Investigating Traditional Systems of Medicine Using Phylogenies Derived from Gene Sequences

Ciira Maina

Dedan Kimathi University of Technology

18th June 2015

Introduction

- Traditional systems of medicine have been a source of several live saving medicines for centuries
- Modern phamaceutical companies have also sought to exploit active compound in traditional medicines.
- A recent success story is the use *Artemisia annua* of to treat malaria

Introduction

- There are various approaches to discovering active drug compounds from traditional medicine
- Collectively we refer to this as **bioprospecting**
- The methods include:
  - Chemical testing for active compounds.
  - in vitro testing of plant extracts.
- Studies have been performed to determine the antiplasmodial action of several plant extracts to develop anti-malaria drugs
- Clarkson *et al.* tested 139 plant extracts for antiplasmodial action and found 23 highly active extracts[1].

---

[1] Clarkson, C. et al (2004). In vitro antiplasmodial activity of medicinal plants native to or naturalised in South Africa. Journal of ethnopharmacology, 92(2), 177-191.

Modern Approach to Traditional Medicine

- Studies suggest the medicinal properties are not randomly distributed among plants[2]
- Certain plants groups are more likely to have bioactive compounds than others
- This suggests that phylogenetics studies of plant species can help in drug discovery

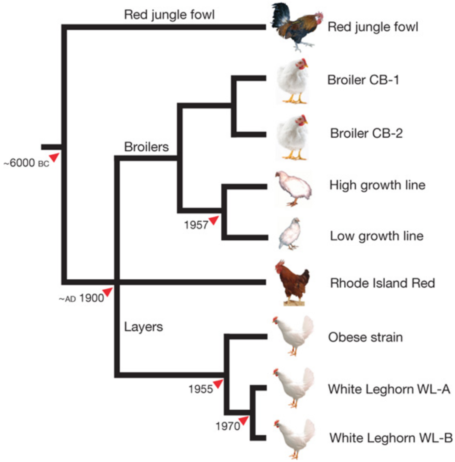
---

[2] Saslis-Lagoudakis CH, et al. (2012) Phylogenies reveal predictive power of traditional medicine in bioprospecting. Proceedings of the National Academy of Sciences 109: 15835-15840.

Modern Approach to Traditional Medicine

- It is costly to randomly test for bioactivity
- A more targeted approach is necessary to find candidates with a high potential of bioactivity
- Combining traditional knowledge of medicinal plants and phylogenetic analysis can uncover neglected species of high promise.

What is a phylogeny?

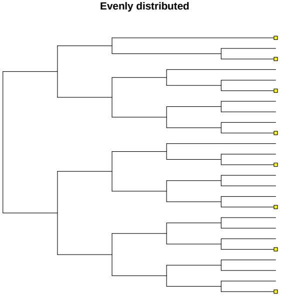


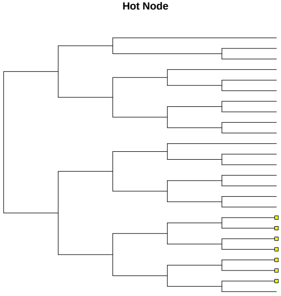
Source: `mikethechickenvet.wordpress.com`

The Idea

- Build phylogenies using DNA sequences
- Use knowledge of traditional use of the plants for medicinal purposes
- Discover any nodes in the phylogeny that are over represented for medicinal plants-hot nodes
- Discover potential bioscreening targets

The Idea



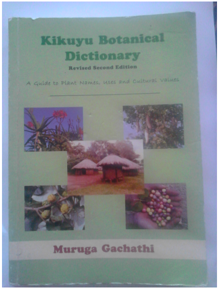


Methods

- We need information on the flora in a given region
- Lists of species used to treat various diseases in the region
- Access to gene sequence data of gene markers that can resolve plant phylogeny
- Methods to build phylogenies from sequence data
- Methods to investigate community structure of the phylogenies.

Study of TM in Central Kenya

- Pioneering work by Muruga Gachathi on a Kikuyu botanical dictionary provides a valuable starting point[3]
- This botanical dictionary contains information on over 400 plants.
- Those with medicinal uses have also been classified in this text with details about the plant part used.



---

[3] Gachathi M (2007) *Kikuyu botanical dictionary: a guide to plant names, uses and cultural values.* Tropical Botany.

