

Research Article

Isolation of BAC Clones Containing Conserved Genes from Libraries of Three Distantly Related Moths: A Useful Resource for Comparative Genomics of Lepidoptera

Yuji Yasukochi,¹ Makiko Tanaka-Okuyama,¹ Manabu Kamimura,¹ Ryo Nakano,^{2,3} Yota Naito,⁴ Yukio Ishikawa,² and Ken Sahara⁴

¹Insect Genome Research Unit, National Institute of Agrobiological Sciences, Owashi 1-2, Tsukuba, Ibaraki 305-8634, Japan

²Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan

³Entomology Research Team, National Institute of Fruit Tree Science, National Agriculture and Food Research Organization, Tsukuba, Ibaraki 305-8605, Japan

⁴Laboratory of Applied Molecular Entomology, Research Institute of Agriculture, Hokkaido University, N9, W9, Kita-ku, Sapporo 060-8589, Japan

Correspondence should be addressed to Yuji Yasukochi, yaskoch@affrc.go.jp

Received 30 June 2010; Accepted 25 October 2010

Academic Editor: Xin-yuan Guan

Copyright © 2011 Yuji Yasukochi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lepidoptera, butterflies and moths, is the second largest animal order and includes numerous agricultural pests. To facilitate comparative genomics in Lepidoptera, we isolated BAC clones containing conserved and putative single-copy genes from libraries of three pests, *Heliothis virescens*, *Ostrinia nubilalis*, and *Plutella xylostella*, harboring the haploid chromosome number, $n = 31$, which are not closely related with each other or with the silkworm, *Bombyx mori*, ($n = 28$), the sequenced model lepidopteran. A total of 108–184 clones representing 101–182 conserved genes were isolated for each species. For 79 genes, clones were isolated from more than two species, which will be useful as common markers for analysis using fluorescence *in situ* hybridization (FISH), as well as for comparison of genome sequence among multiple species. The PCR-based clone isolation method presented here is applicable to species which lack a sequenced genome but have a significant collection of cDNA or EST sequences.

1. Introduction

Bacterial artificial chromosome (BAC) libraries play a critical role in determination of genome organization and chromosome walking. In addition, we have utilized BAC libraries for linkage analysis and physical mapping of Lepidoptera, butterflies and moths. We reported construction of the first lepidopteran BAC library from the silkworm, *Bombyx mori* [1], which was used for characterization of the *Hox* gene cluster [2]. BAC clones isolated with monomorphic sequence tagged sites (STSs) were utilized for finding polymorphisms from flanking regions of the original STSs and construction of BAC contigs covering 22% of the *B. mori* genome, enabling us to localize genes which could not be mapped genetically

[3]. We showed that BAC clones could be used effectively as probes for fluorescence *in situ* hybridization (FISH) in this species despite the limited cytological differentiation of individual chromosomes [2–4] and defined a karyotype for *B. mori* using this technique [4].

BAC libraries have since been constructed for several other lepidopteran species [5–9] and used to reveal longer range genome organization [8–11]. We previously showed that the gene order is well conserved between *B. mori* and the tobacco horn worm, *Manduca sexta*, by mapping *M. sexta* orthologs of 124 conserved and putative single-copy genes using BAC-FISH technology, which is suitable for genetically uncharacterized species [12]. However, *B. mori* and *M. sexta* belong to the same superfamily, Bombycoidea, and analysis

of other major lepidopteran groups is necessary to determine to what extent synteny exists at the chromosomal level among a wide range of Lepidoptera.

