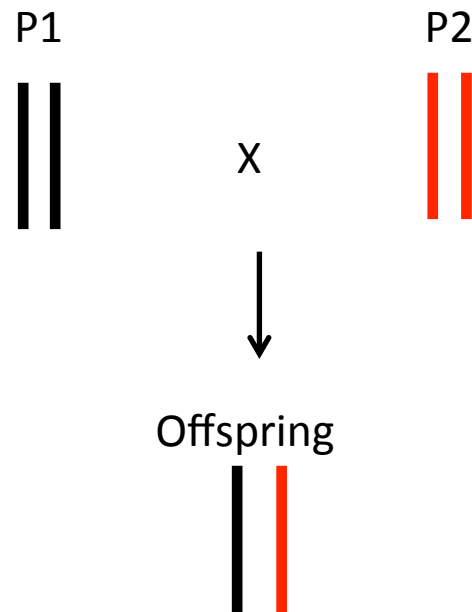


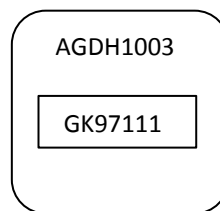
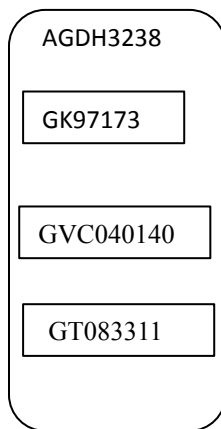
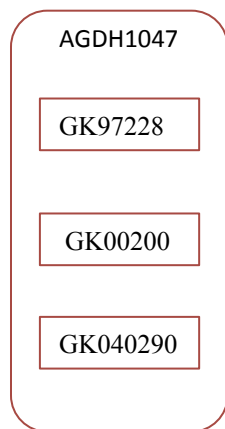
- The pipeline can be used in the analysis of whole genome, parent offspring studies.
- Using a number of steps it can be used to identify SNPs that can distinguish between the two parents. Then use this SNP list to split the aligned reads of the offspring into two genomes.



