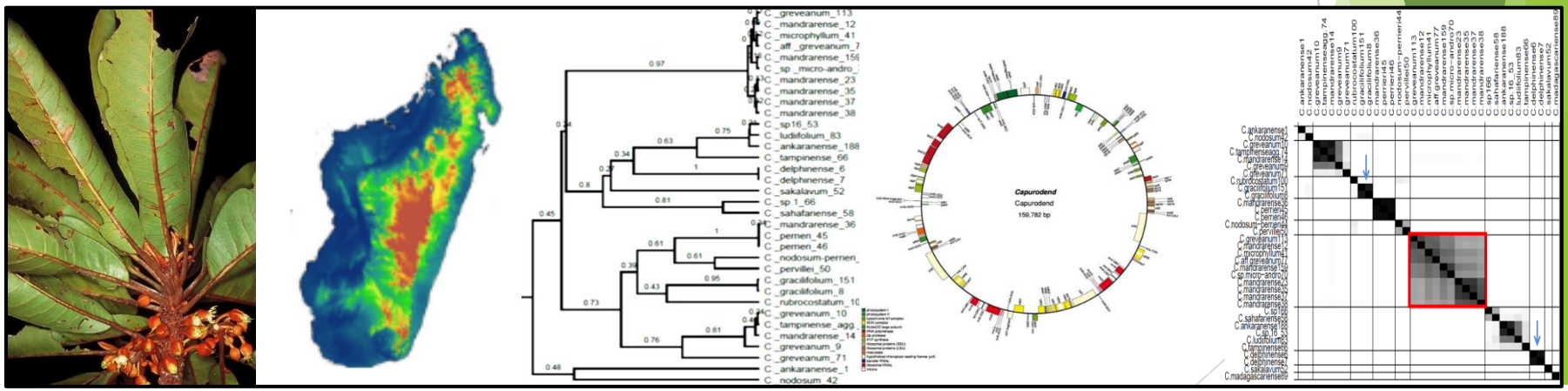
Phylogeny and species delimitation of the Madagascar endemic trees *Capurodendron* using a Gene Capture approach

Carlos G. Boluda

Study supervised by
Yamama Naciri & Laurent Gautier



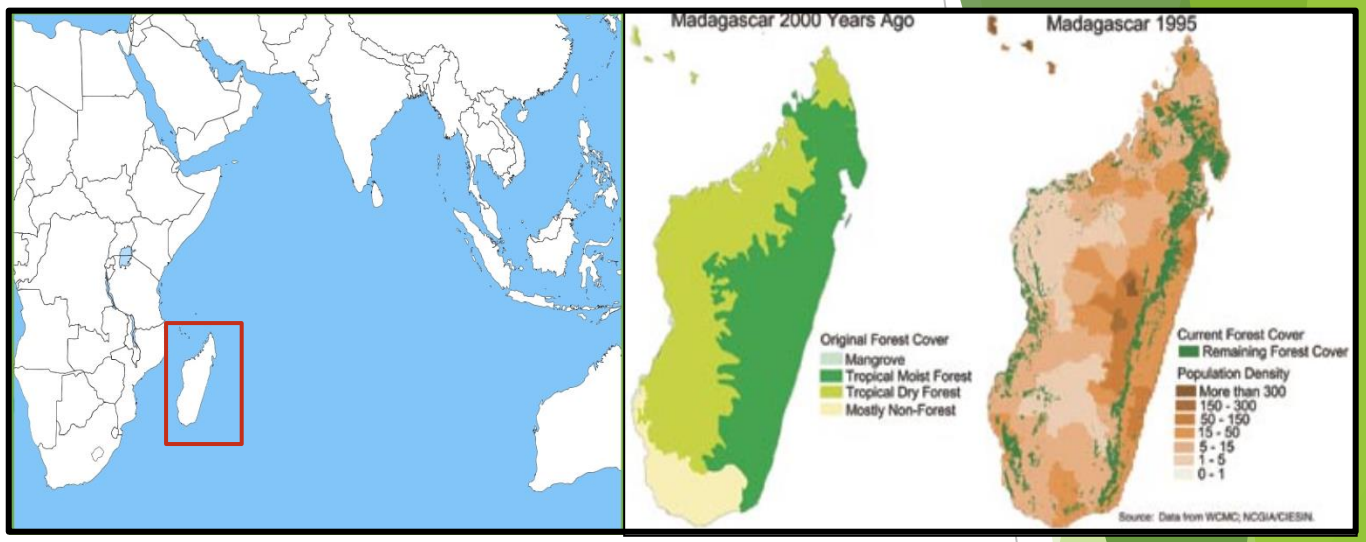
UNIVERSITÉ DE GENÈVE
Fondation Ernst et Lucie Schmidheiny
Franklinia Foundation
FNSNF Swiss National Science Foundation
Conservatoire et Jardin botaniques Genève
UNIVERSITÉ DE GENÈVE
14-October-2019

Introduction

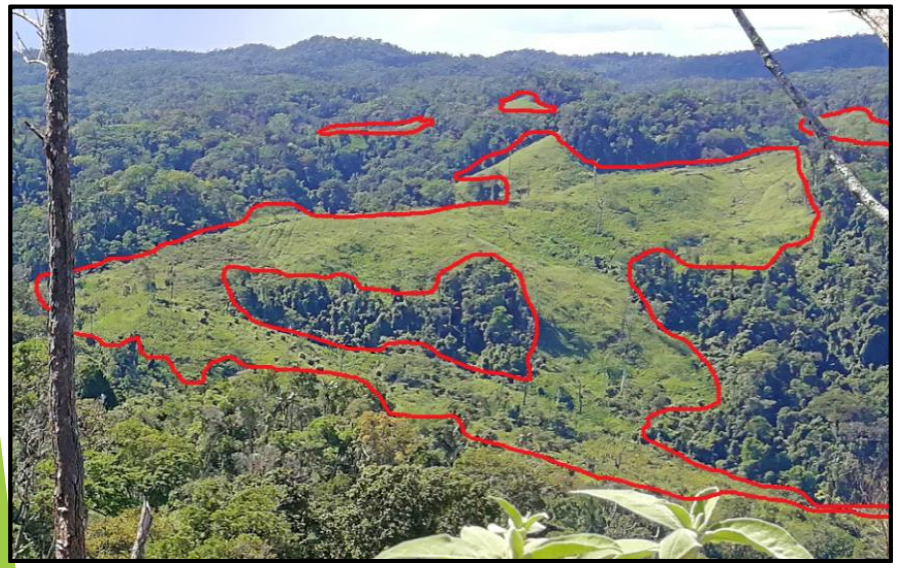
Madagascar:

-Isolated from India ~ 88 mya. (Upper Cretacic).

- 82% of endemic vascular flora.



Deforestation mainly due to:

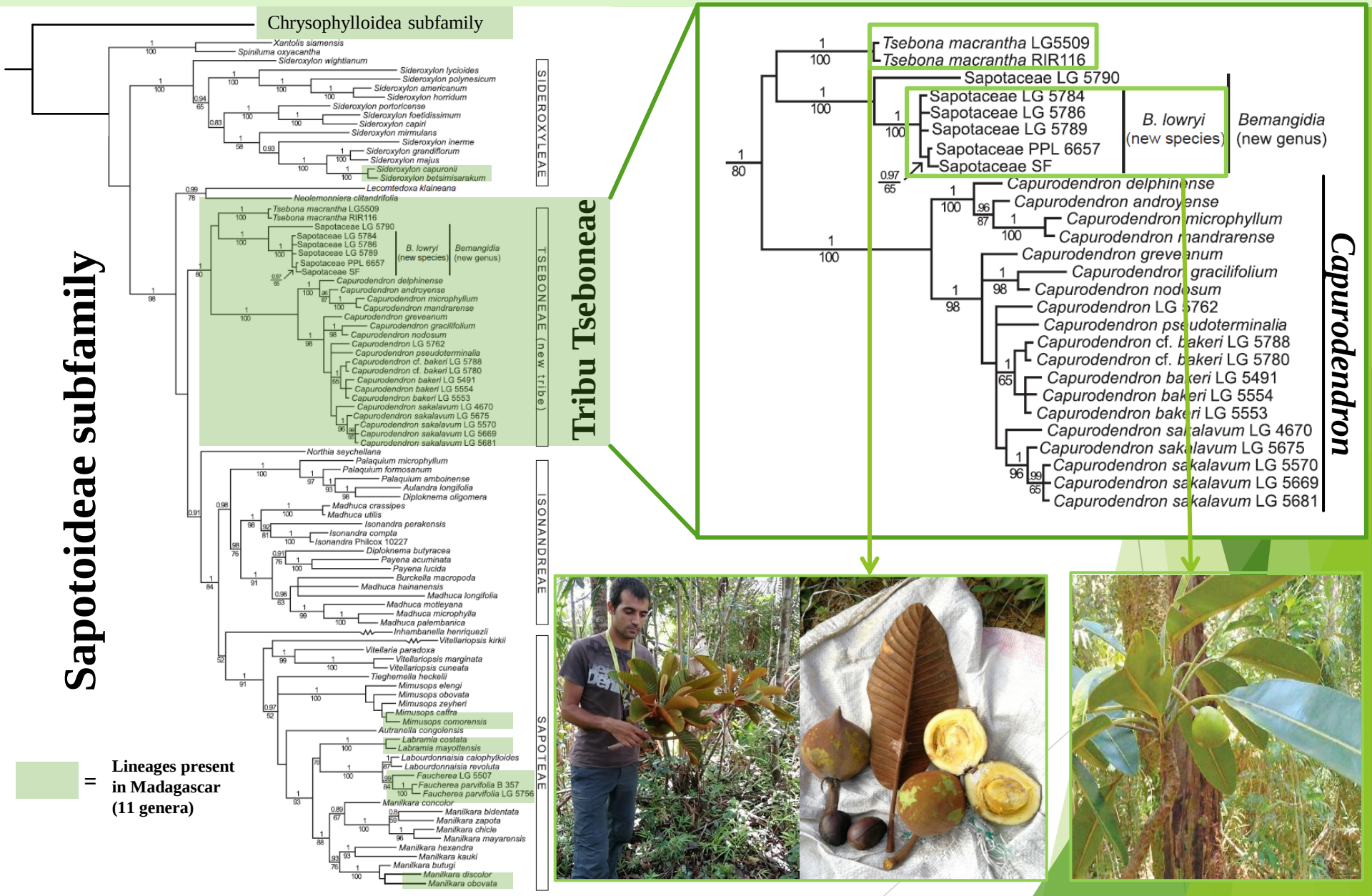


Slash-and-burn agriculture

***Capurodendron* also endangered by:**

Wood felling

Introduction



Tsebona macrantha

Bemangidia lowryi

Gautier *et al.* (2013) Taxon 62:972–983

Introduction

Genus *Capurodendron*:

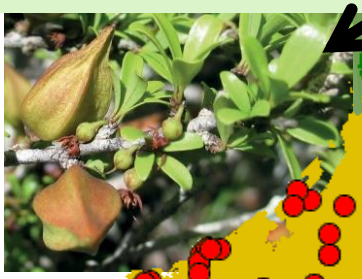
Undescribed morphologies
Species only known from type specimen
Intermediate morphologies
Morpho/species complexes

Undescribed species
Extinct species
Hybridization
Current speciation

?

C. delphinense

C. sahafariense



C. greveanum

C. androyense

Climatic map of Madagascar

C. schatzii

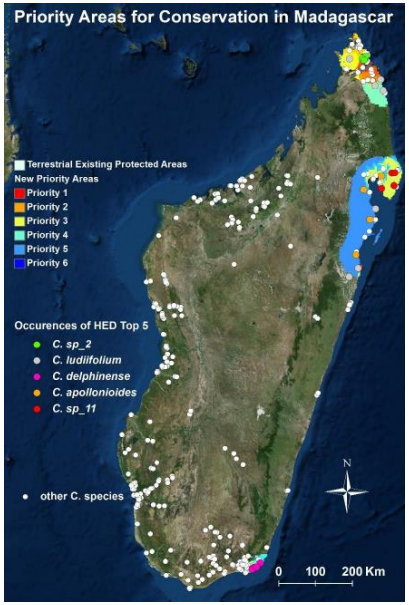
C. tampinense

Objectives

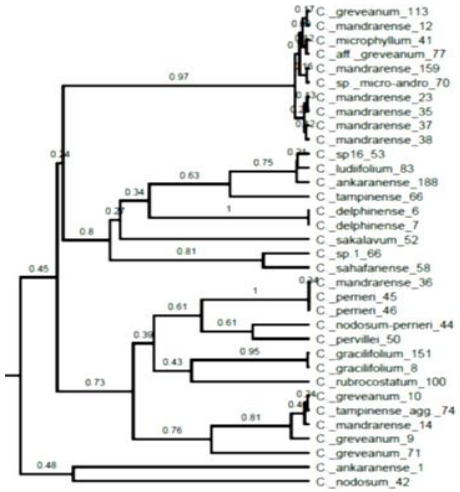
Main objectives:

- 1° Delimit the *Capurodendron* species using phylogenomics.

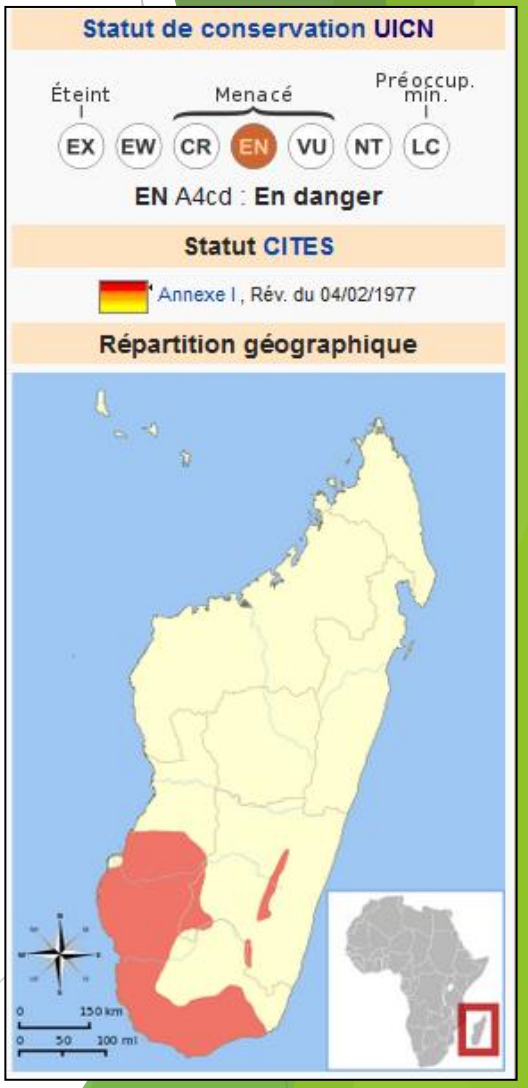
- 2° Estimate the potential distribution of each species and establish the UICN protection categories.



Priority areas for conservation



Species delimitation



Potential distribution, UICN category

Matterials and Methods

Sequences obtention:

Hundreds of genes per specimen are wanted

Undescribed species
Extinct species
Hybridization
Current speciation

?

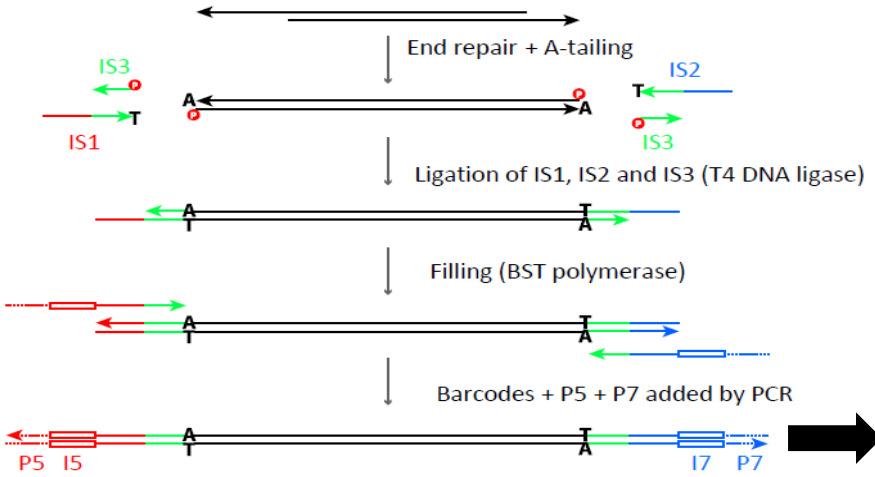
Silica gel samples

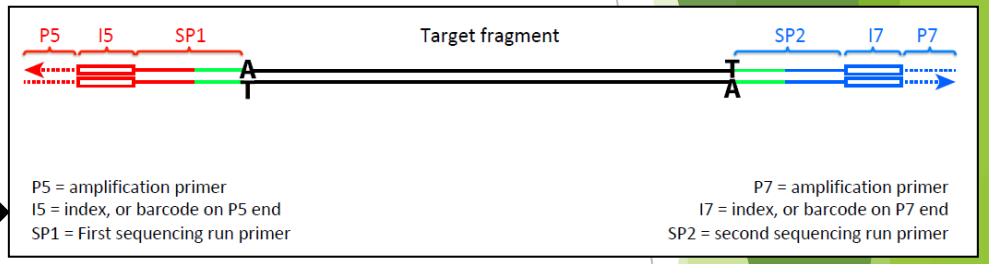
Herbarium samples up to 80 years old

281 specimens:
239 ingroup
42 outgroup

Matterials and Methods

Library construction:





(Kircher 2011)

281 specimens:
239 ingroup
42 outgroup

281 Libraries

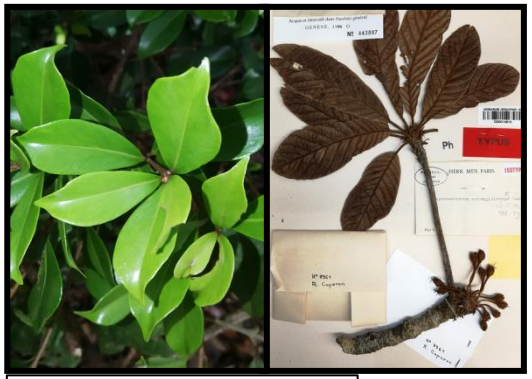
Matterials and Methods

Baits design:

Newly sequenced genomes

Bemangidia lowryi
81 million reads
20x – 40x

C. delphinense
51 million reads
2x – 20x



281 specimens:
239 ingroup
42 outgroup

281 Libraries

Matterials and Methods

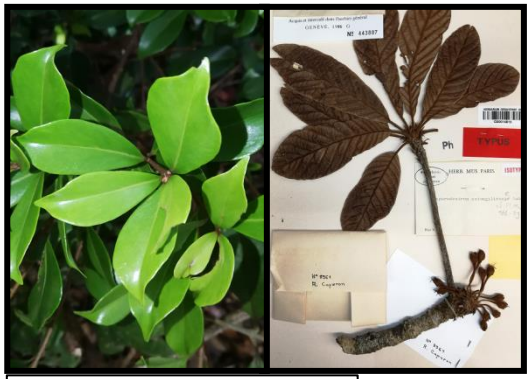
Baits design:

