**Supporting Information for online publication**

1. **Table S1.** Characteristics of twelve microsatellite markers used for parentage analyses.
2. **Table S2.** Recovery of non-overlapping loci between Groups A and B when filtering parameters (r, m, and a) are applied sequentially, and when relaxed.
3. **Table S3.** Overlap in the RAD loci that were obtained while varying different steps of the protocol (size selection and restriction enzymes) and filter parameters (r, m and a) in the populations program from the stacks pipeline.
4. **Table S4.** Number of loci retained in ALL catalog by varying filter parameters (r, m, and a) in populations.
5. **Supplementary ddRAD-seq Protocol**
6. **Supplementary Bioinformatic Pipeline**

|  |
| --- |
| **Table S1.** Characteristics of twelve microsatellite markers used for parentage analyses from 769 variegated fairy-wrens sampled from 2012 to 2016 at Lake Samsonvale, Queensland, Australia. |
| Locus\* | Repeat motif | Forward (F) and Reverse (R) primer sequences | No. of alleles | Size range (bp) |
| *MaLa02* | AGAT | F- AAGCAGCCAAGATGTTAGTTTCC | 32 | 172-316 |
|  |  | R- AATGTGGATCGGCTAAAGAATGC |  |  |
| *MaLa03* | AATG | F- TTGTCTATGTATGACAACGTGGC | 8 | 264-316 |
|  |  | R- GACAGACTGCAATGTTCCTGTAC |  |  |
| *MaLa04* | AATG | F- ACGGACTGAAGTAATACAAGCAC | 8 | 302-346 |
|  |  | R- TGAGCCTTTGAATGTACTCTCAC |  |  |
| *MaLa05* | AAGG | F- ACTTGGAAGGGTGAGGATTTATTC | 20 | 165-275 |
|  |  | R- TGTTCTTCCTTGATGTAACCAACC  |  |  |
| *MaLa06* | ATCC | F- ATCCATCTCCACAGTACCCTTAC | 11 | 152-212 |
|  |  | R- ATCAGGGTGTCACTATGTCCTTG |  |  |
| *MaLa07* | AAGT | F- TATGACCGGTGAGATTCCTGTAG | 7 | 257-289 |
|   |  | R- GCAGTTCTGCTAAACACATATGC |   |  |
| *MaLa08* | ACT | F- AAAGGTCAAATCCATGGGTTGAG | 13 | 200-248 |
|   |  | R- AACATACTCCCTTATCCCATTCAG |   |  |
| *MaLa10* | AATAG | F- AAATCCACCATTGTTTAGGAGGC | 26 | 258-330 |
|   |  | R- TCTTAACAGAGGAATGGGTTTGC |   |  |
| *MaLa13* | ACAG | F- ATTACCACATTTGTCTGCCTGTC | 23 | 290-374 |
|   |  | R- TGACCTGTTTGTAAAGGTTCACC |   |  |
| *MaLa14* | AGAT | F- GTGGCACTGATAAATTTCTATGGC | 21 | 199-309 |
|   |  | R- TCAGTGATGATTTCTTTGGCTGG |   |  |
| *MaLa16* | AAGG | F- AAAGGACAGGGAAAGGGAAGTAG | 14 | 167-223 |
|   |  | R- CTGTGAGTAAAGCCATGGTCAAG |   |  |
| *MaLa18* | AACC | F- AGCAGCAAACAGATATCCAAGTG | 13 | 234-282 |
|   |   | R- TTTCATCTGCATGGTTCCACTG |   |  |
| \* Microsatellites were multiplexed at an annealing temperature of 60°C (*MaLa02*, *MaLa03*, *MaLa07*, and *MaLa18*; *MaLa04*, *MaLa06*, *MaLa08*, and *MaLa10; MaLa05*, *MaLa13*, *MaLa14*, and *MaLa16*). A pigtailed sequence (GTT TCT) was added to the 5’ end of all reverse primers to reduce genotyping error, by adding stability to the PCR reaction. No loci deviated from Hardy-Weinberg equilibrium or were in linkage disequilibrium. |

**Table S2.** Recovery of non-overlapping loci between Groups A and B when filtering parameters (r, m, and a) are applied sequentially, and when relaxed. 797 loci are present in Group A after strict filtering, and 645 loci in Group B (see Table 3 in text). 549 of these loci are overlapping (Table 3), leaving 248 from Group A that are non-overlapping with Group B when strict filter parameters are used. BLASTN was used to determine the number of non-overlapping loci from Group A that were present in Group B, when varying filter parameters.

|  |  |
| --- | --- |
|  |  |
| Non-overlapping loci from Group A present in Group B  | Database  | Loci present in database |
| 241 | Group B catalog | 25521 |
| 121 | Group B filtered r 0.95 | 2169 |
| 33 | Group B filtered r 0.95 m 10 | 1667 |
| 8 | Group B filtered r 0.95 m 10 a 0.25 | 706 |
| 214 | Group B filtered r 0.60 | 4685 |

**Table S3.**Overlap in the RAD loci that were obtained while varying different steps of the protocol (size selection and restriction enzymes) and filter parameters (r, m and a) in the populations program from the stacks pipeline. The first column of tables shows the number of loci recovered before and after filtering for SNPs in Hardy-Weinberg Equilibrium (HWE). In the second column of tables, the diagonal indicates the total number of loci recovered for each treatment. Values above the diagonal represent the percent overlapping loci between groups (relative to the group with the smallest number of loci), while values below the diagonal list the number of loci that were overlapping between groups.

