MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 4770–4784 doi: 10.1111/mec.12889

# Influence of gene flow on divergence dating – implications for the speciation history of *Takydromus* grass lizards

SHU-PING TSENG,*†‡ SHOU-HSIEN LI,* CHIA-HUNG HSIEH,§ HURNG-YI WANG‡¶** and SI-MIN LIN*

*Department of Life Science, National Taiwan Normal University, Taipei 116, Taiwan, †Department of Entomology, National Taiwan University, Taipei 106, Taiwan, ‡Graduate Institute of Clinical Medicine, National Taiwan University, Taipei 100, Taiwan, §Department of Forestry and Nature Conservation, Chinese Culture University, Taipei 111, Taiwan, ¶Institute of Ecology and Evolution, National Taiwan University, Taipei 106, Taiwan, **Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei 100, Taiwan

## Abstract

**Dating the time of divergence and understanding speciation processes are central to the study of the evolutionary history of organisms but are notoriously difficult. The difficulty is largely rooted in variations in the ancestral population size or in the genealogy variation across loci. To depict the speciation processes and divergence histories of three monophyletic *Takydromus* species endemic to Taiwan, we sequenced 20 nuclear loci and combined with one mitochondrial locus published in GenBank. They were analysed by a multispecies coalescent approach within a Bayesian framework. Divergence dating based on the gene tree approach showed high variation among loci, and the divergence was estimated at an earlier date than when derived by the species-tree approach. To test whether variations in the ancestral population size accounted for the majority of this variation, we conducted computer inferences using isolation-with-migration (IM) and approximate Bayesian computation (ABC) frameworks. The results revealed that gene flow during the early stage of speciation was strongly favoured over the isolation model, and the initiation of the speciation process was far earlier than the dates estimated by gene- and species-based divergence dating. Due to their limited dispersal ability, it is suggested that geographical isolation may have played a major role in the divergence of these *Takydromus* species. Nevertheless, this study reveals a more complex situation and demonstrates that gene flow during the speciation process cannot be overlooked and may have a great impact on divergence dating. By using multilocus data and incorporating Bayesian coalescence approaches, we provide a more biologically realistic framework for delineating the divergence history of *Takydromus*.**

*Keywords*: divergence time, gene flow, parapatric speciation, population demography, reptiles, *Takydromus*

*Received 16 May 2013; revision received 7 August 2014; accepted 8 August 2014*

## Introduction

Dating the time of genetic divergence and understanding the speciation process are two central and interconnected issues in the study of the evolutionary history of organisms. However, obtaining accurate estimates of divergence and depicting the speciation process are notoriously difficult because of two primary sources of error: the stochasticity of the coalescent process and the mode of species formation, that is, whether gene flow occurred during speciation (Edwards & Beerli 2000; Wu 2001). Because the divergence time of a gene between

Correspondence: Hurng-Yi Wang, Fax: +886-2-23709820; E-mail: hurngyi@ntu.edu.tw and Si-Min Lin, Fax: +886-2-23912904; E-mail: fish@ntnu.edu.tw

species is the species divergence time ($t$) plus its coalescence time, where the latter is exponentially distributed with a mean equal to 2*Ne* generation times (*Ne* being the effective population size at the time of speciation), divergence dating based on gene trees tends to overestimate species divergence times (Edwards & Beerli 2000), especially for recent divergence events (Carstens & Knowles 2007). Furthermore, variation in *Ne* across different loci can lead to a high variance in divergence time estimations. In the light of this issue, a recently developed Bayesian Markov chain Monte Carlo (MCMC) method, which addresses multispecies coalescence, improves the precision and accuracy of inference of speciation timing (Heled & Drummond 2010).

In spite of variations in coalescence time, most divergence dating methods assume no gene flow during speciation (Heled & Drummond 2010; McCormack *et al.* 2011). If there is gene flow, the divergence between species will be more complex (Leaché *et al.* 2013). For genomic regions associated with reproductive incompatibility, early cessation of gene flow is likely. For regions free from such association, gene flow may continue until relatively late (Osada & Wu 2005). Therefore, a test of gene flow is important not only for divergence dating but also for understanding the mode of speciation. While allopatric speciation assumes no gene flow during divergence, speciation could occur if there is gene flow throughout the process (isolation-with-migration model) (Hey & Nielsen 2004). If we look closer at the timing of gene flow during the speciation process, it can occur at the early stage of divergence (early gene-flow model), or after a long period of allopatric divergence (secondary contact model) (Becquet & Przeworski 2009). In this case, although gene flow occurs, most of the genetic divergence accumulates unimpeded by gene flow in the allopatric stage.

Recently, an efficient and flexible method, approximate Bayesian computation (ABC), has been developed to provide a means to test alternative hypotheses about complex speciation histories using a likelihood search algorithm (Beaumont *et al.* 2002; Beaumont 2010). The ABC method, which provides statistical support for competing hypotheses, is now being widely applied to phylogeographic studies (Beaumont 2010; Bertorelle *et al.* 2010). By using the ABC model selection procedure, the underlying predominant speciation process between diverging sister taxa can be reconstructed confidently with statistical support.

An accurate estimation of divergence time is especially critical in testing hypotheses of diversification that involve different historical events at different time intervals. In Taiwan, a medium-sized island (36 000 $km^2$) located offshore from mainland Asia, orogenic activities and glacial cycles are two major factors affecting the tempo and mode of the speciation and diversification of its inhabitants. Despite a general consensus that mountain lifting fragmented species' ranges and promoted diversification (McKay *et al.* 2010; Lin *et al.* 2011; Wang *et al.* 2011), few detailed studies have directly linked speciation to these events. The East Asian grass lizard genus *Takydromus* Daudin, 1802 (Reptilia: Lacertidae) includes 22 species according to the Reptile Database (http://www.reptile-database.org/) and is widely distributed in the Oriental and eastern Palaearctic regions. Taiwan represents a hot spot of *Takydromus* species diversity, including seven species, of which six are endemic (Lin *et al.* 2002). The most parsimonious explanation is that their common ancestor colonized Taiwan several million years ago and multiple speciation events occurred in situ (Lin *et al.* 2002). Therefore, *Takydromus* may represent a good example to test the association of speciation events with the past geography of the island.

Among endemic *Takydromus* species, *T. viridipunctatus*, *T. luyeanus* and *T. hsuehshanensis* form a clade with reciprocally monophyletic mtDNA lineages (Lin *et al.* 2002; Lue & Lin 2008). *Takydromus viridipunctatus* and *T. luyeanus* are currently separated by the Li-Wu Stream, with a distinct and nonoverlapping distribution (Fig. 1). The presence of this barrier and their restricted ability for long-distance dispersal make it reasonable to suggest that their speciation occurred allopatrically, that is, there was no gene flow during their divergence ($m = 0$). Nevertheless, according to the mtDNA molecular clock provided by Lin *et al.* (2002), the divergence time between the two sibling species would be two million years ago (MYA) (Supporting Information), which was even earlier than the formation of the current habitats of *T. luyeanus* (Chen & Wang 1988) (Fig. 1). Thus, allopatric speciation due to the Li-Wu Stream may be an oversimplified scenario. *T. hsuehshanensis*, unlike its lowland relatives, is found only in the Central Mountain Range (>1800 m altitude) and is distributed allopatrically with *T. viridipunctatus* and *T. luyeanus* (Fig. 1) (Lin *et al.* 2002). Temperature changes with altitude could be the key factor limiting the distribution of *T. hsuehshanensis*, but this hypothesis has never been rigorously tested.

Due to the limited dispersal abilities of Lacertidae lizards (Díaz *et al.* 2000; Hurston *et al.* 2009), it has been suggested that geographical isolation may have played a major role in the divergence process of *Takydromus* species. Nevertheless, the assumption of allopatric speciation has never been rigorously tested. To study the speciation scenario of these *Takydromus* species, including their time of divergence, mode of speciation and historical population demography, we sequenced 20 nuclear loci and combined with the mitochondrial locus published in Lin *et al.* (2002). The species tree and species divergence times were jointly estimated using

