Journal of Insect Science
www.insectscience.org

Nucleotide diversity and selection signature in the domesticated silkworm, *Bombyx mori*, and wild silkworm, *Bombyx mandarina*

Yi Guo[1a], Yi-Hong Shen[1b], Wei Sun[1c], Hirohisa Kishino[2d], Zhong-Huai Xiang[1e], Ze Zhang[1, 3f*]

[1]The Key Sericultural Laboratory of Agricultural Ministry, Southwest University, Chongqing 400715, China
[2]The Laboratory of Biometrics and Bioinformatics, Graduate School of Agriculture and Life Science, The University of Tokyo, 1-1-1, Yayoi, Bunkyo, Tokyo, 113-8657, Japan
[3]The Institute of Agricultural and Life Sciences, Chongqing University, Chongqing 400044, China

Abstract

To investigate the patterns of nucleotide diversity in domesticated silkworm, *Bombyx mori* L. (Lepidoptera: Bombycidae) and its wild relative, Chinese wild silkworm, *Bombyx mandarina* Moore, we sequenced nine nuclear genes. Neutrality test and coalescent simulation for these genes were performed to look at bottleneck intensity and selection signature; linkage disequilibrium (LD) within and between loci was employed to investigate allele association. As a result, *B. mori* lost 33-49% of nucleotide diversity relative to wild silkworm, which is similar to the loss levels found in major cultivated crops. Diversity of *B. mori* is significantly lower than that of *B. mandarina* measured as π_{total} (0.01166 vs. 0.1741) or θ_{w} (0.01124 vs. 0.02206). Bottleneck intensity of domesticated silkworm is 1.5 (in terms of $k = N_b/d$, N_b-bottleneck population size; d-bottleneck duration) with different durations. Gene *DefA* showed signature of artificial selection by all analysis methods and might experience strong artificial selection in *B. mori* during domestication. For nine loci, both curves of LD decay rapidly within 200 bp and drop slowly when distance is > 200 bp, although that of *B. mori* decays slower than *B. mandarina* at loci investigated. However, LD could not be estimated at *DefA* in *B. mori* and at *ER* in both silkworms. Elevated LD observed in *B. mori* may be indicator of selection and demographic events.

Keywords: bottleneck, domestication, nucleotide diversity, selection
Abbreviations: LD, linkage disequilibrium
Correspondence: [a] guo1983752@163.com, [b] cs_shenyh@swu.edu.cn, [c] sunwei077@163.com, [d] kishino@lbm.ab.a.u-tokyo.ac.jp, [e] xbxzh@swu.edu.cn, [f] zezhang@swu.edu.cn, *Corresponding author
Editor: Zhijian Jake Tu was Editor of this paper.
Received: 16 September 2010, **Accepted:** 22 December 2010

ISSN: 1536-2442 | Vol. 11, Number 155

Cite this paper as:
Guo Y, Shen Y-H, Sun W, Kishino H, Xiang Z-H, Zhang Z. 2011. Nucleotide diversity and selection signature in the domesticated silkworm, *Bombyx mori*, and wild silkworm, *Bombyx mandarina. Journal of Insect Science* 11:155 available online: insectscience.org/11.155

Introduction

The typical outcome of domestication of crop plants and animals is a genome-wide loss of genetic diversity (Tanksley and McCouch 1997). Although many forces can shape patterns of genetic diversity and structure in crops and domestic animals, one of the most influential factors is the population bottleneck that occurs during the founding of a new domestic lineage (Eyre-Walker et al. 1998). In addition, natural or artificial selection, selective sweep, population extension and subdivision can often change allele frequency and distribution and thus have a major impact on the organization of genetic diversity within the genome. Therefore, understanding what forces play roles in shaping patterns of genetic variation, decrease in diversity, and target of selection in domesticated species is of importance to genetics of crop and animal domestication.

Population genetics analysis is a powerful approach to disentangling the effects of different evolutionary forces on DNA variation. In the past decade, there have been many efforts to investigate the patterns of genetic diversity and histories of domestication for major crops like maize (Vigouroux et al. 2002; Tenaillon et al. 2004; Wright et al. 2005; Tian et al. 2009), rice (Tang et al. 2006; Caicedo et al. 2007), soybean (Zhu et al. 2003), sunflower (Liu and Burke 2006; Olsen et al. 2006; Kolkman et al. 2007), sorghum (Hamblin et al. 2006), and tomato (Labate et al. 2009). All previous studies demonstrated that major crops have retained 30-67% of nucleotide diversity present in the corresponding wild relatives. Besides effects on overall level of diversity, domestication has led to increase in linkage disequilibrium (LD, nonrandom association of alleles at different sites) throughout a genome via population bottleneck or selection or inbreeding or changes in mating system or their combinations.

Although an increasing number of polymorphism datasets have emerged from different crop species and their wild relatives, there is little data from either domesticated or wild animals. Since animals and plants have striking differences in behavior, life cycle, environment and morphology, it is necessary to describe patterns of nucleotide diversity and extent of LD in a typical domesticated animal.

The domesticated silkworm, *Bombyx mori* L. (Lepidoptera: Bombycidae), originated from the ancient Chinese wild silkworm, *Bombyx mandarina* Moore, and has experienced a history of artificial selection for more than 5,000 years (Xiang et al. 2005). Since the beginning of last century, *B. mori* has been a model organism for the Lepidoptera order of insects because it has many morphological mutants. More than 1,000 inbred domesticated strains are maintained worldwide (Goldsmith et al. 2005). In addition, wild silkworm currently occurs in mulberry fields. *B. mori* and *B. mandarina* differ in many traits including growth rate, flight behavior, egg laying, cocoon size, silk quality, and a kind of mimicry as a result of domestication. These silkworms provide a unique opportunity for study on the evolutionary forces that have resulted in the trait differences between the two species.

Although a genome-wide resequencing of 40 silkworm genomes by Solexa has been reported recently (Xia et al. 2009), information of nucleotide diversity by the traditional sequencing method is still lacking for *B. mori* and *B. mandarina*. Thus, this study is a complementary work which may uncover some findings neglected by the resequencing, especially when pitfalls of the Solexa technique may affect the efficiency of SNP calling (Shendure et al. 2008; Pool et al. 2010). In fact, our study was launched before the silkworm Solexa resequencing project started. Nine nuclear genes were sequenced for two populations of domesticated and wild silkworms. The results suggested that domesticated silkworm lost 33-49% of nucleotide diversity relative to wild silkworms. The reduction of genetic diversity presented in this study is similar to that found in major cultivated crops (Vigouroux et al. 2002; Zhu et al. 2003;Tenaillon et al. 2004; Wright et al. 2005; Tang et al. 2006; Liu and Burke 2006; Olsen et al. 2006; Caicedo et al. 2007; Hamblin et al. 2006; Kolkman et al. 2007;Tian et al. 2009; Labate et al. 2009). We also performed coalescent simulations under a bottleneck model with bottleneck population size (N_b) and duration time (d) to look at bottleneck intensity and detect genes departing from the model. As a result, silkworms might experience more severe bottleneck than crops. One locus, *DefA*, exhibited a departure from neutrality. Furthermore, elevated LD in domesticated silkworms is likely to be caused by bottleneck and inbreeding. The observations presented in this study are somewhat different from that of Solexa resequencing, especially in genetic diversity loss level. The possible reasons for these differences are discussed.

