The last universal common ancestor

Relationship between genomic GC content and optimal growth temperature in Bacteria

Norbert Kopocz;
Supervisor: Allen Rodrigo

December 9, 2009

Table of contents

overview

Why is the relationship between genomic G+C content and the optimal growth temperature so interesting?

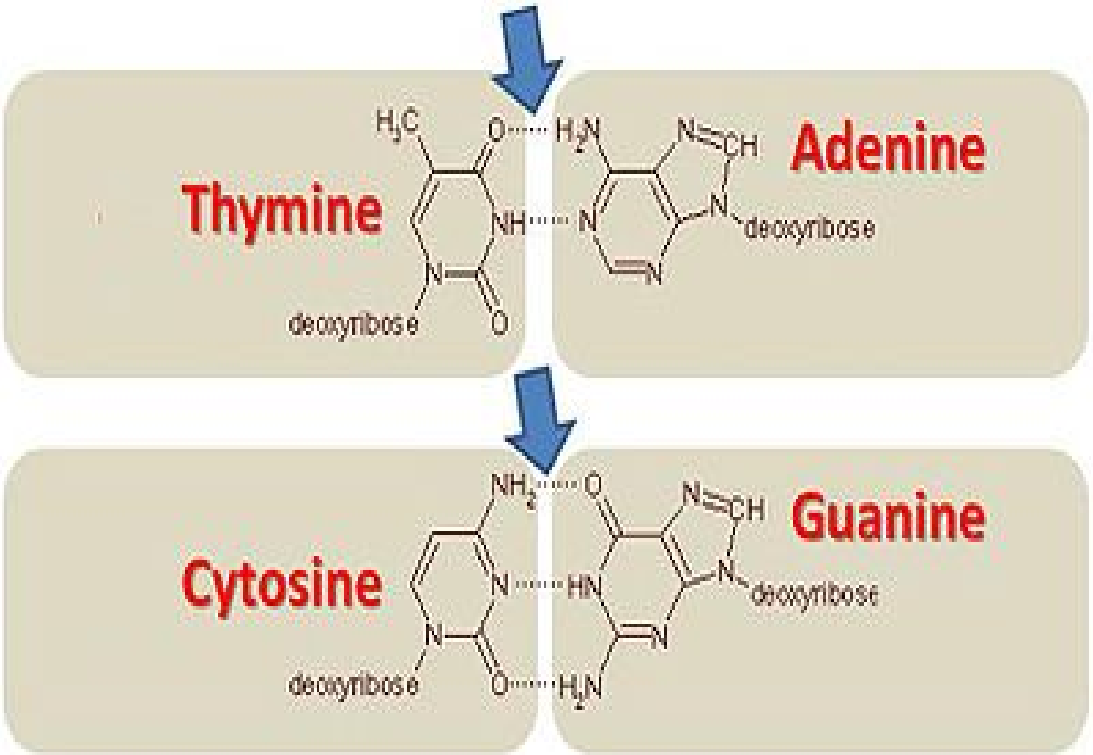
- environmental temperature based mutation of nucleotides

overview

Why is the relationship between genomic G+C content and the optimal growth temperature so interesting?

- environmental temperature based mutation of nucleotides

background

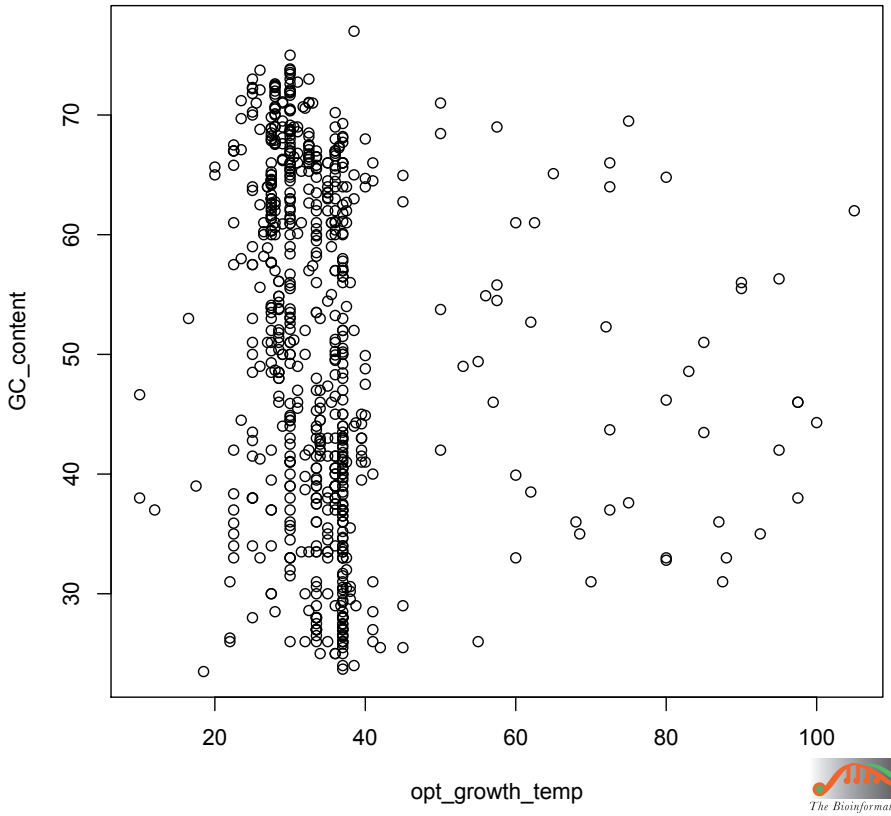


source: http://en.wikipedia.org/wiki/File:AT-GC.jpg

publication by .....

