Genomics 96 (2010) 251–257

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Construction and characterisation of a BAC library made from field specimens of the onchocerciasis vector *Simulium squamosum* (Diptera: Simuliidae)

J.L. Crainey [a], J. Hurst [b], M.D. Wilson [c], A. Hall [d], R.J. Post [a,d,*]

[a] Department of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK
[b] Hurst Bioinformatics Consulting, 13 Rippington Drive, Oxford, OX3 0RH, UK
[c] Noguchi Memorial Institute for Medical Research, University of Ghana, PO LG581, Legon, Accra, Ghana
[d] Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

ARTICLE INFO

Article history:
Received 16 March 2010
Accepted 28 June 2010
Available online 23 July 2010

Keywords:
Bacterial Artificial Chromosome (BAC)
BAC library
Simulium squamosum
Wolbachia
Cytochrome P450
Onchocerciasis

ABSTRACT

A Bacterial Artificial Chromosome (BAC) library was made from wild-caught *Simulium squamosum*, which is an important vector of human onchocerciasis. The library is composed of 12,288 BACs, with an average insert size of 128 kb, and is expected to contain ~1.54 GB of cloned DNA. Random BAC-end sequencing generated over 95 kb of DNA sequence data from which putative *S. squamosum* gene sequences and novel repetitive DNA families were identified, including DNA transposons, retrotransposons and simple sequence repeats (SSRs). The sequence survey also provided evidence of DNA of microbial origin, and dissection of sample blackflies indicated that some of those used to prepare the library were likely to be parasitized by the mermithid *Isomermis lairdi*. Hybridisations with a set of three independent blackfly single-copy genes and two *Wolbachia* genes suggest that the library provides around 13-fold coverage of the *S. squamosum* genome and about 12-fold coverage of its *Wolbachia* endosymbiont.

Introduction

Onchocerciasis (commonly known as river blindness) is a severely debilitating disease caused by infection with the filarial worm *Onchocerca volvulus* (Nematoda: Filarioidea). Affected individuals suffer from skin disease and varying degrees of visual impairment. The WHO estimates that around 37 million people are infected with the disease and that over 800,000 of these are completely blind or have severe visual impairment [1]. Over 99% of these cases occur in Africa where 95% is transmitted by the blood-sucking blackfly *Simulium damnosum* s.l. (Diptera: Simuliidae). *Simulium damnosum* s.l. is not a single species, but a complex of up to 55 different cytospecies (sibling species described on the basis of chromosome variation) which show differences in their ecology, behaviour and competencies as disease vectors [2]. From cytotaxonomic studies, *Simulium squamosum* appears to be chromosomally central within the complex, and it is also an important vector of onchocerciasis making it a suitable model [2].

Although an important disease vector and a target for disease control, relatively little is known about the molecular genetics of the *S. damnosum* complex. DNA sequence deposits for *Simulium* fall far below those for other dipterans of medical importance such as mosquitos (Culicidae) and sandflies (Phlebotominae) (NCBI: http://www.ncbi.nlm.nih.gov/ January 2010) [3], and little is known of genes that would be specifically relevant to disease control. BAC libraries have been used in mosquitoes for the development of tools to track-down and to characterise these sorts of genes, for example those associated with insecticide resistance and pathogen immunity [4–6]. Molecular karyotyping has also been developed in mosquitoes and could be developed for the cytospecies identification of all life-stages of the *S. damnosum* complex [7]. The technique requires the characterisation of chromosomal inversion break-points (in order to design the break-point diagnosing PCR primers) and the successful characterisation of inversion break-points has historically made use of a completely sequenced reference genomic DNA, or a large-insert genome library (such as a BAC library), or both [8–10].

The value of such a library is, however, not just limited to taxonomy and gene discovery and characterisation. Large-insert genomic DNA libraries can be a good starting-point for genome projects. Sequenced or fingerprinted large insert clones (typically BACs) are very often the building blocks used to construct the physical frame-work for complete genome sequencing [11]. For example, BAC libraries were used in the genome projects for *Anopheles gambiae*, *Aedes aegypti*, *Drosophila melanogaster* and the *Wolbachia* endosymbiont of *Brugia malayi* [12–15]. Moreover, a BAC library from *S. squamosum* will also provide a potentially unlimited supply of traceable and renewable genomic DNA from a known cytospecies that would normally require specialist identification and would typically provide only one microgram of DNA per insect.

* Corresponding author. Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. Fax: +44 207 942 5229.
E-mail address: R.Post@nhm.ac.uk (R.J. Post).

