

test_panmixia_among_2pop

September 22, 2021

```
[27]: import dadi
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import scipy.stats as st
import pylab
```

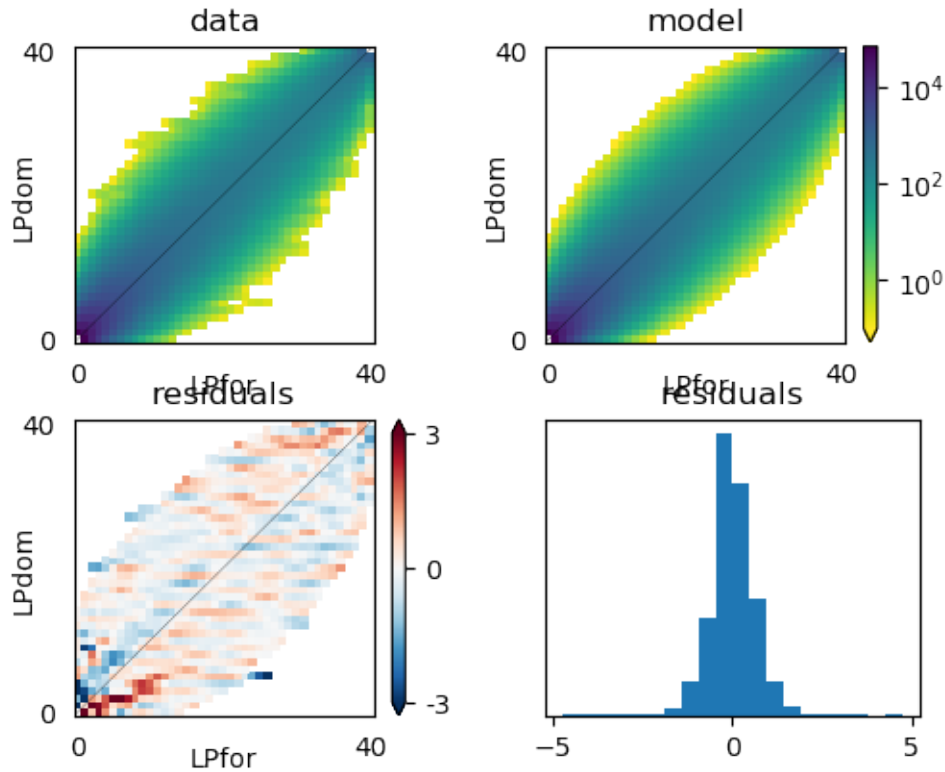
```
[29]: # Parse the VCF file to generate a data dictionary
datafile = '/home/daron/bioInf/wilding/vcf_store/anopheles-rose.3.
↳unifiedGenotyper.cov14x.passQC.merge.pol.vcf.gz'
dd = dadi.Misc.make_data_dict_vcf(datafile, '/home/daron/bioInf/wilding/
↳popstructure/popdifferentiation/dadi/test_panmixia/wilding.popfile.txt')
```

0.1 Test panmixia from LPdom and LPfor

```
[39]: # Extract the joint spectrum asf
pop_ids, ns = ['LPdom', 'LPfor'], [40,40]
fs_lp_dom_for = dadi.Spectrum.from_data_dict(dd, pop_ids, ns)
fs_lp_dom_for.scram = fs_lp_dom_for.scramble_pop_ids()

cmp_fig = pylab.figure(1)
cmp_fig.clear()
dadi.Plotting.plot_2d_comp_multinom(fs_lp_dom_for.scram, fs_lp_dom_for, vmin=0.
↳1, resid_range=3, pop_ids = ('LPdom', 'LPfor'))
cmp_fig.savefig('/home/daron/bioInf/wilding/popstructure/popdifferentiation/
↳dadi/test_panmixia/LPdom_LPfor.model-data.comparison.pdf')
```

Warning: converting a masked element to nan.



```
[40]: ### calculate chi2 in observed data

chi2obs = numpy.sum((fs_lp_dom_for - fs_lp_dom_for.scram)**2/numpy.
    ↳sqrt((fs_lp_dom_for+fs_lp_dom_for.scram/2)))#
chi2obs
```

