PROF. FRANCESC PIFERRER (Orcid ID : 0000-0003-0903-4736)

Article type : Resource Article

A clockwork fish. Age-prediction using DNA methylation-based biomarkers in the European seabass

Running title: Piscine epigenetic clock

Dafni Anastasiadi[1,2] and Francesc Piferrer*[1]

[1]Institut de Ciències del Mar (ICM), Spanish National Research Council (CSIC), Passeig Marítim, 37–49, 08003 Barcelona, Spain.
[2]Current address: The New Zealand Institute for Plant & Food Research, Nelson, New Zealand

*Author for correspondence. Tel. +34-932309567; email: piferrer@icm.csic.es

ORCID numbers
DA: 0000-0002-4871-4649
FP: 0000-0003-0903-4736

Keywords: epigenetic clock, DNA methylation, biomarker, fish, age, growth, stock assessment, conservation

Abstract

Age-related changes in DNA methylation do occur. Taking advantage of this, mammalian and avian epigenetic clocks have been constructed to predict age. In fish, studies on age-related DNA methylation changes are scarce and no epigenetic clocks have been constructed. However, in fisheries and population dynamics studies there is a need for accurate estimation of age, something that is often impossible for some economically important species with the currently available methods. Here, we used the European sea bass, a marine fish where age can be known with accuracy, to construct a piscine epigenetic clock, the first one in a cold-blooded vertebrate. We used targeted bisulfite sequencing to amplify 48 CpGs from four genes in muscle samples and applied penalized regressions to predict age. We, thus, developed an age predictor in fish that is highly accurate (0.824) and precise (2.149 years). In juvenile fish, accelerated growth due to elevated temperatures had no effect in age prediction, indicating that the clock is able to predict the chronological age independently of environmentally-driven perturbations. An epigenetic clock developed using muscle samples accurately predicted age in samples of testis but not ovaries, possibly reflecting the reproductive biology of fish. In conclusion, we report the development of the first piscine epigenetic clock, paving the way for similar studies in other species. Piscine epigenetic clocks should be of great utility for fisheries management and conservation purposes, where age determination is of crucial importance.

Introduction

Age is one of the most important factors in ecology because it affects both individual and population traits. Key characteristics influenced by age in animals include, among others, reproductive maturity, behavior and reproductive success; and in populations, demographic structure and species interactions (Roach & Carey, 2014).

In fisheries management, accurate estimates of age are essential to infer life-history traits and for proper stock assessment (Johnson, McKenna Jr., Dropkin, & Andrews, 2005; Pardo, Cooper, & Dulvy, 2013). Age estimation in fishes has traditionally relied on the analysis of growth marks in hard structures such as scales and otoliths (Campana, 2001). These methods require highly specialized and well-trained personnel, are time-consuming, have low accuracy in some species and are necessarily lethal in the case of otoliths. Further, there are some species of worldwide economic importance such as hake (*Merlucius merlucius*) for which age estimation based on these traditional methods is very difficult or impossible. Consequently, age cannot be evaluated in large-scale stock assessments (EEA, 2017). Thus, alternative methods of age prediction in fish are very much needed.

Aging is associated with changes in DNA methylation, one of the main epigenetic mechanisms. In fish, the first link between aging and DNA methylation was found in humpback salmon (*Oncorhynchus gorbuscha*), where a decrease in the overall number of methylated cytosines with age was observed (Berdyshev, Korotaev, Boiarskikh, & Vaniushin, 1967). Since then, a similar decrease in DNA methylation has been shown in different vertebrates including the rat, mouse and cow (Romanov & Vanyushin, 1981). In general, there is an age-dependent change in DNA methylation that may be summarized in global genomic hypomethylation. This is thought to result from the loss of enzymatic activity of the maintenance DNA methyltransferase 1 (DNMT1) (Heyn et al., 2012; Mugatroyd, Wu, Bockmühl, & Spengler, 2010), accompanied by hypermethylation of specific CpG sites that are not associated with changes in gene expression (Jung & Pfeifer, 2015). However, aside from this “epigenetic drift”, there are DNA methylation changes that are regulated by other DNMTs such as DNMT3b. In this case, DNA methylation changes are of a “clock-like” type, i.e., follow a certain pace, affect specific loci (increase or decrease in methylation) and may entail functional consequences (Paoli-Iseppi et al., 2019).

Variation in DNA methylation is influenced by both intrinsic and extrinsic factors plus stochastic events (Jung & Pfeifer, 2015). Further, DNA methylation at specific loci or genomic regions is also dependent on the tissue examined. Notwithstanding, after a careful examination of age-related changes in DNA methylation in a variety of human tissues and cell lines, a tissue-independent epigenetic clock was first developed in humans (Horvath, 2013). This epigenetic clock included a total of 353 CpG sites that were sufficient to predict the age of an individual much better than with any other methods available at the time such as mitochondrial mutation accumulation (Horvath, 2013; Jarman et al., 2015). Since the basic features of DNA methylation are conserved at least in close-related species or even taxa, the epigenetic clock based on the same informative CpGs identified in humans was found to also work reasonably well in chimpanzees (*Pan troglodytes*) (Horvath, 2013). Since then, species-specific epigenetic clocks have been constructed in chimpanzees (Ito, Udono, Hirata, & Inoue-Murayama, 2018), mice (Han et al., 2018), humpback whales (*Megaptera novaeangliae*) (Polanowski, Robbins, Chandler, & Jarman, 2014), Bechstein's bats (*Myotis bechsteinii*) (Wright et al., 2018), dogs and wolves (Janowitz Koch et al., 2016; Thompson, vonHoldt, Horvath, & Pellegrini, 2017) and in a long-lived seabird (*Ardenna tenuirostris*) (Paoli-Iseppi et al., 2019). Epigenetic clocks are characterized by very high accuracy when compared to more traditional methods of age estimation. However, the patterns of ageing are extensively diverse across the tree of life and the ones of senescence and growth are contrasting (Jones et al., 2014). Thus, it is of great interest to uncover the relationships of DNA methylation with ageing in other vertebrates as well.

In fish, apart from the first study in humpback salmon cited above, more recently a genome-wide study using zebrafish as a model showed an age-dependent hypomethylation of CpGs sites identified by methylation-sensitive enzyme digestion (Shimoda et al., 2014). In addition, a study using Chinook salmon (*Oncorhynchus tshawytscha*) and a candidate gene approach showed that there were changes in DNA methylation that were gene-, tissue- and age-dependent (Venney, Johansson, & Heath, 2016). More recently, in juvenile steelhead trout (*Oncorhynchus mykiss*), hypomethylation was shown to prevail at a later age (Gavery et al., 2019). These results indicate that there is potential for the development of an epigenetic clock in fish, in the same way it has been developed in the species of mammals and birds discussed above. However, birds and mammals are warm-blood animals, which have a lower ^{5m}C content compared to cold-blooded

vertebrates such as fish (Varriale, 2014; Varriale & Bernardi, 2006). ^{5m}C content is inversely related to body temperature. Thus, fish living in polar regions have the highest ^{5m}C content while those living in tropical areas the lowest (Varriale, 2014; Varriale & Bernardi, 2006). This trend accounts for the fact that fish have a higher level of ^{5m}C content when compared to mammals (Varriale, 2014). Furthermore, in contrast to mammals, fish exhibit indeterminate growth and age-increased fecundity, something that has been linked to the delayed senescence and extreme longevity evolved in fish (Reznick, Ghalambor, & Nunney, 2002). On the other hand, in birds and mammals senescence is thought to have evolved as a by-product of determinate growth (Reznick et al., 2002). These important constitutive differences in overall ^{5m}C content, growth and reproduction patterns between warm- and cold-blood animals point out the uncertainty of whether DNA methylation changes of the "clock-like" type may exist in poikilothermic animals.

