

Popgen definitions and tests

gwct

11/4/2019

0. Info

file: tamias_84ind_TargetsOnly_Final.fasta

groups: amoenus, cratericus, minimus, ruficaudus

command:

```
python3 fa_pop.py -i tamias_84ind_TargetsOnly_Final.tfa -g amoenus,cratericus,minimus,ruficaudus  
-p 8 -het
```

1. Overall:

1.1 Definitions

Term	Definition
Total sites	All sites where every individual is either homozygous or heterozygous.
Invariant sites	The number of sites where every individual is homozygous for the same allele.
Polymorphic sites	The number of sites where at least one individual is heterozygous or homozygous for a different allele.
Pi	The expected number of nucleotide differences between two sequences.
Pi per site	Pi divided by the number of sites considered.

1.2 Results:

Type	Count
Total sites	8678441
Invariant sites	8112324
Polymorphic sites	566117
Pi	42810.036
Pi per site	0.00493

2. Within group

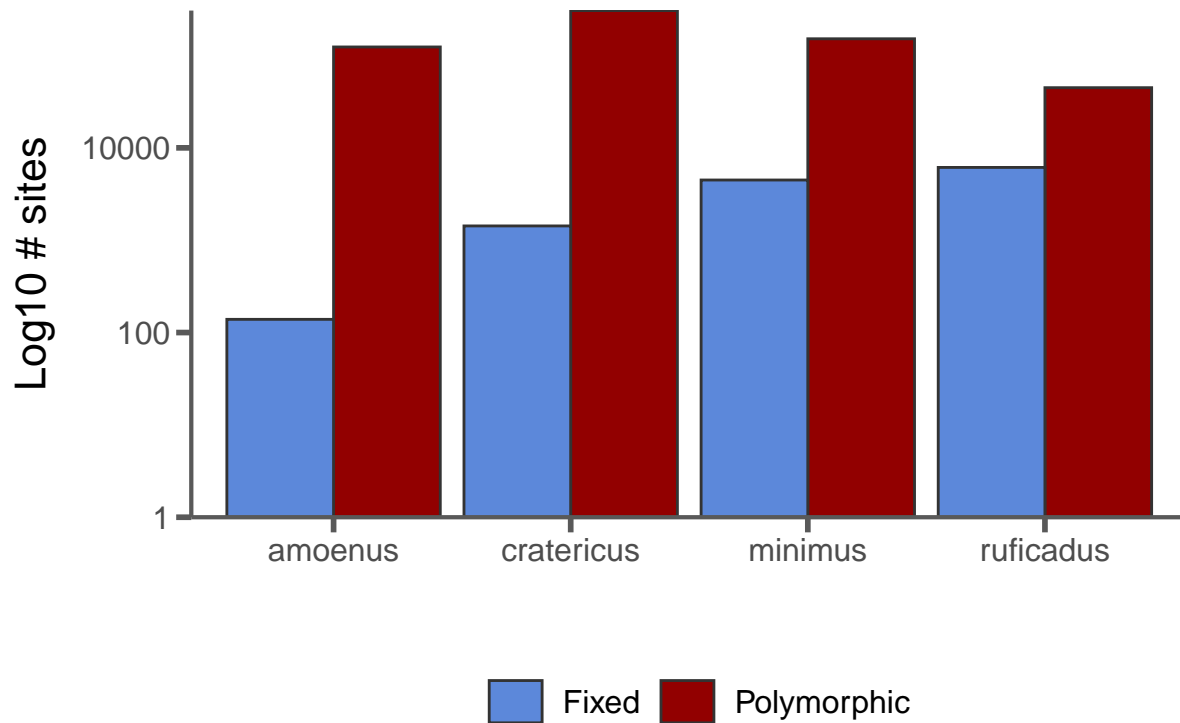
2.1 Definitions

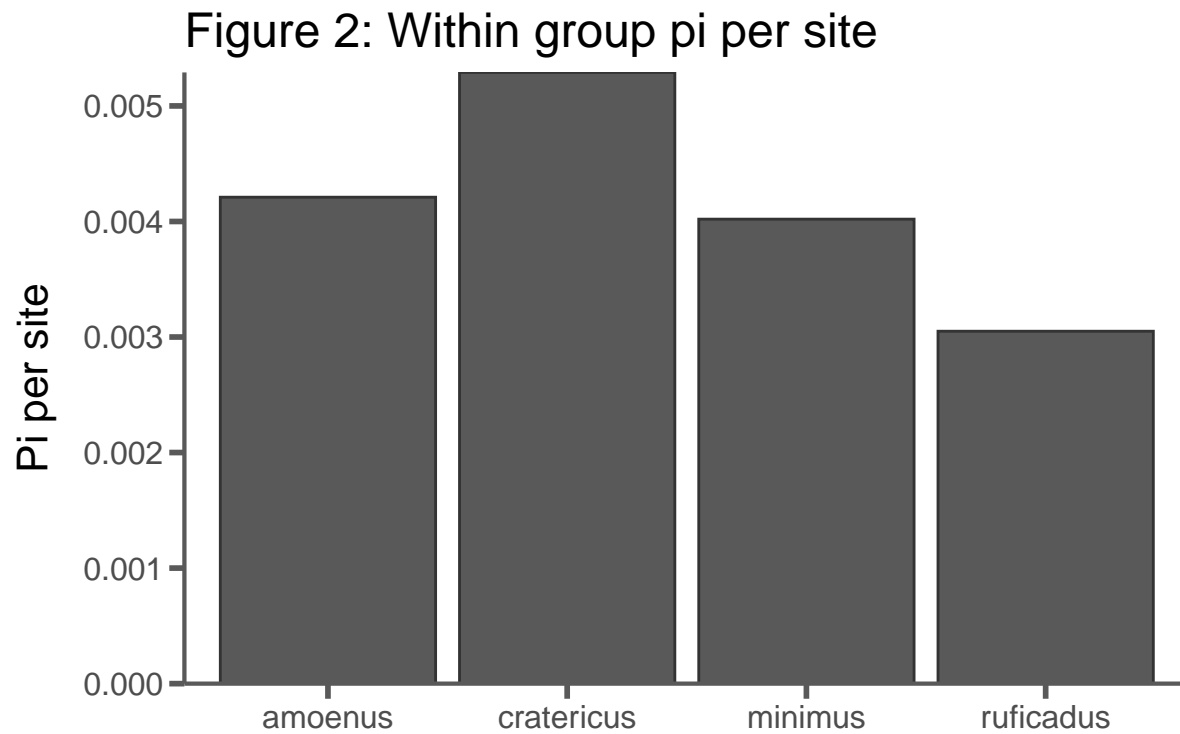
Term	Definition
Fixed site	Sites where all individuals in this group share the same allele.
Polymorphic site	Sites where at least one individual within the group is heterozygous or homozygous for a different allele.
Pi	The expected number of nucleotide differences between two sequences within this group.
Pi per site	Pi divided by the number of sites considered.

