University of North Carolina Asheville
Journal of Undergraduate Research
Asheville, North Carolina
December 2018

Phylogenetic Comparative Analysis of DNA Methylation Rates in Vertebrates

Darcy A. Davis
Cellular and Molecular Biology
The University of North Carolina Asheville
One University Heights
Asheville, North Carolina 28804 USA

Faculty Advisor: Dr. R. Graham Reynolds

Abstract

In addition to playing a role in genomic function, DNA methylation influences evolution by regulating transcription. Technological advances, such as High-Performance Liquid Chromatography (RP-HPLC), have allowed scientists to explore genomic regulatory changes that contribute to species diversity and phenotypic variability. Epigenetic modifications of notable interest include 5-methylcytosine (5mC) and guanine-cytosine content (GC), as they are related to neutral selection on the cellular level. To understand how these regulatory changes evolve in a phylogenetic context, quantitative traits were analyzed phylogenetically by mapping them to separate mitochondrial phylogenies inferred de novo across 28 reptile species, 26 mammal species, and 42 fish species. Previous studies in vertebrates concluded that there was no significant correlation between DNA methylation and environmental stimuli, but these studies did not correct for the non-independence of evolutionarily related species and thus violated a fundamental statistical assumption. To model 5mC and GC, traits were corrected for phylogeny and phylogenetic comparative analyses were run in RStudio®. First, the extent of the phylogenetic non-independence problem was examined by estimating measures of phylogenetic signal for each quantitative trait. Then regressions were repeated from a previous study, following phylogenetic correction, and inferred correlation between our two epigenetic modifications. Finally, a series of evolutionary models were fit to the phylogeny to examine the evolution of these traits across the phylogenies and selected the best fit model using an AICc model selection procedure. Phylogenetic signal was found in 5mC for both mammal and fish species. In reptile species, phylogenetic signal was found in GC but not 5mC, and that phylogenetic correction did not affect results, likely owing to the relatively small number of tips and the lack of phylogenetic signal in one of the traits. The evolution of these traits is best approximated by an Ornstein-Uhlenbeck model, suggesting that local optima exist for these quantitative characters and predicting a loss of phylogenetic signal (convergence or homoplasy). This study is important because the results can be used to understand the modifications to the genome influencing phenotypic diversity.

1. Introduction

1.1 Epigenetic Modifications In Vertebrates

DNA function is plastic; the most important characteristic of DNA is its ability to change structure and composition (and thus function) over time owing to mutation or changes in gene regulation. Epigenetic modifications introduce heritable alterations that, unlike mutations, do not change the DNA sequence but will induce changes in gene functionality. Of notable interest, DNA methylation, or the addition of a methyl group to a DNA base molecule, is a common epigenetic modification. DNA methylation is required to maintain genome stability and is involved with different cellular functions such as transcription inhibition[1,2], genomic imprinting[3], X chromosome inactivation[4], chromatin stabilization[5], and disease states[6]. Over portions of the genome, repetitive elements in the genetic sequence are silenced by DNA methylation[7]. These repetitive sequences remain inactive to prevent interference with transcription and maintain genome stability. Gene-rich regions of the genome are highly methylated, while

transposons are largely unmethylated. DNA methylation mediating transposon control in eukaryotes shows a direct relationship between gene methylation and transcription[1]. Global chemical modifications to cytosine bases are associated with long-term transcriptional repression or activation. DNA methylation makes the gene less available for transcription. Global chemical modifications to cytosine bases are associated with long-term transcriptional repression or activation. In regions of the gene with high GC content, phosphates separate cytosines and guanines (CpG). DNA methyltransferases add a methyl group to the cytosine. CpG methylation has an effect on transcription factor (TF) binding. Transcription factors are often located in the region of the 5' extremity of CpG islands. In a study using massively parallel sequencing to probe the sensitivity of transcription factor binding to DNA modifications in vitro, researchers concluded cytosine methylation within the protein-DNA interface increases binding affinity of TFs and mechanisms of epigenetic control of mRNA production[2].

1.2 DNA Methylation And Vertebrate Evolution

Mechanisms that regulate the expression of genes have accounted for selection among species[8]. Following events of compositional change, gene regulation acts as a mediating factor dictating trait evolution. Reptilian genomes show the compositional pattern of ectothermic vertebrates and do not show CpG islands; while the genome size and the methylation levels are more similar to those of endothermic vertebrate genomes than to those of ectothermic vertebrate genomes[9]. CpGs represent the composition of the genome that are constituted by repetitive cytosine and guanine nucleotides. These regions are frequently methylated in the genome. Regions known as CpG islands show elevated CpG content. These regions are typically associated with the promoter or regulatory regions of the gene[10]. For instance, the 5'-flanking region of the *Pax6* gene, responsible for the development of eyes, is associated with two CpG islands[11]. Changes to the methylation of the *Pax6* gene in olive ridley sea turtles (*Lepidochelys olivacea*) results in aberrant phenotypes. The results suggest and interplay between genetic and epigenetic mechanisms.

Chromosomal variation in reptiles is attributed to the presence of microchromosomes[12]. A microchromosome is a characteristically small and cytogenetically indistinguishable component of the karyotype. Microchromosomes are found in birds, reptiles, and fish, but are absent in mammals[13]. Microchromosomes are GC rich, contain higher frequencies of CpG dimers, and lack repetitive elements[14]. In mammalian genomes that lack characteristic components of microchromosomes, epigenetic drift has been attributed to the global decrease in DNA methylation. Drift, as a result of epigenome maintenance, has been constrained to CpGs across the human genome[15]. A similar divergence of the epigenome associated with age has been observed in American alligators (*Alligator missippiensis*); in which differences in the epigenome are the result of external factors. As with other vertebrates, global DNA methylation declines with age in alligators[16]. The relationship to age is further compounded by the correlation between long-term mercury exposure with DNA methylation. A negative relationship exists between DNA methylation, age, and mercury exposure[16]. Mechanisms that regulate the epigenome are observed across families of vertebrates indicating a retention of function in accordance with a divergence of character trait relatedness. Methylation of the genome is also conserved across vertebrates.

Cytosine methylation of the CpG dinucleotide in reptiles is carried out by three types of DNA methyltransferase (DNMT) enzymes. DNMT3a and DNMT3b methylate DNA de novo, DNMT1 is a maintenance methyltransferase. DNMT1 preferentially methylates hemimethylated DNA[17]. The activity of DMT1 declines with age. The decline is thought to contribute to "drift" in cells[18]. DMRT1 is up-regulated in males during embryonic development. DMRT1 is up-regulated during the thermosensitive period of sex determination of *Trachemys scripta*, and up-regulation influences sex determination during TSD in this species[19]. DNMT3a has no related methyltransferase activity. DNMT3l, a related protein, is essential for imprinting genes in vertebrates[20]. Imprinting has evolved in placental mammals and comparisons of gamete-specific methylation of DNMT3 in vertebrates reveals a link between the existence of DNMT3l and the evolution of imprinting. DNMT enzymes have been speculated to have prokaryotic origins owing to their conservation throughout evolutionary time. Homologs of DNMT enzymes have been identified in fish, birds, and plants[21].