A12DHL X GDDH33



A12DHL X GDDH33

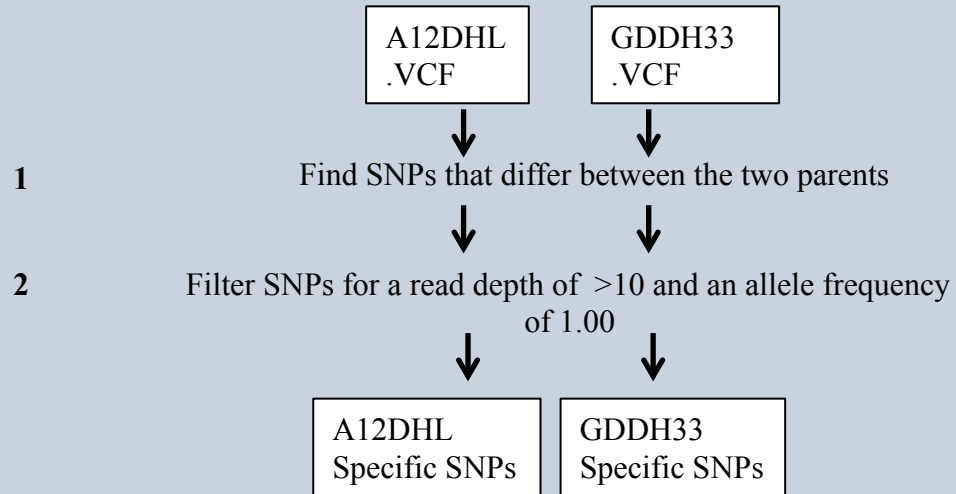
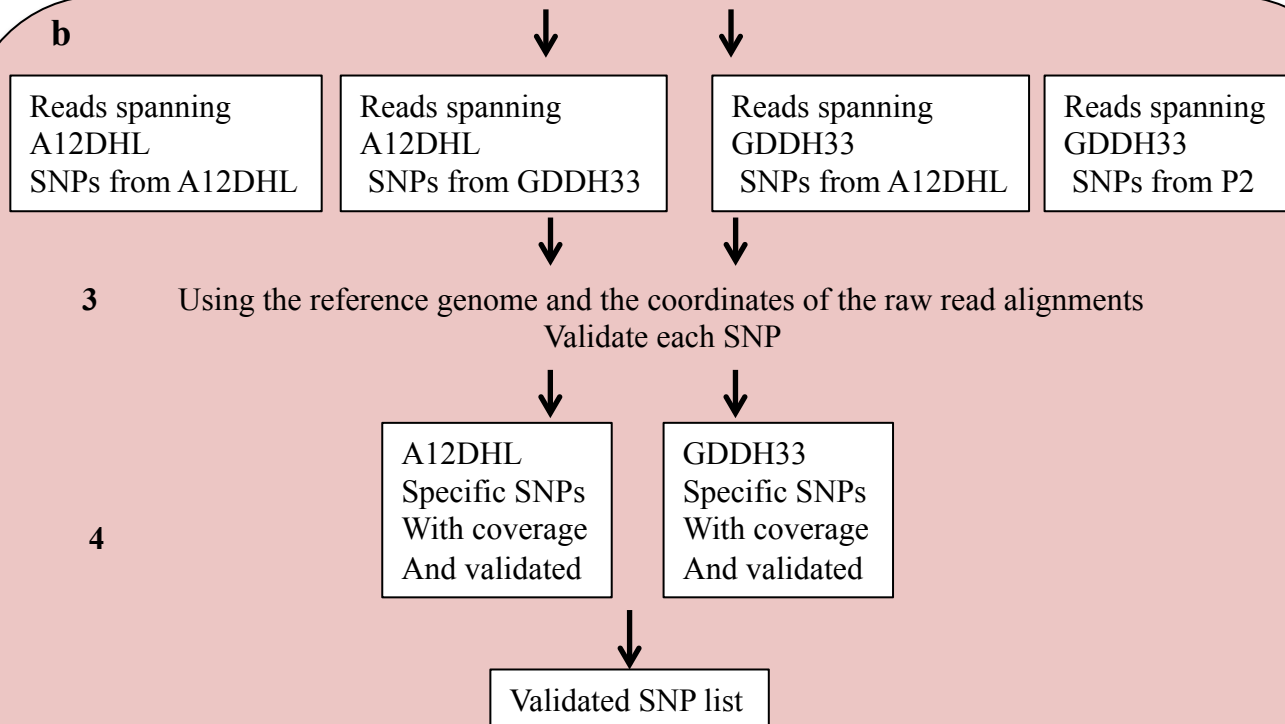


