

Biogeography Revealed by *Mariner*-Like Transposable Element Sequences via a Bayesian Coalescent Approach

Shigeki Nakagome · Yumiko Nakajima ·
Shuhei Mano

Received: 18 February 2013 / Accepted: 19 August 2013 / Published online: 30 August 2013
© Springer Science+Business Media New York 2013

Abstract Genetic diversity of natural populations is useful in biogeographical studies. Here, we apply a Bayesian method based on the coalescent model to dating biogeographical events by using published DNA sequences of wild silkworms, *Bombyx mandarina*, and the domesticated model organisms *B. mori*, both of which categorized into the order of *Lepidoptera*, sampled from China, Korea, and Japan. The sequences consist of the *BmTNML* locus and the flanking intergenic regions. The *BmTNML* locus is composed of cecropia-type *mariner*-like element (MLE) with inverted terminal repeats, and three different transposable elements (TE), including *LIBM*, *BMC1* retrotransposons, and *BmamaT1*, are inserted into the MLE. Based on the genealogy defined by TE insertions/deletions (indels), we estimated times to the most recent common ancestor and these indels events using the flanking, MLE, and indels sequences, respectively. These estimates by using MLE sequences strongly correlated with those by using flanking sequences, implying that cecropia-type MLEs can be used as a molecular clock. MLEs are thought to have transmitted horizontally among different species. By using a pair of published cecropia-type MLE sequences from lepidopteran insect, an emperor moth, and a coral in Ryukyu Islands, we demonstrated dating of horizontal

transmission between species which are distantly related but inhabiting geographically close region.

Keywords Horizontal transmission · *Mariner*-like elements (MLEs) · Molecular clock · Natural history

Introduction

Understanding of species divergence are essential to natural history reconstructions (for example, Avise and Nelson 1989; Baker et al. 1993; Charruau et al. 2011; Driscoll et al. 2007). For example, the wild silkworms, *Bombyx mandarina*, are now widely spreading in the far east Asian region except for the Ryukyu Islands in the southern part of the Japanese archipelago. The domesticated model organism, *B. mori* and *B. mandarina* can be crossed and leave the offspring, and the hybrids are fertile. It has been thought that the origin of *B. mori* was *B. mandarina* inhabiting in China, since they have 28 chromosomes, as same as *B. mori*, while *B. mandarina* in Korea and Japan have 27 chromosomes. It has shown that the divergence between *B. mandarina* from Japan and a *B. mori* strain is around seven million years (Yukuhiro et al. 2002). Dating further details of the divergence events attracts much interest.

Different species, and sometimes even different kingdoms, may share various types of *mariner*-like transposable elements (MLEs; a DNA-type transposon) in their genomes (Robertson 1993; Robertson and MacLeod 1993; Hartl 2001; Bui et al. 2008). Moreover, very similar MLEs are often found in distantly related species (Maruyama and Hartl 1991; Robertson and Lampe 1995; Hartl et al. 1997; Casse et al. 2006). It is thought that each MLE has transposed horizontally across species at various times by moving into the germline of a different organism. Full

S. Nakagome · S. Mano (✉)
Department of Mathematical Analysis and Statistical Inference,
The Institute of Statistical Mathematics, 10-3 Midori-cho,
Tachikawa, Tokyo 190-8562, Japan
e-mail: smano@ism.ac.jp

Y. Nakajima
Functional Genomics Group, Center of Molecular Biosciences
(COMB), Tropical Biosphere Research Center, University of the
Ryukyus, 1 Nishihara, Okinawa 903-0213, Japan

