HÁSKÓLI ÍSLANDS

Phylogeography and phylogenetics of Icelandic groundwater amphipods based on the 16S rRNA gene.

Jaume Cuart Castell

Prof: Snaebjörn Pálsson

Faculty of Life and Environmental Science

University of Iceland, 2010

Phylogeography and phylogenetics of Icelandic groundwater amphipods based on the 16S rRNA gene.

Jaume Cuart Castell

Faculty of Live and Environmental Sciences

University of Iceland

101 Reykjavik, Iceland

Abstract

Icelandic groundwater amphipods survived the glaciations in refugia forming new endemic species such as *Crangonyx islandicus* and *Crymostygius thingvallensis*. However a strong evolutionary pressure has led to a morphological convergence that makes difficult a clear taxonomy classification within this group. The aim of this study is twfold, firstly to assess the position of these species within the superfamily Crangonyctoidae,and secondly to assess and compare the phylogeography of of *C. islandicus* using different fragment sizes (420 and 1200 base paires) of the 16S rRNA gene. A taxonomy of the family Crangonyctoidea using the 16S gene differs from the one based on morphology. A higher variation rate was found along the 1200bps fragment due to a faster evolutionary rate, than in the smaller region.

* E-mail: jaumecuart@hotmail.com

Index

Index of figures

Index of tables

Introduction

Phylogenetic and phylogeographic studies have been increasing in the genetics literature (e.g. Avise 1998; Masta 2000) aiming to provide a insight in the relationship between geological history and biodiversity (Avise et al 2000; Arborgast et al 2001). Combining phylogenetics and biogeography has allowed the examination of the evolutionary processes that shape and maintain genetic diversity (Bermingham and Moritz, 1998). Until recently taxonomy and biodiversity studies were based only on morphological characters, which were used to define ancestor and descendant relationships and to provide a hierarchical foundation for species taxonomy. However morphology provides limited phylogenetically informative characters, particularly among organisms subjected to intensive selection such as in cases of parallel and convergent evolution (Hou et al. 2007) resulting in an under- or over-estimation of the real diversity. Thus genetic studies are being used increasingly to assess phylogeny and geographical distribution between taxa (Lefébeur et al. 2006). Two main trends are observed in this rapidly growing field. At one site DNA barcoding is used to identify and reassign species to a more suitable taxonomic groups using a previously designed set of synapomorphies as markers for each group (Solvalainen et al. 2005) based on the measure of distances between DNA sequences (Hebert et al. 2003). However a single molecular study may not provide enough information to diagnose species differentiation. In the second trend DNA taxonomy is used to classify new species to its corresponding taxa according to a particular divergence in a known gene (Helbert et al. 2003).

Fauna distribution seems to be influenced by significant climate changes which has occurred during millions of years. Species responses to climate change may vary for organisms, depending on their level of cold-tolerance and dispersal capabilities (Schönswetter et al., 2004; Deffontaine et al., 2005; Pinceel et al., 2005). In the boreal regions, the environmental conditions are often very severe, making survival and reproduction more difficult. Fluctuations of population size are common in many boreal species that can have regular cycles or irregular population outbreaks (Sanz 2010). During last period of the Pleistocene most flora and fauna of north hemisphere retreated to lowers latitudes due to glaciations while some species managed to survive in refugia and became genetically insolated and even evolving in new taxa (Vasconceros et al. 2006). Ice age started 2.4 Myr ago and has been characterized by fluctuations in temperature which have resulted in several long glacial periods , separated by short warm interglacial periods .The last glaciations ended about 10000 years ago (e.g.Hewitt et al. 2000).

Groundwaters may have acted as a refugia for many fresh water invertebrates. The groundwaters formed a subterranean habitat with extreme conditions where only highly adapted animals survived. Isolation among refugial populations is expected to promote genetic and phenotypic differentiation as a result of independent adaptation to the local

environments and genetic drift, with consequences for reproductive isolation between discrete refugial lineages and the creation of hybrid contact (Hewitt, 1999). A high diversity at regional scale has been found in those ambient due to:

- Lack of strong competitors and predators.
- Independent evolution.
- Favorable and stable thermal environment conditions.
- Dynamic paleogeography that may have caused numerous changes in the groundwaters distribution.

Groundwaters as subterranean ecosystems are considered to be natural laboratories to study the effect of temporal and spatial isolation on genetic divergence (Culver et al. 1995). Environmental stability, permanent darkness and oligothropy are characteristics of these ecosystems.

Iceland has undergone a number of geological (volcanism, mountain building) and climatic (glaciations, marine inundation) events that has produced a significant effect on evolutionary processes (Chinn and Gemmell, 2004). Furthermore, Iceland has been isolated from other landmasses for 25-40 Myr and during this long period of isolation, an extensive endemic fauna and flora could have survived insulated from other populations. However, endemic species are hardly found in Iceland and species diversity is low compared to other areas at the same latitude. This has been explained by repeated glaciations, which have covered the whole island repeatedly during Ice age (Geirsdóttir et al. 2007), and the difficulties for species to colonize Iceland due to its geographic isolation (Buckland et al. 1986). The glaciations have also affected variation within species, species at high latitudes generally show little variation and shallow genealogies, and it is important to have in mind when analysing genetic variation that population genetic processes require long times to reach equilibrium (Pamilo and Savolainen, 1999). Despite unique nature of the fauna and the natural processes that have shaped it, only limited attention has been directed to the freshwater invertebrate fauna.

Fresh-water crustaceans provide a good example of biodiversity divergence after a glacial period and of biodiversity under-estimation due to a limited morphological variation (Taylor et al., 1998; Witt and Hebert, 2000). This lack of variation is specially found among ground water crustaceans adapted to extreme conditions where these conditions seem responsible of a strong morphological convergence making difficult to get a proper taxonomic distribution (Lefébure et al. 2006). For this reason both DNA barcoding and DNA taxonomy could be of high value in studies of crustacean's biogeography.

Amphipods are crustaceans of the order Malacostrata whose live cycle is characterized by direct development with no independent larval stage, small geographical ranges and high levels of endemism (Porter et al. 2007). Amphipods live in all sorts of habitats

even though most of them are marine and only around 20% of the species are fresh water. The adaption of amphipods to hypogean conditions has involved a series of physiological, morphological and behavioral changes, collectively known as troglomorphism (Oromí et al. 1991; Caccone et al. 2001). These adaptations are common to different taxa, suggesting an evolutionary convergence caused by living under similar selective conditions (Porter et al. 2003). Some of the most characteristic homoplasies or vestigial characters are the eye reduction and the loss of pigmentation (Figure 1 and 2) caused by disuse (Darwin et al. 1860).

