家蚕基因组学国家重点实验室
国家重点基础研究发展计划（973）项目

DOMESTICATED
SILKWORM
GENOME
RESEARCH
2008～2009

主　编　夏庆友　向仲怀

家蚕基因组研究

2008~2009

上

西南师范大学出版社
人民出版社

家蚕基因组学国家重点实验室
国家重点基础研究发展计划（973）项目

DOMESTICATED SILKWORM GENOME RESEARCH

2003～2009

家蚕基因组研究

2008～2009 上

主 编 夏庆友 向仲怀

西南师范大学出版社
人民出版社

DOMESTICATED SILKWORM GENOME RESEARCH 2003～2009

编委会

前言

近年来，中国家蚕基因组计划和功能基因研究发展十分迅速，不但在国内已经成为我国重要动植物功能基因组研究代表之一，在国际上也具有相当的优势和广泛的影响。为了及时总结家蚕基础研究成果，2008 年，我们编辑出版了《家蚕基因组计划 2000～2007》，以家蚕基因组测序、生物信息学分析和平台技术建立等为主，全面总结了 2000～2007 年的 8 年间家蚕基因组计划的主要研究成果。《家蚕基因组计划 2000～2007》的出版，得到业界许多同行专家学者的肯定，并荣获“重庆市首届出版政府奖——优秀图书奖”(2010 年)。得益于大家的支持和鼓励，我们才编辑了《家蚕基因组研究 2008～2009》。同时，我们也希望将家蚕基因组研究成果作为一个系列丛书，将来陆续编辑出版。

家蚕基因组和功能基因组研究，始终围绕着三个中心任务展开，即家蚕基因组计划、家蚕功能基因研究平台和关键技术，以及家蚕功能基因组研究。

首先，就家蚕基因组计划而言，2000 年以来取得了三个标志性成果。第一个是家蚕基因组框架图谱的绘制。如果从 2000 年组织筹划算起，到 2004 年在 *Science* 上发表家蚕基因组框架图，历时 4 年。家蚕基因组框架图谱相关成果，已经收录于《家蚕基因组计划 2000～2007》一书之中。第二个是家蚕基因组精细图谱。该图谱的绘制由中国和日本联合完成。2008 年，中日两国科学家达成一致，由日方完成家蚕 BAC 测序和分子连锁图谱构建，并将日方 3 倍家蚕基因组 WGS 数据并入中国 6 倍 WGS 数据，共同完成了家蚕 9 倍精细图谱，以中方为第一完成者于 2008 年以特刊形式在最重要的昆虫学专业杂志 *Insect Biochemistry and Molecular Biology* 上发表。2009 年在 *Nucleic Acids Research* 上发表了基于家蚕精细图谱的信息库升级版 SilkDBv2.0。第三个是家蚕高精度遗传变异图谱的绘制。家蚕自 5 000 年前由中国野桑蚕驯化而来，形成了大量地理品系和基因突变体材料。在基因组水平上解析家蚕和野桑蚕主要地理品系的遗传变化，具有重要科学价值。为此，我们采用 Illumina-Solexa 测序技术，选取代表性的 29 个家蚕突变品系和 11 个不同地理来源的中国野桑蚕

品系进行了全基因组重测序，共获得 632.5 亿对碱基，测序深度达到 118 层，覆盖了 99.8%的基因组区域。通过数据分析，在家蚕和中国野蚕基因组之间共发现 1 600 万个 SNP 位点、31 万个插入缺失突变和 3.5 万个基因组结构变异。基因组水平的群体遗传分析表明，家蚕由中国野桑蚕而来的驯化变异是单一的驯化事件造成的，过去长期受人们关注的化性特征却并不能完全反映家蚕的起源驯化历程，这是科学家首次在全基因组水平上对家蚕起源进化关系进行的深入探究。最为重要的是，比较分析还发现 1 041 个基因组区域及 354 个蛋白编码基因受到了驯化和人工选择压力的影响，它们主要参与调控蚕的丝蛋白合成、能量代谢、生殖特性、飞行能力等生理行为。该成果于 2009 年 8 月发表于 *Science* 上，国际权威专家对研究成果给予了高度评价。上述这三个标志性成果的取得，意味着家蚕基因组学研究的主体任务已经基本完成。在框架图谱、精细图谱和遗传变异图谱三个阶段，中国科学家都做出了决定性贡献，并牢牢占领了世界优势地位。今后一个时期，家蚕基因组生物学研究将会集中在比较基因组学、遗传变异、人工驯化和 W 染色体全序列分析等方面。有关家蚕基因组框架图谱和信息分析的成果，主要收录于《家蚕基因组计划 2000～2007》一书中，而精细图谱和遗传变异图谱成果，则主要收录于本书之中。

