Analysis of nuclear markers in two species with highly divergent mtDNA lineages in Iceland

ITS in Crangonyx islandicus

EF alpha in Apatania zonella

May 2010

Sonia Sanz Muñoz

LÍF016M Research project in biology for foreign students

Teacher: Snæbjörn Pálsson

University of Iceland Life and Environmental Sciences

1. Abstract

This project is a study of variation in two nuclear markers in two arthropods (*Crangonyx islandicus* and *Apatania zonella*). Both species have been found to have highly divergent mtDNA lineages within Iceland.

Crangonyx islandicus is an endemic groundwater amphipod species recently discovered in Iceland. Based on variation in mtDNA genes, COI and 16S RNA, Kornobis et al, (2010) concluded that this species had survived glaciations periods in sub-glacial refugia. The mtDNA variation defined several monophyletic groups, restricted to different geographic regions in Iceland and which have diverged in Iceland for up to 4 – 5 Million years. This was supported by a correlation between genetic and geographic distances among species. In this study we look at the internal transcribed spacer (ITS) gene, a piece of non-functional RNA situated between structural ribosomal RNAs (rRNA).

The results show different patterns from the mtDNA results in Kornobis et al. (2010), with a major split between two populations, between north and southern Iceland, and different partition among samples in southern populations. The main difference is characterized by a large size difference of the ITS1 region due to insertion or deletion, a highly variable microsatellite was funs within this region.

The second part of the project is based in the study of *Apatania zonella*, a caddisflie (Trichoptera), a circumpolar species which lives at high latitudes, in cold-clear water, streams, lakes and marshes.

This study is a continuation of a previous study "Mitochondrial variation of the caddisflies *Apatania zonella* and *Potamphylax cingulatus*" (Sanz, 2010). The former study showed that Iceland acts as a zone of admixture, where two populations of *A. zonella* with distinct mtDNA types have arrived, from both ends of its range distribution, one from North America and the other from Europe. In this study we use a nuclear marker, EF alpha, in order to know whether the structure obtained by mtDNA in North America and Europe, is reflected in a nuclear marker and also to verify whether individuals in Iceland of different geographic origin as defined by mtDNA have interbred in Iceland.

The results show less clear structure with EF alpha than found in mtDNA. Populations could be sexual or asexual depending on the country. Moreover, results show the rate of evolution of EF alpha is slower than COI's rate.

2. Table of contents

3. List of figures

ITS1 in *Crangonyx islandicus*

4. List of tables

ITS in *Crangonyx islandicus*

EF in *Apatania zonella*

ITS in Crangonyx islandicus

5. Introduction

Quaternary was the last of the five Ice Ages known in Earth's history. The Quaternary period started about 2,4Myr ago and ended about 10kyr ago; it was a period of extensive glaciations in both hemispheres interrupted by warmer interglacial phases (Geirsdóttir et al., 2007). The ice sheets advanced and receded with a roughly 41 - 100kyr cycle which have become increasingly dramatic (Hewitt, 2000). The Quaternary climatic oscillations have driven many changes in species ranges in all parts of the world, this is well known for the species in the Palearctic, where it was necessary for species to move, adapt or go extinct (Hewitt, 2000, 2004; Schmitt, 2007).

The species responses to climate change may vary for different organisms depending on their level of cold-tolerance and dispersal capabilities (Schönswette, 2004; Deffontaine, 2005; Pinceel, 2005). In boreal regions, these responses may have been larger as the environmental conditions are often very severe, making survival and reproduction more difficult. The concept of "refugia" and derivates has been important in thinking about the changing distributions and abundances of organisms especially in response to the climatic changes of the late Quaternary (Bennet and Provan, 2008). In addition to the impact on species range, the climate changes has left its marks on genetical patterns, which is in addition a product of the species' life history and dispersal patterns, geographic history and chance (Avise, 2007; Pauls, 2009). Isolation among refugial populations has further promoted genetic and phenotypic differentiation as a result of independent adaptation to local environments and genetic drift, in some cases with consequences for reproductive isolation between discrete refugial lineages and the creation of hybrid contact (Hewitt, 1999).

During the last glacial maximum (LGM) around 21,000BP, Iceland, a landmass (103,000km²) in the central North Atlantic (Geirsdóttir et al., 2009) where glaciations have been preserved, was almost completely covered by glaciers, as well as northern Europe and North America (Ehlers and Gibbard, 2007). These high-latitude areas and formerly glaciates are characterized by low endemism and species diversity, shallow genetic clades and little diversity within species (Sadler, 1999; Hewitt, 2004).The low species diversity and lack of endemism in the current Icelandic territorial biota (e.g. Buckland et al., 1986) can be explained by the short time since the retreat of the glaciers and the geographical isolation of the island. The inimitable geographic and climatic position is considered to enhance the island's importance in glacial geological and palaeoclimatic research (Geirsdóttir et al., 2007), and may similarly have an importance for phylo- and biogeographical studies.

In this study I look at the phylogeography of amphipods, an order of macroscopic crustaceans of the class Malacostraca which belong to the superorder Peracarida, whose life cycle is characterized by direct development and no independent larval

stage. The amphipod body is segmented throughout and usually laterally compressed, with a more or less curved or hook-like profile (fig. 1). Adult body lengths of freshwater species range 2-40 mm, most commonly between 5 and 15 mm. Most amphipods are marine, but they also inhabit a wide spectrum of freshwater habitats. Freshwater taxa are particularly diversified in relatively cool running waters and subterranean habitats (Väinölä R. et al., 2008). Amphipods are among the animals that characterize subterranean habitats world-wide, with more than 1000 amphipods species described from subterranean waters (Sket, 1999). Due to the diversity of amphipods in subterranean waters and their long presence there, subterranean amphipods are excellent tools for evaluation of biogeographical patterns (Holsinger 1977, 1986, 1993).

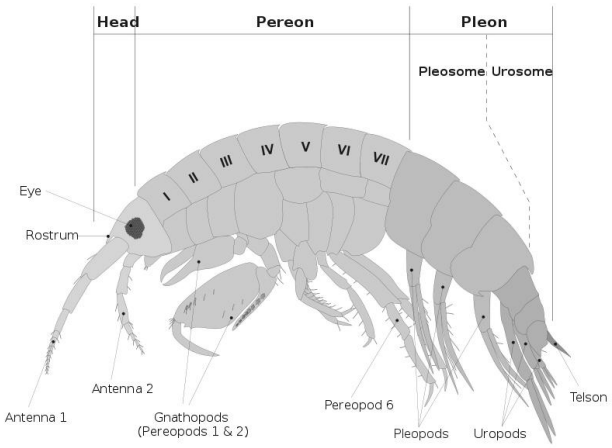


Figure 1. Amphipods anatomy.

Two endemic groundwaters arthropod crustaceans species, *Crangonyx islandicus* and *Crymostygius thingvallensis*, were recently discovered on the mid-Atlantic volcanic island of Iceland (Kornobis et al., 2010). *Crangonyx islandicus* was recorded in South, Southwest, West and Northeast Iceland from numerous springs emerging from relatively young (<10000 years), porous lavas (Svavarsson and Kristjánsson, 2006). The extent of morphological differences from closest relatives, endemism, and genetic patterns (Kornobis et al., 2010) suggests that these two species have survived glaciations periods in sub-glacial refugia. This survival may be explained by geothermal heat linked to volcanic activities, which may have maintained favourable habitats in fissures along the tectonic plate boundary in Iceland during glaciations. (Kornobis et al., 2010).

In Kornobis et al. (2010) five monophyletic groups were found, well supported with both genes COI and 16S (A and A' joined, B and C joined, D, E and F in fig. 2). The genetic variation is geographically structured and each monophyletic group represents a geographically distinct area (fig. 2). In addition to the clear phylogeographic structure, a strong correlation was found between geographic and genetic distances analysed for COI and 16S, separately and combined.

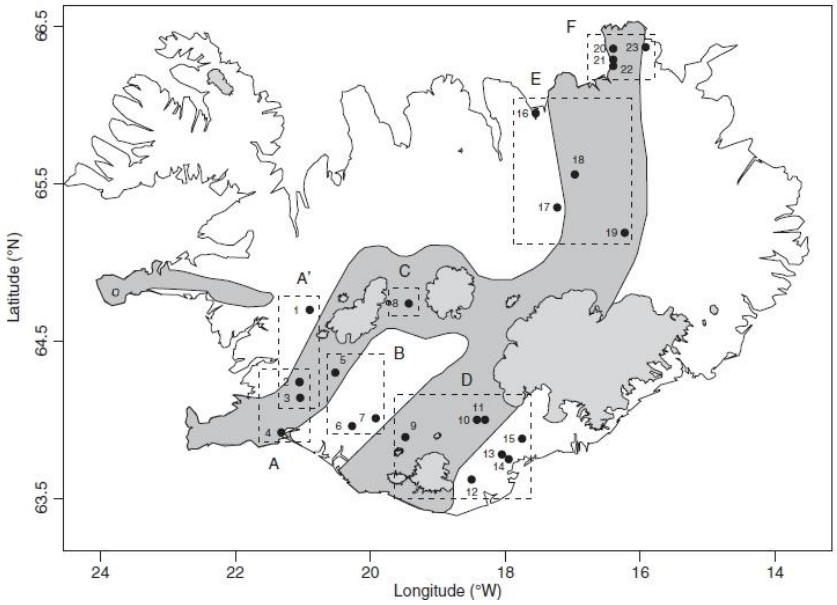


Figure 2. Sampling locations of groundwater amphipods in Iceland. The dotted squares represent the different groups A, A', B, C, D, E and F.

The aim of the project is to compare variation in a nuclear genomic marker with the variation observed in the mitochondrial COI and 16S rRNA genes and see whether it reflects the differences between the different monophyletic groups.

For this study we have chosen the Internal Transcribed Spacer (ITS) region and 5.8S, gene. ITS region is a piece of non-functional RNA situated between structural ribosomal RNAs (rRNA) on a common precursor transcript.

Read from 5' to 3', this polycistronic rRNA precursor transcript contains the 5' external transcribed sequence (5'ETS), 18S rRNA, ITS1, 5.8S rRNA, ITS2, 28S rRNA and finally the 3'ETS (fig. 3). During rRNA maturation, ETS and ITS pieces are excised and degraded because they are non-functional products. Genes encoding ribosomal rRNA and spacers occur in tandem repeats that are thousands of copies long, each separated by regions of non-transcribed DNA termed intergenic spacer (IGS) or non-transcribed spacers (NTS).

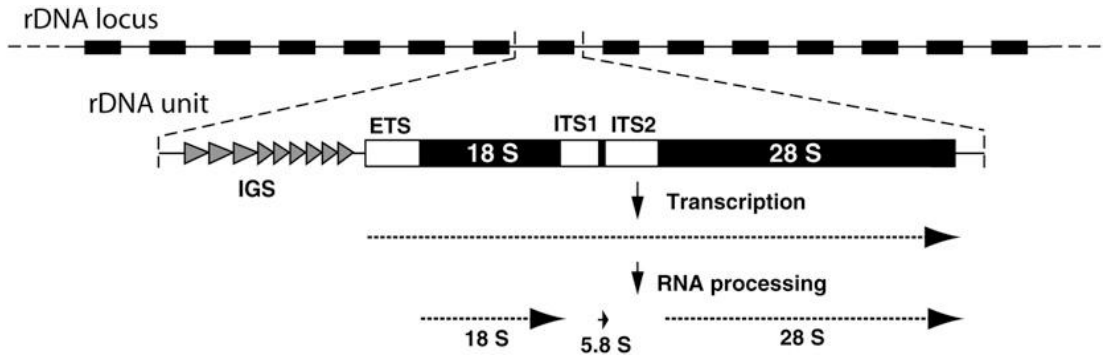


Figure 3. Organization of the rRNA gene.

As ITSs sequences show more divergence than their flanking regions and are easily amplified, they have been routinely used to distinguish related species and to infer phylogenetic relationships from populations to families and even higher taxonomic levels (Han-Lieng, 2006).

In a review (Chu, 2001) showed that:

- ITS1 variations are high among different taxonomic groups of crustaceans.
- Variations between congeneric species appear to be genus-specific, making ITS1 a phylogenetically informative marker in some genera, or a valuable diagnostic tool in others.
- ITS1 is potentially applicable in the analysis of crustacean population structure.
- Variations within individuals may obscure phylogenetic and population analyses and thus must be addressed in individual cases.

6. Material and methods

6.1. Samples

Twelve individuals from *Crangonyx islandicus* were used in this study, sampled at different locations throughout the volcanic zone in Iceland, see Kornobis, 2010.

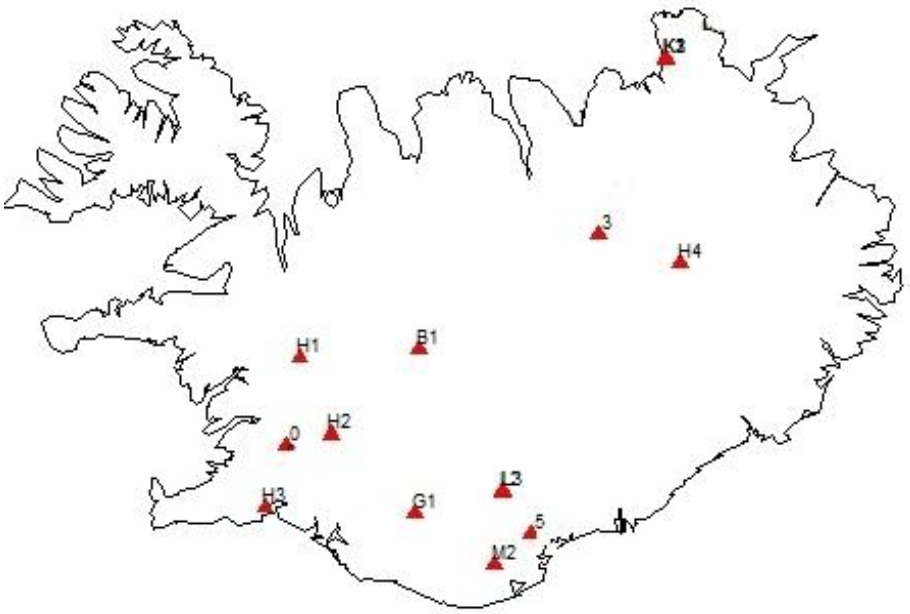


Figure 4. Samples sites of *C. islandicus* in Iceland: (H1)Húsafell, (H2) Miðhúsaskógur, (0) Þingvallavatn, (H3) Hraun Ölfusi, (B1) Svartárbotnar, (G1) Grasleysufjöll, (L3) Lakagígar, (5) Klausturlax, (M2) Hrífunes, Mýrdalssandur, (3) Svartárvatn, (H4) Herðubreiðarlindir, (K3) Klapparós Presthólar.

