



**HAL**  
open science

## Genome skimming at the scale of the whole Alpine ecosystem : insights into genome size, gene family amplifications and functional signatures

France Denoeud, Adriana Alberti, Sébastien Lavergne, Patrick Wincker

### ► To cite this version:

France Denoeud, Adriana Alberti, Sébastien Lavergne, Patrick Wincker. Genome skimming at the scale of the whole Alpine ecosystem : insights into genome size, gene family amplifications and functional signatures. Conference on Plant Genome Evolution, Sep 2019, Sitges, Spain. hal-04440668

**HAL Id: hal-04440668**

**<https://hal.science/hal-04440668>**

Submitted on 8 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Genome skimming at the scale of the whole Alpine ecosystem : insights into genome size, gene family amplifications and functional signatures

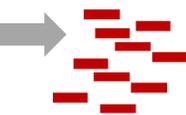
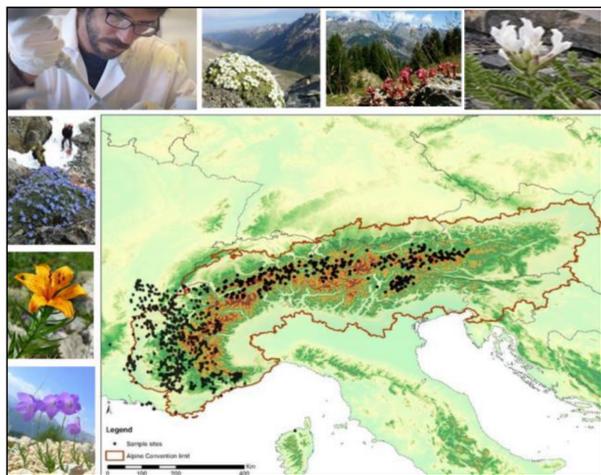
France Denoeud<sup>1</sup>, Adriana Alberti<sup>1</sup>, Sébastien Lavergne<sup>2</sup>, Patrick Wincker<sup>1</sup>

<sup>1</sup> CEA Institut de biologie François-Jacob, Genoscope, Université Paris-Saclay, Evry, France  
<sup>2</sup> Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université J. Fourier, Grenoble, France

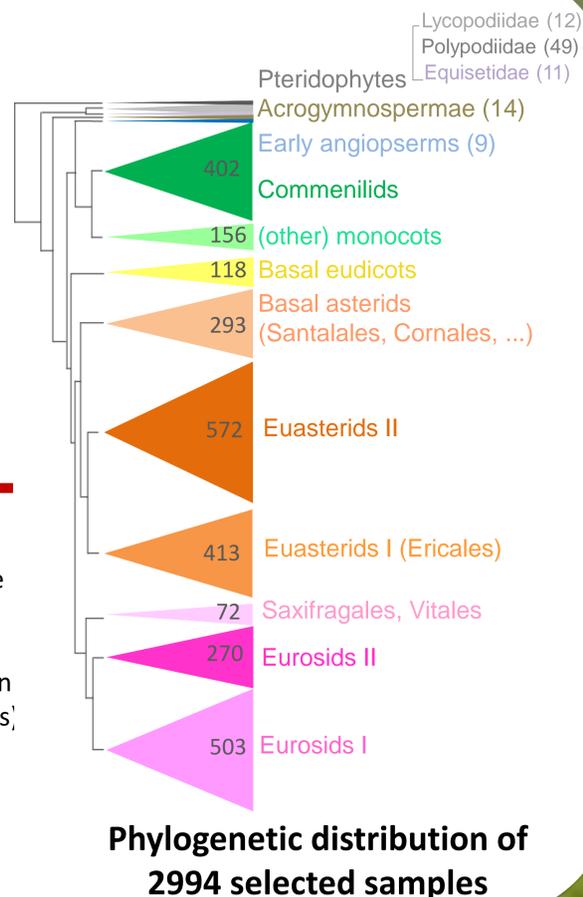
## Introduction

An intensive inventory and sampling of nearly all vascular plant species occurring in the Alps, a biodiversity hotspot harbouring 20% of the European flora was built in the course of the PhyloAlps project (<http://phyloalps.org/>). From a low coverage shotgun sequencing of genomic DNA (genome skimming), the aim was to assemble for each taxon the whole-chloroplast genome and nuclear ribosomal sequences, mainly for phylogenetic and evolutionary purposes. The study presented here aims at taking advantage of this high quality dataset by focusing on nuclear reads.