Materials and Methods

Silkworms

Sixteen domesticated silkworm strains and 15 Chinese wild silkworm samples were used in this study (Table 1). The domesticated accessions were obtained from the Institute of Sericulture and Systems Biology at Southwest University, China, and represented four main geographic strains (i.e., Chinese, Japanese, European, and Tropical) with great genetic diversity in voltine and other characters. The wild silkworm samples were collected from various geographical regions in China. Genomic DNAs were extracted from larva, pupa, or moth for each individual, using a standard phenol-chloroform approach.

Genes studied

To investigate the general patterns of nucleotide diversity in wild and domesticated silkworms, nine loci were chosen based on their functions and locations on chromosomes (Table 2). They were all considered as neutral loci in the study using Solexa sequencing method (Xia et al. 2009). Since α-Amylase (*Amy*), Acetylcholinesterase 2 (*AchE*), and *RpSA* cannot be mapped on chromosomes in the current silkworm genome sequence, the other six loci are located on five respective chromosomes and far from each other even if on the same chromosome. In short, they were selected from at least four of 28 chromosomes and represented genes of different functions including digestion, development, immunity and reproduction. Wingless-1 (*wnt-1*) and ecdysone receptor (*ER*) are involved in development. Defensin A (*DefA*) is a member of the antibacterial peptide family and plays roles in innate immunity. Pheromone biosynthesis activating neuropeptide (*PBAN*) is thought to promote synthesis of pheromone related to reproduction. Glutathione-S-transferase (*GST*) participates in detoxification metabolism and antioxidation process. α-Amylase (*Amy*) and alcohol dehydrogenase (*Adh*) hydrolyze starch and dehydrogenate ethanol during digestion, respectively. Acetylcholinesterase 2 (*AchE*) refers to catalyzing the hydrolysis of the neurotransmitter acetylcholine at cholinergic synapses of the central nervous system, and *RpSA* is ribosomal protein SA gene.

PCR amplification, cloning and sequencing

Primers listed in Table 2 were used to perform PCR amplification for each sample. PCR products were purified using DNA Extract Kit (Omega, www.omega.com) and then cloned into PMD19-T vector (Takara, www.takara-bio.com). Clones were sequenced on both strands using BigDye chemistries (Applied Biosystem, www.appliedbiosystems.com) and Applied Biosystem 3730 DNA sequencer (Invitrogen, www.invitrogen.com) following manufacturer's protocol. Three to five clones for single individual were sequenced to exclude PCR and sequencing errors. All sequences were assembled using SeqMan in DNASTAR software version 5.01 (www.dnastar.com) and edited by Bioedit 7.01 (Hall 1999). Sequences obtained in this study have been deposited with the EMBL/GenBank Data Libraries under accession no. GQ420700-GQ420850, GQ423313-GQ423341, GQ423277-GQ423307, HM132882-HM132933.

Statistics of diversity and recombination

The sequences were aligned by ClustalW 1.81 (Thompson et al. 1997). Insertion/deletions (indels) were excluded from the analysis. For each locus, DNAsp 5.00 (Librado and Rozas, 2009) was utilized to calculate polymorphism parameters including S (number of segregating sites), π_{total} (the mean number of nucleotide differences per site), π_a (average pairwise difference for nonsynonymous sites), π_s (average pairwise difference for synonymous sites) and θ_w (Watterson's estimator of 4$N_e\mu$), as well as π_{silent} (π values for synonymous and noncoding sites)$_,$ θ_{silent} (θ_w values for synonymous and noncoding sites) and $\pi_a/\pi_{s.}$ The minimum number (R_m) of recombination events (Hudson and Kaplan 1985) and the estimator (R) of population recombination parameter (Hudson 1987) were also obtained by DNAsp.

Neutrality tests

To assess neutral prediction and reveal evolutionary history between *B. mori* and *B. mandarina*, two kinds of tests were performed. Tajima's D, which was sensitive to low-frequency variants, was estimated for all sites of each locus by DnaSP. Tajima's D value was usually negative in populations suffering from selection and/or demographic events. Multilocus HKA test were performed using Jody Hey's HKA software (http://genfaculty.rutgers.edu/hey/software).

Coalescent simulation

To investigate bottleneck intensity during silkworm domestication, a single bottleneck model was introduced (Weiss and Haeseler 1998). Coalescent simulation was run using Hudson's ms (Hudson 2002) and the parameters in coalescent simulation (Figure 1) were given as follows:

-N_a and N_p: N$_a$ was the ancestor population size before split of two silkworms and N_p was the population size of present domesticated silkworm. Since previous

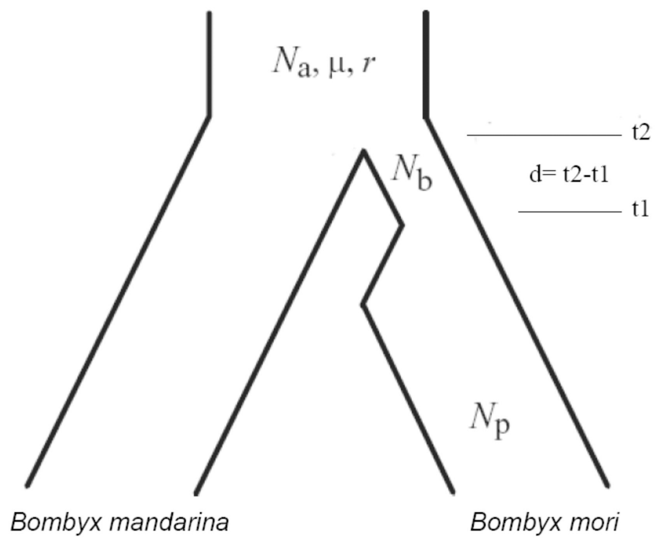


Figure 1. The bottleneck model of coalescent simulation in this study. High quality figures are available online.

studies have shown that there was a litter effect of population size on simulation (Wright et al. 2005; Zhu et al. 2007; Li et al. 2009), we assigned both N_a and N_p to 1000000.

-μ and γ: 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 by LDhat (http://www.stats.ox.ac.uk/~mcvean/LDhat/instructions.html).

