1-2009

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

Brad S. Coates
United States Department of Agriculture, brad.coates@ars.usda.gov

Douglas V. Sumerford
United States Department of Agriculture

Richard L. Hellmich
Follow this and additional works at: http://lib.dr.iastate.edu/ent_pubs

Leslie C. Lewis
United States Department of Agriculture

Follow this and additional works at: http://lib.dr.iastate.edu/ent_pubs

Part of the Agronomy and Crop Sciences Commons, Entomology Commons, and the Plant Breeding and Genetics Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/ent_pubs/124. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

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

Abstract

The European corn borer, Ostrinia nubilalis, is a serious pest of food, fiber, and biofuel crops in Europe, North America, and Asia and a model system for insect olfaction and speciation. A bacterial artificial chromosome library constructed for O. nubilalis contains 36 864 clones with an estimated average insert size of >or=120 kb and genome coverage of 8.8-fold. Screening OnB1 clones comprising approximately 2.76 genome equivalents determined the physical position of 24 sequence tag site markers, including markers linked to ecologically important and Bacillus thuringiensis toxin resistance traits. OnB1 bacterial artificial chromosome end sequence reads (GenBank dbGSS accessions ET217010 to ET217273) showed homology to annotated genes or expressed sequence tags and identified repetitive genome elements, O. nubilalis miniature subterminal inverted repeat transposable elements (OnMITE01 and OnMITE02), and ezi-like long interspersed nuclear elements. Mobility of OnMITE01 was demonstrated by the presence or absence in O. nubilalis of introns at two different loci. A (GTCT)n tetranucleotide repeat at the 5' ends of OnMITE01 and OnMITE02 are evidence for transposon-mediated movement of lepidopteran microsatellite loci. The number of repetitive elements in lepidopteran genomes will affect genome assembly and marker development. Single-locus sequence tag site markers described here have downstream application for integration within linkage maps and comparative genomic studies.

Keywords

Ostrinia nubilalis, bacterial artificial chromosome library, repetitive elements

Disciplines

Agronomy and Crop Sciences | Entomology | Plant Breeding and Genetics

Comments

Rights

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

Brad S. Coates, Douglas V. Sumerford, Richard L. Hellmich, and Leslie C. Lewis

Abstract: The European corn borer, *Ostrinia nubilalis*, is a serious pest of food, fiber, and biofuel crops in Europe, North America, and Asia and a model system for insect olfaction and speciation. A bacterial artificial chromosome library constructed for *O. nubilalis* contains 36 864 clones with an estimated average insert size of ≥120 kb and genome coverage of 8.8-fold. Screening OnB1 clones comprising approximately 2.76 genome equivalents determined the physical position of 24 sequence tag site markers, including markers linked to ecologically important and *Bacillus thuringiensis* toxin resistance traits. OnB1 bacterial artificial chromosome end sequence reads (GenBank dbGSS accessions ET217010 to ET217273) showed homology to annotated genes or expressed sequence tags and identified repetitive genome elements, *O. nubilalis* miniature subterminal inverted repeat transposable elements (OnMITE01 and OnMITE02), and *ezi*-like long interspersed nuclear elements. Mobility of OnMITE01 was demonstrated by the presence or absence in *O. nubilalis* of introns at two different loci. A $(GTCT)_n$ tetranucleotide repeat at the 5′ ends of OnMITE01 and OnMITE02 are evidence for transposon-mediated movement of lepidopteran microsatellite loci. The number of repetitive elements in lepidopteran genomes will affect genome assembly and marker development. Single-locus sequence tag site markers described here have downstream application for integration within linkage maps and comparative genomic studies.

Key words: Ostrinia nubilalis, bacterial artificial chromosome library, repetitive elements.

Résumé : La pyrale du maïs, *Ostrinia nubialis*, est un ravageur important des cultures pour des fins alimentaires, textiles et énergétiques en Europe, en Amérique du Nord et en Asie. De plus, il s'agit d'une espèce modèle pour l'olfaction et la spéciation chez les insectes. Une banque de chromosomes bactériens artificiels a été produite pour l'*O. nubialis* et totalise 36 864 dont la taille moyenne des inserts est d'environ ≥120 kb pour une couverture génomique de 8,8 fois. Un criblage de clones OnB1 conférant une couverture de 2,76 équivalents génomiques a permis de déterminer la position physique de 24 marqueurs de séquence connue dont des marqueurs liés à des caractères écologiques importants ainsi qu'à la résistance à la toxine du *Bacillus thuringiensis*. Les séquences des extrémités de clones chromosomes bactériens artificiels OnB1 (accessions GenBank dbGSS ET217010 à ET217273) présentaient de l'homologie avec des gènes annotés ou des étiquettes de séquences exprimées et ont permis d'identifier des séquences répétitives : des éléments transposables miniatures à répétitions subterminales inversées (OnMITE01 et OnMITE02) ainsi que des éléments nucléaires dispersés de grande taille) de type *ezi*. La mobilité des éléments OnMITE01 a été démontrée par leur présence ou absence au sein d'introns chez deux locus de l'*O. nubialis*. Un motif tétranucléotidique répété $(GTCT)_n$ aux extrémités 5′ des éléments OnMITE01 et OnMITE02 fournit la preuve de la mobilité de locus microsatellites chez les insectes via le mouvement de transposons. Le nombre d'éléments répétitifs au sein des génomes des lépidoptères affectera l'assemblage des génomes et le développement de marqueurs. Les marqueurs de séquence connue à locus unique qui sont décrits dans ce travail se prêtent à des utilisations ultérieures en vue de l'intégration dans des cartes génétiques et pour des études comparatives.

Mots-clés : Ostrinia nubialis, banque de chromosomes bactériens artificiels, éléments répétitifs.

B.S. Coates.[1] USDA-ARS, Corn Insects and Crop Genetics Research Unit, Genetics Laboratory, Iowa State University, Ames, IA 50011, USA.
D.V. Sumerford, R.L. Hellmich, and L.C. Lewis. USDA-ARS, Corn Insects and Crop Genetics Research Unit, Genetics Laboratory, Iowa State University, Ames, IA 50011, USA; Department of Entomology, Iowa State University, Ames, IA 50011, USA.

[1]Corresponding author (e-mail: brad.coates@ars.usda.gov).

Introduction

Many lepidopteran insect species feed on plants that are the source of food, fiber, and biofuel products. Genome information of lepidopterans could lead to novel ways of controlling pest species worldwide (International Lepidopteran Genome Project 2001). Comparative genomics using whole genome sequences (WGSs) or linkage maps will identify conservation among lepidopteran genomes (synteny) or preservation of linear gene order (collinearity: Eckardt 2001; Sankoff and Nadeau 2003). *Bombyx mori* and *Heliconius melpomene* share similar gene content on a given chromosome (synteny), which suggested that the *B. mori* WGS (Mita et al. 2004; Xia et al. 2004) could be used as a reference for detecting lepidopteran genome rearrangements when shared loci are mapped (Heckel et al. 1998; Schmidt 2002; Delseny 2004; Yasukochi et al. 2006; Pringle et al. 2007). Single nucleotide polymorphism (SNP) markers developed from *B. mori* bacterial artificial chromosome (BAC) end sequence (BES) data were placed into 28 linkage groups and provided a connection between BAC physical positions and linkage information that will be needed for reconstruction of this lepidopteran genome (Yamamoto et al. 2006, 2008). Connections between linkage maps and physical constructs allow isolation of genome regions linked to mapped loci among lepidopterans that lack a WGS and can be accomplished by positioning single-locus markers on one or more BAC clones.

Repetitive elements often complicate the assembly of a WGS from shotgun sequence reads (Arabidopsis Genome Initiative 2000; International Human Genome Sequencing Consortium 2001). Highly repetitive regions comprise approximately 45% of the *B. mori* genome (Gage 1974), with a majority of units <500 bp in size (Mita et al. 2004). Proliferation of transposable elements (TEs) contributes to the increased size of eukaryotic genomes (San-Miguel and Bennetzen 1998) and affects cellular function when inserted into gene coding regions (Sommer et al. 1988; Gahan et al. 2001). TEs also have inserted within genome regions that are adjacent to lepidopteran microsatellite repeats. Subsequent oligonucleotide primers used in PCR amplification often anneal to multiple loci within a given genome and produce products comprising multiple fragments derived from more than one independently segregating locus (Van't Hof et al. 2007). Repetitive eukaryotic genomes necessitate the use of large physical constructs such as BAC library clones assembled into a physical map to assist in WGS assembly (Yamamoto et al. 2006, 2008) because short shotgun sequence reads often do not contain a sufficient level of uniqueness to denote a single genome position. BAC libraries and ordered overlapping contigs constructed from individual clones are therefore necessary tools for lepidopteran genome research.

