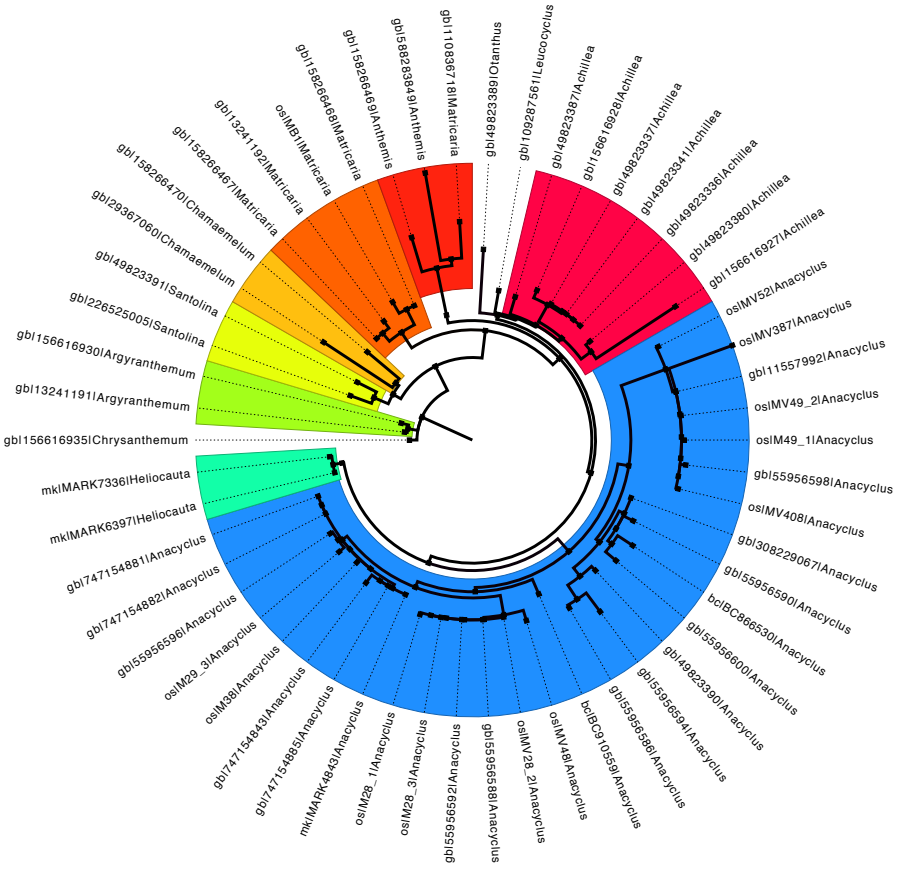
BORRY Maxime

Internship report

November 2015

Phylogeny and automated identification on molecular basis of Anacyclus species using the internal transcribed spacer



Supervisor : Manzanilla Vincent

M1 BEE 2014-2015

Abstract

Anacyclus is a plant genus composed of 14 species and subspecies, part of the Asteraceae family. Among this genus belongs a plant of a particular interest for medicinal uses, *Anacyclus pyrethrum*. This plant, originated from the mediterranean basin, has been used throughout history by the healers of different societies. Nowadays, with the increasing trade of *A. pyrethrum* and the new threats against it, making it and endangered specie, a metabarcoding identification of the plant, and a phylogeny has been developed. New previously un-described hybrids of the genus were sequenced and phylogenetically and morphologically related to their parents, *A. pyrethrum* and *A. radiatus*. The phylogeny showed a good distribution of the *Anacyclus* species into clades for most of the them but three, whose clades were not supported. The marker used, ITS, is efficient for the identification of *A. pyrethrum* in the context of market sample products metabarcoding. The software developed for the identification of Anacyclus species is relevant for a quick diagnosis of market samples composition. Finally, the phylogeny of *Anacyclus* and other closely related Asteraceae genera shows the power and limitations of the ITS marker for this application.

Introduction

Medicinal plants have been used for centuries by humans for their healing properties, from the prehistorical time and even before, by other primates such as monkeys, gorilla, and chimpanzees (Huffman, 2004). The first written records of the use and properties of medicinal plants came with the invention of writing in the ancient Egyptian society, around 3000 to 6000 years ago (Nunn, Ancient Egyptian Medicine.). The use of plants as medical therapeutic options continued to develop throughout the ages, in the ancient Greece (Dioscorides, De Materia Medica), in the Aztec and Maya culture (Berdan et al., 2005), and so on, up until today. In 2006, the World Health Organization estimated the demand the annual demand for medicinal plants at $14 billion: between 65% and 85% of the world population is relying on herbal medicines.

However, it wasn't until the work of Carl von Linné, the Swedish naturalist whose masterpieces «Genera plantarum » and « Species plantarum» were published between 1737 and 1800, that a serious description of the plant genera and species was established, thus allowing a more precise identification and use of medicinal plants.

Yet some of the species described in « Species plantarum » remained misclassified and misidentified long after Linné's work, and taxonomy is still an ongoing debate today.

One of these genera, first named by Linné, is Anacyclus.

This genus, part of Asteraceae (or Compositeae) family of angiosperms and eudicotyledons plants is nowadays recognized to include 12 species and 2 subspecies (Bisby et al., 2010) *(Table 1)* and is

Anacyclus anatolicus
Anacyclus ciliatus
Anacyclus clavatus
Anacyclus homogamos
Anacyclus latealatus
Anacyclus linearilobus
Anacyclus maroccanus
Anancyclus monanthos
Anacyclus monanthos subsp.cyrtolepidioides
Anacyclus nigellifolius
Anacyclys pyrethrum
Anacyclus radiatus
Anacyclus radiatus subsp. coronatus
Anacyclus valentinus

Table 1 : list of the species and subspecies in the Anacyclus genus

native from the Mediterranean basin, though it's been exported and grown in many similar climate regions in the world.