One individual from *Crymostygius thingvallensis* was used in this study, sampled at Þingvallavatn lake.

One individual from *Stygobromus stegerorum*, from USA, was also used in this study.

6.2. Laboratory protocols

6.2.1. DNA extraction

All samples used were already extracted, using chelex extraction, see Kornobis, 2010.

6.2.2. PCR: Polymerase Chain Reaction

Polymerase Chain Reaction is a technique to amplify a single or few copies of a piece of DNA across several orders of magnitude, generating thousands of copies of a particular DNA sequence.

PCR were performed in 10µl volumes containing 4.82µl distilled water, 0.75µl dNTP's (2mM), 1.15µl tween 20 (1%), 1µl taq buffer (10x), 0.5µl BSA (10mg/ml), 0.34µl of each primer (10pM) (forward and reverse), 0.1µl taq polymerase (NEB 5U/µl) and 1µl of template DNA (ca 10-200ng).

Another taq polymerase (dream taq from Fermentas) was also tested in attempting to have better PCR results.

PCR conditions for ITS fragment, using couples of primers described in table 3 involved initial denaturing of 90 seconds at 94°C, then 40 cycles of 20 seconds at 94°C, 40 seconds at different temperatures (see table 1), 1 minute at 72°C and 6 minutes of extension at 72°C.

Table 1. Annealing temperatures used is PCR program.

Primers	Annealing temperature	Time
ITSF – 28SR	56°C	40 sec
ITSF – 28SR	60ºC	40 sec
ITSF – 28SR	63ºC	40 sec
ITSF – ITSR	50ºC	40 sec

Electrophoresis in 1.5% agarose gel was used to observe the result from the PCR, and the samples were run for 20 minutes at 100mV. The PCR products were visualized taking picture of the agarose gel under an UV light. To get a better estimate of the length of the fragments, the concentration of the agarose gel was increased to 3% in some cases and the samples were run for 60 minutes.

6.2.3. Cloning

USER Friendly Cloning kit was used to clone the successfully amplified DNA.

Assembly reaction was made by mixing 0.5µl linearized pNEB206A, 0.5µlUSER enzyme and 5µl PCR sample. The solution was incubated for 15 minutes at 37°C and another 15 minutes at room temperature.

The plasmid was then transformed into bacteria. Our ligation product was mixed with 100µl *Escherichia coli* competent cells and incubated on ice for 30 minutes. Afterwards, the cells were heat shocked for 2 minutes at 37°C and chilled on ice for 5 minutes. One ml of LB media was then added and the mix incubated at 37°C for 1 hour before, it was spread (100µL) on LB + Amp plates medium.

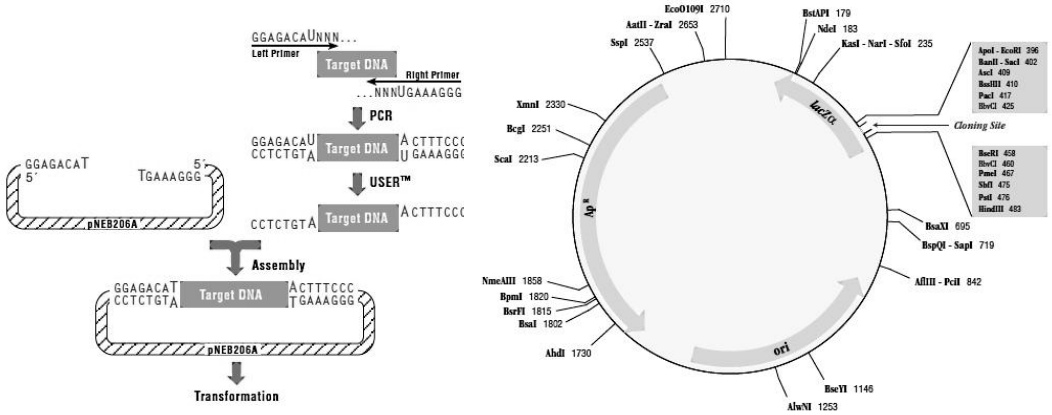


Figure 5. Map of pNEB206A vector.

Finally, the transfected cells were cultured. To recognize the cells that contained the desired insert at the appropriate orientation and isolate these from those not successfully transformed, IPTG and X-gal were added to the plates. The blue-white screening is a molecular technique that allows for the detection of successful ligations in vector-based gene cloning. The competent cells are grown in the presence of X-gal and IPTG, and culturing the cells overnight at 37°C, two different kinds of colonies should grown white ones which are the ones with the insert and the blue ones which haven't got the correct insert.

Three colonies of each individual were picked up and cultured overnight in liquid LB medium (10ml LB media + 10µl Ampicilin 100µg/ml) at 37°C and shaked at 200rpm.

Table 2. Components of LB medium for 1 liter of media.

LB medium
10 gr. Bacto Tryptone
5 gr. Bacto Yeast Extract
5 gr. NaCl
1 l. dH_2O
15 gr. Agar (1.5%Agar) for solid medium

The next step was to purify the overnight bacteria cultures, for what we used the NucleoSpin Plasmid kit. This method consists of washing the DNA with an ethanol solution on a silica membrane filter, and then eluted with water, solving the DNA.

The insertion of the DNA fragment was verified by electrophoresed on 1.5% agarose gel. Sometimes, depending on the length of the fragment another PCR was needed to know if the vector had the correct insert.

6.2.4. Sequencing

Big Dye, a ready reaction mix which contains all the components necessary for the sequencing reaction: Amplitaq DNA polymerase FS, dNTPs, ddNTPsDye terminators pyrophosphatase, Buffer and $MgCl_2$, was used for the sequencing reaction.

Cycle sequencing reaction mix contained 2.65µl ddH_2O, 1.75µl buffer, 1µl big dye, 1.6µl primer and 3µl PCR product. The samples were sequenced using the same forward and reverse primers used for PCR but at 10 folder lower concentration (1pM).

Table 3. Primers used in PCR and sequencing.

Primer	Direction	Sequence (5'–3')	Reference
ITS F	Forward	CACACCGCCCGTCGCTACTACCGAT	Chu et al. 2001
ITS R	Reverse	GCGGCAATGTGCATTCGACATGTGA	Chu et al. 2001
28SR	Reverse	ATCCCTGTTCGGTCGCCCCTACTAA	Hou et al. 2007
ITS int.	Forward	CCACCATACAACTCTGAGCGGTGG	This study
5.8R	Reverse	GACACCACGAGCAGAGATCA	This study
ITSF236cy	Forward	ATAGAGGGTGATGCCGACAG	This study
M13F	Reverse	CGCCAGGGTTTTCCCAGTCACGAC	
M13R	Forward	AGCGGATAACAATTTCACACAGGA	

Sequence reaction involved 10 seconds at 96°C, 25 cycles of 10 seconds at 96°C, 5 seconds at 50°C and 2 minutes at 60°C.

Figure 6. Location of the primers used in PCR and sequencing.

6.2.5. Ethanol precipitation

After sequencing, ethanol precipitation with glycogen used to purify or/and concentrate DNA from aqueous solutions.

The protocol:

- To each 5 - 10µl sequencing reaction add 50µl of a water NaOAc (0.3M) glycogen solution.

- Add 125µl cold (-20°C) 96% EtOH. Mix well by tipping back tubes up and down.
- Spin down at 4,000 rpm during 30 min at 0-4°C. Immediately pour off (most of) the precipitation mix.
- Wrap with 2 kim wipes, and spin down inverted at 300 rpm for 2 min at 0-4°C to remove the rest of the EtOH.
- Carefully, add 200µlof cold (-20°C) 70% EtOH.
- Spin at 4,000 rpm for 5 min at 0-4°C, immediately dump the EtOH and spin again inverted, covered with kin wipes, for 5 min at 300 rpm at 0-4°C.
- Let tubes air dry (in the dark) for 15 min to get rid of lost traces of EtOH (no pellet visible).
- Add 10µl HiDi, vortex and spin down.

The products were run on ABI Prism 3100 Genetic Analyser.

6.3. Sequence variation

Twenty-five different sequences belonging to twelve individuals from *C. islandicus* were obtained: eight sequences of 1800bp, six sequences of 1500bp and eleven sequences of 900bp in length.

The 1500 and 1800bp length sequences included part of the 18S rRNA, ITS1, 5.8S rRNA, ITS2 and part of the 28S rRNA; and the 900bp length included part of the 18S rRNA, ITS1 and part of the 5.8S rRNA.

One sequence from C. thingvallensis of 1800bp was obtained.

The following computer programs were used:

- Chromas Lite (Technelysium Pty Ltd), to correct the sequences by comparing them with the electropherogram.
- BioEdit for read the sequences and align them to see easily segregation sites.
- R (R Development Core Team, 2008) to create the geographical maps (map package).
- MEGA (Tamura et al., 1993-2008) for analyze the sequences obtained by calculating the p-distance.
- DNAsp (Rozas et al., 2009) for generate haplotype data files.
- Network software to construct networks for analyze the sequences obtained by constructing a phylogenetic tree.

Sequences were blasted on GenBank (http://www.ncbi.nlm.nih.gov/) in order to find similar sequences.

7. Results

7.1. *Crangonyx islandicus*

In total twenty-five sequences of *C. islandicus* were sequenced for the ITS gene (table 4).

Table 4. ITS1, 5.8S and ITS2 lengths in samples.

Sample	ITS1 (bp)	5.8S (bp)	ITS2 (bp)	Total length (bp)
H1	686 – 689	142	765	1800
H2	683	142	765	1800
0	689			900
H3	686			900
B1	683	142	765	1500
G1	689			900
L3	689	142	765	1800
5	689			900
M2	683	142	765	1800
3	384	143	763	1500
H4	384	142	765	1500
K3	384	142	763	1500

There are two different groups really well differentiated in length. The ITS1 region is from 683 to 689bp in nine individuals and 384bp in three of them, as is shown in table 4. In figure 7 is showed than the difference in length of 300bp is between the samples from the North (ITS1 of 384bp) and the samples from the South (ITS1 of 683 – 689bp).

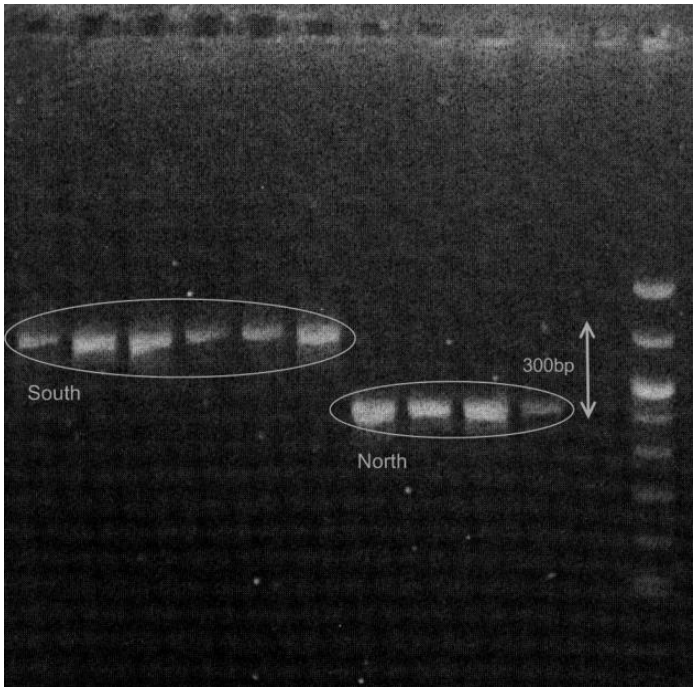


Figure 7. Agarose gel electrophoresis of DNA fragments generated by amplification of *C. islandicus* samples using ITSF/5.8R.

The 300bp insert is a duplication of a region in the ITS1 as is shown in figure 8. This region starts with a microsatellite part in 286bp, with 5 or 6 TGA repeats, and ends with a sequence AGGAAGGC in 561bp, which is not in the original sequence (see appendix).

Figure 8. Schematic figure of 18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2 and 28S rRNA gene with the microsatellites locations and the 300bp insert.

To estimate how this duplication occurred, p-distances were calculated. P-distance is the proportion of nucleotide sites at which two sequences being compared are different. Table 5 shows the p-distances within groups in diagonal, p-distance between groups above and the p-distance net between groups below.

Table 5. Matrix of p-distance within groups in the diagonal, between groups above and the net between groups below.

	1	2	3
1	**0.009**	0.010	0.022
2	0.001	**0.010**	0.023
3	0.015	0.015	**0.006**

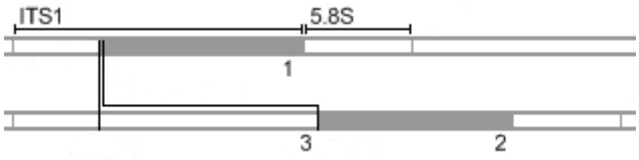


Figure 9 . Part of the sequence where number 1, 2 and 3 are shown the insert and the original sequences.

Comparing the p-distances, one can see that 1 and 2 are more similar than 2 and 3 or 1 and 3. Indicating, assuming neutrality that 3 diverged earlier in time from the others as it has accumulated more mutations.

Table 6. Variable sites in sequences. Position within sequence is given on top. *1= 300bp insert, *2 =TGA repeats, *3=CT repeats, and *4=AC repeats. Numbers in table indicate number of replicated repeat. Vertical lines mark the boundaries between 18S rRNA gene, ITS1, 5.8S rRNA gene and ITS2.

