Protocol for Illumina 2bRAD sample preparation

Galina Aglyamova (aglyamova@austin.utexas.edu)
Mikhail Matz (matz@utexas.edu)

Critical update: August 25 2014 – introduced NNRW adapters to discard PCR duplicates
Minor update: March 20, 2015 – added NNRW to the sequence of reduced-representation (..NG) adapter
Minor update: March 27, 2016 – replaced qPCR with picoGreen quantification in the end

Overview

This is a modification of the protocol described in Wang et al, Nature Methods 2012 2b-RAD paper (doi:10.1038/nmeth.2023) re-designed for Illumina HiSeq and BcgI enzyme (https://www.neb.com/products/R0545-BcgI). BcgI is a relatively frequent cutter, with 75-100k sites per genome, and in our experience is one of the most efficient of the IIb-type restriction enzymes. It produces 36-base fragments with two-base overhangs and is fully heat-inactivatable, both of which facilitate ligation of adaptors. Another enzyme that can be used with this protocol is AlfI (http://www.thermoscientificbio.com/restriction-enzymes/alfi/)

The latest modifications include the use of barcoded ligated adaptors so the samples can be pooled by 12 after ligation, and use of degenerate 5'-adaptor that makes it possible to remove PCR duplicates.

The protocol involves the following steps, of which steps 1-3 are performed within the same tube (or well of a 96-well plate) by consecutively adding reagents.

1. Restriction digest. Genomic DNA is digested with a type IIB restriction enzyme to produce restriction fragments of uniform length.
2. Ligation. Adaptors, barcoded for 3' end and generic for 5', are ligated to the cohesive ends generated by restriction digest. Overnight ligation at +4°C, then heat-inactivate the ligase for 10 min at 65°C.
3. Pool ligations with different 3' barcodes into groups of 12 (because we have 12 different 3' ligation adapters). Determine proportions from gel picture (regular pcr) or by qPCR.
4. Amplification and barcoding of pooled ligations.
5. Purification of the target 170 bp band by gel-electrophoresis (the only purification step in the whole procedure).
6. Quantification and mixing in equal proportions.

Digest

1. Prepare intact, high-quality genomic DNA samples each containing a total of 100-200 ng in 4 µl. This can be accomplished using by ethanol precipitation or by drying under vacuum. Ideally, samples should have the same amount of DNA.
   Note: we have successfully prepared samples from as little as 50 ng.

2. Prepare a digestion master mix. The following recipe is for a single reaction, so multiply by the number of samples plus some small amount for pipetting error.

NEB Buffer #3	0.6 µl
150 µM SAM	0.4 µl
BcgI (1 U μl^{-1})	1.0 µl

**Note: SAM [S-adenosyl-methionine] and enzyme concentrations may differ depending on the manufacturer. If they are different in your case, re-calculate the volumes, make up the difference with nuclease free water if necessary, but keep the single-reaction volume at 6 µl.*

3. Combine 2 µl master mix with each 4 µl DNA sample (6 µl total volume). Cover this preparation with a drop of mineral oil. Incubate at 37°C 1 hr. If using a heat-inactivatable enzyme (e.g. BcgI, AlfI), inactivate the enzyme at 65°C for 20 min. Hold samples on ice.
   **Note: one way to increase digestion efficiency is to add one extra µl of enzyme after the initial 1h incubation and incubate for one more hour before heat-inactivation.*

4. (optional) For each sample, load 1 µl digested DNA on a 1% agarose gel alongside a comparable amount of intact DNA from the same sample to verify the effectiveness of the digest.

The signs of an effective enzyme digest are quite subtle and include (a) a slight downward shift in the high molecular weight DNA band and (b) a subtle smear trailing downward from that band. Quite often these effects are not very clear, but the digest is happening. The key criterion for whether the library is good is the number of PCR cycles required to amplify the target band, later in the protocol – it should be 15 or less for full-representation 2bRAD.

Ligation

In this step adaptors are ligated to the restriction fragments produced above. Note that this is the stage at which reduced tag representation (RTR) must be applied by the choice of adaptor sequences.

