A Vectorette PCR-based approach

for sequencing homologous regions from different genomes

(revised 07/20/04)

This is a more detailed version of the method described in Ko, David, and Akashi "Molecular phylogeny of the *Drosophila melanogaster* subgroup", *Journal of Molecular Evolution* 57:562-573. The vectorette sequences and construction method are modified from a protocol at the Botstein lab website http://genome-www.stanford.edu/group/botlab/protocols/vectorette.html

Introduction

"Vectorette" PCR can be employed to efficiently sequence orthologous gene regions. This method allows amplification of desired DNA fragments using one specific oligonucleotide primer and one primer that extends from a vectorette linker (Riley *et al.* 1990; Arnold and Hodgson 1991). Genomic DNA digestion with a restriction enzyme is followed by ligation of double-stranded vectorettes. The vectorette includes some mismatched bases and a vectorette PCR primer is designed so that synthesis along the complementary strand of the vectorette (from a specific primer) is required before extension can proceed from the vectorette primer (Fig. 1).

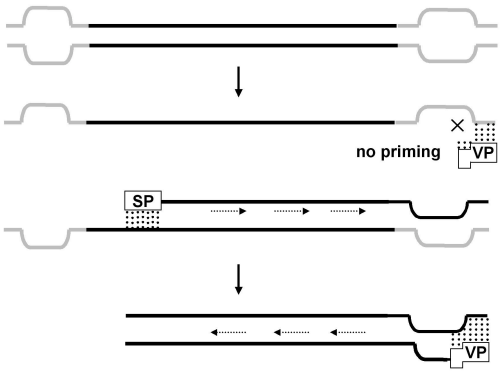

Fig. 1. Vectorette PCR
In the first round of amplification, primer extension only proceeds from the specific primer (SP). Amplification from vectorettes (in gray) does not occur because the vectorette primer (VP) only anneals to the *complement* of the bottom strand of the vectorette. In the second and subsequent rounds of PCR, priming occurs from both the specific primer and the vectorette primer.

Our sequencing strategy begins with a PCR primer designed to recognize a conserved region of a desired gene and we perform PCR reactions with the specific primer and a vectorette primer for multiple vectorette libraries for a given species. A successful primer will often yield an array of fragments with appropriate sizes (in multiples of ~500bp) for DNA sequencing. The vectorette region includes an internal binding site for a sequencing primer which can be used to obtain up to 2kb of sequence (from one strand) from a set of PCR fragments generated from a single specific primer. Primers designed using this sequence are used to sequence the complementary strand. This approach allows us to sequence the 5' and 3' ends of exons and large introns.

The three steps required for this method are described in the following pages: (1) **Extraction of genomic DNA** (2) **Vectorette library construction** (3) **Vectorette PCR** (Fig. 2). If clean, genomic DNA is already available, please start with section 2.