GenBank transcriptome **Newly sequenced genomes**

Manilkara zapota

Bemangidia lowryi
81 million reads
20x – 40x

C. delphinense
51 million reads
2x – 20x



281 specimens:
239 ingroup
42 outgroup

281 Libraries

Matterials and Methods

Baits design:

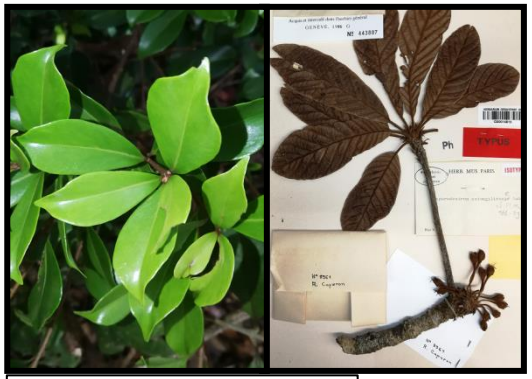
GenBank transcriptome | **Newly sequenced genomes**

Manilkara zapota

Bemangidia lowryi
81 million reads
20x – 40x

C. delphinense
51 million reads
2x – 20x

Baits design



281 specimens:
239 ingroup
42 outgroup

281 Libraries

Baits for 1020 loci:
793 genes
227 microsatellites

Baits: small DNA sequences complementary to a locus that allow us to capture these locus from a genomic DNA solution by hybridization

Baits designed for:

227 microsatellites → Species complexes

532 genes from *Tseboneae* → Species level

262 genes from Angiosperms → Suprageneric level
(Johnson *et al.* 2018)

Camille Christe

Matterials and Methods

Gene Capture

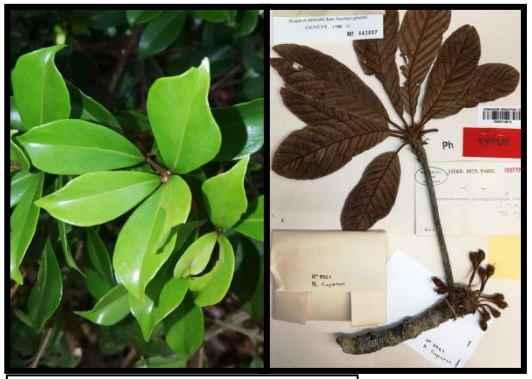
GenBank transcriptome | **Newly sequenced genomes**

Manilkara zapota

Bemangidia lowryi
81 million reads
20x – 40x

C. delphinense
51 million reads
2x – 20x

Baits design



281 spécimens:
239 ingroup
42 outgroup

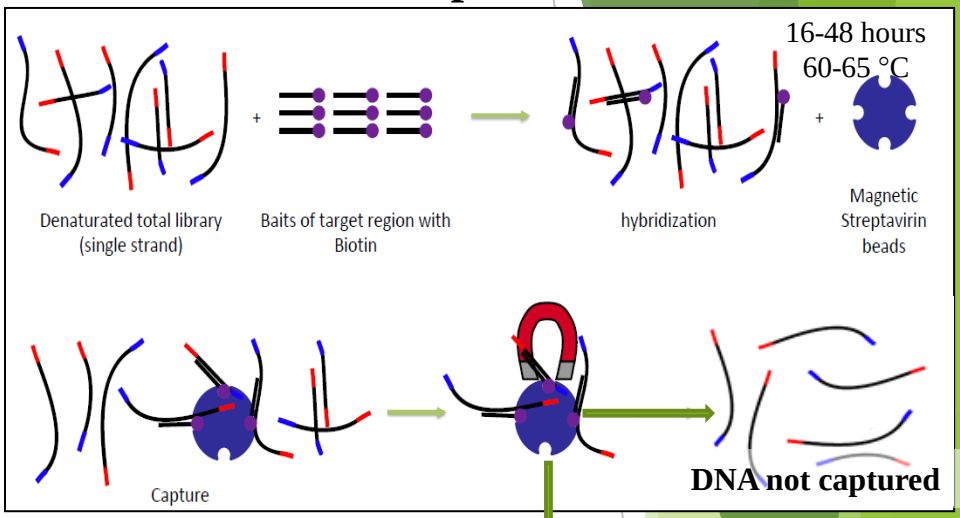
Baits for 1020 loci:
793 genes
227 microsatellites

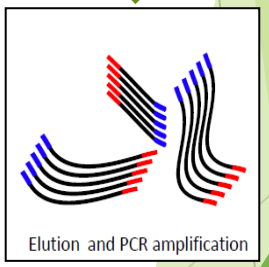
281 Libraries

Gene Capture

Pooled in 7 tubes according to size and baits specificity

Gene Capture





Captured genes

Matterials and Methods

Sequencing

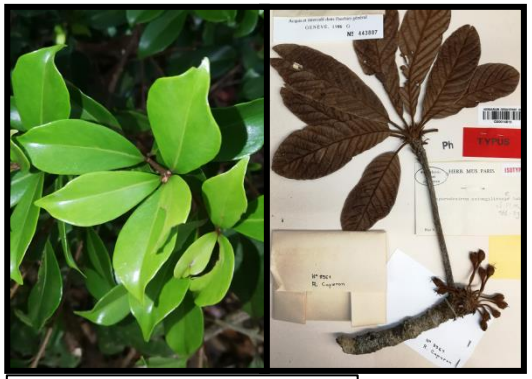
GenBank transcriptome | **Newly sequenced genomes**

Manilkara zapota

Bemangidia lowryi
81 million reads
20x – 40x

C. delphinense
51 million reads
2x – 20x

Baits design



281 spécimens:
239 ingroup
42 outgroup

Baits for 1020 loci:
793 genes
227 microsatellites

281 Libraries → **Gene Capture** → **Captured loci**

Pooled in 5 tubes according to size

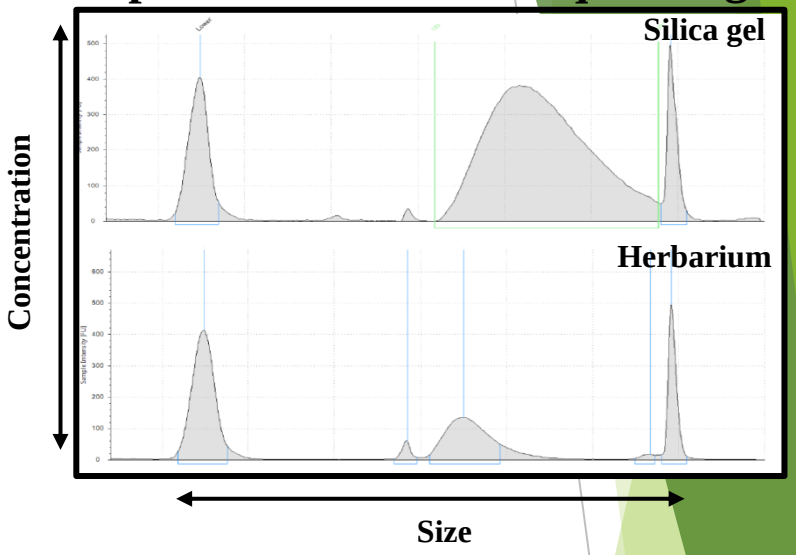
Pooled in 2 tubes according to size

Illumina sequencing

→ **Unicopy genes**

→ **Multicopy genes**

Example of tubes sent to sequencing



Matterials and Methods

Sequencing

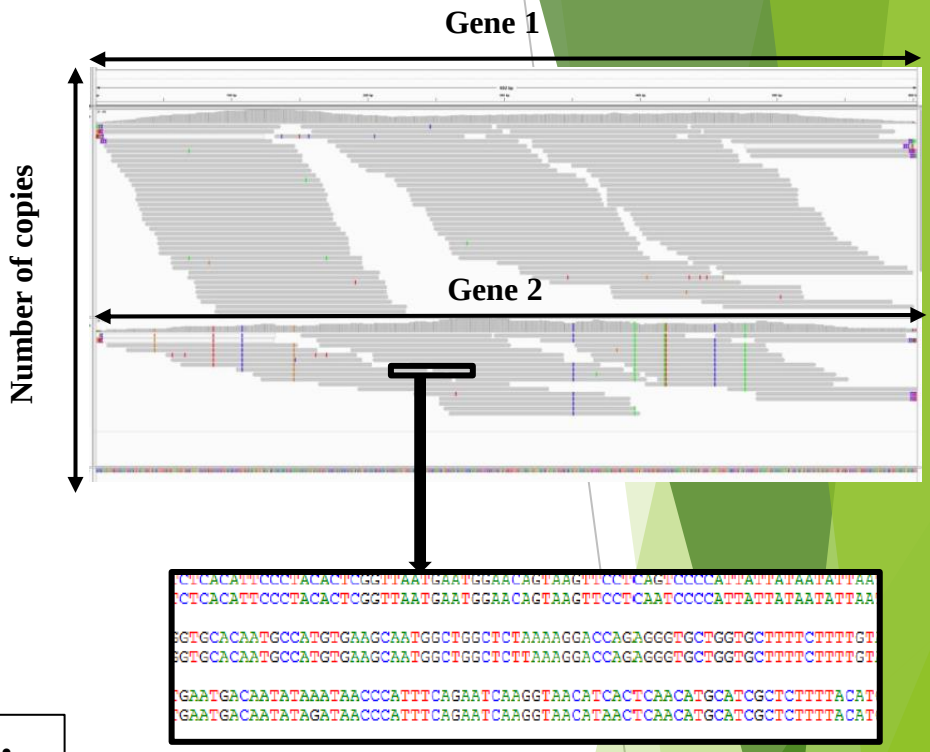
GenBank transcriptome

Manilkara zapota

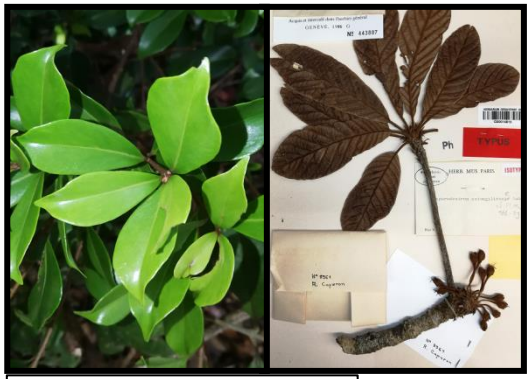
Newly sequenced genomes

Bemangidia lowryi
81 million reads
20x – 40x

C. delphinense
51 million reads
2x – 20x



Baits design



281 spécimens:
239 ingroup
42 outgroup

Baits for 1020 loci:
793 genes
227 microsatellites

Sequences for phylogenomic analyses

281 Libraries → Gene Capture → Captured loci

Pooled in 5 tubes according to size

Pooled in 2 tubes according to size

Illumina sequencing

Unicopy genes

Multicopy genes

Results

Gene Capture has been very efficient ➡ Storing method and collection year had not a strong impact