Among all of the species, one steps out of the genus for its medicinal properties : *Anacyclus pyrethrum (fig 1)*. Already used in the ancient roman era by Magians, religious experts of the Persian empire, under the name parthenium, it was used to treat fevers according to Pliny the Elder and Dioscorides (Annalakshmi et al., 2012). Later imported to Asia, it was and still is intensely used in traditional Ayurvedic medicine : 60 to 80 tones of dry roots are exported from Morocco to Asia every year. (Moroccan « Office des changes »)

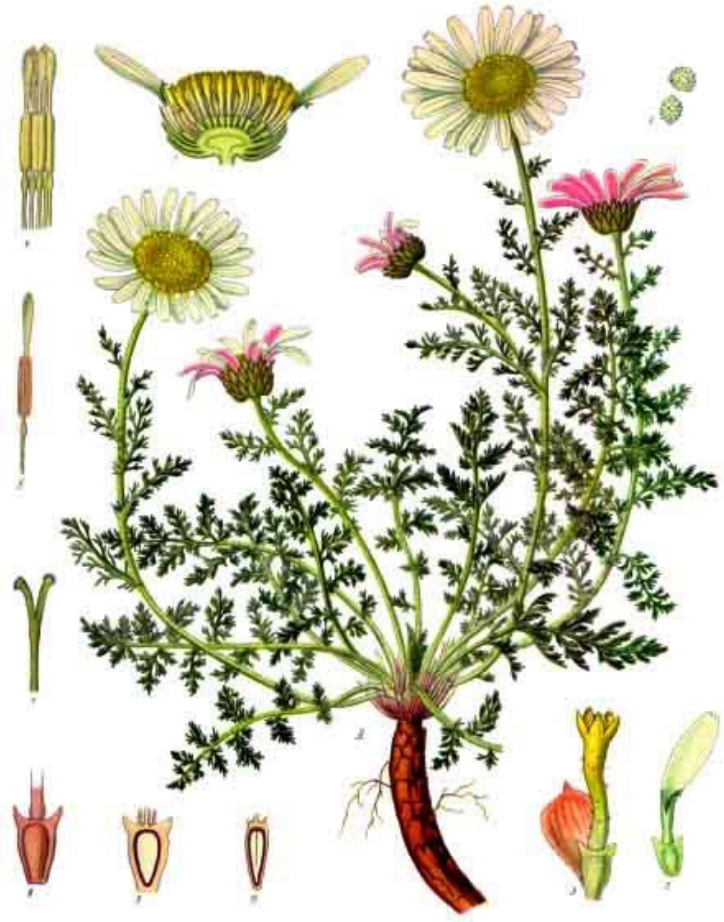


Figure 1 : Anacyclus pyrethrum - © Thomas Schöpke

The root of *Anacyclus pyrethrum* is especially of a particular interest for its numerous active phytochemicals. Among the medicinal properties, the following are described to the day:

- The immunostimulating effect of the hot water-soluble polysaccharide extracts: they induce an increased activity of the reticulo-endothelial system, involved in the phagocytosis of old red blood cells, and an increased number of spleen cells in the mice (Bendjeddou et al., 2003).

- The antidepressant activity was demonstrated on swiss male albino mice by different methods. The *Anacyclus pyrethrum* root extracts showed antidepressant effect by interaction with adrenergic and dopamine receptor (Badhe et al, 2010)
- The memory enhancing activity was shown on albino Wistar rats. The treatment with ethanolic extract of *A. pyrethrum* increased brain cholinesterase level, thus enhancing memory activity (Sujith et al., 2012)
- One of the earliest use of *A. pyrethrum* properties in Ayurvedic medicine: aphrodisiac as demonstrated by Sharma et al. (2009) on albino rats. The aqueous extracts from the roots yielded an increased sperm count in the rat, and an improvement in sexual behavior was observed.
- Last but not least, the local anesthetic effect was demonstrated by Patel et al. (1992) on 200 dentistry patients in a double blind study. Alcoholic extracts were shown to have a similar effect, even at low concentration, compared to the reference anesthetic, xylocaine.

Most of the active phytochemical products of *A. pyrethrum* are extracted from the roots of the plant. Depending on which step of the value chain the product is acquired, a confusion and mix of the product with visually similar others may happen, thus lowering the quality of the final product (Booker et al., 2012). Especially since roots can be very difficult to identify based on their morphology, as they bear very little taxonomic specificities, even more when they are grinded in powders and mixtures, when sold as medicinal preparations in traditional markets. Therefore, other means of identification needed to be applied, such as barcoding (Kool et al., 2012).

The identification of plants at the specie level using molecular data, barcoding, has appeared 20 years ago (Garnock-Jones et. al., 1996) using chloroplast DNA. However, the relatively low taxonomic resolution recovered with traditional chloroplastic markers doesn't allow the identification of samples to specie level (Hollingsworth et al., 2011). Given this constraints, another marker was chosen: the entire Internal Transcribed Spacer (ITS) because of its proven resolution for closely related species, especially in the Asteraceae family (Baldwin, 1991)(Oberprieler, 2004). The ITS is a non coding spacer DNA found in two versions in the plants: ITS1 situated between the 18S and 5.8S rRNA genes, and ITS2 situated between the 5.8S and 26S.

Using this marker, I developed a phylogeny of the Anacyclus genus to verify if barcoding to the specie level was possible, and try to bring more light on the complicated evolutionary history in the *Anacyclus* genus, by the many hybridization events, and the important phenotypic plasticity (Bello et. al., 2013). And for the two new possible *Anacyclus* hybrids discovered, answer the question: what are their parents ?

I also developed a software for visualization as a heatmap of metabarcoding outputs of Anacyclus mixtures.

Material and methods

Plant material

The plant samples are coming from different sources, some were collected directly from the field in Morocco (identifier MV), some were collected in Oslo (identifier MB), some were gathered from different herbarium (Barcelona : identifier BC, Marrakech : identifier MARK). Missing Anacyclus sequences and outgroup sequences were selected from Genbank.

Molecular Biology

DNA extraction

The DNA extraction was performed in the laboratory of the Natural History Museum.

