



HaploPhyle a Graphical Haplotype Network in the Light of External Data

Claire Billot¹, Manuel Ruiz¹, Xavier Perrier¹, Jean-François Rami¹, Gautier Sarah¹, Cédric Hassen¹, Reinhard Simon²

1 – CIRAD, avenue Agropolis, 34398 Montpellier Cedex 5, FRANCE
2 – CIP, Lima, PEROU



Genetic diversity assessment through genotyping is gaining much resolution and currently produces now large quantity of DNA sequences or chains of tightly linked markers. The close linkage between the polymorphisms observed is likely to leave the patterns little affected by recombination. Therefore patterns evolve less quickly (than with unlinked or loosely linked makers) and better reflect past situations.

In the particular context of crops, the past comprises a major event, domestication, which generally involved strong genetic bottlenecks. Pattern analysis of polymorphisms can lead to identification of predominant haplotypes, and inference of ancestral haplotypes vs recombinant haplotypes. Haplotype analysis is likely to enable breakthroughs in crop germplasm analysis.

This project provides the community with new software for analysis of genotyping data, HaploPhyle (pipeline version 1.0, haplophyle.cirad.fr, code source freely available). This Web based pipeline includes haplotype definition, haplotype network analysis and connexion with external data, such as geographic origin, evolutionary history or genetic group assessment. Visualization at once of the haplotypes ordered according to a tree can be performed with a R function.

Welcome on HaploPhyle

General | Haplotyping | Network Construction | External Data Correlation | Haplotype Representation

General Settings

Allelic Data File: Parcourir...

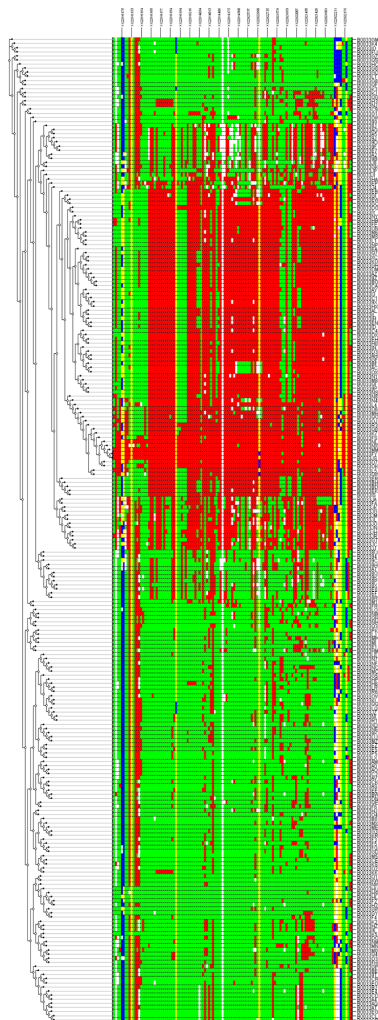
External Data File: Parcourir...

Starting Point: Haplotyping

Submit | Reset

	Germplasm	Haplo1	Haplo2
19	AAACAA	GAAAA	GAAAA
17	AAAGCG	GAGGCG	GAGGCG
18	TGAGCA	AAGGCG	AAGGCG
15	TGAAAC	GAGGCG	GAGGCG
16	AAAGCG	TACAGG	TACAGG
13	TGAGCA	AAGGCG	AAGGCG
11	AAAGCG	TACAGG	TACAGG
12	TGAAAC	GAGGCG	GAGGCG
21	TAAAGC	GAGGCG	GAGGCG
3	TGAGCA	AAGGCG	AAGGCG

Allelic Data



Haplotyping

Haplotyping and haplotype visualisation

Gevalt was used to build the haplotypes. Visualisation at once of the differences between haplotypes can be drawn, by choosing some accessions as references. A R function draws the haplotypes according to a tree.

Welcome on HaploPhyle

General | Haplotyping | Network Construction | External Data Correlation | Haplotype Representation

Haplotyping Options

Haplotyping Program: Gevalt

alignCheck

minMAF: 1E-4

maxMissed: 1

lincov: 0.0

getMSF: 300

getMSL: 10

Submit | Reset

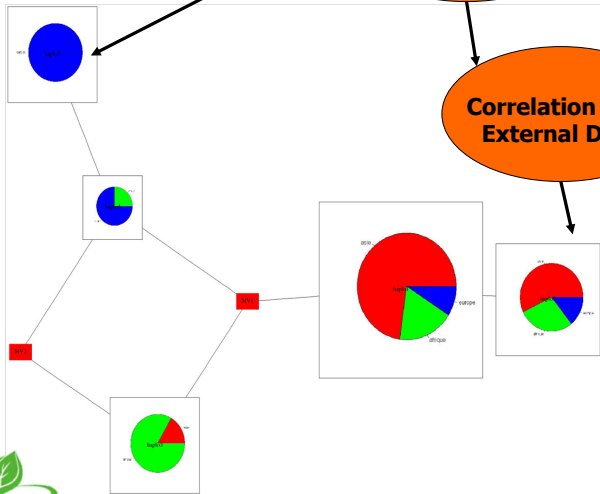
Name	Sequence	Number of haplotypes
haplo17	TGAGCAAGGCGGAGGCGGAAAAAGGAA	1
haplo7	AAAGCGTACAGAGGAGGAGGAGGAAAGGAA	1
haplo1	AAACAAAGAAAAAGAAAAAGAAAAAGAAAA	1
haplo18	TGAGCAAGGCGGAGGCGGAGGAGGAAAAAGAA	1
haplo19	TGAGCAAGGCGGAGGCGGAGGAGGAAAAAGAA	1
haplo10	TGAGCAAGGCGGAGGCGGAGGAGGAAAAAGAA	1
haplo16	TGAGCAAGGCGGAGGCGGAGGAGGAAAAAGAA	1

Clustering

Network building

We created a Jar file implementing for now two algorithms: Minimum Spanning Network and Median Joining Network. In the future, more algorithms will be added enabling more choices to the users. Nodes sizes are proportional to the numbers of germplasm sharing an haplotype and edges length are proportional to the difference between haplotypes.

Network Building



Correlation with External Data

External Data

Haplotype representation according to a tree
Users can set one color for each referent haplotypes, color for common alleles and color for new alleles. Missing data always appear in white. This representation allows users to visualize at once gene integration between two populations.

Correlation with external data

Choose the characteristic for the correlation: Origin

- #F00000 Africa
- #0000FF Europe
- #0000FF Asia
- #0000FF America

Correlation with external data

Nodes of the network are represented by pie-charts taking into consideration the repartition of external information (set by the user) among germplasm owning this haplotype