|  |  |
| --- | --- |
| **Loci Recovered with different filtering parameters during populations:** | **Overlap of loci (HWE filtered) between index groups:** |
| **r 0.95 m 5 a 0.25** |  |  |  | **r 0.95 m 5 a 0.25** |  |  |  |  |  |
| Index group | # loci before HWE | # loci after HWE | % retained |  | Index group | **all** | **A** | **B** | **C** | **D** | **E** |
| all | 685 | 506 | 73.9 |  | **all** | 506 | 83.6 | 75.9 | 80.2 | 1.6 | 1.8 |
| A | 1046 | 937 | 89.6 |  | **I1** | 423 | 937 | 85.3 | 81.9 | 2.4 | 3.3 |
| B | 872 | 787 | 90.3 |  | **I9** | 384 | 671 | 787 | 86.5 | 2.0 | 3.0 |
| C | 1882 | 1707 | 90.7 |  | **I10** | 406 | 767 | 681 | 1707 | 4.9 | 5.1 |
| D | 740 | 699 | 94.5 |  | **I11** | 8 | 17 | 14 | 34 | 699 | 70.0 |
| E | 649 | 607 | 93.5 |  | **I12** | 9 | 20 | 18 | 31 | 425 | 607 |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **r 0.80 m 10 a 0.25** |  |  |  | **r 0.80 m 10 a 0.25** |  |  |  |  |  |
| Index group | # loci before HWE | # loci after HWE | % retained |  | Index group | **all** | **A** | **B** | **C** | **D** | **E** |
| all | 1220 | 742 | 60.8 |  | **all** | 742 | 82.3 | 70.6 | 79.5 | 2.6 | 2.3 |
| A | 1374 | 1171 | 85.2 |  | **I1** | 611 | 1171 | 86.1 | 83.3 | 2.9 | 3.4 |
| B | 1158 | 995 | 85.9 |  | **I9** | 524 | 857 | 995 | 87.2 | 2.4 | 2.8 |
| C | 2353 | 2019 | 85.8 |  | **I10** | 590 | 976 | 868 | 2019 | 5.5 | 5.1 |
| D | 1158 | 1032 | 89.1 |  | **I11** | 19 | 30 | 26 | 57 | 1032 | 78.5 |
| E | 985 | 860 | 87.3 |  | **I12** | 17 | 29 | 24 | 44 | 675 | 860 |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **r 0.95 m 10 a 0.05** |  |  |  | **r 0.95 m 10 a 0.05** |  |  |  |  |
| Index group | # loci before HWE | # loci after HWE | % retained |  | Index group | **all** | **A** | **B** | **C** | **D** | **E** |
| all | 1156 | 910 | 78.7 |  | **all** | 910 | 89.2 | 82.0 | 81.0 | 1.6 | 1.2 |
| A | 1810 | 1664 | 91.9 |  | **I1** | 812 | 1664 | 87.8 | 79.7 | 2.6 | 2.4 |
| B | 1498 | 1398 | 93.3 |  | **I9** | 746 | 1227 | 1398 | 85.9 | 2.3 | 2.3 |
| C | 3311 | 3069 | 92.7 |  | **I10** | 737 | 1326 | 1201 | 3069 | 4.9 | 4.8 |
| D | 1474 | 1401 | 95.0 |  | **I11** | 15 | 36 | 32 | 69 | 1401 | 73.9 |
| E | 1204 | 1145 | 95.1 |  | **I12** | 11 | 27 | 26 | 55 | 846 | 1145 |

|  |
| --- |
| All: 160 samples; Sbf1/Msp1; 450-600 bp. |
| A: 20 samples; Sbf1/Msp1; 450-600 bp (Index 1). |
| B: 20 samples; Sbf1/Msp1; 450-600 bp (Index 9). |
| C: 20 samples; Sbf1/Msp1; 400-700 bp (Index 10). |
| D: 20 samples; SbfI/EcoRI; 450-600 bp (Index 11). |
| E: 20 samples; SbfI/EcoRI; 450-600 bp (Index 12). |

Populations program parameters:

r: minimum percentage of individuals in a population required to process a locus for that population.

m: minimum stack depth required for individuals at a locus

a: minimum minor allele frequency required to process a nucleotide site at a locus (0 < a< 0.5)

**Table S4.** Number of loci retained in ALL catalog by varying filter parameters (r, m, and a) in populations. Filter parameters were applied separately to the entire catalog (including both invariant and variant loci) and to variant loci where only the first SNP was selected per locus. The ALL catalog contains 49662 loci before filtering.

|  |  |
| --- | --- |
|  | filter parameter |
| none | stack depth (m 5) | stack depth (m 10) | minor allele freq (a 0.25) | 5 % missing data (r 0.95) | 20 % missing data (r 0.80) |
| loci removed for which all variant sites were filtered | 544 | 544 | 2301 | 11258 | 32 | 25 |
| loci removed that did not pass sample/population constraints | 4 | 4 | 24192 | 4 | 47160 | 45027 |
| **Loci retained and processed** | **49114** | **49114** | **23169** | **38400** | **2470** | **4610** |
| Loci with variant sites/SNPs (restricted to only the first SNP per locus) | 21451 | 21451 | 14091 | 10737 | 2097 | 3988 |
| % loci retained from catalog | 99% | 99% | 47% | 77% | 5% | 9% |
| % loci with variant sites/SNPs retained from catalog | 43% | 43% | 28% | 22% | 4% | 8% |
| % retained loci with variant sites/SNPs | 44% | 44% | 61% | 28% | 85% | 87% |

**Supplementary ddRAD-seq Protocol**

**DNA extraction**

There are many methods of DNA extraction that would be suitable. Highly quality DNA (i.e. with little degradation) and DNA concentrations around 20-50ng/ul are preferred.