其次，是关于家蚕功能基因组研究平台和关键技术。家蚕基因组计划的主要意义在于获取家蚕全基因组序列、结构信息和遗传变异信息等，以便我们鉴定家蚕所有可能的蛋白质编码基因和具有功能意义的非编码基因等。然而，这不是我们的最终目的，最重要的工作是发现基因、研究基因和利用基因。也就是说，序列分析只是为功能研究服务的。由于基因组和功能基因组研究都是物种全基因组水平上的大规模和系统研究，因此，需要一些大规模的平台和关键技术。如果说结构基因组，即序列测定和分析的理论和技术具有相当程度的普适性，不同的物种之间通用性较强的话，那么，相对而言，功能基因组研究平台技术则具有更多的物种特殊性。因此，家蚕基因组计划完成之后，建立适应于家蚕功能基因组研究的关键平台技术就显得十分的重要。目前为止，我们已基本建立了家蚕发现基因、研究基因和利用基因三个方面的平台。其中，发现基因关键平台技术包括家蚕遗传资源、生物信息分析、蛋白质组学、基于基因芯片的表达谱分析和突变基因定位克隆等，基因功能研究关键技术包括转基因家蚕、家蚕 RNAi 技术和分子生物学一般方法等，基因应用关键技术包括实用品种转基因、基因表达遗传调控等。这些平台和关键技术的主体内容，主要完成于 2000～2007 年期间，相关成果也已收录于《家蚕基因组计划 2000～2007》中。本书则主要涉及一些进一步完善和利用的研究内容。基因的功能鉴定和利用，其本质是基因功能的丧失、获得和调控，而目前已有的技术虽然能满足基本的要求，但仍然存在较大的缺陷。主要的原因是 RNAi 技术应用于家蚕，还存在特异性和适应性方面的不足，使得只有少部分基因能够通过该方法实现基因功能丧失。将来比较重要的发展方向是建立家蚕基因敲除技术。所幸的是，目前我们已经取得了较大进展，但因文

章处于未发表阶段，因此相关内容只能在本书后续的册子中加以总结。总体上分析，我们认为，建立家蚕功能基因组平台和关键技术的主体任务已经完成。未来的主要工作，将进一步集中在家蚕基因的功能研究方面。

最后是家蚕功能基因组研究，这是家蚕基因组和功能基因组研究的主战场。在国家重大基础研究项目“家蚕主要经济性状功能基因组与分子遗传改良研究”(“973”计划：2005CB121000)的支持下，我们初步建立了以家蚕丝蛋白合成、家蚕变态发育、家蚕性别决定和家蚕免疫四个重要生物学性状为主的功能基因组研究体系，并取得了较大的进展。《家蚕基因组计划 2000～2007》一书中收录了主要性状基因作用网络的生物信息学鉴定、关键基因克隆等内容。本书则更多地集中于一些关键基因的功能研究和调控机制分析等方面，研究水平和深度都有很大的进步。除了这四大性状相关功能基因组的研究外，本书还收集了大量有关家蚕重要突变、重要生理生化过程、人工选择与进化、家蚕重要病原微生物基因组和家蚕遗传素材创新等方面的研究论文，也总体上反映了我国家蚕基础研究的整体实力和水平。鉴于家蚕结构基因组和平台建设的主体任务已经完成，家蚕基因组计划的主体工作将集中于功能研究方面，所以，从本集开始，书名将采用《家蚕基因组研究》，以后根据研究论文的数量，以适当的周期陆续编辑出版。

以本书所覆盖的年份(2008 年)开始，“发现基因，研究基因，利用基因”已成为家蚕基因组研究的主旋律。从这个意义上讲，本书将发挥承上启下、继往开来的作用。家蚕基因组研究的主要科学目标，将仍然是振兴蚕丝产业、推进模式昆虫和开拓生物新兴产业。我们在做好扎实的基础工作的同时，还特别希望研究家蚕的人更多一些，研究进展更快一些，研究水平更高一些，研究成果离实际应用更近一些。这也是我们通过本书的出版，最想表达的团队意志和愿望。

本书的出版，得到了西南师范大学出版社和人民出版社的大力支持。特别是西南师范大学出版社周安平社长的工作团队，为本书付出了大量的心血。谨此深表谢意。由于论文收集自不同的国内外杂志，体例格式与现行出版要求不统一之处较多，为保持原貌仅做部分修改。由于编者的水平和能力限制，尚有许多不足之处，敬请读者批评指正。

2010 年 12 月于重庆

DOMESTICATED SILKWORM GENOME RESEARCH 2008~2009

目录

Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*)