Sample	113	120	129	133	137	201	215	226	237	265	276	286*1	286*2	312	322	367	404	410	422	425	426	440	454	467	478
3.2c2	G	A	T	A	T	A	T	G	T	A	G	1	4	A	C	T	A	C	G	A	A	T	A	T	A
3.2c5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
K3.1c1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
K3.1c2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	.	.	.	.	.	.	.
K3.1c3	.	.	.	.	.	.	C	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.
H4.1c1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
M2.6c2	.	.	.	.	.	.	.	.	A	.	.	2	10	.	.	.	.	.	.	.	.	.	.	.	.
B1.1c3	.	.	.	G	.	.	.	.	.	.	.	2	9	.	.	.	G	.	.	.	.	G	.	.	.
H1.1c1	.	.	.	.	.	.	.	.	.	.	.	2	10	.	.	.	.	.	.	.	.	.	.	.	.
H1.1c2	.	.	.	.	.	.	.	A	.	.	.	2	11	.	.	.	.	.	.	.	.	.	C	.	.
H1.1c3	.	.	.	.	.	.	.	.	.	.	.	2	10	.	.	.	.	.	.	G	.	.	.	.	.
H2.1c1	.	T	.	.	.	.	.	.	.	.	.	2	9	.	.	.	.	.	.	.	.	.	.	.	.
L3.7c2	.	.	.	.	.	.	.	.	.	.	.	2	10	.	.	.	.	.	.	.	.	.	.	.	.
L3.7c3	.	.	.	.	.	.	.	.	.	G	.	2	10	.	.	.	.	.	.	.	.	.	.	.	.
0.2c1	.	.	.	.	.	.	.	.	.	.	.	2	11	.	.	.	.	.	.	.	.	.	.	.	.
0.2c2	.	.	.	.	.	.	.	.	.	.	.	2	10	.	.	.	.	.	.	.	.	.	.	.	.
G1.2c2	.	.	.	.	.	.	.	.	.	.	.	2	9	.	T	.	.	.	.	.	.	.	.	.	T
G1.2c3	.	.	.	.	.	.	.	.	.	.	.	2	9	.	.	.	.	.	.	.	.	.	.	.	.
5.4c1	.	.	.	.	.	.	.	.	.	.	.	2	10	.	.	.	.	.	.	.	.	.	.	.	.
5.4c2	.	.	.	.	.	G	.	.	.	.	.	2	10	.	.	.	.	.	.	.	.	.	.	.	.
5.4c4	A	.	.	.	C	.	.	.	.	.	A	2	10	.	.	.	.	.	.	.	G	.	.	.	.
5.4c5	.	.	C	.	.	.	.	.	.	.	.	2	10	.	.	.	.	.	.	.	.	.	.	.	.
H3.10c1	.	.	.	.	.	.	.	.	.	.	.	2	9	.	.	.	.	.	T	.	.	.	.	.	.
H3.10c2	.	.	.	.	.	.	.	.	.	.	.	2	9	.	.	.	.	.	.	.	.	.	.	.	.
H3.10c3	.	.	.	.	.	.	.	.	.	.	.	2	9	.	.	.	.	.	.	.	.	.	.	.	.

Sample	481	500	508	509	514	521	523	528	529	530	536	542	567	569	584	585	710	792	834	885	902	914	924	928	931	932
3.2c2	T	T	A	A	T	T	A	A	C	C	G	T	C	C	C	A	T	T	C	C	A	A	-	C	T	T
3.2c5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	-	.	C	.
K3.1c1	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	T	.	.	.	-	.	.	.
K3.1c2	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	T	.	.	.	-	.	.	.
K3.1c3	.	.	.	-	.	.	.	.	.	.	.	C	.	.	.	.	-	.	T	.	.	.	-	.	.	.
H4.1c1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	-	.	.	.
M2.6c2	.	.	.	.	.	G	C	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	-	.	.	.
B1.1c3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	-	.	.	.
H1.1c1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	G	.	-	.	.	.
H1.1c2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	-	.	.	C
H1.1c3	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	-	A	.	.	.	G	-	.	.	.
H2.1c1	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	-	.	.	.
L3.7c2	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	-	.	.	.	.	.	-	.	.	.
L3.7c3	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	-	.	.	.	.	.	-	.	.	.
0.2c1	.	.	.	.	.	.	.	C	G	.	.	.	.	.	.	.	-	.	.	.	.	.	-	.	.	.
0.2c2	.	.	.	.	.	.	.	C	G	.	.	.	.	.	.	.	-	.	.	.	.	.	-	.	.	.
G1.2c2	C	.	.	.	.	.	.	.	.	A	A	.	.	.	.	.	-	.	.	.	.	.	-	T	.	.
G1.2c3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	T	.	.	C	.	.	.
5.4c1	.	.	.	.	.	.	.	.	.	.	A	C	.	T	.											
5.4c2	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.											
5.4c4	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.											
5.4c5	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.											

H3.10c1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	T	A														
H3.10c2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.														
H3.10c3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.														

Sample	**933**	**941**	**944**	**1000**	**1060**	**1143**	**1155**	**1158**	**1168**	**1170*3**	**1189**	**1194**	**1234**	**1239**	**1265**	**1275**	**1293**	**1379**	**1404**	**1407*4**	**1431**	**1436**	**1474**
3.2c2	C	C	T	T	G	A	C	T	G	7	A	A	G	T	G	G	A	C	C	3	T	T	A
3.2c5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
K3.1c1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
K3.1c2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
K3.1c3	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	T	.	.	.	.	.	.
H4.1c1	.	.	.	.	.	.	.	.	.	8	.	.	.	.	.	.	T	.	.	.	.	.	.
M2.6c2	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	4	.	.	.
B1.1c3	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	T	4	.	.	G
H1.1c1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	4	.	.	.
H1.1c2	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	T	T	.	4	.	.	.
H1.1c3	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	4	.	.	.
H2.1c1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	4	.	.	.
L3.7c2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	4	.	.	.
L3.7c3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	4	.	.	.
0.2c1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	4	C	C	.
0.2c2	.	.	.	.	A	.	T	C	C	.	.	.	.	.	.	.	T	.	.	4	.	.	.
G1.2c2	T	T	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	.	4	.	.	.
G1.2c3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	T	.	.	4	.	.	.

Tables 6 shows all variables sites found in the sequences; from 5.8S rRNA gene samples 5.4 and H3.10 were not sequenced.

In the conservative parts there only few variable sites in comparison with the ITS parts (Table 6). There are 5 segregation points within 18S rRNA gene, 3 within 5.8S rRNA gene, and no one within 28S rRNA gene. Within ITS regions there are 34 segregation points in ITS 1 and 32 in ITS2. The 300bp insert was taken in account as one segregation site, and also the microsatellites regions.

Table 7. Variable sites in 300bp insert. *5=TGA repeats.

Sample	**1*5**	**4**	**5**	**55**	**56**	**68**	**76**	**84**	**95**	**100**	**109**	**112**	**116**	**119**	**213**	**222**	**257**	**278**
M2.6c2	6	T	C	C	A	T	T	A	G	T	T	T	C		G	T	A	A	G
B1.1c3	5	.	G	.	.	.	.	.	.	C	.	.	T		.	.	.	.	.
H1.1c1	5	.	G	.	.	.	.	.	.	.	.	.	.		.	.	.	.	A
H1.1c2	5	.	G	.	.	C	.	.	.	.	.	.	.		.	.	.	.	.
H1.1c3	5	.	G	.	.	.	.	.	.	.	.	.	.		.	G	.	.	.
H2.1c1	5	.	G	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.
L3.7c2	.	.	G	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.
L3.7c3	.	.	.		.	.	G	.	.	.	.	.	.		.	.	.	.	.
0.2c1	5	.	G	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.
0.2c2	.	.	G	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.
G1.2c2	.	.	G	.	.	.	.	.	T	.	.	.	.		.	.	.	.	.
G1.2c3	.	.	G	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.
5.4c1	.	.	G	.	.	.	.	.	.	.	.	.	.		.	.	.	.	.
5.4c2	.	.	G	.	G	.	.	.	.	.	.	.	.		.	.	.	.	.
5.4c4	.	.	G	T	.	.	.	.	.	.	C	.	.		.	G	.	.	.
5.4c5	.	.	G	.	.	.	.	.	.	.	.	G	.		A	.	.	.	.
H3.10c1	.	C	G	.	.	.	.	.	.	.	.	.	.		.	.	G	.	.

H3.10c2	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
H3.10c3	.	C	G	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.

In the 300bp insert there are 18 segregation sites, the first one being at different number of TGA repeats at the start of the duplication.

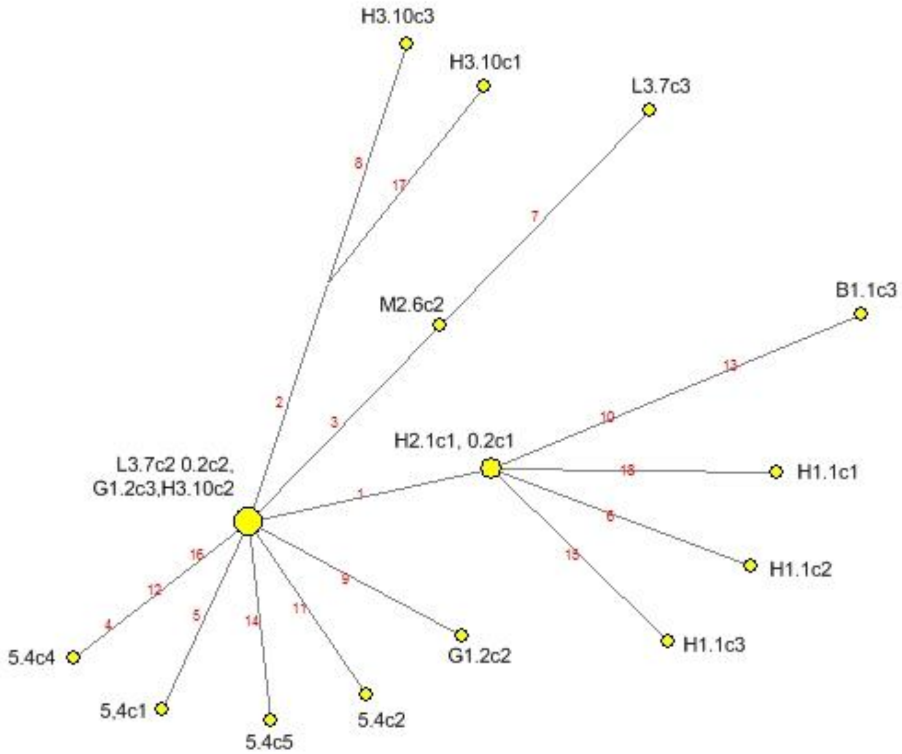


Figure 10. Network based on 300 bp insert within South population.

In figure 10 there are no well differentiated groups, but all 5.4 clones diverge from the same haplotype, and also H1.1 clones from other one. Variation in mutation number 1, referring to the number of TGA repeats, shows the difference between 5 and 6 repeats, as is shown in fig. 11 C.

Looking at the geographical places attending to these variations, we see different patterns depending on the variable sites (fig. 11).

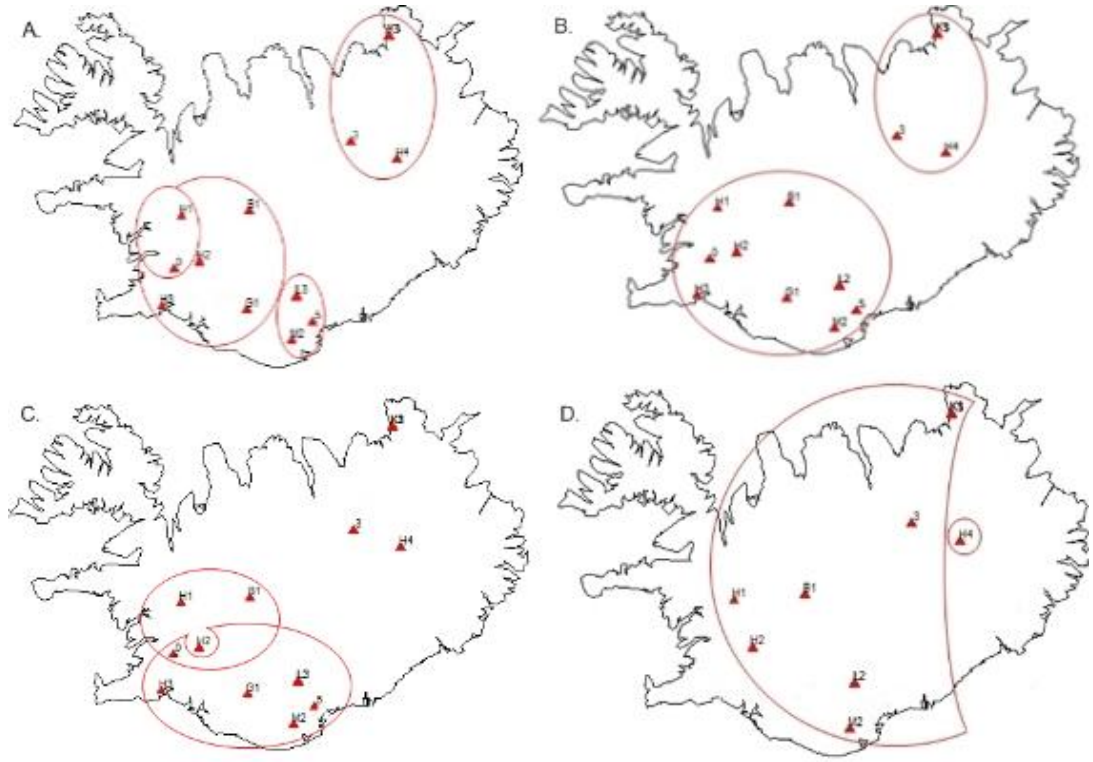


Figure 11. Geographical patterns according to the number of repeats: (A) TGA repetition in 286bp, (B) Insert of 300bp and AC repetition, (C) TGA repetition at the beginning of the 300bp insert, (D) CT repetition.

Network analysis

An allele network was constructed using Network with sequences of *C. islandicus'* individuals of the ITS1 (fig. 12), of the ITS2 (fig. 14) and all sequence (fig.15).

Considering the relationship in ITS1 (fig. 12), a high variation between individuals is seen, even between clones from the same individual, almost all present unique alleles; which can be explained by the relatively fast mutation rate of the ITS1.

The group in the North is really well differentiated from the one in the South due to a big insert of 300bp. Within the North group (A) two individuals share the same allele and one individual display three alleles. The variation is larger within the South group than within the North one: all the samples are singletons with the exception of one allele belonging to two individuals. We can divide the south group in three subgroups (B, C and D) representing three geographical areas. The maximum variation observed between two alleles within those subgroups is twelve mutations.