We extract DNA using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturers protocol with the following changes. To reduce the costs of DNA extraction we use bulk reagents from Qiagen and UPrep® Spin Columns (LPS, catalog # M858143). Approximately 100ul of blood (stored in Queens lysis buffer at RT) is added to 150ul 1X Phosphate Buffered Saline (PBS) and digested with 200ul Buffer AL (Qiagen) and 20ul of 20mg/ml Proteinase K (VWR) overnight at 64 degC. In the final step DNA is eluted in warmed nuclease-free water (Qiagen). Two elutions of 80ul (E1) and 150ul (E2) are performed into separate 1.5ml tubes.

Check the quality of the DNA by running 1-2ul of the extracted DNA on a 1% agarose gel. The DNA should run as one high molecular weight band with little degradation (seen as smearing or the presence of low molecular weight products on the gel).

**Quantify and normalize genomic DNA**

Two microliters of DNA is used to determine the DNA concentration of the first elutions (E1) using the Qubit dsDNA BR Assay Kit and the Qubit® Fluorometer (Life Technologies) following the manufacturers protocol. Those samples with a concentration higher than 30ng/ul are diluted to approximately 25ng/ul with nuclease-free water. For those samples with concentrations lower than 10ng/ul, both elutions are pooled and the DNA concentrated by evaporation using an Eppendorf Vacufuge. Concentrations of these concentrated samples are determined as above and if necessary further concentration or dilution performed.

**Order and prepare adapters and index primers.**

See table 2 below for a list of oligos to order. We order these from IDT (idtdna.com) at a scale of 100nmole with standard desalting.

Resuspend oligos to 100uM in 0.5X AE Buffer (Qiagen). (add buffer at 10X nmoles of primer (in ul). E.g. for a primer with 95.4 nmoles add 954ul 0.5XAE).

For the P1 and P2 adapters you need to anneal the complementary oligos together.

For P1 adapters (Final concentration =5uM):

1. In a PCR tube mix equal volumes of each oligo together to get a final concentration of 50uM of adapter. E.g. 20ul of P1F\_X and 20ul of P1R\_X
2. In a thermocylcer heat the mixture at 80 degC for 1 min
3. Remove tube from the thermocylcer and place on a rack at RT.
4. Allow to cool to RT for at least 30min.
5. Further dilute the adapter to 5uM with 0.5 X AE in 1.5ml tube. E.g Add 40ul of annealed adapter and 360ul 0.5X AE.
6. Store adapters at 4 degC or in freezer for long term storage.

For P2 adapter (Final concentration =25uM): Note: you will need to make several batches of P2 adapter to have enough for each experiment (you need approximately 250ul per 80 samples)

1. In a PCR tube mix equal volumes of each oligo together to get a final concentration of 50uM of adapter. E.g. 50ul of P2F\_MspI and 50ul of P2R\_MspI.
2. In a thermocylcer heat the mixture at 80 degC for 1 min
3. Remove tube from the thermocylcer and place on a rack at RT.
4. Allow to cool to RT for at least 30min.
5. Further dilute the adapter to 25uM with 0.5 X AE in 1.5ml tube. E.g Add 100ul of annealed adapter and 100ul 0.5X AE.
6. Store adapter at 4 degC or in freezer for long term storage.

For the PCR primers (P1\_PCR\_for and P-indexX) you will need to dilute these primers to 5uM.

This is a 1:20 dilution. Eg. 10ul of oligo + 190ul 0.5X AE.

**Digest Samples with *SbfI* and *MspI* and ligate P1 and P2 adapters**

Set up arrangement of samples in 96 well plate:

* Samples should be grouped in groups of 20 (each sample will carry a unique barcode which is part of the P1 adapter) that will be pooled and have unique index sequences added (during PCR enrichment). We have 20 P1 adapters and 12 Index primers, making a total of 240 DNA samples that can be run in one lane.
* If you have samples with large differences in DNA concentration, group samples with similar DNA concentrations together in one index group.
* Prepare a spreadsheet, recording the P1 adapter, index group and plate location for each sample

Example of our plate layout for 160 samples (8 index groups)



Consumables:

96 well PCR plate

Adhesive sealing sheets (ThermoFisher Scientific, AB0558)

strip tubes or single PCR tubes

caps for strip tubes

2ml tubes

1.5ml tubes

10X CutSmart Buffer (NEB) (provided with Enzymes)

20U/ul SbfI-HF (NEB, R3642L)

20U/ul MspI (NEB, R0106S)

10mM ATP (NEB, P0756L)

400U/ul T4 DNA ligase (NEB, M0202L)

gDNA samples (normalized, usually approx. 10-30ng/ul in a total volume of at least 25ul)

5uM P1 adapters (see below)

25uM P2 adapter (see below)

Nuclease-free water (Qiagen, 129115)

(1% agarose gel, 6X loading dye, 1XTAE, 1kb ladder)

Prepare 1:20 dilutions of P1 adapters:

1. This can be done in a strip of PCR tubes to allow you to use the multichannel pipette to dispense the P1 adapters. (You will need 20 tubes if using all the P1 adapters)
2. Label tubes.
3. Add 19ul of nuclease free water to each tube.
4. Add 1ul of P1 adapter to corresponding one tube above. Repeat for all P1 adapters being used.
5. Seal tubes, vortex briefly to mix, centrifuge.