Inheritance of epigenetic markers leave a physical mark (e.g., a methyl group) on DNA as well as altered gene expression. A growing body of literature has investigated DNA methylation in the context of vertebrate evolution[9,22,24,25]. Specifically, the relation between epigenetic modifications and transcription inactivation is of interest because of the relation between gene expression and phenotype. Patterns of methylation (and thus gene regulation) might accompany evolutionary divergence; and such heritable functional changes are speculated to serve as a force driving evolutionary divergence[25]. Therefore, current studies regarding the epigenetic regulation of reproduction will be fundamental to our understanding of expression profiles that are heritable and able to be passed to subsequent generations.

DNA methylation expression patterns are variable between cell types and biological sex. The difference in cellular and molecular components of tissues can contribute to phenotype variations that will persist throughout the lifespan of an offspring suggesting a change in gene transcription. Modifications to promoter regions may play a role in gene regulation. For example, in zebrafish, DNA methylation was found to be associated with down regulation of transcription of genes in the gonads and livers[26]. Genes regulating epigenetic processes were over-expressed in female reproductive tissue. In females, *esr1* was over-expressed in the liver. The promoter of this gene was hypo-methylated[26]. Induced changes to tissue structure introduce variability to the gene transcript by environmental influence[27]. DNA methylation influences cell differentiation and the differentiation of tissues. For example, germline-specific high-CpG-density promoters (HCPs) are hypermethylated in brain tissue, while most HCPs are unmethylated in embryonic stem cells[28].

1.3 Phylogenetic Comparative Methods

Phylogenetic comparative methods (PCMs) enable researchers to analyze correlations between traits in an explicitly evolutionary context. Correlations can be drawn between characters or within a character-by-environment interaction[29]. Importantly, these methods can be used to detect selection in character trait evolution[30]. Recent phylogenetic analyses have focused on how to map character traits onto a phylogenetic tree while accounting for uncertainty in character change[31]. Phylogenetic comparative methods are used to compare traits across species to test hypotheses about trait and evolutionary history. These methods show evolutionary relatedness and work backwards around the problem of phylogenetic uncertainty to answer questions about evolutionary processes driving trait changes. Inferences about patterns of evolution are modeled using a continuous-time Markov-chain that considers all possible character histories. The probability of the instantaneous character state depends only on the current character state transition matrix. This matrix is used to describe these transition rates among the characters. Rates of change are in-turn dependent on the evolutionary model specified[32]. A likelihood function is then used to analyze the fit of the model to the data. The maximum-likelihood (ML) solution suggests the best-fit model of evolution to the data. Further, branch length is incorporated and is used to determine the best-fit model based on evolutionary time (or distance).

A stochastic model of character evolution addresses the problem of uncertainty of phylogenetic trees most accurately when character history is corrected[33]. Stochastic character mapping (SCM) uses a Bayesian Markov-chain Monte Carlo (MCMC) approach to infer rate of change of the character trait across the phylogeny. Likely changes along branches are simulated based on draws from a prior distribution; and multiple changes can be modeled along a branch. Methods described for stochastic character mapping of molecular character history have been applied to address trait uncertainty[29,34]. In the previous study, uncertainty in morphological character history was represented using such a Bayesian method, which accounts for uncertainty in the phylogenetic hypothesis.

Models can be used to study the evolutionary mechanisms by simulating species traits over a phylogeny, and thus explicitly over the evolutionary history of the group of interest. Biological evolution is not constant; rates of evolution vary conditionally in accordance with exposure over an observed period of time. Phylogenetic data stimulation (PDS) can be used in extension to PCM to answer biological questions about relatedness. Simulations are used to make theoretical predictions— PDS methods generate a random number on the phylogeny given an *a priori* model of trait evolution. Common models used for continuous traits include: Brownian motion (BM), which describes a stochastic process of trait evolution, and Ornstein-Uhlenbeck (OU), which describes the existence of optimal trait value(s) towards which traits are evolving. However, parameter values for traits can be dependent on time, which might introduce complication[35]. This method can be extended to model several traits evolving simultaneously on the phylogeny.

Phylogenetic signal is the tendency for related species to resemble each other in trait values[23], meaning that evolved traits are non-independent among lineages with varying evolutionary relatedness. Different measures exist to quantify phylogenetic signal— common indices include Moran's I[37], Aboufeif's C_{mean}[38], Pagel's λ[31], and Blomberg's K[39]. Pagel's λ is the most reliable for continuous trait values that follow a Brownian motion (BM) model of evolution. Pagel's λ has been implemented to measure phylogenetic dependence of observed traits[31]. The coefficient λ accounts for the weight of phylogenetic influence and fits the trait data to the model of evolution[36].

2. Methods

2.1 Phylogeny Inference

Published global cytosine DNA methylation percentages for individual vertebrate taxonomic groups— non-avian reptiles, mammals, and actinopterygian fish species[25,40,41] (Table 1)— were used to model on phylogenies constructed using mitochondrial DNA. Mitochondrially encoded NADH dehydrogenase 2 sequence accessions for each species represented in the methylation dataset were collected from GenBank® using custom R scripts and the APE package in R v3.5.0 running in RStudio®[42]. Selection of the mtDNA sequences was based on the representation in the methylation database and the relative sequence length of the accession (*i.e.*, short sequences were excluded). Mitochondrial sequence data and DNA methylation rates were aligned separately for each taxonomic group (reptiles, mammals, actinopterygian fishes) in a matrix using the ClustalW 2.3 algorithm in the program Geneious 10.5.1 (Biomatters, Auckland, NZ)[43]. A phylogeny was inferred for each vertebrate group using a maximum-likelihood method implemented in the RaxML algorithm plugin in Geneious® v10.4[44]. For each of the three alignments, a GTRGAMMA model and rapid bootstrapping algorithm was used with 1,000 bootstrap replicates followed by a thorough maximum-likelihood search option with 100 independent searches. The resulting tree with support values was then exported as a nexus text file to import into the R environment.

Table 1. Global cytosine DNA methylation rates and mitochondrial sequence data.

Class	Species	GC %	5mC %	Mt-DNA accession
Crocodilia	*Alligator mississippiensis*	48.56	0.96	jf315622
	Crocodylus niloticus	48.44	0.85	dq273697
Testudines	*Caretta caretta*	46.74	0.96	fr694649
	Testudo graeca	45.7	0.77	dq080049
	Trachemys scripta elegans	46.98	1.14	km216748
	Chelydra serpentinae	47.68	1.33	ef122793
	Macrochelys temminckii	48.98	1.11	ef071948
Squamata	*Chlamydosaurus kingii*	44.67	0.93	hq684213
	Furcifer oustaletia	44.49	1.14	af448769
	Python molurus molurus	43.18	0.81	hm581978
	Boa constrictor	41.95	0.68	ab177354
	Walterinnesia aegyptia	42.77	0.8	ay059001
	Natrix tessellata	44.12	1.03	ay870642
	Pantherophis guttatus	41.83	1.29	dq902218
	Hierophis viridiflavus	44.16	1.32	ay487018
	Zamenis lineatus	43.9	1.4	dq902251
	Euprepiophis mandarinus	43.05	1.23	dq902222
	Bothrops jararaca	42.99	1.16	ku194299
	Vipera aspis aspis	43.4	1.2	am944744
	Podarcis muralis	48.18	1.34	ay234145
	Podarcis siculus	47.02	1.47	fj460598
	Gekko gecko	46.05	1.09	jx170698
	Tarentola mauritanica	46.59	0.94	jx041447
	Anguis fragilis	47.6	1.01	fj666559
	Iguana iguana	44.33	1.36	aj278511
	Sceloporus magistera	45.84	0.85	af528741
	Tupinambis teguixin	45.6	0.86	jn700173
	Chelonia mydas	47.38	1	ab012104
Placentals	*Rattus norvegicus*	43.9	0.9	eu104718
	Sciurus vulgaris	39.5	0.6	ku962990
	Homo sapiens	42.8	0.7	dq473645
	Hapalemur griseus	41.4	0.9	kc757397