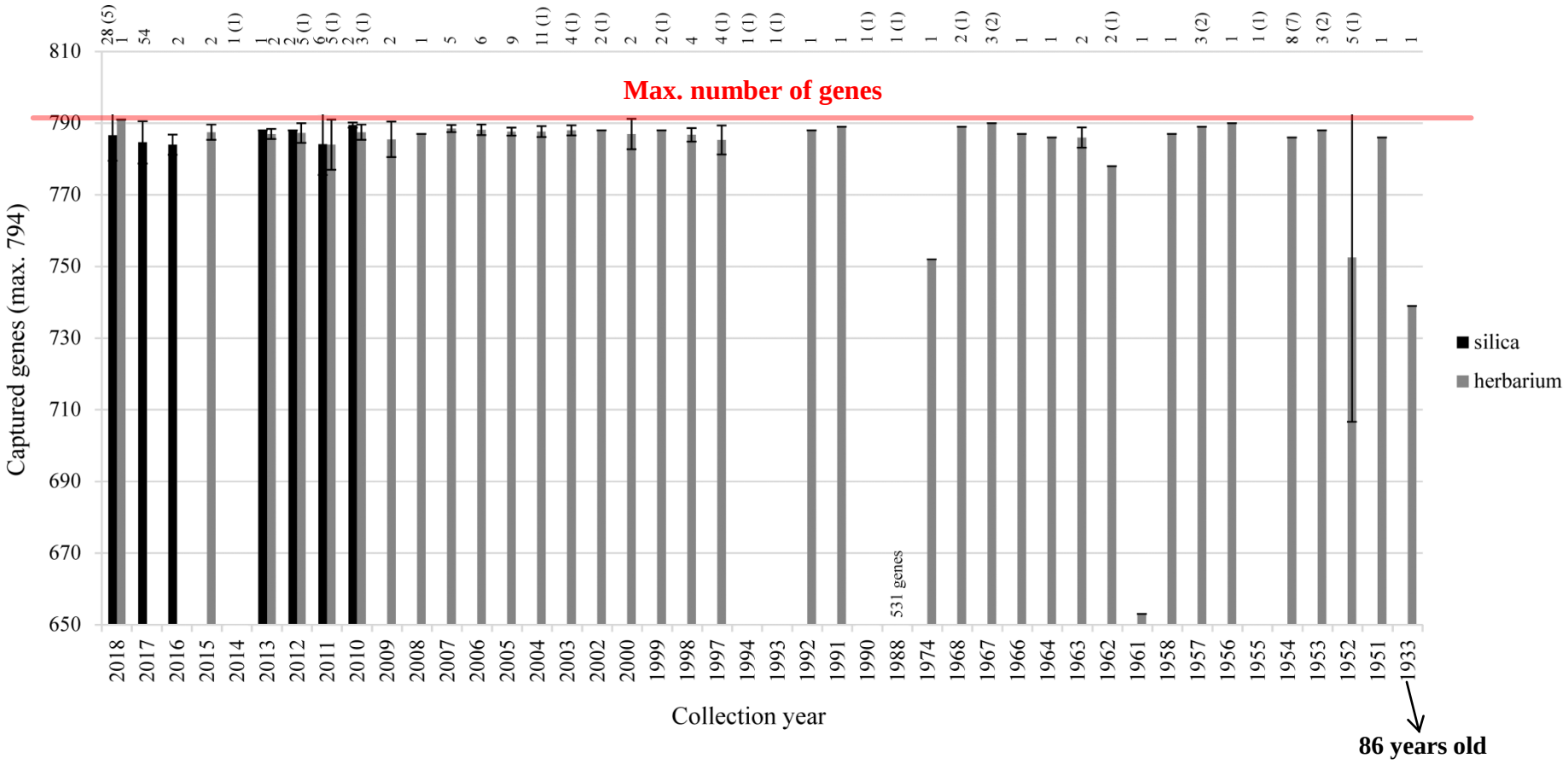


Fig. 2, Number of obtained genes according to the kind of sampling storage (silica-gel or herbarium) and collection year. Values in the upper part indicate the number of analysed samples per year, with the number of failed specimens between brackets.

Results

Gene Capture worked well even in other subfamilies

***= Specimen used for baits design**

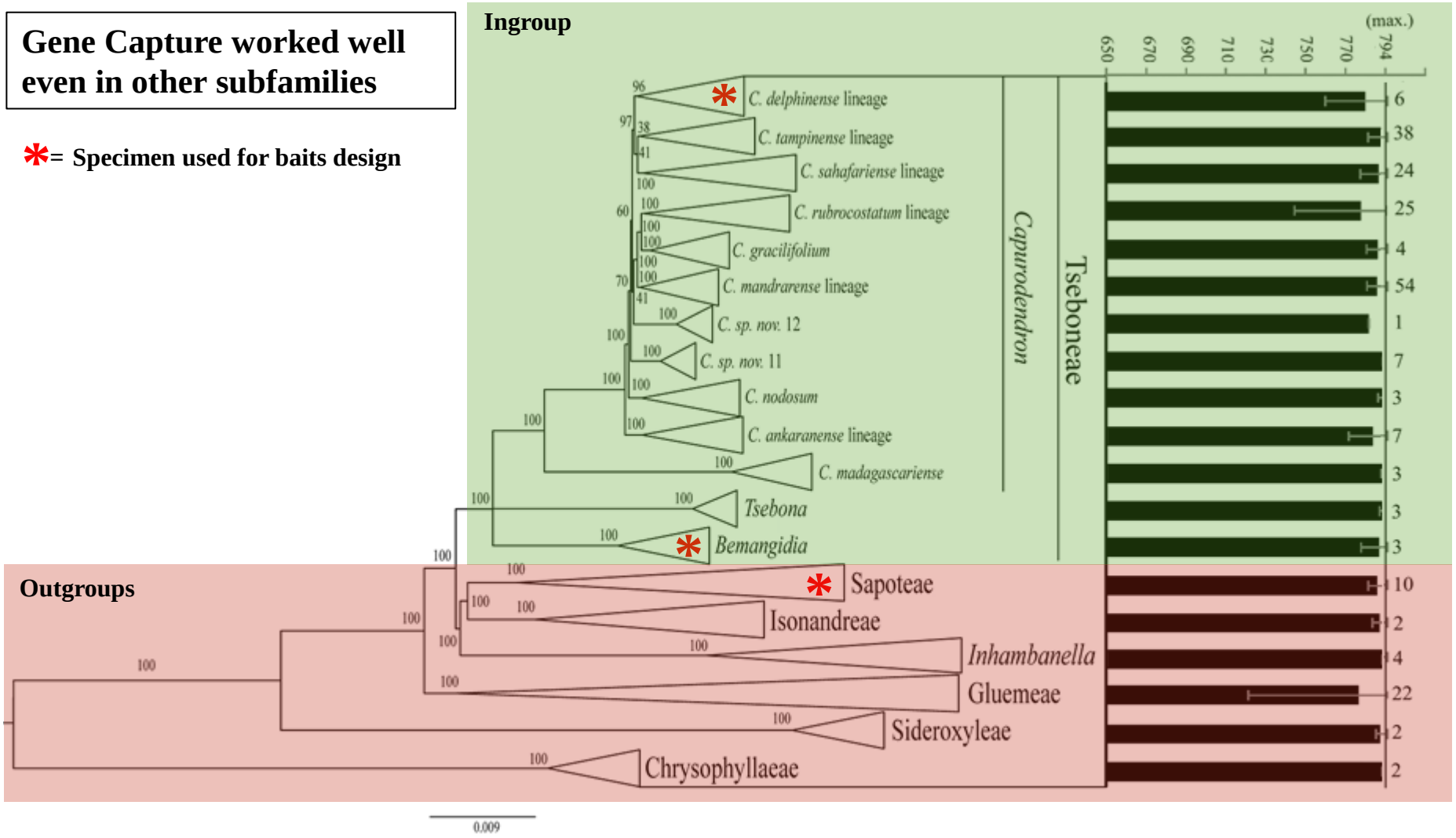
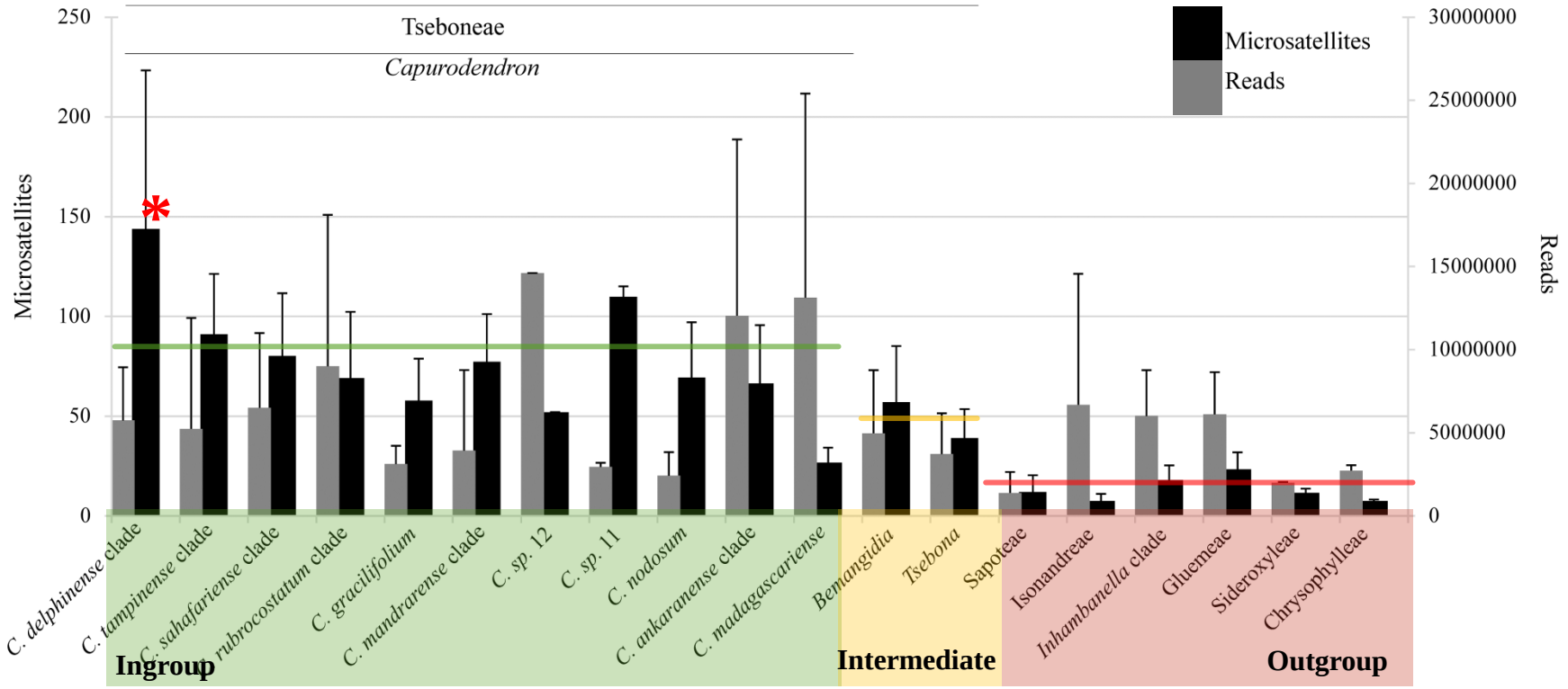