Noctuoidea is the largest family of Lepidoptera and includes many serious and globally distributed agricultural pests. Therefore, we selected the tobacco budworm, *Heliothis virescens*, as a target for our studies since a BAC library [7] and more than 60,000 expressed sequenced tags (ESTs) [13–15] were available. The European corn borer, *Ostrinia nubilalis*, was also appropriate for this analysis due to the availability of a BAC library (<http://www.genome.clemson.edu/services>) and ESTs [16, 17]. *O. nubilalis* belongs to the superfamily, Pyraloidea, which forms a different clade from Bombycoidea and Noctuidae but is closer to them than butterflies (Figure 1) [18]. Another species with available BACs was the diamondback moth, *Plutella xylostella*, which is well known to develop resistance rapidly to a wide variety of insecticides [19–21]. *P. xylostella* belongs to the superfamily, Yponomeutoidea, which is primitive compared with Macrolepidoptera, which contains the other two species, but all belong to the same major group of advanced Lepidoptera, Ditrysia (Figure 1) [18].

Here, we describe isolation of BAC clones containing conserved genes from these three distantly related species, *H. virescens*, *O. nubilalis*, and *P. xylostella*. A total of 458 clones were isolated by PCR-based screening of BAC libraries. These will be a useful resource for comparative genomics in Lepidoptera.

2. Materials and Methods

2.1. cDNA Sequencing of *O. nubilalis*. Total RNA was isolated from tissues and whole bodies of embryos, larvae, pharate pupae, pupae, and adults. cDNA was synthesized using a Super SMART PCR cDNA synthesis kit (Clontech) and cloned into a pGEM-T easy plasmid vector (Promega). A total of twelve cDNA libraries were constructed, and nucleotide sequences of randomly selected clones were determined using an ABI3730 DNA sequencer (Applied Biosystems) (Table S1). The DNA sequences were analyzed using a BLASTx program (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) to estimate their function.

2.2. Cloning of *O. nubilalis* Genes. Single-step or nested PCR amplifications were performed using degenerate primers designed from conserved genes previously mapped in *B. mori*. Single-step PCR amplification was performed using genomic DNA as template except for the small heat shock protein genes (Table S2), for which RT-PCR and 3'-RACE analyses were performed (Table S2). Other sequences were amplified from cDNA by nested PCR (Table S2). PCR products were then cloned into a pGEM-T easy plasmid vector (Promega) and sequenced with an ABI3730 DNA sequencer.

2.3. Selection of Genes for BAC Isolation. Sequences of *H. virescens*, *O. nubilalis*, and *P. xylostella* genes and ESTs were obtained from public databases or by EST sequencing as

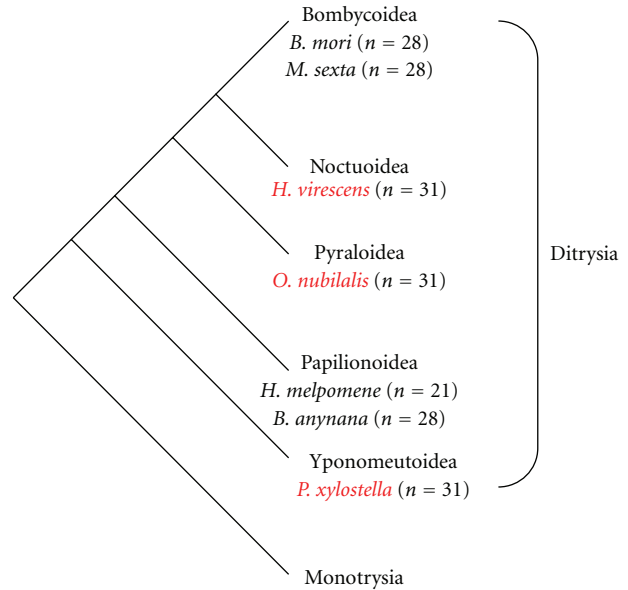


FIGURE 1: Schematic representation of phylogeny of Lepidoptera and haploid chromosome numbers of species referred to in the text. The phylogeny is based on the studies of Regier et al. [18]. Neither Bombycoidea nor Noctuidae are monophyletic, but form a clade clearly distinguishable from Pyraloidea and butterflies including Papilionoidea [18].

described above and used as queries for TBLASTN (sequences with CDS) or TBLASTX (sequences without CDS) searches against assembled genome sequences of *B. mori* [22] using the BLAST tool (<http://kaikoblast.dna.affrc.go.jp/>) associated with a database of the silkworm genome, Kaikobase.

