

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

Latest update (June 2018): removed all interim quantification steps. Instead, equalize your input DNA samples really well (based on Qbit or picogreen, not just Nanodrop) and ensure that their OD 260/230 ratio is ≥ 2 .

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 can be inactivated by heating, 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:

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.
3. Heat-inactivate the ligase and pool the ligations with different 3' barcodes into groups of 12.
4. Amplification and barcoding of pooled ligations.
5. Purification of the target 165 bp band by Pippin or gel-electrophoresis (the only purification step in the whole procedure).

NOTE on experimental design: we strongly recommend including three pairs of genotyping replicates in each 2bRAD experiment. Replicates should be independently prepped DNA samples from the same individual. They will be extremely useful at the analysis step to assess the overall accuracy of genotyping, set quality filtering criteria, and quantify relatedness between samples.

Digest

1. Prepare samples each containing 100 ng of DNA in 4 μ l. Equalize your input DNA samples well (based on qbit or picogreen, not just nanodrop) and ensure that their OD 260/230 ratio is ≥ 2 . Concentrating can be accomplished using by ethanol precipitation or by drying under vacuum.

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

NEB Buffer #3	0.6 μ l
150 μ M SAM	0.4 μ l
H ₂ O	0.5 μ l
BcgI (1 U μ l ⁻¹)	0.5 μ l

**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 final single-reaction volume at 6 μ l.*

3. Combine 2 μ l master mix with each 4 μ l DNA sample (6 μ l total volume). Cover the plate with PCR film, incubate at 37°C (in air incubator or in a thermocycler with heated lid) for 1 hr.
4. If using a heat-sensitive enzyme (e.g. BcgI, AlfI), inactivate the enzyme at 65°C for 10 min then let the thermocycler return to room temp while keeping the heated lid on. Hold samples on ice after this.

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 (iIlBC-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. Pre-mixed adaptors can be stored just like normal oligos.

For generic Adaptor 1, mix 5iIl-NNRW (10 μ M) with the same volume of Anti5iIl-NNRW (10 μ M).

For barcoded Adaptor 2, mix each pair: 3iIlBC(1-12) (10 μ M) with the same volume of antiBC(1-12) (10 μ M).

Incubate at 42°C for 5 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 μ l
10X T4 ligase buffer	2.0 μ l
5 μ M Adaptor 1	1.0 μ l
5 μ M Adaptor 2(1-12)	1.0 μ l
T4 DNA ligase (NEB M0202L)	1.0 μ l
NFW (nuclease-free water)	14.5 μ l

3. Combine 20 μ l master mix with digested DNA (~25 μ l total volume). For 96 well plate format we find it useful to prepare 12 master mixes, each with individual Adapter 2(BC), and distribute it with 8-channel pipette. Keep on ice while mixing.

Incubate at 16°C for AlFI and BcgI, at 4°C for BsaXI, for at least 2 hours and up to overnight.

Then heat at 65°C for at least 20 min to inactivate the ligase (in thermocycler with heated lid to .

Pooling ligations into 12-plex groups

Samples barcoded at the 3' end by different ligation Adapters 2 (1-12) are now pooled into groups of 12: each row of a 96-well plate, 3 µl from each well, is pooled into one sample. The result is 8 pooled samples for the initial 96 well plate.

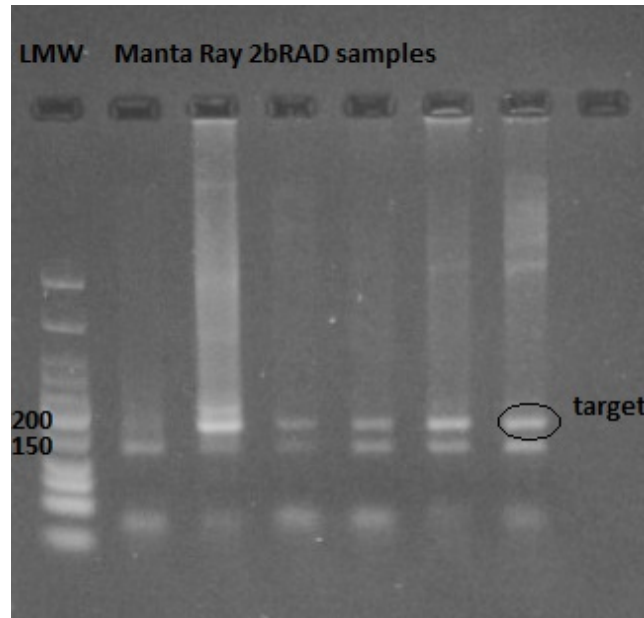
Amplification

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

1. For each reaction prepare the following master mix:

NFW	7.2 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
10x Titanium buffer	2.0 µl
Titanium Taq	0.2 µl (#Clontech 639208)

