

# The space of ultrametric phylogenetic trees

Alex Gavryushkin  
(joint work with Alexei Drummond)

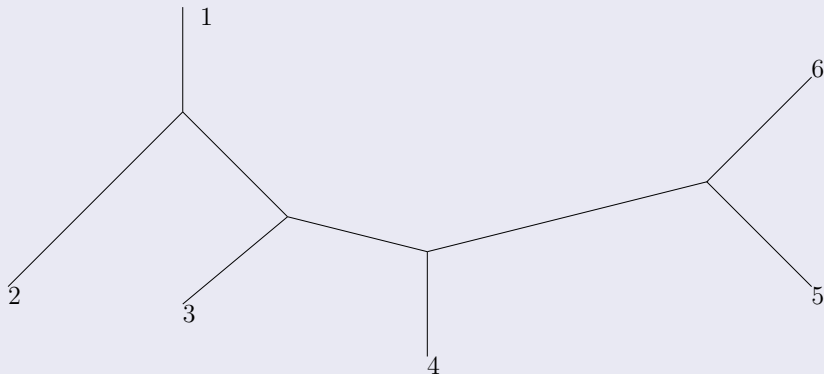


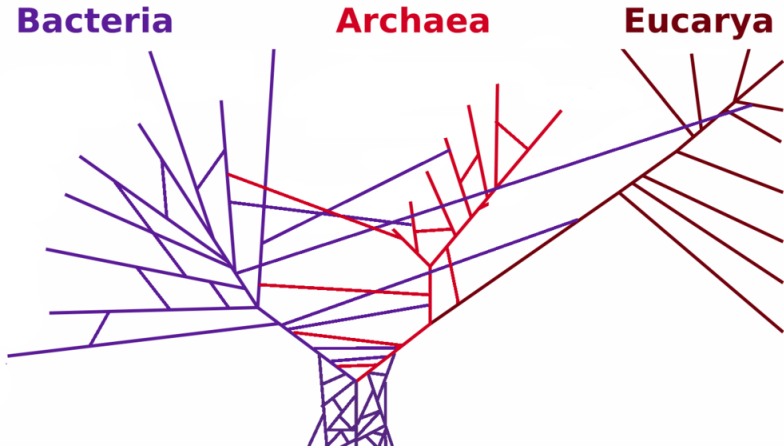
**THE UNIVERSITY OF AUCKLAND**  
**NEW ZEALAND**

18th February 2015

# Unrooted phylogenetic tree

## Definition

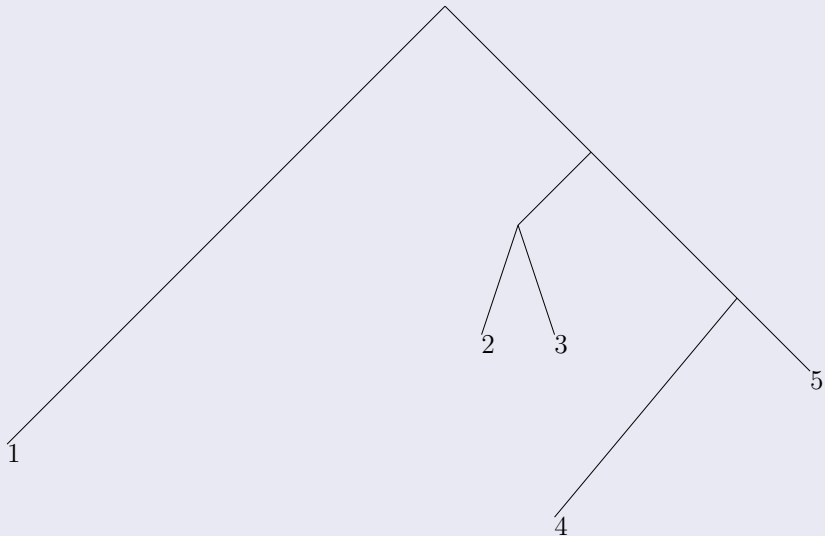




Credit link: [http://commons.wikimedia.org/wiki/File:PhylogeneticTree\\_horizontal\\_transfers.png?uselang=en-gb](http://commons.wikimedia.org/wiki/File:PhylogeneticTree_horizontal_transfers.png?uselang=en-gb)