Qing-you Xia[1,2,*], Yi-ran Guo[3,*], Ze Zhang[1,2,*], Dong Li[1,3,*], Zhao-ling Xuan[3,*], Zhuo Li[3,*], Fang-yin Dai[1], Ying-rui Li[3], Dao-jun Cheng[1], Rui-qiang Li[3,4], Ting-cai Cheng[1,2], Tao Jiang[3], Celine Becquet[5,†], Xun Xu[3], Chun Liu[1], Xing-fu Zha[1], Wei Fan[3], Ying Lin[1], Yi-hong Shen[1], Lan Jiang[3], Jeffrey Jensen[5], Ines Hellmann[5], Si Tang[5], Ping Zhao[1], Han-fu Xu[1], Chang Yu[3], Guo-jie Zhang[3], Jun Li[3], Jian-jun Cao[3], Shi-ping Liu[1], Ning-jia He[1], Yan Zhou[3], Hui Liu[3], Jing Zhao[3], Chen Ye[3], Zhou-he Du[1], Guo-qing Pan[1], Ai-chun Zhao[1], Hao-jing Shao[3,7], Wei Zeng[3], Ping Wu[3], Chun-feng Li[1], Min-hui Pan[1], Jing-jing Li[3], Xu-yang Yin[3], Da-wei Li[3], Juan Wang[3], Hui-song Zheng[3], Wen Wang[3], Xiu-qing Zhang[3], Song-gang Li[3], Huan-ming Yang[3], Cheng Lu[1], Rasmus Nielsen[4,5], Ze-yang Zhou[1,6], Jian Wang[3], Zhong-huai Xiang[1,‡], Jun Wang[3,4,‡]

(*1. The Key Sericultural Laboratory of Agricultural Ministry, College of Biotechnology, Southwest University, Chongqing 400715, China; 2. Institute of Agronomy and Life Sciences, Chongqing University, Chongqing 400044, China; 3. BGL-Shenzhen, Shenzhen 518083, China; 4. Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Kbh Ø, Denmark; 5. Departments of Integrative Biology and Statistics, University of California Berkeley, Berkeley, CA 94720, USA; 6. Chongqing Normal University, Chongqing 400047, China; 7. Innovative Program for Undergraduate Students, South China University of Technology, Guangzhou 510006, China*)

Abstract: A single-base-pair resolution silkworm genetic variation map was constructed from 40 domesticated and wild silkworms, each sequenced to approximately threefold coverage, representing 99.88% of the genome. We identified ~16 million single-nucleotide polymorphisms, many indels, and structural variations. We find that the domesticated silkworms are clearly genetically differentiated from the wild ones, but have maintained large levels of genetic variability, suggesting a short domestication event involving a large number of individuals. We also identified signals of selection at 354 candidate genes that may have been important during domestication, some of which have

* These authors contributed equally to this work. † Present address: Institute for Human Genetics, University of California San Francisco, San Francisco, CA 94143-0794, USA. ‡ To whom correspondence should be addressed. E-mail: wangj@genomics.org.cn (J. W.); xbxzh@swu.edu.cn (Z. X.)

enriched expression in the silk gland, midgut, and testis. These data add to our understanding of the domestication processes, and may have applications in devising pest control strategies and advancing the use of silkworm as efficient bioreactors.

Keywords: Silkworm; SNP; GROSS; Domestication

The domesticated silkworm, *Bombyx mori*, has a mid-range genome size of ~432 Mb[1], is the model insect for the order Lepidoptera, has economically important values (*e.g.*, silk and bioreactors production), and has been domesticated for more than 5,000 years[2]. Because of human selection, silkworms have evolved complete dependence on humans for survival[3] and more than 1,000 inbred domesticated strains are kept worldwide[3]. Archaeological and genetic evidences indicate that the domesticated silkworm originated from the Chinese wild silkworm, *Bombyx mandarina*, that is found throughout Asia, where modern sericulture and silkworm domestication were initiated.

The origin of the domesticated silkworm is a long-standing question that has not been settled by previous limited biochemical and molecular analyses. Two hypotheses suggested a unique domestication but disagreed on the ancestral variety. One hypothesis, based on isoenzyme polymorphism, proposed mono-voltinism as ancestral variety (voltinism represents number of generations per annum), from which bi- and multi-voltine were derived by artificial selection[4]; the other proposed the reverse path considering evidence from archaeology, history, and genetics[5]. An alternative hypothesis based on random amplification of polymorphic DNA indicatede that the ancestral domestic silkworm strains were issued, not from a unique variety, but from mixed geographic locations and ecological types[6]. These theories are conflicting, probably because they were derived from incomplete genetic information. Consequently, we present here a genome-wide detailed genetic variation map in hopes to help reconstruct the silkworm domestication history.

The data consisted of 40 samples from 29 phenotypically and geographically diverse domesticated silkworm lines [categorized by geographical regions[3]: Chinese, Japanese, Tropical, European lineages, and the mutant system], as well as 11 wild silkworms from various mulberry fields in China (Table S1). We sequenced each genome at approximately threefold coverage, after creating single- and paired-end (PE) libraries with inserts of PEs ranging from base pairs 137 to 307 bp[7].

