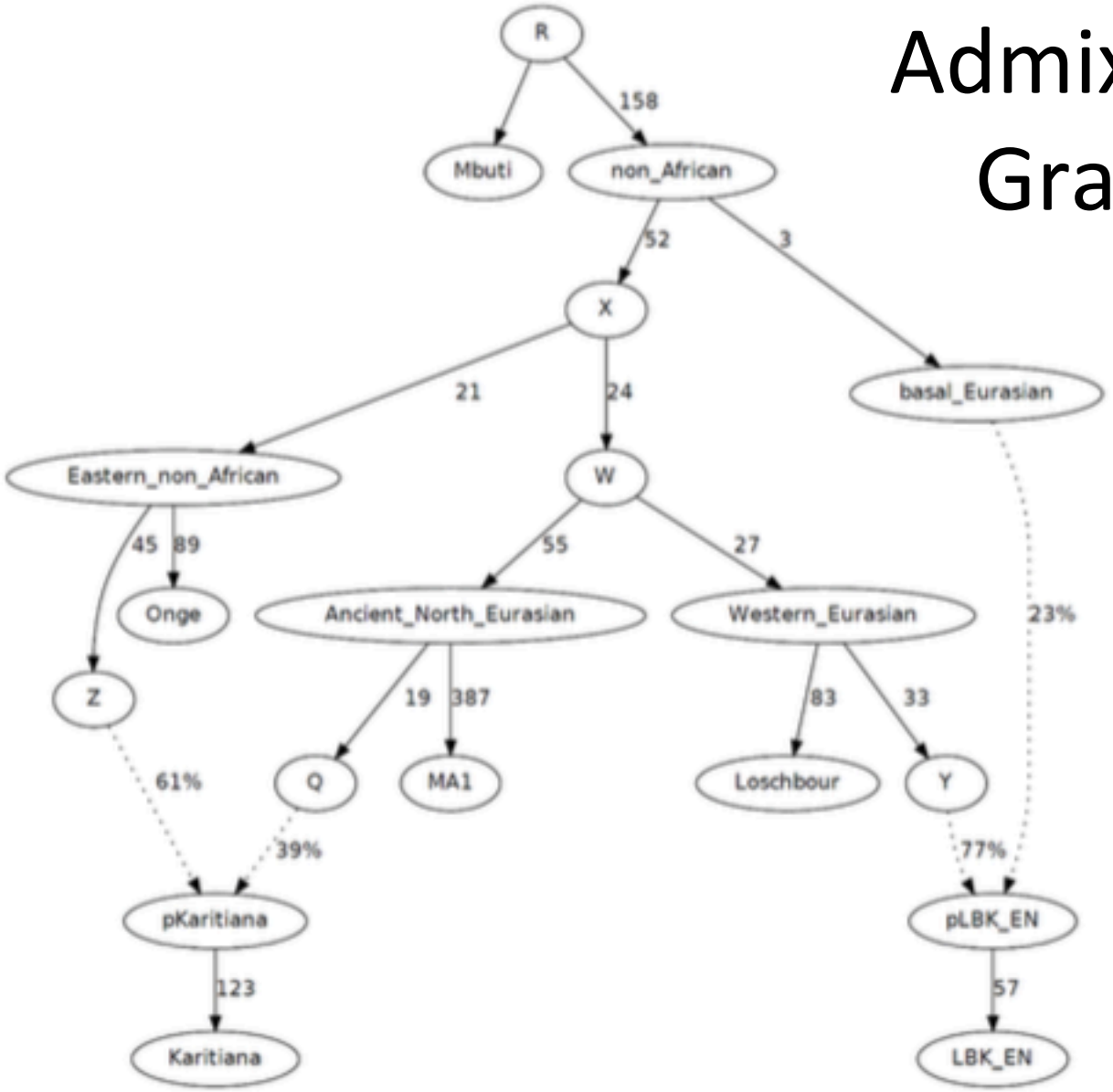
Lesson 12: Admixture Graph

Wednesday August 1, 2018

9:00 – 11:30 am

Figure S8.1: Basic model from ref.4 applied to data of this study.



