6 Lifespan estimation from genomic analysis

Benjamin Mayne and Christopher Faulk

ABSTRACT

Lifespan is a fundamental parameter of population biology and central to wildlife management. Lifespan limits individual reproductive capacity, probability of mortality, population growth rates, dispersal capacity, and viability. The lifespan of most species is unknown and this limits objective wildlife management. On a genomic level it has been shown the CpG density in gene promoters is both associated and predictive of lifespan. In this chapter, we discuss the development of lifespan clocks that produce estimates of animal lifespans from genome sequences. We also discuss the biological and evolutionary mechanisms of CpG density and lifespan. DNA can be readily obtained from species and a lifespan prediction may provide life-history traits. Knowing the lifespan of a species can provide better resources for wildlife management. Our research shows the benefit of using DNA to predict the lifespan of species, such as in predicting lifespan in species that outlive researchers or are extinct. It also provides insight into the evolutionary mechanism of the variability of species lifespan.

INTRODUCTION

Maximum lifespan is a fundamental parameter for management of any wild animal species. Maximum lifespan can be used to assess both extinction risk and the spread of **invasive species** (IUCN 2008; Tabak *et al.* 2018). Unfortunately, lifespan for most species is either unknown or is estimated from a limited number of individuals. In many cases, the individuals have been kept captive and this may lead to lifespans

Copyrighted material

that are artificially extended beyond the normal wild limit. The problem is intensified with long-lived species as they may outlive a generation of researchers. The lack of lifespan information makes it difficult for wildlife managers to decide on appropriate management for most species.

Mortality rates, life expectancy, and longevity (life expectancy at birth) are all determined by maximum lifespan (Hsieh 1991). These parameters are used in a broad spectrum of population biology analyses and minute changes can influence wildlife management. It is imperative the appropriate value for maximum lifespan is used as it can potentially impact the management of a species. For example, in fisheries catch limits are highly dependent upon natural mortality rates. Catch limits can be difficult to determine when mortality rates are poorly defined (Then *et al.* 2014; Kenchington 2014). The orange roughy (*Hoplostethus atlanticus*) population collapsed due to a poorly defined lifespan and consequent excess harvesting (Boyer *et al.* 2001; Clark 2001). Mortality rates can be determined using lifespan, preventing similar situations in the future (Hoenig 2005). A rapid approach for calculating fundamental parameters of life history is critical to better manage populations in a rapidly changing global environment with more species to manage than can be resourced (Healy *et al.* 2019).

Of the many characteristics of the genome, cytosine-phosphate-guanine (CpG) density is one of the less explored (Box 6.1). There are four nucleotide letters used in DNA, and 16 possible combinations of pairs. Within vertebrate genomes, **CpG sites** are the fastest evolving dinucleotides, and their density proximal to genes is under strong evolutionary selection. Similarly, lifespan is a trait that evolves under unusual selective pressure with wide variability in vertebrates and is of great interest in both ecological contexts and human health regarding ageing.

In this chapter, we demonstrate the lifespan clock as a predictor of lifespan in vertebrates (Mayne *et al.* 2020). This chapter focuses on the construction of the first lifespan clock and gives example applications to bowhead whales and marine turtles. We also discuss potential future directions of the work presented and how the reader can develop their own lifespan clock on species not covered in this chapter. This may include non-vertebrate animal groups. We also discuss some of the evolutionary aspects and other genomic features associated with lifespan.