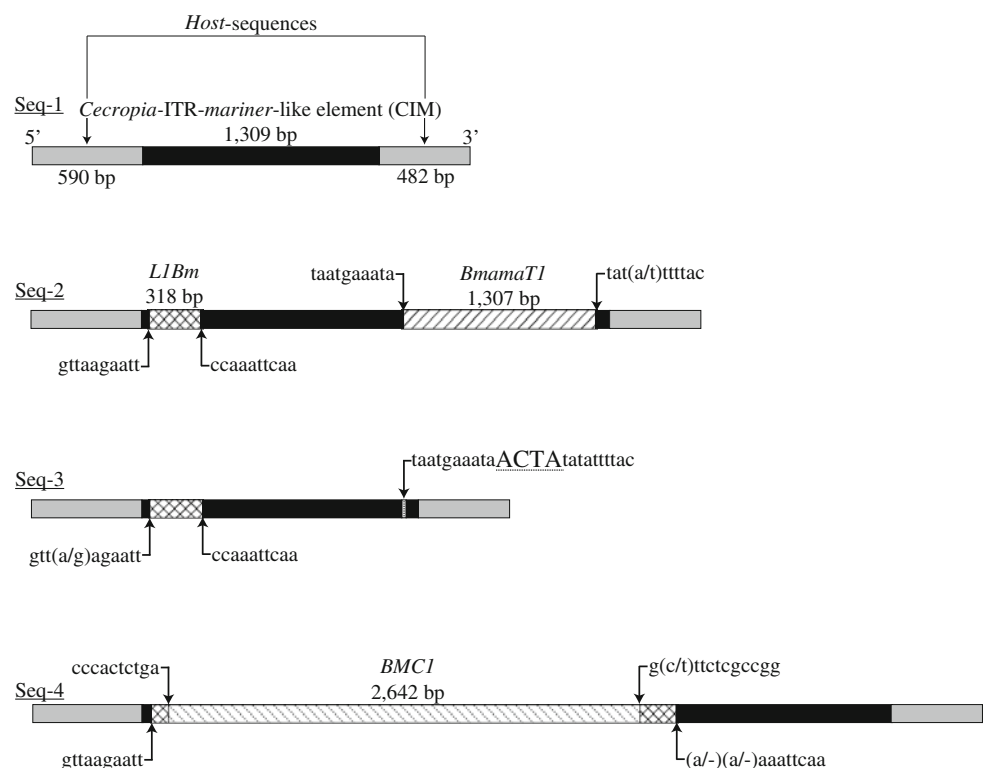
length of MLE sequences can be amplified and isolated from several different species by designed primers based on terminal inverted repeats (TIRs) of MLEs from some species. Therefore, we can hope that dating events between or among species indirectly is possible by investigating sequence divergence (Nakajima et al. 2002; Kawanishi et al. 2007). However, such inferences are possible as long as the molecular clock hypothesis holds (Zuckerkindl and Pauling 1965).

In the previous studies by one of the authors, the full-length of MLEs isolated by using TIR of the MLE from *Hyalophora Cecropia* (Lidholm et al. 1991) from some different species are named as the CIM (Cecropia-ITR-MLE). It was shown that the CIMs including a complete open reading frame for transposase was shared among distantly related species, some lepidopteran insects including an emperor moth (*Attacus atlas*), grasshopper (*Traulia ornata*), and a coral (*Fungia scruposa*), from the Ryukyu Islands (Nakajima et al. 2002). It seemed like that these CIMs have transmitted horizontally among the different species in Ryukyu Islands. CIMs is categorized into *Bmmar2* type of MLE defined by Kumaresan and Mathavan (2004) spreads 19 copies in the genome of *B. mori*. It exists by four copies on the chromosome 6. One of them has a unique tripartite structure composed of three different transposable elements, namely CIM (1,309 bp), *LIBm* (318 bp), and *BMCI* (2,642 bp) retrotransposable elements and we named it as “*BmTNM* locus” in 1999 (Nakajima

et al. 1999). The *BmTNML* including flanking sequence (1,072 bp) denoted *host*-sequences are extracted from population of *B. mori* and *B. mandarina* (Kawanishi et al. 2008). Although the tripartite structure of the *BmTNML* of *B. mandarina* inhabiting various area shows little difference, the sequences have rich polymorphisms and some of them were inserted by new *maT*-type transposon *BmamaT1* (1307 bp) (Kawanishi et al. 2008). The structures of the sequence are summarized in Fig. 1. The advantage to focusing on the *BmTNML* locus is that the locus is uniquely identified in the *Bombix* genome and the genealogy will be the same between *host*-sequences and CIM due to the tight linkage, while patterns of mutations superimposed on the genealogy are independent between them. Therefore, the equality of substitution rate between the estimates given by *host*-sequences and those given by CIM can be evaluated at various time points on the genealogy.