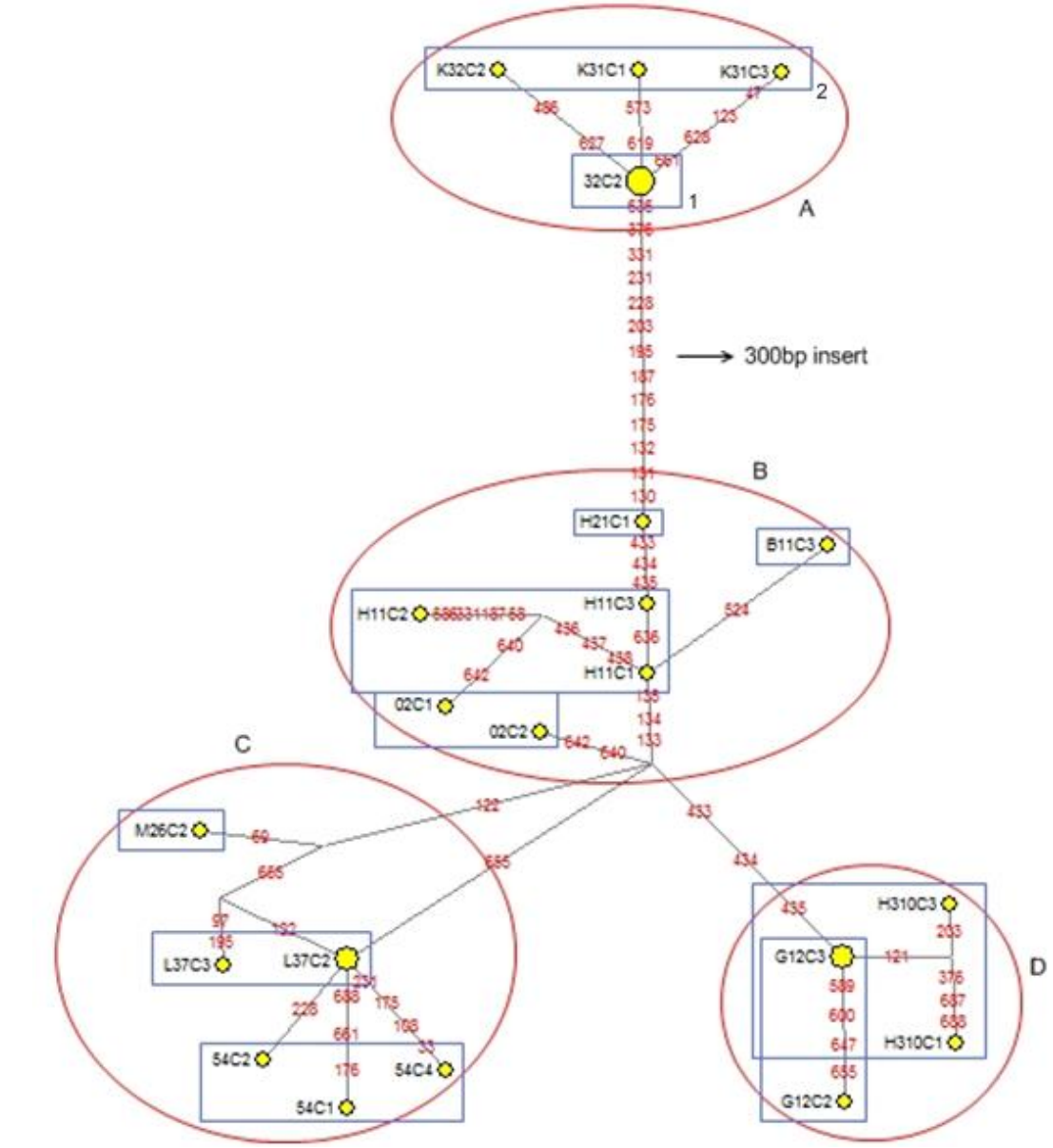


Figure 12. ITS1 allele's network. Geographical groups are delimited by red circles (groups A, B, C and D) and the clones for each individual are delimited by blue squares.

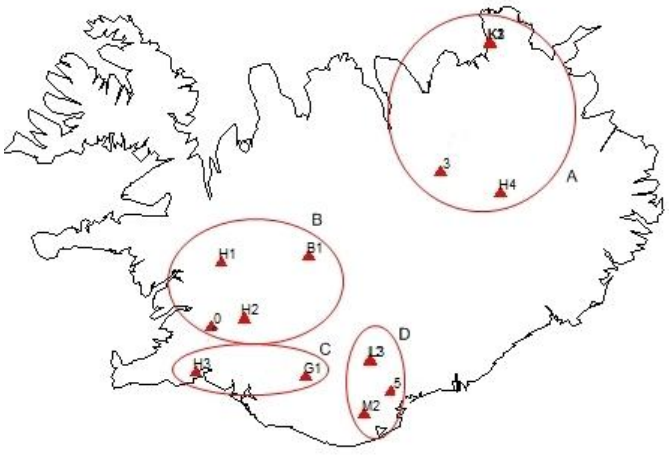


Figure 13. Geographic distribution of the alleles groups defined in Network in figure 10.

The ITS2 network shows unique alleles for each individual and even for each clone, with the exception of one allele which is shared for two clones of the same individual (K3.1).

Moreover, a difference between North and South groups is also represented. The variable sites correspond to one more microsatellite part, CT repeat, in the position 691 and 692 within ITS2, which is in South samples but not in North ones.

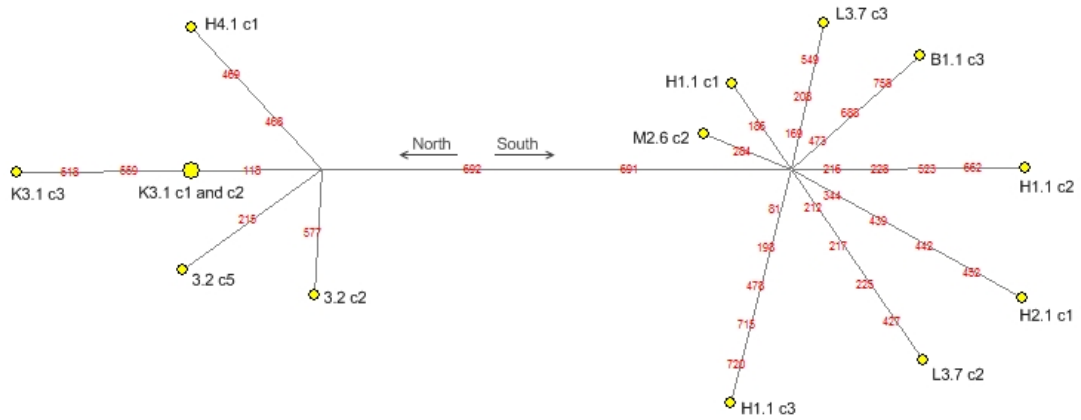


Figure 14. ITS2 allele's network.

- Complete sequence

Network of the complete sequence shows also the difference between North and South based on the 300bp insert. The variable site 286 *2 in table 6 was ignored because with it, the network was messy, possibly due to high mutation rate of number of TGA repeats at 286bp.

Within the North group, two subgroups were found, one included 3.2 and H4.1 samples and the other all clones of K3.1.

Within the South group, there is one haplotype belonging to three different samples and from it evolves the rest of the samples. There is one subgroup quite good differentiated, which groups L3.7, 5.4 and G1.2 samples.

It has to be mentioned that there are just two individuals which have identical clones, 3.2 and 0.2; the rest of the clones within each sample differ at one or more sites.

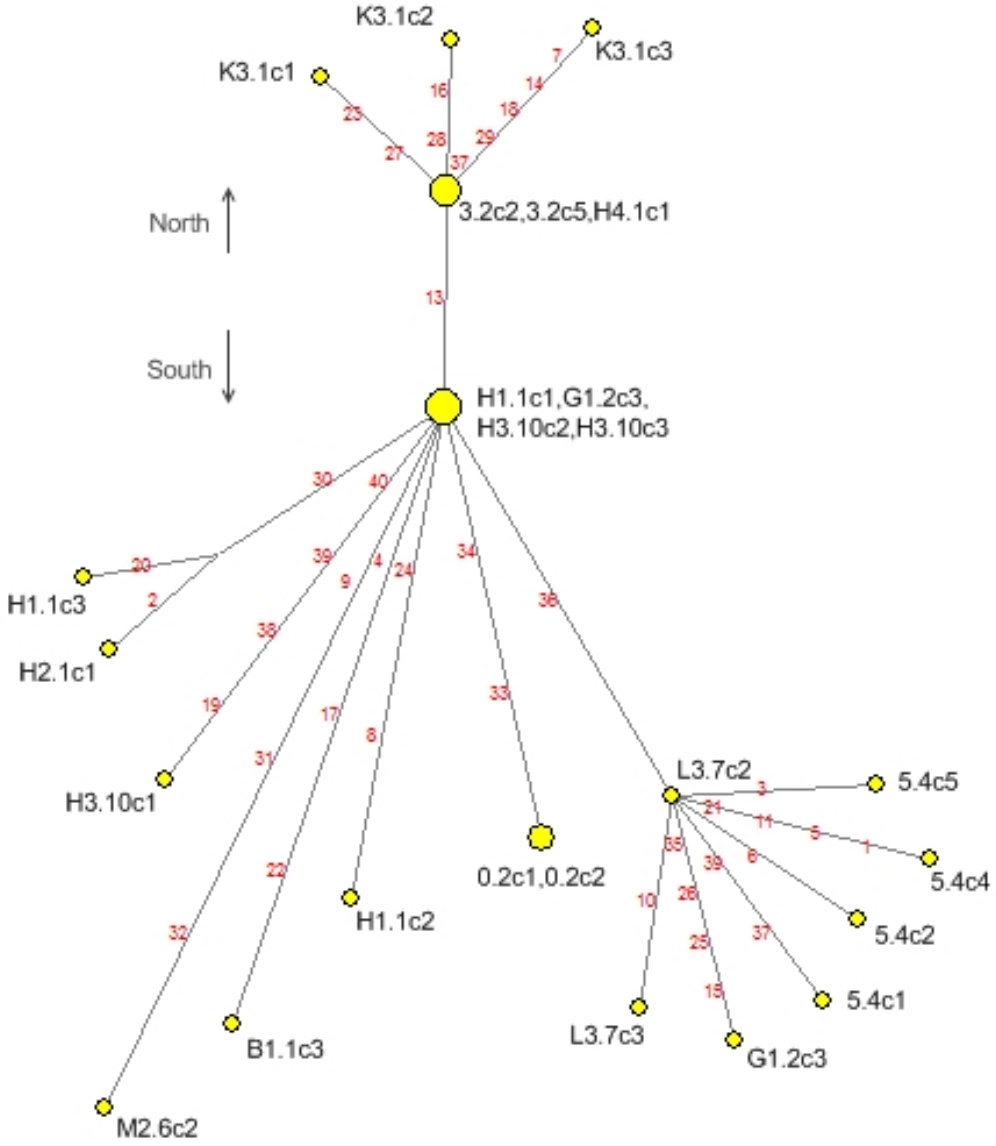


Figure 15 . Allele's network based on the complete sequence.

The results of blast conservative parts in GenBank, 18S rRNA gene, 5.8S rRNA gene and 28S rRNA gene were:

Table 8. Blast results in GenBank using 18S rRNA of *C. islandicus*.

Accession	Description	Taxonomic group	Query coverage	Max ident.
AY826961.1	*Antatelson walkeri*	Gammaridea	100%	99%
AF202978.1	*Bactrutusmucronatus*	Gammaridea	100%	99%
AY826971.1	*Andaniopsis nordlandica*	Gammaridea	100%	98%

Table 9. Blast results in GenBank using 5.8S rRNA of *C. islandicus*.

Accession	Description	Taxonomic group	Query coverage	Max ident.
EU807708.1	*Diporeia hoyi*	Gammaridea	83%	89%
DQ300260.1	*Gammarus tigrinus*	Gammaridea	52%	91%
AY004845.1	*Ampithoe longimana*	Gammaridea	86%	72%

Table 10. Blast results in GenBank using 28S rRNA of *C. islandicus*.

Accession	Description	Taxonomic group	Query coverage	Max ident.
EU80707.1	*Diporeia hoyi*	Gammaridea	100%	94%
AF482752.1	*Eubosmina longispina*	Crustacea	100%	88%
EU370435.1	*Triops cancriformis*	Gammaridea	100%	87%

7.2. *Crymostygius thingvallensis*

One fragment of 1800bp from *C. thingvallensis* was sequenced. The sequence contains part of the 18S rRNA gene, complete ITS1, 5.8S rRNA gene, ITS2 and part of the 28S rRNA gene.

Comparing *C. thingvallensis* sequence with *C. islandicus'* there is high differentiation in ITS1 and ITS2, but 18S rRNA gene, 5.8S rRNA gene and 28S rRNA gene are well conserved.

Table 11. Comparison in length between *C. thingvallensis* and *C. islandicus.*

Species	ITS1	5.8S	ITS2
C. thingvallensis	423 bp	141 bp	955 bp
C. islandicus	384 – 686 bp	142 bp	763 bp

Table 11 shows that 5.8S is a really well conserved region and ITS2 is the most variable region.

ITS regions have several tandem repeats (microsatellites):

In ITS1:

- At 247bp: 22 TC repeats.
- At 532bp, 3 CA repeats.

In ITS2:

- At 842bp, 3 TG repeats.
- At 933bp, 5 GGA repeats.
- At 952bp, 5 CTG repeats.
- At 1199bp, 3 TGG repeats.
- At 1378bp, 16 GGCA repeats.

The results of blast the conservatives parts in GenBank, 18S rRNA gene, 5.8S rRNA gene and 28S rRNA gene were:

Table 12. Blast results in GenBank using 18S rRNA gene.

Accession	Description	Taxonomic group	Query coverage	Max ident.
AY826961.1	*Antatelson walkeri*	Gammaridea	100%	97%
AF202978.1	*Bactrutus mucronatus*	Gammaridea	100%	97%
AY826971.1	*Andaniopsis nordlandica*	Gammaridea	100%	97%

Table 13. Blast results in GenBank using 5.8S rRNA gene.

Accession	Description	Taxonomic group	Query coverage	Max ident.
EU807708.1	*Diporeia hoyi*	Gammaridea	98%	84%
DQ300260.1	*Gammarus tigrinus*	Gammaridea	52%	90%

AY004845.1	*Amphitoe longimana*	Gammaridea	85%	74%

Table 14. Blast results in GenBank using 28S rRNA gene.

Accession	Description	Taxonomic group	Query coverage	Max ident.
EU807707.1	*Diporeia hoyi*	Gammaridea	99%	93%
AF482752.1	*Eubosmina longispina*	Crustacea	99%	89%
EU370435.1	*Triops cancriformis*	Gammaridea	100%	88%

All results from the blast have really high identity, which means the sequences are well conserved between species and even families.