Box 6.1: The evolution of CpG sites

Copyrighted material

Many of the intrinsic factors that determine the lifespans of species are genomic (Kenyon 2010). CpG sites in the genome host epigenetic **DNA methylation** that changes with age unlike nucleotide sequence (see Chapter 7). In other words, the 'C' within a CpG site can be either methylated or unmethylated, and the state of methylation changes over the lifespan of an individual (Kanherkar *et al.* 2014). DNA methylation suppresses gene expression, so intergenic regions and parts of the genome that must be silenced are highly methylated in vertebrates. In contrast, parts of the genome that must be active often have very low levels of DNA methylation, most notably in promoter **CpG islands**. CpG islands exhibit both genetic and epigenetic conservation across the vertebrate radiation. In fact, experiments have shown that CpG density alone is predictive of methylation level (Long *et al.* 2016). This epigenetic mark also has evolutionary consequences. A methylated C can become deaminated over evolutionary time and transition to T, creating a TpG site. Comparative genomics reveals that CpG sites mutate away at a higher rate than any other dinucleotide because of the high frequency of methylated CpG sites in the genome. The result is that the CpG dinucleotide has the lowest observed vs. expected frequency within vertebrate genomes (Fig. 6.1). The case for the outsize importance of the CpG sites is illustrated in humans and our nearest relative, the chimpanzee. The genetic difference between us and chimps averages 0.92% per base pair across the genome; however, at CpG sites that rate jumps to over 15%. In other words, between us and chimps, there is 1 SNP out of a 100 bases on average, but that rises to 1 out of 7 at CpG sites. Because CpG sites mutate rapidly and host DNA methylation that ultimately governs gene expression, they are the both a primary cause of genetic variation and the substrate for both epigenetic and gene regulatory changes (Antequera 2003; Deaton and Bird 2011; Bell *et al.* 2012).

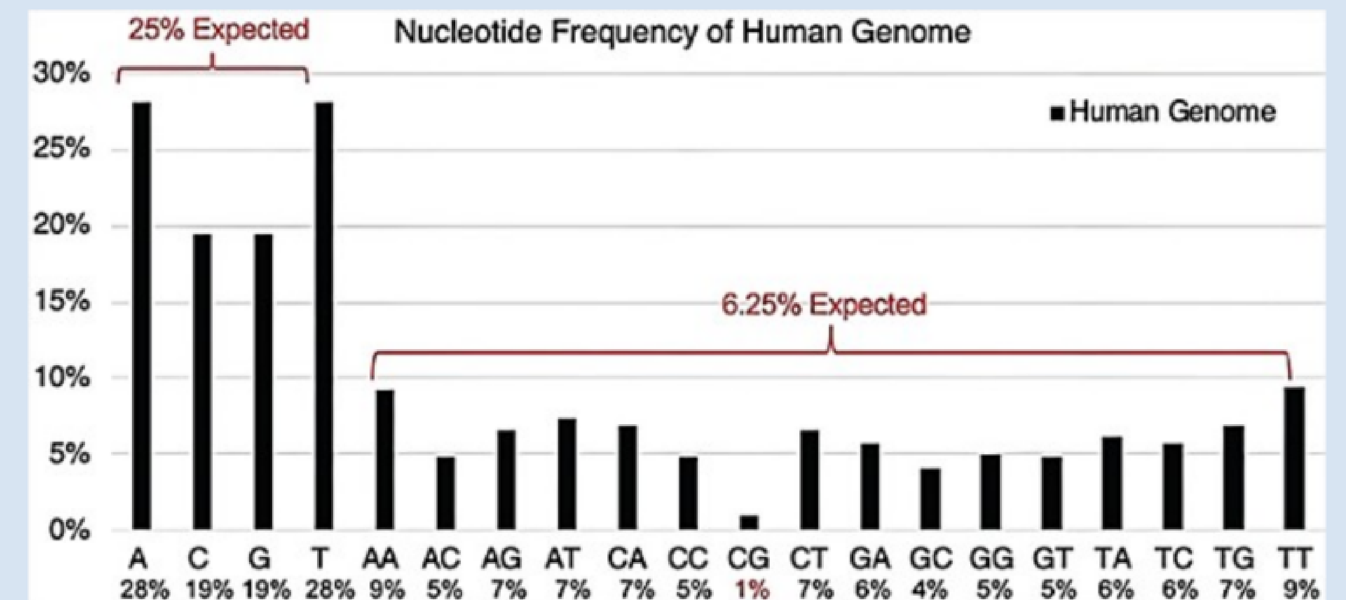

Fig. 6.1. Mutation rate isn't equal for every type of base. The genome is less GC rich than expected by chance and there are fewer CpG sites than any other pair of nucleotides.

Copyrighted material

APPLICATION

The construction of the lifespan clock builds upon two domains of research. The first is the work by McLain & Faulk (McLain and Faulk 2018). McLain & Faulk drew upon multiple databases including the Eukaryotic Promoter Database (EPD), a source of promoter sequences (Dreos *et al.* 2017), the National Center for Biotechnology Information (NCBI) genomes database (https://www.ncbi.nlm.nih.gov/genome/); and the Animal Ageing and Longevity Database (AnAge), a source of lifespans of species (Tacutu *et al.* 2018). McLain & Faulk found the CpG density in 5% of all promoters significantly correlated with lifespan in mammals.