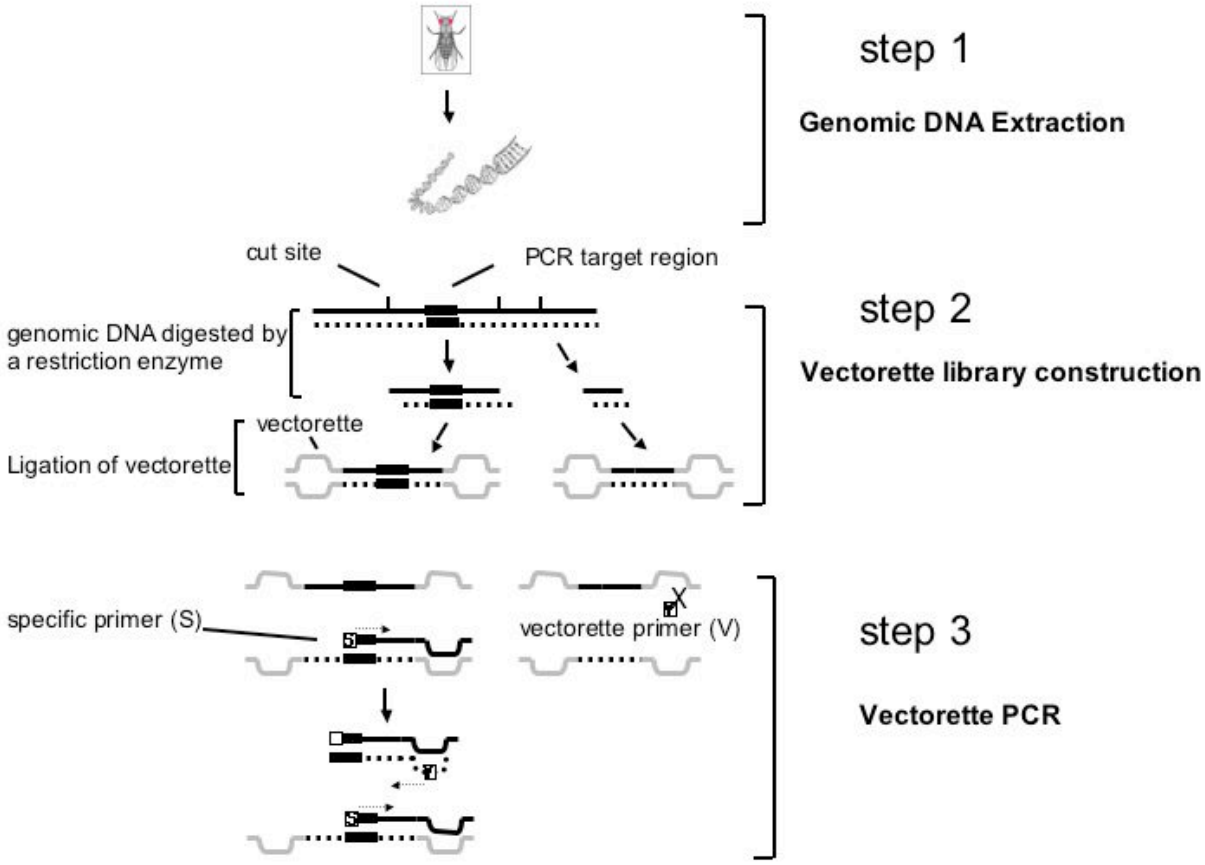

Fig. 2. Genomic DNA extraction, vectorette library construction and vectorette PCR
Genomic DNA was extracted (step 1) and digested with a restriction enzyme followed by the ligation of annealed synthetic oligonucleotides (vectorettes)(step 2). PCR amplification from vectorette libraries employs a specific primer that recognizes region of interest and one primer that anneals to the vectorette.

Step 1: Genomic DNA Extraction

Genomic DNA was extracted from ~100 adult flies. Flies were homogenized and suspended in a 100mM-Tris-HCl (pH 8.0), 50 mM EDTA, 100 mM NaCl, and 2% sodium dodecyl sulfate (SDS) solution. Incubate the solution at 55°C for 3 hours by adding 40 μl of Proteinase K (20mg/ml)(Qiagen) and for another 1 hour by addition of 8 μl of RNase (100mg/ml)(Qiagen).

Table 1. Proteinase Buffer

Chemicals	volume (µl)	Final Conc.
1M Tris-Cl PH 8	36	100mM
0.5M EDTA	36	50mM
5 M NaCl	7.2	100mM
ddH_2O	208.8	
Sum	**288**	
10% SDS	72	2%
Final Total volume	**360**	

Genomic DNA Extraction Protocol

1. Make 288 μl of Proteinase Buffer for 100 individual flies (the gray area in Table 1 which excludes 2% SDS).
2. Add 144 μl of Proteinase Buffer to a 1.5 ml tube.
3. Put 100 individual flies into tube and grind flies with a pestle.
4. Add another 144 μl of Proteinase Buffer (without 2% SDS) to 1.5 ml tube to wash out fly tissue attached to the pestle and tube wall.
5. Add 72 μl 10% SDS (the total volume of Proteinase Buffer is 360 μl). We add SDS after homogenizing flies to avoid bubbling.
6. Add 40 μl Proteinase K (20mg/ml) and incubate at 55 °C for 3 hours. To maximize DNA yield, we mixture with wide-pipette tips every hour.
7. Add 8 μl RNase (100mg/ml). Incubate at 55 °C for 1 hour.
8. After incubation, add 200 μl T.E (to bring reaction mixture up to 600 μl). This is to bring down the DNA conc. in order to reduce DNA loss in the following steps of phenol/chloroform and chloroform extraction, and ethanol precipitation.
9. Phenol/chloroform and chloroform extraction
10. Ethanol precipitation
11. Re-suspend pellets in 300 μl T.E and quantify DNA. We usually obtain about 2 μg DNA/fly.
12. Adjust concentration to 100ng/μl genomic DNA.

