Faculty of Applied Ecology, Agricultural Sciences and Biotechnology

Ardian Sinani

Master Thesis

Development of microsatellite markers to study genetic diversity in populations of the smooth newt and great crested newt in Norway

Master's Degree in Applied and Commercial Biotechnology

01 June 2018

Consent to lending by University College Library	YES ☒	NO ☐
Consent to accessibility in digital archive Brage	YES ☒	NO ☐

Acknowledgement

First, I would like to thank my supervisor Robert C. Wilson for his guidance and valuable advice throughout the entire project. Also, I would like to thank the co-supervisors Lisa Lunde and Arne Linløkken. I am grateful for your never-ending support, excellent and enthusiastic supervision.
Thanks to the teachers and laboratory personnel working at the Inland Norway University of Applied Sciences.

Thanks to all my family members and friends who supported me during the whole duration of my master's studies. A special thanks to the QUOTA program.

I dedicate this thesis to my fiancé, Valentina Krivenjeva. Thank you for your support, assistance and motivation in completing this thesis. Love you!

Table of Contents

ABSTRACT **7**

1. INTRODUCTION **8**

1.1 SMOOTH NEWT AND GREAT CRESTED NEWT 8

1.1.1 Distribution *9*

1.1.2 Life cycle and diet *10*

1.1.3 Habitat choice *11*

1.1.4 Threats and action plan *11*

1.2 GENETIC MARKERS 12

1.3 MICROSATELLITES 13

1.3.1 Mutation of microsatellites *15*

1.3.2 Application of microsatellite markers *16*

1.3.3 Advantages and disadvantages of microsatellites *17*

1.4 ACQUIRING MICROSATELLITES 19

1.4.1 Searching for existing microsatellite markers *19*

1.4.2 Isolating new microsatellite markers *20*

1.5 MULTIPLEX PCR, A COMMON TECHNIQUE FOR GENOTYPING 21

1.5.1 Challenges with Multiplex PCR *21*

1.5.2 Multiplex primer design *22*

1.6 AIM OF THE STUDY 24

2. MATERIAL AND METHODS **25**

2.1 SAMPLING PROTOCOL 26

2.2 DNA ISOLATION OF SMOOTH NEWT AND GREAT CRESTED NEWT 26

2.2.1 Isolation of DNA from tissue *26*

2.2.2 *Isolation of DNA from larval tissue* 27

2.3 DEVELOPMENT OF MICROSATELLITE MARKERS 27

2.3.1 *Selection of microsatellites* 27

2.3.2 *Collecting and processing of the sequence data* 27

2.3.3 *Microsatellites mining* 28

2.4 PRIMER DESIGN 28

2.4.1 *Primer validation* 29

2.4.2 *Agarose gel electrophoresis* 29

2.5 MULTIPLEX DESIGN 30

2.5.1 *Singelplex primer testing* 30

2.5.2 *Visualisation and analyses of the PCR product* 31

2.6 SEQUENCING OF PCR AMPLICONS 32

3. RESULTS 33

3.1 DNA ISOLATION 33

3.1.1 *Isolation of DNA from larval tissue* 33

3.1.2 *Isolation of DNA from tissue* 33

3.2 DEVELOPMENT OF NEW PRIMERS FOR MICROSATELLITE MARKERS 34

3.2.1 *Marker choice* 34

3.2.2 *Quality assesment of the de novo assembly* 35

3.2.3 *Marker identification* 37

3.3 PRIMER DESIGN 38

3.3.1 *Primer validation* 41

3.4 THREE-PRIMER SINGLEPLEX MARKER TESTING 47

3.5 MULTIPLEX OPTIMIZATION AND FRAGMENT ANALYSES 51

3.6 SEQUENCING ANALYSES 55

4. DISCUSSION **57**

4.1 *DE NOVO* ASSEMBLY 57

4.2 DEVELOPMENT OF MICROSATELLITE MARKERS 57

4.2.1 Microsatellite development based on established markers 57

4.2.2 Microsatellite isolation from genomic sequences 59

4.3 SEQUENCING ANALYSIS 61

5. CONCLUSION **62**

6. REFERENCES **63**

APPENDIX **71**

Abstract

The population of amphibians is known to decline in rapid speed and newt species fall in that category. Development of microsatellite markers has the potential to estimate genetic variation in newt populations and investigate the effect of human intrusive, which causes changes to their habitat. The main goal of this study was the development of microsatellite panels for use in both species identification and determination of genetic diversity in populations of smooth newt (Lissotriton vulgaris) and great crested newt (Triturus cristatus) in the southern region of Norway.

Two different strategies were utilized to generate sequences of potential microsatellite loci. The program MPprimer designed primers from already validated microsatellite markers and from sequences identified with GMATo mining tool from publically available genomic sequences. Two primer sets were amplified by a PCR reaction followed by agarose gel electrophoresis. Successfully amplified primers were employed to develop a multiplex PCR with the help of the program Multiplex Manager 1.2. Furthermore, primer pairs with confirmed amplification were tested on singleplex PCR with a three-primer approach. The markers that showed a clear ability for allele calling were grouped and optimized on different multiplex conditions with the help of Applied Biosystems 3130xl Genetic Analyzer accompanied by GeneMapper (v 5.0).

Unfortunately, the number of successfully established markers in this study was low with only six markers showing potential for genotype use. Therefore, we were unable to conduct a genotyping experiment and determine the genetic diversity of newt populations.

1. Introduction

This study transpired at Inland Norway University of Applied Sciences in the Biotechnology Section of the Department of Natural Science and Technology. The focus of this thesis was molecular genetics with the core research on developing novel DNA based genotyping technologies. Main goal was to develop already validated microsatellite markers and identify new markers from publicly available genomic sequences, potentially, creating one set of microsatellite markers with the purpose of developing new panels for genotyping. In this study, attention was drawn to two salamander species that are present in Norway, the smooth newt (*Lissotriton vulgaris*) and the great crested newt (*Triturus cristatus*).

1.1 Smooth newt and Great crested newt

The smooth newt and great crested newt belong to the family Salamandridae, which consists of true salamanders and newts. Salamanders that live in semiaquatic environments are known as newts (Bell & Lawton, 1975; Griffiths & Teunis, 1996). Great crested newt has a moderately slender body, medium-sized legs, and a narrow tail base. The skin is warty in texture. Breeding males develop a jagged crest along the body and tail and this is deeply extended at the base of the tail (*see Fig. 1*). Females have a length of about 16 cm and weigh 6 to 15 g. Males are lighter and are in the range of 14 to 15 cm in length (Edgar & Bird, 2006).

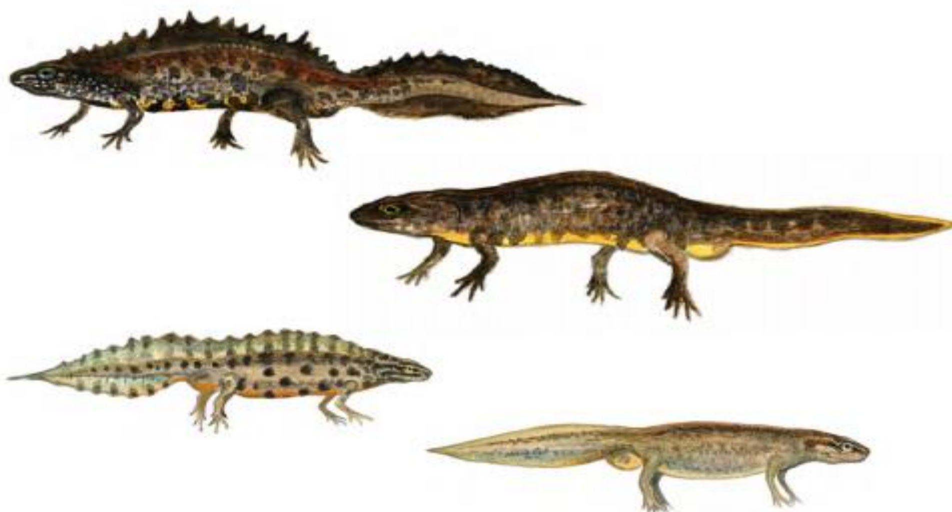


Figure 1*: Images of great crested newts* (Triturus cristatus) *above and smooth newts* (Lissotriton vulgaris) *below; males to the left and females to the right, retrieved from Gustafson* (2011).

Smooth newt females and nonbreeding males are pale brown or olive green, often with two darker stripes on the back. Both sexes have orange bellies, although paler in females covered with black spots. During the breeding season, male smooth newts develop a continuous wavy crest that runs from their heads to their tails (*see Fig. 1*). Both sexes are of similar size, around 10 cm in length (T. Beebee & Griffiths, 2000; Griffiths & Teunis, 1996).

1.1.1 Distribution

The smooth newt habitat ranges from Ireland and Great Britain in the west, through central Europe and Scandinavia, south to Italy, the Balkans and northern and western Turkey, and further east in Ukraine and Russia, from sea level up to 2700 m in altitude (Bell, 1977; Pabijan et al., 2015; Skorinov, Kuranova, Borkin, & Litvinchuk, 2008; Ben Wielstra, Bozkurt, & Olgun, 2015). Great crested newts live mainly in northern and central Europe it is present also in Great Britain. However, it is absent from Ireland, the Iberian Peninsula, southwest France and most of Scandinavia (Babik et al., 2009; B Wielstra, Baird, & Arntzen, 2013). The latest version of newt distribution atlas is available on Societas Europaea Herpetologica (SEH), from where a map showing newt distribution in Norway (*see Fig. 2*) was retrieved (Sillero et al., 2014). Moreover, new updates are expected to be made available during this year (2018).

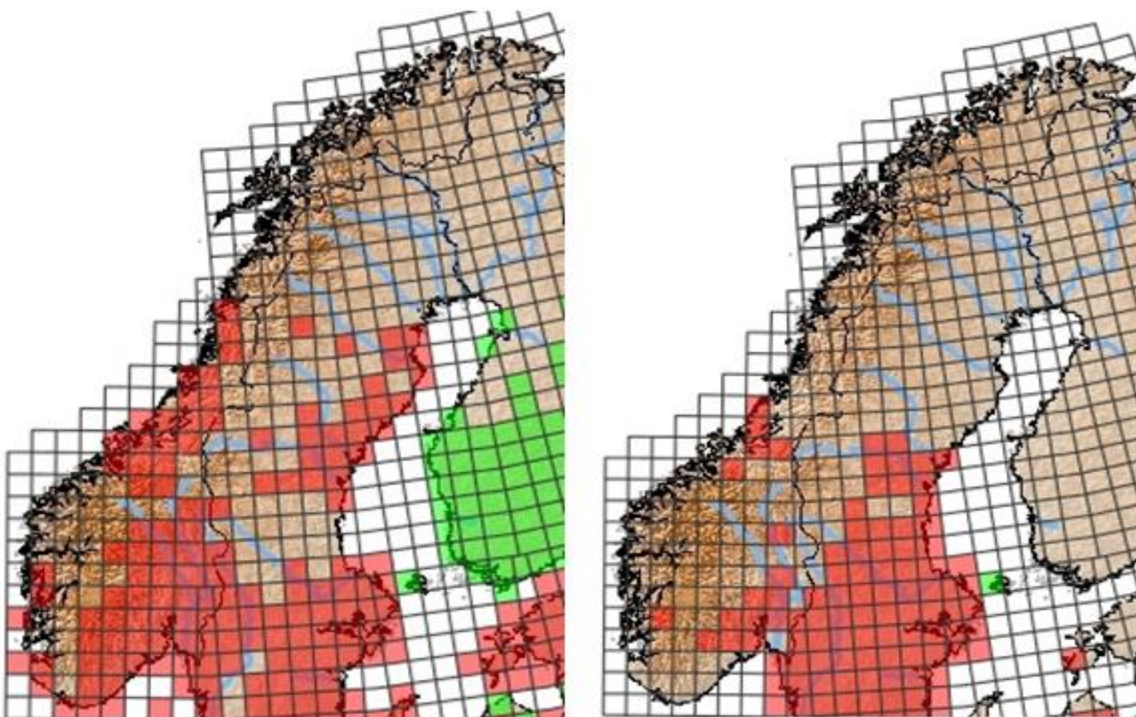


Figure 2: *Distribution of the newt species in Norway (red coloured squares). Smooth newt* (Lissotriton vulgaris)*, left and great crested newt* (Triturus cristatus)*, right* (Sillero *et al., 2014).*

In Norway, as it can be seen above in figure 2, newts live in the eastern region from the Swedish border and around the Oslo fjord to Skien southwest of Oslo stretching up to the middle of Telemark County, and then in Lillehammer in Oppland and Ytre Rendal in Hedmark. There are also some populations around Bergen, Nordmøre, and Trondheim (Dervo, Pedersen, & Bærum, 2016; Dolmen, 1983b).

1.1.2 Life cycle and diet

Newts have a complex life cycle; adults enter small water bodies in early spring, in which they court, mate, lay eggs and where the larvae develop (Babik, Szymura, & Rafiński, 2003; Bell & Lawton, 1975). Mating in newts takes place in water and involves elaborate courtship, which culminates in the transmission of a spermatophore that is deposited on the ground from a male and then picked up by a female (Babik et al., 2003; Halliday, 1990). The female lays 5 to 15 eggs per day wrapped in leaves one by one, and she can lay 200 to 300 eggs per season. In Scandinavia, breeding and egg-laying take place from early April until late June (Malmgren, 2001). The larvae are released from the eggs 2 to 3 weeks after, and the metamorphosed newts will quickly move to the protective terrestrial environment (Dervo et al., 2016). The annual survival rate of adult smooth newts is around 50 %, being rather greater for females than for males. Less reliable data indicate the annual survival rate of juveniles to be around 80 % (Bell, 1977). In another study where great crested newt was colonized to a newly created pond and studied in a six-year period, the survival rate was reported to be 22 % for juveniles and 49 % of adults (Arntzen & Teunis, 1993). Adult newts spend most of their life cycle in terrestrial habitats, searching for food and daytime hiding places, and for hibernation during winter (Jehle & Arntzen, 2000; Malmgren, 2002). In the autumn and winter, newts live in hibernation, mainly in soil caves, under stones or fallen trees. Depending primarily on temperature, overwintering begins in October/November and ends in February/March (Kinne, 2004).

Newts are carnivorous in all stages of life, eating primarily live prey: mostly insects, molluscs, worms, fish, and amphibians (Deban & Wake, 2000). The diet of smooth newt and great crested newt consists mainly of planktonic animals, and to a lesser degree with benthic animals. Zooplankton is important pray for all newt size classes, but the modes of feeding of the species are different (Dolmen & Koksvik, 1983). Considering the fact that both newt species have similar habitat and nutritional requirements, makes their co-existence to be threatened at times of insufficient food supply: great crested newt then begins to use smooth newts as prey (Kinne, 2004).

1.1.3 Habitat choice

The great crested newt is rarely found in the absence of the much common smooth newt, but seems to prefer sites that are relatively large and deep, and that had a high proportion of open water surface (Cooke & Frazer, 1976; Skei, Dolmen, Rønning, & Ringsby, 2006). Smooth newts have lower demands, thus they breed also in small reservoirs of standing water, mainly man-made sites such as deep wheel-ruts and puddles, water-filled clay and gravel-pits, and occasionally in larger dams (Babik et al., 2003). However, both species tend to breed in ponds having abundant aquatic vegetation (Cooke & Frazer, 1976). According to Maletzky et al, (2007) key habitat features are low degree of shading for adults and larvae as well as high density of submerged vegetation for larvae, with no significance of pond depth. Aquatic adults of great crested newt are mainly bottom-living, while smooth newt adults spend much of their time in the open water. The larvae show an opposite pattern (Dolmen, 1983a). On leaving the pond, the newts migrate above ground, covering relative large distances, that takes place in all directions, but with significant preferences towards certain habitat type. They prefer areas with bushes, hedgerows, and trees, avoiding pastures and other open areas (Jehle & Arntzen, 2000). A similar preference of habitats was reported by Vuorio et al, (2015) in a study carried out in the southeast of Finland, where both newt species distinctly favored forests with high understory vegetation cover, adjacent to the ponds. Both species show similar habitat use, with an indication of the great crested newt avoiding open areas and having more need for the shelter provided by vegetation, especially when the distance to pond increases. According to Müllner (2001), the juveniles of both species prefer the same sites as the adults, indicating some hereditary preference for suitable land habitats. Suggesting that terrestrial habitat should be an integrated part of every conservation strategy for newts.

1.1.4 Threats and action plan

Smooth newt is listed in the least concern category of IUCN Red List and included in the Bern Convention and Annex IV EU Natural Habitat Directives. Great crested newt holds the same status. The potential threats to both newt species are habitat loss and fragmentation, chemical pollution, eutrophication and early desiccation of their breeding sites, deforestation, agricultural development, urbanization, introduced fishes and diseases (T. J. Beebee & Griffiths, 2005; Dolmen, Skei, & Blakar, 2008; Martel et al., 2014; Skei et al., 2006). Although many declines are due to habitat loss and overutilization, other, unidentified processes threaten 48 % of rapidly declining species and are driving species most quickly to

extinction (Stuart et al., 2004). Based on the results of his study, Malmgren (2001) suggested that the great crested newt may be more prone to local extinction than previously believed, much due to genetic constraints and possible habitat specialization. According to Denoël (2012), great crested newt is the most threatened newt species, whereas smooth newt has an intermediate decline level. Moreover, the great crested newt is not only regionally threatened but suffers from a global decline in Western Europe. However, (Denoël, Perez, Cornet, & Ficetola, 2013) argues that populations of the common species may follow the fate of the rarest species and should not be neglected in conservation programs. Because environmental trends are leading to a deterioration of aquatic and terrestrial habitat features required by all newt populations.

As part of the conservation strategy, 360 km^2 were established as a reference area for amphibians in central Norway, the hydrography of 341 lentic water bodies was surveyed and the occurrence of the great crested newts and smooth newts in the area was investigated (Skei et al., 2006). The number of great crested newt was doubled in the period 2005 – 2009 compared to the previous 5-year period and further threefold in the next 5-year period from 2010 – 2014. The increase is most likely due to observation effort and the establishment of the service “Species observation” at the Norwegian Biodiversity Information Centre (Dervo et al., 2016). According to the Wildlife Act, newt species are considered wild in Norway and therefore fall under the conservation principle. It is forbidden to capture, hunt, kill or hurt protected wild organisms.

1.2 Genetic markers

The genetic marker usually is a gene or DNA sequence with a known location on a chromosome that can be used to identify individuals or species. It can be described as an observed variation (polymorphism, which may arise due to mutation or alteration in the genomic loci). It is known that all organisms are subject to mutations as a result of normal cellular operations or their interactions with the environment, which leads to genetic variation (Liu & Cordes, 2004). Genetic polymorphism is the occurrence of alternative DNA sequences (alleles) at a locus among individuals, groups or populations, at a frequency greater than 1% (Benavides & Guénet, 2012). In order to be useful to geneticists, this polymorphism must be either recognizable as a phenotypic variation or as a genetic mutation distinguishable through molecular techniques (Liu & Cordes, 2004). Genetic markers are divided into two types, protein, and DNA (molecular). The known protein markers are

hemoglobin used in earlier studies and allozymes as a more common marker for genetic variation studies (Abdul-Muneer, 2014; Liu & Cordes, 2004). Molecular markers can be a useful tool to determine many genetic characteristics such as i) measuring local gene flow and migration, ii) assigning individuals to their most likely population of origin, iii) measuring effective population size through the between-generation comparison of allele frequencies, and iv) detecting past demographic bottlenecks through allele frequency distortions (Jehle & Arntzen, 2002). Based on their transmission and evolutionary dynamics molecular markers can be classified into two groups, mitochondrial DNA and nuclear DNA markers (Park & Moran, 1994). Mitochondrial DNA (mtDNA) markers are inherited only from mother avoiding recombination. Therefore, their genetic effective population size (Ne) is only one quarter compared to nuclear markers (Ferguson & Danzmann, 1998). Restriction enzymes can be used to cut of mtDNA sequence at specific sites, generating restriction fragment length polymorphisms (RFLPs) or sequence analysis of different genes of mtDNA are mainly used to study phylogenetic relationships, pedigree analysis and population differentiation (Abdul-Muneer, 2014). Nuclear marker types that are finding service in the field of genetic studies include randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), single nucleotide polymorphism (SNP), expressed sequence tag (EST) markers, and microsatellites (Arif et al., 2011; Liu & Cordes, 2004).

1.3 Microsatellites

Microsatellites are “simple sequence repeat” (SSR) of a DNA sequence, with a repeat size of 1–6 base pair (bp) repeated several times (*see Fig. 3*), and flanked by regions of non-repetitive unique DNA sequences (Tautz, 1989). They are also called as “short tandem repeat” (STR) DNA sequences (Edwards, Civitello, Hammond, & Caskey, 1991).

>Lv.19, motif $(TCAA)_6$

GGTAAGGGTCATGCCATTTACTAAGGCTTTGTAGCCACCGGGGAGAGAGAGATTGATAGA

AGGGTCAGCAGGAGACTTTTCTCAATCAATCAATCAATCAATCAATCCGGGAATTTGTAA

AGCGCACTACTTACCCGCTAGGGTCTCAAGGCGCTGGGGGGGGGGGGGGAGCTGCAGCTA

CTGGTCGAAGAGCCAGGTCTTGAG

Figure 3*: Example of a microsatellite sequence retrieved from publicly available sequence data. The microsatellite sequence is 124 base pair long, with a tetranucleotide motif (TCAA) repeated six times.*

According to Moore et al. (1991), the microsatellites flanking regions are conserved across species as diverse as primates, artiodactyls, and rodents. Microsatellites are distributed throughout the human genome, accounting for about 3 % of the entire genome. However, their distribution within chromosomes is not quite uniform, they appear less frequently in subtelomeric regions (Koreth, O'Leary, & J, 1996). Most SSRs are found in noncoding regions of the genome, while only 8 % of SSRs are found in coding regions (Ellegren, 2000a). SSRs are classified based on their unit lengths, such as mono-, di-, tri-, tetra-, penta- and hexanucleotide repeats, with dinucleotide as the most common unit repeats (International Human Genome Sequencing, 2001). The overall SSR density is comparable across the human chromosomes. The density of different repeats shows significant variation. With tri- and hexanucleotide repeats being more abundant in exons, whereas other repeats are copious in non-coding regions (Subramanian, Mishra, & Singh, 2003). Another study lead by Temnykh et al.(2001) found out that GC-rich trinucleotides are abundant in exons of the rice genome, whereas AT-rich trinucleotides are disturbed evenly throughout the genomic components, such as coding sequences, untranslated regions (UTRs), introns and intragenic spaces.

As mentioned above, microsatellites are abundant in the genomes of higher eukaryotes, and previously were considered as biological non-functional DNA sequences or "junk DNA" (Nadir, Margalit, Gallily, & Ben-Sasson, 1996). However, thanks to genome sequencing the important role SSRs might play in genomes is being elucidated (Vieira, Santini, Diniz, & Munhoz, 2016). Microsatellites have biological functions that can lead to phenotypic changes. These can be variation in intronic SSRs, which can affect gene transcription, mRNA splicing, or export to the cytoplasm. Moreover, changes in trinucleotide microsatellites in UTR regions or introns can induce heterochromatin-mediated-like gene silencing (Y. C. Li, Korol, Fahima, & Nevo, 2004). There is growing evidence that SSR variation can affect relevant biological processes, such as the regulation of transcription and translation, organization of chromatin, genome size and the cell cycle (Gao et al., 2013; Y. C. Li et al., 2004; Nevo, 2001). In addition, Lawson & Zhang (2008) suggested that SSRs may have an effect on gene expression and may play an important role in contributing to the different expression profiles of housekeeping and tissue-specific genes.

1.3.1 Mutation of microsatellites

Microsatellites are known to be highly polymorphic and well distributed. This is possible due to their high mutation rate, which is considered to be 10^{-2} to 10^{-6} nucleotides per locus per generation (Ellegren, 2000b), compared to unique eukaryotic DNA sequences mutate a rate of approximately 10^{-9} nucleotide per generation (Crow, 1993). Even though SSRs were identified at the beginning of the 1970s, its mutation mechanisms remains poorly understood. Up to date, three possible mechanisms have been proposed: (i) unequal crossing over in meiosis; (ii) retro-transposition mechanism; (iii) strand-slippage replication (Fan & Chu, 2007). Among these mechanisms, strand-slippage replication (*see Fig. 4*) proposed by Levinson & Gutman (1987), appears to be widely regarded as the main pattern of SSR mutation.

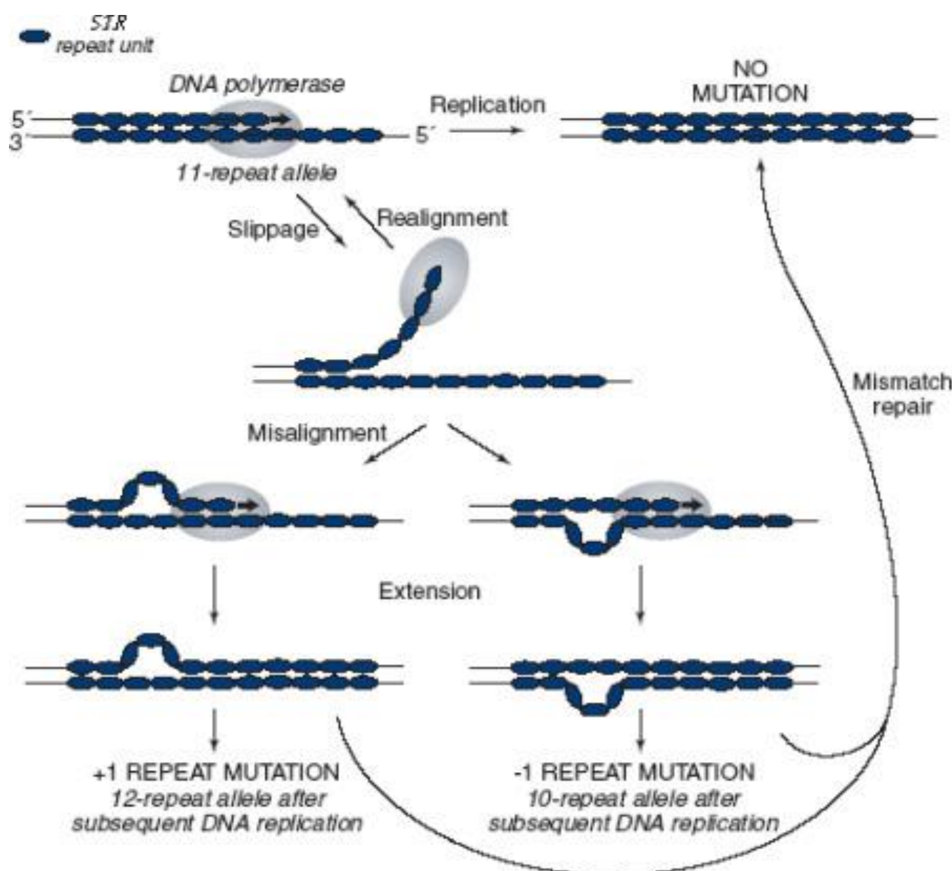


***Figure 4**: Illustration of the strand-slippage mechanism (SSM) at simple sequence repeat (SSR) mutation site. During replication process slippage of the DNA polymerase from the DNA template can occur and replicated strand can reanneal out-of phase, which entails looped-out bases. When DNA replication continues, a shorter or longer product is obtained. These SSM errors are often corrected by exonucleolytic proofreading and pre-dominantly mismatch repair* (Fan & Chu, 2007).

