tissue or whole eggs using a salt-precipitation technique modified from Aljanabi and Martinez (1997) and resuspended in H_2O. DNA concentration and purity were verified using a NanoDrop 8000 Spectrophotometer (Thermo Scientific). DNA extracts were diluted to 2.5 ng/µl.

Two overlapping primer pairs were designed to amplify the entire *N. davidi* mitogenome (Table 4.2). Primers were designed according to the available reference mitogenome (GenBank accession number: NC_023823; Yu et al. 2014). Primer3 (Rozen & Skaletsky 2000) was used to ensure compatible annealing temperatures, appropriate GC content (40–60%), and to minimise secondary structures (hairpins) and primer dimer formation. Overlapping primer pairs are optimal for NGS because the number of sequencing reads tends to follow a normal distribution across the target sequence.

Table 4.2 | Primers designed to amplify the entire mitogenome of red cherry shrimp (*Neocaridina davidi*). Annealing positions according to the polished reference mitogenome assembled in this study (total length = 15,565 bp).

Primer name	Primer sequence (5' → 3')	Annealing position	Amplicon
Shrimp_mtDNA_3F	TTT GGC AGT TCG GTT AGC AG	3048	2
Shrimp_mtDNA_3R	ACA AGG TAG GTG TGC TGG AA	12631	
Shrimp_mtDNA_4F	GTC CGA CCA TTC ATA CTA GAC CT	11700	1
Shrimp_mtDNA_4R	AAT TGA AGC GAG CGG GAG TA	4348	

Polymerase chain reactions (PCRs) were performed in 25 µl, consisting 12.5 µl GoTaq® Long PCR Master Mix (Promega), 2.5 µl (10 µM) each primer, 6.5 µl ddH_2O and 1 µl DNA. Thermal cycle conditions were initial denaturation at 94°C for 2 minutes, followed by 25 cycles of denaturation (92°C, 30 seconds), annealing (60°C, 20 seconds) and extension (65°C, 8 minutes), with a final extension step at 72°C for 10 minutes. Successful amplification was verified on a 0.8% agarose gel. DNA from shrimp of different ages were randomly distributed across PCR plates to minimise plate effects.

4.2.3 Library preparation and sequencing using Illumina NovaSeq

Library preparation and sequencing were delivered by the Genomics Pipelines Group at the Earlham Institute in Norwich, UK. Libraries were constructed following a bespoke protocol based on the Illumina Nextera kit, in which DNA is simultaneously fragmented and tagged with sequencing adapters in a single, rapid, enzymatic reaction. PCR amplicons from the same individual were quantified and pooled to equimolar concentration. Enzymatic reactions were performed to cleave and tag DNA with a universal overhang; 1 ng of pooled, amplicon DNA was combined with 0.9 µl of Nextera reaction buffer and 0.1 µl Nextera enzyme in a 5 µl reaction volume and incubated for 10 minutes at 55˚C.

Individuals were barcoded by adding 2.5 µl of 2 µM custom barcoded P5 (5'-AATGATACGGCGACCACCGAGATCTACAC [barcode] TCGTCGGCAGCGTC-3') and P7 (5'-CAAGCAGAAGACGGCATACGAGAT [barcode] GTCTCGTGGGCTCGG-3') compatible primers, 5 µl 5x Kapa Robust 2G reaction buffer, 0.5 µl 10mM dNTPs, 0.1 µl Kapa Robust 2G enzyme and 10.4 µl water. Libraries were amplified by incubating the sample for 72˚C for 3 minutes, followed by 14 PCR cycles consisting of 95˚C for 1 minute, 65˚C for 20 seconds and 72˚C for 3 minutes.

Libraries were cleaned using magnetic beads to precipitate DNA molecules larger than 200 bp; 20 µl of amplified DNA was mixed with 20 µl of Kapa beads and incubated at room temperature for 5 minutes. Beads were pelleted on a magnetic particle concentrator, the supernatant removed, and two 70% ethanol washes performed. Beads were left to dry for 5 minutes at room temperature before being resuspended with 20 µl of 10 mM Tris-HCl and incubated at room temperature for 5 minutes to elute the DNA. The tube was returned to the concentrator, the beads allowed to pellet, and the aqueous phase containing the size-selected DNA molecules transferred to a new tube.

The size distribution of each purified library was determined by running 3 µl DNA mixed with 18 µl 10 mM Tris-HCl on a Perkin Elmer GX. To generate equimolar libraries, pooled amplicons were size selected on a Sage Science 1.5% BluePippin

cassette recovering molecules between 400–600 bp. Quality control of the size-selected pool was performed by running 1 µl aliquots on a Life Technologies Qubit high sensitivity assay and an Agilent DNA High Sense BioAnalyser chip and the concentration of viable libraries measured using quantitative PCR. For sequencing, library pools were loaded at 120 pM based on an average of the Qubit and quantitative PCR concentrations, both calculated using an average molecule size of 425 bp, and sequenced on the Illumina NovaSeq 6000 with a 2 x 150 bp read metric on a single SP flow cell.

4.2.4 Sequence quality control

Sequences were de-multiplexed and trimmed of adapter sequences and unique barcodes using bcl2fastq2 (Illumina). Quality filtering of sequencing data is a trade-off between removing poor-quality data and retaining enough sequences for downstream processing. Moderate filtering was applied prior to sequence mapping, followed by more stringent quality control for variant calling. Raw and filtered sequences were quality checked using FastQC (version 0.11.7; Andrews 2010) with a focus on base quality scores, GC content and sequence duplication levels. The NGS QC Toolkit (version 2.3.3; Patel & Jain 2012) was used to retain reads if at least 80% of bases had a quality score ≥ Q30 (a 1 in 1000 chance of an incorrect base call). This quality threshold was not extended to the entire read length because read quality tended to decline only towards the end of a read (Supplementary Fig. S4.1).

4.2.5 Mitogenome re-assembly and annotation

Inaccurate reference genomes can lead to false heteroplasmy calls. To verify the existing mitogenome sequence for *N. davidi* (GenBank accession number: NC_023823; Yu et al. 2014), the mitogenomes for every individual were re-assembled using NOVOPlasty - a *de novo* assembler for circular genomes (version 2.7.2; Dierckxsens et al. 2016). The published reference mitogenome was used as a seed. Contigs were aligned to the reference mitogenome using MUSCLE (version 3.8.31; Edgar 2004). Sequences from all individuals were assembled into a single contig of similar length (see results section 4.3.2). One individual (S208) was randomly chosen as the polished

mitogenome reference for downstream analyses. The polished mitogenome was annotated using the MITOS webserver (Bernt et al. 2013). Each protein-coding gene was manually inspected and their start and stop codons adjusted when necessary according to the *N. davidi* mitogenome reference and that belonging to the sister species *N. heteropoda koreana* (GenBank accession number: NC_043865; Park et al. 2019). The polished mitogenome was plotted using Circos (version 0.69-9; Krzywinski et al. 2009).

4.2.6 Sequence mapping

Filtered reads were mapped to the polished mitogenome reference using the BWA-MEM algorithm (version 0.7.17; Li & Durbin 2009) with default settings, and sorted by leftmost position using SAMtools (version 1.9; Li et al. 2009). Read groups were assigned to the mapped reads and PCR duplicates removed using Picard Tools (Broad Institute, Cambridge, US). Coverage was calculated using SAMtools. For variant simulations and heteroplasmy detection, reads were mapped to an extended version of the polished reference mitogenome by copying the first and last 500 bp to opposite ends of the sequence. This accounts for circularity of the mitogenome and avoids an artificial drop in the number of reads due to the arbitrary start and end point, which may result in false negatives.

4.2.7 Haplotype reconstruction of linked SNPs within protein-coding genes

Mapped reads were visualised in Tablet (version 1.19.05.28; Milne et al. 2016), which provides a graphical view of NGS alignments imported as a BAM file. Patterns of variant linkage were apparent in many individuals within a portion of the mitogenome (Fig. 4.1). To deduce the location and most likely cause, linked SNPs were phased into haplotypes using HaploJuice (version 1.5.5; Wong et al. 2018). HaploJuice identifies unique sub-sequences using a sliding window approach on short-read sequences aligned to a reference and can identify up to three haplotypes in each individual. Haplotypes are reconstructed by considering all possible combinations of the three most frequent sub-sequences from each window and choosing the best assembly

according to maximum likelihood. HaploJuice takes a BAM file as input and outputs a single FASTA file for each individual containing three haplotype sequences.

Haplotype sequences were concatenated into a single, multi-individual FASTA file and aligned using MUSCLE. This file was subsequently split into separate alignments for each protein-coding gene (*n* = 13). Focusing on protein-coding genes can help to identify evidence of non-functionality, which is indicative of pseudo heteroplasmy. Partitioning the mitogenome also helped to retain more data for downstream processing because HaploJuice was unable to phase linked SNPs across the entire sequence length. HaploJuice represents regions that cannot be reliably phased with 'N'. This can occur if there is low sequencing coverage for that region or a large distance between variants. For each protein-coding gene, individuals with one or more haplotypes containing Ns were removed from the dataset.

Haplotypes were manually filtered, by editing the alignment in AliView (version 1.26; Larsson 2014), so that only variants called by FreeBayes (version 1.3.1; Garrison & Marth 2012) remained. FreeBayes is optimised for population-wide variant calling. The FreeBayes output was filtered to retain variants with a minor allele frequency (MAF) > 1% (0.01), depth > 1,000x, mean mapping and allele qualities > 30, and where fewer than 90% of minor allele supporting reads occur on either the forward or reverse strand. In addition, all variants needed to occur in more than one individual. Indels were removed because they all involved mononucleotide repeats ≥ 7 and were considered unreliable (possible sequencing error). HaploJuice outputs three sequences per individual, even if there is just one haplotype and all sequences are identical. After variant filtering, duplicate sequences within an individual were removed. Haplotype networks were created in R using 'haploNet{pegas'}' (Paradis 2010) and annotated in Inkscape (version 0.92.2; www.inkscape.org).

4.2.8 Haplotype diversity, identity and functionality

To assess nucleotide diversity among haplotypes within genes, pairwise *p*-distances - the proportion of nucleotide sites at which two sequences differ - were calculated using MEGA (version 10.1.1; Stecher et al. 2020). To identify possible contamination

(non-shrimp haplotypes), haplotypes were identified using nucleotide BLAST queries against the GenBank nucleotide database. Amino acid alignments were visually inspected in AliView and the codon position of each variant was recorded. The number of synonymous versus non-synonymous mutations, transition versus transversion mutations, and stop codons, were counted according to the invertebrate mitochondrial code. A lack of third codon position and transition mutation bias, and the presence of non-synonymous mutations and stop codons are characteristics of NUMT contamination (Triant & DeWoody 2007; Baldo et al. 2011).

4.2.9 Simulations of point heteroplasmy detection using different variant callers

FreeBayes is optimised for population-based variant calling and may be less appropriate for detecting age-related heteroplasmy because variants can occur within a single individual (Jebb et al. 2018). To identify the 'best' variant caller, sequencing reads were generated *in silico* with a list of known variants inserted using the simulator NEAT-genReads (version 2.0; Stephens et al. 2016). The performance of three different callers to detect the known variants was compared.

Firstly, a run-specific sequencing error model was produced using genSeqErrorModel.py with the sequencing reads from individual S208 (used to produce the polished reference mitogenome in this study) as input and all other parameters set to default. Secondly, 150 bp paired-end reads were simulated using genreads.py with the error model from step one, the extended polished reference mitogenome, and a vcf file containing a list of 200 'known' variants to insert as inputs. Variants were randomly chosen for insertion in sets of 50 at a MAF of 1%, 4%, 7% and 10%. Variants were inserted and called exclusively within the region not affected by multiple within-individual haplotypes (amplicon 1; see results section 4.3.3). NEAT-genReads does not have a specific function for setting the MAF. Instead, this was achieved by altering the genotype (GT) field of the input vcf file. For example, for a MAF of 10%, the GT field can be set to 1/0/0/0/0/0/0/0/0/0 and genreads.py run with a ploidy of 10 (*-p* 10). Eight read sets were generated at different coverages: 100x, 250x, 500x, 1,000x, 1,500x, 2,500x, 5,000x and 10,000x. NEAT-genReads outputs a

BAM file, which was used as input for three variant callers; FreeBayes, LoFreq (version 2.1.3.1; Wilm et al. 2012) and VarScan (version 2.4.3; Koboldt et al. 2012).

For each caller and read set, a score between 0 and 1 was assigned. This score was calculated using equation (1), where power is the proportion of known variants detected, accuracy is the number of known variants called within 1% of the true frequency, and false positive rate is the proportion of false variants called, as per Jebb et al. (2018).

$$Score = \frac{power\ x\ accuracy}{1 - false\ positive\ rate} \qquad (1)$$

4.2.10 Point heteroplasmy detection and classification

Heteroplasmy was detected using the best performing variant caller according to the simulation results. Individuals with an average coverage < 1,000x were excluded. Variants were filtered to retain point mutations with a MAF > 1%, depth > 1,000x, average minor allele base quality > 30, and where fewer than 90% of minor allele supporting reads occurred on either the forward or reverse strand. The region affected by multiple within-individual haplotypes (amplicon 2; see results section 4.3.3) was excluded. Filtered heteroplasmies were manually classified by location, and whether they are synonymous or non-synonymous and transition or transversion mutations. Non-synonymous mutations were further classified based on their predicted impact (neutral or deleterious) according to the PROVEAN algorithm using the default threshold of -2.5 (Choi et al. 2012; Choi & Chan 2015). PROVEAN makes predictions based on alignments of homologous protein sequences; deleterious amino acid substitutions are less likely to be evolutionarily conserved across species.

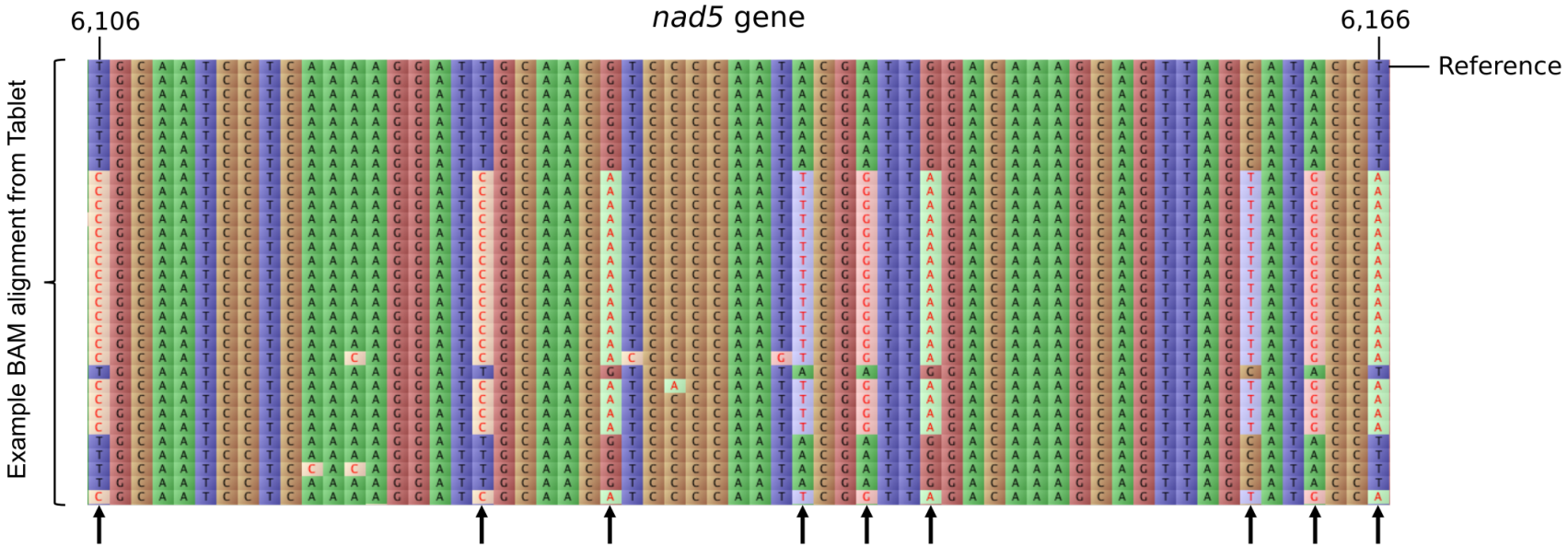


Figure 4.1 | Part of an alignment of Illumina sequencing reads (viewed in Tablet) from an individual red cherry shrimp (*Neocaridina davidi*) mapped to the mitogenome reference. This region was chosen to illustrate linked SNPs (indicated by arrows), which could be phased into additional haplotypes.

4.2.11 Point heteroplasmy and age

The number of variants was compared among shrimp age groups using generalised linear models (GLMs) fitted using 'glm{stats}' (R Core Team 2017a). A Poisson error distribution and log link function were used to account for non-normal residual errors, which are common to count data. Models were checked for overdispersion using 'dispersiontest{AER}' (Kleiber & Zeileis 2008). Influential points were assessed according to Cook's distance (Cook & Weisberg 1982) by examining diagnostic plots using 'plot{graphics}' (R Core Team 2017b). Analyses were performed with/without influential points and with/without the egg age class to control for tissue differences. For both GLMs, the main effect of age was assessed using Chi-Square tests in 'Anova{car}' (Fox & Weisberg 2019).

4.2.12 Heteroplasmy pipeline validation

To validate the heteroplasmy detection pipeline, PCR replicates were performed for a random subset of individuals (n = 8) and subsequently sequenced in separate reactions. For the same subset, sequencing replicates were achieved by performing 2x volume PCRs and sequencing each half in separate reactions. The number of variants called, and their frequencies, were compared across PCR and sequencing replicates.

4.2.13 Statistics

Statistical analyses were performed in R version (3.6.3; R Core Team 2020) using RStudio (RStudio Team 2016). Plots were produced using 'ggplot{ggplot2}' (Wickham 2016).

4.3 Results

4.3.1 Sequence quality control

Three individuals failed to sequence and were removed from the dataset. Sequencing was successful for 180 individuals (148 empirical individuals and 32 technical replicates) with a total of 106,819,320 read pairs after quality filtering. An average 57.5% of raw reads were retained post-filtering and contained an average 91.5% high-quality bases (≥ Q30). Mapping was successful for 172 individuals with an average coverage across the mitogenome of 2,746x (Fig. 4.2). Fourteen individuals with average coverage < 1,000x were removed. Average coverage for the remaining 158 individuals (128 empirical individuals and 30 technical replicates) was 2,938x (range = 1,008–5,620x).

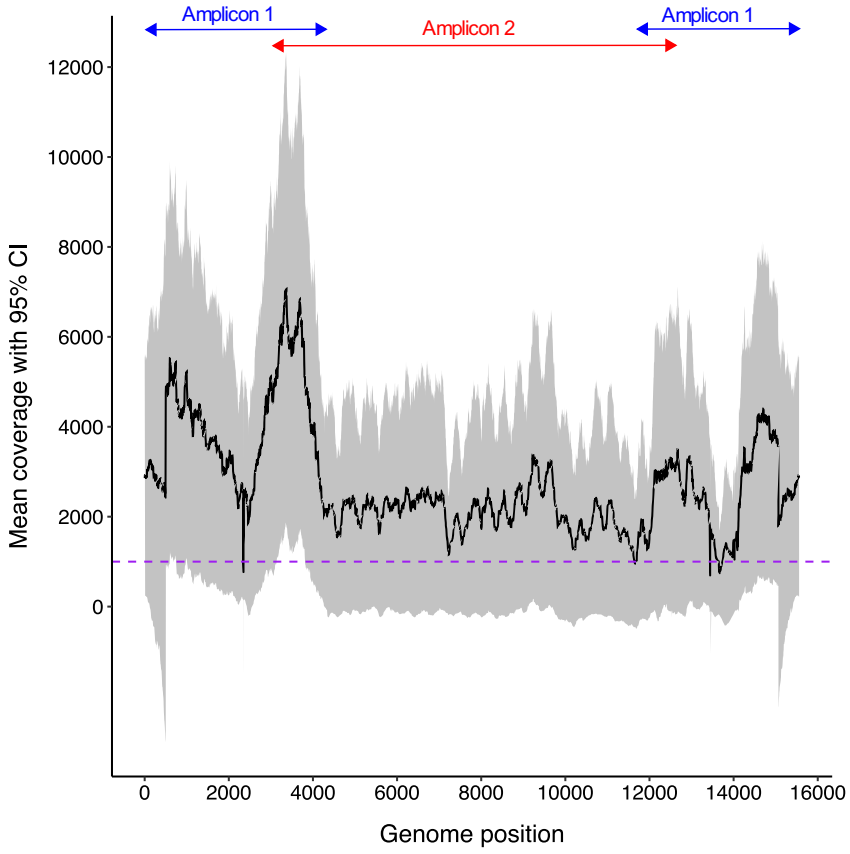


Figure 4.2 | Red cherry shrimp (*Neocaridina davidi*) mitogenome coverage from Illumina NovaSeq data of two overlapping amplicons sequenced from 172 individuals. Reads were mapped using default BWA-MEM settings and PCR duplicates removed. Black line = mean coverage across individuals and grey shaded area = 95% confidence intervals (CI). Horizontal purple line = 1,000x coverage. Approximate amplicon positions shown above.

4.3.2 Mitogenome assembly and annotation

Sequences from all individuals were assembled into a single contig, 15,563–15,567 bp in length. The gene complement and order were identical to the reference *N. davidi* mitogenome. However, *cox1* and *nad5* contained different initiation sequences compared to the *N. davidi* reference mitogenome, but identical to the sister species reference (*N. heteropoda koreana*). One individual was randomly chosen as the polished reference sequence for downstream analyses (S208). A Circos plot of the polished mitogenome reference is provided in Figure 4.3.

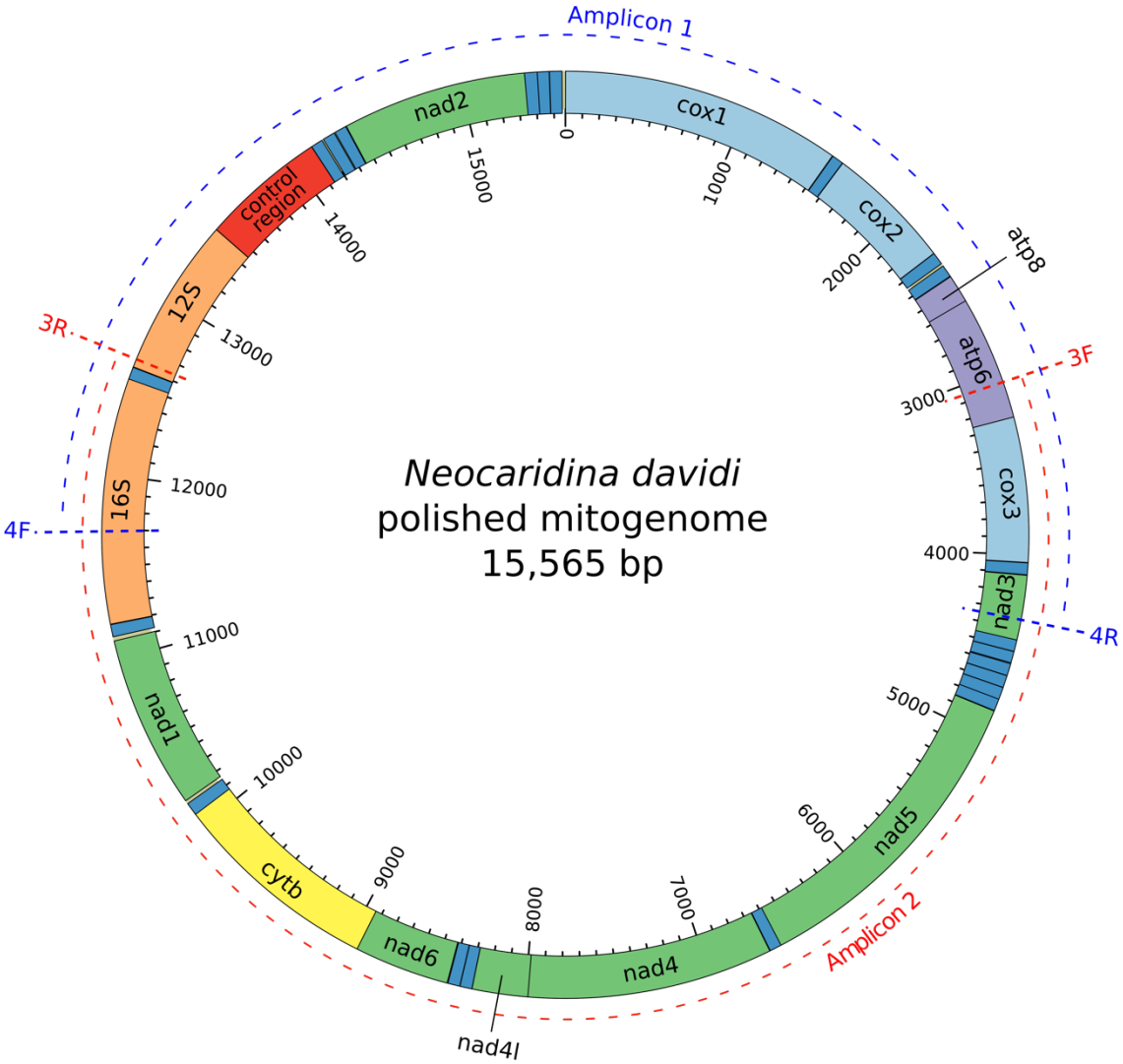


Figure 4.3 | Circos plot of the polished mitogenome sequence assembled in this study for red cherry shrimp (*Neocaridina davidi*). Dotted lines are approximate binding positions of the custom primers used for PCR amplification (Table 4.2). Darker blue regions are tRNAs.

4.3.3 Haplotype diversity

A single haplotype was reconstructed from protein-coding genes *atp8*, *cox3* and *nad3* (Table 4.3) in all individuals. Multiple haplotypes (2–22) were recovered from the remaining ten protein-coding genes (Table 4.3; Fig. 4.4). Haplotype and nucleotide diversity differed between the two amplicons (protein-coding genes amplified by both primer pairs were not considered for comparison). There were one or two haplotypes per protein-coding gene within amplicon 1, with average and maximum pairwise *p*-distances of 0.04% of 0.21%, respectively. Amplicon 2 contained 8–22 haplotypes per protein-coding gene, with average and maximum pairwise *p*-distances of 2.05% and 8.53%, respectively.

For all protein-coding genes with multiple haplotypes (n = 10), most shrimp contained one haplotype (56.1–98.6%; Table 4.4). The most common haplotypes were either the reference sequence or another closely-related haplotype (≤ 2 nucleotides divergent from the reference) (Fig. 4.4), regardless of whether a shrimp contained one or multiple haplotypes (Table 4.4).

Table 4.3 | Haplotype variation within the 13 protein-coding genes of the mitochondrial genome of red cherry shrimp (*Neocaridina davidi*). Haplotypes reconstructed using HaploJuice. Gene *atp6* split according to amplicon.

Gene	Length (bp)	Amplicon	*n* haplotypes	*n* individuals	*n* variable sites	Mean pairwise *p*-distance (%)	Max pairwise *p*-distance (%)
cox1	1521	1	2	132	1	0.03	0.07
cox2	688	1	2	141	1	0.06	0.15
atp8	159	1	1	132	0	-	-
atp6_1	468	1	2	138	1	0.04	0.21
atp6_2	207	1 & 2	2	138	1	0.22	0.48
cox3	786	1 & 2	1	139	0	-	-
nad3	354	1 & 2	1	130	0	-	-
nad5	1728	2	15	82	50	0.91	2.78
nad4	1339	2	22	89	109	2.85	8.07
nad4l	303	2	9	132	14	1.40	4.62
nad6	516	2	8	90	45	3.19	8.53
cytb	1137	2	17	123	74	1.87	6.42
nad1	942	2	14	103	60	2.09	6.05
nad2	1005	1	2	105	1	0.01	0.10

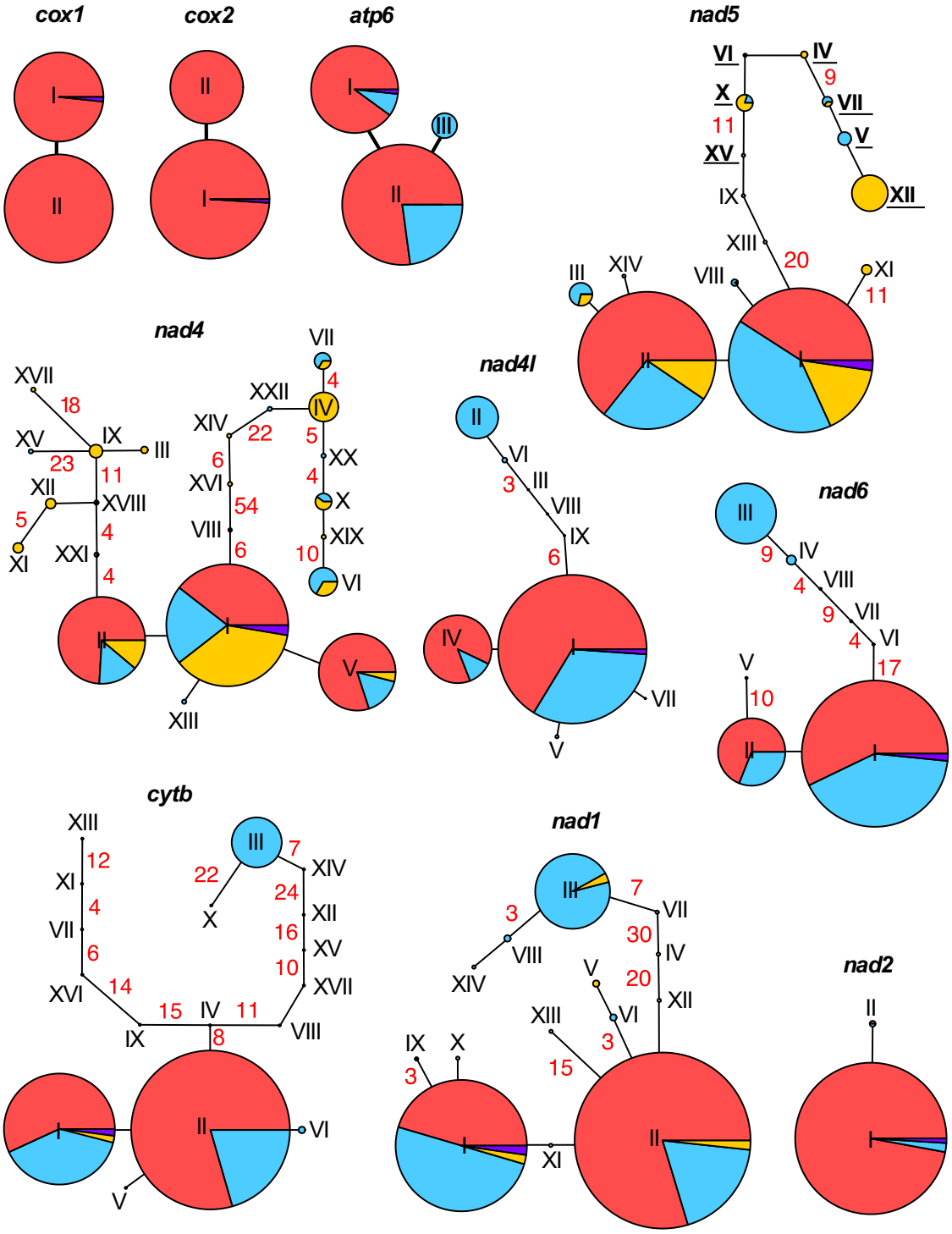


Figure 4.4 | Haplotype networks for mitochondrial protein-coding genes in 82–141 individual red cherry shrimp (*Neocaridina davidi*). Circles shaded by number of haplotypes within an individual: 1 = red, 2 = blue, ≥ 3 = orange. Purple = polished reference mitogenome created in this study. Haplotype IDs in Roman numerals. Numbers in red are the number of nucleotide differences between haplotypes if > 2. Size of circles represents relative haplotype frequency within and not between genes. Haplotype IDs containing a stop codon in bold and underlined.

Table 4.4 | Number and proportion (%) of red cherry shrimp (*Neocaridina davidi*) with 1, 2 or ≥ 3 haplotypes for each mitochondrial protein-coding gene. Gene *atp6* split according to amplicon.

Gene	Amplicon	*n* (%) 1 haplotype	*n* (%) 2 haplotypes	*n* (%) ≥ 3 haplotypes
cox1	1	132 (100)	0	0
cox2	1	141 (100)	0	0
atp8	1	132 (100)	0	0
atp6_1	1	121 (87.7)	17 (12.3)	0
atp6_2	1 & 2	136 (98.6)	2 (1.4)	0
cox3	1 & 2	139 (100)	0	0
nad3	1 & 2	130 (100)	0	0
nad5	2	46 (56.1)	27 (32.9)	9 (11.0)
nad4	2	56 (62.9)	15 (16.9)	18 (20.2)
nad4l	2	98 (74.2)	34 (25.8)	0
nad6	2	56 (62.2)	34 (37.8)	0
cytb	2	88 (71.5)	34 (27.6)	1 (0.8)
nad1	2	67 (65.0)	34 (33.0)	2 (1.9)
nad2	1	103 (98.1)	2 (1.9)	0

Table 4.5 | Number and proportion (in parentheses) of red cherry shrimp (*Neocaridina davidi*) with 1, 2 or ≥ 3 haplotypes that contained either the reference haplotype, or another closely-related haplotype (≤ 2 nucleotides divergent from the reference), for each mitochondrial protein-coding gene. Gene *atp6* split according to amplicon.