-N_b and *d*: The key parameter of the bottleneck model was *k*, where $k = Nb/d$. N_b was the population size of *B. mori* during bottleneck and *d* was the duration of bottleneck. *d* was determined by t2-t1, where t2 was the time domestication occurred and t1 was the time of the domestication termination. The archaeological evidence suggested that *B. mori* split from *B. madnarina* about 5000 years ago (Xiang et al. 2005). 5000 years should be a minimum value as the accurate time of both t2 and t1 was unknown. So we made an implicit assumption that domestication of crops and silkworm occurred at the proximate time and fixed t2 to 7500 and six *d* values (*d* = 200, 500, 1000, 1500, 2000 and 3000) were explored. The maximum *d* value was set to 3000 for two reasons. First, in coalescent simulations for crops, 3000 generation (or year) was the maximum duration of a bottleneck in crop domestication (Zhu et al. 2007). Second, 7500 minus 3000 was 4500, which was around 5000. 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, three diversity statistics were calculated, π, S (the number of segregating sites) and Tajima's D from simulated *B. mori* data. Observed values of π, S and Tajima's D from *B. mori* were compared to simulated values and fit of simulated data to observed data was assessed. To assess fit, levels of acceptance were defined corresponding to a range of ± 20% (Weiss and Haeseler 1998) of π, S or Tajima'S D. The likelihood of each scenario for each locus was obtained by calculating portion of 10000 simulations (Tenaillon et al. 2004; Wright et al. 2005; Zhu et al. 2007; Haudry et al. 2007) fitting the data. Multilocus likelihood values were calculated by multiplying across loci. This method was based on the assumption that loci were independent, which was verified by interlocus linkage disequilibrium.

Analysis of intralocus linkage disequilibrium

Linkage disequilibrium (LD) within genes was measured as r^2 (squared allele frequency correlation, Hill and Robertson, 1968) between pairwise sites over physical distance. For a combined set of all nine loci, DNAsp was used to calculate r^2 and all singletons were excluded from LD analysis. We also evaluated distribution of r^2 for *DefA*, which may show selection signature. r^2 was plotted against pairwise distance and two logarithm curves fitting the data were drawn for *B. mori* and *B. mandarina*, respectively.

Analysis of interlocus linkage disequilibrium

To survey extent of linkage disequilibrium between genes, interlocus LD analysis was performed across nine loci for both *B. mori* and *B. mandarina*. The number of haplotypes was measured and haplotype distribution was identified by DNAsp for each locus. Then, diploid data was generated manually to create an input file for GenePop (Raymond and

Rousset, 1995). In GenePop, each haplotype was considered as an allele and significant tests of interlocus linkage disequilibrium were performed.

Results

Diversity

For nine loci investigated, the average sample sizes were 14.9 for domesticated silkworms and 14.7 for wild silkworms. After excluding indels, the length of alignment sequences varied from 788 to 1678 bp and a total of 9268 bp sequence per individual was sequenced, including 4368 bp of coding region and 4964 bp of noncoding region, respectively (Table 2). In contrast, the number of polymorphic sites in *B. mandarina* was nearly twice than that in *B. mori*; there were 638 polymorphic sites (1 SNP/ 14.5bp) in wild silkworm and 330 segregating sites (1 SNP/28bp) in domesticated silkworm. In Table 3, the value of π_{total} ranged from 0.0006 to 0.02685 with the average of 0.01166 ± 0.00912 per bp in *B. mori*, whereas it varied from 0.0052 to 0.0312 in *B. mandarina* with a mean of 0.01741 ± 0.00951 per bp. Similarly, the estimate of Watterson's θ_{w} fluctuated from 0.0014 to 0.01957 (mean: 0.01124 ± 0.00715) in *B. mori*, compared with that ranging from 0.0098 to 0.0326 (mean: 0.02206 ± 0.00975) in *B. mandarina*. The value of π_{total} in *B. mandarina* was significantly higher than that in *B. mori* (Wilcoxon signed rank test, Wilcoxon 1945, $P < 0.01$, df = 8), as well as θ_{w} (Wilcoxon signed rank test, Wilcoxon 1945, $P < 0.01$, df = 8). *B. mori* harbors only 67% or 51% of genetic diversity relative to *B. mandarina,* measured by π or θ_{w}. Non-zero estimates of $\pi_{\text{a}}/\pi_{\text{s}}$ ranged from 0.033 to 2.49 in *B. mori* and the average was 0.4 ± 0.797*,* whereas the values of $\pi_{\text{a}}/\pi_{\text{s}}$ ranged from 0.022 to 1.339 in *B. mandarina* and the average was 0.321 ± 0.465 (Table 3).

Synonymous and nonsynonymous sites were counted within and between species. There was only one fixed substitution between species. It is not surprising that only one nonsynonymous change occurred at *DefA,* given that domestication has only happened within the last 5,000 years (Lu et al. 2002). Such absence of fixed differences is also the case in rice (Zhu et al. 2007) and sorghum (White et al. 2004). In fact, *B. mori* and *B. mandarina* can be crossed with each other and the hybrid progeny is fertile, even though they are defined as two distinct species (Banno et al. 2004).

We also estimated π and θ_{w} for seven univoltine and five bivoltine domesticated silkworm strains, respectively. Summary statistics of nucleotide diversity did not show difference between univoltine and bivoltine strains (Wilcoxon signed rank test, Wilcoxon 1945, $P > 0.05$). Thus, all strains of *B. mori* were combined as one group in further analysis.

Neutrality test

Tajima's D values in *B. mori* are higher than those in *B. mandarina* for eight of nine loci (Table 3), perhaps resulting from recent bottleneck in domesticated silkworms. This is consistent with the pattern observed in rice (Zhu et al. 2007) and maize (Tenaillon et al. 2004). Tajima's D values are negative at six loci for *B. mori* and at eight loci for *B. mandarina* (Table 3), suggesting that the allele frequency at these loci skewed to rare variants, which may be the result of purifying selection, subdivision or population expansion (Tenaillon et al. 2004). Three loci (*PBAN*, *AchE* and *RpSA*) in *B. mori* and one locus (*RpSA*) in *B. mandarina* showed positive D values, indicating that intermediate-frequency variants were dominant in the populations.

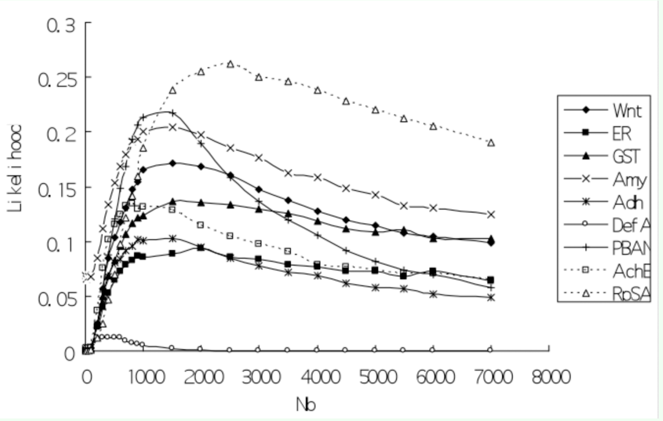


Figure 2. Likelihood curves of 9 loci using the number of segregating sites (S) as fit standard with d=1000. High quality figures are available online.

Three loci, *DefA* , *ER* and *AchE*, exhibited significant Tajima's D values, which were likely due to either selection or demographic events.

The ratio of π_a to π_s at *DefA* exceeded one (1.339) in *B. mandarina*, but was not available in *B. mori* because of the absence of polymorphic sites in the coding region. In fact, a sharp decrease of polymorphism at *DefA* in *B. mori* relative to *B. mandarina* measured by segregating sites was also observed in both whole sequence (4 in *B. mori* vs 85 in *B. mandarina*, Table 3) and coding region (no segregating sites in *B. mor*i and 14 segregating sites in *B. mandarina*, Table 3). Thus, we can make a conclusion that *DefA* is the most likely to be a target of artificial selection in *B. mori*.