[40]: 22779.07788979063

```
[41]: ### calculate chi2 in permuted bootstrap
null_chi2 = []
for i in range(0,1000):
    permutedInFile = "/home/daron/bioInf/wilding/popstructure/
    ↳popdifferentiation/dadi/test_panmixia/dadi_permuted/anopheles-rose.3.
    ↳unifiedGenotyper.cov14x.passQC.merge.pol.LPdom_LPfor.fs.permuted_"+str(i)
    fs_permuted = dadi.Spectrum.from_file(permutedInFile)
    fs_permuted_scram = fs_permuted.scramble_pop_ids()
    chi2 = np.nansum((fs_permuted - fs_permuted_scram)**2/np.
    ↳sqrt((fs_permuted+fs_permuted_scram/2)))
    null_chi2.append(chi2)
```

Warning: converting a masked element to nan.

```
[36]: ### calculate pvalue

# add chi2obs to chi2 null distri to make sure the value will be within the
↳distri
null_chi2_dist = null_chi2
null_chi2_dist = numpy.append(null_chi2_dist, chi20bs)

nbins = 25
#n, bins, _ = plt.hist(null_chi2, nbins)
n, bins = np.histogram(null_chi2_dist, nbins)

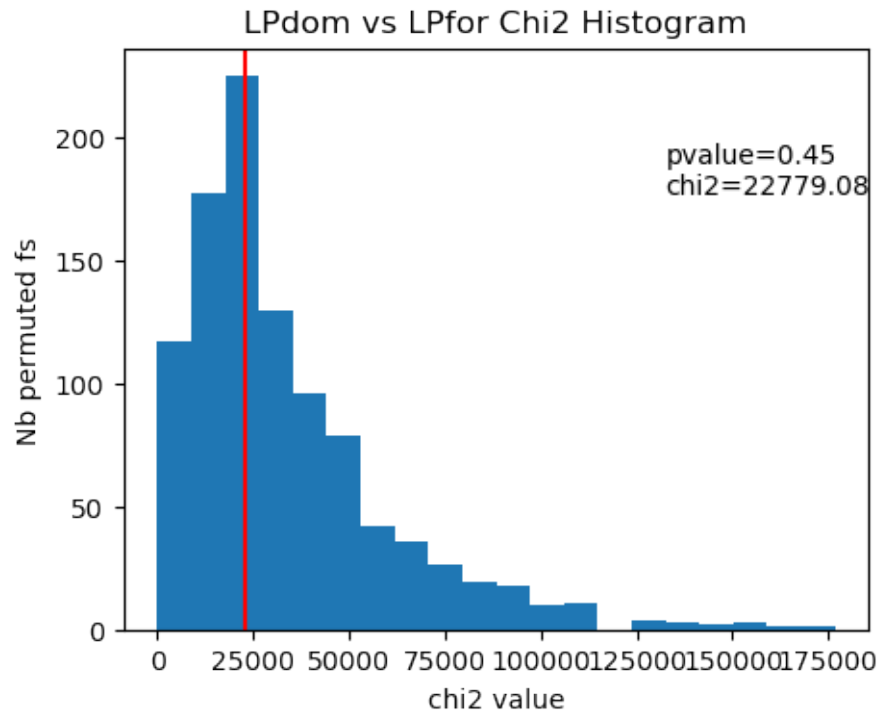
# get the width of each bin
binChi20bsIdx = np.where(bins == bins[bins>=chi20bs][0])[0]

# get the width of each bin
bin_width = bins[1] - bins[0]

# sum over number in each bin and mult by bin width, which can be factored out
integral_given_area = bin_width * sum(n[int(binChi20bsIdx):len(n)])
integral_total = bin_width * sum(n[0:len(n)])
pvalue = integral_given_area/integral_total
pvalue
```