One of most regarded model of microsatellite evolution is stepwise mutation model (SMM) proposed by Ohta & Kimura (1973) in which, upon a mutation the SSR region either gains or loses one repeat unit. A similar model was later suggested by Di Rienzo et al.(1994) called the two-phase model (TPM). Considering this pattern, mutations in the SSR sites can cause changes greater than one repeat unit. Earlier was suggested, that the rate of mutation depends on a sequence specific slippage rate, but is independent of the length of the fragments being synthesized (Schlötterer & Tautz, 1992) or the allele size (Valdes, Slatkin, & Freimer, 1993). Later, Ellegren (2000a) reported that mutation rate within locus increases with allele length, but is not affected by the size difference between an individual's two alleles (allele span). Moreover, long alleles tend to mutate to shorter lengths, preventing infinite growth. This is possible due to incorporation of point mutation in models of microsatellite evolution. For instance, using a Markov chain model, Kruglyak et al.(1998) showed that equilibrium distributions of microsatellite repeat lengths can be explained by a balance between slippage events and point mutations. In theory, selection could also act against long alleles. However, microsatellites are mainly located within noncoding DNA and there are so far no data to suggest possible selective constraints on particularly long alleles (Ellegren, 2000b).

1.3.2 Application of microsatellite markers

Microsatellites are widely used in scientific and applied research. One of the first fields SSRs found application was forensic science. Thanks, to the discovery of minisatellite DNA fingerprinting (Jeffreys, Wilson, & Thein, 1985a, 1985b), the forensic science was revolutionized and became the norm for the determination of relationships for both humans and animals. With usage of the polymerase chain reaction (PCR) process in SSR analyses, Tautz (1989) found out that they show extensive length polymorphisms. Moreover, Tautz proposed “These simple sequence length polymorphisms (SSLP) may be usefully exploited for identity testing, population studies, linkage analysis and genome mapping”. The almost random distribution of microsatellites and their high level of polymorphism was utilized for the construction of genetic map of the mouse genome (Dietrich et al., 1994) and the Généthon human linkage map (Dib et al., 1996). It was this period of time when microsatellites were established as the marker of choice for the identification of individuals and paternity testing (Schlötterer, 2000). Moreover, SSR markers have been used successfully in studies of amphibian effective population sizes and structures, and in assessing the consequences of hybridization (T. Beebee, 2005).

Microsatellite-based techniques can also be applied in genome scans and quantitative trait loci (QTL) mapping in natural populations to investigate the genetic basis of adaptive selection and biodiversity in many species (Rogers & Bernatchez, 2005). Another field SSRs show great potential is conservation and management of wild species. For this purpose, (Banhos, Hrbek, Gravena, Sanaiotti, & Farias, 2008) have characterized the microsatellite loci of the Neotropical harpy eagle, in view of providing a set of molecular tools for the conservation and management of wild and captive harpy eagles. The broad areas of applications of microsatellite markers are depicted below in Figure 5.

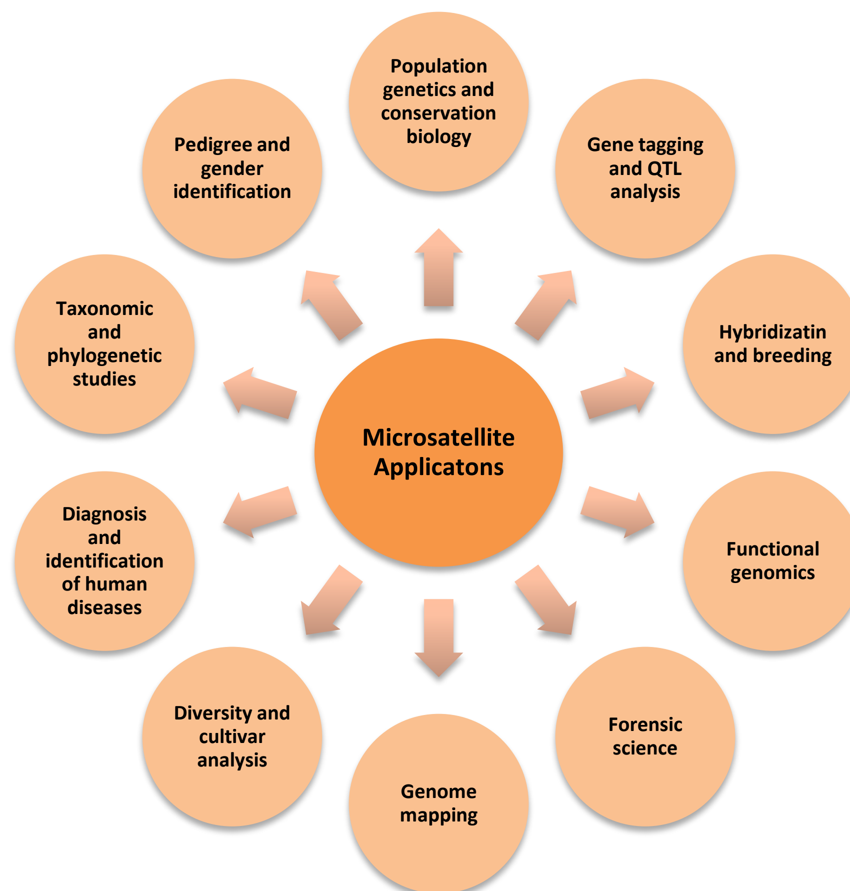


Figure 5*: Applications of microsatellite markers in different areas, modified from Abdul-Muneer* (2014).

1.3.3 Advantages and disadvantages of microsatellites

Microsatellites have many advantages as neutral nuclear marker. They are in Mendelian manner inherited and codominant markers. Other strengths of SSR markers are: their abundance, ubiquity, even genomic distribution, polymerase chain reaction (PCR) typability, small locus size, and high polymorphism (Koreth et al., 1996; Y. C. Li et al., 2004; O'reilly & Wright, 1995). A major advantage of the SSR method is that it can be automated.

In addition, SSR loci are co-dominant markers, what makes them more informative than RAPDs, RFLPs, and AFLPs. Moreover, SSRs show greater polymorphism (Russell et al., 1997). Also, microsatellites exhibit higher levels of polymorphism and abundance in genomic DNA compared with allozyme markers (Schlötterer, 2000). Comparison of SSRs and SNPs in the assessment of genetic relatedness in maize showed that SSRs provide more information on genetic diversity and performed better on estimating relative kinship (Yang et al., 2011). In addition, the success rate of cross-amplification for SSRs in closely related species is typically higher than for SNPs (Guichoux et al., 2011). The high sensitivity of PCR-based SSR analysis advanced forensics and facilitated other research areas, such as the analysis of samples with limited DNA amounts or degraded DNA (Schlötterer & Pemberton, 1998). In another study, the reproducibility of three popular molecular marker techniques (RAPDs, AFLPs, and SSRs) was examined by Jones et al. (1997). They reported that RAPDs were difficult to reproduce. For AFLPs, a single-band difference was observed in one track, whilst SSR alleles were amplified by all laboratories, but small differences in their sizing were obtained. The bottleneck in SSR analysis of the allele sizing is eliminated with the availability of high-throughput capillary sequencers or mass spectrography (Schlötterer, 2000). Moreover, Y. Li et al. (2007) have demonstrated the usefulness of the two multiplexed microsatellite systems for genetic diversity studies on two populations of black tiger shrimp.

Challenges may arise when using microsatellite markers, like with any other technique. Some disadvantages include the appearance of shadow or stutter bands and presence of null alleles, which are existing alleles that are not observed using standard assays (Mohindra, Mishra, Palanichamy, & Ponniah, 2001). The size homoplasy problems in PCR-based microsatellite assays may affect the inference of recent population history (Estoup, Jarne, & Cornuet, 2002). Also, microsatellite flanking regions sometimes contain mutations which may produce identical length variants that could compromise microsatellite population level studies (Zardoya et al., 1996). Due to this mutation, only some cross-species amplification of microsatellites is possible between closely related species (Arif et al., 2011). Each primer set rarely works across taxonomic groups, so primers are usually developed anew for each species (Selkoe & Toonen, 2006). Therefore, new species-specific microsatellite markers have to be isolated whenever a species is studied for the first time, a process that is both time consuming and expensive (Arif et al., 2011).

1.4 Acquiring microsatellites

The development of microsatellite markers can basically be divided into the following stages: (i) prior knowledge of nucleotide sequences in which SSRs occur; (ii) design of oligonucleotides (primers) complementary to the regions flanking the SSR; (iii) validation of primers by PCR and electrophoresis of the product of the reaction, and (iv) detection of polymorphisms among individuals (Mason, 2015). However, there are two possible pathways for generating SSR markers. The first one involves literature review of published work for any existing microsatellite for the target species and closely related species while the second is based on available sequence date and sequencing technologies for isolating new markers.

1.4.1 Searching for existing microsatellite markers

One way of developing microsatellite-based study is to search published literature for existing microsatellites for the target species and closely related species. The availability of microsatellite markers for a given species usually is a combination of past interest in that species or related species and the inherent success rate of microsatellite development for that taxon (Selkoe & Toonen, 2006). The frequency of microsatellite shows differences in regions of the genomes of plants, animals, fungi, and prokaryotes (Tóth, Gáspári, & Jurka, 2000), and this variety of frequencies in the genome dictates the success rate of isolating microsatellite markers (Zane, Bargelloni, & Patarnello, 2002). Flanking regions can be conserved across taxa, allowing cross-species amplification of microsatellite loci using SSRs developed from other species in the same genus or even family, especially for vertebrates such as fishes, reptiles and mammals (Rico, Rico, & Hewitt, 1996). In addition, Moore et al. (1991) found that microsatellites flanking regions were conserved across species as diverse as primates, artiodactyls, and rodents. For instance, Schlotteröer et al. (1991) proved that amplification of homologous loci can be successfully performed in various species of toothed (Odontoceti) and baleen (Mysticeti) whales, even though the estimated divergence in time are 35–40 million years. The success rate of cross-amplification for SSRs in closely related species is high, based on the research done by Sharma et al. (2007), who reported up to 50 % success rate. However, success rate of primers may decrease proportionally to the evolutionary distance between the present species and the species of origin (Primmer, Møller, & Ellegren, 1996). In addition, allelic diversity usually decreases when primers are used in non-source species (Ellegren et al., 1997). Generally speaking, attempting

amplification by using existing primers from related species is less expensive and time-consuming than isolating new ones (Squirrell et al., 2003).

1.4.2 Isolating new microsatellite markers

Microsatellite detection requires available sequence data. Originally, microsatellite mining from sequence databases involves the model of match hit search using Basic Local Alignment Search Tool for nucleotide query sequence (BLASTN) or using tools with similar algorithms (Temnykh et al., 2001). Nowadays, more sophisticated and user-friendly microsatellite mining software's have been developed, such as Tandem-Repeats Finder by Benson (1999), MISA (MIcroSAtellite) by Thiel et al. (2003), SSR Locator by Da Maia et al. (2008) and Genome-wide Microsatellite Analyzing Tool (GMATo) by Wang et al. (2013). A common drawback of such algorithms is that they are heavily biased to mine exact tandem repeats or perfect repeats (Sharma et al., 2007). The process of isolating new microsatellites has been streamlined with technological advances and protocol optimization to make the process cheaper, efficient and more successful (Zane et al., 2002). Microsatellites were developed from both coding and non-coding regions of plant genomes, and several sources were used to search for SSRs, including a variety of DNA libraries (genomic, genomic-enriched for SSR, bacterial artificial chromosome and cDNA libraries), as well as public databases, including expressed sequence tag (EST) databases (Vieira et al., 2016). Earlier, the only possibility to identify sequences harboring SSR motifs was the screening of size-fractionated genomic DNA or of EST libraries (Zane et al., 2002). However, regardless of whether genomic or EST sequences are used for SSR detection, traditional laboratory methods involving cloning, cDNA library construction and Sanger sequencing are costly and time-consuming (Squirrell et al., 2003). Later, next generation sequencing (NGS) techniques have started to be used to identify sequences harboring SSR motifs in non-model species (Allentoft et al., 2009). NGS has been very useful for various studies, including prospecting for new SSR markers. Successors of the Sanger sequencing method include the 454 FLX (Roche), Solexa (Illumina), SOLiD (Applied Biosystems) and HeliScope True Single Molecule Sequencing (Helicos) platforms. Third generation platforms are also currently available, including platforms developed by Pacific Biosciences (PacBio) and Oxford Nanopore, based on new sequencing technology SMRT and MinION sequencing respectively, which have the advantage of producing longer DNA reads (Vieira et al., 2016).

1.5 Multiplex PCR, a common technique for genotyping

The amplification of small fragment of DNA through polymerase chain reaction (PCR) gained popularity with the advent of thermo cyclers (Abdul-Muneer, 2014). The polymerase chain reaction (PCR) was originally developed in 1983 by the American biochemist Kary Mullis. Development of DNA amplification using the PCR technique has opened the possibility of examining genetic changes (Ferguson & Danzmann, 1998). The small amount of tissue required for PCR-based genotyping enables the collection of minimal tissue samples by tail-tip or toe clipping, a procedure that has been shown to be relatively harmless in the species study (Arntzen, Smithson, & Oldham, 1999). In a multiplex PCR, more than one target sequence can be amplified by including more than one set of primers in the reaction. This will save considerable amount of time and effort within the laboratory without the loss of results (Elnifro, Ashshi, Cooper, & Klapper, 2000). A multiplex assay was developed by Freeman et al. (2003) enabling analysis of a series of unlinked, highly polymorphic microsatellite loci to be genotyped in a single PCR, which includes a combination of locus-specific primers. Initially, the assay was composed of five markers, increasing it progressively to 12 such markers. Multiplex PCR is a very powerful and widely used genotyping technique that enables amplification of two or more products in a single reaction. It simultaneously amplifies multiple regions of a DNA template or multiple DNA templates using more than one primer set comprising of forward and reverse primer in single tube (Shen et al., 2010). Multiplexing of microsatellite markers can be performed either using a single dye labeling for amplicons of different size or different dye labeling for amplicons of same size (Arif et al., 2011).

1.5.1 Challenges with Multiplex PCR

The development of an efficient multiplex PCR usually requires prior planning and optimization of reaction conditions and components. For a successful multiplex PCR assay, the concentration of the primers and PCR buffer, a balance between the magnesium chloride and deoxynucleotide concentrations, cycling temperatures, the amount of template DNA and enzymes should be considered (Markoulatos, Siafakas, & Moncany, 2002). In advance, all primer pairs have to be tested in singleplex PCRs to estimate optimal annealing temperature to check for correct amplification of the desired fragments. Moreover, optimal primer concentration and thermo cycling conditions are key features to achieve balanced and stable multiplexing reactions (Sint, Raso, & Traugott, 2012). It is known that fragments targeted by

better performing primer pairs will be amplified preferentially (Markoulatos et al., 2002) or that amplification efficiency of general primers, which target a range of species, can vary between species (Sipos et al., 2007). This can lead to unbalanced amplification strength and differing detection limits among targets within and between multiplex's (Sint et al., 2012). Another obstacle for successful multiplexing can be PCR inhibitors. These inhibitors interact either directly with DNA or with DNA polymerase, preventing the amplification completely or reducing product yield (Wilson, 1997). In a multiplex PCR reaction, it is possible for a specific sequences to suffer from different inhibition effects to different extents, leading to disparity in their relative amplifications (Bessetti, 2007). Increasing the DNA polymerase amount in the reaction and the use of Bovine Serum Albumin (BSA) can help overcoming PCR inhibitors (Bessetti, 2007).

1.5.2 Multiplex primer design

Once the sequences harboring repeat motifs have been identified, suitable primers must be designed. To develop high-quality multiplexed SSRs, stringent selection of markers is necessary (Varshney, Graner, & Sorrells, 2005). This makes primer designing for multiplex PCR a challenging problem and several factors need to be considered such as: miss-priming due to nonspecific binding to non-target DNA templates, primer dimerization, and the inability to separate and purify DNA amplicons with similar electrophoretic mobility (Shen et al., 2010). Primer pairs that amplify fragments of different sizes (e.g. about 100, 200 and 300 bp) should be chosen to permit amplification of several non-overlapping markers with a single dye (Guichoux et al., 2011). Different computer programs that simultaneously identify SSRs and design primers for multiplex have been developed such as MuPlex (Rachlin, Ding, Cantor, & Kasif, 2005), STAMP (Kraemer, Gäbler-Schwarz, & Leese, 2009) and MPprimer (Shen et al., 2010). One of the most regarded programs for reliable multiplex PCR primer design is MPprimer. This program combines the primer design features of Primer3 and the primer specificity evaluation of MFEprimer. The tool creates specific, non-dimerizing primer set combinations with constrained amplicon size (Shen et al., 2010). Genotyping by utilizing fluorescent labels has revolutionized molecular marker-based analysis over the past decade, which has allowed a more rapid data collection compared to earlier methods such as those based on radioactive isotopes/autoradiography and silver staining (Blacket, Robin, Good, Lee, & Miller, 2012; Guichoux et al., 2011). Earlier methods used for resolving microsatellite alleles were based on automated systems using fluorescently labeled PCR fragments that allow very precise allele calling. In these methods, one of the primers used in

the PCR is 5' end labeled with a fluorescent dye (Bonin et al., 2004; Schwengel, Jedlicka, Nanthakumar, Weber, & Levitt, 1994). However, the cost of synthesizing fluorescently labeled primers is a limiting factor for many labs, as labeled primers cost between five and ten times more than unlabeled ones (Missiaggia & Grattapaglia, 2006). To reduce the costs of genotyping with fluorescently labeled microsatellites a novel PCR strategy was proposed by (Oetting et al., 1995), called multiplexing with tailed primers, which employs a forward primer with a 19 bp extension at its 5' end, identical to the sequence of an M13 sequencing primer, a regular reverse primer and a third universal fluorescent labeled M13 primer. A similar, but more advanced method was developed by (Blacket et al., 2012) known as "three primer approach" involving four universal primers (*see Tab. 1*), each labeled with respective fluorescent dye.

Table 1*: Universal primers with their respective multiplex dye (fluorophore), tail sequence, tail length, sequence GC content and primer melting temperature, retrieved from* (Blacket et al., 2012).

Universal Primer	Multiplex dye	Tail Sequence (5'-3')	Tail Length (bp)	GC Content (%)	Primer T_m
Tail A	FAM	GCCTCCCTCGCGCCA	15	80	63 °C
Tail B	VIC	GCCTTGCCAGCCCGC	15	80	57 °C
Tail C	NED	CAGGACCAGGCTACCGTG	18	67	59 °C
Tail D	PET	CGGAGAGCCGAGAGGTG	17	71	59 °C

To ensure the success of co-amplification, it is critical to eliminate primers with potential primer–dimer interactions (Vallone & Butler, 2004). Performing a local blast or using dedicated tools such as Multiplex Manager (Holleley & Geerts, 2009) can help for this purpose (Guichoux et al., 2011).

1.6 Aim of the study

Panels of microsatellite markers have been developed for two newt species present in Norway, smooth newt (*Lissotriton vulgaris*) and great crested newt (*Triturus cristatus*). The majority of the markers are species-specific while others are amenable for genotyping both species. The main goal of this master's thesis was to continue the development of these panels, identifying and adding new microsatellite markers. Two different sources were used for retrieving possible SSR markers. The first source was different published articles, which have isolated and validated microsatellites from closely related newt species. The second source was publically available genomic sequences from the smooth newt. The project entailed also DNA extraction from a variety of sample types including larval tissue (tail clips) and tissue of "road killed" adult newt individuals. Finally, two or three panels had to be established, collectively harboring ca. 12-16 common microsatellite markers and 2-4 species-specific markers. These microsatellite markers would be used for species identification and determination of genetic diversity in populations of both newt species in the southern region of Norway.

2. Material and Methods

One of first material obtained was DNA samples, extracted from larvae tail clips and tissue from "road killed" adult newt individuals, and quantified in a NanoDrop Spectrophotometer. Then sequences harbouring SSRs were selected, either from validated markers or from publicly available sequence data. Primers were designed and initially tested with uniplex. Successful primers were 5' end labelled with universal primers and tested in singleplex. Depending upon the sizes and colour of the labelled primers, they were grouped into multiplexes and optimized in five-coloured laser induced fluorescence capillary electrophoresis system. The schematic workflow of this study is presented in Figure 6 below.

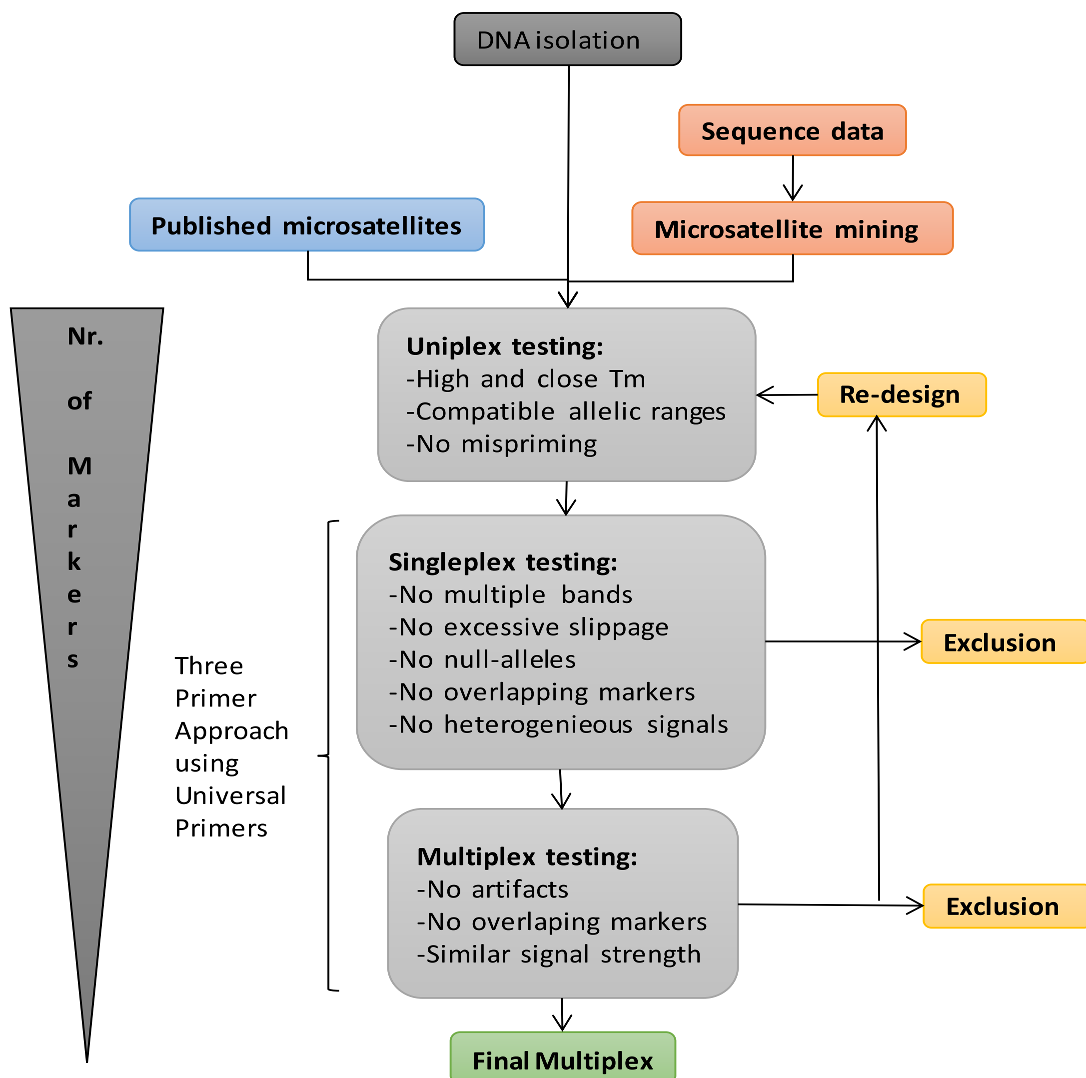


Figure 6: *Schematic of the workflow in this study, modified from Guichoux et al.* (2011).

2.1 Sampling protocol

The samples collection was performed in compliance with the rules set by Directorate for Nature Management (Direktoratet for Naturforvaltning). In the county of Buskerud, eggs were sampled from plastic ribbons (1x40 cm) attached to rods that were put into the bottom on 1 m depth, with the ribbon attachment at 40-50 cm depth, and floating towards the surface. In the study ponds in the county of Hedmark, this method was unsuccessful, and samples were collected by collecting larvae by means of a landing net, removing the tail tip of larvae. The samples were stored in 96 % ethanol at -20 °C. It was collected larvae tails from Solheim (SOL) and Bjørsrud (BJO) and eggs from Lahell (LAH) and Vivelstad (VIV). The samples were collected in two places in each pond and at two occasions, except in LAH where the water level sank due to tapping for irrigation purposes, leaving the egg sampling ribbons in the air before the second sampling. Also, a limited number of adult newts accidentally killed "road killed" were collected in Lier area. These adult newt samples were stored at -20 °C in plastic bags. Sampling was carried out mostly during the breeding period, when newts return to the ponds.

2.2 DNA Isolation of Smooth newt and Great crested newt

Isolation of DNA was performed using E.Z.N.A. ® Tissue DNA Kit protocol (OMEGA BIO-TEK), May 2013. This kit is suitable for the isolation of DNA from up to 30 mg animal tissue or tail snips. However, yields may vary depending on source. Purified DNA can be directly used for most applications such as PCR, Southern blotting, and restriction enzyme digestion.

