Phylogenetic diversity and molecular evolution of nuclear DNA of the genus *Campylopus*

Research report

Iris Tabak Hz University of Applied Sciences 6-6-2018



Phylogenetic diversity and molecular evolution of nuclear DNA of the genus *Campylopus*

Research report

Institute: HZ University of Applied Sciences Company: Naturalis Biodiversity Center Supervisor: Michael Stech Name: I. Tabak, 00068267 Deadline: 06-06-2018 Date: 06-06-2018 Version: 1.0

Abstract

Recent research has shown that many chloroplast markers commonly used for the identification of species in land plants are not useful in the moss genus Campylopus. The only known suitable marker in Campylopus is the nuclear ribosomal ITS1-5.8S-ITS2 region. However, evolutionary patterns of this region seem to be more complicated in *Campylopus* than in other genera, since the internal transcribed spacers ITS1 and ITS2, and possibly also parts within each spacer, may evolve separately. To improve species delimitation in Campylopus, the aims of this study were to (i) identify several Campylopus specimens from Asia based on ITS sequences (ii) compare phylogenetic patterns between different helices of the secondary structures of the ITS and (iii) develop and test new chloroplast primers. The DNA was isolated using three methods: Kit extraction, CTAB extraction and Kingfisher robot extraction. The extracted DNA of the kit and CTAB extraction was amplified using primers M13-ITS2-5.8F/25R, for the ITS2 region, to determine if these methods showed different results. A 96-well plate was extracted using the Kingfisher robot, these samples were amplified with primers M13-ITS-18F/25R for the complete ITS region. The successfully amplified samples were Sanger sequenced, edited and aligned in an already existing alignment containing various Campylopus species. This alignment was used for the identification of several specimens with the complete ITS region and the analysis of the secondary structure of the ITS2 region. The phylogenetic analysis, based on Bayesian Inference, showed that the identification for most species is difficult with only the information of ITS. After phylogenetic analysis of the four main helices of the secondary structure of ITS2, it was concluded that there is no incongruence between them. For designing new chloroplast primers, the genome of Syntrichia ruralis was used. In total 5 primer sets were designed and tested. These primer sets all showed non-specific bands, even after extended testing, resulting in no new chloroplast primers for the genus Campylopus.

Table of Contents

1. Introduction	1
2. Theoretical framework	2
2.1 Bryophytes	2
2.1.1 Life cycle	2
2.2 Bryophyta (mosses)	3
2.3 DNA barcoding	3
2.4 Internal transcribed spacer (ITS)	4
2.5 Secondary structures of ITS	5
2.6 Chloroplast DNA and primer design	6
2.7 Phylogenetic analysis	6
2.7.1 Bayesian inference	6
3. Research method	7
3.1 Sampling	7
3.2 DNA subsampling	7
3.3 DNA extraction	7
3.4 PCR additive test	10
3.5 PCR amplification of the ITS1-5.8S-ITS2 region	11
3.6 Sequencing and sequence analysis	12
3.7 Phylogenetic analyses	13
3.8 Designing and testing new chloroplast primers	14
4. Results	16
4.1 Kit extraction and CTAB extraction and amplification	16
4.2 PCR testing and optimization	17
KingFisher robot extraction and amplification	19
4.3 Sequencing and sequence analysis	20
4.4 Phylogenetic analysis of ITS	21
4.5 Phylogenetic analysis of the secondary structures of ITS2	24
4.5 Chloroplast DNA primer design	25
6. Discussion	28
5. Conclusion	31
Bibliography	

Appendix 1. Sample overview	35
Appendix 2. DNA extraction protocols	
2.1 NucleoSpin plant II extraction	
2.2 CTAB extraction	
2.3 KingFisher extraction robot	45
Appendix 3. Geneious protocol	
Appendix 4. Gel analysis of CTAB and Kit extraction	66
Appendix 5. Standard PCR program ITS2	67
Appendix 6. PCR test with different additives	68
Appendix 7. Phylogenetic trees of helices ITS2	71
Appendix 8. Phylogenetic tree of ITS2	75
Appendix 9. PCR results Hot start PCR	76

1. Introduction

The thesis focused on three research topics in the moss genus *Campylopus*; the identification of *Campylopus* specimens from Asia, the molecular evolution of the Internal Transcribed Spacer (ITS) region, and the development and testing of new chloroplast primers. For these topics, DNA barcoding and phylogenetic analysis were used.

Bryophytes is a collective term for liverworts (Marchantiophyta), mosses (Bryophyta) and hornworts (Anthocerotophyta). Bryophytes are classified as non-vascular land plants and are characterized by a life cycle that alternates between haploid and diploid generations with a dominant gametophyte. This is one of the characteristics that distinguishes them from vascular plants (Vanderpoorten & Goffinet, 2009). The phylum Bryophyta contains approximately 13,000 species, making this phylum the second most diverse of land plants (Goffinet & Shaw, 2008). One of the genera of Bryophyta is *Campylopus*, which now has approximately 150 species (Stech, Sim-Sim, & Kruijer, 2010), with highest species diversity in South America and Asia.

In recent studies it has been shown that molecular markers of the chloroplast genome that are regularly used to distinguish between species in mosses are not variable enough in *Campylopus*. At this moment, only the sequences of the ITS1-5.8S-ITS2 region are available to infer species delimitations and relationships in *Campylopus*. The evolution of the ITS region, however, seems to be more complicated in *Campylopus* compared to other mosses, as the internal transcribed spacers (ITS) 1 and 2 seem to evolve separately, opposed to the generally recognized principle of the 'concerted evolution'. Whether the phylogenetic signal also differs between the different stem-loop regions of the secondary structure within ITS2 remains to be investigated, and new primers for other chloroplast regions as potential DNA barcodes for the genus *Campylopus* need to be tested.

Three main questions were formulated for this project:

"Can Campylopus specimens from Asia be identified based on ITS sequences?"

"Do different parts of ITS2 evolve separately, contrary to concerted evolution?"

"Which chloroplast markers are potentially useful for the identification of species in the genus *Campylopus*?"

To answer these questions, DNA was isolated from moss samples, followed by PCR amplification and Sanger sequencing. These sequences were added to an already existing *Campylopus* dataset comprising mostly samples from South America. The phylogenetic and molecular evolution analysis of the ITS2 region was performed on a selection of samples from the complete dataset. Furthermore, with recently obtained information about the chloroplast genome of *Syntrichia ruralis*, new primers were developed and tested.

2. Theoretical framework

2.1 Bryophytes

Bryophytes is the collective term for mosses, liverworts, and hornworts, which are non-vascular land plants. Even though they all share their small size and independent, dominant gametophyte and dependent sporophyte (see 2.1.1 for the complete life cycle), molecular phylogenetic reconstruction has shown that bryophytes are paraphyletic. Consequently the three main lineages of bryophytes have been separated as phyla, leaving the mosses in the phylum Bryophyta. Looking at all plant groups, the bryophytes are the second most species-rich lineage in the world, ranging from 15,000 to 25,000 species, only exceeded by the flowering plants angiosperms, with approximately 350,000 species. Bryophytes can grow on every continent and in every location which is habitable for plants that rely on photosynthesis (Glime, 2006).

2.1.1 Life cycle

The life cycle of all land plants, including bryophytes and vascular plants, involves the alternation of two distinct multicellular generations, the gametophyte and the sporophyte. For mosses, the gametophyte is the dominant generation (Vanderpoorten & Goffinet, 2009; Graham, 1985). The gametophyte is haploid, meaning that it only has a single set of chromosomes in each nucleus. The sporophyte is diploid, having two copies of chromosomes in each nucleus, and produces spores by meiosis, which is a type of cell division that divides the number of chromosomes in two. The haploid spores germinate into gametophytes, which then produces either male gametes, also called spermatozoids, or female gametes, also called eggs. These male and female gametes combine sexually and grow into new sporophytes, completing the cycle. See figure 1 for an overview of the life cycle of a moss (Graham, 1985).



Figure 1. Diagram of the life cycle of mosses (Graham, 1985).

2.2 Bryophyta (mosses)

Mosses are present in almost every ecosystem, for example, in some places in the world where there are short growing seasons, which limit plant growth, mosses may be dominantly present. Similarly they are found in temperate and tropical rainforests, where they contribute to important ecological functions (Vanderpoorten & Goffinet, 2009). Currently, approximately 12,000 (Vanderpoorten & Goffinet, 2009) to 13,000 (Goffinet & Shaw, 2008) species are recognized, many showing a broad morphological diversity. *Campylopus* has always been considered one of the most species-rich and taxonomically most complex moss genera. It originally comprised 1,000 species, which was brought down to approximately 150 species over a period of almost four decades after taxonomic revisions, by Jan-Peter Frahm (Stech, Sim-Sim, & Kruijer, 2010).

2.3 DNA barcoding

Biological specimens used to be only identified using their morphological features, like their shape, size, and color. In most cases experienced professional taxonomists are needed to do this because of the morphological variations. But even for specialists it can sometimes be almost impossible to identify specimens, for example, if they are damaged or in an immature stage of development. Molecular species identification by DNA barcoding solves these problems (Parmentier, et al., 2013). The concept of DNA barcoding was put in motion by professor Paul Hebert and others at the University of Guelph in Canada in 2003. DNA barcodes are short sequences of DNA, usually between 400 and 800 base pairs, which are specific for each species (Ajmal Ali, et al., 2014; Kress & Erickson, 2008). An ideal DNA barcode should be easy to amplify using polymerase chain reaction (PCR) and should be able to identify all species on the planet with a single primer pair (CBOL Plant Working Group, 2009). For the identification of animals this approach has been very effective using the mitochondrial marker cytochrome oxidase I (CO1). However, this gene does not have the required variability in plants. Due to this, alternative barcoding markers have to be considered. Commonly used markers for the identification of vascular plant species are the chloroplast genes rbcL and matK (Bock & Norris, 2016). However, according to research by the CBOL Plant Working group (2009) it is sometimes not possible to identify to species level, but only to genus level and in some cases only to family level with only one of these markers. The best discrimination was achieved by using both markers together, and some researchers even suggested to use a third marker (trnH-psbA) (Bock & Norris, 2016; CBOL Plant Working Group, 2009). But the majority of the researchers preferred to use a two-gene barcode to avoid the extra cost of sequencing and prevent further delays in implementing a standard barcode for land plants (CBOL Plant Working Group, 2009). The use of DNA barcoding for species identification relies on the expectation that there will be a big online digital library of barcodes that will serve as a standard where unidentified sequences can be matched with. If this is the case, non-taxonomy specialists can send samples to the laboratory where the sample can get sequenced. These samples can then be identified by matching them to the databases (Parmentier, et al., 2013), for example, BOLD (Barcode of Life Database) and GenBank.

However, for bryophytes DNA barcoding has been proven to be a bit more difficult. Recently proposed plastid markers for land plants either tend to be short (*psbA*-t*rn*H spacer), have a lower discrimination capacity at species level (*rbcL*) or need more study regarding the primer design and amplification process (*trnK/matK*) for bryophytes. Even though the optimal combination of the barcoding markers for

bryophytes is not clear yet, several markers have already shown to be useful for species identification in bryophytes (Lang, Kruijer, & Stech, 2014; Stech, et al., 2013), one of which is the nuclear ribosomal ITS region.

2.4 Internal transcribed spacer (ITS)

The ribosomal RNA genes in an eukaryotic cell form a cluster (Kobayashi, 2014). This cluster comprises seven components: 1) the 5' external transcribed spacer, 2) the small subunit 18S gene, 3) the internal transcribed spacer 1 (ITS1), 4) the large subunit 5.8S gene, 5) the internal transcribed spacer 2 (ITS2), 6) the large subunit 26S gene, and 7) the 3'external transcribed spacer (Edger, et al., 2014). See figure 2 for the overview of the cluster.

The part of the ribosomal RNA cluster most frequently used as a molecular marker in phylogenetic analyses at genus and species level is the ITS region. The ITS region is made up of three parts: the ITS1 and ITS2, which separates the genes of the rDNA locus, and the 5.8S gene (figure 2). The length of this region varies between 500 and 750 bp in angiosperms and can go up to 3500 bp in other seed plants (Poczai & Hyvönen, 2009). The ribosomal subunit 5.8S in eukaryotes is highly conserved, while the ITS regions can vary in length caused by indels (Edger, et al., 2014). In comparison to coding regions, spacers evolve more quickly, making them ideal for phylogenetic analysis (Poczai & Hyvönen, 2009).



Figure 2. An illustration of the nucleus organizing region (shown as red on the chromosome) containing the rDNA clusters, whereas each cluster consists of seven main components (Edger, et al., 2014).

The internal transcribed spacers (ITS1 and ITS2) of the nuclear ribosomal RNA gene cluster are the most commonly used nuclear markers for phylogenetic analysis for many eukaryotic groups including most plant families. The reasons for their popularity include that these markers are easy to amplify due to having a high copy number of the gene clusters, they are rapidly evolving markers, meaning there are many variations between closely related species, there are cost-effective methods, and highly conserved primers available (Edger, et al., 2014).

However, there are also disadvantages of using the nuclear ITS region as a barcode. ITS shows not only inherent problems such as the possible presence of paralogous ITS copies or incongruence between ITS1 and ITS2, it also shows difficulty with the amplification of the complete ITS1-5.8S-ITS2 region (Stech, et al., 2013), especially in older plant material with degraded DNA (Telle & Thines, 2008). This, however,

can usually be solved by amplifying ITS1 and ITS2 individually. Up till now, mostly ITS2 has been considered as a DNA barcode marker for plants. However recent studies showed that ITS1, which is generally more variable than ITS2, should also be considered as a DNA barcode marker, especially for closely related species of mosses (Stech, et al., 2013).

2.5 Secondary structures of ITS

The ribosomal RNA cluster as shown in figure 2 is transcribed as a single primary transcript (Mai & Coleman, 1997). The primary transcript is processed to produce mature 18S, 5.8S, and 26S RNA that act as parts of the cytoplasmic ribosomes. During this splicing process, the ITS1 and ITS2 are thus cut out. The maturation and splicing process depends on the secondary structures of ITS1 and ITS2, which implies that the ITS sequences are conserved at some level (Giudicelli, et al., 2017).

The fact that rRNA single stranded chains form secondary structures comprising stem regions and loops correlating with base pairing, opened a new field to analyze phylogenies. The significance of this progress is that the inference of phylogenies would not only be based on the primary sequence information but also on the predicted secondary structures. Being able to predict the secondary structures can improve the sequence alignment, identify non-independent mutations necessary to maintain the structure, (compensatory base-pair changes), and identify parts with different mutation rates (Poczai & Hyvönen, 2009).

Recent studies showed that all examined eukaryote groups shared the same general secondary structure for ITS2 (Joseph, Krauskopf, Vera, & Michot, 1999; Mai & Coleman, 1997). It was concluded that the secondary structure of ITS2 consists of four helixes, where, among plants, the nucleotide sequence evolved the fastest in helix 4, followed by helix 1, see figure 3 for a general prediction of ITS2. The prediction of the secondary structure of ITS1 is harder than of the prediction of ITS2. It seems like ITS1 evolves faster and has less conservative sites than ITS2 (Poczai & Hyvönen, 2009).



Figure 3. Secondary structure of ITS2 (Mai & Coleman, 1997).

2.6 Chloroplast DNA and primer design

At this moment there are no chloroplast primers that are variable enough to distinguish between species for the genus *Campylopus*. Therefore, research has to be done to evaluate new chloroplast marker regions. To be able to develop well-working primers, a few rules are needed: primer length is optimal between 18–22bp, the melting temperature (Tm) in the range of 52–58 °C generally works best, GC content should be between 40–60% and the hairpin Tm and pair dimer Tm should be zero. When looking at a chloroplast genome, the primer sites should be in the conserved regions and in between the primer sites there should be enough variation.

2.7 Phylogenetic analysis

Phylogeny refers to the evolutionary history of species; an important area of focus for molecular phylogenetics, which makes use of genetic sequences to determine or estimate the evolutionary relationships. The relationship obtained from this analysis is usually depicted as a tree-like diagram, also known as a phylogenetic tree (Choudhuri, 2014). A phylogenetic tree is a diagram that represents the lines of evolutionary descent of the living species from their (hypothetical) common ancestors. These trees are useful for organizing knowledge of biological information, for structuring classifications and for providing insight into events that happened during evolution. Usually, phylogenetic trees only show the branching history of common ancestry, unless this is indicated otherwise. The topology, which is the branching pattern, is what is most important. The length of the branches are in some cases irrelevant but usually the branch lengths are meaningful. These trees generally represent either the amount of evolutionary changes in that particular gene sequence or the estimated duration of the evolution of the branches (Baum, 2008). Trees can be oriented in all different ways and yet present the same information. For example, the trees in figure 4 all have the same topology and thus present the same evolutionary relationships (Baum, 2008; McLennan, 2010). They all show sister group relationships, meaning that C and D are sister species, and their ancestor is sister to B. Similarly, the general ancestor of B, C and D is sister to A.



Figure 4. Phylogenetic trees with similar topology, yet different structure (Baum, 2008).

2.7.1 Bayesian inference

Bayesian inference of phylogeny is based on a quantity called the posterior probability of a tree. The posterior probability of a tree can be seen as the probability that the tree is correct in the light of the data and a prior hypothesis. This can be calculated using the Bayes theorem, which is used to combine the prior probability with the likelihood to produce the posterior probability distribution. For example, if there are multiple trees, the one with the highest probability is most likely to be chosen as the best estimate (Huelsenbeck J. P., Ronquist, Nielsen, & Bollback, 2001).

3. Research method

3.1 Sampling

In total, 49 samples morphologically identified as *Campylopus* were analyzed in this project (see appendix 1). However, to be cost efficient, the 96-well extraction plate was filled up with other moss samples, which are not included in this report. Most *Campylopus* samples were collected in Asia. Some of the samples were herbarium samples collected over a period ranging from 1984 to 1999 by several collectors, others were collected in 2011 by Eka Iskandar, a PhD student at Naturalis Biodiversity Center. For a complete sample overview see appendix 1. Sample overview. Because of the extensive research that has been done to the genus *Campylopus*, an alignment of *Campylopus* species was already available. The samples collected in Asia were added to this dataset.

