Use of quantitative real-time polymerase chain reaction to estimate the size of the house-fly *Musca domestica* genome

J. Gao and J. G. Scott  
*Department of Entomology, Comstock Hall, Cornell University, Ithaca, NY, USA*

Abstract

House-flies, *Musca domestica*, are carriers of more than 100 devastating diseases that have severe consequences for human and animal health. A key bottleneck to progress in controlling the devastating human diseases transmitted by house-flies is lack of knowledge of the basic molecular biology of this species. However, before sequencing of the house-fly genome can be seriously considered it is important to know the size of the genome. In this paper, we used quantitative real-time polymerase chain reaction to calculate genome size of the house-fly in side-by-side experiments with *Drosophila melanogaster* (known genome size of 180 Mb). Our results indicate the size of the house-fly genome is 295 ± 10 Mb and that of *D. melanogaster* is 184 Mb. Thus, the house-fly genome is only about 1.6-fold larger than the genome of *D. melanogaster*. This indicates that the size of the house-fly genome makes it an excellent candidate for whole genome sequencing and that quantitative real-time polymerase chain reaction is an accurate method for the estimation of the size of insect genomes.

Keywords: Insecta, genome size, quantitative real-time polymerase chain reaction, *Musca domestica*, *Drosophila melanogaster*.

Introduction

House-flies, *Musca domestica* (L.) are carriers of dozens of devastating diseases that have severe consequences for human and animal health (Scott & Lettig, 1962; Greenberg, 1965; Keiding, 1986), including the transmission of antibiotic-resistant bacteria in hospitals (Boulesteix *et al.*, 2005; Rahuma *et al.*, 2005). Despite the fact that it is a passive vector, a key bottleneck to progress in controlling the devastating human diseases transmitted by house-flies is lack of knowledge of the basic molecular biology of this species. Sequencing of the house-fly genome would provide important inroads to the discovery of novel target sites for house-fly control, understanding of the immune response in this dung-living fly, rapid elucidation of insecticide resistance genes and understanding of numerous aspects of the basic biology of this insect pest. However, it is critically important to have an accurate estimation of the size of the house-fly genome, in order to ascertain if sequencing of the whole genome is feasible. In addition, information on genome sizes could reveal insights into evolutionary processes.

Recently, it was demonstrated that quantitative real-time–polymerase chain reaction (qRT–PCR) was a reliable method for the estimation of genome sizes (Wilhelm *et al.*, 2003). Therefore, we used this technique to determine the size of the house-fly genome. To check the validity of this method for genome size estimation we also used qRT–PCR to measure the size of the *Drosophila melanogaster* genome in side-by-side experiments. Our results indicate the size of the house-fly genome is 295 ± 10 Mb, and is sufficiently small to warrant strong consideration for a whole genome sequencing effort.

Results and discussion

The genome sizes of house-fly and *D. melanogaster*, as determined by real-time PCR, are shown in Table 1. The genome size of house-fly was determined to be 295 ± 10 Mb (mean ± SD) (range: 280–316 Mb). The genome size of *D. melanogaster* was 184 Mb, which is very similar to published values (~ 180 Mb) (range: 179–201 Mb) (Adams *et al.*, 2000). The only previous estimate of the house-fly genome size was made using spectrophotometric analysis of sperm, and suggested the size of the house-fly genome was 950 Mb (Bier & Müller, 1969). It is unclear why this
value is so much higher than what we determined by qRT–PCR. However, because our value for the size of the D. melanogaster genome was very close to the predicted value, we have high confidence in our value for the size of the house-fly genome.

Table 1 shows the genome size and chromosome number of insects whose genomes have been (Adams et al., 2000; Holt et al., 2002), or are currently being sequenced. As is common (Knight, 2002), there is no correlation between genome size and number of chromosomes present. The differences in genome size between the species in Table 1 reflects different rates of DNA loss or gain between species, although the factors responsible for the different size genomes reflects different rates of DNA loss or gain between species, although the factors responsible for the different size genomes is not well understood (Knight, 2002). The genome size of the house-fly is slightly larger than Anopheles gambiae (Holt et al., 2002; Zdobnov et al., 2002), Tribolium castaneum or Apis mellifera (honey bee), and is considerably smaller than Bombyx mori, all of which have had their genomes sequenced. Thus, the size of the house-fly genome is sufficiently small to warrant strong consideration for a whole genome sequencing effort. Our results also confirm the original study (Wilhelm et al., 2003) that demonstrated the utility of qRT–PCR for the determination of genome sizes.

**Experimental procedures**

**Fly strains**

The CS strain (wild-type) of house-fly (Hamm et al., 2002) is maintained in our laboratory. The Canton-S strain (wild-type) of D. melanogaster (from Dr John Ewer, Cornell University, Ithaca, New York, USA) was used side-by-side with house-flies to confirm the accuracy of our genome size estimates.