.... Musto et al.

- correlation over all species
- no positive correlation found
- R = -0.167
  p-Value: < 0.00001
  95% confidence intervall:
  -0.238 to -0.094



publication by .....

.... Musto et al.

analyzing families of prokaryotes:

- 20 prokaryotic families
- 15 out of them with positive correlation
- but only 8 with statistically significance

all in all: "T_{opt} is one of the factors that influences genomic GC in prokaryotes"

publication by .....

.... Musto et al.

analyzing families of prokaryotes:

- 20 prokaryotic families
- 15 out of them with positive correlation
- but only 8 with statistically significance

all in all: "T_{opt} is one of the factors that influences genomic GC in prokaryotes"

publication by .....

.... Wang et al.

- "no significance"

analyzing a dataset of 1065 species:

- separating into 5 temperature groups
  - less than 30 °C
  - 30 °C to 40 °C
  - 40 °C to 50 °C
  - 50 °C to 80 °C
  - greater than 80 °C

publication by .....

.... Wang et al.

- results:
  - average genomic GC is highest in the lowest temperature group (less than 30 °C)
  - significant correlation only in low temperature range

Temp. group [°C]	R	p-value
<= 30	0.29	$< 10^{-6}$
30-40	-0.38	$< 10^{-6}$
40-50	0.14	0.41
50-80	-0.21	0.12
>= 80	0.23	0.25

overview

two different ideas:

- consideration of observed values (Musto et al. and Wang et al.)
- our idea:
  - find a correlation between:
    evolutionary change in GC content
    and evolutionary change in optimal growth temperature

overview

why do we consider the evolutionary change in both values?

- different species - different lifestyle:
  - GC poor e.g.: pathogens or symbionts [1] and species with small genomes [2]
  - GC rich: large genomes [1]

[1] EP Rocha, A. Danchin, Base composition bias might result from competition for metabolic resources
[2] N.A. Moran, Microbial Minimalism: Genome Reduction in Bacterial Pathogens

Data

- 706 species (Archaeabacteria and Prokaryotes)
- genomic GC content and optimal growth temperature
- 16S ribosomal RNA sequences from NCBI

creating tree

- alignment over 706 16S rRNA sequences
- editing alignment using program Squint

creating tree

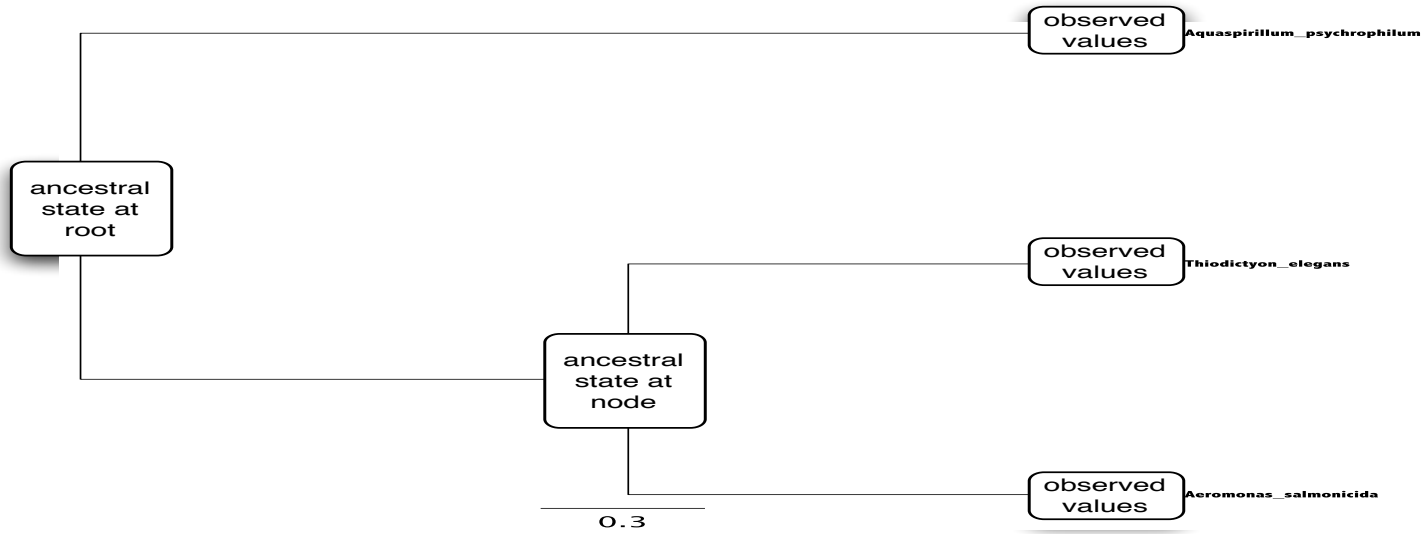
- alignment over 706 16S rRNA sequences
- editing alignment using program Squint
- creating maximum likelihood phylogenetic tree

creating tree

- alignment over 706 16S rRNA sequences
- editing alignment using program Squint
- creating maximum likelihood phylogenetic tree