The second research domain is the methodology commonly used with **epigenetic clocks** (Chapter 7). Epigenetic clocks have been developed for a wide variety of vertebrates and can predict age from the DNA methylation levels of selected CpG sites (De Paoli-Iseppi *et al.* 2017). One of the common methods of developing an epigenetic clock is using an **elastic net regression** model (Friedman*et al.* 2010). Briefly, as it is outside the scope of this chapter, an elastic net regression model can fit a linear model and identify the minimum number of predictors required to predict an outcome. In the case of the lifespan clock the predictors are CpG densities in promoter sequences and the outcome is lifespan. The elastic net regression model is ideal for handling the data presented in this chapter as it can work with many predictors and small sample size. The model of Mayne *et al.* identified 42 promoters conserved across vertebrates as the minimum number of markers required for accurate prediction of lifespan (Mayne *et al.* 2019). These 42 promoters encompass what is referred to as the lifespan clock.

Example: Lifespan prediction of the bowhead whale by genomic analysis

The wildlife management of long-lived species that do not have known lifespan benefits greatly from the lifespan clock. Although it is ideal to use the lifespan clock with reference genomes (see Chapter 9), it is possible to predict lifespan using basic molecular approaches. In this chapter we give an example of how to predict lifespan with and without a **reference genome**.

$$ln\ (maximum\ lifespan) = -4.38996 + (2.57328 \times 4.65) + (-0.92888 \times 4.65) + 2.33508$$

$$maximum\ lifespan = e^{5.59158} = 268\ years$$

Copyrighted material

Table 6.1. The coefficients to be inputted into the lifespan prediction equation for each vertebrate class.

Class	*a*	*b*
Aves	-0.90323	2.14857
Fish	2.14632	-6.58228
Mammalia	-0.92888	2.33508
Reptilia	-0.48958	1.17281

Bowhead whales are thought to be the longest-lived mammal (Keane *et al.* 2015). The oldest aged bowhead whale was found to be 211 years old (George *et al.* 1999). This 211-year-old individual showed no pathological features of age-related diseases, suggesting they could potentially live longer (George *et al.* 1999). The bowhead whale has a reference genome that can be used for lifespan prediction (Keane *et al.* 2015). Lifespan prediction with a reference genome can occur in three main steps. First, the 42 unique promoter sequences are identified using a Basic Local Alignment Search Tool (**BLAST**). BLAST can rapidly identify similar sequences within the genome. The sequences of the promoters used by BLAST are found in the link within the Resources section of this chapter. The second step involves calculating the CpG density. CpG density is the total number of CpG sites divided by the length of the BLAST hit. The final step is to input the CpG densities into the lifespan prediction equation, shown below.

$$ln\ (maximum\ lifespan) = -4.38996 + 2.57328x + ax + b$$

where x is the raw summed CpG density weight per sample and a and b are coefficients relating to the vertebrate class (Table 6.1).

For each promoter the CpG density is multiplied by a unique coefficient that was calculated during the model development. These coefficients can be found in a link within the Resources section. The sum of the product of the CpG densities and coefficients are represented in the equation by x. For the bowhead whale $x = 4.65$. We can then input x and the mammalian specific coefficients (Table 6.1) into the lifespan equation.

As demonstrated with the bowhead whale, a reference genome can provide a straightforward approach to predict lifespan. BLAST is a common bioinformatic tool

Copyrighted material

used to search for similar sequences. CpG densities can be readily calculated once the promoter sequences have been identified and using basic algebraic methods, lifespan can be predicted using the equation above.

Example: Marine turtle lifespan

Marine turtles are both slow growing and long-lived, with the potential to outlive a generation of researchers (Musick and Limpus 1997). Lifespan values for marine turtle species have the potential to provide better predictions for population growth, survival probabilities, and risk of extinction (Hoenig 2005). In this example the lifespans of marine turtles that occur in Australian waters, but lacking a reference genome, was predicted (Mayne *et al.* 2020). First, tissue from each marine turtle was collected for DNA extraction. Any tissue from a somatic cell can be used for lifespan estimation. Of the six marine turtles occurring in Australian waters only the Green sea turtle (*Chelonia mydas*) has a reference genome and lifespan estimate in the AnAge database (Wang *et al.* 2013; Tacutu *et al.* 2018).