Setup digestion-ligation reaction:

1. A standard digestion-ligation reaction has the following:

|  |  |
| --- | --- |
| **Reagents** | **Volume (ul) for one reaction** |
| Nuclease Free water | Up to total volume |
| 10X CutSmart Buffer | Dilute to 1X (vol depends on total vol) |
| 25uM P2 adapter | 2.5 |
| 10mM ATP | 1mM (volume depends on total vol) |
| 20U/ul SbfI-HF | 0.75 |
| 20U/ul MspI | 0.75 |
| 400U/ul T4 DNA ligase | 0.75 |
| Normalized DNA | Approx. 200-500ng (vol depends on conc.) |
| TOTAL | 20-40ul |

**The remainder of this protocol is an example where our reaction volume is 40ul and we add 20ul of DNA at approximately 20ng/ul.**

1. Make a master mix of everything except the P1 adapters and DNA in a 2ml tube.

Make the master mix for a few extra reactions so that you do not run out. We usually do our reactions in plates with 80 samples per plate.

|  |  |  |
| --- | --- | --- |
| Reagents | Volume (ul) for one reaction | Volumes (ul) for Master mix for 80 samples (90 rxns) |
| Nuclease Free water | 6.25 | 562.5 |
| 10X CutSmart Buffer | 4 | 360 |
| 25uM P2 adapter | 2.5 | 225 |
| 10mM ATP | 4 | 360 |
| 20U/ul SbfI-HF | 0.75 | 67.5 |
| 20U/ul MspI | 0.75 | 67.5 |
| 400U/ul T4 DNA ligase | 0.75 | 67.5 |
| TOTAL | 19 | 1710 |

1. Vortex briefly to mix and centrifuge
2. Aliquot 19ul of master mix to each well of a 96 well PCR plate. (note: we do 80 samples per plate and therefore leave column 1 and column 12 empty)

If you would prefer to use a multichannel pipette to aliquot the master mix:
First divide your master mix into a strip of 8 PCR tubes (210ul per tube), and then use this strip of 8 tubes as the reservoir to aliquot 19ul per well using the 8-channel multichannel pipette.
3. Add 1ul of P1 adapter to the corresponding well of the PCR plate (make sure that each group of 20 digestion-ligation reactions receives a unique P1 adapter). If you set the P1 adapters up in a strip of PCR tubes then you can use a multichannel pipette to add the adapters. In our case each column of the PCR plate has a different P1 adapter.
4. Add 20ul of gDNA to each well. (**Note:** every well in the plate will have a different DNA sample).
5. Use a multichannel pipette, set to 30ul, and gently pipette up and down a few times to mix each reaction.
6. Seal the PCR plate well using adhesive seal.
7. Place on the thermocylcer and run the following program:
Choose the preheated lid option and set to 100 degC
37 degC for 30min
20 degC for 60min
hold at 4 degC
8. If desired, select three samples from each plate and check for digestion on a gel:
-randomly select three wells from your PCR plate and remove 2ul of the reaction into a tube containing 2ul of gel loading dye.
-for comparison take 1ul of the corresponding undigested gDNA and add 2ul of gel loading dye.
-Run these on a 1% agarose gel along with a marker. You want to see a smear for the digested sample compared to the mostly intact undigested DNA.



**Pool reactions within Index groups and clean up digested DNA.**

Consumables:

SPRI beads (Agencourt AMPure XP, Beckman Coulter, A63880) or home made MagNA (see protocol at end of methods)

Freshly prepared 70% Ethanol

1.5ml and 2ml tubes

screw-capped tube required by core facility for sample submission

AE Buffer (Qiagen)

Qubit dsDNA BR assay kit (ThermoFisher Scientific, Q32853)

Equipment:

Magnetic separation rack for 1.5ml tubes (e.g. NEB, S1509S)

Qubit Fluorometer (ThermoFisher Scientific)

\*BluePippin (Sage Science)

\* *In our case these are services provided by the Biotechnology Resource Center Genomics Facility at Cornell University (http://www.biotech.cornell.edu/brc/genomics-facility)*

Pool reactions within index groups and clean up DNA.

1. Remove SPRI beads from refrigerator and allow to warm to room temperature.
2. Combine the digestion-ligation reactions from one index group (20 samples with unique adapters) in 2ml tube:
In our example above: all the digestion-ligation reactions from row A in both plate 1 and plate 2 are combined for Index 1
We will have a total of 800ul in the tube (40ul X20).
This will be repeated with row B (index 2), row C, (index 3), etc.
You will have up to 12 tubes each with 800ul of digested-ligated samples.
3. Mix SPRI beads well
4. Add 1.5X volumes of SPRI beads to the 2ml tube with the combined digested-ligated samples (mix well each time you pipette, to ensure that beads are even distributed in the mixture).
In our case we will add 1.2ml SPRI beads to each tube containing 800ul digested-ligated sample.
5. Mix gently by pipetting up and down or turning tube over.
6. Incubate at room temperature for 5min
7. Place tube on the magnetic rack to capture the beads. This may take many minutes. When the beads are captured to the side of the tube and the supernatant is clear you are ready to proceed.
8. Remove and discard the supernatant (keeping the tube on the magnetic rack)
9. Wash the beads by pipetting 2ml 70% Ethanol into the tubes. Keep the tubes on the magnetic rack.
10. Incubate for 1min and then remove and discard the ethanol
11. Repeat the Ethanol wash (steps 9 and 10)
12. Remove as much ethanol as possible and allow the beads to air-dry for 5-10 mins (keep tubes in the magnetic rack).
13. Remove the tube from the magnetic rack and resuspend the beads in 36ul AE buffer. Pipette repeatedly over the side of the tube to resuspend all beads and wash the beads down into the suspension.
14. Incubate at room temperature for 2min.
15. Return the tubes back to the magnetic rack and incubate until the beads are captured on the side of the tube and the supernatant is clear.
16. Transfer 34ul of the supernatant to a new tube (screw capped tube for submission to core facility).

