1-2014

Proliferation and copy number variation of BEL-like long terminal repeat retrotransposons within the Diabrotica virgifera virgifera genome

Brad S. Coates
Iowa State University, brad.coates@ars.usda.gov

Lisa M. Fraser
Iowa State University, lmfraser@iastate.edu

B. Wade French
United States Department of Agriculture

Thomas W. Sappington
Iowa State University, tsapping@iastate.edu

Proliferation and copy number variation of BEL-like long terminal repeat retrotransposons within the Diabrotica virgifera virgifera genome

Abstract
The proliferation of retrotransposons within a genome can contribute to increased size and affect the function of eukaryotic genes. BEL/*Pao*-like long-terminal repeat (LTR) retrotransposons were annotated from the highly adaptable insect species *Diabrotica virgifera virgifera*, the Western corn rootworm, using survey sequences from bacterial artificial chromosome (BAC) inserts and contigs derived from a low coverage next-generation genome sequence assembly. Eleven unique *D. v. virgifera* BEL elements were identified that contained full-length *gag–pol* coding sequences, whereas 88 different partial coding regions were characterized from partially assembled elements. Estimated genome copy number for full and partial BEL-like elements ranged from ~ 8 to 1582 among individual contigs using a normalized depth of coverage (DOC) among Illumina HiSeq reads (total genome copy number ~ 8821). BEL element copy number was correlated among different *D. v. virgifera* populations ($R^2 = 0.9846$), but individual element numbers varied ≤ 1.68-fold and the total number varied by ~ 527 copies. These data indicate that BEL element proliferation likely contributed to a large genome size, and suggest that differences in copy number are a source of genetic variability among *D. v. virgifera*.

Keywords
Genome variation, Integration/excision mutation, Corn rootworm

Disciplines
Animal Sciences | Entomology | Genetics | Genomics

Comments
This article is from *Gene* 534 (2014): 362, doi:10.1016/j.gene.2013.09.100.

Gene 534 (2014) 362–370

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Proliferation and copy number variation of BEL-like long terminal repeat retrotransposons within the *Diabrotica virgifera virgifera* genome

Brad S. Coates [a,b,*], Lisa M. Fraser [c,1], B. Wade French [d,2], Thomas W. Sappington [a,b,3]

[a] United States Department of Agriculture, Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA 50011, USA
[b] Iowa State University, Department of Entomology, Ames, IA 50011, USA
[c] Iowa State University, Biomedical Sciences Department, Ames, IA 50011, USA
[d] United States Department of Agriculture, Agricultural Research Service, North Central Agricultural Research Laboratory, Brookings, SD 57006, USA

ARTICLE INFO

Article history:
Accepted 26 September 2013
Available online 26 October 2013

Keywords:
Genome variation
Integration/excision mutation
Corn rootworm

ABSTRACT

The proliferation of retrotransposons within a genome can contribute to increased size and affect the function of eukaryotic genes. BEL/*Pao*-like long-terminal repeat (LTR) retrotransposons were annotated from the highly adaptable insect species *Diabrotica virgifera virgifera*, the Western corn rootworm, using survey sequences from bacterial artificial chromosome (BAC) inserts and contigs derived from a low coverage next-generation genome sequence assembly. Eleven unique *D. v. virgifera* BEL elements were identified that contained full-length *gag–pol* coding sequences, whereas 88 different partial coding regions were characterized from partially assembled elements. Estimated genome copy number for full and partial BEL-like elements ranged from ~8 to 1582 among individual contigs using a normalized depth of coverage (DOC) among Illumina HiSeq reads (total genome copy number ~8821). BEL element copy number was correlated among different *D. v. virgifera* populations ($R^2 = 0.9846$), but individual element numbers varied ≤1.68-fold and the total number varied by ~527 copies. These data indicate that BEL element proliferation likely contributed to a large genome size, and suggest that differences in copy number are a source of genetic variability among *D. v. virgifera*.

Published by Elsevier B.V.

1. Introduction

Transposable elements (TEs) are mobile "selfish" DNAs that have propagated within and can comprise a significant proportion of eukaryotic genomes (SanMiguel et al., 1996). TEs are categorized as RNA-based class I (retrotransposons) or DNA based class II elements, which are mobilized by reverse transcription of an mRNA-intermediate or by conservative DNA "cut-and-paste" mechanisms, respectively (Kidwell and Lisch, 2001). Based on gene content, mode of integration and phylogenetic relationship among encoded proteins, retrotransposons are further classified into long terminal repeats (LTRs), long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) (together also known as non-LTR), direct inverted repeat (DIR)-like, and Penelope-like elements (Wicker et al., 2007). LTR retroelements retain retroviral-like genes and a conserved mechanism for genome insertion (Boeke and Stoye, 1997). They also contain internal RNA Pol II promoters and termination signals that are necessary for the transcription of an mRNA intermediate containing *gag* and *pol* genes. The *gag* gene encodes a structural virus-like particle (VLP), and the resultant VLP proteins oligomerize to encapsulate the retroelement mRNAs. In contrast to retroviruses, encapsulated retrotransposon mRNAs cannot cross nuclear or cell membranes due to loss of the envelope (*env*) gene. The *pol* gene encodes a multifunctional enzyme with reverse transcriptase (RT), ribonuclease H (RH), protease (PR) and integrase (IN) activities. Based upon the order of genes (or functional domains) LTR retrotransposons are categorized as Ty1/*copia* elements with the order IN-RT, and Ty3/*gypsy* or BEL/*Pao* elements with the RT and IN domains inverted.

LTR retrotransposon integration is directed by the IN domain, and non-random distribution of integrations near gene-rich regions in

Abbreviations: A, alanine; BAC, bacterial artificial chromosome; BRSD, Brookings, South Dakota population; *cad*, cadherin gene; CDD, conserved domain database; CNV, copy number variation; D, aspartic acid; DOC, depth of coverage; E, glutamic acid; *env*, envelope gene; EST, expressed sequence tag(s); *for*1, forager 1 gene; G, glycine; *gag*, group specific antigen gene; HTGS, high throughput genomic sequence; I, isoleucine; IN, integrase protein domain; L, leucine; LINE, long interspersed nuclear element(s); LTR, long terminal repeat; ML, maximum likelihood; ND, non-diapause; ORF, open reading frame(s); PR, protease protein domain; *pol*, polyprotein gene; RH, ribonuclease H protein domain; RPKM, reads per kilobase per million mapped; RT, reverse transcriptase protein domain; rtREV, general reverse transcriptase model; SAM, sequence alignment/map format; SINE, short interspersed nuclear element(s); T, threonine; TE, transposable element; V, valine; VLP, virus-like particle; X, any amino acid; Y, tyrosine.

* Corresponding author at: USDA-ARS, Corn Insects & Crop Genetics, 103 Genetics Laboratory, Iowa State University, Ames, IA, USA. Tel.: +1 515 294 6948.
E-mail addresses: brad.coates@ars.usda.gov (B.S. Coates), lmfraser@iastate.edu (L.M. Fraser), wade.french@ars.usda.gov (B.W. French), tom.sapptington@ars.usda.gov (T.W. Sappington).
[1] Biomedical Sciences, Iowa State University, Ames, IA 50011, USA. Tel.: +1 515 594 4387.
[2] USDA-ARS, North Central Ag Research, 2923 Medary Ave., Brookings, SD 57006, USA. Tel.: +1 605 693 5210.
[3] USDA-ARS, Corn Insects & Crop Genetics, 106 Genetics Laboratory, Iowa State University, Ames, IA, USA. Tel.: +1 515 294 9759.

0378-1119/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.gene.2013.09.100

eukaryotic genomes suggests target-site specificity (Kim et al., 1998). For example, yeast Ty1 and Ty3 retroelements are preferentially located upstream of RNA Pol III promoters, where IN may interact with the Pol III transcriptional machinery or proteins associated with open chromatin (Bachman et al., 2005; Mou et al., 2006; Yieh et al., 2000, 2002). Movement of TEs within a genome can change chromosome structure (Eichler and Sankoff, 2003; Feschotte and Pritham, 2007; Kidwell and Lisch, 2001) and modulate the expression of nearby protein coding genes (Feschotte, 2008). Actively transcribed retrotransposons remain mobile within a genome, and are potent modulators of gene function by producing genomic variation at points of integration. Specifically, the integration of transposable elements (TEs) can cause mutations that give rise to phenotypic variation among individuals within populations (McClintock, 1950; Wendel and Wessler, 2000), and may generate genetic novelties that contribute to local adaptations (Gonzalez et al., 2010). TEs can also cause insertional knockout of a gene following integration within an open reading frame (Gahan et al., 2001), or decreased splicing efficiency and aberrant polyadenylation when integrated in an intron (Beeman et al., 1996; Kaer et al., 2011). Additionally, integrations within transcriptional regulatory elements (promoters and enhancers) can affect the expression of nearby genes (Chung et al., 2007). Understanding TE structure, evolution and interactions within eukaryotic genomes is therefore important in uncovering roles in genome evolution and species adaptation. This is exemplified in the TE-induced mutations that are linked to insecticide resistance traits in arthropods (Aminetzach et al., 2005; Chung et al., 2007; Fabrick et al., 2011; Gahan et al., 2001), and supports the premise that TEs have genome-wide effects on genetic diversity and subsequent phenotypic variation (Kaminker et al., 2002).

