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

1 Insect Genome Research Unit, National Institute of Agrobiological Sciences, Owashi 1-2, Tsukuba, Ibaraki 305-8634, Japan
2 Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan
3 Entomology Research Team, National Institute of Fruit Tree Science, National Agriculture and Food Research Organization, Tsukuba, Ibaraki 305-8605, Japan
4 Laboratory 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 (STSSs) were utilized for finding polymorphisms from flanking regions of the original STSSs 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, Pyrauloidea, which forms a different clade from Bombycoidea and Noctuioidea 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 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 (STs) (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 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 Noctuioidea are monophyletic, but form a clade clearly distinguishable from Pyrauloidea and butterflies including Papilionoidea [18].
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 in 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
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 clones in the three screened species (Table S8). To isolate BAC clones of which were orthologous to known genes and ESTs of *mori* where the interval between markers was relatively long. In all, comparing genomes, we designed 246 additional pairs of new *B. mori* orthologs closely spaced in *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 clones in the three screened species (Table S8). 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).

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

### 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
ancient 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. Igar, Ms. H. Hoshida, and Ms. T. Maeda for their technical assistance.

References


