Amplified restriction fragments for genomic enrichment

(version 2.6b October 2014)

Tom Parchman (tparchma@uwyo.edu)
Zach Gompert (zgompert@uwyo.edu)
Alex Buerkle (buerkle@uwyo.edu)
University of Wyoming

This is a protocol for preparing highly-multiplexed samples for high-throughput sequencing. Genomic DNA is digested with restriction enzymes, barcodes and Illumina adaptor sequences are ligated to the digested fragments, and a subset of these fragments is amplified by PCR. Amplified fragments are purified and size-selected in agarose before being pooled into an Illumina sequencing library (see Fig. 2). This is a flexible protocol that has worked with a wide-variety of taxa and can be modified to target a variable number of fragments for sequencing (see Fig. 3).

Disclaimer: We cannot guarantee your success with this protocol. We also caution that this protocol and other approaches to multiplexed sequencing lead to immense amounts of data, with stochastic variation in sequence coverage across individuals and genetic regions. Importantly, we recommend making a comprehensive plan for analyzing the data before you follow this protocol. If you do not know what you will do with $> 10^8$ barcoded sequencing reads, then this protocol will not be useful.

1 Reagents and Equipment

1. EcoRI (NEB, 20,000 units/mL) R0101L
2. MseI (NEB, 10,000 units/mL) R0525L
3. T4 DNA ligase (NEB, 400,000 units/mL) M0202L
4. BioRad Iproof High Fidelity DNA polymerase Cat no. 172-5301 ($424 for 250 μL [500 units])
5. Molecular Biology grade agarose (suggestion: BioRad 161-3101)
6. 1 M NaCl
7. DMSO
8. Thermal Cyclers
9. 96 well plates and strip caps
10. Pipette tips
11. Agarose gel rig and UV light table
12. Full plate centrifuge
13. Nanodrop or bioanalyzer
14. other reagents listed below

2 Oligonucleotide Sequences

2.1 Adaptors

2.1.1 EcoRI

1. See Table S1 for barcoded EcoRI primer combinations. These adaptor sequences come in pairs and after annealing become one double-stranded adaptor. The barcodes residing in

the adaptors were created using python scripts described in (Meyer *et al.* 2010). To create variability in the initial bases for Illumina sequencing, we created barcodes that were 8, 9, and 10 bases in length. We constructed sets of barcodes so that all barcodes differ by a minimum of 4 bases. This facilitates correction of sequencing errors in barcodes. From the total set, we used an additional python script to select a subset that is balanced in the usage of nucleotides at each position. Further information on the python scripts can be found at (`http://bioinf.eva.mpg.de/multiplex/`).

2. Mix 1 μL of each oligo in a pair (100 μM stock) with 98 μL of water to make 100 μL of 1 pmole/μL (1 μM) of annealed, double-stranded adaptor stock. Heat to 95°C for 5 minutes and slowly cool to room temperature. This annealing step only needs to be performed once for a stock solution of adaptor. Keep the set of adaptors organized in plate format that is convenient for later use in setting up reactions (e.g., Fig. 1).

2.1.2 MseI

Mix 10 μL of the MseI1 and MseI2 oligos (100 μM stock) with 80 μL of water to make 100 μL of 10 pmol/μL (10 μM) stock. Heat to 95°C for 5 minutes and slowly cool to room temp to anneal oligos into double-stranded adaptor. This annealing step only needs to be performed once for a stock solution of adaptor.

These oligos are the same for each fragment. On an Illumina sequencer and with our configuration of the sequencing primer, sequencing only occurs from the EcoRI side, so there is no need to barcode the MseI oligos.

MseI1: 5' `GCAGAAGACGGCATACGAGCTCTTCCGATCTG` 3'

MseI2: 5' `T*ACAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTT*G` 3'

The asterisks between the first three bases in each PCR primer represent phosphothiolation (this is an option when ordering oligos). This prevents exonucleases from acting on the ends of the fragments, and thus protects the product prior to sequencing.

2.1.3 PCR primers

Mix 5 μL of the Illpcr1 and Illpcr2 oligos (100 μM stock) with 90 μL of water to make a working solution (2.5 μM of each oligo).

