An empirical study for revealing dramatic influence of maxSH in PYRAD/IPYRAD

Jie Zhang[1], Francisco Pina-Martins[2], Zushi Jin[3], Yongpeng Cha[1], Zuyao Liu[4], Jianli Zhao[1], and Qingjun Li[1]

[1]Yunnan University
[2]Universidade de Lisboa Faculdade de Ciencias
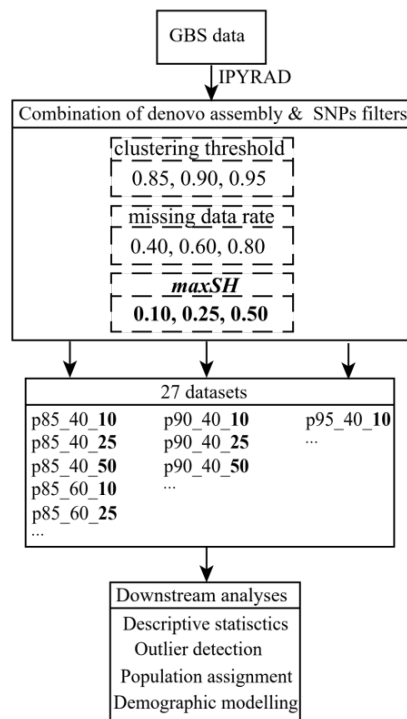[3]Tibet Academy of Agricultural and Animal Husbandry Sciences
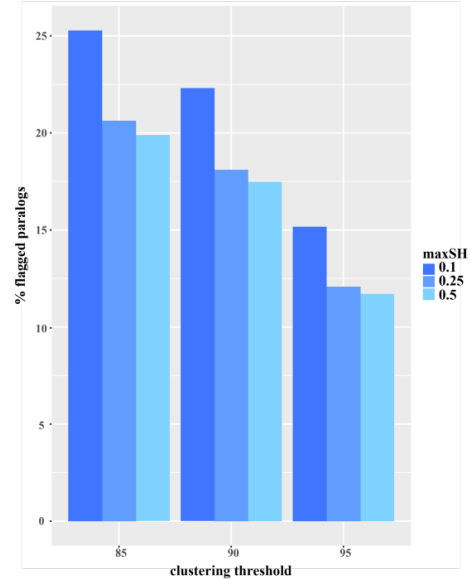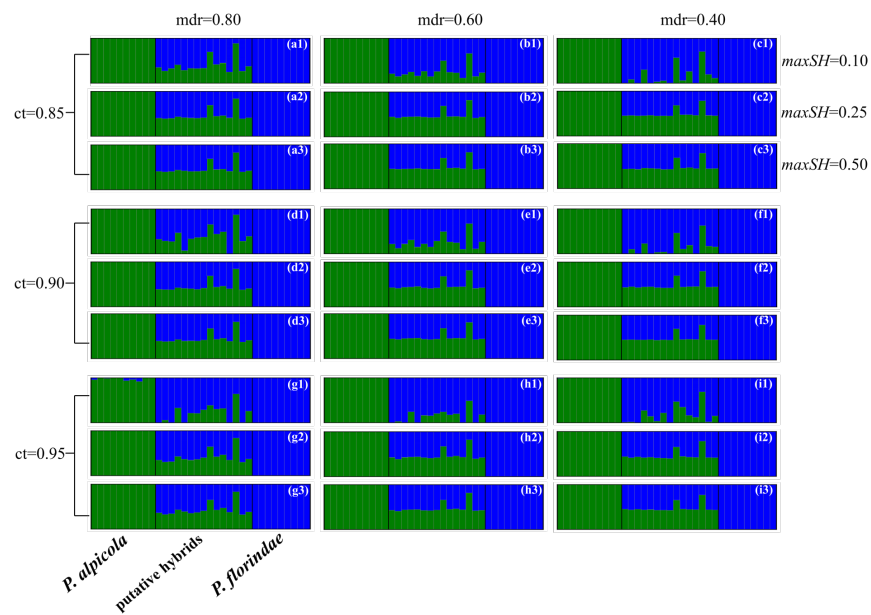[4]University of Bern