European corn borer, *Ostrinia nubilalis*, larvae feed upon cultivated crops in North America, Europe, and Asia. Larval *O. nubilalis* feeding damage is controlled by transgenic crops expressing *Bacillus thuringiensis* toxins but have developed varying levels and modes of tolerance in laboratory environments. We present a preliminary characterization of an *O. nubilalis* BAC library, OnB1, by positioning of sequence tag site (STS) markers and end sequencing. The STS markers that we used encompassed genes linked to *Ostrinia* traits such as pheromone response, diapause duration, and *B. thuringiensis* toxin resistance traits. Furthermore, we show that OnB1 BES data are a resource for lepidopteran comparative genomics and studies of repetitive genome elements. TEs, long interspersed nuclear elements (LINEs), and miniature subterminal inverted repeat TEs (OnMITE01 and OnMITE02) were identified from OnB1 BES reads. We documented mobility and abundance of *O. nubilalis* TEs and suggest their role in the propagation of tetranucleotide repeat microsatellites in lepidopteran genomes.

Materials and methods

BAC library construction

Larval *O. nubilalis* were obtained from the USDA-ARS, Corn Insects and Crop Genetics Research Unit, in Ames, Iowa. The colony is restocked each year using >500 females collected near Ames and consists of bivoltine Z-pheromone strain ecotypes. Larvae were flash frozen in liquid nitrogen and 100 individuals (approximate 1:1 ratio of males to females) were shipped on dry ice to Amplicon Express, Pullman, Washington. There, genomic DNA was prepared, partially digested with *Bam*HI, and size selected for fragments >100 kb, which were then ligated into the *Bam*HI site of vector pEC-BAC1. *Escherichia coli* strain XL1-Blue (Stratagene, La Jolla, California) was transformed and clones selected using the chloramphenical resistance marker. Resistant clones were picked and arrayed onto ninety-six 384-well plates (36 864 total clones) and stored in 10% glycerol at –80 °C.

BAC library pooling strategy

A pooling strategy combined DNA from BAC clones into single samples for PCR screening whereby 16 rows containing 24 clones from each of thirty 384-well plates were combined to form 480 separate row pools (RPs). The 16 RPs from each 384-well plate.s additionally were pooled to form 30 superpools (SPs). Pooling represented approximately one third of the library, PCR-based screening of 1.38×10^9 bp, or approximately 2.76-fold genome coverage. Specifically, 10 µL of glycerol stock from the 24 clones located in each row was combined to create a 240 µL RP. SPs were generated by combining 15 µL of the 16 RPs in a given 384-well plate such that each 240 µL SP corresponded to one 384-well plate. The RPs and SPs were stored in 10% glycerol at –80 °C. This pooling strategy was done for thirty 384-well plates comprising 11 520 clones or approximately 1.52×10^9 bp. The RP and SP glycerol stocks were streaked onto LB agar containing 30 µg/L chloramphenical and grown overnight at 37 °C. RP and SP clones were used to inoculate 800 µL of LB broth containing 30 µg/L chloramphenical and grown overnight at 37 °C. BAC DNA was isolated by alkaline lysis with RNase treatment omitted. BAC DNA was suspended in 200 µL of 0.5× TE buffer, pH. 7.5, and stored at –20 °C prior to use.

Table 1. Eight contigs constructed from 17 *Ostrinia nubilalis* BAC end sequence (BES) reads.

Contig name	OnB1 BES	Read length (bp)	Contig overlap (bp)	SNPs
On-C0001	23A08-F, 23P08-F	365	217	None
On-C0002	23G02-F, 24G01-F	342	285	None
On-C0003	04B19-F, 04F23-F	435	331	186 A/T
				305 A/T
On-C0004	04A19-F, 04P18-F	348	286	174 A/G
				178 T/G
On-C0005	23L08-F, 23L09-F, 23E08-F	570	396	None
On-C0006	23B08-F, 23O08-F	400	319	241 A/C
				362 A/C
On-C0007	23L07-F	410	339	316 C/G
	23L09-F			
On-C0008	04D23-F	469	424	None
	04A23-F			

Note: The read length for each clone and length of overlap of assembled contigs are given in basepair (bp). Putative single nucleotide polymorphism (SNP) loci are indicated by location in the contig followed by alternate nucleotides present. The criteria for contig assembly are ≥95% homology over ≥200 nucleotides and alignment gap penalty = 10.

Marker development and OnB1 PCR screening

PCR screening of SPs and RPs was used as a method to position genes of interest (STSs) on OnB1 clones and to identify homologous genome regions (since PCR markers were single locus). Single-locus STS markers were shown by pedigree analysis for *O. nubilalis* cadherin (Coates et al. 2005*a*), ommochrome binding proteins (obp1 and obp2) (Coates et al. 2005*b*), brainiac/bre5 (Coates et al. 2007), and aminopeptidase 1 (APN1) (Coates et al. 2008). Additional single locus markers were from *OnZ*1 and *OnW*1 (Coates and Hellmich 2003) and kettin (*ket*), triosphosphate isomerase (*tpi*), and lactose dehydrogenase (*ldh*) (Dopman et al. 2005; Malausa et al. 2007). We designed a total of 19 degenerate oligonucleotide primers for PCR amplification of ribosomal loci from Lepidoptera of which 11 were successful in PCRs with *O. nubilalis* genomic DNA as template. Oligonucleotide primers were also designed from OnB1 BES data. Four markers were designed to amplify regions of OnB1 contigs On-C0003, On-0004, On-0006, and On-0007 that contained putative SNP loci (Table 1).

For each STS marker, PCR amplification of SP and appropriate RP BAC template took place in 10 μL reaction volumes containing 2.5 mmol/L MgCl$_2$, 50 μmol/L dNTPs, 5 ng of BAC DNA, 1.8 pmol of each primer, 2 μL of 5× thermal polymerase buffer (Promega, Madison, Wisconsin), and 0.3125U of Go*Taq* DNA polymerase (Promega). A Bio-Rad Tetrad 2 thermocycler used 96 °C for 3 min and then seven initial touchdown (TD) cycles consisting of 96 °C for 20 s, 65 °C for 30 s (–2 °C per cycle for each subsequent cycle), and 72 °C for 1 min. Final PCR amplification was 35 cycles of 96 °C for 20 s, 50 °C for 30 s, and 72 °C for 1 min. PCR products for *O. nubilalis OnZ*1 and *OnW*1, cadherin, *obp*1, *obp*2, and *bre*5 took place according to Coates and Hellmich (2003) and Coates et al. (2005*a*, 2005*b*), 2007), respectively. Entire PCR product volumes were separated on 1.5% agarose gels and positive BAC SPs or RPs was identified by presence–absence of the gene fragment compared with positive control DNA. Once RPs were iden-

tified, the 24 clones from the respective RPs were streaked onto LB agar, cultured overnight, and DNA isolated by alkaline lysis as described previously.

BAC DNA isolation and sequencing

OnB1 clones were streaked from glycerol stocks onto LB agar containing 30 μg/L chloramphenical and grown overnight at 37 °C. Individual clones were used to start seed cultures of 800 μL of LB broth containing 30 μg/L chloramphenical and grown overnight at 37 °C. Larger volume 30 mL LB chloramphenical (30 μg/L) cultures were inoculated with individual seed cultures and grown overnight at 37 °C with shaking at 200 rpm. Cells were harvested by centrifugation at 5000 rpm for 1 min. BAC DNA was purified using the BAC and Low Copy miniprep kits (GerardBIOTECH, Oxford, Ohio) and final pellets suspended in 43 μL of nuclease-free water, 5 μL of 10× buffer D, and 1 μL of *Xba*I and 1 μL *Sal*I restriction enzymes (10 units/μL) (Promega). Digests were incubated overnight at 37 °C. DNA was precipitated in a 5% polyethylene glycol and 0.5 mol/L NaCl solution at 14 000 rpm for 20 min at 4 °C in a centrifuge. Pelleted DNA was washed twice with 0.5 mL of cold 70% ethanol and then dried and suspended in 12 μL of deionized water. Samples were quantified by UV absorbance at 260 nm and diluted to a final concentration of 0.1 μg/μL.