Genes and ESTs showing significant similarity to putative single-copy *B. mori* genome sequences were selected and checked for localization of their *B. mori* orthologs in our previous studies by inheritance-based gene mapping and analysis of BAC contigs [2, 12, 23]. We performed PCR-based linkage analysis of unmapped orthologs with newly designed primers for sequence-tagged sites (STSs) (see in Supplementary Material available online at doi: 10.1155/2011/165894 Table S3) using 22–166 F₂ individuals of *B. mori* from a single pair-mating of a strain C108 female by a strain p50 male, as reported previously [23], for the confirmation of chromosomal locations obtained from Kaikobase.

When PCR products amplified from *B. mori* orthologs were monomorphic in our mapping population, PCR-based screening of a *B. mori* p50 BAC library was performed in the same manner as described elsewhere [24] to confirm whether unmapped *B. mori* orthologs were localized on previously mapped BAC contigs.

2.4. Isolation of BAC Clones. BAC libraries used in this study were an *H. virescens* library, HVB, (Average insert size 171 kb), obtained from the GENE finder Genomic Resources (Texas A&M University, College Station, TX, USA) [7], an *O. nubilalis* library, ON_Ba, (Average insert size 125 kb) and a *P. xylostella* library, PXCDBa, (Average insert

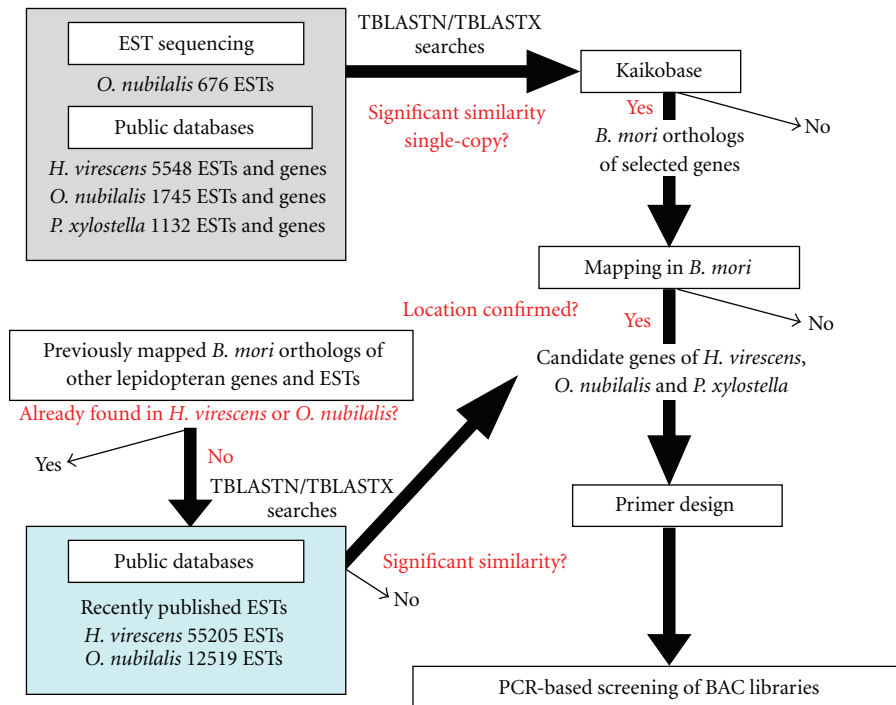


FIGURE 2: Strategy for selection of genes and ESTs in this experiment. Known genes and ESTs [13, 16, 25] were used as queries against a *B. mori* genome database, Kaikobase [20]. ESTs published later [14, 15, 17] were searched when no candidates were found in the first selection.

size 109.4 kb) [9] obtained from the Clemson University Genomics Institute (Clemson, SC, USA). Primer sets used for PCR-based screening of *H. virescens* (Table S5), *O. nubilalis* (Table S6), and *P. xylostella* (Table S7) were designed from genes and ESTs whose *B. mori* orthologs were mapped or localized on mapped BAC contigs as noted above (Table S4).