The superfamily Crangonyctoidea is a widespread groundwater group of amphipods containing most of the North American groundwater amphipods but also some genera present in both Asia and Europe like the genera Stygobromus and Crangonyx. Recently two new species have been discovered in Iceland belonging to this family: *Crymostygius thingvallensis* (Figure 1) and *Crangonyx islandicus* (Figure 2). *C.islandicus* is an endemic specie from the family Crangonyctidae (Svavarsson and Kristjánsson, 2006) and is widely distributed along the volcanic zone in Iceland and has been sampled in springs flowing from lava fields in the geologically youngest areas. *C. thingvallensis* represents a new endemic family of amphipods, the Crysmostygiidae (Kristjánsson and Svavarsson, 2004) and only five individuals have been found in a well at the lake Thingvallatn so far. Many hypothesis have been given to explain the presence of these endemic species in Iceland (Svavarsson and Kristjánsson, 2006), one plausible explanation is that there were amphipods in Iceland before the glaciations which survived in subglacial refugia in Iceland. Geothermal heat might have maintained the groundwater habitat in fissures at the tectonic boundary and in the porous lava bedrock that was constantly built throughout the glacial and interglacial periods (Kristjánsson and Svavarsson, 2007).

Figure.1. Picture of *Crymostygius thingvallensis*.

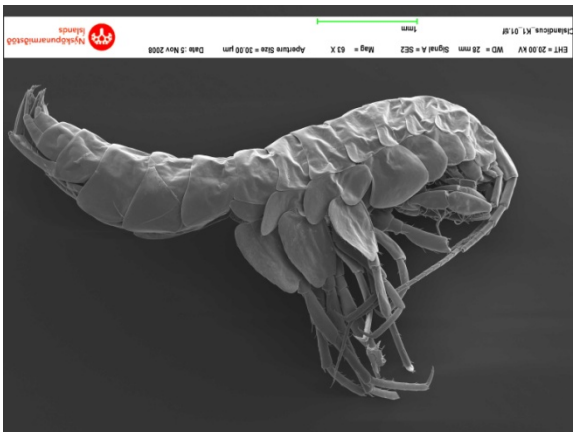

Figure.2. Picture of *Crangonyx islandicus*.

In order to assess phylogeographical and phylogenetic variation mitochondrial genes such as the large subunit 16S ribosomal RNA (figure 3) has been a popular marker. In crustaceans it has been used both at the species and populations levels (Baldwin et al 1998; Tam and Kornfiels et al 1998; Sarver et al 1998; Schubart et al 2000). A 420bps fragment of this gene has been used worldwide for these kinds of studies, and was used by Kornobis et al. (2010) in a study of the geographical variation within *C. islandicus* were six different populations were found according to the mitochondrial variation in the COI gene and the 420bps fragment of the 16S rRNA gene. As small gene fragment may lead to a poor estimation of the genetic variation, a larger region of the 16S rRNA fragment (1200bps) is now more commonly used.

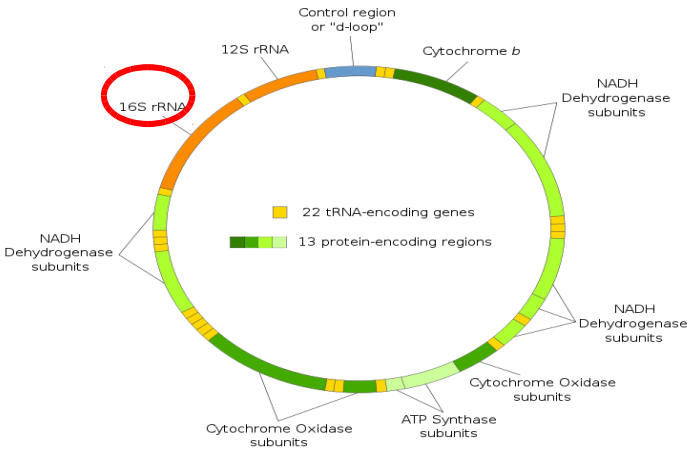

Figure 3. Diagram of the Mitochondrial genome of *C. islandicus*.

Mitochondrial rRNA is the central component of the ribosomes and its function is to decode the mRNA into amino acids and to interact with tRNA during translation by

providing peptidyl transferase activity. Some of the characteristics that make 16S gene useful in phylogenetic studies are:

- The gene is present in almost all organisms.
- Possesses highly conserved sites that may be target of universal PCR primers.
- The gene displays an appropriate level of sequence conservation for the divergences of interest. If there are too many substitutions, then the evolutionary signal becomes saturated, and there is a limit to the depth of the divergences that can be accurately inferred. If there are too little changes (if the gene is too conserved), then there may be little or no change between the evolutionary branching of interest, and it will not be possible to infer close (within species or within genus) relationships.
- The gene is sufficiently large to contain a record of the historical information.
- No recombination. The mtDNA is inherited intact from the mother to its offsprings.

The aim of this study is twofold:

- To estimate the phylogenetic relationship of the Crangonyctidae and related species using a fragment of the 16S rRNA mtDNA gene.
- To evaluate the phylogeographic structure within *Crangonyx islandicus* based on 1200 bases of the 16S rRNA mtDNA gene and to caompare it with the structure obtained by a shorter fragment.

- In addition variation along the 1200 bps fragment is studied and compared to the shorter region of 420bps.

Methods

Sample

One hundred and three *Crangonyx islandicus* specimens were collected and analyzed genetically by Kornobis et al (2010) from 23 sample locations and one specimen of *Crymostygius thingvallensis* (Figure 4) along the geologically most recent volcanic zone of Iceland. The samples were collected in dip-nets after applying electricity and stored in 96% ethanol (Kornobis et al 2010).

Figure.4. Map of Iceland and the locations were *C. isalandicus* (black spots) and *C. thingvallensis* (red spot) were sampled by Kornobis et al. (2010).

In addition to the Icelandic samples 8 species (Table 3 and 4) were used to assess phylogenetic relationship between the Icelandic species with other amphipods. Four out of the eight DNA sequences were obtained from gene bank (Table 3 and 4).

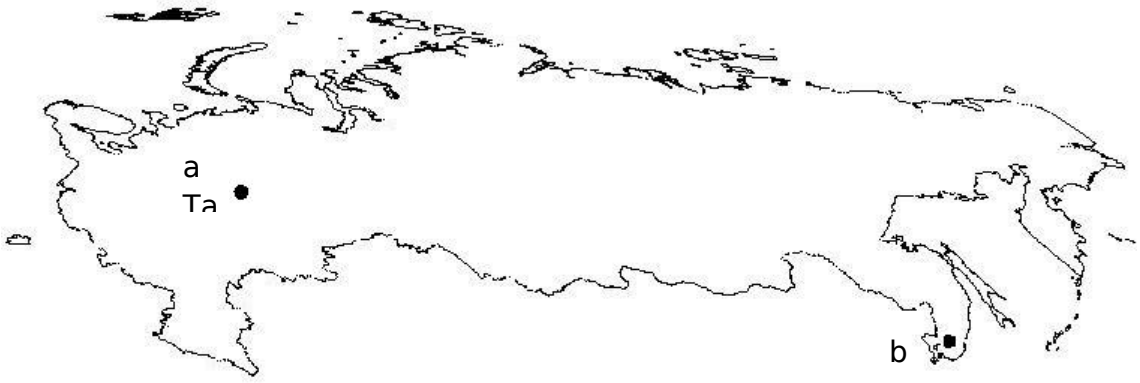

Figure 5. Map of Russia and the location of the sampling of *Crangonyx chlebnikovi* (a) and *Pseudocrangonyx korkishokoorum* (b).