3.2 DNA subsampling

Before the DNA was extracted from the samples, they were subjected to a clean-up procedure. First, one stem of moss was taken and placed into a small flask containing demi water. The flask containing the sample was shaken a few times in order to get rid of most dirt and other contaminants. Then the sample was placed onto a microscopic slide with a drop of demineralized water. This was placed under the binocular and while looking through the binocular, the leaves of the moss were pulled off and cleaned with more demi water. This was done because sometimes there can be dirt, spores or other contaminants on or between the leaves which would result in contamination. After cleaning, the leaves were dried with a tissue paper and placed in a small paper envelope. These envelopes were placed in an air sealed bag containing silica, to keep the samples dry.

3.3 DNA extraction

For the DNA extraction, three methods were used, a kit extraction, CTAB extraction, and KingFisher robot extraction. Initially the kit and CTAB extraction were used to see if the DNA could be extracted and if so, which method showed the most successful extraction. The first extraction that was performed, was the kit extraction following the NucleoSpin plant II protocol, of which a detailed protocol can be found in appendix 2.1. For this method, six samples were used, all varying in age. Since there were several older samples and some samples that were yellow, which makes it harder to have a high yield of extracted DNA with a kit extraction, a CTAB extraction was performed. The full protocol can be found in appendix 2.2. The extraction was performed again with 6 samples, only not with the same samples that had been used for the kit extraction because there was not enough sample left. The kit and CTAB extraction methods are both very time consuming, especially when needing to extract a lot of samples. Therefore, a KingFisher robot extraction was performed on 95 samples. For the full detailed protocol see appendix 2.3.

3.3.1 NucleoSpin Plant II extraction, Macherey Nagel extraction

First, the samples were transferred from the paper envelops to 2.0 ml Eppendorf tubes containing one 7 mm glass bead. The tubes were then placed in liquid nitrogen to make it easier to crush the material. Then the samples were placed in the Qiagen Tissuelyser for 90 sec at a speed of 25 oscillations/s. After all the samples were homogenized, 400 µl lysis buffer and 10 µl RNase were added and the tubes were vortexed thoroughly. This was incubated for 10 minutes at 65 °C. After incubation, a NucleoSpin filter was placed into a new 2 ml collection tube and the lysate was loaded onto this column. The tube was centrifuged for 2 minutes at 11,000 rcf. The flow-through, now containing the DNA, was collected and the filter, containing impurities, was discarded. To the flow-through, 450 µl binding buffer was added and this was mixed by pipetting up and down (approximately 5 times). Then the lysate was loaded onto a NucleoSpin Plant II Column, which was placed into a new 2 ml collection tube. This was centrifuged for 1 minute at 11,000 rcf, the flow-through was then discarded. Then 400 µl of wash buffer was added to the column, this was centrifuged again for 1 minute at 11,000 rcf. The flow-through was discarded again and then 600 µl of wash buffer 2 was added to the column. This was centrifuged again for 1 minute at 11,000 rcf. The flow-through was discarded again and then 200 μ l of wash buffer 2 was added to the column. This was centrifuged for 2 minutes at 11,000 rcf to make sure the wash buffer was completely removed. The NucleoSpin Plant II Column was then placed into a new 1.5 ml Eppendorf microcentrifuge tube. Then 50 μ l of elution buffer (pre-heated to 65 °C) was added onto the membrane. The tubes were incubated for 5 minutes at 65 °C and were then centrifuged for 1 minutes at 11,000 rcf to elute the DNA. This step was performed again for a higher yield of DNA.

3.3.2 CTAB extraction

The crushing of the samples was performed in the same way as the kit extraction. From this step until the end of the extraction everything is handled under the fume hood. After the samples were crushed, 1 ml of CTAB buffer was added. The tubes were then incubated for 1.5 hours at 65 °C on a shakeincubator and were inverted every 15 minutes. After incubation, 450 µl of chloroform-isoamyl alcohol (Sigma-Aldrich) was added and the tubes were centrifuged for 10 minutes at 20,000 rcf. After centrifugation, three layers were visible: The bottom layer, the organic phase, containing the chloroform with the dissolved proteins, lipids and chlorophyll. The middle layer, the interphase, containing the bipolar proteins and molecules and the dissolved green cell residue. The top layer, the water phase, containing the nucleic acids and the dissolved polysaccharides. Then 800 μ l of the water phase was taken and pipetted into a new 2.0 ml Eppendorf tube. To this tube, 450 μl of chloroform-isoamyl alcohol was added and was then inverted for approximately 5 minutes. The tubes were then centrifuged for 10 minutes at 20,000 rcf. Then 550 µl of the water phase was pipetted into a new 2.0 ml Eppendorf tube. To this tube, 550 µl of cold isopropyl alcohol (Sigma-Aldrich) was added and was then inverted for approximately 5 minutes. One sample was left overnight to see if a longer incubation would make a difference in yield. The tubes were then centrifuged for 10 minutes at 8000 rcf. The isopropyl alcohol was pipetted off and the tubes were placed upside down on a tissue paper. The pellet was dissolved in 150 µl TE buffer and incubated for 30 minutes at 37 °C. Then 3 µl RNase was added and the tubes were incubated for 60 minutes at 37 °C. The samples were now ready for use.

3.3.3 KingFisher robot extraction

To a 96 well plate, with Qiagen Collection microtubes, three 3 mm glass beads were added to each tube. Then the dried moss sample was added to the defined microtube. Once one column of 8 samples was filled up, the tubes were closed off with an 8-strip cap. After the plate was completely filled, it was placed in liquid nitrogen and then the samples were crushed using the Qiagen Tissuelyser for 90 seconds at 25 oscillations/s. This step was performed again but then with the plate in the opposite direction. After all the samples had been crushed, 500 μ l of lysis buffer and 10 μ l of RNase were added to each tube. The tubes were then closed and the plate was sealed with tape to make sure the caps would not come off. The plate was then vortexed and centrifuged for 30 seconds at 3700 rpm. The plate was incubated on a pre-heated shake-incubator for 1.5 hours at 56 °C. After incubation, the plate was centrifuged for 15 minutes at 3700 rpm.

During incubation, the plates for the robot were prepared. In total, 6 plates had to be prepared, 5 deep well plates and 1 elution plate. See table 1 for an overview of the plates.

1. Lysis plate (deep well plate)		Wash plate (de	eep well plate)	6. Elution plate	
Lysed sample	400 μl	2. MC3	600 μl	MC6	150 μl
MC2 buffer	400 μl	3. MC4	600 μl		
C-beads	30 µl	4.EtOH 80%	600 μl		
		5. MC5	600 μl		

Table 1. Plate preparation for the KingFisher extraction robot.

After incubation, the samples were pipetted in the lysis plate with buffer and c-beads. The DNA binds to these beads and since the beads are magnetic they stuck to the metal robot arm. This arm moved then to the different plates for the washing steps. All the plates were placed in the extraction robot and the program "Machery_Nagel_Plant_96 KingFisher Flex" was used. Once the robot was finished, the elution plate was divided over a work plate and a stock plate: 15 μ l was pipetted in the work plate and 135 μ l in the stock plate. The stock plate was then stored in the -80 °C freezer.

3.4 PCR additive test

To determine the optimal conditions for the PCR, a few tests with different combinations of additives were performed. This was done as preparation for the PCR amplification of the full 96-well plate, so that no chemicals were wasted with a non-optimal PCR amplification. See table 2 for an overview of which additives and combinations were tested.

Table 2. Overview of the additives and combinations that were tested (- control stands for negative control). The o after 057 stands for overnight, which is the sample that was incubated overnight during the CTAB extraction, as explained in chapter 3.3.2.

		2	3		5	6		8	9	10
004	Betaine	Betaine	DMSO	DMSO	Betaine	Betaine	BSA	BSA	BSA	No
	DMSO	DMSO			BSA	BSA	MgCl ₂	1.0µl	0.5µl	additive
	MgCl ₂		MgCl ₂		MgCl ₂					
0570	Betaine	Betaine	DMSO	DMSO	Betaine	Betaine	BSA	BSA	BSA	No
	DMSO	DMSO			BSA	BSA	MgCl ₂	1.0µl	0.5µl	additive
	MgCl ₂		MgCl₂		MgCl ₂					
-control	Betaine	Betaine	DMSO	DMSO	Betaine	Betaine	BSA	BSA	BSA	No
	DMSO	DMSO			BSA	BSA	MgCl ₂	1.0µl	0.5µl	additive
	MgCl ₂		MgCl ₂		MgCl ₂					

In total 10 different master mixes were made, each for the same two samples, 004 and 0570, and one negative control (-control), which does not contain DNA and should therefore show no band on the gel. For these samples, ITS2 was amplified. The forward primer is called M13-ITS2-5.8F and the reverse is called M13-ITS2-25R. Both ITS primers have a M13 tail attached, this is used for sanger sequencing. See table 3 for the sequences of these primers.

Table 3. The primer sequence of the used primer set for the amplification of the ITS2 region. The M13 tail is marked in bold letters.

Primer	Target region	Direction	Sequence 5'-3'
M13-ITS2-5.8F	ITS	Forward	TGTAAAACGACGGCCAGTGCAACGATGAAGAACGCAGC
M13-ITS2-25R	ITS	Reverse	CAGGAAACAGCTATGACTCCTCCGCTTAGTGATATGC

The volume of PCR buffer, both primers, dNTP's and Taq polymerase remained the same, the only variables were the additives. Depending on the volume after every chemical was added to the mix, milli-Q water was added to get the master mix up to 20μ l. For an example of master mix 1 see table 4 or all the master mixes see appendix 6 'PCR test with different additives'.

Chemical	Brand	Concentration	Reaction (µl)
Milli-Q water	Ultrapure	-	10.8
PCR buffer	Qiagen	10X	2.0
Betaine	Sigma	5M	0.25
DMSO	Ultra	>99.5%	0.2
MgCl ₂	Qiagen	25 mM	1.0
Primer forward	IDT	10 pMol/μl	0.9
Primer reverse	IDT	10 pMol/µl	0.9
dNTP	Qiagen	2.5 mM	0.8
Таq	Qiagen	5 U/µl	0.2
DNA template	-	-	1.0

Table 4. Master mix 1 composition, used to test the additives.

For the PCR program, a program was used that had been tested before, the only adjustment was the annealing temperature which was raised to 65 °C, because this was the optimal annealing temperature for this primer set. See table 5 for the PCR program used for all different master mixes.

Table 5. The PCR program for testing the PCR additives.

Step	Temperature	Time
Initial Denaturation	95°C	0:05:00
Denaturation	95°C	0:00:20
Annealing	65°C	0:00:30
Extension	72°C	0:01:30
Final Extension	72°C	0:07:00
Pause	12°C	∞
Number of cycles	40	

3.5 PCR amplification of the ITS1-5.8S-ITS2 region

The full plate was amplified for the complete ITS region. The forward primer used for the amplification was M13-ITS-18F and the reverse primer M13-ITS-25R. In table 6 the primer sequence of the primer set of ITS is shown.

Table 6. The primer sequence of the used primer set for the amplification of the complete ITS region. The M13 tail is marked in bold letter.

Primer	Target region	Direction	Sequence 5'-3'
M13-ITS-18F	ITS	Forward	TGTAAAACGACGGCCAGTGGAAAGAGAAGTCGTAACAAGG
M13-ITS-25R	ITS	Reverse	CAGGAAACAGCTATGACTCCTCCGCTTAGTGATATGC

See table 7 for the master mix used and see table 8 for the PCR program used for the amplification of ITS.

Chemical	Brand	Concentration	Reaction (µl)
MQ	Ultrapure	-	14.2
PCR buffer	Qiagen	10X	2.0
Primer forward	IDT	10 pMol/μl	0.9
Primer reverse	IDT	10 pMol/μl	0.9
dNTP	Qiagen	2.5 mM	0.8
Таq	Qiagen	5 U/μl	0.2
DNA template	-	-	1.0

Table 7. Overview of master mix used for the amplification of ITS.

Table 8. PCR program of the amplification of ITS.

Step	Temperature	Time
Initial Denaturation	95°C	0:05:00
Denaturation	95°C	0:00:20
Annealing	57°C	0:00:30
Extension	72°C	0:01:00
Final Extension	72°C	0:07:00
Pause	12°C	~
Number of cycles	40	

After the amplification, the samples were analyzed using gel electrophoresis. In case of the full plate, a pre-made 2% agarose E-Gel (Invitrogen; Thermofisher Scientific) containing ethidium bromide was used. For this gel, no ladder was needed. The gel was loaded with 3 μ l sample and 17 μ l demi water. The gel ran for 12 minutes and then a photo was taken with a UV camera. In case of separate samples, a 1% agarose gel was made. Onto this gel 4 μ l sample and in each row 4 μ l of a GeneRuler 1 kb plus DNA ladder (Thermofisher Scientific) was loaded. The gel ran for 40 minutes at 100 V and was then stained in a 1 μ g/ml ethidium bromide (Sigma) solution for 20 minutes.

3.6 Sequencing and sequence analysis

The samples that showed one band at the desired height on the agarose gel were all placed together in a plate. This plate was sent to BaseClear for bidirectional Sanger sequencing using the M13 tails. When the results came back, the sequences were loaded in the software Geneious version 8.1.9 (Kearse, et al., 2012) for editing. The sequences came in two separate files per sample, one for the forward primer and one for the reverse primer, which were assembled using the De Novo Assemble option in Geneious. This formed contigs, meaning the forward and the reverse primer have overlapping regions which together creates a consensus sequence. Then the primer regions and excess were cut off, because these regions are not interesting for research as they may contain errors. Because the excess is just an artifact of the sequencing run and the primer region close to the end of the marker region has no overlapping primer sequences, which makes errors more likely to occur. When the primers were removed, the sequences were checked for gaps and ambiguities. An ambiguity means that the software could not read the signal and gave it a letter which stands for multiple bases, however, in several cases it could be decided which base was present and the respective position by visual inspection. Once the ambiguities were removed as far as possible, the sequences were aligned, using MAFFT multiple alignment in Geneious. Then it was checked whether the remaining ambiguities could be resolved by comparison with the other sequences in the alignment. For a detailed protocol and manual of Geneious see appendix 3. After aligning, the sequences were blasted against GenBank with the BLAST plug in in Geneious, to see if the sequenced material matched the original determination. See figure 5 for the parameters used for the BLAST.

0	Enter unformatted or EASTA sequence		
Database:	nr - GenBank+EMBL+DDBJ+PDB+RefSeq or	•	▼ Add/Remove Databases
Program:	Megablast - fast, high similarity matches (DNA	-	
Results:	Hit table	- ?]
Retrieve:	Matching region	•	
Maximum Hits:	5 🜩		

Figure 5. Example of the parameters used for the BLAST procedure.

After blasting, the *Campylopus* samples that matched the original identification were added to the existing alignment and manually aligned in Geneious, in order to keep the existing gaps and structure. The samples that did not match the identification were excluded from further research for this project.

3.7 Phylogenetic analyses

After the alignment of the complete ITS region was finished, the file was exported as a NEXUS file. The alignment was opened in seqstate version 1.4.1 where the indels were coded using simple indel coding (SIC) (Simmons & Ochoterena, 2000). Phylogenetic analysis was performed using Bayesian inference, with the software MrBayes version 3.2.6 (Huelsenbeck & Ronquist, 2001; Huelsenbeck & Ronquist, 2003) used on CIPRES Science gateway. The number of substitution types was set to 6 with the rates model set at invgamma. In total four runs with four chains were run with 20 million generations and the chains were sampled every 1000th generation. This file was then loaded on CIPRES to run the phylogenetic analysis. After the run was finished multiple new files were created, these files were downloaded and were combined using the commands 'sump' and 'sumt'. This resulted in a new file containing the phylogenetic tree. This tree was edited in FigTree version 1.4.3.

Of the ITS2 region, four analysis were performed. The first step was to determine which parts of the sequence corresponded with the helices. The secondary structure of ITS2 was used for this as the sequence corresponds with the sequence in the alignment. The first helix began shortly after the 5.8S gene and the bases where then followed until the end of helix 1 was reached. This was done for each helix. After the alignment was divided into the four helixes the same process was used as for the phylogenetic analysis of the complete ITS region only with less generation (10 million). The trees were then checked if the clades showed incongruence.

3.8 Designing and testing new chloroplast primers

To be able to design new primers, the chloroplast genome of Syntrichia ruralis was used. This genome was used because there was no complete genome of Campylopus available. Even though Syntrichia belongs to another family, it is the closest genus to Campylopus which has a complete genome. The new primers were designed with the Primer3 software in Geneious. See figure 6 for an example of the input values. The included region is the region where the primers are going to be designed on the genome,

the product size is the length of the full sequence after amplification and the optimal product size is the desired length of the primer. It is possible that the software creates primers that have, for example, a higher or lower GC content than desired or that have a value for the hairpin melting temperature. In that case the input values were adjusted slightly in order to meet the criteria. When the primers were Figure 6. Example of input for designing new primers. designed they were ordered and tested.





For the first PCR amplification a general PCR program for chloroplast primers was used and the optimized master mix was used from the previous tests. See table 9 and 10 for the PCR master mix and program. This master mix was used for all amplifications for the chloroplast primers.

Chemical	Brand	Concentration	Reaction (µl)
MQ	Ultrapure	-	13.4
PCR buffer	Qiagen	10X	2.0
Primer forward	IDT	10 pMol/μl	0.9
Primer reverse	IDT	10 pMol/μl	0.9
dNTP	Qiagen	2.5 mM	1.6
Таq	Qiagen	5 U/μl	0.2
DNA template	-	-	1.0

Table 9. The PCR mix for the first amplification of the new chloroplast primers.