The last base of the Illpcr2 primer is a selective base and is arbitrarily **G**. Different and additional selective bases could be added, as desired. However, there is some evidence that the selective base does not function with high-fidelity polymerase and could be omitted.

Illpcr1 (Forward): 5' `A*ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T` 3'

Illpcr2 (Reverse): 5' `C*AAGCAGAAGACGGCATACGAGCTCTTCCGATCTGTA*A` 3'

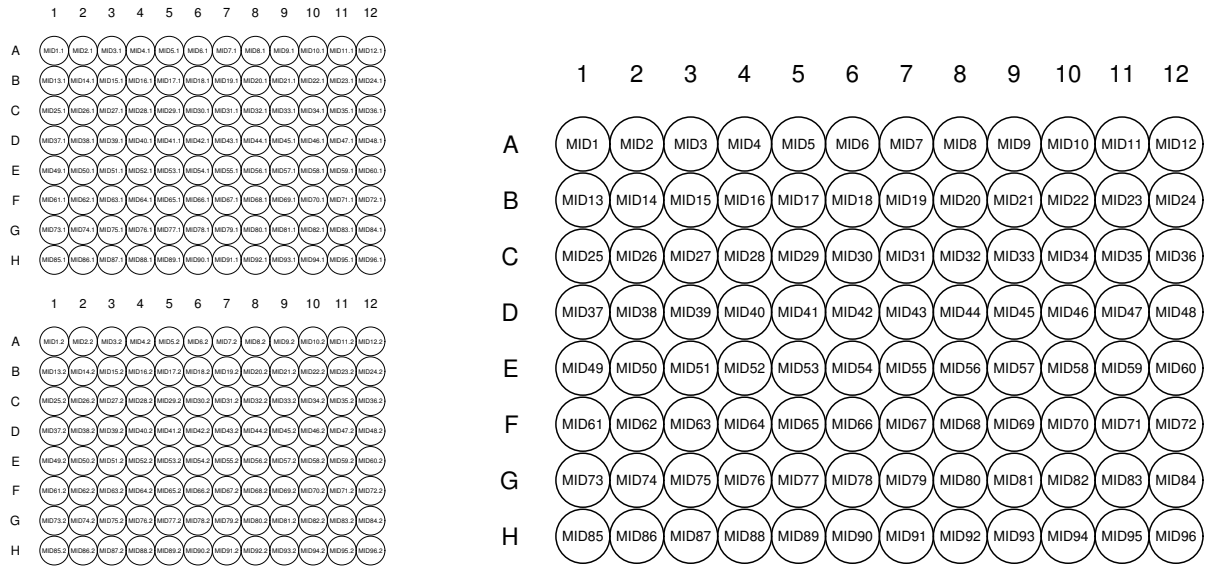


Figure 1: Sample layouts of 96 adaptor oligos on plates. At left are plates as they would be ordered. Oligos pairs are combined, annealed and then stored for use as in the plate on the right.

2.1.4 Illumina sequencing primer

This primer is used during Illumina sequencing, is not part of the library prep, and is presented here only to document its sequence.

5' `ACACTCTTTCCCTACACGACGCTCTTCCGATC` 3'

3 Restriction and Ligation Reactions

Genomic DNA is first digested with two restriction enzymes, EcoRI and MseI. This results in a pool of fragments that have the sticky-ended restriction cut sites on either end, and these ends provide a template for ligation of the customized adaptor sequences. The adaptor sequences contain the Illumina adaptors and primer sequences (and hence provide a binding site for the Illumina PCR primers). After the Illumina adaptor sequences on the EcoRI end, 8–10 base pair barcodes are incorporated that allow the identification of the origin of each sequencing read. The five bases at the 3' end of each adaptor oligo correspond to the ligation site; that is, they match the restriction cut site and have one added, protector base that prevents the cut from occurring again. The restriction digest is run first, followed by the permanent inactivation of enzymes by heating to 65°C. The ligation reaction is then run separately, so digestion of adaptors containing restriction enzyme cut sites cannot occur (a small number of our 768 EcoRI adaptor sequences contain the EcoRI or Mse1 cut sites; likewise for any other potential alternative restriction enzymes) and reactions can occur at their optimum temperatures.

