

D-statistic (ABBA-BABA test) 计算

计算方法

D值计算：

$$D(P_1, P_2, P_3, P_4) = \frac{\sum_{i=1}^n [(1-\hat{p}_{i1})\hat{p}_{i2}\hat{p}_{i3}(1-\hat{p}_{i4}) - \hat{p}_{i1}(1-\hat{p}_{i2})\hat{p}_{i3}(1-\hat{p}_{i4})]}{\sum_{i=1}^n [(1-\hat{p}_{i1})\hat{p}_{i2}\hat{p}_{i3}(1-\hat{p}_{i4}) + \hat{p}_{i1}(1-\hat{p}_{i2})\hat{p}_{i3}(1-\hat{p}_{i4})]}$$

P₁ 和 P₂：待检验群体

P₃：基因交流可能的来源群体

P₄：外群

i：SNP位点

\hat{p}_{i1} ：表示SNP i 在P1中的频率

可理解为在 i 位点处alt的频率， $(1 - \hat{p}_{i1})$ 即ref的频率；

在ABBA和BABA两种模式中，P₄ 都是A，所以总是 $(1 - \hat{p}_{i4})$

引自Durand^[1]

显著性检验：

使用jackknife方法，所有文献都用的这个方法

程序实现

1. 针对每一位点，计算群体中alt的频率

CHROM	POS	A1	A2	A3	B1	B2	B3	O
chr1	100	0/0	0/0	0/1	1/1	0/1	0/1	0/1
chr1	218	1/1	0/1	0/1	0/0	./.	0/1	1/1

对于第一个位点， $\hat{p}_{iA} = \frac{N_{alt}}{N_{total}} = 1/6$ ； $\hat{p}_{iB} = 4/6$ ； $\hat{p}_{iO} = 1/2$

对于第二个位点， $\hat{p}_{iA} = 4/6$ ； $\hat{p}_{iB} = 1/4$ ； $\hat{p}_{iO} = 1$

2. 依据公式得到全基因组的D_{stat}

3. 按照无重叠的滑窗方法计算每一个窗口的D_{stat-w}

4. 使用jackknife做显著性检验^[2]，具体是：

a. 每次从得到的一组D_{stat-w}中去掉一个，剩余的数据组成D_{jack}，每组D_{jack}可以得到一个伪值，最终可以得到一组伪值。 $D_{pseudo} = D_{stat} * N - \bar{D}_{jack} * (N - 1)$ ，其中N 为总的窗口数

b. 计算标准误、z值和p值。

$$D_{stat} \sim N(0,1)$$

$$std_err = \sqrt{\frac{var(D_{pseudo})}{N}}$$

$$Z = \frac{D_{stat}}{std_err}$$

脚本使用

路径:/p299/user/og03/chenquan1609/Resequencing/KF-CQ-B1-20160505-01_honeybee/08.ABBA_BABA

```

1 python Dstat.py -h
2
3 usage: Dstat.py [-h] -v VCF -p1 POP1 -p2 POP2 -p3 POP3 -o OUTGROUP -w WINDOW
4
5 D-statistic(ABBA-BABA test) for SNP
6
7 optional arguments:
8   -h, --help    show this help message and exit
9   -v VCF        population vcf file
10  -p1 POP1       population1 sample list, 1 sam per line
11  -p2 POP2       population2 sample list, 1 sam per line
12  -p3 POP3       population2 sample list, 1 sam per line
13  -o OUTGROUP    outgroup sample list, 1 sam per line
14  -w WINDOW      window for genome-wide scan, larger than LD

```

参考文献

- [1] Durand E Y, Patterson N, Reich D, et al. Testing for ancient admixture between closely related populations[J]. Molecular biology and evolution, 2011, 28(8): 2239-2252.
- [2] Martin S H, Dasmahapatra K K, Nadeau N J, et al. Genome-wide evidence for speciation with gene flow in Heliconius butterflies[J]. Genome Research, 2013, 23(11): 1817-1828.