The Western corn rootworm, *Diabrotica virgifera virgifera* (Coleoptera: Chrysomelidae), is a highly destructive insect pest of cultivated corn in most of the United States and Central Europe (Gray et al., 2009). Subterranean larvae damage roots (Chiang, 1973), which decreases the plant's structural stability and capacity to absorb water and soil nutrients (Kahler et al., 1985). Adult beetles feed on corn silks and can reduce corn pollination rates as well as transmit plant pathogens (Gilbertson et al., 1986; Jensen, 1985). Protection of corn crops from *D. v. virgifera* is difficult in part due to recurring selection of populations for survival when exposed to large-scale pest management practices that rely on different insecticide modes of action, including organochlorine, organophosphate and carbamate chemistries (Gray et al., 2009; Meinke et al., 1998, 2009; Metcalf, 1986; Parimi et al., 2006). Recently, resistance to commercial corn hybrids that express a *Bacillus thuringiensis* (Bt) toxin has been documented in some *D. v. virgifera* populations (Gassmann et al., 2011). The genetic basis of many resistance traits in *D. v. virgifera* remains undetermined, but resistance to the organochlorine insecticide, aldrin, has been linked to a point mutation within the gamma-aminobutyric acid (GABA) receptor (Wang et al., 2013). Organophosphate resistance traits were partially linked to increased carboxylesterase activities in adult *D. v. virgifera* (Miota et al., 1998; Zhou et al., 2003), but the mechanism(s) underlying this adaptation remains unknown. The capacity of *D. v. virgifera* larvae and adults to rapidly adapt to insecticides highlights the need to investigate underlying genetic and genomic mechanisms (Gray et al., 2009; Sappington et al., 2006), with the anticipated benefits of developing sustainable management practices.

The *D. v. virgifera* genome size has been estimated as 2.58 Gb (Coates et al., 2012). Sequence analysis indicated that ~16% of BAC ends consisted of retroelement protein coding sequence (RT, PR, RH and IN protein domains) and that retroelements comprise a large portion of assembled BAC insert sequence (Coates et al., 2012). Mobility of TEs in the *D. v. virgifera* genome was inferred from haplotype variation among cloned DNA sequence, suggesting that TE mobility may contribute to the accumulation of new mutations (Coates et al., 2012). However, TE sequence diversity and phenotypic consequences remain unknown. Here we report the first description of BEL-like LTR retrotransposons in *D. v. virgifera* which may assist future annotation of a genome sequence assembly and clarify *D. v. virgifera* genome evolution in general. Furthermore, we report that BEL-like LTR retrotransposon copy number varies between populations. Although specific integration mutations were not linked to any *D. v. virgifera* phenotype, TE copy number variation suggests that integration may serve as a large source of novel genetic variation in this highly adaptable species.

2. Materials and methods

2.1. Identification and annotation of LTR retrotransposons

A bioinformatic approach was used to predict putative LTR retrotransposon sequences from available *D. v. virgifera* genomic sequence resources. A BAC library constructed from *D. v. virgifera*, called DvvBAC1, was previously described by Coates et al. (2012). High throughput genomic sequence (HTGS) submissions representing contigs derived from full DvvBAC1 inserts were downloaded from the National Center for Biotechnology Information; JQ581035 to JM581042 (Coates et al., 2012), KC248067 to KC248069 (Wang et al., 2013), and KC962414 to KC962431 (Coates et al., unpublished). BAC inserts were sequenced on a Roche 454 sequencer, and read data were assembled into contigs using the Newbler assembler (Roche Applied Sciences, Penzberg, Germany). The entire BAC insert could not be reconstructed into a single contig for all clones, so some HTGS entries are working drafts of unordered fragments (phase 1 submissions; see Coates et al., 2012).

Contigs of genomic DNA were also constructed by assembling short shotgun reads from a low-coverage sequencing of the *D. v. virgifera* genome. Specifically, 50 adult non-diapausing (ND) strain *D. v. virgifera* beetles (Branson, 1976) were obtained from the United States Department of Agriculture — Agricultural Research Service, North Central Agricultural Research Laboratory (USDA-ARS, NCARL), Brookings, SD, and 80 wild adults were collected on 25 July 2012 from a cornfield at the Eastern Soil and Water Research Farm, Brookings, SD (BRSD). Beetles from both populations were maintained separately. Live beetles were pooled by population, starved for 2 days, then ground to a powder in liquid nitrogen. DNA was extracted from ~5 mg of material using Qiagen G20 gravity flow columns according to manufacturer instructions (Qiagen, Valencia, CA). DNA quality and quantity were determined at the Iowa State University DNA Sequence and Synthesis Facility (ISU-DSSF) on an Agilent 2100 BioAnalyzer. Randomly sheared ~500 bp fragments were generated from ~2.0 μg of DNA and used to create L006 (ND strain) and L007 (BRSD) indexed sequencing libraries using the Illumina paired-end sample prep kit according to manufacturer instructions (Illumina, San Diego, CA). Single end 100-cycle reads were used to generate DNA sequence data from approximately equal proportions of each library in a single Illumina HiSeq 2000 flow cell at the ISU-DSSF. FASTQ formatted output was trimmed of read data with PHRED quality scores <20 using the script, TrimmingReads.pl (NGSToolKit, Patel and Jain, 2012). A total of 30 million reads (~3.0 Gbp or ~1.17× genome coverage) from each of the trimmed FASTQ sequence files were loaded in the Velvet assembler (Zerbino and Birney, 2008) and assembled using the de Bruijn graph method with a hash size (*k*-mer) of 21, no coverage cutoff, and a minimum contig length (— min_contig_lgth) of 1000.

Contigs from HTGS accessions and the de novo genome assembly were used to create a local BLAST database (dbDvvContigs). This database was queried with the following BEL/*Pao* LTR retrotransposon and *gag–pol* protein sequences using the tblastn algorithm: *Ninja* (*Drosophila simulans*; BAD01589.1), *Roo* (*Drosophila melanogaster*; AAN87269.1), Bm*Pao* (*Bombyx mori* L09635), TcBEL (*Tribolium castaneum*; XP_001809963.1), AgBEL (*Anopheles gambiae*; CAJ14165), *Max* (*D. melanogaster*; CAD32253.1), BEL (*D. melanogaster*; U23420), *Moose* (*A. gambiae*; AF060859.1), *Suzu* (*Takifugu rubripes*; AF537216.1), *Saci*-1 (*Schistosoma mansoni*; DAA04498.1), and *Caenorhabditis elegans* *Cer*7 (AAB63932.1), *Cer*8 (CAB04994.1) and *Cer*11 (AAA82437.1). Results were filtered for *E*-values $\geq 10^{-30}$ and percent identities ≥35.

Contigs with significant "hits" were retrieved from the Velvet output file using the PERL script "get_selected_sequences.pl" (downloaded at http://alrlab.research.pdx.edu/aquificales/scripts/get_selected_sequences.pl). Longest open reading frames were predicted by ORF Finder (http://www.bioinformatics.org/sms2/orf_find.html) using the parameters of an ATG start codon, minimum length ≥300 codons in standard genetic code and in all six possible reading frames. Each putative BEL/*Pao* ORF as well as sequence encompassing up to 1000 bp upstream of the start codon and 1000 bp downstream of the stop codons were used as input for the program LTR Finder (Zhao and Wang 2007; default parameters). In instances where LTR Finder predicted no LTR retrotransposons, LTR regions were identified by alignment of DNA sequences located immediately upstream and downstream of start and stop codons using ClustalW. HTGS accessions with putative *D. v. virgifera* LTR retrotransposons were loaded into the Artemis genome viewer (Rutherford et al., 2000), and ORFs, direct flanking repeats, and target site duplications were annotated manually. Functional protein domains within putative *D. v. virgifera* BEL/*Pao gag–pol* proteins were identified by query of the conserved domain database (CDD; Marchler-Bauer et al., 2012). ORFs were stored in the file DvvBelORFs.fasta (Supplementary File S1).