- In tubes: ~15mg of plant leaf and/or stem sample was grinded in a Biospec Mini-Beadbeater-1® at 4200 rpm for 50 seconds. After the grinding process, the total DNA was extracted using the Omega Bio-tek E.Z.N.A.® Plant DNA kit according to the manufacturer recommendations.
- In 96 well plates: ~15mg of plant samples were grinded in Retsch MM 301 mixer mill . After the grinding process, the total DNA was extracted using the Qiagen DNAeasy® 96 Plant Kit according to the manufacturer recommendations.

After extraction, DNA samples were kept at -80°C.

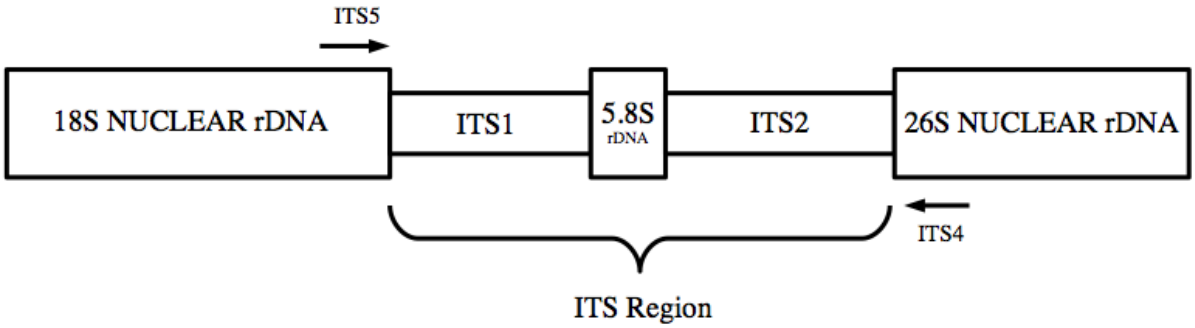


Figure 2 : 18S-26S nuclear ribosomal DNA regions. PCR primers are indicated by the arrows.

PCR amplification

For the the PCR amplification of the whole ITS region, the ITS5 and ITS4 primers (White et al., 1989) were chosen (fig 2). PCR reactions were prepared in a 20μL total volume using 12.5% of Qiagen PCR buffer 10X, 1.875mM of added $MgCl_2$, 0.25μM of dNTP, 0.125mg/mL of BSA, 0.6μM of each primer, 1 unit of ThermoFisher Scientific AmpliTaq® DNA Polymerase, and 1μL of DNA sample. The remainder to the total volume being filled with Milli-Q® purified water.

The PCR reaction itself was performed in a Bio-Rad T100™ Thermal Cycler using the setup described in figure 3.

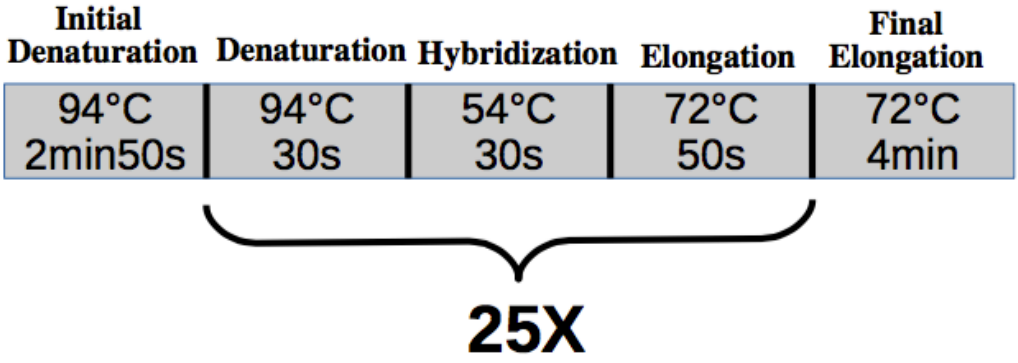


Figure 3 : PCR setup on the Bio-Rad T100™ Thermal Cycler for ITS primers

To check for the lack of non-plant material contamination, a PCR amplification with the primers g and h of the chloroplast trnL intron (Tarberlet et al., 2007) was performed, with the following PCR setup : 10 minutes at 95°C, followed by 35 cycles of 30s at 95°C, 30s at 50°C, and 2 min at 72°C. The PCR mixture and Thermal cycler are the same as those used for the ITS amplification.

Gel electrophoresis

PCR amplification was verified on gel electrophoresis with 1% agarose with 0.004% of Biotium GelRed™ (replacing the ethidium bromide). 4μL of PCR product were mixed with 1μL of New England BioLabs® Gel Loading Dye, Purple (6X). The electrophoresis was run for 30 minutes at 90V. Gels were imaged with a Kodak Gel Logic 200 gel imager and edited with ImageJ.

ExoSAP clean-up

The combination of an **Exo**nuclease and a **S**hrimp **A**lcaline **P**hosphatase allows the removal of the excess of primers and nucleotides leftovers. 2μL of Affymetrix 1X ExoSAP-IT® were mixed with 6μL of successfully amplified PCR product. It was then incubated in a Bio-Rad T100™ Thermal Cycler at 37°C for 45 min, followed by 15 min at 80°C for the inactivation of the ExoSAP.

Cycle sequencing

The cycle sequencing was performed using the ThermoFisher Scientific BigDye® Terminator v3.1 cycle sequencing kit. For each sample, two preparations were realized, one for the forward sequencing primer, another one for the reverse sequencing primer. For each sample, and each primer, the total volume was 9μL divided in 1.8μL of buffer, 0.5μL of sequencing primer (forward

or reverse), 0.4μL of BigDye® v3.1, the remainder to the total volume being filled with Milli-Q® purified water. The cycle sequencing setup is described in figure 4.