2.2.1 Isolation of DNA from tissue

Samples from different individuals of the smooth newt and great crested newt were used for DNA extraction. Only a small part of a tissue was sliced from the newt "road killed" adult individuals and weighted to be around 30 mg. This tissue was added to an Eppendorf tube and DNA was then isolated according to the standard protocol. Samples that did not lyse properly were excluded. The quality and purity of DNA were determined using NanoDrop® ND-1000 Spectrophotometer. DNA isolated successfully was stored at -20 °C.

2.2.2 Isolation of DNA from larval tissue

The larval tissue (tail clips) samples weighed 5 to 20 mg. Therefore, they were directly transferred to an Eppendorf tube. DNA isolation was done using the same protocol mentioned above. Some tiny unsolvable parts were noticed but did not hinder the success rate of the isolation protocol. Quantity and quality of DNA were measured with NanoDrop. The isolated DNA was stored at -20 °C.

2.3 Development of Microsatellite markers

One set of microsatellite markers was initiated from already determined and evaluated microsatellite loci. The published SSR markers are validated either against the species of our interest (smooth newt and great crested newt) or for other closely related species. Generally, a large number of SSR markers are needed in the first stage of the development. Therefore, more markers were selected from different published articles. Another group of microsatellite markers was set to commence from publically available genomic sequence data. This data hold the basic information for identification of new markers.

2.3.1 Selection of microsatellites

Three sets of microsatellite loci from different published articles served as voluntaries for development of SSR markers. First set included seven microsatellites isolated by Nadachowska et al. (2010) for the carpathian newt (*Lissotriton montandoni*). These SSR markers were evaluated successfully for cross-species amplification on the smooth newt species. The second set contained eight microsatellite loci for the greek smooth newt (*Lissotriton vulgaris graecus*) determined by Sotiropoulos et al. (2009). Third selected set holds fifteen microsatellites developed by Drechsler et al. (2013) for the palmate newt (*Lissotriton helveticus*).

2.3.2 Collecting and processing of the sequence data

The National Center for Biotechnology Information (NCBI) is the main database that provides free access to biomedical and genomic information. In the Sequence Read Archive (SRA) are available genomic sequence data of a bio-project (Accession: PRJNA214312) conducted by Institute of Environmental Sciences, Jagiellonian University, Poland. This bio-project includes data of 20 experiments from four *Lissotriton* sister species. The data of eight experiments (Accession: SRX333586, SRX333587, SRX333590, SRX333591, SRX333594, SRX333595, SRX333596, and SRX333597) which contained sequence reads of the smooth

newt species were imported into CLC Genomic Workbench (version 11.0) and a *de novo* sequence assembly was done, creating a single fasta file. The CLC assembler was based on the de Bruijn graphs (Compeau, Pevzner, & Tesler, 2011). *De novo* assembly in CLC Genomics Workbench was performed by setting up various parameters (Mapping Mode = Create simple contig sequences, Automatic bubble size = Yes, Minimum contig length = 200 bp, Automatic word size = Yes, Auto-detect paired distances= Yes, and Perform scaffolding = Yes). A quality assessment was performed for the assembly with web-based QUAST program which is a quality assessment tool for genome assemblies (Gurevich, Saveliev, Vyahhi, & Tesler, 2013).

2.3.3 Microsatellites mining

Genome-wide Microsatellite Analyzing Tool (GMATo) developed by Wang et al. (2013) was used for SSR mining and statistics at genome aspects. Since the GMATo tool requires a single input file containing raw DNA sequences, the fasta file generated by performing *de novo* assembly was utilized in this study. All three parameters were set in default, with the motif length range from 2-10 times, the minimum repeated times at 5 times and the option for highlighting microsatellite. The output files in tabular format listed all SSR loci information and statistical distribution at four classifications.

2.4 Primer design

Primers were designed with MPprimer software (version 1.4) from two sets of microsatellite loci, including the literature reviewed SSR markers and the newly identified microsatellite sequences. The MPprimer program is a well-known tool for designing specific, non-dimerizing primer set constrained amplicon size for multiplex PCR assay. It combines the primer design features of Primer3, dimer checking program named PriDimerCheck, and the primer specificity evaluation of MFEprimer (Shen et al., 2010). The MPprimer software is available in command line program with Linux language (Ubunto_64).

The parameters have been set as following:

- Primer length: 17-30, optimal: 22
- Melting temperature (TM): 59-62 °C, optimal: 60 °C
- G/C content: 30-70 %, optimal: 50 %
- Production size range: 100-400.
- Other parameters were set to default.

The MPprimer program provides also a scoring matrix for multiplex assessment, enabling manual removal of the template sequences that create conflict. The best 15 PSCs (primer set combination) in a user-friendly format were retrieved from MPprimer. A BLAST search was conducted on primers to ensure that the primers were specific to the target region. These primers were eventually ordered from Invitrogen™ using the Custom DNA Oligos service with the purity grade 'desalted'.

2.4.1 Primer validation

The designed primers were tested with uniplex to confirm if they were actually amplifying the target region of the microsatellite sequence. After receiving the primers, first they were dissolved with 0.1x TE-buffer (1 mM Tris-HCl (pH 8.0) & 0.1 mM EDTA) and later were diluted in different concentrations stocks by adding PCR water. Only five DNA template samples were used during the testing, including two smooth newt samples (Sos.16.02, LYSS or Lv.RK5) and two great crested newt samples (LAH_12 or Tc.RK6). To begin with, various PCR protocols were tried to compile a final PCR reaction model that will be used throughout the whole project. PCR was performed with a total volume of 15 µl including the following components: 2.0 mM MgCl2, 1x B1 buffer (Solis Biodyne, 2016), 0.1 µM forward and reverse primer, 0.2 mM dNTPs, 0.05U Hot Firepol® DNA polymerase (Solis Biodyne, 2016) and 3.33 ng genomic DNA template. For each primer pair there was also a non-template control (NTC) included. The PCR was run on Veriti 96 well Thermal Cycler (ABI) with the following program: Initiation activation on 95 °C for 10 minutes, followed by 35 cycles of denaturation on 95 °C for 15 seconds, annealing on 58 °C for 15 seconds and DNA synthesizing on 72 °C for 30 seconds, followed by a single extension step on 72 °C for 15 minutes and a final hold step at 10 °C ∞.

2.4.2 Agarose gel electrophoresis

The PCR products were analysed on a 1.5 % agarose gel 1x TAE (0.04 M Tris-acetate & 0.001 M EDTA) with 1.5 µl ethidium bromide (EtBr, 2 µg/ml) solution per 40 ml gel as fluorescent tag. For each tested primer pair a mix was prepared containing: 5 µl of the sample diluted with 5 µl H2O and 1.0 µl Loading buffer 5x (New England BioLabs® Inc.), which consists of 2.5 % Ficoll®-400, 11 mM EDTA, 3.3 mM Tris-HCl (pH 8.0), 0.017 % SDS and 0.015 % bromophenol blue at a 1x concentration, before loading it on gel wells. A 100 bp DNA Ladder (New England BioLabs® Inc.), which includes 12 different fragments with a size of respectively 1517, 1200, 1000, 900, 800, 700, 600, 500/517, 400, 300, 200 and

100 bp, was used for sizing and approximate quantification of the samples. The electrophoresis was run for 35 minutes at 90 Volts. The gel was visualized with Syngene G: BOX Chemi imaging system, supported by GeneSys software.

2.5 Multiplex design

The primer pairs that showed to be successful after performing gel electrophoresis by amplifying the target sequence regions were used in multiplex designing. In this particular population genetic study the three primer PCR approach according to Blacket et al. (2012) was preferred to end-label the PCR products. The use of four fluorescent dyes (*see Tab. 1*) fixed each on a different universal primer (U) makes possible co-amplifying and detection of greater number of microsatellite loci and size-overlapping markers. In this experimental setup, a universal sequence tail was added to the 5’ end of each forward primer creating so-called "forward tailed primers (Ft)" while the reverse primer was maintained. The tail sequences for each respective forward primer were chosen with the help of Multiplex Manager version 1.2 (Holleley & Geerts, 2009). The same commercial service provider, Invitrogen^{TM} synthesized all modified forward primers that were used in this study.

2.5.1 Singelplex primer testing

The modified forward primers were tested with uniplex together with respective reversee primers, to check if they still amplify the same target sequence region (*see 2.4.1/2.4.2*). Primer pairs that exceeded the uniplex test were deployed for further testing with singleplex. A three primer approach was applied in singleplex, thus the third primer with a fluorophore attached to its 5’ end was introduced as illustrated in Figure 7.

All the components of three-primer singleplex reaction were the same as described above for the unlabelled primer testing (*see 2.4.1*). With exception of the primer concentrations that here was 0.1 µM forward tailed primer (Ft), 0.2 µM reverse primer (R) and 0.1 µM universal primers (U). In addition, 1x Bovine Serum Albumin (BSA) was used for stabilization of the DNA polymerase. It also prevents adhesion of enzymes to the reaction tubes and tip surfaces. The PCR run was performed using the same program described above (*see 2.4.1*).

Markers showing good signal during singleplex testing were selected and classified in groups for multiplex optimization.

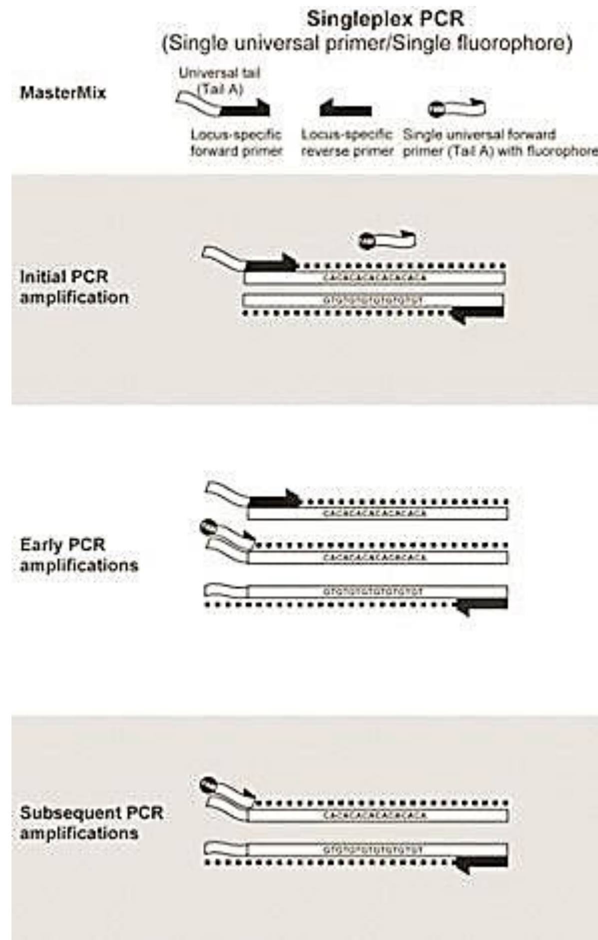


Figure 7: *Singleplex polymerase chain reaction (PCR) with a three primer approach. Amplification of PCR fragments with fluorescently labelled high annealing temperature universal primers and locus-specific tailed forward primers. Labelled universal primers begin to be incorporated into PCR fragments in early PCR cycles, tailed forward primers are exhausted in early cycles and subsequent PCR cycles incorporate fluorophores into PCR fragments* (Blacket et al., 2012).

Multiplex Manager was used to design the multiplexes. Different dyes were deployed for loci with an overlapping size range. The multiplex reactions were carried out in the same manner as the singleplex. However, the primer concentrations may vary between the markers depending on their signal strength.

2.5.2 Visualisation and analyses of the PCR product

The PCR products of the singleplex or multiplex were visualised with capillary electrophoresis. It was transferred 0.5 µl of the PCR product to a mix of 9.5 µl formamide and 0.1 µl 500 Liz™ size standards. The mix was preheated at 95 °C for 2 minutes in Veriti 96 well Thermal Cycler (ABI), and then cooled down by placing it directly on ice. A quick spin was done before loading the samples in the Applied Biosystems 3130xl Genetic Analyzer. The run module was Fragment Analysis, with the following settings: GS75-

300Liz (-250), G5_36cm_POP7_GS500. The pre-run voltage was held at 10.0 kV for 300 s. In addition, the run voltage was maintained at 15.0 kV during 1500 s. The implemented analysing method was “microsatellite default”.

The results were imported and analysed with GeneMapper® software (version 5.1). The electropherograms show the peak of each SSR marker in their size range with particular signal strength. Depending upon the heights of each allele the concentration of the primers was adjusted.

2.6 Sequencing of PCR Amplicons

Randomly were selected seven SSR markers for sequence analysis. These microsatellites were amplified in 15 µl PCR reaction set up using only one genomic DNA sample (Lv.RK5) with respective unlabelled primers. The BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) was used for sequencing. Prior to sequencing, the PCR amplified products were treated with Exonuclease1 (Exo I) to remove excess primer using 2 µl of PCR reagent, 2 µl of 5x sequencing buffer, 0.2 µl of Exo I (20 u/µl) and 5.8 µl of H_2O for a total volume of 10 µl. Samples were incubated at 37 °C for 60 min, 85 °C for 15 min and then held at 10 °C in a thermal cycler. Sequencing reactions were performed following Platt et al. (2007) St_ep method. Amplified fragments were sequenced in both directions (once using the forward primer, and once using the reverse primer) for greater accuracy of the base calls in the overlapping regions (repeated regions). The sequencing product was purified according the NaOAc/EDTA/EtOH precipitation method. The precipitated sequencing products were denatured in 10 µl deionized formamide and run in Applied Biosystems 3130xl Genetic Analyzer using BDv1_1_RapidSeq_POP7_1 run module.

3. Results

In this study, development of microsatellite multiplex panels was attempted, as a genotyping pipeline to determine genetic structure among the smooth newt and great crested newt populations in the southern region of Norway. Isolation of DNA was carried successfully from the larvae (tail clips) and tissue of adult newts "road killed" samples. The DNA samples were quantified with NanoDrop and tested by agarose gel electrophoresis. Microsatellite markers were selected and primers were designed or redesigned before testing them with PCR and gel electrophoresis for amplification. For the successful primer pairs, more testing was carried out using three-primer singleplex PCR approach. Markers that performed well in singleplex were grouped into potential multiplex panels for further testing and optimization.

3.1 DNA isolation

Isolation of DNA was performed successfully from 12 newt individuals, using E.Z.N.A. ® Tissue DNA Kit protocol. The DNA extraction was carried out in three different experiments. The samples were selected from two sources: larval tissue (tail clips) and tissue from adult newt "road killed" individuals.

3.1.1 Isolation of DNA from larval tissue

First, the DNA extraction was carried from four larval tissue samples (tail clips preserved in ethanol). Quantity and quality of DNA was measured with NanoDrop for each elution step and also after pooling together both elution steps *(see Appendix, Table A1).* The purity ratios, 260/280 and 260/230, were in each case to be found in the appropriate interval, 1.8 – 2.0 and 1.8 – 2.2 respectively. A good purity and quite high concentration from each DNA sample were observed. However, the yield was quite small.

3.1.2 Isolation of DNA from tissue

Here the DNA isolation in first case was performed from three *Lissotriton vulgaris* (LISS) tadpole samples (preserved in ethanol). The first sample (LISS 1) did not lyse properly, thus was excluded. The DNA extraction from two other samples (LISS 2 and LISS 3) was successfully performed. In another experiment DNA was extracted from six adult newt "road killed" individuals (one sample was from great crested newt and five from smooth newts). Overall the DNA was isolated successfully but some difficulties were experienced with

Tc.RK6 sample because the tissue was very dry, probably the animal was exposed to the sunlight for too long before collection. The quantity and quality of DNA were determined with NanoDrop *(see Appendix, Table A2 and A3).* Here the purity ratios, 260/280 and 260/230, were nearly in the appropriate interval for each DNA sample, 1.8 – 2.0 and 1.8 – 2.2 respectively. The concentration from each DNA sample was high also the overall yield was quite large. However, significant difference in DNA concentration was observed between the tissue samples preserved in ethanol and those stored at -20 °C in the freezer.

3.2 Development of new primers for microsatellite markers

Generally, a large number of microsatellite markers are needed in the first stage of the development. Therefore, different sources were exploited to generate enough SSR markers. One set of microsatellite markers was initiated from already established microsatellite loci. The published SSR markers are validated either against the species of our interest (smooth newt and great crested newt) or for other closely related species. Another group of microsatellite markers was set to commence from publically available genomic sequence data. This data hold the basic information for identification of new markers.

3.2.1 Marker choice

The first pathway to be considered for developing microsatellite is to search published literature for existing microsatellite loci for the target species and closely related species. The same was done in this study, where a lot of effort was put in the literature review to try and find established microsatellite loci. This effort resulted in selection of three different published articles that served as primary source for development of SSR markers. The first article included seven microsatellites isolated by Nadachowska et al. (2010) for the carpathian newt *(Lissotriton montandoni).* These SSR markers were evaluated successfully for cross-species amplification on smooth newt. The second article contained eight microsatellite loci for the greek smooth newt *(Lissotriton vulgaris graecus)* determined by Sotiropoulos et al. (2009). Third selected published work holds fifteen microsatellites developed by Drechsler et al. (2013) for the palmate newt *(Lissotriton helveticus).* In general, greater number of microsatellite loci is needed in the first stage of marker development. Therefore, a second strategy of bio-mining was employed for new microsatellite markers sequences selected from the SRA archives.

3.2.2 Quality assesment of the *de novo* assembly

Another pathway to secure and add more sequence targets for microsatellite development can be done by searching publicly available sequence data. The National Center for Biotechnology Information (NCBI) is one of the main databases that provide free access to biomedical and genomic information. However, there are limited genomic sequence data for the newt species included in this study. Therefore, the sequence reads of smooth newt species from eight experiments in SRA database were used in this study. These experiments were conducted with ILLUMINA technology (Illumina MiSeq) based on paired amplicon library selected with PCR by using genomic DNA as a source. However, a major drawback of this sequence data was that they were generated by targeting 3' untranslated transcript regions (3' UTR) but we decided to use those date since they were the only available genomic sequence data for the newt species of our interest. Another disadvantage of this data was that the sequence reads were approximately 150 bp in length, which sequence size is not appropriate for multiplex primer design. Thus a *de novo* sequence assembly was performed to generate more suitable sequences. The quality of the assembly was assessed on web-based QUAST program (Gurevich et al., 2013) and the report is presented in Table 2 below.

Table 2*: The quality assessment QUAST* (Gurevich et al., 2013) *report table of the* de novo *assembly showing all statistics based on contigs of size >= 100 base pair (bp), unless otherwise noted. The table also includes the total length of the assembly, Guanine/Cytosine (GC) content, and N50 values*

	Ardian140317.fa.txt
# contigs (>= 0 bp)	70
# contigs (>= 1000 bp)	1
# contigs (>= 5000 bp)	0
# contigs (>= 10000 bp)	0
# contigs (>= 25000 bp)	0
# contigs (>= 50000 bp)	0
Total length (>= 0 bp)	24482
Total length (>= 1000 bp)	1170
Total length (>= 5000 bp)	0
Total length (>= 10000 bp)	0
Total length (>= 25000 bp)	0
Total length (>= 50000 bp)	0
# contigs	70
Largest contig	1170
Total length	24482
GC (%)	45.20
N50	376
N75	288
L50	24
L75	43
# N's per 100 kbp	0.00

As anticipated, the total length of the assembly was relatively short (≈25 kb) composed by 70 contigs, although the N50 value of 376 bp was quite impressive. Likewise, the GC content was satisfactory with an average of 45 %, which is also presented in a diagram based on contig proportion in the following figure.

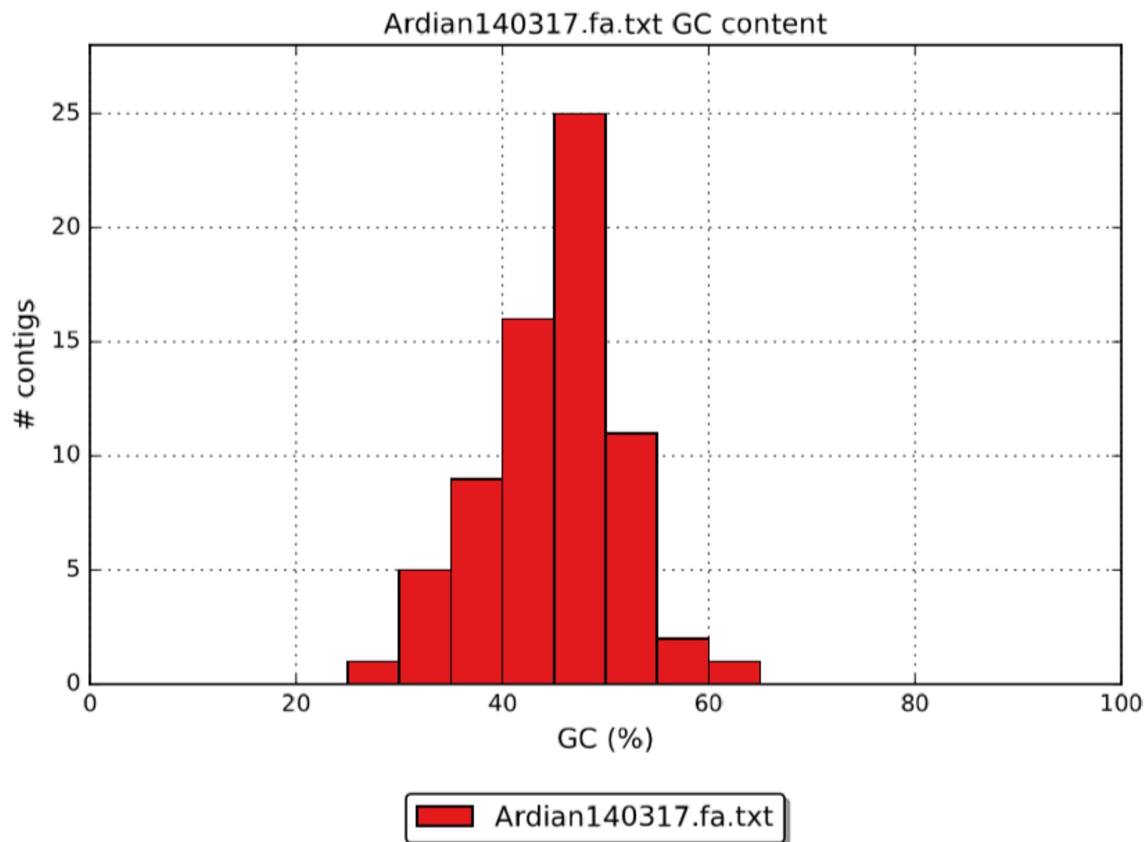


Figure 8*: The diagram showing information about the Guanine/Cytosine (GC) content of the assembly, provided by QUAST tool* (Gurevich et al., 2013*). The Y-axis shows the number of contigs while the X-axis shows the percentage of GC content.*

Another important characteristic that showed great improvement was the contig length which was the main reason to perform the *de novo* assembly. A diagram containing information about the contigs length of the assembly, provided by QUAST tool is presented in Figure 9 (next page).

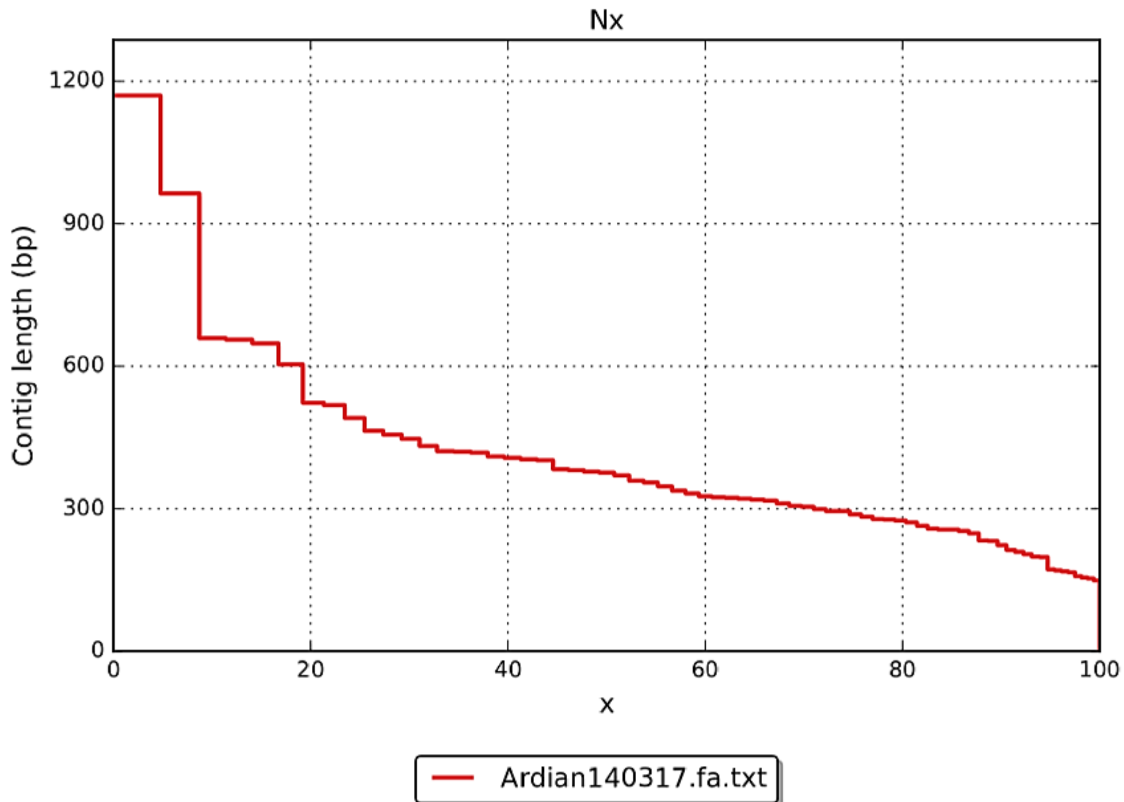


Figure 9*: The diagram showing information about the contigs length of the assembly, provided by QUAST tool* (Gurevich et al., 2013). *The Y-axis shows the contig length in base pair (bp) while the X-axis shows the percentage of the contigs with a particular length.*

The diagram presented in Figure 9 confirms that the majority of the contigs were in the desired length interval (200-500 bp) for multiplex primer design purpose. Thus, the fasta file generated by performing *de novo* assembly was more useful for the following process of bioinformatics analyses for identification of the target sequences of possible new microsatellite loci.

3.2.3 Marker identification

Genome-wide Microsatellite Analyzing Tool (GMATo) developed by Wang et al. (2013) was used for SSR mining and statistics at genome aspects. The GMATo tool is written in the Perl language as command line interface with the Java language providing graphic interface, and it is executable independently in Windows, Linux or Mac OS systems. The Windows package was more user-friendly and easy to use was therefore utilized in this study. Since this tool requires a single input file containing raw DNA sequences in fasta format, only the fasta file of *de novo* assembly was used as more appropriate input file. The output files listed all possible SSR loci information and statistical distribution. Initially, 90 sequences harbouring tandem repeats were selected by screened through the output files manually.