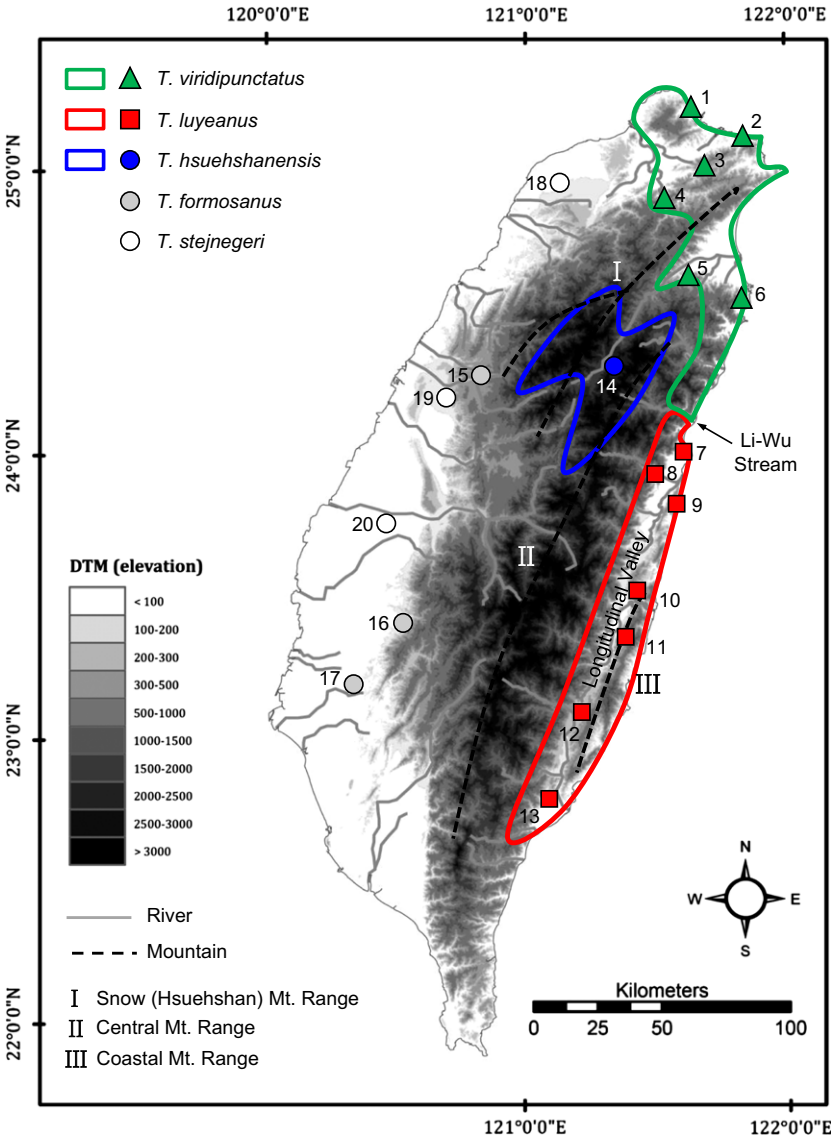


**Fig. 1** Geographical distribution and sampling localities of *Takydromus viridipunctatus* (triangle), *T. luyeanus* (square) and *T. hsuehshanensis* (filled circle), with *T. formosanus* (gray circle) and *T. stejnegeri* (circle) as out-groups. Sample locations: (1) Jinshan; (2) Badouzi; (3) Shiding; (4) Xindian; (5) Niudou; (6) Suao; (7) Qixingtan; (8) Hualien city; (9) Jian; (10) Ruisui; (11) Yuli; (12) Chishang (13) Taitung city; (14) Hehuan Mountain; (15) Dongshih; (16) Dapu; (17) Wushantou; (18) Douliu; (19) Yangmei; (20) Yuanlin.

*BEAST to compare species divergence dates to estimates based on a single gene. In addition, we conducted the ABC procedure to evaluate candidate speciation models/histories and estimate historical demography.

## Materials and methods

### *Sample collection*

A total of 88 grass lizards, including 23 *Takydromus viridipunctatus*, 38 *T. luyeanus*, 13 *T. hsuehshanensis* and 14 *T. formosanus*, were collected during 2009 and 2010 (Fig. 1). Nine *T. stejnegeri* were further included to serve as out-groups. Genomic DNA was extracted from tissues using the EasyPure genomic DNA spin kit (Bioman, Taiwan) according to the manufacturer's instructions and stored at −20 °C until further usage.

### *Amplification, sequencing and haplotype reconstruction*

Twenty nuclear loci, including four exons and 16 introns, were used in this study (Table S1, Supporting Information). A polymerase chain reaction (PCR) mixture was set up in a reaction volume of 20 μL using a GoTaq Flexi DNA polymerase system (Promega, Madison, USA) according to PCR conditions described in detail elsewhere (Molecular Ecology Resources Primer Development Consortium *et al.* 2012; MER database no. 48834 to 48905). All PCR products were sequenced in both directions by Genomics BioSci & Tech Corp. (Taipei, Taiwan) using an ABI3730 autosequencer.

Sequence data from both directions were assembled and checked using SEQUENCHER 4.9 (GeneCodes). Single nucleotide polymorphism (SNP) sites were detected using the 'call the secondary peaks' feature with 50%

threshold, meaning that the secondary peak was higher than the half-height of the primary peak. An SNP was carefully decided by checking signals from both directions in repetitive sequencing reactions and by further PCR-cloning reaction if necessary. Sequences were aligned by CLUSTALW (Thompson *et al.* 1994) implemented in MEGA 5 (Tamura *et al.* 2011) and visually inspected. For mitochondrial 12S rRNA (mt12S), the alignment was adopted from Lin *et al.* (2002). Haplotype reconstruction was performed by PHASE 2.1 (Stephens *et al.* 2001; Stephens & Scheet 2005) implemented in DNASP v5 (Librado & Rozas 2009) for each locus in each species. Each run was set up with MCMC iterations to 100 000 and thinned every 1000 intervals. Samples with phase probabilities lower than 0.6 (approximately 10%) were cloned and sequenced following the protocol provided by Mission Corp (Taipei, Taiwan). The sequences analysed in this study were submitted to GenBank (accession nos JQ746705–JQ747474 and JQ769109–JQ769112).

### *Population genetic analyses*

Population genetic parameters, including the number of segregating sites (S, Watterson 1975), nucleotide diversity, $\pi$/bp (Nei 1987), haplotype diversity, Hd (Nei 1987), and Watterson's estimator of theta per site, $\theta$w (Watterson 1975), were estimated by DNASP v5. This software was also used to perform Tajima's *D* (Tajima 1989) and Fu's *Fs* (Fu 1997) neutrality tests. *P*-values of the above tests were obtained using the implemented simulator with 1000 coalescent simulations assuming no recombination.

We conducted STRUCTURE 2.3.3 (Pritchard *et al.* 2000; Hubisz *et al.* 2009) analyses to check whether individuals cluster according to known species delimitations, to detect further population structure and to identify probable hybrids or migrants. Individuals were clustered based on nuclear haplotype data. The number of clusters (*K*) was evaluated from 1 to 5, and runs were conducted under the admixture model and allele frequencies correlated using an MCMC method with 500 000 iterations and an initial burn-in of 200 000 generations. One data set including three species with the 20 nuclear loci was carried out for 20 runs to assess the degree of variation of the likelihood for each *K*. The best *K* for each species was obtained based on the $\Delta K$ estimated by Evanno's method (Evanno *et al.* 2005).

### *Species-tree construction and divergence dating*

Species tree was estimated using Bayesian inference (*BEAST) implemented in the BEAST ver. 1.7.1 (Heled & Drummond 2010). *BEAST jointly estimates the species tree and all gene trees in one Bayesian MCMC analysis. Therefore, we can estimate TMRCAs and divergence time for each gene in a single run. All nuclear loci combined with mt12S sequences from Lin *et al.* (2002) were used for species-tree construction.

The analysis used an uncorrelated lognormal relaxed clock and a Yule process tree prior. Because *BEAST enables a calibration date on the root node and/or a prior on the substitution rate (Drummond & Rambaut 2007) for divergence dating, we used two approaches for the analysis. First, according to Lin *et al.* (2002), the divergence of five *Takydromus* species took place approximately 5 million years ago (MYA). A normal prior 95CI range from 3.896 to 5.856 MYA (prior mean 4.876 SD = 0.5) was applied to the root node. Second, in a separate analysis, we applied a normal prior 95CI range from 4.29 to $8.21 \times 10^{-3}$ (prior ucld.mean $6.25 \times 10^{-3}$ SD = 1.0) substitutions/site/MY for mt12S based on molecular clock derived from previous studies (Carranza & Arnold 2012; Fontanella *et al.* 2012). The substitution rates of nuclear loci were jointly estimated in both analyses. In each analysis, the MCMC method was run for a total of 300 million generations, sampled every 10 000 steps. Finally, 30 000 tree samples were used to generate a maximum clade credibility (MCC) tree with median node heights using TREEANNOTATOR v1.7.1 after a 10% burn-in.

Because the mt12S data were suspected to support a different topology from the nuclear data for reasons other than the stochasticity of the coalescent process (see Discussion), we also constructed species trees using mt12S and nuclear loci separately. While species-tree constructions are typically carried out using multilocus data, inference from a single locus is possible (Heled & Drummond 2010). Moreover, because species-tree constructions explicitly take the gene divergence within the ancestral population into account, it is preferable for estimating divergence times (Edwards & Beerli 2000). As the method for the whole data set, we separately set root date and the substitution rate prior to conduct the estimation of mt 12S data. For species trees constructed by nuclear loci only, substitution rate priors were adopted from the estimates of whole data set analyses. The root date prior was also applied in a separate analysis. All the rest settings were the same as in the whole data set analysis. For each analysis, MCMC was run for a total of 300 million generations, sampled every 10 000 steps. Finally, 30 000 tree samples were used to generate a MCC tree with median node heights using TreeAnnotator after a 10% burn-in (leaving 3000 trees).