Step 2: Vectorette Library Construction

Vectorette sequences and library construction were modified from a protocol from the Botstein laboratory website (http://genome-www.stanford.edu/group/botlab/index.html). We constructed 15 vectorette libraries for each species using a different restriction enzyme for each library. The 15 restriction enzymes are: *Apo*I, *Bam*HI, *Bcl*I, *Bgl*II, *Bsa*HI, *Bst*BI, *Cla*I, *Eco*RI, *Hpa*II, *Mfe*I, *Nar*I, *Nhe*I, *Spe*I, *Taq*$^{\alpha}$I, and *Xba*I. Please refer to Table 2 for enzyme information.

Table 2. Restriction enzyme information (NEB)

Enzyme	Average fragment size (bp)	Restriction site	Units [U/ml]	Buffer	BSA	Incubation Temp. [°C]	Heat Inact. Temp. [°C]
*Bam*HI	7000	5' G↓GATC C 3' C CATG↑G	20,000	*Bam*HI	✓	37	NO
*Bcl*I	4845	5' T↓GATC A 3' A CTAG↑T	15,000	3	NO	50	NO
*Bgl*II	5493	5' A↓GATC T 3' T CTAG↑A	10,000	3	NO	37	NO
*Bsa*HI	1896	5' GPu↓CG PyC 3' C Py GC↑PuG	10,000	4	✓	37	80
*Bst*BI	2893	5' TT↓CG AA 3' AA GC↑TT	20,000	4	NO	65	NO
*Cla*I	3021	5' AT↓CG AT 3' TA GC↑TA	5,000	4	✓	37	65
Taq$^{\alpha}$I	277	5' T↓CG A 3' A GC↑T	20,000	*Taq*$^{\alpha}$I	✓	65	80
*Hpa*II	515	5' C↓CG G 3' G GC↑C	10,000	1	NO	37	65
*Nar*I	7000	5' GG↓CG CC 3' CC GC↑GG	4,000	1	NO	37	65
*Xba*I	8000	5' T↓CTAG A 3' A GATC↑T	20,000	2	✓	37	65
*Nhe*I	10000	5' G↓CTAG C 3' C GATC↑G	5,000	2	✓	37	65
*Spe*I	10000	5' A↓CTAG T 3' T GATC↑A	10,000	2	✓	37	65
*Eco*RI	4000	5' G↓AATT C 3' C TTAA↑G	20,000	*Eco*RI	NO	37	65
*APO*I	407	5' Pu↓AATT Py 3' Py TTAA↑Pu	4,000	3	✓	50	80
*Mfe*I	2247	5' C↓AATT G 3' G TTAA↑C	10,000	4	NO	37	65

Information above was obtained from the New England BioLabs 2002-03 Catalog & Technical Reference.

2a. Restriction enzyme Digestion

To construct a vectorette library, 20 μg of genomic DNA (in 200 µl T.E) was digested with 20 units of restriction enzyme according to manufacturer's instructions [New England Bio Labs, Inc. (NEB)] for 10-16 hours (10~16× over-digestion).

Protocol for restriction enzyme digestion

1. Add 200 μl genomic DNA (100 ng/μl) to a 1.5 ml tube.
2. Make 50 μl reaction mixture for each 200-μl genomic DNA tube (one example is provided in Table 3).
3. Incubate the tubes at appropriate temperature (refer to Table 2).
4. The digestion reaction takes about 11-16 hrs (10x over-digestion).
5. After digestion, heat inactivate digestion enzyme (if necessary) for 20 minutes at the appropriate temperature (refer to Table 2). This step can be skipped for cases where the restriction enzyme does not recognize the ligated vectorette site (among the enzymes that we use, only *Mfe*I and *Hpa*II require inactivation).
6. Ligation reactions can proceed in these restriction enzyme buffers (without extractions or EtOH precipitation) according to the manufacturer's instructions – NEB (but see p. 8).