3.1 Restriction Digest

1. Place 6 μL of sample DNA in each well of a plate (DNA should ideally be at a minimum concentration of 20 ng/μL and a maximum concentration of 150 ng/μL (note: this is for small

Reagent	Number of samples 1×	150× (for 1 plate)
10× CutSmart Buffer	1	150
1M NaCl	0.52	78
Water	0.73	109.5
MseI	0.1	15
EcoRI	0.25	37.5

Table 1: Reagents and volumes for Restriction Digest master mix I (2.6 μL prepared per sample).

genomes, for large genomes add higher concentrations of DNA). Keep on ice.

2. For each sample prepare master mix I (Table 1), mix by vortexing, and centrifuge.

   For these and all other reactions make sure to prepare an excess of mix to accommodate multiple rounds of pipetting, particularly if you are working with whole plates. Because the enzymes are stored in glycerol and other viscous solutions, a substantial volume is lost through adhesion to the outside of pipette tips. We suggest making 150% of what you think you will need.

3. Add 2.6 μL of the combined master mix I to each sample. Always keep cold once the enzymes have been added.

4. The total reaction volume should be 8.6 μL. Cover and seal the plate, vortex, centrifuge and incubate at 37 °C for 2 hours, followed by 65°C for 20 minutes (to inactivate the enzyme) on a thermalcycler with a heated lid.

3.2 Adaptor Ligation

1. Thaw MseI and EcoRI adaptors. Have these adaptors annealed and easily accessible in plate format (for the EcoRI adaptors) (see Sections 2.1.1 and 2.1.2).

2. Add 1.54 μL of master mix II to each restriction-digested reaction.

3. Add 1 μL of EcoRI adaptors to each well on plate. Note that the MseI adaptor is in master mix I.

4. The total reaction volume should now be 11.14 μL. Cover and seal the plate, vortex, centrifuge and incubate at 16 °C for 2 hours followed by 65°C for 10 minutes on a thermalcycler.

5. Dilute the Restriction-Ligation reaction with 189 μL of 0.1× TE (or water). Store at 4 °C for a month, or -20 °C for longer.

4 PCR amplification

1. This PCR step uses the Illumina PCR primers to amplify fragments that have our adaptors+barcodes ligated onto the ends. To ameliorate stochastic differences in PCR production

Reagent	Number of samples 1×	120× (for 1 plate)
MseI adaptor	1	120
10× T4 buffer	0.322	38.64
1M NaCl	0.05	6
T4 DNA ligase	0.1675	20.1

Table 2: Reagents and volumes for Restriction-Ligation master mix II (1.54 μL prepared per sample).

of fragments in reactions, we run two separate 20 μL reactions per restriction-ligation product, and later combine them. Pairs of individuals can be pooled prior to PCR to reduce time and costs.

Reagent	1× 20 μL Reaction	× 125 (for 1 plate)
Water	10.4	1300
5× Iproof buffer	4.0	500
dNTP (10 mM)	0.4	50
$MgCl_2$ (50mM)	0.4	50
Pre-mixed PCR Primers, 5 μM	1.3	162.5
Iproof taq	0.2	25
DMSO	0.3	37.5
Total MM	17	2125
R/L product	3	NA

Table 3: Reagents and volumes for PCR amplification reactions.

2. Thermalcycler profile for this PCR: 98°C for 30s; 30 cycles of: 98°C for 20s, 60°C for 30s, 72°C for 30s; final extension at 72°C for 10 min.

5 Extra PCR step

1. This PCR step adds Illumina PCR primers, dNTPs, and iproof buffer and executes a final long PCR cycle. The purpose of this step is to attempt to convert single stranded template remaining from the first PCR to double stranded. This typically results in libraries that have a better distribution of fragments in the correct size range, presumably because it removes single stranded molecules that interfere with the precision of the electrophoresis step below.

2. Add 2.125 μL of the mix below directly to each reaction immediately after the PCR step above is completed.

Reagent	20 μL Reaction	$\times$ 125 (for 1 plate)
5$\times$ Iproof buffer	0.425	53.1
dNTP (10 mM)	0.4	50
Pre-mixed PCR Primers, 5 μM	1.3	162.5

Table 4: Reagents and volumes for PCR optional final PCR cycle.