Later, the sequences with "cliff hanging" tandem repeats or flanking regions with repeated sequences were excluded since they were inappropriate for primer designing purpose. Finally, a list composed of 70 sequences showing the potential of new microsatellite loci identification was selected for primer design utility.

3.3 Primer design

Primers were designed with MPprimer software (version 1.4) from two sets of microsatellite loci, including the literature reviewed SSR markers and the newly identified microsatellite sequences. The MPprimer program combines the primer design features of Primer3, dimer checking program named PriDimerCheck, and the primer specificity evaluation of MFEprimer (Shen et al., 2010). These features make it a well-known tool for designing specific, non-dimerizing primer set constrained amplicon size for multiplex PCR assay. The MPprimer software is available in command line program with Linux language (e.g. Ubuntu). A scoring matrix for multiplex assessment provided by the program was used for manual removal of the template sequences that create conflict. Such a matrix is presented in Figure 10 with the scores < 0.2 showing compatible sequence templates whereas the scores ≥ 0.2 indicate incompatible template based on the primer design parameters chosen.

	A	B	C	D	E	F	G
1		sra_data3_contig_26776	sra_data3_contig_80925	sra_data3_contig_52307	sra_data3_contig_19806	sra_data3_contig_10921	sra_data3_contig_21969
2	sra_data3_contig_26776	0	0	0	0	0	0
3	sra_data3_contig_80925	0	0	0	0	0	0
4	sra_data3_contig_52307	0	0	0	0	0	0
5	sra_data3_contig_19806	0	0	0	0	0	0
6	sra_data3_contig_10921	0	0	0	0	0	0
7	sra_data3_contig_21969	0	0	0	0	0	0
8	sra_data3_contig_20817	0	0	0	0	0	0
9	sra_data3_contig_9678	0	0	0	0	0	0
10	sra_data3_contig_23233	0	0	0	0	0	0
11	sra_data3_contig_32884	0	0	0	0	0	0
12	sra_data3_contig_79784	0	0.2	0	0	0	0
13	sra_data3_contig_37268	0	0	0	0	0	0
14	NED_PET	0	0	0	0	0	0
15	sra_data3_contig_19796	0	0	0	0	0	0
16	sra_data3_contig_42256	0	0	0	0	0	0
17	sra_data3_contig_57712	0	0	0.2	0	0	0
18	sra_data3_contig_74545	0	0	0	0	0	0
19	sra_data3_contig_39387	0	0	0	0	0	0
20	sra_data3_contig_45566	0	0	0	0	0	0
21	sra_data3_contig_20377	0	0	0	0	0	0
22	sra_data3_contig_25181	0	0	0	0	0	0
23	sra_data3_contig_51048	0	0	0	0	0	0
24	sra_data3_contig_12831	0	0	0	0	0	0
25	sra_data3_contig_34564	0	0	0	0	0	0
26	sra_data3_contig_23741	0	0	0	0	0	0.04
27	sra_data3_contig_9034	0	0	0	0	0	0
28	sra_data3_contig_40854	0	0	0	0	0	0
29	sra_data3_contig_47133	0	0	0	0	0	0
30	sra_data3_contig_19738	0	0	0	0	0	0
31	sra_data3_contig_30565	0	0	0	0	0	0
32	sra_data3_contig_45151	0	0	0	0	0	0
33	sra_data3_contig_20235	0	0	0	0.2	0	0
34	sra_data3_contig_20573	0	0	0	0	0	0
35	sra_data3_contig_68228	0	0	0	0	0	0
36	sra_data3_contig_13861	0	0	0	0	0	0
37	sra_data3_contig_27547	0	0	0	0	0	0
38	sra_data3_contig_33880	0	0	0.16	0	0	0
39	sra_data3_contig_81260	0	0	0	0	0.16	0
40	sra_data3_contig_22857	0	0	0	0	0	0
41	sra_data3_contig_55814	0	0	0	0	0	0

***Figure 10**: Scoring matrix of the MPprimer program* (Shen et al., 2010) *showing the compatibility of the template sequences. The scores < 0.2 (marked with green circle) show compatible sequence templates whereas the scores ≥ 0.2 (marked with red circle) indicate incompatible template based on the chosen parameters for primer design.*

First, the primers were redesigned based on 30 sequences of microsatellite loci of the established markers selected from three published articles as shown in the following Tables 3, 4 and 5. The redesigning of the primers for the established markers was necessary since they were selected from three different sources and the primer annealing temperatures of the original primers was from 56 °C to 60 °C, which was inadequate for multiplex development.

Table 3*: Characteristics of primer pairs redesigned as part of this thesis for the microsatellite loci described in Nadachowska et al.* (2010)

Locus (GenBank)	Forward & Reverse primer sequence (5’-3’)	Repeat motif	Size range (bp)
GU574493	F: GATTCGCTATCCTGAGGGAGTC R: TTTTTCAGCAAGCCATCCCAAA	$(GA)_3AA(GA)_6AA(GA)_{18}CA(GA)_2CA(GA)_8CA(GA)_2CA(GA)_4$	250-268
GU574494	F: CAGAGCAATTTCTAGGCAAGGT R: GGCGCTATATCAAACTGCAACA	$(TATC)_{10}$	215-247
GU574495	F: CCCTTCAGGCTTGAAGAGAAGA R: ACGTCTTTAGACATGCAGAGGA	$(TATC)_{16}$	229-278
GU574496	F: ACGACAGACATACAGAAAGGCA R: GAGGGAGTGGAAAGAAAAAGCC	$(GATA)_{19}(GACA)_{11}$	214-356
GU574497	F: GCTTTGTGCTGCTACTTACTCC R: ACGGCCTTTACAGATCAGTGAT	$(ATAG)_4G(ATAG)_{24}$	213-274
GU574498	F: AGGCTCAGTTACTTTGACCTGT R: AGACCATTCTTTCTGAGGTATCCT	$(CAGA)_4(TAGA)_{13}$	315-374
GU574499	F: AGGCAGGGTATTTGCGTAGTTA R: GGTCATTTCCACAACAAGCTCA	$(TATC)_{19}$	195-232

Table 4*: Characteristics of primer pairs redesigned as part of this thesis for the microsatellite loci obtained from Sotiropoulos et al.* (2009)

Locus (GenBank)	Forward & Reverse primer sequence (5’-3’)	Repeat motif	Size range (bp)
EU568352	F: CCTCTCATGTGTAATCCTGCCT R: CCCCCAGTAAGAGTGTCACTAC	$(CA)_{12}$	143-185
EU568353	F: TTGGGAGAGCCCTATCTTTGAC R: TGCCAATAGTTTAATCTCTGGCAA	$(CT)_2(GT)_4CG(CT)_2(GT)_7G_4(CT)_2GTTG(GT)_4GGGTC(TG)_3GGTC(TG)_4GGTC(TG)_3T4(TG)_3$	186-205
EU568354	F: ACTTAGTCACAACACAGCCAGA R: GGGGAATAAGGGTGGAGGAAAT	$(CA)_{10}N_{10}(CA)_3AA(CA)_2$	242-250
EU568356	F: AAAAACCCATCGTAAGGTTGGC R: AGTTCATGTGCATATGCTCCAG	$(AT)_2(AC)_6N_4(AC)_4T(CA)_8(TA)_8(CA)_3$	189-205
EU568357	F: CTAGAAATCTCAGCCTGCATGC R: AGATGTGATGTCCTTGGGATCC	$(AC)_5AA(AC)_{10}AA(AC)_4$	151-159
EU568359	F: ATCGTCTACACCTTGGCAGTAG R: CACAACAACACAAGACCTGGTT	$(AC)_2C(AC)_{10}$	309-317

***Table 5**: Characteristics of primer pairs redesigned as part of this thesis for the microsatellite loci described in Drechsler et al* (2013)

Locus (GenBank)	Forward & Reverse primer sequence (5’-3’)	Repeat motif	Size range (bp)
KF442226	F: CAGCTGCAAGCGACGAAG R: GCAGAACTTGTTACCTCCAACC	$(AGTG)_{20}$	90-162
KF442227	F: GGAACCAAGATTCAGCATGGTC R: CTACGGGTAAGTAGTGCGCTAT	$(ATGT)_{12}$	116-188
KF442228	F: GTGCTACCTTTACTTCGGGAGA R: GGAACTGCTTCAATGCCTCTTC	$(ATAG)_{12}$	101-173
KF442229	F: CCAGTCCCAGGCATACTAACAT R: GGTCACCGTGCGCTTTATTAAA	$(AATG)_{10}$	190-206
KF442231	F: TCACTCATTACCAAGTCCTGCT R: ACTTTTGCTGTTCTATGCGACA	$(AGAT)_{19}$	136-151
KF442232	F: TGTTATCAAACCCAAGTCCCCA R: CTGCGGAACTACCTCAAAACTG	$(AACT)_{16}$	198-218
KF442233	F: GTCCATAAGAGCGGTTTCAACC R: CGCATTTAGACCCTCACAGGTA	$(AATC)_{11}$	185-213
KF442234	F: CCATTCACAGAGAGCACTGTTG R: GGGTGATGAGATGCGCTCTATA	$(AGAT)_{10}$	117-134
KF442235	F: CAATGAGTGATGTCATGTGCGA R: GCTTAGGGTAAATAGCCCACCT	$(AATG)_{10}$	166-208
KF442236	F: TAACGGAGCGCGCTATAAAATG R: TCCCCAACTTACAACTCTCCAC	$(ACAT)_{11}$	151-174
KF442238	F: TCAGTGGATGGTCAATCAACCA R: TCAGGAGACAACTCAAGACCTG	$(AATC)_{15}$	164-192
KF442240	F: CCCTGTCAGGTGATTAGCTGTA R: CCACACAACACTACCTCACTCT	$(ATTG)_{15}$	102-152

Seven primer pairs were redesigned (*see Tab. 3*) out of seven sequences of microsatellite loci established by Nadachowska et al. (2010). Another six primer pairs were generated (*see Tab. 4*) out of eight sequences of microsatellite loci determined by Sotiropoulos et al. (2009). In addition, 12 primer pairs were redesigned (*see Tab. 5*) from 15 sequences of microsatellites developed by Drechsler et al. (2013). As mention above, all target sequences were included together in the first set of the primer redesign. Therefore, the sequences that showed conflicts in the scoring matrix were removed one by one. During this process, five target sequences were excluded and from the remaining of the target sequences, 25 primer pairs were redesigned overall in the first set.

In the second set of primer designing were involved 70 target sequences identified by the GMATo microsatellite mining tool. In this occasion, the target sequences showed a greater conflict in the scoring matrix, leading to a larger removal of target sequences during the primer designing process. However, considering the fact that a great number of target

sequences were involved in the start, the remaining of target sequences was satisfactory. This enabled the designing of 42 primer pairs (not shown) in the total of the second set. The primers were Blast-checked against the reference assembly in CLC Genomic Workbench for specificity before ordering them.

3.3.1 Primer validation

The designed primers were subjected to a uniplex validation test to confirm that they generate an amplicon in the expected size range. In this study, PCR reactions were set up using both primer sets (redesigned and/or newly designed primers) to amplify the target microsatellite loci in the genomic DNA samples of smooth newt and great crested newt. Agarose gel electrophoresis was performed to check whether each primer pair could amplify the DNA template under the designed PCR conditions. Successful amplification was confirmed by the presence of bands that were separated by electrophoresis and compared to a 100 bp ladder size standard.

Amplificability testing of the first primer set

The first set of redesigned primer pairs were tested for successful amplification against both newt species involved in this study at two different primer annealing temperatures (Ta) at 58 °C and 52 °C. The template DNA in the first round was of the smooth newt sample and in the second round of the great crested newt sample.

The results of the electrophoresis run after PCR reactions at 58 °C and 52 °C primer annealing temperatures, to analyse the ability of the primer pairs for amplification against smooth newt DNA samples are presented in Table 6 below.

Table 6*: Summary of primer pairs that generated amplicons when tested on smooth newt* (Lissotriton vulgaris) *at two different primer annealing temperatures (Ta) of 58 °C and 52 °C*

Ta of 58 °C	Ta of 52 °C
GU574494*	GU574494*
GU574495	GU574495
GU574496*	GU574496*
GU574497*	GU574497*
GU574498*	GU574498*
KF442227	
KF442234	KF442234
KF442238*	KF442238*
EU568352*	
EU568353	EU568353
EU568357*	

Note: Amplicons of the expected size *

The amplification success rate was rather low, with seven amplicons of the expected size generated out of 25 primer pairs at 58 °C primer annealing temperature. Primer pairs with clear amplicon include GU574494, GU574496, GU574497, GU574498, KF442238, EU568352, and EU568357 (labelled with an * in Table 6). In the electrophoresis runs of the primer pairs GU574495, KF442227, KF442234 and EU568353 non-specific amplification was observed. However, these primer pairs were included in the further testing of marker development. When the Ta was set at 52 °C, more non-specific amplicons were noticed. Those bands were generated out of four primer pairs: KF442236, EU568354, EU568356, and EU568359. Also, the primer pairs EU568352 and EU568357 which gave amplicons of the excepted size at 58 °C Ta, here showed some non-specific amplification. The remainder of the primer pairs which gave a clear band at the Ta of 58 °C did that here as well, except the KF44227 one (*see Appendix, Figure A1*).

The results of the electrophoresis run after PCR reactions at 58 °C and 52 °C Ta respectively, to analyse the ability of the primer pairs for cross-species amplification against great crested newt DNA samples are shown in Table 7 below.

***Table 7**: Summary of primer pairs that generated amplicons when tested on great crested newt* (Triturus cristatus) *at two different primer annealing temperatures (Ta) of 58 °C and 52 °C*

Ta of 58 °C	Ta of 52 °C
GU574495*	GU574495*
GU574497	GU574497
KF442238	KF442238
	EU568352

Note: Amplicons of the expected size *

As anticipated, the success rate of the cross-species amplification at a Ta of 58 °C was very low with only 3 amplicons out of 25 primer pairs. The only primer pairs that generated the amplicon on the predicted size was GU574495 (marked with an * in Table 7), while the bands generated from the primer pairs GU574497 and KF442238 may be regarded as non-specific since the band sizes did not match expected amplicon sizes. Later, when the Ta was set at 52 °C, the only primer pair that showed amplicon on the expected size was GU574495, the primer pair EU568352 gave a rather non-specific amplicon in addition (*see Appendix, Figure A2*). Overall higher level of non-specific amplification was observed at 52 °C primer annealing temperature.

All the primer pairs generating amplicons were selected for further testing in three-primer singleplex. The singleplex in this study was developed according to Blacket et al. (2012).

Possible multiplexes were designed with Multiplex Manager version 1.2 (Holleley & Geerts, 2009) as shown in Figure 11 below.

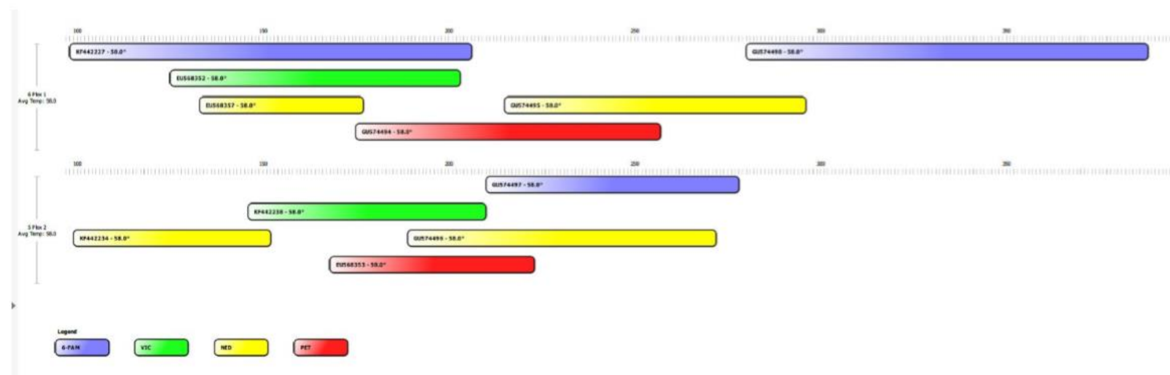


***Figure 11**: Output of Multiplex Manager v-1.2* (Holleley & Geerts, 2009) *showing respective dyes FAM (blue), VIC (green), NED (yellow) or PET (red) for each primer pair in two possible multiplexes (one 6-plex and one 5-plex).*

The output of Multiplex manager *(see Fig. 11)* was used to modify the forward primers by assigning a universal tail to the 5' end as shown in Table 8, before reordering them.

***Table 8**: List of primer pairs for selected microsatellite loci showing the forward tailed (Ft) primers modified with Universal primers (underlined) according to Blacket et al.* (2012) *and respective reverse primers*

Locus + Tail	Forward tailed (Ft) & Reverse primer sequence (5'-3')	Multiplex dye	Size range (bp)
KF442227-A	Ft: GCCTCCCTCGCGCCAGGAACCAAGATTCAGCATGGTC R: CTACGGGTAAGTAGTGCGCTAT	FAM	98-206
GU574498-A	Ft: GCCTCCCTCGCGCCAAGGCTCAGTTACTTTGACCTGT R: AGACCATTCTTTCTGAGGTATCCT	FAM	297-392
EU568352-B	Ft: GCCTTGCCAGCCCGCCCTCTCATGTGTAATCCTGCCT R: CCCCCAGTAAGAGTGTCACTAC	VIC	125-203
EU568357-C	Ft: CAGGACCAGGCTACCGTGCTAGAAATCTCAGCCTGCATGC R: AGATGTGATGTCCTTGGGATCC	NED	133-177
GU574495-C	Ft: CAGGACCAGGCTACCGTGCCCTTCAGGCTTGAAGAGAAGA R: ACGTCTTTAGACATGCAGAGGA	NED	211-296
GU574494-D	Ft: CGGAGAGCCGAGAGGTGCAGAGCAATTTCTAGGCAAGGT R: GGCGCTATATCAAACTGCAACA	PET	197-265
GU574497-A	Ft: GCCTCCCTCGCGCCAGCTTTGTGCTGCTACTTACTCC R: ACGGCCTTTACAGATCAGTGAT	FAM	195-292
KF442238-B	Ft: GCCTTGCCAGCCCGCTCAGTGGATGGTCAATCAACCA R: TCAGGAGACAACTCAAGACCTG	VIC	146-210
KF442234-C	Ft: CAGGACCAGGCTACCGTGCCATTCACAGAGAGCACTGTTG R: GGGTGATGAGATGCGCTCTATA	NED	99-152
GU574496-C	Ft:CAGGACCAGGCTACCGTGACGACAGACATACAGAAAGGCA R: GAGGGAGTGGAAAGAAAAAGCC	NED	196-374
EU568353-D	Ft: CGGAGAGCCGAGAGGTGTTGGGAGAGCCCTATCTTTGAC R: TGCCAATAGTTTAATCTCTGGCAA	PET	168-223

Amplificability testing of the second primer set

The second primer set, containing 42 newly designed primer pairs were tested for amplification on smooth newt and cross-amplification against great crested newt separately. First, all primer pairs were tested on smooth newt DNA template including a non-template control at four different primer annealing temperatures (54 °C, 58 °C, 60 °C, and 62 °C). A summary of the primer pairs that generated amplicons at specific primer annealing temperatures is shown in Table 9 below.

Table 9*: Summary of successful primer pairs in Amplificability testing on smooth newt* (Lissotriton vulgaris) *at four different primer annealing temperatures (Ta) of 54 °C, 58 °C, 60 °C, and 62 °C*

Ta of 54 °C	Ta of 58 °C	Ta of 60 °C	Ta of 62 °C
LvSSR2	LvSSR2	LvSSR2	
		LvSSR6	
LvSSR7	LvSSR7	LvSSR7	
	LvSSR10	LvSSR10	
	LvSSR11	LvSSR11	
	LvSSR12	LvSSR12	
	LvSSR14	LvSSR14	
		LvSSR15	
LvSSR17		LvSSR17	LvSSR17
LvSSR18			
	LvSSR19	LvSSR19	LvSSR19
	LvSSR20	LvSSR20	LvSSR20
	LvSSR21	LvSSR21	
		LvSSR23	
		LvSSR24	
LvSSR25		LvSSR25	
LvSSR27		LvSSR27	
LvSSR29		LvSSR29	
	LvSSR31	LvSSR31	
LvSSR33	LvSSR33	LvSSR33	LvSSR33
	LvSSR34	LvSSR34	LvSSR34
		LvSSR35	
		LvSSR36	
LvSSR37	LvSSR37		
		LvSSR38	
		LvSSR39	
LvSSR40			
LvSSR41		LvSSR41	LvSSR41
LvSSR42		LvSSR42	LvSSR42

The highest amplification success rate on smooth newt sample was obtained at 60 °C primer annealing temperature with 26 amplicons (*see Tab. 9*) out of 42 primer pairs. The amplification rate dropped significantly at 58 °C and 54 °C Ta with 13, 12 amplicons respectively. However, the number of successful primer pairs at 58 °C is expected to be

greater considering the fact that some primer pairs generated bands at 54 °C and 60 °C but not at 58 °C (e.g. LvSSR17, LvSSR25, LvSSR27, LvSSR29, LvSSR41, and LvSSR42). The success rate dropped even more at Ta of 62 °C with only seven amplicons.

In the second occasion, all primer pairs were tested on great crested newt DNA template including a non-template control at two different primer annealing temperatures (58 °C and 60 °C). A summary of the primer pairs that generated amplicons at each respective primer annealing temperatures is shown in Table 10 below.

***Table 10**: Summary of successful primer pairs in Amplificability testing on great crested newt* (Triturus cristatus) *at two different primer annealing temperatures (Ta) of 58 °C and 60 °C*

Ta of 58 °C	Ta of 60 °C
LvSSR2	
LvSSR6	
LvSSR7	
LvSSR10	
LvSSR17	LvSSR17
LvSSR30	
LvSSR31	LvSSR31
LvSSR34	LvSSR34
LvSSR39	

The success rate of primer cross-species amplification testing on great crested newt was quite good, with nine clear amplicons out of 42 primer pairs *(see Tab. 10)* checked at 58 °C primer annealing temperature. However, the success rate was significantly lower at Ta of 60 °C, generating only three bands out of entire second primer set.

Also here the primer pairs that showed clear amplicons were selected for further testing in the process of microsatellite marker development. Multiplex Manager version 1.2 (Holleley & Geerts, 2009) was used as a helping tool to design preliminary multiplexes by assigning a tail sequence according to Blacket et al. (2012) for each forward primer. The modified primers (only forward primers) were ordered anew. After receiving the modified forward tailed primers they were tested in uniplex together with their respective reverse primers. The primer annealing temperature of 58 °C was chosen as more appropriate for further testing, considering that results show greater amplification success rate of the primer testing at this Ta. Likewise, the melting temperature (Tm) of universal primers (*see Tab. 1*) used in this study is around 58 °C. Moreover, higher primer annealing temperatures are known to increase the primers specification. Therefore, a uniplex was performed again for the 27 selected primer pairs at Ta of 58 °C, only on smooth newt DNA samples (*see Fig. 12*). The

testing sample was chosen to be from smooth newt because the second primer set was designed against this newt species.

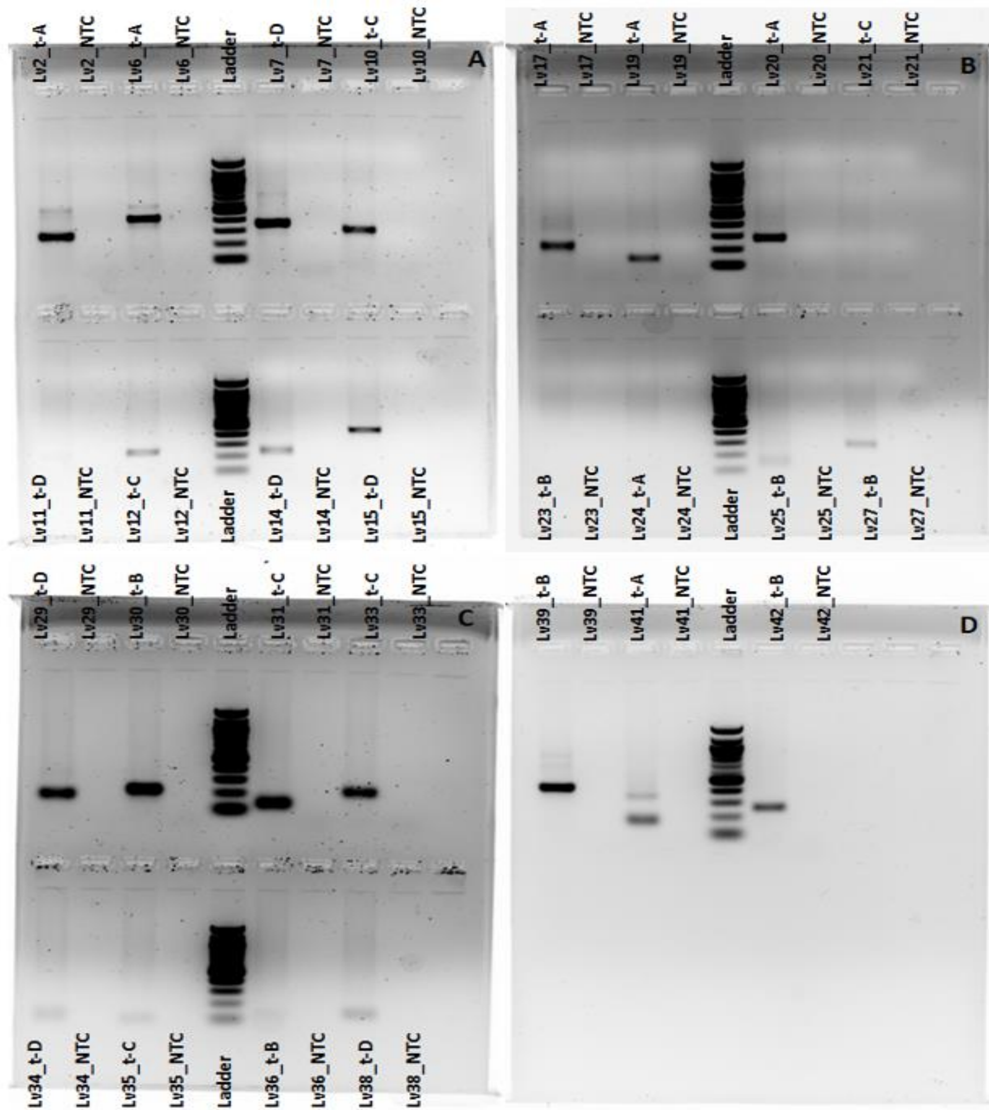


***Figure 12**: Amplificability testing on smooth newt* (Lissotriton vulgaris) *DNA sample (Lv.RK5) at 58°C primer annealing temperature of 27 primer pairs assigned with respective tail sequence (Universal primers) in the 5' end of the forward primer, according to Blacket et al.* (2012).