Here, we used muscle samples of European sea bass (*Dicentrarchus labrax*) of known age to evaluate the construction of a piscine epigenetic clock. The European sea bass is a Perciform fish very important for recreational fisheries and aquaculture (Felip & Piferrer, 2018). We used this marine fish and individuals reared in captivity in order to know the exact age. Building an epigenetic clock in fish implied facing the uncertainties described above (different ^{5m}C content, poikilothermy and indeterminate growth). Therefore, we avoided adding further uncertainties by not using a species from capture (i.e., commercial) fisheries, where age would have to be estimated by otoliths. We amplified the regulatory regions of four genes of the European sea bass using Multiplex Bisulfite Sequencing (MBS) (Anastasiadi, Vandeputte, Sánchez-Baizán, Allal, & Piferrer, 2018). Furthermore, since temperature affects the growth rate in poikilotherms, we tested the effect of high temperature on the performance of the epigenetic clock and evaluated the influence of the number of informative CpGs on the accuracy of the clock. Lastly, we tested whether an epigenetic clock developed in muscle could be applied to the reproductive tissues, testis and ovary.

Materials and Methods

Animals and rearing conditions

Tissue samples of European sea bass were obtained from fish of different ages, with known date of birth, that were reared at the aquarium facilities of either the Institute of Marine Sciences (ICM-CSIC, Barcelona, Spain) or the Institute of Aquaculture (IATS-CSIC, Castellón, Spain). Fish were reared following standard procedures (Díaz, Ribas, & Piferrer, 2013; Morretti, 1999). For a detailed summary of the samples used in this study see **Table S1**. The aquarium facilities (Spanish National Research Council; CSIC,) are authorized for experiments with animals by the Ministry of Agriculture and Fisheries certificate number 08039–46–A according to Spanish legislation (R.D. 223 of March 1988). The treatment of animals was in accordance to the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes (ETS Nu 123, 01/01/91) and the protocol was licensed by the Ethics Committee of the CSIC under the project AGL2016–78710–R.

Fish were sacrificed with an overdose of 2-phenoxyethanol (2PE). Muscle was dissected from male fish of different ages. In addition, for the 1.28, 3.07 and 10.5 age classes, testis was also dissected from the same individuals, while ovaries were dissected from female sibling fish of the same exact age (**Table S1**). In all cases dissection of tissue samples was followed by immediate immersion into liquid nitrogen. In general, juvenile fish were reared at stable natural temperature of 17ºC. After this period, temperature followed natural fluctuations. In order to test the possible effect of accelerated growth due to temperature on DNA methylation, siblings of the fish reared at 17°C were reared at high temperature (21ºC) from 7 to 68 days post fertilization (dpf), which coincides with the thermosensitive period for sea bass sex differentiation (Navarro-Martín, Blázquez, Viñas, Joly, & Piferrer, 2009). For each age class, tissue and temperature combination, four individual fish were considered as biological replicates. In total, 88 different fish were used for this study.

DNA extraction and bisulfite conversion

Genomic DNA extraction was performed by the standard phenol-chloroform-isoamyl alcohol (PCI; 25:24:1) protocol. One microgram of proteinase K (Sigma-Aldrich) and 0.5 µg of ribonuclease A (PureLink RNase A; Life Technologies) were added to the digestion buffer to ensure the absence of proteins and RNA, respectively. The quantity and purity of DNA were measured by the ND-100 spectrophotometer (NanoDrop Technologies). Two micrograms of DNA for each sample were bisulfite converted using the EZ DNA Methylation-Direct™ Kit (Zymo

Research; D5020) according to the instructions of the manufacturer, except from the desulphonation time, which was prolonged to 30 min.

Selection of genes

A list of the genes targeted in this study with their names, abbreviations and tissue of function are listed in **Table S2**. Four genes, *amh-r2*, *fsh-r*, *nr3c1* and *sox9*, were selected for the development of an epigenetic clock in muscle based on the absence of sex-related differences in DNA methylation and major gene expression changes (Anastasiadi et al., 2018; Navarro-Martín, Galay-Burgos, Piferrer, & Sweeney, 2012). For three of them (*amh-r2*, *nr3c1* and *fsh-r*) we had previously constructed MBS libraries (Anastasiadi et al., 2018) and thus, primers were validated and available. The fourth gene was *sox19*, a fish-specific gene known for its involvement in ovarian differentiation (Navarro-Martín et al., 2012), with a baseline expression in the muscle and thus expected to be unrelated with muscle functions.

Eighteen more genes were targeted (**Table S2**) in order to test the inclusion of more CpGs and the influence of the tissue in the construction of the clock. Previous knowledge about the gene functions related to muscle, testis, ovary or gonads was used for selecting the target genes. Genes were ranked according to the number of CpGs per 100 bp aiming for more than 5 CpG/100 bp in the region around the transcription start site (TSS) and a literature survey of relative functional importance The locations of the target regions were identified in the sea bass genome (dicLab1 v1.0c, July 2012) (Tine et al., 2014).

Multiplex bisulfite sequencing libraries

Multiplex bisulfite sequencing (MBS) libraries were prepared as in Anastasiadi et al (2018). Details of the primers used for amplifying the target regions and the corresponding annealing temperature can be found in **Table S3.** Size-selection and normalization were performed following a customized version of the procedure described in (Hosomichi, Mitsunaga, Nagasaki, & Inoue, 2014) using Sera-mag SpeedBeads (Fisher 099981123) (Rohland & Reich, 2012), exactly as described in Anastasiadi et al (2018). After normalization, equal volumes of each amplicon were pooled for each sample. The MBS libraries were prepared in two rounds: 1) Muscle as template for the amplification of 4 genes, resulting in 64 pools of amplicons, 2) Muscle, testis and ovaries as templates for the amplification of 22 genes, resulting in 36 pools of amplicons.

A dual-index strategy was employed to include sample-specific indices, using the i7 indices from Nextera XT index Kit SetA and i5 indices from Nextera XT index Kit SetD (Illumina; FC-131–2001 and FC-131–2004). Following the index PCR, DNA was quantified by the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific) and pooled in equal quantities and further cleaned-up. The final libraries were quantified by real-time qPCR using the Kapa system and sequenced on a MiSeq (Illumina) in 300 bp paired-end mode. The basic bioinformatic analysis were performed exactly as in Anastasiadi et al (2018). Only reads with >5 x coverage were retained.

Statistical analyses: data pre-processing

All statistical analysis were performed using R (v. 3.6.1) (R Core Team, 2015) and Rstudio (v. 1.2.1335) (RStudio Team, 2015). We excluded from the analyses samples where more than 30% of the observations were missing and CpGs where more than 25% of the observations were missing. The rest of missing values were imputed using multivariate imputation by chained equations as implemented in the *mice* package (v. 3.6.0) using the default parameters. For the construction of the epigenetic clock only muscle samples from fish reared at a low temperature were used. Thus, 48 CpGs distributed in 4 genes from 50 valid samples were used for downstream analysis.