Raw short reads were mapped against the refined 432-Mb reference genome from *Dazao*[1] with the program SOAP[8]. We pooled all reads from the 40 complete genomes and identified 15,986,559 ingle-nucleotide polymorphisms (SNPs) using SoapSNP[7,9] (Table S3A). The accuracy of the SNP calling was evaluated with Sequenom (San Diego, California) genotyping of a representative subset of variants in all 40 varieties, resulting in a 96.7% validation rate[7].

We then pooled separately all 29 domesticated strains and all 11 wild varieties and obtained SNP sets for each[7]. The number of SNPs in the domestic versus wild varieties was 14,023,573 and 13,237,865, respectively (Table S3A). To account for the different number of domestic and wild strains, we used the population-size scaled mutation rate θ_S to measure genetic variation[10] (Table S3B). We found that $\theta_{S,\text{domesticated}}$ (0.0108) was significantly smaller than $\theta_{S,\text{wild}}$ (0.0130) [Mann-Whitney U (MWU), $P = 1.10 \times 10^{-7}$], which may reflect differences in effective population size and demographic history (including domestication and artificial se-

lection). The rate of heterozygosity in domesticated strains was more than twofold lower than that of wild varieties(0.0032 versus 0.0080, respectively) (MWU, $P = 3.33 \times 10^{-6}$). This reduction in heterozygosity is most likely due to inbreeding or the bottleneck experienced by domesticated lines.

In addition to SNPs, we also identified 311,608 small insertion-deletions (indels) (Table S4A), a subset of which were validated with polymerase chain reaction[7]. The θ_S values for the indels (Table S4B) were in agreement with a lower effective population size in domesticated versus wild varieties. A mate-pair relationship method[7,11] identified 35,093 structural variants (SVs) among the 40 varieties (Table S5). Over three-fourths of the SVs overlapped with transposable elements (TEs), suggesting that SV events in silkworm are likely due to TE content[12] and mobility[11]. The SNPs, indels, and SVs all contributed to a comprehensive genetic variation map for the silkworm.

To elucidate the phylogeny of silkworms beyond previous studies[6,13,14], we used our identified SNPs to estimate a neighbor-joining tree[7] on the basis of a dissimilarity measure of genetic distance (Figure 1A). This tree represents an average of distances among strains, so lineages cannot be directly interpreted as representing phylogenetic relationships. Instead, the distances may reflect gene flow and other population level processes related to human activities such as ancient commercial trade. the unrooted radial relationship reveals a clear split between the domesticated and wild varieties, and the domestic strains cluster into several subgroups (Figure 1A).

A principle component analysis (PCA)[7] classified the first four eigenvectors as significant (Table S6; Tracy-Widom, $P<0.05$). The first eigenvector clearly separates the domesticated and wild varieties, whereas the second eigenvector divides the domesticated strains into subgroups correlated with voltinism (Figure 1B, top). The third principle component separates D01 and D03, (which are high-silk producing Japanese domesticated strains) from the other domesticated strains, whereas the fourth separates W01 and W04 from the other wild varieties (Figure 1B, bottom). Results of population structure analysis[7] (Figure S3) confirmed the results of the PCA, as well as the neighbor joining analysis. The clear genetic separation between domesticated and wild varieties suggests a unique domestication event and relatively little subsequent gene flow between the two groups.

One puzzling observation is that, although domesticated strains are clearly genetically differentiated from the wild ones, they still harbor ~83% of the variation observed in the wild varieties. This suggests that the population-size bottleneck at domestication only reduced genetic variability mildly[7]; that is, a large number of individuals must have been selected for initial domestication or else domestication occurred simultaneously in many places. To quantify this observation, we fit a simple coalescence-based genetic bottleneck model to the SNP frequency spectrum[7]. The estimated model suggests that the domestication event lead to a 90% reduction in effective population size during the initial bottleneck (Figure S2). Additionally, we observed no excess of low-frequency variants in the domesticated varieties compared with the wild varieties, suggesting that there has not been obvious population growth since the domestication event and that the domestic lines probably have had a generally stable effective population size.

Our measure of pairwise linkage disequilibrium (LD)[7] showed that LD decays rapidly in