Table 3. Example of restriction digestion (*Apo*I)

Chemical	Volume (µl)	Final amount/ conc.	Incub. Temp. (°C)
*Apo*I (4 units/µl)	5	20 units	50
10x NEB 3 buffer	25	1x	
100x BSA	2.5	1x	**Inactivation Temp. (°C)**
ddH20	17.5		
Sum	**50**		80
Genomic DNA (100ng/µl)	200		
Total	**250**		

2b. Vectorette Annealing

Annealed vectorettes were constructed with 4 different overhanging sequences for ligation to the digested genomic DNA. The four vectorette types share the complement of vect 53 oligonucleotide (with some mismatches in the middle region) but differ in the oligonucleotide that contains the *complement* of the 5' overhangs left by the restriction enzymes. Vect 57 CTAG matches the overhang created by *Bam*HI, *Bcl*I, and *Bgl*II. Vect 57 GATC matches *Xba*I, *Nhe*I, and *Spe*I. Vect 57 TTAA matches *Eco*RI, *Apo*I, and *Mfe*I. Finally, vect 57 GC matches *Bsa*HI, *Bst*BI, *Cla*I, *Taq*$^{\alpha}$I, *Hpa*II, and *Nar*I (Fig. 3). Please note the polarity of the oligo sequences.

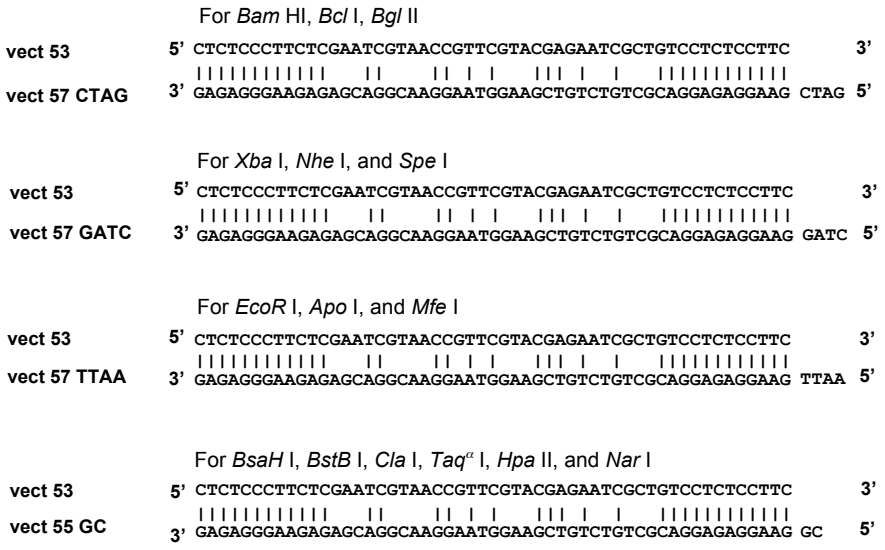

Fig. 3. Double strand vectorettes used for ligation to digested genomic DNA
The vectorette sequences are modified from the Botstein laboratory website at Stanford University (http://genome-www.stanford.edu/group/botlab/index.html).

Table 4. Annealing reaction mixture (for ligation to 20 μg digested genomic DNA)

Chemical	**Volume** (for 20µg DNA)	**Final conc.**
Vect 53 (6.25 µM)	3.2	1 µM
Vect 57 AATT (6.25 µM)	3.2	1 µM
T.E	2.8	
Sum	**9.2**	
$MgCl_2$ (25 mM)	0.8	1 mM
T.E	10	
Total	20	

Vectorette annealing reaction (quantity for ligation to 20 μg DNA)

1. Incubate annealing reaction mixture (9.2 μl, gray area of Table 4) at 65 °C for 5 min. Mix with pipettor. This step is to separate oligonucleotides in the solution.
2. Add 0.8 μl 25 mM $MgCl_2$. Mg^{2+} stabilizes dsDNA.
3. Incubate at 65 °C for 5 min. We use a dry bath.
4. Remove the heat block from the incubator and allow cooling to room temperature (about 1 hour). Annealing of the two vectorette oligos to create "vectorette units" occurs at the stage.
5. Add 10μl T.E to bring the reaction mixture up to 20 μl (final conc. of 1 μM).

2c. Ligation of "vectorette units" to digested genomic DNA

Ligation of appropriate annealed vectorette units to digested genomic DNA is carried out in two steps: (1) annealing of vectorette units to the digested DNA (Fig. 4) followed by (2) ligation of vectorette units to digested DNA by T4 DNA ligase. For the first step, the molar ratio of vectorette to genomic sticky-ends is, on average, 1500. This high ratio promotes annealing of the sticky ends of the digested DNA fragments to vectorette units rather than to other genomic DNA fragments. We did not attempt to maintain the same ratio for different restriction enzymes (but probably should have increased the vectorette concentration for the 4-base cutters). Ligation reaction were performed in restriction endonuclease buffers (2a) supplemented with ATP. These restriction endonuclease buffers are compatible with T4 DNA ligase. However, ligation reactions in NEB buffers *Bam*HI and EcoRI proceed at reduced rates, ~25% and ~50% of rates in 1x ligase buffer (personal communication, NEB technical services). Control vectorette PCR reactions suggest that these rates were sufficient (given the long ligation times) for our construction of Drosophila vectorette libraries.