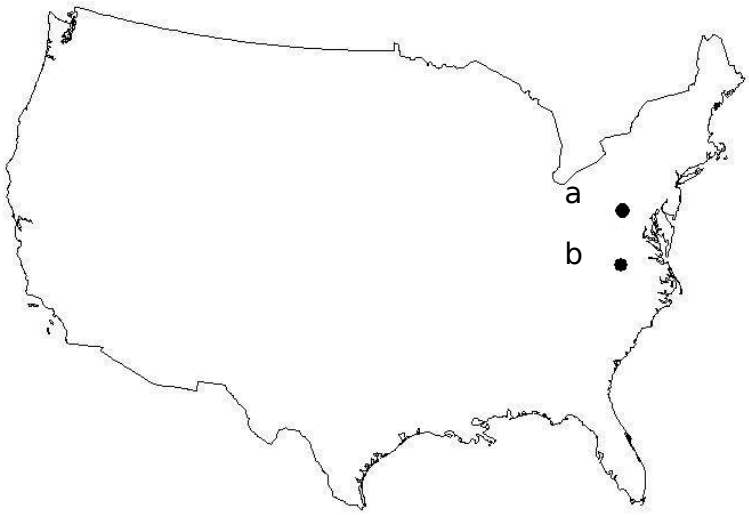

Figure 6. Map of USA and the sampling locations of *Stygobromus gracilipes* (a and b).

DNA extraction

DNA was extracted by Kornobis et al. 2010 by the following protocol:

Chelex protocol (6% (w/v) of the chelating resin in 10mM Tris/1mM EDTA solution with 10 mg/ml of proteinase K (Walsh et al, 1991) was performed. Muscle tissue of each individual was placed in 5% (w/v) Chelex solution and incubated at 65ºC for 2-3 h, warmed to 100ºC and centrifugated at 1300rpm during 2 min. After centrifugation, the DNA supernatant was stored in 5% Chelex at the fridge.

Chelex preserve DNA from denaturing due to DNAases while proteinase K digests all the proteins that remain in the solution.

Polymerase chain reaction

Polymerase chain reaction (PCR) is a technique used to amplify a single or few copies of a particular DNA fragment across several orders of magnitude, generating large number of copies. The method relies on thermal cycling, consisting of cycles of repeated heating and cooling resulting in DNA denaturation and enzymatic replication of the DNA (Figure 2). Primers (short DNA fragments) consisting of sequences complementary to the target region along with a DNA polymerase (after which the method is named) are key components to enable selective and repeated amplification. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified. PCR can be extensively modified to perform a wide array of genetic manipulations (Figure 5).

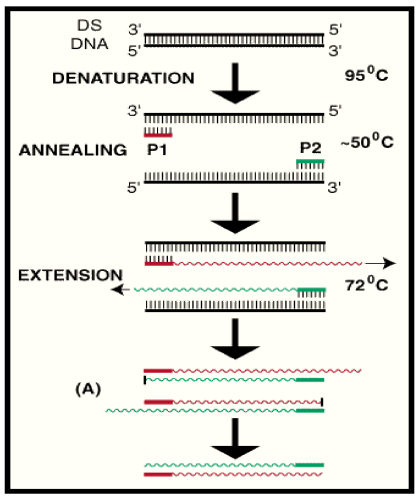

Figure 7. Diagram of PCR (Polymerase Chain Reaction) for a double strain DNA chain showing the steps and approximated temperature.

Two fragments of the 16s ribosomal RNA were amplified using two different forward Primers: 16Stf (MacDonald et al. 2005) and 16Sf77 with the same 16Sbr (Palumbi et al. 1991) reverse primer for both PCRs (Table 1) .The size of the fragments were 420bps and 1200bps (Figure 3).

Table 3. Primers used in PCR amplification and in sequencing reaction.

Primer	Direction	Sequence(5´-3´)
16Stf	Forward	GGT AWH YTR SCY GTG CTA AG
16S F77	Forward	CCG GTT TGA ACT CAG ATC ATC T
16Sbr	Reverse	ATT TAG TGT ACA AAT YGC CCG

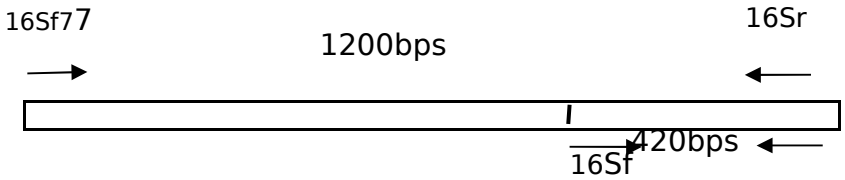

Figure 6. Diagram of the 16S gene fragment amplified by PCR using two forward primers and one reverse primer.

PCR amplification was conducted by using either *Taq* (NEB) and *Dream taq* (Fermentas) in 10 µl of mix conformed by: 1 µl of DNA template ant different concentrations, 4.82 µl of distilled water, 0.75 µl *dd*NTP's (2mM), 1.15 µl of Tween 20% (1M), 1 µl of *taq* buffer, 1 µl BSA (5mg/ml), 0.34 µl of both forward and reverse primers and 0.1 µl of taq polymerase.

Different thermocycle programs were used to achieve a good amplification of the different sequences starting by 4 minutes at 94ºC to denature the DNA chain, and followed by 35 or 40 cycles 30s of 94ºC, 1 min at annealing temperature between 45ºC-60ºC depending on the sample and 45s at 72ºC. A last step of 6 minutes at 72ºC follows after the cycles.

In order to assess if PCR amplifications worked agarose gel (1.5%) electrophoresis were performed at 100 mV during 15 minutes. Results were confirmed by taking pictures under UV light using ethidium bromide to locate the DNA. Size of fragments was assessed using a 100bp DNA ladder.

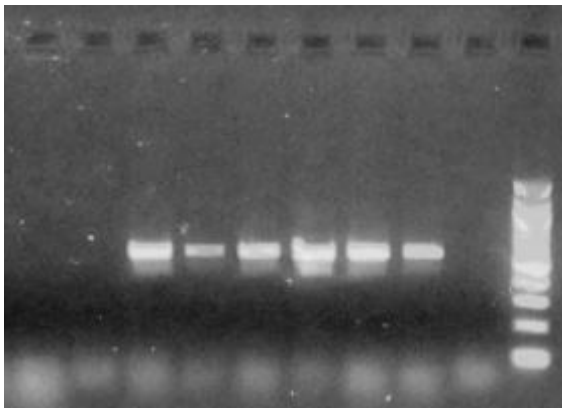

Figure 9. Agarose gel electrophoresis from the PCR amplification of the 16S rRNA gene from *C. islandicus*.

Exosap reaction

An exosap reaction was performed in order to clean the PCR products before the sequencing reaction, by removing dNTPs and primers remaining in the mixture that could interfere with the sequencing reaction.

The exosap reaction was done in 7 µl mixtures, using 5 µl of the amplified DNA, 0.75 µl ddH_20, 0.7 µl of Antartic phosphatase buffer, 0.5 µl Antartic phosphatase and 0.05 µl Exonucleas I.

The Exonucleas I degrades residual single-stranded primers and any strange single-stranded DNA chain in the PCR product while Antartic phosphatase hidrolyzes remaining dNTPs which would interfere with the sequencing reactions.

The reaction takes place in two steps. First 30 minutes at 37ºC for the exosap reaction of suppressing residues and a second step of 15 minutes at 80ºC to inactive enzymes.