### *$IM_A2$ inference*

Based on the relationships derived from species tree (see Results), the $IM_A2$ programme (Hey 2010) was

used to estimate the population genetic parameters of the three *Takydromus* species divergence process, including the effective population sizes of three descendants ($N_V$, $N_L$ and $N_H$), two ancestral populations ($N_{A\text{-}VL}$ and $N_{A\text{-}VLH}$), six directional migration rates among three focal species ($2NM_{VL}$, $2NM_{LV}$, $2NM_{VH}$, $2NM_{HV}$, $2NM_{LH}$ and $2NM_{HL}$), two directional migration rates between ancestral populations of *T. hsuehshanesis* ($2NM_{VL\text{-}H}$ and $2NM_{H\text{-}VL}$) and two divergence times ($T_{div\text{-}VL}$ and $T_{div\text{-}VLH}$). The $IM_A2$ programme assumes no recombination within loci, and gaps would be eliminated in $IM_A2$ inferences (Hey & Nielsen 2004). Prior to the $IM_A2$ analysis, all of the sequence data were trimmed using IMGC, a software that selects nonrecombining regions for IM analysis (Woerner *et al.* 2007). The likelihood-ratio test implemented in $IM_A2$ was used to test the null hypothesis of zero gene flow. The upper boundaries of the prior distributions for each parameter were set on the basis of the results of two preliminary runs for species pairs. Every $IM_A2$ run was carried out using a geometric heating scheme and searched with 25 chains. One genealogy was saved from every 100 steps, and the first 1 000 000 steps were discarded as burn-in. A mean generation time of 1 year was assumed for these *Takydromus* species (J.W. Lin, S.M. Lin, unpublished data; mark–recapture study). Three independent runs in different heating schemes and random start seeds were combined under $IM_A2$ L mode. Finally, 35 000 genealogies (runs of 35 million steps) were used in an $IM_A2$ analysis.

All parameters were scaled by the average substitution rate per gene (μ) as follows: $m = M/\mu$, $t = T^*\mu$, $\theta = 4N\mu$, where $M$ represents migration rate per generation per gene copy, $T$ represents divergence time estimated in years, and $N$ represents the effective population size. The substitution rates were derived from the whole data set calibrated species tree. The geometric means of the substitution rates were used to convert all of the parameters to absolute values.

### *Test of ancient vs. recent gene-flow hypotheses*

To investigate in more detail the divergence history of *T. viridipunctatus* and *T. luyeanus*, we used approximate Bayesian computation (ABC) to make inferences with complex models based on nuclear data. Four candidate scenarios are hypothesized: isolation, isolation-with-migration, early gene flow and secondary contact models (Fig. S1, Supporting Information). The isolation model assumes that *T. viridipunctatus* and *T. luyeanus* started diverging $T_{div}$ generations ago and that these two species accumulated divergence without any gene flow (i.e. allopatric divergence). The isolation with the gene-flow model describes the two species diverging $T_{div}$ generations ago with gene flow from *T. viridipunctatus* to *T. luyeanus* ($4NM_{VL}$) and from *T. luyeanus* to *T. viridipunctatus* ($4NM_{LV}$). The early gene-flow model assumes that they diverged in the presence of gene flow $T_{div}$ to $T_{m1}$ generations ago and subsequently diverged in absolute reproductive isolation (i.e. parapatric divergence). The secondary contact model assumes that the two species diverged $T_{div}$ generations ago and existed without gene flow until $T_{m2}$ generations ago, when they once again experienced gene flow. This scenario can be viewed as allopatric divergence followed by range expansion leading to secondary contact.

The details of the simulation are given in the Supporting Information. In brief, msABC (Pavlidis *et al.* 2011) was used to simulate four candidate scenarios and obtain summary statistics for nuclear data. One million data sets for each of the models were simulated. Posterior probabilities of four models were estimated through a weighted multinomial logistic regression method by 'calmod' (Beaumont 2006). Model selection was performed before estimating the final demographic historic parameters based on posterior probabilities of each model. For the best-supported model, an additional 3 million simulations were performed and combined with the previous one million simulations to estimate the parameters. The posterior probability and distribution of each parameter were obtained using ABCestimator as implemented in ABCtoolbox (Wegmann *et al.* 2010). The mode was chosen as the best point estimation for each parameter.

### *Historical population demography*

The historical population demographies of the three focal species were estimated separately by using loci information with extended Bayesian skyline plots (EBSPs) (Heled & Drummond 2008) implemented in BEAST. The HKY substitution model was used for all introns; while the same model with codon partitions (1+2), 3 was applied to the 4 exon markers. Estimations were carried out with stepwise population model under both strict and uncorrelated lognormal relaxed clock, and the final clock model was chosen using Bayes factor with the criteria of $\log_{10}$ Bayes factor = 3 (Kass & Raftery 1995). Each run was carried out with 300 million MCMC steps, logged every 20 000 steps, with the first 10% discarded as burn-in.

## Results

### *Genetic diversity of Takydromus grass lizards*

All loci were successfully amplified and sequenced from the three focal species except Taky1 from

**Table 1** Average number of chromosomes analysed (*N*), sequence length, number of segregating sites (*S*), haplotype diversity (Hd), nucleotide diversity per site ($\times 10^3$ $\pi$/bp), Watterson's estimator of theta per site ($\times 10^3$ $\theta$w), Tajima's *D* (Taj D) and Fu's *Fs* (FuFs) for three focal species

| Genome | Species | *N* | Length | *S* | Hd | $\pi$/bp | $\theta$w | TajD | FuFs |
|---|---|---|---|---|---|---|---|---|---|
| Nuclear | *T. viridipunctatus* | 20.6 | 892.6 | 7.8 | 0.66 | 2.17 | 2.60 | −0.441 | −0.635* |
| | *T. luyeanus* | 22.3 | 917.5 | 9.1 | 0.70 | 2.41 | 3.08 | −0.697** | −1.800*** |
| | *T. hsuehshanesis* | 18.8 | 898.6 | 5.9 | 0.56 | 2.64 | 2.17 | 0.562* | 1.368** |
| Mitochondria[†] | *T. viridipunctatus* | 14 | 1006 | 19 | 0.89 | 7.59 | 6.43 | 1.064 | 1.873 |
| | *T. luyeanus* | 13 | 1006 | 8 | 0.81 | 1.27 | 3.30 | −1.929* | −6.153*** |
| | *T. hsuehshanesis* | 9 | 1006 | 9 | 0.56 | 0.86 | 1.13 | −0.936 | 0.016 |

[†]Mitochondrial 12S rRNA gene adopted from Lin *et al.* 2002 and Lue & Lin 2008.
*$P < 0.05$; **$P < 10-2$; ***$P < 10^{-3}$; results of the two-tailed Wilcoxon rank-sum test for mean = 0.

*T. luyeanus* and Taky18 from *T. hsuehshanensis* (Table S1, Supporting Information). The lengths of aligned matrixes ranged between 372 and 1228 bp with means of 892.6, 917.5 and 898.6 bp for *T. viridipunctatus*, *T. luyeanus* and *T. hsuehshanensis*, respectively (Table 1). Interspecific indels occurred from mitochondrial 12S rRNA and 14 among the 20 nuclear loci. Three focal species exhibit comparable amount of genetic variability in terms of S, Hd, $\pi$/bp and $\theta$w for nuclear loci. However, for mt12S, *T. viridipunctatus* has genetic variability 2–3 times higher than *T. luyeanus* and *T. hsuehshanensis*. Great genetic variability of *T. viridipunctatus* mt12S is most likely due to population structure. Indeed, phylogeny derived by mt12S indicates that *T. viridipunctatus* from different localities are divided into two subclades, V1 and V2 (Fig. S2, Supporting Information), separated by Snow Mountain (Hsuehshan Mt.) range (Supporting Information).

The results of neutrality tests are also given in Table 1. For nuclear loci, the average Tajima's *D* and Fu's *Fs* are both negative for *T. viridipunctatus*, but only *Fs* (−0.635, $P = 0.02$; Wilcoxon signed-rank test) significantly deviates from zero. For *T. luyeanus*, both the averaged *D* (−0.697, $P < 10^{-2}$) and *Fs* (−1.800, $P < 10^{-3}$) are significantly negative. In contrast, the averaged *D* (0.562, $P = 0.03$) and *Fs* (1.368, $P < 10^{-2}$) are positive for *T. hsuehshanensis*. For mt12S, the *D* and *Fs* are positive for *T. viridipunctatus* possibly because of the existing geographical structure. Tajima's *D* (−1.929) and Fu's *Fs* (−6.153) are all significantly negative for *T. luyeanus*.

## *Dating the time of divergence between species*

The results of a Bayesian clustering analysis are shown in Table S2 (Supporting Information). The mean log-probabilities and $\Delta K$ are largest when $K = 3$, representing three focal species. Therefore, there is no sign of hybrid or misidentified individual (Fig. S3, Supporting Information).

Species trees constructed using root-dating prior and substitution rate prior yielded essentially identical results for all data sets; we therefore present the results from substitution rate prior and put those from root-dating prior in the supplement. For the whole data set, the divergence of *T. viridipunctatus* and *T. luyeanus* was at 0.81 (95% highest posterior density, HPD, 0.57–1.07) million years ago (MYA) (Fig. 2A and Fig. S4A, Supporting Information). The divergence of the three focal species began at 1.40 (95% HPD, 0.97–1.86) MYA. *Takydromus formosanus* is the out-group of the three species and diverging from them at 1.93 (95% HPD, 1.38–2.51) MYA. Species tree derived from mt12S (Fig. 2B and Fig. S4B) produced divergence time estimates considerably more ancient than those from the whole data set. *Takydromus viridipunctatus* and *T. luyeanus* diverged 2.07 MYA (95% HPD, 0.75–3.32). In addition, *T. formosanus* becomes the out-group of above two species instead of *T. hsuehshanensis* ($P < 10^{-3}$; KH-test (Kishino and Hasegawa 1989)). Divergence time estimates derived from the nuclear loci calibrated species tree are close to those from the whole data set with overlapped 95% HPDs (Fig. 2C and Fig. S4C, Supporting Information).

