Please cite this article as: Zhonglai *et. al.*, (2016). Identification of Natural Hybrids by SSR Markers in *Mussaenda*, *Bio-protocol* 6 (13): e1853. DOI: 10.21769/BioProtoc.1853.

bio-protocol www.bio-protocol.org/e1853 Vol 6, Iss 13, Jul 5, 2016

Identification of Natural Hybrids by SSR Markers in *Mussaenda*

Zhonglai Luo[1], Tingting Duan[1, 2], Shuai Yuan[1, 2], Shi Chen[3], Xiufeng Bai[4] and Dianxiang Zhang[1, *]

[1]Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, the Chinese Academy of Sciences, China; [2]University of Chinese Academy of Sciences, China; [3]Beneficial Insects Institute, Fujian Agriculture and Forestry University, China; [4]Department of Cell Biology, the Third Military Medical University, Chongqing, China

*For correspondence: dx-zhang@scbg.ac.cn

[Abstract] Detection of natural hybrids is of great significance for plant taxonomy, reproductive biology, and population genetic studies. Compared with methods depending on morphological characters, molecular markers provide reliable and much more accurate results. This protocol describes approaches employing microsatellite (SSR) markers to identify inter-specific hybrids in *Mussaenda* (Rubiaceae).

Materials and Reagents

A. Consumables
   1. Microfuge tubes (Corning, Axygen®, catalog number: MCT-150-C)
   2. Pipette tips (Corning, Axygen®, catalog number: T-1000-C, T-200-C, T-300)

B. Plant material
   1. Fresh leaves of the putative hybrid individuals, as well as the parental species, were collected and dried in silica gel. It usually takes 3-5 days for complete drying of the leaves.

C. Chemicals
   1. Taq PCR mastermix (Tiangen, catalog number: KT201)
   2. Liquid nitrogen
   3. Absolute ethanol (Sinopharm Chemical Reagent, catalog number: 10009228) (ice cold)
   4. 70% ethanol (ice cold)
   5. 7.5 M ammonium acetate (Sigma-Aldrich, catalog number: A1542-250G)
   6. Chloroform: isoamyl alcohol (24:1) (Sinopharm Chemical Reagent, catalog number: chloroform, 10006818; isoamyl alcohol, 10003218)
   7. ddH_2O (sterile)
   8. 5x Loading buffer with GelRed (Shanghai Generay Biotech, catalog number: GR0205-500)
   9. Cetyltrimethylammonium bromide/Hexadecyl trimethyl-ammonium bromide (CTAB) (Thermo Fisher Scientific, catalog number: ICN19400480)

Please cite this article as: Zhonglai *et. al.*, (2016). Identification of Natural Hybrids by SSR Markers in *Mussaenda*, *Bio-protocol* 6 (13): e1853. DOI: 10.21769/BioProtoc.1853.

10. Ethylenediaminetetraacetic acid (0.5M solution/pH 8.0) (EDTA) (Thermo Fisher Scientific, catalog number: BP2482-500)
11. Polyvinyl pyrrolidone, MW 40,000 (PVP 40)
12. NaCl (Sigma-Aldrich, catalog number: S5886-500G)
13. HCl (Sigma-Aldrich, catalog number: 258148)
14. Tris base (Thermo Fisher Scientific, catalog number: BP152-1)
15. Boric acid (Sigma-Aldrich, catalog number: B0394-100G)
16. Agarose (Biowest, catalog number: 111860)
17. Hi-Di™ formamide (Thermo Fisher Scientific, Applied Biosystems™, catalog number: 4311320)
18. GeneScan 500Liz (Thermo Fisher Scientific, Applied Biosystems™, catalog number: 4322682)
19. CTAB buffer (see Recipes)
20. 1 M Tris (pH 8.0) (see Recipes)
21. TE buffer (see Recipes)
22. 5x TBE (Tris-Borate-EDTA) buffer (Stock) (see Recipes)
23. 1% agarose gel (see Recipes)

Equipment

1. Mortar and pestle
2. Microcentrifuge (Eppendorf, model: 5427R)
3. Water bath (IKA, model: HB10)
4. PCR thermal cycler (EASTWIN, model: ETC-811)
5. Agarose gel electrophoresis system [include PowerPac™ Basic Power Supply for electrophoresis (Bio-Rad Laboratories, catalog number: 1645050)]
6. Genetic Analyzer (Invitrogen, model: ABI PRISM 3100)