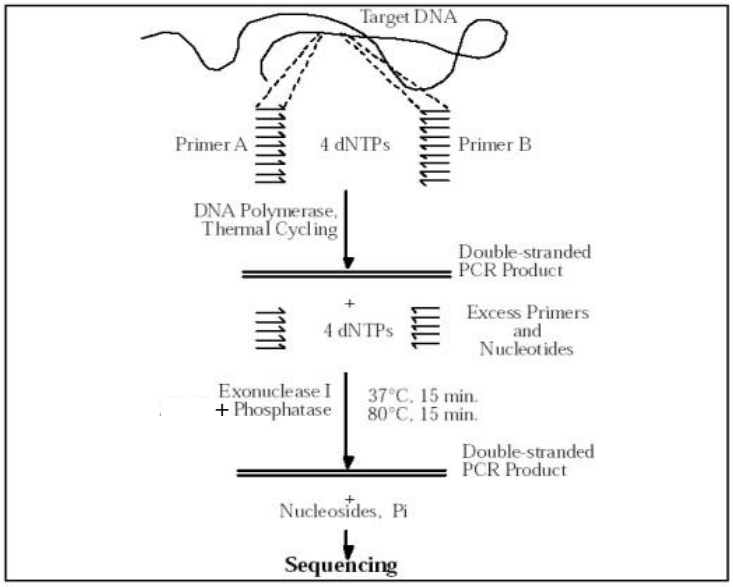

Figure 10. Diagram of the process to obtain a nucleotide sequence from a DNA extraction.

Sequencing

Each locus was sequenced in both directions using 1µM amplification primers to obtain overlapping sequences that allowed us to get the complete fragment.

Sequencing reactions were performed using Big Dye mix containing: Amplitaq DNA polymerase, dNTPs, ddNTPsDye terminators pyrophosphatase, Buffer and $MgCl_2$. The 10 µl volume reaction contained 2.65 µl ddH_2O, 1.75 µl, 1µl big dye, 1.6µl primer and 1µl PCR product, in case of weak PCR amplification more DNA was used.

The sequencing reaction program consisted of 10 seconds at 96°C, 25 cycles of 10 seconds at 96°C, 5 seconds at 50°C and 2 minutes at 60°C.

Ethanol precipitation

Ethanol precipitation was performed before running the samples on ABI Prism 3100 Genetics Analyzer, to purify the sequencing products of reactives. The protocol is the following:

- Add 50 µl of NaOAC (0.3M) to each 5-10 µl sequencing product.

Table 4. Composition of the NaOAC(0.3M) used in the ethanol precipitation

Component	Amount	[ ] in stock	[ ] in reaction
H20	9 ml		
3M NaOAC	1 ml	0.3 M NaOAC	0.081 M
Stock glycogen	50 ml	o.1 mg/ml	0.027 mg

- Add 125 µl of cold (-20ºC) 96% ethanol. Mix by tipping tubes up and down.
- Spin down at 4000 rpm for 30 minutes at 0-4ºC. Immediately pour off the precipitation mix.
- Wrap with two kim wipes, and spin down inverted at 300 rpm for 2 minutes at 0-4ºC to remove the rest of the ethanol.
- Carefully add 200 µl of cold (-20ºC) 70% ethanol.
- Spin down at 4000 rpm for 5 minutes at 0-4ºC, immediately dump the EtOH and spin inverted, covered with kim wipes, at 300 rpm for 5 minutes at 0-4ºC.
- Let tubes air dry in the dark for 15 minutes to get rid of lost traces of EtOH.
- Add 10 µl of HiDi, vortex and spin down.

Data analysis

Genetic distances were calculated both for the short (420bps) and the long (1200bps) fragments. Sequences downloaded from Genebank were added to the ones obtained in the lab in order to have an out group. This were *Maarrka sp, Chydaekata sp* and *Pilbarus sp.*

BioEdit software was used to read the sequences and to align them and to identify segregating sites. This software was also used to complete the 1200bps sequence using the 420bps sequence in the unreadable ending of the sequence.

Mega (Tamura et al. 2008) was used for constructing the phylogenetic tree based on the obtained sequences and to evaluate its reliability with bootstrap values. Neighbor-joining distance model was used to create the trees. The principle of this method is to find pairs of operational taxonomic units (OTUs= neighbors) that minimize the total branch length at each stage of clustering of OTUs starting with a starlike tree of the three taxonomic units. The branch lengths as well as the topology of a tree can quickly be obtained by using this method.

To analyse sequence variation "R" (R Development Core Team, 2008) was used to select unique sequence. The ape package in R was also used to find the best-fitting model to construct the trees (ape packageand was used to select the best evolutionary distance model among the sequences, applying the method developed by Posada (Posada and Crandall, 2001), and then using the maximum likelihood program Phyml (Guindom and Gascuel, 2003) to find the tree which maximizes the likelihood of the observed sequences. R was also used to draw the geographical maps (map package). To assess the variation distribution along the 1200bps 16S rRNA fragment a model was constructed to compare the variable sites between: different *C. isalndicus*; *C. islandicus*

and *C. thingvallensis*; ; *C. islandicus* and *S. gracilipes*; *C. thingvallensis* and *S. gracilipes*.

The best fitting model used to construct the tree was Tamura-Nei (1993) according to the Akaike information criterion. This model corrects for multiple hits, taking into account the differences in substitution rate between nucleotides and the inequality of nucleotide frequencies (Table 3). It distinguishes between transitional substitutions rates between purines and transversional substitution rates between pyrimidines. Equality of substitution rates among sites was assumed so gamma was 0. This model was selected by the "R" package, using the function "phyml" and choosing the one with the maximum liklelihood.

Table 5. The transition probability model according to the TN model (Tamura and Nei 1993). Different substitutions rates are between transversions (beta) and transitions (alpha) are given.

	A	T	C	G
A	-	βg_T	βg_C	$\alpha_1 g_G$
T	βg_A	-	$\alpha_2 g_C$	βg_G
C	βg_A	$\alpha_2 g_T$	-	βg_G
G	$\alpha_1 g_A$	βg_T	βg_C	-

DNAsp5 software was used to calculate the number of variable sites, the number of haplotypes (K), the number of variable sites(S), the haplotype diversity (Hd) and the nucleotide diversity (P*i*),for both 420bps and 1200bps fragments within *Crangonyx islandicus* (Table 5). Also, it was used to create the data set used in Network software to create the unrooted network for *Crangonyx islandicus* using both fragments.

Results

Sequence diversity

For the 1200bps of the 16S rRNA fragment, 19 sequences were obtained. Fifteen were from *C. islandicus* from 12 different locations (Figure 12). One sequence was of *C. thingvallensis* and the other three of *Stygobromus gracilipes* (table 4). For the 420bps of the 16S rRNA fragment, 103 sequences of *C. islandicus* were obtained by Kornobis et al. (2010). 13 more sequences were obtained by PCR amplification and direct sequencing. Three sequences were downloaded from Genebank (Table 4).

Table 4. Data set description for both 420bps and 1200bps fragments.