0888-7543/$ – see front matter.
doi:10.1016/j.ygeno.2010.06.004

Results

Field collection, identification and parasitism of blackfly larvae for a BAC library

In total, over 3300 blackflies were field collected from Boti falls on the River Pawnpawn and were morphologically identified as members of the *S. damnosum* complex. Approximately 3000 of these were preserved in liquid nitrogen and ultimately used for the *S. squamosum* BAC library (see methods for further details). In addition to the specimens collected to prepare the library, ~300 larvae were collected and preserved in Carnoy's solution (ethanol: acetic acid, 3:1) and used as a voucher population for cytological and parasitic infection assessment. Forty-nine voucher specimens were cytotyped and all were identified as *S. squamosum*. Since 1975, a total of 960 cytospecies identifications have been made from this site (80 reported here for the first time from specimens collected between 2006 and 2008, including the 49 larvae collected simultaneously with those used for BAC library construction), and 880 reported previously [16]. Over 98% of these identifications have been *S. squamosum*, with all other specimens identified as *Simulium yahense* (which is cytogenetically and phylogenetically most closely related to *S. squamosum* [17]), but *S. yahense* has never been recorded in November (at the beginning of the dry season) when these specimens were collected [18]. It follows that the DNA used to prepare this library is expected to derive entirely (or almost entirely) from *S. squamosum*. Parasitic examination also revealed that 21% (21/100) of the blackfly larvae were harbouring one or more *Isomermis lairdi* (Nemadota: Mermithidae) [16].

Library size

Following DNA extraction, cloning and colony picking, a BAC library containing a total of 12,288 colonies was arrayed into thirty-two 384-well plates. Copies of this library have been deposited at the London School of Hygiene and Tropical Medicine and the Natural History Museum, London. Fig. 1 shows a pulsed-field electrophoresis gel with ten randomly selected (and restriction enzyme digested) BAC DNA preparations. Size estimations from these digests and a further 58 (not shown) suggest that the average cloned DNA fragment insert-size is ~130 KB, with just one of the BACs appearing to be empty. Based on these digests and on the BAC end-sequencing (see below) a total of 129 BACs have been assessed for the presence of inserts and of these five (3.9%) appear empty. The *S. squamosum* BAC library described here is thus estimated to contain 1.54 GB of cloned DNA.

Library coverage and composition estimates

There is no genome-size estimate for *S. squamosum*, but an independent assessment of genome representation within the BAC library was carried out by hybridising three genes that are expected to be single-copy in *S. squamosum* and two genes from *Wolbachia* (see below). Table 1 provides a list of the putative single-copy genes hybridised to the above BAC library and provides the resultant genomic representation estimates for each gene. For both the CYP10 and CYP6 simuliid genes used for the experiments described here, estimates are based on hybridisations to a complete library, whereas the estimate from the CG8545-gene is extrapolated from a hybridisation to a subsection of the library composed of 4608 clones (37.5% of the library). The two *Wolbachia* genes were each independently hybridised to half the library (6,144 clones).

None of the five genes hybridised to the library hybridised to the same BAC clones and so they can all be regarded as independent, and thus not in close cytological proximity (as already indicated by *in-situ* hybridisation of the CYP6 and CYP10 genes [19]). In summary, the results from these experiments suggest that the *S. squamosum* genome has an approximate 13-fold representation within the library and the *Wolbachia* genome 12-fold (Table 1).

BAC-end sequence survey

A total of 106 BAC-end sequence reads were taken as part of this survey. Ninety of these sequences were paired (from either end of the same BAC) and thus sequence data was obtained from 61 individual BACs. Seven sequence reads contained longer than expected (>400 nucleotides) vector DNA matches suggesting that the four BACs from which they came from did not contain an insert, and therefore it seems that about ~6% of the BACs are empty. Following the removal of these sequences, and of the expected 5' vector sequence from the remaining reads, the survey provides 95,261 nucleotides (with a

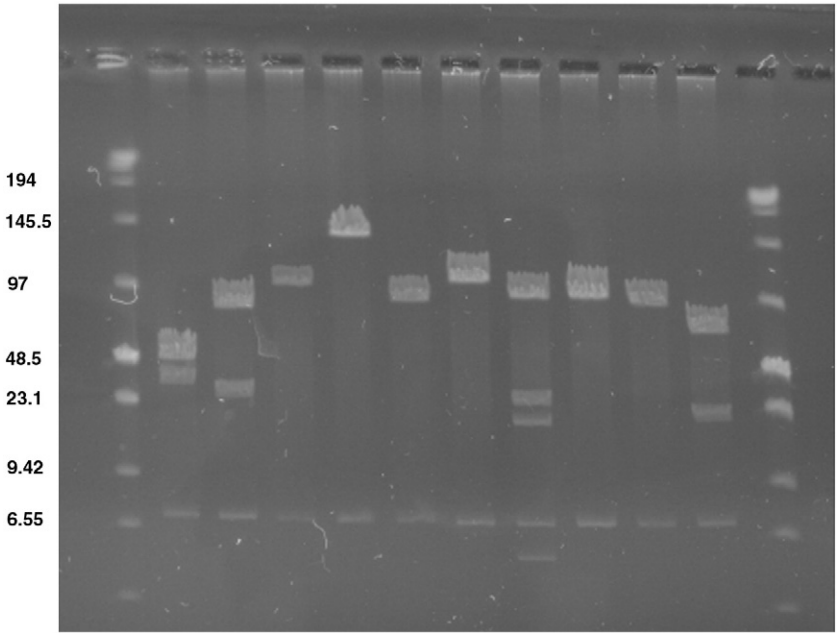

Fig. 1. A pulsed-field electrophoresis gel showing digests of ten randomly selected purified BAC preparations. Sizes of DNA markers are shown in kilobases to the left of the gel and are loaded in the two flanking lanes.