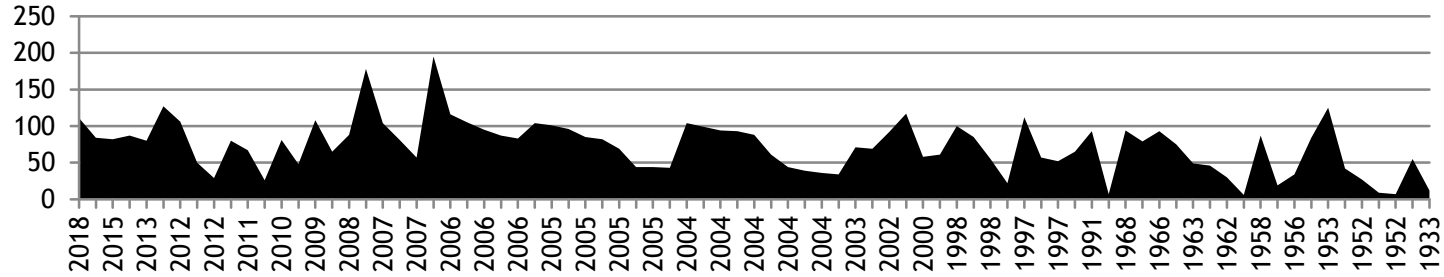
Fig.1 Maximum likelihood tree of 192192 SNPs from519 coding genes and 444 OTUs (222 specimens). The major clades have been collapsed at tribe, genus or infrageneric lineage levels. Bars in the right margin indicate the average number of captured genes per clade and their standard deviation. The number of specimens per clade is indicated at the right margin.

Results

Microsatellite Capture efficency is low → Independent of the number of reads / collection year

*** = Specimen used for baits design**





Number of microsatellites obtained according to collection year (only ingroup)

Results

Microsatellite Capture efficency is low → **Strongly dependent of phylogenetic proximity**

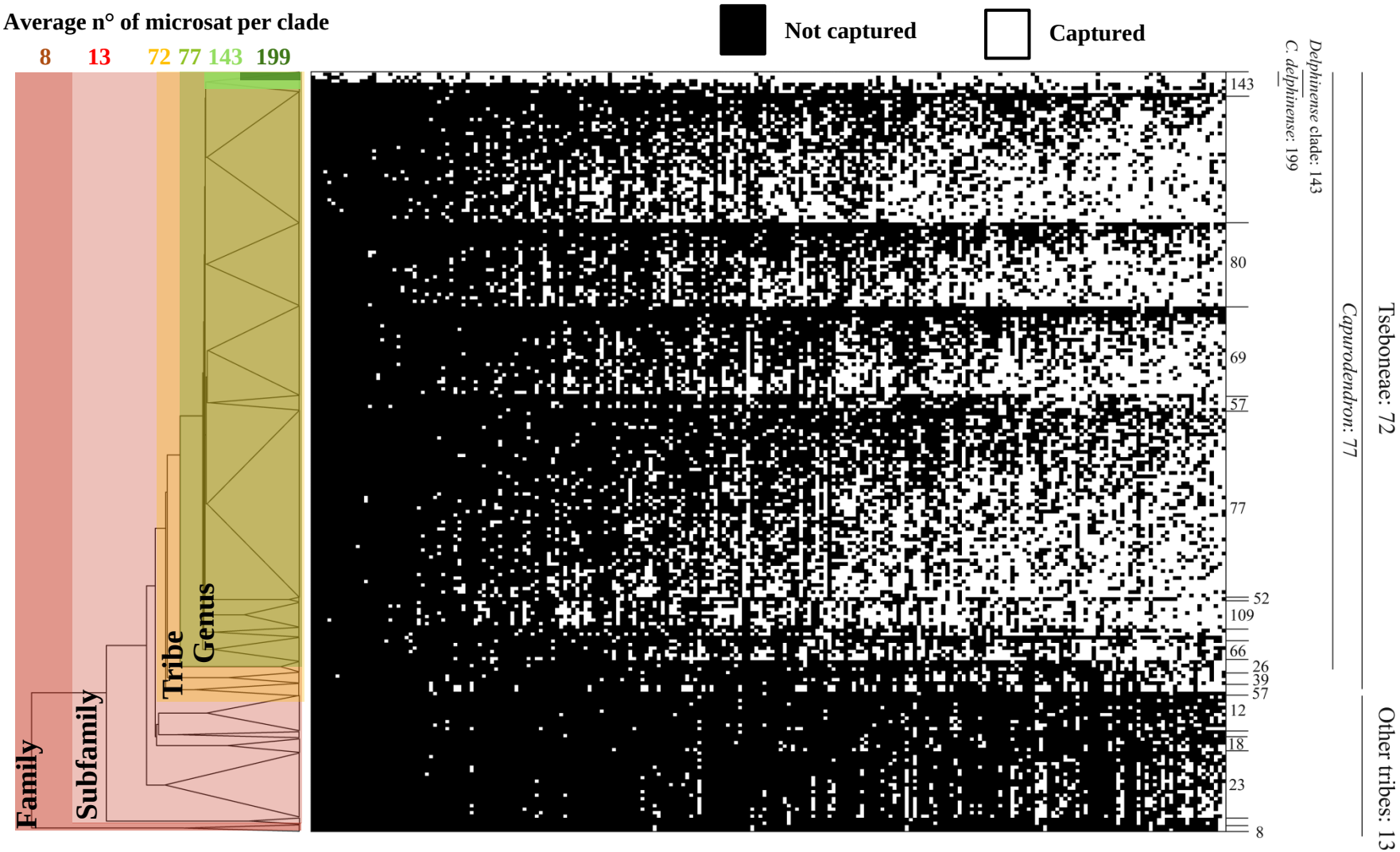
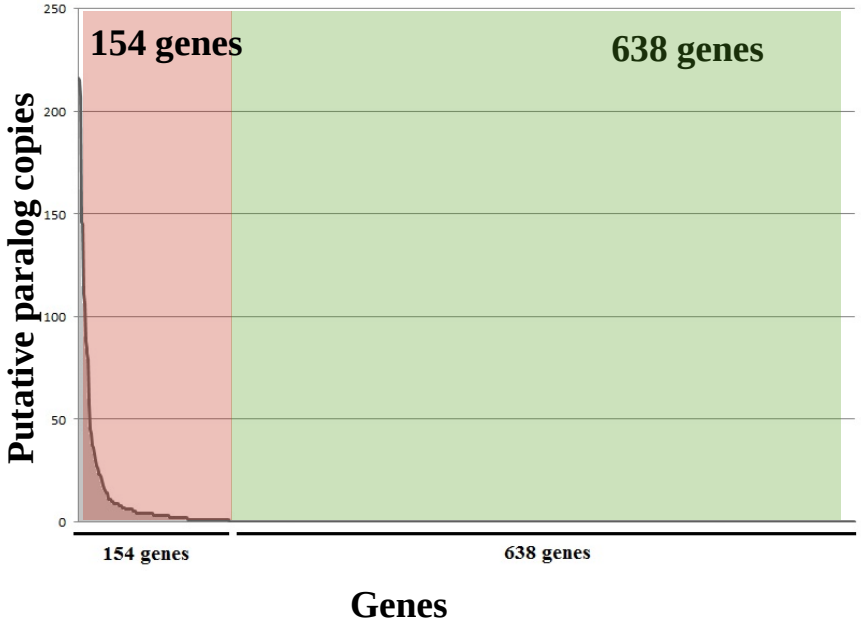


