Chapter 17
Detecting Phenotypic Selection by Approximate Bayesian Computation in Phylogenetic Comparative Methods

Nobuyuki Kutsukake and Hideki Innan

Abstract This chapter discusses the fundamental structure and advantages of the approximate Bayesian computation (ABC) algorithm in phylogenetic comparative methods (PCMs). ABC estimates unknown parameters as follows: (1) simulated data are generated under a suite of parameters randomly chosen from their prior distributions; (2) the simulated data are compared with empirical data; (3) parameters are accepted when the distance between the simulated and empirical data is small; and (4) by repeating steps (1)–(3), posterior distributions of parameters will be gained. Because ABC does not necessitate mathematical expression or analytic solution of a likelihood function, ABC is particularly useful when a maximum-likelihood (ML) estimation is difficult to conduct (a common situation when testing complex evolutionary models and/or models with many parameters in PCMs). As an application, we analysed trait evolution in which a specific species exhibits an extraordinary trait value relative to others. The ABC approach detected the occurrence of branch-specific directional selection and estimated ancestral states of internal nodes. As computational power increases, such likelihood-free approaches will become increasingly useful for PCMs, particularly for testing complex evolutionary models that deviate from the standard models based on the Brownian motion.

N. Kutsukake (✉) · H. Innan
Department of Evolutionary Studies of Biosystems, The Graduate University for Advanced Studies, Kanagawa Hayama 240-0193, Japan
e-mail: kutsu@soken.ac.jp

N. Kutsukake · H. Innan
PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan

17.1 Background

> Evolution is messy. Rates, direction, and mode of evolution vary through time and among
> clades and characters, and this inconstancy itself will often be unpredictable and haphazard.
> Phylogenetic methods that ignore this variation will often produce inaccurate and misleading
> results. As a result, researchers must embrace statistical approaches that assess such variation
> rather than assuming constancy (Losos 2011).

Phylogenetic comparative methods (PCMs) are powerful approaches to test evolutionary models of speciation and trait evolution (Chap. 1). Although PCMs have been widely used in evolutionary biology, it is always important to remember that statistical inferences regarding evolutionary parameters are based on assumptions and hypotheses. In studies of continuous traits, single-rate Brownian motion (BM) has been commonly used as a model for character evolution. BM represents a neutral evolution or evolution tracking a continuously fluctuating optimum value and is a good approximation for describing a *pattern* of trait divergence. However, this does not mean that the BM-like evolutionary mode can always approximate the *process* of divergence well (Estes and Arnold 2007; Gingerich 2009). It is natural to assume that the process comprises heterogeneous evolutionary rates and modes; that is, they vary within lineages, among branches, and within clades (Losos 2011). One approach to coping with this heterogeneity is to test a multiple-rate BM model against a simple single-rate BM model. Researchers can set a model with varying rates for each branch or for mono-phyletic subgroups based on a priori biological hypothesis (O'Meara et al. 2006; Thomas et al. 2006), or set a model without specifying a hypothesis (Venditti et al. 2011). Another approach includes scaling parameters on branch lengths or evo-lutionary rates, and test models of punctuated, accelerating/decelerating, or early burst evolution (Table 17.1; Blomberg et al. 2003; Pagel 1997, 1999; Harmon et al. 2010). Additionally, the local occurrence of stabilising selection can be modelled by using the Ornstein–Uhlenbeck (OU) process (Table 17.1; Chap. 15). Parameters specific to each model (Table 17.1) are incorporated to calculate an expected covariance among the traits of each species. The variance–covariance matrix, a critical component of the likelihood function of PGLS (phylogenetic generalised least squares), is used for estimating unknown parameters via a maximum-likelihood (ML) estimation or a Bayesian approach. With these approaches, it is currently possible to address a wide range of questions. However, some methods are mathematically complex and not always transparent for general users of PCMs. Moreover, traditional approaches may not be well enough estab-lished to test complex evolutionary scenarios with many parameters because the description of the variance–covariance matrix is not straightforward to gain. It is preferable to have a flexible toolkit that allows for testing of such complex evo-lutionary models.

A simulation-based likelihood approach using an approximate Bayesian com-putation (ABC) (Bokma 2010; Slater et al. 2012; Kutsukake and Innan 2013) may

Table 17.1 Examples of parameters used in previous studies of PCMs

Model	Parameters	Biological meaning
Scaling branch length	λ^a	Phylogenetic signal
	δ^a	Temporal rate change
	κ^a	Gradual versus punctuational evolution
ACDC	g^b	The overall rate of acceleration (AC) or deceleration (DC)
Early burst	r^c	The pattern of rate change through time
OU (whole phylogeny)	d^b	A restraint parameter in the OU transformation
OU	α^d	Strength of stabilising selection

[a] Pagel (1997, 1999)
[b] Blomberg et al. (2003)
[c] Harmon et al. (2010)
[d] Hansen (1997), Butler and King (2004)

be one solution. This approach is useful when analytic expressions of likelihood and the ML estimator cannot be gained. In this chapter, we aim (1) to explain the basic structure, method, and caveats of ABC, (2) review the application of ABC to PCMs, (3) provide a protocol for the ABC approach, (4) provide an example using ABC, and (5) discuss the future direction of the ABC approach.

