Molecular Biology and Evolution

Evidence of Selection at Melanin Synthesis Pathway Loci during Silkworm Domestication

Hong-Song Yu,[1] Yi-Hong Shen,[1] Gang-Xiang Yuan,[1] Yong-Gang Hu,[1] Hong-En Xu,[1] Zhong-Huai Xiang,[1] and Ze Zhang*[,1,2]

[1]The Key Sericultural Laboratory of Agricultural Ministry, Southwest University, Chongqing, China

[2]The Institute of Agricultural and Life Sciences, Chongqing University, Chongqing, China

***Corresponding author**: E-mail: zezhang@swu.edu.cn.

Associate editor: Adriana Briscoe

Sequence data from this article have been submitted to GenBank (Accession nos. HM545477 - HM545716, HM856623, HQ338819 - HQ338885, HQ338916- HQ338945, HQ589841-HQ589900).

Research article

Abstract

The domesticated silkworm (*Bombyx mori*) was domesticated from wild silkworm (*Bombyx mandarina*) more than 5,000 years ago. During domestication, body color between *B. mandarina* and *B. mori* changed dramatically. However, the molecular mechanism of the silkworm body color transition is not known. In the present study, we examined within- and between-species nucleotide diversity for eight silkworm melanin synthesis pathway genes, which play a key role in cuticular pigmentation of insects. Our results showed that the genetic diversity of *B. mori* was significantly lower than that of *B. mandarina* and 40.7% of the genetic diversity of wild silkworm was lost in domesticated silkworm. We also examined whether position effect exists among melanin synthesis pathway genes in *B. mandarina* and *B. mori*. We found that the upstream genes have significantly lower levels of genetic diversity than the downstream genes, supporting a functional constraint hypothesis (FCH) of metabolic pathway, that is, upstream enzymes are under greater selective constraint than downstream enzymes because upstream enzymes participate in biosynthesis of a number of metabolites. We also investigated whether some of the melanin synthesis pathway genes experienced selection during domestication. Neutrality test, coalescent simulation, as well as network and phylogenetic analyses showed that tyrosine hydroxylase (TH) gene was a domestication locus. Sequence analysis further suggested that a putative expression enhancer (Abd-B–binding site) in the intron of *TH* gene might be disrupted during domestication. TH is the rate-limiting enzyme of melanin synthesis pathway in insects. Real-time polymerase chain reaction assay did show that the relative expression levels of *TH* gene in *B. mori* were significantly lower than that in *B. mandarina* at three different developmental stages, which is consistent with light body color of domesticated silkworm relative to wild silkworm. Therefore, we speculated that expression change of *TH* gene may contribute to the body color transition from *B. mandarina* to *B. mori*. Our results emphasize the exceptional role of gene expression regulation in morphological transition of domesticated animals.

Key words: silkworm, domestication, body color, melanin synthesis pathway, selection, tyrosine hydroxylase.

Introduction

The domesticated silkworm, *Bombyx mori*, is the model insect for Lepidoptera and is of great economic value because of its silk production. Over 1,000 inbred and mutant lines of domesticated silkworm are kept worldwide. It is believed that *B. mori* originated from the ancient wild silkworm, *Bombyx mandarina*, and has been domesticated and bred for more than 5,000 years (Xiang et al. 2005). During domestication, body color between *B. mandarina* and *B. mori* changed dramatically. *Bombyx mandarina* larvae and adults have black body color, whereas *B. mori* larvae and adults have white body color (supplementary fig. S1, Supplementary Material online).

Melanin, a predominant black pigment in insects, is produced by epidermal cells through the melanin synthesis pathway (Wright 1987). The genes responsible for melanin synthesis have been well studied in many insects, especially in Drosophila *melanogaster* (True 2003; Wittkopp et al. 2003) (fig. 1*a*). The enzyme tyrosine hydroxylase (*TH*) initiates synthesis of melanin by hydroxylating tyrosine to dopa (Wright 1987). Then dopa is converted to dopa melanin (which is black) by the activities of Yellow, Yellow-f, and Yellow-f2 proteins (Walter et al. 1991; Han et al. 2002; Wittkopp et al. 2002a, 2002b). Some dopa can also be converted to dopamine by the dopa decarboxylase (DDC) (Wright 1987), and dopamine serves as a precursor of dopamine melanin (which is brown). Alternatively, *N*-β-alanyl dopamine (NBAD) synthetase can catalyze dopamine to produce NBAD, which is the precursor of yellow sclerotin (Wittkopp et al. 2002a). NBAD can also be converted back into dopamine by an NBAD hydrolase, which is encoded by the *tan* gene (Wittkopp et al. 2003; True et al. 2005). Finally, a family of arylalkylamine-*N*-acetyl transferases catalyzes dopamine to *N*-acetyl dopamine, which is the precursor of colorless sclerotin (Brodbeck et al. 1998). In the swallowtail butterfly, *Papilio xuthus*, its larval cuticular marking is also associated with melanin synthesis pathway genes, such as *TH*, *Ddc*, *yellow*, *ebony*, and *tan* (Futahashi and Fujiwara 2005, 2007; Futahashi et al. 2010).

With the completion of genome project (Xia et al. 2004; Mita et al. 2004; The International Silkworm Genome Consortium 2008), silkworm functional genomics has focused on identifying the responsible genes for numerous silkworm mutants. Recently, the melanin synthesis pathway genes as candidate genes for some body color mutants have been identified. *Yellow* and *ebony* are the responsible genes for the larval color mutant *ch* and *so*, respectively, and *yellow* promotes melanization, whereas *ebony* inhibits melanization (Futahashi et al. 2008). *Dat1* is the responsible gene for the adult color mutant *mln* (Dai et al. 2010). *Tan* is responsible for the larval color mutant *ro* and plays a crucial role in normal black marking formation in *B. mori* (Futahashi et al. 2010). *TH* plays an important role in melanin synthesis producing neonatal larval color and is responsible for the sex-linked chocolate mutant (Liu et al. 2010). However, the molecular mechanism of the silkworm body color transition during domestication is unknown.

In addition, previous studies demonstrated that the upstream genes in plant anthocyanin metabolic pathway evolved more slowly than the downstream genes. This difference in evolutionary rate is due to more functional constraint upon the upstream genes because they participated in several different biochemical pathways (Rausher et al. 1999; Lu and Rausher 2003; Rausher et al. 2008), that is, mutations in the upstream genes generate greater pleiotropic effects than those in downstream genes in the pathway, being more likely to produce a deleterious effect. This pattern of positional rate variation has been confirmed by some studies (Riley et al. 2003; Livingstone and Anderson 2009; Ramsay et al. 2009) but has not been supported by others (Olsen et al. 2002; Flowers et al. 2007; Ramos-Onsins et al. 2008). Whether the position effect exists in insect melanin synthesis pathway remains to be investigated.

In the present study, we focused on the patterns of nucleotide diversity and selection signature for eight melanin synthesis pathway genes in silkworm. We mainly addressed the following two questions: Whether there is a positional variation effect in melanin synthesis pathway of silkworm or not? Is there any evidence of artificial selection acting on the melanin synthesis pathway genes in domesticated silkworm?

Materials and Methods

Sample Collection

A list of samples was given in supplementary table S1 (Supplementary Material online). First, a total of 15 domesticated silkworm strains and 15 wild silkworm samples were used for genetic diversity analysis, then another 60 domesticated silkworm strains and 7 wild silkworm samples were used to confirm the sites under selection. The domesticated accessions were obtained from the silkworm resource pool at the Institute of sericulture and Systems Biology, Southwest University, China and represented four main geographic lines (i.e., Chinese, Japanese, European, and Tropical) with diverse genetic diversity in voltinism and other characteristics. The wild silkworm samples were collected from various geographical regions in China.

Total genomic DNA was extracted from pupa or moth of single individual according to a standard phenol–chloroform protocol (Nagaraja et al. 1995).

Loci Studied

We focused on the nine genes responsible for melanin synthesis in *D. melanogaster* (fig. 1*a*). To identify putative orthologous genes in *B. mori*, we performed a two-round search for each locus. First, we conducted a TBlastN search against *B. mori* genome sequence and annotated genes in SilkDB for each *D. melanogaster* protein (Duan et al. 2010). Second, each hit (*E* value $\leq 10^{-5}$) was in silico translated and utilized as a query for searching the *D. melanogaster* genome and annotated genes in Flybase (Tweedie et al. 2009). If the best hit in this second round search was in accordance with the original *D. melanogaster* gene, the gene was considered as an orthologous gene. An intergenic region unlinked to any melanin synthesis pathway genes was used to exclude demographic effect from bottleneck. The linked upstream and downstream genomic regions of the *TH* gene were used to exclude hitchhiking effects, including genes *BGIBMGA000564*, *BGIBMGA000655*, *BGIBMGA000656*, and the intergenic region between *TH* gene and *BGIBMGA000656*.