ER, which is subject to purifying selection because of significantly negative D values, may experience a relaxation of purifying selection based on different π_a/π_s ratios for domesticated (0.832) and wild species (0.517).

A higher π_a/π_s ratio (2.409) at *PBAN* in *B. mori* was observed. However, this high ratio was based on only one synonymous substitution and three nonsynonymous substitutions that occurred in the partial coding region investigated. In addition, Tajima's D value at *PBAN* in *B. mori* was not significant, suggesting that neutrality could not be ruled out.

In a multilocus HKA test, neutrality was not refused based on nine loci (data not shown). A possible reason is that *B. mori* and *B. mandarina* are closely related species and there is little divergence between the two species. The power of a multilocus HKA test is low. Thus, other analysis methods should be considered.

Coalescent simulation and bottleneck severity

Likelihood curves using *S* (the number of segregating sites) as fit standard for each locus with *d* = 1000 (*d* is the duration of bottleneck) were displayed in Figure 2. Curves of all loci showed apparent peaks except for *DefA*, suggesting that those loci fit well with one bottleneck model. Likelihood values of all loci reached peaks when $N_b < 2500$ (N_b is the population size of *B. mori* during bottleneck) (Figure 2). However, likelihood surface of *DefA* only showed a very slight peak and the likelihood values for this curve were smaller than 0.05, which is the same case at several selected loci of maize also suffering from more severe bottleneck than other neutral loci (Tenaillon et al. 2004). This suggested that *DefA* was subject to the most severe bottleneck. Thus, *DefA* deserves further analysis for selective signature. In addition, likelihood curves based on Tajima's D failed to reach peaks for 7 of 9 loci (Supplementary Figure S1). A possible reason for this is that Tajima's D skewed toward rare alleles. Distribution of Tajima's D for wild silkworms might be never mimicked adequately in simulations (Tenaillon et al. 2004).

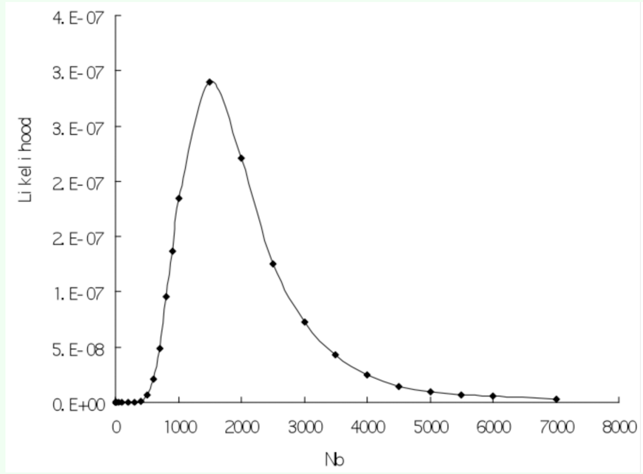


Figure 3. Joint likelihood surfaces of 8 neutral loci using the number of segregating sites (S) as fit standard with d=1000. High quality figures are available online.

Multilocus likelihood values were calculated to find the optimal k (= Nb/d) value. *DefA* was excluded from multilocus likelihood analysis because this locus showed evidence of the most severe bottleneck. Thus, multilocus likelihood curve based on eight neutral loci was obtained. Figure 3 shows multilocus likelihood surface based on S with d = 1000. Interestingly, over six d values, k values with maximum likelihood were all 1.5 using either π or S as fit- of-goodness standard. Because Tajima's D test and π_a/π_s ratios suggested two loci (*ER* and *AchE*) may not evolve neutrally, we repeated multilocus likelihood analysis by removing those two loci. However, the surface changed a little (data not shown), indicating that those two loci fit bottleneck model well and may evolve neutrally.

Test for selection using bottleneck model

To investigate whether one locus evolved neutrally, a likelihood ratio test was performed. Figure 3 shows that for eight neutral loci, multilocus likelihood reached the maximum value when N_b = 1500 (k = 1.5) under the condition that d = 1000. Then, it can be obtained from Figure 2 that likelihood value for *DefA* was 0.0016 when N_b = 1500 (k = 1.5). However, the maximum likelihood value of *DefA* was 0.0117 with N_b = 300 (Figure 2). Thus, –2ln(0.0016/0.0117), an likelihood ratio test statistic for *DefA* was constructed based on S and the condition with d = 1000. The test was significant (P = 0.045, df = 1). Similarly, the likelihood ratio test for *DefA,* using π as fit standard, was also significant (P = 0.017, df = 1). In addition, likelihood ratio tests for the other eight loci were not significant at all. Therefore, *DefA* did not fit the bottleneck model and was a selection target during domestication.

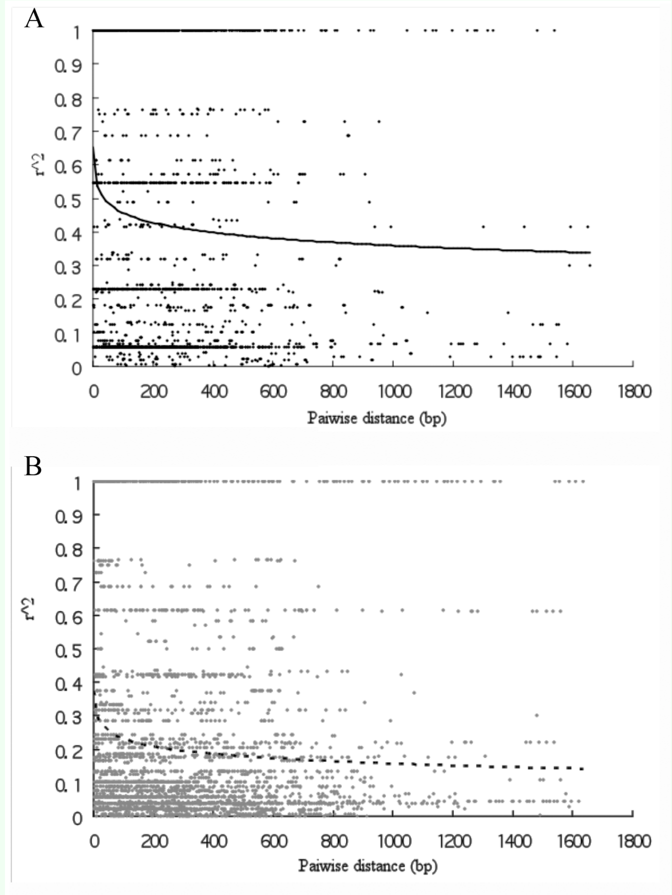


Figure 4. Decay of intralocus linkage disequilibrium against pairwise distance for 9 loci in domesticated silkworm (A) and wild silkworm (B). The solid and dotted lines represent logarithmic curves fitting to the data for *B. mori* and *B. mandarina*, respectively. High quality figures are available online.