Gene sequence data: Biology 101

- Cells of living organisms contain genomes which have genes
- For normal function, a cell must produce certain enzymes and proteins
- Genes code for these proteins
- A gene consists of a sequence of Bases:
    - A,T,G,C

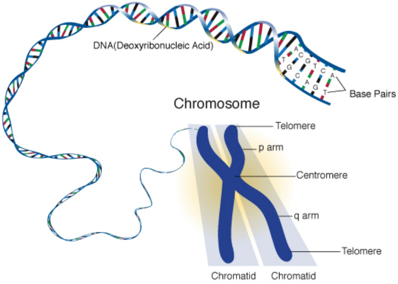


Image courtesy of the National Human Genome Research Institute's

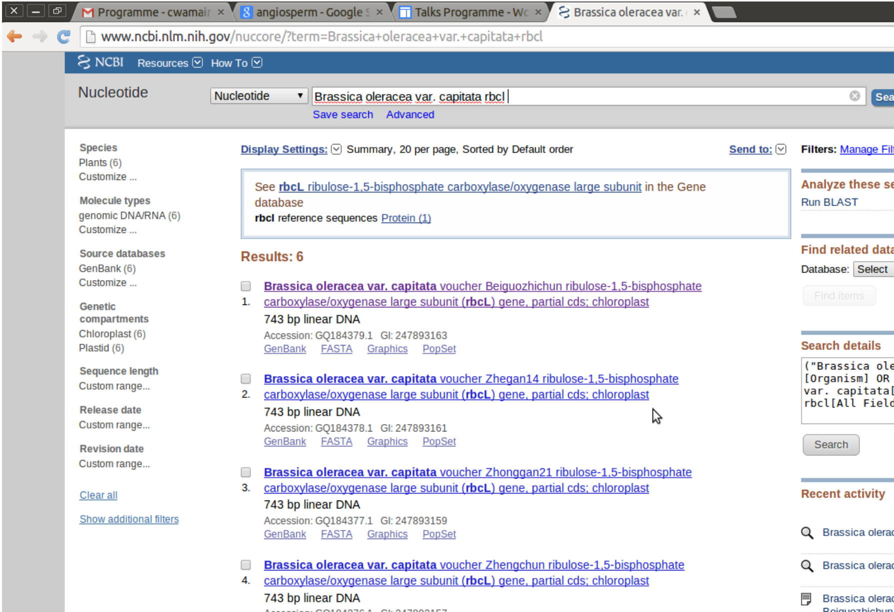
Gene sequence data: Gene Markers

- Gene markers are used to create phylogenies
- They are sequences of DNA which contain enough variability but are present in most organisims
- For most plants the genes *rbcL* and *matK* are good gene markers
- To determine the gene sequence for a particular gene for a given organism, the genome is sequenced from DNA material extracted from the organism

Gene sequence data: Gene Markers

- Consider the cabbage (Mboga), *Brassica oleracea* var. *capitata*
- The sequence for the *rbcL* is available from the National Center for Biotechnology Information (NCBI) database

Gene sequence data: Gene Markers



www.ncbi.nlm.nih.gov/nuccore/?term=Brassica+oleracea+var.+capitata+rbcl

Nucleotide

Nucleotide: Brassica oleracea var. capitata rbcl

Save search Advanced

Species
Plants (6)
Customize ...

Molecule types
genomic DNA/RNA (6)
Customize ...

Source databases
GenBank (6)
Customize ...

Genetic compartments
Chloroplast (6)
Plastid (6)

Sequence length
Custom range...

Release date
Custom range...

Revision date
Custom range...

Clear all

Show additional filters

Display Settings: Summary, 20 per page, Sorted by Default order

Send to:

See **rbcL** ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit in the Gene database
rbcl reference sequences Protein (1)

Results: 6

☐ 1. **Brassica oleracea var. capitata** voucher Beiguozhichun ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (**rbcL**) gene, partial cds; chloroplast
743 bp linear DNA
Accession: GQ184379.1 GI: 247893163
GenBank FASTA Graphics PopSet

☐ 2. **Brassica oleracea var. capitata** voucher Zhegan14 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (**rbcL**) gene, partial cds; chloroplast
743 bp linear DNA
Accession: GQ184378.1 GI: 247893161
GenBank FASTA Graphics PopSet

☐ 3. **Brassica oleracea var. capitata** voucher Zhonggan21 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (**rbcL**) gene, partial cds; chloroplast
743 bp linear DNA
Accession: GQ184377.1 GI: 247893159
GenBank FASTA Graphics PopSet