November 6, 2020

Abstract

Techniques of reduced-representation sequencing (RRS) have revolutionized ecological and evolutionary genomics studies, especially favoring species without reference genome. But it is a great challenge for RRS data to precisely establish homologous loci, which is strongly associated with accuracy of downstream analyses and reliability of biological inferences. *maxSH* is an overlooked parameter with respect to detecting paralogs, belonging to PYRAD/IPYRAD–a prevailing pipeline for genotyping RADseq and GBS data. Using GBS data of two primroses (*Primula alpicola* Stapf and *P. florindae* Ward) and their putative hybrids, as empirical study, we explore the efficiency of *maxSH* on filtering paralogs and its impact on downstream analyses. At the same time, we try to assess if putative hybrids are truly speciated from hybridization. Our study sheds light on the efficiency of *maxSH* on filtering paralogs, and significant effects of *maxSH*, together with clustering threshold and missing data, on downstream analyses of outlier detection, population assignment, and demographic modelling, emphasizing the significance of carefully coping with bioinformatics process. On the other hand, although putative hybrids exhibit a genetic mixture of *P. alpicola* and *P. florindae* according to most STRUCTURE and PCA results, we cannot clearly draw a conclusion on the origin of putative hybrids due to conflicting demographic scenarios mainly resulted from altering *maxSH* value among nine chosen datasets. However, gene flow patterns of most optimal models from multiple *maxSH* values collectively indicate incomplete reproductive isolation between putative hybrids and two primroses, and the existence of indirect introgression between *P. alpicola* and *P. florindae*.
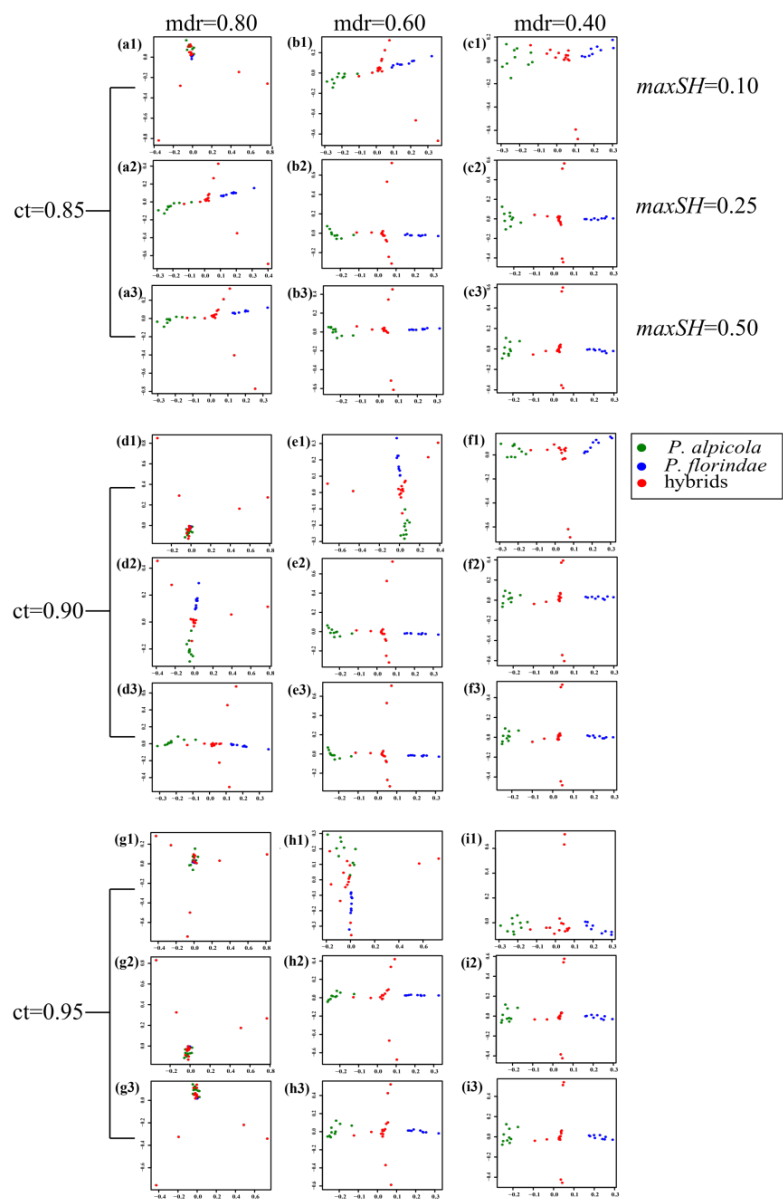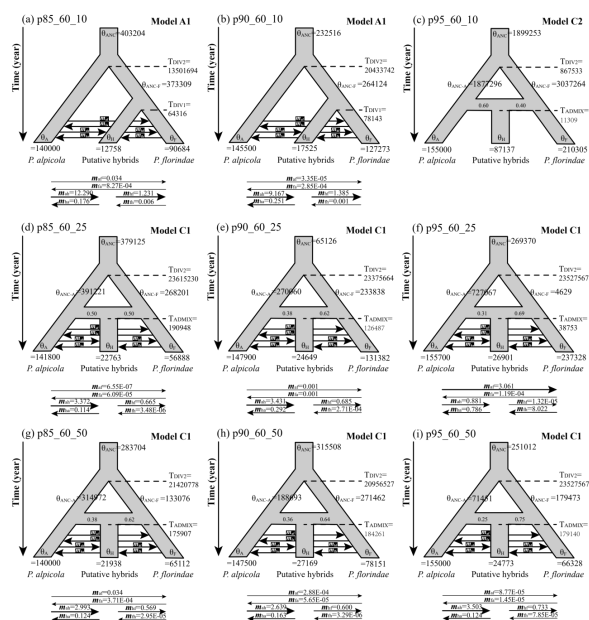
Hosted file