Tab	le	10.	The	PCF	R program	for	the	first	amplification of	ĉ
the	ch	lor	opla	st pi	rimers					

Step	Temperature	Time
Initial Denaturation	95°C	0:03:00
Denaturation	95°C	0:00:30
Annealing	56°C	0:00:30
Extension	72°C	0:00:45
Final Extension	72°C	0:05:00
Pause	12°C	∞
Number of cycles	35	

After the amplification, the samples were analyzed using gel electrophoresis. A 1% agarose gel was made. Onto this gel, 3 μ l of each sample and 3 μ l of a GeneRuler 1 kb plus DNA ladder (Thermofisher Scientific) was loaded. The gel ran for 40 minutes at 100 V and was then stained with a 1 μ g/ml ethidium bromide solution for 20 minutes. Then a picture was taken with an UV-camera.

After this a touchdown PCR was performed. For this method the initial annealing temperature is higher than the optimal melting temperature of the primers and gradually decreases over subsequent cycles until 2-5 °C below the melting temperature of the primers is reached. By using temperatures that are higher than the melting temperature of the primers, the PCR favors the amplification of the template which has the highest primer-template complementarity. For the PCR program of this approach see table 11.

Step	Temperature	Time
1 Initial Denaturation	95°C	0:03:00
2 Denaturation	95°C	0:00:30
3 Annealing	62°C-1°C /cycle	0:00:30
4 Extension	72°C	0:00:45
5 Go to step 2	10x	
6 Denaturation	95°C	0:00:30
7 Annealing	52°C	0:00:45
8 Extension	72°C	0:00:45
9 Go to step 6	20x	
10 Final Extension	72°C	0:05:00
11 Pause	12°C	8
Number of cycles	30	

Table 11. The PCR program of the touchdown PCR for thenew chloroplast primers.

The PCR products of this amplification were also analyzed on a 1% agarose gel, with the same approach as the first amplification.

Another method that was used was a hot start PCR. This method uses a polymerase that is unreactive at room temperature, meaning the amplification process only begins after the polymerase has been activated at 95 °C, which decreases the risk of premature primer annealing. For this PCR program see table 12.

Step	Temperature	Time
Initial Denaturation	98°C	0:01:00
Denaturation	98°C	0:00:10
Annealing	50°C	0:00:10
Extension	72°C	0:00:30
Final Extension	72°C	0:02:00
Pause	12°C	~
Number of cycles	35	

For this amplification the same master mix was used but now with the Thermo Scientific Phire Hot Start II DNA polymerase. The PCR products were again analyzed on a 1% agarose gel.

4. Results

4.1 Kit extraction and CTAB extraction and amplification

To compare both extraction methods, 6 different samples were used in each method (CTAB or kit extraction), because there was not enough plant material of each moss to use the same samples for both extraction methods. All 12 samples were amplified using ITS2 (M13-ITS2-5.8F + M13-ITS2-25R). See figure 7 for the results of this amplification.

For the CTAB extraction four bands are visible, three very strong bands and one weaker band for sample 0570, at an approximate length of 500bp. For the kit extraction only one strong band is visible also at a length of approximately 500bp. The thicker bands just below the ladder are primer dimer. These samples were also diluted 5 times and 10 times to see if there were any compounds present that would inhibit the amplification when not diluted. Compared to the undiluted samples, this did not show improvement. For these result see appendix 4. 'Gel analysis of CTAB and Kit extraction'.



Figure 7. Comparison between two extraction methods. The numbers above the wells represent the sample number. The o after 057 stands for overnight. This sample was incubated overnight during the CTAB extraction opposed to the other samples that were incubated for 30 minutes.

4.2 PCR testing and optimization

For the first PCR amplification, the samples from the CTAB extraction were used and were amplified with the standard program and mix for M13-ITS2-5.8F-25R, see appendix 5.'Standards PCR program ITS2'. Based on these results, adjustments were made. The first parameter that was changed was the extension time. The extension time in the first run was 1 minute and 30 seconds, this was reduced to 1 minute. This did however not show any difference.

Looking at the results in figure 8, it shows very clear bands for 3 samples, but for the other 3 samples only vague bands were visible. This indicated that there was DNA present, only the amplification was not perfect yet. The next step was to look at the annealing temperature. The guideline of the annealing temperature is that it should be within 5°C below the melting temperature of the primers. However, the melting temperatures of the primers used were between 70 and 75 °C, which is quite high for primers. Therefore, a PCR gradient was performed. This is an amplification where the annealing temperature is different for each well. See figure 9 for the results.



Figure 8. Results of the amplification with M13-ITS2 of the samples extraction using the CTAB method, where IT040 stands for the number of PCR amplification. 'la' stands for ladder and bl stands for the negative control.



Figure 9. Gradient PCR for the M13-ITS2 primers ranging from 70°C to 62°C, using sample 0570.

The sample used for this PCR was one of the samples that showed a vague band in the first amplification. The results showed that with an annealing temperature of around 65 °C the brightest band in visible. Therefore, this annealing temperature was used for further amplification.

See table 2, in the methods section, for an overview of which additives and in what combination they were tested. See appendix 6. 'PCR test with different additives' for the master mixes used with the volumes of each chemical and the PCR program. The results of the amplification with all the different mixes were analyzed on a 1% agarose gel, stained with ethidium bromide. See figure 10 for the results.



Figure 10. Results of the additive test with sample 0570 extracted with the CTAB method.

The results show that mixes 4, 6, 9 and 10 have stronger bands compared to the other mixes, where 9 and 10 have the strongest bands. Mix 10 only consisted of buffer with no additive in the mix and for mix 9 there was a small amount of BSA added. Based on these results the following amplifications were performed with mix 10 because this shows to have the most successful amplification.

4.3 KingFisher robot extraction and amplification

A KingFisher robot extraction was performed on 95 moss samples of which 49 samples were originally identified as *Campylopus*. The extracted DNA was then amplified with the primer set M13-ITS-18F and M13-ITS-25R, to obtain the complete ITS region, using the optimized PCR program and master mix. The PCR products were analyzed on a 2% agarose E-Gel, see figure 11. The samples that were originally identified as *Campylopus* are marked by the red boxes. Of the 49 samples, 27 samples were successfully amplified, which is a PCR success rate of 55.1%.



Figure 11. The PCR amplification results of the full plate with M13-ITS-18F-25R, analyzed on a pre-made E-gel.

4.4 Sequencing and sequence analysis

After the amplification of the full plate with M13-ITS-18F-25R, the samples that showed one band were sent to BaseClear for Sanger sequencing, in total 27 *Campylopus* samples were sent. Once the results came back, the sequences were transferred into Geneious. Out of the total 27 samples only 14 could form a contig and were aligned in Geneious. The quality of the other 12 sequences was too low to use for analysis.

The results of the BLAST against Genbank matched with the original identification at genus level for eight samples. Six samples were identified as different genera. See table 13 for the BLAST results of all the sequenced *Campylopus* samples. In the table, only the top hits are shown. For each sample the top 10 consisted of the same genus as the top hit and for some there were multiple species possible. For most samples the top hit had the highest identity percentage and the rest had a lower percentage, hence why they were not presented in the table.

Table 13. Overview of BLAST results against Genbank of the genus Campylopus. The samples of which the identification
matched with Campylopus sequences in Genbank are shown in green, the samples that matched with sequences of other
genera are shown in red.

Name	Genus	Species	Genbank	Species	Query	Identity %
code			Genus		cover %	
IrTa 060	Campylopus -		Campylopus	umbellatus	94	97%
IrTa 068	Campylopus	-	Dicranum	acutifolium	94	97
IrTa 023	Campylopus	japonicus	Campylopus	clavatus	47	90
IrTa 069	Campylopus	-	Dicranum	majus	94	98
IrTa 062	Campylopus	-	Campylopus	umbellatus	94	95
IrTa 078	Campylopus	ericoides	Campylopus	flexuosus	95	98
GD33860	Campylopus	flexuosus	Campylopus	flexuosus	95	98
IrTa 063	Campylopus	-	Campylopus	comosus		
GD33852	Campylopus	fragillis	Campylopus	fragillis	92	92
IrTa 026	Campylopus	sericeodes	Campylopus	gracilis	95	97
GD22184	Campylopus	strictum	Saelania	glaucescens	60	86
IrTa 073	Campylopus	-	Dicranum	majus	94	98
IrTa 081	Campylopus	comosus	Uncultured	-	97	99
			streptophyte			
IrTa 067	Campylopus	-	Dicranum	acutifolium	94	97

The sequences of the matching samples were manually aligned in an existing alignment. This alignment consisted of 129 sequences of *Campylopus* species from all over the world with a focus on South America. This also includes the sequences of *Pilopogon*, which is the sister genus of *Camplyopus* and is used as an outgroup for the phylogenetic analysis. The aim was to add the new samples to this alignment to expand the dataset with samples from Asia. See figure 12 for a part of the alignment.



Figure 12. Part of the existing alignment with the newly sequenced Campylopus samples. This alignment was made and displayed in Geneious.

Each base has a different color to make it easier to align the sequences. All samples were aligned in this alignment, except for sample 'IrTa023 *C. japonicus*' because this sequence showed a large amount of variability in the sequences compared to the alignment and could therefore not be fitted in the alignment. This sample was excluded from this project. The complete alignment, with a length of 2833bp, was used for further phylogenetic analysis.

4.5 Phylogenetic analysis of ITS

The samples were analyzed using the software MrBayes. The phylogenetic tree is presented in figure 13. The numbers on the branches represent the posterior probability, where every value below 0.95 is considered not trustworthy.

Samples IrTa060 and 062 were originally only identified as *Campylopus*. These samples are now part of a clade with *Campylopus umbellatus* with a posterior probability of 0.92. *Campylopus fragilis* is part of the clade with *Campylopus pyriformis*, with a posterior probability of 0.6. *Campylopus ericoides* and *Campylopus flexuosus* are in a clade together, which has a probability of 1. They are also part of a bigger clade with *Campylopus angustiretis, Campylopus sharpie,* and IrTa063 with a probability of 0.91. The latter was originally identified as only *Campylopus. Campylopus sericeodes* is now in a clade with *Campylopus schmidii* with a probability of 0.89.





Figure 13. Phylogenetic tree of the complete alignment including the new samples from Asia (which are marked in red).

4.6 Phylogenetic analysis of the secondary structures of ITS2

The lengths of the parts of the ITS2 alignment corresponding to helices 1–4 of the ITS2 secondary structure is shown in table 14.

Table 14. An overview of the length of the helices.

helix	alignment positions
1	187bp
2	218bp
3	319bp
4	169bp

Each phylogenetic tree from the Bayesian analysis of the four helices had different numbers of clades resolved, part of which were significantly supported (posterior probability above 0.95). In graph 1 the number of clades in each helix is shown. The green bar represents the total number of clades in each helix (with and without support) and the blue bar represents the number of clades with support.



Graph 1. The results of the comparison of phylogenetic analysis of the helices 1–4 of ITS2.

Looking at this graph it is clear that helix 3 shows the most resolution with almost half of the clades being supported. Helix 2 is a close second, with less clades but still almost half of the clades are supported. Helix 1 and 2 show the least amount of resolution with only about a third of their clades supported.

When comparing the separate trees of all the helices, there are no differences between the supported clades in the trees, meaning there are no samples that are in one clade for one helix and in a different clade for another helix. Resulting in no incongruence between the trees. See appendix 7. 'Phylogenetic

trees of helices ITS2' for the phylogenetic trees. Also a tree of the complete ITS2 region was made to use it as a reference for the analysis of the helices which is presented in appendix 8.

4.5 Chloroplast DNA primer design

New chloroplast primers were designed based on the chloroplast genome of *Syntrichia ruralis* with the plug in Primer3 in Geneious. See figure 14 for an example of a primer set in the chloroplast genome.



Figure 14. Part of the chloroplast genome containing one primer set named *trn*K_forward and *chl*B_reverse. This figure was made in Geneious.

The primer pairs were selected based on the genome, they were all selected one by one. By looking at the genome, a selection was made. The primers needed to be in the conserved parts, meaning the primers should be located in the gene (the green parts in figure 14) and they should cover the 'empty' spaces (the blank spaces between the green parts in figure 14).

In total 5 primer pairs were designed with a product size ranging from 544–1081bp. The ideal length of a barcode should be between 400–800bp (Ajmal Ali, et al., 2014). Primer pair *trnK-ch/B*, which has a length of 1081bp, was designed to see if the samples were able to be amplified for a larger region. See table 10 for an overview of the designed primers.

Primer name	Binding site	Sequence	Tm	Length	Direction	GC %	Hairpin TM	Pair Dimer Tm	Product size
atpF_forward	64,767-64,785	CGCACACACTCCCTTTCCA	60,2	19	Forward	57,9	none	none	601 bp
atpH_Reverse	65,348-65,367	GCAGGTCAAGCAGTAGAAGG	58,3	20	Reverse	55	none	none	601 bp
ycf4_Forward	23,242-23,264	GCCCAACAAAAATTACTGATCCT	57,6	23	Forward	39,1	none	none	544 bp
psal_Reverse	23,763-23,785	TTTACCTTCGATTTTTGTGCCTT	57,5	23	Reverse	34,8	none	none	544 bp
chIN_Forward	96,217-96,239	AAGATGTGTATACCTATGCCACG	58,1	23	Forward	43,5	none	none	685 bp
ycf1_Reverse	96,880-69,901	ACCACTTATAGCCACAGTACCA	58,6	22	Reverse	45,5	none	none	685 bp
psbD_Forward	45,062-45,084	GTAATAGACCAGACCAACCTACA	57,1	23	Forward	43,5	none	none	628 bp
trnT_Reverse	45,670-45,689	CTCAGTGGTAGAGTAACGCC	57,4	20	Reverse	55	none	none	628 bp
trnK_Forward	56,573-56,594	AGCTCGTTCTTCTGGTTTTACT	57,7	22	Forward	40,9	none	none	1081 bp
chIB_Reverse	57,633-57,653	TGGTGGTCACGATACAAAAGA	57,2	21	Reverse	42,9	none	none	1081 bp

Table 15. Overview of the designed chloroplast primers.

The binding site presented in table 15 is the place of where the primer is located and the Tm represents the melting temperature of the primers.

These primers have all been tested with the same master mix and program. See figure 15 for the results of the amplification.

A6 A8 A9 H7 A6 A6 **A8** A9 H7 A6 A9 A8 A8 H7 BI A6 A9 H7 trnK-chlB atpF-atpH chIN-ycf1 psbD-trnT vcf4-pal

Figure 15. Amplification test of the new chloroplast primers.

Primer set *atp*F-*atp*H, *ch*/N-*ycf*1 and *psb*D-*trn*T all showed non-specific bands. Non-specific bands can be caused by an annealing temperature that is not suited for the primer set. The first amplification was performed with an annealing temperature of 56 °C. The melting temperature of the primers was around 55 °C, meaning that the annealing temperature of the amplification was too high. Therefore a new amplification was performed with an annealing temperature of 50 °C for the primer set *ch*/N-*ycf*1 and *psb*D-*trn*T. See figure 16 for the results.



Figure 16. PCR results of the amplification with primer set chlN-ycf1 and *psbD-trn*T. Analyzed on a 1% agarose gel Bl stands for the negative contol.

Both primer sets still showed non-specific bands. Another way non-specific bands can be eliminated is by performing a touchdown PCR. For the touchdown PCR, two different programs were used, one with an annealing time of 1:00 minute and one with 0:45 minute. For this amplification primer set *ch*/N-*ycf*1 was used. See figure 17 for the results of this amplification.



Figure 17. Results of the amplification of primer set chlN-ycf1 using a touchdown PCR. BI represents the negative control.

Both programs still gave non-specific bands. The last option to eliminate non-specific bands was to use a Hotstart PCR. This however did not give any results. See appendix 9. 'PCR results Hotstart PCR' results of this amplification.

6. Discussion

The results of the different extraction methods show that for both methods DNA can be extracted, however more samples were successfully extracted with the CTAB extraction. The buffers that were used for the kit extraction were opened in 2012 which may have affected the solutions in a way they did not work optimal. Also not the exact same samples were used for both extraction methods, which makes it harder to compare the results because not all samples are as easily extracted. On top of that, this test was performed with only 6 samples for each method. To determine which method has the most successful extraction it should be performed on a larger scale with the same samples for each method and with fresh buffers.

Even though the CTAB extraction proved to be more successful, the big bulk of samples was extracted using the kingfisher robot, which is slightly different from the kit extraction, because extracting 95 samples with the CTAB method would be too time consuming. In total 49 *Campylopus* samples were extracted using the kingfisher robot. The total ITS region was successfully amplified for 27 of these samples. Most samples that did not show any result were collected in Indonesia in 2011. These samples were not collected and dried the appropriate way for molecular studies, making the DNA degrade faster than usual. Another big part that did not work were herbarium samples. These samples were collected between 1980 and 1999. For older samples it is found difficult to extract DNA with a high yield because of the degrading of the DNA. In research of Särkinen et al. (2012) several extraction methods were tested on herbarium samples, this resulted in the highest yield for the CTAB extraction method but with a slightly lower purity. Therefore it may be useful to extract the DNA of older samples with a CTAB extraction because this generally results in a higher yield of DNA. Another possibility is to try and amplify the extracted DNA with ITS1 and ITS2 separately, since those are smaller fragments.

Of the 27 sequences samples, only 14 could form a contig. Of these, 7 were identified as *Campylopus* when blasted against Genbank. These 7 samples were added to the existing dataset comprising of many *Campylopus* species. Only 25.9% of the samples could be successfully used in phylogenetic analysis. This is quite low. An option for increasing the success rate could be to use a different extraction method which yields in more pure DNA or to add an extra purification step. It is possible that there are too many other compounds that interfere with the DNA.