Template was denatured by incubation at 96 °C for 5 min followed by snap chilling on ice for 5 min. End sequencing of individual BACs was performed with 10 μL (1.0 μg) of *Xba*I- and *Sal*I-digested BAC DNA template, 2 μL of T7-BAC primer (5′-ATACGACTCACTATAGGGCG-3′, 30 pmol/μL), and 8 μL of Dye-Terminator Cycle Sequence kit mastermix (Beckman-Coulter, Fullerton, California). Primer extension reactions were carried out on an Eppendorf Mastercycler (Hamburg, Germany) with an initial denaturation at 96 °C for 2 min followed by 80 cycles of 96 °C for 30 s, 54 °C for 30 s, and 60 °C for 4 min. Primer extension products were purified by ethanol precipitation, suspended in 25 μL of deionized formamide, and separated on a CEQ 8000 genetic analysis system (Beckman-Coulter) with

method LFR-a (denature: 90 °C for 120 s, inject: 2.0 kV for 15 s, and separated: 4.2 kV for 120 min in a 50 °C capillary).

Partial DNA sequence analysis of OnB1 BACs that were identified by PCR screening was performed by the use of the separate primer extension reactions with forward and reverse primers (supplementary material S1[3]). Individual BACs were grown in LB chloramphenical (30 μg/L) cultures and DNA isolated as described above. DNA sequencing was performed with 10 μL (1.0 μg) of XbaI- and SalI-digested BAC DNA template, 2 μL (30 pmol/μL) of the corresponding forward or reverse primer, and 8 μL of Dye-Terminator Cycle Sequence kit mastermix (Beckman-Coulter). Cycle sequencing primer extension reactions and separation on a CEQ8000 system were carried out as previously described.

DNA sequence quality trimming and annotation

Raw sequence data were analyzed and quality of sequence assessed using the PHRED quality parameter. Probability of correct base calls during sequence analysis (PHRED quality parameter $q = -10 \times \log_{10}(p)$ where p is the estimated error probability of a base call) was calculated with the Sequence Analysis software (version 8.0) of the Beckman-Coulter CEQ8000 genetic analysis system. Sequences were trimmed when $q < 20$ (90% base call accuracy) and internal discrepancies were inspected and corrected visually. Vector sequence was automatically trimmed from FASTA-formatted data when exported by the CEQ8000 Sequence Analysis software.

ContigExpress software (Informax, San Francisco, California) was used to assemble contiguous and overlapping sequences among OnB1 BESs. The OnB1 BES contigs and singletons were exported in FASTA format and used as queries molecular database searches. Homology searches were performed for each OnB1 BES against GenBank nr/nt accessions using BLASTn and tBLASTx (current December 2007). Similarity of OnB1 BES reads to expressed sequence tag (EST) and genome survey sequence (GSS) accessions was investigated using BLASTn searches of GenBank dbEST and dbGSS databases. Eight OnB1 contigs and 246 singletons were annotated by using BLASTn of GenBank nucleotide (nr/nt) and EST databases (dbEST). Criteria for positive matches were ≥80% homology over ≥130 nucleotides with E values ≥e^{-10}.

Phylogenetics

GenBank accessions that showed homology to OnB1 BES reads 04K12-F and 27L07-F and pheromone binding protein (pbp) alleles were downloaded in FASTA format (Fig. 1) (supplementary material S3[2]) and imported into the Molecular Evolutionary Genetics Analysis program (version 3.1) (Kumar et al. 2004). ClustalW alignments were performed using default parameters and corrections made manually. Subsequent maximum likelihood phylogenies were constructed using the genetic distances calculated. Bootstrap branch support for each node was generated using 1000 pseudoreplications. A linearized consensus phylogeny produced with nodes contained within ≥400 of 1000 (40%) random trees that were generated.

## Results and discussion

BAC library construction

An *O. nubilalis* BAC library, OnB1, was constructed from an approximately equal proportion of Z-pheromone strain males and females that showed two mating generations per year in field conditions in Ames, Iowa (bivoltine ecotype). The haploid (1N) genome size of *O. nubilalis* is unknown, but an assumption of 500 Mb was made based on the size of the *Galleria melonella* (Rasch 1985) and *B. mori* genomes (Gage 1974). OnB1 consists of genomic DNA inserted into the BamHI site of pECBAC1 with an average size of 120 kb (Amplicon Express communication). A total of 36 864 OnB1clones were arrayed on ninety-six 384-well plates. Sequencing data presented later indicated that 1.85% of clones carry no genomic insert, which suggested that OnB1 contains approximately 4432 Mb, or approximately 8.8-fold, 1N genome coverage. Assuming random probability of insertion for each genomic fragment into the pECBAC1 vector, the 95% and 99% probabilities of recovering a genomic fragment (locus) suggest a minimum of 2.95- and 1.92-fold genome coverage, respectively.

Screening STS markers

PCR screening of BAC library pools is an effective method for isolation of single loci from represented genomes (Bruno et al. 1995; Yim et al. 2007) and was used to test OnB1 representation of the *O. nubilalis* genome. PCR-based screening of 1.38×10^9 bp, or approximately 2.76-fold genome coverage, isolated one or more BAC clones for 24 of 25 *O. nubilalis* STS markers (mean = 1.64 ± 0.70). Some of the genes involved in ecologically important traits were also located on OnB1 clones. The quantitative trait locus (QTL) for male pheromone response (*Resp*) is in proximity (28.1 ± 4.1 cM) to a *tpi* gene marker (Glover et al. 1990; Dopman et al. 2004), and postdiapause development (*Ppd*) time of larvae is determined by male parents (Showers et al. 1972; Glover et al. 1992). BACs carrying Z-chromosome-linked *tpi*, *ket*, and *ldh* genes and the minisatellites marker *OnZ*1 (Coates and Hellmich 2003) were identified (supplementary material S1[2]). Linkage relationships between the markers *tpi*, *ket* and *ldh* and QTL *Resp* and *Ppd* have been determined previously (Dopman et al. 2005). These OnB1 clones carrying Z-chromosome regions likely will assist in the development of additional single-locus Z-chromosome-linked markers and construction of BAC contigs in proximity to QTL *Resp* and *Ppd*. A total of 29 BAC clones were identified that contained 24 of the 25 STS markers (96.3%) that were screened for by PCR. The screen used an approximate one-third coverage of the OnB1 library and was effective in recovery of one to three clones for each STS marker.