Determine concentration after clean-up

Determine the concentration of your DNA using the Qubit dsDNA BR Assay kit and 2ul of cleaned up DNA (from above). (Concentration is usually around 100ng/ul).

Submit to core facility for size selection.

Request size selection in the range of 450-600bp using BluePippin. Note this is a narrow window and to ensure that the same loci are recovered in all index groups it is important to be careful to have the same size range selected for each index group.

**Low cycle PCR to add index groups.**

Consumables

Qubit dsDNA HS assay kit (ThermoFisher Scientific, Q32851)

Phusion® High-Fidelity PCR Master Mix with HF Buffer (NEB, M0531L)

P1\_PCR\_For primer (5uM)

Index primers (5uM)

PCR tubes and caps

Determine concentration of DNA after size selection.

Determine the concentration of the DNA post pippin size selection using 2ul of the size selected DNA and the Qubit dsDNA HS Assay kit. (The concentration is usually around 1ng/ul or less).

Low-cycle PCR

Set up multiple PCR reactions for each index group (the number will depend on the volume of size selected DNA received back after pippin). We aim for approximately 10ng of DNA per PCR reaction. Therefore at around 1ng/ul if you get approximately 30ul total volume you can do 3 PCR reactions. These reactions can be prepared as a master mix with all reagents and then divided between 3 (or more) PCR tubes each with 25ul.

|  |  |  |
| --- | --- | --- |
| Reagent | Volume (ul) for 1 reaction | Volume (ul) for 3 reactions |
| Adapter ligated and size selected DNA | 10 | 30 |
| Phusion Master mix (2X) | 12.5 | 37.5 |
| Primer P1\_PCR\_for (5uM) | 1.25 | 3.75 |
| Index primer (\*P\_Index”X”) (5uM) | 1.25 | 3.75 |

\* use a different Index primer for each group of samples.

Place on the thermocylcer and run the following program:

Choose the preheat option and set to 100 degC

* 98 degC for 30 sec
* 11 cycles of:
98 degC for 5 sec
60 degC for 25 sec
72 degC for 10 sec
* 72 degC for 5min
* Hold at 10 degC

**Pool reactions within Index groups and clean up digested DNA.**

Consumables:

SPRI beads (Agencourt AMPure XP, Beckman Coulter, A63880) or home made MagNA (see protocol at end of methods)

Freshly prepared 70% Ethanol

1.5ml tubes

screw-capped tube required by core facility for sample submission

AE Buffer (Qiagen)

Qubit dsDNA HS assay kit (ThermoFisher Scientific, Q32853)

1 or 2% agarose gel, loading dye, 1kb ladder

Equipment:

Magnetic separation rack for 1.5ml tubes (e.g. NEB, S1509S)

Qubit Fluorometer (ThermoFisher Scientific)

\*BluePippin (Sage Science)

\*Fragment Analyzer (Advanced Analytical)

\* *In our case these are services provided by the Biotechnology Resource Center Genomics Facility at Cornell University (http://www.biotech.cornell.edu/brc/genomics-facility)*

Pool the PCR reactions

Pool the PCR reactions from the same group in 1.5ml tube and run 5ul on a 1 or 2% agarose gel to check for enrichment. There should be a band visible around 500-600bp, but it may be faint.

Clean up DNA post PCR.

1. Remove SPRI beads from refrigerator and allow to warm to room temperature.
2. Mix SPRI beads well
3. Add 0.7X volumes of SPRI beads to the 1.5ml tube with the pooled PCR reactions (mix well each time you pipette, to ensure that beads are even distributed in the mixture).
In our case we will add 52.5ul SPRI beads to each tube containing 75ul PCR product.
4. Mix gently by pipetting up and down or turning tube over.
5. Incubate at room temperature for 5min
6. Place tube on the magnetic rack to capture the beads. This may take many minutes. When the beads are captured to the side of the tube and the supernatant is clear you are ready to proceed.
7. Remove and discard the supernatant (keeping the tube on the magnetic rack)
8. Wash the beads by pipetting 500ul 70% Ethanol into the tubes. Keep the tubes on the magnetic rack.
9. Incubate for 1min and then remove and discard the ethanol
10. Repeat the Ethanol wash (steps 9 and 10)
11. Remove as much ethanol as possible and allow the beads to air-dry for 5-10 mins (keep tubes in the magnetic rack).
12. Remove the tube from the magnetic rack and resuspend the beads in 36ul AE buffer. Pipette repeatedly over the side of the tube to resuspend all beads and wash the beads down into the suspension.
13. Incubate at room temperature for 2min.
14. Return the tubes back to the magnetic rack and incubate until the beads are captured on the side of the tube and the supernatant is clear.
15. Transfer 34ul of the supernatant to a new 1.5ml tube.

Determine concentration of DNA after cleanup.

Determine the concentration of the DNA post PCR using 2ul of the cleaned up DNA and the Qubit dsDNA HS Assay kit. (The concentration may still be low, but you should see an increase in total mass compared to the pre-PCR post-pippin readings).

Submit 4ul DNA from each index group for fragment analysis:

Perform fragment analysis on the DNA post PCR to determine average size of the fragments within each pool. This information will be used to pool the index groups in the final library. We submit our samples to the Biotechnology Resource Center Genomics Facility at Cornell University.

**Pool and Dilute samples to create final library**

Consumables

O.5X AE + 0.1% Tween: 500ul AE buffer (Qiagen, 19077) + 500ul nuclease free water (Qiagen, 129115) + 1ul Tween 20 (Fisher Scientific, BP337-100)

Screw capped tubes for library submission to the core facilty

Using the concentrations determined by Qubit and the average size (provided by the fragment analysis), dilute and pool the index groups to create one library for Illumina sequencing.

