Norwegian University of Life Sciences

Master's Thesis 2024 60 ECTS
Faculty of Biosciences

Evaluating genetic tools to inform conservation efforts for the nationally red-listed European lobster (*Homarus gammarus*)

Erik Sandertun Røed
Master of Science in Genome Science

Master of Science in Genome Science

Evaluating genetic tools to inform conservation efforts for the nationally red-listed European lobster (*Homarus gammarus*)

M60-GS: Article-based thesis

Student: Erik Sandertun Røed [1, 2]

Primary supervisor: Marie Saitou [1]

Co-supervisor: Louise Chavarie [2]

[1] Centre for Integrative Genetics (CIGENE), Faculty of Biosciences, NMBU

[2] Faculty of Environmental Sciences and Natural Resource Management (MINA), NMBU

Table of contents

Abstract

Due to a concerning decline in population abundance, the European lobster (*Homarus gammarus*) has become a vulnerable species in Norway. In response, policymakers have established lobster sanctuaries where catch is prohibited, and biologists closely monitor the recovery of local abundance and population structure. Genetic tools are powerful instruments both for designing and evaluating the impact of such sanctuaries, in part because genetic tools permit non-lethal sampling where more traditional methods may not. Still, any tool applied in monitoring must be reliably evaluated before its potential benefits can be weighed against its cost. In this thesis, two relatively novel genetic methods were evaluated with a focus on their potential application in lobster monitoring in Norway: a novel epigenetic age estimation tool and a multipurpose panel of single nucleotide polymorphisms (SNPs) applied as a tool to detect hybridisation with an invasive species, the American lobster (*H. americanus*). The epigenetic ageing tool was evaluated by testing its ability to accurately predict the longitudinal difference in age between repeated captures of the same lobster, with samples sourced from a lobster sanctuary long-term monitoring program. The genetic tool for detecting hybridisation with *H. americanus* was evaluated *in silico*, by developing a pipeline for simulating the SNP panel; this enabled testing its ability to detect hypothetical hybrids that have not been observed in the wild. The epigenetic ageing method mis-estimated differences in age, but did so systematically, which pointed to useful avenues for future development of epigenetic ageing in lobsters. The SNP panel simulations, on the other hand, were successful in demonstrating that when applied as a tool for detecting hybrids, the multipurpose SNP panel is more sensitive than what has previously been shown with only empirical data. For both these genetic tools, this thesis thereby points to avenues of future research and could contribute to the eventual use of the genetic tools in lobster monitoring.

Preface

To protect the nationally red-listed European lobster, a series of lobster sanctuaries have been formed along the Norwegian Skagerrak coast. In 2021, the smallest and most recent of these sanctuaries was established at Jetéen in the Drøbak sound, near the narrowest point of the Oslofjord. In an ongoing long-term monitoring program, researchers at NMBU's Faculty of Environmental Sciences and Natural Resource Management study the influence of the Jetéen sanctuary on the local lobster population. As of writing, the monitoring program has collected over 1 000 lobster tissue samples intended for genetic analyses. The goal of this thesis, the product of my genome science degree, was to evaluate potential uses of these genetic resources. My research has centred on two existing but relatively novel genetic tools that could be implemented as part of NMBU's monitoring program in the Oslofjord: one tool for ageing lobsters non-lethally, and one for detecting signs of hybridisation with an invasive species.

That said, the use my supervisors and I initially had in mind for this treasure trove of tissue samples was quite different compared to my thesis as it stands now. Our original goal was to investigate why some rare lobsters are blue, a question we hoped would pique some public interest for the lobster to help communicate about its deteriorating conservation status in Norway. As it turns out, blue lobsters are sufficiently rare that we never managed to obtain the necessary genetic samples (and it's quite a difficult phenotype to score properly anyways). I wish to acknowledge and thank Arturo Vera-Ponce de Leon (CIGENE) for his supervision when we still pursued the blue lobster project. And, to my supervisors Marie Saitou and Louise Chavarie: thank you so much for guiding me through the process of identifying new questions to work with, for your engaged supervision and mentorship, and for entrusting me with so much responsibility for the scientific and practical research process leading to this final thesis.

The primary content of this thesis is two journal article drafts, which are summarised and contextualised in the thesis cover. Except where I acknowledge otherwise, this thesis is of course exclusively my own work. Even so, I am very grateful to many others who have assisted me directly or indirectly – my work here has been possible only thanks to them. For the first article draft (Article A), I used tissue samples collected from 2020 to 2023 by Thrond O. Haugen (MINA), Stein R. Moe (MINA), Jonathan E. Colman (MINA), Linda E. Lemmens (MINA), my supervisor Louise Chavarie, and many staff, students, and others who volunteered their time. I was very lucky to join them for about two weeks of fieldwork in total. Jun Soung Kwak (CIGENE) was absolutely instrumental to the subsequent logistics and lab work. To ensure that

Article A got off the ground at all, Jun spent many hours helping me catalogue tissue samples, performed just over half of the DNA extractions and ran a gel electrophoresis for me, and made sure we got the samples shipped abroad for analyses. Once the data were ready, Lars Grønvold (CIGENE) kindly advised me on analyses. For the second article draft (Article B), I want to acknowledge the assistance of Charlie Ellis and Jamie Stevens (University of Exeter) who enthusiastically shared their data, scripts, and thoughts with my supervisors and me. The idea for the second article was developed through correspondence with them. I also want to thank my supervisors' PhD students Celián, Domniki, and Leah for their thoughts and suggestions.

Finally, my time at NMBU would not have been the same without my peers. Thanks to my genome science colleagues Astrid, Lindis, and TPO for great company throughout, to the CIGENE writing group for feedback on my first article draft, and to Pierre and Einar at MINA's R Club for being fellow R enthusiasts. And last but definitely not least: thanks to my close friend Ellen for keeping me level-headed and properly caffeinated over the last two years. It wouldn't have been the same without you!

Erik Sandertun Røed

Ås, May 2024

1. Introduction

The European lobster (*Homarus gammarus*, hereafter either "lobster" or "*H. gammarus*") is a longevous benthic decapod. Native to European waters from the Arctic coast of Norway to the Adriatic, lobsters inhabit an intermediate niche and contribute to structuring the benthic ecosystem via complex trophic interactions (Boudreau & Worm 2012). For centuries, the lobster has also been a culturally and commercially important seafood across much of its pan-European range (Spanier et al. 2015). With the advent of intensive and often poorly regulated modern fisheries, though, many exploited invertebrate stocks have become severely depleted (Anderson et al. 2011). Coastal European waters are amongst the world's most intensively fished (Kroodsma et al. 2018), so the lobster is perhaps unsurprisingly no exception. As of 2021, the lobster is classified as vulnerable in Norway's national red-list of species (Tandberg et al. 2021). Norwegian stocks that historically supported commercial exports have collapsed to sustain only a regulated, seasonal, and primarily recreational fishery (Kleiven et al. 2012, 2022, Spanier et al. 2015). Even then, catchable lobster mortality in intensively trapped Norwegian waters has been reported as high as 83 % over one lobstering season (Wiig et al. 2013). Improved traps and limited catch reporting exacerbate the issue and appear to prevent recovery of fished Norwegian stocks (Kleiven et al. 2012, 2022). In response, Norwegian policymakers have established lobster sanctuaries where all capture of lobster is banned (Knutsen et al. 2022).

Genetic tools can contribute both to the design and evaluation of lobster sanctuary impacts. Broadly speaking, lobster sanctuaries have locally been successful in restoring population abundance (Fernández-Chacón et al. 2021, Moland et al. 2021, Haugen et al. 2023) and diversity (Sørdalen et al. 2018, 2020, Moland et al. 2019). Research on genetic connectivity between sanctuaries in Norway and Sweden suggest that dispersal of larvae from even small sanctuaries could contribute to recruitment over long distances, potentially exporting beneficial local effects (Huserbråten et al. 2013). Studies of genetic connectivity can thereby identify priority source populations to aid sanctuary design (e.g. Pavičić et al. 2020). Simultaneously, genetic tools can identify isolated populations where recruitment could be more prone to harvest of local adult, breeding individuals (e.g. Jørstad et al. 2004, Ellis et al. 2024). Beyond sanctuary design, comprehensive studies of single nucleotide polymorphism (SNP) markers across the native range of the lobster can also inform sourcing of broodstock for stock enhancement programs to ensure maintenance of genetic variation (Ellis et al. 2017, Jenkins et al. 2019b, 2019a, 2020). Still, there are other potentially valuable genetic tools that are not used widely in monitoring of lobster sanctuaries; this thesis aimed to evaluate two of them.

The first overall aim of this thesis was to evaluate the applied reliability of a novel genetic tool for lobster age estimation (Fairfield et al. 2021). Monitoring and maintaining the demographic age composition of a harvested species is a central goal in fisheries management (Francis et al. 2007), but in lobsters age estimation has traditionally necessitated lethal sampling (Sheehy et al. 1996, e.g. 1999, Sheehy & Bannister 2002). As lethal sampling is counter to the intention of a lobster sanctuary, sanctuary monitoring programs are consequently unable to directly monitor the age-demographic response to protection (see e.g. Moland et al. 2021, Haugen et al. 2023). A potential solution was identified by Fairfield et al. (2021), who reported a promising epigenetic age estimation tool based on predictable changes in age-related DNA methylation. Because Fairfield et al. (2021) based their findings on lobsters that were younger than typically catchable lobsters in a fished population, it was unclear if their ageing tool would generalise to older, wild lobsters. Evaluating this was the goal of Article A.

The second overall aim of this thesis was developing simulations to evaluate potential genetic monitoring schemes for invasive introgression in lobsters. Accidental or deliberate introductions of the invasive American lobster (*H. americanus*) have occurred across much of the native range of *H. gammarus* (Jørstad et al. 2011, Stebbing et al. 2012, Øresland et al. 2017, Barrett et al. 2020, Pavičić 2020, Kampouris et al. 2021, Tinlin-Mackenzie et al. 2022). Observations of *H. americanus* have occurred also in Norway, where they are considered a high-risk invasive species (Agnalt et al. 2023). The threat of *H. americanus* is in part its ability to hybridise with *H. gammarus* (van der Meeren et al. 2010), so Ellis et al. (2020) adopted a SNP panel developed for population assignment by Jenkins et al. (2019a) to identify the two species from each other and from their hybrids. The aim of Article B in this thesis was to develop simulations that can estimate how strong the signal of hybridisation will be even after backcrossing with *H. gammarus*, and ultimately prepare a pipeline for determining the necessary sampling scope required to detect different hypothetical levels of genetic contribution from *H. americanus* in the wild, using the SNP panel of Jenkins et al. (2019a) like Ellis et al. (2020) did.

2. Methods

The methods of each article are described in complete detail in the articles themselves, but a short summary is provided here. All analyses, simulations and graphics in the articles were produced with code that is openly available at https://github.com/ErikSRoeed/lobmsc. The repository also includes data availability statements and select datasets that are publicly available.

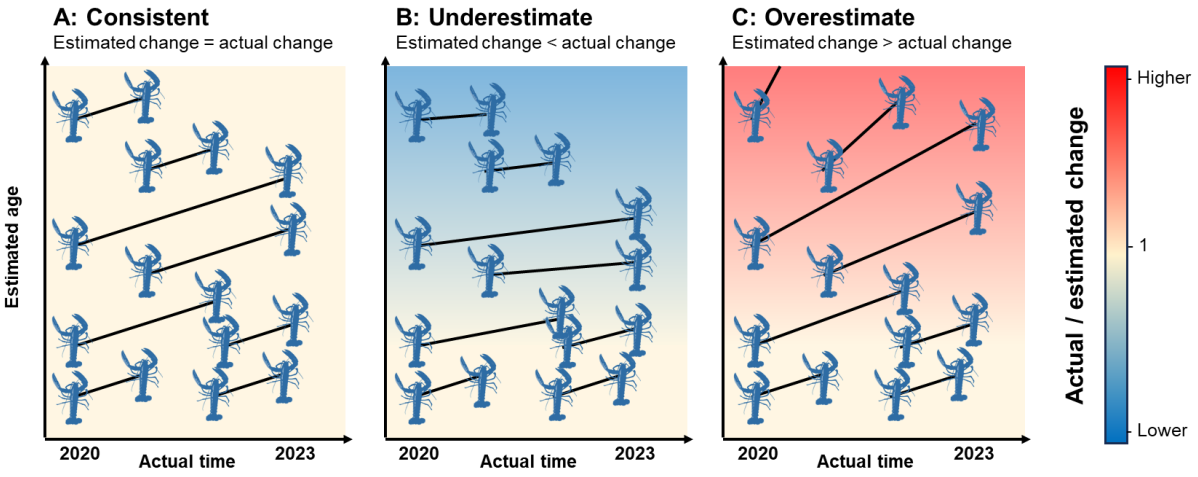

Figure 1 An illustration of hypothetical outcomes in Article A. Lobsters were captured and recaptured at known time intervals (shown as lobster icons connected by lines), and a DNA-based age estimate produced with Fairfield et al.'s (2021) epigenetic ageing method at each time point. If the epigenetic ageing method generalises well to a wild population, the actual over estimated change in age should be one across the age range, as in **A**. If the method generalises poorly, it could either underestimate (**B**) or overestimate (**C**) ages in lobsters that are older than in the original study. Figure adapted with modification from Røed (2024b).

2.1 Summary of methods: Article A

To test how well Fairfield et al.'s (2021) epigenetic ageing method for lobsters might generalise to lobsters of catchable age in the wild, Article A investigated whether age estimates of re-captured lobsters consistently increased by the same amount as the known time elapsed between capture and recapture. Capture and recapture samples were obtained from 48 lobsters through NMBU's long-term monitoring program at the Jetéen lobster sanctuary near Drøbak, Norway (see Haugen et al. 2023). The time from capture to recapture ranged between one and three years. Epigenetic DNA methylation in the genomic region targeted by Fairfield et al. (2021) was obtained for each sample via bisulphite sequencing as described in that study (albeit with minor modifications described in Article A). Age estimates were then produced with Fairfield et al.'s (2021) original statistical model which related age to the degree of methylation at select sites. The relationship between the estimated and actual difference in age was then assessed as illustrated in fig. 1.

The original statistical model of Fairfield et al. (2021) may not have captured relationships between ageing and DNA methylation in the targeted genomic region that are informative in older lobsters. In addition to testing Fairfield et al.'s (2021) original model, Article A additionally interrogated the relationship between ageing and methylation in the 48 Jetéen lobsters sourced here (the "Jetéen data"). First, statistical models were fit to methylation at each

individual methylated site in the target genomic region (an epigenome-wide association scan, or EWAS, as reviewed by Wei et al. 2021) to test which methylated sites were significantly related to the amount of time passed from capture to recapture. A standard machine learning procedure for finding the most age-predictive DNA methylation sites (Anastasiadi & Piferrer 2023) was applied both across all the sites in the Jetéen data and the sites identified as most significant by the EWAS. Finally, Article A assesses the ability of those DNA methylation sites to predict both ageing (i.e. time between capture and recapture) in the Jetéen data and age in the original data of Fairfield et al. (2021).

2.2 Summary of methods: Article B

Article B implemented simulation tools capable of evaluating *in silico* how sensitive the SNP panel of Jenkins et al. (2019a) is to genetic contributions from *H. americanus* in *H. gammarus*, beyond what was demonstrated *in vivo* by Ellis et al. (2020). The simulation tools comprised a core simulation built in SLiM 4 (Haller & Messer 2023) and utilities written in R (R Core Team 2024) to import and export genotype data so that simulations could be initialised with empirical genotypes. In the core simulation, a SNP panel is simulated as an explicit nucleotide sequence (fig. 2A), in principle of any length. The individual genotypes can also be structured into populations connected by migration, e.g. to simulate invasion of *H. americanus* into a population of *H. gammarus*. For the purposes of Article B, simulations were initialised by importing empirical genotype data from *H. gammarus*, *H. americanus*, and hybrids, provided by Ellis et al (2020), including data from Jenkins et al. (2019b).

Once initialised, simulated genotypes "mate" freely to produce offspring representative of, but not identical to, the original empirical data. Article B thereby simulates hybridisation and multiple generations of backcrossing by importing empirical genotypes both from *H. americanus* and *H. gammarus* in a simulated crossing experiment (fig. 2B). To evaluate whether Jenkins et al.'s (2019a) SNP panel distinguishes also backcrossed hybrids from the parental genotypes, Article B finally replicates the population assignment analyses of Ellis et al. (2020) which those authors originally used to detect first generation (i.e. F1) hybridisation.

To interface between SLiM and R, Article B makes use of the R package slimmr v. 1.0.0 (Røed 2024a). While slimmr (previously v. 0.2.0) was originally developed for Røed & Engelstädter (2022), the package was reworked (to v. 1.0.0) from the ground up for the purposes of Article B, and is available to install at https://github.com/ErikSRoeed/slimmr.

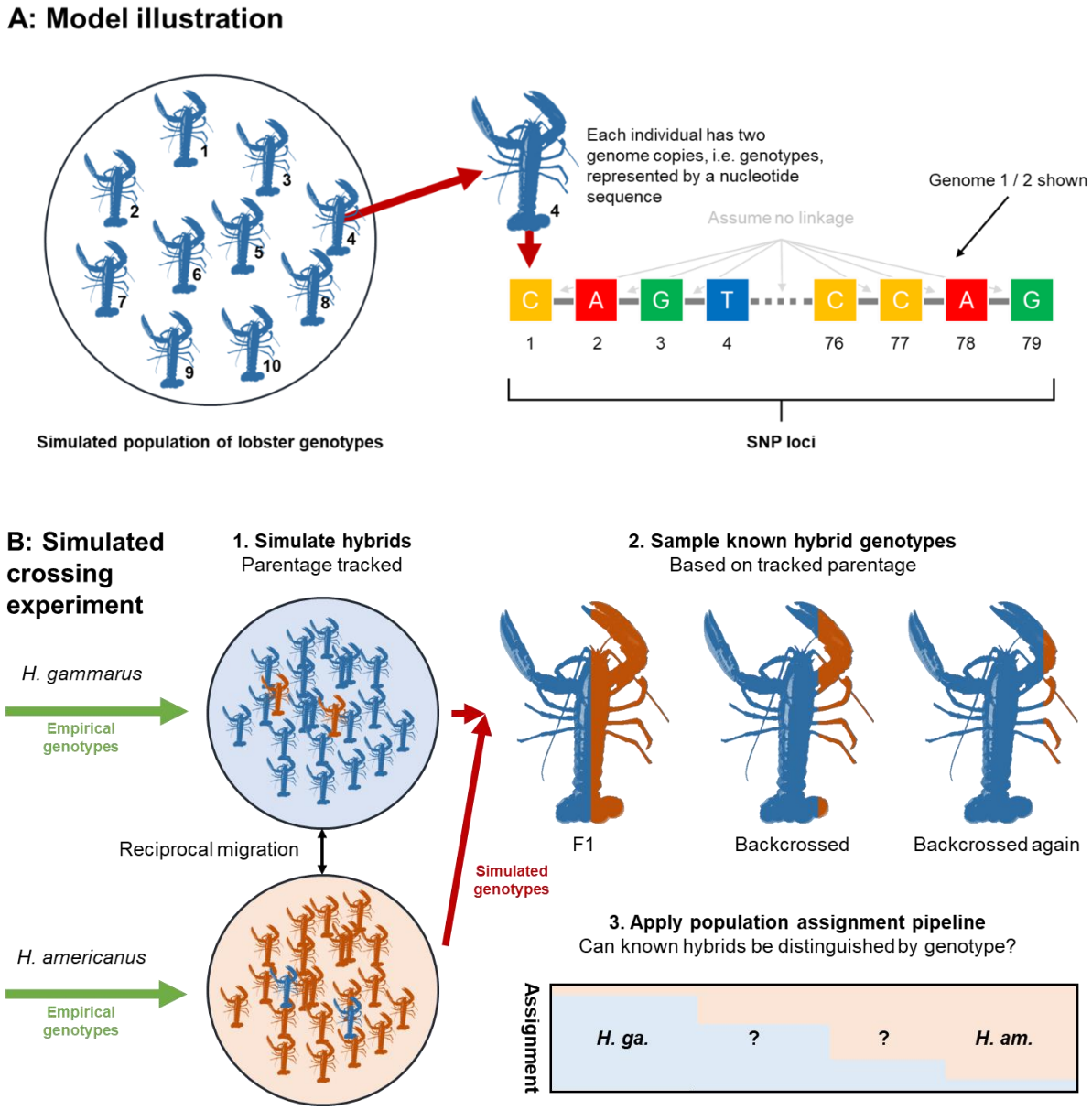

Figure 2 An illustration of the core simulation in Article B. Panel **A** shows a schematic illustration of the foundational model implementation, i.e. how individual genotypes are represented in a simulated population. Nucleotide sequences can be sourced from empirical data to simulate a panel of, in principle, any size. Panel **B** illustrates a simulated crossing experiment in Article B, where empirical data from Ellis et al. (2020) are imported and crossed in silico to produce novel backcrossed genotypes over the SNP panel reported by Jenkins et al. (2019a). To test if the SNP panel can resolve hybrids beyond first generation (F1), Article B implements the same analysis pipeline for population assignment as Ellis et al. (2020).

