

## Phylogenomics: pitfalls and benefits

Venue: University of Neuchâtel, Faculty of Sciences, Emile-Argand 11, UniMail

**March 25<sup>th</sup> Room E213**

**10 am: welcome UniMAIL – room E213**

**10.15 am – Launching lecture by Christian Parisod: Genome evolution: challenges of repeats, recombination and hybridization for phylogenetics using high throughput sequencing**

Retrotransposition, recombination and hybridisation/polyploidy lead to remarkably dynamic genomes. These processes – at the heart of evolution – also complicate inference to be drawn from genome-wide datasets. This will be illustrated by our work on Buckler Mustards, showing an unanticipated history of recurrent whole genome duplication (thus lots of paralogs) coupled with the proliferation of repetitive sequences (thus challenges for assembling genomes).

Bardil et al. 2015. Evolutionary dynamics of retrotransposons following autopolyploidy in the Buckler Mustard species complex. *Plant Journal*

Geiser et al. Manuscript. Recurrent whole genome duplications and retention of clusters of adaptive genes in the Buckler Mustard.

Followed by discussion

**11.00 am – Lecture by Pascal-Antoine Christin: Evolutionary enablers, convergence, and the recurrent assembly of C4 photosynthesis**

The adaptation to changing environments often involves the evolution of traits of impressive complexity, through novel combinations of multiple anatomical and/or biochemical characters. C4 photosynthesis is one of such traits. It increases the productivity of plants growing in warm and arid conditions via a novel biochemical pathway coupled with leaf modifications. The evolutionary origins of independent biochemical and anatomical components can be inferred using phylogenetic tools. Such investigations show that the accumulation of C4-like components over millions of years facilitated the emergence of a C4 physiology in specific groups of plants. The realization of the C4 phenotype was however achieved differently every time, generating the diversity that can be observed nowadays across C4-dominated biomes in tropics and subtropics.

Christin et al. 2012. Adaptive evolution of C4 photosynthesis through recurrent lateral gene transfer. *Current Biology* 22: 445-449.

Christin et al. 2015. Genetic enablers underlying the clustered evolutionary origins of C4 photosynthesis in Angiosperms. *Molecular Biology and Evolution*  
doi:10.1093/molbev/msu410

Followed by discussion

**1.45 pm – Lecture by Christophe Guyeux: Investigating genes effects on phylogeny (topology and bootstraps)**

The amount of completely sequenced chloroplast genomes increases rapidly, leading to the possibility to build large scale phylogenetic trees of plant species. Considering a subset of close plant species, the phylogenetic tree that can be inferred by their core genes is not necessarily well supported, due to the possible occurrence of homoplastic genes that blur phylogenetic signals. To obtain a trustworthy tree still remain reachable if the number of homoplastic genes is low, the problem being to determine the largest subset of core genes that produces the best supported tree. During the session, we will investigate practical aspects of phylogenetic reconstruction, from the annotation stage to the determination of gene effects on both topology and supports.

Côté et al. Manuscript. Systematic investigations of gene effects on both topologies and supports: the *Echinococcus* case.

Followed by discussion

### **3.00 pm – Lecture by Thomas Marcussen: Inference of dated species networks from gene trees**

Most of the phylogenetic methods in use make the implicit assumption that the species phylogeny is a tree. However, many species phylogenies are better described as networks because they contain past events of interlineage gene flow, typically hybridisation followed by homoploid or polyploid hybrid speciation, or introgression. Here I will illustrate problems associated with the reconstruction of phylogenetic networks, and possible solutions, using two angiosperm datasets on *Viola* (violets) and *Aegilops/Triticum* (wheat) as examples.

Marcussen et al. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345: 1250092.

Marcussen et al. 2015. From gene trees to a dated allopolyploid network: insights from the angiosperm genus *Viola* (Violaceae). *Systematic Biology* 64: 84–101.

Followed by discussion

### **4.30 pm: General instructions for practical**

**March 26<sup>th</sup> Room E003**

#### **9 am – Practical (leader: Pascal-Antoine Christin): Inferring evolutionary forces and reconciling gene and species trees**

One gene tree will be analysed and the inferred topology will be compared to the species tree to detect non-vertical inheritance, through topology tests and dating analyses. In addition, the gene tree will be searched for evidence of positive selection using codon models. If time allows, we can then discuss ways of expanding such analyses to bigger (e.g. genomes or transcriptomes) analyses.

#### **1 pm – Practical (leader: Christophe Gueux): Consistent genome annotation and inference of core vs- pan genomes**

Analysis of complete chloroplast sequences, from dogma annotations to trees obtained using subsets of core genes from Apiales. The aim is to present the whole processing: genome annotation, core extraction, core genes alignment, phylogenetic reconstruction based on subsets of core genes, and statistical investigations of gene effects on topologies. If we will have time, we will present some optimization techniques to find the best subset of core genes.

**March 27<sup>th</sup> Room E003**

#### **9 am – Practical (leader: Thomas Marcussen): Inference of hybridization events from genome-wide datasets**

A data set of 275 gene phylogenies will be used to infer past hybridisation events, both homoploid and polyploid, among diploid wheat genome lineages, using both topology-based and branch length-aware methods. The number of methods and tests are currently limited and if time permits, we will discuss further developments in the field.

#### **Synthesis and general discussion involving all lecturers**

How to benefit from and avoid pitfalls with “-omics” datasets, dating and inference of interspecies gene flow?