The TMRCAs and divergence time of each locus were jointly estimated with species trees and are shown in Table 2. For nuclear loci, the TMRCAs of *T. viridipunctatus* showed a great difference from the smallest RPL19 of 0.48 million years (MY) to the largest Taky23 of 4.16 MY, with an average of 0.98 MY and a standard deviation (SD) of 0.88. The TMRCAs of *T. luyeanus* showed little variation among loci, ranging from 0.46 MY (Taky20) to 1.27 MY (RPL19) with an average of 0.68 MY and SD of 0.22. The TMRCAs of *T. hsuehshanensis* ranged from 0.30 MY (Taky6) to 3.02 MY (BDNF) with an average of 0.68 MY and SD of 0.65.

The divergence times for *T. viridipunctatus* and *T. luyeanus* ranged from the smallest J7 of 1.04 MYA to the largest 5.08 MYA of Taky8 with an average of 1.91 MYA. The divergence times between the ancestral

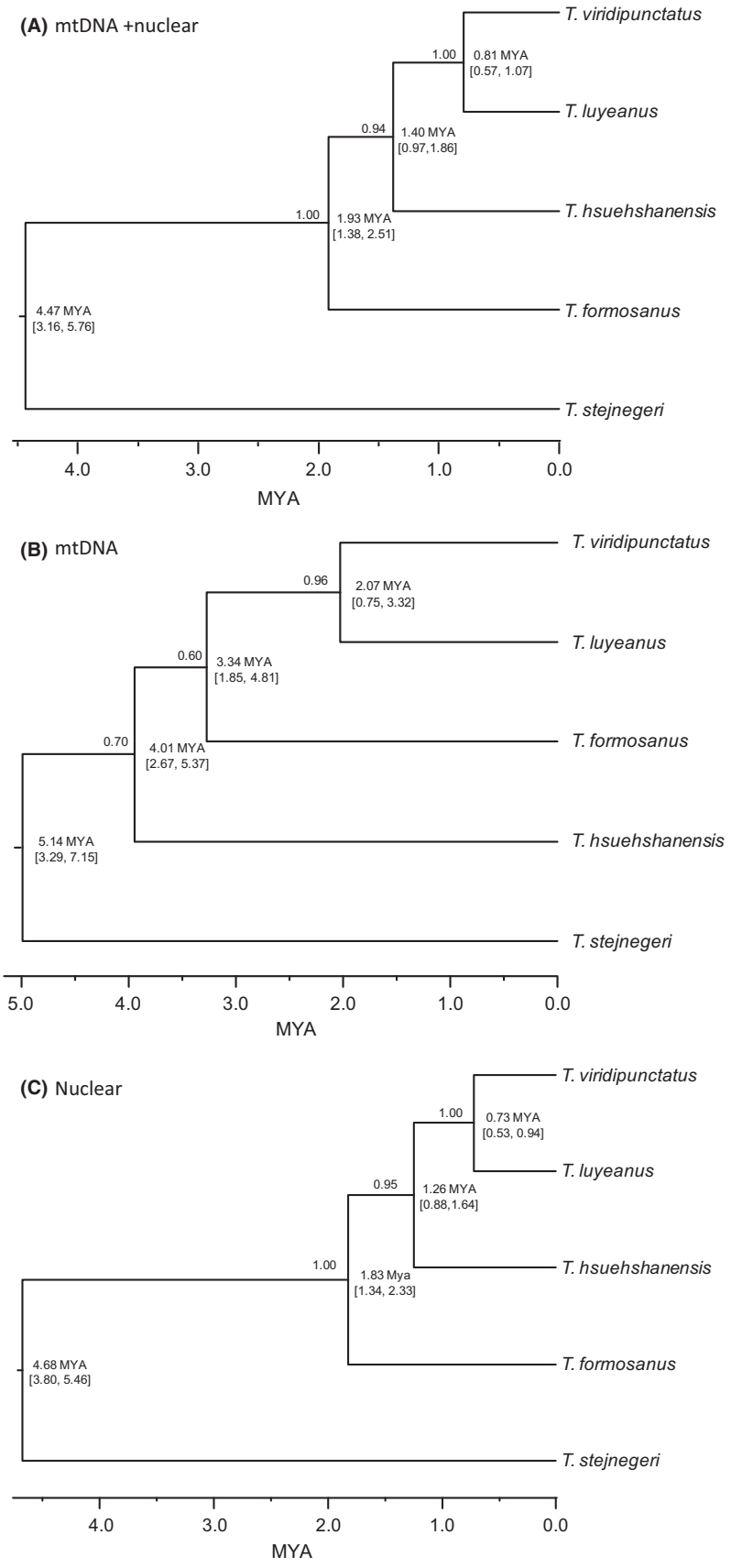


**Fig. 2** Species tree of *Takydromus* constructed by (A) mtDNA and 16 nuclear loci, (B) mtDNA only and (C) nuclear loci only using substitution rate prior. Divergence times are shown at the nodes with 95% highest posterior densities in brackets. Bayesian posterior probabilities are shown above the branches.

**Table 2** List of substitution rates ($\times 10^{-9}$ per site per year), time to the most recent common ancestor (TMRCA) and divergence times in millions of years, with 95% highest posterior density (95HPD) in parentheses

| | | | TMRCA (95HPD) | | | Divergence time (95HPD) | |
|---|---|---|---|---|---|---|---|
| | Prior rate | Posterior rate | *T. viridipunctatus* | *T. luyeanus* | *T. hsuehshanesis* | V v.s. L* | VL v.s. H† |
| BACH1 | 1.25 | 1.14 (0.77, 1.58) | 0.67 (0.33, 0.99) | 0.67 (0.41, 0.95) | 0.38 (0.06, 0.34) | 1.04 (0.64, 1.52) | 1.76 (1.11, 2.49) |
| BDNF | 0.99 | 0.83 (0.49, 1.25) | 0.61 (0.27, 0.96) | 0.63 (0.34, 0.95) | 3.02 (1.75, 4.57) | 1.11 (0.62, 1.72) | 3.02 (1.75, 4.57) |
| DDX1 | 4.93 | 5.12 (3.98, 6.20) | 0.60 (0.29, 0.95) | 0.60 (0.35, 0.85) | 0.47 (0.14, 0.85) | 1.50 (0.89, 2.21) | 8.24 (6.03, 10.63) |
| ETFB | 3.46 | 3.27 (2.28, 4.46) | 0.59 (0.25, 0.97) | 0.70 (0.42, 1.04) | 0.52 (0.10, 1.09) | 1.08 (0.64, 1.61) | 2.10 (1.26, 3.15) |
| FGB | 2.72 | 2.52 (1.88, 3.28) | 0.62 (0.32, 0.98) | 0.57 (0.31, 0.84) | 0.76 (0.26, 1.37) | 1.28 (0.74, 1.95) | 2.98 (1.87, 4.23) |
| J7 | 3.73 | 3.71 (2.76, 4.77) | 0.62 (0.32, 0.98) | 0.57 (0.31, 0.84) | 0.76 (0.26, 1.37) | 1.28 (0.74, 1.85) | 2.98 (1.87, 4.23) |
| PNN | 1.69 | 1.51 (0.98, 2.09) | 0.59 (0.25, 0.97) | 0.63 (0.30, 0.99) | 0.33 (0.06, 0.72) | 2.02 (1.26, 2.93) | 2.02 (1.26, 2.93) |
| R35 | 2.70 | 2.52 (1.74, 3.31) | 0.77 (0.34, 1.28) | 0.72 (0.40, 1.14) | 0.70 (0.19, 1.28) | 1.93 (1.28, 2.71) | 1.93 (1.28, 2.71) |
| RPL13 | 2.95 | 2.81 (1.89, 3.77) | 1.18 (0.76, 1.69) | 1.18 (0.76, 1.69) | 0.43 (0.09, 0.87) | 1.18 (0.76, 1.69) | 2.75 (1.74, 3.96) |
| RPL19 | 3.37 | 3.29 (2.29, 4.24) | 0.48 (0.21, 0.79) | 1.27 (0.74, 1.78) | 0.85 (0.26, 1.44) | 1.27 (0.74, 1.78) | 1.70 (1.07, 2.41) |
| RPS3 | 4.84 | 5.07 (3.81, 6.25) | 1.19 (0.75, 1.71) | 0.52 (0.30, 0.45) | 0.33 (0.08, 0.67) | 1.19 (0.75, 1.71) | 1.83 (1.25, 2.49) |
| Taky2 | 5.01 | 5.33 (3.85, 6.74) | 0.81 (0.43, 1.24) | 0.62 (0.37, 0.87) | 0.85 (0.33, 1.42) | 4.21 (2.73, 5.80) | 4.21 (2.73, 5.80) |
| Taky6 | 2.42 | 2.26 (1.52, 3.05) | 1.17 (0.71, 1.69) | 0.58 (0.32, 0.83) | 0.30 (0.06, 0.62) | 1.17 (0.71, 1.69) | 1.66 (1.05, 2.33) |
| Taky8 | 3.34 | 3.29 (2.32, 4.28) | 0.51 (0.21, 0.92) | 0.61 (0.34, 0.90) | 0.39 (0.13, 0.78) | 5.08 (3.35, 7.46) | 5.08 (3.35, 7.46) |
| Taky20 | 2.33 | 2.12 (1.38, 2.93) | 1.10 (0.66, 1.61) | 0.46 (0.20, 0.74) | 0.41 (0.12, 0.83) | 1.10 (0.66, 1.61) | 1.85 (1.12, 2.73) |
| Taky23 | 4.68 | 4.85 (3.80, 6.00) | 4.16 (2.75, 5.51) | 0.62 (0.40, 0.86) | 0.42 (0.13, 0.78) | 4.16 (2.75, 5.51) | 4.16 (2.75, 5.51) |
| Mean | 3.15 | 3.10 (1.42)‡ | 0.98 (0.88) | 0.68 (0.22) | 0.68 (0.65) | 1.91 (1.32) | 3.02 (1.74) |
| 12S§ | 6.25 | 6.12 (4.86, 7.20) | 1.06 (0.57, 1.59) | 0.32 (0.12, 0.57) | 0.23 (0.05, 0.45) | 2.43 (1.43, 3.47) | 4.32 (3.14, 5.63) |