Statistical analyses: penalized regressions

We used regression analysis to predict age based on the methylation values of the 48 CpGs. We chose to use penalized regressions, as best fitted in our case, where methylation of CpGs have strong multicollinearity and are not independent. Therefore, we evaluated ridge, lasso and elastic net models using the glmnet package (v. 2.0-18) through caret (v. 6.0-84) in a Leave-One-Out-Cross-Validation (LOOCV) context. Pearson's correlations were used to assess the relationships of predicted and chronological age.

Statistical analyses: penalized regressions in young fish and reared at high temperature

The same model was then used to predict age only in the fish of 200 to 600 dpf reared at low temperature and fish from the same lot aged 176 to 500 dpf reared at high temperature. The effect of temperature on age prediction was evaluated using Analysis of Covariance to compare two linear regression models: a) $\text{Age}_{\text{predicted}} = \text{Age}_{\text{chronological}}$ + Temperature + interaction of Age

$_{chronological}$ and Temperature, and b) Age $_{predicted}$ = Age $_{chronological}$. A Welch two sample *t*-test was used to assess whether there were differences in size (length) in fish of 500 days reared at low and high temperature. The correlation of the methylation of the CpGs was evaluated between these two groups of low and high temperature by Pearson's correlation.

Statistical analyses: penalized regressions of more CpGs across tissues

We then wanted to test the inclusion of more CpGs on the clock's construction and the effectiveness of age prediction across tissues. Three age classes were considered, representative of the full spectrum available: 1.28, 3.07 and 10.5 years. We used 299 CpGs on 22 genes (**Table S2**) which resulted in 295 CpGs after filtering. First, we constructed a second version of the epigenetic clock for muscle only since it was our reference tissue of target. We used the same procedure as above and we applied ridge, lasso and elastic net penalized regressions using the caret workflow and chose the best tuning parameters. We evaluated the model performances using LOOCV and selected the model that minimized the prediction error (RMSE). The final model was used to predict the age in the samples of testes and ovaries. The effect of size (length of fish) on age prediction was evaluated using Analysis of Covariance to compare two linear regression models: a) Age $_{predicted}$ = Age $_{chronological}$ + Length + interaction of Age $_{chronological}$ and Length and b) Age $_{predicted}$ = Age $_{chronological}$. The coefficients of each factor for the first full model are reported separately for each tissue in **Table 2**.

Results

Development of the sea bass epigenetic clock

We developed an epigenetic clock to estimate age in sea bass using the DNA methylation levels of 48 CpGs on 4 genes (*amh-r2*, *nr3c1 fsh-r* and *sox9*) in samples of muscle extracted from fish between 0.55 and 10.5 years old. After evaluating three types of penalized regressions, we selected the best model with the lowest Root Mean Squared Error (RMSE, **Fig. S1**) which was an elastic net model (α=0.1, λ=1.22008). Our selected model minimized the prediction error (mean absolute error; MAE) to 2.149 years (**Fig. 1A**). The predicted age showed a high and significant positive correlation with the chronological age with r=0.824 and p=1.97^{e-13} (**Fig. 1A**). The ages predicted with the epigenetic clock were binned into three categories after prediction and were significantly different between young (0 to 3 years), middle (3 to 10 years) and old (more than 10 years) fish

(one-way ANOVA; F-statistic=54.37, *p*-value=5.932^{e-13}; **Fig. 1B**). Length increased with age as expected as regressed to a von Bertalanffy curve (**Fig. 1C**) but had no effect on the epigenetic age prediction when considered as a covariate (F=0.5558, *p*=0.4597). The elastic net shrank the coefficients of the initial 48 CpGs and in the final model 28 of them were non-zero (**Table 1**). Examples of CpGs, the methylation of which was significantly positively or negatively correlated with age, are shown in **Fig. 2.** Despite the fact that not all the correlations of DNA methylation with age were significant (**Table 1**), the methylation of all CpGs was informative for age prediction.

Effect of temperature on the epigenetic age of young fish

Fish, in contrast to mammals or birds, are cold-blooded and have indeterminate growth, which is easily affected by environmental factors, the most important being temperature. In fish elevated higher-than-normal temperatures accelerate growth (Ayala et al., 2000; Navarro-Martín et al., 2009). Therefore, we tested whether accelerated growth due to high temperature exposure influenced the performance of the epigenetic clock. We used young fish (<600 dpf), where the effects of high temperature on growth are more pronounced and reared them at low (17°C) or high (21°C) temperature during the thermosensitive period (7–68 dpf). As expected, fish reared at high temperature were larger than fish reared at low temperature at the same chronological age (**Fig. 3A**). We used the epigenetic clock with the previously tuned parameters to predict age separately in the low and the high temperature group. The correlations of predicted and chronological age were similar in the low (r=0.998, $p<2.2^{e-16}$) and in the high (r=0.996, $p=<2.2^{e-16}$) temperature groups (**Fig. 3B**). The precision was 140.3 and 139.477 days in the low and high temperature groups, respectively. The ANCOVA showed that the effect of temperature was not significant on the prediction of age (F=1.823, *p*=0.174). The linear regression that best fit the data was: $Age_{predicted} = 24.15 + 0.931 * Age_{chronological}$. The methylation levels of CpGs from fish with the same chronological age (500 dpf), but with differences in size due to temperature (Welch two sample *t*-test, t=-3.56, *p*-value=0.012; arrows in **Fig. 3A**), were significantly and positively correlated (Pearson's correlation, ρ=0.999, p-value<2.2^{e-16}; **Fig. 3C**), further supporting that differences in size as a result of temperature perturbations did not hinder finding no differences in predicted age.

Higher number of CpGs and reproductive tissues

We then explored the influence of other target genes and tissue of origin in the performance of the epigenetic clock. Thus, we processed by MBS the regulatory regions of 18 additional genes (**Table S2**), so in total we considered 22 genes. Fish in which more than one tissue was available had an age of 1.28, 3.07 and 10.5 years. This resulted in a total of 299 CpGs to select from in the muscle of these fish. An elastic net penalized regression model with parameters α=0.1 and λ=0.351 (amount of shrinkage) was selected to predict age because it minimized the prediction error to 3.301 years (RMSE, **Fig. S2**). The correlation of predicted with chronological age was high (r=1) and significant ($p<2.2^{e-16}$, **Fig. 4A**). The same final model was used to predict age by regressing the DNA methylation of the same CpGs in the testis and in the ovary. The correlations of predicted and chronological age were still high (r=0.927, $p=1.404^{e-06}$, **Fig. 4B**) in the testis, but were lower (r=0.414, p=0.181, **Fig. 4C**) in the ovary. ANCOVAs between models with or without size (length) as factor showed that there was no significant effect on age prediction in muscle and testis (**Table 2**). However, there was a significant effect of size on the age prediction in the ovaries (**Table 2**). The overall methylation of the 299 CpGs showed no significant differences between the extreme age classes (1.28 vs 10.5 years) in the muscle (Kruskal-Wallis χ^2=0.289, p=0.591; **Fig. S3A**), testis (Kruskal-Wallis χ^2=3.306, p=0.069; **Fig. S3B**) or ovary (Kruskal-Wallis χ^2=0.252, p=0.615; **Fig. S3C**), indicating that differences in the selected CpGs were specific and not due to an overall age-related trend.