Screening was performed in three steps as described previously [24]. The first screening was performed against DNA pools derived from 48 (*H. virescens*), 96 (*O. nubilalis*), and 62 (*P. xylostella*) plates, using a mixture of 384 BAC-DNAs for each plate. When a large number of the plate pools generated positive signals, new primers were designed to improve the specificity for the targeted sequences. The second screening was carried out only in positive plates by amplifying PCR products, using as templates DNA pools for 24 columns and 16 rows, each composed of mixtures of BAC-DNAs located in the same column or rows. Finally, an overnight liquid culture of candidates identified by the preceding two steps was amplified individually to confirm the presence of target sequences.

3. Results

3.1. EST Analysis and Gene Cloning of *O. nubilalis*. When we started this experiment, there were no published *O. nubilalis* EST data. We constructed twelve cDNA libraries from various tissues and stages and determined the sequences of 625 clones randomly selected from them (Table S1). We also attempted to clone *O. nubilalis* orthologs of *B. mori* single-copy genes which we had previously mapped (Table S2) [2, 12, 23]. However, larger-scale EST analyses of *O. nubilalis*

were published later [16, 17], which provided a sufficient number of candidate genes. Thus, we decided not to continue independent EST sequencing and gene cloning.

3.2. Mapping of Conserved Genes in *Bombyx mori*. The strategy we used for the selection of genes used for BAC isolation is summarized in Figure 2. DNA sequences of known genes and ESTs of *H. virescens* [13], *O. nubilalis* [16], and *P. xylostella* [25] were used to find orthologs in *B. mori* by TBLASTN/TBLASTX search against genome sequences in Kaikobase [22]. ESTs of *H. virescens* [14, 15] and *O. nubilalis* [17] were published after the first selection of candidate genes and yielded too many for a manual similarity search using Kaikobase. Instead, TBLASTX search against ESTs in the DNA Data Bank of Japan (<http://www.ddbj.nig.ac.jp/search/top-e.html>) was performed to find *H. virescens* or *O. nubilalis* orthologs when no candidates were found in the first search or ESTs were too short to design PCR primers.

In the previous study, we found some incorrect mapping information in Kaikobase [12] and subsequently had to carry out experimental confirmation of the chromosomal location of *B. mori* orthologs to avoid being misled by false gene translocations. To minimize additional mapping efforts in *B. mori*, we gave priority to finding orthologs of *B. mori* genes which had been confirmed in the previous reports [2, 12, 23]. Putting this limitation on genes for this type of study also leads to the isolation of orthologous BACs from different species, which facilitates comparison among multiple genomes.

More than 800 candidates were identified, which showed significant similarity and seemed to be single-copy in the

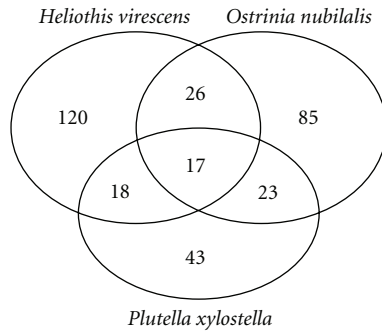


FIGURE 3: Venn diagram showing the total number of genes used for BAC isolation and those corresponding among the three species.

genome of *B. mori*. Two hundred thirty-six of them had been mapped genetically onto a linkage map of *B. mori* or localized to the mapped BAC contigs in our previous reports [2, 12, 23] (Table S4). To improve the resolution for comparing genomes, we designed 246 additional pairs of new primers for *B. mori* orthologs to select and clone BAC probes where the interval between markers was relatively long. In all, we identified 482 putative single-copy conserved genes in *B. mori* which were orthologous to known genes and ESTs of the three species (Table S4).