Admixture Graph

Admixture Graph

- This should be a visually pleasing *summary* of results you have found already using other methods we've learned.

Interpretation and limitations of qpGraph

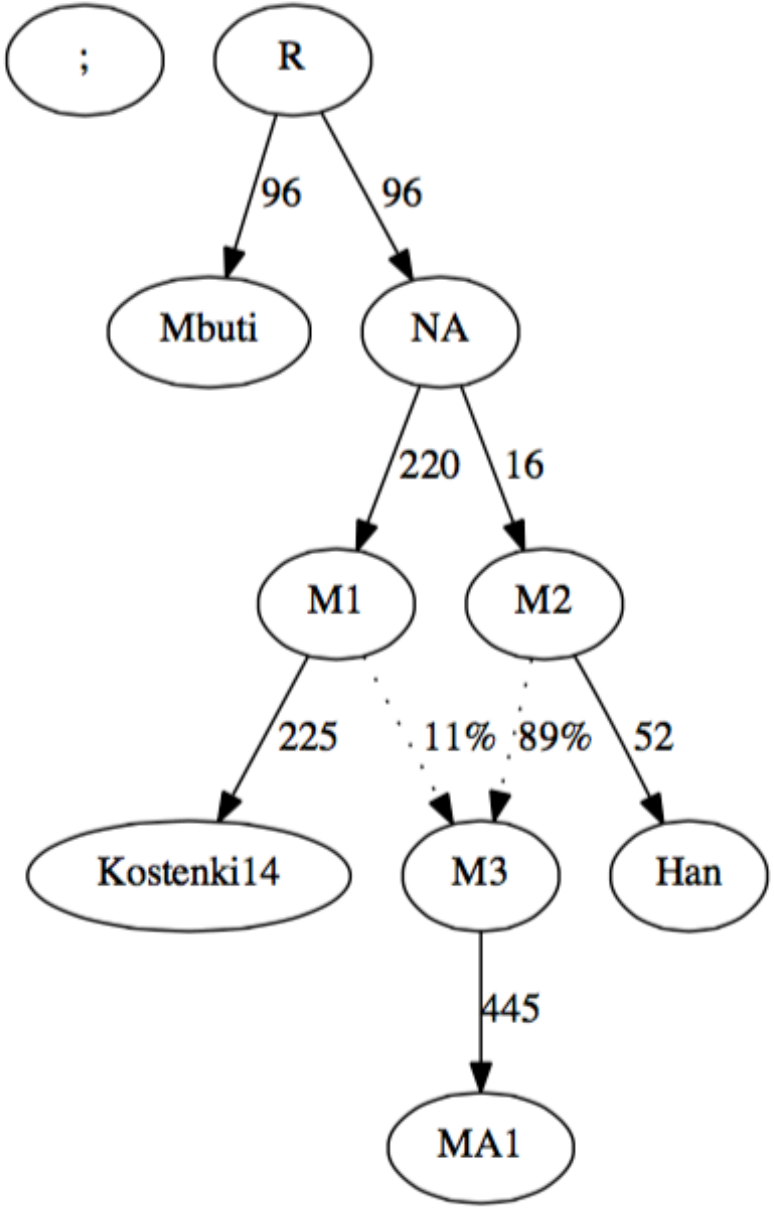
1. A major use of *qpGraph* is to show that a hypothesized phylogeny must be incorrect. This generalizes our *D*-statistic test, which is testing a simple tree on four populations.
2. After fitting parameters, study of which *f*-statistics fit poorly can lead to insights as to how the model must be wrong.
3. Overfitting can be a problem, especially if we hypothesize many admixing events, but only have data for a few populations.