PCR Amplification, Cloning, and Sequencing

All the primer pairs of eight loci were designed based on the silkworm genome sequence (Xia et al. 2004; Mita et al. 2004; The International Silkworm Genome Consortium 2008) using PrimerPriemer 5.0 software and were shown in supplementary table S2 (Supplementary Material online). Eight primer pairs were designed to amplify the whole sequence of *TH* gene because it is the rate-limiting enzyme of melanin synthesis pathway in insects, whereas only one primer pair was designed for the other loci. Each polymerase chain reaction (PCR) contained a total volume of 50 μL, containing 10 ng genomic DNA, 0.2 μM of each primer, 5 μL of 10 × PCR buffer (Mg^{2+} plus), 1 mM deoxyribonucleoside triphosphates and 1 U ExTaq DNA polymerase (Takara, China). The PCR conditions included 4-min initial denaturation at 94 °C, followed by 35 cycles of the following: 30 s denaturation at 94 °C, 40 s annealing at appropriate temperature which was listed in supplementary table S2 (Supplementary Material online), 2 min of extention at 72 °C and a final 10 min extension at 72 °C. The amplified products were purified with agarose gel purification kit (Omega, China) and then cloned into PMD19-T vector (Takara, China). Three to six positive clones were randomly selected and sequenced from each individual to avoid false polymorphism sites resulted from Taq polymerase error (Hilton and Gaut 1998). We employed BigDye 3.1 (Applied Biosystems, USA) chemistry and the recommended cycling conditions. Sequences were read using Applied Biosystems 3730 automated DNA sequencers. All clones were sequenced in both directions. All sequences have been assembled using SeqManII program in DNAStar software (DNASTAR, Madison, WI) and deposited in GenBank (Accession numbers HM545477–HM545716,

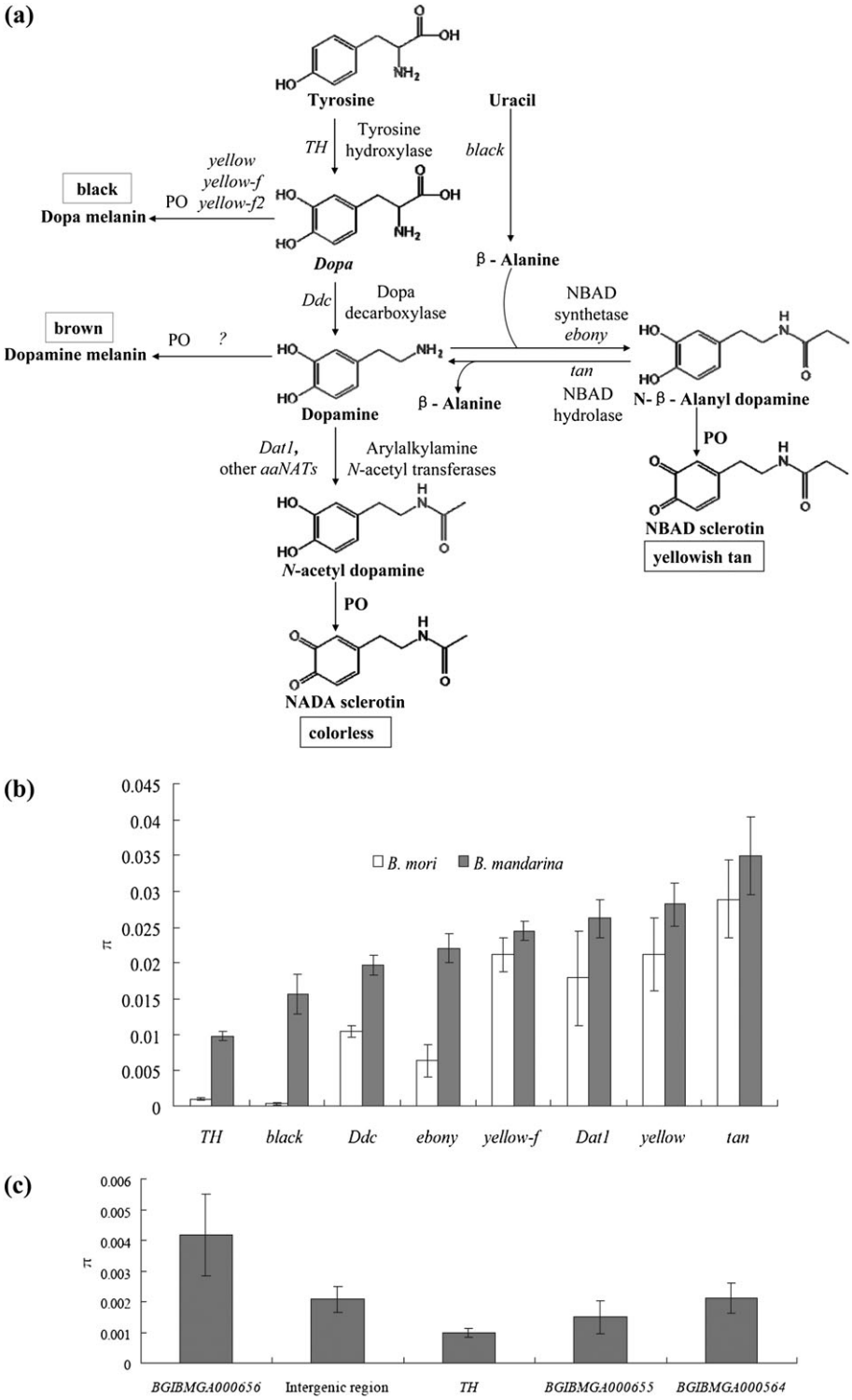

FIG. 1. (*a*) Melanin synthesis pathway in insect (modified from Wittkopp et al. 2003). Pigment precursors are shown in black and bold, enzymes are indicated in black, and the genes that encode them are shown in black and italic. (*b*) Comparison of nucleotide diversity (π) between *B. mori* and *B. mandarina*. The nucleotide diversity (π) was estimated for the whole sequence. (*c*) Comparison of nucleotide diversity (π) of the linked upstream and downstream genomic regions of the *TH* gene in *B. mori*.

HQ338819–HQ338885, HQ338916–HQ338945, HQ589841–HQ589900). We sequenced the homologous sequence of *TH* in *Ernolatia moorei* from the same family Bombycidae of *B. mori* as the outgroup for *TH* phylogenetic analysis. The sequence was deposited in GenBank and the accession number is HM856623.

Sequence Analysis

The DNA sequences were initially aligned using MUSCLE 3.6 (Edgar 2004) and further adjusted by hand. Phylogenetic trees were reconstructed by MEGA version 4.0 (Tamura et al. 2007), using the Neighbor-Joining (NJ) method. The complete deletion option was used in the NJ tree

reconstruction and the stability of internal nodes was assessed by bootstrap analysis with 1,000 resampling replicates. A median-joining network (Bandelt et al. 1999) was also constructed using the program Network version 4.516 (http://www.fluxus-technology.com/sharenet.htm).

Insertion/deletions (indels) were excluded from the analysis. For each locus, DnaSP version 5.10 (Librado and Rozas 2009) was utilized to calculate polymorphism parameters including number of segregating sites (S), Watterson's theta estimator (θ_w) (Watterson 1975), and the nucleotide diversity per site (π) (Tajima 1983) calculated for the whole sequence (π_{total}), for nonsynonymous sites (π_a), for synonymous sites (π_s) and for synonymous and noncoding sites (π_{silent}). θ_w represents the population mutation rate. The minimum number of recombination events (R_m) (Hudson and Kaplan 1985) and linkage disequilibrium (LD) measures (r^2) were also obtained using DnaSP version 5.10 (Librado and Rozas 2009). LD analysis was based on informative polymorphic sites with frequency $\geq$5%. r^2 was plotted against pairwise distance and two logarithm curves fitting to the data were drawn for *B. mori* and *B. mandarina*, respectively.

Tests for Neutrality

To examine deviation from the neutral evolution model and to reveal evolutionary history between *B. mori* and *B. mandarina*, Tajima's D tests (Tajima 1989), D test of Fu and Li (Fu and Li 1993), and McDonald–Kreitman (MK) test (McDonald and Kreitman 1991) were performed using DnaSP version 5.10 (Librado and Rozas 2009). Tajima's D was calculated for all sites based on the discrepancy between the nucleotide diversity per site (π) and Watterson's theta estimator (θ_w). MK test compares the ratio of nonsynonymous to synonymous variation within and between species. G test was used to determine statistical significance and G value was modified by Williams' correction. Hudson–Kreitman–Aguade (HKA) tests were conducted using HKA software (http://genfaculty.rutgers.edu/hey/software#HKA). H test of Fay and Wu (Fay and Wu 2000) was performed using a publicly available web interface (http://www.genetics.wustl.edu/jflab/htest.html). Recombination was analyzed using the composite-likelihood estimator of Hudson (2001) as implemented by LDhat version 2.1 (http://www.stats.ox.ac.uk/~mcvean/LDhat/).

Coalescent Simulations and Likelihood

To investigate bottleneck intensity during the domestication, coalescent simulations were performed for a single bottleneck model (Zhu et al. 2007; Li et al. 2009) using Hudson's ms program (Hudson 2002). The simulations were carried out with wild silkworm (*B. mandarina*) data as the basis of simulations and domesticated silkworm (*B. mori*) data for goodness of fit statistics (Tenaillon et al. 2004; Wright et al. 2005; Zhu et al. 2007; Li et al. 2009). Previous study showed that there was a unique domestication event for silkworm (Xia et al. 2009), so the bottleneck model assumed a single origin during the domestication (supplementary fig. S2, Supplementary Material online). The parameters used for simulation were as follows:

- N_a and N_p: N_a was the ancestor population size before split of wild and domesticated silkworms and N_p was the population size of present domesticated silkworm. As previous studies (Tenaillon et al. 2004; Wright et al. 2005; Zhu et al. 2007; Li et al. 2009) have shown that there was little effect of population size on simulation results, then we assigned both N_a and N_p to be 1,000,000.
- μ and γ: The mutation rate μ for each locus was estimated using θ_{silent} for wild silkworm, where $\mu = \theta/4N_e$. The recombination rate (γ) for each locus was determined by $\gamma = \rho/4N_e$ and ρ values for wild silkworm for each locus were calculated using the composite-likelihood estimator of Hudson (2001) as implemented by LDhat version 2.1 (http://www.stats.ox.ac.uk/~mcvean/LDhat/).
- k, N_b, and d: The key parameter of bottleneck model was bottleneck stringency k, which is the ratio of N_b and d. N_b was the population size during the bottleneck and d was the duration of bottleneck in generations. d was measured as $t2-t1$, which was supposed at time $t2$ generations ago, a single ancestral population of size N_a experienced an instantaneous shift in size to the bottlenecked population size N_b, and at time $t1$ generations ago, the bottleneck population expanded instantaneously to the present population size N_p. Although archaeological evidence suggested that *B. mori* has been domesticated for more than 5,000 years (Xiang et al. 2005), the accurate time of both $t2$ and $t1$ was unknown. So we made an implicit assumption that domestication of crops and silkworm began at almost the same time and fixed $t2$ to 7,500 and six d values (d = 200, 500, 1,000, 1,500, 2,000, and 3,000) were explored. We set maximum d value to 3,000 for two reasons. First, in coalescent simulations for crops, 3,000 generations (or year) was the maximum duration of a bottleneck in crop domestication (Zhu et al. 2007). Second, 7,500 minus 3,000 was 4,500, which was around 5,000 years. In general, the range of k varied from 0.0001 to 7 and 150 combinations (scenarios) for each locus were investigated.