Gene	1 haplotype	2 haplotypes	≥ 3 haplotypes
cox1	132 (100.0)	-	-
cox2	141 (100.0)	-	-
atp8	132 (100.0)	-	-
atp6_1	121 (100.0)	17 (100.0)	-
atp6_2	136 (100.0)	2 (100.0)	-
cox3	139 (100.0)	-	-
nad3	130 (100.0)	-	-
nad5	46 (100.0)	27 (100.0)	9 (100.0)
nad4	55 (98.2)	15 (100.0)	18 (100.0)
nad4l	97 (100.0)	34 (100.0)	-
nad6	57 (100.0)	32 (100.0)	-
cytb	88 (100.0)	34 (100.0)	1 (100.0)
nad1	67 (100.0)	32 (94.1)	2 (100.0)
nad2	103 (100.0)	2 (100.0)	-

4.3.4 Haplotype identity and functionality

All haplotypes were most closely-related to mtDNA from either *N. davidi* or the sister species *N. heteropoda koreana* (BLAST identity ranging from 93.6–99.7%; Supplementary Table S4.1). All variable sites within the protein-coding genes of amplicon 1 were synonymous, transition mutations occurring on the third codon position, except for one non-synonymous, transition mutation in *atp6* affecting the second codon position (Table 4.6). Variables sites within the protein-coding genes of amplicon 2 were less biased towards the third codon position (mean variable sites at first, second and third codon position = 15.1%, 5.0% and 79.9%, respectively) and non-synonymous and transversion mutations occurred in 16.4% and 18.6% of variable sites on average, respectively (Table 4.6). A stop codon was identified in *nad5* (amplicon 2) in 24 individuals (Fig. 4.5).

Table 4.6 | The percentage of variables sites at each codon position and the percentage of non-synonymous (NS) or transversion (Tv) mutations across all haplotypes in the mitochondrial protein-coding genes of 82–141 individual red cherry shrimp (*Neocaridina davidi*). No variable sites were recorded in genes *atp8*, *cox3* or *nad3*. Gene *atp6* split according to amplicon.

Gene	Amplicon	1st codon (%)	2nd codon (%)	3rd codon (%)	NS %	Tv %
cox1	1	0	0	1 (100)	0.0	0.0
cox2	1	0	0	1 (100)	0.0	0.0
atp8	1	-	-	-	-	-
atp6_1	1	0.0	100.0	0.0	100.0	0.0
atp6_2	1 & 2	100.0	0.0	0.0	100.0	0.0
cox3	1 & 2	-	-	-	-	-
nad3	1 & 2	-	-	-	-	-
nad5	2	12.0	2.0	86.0	14.0	20.0
nad4	2	14.7	5.5	79.8	13.8	13.8
nad4l	2	21.4	7.1	71.4	16.7	28.6
nad6	2	22.2	8.9	68.9	33.3	22.2
cytb	2	12.2	1.4	86.5	10.8	12.2
nad1	2	8.3	5.0	86.7	10.0	15.0
nad2	1	0.0	0.0	100.0	0.0	0.0

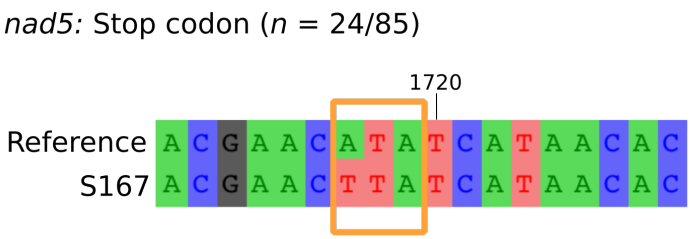


Figure 4.5 | Stop codon identified in a haplotype of the mitochondrial *nad5* gene in red cherry shrimp (*Neocaridina davidi*). The polished reference mitogenome is shown above, with a haplotype sequence from a shrimp sequenced in this study shown below. Sample size is the proportion of individuals found to contain the mutation.

4.3.5 Simulations of point heteroplasmy detection using different variant callers

VarScan and LoFreq showed similar performance in calling known variants in simulated read sets but VarScan was considered the 'best' caller; VarScan achieved a higher score than FreeBayes and LoFreq at all coverages except 2,500x and 10,000x, and consistently achieved the highest power (proportion of true variants called) (Fig. 4.6). VarScan detected ≥ 80% of known variants even at coverages as low as 100x. FreeBayes achieved the lowest score in all read sets, largely based on having a high false positive rate; 3–291 false positives were called at coverages ≤ 1,500x. LoFreq did not make any erroneous calls and VarScan called a maximum of two false positives but only at low coverages (100x and 250x). VarScan was consequently chosen as the variant caller for the analysis on age-related heteroplasmy. Furthermore, a minimum coverage threshold of 1,000x was chosen for downstream analyses; both VarScan and LoFreq achieved ≥ 75% power and accuracy at this coverage and did not call any false positives.

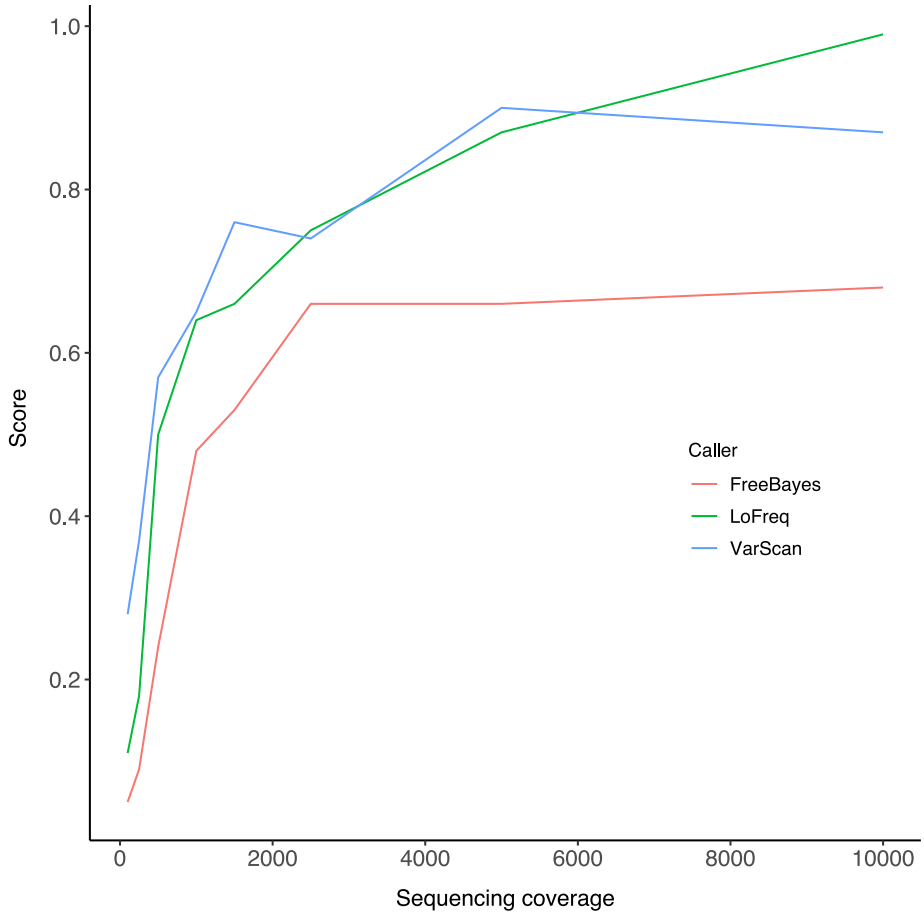


Figure 4.6 | The performance of three variant callers in detecting 200 known variants. Sequencing reads were simulated at eight different coverages using NEAT-genReads and the polished red cherry shrimp (*Neocaridina* davidi) mitogenome. For each caller and coverage, a score (from 0 to 1) was calculated as a function of power, accuracy and false positive rate (see equation 1).

4.3.6 Point heteroplasmy detection and characterisation

In 128 shrimp, 23 filtered point mutations were detected using VarScan across 23 unique positions (Table 4.7). However, after removing a single, outlying individual (see results section 4.3.7), this was reduced to 11 heteroplasmies from 127 shrimp (Table 4.7). Most individuals (*n* = 87; 69%) contained one or more heteroplasmies. Four variants (36%) were only found in one individual, with the remaining heteroplasmies occurring in between 2–58 shrimp (*n* = 7; 64%). Six heteroplasmies (55%) occurred in the non-coding control region. Among the five heteroplasmies in protein-coding genes, just one was a non-synonymous mutation, occurring in *nad2* and predicted to be deleterious (PROVEAN score = -4.0). The majority of heteroplasmies were transition mutations (A:G or C:T; *n* = 8; 73%). The three transversions were either A:C (*n* = 2) or

A:T (*n* = 1) substitutions. No heteroplasmies were detected in the tRNA, rRNA or non-protein-coding genes.

Table 4.7 | The 23 mitochondrial point heteroplasmies detected in 128 red cherry shrimp (*Neocaridina shrimp*). Alternative alleles (Alt) are classified as transition (Ts) or transversion (Tv) mutations. For protein-coding genes, heteroplasmies are synonymous (S) or non-synonymous (NS). All non-synonymous mutations were deemed deleterious with a PROVEAN score less than -2.5. MAF = minor allele frequency. Positions according to the polished mitogenome reference (Ref) assembled in this study. Variants found within the outlying individual in bold and underlined.

Position	Region	Ref	Alt	*n* shrimp	Mean MAF (SD)	Ts/Tv	S/NS
414	***cox1***	**G**	**A**	**1**	**0.01**	**Ts**	**S**
879	***cox1***	**C**	**T**	**1**	**0.01**	**Ts**	**S**
1035	*cox1*	A	G	1	0.01	Ts	S
1660	*cox2*	C	T	2	0.01 (0.00)	Ts	S
1708	***cox2***	**T**	**C**	**1**	**0.01**	**Ts**	**S**
1928	***cox2***	**G**	**A**	**1**	**0.01**	**Ts**	**NS**
2143	***cox2***	**T**	**C**	**1**	**0.02**	**Ts**	**S**
2670	*atp6*	C	T	1	0.01	Ts	S
13426	***12S***	**T**	**C**	**1**	**0.02**	**Ts**	**-**
13429	***12S***	**G**	**A**	**1**	**0.01**	**Ts**	**-**
13516	**control region**	**C**	**T**	**1**	**0.01**	**Ts**	**-**
13521	control region	C	T	2	0.02 (0.00)	Ts	-
13695	**control region**	**C**	**T**	**1**	**0.01**	**Ts**	**-**
13763	**control region**	**A**	**G**	**1**	**0.01**	**Ts**	**-**
13772	control region	T	C	18	0.17 (0.19)	Ts	-
14023	control region	C	T	2	0.23 (0.32)	Ts	-
14030	control region	A	C	1	0.01	Tv	-
14111	control region	T	A	59	0.17 (0.05)	Tv	-
14112	control region	A	T	54	0.07 (0.09)	Tv	-
15087	*nad2*	A	G	3	0.34 (0.12)	Ts	S
15186	***nad2***	**A**	**G**	**1**	**0.01**	**Ts**	**S**
15264	***nad2***	**T**	**A**	**1**	**0.01**	**Tv**	**NS**
15290	*nad2*	G	A	31	0.23 (0.17)	Ts	NS

4.3.7 Point heteroplasmy and age

There was a significant difference in the number of heteroplasmies among shrimp age groups (Poisson GLM: $\chi^2_{(6,128)} = 22.59$; $p < 0.001$) but this effect was driven by a single individual (S14; 40 days old), which was deemed an influential point according to Cook's distance (Supplementary Figures S4.2 and S4.3). The outlying individual

contained an unusually large number of heteroplasmies (*n* = 16) compared to all other individuals, which had between 0–3 variants (Supplementary Fig. S4.3). Twelve heteroplasmies were unique to the outlying individual, all of which existed at low frequency (< 2%). After removing the outlier individual from the dataset, the number of heteroplasmies did not differ among shrimp age groups (Poisson GLM: $\chi^2_{(6,127)}$ = 7.47; p = 0.28; Fig. 4.7) even after removing the egg age class to control for tissue differences (Poisson GLM: $\chi^2_{(5,103)}$ = 7.47; p = 0.19).

Figure 4.7 | The number of point heteroplasmies in seven age groups of red cherry shrimp (*Neocaridina davidi*) (*n* = 127). Boxplots display a mean dot (solid black point), median line, inter-quartile range (IQR) boxes, 1.5*IQR whiskers, and data points.

4.3.8 Heteroplasmy pipeline validation

The number of variants detected by VarScan after filtering was consistent across 71% of both PCR and sequencing replicates (Supplementary Table S4.2). Where the same variant was observed across replicates, the called frequencies were within 1% of each other across 63% of PCR replicates and 100% of sequencing replicates (Supplementary Table S4.3).

4.4 Discussion

This study aimed to investigate whether the number of mtDNA point mutations could be used to differentiate seven age groups of red cherry shrimp sampled from 7–210 days post-fertilisation. Whole mitogenomes were sequenced to an average depth of 2,938x. After removing an outlying individual, eleven unique point mutations were detected in 6 kb of the mitogenome across 127 shrimp. Individual shrimp contained between 0–3 heteroplasmies. The number of heteroplasmies did not differ among shrimp age cohorts. Multiple, within-individual haplotypes were identified across the remaining mitogenome sequence (9.5 kb).

Initial investigation of age-related heteroplasmy in red cherry shrimp revealed a significant difference in the number of point heteroplasmies with age but this was driven by a single individual with a disproportionate number of heteroplasmies (*n* = 16). This individual was deemed an influential point according to Cook's distance. Furthermore, this shrimp was sampled at a young age (40 days) and there is no *a priori* biological reason to expect a larger number of heteroplasmies at this point. Excluding this individual negated the association between heteroplasmy and age. The lack of age-related point heteroplasmy concurs with recent findings in greater mouse-eared bats (Jebb et al. 2018), but is in contrast to observations in humans (Li et al. 2015) and mice (Arbeithuber et al. 2020). Differences in age-related heteroplasmy among species do not appear to be driven by differences in lifespan; mice, for example, accumulate mutations faster than longer-lived humans (Wang et al. 1997; Trifunovic 2006). In this study, shrimp were sampled across ca. 54% of the species' typical lifespan in captivity (Schoolmann & Arndt 2018). If mtDNA mutations accumulate later in life, changes with age will have been missed. However, even if this is the case, a marker of chronological age that only differentiates older adults is unlikely to be of practical use in management or conservation biology. It is also possible to overlook age-related heteroplasmy as a result or measuring heteroplasmy in a tissue or mtDNA region that is less likely to accumulate mtDNA mutations. These possibilities are discussed below.

DNA for all post-hatching shrimp was extracted from walking legs, which are comprised largely of muscle tissue (Cattaert & Edwards 2017). This tissue was chosen because it can be non-destructively sampled, which is an important consideration if a crustacean ageing tool is to be adopted in fisheries management. Post-mitotic, energetically-active tissues (such as muscle) tend to contain higher proportions of mutated mtDNA (Payne & Chinnery 2015) and muscle tissue accumulates mtDNA mutations in various species (Hiona et al. 2010; Naue et al. 2015). This suggests that age-related heteroplasmy in shrimp is unlikely to have been missed as a consequence of tissue choice. However, crustaceans can regrow lost legs (and other appendages) following injury or autotomy, and the effect of tissue regeneration on heteroplasmy is not known. If accumulated mtDNA mutations in leg tissue are removed during limb loss, this would erase any age-related signature. However, limb loss is unlikely to explain the lack of relationship identified here. No shrimp were observed to be missing legs, likely because they were raised in a benign environment with access to plenty of food and no predators. Therefore, there would have been limited (among conspecifics) or no (with predators) fighting that could otherwise cause limb loss. However, moulting also involves the loss and regeneration of some underlying muscle and connective tissue (Götze & Saborowski 2011). The potential impact of moulting on heteroplasmy in leg tissue requires further investigation but age-related patters are unlikely to be completely erased.

By only characterising heteroplasmy in a portion (ca. 39%) of the red cherry shrimp mitogenome, it is possible that age-related patterns elsewhere have been overlooked. This can occur if heteroplasmy is non-random, with mutations occurring preferentially in specific regions. In the current study, age-related heteroplasmy was investigated in the non-coding control region, four protein-coding genes (*nad2*, *cox1*, *cox2* and partial *nad2*), and the 12S rRNA gene. In humans and mice, age-related heteroplasmies tend to be evenly-distributed along the mitogenome or disproportionately high in the control region (Li et al. 2010; Samuels et al. 2013; Arbeithuber et al. 2020). Age-related heteroplasmies have been identified in the control region of human and mouse muscle tissue (Michikawa et al. 1999; Wang et al. 2001; Thèves et al. 2006; Naue et al. 2015; Arbeithuber et al. 2020). In red cherry shrimp, more point heteroplasmies were

detected in the control region (*n* = 6) than the other regions combined (*n* = 5). The control region is hypervariable in other crustaceans, including other shrimp (*Penaeus merguiensis*; Chu et al. 2003) and Atlantic spiny lobsters (*Panulirus argus*; Diniz et al. 2005). These observations indicate that age-related heteroplasmy is unlikely to have been missed in this study as a consequence of focusing on a portion of the mitogenome.

While it was not the focus of this study, some inferences can be made on the most likely cause of heteroplasmy in red cherry shrimp. First, the majority of mutations were transition substitutions (*n* = 8; 73%). A bias towards transitions is consistent with polymerase errors during mtDNA replication (Spelbrink et al. 2000). Similar substitution patterns have been observed in mice (Ma et al. 2018), bats (Jebb et al. 2018), humans (Kennedy et al. 2013) and fruit flies (Itsara et al. 2014), implying that oxidative damage is not a significant contributor to heteroplasmy across diverse species. Oxidative damage typically leads to G:C ⇌ T:A transversions (Cheng et al. 1992). In the current study, just one such mutation (A → C) was observed in the control region of a single individual. Longitudinal sampling would be required to shed light on whether oxidative mutations rarely occur in red cherry shrimp, or whether there is purging of such mutations within individuals, or selective disappearance of shrimp with large amounts of oxidative damage. Most mutations (*n* = 8; 73%) were shared across multiple individuals (among 2–58 shrimp). Intuitively, it can be assumed that shared heteroplasmies are more likely to be inherited, but human studies have shown that somatic mutations can re-occur in unrelated individuals (Samuels et al. 2013). Without information on parental genotypes, and having only measured heteroplasmy in a single tissue from each individual, it is not possible to determine whether the mutations are inherited or somatic. Heteroplasmies found in a single tissue are likely to be somatic, having spread via cell division following a mutation event (Samuels et al. 2013). Mutations found across multiple tissues are probably inherited (Samuels et al. 2013; Li et al. 2015).

Throughout the remaining 9.5 kb of the shrimp mitogenome, up to 44% of shrimp contained one or more additional haplotypes across protein-coding genes. These

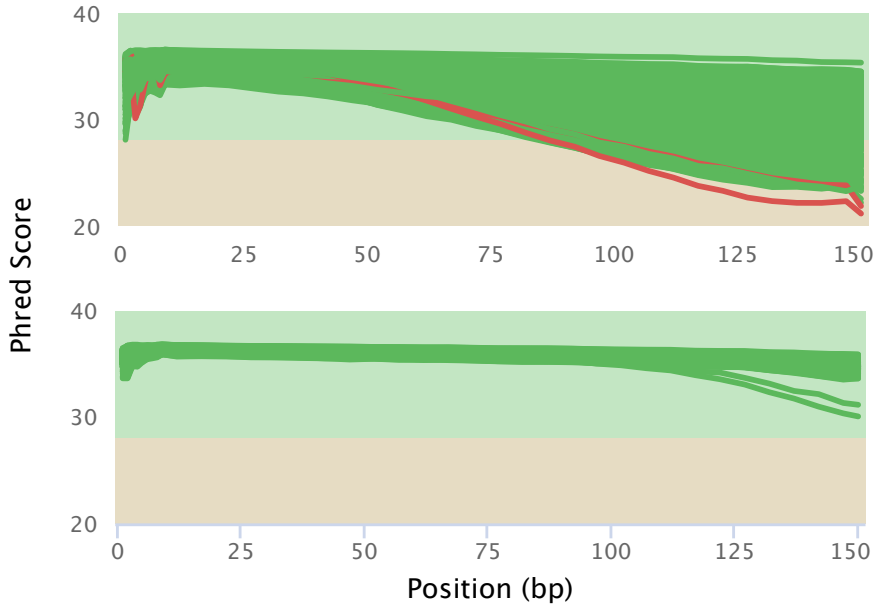
patterns are indicative of inherited heteroplasmy or pseudo heteroplasmy from NUMT, or other, contamination. Authentic heteroplasmy is unlikely due to the large number of variant sites (> 8% of gene length in some cases) and evidence of loss-of-function found in some haplotypes (e.g. a stop codon in *nad5*). All haplotypes were most similar to mtDNA from *Neocaridina* ssp., which rules out contamination from non-shrimp DNA. Instead, features of the additional haplotypes are most consistent with NUMT contamination. First, patterns of haplotype and nucleotide diversity were highly amplicon-specific. For amplicon 1, just one or two closely-related haplotypes were identified. Across amplicon 2, there were between 8–22 unique haplotypes with high levels of sequence divergence (mean and maximum pairwise *p*-distances of 2.05% and 8.53%, respectively). The amplicon-specificity of these patterns suggests they are related to primer properties – i.e. unspecific primer binding in amplicon 2 (in this case, probable co-amplification of nuclear DNA). Second, mitochondrial sequences are typified by an abundance of synonymous, transition mutations occurring on the third codon position (Baldo et al. 2011). Just four mutations were identified in amplicon 1, all of which were transitions and only one was a non-synonymous substitution affecting the second codon position. In amplicon 2, between 10–33.3% of mutations were non-synonymous, up to 28.6% were transversions, as many as 31.1% occurred on the first or second codon positions, and a prevalent stop codon was amplified.

NUMT contamination can interfere with identifying true heteroplasmy, particularly in NGS studies that involve mapping short reads to reference sequences based on sequence similarity. If mitochondrial and NUMT sequences are similar enough to map to the same location, it is difficult to determine the origin of each sequence, resulting in false (either positive or negative) heteroplasmy calls (Maude et al. 2019). Therefore, NUMT contamination needs careful consideration in similar future studies and could easily have been missed in this study. Shrimp mtDNA was amplified using long-range PCR, and this approach is often assumed to avoid co-amplifying NUMTs, which are typically shorter than 1,000 bp (Richly & Leister 2004; Wang et al. 2019). However, NUMTs larger than 7,000–15,000 bp have been identified in mammals (Triant & DeWoody 2007) and invertebrates (Sun & Yang 2016). NUMT contamination is difficult to remove *in silico* and not without compromising sequencing coverage (Duan et al.

2018; see chapter 5 section 5.3.6). Bioinformatic removal is further complicated when NUMTs show large inter-individual variation (Stewart & Chinnery 2015). This appears to be the case for red cherry shrimp, even though individuals are likely to have originated from aquaria stocks of limited genetic diversity. Even popular methods to isolate mtDNA prior to sequencing, such as mtDNA capture hybridisation or density-dependent centrifugation, fail to eliminate NUMTs (Li et al. 2012; Stewart & Chinnery 2015).

Overall, the results suggest that heteroplasmy is not likely to be a viable marker of chronological age in red cherry shrimp. Age-related changes in heteroplasmy were not observed throughout ca. 54% of the red cherry shrimp lifespan and across over a third of the mitochondrial genome, including the hyper-variable control region. Features of the point mutations identified in red cherry shrimp suggest they are not caused by oxidative damage. Amplification of the remaining region appears to have been affected by NUMT contamination. It may be possible to minimise co-amplification of nuclear DNA by designing new primers. However, this is likely to be complicated. A large region was affected by nuclear contamination and there was high divergence in nuclear haplotypes among individuals, which leaves a small window for identifying primers that are specific to mtDNA and conserved across individuals.

4.5 Supplementary information



Supplementary Figure S4.1 | Mean quality scores for Illumina NovaSeq reads of *Neocaridina davidi* mtDNA unfiltered (top) and filtered (bottom) to retain reads if at least 80% of bases had a quality score more than or equal to Q30. Scores for each fastq file were calculated using FastQC and compiled into a single plot using MultiQC (Ewel et al. 2016).

Supplementary Table S4.1 | Top Blast hit for each mitochondrial haplotype reconstructed in this study. All haplotypes were most similar to *Neocaridina* shrimp mtDNA. Frequency = the number of shrimp containing that haplotype. % identity = the percentage of shared nucleotides with the sequence in the GenBank database.

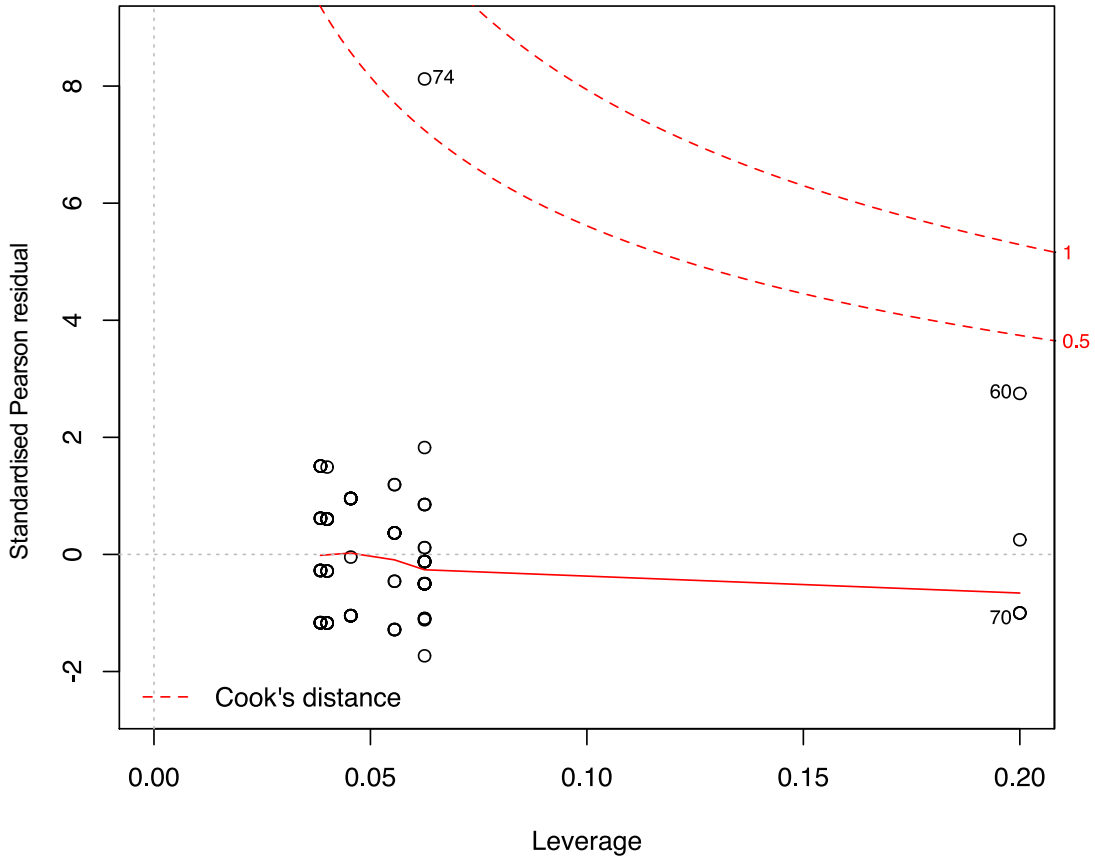
Gene	Haplotype	Frequency	Top BLAST hit	% identity
cox1	I	59	*N davidi*	98.95
cox1	II	73	*N. davidi*	98.88
cox2	I	100	*N. davidi*	98.98
cox2	II	41	*N. davidi*	98.94
atp8	I	132	*N. davidi*	98.74
atp6	I	59	*N. davidi*	98.37
atp6	II	83	*N. davidi*	98.52
atp6	III	17	*N. davidi*	98.37
cox3	I	139	*N. davidi*	98.47
nad3	I	130	*N. davidi*	99.72
nad5	I	43	*N. davidi*	98.67
nad5	II	42	*N. davidi*	98.73
nad5	III	7	*N. davidi*	98.67
nad5	IV	2	*N. davidi*	96.99
nad5	V	4	*N. davidi*	96.64
nad5	VI	1	*N. davidi*	97.05
nad5	VII	11	*N. davidi*	96.59
nad5	VIII	2	*N. davidi*	98.55
nad5	IX	1	*N. davidi*	97.63
nad5	X	5	*N. davidi*	96.99
nad5	XI	3	*N. davidi*	98.26
nad5	XII	3	*N. davidi*	96.59
nad5	XIII	1	*N. davidi*	97.63
nad5	XIV	1	*N. davidi*	98.67
nad5	XV	1	*N. davidi*	97.51
nad4	I	37	*N. davidi*	98.36
nad4	II	27	*N. davidi*	98.36
nad4	III	2	*N. davidi*	97.09
nad4	IV	9	*N. heteropoda koreana*	94.17
nad4	V	25	*N. davidi*	98.36
nad4	VI	9	*N. heteropoda koreana*	94.25
nad4	VII	5	*N. heteropoda koreana*	94.17
nad4	VIII	1	*N. davidi*	98.36
nad4	IX	4	*N. davidi*	97.09
nad4	X	5	*N. heteropoda koreana*	94.10
nad4	XI	3	*N. davidi*	97.39
nad4	XII	3	*N. davidi*	97.76
nad4	XIII	1	*N. davidi*	98.28
nad4	XIV	1	*N. davidi*	94.40
nad4	XV	1	*N. davidi*	95.67
nad4	XVI	1	*N. davidi*	94.85
nad4	XVII	1	*N. davidi*	95.74
nad4	XVIII	1	*N. davidi*	97.61
nad4	XIX	1	*N. heteropoda koreana*	94.10
nad4	XX	1	*N. davidi*	94.10

Supplementary Table S4.1 (continued)

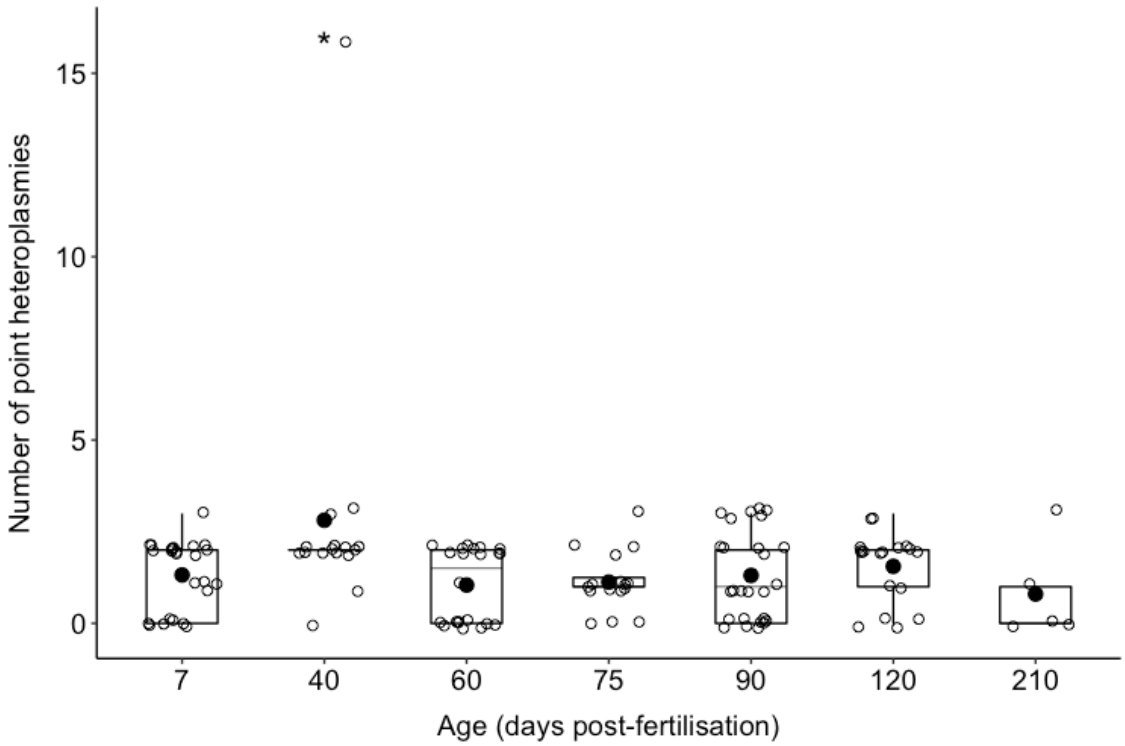
Gene	Haplotype	Frequency	Top BLAST hit	% identity
nad4	XXI	1	*N. davidi*	98.06
nad4	XXII	1	*N. davidi*	94.25
nad4l	I	91	*N. davidi*	98.68
nad4l	II	26	*N. davidi*	96.37
nad4l	III	1	*N. davidi*	97.36
nad4l	IV	42	*N. davidi*	98.35
nad4l	V	2	*N. davidi*	99.34
nad4l	VI	3	*N. davidi*	97.03
nad4l	VII	1	*N. davidi*	98.68
nad4l	VIII	1	*N. davidi*	97.03
nad4l	IX	1	*N. davidi*	96.70
nad6	I	62	*N. davidi*	99.22
nad6	II	29	*N. davidi*	99.03
nad6	III	26	*N. heteropoda koreana*	94.57
nad6	IV	4	*N. heteropoda koreana*	95.93
nad6	V	1	*N. davidi*	97.87
nad6	VI	1	*N. davidi*	95.93
nad6	VII	1	*N. davidi*	95.16
nad6	VIII	1	*N. heteropoda koreana*	93.60
cytb	I	50	*N. davidi*	98.94
cytb	II	73	*N. davidi*	99.03
cytb	III	23	*N. heteropoda koreana*	94.11
cytb	IV	1	*N. davidi*	98.42
cytb	V	1	*N. davidi*	98.94
cytb	VI	3	*N. davidi*	98.86
cytb	VII	1	*N. davidi*	96.57
cytb	VIII	1	*N. davidi*	97.27
cytb	IX	1	*N. davidi*	97.10
cytb	X	1	*N. davidi*	95.07
cytb	XI	1	*N. davidi*	96.39
cytb	XII	1	*N. davidi*	95.51
cytb	XIII	1	*N. davidi*	97.01
cytb	XIV	1	*N. heteropoda koreana*	94.20
cytb	XV	1	*N. davidi*	96.92
cytb	XVI	1	*N. davidi*	97.10
cytb	XVII	1	*N. davidi*	97.10
nad1	I	43	*N. davidi*	98.41
nad1	II	59	*N. davidi*	98.51
nad1	III	25	*N. heteropoda koreana*	94.59
nad1	IV	1	*N. davidi*	96.50
nad1	V	2	*N. davidi*	98.73
nad1	VI	2	*N. davidi*	98.83
nad1	VII	1	*N. davidi*	94.37
nad1	VIII	2	*N. heteropoda koreana*	94.27
nad1	IX	1	*N. davidi*	98.30
nad1	X	1	*N. davidi*	98.51
nad1	XI	1	*N. davidi*	98.41
nad1	XII	1	*N. davidi*	98.41
nad1	XIII	1	*N. davidi*	97.13

Supplementary Table S4.1 (continued)

Gene	Haplotype	Frequency	Top BLAST hit	% identity
nad1	XIV	1	*N. heteropoda koreana*	94.37
nad2	I	103	*N. davidi*	98.61
nad2	II	4	*N. davidi*	98.71



Supplementary Figure S4.2 | Diagnostic plot for a generalized linear model of the number of heteroplasmies with age in red cherry shrimp (*Neocaridina davidi*). Influential points are identified as having a Cook's distance > 0.5 (red dashed line). Influential point 74 corresponds to a 40-day-old shrimp (S14).