17.2 Approximate Bayesian Computation

The ABC framework was originally developed in population genetics (Tavare et al. 1997) and gradually introduced to the disciplines of ecology and evolutionary biology (Beaumont 2010; Bertorelle et al. 2010; Csillery et al. 2010). The major advantage of ABC is that parameter estimation can be performed even when the likelihood of the data cannot be computed, most often due to data complexity. The fundamental structure of ABC is as follows: let x be the observed data and assume an evolutionary model with the parameters η, which we aim to estimate. η could be a vector with multiple parameters.

(1) Determine the prior distributions of all parameters η in x.
(2) For each parameter, generate a random value from the prior distribution. A random set of the parameters at the ith simulation is denoted by η_i'.
(3) Simulate data x_i' using η_i'.
(4) Accept η_i' if x_i' is identical to the observed data x.
(5) Go to (2) until a large number of accepted η' values have been accumulated.

In most cases, the simulation will rarely produce x_i' that is completely identical to the observed data x. To solve this problem, instead of using the full data set, ABC usually employs summary statistics (denoted by s, which usually comprises several types of summary statistics, i.e., $s = [S_1, S_2, \ldots S_n]$) and use $s(x)$ and $s(x_i')$ instead of x and x_i'. Even when summary statistics are used, it may be rare to gain

an $s(x_i')$ value that is exactly the same as $s(x)$. In such a case, we can set a certain range of tolerance. One simple algorithm, a rejection sampling, uses the following relaxed criterion:

$$(4') \text{ Accept } \eta_i' \text{ when } \rho(s(x) - s(x_i')) < \varepsilon,$$

where ε is a tolerance and $\rho(\cdot)$ is a function for calculating a distance between $s(x)$ and $s(x_i')$, representing their similarity. By setting a tolerance, the efficiency of parameter acceptance will dramatically increase in comparison with (4). In addition to this simple rejection sampling algorithm, more sophisticated methods have been proposed to improve data-acceptance efficiency, such as importance sampling or ABC–MCMC (Markov chain Monte Carlo; see Marjoram et al. 2003 and Majoram and Tavare 2006 for details).

17.3 Caveats of ABC

Although Sect. 17.2 provided the fundamental structure of ABCs, several general caveats remain that users should be aware of. Because of space limitations, we will briefly discuss three core points, summary statistic sufficiency, estimation robustness, and model selection, although other general caveats of Bayesian computation apply (see Gelman et al. 2013).

First, the number and choice of summary statistics is critical in ABC (see Beaumont et al. 2002; Csillery et al. 2010; Leuenberger and Wegmann 2010 for details). The use of summary statistics reduces the dimensionality and complexity of the data (Tavare et al. 1997). Accordingly, if researchers use too few summary statistics, too much information is discarded, resulting in a low resolution of the parameter estimation. However, if too many summary statistics are used, the acceptance rate will be so low that it will be computationally intensive. Researchers must examine the performance of their chosen summary statistics before applying ABC to real data, such as by conducting power tests with artificially generated data.

The second point concerns estimation robustness and computational efficiency determined by the level of tolerance ε. If tolerance is not sufficiently small, parameter estimation will be rough (Beaumont 2010), while too severe tolerance is not practical because computational efficiency will be too low. Therefore, researchers are required to set an adequate level of tolerance in ABC, meaning that a certain amount of subjectivity and uncertainty remains in parameter estimation. To address this problem, post hoc correction approaches have been proposed to increase the accuracy of parameter estimation, such as post-sampling regression adjustment using general linear model (ABC–GLM; Leuenberger and Wegmann 2010).

Third, simple model selection cannot be used in the framework of ABC. This difficulty stems from the fact that likelihoods are not calculated, and data acceptance in each model depends on summary statistics and tolerance (Beaumont 2010;

Csillery et al. 2010; Robert et al. 2011). Although this can be problematic in the case of comparing un-nested models, model selection based on posterior distributions is relatively straightforward when models are nested (explained below).

17.4 Application of ABC to PCMs

Three PCM studies have used ABC for analysing trait evolution (Table 17.2; see Rabosky (2009) for an analysis of clade diversification using ABC). These studies aimed to test evolutionary models that are difficult or not straightforward to handle in the traditional framework of PCMs.