For each simulation at each locus, we calculated three diversity statistics, π, S, and Tajima's D from simulated *B. mori* data. Observed values of π, S, and Tajima's D from *B. mori* were compared with simulated values and fit of simulated data to observed data was assessed. To assess fit, we defined levels of acceptance corresponding to a range of 20% of π, S, or Tajima's D. Likelihood of each scenario for each locus was obtained by calculating proportion of 10,000 simulations that fit to the data (Weiss and Von Haeseler 1998). Multilocus likelihood values were calculated by multiplying across loci. This method was based on the assumption that loci were independent.

Real-Time PCR

Total RNA from integument among three main developmental stages (Day 5 of the 5th instar larva, Day 4 of pupation and the adult moth) in both *B. mori* and *B. mandarina* were extracted using TRIzol reagent (Invitrogen, China). The RNA samples were treated with DNase I to remove any contamination of genomic DNA, and then were reversely

transcribed with oligo (dT) and an Moloney murine leukemia virus reverse Transcriptase (Promega, China) according to the manufactures instructions. Real-time PCR was performed with the ABI Prism 7000 sequence detection system (Applied Biosystems, USA) with an SYBR Premix Ex Taq kit (Takara, China) using the manufacturer's recommended procedure to detect the expression levels of *TH* gene in both *B. mori* and *B. mandarina* at the same developmental stages and in the same sampled tissue (integument). The real-time PCR condition was denaturation at 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s, and 60 °C for 31 s. Reactions were run in triplicate using 1 μL of cDNA per reaction. The primers for *TH* and internal control genes are listed in supplementary table S2 (Supplementary Material online), and the eukaryotic translation initiation factor 4A (silkworm microarray probe ID: sw22934) was used as an internal control for normalization of sample loading (Dai et al. 2010).

Results

Identification of Melanin Synthesis Pathway Genes in *B. mori* Genome

All melanin synthesis pathway genes in *D. melanogaster* have orthologs in *B. mori* genome (supplementary table S3, Supplementary Material online). *Yellow-f* and *yellow-f2* have the same orthologous gene (annotated gene *BGIBMGA014224*) in the *B. mori* genome. The full-length cDNAs of some genes have been cloned and characterized in *B. mori*, such as *TH* (AB439286) (Futahashi et al. 2008), *ebony* (AB439000) (Futahashi et al. 2008), yellow (AB438999) (Futahashi et al. 2008), *Ddc* (AF372836), *Dat1* (DQ256382) (Tsugehara et al. 2007), and *tan* (AB499125) (Futahashi et al. 2010), which were in accordance with the results of TBlastN. The mRNA sequence of *yellow-f2* (NM_001043967) was also reported in *B. mori* (Xia et al. 2006). However, when we used its amino acid sequence as a query for searching the *D. melanogaster* genome and annotated genes in Flybase (Tweedie et al. 2009), we found that it has much lower identity with *yellow-f2* gene in *D. melanogaster* than the annotated gene *BGIBMGA014224*. A recent study also suggested that NM_001043967 should be renamed as *yellow-f4-3* rather than *yellow-f2* (Ferguson et al. 2011). So we still consider that *BGIBMGA014224* is the putative orthologous gene of both *yellow-f* and *yellow-f2*.

Nucleotide Diversity and Position Effects of Melanin Synthesis Pathway Genes

Sequences of the eight melanin synthesis pathway loci were obtained from 30 samples, including domesticated silkworm and wild silkworm, respectively. All loci contained at least one indel. After excluding indel polymorphisms, the length of aligned sequence for each locus varied from 858 bp to 7,964 bp, with a total length of 15,801 bp, including 4,583 bp of coding region and 11,218 bp of noncoding region (table 1). All loci contained both coding and noncoding sites, and there were no indels in coding regions. The data set of domesticated silkworm harbored 460 single nucleotide polymorphisms (SNPs) (one SNP per 34.4 bp), whereas there were 1,021 SNPs in wild silkworm taxa (one SNP per 15.2 bp). Of 460 domesticated silkworm SNPs, 59 were synonymous variants (the mutation sites not causing amino acids replacements), 26 were nonsynonymous variants (the mutation sites causing amino acid replacement), and 434 were silent variants (synonymous variants and mutation sites in noncoding region). Of 1,021 wild silkworm SNPs, 115 were synonymous variants, 43 were nonsynonymous variants, and 978 were silent variants.

The detailed statistics of sequence diversity for each locus were summarized in table 1. Estimates of total nucleotide diversity (π_{total}) ranged from 0.00031 (*black*) to 0.02891 (*tan*) with an average value of 0.01340 in *B. mori*, whereas π_{total} of *B. mandarina* varied from 0.00977 (*TH*) to 0.03496 (*tan*) with an average value of 0.02261. Similarly, Watterson's θ (θ_w) of *B. mori* ranged from 0.00069 (*black*) to 0.02885 (*tan*) with a mean value of 0.01337, whereas θ_w of *B. mandarina* varied from 0.01245 (*TH*) to 0.04157 (*tan*) with an average value of 0.02633. The π_{total} of *B. mandarina* was significantly higher than that of *B. mori* (Wilcoxon signed-rank test, $P = 0.01$, degree of freedom [df] $= 7$), as well as θ_w (Wilcoxon signed-rank test, $P = 0.01$, df $= 7$), indicating that about 40.7% or 49.2% of the genetic diversity of wild silkworm was lost in domesticated silkworm measured by π_{total} or θ_w.

Domestication leads to reduction of genetic diversity in domesticated species because of artificial selection and bottleneck. In principle, bottleneck should have an equal effect on all loci. However, it can be seen from figure 1*b* that the reduction of genetic diversity in domesticated silkworm varies greatly among melanin synthesis pathway genes. The ratio of domesticated silkworm nucleotide diversity to wild silkworm diversity (π_{ds}/π_{ws}) at *black* and *TH* is 2% and 10%, respectively. At the rest of the six loci, the ratio is more than 30%. Moreover, to exclude the alternative demographic effect from the bottleneck, we sequenced one intergenic region unlinked to any melanin synthesis pathway genes for wild and domesticated silkworms (see Materials and Methods). The result showed that there is similar nucleotide diversity between *B. mori* and *B. mandarina* (supplementary fig. S3, Supplementary Material online), and that the ratio of nucleotide diversity of domesticated silkworm to that of wild silkworm diversity (π_{ds}/π_{ws}) is 86%. The genetic retention measured by π_{ds}/π_{ws} has been used to correct for functional constraints (Clark et al. 2004). For neutral genes in maize, the genetic retention ranges 60–80%, whereas a lower ratio is indicator of selection (Zhang et al. 2002; Hufford et al. 2007). Therefore, the low levels of genetic retention at loci *black* and *TH* suggested that these loci might experience artificial selection during domestication of silkworm.

To confirm the reduced variability of *TH* was due to selection rather than a hitchhiking effect, we investigated the nucleotide polymorphism of the linked upstream and downstream genomic regions of *TH* gene in domesticated silkworm. Previous studies suggested that genetic diversity of the genomic regions under selection showed a *V*-shaped

Table 1. Summary statistics of sequence diversity of *B. mori* and *B. mandarina*.

Locus	Specie	*n*	Length	*S*	π_{total}	θ_w	S_{coding}	π_a/π_s	π_{silent}	θ_{silent}	R_m	*R*	Tajima's *D*	$D^{※}$	$H^{※}$
TH	*Bombyx. mori*	15	7,964	55	0.00097	0.00212	11	0.806	0.00097	0.00211	0	0.0069	−2.34936***	−3.31159**	0.74633
	B. mandarina	15	7,857	318	0.00977	0.01245	25	0.072	0.01153	0.01509	38	0.0332	−0.95510	−1.56128	1.00088
black	*B. mori*	15	873	2	0.00031	0.00069	0	0.000	0.00058	0.00135	0	0.0000	−1.49051	−1.87275	0.37737
	B. mandarina	15	868	55	0.01569	0.01949	23	0.129	0.02570	0.03129	6	0.0111	−0.84160	−0.45015	1.06767
Ddc	*B. mori*	15	1,072	29	0.01039	0.00832	6	0.000	0.01407	0.01110	2	0.0100	1.04746	0.43095	1.09539
	B. mandarina	15	1,023	83	0.01965	0.02555	9	0.000	0.02686	0.03493	7	0.0839	−0.92740	−0.97162	0.94864
ebony	*B. mori*	15	977	25	0.00639	0.00787	8	0.095	0.00925	0.01070	0	0.0000	−0.78012	0.23919	1.14160
	B. mandarina	15	941	67	0.02203	0.02288	18	0.093	0.03090	0.03113	6	0.0479	0.02695	0.44536	1.17067
yellow-f	*B. mori*	15	1,274	76	0.02112	0.01835	44	0.057	0.04451	0.03575	6	0.0051	0.65841	0.40860	1.11016
	B. mandarina	15	1,270	105	0.02443	0.02543	55	0.056	0.05195	0.04861	20	0.0799	−0.17164	−0.46426	1.14446
Dat1	*B. mori*	15	978	46	0.01789	0.01649	0	0.000	0.02146	0.01978	0	0.0000	0.36572	1.43992*	1.48614
	B. mandarina	15	858	99	0.02621	0.03208	9	0.046	0.02584	0.03298	7	0.0374	−0.69208	−1.14657	1.05784
yellow	*B. mori*	15	1,533	121	0.02119	0.02427	11	0.082	0.02271	0.02685	3	0.0000	−0.55849	−0.90830	1.13084
	B. mandarina	15	1,438	146	0.02813	0.03122	12	0.203	0.03071	0.03564	14	0.0137	−0.43723	−1.10449	0.98212
tan	*B. mori*	15	1,130	106	0.02891	0.02885	5	0.000	0.03319	0.03294	14	0.0021	0.00901	0.82826	1.25425
	B. mandarina	15	1,095	148	0.03496	0.04157	7	0.040	0.04104	0.05365	13	0.0110	−0.70129	−0.81962	1.05545

***$P < 0.001$; **$P < 0.02$; *$P < 0.05$; *n*: sample size; Length: the number of sites aligned sequences, excluding gaps. *S*: the number of segregating sites; π_{total}: the mean number of nucleotide differences per site; θ_w: watterson's estimator of $4N_e\mu$; R_m: estimates of minimum number of recombination events (Hudson and Kaplan 1985); *R*: the estimate of the population recombination parameter per site (Hudson 1987); S_{coding}: number of segregating sites in coding regions; π_a: average pairwise differences calculated on nonsynonymous sites; π_s: average pairwise differences based on synonymous sites; π_a/π_s: ratio of π_a to π_s; π_{silent}: π values for silent sites (synonymous and noncoding sites); θ_{silent}: θ values for silent sites (synonymous and noncoding sites); $D^{※}$, Fu and Li's D; $H^{※}$, Fay and Wu's H.

distribution (Schlötterer 2003). Indeed, our results demonstrated such *V*-shaped distribution of genetic diversity around the *TH* gene (fig. 1*c*). Moreover, *H* test of Fay and Wu also did not detect a hitchhiking effect on *TH* gene (table 1). Therefore, we can conclude that the selection indeed occurred in *TH* gene rather than in its linked regions.