3.3. Isolation of BAC Clones. To isolate BAC clones of *H. virescens*, *O. nubilalis*, and *P. xylostella* using PCR-based screening, we designed primer sets to avoid including putative exon-intron junctions predicted from the alignment of cDNA sequences with genome sequences of *B. mori*. Ultimately, 181, 150, and 101 primer pairs could be used to screen *H. virescens*, *O. nubilalis*, and *P. xylostella* BAC libraries (Tables S5–7) and yielded 188, 163, and 108 BACs for 332 orthologous genes, respectively (Table S8).

We found 24 pairs and two sets of three genes for which orthologs closely spaced in *B. mori* were positive for identical clones in the three screened species (Table S8). Similar colocalization on BAC clones was also observed between *M. sexta* and *B. mori* [12], suggesting microsynteny, that is, conserved fine scale gene order exists among these species, which was recently confirmed by sequence determination of BAC clones [10, 11].

Figure 3 shows a Venn diagram with the correspondence of genes isolated for each species. The relatively small number of clones isolated from *P. xylostella* reflects fewer published ESTs. For 84 genes, clones were isolated from more than two species, which will be useful as universal markers in BAC-FISH analysis. In addition, these clones can be used as a resource for sequence comparison across multiple genomes including *B. mori* which might reveal conserved or specific regulatory elements.

4. Discussion

BAC libraries are highly useful for identifying detailed genome organization across relatively long chromosomal

distances and are now available for several species of Lepidoptera [1, 7–9]. However, few studies using lepidopteran BAC libraries have been published, especially for species other than *B. mori* [8–12]. One reason for this lack is that the technique described here to isolate clones of interest from BAC libraries is not commonly used. For sequenced model organisms like *B. mori*, BAC clones can be easily anchored to ordered genome sequences by their BAC-end sequences, which enables selection of candidate clones *in silico*. In contrast, BAC-end sequencing of nonmodel organisms is not an efficient strategy since it is not suitable for recently developed non-Sanger sequencing technologies, and it takes much cost and labor to ensure sufficient coverage to find sequences of interest. Actually, experimental identification is needed for every target sequence.

The use of PCR-based screening for rapid isolation of BAC clones was a major factor that improved efficiency in this study. *In situ* hybridization using high-density replica (HDR) filters is the most commonly used method to screen BAC libraries. However, HDR filters are usually designed with sufficient redundancy to avoid failures in screening and are too large for the isolation of a minimum number of clones used in studies like ours. PCR-based screening can be carried out using standard thermal cyclers without any special skills, and stepwise changes in the scale of screening using a pooling strategy reduce time and labor. In addition, PCR-based screening can be easily performed for gene sequences downloaded from public databases, whereas DNA probes for *in situ* hybridization either have to be obtained from the original investigators or prepared independently. Thus, we could carefully eliminate genes which were likely to be duplicated in the *B. mori* genome including putative pseudogenes from the candidates for BAC isolation.

On the other hand, preparation of DNA pools for PCR-based screening is laborious, and high efficiency is not accomplished in a small-scale analysis. Ideally, a catalogue linking BACs with located genes should be constructed and published, which will release inexperienced researchers of nonsequenced species from the technical labor of BAC isolation and let them concentrate on functional analysis. The present study is the first step to this final goal in the future.

The species used in this study, *H. virescens*, *O. nubilalis*, and *P. xylostella*, are not closely related to each other (Figure 1), but share a haploid karyotype of $n = 31$. This karyotype is considered the basal number in Lepidoptera since a survey of more than 1,000 species revealed that more than half from many independent lineages carry this chromosome number [26]. The four lepidopteran species for which their chromosome organization has been characterized in detail are either $n = 28$ (*B. mori*, *M. sexta*, *Bicyclus anynana* [27]) or $n = 21$ (*Heliconius melpomene* [28]), indicating that several chromosomal fusions occurred in their lineages. We are now analyzing the chromosomal organization of the tobacco bollworm, *Helicoverpa armigera*, a noctuid species closely related to *H. virescens* using BAC probes from *H. virescens* in parallel with the analysis of *O. nubilalis*. The identification of the karyotypes of the three species used in this experiment will reveal the

ancestral karyotype and chromosome rearrangements which have occurred in each of these representative lepidopteran lineages.