The pictures (A, B, C, and D) of the gels from Figure 12 showed the successful amplification of PCR product from 23 primer pairs. The success rate of the uniplex testing is very good and the inability of four primer pairs (Lv11_t-D, Lv21_t-C, Lv23_t-B, and Lv24_t-A) to amplify the PCR product is understandable, considering the fact that modification of the forward primers almost doubles their sequence length thus increasing the primer melting temperature. Some bands appear to be weaker but this is most probably because of counter-current mobility of EtBr during electrophoresis relative to the mobility of DNA since the weak intensity is also evident for the 100 bp DNA ladder bands.

3.4 Three-primer singleplex marker testing

Primer pairs that passed successfully the uniplex test were deployed for further testing with singleplex. As mention above a three-primer approach was applied in singleplex, thus the third primer (Universal primer) with a fluorophore attached to its 5’ end was introduced to generate the signal when running in Capillary Electrophoresis (CE).

Initially, 11 markers that showed potential to be utilized in genotyping were tested each with a single specific universal primer (U) to evaluate their ability for allele calling by analysing their electropherograms as presented in the following figures (13 and 14).

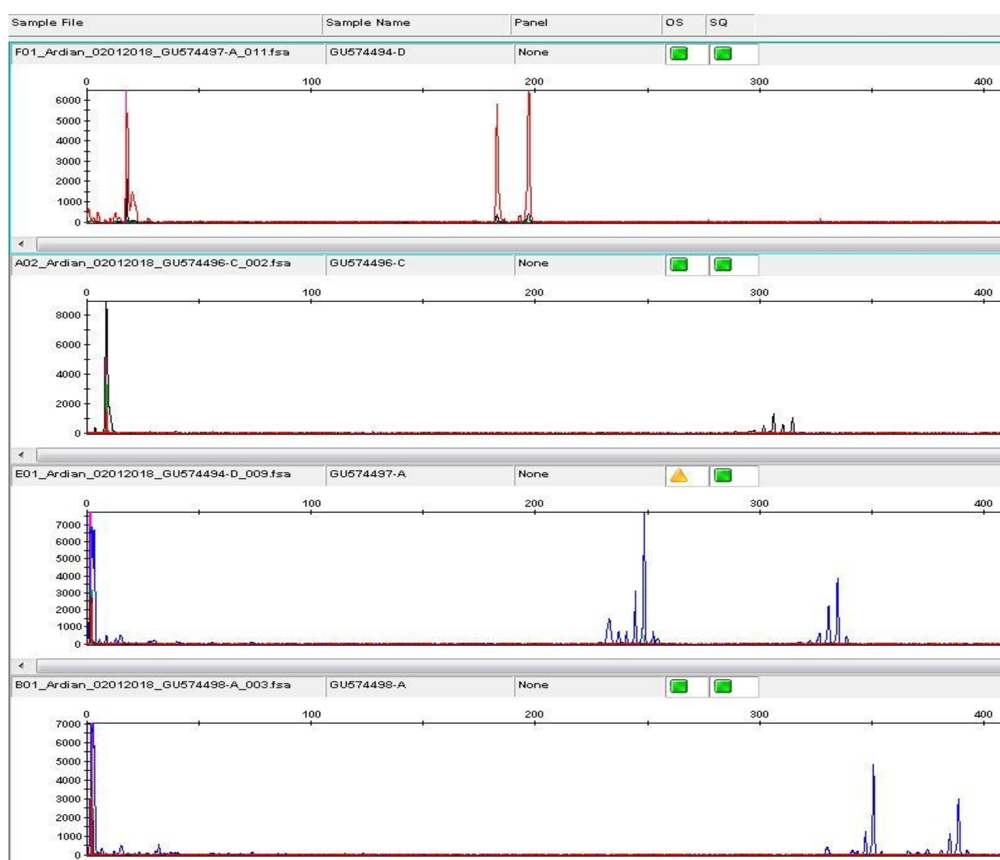


***Figure 13**: Singleplex electropherograms of four successful markers showing clear alleles on smooth newt DNA sample (Lv.RK5). The X-axis shows size (base pair) while the Y-axis shows the intensity in relative fluorescence units (RFU).*

The four markers presented in Figure 13 showed clear heterozygous allele signals. First marker GU574494-D (red) gives two alleles of 183 bp and 197 bp respectively in size. The second marker showing the signal with black colure peaks which represents the yellow dye (NED) emitted from the PCR product of GU574496-C run through CE. As expected, this marker shows alleles of size 306 bp and 315 bp which are in the predetermined interval. Also, the third marker GU574497-A gives clear allele signals on 248 bp and 334 bp length

showing greater allele span with the second allele exceeding the expected size range. The fourth marker (GU574498-A) shows two alleles just inside the predicted interval with 350 bp and 388 bp in size, respectively. Overall these four markers showed great potential for genotype use. However, three of them (GU574494-D, GU574497-A, and GU574498-A) gave signals well above the maximum relative fluorescence units (RFU), which is around 1000 RFU. Considering this, their respective primer concentration should be adjusted during the multiplex optimization. Also, having in mind that GU574497-A and GU574498-A markers gave rise to overlapping alleles should be placed in different multiplexes since they are labelled with the same dye (FAM).

Figure 14*: Singleplex electropherograms showing four markers with no clear alleles, tested on smooth newt DNA sample (Lv.RK5). The X-axis shows size (base pair) while the Y-axis shows the intensity in relative fluorescence units (RFU).*

The electropherograms presented in Figure 14 have been generated from PCR product run in CE of four primer pairs tested in singleplex. In the upper two electropherograms, many artefacts are to be observed making impossible the allele calling from this primer pairs

(KF442227-A and KF442238-B). The bottom two primer pairs (EU568357-C and EU568353-D) showed an unexpected electrophoresis pattern where the first allele peak is shorter than second. Considering the general rule the opposite pattern is expected because amplification of shorter alleles is preferential, the main peak should be located before the smaller peak. Therefore, three primer pairs (KF442227-A, KF442234-C, and GU574496-C), which did not show clear alleles because of too much background noise and artefacts were excluded from further testing.

In the same manner, were tested 23 primer pairs of the second set that passed the uniplex by amplifying the target region of the microsatellite loci (*see Fig. 12*). Some of the primer pairs tested with three-primer singleplex are presented in the electropherograms of Figure 15.

Figure 15*: Singleplex electropherograms of four primer pairs from the second primer set showing clear alleles when tested on smooth newt DNA sample (Lv.RK5). The X-axis shows size (base pair) while the Y-axis shows the intensity in relative fluorescence units (RFU).*

The electropherograms in Figure 15 showed four markers with quite clear alleles. First marker Lv.41-A *(top of the Fig. 14)* gave a strong signal of heterozygous alleles with the first allele size being 169 bp and the second allele on 350 bp showing a great allele span as

anticipated from the uniplex testing in agarose gel electrophoresis (*see Fig. 12*). The second marker Lv.25-B shows stutter peaks which may happen because of the marker dinucleotide repeat motif of 17 repeats (TC x 17). Electrophoresis pattern of the Lv.12-C marker is observed to be in the expected size range with heterozygous alleles of 228 bp and 247 bp in size. Likewise, the marker Lv.14-D gave rise of two alleles inside the expected interval, with 232 bp and 243 bp in size respectively.

Almost all the markers showed allele calling ability even though some background noise was present in some electropherograms most probably because the signal was above the accepted maximum of 1000 RFU limit. Two markers that did not show clear alleles during singleplex testing (Lv.33-C and Lv.17-A) are presented in the fallowing figure.

Figure 16: *The Capillary Electrophoresis pattern of two markers tested with singleplex on smooth newt DNA sample (Lv.RK5). The X-axis shows size (base pair) while the Y-axis shows the intensity in relative fluorescence units (RFU).*

The marker Lv.33-C did not show amplification ability *(see Fig. 16)* even after the second try when tested in singleplex thus was excluded from further testing. Another marker that struggled to give rise of clear alleles was Lv.17-A presented in the below part of Figure 15. This marker showed signals around the predicted interval but many artefacts were present even after reducing the primer concentrations to 0.05 µM Ft, 0.1 µM R and 0.075 µM U primer making quite challenging allele calling.

3.5 Multiplex optimization and fragment analyses

Markers that showed the ability for allele calling during singleplex testing were selected and classified in groups for multiplex optimization. Multiplex Manager was utilized to design the preliminary multiplexes by deploying different dyes for loci with an overlapping size range. The multiplex reactions were performed in the same manner as the singleplex. However, the primer concentrations have been adjusted based on signal strength produced by each marker.

Initially, eight markers from the first set including GU574498-A, GU574497-A, EU568352-B, KF442238-B, EU568357-C, GU574496-C, GU574494-D, and EU568353-D were tried in a multiplex setup at 58°C primer annealing temperature to genotype five smooth newt and two great crested newt individuals *(see Tab. 8)*. The primer concentration was scaled down to 0.05 µM Ft, 0.1 µM R and 0.05 µM U primer for all the markers because the majority of them showed of-scale peaks during singleplex testing.

***Table 8**: Genotyping results of seven newt individuals retrieved from the 8-plex runs in Capillary electrophoresis of the first marker set. Each marker size is shown in base pair (bp) and their corresponding peak height in relative fluorescence units (RFU)*

Multiplex results (JM1)	Smooth Newt (Lv.RK1)		Smooth Newt (Lv.RK2)		Smooth Newt (Lv.RK3)		Smooth Newt (Lv.RK4)		Smooth Newt (Lv.RK5)		Crested Newt (Tc.RK6)		Crested Newt (LH-12)	
	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)
GU574498-A	340; 347	589; 416	354; 358	566; 455	347; 362	568; 383	355; 461	419; 372	351; 388	505; 270				
GU574497-A	248; 326	1437; 519	229; 326	1567; 542	264; 314	1010; 597	243; 331	1185; 550	248; 335	1365; 429	314	1441	314	1150
EU568352-B	138; 151	2539; 2938	138; 151	2497; 997	138; 151	1312; 1947	138; 151	1838; 1905	138; 151	1946; 1621				
KF442238-B	202	782	202	2177	202	212	202	535	202	2230	202	7457	202	5043
EU568357-C	153; 166	344; 1852	153; 166	162; 1072	153; 165	136; 1514	153; 166	136; 1091	153; 165	309; 2036				
GU574496-C	315	671	327	434	315	495	311; 323	320; 253	306; 315	308; 243				
GU574494-D	184; 197	195; 1914	202; 217	159; 1062	184; 197	134; 1386	183; 197	151; 1130	183; 197	748; 2136				
EU568353-D	259; 289	273; 394	259; 285	252; 390	259; 285	208; 340	259; 285	135; 319	259	224				

The multiplexes confirmed previous results retrieved from the singleplex testing. As anticipated four markers including EU568352-B, KF442238-B, EU568357-C, and EU568353-D which did not show clear alleles in singleplex, generated also here a lot of background noise, making more challenging the allele calling. Therefore, the multiplex primer annealing temperature was increased and the panel was tested at Ta of 60 °C and 62 °C. Much clearer electrophoresis pattern was observed at Ta of 60 °C because of greater primer specificity at higher temperature but some background noise was still present. Moreover, the four markers (EU568352-B, KF442238-B, EU568357-C, and EU568353-D) were found to be monomorphic *(see Tab. 8)* in all newt individuals. All these reasons led to exclusion of these four markers from further multiplex optimisation. The other four markers (GU574498-A, GU574497-A, GU574496-C, and GU574494-D) showed better allele calling quality and were observed to be polymorphic. Therefore, they were tested again in a separate

multiplex setup at Ta of 60 °C on five smooth newt individuals as presented in Figure 17 below.

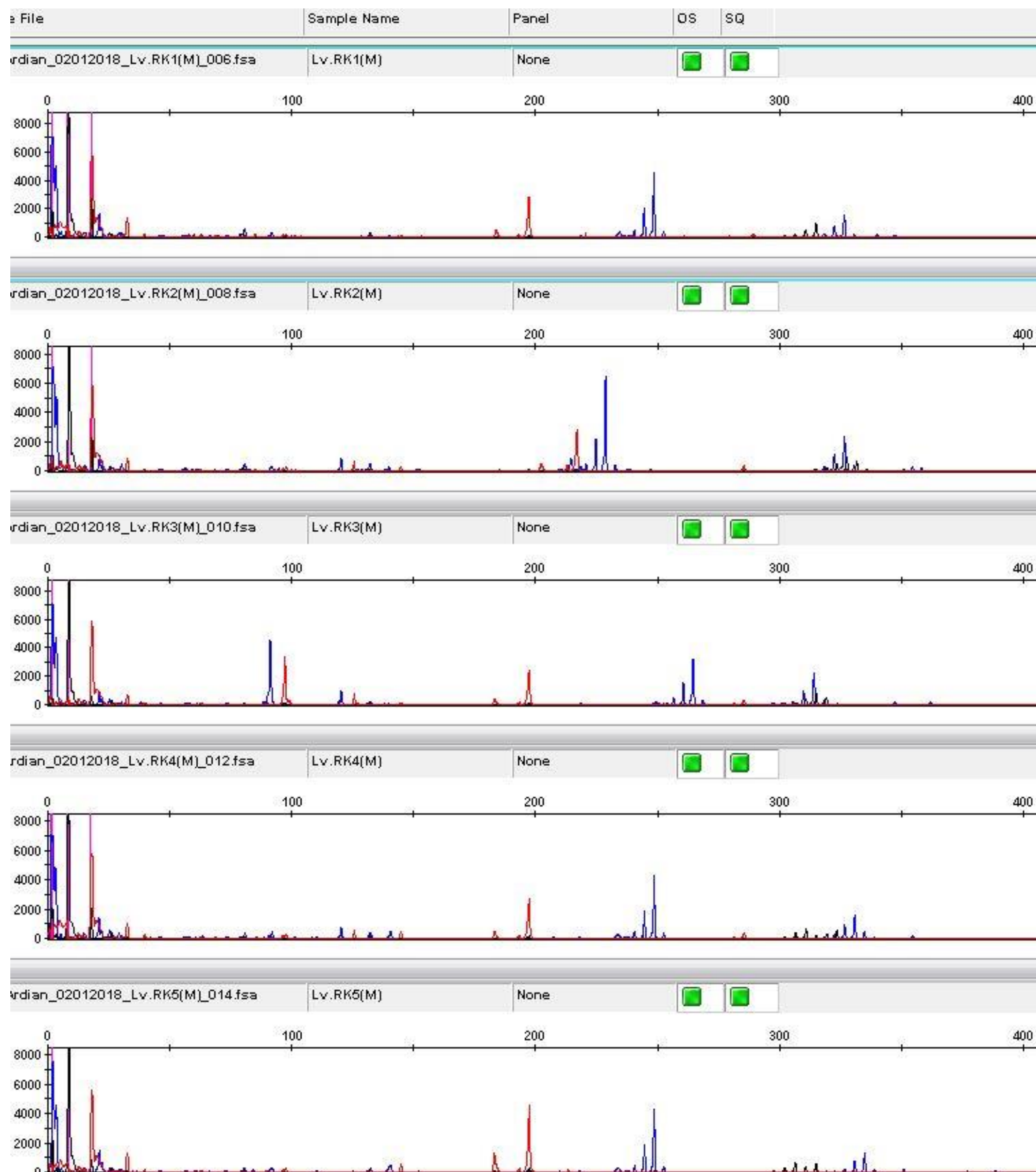


***Figure 17**: The electropherograms of a multiplex panel containing four markers including GU574498-A (blue), GU574497-A (blue), GU574496-C (black), and GU574494-D (red) tested on five smooth newt individuals. The X-axis shows size (base pair) while the Y-axis shows the intensity in relative fluorescence units (RFU).*

As expected, the adapted multiplex setup with only four markers including GU574498-A, GU574497-A, GU574496-C, and GU574494-D confirms their ability for genotyping with clear polymorphic alleles. The marker GU574498-A was observed to be heterozygous in all five smooth newt individuals with overall eight different alleles. Likewise, the marker

GU574497-A gave rise to eight alleles in five heterozygous smooth newt individuals. Moreover, this marker showed homozygous alleles *(see Tab. 8)* on two great crested newt individuals. Another marker showing a great degree of polymorphism was GU574496-C with five different alleles on five smooth newt individuals, of which three were homozygous and two heterozygous. A less polymorphic degree was observed in electrophoresis pattern of GU574494-D marker, where only one smooth newt individual showed alleles in different size from the other four smooth newt individuals. All the smooth newt individuals appeared as heterozygous but this was dubious considering the fact that first allele peak always was observed to be shorter than the second one.

Various multiplex PCR setups were also performed on two preliminary panels containing 20 markers from the second set. The first panel (Newt panel 1) setup was a 12-plex and the second panel (Newt panel 2) was an 8-plex. Initially, both panels were used to genotype six newt individuals (five smooth newts and one great crested newt) at 58 °C primer annealing temperature, which showed the results presented in Table 9.

Table 9*: The genotyping results of six newt individuals retrieved from two multiplex panels (Newt Panel 1- NP 1 and Newt Panel- NP 2), a 12-plex and an 8-plex run in Capillary electrophoresis containing markers of the second set. Each marker size is shown in base pair (bp) and their corresponding peak height in relative fluorescence units (RFU)*

Multiplex results	Smooth Newt (Lv.RK1)		Smooth Newt (Lv.RK2)		Smooth Newt (Lv.RK3)		Smooth Newt (Lv.RK4)		Smooth Newt (Lv.RK5)		Crested Newt (Tc.RK6)	
NP 1	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)	Marker size (bp)	Peak height (RFU)
Lv6_t-A	408	50	408	42	408	81	408	79	408	38	355	130
Lv19_t-A	126; 138	292; 792	130; 134	438; 404	130; 138	425; 357	126; 138	351; 935	125; 138	505; 1345		
Lv20_t-A	268; 270	46; 39	271; 277	36; 63	268;271	57; 37	268; 274	55; 61	268; 274	53; 166		
Lv25_t-B	104; 112	410; 387	106; 122	337; 300	92; 102	242; 219	84; 102	223; 317	108	291	117	1045
Lv27_t-B	272	183	272	221	272	197	272	249	272	229		
Lv39_t-B	415	542	415	721	415	704	415	606	415	559	393; 415	1521; 220
Lv10_t-C	294; 298	233; 255	294; 298	280; 387	294; 298	304; 241	294; 298	87; 322	298	264	295	394
Lv35_t-C	89; 99	984; 2995	89; 99	601; 2381	89; 99	212; 1104	89; 99	767; 2402	89; 99	796; 2401		
Lv7_t-D	345	168	335; 345	104; 224	335; 345	84; 184	335; 346	79; 193	345	224	343; 346	47; 35
Lv14_t-D	231	213	232	207	231	263	231	288	232	241		
Lv15_t-D			426	74	415; 423	52; 73	415; 426	39; 70	415; 428	41; 51		
Lv29_t-D	176; 183	3104; 2113	176; 183	3028; 1212	176; 183	2522; 1138	176; 183	3118; 1392	176; 183	2999; 1283		
NP 2	Smooth Newt (Lv.RK1)		Smooth Newt (Lv.RK2)		Smooth Newt (Lv.RK3)		Smooth Newt (Lv.RK4)		Smooth Newt (Lv.RK5)		Crested Newt (Tc.RK6)	
Lv2_t-A	251	1905	251	2368	248; 251	1014; 1545	248; 251	1064; 1982	250	1976	5 peaks	
Lv41_t-A	169; 350	5260; 423	169; 350	4949; 509	169; 350	3245; 337	169; 350	4203; 325	169; 350	402; 251		
Lv30_t-B	202; 218	460; 4848	202; 218	564; 5864	202; 218	444; 4491	202; 218	315; 3784	218	435	107; 199	1087; 465
Lv36_t-B	130	377	130	373	130	127	130	155	130	169		
Lv42_t-B	257	965	257	848	257	718			257	36		
Lv12_t-C	228; 247	450; 1720	228; 247	797; 2125	228; 247	485; 1384	228; 247	387; 1563	247	1285		
Lv31_t-C	123	4449	123	4556	123	3228	123	3709	123	1566	128	1677
Lv38_t-D	131	6527	131	6616	131	5830	131	5953	131	4142	4 peaks	

The genotyping test of six newt individuals on two panels (a 12-plex and an 8-plex presented in Table 14) confirms the ability of allele calling for the majority of the markers from the second set. Moreover, their electrophoresis pattern was quite clear (*see Fig. 18*) but the signal strength showed great variability, especially between markers with short size

compared to markers with a long size which were labelled with the same colour or even between alleles of heterozygous individuals with great allele span.

***Figure 18**: The electrophoresis pattern of two newt panels from the second marker set, tried on smooth newt individual (Lv.RK5). The first panel (P1, above) was a 12-plex and the second panel (P2, below) was an 8-plex. The X-axis shows size (base pair) while the Y-axis shows the intensity in relative fluorescence units (RFU).*

Therefore, the concentration of the primers was adjusted based on the peak height for each marker, besides the concentration of the universal primers was reduced when markers labelled with the same dye showed high intensity. However, the majority of the markers appeared to be homozygous and monomorphic (*see Tab. 9*). To verify this, the panels were tested for robustness in a greater number of individuals of two different populations (not shown) but still the same pattern was observed. The marker Lv19_t-A showed heterozygous alleles with some degree of polymorphism as did the marker Lv25_t-B. Three markers including Lv6_t-A, Lv39_t-B, and Lv31_t-C were observed to be polymorphic between the two newt species. The markers Lv6_t-A and Lv31_t-C appeared as homozygous in both newt species with alleles 408 bp and 123bp in size on smooth newt individuals whereas the allele size on great crested newt individuals was 355 bp and 128 bp respectively. Likewise, the marker Lv.39_t-B was shown to be homozygous on five smooth newt individuals *(see Tab. 9)* with the allele size of 415 bp whereas on great crested newt individuals appeared as heterozygous with alleles of 393 bp and 415 bp respectively.

3.6 Sequencing Analyses

Direct sequencing was performed for both top and bottom strands of PCR products for seven markers randomly selected from the second primer set (newly isolated) to confirm whether they represented the expected target loci and if they actually harboured tandem sequence repeats. Initially, a BLAST search was conducted to compare nucleotide sequences retrieved from sequencing against the nucleotide collection (nr/nt) database. As expected no significant hits were generated since the primers were designed against the assembly generated from sequenced reads uploaded by authors in Sequence Reads Archive (SRA). Therefore, the respective contig sequences against which the primers were designed for each marker were aligned with two sequences retrieved from sequencing of both top and bottom strands of the PCR products. The alignment program Clustal Omega (1.2.4) was utilized to perform multiple sequence alignments which are to be found in Appendix, while sequence parts harbouring tandem repeats of four markers are presented in Figure 19 below.

```
Lv2-RC          --GGACCTGGGTTTATTATCCGCACTCTCCCCTCTCTT---------TTCGCCGCGCGCG   158
Lv2-F           CTGGAACCTTGCTGGTTATTAGCTACCGTATCTCTCTCTCTCTGTGTGTGGGCGCCCCCC   178
Contig_10921    CTGGAACCTTGCTGGTTATTAGCTACCGTATCTCTCTCTCTCTCTGTTTCAGTGCGCTCC   600

Lv10-RC         TCTCGGCAGGACATGGGATGGACATCTGTCCCCACACCCCCCCACCATCAATGTTCTTCC   233
Lv10-F          TCTCTGCAGGACATACATGGACATCTGTGAGACACACACACACACACTCAATGTTAATGC   234
Contig_20817    TCTCTGCAGGACATACATGGACATCTGTGAGACACACACACACACCATCAATGTTAATGC   539

Lv19-F          AGATGCAGGGAGGGATTACTTTCTCTCAATTCATATTGCAATCAATCAATCCACCGTTGG   115
Contig_26776    AGGGTCAGCAGGAG----ACTTTTCTCAATCAATCAATCAATCAATCAATCCGGGAATTT   116
Lv19-RC         CTATCTCACAGTGTTCTATCTGTTAGGACCCTGTC-CGTGAGGCACCTACCCAGGGATTG   100

Lv39-RC         CACGGCAATAGAGAGAGAGAGAGACATCCCAGGCCACCTCCAACATACAGTGTACACTGCCC   479
Lv39-F          CACGGCAATAGAGAGAGAGAGAGACATCCAGG-CCTCACATCAGTGACATGATCACTTCCCT   309
Contig_9006     CACGGCATTAGAGAGAGAGAGAGACATCCAGGCCCTCACATCAGTGACATGATCACTTCCCT   315
```

***Figure 19**: Multiple sequence alignments (only parts harbouring tandem repeats) of the sequences retrieved from direct sequencing of both top (F) and bottom (RC) strands of PCR products of four markers (Lv2, Lv10, Lv19, and Lv39) against the contig sequence from which respective primers were designed.*

As anticipated, the sequencing results *(see Fig. 19)* showed that the marker sequences matched the sequences of the targeted microsatellite loci. Generally, greater identity was observed between the contig sequence and the sequence of forward sequencing (e.g. markers Lv2, Lv10, and Lv19) because of better sequencing quality with forward primers compared to reverse primers. However, the marker Lv39 was successfully verified from both sequences of forward and reverse sequencing matching with great identity, especially on the

tandem repeat sequence region. The marker sequences obtained from direct sequencing were aligned also against the unassembled reads (*see Appendix, Figure A3-A6*) by performing a blast search (BLASTN 2.8.0+) against SRA database (only against the eight experiments containing reads data utilized in this study), based on a greedy algorithm for aligning DNA sequences developed by Zhang et al. (2000). The mapping of reads against the marker sequence showed great cover throughout the sequence with overlapping sequence reads. The only exception was observed at the ends of contigs which was expected since it is known that the quality of the sequencing is weaker in that part of the sequence. The results support the need of performing *de novo* assembly and in same time confirm the quality of the assembly.