Patterson et al. (2012)

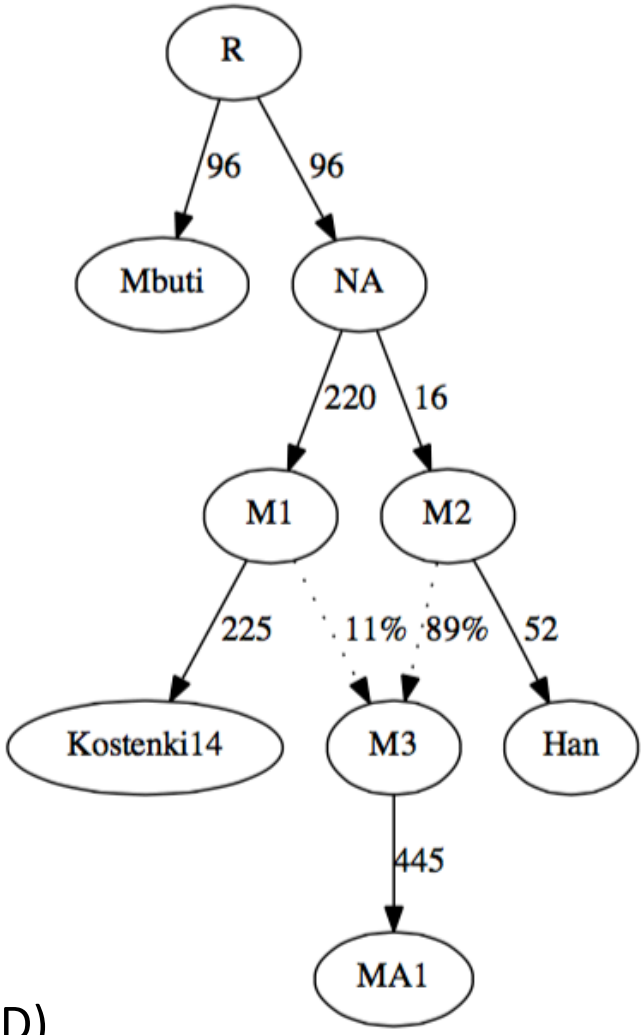
README for AdmixtureGraph

https://github.com/DReichLab/AdmixTools/blob/master/README.QPGRAPH

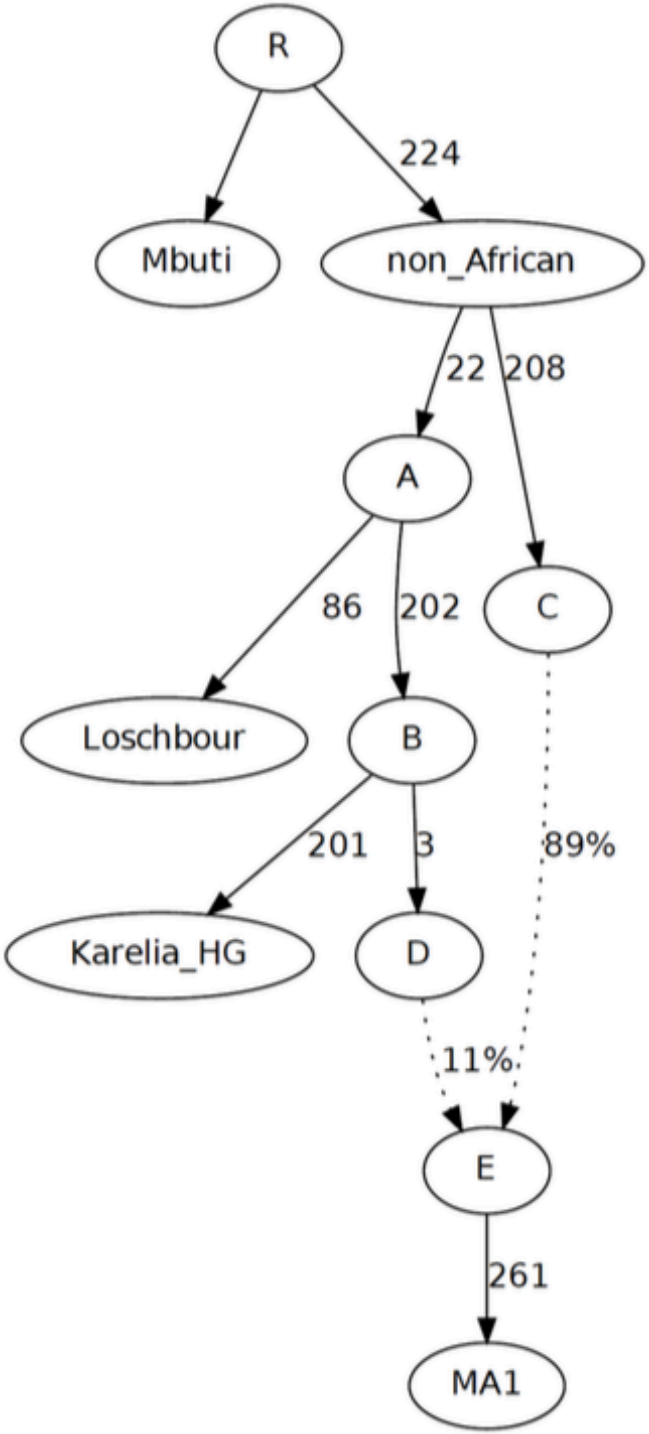
intro_model :: MA1 Han MA1 Han 0.499669 0.500030 0.000361 0.003593 0.101



```
root    R
label   Mbuti   Mbuti
label   MA1     MA1
label   Kostenki14      Kostenki14
label   Han     Han
edge    rm      R       Mbuti
edge    rn      R       NA
edge    ne      NA      M1
edge    ek      M1      Kostenki14
edge    nh      NA      M2
edge    ah      M2      Han
edge    mm      M3      MA1
admix   M3      M1      M2
```