Linkage disequilibrium and recombination

Figure 4 shows the decay of linkage disequilibrium (LD) with nucleotide pairwise distance for pooled data of nine genes in each

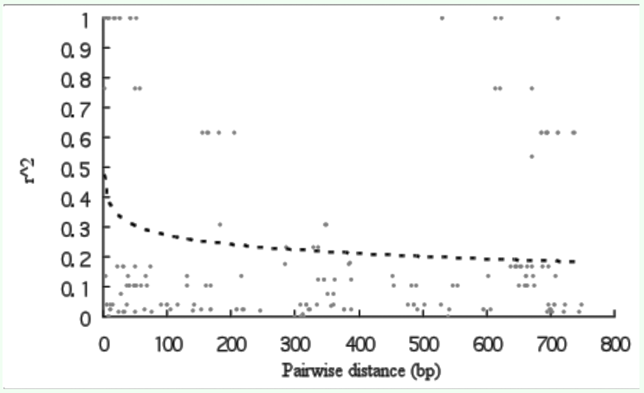
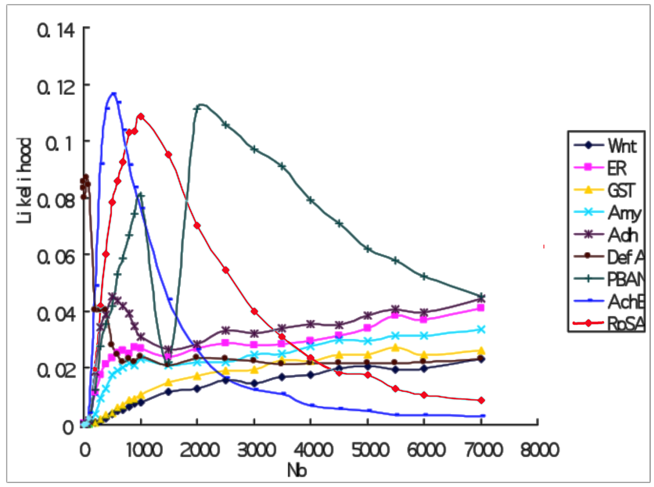


Figure 5. Decay of intralocus linkage disequilibrium against pairwise distance for *DefA* in wild silkworm. The dotted line represents logarithmic curve fitting to the data for *B. mandarina* only. High quality figures are available online.

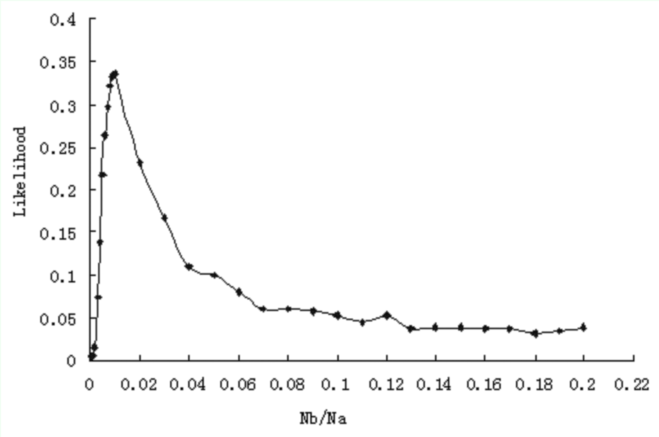
silkworm. r^2 refers to squared allele frequency correlation and was used to fit a logarithm curve across distances for *B. mori* and *B. mandarina*, respectively. Both curves of LD decay rapidly within 200 bp and drop slowly when distance is > 200 bp. Most strikingly, however, *B. mori* shows an extensive LD at a value more than 0.4 within 1,200 bp. In contrast, the curve for *B. mandarina* is below that of *B. mori* and drops to 0.1 within about 1600 bp.

We also analyzed intralocus LD for *DefA* with selection signature. However, the LD pattern at this locus was different from the one obtained for pooled data of nine genes; the curve of LD decay for *DefA* in *B. mandarina* drops to 0.2 within about 500 bp (Figure 5). However, for *DefA* in *B. mori*, there is no pairwise comparison and LD cannot be analyzed.

Estimates of recombination parameter R (Hudson 1987) range from 0 to 0.0866 (mean: 0.0183 ± 0.0288) in *B. mori* and from 0 to 0.0744 (mean: 0.033 ± 0.0262) in *B. mandarina* (Table 3). Similarly, the minimum numbers of recombination events (R_m) range from 0 to 8 in *B. mori* (mean: 2.0 ± 2.8) and from 0 to 12 (mean: 5.6 ± 3.9) in *B. mandarina*.



Supplementary Figure S1: Likelihood curves of 9 loci using Tajima's D (D) as fit standard with *d* = 1000. High quality figures are available online.



Supplementary Figure S2: Joint likelihood curves of 8 neutral loci using S as fit standard without duration of bottleneck. High quality figures are available online.

Interlocus LD for both *B. mori* and *B. mandarina* was analyzed based on all nine loci. However, no significant results were observed in any comparisons and thus, these loci were unlinked. In fact, these loci are either on different chromosomes or on different scaffolds on the same chromosome (Table 2).

Discussion

Nucleotide Diversity

This study presented the nucleotide polymorphism data in domesticated and wild silkworms by traditional sequencing. Wild silkworm, which lives in natural environmental conditions, harbors substantial levels of nucleotide diversity (mean π_{total}: 0.01741±0.00951, Table 3). These observations are comparable to the levels (π: 0.0018-0.0350) found in *Drosophila* (Andolfatto 2005; Matzkin 2008) and butterfly (Counterman et al. 2010). Relative to wild silkworm, domesticated silkworm contains markedly less nucleotide diversity. There was on average 33% or 49% loss of nucleotide diversity measured by π or θ calculated on nine loci in domesticated silkworm in a comparison to wild silkworm.

In contrast, a recent whole genome scan by Solexa resequencing found that domesticated silkworm lost only 17% of nucleotide diversity measured by θ relative to wild silkworm (Xia et al. 2009). However, the nucleotide diversity in *B. mori* was found to be at a very similar level measured as θ between the two studies (0.01124 vs 0.0108: this study vs Solexa). This implies that both the gene set and domesticated silkworm samples used in this study are representative although their numbers are relatively small, respectively. There is, however, the difference of nucleotide diversity in *B. mandarina* between the two studies (0.02206vs 0.0130: this study vs Solexa). Two reasons may lead to this difference. First, the whole genome scan used only 11 wild silkworm representatives from different regions (Xia et al. 2009), whereas this study used 15 wild silkworm samples from different areas in China. Nine of 15 wild samples in this study were collected from the same areas as in the research by Solexa (Xia et al. 2009). In our dataset, average θ value is 0.01878 for nine wild samples used in previous study, which is lower than that (0.02206) for 15 samples. Although the difference is not significant (Wilcoxon signed rank test, Wilcoxon 1945, $W = 1$, $P > 0.05$), different sampling strategies for *B. mandarina* may be in part responsible for the difference in θ value.

Another reason may be that Solexa resequencing underestimated the level of nucleotide diversity in wild silkworm. Sequence reads by Solexa 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 towards the allele found in the reference sequence. As a result, it may underestimate level of nucleotide diversity for divergent sequences. Since one reference genome sequence was 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 the future.