The samples 060 and 062, which were originally only identified as *Campylopus sp.*, are both in a clade with *C. umbellatus*. Only this clade does not show enough support. However they are part of a bigger clade which also includes *C. exasperates*, which is completely supported. Also the BLAST results show that the best match was also *C. umbellatus* for both samples and the second match, with a slightly less identity percentage, was *C. exasperatus*. This also corresponds with the phylogenetic analysis. Based on the combination of these results it seems like the samples can be identified as either *C. umbellatus* or *C. exasperatus*. Sample *C. fragilis* is in a clade together with *C. pyriformis*. However, this clade does not have support at all. Also the BLAST resulted in a top hit with *C. fragilis*, therefore the chances are that the original identification was correct. However, previous research shows that it is difficult to distinguish between *C. pyriformis* and *C. fragilis* both at morphological and molecular levels. On top of that these species are polyphyletic, resulting in the species identified with those names occur on more than one

well-supported clade, which makes it harder to identify which species it actually is (Stech & Wagner, 2005). In other research C. fragilis and C. pyriformis were also located together in a well-supported clade, which led to the question whether C. pyriformis should be included in C. fragilis (Stech, Sim-Sim, & Kruijer, 2010). Based on this information it is not possible to identify the species and therefore additional information is needed. C. flexuosus and C. ericoides are both in the same clade which is fully supported even though they were both morphologically identified as two different species. However the BLAST results showed that C. ericoides had a top match as C. flexuosus, which can indicate that C. ericoides was wrongly identified and should be C. flexuosus. To be able make this identification additional information is needed. Sample IrTa063 was morphologically only determined at genus level as Campylopus sp. The best match of the BLAST is C. comosus. This cannot be matched with the phylogenetic tree. Based on the combination of results it is not possible to identify this sample on species level with the data that is available now. Sample IrTa026 originally identified as C. sericeodes was in a clade together with C. schmidii, but with no support. The BLAST resulted in a top hit with C. gracilis, this species is in a clade close to C. sericeodes which also does not have support. However based on the available data, no identification can be made. Also of *C. sericeodes* there was no matching species in the tree, which can make it difficult to place it in the right clade.

The phylogenetic tree of the complete ITS region showed many clades with no support and some clades were unresolved. Of the dataset also a phylogenetic tree was made for only the ITS2 region since this was used for the secondary structure analysis. The tree for only ITS2 showed that the new samples were located a bit differently in clades in comparison to the complete ITS region. The samples were in clades with similar species and a higher support. Previous research has been done to the phylogenetic differences between ITS1 and ITS2, this resulted in incongruence between these two parts (Gama, Aguirre-Gutiérrez, & Stech, 2017). This can explain why the complete ITS region shows a lower resolution compared to the ITS2 region because there can be contradicting information in ITS1 and ITS2. The different parts of ITS2 shows no incongruence, meaning there is no contradiction within the ITS2 region. However, to be able to determine and solve the problem more research has to be performed to the secondary structures of ITS1 and the comparison of the two regions.

Overall, it is possible that for some samples the initial determination was wrong and they belong to the clade that they are located in now. It is also possible that there is simply not enough information to separate them and hence they cluster together. This problem can, however, not be solved right now given the current data. To be able to solve this problem more research is needed to understand this problem and additional information is needed to support either the morphological identification or the phylogenetic analysis results with, for example, sequences of the chloroplast region.

All the new chloroplast primers showed non-specific bands, this can be caused by several factors. One factor is the annealing temperature, if this is too low multiple or non-specific bands could appear. Another possibility is premature primer annealing, meaning that the reaction already started before the PCR amplification began. Primer set *trnK-ch/B* showed very vague bands, this can be caused by the fact that the extension time was not long enough. An extension time of 45 seconds was used, but the product size of *trnK-ch/B* should be around 1kb so the extension time should be increased to about 1 minute. However, after all these options were tested the primers still showed non-specific bands. There

is a possibility that the primers have multiple binding sites, which results in multiple bands. Further research can be focused on complete genome sequencing of the genus *Campylopus*. This will make it easier to explore new marker regions for this specific genus.

5. Conclusion

Based on the results of the amplification for the ITS2 region of both extraction methods it can be concluded that the CTAB extraction was more successful. The amplification for the complete ITS region of the full plate showed on band at the desired height for 27 samples out of the total of 49 samples. For these samples the DNA was successfully isolated. For the samples that showed multiple bands there can be concluded that there was some kind of contamination present. For the samples that showed no bands at all, additional testing has to be performed. Based on the results of the test with several PCR additives it is concluded that adding additives does not add any important value to the amplification. Therefore, the best option would be to use the master mix without any additive, which is also the most cost-efficient option.

Seven new sequences were added to the existing alignment. Five samples were in clades with no support and the morphological identification did not match the phylogenetic result. Two samples were together in a clade with high support and was also supported by the BLAST results. Consequently, the answer to the question if it is possible to identify *Campylopus* specimens from Asia based on ITS sequences, is that for some samples it is. However for most of these species no identification can be made based on just the ITS sequences. Therefore additional information is needed to be able to identify them.

The new chloroplast primers all showed non-specific bands after amplification. These bands could not be eliminated by using several different approaches for amplification. Therefore, the answer to the question, which chloroplast markers are potentially useful for the identification of species in the genus *Campylopus*, is that at this moment there are no new chloroplast primers that are useful for the identification of species in the genus *Campylopus*.

The phylogenetic analysis of the helices of ITS2 showed that there is no incongruence between the different parts of ITS2. So the answer the question if the different parts of ITS2 evolve separately, contrary to concerted evolution, is no. Based on the results there is no evidence that the helices of ITS2 evolve separately.
Bibliography

- Ajmal Ali, M., Gyulai, G., Hidvégi, N., Kerti, B., Hemaid, F. A., Pandey, A., & Lee, J. (2014). The changing epitome of species identification DNA barcoding. *Saudi Journal of Biological Sciences*, 204-231.
- Baum, D. (2008). Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups. *Nature Education*, p. 190.
- Bock, J. H., & Norris, D. (2016). Forensic plant science. Elsevier Science Publishing Co Inc.
- Butôt, R. (2015). METH007_HTP_DNA_EXTR_PLANT MATERIAL_KINGFISHER. Leiden.
- CBOL Plant Working Group. (2009). A DNA barcode for land plants. *Proceedings of the Natioal Academy* of Sciences, 12794-12797.
- Choudhuri, S. (2014). Bioinformatics for beginners. Maryland: Elsevier Inc.
- Edger, P. P., Tang, M., Bird, K. A., Mayfield, D. R., Conant, G., Mummenhoff, K., . . . Pires, J. C. (2014, July 1). Secondary Structure Analyses of the Nuclear rRNA Internal Transcribed Spacers and Assessment of Its Phylogenetic Utility across the Brassicaceae (Mustards). *Plos one*. Retrieved from Plos one.
- Eurlings, M. (2015). Meth001_CTAB_DNA_EXTR. Leiden: N.D.
- Gama, R., Aguirre-Gutiérrez, J., & Stech, M. (2017). Ecological niche comparison and molecular phylogenysegregate the invasive moss species Campylopus introflexus (Leucobryaceae, Bryophyta) from its closest relatives. *Wiley Ecology and Evolution*, 8017-8037.
- Gibbs, P., & Hiroshi, S. (1996). *What is Occam's Razor?* Retrieved September 13, 2017, from The Physics and Relativity FAQ: http://www.math.ucr.edu/home/baez/physics/General/occam.html
- Giudicelli, G. C., M\u00e4der, G., Silva-Arias, G. A., Zamberlan, P. M., Bonatto, S. L., & Freitas, L. B. (2017).
 Secondary structure of nrDNA Internal Transcribed Spacers as a useful tool to align highly divergent species in phylogenetic studies. *Genetics and Molecular Biology*, 191-199.
- Glime, J. M. (2006). Bryophyte Ecology. Michigan.
- Goffinet, B., & Shaw, A. J. (2008). *Bryophyte Biology Second Edition*. New York: Cambridge University Press.
- Graham, L. E. (1985). The Origin of the Life Cycle of Land Plants: A simple modification in the life cycle of an extinct green alga is the likely origin of the first land plants. *American Scientist*, 179-186.
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. *Science*, 2310-2314.

- Huelsenbeck, J., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, 754-755.
- Huelsenbeck, J., & Ronquist, F. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 1572-1574.
- Joseph, N., Krauskopf, E., Vera, M., & Michot, B. (1999). Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Research*, 4533-4540.
- Kannan, L., & Wheeler, W. C. (2012). Maximum Parsimony on Phylogenetic networks. *Algorithms Mol Biol*, 7-9.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012).
 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.Bioinformatics. *Bioinformatics*, 1647-1649.
- Kobayashi, T. (2014). Ribosomal RNA gene repeats, their stability and cellular senescence. *Proceedings* of the Japan Academy, Ser. B, Physical and Biological Sciences, 119-129.
- Kress, W., & Erickson, D. (2008). DNA barcodes: Genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences*, 2761-2762.
- Lang, A. S., Kruijer, J., & Stech, M. (2014). DNA barcoding of Arctic bryophytes: an example from the moss genus Dicranum (Dicranaceae, Bryophyta). *Polar Biology*, 1157-1169.
- Mai, J. C., & Coleman, A. W. (1997). The Internal Transcribed Spacer 2 Exhibits a Common Secondary Structure. *Journal of Molecular evolution*, 258-271.

McLennan, D. A. (2010). How to read a Phylogenetic Tree. Evolution: Education and Outreach, 506-519.

- NucleoSpin plant II. (2014, July). Retrieved May 2, 2018, from Macherey-Nagel: http://www.mnnet.com/Portals/8/attachments/Redakteure_Bio/Protocols/Genomic%20DNA/UM_gDNAPlant_ NSPlantII.pdf
- Parmentier, I., Duminil, J., Kuzmina, M., Philippe, M., Thomas, D., Kenfack, D., . . . Hardy, O. (2013). How effective are DNA barcodes in the identification of African rainforrest trees? *PLOS one*.
- Poczai, P., & Hyvönen, J. (2009). Nuclear ribosomal spacer regions in plant phylogenetics:. *Molecular Biology Reports*, 1897-1912.
- Poczai, P., & Hyvönen, J. (2009). Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. *Molecular Biology Reports*, 1897-1912.
- Särkinen, T., Staats, M., Richardson, J. E., Cowan, R. S., & Bakker, F. T. (2012). How to Open the Treasure Chest? Optimising DNA Extraction from Herbarium Specimens. *Plos one*.

- Simmons, M. P., & Ochoterena, H. (2000). Gaps as Characters in Sequence-Based Phylogenetic Analyses. *Systematic Biology*, 369-381.
- Stech, M., & Wagner, D. (2005). Molecular Relationships, Biogeography, and Evolution of Gondwanan Campylopus Species (Dicranaceae, Bryopsida). *Taxon*, 377-382.
- Stech, M., Sim-Sim, M., & Kruijer, H. (. (2010). Campylopus Brid. (Leucobryaceae) in Macaronesia revisited. *Tropical Bryology 31*, 154-163.
- Stech, M., Sim-Sim, M., & Kruijer, H. (2010). Campylopus Brid. (Leucobryaceae) in Macaronesia revisited. *Tropical Bryology 31*, 154-163.
- Stech, M., Veldman, S., Larrain, J., Muñoz, J., Quandt, D., Hassel, K., & Kruijer, H. (2013). Molecular Species Delimitation in the Racomitrium canescens Complex (Grimmiaceae) and Implications for DNA Barcoding of Species Complexes in Mosses. *Plos one*, e53134.
- Stokvis, F. (2016). WI_SO003_GENEIOUS. Leiden.
- Straalen, N. v., & Roelofs, D. (2017). *Evolueren wij nog? Alles wat je wilt weten over ontwikkeling en evolutie van ons lichaam*. Amsterdam: Amsterdam University press.
- Telle, S., & Thines, M. (2008). Amplification ofcox2(,620 bp) from 2 mg of Up to 129Years Old Herbarium Specimens, Comparing 19Extraction Methods and 15 Polymerases. *Plos one*, *3*(10).
- Vanderpoorten, A., & Goffinet, B. (2009). *Introduction to Bryology*. New York; USA: Cambridge University Press.

Appendix 1. Sample overview

 Table 16. Overview of the used samples.

Sample	Year	Genus	Species	Location
Code	Collected			
005 IT	1984	Campylopus	austro-subulatus	Indonesia
007 IT	1986	Campylopus	clemensia	Moluccas
008 IT	1990	Campylopus	crispifolius bartr	Indonesia
010 IT	-	Campyopus	comosus	Indonesia
013 IT	2008	Campylopus	goughii	India
014 IT	1985	Campylopus	goughii	Sri Lanka
015 IT	1997	Campylopus	umbellatus	Malaysia
016 IT	1987	Campylopus	austrosubulatus broth	Papua New Guinea
017 IT	1994	Campylopus	clemensiae	Australia
018 IT	1986	Campylopus	clemensiae bartr	Moluccas
019 IT	1986	Campylopus	c. spor	Indonesia
020 IT	1990	Campylopus	crispifolius bartr	Papua New Guinea
021 IT	1987	Campylopus	exasperatus	Papua New Guinea
022 IT	1987	Campylopus	exasperatus	Papua New Guinea
023 IT	1991	Campylopus	japonicus broth	Japan
024 IT	1994	Campylopus	laxitextus	Australia
025 IT	1987	Campylopus	macgragorii broth	Papua New Guinea
026 IT	1986	Campylopus	sericeoides Dix.	Malaysia
029 IT	2011	Campylopus	comosus	Indonesia; West Java
030 IT	2011	Campylopus	comosus	Indonesia; West Java
031 IT	2011	Campylopus	comosus	Indonesia; West Java
032 IT	2011	Campylopus	-	Indonesia; West Java
033 IT	2011	Campylopus	-	Indonesia; West Java
034 IT	2011	Campylopus	-	Indonesia; West Java
058 IT	1992	Campylopus	-	Indonesia
059 IT	1992	Campylopus	-	Indonesia
060 IT	1999	Campylopus	-	Indonesia
061 IT	1991	Campylopus	sp.	Indonesia
062 IT	1995	cf campylopus	-	Indonesia
063 IT	1995	Campylopus	-	Indonesia
064 IT	1995	Campylopus	-	Indonesia

065 IT	1999	Campylopus	sp	Indonesia
066 IT	2003	Campylopus	-	Australia
067 IT	2015	Campylopus	-	Chile
068 IT	2015	Campylopus	-	Chile
069 IT	2015	Campylopus	-	Chile
070 IT	2015	Campylopus	-	Chile
071 IT	2015	Campylopus	-	Chile
072 IT	2015	Campylopus	-	Chile
073 IT	2015	Campylopus	-	Chile
074 IT	2015	Campylopus	-	Chile
075 IT	2015	Campylopus	-	Chile
076 IT	2015	Campylopus	-	Chile
077 IT	2008	Campylopus	-	Costa Rica
078 IT	1985	Campylopus	-	Sri Lanka
079 IT	2011	Campylopus	comosus	Indonesia
081 IT	2011	Campylopus	comosus	Indonesia
GD33860	-	Campylopus	flexuosus	-
GD33852	-	Campylopus	fragilis	-

Appendix 2. DNA extraction protocols

2.1 NucleoSpin plant II extraction

Macherey-Nagel plant kit extraction obtained from Macherey-Nagel website (NucleoSpin plant II, 2014).

NucleoSpin® Plant II

5 NucleoSpin® Plant II protocols

5.1 Genomic DNA from plant

Before starting the preparation:

- Check that Wash Buffer PW2 and RNase A were prepared according to section 3.
- Preheat Elution Buffer PE to 65°C.

Note: The NucleoSpin® Plant II kits include two different lysis buffers for optimal results with most common plant species. Please refer to section 2.5 for choosing the optimal lysis buffer system for your individual plant sample and for information on how to process even more sample material than recommended in the following protocol.

1 Homogenize sample

Homogenize up to 100 mg wet weight or up to 20 mg dry weight (lyophilized) plant material (for homogenization methods see section 2.4).

Homogenize samples

+ 400 µL PL1

+ 10 µL

RNase A

65 °C.

10 min

Proceed with cell lysis using Buffer PL1 (step 2a) or alternatively Buffer PL2 (step 2b).

2a Cell lysis using Buffer PL1

Transfer the resulting powder to a new tube and add 400 µL Buffer PL1. Vortex the mixture thoroughly.

<u>Note</u>: If the sample can not be resuspended easily because for example the plant powder is soaking up too much buffer, additional Buffer PL1 can be added. Note that the volumes of RNase A (step 2a) and Buffer PC (step 4) have to be increased proportionally.

Add 10 µL RNase A solution and mix sample thoroughly.

Incubate the suspension for 10 min at 65°C.

<u>Note</u>: For some plant material it might be advantageous to increase the incubation time to 30 – 60 min.

Proceed with step 3.