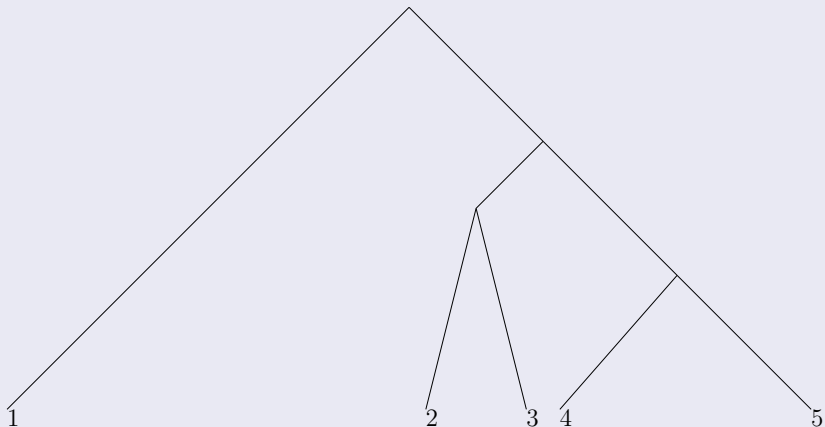
# Rooted phylogenetic tree

## Definition



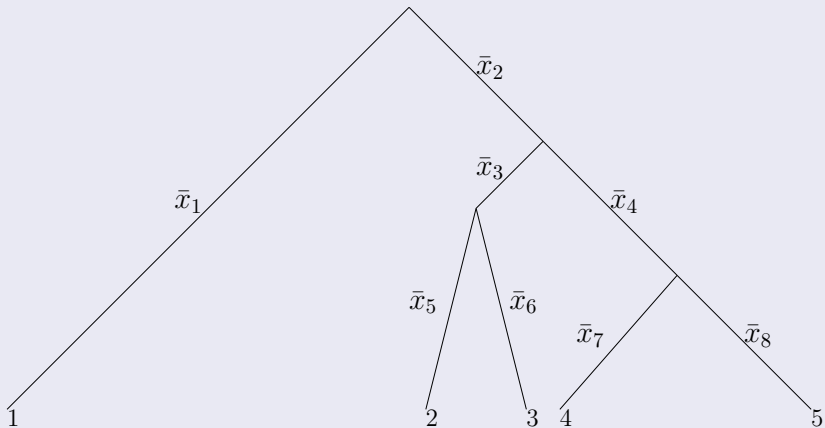
# Equidistant (ultrametric) phylogenetic tree