In principle, domestication leads to loss of genetic diversity in domesticated species. Interestingly, the level of nucleotide diversity loss observed in this study is comparable to the levels observed in major cultivated crops for which loss of genetic diversity is in general 30% - 69% (Table 4). It appears that domestication has a similar quantitative effect on reduction of genetic diversity in diverse domesticated species including plants and animals.

Bottleneck intensity

In our study, *k* values with different domestication durations were all 1.5, contrasting to those in main crops that vary from 2.45 to 5 (Tenaillon et al. 2004; Wright et al. 2005; Haudry et al. 2007; Li et al. 2009). Although a previous study reported *k* = 0.8 for Asian rice (Zhu et al. 2007), probably this small *k* value may be due to high selection pressure on functional regions (Li et al. 2009). Thus, we can conclude that domesticated silkworms experienced more severe bottleneck during domestication than crops. These results indicated that the foundation population size varied from 300 (*d* = 200) to 4500 (*d* = 3000). Given duration of domestication as 1000, the foundation population size was 1500 (Figure 3) for *B. mori*. Although *B. mori* might suffer from more severe bottleneck, the reduction of nucleotide diversity was 33% or 49% relative to *B. mandarina*, a decrease that still falls into the range of diversity loss in crop domestication (Table 4). It is most likely that changes in mating systems may play an important role and contribute to loss of genetic diversity in plant domestication. The transition from open-pollinated to self-fertilization might have led to a decrease in effective population size and recombination rate (Charlesworth et al. 2003). As animals cannot self-fertilize, *B. mori* should maintain a relatively high level of polymorphism than selfing crops. As expected, we observed that the levels of nucleotide diversity (π_{total}: 0.01166±0.00912, Table 3) in domesticated silkworms are somewhat higher than those found in some selfing crop plants. For example, nucleotide diversities (π_{Total}) of selfing crops are 0.0023 in rice (Caicedo et al. 2007), 0.0022 in sorghum (Hamblin et al. 2006) and 0.0013 in soybean (Zhu et al. 2003). Thus, change of mating system may be a factor in determining levels of genetic diversity in crops. Without such a change, *B. mori* maintains relatively higher genetic diversity than crops.

Although a bottleneck model for silkworm domestication was also constructed recently (Xia et al. 2009), it should be pointed out that the bottleneck model neglected duration of bottleneck and, as a result, *k* value was not estimated in the previous study (Xia et al. 2009). Given *k* value with 1.5 for different *d* values, we can infer foundation population size of silkworm and compare it to other domesticated species. We also performed coalescent simulations using a bottleneck model without duration. Based on eight neutral loci, *B. mori* lost 99% population size relative to its ancestor (Supplementary Figure S2) compared with 90% reduction of population size in previous study (Xia et al. 2009). Both decreases in population size of *B. mori* suggest a severe bottleneck during domestication.

Target of selection

One of the goals of this study was to test selection under the bottleneck model. Among nine loci, only *DefA* showed evidence of selection. As one of the factors involved in the innate immune system, *DefA* was involved in resistance to microbes. Strikingly, no polymorphism was observed in *DefA* coding sequences of the domesticated silkworm whereas a normal level of nucleotide diversity in the corresponding sequences was found in the wild silkworm (Table 3). These results can be explained by the fact that *DefA* has experienced a strong artificial selection in the domesticated silkworm. It is most likely that during domestication, human beings selected strongly healthy individuals to raise and breed. As a result, polymorphism at the locus decreased gradually and finally one haplotype with the best fitness was fixed in domesticated silkworm. In contrast, *B. mandarina* was subject to purifying selection rather than such strong artificial selection and thus contained a normal level of polymorphism (π: 0.0202) at *DefA*. Thus, *DefA* may be a target of selection during domestication.

In recent research, a large number of genes involved in immunity, including many unique recognition genes and antimicrobial peptide genes in domesticated silkworm, were identified (Tanaka et al. 2008). Thus, it is interesting to investigate nucleotide diversity at antimicrobial peptide gene loci in silkworms to see whether innate immune systems will be really critical for silkworm domestication in the future.

Although *ER* and *AchE* also showed significant Tajima's D value and *PBAN* harbored high π_a/π_s ratio, neither of them departed from neutrality in a coalescent simulation test. In fact, it is disingenuous to test neutrality just using Tajima's D test because significance of Tajima's D can result from factors including selection and/or demographic event. Similarly, π_a/π_s ratio based on a small segment of coding sequence was weak in the neutrality test. In our study, these three loci fit well with a bottleneck model and thus, they were considered as neutral genes. In contrast, departure from neutrality for *DefA* has been proven by Tajima's D test, π_a/π_s ratio and coalescent simulation. Therefore, selection detection by coalescent simulation is the most reliable, which is also the case in maize (Tenaillon et al. 2004; Wright et al. 2005).

Linkage disequilibrium

In Figure 4, increase in the extent of LD observed in *B. mori* was observed, corresponding to a small mean number of R_m (2.0 ± 2.8) and low recombination parameter (0.0183 ± 0.0288). Reduction of effective population size and/or artificial selection on domestication loci may be responsible for this. It is notable that in both silkworm and crops, Tajiama's D values were higher in domesticated species than those in wild ones (see Results section). Thus, recent bottleneck or inbreeding could not be ignored. However, we cannot distinguish bottleneck and inbreeding in the case of absence of domesticated silkworm landraces. In the future, it would be wise to sequence the genes in silkworm landraces in order to investigate corresponding diversity and LD patterns.

Compared with Figure 4A, *DefA* showed absence of decay of LD in *B. mori*. In fact, *B. mori* was extremely homozygous at *DefA* (Table 3). Thus, it can concluded that selection resulted in lower recombination at this locus and should be responsible for extensive LD in *B. mori*.

As in an outcrosser, r^2 declines to 0.10 and 0.4 within 1.6 kbp in wild silkworm and domesticated silkworm, respectively. Similarly, r^2 declines to < 0.10 within 1.0 kbp in maize, a crop that is highly outcrossed (Tenaillon et al. 2001; Remington et al. 2001). However, selfing crops showed somewhat different patterns in LD decay. For example, there is little decline in LD over distances as great as 50 kbp in soybean (Zhu et al. 2003); LD in rice approaches r^2 = 0.10 only after about 100 kbp (Garris et al. 2003). Thus, LD patterns of both silkworms in this study are similar to those of outcrossing crops.

Conclusions

Diversity of *B. mori* was significantly lower than that of *B. mandarina* measured as π_{total} (0.01166 vs. 0.1741) or θ_{w} (0.01124 vs. 0.02206). In general, *B. mori* lost 33-49% of nucleotide diversity relative to wild silkworm and the reduction of genetic diversity is similar to the levels found in major cultivated crops. *B. mori* suffered from more severe bottleneck than crops. *DefA* showed evidence of selection using either a neutrality test or coalescent simulation. As a kind of immunity-related gene, *DefA* perhaps experienced strongly artificial selection and became highly conserved in *B. mori* during domestication. Elevated LD in *B. mori* was observed and may be an indicator of selection and demographic events. LD patterns of both silkworms in this study are similar to those of outcrossing crops.