The green sea turtle genome was used to BLAST the lifespan promoters and to design PCR primers. Primers were then optimised using DNA from each marine turtle species. By optimising primers for closely related species negates the need to have a reference genome for each species. PCR was used to amplify the promoters of the marine turtles that were then subject to Sanger sequencing. Sanger sequencing enables to calculate the CpG density of each promoter. Once the CpG density has been determined for each promoter the process above can be used without deviation. Table 6.2 details the lifespan prediction for each marine turtle species. In this chapter we have shown how lifespan can be predicted with and without a reference genome. The bowhead whale and marine turtles are examples of long-lived species where access to estimates of their lifespan can contribute significantly to their management.

Copyrighted material

Table 6.2. The lifespan predictions of marine turtles using sanger sequencing to determine CpG density.

Species	Prediction
Leatherback sea turtle (*Dermochelys coriacea*)	90.4
Loggerhead sea turtle (*Caretta caretta*)	62.8
Olive Ridley sea turtle (*Lepidochelys olivacea*)	54.3
Hawksbill sea turtle (*Eretmochelys imbricata*)	53.2
Flatback sea turtle (*Natator depressus*)	50.4

FUTURE DIRECTIONS

Lifespan prediction from genomics has many future directions with implications that could benefit the management of wildlife. The lifespan clock was developed with five classes of vertebrate species. Multiple lifespan clocks can be developed for specific taxonomic groups (Box 6.2). Lifespan clocks calibrated to taxonomic groups of interest may increase the accuracy of clock predictions for species within these groups. One of the challenges of developing lifespan clocks for a broader range of species is the lack of reliable lifespan data for calibration. Lifespans are often difficult to obtain due to the longevity of long-lived species. A workaround may be to develop a lifespan clock on short-lived species that can be held in captivity. The subsequent clock that is developed could potentially be extrapolated to long-lived species. Granted, extrapolating any model beyond the confines of the training can result in false-positives and false-negatives (Forbes *et al.* 2008). It should be noted that the purpose of any model is to provide guidance to a challenge or problem. A lifespan clock may not be able to provide the exact number of years a species can live to but may be able to provide potential brackets or ranges.

So far in this chapter, the reader has had to have significant bioinformatic and machine learning knowledge to determine lifespan for their species of interest. However, it should be acknowledged that most researchers may not have the time or expertise to carry out the analysis. A future project for this work is to automate the method, through a web-based upload. Here, users could upload the reference genome and receive a species lifespan prediction. This would provide simple efficient method alternative to performing the bioinformatic analysis by oneself.

Box 6.2: How to develop a novel lifespan clock

Copyrighted material

Step 1: Assemble a dataset of known lifespan species with reference genomes.

There are three components that are required to develop a novel lifespan clock: reference genomes; well-defined lifespans, most likely to be obtained from sources such as AnAge (Tacutu *et al.* 2018), promoter sequences of the group of species that will be used to determine CpG density. Some examples of promoter databases include EPD and the Ensembl genome database (Hubbard *et al.* 2002; Tacutu *et al.* 2018).

Step 2: Determine promoter CpG density.

Once the list of species with reference genomes and well-defined lifespans has been collated, the next step is to determine the CpG density of the promoters for each species. This can be done using software such as BLAST. In the original studies, a BLAST hit was considered with an identity of identity >70% (McLain and Faulk 2018; Mayne *et al.* 2019). Once the BLAST hit sequences have been determined, CpG density can be calculated by dividing the frequency of CpG sites by the BLAST hit length. This part is the most computationally expensive as it requires the download of potentially large genomes and creating temporary BLAST databases to search for promoter sequences.

Step 3: Machine learning to develop a lifespan clock

The final part to develop a lifespan predictor is to calibrate and test the model. Typically, 70% of the samples should be randomly assigned to a training dataset and the remaining to a validation dataset. Since lifespan is a continuous outcome, a regression algorithm can be used. In the original study, an elastic net regression, implemented in the glmnet R package was used to develop the lifespan predictor (Friedman *et al.* 2010; Mayne *et al.* 2019).