Supplementary Figure S4.3 | The number of point heteroplasmies in seven age groups of red cherry shrimp (*Neocaridina davidi*) (*n* = 128). Data point marked with an asterisk was deemed an influential point according to Cook's distance (Supplementary Fig. S4.2). Boxplots display a mean dot (solid black point), median line, inter-quartile range (IQR) boxes, 1.5*IQR whiskers, and data points.

Supplementary Table S4.2 | The number of heteroplasmies called across PCR and NovaSeq sequencing (Seq) replicates in a portion of the mitogenome of eight red cherry shrimp (*Neocaridina davidi*). Counts identical across either replicate are in bold and underlined. Empty cells are where an individual failed sequencing or mapping quality control.

Individual ID	PCR 1	PCR 2	Seq 1	Seq 2
S10	2	3	**3**	**3**
S50	**3**	**3**	**2**	**2**
S61	**0**	**0**	2	0
S144	**0**	**0**	**0**	**0**
S178	**2**	**2**	**2**	**2**
S204	**0**	**0**	0	-
S238	1	-	0	1
S239	3	2	**3**	**3**

Supplementary Table S4.3 | The minor allele frequencies of heteroplasmies called across PCR and NovaSeq sequencing (Seq) replicates in a portion of the mitogenome of six red cherry shrimp (*Neocaridina davidi*). Frequencies are in bold and underlined when the same variant was called across either replicate within a frequency of 1% (0.01) of each other. Empty cells are where an individual failed sequencing or mapping quality control.

Individual ID	position	Ref	Alt	PCR 1	PCR 2	Seq 1	Seq 2
S10	13772	T	C	0.42	0.38	**0.44**	**0.43**
S10	14111	T	A	0.07	-	**0.06**	**0.07**
S10	15290	G	A	0.41	0.38	**0.42**	**0.43**
S50	13516	C	T	0.01	-	-	-
S50	14030	A	C	0.01	-	-	-
S50	14111	T	A	0.18	0.20	**0.18**	**0.17**
S50	14112	A	T	**0.04**	**0.04**	**0.04**	**0.04**
S61	14111	T	A	-	-	0.19	-
S61	14112	A	T	-	-	0.04	-
S178	14111	T	A	**0.17**	**0.17**	**0.19**	**0.19**
S178	14112	A	T	**0.04**	**0.04**	**0.04**	**0.04**
S238	15290	G	A	0.50	-	-	0.45
S239	13772	T	C	0.02	-	**0.02**	**0.01**
S239	14111	T	A	**0.07**	**0.06**	**0.06**	**0.07**
S239	15290	G	A	**0.03**	**0.02**	**0.02**	**0.02**

Chapter 5

Age-related heteroplasmy and the complexity of sequencing mitochondrial DNA in European lobsters (*Homarus gammarus*)

5.0 Abstract

Understanding the age structure of animal populations is crucial for predicting population dynamics and forms the basis of sustainable fisheries management. Crustaceans are difficult to age because they moult throughout their lives and their growth rate varies according to age and the environment. In humans and some vertebrate animal models, mitochondrial DNA (mtDNA) mutations accumulate throughout life, and could be a useful marker of age. This study investigates whether the number of mtDNA mutations varies with age in European lobsters (*Homarus gammarus*). The mitogenomes of lobsters belonging to six age groups (spanning 1 day to ≥ 4 years old) were amplified and sequenced to high depth. The mitogenome contains a large tandem duplication (8.9 kb) and issues were encountered involving reads mis-mapping in this region. In the remaining region, some, if not all, lobsters contained multiple haplotypes, and there was evidence that these additional haplotypes belong to nuclear mitochondrial DNA (NUMTs). A range of bioinformatic approaches aimed at eliminating mis-mapping and removing NUMTs was attempted. Unfortunately, it was not possible to overcome mis-mapping without compromising sequencing depth and so the duplicated region was excluded from further analyses. For the remaining region, increasing the mapping seed length successfully reduced NUMT-derived reads. After thorough filtering of mutations to remove residual false positives, 118 point mutations were identified in 143 individuals across 9.8 kb of the mitogenome. No difference in the number of mtDNA mutations was observed among age groups. This, combined with the apparent complexity of mapping and sequencing mtDNA in European lobsters, suggests that heteroplasmy is not a viable marker of chronological age in this species.

5.1 Introduction

Accurate determination of animal age is crucial for wildlife management and conservation and forms the basis of predicting population dynamics for sustainable fisheries management (Campana 2001). Some animals can be aged using morphological features, such as counting growth rings in fish otoliths (Panella 1971), but many species do not have physical features that change with age. Molecular markers of age could provide a solution to this problem and are the subject of growing attention among conservation ecologists (reviewed by Jarman et al. 2015; De Paoli-Iseppi 2017). Development of such markers involves the identification of features of DNA or RNA, or associated molecules, that changes consistently with age among individuals. Examples include telomere length attrition (e.g. Haussmann & Vleck 2002), levels of DNA methylation (e.g. De Paoli-Iseppi et al. 2019) and the accumulation of mitochondrial DNA (mtDNA) mutations (e.g. Thèves et al. 2006). Age-related changes in telomere length and DNA methylation have been the focus of a large number of studies (reviewed by Dunshea et al. 2011 and De Paoli-Iseppi 2017, respectively) but little attention has been given to mtDNA mutation accumulation (Jebb et al. 2018).

Mitochondria are the only animal organelle to contain their own genetic information. Mitochondrial DNA is a short (15–17 kb), circular molecule that exists in multiple copies per cell (up to several thousand) (Stewart & Chinnery 2015). The mitogenome includes 13 protein-coding genes essential for the electron transport chain, 22 translation RNAs (tRNAs) and two ribosomal RNAs (rRNAs) necessary for DNA translation, and a non-coding control region. Mitochondrial DNA is predominantly maternally inherited and mutates relatively rapidly (10–20 times faster than nuclear DNA: Payne & Chinnery). Consequently, mtDNA has been widely used in animal phylogenetic and phylogeography studies (Hurst & Jiggins 2005). However, such studies typically rely on the assumption that all mtDNA copies within an individual are identical (Mao et al. 2014). With the development of next-generation sequencing (NGS), it is increasingly recognised that many animals carry multiple differing mitogenomes, known as heteroplasmy. Different copies of mtDNA can differ at individual nucleotides (point heteroplasmy) or in length (length heteroplasmy).

Heteroplasmy may originate from the inheritance of two or more different mtDNA copies and/or *de novo* somatic mutation generating novel variation (Lawless et al. 2020). Inherited heteroplasmy usually occurs via maternal transmission of heteroplasmic mtDNA in egg cells and, rarely, because of the inheritance of paternal mtDNA (Polovina et al. 2020). Multiple mechanisms exist to destroy sperm mtDNA to reduce the spread of potentially harmful mutations (Polovina et al. 2020). Somatic mutations may result from oxidative damage or polymerase errors during mtDNA replication (Lawless et al. 2020).

Heteroplasmy has been shown to accumulate over time within individuals in humans and model vertebrate species, including rhesus monkeys (*Macaca mulatta*) and mice (*Mus musculus*) (Larsson 2010 and references therein; Li et al. 2015; Arbeithuber et al. 2020). This build-up of mtDNA mutations is thought to occur due to accumulating oxidative damage or via the clonal expansion of early, pre-existing mutations (Lawless et al. 2020). As a by-product of its role in cellular respiration, mtDNA produces large quantities of reactive oxygen species (ROS) leaving it exposed to oxidative damage (Trifunovic 2006). This damage can lead to a 'vicious cycle' of reduced respiratory efficiency, causing enhanced ROS production and further mtDNA mutations (Harman 1972; Payne & Chinnery 2015). Alternatively, clonal expansion of existing mutations can occur in dividing or non-dividing cells. In dividing cells, there is unequal, random partitioning of mtDNA molecules to the daughter cells, which leads to shifts in the level of heteroplasmy over cell generations (Stewart & Chinnery 2015). In non-dividing cells, mtDNA continuously self-replicates and there may be preferential replication of mutant copies (Stewart & Chinnery 2015).

In humans, the age-related increase of mtDNA mutations has been suggested as a tool for estimating age in forensic investigations (Meissner & Ritz-Timme 2010). Mitochondrial DNA is particularly well-suited as a molecular marker of age because it has a high copy number per cell and its robust circular shape means it is possible to extract genetic material from degraded tissues (Just et al. 2015). The cost of sequencing mtDNA is another key consideration, which makes this potential marker attractive for studies of wild animals (Jarman et al. 2015). The small size of mtDNA

makes it possible to screen for heteroplasmy within whole genomes at a relatively low cost. Furthermore, in humans, it has been possible to estimate age using the frequency of single mutations: for example, a common mtDNA deletion of 4,977 bp (Meissner et al. 1999; von Wurmb-Schwark et al. 2002) and a single point mutation (A189G: Thèves et al. 2006) accumulate with age across unrelated individuals. Ageing assays based on small regions of mtDNA could drastically reduce the costs of heteroplasmy-based markers of age and allow for higher-throughput screening. However, despite the potential of heteroplasmy-based markers of age in animal ecology and conservation, just one study to date has explored the link between heteroplasmy and age in a wild animal (Jebb et al. 2018); no difference in the number of mtDNA point mutations was observed among different age groups of greater mouse-eared bats (*Myotis myotis*: Jebb et al. 2018).

The slow progress in investigating age-related heteroplasmy in different species may be due to the technical and/or analytical challenges of identifying mtDNA mutations. Most mtDNA mutations exist at low frequency (< 5%; Jebb et al. 2018) but traditional Sanger sequencing methods can only detect mutations with a frequency of ca. 15% (Duan et al. 2018). For this reason, heteroplasmy was once thought to be rare (Stewart & Chinnery 2015). Next-generation sequencing is not limited by sensitivity and is able to detect even very low-level heteroplasmy (ca. 1%; Duan et al. 2018). However, NGS comes with its own challenges. Sequencing error is common to all sequencing technologies but is typically higher for NGS (Wang et al. 2012). Failing to account for sequencing error can lead to false positive variant calls. Finally, both Sanger and NGS carry the risk of co-amplifying NUMTs, which can be difficult to avoid (Li et al. 2012; Duan et al. 2018) and can artificially inflate polymorphism estimates.

The present study aims to investigate whether the number of mtDNA point mutations can be used as a proxy marker of chronological age in European lobsters (*Homarus gammarus*). This species is harvested throughout its range, and lobster fishing plays a crucial role in supporting regional economies and livelihoods in local fishing communities (Ellis et al. 2015). Like other crustaceans, the European lobster is very difficult to age accurately (Vogt 2012; Kilada & Driscoll 2017). Throughout life,

crustaceans moult and regenerate their exoskeleton and therefore lose many of their hard, calcified structures that might otherwise accumulate measurable growth rings (Kilada & Driscoll 2017). Crustacean stock assessments currently rely on size-structured models based on measurements of carapace length (CL) (Vogt 2012; Kilada & Driscoll 2017). However, CL is weakly correlated with age in crustaceans due to environmental and age-related effects on growth rate (Vogt 2012; Kilada & Driscoll 2017). European lobster stock declines in Norway (Agnalt et al. 2007) and the Mediterranean (Spanier et al. 2015) have set a precedent for ensuring that stocks are managed sustainably. To this end, an accurate method for estimating age in European lobsters (and other crustaceans) would have considerable benefits for the conservation ecology of this economically-important species.

Point heteroplasmy was characterised across the mitochondrial genome of lobsters using deep NGS. However, technical issues were identified with these data, which confounded the analysis of age-related heteroplasmy. Namely, some lobsters contained linked single-nucleotide polymorphisms (SNPs), where neighbouring SNPs occur at similar frequency, throughout the mitogenome. Such patterns are suggestive of genuine heteroplasmy (e.g. inheritance of more than one mitogenome) or pseudo heteroplasmy (e.g. mis-mapping or NUMTs). Therefore, the first aim of this study was to determine the most likely cause of linked SNPs. The second aim was to investigate whether reads containing linked SNPs could be removed using alternative mapping methods. The third aim was to compare the number of mtDNA point mutations across six lobster age cohorts. This study is one of the first to investigate age-dependent heteroplasmy outside model organisms (Jebb et al. 2018).

5.2 Methods

5.2.1 Study species and sampling

Tissue samples (claws) were collected from five cohorts of known-age European lobsters (0–2 years) reared at the National Lobster Hatchery (NLH) in Cornwall, UK (Table 5.1). Unknown-age, wild lobsters, caught off the coast of Cornwall, were sampled by clipping the terminal end of an antenna (Table 5.1). All wild-caught lobsters were above the minimum landing size (89–136 mm CL) and therefore estimated to be mature individuals ≥ 4 years of age. This estimated minimum age is based on size-at-age data from previous mark-recapture studies (Bannister & Addison 1998; Uglem et al. 2005; Schmalenbach et al. 2011).

Table 5.1 | Sample demography of the 156 European lobsters (*Homarus gammarus*) sequenced in this study. Sample sizes for the 143 individuals that passed sequencing and mapping quality control, and used for the analysis on age-related heteroplasmy, in parentheses. Ages are time post-hatching. Error in age estimates arises for individuals that were graded by moult stage rather than hatch date. Lobsters ≥ 7 months-old reared at the National Lobster Hatchery (NLH) were deployed into sea-based containers (Daniels et al. 2015) off the coast of Cornwall, UK, ca. 1-month post-hatching. CL = carapace length. All lobsters were sampled in 2018.

Source	Age at sampling	Age uncertainty	Tissue	*n*
NLH	1 day	0	Claw	29 (27)
NLH	8 weeks	0	Claw	27 (26)
NLH	7 months	± 14 days	Claw	25 (25)
NLH	1 year	± 14 days	Claw	19 (19)
NLH	2 years	± 14 days	Claw	27 (25)
Wild caught, Cornwall	≥ 4 years	Unknown	Antenna	29 (21)

5.2.2 DNA extraction and long-range mtDNA amplification

Genomic DNA was extracted from ca. 2 mm^3 of tissue using a salt-precipitation technique modified from Aljanabi and Martinez (1997) and resuspended in H_2O. DNA

concentration and purity were verified using a NanoDrop 8000 Spectrophotometer (Thermo Scientific). DNA extracts were diluted to 25 ng/µl in ddH_2O.

Two long-range primer pairs were designed to amplify the entire *H. gammarus* mitogenome (Table 5.2). Primers were designed according to the reference mitogenome available at the time (GenBank accession number: KC107810; Shen et al. 2013). Primer3 (Rozen & Skaletsky 2000) was used to ensure compatible annealing temperatures, appropriate GC content (40–60%), and to minimise secondary structures (hairpins) and primer dimer formation. Overlapping primers are optimal for NGS because the number of sequencing reads tends to follow a normal distribution across the target sequence. Multiple attempts were made to develop overlapping primers for the European lobster mitogenome but without successful amplification. Therefore, the two primer pairs were the reverse complement of each other to complete amplification of the circular molecule. Gan et al. (2019) later discovered that the reference mitogenome used for primer design (and the only reference sequence available at the time) incorrectly reported missing genes (*nad2* and a few tRNAs) and failed to identify a large tandem duplication (Fig. 5.1). Subsequent analysis steps involved the revised European lobster mitogenome (GenBank accession number: MH747083).

Table 5.2 | Primers designed to PCR amplify the entire European lobster (*Homarus gammarus*) mitogenome. Annealing positions according to the reference sequence assembled by Gan et al. 2019 (GenBank accession number: MH747083). Total length = 20,303 bp.

Primer name	Primer sequence (5' → 3')	Annealing position	Annealing (°C)	Extension (minutes)
Hom_mtDNA_5F Hom_mtDNA_5R	ACC TGC CGT AAC TAG AGT GG AAG GGC CGC GGT ATT TTA AC	9841 15745	60	6
Hom_mtDNA_9F Hom_mtDNA_9R	GTT AAA ATA CCG CGG CCC TT CCA CTC TAG TTA CGG CAG GT	15726 9860	62	8

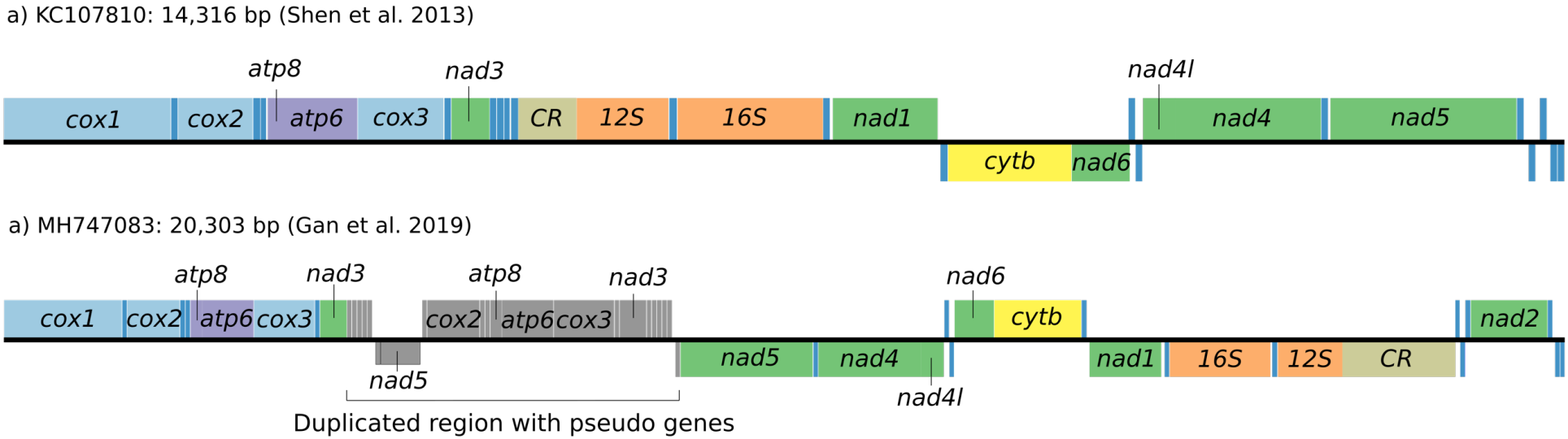


Figure 5.1 | Gene arrangements for two mitogenome sequences for the European lobster (*Homarus gammarus*). The mitogenome assembled by Shen et al. 2013 (a) was used for primer design in this study. This reference mitogenome was later found by Gan et al. 2019 to have incorrectly reported missing genes (*nad2* and a few tRNAs) and failed to identify a large tandem duplication. Subsequent analysis steps involved the revised European lobster mitogenome (b). Transfer RNAs are in darker blue, or unlabelled grey sections (within the duplication). Gene arrangements created using OGDRAW (version 1.3.1; Greiner et al. 2019).

Polymerase chain reactions (PCRs) were performed in 25 µl, consisting 12.5 µl GoTaq® Long PCR Master Mix (Promega), 2.5 µl (10 µM) each primer, 6.5 µl ddH_2O and 1 µl DNA. Thermal cycle conditions were initial denaturation at 94°C for 2 minutes, followed by 25 cycles of denaturation (92°C, 30 seconds), annealing (20 seconds) and extension (65°C), with a final extension step at 72°C for 10 minutes. Optimised annealing temperatures and initial extension times are provided for each primer pair in Table 5.2. Successful amplification was verified on a 0.8% agarose gel. DNA samples from lobsters of different ages were randomly distributed across PCR plates to minimise plate effects.

5.2.3 Library preparation and sequencing using Illumina NovaSeq

Library preparation and sequencing were delivered by the Genomics Pipelines Group at the Earlham Institute in Norwich, UK. Libraries were constructed following a bespoke protocol based on the Illumina Nextera kit, in which DNA is simultaneously fragmented and tagged with sequencing adapters in a single, rapid, enzymatic reaction. PCR amplicons from the same individual were quantified and pooled to equimolar concentration. Enzymatic reactions were performed to cleave and tag DNA with a universal overhang; 1 ng of pooled amplicon DNA was combined with 0.9 µl of Nextera reaction buffer and 0.1 µl Nextera enzyme in a 5 µl reaction volume and incubated for 10 minutes at 55˚C.

Individuals were barcoded by adding 2.5 µl of 2 µM custom barcoded P5 (5'-AATGATACGGCGACCACCGAGATCTACAC [barcode] TCGTCGGCAGCGTC-3') and P7 (5'-CAAGCAGAAGACGGCATACGAGAT [barcode] GTCTCGTGGGCTCGG-3') compatible primers, 5 µl 5x Kapa Robust 2G reaction buffer, 0.5 µl 10 mM dNTPs, 0.1 µl Kapa Robust 2G enzyme and 10.4 µl water. Libraries were amplified by incubating the sample for 72˚C for 3 minutes, followed by 14 PCR cycles consisting of 95˚C for 1 minute, 65˚C for 20 seconds and 72˚C for 3 minutes.

Libraries were cleaned using magnetic beads to precipitate DNA molecules larger than 200 bp. 20 µl of amplified DNA was mixed with 20 µl of Kapa beads and incubated at room temperature for 5 minutes. Beads were pelleted on a magnetic particle

concentrator, the supernatant removed, and two 70% ethanol washes performed. Beads were left to dry for 5 minutes at room temperature before being resuspended with 20 µl of 10 mM Tris-HCl and incubated at room temperature for 5 minutes to elute the DNA. The tube was returned to the concentrator, the beads allowed to pellet, and the aqueous phase containing the size-selected DNA molecules transferred to a new tube.

The size distribution of each purified library was determined by running 3 µl DNA mixed with 18 µl 10 mM Tris-HCl on a Perkin Elmer GX. To generate equimolar libraries, pooled amplicons were size selected on a Sage Science 1.5% BluePippin cassette recovering molecules between 400–600 bp. Quality control of the size-selected pool was performed by running 1 µl aliquots on a Life Technologies Qubit high sensitivity assay and an Agilent DNA High Sense BioAnalyser chip and the concentration of viable libraries measured using quantitative PCR. For sequencing, library pools were loaded at 120 pM based on an average of the Qubit and quantitative PCR concentrations, both calculated using an average molecule size of 425 bp, and sequenced on the Illumina NovaSeq 6000 with a 2 x 150 bp read metric on a single SP flow cell.

5.2.4 Sequence quality control

Sequences were de-multiplexed and trimmed of adapter sequences and unique barcodes using bcl2fastq2 (Illumina). Raw and filtered sequences were quality checked using FastQC (version 0.11.7; Andrews 2010) with a focus on base quality scores, GC content and sequence duplication levels. The NGS QC Toolkit (version 2.3.3; Patel & Jain 2012) was used to retain reads if at least 80% of bases had a quality score ≥ Q30 (a 1 in 1000 chance of an incorrect base call).

5.2.5 Initial sequence mapping

Filtered reads were mapped to the revised mitogenome reference (MH747083) using the BWA-MEM algorithm (version 0.7.17; Li & Durbin 2009) with default settings and sorted by leftmost position using SAMtools (version 1.9; Li et al. 2009). Read groups

were assigned to the mapped reads and PCR duplicates removed using Picard Tools (Broad Institute, Cambridge, US). Coverage was calculated using SAMtools. Mapping qualities were calculated and plotted using Qualimap (version v2.2.1; Okonechnikov et al. 2016).

5.2.6 Identifying linked SNPs across the mitogenome

Mapped reads were visualised in Tablet (version 1.19.05.28; Milne et al. 2016), which provides a graphical view of NGS alignments imported as a BAM file. A large number of linked SNPs were visible throughout the entire mitogenome in most, if not all, individuals (Fig. 5.2). Linked SNPs were hereafter explored in the duplicated region (*cox2*:*nad5* inclusive; 8.9 kb) and remaining region (11.4 kb) separately.

5.2.7 Investigating the cause of linked SNPs *within* the duplicated region

Accurately mapping reads to sequences containing duplications is complicated if reads are not long enough to span the duplication. For example, reads belonging to the functional sequence could incorrectly map to the duplicated, pseudo copy or vice versa. To determine if linked SNPs within the duplicated region were, at least partly, due to mis-mapping of reads, a pairwise alignment of the two parts of the duplicated region was manually created in AliView (version 1.26; Larsson 2014) and the reference genome positions of duplication-specific SNPs recorded (hereafter 'duplication SNPs') (Fig. 5.3). Duplication SNPs were inspected in Tablet for patterns of variant linkage. For two individuals with obvious signs of linkage among duplication SNPs (L240 and L311), linkage was further visualised by using the heteroplasmy option in NOVOPlasty (version 3.7; Dierckxsens et al. 2016; Dierckxsens et al. 2020), with the minor allele frequency (MAF) set to 1%. NOVOPlasty is able to infer linkage among some variants as part of the variant calling process and outputs a text file (links.txt) that lists all the detected links. To visualise these links, the output file was used as input for Circos (version 0.69-9; Krzywinski et al. 2009). Circos plots for each individual were created with a) the original links.txt file and b) the links.txt file with duplication SNPs removed. If links are removed following removal of the duplication SNPs, this suggests they are the result of mis-mapping.

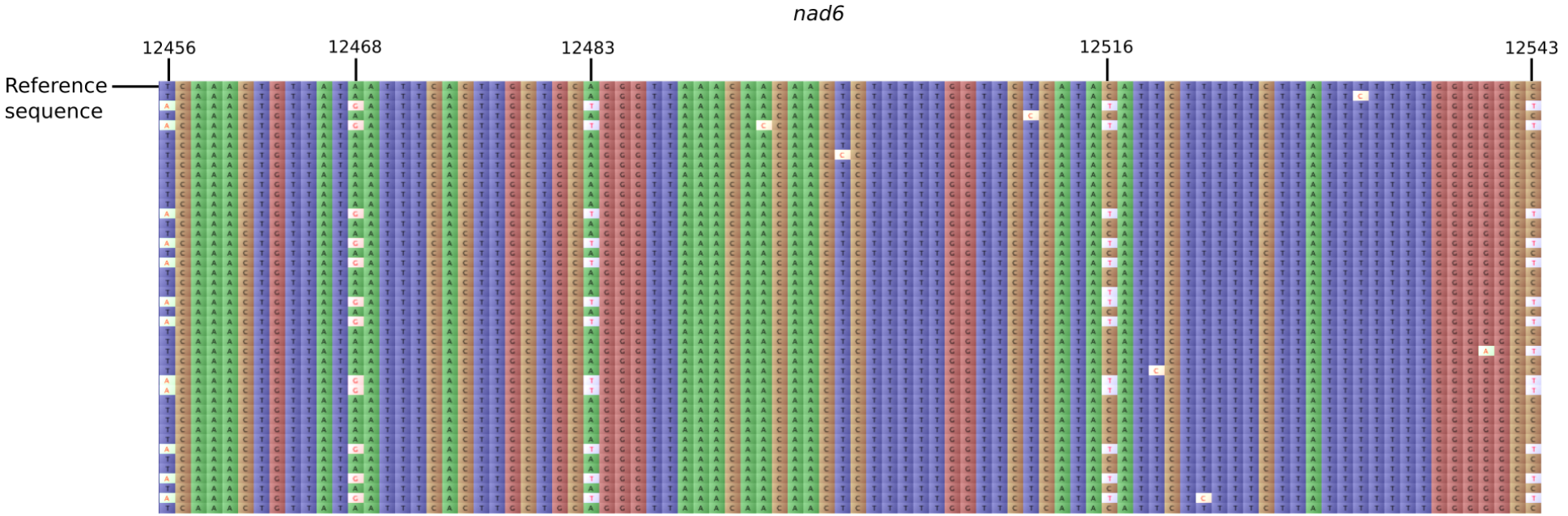


Figure 5.2 | Part of an alignment of Illumina sequencing reads (viewed in Tablet) from an individual European lobster (*Homarus gammarus*) mapped to the mitogenome reference. This region was chosen as an example to illustrate linked SNPs (numbered positions). These same five positions were later used to quantify the effects of different mapping strategies aimed at removing reads containing linked SNPs (see methods section 5.2.11).

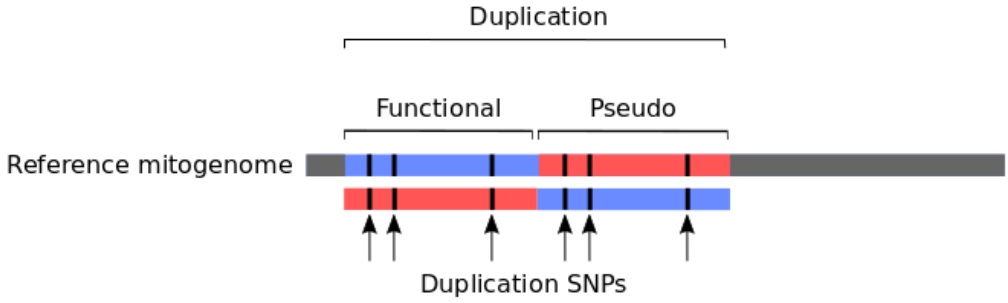


Figure 5.3 | Schematic to demonstrate pairwise alignment of the duplicated region of the European lobster (*Homarus gammarus*) mitogenome for identifying SNPs between the functional and pseudo sequences ('duplication SNPs').

5.2.8 Haplotype reconstruction of protein-coding genes to identify the cause of linked SNPs *outside* of the duplicated region

For the remaining mitogenome sequence outside of the duplicated region, linked SNPs were phased into haplotypes using HaploJuice (version 1.5.5; Wong et al. 2018). HaploJuice takes a BAM file as input and outputs a single FASTA file for each individual containing three haplotypes. Haplotypes were concatenated into a single, multi-individual FASTA file and aligned using MUSCLE (version 3.8.31; Edgar 2004). This file was subsequently split into separate alignments for each protein-coding gene (n = 7). Focusing on protein-coding genes can help to identify evidence of non-functionality, which is indicative of pseudo heteroplasmy. Partitioning the mitogenome also helped to retain more data for downstream processing because HaploJuice was unable to phase linked SNPs across the entire sequence length. HaploJuice represents regions that cannot be reliably phased with 'N'. This can occur if there is low sequencing coverage for that region or a large distance between variants. For each protein-coding gene, individuals with one or more haplotypes containing Ns were removed from the dataset.

Haplotypes were manually filtered, by editing the alignment in AliView, so that only variants called by FreeBayes (version 1.3.1; Garrison & Marth 2012) remained. FreeBayes is optimised for population-wide variant calling. The FreeBayes output was filtered to retain variants with a MAF > 1%, depth > 1,000x, mean mapping and allele

qualities > 30, and where fewer than 90% of minor allele supporting reads occur on either the forward or reverse strand. In addition, all variants were required to occur in more than one individual. Indels were removed because they all involved mononucleotide repeats ≥ 7 bp in length and were considered unreliable (possible sequencing error). HaploJuice outputs three sequences per individual, even if there is just one haplotype and all sequences are identical. After variant filtering, duplicate sequences within an individual were removed.

Haplotype networks were created in R using 'haploNet{pegas'}' (Paradis 2010) and annotated in Inkscape (version 0.92.2; www.inkscape.org). Because of the large number of unique haplotypes (see results section 5.3.4), haplotypes that occurred in only one individual were excluded from haplotype networks for visual clarity.

5.2.9 Haplotype diversity, identity and functionality

To assess nucleotide diversity among haplotypes, pairwise *p*-distances - the proportion of nucleotide sites at which two sequences differ - were calculated using MEGA (version 10.1.1; Stecher et al. 2020). To identify possible contamination from non-lobster DNA, haplotypes were identified using nucleotide BLAST queries against the GenBank nucleotide database. Amino acid alignments were visually-inspected in AliView and the codon position of each variant was recorded. The number of synonymous, non-synonymous, transition and transversion substitutions, and stop codons and indels were quantified according to the invertebrate mitochondrial code. A lack of third codon position and transition mutation bias, and the presence of non-synonymous mutations, stop codons and/or frameshift mutations, are characteristics of NUMT contamination (Triant & DeWoody 2007; Baldo et al. 2011).

5.2.10 Alternative mapping methods to try and remove linked SNPs

An overview of the following workflow (5.2.10 and 5.2.11) is provided in Fig. 5.5. For two individuals (L240 and L311) with obvious signs of linked SNPs throughout the mitogenome (using Tablet), and high mitogenome coverage, attempts were made to remove linked SNPs using three different strategies:

1) Repeating sequence mapping using alternative mappers to BWA-MEM

The alternative mappers used were Bowtie2 using both the '*--very-fast*' and '*--very-sensitive*' options (version 2.3.5.1; Langmead & Salzberg 2012), Yara using both the '*-y full*' or '*-y low*' options (version 0.9.11; Siragusa 2015) and Stampy with default settings (version 1.0.32; Lunter & Goodson 2011). These mappers use different algorithms compared to BWA-MEM and claim to be more accurate.

2) Filtering the mapped reads following BWA-MEM mapping

Mapped reads were filtered in two different ways: a) removing pairs where one of the reads mapped to the wrong location and/or in the wrong orientation (improperly-paired reads) using the '*view -f 2*' option of SAMtools and b) removing reads that map equally well to more than one location (multi-mapped reads) by discarding reads with the BWA-MEM flags XA:Z: and SA:Z using the grep command. For the duplicated region, an additional filtering step to remove reads with a mapping quality less than Q60 was performed using the '*view -q 60*' option of SAMtools. This decision was based on the apparently lower mapping quality across this region (see results Fig. 5.7).