3. Results and discussion

3.1 Summary of results and discussion: Article A

In each of the 48 Jetéen lobsters sampled in Article A, DNA methylation was measured successfully across the genomic region targeted by Fairfield et al. (2021), and the overall trend in methylation from capture to recapture, i.e. with increasing age, was as expected. However, the age estimates predicted by the age-methylation relationship at methylated sites identified by Fairfield et al. (2021) were systematically underestimated by on average ca. one and a half years (reminiscent of the illustration in fig. 2B).

Assessing whether other methylated sites may be more predictive in the Jetéen lobsters, the EWAS identified several sites in the target genomic region where methylation was significantly associated with ageing. Moreover, the machine learning models trained on the Jetéen data performed somewhat better than Fairfield et al.'s (2021) original model, but no combination of methylated sites was found that did not underestimate the difference in age between capture and recapture. Yet, notably, the combinations of methylated sites that were least error-prone in the Jetéen data were predictive of age in the original data of Fairfield et al. (2021).

As such, the most important finding of Article A is that the underlying assumptions of the models relating age (or ageing) to methylation in both Fairfield et al. (2021) and Article A appeared valid in the young lobsters Fairfield et al. (2021) studied, but not in the wild Jetéen lobsters. Fairfield et al. (2021) in fact identified this as a potential concern in their study; the underlying model assumes the relationship between age and methylation at each methylated site used to predict age is linear. That assumption clearly held for young lobsters in the original study, and Article A recapitulated the same results using Fairfield et al.'s (2021) data. But the systematic underestimation of ages in the older Jetéen lobsters suggested, on the contrary, that the relationship between age and methylation is decelerated in older animals. This is reminiscent of the relationship between methylation and ageing in humans (Horvath & Raj 2018, Snir et al. 2019). These findings can inform the choice of statistical model used to relate age and DNA methylation in future development of epigenetic ageing for lobsters.

3.2 Summary of results and discussion: Article B

Article B established a freely available and extensible simulation pipeline for simulating SNP data. To validate that the core simulation imposed no bias on the simulated genotypes, Article

B first neutrally simulated empirical data for Jenkins et al.'s SNP panel (2019a) sourced from Ellis *et al*. (2020). Simulated genotypes clustered with their empirical source material, verifying that the simulations functioned as expected and that results of the subsequent simulated crossing experiment (fig. 2B) should be representative of the empirical data it was initialised from.

In the simulated crossing experiment, genotypes were assigned their correct level of crossing or backcrossing with *H. americanus* in at least 93 % of cases when accounting for one generation of backcrosses. The probabilities of correct assignments were also high compared to false positive or false negative detections of *H. americanus* ancestry. Results were less consistent when accounting for a second generation of backcrosses, but generally Article B suggested the 79 SNPs of Jenkins et al. (2019a) are more powerful at detecting hybrid specimens than previously shown by Ellis et al. (2020). The flexible simulation tools provided in Article B further provide a framework for planning monitoring schemes using SNP genotyping.

3.3 Synthesis and concluding remarks

Sustainable management of European lobster stocks in Norway can benefit from the application of genetic tools (e.g. Jørstad et al. 2004, 2011, Huserbråten et al. 2013). In a sanctuary monitoring setting, this is not least because genetic sampling can be performed non-lethally (Butler 2017). Still, practical barriers to the application of genetic tools could include a need for additional validation (e.g. Fairfield et al. 2021) or a scarcity of empirical data to evaluate their performance against (e.g. Ellis et al. 2020). Over two articles, the aims of this thesis were to contribute towards the evaluation of two genetic tools for lobster management; ultimately, to facilitate their practical use. The epigenetic ageing tool of Fairfield et al. (2021) did unfortunately not appear to predict accurately enough to be useful in sanctuary monitoring at present. Still, Article A identified avenues of future research that may contribute to building on the foundation Fairfield et al. (2021) has laid for DNA-based and non-lethal ageing in lobsters. On the other hand, Article B showed that the SNP panel published by Jenkins et al. (2019a) might in fact be an even more sensitive tool for detecting hybridisation with *H. americanus* than Ellis et al. (2020) could observe with available empirical data. Article B further provides easily extensible tools that can be used to plan and optimise monitoring schemes in the wild. As such, both articles that comprise the main body of this thesis identified actionable avenues of future research that can contribute towards the sustainable management and conservation of the nationally red-listed European lobster.

5. References

Agnalt, A.-L., Falkenhaug, T., Glenner, H., Husa, V., Jelmert, A. & Mortensen, S. 2023. Storkrepser: Vurdering av amerikahummer *Homarus americanus* for Fastlands-Norge med havområder. Fremmedartslista 2023. Artsdatabanken. https://artsdatabanken.no/lister/fremmedartslista/2023/1681

Anastasiadi, D. & Piferrer, F. 2023. Bioinformatic analysis for age prediction using epigenetic clocks: Application to fisheries management and conservation biology. Frontiers in Marine Science, **10**: 1096909. doi:10.3389/fmars.2023.1096909.

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. doi:10.1371/journal.pone.0014735.

Barrett, C.J., Cook, A., Stone, D., Evans, C., Murphy, D., Johnson, P., Thain, M., Wyn, G., Grey, M., Edwards, H., Quigley, D. & Stebbing, P.D. 2020. A review of American lobster (*Homarus americanus*) records around the British Isles: 2012 to 2018. Hydrobiologia, **847**: 3247–3255. doi:10.1007/s10750-020-04326-7.

Boudreau, S. & Worm, B. 2012. Ecological role of large benthic decapods in marine ecosystems: a review. Marine Ecology Progress Series, **469**: 195–213. doi:10.3354/meps09862.

Butler, M.J. 2017. Collecting and processing lobsters. Journal of Crustacean Biology, **37**: 340–346. doi:10.1093/jcbiol/rux021.

Ellis, C., Hodgson, D., Daniels, C., Collins, M. & Griffiths, A. 2017. Population genetic structure in European lobsters: implications for connectivity, diversity and hatchery stocking. Marine Ecology Progress Series, **563**: 123–137. doi:10.3354/meps11957.

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. doi:10.1038/s41598-020-64692-z.

Ellis, C.D., Paris, J.R., Jenkins, T.L., Stralen, M.R. van, Steins, N.A., Schotanus, J. & Stevens, J.R. 2024. Genetic divergence and adaptation of an isolated European lobster population in the Netherlands. ICES Journal of Marine Science: fsae059. doi:10.1093/icesjms/fsae059.

Fairfield, E.A., Richardson, D.S., Daniels, C.L., Butler, C.L., Bell, E. & Taylor, M.I. 2021. Ageing European lobsters (*Homarus gammarus*) using DNA methylation of evolutionarily conserved ribosomal DNA. Evolutionary Applications, **14**: 2305–2318. doi:10.1111/eva.13296.

Fernández-Chacón, A., Buttay, L., Moland, E., Knutsen, H. & Olsen, E.M. 2021. Demographic responses to protection from harvesting in a long-lived marine species. Biological Conservation, **257**: 109094. doi:10.1016/j.biocon.2021.109094.

Francis, R.C., Hixon, M.A., Clarke, M.E., Murawski, S.A. & Ralston, S. 2007. Ten Commandments for Ecosystem-Based Fisheries Scientists. Fisheries, **32**: 217–233. doi:10.1577/1548-8446(2007)32[217:tcfbfs]2.0.co;2.

Haller, B.C. & Messer, P.W. 2023. SLiM 4: Multispecies Eco-Evolutionary Modeling. The American naturalist, **201**: E127–E139. doi:10.1086/723601.

Haugen, T.O., Colman, J.E. & Moe, S.R. 2023. Torsk og hummer i indre Oslofjord: Effekter av vernetiltak. MINA fagrapport 83. Norwegian University of Life Sciences, Ås, Norway.

Horvath, S. & Raj, K. 2018. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nature Reviews Genetics, **19**: 371–384. doi:10.1038/s41576-018-0004-3.

Huserbråten, M.B.O., Moland, E., Knutsen, H., Olsen, E.M., André, C. & Stenseth, N.Chr. 2013. Conservation, Spillover and Gene Flow within a Network of Northern European Marine Protected Areas. PLoS ONE, **8**: e73388. doi:10.1371/journal.pone.0073388.

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. doi:10.1002/ece3.6304.

Jenkins, T.L., Ellis, C.D. & Stevens, J.R. 2019a. SNP discovery in European lobster (*Homarus gammarus*) using RAD sequencing. Conservation Genetics Resources, **11**: 253–257. doi:10.1007/s12686-018-1001-8.

Jenkins, T.L., Ellis, C.D., Triantafyllidis, A. & Stevens, J.R. 2019b. 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. doi:10.1111/eva.12849.

Jørstad, K.E., Agnalt, A.-L. & Farestveit, E. 2011. The Introduced American Lobster Homarus americanus in Scandinavian Waters. *In* In the Wrong Place - Alien Marine Crustaceans: Distribution, Biology and Impacts. *Edited by* B.S. Galil, P.F. Clark & J.T. Carlton. Springer, Dordrecht. pp. 625–638. doi:10.1007/978-94-007-0591-3_22.

Jørstad, K.E., Prodöhl, P.A., Agnalt, A.-L., Hughes, M., Apostolidis, A.P., Triantafyllidis, A., Farestveit, E., Kristiansen, T.S., Mercer, J. & Svåsand, T. 2004. Sub-arctic Populations of European Lobster, *Homarus gammarus*, in Northern Norway. Environmental Biology of Fishes, **69**: 223–231. doi:10.1023/b:ebfi.0000022899.52578.37.

Kampouris, T., Gkafas, G., Sarantopoulou, J., Exadactylos, A. & Batjakas, I. 2021. An American in the Aegean: first record of the American lobster *Homarus americanus* H. Milne Edwards, 1837 from the eastern Mediterranean Sea. BioInvasions Records, **10**: 170–180. doi:10.3391/bir.2021.10.1.18.

Kleiven, A.R., Espeland, S.H., Stiansen, S., Ono, K., Zimmermann, F. & Olsen, E.M. 2022. Technological creep masks continued decline in a lobster (*Homarus gammarus*) fishery over a century. Scientific Reports, **12**: 3318. doi:10.1038/s41598-022-07293-2.

Kleiven, A.R., Olsen, E.M. & Vølstad, J.H. 2012. Total Catch of a Red-Listed Marine Species Is an Order of Magnitude Higher than Official Data. PLoS ONE, **7**: e31216. doi:10.1371/journal.pone.0031216.

Knutsen, J.A., Kleiven, A.R., Olsen, E.M., Knutsen, H., Espeland, S.H., Sørdalen, T.K., Thorbjørnsen, S.H., Hutchings, J.A., Fernández-Chacón, A., Huserbråten, M., Villegas-Ríos, D., Halvorsen, K.T., Kleiven, P.J.N., Langeland, T.K. & Moland, E. 2022. Lobster reserves as a management tool in coastal waters: Two decades of experience in Norway. Marine Policy, **136**: 104908. doi:10.1016/j.marpol.2021.104908.

Kroodsma, D.A., Mayorga, J., Hochberg, T., Miller, N.A., Boerder, K., Ferretti, F., Wilson, A., Bergman, B., White, T.D., Block, B.A., Woods, P., Sullivan, B., Costello, C. & Worm, B. 2018. Tracking the global footprint of fisheries. Science, **359**: 904–908. doi:10.1126/science.aao5646.

Moland, E., Carlson, S.M., Villegas-Ríos, D., Wiig, J.R. & Olsen, E.M. 2019. Harvest selection on multiple traits in the wild revealed by aquatic animal telemetry. Ecology and Evolution, **9**: 6480–6491. doi:10.1002/ece3.5224.

Moland, E., Fernández-Chacón, A., Sørdalen, T.K., Villegas-Ríos, D., Thorbjørnsen, S.H., Halvorsen, K.T., Huserbråten, M., Olsen, E.M., Kleiven, P.J.N., Kleiven, A.R., Knutsen, H., Espeland, S.H., Freitas, C. & Knutsen, J.A. 2021. Restoration of Abundance and Dynamics of Coastal Fish and Lobster Within Northern Marine Protected Areas Across Two Decades. Frontiers in Marine Science, **8**: 674756. doi:10.3389/fmars.2021.674756.

Øresland, V., Ulmestrand, M., Agnalt, A.-L. & Oxby, G. 2017. Recorded captures of American lobster (*Homarus americanus*) in Swedish waters and an observation of predation on the European lobster (*Homarus gammarus*). Canadian Journal of Fisheries and Aquatic Sciences, **74**: 1503–1506. doi:10.1139/cjfas-2016-0532.

Pavičić, M. 2020. First record of American lobster, *Homarus americanus* (H. Milne Edwards, 1837), in the Mediterranean Sea. BioInvasions Records, **9**: 83–88. doi:10.3391/bir.2020.9.1.11.

Pavičić, M., Žužul, I., Matić-Skoko, S., Triantafyllidis, A., Grati, F., Durieux, E.D.H., Celić, I. & Šegvić-Bubić, T. 2020. Population Genetic Structure and Connectivity of the European Lobster *Homarus gammarus* in the Adriatic and Mediterranean Seas. Frontiers in genetics, **11**: 576023. doi:10.3389/fgene.2020.576023.

R Core Team. 2024. R: A Language and Environment for Statistical Computing. https://www.R-project.org/.

Røed, E.S. 2024a. slimmr: SLiM Models in R. R package. https://github.com/ErikSRoeed/slimmr

Røed, E.S. 2024b. Labs to lobsters - promoting DNA tools in fisheries monitoring. Presentation, Reef Check Australia Coast to Coral Online. https://youtu.be/71c6nO2S5aw.

Røed, E.S. & Engelstädter, J. 2022. Cytoplasmic incompatibility in hybrid zones: infection dynamics and resistance evolution. Journal of Evolutionary Biology, **35**: 240–253. doi:10.1111/jeb.13974.

Sheehy, M.R.J. & Bannister, R.C.A. 2002. Year-class detection reveals climatic modulation of settlement strength in the European lobster, *Homarus gammarus*. Canadian Journal of Fisheries and Aquatic Sciences, **59**: 1132–1143. doi:10.1139/f02-083.

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. doi:10.1139/f99-116.

Sheehy, M.R.J., Shelton, P.M.J., Wickins, J.F., Belchier, M. & Gaten, E. 1996. Ageing the European lobster *Homarus gammarus* by the lipofuscin in its eyestalk ganglia. Marine Ecology Progress Series, **143**: 99–111. doi:10.3354/meps143099.

Snir, S., Farrell, C. & Pellegrini, M. 2019. Human epigenetic ageing is logarithmic with time across the entire lifespan. Epigenetics, **14**: 912–926. doi:10.1080/15592294.2019.1623634.

Sørdalen, T.K., Halvorsen, K.T., Harrison, H.B., Ellis, C.D., Vøllestad, L.A., Knutsen, H., Moland, E. & Olsen, E.M. 2018. Harvesting changes mating behaviour in European lobster. Evolutionary Applications, **11**: 963–977. doi:10.1111/eva.12611.

Sørdalen, T.K., Halvorsen, K.T., Vøllestad, L.A., Moland, E. & Olsen, E.M. 2020. Marine protected areas rescue a sexually selected trait in European lobster. Evolutionary Applications, **13**: 2222–2233. doi:10.1111/eva.12992.

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.-L., 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. doi:10.1093/icesjms/fsv066.

Stebbing, P., Johnson, P., Delahunty, A., Clark, P., McCollin, T., Hale, C. & Clark, S. 2012. Reports of American lobsters, *Homarus americanus* (H. Milne Edwards, 1837), in British waters. BioInvasions Records, **1**: 17–23. doi:10.3391/bir.2012.1.1.04.

Tandberg, A., Djursvoll, P., Falkenhaug, T., Glenner, H., Meland, K. & Walseng, B. 2021. Krepsdyr: Vurdering av hummer *Homarus gammarus* for Norge. Rødlista for arter 2021. Artsdatabanken. https://artsdatabanken.no/lister/rodlisteforarter/2021/14133

Tinlin-Mackenzie, A., Ellis, C., Lodola, A., Martin-Ruiz, C., Stevens, J. & Fitzsimmons, C. 2022. New kid on the block: first record of juvenile American lobster, *Homarus americanus* H. Milne Edwards, 1837, in European waters. BioInvasions Records, **11**: 901–1000. doi:10.3391/bir.2022.11.4.17.

van der Meeren, G.I., Støttrup, J., Ulmestrand, M., Øresland, V., Knutsen, J.A. & Agnalt, A.-L. 2010. NOBANIS Invasive Alien Species Fact Sheet - *Homarus americanus*. Online Database of the European Network on Invasive Alien Species – NOBANIS.

https://www.nobanis.org/globalassets/speciesinfo/h/homarus-americanus/homarus_americanus.pdf

Wei, S., Tao, J., Xu, J., Chen, X., Wang, Z., Zhang, N., Zuo, L., Jia, Z., Chen, H., Sun, H., Yan, Y., Zhang, M., Lv, H., Kong, F., Duan, L., Ma, Y., Liao, M., Xu, L., Feng, R., Liu, G., Project, T.E. & Jiang, Y. 2021. Ten Years of EWAS. Advanced Science, **8**: 2100727. doi:10.1002/advs.202100727.

Wiig, J.R., Moland, E., Haugen, T.O. & Olsen, E.M. 2013. Spatially structured interactions between lobsters and lobster fishers in a coastal habitat: fine-scale behaviour and survival estimated from acoustic telemetry. Canadian Journal of Fisheries and Aquatic Sciences, **70**: 1468–1476. doi:10.1139/cjfas-2013-0209.

Like clockwork? A longitudinal test of an epigenetic clock in wild European lobsters (*Homarus gammarus*)

Erik Sandertun Røed [1, 2]

[1] Centre for Integrative Genetics, Faculty of Biosciences, NMBU

[2] Faculty of Environmental Sciences and Natural Resource Management, NMBU

Abstract

A reliable, non-lethal age estimation method is needed to inform sustainable management of European lobster fisheries, which are at a historic low. To this end, a study recently found that ribosomal DNA methylation at 46 loci can predict age accurately and precisely in young hatchery-reared lobsters. While promising, this "epigenetic clock" now requires validation in older and ideally wild lobsters, which are challenging to source reliably at known age. To circumvent the challenge of sourcing known-age individuals, we sourced samples from an ongoing long-term capture-recapture monitoring program. By measuring methylation at known intervals of up to three years in 48 lobsters, we tested whether the epigenetic clock reliably separates longitudinal replicate samples by the known age interval from capture to recapture. To search for differences in the methylation-aging relationship between our sample and the original study, we also performed a longitudinal epigenome-wide association scan for ageing-associated loci and constructed penalised regression models to predict the time elapsed between capture and recapture. The original epigenetic clock systematically underestimated capture-recapture intervals by on average 18.1 months. Other combinations of loci identified by the association scan and penalised regression models only somewhat better predicted ageing. More encouragingly, the loci identified as ageing-associated here were transferable as valid epigenetic clocks in the dataset of the original study, implying transferability between populations, age brackets, and tissues. Moreover, the systematic underestimation of ageing suggested that the ageing-methylation relationship in European lobster may be non-linear, which we argue is plausible and a useful consideration for future research.

Introduction

Fisheries for European lobster (*Homarus gammarus*, hereafter "lobster") are in decline (Kleiven et al. 2012, 2022, Sundelöf et al. 2013). To inform more sustainable lobster management practices, biologists must monitor the distribution of age classes in lobster populations so fluctuations in population recruitment- and replenishment can be related to the abundance of breeding adults, environmental variables, or anthropogenic factors such as historical harvest pressure (Sheehy 2001, Sheehy & Bannister 2002). However, traditional methods for lobster age estimation are laborious and require lethal sampling (Sheehy et al. 1996, Huntsberger et al. 2020). These constraints of traditional methods restrict their applicability, so lobster age distributions are rarely monitored in recovering or protected populations (see e.g. Moland et al. 2021). In lieu of ageing methods, body size can serve as an indirect indicator of age (Uglem et al. 2005). The generality of the age-size association in lobster, though, is questionable both due to indeterminate growth (Sheehy et al. 1999) and due to being directly influenced by fisheries activity (Sørdalen et al. 2022). Consequently, body size is best avoided as an indirect age-proxy to avoid mismanagement (Fonseca & Sheehy 2007). Given the shortcomings of traditional ageing methods and the inconsistent age-size relationship, demand is high for a lobster age estimation method that is non-lethal.

Because DNA can be sourced at minimal harm to lobsters from autotomising tissues such as antennae (Butler 2017), genetic methods are a promising avenue for circumventing the issues with traditional lobster ageing (Silva et al. 2019, Fairfield 2021). Among the most promising genetic ageing tools in molecular ecology to date are *epigenetic clocks* (Dunshea et al. 2011, Clercq et al. 2023, Pepke 2023). An epigenetic clock predicts age from a weighted sum of age-related epigenetic methylation levels at select cytosine-guanine dinucleotide (CpG) loci. The age-related CpG loci are typically identified from methylation array or bisulphite sequencing data, using machine learning (Anastasiadi & Piferrer 2023, Piferrer & Anastasiadi 2023). The first epigenetic clocks were developed to predict age in humans, e.g. for forensic applications (Bocklandt et al. 2011, reviewed in Horvath & Raj 2018, Rutledge et al. 2022), but in the following decade epigenetic clocks were also reported in studies of dozens animals and at least one plant (Gardner et al. 2022, Clercq et al. 2023, Tangili et al. 2023, Brink et al. 2024). In the context of fisheries management and conservation, epigenetic clocks are already emerging in commercially or culturally valuable species such as European seabass (Anastasiadi & Piferrer 2020), red grouper, Northern red snapper (Weber et al. 2022), and golden perch (Mayne et al. 2023). The combination of applicability in many taxa with non-lethal sampling and often high

reliability has promoted the development and use of epigenetic clocks as ageing tools in fisheries management generally (Piferrer & Anastasiadi 2023), including in management of lobster species (Silva et al. 2019).