Fig. 1. Maximum likelihood phylogeny showing the relationship among miniature subterminal inverted repeat transposable elements (MITEs) from Lepidoptera that show homology to OnMITE01 and OnMITE02 from *Ostrinia nubilalis*. Branch support is indicated by the number of bootstrap pseudoreplications producing each node (n = 1000). Accessions from species included in the tree are from *Bombyx mori*, *Coenonympha hero*, *Drupadia theda*, *Manduca sexta*, *Plutella xylostella*, and *Utetheisa ornatrix*.
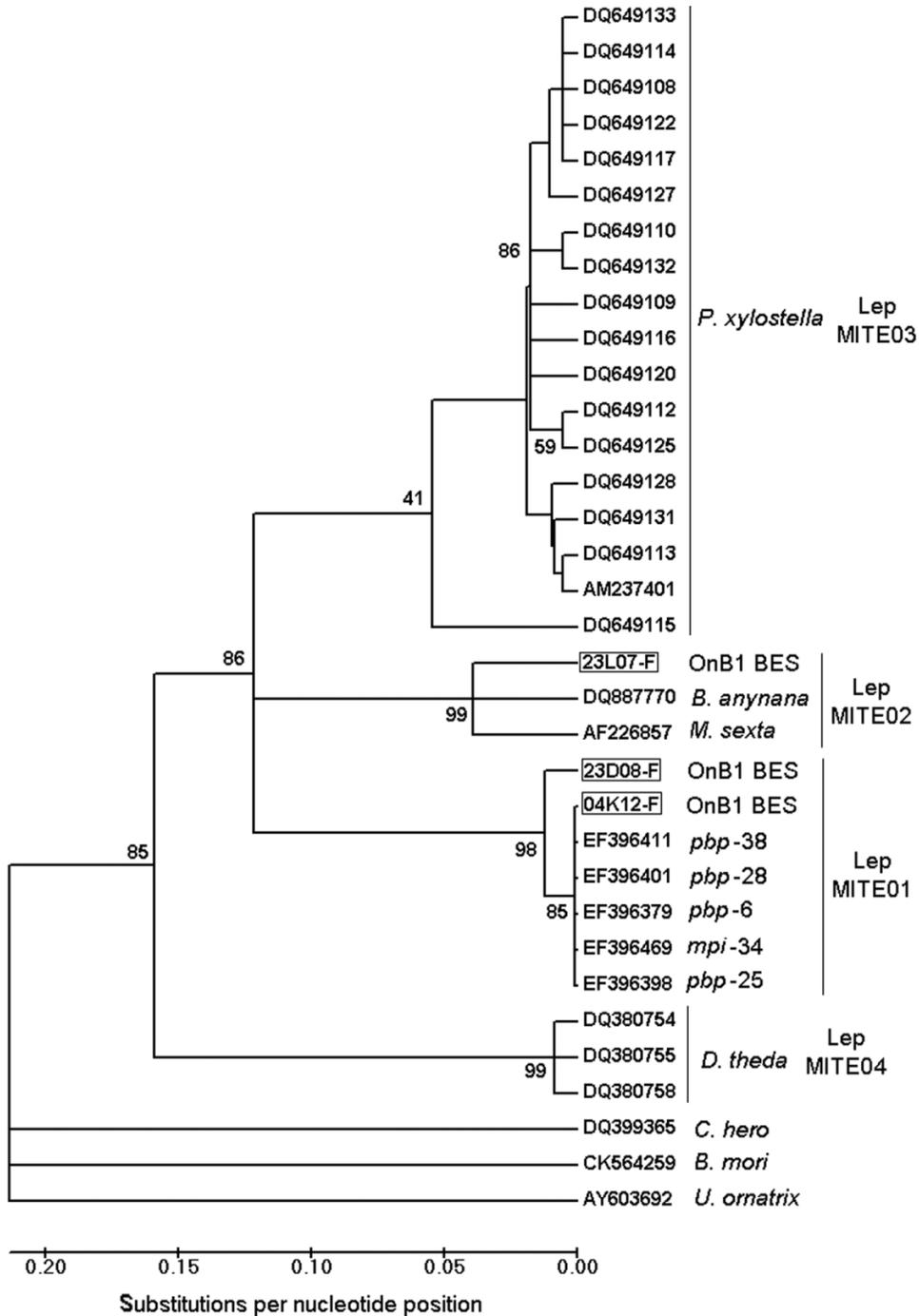
Annotation, contig assembly, and redundancy of BESs

A total of 264 GSSs were obtained from 294 randomly selected OnB1 clones using the T7 sequencing primer (90% success rate). Vector trimming identified five sequences that were vector only (1.87%). The remaining 264 sequences gave a total of 101 486 nucleotide calls for an average read length of 381.5 ± 119.9 nt with PHRED scores of >20 (Ewing et al. 1998). OnB1 sequences are accessible through the GenBank GSS database (dbGSS) (accessions ET217010 to ET217273). BLASTn search results indicated that 10 OnB1 BES reads matched lepidopteran genes or ESTs (supplementary material S2[2]). Eight overlapping or contiguous regions (contigs) were constructed from OnB1 BES reads with two or three sequences per contig (mean = 2.13 ±

Table 2. Homology of *Ostrinia nubilalis* BAC end sequence (BES) reads to previously described *Ostrinia* genes in the GenBank nr/nt database.

Queried sequence		Homologous sequence				BLASTn statistics	
OnB1 BES	Positions	Gene	Accession	Positions	Gene location	Match	E
04K12-F	25–274	*pbp*-38	EF396411	682–943	Intron 2	249/262	$1e^{-114}$
	25–274	*pbp*-28	EF396401	683–944	Intron 2	249/262	$1e^{-114}$
	25–274	*pbp*-6	EF396379	682–943	Intron 2	249/262	$1e^{-114}$
	25–274	*pbp*-25	EF396398	682–945	Intron 2	249/264	$7e^{-112}$
	28–286	*mpi*-22	EF396469	532–254	Intron	255/279	$9e^{-111}$
23L07-F	1–329	*chit*1	DQ88770	492–820	Intron 2	320/329	$2e^{-154}$
23N03-F	43–264	*ezi*-Δ11α	EF113398	598–819	5′-UTR	218/222	$7e^{-105}$
	43–264	*ezi*-Δ11α	EF113399	2197–1977	5′-UTR	210/222	$6e^{-91}$
	43–264	*ezi*-Δ11α	EF113400	1592–1372	5′-UTR	210/222	$6e^{-91}$
23M11-F	24–253	*ezi*-Δ11α	EF113398	1870–2097	CDS	207/232	$1e^{-73}$
	24–253	*ezi*-Δ11α	EF113399	2794–3021	CDS	207/232	$1e^{-73}$
	24–253	*ezi*-Δ11α	EF113400	2189–2416	CDS	207/232	$1e^{-73}$
	24–253	*ezi*-Δ11α	EF113401	1327–1554	CDS	207/232	$1e^{-73}$
	24–253	*ezi*-Δ11α	EF113402	1143–1370	CDS	207/232	$1e^{-73}$

Note: The criteria for match are ≥90% homology over ≥200 nucleotides with E values ≥e^{-20}. Homologous genes are *Ostrinia* pheromone-binding protein (*pbp*), mannose phosphate 6-isomerase (*mpi*), chitinase (*chit*1), and the retrotransposon-like long interspersed nuclear element (LINE) *ezi*- located upstream of the Δ11α-desaturase gene.

0.35) (Table 1). The remaining 249 OnB1 BESs were unique nonredundant singletons. Nucleotide mismatch among four contigs identified seven SNPs.

A high level of *O. nubilalis* genome redundancy was indicated by assembly of eight OnB1 contigs from 264 BES reads. Similar regions of the contig assemblies showed seven SNPs (Table 1) and also extended regions of dissimilarity, which suggested that members originated from different genome locations. The overlapping regions of eight contigs constructed from OnB1 BES reads ranged in size from 217 to 424 bp (Table 2). The full lengths of homologous sequences are not known, since the regions of similarity extended outside the BES reads and additional sequencing is required to determine the sizes of these putative repetitive genome regions. Genome redundancy also was observed by OnB1 screening of PCR-based markers designed to amplify contig regions with polymorphic SNPs. Each marker was PCR amplified from ≥4 of 384 randomly selected clones (shared sequence encountered every 11.52 Mb) and indicated that OnB1 contigs may represent repetitive elements (data not shown). Approximately 45% of the *B. mori* genome is composed of repetitive DNA (Gage 1974) with an average size of <500 bp and present at 2.5 to 3 kb intervals (Mita et al. 2004). Yamamoto et al. (2006, 2008) used data from 115 968 BES reads to design 3840 nonredundant single-locus *B. mori* markers (successful marker development from 3.3% of BESs). Yamamoto et al. (2006, 2008) and OnB1 data suggested that BES data may not be sources for efficient single-locus PCR marker development among lepidopterans owing to the prevalence of repetitive DNA elements. Additionally, repeated and duplicated genome regions within the libraries likely may complicate the assembly of BAC contigs (Lin et al. 2000).

Annotation of repetitive genome elements

BLASTn searches of the GenBank nr/nt database indicated that OnB1 sequences 04K12-F, 23L07-F, 23N03-F, and 23M11-F had significant similarity to previously de-scribed *Ostrinia* genes (Table 2). The 23N03-F (ET217126) and 23M11-F (ET217161) sequence reads both shared similarity to an *Ostrinia* genome region encoding a transposase. The 04K12-F (ET217030) sequence had homology to introns of *pbp* and mannose 6-phosphate isomerase (*mpi*) alleles, as did the 23L07-F sequence (ET217030) to intron 2 of the *Ostrinia furnacalis* chitinase gene. Further investigations, presented below, indicated that the four sequences belonged to two types of repetitive genomic elements: LINEs related to the *Ostrinia ezi*-element (Xue et al. 2007) and mobile MITE-like sequences.