Software

1. GeneMarker version 2.4.0 (http://www.softgenetics.com/GeneMarker.html)
2. FSTAT version 2.9.3 (http://www2.unil.ch/popgen/softwares/fstat.htm)
3. GenAlEx version 6.2 (http://biology-assets.anu.edu.au/GenAlEx/Download.html)
4. STRUCTURE version 2.2 (http://pritchardlab.stanford.edu/software/structure2_2.html)
5. NewHybrids version 1.1 beta (http://ib.berkeley.edu/labs/slatkin/eriq/software/software.htm#NewHybs)

Procedure

1. DNA extraction and polymerase chain reaction

   Total DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) method

Please cite this article as: Zhonglai *et. al.*, (2016). Identification of Natural Hybrids by SSR Markers in *Mussaenda*, *Bio-protocol* 6 (13): e1853. DOI: 10.21769/BioProtoc.1853.

(Doyle, 1991). Ten pairs of SSR markers were developed by Duan and Zhang (2014) (AC30, CT12, CT59, CT113, CT17, CT60, CT48, CT135, CT142, CT91) from *M. pubescens* were selected for testing the gene flow between *M. pubescens* var. *alba* and *M. kwangtungensis.* The forward of each primer pair was labeled with fluorochrome (FAM, HEX, ROX, TAMRA). The information of the primers is shown in Table 1.

Table 1. Characteristics of the 10 pairs of SSR primers employed in this study. Presented for each locus are the forward (F) and reverse (R) primer sequences, repeat motif, size of the original fragment (bp), annealing temperature (*Ta*), and GenBank accession number. Td, touchdown.

Name	Primer sequence (5'–3')	Repeat	Size (bp)	*Ta* (°C)	GenBank accession number
AC30	F: 5'FAM-GAAAATCCAAGAAACACAT3' R: GACAACTCACAAGCCACTC	(TG)5TT(TG)4	435-463	Td	KF740515
CT48	F: 5'FAM-GGTAAAAAAAAGGATGGAGA3' R: ATGGTATTGCGAGATGGAAAA	(CT)19	316-350	53	KF740517
CT113	F: 5'FAM-AACATACAGACCCAAGCC3' R: AAGCACCTACGAACTCCC	(GA)9	276-356	Td	KF740518
CT17	F: 5'ROX-CACAAAAAAGTAAACGCATA3' R: CTCCCCTCTCACTGTAGAGAG	(TC)6TT(TC)4	285-323	56	KF740519
CT142	F: 5'ROX-CACTGGAGAAGAAAAGCG3' R: GCATGTGCATATACCCGA	(CT)17	255-311	Td	KF740520
CT59	F: 5'ROX-ATTCCAGACACTTACTCACAGC3' R: TGCAAACATACTTGATCCTACC	(CT)11	251-299	56	KF740521
CT12	F: 5' HEX-CAAACTCGCTTCAAAAAAGTGACCATT3' R: CAAAAAGCCAACTAAGCTACGACCATG	(CT)10	223-249	56	KF740522
CT60	F: 5' HEX-CCTATATACTTGGTCTTGTGGT3' R: CAGAACTATCTTATCTGTTGCC	(TC)10	204-276	58	KF740523
CT135	F: 5' HEX-CAAAGCAAAGGATAGTAGGA3' R: GTTGACAGATGCTGGTAATG	(AG)22	181-257	Td	KF740524
CAA112	F: 5' TAMRA –GCAGGCTAGCTATTTCTCC3' R: TGTTCCTTCCCTTTGTTTT	(TTC)6	152-178	53	KF740527

DNA amplifications were carried out in 10 µl reaction mixtures:

5 µl Taq PCR mastermix

0.2 µM of each primer

3.6 µl ddH_2O

20-40 ng of genomic DNA

The amplification conditions were as follows:

a. 94 °C, 4 min (initial denaturation)
b. 94 °C, 30 sec (denaturation)
c. 60 °C, 30 sec (annealing)
d. 72 °C, 30 sec (extension)
   7 cycles from step (b) to (d)
e. 94 °C, 30 sec (denaturation)
f. 53 °C, 30 sec (annealing)
g. 72 °C, 30 sec (extension)

Please cite this article as: Zhonglai *et. al.*, (2016). Identification of Natural Hybrids by SSR Markers in *Mussaenda*, *Bio-protocol* 6 (13): e1853. DOI: 10.21769/BioProtoc.1853.