Indeed, in the lobster, Fairfield et al. (2021) recently discovered that an epigenetic clock based on 46 CpG loci in the ribosomal DNA (rDNA) is predictably related to age. The rDNA was targeted as it has been identified as a potential species-independent epigenetic clock region, which becomes comparatively hypermethylated with greater age (Wang & Lemos 2019). Fairfield et al. (2021) reported that the epigenetic clock produced reliable age estimates in hatchery-reared lobsters of known age up to four years (estimated and known age $R^2 = 0.98$), even in a more cost-effective, reduced implementation ($R^2 = 0.95$). However, any age estimation method is trustworthy only within the age range for which the method is validated (Beamish & McFarlane 1983, Campana 2001). Therefore, as Fairfield et al. (2021) emphasise, a prerequisite for applying their findings in lobster management is to validate the epigenetic clock in animals across a range of ages representative of fished populations, ideally outside hatchery conditions.

The gold-standard for validation of an age estimation method is to test it on known-age individuals across the age range of interest (Campana 2001). However, obtaining known-age lobsters across their natural age range via hatchery-rearing or release-recapture programs demands considerable research continuity, because lobsters are one of the most longevous decapod species (Vogt 2019) and often live for 20 - 30 years even in fisheries (Sheehy et al. 1999, Sheehy & Bannister 2002). Validating a lobster ageing method with known-age lobsters across the species' natural age range is therefore likely to be prohibitively costly and time-consuming. Where acquiring known-age animals is impracticable, an ageing method could be validated longitudinally instead, by confirming that age estimates increase as expected over known time intervals (Campana 2001). In the context of epigenetic ageing, repeat-measurement has in fact recently been identified as an avenue for refining age prediction (Piferrer & Anastasiadi 2023), and has already been used to corroborate epigenetic clocks in e.g. green turtles (Mayne et al. 2022), roe deer (Lemaître et al. 2022), and the short-tailed shearwater (Paoli-Iseppi et al. 2019). If the epigenetic clock reported by Fairfield et al.(2021) were similarly validated longitudinally by age estimating wild lobsters at known time intervals, this should satisfy the pressing need for validation without necessitating known-age animals. Hence, in this study, we sought to longitudinally validate this lobster epigenetic clock.

We took advantage of an ongoing long-term mark-recapture program to obtain repeated longitudinal samples from individual wild lobsters. Our primary objective was to test whether Fairfield et al.'s (2021) rDNA epigenetic clock would precisely and accurately predict the known age intervals between repeated samples. In other words, we asked whether the predictive relationship between rDNA methylation and chronological age in lobsters is consistent both within and between individuals in an independent dataset from a wild population. Assuming the hypothesis that it is, we tested two predictions. First, within lobsters, replicate age estimates should on average agree with the known time interval between captures. Second, between lobsters, the rate of change of epigenetic age estimates should remain consistent across estimate values and potential covariates such as sex and size. To gauge the reproducibility of epigenetic clocks across age brackets and between tissues, and to investigate the relationship between rDNA methylation and ageing in older lobsters, we also scanned our data for age-associated CpG loci and tested their power to predict age in the original dataset of Fairfield et al. (2021).

Methods

Sample collection

Lobster samples were sourced from the ongoing long-term lobster mark-recapture program in compliance with all applicable laws and regulations under research exemptions granted by the Norwegian Directorate of Fisheries (permit reference numbers 20/11501, 21/6317, and 23/11763). The design and timing of the sampling was governed by the mark-recapture program, which is exhaustively described elsewhere (for details, see: Sharon 2022, Vold 2022, Haugen et al. 2023). In brief, all samples were collected between September 2020 and December 2023 from three locations in and around the Jetéen lobster sanctuary in the inner Oslofjord, Norway. One annual sampling period took place in September, before the regular South-East Norwegian lobstering season in October and November. From 2021, another annual sampling period took place in December after the lobstering season.

In each sampling period, standard parlour traps were set and hauled at daily intervals for five consecutive days. Catch recovered from each trap was processed at the location of capture and promptly re-released nearby. Lobsters above a total length of 15 cm (i.e. the length from anterior tip of rostrum to posterior margin of telson) were tagged with a uniquely numbered 60 mm FD-94 T-Bar tag (Floy-Tag) at the left ventral surface of the first abdominal segment, unless re-captured with an existing tag. The sex and total length of each lobster were

recorded, and a non-destructive tissue sample collected from the terminal margin of one uropod with surgical scissors sterilized in ≥ 70 % EtOH. To obtain longitudinal samples separated by known time intervals, replicate tissue samples were collected in September and December of 2023 from lobsters that were re-captured with an existing tag. The replicate samples were consistently collected from a previously un-sampled uropod to avoid any influence of tissue re-growth on rDNA methylation. A few lobsters were serendipitously re-sampled in 2022 and included on the assumption that the second tissue sample was most likely re-sampled from an un-sampled uropod. To prevent DNA degradation, all tissue samples were stored in 96 % EtOH at 4 °C for up to half a year, or -20 °C for longer term storage.

Sample selection

Records of lobster encounters from the monitoring program were filtered to select reliably identified lobsters that had been sampled at least twice. To limit the influence of technical error, we excluded individuals with repeat samples separated by less than 350 days, given an expected mean epigenetic ageing precision of ca. 50 days (Fairfield et al. 2021). Then, from a total of 48

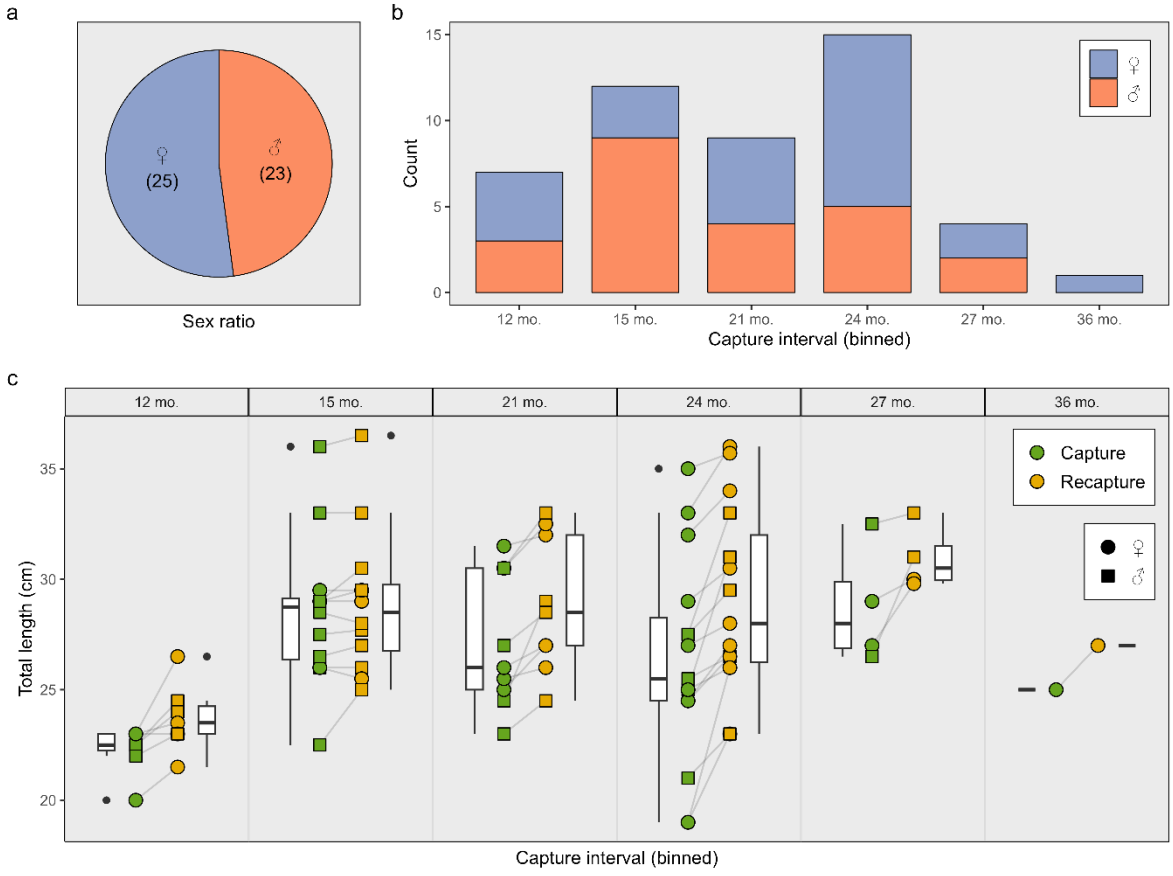

Figure 1 Lobsters (N = 48) included in this study are summarised in **a)** a pie chart of sex ratio, **b)** pie charts of sex ratios and counts per capture interval, and **c)** dot and box plots of total length. Lines connect the two replicate measurements of each lobster at capture and recapture. Mo.: Months.

lobsters, we selected two replicate samples separated by the greatest available time interval. We prioritised including lobsters with greater recapture intervals, and an unbiased distribution of sex and total length. Sex might influence age-related epigenetic changes in both vertebrates (Rubbi et al. 2022, Horvath et al. 2023) and invertebrates (Brink et al. 2024), whereas both sex and total length are potential age covariates in lobster (Sheehy et al. 1999, Uglem et al. 2005) that were feasible to control to increase the likelihood of sampling a broad age range.

A final selection of 48 lobsters was drawn from in total 73 reliably identified individuals with repeat samples that satisfied the minimum 350-day capture interval criterion. 41 lobsters were recaptured at intervals of 15 months and above, with a maximum interval of 1080 days (i.e. 36 months; fig. 1b). These 41 individuals were all included in the final selection. The remaining seven individuals were selected arbitrarily to increase the representation of smaller lobsters, from a total of 32 reliably identified individuals recaptured at 12-month intervals. The final sex ratio was balanced (fig. 1a). Total lengths ranged from 19.0 cm – 36.0 cm at capture and 21.5 cm – 36.5 cm at recapture (fig. 1c), which is representative of lobsters trapped at the sampling locality (Haugen et al. 2023) and in lobster sanctuaries elsewhere in Norway (Thorbjørnsen et al. 2018, Fernández-Chacón et al. 2020).

DNA extraction and quality control

From each tissue sample, genomic DNA (gDNA) was extracted from a small piece of soft tissue enclosed in exoskeleton using the DNeasy Blood & Tissue Kit (Qiagen) with the following adjustments to the manufacturer's protocol. To improve yield from the chitinous exoskeleton, the Proteinase K lysis step was extended to 90 minutes on a shaker-incubator at 1 000 rpm; remaining tissue was pelleted by pulse centrifugation and only the supernatant retained for gDNA purification. Purified gDNA was eluted in 60 µl of elution buffer pre-heated to 56 °C. Fragment sizes were checked by running 1 µl gDNA aliquots on an agarose gel stained with RedSafe nucleic acid staining solution (Bulldog-Bio) and imaged with the ChemiDoc XRS+ (BioRad). Finally, a 2 µl aliquot was drawn from each sample to measure concentration and ultraviolet absorbance on the NanoDrop Eight (Thermo Fisher Scientific), using the double stranded DNA setting in the manufacturer's software.

Quantification of rDNA methylation

Quantification of rDNA methylation by targeted bisulphite sequencing was performed by Zymo Research as described by Fairfield et al. (2021), except that the primers designed and validated for that study were repurposed here without need for re-design or re-validation. Sequencing

reads were aligned against Fairfield et al.'s (2021) partial sequences of the 18S, ITS1, ITS2, and 28S regions of the lobster rDNA. Zymo's bioinformatic pipeline for quantification of DNA methylation made use of the quality control utility FastQC v. 0.11.9 (Andrews 2020), the read trimmer Cutadapt (Martin 2011) implemented via Trim Galore! v. 0.6.6 (Krueger 2020), the bisulphite sequencing aligner Bismark v. 0.23.0 (Krueger & Andrews 2011), and the methylation extractor MethylDackel v. 0.5.2 (Ryan 2021). Results were aggregated with MultiQC v. 1.11 (Ewels et al. 2016). Methylation levels were reported as β-values, i.e. the observed fraction of methylated cytosines per CpG across all copies of the site. Except where stated otherwise, we logit-transformed methylation levels from β-values to M-values for all analyses to avoid heteroscedasticity at extreme values (Du et al. 2010).

Differential mean methylation analysis

Without initially focusing on age prediction *per se*, we first assessed whether the recapture samples were on average hypermethylated relative to the capture samples, as is expected with greater age (Wang & Lemos 2019). Overall methylation of each sample was calculated simply as the mean M-value across all CpG loci. The mean M-values were then regressed with a linear mixed effect model over months-since-first-capture (i.e. zero at capture) for each sample; effect estimates can be converted to β-values with little bias if needed (Xie et al. 2018). We specified models including time since first capture, total length, and sex as possible fixed effects, with and without interaction between total length and sex. Lobster identity and the month of capture nested within year of capture were fit as crossed random effects. To select the best model of average methylation with time, singular random effects were first removed (Barr et al. 2013). The remaining models were scored by the corrected Akaike Information Criterion (AIC_c) (Akaike 1974, Hurvich & Tsai 1989).

Estimation of sample ages and lobster age intervals

Next, we evaluated the reliability of the epigenetic clock model described by Fairfield et al. (2021). First, the rDNA methylation age in months of each sample was estimated as the linear combination of 46 elastic net coefficients multiplied by their respective CpG site β-values (Fairfield et al. 2021). Because Fairfield et al. (2021) reported no model intercept, we concerned ourselves only with the difference between capture and recapture and therefore centred the age estimates by subtracting the median of all capture sample age estimates. The change in estimated age between capture and recapture was tested with a linear mixed effect model: age

estimates were regressed with capture versus recapture, total length, sex, and the interaction of total length and sex as fixed effects, and lobster identity as a random effect.

For each lobster, the interval between the estimated age at capture and that at recapture was compared to the known interval between the two sampling time points. The mean difference between estimated and known intervals was tested with a paired t-test; the model is corroborated if the estimated age intervals on average match the known capture intervals with reasonable accuracy. Moreover, if estimation error is unbiased for (estimated) age and the covariates (i.e. sex and total length), the model ages consistently between individuals. We tested this by modelling the interval error as a multiple regression of the age estimate at capture, mean total length, sex, and the interaction of mean total length and sex. To uniformly scale interval errors, they were represented as relative errors, i.e. the absolute error of the estimated interval divided by the known interval.

EWAS of CpG loci *associated with months-since-first-capture*

To find which CpG loci were most associated with ageing in our data, we modelled the longitudinal relationship between methylation and ageing at each individual CpG in a longitudinal epigenome-wide association scan (EWAS) (for relevant examples, see: Lemaître et al. 2022, Refn et al. 2023). Specifically, the M-value at each CpG was modelled as a response of months-since-first-capture in a linear mixed effect model. Individual differences in baseline methylation at each locus were accounted for by including lobster identity as a random effect; this is why the level of methylation, not the amount of ageing, was taken as the response variable (following e.g. Wang et al. 2018, Refn et al. 2023). To remove as much ageing-extraneous variability as possible, we also included the maximal fixed and random effect structure described for the differential methylation analysis above, without model selection or removal of singular random effects. We finally applied a relatively relaxed (cf. Saffari et al. 2018) significance adjustment and threshold by adjusting the false discovery rate (Benjamini & Hochberg 1995) to $p = 0.05$, as the results of the EWAS were primarily used as feature pre-selection for penalised regression models.

Predicting months-since-first-capture with penalised regression models

Changes in methylation at CpG loci other than those selected in lobsters less than four years old by Fairfield et al. (2021) might better predict ageing in older individuals. To investigate this, we trained two penalised regression models *de novo* to predict the relative amount of ageing, i.e. months-since-first-capture, of each sample included in this study. The CpG loci available as

predictors were unrestricted in the first model, and then restricted in the second model to only the significantly ageing-associated CpG loci from the EWAS. These two models were compared to each other and the original epigenetic clock (Fairfield et al. 2021) by the relative error of estimated age intervals.

Each of the two models was constructed following a typical epigenetic clock building procedure, e.g. as described by Anastasiadi & Piferrer (2023). A quarter of the samples were first held out as a testing set, stratified for months-since-first-capture. Based on the remaining three quarters of samples, i.e. the training set, the available CpG loci were then filtered to remove near-zero-variance predictors using the default settings of the nearZeroVariance function in the caret R package (Kuhn 2008). Similarly, highly correlated CpG loci were filtered using caret's findCorrelation function with exact re-calculation of correlations upon removal of a predictor, to retain more CpG loci, and otherwise default settings. The filtered predictor CpG loci in the training set were then scaled and centred to a mean of zero and standard deviation of one. An identical transformation was applied blindly to the predictors in the test set. Models were then trained across a $\log_{10}$ sequence of 100 λ penalty values from 0.001 to 1000 with 10 repeated 10-fold cross validations using caret's train function and the glmnet model implementation (Friedman et al. 2010). For consistency across models, the α-parameter was held at one-half, i.e. elastic net regression (Zou & Hastie 2005). This choice of α aligned with Fairfield et al. (2021) and is a reasonably conventional default for epigenetic clock building both in the biomedical sciences (e.g. Horvath 2013, Bernabeu et al. 2023) and in molecular ecology (e.g. Wilkinson et al. 2021, Lemaître et al. 2022, Robeck et al. 2023). Optimal λ values, i.e. the final model tunings, were selected by minimising root mean square error (RMSE).

Predicting known age using CpG loci *associated with months-since-first-capture*

Last, to independently control the age-association of CpG loci selected by the EWAS and elastic net regression, epigenetic clocks restricted to these loci were trained on the known-age data published by Fairfield et al. (2021). These methylation data were first converted from β-values to M-values; to assess the influence this conversion, we also re-trained an epigenetic clock constrained to the original 46 loci (Fairfield et al. 2021). Many loci in the 28S region were missing from the known-age methylation data provided by Fairfield et al. (2021) (likely a file conversion incident); those CpG loci were disregarded without replacement. Otherwise, epigenetic clock models were trained as described above. The known-age methylation dataset of Fairfield et al. (2021) and the capture-recapture methylation dataset collected in this study were

finally combined. A principal components analysis (PCA) over the CpG loci with non-zero coefficients in each of the epigenetic clocks was used to assess whether most of the variation in methylation at those CpG loci was explained by a shared time-like dimension in both datasets.

Data analysis and visualisation

Data processing, analyses, and graphics were implemented with R v. 4.3.3 (R Core Team 2024), unless stated otherwise. For data wrangling we made use of the following R packages from the tidyverse ecosystem (Wickham et al. 2019): dplyr v. 1.1.4 (Wickham et al. 2023), lubridate v. 1.9.3 (Grolemund & Wickham 2011), tidyr v. 1.3.1 (Wickham et al. 2024), magrittr v. 2.0.3 (Bache & Wickham 2022), readxl v. 1.4.3 (Wickham & Bryan 2023), and stringr v. 1.5.1 (Wickham 2023). Graphics were produced using the R packages ggplot2 v. 3.5.0 (Wickham 2016), ggh4x v. 0.2.8 (Brand 2024), ggnewscale v. 0.4.10 (Campitelli 2024), ggvenn v. 0.1.10 (Yan 2023), ggsignif v. 0.6.4 (Ahlmann-Eltze & Patil 2021), and patchwork v. 1.2.0 (Pedersen 2024), with colour palettes sourced from viridis v. 0.6.5 (Garnier et al. 2024) and RColourBrewer v. 1.1.3 (Harrower & Brewer 2003, Neuwirth 2022). Additionally, linear mixed effect models were implemented in R with the function lmer of package lmerTest v. 3.1.3 (Kuznetsova et al. 2017) and the corrected Akaike Information Criterion calculated with the function AICc of R package MuMIn v. 1.47.5 (Barton 2023). All penalised regressions were built using the R package rsample v. 1.2.1 (Frick et al. 2024) in addition to caret v. 6.0.94 (Kuhn 2008). The scripts used to perform the analyses described here and produce all the supporting graphics are publicly available at https://github.com/ErikSRoeed/lobmsc.

Results

Quantification of rDNA methylation

Quantification of methylation was successful with sufficient coverage (i.e. ≥ 10×, as per Fairfield et al. 2021) in all 48 × 2 samples at 442 CpG loci across the 18S, ITS1, ITS2, and 28S rDNA regions (fig. 2a), which is 87 more loci than available to originally train the lobster epigenetic clock (Fairfield et al. 2021). Bisulphite conversion rates were 99.7 % across all samples, satisfying the typical conversion target of 99 % (Anastasiadi & Piferrer 2023). The mean coverage was 3804 and the minimum coverage of any CpG in any sample was 109 (fig. S1), so no loci, samples, or lobsters were excluded from downstream analyses. M-values across all CpG loci ranged from -8.27 ($\beta < 10^{-2}$) to 3.43 ($\beta = 0.92$) (fig. 2a - 2c).