Acknowledgements

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. 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 Programme of Introducing Talents of Discipline to Universities (B07045).

References

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437(7062): 1149-1152.

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(1): 96-101.

Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, *et.al.* 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics* 3: e163.

Charlesworth D, Bartolomé C, Schierup MH, Mable BK. 2003. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. *Molecular Biology and Evolution.* 20(11): 1741-1753.

Counterman B, Araujo-Perez F, Hines H, Baxter S, Morrison C, Lindstrom1 D, Papa R, Ferguson L, Joron M, Ffrench-Constant R, *et al*. 2010. Genomic hotspots for adaptation: the population genetics of mullerian mimicry in *Heliconius erato*. *PLoS Genetics* 6: e1000796.

Eyre-Walker AR, Gaut L, Hilton H, Eldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. *Proceedings of the National Academy of Sciences* 95(8): 4441-4446.

Garris AJ, McCouch SR, Kresovich S. 2003. Population structure and its effects on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa L.*). *Genetics* 165(2): 759-769.

Goldsmith MR, Shimada T, Abe H. 2005. The genetics and genomics of the silkworm. *Bombyx mori*. *Annual Review of Entomology* 50: 71-100.

Hall TA. 1999. Bioedita user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.

Hamblin MT, Casa AM, Sun H, Murray SC, Paterson AH, Aquadro CF, Kresovich S. 2006. Challenges of Detecting Directional Selection After a Bottleneck: Lessons From *Sorghum bicolor. Genetics* 173(2): 953-964.

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glemin S, David J. 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Molecular Biology and Evolution*. 24(7): 1506-1517.

Hill WG, Robertson A. 1968. Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 38(6): 226-231.

Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111(1): 147-164.

Hudson RR. 1987. Estimating the recombination parameter of a finite population model without selection. *Gentical Research* 50(3): 245-250.

Hudson, RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18(2): 337-338.

Kolkman JM, Berry ST, Leon AJ, Slabaugh MB, Tang S, Gao W, Shintani DK, Burke JM, Knapp SJ. 2007. Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics* 177(1): 457-468.

Labate JA, Robertson LD, Baldo AM. 2009. Multilocus sequence data reveal extensive departures from equilibrium in domesticated tomato (*Solanum lycopersicum L.*). *Heredity* 103(3): 257-267.

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

Li X, Tan L, Zhu Z, Huang H, Liu Y, Hu S, Sun C. 2009. Patterns of nucleotide diversity in wild and cultivated rice. *Plant systematics and evolution* 281: 97-106.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11): 1451-1452.

Liu A, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* 173(1): 321-330.

Lu C, Yu SH, Xiang ZH. 2002. Molecular systematic studies on Chinese *mandarina* silkworm (*Bombyx manderina M.*) and domestic silkworm (*Bombyx mori L.*). *Scieatia Agricultura Sinica* 35(1): 94-101.

Matzkin LM. 2008. The molecular basis of host adaptation in Cactophilic *Drosophila*: molecular evolution of a Glutathione S-Transferase gene (*GstD1*) in *Drosophila mojavensis*. *Genetics* 178(2): 1073-1083.

Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan MD. 2006. Selection under domestication: evidence for a sweep in the rice *waxy* genomic region. *Genetics* 173(2): 975-983.

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

Raymond M, Rousset F. 1995. GENEPOP (version 1.2), population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86(3): 248-249.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler IV ES. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences* 98(20): 11479-11484.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nature Biotechnology* 26(10): 1135-1145.

Tanaka H, Ishibash J, Fujita K, Nakajima Y, Sagisaka A, Tomimoto K, Suzuki N, Yoshiyama M, Kaneko Y, Iwasaki T, *et al.* 2008. A genome-wide analysis of genes and gene families involved in innate immunity of *Bombyx mori*. *Insect Biochemistry and Molecular Biology* 38(12): 1087-1110.

Tang T, Lu J, Huang J, He J, McCouch SR, Shen Y, Kai Z, Purugganan MD, Shi S, Wu CI. 2006. Genomic variation in rice: genesis of highly polymorphic linkage blocks during domestication. *Plos Genetics* 2: e199.

Tanksley S, McCouch SR. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277(5329): 1063-1066.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. 2001. Patterns Nature of the Drugs of the Peal Bag in Songs (*Zea mays ssp. mays L.*). *Proceedings of the National Academy of Sciences* 98(16): 9161–9166.

Tenaillon MI, U'ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. *Molecular Biology and Evolution* 21(7): 1214-1225.

Thompson, JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25(24): 4876-4882.

Tian F, Stevens NM, Buckler IV ES. 2009. Tracking footprints of maize Higginson syringe. *Proceedings of the National Academy of Sciences* 106(Supplement 1): 9979-9986.

Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proceedings of the National Academy of Sciences* 99(15): 9650-9655.

Weiss, G, Haeseler A. 1998 Inference of population history using a likelihood approach. *Genetics* 149(3): 1539-1546.

Wilcoxon F. 1945. Individual comparisons by ranking methods. *Biometrics* 1(6): 80-83.

White GM, Hamblin MT, Kresovich S. 2004. Molecular evolution oagronomic cropreplacement evolution. *Molecular Biology and Evolution* 21(4): 716-723.

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(5726): 1310-1314.

Xia Q, Guo Y, Zhang Z, Li D, Xuan Z, Li Z, Dai F, Li Y, Cheng D, Li R, *et al*. 2009. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science* 326(5951): 433-436.

Xiang ZH, Huang JT, Xia JG, Lu C. 2005. *Biology of Sericulture*. China Forestry Publishing House.

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S. 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice.

Molecular Biology and Evolution 24(3): 875-888.

Zhu YL, Song QJ, Hyten DL, Von Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB. 2003. Single-nucleotide polymorphisms in soybean. *Genetics* 163(3): 1123-1134.

Table 2. Summary of genes surveyed and primer sequences employed in this study.

Gene	Chromosome location	Alignment length (bp)			Primer Sequences	Functional Association
		Total	Coding	Noncoding		
Wnt-1	4	916	529	387	AATCGCTTGGACTCGCTCG	Wingless-1
					GGGACCGTTATCTGGCTTCG	
ER	10	788	336	452	CGGTTCGCTCGCTCTATGT	Ecdysone Receptor
					ACCTGTTCGCACGCTGTTC	
GST	6	794	150	644	TTCTGTTATTTCCTCCCTTAG	Glutathione-S-transferase
					ACAAACCACCGTTATTACTCAT	
Amy	nscaf2827*	1015	250	765	TCAGTCGCAACGAATACA	α-Amylase
					CATCGCTGCCTACTACAA	
Adh	11	1678	1131	547	GCAATGGTCGTGGTCTAACTC	Alcohol Dehydrogenase
					TGGCTGGATAGGCTCAAT	
DefA	4	809	148	661	GAACCAAAGCCATTTGCT	Defensin A
					TTGTGGACTCACTCGCAGA	
PBAN	11	983	240	743	CAAAGAAGATCATCTTCACCCCC	Pheromone Biosynthesis
					GTACGAAAGCTCCCTTCCCAGTC	Activating Neuropeptide
AchE	nscsaf2877*	1187	1187	0	ACTCGCAGACTACTTCAGA	Acetylcholinesterase 2
					ATTGATAGGACGAACTTGA	
RpSA	nscsaf3045*	1098	333	765	AACGATGTCGGGAGGATTA	ribosomal protein SA
					ATGGGTTGATGGTCTTGTGC	
Total		9268	4304	4964		