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 14 cycles
3. Load 5 µl on a 2% agarose gel alongside LMW ladder (NEB N3233S) or other marker that has 150 and 200 bp bands - see next page.



NB: There should be no substantial variation in product amount among samples at this stage (unlike this gel photo); if there is, something went wrong with DNA quantity/quality of some original samples.

Confirm that all samples have a visible band at ~160-170 bp. You might also see a band below 150 bp, which is an artifact from the carried-over ligase (make sure to heat-inactivate the ligase before mixing the PCR reaction!)

If the product looks over-amplified (as bright or brighter than the second lane in the gel photo above), re-do amplification with two less cycles. If the 165 b product is visible but barely, simply add two more cycles to the same reactions.

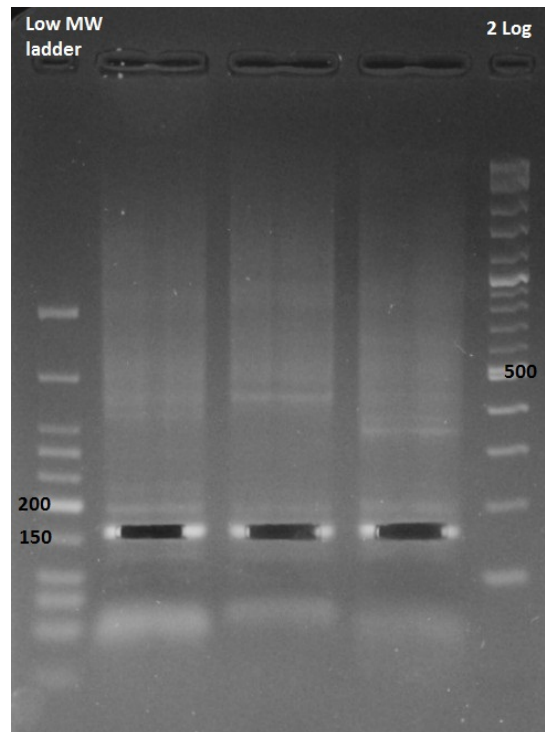
If the 165 b product is not visible at all after 14 cycles, do not add more cycles. Instead, optimize previous steps: make sure that ligase and restriction enzyme did not expire, consider adding an extra purification step during DNA isolation, and make sure you start with at least 50-100 ng on DNA. Even though weak samples could be amplified with more PCR cycles, such samples would have poor representation of allelic diversity, i.e. will show lower counts after removal of PCR duplicates, and will often miss a second allele in a heterozygote.

If all is well (all samples have visible 165 bp band of the same intensity), all reactions can be pooled together and the target band extracted with PippinPrep or BluePippin (155-170 bp range). Most sequencing core facilities would do it for you. If not, gel-purify the band as described below.

Gel-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 30-50 μ l, or simply tape together two wells.
2. Load 30-50 μ l of sample (40 μ l sample + 10 μ l loading dye) alongside LMW 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 (<20 seconds) on a UV or (possibly longer) on appropriate for your DNA dye black-light transilluminator set at low intensity to verify the presence of target band and adequate separation of molecular weight standards to resolve bands at ~165 bp and (possibly) below 150 bp. Typically ~5 cm run distance is sufficient. Photograph.
4. Cut out target ~165 bp band 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.



**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 μ l 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 μ l) into a new tube. This material is now ready for sequencing.