☐ 4. **Brassica oleracea var. capitata** voucher Zhengchun ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (**rbcL**) gene, partial cds; chloroplast
743 bp linear DNA

Filters: Manage Fil

Analyze these se
Run BLAST

Find related data
Database: Select
Find items

Search details

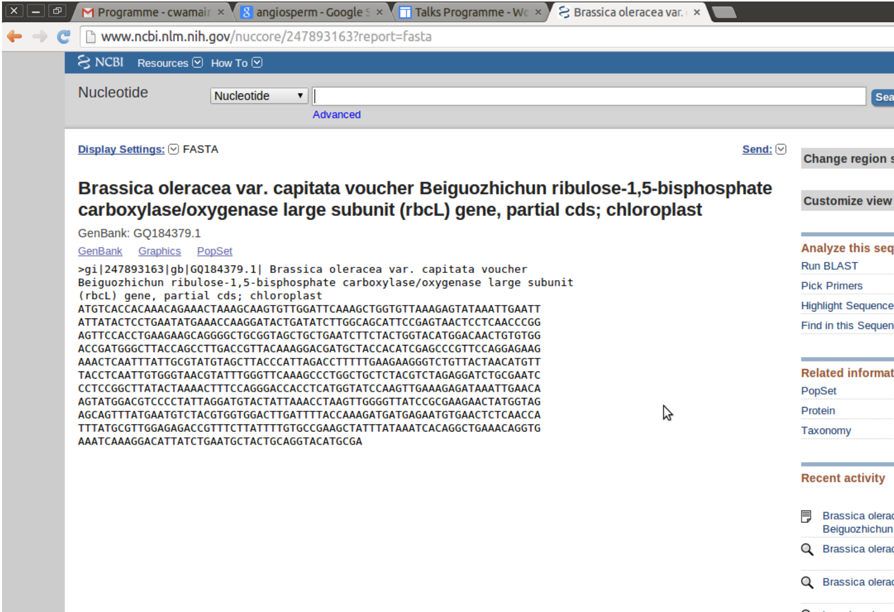
```
("Brassica ole
[Organism] OR
var. capitata[
rbcl[All Field
```

Search

Recent activity

Brassica olerac
Brassica olerac
Brassica olerac
Beiguozhichun

Gene sequence data: Gene Markers



Brassica oleracea var. capitata voucher Beiguozhichun ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene, partial cds; chloroplast

GenBank: GQ184379.1

GenBank Graphics PopSet

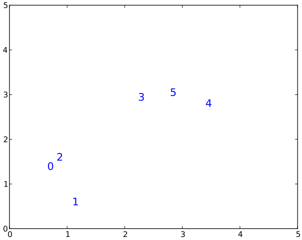
```
>gi|247893163|gb|GQ184379.1| Brassica oleracea var. capitata voucher
Beiguozhichun ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit
(rbcL) gene, partial cds; chloroplast
ATGTCACCACAAACAGAAACTAAAGCAAGTGTTGGATTCAAAGCTGGTGTTAAAGAGTATAAATTGAATT
ATTATACTCCTGAATATGAAACCAAGGATACTGATATCTTGGCAGCATTCCGAGTAACTCCTCAACCCGG
AGTTCCACCTGAAGAAGCAGGGGCTGCGGTAGCTGCTGAATCTTCTACTGGTACATGGACAACTGTGTGG
ACCGATGGGCTTACCAGCCTTGACCGTTACAAAGGACGATGCTACCACATCGAGCCCGTTCCAGGAGAAG
AAACTCAATTTATTGCGTATGTAGCTTACCCATTAGACCTTTTTGAAGAAGGGTCTGTTACTAACATGTT
TACCTCAATTGTGGGTAACGTATTTGGGTTCAAAGCCCTGGCTGCTCTACGTCTAGAGGATCTGCGAATC
CCTCCGGCTTATACTAAAACTTTCCAGGGACCACCTCATGGTATCCAAGTTGAAAGAGATAAATTGAACA
AGTATGGACGTCCCCTATTAGGATGTACTATTAAACCTAAGTTGGGGTTATCCGCGAAGAACTATGGTAG
AGCAGTTTATGAATGTCTACGTGGTGGACTTGATTTTACCAAAGATGATGAGAATGTGAACTCTCAACCA
TTTATGCGTTGGAGAGACCGTTTCTTATTTTGTGCCGAAGCTATTTATAAATCACAGGCTGAAACAGGTG
AAATCAAAGGACATTATCTGAATGCTACTGCAGGTACATGCGA
```