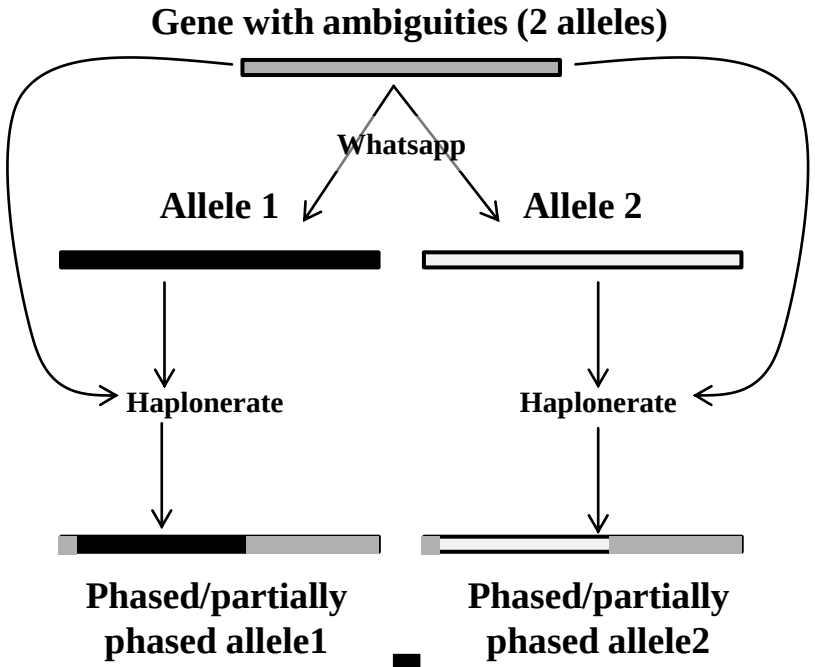
Fig. 4. Matrix showing falled microsatellites (black squares) and obtained (white squares) for each specimen and phylogenetic lineages. Numbers in the right margin indicate the average number of microsatellites succesfully obtained.

Results

Gene selection



Phasing



Phylogeny

Cohalescence (BEAST):

- -Sequences -34 genes
- -444 OTUs -110574 bp
- -72576 computing hours

Same tree topology

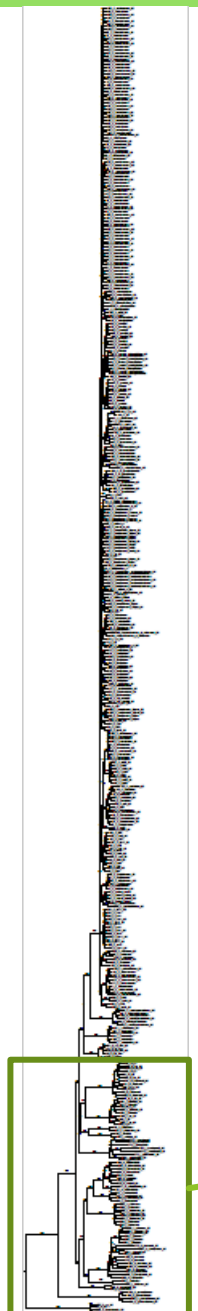
Maximum Likelihood (RaxML):

- -SNPs -519 genes
- -444 OTUs -192000bp
- -4838 computing hours

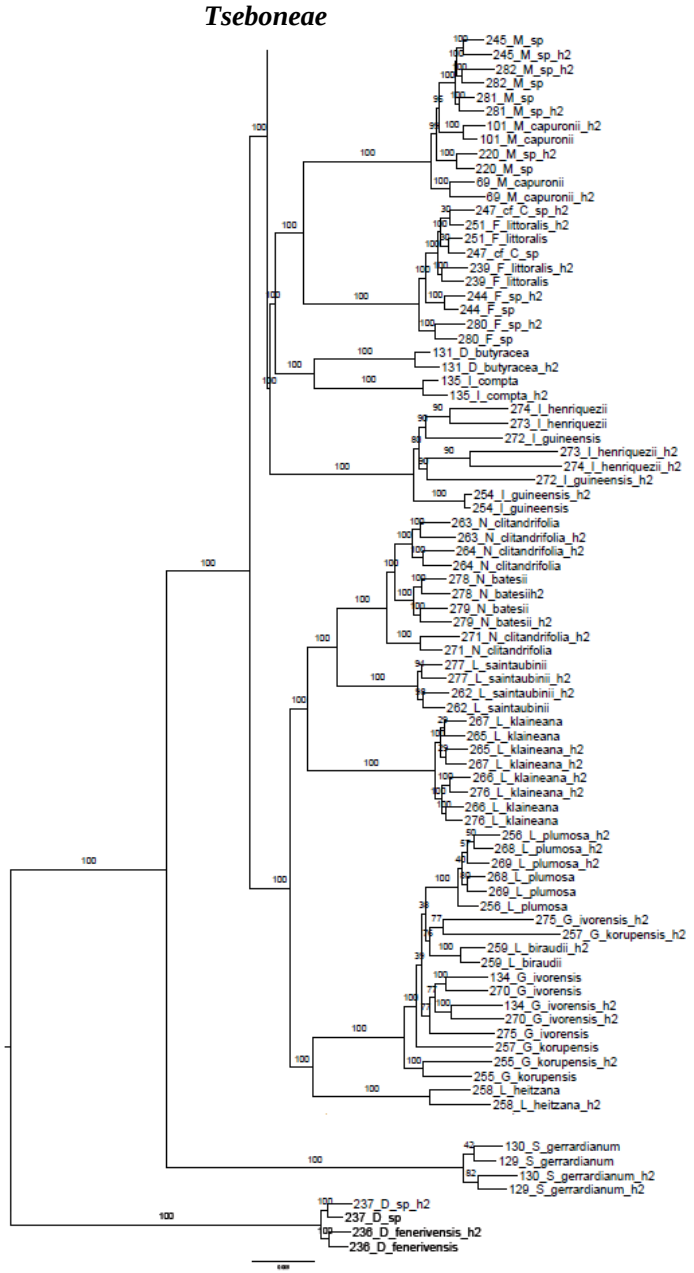
Results

Baits worked well in all the tribes

New projects has been started using the same baits



RaxML tree 519 genes



Tseboneae

Sapoteae

Isonandreae

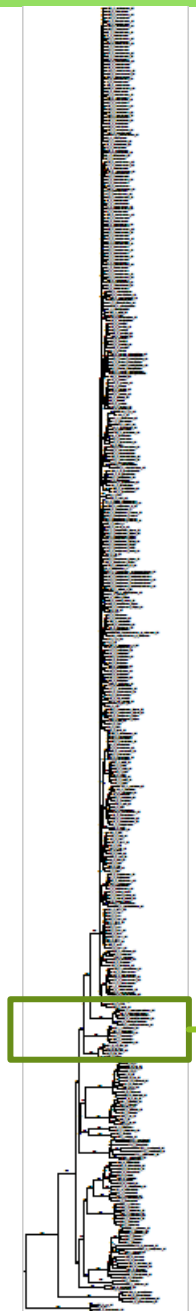
Inhambanella

Gluemeae

Sideroxyleae

Chrysophylleae

Results



Capurodendron

100 89 *C. madagascariense* H2
100 89 *C. madagascariense*
100 185 *C. madagascariense*
100 185 *C. madagascariense* H2
100 94 *C. madagascariense* H2
94 *C. madagascariense*

C. madagascariense

77 121 *B. sp. nov.* H2
100 121 *B. sp. nov.*
122 *B. sp. nov.* H2
122 *B. sp. nov.*

Bemangidia sp. nov.