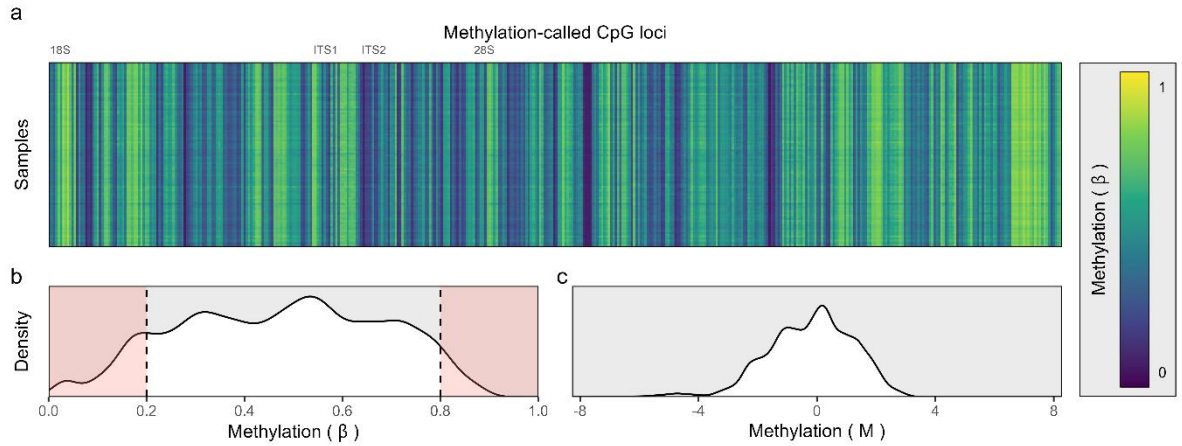

Figure 2 Methylation was quantified at 442 CpG loci across the ribosomal DNA. Panel **a)** is a heatmap of methylation β-values across all 96 samples (arbitrarily ordered, two per lobster), and panel **b)** gives the total distribution of β-values. The shaded areas bordered by dotted lines indicate value ranges that are prone to heteroscedasticity and therefore could be poorly represented in Gaussian statistical models (Du et al. 2010). Finally, panel **c)** shows the distribution of methylation values after logit-conversion to M-values following Du et al. (2010).

Table 1: Linear mixed models of mean methylation ranked by AIC_c. Mean M-value was the response in all models. Formulae are written in the nomenclature of the R package lmerTest (Kuznetsova et al. 2017). Abbreviations are DF: Degrees of freedom and AIC_c: Corrected Akaike Information Criterion.

Model	**DF**	**AIC_c**	**ΔAIC_c**	**Note**
Months + (1 | Lobster identity)	4	-152.01	0	Selected (next $\Delta AIC_c > 2$)
Months + Sex + (1 | Lobster identity)	5	-145.13	6.88	
Months + Length + (1 | Lobster identity)	5	-142.16	9.85	
Months + Length + Sex + (1 | Lobster identity)	6	-135.14	16.88	
Months + Length * Sex + (1 | Lobster identity)	7	-124.59	27.42	

Mean differential methylation analysis

On average, the within-individual mean methylation across all CpG loci increased from capture to recapture (fig. 3a and 3b). At capture, the within-individual mean M-value ranged from -0.43 (β = 0.44) to 0.09 (β = 0.52); the mean of means was -0.24 (β = 0.47) with a standard deviation (SD) of 0.11 (β = 0.02). At recapture, these values ranged from -0.45 (β = 0.44) to 0.08 (β = 0.52) with a mean of -0.20 (β = 0.48) and SD of 0.12 (β = 0.02). The best fit linear mixed effect model of M-values as a response to months-since-first-capture included only months-since-first-capture as a fixed effect (table 1). The random effect of months nested within years was excluded, as it was singular. The increase in methylation with months-since-first-capture was

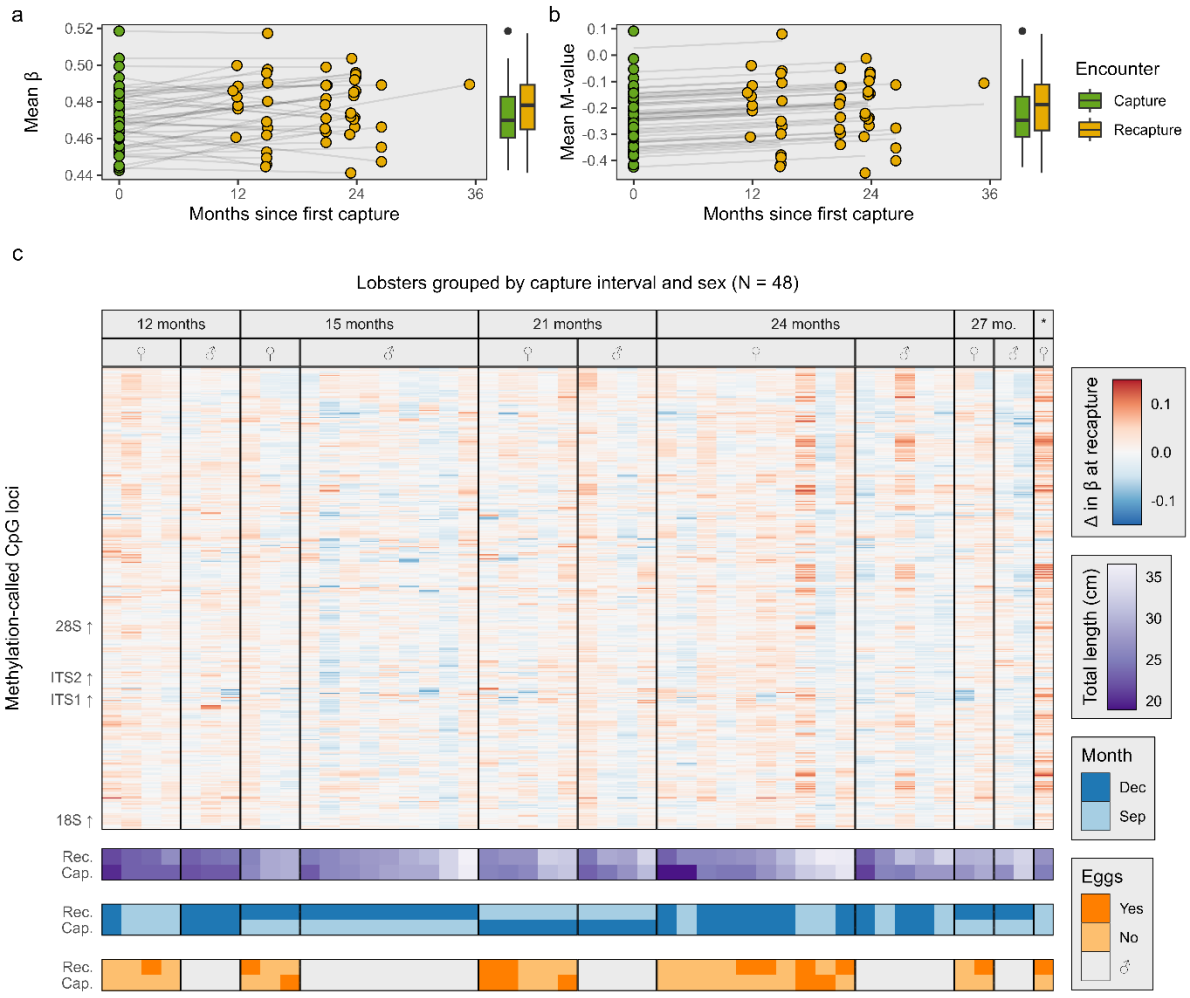

Figure 3 Change in methylation from capture to recapture is illustrated **a)** as a scatterplot of mean methylation (β) versus months-since-first-capture where lines connect replicate observations of the same lobster. For comparison, panel **b)** shows change in mean M-values with lines connecting predictions for each lobster from the best fit linear mixed model. The increase in methylation with months since first capture was significant ($p < 0.01$). Panel **c)** is a heatmap of change in methylation (β) at each locus, from capture to recapture. The presence or absence of eggs in females is indicated here but was not analysed. Abbreviations are Rec: Recapture, Cap: Capture, Mo: Months, and *: 36 months.

statistically significant (effect size = 0.0018, standard error, SE = 0.0006, $p < 0.01$), in accord with the expectation that rDNA is generally hypermethylated with age (Wang & Lemos 2019).

Despite the observed trend of rDNA hypermethylation with time, there was considerable between- and within individual variation in the difference of methylation from capture to recapture (fig. 3a and 3c). In the lobster with the greatest increase in mean methylation, the mean M-value rose by 0.22 ($\beta = 0.03$) from capture to recapture, but methylation still decreased at 73 CpG loci. Conversely, the lobster with the greatest decrease in mean methylation saw an

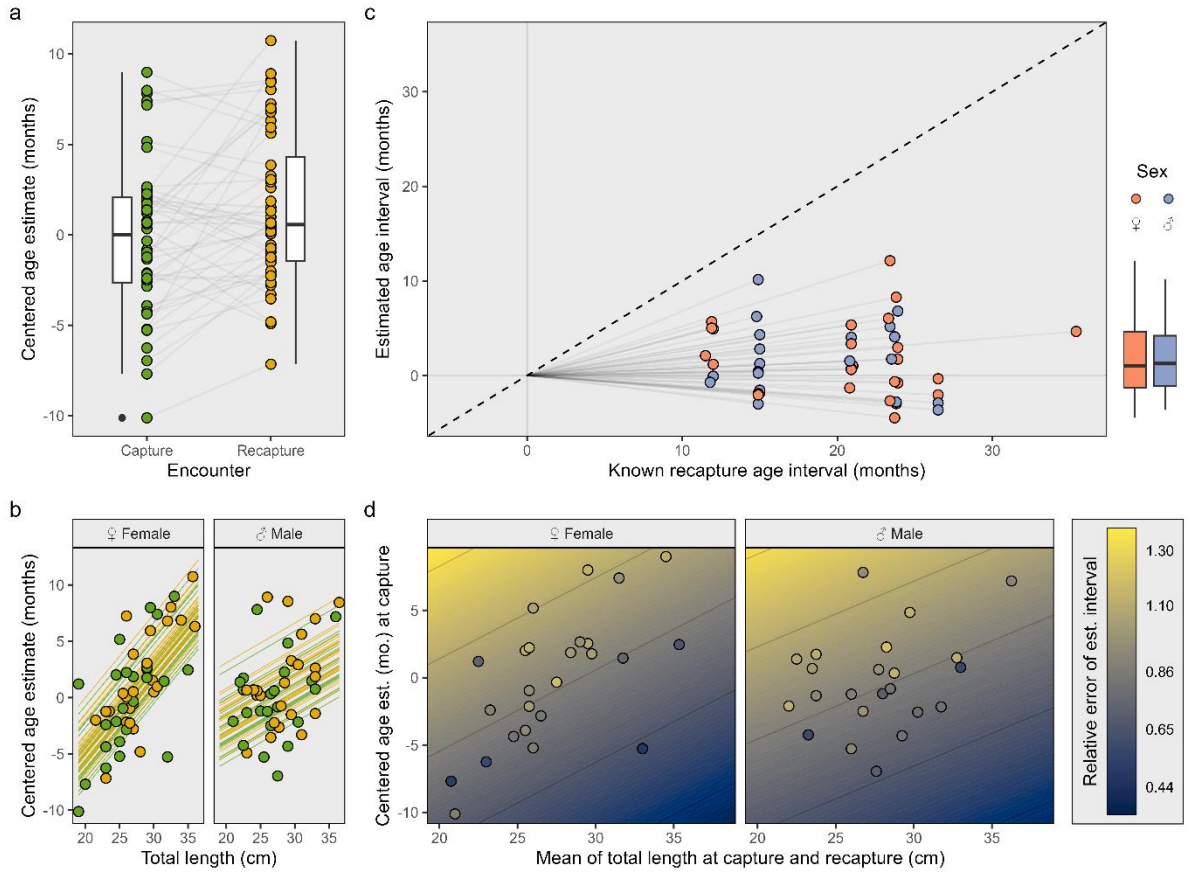

Figure 4 Age estimates of wild adult lobsters at capture and recapture are shown in panel **a)** by a dot and box plot. Observations of the same lobster are connected by lines. The effect of encounter was not significant ($p > 0.1$) in the linear mixed effect model illustrated in panel **b)**, where lines are the capture and recapture random intercept predictions per individual with colouration as in panel **a**. Panel **c)** presents the estimated age intervals, i.e. recapture minus capture, compared to known age intervals. Finally, panel **d)** shows actual relative errors (dots) over relative errors predicted in a multiple regression (coloured contour plot). Abbreviations are Est: Estimate and Mo: Months.

increase in methylation at 38 CpG loci while its mean methylation decreased by 0.13 ($\beta = 0.02$) from capture to recapture.

Estimation of sample ages and lobster age intervals

Although mean methylation changed significantly with time, the ages predicted with the existing epigenetic clock (Fairfield et al. 2021) systematically underestimated the known age intervals between capture and recapture (fig. 4). The median-centred age-estimates ranged from -10.1 to 9.0 (mean ± SD = -0.2 ± 4.3) at capture and -7.2 to 10.7 (mean ± SD = 1.5 ± 4.2) at recapture (fig. 4a), but a mean effect size of 0.7 months attributed to the capture versus recapture fixed effect was not statistically significant in the linear mixed model (SE = 0.6, $p > 0.1$). Total length, however, had a significant effect (effect size = 0.7, SE = 0.2, $p < 10^{-5}$) (fig. 4b). Neither sex (effect size = 10.3, SE = 5.8, $p < 0.1$) nor the interaction of total length and sex (effect size

= -0.4, SE = 0.2, $p < 0.1$) had a significant effect at $p < 0.05$ level, but both were marginally significant (fig. 4b). Given the subtle mean change in age estimates from capture to recapture, the estimated age intervals were on average positive but 18.1 months shorter than the known age intervals (paired t-test, t = -18.5, degrees of freedom, DF = 47, $p < 10^{-15}$) (fig. 4c). There was no evidence for a difference of estimated age intervals between the sexes (Welch two-sample t-test, t = -0.003, DF = 46, $p > 0.5$).

The relative errors of the estimated age interval were biased (fig. 4d). In the linear model of relative errors (adjusted $R^2 = 0.27$, $p < 0.01$), greater estimated age at capture was a significant predictor of greater relative error (effect size = 0.032, SE = 0.007, $p < 10^{-4}$). Mean total length was marginally significant (effect size = -0.019, SE = 0.011, $p < 0.1$) while sex (effect size = -0.10, SE = 0.39, $p > 0.5$) and the interaction of mean total length and sex (effect size = 0.004, SE = 0.014, $p > 0.5$) were both not significant.

EWAS of CpG loci *associated with months-since-first-capture*

Months-since-first-capture was significantly predictive of methylation at 28 CpG loci in the EWAS (fig. 5a). The CpG loci with a significant effect of months-since-first-capture were located in the 18S, ITS2, and 28S regions. Overall, the effect size of months-since-first-capture was positive at most CpG loci and especially at the CpG loci where the effect was significant (fig. 5b). In a PCA over M-values at the 28 CpG loci with a significant effect of months-since-first-capture, the first two principal components (PC) explained 74.2 % of variation in methylation (fig. 5c). There was a significant directionality from capture to recapture at both PC1 (paired t-test, mean difference = 0.43, t = 5.4, DF = 47, $p < 10^{-5}$) and PC2 (paired t-test, mean difference = 0.08, t = 2.2, DF = 47, $p < 0.05$), although not all individuals conformed to it. There were 10 individuals that moved opposite of the PC1 trend: five males and five females captured at intervals of 14.8 to 26.5 months with mean total length (i.e. mean of capture and recapture) of 30.2 ± 2.7 cm (mean ± SD) and growth of 1.3 ± 1.1 cm (mean ± SD).

Predicting months-since-first-capture with penalised regression models

For comparison against the age interval estimates of the original (hereafter “Fairfield”) model of Fairfield et al. (2021), we trained two elastic net regression models to predict months-since-first-capture (fig. 6). The first model (hereafter “Elastic Net”, abbreviated “EN”) training process allowed all 442 CpG loci as predictors, of which 378 were retained after filtering for near-

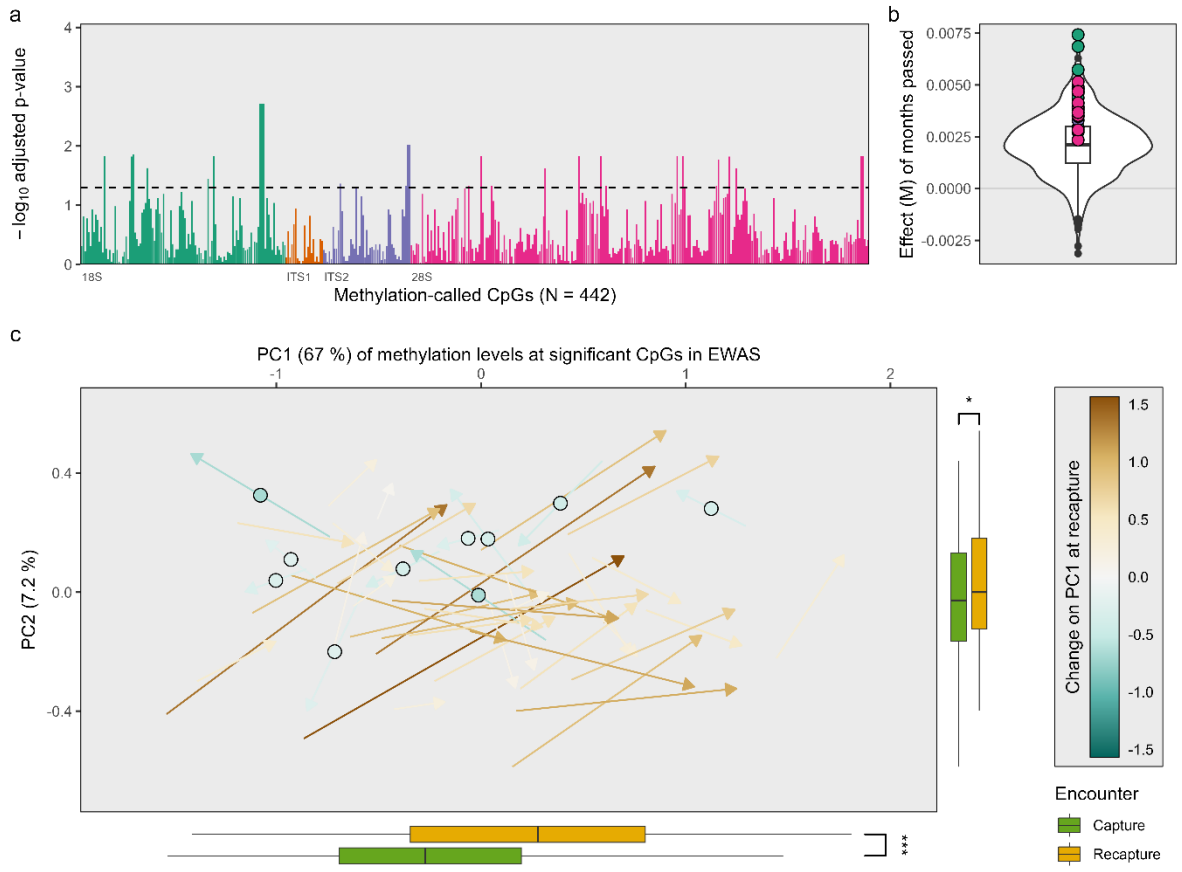

Figure 5 The results of an EWAS of methylation as a response to time are shown in panel **a)** with a Manhattan plot of adjusted p-values per CpG locus. CpG loci (bars) are coloured by region. The horizontal dotted line indicates the p = 0.05 significance threshold. Panel **b)** is a violin plot of effect size, with dots for loci where methylation M-value was significantly predicted by months-since-first-capture ("months passed"). Dots are coloured by region as in panel **a**. Last, panel **c)** illustrates a principal component analysis of methylation at the significant CpG loci. Arrows connect capture and recapture samples of each individual lobster and are coloured according to change on principal component (PC) 1. Individuals that moved opposite the trend on PC1 are labelled with dots. Paired t-test significance levels are *: $p < 0.05$ and ***: $p < 0.001$.

zero variance and multicollinearity. The second model (hereafter "EWAS + Elastic Net", abbreviated "EWAS + EN") training process allowed as predictors only the 28 CpG loci identified as significantly associated with months-since-first-capture in the EWAS (fig. 5); 24 were retained after filtering. At its optimal value of $\lambda = 1.874$, the Elastic Net model selected 28 CpG loci as predictors, with a roughly even split of positive and negative coefficients (fig. 6a and 6f). Three CpG loci were selected in the EWAS + Elastic Net model ($\lambda = 1.874$), which were a subset of the CpG loci selected by the Elastic Net model (fig. 6a and 6h). Only three CpG loci were shared between the Fairfield and Elastic Net models, and none were shared between the Fairfield and EWAS + Elastic Net models (fig. 6a).