This method is first applied to the training dataset to identify the promoters that can predict lifespan. These promoters are then tested on the validation data. The performance of the model can be assessed by comparing the known and predicted lifespans with Pearson correlations, absolute and relative error rates. It is important to ensure the model is not overfitted. This can be done by comparing the correlations, absolute and relative error rates between the training and validation data. Ideally, no difference should be observed. The method described here can be carried out on any taxonomic group to develop novel lifespan clocks.

DISCUSSION

Can a lifespan clock be developed for invertebrates?

CpG methylation is sparse within invertebrate genomes, despite their overall greater genome wide CpG density. To understand why, it is helpful to examine mutation in the context of evolutionary pressure. The ancestral animal genome, prior to the divergence of Chordata from other phyla, probably had CpG sites in similar frequency to all other dinucleotides (Fig. 6.2) Within the newly evolved vertebrate clade, DNA methylation also evolved as a means of gene control and transposon suppression (Box 6.3). Therefore, most CpG sites became methylated, except for ones in the promoter regions nearby genes since that would have suppressed important gene expression, CpG sites near genes remained unmethylated and were spared from deamination mutations. Over time, the methylated CpG sites outside of promoters mutated rapidly away leaving the present pattern of CpG islands within vertebrates. Thus, the accumulation of CpG islands was a consequence, not a cause, of evolution (Sharif *et al.* 2010). For this reason, it is unlikely that a 'lifespan clock' could be built for species with very sparse DNA methylation using CpG density.

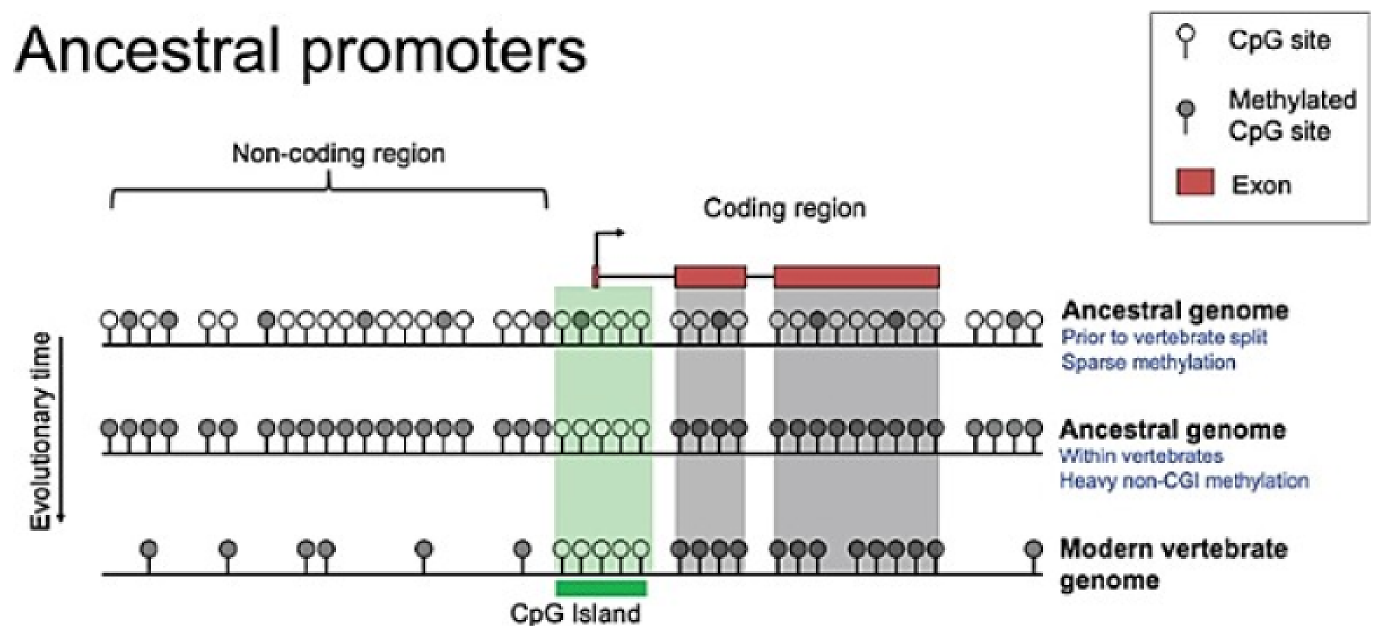

Fig. 6.2. Ancestral animal genomes had sparse methylation with no deviation in expected CpG site frequency. During vertebrate evolution, heavy DNA methylation became standard at CpG sites. Due to rapid mutation of methylated CpGs, the genome lost most CpG sites outside of promoter CpG islands where they are generally unmethylated. Note that CpG sites within coding regions are under negative selective pressure to avoid coding substitutions in amino acids.