Fragment	Specie	Nº	Haplotypes	Source
1200bps	*C. islandicus*	15	14	PCR
	C. thingvallensis	1	1	PCR
	S. gracilipes	3	3	PCR
420bps	*C. islandicus*	103	17	PCR
	C. thingvallensis	1	1	PCR
	S. gracilipes	5	3	PCR
	Crangonyx chelbnikov	3	2	PCR
	Crango. pseudogracilis	1	1	PCR
	Pseudocran. korkishkoor	2	1	PCR
	Symurella ambulance	1	1	PCR
	Maarrka sp	1	1	Genebank
	Chydaekata sp	1	1	Genebank
	Pilbarus	1	1	Genebank

Table 5. Description of the C*rangonyx islandicus* sequences used in the study. Length is the number of bases of the fragment; Nº is the number of sequences obtained; K is the number of haplotypes; S is the number of variable sites; Hd is the haplotype diversity; P*i* is the nucleotide diversity.

Specie	Length	Nº	K	S	Hd	P*i*
Crangony x islandicus	1045	15	14	64	0.094	0.020
	414	103	17	24	0.098	0.011

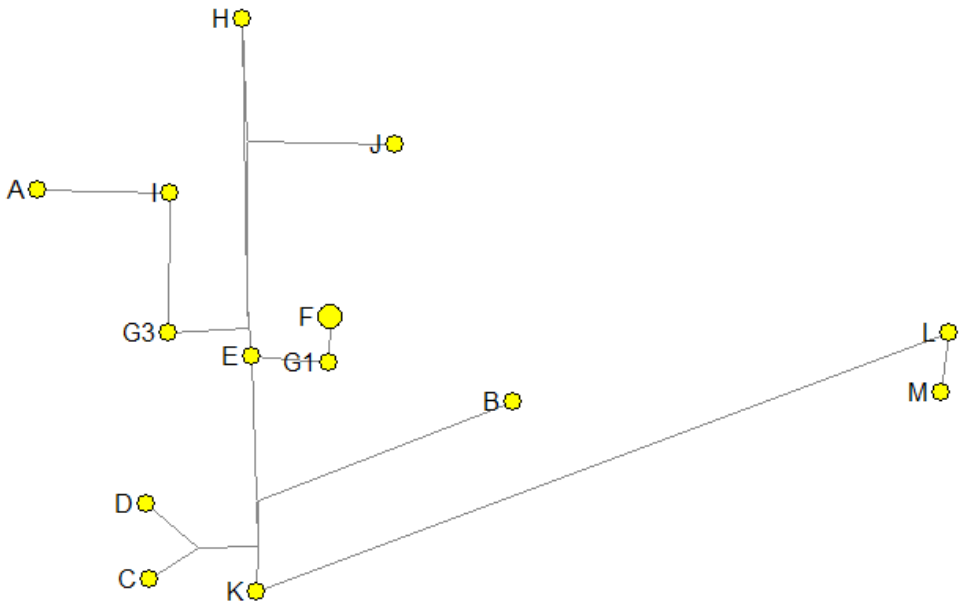

Figure 11. Unrooted network of *C. islandicus* using the 1200bps fragment from 16S gene. "M" and "L" docs show individuals from the north "F" zone described by Kornobis et al (2010). Distances between haplotypes corresponding to number of substitutions. Names of haplotypes (A-M) refer to different haplotypes found at the same location, shown in figure 12. Numbers G1 and G3 refer to different haplotypes found at the same location.

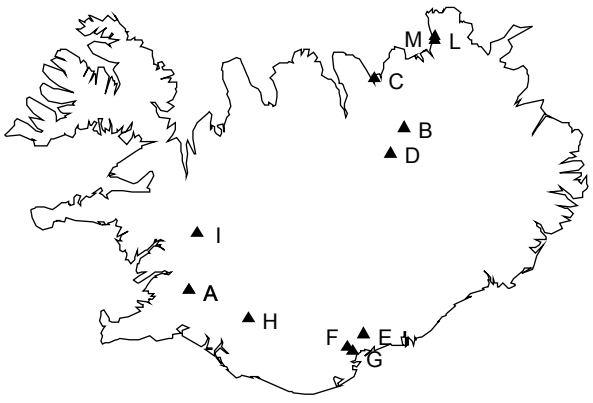

Figure 12. Map of Iceland showing the sampling location of the individuals sampled for the 1200bps fragment.

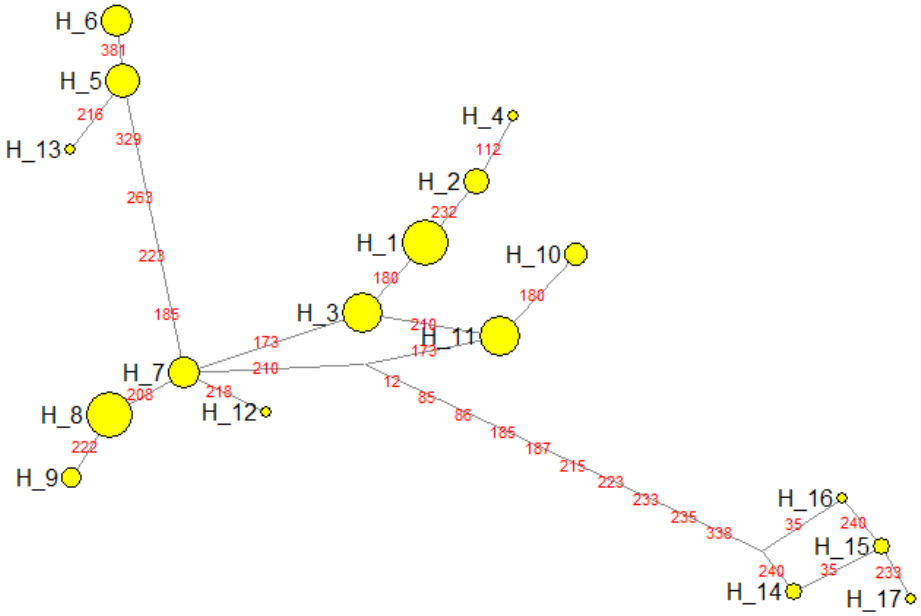

Figure 13. Unrooted network of *C. islandicus* using the 420bps fragment from 16S gene. "H_14" to "H_17" docs show individuals from the north "F" zone described by Kornobis et al (2010). The size of the haplotype circle refers to the number of individual that share that sequence.

Table 6. *Crangonyx islandicus* haplotypes for the 420bps fragment. The sequences were obtained by direct sequencing by Kornobis et al. (2010).

