Universal Fast Walking for direct and versatile determination of flanking sequence

Kyl V. Myrick*, William M. Gelbart

Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

Received 10 September 2001; received in revised form 6 December 2001; accepted 21 December 2001
Received by D. Finnegan

Abstract

We report a highly compact system for accelerating direct genome walking. Unlike previous walking techniques, our strategy does not rely on restriction enzymes or ligases, and is therefore unaffected by the availability of useful restriction sites in the flanking region. A complete circumvention of molecular cloning steps qualifies this method for sequencing genome segments that are regarded unclonable, and thus unsequenceable by the traditional methods. A premium was placed on economy of design: the system comprises just four direct reagent additions, in microliter-scale volumes, over the course of a 6-h procedure. The walk range in this method is directly related to the capabilities of the associated polymerase blend, indicating that it can achieve in excess of 35 kilobases per reaction. It also produces a DNA fingerprint that is distinctive to the flanking sequence. Despite the complexity of banding patterns in these fingerprints, we observed that the reaction products were directly sequenceable. In view of its speed, reliability and generality, we term the described method Universal Fast Walking.

q 2002 Elsevier Science B.V. All rights reserved.

Keywords: Genome walking; Transposon mapping; DNA fingerprint; DNA sequencing

1. Introduction

The practice of genome walking – the systematic determination of sequence adjacent to a known region – is integral to the molecular study of chromosomes and large genetic elements. For this, investigators frequently utilize inverse PCR (iPCR), comprised of restriction digestion, ligation-circularization and amplification for molecules spanning the known/unknown boundary (Ochman et al., 1988; Triglia et al., 1988). Over the years, alternative flanking sequence amplification strategies offering various improvements have emerged, including cassette ligation-anchored PCR (Mueller and Wold, 1989), vectorette PCR (Riley et al., 1990), panhandle PCR (Jones and Winistorfer, 1992; Megonigal et al., 2000), Alu-PCR (Puskas et al., 1994), thermal asymmetric interlaced PCR (TAIL-PCR) (Liu and Whittier, 1995), and EPTS/LM-PCR (Schmidt et al., 2001). Though each of these methods has shown some success, their overall effectiveness has been limited by one or more of the following: (i) reduced specificity; (ii) a limited walk range; (iii) a dependence on a favorable restriction map; (iv) a dependence on ligation efficiency; (v) a requirement for molecular cloning; and (vi) excessive failure rates.

In this study, we describe Universal Fast Walking (UFW), a system that addresses all of the above drawbacks, through its high specificity; abolition of restriction enzymes, ligases, and molecular cloning; intensive procedural and spatial economy; and general reliability – providing a number of advantages that, to our knowledge, have not previously been encountered as a whole.

2. Material and methods

2.1. DNAs and oligonucleotides

Genomic DNAs were isolated from Drosophila adults by a LiCl procedure, as previously described (Huang et al., 2000); custom oligonucleotides were obtained from Sigma-Genosys (The Woodlands, TX), and the 1 kb DNA marker ladder was from Life Technologies (Rockville, MD).
2.2. The UFW process

UFW (Fig. 1, and Tables 1 and 2 for specific reaction conditions) is a 5–6 h series of reactions, performed as a one-tube assay under a single thermal cycler program. The method begins with a primer-directed first strand synthesis, followed by destruction of the first primer with exonuclease I (exol) (United States Biochemical, Cleveland, OH).
Table 1
UFW: standard reaction conditions