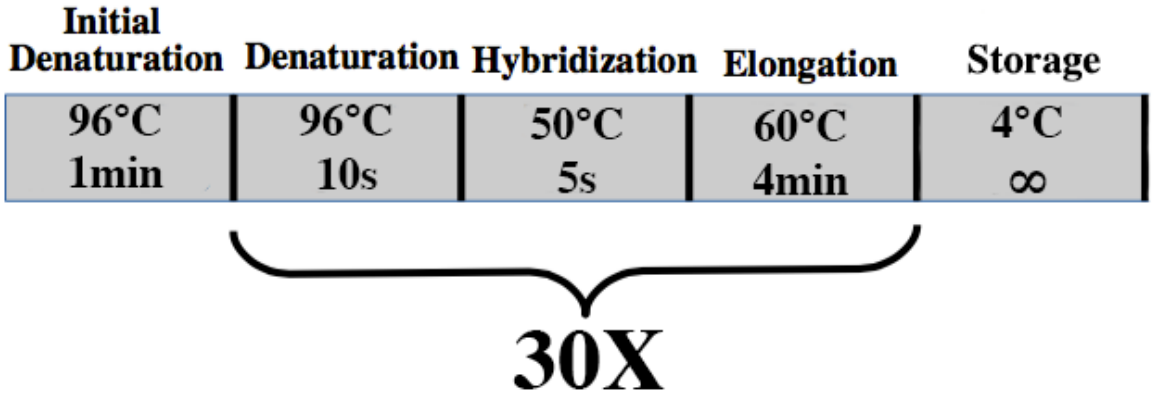


Figure 4 : BigDye® Terminator v3.1 cycle sequencing setup

Ethanol precipitation

After sequencing, samples were purified using ethanol. 1μL of 125mM EDTA, 1μL of 3M NaAc, and 25μL of 96% Ethanol were added to each sample. After being vortexed and let still on the bench for 15 minutes, samples were centrifuged for 25 minutes at 5500 rpm at 4°C. Supernatant was removed by centrifuging the samples upside down for 15 seconds at 400pm. 35μl of 70% Ethanol was then added to each sample, followed by a centrifugation step of 20 minutes at 5500 rpm and 4°C. The supernatant was removed by repeating the previous inverted sample step. To get rid of the remaining ethanol droplets that might interfere with downstream applications, samples were centrifuged one last time in a vacuum centrifuge for 2 minutes.

Resuspension and capillary electrophoresis

Samples were resuspended in 10μL of ThermoFisher Scientific HIghly DeIonized Formamide (Hi-Di™) and immediately processed on the ABI PRISM « 3130xl Genetix Analyzer » automated capillary electrophoresis machine.

BioInformatics and data analysis

Sanger sequencing assembly and processing

Trace files of the sequences were assembled and corrected with CodonCode Aligner V5.1.5. Each successfully sequenced sample was manually checked for ambiguity. Ambiguity were either solved manually if possible, or deleted with the sequence before/after if necessary. To confirm

identification, each successfully sequenced sample was blasted (blastn) against the GenBank® nucleotide collection (nr/nt).

Selection of Asteraceae samples from GenBank® nucleotide collection

For the outgroup and, missing or poorly represented species, sequences deposited in GenBank® were added to the analysis. 37 sequences of the ITS region were selected, representing 26 different species. Among those sequences, 26 display the full ITS region (ITS1, 5.8S, and ITS2) and 11 are the result of the concatenation of the ITS1 and ITS2 sequence from the same sample, resulting in a sequence missing 5.8S.

Alignment

The sequence matrix was visualized using AliView 1.17.1 and sequences were aligned using the built-in implementation of MUSCLE. The matrix was manually checked and edited if necessary.

Phylogeny

To verify the isolation of the *Anacyclus* genus from the outgroup, a pairwise distance matrix was realized using the « dist.dna » function from the R package « Analysis of Phylogenetics and Evolution » (APE 3.3).

In order to perform the maximum likelihood analysis, the best evolution model was first chosen by running jModelTest 2.1.7. The Maximum likelihood analysis was conducted by PhyML 3.0 with the best model chosen with AIC by jModelTest, with 100 non parametric bootstrap replicates.

Finally, after completion of the analysis, the phylogenetic tree was edited with FigTree 1.4.2 for readability and publication purposes.

Metabarcoding identification and heatmap visualization

To develop the metabarcoding identification and heatmap visualization software, Python 2.7.10 and R 3.2.1 were used, as well as other external softwares such as the Bowtie2 short-read aligner, and Samtools 1.2. The database for identification of NGS samples was assembled with a custom python script using local and GenBank® sequences. As no *Anacyclus* mixture was available for sequencing to test this software, a virtual mixture was created using wgsim to simulate IonTorrent reads.

Further details about this software will be released in an upcoming publication.

Results

PCR and electrophoresis

The result of ITS PCR and gel are shown in figure 5. Amplification was successful for 22 samples out of 24. There was no evidence on any non-plant material on the trnL intron gel.