	Galeopterus variegatus	40.6	0.9	aj428849
	Oryctolagus cuniculus	44.3	0.9	aj001588
	Procavia capensis	41	0.7	ab096865
	Balaenoptera physalus	41.3	0.9	kc572860
	Physeter catodon	41.9	1.1	ku891394
	Sus scrofa	44.6	0.9	kj782448
	Equus caballus	42.8	1	ku575247
	Hipposideros galeritus	41.4	0.9	ay504532
	Crocidura russula	41.4	0.7	ay769264
	Noctilio albiventris	43.3	0.6	ay504576
	Myotis lucifugus	43.5	1	ay504565
	Nycteris hispida	42.9	0.9	ay504544
	Canis lupus familiaris	41.1	0.7	ay729880
	Panthera uncia	41.5	0.9	kp202269
	Erinaceus europaeus	45.5	0.5	af513818
	Didelphis virginiana	39.2	0.3	z29573
Monotremes	*Ornithorhynchus anatinus*	48.5	1.2	x83427
	Tachyglossus aculeatus	48.9	1	aj303116
Marsupials	*Macropus rufus*	41.7	0.4	jn967007
	Macropus robustus	41.2	0.4	y10524
	Vombatus ursinus	40.9	0.3	af343893
	Monodelphis domestica	39.1	0.3	aj508398
Actinopterygians	*Jordanella floridae*	41.43	1.1	ay902108
	Ophiodon elongatus	44.31	1.78	ay225719
	Scorpaena guttata	41.27	1.48	jq088494
	Notopterus notopterus	44.96	1.22	ap008925
	Pantodon buchholzi	45.66	1.77	ab035229
	Sardina pilchardus	47.12	1.4	ap009233
	Danio rerio	39.19	1.35	km244705
	Carassius auratus auratus	39.53	1.44	jx183457
	Oncorhynchus keta	45.75	1.49	ap010773
	Merluccius merluccius	48.69	2.18	fr751402
	Gadus morhua	48.61	2.37	hg514359
	Arctogadus glacialis	48.13	2.74	am919429
	Boreogadus saida	48.48	2.22	dq356936
	Mullus barbatus	48.86	2.19	aj491821
	Capros aper	46.69	1.87	ap009159
	Aphyolebias peruensis	45.7	1.5	af092407
	Holacanthus passer	44.1	1.43	kp965872
	Aphanius fasciatus	43.17	1.62	af449313
	Xiphophorus maculatus	41.28	1.39	ef017600
	Fundulus heteroclitus	42.8	1.76	kj878751
	Cottoperca gobio	43.65	1.98	jn186884
	Bovichtus diacanthus	41.95	1.83	kf412875
	Pseudochaenichthys georgianus	44.9	2.22	hm165672
	Neopagetopsis ionah	43.8	2.08	hm165754
	Chaenocephalus aceratus	44.27	2.1	hm166185
	Chionodraco rastrospinosus	43.65	2.28	hm165958
	Chionodraco hamatus	43.9	2.4	hq170102
	Champsocephalus esox	45.54	2.22	hq170096
	Notothenia rossii	44.52	1.78	ay256567
	Notothenia coriiceps	44.4	1.87	fj647714
	Dissostichus mawsoni	44.09	1.85	dq184498
	Trematomus bernacchii	43.59	1.82	fj647717
	Trematomus newnesi	44.57	1.82	dq184506
	Lepidonotothen nudifrons	43.74	1.79	kp745380

	Patagonotothen guntheri	44.08	1.69	kf412892
	Lepidonotothen squamifrons	43.79	1.87	kp745376
	Lepidonotothen kempi	43.31	1.64	kf412886
	Gobionotothen marionensis	44.32	1.92	kp745410

2.2 Phylogenetic Comparative Analyses

2.2.1 stochastic character mapping

R v3.5.0 in Rstudio® was used to address phylogenetic non-independence of constructed phylogenies. Two continuous character traits, 5mC: the methylated form of the DNA based cytosine and GC: the percentage of nitrogenous based that are either guanine or cytosine, were stochastically mapped to phylogenetic trees for each class of vertebrates using the *make.simmap* function in the R package PHYTOOLS[45]. Each tree contained a fixed number of species (reptiles: n = 28; mammals: n = 26; fishes: n = 42). A continuous-time Markov chain was used to the model evolutionary history of the character change[29]. Possible character histories for each trait were sampled such that specific history varied with the distribution of the posterior probabilities. Joint reconstruction of continuous character data across nodes on a matrix were plotted from their joint posterior probability distribution[29]. Changes along edges of the tree were simulated using rejection. Wait times for the changes between states were drawn on exponential distribution with rate. Further changes were stimulated if waiting time was shorter than total branch length. Successful stimulation of stochastic history was achieved when branch nodes matched. Stochastic character maps were then plotted on the phylogenetic tree to visualize the different character state probabilities through time using the *plotSimmap* function in phytools[45,46]. Ancestral states were estimated for internal nodes using ML then interpolated along the branches of the tree[47]. A continuous color gradient was applied to the trait maps to visualize character state-change through time.

2.2.2 phylogenetic signal

To ascertain whether a phylogenetic correction was necessary for subsequent analyses, a series of methods were used to test for autocorrelation of phylogenetic distance and character traits. Correlograms were constructed for each trait (GC, 5mC) for each taxonomic group (reptiles, mammals, fishes) as a method for assessing lag distance. The distance between trait values, or lag, represents a correlation between plot points. The PHYLOSIGNAL package in R was then used to analyze measures of phylogenetic signal in a simulated phylogenetic context using actual continuous trait data (5mC and GC)[49]. Global measures of autocorrelation were used to indicate the presence of phylogenetic signal in the simulated phylogenies. Phylogenetic signal indices Moran's I[37], Aboufeif's Cmean[38], Pagel's λ[31], and Blomberg's K[39] were evaluated based on spatial autocorrelation within the context of the phylogenetic trees to establish an informative measure of phylogenetic signal. Signal indices were based on phylogenetic independent contrasts assuming Brownian motion. Moran's I describes the relationship of a trait variation to the phylogeny[37], Aboufeif's Cmean describes independence among traits based on closely related species[38], Pagel's λ measures phylogenetic dependence of traits[31], and Blomberg's K represents phylogenetic signal strength as a ratio[39]. Statistical non-independence was assumed for 5mC and GC traits when determining phylogenetic relatedness. Based on the results of this simulation study, Pagel's λ was selected as the autocorrelation index because of experimental small type I error across all sizes of phylogenies for test of phylogenetic signal[36]. The index assumes a Brownian motion (BM) model of evolution[31]. A value of zero indicates phylogenetic independence; values of one indicated distribution under BM and phylogenetic signal. The R package GEIGER was used to estimate Pagel's λ and the function *fitContinuous* was then used to correct for phylogenetic autocorrelation[49]. Character traits were individually tested for phylogenetic signal in each phylogeny using Pagel's λ, resulting in nine total tests of signal (2 trait sets for each of 3 phylogenies).

