FULL LENGTH RESEARCH PAPER

Isolation and characterization of the RanGAP gene in the mosquito Aedes aegypti

SUNG-JAE CHA1, NEIL LOBO2, BECKY DEBRUYN2, & DAVID W. SEVERSON2

1Department of Molecular Microbiology and Immunology, Johns Hopkins School of Public Health, Malaria Research Institute, Baltimore, MD, USA, and 2Department of Biological Sciences, Center for Tropical Disease Research and Training, University of Notre Dame, Notre Dame, IN, USA

Abstract
A duplicated 3’-truncated version of RanGAP was previously identified as Segregation distorter (Sd), the meiotic drive gene in Drosophila melanogaster. Here we report the cloning and characterization of the complete gene sequence for the RanGAP homolog from the mosquito Aedes aegypti. The 1995 bp cDNA sequence consists of a 113 bp 5’UTR and 130 bp 3’UTR, and encodes a 583 amino acid protein with high sequence identity with RanGAP homologues of several species. A 20,125 bp genomic DNA sequence contains the complete RanGAP gene, consisting of three exons and two introns. Intron 2 comprises 18,082 bp and contains multiple repetitive elements as well as putative coding regions. The RanGAP locus was mapped to the q-arm of chromosome 2. Because the meiotic drive gene (M^D) in A. aegypti was previously shown to be tightly linked with the sex determining locus on chromosome 1, RanGAP is likely not the M^D gene.

Keywords: Segregation distorter, meiotic drive, sex ratio distortion, linkage mapping, genomic organization, repetitive elements

Introduction
Ran is a GTP-binding protein that is a member of the Ras-related GTP-binding protein superfamily (Drivas et al. 1990). It is involved in cell cycle control, mitotic spindle formation, post-mitotic nuclear envelope assembly, and is an important component of the nuclear transport machinery (Sazer and Dasso 2000, Dasso 2001, Moore 2001). It is an abundant protein localized mainly in the nucleus, and was first cloned from a human tetratocarcinoma cDNA library (Drivas et al. 1990). Ran has GTPase activities and cycles between a GTP and a GDP bound form. It tightly binds guanine nucleotides and slowly hydrolyses GTP (Klebe et al. 1995). The products of two major cofactors of Ran regulate its GTPase activities. These include RanGEF (Ran guanine nucleotide exchange factor; also known as RCC1) (Bischoff and Pousttangle 1991) and Ran-GTPase activating protein (RanGAP) (Bischoff et al. 1994). RanGAP is localized inside the nucleus, whereas RanGAP accumulates in the cytosol. Cytoplasmic localization of RanGAP and the nuclear localization of RanGEF establish a concentration gradient of Ran-GTP across the nuclear envelope that is critical for proper Ran-mediated nuclear transport (Kusano et al. 2002).

A mutated RanGAP gene has been shown to be a primary component of the well-characterized meiotic drive system (Segregation distorter or SD) in Drosophila melanogaster (Merrill et al. 1999). The SD complex has been resolved into three major components: the distorter gene (Sd), an enhancer of distortion (E(Sd)) and the responder locus (Rsp). These reflect strong linkage disequilibrium due to suppression of recombination within the peri-centromeric region of D. melanogaster chromosome two (Palopoli and Wu 1996). Sd individuals carry the wild type RanGAP (6.5 kb EcoRI fragment) and a 3’ truncated tandem duplication (5 kb EcoRI fragment) (Merrill et al. 1999). The duplicated copy

Correspondence: D. W. Severson, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556-5645, USA. Tel: 1 574 631 3826. Fax: 1 574 631 7413. E-mail: severson.1@nd.edu

ISSN 1042-5179 print/ISSN 1029-2365 online © 2006 Taylor & Francis
DOI: 10.1080/10425170600805540
is expressed and encodes a truncated RanGAP missing 234 amino acids at the C-terminus, yet retains essentially normal enzymatic activity (Kusano et al. 2001). Transgenic studies showed that the 3' truncated version of RanGAP is the effector gene for the SD system (Merrill et al. 1999). Males heterozygous for an Sd carrying chromosome and a sensitive Rsp carrying chromosome show sperm dysfunction. Dysfunction is evidenced as the failure of proper chromatin condensation in sensitive Rsp bearing spermatids, leading to subsequent defects in spermatid elongation and maturation (Tokuyasu et al. 1977). Kusano et al. (2001) suggested that the 3' truncated RanGAP is missing one of the two normal nuclear export sequences (NES), thus disrupting its exportation from the nucleus. The truncated RanGAP is mislocalized to the nucleus, which increases Ran-GTP hydrolysis and disrupts normal Ran-dependent functions. The aberrant nuclear accumulation of even wild type RanGAP results in segregation distortion, although the exact mechanisms for distortion are unclear (Kusano et al. 2002).