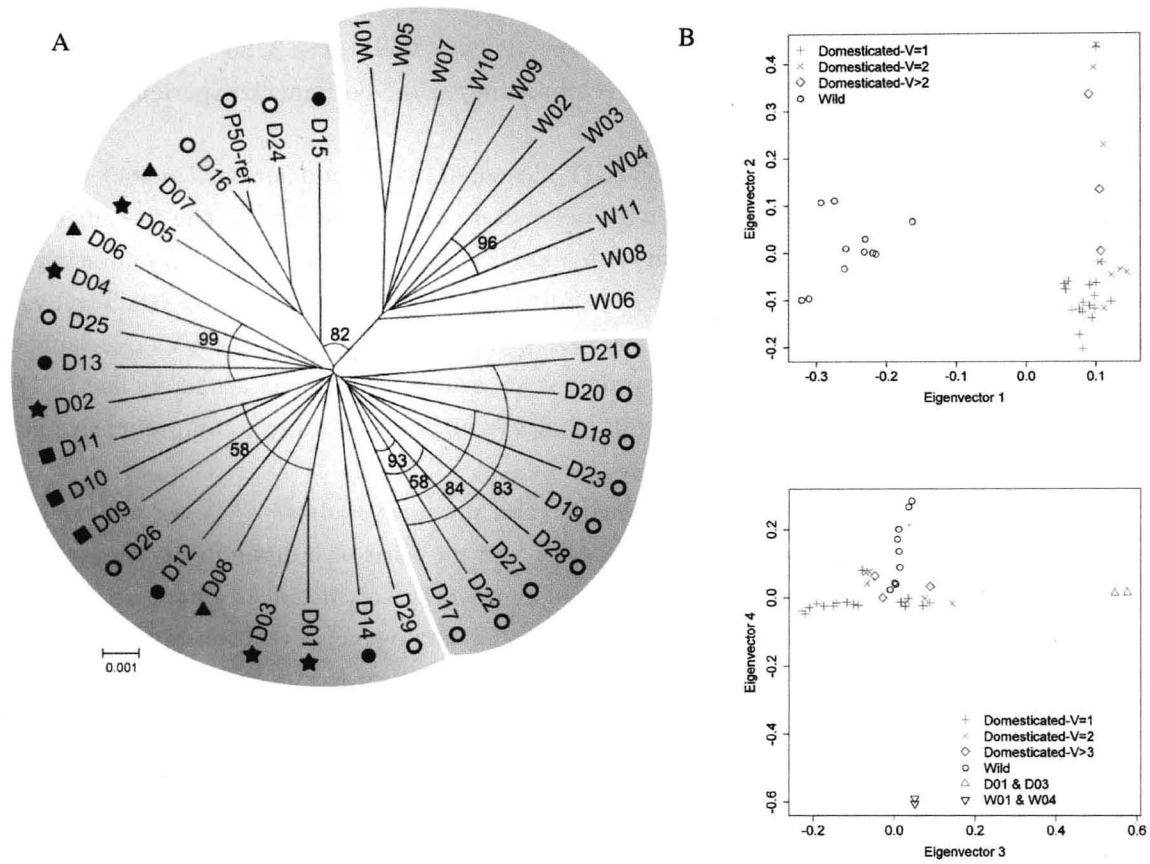

Figure 1. Silkworm phylogeny and population structure from PCA. (A) A Neighbor-Joining tree from genomic SNPs, bootstrapped with 1,000 replicates (bootstrap values less than 100 are shown on arcs; those equal to 100 are not shown); green for all wild varieties; others are domesticated strains separated into three groups (purple, red, and yellow). Domesticated strains are denoted by a combination of symbols representing silkworm systems (hollow circles, Chinese; stars, Japanese; triangles, tropical; squares, European; filled circles, mutant system) and sample IDs (D01 to D29 and P50-ref for the reference genome of Dazao). Wild varieties are indicated by their IDs (W01 to W11). Scale bar, frequencies of base-pair differences. (B) PCA results of the first four statistically significant components (Tracy-Widom, $P<0.05$). (Top) the first eigenvector separates domesticated and wild varieties, and the second divides the domesticated strains into subgroups. (Bottom) the third eigenvector separating the high-silk production Japanese domesticated strains D01 and D03 from the other domesticated strains, and the fourth separates the wild varieties W01 and W04 from the other wild varieties

silkworms, with correlation coefficient r^2 decreasing to half of its maximum value at distances of ~46 and 7 base pairs for the domesticated and wild varieties, respectively (Figure S1). The fast decay of LD implies that regions affected by selective sweeps are probably relatively small. To detect regions with significant (Z test, $P<0.005$) signatures of selective sweep, we measured SNP variability and frequency spectrum following a genome-wide sliding window strategy[7] (Figure 2A). Though the significance of our Z-tests[7] cannot be interpreted literally due to correlations in LD and shared ancestral history between the two populations, they Z test suggest differences in frequency spectra and amounts of variability between the two groups. We termed the candidate regions genomic regions of selective signals (GROSS).

We identified a total of 1,041 GROSS[7], covering 12.5 Mb (2.9%) of the genome, which may reflect genomic footprints left by artificial selection during domestication. A re-

gion affected by selective sweep typically has an elevated level of LD[15,16], and in our GROSS, the level of LD among SNP pairs less than 20-kb apart was 2.3 times higher than genome average (Figure 2B), consistent with the hypothesis that selection is affecting these regions. In all these regions, divergence levels[7] between the domesticated and wild groups were also elevated (Figure 2C), confirming the differentiation of the two subpopulations.