Retrotransposon-like LINEs

The OnB1 BESs 23N03-F and 23M11-F showed homology to an *O. nubilalis* reverse transcriptase (RT) and Δ11α-desaturase gene fusion locus (Xue et al. 2007). The 23N03-F sequence (264 bp) has 94.6% homology to a 222 bp region of the first 21 amino acids of the RT and 157 bp of the upstream 5′-UTR. Similarly, a 230 bp region of the 539 bp 23M11-F OnB1 sequence has approximately 90% homology to the C-terminal end of the RT (amino acids 222–363 of the translated peptide AB045235). RT genes typically are encoded by retrotransposons. The *B. mori* genome is highly repetitive with several active and degraded retrotransposons (Mita et al. 2004; Xia et al. 2004) and might suggest that sequence reads 23N02-F and 23M11-F represent portions of a transposon-like sequence. Xue et al. (2007) indicated BLAST evidence relating the RT upstream of the *Ostrinia* Δ11α-desaturase genes to LINEs from *Caenorhabditis elegans* and renamed the *Ostrinia* LINE *ezi*-. LINEs are autonomous non-long terminal repeat retrotransposons that typically are >5 kb and encode an RT and an endonuclease required for genome propagation by RNA intermediates (class I TEs). The copy number of the *B. mori* LINE element *kaikoga* was shown to be ≥200 copies based on WGS searches (Xue et al. 2007), and the *ezi*-Δ11α LINE has proliferated in the *Ostrinia* genome based on desaturase and OnB1 BES data. Screening of 360 OnB1 clones with pri-

Fig. 2. Alignment of *Ostrinia* miniature subterminal inverted repeat transposable elements (MSITEs). (A) OnMITE01-like sequences from OnB1 BES 04K12-F (ET217030) and 23D08 (ET217149) and introns of *O. nubilalis* and pheromone-binding protein (*pbp*-25, EF396398) and mannose phosphate isomerase genes (*mpi*-22, EF396469). (B) Alignment of OnMITE02-like sequences from OnB1 BES 23L07-F (ET217130) *O. furnacalis* chitinase *chit*1, Of DQ887770. A polymorphic (GTCT)$_n$ microsatellite is highlighted and stem (S) and subterminal inverted repeat (SIR) structures are labeled accordingly with arrows indicating direction.

(A) OnMITE01

```
OnB1_04K12-F ET217030   (1) CATACAGGTATTAAGATGACACGATAAATCCCAACTAATATTATAAATGCGAAAGTAAC
OnB1_20D08-F ET217149       ***********TAATCCCAAC---TAATCCCAACTAATATTATAAATGCGAAAGTAAC
On_pbp-25    EF396398  (661) TG-TCAGGCCTCCA-TCCCAAC-TTATATCCCAACTAATATTATAAATGCGAAAGTAAC
On_mpi-22    EF396411   (65) AGCTCTGGTAGTTC-TGC-AA-----CATCCCAACTAATATTATAAATGCGAAAGTAAC

OnB1_04K12-F ET217030   (61) TCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT--------------------GTTACG-CT
OnB1_20D08-F ET217149        TCTGTCTGTCTGTCTGTCT-----------------------------GTTACGACT
On_pbp-25    EF396398  (718) TCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT--------GTTACG-CT
On_mpi-22    EF396411  (116) TCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTTACG-CT
                                                                                      ___

OnB1_04K12-F ET217030   (99) TTCCC-G-CTT-AAACCTCG-C-AACCGATTTT-G-ATGAAATTT-GGCAT-AGAGATA-
OnB1_20D08-F ET217149        TTCCCCGACTTAAAACCTAGACAAACCGAATTTTGAATGAAATTTTGGGATAAGAGATAA
On_pbp-25    EF396398  (770) TTCCC-G-CTT-AAACCTCG-C-AACCGATTTT-G-ATGAAATTT-GGCAT-AGAGATA-
On_mpi-22    EF396411   176) TTCCC-G-CTT-AAACCTCG-C-AACCGATTTT-G-ATGAAATTT-GGCAT-AGAGATA-
                SIR ──→      ── S1a────→         ──S1b─→    ←─S1b──      ←──S1a──
OnB1_04K12-F ET217030  (159) GTTTGAGTCC-CGGGAAAGAACATAGGATAGTTTTTATCCCGGTTTTTGAAACAGGG--
OnB1_20D08-F ET217149        GTTTGAGTCCCGGGAAAGAAACATAGGATAGTTTTTATCCCGGTTTTTGAAAACAGGGA
On_pbp-25    EF396398  (830) GTTTGAGTCC-CGGGAAAGAACATAGGATAGTTTTTATCCCGGTTTTTGAAACAGGG--
On_mpi-22    EF396411  (236) GTTTGAGTCC-CGGGAAAGAACATAGGATAGTTTTTATCCCGGTTTTTGAAACAGGG--
                ─────────                                   ──S2──→   ←── S2──
OnB1_04K12-F ET217030  (219) ACGCGCGCGATA-AAGTTTTTCTGTGACAGACAAAATTCCACGCGGGCGAAGCCGCGGGC
OnB1_20D08-F ET217149        ACGACGACGGTATAAGTTTTT*************************************
On_pbp-25    EF396398  (890) ACGCGCGCGATA-AAGTTTTTCTGTGACAGACAAAATTCCACGCGGGCGAAGCCGCGGGC
On_mpi-22    EF396411  (296) ACGCGCGCGATA-AAGTTTTTCTGTGACAGACAAAATTCCACGCGGGCGAAGCCGCGGGC
                ─S3a─→        ──── S3b──→       ←── S3b──   ←─ S3a──
OnB1_04K12-F ET217030  (279) GGAAAGCTAGTAGATAATAAAAGATAAGATC--TGTGAAGGCCAACAT—GTTTTTTGAGG
OnB1_20D08-F ET217149        ****************************************************
On_pbp-25    EF396398  (950) GGAAAGCTAGTTTTAAAATAAATCTTTCTACATTTTGAACTGTCTCATAGCTACATTCTA
On_mpi-22    EF396411  (356) GGAAAGCTAGTAAGAGATAAAAGTTGTTTA---TTTCCTTTCAATCATAGGGTGTTGGTT
                SIR ←───
```

(B) OnMITE02

```
OnB1_23L07-F ET217130    (1) ********************TATAATAAATCTGTAGAGAGGTCAATTCTGTACATGAA
Of_chit1 DQ887770      (470) GTTTGGTCCAAAAATCTATACTTATAATAAACCTGTAGAGAGGTCAATTCTGTACATGAA

OnB1_23L07-F ET217130   (39) ATATATTTTCAAAATTACTATCAGGGGGTGATTAGTGATCGATACTGATGCCAAAAATGC
Of_chit1 DQ887770      (530) ATATATTTTCAAAATAACTATCAGGGGGTGATTAGTGATCGATACTGATGCCAAAAATGC
                                                                              ─ SIR→
OnB1_23L07-F ET217130   (99) AATCAGTAAAATTTTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTATGTTCCTTATAGAAACA
Of_chit1 DQ887770      (590) AATCAGTAAAATTTTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTATGTTCCTTATAGAAACA

OnB1_23L07-F ET217130   159) AAAACTACTCGACGGATTTTAACGAAACTTGGTACAATTATTCTTCATACTCCTGGGCAG
Of_chit1 DQ887770      (650) AAAACTACTCGACGGATTTTAACGAAACTTGGTGCAATTATTCTTCATACTCCTGGGCAG
                                              ──S1ab     ←─ S1a──        ───
OnB1_23L07-F ET217130  (219) GTTATAGGATACTTAGGAATTCCC---ACGGGAACGGGCATTAGC-----TGGAAAATCC
Of_chit1 DQ887770      (710) GTTATAGTATACTTAGGAATTCCC---ACGGGAACGGGAATTAGC-----GGGGAAATCC
                ─S2a─→      ──S2b──→                           ←─ S2b──
OnB1_23L07-F ET217130  (271) TTTTGTAT--GAAAAATCTAAACCGCTTAAGTTAGACGCTTGAAATTTGGCAAGCAGGTA
Of_chit1 DQ887770      (762) TTTTGTAT--GAAAAATCTAAACCGCTTAAGTTAAACGCTTGAAATTTGGCATGCAGGTA
                ←─S2a──                                        ←─ SIR──
```

mers designed to PCR amplify the *ezi*-Δ11α elements identified 113 putative *O. nubilalis* LINE-like elements in an approximately 0.184-fold genome equivalent sampling (supplementary material S1[2]). The library screening likely gave an underrepresentation of the *ezi* copy number owing to sampling error but still suggested that approximately ≥600 related loci may be present in the 1*N O. nubilalis* genome.

