SHORT COMMUNICATION

One-way sequencing of multiple amplicons from tandem repetitive mitochondrial DNA control region

JIAWU XU & DINA M. FONSECA

Center for Vector Biology, Rutgers University, New Brunswick, NJ 08901, USA

(Received 25 May 2011; revised 28 September 2011; accepted 7 October 2011)

Abstract
Repetitive DNA sequences not only exist abundantly in eukaryotic nuclear genomes, but also occur as tandem repeats in many animal mitochondrial DNA (mtDNA) control regions. Due to concerted evolution, these repetitive sequences are highly similar or even identical within a genome. When long repetitive regions are the targets of amplification for the purpose of sequencing, multiple amplicons may result if one primer has to be located inside the repeats. Here, we show that, without separating these amplicons by gel purification or cloning, directly sequencing the mitochondrial repeats with the primer outside repetitive region is feasible and efficient. We exemplify it by sequencing the mtDNA control region of the mosquito Aedes albopictus, which harbors typical large tandem DNA repeats. This one-way sequencing strategy is optimal for population surveys.

Keywords: Aedes albopictus, concerted evolution, repetitive DNA, sequencing strategy

Introduction
Repetitive DNA sequences not only constitute a large proportion of eukaryotic nuclear genomes (Britten and Kohne 1968), but also occur in mitochondrial DNA (mtDNA) control regions of many animals (Zhang and Hewitt 1997; Lunt et al. 1998), such as fruit flies (Drosophila), mosquitoes (Aedes), and rabbits (Oryctolagus). This noncoding region contains the origin of mtDNA replication and promoter for transcription initiation, thus playing a functionally important role in mitochondrial genome. In insects, the control region is also called “A + T-rich region” due to its extremely high A + T nucleotide content (> 85%). Because this region experiences high rate of evolution in both nucleotide substitutions and insertion/deletion (indel) events, it is generally the most variable area in mtDNA, therefore a potential marker for intraspecific genetic surveys (Simon et al. 1994), although this is not always true for insects (Zhang and Hewitt 1997).

The size of the control region varies considerably in different taxa, and the length variations among closely related taxa mainly result from tandem repetitions (both the size and copy number). For example, due to the variable copy numbers of a 181 bp repetitive unit (three to six copies, near the 12S gene) and a 43 bp repetitive unit (three to four copies, near the tRNA1le gene), the control region of the yellow fever mosquito Aedes aegypti varies from 1480 to 2070 bp in length (Rondan Dueñas et al. 2006). While in its relative, the Asian tiger mosquito Aedes albopictus, there exist five copies of a ~190 bp repeat (Repeat I 1-5, namely RI1-RI5, near the 12S gene; Figure 1B) and four copies of a ~42 bp repeat (near the tRNA1le gene) in this region, with a total length of 1772 bp (GenBank AY072044). The copy number of a repetitive unit may also vary within a mitochondrial genome, resulting in mtDNA length heteroplasy. Tandem repetitions often follow a process of concerted evolution, i.e. sequence similarity of the
repetitive units is greater within than between species or lineages (Zhang and Hewitt 1997). Both high A+T content and tandem repeats challenge the amplification and sequencing of this region, limiting its applications to population genetic studies. In some cases like the above two mosquitoes, due to the difficulty of long amplification (see Material and methods), one polymerase chain reaction (PCR) primer has to be located in the repetitive DNA, which will produce more than one amplicon due to multiple binding sites.

For multiple amplicons from one reaction to be sequenced, they usually need to be isolated either by gel purification or by cloning. However, in the case of mitochondrial repetitive DNA, all amplicons actually overlap at one end but extend differently at the other end, because there is a unique binding site for the primer outside the repetitive DNA and diverse binding sites for the primer in the repetitive DNA (Figure 1B). In principle, the multiple amplicons could be directly sequenced with the former primer, thus avoiding the laborious gel purification or cloning procedures. In this study, we test the feasibility and efficiency of this one-way sequencing strategy for sequencing the mitochondrial control region of the mosquito _A. albopictus_, aiming at exploring the variability and usefulness of this region for population genetic analyses. For comparison, the complete mitochondrial cytochrome c oxidase I (COI) gene was also sequenced.