2.2. Phylogenetic reconstruction

Reverse transcriptase (RT) domains were manually trimmed from amino acid sequences of previously described BEL/*Pao* LTR retrotransposons (see accessions in the previous section; representing *Pao*, *Saci*-1, BEL, *Tas* and *Suzu* subfamilies). Additionally, the RT domains were trimmed from the derived protein coding sequence of 37 putative BEL/*Pao*-like elements predicted within assembled *D. v. virgifera* contigs. Sequences were loaded into the MEGA 5.0 alignment utility (Tamura et al., 2011), and a multiple sequence alignment was constructed using ClustalW default parameters and the Gonnet protein weight matrix. The gamma shape and best model parameters were estimated by maximum likelihood (ML) using MEGA 5.0. Phylogenetic trees were constructed from exported .meg files using the ML method using complete deletion of gapped data under the general reverse transcriptase model (rtREV; Dimmic et al., 2002) + G (gamma distributed rate variation). Bootstrap support at each node was based on 1000 pseudoreplicates of the aligned data (Felsenstein, 1985). The RT domain from the *D. melanogaster gypsy* element (P10401.1) served as the outgroup.

2.3. Estimation of BEL copy number and expression

BEL-like retrotransposon copy number within the *D. v. virgifera* genome was estimated by mapping whole genome shotgun sequence read data from the ND laboratory strain and BRSD field population to 99 contigs in the DvvBelORFs.fasta file (Supplementary File S1). The cDNA sequence from the *D. v. virgifera* cadherin (*cad*; GenBank accession EF531715.1) and forager 1 (*for*1; DQ913742.1) genes were used as internal references for normalizing mapped Illumina read data derived from the ND strain and BRSD field population libraries, and provided single-copy controls to normalize the estimates of BEL retrotransposon copy number. Specifically, 50 million Illumina reads from ND strain and BRSD FASTQ files (previously trimmed of segments with PHRED quality scores <20) were aligned to the 99 *D. v. virgifera* contigs encoding BEL/*Pao*-like ORFs as well as to the *cad* and *for*1 cDNA sequences using the short read aligner Bowtie (Langmead et al., 2009; mismatches allowed (−v) 3; output in SAM format (−S)). The copy numbers of LTR retrotransposons were estimated from a ratio of the average coverage of mapped Illumina reads across each contig to that of the single-copy genes *cad* and *for*1 (i.e., mean coverage of contig containing a BEL-like element/mean coverage of single-copy gene). Copy number was also estimated as reads per kilobase per million mapped (RPKM) according to Tenaillon et al. (2011), and RPKM values for BEL elements were normalized using the ratio of RPKM to single-copy genes as described above.

Variation in copy number of each BEL element between ND strain and BRSD populations was estimated by plotting the estimated ratio-normalized copy number for each of the 99 BEL-containing contigs. This was performed separately for the mean read depths and RPKM estimates for each library. Estimates of BEL copy number normalized by the *cad* gene were used to calculate a ratio of BEL copy number within each population based on mean read depth. A ratio was similarly calculated using RPKM values. The ratio of copy number between populations was regressed on the log2 scaled difference between the ND strain vs. BRSD population.

Expressed sequence tag (EST) data for *D. v. virgifera* head and midgut tissues were downloaded from GenBank dbEST accessions EW761110.1–EW777358.1 (Knolhoff et al., 2010), and CN497248.1–CN498776.1 (Siegfried et al., 2005), and used to construct a blast database (dbDvEST). dbDvEST was queried with translated protein sequence from *D. v. virgifera* genomic contigs encoding BEL/*Pao* element *gag–pol* functional domains using the tblastn algorithm, and results were filtered for *E*-values $\geq 10^{-25}$ and percent identities ≥40. ESTs showing >98% similarity to a characterized BEL element from the non-diapausing strain were considered to be derived from the same mobile and multicopy locus.

3. Results and discussion

3.1. Identification and annotation of LTR retrotransposons

TE integrations have been implicated in the alteration of gene structure and function that has led to the evolution of insecticide resistance traits. Initial annotation of TE structure and abundance are fundamental for future studies of genome evolution and investigations of TE roles in shaping genetic variability in a species. To address this knowledge gap in the highly adaptable corn pest *D. v. virgifera*, we characterized the diversity and prevalence of BEL-like LTR retrotransposons within its genome. BEL/*Pao*, Ty1/*copia* and Ty4/*gypsy* groups are the main types of LTR-retrotransposon among animal species (Xiong and Eickbush, 1990), where BEL/*Pao* retroelements represent the most recently described clade identified through phylogenetic analysis of *gag–pol* coding regions (Frame et al., 2001). BEL/*Pao* element diversity remains relatively unknown, and their description herein provides insight into BEL/*Pao* structure, functional constraint, and diversification in eukaryotic genomes.

BEL/*Pao* elements were successfully identified and annotated from an assembly of the *D. v. virgifera* genome from low coverage read data and BAC full insert sequences. *De novo* assembly of an ~1.17× coverage depth of the *D. v. virgifera* genome from the ND strain (NCBI BioProject PRJNA205954; BioSample SAMN02183419: ND_pool2012; NCBI short read archive (SRA) run SRR921383) and a Brookings, SD field population (NCBI BioProject PRJNA206802: BioSample SAMN02190085; SRA run SRR921430) resulted in the assembly of 11.7 million reads into 4173 contigs (314,773 bp total contig length; range 1000 to 73,314 bp; N50 2646 bp). Due to the use of low genome coverage and unpaired read data during assembly, scaffolding or assembling larger portions of the *D. v. virgifera* genome were not attempted. Assembly parameters and number of reads affect contig number and length. Although not investigated in this study, the relaxation of assembly parameters probably increased contig length and decreased contig number, but also increased the probability of assembling non-homologous reads. Query of the 4173 *D. v. virgifera* genomic contigs with previously described BEL/*Pao gag–pol* protein sequences resulted in "hits" to 99 contigs with amino acid similarities ≥37.6% (contig size 1482 to 9296 bp; ORF size 434 to 1866 amino acids; Supplementary File S1). Similarly, queries of previously assembled BAC inserts (Coates et al., 2012; Wang et al., 2013) identified regions of clones 213A05 and 278L20 with ≥38% identity to the *T. castaneum*

BEL12_AG TcBEL LTR transposon (GenBank accession XP_001809963.1; remaining results not shown).

A total of 11 full-length and 88 partial BEL/*Pao* coding sequences were identified from *D. v. virgifera* genomic contigs and BAC inserts. Neither the LTR Finder utility nor alignment of flanking sequences with ClustalW predicted direct repeats upstream or downstream of the ORF for most elements, and was likely due to difficulties in assembly of non-unique sequence from the directly repeated LTR sequences. In instances where LTRs assembled, unique block deletions and single nucleotide polymorphisms were present, suggesting an increased age since integration. Full-length ORFs encoded putative *gag–pol*-like peptides with functional domains in the order GAG-PR-RT-RH-IN (Supplementary File S1), which is consistent with other known BEL/*Pao* and Ty3/*gypsy* elements (Frame et al., 2001; Xiong and Eickbush, 1990), while differing from the Ty1/*copia* order of GAG-PR-IN-RT-RH. An LTR promoter/primer binding site (PBS) predicted to bind $tRNA^{LysTTT}$ was identified upstream of the GAG domain. Additionally, a 15 bp polypurine tract (PPT), which functions during reverse transcription of *gag–pol* mRNAs, was annotated (Fig. 1). All putative translated *D. v. virgifera gag*-domains contained a characteristic BEL/*Pao* triple Cys–His box motif ($CX_2CX_{3-4}HX_4C$, $CX_2CX_{3-4}HX_{4-5}C$, and $CX_{2-4}CX_3HX_4H$) (Fig. 2A), and not the single zinc fingers found in Ty3/*gypsy* and Ty1/*copia gag*-domains (Bowen and McDonald, 1999; Copeland et al., 2005; Frame et al., 2001). BEL/*Pao* protease (PR) domains are characterized by unique ALLDXG and LIGXD active site motifs (Abe et al., 2001), and *D. v. virgifera* elements varied at the first position of the ALLDXG motif (Fig. 2B). BEL/*Pao* RT domains are presumably derived from Ty3/*gypsy* elements (Frame et al., 2001) and share a unique TVDN active site instead of F/YXDD found among other LTR-retrotransposons (Felder et al., 1994; Frame et al., 2001). The *D. v. virgifera* BEL/*Pao* element RT active site, YVDD, is similar to that of the BEL-like *Ninja* element from *D. simulans* (Ogura et al., 1996) and Bm*Pao* from *B. mori* (Xiong et al., 1993) (Fig. 3). Seven conserved amino acid blocks characteristic of the BEL/*Pao* element RT domain (Xiong et al., 1993) were identified in the putative *D. v. virgifera* BEL/*Pao* elements as well (Fig. 3). Finally, IN domains of *D. v. virgifera* elements contained a BEL/*Pao*-like Cys–His box and a conserved catalytic triad ($DX_{65}DX_{53}E$) (Fig. 4; Copeland et al., 2005). The spacing between catalytic residues within *D. v. virgifera* IN domains was $DX_{65}DX_{49\ to\ 53}E$ among all elements, except for those from contig0034 and contig2516 that had a $DX_{65}DX_{-41}E$ and $DX_{65}DX_{-40}E$ spacing, respectively (Fig. 4).