2.2.3 phylogenetic regression

Two different regression models were fit for the relationship between 5mC and GC between each class to check the type I error. Regression models were applied in which permutations of the values of **y** were based on a single predictor[50]. Regression model I assumed a fixed value of the predictor. Parametric tests of significance in this model are through the origin. Pearson's product-moment correlation coefficient (r) was used to analyze the strength of the association between variables. The sample population was statistically independent for the application of this

correlation. Regression model II, also called the phylogenetically-corrected regression, assumes the predictor and response vary jointly following a multivariate BM process[47]. Traits were not log-transformed prior to the regression. However, standardized phylogenetic independent contrast for both traits was computed. Variance of standard was computed as an index to evaluated the fit of the tree to the data. Contrast was compared with the values obtained after the data was permutated randomly across the tips of the tree. Permutations are independent of relationship to the phylogeny[39]. Residuals are distributed normally with covariances proportional to branch length[51]. A least-square linear regression through the origin was used for the contrast.

2.2.4 evolutionary models

Evolutionary models of continuous character evolution were fit to our phylogeny to test the following evolutionary models of speciation and trait evolution: Brownian motion (BM)[52], white-noise (White)[53], Early-burst (EB)[54], and Ornstein-Uhlenbeck (OU)[55]. Models for continuous character data include Brownian Motion (BM) and Ornstein-Uhlenbeck (OU). The Brownian motion model has been described as a "random walk." Trait evolution occurs over a contingency. In the Ornstein-Uhlenbeck model, trait evolution is attached towards a central, optimum value. The Early Burst model is applicable for studies of adaptive radiation. In this model, trait evolution is rapid during early stages and subsequently slows down[54]. However, few studies have revealed an Early Burst model of evolution[56]. Models for estimates of divergence time include white-noise and CIR. Both models function in the Bayesian framework. The white-noise model is non-autocorrelative and an alternative form of relaxed clock models[53]. Akaike information criterion was used to select the model that was the most efficient approximation. The best fit model was selected using AICc optimized under the function *fitContinuous* in the R package GEIGER[49].

2.3 Statistical Analyses

R v3.4.1 in Rstudio® was used to evaluate statistical non-independence between continuous traits, 5mC and GC, and habitat parameters. Global cytosine DNA methylation percentages for fish species were analyzed at one factor level for habitat based on classification in FishBase®. Habitat levels included polar, temperature, subtropical, and tropical. Water temperature was assumed as the mean body temperature for each species. Ectotherms are acclimated to water temperature. In teleost fish, species can inhabit a wide range of different temperatures[57]. Differences in temperature affects metabolism; transcriptional regulation has been implicated in relation to metabolism in mammals[58]. Understanding regulatory factors, such as temperature, that influence gene expression is of interest.

3. Results

3.1 Variation In DNA Methylation Across Phylogeny

The continuous characters GC and 5mC were mapped to the phylogenies using stochastic character mapping (Fig. 2). DNA methylation rates and mtDNA sequences were analyzed for 28 reptile species; two representatives from order Crocodilia, five representatives from order Testudines, and twenty-one representatives from order Squamata after eliminating MT-ND2 sequences that were not associated with the entire genome. Through the phylogenetic correction of non-independence, the trait map indicated order Crocodilia was more GC rich, while Serpentes was more GC poor compared to other reptiles. Distribution of 5mC was clade-dependent (Fig. 2a). For the mammalian class, 26 species were analyzed; two representatives from order Monotremata, four representatives from order Marsupialia, and twenty representatives from order Placentalia were present in the phylogeny. Order Monotremata was more GC and 5mC poor compared to other mammals. Two of the species of marsupials (*Monodelphis domestica* and *Vombatus ursinus*) were more GC and 5mC rich than other species in the order. Order Monotremata was more GC and 5mC poor compared to other mammals (Fig. 2b). Forty-two representative fish species were present in the phylogeny. Order Cypriniformes was more GC poor compared to other fishes. Order Gadiformes was more GC and 5mC rich (Fig. 2c).

To investigate the relationship between DNA methylation rates and phylogeny, autocorrelational analyses were performed on the data. In reptiles, our autocorrelation analyses and Pagel's lambda suggest phylogenetic signal was present in the GC character trait and absent in 5mC (GC: λ= 1.00, P= 0.00100, logL= -48.8; 5mC: λ= 0.235, P= 0.181, logL= 5.65; Table 2). Subsequence autocorrelational analyses on mammals and fishes indicated phylogenetic signal was present in the 5mC character trait; mammals: (GC: λ= 0.375, P= 0.320, logL= -60.5; 5mC: λ= 0.746, P= 0.00160,

logL= -1.70; Table 2), fishes: (GC: λ= 6.61E-05, P= 1.00, logL= -96.6; 5mC: λ= 0.755, P= 0.0858, logL= -17.8; Table 2).

Table 2. Autocorrelation analyses of continuous character trait data using Pagel's lambda.

Class	Data	Lambda	P	log-likelihood
Reptilian	5mC	0.235	0.181	5.65
	GC	1.00	0.00100	-48.8
Mammalian	5mC	0.746	0.00160	-1.70
	GC	0.375	0.320	-60.5
Fishes	5mC	0.755	0.0858	-17.8
	GC	6.61E-05	1.00	-96.6

Autocorrelation was further explored through correlograms (Fig. 1,3). The correlogram of GC presents a strong positive autocorrection for short lags and negative autocorrelation for medium lags in reptiles (Fig. 1a). Autocorrelation analyses with Pagel's lambda suggested phylogenetic signal in the GC character trait for only the reptile family. In this family, closely related species are highly correlated for the trait. In mammals, the phylogenetic correlogram is relatively flat (Fig. 1b). The sample is random and nonsignificant. A positive autocorrelation for long lags is seen in fishes (Fig. 1c).

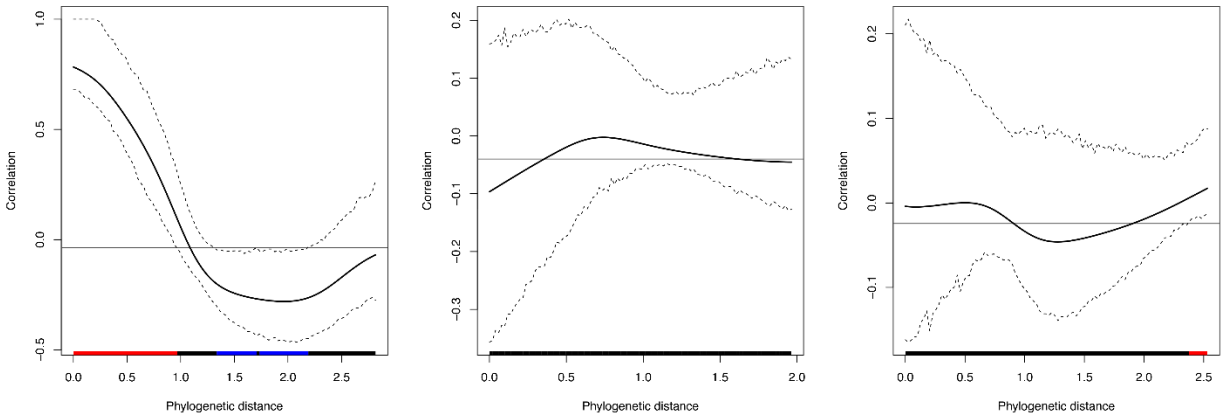

Figure 1. Phylogenetic correlogram for GC.