In the melanin synthesis pathway (fig. 1*a*), the eight core structural genes can be divided into two groups. *TH* and *black* encode the two earliest-acting enzymes in the pathway, processing tyrosine and uracil into the melanin precursors, respectively, whereas *Ddc*, *yellow*, *yellow-f*, *Dat1*, *ebony*, and *tan* are the downstream genes. To study the relationship between nucleotide polymorphism and the position of the melanin synthesis pathway genes, we evaluated whether nucleotide polymorphisms are affected by the position of the genes in the pathway. An average value of nucleotide diversity of upstream genes (*black* and *TH*) in *B. mori* is 0.00064, whereas an average value of diversity of downstream genes is 0.01765. The upstream genes have significantly lower nucleotide polymorphisms than the downstream genes in *B. mori* (Student's *t*-test, $P = 0.0310$, df = 6). Similarly, an average value of nucleotide diversity of the upstream genes in *B. mandarina* is 0.01273 and significantly smaller than the corresponding average value of 0.02590 at the downstream loci (Student's *t*-test, $P = 0.0207$, df = 6). This suggested that the position of melanin synthesis pathway does affect the distribution of selective constraint along it. More specifically, the upstream genes have higher levels of selective constraint than the downstream genes.

Neutrality Test

To check whether the nucleotide polymorphism data fit the neutral equilibrium model, we performed Tajima's *D* test (Tajima 1989) and *D* test of Fu and Li (Fu and Li 1993) for each gene. Significant negative Tajima's *D* value was only observed at the *TH* locus in *B. mori*, where the neutral model was rejected at $P < 0.001$ (table 1). Significant negative *D* value of Fu and Li ($P < 0.02$, table 1) was found at the *TH* locus. A significant positive value of Fu and Li ($P < 0.05$, table 1) was also found at *Dat1* locus. However, it was not significant after the Bonferroni correction for multiple comparisons. A significant negative Tajima's *D* test statistic or D test statistic of Fu and Li indicates an excess of low frequency of polymorphism, which is consistent with directional selection or population expansion (Tajima 1989). It is interesting to note that four of eight loci in *B. mori* and seven of eight loci in *B. mandarina* had negative Tajima's *D* values (table 1), and three of eight loci in *B. mori* and seven of eight loci in *B. mandarina* had negative values of Fu and Li (table 1), suggesting that both domesticated and wild silkworms were under directional selection. We found that four loci (*Ddc*, *Dat1*, *yellow-f*, and *tan*) in *B. mori* and one locus (*ebony*) in *B. mandarina* had positive Tajima's *D* and *D* values of Fu and Li, and another locus (*ebony*) in *B. mori* also had positive values of Fu and Li, which indicated balancing selection or a population subdivision event because the long-term maintenance of distinct haplotypes can lead to an excess of intermediate frequency variants. *H* statistic of Fay and Wu (Fay and Wu 2000) was calculated for the whole sequence of all locus in both *B. mori* and *B. mandarina* population. The *H* statistic determines the level of high-frequency variants to detect hitchhiking. However, none of the *H* statistics were significant (table 1).

Multilocus HKA test was performed to examine whether levels of polymorphism and divergence across loci would be correlated, as expected under the neutral model of molecular evolution. No significant departure from neutrality was detected between *B. mori* and *B. mandarina* (chi-square = 10.28, $P = 0.741$, df = 14).

We also performed HKA test on the 5′-noncoding, coding, intron, and 3′-noncoding sequence to investigate whether a particular region of *TH* may have been the target of selection because of the Tajima's *D* test significant result.

However, neutrality was still not rejected between *B. mori* and *B. mandarina* (chi-square = 7.41, $P = 0.492$, df = 6). In fact, there were no fixed substitutions in coding regions between species except for *TH* that just had one fixed synonymous substitutions and no fixed nonsynonymous substitution, so we could not perform the MK test between *B. mori* and *B. mandarina*. Such absence of fixed differences is also the case in Drosophila (Inomata and Yamazaki 2002), sorghum (White et al. 2004), and rice (Zhu et al. 2007). Only 5,000 years after domestication may be main reason for this result (Xiang et al. 2005). Actually, *B. mori* and *B. mandarina* can be crossed each other and the hybrid progeny are fertile although they were defined as two distinct species (Banno et al. 2004). When the MK test was performed for *TH* gene between *B. mori* and *Ernolatia moorei* (one species from the same family Bombycidae), we found an excess of replacement changes ($G = 22.832$, $P < 0.001$). Similar result was also obtained for *TH* gene between *B. mandarina* and *E. moorei* ($G = 8.779$, $P < 0.01$). These significant results may be false positive because *E. moorei* is relatively distant from both *B. mori* and *B. mandarina*. We tried our best to find an appropriate outgroup for the HKA and MK analyses. However, the problem is that there are a few extant close species to *B. mori*. That the efficacy of multilocus HKA and MK tests is low may be due to lack of a proper outgroup. Thus, coalescent simulation was performed instead in this study.

Coalescent Simulation

Likelihood curves using π as fit standard for each locus with $d = 1{,}000$ were displayed in figure 2*a*. Six of eight loci harbored noticeable peaks, suggesting that those loci were well fitted by a bottleneck model, and the maximum likelihood values reached peaks within $N_b < 5{,}000$. However, likelihood surfaces of *TH* and *black* just showed very slight peaks and the likelihood values for these curves were smaller than 0.05, especially for *TH*, indicating a departure from the bottleneck model (Haudry et al. 2007).

Multilocus likelihood values were calculated to find the optimal *k* value. *TH* and *black* were excluded from multilocus likelihood analysis, and thus, multilocus likelihood curves based on six neutral loci were obtained. Previous studies also excluded the loci deviating from the bottleneck model in multilocus likelihood analysis (Tenaillon et al. 2004; Zhu et al. 2007). In figure 2*b*, a multilocus likelihood surface based on π with $d = 1{,}000$ is displayed.

To identify loci that may experience selection using bottleneck model, likelihood ratio tests (LRT) were performed based on $\chi^2 = -2\ln(L_0/L_1)$, in which L_0 is the likelihood value for locus with the optimal bottleneck stringency *k* value ($k = N_b/d$) when multilocus likelihood reached to the maximum value, whereas L_1 is the maximum likelihood value for each locus. Figure 2*b* showed that for six neutral loci, multilocus likelihood reached to the maximum value when $N_b = 1{,}500$ ($k = 1.5$) under the condition with $d = 1{,}000$, and in figure 2*a*, likelihood value for *TH* with $N_b = 1{,}500$ was 0.00025. By contrast, the maximum likelihood value of *TH* was 0.00481 with $N_b = 300$ (fig. 2*a*). Then LRT for *TH* with $N_b = 1{,}500$ based on π indicated that these two likelihood values were significant different ($\chi^2 = 5.91$, $P = 0.015$, df = 1). Similarly, likelihood value for *black* with $N_b = 1{,}500$ was 0.00005, whereas the maximum likelihood value of *black* was 0.0363 with $N_b = 100$ (fig. 2*a*). Then LRT for *black* with $N_b = 1{,}500$ based on π indicated that these two likelihood values were significant different ($\chi^2 = 13.17$, $P = 0.00028$, df = 1). In addition, LRT for other six loci were not significant at all. Therefore, *TH* and *black* did not fit this bottleneck model and might be selection targets during domestication.

LD and recombination

The results of the LD with nucleotide pairwise distance are summarized in supplementary figure S4 (Supplementary Material online). r^2, which is squared allele frequency correlation, was pooled from eight genes and used to fit logarithm curves across distances for *B. mori* and *B. mandarina*, respectively. It is obvious that fitting curve in *B. mori* was above that in *B. mandarina*. Both curves of LD decay rapidly within 200 bp and drops slowly when distance is >200 bp. The average r^2 in *B. mandarina* drops rapidly for the first 200 bp to a value of 0.22 and a value of 0.10 within 1,600 bp, whereas the value in the domesticated silkworm declines much slowly to a value of only 0.54 within 200 bp and remains a value of >0.40 by 1,600 bp. The higher level of LD in *B. mori* is consistent with an increase in LD due to domestication.

We also analyzed intralocus LD for *TH* and *black* that may show selection signatures. However, patterns of LD in these loci were different from that in supplementary figure S4 (Supplementary Material online). A curve of LD decay for *B. mandarina TH* is shown in supplementary figure S5 (Supplementary Material online). The curve drops to 0.20 within 150 bp and to 0.10 within 1,400 bp; whereas there is no pairwise comparison sites and LD cannot be analyzed in *B. mori*. Similarly, curve of LD decay for *B. mandarina black* is shown and drops to 0.15 within 600 bp, whereas there is no pairwise comparison sites and LD cannot be analyzed in *B. mori*.

Estimates of population recombination parameter *R* (Hudson 1987) range from 0 to 0.0069 (mean: 0.0035 ± 0.0049) in *B. mori* and from 0.011 to 0.0839 (mean: 0.0398 ± 0.0293) in *B. mandarina* (table 1). Similarly, the minimum numbers of recombination events (R_m) range from 0 to 14 in *B. mori* (mean: 3.1 ± 4.9) and from 6 to 38 (mean: 13.9 ± 10.9) in *B. mandarina*.

Haplotype Structure and Phylogenetic Analyses

Previous studies suggested that if a single preferred allele at a domestication locus becomes fixed in domesticated species and subsequently does not experience any mutations, then the extant preferable haplotypes forms a single clade in phylogenetic analysis including alleles from the corresponding progenitor (Wang et al. 1999; Clark et al. 2004; Sigmon and Vollbrecht 2010). In contrast, alleles at a neutral locus in domesticated species scatter among its progenitor alleles on a phylogenetic tree (Goloubinoff et al. 1993; Hanson et al. 1996; Hilton and Gaut 1998; Sigmon and Vollbrecht 2010).