Box 6.3: The ageing epigenome

As animals age, the strongly defined epigenetic pattern of DNA methylation degrades, with most highly methylated CpG sites gradually becoming demethylated, while relatively unmethylated CpG islands increase in methylation. In fact, this pattern is predictable enough that careful selection of CpG sites has allowed the development of epigenetic clocks to measure the ageing process by the change in methylation over chronological time (Wagner

Copyrighted material

Copyrighted material

2017). Since the methylation of CpG sites increases as an animal ages, the evolution of CpG sites becomes crucial to understanding ageing. Within mammals, coding regions of genomes are quite similar, yet lifespan varies widely. The epigenetic clock ticks according to longevity. Evidence supporting this case has been shown with mice carrying a complete copy of human chromosome 21. When epigenetic 'clock sites' on hum_chr21 were assessed, the human chromosome aged at the rate of the host mouse, suggesting that trans-factors govern ageing rate (Lowe *et al.* 2018). Similarly, dog epigenomes age at drastically different rates depending on the lifespan of the breed, despite being extremely genetically similar overall. Therefore, strong evidence indicates that ageing-associated methylation marks are a dynamic molecular readout of lifespan variation among different mammalian species.

What is the role of CpG density with lifespan?

Very CpG dense promoter regions tend to be hypomethylated and resist the natural increase in DNA methylation as we age. We hypothesise that CpG density itself is under selective pressure in some genes to maintain expression stability for a long life, in effect 'guarding' genes necessary for the continued homeostasis of organisms during their extended lifespan (Gardiner-Garden and Frommer 1987; Antequera 2003; Mayne *et al.* 2019). Ultimately, CpG density links both genetic and epigenetic variation across both evolutionary time across species and within the lifespan of a single individual.

What other genomic features are associated with lifespan?

The clustering of genomic features in general has been understudied with respect to lifespan. For example, CpG density itself has only recently been examined in relation to gene function. A recent paper by Boukas *et al.* demonstrated that CpG density is highest in genes most intolerant to **loss of function (LOF) mutations** (Boukas *et al.* 2020). They conclude that 'high CpG density is not merely a generic feature of human promoters but is preferentially encountered at the promoters of the most selectively constrained genes, calling into question the prevailing view that CpG islands are not subject to selection'.

Similarly, other genomic features besides coding regions should be assessed for their roles in the evolution of lifespan. The best example of this is the activity and consequences of transposons. While animals all have in the range of 10 to 20 thousand genes, the number of transposons in humans and mice is on the order of 4 million, or about 40–50% of the genome (Canapa *et al.* 2015). Interestingly, though most animals have about the same proportion of transposons, it appears that each clade has a unique expansion and complement of transposons (Platt *et al.* 2018). They also cluster in the genome in non-random ways and affect nearby gene expression, and are understudied in terms of lifespan due to difficulty in sequence mapping. Other features under natural selection for density have been studied but not in the context of lifespan include transcription factor motifs, SNP density in molecular evolution, target motifs for iron-binding proteins, and even olfactory receptor DNA binding motifs (Zhao *et al.* 2003; Sonntag *et al.* 2004; Kim *et al.* 2007; Faulk and Kim 2009).

CONCLUDING REMARKS

Rapid prediction of lifespan has the potential to advance many aspects of wildlife management. A lifespan clock reduces the need to monitor species from birth to death to obtain a lifespan estimate. Therefore, parameters relating to population dynamics can be immediately determined and evaluations of populations more accurately assessed. Advancements in lifespan clocks are still required, including expanding into invertebrates and increasing automation. In addition, more work is required to better understand the evolutionary aspects and basic biology relating lifespan and the genome.