4. Discussion

4.1 *De novo* assembly

The implementation of NGS approaches enormously improved the success rate of the development of new microsatellite loci for many non-model species. However, there is a lack of available sequence data especially for the species with large genome and high proportion of repetitive DNA regions as this one studied here (Salamanders). In the present study, a *de novo* assembly was constructed by using a limited data set with CLC assembler. The N50 measure gives a very good indication of how well the assembly has succeeded, which is dependent on the scaffold (or contig) length as well as its numbers (Mäkinen, Salmela, & Ylinen, 2012). The greater N50 value usually refers to the better assembly, while the erroneously joined reads may also produce the high value of N50 (Axelson-Fisk, 2015). A quality assessment in our assembly with QUAST web-based program showed a very low N50 value (376), which value would be considered as inacceptable for the vast majority of bioinformatics analyses in terms of genome assembly evaluation. However, having in mind that our assembly was generated only from sequence reads generated from different spots of 3' UTRs this N50 value is reasonable. Nevertheless, considering the fact that our goal of performing *de novo* assembly was mainly to increase the contig length of the available sequence data, we can evaluate it as quality assembly. This assembly showed to be very useful by generating sufficient contigs with appropriate length in a user-friendly file for use in multiplex primer design for development of microsatellite markers.

4.2 Development of microsatellite markers

4.2.1 Microsatellite development based on established markers

One of the objectives of this study included the redesigning of microsatellite primers from already established markers. The sequences of 30 microsatellite loci selected from three published articles, served as voluntaries for the redesign of 25 primer pairs with MPprimer program. The first article contributed with seven microsatellites isolated by Nadachowska et al. (2010) for the carpathian newt *(Lissotriton montandoni)*. These SSR markers were evaluated successfully for cross-species amplification on smooth newt. From the second article were successfully utilized six microsatellite loci of the greek smooth newt *(Lissotriton vulgaris graecus)* determined by Sotiropoulos et al. (2009). Another 12 markers developed

by Drechsler et al. (2013) for the palmate newt *(Lissotriton helveticus)* were added to complete the first primer set. Primers were tested on the tissue samples of both newt species at two different primer annealing temperatures (52 °C and 58 °C). When tested on smooth newt samples 11 primer pairs generated amplicons whereas the testing on great crested newt samples showed only 3 primer pairs to work. As anticipated, primer testing at Ta of 58 °C showed to be the appropriate primer annealing temperature for the uniplex PCR reactions, whereas the primers tried at Ta of 52 °C were observed to give rise of unspecific amplicons. Onward, the primer pairs with confirmed amplification were tested on singleplex PCR by employing a three-primer approach according to Blacket et al. (2012) utilizing four fluorescently labeled universal primers to give rise of detectably amplicons when run on Capillary Electrophoresis. The number of successful primers dropped to eight after the singleplex testing. However, the singleplex examination showed only four markers with clear ability for allele calling whereas from other four markers were observed dubious alleles. Nevertheless, all eight markers were tried in various multiplex setups which in the end confirmed that only four markers including GU574498-A, GU574497-A, GU574496-C, and GU574494-D *(see Fig. 16)* show clear allele calling ability with a good polymorphism degree. It is important to mention that the four successful markers are established from the sequences of microsatellite loci obtained from Nadachowska et al. (2010). Nadachowska and co-authors developed microsatellites for genetic study on the *L. montandoni* species, which they successfully tested for amplification on *L. vulgaris* species. This indicates that the other closely related species, the *L. vulgaris graecus* and the *L. helveticus* from the studies of (Sotiropoulos et al., 2009) and (Drechsler et al., 2013) respectively, could have larger genetic distance from the *L. vulgaris* species studied here. This kind of significant and negative relationship between microsatellite amplification performance and evolutionary distance between the original species and the tested species was reported previously by Primmer et al. (1996). The above assertation gains support from the results of this study considering the fact that none of the four markers established in the present study showed to work on cross-species amplification testing against the phylogenetically more distant *T. cristatus* species. A similar finding was reported by Krupa et al. (2002) but in the opposite direction. The authors demonstrated that most of the determined loci for *T. cristatus* are useful for the cross-species study of all large-bodied species of the genus, but showed very low success rate when tried in species with greater genetic divergence such as *L. alpestris* and *L. vulgaris*. The inefficiency of microsatellite markers developed from loci of closely related species, the *L. helveticus* and the *L. vulgaris graecus* could be partly affected by the environmental factors,

considering the geographical distance between Norway and France or Balkan Peninsula. The high substitution rate of microsatellite markers provokes primer binding sites or their flanking sequences to have a few mutations for the species that diverged a long time ago (Zardoya et al., 1996), even geographically. A variable success rate of microsatellite cross-species amplification was previously reported by Nadachowska et al. (2010), who suggested that the utility of their established markers may be limited to geographically restricted groups of the smooth newt. Their prediction could be confirmed by the data of the present study with only four successful markers on smooth newt populations in the southern region of Norway out of seven microsatellite loci established for population study of the same newt species in Poland. Since, the minimum required number of markers for the genotyping purpose was not reached, it was necessary to explore other possible strategies for additional markers.

4.2.2 Microsatellite isolation from genomic sequences

The second strategy utilized for identification of microsatellite loci was based on available sequence data. In general, a lack of sequence data of the newt species has been encountered. Large genome size and repetitive DNA content are widely cited as the main challenges for genome assembly. These two reasons seem to be the main drawback also for genome sequencing of newt species. In addition, the impact of these two factors could dependent on the genomic distribution of repetitive DNA and single-copy sequences (Keinath et al., 2015). Recently, the significant advance in sequencing technology and the determination to go through such challenging process enabled the sequencing of a few salamander species genome mainly on model organisms with unique regenerative abilities. Keinath et al. (2015) have developed chromosome-targeted sequencing approach which they implemented successfully by capturing and sequencing two smallest *Ambystoma mexicanum* chromosomes. Based on the assembled sequence data the authors estimated the axolotl genome to be ten times greater than the human genome. In another study conducted by researchers at Karolinska Institutet have managed to sequence the entire genome of the Iberian ribbed newt *(Pleurodeles waltl)*, which is a full six times greater than the human genome (Elewa et al., 2017). For the newt species studied in the present study there are only a few experiments with sequence reads available up to date. Therefore, this sequence data were utilized for new microsatellite loci identification although they posed a high risk of failure since they were generated by targeting 3’ untranslated transcript regions (3’ UTR). However, Tang et al. (2008) previously have reported identifying a large number of

polymorphic SSRs using publicly available EST sequences of potato, tomato, rice, *Arabidopsis*, *Brassica*, and chicken even at 3' UTR sites.

Initially, the sequence reads were assembled by performing *de novo* assembly, which resulted in more appropriate contig length for multiplex primer design purpose. The use of these data was very promising in terms of successful identification of sequences harboring tandem repeats and primer designing. Moreover, a satisfactory number of markers showed to have strong ability of clear allele calling for genotype use. However, when tested for robustness in greater number of newt individual's, happened what we feared most, only two markers showed some degree of polymorphism.

Considering the fact that the sequences harboring tandem repeats were selected with a mining tool the length of tandem repeats for majority of the markers was ≤15 base pair. This was due to the drawback of mining tools which are heavily biased to mine exact tandem repeats or perfect repeats (Sharma et al., 2007). In generally, it is a consensus that microsatellites containing a larger number of repeats are more polymorphic, one may suppose that the observed monomorphic was partly affected by the short number of repeats. Temnykh et al.(2001) determined the frequency and distribution of different simple sequence repeats in the rice genome by categorized microsatellite loci into two groups based on the length of the repeat motif. Class I, or hyper-variable markers, consisted of SSRs ≥20 bp, and Class II, or potentially variable markers, consisted of SSRs ≥12 bp <20 bp. However, polymorphism has been observed in microsatellites with as few as five repeats (Karsi et al., 2002). Similarly, Tang et al. (2008) were surprised, when they observed higher frequency of polymorphic in the short SSRs.

The expansions of SSR stationed in the 3′-UTRs can induce transcription slippage which leads to expanded mRNA product, moreover this can be accumulated as nuclear foci, which can disrupt splicing and possibly other cellular function (Y. C. Li et al., 2004). In addition, microsatellites occurring in the 3′-UTR could affect gene expression through their influence on the stability of transcribed products. The GA-rich repetitive DNA segment in the 3′-UTR of the chicken elastin gene was found to have such a role (Hew, Lau, Grzelczak, & Keeley, 2000). Recently, more and more pivotal roles of SSR markers situated in the 3' UTR are revealed, leading us to conclude that such markers are not appropriate for genotyping purpose, since they are highly likely to be under selection.

4.3 Sequencing Analysis

The sequencing results verified the microsatellite markers by confirming that the sequences were harboring the expected motif of the tandem repeats. Moreover, the alignment of the marker sequence against the contigs of the assembly from which respective primers were designed revealed an unexpected degree of polymorphism.

Considering the fact that the original sequence reads were generated from samples of newt population of a great geographical distance from the ones studied in the present study. We suggested that these markers may show a small degree of polymorphism only between newt populations with great geographical distance restricted. Another assumption that can be driven from this is the possibility of a very low number of alleles for these markers.

The alignments of the markers sequence with the contigs against which the primers were designed as a wet-lab-based assessment of the quality of the de novo assembly can give the needed ground support. However, the quality of the sequencing could have been improved and a greater number of the sequenced markers would have increased the support of this assumption. Nevertheless, the blast search of the markers sequence against the unassembled reads confirms the need for assembled reads since the designing of primers from the original unassembled reads would generate markers of the same size which are unsuitable for multiplex panel development.

5. Conclusion

The main goal of this study was the development of microsatellite panels for use in both species identification and determination of genetic diversity in populations of smooth newt (Lissotriton vulgaris) and great crested newt (Triturus cristatus) in the southern region of Norway. Two different strategies were utilized to generate sequences of potential microsatellite loci. Initially, 30 established markers of newt species studied here or their closely related newt species were selected from different published articles. The first primer set was comprised of 25 primer pairs redesigned from prior selected microsatellite loci. However, a very low success rate of established markers was achieved with only four markers that showed a clear ability for allele calling after the optimization on different multiplex conditions. The results suggest that a greater number of evaluated loci would increase the chance of developing multiplex panels. Therefore, 70 sequences were identified with GMATo mining tool from publically available genomic sequences of the smooth newt species to isolate and add new markers. A second primer set containing 42 primer pairs newly designed from identified sequences was set up and tested. Nevertheless, these markers showed an extremely low success rate with only two polymorphic markers while 18 other markers were observed to be monomorphic. Since these markers were established from sequences of the 3’ UTRs it is highly likely to be under selection. Finally, two or three panels had to be established, collectively harboring ca. 12-16 common microsatellite markers and 2-4 species-specific markers. Unfortunately, the number of successfully established markers in this study was low with only six markers showing potential for genotype use. Therefore, we were unable to conduct a genotyping experiment and determine the genetic diversity of newt populations.

6. References

Abdul-Muneer, P. (2014). Application of microsatellite markers in conservation genetics and fisheries management: recent advances in population structure analysis and conservation strategies. *Genetics research international, 2014*.

Allentoft, M. E., Schuster, S., Holdaway, R., Hale, M., McLay, E., Oskam, C. L., . . . Bunce, M. (2009). Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *BioTechniques, 46*(3), 195-200.

Arif, I. A., Khan, H. A., Bahkali, A. H., Al Homaidan, A. A., Al Farhan, A. H., Al Sadoon, M., & Shobrak, M. (2011). DNA marker technology for wildlife conservation. *Saudi Journal of Biological Sciences, 18*(3), 219-225.

Arntzen, J., Smithson, A., & Oldham, R. (1999). Marking and tissue sampling effects on body condition and survival in the newt Triturus cristatus. *Journal of Herpetology*, 567-576.

Arntzen, J., & Teunis, S. (1993). A six year study on the population dynamics of the crested newt (Triturus cristatus) following the colonization of a newly created pond. *Herpetological Journal, 3*(3), 99-110.

Axelson-Fisk, M. (2015). Implementation of a Comparative Gene Finder *Comparative Gene Finding* (pp. 311-324): Springer.

Babik, W., Pabijan, M., Arntzen, J., Cogalniceanu, D., Durka, W., & Radwan, J. (2009). Long-term survival of a urodele amphibian despite depleted major histocompatibility complex variation. *Molecular Ecology, 18*(5), 769-781.

Babik, W., Szymura, J., & Rafiński, J. (2003). Nuclear markers, mitochondrial DNA and male secondary sexual traits variation in a newt hybrid zone (Triturus vulgaris× T. montandoni). *Molecular Ecology, 12*(7), 1913-1930.

Banhos, A., Hrbek, T., Gravena, W., Sanaiotti, T., & Farias, I. P. (2008). Genomic resources for the conservation and management of the harpy eagle (Harpia harpyja, Falconiformes, Accipitridae). *Genetics and Molecular Biology, 31*(1), 146-154.

Beebee, T. (2005). Conservation genetics of amphibians. *Heredity, 95*(6), 423.

Beebee, T., & Griffiths, R. A. (2000). Amphibians and reptiles: A natural history of the British herpetofauna. The New Naturalist series: London: HarperCollins.

Beebee, T. J., & Griffiths, R. A. (2005). The amphibian decline crisis: a watershed for conservation biology? *Biological Conservation, 125*(3), 271-285.

Bell, G. (1977). The life of the smooth newt (Triturus vulgaris) after metamorphosis. *Ecological Monographs, 47*(3), 279-299.

Bell, G., & Lawton, J. (1975). The ecology of the eggs and larvae of the smooth newt (Triturus vulgaris (Linn.)). *The Journal of Animal Ecology*, 393-423.

Benavides, F. J., & Guénet, J.-L. (2012). Chapter 1.4 - Mouse Genomics A2 - Hedrich, Hans J *The Laboratory Mouse (Second Edition)* (pp. 57-90). Boston: Academic Press.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research, 27*(2), 573.

Bessetti, J. (2007). An introduction to PCR inhibitors. *J. Microbiol. Meth, 28*, 159-167.

Blacket, M. J., Robin, C., Good, R. T., Lee, S. F., & Miller, A. D. (2012). Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. *Molecular Ecology Resources, 12*(3), 456-463.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology, 13*(11), 3261-3273.

Compeau, P. E., Pevzner, P. A., & Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nat Biotechnol, 29*(11), 987-991.

Cooke, A. S., & Frazer, J. F. D. (1976). Characteristics of newt breeding sites. *Journal of Zoology, 178*(2), 223-236.

Crow, J. F. (1993). How much do we know about spontaneous human mutation rates? *Environ Mol Mutagen, 21*(2), 122-129.

Da Maia, L. C., Palmieri, D. A., De Souza, V. Q., Kopp, M. M., de Carvalho, F. I. F., & Costa de Oliveira, A. (2008). SSR Locator: Tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *International Journal of Plant Genomics, 2008*.

Deban, S. M., & Wake, D. B. (2000). Aquatic feeding in salamanders. *Feeding: form, function and evolution in tetrapod vertebrates*, 65-94.

Denoël, M. (2012). Newt decline in Western Europe: highlights from relative distribution changes within guilds. *Biodiversity and Conservation, 21*(11), 2887-2898.

Denoël, M., Perez, A., Cornet, Y., & Ficetola, G. F. (2013). Similar local and landscape processes affect both a common and a rare newt species. *PLoS One, 8*(5), e62727.

Dervo, B. K., Pedersen, C., & Bærum, K. M. (2016). Tap av ynglelokaliteter for storsalamander i Norge.

Di Rienzo, A., Peterson, A., Garza, J., Valdes, A., Slatkin, M., & Freimer, N. (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences, 91*(8), 3166-3170.

Dib, C., Fauré, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., . . . Seboun, E. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature, 380*(6570), 152.

Dietrich, W. F., Miller, J. C., Steen, R. G., Merchant, M., Damron, D., Nahf, R., . . . Dredge, R. D. (1994). A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genetics, 7*(2S), 220.

Dolmen, D. (1983a). Diel rhythms and microhabitat preference of the newts Triturus vulgaris and T. cristatus at the northern border of their distribution area. *Journal of Herpetology*, 23-31.

Dolmen, D. (1983b). Growth and size of Triturus vulgaris and T. cristatus (Amphibia) in different parts of Norway. *Ecography, 6*(4), 356-371.

Dolmen, D., & Koksvik, J. I. (1983). Food and feeding habits of Triturus vulgaris (L.) and T. cristatus (Laurenti)(Amphibia) in two bog tarns in central Norway. *Amphibia-reptilia, 4*(1), 17-24.

Dolmen, D., Skei, J. K., & Blakar, I. (2008). Scandinavian amphibians: their aquatic habitat and tolerance. *Fauna norvegica, 26*.

Drechsler, A., Geller, D., Freund, K., Schmeller, D. S., Künzel, S., Rupp, O., . . . Steinfartz, S. (2013). What remains from a 454 run: estimation of success rates of microsatellite loci development in selected newt species (Calotriton asper, Lissotriton helveticus, and Triturus cristatus) and comparison with Illumina-based approaches. *Ecology and Evolution, 3*(11), 3947-3957.

Edgar, P., & Bird, D. R. (2006). Action plan for the conservation of the crested newt Triturus cristatus species complex in Europe. *Council of the European Union, Strassbourg, Germany*, 1-33.

Edwards, A., Civitello, A., Hammond, H. A., & Caskey, C. T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American journal of human genetics, 49*(4), 746.

Elewa, A., Wang, H., Talavera-López, C., Joven, A., Brito, G., Kumar, A., . . . Simon, A. (2017). Reading and editing the Pleurodeles waltl genome reveals novel features of tetrapod regeneration. *Nature Communications, 8*(1), 2286.

Ellegren, H. (2000a). Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics, 24*, 400.

Ellegren, H. (2000b). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet, 16*(12), 551-558.

Ellegren, H., Moore, S., Robinson, N., Byrne, K., Ward, W., & Sheldon, B. C. (1997). Microsatellite evolution--a reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Mol Biol Evol, 14*(8), 854-860.

Elnifro, E. M., Ashshi, A. M., Cooper, R. J., & Klapper, P. E. (2000). Multiplex PCR: optimization and application in diagnostic virology. *Clinical microbiology reviews, 13*(4), 559-570.

Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol, 11*(9), 1591-1604.

Fan, H., & Chu, J.-Y. (2007). A Brief Review of Short Tandem Repeat Mutation. *Genomics, Proteomics & Bioinformatics, 5*(1), 7-14.

Ferguson, M. M., & Danzmann, R. G. (1998). Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting? *Canadian Journal of Fisheries and Aquatic Sciences, 55*(7), 1553-1563.

Freeman, B., Smith, N., Curtis, C., Huckett, L., Mill, J., & Craig, I. (2003). DNA from buccal swabs recruited by mail: evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. *Behavior genetics, 33*(1), 67-72.

Gao, C., Ren, X., Mason, A. S., Li, J., Wang, W., Xiao, M., & Fu, D. (2013). Revisiting an important component of plant genomes: microsatellites. *Functional Plant Biology, 40*(7), 645-661.

Griffiths, R., & Teunis, B. (1996). *Newts and salamanders of Europe*: T & AD Poyser London.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., . . . Salin, F. (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources, 11*(4), 591-611.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics, 29*(8), 1072-1075.

Gustafson, D. (2011). *Choosing the best of both worlds* (Vol. 2011).

Halliday, T. R. (1990). The Evolution of Courtship Behavior in Newts and Salamanders. In P. J. B. Slater, J. S. Rosenblatt, & C. Beer (Eds.), *Advances in the Study of Behavior* (Vol. 19, pp. 137-169): Academic Press.

Hew, Y., Lau, C., Grzelczak, Z., & Keeley, F. W. (2000). Identification of a GA-rich sequence as a protein-binding site in the 3′-untranslated region of chicken elastin mRNA with a potential role in the developmental regulation of elastin mRNA stability. *Journal of Biological Chemistry, 275*(32), 24857-24864.

Holleley, C. E., & Geerts, P. G. (2009). Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques, 46*(7), 511-517.

International Human Genome Sequencing, C. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*, 860.

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985a). Hypervariable 'minisatellite' regions in human DNA. *Nature, 314*(6006), 67-73.

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985b). Individual-specific 'fingerprints' of human DNA. *Nature, 316*(6023), 76-79.

Jehle, R., & Arntzen, J. (2000). Post-breeding migrations of newts (Triturus cristatus and T. marmoratus) with contrasting ecological requirements. *Journal of Zoology, 251*(3), 297-306.

Jehle, R., & Arntzen, J. (2002). Microsatellite markers in amphibian conservation genetics. *Herpetological Journal, 12*, 1-9.

Jones, C., Edwards, K., Castaglione, S., Winfield, M., Sala, F., Van de Wiel, C., . . . Daly, A. (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular breeding, 3*(5), 381-390.

Karsi, A., Cao, D., Li, P., Patterson, A., Kocabas, A., Feng, J., . . . Liu, Z. (2002). Transcriptome analysis of channel catfish (Ictalurus punctatus): initial analysis of gene expression and microsatellite-containing cDNAs in the skin. *Gene, 285*(1), 157-168.

Keinath, M. C., Timoshevskiy, V. A., Timoshevskaya, N. Y., Tsonis, P. A., Voss, S. R., & Smith, J. J. (2015). Initial characterization of the large genome of the salamander Ambystoma mexicanum using shotgun and laser capture chromosome sequencing. *Scientific Reports, 5*, 16413.

Kinne, O. (2004). Successful re-introduction of the newts Triturus cristatus and T. vulgaris. *Endangered Species Research, 1*, 25-40.

Koreth, J., O'Leary, J. J., & J, O. D. M. (1996). Microsatellites and PCR genomic analysis. *J Pathol, 178*(3), 239-248.

Kraemer, B., Gäbler-Schwarz, H., & Leese, M. (2009). Pöhlmann and Frickenhaus (2009). STAMP: Extensions to the STADEN sequence analysis package for high throughput interactive microsatellite marker design. *BMC Bioinformatics, 10*, 41.

Kruglyak, S., Durrett, R. T., Schug, M. D., & Aquadro, C. F. (1998). Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences, 95*(18), 10774-10778.

Krupa, A., Jehle, R., Dawson, D., Gentle, L., Gibbs, M., Arntzen, J., & Burke, T. (2002). Microsatellite loci in the crested newt (Triturus cristatus) and their utility in other newt taxa. *Conservation genetics, 3*(1), 85-87.

Lawson, M. J., & Zhang, L. (2008). Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5'-UTR region. *Gene, 407*(1-2), 54-62.

Levinson, G., & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol, 4*(3), 203-221.

Li, Y., Wongprasert, K., Shekhar, M., Ryan, J., Dierens, L., Meadows, J., . . . Lyons, R. E. (2007). Development of two microsatellite multiplex systems for black tiger shrimp Penaeus monodon and its application in genetic diversity study for two populations. *Aquaculture, 266*(1-4), 279-288.

Li, Y. C., Korol, A. B., Fahima, T., & Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol, 21*(6), 991-1007.

Liu, Z. J., & Cordes, J. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture, 238*(1-4), 1-37.

Mäkinen, V., Salmela, L., & Ylinen, J. (2012). Normalized N50 assembly metric using gap-restricted co-linear chaining. *BMC Bioinformatics, 13*(1), 255.

Maletzky, A., Kyek, M., & Goldschmid, A. (2007). *Monitoring status, habitat features and amphibian species richness of crested newt (Triturus cristatus superspecies) ponds at the edge of the species range (Salzburg, Austria).* Paper presented at the Annales de Limnologie-International Journal of Limnology.

Malmgren, J. C. (2001). *Evolutionary ecology of newts.* Örebro universitetsbibliotek.

Malmgren, J. C. (2002). How does a newt find its way from a pond?: Migration patterns after breeding and metamorphosis in great crested newts (Triturus cristatus) and smooth newts (T-vulgaris). *Herpetological Journal, 12*(1), 29-35.

Markoulatos, P., Siafakas, N., & Moncany, M. (2002). Multiplex polymerase chain reaction: a practical approach. *Journal of clinical laboratory analysis, 16*(1), 47-51.

Martel, A., Blooi, M., Adriaensen, C., Van Rooij, P., Beukema, W., Fisher, M. C., . . . Pasmans, F. (2014). Recent introduction of a chytrid fungus endangers Western Palearctic salamanders. *Science, 346*(6209), 630-631.

Mason, A. S. (2015). SSR genotyping. *Methods Mol Biol, 1245*, 77-89.

Missiaggia, A., & Grattapaglia, D. (2006). Plant microsatellite genotyping with 4-color fluorescent detection using multiple-tailed primers. *Genet Mol Res, 5*(1), 72-78.

Mohindra, V., Mishra, A., Palanichamy, M., & Ponniah, A. (2001). Cross-species amplification of Catla catla microsatellite locus in Labeo rohita.

Moore, S., Sargeant, L., King, T., Mattick, J., Georges, M., & Hetzel, D. (1991). The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics, 10*(3), 654-660.

Müllner, A. (2001). Spatial patterns of migrating great crested newts and smooth newts: the importance of the terrestrial habitat surrounding the breeding pond. *Rana, Sonderheft, 4*, 279-293.

Nadachowska, K., Flis, I., & Babik, W. (2010). Characterization of microsatellite loci in the Carpathian newt (Lissotriton montandoni). *The Herpetological Journal, 20*(2), 107-110.

Nadir, E., Margalit, H., Gallily, T., & Ben-Sasson, S. A. (1996). Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. *Proceedings of the National Academy of Sciences, 93*(13), 6470-6475.

Nevo, E. (2001). Evolution of genome-phenome diversity under environmental stress. *Proc Natl Acad Sci U S A, 98*(11), 6233-6240.

O'reilly, P., & Wright, J. M. (1995). The evolving technology of DNA fingerprinting and its application to fisheries and aquaculture. *Journal of Fish Biology, 47*(sA), 29-55.

Oetting, W. S., Lee, H., Flanders, D., Wiesner, G., Sellers, T., & King, R. (1995). Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics, 30*(3), 450-458.

Ohta, T., & Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetics Research, 22*(2), 201-204.

Pabijan, M., Zieliński, P., Dudek, K., Chloupek, M., Sotiropoulos, K., Liana, M., & Babik, W. (2015). The dissection of a Pleistocene refugium: phylogeography of the smooth newt, Lissotriton vulgaris, in the Balkans. *Journal of Biogeography, 42*(4), 671-683.

Park, L. K., & Moran, P. (1994). Developments in molecular genetic techniques in fisheries. *Reviews in fish biology and fisheries, 4*(3), 272-299.

Platt, A. R., Woodhall, R. W., & George, A. L. (2007). Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *BioTechniques, 43*(1), 58.

Primmer, C., Møller, A., & Ellegren, H. (1996). A wide-range survey of cross-species microsatellite amplification in birds. *Molecular Ecology, 5*(3), 365-378.

Rachlin, J., Ding, C., Cantor, C., & Kasif, S. (2005). MuPlex: multi-objective multiplex PCR assay design. *Nucleic Acids Research, 33*(suppl_2), W544-W547.