Meiotic drive has previously been reported in two mosquito species, Aedes aegypti and Culex pipiens (Sweeney and Barr 1978, Hickey and Craig 1996a,b). The meiotic drive gene (M) in A. aegypti is tightly linked with the male-determining allele (M) at the sex locus on chromosome 1, and its product acts to cause fragmentation of female determining gametes (m) carrying a susceptible responder allele (m'). Because of the potential for effecting population replacement and similarities to the SD system of D. melanogaster (Wood and Newton 1991; Lyttle 1993), we have initiated efforts to identify and characterize the A. aegypti RanGAP homolog. That is, the general inability to control mosquito-borne disease transmission, mosquito population replacement strategies using genetic manipulation techniques to produce incompetent mosquito vectors are actively being evaluated (Adelman et al. 2002, Moreira et al. 2002). Population replacement with mosquitoes using genetic manipulation techniques to produce incompetent mosquito vectors are actively being evaluated (Adelman et al. 2002, Moreira et al. 2002). Population replacement with mosquitoes using genetic manipulation techniques to produce incompetent mosquito vectors are actively being evaluated (Adelman et al. 2002, Moreira et al. 2002).

### Materials and methods

#### Degenerate primer design

We performed a multiple sequence alignment of RanGAP homologues from six species (Anopheles gambiae: GenBank accession no. XM317204, D. melanogaster: GenBank accession no. AP143860, Homo sapiens: GenBank accession no. BC041396, Mus musculus: GenBank accession no. AAH14855, Xenopus laevis: GenBank accession no. O13066, Saccharomyces cerevisiae: GenBank accession no. CAA90206) to find conserved sequence domains for designing degenerate primers. The most prominent features of the RanGAP proteins are the leucine-rich repeats (LRR) that constitute the major part of the protein sequence. LRRs have been found in proteins with different functions and intracellular localizations, and participate in protein–protein interactions (Kobe and Deisenhofer 1994). We selected two LRR regions for designing degenerate primers flanking the 5’- and 3’-ends (Aa-RanGAP_F 5’-ACRCTSGGC-GTGRGCSGC-GCAA-3’, Aa-RanGAP_R 5’-GTtT-TSACKCARRAGTCKCGAAA-3’), for which the maximum degeneracy was 32.

#### cDNA isolation and characterization

Total RNA was isolated from 10 (A. aegypti Liverpool SB strain: five males and five females) mosquitoes in 1 ml of TRIZOL™ reagent (Invitrogen) following the supplier’s protocol. RT-PCR was performed using the Access RT-PCR system (Promega) following the supplier’s protocol with 40 pmol of forward and reverse degenerate primers and 500 ng of total RNA per 50 μl reaction. The first strand cDNA was synthesized with 5 U of AMV reverse transcriptase at 48°C for 45 min. The AMV reverse transcriptase was inactivated by 2 min incubation at 94°C, and then PCR reactions were performed: 95°C for 30 s, 58°C for 1 min, 68°C for 2 min for 40 cycles. The final cycle had an extension time of 7 min at 68°C.