1. Prepare double stranded adaptors by combining each pair of primers (illBC-ii, antiBC-ii). The oligonucleotide combination used for each alternative 2b-RAD preparations and sequences of each oligo are provided at the end of this document.
   For generic Adaptor 1, mix 5Ill-NNRW (10 µM) with the same volume of Anti5ill-NNRW (10 µM).
   For barcoded Adaptor 2, mix each par: 3IllBC(1-12) (10 µM) with the same volume of antiBC(1-12) (10 µM).
   Incubate at 42°C for 10 minutes then keep at room temperature until ligation.

2. Prepare 12 master mixes for ligations (one for each barcoded 3' primer). This recipe is for a single reaction, so scale up as needed.

10 mM ATP	0.5 µl
10X T4 ligase buffer	2.0 µl
5 µM Adaptor 1	1.0 µl
5 µM Adaptor 2(1-12)	1.0 µl
T4 DNA ligase (NEB M0202L)	1.0 µl
NFW (nuclease-free water)	14.5 µl

3. Combine 20 µl master mix with the remaining 5 µl of digested DNA (25 µl total volume). Incubate at 16°C for AlfI and BcgI, at 4°C for BsaXI, for at least 2 hours.
   When set up ligation in 96 well plate format we find it useful to prepare 12 master mixes, each with individual Adapter 2(BC), and aliquot it with multi channel pipette. Keep on ice while mixing.

   Incubate in 4°C overnight;

   Then heat at 65^{o}C for 20 min to inactivate the ligase.

Pooling ligations into 12-plex groups
If you used the same 3' barcoded adapters for all samples skip this step, since you will not be pooling samples at the next stage.

Samples barcoded at the 3' end by different ligation Adapters 2 (1-12) can be pooled into groups of 12. Each such group (you will end up with 8 such groups in a 96 well plate) works as a single sample sample in the subsequent Amplification and Purification, so pooling saves quite a bit of effort.

However, to ensure that we are pooling equivalent amounts of each initial sample into a group, it is necessary to quantify them using one of the following procedures:
i. Use standard pcr (see next chapter **Amplification)** with 12-14 cycles and estimate the brightness of the bands on a gel. Mix 5µl (most efficient), 10, 15 or 20µl of each ligation in groups of twelve. Label Groups 1-8.
ii. Alternatively, set up quantitative PCR:

	(volumes given in µl)
H_2O	4.45
SYBR Green mix	7.5
10 mpx primer(10 µM)	0.15
any Ill-Rad-bc primer (1 µM)	1.5
IC1-P5 primer (10 µM)	0.2
IC1-P5 primer (10 µM)	0.2

a. Add 1µl of Ligation, either undiluted or diluted 10 times, to appropriate well as a template. Amplify in a qPCR machine with some NTC (no-template control) using the following profile:

95°C 5 min, (95°C 40 sec, 65°C 2 min, 72°C 30 sec) X 35 cycles

b. Arrange the data in Excel in the form of a table with four columns: sam (sample name), lane (group number), conc (DNA dilution; use 0.066 for undiluted (0.066=1/15) and 0.0066 for 10-fold diluted template), and ct (qPCR result for this sam-conc combination). There must be at least two technical replicates for each combination of sam-conc (i.e. two rows with the same sam and conc and different ct values). The order of columns and rows does not matter, but the names of the columns do matter (note that they are case sensitive).

Export the data from Excel as comma-separated values (.csv). Open script mix_illumina_qpcr.R in R and follow the instructions given in the comments within the script.

Amplification

In this step, the constructs produced by ligation are amplified using a set of four primers that introduce sample (or pooled sample) specific barcodes and the annealing sites for HiSeq amplification and sequencing primers.

Very Important: if you find that it takes more than 15 cycles to amplify a visible target band, consider optimizing previous steps (DNA isolation, digestion, and ligation), since over-amplified RAD samples tend to lose heterozygosity.

Prepare a test-scale PCR to determine optimum cycle number and evaluate relative yield across samples.