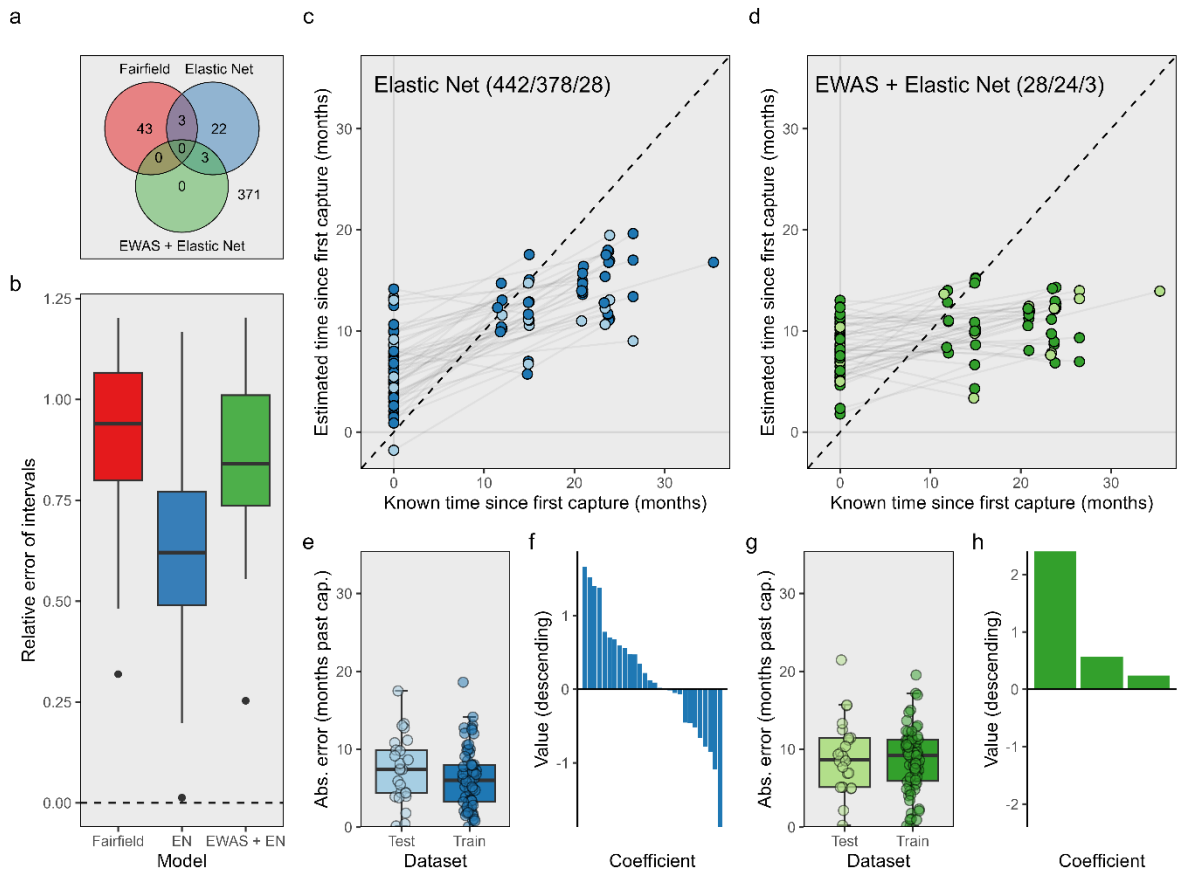

Figure 6 Penalised regressions of months-since-first-capture are compared to the original epigenetic clock in panel **a)**, a Venn diagram of non-zero elastic net coefficients, and panel **b)**, a box plot of relative error of estimated age intervals. Colouration is consistent throughout. Panels **c)** and **d)** show predictions from the EN and EWAS + EN model, respectively. Numbers in parentheses are (available CpG loci / after filtering / with non-zero coefficient), and the dotted diagonal indicates equality of known and estimates values. Dots are coloured to indicate whether they belong to the training or testing dataset as per colouration in panels **e)** and **g)**, which show box plots of absolute prediction error. Panels **f)** and **h)** finally show the coefficient values of each model, arranged in descending order. EN: Elastic Net, Abs. error: Absolute error, and Cap: Capture.

Choice of model significantly influenced the relative error of predicted intervals (ANOVA, DF = 2, F = 24.3, $p < 10^{-9}$). The Fairfield and EWAS + Elastic Net models were not significantly different (*post hoc* Tukey multiple comparison of means, $p > 0.5$), whereas the Elastic Net model interval relative error was 0.27 ($p < 10^{-7}$) and 0.23 ($p < 10^{-6}$) lower than the Fairfield model and EWAS + Elastic Net models, respectively. Nonetheless, the Elastic Net model still systematically underestimated the known age intervals by 12.7 ± 5.6 months (MAE ± SD) (fig. 6b).

The systematic error of estimated age intervals appeared to be driven by overestimates of months-since-first-capture at capture and underestimates at recapture (fig. 6c, 6d). Accordingly, the Elastic Net model yielded a high mean absolute error (MAE) = 8.9 months (fig. 6e)

and RMSE = 10.2 months, but a low $R^2 = 0.23$. By comparison, an oft cited benchmark for a true epigenetic clock is a Pearson correlation (i.e. R) of 0.80 (Horvath & Raj 2018, Anastasiadi & Piferrer 2023). For the considerably sparser EWAS + Elastic Net model, the same metrics were MAE = 8.9 months, RMSE = 9.8 months, and $R^2 = 0.21$.

Predicting known age using CpG loci *associated with months-since-first-capture*

Finally, we investigated whether the CpG loci identified as most ageing-associated here also functioned as epigenetic clocks in the independent dataset of known-age lobsters from Fairfield et al. (2021). As a benchmark for comparison, an epigenetic clock constrained to use the CpG loci included in Fairfield et al.'s (2021) original epigenetic clock unsurprisingly recapitulated their results closely, notwithstanding the lack of nine CpG loci in the known-age data and the conversion to M-values (fig. 7a). From 37 available CpG loci, 30 were retained after filtering and 25 had non-zero coefficients at the optimal value of $\lambda = 0.231$ where RMSE = 3.6 months, MAE = 2.8 months, and $R^2 = 0.96$ (c.f. $R^2 = 0.98$ in Fairfield et al. 2021). Next, only 14 of the CpG loci with non-zero coefficients in the Elastic Net (i.e. without EWAS CpG selection) model of months-since-first-capture were available in the known-age data. All 14 were retained after filtering and had non-zero elastic net coefficients at $\lambda = 0.267$ in an epigenetic clock (fig. 7c) with RMSE = 5.5 months, MAE = 4.5 months, and $R^2 = 0.89$. A last epigenetic clock (fig. 7e) was constrained to use the 28 CpG loci that were significantly associated with months-since-first-capture in the EWAS (fig. 5), of which 18 were available in the known-age data. Four were filtered, and nine retained non-zero coefficients at the optimal value of $\lambda = 0.705$. This last epigenetic clock based on the EWAS CpG loci had the highest RMSE (6.4 months) and MAE (5.2 months), and the lowest R^2 (0.86).

In PCAs of methylation at the CpG loci included in each epigenetic clock, the two first principal components explained over 80 % of variability in the methylation data (fig. 7b, 7d, and 7f). A distinct, clustered distribution of known-age individuals was shared between the first two principal components of the Fairfield (fig. 7b) and Elastic Net (fig. 7d) CpG loci, though neither PC1 separated capture from recapture samples in the data from this study. On the other hand, the two first principal components of methylation at the EWAS CpG loci (fig. 7f) displayed a clear directionality from younger to older and had a statistically significant directionality from capture to recapture on PC1 (paired t-test, mean difference = -0.12, DF = 47, $t = -3.8$, $p < 0.001$). There was some overlap between the datasets where the oldest known-age lobsters gave way to capture samples from wild-caught lobsters of unknown age.

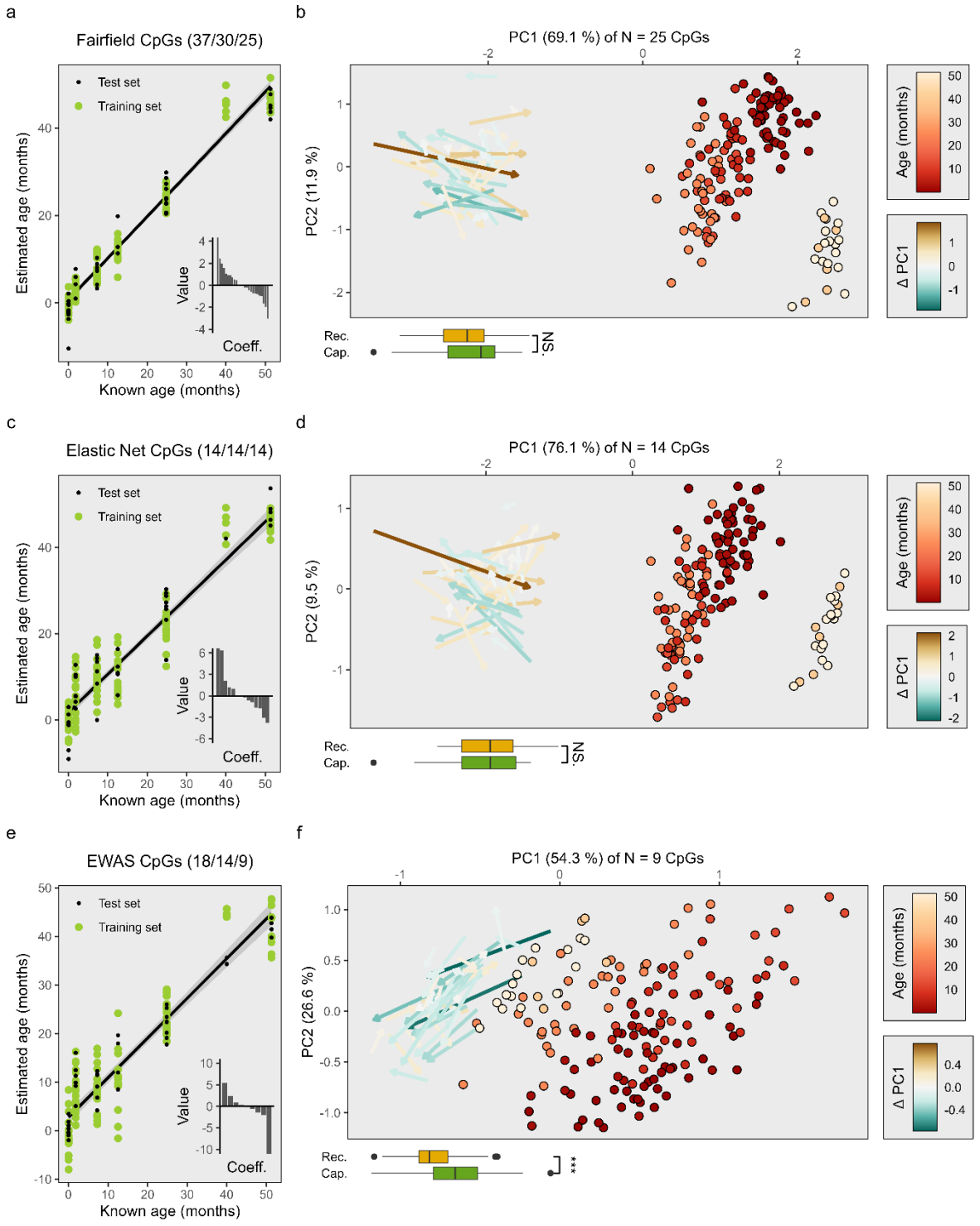

Figure 7 Predictions of epigenetic clocks trained on known-age data (Fairfield et al. 2021) are shown in panels **a)**, **c)**, and **e)**: numbers in parentheses are CpG loci (available / after filtering / with non-zero coefficients), and insets give non-zero coefficients (descending). Lines are linear models fit to the training set predictions, with 95 % confidence intervals. Panels **b)**, **d)**, and **f)** present the two first principal components from PCAs of methylation at the CpG loci with non-zero coefficients. Lobsters of known age (Fairfield et al. 2021) are shown as dots, and lobsters of unknown age (this study) as arrows from capture (Cap.) to recapture (Rec.). The CpG loci in panels **a, b)** were restricted to the 46 originally selected by Fairfield et al. (2021), of which 37 were available. In panels **c, d)** CpG loci were restricted to 28 CpG loci selected via elastic net (fig. 6; 14 available), and in panels **e, f)** restricted to the 28 significant CpG loci in the EWAS (fig. 5; 18 available). Paired t-test results were N.S.: $p > 0.05$ and ***: $p < 0.001$.

Discussion

With capture-recapture samples separated by up to three years, this study aimed to longitudinally test the epigenetic clock discovered in lobsters by Fairfield et al. (2021). If age intervals predicted by this epigenetic clock corresponded to the known age intervals between capture and recapture, this would validate the epigenetic clock as a valuable ageing method for lobster fisheries management (Campana 2001). Fairfield et al. (2021) originally trained and validated the epigenetic clock in lobsters up to four years of age. The lobsters included in our sample were likely older, so we have contributed a first attempt at validation for older age brackets, in the wild. Given the size range from ca. 20 – 36 cm total length, the smallest and largest lobsters in our sample were likely at least ca. 4 and 10 years old, respectively (Uglem et al. 2005). It is noteworthy, too, that these minimum age estimates based on body size should be considered conservative, as lobsters of comparable size could be decades old (Sheehy et al. 1999).

Our results indicated that rDNA mean methylation increased from capture to recapture (fig. 3), which is associated with increasing age (Wang & Lemos 2019). Nevertheless, none of our primary predictions about the epigenetic clock held. First, known age intervals were systematically underestimated (fig. 4c). Despite a small average difference in age estimates from capture to recapture (fig. 4a), the increase was insignificant when length and sex were accounted for as covariates (fig. 4b). Second, relative age interval errors were greater for lobsters with higher age estimates at capture (fig. 4d), contrary to our prediction that error ought to be unbiased. These results were not entirely unexpected, given the extent to which we likely extrapolated beyond the age range of the original training set (Mayne et al. 2021a). Moreover, some inconsistencies could also be explained by environmental or physiological differences between capture and recapture, e.g. in the individuals misaligned with the trend in fig. 5c (Poganik et al. 2023). Even so, our results stood in striking contrast to precise and accurate age estimates in young, hatchery-reared lobsters (Fairfield et al. 2021). Fairfield et al. (2021) identified two unanswered questions that further shed light on the discrepancy between our results and theirs, while pointing towards useful findings in this study and potential future avenues of research.

It was not established whether the predictive relationship between rDNA methylation and age in lobsters is tissue-independent (Fairfield et al. 2021). And, whereas the epigenetic clock was originally trained on rDNA methylation in claw tissue (Fairfield et al. 2021), we sourced DNA from uropod tissue for consistency with previously collected samples. As the

animals in our study were re-released after sampling, uropod tissue was also a minimally invasive choice. Earlier age estimates from antennal tissue in wild lobsters were remarkably low compared to expectation (Fairfield et al. 2021), which could have been symptomatic of poor cross-tissue applicability (as in e.g. Anastasiadi & Piferrer 2020). If so, our choice of tissue may have influenced our results. However, our results did not seem to indicate that tissue-specificity was a primary cause of the underestimated and biased age interval predictions we observed. Few of the CpG loci identified as most ageing-associated in our data overlapped with the original 46 CpG loci of Fairfield et al. (2021), yet they were highly predictive of age in Fairfield et al.'s dataset ($R^2 > 0.85$, fig. 7c and 7e). Nine CpG loci with non-zero elastic net coefficients selected in the original data also had a time-like pattern of samples on the first principal component, across both datasets (fig. 7f). As multiple combinations of CpG loci could constitute valid epigenetic clocks (Porter et al. 2021), future work should re-sample CpG loci in Fairfield et al.'s (2021) data to confirm that we were unlikely to select predictive epigenetic clocks across tissues by chance. Nevertheless, while none of the epigenetic clock models we tested predicted age intervals well in our data, the CpG loci we identified as least error-prone were valid epigenetic clocks in an independent dataset of methylation from another tissue.

Another possible contributing factor to our observations was that methylation at the CpG loci in the epigenetic clock either plateaued with age or had a non-linear relationship with age in older animals, contrary to the assumption of lifelong linearity (Fairfield et al. 2021). It did not appear that epigenetic clock methylation had plateaued or saturated in our sample. Estimated age was positively associated with total length, especially in females (fig. 4b). Given larger lobsters are usually older (Uglem et al. 2005), this corroborated that there was some non-negligible association between epigenetic clock methylation and age. The significant time-directionality on the first principal component through the methylome space of CpG loci identified in the EWAS also implied a methylation signal of ageing (fig. 5c). Time-directionality also appeared to carry over to Fairfield et al.'s (2021) original dataset (fig. 7f), where the pattern of samples was reminiscent of earlier observations in green turtles (Mayne et al. 2022). Nonetheless, the systematic underestimation of known age intervals (fig. 4a and 4c), coupled with greater underestimation at higher estimates (fig. 4d), and our inability to accurately predict months-since-first-capture from our own data (fig. 6), suggested that the gradient of age estimates declined with greater age. The sum of this evidence indicated that, while methylation change with age probably did not plateau in our sample, it could have plausibly been non-linear with age.

Counter to this, Fairfield et al. (2021) reasoned that if the methylation-age relationship is non-linear in lobsters, their epigenetic clock should overestimate ages in older individuals. As they argue, their observation that the epigenetic clock instead appeared to underestimate age in wild lobsters would then contradict non-linear epigenetic clock methylation in lobsters. However, we agree that Fairfield et al. (2021) were quite right to initially identify methylation non-linearity with age as a plausible concern; we speculate that a non-linear relationship between epigenetic clock methylation and chronological age would in fact explain their results and unify them with ours.

As Fairfield et al. (2021) discussed, non-linear epigenetic clock methylation is well-documented in humans (Horvath & Raj 2018). In humans, each added unit of age confers a logarithmically declining change in epigenetic clock methylation, which consequently changes most rapidly in young age (Snir et al. 2019). Snir et al. (2019) noted that this consistent logarithmic methylation deceleration may well explain observations that some epigenetic clocks "systematically underestimate" ages in older human tissues (Khoury et al. 2018, Marioni et al. 2019). Epigenetic clocks in multiple longevous species similarly have a propensity to overestimate younger individuals and underestimate older ones, including in cetaceans (Polanowski et al. 2014, Bors et al. 2020, Robeck et al. 2021, Peters et al. 2022), pinnipeds (Robeck et al. 2023), and trees (Gardner et al. 2022), although this does not seem to be a universal trend (Mayne et al. 2021b, 2022). It is not clear whether longevous arthropods might have an early-life phase of accelerated change in epigenetic clock methylation, because studies of ageing-related methylation in arthropods have focused on shorter-lived species such as hymenopterans (Cardoso-Júnior et al. 2018, Morandin et al. 2019, Renard et al. 2023, Brink et al. 2024) and the crustacean *Daphnia magna* (Hearn et al. 2021). Still, we can entertain the hypothesis that lobster methylation may also change more rapidly early in life. If so, an epigenetic clock trained only on young individuals would overestimate the amount of methylation change per unit of ageing in older individuals and, consequently, underestimate their age. This would be compatible both with our observations and the seemingly underestimated ages Fairfield *et* al. (2021) reported in wild lobsters, when using an epigenetic clock trained on methylation in young, hatchery-reared animals. If this hypothesis is accurate, we suggest non-linear modelling of ageing-related methylation could be a useful approach in future endeavours to identify and validate epigenetic clocks in lobsters or other longevous decapods.

Concluding remarks

To enhance monitoring of a declining fishery, we have leveraged a long-term longitudinal sampling program in an attempt to validate an epigenetic clock that enables non-lethal age estimation in European lobsters (*Homarus gammarus*). While the findings presented here unfortunately provided no such validation, they did shed additional light on two unanswered questions remarked by Fairfield et al. (2021). First, our findings suggested that ageing-related CpG loci in rDNA are transferable between tissues sampled from claw and uropod. This preliminary evidence of tissue-independence is a hopeful sign that perhaps one single, unified epigenetic clock could be built to serve the wider lobster management and conservation research community, who sample a variety of anatomical locations for practical reasons (Butler 2017, Jenkins et al. 2019). Second, we hypothesised based on our evidence that rDNA methylome change in the lobster is non-linear with age, which could serve as a useful consideration in future work. Still, additional work does indeed seem necessary to construct an epigenetic clock able to capture the relationship between rDNA methylation and age across the lobster lifespan. We enthusiastically encourage any such initiative. Finally, we hope that our longitudinal dataset of rDNA methylation in 48 lobsters can become a useful benchmark for validation of new epigenetic clocks.

Acknowledgements

The author wishes to thank Dr. Jun Soung Kwak (CIGENE, The Norwegian University of Life Sciences NMBU) for assistance with laboratory work, sample management, and logistics. Notably, to accelerate the sample shipment to Zymo Research for processing, Dr. Kwak kindly performed just above half of the DNA extractions and ran the electrophoresis gel. Furthermore, thanks are owed to Dr. Lars Grønvold (CIGENE) for valuable consultation on data analysis and visualisation. Many thanks are also well deserved by the numerous staff, students, and volunteers of NMBU's lobster sanctuary mark-recapture program, who have since 2020 diligently collected samples that were used in this study. Not only did they kindly invite the author's participation in roughly two weeks of fieldwork spread over three fieldwork seasons, but also helpfully adjusted their sampling protocols to facilitate this research. Finally, thanks to the many technical and administrative staff at NMBU whose assistance and advice ensured the logistics of this project ran smoothly despite unexpected delays and challenges.