Place in the tree	Number of haplotypes	Individuals
H_1	16	0.1 0.15 0.5 0.6 0.8 0.9 B1.1 H1.2 S1.2 v1 V10 v2 v4 V7 V8 V9
H_2	5	0.11 0.12 0.4 H1.1 H1.3
H_3	13	0.18 0.2 H2.5 H2.7 H3.1 H3.2 H3.4 H3.7 S1.1 S1.3 S1.4 V11 v3
H_4	1	0.3
H_5	9	1.1 1.2 1.3 1.4 2.1 H4.1 H4.2 H4.3 H4.5
H_6	8	2.2 2.3 2.4 3.1 3.2 3.3 3.5 3.6
H_7	8	4.1 4.2 4.3 4.4 L2.2 L3.3 L3.4 L3.6
H_8	16	5.2 5.3 6.1 6.2 6.3 L2.1 L3.1 L3.2 L3.5 L3.7 L3.8 M2.1 M2.2 M2.3 M2.4 M2.5
H_9	3	5.6 5.7 6.6
H_10	4	7.1 7.3 L1.3 L1.4

H_11	12	7.4 7.6 7.8 H2.1 H2.2 H2.3 H2.4 H2.6 H2.8 L1.1 L1.2 L1.5
H_12	1	G1.1
H_13	1	H4.4
H_14	2	K1.1 K1.2
H_15	2	K2.1 M1.1
H_16	1	K3.1
H_17	1	M1.2

The unrooted three networks of the *Crangonyx islandicus* for both fragments of the 16S rRNA gene (Figure 11 and 13) show the distribution of the different sub-populations along Iceland as described in Kornobis et al. (2010). The “F” described by Kornobis is clearly appreciable in both networks (groups “M” and “L” in the 1200bps network and groups H_14 to H_17 in the 420bps network) as a more distanced group due a higher number of variable sites. Table 6 shows the number and name of the individuals that compose each haplotype as well as its position in the network.

In the 420bps network is also clearly appreciable the group “E” described by Kornobis et al. (2010) corresponding to the haplotypes H_5, H_6 and H_13. This group is not clear in the 1200bps network.

Phylogenetic relationship of Crangonyctidae based on 16S rRNA gene

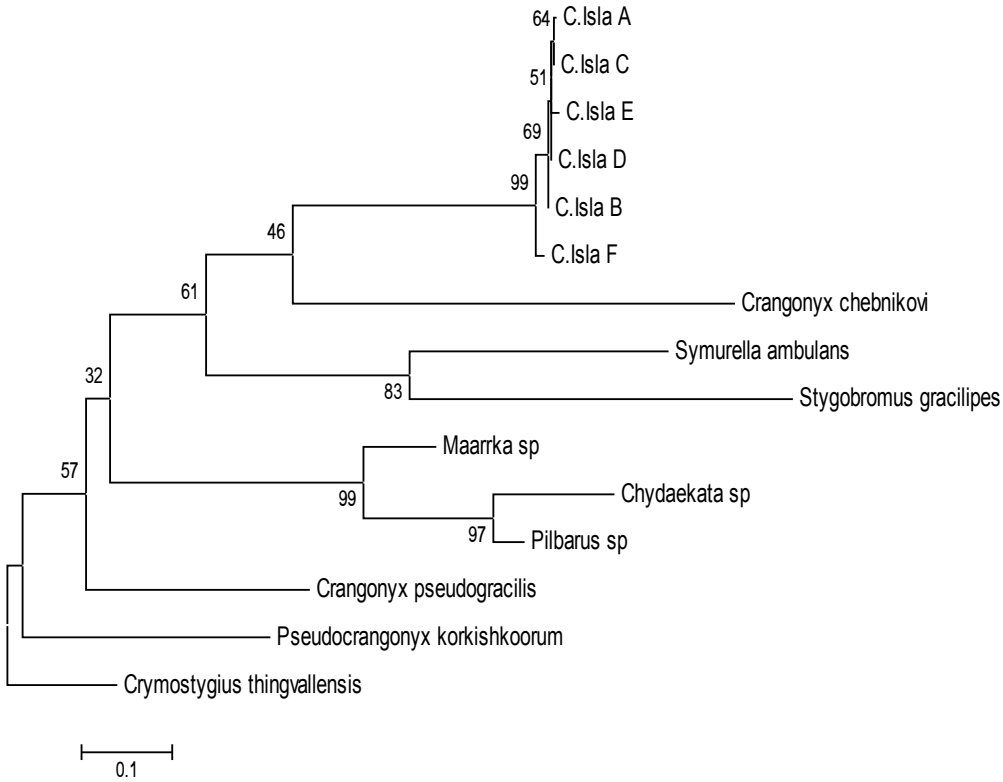

Figure 14. Phylogenetic tree based on the 420bps fragment of the 16S rRNA gene. The tree was constructed with Mega software using the Tamura-Nei (1993) model.

Figure 14 shows the phylogenetic relationship between *Crangonyx islandicus* from the different locations assessed by Kornobis et al. (2010) and different species of Amphipods. The tree was rooted with *Crymostygius thingvallensis* as outgroup. The specie *Crangonyx pseudogracilis* was found to be outside of the Crangonyctidae family according to the 16S gene, however the bootstrap value is low , suggesting that more data is needed for a reliable estimate of this split.

Phylogeography of *Crangonyx islandicus* based on 16S rRNA gene.

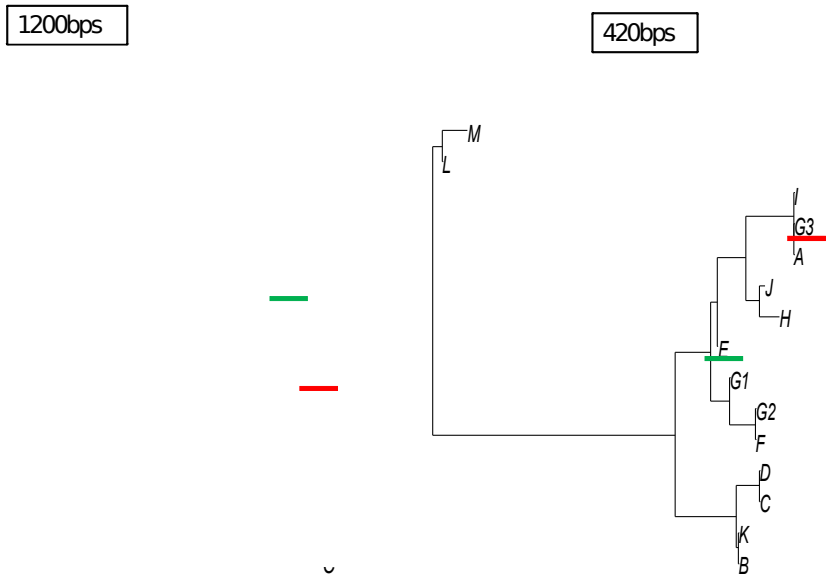

Figure 15. Comparison of the phylogenetic trees of Crangonyx islandicus using the 420bps and 1200bps fragment. Stygobromus gracilipes was used as outgroup and deleted from the final image.

The same fifteen individuals were sequenced both for the 1200bps and 420bps fragments and two different trees were constructed following the TN93 model using *Stygobromus gracilipes* as outgroup. As expected clear differentiation was found between “M” and “L” individual from the northern ”F” area described by Kornobis et al (2010) and the rest of individual.

Small differences were found comparing both trees and comparing the results with the distribution described by Kornobis et al. (2010). A different position inside the tree was found for two out of fifteen sequences (sequences E and G3) when more information was added using the long fragment due to a higher number of variable sites along the 1200bps fragment.

The sequence G3 was found to have a closer relationship with species described in the “A” zone by Kornobis et al. (2010) than with the other individuals from the G sampling zone.

Variation along the 1200bps fragment of the 16S rRNA gene

Distance between sequences were studied both intra-specific for *C. islandicus* and between species to compare the evolution rate between the 420bps and the 1200bps fragment. Distances between species were studied using C. thingvallensis, S. gracilipes and two samples of C. islandicus: one from the A population and the other from the F population described by Kornobis et al. (2010).