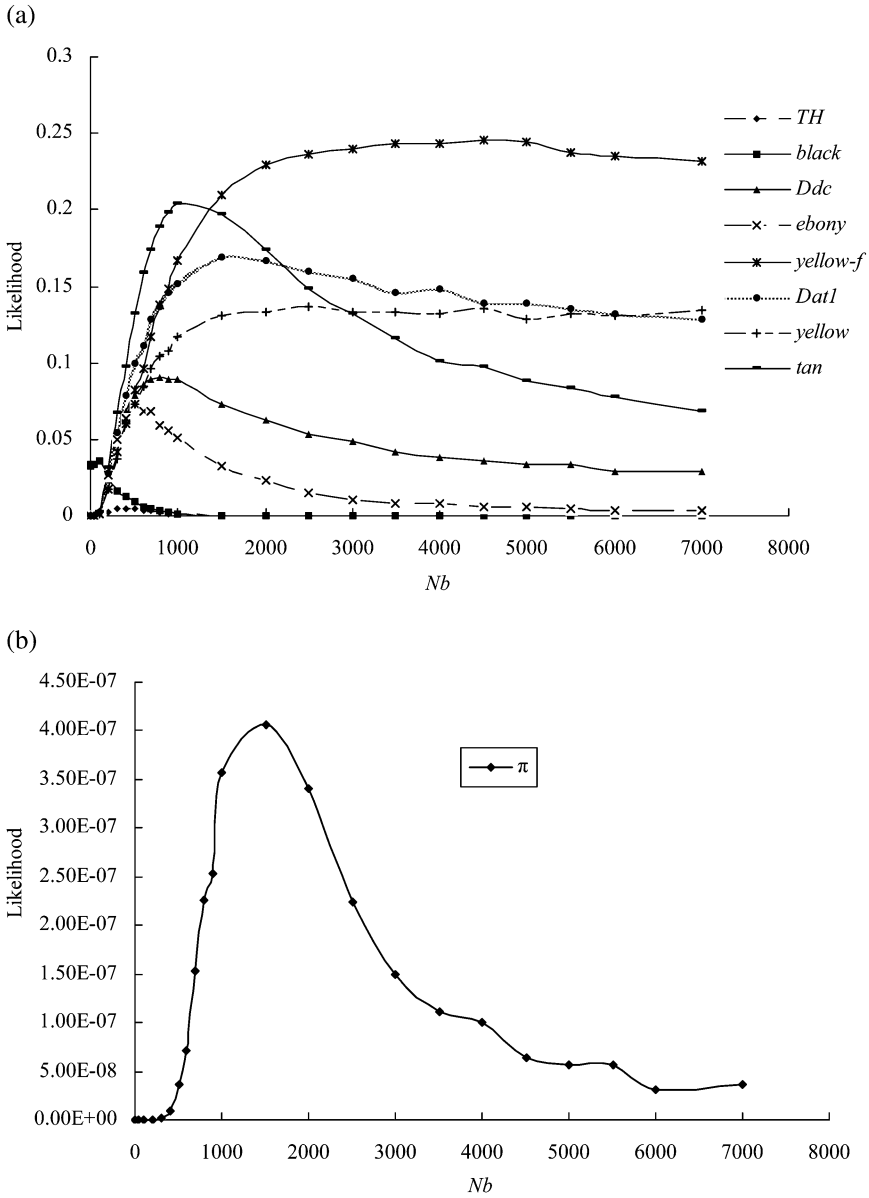

Fig. 2. (*a*) Respective likelihood curves of eight loci. The likelihood curves are from a coalescent simulation, *TH* and *black* depart significantly from the bottleneck model. (*b*) joint likelihood surfaces among six neutral loci using nucleotide diversity (π) as fit standard based on $d = 1{,}000$.

Thus, analyses of phylogeny and haplotype structure provide information about whether loci are domestication loci or not.

The results of the haplotype structure and phylogenetic analyses indicated that almost all loci show the patterns consistent with a neutral locus except for *TH*. So, we will just discuss the results for *TH* locus here. The median-joining network analysis for *TH* gene showed strikingly different genealogical patterns between *B. mori* and *B. mandarina* (fig. 3*a*). There were 15 haplotypes in wild silkworm, and the haplotypes exhibited several divergent clusters. Thirteen haplotypes existed in domesticated silkworm and three silkworm strains (D1, D11, and D13) shared common haplotype. All the haplotypes in domesticated silkworm were confined to a single cluster for *TH* gene, which is different from the other seven genes, such as *tan* (supplementary fig. S6, Supplementary Material online).

To root the phylogenetic tree reconstructed by *TH* sequences, we sequenced the corresponding homolog of a related species of *E. moorei* from the same family of *B. mori* as an outgroup. Phylogenetic analysis for *TH* locus clearly suggested that all domesticated silkworm strains fell predominantly into one clade and were a subset of the wild silkworm, although there is a long branch to outgroup (fig. 3*b*). It should be pointed out that *E. moorei* is a relatively close and extant species to *B. mori*. Taken together, both haplotype structure and phylogenetic analyses showed that *TH* may be a candidate domestication locus.

Nuleotide Diversity of *TH* Locus

To visualize variation in polymorphism throughout the 7.8-Kb *TH* region, a sliding window analysis was performed. The results showed that *TH* experienced selection

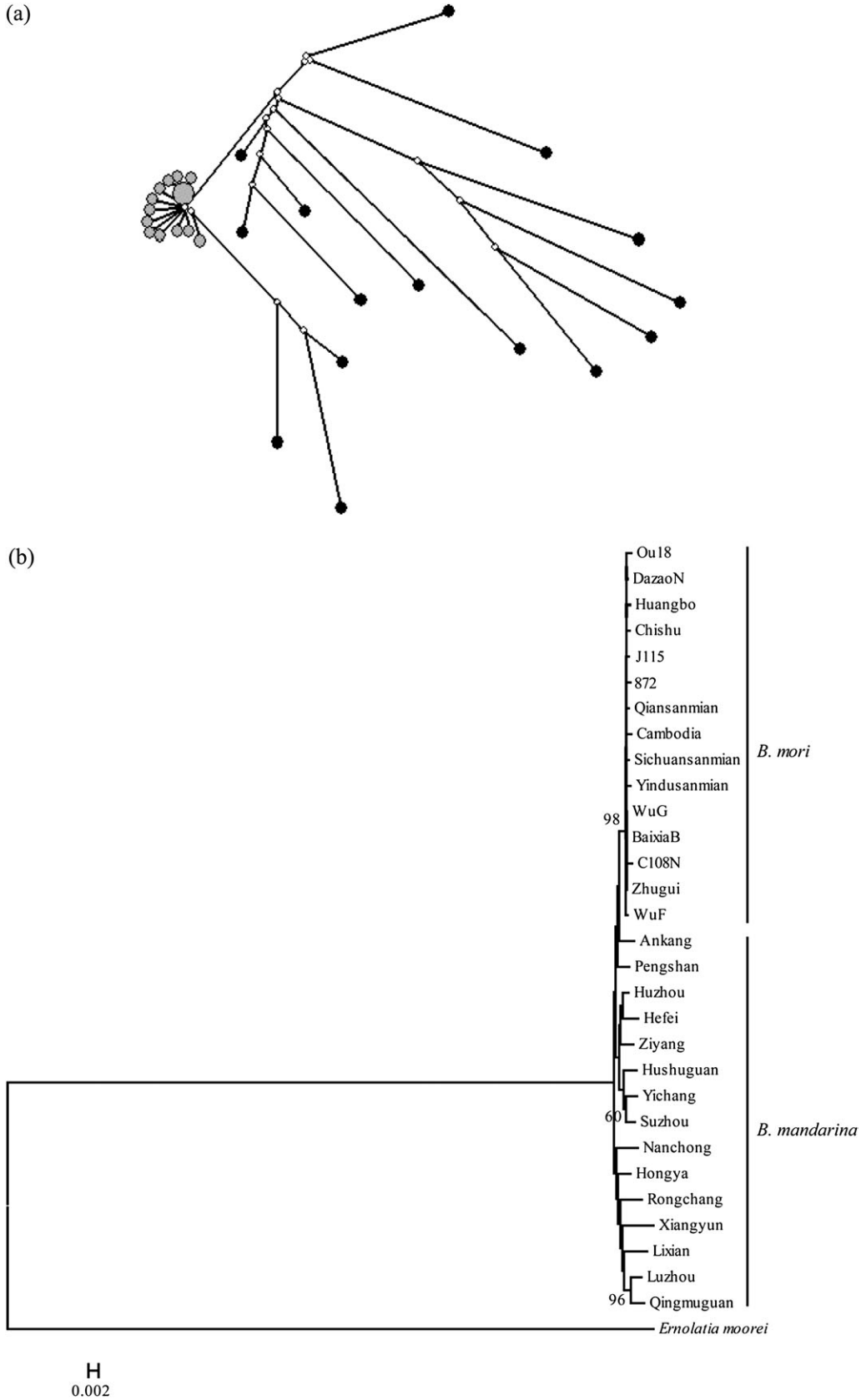

Fig. 3. Median-joining networks based on the haplotypes (*a*) and NJ tree (*b*) of *TH* gene. Gray nodes represent haplotypes from domestic silkworms, whereas black ones indicated haplotypes from wild silkworms. The homologous sequence of *Ernolatia moorei* from the same family Bombycidae was used as outgroup and bootstrap values are shown for each node when >50%.