*Divergence time between *T. viridipunctatus* and *T. luyeanus*.
†Divergence time between *T. hsuehshanesis* and common ancestor of *T. viridipunctatus* and *T. luyeanus*.
‡Standard deviation.
§Adopted from Lin *et al.* (2002) and Lue & Lin (2008).

populations of *T. viridipunctatus* + *T. luyeanus* and *T. hsuehshanensis* ranged from the smallest Taky6 of 1.66 MYA to the largest 8.24 MYA of DDX1, with an average of 3.02 MYA. For mt12S, the TMRCAs for *T. viridipunctatus*, *T. luyeanus* and *T. hsuehshanensis* were 1.06, 0.32 and 0.23 MYA, respectively. The divergence time between *T. viridipunctatus* and *T. luyeanus* was 2.43 MYA. The divergence of the three species began 4.32 MYA.

### *Speciation with gene flow*

$IM_A2$ analysis could be applied to estimate parameters associated with two time stages: before the speciation of *T. viridipunctatus* and *T. luyeanus* (early stage) and after this event (late stage). The speciation of *T. hsuehshanensis* occurred approximately 1.31 MYA, and the speciation of *T. viridipunctatus* and *T. luyeanus* took place approximately 0.85 MYA (Fig. 3). In the early stage, migration was detected from the ancestral population of *T. viridipunctatus* and *T. luyeanus* ($ANC_{VL}$) to *T. hsuehshanensis*, with a magnitude of 0.181 ($P < 0.05$) in the *2NM* term. In the late stage, migration was detected from *T. viridipunctatus* to *T. luyeanus* (0.016, $P < 0.05$) and from *T. viridipunctatus* to *T. hsuehshanensis* (0.015, $P < 10^{-2}$). The estimated population parameters, including effective population size, time of divergence and gene flow, are given with their distributions in Table S3 and Fig. S5 (Supporting Information). The effective population size of *T. hsuehshanensis* (105 640) was smaller than that of *T. viridipunctatus* (201 422) and *T. luyeanus* (259 735). The population size of the common ancestor of these three species was 944 880.

### *Evolutionary scenario for T. viridipunctatus and T. luyeanus*

$IM_A2$ suggested the existence of gene flow during the formation of the three focal species. However, migration could have occurred at different stages of speciation. To reveal the divergence history, four models were proposed and simulated by msABC, that is, isolation (I), isolation with migration (IM), early gene flow (E) and secondary contact (SEC) (Fig. S1, Supporting Information; see Materials and Methods). Because too many parameters need to be considered in the three-species estimation, we focused only on *T. viridipunctatus* and *T. luyeanus* in this analysis. For each model, 1 million simulations were conducted by msABC, based on the observed 38 summary statistics listed in Table S4 (Supporting Information). The prior settings were shown in Table S5 (Supporting Information). Among the four

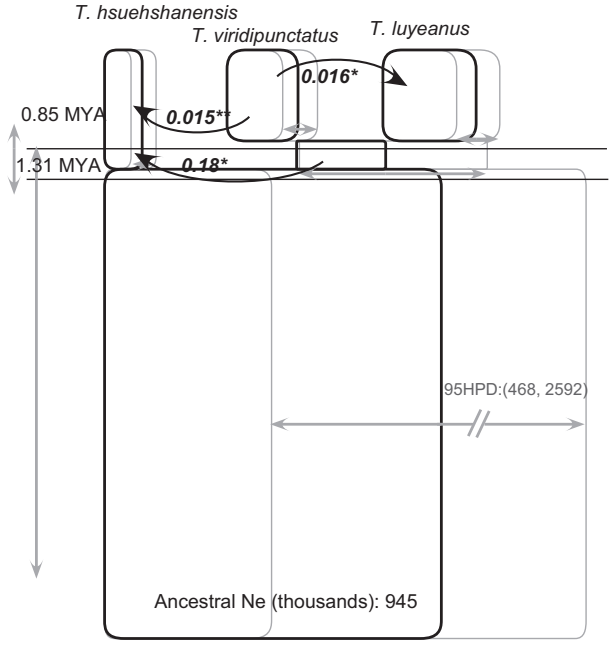


**Fig. 3** The speciation scenario of *Takydromus* simulated by 20 nuclear loci using $IM_A2$. Each species is represented by a box, the height of which refers to how long it has lasted and the width to its effective population size. Grey double-headed arrows denote 95% highest posterior density intervals. Migration arrows represent the population migration rate (2NM) from the source population to the receiving population. Only statistically significant migration rates according to a likelihood-ratio test (Nielsen & Wakeley 2001) are shown. *: $P < 0.05$; **: $P < 0.01$.

models, the posterior probabilities of the early gene-flow model are >0.99 under different tolerance rates (Table 3). The Bayes factors between early gene-flow and the other models are $6 \times 10^3$, $7 \times 10^6$ and $4 \times 10^5$ for $B_{E/I}$, $B_{E/IM}$ and $B_{E/SEC}$, respectively.

To test the effect of incorrect haplotype reconstruction, we removed the linkage-associated summary statistics (*i.e.* $Z_{ns}$) and repeated the model selection procedure. The early gene-flow model was overwhelmingly preferred over other models, with Bayes factors of $4 \times 10^3$, $2 \times 10^6$ and $2 \times 10^5$ for $B_{E/I}$, $B_{E/IM}$ and $B_{E/SEC}$, respectively (detail not shown). In conclusion, the early gene-flow model is strongly supported over the other evolutionary scenarios in regard to different tolerance rates and choices of summary statistics.

To estimate their speciation history in detail, an additional three million simulations under the early gene-flow model were generated by msABC and combined with the previous one million simulations. The divergence of *T. viridipunctatus* and *T. luyeanus* ($T_{div}$) took place approximately 1.83 (95% HPD 0.28–7.76) MYA with continuous migrations between them ($T_{m1}$) until 0.69 (95% HPD 0.21–1.09) MYA (Table 4, Fig. 4). During this period, the average migrants per generation from *T. viridipunctatus* to *T. luyeanus* ($4NM_{VL}$) and *vice versa* ($4NM_{LV}$) were approximately 20.5 (95% HPD 2.4–29.5) and 11.6 (95% HPD 0.3–27.3), respectively. The effective population sizes of *T. viridipunctatus* ($N_V$) and *T. luyeanus* ($N_L$) were similar, with 236 181 for the former and

**Table 3** Posterior probabilities of different demographic models under different tolerance rates

| | Tolerance rate | | | |
|---|---|---|---|---|
| | 0.001 | 0.005* | 0.01 | 0.05 |
| Isolation | $5.64 \times 10^{-5}$ | $1.63 \times 10^{-4}$ | $6.74 \times 10^{-5}$ | $1.92 \times 10^{-5}$ |
| Isolation with migration | $2.67 \times 10^{-8}$ | $1.40 \times 10^{-7}$ | $2.08 \times 10^{-7}$ | $3.96 \times 10^{-9}$ |
| Early gene flow | 0.999944 | 0.999835 | 0.999932 | 0.999981 |
| Secondary contact | $7 \times 10^{-8}$ | $2.23 \times 10^{-6}$ | $3.18 \times 10^{-7}$ | $8.67 \times 10^{-10}$ |

*Threshold used for Bayes factor calculations.

**Table 4** Parameter estimates in approximate Bayesian computation under the early gene-flow model

| | $N_V$ | $N_L$ | $N_A$ | $4NM_{VL}$ | $4NM_{LV}$ | $T_{div}$ | $T_{m1}$ |
|---|---|---|---|---|---|---|---|
| Mode | 236 181 | 256 282 | 809 045 | 20.5 | 11.6 | 1 825 780 | 687 840 |
| HPD95Lo | 153 268 | 185 932 | 175 883 | 2.4 | 0.3 | 281 843 | 202 270 |
| HPD95Hi | 316 581 | 329 144 | 999 997 | 29.5 | 27.3 | 7 755 160 | 1 093 056 |

Mode: posterior mode in the estimation; HPD95Lo: the lower bound of the estimated 95% highest posterior density (HPD) interval; HPD95Hi: the upper bound of the estimated 95% HPD interval; $N_V$ and $N_L$: long-term equilibrium effective population sizes of *T. viridipunctatus* and *T. luyeanus*, respectively; $N_A$: effective population size of ancestral species; $4NM_{VL}$ and $4NM_{LV}$: population migration rates per generation from *T. viridipunctatus* to *T. luyeanus* and from *T. luyeanus* to *T. viridipunctatus* forward in time, respectively; $T_{div}$: divergence time; $T_{m1}$: migration stop time.