100 104b *B. lowry*
104b *B. lowry* H2

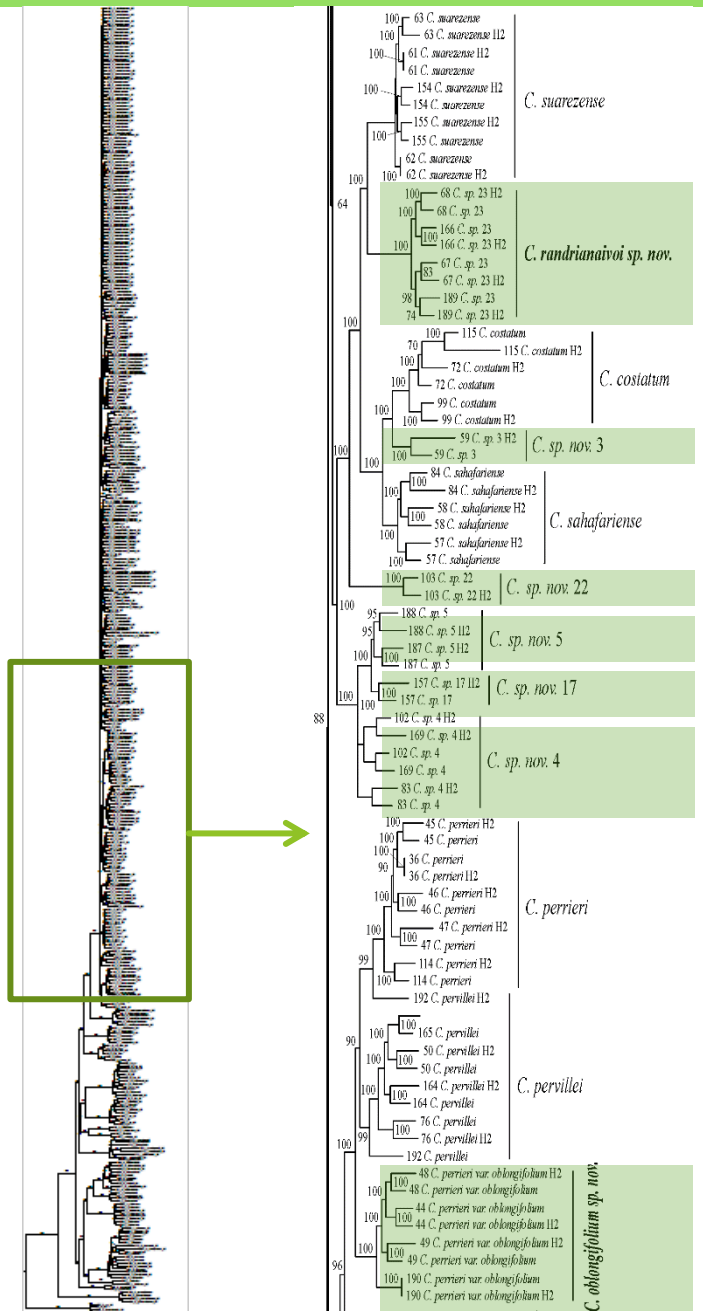
Bemangidia lowry

100 137 *T. macrantha*
100 137 *T. macrantha* H2
214 *T. macrantha* H2
100 214 *T. macrantha* H2
124 *T. macrantha*
100 124 *T. macrantha* H2

Tsebona macrantha

69
100

Results



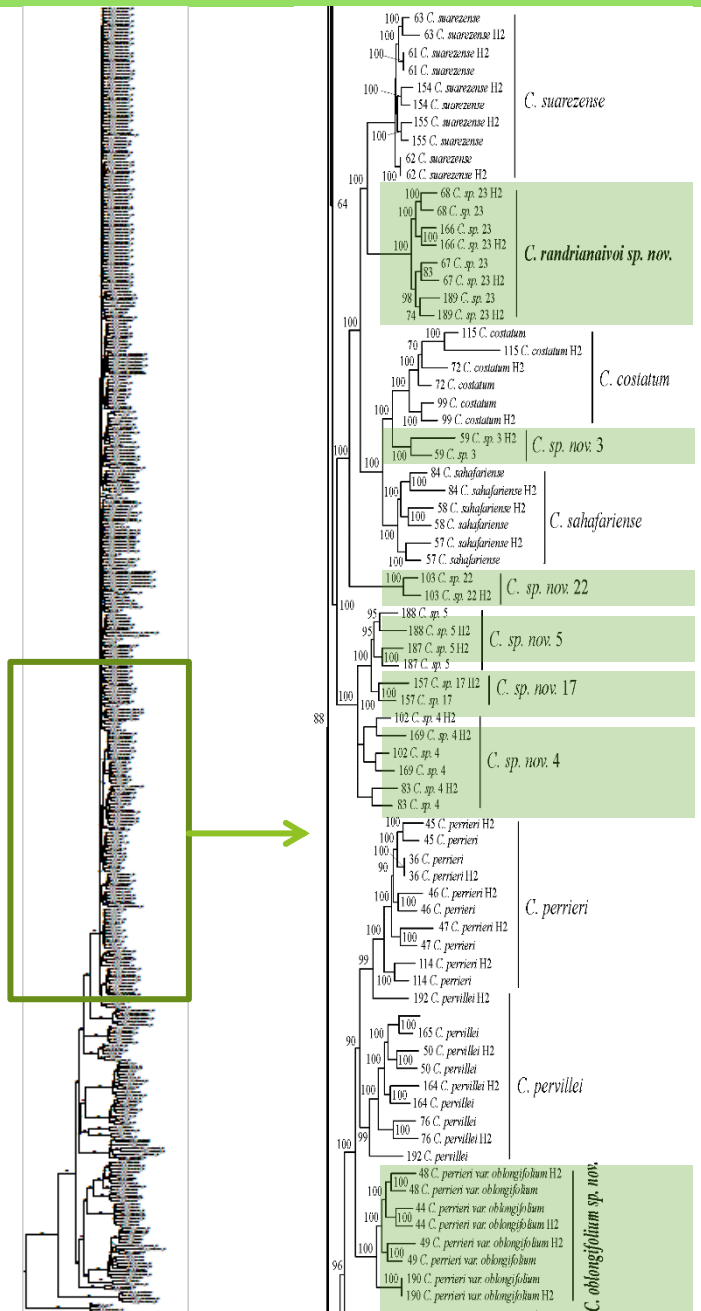
~ 49 clades/morphologies candidates for new species

Only 26 species described

23 new species?

= new species

Results



Described species

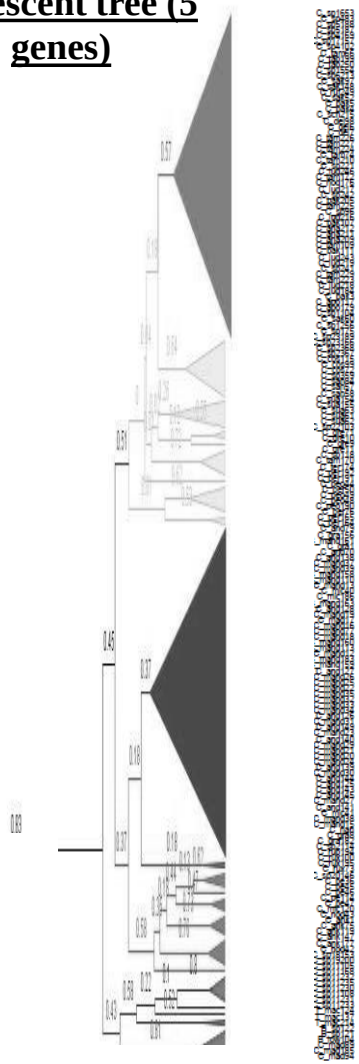
Undescribed species

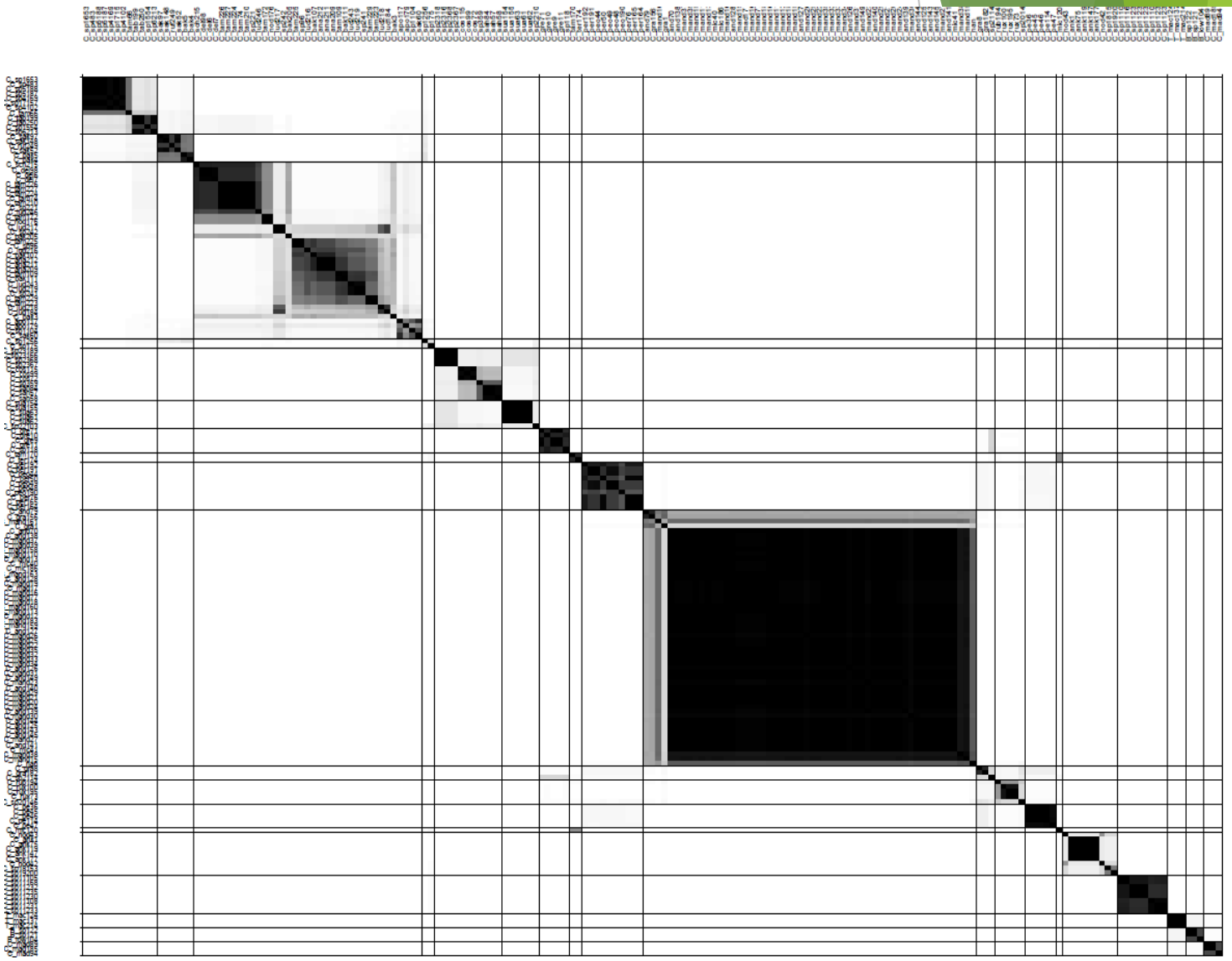
= new species

Results

STACEY analysis for species delimitation

Multi-species coalescent tree (5 genes)