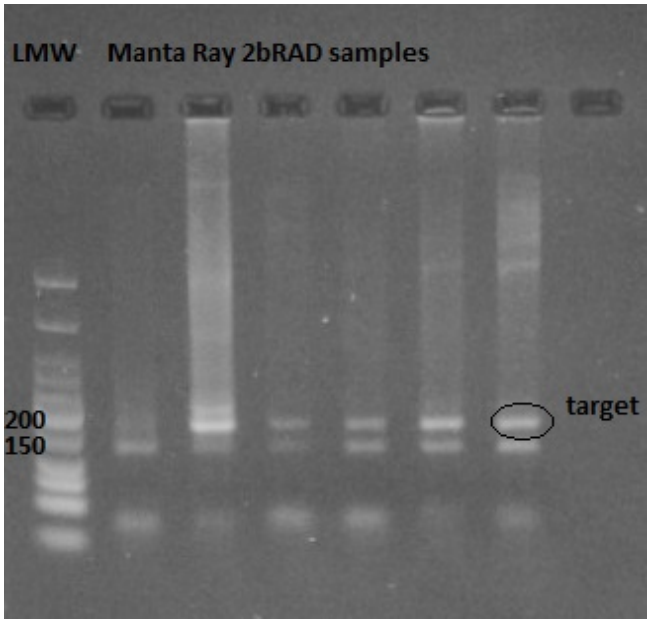
1. For each reaction prepare the following master mix:

NFW	5.2 µl
2.5 mM dNTP	2.5 µl
10 µM IC1-P5	0.4 µl
10 µM IC2-P7	0.4 µl
10 µM Mpx primer	0.3 µl
5X HF buffer	4.0 µl
Phusion polymerase	0.2 µl #NEB M0530L (or M0530S)

2. Combine 13 µl master mix with 4 µl ligation, add 3 µl of 1 µM Ill-Rad-bc (barcode-bearing) primer to each reaction, and amplify as follows: 70°C 30 sec then (95°C 20 sec, 65°C 3 min, 72°C 30 sec) X *N* cycles

3. Sample 5 µl from each reaction at few time points. For example, at 8, 10 and 12 cycles. Load these products on a 2% agarose gel with a low molecular weight marker to confirm molecular weight of PCR product.

4. Select the minimum number of cycles that produces a visible, but not over-amplified, product at ~170 bp. You might also see a 130bp band, which this seems to be an artifact from the carried-over ligase (make sure to heat-inactivate the ligase before mixing the PCR reaction!).



5. Prepare the following master mix for each sample:

NFW	13 µl
2.5 mM dNTP	6 µl
10 µM Mpx primer	0.75 µl
10 µM IC1-P5	1 µl
10 µM IC2-P7	1 µl
5X HF buffer	10 µl
Phusion polymerase	0.5 µl #NEB M0530L (or M0530S)

6. Combine 30 µl master mix with 10 µl ligation (group 1-8), add 7.5 µl of 1 µM Ill-Rad-bc (barcode-bearing) primer individually to each reaction, and amplify using the temperature profile as in 2 above with the cycle number as determined in 4.

Purification

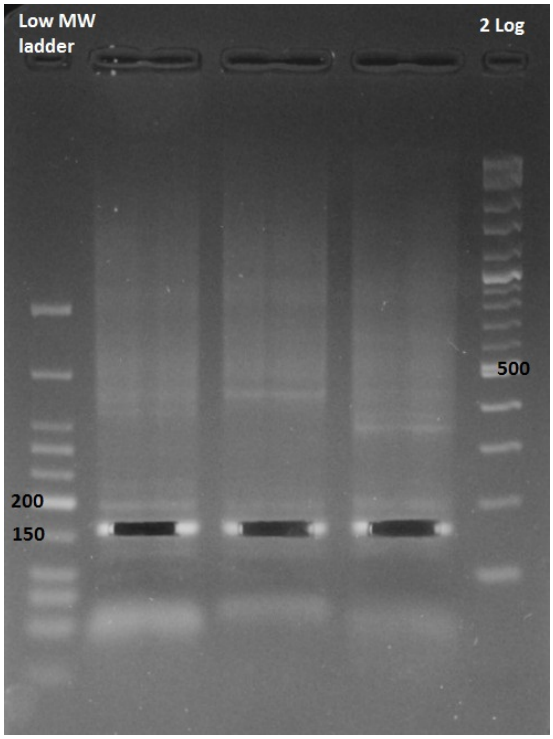
In this step the target band is gel-extracted to exclude high-molecular weight fragments and any chaff that may emerge during PCR (e.g., primer dimers).