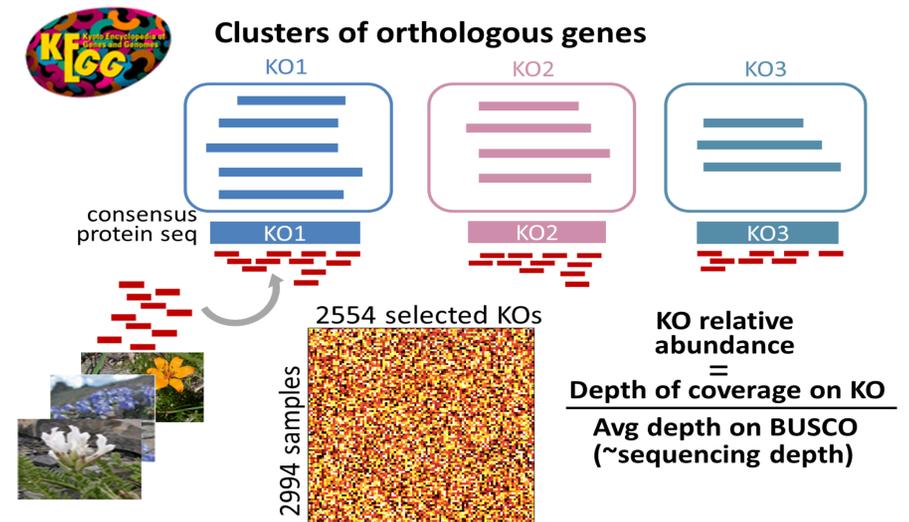
## Sampling ~ 4000 species of vascular plants in the Alps



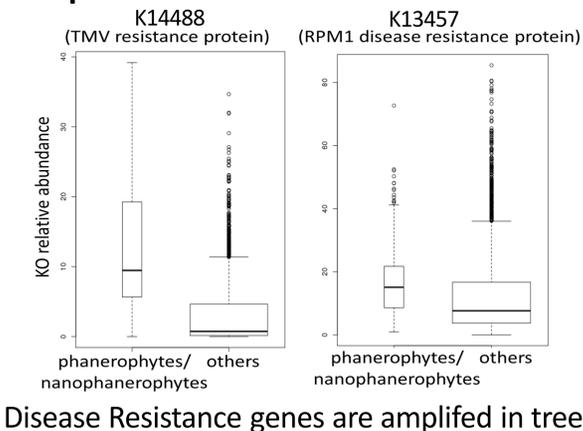
Low coverage sequencing  
~1 illumina paired-end run  
(5 million reads, per sample)



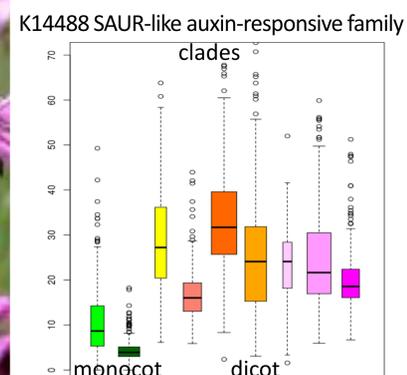
## KO abundance quantification



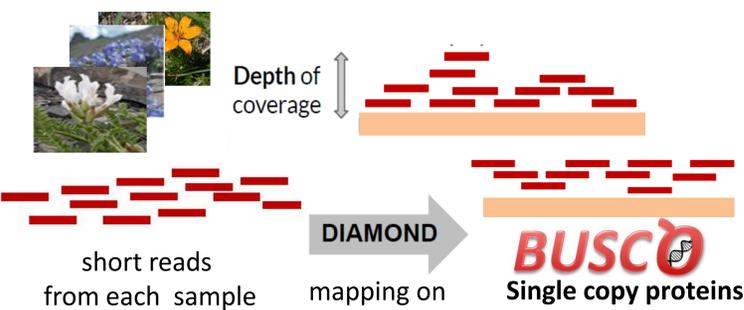
## Identification of gene family amplifications associated to life traits



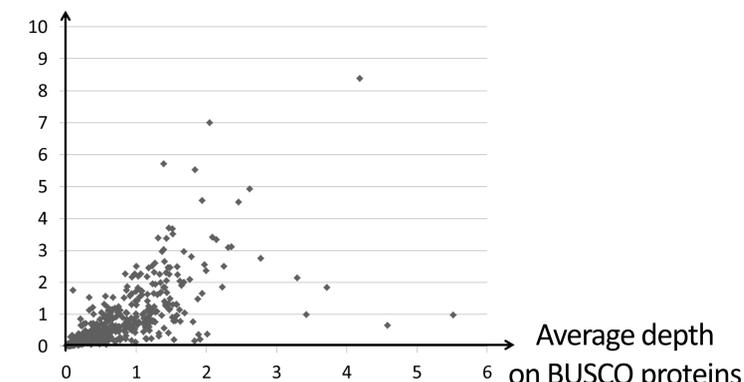
## Identification of functional signatures for phylogenetic clades