Figure 1 a) correlogram for 28 non-avian reptile species. The solid gray line indicates the expected value under the null hypothesis of no autocorrelation. 95% confidence intervals are represented as dashed lines. The red line at the bottom indicates significant autocorrelation at a given phylogenetic distance, b) mammals, c) fishes.

The correlogram of 5mC exhibits a positive autocorrelation for short lags in reptiles (Fig. 3a). The autocorrelation analyses suggest phylogenetic signal is present in the 5mC character trait for mammals and fishes. In mammals, there is nonsignificant autocorrelation (Fig. 3b). Positive autocorrelation is observed for short and medium lags in fishes (Fig. 3c).

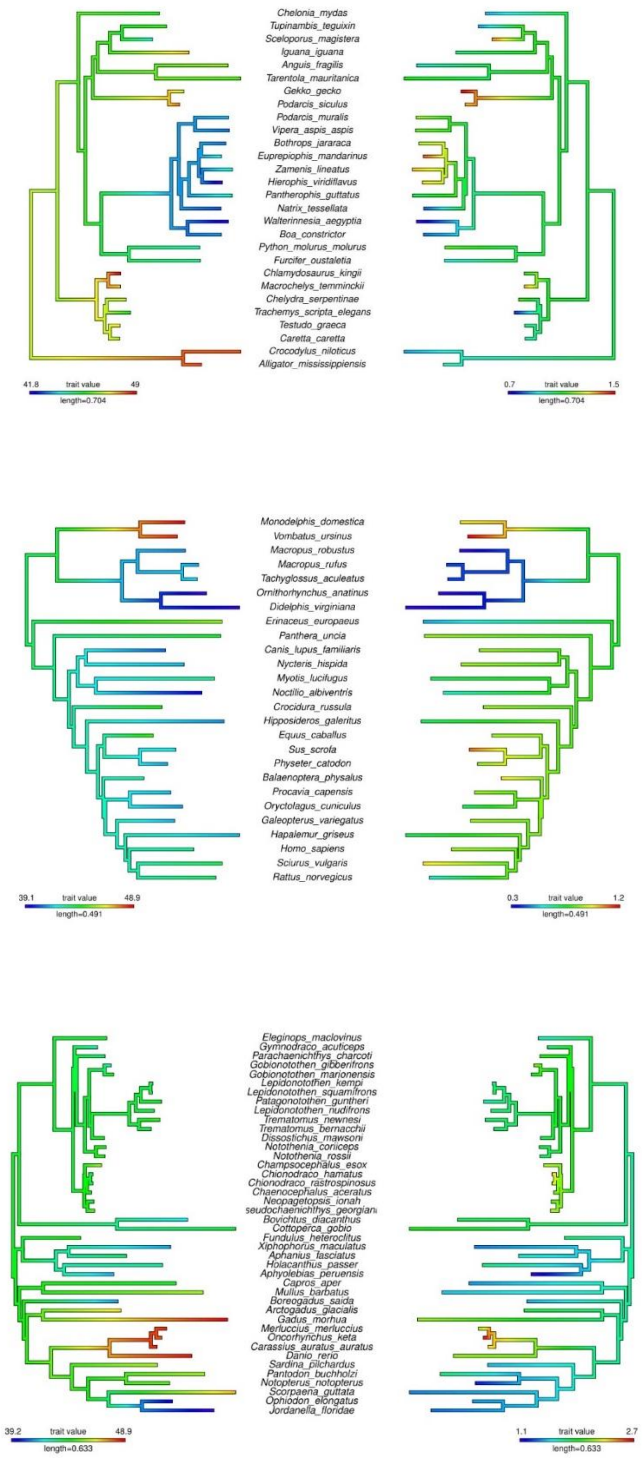

Figure 2. Phylogenetic comparison of continuous character evolution.

Figure 2 a) continuous character evolution for GC (left) and 5mC (right) across major groups of non-avian reptiles. Note the position of *Chelonia mydas*, there was low support for this topological placement owing to the quality of the sequence data, b) mammals, c) fishes.

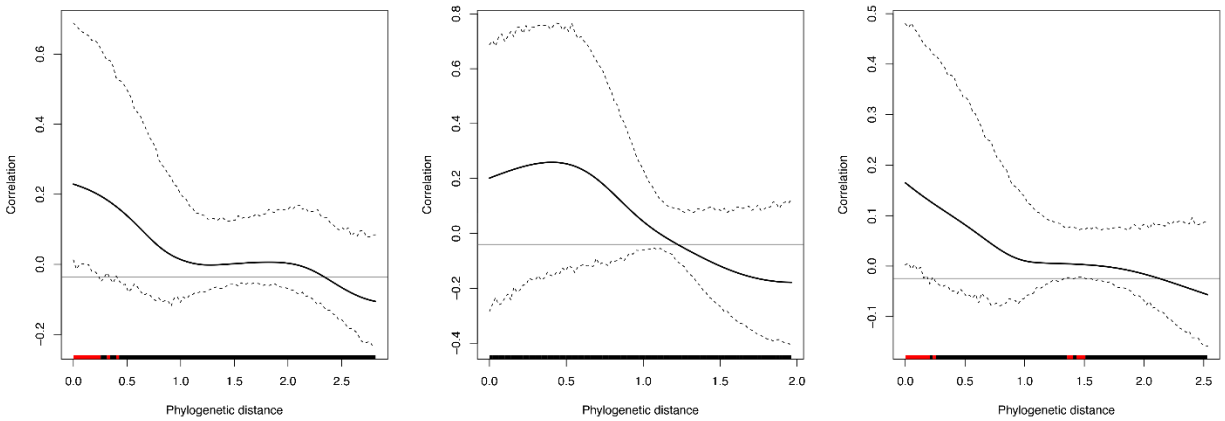

Figure 3. Phylogenetic correlogram for 5mC.

Figure 3 a) correlogram for 28 non-avian reptile species. The solid gray line indicates the expected value under the null hypothesis of no autocorrelation. 95% confidence intervals are represented as dashed lines. The red line at the bottom indicates significant autocorrelation at a given phylogenetic distance, b) mammals, c) fishes.

To determine the representative phylogenetic signal index, traits were stimulated under increasing values of Brownian motion. At complete randomness through the contingency, lambda was the most efficient method for measuring phylogenetic signal under increasing values of Brownian motion in reptiles (Fig. 4a). The strength of the lambda index in the mammal and fish model was not as strong (Fig. 4b, 4c).