The first PCM study using ABC was performed by Bokma (2010). This study aimed to separate the effects of parameters of cladogenetic evolution and of anagenetic evolution on trait disparity. There are wide variations in trait disparity among clades, and understanding the ecological and evolutionary causes of this phenomenon has been a central interest in macroevolutionary studies. It is likely that a trait disparity is positively correlated with a parameter of anagenetic evolution (i.e. rate parameter of BM). Additionally, the number of speciation events may be positively correlated with the trait disparity because the interval between speciation events should be smaller in larger clades, automatically resulting in larger trait variance (Ricklefs 2004, 2006; Purvis 2004). Thus, the effects of anagenetic and cladogenetic evolution on trait disparity are confounded and difficult to separate. To solve this problem, Bokma (2010) applied ABC to trait disparity in passerine data (originally used in Ricklefs 2004) to investigate the relative importance of those parameters with respect to anagenetic evolution, cladogenetic evolution, and their combination. Bokma (2010) conducted a simulation using a model comprising two free parameters (anagenetic and cladogenetic parameters) and estimated those parameters using phenotypic variance as summary statistics. They found that gradual anagenetic change was more important than a combination of anagenetic and cladogenetic evolution in terms of explaining the phenotypic divergence observed in this group.

In another study, Slater et al. (2012) used a mixture of Markov-chain Monte Carlo (MCMC) and ABC to infer the speciation/extinction rates and parameters of trait evolution in an incompletely sampled phylogeny. This study was motivated by a common problem in PCM studies: it is usually difficult to perform complete sampling of a phylogeny and collect trait data from all species. Data incompleteness could reduce the accuracy of statistical inference of evolutionary parameters and the power to test models. Nevertheless, ABC can handle such incomplete data relatively well because data are transformed into summary statistics so that it is not always necessary to sample all data (as long as collected data are unbiased and represent each clade). Using this unique characteristic of ABC, Slater et al. (2012) developed a new framework with which to test evolutionary models using incomplete data. First, speciation and extinction parameters were

Table 17.2 Summary of three PCM studies using ABC

Study	Parameters	Reason for using ABC	Summary statistics	Empirical example
Bokma (2010)	Anagenetic and cladogenetic evolution parameters	No analytical solution for an evolutionary model	Phenotypic variance	Parameter estimation of phenotypic variance in passerine birds
Slater et al. (2012)	Speciation rate, extinction rate, ancestral state, BM rate	Incompletely sampled phylogeny and trait data	Mean and variance for terminal lineages	Comparison of terrestrial carnivores and pinnipeds
Kutsukake and Innan (2013)	See Table 17.3 for details	Branch-specific directional selection	Direct (and full) likelihood	Detection of directional selection in human brain size

estimated from known data by MCMC. They conducted simulations to generate a phylogeny under a birth/death process using sampled parameter values. Next, using ABC, they simulated trait evolution according to BM rates on the generated phylogeny using the rate parameter(s) of BM and sampled the trait value at the root. Finally, summary statistics (mean and variance) calculated from the simulated trait data were compared with those calculated from the real data of each clade. This study applied this algorithm to test the difference in BM rate parameters of the body size between pinnipeds and terrestrial carnivores. This study was interested in estimating two to three parameters, namely the trait value at the root, and one or two rate parameters assigned to those two groups. Given that a model with one rate parameter is nested within a model with two rate parameters, this study used a posterior probability to select the two models. The ABC approach revealed that, contrary to expectation, rate parameters did not differ between the two groups.

Testing multi-BM models is also possible in the traditional framework of PCMs (O'Meara et al. 2006; Thomas et al. 2006), but it requires the assumption of complete sampling. Slater et al. (2012) demonstrated that ABC handles this technical problem quite well and stated that their method is applicable to other existing evolutionary models, thereby providing an important first step in testing complex evolutionary models by ABC. This method was implemented in the software *MECCA* (Modeling Evolution of Continuous Characters using ABC)

Third, our recent study (Kutsukake and Innan 2013) used ABC under an evolutionary model with heterogeneous evolutionary modes and rates within a phylogeny (assuming that the phylogeny is known). Although this algorithm is designed to model wide ranges of trait evolution, one useful application is to detect directional selection occurring locally within the phylogeny. In molecular evolution studies, there is a popular approach to measure branch-specific directional selection (e.g. dN/dS or Ka/Ks) (Li 1997; Yang 2006). Motivated by this approach, our model was designed to incorporate branch-specific selection parameters and

allowed statistical testing and measurement of the intensity of selection. We considered the BM model (with slight modifications) to be a null neutral model. This simplest model can be extended by adding as many branch-specific selection parameters as desired for setting an alternative.

Using this framework, Kutsukake and Innan (2013) showed a simple example analysis of data on brain volume among four species of great apes. In total, three parameters were involved: the trait value at the most recent common ancestor, the background evolutionary rate, and the strength of directional selection at the human lineage (the method will be explained later). It was shown that the trait evolution on the branch reaching to humans significantly deviated from the BM mode, exhibiting strong evidence of branch-specific directional selection.

Although only a few PCM studies have used ABC thus far, ABC has a great potential to be applied to a wide range of problems and settings. Since phenotypic evolution consists of a process in which trait values increase or decrease, listing all parameters making up this process should be sufficient to build any complex evolutionary models. Of PCM studies using ABC, only our model is flexible enough to be applied to various evolutionary processes at any time point; these can be deviate from the simple BM process and include various modes and intensities of selection on different branches. Another strength is that our framework employs direct likelihood as a summary statistic, by which it is possible to avoid the problem of which and how many summary statistics should be used in ABC (see Sect. 17.3). Furthermore, intraspecific variation, a factor that has been overlooked but is known to affect to parameter estimation in PCMs (Garamszegi and Møller 2010; Chap. 7), and uncertainty in phylogeny can be taken into account (see below and Table 17.3). Below, we describe the model and algorithm of our approach in detail.

