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

下

西南師範大學出版社
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

2008～2009

家蚕基因组研究

2008～2009 下

主 编 夏庆友 向仲怀

西南师范大学出版社
人民出版社

图书在版编目(CIP)数据

家蚕基因组研究2008～2009/夏庆友,向仲怀主编.—重庆:西南师范大学出版社,2011.3
ISBN 978-7-5621-5186-9

Ⅰ.①家… Ⅱ.①夏… ②向… Ⅲ.家蚕-基因组-研究 Ⅳ.S885.9

中国版本图书馆CIP数据核字(2011)第038074号

家蚕基因组研究2008～2009

主　编　夏庆友　向仲怀

出 版 人:周安平
责任编辑:杜珍辉　郑持军　杨光明　伯古娟
封面设计:王　煤　帖　鑫
版式设计:梅木子
出版发行:西南师范大学出版社
地址:重庆市北碚区天生路2号
邮编:400715
http://www.xscbs.com
人民出版社
地址:北京市朝阳门内大街166号
邮编:100706
http://www.peoplepress.net
经　　销:全国新华书店
印　　刷:重庆康豪彩印有限公司
开　　本:889mm×1194mm 1/16
印　　张:52.75
字　　数:1400千字
版　　次:2011年5月　第1版
印　　次:2011年5月　第1次印刷
书　　号:ISBN 978-7-5621-5186-9

定　　价:360.00元(上下册)

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

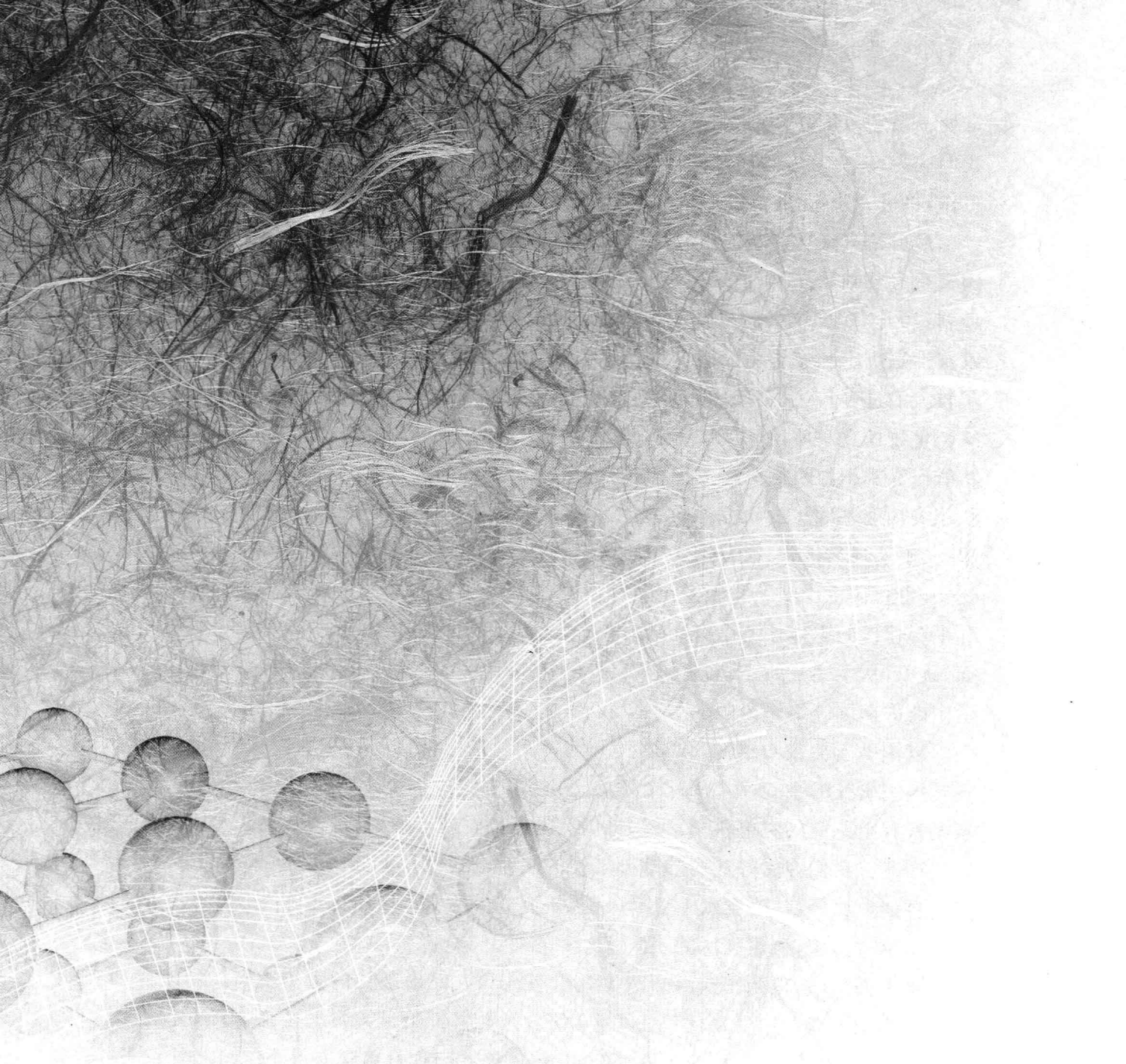
章处于未发表阶段，因此相关内容只能在本书后续的册子中加以总结。总体上分析，我们认为，建立家蚕功能基因组平台和关键技术的主体任务已经完成。未来的主要工作，将进一步集中在家蚕基因的功能研究方面。