<table>
<thead>
<tr>
<th>Starting mix</th>
<th>ExoI-digest I</th>
<th>Tagged-random priming</th>
<th>ExoI-digest II</th>
<th>Nested PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 µl genomic DNA (50–100 ng)</td>
<td>Add a pre-mix of:</td>
<td>Add a pre-mix of:</td>
<td>Add a pre-mix of:</td>
<td>Add a pre-mix of:</td>
</tr>
<tr>
<td>3 µl 10× PCR buffer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.8 µl 10 mM dNTP mix</td>
<td>0.5 µl exoI (5 units)</td>
<td>1 µl (15 pmol) primer 2</td>
<td>1 µl exoI (10 units)</td>
<td>1 µl (10 pmol) primer 3</td>
</tr>
<tr>
<td>0.6 µl Takara LA-Taq primer 1</td>
<td>0.5 µl 10× PCR buffer</td>
<td>0.5 µl 10× PCR buffer</td>
<td>0.5 µl 10× PCR buffer</td>
<td>1 µl (10 pmol) primer 4</td>
</tr>
<tr>
<td>0.5 µl exoI (2.5 units)</td>
<td>3.2 µl 10 mM dNTP mix</td>
<td>3.5 µl dH2O</td>
<td>0.5 µl 10× PCR buffer</td>
<td>2.5 µl dH2O</td>
</tr>
<tr>
<td>20.1 µl PCR buffer</td>
<td>0.5 µl dH2O</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cold-start, denature at 95°, 3 min; 51°, 30 s; 68°, 15 s; pause at 37°</td>
<td>37°, 30 min; pause at 4°</td>
<td>Cold-start, denature 94°, 5 min; 18°, 1 s; 68°, 1 s at 0.02° per s for ramping; pause at 4°</td>
<td>37°, 45 min; 68°, 15 min; 80°, 15 min; 95°, 3 min; 60°, 1 s; 43°, 1 s at 0.03° per s for ramping; 68°, 2 min; pause at 4°</td>
<td>Cold-start, denature 95°, 2.5 min, then run 36 cycles of: 95°, 30 s; 51°, 30 s; 68°, 1.5 min, adding 2 s/cycle. A final extension is at 68°, 7 min</td>
</tr>
</tbody>
</table>

Comments:
Specific primer-directed first-strand synthesis. Walk distance is controlled primarily by the polymerase extension time. Annealing temperature is adjustable for the particular primer. Long-distance polymerases are preferred for all walk-lengths, on account of yield and consistency.

Comments:
First primer destruct. This and subsequent additions may be from pre-mixes, for high throughput.

Comments:
Random binding of 5'-tagged primer. Cold-start eliminates the need to heat-inactivate exol of the previous step.

Comments:
Second primer destruct simultaneously with first-strand trimming. Sequence conversion at the 3' end of the first strand. First-strand lariat formation by intrastand annealing and extension

Comments:
Amplification- and first-strand extension times may be jointly increased for longer walks. Nesting removes lariat stem. Annealing temperature is adjustable for Tm. Listed parameters are for high throughput insertion element mapping.

Table 2
P-element UFW primers

<table>
<thead>
<tr>
<th>5'-transposon end</th>
<th>3'-transposon end</th>
</tr>
</thead>
<tbody>
<tr>
<td>P5-1: gaattaattcaacctgcatcagc</td>
<td>P3-1: ggaattaaccaaccccaagg</td>
</tr>
<tr>
<td>P5-2: atccgacctgaggttaacggtgggggn</td>
<td>P3-2: caacactatactctgcmcmcmcmcmc</td>
</tr>
<tr>
<td>P5-3: ctctcagcatcagctgctgctgctgctgtc</td>
<td>P3-3: gcgcactataactacagacac</td>
</tr>
<tr>
<td>P5-4: gcgcacacacgccctgcctcctc</td>
<td>P3-4: ctctcagcatcagctgctgctgctgctgtc</td>
</tr>
<tr>
<td>P5-5: caacactatactctgctgctgctgctgctgctgctg</td>
<td>P3-5: caacactatactctgctgctgctgctgctgctgctg</td>
</tr>
</tbody>
</table>

* Each primer sequence is shown 5'- to -3', and is named in the order of usage within UFW.
format. A further criterion was maintaining sufficient amplification specificity to allow direct sequencing without time-consuming agarose gel purification.