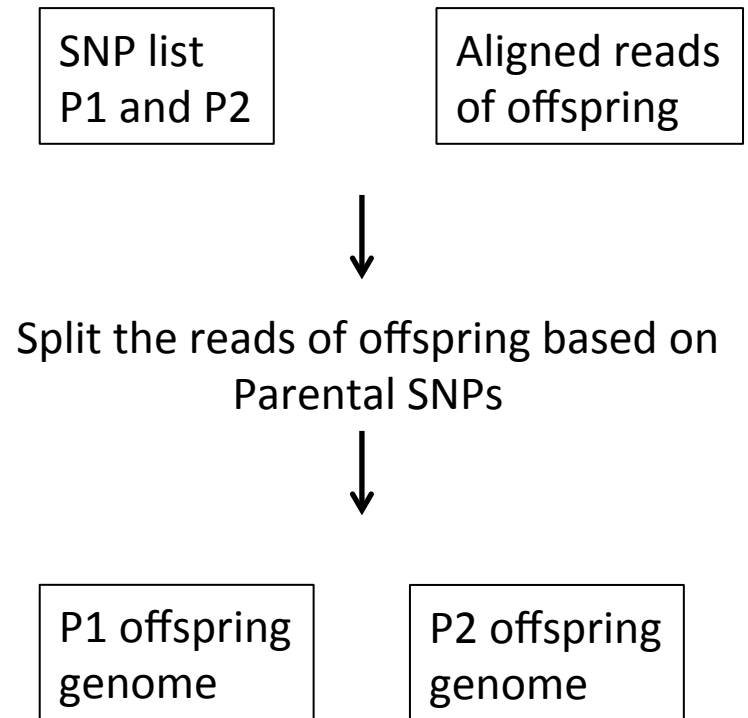
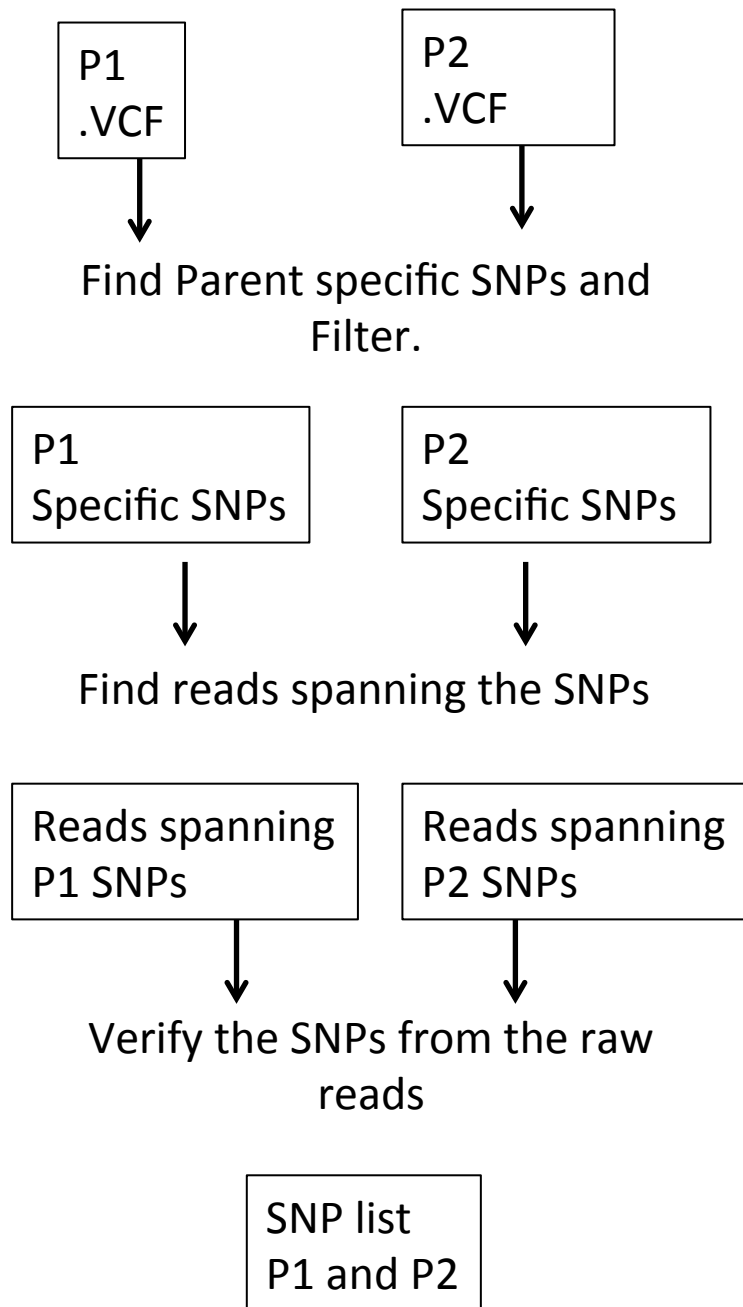
## Program Inputs

- Parent 1 .vcf
  - Parent2 .vcf
  - Parent 1.SAM
  - Parent2.SAM
  - Reference.fasta
  - Offspring reads.SAM
- Before beginning, there must be a .SAM file containing the Aligned reads for P1, P2 and the offspring. This can be Completed using any alignment software.
- There must also be a .vcf file containg SNP information
- For P1 and P2, these must be called using the same reference
- As the read alignemnt and can be called using any software