3. Thermalcycler profile for this PCR cycle: 98°C for 3 min; 60°C for 2 mins, and 72°C for 10 min.

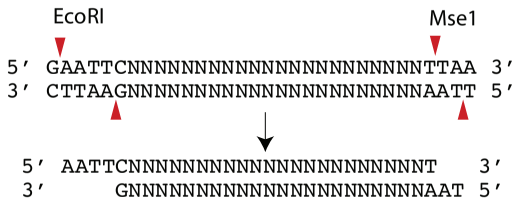
6 Gel Purification

This step involves extracting a desired size range of fragments using gel electrophoresis. The size range can be adjusted to modify the number of fragments expected for sequencing. Because techniques such as this typically result in a negative relationship between fragment size and number, larger fragments should decrease the final number of fragments in the template and increase the coverage depth of these regions after sequencing. To eliminate variation in fragments among individuals, it is recommended to pool all PCR reactions (they already have the barcodes incorporated) before this step. We now use the BluePippin for this step. Purify 90-120 μl of sample with the BluePippin in the Genomics Core Lab. Your sample should be provided in three to four 0.2 ml tubes in 30 μl aliquots. Have your sample quantified on the BioAnalyzer with the high sensitivity chip after purification.

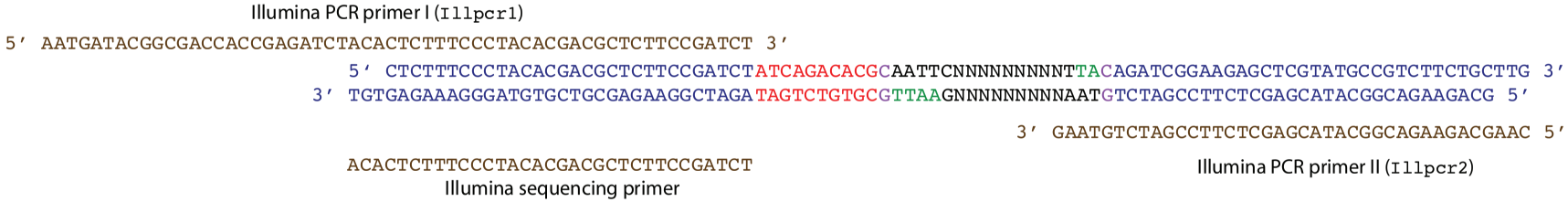
7 Preparing final template for Illumina sequencing

Use a Nanodrop or similar spectrophotometer to measure the DNA concentration of purified library. Add the same amount of DNA from each gel purification to a 2 mL tube. A total concentration of 25μL/ng – 100μL/ng is ideal for sending off for Illumina sequencing, although library size requirements vary.

1. Digest double-stranded DNA with EcoRI and Mse1.



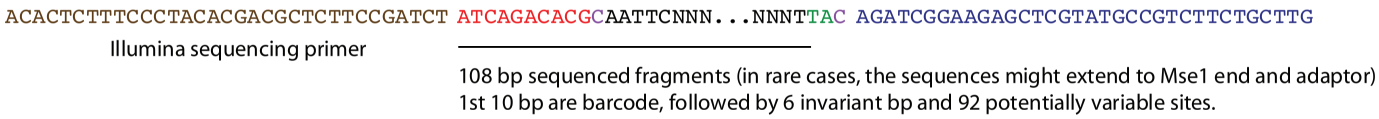
2. Ligate adaptors to fragments. Adaptors include adaptor sequence, barcode, cutsite, and protector base (adaptors in color; EcoRI on left, Mse1 on right).



3. Amplify fragments with Illumina PCR primers.

4. Gel purify PCR product in the desired size range (300-400bp).

5. Illumina sequencing.



***Note only fragments with the EcoRI adaptor on one side and the Mse1 adaptor on the other will be sequenced due to the bridge-PCR method in Illumina sequencing.

Figure 2: A schematic of the steps involved in the Restriction-Ligation and PCR steps involved in creating the seqeuncing library.