7.3. *Stygobromus stegerorum*

One fragment of 500bp from *S. stegerorum* was sequenced.

Blasting the complete sequence the most similar results were:

Table 15. Blast results in GenBank using the complete sequence of *S. stegerorum*.

Accession	Description	Taxonomic group	Query coverage	Max ident.
EU807707.1	*Diporeia hoyi*	Crustacea, Malacostraca	22%	94%
AF482752.1	*Eubosmina longispina*	Crustacea, Branchiopoda	21%	88%
EU370435.1	*Triops cancriformis*	Crustacea, Branchiopoda	21%	87%

All matches correspond to the last part of the sequences, the 28S rRNA gene. There is no match in the rest of the sequence, neither in 18S rRNA gene or in 5.8S rRNA gene, which is supposed to be the most conserved parts. But actually if there's no sequence to compare with, we cannot know if the sequence starts in 18S rRNA gene or just in ITS2.

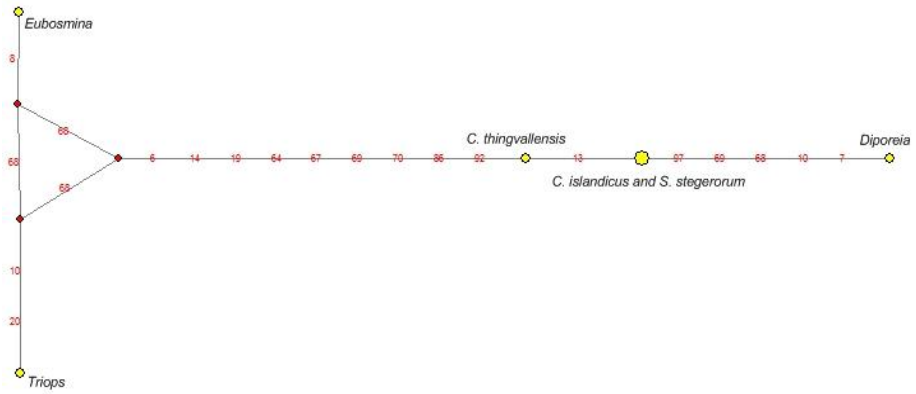


Figure 16. Network based on part of 28S rRNA gene with sequences from *C. islandicus*, *C. thingvallensis*, *S. stegerorum*, *T. cancriformis*, *E. longispina* and *D. hoyi*.

7.4. Unknown sequences

PCR amplifications using the ITS primers resulted in several cases in two bands of different length, the small one around 600 base pairs and the larger one about 1800 bases.

The species where this second band was found were in *Crangonyx islandicus* (Iceland), *Synurella ambulans* (Slovenia), *Stygobromus gracilipes* (USA), *Stygobromus stegerorum* (USA), *Crangonyx chlebnikov* (Russia) and *Synurella jakutana* (Russia).

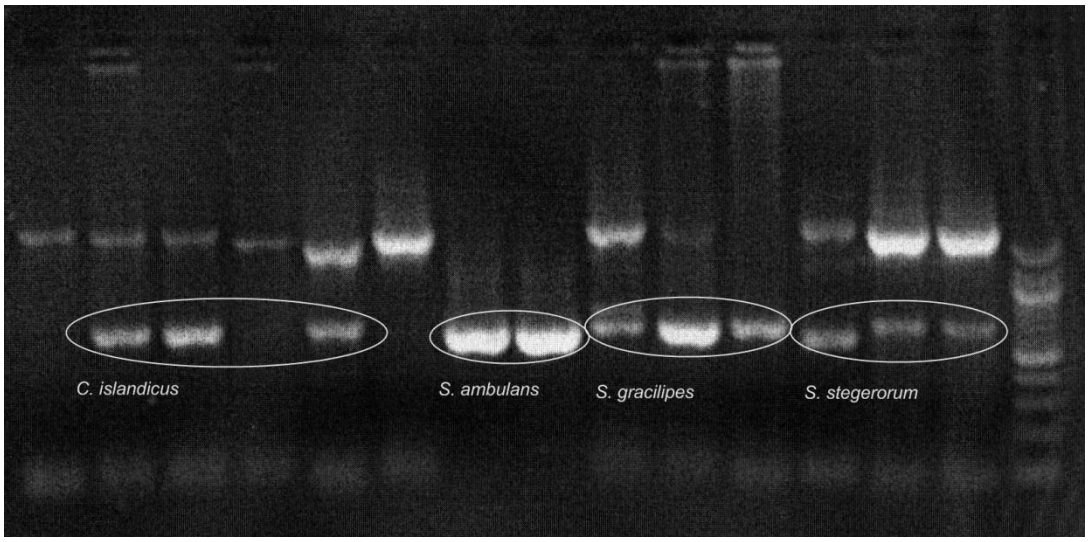


Figure 17. Agarose electrophoresis gel of DNA fragments generated by amplification of amphipod's samples using ITSF/28SR.

Two bands were cut from the gel and purified with the NucleoSpin Extract II kit, following the Protocol for DNA Extraction from agarose gels. Sequences of these fragments with the ITS primer where blasted on GenBank (www.ncbi.nlm.nih.gov).

The larger fragment gave sequences similar to amphipods as expected and the shorter fragment gave us sequences of unknown origin.

The amphipods from where we got these sequences were from *C. islandicus, C. chlebnikov, S. ambulans* and *S. jakutana*.

In total nineteen sequences of unknown origin were obtained from twelve individuals, defining four distinct types as shown in Network of fig. 18.

Table 16. List of countries and the amphipod's species where unknown sequences were found.

Sample	Country	Species (amphipod)	Seq. type
0.12	Iceland	*C. islandicus*	1
0.18	Iceland	*C. islandicus*	4
0.2	Iceland	*C. islandicus*	3
Vi3	Iceland	*C. islandicus*	1
7.6	Iceland	*C. islandicus*	3
H2.1	Iceland	*C. islandicus*	1
H3.1	Iceland	*C. islandicus*	1
S1.2	Iceland	*C. islandicius*	1
S.amb 2.1	Slovenia	*S. ambulans*	2
R6.1	Russia	*S. jakutana*	3
R9.1	Russia	*C. chlebnikovi*	3

The complete sequences are around 630 bp in length, including part of the 18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2 and part of the 28S rRNA. Looking at the length of the different parts we get:

Table 17. ITS1, 5.8S rRNA gene and ITS2 length in sequences.

Seq. type	ITS1	5.8S	ITS2
1	65 bp	152 bp	162 – 163 bp
2	65 bp	152 bp	161 bp
3	89 bp	147 bp	135 bp
4	76 bp	150 bp	158 bp

The ITS parts are completely different between each sequence type, and it was impossible to align them; nevertheless, 18S rRNA gene, 5.8S rRNA gene and 28S rRNA gene are more similar.

Blasting each sequence type in GenBank, the most similar results using the complete sequence are:

Table 18. Blast results in GenBank using the complete sequence type 1.

Accession	Description	Taxonomic group	Query coverage	Max ident.
AF087098.1	*Compsopogon coeruleus*	Rodophyta	68%	82%
EF158845.1	*Homalogastra setosa*	Ciliophora	68%	90%
GQ281532.1	Uncultured freshwater alveolate	Alveolata	68%	70%
AY102174.1	*Tokophyra quadripartita*	Ciliophora	23%	84%
AF043981.1	*Lepeophtheirus salmonis*	Arthropoda	32%	78%

Table 19. Blast results in GenBank using the complete sequence type 2.

Accession	Description	Taxonomic group	Query coverage	Max ident.
AF087098.1	*Compsopogon coeruleus*	Rodophyta	58%	87%
AF043980.1	*Lepeophteirus salmonis*	Arthropoda	73%	74%
GQ281532.1	Uncultured freshwater alveolate	Alveolata	39%	78%
DQ270014.1	*Cryptocaryon irritans*	Ciliophora	55%	81%
EU047813.1	*Chilodonella uncinata*	Ciliophora	37%	77%

Table 20. Blast results in GenBank using the complete sequence type 3.

Accession	Description	Taxonomic group	Query coverage	Max ident.
EU490182.1	Uncultured basidiomycete	Fungi	88%	70%
AF149979.1	*Paramecium tetraurelia*	Ciliophora	88%	70%
EF114296.1	*Urocentrum turbo*	Ciliophora	76%	72%
EU262622.1	*Pseudocohnilembus persalinus*	Ciliophora	53%	75%
FJ553738.1	*Oxytrichidae sp.*	Ciliophora	56%	94%

Table 21. Blast results in GenBank using the complete sequence type 4.

Accession	Description	Taxonomic group	Query coverage	Max ident.
FJ810408.1	*Carchesium polypinum*	Ciliophora	63%	71%
EU262622.1	*Pseudocohnilembus persalinus*	Ciliophora	33%	78%
EF114293.1	*Urocentrum turbo*	Ciliophora	64%	87%
EF174294.1	*Pseudokeronopsis carnea*	Ciliophora	29%	78%
AJ491193.1	*Begonia fenicis*	Viridiplantae	27%	79%

Afterwards, the sequence without ITS1 and ITS2 were blasted:

Table 22. Blast results in GenBank using sequence type 1 without ITS.

Accession	Description	Taxonomic group	Query coverage	Max ident.
AF087098.1	*Compsopogon coeruleus*	Rodophyta	38%	92%
AY102174.1	*Tokophrya quadripartita*	Ciliophora	36%	84%
AY969626.1	Uncultured ciliate	Ciliophora	26%	92%
FJ873805.1	*Chilodonella cyprini*	Ciliophora	26%	91%
DQ270016.1	*Ichthyophthirius multifiliis*	Ciliophora	37%	91%

Table 23. Blast results in GenBank using sequence type 2 without ITS.

Accession	Description	Taxonomic group	Query coverage	Max ident.
AF087098.1	*Compsopogon coeruleus*	Rodophyta	47%	93%
GQ281532.1	Uncultured freshwater alveolate	Alveolata	45%	96%
DQ270014.1	*Cryptocaryon irritans*	Ciliophora	75%	85%
AY513753.1	*Pseudocohnilembus hargisi*	Ciliophora	39%	84%
DQ241755.1	*Halteria grandinella*	Ciliophora	37%	84%

Table 24. Blast results in GenBank using sequence type 3 without ITS.

Accession	Description	Taxonomic group	Query coverage	Max ident.
AY969670.1	Uncultured alveolate	Alveolata	36%	86%
DQ777741.1	*Apokeronopsis bergeri*	Ciliophora	36%	86%
FJ553738.1	*Oxytrichidae sp.*	Ciliophora	63%	96%
AF508780.1	*Uroleptus pisces*	Ciliophora	79%	93%
AF508766.1	*Paraurostyla viridis*	Ciliophora	77%	94%

Table 25. Blast results in GenBank using sequence type 4 without ITS.

Accession	Description	Taxonomic group	Query coverage	Max ident.
EU047813.1	*Chilodonella uncinata*	Ciliophora	36%	81%
M98365.1	*Chaenea vorax*	Ciliophora	24%	91%
EF174297.1	*Pseudokeronopsis carnea*	Ciliophora	37%	81%
AY378113.1	*Hartmannula derouxi*	Ciliophora	37%	81%
GQ281532.1	Uncultured freshwater alveolate	Alveolata	47%	87%

Query coverage in almost every case is less, but the maximum identity is higher in all of them, which means the most conservative part is 18S rRNA gene, 5.8S rRNA gene and 28S rRNA gene, because the identity is higher when it's just blast these parts. Anyway, some matches are found when ITS' parts are also blasted, but with not really high degree of identity.

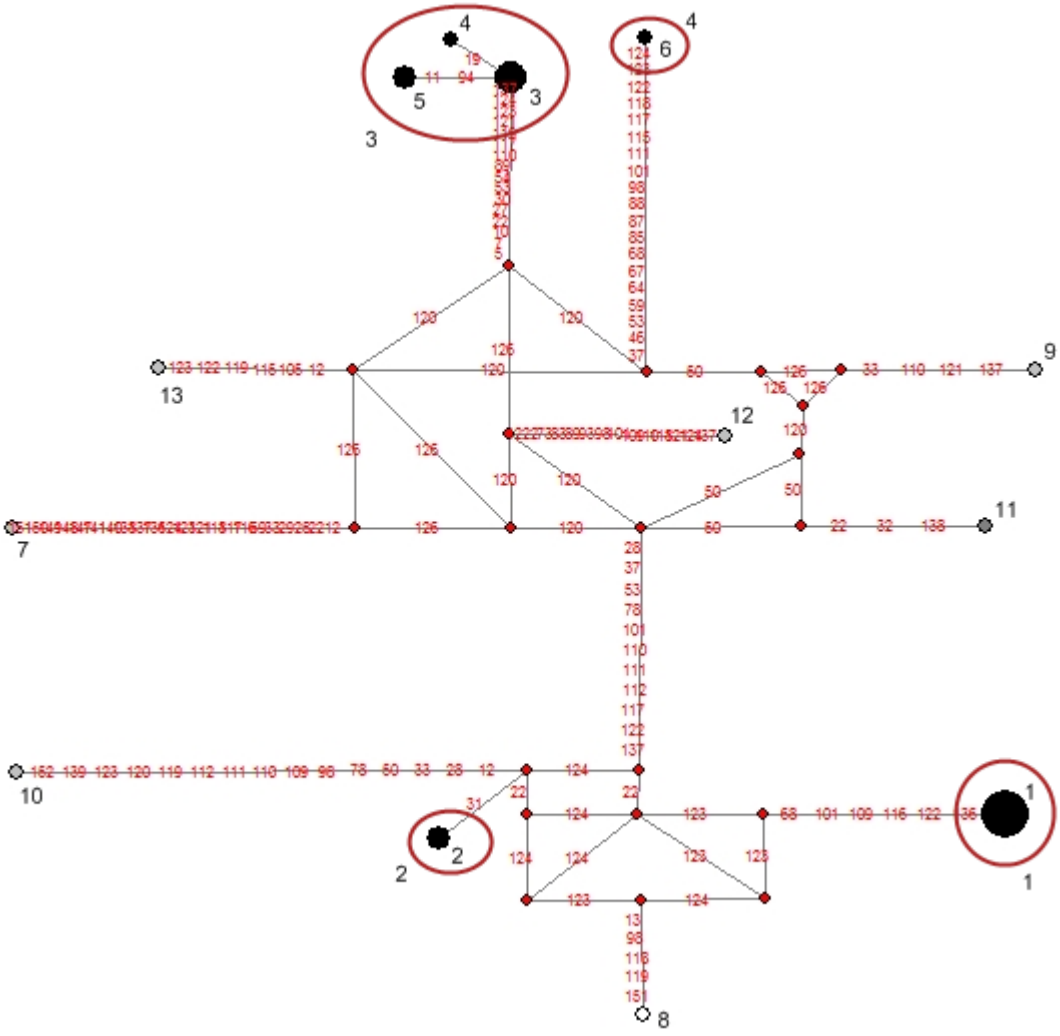


Figure 18. Network of phylogeny based on 5.8S rRNA gene sequence. Samples that I have sequenced are in black, ciliates samples from GenBank are in light grey, fungi sample is in dark grey and rodophyta sample is in white. Sequences types are delimited by red circles. Numbers in red denote the variable sites in sequences. Numbers next to the dots refer to sequences given in table 27. Number outside red circles refer to the unknown taxonomic group (table 26).