The RT-PCR products were cloned using the TOPO TA cloning™ Kit (Invitrogen) following the supplier’s protocol. Cycle sequencing was performed using the ABI Prism Big Dye Terminator Kit v. 3.1 (PE Applied Biosystems) and an ABI PRISM 3700 Genetic Analyzer (Biosystems) with M13 primers. Sequence data were submitted to the TBLASTx program (Altschul et al. 1997) to verify identification as the A. aegypti RanGAP homolog. The 5’- and 3’-end sequences were determined by 5’- and 3’- RACE (rapid amplification of cDNA end), respectively. Full-length cDNA was synthesized using the GeneRacer™ RACE Ready cDNA Kit (Invitrogen) following the supplier’s protocol. 1000 ng of total RNA was treated with calf intestinal phosphatase (CIP) to remove 5’ phosphates, and then with tobacco acid pyrophosphatase (TAP) to remove the 5’ mRNA cap structure, which leaves a 5’ phosphate required for ligation to the GeneRacer RNA Oligo (5’-CGACUGGAGCAGGAGACUGAC-AUGGACUGAAGGAGGAAA-3’). The GeneRacer RNA Oligo was ligated to the 5’ end of the decapped mRNA using T4 RNA ligase. The ligated mRNA was reverse-transcribed using SuperScript II reverse transcriptase and the GeneRacer Oligo dT
Primer (5'-GCTGTCAACGATACGCTACGTAACGCGATACAGT(T)18-3') to synthesize complete cDNA carrying GeneRacer kit-specific oligonucleotide sequences at the 5'-and 3'-ends. The 5', 3'-cDNA ends were amplified with gene specific primers: 5'-primer (5'-AC-RanGAP_5'-RACE-R 5'-ATGCTTTC CGACAGGCGCTAATG-3'), 3'-primer (5'-RanGAP_3'-RACE-F 5'-GCCTACGAGTCGCGAAGCCCAACA-3') and GeneRacer primers: 5'-primer (5'-GCACGGACACUUGCAUGAGCGA-3'; position 9–33 of GeneRacer RNA Oligo), 3'-primer (5'-GCTGTCAACGATACGCTACGTAACG-3'; position 1–25 of GeneRacer Oligo dT primer). After 20-fold dilution of full length RACE cDNA, 11 was added to a 25 l PCR reaction (35 cycles of 94°C for 1 min, 63°C for 1 min and 72°C for 2 min) with 5 min preheating at 94°C and 10 min final extension at 72°C. Amplified 5'- and 3'-RACE DNA was cloned using the TOPO TA cloning Kit (Invitrogen) following the supplier's protocol. Cycle sequencing was performed using the ABI Prism Big Dye Terminator Kit v. 3.1 (PE Applied Biosystems) and an ABI PRISM 3700 Genetic Analyzer with M13 primers. The complete cDNA sequence was submitted to the TIGR index, v. 4.0 (http://www.tigr.org/tigrscripts/tgi/T_index.cgi?species=a_aegypti) and GeneRacer Kit-specific oligonucleotides were submitted to the BLASTN program against the TIGR complete genomic sequence was submitted to the CENSOR SERVER (Jurka et al. 1996; http://www.girinst.org/Censor_Server.html) using the vertebrate database. The complete genomic sequence was submitted to two gene finding programs: GENSCAN (http://genes.mit.edu/GENSCAN.html) using the vertebrate database and FGENESH (http://www.softberry.com/berry.phtml?topic=genshes&group=programs&subgroup=gfindex) using the A. gambiae database. Predicted genes were submitted to Pfam (http://www.sanger.ac.uk/Software/Pfam/) for detection of conserved domain families. To detect transposable elements, the complete genomic sequence was submitted to the CENTILOG at (Severson 1997). Linkage associations were determined using Mapmaker/Exp (3.0b) with a minimum LOD threshold of 3.0 (Lander and Botstein 1989). Map distances were calculated as Kosambi centiMorgan distances (Kosambi 1944).

Genetic mapping

To identify informative strains for mapping, the partial RanGAP cDNA clone was radiolabeled and hybridized to EcoRI-digested genomic DNA from several A. aegypti strains. After selecting two informative strains, strain-diagnostic RFLP markers (Severson et al. 2002) were identified on each chromosome. F1 intercross mapping populations were then prepared as previously described (Severson et al. 1993). From one mapping population, 48 males and 48 females were subjected to individual genomic DNA extractions and Southern transfer following our standard protocols (Severson 1997). Linkage associations were determined using Mapmaker/Exp (3.0b) with a minimum LOD threshold of 3.0 (Lander and Botstein 1989). Map distances were calculated as Kosambi centiMorgan distances (Kosambi 1944).

Results

Full-length RanGAP cDNA characterization

We successfully amplified a 656 bp fragment using degenerate primers and RT-PCR. TBLASTX analysis showed that the PCR product had 63 and 80% sequence identities to the RanGAP homolog in H. sapiens and A. gambiae, respectively. The 5'- and 3'-end sequences of the cDNA were isolated using 5'-, 3'-RACE and sequenced using M13 primers. The complete cDNA sequence consisted of 1,995 bp and included a 113 bp 5' untranslated region (UTR) and a 130 bp 3' UTR (Figure 1). BLASTN analysis against the A. aegypti TIGR EST database identified one tentative consensus sequence (TC31112), that showed 99% sequence identity with our RanGAP cDNA clone and contained a polyadenylation signal near the 3'-end. No significant match with the 5'-cDNA end sequence was detected.