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

Characterization of the mitochondrial genome of the Chinese wild mulberry silkworm, *Bombyx mandarina* (Lepidoptera: Bombycidae)

Min-hui Pan[1], Quan-you Yu[1], Yu-ling Xia[1], Fang-yin Dai[1], Yan-qun Liu[2], Cheng Lu[1,*], Ze Zhang[1], Zhong-huai Xiang[1]

(1. *Key Sericultural Laboratory of Agricultural Ministry, Southwest University, Chongqing 400716, China*; 2. *College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang 110161, China*)

Abstract: The complete mitochondrial genome of Chinese *Bombyx mandarina* (Ch*Bm*) was determined. The circular genome is 15,682 bp long, and contains a typical gene complement, order, and arrangement identical to that of *Bombyx mori* (*B. mori*) and Japanese *Bombyx mandarina* (Ja*Bm*) except for two additional tRNA-like structures: $tRNA^{Ser(TGA)}$-like and $tRNA^{Ile(TAT)}$-like. All protein-coding sequences are initiated with a typical ATN codon except for the COI gene, which has a 4-bp TTAG putative initiator codon. Eleven of 13 protein-coding genes (PCGs) have a complete termination codon (all TAA), but the remaining two genes terminate with incomplete codons. All tRNAs have the typical clover-leaf structures of mitochondrial tRNAs, with the exception of $tRNA^{Ser(TGA)}$-like, with a four stem-and-loop structure. The length of the A+T-rich region of Ch*Bm* is 484 bp, shorter than those of Ja*Bm* (747 bp) and *B. mori* (494～499 bp). Phylogenetic analysis among *B. mori*, Ch*Bm*, Ja*Bm*, and *Antheraea pernyi* (*Anpe*) showed that *B. mori* is more closely related to Ch*Bm* than *JaBm*. The earliest divergence time estimate for *B. mori* - Ch*Bm* and *B. mori* - JaBm is about 1.08±0.18～1.41±0.24 *mya* and 1.53±0.20～2.01±0.26 *mya*, respectively. Ch*Bm* and Ja*Bm* diverged around 1.11±0.16～1.45±0.21 *mya*.

Keywords: Chinese *Bombyx mandarina*; mtDNA; Phylogeny; Origin of *Bombyx mori*

1 Introduction

Metazoan mitochondrial DNA (mtDNA) is a double-stranded, circular molecule, ranging in size from 14 to 39 kb, that encodes 13 protein coding genes (PCGs), 2 rRNA genes, and 22 tRNA genes [1]. Although the gene content of the metazoan mitogenome is highly con-

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

served, some exceptions have been reported. For example, *Chrysomya putoria* (Diptera) and *Bombus ignites* (Hymenoptera) contain a variable number of tRNA genes[2,3]. The insect mitochondrial genome contains an A+T− rich control region that has a higher level of sequence and length variability than other regions of the genome[4,5,6].