3.2. Phylogenetic reconstruction

Phylogenetic reconstruction of the BEL/*Pao* lineage was conducted using RT domains from 37 unique *D. v. virgifera* elements, as well as representatives from five BEL/*Pao* subfamilies (*Pao*, *Sinbad*/*Saci*-1, BEL, *Tas*, and *Suzu*) previously identified by phylogenetic analysis of RT domains by Copeland et al. (2005). A 157 amino acid long consensus alignment was generated (Supplementary File S2), from which a gamma parameter for among site rate variation was estimated by maximum likelihood as 1.2832 (likelihood = −13,138.012). The resulting ML tree placed constituent members for the RT domain alignment into the major subfamilies of the BEL/*Pao* LTR retrotransposons; *Pao*, BEL, *Saci*-1, *Tas*, and *Suzu* (Fig. 5). Tree topology was similar to that previously described among the major subfamilies defined by Copeland et al. (2005). Thirty six of the 37 unique *D. v. virgifera* RT domains were grouped with previously described BEL subfamily members *Max*, *Moose*, *Roo*, and *T. castaneum* BEL12-AG (TcBEL) with 72% of bootstrap iterations supporting this clade. Subgroups within the *D. v. virgifera* BEL-lineage lacked strong bootstrap support, perhaps reflecting low sequence homology and thus rapid sequence divergence among RT domains.

3.3. Estimation of BEL copy number and expression

3.3.1. Genome copy number variation between populations

Ty3/*gypsy* elements are typically more prevalent in eukaryotic genomes than BEL/*Pao*-like and Ty1/*copia* elements with intermediate and low copy numbers, respectively (Felder et al., 1994; Xiong et al., 1993). Estimating TE numbers from species that lack whole genome sequence assemblies presents unique challenges. NGS data has been used to determine copy number variants (CNVs) in genomes by mapping reads to a reference genome, but this is an evolving field where best practices or pipelines are not yet established (Teo et al., 2012). Depth of coverage (DOC) estimates have been applied to predict CNVs by assuming that sequence data is generated uniformly across a genome (following a Poisson distribution) and depth of aligned reads is proportional to genome copy numbers (Xie and Tammi, 2009). Our DOC analyses of the 99 contigs with BEL coding regions resulted in mean read depths among individual BEL elements ranging from 12.3 to 2024.5 (normalized 9.6 to 1581.6) and 11.0 to 1842.5 (normalized 8.3 to 1395.8) from ND strain and BRSD libraries, respectively (Supplementary Table S3). GC nucleotide content can lead to under- or overrepresentation of sequence reads acquired by NGS technologies, and the need for GC-corrections has been proposed (Abyzov et al., 2011). However, the DOC (mean read depth) for the single copy *D. v. virgifera cad* gene (35.68% GC content) of 1.28 and 1.32 respectively for ND strain and BRSD libraries was not significantly different than the 1.18 and 1.41 mean read depths for the single copy *D. v. virgifera for*1 gene (41.79% GC), respectively estimated from ND strain and BRSD libraries. Because the DOC observed between single copy genes was similar, we reasoned

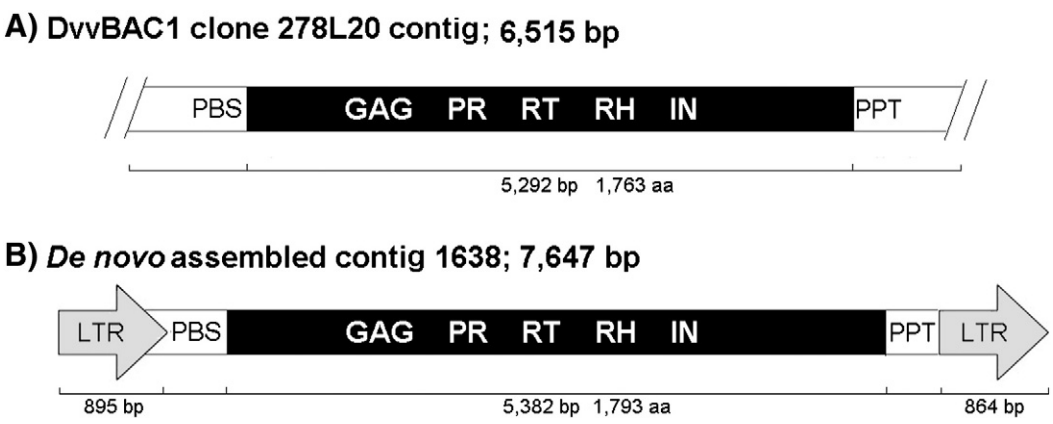

Fig. 1. Structural arrangement of BEL subfamily LTR retrotransposons in the *D. v. virgifera* genome. A) A BEL-like element from BAC clone 278L20 that lacks long-terminal repeats (LTRs), and B) a full-length BEL retroelement derived from a de novo assembly of the *D. v. virgifera* genome. Primer binding sequence (PBS; CGC TGG AAA GAG ACA TTG) and polypurine tract (PPT; GCA AAA TAG AAA GTG) nucleotide sequences are indicated, as well as group specific antigen (GAG; pfam03564), aspartic protease (PR; pfam05585), reverse transcriptase (RT; pfam00078), ribonuclease H (RH; pfam00075), and integrase (IN; pfam00665) protein coding domains. LTRs shown as arrows indicate direction of LTR direct repeats.