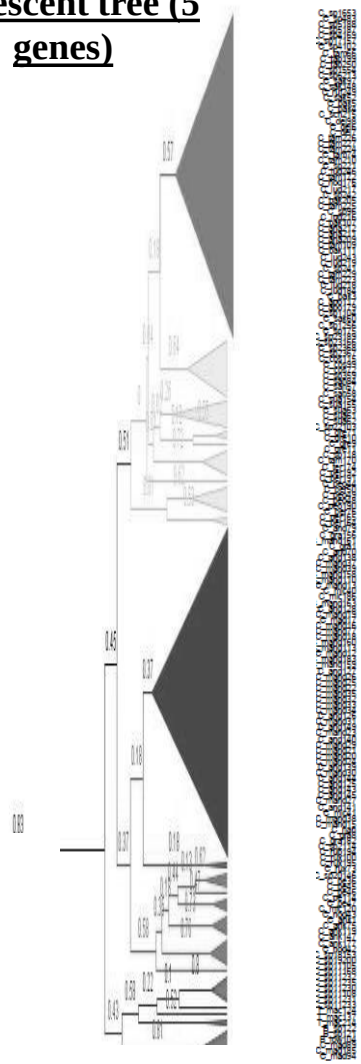
Putative species are indicated in black squares

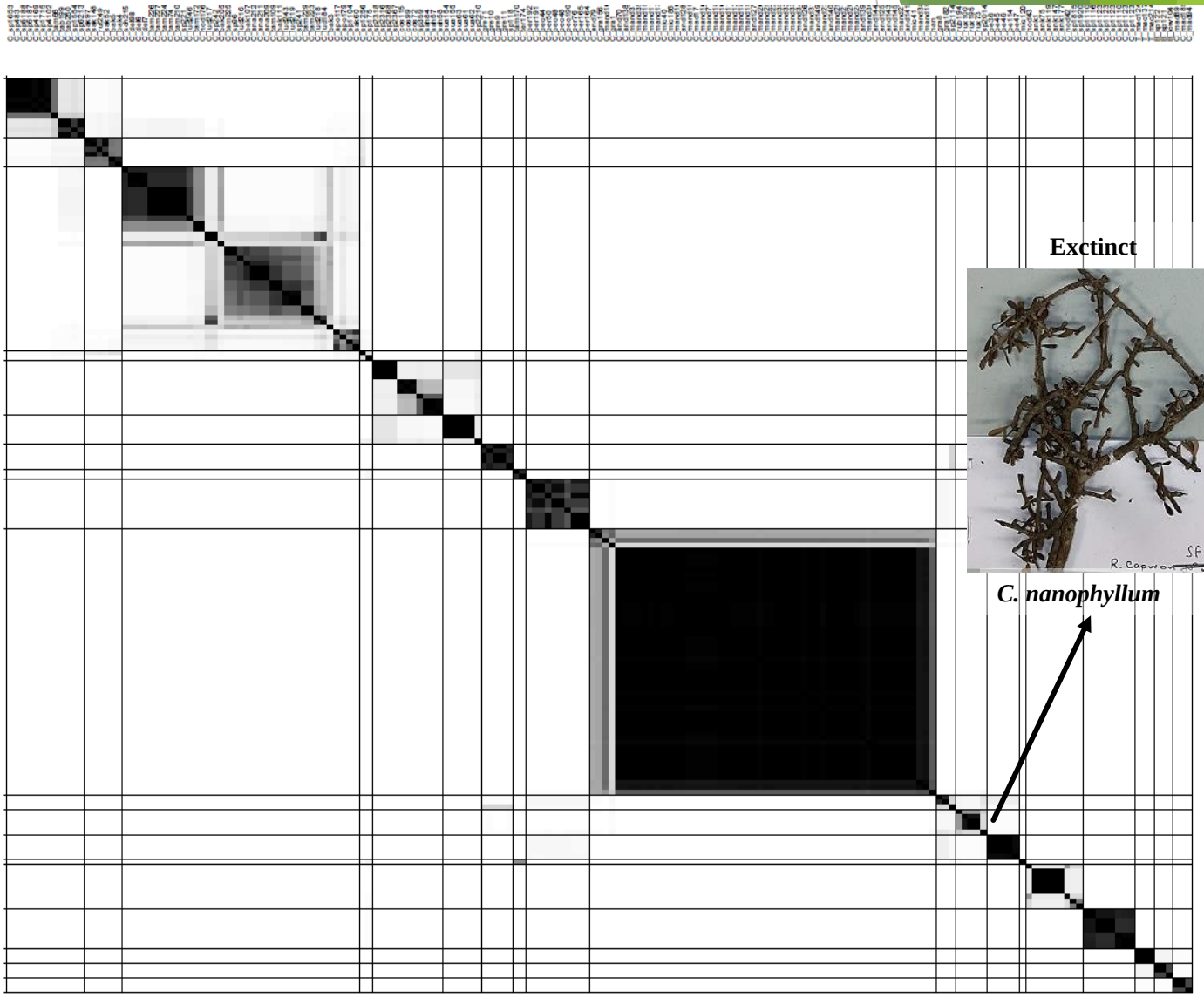
Conspecificity: 0% 50% 100%

Results

STACEY analysis for species delimitation

Multi-species coalescent tree (5 genes)

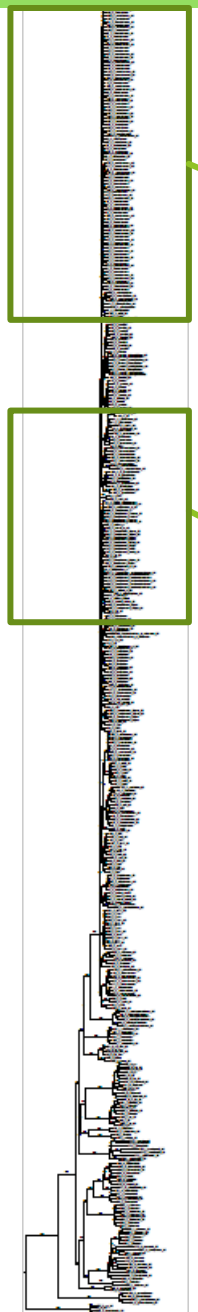


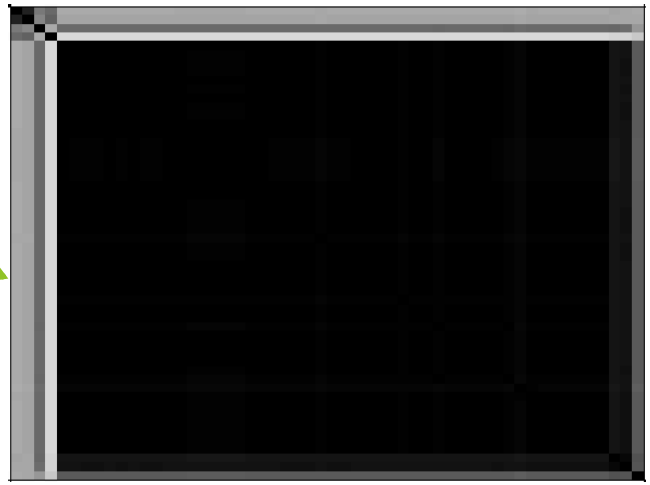


Putative species are indicated in black squares

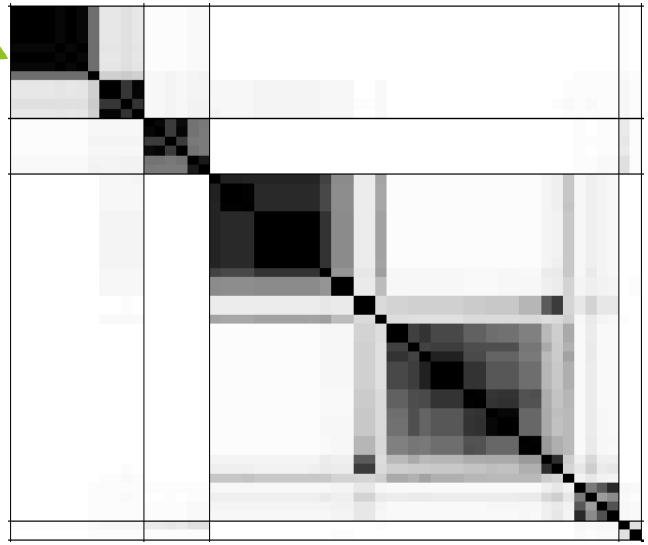
Conspecificity: 0% 50% 100%

Results





Arid species complex



Eastern species complex

Main Conclusions

- **The Gene Capture has been very efficient.**
- **Designed baits can be aplied to all Sapotaceae family.**
- **Microsatellite capture only works well in the reference species.**
- ***Tseboneae* tribe may contain up to 23 undescribed species.**

Aknowledgements

Yamama Naciri

Richard Randrianaivo

Aina Randriarisoa

Carlos Galan Boluda

Camille Christe

Laurent Gautier

Conservatoire et Jardin botaniques Genève

FNSNF Swiss National Science Foundation

Université de Genève

Fondation Ernst et Lucie Schmidheiny

Thank you!

Contact: Carlos.g.boluda@gmail.com