Bombyx mori (mulberry silkworm) is the only truly domesticated insect. It is thought that *B. mori* was domesticated about 5,000～10,000 years ago from the ancient *Bombyx mandarina*, the wild mulberry silkworm which is now distributed throughout areas of Asia[7]. However, there is great diversity in wild mulberry silkworms. For example, two types of wild silkworms with a different number of chromosomes per haploid genome are currently found in mulberry fields. One, living in Japan and some regions of Korea, has 27 chromosomes per haploid genome[8,9]. The second type, found in China, has 28 chromosomes per haploid genome, like the domesticated silkworm, *B. mori*[10,11]. Although, a few studies have shown that *B. mandarina* is probably a close relative of *B. mori*[12,13], the question of whether *B. mori* has a mono-or polyphyletic origin remains unclear.

The use of mtDNA has become popular in phylogenetics and population genetic studies at the genomic level[14] because mitochondrial DNA is maternally inherited and has a higher mutation rate than the nuclear genome[15,16]. Recently, Yukuhiro *et al*.[12] have sequenced the mitochondrial genome of Japanese *B. mandarina* (Ja*Bm*) and found significant sequence divergence between the Ja*Bm* and *B. mori* mitochondrial genomes. Based on the referenced nucleotide substitution rate taken from the beetle mitochondrial *nad5* gene, the estimate of *B. mori* and Ja*Bm* divergence was about 7.1 *mya*, well before domestication[12]. Here we present the complete mitochondrial genome sequence of the Chinese *B. mandarina* (Ch*Bm*) and compare it with those of the Ja*Bm* and *B. mori*. Using the available complete mtDNA sequences of several *B. mori* strains, wild silkworms, and *Antheraea pernyi* (Lepidoptera: Saturniidae), phylogenetic analyses were carried out based on the concatenated amino acid sequences of protein coding genes. We found that, consistent with chromosome number data, the domesticated silkworm, *B. mori*, is more closely related to Ch*Bm* than to Ja*Bm*.

2 Materials and methods

2.1 *Insect and mtDNA extraction*

Wild silkworms were collected from Ankang in the Shanxi Province of China. Mitochondrial DNA was extracted from pupaes using a standard protocol for genomic DNA extraction[17].

2.2 *Primer design, PCR, and sequencing*

The primer pairs used for PCR amplification were designed by Oligo software (version 6.0) based on the complete mtDNA sequence of the Japanese wild silkworm[12]. PCR was performed using the following cycling protocol: 94 ℃ for 2 min (initial denaturation); 35 cycles of 94 ℃ for 30 s (denaturation), 34 ℃～50 ℃ for 40 s (annealing), and 72 ℃ for 45～60 s (extension); and 72 ℃ for 10 min (final elongation). The amplification products were isolated using the QIAquick Gel Extraction Kit (Qiagen, Germany) prior to direct sequencing with the ABI PRISM™ 310 Genetic Analyzer (PE Applied Biosystems, USA).

2.3 *Sequence analysis*

Ch*Bm* mitochondrial genes and A + T − rich region were found by aligning the Ch*Bm* mtDNA sequence with homologous regions of full-length insect mitochondrial sequences using CLUSTAL X[18]. Nucleotide sequences of the protein-coding genes were translated based on the invertebrate mtDNA genetic code. The tRNA genes were located based on comparisons with the mitochondrial tRNA genes of other insects, including Ja*Bm*, *B. mori*, and *Bombus ignitus* [3,12], and were confirmed by the respective anticodon sequences. Secondary structures of the tRNA genes were predicted using tRNAscan-SE program[19]. Protein coding genes and ribosomal RNA genes were identified by their similarity to published *B. mori* mitochondrial sequences. The sequence data have been deposited in GenBank under accession NO. AY301620.

2.4 *Phylogenetic analysis*

The complete mtDNA sequences of four *B. mori* strains (AB070264 for C108, AB083339 for Aojuku, AY048187 for Xiafang, and AF149768 for Backokjam), Ja*Bm* (NC_003395), and Antheraea pernyi (NC_004622) were retrieved from GenBank for comparison with Ch*Bm*. The amino acid sequences of the 13 protein-coding genes were concatenated for phylogenetic analysis. Concatenated sequences from the seven mitochondrial genomes were aligned with Clustal X, and Neighbor-joining (NJ) trees were reconstructed by MEGA version 2.1[20]. Poisson correction distances were used for NJ. The robustness of the NJ trees was evaluated by bootstrapping (1,000 resamplings for NJ). The maximum likelihood (ML) tree was reconstructed by quartet puzzling using TREE-PUZZLE 5.0 [21]. A substitution model mtREV 24 with two rates (1 invariable + 1 variable) [22] was used in the ML analysis and the fraction of invariable sites was estimated from the data set.