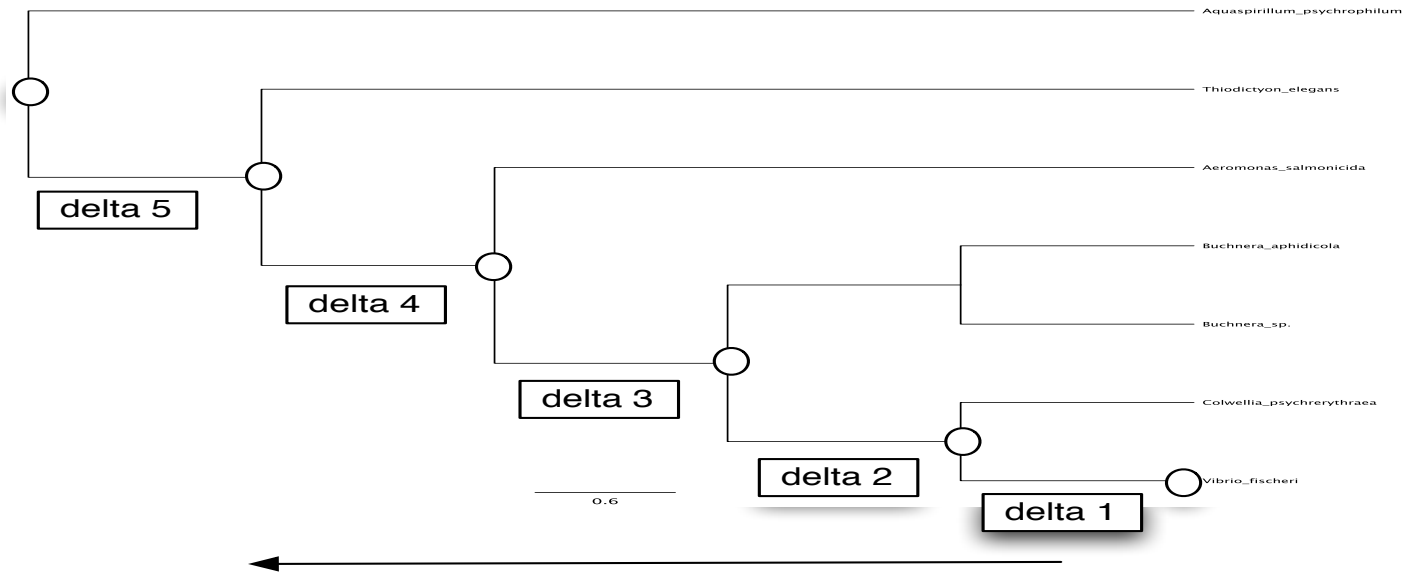
ancestral states

- ancestral states: squared change parsimony method
  - ancestral states for opt. growth temperature
  - and genomic GC content
- Furthermore: program to calculate 16S rRNA GC content



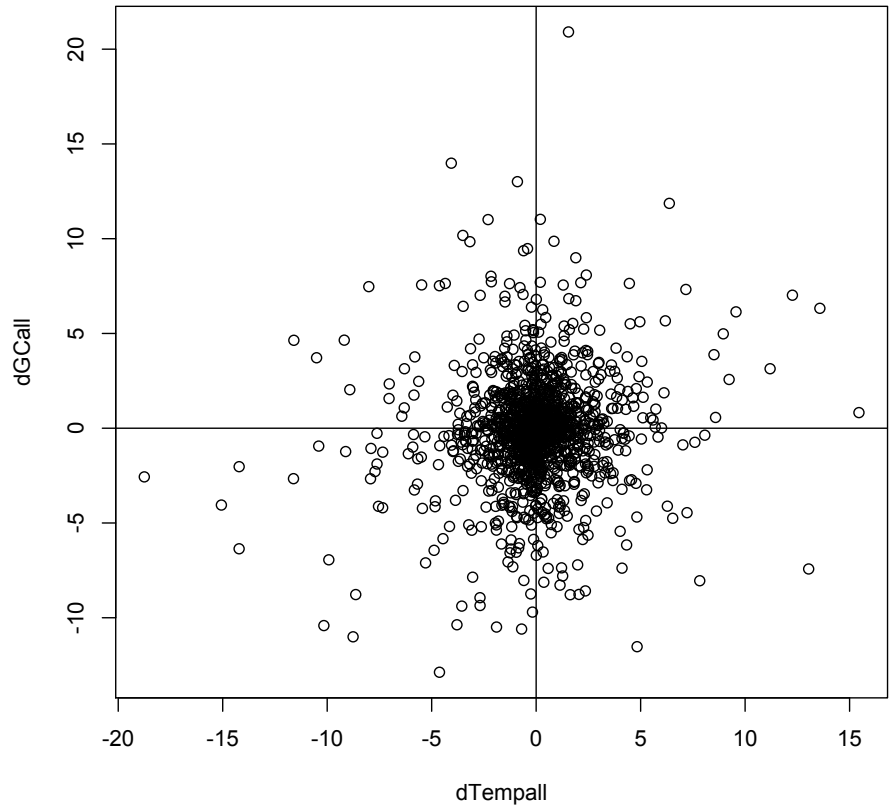
delta values

- set ancestral state values on right position in tree
- calculating delta values using java, jebl library