References

Ahlmann-Eltze, C. & Patil, I. 2021. ggsignif: R Package for Displaying Significance Brackets for "ggplot2." PsyArxiv. doi:10.31234/osf.io/7awm6.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, **19**: 716–723. doi:10.1109/tac.1974.1100705.

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. doi:10.1111/1755-0998.13111.

Anastasiadi, D. & Piferrer, F. 2023. Bioinformatic analysis for age prediction using epigenetic clocks: Application to fisheries management and conservation biology. Frontiers in Marine Science, **10**: 1096909. doi:10.3389/fmars.2023.1096909.

Andrews, S. 2020. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bache, S.M. & Wickham, H. 2022. magrittr: A Forward-Pipe Operator for R. R package. https://CRAN.R-project.org/package=magrittr.

Barr, D.J., Levy, R., Scheepers, C. & Tily, H.J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, **68**: 255–278. doi:10.1016/j.jml.2012.11.001.

Barton, K. 2023. MuMIn: Multi-Model Inference. R package. https://CRAN.R-project.org/package=MuMIn.

Beamish, R.J. & McFarlane, G.A. 1983. The Forgotten Requirement for Age Validation in Fisheries Biology. Transactions of the American Fisheries Society, **112**: 735–743. doi:10.1577/1548-8659(1983)112<735:tfrfav>2.0.co;2.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), **57**: 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Bernabeu, E., McCartney, D.L., Gadd, D.A., Hillary, R.F., Lu, A.T., Murphy, L., Wrobel, N., Campbell, A., Harris, S.E., Liewald, D., Hayward, C., Sudlow, C., Cox, S.R., Evans, K.L., Horvath, S., McIntosh, A.M., Robinson, M.R., Vallejos, C.A. & Marioni, R.E. 2023. Refining epigenetic prediction of chronological and biological age. Genome Medicine, **15**: 12. doi:10.1186/s13073-023-01161-y.

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. doi:10.1371/journal.pone.0014821.

Bors, E.K., Baker, C.S., Wade, P.R., O'Neill, K.B., Shelden, K.E.W., Thompson, M.J., Fei, Z., Jarman, S. & Horvath, S. 2020. An epigenetic clock to estimate the age of living beluga whales. Evolutionary applications, **14**: 1263–1273. doi:10.1111/eva.13195.

Brand, T. van den. 2024. ggh4x: Hacks for "ggplot2." R package. https://CRAN.R-project.org/package=ggh4x.

Brink, K., Thomas, C.L., Jones, A., Chan, T.W. & Mallon, E.B. 2024. Exploring the ageing methylome in the model insect, *Nasonia vitripennis*. BMC Genomics, **25**: 305. doi:10.1186/s12864-024-10211-7.

Butler, M.J. 2017. Collecting and processing lobsters. Journal of Crustacean Biology, **37**: 340–346. doi:10.1093/jcbiol/rux021.

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. doi:10.1111/j.1095-8649.2001.tb00127.x.

Campitelli, E. 2024. ggnewscale: Multiple Fill and Colour Scales in "ggplot2." R package. https://CRAN.R-project.org/package=ggnewscale.

Cardoso-Júnior, C.A.M., Guidugli-Lazzarini, K.R. & Hartfelder, K. 2018. DNA methylation affects the lifespan of honey bee (*Apis mellifera* L.) workers – Evidence for a regulatory module that involves vitellogenin expression but is independent of juvenile hormone function. Insect Biochemistry and Molecular Biology, **92**: 21–29. doi:10.1016/j.ibmb.2017.11.005.

Clercq, L.L., Kotzé, A., Grobler, J.P. & Dalton, D.L. 2023. Biological clocks as age estimation markers in animals: a systematic review and meta-analysis. Biological Reviews **98**: 1972-2011. doi:10.1111/brv.12992.

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L. & Lin, S.M. 2010. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics, **11**: 587. doi:10.1186/1471-2105-11-587.

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. doi:10.1111/j.1755-0998.2010.02976.x.

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. doi:10.1093/bioinformatics/btw354.

Fairfield, E.A. 2021. Ageing the unageable: investigating novel methods of ageing crustaceans. PhD thesis. University of East Anglia, Norwich, United Kingdom.

Fairfield, E.A., Richardson, D.S., Daniels, C.L., Butler, C.L., Bell, E. & Taylor, M.I. 2021. Ageing European lobsters (*Homarus gammarus*) using DNA methylation of evolutionarily conserved ribosomal DNA. Evolutionary Applications, **14**: 2305–2318. doi:10.1111/eva.13296.

Fernández-Chacón, A., Villegas-Ríos, D., Moland, E., Baskett, M.L., Olsen, E.M. & Carlson, S.M. 2020. Protected areas buffer against harvest selection and rebuild phenotypic complexity. Ecological Applications, **30**: e02108. doi:10.1002/eap.2108.

Fonseca, D.B. & Sheehy, M.R. 2007. Does size matter? A cautionary experiment on overoptimism in length-based bioresource assessment. Canadian Journal of Fisheries and Aquatic Sciences, **64**: 996–1008. doi:10.1139/f07-076.

Frick, H., Chow, F., Kuhn, M., Mahoney, M., Silge, J. & Wickham, H. 2024. rsample: General Resampling Infrastructure. R package. https://CRAN.R-project.org/package=rsample.

Friedman, J., Hastie, T. & Tibshirani, R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, **33**(1)**:** 1-22. doi:10.18637/jss.v033.i01.

Gardner, S.T., Bertucci, E.M., Sutton, R., Horcher, A., Aubrey, D. & Parrott, B.B. 2022. Development of DNA methylation-based epigenetic age predictors in loblolly pine (*Pinus taeda*). Molecular Ecology Resources, **23**: 131–144. doi:10.1111/1755-0998.13698.

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedr, A., Sciain, Marc, Schere & Cédric. 2024. viridis(Lite) - Colorblind-Friendly Color Maps for R. R package. https://sjmgarnier.github.io/viridis/.

Grolemund, G. & Wickham, H. 2011. Dates and Times Made Easy with lubridate. Journal of Statistical Software, **40**(3): 1-25. doi:10.18637/jss.v040.i03.

Harrower, M. & Brewer, C.A. 2003. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. The Cartographic Journal, **40**: 27–37. doi:10.1179/000870403235002042.

Haugen, T.O., Colman, J.E. & Moe, S.R. 2023. Torsk og hummer i indre Oslofjord: Effekter av vernetiltak. MINA fagrapport 83. Norwegian University of Life Sciences, Ås, Norway.

Hearn, J., Plenderleith, F. & Little, T.J. 2021. DNA methylation differs extensively between strains of the same geographical origin and changes with age in *Daphnia magna*. Epigenetics & Chromatin, **14**: 4. doi:10.1186/s13072-020-00379-z.

Horvath, S. 2013. DNA methylation age of human tissues and cell types. Genome Biology, **14**: R115. doi:10.1186/gb-2013-14-10-r115.

Horvath, S., Haghani, A., Zoller, J.A., Lu, A.T., Ernst, J., Pellegrini, M., Jasinska, A.J., Mattison, J.A., Salmon, A.B., Raj, K., Horvath, M., Paul, K.C., Ritz, B.R., Robeck, T.R., Spriggs, M., Ehmke, E.E., Jenkins, S., Li, C. & Nathanielsz, P.W. 2023. Pan-primate studies of age and sex. GeroScience, **45**: 3187–3209. doi:10.1007/s11357-023-00878-3.

Horvath, S. & Raj, K. 2018. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nature Reviews Genetics, **19**: 371–384. doi:10.1038/s41576-018-0004-3.

Huntsberger, C.J., Kilada, R., Ambrose, W.G. & Wahle, R.A. 2020. Age-at-size relationships of the American lobster (*Homarus americanus*) from three contrasting thermal regimes using gastric mill band counts as a direct aging technique. Canadian Journal of Fisheries and Aquatic Sciences, **77**: 1733–1740. doi:10.1139/cjfas-2020-0018.

Hurvich, C.M. & Tsai, C.-L. 1989. Regression and time series model selection in small samples. Biometrika, **76**: 297–307. doi:10.1093/biomet/76.2.297.

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. doi:10.1111/eva.12849.

Khoury, L.E., Gorrie-Stone, T., Smart, M., Hughes, A., Bao, Y., Andrayas, A., Burrage, J., Hannon, E., Kumari, M., Mill, J. & Schalkwyk, L.C. 2018. Properties of the epigenetic clock and age acceleration. bioRxiv: 363143. doi:10.1101/363143.

Kleiven, A.R., Espeland, S.H., Stiansen, S., Ono, K., Zimmermann, F. & Olsen, E.M. 2022. Technological creep masks continued decline in a lobster (*Homarus gammarus*) fishery over a century. Scientific Reports, **12**: 3318. doi:10.1038/s41598-022-07293-2.

Kleiven, A.R., Olsen, E.M. & Vølstad, J.H. 2012. Total Catch of a Red-Listed Marine Species Is an Order of Magnitude Higher than Official Data. PLoS ONE, **7**: e31216. doi:10.1371/journal.pone.0031216.

Krueger, F. 2020. Trim Galore! https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Krueger, F. & Andrews, S.R. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics, **27**: 1571–1572. doi:10.1093/bioinformatics/btr167.

Kuhn, M. 2008. Building Predictive Models in R Using the caret Package. Journal of Statistical Software, **28**(5): 1-26. doi:10.18637/jss.v028.i05.

Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, **82**(13): 1-26. doi:10.18637/jss.v082.i13.

Lemaître, J., Rey, B., Gaillard, J., Régis, C., Gilot-Fromont, E., Débias, F., Duhayer, J., Pardonnet, S., Pellerin, M., Haghani, A., Zoller, J.A., Li, C.Z. & Horvath, S. 2022. DNA methylation as a tool to explore ageing in wild roe deer populations. Molecular Ecology Resources, **22**: 1002–1015. doi:10.1111/1755-0998.13533.

Marioni, R.E., Suderman, M., Chen, B.H., Horvath, S., Bandinelli, S., Morris, T., Beck, S., Ferrucci, L., Pedersen, N.L., Relton, C.L., Deary, I.J. & Hägg, S. 2019. Tracking the Epigenetic Clock Across the Human Life Course: A Meta-analysis of Longitudinal Cohort Data. The Journals of Gerontology: Series A, **74**: 57–61. doi:10.1093/gerona/gly060.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, **17**: 10–12. doi:10.14806/ej.17.1.200.

Mayne, B., Berry, O. & Jarman, S. 2021a. Optimal sample size for calibrating DNA methylation age estimators. Molecular Ecology Resources, **21**: 2316–2323. doi:10.1111/1755-0998.13437.

Mayne, B., Espinoza, T., Crook, D.A., Anderson, C., Korbie, D., Marshall, J.C., Kennard, M.J., Harding, D.J., Butler, G.L., Roberts, B., Whiley, J. & Marshall, S. 2023. Accurate, non-destructive, and high-throughput age estimation for Golden perch (*Macquaria ambigua* spp.) using DNA methylation. Scientific Reports, **13**: 9547. doi:10.1038/s41598-023-36773-2.

Mayne, B., Espinoza, T., Roberts, D., Butler, G.L., Brooks, S., Korbie, D. & Jarman, S. 2021b. Nonlethal age estimation of three threatened fish species using DNA methylation: Australian lungfish, Murray cod and Mary River cod. Molecular Ecology Resources, **21**: 2324–2332. doi:10.1111/1755-0998.13440.

Mayne, B., Mustin, W., Baboolal, V., Casella, F., Ballorain, K., Barret, M., Vanderklift, M.A., Tucker, A.D., Korbie, D., Jarman, S. & Berry, O. 2022. Age prediction of green turtles with an epigenetic clock. Molecular Ecology Resources, **22**: 2275–2284. doi:10.1111/1755-0998.13621.

Moland, E., Fernández-Chacón, A., Sørdalen, T.K., Villegas-Ríos, D., Thorbjørnsen, S.H., Halvorsen, K.T., Huserbråten, M., Olsen, E.M., Kleiven, P.J.N., Kleiven, A.R., Knutsen, H., Espeland, S.H., Freitas, C. & Knutsen, J.A. 2021. Restoration of Abundance and Dynamics of Coastal Fish and Lobster Within Northern Marine Protected Areas Across Two Decades. Frontiers in Marine Science, **8**: 674756. doi:10.3389/fmars.2021.674756.

Morandin, C., Brendel, V.P., Sundström, L., Helanterä, H. & Mikheyev, A.S. 2019. Changes in gene DNA methylation and expression networks accompany caste specialization and age-related physiological changes in a social insect. Molecular Ecology, **28**: 1975–1993. doi:10.1111/mec.15062.

Neuwirth, E. 2022. RColorBrewer: ColorBrewer Palettes. R package. https://CRAN.R-project.org/package=RColorBrewer.

Paoli-Iseppi, R.D., 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. doi:10.1111/1755-0998.12981.

Pedersen, T.L. 2024. patchwork: The Composer of Plots. R package. https://CRAN.R-project.org/package=patchwork.

Pepke, M.L. 2023. Telomere length is not a useful tool for chronological age estimation in animals. BioEssays: e2300187. doi:10.1002/bies.202300187.

Peters, K.J., Gerber, L., Scheu, L., Cicciarella, R., Zoller, J.A., Fei, Z., Horvath, S., Allen, S.J., King, S.L., Connor, R.C., Rollins, L.A. & Krützen, M. 2022. An epigenetic DNA methylation clock for age estimates in Indo-Pacific bottlenose dolphins (*Tursiops aduncus*). Evolutionary Applications, **16**: 126–133. doi:10.1111/eva.13516.

Piferrer, F. & Anastasiadi, D. 2023. Age estimation in fishes using epigenetic clocks: Applications to fisheries management and conservation biology. Frontiers in Marine Science, **10**: 1062151. doi:10.3389/fmars.2023.1062151.

Poganik, J.R., Zhang, B., Baht, G.S., Tyshkovskiy, A., Deik, A., Kerepesi, C., Yim, S.H., Lu, A.T., Haghani, A., Gong, T., Hedman, A.M., Andolf, E., Pershagen, G., Almqvist, C., Clish, C.B., Horvath, S., White, J.P. & Gladyshev, V.N. 2023. Biological age is increased by stress and restored upon recovery. Cell Metabolism, **35**: 807-820.e5. doi:10.1016/j.cmet.2023.03.015.

Polanowski, A.M., Robbins, J., Chandler, D. & Jarman, S.N. 2014. Epigenetic estimation of age in humpback whales. Molecular Ecology Resources, **14**: 976–987. doi:10.1111/1755-0998.12247.

Porter, H.L., Brown, C.A., Roopnarinesingh, X., Giles, C.B., Georgescu, C., Freeman, W.M. & Wren, J.D. 2021. Many chronological aging clocks can be found throughout the epigenome: Implications for quantifying biological aging. Aging Cell, **20**: e13492. doi:10.1111/acel.13492.

R Core Team. 2024. R: A Language and Environment for Statistical Computing. https://www.R-project.org/.

Refn, M.R., Andersen, M.M., Kampmann, M.-L., Tfelt-Hansen, J., Sørensen, E., Larsen, M.H., Morling, N., Børsting, C. & Pereira, V. 2023. Longitudinal changes and variation in human DNA methylation analysed with the Illumina MethylationEPIC BeadChip assay and their implications on forensic age prediction. Scientific Reports, **13**: 21658. doi:10.1038/s41598-023-49064-7.

Renard, T., Martinet, B., Araujo, N.D.S. & Aron, S. 2023. DNA methylation extends lifespan in the bumblebee *Bombus terrestris*. Proceedings of the Royal Society B, **290**: 20232093. doi:10.1098/rspb.2023.2093.

Robeck, T.R., Fei, Z., Haghani, A., Zoller, J.A., Li, C.Z., Steinman, K.J., Dirocco, S., Staggs, L., Schmitt, T., Osborn, S., Montano, G., Rodriguez, M. & Horvath, S. 2021. Multi-Tissue Methylation Clocks for Age and Sex Estimation in the Common Bottlenose Dolphin. Frontiers in Marine Science, **8**: 713373. doi:10.3389/fmars.2021.713373.

Robeck, T.R., Haghani, A., Fei, Z., Lindemann, D.M., Russell, J., Herrick, K.E.S., Montano, G., Steinman, K.J., Katsumata, E., Zoller, J.A. & Horvath, S. 2023. Multi-tissue DNA methylation aging clocks for sea lions, walruses and seals. Communications Biology, **6**: 359. doi:10.1038/s42003-023-04734-0.

Rubbi, L., Zhang, H., Feng, J., He, C., Kurnia, P., Ratan, P., Tammana, A., House, S., Thompson, M., Farrell, C., Snir, S., Stahler, D., Ostrander, E.A., vonHoldt, B.M. & Pellegrini, M. 2022. The effects of age, sex, weight, and breed on canid methylomes. Epigenetics, **17**: 1497–1512. doi:10.1080/15592294.2022.2069385.

Rutledge, J., Oh, H. & Wyss-Coray, T. 2022. Measuring biological age using omics data. Nature Reviews Genetics, **23**: 715–727. doi:10.1038/s41576-022-00511-7.

Ryan, D. 2021. MethylDackel. https://github.com/dpryan79/MethylDackel.

Saffari, A., Silver, M.J., Zavattari, P., Moi, L., Columbano, A., Meaburn, E.L. & Dudbridge, F. 2018. Estimation of a significance threshold for epigenome-wide association studies. Genetic Epidemiology, **42**: 20–33. doi:10.1002/gepi.22086.

Sharon, T. 2022. Determinants of habitat utilization patterns of European lobster (*Homarus gammarus*) in the inner Oslo fjord, Norway. Master's thesis. Norwegian University of Life Sciences, Ås, Norway.

Sheehy, M.R.J. 2001. Implications of protracted recruitment for perception of the spawner–recruit relationship. Canadian Journal of Fisheries and Aquatic Sciences, **58**: 641–644. doi:10.1139/f01-047.

Sheehy, M.R.J. & Bannister, R.C.A. 2002. Year-class detection reveals climatic modulation of settlement strength in the European lobster, *Homarus gammarus*. Canadian Journal of Fisheries and Aquatic Sciences, **59**: 1132–1143. doi:10.1139/f02-083.

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. doi:10.1139/f99-116.

Sheehy, M.R.J., Shelton, P.M.J., Wickins, J.F., Belchier, M. & Gaten, E. 1996. Ageing the European lobster *Homarus gammarus* by the lipofuscin in its eyestalk ganglia. Marine Ecology Progress Series, **143**: 99–111. doi:10.3354/meps143099.

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. doi:10.1007/s11160-019-09573-z.

Snir, S., Farrell, C. & Pellegrini, M. 2019. Human epigenetic ageing is logarithmic with time across the entire lifespan. Epigenetics, **14**: 912–926. doi:10.1080/15592294.2019.1623634.

Sørdalen, T.K., Halvorsen, K.T. & Olsen, E.M. 2022. Protection from fishing improves body growth of an exploited species. Proceedings of the Royal Society B, **289**: 20221718. doi:10.1098/rspb.2022.1718.

Sundelöf, A., Bartolino, V., Ulmestrand, M. & Cardinale, M. 2013. Multi-Annual Fluctuations in Reconstructed Historical Time-Series of a European Lobster (*Homarus gammarus*) Population Disappear at Increased Exploitation Levels. PLoS ONE, **8**: e58160. doi:10.1371/journal.pone.0058160.

Tangili, M., Slettenhaar, A.J., Sudyka, J., Dugdale, H.L., Pen, I., Palsbøll, P.J. & Verhulst, S. 2023. DNA methylation markers of age(ing) in non-model animals. Molecular Ecology, **32**: 4725–4741. doi:10.1111/mec.17065.

Thorbjørnsen, S., Moland, E., Huserbråten, M., Knutsen, J., Knutsen, H. & Olsen, E. 2018. Replicated marine protected areas (MPAs) support movement of larger, but not more, European lobsters to neighbouring fished areas. Marine Ecology Progress Series, **595**: 123–133. doi:10.3354/meps12546.

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. doi:10.1651/c-2448.

Vogt, G. 2019. A compilation of longevity data in decapod crustaceans. Nauplius, **27**: e2019011. doi:10.1590/2358-2936e2019011.

Vold, M. 2022. Short-term effects of a marine sanctuary on the local lobster population in the Oslo fjord, Norway. Master's thesis. Norwegian University of Life Sciences, Ås, Norway.

Wang, M. & Lemos, B. 2019. Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. Genome Research, **29**: 325–333. doi:10.1101/gr.241745.118.

Wang, Y., Pedersen, N.L. & Hägg, S. 2018. Implementing a method for studying longitudinal DNA methylation variability in association with age. Epigenetics, **13**: 866–874. doi:10.1080/15592294.2018.1521222.

Weber, D.N., Fields, A.T., Patterson, W.F., Barnett, B.K., Hollenbeck, C.M. & Portnoy, D.S. 2022. Novel epigenetic age estimation in wild-caught Gulf of Mexico reef fishes. Canadian Journal of Fisheries and Aquatic Sciences, **79**(1): 1–5. doi:10.1139/cjfas-2021-0240.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer, Cham. ISBN: 978-3-319-24277-4. doi:10.1007/978-3-319-24277-4.