Root – starting point of your tree (root R)
Label – Each of your tips of tree (label PopID PopID)
Edge – each branch of tree (edge edgename node1 node2/tiplabel)
Admix – two lines indicating admixture (admix Target S1 S2)



Activity

Can you write the graph in text form that the program qpGraph will interpret?

Try to run it using

qpGraph -p PARFILE -g GRAPHFILE -o GRAPHFILE.out -d GRAPHFILE.dot > GRAPHFILE.log

Then, to make a PS file:
dot -Tps GRAPHFILE.dot > GRAPHFILE.ps

Then, to make a PDF file:
ps2pdf GRAPHFILE.ps GRAPHFILE.pdf

Did you get the same graph?

How to use qpGraph

- Start with a base graph – dependent on which populations may be relevant, previously published AdmixtureGraphs, etc.
- Add one new population or individual at a time, trying it as an added branch to all possible branches (edge), or as a mixture of all possible pairs of two branches (admix).
- Look to see which trees have |maxf|<3.
- Use these trees to add further new populations or individuals.

Searching the Space

- Look at the PDF files beginning with "Ami.Tianyuan.A2_*pdf" on my github page, under "Lesson12/".
  - fALL = all pairs of admixtures
  - nALL = adding as single branch
  - Top right corner lists worst f-stat
  - Let's find the best tree(s)!

Searching the Space - Activity

- Now look at the PDFs beginning with “Malta1.Ami.Tianyuan*pdf”. What are the best trees here?

Problem of qpGraph

- Manually searching the tree space.
- Order of addition limits direction of gene flow.