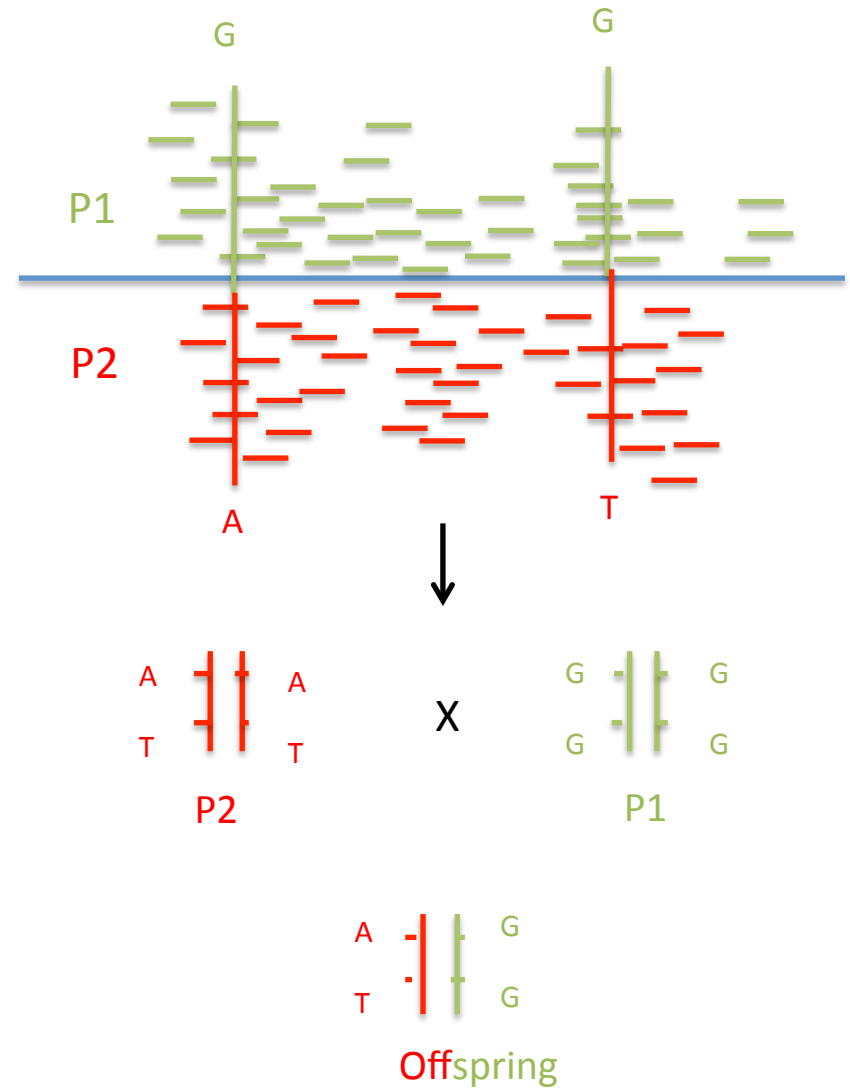
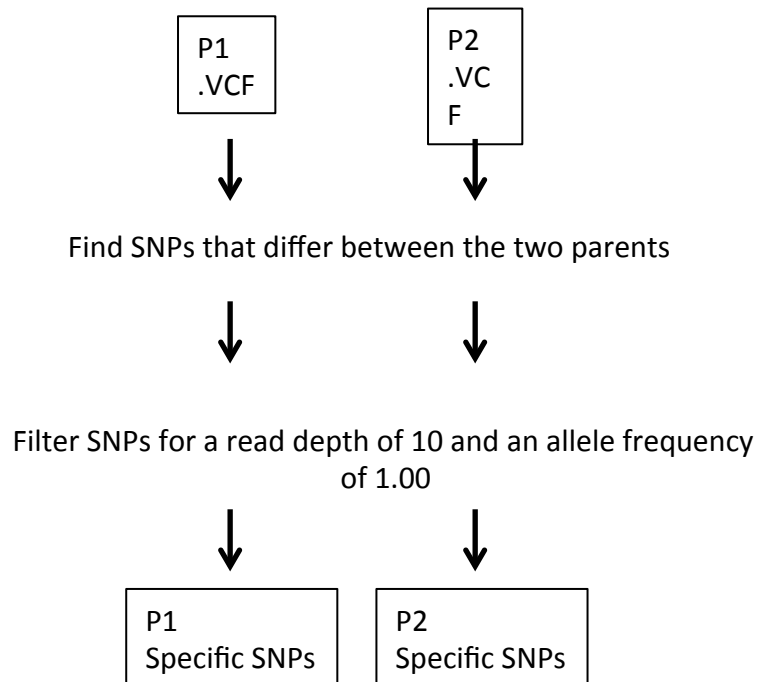
## Program outputs