The concatenated amino acid sequences of the *COI* + *COII* genes from the seven mtDNA genomes were aligned with Clustal X 1.8. Pairwise genetic distances were calculated by MEGA version 2.1 using the Kimura two-parameter model for nucleotide substitution[23]. The molecular clock was based on the substitution rate of $(7.8\sim10.2)\times10^{-9}$ per site per year[24] in the *COI* + *COII* genes of the swallowtail butterfly of the genus *Papilio* (Lepidoptera: Papilionidae). Estimates of the divergence time (T) between the domesticated silkmoth, *Bombyx mori*, and its close relative species were derived from the molecular clock (r) and genetic distance (K) using the simple equation $T = K/(2r)$[25].

3 Results

3.1 *Genome structure, organization and composition*

The complete nucleotide sequence of Chinese *B. mandarina* (Ch*Bm*) mtDNA was determined and submitted to GenBank (*accession no.* AY301620). The 15,682-bp genome contains 13 protein coding genes, 2 rRNA genes, 22 tRNA genes, and an A + T-rich region. Gene content and arrangement were identical to those of *B. mori* and Japanese *B. mandarina* (Ja*Bm*)[12], except for two additional tRNA-like structures in Ch*Bm*: $tRNA^{Ser(UGA)}$-like and $tRNA^{Ile(UAU)}$-like. Excluding the tRNA-like sequences, Ch*Bm* mitochondrial genes over-

lap by a total of 45 bp at 11 locations (Table 1). Ch*Bm* mitochondrial genes are separated by 875 bp of the intergenic sequence, which is spread over 17 regions (391 bp) and the A + T – rich region (484 bp). The intergenic spacers ranges in size from 1 to 59 bp, with the longest spacer located between $tRNA^{His}$ and ND4 (Table 1).

Table 1. Organization of the mitochondrial genome of Ch*Bm*, Ja*Bm* and *B. mori* C108

Genes or region	Location	Size (bp)	Start Codon	Stop Codon	3′ Spacer[a]	Strand[b]	Ja*Bm* size (bp)	*B. mori* C108 size (bp)
COIII	Join (15134-15682, 1-240)	789	ATG	TAA	2	H	798	798
$tRNA^{Gly}$	243-311	69			3	H	66	66
ND3	315-665	351	ATA	TAA	53	H	351	351
$tRNA^{Ala}$	719-786	68			50	H	64	68
$tRNA^{Arg}$	837-900	64			0	H	64	64
$tRNA^{Asn}$	901-967	67			0	H	67	65
$tRNA^{Ser}$	968-1034	67			10	H	67	67
$tRNA^{Glu}$	1045-1107	63			–1	H	64	65
$tRNA^{Phe}$	1107-1173	67			4	L	67	67
ND5	1178-2896	1719	ATT	TAA	16	L	1,719	1,719
$tRNA^{His}$	2913-2981	69			59	L	67	67
ND4	3041-4384	1344	ATG	TAA	–1	L	1,341	1,341
ND4L	4384-4674	291	ATG	TAA	4	L	291	291
$tRNA^{Thr}$	4679-4744	66			–1	H	66	65
$tRNA^{Pro}$	4744-4812	69			1	L	68	66
ND6	4814-5344	531	ATA	TAA	55	H	531	531
Cytb	5400-6557	1158	ATA	TAA	–1	H	1,158	1,152
$tRNA^{Ser}$	6557-6622	66			25	H	66	66
ND1	6648-7592	945	ATA	TAA	–6	L	945	945
$tRNA^{Leu}$	7587-7657	71			33	L	71	70
16S rRNA	7691-9040	1350			0	L	1,377	1,378
$tRNA^{Val}$	9041-9111	71			0	L	72	69
12S rRNA	9112-9895	784			0	L	783	783
A + T rich region	9896-10379	484			0	H	747	494
$tRNA^{Ser(UGA)}$ *-like*	9965-10055	91	–	–	–	H	–	–
$tRNA^{Met}$	10380-10447	68			–2	H	68	68
$tRNA^{Ile}$	10446-10511	66			–3	H	66	66
$tRNA^{Gln}$	10509-10577	69			48	L	69	69
$tRNA^{Ile(UAU)}$ *-like*	10570-10639	70	–	–	–14	H	–	–
ND2	10626-11648	1023	ATA	TAA	5	H	1,023	1,023
$tRNA^{Trp}$	11654-11723	70			–8	H	70	70
$tRNA^{Cys}$	11716-11782	67			6	L	67	67
$tRNA^{Tyr}$	11789-11854	66			17	L	66	66
COI	11872-13402	1531	TTAG	T+ +	0	H	1,531	1,531
$tRNA^{Leu}$	13403-13469	67			0	L	67	67
COII	13470-14151	682	ATG	T+ +	0	H	682	682
$tRNA^{Lys}$	14152-14222	71			–1	H	71	71
$tRNA^{Asp}$	14222-14288	67			0	H	67	67
ATP8	14289-14450	162	ATA	TAA	–7	H	162	162
ATP6	14444-15121	678	ATG	TAA	–14	H	678	678