3) Using BWA-MEM with different minimum seed lengths (*k*)

BWA-MEM mapping was repeated using 13 different minimum seed lengths (*k*) in increments of 10 from 29–139 by altering the '*-k*' option according to the desired seed length. BWA-MEM operates using a seed-and-extend approach. A long substring of nucleotides within a read, called a seed, must match the reference according to a specified length before mapping can take place. Increasing *k* effectively reduces the number of mismatches allowed before a read is discarded (Fig. 5.4). For all mapping strategies, reads were sorted by leftmost position using SAMtools, and read groups were assigned and PCR duplicates removed using Picard Tools.

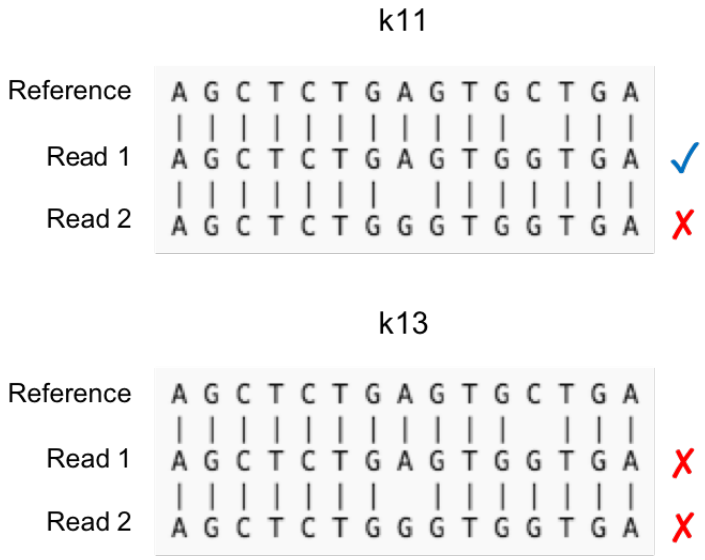


Figure 5.4 | Schematic to demonstrate that increasing the minimum mapping seed length (*k*), in this example from 11 to 13, reduces the number of sequencing reads mapped to the reference.

5.2.11 Quantifying the effects of alternative mapping methods on the frequency of linked SNPs and mitogenome coverage

The effects of the alternative mapping approaches on the presence of linked SNPs were quantified for the duplicated region and the remaining mitogenome separately. For both regions, the effects were quantified for a small subset of linked SNPs (Fig. 5.2 and Fig. 5.8). These sites were chosen because of their close proximity, which meant that the impact of different mapping methods on patterns of variant linkage could be easily visualised in Tablet. For the duplicated region, sites involving duplication SNPs were chosen because these SNPs were likely to be caused my mis-mapping (Fig. 5.8). The average alternative allele frequency (AAF) was calculated across the chosen linked SNPs for each mapping method by counting the number of reads supporting the reference or alternative allele for each linked SNP position using SAMtools mpileup. The average coverage was calculated across the region in question (duplicated or remaining mitogenome) using SAMtools. Removal of linked SNPs was deemed successful if the average AAF of linked SNPs dropped below 0.5% and the average coverage for the region in question remained > 1,000x.

5.2.12 Validating new mapping method

For any mapping strategy that successfully removed linked SNPs, the approach was extended to the full list of individuals and PCR duplicates were removed using Picard Tools. To validate if the revised mapping method removed linked SNPs, the new BAM files were visualised in Tablet for a random 5% subset of individuals (*n* = 7; L65, L195, L198, L237, L259, L281, L306), plus the original two individuals used for quantifying the effects of the different mapping approaches (L240 and L311), and the entire mitogenome was investigated for evidence of linked SNPs.

5.2.13 Point heteroplasmy detection *outside* of duplicated region and variant filtering

Prior to variant calling, mapping was performed using BWA-MEM with *k* = 119, which successfully reduced linked SNPs outside of the duplicated region (see results section 5.3.6b). Reads were mapped to an extended version of the reference mitogenome by copying the first and last 500 bp to opposite ends of the sequence. This accounts for circularity of the genome and avoids an artificial drop in the number of reads due to the arbitrary start and end point. The duplicated region was excluded from variant calling because linked SNPs within this region could not be removed without compromising coverage (see results section 5.3.6a). Individuals with an average coverage < 1,000x for the remaining mitogenome sequence were omitted from the dataset.

Variant calling was performed using VarScan (version 2.4.3; Koboldt et al. 2012). VarScan consistently performs well at calling low frequency variants with high accuracy and sensitivity, and a low false positive rate (Jebb et al. 2018; see section 4.3.5 in chapter 4). Variants were filtered to retain point mutations with a MAF > 1%, depth > 1,000x, average minor allele base quality > 30, and where fewer than 90% of minor allele supporting reads occurred on either the forward or reverse strand.

Linked SNPs outside of the duplicated region were reduced but not completely eliminated (see results section 5.3.6b). Therefore, point heteroplasmies found in more

than one individual (shared) were removed so that only 'private' heteroplasmies remained. Shared heteroplasmies are more likely to have a common origin such as inherited heteroplasmy or NUMT contamination (Yuan et al. 2020). Private SNPs within 150 bp of a neighbouring SNP with similar frequency (within 1%), and therefore possibly linked, were also excluded.

Point heteroplasmies in *cox1* were also removed from the dataset; *cox1* was the only protein-coding gene outside of the duplicated region to contain heteroplasmies with evidence of loss-of-function. Three private stop codons were observed in *cox1*, including one at high frequency (50%). In addition, NOVOPlasty identified *cox1* SNPs that were in linkage with duplication SNPs (Fig. 5.9). These observations point towards a possible, and previously unreported, duplication involving *cox1*.

5.2.14 Point heteroplasmy characterisation and distribution

Filtered heteroplasmies were classified as occurring in protein-coding, non-protein coding or non-coding regions. Protein-coding mutations were further partitioned into synonymous or non-synonymous amino acid changes. Non-synonymous mutations were classified based on their predicted impact (neutral or deleterious) according to the PROVEAN algorithm using the default threshold of -2.5 (Choi et al. 2012; Choi & Chan 2015). The frequency distribution of mutations was compared among the different mitochondrial partitions using Kolmogorov-Smirnov (KS) or Mann-Whitney U tests in R with the functions 'ks.test{stats}' and 'wilcox.test{stats}', respectively (R Core Team 2017a).

To determine whether different regions of the mitogenome were enriched for heteroplasmy, mutations were binned into tRNAs, rRNAs, protein-coding genes, non-protein coding genes and non-coding regions. The number of point heteroplasmies in each bin was compared to the remaining mitogenome sequence with Pearson's Chi-Square tests using 'chisq.test{stats}'.

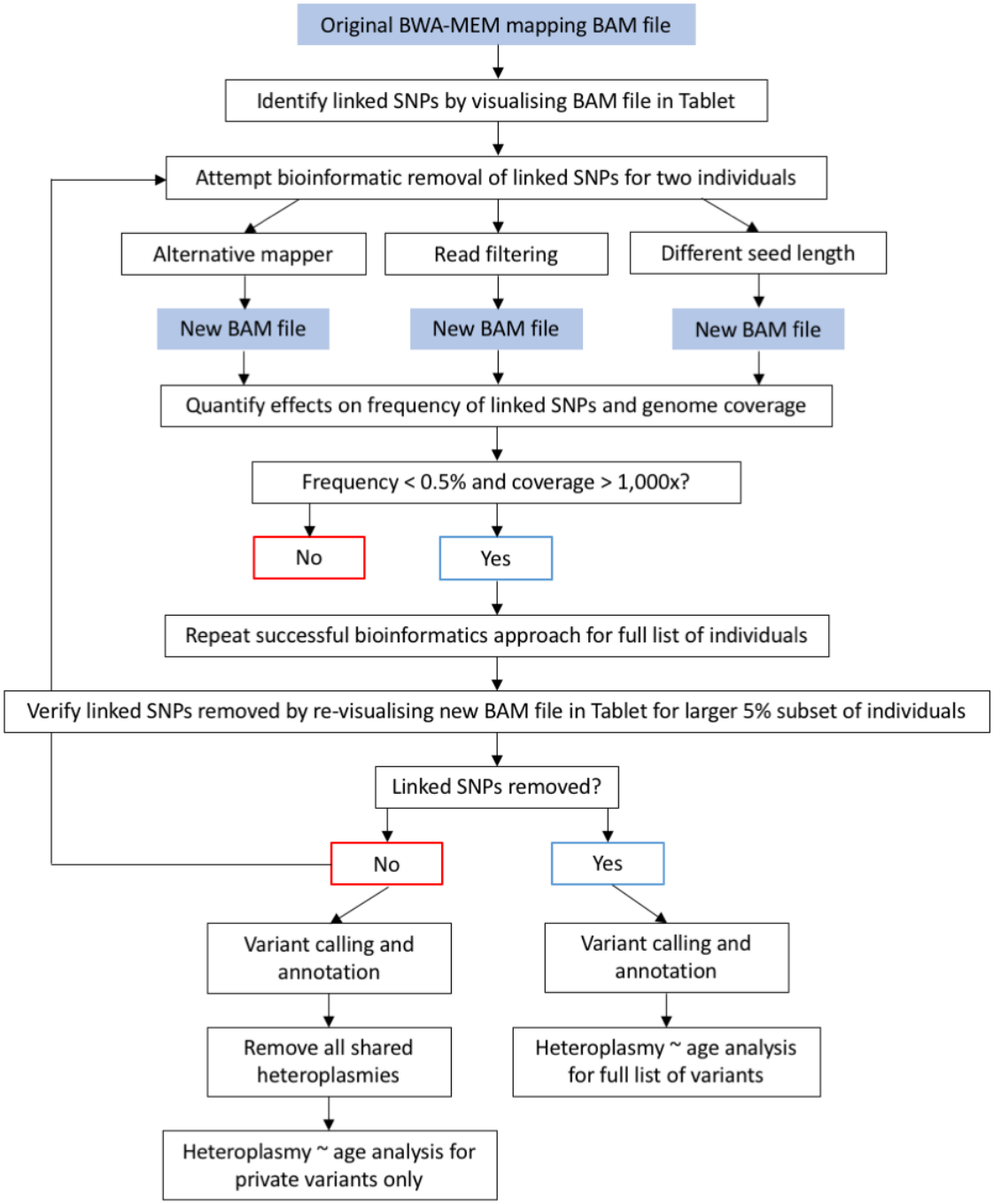


Figure 5.5 | The workflow undertaken to quantify and visualise the effects of different mapping strategies on the presence of linked SNPs in mitogenome NGS data. Linked SNPs may indicate the presence of pseudo heteroplasmy and need to be minimised prior to any analysis on the relationship between genuine heteroplasmy and age. If linked SNPs cannot be removed entirely, an analysis based on heteroplasmies found in one individual only (private heteroplasmies) should reduce false positives arising from pseudo heteroplasmy.

5.2.15 Point heteroplasmy and age

The number of mtDNA point mutations was compared among lobster age groups using generalised linear models (GLMs). Maximal GLMs contained significant overdispersion ($p < 0.001$) according to 'dispersiontest{AER}' (Kleiber & Zeileis 2008), making a Poisson distribution unsuitable. Therefore, GLMs were fitted with a negative binomial distribution and a log link function using 'glm.nb{MASS}' (Venables & Ripley 2002). Influential points were assessed according to Cook's distance (Cook & Weisberg 1982) by examining diagnostic plots using 'plot{graphics}' (R Core Team 2017b). Analyses were performed with/without the wild lobster age class to control for tissue differences. For both GLMS, the main effect of age was assessed using Chi-Square tests in 'Anova{car}' (Fox & Weisberg 2019).

There was large size variation among the wild lobster cohort and these individuals may differ markedly by age (Sheehy et al. 1999). To assess whether heteroplasmy varies with size (and therefore age) in wild lobsters, the cohort was divided into small (89–100 mm CL) and large (102–136 mm CL) lobsters and the number of heteroplasmies compared between size classes using 'wilcox.test{stats}'.

5.2.16 Statistics

Statistical analyses were performed in R version (3.6.3; R Core Team 2020) using RStudio (RStudio Team 2016). Plots were produced using 'ggplot{ggplot2}' (Wickham 2016).

5.3 Results

5.3.1 Sequence quality control

Four individuals failed to sequence and were removed from the dataset. Sequencing was successful for 152 lobsters with a total of 208,929,622 read pairs after quality filtering. An average 87.2% of raw reads were retained post-filtering, all of which contained bases with an average quality ≥ Q30.

5.3.2 Initial sequence mapping

Mapping using default BWA-MEM settings was successful for all lobsters with an average coverage of 2,410x (range = 137–8,155x) after removal of PCR duplicates (Fig. 5.6). Mapping quality was noticeably lower across the duplicated region compared to the remaining mitogenome (Fig. 5.7).

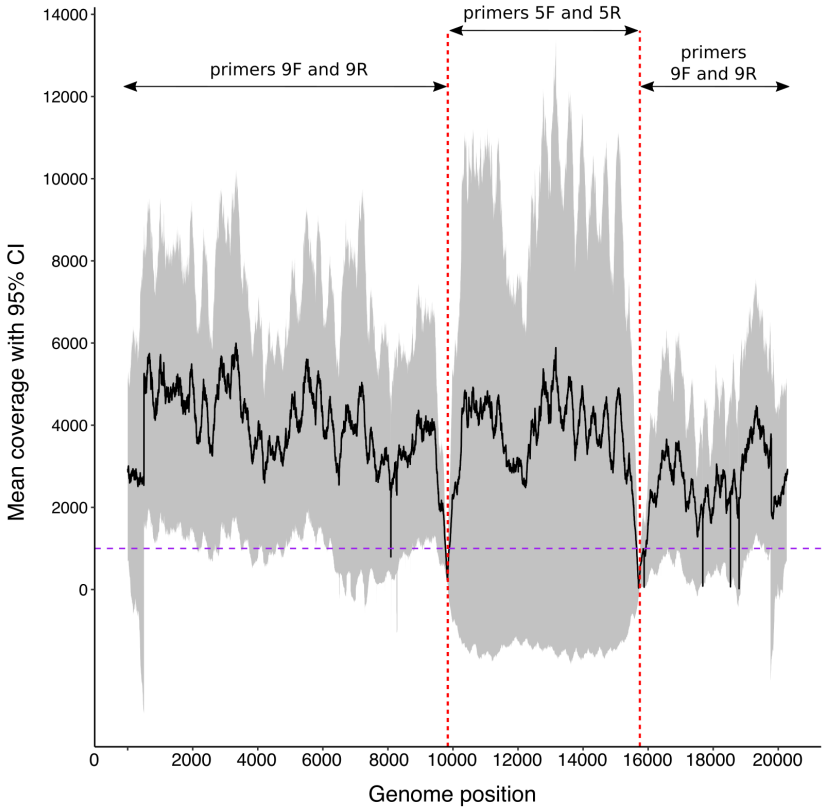


Figure 5.6 | European lobster (*Homarus gammarus*) mitogenome coverage from Illumina NovaSeq data of two amplicons. Reads were mapped using default BWA-MEM settings and PCR duplicates removed. Black line = mean coverage across individuals (*n* = 152) and grey shaded area = 95% confidence intervals (CI). Vertical red lines are primer binding sites. Horizontal purple line = 1,000x coverage.

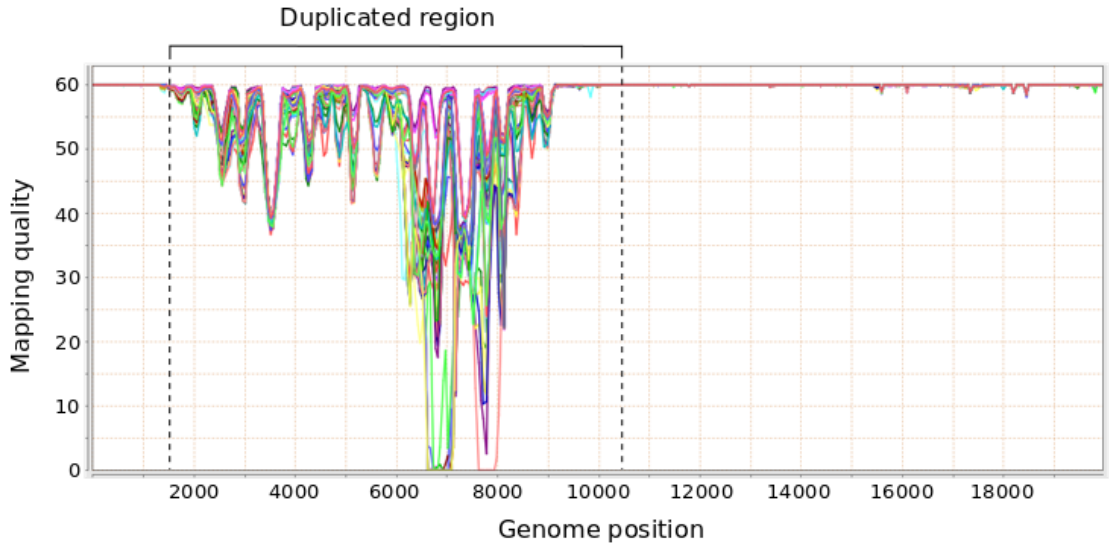


Figure 5.7 | Mean mapping quality (phred-scaled) of Illumina NovaSeq reads aligned to the European lobster (*Homarus gammarus*) mitogenome using BWA-MEM. Each coloured line represents a single lobster (*n* = 152). Vertical dotted lines delimit approximate location of the duplicated region. Mapping qualities were calculated and plotted using Qualimap.

5.3.3 Mis-mapping of reads *within* the duplicated region

Pairwise alignment of the duplicated region identified 88 SNPs (duplication SNPs) between the functional sequence and its pseudo counterpart (Supplementary Table S5.1), equivalent to 97.7% sequence similarity. Indel differences were not considered as they all occurred in mononucleotide repeats ≥ 5 bp in length. Linkage between some of the duplication SNPs was apparent in Tablet (Fig. 5.8). NOVOPlasty correctly identified linked SNPs within the duplicated region, which were visualised in Circos (Fig. 5.9a). A large number of links were removed following removal of the duplication SNPs, confirming that at least some of the linked SNPs were caused by mis-mapping (Fig. 5.9b).

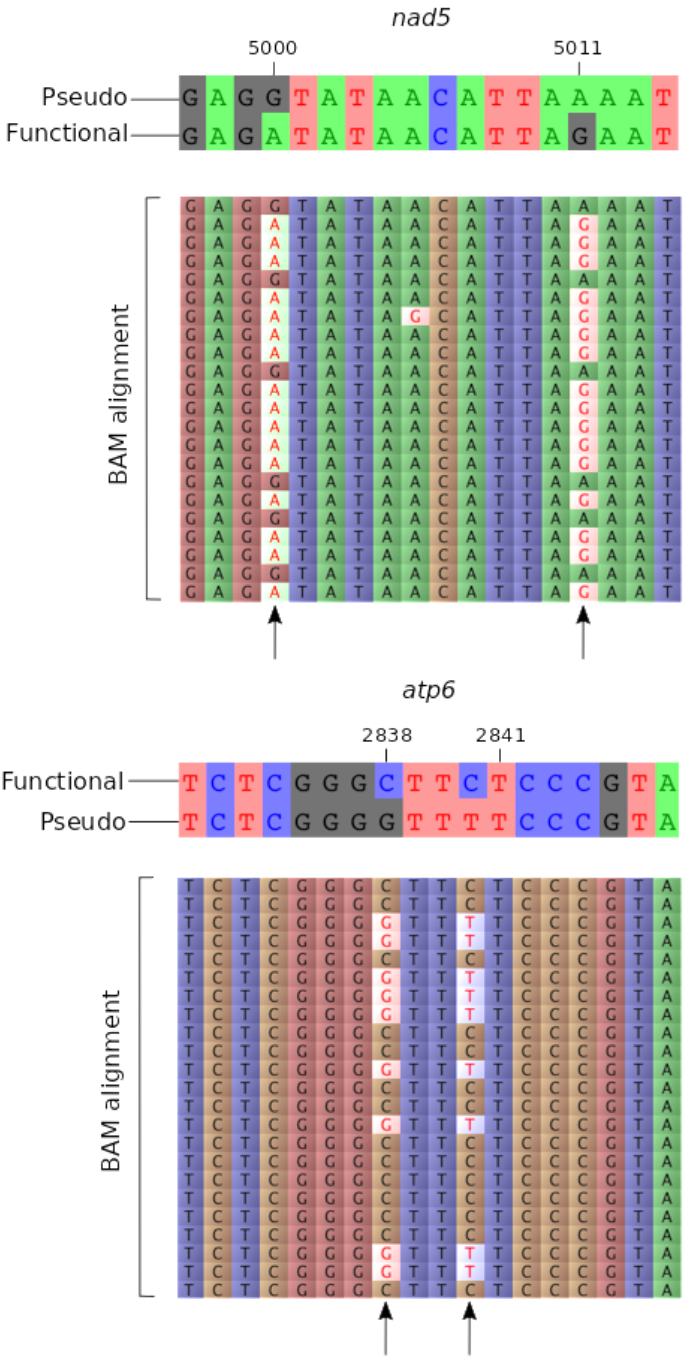
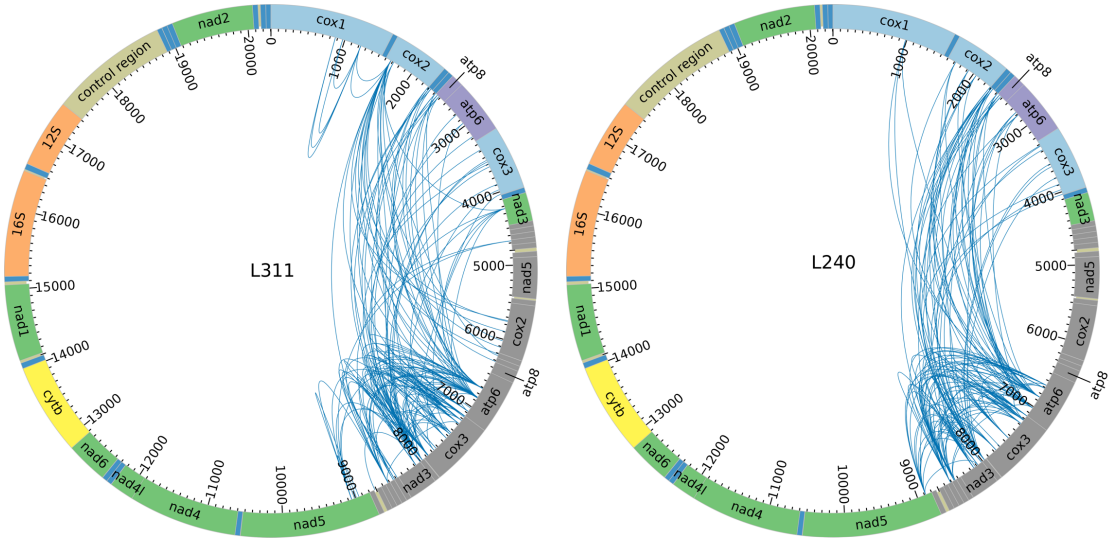


Figure 5.8 | An alignment of the European lobster (*Homarus gammarus*) mitogenome reference sequence viewed in AliView (above) and the BAM file of an individual European lobster alignment viewed in Tablet (below) for two duplicated genes (*nad5* and *atp6*). Each row in the Tablet output represents part of a 150 bp Illumina sequencing read. Arrows indicate linked SNPs caused by mis-mapping between the functional and pseudo sequences of each duplicated gene. These same four positions were used to quantify the effects of different mapping strategies aimed at removing reads containing linked SNPs in the duplicated region (see methods section 5.2.11).

a) Unfiltered



b) Duplication SNPs removed

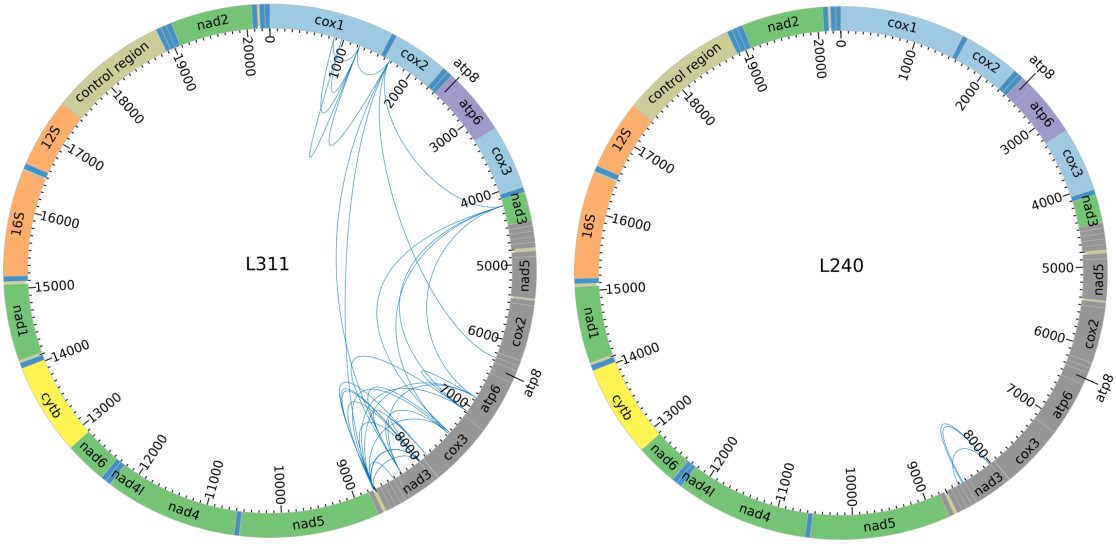


Figure 5.9 | Circos plots of the mitogenome of two European lobsters (*Homarus gammarus*). Blue lines represent linked SNPs detected by the heteroplasmy caller NOVOPlasty. Circos plots are shown a) before filtering of any linked SNPs and b) after removing linked SNPs that are known to differ between the functional and pseudo sequence (grey region) of the mitochondrial duplication. Darker blue or unlabelled grey sections are transfer RNAs.

5.3.4 Haplotype diversity *outside* of the duplicated region

For the seven protein-coding genes outside of the duplicated region, 23–88 unique haplotypes were reconstructed, with average and maximum pairwise *p*-distances of 1.41% and 8.43%, respectively (Table 5.3). Between 45–60% of lobsters contained one haplotype, up to 42% contained two haplotypes and as many as 20% contained three or more (Table 5.4). The most common haplotypes were either the reference sequence or another closely-related haplotype (≤ 3 nucleotides divergent from the reference) (Fig. 5.10), regardless of whether the lobster contained one or multiple haplotypes (Table 5.5).

Table 5.3 | Haplotype variation within seven protein-coding genes of the mitochondrial genome (excluding the duplicated region) of the European lobster (*Homarus gammarus*). Haplotypes reconstructed from Illumina NovaSeq data using HaploJuice.

Gene	Length (bp)	*n* haplotypes	*n* individuals	*n* variable sites	Mean pairwise *p*-distance (%)	Max pairwise *p*-distance (%)
cox1	1537	38	123	53	0.23	0.72
nad4	1341	69	103	126	1.99	8.43
nad4l	303	23	123	23	2.30	6.93
nad6	519	49	133	41	2.03	6.36
cytb	1135	88	139	81	1.72	5.99
nad1	939	75	110	60	1.25	5.01
nad2	1002	58	133	27	0.35	0.80

Table 5.4 | Number and proportion of European lobsters (*Homarus gammarus*) with 1, 2 or ≥ 3 haplotypes for seven mitochondrial protein-coding genes.

Gene	*n* (%) 1 haplotype	*n* (%) 2 haplotypes	*n* (%) ≥ 3 haplotypes
cox1	73 (59.3)	37 (30.1)	13 (10.6)
nad4	55 (53.4)	32 (31.1)	16 (15.5)
nad4l	74 (60.2)	40 (32.5)	9 (7.3)
nad6	76 (57.1)	46 (34.6)	11 (8.3)
cytb	68 (48.9)	59 (42.4)	12 (8.6)
nad1	49 (44.5)	41 (37.3)	20 (18.2)
nad2	64 (48.1)	43 (32.3)	26 (19.5)

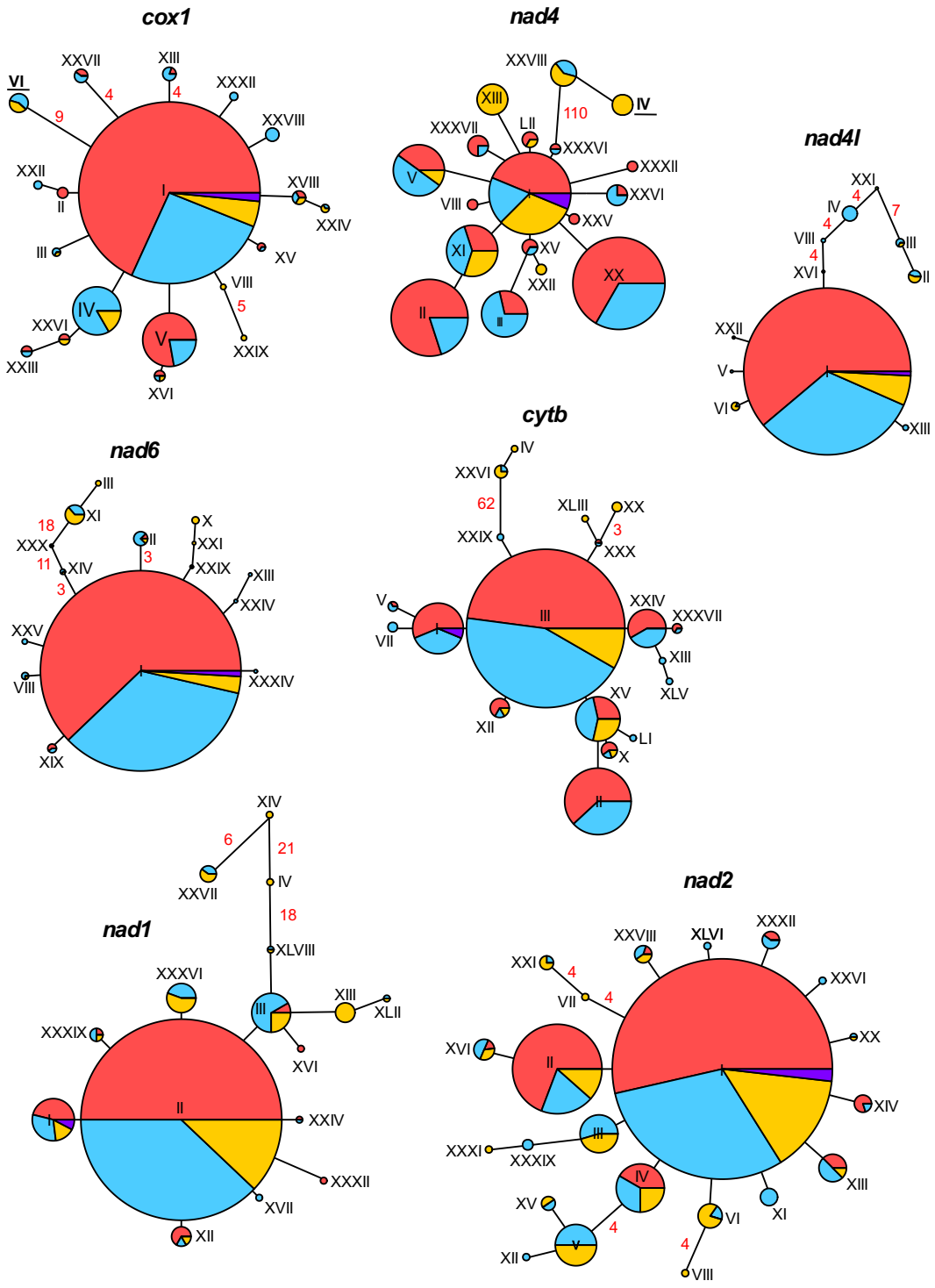


Figure 5.10 | Haplotype networks for mitochondrial protein-coding genes in 103–139 individual European lobsters (*Homarus gammarus*). Circles shaded by number of haplotypes within an individual: 1 = red, 2 = blue, ≥ 3 = orange. Purple = reference mitogenome. Haplotype IDs in Roman numerals. Numbers in red are the number of nucleotide differences between haplotypes if > 2. All indels considered a single mutational step regardless of size. Size of circles represents relative haplotype frequency within and not between genes. Haplotype IDs containing stop codons or frameshift mutations in bold and underlined. Only haplotypes found in > 1 individual are shown.

Table 5.5 | Number and proportion (in parentheses) of European lobsters (*Homarus gammarus*) with 1, 2 or ≥ 3 haplotypes that contained either the reference haplotype, or another closely-related haplotype (≤ 3 nucleotides divergent from the reference), for seven mitochondrial protein-coding genes.

Gene	1 haplotype	2 haplotypes	≥ 3 haplotypes
cox1	68 (93.2)	35 (94.6)	13 (100.0)
nad4	55 (100.0)	31 (96.9)	16 (100.0)
nad4l	74 (100.0)	40 (100.0)	9 (100.0)
nad6	76 (100.0)	46 (100.0)	11 (100.0)
cytb	68 (100.0)	59 (100.0)	11 (91.7)
nad1	48 (98.0)	41 (100.0)	20 (100.0)
nad2	64 (100.0)	43 (100.0)	24 (92.3)

5.3.5 Haplotype identity and functionality

All haplotypes were most closely-related to mtDNA from either *H. gammarus* (*n* = 394) or the sister species *H. americanus* (*n* = 6) with BLAST identity ranging from 92.8–100% (Supplementary Table S5.2). There was evidence of loss-of-function in some of the haplotypes belonging to three protein-coding genes; frameshift mutations in *cox1* and *nad1*, and a stop codon in *nad4* (Fig. 5.11). Most mutations occurred at the third codon position (mean = 78.8%), followed by the first (mean = 15.7%) and second (mean = 5.5%) (Table 5.6).