## Definition

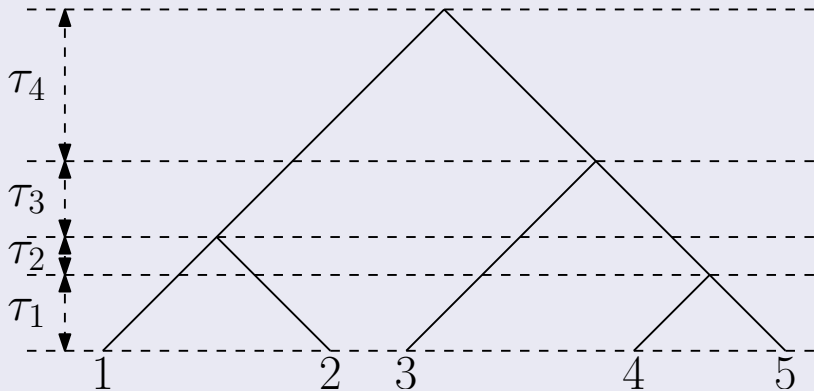


# Equidistant phylogenetic tree with parameters

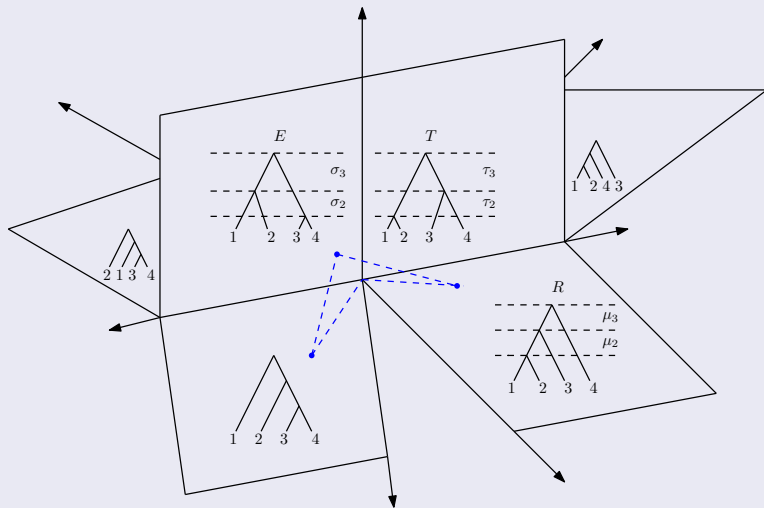
## Definition



## Definition

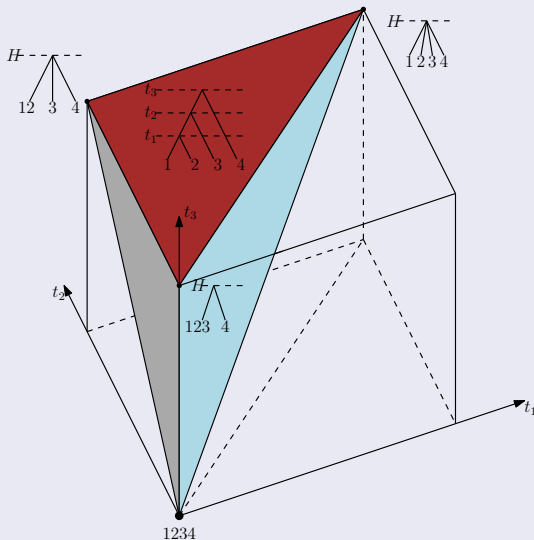


## Definition

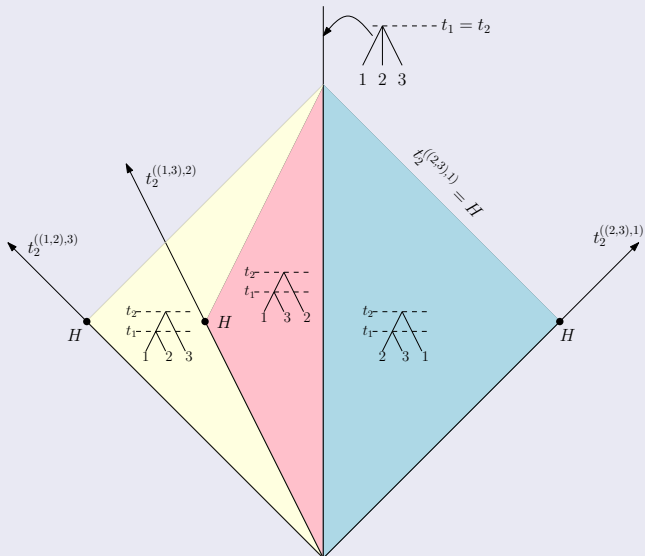




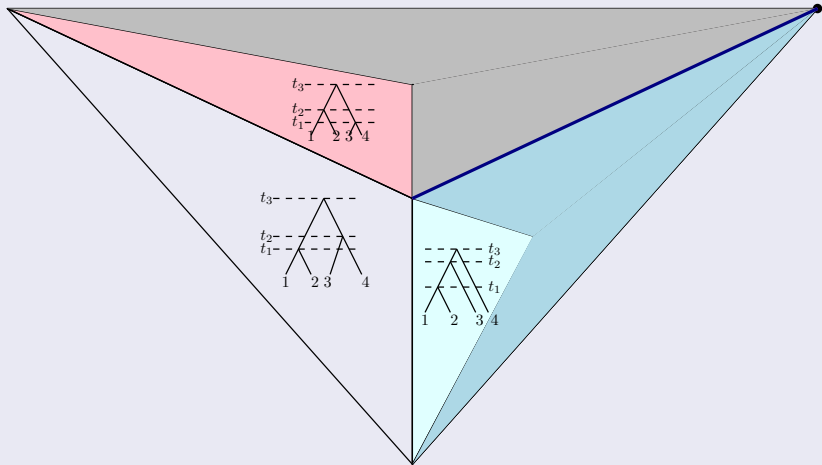
## Definition



## Definition



## Definition



- 1 Bayesian MCMC: Mixing rate, access time, efficient proposals.
- 2 Summarising posterior: No need to introduce several random variables on different probability spaces, no need to fit inconsistent data together.
- 3 Interesting algorithmic/data structures problems: How to solve NP-complete problems on real computers for real data (Chris and Erick can compute SPR-distance).
- 4 Interesting geometries: “Every new example of a non-trivial simplicial complex of non-positive curvature is a big deal.”

Geodesic is a short for shortest path.

### Theorem (G and Drummond [6])

*$\tau$ -space has unique geodesics.*

The reason this is true is pretty much the same as why this is true in BHVspace [2].

### Theorem (G and Drummond [6])

*Geodesics in  $\tau$ -space are efficiently computable.*

(Assuming  $\mathcal{O}(n^4)$  is efficient.)

The reason this is true is pretty much the same as why this is true in BHVspace [5].

## Definition

A metric space is called *nice* if most statisticians would like it.

Examples of nice metric spaces include real line, Euclidean space, and its nice subspaces.

Examples of not nice metric spaces include all non-measurable subsets of a Euclidean space, all nowhere dense subsets of a Euclidean space, and most importantly the spaces where it is hard to define a random variable.

## Theorem (Billera, Holmes, and Vogtmann [2])

*The space of phylogenetic trees is a nice space.*