2b	Cell lysis using Buffer PL2		
	Transfer the resulting powder to a new tube and add 300 μL Buffer PL2. Vortex the mixture thoroughly.		+ 300 µL PL2
	<u>Note</u> : If the sample can not be resuspended easily because for example the plant powder is soaking up too much buffer, additional Buffer PL2 can be added. Note that the volumes of RNase A, Buffer PL3 (step 2b), and Buffer PC (step 4) have to be increased proportionally.	Ĩ	+ 10 μL RNase A
	Add 10 µL RNase A solution and mix sample thoroughly.	V	65 °C,
	Incubate the suspension for 10 min at 65 °C.		TO MIN
	<u>Note</u> : For some plant material it might be advantageous to increase the incubation time to 30–60 min.		+75 ul DI 9
	Add 75 µL Buffer PL3, mix thoroughly and incubate for 5 minutes on ice to precipitate SDS completely.		on ice 5 min
	Proceed with step 3.		
3	Filtration / Clarification of crude lysate		
	Place a NucleoSpin [®] Filter (<u>violet</u> ring) into a new Collection Tube (2 mL) and load the lysate onto the column. Centrifuge for 2 min at 11,000 x g, collect the clear flow-through and discard the NucleoSpin [®] Filter.	윩	
	If not all liquid has passed the filter, repeat the centrifuga- tion step.		
	If a pellet is visible in the flow-through, transfer the clear supernatant to a new 1.5 mL microcentrifuge tube (not provided).	Ċ	11,000 x <i>g</i> , 2 min
	Alternatively, centrifuge the crude lysate for 5 min at 11,000 x g and transfer the supernatant to a new tube or pass the precleared supernatant through the NucleoSpin [®] Filter to remove solid particles completely.		
4	Adjust DNA binding conditions	9	
	Add 450 µL Buffer PC and mix thoroughly by pipetting up and down (5 times) or by vortexing.	V	+ 450 µL PC

5	Bind DNA		
	Place a NucleoSpin [®] Plant II Column (<u>green</u> ring) into a new Collection Tube (2 mL) and load a maximum of 700 µL of the sample.	1000	Load lysate
	Centrifuge for 1 min at 11,000 x g and discard the flow-through.	0	11,000 x g,
	The maximum loading capacity of the NucleoSpin® Plant II Column is 700 µL. For higher sample volumes repeat the loading step.	0	1 min
6	Wash and dry silica membrane		+ 400 µL PW1
	1* wash		
	Add 400 µL Buffer PW1 to the NucleoSpin [®] Plant II Column. Centrifuge for 1 min at 11,000 x g and discard flow-through.		$\frac{11,000 \times g}{1 \text{ min}}$
	<u>Note</u> : Although washing with Buffer PW1 increases purity it can in some cases slightly reduce the final yield.		+ 700 µL PW2
	2 nd wash	Ø	11.000 x a.
	Add 700 μ L Buffer PW2 to the NucleoSpin [®] Plant II Column. Centrifuge for 1 min at 11,000 x g and discard flow-through.	Ċ	1 min
	3 rd wash		+ 200 µL PW2
	Add another 200 µL Buffer PW2 to the NucleoSpin® Plant II Column. Centrifuge for 2 min at 11,000 x g in order to remove wash buffer and dry the silica membrane completely.		11,000 x <i>g</i> , 2 min
7	Elute DNA		
	Place the NucleoSpin [®] Plant II Column into a new 1.5 mL microcentrifuge tube (not provided).		+ 50 μL PE 65 °C, 5 min
	Pipette 50 μ L Buffer PE (65 °C) onto the membrane. Incubate the NucleoSpin [®] Plant II Column for 5 min at 65 °C. Centrifuge for 1 min at 11,000 x g to elute the DNA.		11,000 x <i>g</i> , 1 min
	Repeat this step with another 50 μL Buffer PE (65 °C) and elute into the same tube.	Ò	+ 50 µL PE 65 °C,
	<u>Note</u> : To achieve maximum yield or higher concentrations refer to section 2.6 for alternative elution procedures.		5 min
	Elution Buffer PE does not contain EDTA. If DNA degradation		11,000 x g, 1 min

is observed after storage of purified DNA, adjust EDTA in Buffer PE to 1 mM before elution.

39

2.2 CTAB extraction

Obtained from Naturalis Biodiversity intranet (Eurlings, 2015)

TITLE: METH001_CTAB_DNA_EXTR

1 INTRODUCTION

1.1 SCOPE

Lab users who want to perfom a DNA extraction from plant tissue following the procedure described by Doyle JJ, Doyle JL (1987).

1.2 OBJECTIVE

This method is meant to be used for extracting DNA from plant and fungal tissue by using a detergent (CTAB) lysis, an halogenic/organic (Chloroform-Isoamyl alcohol) purification and alkonolic/salt (Isopropyl alcohol/NH₄ acetate) precipitation. In general this method yields more DNA than kit (column) based methods.

1.3 DEFINITIONS AND ABBREVIATIONS

RT: Room Temperature

1.4 STATUS

The status and management of this document are regulated in the AODocs platform, the foundation wide accessibility is provided on <u>N=Info</u>.

2 RESPONSIBILITIES & PEOPLE

2.1 **RESPONSIBILITIES**

FUNCTION	ROLE / RESPONSIBILITIES	NAME
(Senior) Research Technician	Author, Submitter, Quality Controller. Maintain the procedure, train new users, address safety issues.	Marcel Eurlings and all molecular trained technicians
Lab workers (students, guests)	Follow the instructions, note down deviations in lab notebook and give feedback to the trainer and/or author/submitter.	
Head Laboratories	End responsible for all operational methods and its usability.	

2.2 LIST OF PEOPLE ADDRESSED:

SHEET004_ACCESSIONS_NBC_LABWORKERS

3 FLOW CHART

NA

4 ADDITIONAL AIDS AND RESOURCES

mortar and pestle or Retsch Mill (MM200, Retsch): WI_AP008_BEAT-02

eppendorf tubes 1,5 and 2,0 ml

glass beads 7mm

4.1 REAGENTS, SOLUTIONS, KITS AND CHEMICALS

2xCTAB extraction buffer: WI_CS008_CTAB_EXTRACTION_BUFFER

Chloroform-Isoamyl alcohol !!! (24:1) RT

Chloroform

Isopropyl alcohol !!! (cold, -20 °C)

2-mercapto-ethanol !!!

1xTE buffer (10mM Tris pH8 / 1mM EDTA)

RNAse

5M NH₄ acetate

100% Ethanol !!! (cold, -20 °C)

76% Ethanol, 10 mM NH₄ Acetate

Liquid nitrogen **!!!** and container

!!! Safety warning: Only allowed to be used by people instructed by an experienced technician or lab user.

4.2 EQUIPMENT

Shaker

Heatblock

Safety glasses (when working with liquid nitrogen)

Sterilizer: Used for sterilization of forceps and knives

Waterbath: optional for pre-heating and/or incubating CTAB buffer

5 METHOD

5.1 TISSUE GRINDING & LYSIS

5.1.1 Preparation:

- Preheat a heat block (and optionally also a waterbath) at 65°C in the fume hood and switch on the FST Sterilizer.
- Prepare a master-mix of CTAB buffer with 2% (20 μL per ml) 2-mercapto-ethanol.
 Make sure you prepare 10% extra.
- Pre-warm the 2xCTAB extraction buffer (e.g. in the heat block or pre-warmed waterbath).

5.1.2 Grinding with mortar and pestle:

Step 5 is only applicable when using liquid nitrogen.

- 4. Put sample material in the mortar and (optional) add a little bit of sterile sand.
- 5. Carefully add liquid nitrogen, hold pestle in the nitrogen to cool, wait until most of the nitrogen has evaporated and grind the sample material. (work with safety glasses).
- 6. Put the sample in a 1.5 ml tube with a (cooled) spatula, close the tube and put it in the container with the liquid nitrogen. Clean grinder, pestle and spatula before grinding the next sample or use a fresh set. The spatula can be cleaned in between usage using ethanol and a sterilizer.
- After all samples are ground, carefully open the tubes and add 1 ml prepared CTAB buffer. Incubate for 1 hour at 65°C shaking constantly and also mix the whole tube every 10-15 minutes.

5.1.3 Grinding with Beater

Check before you start if the containers to be used in the beater are stored in the freezer! Instructions for the beater you will find in WI_AP008_RETSCH_MILL

4. Mark 2 ml eppendorf tubes for all your samples, add a glass bead (7mm) and a little bit of sterile sand to every tube.

- 5. Add tissue (parts) to the tube and put it in the liquid nitrogen.
- 6. After all samples are prepared put them in the small or big containers depending on the number of samples you have.
- 7. After all samples are ground, carefully open the tubes and add 1 ml prepared CTAB

buffer. Incubate for 1 hour at 65°C shaking constantly and also mix the whole tube

every 10-15 minutes.

5.2 CHLOROFORM EXTRACTION

5.2.1 Removal of proteins

- 8. Add 450 μl chloroform-isoamyl alcohol (24:1) in fume hood, mix by inverting for 5 minutes, under fume hood.
- 9. Centrifuge in the fume hood for 10 min. at 20.000g.

Three phases are visible:

- Bottom- Organic phase: contains chloroform with dissolved proteins, lipids and chlorophyll.
- Middle- Interphase: contains bipolar proteins and molecules and dissolved green cell residue.
- Top- Water phase: contains nucleic acids and dissolved polysaccharides.
 - 10. Carefully take 800 μ L of the water phase and put in a new tube. Do not touch the inter-phase.
 - 11. Add 450 μL chloroform-isoamyl alcohol (24:1). Mix by inverting for 5 min. in the fume hood.
 - 12. Centrifuge samples for 10 min. at 20.000g.
 - 13. Carefully take 550 μl of the water phase and put it in a new 2 ml tube.

5.2.2 DNA precipitation

- 14. Add 550 μl cold isopropyl alcohol to the new tube containing 550 μl of the water phase. Mix by carefully inverting for 5 min. at room temperature, white flakes can appear in case of a high yield high molecular DNA. Leaving it longer (overnight is optional) could result in a higher yield.
- 15. Centrifuge for 10 min. at 8000g.
- 16. Drain the isopropyl alcohol. Place the tubes upside-down on a tissue to pour off as much of the isopropyl alcohol as possible.

Volumes for step 17 to 19 are doubled in some cases, end concentration will

be exactly the same however!

17. Dissolve the pellets in 150 μ l TE, put at 37°C for 30 min., then add 3 μ L RNAse, and leave at 37°C for 30-60 min.

~Possible stop~

Keep samples in 4°C fridge overnight

18. Put the samples at 37°C for 30 min. to dissolve the DNA before continuing

5.3 EXTRA PURIFICATION (OPTIONAL)

5.3.1 Precipitation in high salt concentration – extra purification

- 19. Add 150 μ l 5M NH₄ acetate (final concentration 2,5M). Add 750 μ l (2,5 volume) of 100% cold ethanol.
- 20. Put in -20°C freezer for a minimum of 20 min. Leaving it overnight is optional and could result in a higher yield.
- 21. Centrifuge for 10 min. at 20.000g at 4°C, carefully drain ethanol.
- 22. Wash with 500 μ l washing buffer (76% ethanol, 10 mM NH₄ acetate), swerve carefully
- 23. Centrifuge for 10 min at 20.000g at 4°C, and carefully drain ethanol.
- 24. Optionally for extra cleaning step 22-23 are repeated.
- 25. Spin the tubes again for 5 minutes at 20.000g at 4°C, and carefully drain ethanol
- 26. Place the tubes upside-down on a tissue to drain remaining ethanol. Make sure all ethanol is evaporated.
- 27. Add 50-100 μL TE and put at 37°C for at least an hour to dissolve.

6 RELATED/ASSOCIATED DOCUMENTS

WI_CS008_CTAB_EXTRACTION_BUFFER

WI_AP008_BEAT-02

7 LITERATURE

Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf

260 tissue. Phytochem Bull 1987; 19: 11-15.

Takakura K1, Nishio T.J Safer DNA extraction from plant tissues using sucrose buffer and glass fiber filter.Plant Res. 2012 Nov;125(6):805-7. doi: 10.1007/s10265-012-0502-x. Epub 2012 Jun 14.

2.3 KingFisher extraction robot

Protocol obtained from Naturalis Biodiversity intranet (Butôt, 2015)

TITLE: METH007_HTP_DNA_EXTR_PLANT MATERIAL_KINGFISHER

1 INTRODUCTION

The NucleoMag 96 Plant procedure is based on reversible adsorption of nucleic acids to paramagnetic beads under appropriate buffer conditions. It can be used for plant and fungal tissue samples.

1.1 SCOPE

This document applies when a high throughput DNA extraction from plant and fungal material should be performed by using the KingFisher extraction robot.

1.2 OBJECTIVE

This method describes the process of a high throughput DNA extraction from plant and fungal material by using the KingFisher extraction robot.

1.3 DEFINITIONS AND ABBREVIATIONS

- dH ₂ O	: Distilled water
- EtOH	: Ethanol
- KF	: KingFisher extraction robot
- LN2	: Liquid Nitrogen
- milliQ water	: Ultrapure water
- MN	: Machery Nagel
- sterilizer	: Apparatus used for sterilization of preparation tools
- tissue lyser	: Apparatus used for grinding plant- en fungal material

1.4 STATUS

All SOPS should have a table as shown below to define the operational status of the document. Only SOPS signed off by both the Laboratory Manager and the Quality Officer are officially in use.

VERSION	OPERATIONAL DATE	HEAD LABORATORIES	QUALITY CONTROLLER
METH007.02		Arjen Speksnijder	Marcel Eurlings
1.5 VERSION HIST	ORY	• •	•

Changes in the latest version of a document should be indicated in the text exactly were the change is made. Make the indication by using brackets and the words "change version" and the document name with the old version number. E.g. (change version PROC001.00)

2.1 **RESPONSIBILITIES**

FUNCTION	RESPONSIBILITIES
Technicians	Is able to follow the procedure, gives feedback in case of important changes.
Author	Revise this procedure when there are essential updates.
Head Laboratories	Is end responsible for operational all methods

2.2 LIST OF PEOPLE ADRESSED:

See PROC001_HANDLING_DOCUMENTS

3 FLOW CHART

4 ADDITIONAL AIDS AND RESOURCES

4.1 REAGENTS, SOLUTIONS, KITS AND CHEMICALS

- MN NucleoMag Plant kit (Cat. no. 744400.24):
 - Lysis Buffer: MC1
 - Binding buffer: MC2 (contains isopropanol)
 - Wash buffers: MC3, MC4 (contains sodium perchlorate + ethanol)
 - Wash buffers: 80% EtOH, MC5
 - Elution Buffer: MC6
 - C-Beads
 - RNase 30 mg/bottle
- 0 LN₂
- o milliQ water

4.2 EQUIPMENT

Subsampling and lysis:

- Qiagen Collection Microtubes (Cat. No. 19560) + Caps (Cat. No:19566)
- Tissue Lyser
- Glass beads (3 mm)
- o Forceps
- o Sterilizer
- o Vortex
- Centrifuge
- Shake-incubator (56°C)

DNA extraction

- Deepwell 96 plate (Cat. No. 95040450)
- o KingFisher 200 μl 96 plate (Cat. No. 97002540)
- KingFisher 96 tipcomb for deepwell magnet (Cat. No. 97002534)
- o KF Extraction Robot
- ο Filtertips 200 μl

DNA storage

- Eppendorf twin.tec PCR plate (Cat. No. 951020460)
- Thermo 8-strip flat caps (Cat. No. AB-0784)
- Micronic loborack 96 (Cat. No. MP75117BC)
- Micronic piercable TPE capluser 96 (Cat. No. MP53001)
- o CyBi-SELMA Liquid Handler
- ο Cybio Selma 250 μl tips (Cat. No. OL3800-25-559-N)

5 METHOD

5.1 SUBSAMPLING AND LYSIS OF PLANT AND FUNGAL MATERIAL



Pre-treatment of the Qiagen Collection Microtube Caps (Cat. No:19566) before use.

In order to prevent the caps will pop off the tubes, the caps should be washed in ethanol 96%.

- 1. Put the required amount of caps into the empty BOTTLE 2
- 2. Pour the ethanol from BOTTLE 1 into BOTTLE 2
- 3. Shake BOTTLE 2 with the caps thoroughly
- 4. Pour the ethanol from BOTTLE 2 into BOTTLE 1
- 5. Dry the caps with tissues

5.1.1 Subsampling and lysis of plant material (inclusive mosses)

Pre-treatment in case of DNA extraction from mosses: the samples need to be washed and dried, in order to get rid of the impurities present, such as sand, clay, algae and other plant material, before being placed in the Collection Microtubes.

- 1. Sterilize the forceps before each sample by using the sterilizer and add a piece of dried plant material to the defined Collection Microtube.
- 2. When the tubes of one row are filled-up with plant material add **three** 3 mm glass beads to each tube and close the tubes with the 8-strip caps.
- 3. Repeat step 1 and 2 for the rest of the plate (Optional: store the plate at -20°C).
- 4. When the Microtube Plate is filled-up, centrifuge the plate (20-30 sec, 3700 rpm).
- 5. Use LN₂ to freeze the tubes (20 sec). and put the Microtube Plates in the Tissue Lyser.
- 6. Shake the Microtube Plates (1,5 min., 25/s).
- 7. Repeat step 5 and 6 with the Microtube Plate in a different orientation in the Tissue Lyser.
- 8. Encode the 8-strip caps from 1 to 12.
- 9. Remove carefully the caps and place them on a tissue paper.
- 10. Add 500 μl MC1 Lysisbuffer + 10 μl RNase to each tube and close the tubes.

Preparation of RNase: add 2,5 ml milliQ to a bottle with 30 mg RNase (sufficient for 2 plates).

- 11. Lay a carton sleeve on top of the tubes (this will prevent opening of the tubes during incubation) and close the Microtube Plate.
- 12. Seal the Microtube Plate with tape.
- 13. Vortex (short) and centrifuge (20-30 sec., 3700 rpm).
- 14. Place the Microtube Plate on a pre-heated shake-incubator (56°C, 250 rpm).
- 15. Incubate (at least for 1 hour).
- 16. Centrifuge the Microtube Plate (20 min., 3700 rpm).
- 17. Perform a KingFisher DNA extraction by using the MN Plant Kit (see 5.2).

5.1.2 Subsampling and lysis of fungal material

 Sterilize the forceps before each sample by using the sterilizer and add a piece of fungal material (in CTAB) to the defined tube.

(use a plate sealed with individual rubber caps to prevent contamination with spores)

- 2. When the tubes of one row are filled-up with fungal material add **two** 3 mm glass beads to each tube and close the tubes with individual rubber caps.
- 3. Repeat step 1 to 2 for the rest of the plate (Optional: store the plate at -20°C).
- 4. Centrifuge (20-30 sec., 3700 rpm) and remove carefully the rubber caps (column by column).
- 5. Add 500 μl MC1 Lysisbuffer + 10 μl RNase to each tube and close the tubes of that column. Preparation: add 2,5 ml milliQ to a bottle of 30 mg RNase (sufficient for 2 plates).
- 6. Lay a carton sleeve on top of the tubes (this will prevent opening of the tubes during incubation) and close the Microtube Plate.
- 7. Seal the Microtube Plate with tape.
- 8. Put the Microtube Plates in the TissueLyser.
- 9. Shake the Microtube Plates (1,5 min., 25/s).
- 10. Repeat step 8 and 9 with the Microtube Plate in a different orientation in the Tissue Lyser.
- 11. Centrifuge (20-30 sec., 3700 rpm).
- 12. Place the Microtube Plate on a pre-heated shake-incubator (56°C, 250 rpm).
- 13. Incubate (at least for 1 hour).
- 14. Centrifuge the Microtube Plate (20 min., 3700 rpm).
- 15. Perform a KingFisher DNA extraction by using the MN Plant Kit (see 5.2).