[36]: 0.4465534465534466

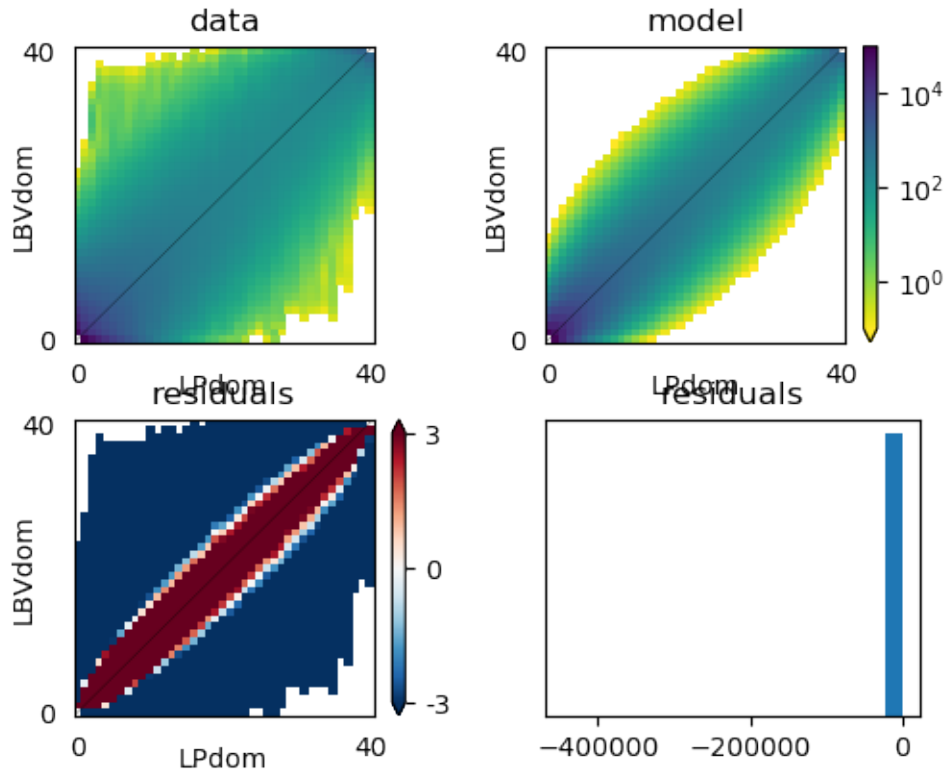
```
[38]: # hist of null distribution
plt.rcParams.update({'figure.figsize':(5,4), 'figure.dpi':100})
n, bins, _ = plt.hist(null_chi2, bins=20)
bin_width = bins[1] - bins[0]
plt.axvline(x=chi20bs, color="red")
x = np.amax(bins)-(bin_width*5)
y = np.unique(np.sort(n))[len(np.unique(n))-2]
text = "pvalue="+str(round(pvalue, 2))+"\n"+"chi2="+str(round(chi20bs, 2))
plt.text(x,y,text)
plt.gca().set(title='LPdom vs LPfor Chi2 Histogram', ylabel='Nb permuted fs',
↳xlabel='chi2 value')
plt.savefig('/home/daron/bioInf/wilding/popstructure/popdifferentiation/dadi/
↳test_panmixia/LPdom_LPfor.chi2Distri.pdf')
```



0.2 Test panmixia from LBVdom and LPdom

```
[42]: pop_ids, ns = ['LBVdom', 'LPdom'], [40,40]
fs_lbv_lpdom = dadi.Spectrum.from_data_dict(dd, pop_ids, ns)
fs_lbv_lpdom.scram = fs_lbv_lpdom.scramble_pop_ids()

dadi.Plotting.plot_2d_comp_multinom(fs_lbv_lpdom.scram, fs_lbv_lpdom, vmin=0.1,
↳ resid_range=3, pop_ids = ('LBVdom', 'LPdom'))
```



```
[43]: chi20bs = numpy.sum((fs_lbv_lpdom - fs_lbv_lpdom.scram)**2/numpy.
      ↪sqrt((fs_lbv_lpdom+fs_lbv_lpdom.scram/2)))
      chi20bs
```

[43]: 18457850.560944583

```
[45]: # calculate chi2 in permuted bootstrap
null_chi2 = []
for i in range(0,1000):
    permutedInFile = "/home/daron/bioInf/wilding/popstructure/
    ↪popdifferentiation/dadi/test_panmixia/dadi_permuted/anopheles-rose.3.
    ↪unifiedGenotyper.cov14x.passQC.merge.pol.LBVdom_LPdom.fs.permuted_"+str(i)
    fs_permuted = dadi.Spectrum.from_file(permutedInFile)
    fs_permuted.scram = fs_permuted.scramble_pop_ids()
    chi2 = numpy.nansum((fs_permuted - fs_permuted.scram)**2/numpy.
    ↪sqrt((fs_permuted+fs_permuted.scram/2)))
    null_chi2.append(chi2)
```