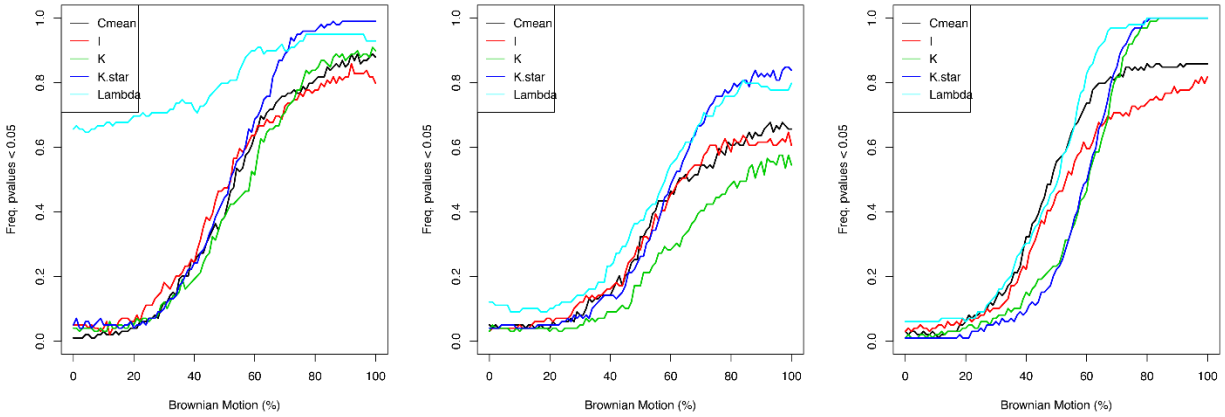

Figure 4. Phylogenetic signal test.

Figure 4 a) response to stimulate trait values under increasing values of Brownian motion on the reptile phylogeny. Pagel's lambda clearly shows the most power to resolve phylogenetic signal for the phylogenies, b) simulated trait values on mammalian phylogeny, c) fish phylogeny.

To reaffirm results of previous correlational studies between GC and 5mC rates in reptiles, linear regressions between each family were conducted. Previous studies concluded there was no significant correlation between DNA methylation in reptiles[41]. Reptiles show variability in DNA methylation rates when corrected for phylogeny (regression1: R^2=0.00, df =26, P=0.87, regression2: R^2=0.008, df=26, P=0.65; Fig. 5a, 6a). Subsequent analyses in mammals indicated the relationship between GC and 5mC in vertebrates was a positive, linear correlation[25]. There was a significant relationship between GC and 5mC rates in mammals for linear and phylogenetically corrected regressions (regression1: R^2=0.2939, df =24, P=0.004226, regression2: R^2=0.2228, df=22, P=0.01492; Fig. 5b, 6b).

The same paper presented a positive, linear correlation between GC and 5mC rates in fishes[25]. There was a significant relationship between GC and 5mC rates in fishes for linear and phylogenetically corrected regressions (regression1: R^2=0.2752, df =40, P=0.0003615, regression2: R^2=0.08692, df =40, P=0.05804; Fig. 5c, 6c). Methylation may linearly increase with GC content; however, phylogenetic factors may also influence trait selection subject to exploration.

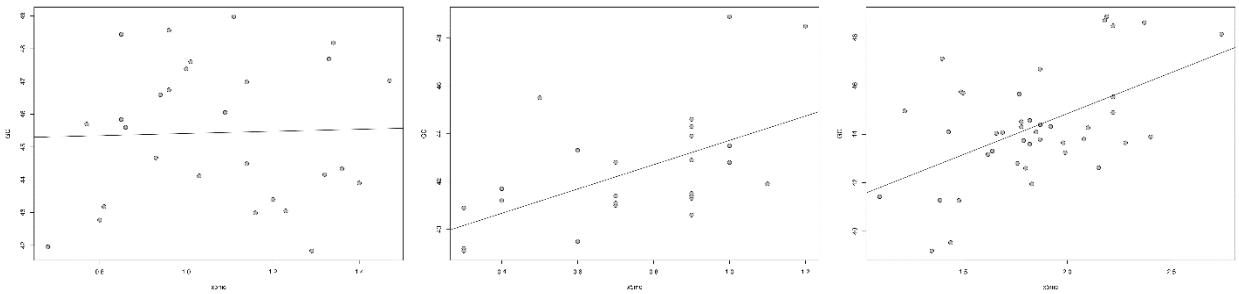

Figure 5. Linear model.

Figure 5 a) linear relationship between 5mC and GC in reptiles (R^2=0.001, df =26, P=0.87), showing no correlation, b) linear relationship between 5mC and GC in mammals (R^2=0.2939, df =24, P=0.004226), c) linear relationship between 5mC and GC in fishes (R^2=0.2752, df =40, P=0.0003615).

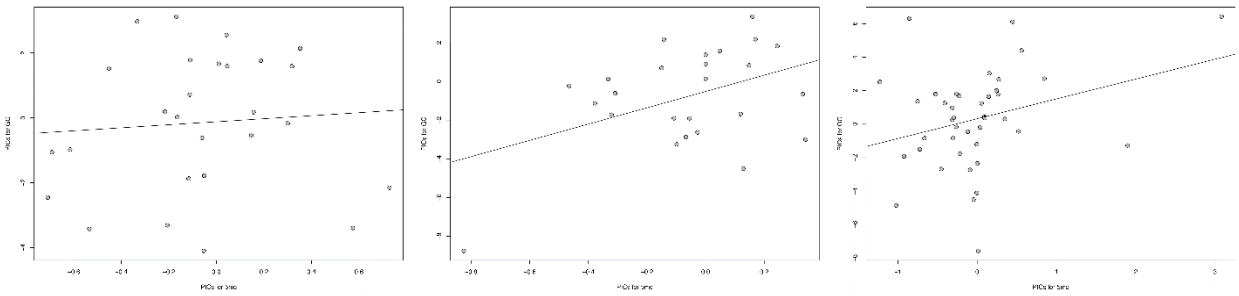

Figure 6. Phylogenetic correction.

Figure 6 a) phylogenetically-correlated regression for the relationship between 5mC and GC in reptiles (R^2=0.008, df=26, P=0.60), the statistically appropriate approach, also showing no relation, b) phylogenetically-correlated regression for the relationship between 5mC and GC in mammals (R^2=0.2228, df=22, P=0.01492), c) phylogenetically-correlated regression for the relationship between 5mC and GC in fishes (R^2=0.08692, df =40, P=0.05804).

The model of evolution that best predicts the trait selection for the reptile sample with AIC_C was OU (L = 5.54, AIC_C = -4.01, Δ_i= 0.00, w_i = 0.73; Table 3). Mammals are also represented under OU (L=7.45, AIC_C = -7.81, Δ_i= 0.00, w_i = 3.90E-01; Table 3). Fishes are best represented by BM (L=-2.34, AIC_C = 11.32, Δ_i= 0.00, w_i = 9.99E-01; Table 3). Families best represented by the OU model of trait evolution are attracted towards a central, optimum value. Traits occur over a contingency in the fishes that are best represented by the BM model of trait evolution.

Table 3. Best-fit evolutionary models for the continuous character data (# iterations = 50).