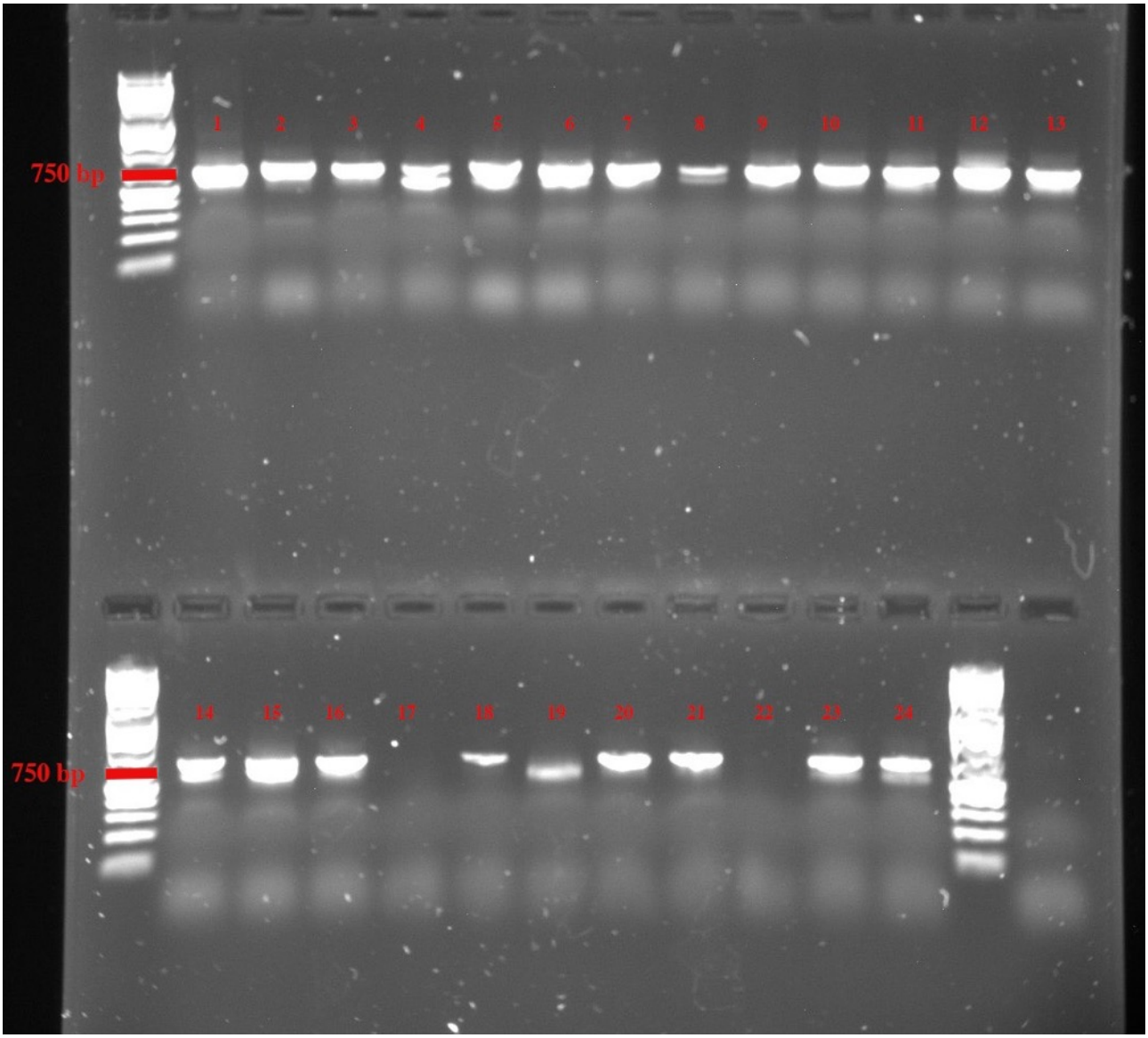


Figure 5 : Gel electrophoresis of 1 μL of ITS PCR products

Phylogeny

As the distance matrix proved a good isolation of the Anacyclus genus among the other genera of the Asteraceae family (fig 6), the analysis of the phylogeny was carried on.

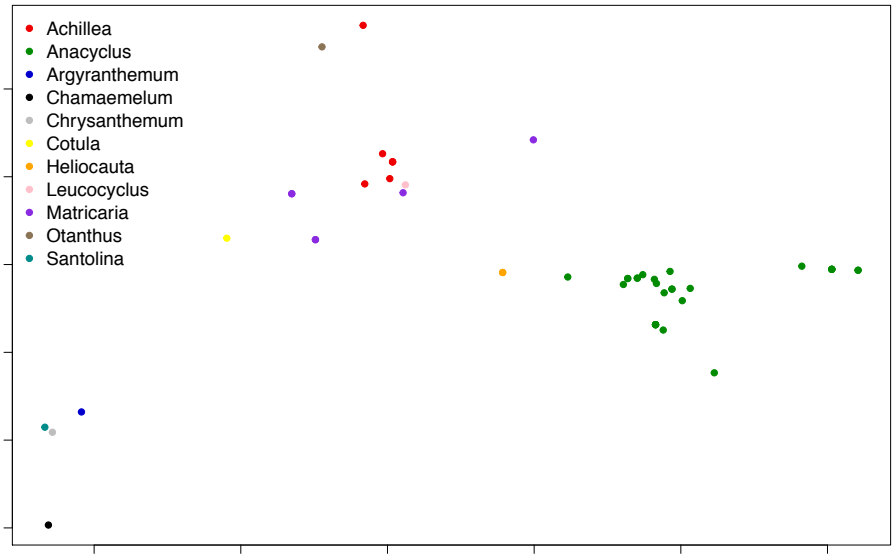


Figure 6 : Multidimensional scaling (MDS) of the pairwise distance matrix using the Jukes and Kantor (JC69) DNA evolution model for 11 Asteraceae genera representing 51 samples.

The SYM+G maximum likelihood model was chosen, based on the lowest AIC. To root the tree, the closest genera to Anacyclus were added to the phylogeny.