- Distinguishing SNPs (position, base in parent 1 and base in parent 2
- Verification file, showing the reads supporting the SNP.
- Two SAM files containing the offspring reads fro each parent aligned to the reference genome

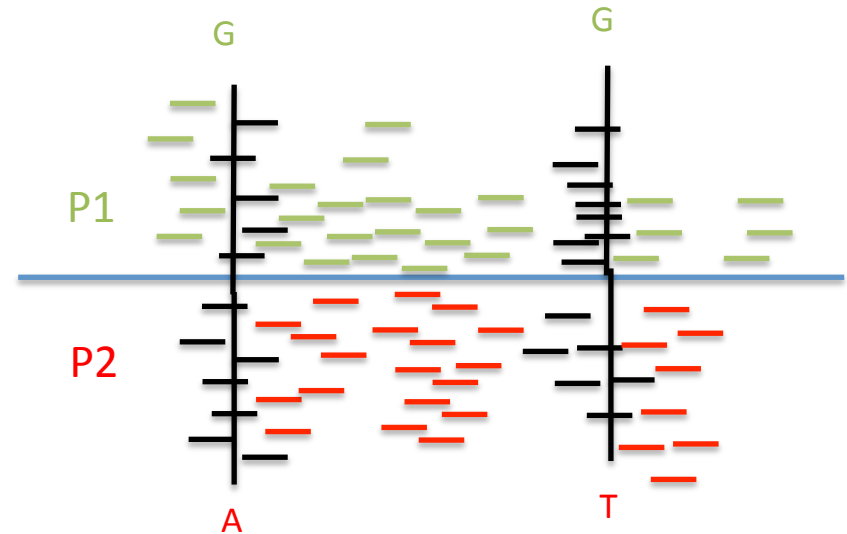
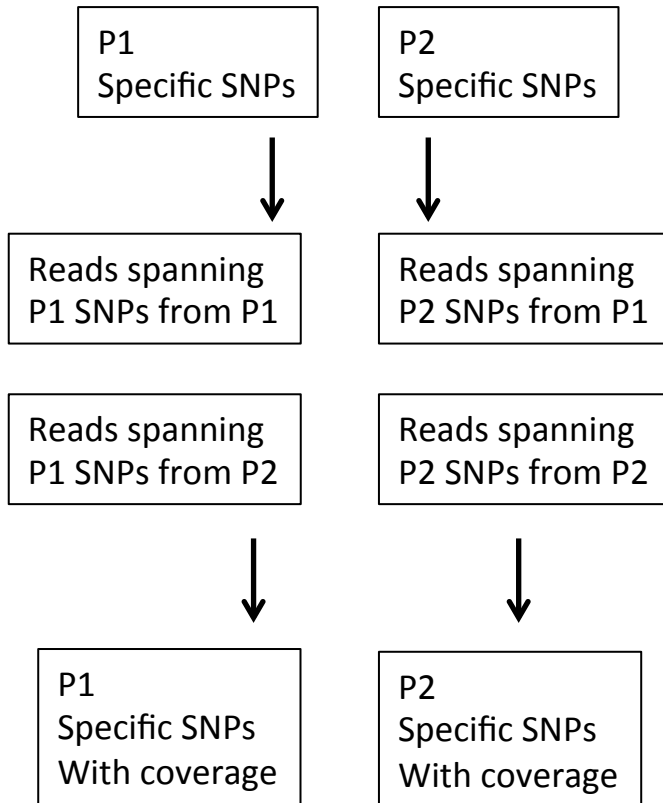
**a****b**