The complete RanGAP cDNA was deduced to encode a 583 amino acid protein (Figure 1). BLASTP
queries showed that *A. aegypti* RanGAP has high sequence identity with the RanGAP homolog in several other species: 73% with *A. gambiae*; 39% with *D. melanogaster*; 52% with *X. laevis*; 53% with *H. sapiens*; 31% with *S. cerevisiae* and 52% with *M. musculus*, respectively. In addition, BLASTP queries against the *D. melanogaster* and *A. gambiae* databases identified only a single best match in each organism with the *A. aegypti* sequence. Multiple RanGAP sequence alignments from 6 species were performed using ClustalX (Figure 2). Alignments show the conserved sequence domain of one putative nuclear localization sequence (NLS) and two NES critical for the cytosolic localization of RanGAP. The complete RanGAP cDNA and genomic sequences were submitted to GenBank (GenBank accession nos. DQ138953 and DQ177444).

**Genomic structure of the RanGAP gene**

Using the 656 bp degenerate RT-PCR product as a probe, we isolated a BAC clone containing the RanGAP gene and determined the complete gene sequence by primer walking. Pair-wise alignment of the cDNA and genomic DNA sequences indicated that the RanGAP gene consists of three exons and two introns. The exon–intron boundaries and the respective exon and intron sizes are shown in Table I. All exon–intron junctions contain the conserved GT/AG dinucleotide motifs. The complete genomic DNA sequence (20,125 bp) has similarity to six *A. aegypti* ESTs and to several putative transposable elements (Figure 3). Neither of the gene finding programs (GENSCAN and FGENESH) correctly identified the RanGAP coding sequence. Both resulted in sets of inaccurate predictions that overlapped the RanGAP open reading frame (ORF), the EST hits, as well as the transposon ORFs. Pfam searches of predicted genes confirmed that the gene finding algorithm in GENSCAN interpreted repetitive elements, likely associated with transposons, as putative genes. Of six ESTs that had significant similarity (< \(E^{-100}\)) to this genomic region, TC66147 matched the 3' end of the RanGAP gene, while the other five were located within the ca. 18 kb second intron. Three overlapping consensus EST sequences (TC52230, TC51913 and TC51912) had significant similarity to only a portion of the *A. aegypti* nanos gene (AY878073), while TC57031 had low similarity to the 5' end of an *A. gambiae* gene (ENSANGG00000023523). The intron region showing partial similarity with TC52296 also included several repetitive sequences indicating that it is probably a degenerate region or pseudogene.
The RanGAP gene in the mosquito Aedes aegypti
Table I. Exon–intron junctions of the RanGAP gene. Coding sequences are shown in uppercase letters and noncoding regions in lowercase letters. The conserved GT/AG exon/intron junctions are shown in bold. Amino acid residues are indicated with respect to each boundary and the stop codon (TGA) is indicated by italics. Numbers refer to the corresponding positions in the RanGAP cDNA starting with +1 at the adenosine defining the initiation codon ATG.

<table>
<thead>
<tr>
<th>No.</th>
<th>Exon size (bp)</th>
<th>Splice-donor</th>
<th>Intron size (bp)</th>
<th>Splice-acceptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>222</td>
<td>109 CCGAGGtaagggaa...Pro Glu</td>
<td>70</td>
<td>110...tatccggaagCCAAGG Pro Arg</td>
</tr>
<tr>
<td>22</td>
<td>324</td>
<td>433 CTGCAGggaagttt t...Leu Gln</td>
<td>18082</td>
<td>434...tttt ttagGAACCG Glu Pro</td>
</tr>
<tr>
<td>33</td>
<td>1429</td>
<td>1753 3 TAGggaagtcatca...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chromosome location of the RanGAP locus

We selected two informative strains, Liverpool and RED strain, for preparing segregating populations to determine linkage associations between RanGAP and marker loci distributed across the genome. The parental cross was set up with one Liverpool male and five RED females. A single F1 egg paper was hatched and mass mated to produce an F2 generation. Liverpool and RED strain diagnostic markers were identified for each chromosome by screening bulked genomic DNA digests of each strain and included: LF397 and LF198 on chromosome 1; LF169, RpL17A on chromosome 2; and LF168 on chromosome 3. The individual F2 progenies were then genotyped for each marker and for RanGAP, and the data were analyzed using Mapmaker/Exp (3.0b). The RanGAP locus mapped to the q-arm of chromosome 2, about 25.5 cM from the RpL17A locus (Figure 4). Because the \( M^D \) locus is tightly linked with the sex-determining locus on chromosome 1 (Newton et al. 1978, Hickey and Craig, 1996a,b), the RanGAP gene is, therefore, likely not responsible for the meiotic drive phenotype in \( A. aegypti \).