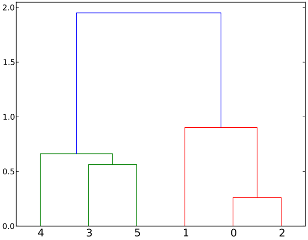
Building Phylogenetic Trees from Gene sequence data

- Given gene sequences of gene markers from organisms of interest we can construct phylogenetic trees
- We download these sequences for the NCBI database `http://www.ncbi.nlm.nih.gov/genbank/`
- Scripts can do this automatically
- The first step is to align the sequences using multiple sequence alignment software
- In this work we use the ClustalW2 program from the European Bioinformatics Institute

Building Phylogenetic Trees from Gene sequence data

- With aligned sequences we can then build phylogenies using agglomerative clustering techniques





Building Phylogenetic Trees from Gene sequence data

- Or we can use probabilistic methods to infer the phylogeny form the gene sequences
- We can use maximum likelihood methods RAxML as well as other open source programs such as GARLI, PHYML
- Bayesian approaches MrBayes

Investigating community structure

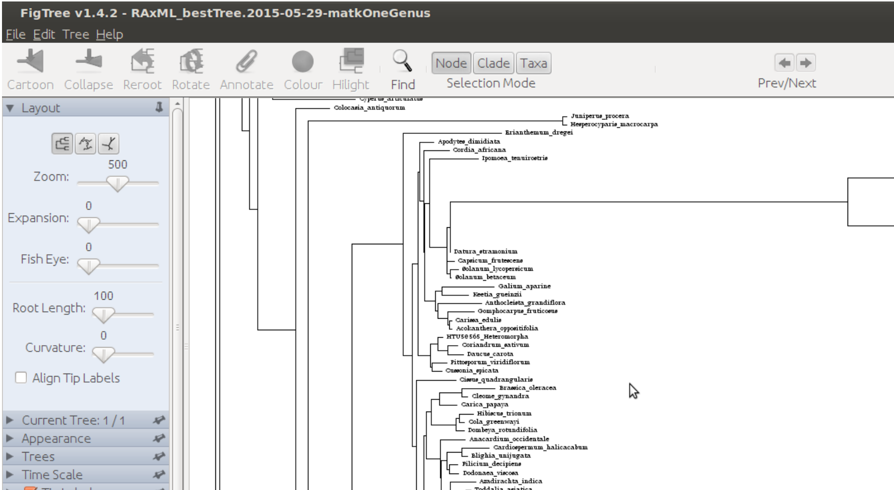
- Once we have a phylogenetic hypothesis we can explore the community structure[4]
- If we have a group of plants that treat a particular disease we can explore their distribution in the phylogenetic tree
- Compute statistics such as Mean Phylogenetic Distance
- See whether the MND is significantly smaller than would be expected at random
- Use this information to discover hot nodes and potential candidates, we use picante an R implementation of Phylocom

---

[4] Webb CO, et al. (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics 24: 2098-2100

Results

- We use the *matK* gene
- Phylogenetic tree of approximately 130 taxa obtained using RAxML



Zoom: 500

Expansion: 0

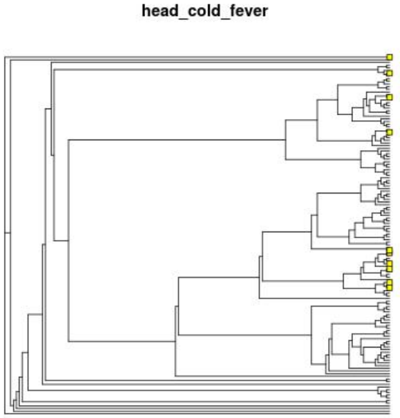
Fish Eye: 0

Root Length: 100

Curvature: 0

☐ Align Tip Labels

Community structure for Headache, cold and fever



- Observed MPD 0.97, Average MPD 2.43, p-value 0.062

Community structure for Cough, chest and pneumonia

- Observed MPD 0.96, Average MPD 2.43, p-value 0.039

Conclusions and future directions

- We have explored using phylogenetic knowledge to investigate Traditional medicine
- Preliminary results are encouraging
- Explore Bayesian approaches to phylogenetic tree construction.
- Add more taxa and explore other gene markers
- Collaboration with chemists, botanists and herbalists

Thank You