# 1 Find Distinguishing SNPs and Filter



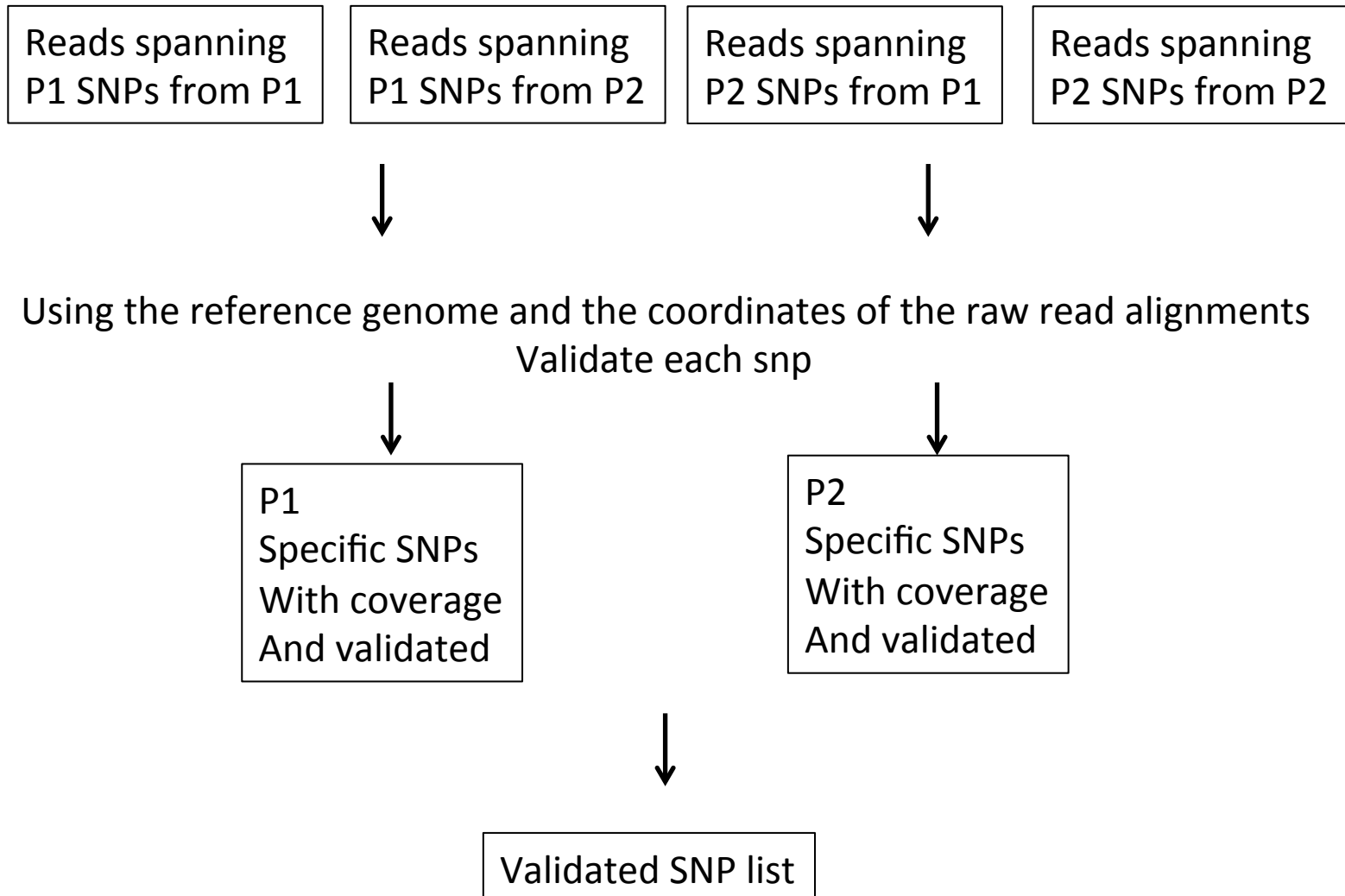
## 2 Find Reads from Both parents that span each SNP