Rico, C., Rico, I., & Hewitt, G. (1996). 470 million years of conservation of microsatellite loci among fish species. *Proc. R. Soc. Lond. B, 263*(1370), 549-557.

Rogers, S. M., & Bernatchez, L. (2005). Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (Coregonus clupeaformis). *Mol Ecol, 14*(2), 351-361.

Russell, J., Fuller, J., Macaulay, M., Hatz, B., Jahoor, A., Powell, W., & Waugh, R. (1997). Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theoretical and Applied Genetics, 95*(4), 714-722.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma, 109*(6), 365-371.

Schlötterer, C., & Pemberton, J. (1998). The use of microsatellites for genetic analysis of natural populations—a critical review *Molecular approaches to ecology and evolution* (pp. 71-86): Springer.

Schlötterer, C., & Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Research, 20*(2), 211-215.

Schlotteröer, C., Amos, B., & Tautz, D. (1991). Conservation of polymorphic simple sequence loci in cetacean species. *Nature, 354*(6348), 63.

Schwengel, D. A., Jedlicka, A. E., Nanthakumar, E. J., Weber, J. L., & Levitt, R. C. (1994). Comparison of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. *Genomics, 22*(1), 46-54.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters, 9*(5), 615-629.

Sharma, P. C., Grover, A., & Kahl, G. (2007). Mining microsatellites in eukaryotic genomes. *Trends in Biotechnology, 25*(11), 490-498.

Shen, Z., Qu, W., Wang, W., Lu, Y., Wu, Y., Li, Z., . . . Zhang, C. (2010). MPprimer: a program for reliable multiplex PCR primer design. *BMC Bioinformatics, 11*(1), 143.

Sillero, N., Campos, J., Bonardi, A., Corti, C., Creemers, R., Crochet, P.-A., . . . Gonçalves, J. (2014). Updated distribution and biogeography of amphibians and reptiles of Europe. *Amphibia-reptilia, 35*(1), 1-31.

Sint, D., Raso, L., & Traugott, M. (2012). Advances in multiplex PCR: balancing primer efficiencies and improving detection success. *Methods in Ecology and Evolution, 3*(5), 898-905.

Sipos, R., Szekely, A. J., Palatinszky, M., Revesz, S., Marialigeti, K., & Nikolausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiol Ecol, 60*(2), 341-350.

Skei, J. K., Dolmen, D., Rønning, L., & Ringsby, T. H. (2006). Habitat use during the aquatic phase of the newts Triturus vulgaris (L.) and T. cristatus (Laurenti) in central Norway: proposition for a conservation and monitoring area. *Amphibia-reptilia, 27*(3), 309-324.

Skorinov, D. V., Kuranova, V. N., Borkin, L. J., & Litvinchuk, S. N. (2008). Distribution and conservation status of the smooth newt (Lissotriton vulgaris) in Western Siberia and Kazakhstan. *Russian Journal of Herpetology, 15*(2), 157-165.

Sotiropoulos, K., Eleftherakos, K., Tsaparis, D., Kasapidis, P., Magoulas, A., & Legakis, A. (2009). New polymorphic microsatellite loci for the Greek smooth newt, Lissotriton vulgaris graecus, and their utility in the nominotypical subspecies. *Molecular Ecology Resources, 9*(1), 292-295.

Squirrell, J., Hollingsworth, P., Woodhead, M., Russell, J., Lowe, A., Gibby, M., & Powell, W. (2003). How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology, 12*(6), 1339-1348.

Stuart, S. N., Chanson, J. S., Cox, N. A., Young, B. E., Rodrigues, A. S., Fischman, D. L., & Waller, R. W. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science, 306*(5702), 1783-1786.

Subramanian, S., Mishra, R. K., & Singh, L. (2003). Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biology, 4*(2), R13.

Tang, J., Baldwin, S. J., Jacobs, J. M., van der Linden, C. G., Voorrips, R. E., Leunissen, J. A., . . . Vosman, B. (2008). Large-scale identification of polymorphic microsatellites using an in silico approach. *BMC Bioinformatics, 9*(1), 374.

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research, 17*(16), 6463-6471.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., & McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome research, 11*(8), 1441-1452.

Thiel, T., Michalek, W., Varshney, R., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theoretical and Applied Genetics, 106*(3), 411-422.

Tóth, G., Gáspári, Z., & Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome research, 10*(7), 967-981.

Valdes, A. M., Slatkin, M., & Freimer, N. B. (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics, 133*(3), 737-749.

Vallone, P. M., & Butler, J. M. (2004). AutoDimer: a screening tool for primer-dimer and hairpin structures. *BioTechniques, 37*(2), 226-231.

Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology, 23*(1), 48-55.

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. d. F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and Molecular Biology, 39*(3), 312-328.

Vuorio, V., Tikkanen, O.-P., Mehtätalo, L., & Kouki, J. (2015). The effects of forest management on terrestrial habitats of a rare and a common newt species. *European journal of forest research, 134*(2), 377-388.

Wang, X., Lu, P., & Luo, Z. (2013). GMATo: A novel tool for the identification and analysis of microsatellites in large genomes. *Bioinformation, 9*(10), 541-544.

Wielstra, B., Baird, A., & Arntzen, J. (2013). A multimarker phylogeography of crested newts (Triturus cristatus superspecies) reveals cryptic species. *Molecular Phylogenetics and Evolution, 67*(1), 167-175.

Wielstra, B., Bozkurt, E., & Olgun, K. (2015). The distribution and taxonomy of Lissotriton newts in Turkey (Amphibia, Salamandridae). *ZooKeys*(484), 11.

Wilson, I. G. (1997). Inhibition and facilitation of nucleic acid amplification. *Applied and environmental microbiology, 63*(10), 3741.

Yang, X., Xu, Y., Shah, T., Li, H., Han, Z., Li, J., & Yan, J. (2011). Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. *Genetica, 139*(8), 1045.

Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Mol Ecol, 11*(1), 1-16.

Zardoya, R., Vollmer, D. M., Craddock, C., Streelman, J. T., Karl, S., & Meyer, A. (1996). Evolutionary conservation of microsatellite flanking regions and their use in

resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proc. R. Soc. Lond. B, 263*(1376), 1589-1598.

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J Comput Biol, 7*(1-2), 203-214.

Appendix

The DNA quantification

***Table A1**: NanoDrop ND-1000 Spectrophotometer DNA quantification*

Sample ID	260/280	260/230	ng/µl	Elution step
Sos.16.01	1.95	2.17	85.0	First elution step (with 50 µl elution buffer)
Sos.16.02	1.99	2.29	155.3	
Sos.16.10	1.87	1.72	74.6	
Sos.16.11	1.88	1.69	79.7	
Sos.16.01	1.88	1.69	79.7	Second elution step (with 30 µl elution buffer)
Sos.16.02	1.86	2.02	34.7	
Sos.16.10	1.80	1.82	23.2	
Sos.16.11	1.90	2.12	74.0	
Sos.16.01	1.97	2.07	75.0	Both elution steps pooled together
Sos.16.02	2.00	2.18	110.9	
Sos.16.10	1.97	2.03	50.3	
Sos.16.11	1.96	2.17	91.5	

***Table A2**: NanoDrop quantification of DNA samples isolated from two smooth newt individuals*

Sample ID	260/280	260/230	ng/µl	Elution step
LISS 2	1.94	2.34	1103.1	First elution step (100 µl el-buffer)
LISS 3	1.92	2.35	1085.6	
LISS 2	2.02	2.31	438.5	Second elution step (70 µl el-buffer)
LISS 3	1.99	2.33	209.9	
LISS 2	2.00	2.14	44.2	Third elution steps (30 µl el-buffer)
LISS 3	1.96	2.35	70.8	

***Table A3**: NanoDrop quantification of DNA samples isolated from five smooth newts and one great crested newt individual*

Sample ID	260/280	260/230	ng/µl	Elution step
Lv.RK1	1.88	2.30	514.1	First elution step (with 100 µl elution buffer)
Lv.RK2	1.89	2.31	453.8	
Lv.RK3	1.91	2.30	242.9	
Lv.RK4	1.90	2.28	203.5	
Lv.RK5	1.89	2.33	549.8	
Tc.RK6	1.92	2.28	295.8	
Lv.RK1	1.92	2.29	35.6	Second elution step (with 50 µl elution buffer)
Lv.RK2	1.89	2.29	145.0	
Lv.RK3	1.89	2.32	165.2	
Lv.RK4	1.73	1.97	28.1	
Lv.RK5	1.91	2.21	101.3	
Tc.RK6	1.91	2.30	221.7	
Lv.RK1	1.92	2.29	352.3	Both elution steps pooled together
Lv.RK2	1.92	2.30	352.6	
Lv.RK3	1.92	2.28	231.2	
Lv.RK4	1.91	2.31	145.3	
Lv.RK5	1.92	2.31	405.3	
Tc.RK6	1.94	2.27	264.9	

Primer validation

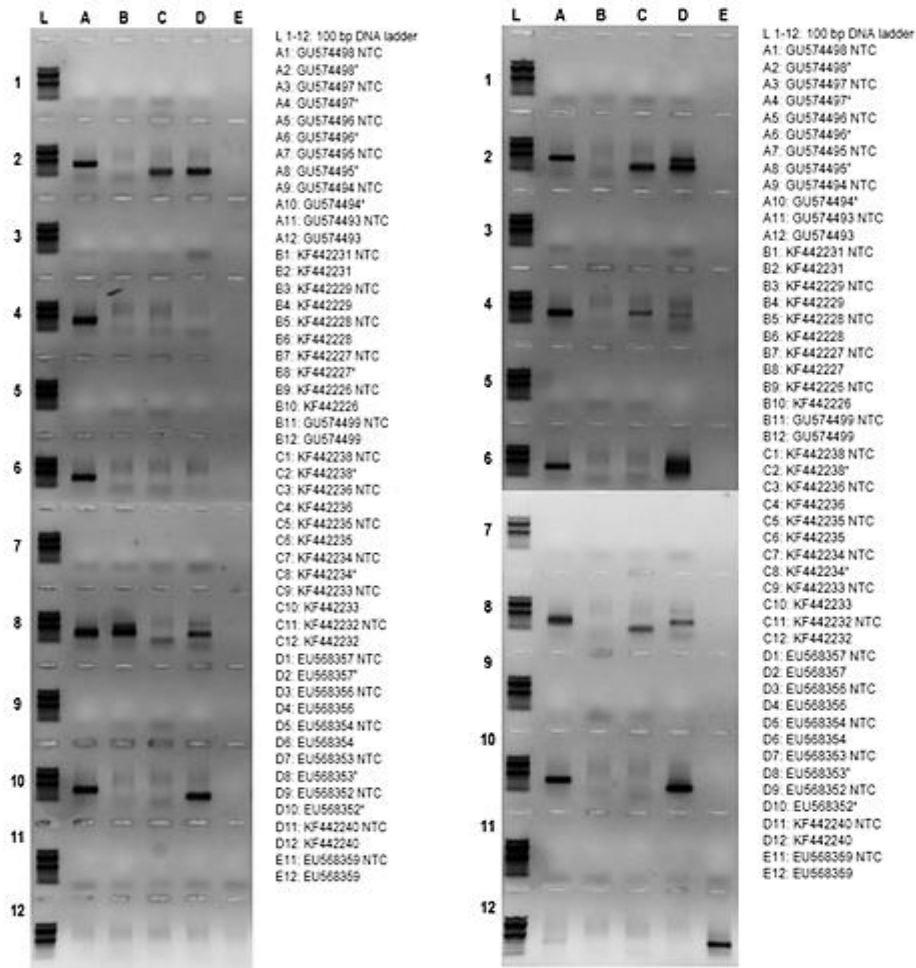


Figure A1**: Amplification testing of primers on smooth newt* (Lissotriton vulgaris) *DNA samples at 58 °C (left) and 52 °C (right) primer annealing temperatures. The primer pairs that generated amplicons are labelled with an asterisksign.*

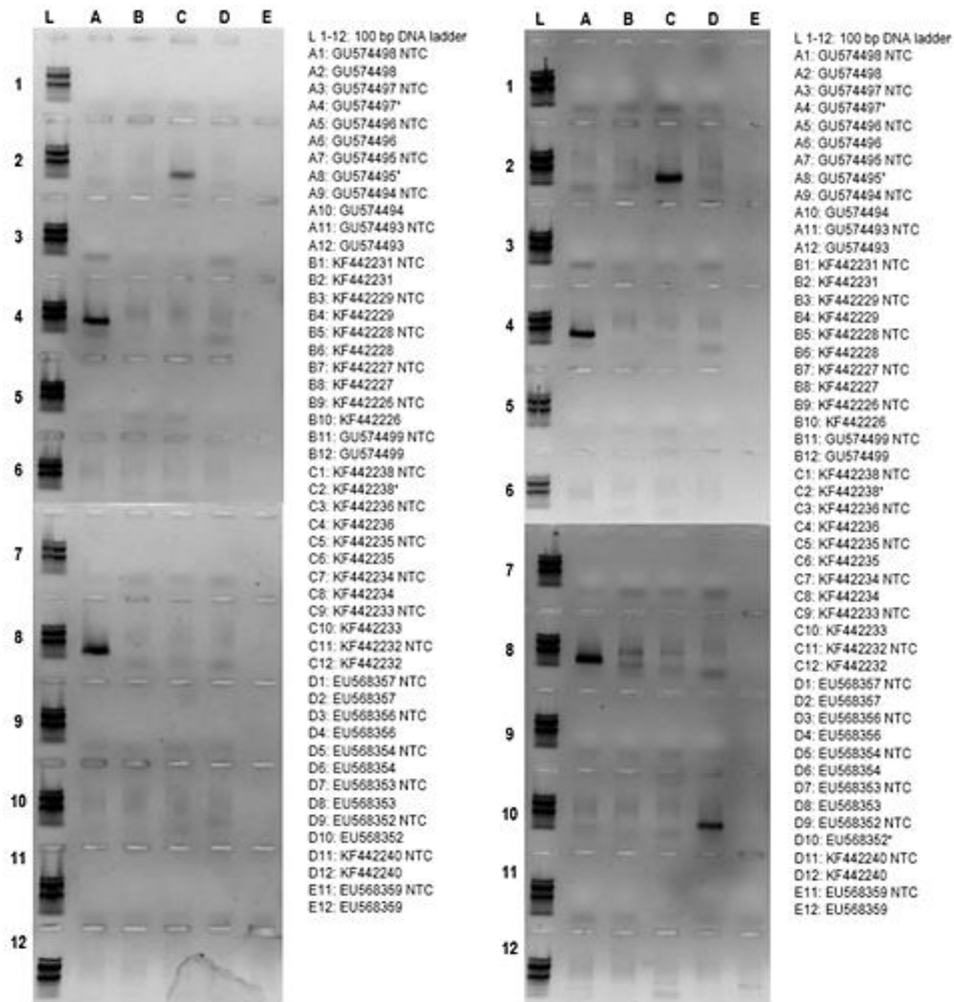


***Figure A2**: Cross-species amplification testing of primers on great crested newt* (Triturus cristatus) *DNA samples at 58 °C (left) and 52 °C (right) primer annealing temperatures. The primer pairs that generated amplicons are labelled with an asterisk*sign.*

Multiple sequence alignments of the sequences retrieved from direct sequencing of both top (F) and bottom (RC) strands of PCR products of seven markers (Lv2, Lv7, Lv10, Lv19, Lv25, Lv31, and Lv39) against the contig sequence from which the respective primers were designed. The highlighted sequence represents tandem repeats.

```
CLUSTAL O (1.2.4) multiple sequence alignment


Lv2-RC            ------------------------------------------------------------ 0
Lv2-F             ------------------------------------------------------------ 0
Contig_10921      TAACGCCGCATATCAGCGACGACCACATTCCTATGGATAGTCATTAGGACAATGTCTCAG 60


Lv2-RC            -TATATAAAAGGCGGCGAGAAATAATGCCTC----------------------------- 30
Lv2-F             ------------------------------------------------------------ 0
Contig_10921      TTATCTACTCGGGATAAAACTATAATTACGACTTTTAGTATGGTTCCCAAATTATATGAT 120


Lv2-RC            ------------------------------------------------------------ 30
Lv2-F             ------------------------------------------------------------ 0
Contig_10921      GTACCTCCTATTCTTGGCATCAAAGGTCCACTCACTCCCTCTTTGTTCGATCACCATTTA 180


Lv2-RC            ------------------------------------------------------------ 30
Lv2-F             ------------------------------------------------------------ 0
Contig_10921      CCTGATCGGGATTTCACATCCGTTCACATAAACAACTCTCAACTAGGAGATGCATTTACC 240


Lv2-RC            ------------------------------------------------------------ 30
Lv2-F             ------------------------------------------------------------ 0
Contig_10921      AAGATGGAACAGGAACACAAGGCACACGGTAATTCTATAAGAATTAGGATATTTATTGAC 300


Lv2-RC            ------------------------------------------------------------ 30
Lv2-F             ------------------------------------------------------------ 0
Contig_10921      ACCACATGCAAATCACATCACTATGCAACAATAATTTCTACACAGGAAAGATAAAGATTA 360


Lv2-RC            ------------------------------------------------------------ 30
Lv2-F             ------------------------------------------------------------ 0
Contig_10921      GAATGCGCGCTAATTGGGAAAAGGGCGGAGGTCACCTGGTTAATATGAAGTCGGGACTCG 420


Lv2-RC            ------------------------------TGCTCTGTGTTGTCCGCGCGGGACACCC-- 58
Lv2-F             --ATTTTTTTGGCTCCTTCTCGTCTCGGCGACTCCTGAGCTCGTCTGGTTGTCCACCCCA 58
Contig_10921      GCAGTCATCTAGGTTCCCTCTTTGTCTCTGTGTCCTGGCTCGTCTCGCAGTCACACCCCA 480
                                                    ***                *****

Lv2-RC            -------CACGCACACAAAATCATTGTCAGCCTAGCAAACGTCTTGTTCCCCGGACGG-- 109
Lv2-F             GCCGTCACAAGTACACGTCTGCACCGCTAGGCTCAACCTCGTCTGCTTCCCTCCAGGCCA 118
Contig_10921      GCCGTCACAAGTACACGTCTGCACCGCTAGGCTCAACCTCGTCTGCTTCCCTCCAGGGCA 540
                         ** * ****     **  *  ** **      *****  *****   * *

Lv2-RC            --GGACCTGGGTTTATTATCCGCACTCTCCCCTCTCTT---------TTCGCCGCGCGCG 158
Lv2-F             CTGGAACCTTGCTGGTTATTAGCTACCGTATCTCTCTCTCTCTGTGTGTGGGCGCCCCCC 178
Contig_10921      CTGGAACCTTGCTGGTTATTAGCTACCGTATCTCTCTCTCTCTCTGTTTCAGTGCGCTCC 600
                    *** *   * *  ****  **   *    ******           *    ** * *

Lv2-RC            CTATAAATCCACAAAA--AAACCCCAAAA---------------------------     186
Lv2-F             CCCAATATCTCCCCCAC-AGTCTCTCAAAAAAGAGCGGGGGG--------------     219
Contig_10921      GCGCTAATTCTCCTTCAGTCTTTCTCAGAGAGTCTCGGTCTTAAATCTTTCTCGGA     656
                        **   *           * ** *
```

```
CLUSTAL O (1.2.4) multiple sequence alignment


Lv7-RC          ------------------------------------------------------------      0
Lv7-F           ------------------------------------------------------------      0
Contig_19806    CTCGCATACAACCTAAACCACACACAACCTCCCACATGCTGAACACAATGCACACACCAA      60


Lv7-RC          --------------------------------TAAACACAACTATACCCCACTCTGAACA      28
Lv7-F           ------------------------------------------------------------      0
Contig_19806    ACTGCACATGAGCGCCACCAGCCACACACTCTAAAACACAACTCTCCCACACTCTGAACA      120


Lv7-RC          ATATACTCACA---CAAACAGCACCAGTGCAGCAGCCGCGCAGTCACACGCTAAATCGCA      85
Lv7-F           -----------------AGGAAGGGGGCCAGACACGCTGTAGAGATCACGCTAAATCGCA      43
Contig_19806    CTACACACACACACCAAACAGCACCAGTGCACCAGCCGCACAGTCACACGCTAAATCGCA      180
                                 *        *     **  *         **************

Lv7-RC          CACAACCTCCCACACGCTGAACCACAACACATATCAAACTCTCCAAAGCACAACAACTAT      145
Lv7-F           CACAACCTCCCACACGCTGAACCACAACACATATCAAACTCTCCAAAGCACAACAACTAT      103
Contig_19806    CACAACCTCCCACACGCTGAACCACAACACATATTAAACTCTCCAAAGCACAACAACTAT      240
                ********************************** *************************

Lv7-RC          ACACTAGTCCTTACATACACCCATACACACTGGACCGCTAAAACTCACACACCAGCCACA      205
Lv7-F           ACACTAGTCCTTACATACACCCATACACACTGGACCGCTAAAACTCACACACCAGCCACA      163
Contig_19806    ACACTAGTCCTTACATACACCCGCACACACTGGACAACTAAAACTCACACACCAGCCACA      300
                **********************  ***********  ***********************

Lv7-RC          CACGCCGAACCACATGCATCAAGACAAACTCACAGCGTCCATAAACTCCTCCACCACGCA      265
Lv7-F           CACGCCGAACCACATGCATCAAGACAAACTCACAGCGTCCATAAACTCCAAACCACGCAC      223
Contig_19806    CACGCCGAACCACATGCATCAAGACAAACTCACAGCGTCCATAAACTCCAAACCACGCAC      360
                *************************************************    *

Lv7-RC          CGATATCCTATAGACTTTCCTGTCACAAAAGACATACGCCCCAATAACT-----------      314
Lv7-F           GTATGTGTAGACACCAAACTA----CAAAAGACATACCCACATGAATCTACACAAAAGGC      279
Contig_19806    ATATGAATAGACACCAAACTA----CAAAAGACATACACACATGAATCTACACAAAAGGC      416
                  **          *   *      ************ * *     * **

Lv7-RC          ----------------------------------------  314
Lv7-F           AAACAGGGGTAA----------------------------  291
Contig_19806    AAACAGGGGTAAGCACACTAATGCTCACACAGGAGCACAT  456

CLUSTAL O (1.2.4) multiple sequence alignment


Lv10-RC         ------------------------------------------------------------ 0
Lv10-F          ------------------------------------------------------------ 0
Contig_20817    GTCACGTTCCACCATGGTTGCAAGGTCAGGCAACGGTCCATGGACCGCTGGTGCCCTGCA 60


Lv10-RC         ------------------------------------------------------------ 0
Lv10-F          ------------------------------------------------------------ 0
Contig_20817    ACACATAGGTGATCAAAGAAAAATTGTCAGTCCCCACACCTATGTTGCATTGACCATACT 120


Lv10-RC         ------------------------------------------------------------ 0
Lv10-F          ------------------------------------------------------------ 0
Contig_20817    CAACATCACATGCGATGTCATGACGGGAAACAAGTACTTGATGTGTGTGCCGTTGTATAT 180


Lv10-RC         ------------------------------------------------------------ 0
Lv10-F          ------------------------------------------------------------ 0
Contig_20817    GGACATATGTGTCACCAACATGCCCATCATGCAACATCTCAGTTCAATGTGTGTACACAA 240


Lv10-RC         ------------------------------------------------------------ 0
```

```
Lv10-F              ------------------------------------------------------------ 0
Contig_20817        CAAGCTCAATGTTGTGCTATGGCATTTCATTAGGCTGTGTAACATTACCTATGTACTGCT 300


Lv10-RC             -----CTTACTACACCCCCCCCTCACTACAAGTGAGGACAGATGTGTCCCAGGTA-TCCC 54
Lv10-F              ------ACCACTTTGGGGGTGGTTTTTTTTTTTTGGGCCGTTAGTGTGCCAGGTATCCAG 54
Contig_20817        GTCACTGCCCAAACACACACCACAGTGACAAGTGAGGACAGATGTGTCCCAGGTA-TCAC 359
                                                    *  ** *    **** *******  *

Lv10-RC             AGCTCTCCGGAAGGTAGATTCCCAGGGTCCCTGGATTTCCACATGGACGCTGGCATGACA 114
Lv10-F              CGGGGGGGGGAGGTAGCCTTCCCAGGGACCCTGGTTTGCCACATGGACACTGGCATGACA 114
Contig_20817        AGCTCTCCGGAAGGTAGATTCCCAGGGACCCTGGATTGCCACATGGACACTGGCATGACA 419
                     *      *** *     ********* ****** ** ********** ***********

Lv10-RC             TGGCATATCATGGACATGTCAATAAACATCCATCCCC-CCACAGGTAGTGGCCCCTGTAG 173
Lv10-F              TGGCATATCATGGACATGCCAATAAACAACCATGACTACCACAGGTAGTGGCCCCTGTAG 174
Contig_20817        TGGCATATCATGGACATGCCAATAAACAACCATGACCACCACAGGTAGTGGCCCCTGTAG 479
                    ****************** ********* ****  *  **********************

Lv10-RC             TCTCGGCAGGACATGGGATGGACATCTGTCCCCACACCCCCCCACCATCAATGTTCTTCC 233
Lv10-F              TCTCTGCAGGACATACATGGACATCTGTGAGACACACACACACACTCAATGTTAATGC 234
Contig_20817        TCTCTGCAGGACATACATGGACATCTGTGAGACACACACACACACCATCAATGTTAATGC 539
                    **** *********     *            ***** * * ***  ********  * *

Lv10-RC             CCTCCCCCACAGGGTTGGGATTTTCTGAAAAAAAAAAACCCCAAA-ACACC---------- 282
Lv10-F              ACACACACAATGGATGGGGGAAACTGTCTGACATATGGCACATGCATATCAGTCAG---- 290
Contig_20817        ACACACCACATGGTTGTGGA-AACTGTCTCACATTTGGCACATGGGAATCAGAGACATCT 598
                     * * *     ** *  **            * *     * **      * *

Lv10-RC             ------------------------------------------------------------ 282
Lv10-F              ------------------------------------------------------------ 290
Contig_20817        GAAACACAACATGCACACAGACTGTTTATGTACGTCACAATTGTGGCAATGGGTCACTCT 658


Lv10-RC             ------------------------------------------------------------ 282
Lv10-F              ------------------------------------------------------------ 290
Contig_20817        GTGATGTACATGCCATCCCACCTACGACCCGTGTGGTTTGTGACATGACATACACATATG 718


Lv10-RC             ------------------------------------------------------------ 282
Lv10-F              ------------------------------------------------------------ 290
Contig_20817        CACACAGGAGATGTGACCATGCCACTTGCTCTGAACTGACTAGGTGCACATCACAGTGGT 778


Lv10-RC             ------------------------------------------------------------ 282
Lv10-F              ------------------------------------------------------------ 290
Contig_20817        CATGTGCACAACACACAAAGGATTCTGTCCCCATAAGCACGTGCACAGCCTGGGAGTGGC 838


Lv10-RC             ------------------------------------------------------------ 282
Lv10-F              ------------------------------------------------------------ 290
Contig_20817        TGCCCCCAAGTGACACCCAACAAACCACATGTTGACCCACACACACAAACACAGGTAACG 898


Lv10-RC             ------------------------------------------------------------ 282
Lv10-F              ------------------------------------------------------------ 290
Contig_20817        CAGCCTTATCTGACTTAGGACACAGTGTGGCACAGGCGGTGGTGTTTTGGTTGTGCACCG 958


Lv10-RC             ------   282
Lv10-F              ------   290
Contig_20817        GAATCC   964
```