T**able 5.6 |** The percentage of variables sites at each codon position and the percentage of non-synonymous (NS) or transversion (Tv) mutations across all mitochondrial haplotypes in seven protein-coding genes belonging to 103–139 European lobsters (*Homarus gammarus*).

Gene	1st codon (%)	2nd codon (%)	3rd codon (%)	NS (%)	Tv (%)
cox1	17.1	5.7	77.1	17.1	8.6
nad4	11.3	5.6	83.1	15.3	13.1
nad4l	8.7	8.7	82.6	8.7	13.0
nad6	17.1	7.3	75.6	19.5	17.1
cytb	7.4	7.4	85.2	11.1	14.6
nad1	15.3	0.0	84.7	5.1	8.5
nad2	33.3	3.7	63.0	29.6	17.9

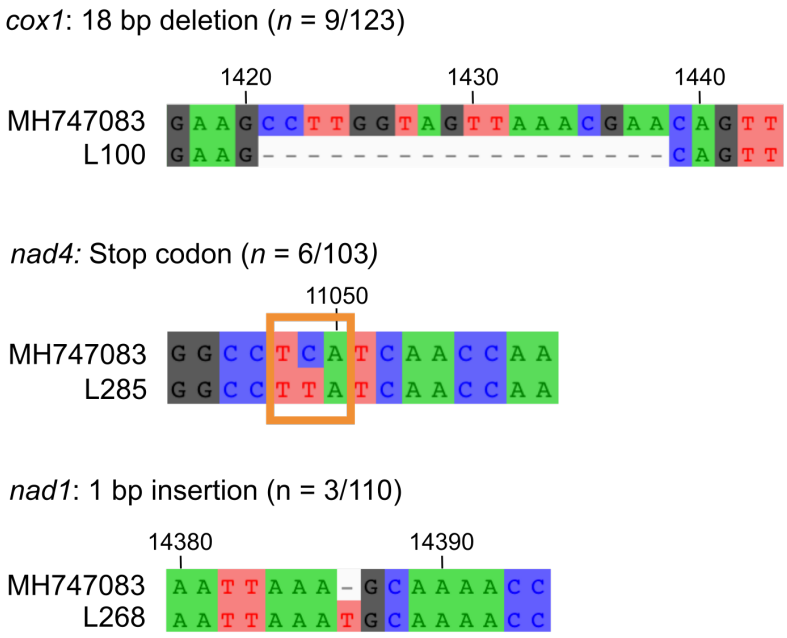


Figure 5.11 | Evidence of loss-of-function in mitochondrial haplotypes of three protein-coding genes assembled from Illumina NovaSeq data of European lobsters (*Homarus gammarus*). For each gene, the sequence for the reference mitogenome is shown above with a haplotype from an individual lobster sequenced in this study below. Sample sizes are the proportion of lobsters found to contain the mutation.

5.3.6 Effects of alternative mapping methods on the frequency of linked SNPs and mitogenome coverage

a) *Duplicated region*

Two individuals (L240 and L311) were chosen to investigate the effects of alternative mapping methods on the presence of four linked SNPs (Fig. 5.8), and average mitogenome coverage, across the *duplicated region*. These linked SNPs appeared to be caused by mis-mapping between functional and pseudo sequences (Fig. 5.8). For both individuals, linked SNPs were most effectively reduced by removing multi-mapped reads following BWA-MEM mapping (Table 5.7). This filtering step removed all four linked SNPs for one individual and reduced the average AAF of the same variants to < 0.5% for the other. However, removing multi-mapped reads came at the cost of reducing average coverage by 64–74% compared to the default BWA-MEM mapping. The average coverage fell below the 1,000x threshold for one of the two individuals (1677x for L240 and 828x for L311). No other BWA-MEM filtering step, minimum BWA-MEM seed length, or alternative mapper successfully eliminated mis-mapping for both individuals (average AAF > 1%) (Table 5.7; Fig. 5.12).

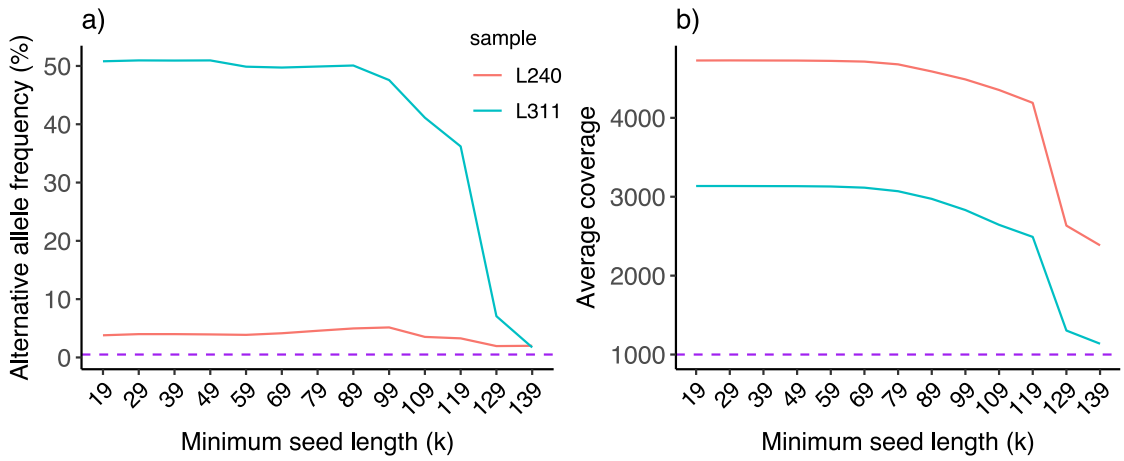


Figure 5.12 | Illumina NovaSeq reads belonging to two European lobster (*Homarus gammarus*) individuals, L240 and L311, were mapped to the reference mitogenome using the BWA-MEM algorithm with 13 different minimum seed lengths (*k*). The effects of *k* on a) the average alterative allele frequency (AAF) of four linked SNPs and b) the average coverage across the entire duplicated region are shown. Horizonal purple lines display a) an AAF of 0.5% and b) an average coverage of 1,000x.

Table 5.7 | The effects of different mapping methods on the read counts of reference (REF) and alternative alleles (ALT) at four positions (POS) in the duplicated region of the mitogenome belonging to two European lobsters (*Homarus gammarus*; L240 and L311). Alternative alleles at positions 2838 and 2843, and 5000 and 5011, were linked, and appeared to be caused by mis-mapping of functional and pseudo sequences (Fig. 5.8). The frequency of the alternative allele (AAF) is averaged across all four positions. Mean coverage is across the entire duplicated region. For each lobster, rows are sorted in order of descending mean AAF.

Mapping method	POS 2838		POS 2843		POS 5000		POS 5011		mean AAF (%)	mean coverage
	REF (C)	ALT (G)	REF (C)	ALT (T)	REF (G)	ALT (A)	REF (A)	ALT (G)		
Lobster L240										
Bowtie2 --*very-fast*	3773	126	3801	124	3496	158	3423	224	4.20	4814x
Bowtie2 --*very-sensitive*	3776	121	3804	120	3495	159	3530	224	4.15	4815x
BWA-MEM	5899	194	5928	194	4810	213	4763	283	4.05	4723x
BWA-MEM with improperly-paired reads removed	5899	194	5928	194	4810	213	4763	283	4.05	4723x
Yara -*y full* or -*y low*	6286	193	6325	164	4950	163	4915	279	3.53	4808x
Stampy	5805	123	5831	124	4518	141	4485	229	3.03	4862x
BWA-MEM with minimum mapping quality of Q60	5790	1	5814	4	4738	60	4694	151	1.13	3609x
BWA-MEM with multi-mapped reads removed	208	0	211	1	1981	7	2448	1	0.23	1677x

Table 5.7 (continued)

Mapping method	POS 2838		POS 2843		POS 5000		POS 5011		mean AAF (%)	mean coverage
	REF (C)	ALT (G)	REF (C)	ALT (T)	REF (G)	ALT (A)	REF (A)	ALT (G)		
Lobster L311										
Bowtie2 --*very-sensitive*	2093	707	2136	691	475	1858	453	1969	52.65	3136x
Bowtie2 --*very-fast*	2096	705	2139	690	477	1855	453	1963	52.60	3221x
Yara -*y full* or -*y low*	3130	1114	3183	1041	673	2388	642	2579	52.23	3171x
BWA-MEM	3037	867	3090	833	648	2451	625	2612	50.80	3136x
BWA-MEM with improperly-paired reads removed	3037	867	3090	833	648	2451	625	2612	50.80	3136x
Stampy	2929	640	2976	611	618	2226	597	2411	48.35	3171x
BWA-MEM with minimum mapping quality of Q60	2984	2	3033	0	637	428	625	706	23.30	2031x
BWA-MEM with multi-mapped reads removed	129	0	130	0	278	0	332	0	0	828x

b) Mitogenome *excluding* duplicated region

The same two individuals (L240 and L311) were chosen to investigate the effects of alternative mapping methods on the presence of five linked SNPs (Fig. 5.2), and average mitogenome coverage, for the region *outside* of the duplication. For both individuals, the five linked SNPs were not successfully removed using three alternative mappers or filtering the reads following BWA-MEM mapping (average AAF ≥ 1.76%; Table 5.8). However, increasing *k* to 119 reduced the average AAF of the five linked SNPs to below 0.5% (Fig. 5.13a) while retaining an average coverage > 1,000x (Fig. 5.13b). Average coverage dropped below 1,000x for one of the individuals (L311) at *k* > 119 (Fig. 5.13b).

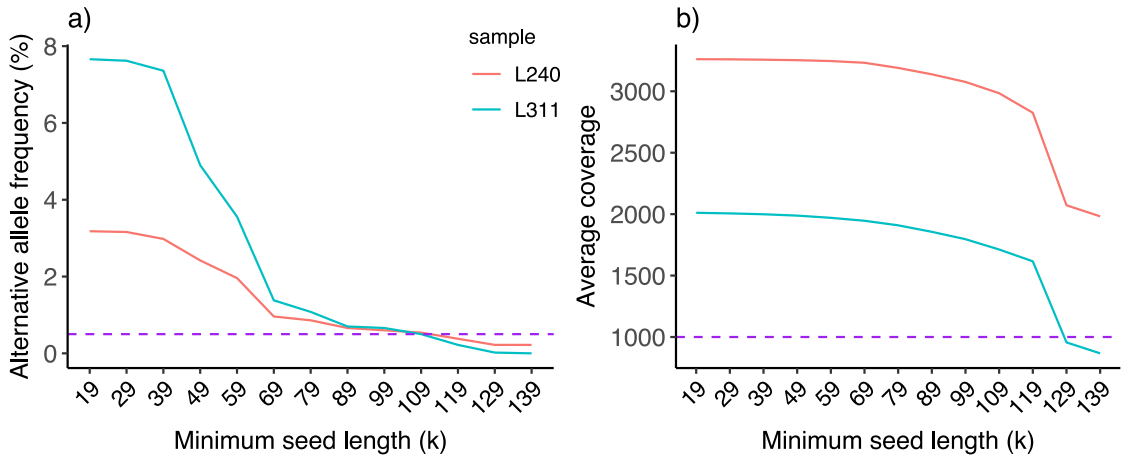


Figure 5.13 | Illumina NovaSeq reads belonging to two European lobsters (*Homarus gammarus*), L240 and L311, were mapped to the reference mitogenome using the BWA-MEM algorithm with 13 different minimum seed lengths (*k*). The effects of *k* on a) the average alterative allele frequency (AAF) of five linked SNPs and b) the average coverage outside of the duplicated region are shown. Horizonal purple lines display a) an AAF of 0.5% and b) an average coverage of 1,000x.

Table 5.8 | The effects of different mapping methods on the read counts of reference (REF) and alternative alleles (ALT) at five positions (POS) outside of the duplicated region of the mitogenome belonging to two European lobsters (*Homarus gammarus*): L240 and L311. The five alternative alleles were linked (Fig. 5.2). The frequency of the alternative allele (AAF) is averaged across all five positions. Mean coverage is across the mitogenome excluding the duplicated region. For each lobster, rows are sorted in order of descending mean AAF.

Mapping method	12456		12468		12483		12516		12543		mean AAF (%)	mean coverage
	REF (T)	ALT (A)	REF (A)	ALT (G)	REF (A)	ALT (T)	REF (C)	ALT (T)	REF (C)	ALT (T)		
Lobster L240												
Bowtie2 *--very-sensitive*	1770	61	1801	58	1781	53	1687	59	1818	64	3.22	3143x
bwa-mem	2323	80	2364	77	2360	71	2283	76	2475	86	3.20	3261x
bwa-mem with improperly-paired reads removed	2323	80	2364	77	2360	71	2283	76	2475	86	3.20	3089x
bwa-mem with multi-mapped reads removed	2322	80	2363	77	2359	71	2282	76	2474	86	3.20	3220x
Stampy	2275	77	2315	74	2308	69	2231	75	2421	82	3.18	3173x
Bowtie2 *--very-fast*	1773	50	1803	48	1784	43	1688	52	1820	57	2.74	3132x
Yara *-y full* or *-y low*	2377	28	2375	73	2372	22	2291	36	2484	41	1.76	3083x
Lobster L311												
Stampy	976	78	1005	84	1009	76	950	82	1054	100	7.74	1943x
bwa-mem	1009	81	1045	87	1045	78	991	85	1098	101	7.68	2011x
bwa-mem with improperly-paired reads removed	1009	81	1045	87	1045	78	991	85	1098	101	7.68	1864x

Table 5.8 (continued)

Mapping method	12456		12468		12483		12516		12543		mean AAF (%)	mean coverage
	REF (T)	ALT (A)	REF (A)	ALT (G)	REF (A)	ALT (T)	REF (C)	ALT (T)	REF (C)	ALT (T)		
Lobster L311												
bwa-mem with multi-mapped reads removed	1009	81	1045	87	1045	78	991	85	1098	101	7.68	1982x
Bowtie2 --*very-sensitive*	810	62	848	67	838	60	776	64	854	78	7.42	1920x
Bowtie2 --*very-fast*	812	48	849	53	840	47	778	52	69	856	6.12	1905x
Yara -*y full* or -*y low*	1018	20	1055	42	1053	18	1006	25	1120	40	2.64	1867x

5.3.7 Validating new mapping method

The reads for all individuals were subsequently re-mapped to the reference mitogenome using BWA-MEM with *k* = 119. Re-mapping was successful for 152 lobsters with an average coverage of 3,184x (range = 76–10,086x). However, when re-examining the new alignments for a random 5% subset of individuals (plus L240 and L311), linked SNPs outside the duplicated region were still visible (Fig. 5.14). Therefore, although increasing *k* to 119 *reduced* linked SNPs outside of the duplicated region (Fig. 5.13a) it did not appear to remove all of them for all individuals (Fig. 5.14).

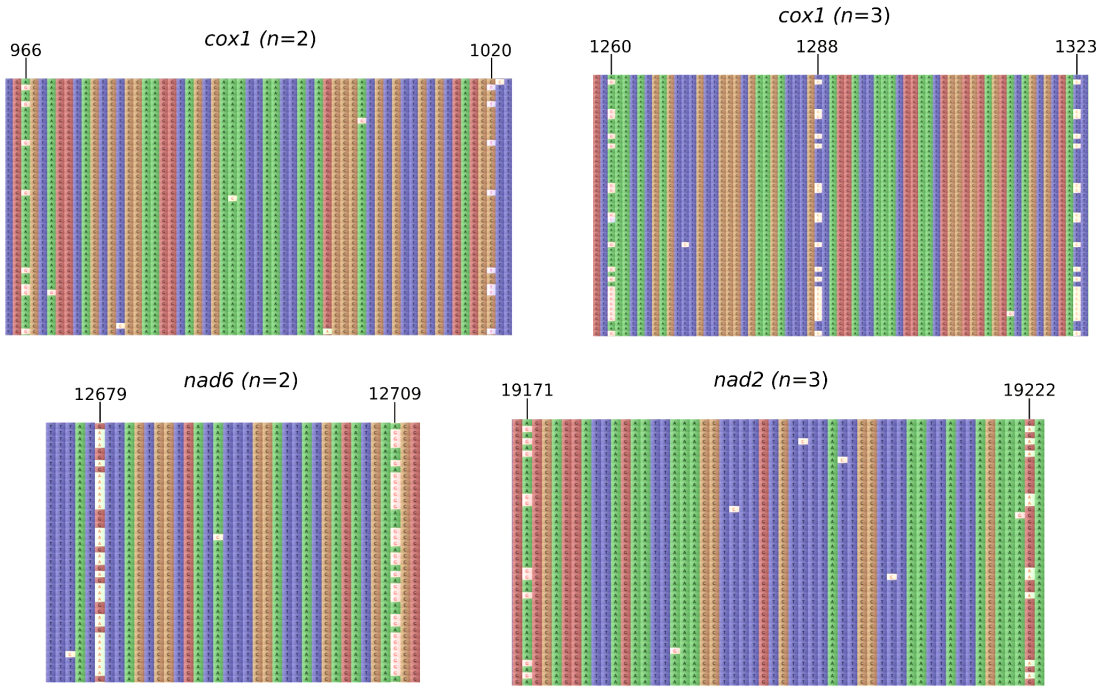


Figure 5.14 | Illumina NovaSeq reads of the European lobster (*Homarus gammarus*) mitogenome were re-mapped to the reference using a minimum seed length of 119. The BAM file alignments of seven lobsters were subsequently visualised in Tablet. Above are screenshots of some of the linked SNPs (numbered positions) observed. Samples sizes are the number of individuals found to contain the same linked SNPs.

5.3.8 Point heteroplasmy characterisation and distribution

After excluding the duplicated region and *cox1*, removing SNPs found in more than one individual, and excluding lobsters with low sequencing coverage (< 1,000x; $n = 9$), 118 private point heteroplasmies were found in 143 individuals across 9.8 kb of the mitogenome (Supplementary Table S5.3). The majority of SNPs were transition substitutions (94%; $n = 111$). Among the seven transversions, four were A:C substitutions and the remaining three were either A:T, C:G or G:T mutations. Seventy mutations occurred in protein-coding genes, 47.1% ($n = 33$) of which resulted in non-synonymous amino acid changes. The majority of non-synonymous mutations (57.6%; $n = 19$) were predicted to be deleterious (PROVEAN score ≤ -3.1).

Most heteroplasmies were at low frequency with 83% of mutations ($n = 98$) occurring at a frequency of ≤ 5% (Fig. 5.15). There was no significant difference in the frequency distribution of mutations in protein-coding or non-protein coding versus non-coding regions (Mann-Whitney $U \geq 249$; $p \geq 0.584$) or between protein-coding and non-protein coding genes (Mann-Whitney $U = 1057$; $p = 0.750$). Within protein-coding genes, synonymous mutations were at lower frequency on average than non-synonymous mutations (KS test $D = 0.346$; $p = 0.022$; Fig. 5.16). Among non-synonymous mutations, there was no significant difference in the frequency distribution of neutral versus deleterious SNPs (KS test $D = 0.293$; $p = 0.403$).

Point heteroplasmies were evenly-distributed across the mitogenome. No mitochondrial partition (tRNAs, rRNAs, protein-coding, non-protein coding and non-coding regions) showed significant enrichment for heteroplasmy compared to the remaining mitogenome sequence ($\chi^2_{(1,118)} \geq 0.001$; $p \geq 0.205$).

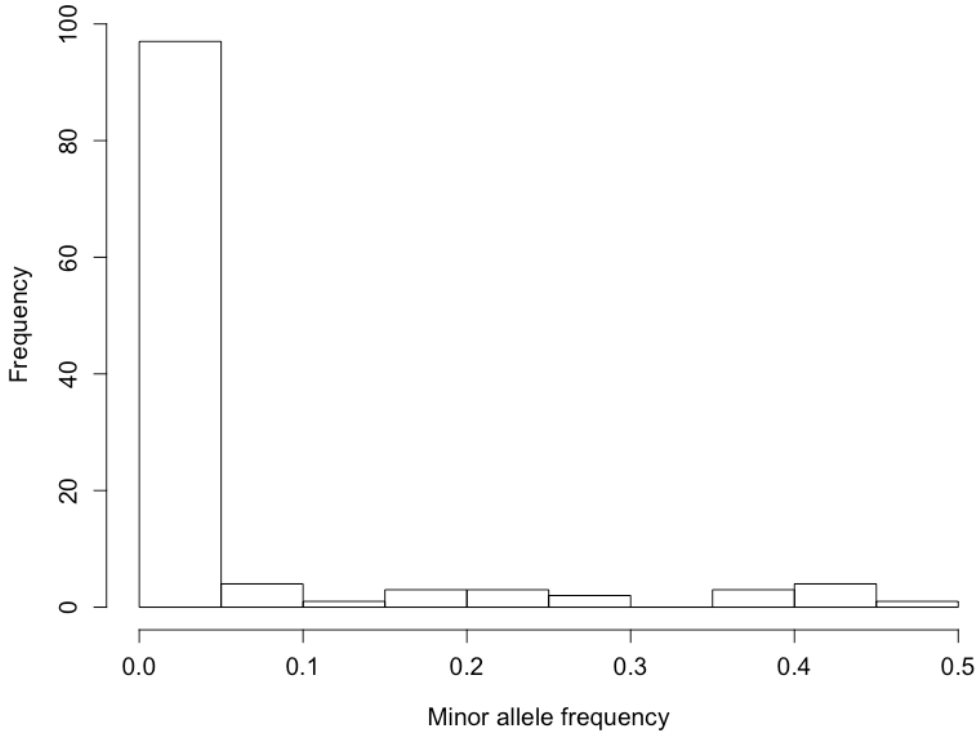


Figure 5.15 | The frequency distribution of the minor allele frequencies of 118 private point heteroplasmies identified in 143 European lobsters (*Homarus gammarus*).

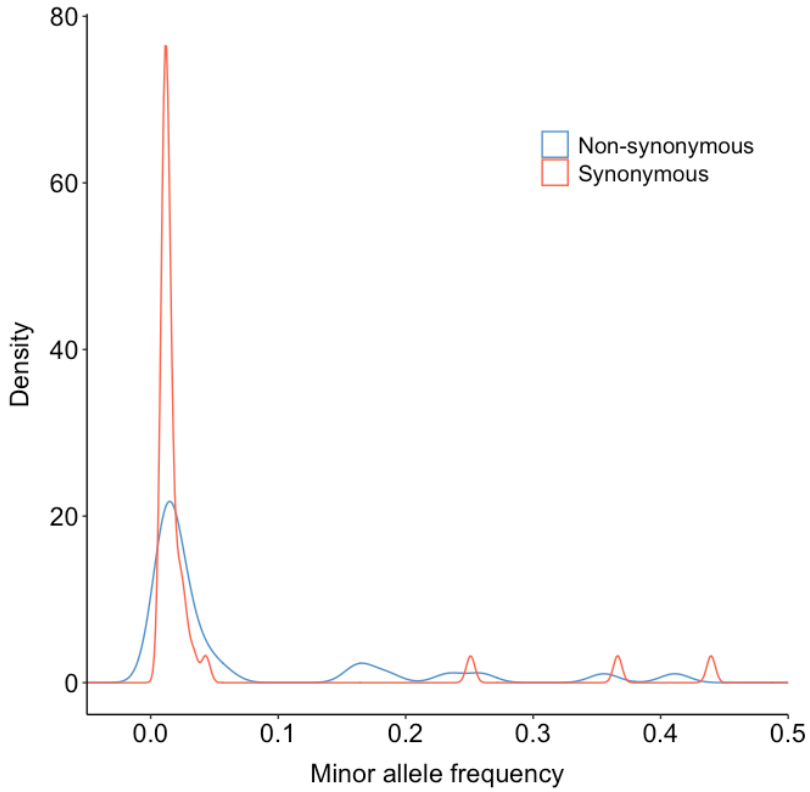


Figure 5.16 | Density plot of minor allele frequencies for synonymous (*n* = 37) and non-synonymous (*n* = 33) mutations identified in mitochondrial DNA from 143 European lobsters (*Homarus gammarus*).

5.3.9 Point heteroplasmy and age

Most of the 143 individuals analysed did not contain any point heteroplasmies (57.3%; *n* = 82). The remaining individuals (*n* = 61) had between 1–8 heteroplasmies. The number of heteroplasmies did not differ among lobster age groups (negative binomial GLM: $\chi^2_{(5,143)} = 1.679$; $p = 0.892$; Fig. 5.17) even after removing the wild lobster age class to control for tissue differences (known-age lobster claws vs wild lobster antennae) (negative binomial GLM: $\chi^2_{(4,122)} = 0.960$; $p = 0.916$). There was no significant difference in the number of heteroplasmies between small (89–100 mm CL) and large, wild lobsters (102–136 mm CL) (Mann-Whitney $U = 68.5$; $p = 0.306$; Fig. 5.18).

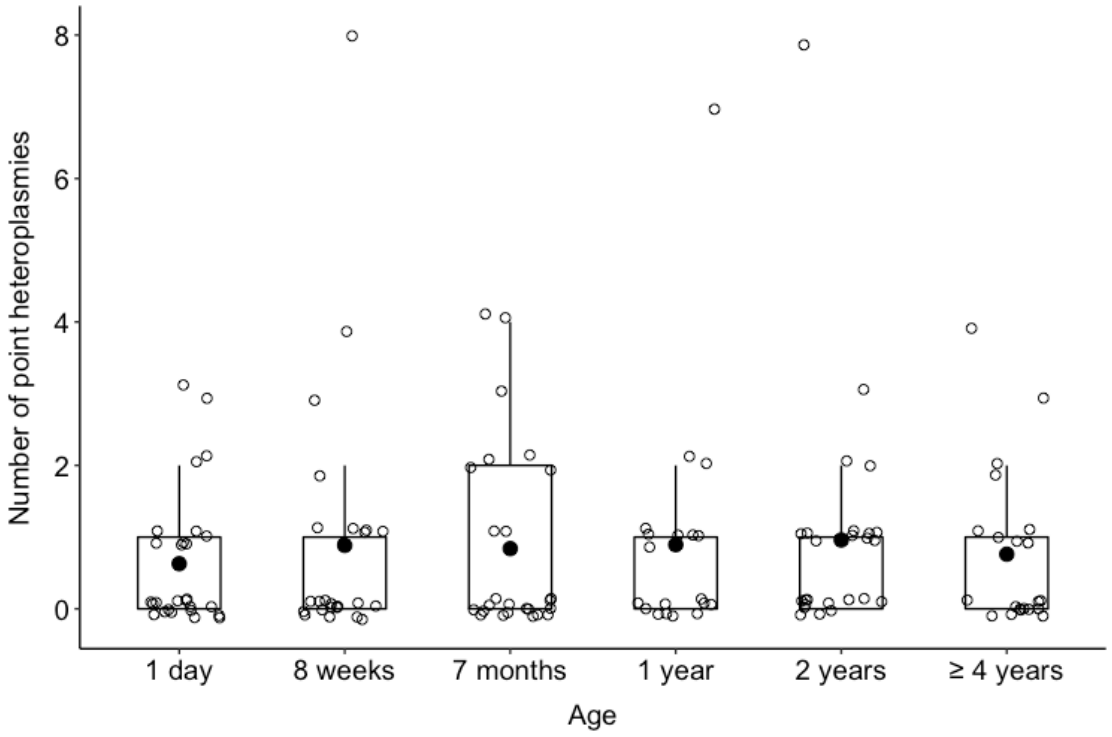


Figure 5.17 | The number of private point heteroplasmies in six age groups of the European lobster (*Homarus gammarus*) (*n* = 143). Boxplots display a mean dot (solid black point), median line, inter-quartile range (IQR) boxes, 1.5*IQR whiskers, and data points.

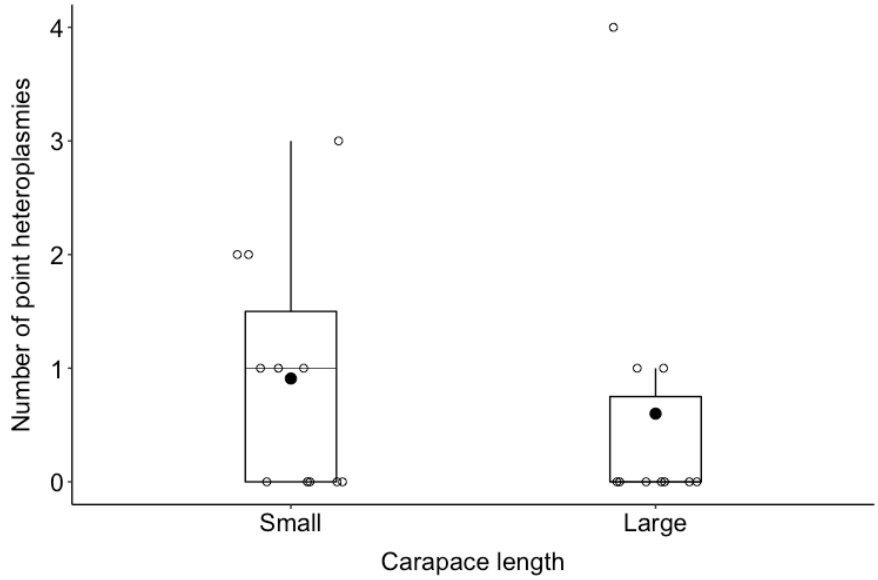


Figure 5.18 | The number of private point heteroplasmies in small (89–100 mm CL) and large (102–136 mm CL) wild-caught European lobsters (*Homarus gammarus*). Boxplots display a mean dot (solid black point), median line, inter-quartile range (IQR) boxes, 1.5*IQR whiskers, and data points.

5.4. Discussion

This study attempted to determine whether the number of mtDNA point mutations could be used to differentiate age groups of European lobsters. Across the entire mitogenome, sequencing data for some, if not all, lobsters were affected by spurious reads. Bioinformatic removal of spurious reads was attempted, but without complete success. After removing the duplicated region and *cox1*, and rigorous SNP filtering to reduce possible false positives in the remaining region (9.8 kb representing 48.3% of the mitogenome), 118 private point heteroplasmies were identified in 143 lobsters. The number of heteroplasmies did not differ among six age groups sampled from 1 day to ≥ 4 years post-hatching.

Most lobsters did not contain point heteroplasmies (57.3%; *n* = 82) and the remaining lobsters (*n* = 61) had 1–8 mtDNA point mutations. The majority of heteroplasmies were transition mutations (*n* = 111; 94%). This transition bias is consistent with studies in other species (e.g. Ma et al. 2018; Jebb et al. 2018) and suggests that heteroplasmy in lobsters is primarily caused by polymerase errors during replication as opposed to oxidative damage (Spelbrink et al. 2000). The observed absence of age-related changes to point heteroplasmy in European lobsters concurs with research on greater mouse-eared bats (Jebb et al. 2018), but contrasts with studies in humans (Li et al. 2015) and mice (Arbeithuber et al. 2020). Differences among species do not appear to be driven by differences in lifespan; mice (*Mus musculus*), for example, accumulate mtDNA mutations faster than longer-lived humans (Wang et al. 1997; Trifunovic 2006). Comparisons across studies/species could be confounded by using DNA from alternative tissues, focusing on different relative age ranges, investigating different regions of the mitogenome, or setting different variant thresholds. For example, if heteroplasmy is quantified in a tissue or mtDNA region with a low mutation rate, in individuals representing a narrow age range, or if variant thresholds are too stringent to detect rare mutations, age-related changes could be missed. These possibilities are discussed below.

A priority for developing ageing methods for ecology and conservation is to test tissues that can be non-lethally sampled, such as the lobster claws and antennae used in this study. In humans, mtDNA mutations accumulate in a wide range of tissues (Li et al. 2015), suggesting that age-related point heteroplasmy in lobsters is unlikely to have been overlooked based on tissue choice. In addition, DNA for five of the six lobster age groups was sampled from claw tissue, which consists largely of connective and muscle tissue (Götze & Saborowski 2011). Muscle tissue consistently contains among the highest levels of heteroplasmy (Payne & Chinnery 2015) and increasing amounts of heteroplasmy have been observed in ageing muscle tissue from various species (Hiona et al. 2010; Naue et al. 2015). However, lobsters can lose claws or antennae as a result of injury or autotomy during fights or predator exposure. Lost appendages can subsequently be regrown. It is not known if any accumulated mtDNA mutations are also lost during limb loss. Limb loss is unlikely to have occurred in the five lobster age groups sampled at ≤ 2 years of age; these individuals were reared in a hatchery or enclosed sea-containers with each lobster occupying its own compartment (Daniels et al. 2015). Subsequently, these lobsters were not involved in fights among conspecifics or exposed to predators. However, moulting (which would have occurred in most lobsters sampled in this study) also involves the loss and regeneration of some underlying muscle and connective tissue (Götze & Saborowski 2011). The extent to which mtDNA mutation accumulation is erased during limb loss or moulting may depend on the amount of original tissue remaining and/or whether heteroplasmy accumulates in the cells that proliferate to form new tissue. It would be interesting to investigate whether heteroplasmy varies with age in lobster tissues that are not affected by limb loss/moulting. However, ageing assays based on lethally-sampled tissues are of less value to fisheries scientists and conservationists.

The age range of the wild-caught lobsters sampled in this study is unknown and age-related heteroplasmy could have been missed as a result of sampling a narrow age range. The European lobster is a long-lived species that is difficult to track in the wild. Obtaining DNA from old individuals of known age in such species is therefore challenging. However, due to the large spread of sizes (89–136 mm CL), the wild lobsters are predicted to represent a wide age range (Sheehy et al. 1999). Despite this,

there was no significant difference in the number of point heteroplasmies between small, 'young' lobsters (89–100 mm CL) and large, 'old' lobsters (102–136 mm CL).