Table 26. List of haplotypes, their position in the tree and their corresponding taxonomic group.

Position in the tree	Samples	Taxonomic group
1	0.12c1,0.12c2,0.12c3,H2.1c1, H3.1c1, H3.1c2, S1.2, Vi3c3	Unknown, seq. type 1
2	S.amb 2.1 c1, S.amb 2.1 c2	Unknown, seq. type 2
3	0.2c1, 0.2c3, R91c1, R9.1c2	Unknown, seq. type 3
4	R6.1c1	Unknown, seq. type 3
5	R6.1c2, 7.6c1	Unknown, seq. type 3
6	0.18c1	Unknown, seq. type 4
7	*Carchesium polypinum*	Ciliophora
8	*Compsopogon coeruleus*	Rodophyta
9	*Homalogastra setosa*	Ciliophora
10	Uncultured freshwater alveolate	Alveolata
11	Uncultured basidiomycete	Fungi
12	*Paramecium tetraurelia*	Ciliophora
13	*Urocentrum turbo*	Ciliophora

In fig. 18, based just in 5.8S rRNA gene, the four sequence types are perfectly set; so it's easier to compare them with sequences from GenBank.

Sequences types 1 and 2 are more closely related to each other than with 3 and 4; moreover, both of them are more similar to Rodophyta.

Sequence type 3 and 4 are more closely related with Ciliophora than with other sequences having twenty-three and twenty-five mutation sites of difference respectively; but on the other hand, twenty-four and twenty-five mutation sites is the difference between them and Fungi.

8. Discussion

8.1. *Crangonyx islandicus*

Four different populations within *C. islandicus* were found based on ITS1. The most important difference is the 300bp insert which appears in all individuals from South Iceland.

The 300 bp insert is a duplication, which could have evolved as is shown in figure 19. The duplication starts with a microsatellite zone of TGAi in 286 bp and ends around 558 bp with GAAAGGAAGGC sequence in "X" (green arrow in fig. 19) which is not in the initial sequence. Could be that this part was in the initial sequence but it is still only in the duplication part.

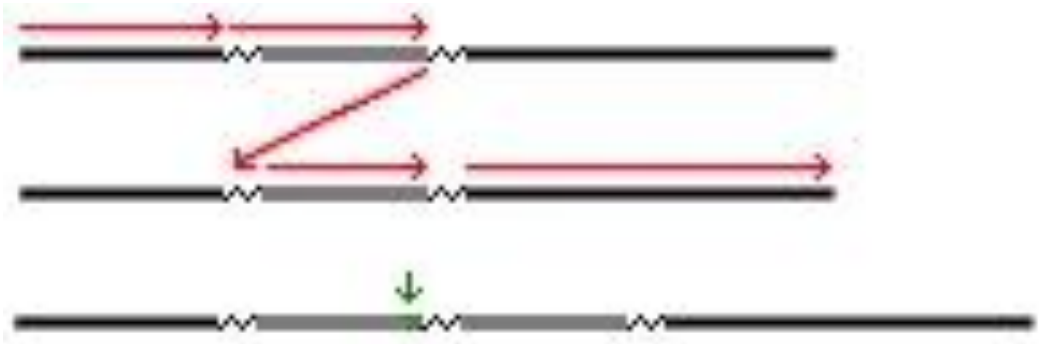


Figure 19. Schematic figure of how the duplication could have occurred.

P-distance (table 5) shows that 1 and 2 (in fig. 9), here referred to as N1 and S1, are more closely related to each other than to 3 (S2). Based on that, there are two possible hypotheses:

- A. This hypothesis is favored if neutral:

  A duplication occurs before the split of South and North, with ancestor of 1 separated from the ancestor of 2. Lineage 1 starts to diverge from lineage 2 due to mutations with time. After the split of North and South, N1 and S1 start to diverge from each other creating two different lineages, and continue to diverge from, lineage 2 in North and South. At some point, the North group lineage (N2) is lost. Finally, the result is two groups N1 in North and S1 and S2 in south.

- B. This hypothesis is favored if selection acted against changes in N1 and S1:

  Duplication occurs in South creating S1 and S2 after the split of North and South. S2 started to mutate faster than S1, which evolve at the same rate than N1.

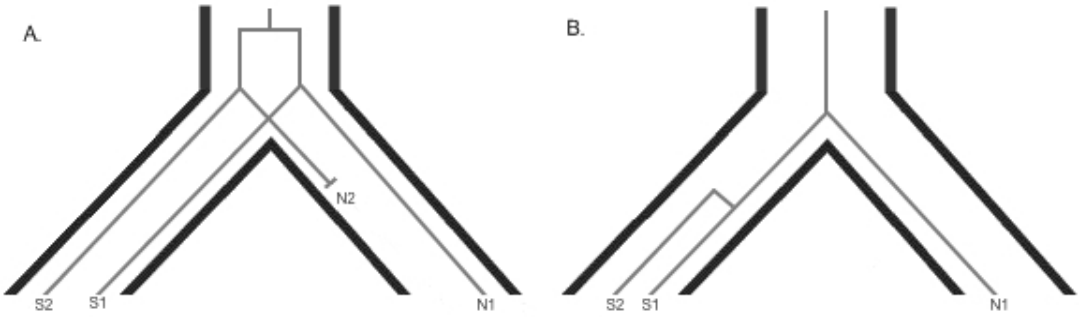


Figure 20. Hypothesis of how duplication appears in lineage. In A duplication occurs before the split of the two groups South and North, one lineage (S2) is lost. In B, duplication occurs within the South group.

This higher similarity of S1 and N1 could have a functional explanation. The fragment, although not transcribed, may have a functional role such as a binding site and be conserved by selection. The second copy may have relaxed constraints, allowing faster evolution.

There are two groups in the North, corresponding to two different geographical areas obtained also in Kornobis et al., 2010. These groups are E and F in fig. 21, and 1 and 2 in fig. 12, respectively.

There are three groups in the South, corresponding to three different geographical areas, but these groups are different than the groups obtained in Kornobis et al., 2010.

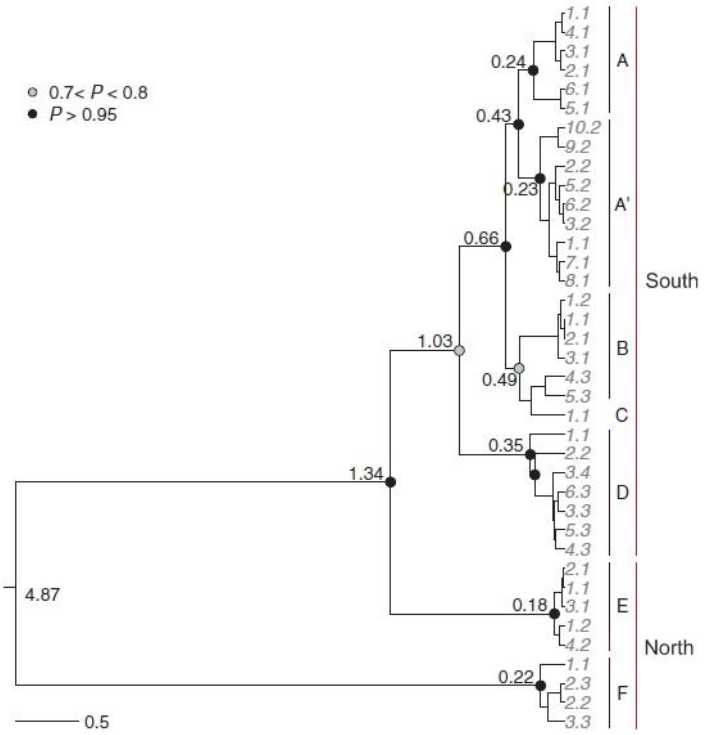


Figure 21 .Bayesian phylogenetic tree of *C. islandicus* populations in Iceland reconstructed from unique COI and 16S haplotype (Kornobis et al., 2010).

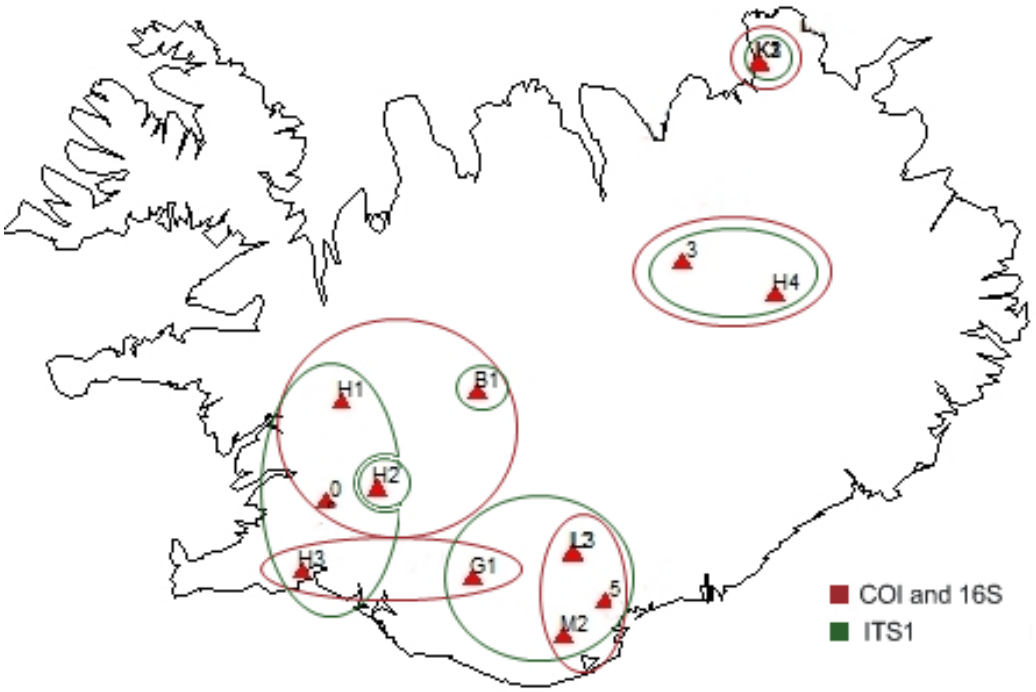


Figure 22. Geographical comparison of groups based on COI and 16S and ITS1.

This difference could be because of a different mutation rate between COI gene and ITS1. The first division was between the population from the North and the South and mitochondrial and nuclear genome show the same results because it happened a long time ago; but within South population the division time has been shorter and the results are different due to highly mutation rate in mitochondrial genome.

8.2. *Crymostygius thingvallensis*

The comparison between *C. islandicus* and *C. thingvallensis* shows that the functional parts of the sequence we get, 18S rRNA gene, 5.8S rRNA gene and 28S rRNA gene, are well conserved between species even between different families.

On the other hand, looking at ITS1 and ITS2, the mutation rate is really high and there are no matches between them. It could be because ITS is a piece of non-functional rRNA gene.

All microsatellites parts are in ITS regions and not in the conservative parts possibly also due to the functional/non-functional selection in the rRNA.

ITS can evolve and mutate faster than 18S rRNA gene, 5.8S rRNA gene and 28S rRNA gene, which have an important function and they have to be conservative parts.

This is supported with the results of blasting the conservative parts of the sequence, which show really high degree of identity.

Anyway, it could be a good idea try to get more sequences from *C. thingvallensis* and compare within them to know how fast evolve ITS between individuals and even between different clones of the same individual.

8.3. Unknown sequences

The unknown sequences which were found in PCR products in amphipod's samples were analyzed attempting to find out which species they were. After blasting the sequences, some matches were found, mostly ciliophora, but also rodophyta and fungi species.

The fact these sequences were found in samples of different species as *Crangonyx islandicus, Synurella ambulans, Stygobromus gracilipes, Stygobromus stegerorum, Crangonyx chlebnikov and Synurella jakutana* from different countries as Iceland Slovenia, USA and Russia led us to propose several hypothesis, which are not mutually exclusive:

- Some kind of contamination in the water or plastware.
- Microscope animals that were eaten by the amphipods and they were still on the throat.
- A parasite sticks in the amphipod when they were caught. It is known that some parasites can evolve together with the host over time; but this possibility is not really feasible because the different types of sequences are not differentiated between countries, and sequence type 3, for example, belong to samples from Russia and Iceland. However this could instance could be a contamination.

Trying to know better what it is, electronic microscope was used to observe the surface of the amphipods and try to find a match for sequences. Some unknown structures were found e.g. fig. 23.

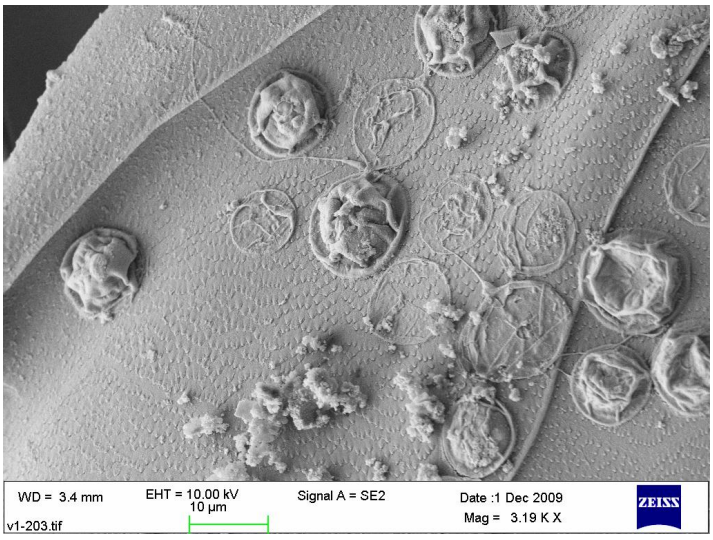


Figure 23. Unknown structures found in amphipod's surface.