* Scaffold ID from SilkDB (http://silkworm.swu.edu.cn/silkdb/)

Table 3. Summary of diversity in *B. mori* and *B. mandarina.*

Locus	Species	n	S	π_{total}	θ_w	S_{coding}	π_{silent}	θ_{silent}	π_a	π_s	π_a/π_s	Tajima's D	R_m	R
Wnt-1	Domesticated	16	31	0.0082	0.0102	21	0.01209	0.0135	0.0022	0.0358	0.059	-0.77	1	0.0406
	Wild	15	45	0.0092	0.0151	26	0.01464	0.02396	0.0023	0.0346	0.066	-1.67	3	0.0744
ER	Domesticated	16	14	0.0032	0.0054	7	0.0035	0.00506	0.0025	0.003	0.832	-1.58	0	0.0866
	Wild	15	25	0.0052	0.0098	9	0.00622	0.0109	0.0032	0.0061	0.517	-1.92*	0	0.0313
GST	Domesticated	16	48	0.016	0.0177	7	0.01729	0.01995	0.0046	0.0418	0.107	-0.4	5	0.0174
	Wild	15	75	0.0212	0.0285	6	0.02379	0.03702	0.0033	0.0338	0.097	-1.12	7	0.0091
α-Amy	Domesticated	15	60	0.0155	0.0181	8	0.01867	0.02208	0.0025	0.0271	0.091	-0.62	8	0.0067
	Wild	15	107	0.0246	0.0302	8	0.029	0.0348	0.0007	0.0311	0.022	-0.81	10	0.0232
Adh	Domesticated	14	29	0.0043	0.0054	8	0.00809	0.00956	0.0007	0.0103	0.065	-0.89	0	0.0094
	Wild	14	73	0.0089	0.0137	22	0.01696	0.02562	0.0012	0.0202	0.057	-1.55	2	0.0131
DefA	Domesticated	16	4	0.0006	0.0014	0	0.00072	0.00174	0	0	—	-1.83*	0	0
	Wild	15	85	0.0202	0.0322	14	0.02066	0.03312	0.0158	0.0118	1.339	-1.63	4	0
PBAN	Domesticated	16	56	0.0228	0.0182	4	0.02676	0.02154	0.0073	0.003	2.409	1.09	0	0.0014
	Wild	15	98	0.0312	0.0326	10	0.03639	0.03886	0.01	0.0132	0.76	-0.18	12	0.0693
RpSA	Domesticated	14	70	0.02685	0.01957	4	0.03221	0.02404	0	0.01084	0	1.64915	2	0.0016
	Wild	14	93	0.02731	0.02659	8	0.0349	0.03386	0.00058	0.0209	0.027	0.12187	7	0.0279
AchE	Domesticated	11	18	0.0075	0.0052	18	0.02969	0.01909	0.00099	0.02969	0.033	1.99589*	2	0.0009
	Wild	14	37	0.00892	0.00982	37	0.03886	0.04219	0.00016	0.03859	0.004	-0.39617	5	0.0496
Total/	Domesticated		330	0.01166	0.01124	77	0.01656	0.01517	0.00231	0.01795	0.4	-0.21726	2	0.0183
Average	Wild		638	0.01741	0.02206	140	0.0246	0.03115	0.00414	0.02337	0.321	-0.90429	5.6	0.0331

* p <0.05; n: sample size,; S: the number of segregating sites; πtotal: the mean number of nucleotide differences per site based on entire locus; θw: watterson's estimator of 4Neμ based on entire locus; πsilent: π values for silent sites (synonymous and noncoding sites); θsilent: θ values for silent sites (synonymous and noncoding sites); Rm: estimates of minimum number of recombination events (Hudson and Kaplan 1985); R: the estimate of the population recombination parameter per site (Hudson 1987); Scoding: number of segregating sites in coding regions; πa: average pairwise differences calculated on nonsynonymous sites; πs: average pairwise differences based on synonymous sites; πa/πs: ratio of πa to πs;

Table 1. Characters and source of materials used in this study

Species	No.	Strain/population name	System/location	Voltine*
Bombyx mori	S1	Zhugui	Zhejiang, China	CV1
	S2	Huangbo	China	CV1
	S3	Sichuansanmian	Sichuan, China	CV1
	S4	Qiansanmian	Guizhou, China	CV1
	S5	Yindusanmian	India	T
	S6	Chishu	Japan	JV1
	S7	Ou18	Europe	EV1
	S8	Yi16	Europe	EV1
	S9	J115	Japan	JV2
	S10	872	Japan	JV2
	S11	DazaoN	China	CVd
	S13	WuG	Ukraine	EV2
	S14	WuF	Ukraine	EV2
	S15	Cambodia	Cambodia	T
	S16	BaixiaB	China	CV3
	S17	C108N	China	CV2
Bombyx mandarina	Y1	Huzhou	Huzhou, Zhejiang	Wild
	Y2	Nanchong	Nanchong, Sichuan	Wild
	Y3	Ankang	Ankang, Shanxi	Wild
	Y4	Yichang	Yichang, Hubei	Wild
	Y5	Luzhou	Luzhou, Sichuan	Wild
	Y6	Lixian	Lixian, Hunan	Wild
	Y7	Suzhou	Suzhou, Jiangsu	Wild
	Y8	Qingmuguan	Qingmuguan, Chongqing	Wild
	Y9	Rongchang	Rongchang, Chongqing	Wild
	Y11	Ziyang	Ziyang, Sichuan	Wild
	Y12	Pengshan	Pengshan, Sichuan	Wild
	Y13	Hongya	Hongya, Sichuan	Wild
	Y14	Xiangyun	Xiangyun, Yunnan	Wild
	Y15	Hushuguan	Hushuguan, Jiangsu	Wild
	Y16	Hefei	Hefei, Anhui	Wild

* CV1: Chinese univoltine; CV2: Chinese biovoltine; CV3: Chinese trivoltine; CVd: Chinese multi-voltine; JV1: Japanese univoltine; JV2: Japanese biovoltine; EV1: European univoltine; EV2: European biovoltine; T: Tropical

Table 4. Reduction of diversity in domesticated species relative to their wild species.

Species	Ratio of diversity reduction	Measure	Reference
Silkworm	33% or 49%	π or θ_w	This Study
Maize	43%	θ_w	Wright et al. 2005
Maize	38%	θ_{silent}	Tenaillon et al. 2004
Sunflower	60% or 50%	π or θ_w	Liu and Burke 2006,
Asian Rice (*indica*)	63% or 57%	π or θ_w	Caicedo et al. 2007
Asian Rice (*japonica*)	69% or 67%	π or θ_w	Caicedo et al. 2007
Rice (*indica*+ *japonica*)	30%	θ_w	Li et al. 2009