Acknowledgments

This paper was funded by the Program for Promotion of Basic Research Activities for Innovative Biosciences (PRO-BRAIN). The authors are grateful to Dr. M. R. Goldsmith for critical reading of the paper. They also thank Ms. H. Takahashi, Ms. E. Igari, Ms. H. Hoshida, and Ms. T. Maeda for their technical assistance.

References

- [1] C. Wu, S. Asakawa, N. Shimizu, S. Kawasaki, and Y. Yasukochi, "Construction and characterization of bacterial artificial chromosome libraries from the silkworm, *Bombyx mori*," *Molecular and General Genetics*, vol. 261, no. 4-5, pp. 698–706, 1999.
- [2] Y. Yasukochi, L. A. Ashakumary, C. Wu et al., "Organization of the Hox gene cluster of the silkworm, *Bombyx mori*: a split of the Hox cluster in a non-*Drosophila* insect," *Development Genes and Evolution*, vol. 214, no. 12, pp. 606–614, 2004.
- [3] Y. Yasukochi, L. A. Ashakumary, K. Baba, A. Yoshido, and K. Sahara, "A second-generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects," *Genetics*, vol. 173, no. 3, pp. 1319–1328, 2006.
- [4] A. Yoshido, H. Bando, Y. Yasukochi, and K. Sahara, "The *Bombyx mori* karyotype and the assignment of linkage groups," *Genetics*, vol. 170, no. 2, pp. 675–685, 2005.
- [5] E. D'Alençon, P. Piffanelli, A.-N. Volkoff et al., "A genomic BAC library and a new BAC-GFP vector to study the holocentric pest *Spodoptera frugiperda*," *Insect Biochemistry and Molecular Biology*, vol. 34, no. 4, pp. 331–341, 2004.
- [6] B. S. Coates, D. V. Sumerford, R. L. Hellmich, and L. C. Lewis, "Repetitive genome elements in a European corn borer, *Ostrinia nubilalis*, bacterial artificial chromosome library were indicated by bacterial artificial chromosome end sequencing and development of sequence tag site markers: implications for lepidopteran genomic research," *Genome*, vol. 52, no. 1, pp. 57–67, 2009.
- [7] C. Wu, D. Proestou, D. Carter et al., "Construction and sequence sampling of deep-coverage, large-insert BAC libraries for three model lepidopteran species," *BMC Genomics*, vol. 10, article 283, 2009.
- [8] R. Papa, C. M. Morrison, J. R. Walters et al., "Highly conserved gene order and numerous novel repetitive elements in genomic regions linked to wing pattern variation in *Heliconius* butterflies," *BMC Genomics*, vol. 9, article 345, 2008.
- [9] S. W. Baxter, M. Chen, A. Dawson et al., "Mis-spliced transcripts of nicotinic acetylcholine receptor $\alpha 6$ are associated with field evolved spinosad resistance in *Plutella xylostella*," *PLoS Genet*, vol. 6, no. 1, Article ID e1000802, 2010.
- [10] E. D'Alençon, H. Sezutsu, F. Legeai et al., "Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 17, pp. 7680–7685, 2010.
- [11] L. Ferguson, S. Failee, N. Chamberlain et al., "Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the *HmYb/Sb* locus," *Molecular Ecology*, vol. 19, supplement 1, pp. 240–254, 2010.
- [12] Y. Yasukochi, M. Tanaka-Okuyama, F. Shibata et al., "Extensive conserved synteny of genes between the karyotypes of *Manduca sexta* and *Bombyx mori* revealed by BAC-FISH mapping," *PLoS One*, vol. 4, no. 10, Article ID e7465, 2009.
- [13] K. S. Shelby and H. J. R. Popham, "Analysis of ESTs generated from immune-stimulated hemocytes of larval *Heliothis virescens*," *Journal of Invertebrate Pathology*, vol. 101, no. 2, pp. 86–95, 2009.