2.2 Results

Group	Fixed	Polymorphic	Pi	Pi per site
amoenus	139	123722	36545.143	0.00421
cratericus	1427	306711	45904.983	0.00529
minimus	4486	151693	34913.438	0.00402
ruficadus	6137	44810	26493.5	0.00305

Figure 1: Within group site counts





3. Between group

3.1 Definitions

Term	Definition
Shared fixed differences	Sites where these two groups are fixed for one allele that is different from other groups.
Paired fixed differences	Sites where these two groups are fixed for different alleles relative only to each other.
Paired Pi	The expected number of nucleotide differences between two sequences among both groups.
Paired Pi per site	Paired pi divided by the number of sites considered.

3.2 Results

Shared fixed differences:

	amoenus	cratericus	minimus	ruficadus
amoenus	-	3885	4651	18465
cratericus		-	1460	6180
minimus			-	42258
ruficadus				-

Paired fixed differences:

amoenus	cratericus	minimus	ruficadus	
amoenus	-	214550	319622	405967
cratericus		-	150820	228071
minimus			-	368455
ruficadus				-

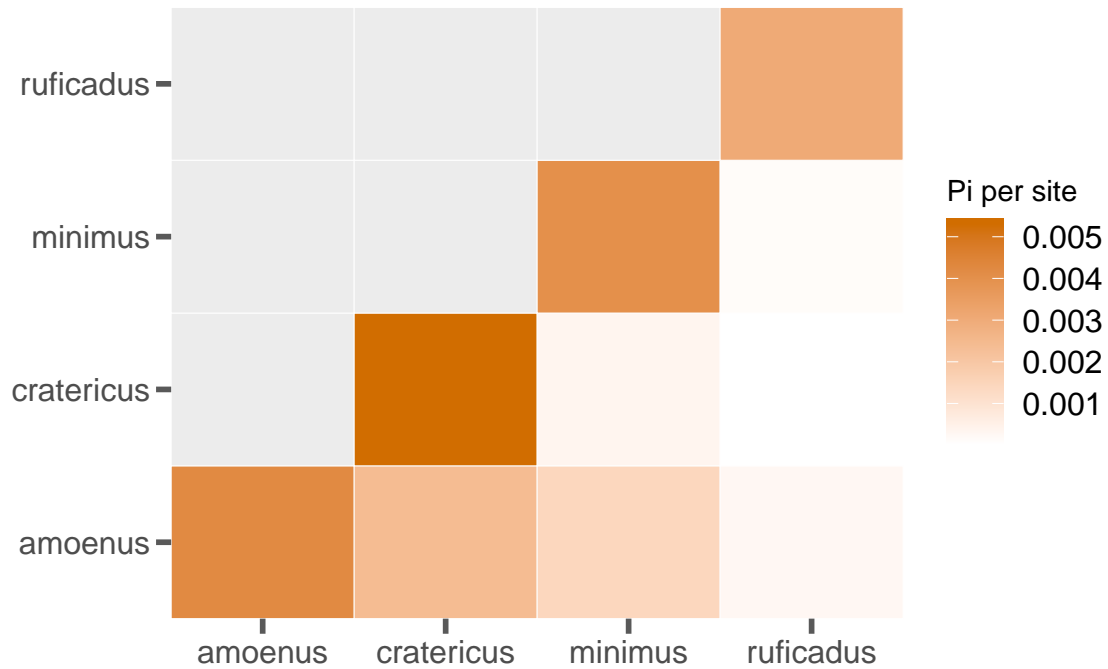
Paired Pi:

amoenus	cratericus	minimus	ruficadus	
amoenus	-	20971.50769	12422.65217	2602.88889
cratericus		-	3264.37838	318.53333
minimus			-	1396.0
ruficadus				-

Paired Pi per site:

amoenus	cratericus	minimus	ruficadus	
amoenus	-	0.00242	0.00143	0.0003
cratericus		-	0.00038	4e-05
minimus			-	0.00016
ruficadus				-

Figure 3: Pi per site between groups



Diagonal is within group pi per site.

4. Other considerations/things I'm working on

1. Why is Pi per site higher within groups than between groups? This doesn't make sense to me...
2. Currently only considering sites without missing info.
3. Using observed heterozygosity instead of expected.