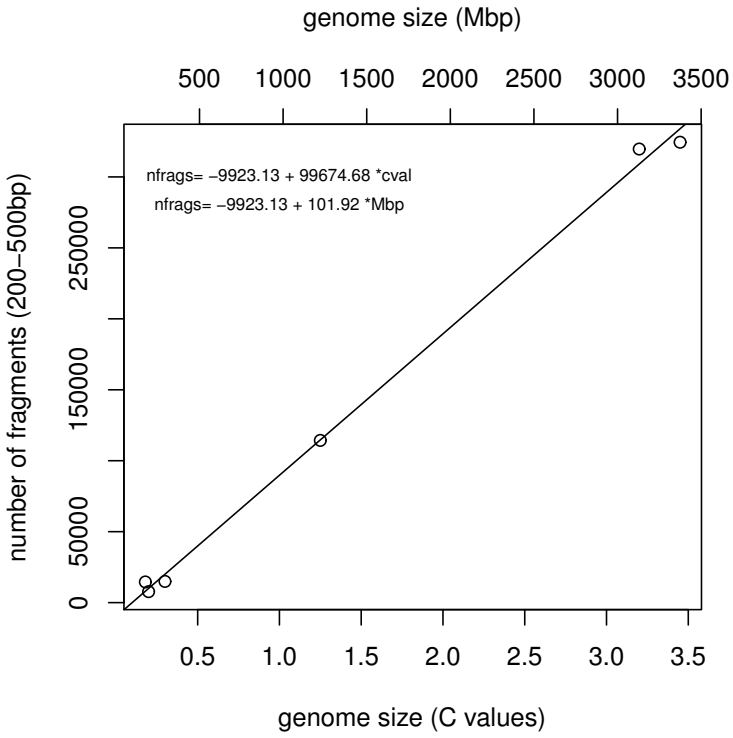


Figure 3: Number of fragments (200–500 bp) that are expected to be generated by digesting genomes of differing size. These expectations were derived by using Perl to digest reference genomes for *Drosophila*, chicken, mouse, *Arabidopsis*, *Tribolium* and human, and to count fragments within the 200–500 bp size range. In practice, a larger number of fragments is recovered and assembled in population studies.

Table S1: Example EcoRI adaptor sequences with barcodes added. The barcodes (in uppercase letters) are located between the adaptor sequence and the restriction site sequence (see Fig. 2). The full list of barcodes can be found in the supplemental file annealed_barcode_out_768.csv

Name	Sequence (5' to 3')
Mid1.1	ctctttccctacacgacgctcttccgatctACGAGTGCGTc
MID1.2	aattgACGCACTCGTagatcggaagagcgtcgtgtagggaaagagtgt
MID2.1	ctctttccctacacgacgctcttccgatctACGCTCGACAc
MID2.2	aattgTGTCGAGCGTagatcggaagagcgtcgtgtagggaaagagtgt
MID3.1	ctctttccctacacgacgctcttccgatctAGACGCACTCc
MID3.2	aattgGAGTGCGTCTagatcggaagagcgtcgtgtagggaaagagtgt
MID4.1	ctctttccctacacgacgctcttccgatctAGCACTGTAGc
MID4.2	aattgCTACAGTGCTagatcggaagagcgtcgtgtagggaaagagtgt
MID6.1	ctctttccctacacgacgctcttccgatctATATCGCGAGc
MID6.2	aattgCTCGCGATATagatcggaagagcgtcgtgtagggaaagagtgt
MID7.1	ctctttccctacacgacgctcttccgatctCGTGTCTCTAc

Name	Sequence (5' to 3')
MID7.2	aattgTAGAGACACGagatcggaagagcgtcgtgtagggaaagagtgt
MID8.1	ctctttccctacacgacgctcttccgatctCTCGCGTGTCc
MID8.2	aattgGACACGCGAGagatcggaagagcgtcgtgtagggaaagagtgt
MID10.1	ctctttccctacacgacgctcttccgatctTCTCTATGCGc
MID10.2	aattgCGCATAGAGAagatcggaagagcgtcgtgtagggaaagagtgt