17.5 ABC Algorithm in Kutsukake and Innan (2013)

We herein explain the algorithm of Kutsukake and Innan (2013) in more details (see also Fig. 17.1a for a simplified structure of this algorithm; an example program written in the C language in shown in the Online Practical Material, http://www.mpcm-evolution.com).

A necessary data set for applying this algorithm is the same as those for other studies. First, trait data are required for each species. Intraspecific variation can also be incorporated by setting a distribution of the trait. Any kind of distribution can be handled, from a regular quantitative trait that likely follows a simple normal distribution to a trait with a discrete distribution. The phylogeny ψ (topology and branch length τ) of the species is also needed and is assumed to be known (this assumption could be relaxed as phylogenetic uncertainty can be taken into account in the algorithm; see Table 17.3). $\varLambda$ represents all other parameters involved. At minimum, may comprise the trait value of the most recent common ancestor (θ_0) and evolutionary rate μ. The evolutionary rate, the number of evolutionary events (i.e. mutation and

Table 17.3 Main parameters used in a simulation-based likelihood approach by Kutsukake and Innan (2013)

| Parameter	Biological meaning	Notes
Known parameters		
ψ	Species tree (topology and branch lengths)	When topology and/or branch length includes uncertainty, it is possible to consider those uncertainties by using a randomly chosen topology and/or a set of random values for branch length in each simulation
τ_i	Length of the ith branch on the tree in a given unit (e.g. time or genetic distance)	
n	Number of nodes	
$\Omega = \{\Omega_1, \Omega_2, \Omega_3, \ldots \Omega_n\}$	Observed nodal data (e.g. mean)	This setting assumes data contain no measurement errors. Measurement errors can be incorporated by considering that σ contains both intraspecific variation and measurement errors, or by setting a new parameter set to represent measurement error. Internal nodes such as fossil data can also be used
$\sigma = \{\sigma_1, \sigma_2, \sigma_3, \ldots \sigma_n\}$	Observed nodal data on intraspecific variation (standard deviation or other parameters)	
A parameter set of interests (Λ)		
θ_0	Phenotype of the MRCA	
μ	Evolutionary rate (the number of a phenotypic change) per time unit (e.g. a million years, a generation) and genetic distances	μ^+ and μ^- for increasing or decreasing a trait value. It is possible to assume different evolutionary rates for each branch i by setting μ_i^+, μ_i^-
Other factors		
ϕ	Changes in phenotypic values caused by a single event of evolution; this parameter can be set arbitrarily	ϕ^+, ϕ^- for different effects for increasing or decreasing trait value. For example, it is possible to model a larger evolutionary effect increasing a trait than one decreasing a trait by setting $\phi^+ > \phi^-$. In the case that the evolutionary effect differs among branches, a different value (ϕ_i^+, ϕ_i^-) can be used

(continued)

Table 17.3 (continued)

Parameter	Biological meaning	Notes
$s_i = f(\mu, \tau)$ or $f(\mu, \tau, \ldots)$	Number of fixed mutations that increase (s_i^+) or decrease (s_i^-) the phenotypic value at a certain branch i	A random integer from a Poisson distribution with mean $\mu_i^+ \tau_i$ (or $\mu_i^- \tau_i$) in the case of BM evolution. If one wants to test more complex evolutionary rate in which the evolutionary rate is a function of the branch length or time from root, such as accelerating or decelerating evolutionary rate, additional parameters can be used in this function
$\Theta = \{\theta_1, \theta_2, \theta_3, \ldots \theta_n\}$	Simulated data of n descendent species	

fixation), is denoted by μ_i at branch i, whose length is τ_i. Thus, the expected number of evolutionary events increasing and/or decreasing the trait value is $\mu_i \tau_i$. The effect of each evolutionary event on the trait is denoted by ϕ, which should follow a certain distribution. By setting branch-specific $\mu_i \tau_i$, it is possible to model the density distribution of the phenotypic change for each branch and to test wide ranges of evolutionary models such as branch- or clade-specific directional selection (see below) or the OU process (see Kutsukake and Innan 2013 for details). This setting enables researchers to treat those models and a null model (neutral evolution) as nested models, allowing one to avoid the complicated problem of model selection in ABCs (see Sect. 17.3).

Under this framework, it is possible to apply the ABC algorithm along the following four steps. The steps correspond to the general ABC procedure in Sect. 17.2.

Step 1 *Determine the prior distribution of each parameter.* An advantage of Bayesian statistics is that it enables the setting of informative (i.e. strong) prior distributions based on prior biological knowledge.

Step 2 *Choose a random value for each parameter from the prior distribution.* Parameters used in the simulation (Λ') are randomly chosen from their prior distributions.