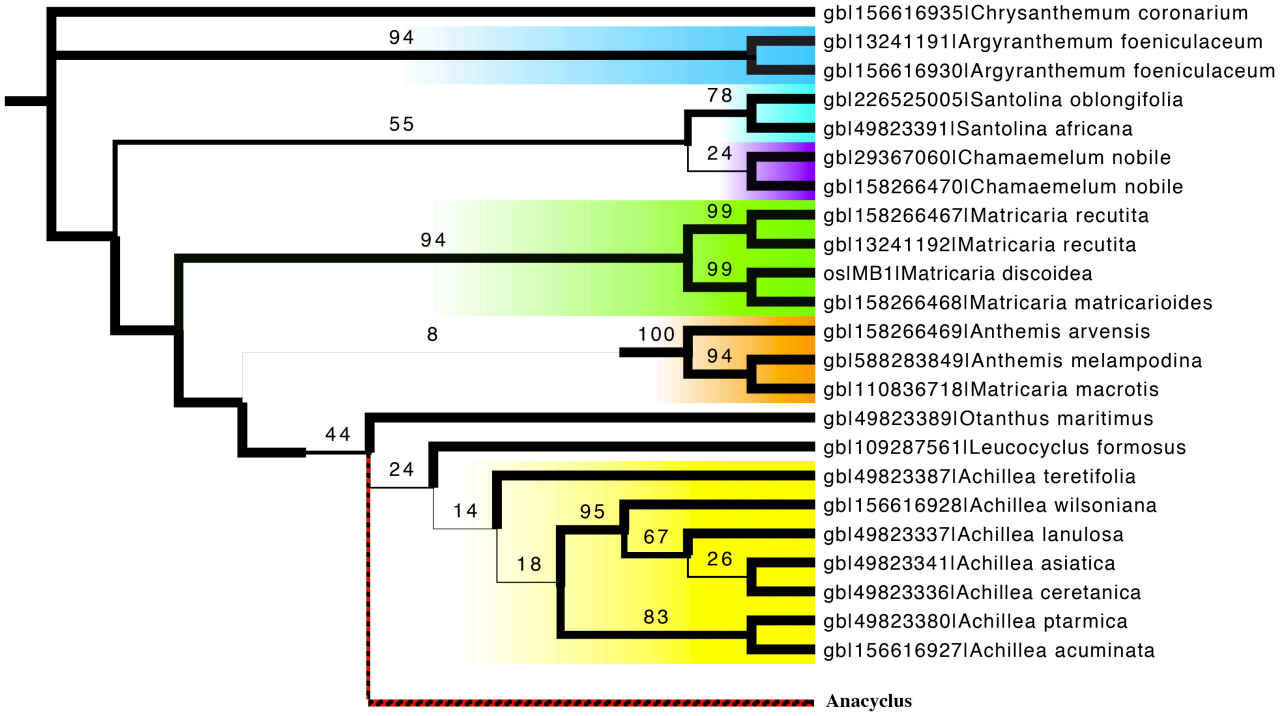


Figure 7a : Root of the Phylogenetic tree constructed using the SYM+G model in maximum likelihood for 54 samples in 11 genera. Each color represents a clade, the number on the branches represents the non parametric bootstrap value. The branches thickness is proportional to the bootstrap value.

This phylogeny confirms the hypothesis of Oberprieler et. al. (2006) about the reclassification of *Matricaria macrotis* as *Anthemis macrotis*. All of the different species are correctly placed in their respective genera clades. For the clarity of the phylogenetic tree, the root of the tree (fig 7a) is presented separately from the rest of the tree, presenting the *Anacyclus* genus clade (fig 7b).

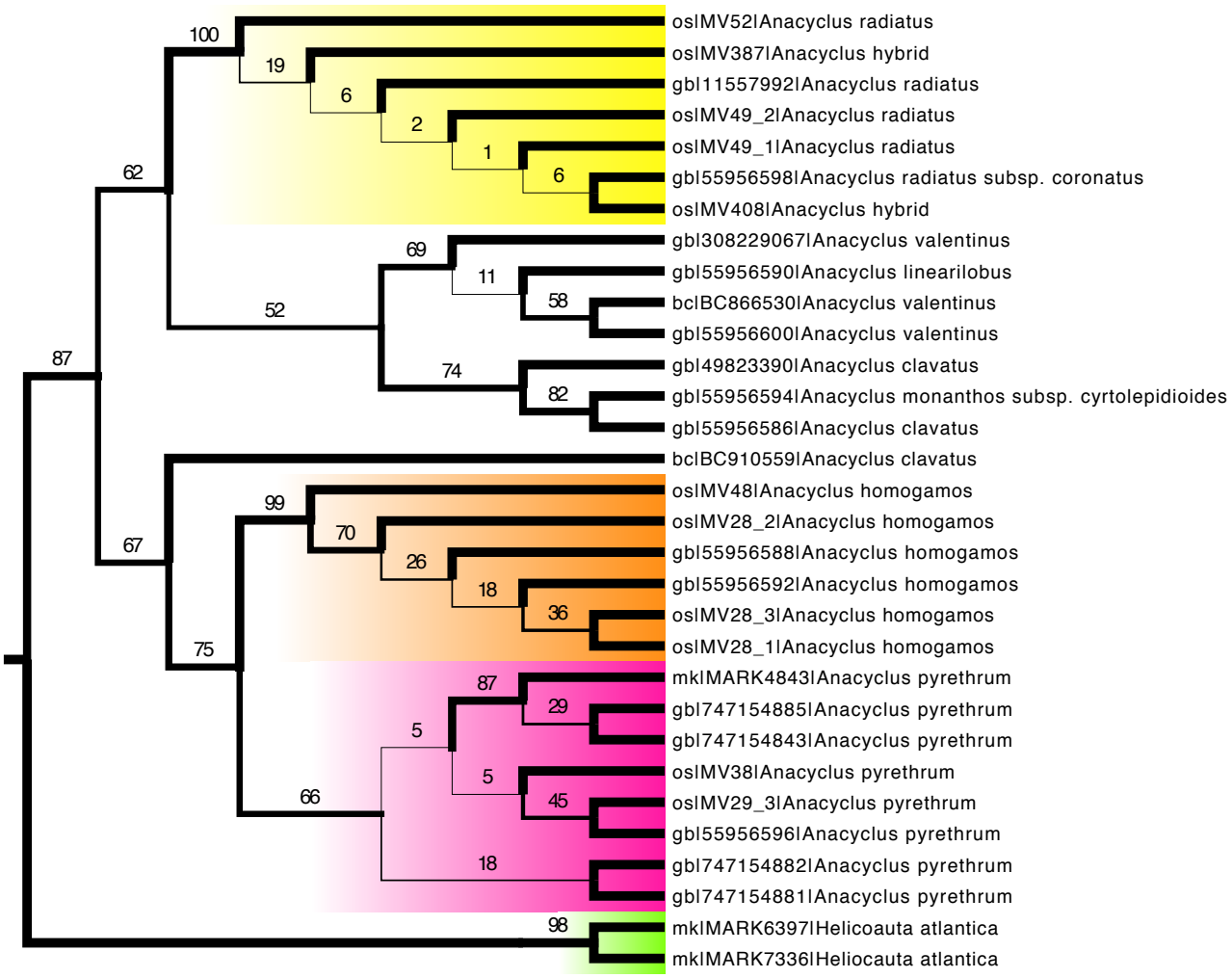


Figure 7b : Second half of the Phylogenetic tree, representing the Anacyclus clade. The whole tree is constructed using the SYM+G model in maximum likelihood for 54 samples in 11 genera. Each color represents a clade, the number on the branches represents the non parametric bootstrap value. The branches thickness is proportional to the bootstrap value.