Oligonucleotide sequences for Illumina HiSeq

Name	Sequence (5' – 3')
5ILL-NNRW	CTACACGACGCTCTTCCGAT CTNNRWCCNN
Anti5ill-NNRW	GGWYNNAGATCGG/3InvdT ¹ /
3ILL-NN	CAGACGTGTGCTCTTCCGAT CTNN
anti-ILL	AGATCGGAAGAGC/3InvdT ¹ /
5ILL-NG ²	CTACACGACGCTCTTCCGATCTNNRWCCNG
3ILL-NG ²	CAGACGTGTGCTCTTCCGATCTNG
ILL-Mpx ³	AATGATACGGCGACCAACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT
ILL-RAD-bc ⁴	CAAGCAGAAGACGGCATACGAGAT [barcode] ⁵ GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
IC1-P5	AATGATACGGCGACCAACCGA
IC2-P7	CAAGCAGAAGACGGCATACGA

1 InvdT: inverted dT to prevent extension by DNA polymerase.

2 These two adaptors 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 is a standard Illumina "universal" primer in TrueSeq v.3 (configuration 5' P5-index2-R1primingSite 3')

This oligo can be substituted for oligos bearing additional barcode, for example:

TruSeq_Un1	AATGATACGGCGACCAACCGAGATCTACAC ATCAGG ACACCTCTTCCCTACACGACGCTCTTCCGATCT
TruSeq_Un2	AATGATACGGCGACCAACCGAGATCTACAC ACTTGA ACACCTCTTCCCTACACGACGCTCTTCCGATCT
TruSeq_Un3	AATGATACGGCGACCAACCGAGATCTACAC GGCTAC ACACCTCTTCCCTACACGACGCTCTTCCGATCT

4 "bc" stands for "barcode", and is typically replaced by the barcode number (or other barcode-specific identifier) in the actual name of the primer. This primer can be substituted for a standard TruSeq v.3 barcoded oligo of the configuration 5' P7-Index1-R2primingSite 3'.

5 barcode: 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)

Secondary-barcoded adapters for Ligation (to be used instead of 3ILL-NN and anti-ILL)

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/

Barcoded PCR primers (can be substituted for 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
ILL-RAD13	CAA GCA GAA GAC GGC ATA CGA GAT TTG ACT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD14	CAA GCA GAA GAC GGC ATA CGA GAT GGA ACT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD15	CAA GCA GAA GAC GGC ATA CGA GAT TGA CAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD16	CAA GCA GAA GAC GGC ATA CGA GAT GGA CGG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD17	CAA GCA GAA GAC GGC ATA CGA GAT CTC TAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD18	CAA GCA GAA GAC GGC ATA CGA GAT GCG GAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD19	CAA GCA GAA GAC GGC ATA CGA GAT TTT CAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD20	CAA GCA GAA GAC GGC ATA CGA GAT GGC CAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD21	CAA GCA GAA GAC GGC ATA CGA GAT CGA AAC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD22	CAA GCA GAA GAC GGC ATA CGA GAT CGT ACG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD23	CAA GCA GAA GAC GGC ATA CGA GAT CCA CTC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD24	CAA GCA GAA GAC GGC ATA CGA GAT GCT ACC GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD25	CAA GCA GAA GAC GGC ATA CGA GAT ATC AGT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD26	CAA GCA GAA GAC GGC ATA CGA GAT GCT CAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD27	CAA GCA GAA GAC GGC ATA CGA GAT AGG AAT GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD28	CAA GCA GAA GAC GGC ATA CGA GAT CTT TTG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD29	CAA GCA GAA GAC GGC ATA CGA GAT TAG TTG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD30	CAA GCA GAA GAC GGC ATA CGA GAT CCG GTG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD31	CAA GCA GAA GAC GGC ATA CGA GAT ATC GTG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD32	CAA GCA GAA GAC GGC ATA CGA GAT TGA GTG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD33	CAA GCA GAA GAC GGC ATA CGA GAT CGC CTG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD34	CAA GCA GAA GAC GGC ATA CGA GAT GCC ATG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT

[illegible]

ILL-RAD90	CAA GCA GAA GAC GGC ATA CGA GAT TAT AGA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD91	CAA GCA GAA GAC GGC ATA CGA GAT TTT GCA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD92	CAA GCA GAA GAC GGC ATA CGA GAT GTG CCA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD93	CAA GCA GAA GAC GGC ATA CGA GAT CTA ACA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT
ILL-RAD94	CAA GCA GAA GAC GGC ATA CGA GAT ATA GAA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG AT