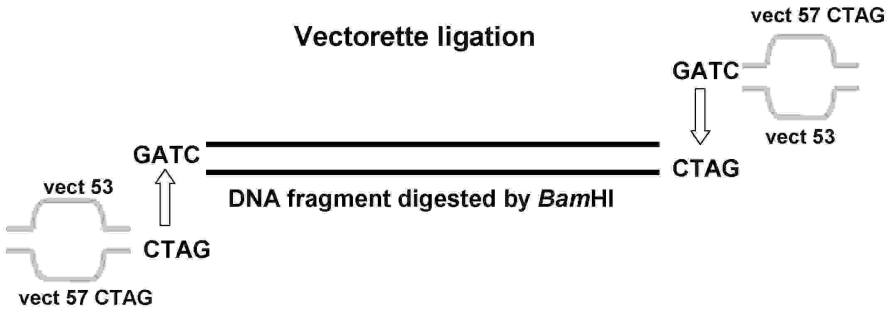

Fig.4. Ligation of vectorette to digested genomic DNA
An example of vectorette ligation shows that vectorette 57 CTAG matches the overhang created by *Bam*HI of a digested DNA fragment.

Table 5. Reaction mixture for ligation (20 μg digested genomic DNA)

Chemical	**Volume** (µl)	**Final amount/conc.**
Digested DNA mixture (0.08 µg/µl)	250	20 µg
Vectorette (1µM)	20	20 pmol
10× ligase buffer (containing 10 mM ATP)	5	1x
ATP (10 mM)	25	1 mM
T4 Ligase (400 U/µl) NEB	2	800 U
Total	302	

Annealing of the vectorette units is followed by the addition of the ligase, ligase buffer and ATP as described below. The 250 µl digested DNA mixture contains 1x buffer (compatible with ligase) but does not contain ATP. 5 µl of 10x ligase buffer and 25 µl of 10 mM ATP is added to the solution to bring the final buffer concentration to 1x and the final ATP concentration to 1 mM in a ~300 µl reaction.

Protocol for ligation reaction

1. Add 20 μl annealed vectorette units (1μM annealing procedure described above) to 250 μl digested genomic DNA (from 2a).
2. Incubate at 65 °C for 5 min to separate "sticky ends" of both vectorettes and genomic DNA.
3. Take the block out of the dry bath and let it cool to room temperature (about 30 min). Annealing of digested DNA and vectorette units occurs at this step.
4. Cool the block to 4 °C.
5. Add 32 μl ligation mixture [including ligase buffer (5 μl), ATP (25 μl), and ligase (2 μl), see Table 5].
6. Incubate at 16 °C for 12-16 hours to ensure complete ligation.
7. After ligation, incubate at 65 °C for 20 min to heat inactivate the ligase.
8. Purify DNA by phenol/chloroform and chloroform extractions and ethanol precipitation. Vectorette PCR can be performed without this step, but long-term stability of the vectorette library should be enhanced by the clean-up.
9. Resuspend the vectorette-ed DNA in 800 μl T.E to a final concentration of 20 ng/μl. We assume 80% DNA recovery rate after phenol/chloroform and chloroform extraction and ethanol precipitation.

Step 3: Vectorette PCR Amplification

Vectorette PCR with multiple vectorette libraries provide an efficient way to obtain sequences of orthologous genes from a group of closely related species. The strategy of vectorette PCR sequencing requires: (1) vectorette PCR using multiple vectorette libraries and (2) sequencing of multiple PCR products (of different sizes) using a vectorette primer.

Vectorette PCR amplification is performed with multiple vectorette libraries using one specific primer and one vectorette primer. The restriction enzymes selected for vectorette library construction have different restriction sites. Some restriction enzymes are selected for digesting DNA into smaller fragment sizes (~ 500 bp or less, on average) while others are selected for digesting DNA into larger fragment sizes (>1000 bp and up to 10 kb, on average, see Table 2). The sizes of PCR products amplified from these libraries depend both on the priming position of the specific primer and positions of the restriction sites in the target region. For those restriction sites less than 2 kb away from the priming position of the specific primer, the DNA fragments can be successfully amplified (we have not attempted "long PCR"). When multiple vectorette libraries are used in vectorette PCR reactions, different sizes of PCR products are expected because the restriction sites are different among these libraries.