In this study, we estimated the times to the most recent common ancestor (MRCA) and the TE insertion/deletion (named as indel hereafter) events on the *BmTNML* locus. These estimates by using CIM sequences strongly correlated with those by using flanking sequences, implying that this type of MLEs can be used as a molecular clock for dating evolutionary events among species. By using the CIM sequences from *A. atlas* and *F. scruposa* we estimated the time of the horizontal transmission event between these species.

Fig. 1 Genomic structures of the *BmTNML* locus consisting of *host*-sequences, CIM, *LIBm*, *BmamaT1*, and *BMCI*. The upstream and downstream sequences of the inserted sites are shown in the *flanking regions* of TEs. The polymorphisms in these sequences are denoted as the *parenthesis* including two alleles. The excision of *BmamaT1* is shown as a *dotted line* with the footprint “ACTA” and the flanking sequences “taatgaata” and “tatattttac”



Methods

A basic local alignment search tool (BLAST) search (<http://kaikoblast.dna.affrc.go.jp/>) was conducted against the *B. mori* genome using the common TIR sequence flanking MLEs (Robertson 1993; Robertson and MacLeod 1993) as a probe. At a threshold of more than 95 % sequence homology, we found that 19 kinds of CIMs were inserted in the *B. mori* genome. In this locus, 11 publicly available sequences from *B. mandarina* and two from *B. mori* are obtained from the National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov/>). Sequences are categorized into four types (from Seq-1 to Seq-4) according to indels of TEs. Geographic labels (and accession numbers) of the sequences are as follows: Seq-1 China (AB3630105), Seq-2 Korea (AB363006), Japan (AB363010, AB363014, AB473770, AB473771), Seq-3 Japan (AB363021, AB363024, AB473769, AB473773), and Seq-4 China (AB473763), *B. mori* Japan (AB363029), *B. mori* (strain name DAIZO) derived from *B. mori* genome. The *BmTNML* locus is intergenic and no significant deviation from the selective neutrality was detected according to Tajima's test (Tajima 1989).

Indels of TEs can be present as unique event polymorphisms (UEPs), because these events occur exactly once in the evolutionary history. Therefore, the genealogy is specifically defined by UEPs. Figure 2a shows the genealogy. T_1 – T_6 represent MRCAs of the sequences included in each lineage. The number of segregating sites within sample sequences in a lineage was counted in each segment of *host*-sequence, CIM, *LIBm*, *BmamaT1*, or *BMCI*, respectively. Alignment gaps were excluded from subsequent analyses. The lineage is defined as the descendant leaves of each internal node of the genealogy. For example, the descendants of the T_2 node consist of sequences belonging to Seq-4, Seq-2, and Seq-3 (Fig. 1). When a given leaf consisted of multiple sequences, the number of segregating sites in the sequences was counted. We see that (1) CIM was already inserted in the *BmTNML* locus at the root of the genealogy. (2) The tripartite structure of CIM, *LIBm*, and *BMCI* (Fig. 1), characterized the *B. mori* sequences and the *B. mandarina* sequence from China. (3) The *BmamaT1* deletion was observed in the sub-lineage (Seq-3) of the *BmamaT1* inserted sequences. It suggests that the wild silkworm comes to Japan from China through Korea