Note: [a] Negative numbers indicate that adjacent genes overlap. [b] H, heavy strand; L, light strand

The nucleotide composition of the Ch*Bm* mitogenome (Table 2) is biased toward adenine and thymine (81.59%), as are the genomes of Ja*Bm* (81.68%) and *B. mori* (81.32%). The A+T content of the Ch*Bm* A+T− rich region is 94.42%, lower than that of both Ja*Bm* (95.18%) and *B. mori* (95.39%). The AT skew (Perna and Kocher, 1995) is slightly positive and the GC skew is negative in the mitogenomes of Ch*Bm*, Ja*Bm* and *B. mori*, indicating an obvious bias toward the use of As and Cs, but the possible reasons for this bias are not well understood.

Table 2. Skewed nucleotide composition in regions of the Ch*Bm*, Ja*Bm* and *B. mori* C108 mitochondrial genomes

Species	Length (bp)	A%	G%	T%	C%	A+T %	AT skew	GC skew
Whole mtDNA								
B. mori(C108)	15,643	43.05	7.32	38.27	11.36	81.32	0.058	−0.215
Ja*Bm*	15,928	43.08	7.21	38.60	11.11	81.68	0.055	−0.213
Ch*Bm*	15,682	43.11	7.40	38.48	11.01	81.59	0.057	−0.196
PCGs								
B. mori(C108)	11,170	42.91	8.17	36.66	12.27	79.57	0.079	−0.201
Ja*Bm*	11,194	42.78	8.14	36.86	12.22	79.64	0.074	−0.200
Ch*Bm*	11,197	42.83	8.26	37.04	11.87	79.87	0.072	−0.179
tRNAs								
B. mori(C108)	1,462	42.20	7.80	39.33	10.67	81.53	0.035	−0.155
Ja*Bm*	1,463	41.90	7.79	39.71	10.59	81.61	0.026	−0.152
Ch*Bm*	1,472	41.78	7.81	39.95	10.46	81.73	0.022	−0.145
rRNA								
B. mori(C108)	2,158	43.74	4.59	41.06	10.61	84.84	0.032	−0.397
Ja*Bm*	2,160	43.89	4.63	41.30	10.19	85.19	0.030	−0.375
Ch*Bm*	2,134	43.86	4.78	41.50	10.31	84.91	0.028	−0.366
A+T rich region								
B. mori(C108)	499	44.69	1.60	50.70	3.01	95.39	−0.063	−0.305
Ja*Bm*	747	45.52	2.41	49.67	2.41	95.18	−0.043	0.000
Ch*Bm*	484	46.49	2.69	47.93	2.89	94.42	−0.015	−0.036

3.2 Protein-coding genes

The 13 protein-coding genes in the mitogenome of Ch*Bm* begin with typical ATN codons: 6 have putative ATA initiator codons, 5 use putative ATG initiation codons, and ND5 has a putative ATT initiator. However, *COI* may use TTAG, a 4-bp initiation codon that is also found in Ja*Bm* [12].

The 13 PCGs of Ch*Bm* have characteristics very similar to those of Ja*Bm* and *B. mori*. Eleven of the protein-coding genes use a putative TAA stop codon, while *COI* and *COII* use an incomplete stop codon (T). Like the mtDNA of Ja*Bm*, the Ch*Bm* Cytb gene is 6 bp longer than that of *B. mori* C108 [12]. The Ch*Bm* mitochondrial *ND4* gene is 3 bp longer than those of Ja*Bm* and *B. mori* C108. The remaining 11 PCGs are the same length in all three genomes (Table 1).