Wickham, H. 2023. stringr: Simple, Consistent Wrappers for Common String Operations. R package. https://CRAN.R-project.org/package=stringr.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. 2019. Welcome to the Tidyverse. Journal of Open Source Software, **4**(43): 1686. doi:10.21105/joss.01686.

Wickham, H. & Bryan, J. 2023. readxl: Read Excel Files. R package. https://CRAN.R-project.org/package=readxl.

Wickham, H., Francois, R., Henry, L., Muller, K. & Vaughan, D. 2023. dplyr: A Grammar of Data Manipulation. R package. https://CRAN.R-project.org/package=dplyr.

Wickham, H., Vaughan, D. & Girlich, M. 2024. tidyr: Tidy Messy Data. R package. https://CRAN.R-project.org/package=tidyr.

Wilkinson, G.S., Adams, D.M., Haghani, A., Lu, A.T., Zoller, J., Breeze, C.E., Arnold, B.D., Ball, H.C., Carter, G.G., Cooper, L.N., Dechmann, D.K.N., Devanna, P., Fasel, N.J., Galazyuk, A.V., Günther, L., Hurme, E., Jones, G., Knörnschild, M., Lattenkamp, E.Z., Li, C.Z., Mayer, F., Reinhardt, J.A., Medellin, R.A., Nagy, M., Pope, B., Power, M.L., Ransome, R.D., Teeling, E.C., Vernes, S.C., Zamora-Mejías, D., Zhang, J., Faure, P.A., Greville, L.J., M., L.G.H., Flores-Martínez, J.J. & Horvath, S. 2021. DNA methylation predicts age and provides insight into exceptional longevity of bats. Nature Communications, **12**: 1615. doi:10.1038/s41467-021-21900-2.

Xie, C., Leung, Y.-K., Chen, A., Long, D.-X., Hoyo, C. & Ho, S.-M. 2018. Differential methylation values in differential methylation analysis. Bioinformatics, **35**: 1094–1097. doi:10.1093/bioinformatics/bty778.

Yan, L. 2023. ggvenn: Draw Venn Diagram by “ggplot2.” R package. https://CRAN.R-project.org/package=ggvenn.

Zou, H. & Hastie, T. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), **67**: 301–320. doi:10.1111/j.1467-9868.2005.00503.x.

Article A: Supplement

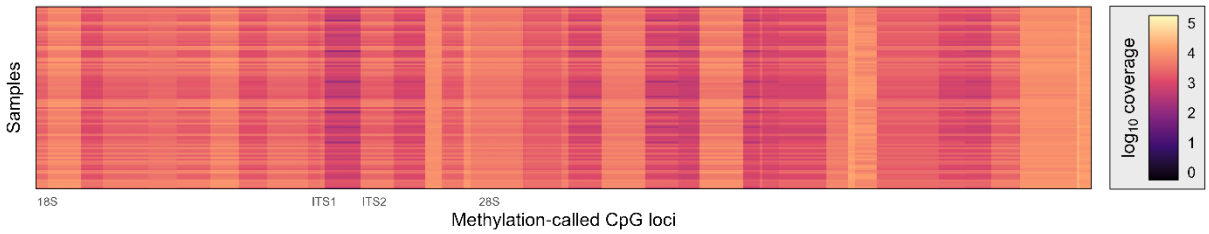

Figure S1 For all samples (two per lobster, i.e. 96 samples), coverage at 442 CpG loci is illustrated in a heatmap. The rDNA regions are indicated on the horizontal axis.

Simulating population assignment SNPs to inform monitoring of invasive introgression in European lobster (*Homarus gammarus*)

Erik Sandertun Røed [1, 2]

[1] Centre for Integrative Genetics, Faculty of Biosciences, NMBU

[2] Faculty of Environmental Sciences and Natural Resource Management, NMBU

Abstract

Distinguishing the European lobster (*Homarus gammarus*) from its transatlantic cousin American lobster (*H. americanus*) or their hybrid offspring is important for monitoring the invasive potential of *H. americanus* in the native range of *H. gammarus*. Morphological indicators are not sufficiently reliable for this purpose, but a recent study showed that a panel of 79 SNPs reliably separates the two species from each other and from their first-generation hybrid offspring. Because the SNP panel originally was developed for another purpose, it is unclear how effective it is at detecting invasive *H. americanus* introgression more broadly. Here, we describe simulations developed to explore the potential of the panel as a tool to monitor invasive introgression in populations of *H. gammarus*. The underlying code applies open-source software and is freely available to extend. We demonstrate that the simulations recapitulate empirical data and finally show, as an example application, that the SNP panel should also be able to detect introgression after at least one generation of backcrossing.

Introduction

Fishery biologists in Europe must distinguish the European lobster (*Homarus gammarus*) from the American lobster (*H. americanus*), a closely related but invasive species (van der Meeren et al. 2010). Larger and more aggressive, the alien *H. americanus* could threaten *H. gammarus* through direct competition and predation (Øresland et al. 2017), spread of disease (Davies & Wootton 2018), or maladaptive introgression via hybridisation (Todesco et al. 2016). Yet, the morphological features often used to identify *H. americanus* or potential hybrids in European waters (see e.g. Stebbing et al. 2012) are not always diagnostic (Jørstad et al. 2007, 2011). Instead, genetic markers can be used to reliably separate the two species.

Most notably, Ellis et al. (2020) demonstrated species assignment with virtually perfect fidelity using a multipurpose panel of 79 single nucleotide polymorphisms (SNPs) originally described by Jenkins et al. (2019a). Jenkins et al.'s (2019a) 79 SNPs separated 38 *H. americanus* from 1 278 *H. gammarus*; the two species were also separated from 30 first generation filial (F1) hybrid larvae sourced from an ovigerous *H. americanus* female captured in Sweden. As Ellis et al. (2020) noted, a report that F1 hybrids can be fertile (J. Kittaka, pers. comm. to van der Meeren et al. 2008, 2010) is cause for concern that backcrossing in the wild could lead to pervasive hybridisation. Sensitive genetic assignment of hybrids is consequently useful in a monitoring context given mounting evidence of hybridisation in the wild (Øresland et al. 2017, Barrett et al. 2020, Ellis et al. 2020). That said, it is unclear if Jenkins et al.'s (2019a) 79 SNPs distinguish hybrids as well across the intraspecific genetic cline of *H. gammarus* (see Jenkins et al. 2019b), or in backcrossed hybrids (Ellis et al. 2020). Therefore, estimating the false positive- or negative rates of invasive group assignment with Jenkins et al.'s (2019a) 79 SNPs, as well as optimal sampling strategies to detect introgression, would be useful.

In this brief report, we describe and make available a set of utility scripts written to simulate Jenkins et al.'s (2019a) 79-SNP panel, as used for species- and F1 hybrid assignment by Ellis et al. (2020). These scripts could be used to simulate empirical or hypothetical genotypes, generate *in silico* crosses and backcrosses of the SNP panel, and to estimate the sensitivity and specificity of hypothetical introgression monitoring schemes in order to optimise sample sizes. By virtue of the underlying software, these utilities could also be extended beyond the scope described here. To exemplify one use of the simulation pipeline we present, we apply it to the question of whether the Jenkins et al.'s (2019a) 79 SNP panel can detect hybridisation in backcrossed hybrids.

Methods

The core simulation was implemented in the population genetic software SLiM 4 (Haller & Messer 2023) with a simple script written in Eidos (Haller 2016). Fundamentally, the simulation is a Wright-Fisher model (Fisher 1923, Wright 1931) where an arbitrary number of discrete, diploid individuals are each characterized by two copies of an explicit nucleotide sequence. The between-nucleotide recombination rate is set to one-half; each nucleotide therefore segregates independently, and the modelled sequence behaves like a SNP panel in linkage equilibrium. For the purposes described here we assumed mutations were negligible, and a sequence of 79 nucleotides represented Jenkins et al.'s (2019a) 79 SNPs which were used to distinguish *H. americanus* from *H. gammarus* (Ellis et al. 2020).

The simulated SNP genotypes can be initialised from empirical data relative to an arbitrary reference sequence. Afterwards, simulated genotypes can be exported via variant calling against the same reference. Moreover, in the core simulation, populations of imported genotypes can be tagged with their true fraction of *H. americanus* ancestry. Parentage is then tracked throughout so that e.g. simulated F1 hybrid genotypes (*H. americanus* ancestry = 0.5) or F1 × *H. gammarus* backcross genotypes (*H. americanus* ancestry = 0.25) can be identified irrespective of the genotype itself. For ease of analysis in the R programming language, we implemented genotype import and export functions based on the packages vcfR v. 1.15.0 (Knaus & Grünwald 2017) and adegenet v. 2.1.10 (Jombart & Ahmed 2011) in R v. 4.3.3 (R Core Team 2024), built around adegenet's "genind" genotype objects. For the examples described below, we sourced real genotypes for Jenkins et al.'s (2019a) 79 SNP panel from a set of 1 591 *H. americanus*, *H. gammarus*, and F1 hybrid genotypes provided by Ellis et al. (2020) (including data from Jenkins et al. 2019b), but excluded individuals with missing data.

With a minimal model implementation, we controlled that the core simulation produces genotypes representative of imported data in the absence of simulated introgression. The minimal model implementation lacked population structure or introgression; each replicate simply allowed genotypes to neutrally shuffle for 100 generations. For simplicity we focused on empirical data from *H. gammarus* sampled at 38 Atlantic locations, *H. americanus* sampled at two locations, and one clutch of F1 hybrids, to the exclusion of *H. gammarus* sampled outside the Atlantic (as grouped by Ellis et al. 2020). Empirical data from each location were used to initialise two replicate simulations of 250 genotypes. From a final selection of 866 empirical genotypes (fig. S1), this produced 20 500 simulated genotypes for genetic clustering analyses.

Then, for the above-mentioned example application, we tested whether Jenkins et al.'s (2019a) 79 SNPs can distinguish *H. gammarus* and *H. americanus* also from backcrossed hybrids. A simulated crossing experiment was run with two populations of 1000 genotypes: one initialised from *H. gammarus* genotypes (16 empirical genotypes from Singlefjord, Norway), and the other from 20 *H. americanus* genotypes. After 100 generations of neutral burn-in, reciprocal migration began at a rate of 0.001 for five generations to permit hybridisation. In each of those five generations, hybrid genotypes with a 0.125, 0.25, 0.5, 0.75, or 0.875 fraction of known *H. americanus* parentage were recorded for genetic clustering. In the fifth and final generation, all remaining pure *H. gammarus* and *H. americanus* genotypes were also recorded.

For ease of comparison with the empirical results of Ellis et al. (2020), our genetic assignment analyses followed theirs by adapting scripts they kindly provided. In short, genetic separation between species was visualised with a cross-validated discriminant analysis of principal components (DAPC; Jombart et al. 2010) implemented in R using the adegenet package. We restricted the maximum number of DAPC axes to seven, i.e. the number of effective populations in the underlying empirical data (Ellis et al. 2020, 2023, 2024) minus one (Thia 2023). Genetic clustering for species and hybrid assignment was performed via maximum-likelihood estimation using the Snapclust algorithm (Beugin et al. 2018) with optional modelling of hybrids enabled, also implemented in R with adegenet. Modelling of hybrids was parameterised for the expected number of backcrosses in each set of genotypes.

From simulations to analyses, all scripts necessary to run the computations described here are available at https://github.com/ErikSRoeed/lobmsc to use or extend freely, alongside the empirical data necessary to replicate or build on the examples (reproduced with the permission of Ellis et al. 2020). Besides the above-mentioned dependencies, the scripts require the R packages slimmr (Røed 2024), poppr (Kamvar et al. 2014), dplyr (Wickham et al. 2023), tidyr (Wickham et al. 2024), purrr (Wickham & Henry 2023), stringr (Wickham 2023a), magrittr (Bache & Wickham 2022), and forcats (Wickham 2023b) for analysis and data wrangling, and ggplot2 (Wickham 2016), patchwork (Pedersen 2024), RColorBrewer (Harrower & Brewer 2003, Neuwirth 2022), and viridis (Garnier et al. 2024) for visualisation.

Results and discussion

The minimal simulations propagating 866 empirical genotypes of *H. americanus*, F1 hybrids, and *H. gammarus* from 38 Atlantic locations, confirmed that the core simulation behaves as

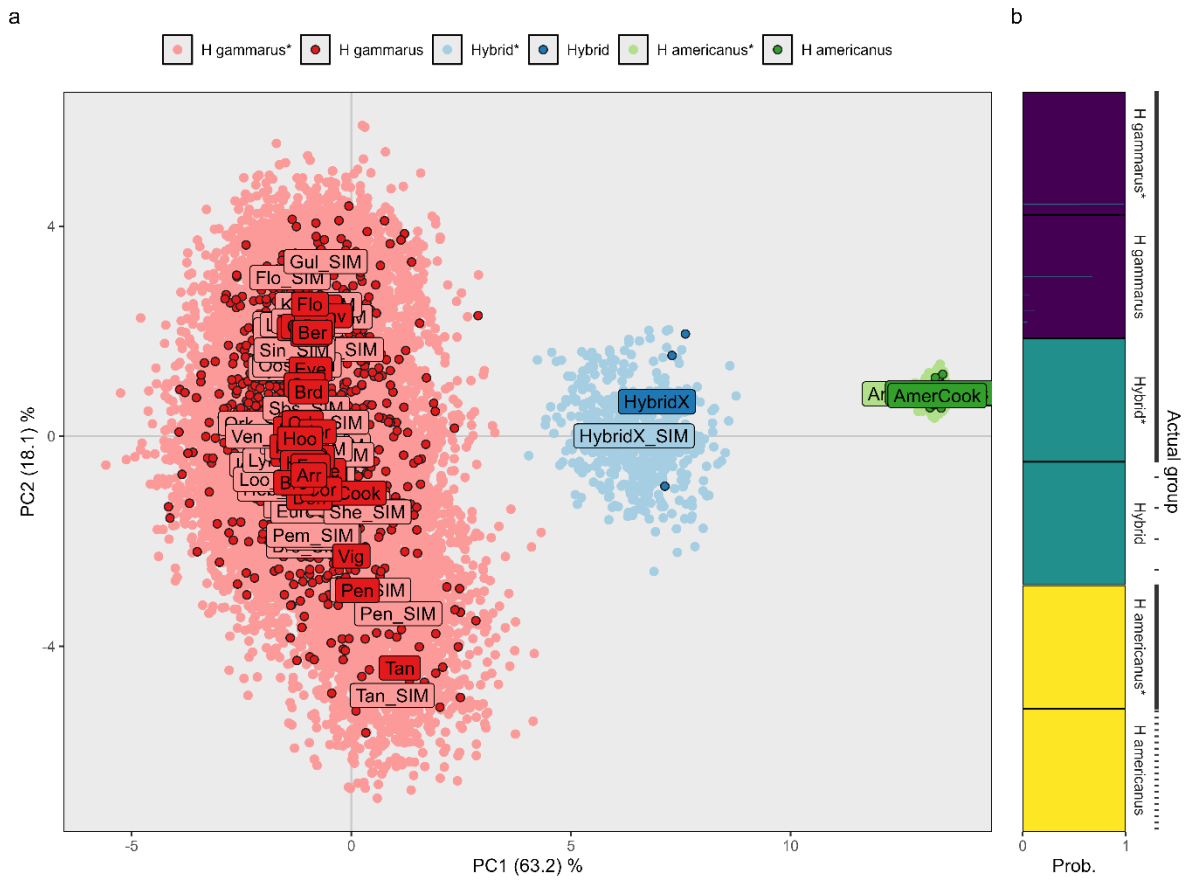

Figure 1 Genetic clustering of 20 500 simulated (labelled with *) and 866 empirical genotypes shown in **a)**, a plot of the two first principal components from DAPC, and **b)** a membership composition plot with stacked membership probabilities (prob.) from Snapclust. Labels in panel **a)** give the centroid of each set of genotypes. Empirical genotypes are named for the geographical origin of the source material as per Jenkins et al. (2019b), and centroids of data simulated from each set of empirical data are named similarly, with the suffix "_SIM". Hybrids were F1. Ticks above the group labels in **b)** indicate the number of individual genotypes in each actual group, which are collected and scale-free for clarity given the large number of simulated genotypes.

expected (fig. 1). In the DAPC, simulated genotypes (hereafter labelled with an asterisk, e.g. *H. gammarus**, or F1 hybrids*) grouped with their empirical source genotypes. The *H. gammarus**, *H. americanus** and F1 hybrid* genotypes clearly separated along the first principal component (PC) and recapitulated the empirical results of Ellis et al. (2020) closely (fig. 1a). Reflecting the substantial Atlantic cline of genetic variation in the empirical *H. gammarus* data (fig. S1; Jenkins et al. 2019b), *H. gammarus** genotypes similarly separated along the second PC. Snapclust unsurprisingly confirmed the impression from the DAPC with membership assignments of simulated genotypes coinciding with the empirical data (fig. 1b). These results verified that our simulations produced sensible approximations of real genotype data at Jenkins et al.'s (2019a) 79 SNPs.

Table 1 Number of simulated genotypes (labelled with * for consistency) from a simulated crossing experiment, grouped by their fraction of ancestry from *H. americanus** genotypes.

H. americanus* **ancestry**	**Equivalent cross**	**Count**
0.000	*H. gammarus** × *H. gammarus**	5922
0.125	*H. gammarus** × (*H. gammarus** × F1*)	1813
0.250	*H. gammarus** × F1*	1251
0.500	*H. gammarus** × *H. americanus** (i.e. F1 hybrids)	1675
0.750	*H. americanus** × F1*	1454
0.875	*H. americanus** × (*H. americanus** × F1*)	2020
1.000	*H. americanus** × *H. americanus**	5427
Total count		**19 562**

To investigate whether the same 79 SNPs might reliably detect also backcrossed hybrid genotypes, we also simulated 19 562 genotypes of *H. gammarus**, H. *americanus**, F1 hybrids* and four levels of backcrossing (table 1). These genotypes were all entirely simulated from empirical *H. gammarus* and *H. americanus* genotypes, so that simulated crossed genotypes were not initialised from empirical hybrid data (unlike in fig. 1). Still, the DAPC and Snapclust group membership of 1 675 F1 hybrids* from *H. gammarus** × *H. americanus** simulated crosses matched both their empirical and simulated-from-empirical counterparts shown in fig. 1 (fig. S2). The empirical F1 hybrid genotypes were sourced in Sweden (Ellis et al. 2020), i.e. from within the same Scandinavian effective population as the Norwegian *H. gammarus* genotypes used to initialise the crossing simulations (Jenkins et al. 2019b, Ellis et al. 2023). The shared clustering reaffirmed expectations as such, but for future studies it might be interesting to explore simulated crosses initialised from *H. gammarus* genotypes sourced elsewhere along the *H. gammarus* genetic cline.

For the Scandinavian population, genetic clustering of the simulated cross genotypes suggested that Jenkins et al.'s (2019a) 79 SNPs should still reliably detect *H. americanus* introgression after at least one generation of backcrossing (fig. 2a - 2c). When challenged with first-generation backcrossed hybrid genotypes (i.e. *H. gammarus** or *H. americanus** × F1*), Snapclust correctly assigned 93 % and 99 % of backcrosses with *H. gammarus** and *H. americanus**, respectively (fig. 2a). Simultaneously, over 96 % of the remaining genotypes were assigned their true level of *H. americanus** ancestry. Within each assigned group, the membership probability of true positive group members mostly exceeded 0.95 (i.e. the equivalent of $p < 0.05$), whereas the membership probability of misassignments were lower (fig. 2b and 2c).