Age-related changes in point heteroplasmy could be overlooked as a consequence of characterising heteroplasmy in a portion (ca. 48%) of the European lobster mitogenome, and/or rigorous SNP filtering. There was evidence of mis-mapping across the duplicated region (8.9 kb) of the European lobster mitogenome, which could influence polymorphism estimates if unresolved. Mis-mapping was not corrected by using alternative mapping methods. Interestingly, there was evidence that a duplication of the *cox1* gene may also have occurred within the lobster mtDNA genome, which warrants further investigation. The duplicated region and *cox1* were subsequently removed from the analysis on heteroplasmy with age. It is unlikely that age-related heteroplasmy was missed as a consequence of excluding these regions. In other species, age-related heteroplasmies tend to be evenly-distributed along the mitogenome or elevated in the control region (Arbeithuber et al. 2020). In lobsters, no mitochondrial partition (including the control region) showed significant enrichment for heteroplasmy. Across the remaining 9.8 kb of mtDNA, most lobsters contained multiple haplotypes, which suggests either genuine or pseudo heteroplasmy. Features of these additional haplotypes suggest they were most likely NUMTs (see below). Therefore, SNPs found in more than one individual were excluded - the rationale being that NUMT-derived SNPs typically re-occur in multiple individuals (Yuan et al. 2020). As consequence of removing shared heteroplasmies, it is possible that real mtDNA variants were also removed. Even so, the number of false negatives is not expected to differ with age.

For the seven protein-coding genes outside of the duplicated region and *cox1*, multiple haplotypes were reconstructed within many individuals. All haplotypes were most-closely related to mtDNA from *H. gammarus* or its American counterpart (*H. americanus*), which rules out contamination from non-lobster DNA. Alternative explanations are inherited heteroplasmy or NUMT contamination. The high levels of sequence divergence present in some haplotypes compared to the reference would suggest that inheritance-based mechanisms are unlikely. In addition, paternal leakage

of mtDNA is rare (Polovina et al. 2020). However, hybridisation can facilitate paternal leakage by weakening the sperm mtDNA recognition system (Mastrantonio et al. 2019), which exists to trigger the destruction of sperm mitochondria (Polovina et al. 2020). Imported American lobsters have been known to hybridise with *H. gammarus* following their escape or release into the range of European lobsters (Ellis et al. 2020). A small minority of haplotypes (*n* = 6; 1.5%) were most similar to mtDNA from *H. americanus*, albeit with *H. gammarus* a close second (≤ 1% sequence difference; Supplementary Table S5.2), suggesting that hybridisation may have occurred. However, additional haplotypes were found in 46.9% of lobsters on average and it is unlikely that the incidence of hybridisation is that high. Nonetheless, a recently-developed panel of 79 SNPs provides powerful molecular identification of *Homarus* hybrids and it would be interesting to apply this method to the lobsters in this study (Ellis et al. 2020). Assuming that the incidence of hybridisation in this study is low, the most likely explanation for the additional haplotypes is NUMT contamination. Some haplotypes contained evidence of loss-of-function, consistent with NUMTs. For example, a stop codon was identified in *nad4* and frameshift mutations were seen in *cox1* and *nad1*. Another key diagnostic of NUMTs is the loss of third codon and transition mutation bias typically seen in functional genes. Across all haplotypes, most mutations were transitions (mean = 86.7%) occurring on the third codon position (mean = 78.8%). It is not unheard of for NUMTs to retain features of their functional counterpart; NUMT sequences can reflect a past mitochondrial genome that existed before the integration event and have subsequently been referred to as 'DNA fossils' (Bensasson et al. 2001).

The SNP filtering procedure adopted in this study should limit any false positives arising from NUMT contamination. Even if NUMT-derived SNPs remain, the number of false positives is unlikely to differ among age groups unless there is more nuclear DNA (and therefore more NUMT-derived mutations exceeding the variant thresholds) in older age groups. Mitochondrial DNA copy number has been shown to decline with age in the muscle tissue of other animals (e.g. Barazzoni et al. 2000; Hartmann et al. 2011; Wachsmuth et al. 2016), which would reduce the ratio of mitochondrial versus nuclear DNA with age. However, the number of mtDNA mutations was consistent

across age groups and so if older lobsters contained more NUMT-derived mutations this would mean they carried proportionately fewer mtDNA mutations. This is unlikely because a decline in mtDNA mutations with age has not been observed in any other species.

Failing to avoid or control for NUMT contamination and mis-mapping can result in inaccurate polymorphism estimates. NUMTs are common in crustaceans (Song et al. 2008), which may be related to their large genomes (Alfsnes et al. 2017; Ding et al. 2018). Therefore, NUMTs need careful consideration in studies based on crustacean mtDNA. A commonly-cited method for avoiding co-amplification of NUMTs is long-range PCR (as adopted in this study). This approach only works if NUMT sequences are considerably shorter than the target, mitochondrial amplicon. However, there are increasing reports of large NUMTs in vertebrates (e.g. Triant & DeWoody 2007) and invertebrates (e.g. Sun & Yang 2016), and this study is not the first to demonstrate the ineffectiveness of long-range PCR to avoid NUMTs (Wang et al. 2020). In this study, a range of alternative mapping methods aimed at removing NUMTs and avoiding mis-mapping were tested, but none were completely successful. Aside from *in silico* removal, other methods for removing NUMT contamination rely on isolating mtDNA prior to sequencing but these methods also fail to completely eliminate NUMTs (Li et al. 2012; Stewart & Chinnery 2015). Mis-mapping of reads across tandem duplications could be reduced using long-read sequencing, assuming reads are long enough to span the two repeats (e.g. Gan et al. 2019). However, long-read sequencing comes at the cost of higher sequencing error and so these technologies are currently not suitable for detecting low-frequency heteroplasmy.

In conclusion, no difference in the number of mtDNA mutations was observed among six lobster age groups. This study was based on DNA from lobster claws or antennae but it remains unknown whether limb loss or moulting could influence age-related heteroplasmy in these body parts. To conclusively rule-out mtDNA mutation accumulation in European lobsters, similar studies are needed based on tissues that are not affected by regeneration. Such tissues were not considered here as they would require lethal sampling. Mis-mapping of reads across the duplicated region and

probable NUMT contamination meant that thorough SNP filtering was required, which may have simultaneously removed real variants. However, even if age-related heteroplasmy occurs and has been missed, this study highlights the difficulty of assessing heteroplasmy in mitochondrial genomes with duplications and/or nuclear pseudogenes. This complexity alone suggests that heteroplasmy is unlikely to be a viable marker of age in European lobsters.

5.5 Supplementary information

Supplementary Table S5.1 |The 88 SNPs between the functional and pseudo sequence of the duplicated region of the European lobster (*Homarus gammarus*) mitogenome, the number of individuals (*n*) containing the alternative allele and its average frequency (Freq). For example, for the functional sequence, *n* refers to the number of individuals containing the pseudo sequence allele and 'Freq' refers to the average frequency of that allele across individuals. Allele counts and frequencies calculated following variant calling and filtering using VarScan.

	Functional sequence				Pseudo sequence			
Gene/tRNA	Position	Allele	*n*	Freq %	Position	Allele	*n*	Freq %
trnL2	1556	T	4	1.1	5457	C	7	1.2
trnL2	1571	C	0	-	5472	T	10	1.3
cox2	1604	T	2	1.0	5505	C	6	1.2
cox2	1639	G	0	-	5540	A	13	1.1
cox2	1641	G	3	1.1	5542	A	12	1.1
cox2	1686	G	36	1.5	5587	A	64	5.3
cox2	1717	G	45	2.6	5618	A	19	1.2
cox2	1732	G	72	1.9	5633	C	29	1.4
cox2	1843	C	79	1.9	5744	G	46	2.1
cox2	1855	T	50	1.5	5756	A	72	5.4
cox2	1921	A	42	1.5	5822	G	65	5.2
cox2	1926	T	41	1.5	5827	C	49	1.7
cox2	1945	C	35	1.7	5846	T	29	1.6
cox2	1972	C	21	1.3	5873	T	40	1.5
cox2	1993	T	20	1.2	5894	C	49	5.6
cox2	2021	A	34	1.5	5922	G	25	1.5
cox2	2043	C	55	4.9	5944	T	19	1.8
cox2	2122	G	42	2.4	6023	A	43	1.5
cox2	2214	T	26	1.4	6114	C	56	4.4
cox2	2245	T	13	1.8	6145	G	19	1.3
cox2	2286	T	34	1.6	6186	C	14	1.4
trnK	2309	C	57	3.4	6209	A	29	1.6
trnK	2331	A	41	3.3	6231	G	39	1.8
trnD	2416	T	35	1.7	6315	C	29	2.2
atp8	2450	G	34	1.7	6349	A	0	-
atp8	2478	T	56	2.3	6377	C	0	-
atp6	2607	C	45	3.6	6506	T	33	3.7
atp6	2692	C	63	4.6	6591	T	0	-
atp6	2769	G	28	1.5	6668	A	26	2.1
atp6	2814	A	62	3.0	6713	G	26	6.3
atp6	2838	C	76	4.9	6737	G	5	15.8
atp6	2841	C	72	4.5	6740	T	1	1.2
atp6	2864	C	2	1.4	6763	G	2	2.0
atp6	2932	T	1	1.4	6831	C	3	2.3
atp6	2950	T	94	7.4	6849	C	55	2.0
atp6	3114	C	9	1.4	7013	T	12	3.1
atp6	3122	G	98	9.5	7021	A	35	1.5
atp6	3217	T	97	2.7	7116	C	29	5.0

Supplementary Table S5.1 (continued)

	Functional sequence				Pseudo sequence			
Gene/tRNA	Position	Allele	*n*	Freq %	Position	Allele	*n*	Freq %
cox3	3253	G	42	2.0	7152	A	3	1.2
cox3	3254	T	67	1.9	7153	C	0	-
cox3	3326	G	47	3.2	7225	A	48	3.3
cox3	3382	G	92	5.9	7281	A	10	2.1
cox3	3402	C	0	-	7300	T	3	1.8
cox3	3411	T	92	4.2	7309	C	37	2.7
cox3	3439	A	10	2.5	7337	G	14	3.4
cox3	3775	T	12	4.7	7673	C	14	4.3
cox3	3823	C	63	11.5	7721	A	57	2.2
cox3	3824	G	22	5.5	7722	A	69	2.9
cox3	3836	G	77	3.6	7734	A	77	2.3
cox3	3868	C	80	8.7	7766	T	54	2.3
cox3	3894	C	76	2.4	7792	T	0	-
cox3	3971	G	87	2.9	7869	A	61	1.8
cox3	4035	G	81	4.8	7933	A	32	1.8
trnG	4063	G	3	1.4	7961	A	1	3.0
trnG	4071	G	100	7.0	7969	A	15	1.5
nad3	4181	T	50	1.5	8079	C	26	3.3
nad3	4248	G	87	3.1	8133	A	44	1.4
nad3	4234	G	83	2.6	8147	A	88	3.2
nad3	4372	C	0	-	8271	A	65	2.9
trnA	4481	G	81	2.0	8380	A	82	2.2
trnA	4506	G	80	1.9	8405	A	85	2.1
trnR	4541	G	83	2.0	8440	A	80	2.0
trnN	4602	G	37	1.3	8501	C	41	4.0
trnN	4621	A	66	1.5	8520	G	43	4.2
trnN	4639	G	32	4.8	8538	A	14	2.6
trnN	4649	T	36	1.2	8548	C	23	6.2
trnS1	4672	C	50	6.7	8571	T	5	1.0
trnS1	4684	G	24	2.0	8583	A	0	-
trnE	4793	C	0	-	8692	T	0	-
NA	4822	G	0	-	8720	A	0	-
trnF	4841	T	13	2.1	8739	C	0	-
trnF	4844	G	0	-	8742	A	0	-
trnF	4893	T	8	17.7	8791	A	0	-
nad5	8848	G	24	2.3	4950	A	0	-
nad5	8884	A	11	4.1	4986	G	2	1.2
nad5	8898	A	5	1.3	5000	G	80	5.8
nad5	8909	G	9	1.2	5011	A	57	7.4
nad5	8951	T	11	1.8	5053	C	0	-
nad5	8986	T	28	7.2	5088	C	0	-
nad5	8992	A	1	1.2	5094	G	42	7.0
nad5	9007	T	0	-	5109	A	0	-
nad5	9015	G	28	6.0	5117	A	0	-
nad5	9022	T	0	-	5124	C	0	-
nad5	9058	A	16	4.8	5160	G	25	2.1
nad5	9069	G	0	-	5171	A	7	1.2
nad5	9086	T	16	1.3	5188	C	60	5.7
nad5	9230	C	1	1.5	5332	T	3	2.8
nad5	9263	A	0	-	5365	G	46	5.3

Supplementary Table S5.2 | Top Blast hit for each mitochondrial haplotype (Hap) reconstructed in this study. Species identity was either *Homarus gammarus* (-) or *H. americanus* (americanus). Freq = the number of individuals containing that haplotype. % identity = the percentage of shared nucleotides with the sequence in the GenBank database. For haplotypes with highest sequence similarity to *H. americanus*, % identity for *H. gammarus* is provided in parentheses. Haplotypes in bold and underlined contained stop codons or frameshift mutations.

Gene	Hap	Freq	Species	% identity	Gene	Hap	Freq	Species	% identity
cox1	I	68	-	100.0	cytb	XXII	1	-	98.2
cox1	II	4	-	99.9	cytb	XXIII	1	-	99.8
cox1	III	3	-	99.9	cytb	XXIV	12	-	99.8
cox1	IV	18	-	99.9	cytb	XXV	1	-	97.8
cox1	V	20	-	99.9	cytb	XXVI	4	-	94.3
cox1	**VI**	**7**	**-**	**98.2**	cytb	XXVII	1	-	99.7
cox1	VII	1	-	99.5	cytb	XXVIII	1	-	94.3
cox1	VIII	2	-	99.9	cytb	XXIX	2	-	99.7
cox1	IX	1	-	99.5	cytb	XXX	2	-	99.8
cox1	X	1	-	99.5	cytb	XXXI	1	-	99.6
cox1	XI	1	-	99.5	cytb	XXXII	1	-	97.9
cox1	XII	1	-	99.5	cytb	XXXIII	1	-	99.4
cox1	XIII	5	-	99.7	cytb	XXXIV	1	-	99.9
cox1	XIV	1	-	99.7	cytb	XXXV	1	-	99.2
cox1	XV	3	-	99.9	cytb	XXXVI	1	-	99.7
cox1	XVI	4	-	99.9	cytb	XXXVII	3	-	99.7
cox1	XVII	1	-	99.7	cytb	XXXVIII	1	-	99.9
cox1	XVIII	5	-	99.9	cytb	XXXIX	1	-	99.4
cox1	**XIX**	**1**	**-**	**98.3**	cytb	XL	1	-	99.5
cox1	XX	1	-	99.7	cytb	XLI	1	-	99.7
cox1	XXI	1	-	99.9	cytb	XLII	1	-	97.3
cox1	XXII	3	-	99.9	cytb	XLIII	2	-	99.7
cox1	XXIII	4	-	99.8	cytb	XLIV	1	-	99.7
cox1	XXIV	3	-	99.9	cytb	XLV	2	-	99.7
cox1	XXV	1	-	99.9	cytb	XLVI	1	-	99.5
cox1	XXVI	4	-	99.9	cytb	XLVII	1	-	95.2
cox1	XXVII	5	-	99.7	cytb	XLVIII	1	-	97.4
cox1	XXVIII	5	-	99.9	cytb	XLIX	1	-	97.7
cox1	XXIX	2	-	99.6	cytb	L	1	-	95.5
cox1	XXX	1	-	99.8	cytb	LI	2	-	99.7
cox1	XXXI	1	-	99.9	cytb	LII	1	-	96.6
cox1	XXXII	3	-	99.9	cytb	LIII	1	-	99.7
cox1	XXXIII	1	-	99.9	cytb	LIV	1	-	96.0
cox1	XXXIV	1	-	99.9	cytb	LV	1	-	99.4
cox1	XXXV	1	-	99.9	cytb	LVI	1	-	99.7
cox1	XXXVI	1	-	99.9	cytb	LVII	1	-	99.1
cox1	XXXVII	1	-	99.7	cytb	LVIII	1	-	99.8
cox1	**XXXVIII**	**1**	**-**	**98.1**	cytb	LIX	1	-	99.7
nad4	I	16	-	100.0	cytb	LX	1	-	99.4
nad4	II	15	-	99.9	cytb	LXI	1	-	99.5
nad4	III	9	-	99.3	cytb	LXII	1	-	99.8
nad4	**IV**	**4**	**americanus**	**92.9 (91.9)**	cytb	LXIII	1	-	98.8
nad4	V	10	-	99.3	cytb	LXIV	1	-	96.8
nad4	VI	1	-	95.5	cytb	LXV	1	-	99.8
nad4	VII	1	-	99.9	cytb	LXVI	1	-	94.4

Supplementary Table S5.2 (continued)

Gene	Hap	Freq	Species	% identity	Gene	Hap	Freq	Species	% identity
nad4	VIII	2	-	99.9	cytb	LXVII	1	-	96.3
nad4	IX	1	-	99.2	cytb	LXVIII	1	-	96.5
nad4	X	1	-	99.5	cytb	LXIX	1	-	98.1
nad4	XI	10	-	99.9	cytb	LXX	1	-	99.8
nad4	XII	1	-	99.8	cytb	LXXI	1	-	99.8
nad4	XIII	6	-	99.9	cytb	LXXII	1	-	99.5
nad4	XIV	1	-	97.8	cytb	LXXIII	1	-	99.7
nad4	XV	3	-	99.9	cytb	LXXIV	1	-	94.2
nad4	XVI	1	-	99.6	cytb	LXXV	1	-	95.5
nad4	XVII	1	-	95.8	cytb	LXXVI	1	-	99.6
nad4	XVIII	1	-	97.3	cytb	LXXVII	1	-	95.2
nad4	XIX	1	-	98.1	cytb	LXXVIII	1	-	96.7
nad4	XX	18	-	99.9	cytb	LXXIX	1	-	94.0
nad4	XXI	1	-	98.3	cytb	LXXX	1	-	99.8
nad4	XXII	2	-	99.9	cytb	LXXXI	1	-	94.2
nad4	XXIII	1	-	96.6	cytb	LXXXII	1	-	95.5
nad4	XXIV	1	-	97.8	cytb	LXXXIII	1	-	97.7
nad4	XXV	2	-	99.9	cytb	LXXXIV	1	-	99.4
nad4	XXVI	4	-	99.9	cytb	LXXXV	1	-	99.6
nad4	XXVII	1	-	96.3	cytb	LXXXVI	1	-	94.7
nad4	XXVIII	5	americanus	93.1 (92.0)	cytb	LXXXVII	1	-	99.6
nad4	XXIX	1	-	98.9	cytb	LXXXVIII	1	-	97.4
nad4	XXX	1	-	95.2	nad1	I	13	-	100.0
nad4	XXXI	1	-	99.3	nad1	II	62	-	99.9
nad4	XXXII	2	-	99.9	nad1	III	12	-	99.8
nad4	XXXIII	1	-	98.1	nad1	IV	2	-	97.7
nad4	XXXIV	1	-	97.5	nad1	V	1	-	97.4
nad4	XXXV	1	-	99.9	nad1	VI	1	-	96.9
nad4	XXXVI	2	-	99.9	nad1	VII	1	-	99.9
nad4	XXXVII	4	-	99.9	nad1	VIII	1	-	97.8
nad4	XXXVIII	1	-	97.7	nad1	IX	1	-	99.2
nad4	XXXIX	1	-	99.9	nad1	X	1	-	98.6
nad4	XL	1	-	99.9	nad1	XI	1	-	99.8
nad4	XLI	1	-	99.6	nad1	XII	6	-	99.8
nad4	XLII	1	-	97.8	nad1	XIII	6	-	99.7
nad4	XLIII	1	-	98.1	nad1	XIV	2	-	99.7
nad4	XLIV	1	-	93.8	**nad1**	**XV**	**1**	**-**	**98.2**
nad4	XLV	1	-	97.3	nad1	XVI	2	-	99.7
nad4	XLVI	1	-	99.9	nad1	XVII	2	-	99.8
nad4	XLVII	1	-	99.0	nad1	XVIII	1	-	99.2
nad4	XLVIII	1	-	95.0	nad1	XIX	1	-	99.4
nad4	XLIX	1	-	99.9	nad1	XX	1	-	99.3
nad4	L	1	-	98.7	nad1	XXI	1	-	97.1
nad4	LI	1	-	98.9	nad1	XXII	1	-	98.0
nad4	LII	3	-	99.9	nad1	XXIII	1	-	99.7
nad4	LIII	1	-	95.2	nad1	XXIV	2	-	99.8
nad4	LIV	1	-	99.9	nad1	XXV	1	-	99.7
nad4	LV	1	-	93.1	nad1	XXVI	1	-	99.3
nad4	**LVI**	**1**	**-**	**93.2**	nad1	XXVII	5	-	95.0
nad4	LVII	1	-	96.1	nad1	XXVIII	1	-	99.4
nad4	LVIII	1	-	99.9	nad1	XXIX	1	-	98.5
nad4	LIX	1	-	93.4	nad1	XXX	1	-	97.1
nad4	LX	1	-	99.5	nad1	XXXI	1	-	99.2
nad4	LXI	1	-	97.8	nad1	XXXII	2	-	99.8

Supplementary Table S5.2 (continued)

Gene	Hap	Freq	Species	% identity	Gene	Hap	Freq	Species	% identity
nad4	LXII	1	-	94.3	nad1	XXXIII	1	-	99.0
nad4	LXIII	1	-	99.9	nad1	XXXIV	1	-	99.6
nad4	**LXIV**	**1**	**americanus**	**92.8 (92.0)**	nad1	XXXV	1	-	95.3
nad4	LXV	1	-	99.4	nad1	XXXVI	9	-	99.8
nad4	LXVI	1	-	99.4	nad1	XXXVII	1	-	99.5
nad4	LXVII	1	-	97.7	**nad1**	**XXXVIII**	**1**	**-**	**98.9**
nad4	LXVIII	1	-	99.9	nad1	XXXIX	4	-	99.7
nad4	LXIX	1	-	98.8	nad1	XL	1	-	99.7
nad4l	I	123	-	100.0	nad1	XLI	1	-	98.8
nad4l	II	9	americanus	94.1 (93.1)	nad1	XLII	2	-	99.7
nad4l	III	6	americanus	94.4 (93.4)	nad1	XLIII	1	-	96.5
nad4l	IV	11	-	97.0	nad1	XLIV	1	-	98.4
nad4l	V	2	-	99.7	nad1	XLV	1	-	99.7
nad4l	VI	6	-	100.0	nad1	XLVI	1	-	99.8
nad4l	VII	1	-	98.0	nad1	XLVII	1	-	98.3
nad4l	VIII	3	-	98.4	nad1	XLVIII	2	-	99.6
nad4l	IX	1	-	99.7	nad1	XLIX	1	-	95.2
nad4l	X	1	-	97.7	**nad1**	**L**	**1**	**-**	**99.8**
nad4l	XI	1	-	96.4	nad1	LI	1	-	99.7
nad4l	XII	1	-	95.1	nad1	LII	1	-	99.9
nad4l	XIII	4	-	99.7	nad1	LIII	1	-	99.8
nad4l	XIV	1	-	97.7	nad1	LIV	1	-	98.5
nad4l	XV	1	-	98.4	nad1	LV	1	-	97.0
nad4l	XVI	2	-	99.7	nad1	LVI	1	-	98.6
nad4l	XVII	1	-	99.7	nad1	LVII	1	-	99.5
nad4l	XVIII	1	-	100.0	nad1	LVIII	1	-	99.8
nad4l	XIX	1	-	96.7	nad1	LIX	1	-	97.6
nad4l	XX	1	-	96.7	nad1	LX	1	-	99.4
nad4l	XXI	2	americanus	96.0 (95.7)	nad1	LXI	1	-	99.4
nad4l	XXII	2	-	99.7	nad1	LXII	1	-	99.8
nad4l	XXIII	1	-	98.0	nad1	LXIII	1	-	97.3
nad6	I	115	-	100.0	nad1	LXIV	1	-	99.5
nad6	II	8	-	99.4	nad1	LXV	1	-	98.3
nad6	III	3	-	93.8	nad1	LXVI	1	-	98.7
nad6	IV	1	-	99.6	nad1	LXVII	1	-	99.7
nad6	V	1	-	95.8	nad1	LXVIII	1	-	97.9
nad6	VI	1	-	99.6	nad1	LXIX	1	-	97.8
nad6	VII	1	-	97.5	nad1	LXX	1	-	96.3
nad6	VIII	4	-	99.8	nad1	LXXI	1	-	99.2
nad6	IX	1	-	99.0	nad1	LXXII	1	-	98.5
nad6	X	4	-	99.6	nad1	LXXIII	1	-	99.8
nad6	XI	11	-	94.0	nad1	LXXIV	1	-	99.9
nad6	XII	1	-	98.8	nad1	LXXV	1	-	98.3
nad6	XIII	2	-	99.2	nad2	I	64	-	100.0
nad6	XIV	3	-	99.4	nad2	II	26	-	99.9
nad6	XV	1	-	99.6	nad2	III	11	-	99.9
nad6	XVI	1	-	97.3	nad2	IV	14	-	99.9
nad6	XVII	1	-	98.8	nad2	V	12	-	99.5
nad6	XVIII	1	-	99.8	nad2	VI	7	-	99.9
nad6	XIX	5	-	99.8	nad2	VII	2	-	99.6
nad6	XX	1	-	97.5	nad2	VIII	2	-	99.5
nad6	XXI	2	-	99.8	nad2	IX	1	-	99.3
nad6	XXII	1	-	98.5	nad2	X	1	-	99.3
nad6	XXIII	1	-	96.0	nad2	XI	5	-	99.9

Supplementary Table S5.2 (continued)

Gene	Hap	Freq	Species	% identity	Gene	Hap	Freq	Species	% identity
nad6	XXIV	2	-	99.6	nad2	XII	2	-	99.3
nad6	XXV	3	-	99.8	nad2	XIII	8	-	99.9
nad6	XXVI	1	-	99.6	nad2	XIV	5	-	99.9
nad6	XXVII	1	-	94.2	nad2	XV	4	-	99.4
nad6	XXVIII	1	-	96.0	nad2	XVI	6	-	99.8
nad6	XXIX	2	-	99.8	nad2	XVII	1	-	99.9
nad6	XXX	2	-	97.3	nad2	XVIII	1	-	99.9
nad6	XXXI	1	-	97.9	nad2	XIX	1	-	99.3
nad6	XXXII	1	-	98.1	nad2	XX	2	-	99.9
nad6	XXXIII	1	-	94.6	nad2	XXI	4	-	99.2
nad6	XXXIV	2	-	99.8	nad2	XXII	1	-	99.6
nad6	XXXV	1	-	97.7	nad2	XXIII	1	-	99.5
nad6	XXXVI	1	-	95.6	nad2	XXIV	1	-	99.7
nad6	XXXVII	1	-	99.2	nad2	XXV	1	-	99.7
nad6	XXXVIII	1	-	99.6	nad2	XXVI	2	-	99.9
nad6	XXXIX	1	-	99.8	nad2	XXVII	1	-	99.7
nad6	XL	1	-	96.7	nad2	XXVIII	5	-	99.8
nad6	XLI	1	-	98.5	nad2	XXIX	1	-	99.4
nad6	XLII	1	-	98.5	nad2	XXX	1	-	99.8
nad6	XLIII	1	-	96.7	nad2	XXXI	2	-	99.7
nad6	XLIV	1	-	97.9	nad2	XXXII	5	-	99.9
nad6	XLV	1	-	95.4	nad2	XXXIII	1	-	99.8
nad6	XLVI	1	-	98.5	nad2	XXXIV	1	-	99.8
nad6	XLVII	1	-	96.5	nad2	XXXV	1	-	99.7
nad6	XLVIII	1	-	96.5	nad2	XXXVI	1	-	99.7
nad6	XLIX	1	-	98.7	nad2	XXXVII	1	-	99.8
cytb	I	16	-	100.0	nad2	XXXVIII	1	-	99.8
cytb	II	21	-	99.7	nad2	XXXIX	3	-	99.8
cytb	III	50	-	99.9	nad2	XL	1	-	99.6
cytb	IV	2	-	94.1	nad2	XLI	1	-	99.7
cytb	V	3	-	99.9	nad2	XLII	1	-	99.4
cytb	VI	1	-	98.4	nad2	XLIII	1	-	99.8
cytb	VII	3	-	99.9	nad2	XLIV	1	-	99.9
cytb	VIII	1	-	99.7	nad2	XLV	1	-	99.6
cytb	IX	1	-	95.2	nad2	XLVI	2	-	99.9
cytb	X	5	-	99.7	nad2	XLVII	1	-	99.3
cytb	XI	1	-	99.4	nad2	XLVIII	1	-	99.4
cytb	XII	6	-	99.8	nad2	XLIX	1	-	99.4
cytb	XIII	2	-	99.7	nad2	L	1	-	99.2
cytb	XIV	1	-	98.9	nad2	LI	1	-	99.5
cytb	XV	14	-	99.8	nad2	LII	1	-	99.6
cytb	XVI	1	-	97.4	nad2	LIII	1	-	99.4
cytb	XVII	1	-	97.2	nad2	LIV	1	-	99.6
cytb	XVIII	1	-	99.2	nad2	LV	1	-	99.6
cytb	XIX	1	-	98.6	nad2	LVI	1	-	99.8
cytb	XX	3	-	100.0	nad2	LVII	1	-	99.7
cytb	XXI	1	-	94.5	nad2	LVIII	1	-	99.3

Supplementary Table S5.3 | The 118 private point heteroplasmies detected in 143 European lobsters (*Homarus gammarus*). Alternative alleles (Alt) are classified as transition (Ts) or transversion (Tv) mutations. For protein-coding genes, heteroplasmies are synonymous (S) or non-synonymous (NS). NS mutations were deemed deleterious if they had a PROVEAN score less than -2.5. MAF = minor allele frequency. Positions according to the reference mitogenome (Ref) available on GenBank: MH747083.

Position	Region	Ref	Alt	MAF	Ts/Tv	S/NS	Effect
10617	*nad4*	G	A	0.23	Ts	NS	Neutral
10654	*nad4*	G	A	0.03	Ts	S	-
10709	*nad4*	A	G	0.03	Ts	S	-
10755	*nad4*	G	A	0.01	Ts	NS	Deleterious
10769	*nad4*	A	G	0.01	Ts	S	-
10872	*nad4*	T	C	0.04	Ts	NS	Deleterious
10990	*nad4*	A	G	0.03	Ts	NS	Deleterious
11243	*nad4*	G	A	0.01	Ts	S	-
11415	*nad4*	A	G	0.01	Ts	NS	Neutral
11687	*nad4*	A	T	0.01	Tv	S	-
11744	*nad4*	T	C	0.44	Ts	S	-
11849	*nad4*	T	C	0.01	Ts	S	-
12111	*nad4l*	T	C	0.01	Ts	S	-
12188	*nad4l*	A	G	0.26	Ts	NS	Neutral
12232	*tRNA*	T	C	0.44	Ts	-	-
12517	*nad6*	A	G	0.05	Ts	NS	Neutral
12579	*nad6*	T	C	0.02	Ts	S	-
12709	*nad6*	A	G	0.03	Ts	NS	Neutral
12774	*nad6*	A	G	0.03	Ts	S	-
12992	*cytb*	A	G	0.01	Ts	S	-
13060	*cytb*	T	C	0.01	Ts	NS	Neutral
13199	*cytb*	A	G	0.01	Ts	S	-
13266	*cytb*	T	C	0.16	Ts	NS	Deleterious
13267	*cytb*	T	C	0.16	Ts	NS	Deleterious
13295	*cytb*	A	G	0.02	Ts	S	-
13307	*cytb*	G	A	0.02	Ts	S	-
13382	*cytb*	C	T	0.02	Ts	S	-
13495	*cytb*	G	A	0.01	Ts	NS	Deleterious
13576	*cytb*	G	A	0.18	Ts	NS	Deleterious
13598	*cytb*	A	G	0.01	Ts	S	-
13640	*cytb*	T	C	0.01	Ts	S	-
13799	*cytb*	T	C	0.01	Ts	S	-
13817	*cytb*	A	G	0.01	Ts	S	-
13953	*cytb*	G	A	0.06	Ts	NS	Neutral
13967	*cytb*	T	C	0.37	Ts	S	-
13974	*cytb*	G	A	0.02	Ts	NS	Neutral
14091	-	A	G	0.13	Ts	-	-
14217	*nad1*	A	G	0.01	Ts	S	-
14221	*nad1*	A	G	0.01	Ts	S	-
14288	*nad1*	G	A	0.01	Ts	NS	Deleterious
14311	*nad1*	A	G	0.01	Ts	S	-
14332	*nad1*	T	C	0.01	Ts	S	-
14502	*nad1*	A	G	0.02	Ts	S	-

Supplementary Table S5.3 (continued)

Position	Region	Ref	Alt	MAF	Ts/Tv	S/NS	Effect
14542	*nad1*	T	C	0.01	Ts	S	-
14629	*nad1*	G	A	0.25	Ts	S	-
14728	*nad1*	A	G	0.01	Ts	S	-
14825	*nad1*	G	A	0.03	Ts	NS	Neutral
14857	*nad1*	T	C	0.01	Ts	S	-
15066	-	G	A	0.04	Ts	-	-
15226	*16S*	C	T	0.01	Ts	-	-
15242	*16S*	A	G	0.04	Ts	-	-
15334	*16S*	A	G	0.42	Ts	-	-
15370	*16S*	T	C	0.48	Ts	-	-
16522	*tRNA*	A	G	0.01	Ts	-	-
16537	*tRNA*	G	A	0.02	Ts	-	-
16628	*12S*	G	A	0.02	Ts	-	-
16631	*12S*	A	G	0.01	Ts	-	-
16633	*12S*	T	C	0.01	Ts	-	-
16635	*12S*	A	G	0.01	Ts	-	-
16755	*12S*	T	C	0.01	Ts	-	-
16804	*12S*	T	C	0.01	Ts	-	-
16855	*12S*	G	A	0.06	Ts	-	-
16869	*12S*	A	G	0.04	Ts	-	-
16876	*12S*	A	G	0.01	Ts	-	-
16899	*12S*	T	C	0.02	Ts	-	-
16930	*12S*	A	G	0.01	Ts	-	-
16950	*12S*	A	G	0.01	Ts	-	-
16956	*12S*	T	C	0.02	Ts	-	-
17008	*12S*	A	G	0.01	Ts	-	-
17062	*12S*	T	C	0.01	Ts	-	-
17225	*12S*	T	C	0.01	Ts	-	-
17369	*12S*	T	C	0.02	Ts	-	-
17380	*12S*	T	C	0.03	Ts	-	-
17588	Control region	A	G	0.01	Ts	-	-
17778	Control region	C	T	0.38	Ts	-	-
17791	Control region	A	C	0.08	Tv	-	-
18076	Control region	C	G	0.23	Tv	-	-
18079	Control region	T	C	0.25	Ts	-	-
18143	Control region	T	C	0.01	Ts	-	-
18156	Control region	C	A	0.05	Tv	-	-
18180	Control region	T	C	0.01	Ts	-	-
18203	Control region	A	G	0.01	Ts	-	-
18284	Control region	T	C	0.05	Ts	-	-
18292	Control region	T	C	0.02	Ts	-	-
18345	Control region	T	C	0.01	Ts	-	-
18347	Control region	T	C	0.03	Ts	-	-
18455	Control region	T	C	0.01	Ts	-	-
18709	Control region	G	A	0.02	Ts	-	-
18714	Control region	A	G	0.01	Ts	-	-
18865	Control region	T	C	0.01	Ts	-	-
18911.1	*tRNA*	C	A	0.01	Tv	-	-
18911.2	*tRNA*	C	T	0.01	Ts	-	-

Supplementary Table S5.3 (continued)

Position	Region	Ref	Alt	MAF	Ts/Tv	S/NS	Effect
19029	*tRNA*	G	A	0.01	Ts	-	-
19046	*tRNA*	C	T	0.03	Ts	-	-
19086	*nad2*	T	C	0.01	Ts	NS	Neutral
19175	*nad2*	G	A	0.36	Ts	NS	Deleterious
19300	*nad2*	A	G	0.01	Ts	S	-
19315	*nad2*	A	G	0.04	Ts	S	-
19322	*nad2*	C	T	0.02	Ts	NS	Neutral
19341	*nad2*	G	A	0.01	Ts	NS	Neutral
19373	*nad2*	G	A	0.01	Ts	NS	Deleterious
19380	*nad2*	C	T	0.01	Ts	NS	Deleterious
19382	*nad2*	T	C	0.02	Ts	NS	Deleterious
19383	*nad2*	T	C	0.01	Ts	NS	Deleterious
19443	*nad2*	T	C	0.41	Ts	NS	Deleterious
19445	*nad2*	A	G	0.01	Ts	NS	Deleterious
19551	*nad2*	C	T	0.01	Ts	NS	Deleterious
19641	*nad2*	T	C	0.01	Ts	NS	Deleterious
19646	*nad2*	G	A	0.01	Ts	NS	Deleterious
19654	*nad2*	A	G	0.01	Ts	S	-
19711	*nad2*	T	C	0.01	Ts	S	-
19714	*nad2*	C	T	0.01	Ts	S	-
19746	*nad2*	C	T	0.02	Ts	NS	Neutral
19759	*nad2*	T	C	0.01	Ts	S	-
19886	*nad2*	C	A	0.03	Tv	NS	Neutral
19889	*nad2*	T	C	0.01	Ts	S	-
19898	*nad2*	C	T	0.01	Ts	NS	Deleterious
19951	*nad2*	G	T	0.01	Tv	S	-

Chapter 6

General discussion

6.1 Synopsis

Crustaceans are extremely difficult to age and the consequent lack of data on the age structure of stocks makes it impossible to reliably predict population growth or decline for sustainable fisheries management. The aim of this thesis was to investigate whether crustacean age can be estimated using DNA-based markers. Three potential markers of age were investigated in laboratory-reared cohorts of red cherry shrimp (*Neocaridina davidi*) and/or the economically-important European lobster (*Homarus gammarus*). Firstly, this research assessed whether genome-wide levels of methylation could differentiate age groups of lobsters and shrimp. Secondly, site-specific methylation was quantified in lobster ribosomal DNA (rDNA) to investigate whether age can be predicted using percentage methylation at a combination of loci within this region. Lastly, the mitochondrial genomes of both species were sequenced to high depth to determine whether the number of mitochondrial DNA (mtDNA) mutations varies with age. The results from these studies are summarised below and potential avenues for future research are discussed.