**DNA isolation**

Genomic DNA was isolated from a single male CS house-fly, and from 30 Canton-S D. melanogaster, using the protocol developed by J. Rehm, Berkeley Drosophila Genome Project (http://www.fruitfly.org). Genomic DNA solution was run on 0.8% agarose gel with 1× TAE. Fragments from 15 to 50 kb were recovered using a QIAEX II gel extraction kit (Qiagen, Valencia, CA, USA). The purity was regarded as acceptable when the ratio A260/A280 was between 1.8 and 2.0. One A260 unit was assumed to correspond to a DNA concentration of 50 μg/ml.

**Amplification of CYP6D1 and Cyp6a2**

CYP6D1 is a single-copy gene (Tomita et al., 1995) and was used as the target DNA fragment in the genome of house-fly. Cyp6a2 is a single-copy gene (Waters et al., 1992; Adams et al., 2000) and was used as the target DNA fragment in the genome of D. melanogaster. PCR products for CYP6D1 or Cyp6a2 were generated (primers are given in Table 2) in an iCycler™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) with the respective genomic DNA as the template. The PCR reaction contained 5.4 μM each forward and reverse primer, 1× PCR ReddyMix™ Master Mix (ABgene, Rochester, NY, USA) and 20–50 ng genomic template DNA were used in a total volume of 25 μl. The PCR protocol consisted of an initial denaturation step of 95 °C for 3 min, followed by 35 amplification cycles with 30 s at 95 °C, 30 s at 55 °C and 1 min at 72 °C and a final extension step of 72 °C for 10 min. PCR products were detected by running on a 1.2% agarose gel, and were purified using QIAEX II Gel Extraction Kit (Qiagen) according to the manufacturer’s instructions. PCR fragments were cloned into pCR® 2.1-TOPO® vector using TOPO® TA cloning kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer’s instructions. Plasmid DNA was isolated using a QIAprep spin mini-prep kit (Qiagen) and was eluted in double distilled water. Quality and concentration of all purified standard DNA samples were determined by ultraviolet spectroscopy as described above. PCR products and plasmid DNA having the expected insert were confirmed by sequencing at the Cornell Biotechnology Resource Center.

**Quantitative real-time polymerase chain reaction**

The amount of target sequence in genomic DNA was quantified using real-time PCR. The target sequences were amplified with

<table>
<thead>
<tr>
<th>Target</th>
<th>Name</th>
<th>Sequence 5′→3′</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyp6a2</td>
<td>CYP6A2F1</td>
<td>TGCCCTCGGAGTTAGTGA</td>
</tr>
<tr>
<td>CYP6A2F2</td>
<td>TGGTTTCGCAACATGAGG</td>
<td></td>
</tr>
<tr>
<td>CYP6A2R1</td>
<td>TGGTTTGGGTCCCTC</td>
<td></td>
</tr>
<tr>
<td>CYP6A2R2</td>
<td>CCGCTCTTCAGAACGCT</td>
<td></td>
</tr>
<tr>
<td>CYP6D1</td>
<td>CYP6D1F1</td>
<td>GAGTGGTATAGGGCACCACATA</td>
</tr>
<tr>
<td>CYP6D1F5</td>
<td>CGCATCATGGTCCGGTATAT</td>
<td></td>
</tr>
<tr>
<td>CYP6D1R5</td>
<td>CCAATGGCGTCCCTTGGT</td>
<td></td>
</tr>
<tr>
<td>CYP6D1R5</td>
<td>TCACTCAAGACAGAAGATTGT</td>
<td></td>
</tr>
</tbody>
</table>

© 2006 The Authors

the primer pairs CYP6D1F1/R1 for house-fly, and CYP6A2F1/R1 for D. melanogaster. Real-time PCR reactions were carried out using an ABI Prism 7900 Sequence Detector (Applied Biosystems, Foster City, CA) in MicroAmp optical 384-well reaction plates with optical covers, according to manufacturer's instructions. SYBR Green PCR Core Reagent (Applied Biosystems) was used for the PCR reactions. The template DNA was either PCR generated DNA or genomic DNA (sample). Each run contained a series of standards (standard DNA as template with seven 10-fold dilutions) and the unknowns (genomic DNA as template) for one target (Cyp6a2 or CYP6D1). A standard curve (of the C\textsubscript{T} values vs. the amount of DNA) was used to calculate the concentrations of the unknowns based on the C\textsubscript{T} values according to the manufacturer's instructions (User Bulletin no. 2, ABI Prism 7700 Sequence Detection System). For each experiment there were six determinations per sample. The entire experiment was replicated 10 times.

\textit{Calculation of the genome size}

The genome size (GS, or the number of base pairs per genome) was calculated using the formula GS = C × N\textsubscript{A} / M\textsubscript{BP} (Wilhelm et al., 2003). In this formula C = mass of template DNA × copy number of the target sequence, N\textsubscript{A} is Avogadro's number and M\textsubscript{BP} is the average mass of a base pair (660 g/mol).

\textbf{Acknowledgements}

We thank J. Ewer for providing the D. melanogaster and C. Leichter for technical assistance. This work was supported by the Daljit S. and Elaine Sarkaria Professorship and Hatch Project 414.

\textbf{References}