DISCUSSION TOPICS

1. What is maximum lifespan? Discuss the definitions of lifespan, and how different definitions may affect interpretation of the estimates provided by lifespan-predicting models.
2. What considerations need to be made when assembling genomes and lifespan data for calibration of a lifespan prediction model?
3. Lifespan is one life history characteristic that can be estimated from genome sequence analysis. What other life history parameters do you expect might be predictable from similar models?

RESOURCES

Promoter sequences and coefficients to predict lifespan:
https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-019-54447-w/MediaObjects/41598_2019_54447_MOESM3_ESM.xlsx

REFERENCES

Copyrighted material
Copyrighted material

Antequera F (2003) Structure, function and evolution of CpG island promoters. *Cellular and Molecular Life Sciences CMLS* **60**(8), 1647–1658. doi:10.1007/s00018-003-3088-6

Bell CG, Wilson GA, Butcher LM, Roos C, Walter L, *et al.* (2012) Human-specific CpG 'beacons' identify loci associated with human-specific traits and disease. *Epigenetics* **7**(10), 1188–1199. doi:10.4161/epi.22127

Bjornsson HT, Hansen KD (2020) Promoter CpG density predicts downstream gene loss-of-function intolerance. *The American Journal of Human Genetics* **107**(3), 487–498. doi:10.1016/j.ajhg.2020.07.014

Boyer D, Kirchner CH, McAllister MK, Staby A, Staalesen BI (2001) The orange roughy fishery of Namibia: lessons to be learned about managing a developing fishery. *South African Journal of Marine Science* **23**(1), 205–221.

Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E (2015) Transposons, genome size, and evolutionary insights in animals. *Cytogenetic and Genome Research* **147**(4), 217–239. doi:10.1159/000444429

Clark M (2001) Are deepwater Fisheries sustainable? – the example of orange roughy (*Hoplostethus atlanticus*) in New Zealand. *Fisheries Research* **51**, 123–135.

Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. *Genes & Development* **25**(10), 1010–1022. doi:10.1101/gad.2037511

Dreos R, Ambrosini G, Groux R, Cavin Périer R, Bucher P (2017) The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic Acids Research* **45**, D51–D55. doi:10.1093/nar/gkw1069

Faulk CD, Kim J (2009) YY1's DNA-Binding motifs in mammalian olfactory receptor genes. *BMC Genomics* **10**(1), 576. doi:10.1186/1471-2164-10-576

Forbes VE, Calow P, Sibly RM (2008) The extrapolation problem and how population modeling can help. *Environmental Toxicology and Chemistry* **27**(10), 1987–1994. doi:10.1897/08-029.1

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

Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. *Journal of Molecular Biology* **196**(2), 261–282. doi:10.1016/0022-2836(87)90689-9

George J, Bada J, Zeh J, Scott L, Brown SE, *et al.* (1999) Age and growth estimates of bowhead whales (*Balaena mysticetus*) via aspartic acid racemization. *Canada Journal of Zoology* **77**(4), 571–580.

Healy K, Ezard TH, Jones OR, Salguero-Gómez R, Buckley YM (2019) Animal life history is shaped by the pace of life and the distribution of age-specific mortality and reproduction. *Nature Ecology & Evolution* **3**(8), 1217–1224. doi:10.1038/s41559-019-0938-7

Hoenig J (2005) Empirical use of longevity data to estimate mortality rates. SEDAR33RD17. SEDAR, North Charleston, SC, USA.

Hsieh JJ (1991) A general theory of life table construction and a precise abridged life table method. *Biometrical journal, Biometrische Zeitschrift* **33**(2), 143–162. doi:10.1002/bimj.4710330204

Hubbard T, Barker D, Birney E, Cameron G, Chen Y, *et al.* (2002) The Ensembl genome database project. *Nucleic Acids Research* **30**(1), 38–41. doi:10.1093/nar/30.1.38

IUCN (2008) *Strategic Planning for Species Conservation: A Handbook, The Species Conservation Planning Task Force*. IUCN Species Survival Commission, Gland.