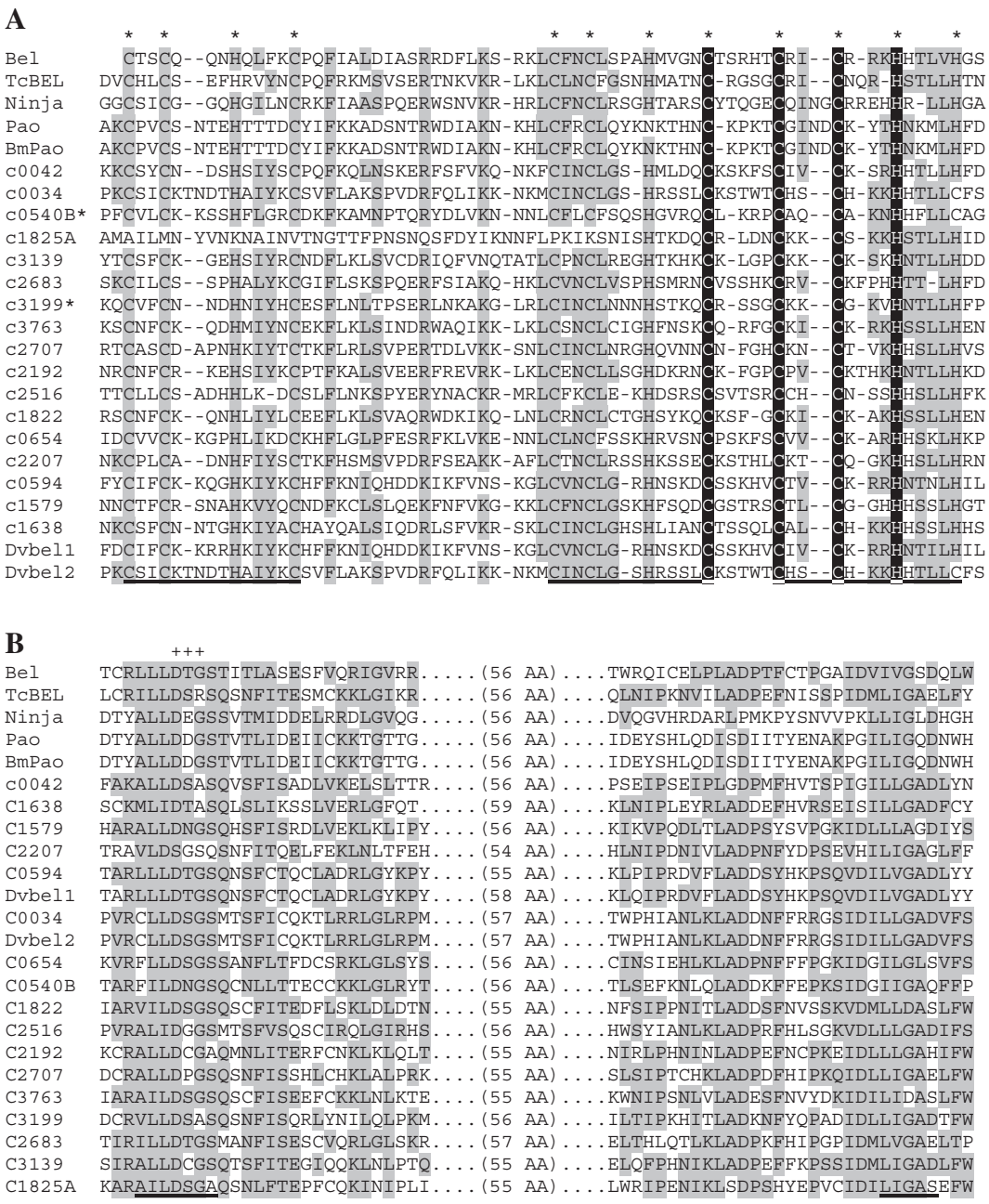

Fig. 2. Multiple alignment of amino acid sequences of *D. v. virgifera* BEL-like retrotransposons. A) Conserved triple Cys–His box motifs ($CX_2CX_{3-4}HX_4C$, $CX_2CX_{3-4}HX_{4-5}C$, and $CX_{2-4}CX_3HX_4H$) of zinc fingers within the *gag* gene-encode nucleocapsid protein are underlined, and B) ALLDXG and LIGXD motifs within the protease (PR) domain are underlined and catalytic residues are marked with a (+). Residues showing 100% conservation are shaded in black, and chemically similar amino acids in at least half of the sequences are shaded in gray. BEL (*D. melanogaster*; T13250); TcBEL (*T. castaneum*; XP_001809963.1); *Ninja* (*D. melanogaster*; G002006), BmPao (*B. mori*; L09635), and numbers preceded by a "c" indicate the de novo assembled *D. v. virgifera* contig from which the sequence was derived.

that GC correction was not required. Subsequent normalization of data based on DOC for the single copy *cad* gene suggested that individual contigs containing BEL elements have copy numbers from ~10 to 1582. Based on the difference in total genome copy number between the ND strain ($n = 8821$) and the BRSD population ($n = 8294$), we estimated a total CNV between populations of ~527.

Fold difference in DOC (mean read depth) estimated between individual BEL-like contigs ranged from 1.278 to −1.688 between ND and BRSD populations (Fig. 6). The greatest absolute difference in the estimated number of an individual BEL element between the ND strain ($n = 45$) and the BRSD field population ($n = 36$) was for contig3376 (−1.688 fold change). The estimated relative BEL element copy numbers were similar between populations ($R^2 = 0.9846$) (Fig. 6). TE family member copy numbers in *Zea mays* and *Zea luxurians* genomes are also highly correlated (Tenaillon et al., 2011), whereas Hawkins et al. (2006) described differential amplification of *copia*-like elements among species of *Gossypium*. Correlation between copy numbers of BEL-like contigs in *D. v. virgifera* does not conflict with the high total CNV between populations described above, rather it suggests that copy numbers for variant BEL elements (represented as individual contigs) have not diverged significantly over evolutionary time. Transposition of the BEL-like *Roo* element of *D. melanogaster* causes intraspecific variation in copy number (Papaceit et al., 2007), and associated genome-wide effects of such mutations were negatively correlated with overall fitness within replicated *D. melanogaster* lines (Pasyukova et al., 2004). Our data suggest that BEL element CNV is likely present. Identifying the

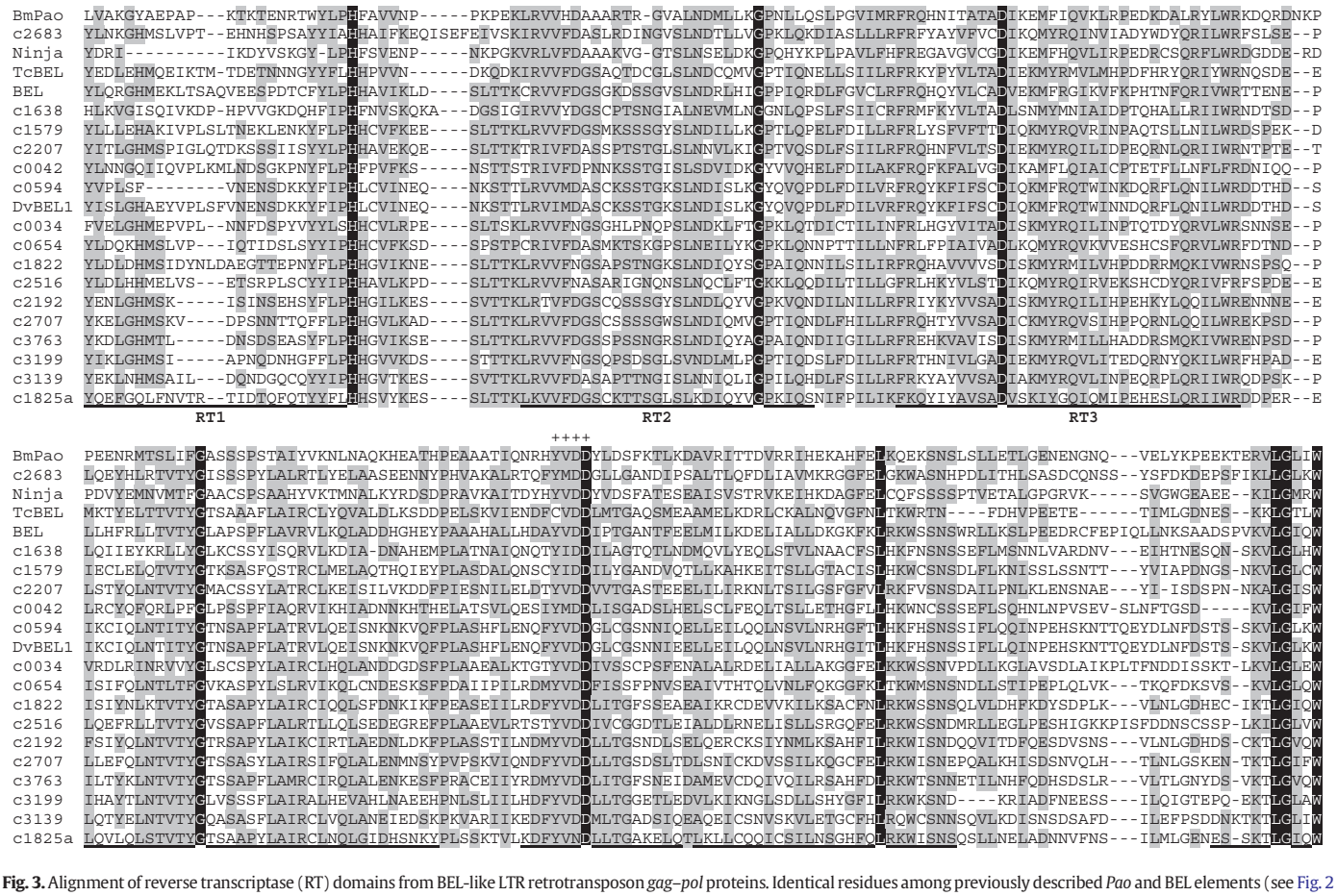

Fig. 3. Alignment of reverse transcriptase (RT) domains from BEL-like LTR retrotransposon *gag–pol* proteins. Identical residues among previously described *Pao* and BEL elements (see Fig. 2 for accession numbers) and derived *D. v. virgifera* peptides are shaded in black, and chemically similar amino acids in at least half of the sequences are shaded in gray. The RT catalytic domains are marked (+), and seven conserved RT blocks (Xiong et al., 1993) are underlined.