Step 3 *Let trait values evolve by simulation.* Simulation of the trait evolution of phylogeny ψ is conducted using Λ'. As a result, simulated values Θ of traits of n species are gained.

Step 4 *Calculate the likelihood by comparing simulated data with the real data and determine whether that parameter set is accepted or rejected.* Each simulated value θ_i is compared with the real value Ω_i. By comparing n species, a joint probability (full likelihood) $\Pr(\Omega|\Theta) = \Pr(\Omega_1, \Omega_2, \Omega_3, \ldots, \Omega_n | \theta_1, \theta_2, \theta_3, \ldots, \theta_n)$ will be calculated. In the case that this probability is computationally very difficult to gain, a composite

Fig. 17.1 a An illustrative
example of the ABC
approach. A trait simulation
can be conducted based on a
parameter set (Λ) on a
phylogeny with a discrete
timescale. The number of
simulation rounds is shown in
parenthesis. A number (x) of
simulations were conducted
until enough samples were
collected to infer posterior
distribution. **b** The trait value
can be simulated by the
BM-like evolutionary mode
(a *lower branch*) or
directional selection to
increase trait value (an *upper
branch*). Vertical axis
indicates a trait value, and
white arrows indicate
direction of trait evolution for
each branch, with its length
corresponding to the number
of evolutionary events
increasing or decreasing trait
values
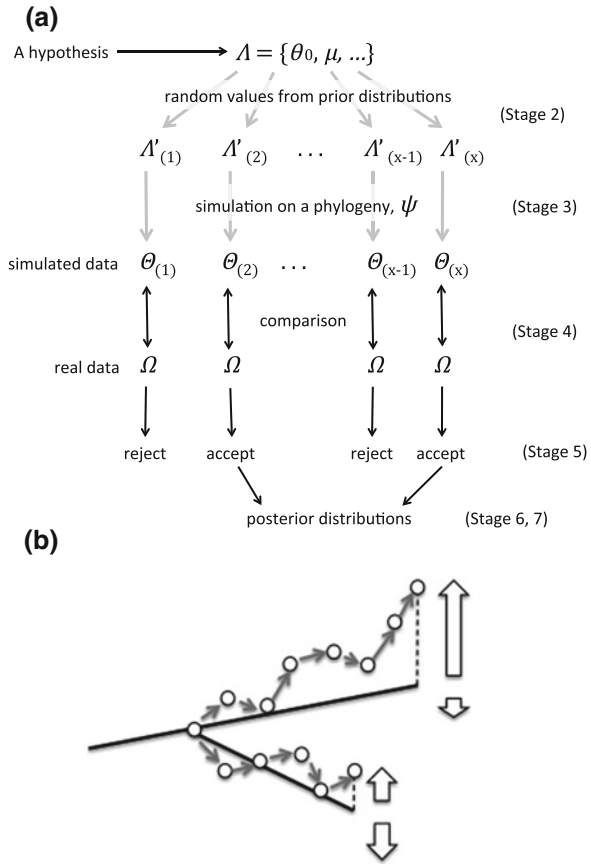
$$\text{likelihood } \Pr(\Omega|\Theta) = \prod_{i=1}^{n} \Pr(\Omega_i|\theta_i) \text{ may be used as an approximate proxy,}$$

which should work fairly well as long as the sampled species are rea-
sonably diverged. Here, one important advantage in our framework is that
the choice of summary statistics is avoided. Although previous studies
have used means or variance as summary statistics (Table 17.2), this study
employed a more straightforward method by using the direct likelihood
instead of a summary statistic. The use of likelihood as a summary sta-
tistics may not be common in standard ABC, where data are usually
represented by a set of summary statistics and likelihood cannot be ana-
lytically computed. Note that we can compute the likelihood of the
observed data given a "simulated data set". In our framework, given a
parameter set, a single run of random simulation provides a simulated data
set representing a single realisation of the random process. Then, the
likelihood is computed given this simulated data set. This is different from
the standard ML estimation that requires the likelihood given a parameter

set. Intraspecific variation in the trait data can be considered by using the probability density function when calculating $\Pr(\Omega_i|\theta_i)$, the probability of gaining Ω given θ. Thus, by assuming a certain distribution of intraspecific variation, we can evaluate the similarity between the simulated and real data using the likelihood. As a consequence, our algorithm can avoid the common problem of choosing appropriate summary statistics.

Using $\Pr(\Omega|\Theta)$, acceptance of Λ' can be determined. As described above, there are several methods of judgment (Marjoram et al. 2003; Majoram and Tavare 2006); the researchers determine which method will be used. However, we caution users on choosing the acceptance threshold. A strict threshold increases the precision of parameter estimation, but if it is too strict, the computational load may be too large. A reasonable choice is desired so as to gain posterior distributions within a realistic computation time (see Sect. 17.3).