6.2 Genome-wide (global) methylation in lobsters and shrimp (chapter 2)

Global DNA methylation has been shown to decline throughout life in a range of animals (Berdyshev et al. 1967; Wilson et al. 1987; Fuke et al. 2004; Gryzińska et al. 2013; Lian et al. 2015). These patterns suggest that genome-wide levels of methylation could be used to estimate animal age. However, little is known about the prevalence of, and age-related changes in, global DNA methylation in crustaceans. Global DNA methylation was quantified in different age groups of lobsters and shrimp using a widely-used, enzyme-linked immunosorbent assay (ELISA) kit. There was no difference in global DNA methylation between larval or adult lobsters. Among all shrimp life-cycle stages (eggs, larvae, juveniles and adults), shrimp larvae had higher levels of global DNA methylation compared to juveniles and adults, but no other pairwise comparisons were significant. These results suggest that global DNA methylation is not a useful predictor of crustacean age.

However, this study exposed issues with the reliability of the ELISA kit used (MethylFlash by Epigentek). Specifically, there was high variability in measured absorbance between technical replicates and among methylated controls across runs. This appeared to be a problem with the provided controls themselves and, despite communication with the kit provider, this could not be satisfactorily resolved. When considered alongside reported issues with the kit from other users (e.g. Price 2009; Lisanti 2013), a decision was made not to invest further resources (time and money) in the method. While there are other methods for quantifying global DNA methylation (Kurdyukov & Bullock 2016), these are either more technically difficult or costly, which are important considerations when developing an ageing assay for animal ecology/fisheries management. In addition, given the promising results from the rDNA analysis (see section 6.3 below), it was not considered worthwhile to pursue these alternative methods of assessing global DNA methylation for age estimation.

6.3 Site-specific rDNA methylation in lobsters (chapter 3)

In recent years, site-specific methylation has proven to be one of the best molecular predictors of chronological age in a range of animals (e.g. Horvath 2013; Polanowski et al. 2014; De Paoli-Iseppi et al. 2019; Anastasiadi & Piferrer 2020) and was subsequently a priority for investigation in crustaceans. However, in the early stages of this PhD research, studies were mostly based on quantifying site-specific methylation at candidate genes first identified in humans. This approach was unlikely to be applicable to crustaceans because of the evolutionary distance between humans and crustaceans. Even if orthologous sequences do occur in crustaceans, the lack of crustacean genomic data (Yuan et al. 2016) meant it was not possible to demonstrate their existence. Alternatively, DNA methylation can be measured for all genomic loci, but these data are extremely expensive to obtain (Kurdyukov & Bullock 2016) and identifying differentially-methylated regions depends on a reference genome for read mapping. Fortuitously, in 2019, a study in mice demonstrated that levels of site-specific methylation in the evolutionarily-conserved rDNA may be a universal marker of age (Wang & Lemos 2019).

Using bisulphite sequencing, site-specific methylation was measured across 5.2 kb of lobster rDNA sequence in individuals that ranged in known age from 0–2 years old. This was the first time that methylation status had been investigated as a tool for estimating age in crustaceans, and one of the first (if not the first) to investigate rDNA methylation as a marker of age in a wild animal. After independent analysis of methylation with age at each locus, the ten 'best' loci were incorporated into a multiple linear regression model with predicted age plotted against known age. The model predicted lobster age with high accuracy and precision (R^2 = 0.91, standard deviation = 2.02 months). This research has highlighted the exciting potential of rDNA-based methylation for ageing lobsters.

However, the ageing model was calibrated for a relatively narrow age range (0–2 years old), and uncertainty remains over the applicability of rDNA methylation for estimating age in older lobsters. The ten selected loci were used to estimate age in a sample of wild lobsters of unknown age. Previous mark-recapture studies from the 1980s–early 2000s (Bannister & Addison 1998; Uglem et al. 2005; Schmalenbach et al. 2011) suggest that the wild lobsters should be at least 4 years old based on their size. However, rDNA-estimated age was lower than expected for these lobsters (mean = 25.0 months; range = 15.0–41.5 months). Potential explanations include 1) differences in methylation levels across tissues (claw and antennal DNA was sampled from known-age and wild lobsters, respectively), 2) overfitting of the model to the small age range of known-age lobsters and/or 3) non-linear changes in rDNA methylation levels with age in older lobsters.

Firstly, to assess potential tissue effects, sourcing claw DNA from older, wild lobsters is a recommended priority to assess whether the rDNA ageing model provides a higher (expected) estimated age when the sampled tissue is consistent between known-age and unknown-age lobsters. In addition, a multi-tissue comparison of methylation changes with age using known-age lobsters would help to determine which tissue, or combination of tissues, provides the best explanatory power and precision. Testing lobster tissues that can be non-destructively sampled without compromising the commercial value of wild catch (e.g. pleopods or antennae) would be particularly

relevant to fisheries management. Secondly, the known-age lobsters used in this study were sourced from the National Lobster Hatchery, which, at the time of this study, had reared and sampled lobsters from sea-based containers for up to 2 years of age. Lobsters from these cohorts are still available and further sampling of these individuals (now 4 years old) is now possible to calibrate the ageing model across a wider range of known ages.

Testing molecular markers of age across a wide range of known-age individuals is likely to be a common problem for many of the most economically-valuable crustaceans, which tend to have (or are predicted to have) long lifespans (Vogt 2019). The European lobster has a predicted lifespan of 42–72 years (Sheehy et al. 1999). Possible sources of genetic material from known-age lobsters include 1) previous tag and recapture experiments and 2) captive-reared individuals. However, there are no known, long-running, tag and recapture experiments through which genetic material has been collected, preserved and stored. In a review on crustacean ageing it was stated that "Reports on decapods marked for a decade and more are rare" (Vogt 2012). Efforts were also made to acquire lobster samples from individuals kept in sea-life centres. Public aquariums often house unusually large, and potentially very old, lobsters. However, in each case the lobster had been wild-caught at an already unknown age. Highlighting just how difficult it is to find an old lobster of known-age, a research team in Canada offered a $10,000 reward for one such individual but were unsuccessful (DFO 2017). In contrast, the red cherry shrimp system established for this PhD research allows for molecular markers of age to be tested across the entire lifespan of a crustacean under controlled environmental conditions. Investigating age-related rDNA methylation in red cherry shrimp (or another relatively short-lived crustacean: see Vogt et al. 2019 for a compilation of crustacean lifespan data) is therefore needed to better understand methylation status across the full lifespan of a crustacean.

Site-specific DNA methylation can also vary according to the environment (Martin & Fry 2018). To test the broader applicability for methylation-based ageing assays to lobster stocks, it is important to determine whether rDNA methylation varies among geographic areas with different environmental parameters (e.g. water temperature

and salinity). A study in which lobsters of the same age and genetic background (e.g. hatchery-sourced juveniles reared from the same mother) are placed in sea-based containers in areas with different environmental conditions, and then sampled in unison at different known ages is required to test this. This would help determine how much variation in methylation is caused by environmental differences; an important step if the assay is to be widely used across lobster populations in different areas. In addition, experimental manipulations involving the shrimp, whereby rates of age-related DNA methylation are compared among different conditions (e.g. aquaria with different water temperatures or food availability), could be a useful and tractable experimental approach.

Studies that assess the potential scalability of site-specific, methylation-based markers of age are also needed. For this research, the cost of bisulphite sequencing (including primer design, bisulphite conversion and bioinformatics) equated to ca. £100 per individual (provided by Zymo Research Corporation). This is likely to be too costly for estimating age structure in stock assessments. However, this cost could be reduced considerably by 1) using the same panel of amplicons developed in this study (therefore removing the need for primer design), 2) sequencing fewer amplicons (i.e. only sequencing amplicons containing loci with a significant relationship with age across 0–2 year-old lobsters), 3) analysing the sequencing data using freely-available software (e.g. Bismark: Krueger & Andrews 2011) and 4) performing some of the sample preparation in-house (i.e. not paying an external company to provide the full service). For example, commercially-available kits exist for bisulphite conversion of DNA (e.g. EZ DNA Methylation-Gold Kit by Zymo Research) for as little as ca. £2.50 per individual. The only other consideration would be the cost of next-generation sequencing (NGS), which is increasingly affordable but would depend on the number of individuals/amplicons to be measured, and desired coverage.

6.4 Mitochondrial DNA mutations in shrimp and lobsters (chapters 4 and 5, respectively)

The age-related accumulation of mtDNA mutations observed in humans and vertebrate models (Larsson 2010 and references therein; Arbeithuber et al. 2020) has been proposed as a predictor of human age (Meissner & Ritz-Timme 2010). Prior to this study, mtDNA mutation accumulation had been explored in just one wild animal (Jebb et al. 2018) and never in a crustacean. Whole lobster and shrimp mitogenomes were sequenced to high depth using NGS with the aim of comparing the number of mtDNA mutations across different age groups. However, quantifying these mtDNA mutations proved to be analytically extremely challenging. For both species, sequencing data were confounded by the presence of multiple, within-individual haplotypes. Features of some of these additional haplotypes (e.g. high levels of sequence divergence and evidence of loss-of-function) point towards co-amplification of nuclear DNA of mitochondrial origin (NUMTs). The lobster data were further complicated by mis-mapping of sequencing reads across a tandem duplication. Further research is needed to verify the existence of NUMTs. Such work could involve re-sequencing using isolated mtDNA, or mapping reads to nuclear genomes (when they become available). If individuals lack multiple haplotypes when sequencing isolated mtDNA, or if NGS reads map with greater affinity to nuclear compared to mtDNA, this would help to confirm NUMTs. If NUMTs are not confirmed, the next most-likely explanation is inheritance of multiple mitogenomes (either exclusively from the mother or a mixture of mother-father mtDNA). Studies involving the shrimp, in which mtDNA from offspring and parents is sequenced and compared could help to confirm or exclude inheritance-based mechanisms.

Identifying the source of additional mtDNA haplotypes has important implications. If NUMTs are the cause, this needs careful consideration for any study based on crustacean mtDNA, whether for developing an ageing assay or another purpose. For example, mtDNA is commonly used for investigating population genetic structure, resolving taxonomies and species identification in crustaceans (and other animals) (e.g. Song et al. 2008; Matzen da Silva et al. 2011). However, NUMTs appear to be

common in crustaceans (Song et al. 2008) and failing to account for them could lead to misinterpretations of population genetic structure, or species diversity or identity (e.g. Hlaing et al. 2009; Matzen da Silva et al. 2011; Raupach & Radulovici 2015). Comparing mitochondrial diversity among individuals is especially challenging if NUMT sequences show large inter-individual variation, which appeared to be the case for both lobsters and shrimp.

There was no evidence of mtDNA mutation accumulation in either species, although the issues outlined above meant it was not possible to investigate this phenomenon throughout the entire mitogenome of either species and rigorous filtering of variants was required for the lobster dataset. Perhaps the best evidence that mtDNA mutation accumulation is unlikely to be a useful predictor of crustacean age comes from the study on shrimp (chapter 4). Across 6 kb of mtDNA sequence (ca. 39% of the mitogenome), which was unaffected by (probable) NUMTs, the number of mtDNA mutations did not differ among seven known-age cohorts sampled 7–210 days post-fertilisation (spanning ca. 54% of the species' typical lifespan in captivity).

It is possible that any age-related changes in mtDNA mutation accumulation were missed, most likely as a result of tissue choice. Lobster DNA was sampled from claws or antennae and shrimp DNA from legs. All of these tissues undergo tissue regeneration, either following limb loss or during moults. It is not known how these processes influence mtDNA mutation accumulation but the shrimp system provides an ideal opportunity to explore this. For all of the shrimp sequenced in this study, the remaining specimen has been preserved in ethanol. Therefore, DNA from alternative tissues could be readily extracted from existing samples, sequenced using the primer pair designed in this study that did not co-amplify NUMTs (primers Shrimp_mtDNA_4F and 4R), and the number of mtDNA mutations compared among the same age groups analysed in this study. Of particular interest would be DNA from tissues that are less affected by tissue regeneration/moulting (e.g. different organs).

However, even if age-related heteroplasmy occurs and has been missed (as a result of excluding regions of the mitogenome or tissue choice), this study highlights the

difficulty of assessing heteroplasmy in mitochondrial genomes with duplications and/or nuclear pseudogenes. This complexity alone suggests that heteroplasmy is unlikely to be an easy, viable marker of age in crustaceans for stock assessments.

6.5 Overall conclusion

This research has taken the quest to 'age the unageable' several steps forward by exploring several previously-unexplored potential molecular markers of crustacean age. Firstly, it can be concluded that neither genome-wide methylation measured using an ELISA kit, or the number of mtDNA mutations quantified using short-read NGS, provide easy, viable markers of crustacean age. On the other hand, site-specific methylation of rDNA holds considerable promise for estimating crustacean age. Further studies are now urgently needed to confirm the reliability and viability of rDNA methylation for ageing crustaceans, as well as to better understand the importance of tissue type used and the environment in measures of methylation. Such research could have significant impacts for the management and conservation of economically-important species.

References

Agnalt, A., Kristiansen, T.S. & Jørstad, K.E. (2007) Growth, reproductive cycle, and movement of berried European lobsters (*Homarus gammarus*) in a local stock off southwestern Norway. *ICES Journal of Marine Science* **64**, 288–297.

Agnalt, A., Farestveit, E., Gundersen, K., Jørstad, K.E. & Kristiansen, T.S. (2009) Population characteristics of the world's northernmost stocks of European lobster (*Homarus gammarus*) in Tysfjord and Nordfolda, northern Norway. *New Zealand Journal of Marine and Freshwater Research* **43**, 47–57.

Agrelius, T., Dudycha, J.L. & Morris, J.T. (2018) Global DNA cytosine methylation variation in *Spartina alterniflora* at North Inlet, SC. *PLOS ONE* **13**, e0203230.

Ahyong, S.T., Lowry, J.K., Alonso, M., Bamber, R.N., Boxshall, G.A., Castro, P., Gerken, S., Karaman, G.S., Goy, J.W., Jones, D.S., Meland, K., Rogers, C.R. & Svavarsson, J. (2011) Subphylum Crustacea Brünnich, 1772. Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. *Zootaxa* **3148**, 165–191.

Alfsnes, K., Leinaas, H.P. & Hessen, D.O. (2017) Genome size in arthropods; different roles of phylogeny, habitat and life history in insects and crustaceans. *Ecology and Evolution* **7**, 5939–5947.

Aljanabi, S. & Martinez, I. (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR- based techniques. *Nucleic Acids Research* **25**, 4692–4693.

Anastasiadi, D. & Piferrer, F. (2020) A clockwork fish: Age prediction using DNA methylation-based biomarkers in the European seabass. *Molecular Ecology Resources* **20**, 387–397.

Anderson, S.C., Flemming, J.M., Watson, R. & Lotze, H.K. (2011) Rapid global expansion of invertebrate fisheries: trends, drivers, and ecosystem effects. *PLOS ONE* **6**, e14735.

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

Arbeithuber, B., Hester, J., Cremona, M.A., Stoler, N., Zaidi, A., Higgins, B., Anthony, K., Chiaromonte, F., Diaz, F.J. & Makova, K.D. (2020) Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues. *PLOS Biology* **18**, e3000745.

Baldo, L., de Queiroz, A., Hedin, M., Hayashi, C.Y. & Gatesy, J. (2011) Nuclear-mitochondrial sequences as witnesses of past interbreeding and population diversity in the jumping bristletail *Mesomachilis*. *Molecular Biology and Evolution* **28**, 195–210.

Bannister, R.C.A. & Addison, J.T. (1998) Enhancing lobster stocks: a review of recent European methods, results, and future prospects. *Bulletin of Marine Science* **62**, 369–387.

Barazzoni, R., Short, K.R. & Nair, K.S. (2000) Effects of aging on mitochondrial DNA copy number and cytochrome *c* oxidase gene expression in rat skeletal muscle, liver, and heart. *The Journal of Biological Chemistry* **275**, 3343–3347.

Barrett, A., Arbeithuber, B., Zaidi, A., Wilton, P., Paul, I.M., Nielsen, R. & Makova, K.D. (2019) Pronounced somatic bottleneck in mitochondrial DNA of human hair. *Philosophical Transaction of the Royal Society B: Biological Sciences* **375**, 20190175.

Barrett, E.L.B., Burke, T.A., Hammers, M., Komdeur, J. & Richardson, D.S. (2013) Telomere length and dynamics predict mortality in a wild longitudinal study. *Molecular Ecology* **22**, 249–259

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48.

Beal, A.P., Kiszka, J.J., Wells, R.S. & Eirin-Lopez, J.M. (2019) The bottlenose dolphin epigenetic aging tool (BEAT): A molecular age estimation tool for small cetaceans. *Frontiers in Marine Science* **6**, 561.

Becker, C., Dick, J.T.A., Cunningham, E.M., Schmitt, C. & Sigwart, J.D. (2018) The crustacean cuticle does not record chronological age: new evidence from the gastric mill ossicles. *Arthropod Structure & Development* **47**, 498–512.

Beddington, J.R., Agnew, D.J. & Clark, C.W. (2007) Current problems in the management of marine fisheries. *Science* **316**, 1713–1716.

Béné, C., Arthur, R., Norbury, H., Allison, E.H., Beveridge, M., Bush, S., Campling, L., Leschen, W., Little, D., Squires, D., Thilsted, S.H., Troell, M. & Williams, M. (2016) Contribution of fisheries and aquaculture to food security and poverty reduction: assessing the current evidence. *World Development* **79**, 177–196.

Bensasson, D., Zhang, D.X., Hartl, D.L. & Hewitt, G.M. (2001) Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends in Ecology and Evolution* **16**, 314–321.

Berdyshev, G.D., Korotaev, G.K., Boiarskikh, G.V. & Vaniushin, B.F. (1967) Nucleotide composition of DNA and RNA from somatic tissues of humpback salmon and its changes during spawning. *Biokhimiia* **32**, 988–993.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M. & Stadler, P.F. (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**, 313–319.

Bize, P., Criscuolo, F., Metcalfe, N.B., Nasir, L. & Monaghan, P. (2009) Telomere dynamics rather than age predict life expectancy in the wild. *Proceedings of the Royal Society of London Series B: Biological Sciences* **276**, 1679–1683.

Bocklandt, S., Lin, W., Sehl, M.E., Sánchez, F.J., Sinsheimer, J.S., Horvath, S. & Vilain, E. (2011) Epigenetic predictor of age. *PLOS ONE* **6**, e14821.

Bondad-Reantaso, M.G., Subasinghe, R.P., Josupeit, H., Cai, J. & Zhou, X. (2012) The role of crustacean fisheries and aquaculture in global food security: past, present and future. *Journal of Invertebrate Pathology* **110**, 158–165.

Boudreau, S.A. & Worm, B. (2012) Ecological role of large benthic decapods in marine ecosystems: a review. *Marine Ecology Progress Series* **469**, 195–213.

Bower, J.E., Cooper, R.D. & Beebe, N.W. (2009) Internal repetition and intraindividual variation in the rDNA ITS1 of the *Anopheles punctulatus* group (Diptera: Culicidae): multiple units and rates of turnover. *Journal of Molecular Evolution* **68**, 66–79.

Brunel, T. & Piet, G.J.J. (2013) Is age structure a relevant criterion for the health of fish stocks? *Journal of Marine Science* **70**, 270–283.

Brunk, U.T. & Terman, A. (2002) Lipofuscin: mechanisms of age-related accumulation and influence on cell function. *Free Radical Biology & Medicine* **33**, 611–619.

Butler, M.J. (2017) Collecting and processing lobsters. *Journal of Crustacean Biology* **37**, 340–346.

Campana, S.E. (2001) Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology* **59**, 197–242.

Cattaert, D. & Edwards, D.H. (2017) *Control of Locomotion in Crustaceans*. Oxford University Press.

CEFAS (2018) *Lobster (Homarus gammarus) CEFAS stock status report 2017*. Centre for Environment, Fisheries and Aquaculture Science.

Chen, S., McKinney, G.J., Nichols, K.M. & Sepúlveda, M.S. (2014) *In silico* prediction and *in vivo* validation of *Daphnia pulex* microRNAs. *PLOS ONE* **9**, e83708.

Cheng, K.C., Cahill, D.S., Kasai, H., Nishimura, S. & Loeb, L.A. (1992) 8-hydroxyguanine, an abundant form of oxidative DNA damage, causes G → T and A → C substitutions. *Journal of Biological Chemistry* **267**, 166–172.

Choi, Y. & Chan, A.P. (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747.

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. & Chan, A.P. (2012) Predicting the functional effect of amino acid substitutions and indels. *PLOS ONE* **7**, e46688.

Christensen, B.C., Houseman, E.A., Marsit, C.J., Zheng, S., Wrensch, M.R., Wiemels, J.L., Nelson, H.H., Karagas, M.R., Padbury, J.F., Bueno, R., Sugarbaker, D.J., Yeh, R., Wiencke, J.K. & Kelsey, K.T. (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLOS Genetics* **5**, e1000602.

Chu, K.H., Li, C.P. & Ho, H.Y. (2001) The first internal transcribed spacer (ITS-1) of ribosomal DNA as a molecular marker for phylogenetic and population analyses in Crustacea. *Marine Biotechnology* **3**, 355–361.

Chu, K.H., Li, C.P., Tam, Y.K. & Lavery, S. (2003) Application of mitochondrial control region in population genetic studies of the shrimp *Penaeus*. *Molecular Ecology Notes* **3**, 120–122.

Clavelle, T., Lester, S.E., Gentry, R. & Froehlich, H.E. (2019) Interactions and management for the future of marine aquaculture and capture fisheries. *Fish and Fisheries* **20**, 368–388.

Cook, R.D. & Weisberg, S. (1982) *Residuals and Influence in Regression*. Chapman & Hall.

Crowley, C.E., Gandy, R.L., Daly K.L. & van Vleet, E.S. (2014) Problems associated with a lipofuscin extraction method used to age blue crabs *Callinectes sapidus* cultured in Florida, USA. *Aquatic Biology* **21**, 85–92.

Daan, N., Gislason, H., Pope, J.G. & Rice, J.C. (2011) Apocalypse in world fisheries? The reports of their death are greatly exaggerated. *ICES Journal of Marine Science* **68**, 1375–1378.

Daniels, C.L., Wills, B., Ruiz-Perez, M., Miles, E., Wilson, R.W. & Boothroyd, D. (2015) Development of sea based container culture for rearing European lobster (*Homarus gammarus*) around South West England. *Aquaculture* **448**, 186–195.

De Paoli-Iseppi, R., Deagle, B.E., McMahon, C.R., Hindell, M.A., Dickinson, J.L. & Jarman, S.N. (2017) Measuring animal age with DNA methylation: from humans to wild animals. *Frontiers in Genetics* **8**, 106.

De Paoli-Iseppi, R., Deagle, B.E., Polanowski, A.M., McMahon, C.R., Dickinson, J.L., Hindell, M.A. & Jarman, S.N. (2019) Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. *Molecular Ecology Resources* **19**, 411–425.

DFO (2017) Growth bands in lobsters, crabs and shrimp reveal age. http://www.dfo-mpo.gc.ca/science/publications/article/2012/10-30-12-eng.html

Dierckxsens, N., Mardulyn, P. & Smits, G. (2016) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* **45**, e18.

Dierckxsens, N., Mardulyn, P. & Smits, G. (2020) Unraveling heteroplasmy patterns with NOVOPlasty. *NAR Genomics and Bioinformatics* **2**, lqz011.

Ding, Y.R., Li, B., Zhang, Y.J., Mao, Q.M. & Chen, B. (2018) Complete mitogenome of *Anopheles sinensis* and mitochondrial insertion segments in the nuclear genomes of 19 mosquito species. *PLOS ONE* **13**, e0204667.

Diniz, F.M., Maclean, N., Ogawa, M., Cintra, I.H.A. & Bentzen, P. (2005) The hypervariable domain of the mitochondrial control region in Atlantic spiny lobsters and its potential as a marker for investigating phylogeographic structuring. *Marine Biotechnology* **7**, 462–473.

Duan, M., Tu, J. & Lu, Z. (2018) Recent advances in detecting mitochondrial DNA heteroplasmic variations. *Molecules* **23**, 323.

Dunshea, G., Duffield, D., Gales, N., Hindell, M., Wells, R.S. & Jarman, S.N. (2011) Telomeres as age markers in vertebrate molecular ecology. *Molecular Ecology Resources* **11**, 225–235.

Dyomin, A.G., Koshel, E.I., Kiselev, A.M., Saifitdinova, A.F., Galkina, S.A., Fukagawa, T., Kostareva, A.A. & Gaginskaya, E.R. (2016) Chicken rRNA Gene Cluster Structure. *PLOS ONE* **11**, e0157464.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797.

Ekblom, R. & Galindo, J. (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1–15.

Ellis, C.D., Hodgson, D.J., Daniels, C.L., Boothroyd, D.P., Bannister, R.C.A. & Griffiths, A.G.F. (2015) European lobster stocking requires comprehensive impact assessment to determine fishery benefits. *ICES Journal of Marine Science* **72**, i35–i48.

Ellis, C.D., Jenkins, T.L., Svanberg, L., Eriksson, S.P. & Stevens, J.R. (2020) Crossing the pond: genetic assignment detects lobster hybridisation. *Scientific Reports* **10**, 7781.

Elmore, L.W., Norris, M.W., Sircar, S., Bright, A.T., McChesney, P.A., Winn, R.N. & Holt S.E. (2008) Upregulation of telomerase function during tissue regeneration. *Experimental Biology and Medicine* **233**, 958–967.

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

Fang, X., Thornton, C., Scheffler, B.E. & Willett, K.L. (2013) Benzo [a] pyrene decreases global and gene specific DNA methylation during zebrafish development. *Environmental Toxicology* **36,** 40–50.

FAO (1998) *Introduction to Tropical Fish Stock* Assessment. FAO Fisheries Department.

FAO (2002) *A Fishery Manager's Guidebook: Management Measures and their Application*. FAO Fisheries Department.

FAO (2020) *The State of World Fisheries and Aquaculture 2020: Sustainability in Action*. FAO Fisheries Department.

Feng, C.Z., Ding, X., Li, Y., Chen, S., Yang, F. & Yang, W.J. (2007) The DNA methyltransferase-2 gene in the prawn *Macrobrachium rosenbergii*: characteristics and expression patterns during ovarian and embryonic development. *Zoological Science* **24**, 1059–1065.

Field, A.E., Robertson, N.A., Wang, T., Havas, A., Ideker, T. & Adams, P.D. (2018) DNA methylation clocks in aging: categories, causes, and consequences. *Molecular Cell* **71**, 882–895.

Fox, J. & Weisberg, S. (2019) *An R Companion to Applied Regression*. SAGE.

Friedman, J., Hastie, T., Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**, 1–22.

Fuke, C., Shimabukuro, M., Petronis, A., Sugimoto, J., Oda, T., Miura, K., Miyazaki, T., Ogura, C., Okazaki, Y. & Jinno, Y. (2004) Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Annals of Human Genetics* **68**, 196–204.

Gan, H.M., Gradnjean, F., Jenkins, T.L. & Austin, C.M. (2019) Absence of evidence is not evidence of absence: Nanopore sequencing and complete assembly of the European lobster (*Homarus gammarus*) mitogenome uncovers the missing *nad2* and a new major gene cluster duplication. *BMC Genomics* **20**, 335.

Garrison, E. & Marth, G. (2012) Haplotype-based variant detection from short-read sequencing. http://arxiv.org/abs/1207.3907

Gibbons, W.J. & Andrews, K.M. (2004) PIT tagging: simple technology at its best. *BioScience* **54**, 447–454.

Godwin, R.M., Frusher, S., Montgomery, S.S. & Ovenden, J. (2011) Telomere length analysis in crustacean species: *Metapenaeus macleayi, Sagmariasus verreauxi,* and *Jasus edwardsii*. *ICES Journal of Marine Science* **68**, 2053–2058.

Gomes, M.V.M., Toffoli, L.V., Arruda, D.W., Soldera, L.M., Pelosi, G.G., Neves-Souza, R.D., Freitas, E.R., Castro, D.T. & Marquez, A.S. (2012) Age-related changes in the global DNA methylation profile of leukocytes are linked to nutrition but are not associated with the MTHFR C677T genotype or to functional capacities. *PLOS ONE* **7**, e52570.

González, M.M., Ramos, A., Aluja, M.P. & Santos, C. (2020) Sensitivity of mitochondrial DNA heteroplasmy detection using next generation sequencing. *Mitochondrion* **50**, 88–93.

Götze, S. & Saborowski, R. (2011) Proteasomal activities in the claw muscle tissue of European lobster, *Homarus gammarus*, during larval development. *Journal of Comparative Physiology B: Biochemical, Systems, and Environmental Physiology* **181**, 861–871.

Greiner, S., Lehwark, P. & Bock, R. (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* **47**, W59–W64.

Grönniger, E., Weber, B., Heil, O., Peters, N., Stäb, F., Wenck, H., Korn, B., Winnefeld, M. & Lyko, F. (2010) Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLOS Genetics* **6**, e1000971.

Gryzińska, M., Blaszczak, E., Strachecka, A. & Jezewska-Witkowska, G. (2013) Analysis of age-related global DNA methylation in chicken. *Biochemical Genetics* **51**, 554–563.

Gryzińska, M., Jakubczak, A., Listos, P., Dudko, P., Abramowicz, K. & Jeżewska-Witkowska, G. (2016) Association between body weight and age of dogs and global DNA methylation. *Praca oryginalna* **72**, 64–67.

Guarasci, F., D'Aquila, P., Mandalà, M., Garasto, S., Lattanzio, F., Corsonello, A., Passarino, G. & Bellizzi, D. (2018) Aging and nutrition induce tissue-specific changes on global DNA methylation status in rats. *Mechanisms of Ageing and Development* **174**, 47–54.

Guevara, E.E., Lawler, R.R., Staes, N., White, C.M., Sherwood, C.C., Ely, J.J., Hopkins, W.D. & Bradley, B.J. (2020) Age-associated epigenetic change in chimpanzees and humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **375**, 20190616.