5.2 DNA EXTRACTION USING THE KINGFISHER EXTRACTION ROBOT (PLANT KIT)

- 1. Prepare the lysis-, wash- and elution plates according to table 1 (see below).
- 2. Use the "Machery_Nagel_Plant_96 KingFisher Flex" program.
- 3. For how to use the KingFisher extraction robot see *WI_AP007_KF-01.
- 4. When the KF is finished the eluate can be divide among a DNA Work Plate and a DNA Storage Plate. For this you can use the CyBio Selma Liquid Handler see *WI_AP..._SELMA-01.
 - DNA Work Plate contains 15 μ l of each sample and should be placed in the fridge.
 - DNA Storage Plate contains the rest (± 135 $\mu l)$ of each sample and should be placed in the -80°C DNA-bank.
- 5. Scan the NCBN plate with the BioMicroLab Sample Scan-Mini see *WI_AP..._SCAN-01.

Table 1: Quantities per well of used buffers, beads and lysed samples (Marchery-Nagel Plant Kit)

Lysis plate (deepwell	plate)
Lysed sample	400 µl
MC2	400 μl
C-beads	30 µl

Wash plates (deepwell plates)		
MC3	600 μl	
MC4	600 μl	
80% EtOH	600 μl	
MC5	600 μl	

Elution plate (200 µl plate)	
MC6	150 µl

6 RELATED/ASSOCIATED DOCUMENTS

WI AP007 KF-01

FORM006_APP_LOG_KF-01

WI AP028 SELMA-01

WI_AP..._SCAN-01

7 LITERATURE

-

Appendix 3. Geneious protocol

Protocol obtained from Naturalis Biodiversity Center intranet (Stokvis, 2016).

TITLE: WI_SO003_GENEIOUS

1 INTRODUCTION

1.1 PURPOSE OF SOFTWARE

Editing and quality control of raw sequence read data. Storage and export of edited data.

1.3 DEFINITIONS AND ABBREVIATIONS

BLAST: Basic Local Alignment Search Tool

BOLD: Barcode of Life Data Systems

1.4 STATUS

The status and management of this document are regulated in the AODocs platform, the foundation wide accessibility is provided on <u>N=Info</u>.

2 **RESPONSIBILITIES & PEOPLE**

2.1 **RESPONSIBILITIES**

FUNCTION	ROLES and RESPONSIBILITIES
Lab workers (Researchers, Students, Guests)	Is able to use this document if contributing to the DNA barcoding work process
(Senior) Research Technicians	Knows these document instructions and is able to instruct other lab workers who need to use the Geneious program.
Head Laboratories	Has final responsibility for all laboratory workflow and instructions

2.2 LIST OF PEOPLE ADDRESSED

3 ADDITIONAL AIDS AND RESOURCES

AliView sequence alignment editor

BioEdit sequence alignment editor

http://www.geneious.com/tutorials

4 INSTRUCTIONS

4.1 CHANGE DATA STORAGE LOCATION

Choose a locati	ion to store your data:		
C:\Temp\FS\G	ieneious		Browse
🥅 Help us imp	prove Geneious by sending a	nonymous u <mark>sa</mark> ge <mark>in</mark> form	ation
Privacy Polic	<u>v</u> Activate a License	Request Trial	Trial Later

- Change the data storage location to `C:\Temp\[Name]\Geneious', do not use spaces or special characters!

- The location can be changed later in $\verb|Tools|Preferences|General, restart Geneious and reinstall plugins to apply$

4.2 ACTIVATE LICENCE*

*Only on computers connected to the Naturalis network.

O Use license key			
Licensee name:			
			*
License Key:			
			 -
Ose floating license server			
Server:	1 <mark>45,136,2</mark> 4	1.79	
Port:	27001		
) Use Sassafras KeyServer			

-\Help\Activate License...

- \mathbf{O} Use floating license server
- -\Server: 145.136.241.79
- -\Port: 27001
- \Ok

4.3 CONNECT TO DATABASE

Geneious 8.1.8 - For non-commercial use only	
File Edit View Tools Sequence Annotate & Predict Help	
Q₂ → Q₂ Align/Assemble Image: Comparison Image:	9 ? Ip Support Help
Sources	Set up a new database connection
Sources	Set up a new database connection Connection Method: Direct SQL Connection Geneious Database Server Username: geneious Password: Password: Host: 145.136.242.8 Database Name: geneious Port: 3,306 🕆 V Custom Database Type: MySQL V Connect account on start up Use your own driver Portiver Jara file: Driver .jar file: ServerDatabasePlugin/mysql-connector-java-5.1.22-bin.jar Driver Class Name: com.mysql.jdbc.Driver Connection URL Prefix: jbcc:mysql:// Specify a custom connection URL Specify a custom connection URL Use serializable transactions Use serializable transactions
	Fewer Options Reset to Defaults OK Cancel

- In left window, right click on `Shared Databases' and select `Connect to a database'

- -\Username: geneious
- -\Password: mkjakhf39rtt7dhs890eajln
- -\Host: 145.136.242.8
- -\Database Name: naturalis-geneious
- -\Port: 3,306
- -\Database Type: MySQL
- \Driver .jar file, also available in \B:\DNA Barcoding\Geneious
- $\backslash \mathbf{\sc D}$ Connect account on startup
- \Ok

4.4 PLUGINS

- Install the following plugins in \Tools\Plugins\Install:

-Mafft Multiple Alignment

- PHYML

4.5 FOLDER STRUCTURE



- Create a new project folder in Geneious `\geneious\Projects high throughput', according to the following folder structure: \[project number] - [project name]\[marker]\[project number]-[plate number]_[extraction plate number]_[PCR number] e.g. \16021 - Heteropoda\COI\16021-01_NCBN000907_FS290 (WI SP006 FOLDER STRUCTURE DNA BARCODING DRIVE)

- Change folder or file colour (to be edited, edited, uploaded to BOLD) by right clicking on the folder \Set Folder Colour...

4.6 READS ASSEMBLY

- From the relevant BaseClear folder on e.g. B:\DNA Barcoding\, select the .ab1 files and drag them to the relevant Geneious plate folder use \File\Import\From File...\ to select the .ab1 files

- select all . ab1 files in the Geneious plate fiolder (Ctrl+a)

-\Align Assemble\De Novo Assemble ...

Data	
Assemble by: 1st 👻	part of name, separated by - (Hyphen) -
Assemble e	each sequence list separately
/ethod	
lealou	
Sensitiv	vity: Highest Sensitivity / Slow 💽 👔
Memory F	Required: Between 25 MB and 27 MB of 5.5 GB
rim Sequences	Results
	Ascembly Name /Deads Name} Ascembly
Use existing trim regions	Save assembly report
Remove existing trim regions from sequences	Save list of unused reads
	I Save in sub-folder
Trim sequences Options	
O not trim	Save contigs (V Maximum 1,000 -)
	Dave consensus sequences COUDIS

- $\$ Assemble by 1st part of name, separated by (Hyphen)
- \Sensitivity: Highest Sensitivity Slow
- \mathbf{V} Save assembly report
- $\mathbf{\nabla} \mathbf{Z}$ Save in sub-folder
- $\$ Save contigs
- -\🗹 Maximum 1,000
- -\ Do not trim
- \Ok
- A new folder called 'Reads Assembly' will be created, containing the contigs

These settings will be saved for future use of Geneious



- From the folder 'Reads Assembly', select 'Assembly Report'

- In the bottom window, select the `not assembled' link e.g. `6 reads' and non-assembled reads will be selected in the plate folder

- Right-click on the selected reads \Set Document Color\ and select the colour red

4.7 PRIMERS

In order to easily find and remove the primers from the reads files when editing contigs, they can be visualised in Geneious.

- In the left window in Geneious, open the green folder 'Primers'

-\Sequence\New Sequence...

					Convert data		And the second second	Contraction front format
A State A		0 0 0			sections sont preventege	n Pagna-more	at runnes oegenen	Reconcertific press
· ··· · · · ·	· · · ·			Q, 🗇 🖌 🙀 Search	B Z T	A	I III 121 . INT. M.	The second secon
Parward BLACT Agents Algn/A	asentile Tree Primers Cloring	Bad Up Support rep		Artebrand	Plakten	A A	1 m M	Softeren en Zoeken en
Sources	10	Contrast Contracto		0.01.90 3696,060		4 · · · ·		Googensen. 2 Meren velederen
Secole Day ments (621)		Name A	Seque Organis	 Modified 	Centrolid + Lettertyp	* 14 UB	ining a cettat a	30per Lesen Bewerken
Deleted Dems (379)		M13R dirdR2	46 -	28 Mar 2012 3:38 PM *	D25 •	fr TGT	TAAAACGACGGCCAGTTAA	TTTACGATCAATTCATTC
ned Catabases		MISR-COLEXIR	43 +	28 Mar 2012 3:38 PM	A A		c	D
peneious (8)		MLIR+IC02191	43 -	28 Mar 2012 3:38 PM	1 Drimor	Incode	dir	anguages 5' - 3'
Analyses (3)		M13R-HC02198	43 -	24 Apr 2013 1:08 PM	3 14135 (COL480	cos		
Contaminations (224)		MIR GERT	40 -	24 Apr 2013 198 PM	2 4125 000 1490	0.00		TO FRANK WOOD WOOD TO FUNCTION CALIFORNIA LATE VO
Primers (48)		 MLR OLL (eversed) 	37	25 Jan 2017 9:39 AM	a Wise-CepF1	COL		TOTAMACONCOGCCASTATTCAACCAATCATAAASATATTGG
RLASSE BOT 175 (3324)		MITRAF(#)	17.	29 May 2012 2-18 PM	4 M13R-HCO2198	COI	8	CASGARACASCIATGACTARACTICASGGIGACCAARAAATCA
C RLASSE Bot Mark (2467)		MIR-bd.ar (everged)	37 -	21 Dec 2016 12 32 PM	5 M13R-LepR1	COI	R	CAGGAAACAGCTATGACTAAACTTCTSGATGTCCAAAAAATCA
CASE 200 COI (10313, 357 mmre	e0	 M13R-briefR (reversed) 	40 -	16 Jan 2017 3:15 PM	6 M13F-LepFoF	COI	8	LepF1/LC01490
NGS Projects (87)		🐨 MF1	22 -	07 May 2012 30:06 AM	7 MI3R-LepFolk	COI	8	lepR1/HC02198
Cither Projects (Kevin) (48)	S 59	★ M01-MR1	20 -	07 May 2012 10:06 AM	8 M13F-BirdF1	COL	F	TOTAAAACGACGGCCAGTTTCTCCAACCACAAAGACATTGGCAC
Projects DNA barcoding (15130, 538)	9 unread	➡ H.epF1	22 -	24 Jul 2012 5:09 PM	9 MI2R RindR1	001	8	***************************************
Projects high twoughput (3)	The second se	 MLepR1 	20 .	24 34 2012 5:09 /44	10 14118 Bird93	100		
11011 - Root essentieled Corp G	ent (209, 1 several)	OD0-HC02241	21 -	28 Mar 2012 3:38 PM	AV HARN-BEOK2	0.00		CANADA AND AND AND AND AND AND AND AND AN
B 15036 - A LEC and Macan presser	ar yum species (1125)	- 0004001512	20 -	20 Mar 2012 3:30 PM	11 W15P-8#d+	001		Birdri/LC01490
 gar Ltdbk - Lans Canary Island prog 	et (al.d. 328 usread)	- ALTIGLE	-	014-02014-01124	12 M13R-Bird+	COI	R	BirdR1/BirdR2/HC02198
III D IIIII - Anadore Pers part harrow	along Cassudria Netternal Press			The way to be an	13 M13F-0DO-LC01490d	COL	1	TGTARAACGACGGCCAGTTTTCTACWAACCAYAAAGATATTGG
In the second second (S75)	lew Sequence	and these sections with the sec	- 100	- X-	14 M13R-ODO-HCO2198d	COI	8	CAGGADACASCTAYGACTAAACTTCW9GRTGTCCAAARAATCA
- B store - Georges sortes rad					15 DDO-cocktail-F	COI	F	0D0 LC01490d s1/LepF1 s1/LC01490 s1
and a start of	Type or paste in sequence:				16 000-cocktail-R	001		000 H0021984 +1/tan81 +1/H002198 +1
O BE MARK IN	1-40 TOTAAAACO	ACOOCCRUTTAATTIACGATCAATTCATIC		*	17 Upi Maihad1	Miel COL		
Bank America					10 Uni Marthauth	1000	121	
H 10032-02 NORMOOD					15 Oni-Minibarks	Marin-COI		CAPONANCALCTATORCONDATCATARTORADOCATONIC
18 00 150 12 03 NC5N0000					19 M135-IErcO1490	COI		TOTAMAACGACGGCCAGTTITCIACIAAYCAYAARGAYATTGG
8-69-020					20 M13R-igHCO2198	COI	R	CASGANACASCTATORCTALACYTCISGRIGICCEAARAAYCA
- D 18053 - Cartnerius and Russ					21 M13F-175E	ITS	£.	TOTARAACGACGGCCAGTACGAATTCATGGTCCGGTGAAGTGTFC
🖶 💭 38058 - Tetractinelida (de Ve					22 M13R-265E	ITS	R	CASGRARCASCTATGACTAGRATTCCCCCGGTTCGCCGCTTAC
(8) 👹 36066 - Pheropoda CCE (129					23 M13F-5.8/2	ITS2	F	TOTARACOACOGCCAOTOCCT000COTCACOC
Image: Signature State Stat					24 M138-5.8/1	1751		CAGGAAACAGCTATGACGTTGCCGAGAGTCGT
🖶 📁 15571 - Neptici,kodes (1567)					25 MINE-matk-of	mate	1	007333330030000000000733777300370337903750
🖶 🎒 20078 - check (204)					26 MILLS mate MALD	matif		21 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
B 18076 - Nephrolepis (179)					27 material	marker.		
🗊 🗊 30001 - Testudo Herrinarsi 🚺					27 materia	mate		TAATTACUATCAATTCATTC
Ants (R. 180/18) (28)					28 matk-MALP	matx	R	ACAAGAAAGTCGAAGTAT
B 38087 - Path 20 (B. Heekbern)					29			
speedlest (360, 189 unreal					30			
Thinks (13)					31			
united (1993)					32			
					33			
ane l				•	34 M13F	Universal	e .	207333320000000007
Genome		Name: M13F-matK-of			15 14110	tinisseral		CAGGAAA (CAGCERARGA (
Nucleoside	De	scriptor: Column			10 14110 (3-10)	000	2	
Paplet.				12		Mark Cost	2	CHEPARATCALA TALATO
Protein		a Bausar I of cruze		N	37 M15P-Ohi-MinibarF1	Mini-C0I	-	TOTARAACOACOGCCADTTCCRCTAATCACAARGATATTGGTAC
PubMed		Type: Primer - Owner	teristics		38 M13R-Uni-MinibarR1	Mini-COI	R	CASGAAACASCTATGACGAABATCATAATGAAGGCATGASC
51P		the party of the second			39 M13F-Bird-plus	COI	1	BirdF1/LC01490
Structure		Cirolar			40 M13R-Bird-plus	COI	8	8irdR1/8irdR2/HCO2198
Taxanami					41 MI3F-HanfF	COI	F	
(Acceleration)				and a second sec	the second se	Manager and strength of the State of	A statute and a second a second	and the second se
hat.				OC Cancel	Seq options	SSECEN ROLL	D Sneet PCK miles PC	Commers PLX programa Cal

- From your PCR sheet (e.g. <u>PCR_template_v4.3.xls</u>), from tab 'PCR primers', copy a primer sequence and paste it in the 'New Sequence' window in Geneious

- Fill in \Name: \e.g `M13F-matK-xf' and select \Type: \Primer

- After creating the reverse primer, double click on it, select ${
ot\!\!\!\!\! ilde{ R.C.} }_{\backslash Ok}$

4.8 EDITING CONTIGS

- In the 'Reads Assembly' folder, double click on a contig and the contig window will open

- \Graphs \ 🗹 Qual

- Zoom to basepair level by clicking the magnifying glass button on the right

e012521570 Nep bro MI059 pitA (Contig)		C.	- S -×
de Edit View Sequence Annotate&Predict Help			
Conte View Dobplot Virtual Gel Lengths Graph Text View Jofo			
🗇 🕸 🕞 Extract 💋 R.C. 🤡 Translate 🏨 Additidit Annotation 🛛 🖉 Alon Edilog 🚔 Annotate & Predit 🚓	Primer Design 🙀 Sane		SQ (
Consensus	INTOTOTIĞANSCHECHTÜTATANACGGÜTANDACCCAÜCTGGTTATGÜNGCACANCTĞGATADCANAÜCHCHTAGASĞTTTG	PP 200 ⊕ ₩ J Graphs	P 31 1
	monthere and the month of the second of the	Show Graphs Dop	ort
De 14012521570_Nep_bro_ME059_pstA-C11_M13F_M13R_M13-FP (P3591)_X170121003.ab1	TTCTGTTGAAGCTCCTTCTATAAACGGGTAACACCCATCTGGTTATGCAG B ACAACTGGATACCAAATCTCTTAGAGGTTGG	C Content 100	10-17 E
	Anna and a sharp a same and a same	Sking Window Szer 48	9
e= HV e4012521510_Nep_tro_ME059_ptbA-C11_M13F_M13R-pUC(-40; (*3552)_X170121119 ab1	TTT THE THE CAADE TCCTTETE TAAACGGGTAACECCENTE TGGTTETGCAGCECAAETGGETED CAAATETETTAGEGGTTEG	Chronatogram	00
unor before bees 1 in 3 sequences. Mouse in e401201570_Nep_bro_M0059_pable C11_H13F_M13F_M13FFP (F3591)_3170121613.abd	· · · · · · · · · · · · · · · · · · ·	100 00 00 00 00 00 V	- 18

- Verify the reads (trace files) FWD and REV directions by hovering over their names and checking the last part of their file names. In this case the FWD primer name ends with `M13-FP', so the directions are correct.

