Available online at www.sciencedirect.com
ScienceDirect
Gene 411 (2008) 27–37
GENE
www.elsevier.com/locate/gene

Multiple tandem gene duplications in a neutral lipase gene cluster in *Drosophila*

Irene Horne, Victoria S. Haritos *

CSIRO Entomology, GPO Box 1700, Canberra, ACT 2601, Australia

Received 15 October 2007; received in revised form 22 December 2007; accepted 2 January 2008
Available online 11 January 2008
Received by C. Feschotte

Abstract

We have examined a highly dynamic section of the *Drosophila melanogaster* genome which contains neutral lipase family genes that have undergone multiple tandem duplication events. We have identified the orthologous clusters, encoding between five and eight apparently functional lipases, in other *Drosophila* genomes: *yakuba*, *ananassae*, *pseudoobscura*, *virilis*, *mojavensis*, *persimilis*, *grimshawi* and *willistoni*. We examined their gene structure, duplication and pseudogene formation, and the presence of transposable elements. Based on phylogenetic comparisons, the lipase genes contained in each of the clusters fall into four distinct clades. Clades I and II have distinct evolutionary constraints to clades III and IV. Multiple gene duplications have occurred in different lineages of clades I and II while clades III and IV contain a single lipase gene from each species. Compared with lipases from other clades, clade IV genes contain an additional 3' domain of tandemly repeated sequence of varying length and composition, and a substitution in the residue adjacent to the key catalytic serine in the encoded proteins. A comparison of non-synonymous to synonymous nucleotide substitution (dN/dS) rates within each clade showed the highest rate of divergence was between paralogous lipase gene pairs suggesting selection pressure on duplicated genes. Analysis of the encoded lipase protein sequences within each species using PAML identified positively selected sites; structure homology modeling based on human pancreatic lipase indicated many of these residues formed part of the active site of the enzyme. As some of the cluster lipase genes are known to be expressed in the insect midgut and respond to changes in dietary components, we propose that the lipase cluster has undergone dynamic evolutionary changes to maximize absorption of lipid nutrients from the diet.

Keywords: Gene duplication; Adaptive radiation; Gene cluster; Lipase; *Drosophila*; Positive selection

1. Introduction

Supergene families arise in genomes as a result of multiple gene duplication events. Such duplication events are important occurrences in evolution as they contribute to diversification of species and gene duplication is a key mechanism for evolving new activities of existing genes and their products (Long and Thornton, 2001). Generally, duplicated genes are arranged in a cluster on the genome but can also be dispersed in the genome due to chromosomal rearrangement events occurring after duplication. Duplication of genes may relax selective constraint on one paralog due to redundancy; that gene can undergo independent evolution via mutations that may be deleterious and ultimately lead to loss of function, or more rarely, acquire beneficial mutations that lead to new functions (Ohta, 1989). However, under the duplication–degeneration–complementation (DDC) or subfunctionalization model, both paralogs may initially perform the ancestral gene activity but then both undergo relaxed selection and accumulate degenerative mutations such that ultimately neither gene alone substitutes for the ancestral activity (Hughes, 1994; Force et al., 1999). Occasionally, after the duplication process one of the genes acquires nucleotide substitutions or frameshift mutations that render them non-functional and are then known as pseudogenes (Balakirev and Ayala, 2003).

Abbreviations: dN/dS, non-synonymous to synonymous nucleotide substitution; LRT, likelihood ratio test; PAML, phylogenetic analysis by maximum likelihood; RMSD, root mean square deviation, D-D-C, duplication–degeneration–complementation.

* Corresponding author. Tel.: +61 2 6246 4245; fax: +61 2 6246 4000.
E-mail address: victoria.haritos@csiro.au (V.S. Haritos).

0378-1119/$ - see front matter.
doi:10.1016/j.gene.2008.01.002

Organisms require efficient metabolic processes to acquire nutrients such as carbohydrates, fats and proteins from the diet as they are critical for survival. The primary organ for absorption of these nutrients in insects, like mammals, is the intestinal tract. Different sections of the insect midgut are responsible for the absorption of components of the diet. That is, the absorption of fatty acids occurs primarily in the anterior portion of the midgut following their release from dietary lipids by the action of lipases that are secreted into the midgut lumen (Canavoso et al., 2001). A distinguishing feature of insect midgut from mammalian intestine is the gradient of lipolytic activities in different regions of the midgut leading to preferential adsorption of fatty acids at particular sites (Kirfel and Komnick, 1999; Weintraub and Tietz, 1973). As the quality and quantity of lipid nutrients in the diet can have profound effects on fecundity and survival (Simmons and Bradley, 1997) and hydrolysis of lipids is the first step in fat absorption, lipases are crucial enzymes for acquiring lipid from the diet.