Class	Model	log-likelihood	AIC	AIC_c	Delta AIC_c (Δ_i)	Akaike weight (w_i)
Reptilian	BM	3.67	-1.35	-0.349	3.73	0.113
	White	3.67	-1.35	-0.349	3.73	0.113
	EB	3.67	0.651	2.39	5.73	0.0416
	OU	5.54	-5.08	-4.08	0.00	0.732
Mammalian	BM	7.45	-8.90	-7.81	0.0000120	3.90E-01
	White	-1.32	8.63	9.73	17.5	6.07E-05
	EB	7.87	-7.75	-5.84	1.16	2.19E-01
	OU	7.45	-8.90	-7.81	0.00	3.90E-01
Fishes	BM	-2.34	10.7	11.3	0.00	9.99E-01
	White	-14.8	35.6	36.2	24.9	3.95E-06
	EB	-14.8	37.6	38.6	26.9	1.45E-06
	OU	-8.87	23.7	24.4	13.1	1.46E-03

A one-way ANOVA between temperature conditions was conducted to compare the effect of DNA methylation rates on habitat parameters in fishes for GC and 5mC. There were no statistically significant differences between group means as determined by one-way ANOVA for GC ($F=0.865$, $df=3$, $P=0.468$). There was a significant effect of 5mC on habitat at the $p<0.05$ level for the four conditions ($F=8.479$, $df=3$, $P=0.000194$).

4. Discussion

Among reptiles, there was dependence of DNA methylation rates as a result of a phylogenetic relationship in GC, but not in 5mC. This indicated that closely relates species share comparable GC content in their genomes as a result of evolutionary similarity. Patterns of similarity are best analyzed through quantitative, phylogenetic analysis. The reptilian model has been imposed as a historical sequence for the evolution of quantitative traits, including squamate viviparity and matrotrophy. Viviparity, derived from independent origins, is more similar between squamate clades than that of mammals and fishes[59]. A modified scenario for the evolution of DNA methylation is suggested to appear in conjunction with reptile speciation. The rate of chromosome change was been correlated to the number of living reptile species[12]. DNA methylation has been proposed as an epigenetic mechanism used to preserve chromosome stability[60]. DNA methylation is thus a mediating factor dictating the propensity of the chromosomes to acquire variation.

Monodelphis domestica and *Vombatus ursinus* are represented in a separate clade from the other marsupials in the phylogeny. Previous research has indicated monotremes are more GC rich and 5mC rich in comparison to other mammals[25].The results indicate monotremes are more GC poor and 5mC poor than other mammals.

Fishes in the same genus share similar DNA methylation rates regardless of body temperature. Species of genera *Lepidonotothen* have different habitat parameters. *Lepidonotothen nudifrons* is a polar fish, while *Lepidonotothen squamifrons* and *Lepidonotothen kempi* are temperature fish according to FishBase®. Phylogenetic proximity was more important than body temperature when determining character trait relatedness. A previous study concluded fishes that belong to the same genus and share the same body temperature showed similar GC and 5mC levels[40]. Their analyses did not account for a fundamental statistical assumption that would have corrected for phylogenetic non-independence.

While phylogenetic signal was exclusively indicated for the GC character trait in reptiles, there was autocorrelation of both traits at low phylogenetic distance. With increasing phylogenetic distance, phylogenetic signal of traits decreases. At high phylogenetic distance, we see no autocorrelation. These findings indicate closely related species are highly correlated. The phylogenetic relationships between these species thus indicate statistical non-independence of the trait values. There was nonsignificant autocorrelation of both traits in mammals across all phylogenetic distance. In fishes, there was autocorrelation for GC at high phylogenetic distance; however, phylogenetic signal was not exhibited by this trait in successive tests.

The phylogenetic signal that is strongly exhibited by the GC trait value in reptiles and 5mC trait value in mammals and fishes is most effectively measured using lambda values. Lambda recovers phylogenetic signal for the trait value at low levels of phylogenetic signal imparted under a Brownian motion simulation of trait evolution.

Previous studies have determined no significant correlation exists between DNA methylation rates; however, the results were statistically invalid because they did not correct for phylogeny[25,41]. In reptiles, no statistically significant correlation was identified when corrected for phylogeny, thus indicating a variability in DNA methylation rates between reptiles. However, through events of speciation, there are instances of diversification that pull groups towards a central optimal value (the Ornstein-Uhlenbeck Model). The optimal values for each lineage appear to co-evolve independently. There is indication that the optimum value for Crocodilians is a tendency towards low 5mC and high GC, while Testudines exhibit a weaker autocorrelation of GC richness. Squamates, specifically Serpentes, evolved towards high 5mC and low GC.

In mammals and fishes, a statistically significant correlation was identified for the linear regression and when corrected for phylogeny, thus indicating a correlating factor between these rates. However, identification of phylogenetic signal indicates that speciation impacts events of diversification in these families. Mammals show a pull toward a central optimum value. There is indication that the optimum value for Monotremes is a tendency towards low 5mC and GC.

Habitat conditions influence global cytosine methylation in fishes. GC content is not supported by habitat selection. Previous studies have indicated a connection between DNA methylation and environment[25]. However, the phylogeny influences the value of these traits. There was phylogenetic signal in the 5mC character trait in fishes exclusively. This dimensionality suggests DNA methylation does not function in isolation of environmental or phylogenetic factors.

5. Acknowledgements

The author wishes to express their appreciation to faculty for their support, encouragement, and constructive criticism which has contributed to both context and presentation of the article.

6. References

1. Bird, A. 2002. DNA methylation patterns and epigenetic memory. Genes & Development 16: 6-21.
2. Kribelbauer, J. F., O. Laptenko, S. Chen,…, C. Prives, R. S. Mann, and H. J. Bussemarker. 2017. Quantitative Analysis of the DNA methylation sensitivity of transcription factor complexes. Cell Reports 19(1): 2383-2395.
3. Paulsen, M., and A. C. Ferguson-Smith. 2001. DNA methylation in genomic imprinting, development, and disease. The Journal of Pathology 195: 97-110.
4. Csankovszki, G., A. Nagy, and R. Jaenisch. 2001. Syngergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. Journal of Cell Biology 153(4): 773-783.
5. Robertson, K. D. 2002. DNA methylation and chromatin- unraveling the tangled web. Oncogene 21: 5361-5379.
6. Richardson, B., and R. Yung. 1999. Role of DNA methylation in the regulation of cell function. Journal of Laboratory and Clinical Medicine 134(4): 333-340.
7. Ambrosi, C., M. Manzo, and T. Baubec. 2017. Dynamic and context-dependent role of DNA methylation. Journal of Molecular Biology 429(1): 1459-1475.
8. Vilgalys, T. P., J. Rogers, C. Jolly, S. Mukherjee, J. Tung, and Baboon Genome Analysis Consortium. 2018. Evolution of DNA methylation in Papio baboons. bioRxiv, 400093.
9. Jabarri, J., S. Caccio, J. P. de Barros, J. Desgres, and G. Bernardi. 1997. Evolutionary changes in CpG and methylation levels in the genome of vertebrates. Gene 205: 109-118.
10. Piferrer, F. 2013. Epigenetics of sex determination and gonadogenesis. Developmental Dynamics 242(4): 360-370.
11. Martín-del-Campo, R., A. Bárcenas-Ibarra, I. Sifuentes-Romero, R. Llera-Herrera, and A. García-Gasca. 2018. Methylation status of the putative Pax6 promoter in olive ridley sea turtle embryos with eye defects: An initial approach. Mechanisms of development.
12. Olmo E. 2005. Rate of chromosome changes and speciation in reptiles. Genetica 125:185 203.
13. Fillon, V. 1998. The chicken as a model to study microchromosomes in birds: a review. Genetics Selection Evolution 30(3): 209–19.