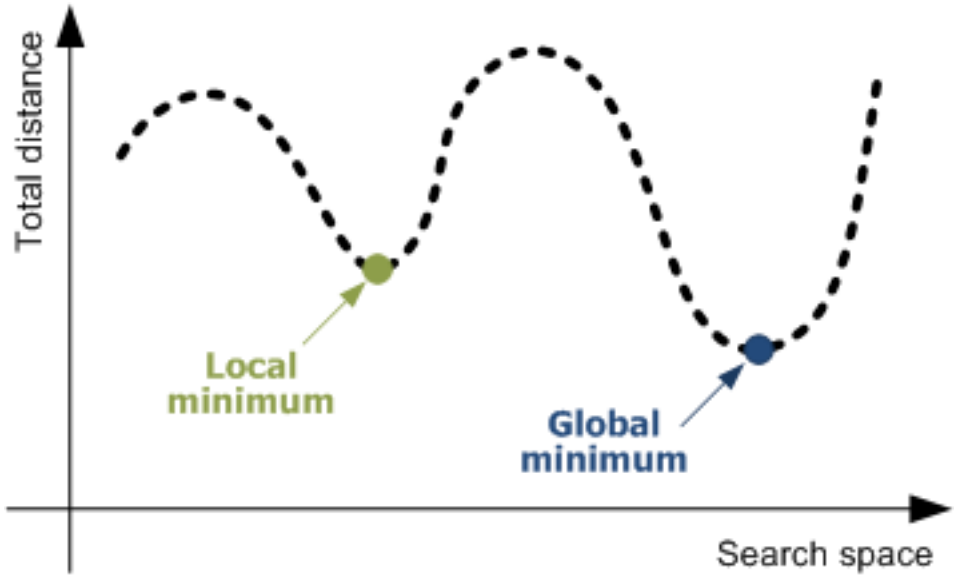
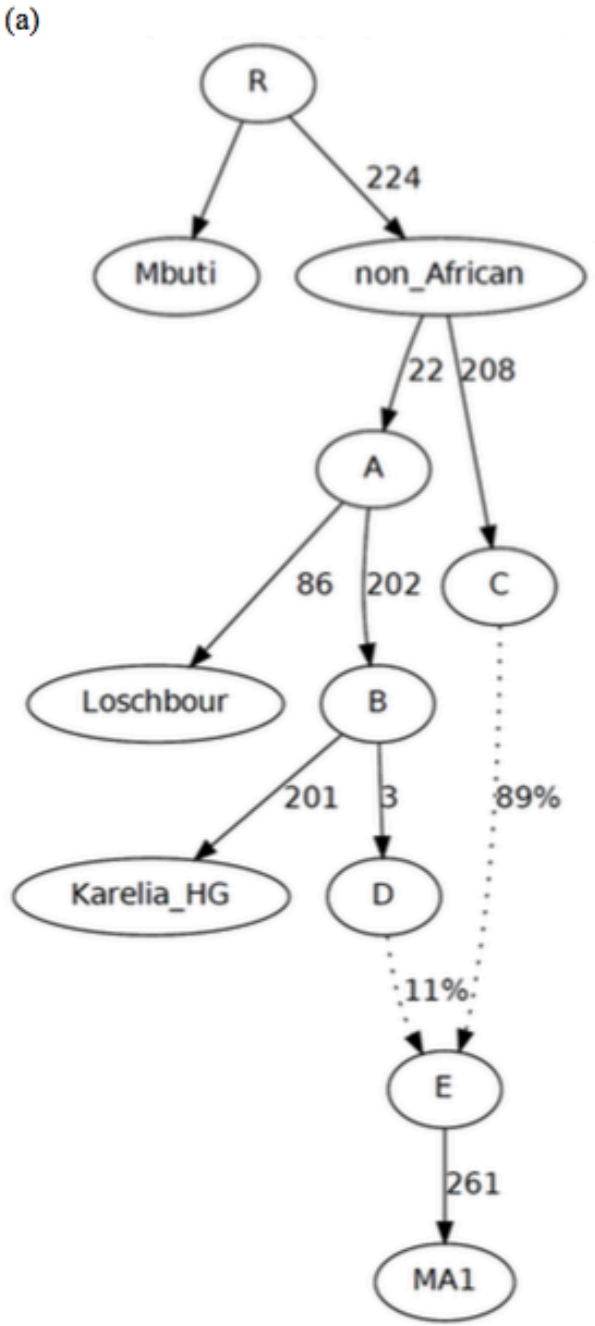
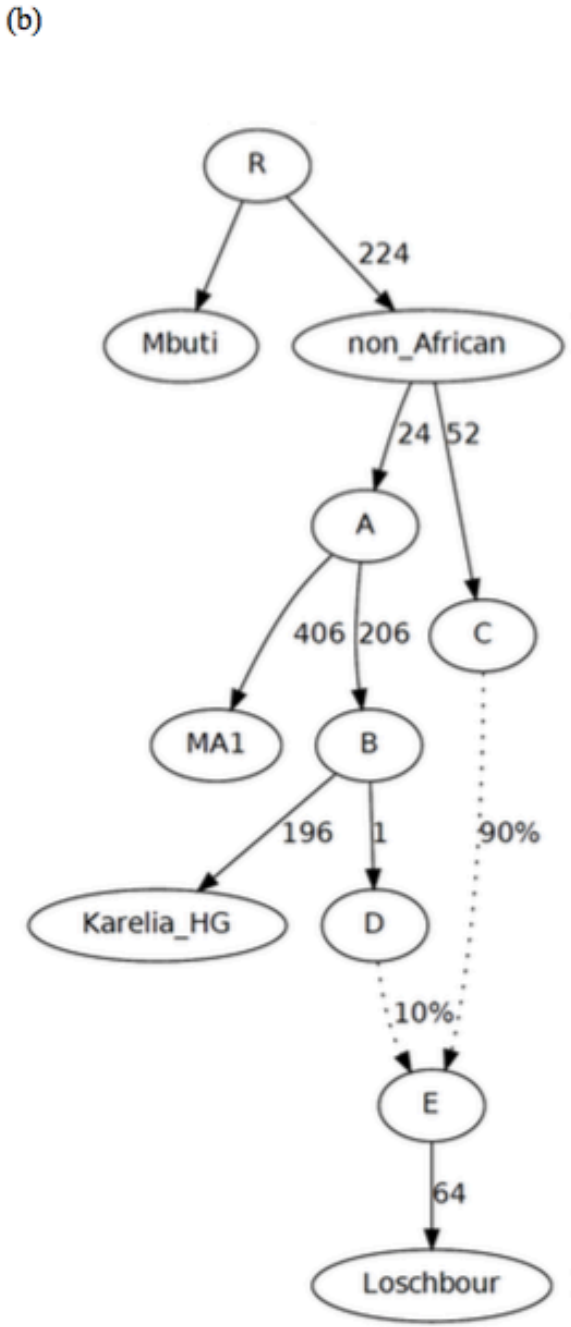