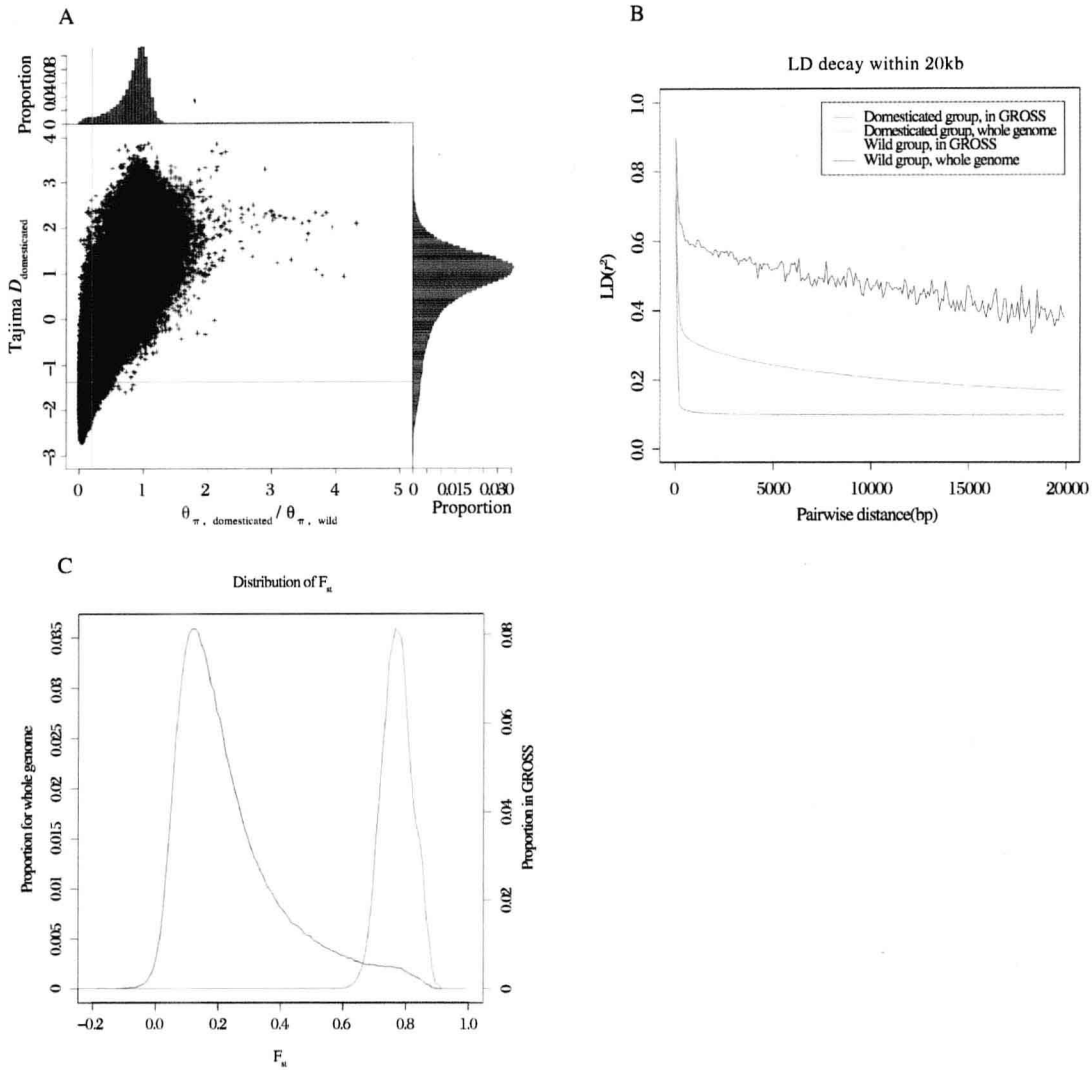

Figure 2 Genomic regions of selective signals (GROSS). (A) Two-dimensional distribution for $\theta_{\pi,\text{domesticated}}/\theta_{\pi,\text{wild}}$ and Tajima's D for domesticated silkworms. 5-kb windows, data points of which locate to the left of the vertical red line (corresponding to Z test $P<0.005$) and below the horizontal red line (also Z test $P<0.005$) were picked out as building blocks of GROSS. (B) Linkage disequilibrium (LD) in GROSS. For domesticated silkworms, LD decays much more slower in GROSS than in the whole genome, whereas for wild varieties, no obvious change in the pattern was observed. (C) Distribution of divergence between domesticated and wild groups in GROSS versus whole genome (F_{st})[7]

B. mori has experienced intense artificial selection, represents a completely domesticated insect[3], and has become totally dependent on humans for survival. Artificial selection has also enhanced important economic traits such as cocoon size, growth rate, and digestion efficiency[3]. Moreover, compared to its wild ancestor *B. mandarina*, *B. mori* has gained

some representative behavioral characteristics (such as tolerance to human proximity and handling, as well as extensive crowding) and lost other traits (such as flight, predators, and diseases avoidance). However, to date no genes have been identified as domestication genes under artificial selection. Within GROSS, we identified 354 protein-coding genes that represent good candidates for domestication genes (Table S9). Their Gene Ontology annotation[17] showed the most representation in the category of "binding" and "catalytic" in molecular function, as well as "metabolic" and "cellular" in biological process (Figure S4).