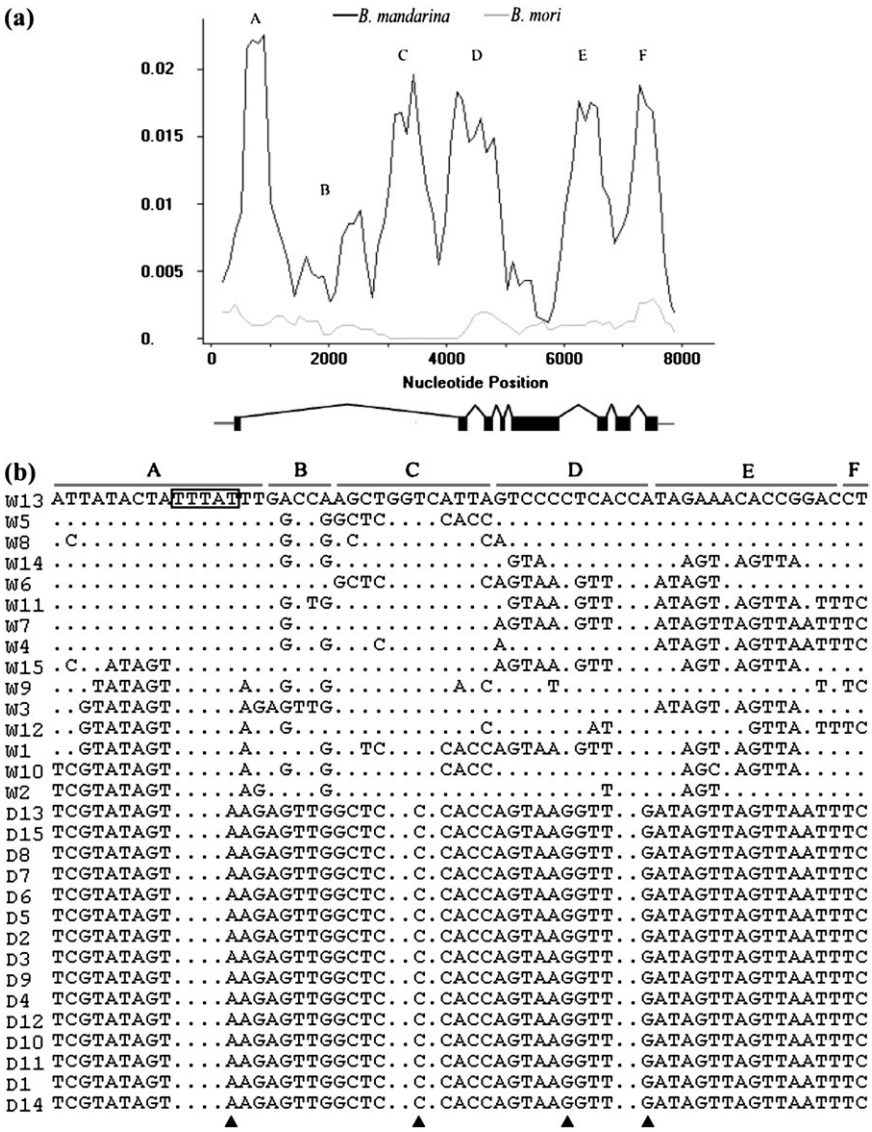

Fig. 4. Patterns of nucleotide diversity at *TH* locus. (*a*) Sliding window analysis of nucleotide diversity (π) at *TH* is shown for *B. mandarina* (black line) and *B. mori* (gray line). π was calculated for segments of 400 bp at 100-bp intervals. Within the gene cartoon, the black rectangles indicate coding regions, and the black lines represent noncoding regions. (*b*) Nucleotide mutations and fixed sites in the *TH* gene. ▲ designates the fixed sites, and the open rectangle indicates the motif of Abd-B–binding sites.

throughout the entire gene sequence, especially in noncoding region (fig. 4*a*, supplementary table S4 and fig. S7a, Supplementary Material online). There were five regions (A–F) with high nucleotide diversity difference between *B. mori* and *B. mandarina* (fig. 4*a*). Sequence alignment suggested that there were some sites with the same nucleotide sequence among the five regions in domesticated silkworm, whereas high nucleotide polymorphism existed in wild silkworm (fig. 4*b*). There were three fixed sites between *B. mori* and *B. mandarina* in the introns of *TH* gene, whereas there was only one fixed substitution in the coding region (fig. 4*b*).

Four fixed sites between *B. mori* and *B. mandarina* are located in the position 854, 3,349, 4,608, and 5,631 of *TH* locus, respectively. The former two fixed sites are located in the first intron, and the third one is located in the second intron. Although the last fixed site is located in the exon, that codon (CCA) mutated to codon (CCG) does not cause an amino acid change (fig. 4*b*). To investigate the contents of surrounding regions of the former three fixed sites between *B. mori* and *B. mandarina*, the MatInspector professional program (www.genomatix.de) was used to predict transcription factor binding sites. Interestingly, we found that there was a five-base-pair motif TTTAT around the first fixed site in all *B. mandarina* strains, whereas the five-base-pair motif TTTAT mutated to TTTAA and fixed in all *B. mori* samples (fig. 4*b*). Abd-B and paralogous HOX proteins have been shown to preferentially bind to the five-base-pair motif TTTAT or TTTAC, and with some affinity to the shorter motif TTAT in Drosophila (Ekker et al. 1994). In the second fixed site, core sequence GGTC for retinoid x receptor (RXR) heterodimer binding sites mutated to GGCC in *B. mori* (fig. 4*b*). In the third fixed site, there was no gain or loss of transcription factor binding sites for the mutation from C to G in *B. mori* (fig. 4*b*). To date,

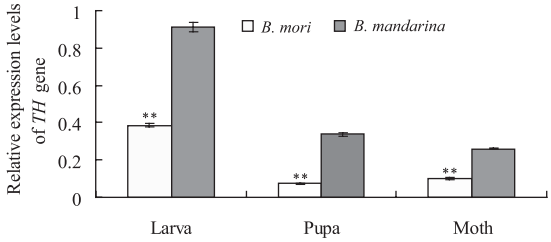

FIG. 5. Comparison of the relative expression levels of *TH* gene between *B. mori* and *B. mandarina* by Real-time PCR at three developmental stages the 5th day of 5th instar larva, the 4th day of pupation and the adult moth. Expression levels were normalized to the expression of the eukaryotic translation initiation factor 4A gene. The number of assay replicates is 3 ($n = 3$). $^{**}P < 0.0001$, Student's *t*-test.

there has not been any report to show association between RXR heterodimer binding sites and insect pigmentation.

The loss of a putative Abd-B–binding site (motif TTTAT vs. TTTAA) observed in *B. mori* may be the result of a small sample size. To further confirm this result in a larger sample, we sequenced the region of a putative Abd-B–binding site for another 60 domesticated silkworm strains and 7 wild silkworm samples. The result indeed showed that all domesticated silkworm strains had TTTAA haplotype, whereas TTTAT existed in all wild silkworm (fig. S8, Supplementary Material online).

Real-Time PCR

Previous studies have shown that mutational inactivation of a key Abd-B–binding site could cause male-specific pigmentation to be lost in Drosophila (Jeong et al. 2006; Williams et al. 2008) and use of RNAi with *TH* gene prevented pigmentation and hatching, and feeding of a *TH* inhibitor also suppressed larval pigmentation in *B. mori* (Liu et al. 2010). We compared the expression of *TH* gene between *B. mori* and *B. mandarina* by real-time PCR at three developmental stages: Day 5 of the 5th instar larvae, Day 4 of pupation and the adult moths. The results showed that the relative expression levels of *TH* gene in *B. mori* were significantly lower than that in *B. mandarina* among three developmental stages (Student's *t*-test, $P < 0.0001$) (fig. 5), which is consistent with light body color of domesticated silkworm relative to wild silkworm.

Discussion

Genetic Diversity

There was a large decrease in nucleotide diversity during domestication of silkworm, *B. mori*. The mean of nucleotide diversity in the eight melanin synthesis pathway genes in *B. mori* is $\pi_{total} = 0.01340$ and $\theta_w = 0.01337$. Relative to *B. mori*, wild silkworm, *B. mandarina*, harbors much higher nucleotide diversity with an average value of $\pi_{total} = 0.02261$ and $\theta_w = 0.02633$. About 40.7% or 49.2% of the genetic diversity of wild silkworm was lost in domesticated silkworm measured by π_{total} or θ_w. These results are quite different from that of a recent study using Solexa sequencing (Xia et al. 2009) in which only 17% loss of genetic variation measured by θ was observed in *B. mori* during the domestication. Our results about loss of genetic diversity are comparable with that observed in other domesticated species, for example, 43% loss in maize (Wright et al. 2005), 50–60% loss in sunflower (Liu and Burke 2006), 69–84% loss in wheat (Haudry et al. 2007), and 57–69% loss in rice (Caicedo et al. 2007). It is well known that *B. mori* is a completely domesticated insect that has experienced intense artificial selection and inbreeding (Goldsmith et al. 2005). *B. mori* would lose a large amount of genetic diversity. The nucleotide diversity in *B. mori* was found to be at similar level measured as θ between the two studies (0.01337 vs. 0.01080: this study vs. Solexa), suggesting that domesticated silkworm samples used in this study are representative although the sample number is relatively small. However, nucleotide diversity in *B. mandarina* between the two studies is different (0.02633 vs. 0.0130: this study vs. Solexa). Therefore, underestimation of genetic diversity loss in domesticated silkworm in Solexa sequencing study may be due to underestimate of the nucleotide diversity in wild silkworm because of sampling and limitation of Solexa sequencing.

First, the whole genome scan used only 11 wild silkworm samples (Xia et al. 2009), whereas this study used 15 wild silkworm samples. Ten of 15 wild samples in this study were collected from the same areas as in the research using Solexa sequencing (Xia et al. 2009). In our data set, average θ value is 0.02323 for ten wild samples used in previous study using Solexa sequencing (Xia et al. 2009), which is significantly lower (Wilcoxon signed-rank test, $P < 0.05$, $df = 7$) than that (0.02633) for 15 samples. Therefore, different sampling strategies for *B. mandarina* may be in part responsible for the difference in θ value.

The other reason may be that Solexa sequencing underestimated the level of nucleotide diversity in wild silkworm. Sequence reads using Solexa sequencing are usually short (~75 bp). For some mapping algorithms, sequence reads with more than one or two differences from a reference genome will not be placed (Li et al. 2008). As Pool et al. (2010) pointed out, this makes the mapping of alleles that are different from the reference genome less probable than for a reference-matching allele, causing a bias in allele frequency toward the allele found in the reference sequence. As a result, it may underestimate the level of nucleotide diversity for divergent sequences. Because one reference genome sequence is available only for *B. mori* and used for calling SNPs in both *B. mori* and *B. mandarina*, underestimation of nucleotide diversity in *B. mandarina* can be envisioned. Thus, the available reference genome sequence of *B. mandarina* will be a key factor to correct this bias in future.

Strikingly, *TH* that was shown to be under artificial selection in this study was not found to be under selection in a recent study using Solexa sequencing (Xia et al. 2009). This indicates the necessity of use of more reliable sequencing in fine detection of artificial selection signature for genomic regions of interest in domesticated organisms.

It is worth noting that *TH* gene is located on Z chromosome and the reduced variability of *TH* gene may be because sex chromosome has a smaller population size. Actually, this is not obviously a strong factor in the present data because the most variable gene, *tan*, is the only other one that is sex linked.

Association between Variation and Position in the Silkworm Melanin Synthesis Pathway

Because Rausher et al. (1999) showed that the upstream genes in the anthocyanin pathway evolve more slowly than downstream genes, several papers on this pattern of evolutionary rate variation in metabolic pathway have been published (Olsen et al. 2002; Riley et al. 2003; Flowers et al. 2007; Ramos-Onsins et al. 2008; Livingstone and Anderson 2009; Ramsay et al. 2009). An FCH was proposed to explain a pattern of evolutionary rate variation: Mutations in the upstream genes would generate greater pleiotropic effects than those in downstream genes in the pathway, being therefore more possibly to produce a deleterious effect (Rausher et al. 1999). Some studies did support this hypothesis (Riley et al. 2003; Livingstone and Anderson 2009; Ramsay et al. 2009). However, the hypothesis was not supported by some other studies (Olsen et al. 2002; Flowers et al. 2007; Ramos-Onsins et al. 2008). In addition, almost all the studies used divergence data between species except for one that used multilocus analysis of variation in phenylpropanoid pathway genes of *Arabidopsis thaliana* (Ramos-Onsins et al. 2008). Nevertheless, the polymorphism data from *Arabidopsis* did not support FCH (Ramos-Onsins et al. 2008). In this study, we found that the upstream genes have significantly lower levels of genetic diversity than the downstream genes in the silkworm melanin synthesis pathway, suggesting that the upstream genes were subject to stronger selective constraint than the downstream genes. Thus, our study for the first time used the genetic variation data to support the FCH.

It is well known that gene duplication creates homologous redundancy, resulting in relaxation of functional constraint (Ramsay et al. 2009). Thus, gene copy number could explain why certain lineages show FCH and others do not. If homologous redundancy is a reason for our results, most downstream genes should be duplicates. We examined the gene copy number for each of the silkworm melanin synthesis genes and found that all the genes except for *yellow* genes are single copy genes. Therefore, gene copy number cannot explain FCH observed in this study. In other words, FCH does hold in the silkworm melanin synthesis pathway.

Evidence of Selection

To investigate whether artificial selection contributes to color pattern diversification, we examined nucleotide diversity within and between *B. mori* and *B. mandarina* in eight melanin synthesis pathway genes, which play a key role in cuticular pigmentation of insect. For *black* locus, 98.0% of nucleotide diversity in wild silkworm lost in domesticated silkworm measured by π_{total} (table 1, fig. 1*b*), and there is no pairwise comparison and LD cannot be analyzed for *black* in *B. mori*. Coalescent simulation test showed that *black* also deviated from the bottleneck model used in this study. However, Tajima's *D* test for *black* was not significant, and haplotype structure and phylogenetic analysis showed that domesticated silkworm alleles scattered among wild silkworm. For *TH* locus, there was a lot of evidence for artificial selection. First, 89.6% of nucleotide diversity for *TH* in wild silkworm lost in domesticated silkworm measured by π_{total} (table 1, fig. 1*b*), and the standing variation in *B. mori* was mainly part of natural variation *B. mandarina* had. More definitely, among parsimony informative sites at *TH* locus in *B. mori*, there was just one new variation that was not from *B. mandarina* and accumulated after domestication. There were 126 parsimony informative sites for *TH* in *B. mandarina* but 118 of which fixed in *B. mori* in a haplotype pattern. At other eight sites, *B. mori* showed singleton variable sites, and all eight mutations were from *B. mandarina* (supplementary fig. S7b, Supplementary Material online). These results reflect the signature of selection during the domestication. Moreover, LD cannot be analyzed for *TH* in *B. mori* because there are no pairwise comparisons.

Second, significant negative Tajima's *D* value was only observed at the *TH* locus for *B. mori*, where the neutral model was rejected at $P < 0.001$ (table 1). A significant negative Tajima's *D* test statistic indicates an excess of low frequency of polymorphism, which is consistent with directional selection or population expansion (Tajima 1989). Third, coalescent simulation test showed that *TH* deviated from the bottleneck model and was selection target during the domestication. The effect of coalescent simulation by neutrality test is strong, which is also used in analysis of crop evolution (Tenaillon et al. 2004; Wright et al. 2005; Zhu et al. 2007; Li et al. 2009). Fourth, haplotype structure and phylogenetic analyses of *TH* sequences suggested that all domesticated silkworm strains fell predominantly into one clade and were subsets of wild silkworm. Last, both demographic effects from a bottleneck and the hitchhiking effect were excluded. Taken together, *TH* is a likely candidate domestication locus in silkworm.

Moreover, noncoding region of *TH* may have been a target of selection (fig. 4*a*, supplementary table S4 and fig. S7a, Supplementary Material online). Similar effects of selection were found in *tb1* and *ramosa1* loci during maize domestication (Wang et al. 1999; Sigmon and Vollbrecht 2010). Notably, a potential Abd-B–binding site TTTAT in *B. mandarina* wild silkworm mutated to TTTAA in *B. mori* during the domestication. Abd-B directly activates the expression of *yellow* gene in the male *D. melanogaster* A5 and A6 segments through binding sites (TTTAT or TTTAC) in a specific *cis*-regulatory element, and mutational inactivation of a key Abd-B–binding site could cause male-specific pigmentation to be lost (Jeong et al. 2006; Williams et al. 2008). Real-time PCR indeed showed that the relative expression levels of *TH* gene in *B. mandarina* were significantly higher than that in *B. mori* at three developmental stages (Student's *t*-test, $P < 0.0001$) (fig. 5), which is consistent with light body color of domesticated silkworm relative to wild silkworm. TH is the rate-limiting

enzyme of melanin synthesis pathway in insects, and the activity of TH directly affected insect body color. In Drosophila, null mutation of *TH* gene resulted in unpigmented embryos and loss of the ability to hatch (Pendleton et al. 2002). In the swallowtail butterfly, *Papilio xuthus*, inhibition of TH activity prevented in vitro pigmentation completely (Futahashi and Fujiwara 2005). In *B. mori*, use of RNAi with the *TH* gene prevented pigmentation and hatching, and feeding of a TH inhibitor also suppressed larval pigmentation (Liu et al. 2010). Significant difference of expression level of *TH* gene may be a reason for body color change between *B. mori* and *B. mandarina* during silkworm domestication. However, further studies are required to confirm whether the different expression patterns of *TH* gene are caused by mutation of Abd-B–binding sites.

It has been proposed that morphological evolution relies predominantly on changes in the architecture of gene regulatory networks and in particular on functional changes within *cis*-regulatory elements (Carroll 2000, 2005; Prud'homme et al. 2007). Recent studies have provided direct evidence for the role of *cis*-regulatory element change in morphological evolution (Gompel et al. 2005; Jeong et al. 2006; Prud'homme et al. 2006; Williams et al. 2008; Wittkopp et al. 2008a, 2008b). Insect body color could evolve by altering activity or spatial deployment of transcription factors that regulate melanin synthesis pathway genes and/or by change in *cis*-regulatory elements of melanin synthesis pathway genes themselves. The data presented in this study demonstrated that *TH* of the silkworm melanin synthesis pathway genes underwent artificial selection during the domestication; furthermore, selection targets are located in the first intron. The target is likely a binding site of the Abd-B–transcriptional factor. The expression data of *TH* gene is consistent with the scenario of artificial selection for light body color of domesticated silkworm. Thus, we speculated that the loss of a key Abd-B–binding site at the *TH* locus may contribute to the body color transition from *B. mandarina* to *B. mori*. Taken together, our results emphasize the exceptional role of gene expression regulation in morphological transition of domesticated animals (Carroll 2000, 2005).

Supplementary Materials

Supplementary tables S1–S4 and figures S1–S8 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Dr Zhou-He Du for help in collecting wild silkworm samples and all other members of Zhang's group for their laboratory assistance. We are also grateful to three anonymous reviewers for a number of insightful comments that improved this manuscript. We thank the associate editor, Dr Adriana Briscoe, for her help in polishing the English of our manuscript. This work was supported by the Hi-Tech Research and Development (863) Program of China (2006AA10A117), a grant from National Science Foundation of China (No. 30671587), and the Program of Introducing Talents of Discipline to Universities (B07045). The sample of *Ernolatia moorei* used in this study was collected by Stefan Naumann and kindly provided by Charles Mitter and Jerome Regier. They were supported by grants from the U.S. National Science Foundation (N.S.F.—USA no. 0531769 and 0531626; DEB-0212910).

References

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 16:37–48.

Banno Y, Nakamura T, Nagashima E, Fujii H, Doira H. 2004. M chromosome of the wild silkworm, *Bombyx mandarina* ($n = 27$), corresponds to two chromosomes in the domesticated silkworm, *Bombyx mori* ($n = 28$). *Genome* 47:96–101.

Brodbeck D, Amherd R, Callaerts P, Hintermann E, Meyer UA, Affolter M. 1998. Molecular and biochemical characterization of the *aaNAT1* (*Dat*) locus in *Drosophila melanogaster*: differential expression of two gene products. *DNA Cell Biol*. 17:621–633.

Caicedo AL, Williamson SH, Hernandez RD, et al. (12 co-authors). 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *Plos Genet*. 3:1745–1756.

Carroll SB. 2000. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 101:577–580.

Carroll SB. 2005. Evolution at two levels: on genes and form. *Plos Biol*. 3:1159–1166.

Clark RM, Linton E, Messing J, Doebley JF. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc Natl Acad Sci U S A*. 101:700–707.

Dai FY, Qiao L, Tong XL, Cao C, Chen P, Chen J, Lu C, Xiang ZH. 2010. Mutations of an arylalkylamine-N-acetyltransferase, *Bm-iAANAT*, are responsible for silkworm melanism mutant. *J Biol Chem*. 285:19553–19560.

Duan J, Li RQ, Cheng DJ, et al. (11 co-authors). 2010. SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology. *Nucleic Acids Res*. 38:D453–D456.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.

Ekker SC, Jackson DG, Von Kessler DP, Sun BI, Young KE, Beachy PA. 1994. The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. *EMBO J*. 13:3551–3560.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics*. 155:1405–1413.

Ferguson L, Green J, Surridge A, Jiggins CD. Forthcoming 2011. Evolution of the insect *yellow* gene family. *Mol Biol Evol*. 28:257–272.

Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt PS, Eanes WF. 2007. Adaptive evolution of metabolic pathways in *Drosophila*. *Mol Biol Evol*. 24:1347–1354.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.

Futahashi R, Banno Y, Fujiwara H. 2010. Caterpillar color patterns are determined by a two-phase melanin gene prepatterning process: new evidence from *tan* and *laccase2*. *Evol Dev*. 12:157–167.

Futahashi R, Fujiwara H. 2005. Melanin-synthesis enzymes coregulate stage-specific larval cuticular markings in the swallowtail butterfly, *Papilio xuthus*. *Dev Genes Evol*. 215:519–529.

Futahashi R, Fujiwara H. 2007. Regulation of 20-hydroxyecdysone on the larval pigmentation and the expression of melanin synthesis enzymes and *yellow* gene of the swallowtail butterfly, *Papilio xuthus*. *Insect Biochem Mol Biol*. 37:855–864.

Futahashi R, Sato J, Meng Y, et al. 14 co-authors. 2008. *yellow* and *ebony* are the responsible genes for the larval color mutants of the silkworm *Bombyx mori*. *Genetics* 180:1995–2005.

Goldsmith MR, Shimada T, Abe H. 2005. The genetics and genomics of the silkworm, *Bombyx mori*. *Annu Rev Entomol*. 50:71–100.

Goloubinoff P, Paabo S, Wilson AC. 1993. Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proc Natl Acad Sci U S A*. 90:1997–2001.

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. 2005. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* 433:481–487.

Han Q, Fang J, Ding H, Johnson JK, Christensen BM, Li J. 2002. Identification of *Drosophila melanogaster* yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes. *Biochem J*. 368:333–340.

Hanson MA, Gaut BS, Stec AO, Fuerstenberg SI, Goodman MM, Coe EH, Doebley JF. 1996. Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics* 143:1395–1407.

Haudry A, Cenci A, Ravel C, et al. (11 co-authors). 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol Biol Evol*. 24:1506–1517.

Hilton H, Gaut BS. 1998. Speciation and domestication in maize and its wild relatives: evidence from the *globulin-1* gene. *Genetics* 150:863–872.

Hudson RR. 1987. Estimating the recombination parameter of a finite population model without selection. *Genet Res*. 50:245–250.

Hudson RR. 2001. Two-locus sampling distributions and their application. *Genetics* 159:1805–1817.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.

Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164.

Hufford KM, Canaran P, Ware DH, McMullen MD, Gaut BS. 2007. Patterns of selection and tissue-specific expression among maize domestication and crop improvement loci. *Plant Physiol*. 144:1642–1653.

Inomata N, Yamazaki T. 2002. Nucleotide variation of the duplicated *amylase* genes in *Drosophila kikkawai*. *Mol Biol Evol*. 19:678–688.

Jeong S, Rokas A, Carroll SB. 2006. Regulation of body pigmentation by the abdominal-B Hox protein and its gain and loss in Drosophila evolution. *Cell* 125:1387–1399.

Li H, Ruan J, Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*. 18(11):1851–1858.

Li XR, Tan LB, Zhu ZF, Huang HY, Liu Y, Hu SN, Sun CQ. 2009. Patterns of nucleotide diversity in wild and cultivated rice. *Plant Syst Evol*. 281:97–106.

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

Liu AZ, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics*. 173:321–330.

Liu C, Yamamoto K, Cheng TC, et al. (19 co-authors). 2010. Repression of tyrosine hydroxylase is responsible for the sex-linked chocolate mutation of the silkworm, *Bombyx mori*. *Proc Natl Acad Sci U S A*. 107:12980–12985.

Livingstone K, Anderson S. 2009. Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. *J Hered*. 100:754–761.

Lu Y, Rausher MD. 2003. Evolutionary rate variation in anthocyanin pathway genes. *Mol Biol Evol*. 20:1844–1853.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654.

Mita K, Kasahara M, Sasaki S, et al. (21 co-authors). 2004. The genome sequence of silkworm, *Bombyx mori*. *DNA Res*. 11:27–35.

Nagaraja GM, Nagaraju J. 1995. Genome fingerprinting of the silkworm, *Bombyx mori*, using random arbitrary primers. *Electrophoresis* 16:1633–1638.

Olsen KM, Womack A, Garrett AR, Suddith JI, Purugganan MD. 2002. Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. *Genetics* 160:1641–1650.

Pendleton RG, Rasheed A, Sardina T, Tully T, Hillman R. 2002. Effects of tyrosine hydroxylase mutants on locomotor activity in *Drosophila*: a study in functional genomics. *Behav Genet*. 32:89–94.

Pool J, Hellmann I, Jensen J, Nielsen R. 2010. Population genetic inference from genomic sequence variation. *Genome Res*. 20(3):291–300.

Prud'homme B, Gompel N, Carroll SB. 2007. Emerging principles of regulatory evolution. *Proc Natl Acad Sci U S A*. 104:8605–8612.

Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB. 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* 440:1050–1053.

Ramos-Onsins SE, Puerma E, Balana-Alcaide D, Salguero D, Aguade M. 2008. Multilocus analysis of variation using a large empirical data set: phenylpropanoid pathway genes in *Arabidopsis thaliana*. *Mol Ecol*. 17:1211–1223.

Ramsay H, Rieseberg LH, Ritland K. 2009. The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. *Mol Biol Evol*. 26:1045–1053.

Rausher MD, Lu YQ, Meyer K. 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *J Mol Evol*. 67:137–144.

Rausher MD, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol*. 16:266–274.

Riley RM, Jin W, Gibson G. 2003. Contrasting selection pressures on components of the Ras-mediated signal transduction pathway in *Drosophila*. *Mol Ecol*. 12:1315–1323.

Schlötterer C. 2003. Hitchhiking mapping—functional genomics from the population genetics perspective. *Trends Genet*. 19:32–38.

Sigmon B, Vollbrecht E. 2010. Evidence of selection at the *ramosa1* locus during maize domestication. *Mol Ecol*. 19:1296–1311.

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.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*. 24:1596–1599.

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. *Mol Biol Evol*. 21:1214–1225.

The International Silkworm Genome Consortium. 2008. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem Mol Biol*. 38:1036–1045.

True JR. 2003. Insect melanism: the molecules matter. *Trends Ecol Evol*. 18:640–647.

True JR, Yeh SD, Hovemann BT, Kemme T, Meinertzhagen IA, Edwards TN, Liou SR, Han Q, Li JY. 2005. *Drosophila tan* encodes a novel hydrolase required in pigmentation and vision. *Plos Genet*. 1:551–562.

Tsugehara T, Iwai S, Fujiwara Y, Mita K, Takeda M. 2007. Cloning and characterization of insect arylalkylamine N-acetyltransferase from *Bombyx mori*. *Comp Biochem Phys B Biochem Mol Biol*. 147:358–366.

Tweedie S, Ashburner M, Falls K, et al. (11 co-authors). 2009. FlyBase: enhancing *Drosophila* gene ontology annotations. *Nucleic Acids Res*. 37:D555–D559.

Walter MF, Black BC, Afshar G, Kermabon AY, Wright TR, Biessmann H. 1991. Temporal and spatial expression of the *yellow* gene in correlation with cuticle formation and dopa decarboxylase activity in *Drosophila* development. *Dev Biol*. 147:32–45.

Wang RL, Stec A, Hey J, Lukens L, Doebley J. 1999. The limits of selection during maize domestication. *Nature* 398:236–239.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 7:256–276.

Weiss G, Von Haeseler A. 1998. Inference of population history using a likelihood approach. *Genetics* 149:1539–1546.

White GM, Hamblin MT, Kresovich S. 2004. Molecular evolution of the phytochrome gene family in sorghum: changing rates of synonymous and replacement evolution. *Mol Biol Evol*. 21:716–723.

Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB. 2008. The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* 134:610–623.

Wittkopp PJ, Carroll SB, Kopp A. 2003. Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *Trends Genet*. 19:495–504.

Wittkopp PJ, Haerum BK, Clark AG. 2008a. Independent effects of cis- and trans-regulatory variation on gene expression in *Drosophila melanogaster*. *Genetics* 178:1831–1835.

Wittkopp PJ, Haerum BK, Clark AG. 2008b. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet*. 40:346–350.

Wittkopp PJ, True JR, Carroll SB. 2002a. Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* 129:1849–1858.

Wittkopp PJ, Vaccaro K, Carroll SB. 2002b. Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Curr Biol*. 12:1547–1556.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection of the maize genome. *Science* 308:1310–1314.

Wright TR. 1987. The genetics of biogenic amine metabolism, sclerotization, and melanization in *Drosophila melanogaster*. *Adv Genet*. 24:127–222.

Xia AH, Zhou QX, Yu LL, Li WG, Yi YZ, Zhang YZ, Zhang ZF. 2006. Identification and analysis of YELLOW protein family genes in the silkworm, *Bombyx mori*. *BMC Genomics*. 7:195–204.

Xia QY, Guo YR, Zhang Z, et al. (60 co-authors). 2009. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science* 326:433–436.

Xia QY, Zhou ZY, Lu C, et al. (93 co-authors). 2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* 306:1937–1940.

Xiang ZH, Huang JT, Xia JG, Lu C. 2005. Biology of sericulture. Beijing: China Forestry Publishing House.

Zhang L, Peek AS, Dunams D, Gaut BS. 2002. Population genetics of duplicated disease-defense genes, *hm1* and *hm2*, in maize (*Zea mays* ssp. *mays L.*) and its wild ancestor (*Zea mays* ssp. *parviglumis*). *Genetics* 162:851–860.

Zhu QH, Zheng XM, Luo JC, Gaut BS, Ge S. 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol Biol Evol*. 24:875–888.