Most of the Anacyclus species are distributed in clades. Only three species, are not placed in their own supported clades.

Heliocauta atlantica, for a very long time misclassified as an *Anacyclus*, is the only representative of the *Heliocauta* genus, and is the sister group of the *Anacyclus* genus.

Heatmap visualization

The output of the metabarcoding visualization sofware if presented in figure 8.

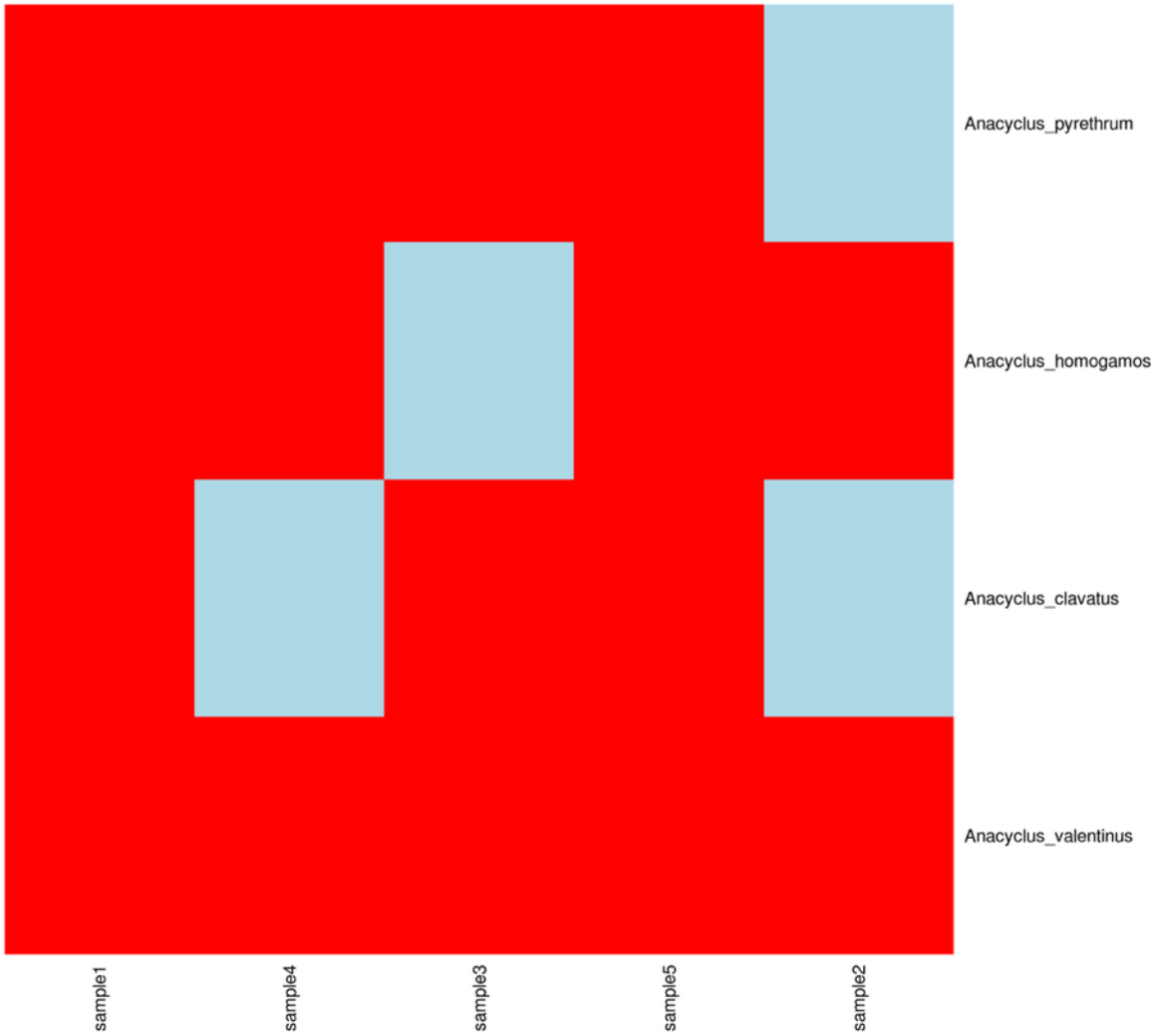


Figure 8 : Presence-absence heatmap of Anacyclus species. A red square signifies the presence of the specie in the sample, a blue square is the absence of the specie.

For each sample, the heatmap visualization software output is correctly matching and representing the composition of the original virtual sample.

Discussion

With the increasing number of commercial exchanges of Anacyclus roots (Moroccan « Office des changes »), one needs to be able to correctly identify the traded samples, especially in the light of the recent reclassification of *A. pyrethrum* as an endangered specie (Rankou and et al. unpublished). Barcoding is one way to approach is issue, and to track the trade of CITES (the Convention on International Trade in Endangered Species of Wild Fauna and Flora) and IUCN red listed species.

However, the phylogeny presented here shows a few weaknesses of the ITS marker : for three of the Anacyclus species, *A. valentinus*, *A. linearilobus*, and *A. clavatus*, the phylogenetic analyses were not able to separate them on a molecular basis, and therefore their clades are not correctly resolved here. One possible solution to this problem would be to use a combination of different maker to allow differentiation of those very closely related species. However, for the main specie of medicinal interest, *Anacyclus pyrethrum,* the phylogeny shows a clear separation from the other species, allowing the usage of ITS marker for its identification in metabarcoding applications.