Discussion

In this study we constructed the first age predictor in fish based on an epigenetic clock. This piscine epigenetic clock involves age prediction via regressions of the DNA methylation levels of 48 CpGs. The predicted age highly correlates with the chronological age and the clock has good precision. In comparison to the clocks for warm-blooded animals developed so far, our piscine clock performs fairly well (**Table 3**). The precision of our clock was measured by the mean absolute error, which indicates the average absolute difference between the observed and the predicted values, and thus demonstrates the calibration efficiency.

The first epigenetic clock was constructed in humans using DNA methylation data obtained from microarrays on 21,369 CpGs from 8,000 samples. Following penalized regressions, a subset of

353 CpGs was used to predict age across tissues with 0.96 accuracy (Horvath, 2013). In subsequent studies, researchers took advantage of this knowledge in humans in order to target specific CpGs already known to associate with age. In our study, the initial dataset consisted of only 48 CpGs located in 4 genes which had showed no changes with extrinsic or intrinsic influences in previous studies (Anastasiadi et al., 2018) or known to be irrelevant to the core functions of muscle (Navarro-Martín et al., 2012), our target tissue. For the variant of the clock, the initial number was 299 CpGs which allowed selecting CpGs from a wider spectrum of candidate genes and at the end age prediction showed higher accuracy than the first model. Both approaches were considered valid and the workflow of choice for development of future clocks should, thus, depend on the compensation of allocation of resources during the first steps. Obtaining more candidate initial CpGs can provide the models with more CpGs to select from in order to accurately make predictions. However, even with 48 CpGs, for which no previous knowledge regarding age-related DNA methylation changes existed in this species, we were able to predict age. Here, we interrogated a small fraction of the genome. Nevertheless, it should be noticed that in elderly humans only 20% of the differentially methylated regions overlap with promoters and exons (Heyn et al., 2012) and that changes in DNA methylation with age are enriched in other parts of the genome such as enhancers and intragenic regions (Weidner & Wagner, 2014). In this study, an age effect was identified in the CpGs of the regulatory regions of our target genes.

Blood has been the preferred target tissue in mice, chimpanzees and short-tailed shearwaters, while skin was targeted in the humpback whales and wing punches in the Bechstein's bats due to the easiness of accessibility. In our case, the most intuitive tissue to target was muscle, since it is the edible part of the fish and it is easily accessible. However, muscle was one of the tissues that performed worst in the human epigenetic clock, probably due to its plasticity and its high stem cell content (Horvath, 2013). Despite this, we could build the sea bass epigenetic clock based on muscle samples. We tested the epigenetic age prediction in two reproductive tissues, testis and ovary, and found that it performs relatively well in testis but fails to accurately predict age in the ovaries. Since in fish growth is indeterminate and fecundity increases with age (Reznick et al., 2002), the ovaries may not reflect the chronological age in terms of DNA methylation, because they are expected to maintain their functional properties as repositories of germ cells regardless of age. Nevertheless, it could also be possible that the lower performance of the clock in the ovary

was due to the fact that ovaries came from fish different from those where muscle and testes were obtained. This calls for caution and indicates that the influence of genetics and tissue of origin on DNA methylation deserve more research. Importantly, an epigenetic clock in muscle involves lethal sampling. Although it can be useful to estimate fish age from muscle samples when fish organs are routinely dissected, for example during scientific surveys, it will be extremely beneficial to predict age via non-lethal samplings as well. Thus, further studies on piscine epigenetic clocks should also focus on tissues accessible via non-lethal techniques, such as finclips, gills or scales. The fins are ideal tissues for non-lethal sampling and have been used for years for all types of studies, therefore, an epigenetic clock based on fin should be the main target for future studies. The use of scales and otoliths obtained from the same fish would allow the cross-calibration with DNA methylation-based methods as the one presented here.

In other vertebrates, the epigenetic clocks have distinct tick rates in young and adult animals. For example, in humans the tick rate is high until adulthood is reached and then it slows down and stays constant (Horvath, 2013). In the short-tailed shearwaters, a larger variability in DNA methylation was found in younger ages, although the few samples of young birds was a limiting factor for this study (Paoli-Iseppi et al., 2019). In the European sea bass, on the contrary, we had a large number of juvenile individuals of less than 1.5 years. We regressed the DNA methylation of the 48 CpGs using the tuning parameters of our final model to predict the age of 200 to 600 dpf fish, separately from the adults and this model accurately predicted age. Moreover, we exposed siblings (to avoid any effect of genetic differentiation) to elevated temperature to study the effect of accelerated growth on age prediction. Growth in the high temperature was accelerated as expected (Navarro-Martín et al., 2009) and thus fish from the high temperature group were size-matched with fish chronologically younger from the low temperature group. In animals with indeterminate growth such as fish, size is an important characteristic that relates to other key life-history traits such as the onset of sex differentiation (Blázquez, Carrillo, Zanuy, & Piferrer, 1999; Díaz et al., 2013). Age prediction in the high temperature group resulted in similar accuracy and precision as in the low temperature group. This means that the epigenetic clock presented here is able to predict the chronological age of the sea bass independently of their size or the environmental conditions they have experienced, at least in terms of temperature, which is the main varying abiotic factor in the ocean.

In fisheries stock assessment, as well as for fish ecology, age constitutes one of the most prominent biological variables (Campana, 2001). The age of fish is used to calculate important life-history traits, such as growth parameters and natural mortality. Age and size are the most informative measures for the assessment of the demographic structure of fish stocks. However, the collection of ageing data is expensive and limited for many fish stocks, while for many species age data are directly not available (Maunder & Piner, 2015). The standard methods of age determination in fish, i.e., otolith- and scale-analysis are laborious and need to be assessed by experts, something that can lead to team-dependent discrepancies (Kolody, Eveson, & Hillary, 2016). Further, age estimation by otolith may give very poor estimates in some species and does not work in fish inhabiting tropical areas where growth seasonality is little or absent (Sparre & Venema, 1998). Thus, epigenetic age determination in fish, as demonstrated in this study in the European sea bass, could be an ideal replacement for traditional age determination methods. The epigenetic clock is highly accurate in other vertebrates as well as in fish as shown here. Using low cost protocols of targeted bisulfite sequencing, such as the one used in this study, the process is highly cost-effective and can be applied to any organism without the necessity of having a sequenced genome.

Here, we have shown the existence of a clock in the muscle, a tissue routinely accessible during samplings. However, since it has been shown that multi-tissue epigenetic clocks exist in mammals, it should be tested whether the same or a similar panel of CpGs is able to predict age using fins, gills or scales, all of them accessed without the need to sacrifice the fish and requiring low expertise to dissect them. In conclusion, piscine epigenetic clocks such as the one presented here hold great potential for use in fish population studies and fisheries science to accurately estimate fish age. In this respect then, the present study paves the way for similar efforts in other species.