Considering published expression profiles performed on different tissues in fifth-instar day 3 of Dazao with genome-wide microarray[18], we found that 159 of our GROSS genes exhibit differential expression. Of these, 4, 32, and 54 genes are enriched in tissues of silk gland, midgut, and testis, respectively (Figure S5). Among the genes enriched in the silk gland is silk gland factor-1 (*Sgf-1*), a homolog of a *Drosophila melanogaster Fkh* gene. *Sgf-1* regulates the transcription of the *B. mori* glue protein-encoding *sericin-1* gene and of three fibroin genes encoding fibroin light chain, fibroin heavy chain and *fhx/P25*[19,20]. Another silk gland enriched gene, *BGIBMGA005127*, homologous to the *Drosophila sage* gene, was overexpressed fourfold in a high-silk strain compared with *Dazao* (Figure S6). In *Drosophila*, the products of *Fkh* and *sage* genes cooperate to regulate the transcription of the glue genes *SG1* and *SG2*, which are crucial for the synthesis and secretion of glue proteins[21,22]. Additionally, midgut- and testis- enriched genes suggest that genes involved in energy metabolism and reproduction have also been under artificial selection during domestication[7]. Specifically, we identified three likely candidate for artificial selection: (i) *NM 001130902* is homologous to paramyosin protein in *Drosophila* and may be related to flight[23]; (ii) *NM 001043506* is homologous to fattyacyl desaturase (*desat1*) in *Drosophila*, which is related to courtship behaviors, because mutations in *desat1* can change the pattern of sex pheromones production and discrimination[24]; finally, (iii) *BGIBMGA000972* is homologous to tyrosine-protein kinase *Btk29A* in *Drosophila*, which is involved in male genitalia development[25].

In sericulture, silkworms are typically categorized by their geographic origins[3]. Voltinism, which results from adaption to ecological conditions, and geographic systems have been central to previous studies of silkworm origin and domestication[4~6]. Our findings indicate that a unique domestication event occurred and, although voltinism correlates with genetic distances, major genetically cohesive strains cannot be identified on the basis of voltinism. We observed no correlation between longitudes of the sample origins and any of the principle components, but we did find a significant correlation between the latitudes and eigenvectors 2 and 4 in the PCA (Table S7). Although this correlation might be due to isolation by distance, this result also agrees with previous studies suggesting that climate affects silkworm biology[2].

The silkworm data reported here represents the largest body of genome sequences for a lepidopteran species and offers a source of near-relatives in this clade for comparative genomic analysis. We further proposed a set of candidate domestication genes that, in addition to being putatively under artificial selection, also show higher expression levels in tissues important for silkworm economic traits. Because a proportion of the GROSS genes were probably important in domestication, functional verification of these candidate genes may enable a comprehensive understanding of the differences of biological characteristics between *B. mori* and *B. mandarina*. Moreover, domesticated silkworms have been used as bioreactors[26,27],

and such an effort may provide useful clues to help improve the capacity and capability of silkworm to produce foreign proteins[26]. These findings may also aid in the understanding of how to enhance traits of interest in other organisms in an environmentally safe manner and, because the wild silkworm is a destructive pest, allow new approaches for pest control.

References and notes

[1]The International Silkworm Genome Consortium, *Insect Biochem. Mol. Biol.* 38, 1036 (2008).

[2]Z. Xiang, J. Huang, J. Xia, C. Lu, *Biology of Sericulture*. (China Forestry Publishing House, Beijing, 2005).

[3]M. R. Goldsmith, T. Shimada, H. *Abe*, *Annu. Rev. Entomol.* 50, 71 (2005).

[4]N. Yoshitake, *J. Sericult. Sci. Japan* 37,83(1967).

[5]Y. Jiang, *Agric Archaeol.* 14, 316 (1987).

[6]C. Lu, H. Yu, Z. Xiang, *Agric. Sci.* China 1, 349 (2002).

[7]Materials and methods are available as supporting material on *Science* Online.

[8]R. Li, Y. Li, K. Kristiansen, J. Wang, *Bioinformatics* 24, 713 (2008).

[9]R. Li *et al.*, *Genome Res.* 19, 1124 (2009).

[10]G. A. Watterson, *Theor. Popul. Biol.* 7, 256 (1975).

[11]J. Wang *et al.*, *Nature* 456, 60 (2008).