Warning: converting a masked element to nan.

```
[46]: ### calculate pvalue

# add chi2obs to chi2 null distri to make sure the value will be within the
↳distri
null_chi2_dist = null_chi2
null_chi2_dist = numpy.append(null_chi2_dist, chi20bs)

nbins = 25
#n, bins, _ = plt.hist(null_chi2, nbins)
n, bins = np.histogram(null_chi2_dist, nbins)

# get the width of each bin
binChi20bsIdx = np.where(bins == bins[bins>=chi20bs][0])[0]

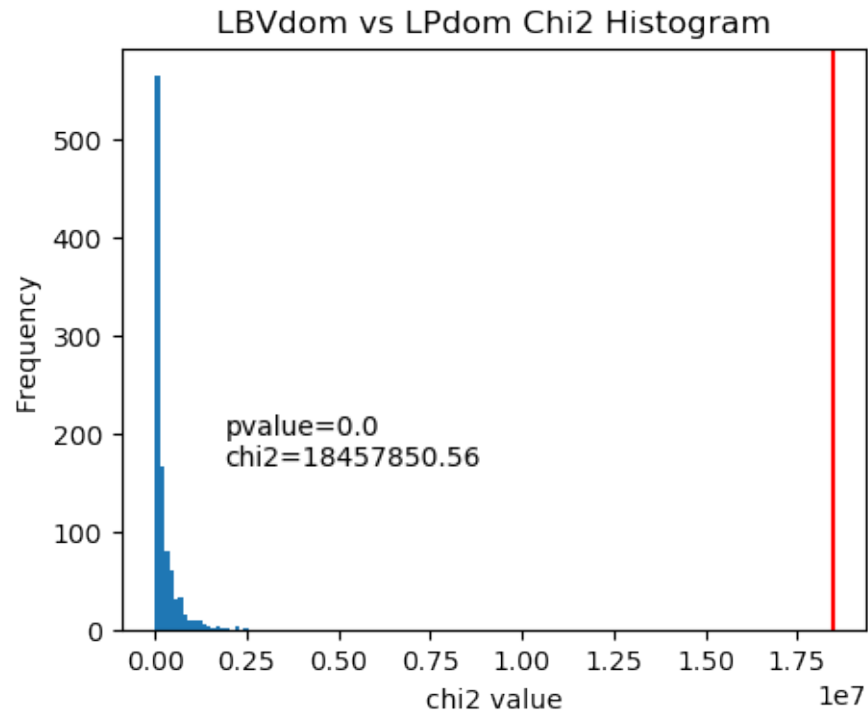
# get the width of each bin
bin_width = bins[1] - bins[0]

# sum over number in each bin and mult by bin width, which can be factored out
integral_given_area = bin_width * sum(n[int(binChi20bsIdx):len(n)])
integral_total = bin_width * sum(n[0:len(n)])
pvalue = integral_given_area/integral_total
pvalue
```

[46]: 0.0

```
[47]: # hist of null distribution
plt.rcParams.update({'figure.figsize':(5,4), 'figure.dpi':100})
n, bins, _ = plt.hist(null_chi2, bins=20)
bin_width = bins[1] - bins[0]
plt.axvline(x=chi20bs, color="red")
x = np.amax(bins)-(bin_width*5)
y = np.unique(np.sort(n))[len(np.unique(n))-2]
text = "pvalue="+str(round(pvalue, 2))+"\n"+"chi2="+str(round(chi20bs, 2))
plt.text(x,y,text)
plt.gca().set(title='LBVdom vs LPdom Chi2 Histogram', ylabel='Frequency',
↳xlabel='chi2 value')
```

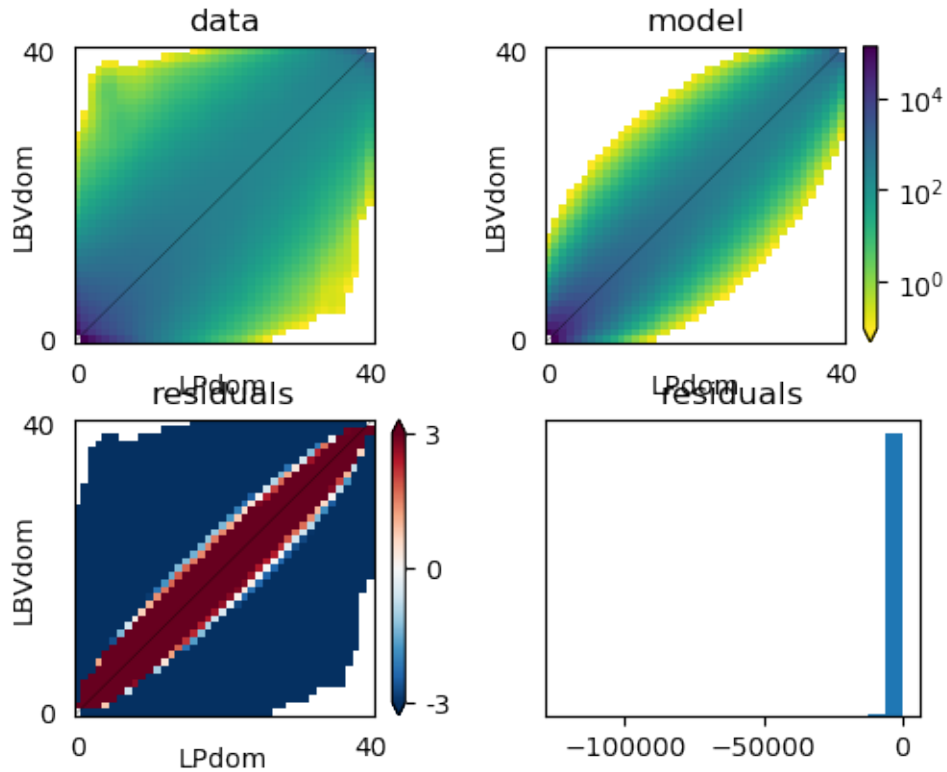
```
[47]: [Text(0, 0.5, 'Frequency'),
Text(0.5, 0, 'chi2 value'),
Text(0.5, 1.0, 'LBVdom vs LPdom Chi2 Histogram')]
```