Follow the calculations shown in the table below calculate the current molarity of each index group. Then dilute each index group to 2nM (this can be done in any volume, but 20ul is usually sufficient) with 0.5XAE containing 0.1%Tween.

For an example see “index 1” entry in the table below.

Finally combine equal volumes of each normalized index together in one screw capped tube. (e.g. add 20ul of each normalized index (2nM). If combining 12 index groups you will have a total of 240ul)

**Submit for Illumina sequencing**

Instrument: HiSEQ 2500

Run Length: 100 single end

These ddRAD experiments have a low 5' complexity, therefore make sure to request that the library is run with the addition of the Illumina PhiX control (>5%). Our libraries are usually run with the addition of 15% PhiX.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Enter data | Calculations | dilute to 2nM |
|  | **Size (bp)**  | **conc (ng/ul)** | **MW (g/mol)** | **convert conc to g/L** | **Molarity of index** **(M, mol/L)** | **Molarity of Index** **(nM, nmol/L)**  | **vol (ul) of index** | **0.5X AE/0.1% Tween to add (ul)** | **total volume (ul)** |
|  | *Avg bp from frag analysis output* | *Conc. in ng/ul from Qubit* | *= Size (bp) X 660Da (g/mol)* | *= conc (ng/ul) ÷ 1000* | *= conc (g/L) ÷ MW* | *= M X 109* | *= 2nM (desired molarity) X total Vol ÷ current molarity (nM )of index* | *=total volume - volume of index to be added* |  |
| index 1 | 590 | 1.63 | 389400.00 | 1.63E-03 | 4.19E-09 | 4.19 | 9.6 | 10.4 | 20.0 |
| index 2 |  |  |  |  |  |  |  |  |  |
| index 3 |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

Table 1: Example of calculations required to normalize individual index groups

Table 2: Oligos for ddRAD-seq

|  |  |  |  |
| --- | --- | --- | --- |
| **Sequence Name** | **Sequence** | **barcode length** | **Bases** |
| P1F\_1 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ATC ACG TGC\* A | 6 | 43 |
| P1F\_2 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CGA TGT TGC\* A | 6 | 43 |
| P1F\_3 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TTA GGC TGC\* A | 6 | 43 |
| P1F\_4 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TGA CCA TGC\* A | 6 | 43 |
| P1F\_5 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ACA GTG TGC\* A | 6 | 43 |
| P1F\_6 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GCC AAT TGC\* A | 6 | 43 |
| P1F\_15 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CTT GAT GC\*A | 5 | 42 |
| P1F\_16 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TCA CCT GC\*A | 5 | 42 |
| P1F\_17 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CTA GCT GC\*A | 5 | 42 |
| P1F\_18 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ACA AAT GC\*A | 5 | 42 |
| P1F\_20 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AGC CCT GC\*A | 5 | 42 |
| P1F\_21 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTA TTT GC\*A | 5 | 42 |
| P1F\_22 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CTG TAT GC\*A | 5 | 42 |
| P1F\_23 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AGC ATT GC\*A | 5 | 42 |
| P1F\_24 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ACT ATT GC\*A | 5 | 42 |
| P1F\_26 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CTTGCTT TGC\* A | 7 | 44 |
| P1F\_27 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ATGAAAC TGC\* A | 7 | 44 |
| P1F\_28 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AAAAGTT TGC\* A | 7 | 44 |
| P1F\_29 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GAATTCA TGC\* A | 7 | 44 |
| P1F\_30 | ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGACCTA TGC\* A | 7 | 44 |
| P1R\_1 | /5Phos/CGT GAT AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 42 |
| P1R\_2 | /5Phos/ACA TCG AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 42 |
| P1R\_3 | /5Phos/GCC TAA AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 42 |
| P1R\_4 | /5Phos/TGG TCA AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 42 |
| P1R\_5 | /5Phos/CAC TGT AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 42 |
| P1R\_6 | /5Phos/ATT GGC AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 42 |
| P1R\_15 | /5Phos/TCA AGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA A\*A |  | 41 |
| P1R\_16 | /5Phos/GGT GAA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA A\*A |  | 41 |
| P1R\_17 | /5Phos/GCT AGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA A\*A |  | 41 |
| P1R\_18 | /5Phos/TTT GTA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA A\*A |  | 41 |
| P1R\_20 | /5Phos/GGG CTA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA A\*A |  | 41 |
| P1R\_21 | /5Phos/AAT ACA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA A\*A |  | 41 |
| P1R\_22 | /5Phos/TAC AGA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA A\*A |  | 41 |
| P1R\_23 | /5Phos/ATG CTA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA A\*A |  | 41 |
| P1R\_24 | /5Phos/ATA GTA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GTA A\*A |  | 41 |
| P1R\_26 | /5Phos/AAGCAAG AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 43 |
| P1R\_27 | /5Phos/GTTTCAT AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 43 |
| P1R\_28 | /5Phos/AACTTTT AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 43 |
| P1R\_29 | /5Phos/TGAATTC AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 43 |
| P1R\_30 | /5Phos/TAGGTCC AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT AA\*A |  | 43 |
| P2F\_MspI | /5Phos/CGA GAT CGG AAG AGC GAG AAC AA |  | 34 |
| P2R\_MspI | GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T |  | 34 |
| P2F\_EcoRI # | /5phos/AAT TAG ATC GGA AGA GCG AGA ACA A |  | 25 |
| P1\_PCR\_for | AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC\* T |  | 58 |
| P\_index1 | CAA GCA GAA GAC GGC ATA CGA GAT CGT GAT GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index2 | CAA GCA GAA GAC GGC ATA CGA GAT ACA TCG GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index3 | CAA GCA GAA GAC GGC ATA CGA GAT GCC TAA GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index4 | CAA GCA GAA GAC GGC ATA CGA GAT TGG TCA GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index5 | CAA GCA GAA GAC GGC ATA CGA GAT CAC TGT GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index6 | CAA GCA GAA GAC GGC ATA CGA GAT ATT GGC GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index7 | CAA GCA GAA GAC GGC ATA CGA GAT GAT CTG GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index8 | CAA GCA GAA GAC GGC ATA CGA GAT TCA AGT GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index9 | CAA GCA GAA GAC GGC ATA CGA GAT CTG ATC GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index10 | CAA GCA GAA GAC GGC ATA CGA GAT AAG CTA GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index11 | CAA GCA GAA GAC GGC ATA CGA GAT GTA GCC GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |
| P\_index12 | CAA GCA GAA GAC GGC ATA CGA GAT TAC AAG GTG ACT GGA GTT CAG ACG TGT G\*C |  | 53 |