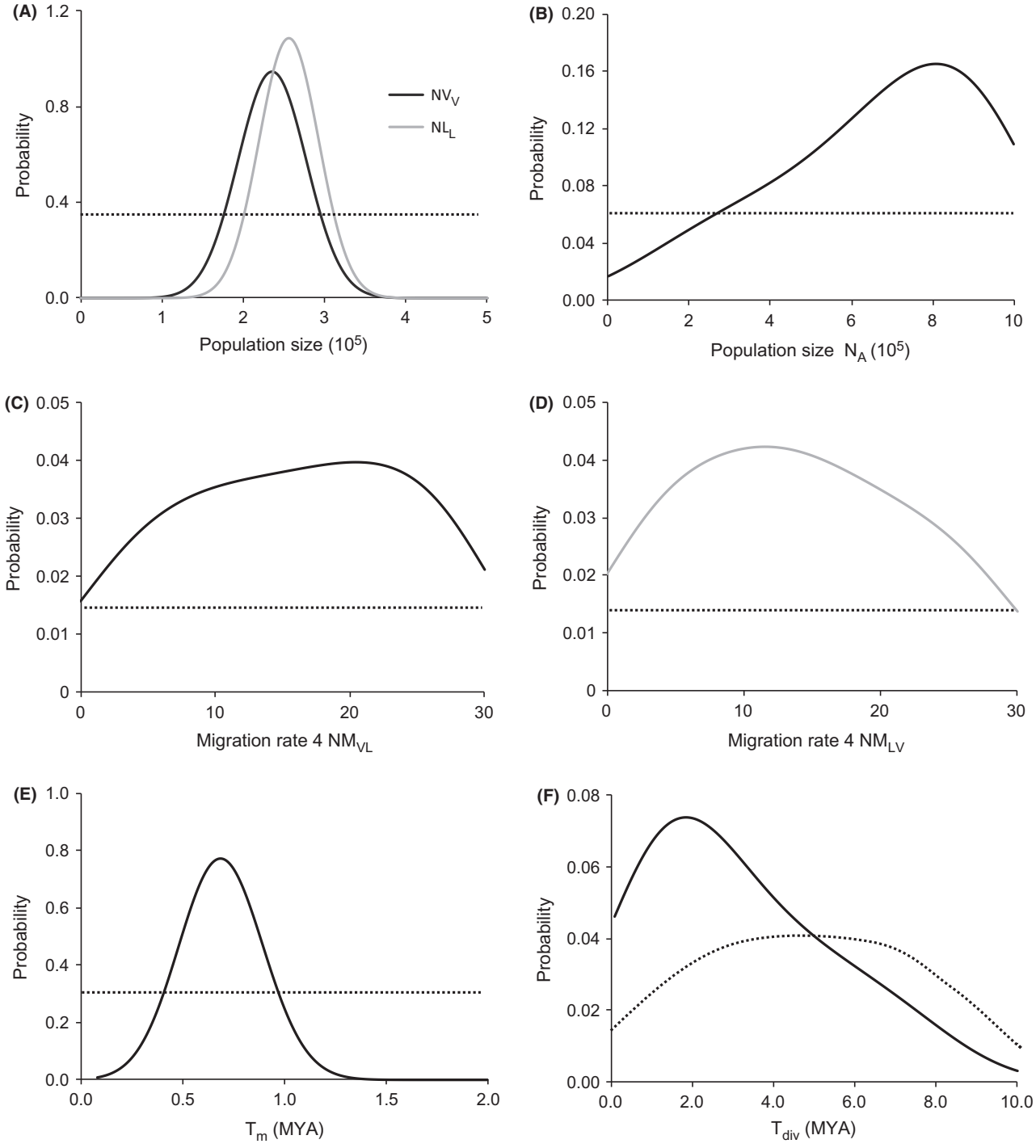


**Fig. 4** Prior (dotted lines) and posterior (continuous lines) distributions of the parameters involved in the early gene-flow model, with population sizes (A and B), migration rates in individuals (C and D) and times in years ago (E and F). $N_V$ and $N_L$: long-term equilibrium effective population size of *T. viridipunctatus* and *T. luyeanus*, respectively; $N_A$: effective population size of ancestral species; $4NM_{VL}$ and $4NM_{LV}$: population migration rate per generation from *T. viridipunctatus* to *T. luyeanus* and from *T. luyeanus* to *T. viridipunctatus* forward in time, respectively; $T_{div}$: divergence time; $T_{m1}$: migration stop time.

256 282 for the latter, while the effective population size of the entire ancestral population ($N_A$) was about four times the current populations ($N_A$ = 809 045).

### *Population demography*

Significant deviation from zero in neutrality tests may indicate the influence of historical population

demography. While negative Tajima's *D* and Fu's *Fs* suggest a recent population expansion of *T. luyeanus*, positive *D* and *Fs* imply a population shrinkage scenario of *T. hsuehshanensis*. We next used the Bayesian approach to infer the historical population demography of the three species. According to an extended Bayesian skyline plot (EBSP), the population size of *T. viridipunctatus* remained constant over time (Fig. 5A). A tenfold population increase beginning at 0.35 MYA was found in *T. luyeanus* (Fig. 5B). In *T. hsuehshanensis*, a population decrease after 0.10 MYA was observed (Fig. 5C). In summary, two lines of evidence, the frequency spectrum and EBSP, support a recent population expansion of *T. luyeanus* and reduction in *T. hsuehshanensis*.

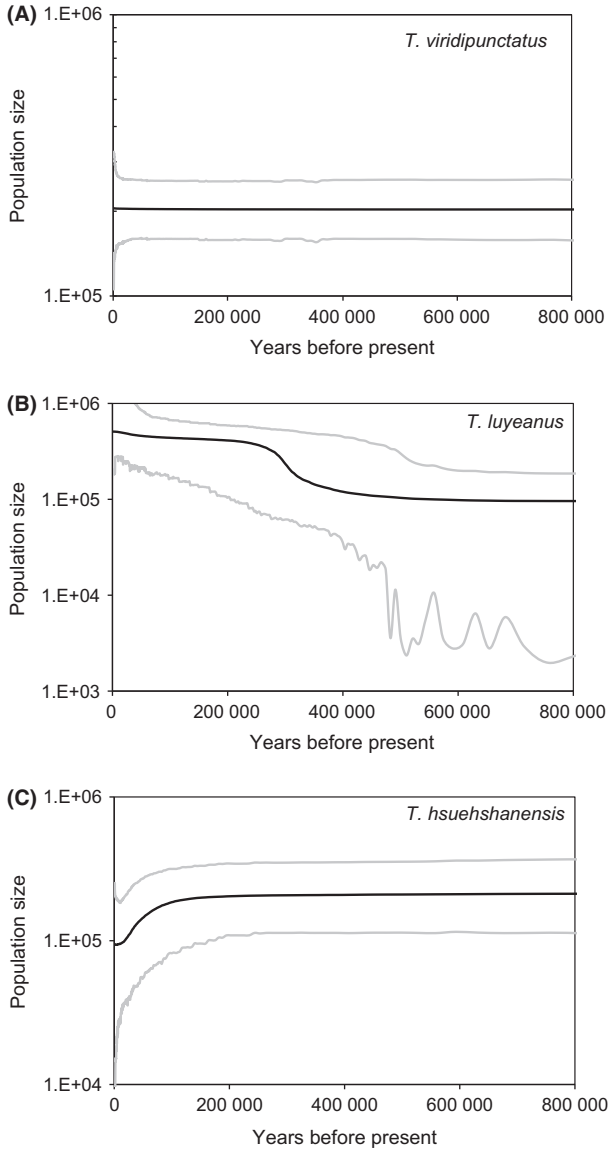


**Fig. 5** Historical demography of *Takydromus viridipunctatus* (A), *T. luyeanus* (B) and *T. hsuehshanesis* (C) inferred from nuclear loci by an extended Bayesian skyline plot. The black line represents median estimates, and grey lines represent 95% highest posterior densities.

## Discussion

### *Gene flow at an early stage of divergence accounts for variations in divergence dating among genes*

The incongruence across different markers might be due to variation in gene divergence time (*t*) or variation in coalescence time, which has been discussed in literature (Osada & Wu 2005). Because genes may vary in *Ne*, the deduced divergence times may differ even though they have a similar divergence history. Thus, the variation in *Ne* combined with stochasticity in the coalescent process might account for the differences in divergence times derived from individual genes vs. species-tree approaches. However, this explanation is based on the assumption that there is no gene flow during the speciation process (Knowles 2009; Heled & Drummond 2010), that is, fixed *t* across loci. If this assumption is violated, *t* will not be constant because certain genes are able to diffuse at an early stage of speciation, and the initial divergence time may be underestimated.

