Locating OriCs within a Genome

Javin deMello-Folsom
Problem?

- Is there a way to identify the location of potential OriCs within a given genome?
Approach!

- Skew diagrams
  - OriCs found within the valleys
- Define way to find valleys
- Find index of lowest point in those valleys
S. solfataricus Skew Diagram

- **Backrange**
- **Skewdrop**

Nucleotide Index

Skew (\#G - \#C)
Data

- *S. solfataricus* genome
  - NCBI
- Toy sequences
  - Self-made
  - Varying degree of complexity
```python
All.Valleys = []
For i in range(len(seqskew) - backrange):
    If (seqskew[i] + skewdrop) < max(backrange):
        All.Valleys = All.Valleys + seqskew[i]

Valleys = []
NewVal = []
For i in range(len(All.Valleys)-1):
    If All.Valleys[i] == All.Valleys[i+1],[i+1]-1,[i+1]+1:
        NewVal = NewVal + [All.Valleys[i]]
    Else:
        NewVal = NewVal + [All.Valleys[i]]
    Valleys = Valleys + [NewVal]
    NewVal = []

MinInd = 0
OriCs = []
For i in range(len(Valleys)):
    MinInd = Valleys[i].index(min(Valleys[i]))
    OriCs = OriCs + [Indicies[i][MinInd]]
```

Create a list of all potential OriC locations
Separate those into individual groupings
Find lowest value in each grouping
Results

Python 3.6.1 (default, Dec 2015, 13:05:11)
[GCC 4.8.2] on linux

The genome sequence is 1300 nucleotides long.

Progress:
0
1000
OriC count: 4
Wrote to Toy3.png
Time elapsed: 0.01550372838973999 minutes
OriCs: [295, 498, 760, 1097]
Results

- Possible to identify valleys (OriCs)
- Depends on given parameters: backrange, skewdrop