28 cycles from step (e) to (g)

h. 72 °C, 10 min (final extension)

i. Keep sample at 4 °C until further processing

The PCR products were determined on an ABI PRISM 3100 Genetic Analyzer in the Sangon Biotech (Shanghai) Co., Ltd., and 1 μl products (10-fold diluted) were mixed with 15 μl Hi-Di™ Formamide probe, which included 2% GeneScan 500Liz. The amplified SSR fragment data were collected using GeneMarker version 2.4.0 software.

2. Data analysis

The software FSTAT 2.9.3 (Goudet, 2001) and GenAlEx 6.2 (Peakall and Smouse, 2006) were used for calculating the alleles and private alleles, respectively. We used the program STRUCTURE 2.2 (Pritchard *et al*., 2000; Falush *et al*., 2003 and 2007) for population genetic structure analysis and hybrid detection. This method implements a Bayesian model-based clustering approach. We supposed all the loci were independent, following the Hardy–Weinberg equilibrium and linkage equilibrium. The program was run 10 times for each K-value, ranging 1-5, with 100,000 replicates for burn-in and 40,000 iterations. Out of the five runs for k = 2, the run with the highest likelihood value was selected to assign the posterior membership coefficients (Evanno *et al.*, 2005).

The entire data set is arranged as a matrix in a single file, in which the data for individuals are in rows, and the loci are in columns. For a diploid organism, data for each individual can be stored either as 2 consecutive rows, where each locus is in one column, or in one row, where each locus is in two consecutive columns. Missing data should be indicated by a number that doesn't occur elsewhere in the data.

The software NewHybrids version 1.1 beta (Anderson and Thompson, 2002) was used to identify hybrids in the individuals sampled. The program calculates the posterior probability that fall into each of the six genotype categories, Parent 1, Parent 2, F1 hybrids, F2 hybrids, back-cross generation to Parent 1, and back-cross generation to Parent 2.

For FSTAT, GenAlex and NewHybrids, we used the default value in each software. For STRUCTURE, we chose admixture model, and set the parameters as we mentioned in the original article (Luo *et al.*, 2015), other parameters we used the default values.

Representative data

A representative example of data can be found in our published work (Luo *et al.*, 2015). Figure 1 shows genotype class analysis of all sampled individuals based on the programs NewHybrids and STRUCTURE using SSR data.

Please cite this article as: Zhonglai *et. al.*, (2016). Identification of Natural Hybrids by SSR Markers in *Mussaenda*, *Bio-protocol* 6 (13): e1853. DOI: 10.21769/BioProtoc.1853.