Sequence information can be efficiently obtained because one end of each PCR product is the vectorette region. Common vectorette primers can be used in sequencing reactions in one direction for all of the PCR products. With sequence information on one strand, other sequence primers can be designed to obtain sequence from the complementary strand.

For a successful vectorette PCR, a good specific PCR primer is critical. A specific PCR primer can be designed from a conserved region of the target gene given at least one sequence from an outgroup. This method may be useful for molecular evolution and systematics studies that require sequences of a number of orthologous gene regions from a group of closely related species.

We employed all 15 vectorette libraries plus one negative control (PCR reaction mixture + primers without DNA) in a single run of PCR amplification with a given specific primer and a vectorette primer (C20, Fig. 5) for a given species. The vectorette PCR reaction mixture is given in Table 6. To obtain sufficient quantities of PCR fragments for DNA sequencing, PCR products can also be re-amplified using the initial specific primer and a second vectorette primer (B21, Fig. 5). Non-specific binding can be reduced by using a vectorette primer internal to that used in the initial amplification.

Table 6. Vectorette PCR reaction mixture

Chemical	reaction sample to be divided into 3 tubes (μl)	Final amount/ conc.
dd H_2O	17.7	
10x Q10 buffer (Qiagen)	6	1x
10mM dNTP (including AGCT)	1.2	200 μM
Vect. primer C20 (6.25μM)	2.4	0.25 μM
Specific primer (6.25μM)	2.4	0.25 μM
Taq (5 units/μl) [HotStarTaq (Qiagen)]	0.3	1.5 unit
Vect DNA (2ng/μl)	30	60 ng
Total	60	

Vectorette PCR Protocol

1. Prepare 30 μl vectorette-DNA template (2ng/μl) in a 1.5 ml microcentrifuge tube for 15 vectorette libraries and one tube with only 30 μl ddH_2O as negative control (total of 16 reaction tubes).
2. Add 30 μl PCR reaction mixture to each tube (total 60 μl, see Table 6). A master mixture for these 16 reaction samples is recommended.
3. For each sample tube, distribute 20 μl of the reaction mixture to three 0.2 ml thermal cycler PCR tubes labeled H, M, L (high, medium, and low temperatures). In total, there will be 16 × 3 = 48 reaction tubes ready for PCR (Fig. 6).
4. Set the 48 thermal cycler tubes on a gradient thermal cycler and proceed with step down PCR reaction (Fig. 7).

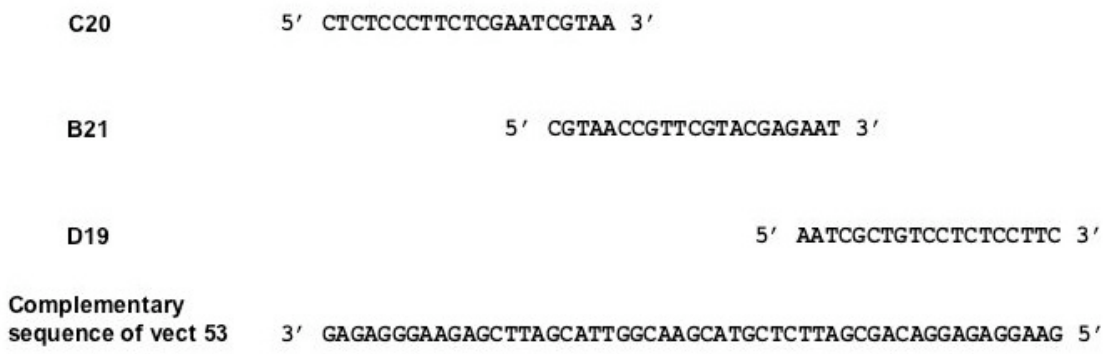

Fig. 5. Vectorette primer information
Vectorette primers are designed to anneal to the complementary sequence of vect 53 (see Fig. 3) that are produced in the initial PCR cycles. Vectorette primer C20 was used for vectorette PCR. B21 was used for re-amplification to yield sufficient quantities of PCR fragments for DNA sequencing. D19 was used for DNA sequencing.