14. Hillier, L.W., W. Miller, E. Birney, et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 432: 695–716.
15. Shah, S., A. F. McRae, R. E. Marioni, S. E. Harris, J. Gibson, A. K. Henders, ... and L. Murphy. 2014. Genetic and environmental exposures constrain epigenetic drift over the human life course. Genome research gr-176933.
16. Nilsen, F. M., B. B. Parrott, J. A. Bowden, B. L. Kassim, S. E. Somerville, T. A. Bryan, ... and S. E. Long. 2016. Global DNA methylation loss associated with mercury contamination and aging in the American alligator (Alligator mississippiensis). Science of The Total Environment 545: 389-397.
17. Parrott, B. B., S. Kohno, J.A. Cloy-McCoy, and L. J. Guillette Jr. 2014. Differential incubation temperatures result in dimorphic DNA methylation patterning of the SOX9 and aromatase promoters in gonads of alligator (Alligator mississippiensis) embryos. Biology of reproduction 90(1): 2-1.
18. Jones, M. J., S. J. Goodman, and M. S. Kobor. 2015. DNA methylation and healthy human aging. Aging Cell 14: 924–932.
19. Murdock, C., and T. Wibbels. 2003. Expression of DMRT1 in a turtle with temperature dependent sex determination. Cytogenetic and Genome Research 101: 3-4.
20. Yokomine, T., K. Hata, M. Tsudzuki, and M. Sasaki. 2006. Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. Cytogenet Genome Res 113:75–80.
21. Jurkowski T. P, and A. Jeltsch. 2011. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. PlOS One 6: e28104.
22. Colot, V., and J-L. Rossignol. 1999. Eukaryotic DNA methylation as an evolutionary device. BioEssays 21: 402-411.
23. Rolshausen, G., T. J. Davies, and A. P. Hendry. 2018. Evolutionary rates standardized for evolutionary space: perspectives on trait evolution. Trends in Ecology & Evolution 33(6): 379-389.
24. Li, Z. and X. Wan. 2018. Long-term evolutionary DNA methylation dynamic of protein-coding genes and its underlying mechanism. Gene 677: 96-104.
25. Varriale, A. 2014. DNA methylation, epigenetics, and evolution in vertebrates: facts and challenges. International Journal of Evolutionary Biology 2014(2): 1-7.
26. Laing, L.V., J. Vianna, E. L. Dempster, T. M. Uren Webster, R. van Aerle, J. Mill, and E. M. Santos. 2018. Sex specific transcription and DNA methylation profiles of reproductive and epigenetic associated genes in the gonads and livers of breeding zebrafish. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 222(1): 16-25.
27. Burdge, G. C., Hanson, M. A., Slater-Jefferies, J. L., and K. A. Lillycrop. 2007. Epigenetic regulation of transcription: A mechanism for inducing variations in phenotype (fetal programming) by differences in nutrition during early life? The British Journal of Nutrition 97(6): 1036–1046.
28. Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., … Lander, E. S. 2008. Genome scale DNA methylation maps of pluripotent and differentiated cells. Nature 454(7205): 766–770.
29. Huelsenbeck, J. P., R. Nielsen, J. P. Bollback. 2013. Stochastic mapping of morphological characters. Systematic Biology 52(2):131-158.
30. Harvey, P. H. and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.
31. Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401(1): 877-884.
32. Currie, T. E. and A. Meade. 2014. Keeping Yourself Updated: Bayesian Approaches in Phylogenetic Comparative Methods with a Focus on Markov Chain Models of Discrete Character Evolution. In: Garamszegi L. (eds) Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Springer, Berlin, Heidelberg.
33. Bollback, J. P. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7(1): 88.
34. Nielsen, R. 2002. Mapping mutations on phylogenies. Systematic Biology 51: 729-739.
35. Paradis, E. 2014. Simulation of phylogenetic data. In: Garamszegi L. (eds) Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Springer, Berlin, Heidelberg.
36. Munkemuller, T., S. Lavergne, B. Bzeznik, S. Dray, T. Jombart, K. Schiffers, and W. Thuiller. 2012. How to measure and test phylogenetic signal. Methods in Ecology and Evolution 3: 743-756.

37. Gittleman, J. L. and M. Kot. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. Systematic Zoology 39: 227-241.
38. Abouheif, E. 1999. A method for testing the assumption of phylogenetic independence in comparative data. Evolutionary Ecology Research 1: 895-909.
39. Blomberg, S. P., T. Gardland, Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57(4): 717-745.
40. Varriale, A. and G. Bernardi. 2006. DNA methylation and body temperature in fishes. Gene 385(1): 111-121.
41. Varriale, A. and G. Bernardi. 2006. DNA methylation in reptiles. Gene 385(1): 122-127.
42. Paradis, E., K. Schliep, and R. Schwarts. 2018. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 1: 3.
43. Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, and J. D. Thompson. 2007. Clustal W and clustal X version 2.0. Bioinformatics 23(21): 2947-2948.
44. Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22(1): 2688-2690.
45. Revell, L. J. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3(2): 217-223.
46. Revell, L. J. 2013. Two new graphical methods for mapping trait evolution on phylogenies. Methods in Ecology and Evolution 4(8): 754-759.
47. Felsenstein, J. 1985. Phylogenies and the comparative method. The American Naturalist 125(1): 1-15.
48. Keck, F., F. Rimet, A. Bouchez, and A. Franc. 2016. Phylosignal: an R package to measure, test, and explore the phylogenetic signal. Ecology and Evolution 6(9): 2774-2780.
49. Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating evolutionary radiations. Bioinformatics 24:129-131.
50. Legendre, P. and Y. Desdevises. 2009. Independent contrasts and regression through the origin. Journal of Theoretical Biology 259(1): 727-743.
51. Hansen, T. F. 2014. Use and Misuse of Comparative Methods in the Study of Adaptation. In: Garamszegi L. (eds) Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Springer, Berlin, Heidelberg.
52. Brown, R., J. Durbin, and J. Evans. 1975. Techniques for Testing the Constancy of Regression Relationships over Time. Journal of the Royal Statistical Society. Series B (Methodological) 37(2): 149-192.
53. Lepage, T., D. Bryant, H. Philippe, and N. Lartillot. 2007. A general comparison of relaxed molecular clock models. Molecular Biology and Evolution 24(12): 2669-2680.
54. Simpson, G. G. 1952. Periodicity in vertebrate evolution. Journal of Paleontology 26(3): 359-370.
55. Uhlenbeck, G. E. and L. S. Ornstein. 1930. On the theory of the Brownian motion. Physical Review 36: 823-841.
56. McMahon, D.P, A. Hayward, and J. Kathlrithamby. 2011. The first molecular phylogeny of strepsiptera (insecta) reveals an early burst of molecular evolution correlated with the transition to endoparasitism. PLOS one 6(6): e21206.
57. Johnston, I. A., and J. E. F.F. Dunn. 1987. Temperature acclimation and metabolism in ectotherms with particular reference to teleost fish. In Symposia of the Society for Experimental Biology 41: 67-93. Cambridge University Press Cambridge.
58. Towle, H. C. 1995. Metabolic Regulation of gene transcription in mammals. The Journal of Biological Chemistry 270: 23235-23238.
59. Blackburn, D. G. 2006. Squamate reptiles as model organisms for the evolution of viviparity. Herpetological Monographs 20(1): 131-146.
60. Phillips, T. 2008. The role of methylation in gene expression. Nature Education 1(1):116.