1. Prepare a 2% agarose gel using TBE or TB. Use a wide comb that can accommodate 50 µl sample loading plus 10 µl loading dye, or tape together two wells if required.

2. Load the entire volume of each reaction alongside a low-molecular weight ladder. Run gel at low voltage for a long time until bands at 150bp and 200bp will clearly resolved. In our experience 100V for 70 minutes produce the good separation.

3. View the gel briefly (<30 seconds) on a UV or appropriate for your DNA dye black-light transilluminator set at low intensity to verify the presence of target bands and adequate separation of molecular weight standards to resolve bands at ~170 bp and ~130 bp. Typically ~5 cm run distance is sufficient. Photograph.

4. Cut out target band of 170 bp in a narrow gel slice, avoiding the edges of the lane (i.e., cut out the middle 70-75% of the band). Cut just inside the bottom boundary of the target band to avoid getting anything smaller than 170bp.



**Note: at this stage a commercial gel-extraction kit can be substituted for the following three steps, if you feel more confident this way. In our practice, simply soaking the gel slice in water overnight, as described below, works just fine.*

5. Transfer each gel slice into a 1.5 ml microcentrifuge tube and add 20 µl NFW.

6. Make sure gel slice is in contact with water (cut or break it into a few, say 4-5, smaller pieces to make sure they sit comfortably at the bottom of the tube). Hold overnight at 4°C.

7. The following day transfer the eluate (~15 µl) into a new tube. This material is now ready for sequencing, pending qPCR quantification (see the writeup and scripts under "Quantifying samples for sequencing on the same Illumina HiSeq lane")

Oligonucleotides

Name	Sequence (5' – 3')
5ILL-NNRW	CTACACGACGCTCTTCCGATCTNNRWCCNN
Anti5ill-NNRW	GGWYNNAGATCGG/3InvdT[1]/
5ILL-NG[2]	CTACACGACGCTCTTCCGATCTNNRWCCNG
3ILL-NG[2]	CAGACGTGTGCTCTTCCGATCTNG
ILL-Mpx[3]	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT
ILL-RAD-bc[4]	CAAGCAGAAGACGGCATACGAGAT [barcode][5] GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
IC1-P5	AATGATACGGCGACCACCGA
IC2-P7	CAAGCAGAAGACGGCATACGA

1: InvdT: inverted dT to prevent extension by DNA polymerase.
2: These adaptor oligos can be used to reduce representation of the 2b-RAD tags down to approximately 1/16th of the total number. This is a useful cost-lowering trick for applications such as basic population genetics, relatedness analysis, or QTL mapping, the power of which would max out already at a few hundred polymorphic markers.
3: This oligo can be substituted for a standard Illumina - 'universal' primer in TrueSeq v.3 (configuration 5' P5-index2-R1primingSite 3')
4: This oligo can be substituted for a standard TruSeq v.3 barcoded oligo of the configuration 5' P7-Index1-R2primingSite 3'.
5: 'barcode' is a 6-base sequence easily distinguishable from other sequences on that same sequencing run. A list of good working barcodes can be found here:
https://wikis.utexas.edu/display/GSAF/Illumina+-+all+flavors
BUT NOTE that the barcode sequences listed at that site are in fact REVERSE COMPLEMENTS relative to what needs to be written in the ILL-RAD-bc oligo (we had so many fun moments because of that)

"Adaptor 2" oligos with in-read barcodes:

3illBC1	CAGACGTGTGCTCTTCCGATCT ACAC NN
3illBC2	CAGACGTGTGCTCTTCCGATCT GTCT NN
3illBC3	CAGACGTGTGCTCTTCCGATCT TGGT NN
3illBC4	CAGACGTGTGCTCTTCCGATCT CACT NN
3illBC5	CAGACGTGTGCTCTTCCGATCT GATG NN
3illBC6	CAGACGTGTGCTCTTCCGATCT TCAC NN
3illBC7	CAGACGTGTGCTCTTCCGATCT CTGA NN
3illBC8	CAGACGTGTGCTCTTCCGATCT AAGC NN
3illBC9	CAGACGTGTGCTCTTCCGATCT GTAG NN
3illBC10	CAGACGTGTGCTCTTCCGATCT GACA NN
3illBC11	CAGACGTGTGCTCTTCCGATCT GTGA NN
3illBC12	CAGACGTGTGCTCTTCCGATCT AGTC NN
antiBC1	GTGT AGATCGGA/3InvdT/
antiBC2	AGAC AGATCGGA/3InvdT/
antiBC3	ACCA AGATCGGA/3InvdT/

antiBC4	AGTG AGATCGGA/3InvdT/
antiBC5	CATC AGATCGGA/3InvdT/
antiBC6	GTGA AGATCGGA/3InvdT/
antiBC7	TCAG AGATCGGA/3InvdT/
antiBC8	GCTT AGATCGGA/3InvdT/
antiBC9	CTAC AGATCGGA/3InvdT/
antiBC10	TGTC AGATCGGA/3InvdT/
antiBC11	TCAC AGATCGGA/3InvdT/
antiBC12	GACT AGATCGGA/3InvdT/

Some possible sequences of barcoded PCR primers (standard TruSeq v.3 oligos: 5′ P7-**Index1**-R2primer 3′)

ILL-RAD01	CAA GCA GAA GAC GGC ATA CGA GAT **CGT GAT** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD02	CAA GCA GAA GAC GGC ATA CGA GAT ACA TCG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD03	CAA GCA GAA GAC GGC ATA CGA GAT GCC TAA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD04	CAA GCA GAA GAC GGC ATA CGA GAT TGG TCA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD05	CAA GCA GAA GAC GGC ATA CGA GAT CAC TGT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD06	CAA GCA GAA GAC GGC ATA CGA GAT ATT GGC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD07	CAA GCA GAA GAC GGC ATA CGA GAT GAT CTG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD08	CAA GCA GAA GAC GGC ATA CGA GAT TCA AGT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD09	CAA GCA GAA GAC GGC ATA CGA GAT CTG ATC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD10	CAA GCA GAA GAC GGC ATA CGA GAT AAG CTA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD11	CAA GCA GAA GAC GGC ATA CGA GAT GTA GCC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD12	CAA GCA GAA GAC GGC ATA CGA GAT TAC AAG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT

Quantification for mixing on the same HiSeq lane

Run Quant-IT picogreen DS DNA assay (Life Technologies P7589) to determine the final concentrations of the eluted product in order to mix libraries in equal proportions.

PicoGreen assay Protocol:

1) Place 100ul 1X TE into all first column wells except B1.

2) Add 150ul of DNA standard (@ 2ug/ml, which is the same as 2ng/ul) into B1.

3) Serially dilute standards by taking 50ul of B1, mixing into C1, taking 50ul of C1, mixing into D1, and so on until taking 50ul from H1 and throwing it out.

4) To all sample wells, add 98ul of 1X TE.

5) Add 2ul sample DNA to sample wells.

6) Mix Pico Green Master mix: 99.5ul 1XTE + 0.5ul PicoGreen for one sample. Multiply accordingly (plus 8 wells for DNA standard).

7) Add 100ul of master mix to all standard and sample wells, bringing up final volumes in each well to 200.

8) Read the fluorescence (excitation 480nm, emission 520nm). We use SpectraMax M2 plate reader and Costar assay plates 96 well, no lid, flat bottom, non-treated black with black bottom (Corning 3650) or clear bottom (Corning 3631).

9) Save the data into txt file, assemble the results in Excel in two-column form – well, reading - save it as comma-delimited (.csv) file. The file must contain all A1-H1 wells (blank and calibrators) plus an arbitrary number of sample wells, in any order. See file picogreen.csv as an example.

10) Use picogreen.R script to calculate sample concentrations (ng/ul in the original sample).