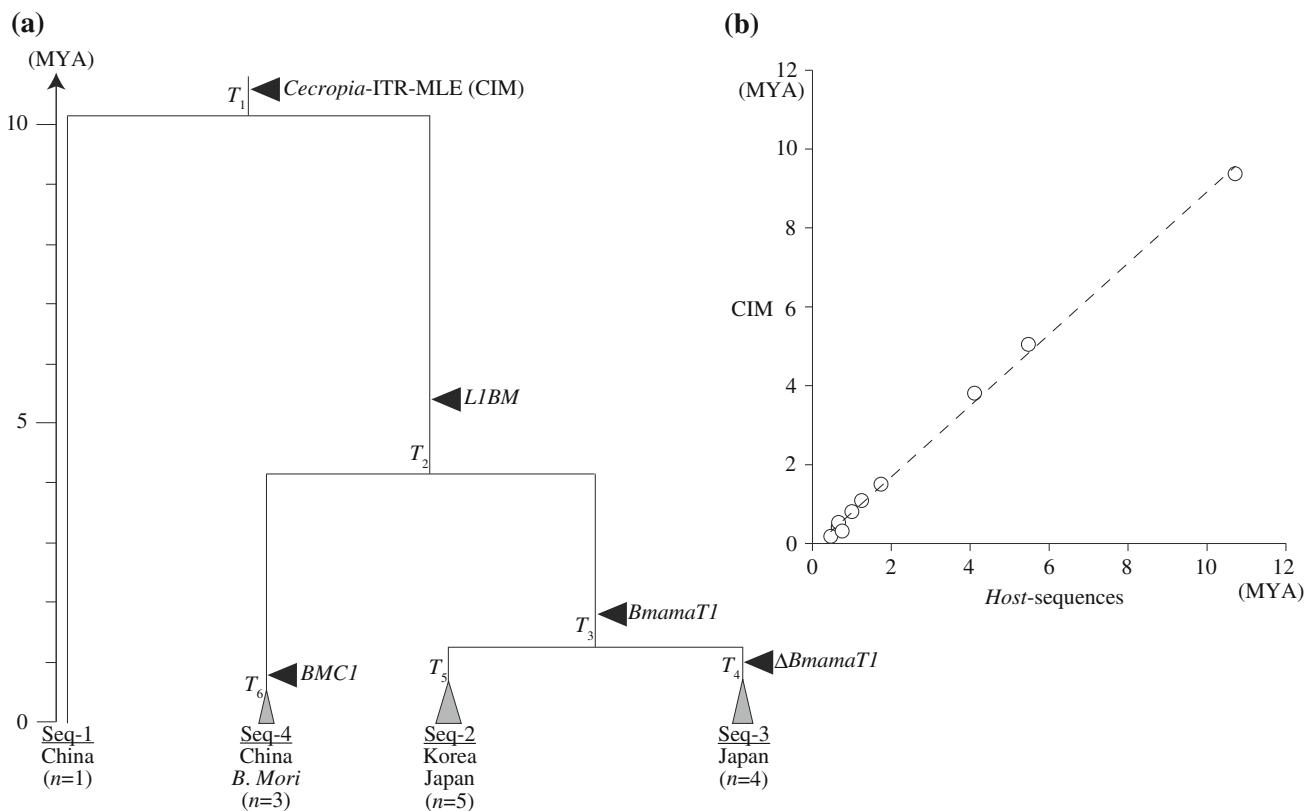


Fig. 2 a The genealogy of *B. mandarina* population using *host*-sequences at the *BmTNML* locus. Black triangles represent the insertion or the excision ($\Delta BmamaT1$) (Kawanishi et al. 2008) of TEs. For Seq-1, the lineage consists of a single sequence, and none of the

coalescent estimate for $TMRC\Delta$ is obtained in the lineages. Times are scaled by MYA (see Table 1). **b** The correlation of estimated times scaled by MYA between *host*-sequences and CIM. Circles indicate the $TMRC\Delta$ and $T\Delta$ referred from Table 1

and the silkworm domestication occurred in China, as is historically known (Xiang et al. 2005).

Based on the coalescent model (Kingman 1982; Hudson 1983; Tajima 1983), we estimated the ages of TE indels and the TMRCA (times to the MRCA) of the lineages by a Bayesian method. Given, the number of segregating sites in a sample of size n (S_n), coalescent times were estimated by the rejection-sampling method (Ripley 1987). From the Bayes rule, the posterior distribution of the coalescent times $T = (T_n, T_{n-1}, \dots, T_2)$ is $P(T|S_n) \propto f(S_n|T)\pi(T)$. The likelihood of $f(S_n|T)$ follows a Poisson distribution, and T_i , which is the waiting time of coalescence between i sequences, follows an exponential distribution with parameter $i(i-1)/2$. We denoted the joint distribution of waiting times T_i by $\pi(T)$. We assumed from the infinitely-many-sites model that each gene is a sequence of completely linked sites and that every mutation occurs at a site different from the sites of the previous mutations (Watterson 1975). We obtained the posterior estimate of TMRCA of the sample using the following algorithm (Tavaré et al. 1997):

Algorithm 1 To simulate from the joint density of T_n, T_{n-1}, \dots, T_2 given $S_n = k$. The algorithm is given by replacing θ in Algorithm 7.3 of Tavaré (2004) with Watterson's estimator θ_w (Watterson 1975). We repeated the simulation until 1,000,000 simulated samples were obtained.

We investigated the times to a TE indel event ($T\Delta$) as observed in sample sequences using the conditional genealogy given the UEP. Given that the TE indel is represented b times in the sample ($1 \leq b < n$), the UEP property requires that the b sequences coalesce together before any of the non-TE indel sequences share any common ancestors with them. We modified the rejection-sampling method of "Algorithm 1" according to previous studies (Slatkin and Rannala 1997; Tavaré 2004), and let $TMRC\Delta$ as the TMRCA given that the sequences have the UEP.

Algorithm 2 To simulate coalescent times from conditional distributions of $T\Delta$ and $TMRC\Delta$ given m additional mutations have occurred in a linked region containing the UEP. The algorithm is given by replacing θ in Algorithm 8.2 of Tavaré (2004) with θ_w .

We repeated the simulation until the accepted samples reached 1,000,000.

We computed the number of segregating sites among the sequences corresponding to each of the MRCAs (T_1 – T_6 in Fig. 2a). The TMRCA in T_1 or T_5 was estimated using algorithm 1, because of the fixation of CIM or unconditional with TE. The other four TMRCA were conditional with the insertions of *LIBm*, *BmamaT1*, and *BMCI*, or with the deletion of *BmamaT1*. Algorithm 2 was used to estimate the $TMRC\Delta$ and $T\Delta$.

Results and Discussion

The posterior distributions of estimated times at all points of MRCAs and UEPs were overlapped among *host*-sequences, CIM, and TEs (Table 1). It was suggested that the MRCA of the *B. mandarina* population existed 10.713 MYA and 95 % credible interval (CI) 6.701–15.642. This estimate was nearly identical to that using CIM (9.373 MYA and 95 %CI 5.916–13.579). Subsequent lineages were characterized by the *LIBm* (T_2), *BmamaT1* (T_3), and *BMCI* (T_6) insertions and the *BmamaT1* deletion (T_4) (Fig. 2a). One of the UEPs, *BMCI*, defines the origin of the *B. mori* population that shares a common ancestor with *B. mandarina* from China. This relationship is consistent with previous studies (MinHui et al. 2008; Li et al. 2010). We suggest that their common ancestor dates to 0.468 MYA. The estimate is much older than the time of domestication 5,000 years ago (Xiang et al. 2005), while much more recent than the 7.1 MYA divergence between *B. mandarina* from Japan and *B. mori* (Yukuhiro et al. 2002). The estimated times of $TMRC\Delta$ and $T\Delta$ from *host*-sequences were strongly correlated with those using CIMs ($r^2 = 0.9972$) along the lineages of the genealogy (Fig. 2b). This strong correlation between estimate by using selectively neutral sequence and MLE sequence supports that CIM is a useful molecular clock.

The horizontal transmission of the CIM has been observed in an emperor moth (*A. atlas*: AB006464) and a coral (*F. scruposa*: AB055188) from the Ryukyu Islands (Nakajima et al. 2002). The two sequences differ at six sites. We estimated the time to the insertion event by assuming that the number of mutations follow the Poisson distribution with parameter $2ut$ with the mutation rate u followed the Gamma distribution with shape parameter and the scale parameter are α and β , respectively, and the mean was the *B. mandarina* estimate (2.759×10^{-3} mutations/site/million years \times 1,223 bp; see "Note" in Table 1). Integrating out the mutation rate, the marginal likelihood of the time (t) was obtained as $L(t|n) = \frac{(2t)^n \Gamma(n+\alpha)}{n! \Gamma(\alpha) \beta^\alpha (2t + \frac{1}{\beta})^{n+\alpha}}$, where n is the number of nucleotide differences between the two sequences. The maximum likelihood estimator of the time is $n/2\alpha\beta$, with asymptotic variance $(n-1)(\alpha+n-1)/2\alpha(\alpha+2)$. The maximum likelihood estimate was 0.89 MYA with a 95 % confidence interval 0–3.23 MYA (the variance equals the mean in the Gamma distribution) or 0–3.88 MYA (the variance equals ten times larger than the mean). These two species belong to different phyla of Arthropoda and Cnidaria. This estimate is much more recent than the time of existence of their MRCA.

Here we demonstrate that CIMs can be used as a molecular clock to date horizontal transmissions. But our justification comes from analyses of only the single locus

Table 1 Coalescent times of *host*-sequences, *Cecropia*-ITR-MLE (CIM), *LIBm*, *BmamaTI*, and *BMCI* in the *B. mandarina* genealogy

MRCA	n^a	Host-sequences			CIM		LIBm		BmamaTI		BMCI	
		S_n (θ_w) ^c	Times (MYA) [95 % CI ^d]	Times (MYA) [95 % CI]	S_n (θ_w)	Times (MYA) [95 % CI]	S_n (θ_w)	Times (MYA) [95 % CI]	S_n (θ_w)	Times (MYA) [95 % CI]	S_n (θ_w)	Times (MYA) [95 % CI]
T_1	13	101 (32.547)	10.713 [6.701–15.642]	9.373 [5.916–13.579]	114 (36.736)	–	–	–	–	–	–	–
T_2												
<i>TMRCAL</i>	12	46 (15.232)	4.116 [2.390–6.421]	3.808 [2.241–5.871]	52 (17.219)	13 (4.305)	3.104 [1.489–5.595]	–	–	–	–	–
<i>TA</i>			5.475 [2.819–10.719]	5.048 [2.636–9.828]			4.357 [1.873–9.293]	–	–	–	–	–
T_3												
<i>TMRCAL</i>	9	15 (5.519)	1.254 [0.610–2.199]	1.087 [0.536–1.888]	16 (5.887)	9 (3.311)	2.204 [0.960–4.153]	–	–	–	–	–
<i>TA</i>			1.744 [0.763–3.661]	1.505 [0.667–3.142]			3.168 [1.252–6.956]	–	–	–	–	–
T_5	5	5 (2.400)	0.660 [0.250–1.326]	0.422 [0.150–0.879]	4 (1.920)	0 (0.000)	N.A.	0.936 [0.423–1.709]	9 (4.320)	–	–	–
T_4												
<i>TMRCAL</i>	4	7 (3.818)	0.666 [0.260–1.284]	0.535 [0.209–1.031]	7 (3.818)	4 (2.182)	1.045 [0.322–2.272]	–	–	–	–	–
<i>TA</i>			1.002 [0.355–2.301]	0.805 [0.285–1.848]			1.675 [0.484–4.136]	–	–	–	–	–
T_6												
<i>TMRCAL</i>	3	5 (3.333)	0.468 [0.146–0.994]	0.178 [0.041–0.440]	3 (2.000)	2 (1.333)	0.396 [0.072–1.094]	–	–	5 (3.333)	0.181 [0.056–0.386]	–
<i>TA</i>			0.760 [0.218–1.882]	0.315 [0.069–0.851]			0.757 [0.135–2.192]	–	–	–	0.293 [0.084–0.726]	–

Times were measured in units of $2 N_e$ generations, and the effective population size was estimated by dividing the Watterson's estimator (θ_w) by the mutation rate. We used 2.759×10^{-9} as the per-site per-year mutation rate; this value was calculated from the average number of substitutions and was corrected by the Tajima–Nei substitution model (Tajima and Nei 1984) at three nuclear gene loci (*snf*, *RplI215*, and *Pdg*) between Diptera (*Drosophila melanogaster*: NM_078490, NM_078569, and M80598) and *Lepidoptera* (*B. mori*: DQ202313, M55993, and NM_001047060; *Heliothis virescens*: GQ265660 and GQ265618) assuming a Diptera–Lepidoptera divergence time of 190 MYA (Grimaldi and Engel 2005). Coalescent dates then were translated to chronological time using “ $2 \times N_e \times T$ (coalescent time units) \times the 1-year inter-generation time.” In this table, – denotes that there is no MRCA in a transposable element, and N.A. denotes that coalescent times can not be assessed due to no-polymorphic sites

^a Number of sequences used to compute the summary statistics “S”

^b Number of segregating sites

^c Watterson's theta

^d 95 % credible intervals in the posterior distributions of coalescent times

of the single species. Further study about other loci of CIMs and also about other MLEs to confirm the usefulness of MLEs as a molecular clock is needed. Since MLEs are easily amplified and identified without prior information about surrounding genomic sequences of the organisms, our method can be expanded to the analysis of other transposable elements and applied to diverse organisms and will provide crucial insights of natural history reconstructions in biogeography.

Acknowledgments We would like to thank Drs. Hideaki Maekawa and Yuichi Kawanishi for their helpful discussions and comments to this study. S.N. has been supported in part by a Grant-in-Aid for the Japan Society for the Promotion of Science (JSPS) Research fellow (24-3234). Y.N. and S.M. has been supported by KAKENHI 19658023.

References

- Avise JC, Nelson WS (1989) Molecular genetic relationships of the extinct dusky seaside sparrow. *Science* 243:646–648
- Baker CS, Perry A, Bennister JL, Weinrich MT, Abernethy RB, Calambokidis J, Liens J, Lambertsen RH, Urbán Ramírez J, Vasquez O, Clapham PJ, Alling A, O'Brien S, Palumbi SR (1993) Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. *Proc Natl Acad Sci USA* 90:8239–8243
- Bui QT, Casse N, Leignel V, Nicolas V, Chenais B (2008) Widespread occurrence of mariner transposons in coastal crabs. *Mol Phylogenet Evol* 47:1181–1189
- Casse N, Bui QT, Nicolas V, Renault S, Bigot Y, Lallier M (2006) Species sympatry and horizontal transfers of mariner transposons in marine crustacean genomes. *Mol Phylogenet Evol* 40:609–619
- Charreau P, Fernandes PC, Orozco-Terwengel P, Peters J, Hunter L, Ziaie H, Jourabchian A, Jowkar H, Schaller G, Ostowski S, Vercammen P, Grange T, Schlötterer C, Kotze A, Geigl E-M, Walzer C, Burger PA (2011) Phylogeography, genetic structure and population divergence time of cheetahs in Africa and Asia: evidence for long-term geographic isolates. *Mol Ecol* 20:706–724
- Driscoll CA, Menotti-Raymond M, Roca AL, Hupe K, Johnson WE, Geffen E, Harley EH, Delibes M, Pontier D, Kitchener AC, Yamaguchi N, O'Brien SJ, Macdonald DW (2007) The near eastern origin of cat domestication. *Science* 317:519–523
- Grimaldi D, Engel MS (2005) *Evolution of the insects*. Cambridge University Press, New York
- Hartl DL (2001) Discovery of the transposable element mariner. *Genetics* 157:471–476
- Hartl DL, Lozovskaya ER, Nurminsky DI, Lohe AR (1997) What restricts the activity of mariner-like transposable elements. *Trends Genet* 13:197–201
- Hudson RR (1983) Properties of a neutral allele model with intragenic recombination. *Theor Popul Biol* 23:183–201
- Kawanishi Y, Takaishi R, Banno Y, Fujimoto H, Nho SK, Maekawa H, Nakajima Y (2007) Sequence comparison of mariner-like elements among the populations of *Bombyx mandarina* inhabiting China, Korea and Japan. *J Insect Biotechnol Sericol* 76(2):79–87
- Kawanishi Y, Takaishi R, Morimoto M, Banno Y, Kab Nho S, Maekawa H, Nakajima Y (2008) A novel maT-type transposable element, BmamaT1, in *Bombyx mandarina*, homologous to the *B. mori* mariner-like element Bmmar6. *J Insect Biotechnol Sericol* 77:45–52
- Kingman JFC (1982) On the genealogy of large populations. *J Appl Probability* 19:27–43
- Kumaresan G, Mathavan S (2004) Molecular diversity and phylogenetic analysis of mariner-like transposons in the genome of the silkworm *Bombyx mori*. *Insect Mol Biol* 13(3):249–271
- Li D, Guo Y, Shao H, Tellier LD, Wang J, Xiang Z, Xia Q (2010) Genetic diversity, molecular phylogeny and selection evidence of the silkworm mitochondria implicated by complete resequencing of 41 genomes. *BMC Evol Biol* 10:81
- Lidholm DA, Gudmundsson GH, Boman HG (1991) A highly repetitive, mariner-like element in the genome of *Hyalophora cecropia*. *J Biol Chem* 266(18):11518–11521
- Maruyama K, Hartl DL (1991) Evolution of the transposable element mariner in *Drosophila* species. *Genetics* 128:319–329
- MinHui P, QuanYou Y, YuLing X, YanQun L, Cheng L, Zhang Z, Xiang ZH (2008) Characterization of mitochondrial genome of Chinese wild mulberry silkworm, *Bombyx mandarina* (Lepidoptera: Bombycidae). *Sci China Ser C* 51:693–701
- Nakajima Y, Hashido K, Tsuchida K, Takada N, Shiino T, Maekawa H (1999) A novel tripartite structure comprising a mariner-like element and two additional retrotransposons found in the *Bombyx mori* genome. *J Mol Evol* 48:577–585
- Nakajima Y, Fujimoto H, Negishi T, Hashido K, Shiino T, Tsuchida K, Hidaka M, Takada N, Maekawa H (2002) Possible horizontal transfer of mariner-like sequences into some invertebrates including Lepidopteran insects, a grasshopper and a coral. *J Insect Biotechnol Sericol* 71:109–121
- Ripley BD (1987) *Stochastic simulation*. Wiley, New York
- Robertson HM (1993) The mariner transposable element is widespread in insects. *Nature* 362:241–245
- Robertson HM, Lampe DJ (1995) Distribution of transposable elements in arthropods. *Annu Rev Entomol* 40:333–357
- Robertson HM, MacLeod EG (1993) Five major subfamilies of mariner transposable elements in insects, including the mediterranean fruit fly, and related arthropods. *Insect Mol Biol* 2:125–139
- Slatkin M, Rannala B (1997) Estimating the age of alleles by use of intra-allelic variability. *Am J Hum Genet* 60:447–458
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol* 1:269–285
- Tavaré S (2004) Ancestral inference in population genetics. In: *Lectures on probability theory and statistics, Ecole de d'Eté de Probabilités de Saint-Flour XXXI-2001*. Springer, Heidelberg
- Tavaré S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics* 145:505–518
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276
- Xiang ZH, Huang JT, Xia JG, Lu C (2005) *Biology and sericulture*. China For Publ House, Beijing
- Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, Banno Y (2002) Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkworm, *Bombyx mandarina*, and its close relative, the domesticated silkworm, *Bombyx mori*. *Mol Biol Evol* 19(8):1385–1389
- Zuckerandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) *Evolutionary genes and proteins*. Academic Press, New York, pp 97–166