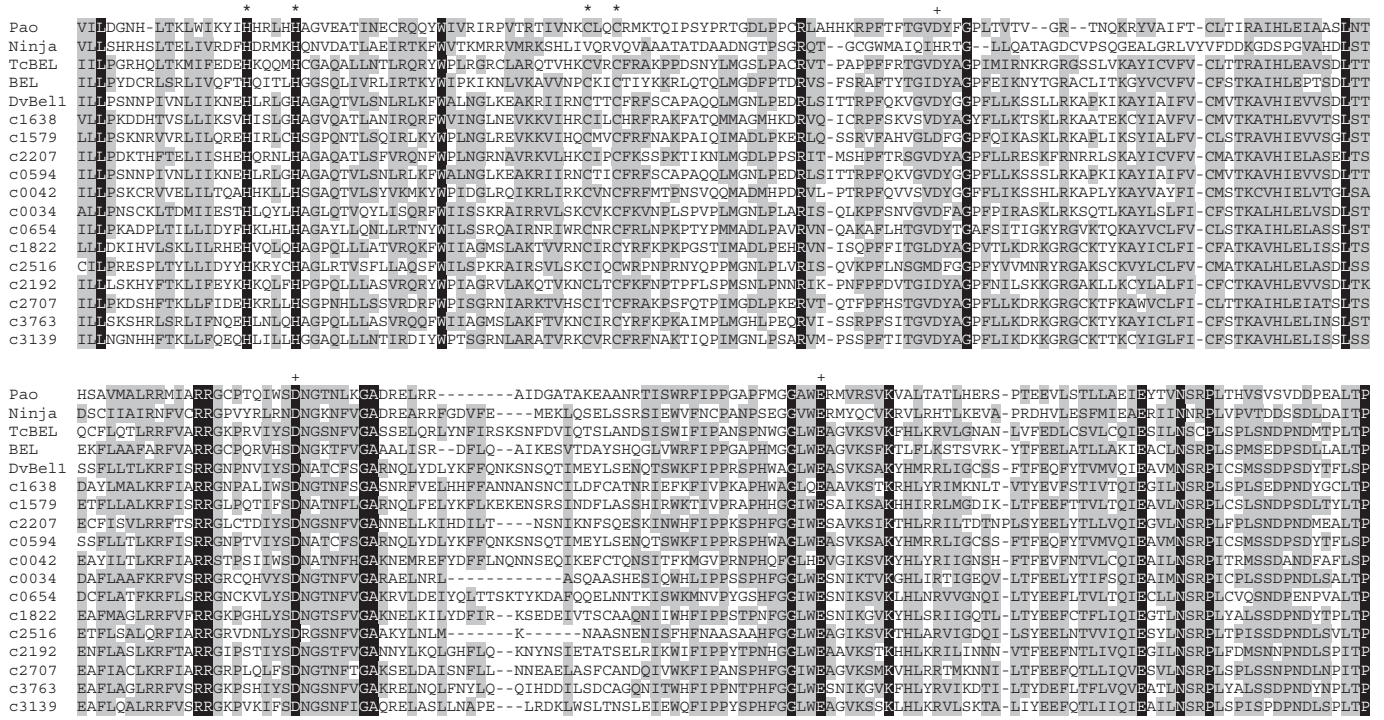

Fig. 4. Alignment of the integrase (IN) domain from BEL-like LTR retrotransposon gag–*pol* proteins. Identical residues among previously described *Pao* and BEL elements (see Fig. 2 for accession numbers) and derived *D. v. virgifera* peptides are shaded in black, and chemically similar amino acids are in gray. Residues of a conserved Cys–His box motif of a zinc-finger are indicated with (*), and those within the catalytic domain are marked (+).

effects of significant changes in BEL element number in *D. v. virgifera* remains difficult since integration points as well as any accompanying effects on phenotype remain unknown. Association of TE proliferation and different phenotypes will be the focus of future research.

3.3.2. Transcription of BEL elements

Expressed sequence tags represent gene transcripts in cDNA libraries constructed from a specific tissue, developmental stage, phenotype, or experimental treatment group (Adams et al., 1991). Published *D. v. virgifera* EST datasets from adult head (Knolhoff et al., 2010) and larval midgut tissues (Siegfried et al., 2005) were examined for evidence of BEL element transcription. Results revealed that 18 transcripts expressed in the head (EW765187, EW765400, EW766938, EW767051, EW767234, EW767949, EW768879, EW768975, EW769268, EW770478, EW770495, EW770524, EW770537, EW771419, EW771663, EW771738, EW777129) had ≥40% identity (*E*-values $\leq 1 \times 10^{-36}$) to *D. v. virgifera* of BEL LTR retrotransposon proteins. The dbEST accessions EW767051.0, EW770495.1 and EW771419.1 showed ≥98.7% nucleic acid similarity to the BEL element assembled in contig0042, suggesting that this BEL element is actively transcribed in the genome. None of the other transcripts showed high similarity to BEL-like retrotransposons assembled from ND or BRSD libraries. Since the DOC (estimated copy number) of BEL-like elements represented by contig0042 was high, this suggests that transcription may be very low from any given element, and that overall these transcripts are rare within EST datasets. This further suggests that genomic silencing may not be completely effective in suppressing the proliferation of TE integrations in the genome (Lisch and Bennetzen, 2011; Slotkin and Martienssen, 2007). Alternatively, TEs may have been domesticated after integration into genes by subsequent gain of function within gene regulatory networks (Feschotte, 2008; Volff, 2006). The role that transcribed LTR elements play in *D. v. virgifera* genome function remains unknown, although co-option into various gene pathways might explain the observed tolerance of TE proliferation in this species.

In contrast to our ability to assign the origin of 3 transcripts to a particular BEL element, the remaining 15 putative BEL LTR retrotransposon-derived ESTs showed ≤77.8 similarities to BEL elements identified in our low coverage de novo assembly. This might result from sequence differences in BEL elements in the populations from Illinois used to construct the head EST library (Knolhoff et al., 2010). A more likely explanation may reside in the incomplete assembly of BEL elements, as highlighted by the 88 partial elements annotated in the current study. The low-coverage genome assembly may be insufficient for the assembly of low copy number BEL elements. Regardless, search results from EST data resources suggest that LTR retrotransposons are transcribed and may still be actively mobile in the *D. v. virgifera* genome.

3.4. Conclusions

BEL-like LTR retrotransposons have proliferated to a high copy number within the *D. v. virgifera* genome, and their presence in EST sequence data suggests that several elements are actively transcribed and may remain mobile. Variation in BEL copy number was detected between two populations, which further suggests that these retroelements are capable of modifying genome structure and may have consequences for gene function. The role of host suppression in TE proliferation, as well

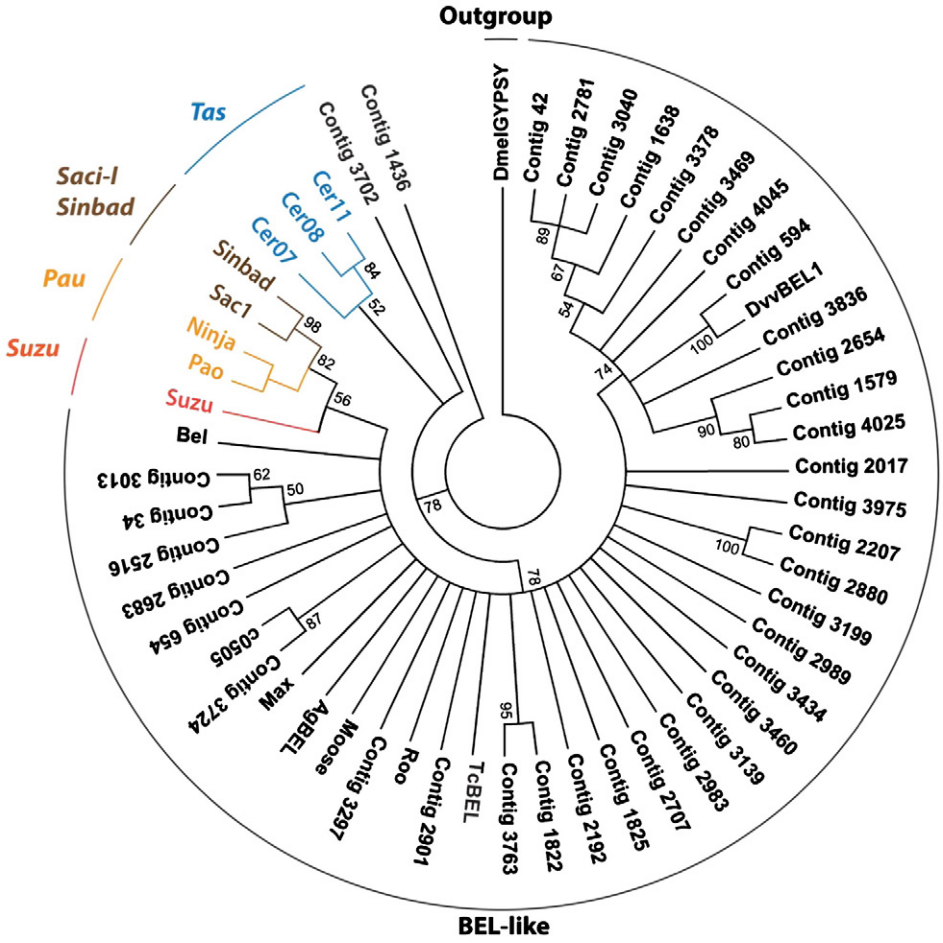

Fig. 5. Phylogenetic reconstruction of *D. v. virgifera* within the BEL/*Pao*-like group of LTR retrotransposons. Maximum-likelihood tree is based upon evolutionary distances computed by the rtREV based table and gamma shape parameter of 1.2832. Bootstrap support is indicated at each node and is given as the percentage of 1000 bootstrap pseudoreplicate trees in which the taxa were clustered. Total branch length is 28.23.