A straightforward procedure was thus devised consisting entirely of direct additions of microliter volumes into microliter volumes, within a standard 96-well PCR tray from start to finish (Table 1). In the most typical form of this method, first strand extension is 15 s, yielding 0.8–1.5 kilobase (kb) bands (Fig. 2), a size range appropriate for our laboratory’s mapping requirements. However, since the walking range of this method is directly related to the capabilities of the polymerase blend, there should be no reason that much longer (35–50 kb) final products could not be obtained by coordinately increasing the first-strand and PCR extension times, using the same long-distance polymerases from this study.

Following primer destruction by the single-strand-specific enzyme, exoI, sequence modification of the 3’ ends of the first strands is accomplished without ligation, by essentially random annealing of primer 2, an oligonucleotide having ten random bases at its 3’ end, but having a specific 5’ sequence, based on a motif from the known end of the first strand (Figs. 1 and 3A). A second addition of exoI removes free primer, and concurrently removes the 3’ end of the first strand until digestion arrives at the point of contact with the primer bound nearest that end, and hence farthest from the walk origin. The resulting 5’ overhang is then filled in by the still-active polymerase, thus converting the first-strand ends to a new sequence, and setting up lariat formation by intrastrand annealing (Fig. 1). Importantly, these termini modifications do not involve restriction cutting, making this technique independent of the neighboring restriction sites, an advantage when working in genome regions poor in restriction sites. In this regard, since it avoids restriction enzyme based cloning – and molecular cloning altogether – we believe this method will also be useful for evaluating genome segments that are poorly represented in whole genome libraries, or completely unclonable and thus conventionally unsequenceable. We term our method UFW, to reflect this directness and generality. As an example, Fig. 3 shows the determination of additional flanking sequence for a GenBank Drosophila scaffold entry, AE003078, originally submitted with an appreciable gap.

We expected that for primer selection in the final PCR step, in theory a number of combinations could generate amplicons, i.e. primers 1 and 3 individually or together, primers 1 and 4 together, or primers 3 and 4 together (Fig. 1); however, the last primer pair, in addition to being fully nested, favors the desired removal of the lariat stem during amplification.

### 3.2. DNA fingerprinting

Interestingly, banding patterns in agarose gels were somewhat complex, as well as reproducible (Fig. 3B), and indeed are a form of DNA fingerprinting that distinguishes between the different flanking regions (Fig. 2). We also find the DNA fingerprinting to be useful as a quality control indicator for the UFW process. This feature is probably best explained by a combination of a component of template site preferences in the annealing of primer 2, and polymerase pause sites (the polymerases used in this study, though

---

**Fig. 2.** UFW DNA fingerprints at high throughput. Agarose gel analysis of 96 different short walks from P-element insertion sites in Drosophila genomic DNA. The banding pattern in each lane is distinctive to the particular walk. Reactions were performed in a single PCR tray from start to finish. Lanes M: 1 kb ladder markers.
**Fig. 3. UFW determination of new sequence.**

(A) Excerpt from a previous *Drosophila* genome project entry, GenBank accession no. AE003078, shown here starting from 5' nucleotide 4651, and containing a sequencing gap; nucleotides in the excerpt are unnumbered due to upward base-count uncertainty presented by the gap. Primers from this study are placed above their respective target segments. UFW primers: oligonucleotides 1 through 5, numbered according to the scheme in Fig. 1. Direct PCR primers: oligonucleotides 4 and 6.

(B) Agarose gel analysis of UFW and direct PCR products spanning the original sequencing gap. Direct PCR was performed on 100 ng *Drosophila* genomic DNA, with the Takara LA-PCR polymerase blend (PanVera) according to manufacturer instructions. Primer annealing was at 60°C, with an extension time of 5 min in each cycle. UFW results from two separate reactions document DNA fingerprint reproducibility. Lane M: 1 kb ladder markers.