Δ genomic GC vs. Δ temp

over all ancestral states:

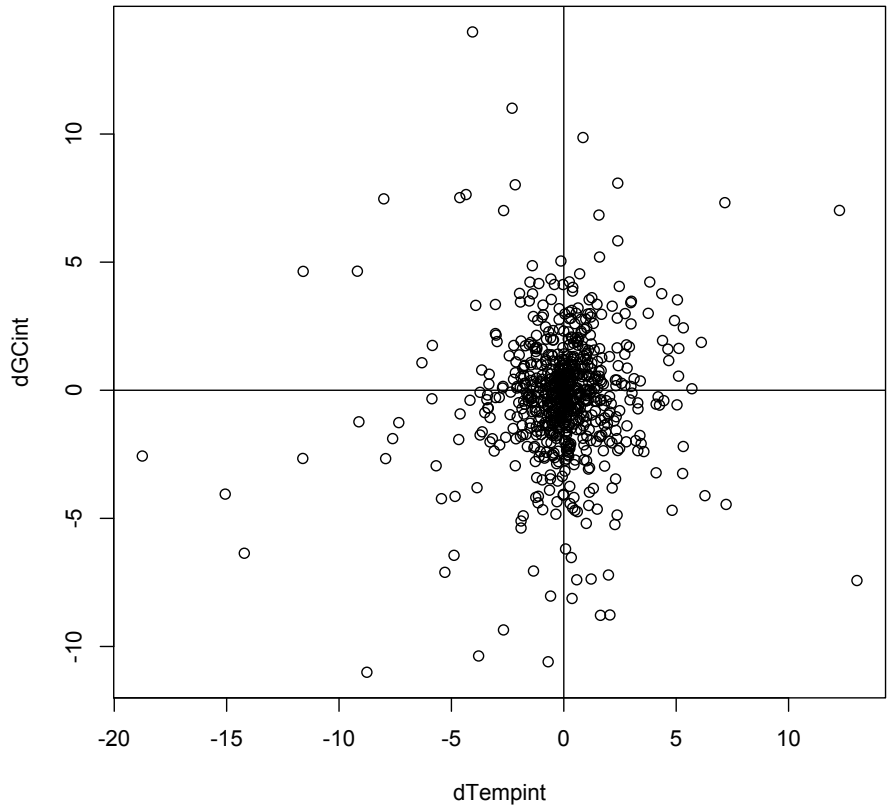


correlation:

- R = 0.104
- p-value: < 0.0001
- 95% confidence intervall: 0.052 to 0.155
- degrees of freedom: 1408

Δ genomic GC vs. Δ temp

over internal nodes:

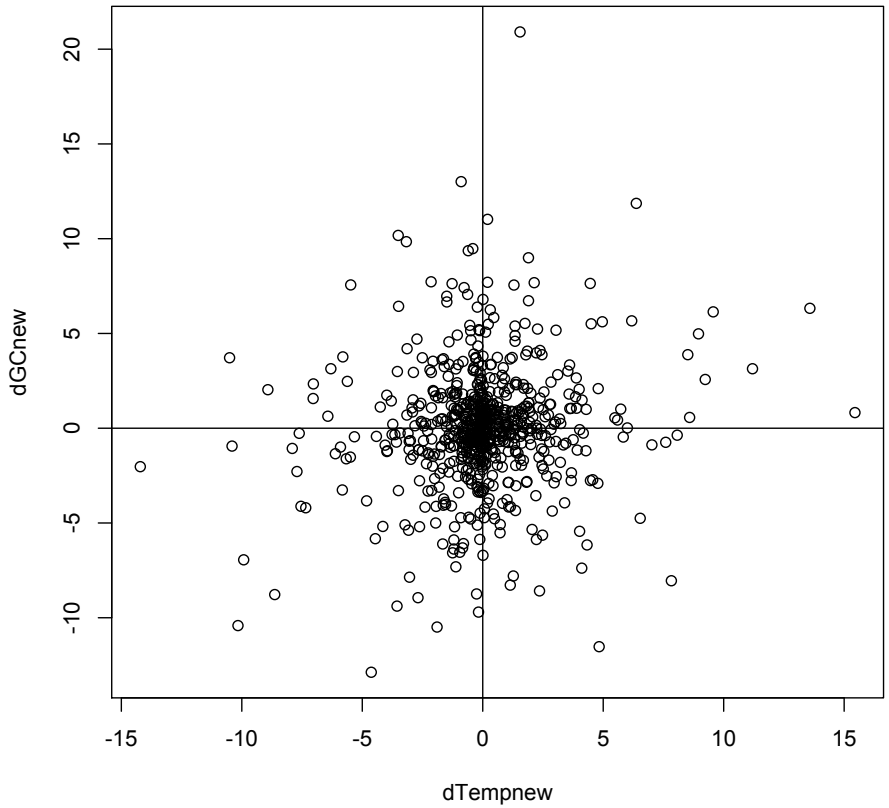


correlation:

- R = 0.068
- p-value: 0.06966
- 95% confidence intervall: -0.0055 0.14158
- degrees of freedom: 702

Δ genomic GC vs. Δ temp

over external nodes:

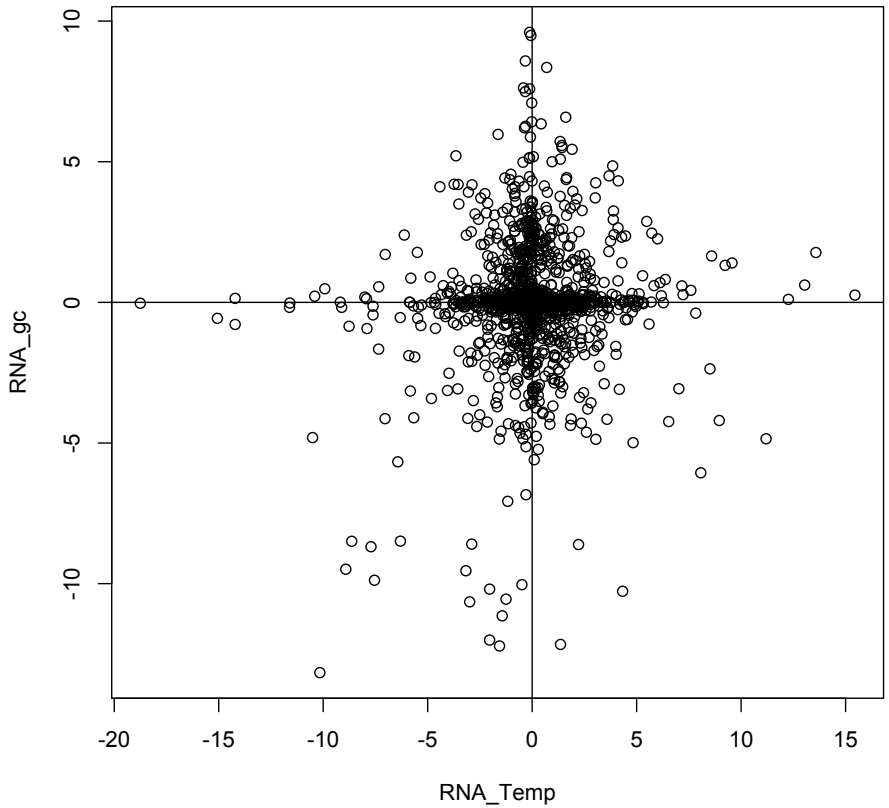


correlation:

- R = 0.128
- p-value: 0.0006
- 95% confidence intervall: 0.055 to 0.2
- degrees of freedom: 704

Δ rRNA GC content vs. Δ temp

over all ancestral states (rRNA):

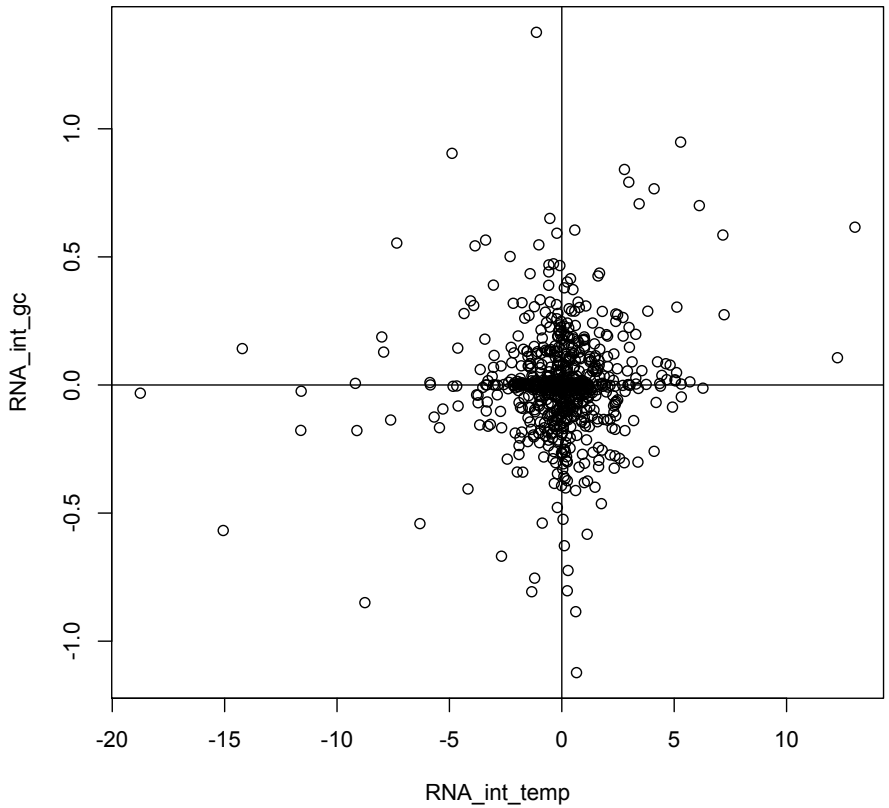


correlation:

- R = 0.094
- p-value: 0.0004
- 95% confidence intervall: 0.042 to 0.145
- degrees of freedom: 1408

Δ rRNA GC content vs. Δ temp

over internal nodes (rRNA):