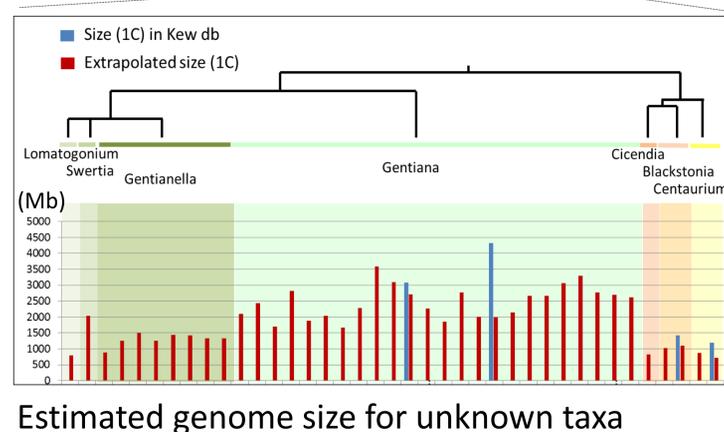
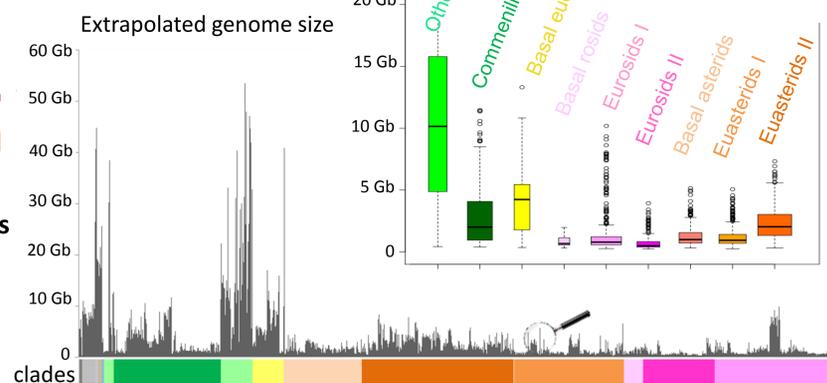
## Genome size prediction



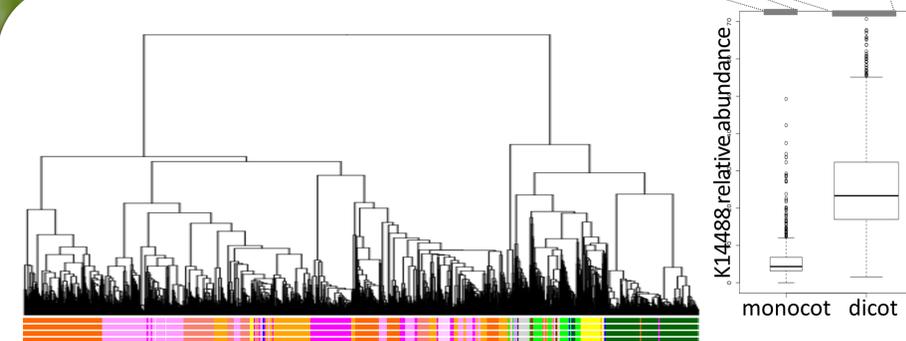
Sequencing depth from Royal Botanic Gardens Kew Plant DNA C-values Database



Use of average depth on BUSCO proteins to extrapolate genome size for all samples



Estimated genome size for unknown taxa



Clustering of samples based on relative abundance of KOs

## Conclusion

- We extrapolated genome sizes by mapping the reads on the set of single copy genes in BUSCO plants. These genome sizes will be compared with species life traits and KO abundance profiles. Preliminary results show a strong correlation of genome size with the number of copies of histone genes.
- We showed that the clustering of the samples based on KO abundance is well related to the phylogeny. We identified functional signatures (i.e. specific gene family amplifications) for phylogenetic clades.
- We identified gene functions that are amplified in species sharing similar life traits (for instance R genes in trees). We are investigating gene functions that are correlated to life traits such as light in order to identify genome amplification events that may have allowed some plant lineages to adapt and colonize the stressful environments of high altitude.