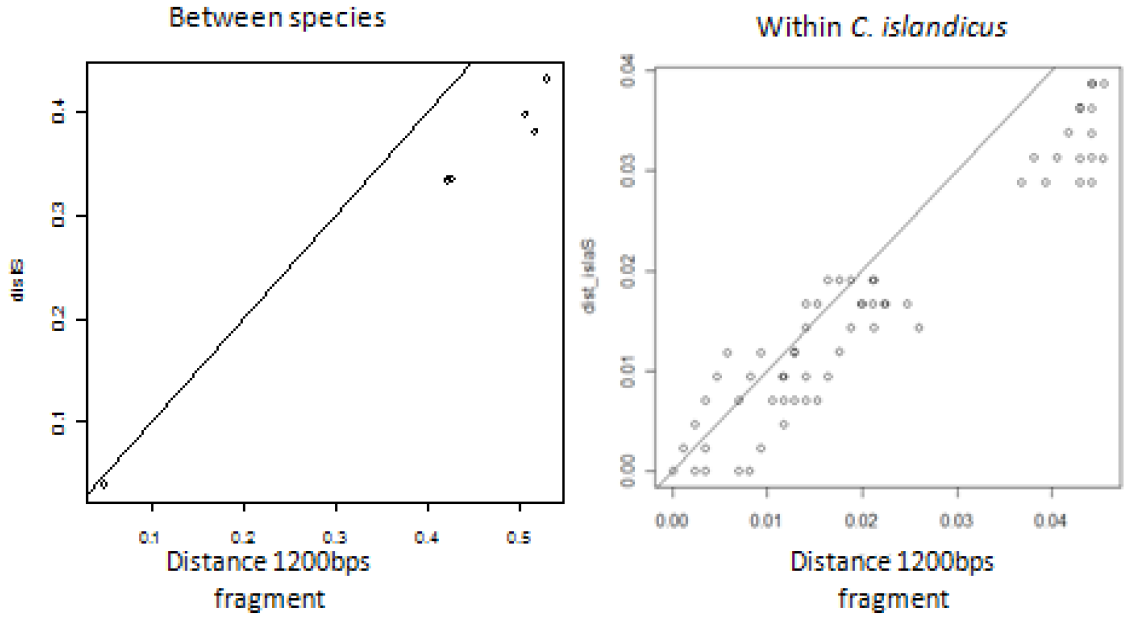

Figure 16. Comparison of sequences distances between the 420bps (Y-axis) and the 1200bps (X-axis) fragment both interspecies (left) and within species (right) for *C. islandicus*. A higher evolutionary rate was found for the 1200bps fragment in both cases.

A faster evolutionary rate was found along the 1200bps fragment. In the comparison between distances, the long fragment showed a higher value both intra- and inter-specific.

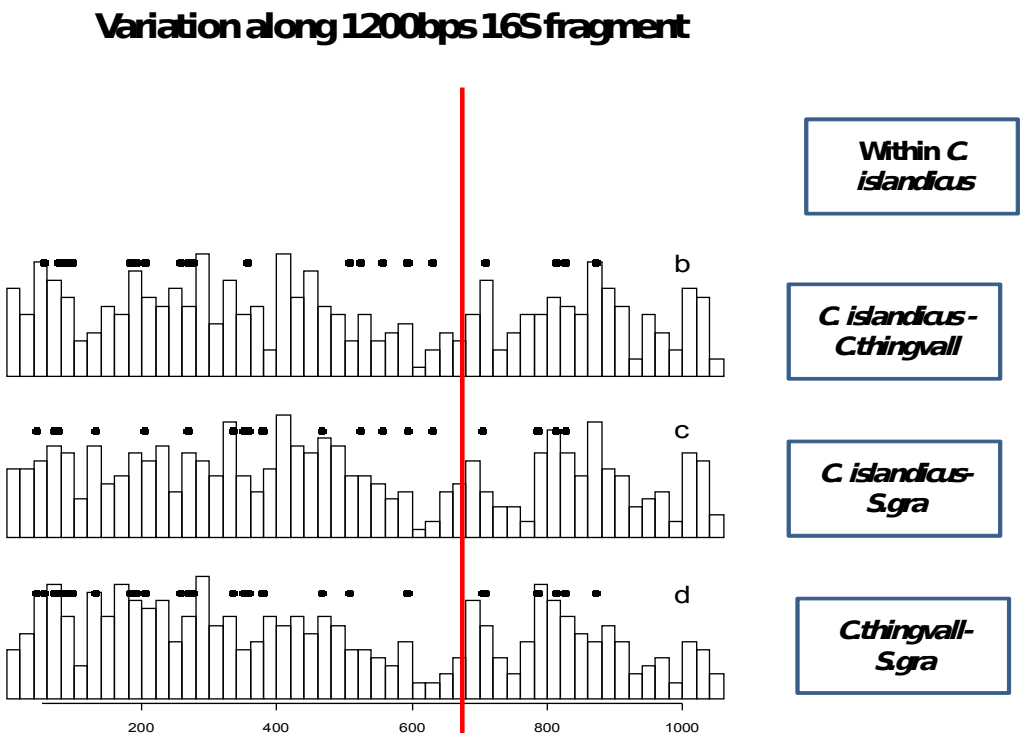

Figure 17. Diagram showing the variable sites both within C. islandicus and between species. The bars show the number of variable sites each 20bps. The black docs show the insertions and deletions in each 20bps.

A higher number of variable sites were found in the 700bps from the 5'side corresponding to the 1200bps fragment. No deletions or insertions were found within *C. islandicus* for the 420bps fragment. A higher number of gaps were found along the first 700bps within and between species.

Discussion

Phylogenetic analysis of Crangonyctidae

The phylogenetic tree shows a clear differentiation between *C. islandicus* and the other species of the family Crangonyctidae. That may be explained by an early separation of ancestor populations well-before Iceage.

The classification of *Crangonyx pseudogracilis* out of the Crangonyctidae family could be a good example of divergence between morphology based on taxonomy and DNA barcoding. The small bootstrap value make this conclusion though not very trustful as it is not possible to assess the real position of this specie within the superfamily according to the 420 bps fragment of the 16S rRNA gen.

A combination between morphology based on taxonomy and DNA barcoding could be the best way to assess the phylogeny of a taxa since it seems not possible to guarantee that the phylogeny based on the 16S gene is better than the classic morphology based on taxonomy. The 1200bps fragment could be used to verify this conclusion unfortunately it was not possible to obtain this sequence by direct sequencing. It would also be useful to use different marker, such as the mitochondrial COI, to compare the results and verify this clustering.

Phylogeography of *Crangonyx islandicus* based on 16S gene.

The high genetic diversity observed within *C. islandicus* differs from the low diversity observed in other high latitude species (Sadler 1999; Hewitt 2004), which are commonly characterized by shallow genealogies within regions. Such patterns have been explained by repeated population contractions towards southern refugia during the cold periods of the Ice Age and expansion phases towards Arctic regions during warmer periods. The divergence within the different *C. islandicus* populations may be explained by the constriction and insulation of the ancestor populations during the glaciations leading to a genetic divergence between refugia sub-populations. This can clearly be observed comparing the individual from the "L" and "M" zones included in the "F" sub-population as described by Kornobis et al (2010). Those individuals have diverged earlier from the other sub-populations, forming the first insulated sub-population. After that the other groups have been formed, possibly in different insolated refugia.