- [14] H. Vogel, A. J. Heidel, D. G. Heckel, and A. T. Groot, "Transcriptome analysis of the sex pheromone gland of the noctuid moth *Heliothis virescens*," *BMC Genomics*, vol. 11, no. 1, article 29, 2010.
- [15] G. Govind, O. Mittapalli, T. Griebel, S. Allmann, S. Böcker, and I. T. Baldwin, "Unbiased transcriptional comparisons of generalist and specialist herbivores feeding on progressively defenseless *Nicotiana attenuata* plants," *PLoS One*, vol. 5, no. 1, Article ID e8735, 2010.
- [16] B. S. Coates, D. V. Sumerford, R. L. Hellmich, and L. C. Lewis, "Mining an *Ostrinia nubilalis* midgut expressed sequence tag (EST) library for candidate genes and single nucleotide polymorphisms (SNPs)," *Insect Molecular Biology*, vol. 17, no. 6, pp. 607–620, 2008.
- [17] C. Khajuria, Y. C. Zhu, M.-S. Chen et al., "Expressed sequence tags from larval gut of the European corn borer (*Ostrinia nubilalis*): exploring candidate genes potentially involved in *Bacillus thuringiensis* toxicity and resistance," *BMC Genomics*, vol. 10, article 286, 2009.
- [18] J. C. Regier, A. Zwick, M. P. Cummings et al., "Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study," *BMC Evolutionary Biology*, vol. 9, no. 1, article 280, 2009.
- [19] A. M. Shelton, J. A. Wyman, N. L. Cushing et al., "Insecticide resistance of diamondback moth (Lepidoptera: Plutellidae) in North America," *Journal of Economic Entomology*, vol. 86, no. 1, pp. 11–19, 1993.
- [20] B. E. Tabashnik, N. L. Cushing, N. Finson, and M. W. Johnson, "Field development of resistance to *Bacillus thuringiensis* in diamondback moth (Lepidoptera: Plutellidae)," *Journal of Economic Entomology*, vol. 83, no. 5, pp. 1671–1676, 1990.
- [21] J. Z. Zhao, Y. X. Li, H. L. Collins et al., "Monitoring and characterization of diamondback moth (Lepidoptera: Plutellidae) resistance to spinosad," *Journal of Economic Entomology*, vol. 95, no. 2, pp. 430–436, 2002.
- [22] International Silkworm Genome Consortium, "The genome of a lepidopteran model insect, the silkworm *Bombyx mori*," *Insect Biochemistry and Molecular Biology*, vol. 38, no. 12, pp. 1036–1045, 2008.
- [23] Y. Yasukochi, "A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1,018 molecular markers," *Genetics*, vol. 150, no. 3, pp. 1513–1525, 1998.
- [24] Y. Yasukochi, "PCR-based screening for bacterial artificial chromosome libraries," *Methods in Molecular Biology*, vol. 192, pp. 401–410, 2002.
- [25] J. H. Eum, Y. R. Seo, S. M. Yoe, S. W. Kang, and S. S. Han, "Analysis of the immune-inducible genes of *Plutella xylostella* using expressed sequence tags and cDNA microarray," *Developmental and Comparative Immunology*, vol. 31, no. 11, pp. 1107–1120, 2007.
- [26] R. Robinson, *Lepidoptera Genetics*, Pergamon, Oxford, UK, 1971.

- [27] P. Beldade, S. V. Saenko, N. Pul, and A. D. Long, "A gene-based linkage map for *Bicyclus anynana* butterflies allows for a comprehensive analysis of synteny with the lepidopteran reference genome," *PLoS Genetics*, vol. 5, no. 2, Article ID e1000366, 2009.
- [28] E. G. Pringle, S. W. Baxter, C. L. Webster, A. Papanicolaou, S. F. Lee, and C. D. Jiggins, "Synteny and chromosome evolution in the lepidoptera: evidence from mapping in *Heliconius melpomene*," *Genetics*, vol. 177, no. 1, pp. 417–426, 2007.