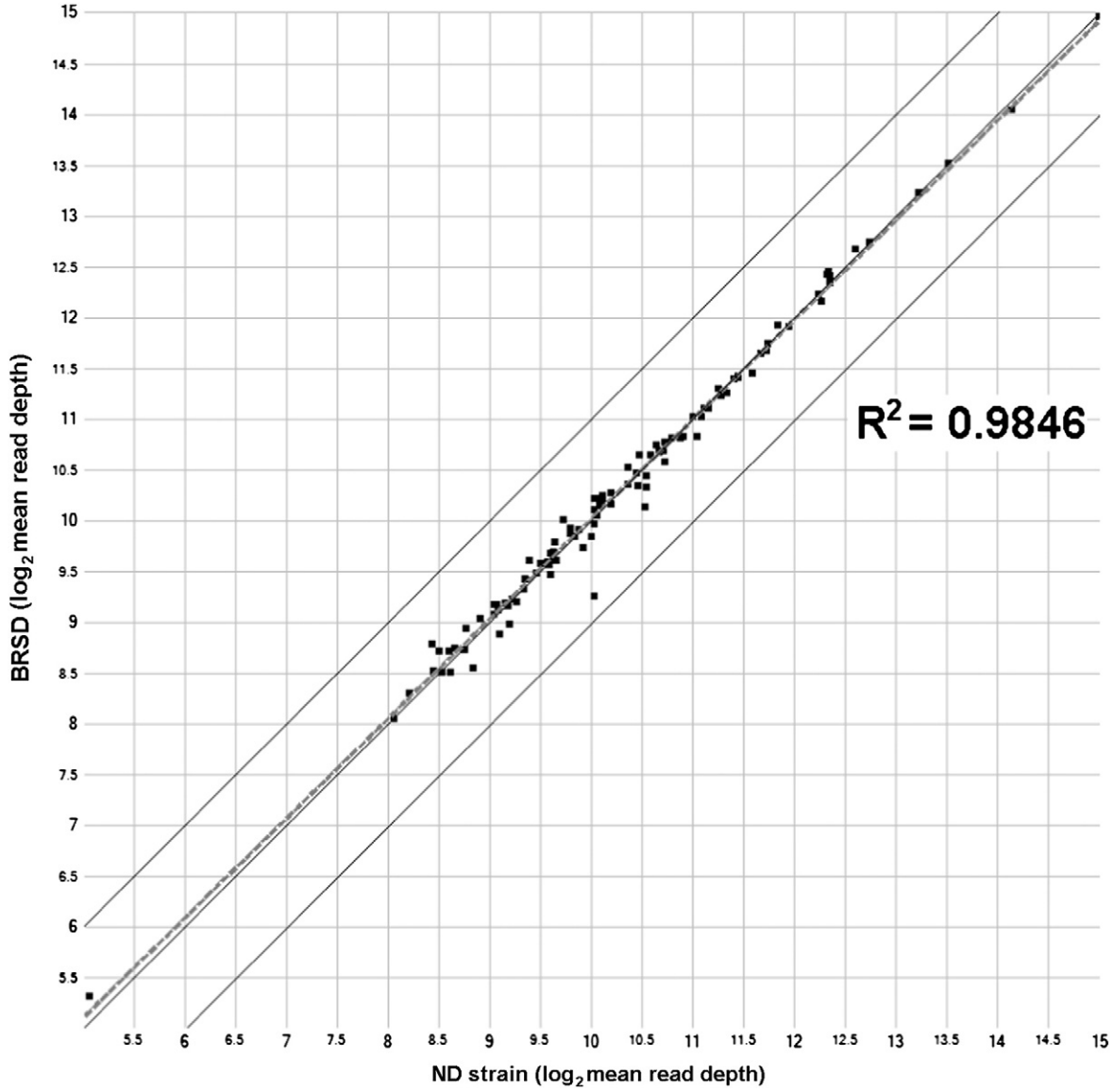

Fig. 6. Estimation of differential BEL-like LTR retrotransposon copy number between *D. v. virgifera* populations. Depth of coverage (DOC) estimates were generated from reads acquired from the non-diapausing (ND) laboratory strain and a field-collected population from Brookings, SD (BRSD). The mean depth of reads mapped to a total of 99 contigs from a low coverage de novo genome assembly. Read depth of putative BEL elements was compared between populations, where a measure of ~1.0 along the diagonal indicates elements of approximate equal copy number.

as the potential function of TEs in the creation of genetic novelty and the subsequent evolution of *D. v. virgifera* phenotypes, will be the subjects of future research efforts.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2013.09.100.

Conflict of interest

The authors have no conflicts of interest to declare.

Acknowledgments

This research was a joint contribution from the USDA, Agricultural Research Service (CRIS Project 3625-22000-017-00D), and the Iowa Agriculture and Home Economics Experiment Station, Iowa State University, Ames, IA (Project 3543). The authors would like to thank USDA-ARS Technicians, Robert Gunnarson and Chad Nielson, for performing associated laboratory protocols, and collecting and rearing of beetles. This article reports the results of research only. Mention of a proprietary product does not constitute an endorsement or a recommendation by USDA for their use.

References

Abe, H., et al., 2001. Novel *Pao*-like retrotransposons (*Kamikaze* and *Yamato*) from the silkworm species *Bombyx mori* and *B. mandarina*: common structural features of *Pao*-like elements. Mol. Genet. Genomics 265, 375–385.

Abyzov, A., et al., 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21, 974–984.

Adams, M.D., et al., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252, 1651–1656.

Aminetzach, Y.T., MacPherson, J.M., Petrov, D.A., 2005. Pesticide resistance via transposon-mediated adaptive gene truncation in *Drosophila*. Science 309, 764–767.

Bachman, N., Gelbart, M.E., Tsukiyama, T., Boeke, J.D., 2005. TFIIIB subunit Bdp1p is required for periodic integration of the Ty1 retrotransposon and targeting of Isw2p to *S. cerevisiae* tDNAs. Genes Dev. 19, 955–964.

Beeman, R.W., et al., 1996. Woot, an active *gypsy*-class retrotransposon in the flour beetle, *Tribolium castaneum*, is associated with a recent mutation. Genetics 143, 417–426.

Boeke, J.D., Stoye, J.P., 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, pp. 343–435.

