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

Table 1. The genome size of *Musca domestica* and *Drosophila melanogaster* determined in this study, compared to 11 representative insect species that have had their genomes sequenced (or are in the process of being sequenced)

Species	Genome size (Mb)	Chromosome no. (n)
*Musca domestica**	295 ± 10	6
*Drosophila melanogaster**	184 ± 9	4
Aedes aegypti	800	3
Anopheles gambiae†	278	3
Apis mellifera	200	16
Bombyx mori	530	28
Culex pipiens	540	3
Drosophila melanogaster‡	180	4
D. yakuba	180	4
D. persimilis	150	4
D. simulans	150	4
D. pseudoobscura	125	4
Rhodnius prolixus	670	11
Tribolium castaneum	200	10

*Determined in this paper.
†Holt *et al.* (2002).
‡Adams *et al.* (2000).
Other values from NCBI Genome Project database.

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 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.*, 2005) 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 $1 \times$ TAE. Fragments from 15 to 50 kb were recovered using a QIAEX ∏ gel extraction kit (Qiagen, Valencia, CA, USA). The elution was carried out with double-distilled water. Quality and concentration of all DNA samples were determined by recording a ultraviolet absorption spectrum between 220 and 320 nm with a DU® 640 Spectrophometer (Beckman, Columbia, MD, USA). The purity was regarded as acceptable when the ratio $A260/A_{280}$ was between 1.8 and 2.0. One A_{260} 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 0.4 μM each forward and reverse primer, $1 \times$ 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 ∏ 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 2. Primers used in this study. All primers were from Integrated DNA Technologies Inc. (Coralville, IA, USA)

Target	Name	Sequence 5′→3′
Cyp6a2	CYP6A2F1	TGCCTTCGGCATTGAGTGTA
	CYP6A2F2	TGATTTCCGCACCATGGG
	CYP6A2R1	TCGTTGAGGGCCTTCCG
	CYP6A2R2	TGCCCTTCTCAATCACAAGCT
CYP6D1	CYP6D1F1	GAATCGGATAAGGCCACCATA
	CYP6D1F5	CGACATCATTGGTTCGGTGATAT
	CYP6D1R1	CCAAATGGCCGTTCCTTGT
	CYP6D1R5	TCACTCAAGACACGAAATTCATTGT

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_T values vs. the amount of DNA) was used to calculate the concentrations of the unknowns based on the C_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.

Calculation of the genome size

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

Acknowledgements

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.

Bier, K. and Müller, W. (1969) DNA-Messungen bei Insekten und eine Hypothese über retardierte Evolution und besonderen DNA-Reichtum in Tierreich. *Biol Zentralbl* **88**: 425–449.

Boulesteix-.G., Le-Dantec-.P., Chevalier-.B., Dieng-.A., Niang-.B. and Diatta-.B. (2005) Role de *Musca domestica* dans la transmission des bacteries multiresistantes dans un service de reanimation en Afrique subsaharienne. *Ann Fr Anesth Reanim* **24**: 361–365.

Greenberg, B. (1965) Flies and disease. *Sci Am* **213**: 92–99.

Hamm, R., Shono, T. and Scott, J.G. (2005) A cline in frequency of autosomal males is not associated with insecticide resistance in house fly (Diptera: Muscidae). *J Econ Entomol* **98**: 171–176.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R. *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129–149.

Keiding, J. (1986) *The House Fly: Biology and Control*. WHO *Vector Control Series* 63, World Health Organization, Geneva, Switzerland.

Knight, J. (2002) All genomes great and small. *Nature* **417**: 374–376.

Rahuma, N., Ghenghesh, K.S., Ben-Aissa, R. and Elamaari, A. (2005) Carriage by the housefly (*Musca domestica*) of multiple-antibiotic-resistant bacteria that are potentially pathogenic to humans, in hospital and other urban environments in Misurata, Libya. *Ann Trop Med Parasitol* **99**: 795–802.

Scott, H.G. and Lettig, K.S. (1962) *Flies of Public Health Importance and Their Control*. US Government Printing Office, Washington, DC.

Tomita, T., Liu, N., Smith, F.F., Sridhar, P. and Scott, J.G. (1995) Molecular mechanisms involved in increased expression of a cytochrome P450 responsible for pyrethroid resistance in the housefly, *Musca domestica*. *Insect Mol Biol* **4**: 135–140.

Waters, L.C., Zelhof, A.C., Shaw, B.J. and Ch'ang, L.Y. (1992) Possible involvement of the long terminal repeat of transposable element 17.6 in regulating expression of an insecticide resistance-associated P450 gene in *Drosophila*. *Proc Natl Acad Sci USA* **89**: 4855–4859.

Wilhelm, J., Pingoud, A. and Hahn, M. (2003) Real-time PCR-based method for the estimation of genome sizes. *Nucleic Acids Res* **31**: e56.

Zdobnov, E.M., von-Mering, C., Letunic, I., Lorrents, D., Suyama, M., Copley, R.R. *et al.* (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**: 149–159.