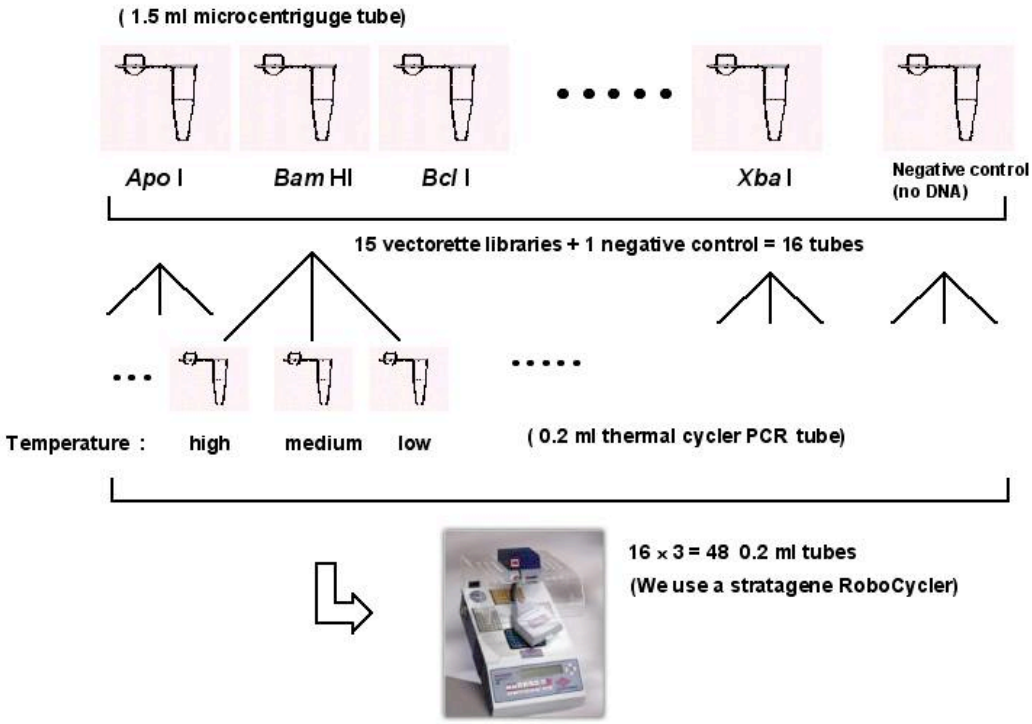

Fig. 6. Setting up a vectorette PCR reaction
15 vectorette libraries plus one negative control in a single run of PCR amplification with a given specific primer and a vectorette primer for a given species. For each tube, reaction mixtureture (60 μl) was distributed evenly to three 0.2 ml thermal cycler PCR tubes labeled H, M, L (high, medium, and low temperatures). In total, there will be 16 × 3 = 48 reaction tubes ready for PCR.

PCR cycles
1 cycle (95°C for 14 min)
(to activate HotStarTaq)
↓
5 cycles (95°C for 1 min, 63~72°C for 1 min, 72°C for 2 min)
↓
5 cycles (95°C for 1 min, 59~68°C for 1 min, 72°C for 2 min)
↓
15 cycles (95°C for 45 sec, 55~64°C for 1 min, 72°C for 2 min)
↓
15 cycles (95°C for 45 sec, 51~60°C for 1 min, 72°C for 2 min)
↓
1 cycle (72°C for 10 min)

Fig. 7. PCR cycles
Initial PCR cycles employed higher annealing temperatures [approximately 10 °C above the melting temperature (T_m) of the primer-template complexes] and the annealing temperature was gradually lowered in steps of 4°C. Early PCR cycles with high annealing temperatures are designed for specificity while later cycles with reduced annealing temperatures produce greater yields. This is a modified form of "touchdown" PCR referred to as "step-down" PCR (Hecker and Roux 1996).

Vectorette PCR Results

Figure 8 shows products of vectorette PCR from the *Adhr* locus in *D. mimetica*. Subsequent DNA sequencing confirmed that vectorette PCR successfully amplified all expected products within 1.5 kb of the specific primer. Desired and spurious PCR products can be distinguished using the temperature gradient. The ratio of target to spurious fragments increases with annealing temperature (see *Bam*HI and *Bst*BI lanes in the figure below).