MITEs

MITEs are short high copy number mobile DNA regions

characterized by high A+T content, pronounced secondary structure, lack of an internal protein coding sequence, and the presence of terminal inverted repeats (TIRs) (Wessler et al. 1995). MITEs are found ubiquitously within genomes and have been characterized from fruit fly (Wilder and Hollocher 2001) and mosquito (Tu 2000) and within introns of the cytochrome p450 monooxygenase genes from *Helicoverpa zea* (Chen and Li 2007). MITEs may represent nonautonomous TEs depleted of an internal coding sequence. Cut-and-paste movements of MITEs occur by DNA intermediates (class II TEs) and is mediated by *trans*-acting transposases that are encoded by related autonomous TEs (Dufresne et al. 2007). Quesneville et al. (2006) identified 304 autonomous *P*-element-like TEs within the *Anopheles gambiae* WGS that may facilitate transposition of 2570 related *P*-MITEs, which indicated that MITEs have proliferated to a greater degree within genomes compared with related autonomous elements.

Two MITE-like sequences, OnMITE01 and OnMITE02, were predicted from OnB1 BES data. OnMITE01 was found within two BES reads, 04K12-F and 23D08-F, where only the 04K12-F sequence provided a full 236 bp copy. The 328 bp OnMITE02 is present in the BES 23L07-F. Characteristic of other MITEs, OnMITE01 and OnMITE02 have high A+T contents (55.5% and 63.8%, respectively) and were shown to PCR amplify from 65% and 56% of BACs, respectively. OnMITE secondary structure composed of two or three adjacent stem loop structures was indicated by Mfold output (supplementary material S1, S2, and S3[2]) (Fig. 2) (Zuker et al. 1999). Lastly, no internal OnMITE1 or OnMITE02 protein coding sequences >50 amino acids were predicted using BLASTx or tBLASTx searches of GenBank. In contrast with most other MITEs, OnMITE01 and OnMITE02 lack characteristic TIRs (Fig. 2). Omission of TIRs from MITE-like sequences has been previously shown in the *Aedes aegypti microuli* element (Tu and Orphanidis 2001) where transposition was hypothesized to occur from subterminal inverted repeats (SIRs). SIRs may be derived from TIRs that have undergone recombination or incurred mutations or represent intermediates during genesis of novel MITE-like TIRs (Tu and Orphanidis 2001). OnMITE01 and OnMITE02 appeared to lack TIRs wherein OnMITE01 and OnMITE02, respectively, have the SIR sequences CTTTCC/GGAAAG and TGCCAAA/TTTGGCA (Fig. 2).

Sequences outside the MITE-like regions of 04K12 and 20D08 showed no similarity and suggested insertion of OnMITE01 into more than one genome region. Mobility of nonautonomous MITEs in the *Ostrinia* genome was inferred by location at multiple loci and presence–absence at a single locus. Searches of the GenBank nucleotide collection (nt/nr) indicated homology between OnMITE01 and *O. nubilalis pbp* (alleles -6, -25, -28, and -38 from accessions EF396379, EF396398, EF396401, and EF396411, respectively; ID ≥ 94%, *E* value ≤ 7 × 10^{-112}), and *mpi* (allele –22 from accession EF396469; ID = 97%; *E* value ≤ 2 × 10^{-120}).

OnMITEs insertions into *Ostrinia pbp*, *mpi*, and chitinase genes were within introns, whereas OnMITEs from OnB1 BES data may be located in nongenic regions. Quesneville et al. (2006) found that 14.9% of all *A. gambiae P* MITEs were within introns compared with 0.04% within exons.

Prevalence of MITE insertion within introns, as compared with peptide coding regions, may have resulted from preference of *trans*-acting endonucleases for genome regions with high AT content (Tu 2001). Cut-and-paste movement of class II TEs results in "jumping" of respective DNA segments in or out of a locus, with the only remnant of insertion being an inverted terminal repeat created by staggered endonuclease cleavage and DNA polymerase fill-in. Target site cleave and insertion takes place at AT/TA dinucleotides for OnMITE01 and OnMITE02 (Fig. 2). The AT content of *pbp* and *mpi* introns are 68.8% and 70.3%, respectively, compared with 58.9% and 50.7% of the flanking exons. Similarly, the *O. furnacalis chit*1 intron 1 has 66.3% and flanking exons have 53.4% AT content (OnMITE sequence omitted in all comparisons). Similar to mosquito MITEs (Tu 2001), OnMITEs may insert within introns at high frequency owing to higher prevalence of AT/TA dinucleotides, but the target site specificity of the *trans*-acting endonuclease will be required to confirm target site preference.

Approximately 83% of *P* MITEs are located outside gene coding regions in the *A. gambiae* WGS. Insertions into exon regions likely are highly selected against owing to disruption of encoded proteins, with only 0.4% found disrupting exons (Quesneville et al. 2006). Insertion into transcribed regions does take place and was indicated from *B. mori* and *Bicyclus anynana* ESTs (supplementary material S2[2]). The presence of OnMITEs within introns of *Ostrinia pbp*, *mpi*, and chitinase genes may not be selectively neutral, considering that Quesneville et al. (2006) showed that mosquito MITE prevalence within introns is significantly lower than expected by random genomic distribution. OnMITE01 was present within intron 2 in 4 of 83 *O. nubilalis pbp* alleles (0.046%, *pbp*-6, -25, -28, and -38: Willett and Harrison 1999; Dopman et al. 2005; Malausa et al. 2007). Similarly, 1 of 31 *O. nubilalis mpi* alleles (*mpi*-22) had the OnMITE01 insertion (0.032%; Malausa et al. 2007). A *pbp* phylogeny constructed from intron regions flanking the insertion indicated OnMITE01 jumped into a single allele lineage (supplementary material S3[2]). One member of the same lineage has not undergone OnMITE01 insertion (*pbp*-34, EF396407). Since flanking intron and exon regions of *pbp*-34 have not diverged from *pbp*-6, -25, -28, and -38 alleles that contained OnMITE01, recent insertion or excision of OnMITE01 from a single *pbp* allele lineage was inferred.

Lepidopteran (GTCT)$_n$ microsatellites hitchhiking within MITEs

Genome regions flanking microsatellite loci within and between lepidopteran genomes often share sequence similarities and were shown to be paralogs that belonged to multilocus microsatellite families (Anderson et al. 2007; Meglécz et al. 2007; Van't Hof et al. 2007). The origin of microsatellite flanking sequence similarity remains unknown (Meglécz et al. 2007) but has been hypothesized to result from association of minisatellites and mobile DNA elements (Zhang 2004; Van't Hof et al. 2007). Microsatellite family members are associated with interspersed repetitive elements of *Drosophila melanogaster* and *A. gambiae* (Meglécz et al. 2007) as well as different TE families (Hoekstra et al. 1997; Fagerberg et al. 2001). Omitted from discussion to this point was

the presence of $(GTCT)_n$ tetranucleotide microsatellite repeat arrays near the 5′ end of OnMITE01 and OnMITE02 (Fig. 2). $(GTCT)_n$ microsatellites were also found within OnMITE-like regions in several lepidopteran species including in the 5′-UTR of the *M. sexta* juvenile hormone binding protein precursor (AF226857) and transcribed genome regions in *B. anyanana* wing (DY763404 and DY664579) and *B. mori* embryonic EST accessions (CK564259) (supplementary material S2[2]). Genome regions flanking lepidopteran $(GTCT)_n$ microsatellite loci from *Drupadia theda*, *Plutella xylostella*, and *Utetheisa ornatrix* also showed similarity to OnMITE01 (supplementary material S2[2]). A single variant microsatellite repeat array, $(GT)_9$, was observed from a *Coenonympha hero* sequence in the position homologous to $(GTCT)_n$ repeats of all other MITEs.

Microsatellite loci from *D. theda* and *P. xylostella* provided the largest intraspecific microsatellite data sets (Fig. 1). We previously showed two lines of evidence that suggested that OnMITE01 and OnMITE02 are mobile; the two elements were multilocus as indicated from divergent sequences 5′ and 3′ of AT or TA target site duplications, and active transposition was inferred by the presences–absence within introns of *pbp* and *mpi* alleles. Similar to On-MITEs, the *D. theda* and *P. xylostella* $(GTCT)_n$ microsatellite flanking regions showed little or no homology outside the putative target site duplications (alignments not shown), which suggested that the "core" MITE-like sequences are located at different genome positions. In contrast, MITE-like "core" regions that immediately flanked *D. theda* and *P. xylostella* $(GTCT)_n$ microsatellites showed ≤0.05 nucleotide substitutions per nucleotide site within each species (Fig. 1). Active transposition of *D. theda* and *P. xylostella* $(GTCT)_n$ microsatellite loci could not be shown, since homologous sequences 5′ and 3′ of target site duplications that also lacked the putative MITE were not present in GenBank. Analogous to OnMITE01 and On-MITE02, the *D. theda* and *P. xylostella* $(GTCT)_n$ microsatellite presented in Fig. 1 may be associated with interspersed repetitive elements. Proliferation of lepidopteran MITEs (LepMITEs) that also carry microsatellite repeats within respective genomes may be responsible for the difficulties that have been incurred during single-locus marker development in several species (Zhang 2004).