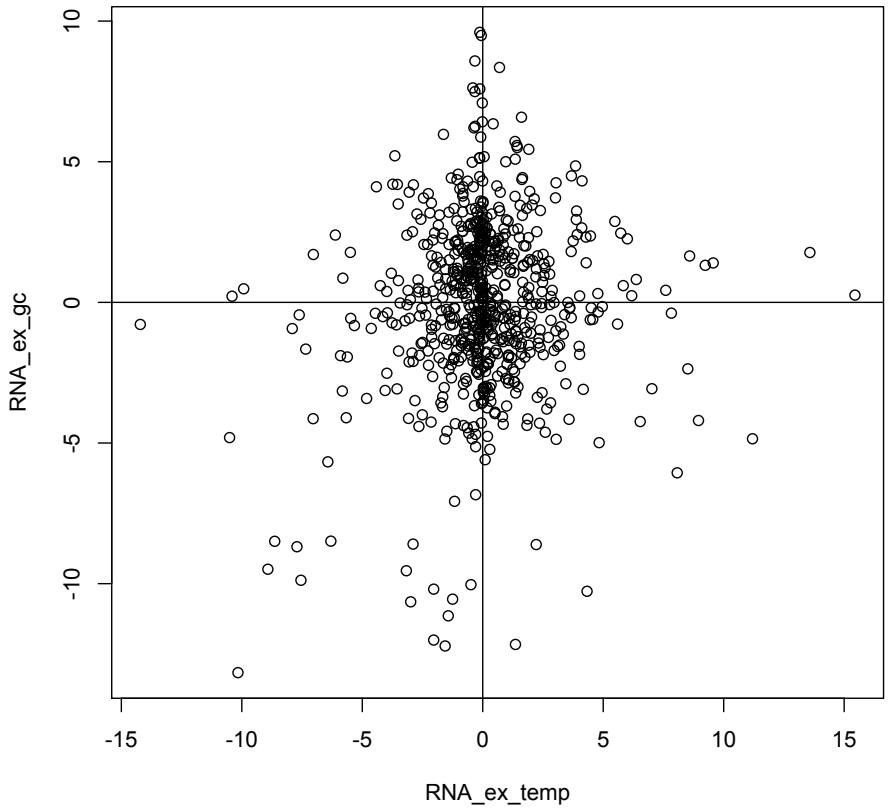


correlation:

- R = 0.115
- p-value: 0.002
- 95% confidence intervall: 0.0416 to 0.1874
- degrees of freedom: 702

Δ rRNA GC content vs. Δ temp

over external nodes (rRNA):



correlation:

- R = 0.12
- p-value: 0.001
- 95% confidence intervall: 0.0464 to 0.1919
- degrees of freedom: 704

summary

- new way to figure out a relationship using:
  - phylogenetic background
  - evolutionary change in both values
  - significance

??Relationship??

- Yes, we can ..... say there is a relationship between:
  - genomic GC content and T_{opt}
  - rRNA GC content and T_{opt}