The neutral lipase gene superfamily (Pf00151) is restricted to chordates and arthropods (IH, unpublished data) and their proteins catalyse the removal of fatty acid from triacylglycerols, phospholipids and other acylglycerols. Microarray analysis has identified three lipase genes from the neutral lipase family (*CG6271*, *CG6277* and *CG6283*) whose expression in larvae is strongly downregulated when placed on a sugar diet but not when starved (Zinke et al., 2002). *In situ* hybridization analysis conducted by the group showed the expression of two of the lipase genes (*CG6271* and *CG6277*) in the anterior portion of the midgut, the section of the midgut in which most fat is absorbed from the diet. The three lipase genes were shown to reside within a single gene cluster on the *D. melanogaster* genome (Zinke et al., 2002) but nothing beyond this finding has been reported. The second to last gene in the cluster (*CG6295*) is also expressed in the midgut but is regulated by the presence of dietary fatty acids and sterol regulatory protein in the midgut (Kunte et al., 2006).

Drosophila melanogaster is the first insect whose genome was sequenced in full and it has proven to be a useful tool in insect genomics and genetic studies. With the recent release of the *D. pseudoobscura* genome (Richards et al., 2005) and more genomic data becoming available for other *Drosophila* species, it is now possible to examine genome evolution, evolutionary mechanisms and functionally important yet rapidly evolving sections of the genome. *Drosophila* is useful in examining evolutionary events due to the abundant information available on evolutionary relationships between species (Powell and DeSalle, 1995; Schawaroch, 2002; Tamura et al., 2004) as well as rates of gene duplication, nucleotide substitution, transposable element movement and DNA loss (Petrov and Hartl, 1998; Kaminker et al., 2002).

We examined the evolution of a lipase gene cluster, which incorporates the four lipase genes collectively identified by Zinke et al. (2002) and Kunte et al. (2006), by comparing the orthologous genomic regions in detail for six *Drosophila* species — *D. melanogaster*, *D. yakuba*, *D. ananassae*, *D. pseudoobscura*, *D. virilis*, and *D. mojavensis* (Fig. 1A). We found each *Drosophila* genome examined has an orthologous cluster of full-length lipase genes that have undergone dynamic changes with duplications, and evidence of pseudogene formation in some, indicating that this section of the genome is changing rapidly probably in response to changes in the diet.

2. Materials and methods

2.1. Detection of lipase gene clusters in Drosophila spp. genomes

The genomes were selected for analysis based on their close relationship within the *Sophophora* to *D. melanogaster* (e.g. *D. yakuba*), or that they were of intermediate evolutionary distance between *D. melanogaster* and *D. pseudoobscura* (*D. ananassae*). *D. virilis* and *D. mojavensis* we included as members of the *Drosophila*, a distinct subgenus to the Sophophorans, and *D. mojavensis* is also of interest because of its well known dietary specialization on cacti of the Sonoran desert. The genomes of, *D. persimilis*, *D. grimshawi* and *D. willistoni* were also examined for the presence of an orthologous cluster and genes within but were not included in further analysis. At the time of investigation, the sequence of the genome of *D. erecta* was incomplete in the region of the lipase cluster and therefore no further analysis was performed.

Detection of orthologous lipase genes in the other *Drosophila* genomes was accomplished by similarity searches. We have adopted a simplified nomenclature for the *Drosophila* cluster lipase genes and gene products investigated in this study which numbers the presumed active lipases sequentially as they appear from the 5' to 3' end of the cluster. The *Drosophila* lipase genes and products are described by DxxxLy where xxx is the first three letters of the species name, y is the lipase number as shown in Table 1. The DmelL1 protein sequence (translated from *CG6271*) was used in a TBLASTN (Altschul et al., 1997) analysis against the *D. pseudoobscura*, *D. yakuba, D. persimilis*, *D. erecta*, *D. grimshawi* and *D. willistoni* (http://www.ncbi.nlm.nih.gov/BLAST/), *D. ananassae*, *D. virilis* and *D. mojavensis* genomes (http://genome.ucsc.edu/cgi-bin/hgBLAT). Genomic regions containing significant sequence similarity in multiple regions were then examined for genes using FGENESH on Softberry (http://www.softberry.com/cgi-bin/programs/gfin/fgenesh) as well as Genscan (Burge and Karlin, 1997) using parameters for *Drosophila*. Each of the gene products predicted was then subjected to a BLASTP search (Altschul et al., 1997) to confirm they possessed sequence similarity with members of the neutral lipase family (Pfam00151). Flanking regions were examined to confirm the end of each cluster. The nucleotide sequence of a single member (*DmelL1*, *DyakL1*, *DanaL2*, *DpseL3*, *DvirL4* and *DmojL4*) from each cluster was then used in a nucleotide BLAST against its respective genome to determine if gene duplications had occurred but were contained outside of the cluster region. Only genes within the clusters gave significant matches, suggesting that duplications occurred only within the cluster. Only a single nucleotide sequence was required for BLAST analysis to locate the genomic region or to examine gene duplication outside clusters due to the high sequence identity (~60%) of gene members within the cluster.