Phylogenetic analysis showed that four closely related MITE-like elements are present among lepidopteran genomes (LepMITE01 to LepMITE04). Sequence comparison indicated ≤0.17 nucleotide substitutions per nucleotide site within MITE families, and LepMITE families 03 and 04 appeared to be exclusively present within *P. xylostella* and *D. theda* genomes, respectively. The low level of sequence divergence among LepMITE families suggested that they were derived from a single ancestral element. Our comparative analysis suggested that a class of lepidopteran microsatellites are contained within families of related MITE sequences and are multilocus owing to hitchhiking within these mobile units. Tetranucleotide microsatellite repeat arrays previously were encountered within a microsatellite-inducing mobile element, *mini-me*, from *Drosophila*, which was hypothesized to function in the dispersal and genesis of new microsatellite loci (Wilder and Hollocher 2001). The origin of lepidopteran microsatellites encoded by mobile MITES is

unknown, but presence in multiple species suggests that entrance may have occurred early in the lepidopteran lineage. The total number of lepidopteran MITE families that contain mobile $(CTGT)_n$ microsatellite repeats, or other microsatellite repeats, is unknown. The presence of a $(GT)_9$ repeat in the putative *C. hero* MITE sequence suggested that $(CTGT)_n$ microsatellites may have undergone substitution mutation such that other microsatellite repeat motifs also may be present within mobile MITEs and will provide an area of new research in lepidopteran genomics.

Conclusions

BAC end sequencing and STS marker screening were used to characterize the *O. nubilalis* BAC library, OnB1. Representation of OnB1 was demonstrated through identification of 20 autosomal and 4 Z-chromosome-linked genes and included genes shown to be candidate *B. thuringiensis* toxin resistance genes. The OnB1 library was shown to contain clones representative of the Z chromosome involved in the determination of male *Resp* and larval *Ppd*. Therefore, if linkage mapping efforts improve QTL intervals, we expect OnB1 to serve as a genomic resource in the positional cloning of the genes underlying those QTL. Our BES data have proven to be a resource for the identification of repetitive genome elements from Lepidoptera. Two *ezi*-like LINEs were identified and shown to be multicopy within OnB1. Two microsatellite-containing MITE-like sequences from OnB1 BES were shown to be mobile with the *Ostrinia* genome and had homology to other lepidopteran microsatellite sequences. In the current study, we proposed a connection between MITEs and mobile microsatellites within the genome of Lepidoptera. The presence of multilocus microsatellite families in Lepidoptera has hindered the development of single-locus markers for population and linkage analysis. Development of a repetitive element bank for Lepidoptera for use with the RepeatMasker2 algorithm (www. repeatmasker.org/webrepeatmaskerhelp.html#references) may be an effective method by which to avoid the design of PCR primers within known repetitive elements but is contingent on the future description of additional mobile element families.

Acknowledgements

References

Anderson, S.J., Gould, P., and Freeland, J.R. 2007. Repetitive flanking sequences (ReFS): novel molecular markers from microsatellite families. Mol. Ecol. Notes, **7**: 374–376. doi:10.1111/j.1471-8286.2006.01628.x.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature (Lond.), **408**: 796–815. doi:10.1038/35048692. PMID:11130711.

Bruno, W.J., Knill, E., Balding, D.J., Bruch, D.C., Doggett, N.A., Sawhill, W.W., et al. 1995. Efficient pooling designs for library

screening. Genomics, **26**: 21–30. doi:10.1016/0888-7543(95)80078-Z. PMID:7782082.

Chen, S., and Li, X. 2007. Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome P450 genes. BMC Evol. Biol. **7**: 46. doi:10.1186/1471-2148-7-46. PMID:17381843.

Coates, B.S., and Hellmich, R.L. 2003. Two sex-chromosome-linked microsatellite loci show geographic variance among North American *Ostrinia nubilalis*. J. Insect Sci. **3**: 29. PMID:15841245.

Coates, B.S., Sumerford, D.V., Hellmich, R.L., and Lewis, L.C. 2005*a*. Sequence variation in the cadherin gene of *Ostrinia nubilalis*: a tool for field monitoring. Insect Biochem. Mol. Biol. **35**: 129–139. doi:10.1016/j.ibmb.2004.10.008. PMID:15681223.

Coates, B.S., Hellmich, R.L., and Lewis, L.C. 2005*b*. Two differentially expressed ommochrome-binding protein-like genes (*obp*1 and *obp*2) in larval fat body of the European corn borer, *Ostrinia nubilalis*. J. Insect Sci. **5**: 19. PMID:16341251.

Coates, B.S., Hellmich, R.L., Sumerford, D.V., and Lewis, L.C. 2007. A β-1,3-galactosyltransferase and brainiac/bre5 homolog expressed in the midgut did not contribute to a Cry1Ab toxin resistance trait in *Ostrinia nubilalis*. Insect Biochem. Mol. Biol. **37**: 346–355. doi:10.1016/j.ibmb.2006.12.008. PMID:17368198.

Coates, B.S., Sumerford, D.V., and Lewis, L.C. 2008. Segregation of European corn borer, *Ostrinia nubilalis*, aminopeptidase 1, cadherin, and bre5-like alleles from a colony resistant to *Bacillus thuringiensis* Cry1Ab toxins are not associated with F_2 larval weights when fed a diet containing Cry1Ab. J. Insect Sci. **8**: 21.

Delseny, M. 2004. Re-evaluating the relevance of ancestral shared synteny as a tool for crop improvement. Curr. Opin. Plant Biol. **7**: 126–131. doi:10.1016/j.pbi.2004.01.005. PMID:15003211.

Dopman, E.B., Bogdanowicz, S.M., and Harrison, R.G. 2004. Genetic mapping of sexual isolation between *E* and *Z* pheromone strains of the European corn borer (*Ostrinia nubilalis*). Genetics, **167**: 301–309. doi:10.1534/genetics.167.1.301. PMID:15166156.

Dopman, E.B., Perez, L., Bogdanowicz, S.M., and Harrison, R.G. 2005. Consequences of reproductive barriers for genealogical discordance in the European corn borer. Proc. Natl. Acad. Sci. U.S.A. **102**: 14706–14711. doi:10.1073/pnas.0502054102. PMID:16204000.

Dufresne, M., Hau-Van, A., Abd el Wahab, H., M'Barek, S.B., Vasnier, C., Teysset, L., et al. 2007. Transposition of a fungal miniature inverted-repeat transposable element through the action of a Tc1-like transposase. Genetics, **175**: 441–452. doi:10.1534/genetics.106.064360. PMID:17179071.

Eckardt, N.A. 2001. Everything in its place: conservation of gene order among distantly related plant species. Plant Cell, **13**: 723–725. doi:10.1105/tpc.13.4.723. PMID:11283330.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. **8**: 175–185. PMID:9521921.

Fagerberg, A.J., Fulton, R.E., and Black, W.C. 2001. Microsatellite loci are not abundant in all arthropod genomes: analysis in the hard tick, *Ixodes scapularis* and the yellow fever mosquito, *Aedes aegypti*. Insect Mol. Biol. **10**: 225–236. doi:10.1046/j.1365-2583.2001.00260.x. PMID:11437914.

Gage, L.P. 1974. The *Bombyx mori* genome: analysis by DNA reassociation kinetics. Chromosoma, **45**: 27–42. doi:10.1007/BF00283828. PMID:4837969.

Gahan, L.J., Gould, F., and Heckel, D.G. 2001. Identification of a gene associated with Bt resistance in *Heliothis virescens*. Science (Washington, D.C.), **293**: 857–860. doi:10.1126/science.1060949.

Glover, T., Campell, M., Robbins, P., and Roelofs, W. 1990. Sex-linked control of sex pheromone behavioral responses in European corn borer moths (*Ostrinia nubilalis*) confirmed with TPI marker gene. Arch. Insect Biochem. Physiol. **15**: 67–77. doi:10.1002/arch.940150202.