(C) UFW-derived sequence extending into the original gap of the GenBank entry. New sequence is denoted in uppercase. UFW and direct PCR amplicons yielded identical sequence.
thermophilic, are expected to be partially active during the 37°C post-synthetic step). Here, it is important to note: (i) the specificity imparted by primer nestings at both the PCR and sequencing levels; and (ii) the fact that the variously sized amplicons from a particular UFW reaction are related, and share a common terminus for sequence priming. Thus, despite the banding complexity, UFW molecules can be sequenced as directly as the standard specific PCR product, without specialized enrichments for the desired amplicon. We typically observe no losses in read-lengths or general quality of UFW sequence electropherograms in comparison to traces from conventional PCR fragments, although it is conceivable that targets with long GC repeats or high secondary structure could occasionally hamper UFW-related amplification or sequencing, to the extent that the currently available polymerases are affected by these structures.

3.3. Potential applications for UFW

UFW has become a preferred technique in our laboratory’s ongoing transposon studies, since it generally permits rapid evaluation of junction sequence (Table 3). We also point out that walking out from certain types of more widely dispersed sites, such as multiple transposon inserts – even if the respective flanking sequences are unique – is expected to be somewhat more challenging for UFW (in the absence of additional isolations), since direct sequencing would then yield composite traces. In that event, UFW may be supplemented with the appropriate standard techniques, such as gel band extraction or molecular cloning of amplicons.

At the time of this writing, UFW appears to be the most widely applicable approach for a first walk into any region when starting from a unique point of origin – and in principle, vis-à-vis competing, almost identical origins, a well-placed nucleotide difference, however small, in the first-strand primer design is likely to provide the specificity required of UFW. Further, since the maximum walk length is set by the most distal, not proximal, random binding site (Fig. 1), UFW is truly ‘polymerase-driven’, and therefore the persistent advances in enzyme technology should empower this method to span ever longer intervals between the available UFW-accessible loci.

Besides direct genome walking, other potential uses of UFW in biomedically important problems include high-throughput mapping of genome-wide insertional mutagenesis for functional genomics, identifying vector integration sites for gene therapy studies, and tracking of viral replication by detecting the insertional activity that accompanies productive infection, as with retroviruses. It should also be of practical interest that the manipulations for running UFW are conducive to automation.

In conclusion, we have developed a novel system, Universal Fast Walking, for flanking sequence determinations. UFW is non-ligational, eliminating concern over DNA ligase efficiencies. It is also non-restrictional and free of molecular cloning, which lends our procedure to very high throughputs, and to the sequencing of traditionally unclonable/unsequenceable genome regions. The walking distance is not limited by the nature of the region’s restriction map, but is a function of the polymerase, putting the present ‘reach’ of this method at 50 kb per walk, given the performance of the currently available enzymes. The operational efficiencies presented by this technique, as well as its consistency and range, have made UFW a method of choice in our day-to-day mapping efforts.

Acknowledgements

Sequence reported in this study has been deposited in GenBank, accession no. AY043178. This study was supported by funding from Grant R37 GM28669-21 of the National Institute of General Medical Sciences of the NIH. We thank Stephanie Mohr for critical reading of the manuscript, and François Huet for helpful discussion.

References


Table 3

<table>
<thead>
<tr>
<th>Transposon line</th>
<th>5' junction</th>
<th>3' junction</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP(2)0616</td>
<td>gttattcatgtGGCCAGCTATCATTTGAAATTCA AGTAAAGTTGGCAGCCCCCTG GCCGTCTGAAAA----</td>
<td>gttattcatgtAGCTGGCAGCTCCACTGGCTAGTG TAACGCTTTGATCATTTGAATTAGCCGTTACA----</td>
</tr>
<tr>
<td>EP(2)2479</td>
<td>gttattcatgtTTTCCACCTGTCTGCTCCTC TCCAAATGTGGCAACAAAAATGAGAGGCTTAT----</td>
<td>gttattcatgtGGCTGAAAGATGTTCAATTGAGA CAATCAACATTTGGAAGGTITTITAAATATTTT----</td>
</tr>
</tbody>
</table>

* Excerpts of UFW-derived genomic sequence (uppercase lettering) flanking both ends of selected Drosophila transposons from the EP collection (Rorth et al., 1998). Sequence is validated against the established iPCR determinations that accompany the EP collection.