CLUSTAL O (1.2.4) multiple sequence alignment

```
Lv19-F              ----TTGGTTATTCCTCCCCCGGGGGGG-GGGAGGAGGAGGGTTGGTCAAAGGGGGGGGT 55
Contig_26776        GGTAAGGGTCATGCCATTTACTAAGGCTTTGTAGCCACCGGGGAGAGAGAGATTGATAGA 60
```

```
Lv19-RC           -----------------GCACCAAGAA--AAAAAAAAAATCCTGGGTCCATGTGGCGGTT 41
                                     *   *       *           *    *    *

Lv19-F            AGATGCAGGGAGGGATTACTTTCTCTCAATTCATATTGCAATCAATCAATCCACCGTTGG 115
Contig_26776      AGGGTCAGCAGGAG----ACTTTTCTCAATCAATCAATCAATCAATCAATCCGGGAATTT 116
Lv19-RC           CTATCTCACAGTGTTCTATCTGTTAGGACCCTGTC-CGTGAGGCACCTACCCAGGGATTG 100
                                     *  *   *     *      *   * * * **     *

Lv19-F            AAATTCCCGGTGAAAAAATT-CGGGCTTCCATAGGAC------GCACTAAGCGGGGGTGT 168
Contig_26776      GTA----AAGCGCACTACTTACCCGCTAGGG----TC------TCAAGGCGCTGGGGGGG 162
Lv19-RC           CAATAAGAATTGCGAAAACACCCCACTAGCTTCCCCCTTCGGTTCACTTCTCCTCCACTC 160
                    *        *    *    *   **         *       **     *

Lv19-F            GTGTTGTCCTTTTAAATAGGGGGGGGTGGGAGGGACTCCCTGTGTTTTTTTGTTGTTGTGG 228
Contig_26776      GGGGGGGAGCTGCAGCTACTGGTCG------AAGAGCCAGGTCTTGAG------------ 204
Lv19-RC           CCCATCCCGTTACTTCTCAACCCCCGAAAAAAAAAACCAAAG------------------ 202
                            *     *                    **

Lv19-F            GG      230
Contig_26776      --      204
Lv19-RC           --      202
```

CLUSTAL O (1.2.4) multiple sequence alignment

```
Contig_41708      CACTGTCATTCTTCATGTCTCACGTGTACTGTTTAGCCTATTGTTGCATCTCTCTCTCTG       60
Lv25-F            ------------------------------------------------------------       0
Lv25-RC           ------------------------------------------------------------       0

Contig_41708      TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCACTCCTCCACACTTTCACTCTTCCT       120
Lv25-F            --------------------------------TTCTGGCGTCAACGCGGGGGAAAAAAAT       28
Lv25-RC           -------------------------------GGCGCCGCCGCCACGCAATAAAAAAAAAA       29
                                                      *  *  ** *

Contig_41708      AACGACTGGCACCTAATCACC------------------------TAATTGGCAAACCTG       156
Lv25-F            GCGGTGGGGCTTTTGCTACCCCTTTTGTGG------------------------------       58
Lv25-RC           AAAGAGGGGGCCGCCCCCCCCCTTTTTTAAAACGAAAAAAACCTGTGGTCTCCAATGCTG       89
                     *   **          **

Contig_41708      GTGCACCACCCAATTGGCTCCAAATACCACAAGTATTTGCACAAATTTGTCTTAAATAAA       216
Lv25-F            ------------------GGGGGGGAAAAAAAAACTACTCCC-TATG-CTTTTCGACTCC       98
Lv25-RC           CTACTACTACTACTGCCCGCCTACTAAAAAAAAACTACGACG-ACTT-TTTTGTCCCCCC       147
                                           *  * **   *         *   * *

Contig_41708      AATTGTAGATGCCATGTTACTACCTCTCCTTTCTGCTAAATAATGATAAAAAATAAATTG       276
Lv25-F            GAATGA---TCCTC---GGAGTCCTG-ATTGATAGGGCTGATAAGAAG----CAGTTTTG       147
Lv25-RC           TCTTATGGGGGGGTGGTGAAGTCCTCTATTGTTTTCTCCCCTTTCACC----ACGTGTTG       203
                     *                  ***     *               *          ***

Contig_41708      GCCAAGAAGGAAAGGACAGGTAAGACGACT   306
Lv25-F            GGGGTTGAAAAGACAACAT-----------   166
Lv25-RC           CGTCGGGGGAAAAAAAGGGG----------   223
```

CLUSTAL O (1.2.4) multiple sequence alignment

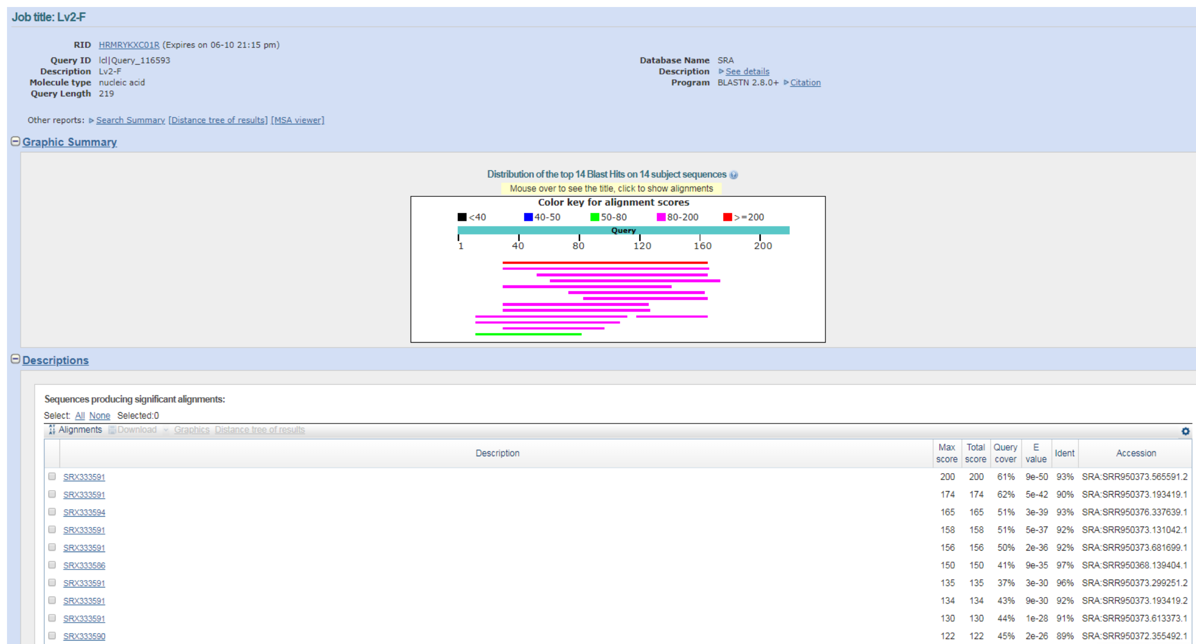
```
Lv31-F            ------------------------------------------------------------ 0
Lv31-RC           AAAAAATTAAAAAAACTTTAAAGGGGGCCCCCCCCCCCCCCTTTTTTTAAAAACACAAAA 60
Contig_57228      -------------------------------------------TTAGTCAGGTGGGCGTA 17

Lv31-F            --ACCAGCTGCACTACCCACTGCACGCTGCTAATCATCCGGAGAACGTGGCCCAGGCAGG 58
Lv31-RC           AAAAAAATACGACTCCCCTGTCCCCTTTTACAATGT--ATGTGCTGGGGCCAAACTAAGC 118
Contig_57228      GAAGCAA---------TTCACACCCCCTTTCAATGT--ATGTGCTGGCCAAACTAG--TA 64
                    *  *                * *  *   ***      * *   *
```

```
Lv31-F            AAGAGTAGGCATACGTCGC----A--CACA--TTCACACCCATTCAGGG--GGGAGGGGA 108
Lv31-RC           AAAAATGGTCCAAAATTTTCCGGACCCACCCTTACCCCCCCCCACGAACGCCACCGCTAA 178
Contig_57228      A----AATGCCAAATTTCACCACCACCACC--ACCCCCTCACCTCTAAC---ACCGGGAC 115
                  *        *  *  *          ***     * *  *    *          *

Lv31-F            GGTTGAGGGGAGAAAAAAAAGGGGGGGGGCCCCCACCAAAAAAAAAATATAAAAATAAAA 168
Lv31-RC           AACACCCCGGGGGAACCGGAGACCAAAAGACC---------------------------- 210
Contig_57228      GTGTCCAGGCAGGAAGTGGATCTCCACCATTCCTCCGCA-CCGCCGCTCTGCACAGTG-- 172
                          *  * **    *           *

Lv31-F            A 169
Lv31-RC           - 210
Contig_57228      - 172
```

CLUSTAL O (1.2.4) multiple sequence alignment

```
Lv39-RC          TTCAAAAAAAAATTTTCCCCCAAAGGCTTCTTTTTCAAAGGGCCCGGGGGAAAAAATAGA  60
Lv39-F           -------------------------GGGCCTGTTGCAAATGAGAC-----------TGCC  24
Contig_9006      --------------------ATTCAGAGTCACTGACAAACCAGCA-----------CTCA  29
                                          *   *  *  ****

Lv39-RC          AAGAAAACCTGCCCAAAAAAAAAAAAATTGGCACAATAGGATGGAAACCAAATTCCAACCA  120
Lv39-F           AAGGAAAATTGCACATAG----------------------------------ATGAACAA  50
Contig_9006      TCAAGAAGTTCCCCAAGC----------------------------------TCTTTCAA  55
                      **  * * **                                          * *

Lv39-RC          CCCTTGGTTGGGCCGCCAAACCACCCACCCGGAAAAAAAATTCCCCTCGTTTTCC-CCAC  179
Lv39-F           ATCACACCTGTGGCGC---------AACACCACCGAAAAATCCCGTTCCACCTGCAAGAA  101
Contig_9006      GGCTTGGGAAAAATGA---------AGAACTGCCAAAAAATCCCGTTCCACCTGCAAGAA  106
                   *           *              *     ****** **  **    * *   *

Lv39-RC          CCCTGCCAAAAGAAAAGCCCCGTTGGGAAGAAAAGGGAAACCTCCCAAAGAACCCTTGCC  239
Lv39-F           GCCGTGGAGAAGGAACTCCAAGACCTGCTGGAAGCCGACATCATCGA-------------  148
Contig_9006      GCCGTGGAGAAGGAACTCCAAGACCTGCTGGAAGCAGACATCATCGA-------------  153
                  **    * *** **  **  *    *  * **   ** * *  * *

Lv39-RC          TGGGAAAAGGCCCGGACCAATCCAATTCGAAGGAGGAAAACAAGGAATTGGGACCCCAAA  299
Lv39-F           ---GAGAACAGATGGACCAACTC-------------------------------------  168
Contig_9006      ---GAGAACGGATGGACCAACTC-------------------------------------  173
                    ** **     *******  *

Lv39-RC          CCTCCCCATGGGGGTCCTCCCCCCCGGGTTCCGGTGGGGTGGGTTCCCCCCAAAAGAAAA  359
Lv39-F           ------CATGGGT-------------CTCCCCGGTCGTGGTGGTCCCAAAGAAAGGA---  206
Contig_9006      ------CATGGGT-------------CTCCCCGGTCGTGGTGGTCCCAAAGAAAGGA---  211
                       ******                  ***** * *  *** **    *** **

Lv39-RC          AGGGAGGAAAAACCTGGCCCGGAAGTTCCGTTCTGTTGTGTAGACCATGAGGGCGCCCAA  419
Lv39-F           ---------GAAACTGGCCG-------AGTCCGTCTGTGTGTAGACATGAGGGCGCCCAA  250
Contig_9006      ---------GAAACTGACCG-------AGTCCGTCTGTGTGTAGACATGAGGGCGCCCAA  255
                           ** *** **              * *         ***************

Lv39-RC          CACGGCAATAGAGAGAGAGAGACATCCCAGGCCACCTCCAACATACAGTGTACACTGCCC  479
Lv39-F           CACGGCAATAGAGAGAGAGAGACATCCAGG-CCTCACATCAGTGACATGATCACTTCCCT  309
Contig_9006      CACGGCATTAGAGAGAGAGAGACATCCAGGCCCTCACATCAGTGACATGATCACTTCCCT  315
                 ******* *******************  * ** *     *   ***   *    * **

Lv39-RC          CCGCCC-----------------------------------------------------T  486
Lv39-F           CAATGGAGCCAAGATCTTTTAAA-ACTGGACCTCAACAAAGGATACCACCAACTAGAACT  368
Contig_9006      CAATGGAGCCAAGATCTTTTCAAAACTTGACCTCAACAGAGGATACCACCAACTAGAACT  375
                 *                                                          *

Lv39-RC          CCCC------GCTCGATCCACCCCAATTTTCTTCTAGGGCGACCGCCCCTT------     531
Lv39-F           GGACGAAAGG-----------------------------------------------     378
Contig_9006      GCACGAAAGGTCGAGATACATCACAACATTCTCGACACACCTTGGACTATTCAGATA     432
                    *
```

The sequences of four markers (Lv2, Lv7, Lv10, and Lv39) obtained from direct sequencing aligned against the unassembled reads



Job title: Lv2-F

RID HRMRYKXC01R (Expires on 06-10 21:15 pm)

Query ID lcl|Query_116593
Description Lv2-F
Molecule type nucleic acid
Query Length 219

Database Name SRA
Description See details
Program BLASTN 2.8.0+ Citation

Other reports: Search Summary [Distance tree of results] [MSA viewer]

Graphic Summary

Distribution of the top 14 Blast Hits on 14 subject sequences
Mouse over to see the title, click to show alignments
Color key for alignment scores
<40 40-50 50-80 80-200 >=200
Query
1 40 80 120 160 200

Descriptions

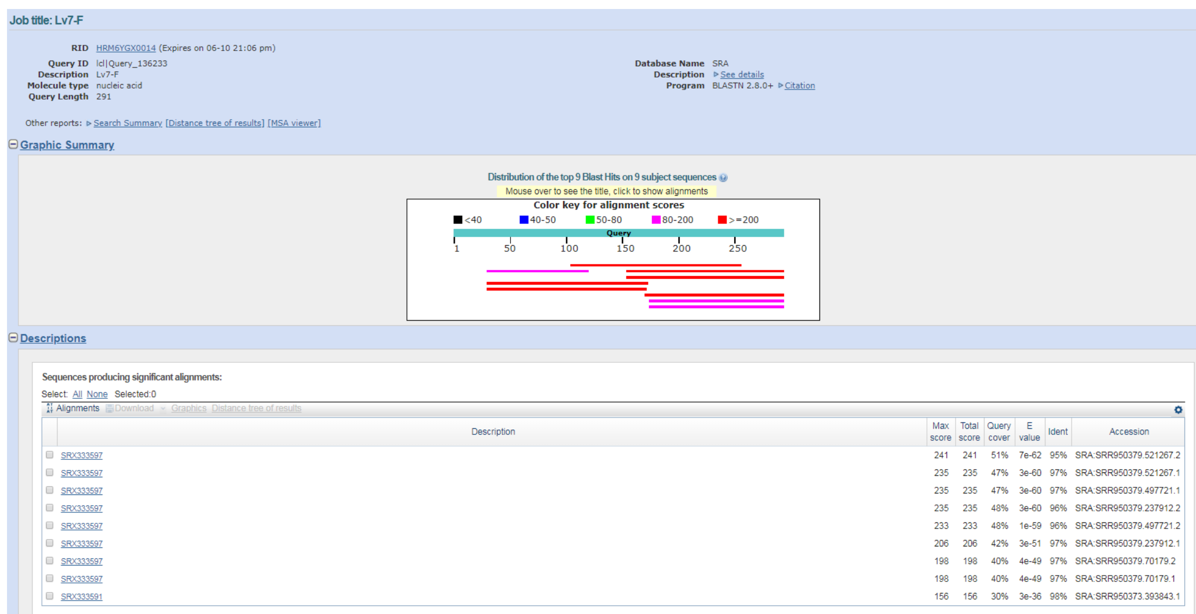
Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
SRX333591	200	200	61%	9e-50	93%	SRA:SRR950373.565591.2
SRX333591	174	174	62%	5e-42	90%	SRA:SRR950373.193419.1
SRX333594	165	165	51%	3e-39	93%	SRA:SRR950376.337639.1
SRX333591	158	158	51%	5e-37	92%	SRA:SRR950373.131042.1
SRX333591	156	156	50%	2e-36	92%	SRA:SRR950373.681699.1
SRX333586	150	150	41%	9e-35	97%	SRA:SRR950368.139404.1
SRX333591	135	135	37%	3e-30	96%	SRA:SRR950373.299251.2
SRX333591	134	134	43%	9e-30	92%	SRA:SRR950373.193419.2
SRX333591	130	130	44%	1e-28	91%	SRA:SRR950373.613373.1
SRX333590	122	122	45%	2e-26	89%	SRA:SRR950372.355492.1

***Figure 3**: The sequence of the marker Lv2 aligned against unassembled reads.*



Job title: Lv7-F

RID HRM6YGX0014 (Expires on 06-10 21:06 pm)

Query ID lcl|Query_136233
Description Lv7-F
Molecule type nucleic acid
Query Length 291

Database Name SRA
Description See details
Program BLASTN 2.8.0+ Citation

Other reports: Search Summary [Distance tree of results] [MSA viewer]

Graphic Summary

Distribution of the top 9 Blast Hits on 9 subject sequences
Mouse over to see the title, click to show alignments
Color key for alignment scores
<40 40-50 50-80 80-200 >=200
Query
1 50 100 150 200 250

Descriptions

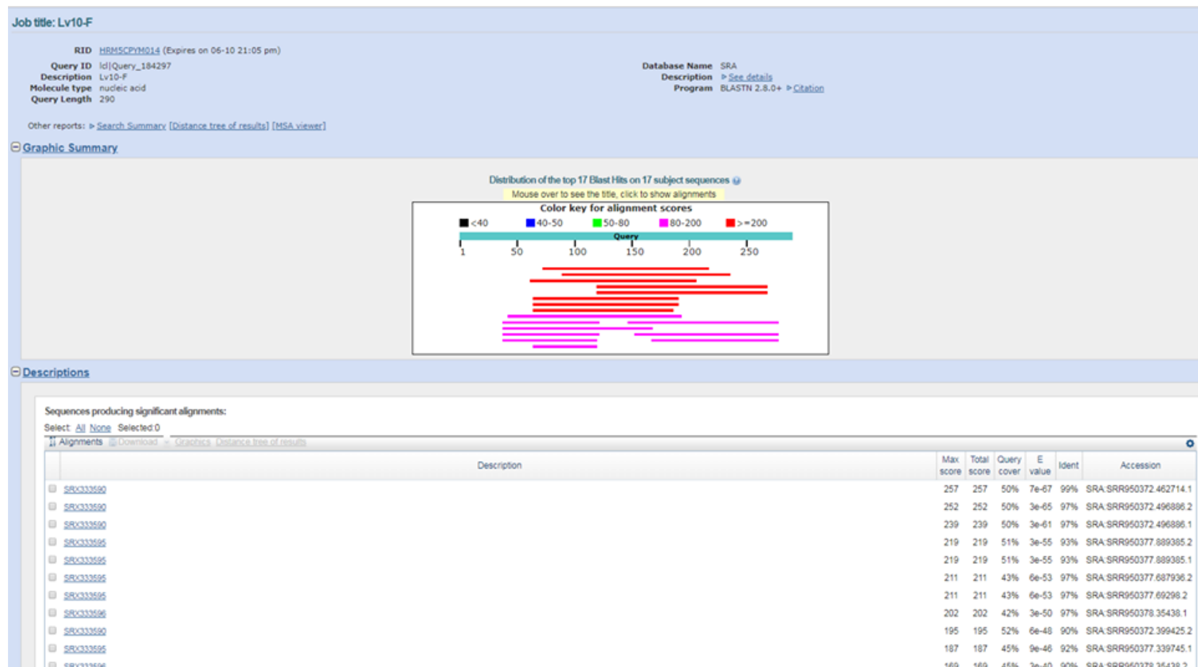
Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download Graphics Distance tree of results

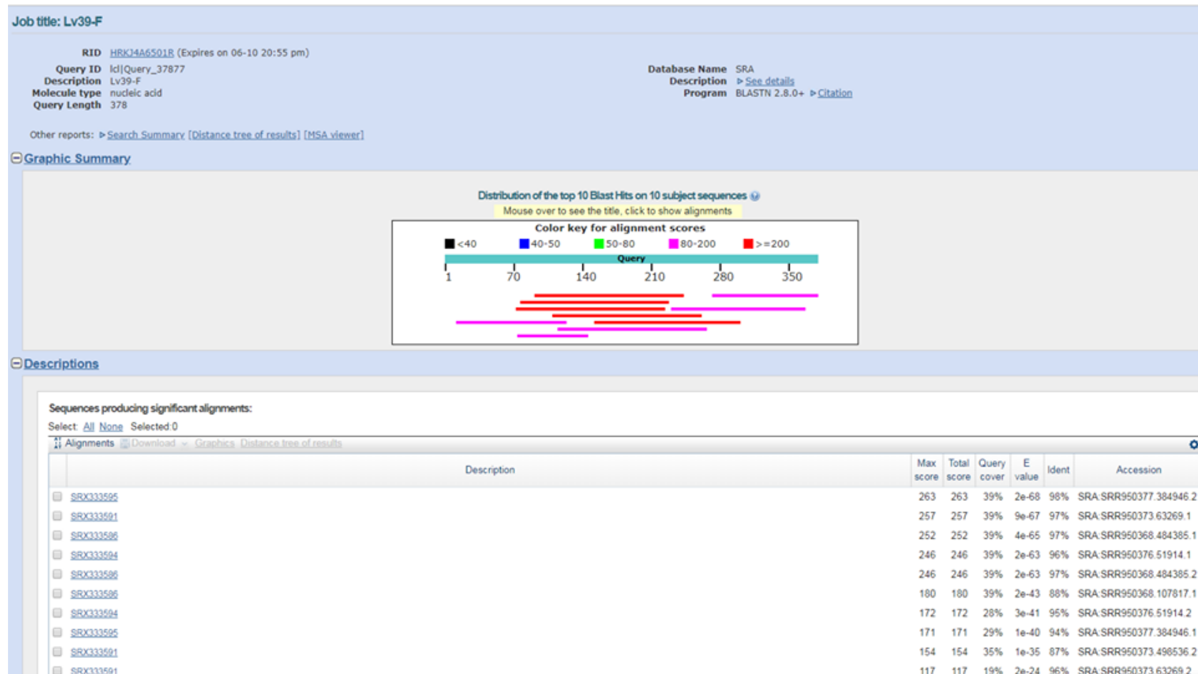
Description	Max score	Total score	Query cover	E value	Ident	Accession
SRX333597	241	241	51%	7e-62	95%	SRA:SRR950379.521267.2
SRX333597	235	235	47%	3e-60	97%	SRA:SRR950379.521267.1
SRX333597	235	235	47%	3e-60	97%	SRA:SRR950379.497721.1
SRX333597	235	235	48%	3e-60	96%	SRA:SRR950379.237912.2
SRX333597	233	233	48%	1e-59	96%	SRA:SRR950379.497721.2
SRX333597	206	206	42%	3e-51	97%	SRA:SRR950379.237912.1
SRX333597	198	198	40%	4e-49	97%	SRA:SRR950379.70179.2
SRX333597	198	198	40%	4e-49	97%	SRA:SRR950379.70179.1
SRX333591	156	156	30%	3e-36	98%	SRA:SRR950373.393843.1

***Figure 4**: The sequence of the marker Lv7 aligned against unassembled reads.*



Description	Max score	Total score	Query cover	E value	Ident	Accession
SRX333590	257	257	50%	7e-67	99%	SRA:SRR950372.462714.1
SRX333590	252	252	50%	3e-65	97%	SRA:SRR950372.496886.2
SRX333590	239	239	50%	3e-61	97%	SRA:SRR950372.496886.1
SRX333595	219	219	51%	3e-55	93%	SRA:SRR950377.889385.2
SRX333595	219	219	51%	3e-55	93%	SRA:SRR950377.889385.1
SRX333595	211	211	43%	6e-53	97%	SRA:SRR950377.687936.2
SRX333595	211	211	43%	6e-53	97%	SRA:SRR950377.69298.2
SRX333596	202	202	42%	3e-50	97%	SRA:SRR950378.35438.1
SRX333590	195	195	52%	6e-48	90%	SRA:SRR950372.399425.2
SRX333595	187	187	45%	9e-46	92%	SRA:SRR950377.339745.1

***Figure 5**: The sequence of the marker Lv10 aligned against unassembled reads.*



Description	Max score	Total score	Query cover	E value	Ident	Accession
SRX333595	263	263	39%	2e-68	98%	SRA:SRR950377.384946.2
SRX333591	257	257	39%	9e-67	97%	SRA:SRR950373.63269.1
SRX333586	252	252	39%	4e-65	97%	SRA:SRR950368.484385.1
SRX333594	246	246	39%	2e-63	96%	SRA:SRR950376.51914.1
SRX333586	246	246	39%	2e-63	97%	SRA:SRR950368.484385.2
SRX333586	180	180	39%	2e-43	88%	SRA:SRR950368.107817.1
SRX333594	172	172	28%	3e-41	95%	SRA:SRR950376.51914.2
SRX333595	171	171	29%	1e-40	94%	SRA:SRR950377.384946.1
SRX333591	154	154	35%	1e-35	87%	SRA:SRR950373.498536.2
SRX333591	117	117	19%	2e-24	96%	SRA:SRR950373.63269.2

***Figure 6**: The sequence of the marker Lv39 aligned against unassembled reads.*