Glover, T.J., Robbins, P.S., Eckenrode, C.J., and Roelofs, W.L. 1992. Genetic control of voltinism characteristics in European corn borer races assessed with a marker gene. Arch. Insect Biochem. Physiol. **20**: 107–117. doi:10.1002/arch.940200203.

Heckel, D.G., Gahan, L.J., Daly, J.C., and Trowell, S. 1998. A genomic approach to understanding *Heliothis* and *Helicoverpa* resistance to chemical and biological insecticides. Philos. Trans. R. Soc. Lond. B Biol. Sci. **353**: 1713–1722. doi:10.1098/rstb.1998.0323.

Hoekstra, R., Criado-Fornelio, A., Fakkeldij, J., Bergman, J., and Roos, M.H. 1997. Microsatellites of the parasitic nematode *Haemonchus contortus*: polymorphism and linkage with a direct repeat. Mol. Biochem. Parasitol. **89**: 97–107. doi:10.1016/S0166-6851(97)00108-4. PMID:9297704.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature (Lond.), **409**: 860–921. doi:10.1038/35057062. PMID:11237011.

International Lepidopteran Genome Project. 2001. Available from http://papilio.ab.a.u-tokyo.ac.jp/lep-genome/proposal.pdf.

Kumar, S., Tamura, K., and Nei, M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5**: 150–163. doi:10.1093/bib/5.2.150. PMID:15260895.

Lin, Y., Draye, X., Qian, X., Ren, S., Zhu, L., Tomkins, J., et al. 2000. Locus-specific contig assembly in highly-duplicated genomes, using the BAC–RF method. Nucleic Acids Res. **28**: e23. doi:10.1093/nar/28.7.e23. PMID:10710440.

Malausa, T., Leniaud, L., Martin, J.F., Audiot, P., Bourguet, D., Ponsard, S., et al. 2007. Molecular differentiation at nuclear loci in French host races of the European corn borer (*Ostrinia nubilalis*). Genetics, **176**: 2343–2355. doi:10.1534/genetics.107.072108. PMID:17603115.

Meglécz, E., Anderson, S.J., Bourguet, D., Butcher, R., Caldas, A., Cassel-Lundhagen, A., et al. 2007. Microsatellite flanking region similarities among different loci within insect species. Insect Mol. Biol. **16**: 175–185. doi:10.1111/j.1365-2583.2006.00713.x. PMID:17298557.

Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., et al. 2004. The genome sequence of silkworm, *Bombyx mori*. DNA Res. **11**: 27–35. doi:10.1093/dnares/11.1.27. PMID:15141943.

Pringle, E.G., Baxter, S.W., Webster, C.L., Papanicolaou, A., Lee, S.F., and Jiggins, C.D. 2007. Synteny and chromosome evolution in the Lepidoptera: evidence from mapping in *Heliconius melpomene*. Genetics, **177**: 417–426. doi:10.1534/genetics.107.073122. PMID:17603110.

Quesneville, H., Nouaud, D., and Anxolabéhère, D. 2006. *P* elements and MITE relatives in the whole genome sequence of *Anopheles gambiae*. BMC Genomics, **7**: 214. doi:10.1186/1471-2164-7-214. PMID:16919158.

Rasch, E.M. 1985. DNA "standards" and the range of accurate DNA estimates by Feulgen absorption microspectrophotometry. *In* Advances in microscopy. *Edited by* R.R. Cowden and S.H. Harrison. Alan R Liss, New York. pp. 137–166.

Sankoff, D., and Nadeau, J.H. 2003. Chromosome rearrangements in evolution: from gene order to genome sequence and back. Proc. Natl. Acad. Sci. U.S.A. **100**: 11188–11189. doi:10.1073/pnas.2035002100. PMID:14506293.

SanMiguel, P., and Bennetzen, J.L. 1998. Evidence that a recent increase in *maize* genome size was caused by the massive amplification of intergene retrotransposons. Ann. Bot. (Lond.), **82**: 37–44. doi:10.1006/anbo.1998.0746.

Schmidt, R. 2002. Plant genome evolution: lessons from comparative genomics at the DNA level. Plant Mol. Biol. **48**: 21–37. doi:10.1023/A:1013788306201. PMID:11860210.

Showers, W.B., Brindley, T.A., and Reed, G.L. 1972. Survival and diapause characteristics of hybrids of three geographical races of the European corn borer. Ann. Entomol. Soc. Am. **65**: 450–457.

Sommer, H., Bonas, U., and Saedler, H. 1988. Transposon-induced alterations in the promoter region affect transcription of the chalcone synthase gene of *Antirrhinum majus*. Mol. Gen. Genet. **211**: 49–55. doi:10.1007/BF00338392.

Tu, Z. 2000. Molecular and evolutionary analysis of two divergent subfamilies of a novel miniature inverted repeat transposable element in the yellow Fever mosquito, *Aedes aegypti*. Mol. Biol. Evol. **17**: 1313–1325. PMID:10958848.

Tu, Z. 2001. Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. Proc. Natl. Acad. Sci. U.S.A. **98**: 1699–1704. doi:10.1073/pnas.041593198. PMID:11172014.

Tu, Z., and Orphanidis, S.P. 2001. Microuli, a family of miniature subterminal inverted repeat transposable elements (MSITEs): transposition without terminal inverted repeats. Mol. Biol. Evol. **18**: 893–895. PMID:11319273.

Van't Hof, A.E., Brakefield, P.M., Saccheri, I.J., and Zwaan, B.J. 2007. Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. Heredity, **98**: 320–328. doi:10.1038/sj.hdy.6800944. PMID:17327875.

Wessler, S.R., Bureau, T.E., and White, S.E. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. **5**: 814–821. doi:10.1016/0959-437X(95)80016-X. PMID:8745082.

Wilder, J., and Hollocher, H. 2001. Mobile elements and the genesis of microsatellites in dipterans. Mol. Biol. Evol. **18**: 384–392. PMID:11230539.

Willett, C.S., and Harrison, R.G. 1999. Pheromone binding proteins in the European and Asian corn borers: no protein change associated with pheromone differences. Insect Biochem. Mol. Biol. **29**: 277–284. doi:10.1016/S0965-1748(99)00003-X. PMID:10319441.

Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., et al. 2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science (Washington, D.C.), **306**: 1937–1940. doi:10.1126/science.1102210. PMID:15591204.

Xue, B., Rooney, A.P., Kajikawa, M., Okada, N., and Roelofs, W.L. 2007. Novel sex pheromone desaturases in the genomes of corn borers generated through gene duplication and retroposon fusion. Proc. Natl. Acad. Sci. U.S.A. **104**: 4467–4472. doi:10.1073/pnas.0700422104. PMID:17360547.

Yamamoto, K., Narukawa, J., Kadone-Okuda, K., Nohata, J., Sasanuma, M., Suetsugu, Y., et al. 2006. Construction of a single nucleotide polymorphism linkage map for the silkworm, *Bombyx mori*, based on bacterial artificial chromosome end sequences. Genetics, **173**: 151–161. doi:10.1534/genetics.105.053801. PMID:16547112.

Yamamoto, K., Nohata, J., Kadono-Okuda, K., Narukawa, J., Sasanuma, M., Sasanuma, S., et al. 2008. A BAC-based integrated linkage map of the silkworm *Bombyx mori*. Genome Biol. **9**: R21. doi:10.1186/gb-2008-9-1-r21. PMID:18226216.

Yasukochi, Y., Ashakumary, L.A., Baba, K., Yoshido, A., and Sahara, K. 2006. A second-generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects. Genetics, **173**: 1319–1328. doi:10.1534/genetics.106.055541. PMID:16547103.

Yim, Y.S., Moak, P., Sanchex-Villeda, H., Musket, T.A., Close, P., Klein, P.E., et al. 2007. A BAC pooling strategy combined with PCR-based screenings in a large, highly repetitive genome enables integration of the maize genetic and physical maps. BMC Genomics, **8**: 47. doi:10.1186/1471-2164-8-47. PMID:17291341.

Zhang, D.X. 2004. Lepidopteran microsatellite DNA: redundant but promising. Trends Ecol. Evol. **19**: 507–509. doi:10.1016/j.tree.2004.07.020. PMID:16701315.

Zuker, M., Mathews, D.H., and Turner, D.H. 1999. Turner algorithms and thermodynamics for RNA secondary structure prediction: a practical guide in RNA biochemistry and biotechnology. *Edited by* J. Barciszewski and B.F.C. Clark. NATO ASI Series. Kluwer Academic Publishers, Boston, Mass.