Aside from age, another essential parameter for assessing the reproductive potential of fish stocks is the sex ratio. In some species, there are sex-specific growth rates or mortality, while sex ratios affect the egg production and fertilization rates. Sex, in some species with no external sexual dimorphism, is assessed by histology, which is time-consuming and in the case of juvenile fish is impossible. In the European sea bass, a method based on epigenetic markers, similar to the epigenetic clocks for age, has been developed that predicts sex with ~90% accuracy (Anastasiadi et al., 2018). Sex and age predictors based on DNA methylation could be easily combined in a

MBS-based approach. Together, these methods for sex and age prediction demonstrate the potential of integrating epigenetic markers into fisheries science in order to replace traditional time-consuming methods for estimating key population parameters.

Acknowledgments

We thank Dr. Ana Gómez (IATS-CSIC) for providing tissue samples, Elvira Martínez for assistance with fish husbandry, Marta Lombó and Margarita Metallinou for suggestions to improve the PCR conditions, Sílvia Joly for sample processing and Dr. Federico Quattrocchi for help with the statistical models. DA was supported by a Ph.D. scholarship from the Spanish Government (BES–2011–044860). Research supported by Spanish Ministry of Economy and Competitiveness "Epimark" (ref. AGL2016–78710–R) to FP.

References

Anastasiadi, D., Piferrer, F. (2019), A clockwork fish. Age-prediction using DNA methylation-based biomarkers in the European seabass, v2, Dryad, Dataset, doi: 10.5061/dryad.m0cfxpnz4.

Anastasiadi, D., Vandeputte, M., Sánchez-Baizán, N., Allal, F., & Piferrer, F. (2018). Dynamic epimarks in sex-related genes predict gonad phenotype in the European sea bass, a fish with mixed genetic and environmental sex determination. *Epigenetics*, *13*(9), 988–1011. doi: 10.1080/15592294.2018.1529504

Ayala, M. D., López-Albors, O., Gil, F., Latorre, R., Vázquez, J. M., García-Alcázar, A., … Moreno, F. (2000). Temperature effect on muscle growth of the axial musculature of the sea bass *Dicentrarchus labrax* L. *Anatomia, Histologia, Embryologia*, *29*(4), 235–242. doi: 10.1046/j.1439-0264.2000.00262.x

Berdyshev, G. D., Korotaev, G. K., Boiarskikh, G. V., & Vaniushin, B. F. (1967). Nucleotide composition of DNA and RNA from somatic tissues of humpback and its changes during spawning. *Biokhimiiâ˘(Moscow, Russia)*, *32*(5), 988–993.

Blázquez, M., Carrillo, M., Zanuy, S., & Piferrer, F. (1999). Sex ratios in offspring of sex-reversed sea bass and the relationship between growth and phenotypic sex differentiation. *Journal of Fish Biology*, *55*(5), 916–930. doi: 10.1111/j.1095-8649.1999.tb00730.x

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*(2), 197–242. doi: 10.1111/j.1095-8649.2001.tb00127.x

Díaz, N., Ribas, L., & Piferrer, F. (2013). The relationship between growth and sex differentiation in the European sea bass (*Dicentrarchus labrax*). *Aquaculture*, *408–409*, 191–202. doi: 10.1016/j.aquaculture.2013.05.032

EEA. (2017). *Environmental indicator report 2017 — In support to the monitoring of the 7th Environment Action Programme* (Publication No. 21). Retrieved from European Environment Agency website: https://www.eea.europa.eu/publications/environmental-indicator-report-2017

Felip, A., & Piferrer, F. (2018). The Induction of Polyploidy, Gynogenesis, and Androgenesis in the European Sea Bass. In *Sex Control in Aquaculture* (pp. 347–358). doi: 10.1002/9781119127291.ch16

Gavery, M. R., Nichols, K. M., Berejikian, B. A., Tatara, C. P., Goetz, G. W., Dickey, J. T., … Swanson, P. (2019). Temporal Dynamics of DNA Methylation Patterns in Response to Rearing Juvenile Steelhead (*Oncorhynchus mykiss*) in a Hatchery versus Simulated Stream Environment. *Genes*, *10*(5), 356. doi: 10.3390/genes10050356

Han, Y., Eipel, M., Franzen, J., Sakk, V., Dethmers-Ausema, B., Yndriago, L., … Wagner, W. (2018). Epigenetic age-predictor for mice based on three CpG sites. *ELife*, *7*, e37462. doi: 10.7554/eLife.37462

Heyn, H., Li, N., Ferreira, H. J., Moran, S., Pisano, D. G., Gomez, A., … Esteller, M. (2012). Distinct DNA methylomes of newborns and centenarians. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(26), 10522–10527. doi: 10.1073/pnas.1120658109

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

Hosomichi, K., Mitsunaga, S., Nagasaki, H., & Inoue, I. (2014). A Bead-based Normalization for Uniform Sequencing depth (BeNUS) protocol for multi-samples sequencing exemplified by HLA-B. *BMC Genomics*, *15*, 645. doi: 10.1186/1471-2164-15-645

Ito, H., Udono, T., Hirata, S., & Inoue-Murayama, M. (2018). Estimation of chimpanzee age based on DNA methylation. *Scientific Reports*, *8*(1), 9998. doi: 10.1038/s41598-018-28318-9

Janowitz Koch, I., Clark, M. M., Thompson, M. J., Deere-Machemer, K. A., Wang, J., Duarte, L., … vonHoldt, B. M. (2016). The concerted impact of domestication and transposon insertions on methylation patterns between dogs and grey wolves. *Molecular Ecology*, *25*(8), 1838–1855. doi: 10.1111/mec.13480

Jarman, S. N., Polanowski, A. M., Faux, C. E., Robbins, J., De Paoli-Iseppi, R., Bravington, M., & Deagle, B. E. (2015). Molecular biomarkers for chronological age in animal ecology. *Molecular Ecology*, *24*(19), 4826–4847. doi: 10.1111/mec.13357

Johnson, J. H., McKenna Jr., J. E., Dropkin, D. S., & Andrews, W. D. (2005). A novel approach to fitting the von Bertalanffy relationship to a mixed stock of Atlantic sturgeon harvested off the New Jersey Coast. *Northeastern Naturalist*, *12*(2), 195202. doi: 10.1656/1092-6194(2005)012[0195:ANATFT]2.0.CO;2

Jones, O. R., Scheuerlein, A., Salguero-Gómez, R., Camarda, C. G., Schaible, R., Casper, B. B., … Vaupel, J. W. (2014). Diversity of ageing across the tree of life. *Nature*, *505*(7482), 169–173. doi: 10.1038/nature12789

Jung, M., & Pfeifer, G. P. (2015). Aging and DNA methylation. *BMC Biology*, *13*, 7. doi: 10.1186/s12915-015-0118-4

Kolody, D. S., Eveson, J. P., & Hillary, R. M. (2016). Modelling growth in tuna RFMO stock assessments: Current approaches and challenges. *Fisheries Research*, *180*, 177–193. doi: 10.1016/j.fishres.2015.06.016

Maunder, M. N., & Piner, K. R. (2015). Contemporary fisheries stock assessment: many issues still remain. *ICES Journal of Marine Science*, *72*(1), 7–18. doi: 10.1093/icesjms/fsu015

Morretti, A. (1999). *Manual on Hatchery Production of Seabass and Gilthead Seabream*. Food & Agriculture Org.

