

# The GALAH survey: Stellar chemical tagging using phylogenetic trees

Klemen Čotar\*, Tomaž Zwitter\* and the Galah team

\*Faculty of mathematics and physics, University of Ljubljana, Slovenia



Univerza v Ljubljani  
Fakulteta za matematiko in fiziko

**Objective:** The GALactic Archaeology with HERMES (GALAH) spectroscopic survey aims to provide stellar spectra for 1 million stars in four different wavelength intervals. They are selected in a way that it should be possible to obtain more than 20 different chemical abundances per star. Chemical composition of the stars can be used to find (tag) members of long dissipated stellar clusters (De Silva et al. 2015, MNRAS, 449, 2604).

Our method uses abundances of 13 elements (determined by Cannon (Ness et al. 2015, ApJ, 808, 1) method) to assess the similarity in chemical composition of stars. The similarity information is then used to construct a phylogenetic tree that gives us a visual representation of similarity in composition between individual stars. The tree is then traversed in a direction from tips to the root of the tree to further analyze kinematic relations between stars with similar chemical composition.

## Datasets and filtering

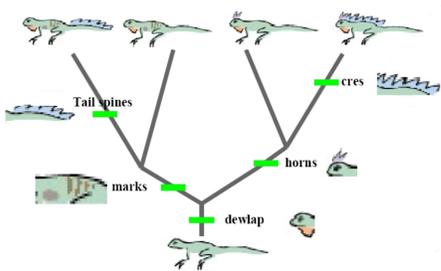
For this analysis we used GALAH radial velocities and abundances determined by the Cannon (pipeline version 1.2) approach for more than 200k stars. The positions and proper motions are taken from the Gaia-TGAS objects, which greatly reduces the final number of useful objects as GALAH usually observes dimmer objects ( $12 < G < 14$ ) that were not observed by the Hipparcos satellite.

In order to select a reliable dataset, we applied multiple filters that gave us the best possible set of objects to work with. We selected objects with successfully reduced spectra and reliable Cannon analysis. To further increase the reliability, we removed the objects that are known to produce anomalous abundance values (binary objects, low SNR observations etc.). In some analysis parameters reduced from the repeated observations of the same objects were merged together.

After applying the data cuts and merging repeated observations we are left with 8300 unique objects.

## Phylogenetic representation

A phylogenetic or evolutionary tree is a diagram that shows evolutionary relations among our data, where the position on the tree is based upon similarities and differences in the parameters that represent our observations. The trees were invented to show the evolution of the DNA sequence, where similarity is defined as number of matching nucleotides (Lemey P., 2009, The phylogenetic handbook).



**Image:** A simple cartoon representation of a phylogenetic tree that tries to show possible evolutionary changes and paths of 4 different species of lizards (Baum et al., 2008, Nature Education, 1, 191).

## Similarity between chemical composition of stars

In our case distances between stellar abundances were determined in a mathematical way. We tested different algorithms to compute distance between objects:

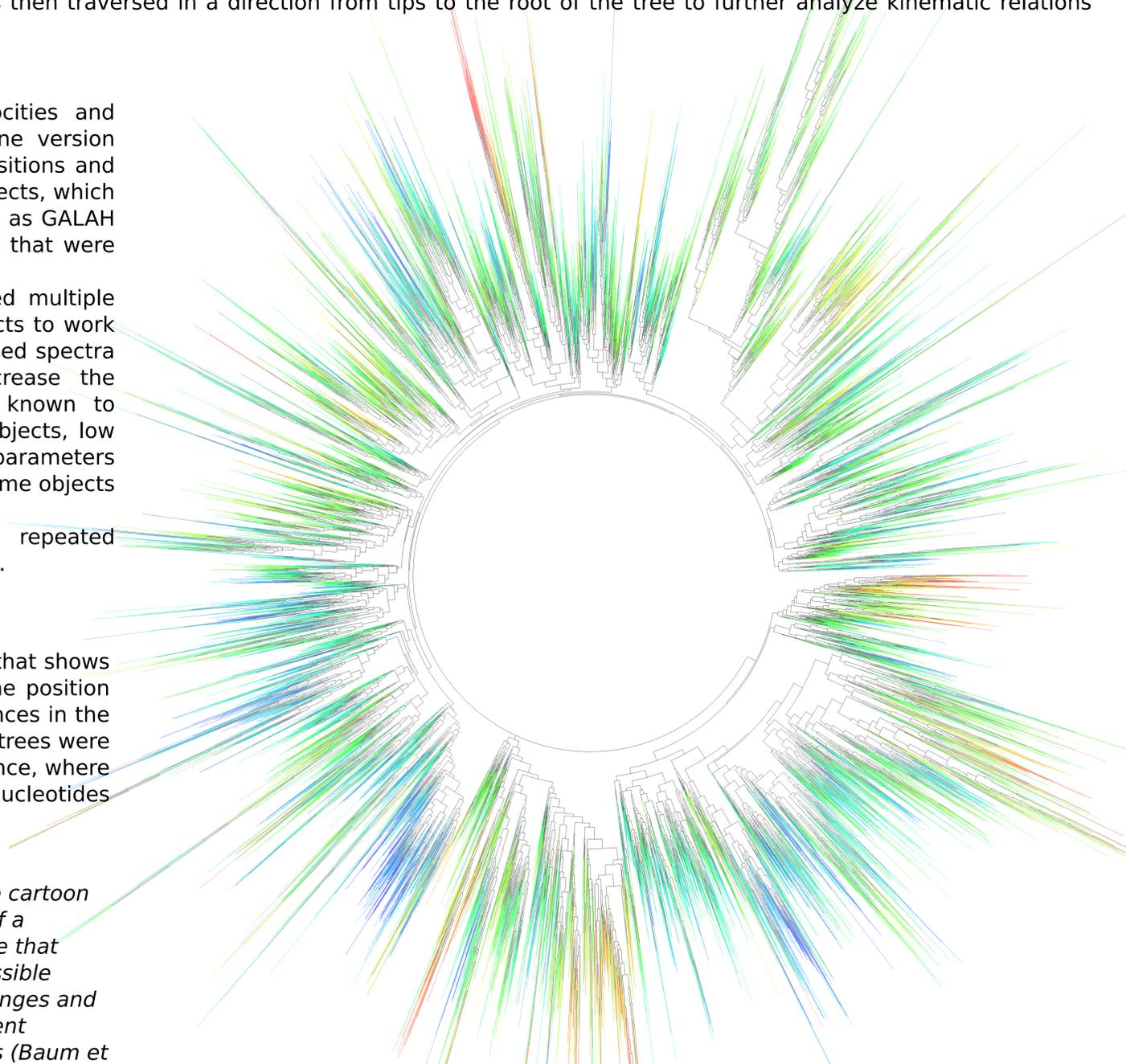
- euclidean (best for clustering repeats)
- manhattan
- sorensen
- canberra (had the smallest number of nodes between repeated observations).

The only information that we used to determine the distance between stars were abundance values itself. They were standardized before the computation in order to remove the differences in mean and variance value between individual abundances. No additional information about physical properties or kinematic information was used.

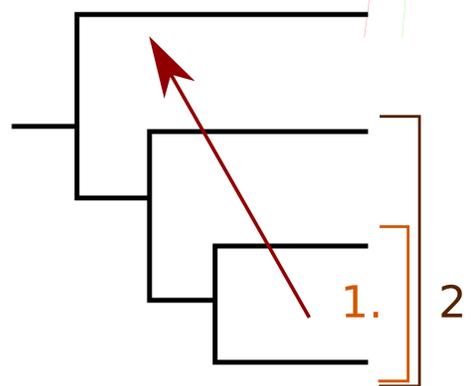
## Determining properties of the tree

Based on how the trees are constructed, we know that objects with identical or very similar chemical composition should be located together at the top of the tree (Jofrè et al. 2017, MNRAS, 467, 1140). Our selected set of observations has nearly 500 repeated objects and some members of known stellar clusters. With them we can estimate:

- how many of the repeated observations are located together (14% in our case)
- what is the expected average number of nodes between objects to be still considered to have similar composition (20 nodes in our case)
- do cluster members have chemical composition similar enough to get clustered in one part of the tree



**Image up:** One of the constructed trees where tips of the tree are colored by the metallicity of the star that ended up at that location. It can be observed that the algorithm combines objects with similar metallicity (without any prior knowledge about the exact value of metallicity parameter that is plotted in colors).



**Image left:** Schematic representation of the tree traversing steps. The red arrow shows the direction of traversing (leaves to root) and numbers indicate which stars are evaluated at different steps. Stellar objects are located at the tips of the branches shown in black.

## Traversing the tree structure

The tree structure may be regarded as any other dimensional reduction and/or clustering technique with the difference that here we are given an exact path or relation between any two objects in our dataset.

The path we are following is represented and explained by above image. At every step we assume that the objects have very similar chemical composition, but can have different position and velocity through the Galaxy. To calculate full 6D position and velocity information, Gaia observation are used. This it is used to select object with similar kinematics and possibly find stellar clusters.

## Future work

Some of the topics I would like to investigate using those stellar phylogenetic tree structures are:

- definition of chemically homogeneous clusters or streams of stars
- blind search for long dissipated stellar clusters
- possible hints to the chemical enrichment history and evolution