Bowen, N.J., McDonald, J.F., 1999. Genomic analysis of Caenorhabditis elegans reveals ancient families of retroviral-like elements. Genome Res. 9, 924–935.
Branson, T.F., 1976. The selection of a non-diapausing strain of Diabrotica virgifera (Coleoptera: Chrysomelidae). 19, 148–154.
Chiang, H.C., 1973. Bionomics of the Northern and Western corn rootworms. Ann. Rev. Entomol. 18, 47–72.
Chung, H., et al., 2007. *Cis*-regulatory elements in the *Accord* retrotransposon results in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene Cyp6g1. Genetics 175, 1071–1077.
Coates, B.S., et al., 2012. Distribution of genes and repetitive elements in the *Diabrotica virgifera virgifera* genome estimated using BAC sequencing. J. Biomed. Biotechnol. 604076.
Copeland, C.S., Mann, V.H., Morales, M.E., Kalinna, B.H., Brindley, P.J., 2005. The *Sinbad* retrotransposon from the genome of the human blood fluke, *Schistosoma mansoni*, and the distribution of related *Pao*-like elements. BMC Evol. Biol. 5, 20.
Dimmic, M.W., Rest, J.S., Mindell, D.P., Goldstein, R.A., 2002. rtREV: an amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. J. Mol. Evol. 55, 65–73.
Eichler, E.E., Sankoff, D., 2003. Structural dynamics of eukaryotic chromosome evolution. Science 301, 793–797.
Fabrick, J.A., Mathew, L.G., Tabashnik, B.E., Li, X., 2011. Insertion of an intact CR1 retrotransposon in a cadherin gene linked with Bt resistance in the pink bollworm, *Pectinophora gossypiella*. Insect Mol. Biol. 20, 651–665.
Felder, H., et al., 1994. *Tas*, a retrotransposon from the parasitic nematode *Ascaris lumbricoides*. Gene 149, 219–225.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Feschotte, C., 2008. Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. 9, 397–405.
Feschotte, C., Pritham, E.J., 2007. DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41, 331–368.
Frame, I.G., Cutfield, J.F., Poulter, R.T.M., 2001. New BEL-like LTR-retrotransposons in *Fugu rubripes*, *Caenorhabditis elegans*, and *Drosophila melanogaster*. Gene 263, 219–230.
Gahan, L.J., Gould, F., Heckel, D.G., 2001. Identification of a gene associated with Bt resistance in *Heliothis virescens*. Science 293, 857–860.
Gassmann, A.J., Petzold-Maxwell, J.L., Keweshan, R.S., Dunbar, M.W., 2011. Field-evolved resistance to Bt maize by Western corn rootworm. PLoS One 6, e22629.
Gilbertson, R.L., Brown, W.M., Ruppel, E.G., Capinera, J.L., 1986. Association of corn stalk rot *Fusarium* spp. and Western corn rootworm beetles in Colorado. Phytopathology 76, 1309–1314.
Gonzalez, J., Karasov, T., Messer, P.W., Petrov, D.A., 2010. Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. PLoS Genet. 6, e1000905.
Gray, M.E., Sappington, T.W., Miller, N.J., Moeser, J., Bohn, M.O., 2009. Adaptation and invasiveness of Western corn rootworm: intensifying research on a worsening pest. Ann. Rev. Entomol. 54, 303–321.
Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A., Wendel, J.F., 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res. 16, 1252–1261.
Jensen, S.G., 1985. Laboratory transmission of maize chlorotic mottle virus by three species of corn rootworms. Plant Dis. 69, 864–868.
Kaer, K., Branovets, J., Hallikma, A., Nigumann, P., Speek, M., 2011. Intronic L1 retrotransposons and nested genes cause transcriptional interference by inducing intron retention, exonization and cryptic polyadenylation. PLoS ONE 6, e26099.
Kahler, A.L., Olness, A.E., Sutter, O.R., Dybing, C.D., Devine, O.J., 1985. Root damage by Western corn rootworm and nutrient content in maize. Agron. J. 77, 769–774.
Kaminker, J.S., et al., 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. Genome Biol. 3 (research 0084–0084.20).
Kidwell, M.G., Lisch, D.R., 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. Evol. Int. J. Org. Evol. 55, 1–24.
Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A., Voytas, D.F., 1998. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res. 8, 464–478.
Knolhoff, L.M., et al., 2010. Microarray analysis yields candidate markers for rotation resistance in the Western corn rootworm beetle, *Diabrotica virgifera virgifera*. Evol. Appl. 3, 17–27.
Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25.
Lisch, D., Bennetzen, J.L., 2011. Transposable element origins of epigenetic gene regulation. Curr. Opin. Plant Biol. 14, 156–161.
Marchler-Bauer, A., et al., 2012. CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res. 41, D384–D.
McClintock, B., 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. USA 36, 344–355.
Meinke, L.J., Siegfried, B.D., Wright, R.J., Chandler, L.D., 1998. Adult susceptibility of Nebraska Western corn rootworm (Coleoptera: Chrysomelidae) populations to selected insecticides. J. Econ. Entomol. 91, 594–600.
Meinke, L.J., et al., 2009. Western corn rootworm (*Diabrotica virgifera virgifera* LeConte) population dynamics. Agric. For. Entomol. 11, 29–46.
Metcalf, R.L., 1986. The ecology of insecticides and the chemical control of insects. In: Kogan, M. (Ed.), Ecological Theory and Integrated Pest Management Practice. Wiley, NY, pp. 251–297.
Miota, F., et al., 1998. Mechanisms of methyl and ethyl-parathion resistance in the Western corn rootworm (Coleoptera: Chrysomelidae). Pestic. Biochem. Physiol. 61, 39–52.
Mou, Z., Kenny, A.E., Curcio, M.J., 2006. Hos2 and Set3 promote integration of Ty1 retrotransposons at tRNA genes in *Saccharomyces cerevisiae*. Genetics 172, 2157–2167.
Ogura, K., Takechi, S., Nakayama, T., Yamamoto, M.T., 1996. Molecular structure of the transposable element *Ninja* in *Drosophila simulans*. Genes Genet. Syst. 71, 1–8.
Papaceit, M., et al., 2007. The dynamics of the *Roo* transposable element in mutation-accumulation lines and segregating populations of *Drosophila melanogaster*. Genetics 177, 511–522.
Parimi, S., Meinke, L.J., French, B.W., Chandler, L.D., Siegfried, B.D., 2006. Stability and persistence of aldrin and methyl-parathion resistance in Western corn rootworm populations (Coleoptera: Chrysomelidae). Crop. Prot. 25, 269–274.
Pasyukova, E.G., Nuzhdin, S.V., Morozova, T.V., MacKay, T.F., 2004. Accumulation of transposable elements in the genome of *Drosophila melanogaster* is associated with a decrease in fitness. J. Hered. 95, 284–290.
Patel, R.K., Jain, M., 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE 7, e30619.
Rutherford, K., et al., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16, 944–945.
SanMiguel, P., et al., 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274, 765–768.
Sappington, T.W., Siegfried, B.D., Guillemaud, T., 2006. Coordinated *Diabrotica* genetics research: accelerating progress on an urgent insect pest problem. Am. Entomol. 52, 90–97.
Siegfried, B.D., Waterfield, N., Ffrench-Constant, R.H., 2005. Expressed sequence tags from *Diabrotica virgifera virgifera* midgut identify a coleopteran cadherin and a diversity of cathepsins. Insect Mol. Biol. 14, 137–143.
Slotkin, R.K., Martienssen, R., 2007. Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8, 272–285.
Tamura, K., et al., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Tenaillon, M.I., Hufford, M.B., Gaut, B.S., Ross-Ibarra, J., 2011. Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. Genome Biol. Evol. 3, 219–229.
Teo, S.M., Pawitan, Y., Chee Seng, K., Kee Seng, C., Salim, A., 2012. Statistical challenges associated with detecting copy number variations with next-generation sequencing. Bioinformatics 28, 2711–2718.
Volff, J.N., 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 28, 913–922.
Wang, H., Coates, B.S., Chen, H., Sappington, T.W., Guillemaud, T., Siegfried, B.D., 2013. Role of a gamma-aminobutryic acid (GABA) receptor mutation in the evolution and spread of *Diabrotica virgifera virgifera* resistance to cyclodiene insecticides. Insect Mol. Biol. 22, 473–484.
Wendel, J.F., Wessler, S.R., 2000. Retrotransposon-mediated genome evolution on a local ecological scale. Proc. Natl. Acad. Sci. U. S. A. 97, 6250–6252.
Wicker, T., et al., 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8, 973–982.
Xie, C., Tammi, M.T., 2009. CNV-Seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinforma. 10, 80.
Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9, 3353–3362.
Xiong, Y., Burke, W.D., Eickbush, T.H., 1993. *Pao*, a highly divergent retrotransposable element from *Bombyx mori* containing long terminal repeats with tandem copies of the putative R region. Nucleic Acids Res. 21, 2117–2123.
Yieh, L., Kassavetis, G., Geiduschek, E.P., Sandmeyer, S.B., 2000. The *Brf* and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the *gypsy*-like element, Ty3. J. Biol. Chem. 275, 29800–29807.
Yieh, L., Hatzis, H., Kassavetis, G., Sandmeyer, S.B., 2002. Mutational analysis of the transcription factor IIIB-DNA target of Ty3 retroelement integration. J. Biol. Chem. 277, 25920–25928.
Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.
Zhao, X., Wang, H., 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nuc. Acids Res. 35, W265–W268.
Zhou, X., Scharf, M.E., Meinke, L.J., Chandler, L.D., Siegfried, B.D., 2003. Characterization of general esterases from methyl-parathion resistant and susceptible populations of Western corn rootworm (Coleoptera: Chrysomelidae). J. Econ. Entomol. 96, 1855–1863.