Step 5 *Obtain posterior distributions of the parameters and assess the importance of each parameter.* Repeat Steps (2)–(4) until a sufficient number of parameters is accepted. This usually requires intensive iterative computation, particularly when there are many parameters. Posterior distributions and credible intervals can be used to judge whether each parameter is different from a specific value (e.g. a value expected under a null model) and to accordingly test which evolutionary models are supported.

17.6 Detecting Branch-Specific Directional Selection

In this ABC approach, it is possible to let phenotypic traits evolve via directional selection only in certain branches (Fig. 17.1b). There are various ways to model such local occurrence of directional selection. Kutsukake and Innan (2013) introduced a simple model in which the direction and intensity of selection is parameterised by a single parameter, k. That is, it is assumed that if selection favours the increase in a trait at a branch with $\tau\mu$, the expected numbers to increase and decrease the trait are given by $k\tau\mu$ and $\tau\mu/k$, respectively. When k equals 1, the result is the same as neutral evolution, as the numbers of evolutionary events increasing and decreasing the trait value are identical on average.

The model using k is not the only universal way to model directional selection. Other types of directional selection can be flexibly modelled by setting new parameters. For example, directional selection in one direction can co-occur with constant occurrence of neutral evolution, including evolution in the opposite direction. In such cases, researchers can set a new parameter, and evolutionary events increasing a trait value can be multiplied by that parameter while those decreasing the trait value will be untouched. In other cases, consistent directional selection of a trait may start to operate from a certain intermediate point of a branch because of alterations in a fitness landscape by environmental changes. In such cases, researchers can create a model in which a trait value evolves neutrally

until the intermediate point of the branch, and the evolutionary rate is multiplied by a parameter of directional selection in the rest of the branch.

Given that there are several ways to model directional selection and to represent the relative intensity of selection, it is useful to quantify the expected change in the trait value for each branch, which is denoted by Δ_{sel}. In the case of using k, the Δ_{sel} value can be calculated by $\mu_i^+ \tau_i k - \mu_i^- \tau_i / k$ (Kutsukake and Innan 2013). To test the presence of directional selection, the posterior distribution of Δ_{sel} can be compared to zero or to a maximum value of the trait change under a pure neutral model (denoted by Δ_{null} in Kutsukake and Innan 2013). The advantage of using Δ_{sel} is that it can cancels out the confounding relationship between μ and k (or other parameters for directional selection); that is, k is inversely correlated with μ because those two parameters compensate for each other to let a phenotypic value increase (or decrease) towards a certain value. In other words, a low evolutionary rate must necessitate strong directional selection, whereas weak directional selection may be sufficient when evolutionary rate is large. In such a situation, the most meaningful quantity should be Δ_{sel} rather than μ or k.

17.7 Application to Weevil Rostrum Demonstrating Branch-Specific Direction Selection

We analysed weevil rostrum evolution to show how our framework can be applied to a complex model of trait evolution with branch-specific directional selection (see OPM). Toju and Sota (2006) studied interspecific variation in the rostrum, an organ used to excavate host plant fruits to lay eggs inside, among seven species of weevils. They investigated two nearly isolated subpopulations of one of those species, *Curculio camelliae*. Interestingly, their interpopulation comparison between the subpopulations showed a positive correlation between the thickness of the camellia fruit walls and the weevil rostrum length, suggesting that an arms race occurred between those two traits. Toju and Sota (2006) applied PGLS (Chaps. 5 and 6) and showed significant effects of two scaling parameters, κ and δ (Pagel 1999). The significant effects of these parameters indicate punctuated evolution and the large effect of a short branch. They also estimated the ancestral states by the ML approach (Schulter et al. 1997). As a result, the rostrum length of the common ancestor (an internal node C in Fig. 17.2a) between *C. camelliae* and its sister species (*C.* species on *C. sasanqua*) was estimated at 8.12, an approximately intermediate value of their descendent species (Fig. 17.2a). This estimated value suggests that the rostrum length should have increased in *C. camelliae* but decreased in its sister species.

Note that these results reflect BM-based models, in which a uniform rate of evolution was applied to the entire tree, and directional selection on the lineage of *C. camelliae* was not specifically incorporated. Therefore, the estimated ancestral states of *C. camelliae* may have been overestimated. We re-analysed the data

reported by Toju and Sota (2006) and tested the model showing that branch-specific directional evolution has increased the rostrum length in the lineage leading to *C. camelliae*. In the ABC, we conducted trait simulation by setting three parameters: the ancestral value, background (neutral) evolutionary rate, and intensity of directional selection k. We assumed a Gaussian distribution of rostrum length and calculated the probability for the ith species by considering intraspecific variation σ as follows:

$$\Pr(\Omega_i | \theta_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(\theta_i - \Omega_i)^2}{2\sigma_i^2}\right]. \qquad (17.1)$$

We accepted data by the criterion that the probability of acceptance is proportional to a direct likelihood (Kutsukake and Innan 2013). Using this equation, we can avoid the common problem of the tolerance (see Sect. 17.3).