Mugatroyd, C., Wu, Y., Bockmühl, Y., & Spengler, D. (2010). The Janus face of DNA methylation in aging. *Aging (Albany NY)*, *2*(2), 107–110. doi: 10.18632/aging.100124

Navarro-Martín, L., Blázquez, M., Viñas, J., Joly, S., & Piferrer, F. (2009). Balancing the effects of rearing at low temperature during early development on sex ratios, growth and maturation in the European sea bass (*Dicentrarchus labrax*). *Aquaculture*, *296*(3–4), 347–358. doi: 10.1016/j.aquaculture.2009.07.022

Navarro-Martín, L., Galay-Burgos, M., Piferrer, F., & Sweeney, G. (2012). Characterisation and expression during sex differentiation of Sox19 from the sea bass Dicentrarchus labrax. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *163*(3), 316–323. doi: 10.1016/j.cbpb.2012.08.004

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*(2), 411–425. doi: 10.1111/1755-0998.12981

Pardo, S. A., Cooper, A. B., & Dulvy, N. K. (2013). Avoiding fishy growth curves. *Methods in Ecology and Evolution*, *4*(4), 353–360. doi: 10.1111/2041-210x.12020

Polanowski, A. M., Robbins, J., Chandler, D., & Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources*, *14*(5), 976–987. doi: 10.1111/1755-0998.12247

R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.R-project.org

Reznick, D., Ghalambor, C., & Nunney, L. (2002). The evolution of senescence in fish. *Mechanisms of Ageing and Development*, *123*(7), 773–789. doi: 10.1016/S0047-6374(01)00423-7

Roach, D. A., & Carey, J. R. (2014). Population Biology of Aging in the Wild. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 421–443. doi: 10.1146/annurev-ecolsys-120213-091730

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946. doi: 10.1101/gr.128124.111

Romanov, G. A., & Vanyushin, B. F. (1981). Methylation of reiterated sequences in mammalian DNAs. Effects of the tissue type, age, malignancy and hormonal induction. *Biochimica Et Biophysica Acta*, *653*(2), 204–218.

RStudio Team. (2015). *RStudio: Integrated Development Environment for R*. Retrieved from http://www.rstudio.com/

Shimoda, N., Izawa, T., Yoshizawa, A., Yokoi, H., Kikuchi, Y., & Hashimoto, N. (2014). Decrease in cytosine methylation at CpG island shores and increase in DNA fragmentation during zebrafish aging. *AGE*, *36*(1), 103–115. doi: 10.1007/s11357-013-9548-5

Sparre, P., & Venema, S. C. (1998). *Introduction to Tropical Fish Stock Assessment: Manual*. Food and Agriculture Organization of the United Nations.

Thompson, M. J., vonHoldt, B., Horvath, S., & Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. *Aging*, *9*(3), 1055–1068. doi: 10.18632/aging.101211

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., … Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, *5*, 5770. doi: 10.1038/ncomms6770

Varriale, A. (2014). DNA Methylation, Epigenetics, and Evolution in Vertebrates: Facts and Challenges. *International Journal of Evolutionary Biology*, *2014*, e475981. doi: 10.1155/2014/475981

Varriale, A., & Bernardi, G. (2006). DNA methylation and body temperature in fishes. *Gene*, *385*, 111–121. doi: 10.1016/j.gene.2006.05.031

Venney, C. J., Johansson, M. L., & Heath, D. D. (2016). Inbreeding effects on gene-specific DNA methylation among tissues of Chinook salmon. *Molecular Ecology*, *25*(18), 4521–4533. doi: 10.1111/mec.13777

Weidner, C. I., & Wagner, W. (2014). The epigenetic tracks of aging. *Biological Chemistry*, *395*(11), 1307–1314. doi: 10.1515/hsz-2014-0180

Wright, P. G. R., Mathews, F., Schofield, H., Morris, C., Burrage, J., Smith, A., … Hamilton, P. B. (2018). Application of a novel molecular method to age free-living wild Bechstein's bats. *Molecular Ecology Resources*, *18*(6), 1374–1380. doi: 10.1111/1755-0998.12925

Data accessibility statement

All data related to this publication are publicly available: mapped reads and methylation values of each CpG for each sample can be found in https://doi.org/10.5061/dryad.m0cfxpnz4

Author Contributions

DA designed and performed research, analyzed data and wrote the article. FP conceived and designed research, contributed reagents and wrote the article.

Tables

Table 1. CpGs with non-zero coefficients in the final elastic net model

Gene	Chromosome	Position	Elastic Net Coefficients	Correlation	*p*-value
amh-r2	LG1A	20275423	0.651	0.390	0.005
	LG1A	20275622	0.003	0.228	0.111
fsh-r	LG1B	11628807	0.022	0.131	0.364
	LG1B	11628892	-0.126	-0.372	0.008
	LG1B	11628926	-6.902	-0.449	0.001
nr3c1	LG2	10393123	0.016	0.141	0.329
	LG2	10393151	0.163	0.094	0.517
	LG2	10393169	-0.058	-0.116	0.424
	LG2	10393197	0.655	0.204	0.155
	LG2	10393230	0.002	-0.005	0.974
	LG2	10393251	0.019	-0.022	0.877
	LG2	10393253	-0.099	-0.048	0.743
	LG2	10393322	0.020	0.024	0.867
	LG2	10393372	0.313	0.136	0.347
	LG2	10393403	-0.161	-0.153	0.289
	LG2	10393443	-0.275	-0.052	0.722
sox19a	LG3	78269	-3.826	-0.410	0.003
	LG3	78301	0.225	0.478	0.000
	LG3	78394	-0.008	-0.002	0.988
	LG3	78426	0.135	0.212	0.139
	LG3	78429	0.052	0.394	0.005
	LG3	78447	0.018	0.217	0.129
	LG3	78460	0.063	0.230	0.109
	LG3	78511	0.042	0.106	0.465
	LG3	78514	0.175	0.276	0.053
	LG3	78570	-0.044	-0.002	0.987
	LG3	78590	-0.003	0.107	0.461
	LG3	78593	0.083	0.334	0.018

Correlation coefficients and their corresponding p-values are reported as tested by Pearson's correlation of the DNA methylation of each CpG with the chronological age

Table 2. Analysis of Covariance for evaluating the effect of length on age prediction

Tissue	Model	Residual df	Residual SS	SS	F	*p*-value
Muscle	Model 1	8	0.009			
	Model 2	10	0.013	-0.003	1.544	0.271
Testis	Model 1	8	6.445			
	Model 2	10	8.999	-2.554	1.585	0.263
Ovary	Model 1	8	13.377			
	Model 2	10	35.083	-21.706	6.491	**0.021**

Model 1: $Age_{predicted} = Age_{chronological} * Length$

Model 2: $Age_{predicted} = Age_{chronological}$

Abbreviations: df, degrees of freedom; SS, sums of squares

Table 3. Epigenetic clocks for age prediction in vertebrates

Vertebrate	Species	Tissue	Method	Initial CpGs	Final CpGs	Accuracy	Precision	Reference
Mammal	*Homo sapiens*	Multi-tissue	Microarrays	21369	353	0.96	3.60	Horvath (2013)
Mammal	*Megaptera novaeangliae*	Skin	Pyrosequencing	37	3	0.79	2.99	Polanowski et al (2014)
Mammal	*Mus musculus*	Blood	Pyrosequencing	71	3	0.95	5.29	Han et al (2018)
Mammal	*Pan troglodytes*	Blood	Pyrosequencing	14	4	0.73	5.43	Ito et al (2018)
Mammal	*Myotis bechsteinii*	Wing punches	Pyrosequencing	7	7	0.58	2.08	Wright et al (2018)
Mammal	*Canis familiaris*	Blood	RRBS	252,240	41	1.00	0.05	Thompson et al (2017)
Mammal	*Canis lupus*	Blood	RRBS	252,240	67	0.97	0.04	Thompson et al (2017)
Bird	*Ardenna tenuirostris*	Blood	DREAM	2338	7	0.59	2.81	Paoli-Iseppi et al (2019)
Fish	*Dicentrarchus labrax*	Muscle	MBS	48	28	0.82	2.15	This study