A+T content skews slightly positive, and G+C skews negatively, in the 13 PCGs in the

mitogenomes of Ch*Bm*, Ja*Bm* and *B. mori* (Table 2). Although A and T nucleotides are overrepresented in all three codon positions of each mitochondrial genome, each codon position has a different AT/GC bias. In the first codon position, Ch*Bm*, Ja*Bm*, and *B. mori* are biased toward the use of A and G, A and C, and T and C, respectively. The second codon position of the three mitogenomes is biased toward T and G. In the third codon position, Ch*Bm* and *B. mori* are biased toward the use of T and G, while there is a preference for T and C in Ja*Bm* (Table 3).

Table 3. Codon usage in the protein-coding genes of Ch*Bm*, Ja*Bm* and *B. mori* C108

Species	First position				Second position				Third position			
	A	T	G	C	A	T	G	C	A	T	G	C
Ch*Bm*	38.2	36.9	16.1	8.8	22.9	50.7	13.4	13.0	40.6	52.7	3.4	3.3
Ja*Bm*	37.9	36.8	10.9	14.4	25.3	47.9	16.1	10.7	41.5	47.7	4.2	6.6
B. mori C108	37.3	37.6	9.4	15.7	22.2	48.6	15.9	13.3	43.6	49.2	4.5	2.7

3.3 Transfer RNA and ribosomal RNA genes

tRNA gene structure was predicted by the tRNAscan-SE Search Server (Lowe and Eddy, 1997). The results revealed different $tRNA^{Gly}$, $tRNA^{Val}$, $tRNA^{Ala}$, $tRNA^{His}$, and $tRNA^{Glu}$ sequences in the *ChBm* and Ja*Bm* mitochondrial genomes, and predicted different $tRNA^{Gly}$, $tRNA^{Leu(TAG)}$, $tRNA^{Pro}$, $tRNA^{Val}$, $tRNA^{His}$, and $tRNA^{Glu}$ structures in the two genomes (Figure 1). These changes are primarily the result of insertions and deletions.

There are two more tRNA-like sequences in Ch*Bm* mitogenome than in Ja*Bm* and *B. mori* (Table 1). The $tRNA^{Ser(TGA)}$-like sequence is encoded on the H strand, which is located within the A+T-rich region (9,965~1,055 bp), and forms a structure with four stem-loops and one big loop. The $tRNA^{Ile(ATA)}$-like sequence is located between nucleotides 10,570 and 10,639, which overlaps the $tRNA^{Gln}$ (9 bp) and *ND2* (13 bp) genes, and forms the typical clover-leaf structure (Figure 2). Although the gene content of the metazoan mitogenome is highly conserved, some exceptions have been reported. For example, *Chrysomya putoria*, which belongs to the order Diptera, has an extra $tRNA^{Ile}$ (Junqueira *et al.*, 2004.), while several species of Hymenoptera contain variable numbers of tRNA genes[3,26,27]. However, this is the first report of a variable number of tRNA genes in Lepidoptera.

The two rRNA genes in Ch*Bm* are located between $tRNA^{Leu}$ and the A+T-rich region. The size of Ch*Bm* *lr*RNA (1,350 bp) is shorter than the *lr*RNA of Ja*Bm* (1,377 bp) and *B. mori* (1,378 bp). Conversely, the size of *sr*RNA is 784 bp, and is similar to that of Ja*Bm* (783 bp) and *B. mori* (783 bp) (Table 1). The Ch*Bm* bias toward the use of A and C in the tRNA and rRNA genes is reflected in Ja*Bm* and *B. mori* (Table 2).