On the two new putative hybrids, one can observe, according to the morphological characters, and the phylogenetic analysis results, that one of the parents is *A. radiatus*. The ITS copy has been captured from *A. radiatus*. And according to the morphological observations, the other parent might be *A. pyrethrum* (unpublished data). The hybridization events in the Asteraceae family are common, and they main explain the important observed phenotypic plasticity. However it would be necessary to test the influence of epigenetic factors as well to conclude anything on hybridization and plasticity.
More sequencing should be done to improve the bootstrap value and thus support the phylogeny, . Missing *Anacyclus* species (because of their field inaccessibility for endemic species, or sequencing failures) should also be included in further phylogenetic studies. More representative samples per specie from the different origins should eventually be included to take into account the geographic distribution of the genus.

Last but not least, regarding the heatmap visualization software, the next step will be the implementation of PCR differential amplification correction softwares and solutions to allow a semi-quantitative output.

Acknowledgments

First of all I'd like to thank Vincent for his warm welcome, his help, the review of this report, and his friendship. Thanks a lot also to Sugir for being my host and my guide in my first days in Oslo, and for the help with the alignments.
Thanks also to Audun, Anca and Madelina for their support in the lab.
Last but not least, thanks a lot to Hugo and Anneleen for welcoming me in at the museum in their lab, and helping me deal with the norwegian administration.

References

- Annalakshmi, R., Uma, R., Chandran, G. S., & Muneeswaran, A. (2012). A treasure of medicinal herb-Anacyclus pyrethrum A review. Indian Journal of Drugs and Diseases, 1(3), 59-67.

- Badhe, S. R., Badhe, R. V., Ghaisas, M. M., Chopade, V. V., & Deshpande, A. D. (2010). Evaluations of antidepressant activity of Anacyclus pyrethrum root extract. International Journal of Green Pharmacy, 4(2), 79.

- Baldwin, B. G. (1992). Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the compositae. Molecular Phylogenetics and Evolution, 1(1), 3–16. http://doi.org/10.1016/1055-7903(92)90030-K

- Bello, M. A., Álvarez, I., Torices, R., & Fuertes-Aguilar, J. (2013). Floral development and evolution of capitulum structure in Anacyclus (Anthemideae, Asteraceae). Annals of Botany, mcs301. http://doi.org/10.1093/aob/mcs301

- Bendjeddou, D., Lalaoui, K., & Satta, D. (2003). Immunostimulating activity of the hot water-soluble polysaccharide extracts of Anacyclus pyrethrum, Alpinia galanga and Citrullus colocynthis. Journal of ethnopharmacology, 88(2), 155-160.

- Berdan, F. F. (1982). The Aztecs of central Mexico: An imperial society.

- Bisby FA, Roskov YR, Orrell TM, Nicolson D, Paglinawan LE, Bailly N, Kirk PM, Bourgoin T, Baillargeon G., eds (2010). Species 2000 & ITIS Catalogue of Life: 2010 Annual Checklist. Digital resource at http://www.catalogueoflife.org/annual-checklist/2010. Species 2000: Reading, UK.

- Booker, A., Johnston, D., & Heinrich, M. (2012). Value chains of herbal medicines—Research needs and key challenges in the context of ethnopharmacology. Journal of Ethnopharmacology, 140(3), 624–633. http://doi.org/10.1016/j.jep.2012.01.039

- Garnock-Jones, P. J., Timmerman, G. M., & Wagstaff, S. J. (1996). Unknown New Zealand angiosperm assigned toCunoniaceae using sequence of the chloroplastrbcL gene. Plant Systematics and Evolution, 202(3-4), 211–218. http://doi.org/10.1007/BF00983383

- Huffman, M. A., & Hirata, S. (2004). An experimental study of leaf swallowing in captive chimpanzees: insights into the origin of a self-medicative behavior and the role of social learning. Primates, 45(2), 113-118.

- Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and Using a

Plant DNA Barcode. PLoS ONE, 6(5), e19254. http://doi.org/10.1371/journal.pone.0019254

- Kool, A., de Boer, H. J., Krüger, Å., Rydberg, A., Abbad, A., Björk, L., & Martin, G. (2012). Molecular Identification of Commercialized Medicinal Plants in Southern Morocco. PLoS ONE, 7(6), e39459. http://doi.org/10.1371/journal.pone.0039459

- Nunn, J. F. (2002). Ancient egyptian medicine. University of Oklahoma Press.

- Oberprieler, C. (2004). On the Taxonomic Status and the Phylogenetic Relationships of Some Unispecific Mediterranean Genera of Compositae-Anthemideae I. Brocchia, Endopappus and Heliocauta. Willdenowia, 34(1), 39–57.

- Oberprieler, C., & Vogt, R. (2006). The Taxonomic Position of Matricaria macrotis (Compositae-Anthemideae). Willdenowia, 36(1), 329–338.

- Patel, V. K., Patel, R. V., Venkatakrishna-Bhatt, H., Gopalakrishna, G., & Devasankariah, G. (1992). A clinical appraisal of Anacyclus pyrethrum root extract in dental patients. Phytotherapy Research, 6(3), 158-159.

- Sharma, V., Thakur, M., Chauhan, N. S., & Dixit, V. K. (2009). Evaluation of the anabolic, aphrodisiac and reproductive activity of Anacyclus pyrethrum DC in male rats. Sci Pharm, 77, 97-110.

- Sujith, K., Darwin, C. R., & Suba, V. (2012). Memory-enhancing activity of Anacyclus pyrethrum in albino Wistar rats. Asian pacific journal of tropical disease, 2(4), 307-311.

- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research, 35(3), e14. http://doi.org/10.1093/nar/gkl938

- T. J. White, T. D. B. (1989). In Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. Sninsky J and White T. San Diego: CA Academic Press, 38, 315–322. http://doi.org/10.1016/B978-0-12-372180-8.50042-1