`manuscript.pdf` available at https://authorea.com/users/373751/articles/491358-an-empirical-study-for-revealing-dramatic-influence-of-maxsh-in-pyrad-ipyrad

```
                        ┌─────────────┐
                        │  GBS data   │
                        └─────────────┘
                            │ IPYRAD
                            ▼
┌─────────────────────────────────────────────────────┐
│ Combination of denovo assembly &  SNPs filters        │
│   ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐                         │
│     clustering threshold                              │
│   │  0.85, 0.90, 0.95        │                        │
│   └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘                         │
│   ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐                         │
│     missing data rate                                 │
│   │  0.40, 0.60, 0.80        │                        │
│   └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘                         │
│   ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐                         │
│        maxSH                                          │
│   │  0.10, 0.25, 0.50        │                        │
│   └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘                         │
└─────────────────────────────────────────────────────┘
```

27 datasets

p85_40_**10**	p90_40_**10**	p95_40_**10**
p85_40_**25**	p90_40_**25**	...
p85_40_**50**	p90_40_**50**	
p85_60_**10**	...	
p85_60_**25**		
...		

Downstream analyses

Descriptive statisctics
Outlier detection
Population assignment
Demographic modelling

2

mdr=0.80　　　　mdr=0.60　　　　mdr=0.40

(a1)	(b1)	(c1) *maxSH*=0.10
(a2)	(b2)	(c2) *maxSH*=0.25
(a3)	(b3)	(c3) *maxSH*=0.50

ct=0.85

(d1)	(e1)	(f1)
(d2)	(e2)	(f2)
(d3)	(e3)	(f3)

ct=0.90

(g1)	(h1)	(i1)
(g2)	(h2)	(i2)
(g3)	(h3)	(i3)

ct=0.95

P. alpicola　　putative hybrids　　*P. florindae*

Hosted file

`table.pdf` available at https://authorea.com/users/373751/articles/491358-an-empirical-study-for-revealing-dramatic-influence-of-maxsh-in-pyrad-ipyrad