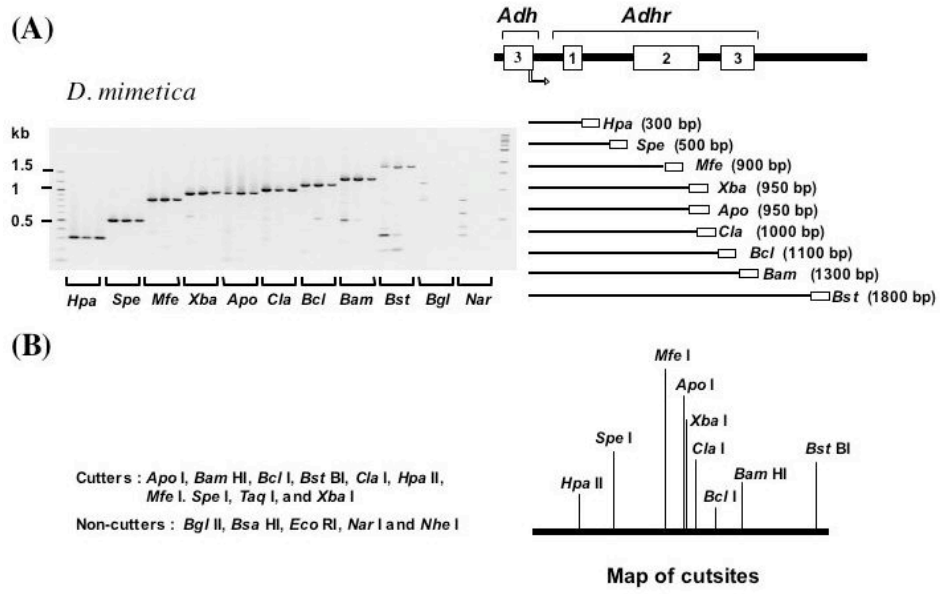

Fig. 8. **(A)** Vectorette PCR of the *Alcohol dehydrogenase related* (*Adhr*) gene region in *Drosophila mimetica*. Gel images show vectorette PCR fragments from amplifications using a specific primer that recognizes the end of the last exon of *Adh*. Amplifications were performed at three different annealing temperatures (shown from left to right for annealing temperatures from low to high for each vectorette library). Size standards are shown in the right- and left-most lanes of each gel. Vectorette PCR products are shown in increasing order of size and the corresponding regions of the *Adhr* region are depicted to the right of each gel image. The empty box at the right end of each PCR product represents the vectorette sequence. Results are also shown for two libraries that were not expected to yield fragments (*Bgl* II and *Nar* I). **(B)** The map of restriction enzyme cut sites from the known sequence of *Adhr* in *D. mimetica*. All expected fragments were successfully amplified.

References

Arnold C, Hodgson IJ (1991) Vectorette PCR: a novel approach to genomic walking. PCR Methods Appl 1:39-42

Hecker KH, Roux KH (1996) High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. Biotechniques 20:478-85

Ko, W-Y, Ryan MD, Akashi H (2003) Molecular Phylogeny of the *Drosophila melanogaster* species subgroup. J Mol Evol 57:562-573

Riley J, Butler R, Ogilvie D, Finniear R, Jenner D, Powell S, Anand R, Smith JC, Markham AF (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res 18:2887-90

Acknowledgments

This research was supported by grants from the Alfred P. Sloan Foundation grant and the National Science Foundation (DEB-0235472) to HA.

Updates to the protocol

Listed below are some corrections and modifications made in the July 20, 2004 revision.

1. In the protocol for restriction enzyme digestion (page 5), phenol/chloroform and chloroform extraction and ethanol precipitation were removed. This step is unnecessary because ligation reactions can be performed in the restriction endonuclease buffers (supplemented with ATP).

2. Table 5 (page 8) was modified for ligation reactions in the restriction endonuclease buffers. In addition, the final amount of vectorette was corrected to 20 pmol.

3. In Tables 4 and 5 (page 6 and 8, respectively) and the protocols for vectorette annealing and ligation reactions (page 7 and 9, respectively), the concentration of vectorettes were corrected from "mM" to "μM" (i.e., 6.25 mM was changed to 6.25 μM for the initial vectorette concentration and 1mM was changed to 1 μM for the final vectorette concentration).

5. In Table 6, the final amount of Taq was corrected to 1.5 units.

6. In Fig. 5, the complementary sequence of vect 53 is shown (to illustrate the relative positions to which vectorette primers anneal).