Figure S8.7: Alternative phylogenies that fit the data in which (a) MA1 is admixed, or (b) Loschbour is admixed.

Loschbour, then Mal'ta1





Mal'ta1, then Loschbour

Things to Consider:

- qpGraph is dependant on your hypotheses, so it **needs to be done at the end**.
- Depending on who you include and how many admixture events you add, you can almost always get a tree that works. Thus, the key is keeping it as **simple** as possible and **thoroughly searching within the space you JUSTIFY**.
- It is good at **eliminating specific models** you want to emphasize are NOT correct.
- It is good for summarizing major conclusions of your study, but an **important caveat is to remember there are likely other models not explored**.

Rest of Course

- Test on Friday
- Mini-projects on Monday (10 am? Afternoon?)
- mtDNA course Tuesday to Thursday with Albert

Exam

- Written test, one worksheet of notes.
- 75 minutes, 10 am.
- Predominantly Lessons 1 to 9.
  - Familiarity with major points and ancient individuals used in the Haak paper.
  - Understanding contamination estimates, sequencing, data processing that occurs in preparing a dataset.
  - Constructing the appropriate f3-, D-, and qpAdm analyses to answer your question.
  - Interpretation of f3-, D-, and qpAdm results
  - Developing hypotheses using PCA
  - Familiarity with uniparental markers and what you can learn from them.

Questions?

- Office Hours
  - Rest of class today
  - Thursday 2-3 pm