Abbreviations: RRBS, Reduced Representation Bisulfite Sequencing; DREAM, Digital Restriction Enzyme Analysis of Methylation; MBS, Multiplex Bisulfite Sequencing

Figure legends

Figure 1. Accuracy and precision of age prediction in fish muscle. **A**) Predicted age based on the DNA methylation of 48 CpGs obtained by applying elastic net regression as compared to chronological age. Accuracy is measured by Pearson's correlation coefficient (r) and its significance level (*p*-value). Precision of the clock calculated by Leave-One-Out-Cross-Validation (LOOCV) as mean absolute error (MAE) in years. **B**) Distribution of the predicted age binned into three age categories after prediction: 0 to 3, 3 to 10 and more than 10 years old represented by boxplots, where boxes show the interquartile distribution (IQR), the upper whisker=Q3+1.5*IQR, the lower whisker=Q1-1.5*IQR and the black lines show the median. Letters (a, b and c) indicate that there are differences evaluated by Analysis of Variance**. C)** Relationship of chronological age (days) with length (cm) for the samples used in the construction of the clock adjusted by the von Bertalanffy curve.

Figure 2. DNA methylation of individual CpGs relates with chronological age. Examples of individual CpGs selected based on the significance of the correlations of their DNA methylation with age. Colors indicate the localization of CpGs on different genes. Lines indicating the linear regressions of DNA methylation to age are added for visual purposes. The full list of CpGs with the elastic net (final model) and Pearson's correlation coefficients can be found in **Table 1**.

Figure 3. Effect of accelerated growth due to temperature on the age prediction of juvenile fish. **A**) Relationship between chronological age (days) and length (cm) in the low (blue) and the high (red) temperature group represented by linear regression lines. The arrows indicate fish from the same age of 500 days which had differences in length as assessed by Welch two sample t-test**. *** = $p < 0.001$. **B**) Correlation of predicted age by elastic net regression of the DNA methylation of the 48 CpGs and chronological age. The Pearson's correlation coefficients (r) and their significance level (*p*) are reported for fish reared at low (blue) and high (red) temperature. **C**) Relationship of DNA methylation in 500 days old fish reared at low and high temperature. Each point represents an individual CpG colored according to the gene where it is located. Both axes are log-scaled.

Figure 4. Age prediction using 295 CpGs in **A**) muscle, **B**) testis, and **C**) ovary. Elastic net regressions of the DNA methylation of the 298 CpGs were used to predict age in the muscle. The

final model, with 63 CpGs with non-zero coefficients, was used to predict age in the testis and ovary. Age predictions was evaluated by Pearson's correlation (r) and its significance level (*p*). In (A), the original dots were almost fully overlapping, thus, they were minimally displaced in order to be visible.