One idea would be try to get more sequences of more samples from different species even countries and see if they evolve similar like the amphipod or if there is some relation between them constructing a phylogenetic tree.

EF alpha in Apatania zonella

9. Introduction

According to the results of my previous project (Sanz, 2010) "Mitochondrial variation of the caddisflies *Apatania zonella* and *Potamophylax cingulatus*", Iceland acts as a region, or a putative hybrid zone for *A. zonella*, where two distinct populations have arrived to Iceland from both ends of its range distribution; from Europe and from North America.

The previous study was realized with the mitochondrial cytochrome c oxidase subunit I (COI).As it, is only maternally inherited it does not give information on whether the two populations interbreed in Iceland; so this second part of the project is based on a nuclear marker in order to know if they are interbreeding or not.

The nuclear marker used is a fragment of the translation elongation factor 1α (EF-1α), which is a highly conserved core component of the translation machinery that is shared by all cellular life (Keeling and Inagaki, 2004). It is part of a large superfamily of GTPases that are involved in translation initiation, elongation and termination, as well as several other cellular functions. Because of its high level of conservation and seemingly ubiquitous distribution, eukaryotic EF-1α has been used to examine a variety of evolutionary questions and also served as an important molecular marker for determining phylogenetic relationships among eukaryotes (Keeling and Inagaki, 2004).

10. Material and methods

10.1. Samples

Seventeen specimens of *A. zonella* were studied in this project.

The samples were from different countries: six of them were form Iceland, one from Alaska, two from Norway and nine from Greenland.

Figure 24. Sample sites of *A.zonella* in Iceland: (A) Aðalvík, (B) Bjarnadalsá, (C) Staðará, (D) Ytri Rangá, (E) Hofsá, (F) Blautakvísl.

Figure 25. Sample site of *A. zonella* in Alaska: (A) Galbraith lake.

Figure 26. Samples site of *A. zonella* in Norway: (A) Dofrefjell.

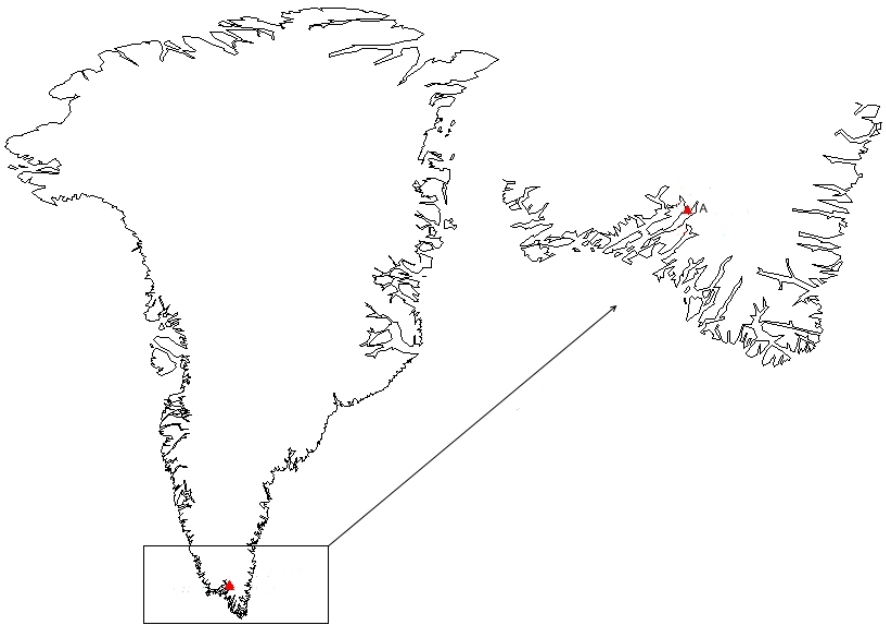


Figure 27. Sample site of *A. zonella* in Greenland: (A) Brattahlid.

10.2. Laboratory protocols

10.2.1. DNA extraction

The seventeen samples were already extracted, with the chelex method.

10.2.2. PCR: Polymerase Chain Reaction

PCR conditions for EF alpha fragment, using primes EF-1F and EF-1R (see table 27) involved initial denaturing of 5 minutes at 94°C, then 40 cycles of 20 seconds at 94°C, 45 seconds at 55°C, 1 minute at72°C and 6minutes of extension at 72°C.

Electrophoresis was used in order to evaluate the result from the PCR, exactly as was described in 6.2.2.

10.2.3. Exosap reaction

Successfully amplified DNA, was taken to an exosap reaction, which is used to remove unwanted dNTP's and primers, and also residual single-stranded primers and extraneous single-stranded DNA produced by the PCR.

Exosap mix containing 0,75µl ddH_2O, 0,7µl Antarctic phosphatase buffer, 0,5µl Antarctic phosphatase and 0,05µl ExoI, was added to 6µl of the PCR product.

The mix was warmed in PCR machine for 30 minutes at 38°C for treatment and then 15 minutes at 80°C to inactive enzymes.

The exosap reaction is used just for direct sequencing, and is not required for samples which are cloned as described in next section.

10.2.4. Cloning

TOPO TA Cloning kit for sequencing was used to clone four of the successfully amplified DNA samples.

Assembly reaction was made by mixing 1µl salt, 0,125µl TOPO vector, 1,875µl water and 3µl PCR sample. The samples were incubated for 20 minutes at room temperature.

The plasmid was then transformed into bacteria. The ligation product was mixed with 25µl of Chemically Competent *Escherichia coli* and incubated on ice for 30 minutes. Afterwards, the cells are heat shocked for 45 seconds at 42°C and immediately transfer the tubes to ice. LB medium (250µl) (table 2) was added and the cells incubated for 1 hour at 37°C before spreading the whole mix on LB + kanamycin plates medium. Finally, the transfected cells were cultured overnight at 37°C to obtain the colonies.

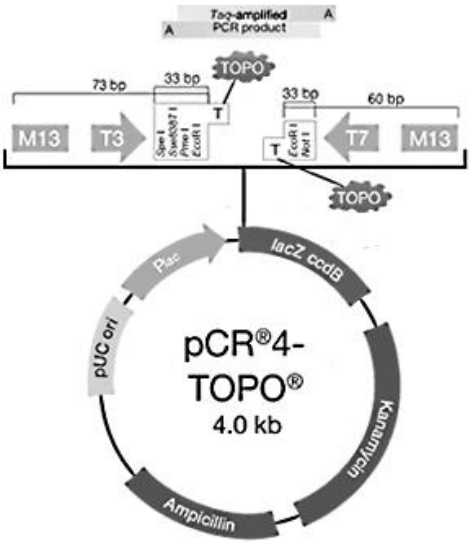


Figure 28. Map of PCR4 TOPO-vector

The next two steps of the cloning are exactly as the ones explained in 6.2.3.

To check if our fragment of interest is in the vector we use the restriction enzyme EcoRI. EcoRI mix contains 2µl buffer, 0,5µl EcoRI, 12,5µl ddH_2O and 5µl DNA. The restriction needs 1 hour at 37°C, the fragments were then electrophoresed on agarose gel. If two bands are visualized, the fragment is inserted in the vector but if we have just one band, the fragment was not inserted.

10.2.5. Sequencing

The sequencing reaction is the same that the one described in section 6.2.4.

Table 27. Primers used in PCR and sequencing.

Primer	Direction	Sequence (5'–3')
EF alpha 1F	Forward	ATCGAGAGGTTCGAGAARGARGC
EF alpha 1R	Reverse	CCAYCCCTTRAACCANGGCAT

10.2.6. Ethanol precipitation

The ethanol precipitation protocol is the same that I followed in 6.2.5.

10.3. Sequence variation

Seventeen individuals from different localities and with different haplotypes found in the first project were directly sequenced without cloning for the length of 400 base pairs.

Four individuals were selected for cloning and we obtained a total of ten more sequences, three clones from two individuals and two belonging to the other two (table 30).

Table 28. Samples places and COI haplotypes.

Sample	Place	COI haplotype
T13	Aðalvik, Iceland	European
T21	Hofsá, Iceland	European
T22	Ytri Rangá, Iceland	European
T48	Bjarnadalsá, Iceland	European
T25	Blautakvísl, Iceland	American
T47	Staðará, Iceland	American
AzA	Galbraith lake, Alaska	American
ApZo2	Dofrefield, Norway	European
ApZo3	Dofrefield, Norway	European
AzB1.1	Brattahlid, Greenland	American
AzB1.2	Brattahlid, Greenland	American
AzB1.5	Brattahlid, Greenland	American
AzB1.7	Brattahlid, Greenland	American
AzB1.8	Brattahlid, Greenland	American
AzB1.9	Brattahlid, Greenland	American
AzB1.10	Brattahlid, Greenland	American
AzB2.1	Brattahlid, Greenland	American

The computer programs used were the same as described in section 6.3.

11.Results

Seventeen sequences were obtained for the EF alpha gene with 400 base pairs in length with direct sequencing, without cloning. Four individuals were selected for cloning and we obtained a total of ten more sequences, three clones from two individuals and two belonging to the other two.

All sequences were aligned by eye and four variable sites were observed (table 29), where BioEdit put nucleotide N. By inspecting the electropherogram (figure 29) we could read the nucleotides of each allele.

Table 29. Variable sites in *A. zonella* samples, where R=G or A, Y=C or T and K=G or T.

Sample	89	233	257	267
T13	R	Y	R	T
T21	R	Y	R	T
T22	R	Y	R	T
T48	R	Y	R	T
T25	R	C	R	K
T47	R	Y	R	T
AzA	A	C	G	T
ApZo2	R	C	G	T
ApZo3	G	C	G	T
AzB1.1	A	C	G	T
AzB1.2	R	Y	R	T
AzB1.5	R	Y	R	T
AzB1.7	G	C	G	T
AzB1.8	R	Y	R	T
AzB1.9	R	C	R	T
AzB1.10	R	Y	R	T
AzB2.1	R	Y	R	T

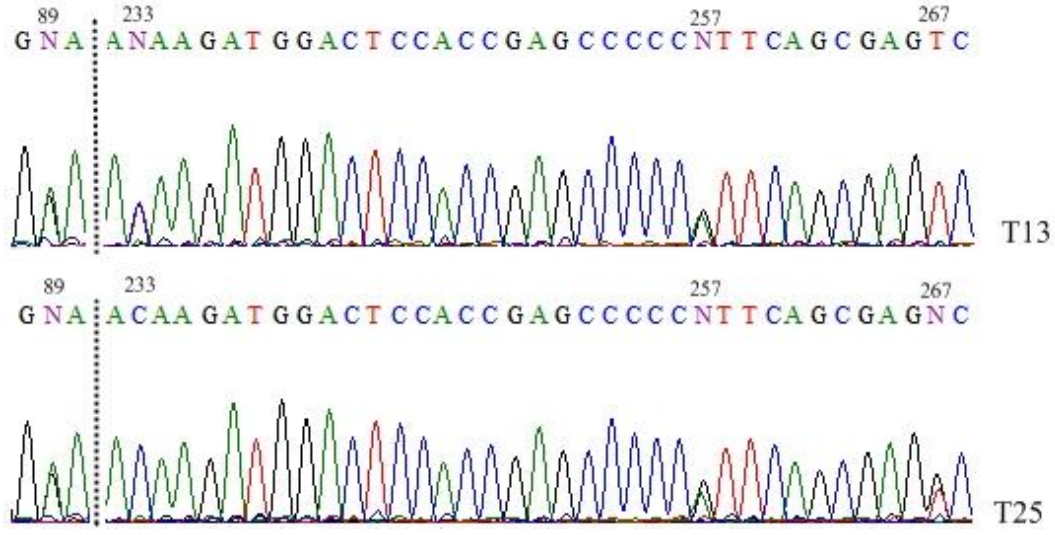


Figure 29. Electropherogram of variable sites.

In order to know the corresponding sequence for each allele, we cloned four samples, the most variable ones, before sequencing them (see table 30 and figure 30).

Table 30. Variable sites in *A. zonella* clones, where R=G or A, Y=C or T and K=G or T.

Sample	Clone	89	233	257	267
T13		R	Y	R	T
	c1	G	C	G	
	c2	A	T	A	
T25		R	C	R	K
	c1	A		A	G
	c2	G		G	T
AzB1.2		R	Y	R	T
	c1	A	T	A	
	c2	G	C	G	
AzB1.9		R	C	R	T
	c1	G		G	
	c2	A		A	

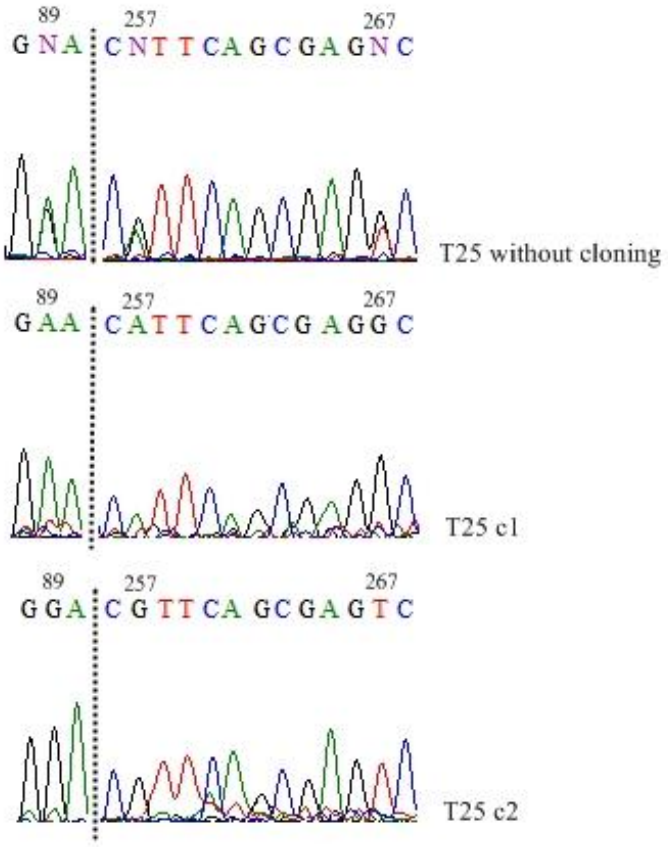


Figure 30. Electropherogram comparison between T25 without cloning and its clones (c1 and c2) in the variable sites.

Once we know the two different alleles of each individual that we have cloned, we know the haplotypes of the samples.