Discussion

We have identified and characterized a 1995 bp cDNA and 20,125 bp genomic sequence as the RanGAP homolog in \( A. aegypti \), because of its potential as the meiotic driver gene \( M^D \). Sequence alignment with RanGAP homologues in six other species indicates high sequence identities with the RanGAP homolog in \( A. aegypti \). In addition, BLASTP analyses identified the \( D. melanogaster \) and \( A. gambiae \) RanGAP genes as the single best matches to the \( A. aegypti \) sequence. The conserved LRR_RI motifs are prominent features of the RanGAP proteins, and participate in protein–protein interactions (Kobe and Deisenhofer 1994). Both an NLS and two NESs are conserved, with identities of 72 and 61% with \( A. gambiae \), respectively. Therefore, \( A. aegypti \) RanGAP has high functional domain sequence identities for RanGAP activities and its NLS/NES ratio predicts the cytoplasmic localization of the gene product critical for the \( Run \) pathway (Kusano et al. 2002).

The RanGAP gene in \( A. aegypti \) is slightly greater than 20 kb, compared with only 6.5 kb in \( D. melanogaster \). Although the length of RanGAP coding sequences in the two species are similar at \( \sim 2 \) kb, the largest intron in \( A. aegypti \) is \( \sim 4.5 \) -fold larger that that observed in \( D. melanogaster \). This phenomenon is not surprising given that genome size in \( A. aegypti \) is at least 4.8-fold larger than \( D. melanogaster \) (Severson et al. 2004b). Presence of a large 18 kb intron, harboring several repetitive elements, negatively influenced the ability of GENSCAN and FGENESH gene finding programs to accurately predict the RanGAP coding sequence in the \( ca. \) 20 kb genomic sequence. In particular, the observed frequency of sequences with high identities to transposons reflects the highly repetitive nature of the \( A. aegypti \) genome, and indicates the expected complexity associated with efforts to correctly annotate the entire genome sequence. Of note, is the absence of the SINE element, \( Feilai \), within the RanGAP gene as this element is ubiquitous throughout the genome with an estimated 59,000 copies (Tu 1999). Regions with high similarity to \( A. aegypti \) ESTs suggest the presence of possible transcripts within the large exon. However, these regions only partially match functional genes, indicating that they likely do not correspond to complete functional transcript units, but instead reflect the long term effects of transposon mobilization and movement that includes partial gene duplication.

Our genetic mapping results show that RanGAP is located on chromosome 2q in \( A. aegypti \). This indicates that RanGAP is likely not responsible for meiotic drive in \( A. aegypti \), because the \( M^D \) locus has...
been mapped to chromosome 1 (Hickey and Craig 1966a,b, Newton et al. 1978).

Although the observed sperm dysfunction of the *A. aegypti* and *D. melanogaster* systems may look similar, the mode of action in each drive system can be different. Sperm dysfunction in SD flies is due to interference with proper chromatin condensation during spermatogenesis (Tokuyasu et al. 1977). Defective sperm fail to undergo the normal lysine-rich to arginine-rich histone transition and generally do not separate properly into individual sperm, which remain syncytial spermatids with tightly coiled tails (Lytte 1993). With *A. aegypti*, the *M^D* phenotype is fragmentation of the female determining chromosome, wherein isochromatid breakage occurs on chiasmic arms of the bivalent at or before the diplotene stage (Wood and Newton 1991). However, it is reasonable to predict that *M^D* is a gene involved in cell division and chromosome assembly during sperm generation. Because the *Ran* signaling pathway is deeply involved in cell division and chromosome formation (Bischoff et al. 1994, Kusano et al. 2002), *MD* may be a mutant version of another co-factor in the *Ran* signaling pathway, although it may be due to novel gene effects.

The complete genome sequences of the *D. melanogaster* and the African malaria mosquito *A. gambiae* have been reported (Adams et al. 2000, Holt et al. 2002). Using this genome information, an *in silico* comparative genome analysis to *A. aegypti* was performed and identified significant whole chromosome arm conservation between all three species (Severson et al. 2002a). Our studies showed that the *RanGAP* gene is located on chromosome 2q in *A. aegypti*, while gene database searches indicated that the homolog is located on chromosome 2L in *D. melanogaster* and chromosome 3R in *A. gambiae*. However, the *M^D* locus in *A. aegypti* is tightly linked with the sex determining locus on chromosome 1 (Newton et al. 1976). Therefore, the *M^D* homolog is most likely located on chromosome 3R or the X in...
A. gambiae (Severson et al. 2004a). The anticipated availability of the complete A. aegypti genome sequence (Severson et al. 2004b) should facilitate our efforts to identify and characterize the $M^D$ system.

Acknowledgements

This work was supported by grant PO1-AI45123 and contract HHSN266200400039C from NIH NIAID.

References