The obtained posterior distributions favoured our model of branch-specific directional selection over the null model based on neutral evolution. The posterior distribution of k did not overlap with 1, the value indicating neutral evolution (Fig. 17.2b). The ancestral value estimated by this branch-specific directional selection model was much smaller than that estimated by ML estimator, not only in the ancestor of *C. camelliae* but also in other internal nodal species (Fig. 17.2b). These differences are obviously due to the incorporation of branch-specific directional selection. This estimation supports the idea that a co-evolutionary arms race promoted the evolution of an exaggerated trait.

Although an OU model can also be applied to the branch to the Japanese species, we believe that applying directional selection is more suitable in this case because there should be no adaptive optimum in an arms race, and phenotypic shifts should be unidirectional.

17.8 Further Applications

An advantage of the ABC approach is that researchers can flexibly test complicated evolutionary models even when their likelihood cannot be computed. This advantage meets recent demands of PCMs as comparative approaches are currently applied to wide ranges of biological problems and complicated models. Another advantage of ABC is that trait simulation provides a good opportunity to critically consider each stage of an evolutionary event. Using simulations, researchers can create models focusing on evolutionary processes, rather than evolutionary patterns. However, user should be aware of the caveats of ABCs (discussed in Sect. 17.3). Careful calibration of summary statistics, tolerance, power analysis, and well-designed model settings are necessary. The ABC approach itself is also rapidly developing, and we recommend that ABC users follow the ongoing improvements and debates.
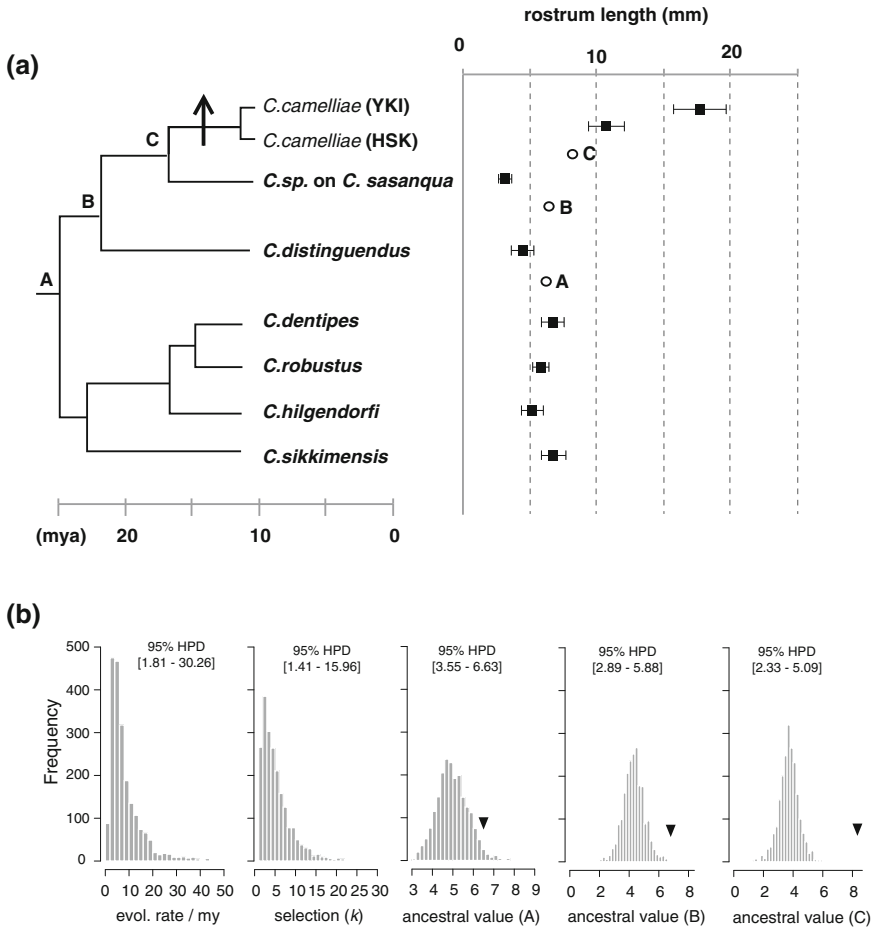
Fig. 17.2 a Phylogeny and trait values (rostrum length; mean and 1 SE) of seven species of weevils (*C. camelliae* were sampled in two locations: YKI and HSK). Phylogenetic relationship of these species was estimated by mitochondrial COI gene sequences (Toju and Sota 2006). The trait values indicated by *white circles* correspond to the estimated values of internal nodes *A*, *B*, and *C* by Toju and Sota (2006). We tested an evolutionary model in which directional selection to increase a trait value has occurred between internal nodes *C* of the ancestor of *C. camelliae* (indicated by a *thick line* with an *upward arrow*). Simulations were performed by a discrete timescale using a million years. **b** Posterior distributions of parameters and 95 % confidence intervals based on 2,000 accepted parameter sets (ancestral value at internal nodes *A*, *B*, and *C*, evolutionary rate, and selection) of our ABC analysis. We did not accept simulated data with a negative trait value. *Black inverted triangles* indicate trait values estimated by Toju and Sota (2006). In this analysis, ϕ was set as an exponential distribution with a mean of 0.05 mm, meaning that one evolutionary change results in a trait change whose magnitude is a random value from an exponential distribution whose average is 0.05 mm. Prior distributions were set as follows: MRCA $\sim U(3.15, 9)$, evolutionary rate per a million year $\sim U(0, 50)$, and $k \sim U(0.0001, 30)$, where $U(\cdot)$ indicates a uniform distribution