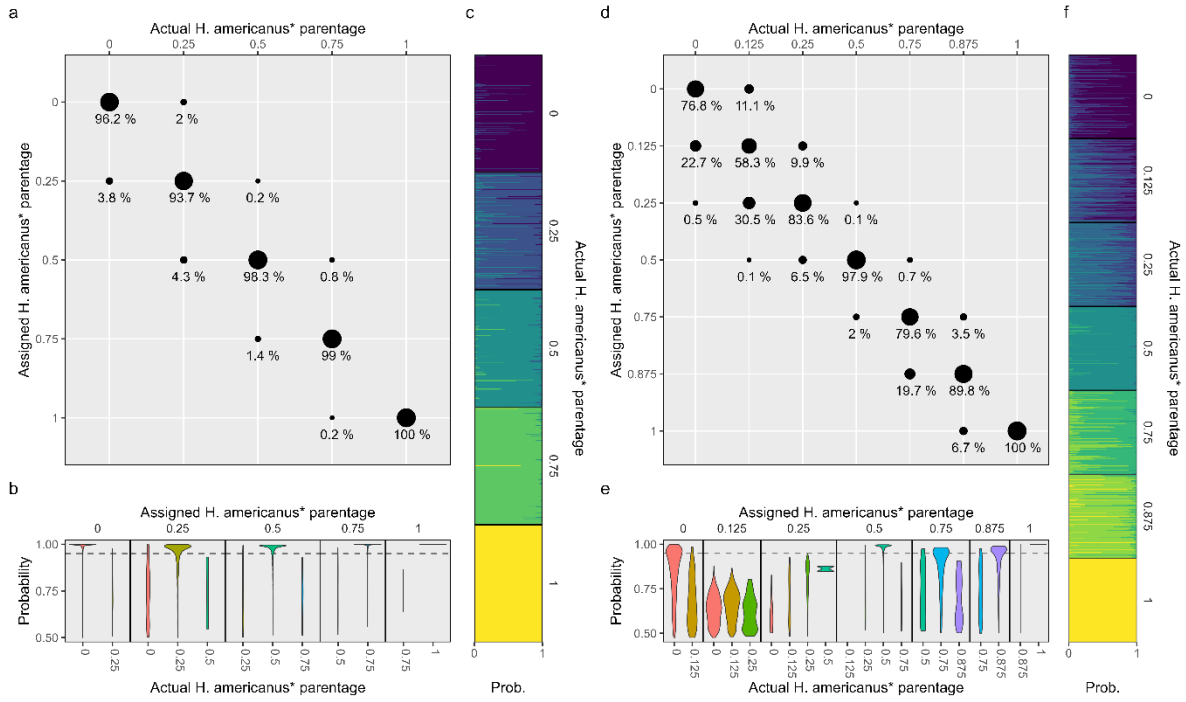

Figure 2 Snapclust assignments with modelling for one (panels **a, b, c**) and two (panels **d, e, f**) backcross generations. All genotypes are simulated (labelled *); pure *H. gammarus** have *H. americanus** parentage of 0, F1 hybrids* have 0.5, and so on. Panels **a, d)** shows scaled dots indicating the proportion of genotypes with majority assignment to the different assigned groups. Larger dots are lighter, and assignments on the diagonal match the actual group. The numbers under each dot give the percentage of actual group members with majority assignment to each assigned group and sum to 100 % vertically. Next, panels **b, e)** are violin plots of membership probabilities for each actual group within each assigned group. The horizontal dotted line indicates a membership probability of 0.95. Finally, panels **c, e)** are membership composition plots with groups scaled to equal vertical height.

When a second generation of backcrossing was included, group assignments were less accurate (fig. 2d – 2f). Broadly speaking, most genotypes were still assigned their true level of *H. americanus** ancestry, but especially hybrid genotypes doubly backcrossed with *H. gammarus** were susceptible to misassignment (correct assignment rate = 58.3 %, fig. 2d). In the applied context of monitoring for *H. americanus* introgression, it was particularly noteworthy that 11.1 % of true double backcross hybrids* were assigned to the *H. gammarus** group; *in vivo*, these would be false negatives. Conversely, the false positive rate among true *H. gammarus** genotypes was 22.7 %. In line with these false positive- and negative rates, correct membership probabilities skewed lower and wrong membership probabilities skewed higher, compared to the analysis including only one generation of backcrosses (fig. 2e and 2f c.f. fig. 2b and 2c). Still, across all groups with a true non-zero *H. americanus** contribution, the majority (at least 88.9 %, i.e. where *H. americanus** ancestry = 0.125) of genotypes were not

mistaken for pure *H. gammarus**. Future work could simulate sampling schemes in different scenarios of introgression to further explore the ramifications of backcrossing for the ability of Jenkins et al.'s (2019a) 79 SNPs to detect introgression in an applied context. For instance, assuming backcrosses could occur in the wild, how confident should we be in a negative result?

The greatest benefit of having implemented these simulations in SLiM 4 with an interface to R is precisely that future work easily can extend the core simulation. To model the spread of introgression in the wild, a non-Wright-Fisher model implementation (Haller & Messer 2019) could account for phenomena such as preferential mate choice in *H. gammarus* (van der Meeren et al. 2008) and the potential effect of differential fertility in hybrids (Talbot et al. 1983, 1984, c.f. J. Kittaka, pers. comm. to van der Meeren et al. 2008, 2010). Where more voluminous *H. americanus* releases or findings of juvenile *H. americanus* have caused concern of establishment (Barrett et al. 2020, Tinlin-Mackenzie et al. 2022), simulations could also model larval dispersal by implementing life stages and continuous space (Haller & Messer 2019, 2023). A different SNP panel could alternatively be simulated simply by changing the size of the simulated nucleotide sequence and importing different empirical data. Alternatively, the focal SNP panel here (Jenkins et al. 2019a) has also been shown to enable parentage assignment (Wit et al. 2021). The scripts we share could be used to explore that topic instead, or indeed examine other potential uses to maximise the benefit of a relatively cost-effective genetic tool and promote its use in the applied management of *H. gammarus*.

Concluding remarks

Here, we have shared a set of scripts for easily simulating Jenkins et al.'s (2019a) panel of 79 SNPs that can detect invasive introgression of *H. americanus* in populations of *H. gammarus* (Ellis et al. 2020). As the core simulation uses the powerful population genetic software SLiM 4 (Haller & Messer 2023), the scripts we share can easily be extended to relax the underlying model assumptions. Still, the simulations importantly depend on empirical genotypes and output analogous ones, so results produced with them reflect a real tool with which fisheries biologists can monitor populations in practice. We verified that the simulations conformed to the empirical data they were primed with, and that they can extend such empirical data to new questions that are relevant in a practical context. As an example, we demonstrated that the 79 SNPs reported by Jenkins et al. (2019a) can detect introgression reliably not only in F1 hybrids (Ellis et al. 2020), but also after at least one generation of backcrossing.

Acknowledgements

The author is grateful to Dr. Charlie Ellis and Prof. Jamie Stevens, who have kindly shared data, scripts and advice on analyses and research directions that were instrumental to the commencement of this project. Their much-appreciated involvement is still valuable to the ongoing development of the simulations described here.

References

Bache, S.M. & Wickham, H. 2022. magrittr: A Forward-Pipe Operator for R. R package. https://CRAN.R-project.org/package=magrittr.

Barrett, C.J., Cook, A., Stone, D., Evans, C., Murphy, D., Johnson, P., Thain, M., Wyn, G., Grey, M., Edwards, H., Quigley, D. & Stebbing, P.D. 2020. A review of American lobster (*Homarus americanus*) records around the British Isles: 2012 to 2018. Hydrobiologia, **847**: 3247–3255. doi:10.1007/s10750-020-04326-7.

Beugin, M., Gayet, T., Pontier, D., Devillard, S. & Jombart, T. 2018. A fast likelihood solution to the genetic clustering problem. Methods in Ecology and Evolution, **9**: 1006–1016. doi:10.1111/2041-210x.12968.

Davies, C.E. & Wootton, E.C. 2018. Current and emerging diseases of the European lobster (*Homarus gammarus*): a review. Bulletin of Marine Science, **94**: 959–978. doi:10.5343/bms.2017.1142.

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. doi:10.1038/s41598-020-64692-z.

Ellis, C.D., MacLeod, K.L., Jenkins, T.L., Rato, L.D., Jézéquel, Y., Pavičić, M., Díaz, D. & Stevens, J.R. 2023. Shared and distinct patterns of genetic structure in two sympatric large decapods. Journal of Biogeography, **50**: 1271–1284. doi:10.1111/jbi.14623.

Ellis, C.D., Paris, J.R., Jenkins, T.L., Stralen, M.R. van, Steins, N.A., Schotanus, J. & Stevens, J.R. 2024. Genetic divergence and adaptation of an isolated European lobster population in the Netherlands. ICES Journal of Marine Science: fsae059. doi:10.1093/icesjms/fsae059.

Fisher, R.A. 1923. On the dominance ratio. Proceedings of the Royal Society of Edinburgh, **42**: 321–341. doi:10.1017/s0370164600023993.

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedr, A., Sciain, Marc, Schere & Cédric. 2024. viridis(Lite) - Colorblind-Friendly Color Maps for R. R package. https://sjmgarnier.github.io/viridis/.

Haller, B. 2016. Eidos: A Simple Scripting Language. http://benhaller.com/slim/Eidos_Manual.pdf.

Haller, B.C. & Messer, P.W. 2019. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. Molecular Biology and Evolution, **36**: 632–637. doi:10.1093/molbev/msy228.

Haller, B.C. & Messer, P.W. 2023. SLiM 4: Multispecies Eco-Evolutionary Modeling. The American naturalist, **201**: E127–E139. doi:10.1086/723601.

Harrower, M. & Brewer, C.A. 2003. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. The Cartographic Journal, **40**: 27–37. doi:10.1179/000870403235002042.

Jenkins, T.L., Ellis, C.D. & Stevens, J.R. 2019a. SNP discovery in European lobster (*Homarus gammarus*) using RAD sequencing. Conservation Genetics Resources, **11**: 253–257. doi:10.1007/s12686-018-1001-8.

Jenkins, T.L., Ellis, C.D., Triantafyllidis, A. & Stevens, J.R. 2019b. 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. doi:10.1111/eva.12849.

Jombart, T. & Ahmed, I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, **27**: 3070–3071. doi:10.1093/bioinformatics/btr521.

Jombart, T., Devillard, S. & Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics, **11**: 94–94. doi:10.1186/1471-2156-11-94.

Jørstad, K.E., Agnalt, A.-L. & Farestveit, E. 2011. The Introduced American Lobster *Homarus americanus* in Scandinavian Waters. *In* In the Wrong Place - Alien Marine Crustaceans: Distribution, Biology and Impacts. *Edited by* B.S. Galil, P.F. Clark & J.T. Carlton. Springer, Dordrecht. pp. 625–638. doi:10.1007/978-94-007-0591-3_22.

Jørstad, K.E., Prodohl, P.A., Agnalt, A.-L., Hughes, M., Farestveit, E. & Ferguson, A.F. 2007. Comparison of genetic and morphological methods to detect the presence of American lobsters, *Homarus americanus* H. Milne Edwards, 1837 (Astacidea: Nephropidae) in Norwegian waters. Hydrobiologia, **590**: 103–114. doi:10.1007/s10750-007-0762-y.

Kamvar, Z.N., Tabima, J.F. & Grünwald, N.J. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, **2**: e281. doi:10.7717/peerj.281.

Knaus, B.J. & Grünwald, N.J. 2017. vcfr: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources, **17**: 44–53. doi:10.1111/1755-0998.12549.

Neuwirth, E. 2022. RColorBrewer: ColorBrewer Palettes. R package. https://CRAN.R-project.org/package=RColorBrewer.

Øresland, V., Ulmestrand, M., Agnalt, A.-L. & Oxby, G. 2017. Recorded captures of American lobster (Homarus americanus) in Swedish waters and an observation of predation on the European lobster (Homarus gammarus). Canadian Journal of Fisheries and Aquatic Sciences, **74**: 1503–1506. doi:10.1139/cjfas-2016-0532.

Pedersen, T.L. 2024. patchwork: The Composer of Plots. R package. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2024. R: A Language and Environment for Statistical Computing. https://www.R-project.org/.

Røed, E.S. 2024. slimmr: SLiM Models in R. R package. https://github.com/ErikSRoeed/slimmr

Stebbing, P., Johnson, P., Delahunty, A., Clark, P., McCollin, T., Hale, C. & Clark, S. 2012. Reports of American lobsters, *Homarus americanus* (H. Milne Edwards, 1837), in British waters. BioInvasions Records, **1**: 17–23. doi:10.3391/bir.2012.1.1.04.

Talbot, P., Hedgecock, D., Borgeson, W., Wilson, P. & Thaler, C. 1983. Examination of spermatophore production by laboratory-maintained lobsters (*Homarus*). Journal of the World Mariculture Society, **14**: 269–278. doi:10.1111/j.1749-7345.1983.tb00083.x.

Talbot, P., Thaler, C. & Wilson, P. 1984. Spawning, egg attachment and egg retention in captive lobsters (*Homarus americanus*). Aquaculture, **37**: 239–249. doi:10.1016/0044-8486(84)90157-1.

Thia, J.A. 2023. Guidelines for standardizing the application of discriminant analysis of principal components to genotype data. Molecular Ecology Resources, **23**: 523–538. doi:10.1111/1755-0998.13706.

Tinlin-Mackenzie, A., Ellis, C., Lodola, A., Martin-Ruiz, C., Stevens, J. & Fitzsimmons, C. 2022. New kid on the block: first record of juvenile American lobster, *Homarus americanus* H. Milne Edwards, 1837, in European waters. BioInvasions Records, **11**: 901–1000. doi:10.3391/bir.2022.11.4.17.

Todesco, M., Pascual, M.A., Owens, G.L., Ostevik, K.L., Moyers, B.T., Hübner, S., Heredia, S.M., Hahn, M.A., Caseys, C., Bock, D.G. & Rieseberg, L.H. 2016. Hybridization and extinction. Evolutionary Applications, **9**: 892–908. doi:10.1111/eva.12367.

van der Meeren, G., Chandrapavan, A. & Breithaupt, T. 2008. Sexual and aggressive interactions in a mixed species group of lobsters *Homarus gammarus* and *H. americanus*. Aquatic Biology, **2**: 191–200. doi:10.3354/ab00050.

van der Meeren, G.I., Støttrup, J., Ulmestrand, M., Øresland, V., Knutsen, J.A. & Agnalt, A.-L. 2010. NOBANIS Invasive Alien Species Fact Sheet - *Homarus americanus*. Online Database of the European Network on Invasive Alien Species – NOBANIS. https://www.nobanis.org/globalassets/speciesinfo/h/homarus-americanus/homarus_americanus.pdf.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer, Cham. ISBN: 978-3-319-24277-4. doi:10.1007/978-3-319-24277-4.

Wickham, H. 2023a. stringr: Simple, Consistent Wrappers for Common String Operations. R package. https://CRAN.R-project.org/package=stringr.

Wickham, H. 2023b. forcats: Tools for Working with Categorical Variables (Factors). R package. https://CRAN.R-project.org/package=forcats.

Wickham, H., Francois, R., Henry, L., Muller, K. & Vaughan, D. 2023. dplyr: A Grammar of Data Manipulation. R package. https://CRAN.R-project.org/package=dplyr.

Wickham, H. & Henry, L. 2023. purrr: Functional Programming Tools. R package. https://CRAN.R-project.org/package=purrr.

Wickham, H., Vaughan, D. & Girlich, M. 2024. tidyr: Tidy Messy Data. R package. https://CRAN.R-project.org/package=tidyr.

Wit, P.D., Svanberg, L., Casties, I., Eriksson, S.P., Sundell, K. & André, C. 2021. Single nucleotide polymorphisms are suitable for assessing the success of restocking efforts of the European lobster (*Homarus gammarus*, L.). Conservation Genetics Resources, **14**: 47–52. doi:10.1007/s12686-021-01234-3.

Wright, S. 1931. Evolution in Mendelian populations. Genetics, **16**: 97–159. doi:10.1093/genetics/16.2.97.

Article B: Supplement

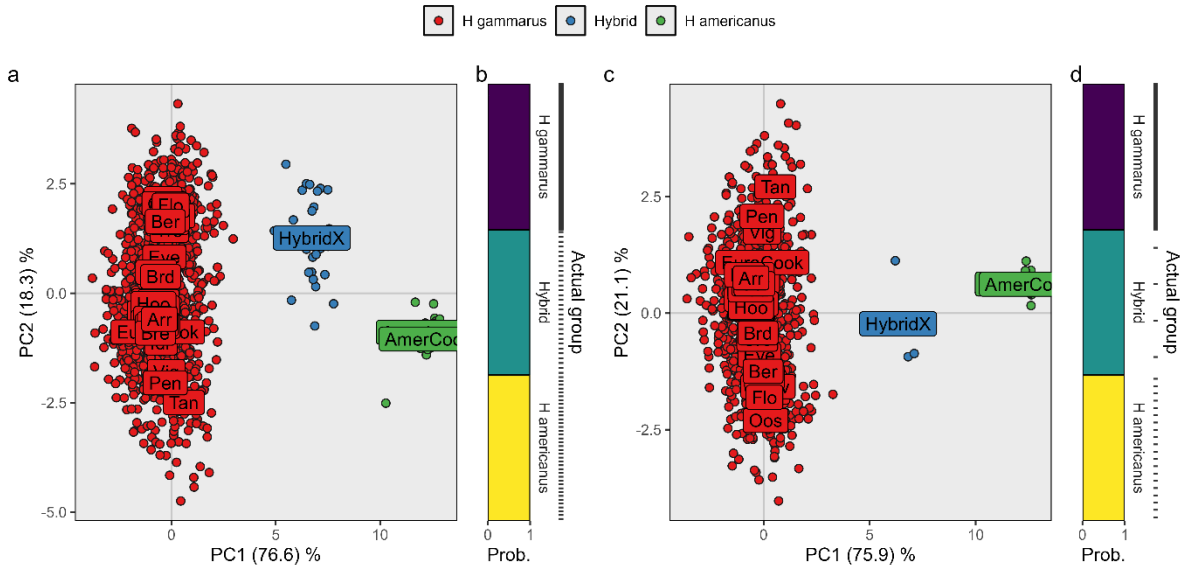

Figure S1 The empirical data used in this study, shown in DAPC (panels **a, c**) and Snapclust membership composition plots (panels **b, d**) with (**a**, **b**) and without (**c**, **d**) genotypes missing data at one or more loci. Data is sourced from Ellis et al. (2020) with labels as per Jenkins et al. (2019b). Ticks above the Snapclust compoplots show the number of group members. Note that PC2 is flipped around the horizontal axis in panel **c)** relative to **a)**.

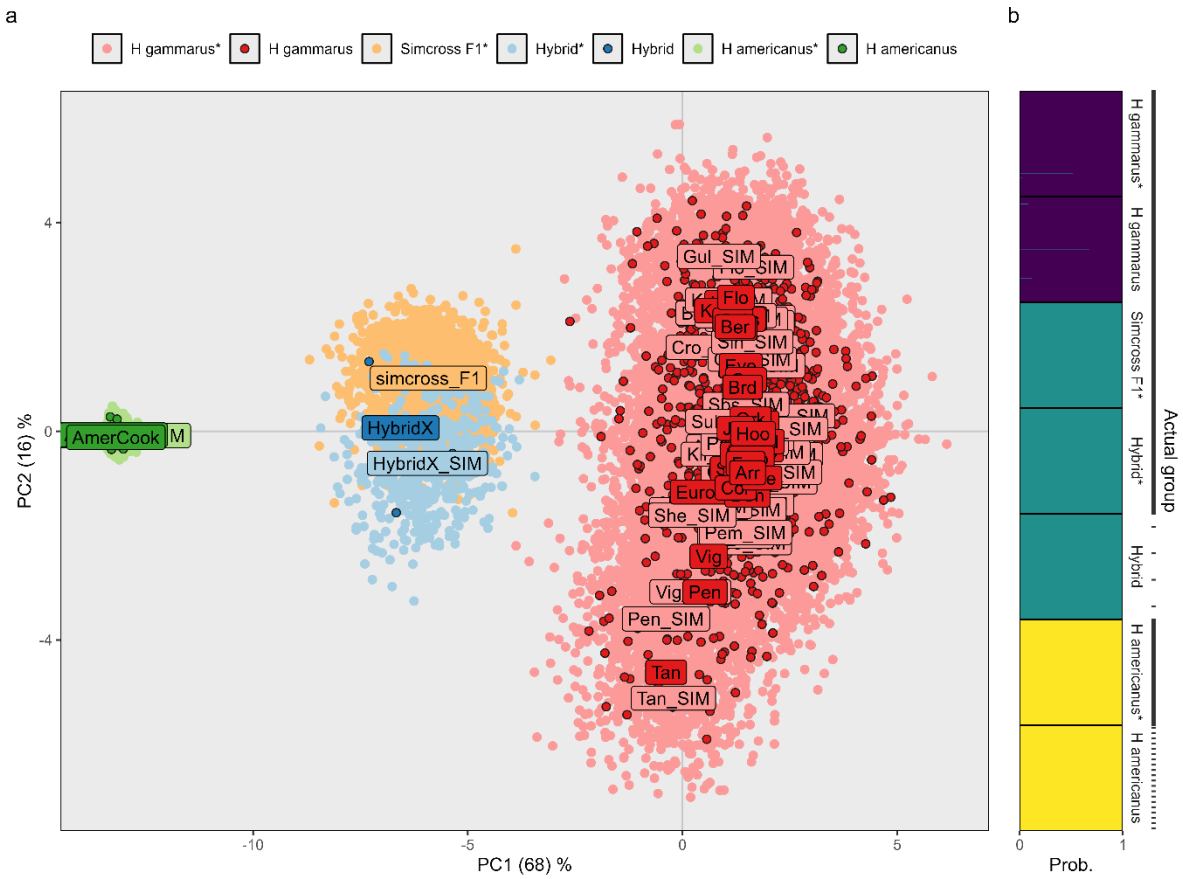

Figure S2 F1 hybrid genotypes generated from within-simulation crosses of *H. americanus** and *H. gammarus** genotypes ("Simcross F1*") are compared to F1 hybrid* genotypes simulated from empirical F1 hybrid data, and empirical and simulated data from *H. americanus* and *H. gammarus*. Simulated data is labelled with an asterisk in captions, and/or labelled with the suffix "_SIM". Panel **a)** is a DAPC plot and panel **b)** a grouped Snapclust membership composition plot scaled freely on the vertical axis to show all groups irrespective of the number of members (shown as ticks above the actual group labels). Note that F1 hybrids assign similarly irrespective of whether they are empirical, simulated directly from empirical data, or are crosses of simulated *H. americanus* and *H. gammarus* genotypes.

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway