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/*gyspy* 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.

Download English Version:

https://daneshyari.com/en/article/5906027

Download Persian Version:

https://daneshyari.com/article/5906027

Daneshyari.com