\* = Phosphorothioate Bond, inhibits exonuclease degradation

/5Phos/ = 5' Phosphorylation, inhibits degradation by 3' exonulceases and blocks extension by DNA polymerases

#Note: P2F\_EcoRI is combined with P2R\_MspI to create a P2 adapter for use with EcoRI digested DNA. It is only needed if you are going to use EcoRI instead of MspI in the digest.

**Homemade MagNA:**

This protocol is taken from Rohland, N. and Reich, D. (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture.Genome Res. May;22(5):939-46. doi: 10.1101/gr.128124.111. and a detailed protocol can be found: https://ethanomics.files.wordpress.com/2012/08/serapure\_v2-2.pdf

Consumables:

Sera-Mag SpeedBeads Carboxyl Magnetic Beads Hydrophobic, GE Healthcare (VWR, 10204-670)

PEG-8000 (VWR, 0159-500G)

0.5M EDTA, pH 8.0 (Fisher Scientific, BP2482-1)

1.0M Tris, pH 8.0 (Fisher Scientific, BP1758-500)

Tween 20 (Fisher Scientific, BP337-100)

5M NaCl (Sigma, 59222C-500ML)

GeneRuler Ultra Low Range DNA Ladder (Thermo Scientific, SM1211)

Nuclease free water

Prepare 1XTE (In a 50ml conical using sterile solutions add the following):

* 500 ul 1M Tris, pH8
* 100 ul 0.5M EDTA, pH 8
* fill to 50ml with nuclease free water

Wash Sera-Mag Beads:

1. Mix the Sera-Mag beads well and transfer 1ml to a 1.5ml tube
2. Place the tube on the magnet stand and wait a couple of minutes until all the beads are collected on the magnet and the supernatant is clear.
3. Remove the supernatant and discard.
4. Remove the tube from the magnet, add 1 ml 1XTE, mix and return tibe to magnet.
5. Repeat the wash the 1XTE.
6. Finally remove the tube from the magnet and resuspend the beads in 1 ml 1XTE.

Prepare the MagNA beads:

1. Add 9g PEG-8000 to a new 50 ml conical tube.
2. Add 10ml 5M NaCl
3. Add 500 ul 1M Tris, pH 8
4. Add 100 ul 0.5M EDTA, pH 8
5. Fill conical to approximately 49ml with nuclease free water
6. Mix for about 3-5 min until the solution becomes clear (you will need to stop shaking and wait for the bubbles to clear)
7. Add 27.5 ul Tween 20
8. Mix the 1 ml SeraPure + TE solution from above and add to the conical.
9. Make up to 50 ml with nuclease free water and gently mix until evenly distributed.
10. Test using aliquots of ladder (you can compare with commercial Ampure XP).
11. Wrap 50 ml tube in foil and store at 4 degC.
12. If storing for an extended period of time – retest monthly.

To test the MagNA:

1. Mix 6ul of GeneRuler with 54 ul nuclease free water
2. Aliquot 20ul into three 1.5ml tubes.
Mix the MagNA well to ensure that the beads are evenly distributed.
3. To the first tube add 18 ul MagNA (0.9X vol), to the second tube add 30 ul MagNA (1.5X vol), keep the remaining GeneRuler for the control.
4. Incubate the tubes with beads for 5min at RT
5. Place the tubes with beads on a magnet stand and wait until beads have collected at the magnet and the supernatant is clear.
6. Remove the supernatant.
7. Wash the beads twice with 500 ul freshly prepared 70% Ethanol. Keep the tubes on the magnet, add the ethanol, allow to stand for 1 min and then remove the ethanol and repeat.
8. Remove as much ethanol as possible and air dry the beads for approximately 5 min.
9. Remove the tube from the magnet stand and resupend the beads in 20 ul nuclease free water.
10. Allow to stand for 2 min
11. Place tubes on magnet stand and remove supernatant to a new 1.5ml tube.
12. Add 4ul 6X laoding dye to all the tubes (incl. the 20ul control) and run on a 1.5% or 2% agarose gel.
13. For 0.9X only the top band (300bp) is retained and for 1.5X the 100bp band is lost, but the 150bp, 200bp and 300bp bands are retained.