0.3 Test panmixia from LBVdom and LPfor

```
[48]: pop_ids, ns = ['LBVdom', 'LPfor'], [40,40]
fs_lbv_lpdom = dadi.Spectrum.from_data_dict(dd, pop_ids, ns)
fs_lbv_lpdom.scram = fs_lbv_lpdom.scramble_pop_ids()

dadi.Plotting.plot_2d_comp_multinom(fs_lbv_lpdom.scram, fs_lbv_lpdom, vmin=0.1,
↳resid_range=3, pop_ids = ('LBVdom', 'LPdom'))
```



```
[49]: chi20bs = numpy.sum((fs_lbv_lpdom - fs_lbv_lpdom.scram)**2/numpy.
      ↪sqrt((fs_lbv_lpdom+fs_lbv_lpdom.scram/2)))
chi20bs
```

[49]: 30468962.87874197

```
[50]: # calculate chi2 in permuted bootstrap
null_chi2 = []
for i in range(0,100):
    permutedInFile = "/home/daron/bioInf/wilding/popstructure/
    ↪popdifferentiation/dadi/test_panmixia/dadi_permuted/anopheles-rose.3.
    ↪unifiedGenotyper.cov14x.passQC.merge.pol.LBVdom_LPdom.fs.permuted_"+str(i)
    fs_permuted = dadi.Spectrum.from_file(permutedInFile)
    fs_permuted.scram = fs_permuted.scramble_pop_ids()
    chi2 = numpy.nansum((fs_permuted - fs_permuted.scram)**2/numpy.
    ↪sqrt((fs_permuted+fs_permuted.scram/2)))
    null_chi2.append(chi2)
```

Warning: converting a masked element to nan.


```
[51]: ### calculate pvalue

# add chi2obs to chi2 null distri to make sure the value will be within the
↳distri
null_chi2_dist = null_chi2
null_chi2_dist = numpy.append(null_chi2_dist, chi20bs)

nbins = 25
#n, bins, _ = plt.hist(null_chi2, nbins)
n, bins = np.histogram(null_chi2_dist, nbins)

# get the width of each bin
binChi20bsIdx = np.where(bins == bins[bins>=chi20bs][0])[0]

# get the width of each bin
bin_width = bins[1] - bins[0]

# sum over number in each bin and mult by bin width, which can be factored out
integral_given_area = bin_width * sum(n[int(binChi20bsIdx):len(n)])
integral_total = bin_width * sum(n[0:len(n)])
pvalue = integral_given_area/integral_total
pvalue
```

[51]: 0.0

```
[52]: # hist of null distribution
plt.rcParams.update({'figure.figsize':(5,4), 'figure.dpi':100})
n, bins, _ = plt.hist(null_chi2, bins=20)
bin_width = bins[1] - bins[0]
plt.axvline(x=chi20bs, color="red")
x = np.amax(bins)-(bin_width*5)
y = np.unique(np.sort(n))[len(np.unique(n))-2]
text = "pvalue="+str(round(pvalue, 2))+"\n"+"chi2="+str(round(chi20bs, 2))
plt.text(x,y,text)
plt.gca().set(title='LBVdom vs LPfor Chi2 Histogram', ylabel='Frequency',
↳xlabel='chi2 value')
```

```
[52]: [Text(0, 0.5, 'Frequency'),
Text(0.5, 0, 'chi2 value'),
Text(0.5, 1.0, 'LBVdom vs LPfor Chi2 Histogram')]
```