3.4 A+T-rich region

The size of the A+T-rich region in insects varies considerably in different taxa, and even within the same species. Length variations between closely related taxa are generally the result of different numbers of tandem repeats in the A+T-rich region [28]. A striking difference between the mitochondrial genomes of the Chinese and Japanese wild silkworms is the

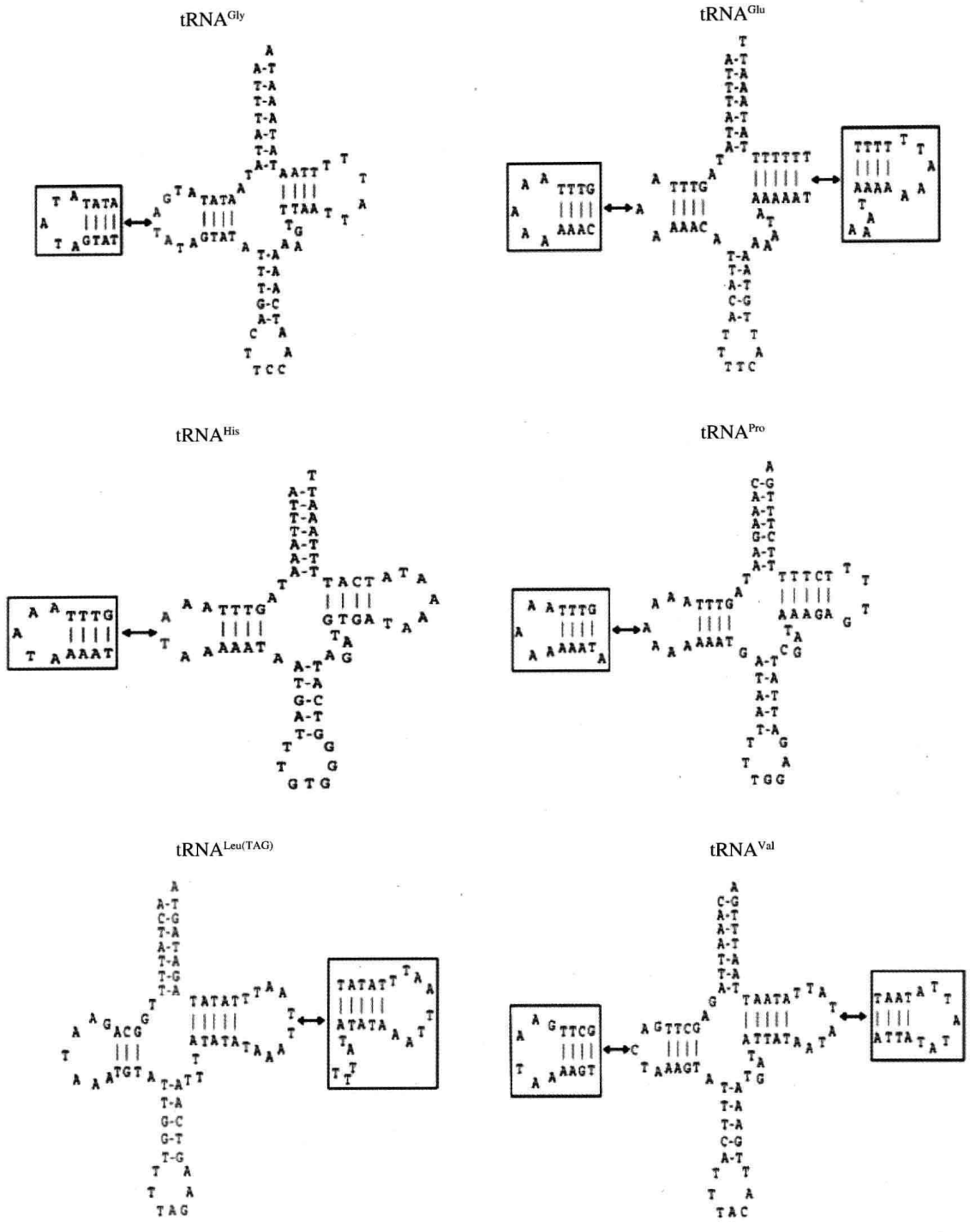


Figure 1. Differences in mitochondrial tRNA secondary structure between Ch*Bm* and Ja*Bm*. Clover-leaf structures of Ch*Bm* tRNAs are shown, with boxed sections indicating those regions that are modified in Ja*Bm* tRNAs

length of the A + T − rich region (Table 1). The former has a much shorter A + T − rich region (484 bp) than the latter (747 bp). It must be noted that the length of the A + T − rich region of the Ch*Bm* mitochondrial genome is similar to that of *B. mori* (494～499bp)[12,29]. The A + T − rich region of Ja*Bm* has three tandem repeats of a 126-bp element, but only one of the repeat elements is found in the A + T − rich regions of Ch*Bm* and *B. mori* [12,13].