- Use ØR.C. to change the directions when necessary

Geneious 8.1.8 - For non-commercial use only			0 8 2			
File Edit Vew Tools Senamer Associate & Eventert Hele						
				and the second se		
Q · · · · · · · · · · · · · · · · · · ·		Q	search			
Back Forward BLAST Agents Align/Assemble Tree Primers Clo	ning Back Up Support Help					
Sources	N		14f 50 selected			
(I)- 📁 Local (0)	V Name A	Seare Organism	Modified 175			
() Disample Documents (623)	a line in a	Contraction of the second		the second se		
🗄 🖬 Deleted Tierrs (179)	MIR COLEXE	43 -	28 Mar 2012 3:38 PM			
i Shared Catabases	MI3R+C02151	43	25 Mar 2012 3:36 PM			
(i) (D geneious (8)	MITC/anD1	43	14 Apr 2011 1-08 PM	M13F-matK-xf (Sequence)		
E-C Analyses (8)	M132 (81 Severand)	37 .	26 Jay 2017 9-99 AM			
Gontaminations (224)	 M13R-matK3FGM (reversed) 	42 .	04 Aug 2015 11:23 AM	He bol Yew Sepance Annotate & Hedict He	P	
Primers (S0)	🖓 🖛 HISR-netK-HALP	35 +	19 Jan 2017 1-11 PH	Sequence View Annotations Delpist (Self) DNA Fold	Text.View Infe	
RLASSE BOT ITS (1320	MISR-sUC(-40)	17 +	29 Mar 2012 3:38 PM	On all Contract of D.C. of Translate all and	Edit Acceptation 🧳 May Editors 🧰 Acceptate & Dealert 🍂 S	toner Destron III Source II
TANGE BOCHARK (249-7)	MISR-bd.ar (everged)	37 +	21 Dec 2016 12:32 PM	A A Course State & search date	An an and a substant at the	and strap. We see
- go takon and an anti-	 MI3R-briHR (eversed) 	40 +	16 Jan 2017 3:15 PM			P P 20 0 % ++ P 21
(i) (ii) (iii) (iii) (iii) (iii) (iii) (iii)	➡ M ² 1	22 -	07 May 2012 30:06 AM			General
Denie de Projecta parteri (160)	▼ 391-P5[1	20 -	07 May 2012 30:06 AM	BORNAR BOOK CONCERNMENT	A REAL PROPERTY AND A REAL	Colorest BERGE
S - C Projects Crick barb data data d 20	→ MLepF1	22 -	24 3.4 2012 5:09 PM	TO TAXAR OCCUPENCE AND TAXE	A SUGAL CARLES ALLO	
a project reprint and the first first and a		21	24 JU 2012 505 PM			I Cranter Ontaria >
The second secon	- 000+001514	24	20 May 2012 3:30 PM			A adva
In 193 1932 - Lane Control Mand county (2178 - 338 increased)	 dd-fielF 	44 -	26 34 2016 2:38 PM			V Arrestators Cotors > . 9%
10 100 1011 - Analysis has and have been been been been been been	+ md-m#-M13F	38 -	03 Aug 2016 6:13 PM	Out which are a second configure or emplotion, or which a	restor to some in Chirdelik bil shisk to some tool.	The second secon
in all 18001 - Heterconde (1970)	rbd-infR-M13R-rc (reversed)	41 -	27 3J 2016 11:99 AM	Contraction in the second consideration of the second		
0-10 36024 - Georgea Bornes Zac (240)	Tht21375R-M13R-rc	43 -	26 Jul 2016 3:11 PM	M138-matk-M2(P (Securery e)		
- M MO32 - Guyana stat barcading (0)	🖛 T3	17 -	28 Mar 2012 3:38 PM			
in the matrix (0)	 T2pronoter 	20 -	28 Mar 2012 3:38 PM	File Edit Vew Sequence Annolate & Wedict File	The second s	
B 16032-01 MCBM000003 P5289 (190)	Une-MinibarF1	44 -	28 Mar 2012 3:38 PM	Sequence Vew Annotations Dotplot (Self) DNA Re	A Neverse Complement	
- (IR Reads Assembly (89)	Un Minbark1	41 +	28 Mar 2012 3:38 PM -	A sh Calabar A B. W Transists sh a	Contraction in the second second second second	History 52 (D)
III 16032-02_NCBN000904_P5288 (297)	·		VIEW	A A C	Contact and retreated companies contraction region	
160 100 100 100 100 100 100 100 100 100	Seguence View Acceptations Datalet (Leff)	CNA Faild Text Vew 2-fa			Extraction name:	
(B) (D) (5:11 (5:21)	A A C	An and the second second			· Reverse consiment entre sequence	General
🛞 💋 18053 - Certinarius and Russula (141)	Qa alle Ge Extract 22 R.C. 💽 Translate	SP Ass/Edt Annotation	>> 66 01 00 19 7			Friere Allia
🕀 📁 16058 - Tetractinelida (de Voogd) (193)		P P 100	N. 46 JP 22	Distant Report	C Actively link source and extracted documents	
Image:		Gar	And a state of the			Gracha Cotors >
Im 26067 - Underground Fungal networks (275)		Contraction of the local sectors of the local secto	and the second se		OK C	and D
ISS71 - Nepticukides (1567, 214 unitead)		Calors: ACC.	• cat @			Anotators Optices > %
🖶 🎫 18078 - Check (204)		and all the second second	in the second second	Child did as a second poster or section or particular	and a state of the state of the state of the state of the	
(II- III 10076 - Nephroleps (179)		(V) Graphs	ciptions >		April & South St. Co. and Mr. Color Without Color	
e4012517258_Way_aka_F5289_matK (Contig)		-		and the second se		
The Lab Many Semigraphy Learning & Reading Links						
The same ten supporter removes a ready resp						
Control View Annotations Dolptot Vintual Gel Lengths Graph Text View 1	infe)					
🕼 🕸 🕞 Extract 🧱 R.C. 🚱 Translate 🏟 Add/Edit Arrostation 🌙	🕈 Allow Editing 🚔 Annotate & Predict 🥠 Primer	Design 🛛 📓 Sinver				E (1)
	10 50	1	0.20	10 AN 71	20 BL 80 BL	
Comensus	ATTTECTTTTAGAGGACAA	ATTICACATIT	AATTATGTATTAG	ATATACTAATACTCCACTCCATTCA	ATCTGGARATCTTGGTTCARACTCTTCG	CTAC
				0 0	F1 8 57	6raprs 🕌
	had and we	Ph rola IL	A state	me and All -A.Da.	H man A H M H	Show Graphs Export.
	DOC 0000 XX 3000000	REXXXX	X CONTROLO	CONSTRUCTION CONTRACTOR	About About the sale to	
IN HUARMINESTICS Was also CODE materials material	A REPORT STREET ACACCACAA	ARRECOACARES.	LATTATCTATEAC	A TARACTA A TACTOCACTOCATICS	A POPOCA & & POPPOCEPTA & A PPOPPOC	
Contraction and and and and and and a full	CALL SCIENCE AND ACKA			ALTA INCIDENTION INCIDENTION	and a stand and a stand	
		17			AAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	Kittin Prase por
			mohm	20000000000000000000000000000000000000	XVVX IV - VXVVV IVVV IVVV	Searce Window Size: 48-5- 48-5-
De NO e4012517258_Way_aka_F8289_matKH11_matKat1	TERMONTTTTTTTTTTANAGGARAA	TTTCCACATET	AATTATGTATTAG	ATA7ACTAATACTCCACTCCATICJ	ITCTGGAAATCTTGGTTCAAACTCTTCG	CTAC Dimensionan (A) (A)
		750				CALL THE THE THE THE O
×						1 10 4 10 0 10 6 10 1 10 0mm * 4
Cursor before base 836 in 3 sequences.						
		A DESCRIPTION OF	The second second	the second se	a suggestion of the suggestion	A
					A COMPANY OF A COM	10 10 10 10 10 10 10 10 10 10 10 10 10 1

- Open the primer windows to help find and remove the primers from the reads files

- When editing, pay attention to 3 things: 1. confidence (blue columns behind chromatograms), 2. gaps in consensus (upper sequence) and 3. ambiguities in consensus

- when part of a 'bad' read file causes a lot of errors, it can be deleted individually

- save (Ctrl+s) when finished with a contig and close (click Yes when asked to make changes in the original sequence)

- \Set Document Color... \ high quality green, medium quality (1.5-3.0% ambiguous bases and/or 25-40% missing data), low quality (>3.0% ambiguous bases and/or 40% missing data), contaminated/NUMT

4.9 EDITING NON-ASSEMBLED READS

If only one of the two reads is of high quality, it can still be used for further analyses.

- Go to the plate folder with selected non-assembled reads
- Check the quality of red coloured reads and edit if possible (reads with a HQ% >1 can mostly be edited)
- \Set Document Color... \ high quality green, medium quality, low quality, contaminated/NUMT
- Sort by \color\, select all green and yellow reads
- -\Align Assemble\De Novo Assemble ...

- The `Assembly report' will open automatically in the bottom window

- Click on the `not assembled' link and non-assembled reads will be selected

- Copy the non-assembled green and yellow reads (Ctrl+c) and paste them in the folder `Reads Assembly' (Always leave all original reads in the plate folder intact!)

- Cut the new contigs and assembly report from Reads Assembly 2 (Ctrl+x) and paste them into the 'Reads Assembly' folder (Ctrl+v) and keep both alignment files

4.10 ALIGNMENT

4.10.1 Alignment of consensus sequences



- Select all green and yellow contigs

-\Align Assemble\Multiple align...

 $- \mathbf{O}$ Create alignment of consensus sequences only

- \MAFFT Alignment

- \Algorithm: Auto

-\Scoring matrix: 200PAM k=2

- -\Gap open penalty: 1.53
- -\Offset value: 0

- $\$ Automatically determine sequences' direction

- \Consensus options
- \Threshold:Highest Quality
- \🗹 Assign Quality: Highest

- OkOk, If an error occurs, change the data storage location (paragraph 4.1)

4.10.2 Alignment of all sequences

- In the folder 'Reads assembly', deselect all contigs

- Select both the file 'Nucleotide alignment' and the non-aligned green and yellow reads

- -\Align Assemble\Multiple align...
- O Create alignment of all sequences (keep existing gaps)
- \MAFFT Alignment
- \Algorithm: Auto
- -\Scoring matrix: 200PAM k=2
- -\Gap open penalty: 1.53
- -\Offset value: 0
- \ Automatically determine sequences' direction
- \0k

- Double click the new nucleotide alignment (e.g. 'Nucleotide alignment 2') to open it in new window

e4012521512 Nep aler MEDS pubA Contig				×	0+ 20. e4012521510_Nep_cor_ME051_psbA 0+ 21. e4012521517_cor_lia_ME059_psbA	GGGEGGAEGEAGEEAAGEG	GATEAAGGEAATG	
File Culit Vern Sequence Annotate & Fredict Help					CP 22 e4012521510_cor_lia_ME059_psbA CP 23 e4012521543_np_cis_ME059_psbA	GGGEGGAEGTAGEEAAGTG	GATCAAGGCAATG GATCAAGGCAATG	
Contg View Annotations Dolplot Virtual Gel Lengths Graph Text View	Infe				C+ 24 e4012521554 cor Na ME059 psbA C+ 25 e4012521497 Nep cor ME059 psbA	GGGEGGAEGEAGEEAAGEG	GATEAAGGEAATG	
🕼 🕸 🕞 Extract 🧟 R.C. 😵 Translate 🏟 Add.Eist Avvolution 🔰	🖋 Allen Editing 🚔 Annotate & Predict 📌 Primer Des	on Seve	E	01	C= 20. +4012521532 exa ata ME059 psbA C= 27. +4012521553 hip_cis_WE059_psbA	GGGGGGGACGTAGCCAAGING GGGCGGACGTAGCCAAGING	GATCANGGCAATG GATCANGGCAATG	
Contensus	AGECAAGING-ATCAAGGCAATG	S S 200 - ™ Graphs		5	D= 29. e4012521500_cor_e8_MED56_p8bA D= 39. e4012521533_Nep_cor_MED56_p8bA D= 30. e4012521545_Nep_cor_MED56_p8bA	GGGEGGAEGTAGEEAAGTG GGGEGGAEGTAGEEAAGTG	GATEANGGEAATG GATEANGGEAATG	
.00	attack the march	2 Show Graphs	Export		De 31. e4012521557_Nep_lau_MED55_psbA De 32. e4012521519_lau_Fil_MED59_psbA De 33. e4012521519_lau_Fil_MED59_psbA	GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	GATCANGGCAATG GATCANGGCAATG GATCANGGCAATG	
C+ MV e4012521512_Nep_atr_ME059_pe64-C12_N13F_M. 2.07	AGCCAAGEG ATCAAGGCAAB	C GC Content	100 -		De 34 e4012521499 Nep dic ME055 psbA De 35 e4012521511 Nep dic ME055 psbA De 36 e4012521510 per ata ME059 psbA	GGGGGGACGTAGCCAAGTG GGGGGGACGTAGCCAAGTG GGGCGGACGTAGCCAAGTG	GATCAAGGCAATG GATCAAGGCAATG GATCAAGGCAATG	
	Amath and Marine Marinet	Skilling Window State	40 [5]	2	0+ 37. e4012521500_Pta_hot_WE059_psbA 0+ 38. e4012521512_Nep_abr_ME059_psbA	GGGEGGAEGEAGEEAAGEG	GATCARGGEAATG	
D+ H0 e4012521512_Nip_atr_ME059_bsb4-C12_W13F_MCGT	ANCCAAGEGNATCANGGCAAT	🖉 Ovomatogram	40-0	o,	Go to referenced sequence (hold down Ctil	to open in new window)	0.07	
4 Current Palitone hasee 504 (rectining home 504).		WA WE WE WI	V Qual +	<u>1</u>	Current Indices Issue 1 (opticited house 1). Novum in add/2010	4 12 Mars after MEXIS could	· ·	

- Check alignment for correctness by opening the contig window and alignment side-by-side

- When you click on the blue arrow 📴 to the left of the sequence name in the alignment, while holding down Ctrl, the contig will open in a new window



- Check the alignment of coding genes (e.g. COI, matK, rbcL) for gaps (shown as a dash `-', or search the alignment using Ctrl+f for `-')

4.11 CHECK FOR CONTAMINATIONS AND MIX-UPS WITH A MAXIMUM LIKELIHOOD

TREE

4.11.1 Contaminations

- Select the relevant nucleotide alignment (e.g. 'Nucleotide alignment 2')

Substitution model:	GTR 👻	012345	
Branch Support:	No Bootstrapping/Likelihood 👻 Number	of bootstraps:	0
Transition / transversion ratio:	4 *	Fixed	ł
Proportion of invariable sites:	0	Fixed Stimated	1
Number of substitution rate categories:	4		
Gamma distribution parameter:	0	Fixed Stimated	ł
Optimize:	Topology/length/rate 🗸		
Topology Search:	NNI (Default, fast) 👻		
The second second			

- \Tree
- \mathbb{PHYML}
- -\Substitution model: GTR
- \No Bootstrapping Likelikhood
- \Ok

Leave the rest rest of the settings default (the analysis will take a few minutes)

Comparing the second seco
- Memory can be freed by clicking the memory field at the bottom left of the left screen
- A Maximum Likelihood tree is created (e.g. 'ML tree of Nucleotide alignment 2')
- Double click on the ML tree to open the tree window
- Adjust \Font
- Right window
- $\backslash \mathbf{Z}$ Enable all layouts for unrooted trees

- $\$ Order Branches

- \Ordering: decreasing



This tree has a clear indication of contamination, shown by the long branch.

ne tak vew septence Anneuer Crivedici Hop	Construction of the state of th
Contg Vew Annotations Dolptot, Whual Gel Lengths Graph Text Vew Info	🖕 🔶 🗋 🛔 https://blat.orbl.ahm.mh.gov/Blat.cg/CMBe/WebSLLW 🛛 🖉 🔍 Jorism
A Ber Größer Z. C. Transfer Be Add Annotation / Alex Eding Annotation Annotation (Particle Annotation Annotation) Alex Eding Annotation (Particle Annotation) Prove Despin	1 1
Diffee AT Graph	Strandard Musleadide BLAST
Skilling Window Statt 188	Matte blasts blasts minute blasts
W Oronatogram (4) -	Enfer Query Sequence BLASTN programs search nucleotide databases using a nucleotide guery martine Statitude Bookman Basitanas Bookman
Ti Coverage (1)	Enter accession number(s), gi(s), or FASTA sequence(s) 😡 Clear Query subrange 🚱
□ (m) (m) (m) □ (m)	Core upload file Idadema
an an an an an an an	Job Title
Com. SAAATAAATCAACAGAGGGGGGGGGGGGCGGACCTAGGCAAGGGATCAAGGCAARG	Align two or more sequences ()
Way An Way and a	Choose Search Set
TAATAAATCAACAAGAGGGGCCGACGTAGCCAAGCCAAG	Durahase Okuman generic + transcript OMouse genome + transcript ® Ohives (in etc.) Noganiam Corrections (invin) Durahase Corrections name in the constraints will be suggested Durahase Durahas
	Enter organism common name, binomial, ertaxid. Only 20 top tasa will be shown 🥹
	Umitto Discourses from two material
	Continued Control
	Deserve Coloring
	Concent serementer: Optimize for More dissimilar sequences (inceptions megaflast) Osnevelasti sindar sequences (distor) Choose a BLAST algorithm. @
	BLAST Search database Nucleotide collection (unit) using Megablast (Optimize for highly similar sequences) Deoversuits in a new window
e commente de la constante des las las las las las las las las las la	Algorithm parameters Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign Detree default events carmeters

- Select the consensus of the contaminated sequence, copy it (Ctrl+c) and paste it into the FASTA window on the <u>NCBI GenBank BLAST page</u>

- \① Highly similar sequences (megablast)\BLAST\ a search window will appear, followed by a results window after ± 1 minute

- If the `Sequences producing significant alignments' do not match your aimed taxon, change the document color of the contaminated sequence to pink (contaminated/NUMT) in Geneious

4.11.2 Mix-ups

- Create a new alignment without contaminated sequences (paragraph 4.7) and a new ML tree (paragraph 4.8.1)



In this tree, there are no exceptionally long branches, indicating there are no contaminations. Furthermore, similar taxa names end up together, which indicates there are no mix-ups

-Save the tree as .pdf \File \Save As Image File... \Width: 1000 \Height: 1500 \Format: PDF Image

4.12 CHECK CODING* GENES FOR STOP CODONS

*For non-coding genes (e.g. 16S and ITS), proceed to paragraph 4.10

4.12.1 Full-length sequences

- Translate Translate entire sequences Genetic code (Invertebrate Mitochondrial for COI) Translation frame 2 for Folmer based COI Treat first codon as start of coding region/Modify sequence names

- Search for stop codons, depicted as `*' or use Ctrl+f for `*'

4.12.2 Non-full-length sequences

- Select relevant nucleotide alignment
- -\File\Export\Selected Documents...\Files of type: FASTA\Export\
- -\ \bigcirc Upper case\ \bigcirc Ns\
- Open with <u>AliView</u>
- \View\Show as translation (Ctrl+t or translation button)
- Reading frame (2 for COI), Translation table (5. Invertebrate Mitochondrial for COI)
- \sum button for stop codon count



4.13 EXPORT FASTA

- Select relevant nucleotide alignment
- -\File\Export\Selected Documents...\Files of type: FASTA\Export\
- \mathbf{O} Upper case
- \bigcirc gaps (-) \land for non-coding data \bigcirc Ns \land for coding data e.g. COI, matK, rbcL

4.13.1 Split FASTA into individual FASTA files (optional)

- Open the relevant FASTA in BioEdit
- -\File\Export\Split into individual fasta files\

5 ADDITIONAL PROCEDURES

WI SP006 FOLDER STRUCTURE DNA BARCODING DRIVE

6 CALCULATIONS

7 QUALITY CONTROL

8 RELATED/ASSOCIATED DOCUMENTS

http://www.geneious.com/tutorials

PCR_template_v4.3.xls

9 LITERATURE

PhyML 3.0:

"New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0."

Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O.

Systematic Biology, 59(3):307-21, 2010.

Katoh, Standley 2013 (Molecular Biology and Evolution 30:772-780)

MAFFT multiple sequence alignment software version 7: improvements in performance and usability.

(outlines version 7)



Appendix 4. Gel analysis of CTAB and Kit extraction

Figure 18. Results of the amplification with ITS for the kit extraction, undiluted, 5x diluted and 10x diluted.

003 004 006 0	50 057 057	La 003	004 006 05	0 057 057	La
		-			
		Ξ			Ξ
				N	I
		=_			III
Undiluted		•	5x diluted		
003 004 006 0	50 057 057	1			
003 104 000 0	50 .057 037	<u>La</u> (1013		i cero cero	
1			(1) · ·		
A Cate		_			
and a second sec		the second se			
1 Dellar					
					• •
10x diluted					• •

Figure 19. Results of the amplification with ITS for the CTAB extraction, undiluted, 5x diluted and 10x diluted.

All samples were analyzed on a 1% agarose gel and stained with ethidium bromide.

Appendix 5. Standard PCR program ITS2

PCR reaction mix:		ITS-br	ITS-bryophytes	
Brand	Chemical	Conc.	reaction (µL)	
Ultrapure	mQ		12,8	
	Buffer	10 X	2,0	
Qiagen	MgCI2		1,0	
	Betaine		0,25	
	DMSO		0,2	
	primer 28	10 pMol/ul	0,90	
	primer 29	10 pMol/ul	0,90	
	dNTP	2, mM	0,8	
	Таq		0,20	
	Template		1	
	Total		20,0	

ITS 5.8F-25R PCR program: Number of cycles: 40 Temp. Time 95 °C Init. Denature 0:05:00 95 °C Denature 0:00:20 Annealing 57 °C 0:00:30 Extension 72 °C 0:01:30 Final Extension 72 °C 0:07:00 12 °C Pause ∞

Figure 20. PCR program used for the first amplification of ITS2.
Appendix 6. PCR test with different additives

10 different master mixes were used to test several PCR additives.

1			2		
Chemical	Reaction (ul)	Per5+2	Chemical	Reaction (ul)	Per5+2
mQ	10.8	75.5	mQ	11.7	82.2
Buffer	2.0	14.0	Buffer	2.0	14
MgCl2	1.0	6.7			
Betaine	0.25	1.8	Betaine	0.25	1.8
DMSO	0.2	1.4	DMSO	0.2	1.4
Primer F	0.9	6.3	Primer F	0.9	6.3
Primer R	0.9	6.3	Primer R	0.9	6.3 5.6
dNTP	0.8	5.6	dNTP	0.8	
Tag 0.2 1.4		Taq	0.2	1.4	
3					
3 Chemical	Reaction (ul)	Per5 +2	4 Chemical	Reaction (ul)	Per5+2
3 Chemical mQ	Reaction (ul)	Per5+2 77.2	4 Chemical mQ	Reaction (ul)	Per5+2 83.9
3 Chemical mQ Buffer	Reaction (ul) 11 2.0	Per5+2 77.2 14	4 Chemical MQ Buffer	Reaction (ul) 12 2.0	Per5+2 83.9 14
3 Chemical mQ Buffer MgCl2	Reaction (ul) 11 2.0 1.0	Per5+2 77.2 14 6.7	4 Chemical mQ Buffer	Reaction (ul) 12 2.0	Per5+2 83.9 14
3 Chemical mQ Buffer MgCl2 DMSO	Reaction (ul) 11 2.0 1.0 0.2	Per5+2 77.2 14 6.7 1.4	4 Chemical MQ Buffer DMSO	Reaction (ul) 12 2.0 0.2	Per5+2 83.9 14 1.4
3 Chemical mQ Buffer MgCl2 DMSO Primer F	Reaction (ul) 11 2.0 1.0 0.2 0.9	Per5+2 77.2 14 6.7 1.4 6.3	4 Chemical MQ Buffer DMSO Primer F	Reaction (ul) 12 2.0 0.2 0.9	Per5+2 83.9 14 1.4 6.3
3 Chemical mQ Buffer MgCl2 DMSO Primer F Primer R	Reaction (ul) 11 2.0 1.0 0.2 0.9 0.9	Per5+2 77.2 14 6.7 1.4 6.3 6.3	4 Chemical MQ Buffer DMSO Primer F Primer R	Reaction (ul) 12 2.0 0.2 0.9 0.9	Per5+2 83.9 14 1.4 6.3 6.3
3 Chemical mQ Buffer MgCl2 DMSO Primer F Primer R dNTP	Reaction (ul) 11 2.0 1.0 0.2 0.9 0.9 0.9 0.8	Per5+2 77.2 14 6.7 1.4 6.3 6.3 6.3 5.6	4 Chemical mQ Buffer DMSO Primer F Primer R dNTP	Reaction (ul) 12 2.0 0.2 0.9 0.9 0.9 0.8	Per5+2 83.9 14 1.4 6.3 6.3 5.6

Chemical	Reaction (ul)	Per5 +2	
mQ	10	69.9	
Buffer	2.0	14	
MgCl2	1.0	6.7	
BSA	1.0	7	
<u>Betaine</u>	0.25	1.8	
Primer F	0.9	6.3	
Primer R	0.9	6.3	
dNTP	0.8	5.6	
Taq	0.2	1.4	

76.6
14
7
1.8
6.3
6.3
5.6
1.4

7			8		
Chemical	Reaction (ul)	Per5+2	Chemical	Reaction (ul)	Per5+2
mQ	10.2	71.6	mQ	11.2	78.3
Buffer	2.0	14	Buffer	2.0	14
MgCl2	1.0	6.7			
BSA	1.0	7	BSA	1.0	7
		-			
Primer F	0.9	6.3	Primer F	0.9	6.3
Primer R	0.9	6.3	Primer R	0.9	6.3
dNTP	0.8	5.6	dNTP	0.8	5.6
Taq	0.2	1.4	Tag	0.2	1.4
	2				

9			10		
Chemical	Reaction (ul)	Per5+2	Chemical	Reaction (ul)	Per5 +2
mQ	11.7	81.8	mQ	12.2	85.3
Buffer	2.0	14	Buffer	2.0	14
BSA	0.5	3,5			
Primer F	0.9	6.3	Primer F	0.9	6.3
Primer R	0.9	6.3	Primer R	0.9	6.3
dNTP	0.8	5.6	dNTP	0.8	5.6
Tag	0.2	1.4	Tag	0.2	1.4
2					-

Figure 21. All the different master mixes that were used to test the additives.

The PCR program that was used to test the additives.

PCR program:	ITS2 Iris		
Number of cycles:	40		
Init. Denature Denature Annealing Extension Final Extension Pause	Temp. 95 °C 95 °C 65 °C 72 °C 72 °C 72 °C	Time 0:05:00 0:00:20 0:01:30 0:01:30 0:07:00 ∞	

Figure 22. The PCR program used for the additive test.

	Pa_Pilopogon_atricanus		
	P_5quipayae_B_102		
			C040_silfer_Brazil 0.08 C078_pilfer_lameilabus_Peru
			C094_pillor_Peru
			Campyopus_enioldes_Irta078
-	1		Campylopus_flexuosus_G033860
			Campylopus_huallagensis_EI442_
			Campyopos_sentedoes_inauzo
			C004_thwalesii_South_Africa
			C009_clavatus_South_Africa
			C012_pitef_Costa_kita C016_crateris_Reunion
			C019_dichrosting_Brazil
			Coze_seriumaturi_twataa
			C027_schmidi_Sil_Lanka C028_angustratis_Guyana
			C029_angustiretis_Paraguay
			C032_griseus_Brazil
			C041_colongus_Peru C041_colongus_Brazi
			C043_germatus_Brazil
		0.96	C052_johanismeyen_DR_Congo
			C059_gradiis_Srl_Lanka
			Cost_sharpi_Peru
			C084_widgreni_Guyana C064_widgreni_Guyana
			C067_subjugorum_Venezuela
			Coro_forum_court_entra
			C075 hunlingenet Colombia
			C0/3_incassabus_C0/incassabus_Chile
			C085_alkidovirens_Peru C085_alkidovirens_Peru
			C002_decani_Zimbabwe
			Cutos_lankinus_indonesia C098_perpusitus_Cameroon
			C080_reflexisehus_Costa_Rica
			Cool and the second sec
			C115_Isemilinerits_Brazil C119_schmidl Reulon
			C120_pillet_Reurion
			0.53 C126_pillor_Social_entra
			C134_piller_thay C135_piller_Amera
			C136_piller_Acces
			C143_piller_Cape_Verde
			C146_piller_Socota
			C159_pilfer_Honduras
			C161_dexuosus_France C167_cubensis_Surinam
			Citig facelous_South_Africa
			C105_bblogus_Peru
			C190_tragliformis_Brazi
			C195_fragilis_Ecuador
			C200_pyrtomis_USA C201_feruosus_China
			C205_flexuosus_incacorralis_Venezuela
			C215_piller_Madeira
			C234 sublidatus Madeira
			C232_bitle_Modeira
			C240_subulbus_Madetra C245_t_pildr_Acores
			C246_1_piller_Acores
			C246_0_Units_Cooles C254_8_Units_Acores
			C266_fragilis_Acores
			C_arctocarpus_SP_26
			C_nexuosus_M0_30 C_pupureccaulis.New_Zealand_159
			1 P_guadaloupensis,P_99
			r_seve_c_roo
			C007 australia C054_carolinae_lvory_Coast
			0.51 C014_australia_
			0.88 C185 robilitade CR wanda
			0.7 C037 richardi Venezuela
			0.56 C057_pmfe_Mexico
			CI37_pnite_USA
			C199 Tolly Tolly C199 Tolly Control C199
			068 Collaphia Line Collaphia
			C062_talutensis_Colombia Campylogus_IfTa_062_umbellatus_
			0.98 Campylopus J/Tab60_umballatus_
			Col1_exasperatus_indonesia Col1_exasperatus_Philippines
			C002_Involutas_China
			0.95 C149_biolor_Australia
			C100_biter_Australia
			C122_piller_Reunion
			C236_pilfer_Madeira
			C097_pyriformis_South_Africa
			0.6 C179_pyrformis_Paraguay
			0.69 C197_pytformis_South_Africa
			C025_fragilis_zollingerianus_indonesia
			0.85 0.69 C177_tallulensis_Honduras
			C000_introfexus_South_Africa
			C085_introflexus_Australia C124_introflexus_Nathariants
			C217_introflexus_France
			0.00
			0.99 CiCh1_introflexus_Chile C_introflexus_Trin_174
			Cicht_introferus_Chie C_introferus_Trin_174
			Cicht_introfexus_Chile Cicht_introfexus_Chile Cicht_introfexus_New_Zoaland Cicht_introfexus_USA
		_	699 GiCh1_Initoflexus_Chile C_intoflexus_Trim_174 CGP2_intoflexus_Germany CGP1_intoflexus_USA

Appendix 7. Phylogenetic trees of helices ITS2

Figure 23. Phylogenetic tree Helix 1 of ITS2.

Pa_Pilopogon_africanus			
P_laevis_B_100 P_tiouloavae B 101 CF richardii			
P_tiquipayae_B_102	C018 richardii Colombia		
	C037_richardii_Venezuela C046_richardii_Peru		
		campylopus_l cres C008_umbellatus_Indonesia	irTa060_umbellatus_
	L	0.84 Campytopus_IrTa C017_exasperatu	_062_umbellafus_ is_Philippines
			C083_incrassatus_Chile
			Campylopus_flexuosus_0D33860 Campylopus_fragilis_0D33852
			Campylopus_huallagensis_El442_ Campylopus_sericeodes_IrTa026
			Campylopus_tallulensis_EH020_ C004_thwaitesii_South_Africa
			C005_thwallesii_South_Africa C009_clavatus_South_Africa
			C012_pilifer_Costa_Rica C016_crateris_Reunion
			C027_schmidil_Sri_Lanka C030_cuspidatus_Colombia
			C052_johannismoyeri_DR_Congo
			C059_gracilis_Sri_Lanka G057_pil/Ber_Mexico
		0.95	C064_wagrenir_ouyana
		0.85	C070_nageimerus_rniippines C073_acuminatus_kirkii_Australia
			C095_laxitextus_indonesia
			C119_schild_Reunion
			C134_piller_lary
			C136_DIMF_ACTOR
			C138_pillfer_Cape_Verde
			C159_piller_Honduras C161_flexuosus_France
			C167_cubensis_Surinam C170_pauper_Bolivia
			C189_fragilis_Ecuador C189_fragilis_Canary_is
			C190_fragilformis_Brazil C191_goughii_Sri_Lanka
			U-59C195_fragilis_Ecuador C200_pyriformis_USA
			C201_flexuosus_China C205_flexuosus_incacorralis_Venezuela
			C210_flexuosus_incacorralis_Colombia C225_oilifer_Madeira
			C235_pilifer_Madeira C236_pilifer_Madeira
			C241_pilfer_Madeira C245_1_pilfer_Azores
			C246_1_pillfor_Azores C240_0_pillfor_Azores
			C254_8_pilfer_Azores C266_tragilis_Azores
			C_arctocarpus_MG_27 C_arctocarpus_SP_28
			C_purpureocaulis_New_Zealand_158
			C001_carolinae_Brazil
			C054_carolinae_tvory_Coast
			Contraction Contra
			C014_australis_Australia C028_angustiretis_Guyana
			C062_tallulensis_Colombia
			C055_huallagensis_Colombia
			C031 C115 Jamellinervis_Brazil
			C094_pillfer_Peru
			C169_faccidus_South_Africa
			C041_oblongus_Peru
			0-03 C060_heterostachys_Peru C063_sharpii_Peru
			C021_robilardei_8outh_Africa
			C185_robiliardei_cf_Rwanda
			C048_anderssonil_Bolivia
			C185_oblongus_Peru
			C120_plifer_Reunion C122_plifer_Reunion
			C132_plifer_Socotra C132_plifer_France
			C146_pilifer_Socotra
			C023_savannarum_Rwanda
			C150_savannarum_Colombia
			C_flexuosus_MG_38 C043_gemmatus_Brazil
			C067_subjugorum_Venezuela 0.88 0-750086_albidovirens_Peru
			C108_nivalis_Rwanda C098_reflexisetus_Costa_Rica
			0.86 C234_subulatus_Madeira C239_subulatus_Madeira
			C240_subulatus_Madeira C025_fragilis_zollingerianus_Indonesia
			C089_cryptopodioides_Brazil_occultus_R C071_fragilis_Australia
			C097_pyiformis_South_Africa C177_tallulensis_Honduras
			C179_pyrtformis_Paraguay C197_pyrtformis_South_Africa
			C_Introflexus_Trin_174 C_pyriformis_New_Zealand_157
			C000_introflexus_South_Africa
			U.59 CiCht_introflexus_Chile CiDe2_introflexus_Germany
			0.82 CIN21_introflexus_New_Zeatand CIUs1_introflexus_USA
			0.58 C005_Introflexus_Australia C217_Introflexus_France
		0.03	

Figure 24. Phylogenetic tree Helix 2 of ITS2.



Figure 25. Phylogenetic tree Helix 3 of ITS2.



Figure 26. Phylogenetic tree helix 4 of ITS2.



Appendix 8. Phylogenetic tree of ITS2

Figure 18. the phylogenetic tree of the ITS2 region with the new samples.

Appendix 9. PCR results Hot start PCR



Figure 28. Results of the amplification of the normal hot start amplification and the touchdown hot start amplification. analyzed on a 1% agarose gel.