**Materials and methods**

The mosquitoes used in this study were collected from Kyoto, Japan (N = 5), Mercer County (N = 5) and Monmouth County (N = 5), NJ, the USA, and deposited at Rutgers Center for Vector Biology. Genomic DNA from individual mosquitoes was extracted with DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA) following the manufacturer’s protocol. We have attempted to amplify the entire control region using primers in the flanking genes but failed, even having met the specific requirements for amplifying long A+T-rich DNA (Rondan Dueñas et al. 1999). We then amplified partial control region (adjacent to the 12S gene) with a modified primer SR-J-14612 (5'-AGGGTATCTATCCTAGTTTA-3', priming in the 12S gene; Simon et al. 1994) and a newly designed primer ATrichR (5'-AGGGGTAA-TTTAATAAGTTT-3', priming in the tandem repetitions of control region). PCR amplification was performed in 20 μl reactions containing 1 × reaction buffer, 2 mM of MgCl₂, 0.33 mM of each dATP–dTTP, and 0.1 mM of each dCTP–dGTP (Rondan Dueñas et al. 1999), 0.2 μM of each primer, 0.2 μg/ml of bovine serum albumin (BSA), 1 unit of Taq DNA Polymerase (Applied Biosystems, Foster City, CA, USA), and 10–50 ng of genomic DNA. The amplification program consisted of 94°C for 180 s, 35 cycles of 94°C for 20 s, 55°C for 45 s, 70°C for 120 s, and a final extension at 70°C for 300 s. The complete COI gene was amplified with primers COIF (5’-TTTACAATTTATCGCCTAAACTTC-3’) and COIR (5’- CATTGCACTAATCTGCCATA-3’), located in the flanking genes tRNA^{Tyr} and tRNA^{Leu}, respectively, and designed from consensus sequence of mosquito complete mitochondrial genomes available in GenBank. The reaction composition and PCR program are the same as those for the control region except that the concentration of each dNTP is 0.2 mM and extension temperature is 72°C. The PCR products were treated with ExoSAP-IT® (USB Corporation, Cleveland, OH, USA) and cycle sequenced for analyzing on ABI 3100 automated sequencer. The control region was cycle sequenced with a single primer SR-J-14612, and COI was
sequenced with both PCR primers and DNA barcoding primers (LCO1490 and HCO2198) of Folmer et al. (1994). Sequences were edited with Sequencher 4.10 (Gene Codes). Sequence alignment, nucleotide composition, and genetic distance were analyzed with MEGA 4 (Tamura et al. 2007). The sequences of haplotypes have been deposited in GenBank (JQ004524–JQ004529).

Results and discussion

The amplification of the control region generated four clear amplicons roughly 500, 700, 900, and 1100 bp in length (Figure 1A), corresponding to the four expected bands of 518, 706, 896, and 1085 bp spanning the 12S gene and control region, with primer ATrichR targeting at repetitions RI1–RI4, respectively (Figure 1B). In addition, a weak band around 1300 bp, matching the amplicon when ATrichR is priming at repetition RI5, can also be observed in some reactions. As expected, the COI reaction produced a clear band about 1600 bp (not shown). After gel electrophoresis on ABI 3100 sequencer, clean sequences of up to 800 bp were obtained for the control region, with exception of the ATrichR priming and starting regions (~16 bp), which were somewhat ambiguous (Figure 2). This primer matches well with each of the five repetitions, although there is one mismatch site (in the middle of the primer) between the primer and repetitions RI1–4 (GenBank AY072044).

After excluding the 12S gene (338 bp, no variation among all sequences) and ambiguous primer regions, we obtained a 495 bp sequence from the control region, with no indels observed. In 15 mosquitoes (or sequences), we identified four variable sites defining four haplotypes (H1-H4), all of which are present in Kyoto population, whereas all New Jersey mosquitoes share a single haplotype (H3). The average distance (uncorrected p distance) between haplotypes is 0.004 (0.002–0.008). The A + T content for this segment is 91.1%, similar to that of the whole control region (91.6%; GenBank AY072044). In the 1537 bp COI gene, we detected two variable sites (one on the first codon position and the other on the third codon position) defining two haplotypes, one in Kyoto population (H1) and the other in New Jersey population (H2), with a between haplotypes distance of 0.001, which is one-fourth of that of the control region. The A + T content for this gene is 69.9%, and that for the first, second, and third codon positions is 59.2%, 59%, and 91.2%, respectively. If only the third codon positions are considered, the between haplotype distance of 0.002 is still lower than that of the control region, underlying the usefulness of the latter as a population genetic marker.

Different strategies have been devised to sequence the repetitive DNA. For sequencing the mitochondrial control regions of leaf beetles (Chrysomela and Goniocenta), which contain 12–17 copies of 107–159 bp tandem repetitions, Mardulyn et al. (2003) employed a cloning-transposon insertion strategy to avoid priming at highly similar repetitive units and thus be free from multiple amplicons. In order to sequence a 964 bp repetitive region (five repeat units R1–R5, R1 = 180 bp, R2 = R3 = R4 = 249 bp, R5 = 37 bp) in a nuclear gene of Schistosoma mansoni tegumental antigen, Abath and Holder (1995) used a strategy based on partial restriction enzyme cleavage, as well as cloning and gel purification. These methods are

Figure 2. Chromatogram of a sequence from one-way sequencing with primer SR-J-14612. The line shows the boundary between the 12S gene and the control region. Arrows indicate the priming sites or regions.
labor-intensive and expensive, which limit their applications to large sample data that are typically required for population surveys.

The above one-way sequencing strategy is optimal for population analyses, if repetitive regions are the sequencing targets. It avoids the laborious and costly gel purification and cloning procedures, thus also overcoming the issues of cloning artifacts or PCR errors. Moreover, sequences are easily proofread and aligned due to their high similarity within populations. Although this sequencing method is feasible and efficient for sequencing tandem repetitive mtDNA, its applications to nuclear repetitive regions are not explored. Expectably, both the allelic variation (especially indels) at one locus and potentially multiple loci variation will largely limit its usage in nuclear genomes.

Acknowledgements

This work was funded by Cooperative Agreement between USDA-ARS and Rutgers University (USDA-ARS-58-6615-8-105), which is entitled “Area-wide Management of the Asian Tiger Mosquito.”

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References