Both the $IM_A2$ and msABC inferences suggest that there was gene flow during the speciation of *Takydromus* species, even though the degree of migration estimated by different approaches varies substantially. Therefore, variation in *Ne* cannot completely account for the differences derived from the different markers and methods. Due to their limited dispersal ability, geographical isolation may have played a major role in the divergence process. Overall, our results reveal a more complex situation than the simplified allopatric speciation model. This study also strengthened the recent aspect (e.g. Heled *et al.* 2013; Leaché *et al.* 2013) that gene flow during the speciation process cannot be overlooked and may have great impact on divergence dating.

### *Speciation history of Takydromus in Taiwan*

We propose an evolutionary history of three *Takydromus* species in Taiwan based on the msABC simulations and our knowledge of past geography. The simulated divergence of *T. viridipunctatus* and *T. luyeanus* was approximately 1.8 MYA (Table 4), which is close to those derived from mt12S species tree (Fig. 2B). Although the separation of *T. hsuehshanensis* was not simulated, judging by gene and species divergences, it should have occurred earlier than the divergence of *T. viridipunctatus* and *T. luyeanus*. Therefore, the speciation of

*T. hsuehshanensis* may have been near the time of the divergence of its mt12S (4.32 MYA, Table 2; 4.01 MYA, Fig. 2B), which was during the uplift of Snow Mountain (Hsuehshan Mt.), about 3–5 MYA (Chen *et al.* 2000). It is reasonable to propose that the split of *T. hsuehshanensis* was initiated by the formation of Snow Mountain. During the early stages of their speciation, a substantial amount of gene flow was detected. This scenario is similar to the parapatric mode of speciation. Under this condition, the divergence between the genomes of the two species would be in a mosaic pattern (Wu 2001) because regions corresponding to the speciation process would be differentiated, whereas the rest of the genome would not. Thus, different genomic regions would have different levels of divergence and TMRCA. For example, the TMRCA of *T. viridipunctatus* ranged from 0.48 to 4.16 MY, suggesting that different parts of the genome have different histories of divergence.

According to the results of the simulation, migration between *T. viridipunctatus* and *T. luyeanus* stopped approximately 0.69 MYA, which is similar to the estimated divergence time by the species-tree approach (Fig. 2A and C). Because the species-tree method assumes no gene flow during speciation, divergence dating based on this approach should be close to the time when the two diverging groups were completely isolated (Heled *et al.* 2013). That period also corresponds to the formation of southeastern Taiwan, including the Coastal Range and Longitudinal Valley (Chen & Wang 1988), the current habitats of *T. luyeanus*. We proposed that *T. luyeanus* colonized these areas right after they formed, and the two species have been separated ever since by the Li-Wu Stream. Although the stream may not have been an effective barrier during the early stage of speciation, it creates a sharp division in their current distributions and may have contributed to the final step of isolation.

Although $IM_A2$ and msABC both support gene flow during speciation, they yielded very different divergence dates, most likely due to different amounts of estimated gene flow. Because the IM model may oversimplify the real situation (assuming that the migration rate is constant during divergence) and the time frame simulated by msABC is synchronous with known geographical changes, we suggest that the latter may be a more probable scenario.

### *Mitochondrial vs. nuclear divergence*

Although the branching pattern for the mitochondrial phylogeny was not highly supported, species trees derived from mt12S and nuclear loci showed substantial difference in divergence dating and species phylogenies. There are several possible explanations for this incongruence between mt12S and nuclear loci, including incomplete lineage sorting, homoplasy or saturation of substitutions (Felsenstein 1978; Maddison & Knowles 2006; Brandley *et al.* 2011, 2012). Except for these well-documented sources of conflict, an alternative explanation is due to differences in divergence history among mitochondrial and most of the nuclear loci. This phenomenon might be due to male-biased dispersal, which is a common feature in many lizards (Doughty *et al.* 1994; Olsson *et al.* 1996; Massot *et al.* 2003). Indirect evidence from Table 2 also suggests asymmetrical sex dispersal. Because the effective population size of mitochondrial genomes is only one-fourth as large as nuclear genes, the TMRCA estimate from the former should be shorter than the latter. Indeed, the TMRCAs of mt12S for *T. luyeanus* (0.32 MY) and *T. hsuehshanensis* (0.23 MY) are less than half of their nuclear averages. Nevertheless, due to population structure (Fig. S2, Supporting Information), the TMRCA of mt12S for *T. viridipunctatus* (1.06 MY) is more ancient than the average TMRCA derived from nuclear loci (0.98 MY). While the mitochondrial gene shows great population structure, only a small amount of differentiation is observed in the nuclear genome of *T. viridipunctatus* (Fst = 0.79 vs. 0.08; see Supporting Information). This observation implies unequal gene flow in the two genomes, as expected under male-biased dispersal. In addition to asymmetrical sex dispersal, postzygotic isolation caused by mitochondrial-nuclear incompatibility may also have contributed to degrees of divergence in mitochondrial vs. nuclear genomes. Mitochondria are the main source of cellular energy production (mitochondrial bioenergetics) and play a major role in cell metabolism and whole organism development. Mitochondrial- and nuclear-encoded proteins interact closely with each other in the electronic transporter complex (ETC.), possibly leading to intergenomic co-adaptation (Rand *et al.* 2004; Gershoni *et al.* 2009). Living at high altitudes would require *T. hsuehshanensis* to increase its mitochondrial aerobic energy production, as has been demonstrated for both vertebrates (Portner 2004) and invertebrates (Sommer & Pörtner 2002). This process would promote the divergence of the mitochondrial genome. It is possible that the divergence of these species approximately 4 MYA was initiated by their mitochondrial divergence, which may have subsequently contributed to the divergence of their nuclear counterparts. Because only mitochondria-associated genes were differentiated, gene flow in other parts of the genome was still possible. As a result, we observed a mosaic pattern of divergence in their genomes. This hypothesis can be tested by studying genes involved in ETC. and randomly selected loci. The former category should have a more ancient history of divergence than the latter.

*Historical population demography*

The genetic variation among these *Takydromus* species also suggests the influence of geography and climatic change on their population demography. The EBSP analyses suggest that the population of *T. luyeanus* has increased in size since the Middle Pleistocene (0.50–0.40 MYA), which was close to the time its current habitat was formed. Thus, the colonization of a new territory may have promoted its population expansion. In contrast to *T. luyeanus*, the population of *T. hsuehshanensis* has been decreasing since the last glacial epoch, which began 110 kilo years before present (YBP), reached a maximum decrease at 24–18 kilo YBP (last glacial maxima; LGM) and ended approximately 10 kilo YBP (Tsukada 1966; Liew & Chung 2004). Population reduction became more severe as temperatures decreased towards the LGM when the average temperature was approximately 6–7 °C lower than at present (Tsukada 1967).

Many studies have attributed population size changes to the effects of glacial-interglacial cycles in Taiwan (Lin *et al.* 2008, 2011; Huang & Lin 2011). In our case, temperature fluctuation did not seem to play an important role in the population demography of the two lowland lizards. For *T. luyeanus*, low temperatures in the last glacial period might have only slightly retarded the rate of population growth, but the overall trend has been towards positive growth. For *T. viridipunctatus*, the population size remained constant during the last 0.8 MY. On the other hand, lower temperatures may have contributed to a population decrease for *T. hsuehshanensis* living at high altitude.

## Conclusion

This study provides the first comprehensive assessment of the speciation process and divergence history of *Takydromus* in East Asia. Due to their limited dispersal ability, it has been suggested that geographical isolation may have played a major role in the divergence process in these species. Nevertheless, the current study reveals a far more complex scenario than a simple allopatric speciation model. Divergence dating based on individual genes revealed an earlier split than those based on the species-tree method. In contrast to previous assumptions, the simulation results suggest that the stochasticity of the coalescent process during speciation cannot completely account for this discrepancy. High variations in divergence among loci may be a result of migration between diverging species and suggest a parapatric mode of speciation.

The divergence of three focal *Takydromus* species was initiated at least 4 MYA, most likely resulting from the rise of Snow Mountain (Hsuehshan). Between *T. viridipunctatus* and *T. luyeanus*, gene flow persisted during three-fifths of their divergence and did not cease until 0.69 MYA, coincident with the formation of the current range of *T. luyeanus*, which includes the Coastal Range and Longitudinal Valley (Fig. 1). After *T. luyeanus* moved into these areas, the two species were separated by the Li-Wu Stream, which prohibited gene flow between them. The population size of *T. viridipunctatus* remained constant over the last 0.8 MY, while a demographic expansion of *T. luyeanus* beginning between 0.40 and 0.50 MYA was revealed, which was most likely promoted by its colonization of new habitats. We observed a population reduction in *T. hsuehshanensis* beginning at 0.1 MYA, which coincides with the beginning of the last glacial epoch. Therefore, temperature fluctuation may greatly influence the populations of this highland lizard.

## Acknowledgements

We wish to express our sincere thanks to Dr. Pavlos Pavlidis for providing vital advice on the ABC analysis. We also thank Prof. Chung-Ping Lin for access to his high-performance computation facility. Some of the samples were collected with the assistance of Yen-Po Lin, Chung-Wei You and Jen-Chieh Wang. This work was supported by the National Science Council, Taiwan (NSC- 99-3112-B-002-048), to H.-Y.W. and (NSC 97-2621-B-003-007-MY3) to S.M.L.

## References

Beaumont MA (2006) *Joint Determination of Topology, Divergence Time, and Immigration in Population Trees*. McDonald Institute for Archaeological Research, Cambridge.

Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.

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

Becquet C, Przeworski M (2009) Learning about modes of speciation by computational approaches. *Evolution*, **63**, 2547–2562.

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

Brandley MC, Wang Y, Guo X *et al.* (2011) Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of *Plestiodon (Eumeces)* lizards. *Systematic Biology*, **60**, 3–15.

Brandley MC, Ota H, Hikida T *et al.* (2012) The phylogenetic systematics of blue-tailed skinks (*Plestiodon*) and the family Scincidae. *Zoological Journal of the Linnean Society*, **165**, 163–189.

Carranza S, Arnold EN (2012) A review of the geckos of the genus *Hemidactylus* (Squamata: Gekkonidae) from Oman based on morphology, mitochondrial and nuclear data, with descriptions of eight new species. *Zootaxa*, **3378**, 1–95.

Carstens BC, Knowles LL (2007) Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology*, **16**, 619–627.

Chen WS, Wang Y (1988) Development of deep-sea fan systems in Coastal Range basin, Eastern Taiwan. *Science Reports of the National Taiwan University ACTA Geologica Taiwanica*, **26**, 37–56.

Chen WS, Erh CH, Chen MM *et al.* (2000) The evolution of foreland basins in the western Taiwan: evidence from the Plio-Pleistocene sequences. *Bulletin of the Central Geological Survey*, **13**, 137–156.

Díaz JA, Carbonell R, Virgós E, Santos T, Tellería JL (2000) Effects of forest fragmentation on the distribution of the lizard *Psammodromus algirus*. *Animal Conservation*, **3**, 235–240.

Doughty P, Sinervo B, Burghardt GM (1994) Sex-biased dispersal in a polygynous lizard, *Uta stansburiana*. *Animal Behaviour*, **47**, 227–229.

Drummond A, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Edwards S, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, **27**, 401–410.

Fontanella FM, Olave M, Avila LJ, Morando M (2012) Molecular dating and diversification of the South American lizard genus *Liolaemus* (subgenus *Eulaemus*) based on nuclear and mitochondrial DNA sequences. *Zoological Journal of the Linnean Society*, **164**, 825–835.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Gershoni M, Templeton AR, Mishmar D (2009) Mitochondrial bioenergetics as a major motive force of speciation. *BioEssays*, **31**, 642–650.

Heled J, Drummond A (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570–580.

Heled J, Bryant D, Drummond AJ (2013) Simulating gene trees under the multispecies coalescent and time-dependent migration. *BMC Evolutionary Biology*, **13**, 44.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Huang JP, Lin CP (2011) Lineage-specific late pleistocene expansion of an endemic subtropical gossamer-wing damselfly, *Euphaea formosa*, in Taiwan. *BMC Evolutionary Biology*, **11**, 94.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Hurston H, Voith L, Bonanno J *et al.* (2009) Effects of fragmentation on genetic diversity in island populations of the Aegean wall lizard *Podarcis erhardii* (Lacertidae, Reptilia). *Molecular Phylogenetics and Evolution*, **52**, 395–405.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of molecular evolution*, **29**, 170–179.

Knowles LL (2009) Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. *Systematic Biology*, **58**, 463–467.

Leaché AD, Harris RB, Rannala B, Yang Z (2013) The influence of gene flow on species tree estimation: a simulation study. *Systematic Biology*, **63**, 17–30.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Liew PM, Chung NJ (2004) Vertical migration of forests during the last period in subtropical Taiwan. *Western Pacific Earth Sciences*, **1**, 405–414.

Lin SM, Chen CA, Lue KY (2002) Molecular phylogeny and biogeography of the grass lizards genus *Takydromus* (Reptilia: Lacertidae) of East Asia. *Molecular Phylogenetics and Evolution*, **22**, 276–288.

Lin RC, Yeung CK, Li SH (2008) Drastic post-LGM expansion and lack of historical genetic structure of a subtropical fig-pollinating wasp (*Ceratosolen* sp. 1) of *Ficus septica* in Taiwan. *Molecular Ecology*, **17**, 5008–5022.

Lin HD, Chen YR, Lin SM (2011) Strict consistency between genetic and topographic landscapes of the brown tree frog (*Buergeria robusta*) in Taiwan. *Molecular Phylogenetics and Evolution*, **62**, 251–262.

Lue KY, Lin SM (2008) Two new cryptic species of *Takydromus* (Squamata: Lacertidae) from Taiwan. *Herpetologica*, **64**, 379–395.

Maddison WP, Knowles L (2006) Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, **55**, 21–30.

Molecular Ecology Resources Primer Development Consortium, Abelló P, Ai W *et al.* (2012) Permanent genetic resources added to molecular ecology resources database 1 April 2012 – 31 May 2012. *Molecular Ecology Resources*, **12**, 972–974.

Massot M, Huey RB, Tsuji J, van Berkum FH (2003) Genetic, prenatal, and postnatal correlates of dispersal in hatchling fence lizards (*Sceloporus occidentalis*). *Behavioral Ecology*, **14**, 650–655.

McCormack JE, Heled J, Delaney KS, Peterson AT, Knowles LL (2011) Calibrating divergence times on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. *Evolution*, **65**, 184–202.

McKay B, Mays H, Peng YW *et al.* (2010) Recent range-wide demographic expansion in a Taiwan endemic montane bird, Steere's Liocichla (*Liocichla steerii*). *BMC Evolutionary Biology*, **10**, 71.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University, New York.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Olsson M, Gullberg A, Tegelström H (1996) Malformed offspring, sibling matings, and selection against inbreeding in

the sand lizard (*Lacerta agilis*). *Journal of Evolutionary Biology*, **9**, 229–242.
Osada N, Wu CI (2005) Inferring the mode of speciation from genomic data: a study of the great apes. *Genetics*, **169**, 259–264.
Pavlidis P, Laurent S, Stephan W (2011) msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis. *Molecular Ecology Resources*, **10**, 723–727.
Portner HO (2004) Climate variability and the energetic pathways of evolution: the origin of endothermy in mammals and birds. *Physiological and Biochemical Zoology*, **77**, 959–981.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
Rand DM, Haney RA, Fry AJ (2004) Cytonuclear coevolution: the genomics of cooperation. *Trends in Ecology and Evolution*, **19**, 645–653.
Sommer A, Pörtner HO (2002) Metabolic cold adaptation in the lugworm *Arenicola marina*: comparison of a White Sea and a North Sea population. *Marine Ecology-Progress Series*, **240**, 171–182.
Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, **76**, 449–462.
Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.
Tsukada M (1966) Late Pleistocene vegetation and climate in Taiwan (Formosa). *Proceedings of the National Academy of Sciences USA*, **55**, 543–584.
Tsukada M (1967) Vegetation in subtropical Formosa during the Pleistocene glaciations and the Holocene. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **3**, 49–64.
Wang CF, Hsieh CH, Lee SC, Wang HY (2011) Systematics and phylogeography of the Taiwanese endemic minnow *Candidia barbatus* (Pisces: Cyprinidae) based on DNA sequence, allozymic, and morphological analyses. *Zoological Journal of the Linnean Society*, **161**, 613–632.
Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.
Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, **11**, 1–7.
Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851–1853.
Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

S.M.L. and S.H.L. originally formulated the idea. S.M.L. designed the research, applied the grant and organized the programme. S.H.L. took charge of molecular facilities and techniques. S.P.T. performed the experiment, including experimental design, field collection, molecular works and data analyses. C.H.H. helped data analyses; H.Y.W. carefully checked the data and finalized the analyses. S.M.L., S.P.T. and H.Y.W. wrote the manuscript.

## Data accessibility

DNA sequences: GenBank accession numbers KM487140-KM487188, JQ746705–JQ747474 and JQ769109–JQ769112. Sequence alignments used in this study and tree files are deposited at DRYAD entry doi: 10.5061/dryad.4p7n5.

## Supporting information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Four demographic models simulated by msABC.

**Fig. S2** Bayesian phylogeny tree of *Takydromus* lizards inferred from mitochondrial 12S rRNA sequence data.

**Fig. S3** Population structure of *Takydromus* species derived from Structure analyses.

**Fig. S4** Species tree of *Takydromus* constructed by (A) mtDNA and 16 nuclear loci (B) mtDNA only (C) nuclear loci only using root-dating prior.

**Fig. S5** Posterior probability distribution of parameters, with population sizes (A and B), migration rates in individuals (C and D), and divergence times in years ago (E and F) in $IM_A2$ analysis.

**Table S1** Number of chromosomes analyzed (N), sequence length, number of segregating sites (S), haplotype diversity (Hd), nucleotide diversity per site ($\times 10^3$ $\pi$/bp), Waterson's estimator of Theta per site ($\times 10^3$ $\theta$w), Tajima's D (Taj D), Fu's Fs (FuFs), and Fay and Wu's H (FWH) for three focal species.

**Table S2** The estimated mean log probability of the data *standard deviation* of *LnP*(*K*) and Delta *K* for *Takydromus* species inferred by Bayesian methods (Hubisz *et al.* 2009; Pritchard *et al.* 2000).

**Table S3** Parameter estimates in $IM_A2$ analysis.

**Table S4** The summary statistics used in the ABC analysis.

**Table S5** Prior distributions of each demographic model used in ABC analysis.

**Appendix S1** Methods.