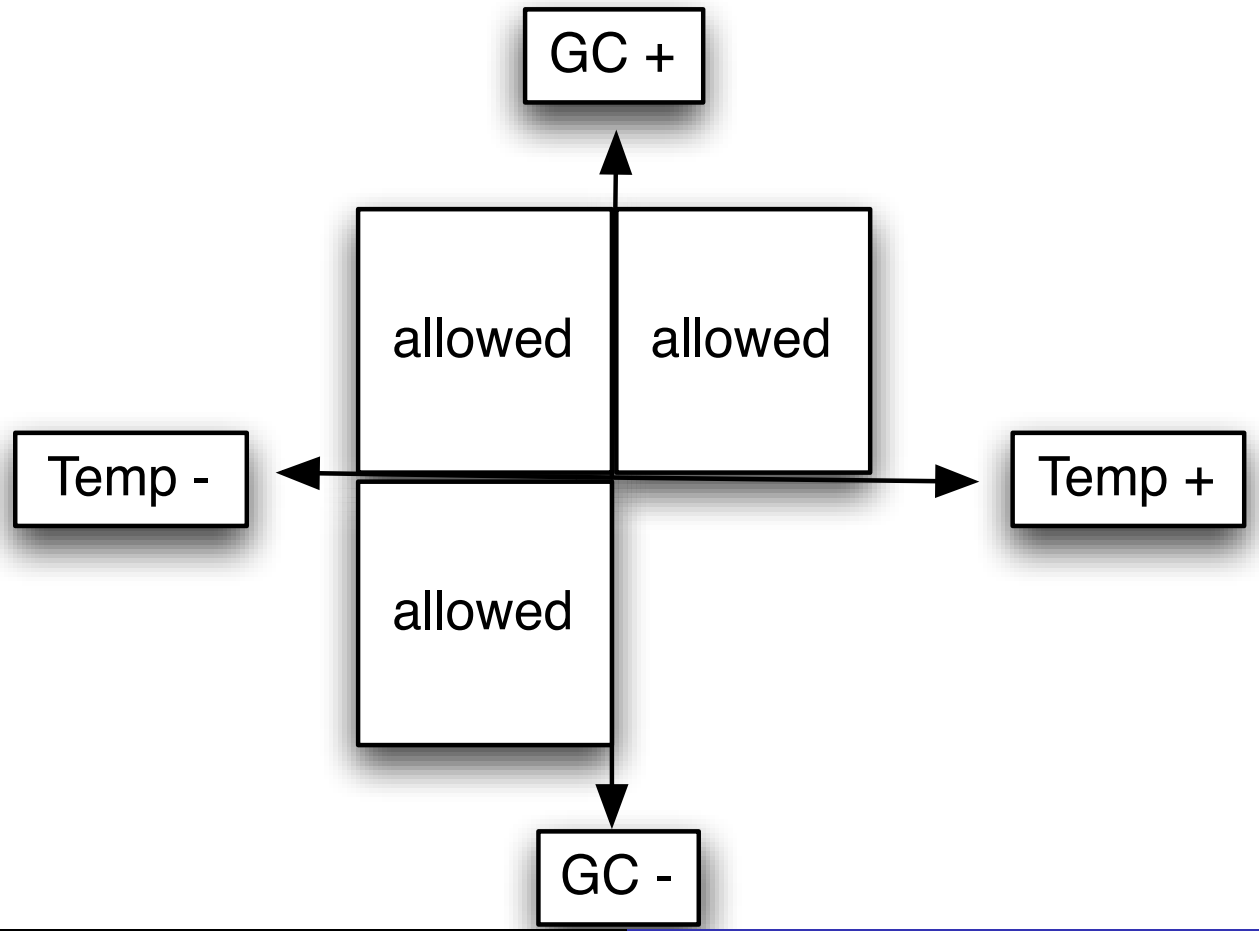
Acknowledgements

Thank you very much for your attention!

- Allen Rodrigo
- Peter Tsai
- Sibon Li
- Kevin Chang
- Bioinformatics Institute

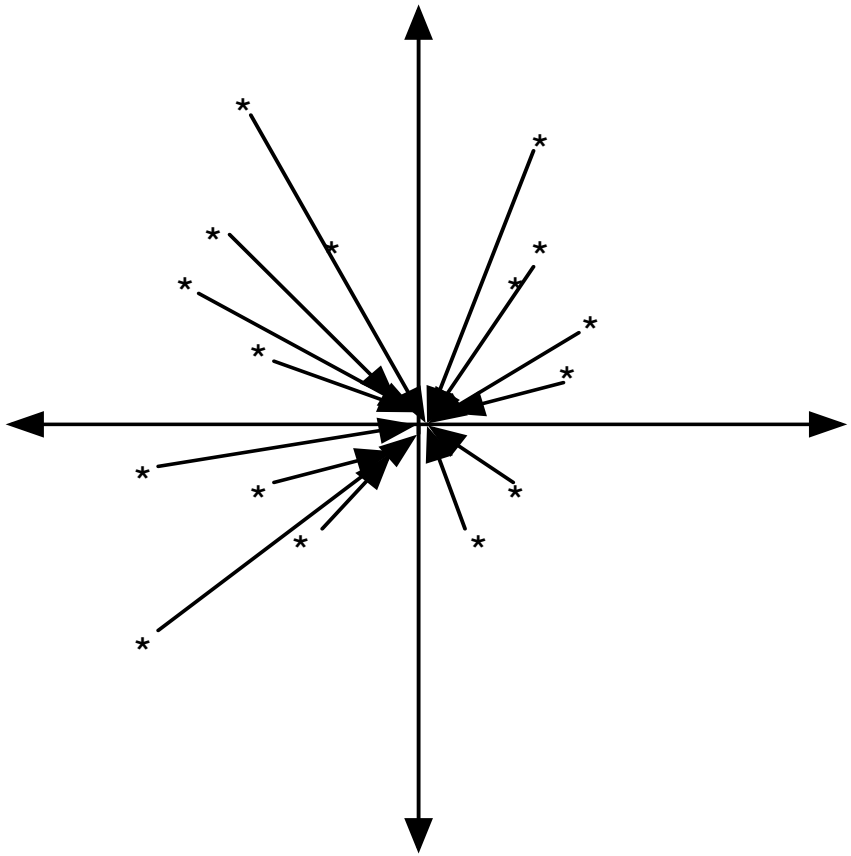
overview

our idea:

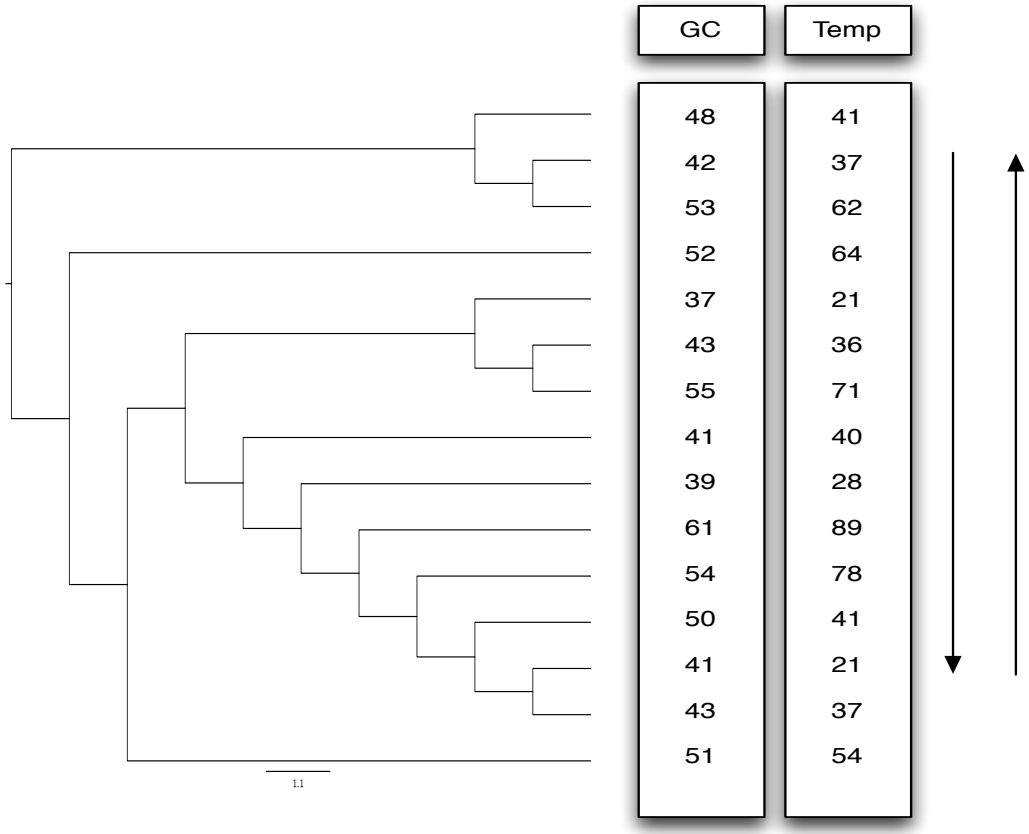


distances

developing a permutation-test of the distances to the origin

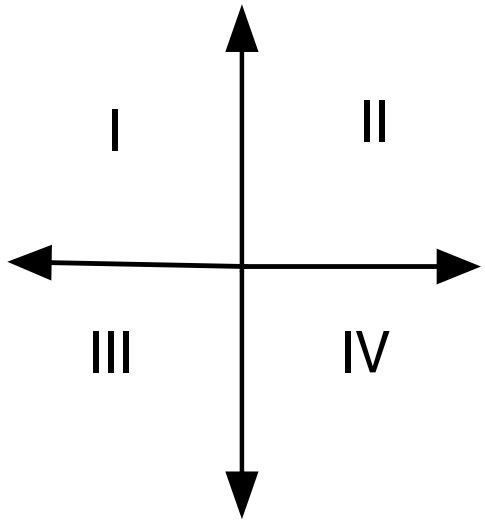


how works my permutation test?

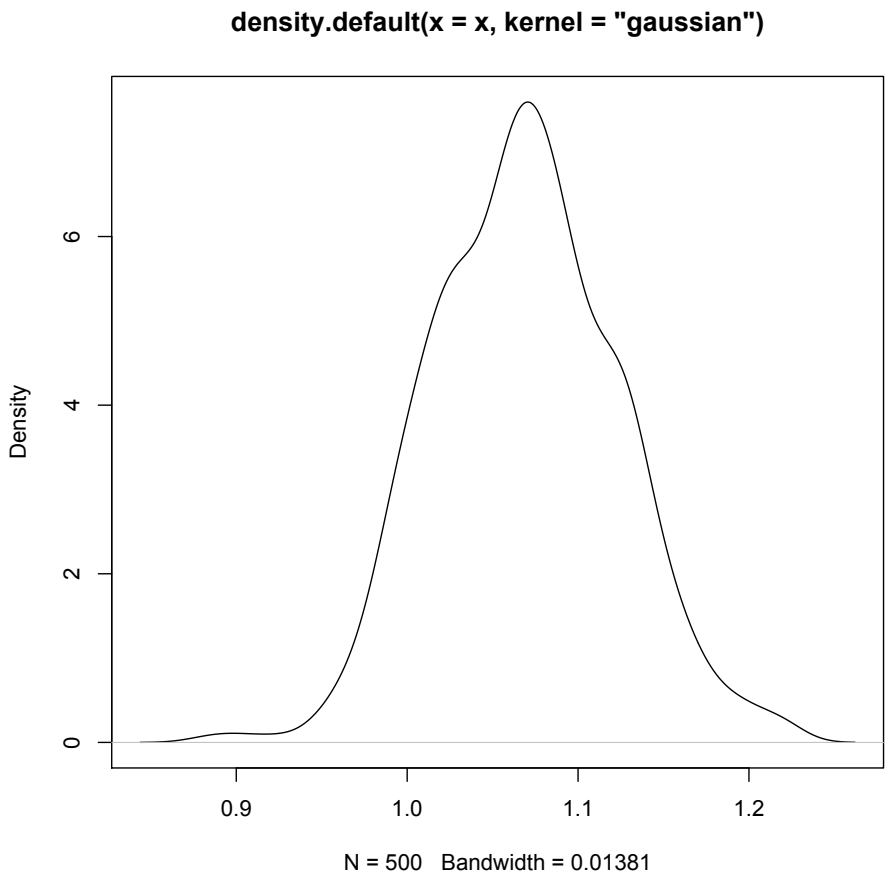


statistic

- permutation test of the distances to the origin
- distance to origin:
  $d = \sqrt[2]{\Delta GC^2 + \Delta Temp^2}$
- $S = \frac{\overline{d}_{IV}}{\overline{d}_{I,II,III}}$



statistic



- number of permutations: N = 500
- observed value for distance proportion S = 0.909
- significant for $\alpha = 1$ %