## Theorem (G and Drummond [6])

*The space of equidistant phylogenetic trees is a nice space.*

Theorem (G and Drummond [6])

*t-space is not a very nice space.*

That is,

Theorem (G and Drummond [6])

*Geodesics in t-space are hard to compute. Possible but hard.*

Hard here means that we (Alexei and I) don't know how.

## Definition

A geodesic metric space is called *nice* if it is a convex path-connected subspace of a computable metric space with unique geodesics of the same dimension.

## Theorem (G and Drummond [6])

$\tau$ -space is an efficiently computable cubical complex with unique geodesics.

## Conjecture (G and Drummond [6])

$t$ -space is a simplicial complex with unique geodesics, which are NP-hard to compute.

## Corollary

Both  $\tau$ -space and  $t$ -space are nice.



# Thank you for your attention!



Philipp Benner, Miroslav Bačák, and Pierre-Yves Bourguignon.  
Point estimates in phylogenetic reconstructions.  
*Bioinformatics*, 30(17):534–540, 2014.



Louis J Billera, Susan P Holmes, and Karen Vogtmann.  
Geometry of the space of phylogenetic trees.  
*Advances in Applied Mathematics*, 27(4):733–767, 2001.



Joseph Heled and Remco R Bouckaert.  
Looking for trees in the forest: summary tree from posterior samples.  
*BMC evolutionary biology*, 13(10):221, 2013.



Megan Owen and J Scott Provan.  
A fast algorithm for computing geodesic distances in tree space.  
*IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 8(1):2–13, 2011.



Chris Whidden and Frederick A. Matsen IV.  
Quantifying MCMC exploration of phylogenetic tree space.  
*Systematic Biology*, doi:10.1093/sysbio/syv006, 2015.



Alex Gavryushkin and Alexei Drummond.  
The space of ultrametric phylogenetic trees.  
*arXiv preprint arXiv:1410.3544*, 2014.