Han, Y., Eipel, M., Franzen, J., Sakk, V., Dethmers-Ausema, B., Yndriago, L., Izeta, A., de Hann, G., Geiger, H. & Wagner, W. (2018). Epigenetic age-predictor for mice based on three CpG sites. *eLife* **7**, e37462.

Harman, D (1972) The biologic clock: the mitochondria? *Journal of the American Geriatrics Society* **20**, 145–147.

Hartmann, N., Reichwald, K., Wittig, I., Dröse, S., Schmeisser, S., Lück, C., Hahn, C., Graf, M., Gausmann, U., Terzibasi, E., Cellerino, A., Ristow, M., Brandt, U., Platzer, M. & Englert, C. (2011) Mitochondrial DNA copy number and function decrease with age in the short-lived fish *Nothobranchius furzeri*. *Aging Cell* **10**, 824–831.

Hartnoll, R.G. (2001) Growth in Crustacea – twenty years on. *Hydrobiologia* **449**, 111–122.

Haussmann, M.F. & Vleck, C.M. (2002) Telomere length provides a new technique for aging animals. *Oecologia* **130**, 325–328.

Herbstman, J.B., Wang, S., Perera, F.P., Lederman, S.A., Vishnevetsky, J., Rundle, A.G., Hoepner, L.A., Qu, L. & Tang, D. (2013) Predictors and consequences of global DNA methylation in cord blood and at three Years. *PLOS ONE* **8**, e72824.

Hilborn, R. (1992) Current and future trends in fisheries stock assessment and management. *South African Journal of Marine Science* **12**, 975–988.

Hilborn, R. (2007) Reinterpreting the state of fisheries and their management. *Ecosystems* **10**, 1362–1369.

Hilborn, R. & Walters, C.J. (1992) *Quantitative Fisheries Stock Assessment. Choice, Dynamics & Uncertainty*. Chapman & Hall.

Hilborn, R., Amoroso, R.O., Anderson, C.M., Baum, J.K., Branch, T.A., Costello, C., de Moor, C.L., Faraj, A., Hively, D., Jensen, O.P., Kurota, H., Little, L.R., Mace, P., McClanahan, T., Melnychuk, M.C., Minto, C., Osio, G.C., Parma, A.M., Pons, M., Segurado, S., Szuwalski, C.S., Wilson, J.R. & Ye, Y. (2020) Effective fisheries management instrumental in improving fish stock status. *Proceedings of the National Academy of Sciences* **117**, 2218–2224.

Hiona, A., Sanz, A., Kujoth, G.C., Pamplona, R., Seo, A.Y., Hofer, T., Someya, S., Miyakawa, T., Nakayama, C., Samhan-Arias, A.K., Servais, S., Barger, J.L., Portero-Otín, M., Tanokura, M., Prolla, T.A. & Leeuwenburgh, C. (2010) Mitochondrial DNA mutations induce mitochondrial dysfunction, apoptosis and sarcopenia in skeletal muscle of mitochondrial DNA mutator mice. *PLOS ONE* **5**, e11468.

Hlaing, T., Tun-Lin, W., Somboon, P., Socheat, D., Setha, T., Min, S., Chang, M.S. & Walton, C. (2009) Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. *BMC genetics* **10**, 11.

Holland, B.S. & Copenhaver, M.D. (1987) An improved sequentially rejective Bonferroni test procedure. *Biometrics* **43**, 417.

Hovrath, D. (2013) DNA methylation age of human tissues and cell types. *Genome Biology* **14**, R115.

Huan, T., Chen, G., Liu, C., Bhattacharya, A., Rong, J., Chen, B.H., Seshadri, S., Tanriverdi, K., Freedman, J.E., Larson, M.G., Murabito, J.M., & Levy, D. (2018)

Age-associated microRNA expression in human peripheral blood is associated with all-cause mortality and age-related traits. *Aging Cell* **17**, e12687.

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

Ito, H., Udono, T., Hirata, S. & Inoue-Murayama, M. (2018) Estimation of chimpanzee age based on DNA methylation. *Scientific Reports* **8**, 9998.

Itsara, L.S., Kennedy, S.R., Fox, E.J., Yu, S., Hewitt, J.J., Sanchez-Contreras, M., Cardozo-Pelaez, F. & Pallanck, L.J. (2014) Oxidative stress is not a major contributor to somatic mitochondrial DNA mutations. *PLOS Genetics* **10**, e1003974.

Jarman, S.N., Polanowski, A.M., Faux, C.E., Robbins, J., De Paoli-Iseppi, R., Bravington, M. & Deagle, B.E. (2015) Molecular biomarkers for chronological age in animal ecology. *Molecular Ecology* **24**, 4826–4847.

Jebb, D., Foley, N.M., Whelan, C.V., Touzalin, F., Puechmaille, S.J. & Teeling, E.C. (2018) Population level mitogenomics of the long-lived greater mouse-eared bat, *Myotis myotis*, reveals dynamic heteroplasmy and challenges the free radical theory of ageing. *Scientific Reports* **8**, 13634.

Jenkins, T.L., Ellis, C.D., Triantafyllidis, A. & Stevens, J.R. (2019) Single nucleotide polymorphisms reveal a genetic cline across the north-east Atlantic and enable powerful population assignment in the European lobster. *Evolutionary Applications* **12**, 1881–1899.

Jenkins, T.L., Ellis, C.D., Durieux, E.D.H., Filippi, J., Bracconi, J. & Stevens, J.R. (2020) Historical translocations and stocking alter the genetic structure of a Mediterranean lobster fishery. *Ecology and Evolution* **10**, 5631–5636.

Jones, M.J., Goodman, S.J. & Kobor, M.S. (2015) DNA methylation and healthy human aging. *Aging Cell* **14**, 924–932.

Ju, S., Secor, D.H. & Harvey, H.R. (1999) Use of extractable lipofuscin for age determination of blue crab *Callinectes sapidus*. *Marine Ecology Progress Series* **185**, 171–179.

Just, R.S., Irwin, J.A. & Parson, W. (2015) Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Science International: Genetics* **18**, 131–139.

Kennedy, S.R., Salk, J.J., Schmitt, M.W. & Loeb, L.A. (2013) Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. *PLOS Genetics* **9**, e1003794.

Kenny, N.J., Sin, Y.W., Shen, X., Zhe, Q., Wang, W., Chan, T.F., Tobe, S.S., Shimeld. S.M., Chu, K.H. & Hui, J.H.L. (2014) Genomic sequence and experimental tractability of a new decapod shrimp model, *Neocaridina denticulata*. *Marine Drugs* **12**, 1419–1437.

Kilada, R. & Driscoll, J.G. (2017) Age determination in crustaceans: a review. *Hydrobiologia* **799**, 21–36.

Kilada, R.W., Campana, S.E. & Roddick, D. (2009) Growth and sexual maturity of the northern propellerclam (*Cyrtodaria siliqua*) in eastern Canada, with bomb radiocarbon age validation. *Marine Biology* **156**, 1029–1037.

Kilada, R., Sainte-Marie, B., Rochette, R., Davis, N., Vanier, C. & Campana, S. (2012) Direct determination of age in shrimps, crabs, and lobsters. *Canadian Journal of Fisheries and Aquatic Sciences* **69**, 1728–1733.

Kinser, H.E. & Pincus, Z. (2020) MicroRNAs as modulators of longevity and the aging process. *Human Genetics* **139**, 291–308.

Klapper, W., Kühne, K., Singh, K.K., Heidorn, K., Parwaresch, R. & Krupp, G. (1998) Longevity of lobsters is linked to ubiquitous telomerase expression. *FEBS Letters* **439**, 143–146.

Kleiber, C. & Zeileis, A. (2008) *Applied Econometrics with R*. Springer-Verlag.

Koboldt, D., Zhang, Q., Larson, D., Shen, D., McLellan, M., Lin, L., Miller, C., Mardis, E., Ding, L. & Wilson, R. (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* **22**, 568–676.

Krueger, F. & Andrews, S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572.

Krzywinski, M.I., Schein, J.E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. & Marra, M.A. (2009) Circos: An information aesthetic for comparative genomics. *Genome Research* **19**, 1639–1645.

Kurdyukov, S. & Bullock, M. (2016) DNA methylation analysis: choosing the right method. *Biology* **5**, 3.

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

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics* **30**, 3276–3278.

Larsson, N. (2010) Somatic mitochondrial DNA mutations in mammalian aging. *Annual Review of Biochemistry* **79**, 683–706.

Lawless, C., Greaves, L., Reeve, A.K., Turnbull, D.M. & Vincent, A.E. (2020) The rise and rise of mitochondrial DNA mutations. *Open Biology* **10**, 200061.

Lenth, R.V. (2016) Least-squares means: the R package lsmeans. *Journal of Statistical Software* **69**, 1–33.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Processing Subgroup. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Li, M., Schonberg, A., Schaefer, M., Schroeder, R., Nasidze, I. & Stoneking, M. (2010) Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *American Journal of Human Genetics* **87**, 237–249.

Li, M., Schroeder, R., Ko, A. & Stoneking, M. (2012) Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic Acids Research* **40**, e137.

Li, M., Schröder, R., Ni, S., Madea, B. & Stoneking, M. (2015) Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proceedings of the National Academy of Sciences* **112**, 2491–2496.

Lian, S., He, Y., Li, X., Zhao, B., Hou, R., Hu, X., Zhang, L. & Bao, Z. (2015) Changes in global DNA methylation intensity and *DNMT1* transcription during the aging process of scallop *Chlamys farreri. Journal of Ocean University of China* **14**, 685–690.

Lisanti, S. (2013) Changes in DNA methylation patterns in mammals with senescence, ageing and energy restriction. PhD thesis, University of Newcastle Upon Tyne.

Lowe, R., Barton, C., Jenkins, C.A., Ernst, C., Forman, O., Fernandez-Twinn, D.S., Bock, C., Rossiter, S.J., Faulkes, C.G., Ozanne, S.E., Walter, L., Odom, D.T., Mellersh, C. & Rakyan, V.K. (2018) Ageing-associated DNA methylation dynamics are a molecular readout of lifespan variation among mammalian species. *Genome Biology* **19**, 22.

Lunter G. & Goodson, M. (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research* **21**, 936–939.

Ma, H., Lee, Y., Hayama, T., Dyken, C.V., Marti-Gutierrez, N., Li, Y., Ahmed, R., Koski, A., Kang, E., Darby, H., Gonmanee, T., Park, Y., Wolf, D.P., Kim, C.J. & Mitalipov, S. (2018) Germline and somatic mtDNA mutations in mouse aging. *PLOS ONE* **13**, e0201304.

Mao, X., Dong, J., Hua, P., He, G., Zhang, S. & Rossiter, S.J. (2014) Heteroplasmy and ancient translocation of mitochondrial DNA to the nucleus in the chinese horseshoe bat (Rhinolophus sinicus) complex. PLOS ONE 9, e98035.

Maegawa, S., Hinkal, G., Kim, H.S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L.A. & Issa, J.J. (2010) Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Research* **20**, 332–340.

Martin, E.M. & Fry, R.C. (2018) Environmental influences on the epigenome: exposure-associated DNA methylation in human populations. *Annual Review of Public Health* **39**, 309–333.

Mastrantonio, V., Urbanelli, S. & Porretta, D. (2019) Ancient hybridization and mtDNA introgression behind current paternal leakage and heteroplasmy in hybrid zones. *Scientific Reports* **9**, 19177.

Matzen da Silva, J., Creer, S., dos Santos, A., Costa, A.C., Cunha, M.R., Costa, F.O. & Carvalho, G.R. (2011) Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). *PLOS ONE* **6**, e19449.

Maude, H., Davidson, M., Charitakis, N., Diaz, L., Bowers, W.H.T., Gradovich, E., Andrew, T. & Huntley, D. (2019) NUMT confounding biases mitochondrial

heteroplasmy calls in favour of the reference allele. *Frontiers in Cell and Developmental Biology* **7**, 201.

Maunder, M.N. & Piner, K.R. (2014) Contemporary fisheries stock assessment: many issues still remain. *ICES Journal of Marine Science* **72**, 7–18.

Maunder, M.N. & Punt, A.E. (2013) A review of integrated analysis in fisheries stock assessment. *Fisheries Research* **142**, 61-74.

Meissner, C. & Ritz-Timme, S. (2010) Molecular pathology and age estimation. *Forensic Science International* **203**, 34–43.

Meissner, C., von Wurmb-Schwark, N., Schimansky, B. & Oehmichen, M. (1999) Estimation of age at death based on quantitation of the 4977-bp deletion of human mitochondrial DNA in skeletal muscle. *Forensic Science International* **105**, 115–124.

Mengel-From, J., Thinggaard, T., Dalgård, C., Kyvik, K.O., Christensen, K. & Christiansen, L. (2014) Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. *Human Genetics* **133**, 1149–1159.

Mhanni, A.A. & McGowan, R.A. (2004) Global changes in genomic methylation levels during early development of the zebrafish embryo. *Development Genes and Evolution* **214**, 412–417.

Michikawa, Y., Mazzucchelli, F., Bresolin, N., Scarlato, G. & Attardi, G. (1999) Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. *Science* **286**, 774–779.

Miller, F.J., Rosenfeldt, F.L., Zhang, C.F., Linnane, A.W. & Nagley, P. (2003) Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. *Nucleic Acids Research* **31**, e61.

Milne, I., Bayer, M., Stephen, G., Cardle, L. & Marshall, D. (2016) Tablet: visualizing next-generation sequence assemblies and mappings. *Methods in Molecular Biology* **1374**, 253–268.

MMO (2019) *UK Sea Fisheries Annual Statistics Report 2018*. Marine Management Organisation.

Monaghan, P. & Haussmann, M. (2006) Do telomere dynamics link lifestyle and lifespan? *Trends in Ecology and Evolution* **21**, 47–53.

Montier, L.L.C., Deng, J.J. & Bai, Y. (2009) Number matters: control of mammalian mitochondrial DNA copy number. *Journal of Genetics and Genomics* **36**, 125–131.

Myers, R.A. & Worm, B. (2003) Rapid worldwide depletion of predatory fish communities. *Nature* **423**, 280–283.

Mykles, D.L. & Hui, J.H.L. (2015) *Neocaridina denticulata*: A decapod crustacean model for functional genomics. *Integrative and Comparative Biology* **55**, 891–897.

Naue, J., Hörer, S., Sänger, T., Strobl, C., Hatzer-Grubwieser, P., Parson, W. & Lutz-Bonengel, S. (2015) Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. *Mitochondrion* **20**, 82–94.

Nur, F.A.H. & Christianus, A. (2013) Breeding and life cycle of *Neocaridina denticulate sinensis* (Kemp, 1918). *Asian Journal of Animal and Veterinary Advances* **8**, 108–115.

Nussey, D.H., Froy, H., Lemaitre, J-F., Gaillard, J-M. & Austad, S.N. (2013) Senescence in natural populations of animals: Widespread evidence and its implications for bio-gerontology. *Ageing Research Reviews* **12**, 214–225.

O'Donovan, V. & Tully, O. (1996) Lipofuscin (age pigment) as an index of crustacean age: correlation with age, temperature and body size in cultured juvenile *Homarus gammarus* L. *Journal of Experimental Marine Biology and Ecology* **207**, 1–14.

Okonechnikov, K., Conesa, A. & García-Alcalde, F. (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294.

Panella, G. (1971) Fish otoliths: daily growth layers and periodical patterns. *Science* **173**, 1124–1127.

Paradis, E. (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**, 419–420.

Park, J., Kim, Y., Kwon, W., Xi, H. & Park, J. (2019) The complete mitochondrial genome of *Neocaridina heteropoda koreana* Kubo, 1938 (Decapoda: Atyidae). *Mitochondrial DNA Part B* **4**, 2332–2334.

Patel, R.K. & Jain, M. (2012) NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLOS ONE* **7**, e30619.

Pauly, D. & Morgan, G.R. (1987) *Length-Based Methods in Fisheries Research*. International Centre for Living Aquatic Resources Management.

Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. (1998) Fishing down marine food webs. *Science* **279**, 860–863.

Pauly, D., Christensen, V., Guénette, S., Pitcher, T.J., Sumaila, U.R., Walters, C.J., Watson, R. & Zeller, D. (2002) Towards sustainability in world fisheries. *Nature* **418**, 689–695.

Payne, B.A.I. & Chinnery, P.F. (2015) Mitochondrial dysfunction in aging: Much progress but many unresolved questions. *Biochimica et Biophysica Acta* **1847**, 1347–1353.

Pérez-Barbería, F., Duff, E., Brewer, M., & Guinness, F. (2014) Evaluation of methods to age Scottish red deer: The balance between accuracy and practicality. *Journal of Zoology* **294**, 180–189.

Phillips, B.F. (2006) *Lobsters: Biology, Management, Aquaculture and Fisheries*. Blackwell Publishing.

Picard, R.R. & Dennis Cook, R. (1984) Cross-validation of regression models. *Journal of the American Statistical Association* **79**, 575–583.

Pincus, Z., Smith-Vikos, T. & Slack, F.J. (2011) MicroRNA predictors of longevity in *Caenorhabditis elegans*. *PLOS Genetics* **7**, e1002306.

Polanowski, A.M., Robbins, J., Chandler, D. & Jarman, S.N. (2014) Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources* **14**, 976–987.

Polovina, E., Parakatselaki, M. & Ladoukakis, E.D. (2020) Paternal leakage of mitochondrial DNA and maternal inheritance of heteroplasmy in *Drosophila* hybrids. *Scientific Reports* **10**, 2599.

Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L.A., Barrio-Amorós, C.L., Salazar-Valenzuela, D. & Prost, S. (2018) Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *Gigascience* **7**, giy033.

Pope, K.L., Lochmann, S.E. & Young, M.K. (2010) *Methods for Assessing Fish Populations* in *Inland Fisheries Management in North* America. American Fisheries Society.

Price, E.M.W. (2009) DNA methylation in human development: methodologies and analytics for genome-wide studies. PhD thesis, University of British Columbia.

Punt, A.E., Huang, T.C. & Maunder, M.N. (2013) Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. *ICES Journal of Marine Science* **70**, 16–33.

Quinn, T.J. & Collie, J.S. (2005) Sustainability in single-species population models. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 147–162.

R Core Team (2017a) *stats: The R Stats Package*. The R Foundation.

R Core Team (2017b) *R: A Language and Environment for Statistical Computing*. The R Foundation.

R Core Team (2020) *Graphics: The R Graphics Package*. The R Foundation.

Raupach, M.J. & Radulovici, A.E. (2015) Looking back on a decade of barcoding crustaceans. *ZooKeys* **539**, 53–81.

Richly, E. & Leister, D. (2004) NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution* **21**, 1081–1084.

Rodríguez-Pena, E., Verísimo, P., Fernández, L., González-Tizón, A., Bárcena, C. & Martínez-Lage, A. (2020) High incidence of heteroplasmy in the mtDNA of a natural population of the spider crab *Maja brachydactyla*. *PLOS ONE* **15**, e0230243.

Rooney, J.P., Ryde, I.T., Sanders, L.H., Howlett, E.H., Colton, M.D., Germ, K.E., Mayer, G.D., Greenamyre, J.T. & Meyer, J.N. (2015) PCR based determination of mitochondrial DNA copy number in multiple species. *Methods in Molecular Biology* **1241**, 23–38.

Rossignol, R., Faustin, B., Rocher, C., Malgat, M., Mazat, J. & Letellier, T. (2003) Mitochondrial threshold effects. *Biochemical Journal* **370(Pt 3)**, 751–762.

Rötzer, M.A.I.N. & Haug, J.T. (2015) Larval development of the European lobster and how small heterochronic shifts lead to a more pronounced metamorphosis. *International Journal of Zoology* **2015**, 345172.

Rozen, S. & Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology* **132**, 365–386.

RStudio Team (2016) *RStudio: Integrated Development for R*. RStudio, PBC.

Samuels, D.C., Li, C., Li, B., Song, Z., Torstenson, E., Clay, H.B., Rokas, A., Thornton-Wells, T.A., Moore, J.H., Hughes, T.M., Hoffman, R.D., Haines, J.L., Murdock, D.G., Mortlock, D.P. & Williams, S.M. (2013) Recurrent tissue-specific mtDNA mutations are common in humans. *PLOS Genetics* **9**, e1003929.

Santibanez-Koref, M., Griffin, H., Turnbull, D.M., Chinnery, P.F., Herbert, M. & Hudson, G. (2019) Assessing mitochondrial heteroplasmy using next generation sequencing: a note of caution. *Mitochondrion* **46**, 302–306.

Schmalenbach, I., Mehrtens, F., Janke, M. & Buchholz, F. (2011) A mark-recapture study of hatchery-reared juvenile European lobsters, *Homarus gammarus*, released at the rocky island of Helgoland (German Bight, North Sea) from 2000 to 2009. *Fisheries Research* **108**, 22–30.

Schoolmann, G. & Arndt, H. (2018) Population dynamics of the invasive freshwater shrimp *Neocaridina davidi* in the thermally polluted Gillbach stream (North Rhine-Westphalia, Germany). *Limnologica* **71**, 1–7.

Schübeler, D. (2015) Function and information content of DNA methylation. *Nature* **517**, 321–326.

Shabani, M., Borry, P., Smeers, I. & Bekaert, B. (2018) Forensic epigenetic age estimation and beyond: ethical and legal considerations. *Trends in Genetics* **34**, 489–491.

Sharp, W.C., Lellis, W.A., Butler, M.J., Herrnkind, W.F., Hunt, J.H., Pardee-Woodring, M. & Matthews, T.R. (2000) The use of coded microwire tags in mark-recapture studies of juvenile Caribbean spiny lobster, *Panulirus argus*. *Journal of Crustacean Biology* **20**, 510–521.

Sheehy, M.R.J., Bannister, R.C.A., Wickins, J.F. & Shelton, P.M.J. (1999) New perspectives on the growth and longevity of the European lobster (*Homarus gammarus*). *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 1904–1915.

Shen, H., Braband, A. & Scholtz, G. (2013) Mitogenomic analysis of decapod crustacean phylogeny corroborates traditional views on their relationships. *Molecular Phylogenetics and Evolution* **66**, 776–789.

Sheridan, M., O'Connor, I. & Henderson, A.C. (2016) Investigating the effect of molting on gastric mill structure in Norway lobster (*Nephrops norvegicus*) and its potential as a direct ageing tool. *Journal of Experimental Marine Biology and Ecology* **484**, 16–22.

Sheridan, M., Durán, J., del Mar Gil, M., Pastor, E. & O'Connor, I. (2020) Can crustacean cuticle preserve a record of chronological age? Investigating the utility of decapod crustacean eyestalks for age determination of two European spider crab species. *Fisheries Research* **224**, 105467.

Silva, C.N.S., Villacorta-Rath, C., Woodings, L.N., Murphy, N.P., Green, B.S., Hartmann, K., Gardner, C., Bell, J.J. & Strugnell, J.M. (2019) Advancing our understanding of

the connectivity, evolution and management of marine lobsters through genetics. *Reviews in Fish Biology and Fisheries* **29**, 669–687.

Siragusa, E. (2015) Approximate string matching for high-throughput sequencing. PhD thesis, Free University of Berlin.

Slieker, R.C., Relton, C.L., Gaunt, T.R., Slagboom, P.E. & Heijmans, B.T. (2018) Age-related DNA methylation changes are tissue-specific with *ELOVL2* promoter methylation as exception. *Epigenetics & Chromatin* **11**, 25.

Smith, M.T. & Addison, J.T. (2003) Methods for stock assessment of crustacean fisheries. *Fisheries Research* **65**, 231–256.

Snir, S., Farrell, C. & Pellegrini, M. (2019) Human epigenetic aging is logarithmic with time across the entire lifespan. *Epigentics* **14**, 912–926.

Song, H., Buhay, J.E., Whiting, M.F. & Crandall, K.A. (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences* **105**, 13486–13491.

Spanier, E., Lavalli, K.L., Goldstein, J.S., Groeneveld, J.C., Jordaan, G.L., Jones, C.M., Phillips, B.F., Bianchini, M.L., Kibler, R.D., Díaz, D., Mallol, S., Goñi, R., van Der Meeren, G.I., Agnalt, A., Behringer, D.C., Keegan, W.F. & Jeffs, A. (2015) A concise review of lobster utilization by worldwide human populations from prehistory to the modern era. *ICES Journal of Marine Science* **72**, i7–i21.

Spelbrink, J.N., Toivonen, J.M., Hakkaart, G.A.J., Kurkela, J.M., Cooper, H.M., Lehtinen, S.K., Lecrenier, N., Back, J.W., Speijer, D., Foury, F. & Jacobs, H.T. (2000) *In vivo* functional analysis of the human mitochondrial DNA polymerase PLOG expressed in cultured human cells. *Journal of Biological Chemistry* **275**, 24818–24828.

Spiers, H., Hannon, E., Wells, S., Williams, B., Fernandes, C., & Mill, J. (2016) Age-associated changes in DNA methylation across multiple tissues in an inbred mouse model. *Mechanisms of Ageing and Development* **154**, 20–23.

Stecher, G., Tamura, K. & Kumar, S. (2020) Molecular evolutionary genetics analysis (MEGA) for macOS. *Molecular Biology and Evolution*, msz312.

Steneck, R.S., Hughes, T.P., Cinner, J.E., Adger, W.N., Arnold, S.N., Berkes, F., Boudreau, S.A., Brown, K., Folke, C., Gunderson, I., Olsson, P., Scheffer, M., Stephenson, E., Walker, B., Wilson, J. & Worm, B. (2011) Creation of a gilded trap by the high economic value of the Maine lobster fishery. *Conservation Biology* **25**, 904–912.

Stephens, Z.D., Hudson, M.E., Mainzer, L.S., Taschuk, M., Weber, M.R. & Iyer, R.K. (2016) Simulating next-generation sequencing datasets from empirical mutation and sequencing models. *PLOS ONE* **11**, e0167047.

Stewart, J.B. & Chinnery, P.F. (2015) The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nature Reviews Genetics* **16**, 530–542.

Stewart, J.B. & Larsson, N. (2014) Keeping mtDNA in shape between generations. *PLOS Genetics* **10**, e1004670.

Stubbs, T.M., Bonder, M.J., Stark, A., Krueger, F., BI Ageing Clock Team, von Meyenn, F., Stegle, O. & Reik, W. (2017) Multi-tissue DNA methylation age predictor in mouse. *Genome Biology* **18**, 68.

Sumaila, U.R. & Tai, T.C. (2020) End overfishing and increase the resilience of the ocean to climate change. *Frontiers in Marine Science* **7**, 523.

Sun, X. & Yang, A. (2016) Exceptionally large mitochondrial fragments to the nucleus in sequenced mollusc genomes. *Mitochondrial DNA* **27**, 1409–1410.

Tan, M.H., Gan, H.M., Lee, Y.P., Bracken-Grissom, H., Chan, T., Miller, A.D. & Austin, C.M. (2019) Comparative mitogenomics of the Decapoda reveals evolutionary heterogeneity in architecture and composition. *Scientific Reports* **9**, 10756.

Tang, B., Zhou, K., Song, D., Yang, G. & Dai, A. (2003) Molecular systematics of the Asian mitten crabs, genus *Eriocheir* (Crustacea: Brachyura). *Molecular Phylogenetics and Evolution* **29**, 309–316.

Taylor, R.W. & Turnbull, D.M. (2005) Mitochondrial DNA mutations in human disease. *Nature Reviews Genetics* **6**, 389–402.

Thèves, C., Keyser-Tracqui, C., Crubézy, E., Salles, J.P., Ludes, B. & Telmon, N. (2006) Detection and quantification of the age-related point mutation A189G in the human mitochondrial DNA. *Journal of Forensic Sciences* **51**, 865–873.

Thompson, M.J., vonHoldt, B., Horvath, S. & Pellegrini, M. (2017) An epigenetic aging clock for dogs and wolves. *Aging* **9**, 1055–1068.

Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267–288.

Townsend, E. (2011) *Lobster: A Global History*. Reaktion Books.

Triant, D.A. & DeWoody, J.A. (2007) The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematics and phylogeography. *Journal of Mammalogy* **88**, 908–920.

Trifunovic, A. (2006) Mitochondrial DNA and ageing. *Biochimica et Biophysica Acta* **1757**, 611–617.

Tully, O., Roantree, V. & Robinson, M. (2001) Maturity, fecundity and reproductive potential of the European lobster (*Homarus gammarus*) in Ireland. *Journal of the Marine Biological Association of the UK* **81**, 61–68.

Uglem, I., Belchier, M. & Svåsand, T. (2005) Age determination of European lobsters (*Homarus gammarus* L.) by histological quantification of lipofuscin. *Journal of Crustacean Biology* **25**, 95–99.

Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S*. Springer.

Vila, Y., Medina, A., Megina, C., Ramos, F. & Sobrino, I. (2000) Quantification of the age-pigment lipofuscin in brains of known-age, pond-reared prawns *Penaeus japonicus* (Crustacea, Decapoda). *Journal of Experimental Zoology* **286**, 120–130.

Vogt, G. (2012) Ageing and longevity in the Decapoda (Crustacea): a review. *Zoologischer Anzeiger* **251**, 1–25.

Vogt, G. (2019) A compilation of longevity data in decapod crustaceans. *Nauplius* **27**, e2019011.

Vogt, G., Falckenhayn, C., Schrimpf, A., Schmid, K., Hanna, K., Panteleit, J., Helm, M., Schulz, R. & Lyko, F. (2015) The marbled crayfish as a paradigm for saltational speciation by autopolyploidy and parthenogenesis in animals. *Biology Open* **4**, 1583–1594.

von Wurmb-Schwark, N., Higuchi, R., Fenech, A.P., Elfstroem, C., Meissner, C., Oehmichen, M. & Cortopassi, G.A. (2002) Quantification of human mitochondrial DNA in a real time PCR. *Forensic Science International* **126**, 34–39.

Wachsmuth, M., Hübner, A., Mingkun, L., Burkhard, Madea. & Stoneking, M. (2016) Age-related and heteroplasmy-related variation in human mtDNA copy number. *PLOS Genetics* **12**, e1005939.

Wang, M. & Lemos, B. (2019) Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome Research* **29**, 325–333.

Wang, D., Xiang, H., Ning, C., Liu, H., Liu, J., Zhao, X. (2020) Mitochondrial DNA enrichment reduced NUMT contamination in porcine NGS analyses. *Briefings in Bioinformatics* **21**, 1368–1377.

Wang, E., Wong, A. & Cortopassi, G. (1997) The rate of mitochondrial mutagenesis is faster in mice than humans. *Mutation Research* **377**, 157–166.

Wang, M., Marinotti, O., Zhong, D., James, A.A., Walker, E., Guda, T., Kweka, E.J., Githure, J. & Yan, G. (2013) Gene expression-based biomarkers for *Anopheles gambiae* age grading. *PLOS ONE* **8**, e69439.

Wang, X.V., Blades, N., Ding, J., Sultana, R., & Parmigiani, G. (2012). Estimation of sequencing error rates in short reads. *BMC Bioinformatics* **13**, 185.

Wang, Y., Michikawa, Y., Mallidis, C., Bai, Y., Woodhouse, L., Yarasheski, K.E., Miller, C.A., Askanas, V., Engel, W.K., Bhasin, S. & Attardi, G. (2001) Muscle-specific mutations accumulate with aging in critical human mtDNA control sites for replication. *Proceedings of the National Academy of Sciences* **98**, 4022–4027.

Warner, A.H. & Bagshaw, J.C. (1984) Absence of detectable 5-methylcytosine in DNA of embryos of the brine shrimp, *Artemia*. *Developmental Biology* **102**, 264–267.

Watson, R.A., Cheung, W.W.L., Anticara, J.A., Sumaila, U.R., Zeller, D. & Pauly, D. (2012) Global marine yield halved as fishing intensity redoubles. *Fish and Fisheries* **14**, 493–503.

Weidner, C.I., Lin, Q., Koch, C.M., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D.O., Jöckel, K-H., Erbel, R., Mühleisen, T.W., Zenke, M., Brümmendorf, T.H. & Wagner, W. (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology* **15**, R24.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Wilm, A., Aw, P.P.K., Bertrand, D., Yeo, G.H.T., Ong, S.H., Wong, C.H., Khor, C.C., Petric, R., Hibberd, M.L. & Nagarajan, N. (2012) LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research* **40**, 11189–11201.

Wilson, V.L., Smith, R.A., Ma, S. & Cutler, R.G. (1987) Genomic 5-methyldeoxycytidine decreases with age. *Journal of Biological Chemistry* **262**, 9948–9951.

Wong, T.F.K., Ranjard, L., Lin, Y. & Rodrigo, A.G. (2018) HaploJuice: accurate haplotype assembly from a pool of sequences with known relative concentrations. *BMC Bioinformatics* **19**, 389.

Worm, B, Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., Jackson, J.B.C., Lotze, H.K., Micheli, F., Palumbi, S.R., Sala, E., Selkoe, K., Stachowicz, J.J. & Watson, R. (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science* **314**, 787–790.

Wright, P.G.R., Mathews, F., Schofield, H., Morris, C., Burrage, J., Smith, A., Dempster, E.L. & Hamilton, P.B. (2018) Application of a novel molecular method to age free-living wild Bechstein's bats. *Molecular Ecology Resources* **18**, 1374–1380.

Yu, Y.Q., Yang. W.J. & Yang, J.S. (2014) The complete mitogenome of the Chinese swamp shrimp *Neocaridina denticulata sinensis* Kemp 1918 (Crustacea: Decapoda: Atyidae). *Mitochondrial DNA* **25**, 204–205.

Yuan, Y., Ju, Y.S., Kim, Y., Li, J., Wang, Y., Yoon, C.J., Yang, Y., Martincorena, I., Creighton, C.J., Weinstein, J.N., Xu, Y., Han, L., Kim, H., Nakagawa, H., Park, K., Campbell, P.J., Liang, H. & PCAWG Consortium (2020) Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nature Genetics* **52**, 342–352.

Zeng, Y. & Chen, T. (2019) DNA methylation reprogramming during mammalian development. *Genes* **10**, 10040257.