**Bioinformatic pipeline for analysis of ddRAD-seq data for parentage analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **STEP** | **Description** | **Program, link, examples of commands used** | **Notes** |
| 1 | Download data |  |  |
| 2 | Check quality of data | fastqc ./*name-of-file*.fastq.gz<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> | Can be run on zipped dataOutput in html |
| 3 | Unzip data |  |  |
| 4 | Trim and filter reads | FASTQ Quality trimmer and Filter (<http://hannonlab.cshl.edu/fastx_toolkit/>):Fastx\_trimmer -Q33 -l 97 -i ./*name-of-file*.fastq -o ./sampe1\_t.fastq(Trim lower quality bases at the end of each read, in this case the last 4bp) Repeat for all filesfastq\_quality\_filter -Q33 -v -q 10 -p 100 -i ./sampe1\_t.fastq -o ./sample1\_tf.fastq(discard all reads with Phred score of 10 or less) Repeat for all filesfastq\_quality\_filter -Q33 -v -q 20 -p 95 -i ./sample1\_tf.fastq -o ./sample1\_tff.fastq(discard reads with Phred 20 in more than 5% of the reads) Repeat for all files. | These parameters will depend on the quality of the sequencing data (see fastqc results above) |
| 5 | Demultiplex | Process\_radtags program from the stacks pipeline (<http://catchenlab.life.illinois.edu/stacks/comp/process_radtags.php>)process\_radtags -p ./sample\_raw -b ./index1.txt -o ./demultfilter -e sbfI -c -q -E phred33 --inline\_null -i fastq --adapter\_1 GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG --adapter\_mm 1 --filter\_illuminaRepeat for all index groups (with different adapter sequences and barcode files) | First move trimmed and filtered data into a raw data folder and create the appropriate barcode text files. |
| 6. | Align reads | Use denovo\_map.pl program from the stacks pipeline (<http://catchenlab.life.illinois.edu/stacks/comp/denovo_map.php>) denovo\_map.pl -m 5 -M 5 -n 5 -T 15 -B radtags -S -b 1 -t -D "radtags" -i 1 -o /denovo \-s /demultfilter/indiv1.fq \-s /demultfilter/indiv2.fq \-s /demultfilter/indiv3.fq \etc. | Optimize parameters depending on the level of polymorphism in the study organism |
| 6 | Apply corrections | Use rxstacks program from the stacks pipeline(<http://catchenlab.life.illinois.edu/stacks/comp/rxstacks.php>) rxstacks -b 1 -P /denovo -o /post\_rxstacks\_assembly --conf\_lim 0.25 --prune\_haplo --model\_type bounded --bound\_high 0.1 --lnl\_lim -300.0 --lnl\_dist -t 8 –verbose |  |
| 7 | Determining the distribution of catalog loci log likelihoods | Bucket the log likelihood values:cat batch\_1.rxstacks\_lnls.tsv | grep -v "^#" | awk '{bucket=(int($2)); lnls[bucket] += 1} END { for (bucket in lnls) print bucket, "\t", lnls[bucket]}' | sort -n > lnls.tsvPlot these values and choose an appropriate threshold that will retain enough loci |  |
| 8 | Rerun rxstacks with new log likelihood threshold value | rxstacks -b 1 -P /denovo -o /post\_rxstacks\_assembly --conf\_lim 0.25 --prune\_haplo --model\_type bounded --bound\_high 0.1 --lnl\_lim -40.0 --lnl\_dist -t 8 –verbose | Use the lnl\_lim value determined in step 7 (e.g. -40) |
| 9 | Run cstacks (create catalog) and sstacks (search putative loci against catalog) | <http://catchenlab.life.illinois.edu/stacks/comp/cstacks.php>cstacks -b 1 –o /post\_rxstacks\_assembly -p 15 -n 5 \-s /post\_rxstacks\_assembly/indiv1 \-s /post\_rxstacks\_assembly/indiv2 \-s /post\_rxstacks\_assembly/indiv3 \etc.<http://catchenlab.life.illinois.edu/stacks/comp/sstacks.php>sstacks -b 1 -c /post\_rxstacks\_assembly/batch\_1 -p 15 -o /post\_rxstacks\_assembly -s /indiv1sstacks -b 1 -c /post\_rxstacks\_assembly/batch\_1 -p 15 -o /post\_rxstacks\_assembly -s /indiv2sstacks -b 1 -c /post\_rxstacks\_assembly/batch\_1 -p 15 -o /post\_rxstacks\_assembly -s /indiv3etc. |  |
| 10 | Compute population genetics statistics and output SNPs | Using the population program from the stacks pipeline <http://catchenlab.life.illinois.edu/stacks/comp/populations.php> /populations -b 1 -P /post\_rxstacks\_assembly -M /popmap.txt -r 0.95 -p 1 -m 10 -t 8 --write\_single\_snp --structure --vcf --genepop -a 0.25Note: here you can change the r (minimum percentage of individuals in population to process a locus), m (minimum stack depth),and a (minimum minor allele frequency) parameters to recover more loci.  | First prepare a population map (tab separated txt file listing all individuals in the population) |
| 11 | Remove loci that are not in Hardy-Weinberg equilibrium | Using vcftools. <http://vcftools.sourceforge.net/man_latest.html>vcftools --vcf ./batch\_1.vcf --hwe 0.05 --recode --out filtered\_1.vcf &For populations exhibiting substructure (i.e. those that are closely related or highly inbred), a large number of loci may deviate from Hardy-Weinberg equilibrium, and thus be filtered during this step. For these cases, this step can be skipped in order to retain a larger number of SNP loci. These loci can be used for parentage inference, but should be interpreted with caution (see CERVUS tutorial for more information). |  |
| 12 | Convert vcf file to genepop format | Use PDGSpider: <http://www.cmpg.unibe.ch/software/PGDSpider/> | Note: a genepop file can be imported into the Cervus program using the “tools” function. |