Kanherkar RR, Bhatia-Dey N, Csoka AB (2014) Epigenetics across the human lifespan. *Frontiers in Cell and Developmental Biology* **2**, 49. doi:10.3389/fcell.2014.00049

Keane M, Semeiks J, Webb AE, Li YI, Quesada V, *et al.* (2015) Insights into the evolution of longevity from the bowhead whale genome. *Cell Reports* **10**(1), 112–122. doi:10.1016/j.celrep.2014.12.008

Kenchington TJ (2014) Natural mortality estimators for information-limited fisheries. *Fish and Fisheries* **15**(4), 533–562. doi:10.1111/faf.12027.

Kenyon CJ (2010) The genetics of ageing. *Nature* **464**(7288), 504–512. doi:10.1038/nature08980

Kim JD, Faulk C, Kim J (2007) Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. *Nucleic Acids Research* **35**(10), 3442–3452. doi:10.1093/nar/gkm235

Long HK, King HW, Patient RK, Odom DT, Klose RJ (2016) Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Research* **44**(14), 6693–6706. doi:10.1093/nar/gkw258

Lowe R, Barton C, Jenkins CA, Ernst C, Forman O, *et al.* (2018) Ageing-associated DNA methylation dynamics are a molecular readout of lifespan variation among mammalian species. *Genome Biology* **19**, 1. doi:10.1186/s13059-018-1397-1

Mayne B, Berry O, Davies C, Farley J, Jarman S, (2019) A genomic predictor of lifespan in vertebrates. *Scientific Reports* **9**(1), 17866. doi:10.1038/s41598-019-54447-w

Mayne B, Tucker AD, Berry O, Jarman S, (2020) Lifespan estimation in marine turtles using genomic promoter CpG density. *PLoS One* **15**, 7. doi:10.1371/journal.pone.0236888

McLain AT, Faulk C (2018) The evolution of CpG density and lifespan in conserved primate and mammalian promoters. *Aging (Albany NY)* **10**(4), 561–572. doi:10.18632/aging.101413

Musick JA, Limpus CJ (1997) Habitat utilization and migration in juvenile sea turtles. In *The Biology of Sea Turtles, Boca Raton*. pp. 137–163, CRC Press, Costa Rica.

De Paoli-Iseppi R, Deagle BE, McMahon CR, Hindell MA, Dickinson JL, *et al.* (2017) Measuring animal age with DNA methylation: From humans to wild animals. *Frontiers in Genetics* **8**, 106. doi:10.3389/fgene.2017.00106

Platt RN, Vandewege MW, Ray DA (2018) Mammalian transposable elements and their impacts on genome evolution. *Chromosome Research* **26**(1–2), 25–43. doi:10.1007/s10577-017-9570-z

Sharif J, Endo TA, Toyoda T, Koseki H, (2010) Divergence of CpG island promoters: a consequence or cause of evolution? *Development, Growth & Differentiation* **52**(6), 545–554. doi:10.1111/j.1440-169X.2010.01193.x

Sonntag DM, de Boer J, Medvedovic M, Baxter CS, LeMasters G, *et al.* (2004) Mutational biases associated with potential iron-binding DNA motifs in rodent lacI and human p53 mutational databases. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **550**(1–2), 73–88. doi:10.1016/j.mrfmmm.2004.02.004

Copyrighted material

Copyrighted material

Tabak MA, Webb CT, Miller RS (2018) Propagule size and structure, life history, and environmental conditions affect establishment success of an invasive species. *Scientific Reports* **8**(1), 10313. doi:10.1038/s41598-018-28654-w

Tacutu R, Thornton D, Johnson E, Budovsky A, Barardo D, *et al.* (2018) Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research* **46**, 1083–1090. doi:10.1093/nar/gkx1042.

Then AY, Hoenig JM, Hall NG, Hewitt DA, Jardim HeE (2014) Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. *ICES Journal of Marine Science* **72**(1), 82–92. doi:10.1093/icesjms/fsu136

Wagner W (2017) Epigenetic aging clocks in mice and men. *Genome Biology* **18**(1), 107. doi:10.1186/s13059-017-1245-8

Wang Z, Pascual-Anaya J, Zadissa A, Li W, Niimura Y, *et al.* (2013) The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nature Genetics* **45**, 701. doi:10.1038/ng.2615

Zhao Z, Fu YX, Hewett-Emmett D, Boerwinkle E (2003) Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* **312**, 207–213. doi:10.1016/s0378-1119(03)00670-x

Copyrighted material