We believe that ABC-based PCMs have many developmental possibilities and can be applied to broad ranges of evolutionary questions and empirical data. As our example showed, one can model branch-specific directional selection by setting a specific value of the evolutionary rate (μ; Table 17.3) at each branch. As an extension of this directional selection model, an acceleration of selection pressure, often witnessed during co-evolutionary arms races, can be also analysed. With further modifications ϕ (Table 17.3), one can model an evolutionary scenario in which different species have a different degree of trait change stemming from one evolutionary event. Such a situation is common in analyses of size traits or body mass, in which the degree of trait change (e.g. an increase/decrease in body mass) positively correlates with its species trait value (e.g. body mass). Furthermore, changes in the evolutionary rate and mode in the middle of a branch, hybridisation, and cultural evolution can also be flexibly incorporated in this framework.

Finally, we should note that our ABC approach may not fit to a ready-made software or statistical library because flexibility is the most important advantage of the ABC. It is ideal that each user writes programs for evolutionary models that the given researcher would like to test (see OPM).

References

Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. Ann Rev Ecol Evol Syst 41:379–406

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics 162:2025–2035

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Mol Ecol 19:2609–2625

Blomberg SP, Garland T Jr, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745

Bokma F (2010) Time, species, and separating their effects on trait variance in clades. Syst Biol 59:602–607

Butler MA, King AA (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am Nat 164:683–695

Csillery K, Blum MG, Gaggiotti OE, François O (2010) Approximate Bayesian computation (ABC) in practice. Trends Ecol Evol 25:410–418

Estes S, Arnold SJ (2007) Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. Am Nat 169:227–244

Garamszegi LZ, Møller AP (2010) Effects of sample size and intraspecific variation in phylogenetic comparative studies: a meta-analytic review. Biol Rev 85:797–805

Gelman A, Carlin JB, Stern HS, Rubin DB (2013) Bayesian data analysis, 3rd edn. CRC Press, London

Gingerich PD (2009) Rate of evolution. Ann Rev Ecol Evol Syst 40:657–675

Hansen TF (1997) Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351

Harmon LJ, Losos JB, Jonathan Davies T, Gillespie RG, Gittleman JL et al (2010) Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396

Kutsukake N, Innan H (2013) Simulation-based likelihood approach for evolutionary models of phenotypic traits on phylogeny. Evolution 67:355–367

Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. Genetics 184:243–252

Li W-H (1997) Molecular evolution. Sinauer Associates, Sunderland, Massachusetts

Losos JB (2011) Seeing the forest for the trees: the limitations of phylogenies in comparative biology. Am Nat 177:709–727

Marjoram P, Tavare S (2006) Modern computational approaches for analysing molecular genetic variation data. Nat Genet Rev 7:759–770

Marjoram P, Molitor J, Plagnol V, Tavare S (2003) Markov chain Monte Carlo without likelihoods. Proc Natl Acad Sci USA 100:15324–15328

O'Meara BC, Ane CM, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution in different groups using likelihood. Evolution 60:922–933

Pagel M (1997) Inferring evolutionary processes from phylogenies. Zool Scr 26:331–348

Pagel M (1999) Inferring the historical patterns of biological evolution. Nature 401:877–884

Purvis A (2004) Evolution: how do characters evolve? Nature 432:166

Rabosky DL (2009) Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. Syst Biol 58:629–640

Ricklefs RE (2004) Cladogenesis and morphological diversification in passerine birds. Nature 430:338–341

Ricklefs RE (2006) Time, species, and the generation of trait variance in clades. Syst Biol 55:151–159

Robert CP, Cornuet JM, Marin JM, Pillai NS (2011) Lack of confidence in approximate Bayesian computation model choice. Proc Natl Acad Sci USA 108:15112–15117

Schulter D, Price T, Mooers AO, Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711

Slater GJ, Harmon LJ, Wegmann D, Joyce P, Revell LJ, Alfaro ME (2012) Fitting models of continuous trait evolution to incompletely sampled comparative data using approximate Bayesian computation. Evolution 66:752–762

Tavare S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. Genetics 145:505–518

Toju H, Sota T (2006) Phylogeography and the geographic cline in the armament of a seed-predatory weevil: effects of historical events vs. natural selection from the host plant. Mol Ecol 15:4161–4173

Thomas GH, Freckleton RP, Szekely T (2006) Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proc R Soc Lond B 273:1619–1624

Venditti C, Meade A, Pagel M (2011) Multiple routes to mammalian diversity. Nature 479:393–396

Yang Z (2006) Computational molecular evolution. Oxford University Press, Oxford, UK