men_13111_f1-4.pptx

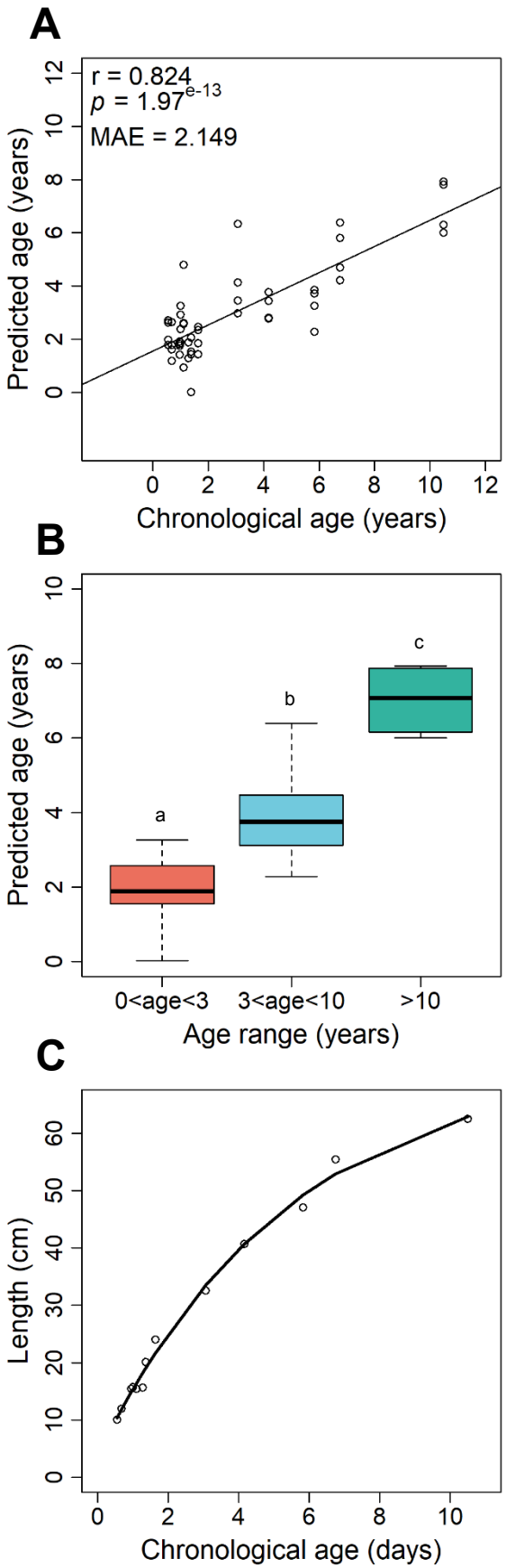

Figure 1

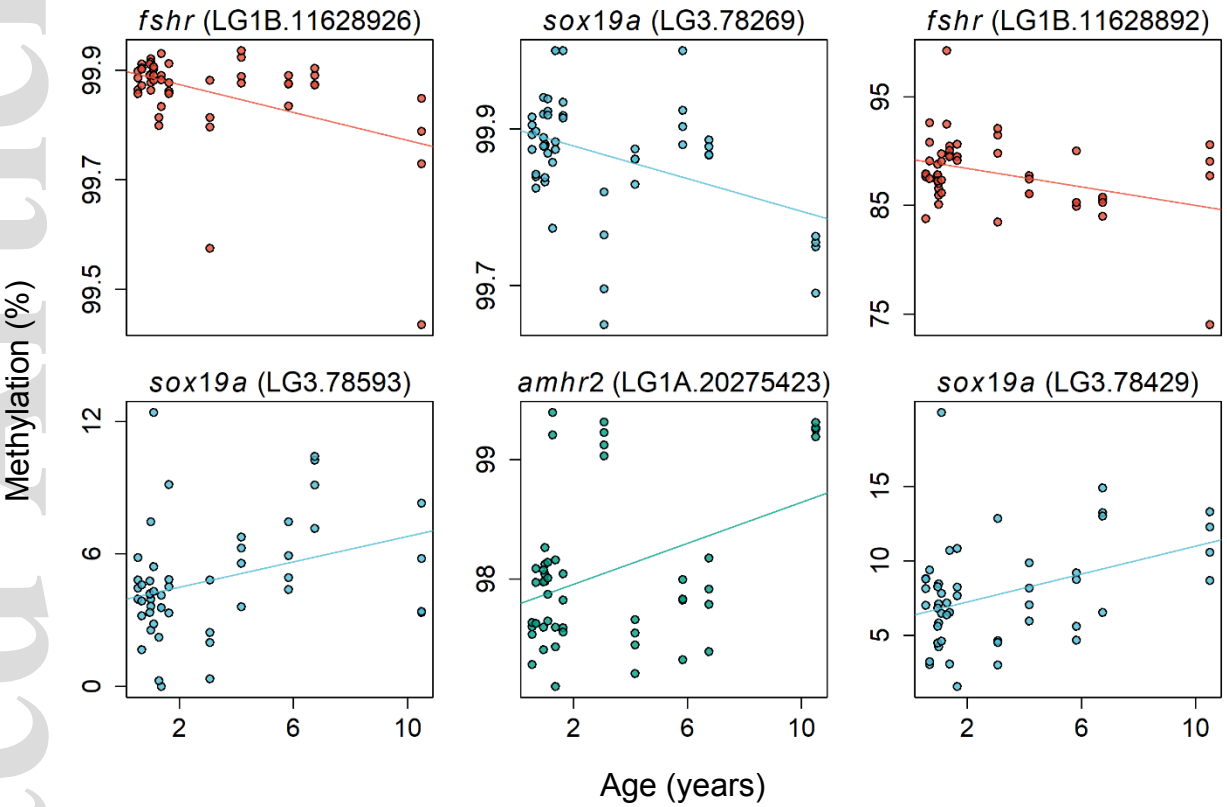

Figure 2

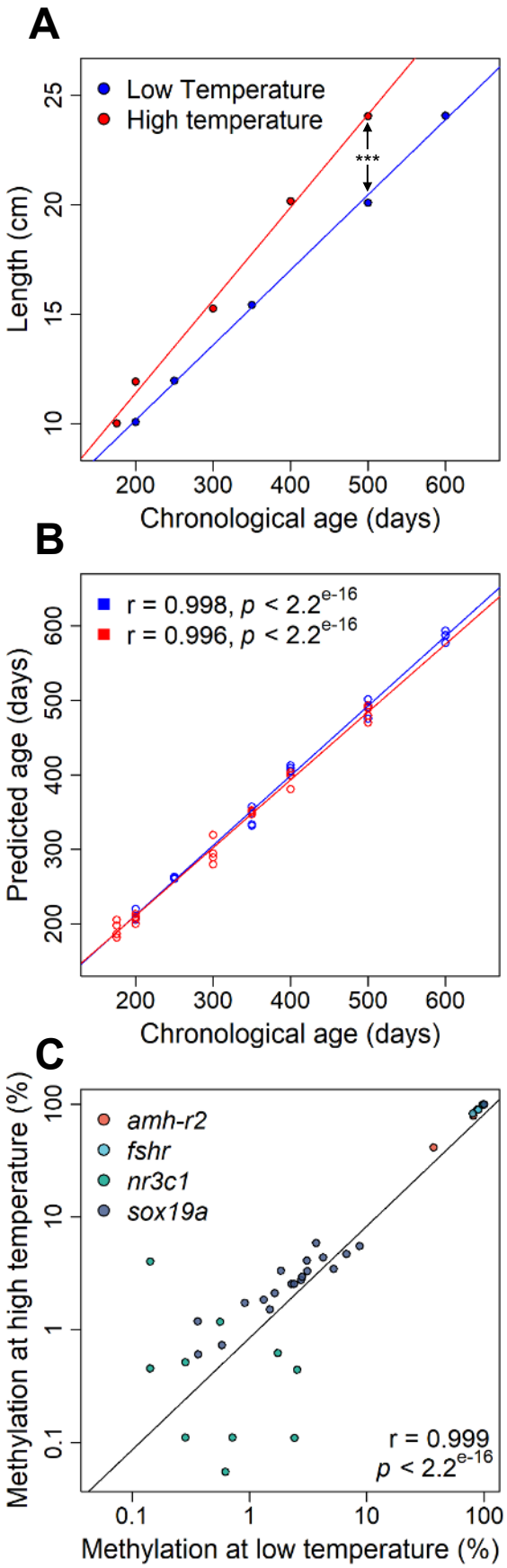

Figure 3

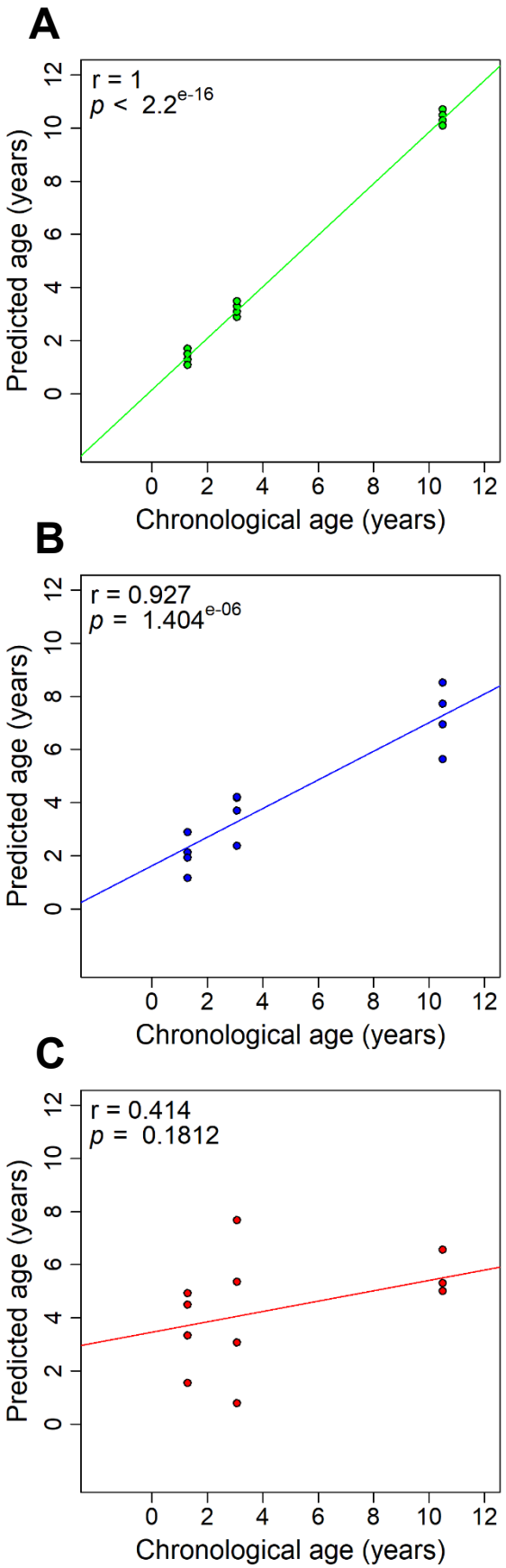

Figure 4