The five possible haplotypes we have got are:

1. ACGT
2. ATAT
3. GCGT
4. ACAT
5. ACAG

Looking at the network based on the different haplotypes of the individuals, we can compare them with respect to the countries of origin (fig.31) or to the COI haplotypes, obtained in the first project (fig. 32).

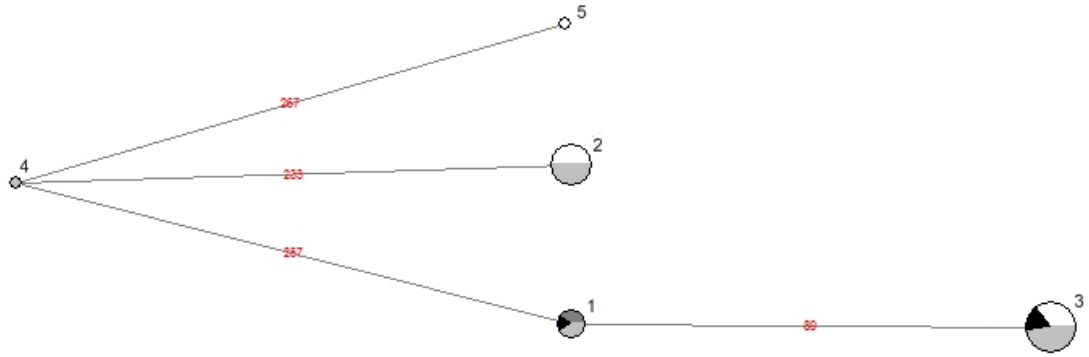


Figure 31. Tree of repartition of *A. zonella* with the software Network. Norway is in black, Iceland is in white, Alaska is in dark grey and Greenland is in light grey.

Haplotype 2 was just found in Greenland and Iceland; haplotype 1 is found in all studied countries except in Iceland and haplotype 3 in all except Alaska.

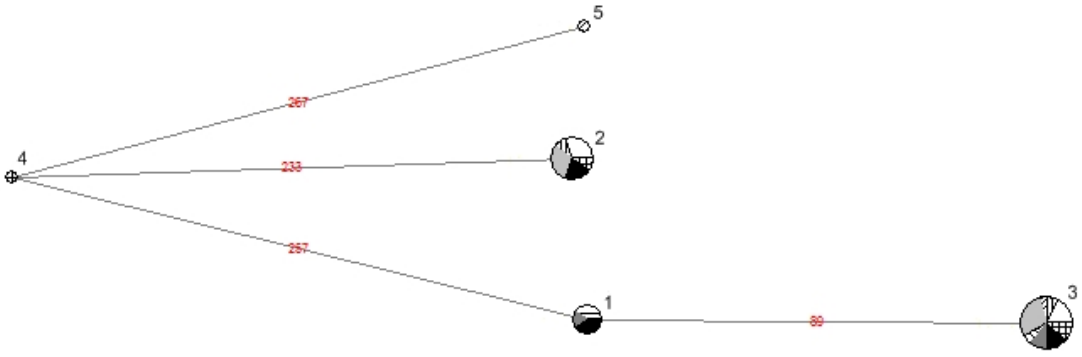


Figure 32. Tree of repartition of *A. zonella* with the software Network. Haplotypes based in COI are presented with different markings; A': white, B': vertical lines, C': backward diagonal lines, D': light grey, E': horizontal lines, F': dark grey, G': forward diagonal lines, H': black and NA: cross lines.

Table 31. List of samples with EF α genotype and COI haplotype.

Sample	EF α genotype	COI haplotype	Country
T13	23	A’	Iceland
T21	23	A’	Iceland
T22	23	B’	Iceland
T48	23	A’	Iceland
T25	35	C’	Iceland
T47	23	D’	Iceland
AzA	11	E’	Alaska
ApZo2	13	F’	Norway
ApZo3	33	G’	Norway
AzB1.1	11	H’	Greenland
AzB1.2	23	D’	Greenland
AzB1.5	23	H’	Greenland
AzB1.7	33	D’	Greenland
AzB1.8	23	D’	Greenland
AzB1.9	34	NA	Greenland
AzB1.10	23	NA	Greenland
AzB2.1	23	H’	Greenland

Within Icelandic population we have found just two different genotypes, both heterozygotes: 23 is found in five samples and 35 in one of them.

Otherwise, genotype 23 is the most common one, found in five of the specimens from Greenland.

Looking at the alleles, number 3 is the most common, being present in all samples except in two homozygotes samples, 11, the one from Alaska and one from Greenland.

Table 32. Frequencies of genotypes, with the COI type.

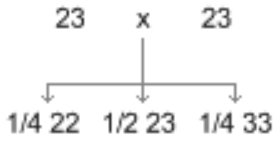
Genotype	Norway	Iceland	Greenland	Alaska	Total
11			1H’	1E’	2
12	1F’				1
23		3A’,1B’1D’	2D’, 2H’, 1NA		10
33	1G’		1D’		2
34			1NA		1
35		1C’			1
Total	2	6	8	1	17

12. Discussion

- **EF α genotype:**

The results show that genotype 23 is the most common one, belonging to almost every individual from Iceland and Greenland. Two individuals were found with genotype 33, but no one with 22.

Sexual reproduction should results in segregation of the 2 and 3 alleles, as when two heterozygotes 23 would mate:



but there were no individual found with genotype 22.

This means if they were sexual, all kind of genotypes has to been shown, because of the mixture of all possibilities. The probability to detect the different types depends though on the relative frequencies of the alleles, if an allele rare it mainly occurs in heterozygote state.

Under asexual reproduction, fewer genotypic combinations are expected because they don't interbreed and the inheritance is only maternal. Diversity arises from rare mutations but could also carry information about past sexual mixing, e.g. offspring from hybriditation of distinct lineages such as from different continents can be asexual. An example could American type A1/A2 mating with European type E1/E1 could give asexual A1/E1.

As is shown in the results, there are six different genotypes, just two of them homozygotes, 11 and 33 and four heterozygotes, 12, 23, 34 and 35. The proportion of haplotype 3 in the population is 16/34, which is almost half of the population; while haplotype 1 is just 5/34.

One hypothesis could be they are both, sexual or asexual, depending on the countries or environmental factors; because there are just 0.5 – 1% males, so maybe if females cannot find a male they are asexual but if they found one they are sexual.

- **EF α genotype and COI haplotype**

There is no relation between EF α genotype and COI haplotype.

A lot of variable sites and different haplotypes were found looking at the COI gene, but EF α gene is really well conserved, with only four featured variable sites.

One explanation could be the mutation rate is totally different between the genes; the mitochondrial genome, COI, evolve faster than the nuclear one, EF α.

Anyway, it would be necessary get more sequences to continue the study and take a final conclusion. Seventeen samples are not enough to see how the EF α evolve and to get a proper estimate of its allelic frequencies within the population.

13. Acknowledgments

I wanted to say thank you, firstly, to Snæbjörn Pálsson for accepting me in his group of work and give me the opportunity to start my scientific career in a laboratory with this project.

Of course, I want to thank very much Etienne Kornobis for all the things that you have taught to me about genetics, explain me every protocol to follow in the laboratory and also being so patient with me and all my computer problems.

Finally, also thanks to University of Salamanca and University of Iceland for their agreement, because of it I was able to get the Erasmus grant and work on this project in Iceland.

14. References

Avise J., 2007. Twenty-five key evolutionary insights from the phylogeographic revolution in population genetics. In: Phylogeography of Southern European refugia 7-21. Springer.

Bennet K. and Provan J., 2008. What do we mean by "refugia"? Quaternary Science Reviews 27: 2449-2455.

Chu et al., 2001. The first Internal Transcribed Spacer (ITS1) of ribosomal DNA as a molecular marker for phylogenetic and population analyses in Crustacea. Marine biotechnology, 3: 355-361.

Deffontaine V. et al., 2005. Beyond the Mediterranean peninsulas: evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). Molecular Ecology, 12, 1727-1739.

Geirsdóttir Á. Et al., 2007. Glaciation, erosion, and landscape evolution of Iceland. Journal of Geodynamics 43, 170-186.

Geirsdóttir et al., 2009. Holocene and latest Pleistocene climate and glacier fluctuations in Iceland. Quaternary Science Reviews 28: 2107-2118.

Han-Liang et al., 2006. Study of sequences of ribosomal DNA Internal Transcribed Spacers of clam belonging to the Veneridae family (Molusca: Bivalvia). Acta Genetica Sinica, 33: 702-710.

Hewitt G., 1999. Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society 68: 87-112.

Hewitt G., 2000. The genetic legacy of the Quaternary ice ages. Nature vol. 405.

Hewitt G. 2004. Genetic consequences of climatic oscillations in the Quaternary. The royal society. 359, 183-195.

Keeling P. and Inagaki Y., 2004. A class of eukaryotic GTPase with a punctuate distributrion suggesting multiple functional replacements of translation elongation factor 1α. PNAS vol.101 no.43.

Kornobis et al., 2010. Molecular evidence of the survival of subterranean amphipods (Arthopoda) during Ice Age underneath glaciers in Iceland. Molecular Ecology.

Pauls S. et al., 2009. Patterns of population structure in two closely related, partially sympatric caddisflies in Eastern Europe: historic introgression, limited dispersal and cryptic diversity. The North American Benthological Society vol. 28 : 517-536.

Pinceel J. et al., 2005. Rangwide phylogeography of a terrestrial slug in Europe: evidence for Alpine refugia and rapid colonization after the Pleistocene glaciations. Molecular Ecology, 14, 1133-1150.

Schmitt T., 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. Frontiers in zoology 4:11

Schönswetter P. et al., 2004. Molecular evidence for glacial refugia of mountains plants in the European Alps. Molecular Ecology, 14, 3547-3555.

Svavarsson J. and Kristjánsson B., 2006. *Crangonyx islandicus* sp. Nov., a subterranean freshwater amphipod (Crustacea, Amphipoda, Crangonyctidae) from springs in lava fields in Iceland. Zootaxa 1365: 1-17.

Väïnölä R. et al., 2008. Global diversity of amphipods (Amphipoda;Crustacea) in freshwater. Hydrobiologia, 595: 241-255.

15.Appendix

- **H1.1 c1 sequence:**

Part of 18S rRNA gene:
[CACACCGCCCGTCGCTACTACCGATTGGGTGTACCAGTGAGAGCCTTG
GACTGGCGTCCGCTGCCTTCGGGCAAGCGGGTCCGACGGAAAGAGGTC
CGAACTGGTGCACTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCCGTA
GGTGAACCTGCGGAAGGATCATTA]

ITS1:
[CCGGTTCTAAAAGCAATATGATACAGAGCTAAACGCTCGCACAGCTTC
TTGTTGTTGGGTCTCTTGATAGAGGGTGATGCCGACAGGGCGGCGGCAT
CAAAACCCCGAAAGGATGT(286bp)TGATGATGATGATGACGGCAAATGC
GCCGCTGGTGTTGCATTGTTACCTGGCATGGTGGGTGGTTCAGTTACTAC
CTACCAGCAGGTCAAAGGGACTGTGACCGATCGTGGCCGAGGAGAGAA
CCTACTCGGACGATGCTAAACAAACGAATCTCTGGCCCGATTATAAAAA
TGACTCACCCAAGCAACCGACCCTGGCGTGGAAGCAGTGACCGGTCAA
CATACCGGCACGGTTTCTCGCCTAAGTCTAACCACCATACAACTCTGAG
CGGTGAA(561bp)AGGAAGGC(570bp)TGATGATGATGATGATGATGATGA
TGATGACGGCAAATGCGCCGCTGGTGTTGCATTGTAACCTGGCATGGTG
GGTGGTTCAGTTACTACCTACCAGCAGGTGAAAGGGACTGTGACCGATC
GTGGCCGAGGAGAGAACCTACTCGGACGATGCTAAACAAACGAATCTC
TGGCCCAATTATAAAAATGACCTCCCCCAAGCAACCGACCCTGGCGTGG
AAGCAGTGACCGGTCAACATACCGGCACGGTTTCTCGCCTAAGTCTAAC
CACCATACAACTCTGAG]

5.8S rRNA gene:
[CGGTGGATCACTTGGCCTGTGGATGCTATGAAGACCATAGCTAAATGT
GAGAACGGCAGCGAGTCGTTGCTATATGCGCTGTCTCCCTTCTCTCATAT
GTCGAATGCACATTGCGCCCCACCAGACAGGTCGCTACATGGAA]

ITS2:
[GGATGAAATCTGTGAGGCTCGTGACAAACGCGATTGTGTGACCCCCGA
GTGACGACGGGGGACTCAGAAGCACATGAGTGTGCGTCTCGAGCCTGT
GTCATAGTAATCCGTTCCGACCGATAAGGTGGGGCACTTCTAGTCGAGT
GTGGTGGATCGAAGACCGGTGATCTCTGCTCGTGGTGTCCGCACGGGTT
CAGAAGTAAAGCCGTGCTCTTCTCTGTTTCTGTTGTGGCAATTAGGGTCA
ACATCAGTTGTGCCCGATCAAGCATCTGCCAGACTTGTTACCTCTGTGC
ATGGCCTCACATCCGCGTGACGTGCTGTGTGCATTGTAGGTCCATGTTG
GTAAGCACTATGAGCGGCCAGGGCACTGCAGCACACTTCACTGGGTGC
ACGAGTGCGCATCGGTCCAATATCGAAGGTGTATGAGGATATGCTTGCT
CTCTCGGTGTCGACTCTCTCTCTCTCTGTGACTCTAAAAACTGGCGTTGA
GCTGGGCGGCCCATAAAGGCAAACTGTCTGTCATATCCGGTCACTCGCG
ATTCGTGGTATTGGCAGGCAGAAAAATTGCGCGGGTCGTTCAAGCAGTG

TCTGGGTTCTGGCATGCCCAGCAGGATGGATTTCTGTTGATACGAGGCC
TCGTGAGAGGTCTCTGAGCACACAAGACTAGAGAGAGTCACCAGCCCA
CACACGTTGAAGACCCGATATTTGAGATTGCATAGGCAGGCACTCGATA
TATCCTCATGCCCTTTGTACTATGTCGCC]

Part of 28S rRNA gene:
[ACCACCTCGACTTAGAAGAGATTACCCGCTGAACTTAAGCATATCAGT
AAGCGGAGGAAAAGAAACCAATTCGGGATTCCCTTAGTAGGGGCGACC
GAACAGGGAT]