[12]Biology Analysis Group *et al.*, *Science* 306,1937(2004).

[13]Q. Xia, Z. Zhou, C. Lu, Z. Xiang, *Acta Entomol. Sinica* 41, 32 (1998).

[14]M. Li et *al.*, *Genome* 48, 802 (2005).

[15]R. Nielsen, *Annu. Rev. Genet.* 39, 197 (2005).

[16]M. Slatkin, *Nat. Rev. Genet.* 9, 477 (2008).

[17]J. Ye *et al.*, *Nucleic Acids Res.* 34, W293 (2006).

[18]Q. Xia *et al.*, *Genome Biol.* 8, R162 (2007).

[19]B. Horard, E. Julien, P. Nony, A. Garel, P. Couble, *Mol. Cell. Biol.* 17, 1572 (1997).

[20]V. Mach et al., *J. Biol. Chem.* 270, 9340 (1995).

[21]E. W. Abrams, W. K. Mihoulides, D. J. Andrew, *Development* 133, 3517 (2006).

[22]T. R. Li, K. P. White, *Dev. Cell* 5, 59 (2003).

[23]H. Liu *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* 102,10522(2005).

[24]F. Marcillac, Y. Grosjean, J. F. Ferveur, *Proc. Biol. Sci.* 272, 303 (2005).

[25]K. Baba *et al.*, *Mol. Cell. Biol.* 19, 4405 (1999).

[26]S. Maeda, *Annu. Rev. Entomol.* 34, 351 (1989).

[27]S. Maeda *et al.*, *Nature* 315, 592 (1985).

[28]We thank two anonymous referees, L. Goodman, L. Bolund and K. Kristiansen for providing valuable comments. This work was supported by the Ministry of Science and Technology of China (grants 2005CB121000, 2007CB815700, 2006AA10A117, 2006AA10A118, 2006AA02Z177, 2006AA10A121), the Ministry of Education of China (Program for Changjiang Scholars and Innovative Research Team in University, grant IRT0750), Chongqing Municipal Government, the 111 Project (grant B07045), the National Natural Science Foundation of China (grants 30725008; 30890032; 90608010), the International Science and Technology Cooperation Project(grant 0806), the Chinese Academy of Science (grant GJHZ0701－6), the Danish Platform for Integrative Biology, the Ole RØmer grant from the Danish Natural Science Research Council, and the Solexa project (grant 272－07－0196). Raw genome data are deposited in the National Center for Biotechnology Information/Short Read archive database with accession number SRA009208; silkworm genetic variations, GROSS information and microarray data can be found in http://silkworm.swu.edu.cn/silkdb/resequencing.html.

Supporting Online Materials for

Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*)

Qing-you Xia, Yi-ran Guo, Ze Zhang, Dong Li, Zhao-ling Xuan, Zhuo Li, Fang-yin Dai, Ying-rui Li, Dao-jun Cheng, Rui-qiang Li, Ting-cai Cheng, Tao Jiang, Celine Becquet, Xun Xu, Chun Liu, Xing-fu Zha, Wei Fan, Ying Lin, Yi-hong Shen, Lan Jiang, Jeffrey Jensen, Ines Hellmann, Si Tang, Ping Zhao, Han-fu Xu, Chang Yu, Guo-jie Zhang, Jun Li, Jian-jun Cao, Shi-ping Liu, Ning-jia He, Yan Zhou, Hui Liu, Jing Zhao, Chen Ye, Zhou-he Du, Guo-qing Pan, Ai-chun Zhao, Hao-jing Shao, Wei Zeng, Ping Wu, Chun-feng Li, Min-hui Pan, Jing-jing Li, Xu-yang Yin, Da-wei Li, Juan Wang, Hui-song Zheng, Wen Wang, Xiu-qing Zhang, Song-gang Li, Huan-ming Yang, Cheng Lu, Rasmus Nielsen, Ze-yang Zhou, Jian Wang, Zhong-huai Xiang *, Jun Wang *

This word file includes:

1 Materials and methods

1.1 Sample collection

In order to include major silkworm systems kept in the laboratories worldwide, we collected strains from diverse geographic regions, such as China, Japan, Europe and tropical areas (mostly southeast Asian: India, Cambodia and Laos), as well as silkworms from the mutant system. All 29 domesticated samples listed in Table S1 are from the Institute of Sericulture and Systems Biology in Southwest University of China. Two important developmental characteristics, voltinism (number of generations per year) and moltinism (number of larval molts per generation), and sex were recorded for each of those 29 domesticated silkworms. Of these, 18 are monovoltine, 8 are bivoltine and others are polyvoltine. We also captured 11 wild silkworms from mulberry fields in China, facilitating the comparative analysis between domesticated and wild groups.

* Corresponding author. E-mails: wangj@genomics.org.cn (J. W.) and xbxzh@swu.edu.cn (Z. X.)