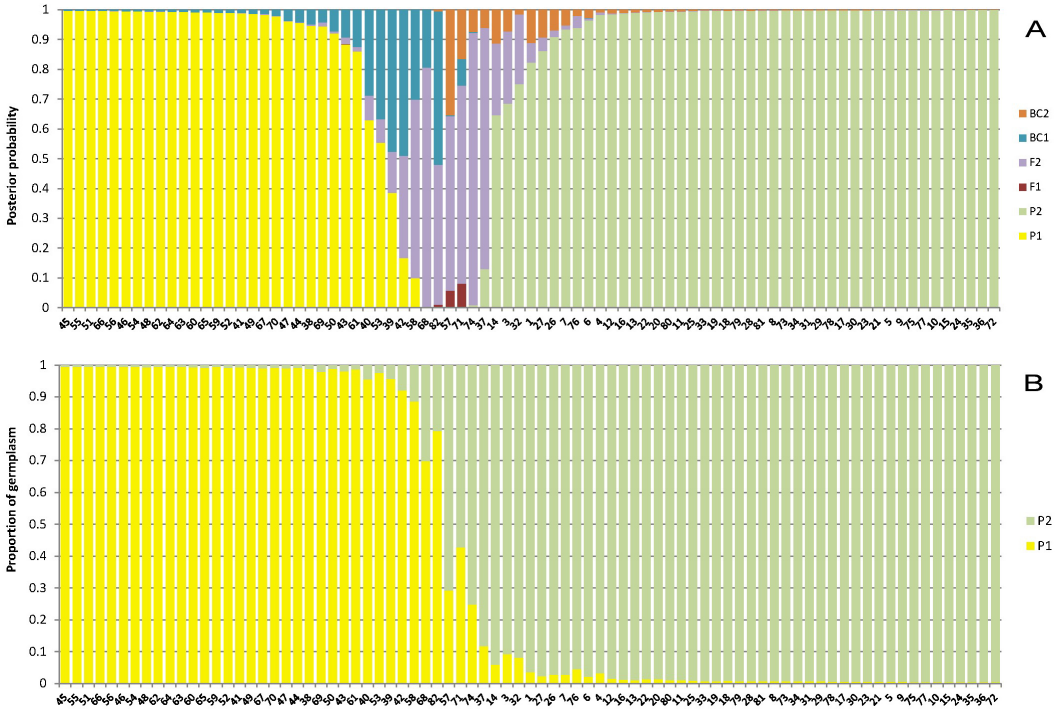

Figure 1. Genotype class analysis of all sampled individuals based on the programs NewHybrids (A) and STRUCTURE (B) using SSR data. A. NewHybrids calculated posterior probability that falls into each of six genotype categories: P1 (Parent 1), P2 (Parent 2), F1 (F1 hybrids), F2 (F2 hybrids), BC 1 (backcross generation to Parent 1), BC 2 (backcross generation to Parent 2). B. Structure of 81 accessions based on 13 SSR loci (K = 2). Each bar represents one individual, P1 = *M. kwangtungensis*, P2 = *M. pubescens* var. *alba*.

Recipes

1. CTAB buffer, 100 ml
   2.0 g CTAB
   10.0 ml 1 M Tris (pH 8.0)
   4.0 ml 0.5 M EDTA (pH 8.0)
   28.0 ml 5 M NaCl
   40.0 ml H_2O
   1 g PVP 40
   Adjust all to pH 8.0 and make up to 100 ml with ddH_2O.
2. 1 M Tris pH 8.0, 1 L
   121.1 g Tris
   700 ml ddH_2O
   Dissolve Tris in ddH_2O and bring to 900 ml.
   Adjust pH to 8.0 by adding 40-50 ml of concentrated HCl.

3. TE buffer for 100 ml
   10 mM 1 ml of 1 M Tris (pH 8.0)
   1 mM 0.2 ml of 0.5 M EDTA
4. 5x TBE (Tris-Borate-EDTA) buffer (Stock)
   54 g Tris base
   27.5 g boric acid
   20 ml 0.5 M EDTA (pH 8.0)
   Adjust pH to 8.3 by HCl.
   This stock solution can be diluted to 1x prior to use in electrophoresis.
5. 1% agarose gel
   Dissolve 1 g agarose in 100 ml 1x TBE

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31000109, 31370269, 31170184, U1202261) and the International Foundation for Science (AD/22476). This protocol was modified from our previous work, Luo, Z. L., Duan, T. T., Yuan, S., Chen, S., Bai, X. F., and Zhang, D. X. (2015) Reproductive isolation between sympatric sister species, *Mussaenda kwangtungensis* and *M. pubescens* var. *alba*. J Integr Plant Biol 57: 859-870.

References

1. Anderson, E. C. and Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 160(3): 1217-1229.
2. Doyle, J. (1991). DNA protocols for plants-CTAB total DNA isolation. In: Hewitt, G. M. and Johnston, A. (eds). *Molecular techniques in taxonomy.* Springer, 283-293.
3. Duan, T. and Zhang, D. (2014). Fourteen additional microsatellite markers for *Mussaenda pubescens* and cross-species amplification. *J Genet* 93(2): e44-47.
4. Evanno, G., Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 14: 2611-2620.
5. Falush, D., Stephens, M. and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4): 1567-1587.
6. Falush, D., Stephens, M., and Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null allele. *Mol Ecol Notes* 7: 574-578.
7. Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3) (updated from Goudet, J. (1995). FSTAT (version 1.2): a computer program to calculate F-statistics. J. Hered. 86: 485-486). Available at: http://www.unil.ch/izea/softwares/fstat.html.

8. Luo, Z. L., Duan, T. T., Yuan, S., Chen, S., Bai, X. F., and Zhang, D. X. (2015) Reproductive isolation between sympatric sister species, *Mussaenda kwangtungensis* and *M. pubescens* var. *alba*. *J Integr Plant Biol* 57: 859-870.
9. Peakall, R. and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* 28(19): 2537-2539.
10. Pritchard, J., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959
11. Shaw, J., Lickey, E. B., Schilling, E. E. and Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot* 94(3): 275-288.