Identify reads for each parent that can  
Verify the SNP

Identify SNP positions that have coverage in  
Both parents

### 3 Validate the SNPs





## Outputs

### .verify

```
>C1;;5718;;HISEQ:183:C37UYACXX:7:1205:12409:10531_1:N:0:GCCAAT;;5789;;A:T
:::;>C1;;5700;;HISEQ:183:C37UYACXX:8:1112:6117:48546_1:N:0:GTGAAA;;5789;;A:T
5718  GATGTGTTTGTGGGTGGAGGTGAATATGTGAGTATTTTTGTATAATTAAATTAATAAAAAATTTTTTTTTT-T-AGTTTGATGAAGAGTTTTTGATT
5718  GATGTGTTTCGTTGGTGGAGGTGAATACGTGAGTATTCCTTGTAATAATTCAATTAATAAAAACTTTTTTTTTT-A-AGTCTGATGAAGAGCTTCTGATCA
5718  GATGTGTTTGTGGGTGGAGGTGAATATGTGAGTATTTTTGTATAATTAAATTAATAAAAAATTTTTTTTTT-A-AGTTT
```

### .snps

Chr	Pos	OtherPbase	Thispbase
1	1682	C	A
1	3120	A	T

1298356

802183



Find Parent specific SNPs and  
Filter.

330860

116690



Find reads spanning the SNPs

310816

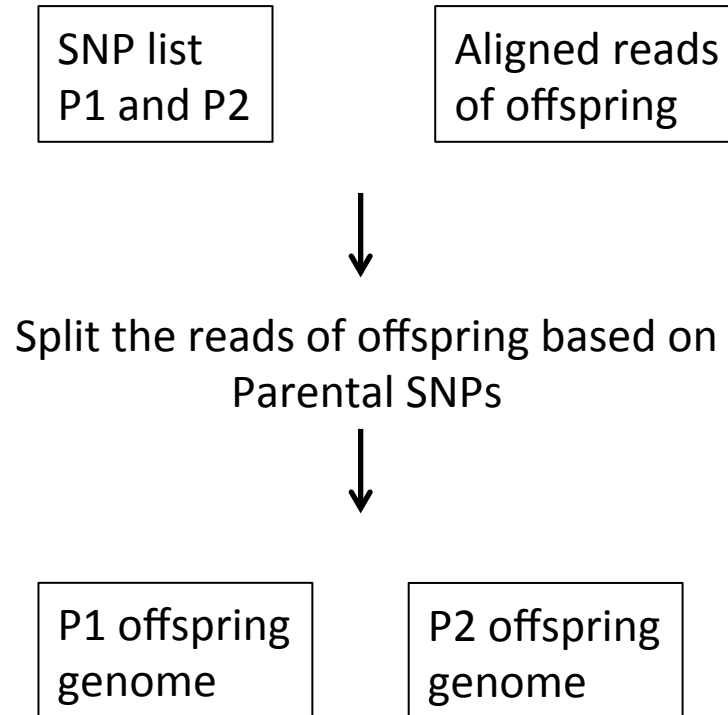
100255



Verify the SNPs from the raw  
reads

320339

#### 4 Split the offspring reads into parental origin genomes



12 SNP scenarios

Strand conflict on  
Different strand

Parental SNP

P1	P2
G	C
C	G
A	T
T	A
G	T
T	G
C	A
A	C

Strand conflict on  
Same strand

P1	P2
G	A
A	G
C	T
T	C

Possibility in  
Offspring

P1	P2
G	C/T
C/T	G

P1	P2
G	A
C/T	T

---

In Parent		Possibility in Offspring			
		Top strand		Bottom strand	
P1	P2	P1	P2	P1	P2
G	C	G	C/T	C/T	C/T
C	G	C/T	G	G	G
A	T	A	T	T	T
T	A	T	A	A	A
G	T	G	T	T	C/T
T	G	T	G	G	A
C	A	C/T	A	A	G
A	C	A	C/T	C/T	T

---