The differences in the comparison between the 420bps and 1200bps trees may be explained by higher variation rate among the 1200bps fragment which could lead to a better insight of the evolutionary history of the specie. Adding more information is giving clearer classification although it may not be necessary to spend resources, as not much is gained, in acquiring the 1200bps sequences when the 420bps sequence has already been used.

Variation of the 1200bps fragment

The higher variation among the first 700bps sequence respect to the following 500bps were the 420bps fragment is included may be explained by the assumption that the less conservative sequences have a less important role in the final function of the gene. If the mutation happens in a very conservative region it may selected against. The selection against non-synonymous substitutions at the DNA level is called "purifying selection", it refers to the probability that a substitution in a base is translated in the same aminoacid. If it doesn't happen the protein may not have the proper function, resulting in a lower fitness of its carrier. The evolutionary distance based on synonymous substitutions is expected to be greater than the distance based on non-synonymous substitutions due to the fact that the non-synonymous are not passed to the offspring. This is true for proteins coding regions, however this study focussed on variation in a RNA gene which is not coding for amino acids but nevertheless, regions with the gene such as in basal stems of hairpins may be more variable than binding sites in the loops of the hairpins.

The region of the gene corresponding to the 420bps fragment seems to follow this assumption since it was found to be more conservative. This may be important to establish the size of the fragment to amplify in further studies. If the phylogeny within species is to be studied the 420bps can give not enough information to really asses so the 1200bps can be a good option since it can give us more information. Meanwhile if we want to study the phylogeny between species, at genera or family level, the 420bps may be enough to have a good criterion to assess this relationship due to that the 1200bps seems to be evolving faster and we can have a saturation of information not being possible to discriminated between conservative sites and sites that have mutated and then mutated again to the same nucleotide.

Acknowledgements

I thank Snaebjörn Pálsson for giving me the opportunity to work in his group and teaching me things that I could never learn on my own. To everyone working in the labs expecially to Vanessa and Sonia for helping me and keeping me company. I want to thank especially Etienne Kornobis for explaining me so many things, not getting annoyed every time that I interrupted him and for being very nice with us. Also thanks to the University of Iceland and to the Erasmus scholarship for making my stance possible in Iceland.

References

Arbogast BS, Kegany, G.J. 2001. Comparative phylogeography as an integrative approach to historical biogeography. Jour of Biogeogra. 14: 1153-1158.

Avise JC, DeEtte Walker, Glenn C. Johns. 1998. Speciation Durations and Pleistocene Effects on Vertebrate Phylogeography. Proceedings: Biological Sciences, 265:1707-1712

Avise, JC. Nelson, WS. Bowen, BW. Walker, D. Phylogeography of colonially nesting seabirds, with special reference to global matrilineal patterns in the sooty tern (*Sterna fuscata*). Molecular Ecology. 9.11: 1783 – 1792

Buckland PC, Perry DW, Gíslason GM, Dugmore AJ (1986) The pre-Landna´m fauna of Iceland: a palaeontological contribution. Boreas, 15, 173–184.

Buhay, J.E. Crandall, K.A.2005. Subterranean phylogeography of fresh-water crayfishes shows extensive gene flow and surprisingly large population sizes. Am Nat 142, S65-S67.

Caccone, A. Sbordoni, V. 2001. Molecular biogeography of cave life: a study using mitochondrial DNA for Bathysciine beetles. Evolution 55, 122-130.

Culver, C. kane, T.C. Fong, D.W. 1995. Adaption and Natural Selection in Caves: the evolution of *Gammarus minus*. Harvard university press.

Darwin CR (1860) The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, revised edn. D. Appleton, New York.

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.

Geirsdó ttir Á , Miller GH, Andrews JT. 2007. Glaciation, erosion, and landscape evolution of Iceland. Journal of Geodynamics, 43, 170–186.

Hebert, P. D. N. Cywinska, A. Ball, S.L. deWaar, J.R., 2003. Biologycal identification through DNA barcodes. Proc. R. Soc. Lond. B. Bio. Sci. 270, 313-321.

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

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London B, 359, 183–195.

Kristjánsson, B.K. Svavarsson, J. 2007. Subglacial Refugia in Iceland enable groundwaters Amphipods to survive glaciations. Am Nat. 170, 292-296.

Kristjánsson, B.K. Svavarsson, J. 2004. Crymostygidae, a new subterranean freshwater gammaridean amphipods (Crustacea) recorded from subartic Europe. J Nat His 38, 1881-1894.

Kornobis, E. Pálsson, S. Krisstjánsson, B.K. (2010) Molecular evidences of the survival of subterranean amphipods (Arthropoda) during Ice Age underneath glaciers in Iceland. Mol Eco.

Lefébeur, T. Douday, C.J. Gouy, M. Gilbert, J. 2006.Relationship between morphological and molecular divergence within Crustacea: Proposal of a molecular threshold to species delimitation. Mol Phylogenet Evol. 40, 435-447.

Lefébure T, Douady CJ, Malard F et al. (2007) Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (Niphargus rhenorhodanensis). Molecular Phylogenetics and Evolution, 42, 676–686.

MacDonald KS, Yampolsky L, Duffy JE (2005) Molecular and morphological evolution of the amphipod radiation of Lake Baikal. Molecular Phylogenetics and Evolution, 35, 323–343.

Oromí, P. Martín, J.L. Medina, A.L. 1991. The evolution of the hypogean fauna in the Canary Islands.

Palumbi S, Martin A, Romano S, McMillan WO (1991) The Simple Fool's Guide to PCR, version 2.0. University of Hawaii, Honolulu. Paradis E, Claude

Pamilo P. and Slovainen O., 1999. Post-glacial colonization, dirft, local selection and conservation value of populations: a northern perspective. Hereditas 130, 229-238.

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.

Porter ML ,2007 .Subterranean biogeography: what have we learned from molecular techniques? Journal of Cave and Karst Studies, 69, 179–186.

Porter, M.L. Crandal, K.A. 2003. Lost along the way: the significance of evolution in reverse. Trends Eco Evol 18, 541-547.

R Development Core Team, 2004. R. a language and environment for statistical computing. Foundation for statistical computing. Vienna.

Sadler JP (1999) Biodiversity on oceanic islands: a palaeoecological assessment. Journal of Biogeography, 26, 75–87.

Salvolainen, V. Cowarn, R.S. 2005. Towards writing the encyclopedia of live: an introduction to DNA barcoding. Philos. Trans. R. Soc. B 360, 1805-1811.

Sanz S. 2010.Mitochondrial variation of the caddiflies *Apatadia zonella* and *Potamophylax cingulatus*. University of Iceland.

Susan E. Masta Mitochondrial Sequence Evolution in Spiders: Intraspecific Variation in tRNAs Lacking the TΨC Arm Mol Biol Evol 17: 1091-1100.

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

Tajima, F, 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123: 585-595.

Witt, J.D.S. Herbert, P.D.N. 2000. Cryptic species diversity and evolution in amphipod genus Hyalella within central glaciated North America. A molecular phylogenetic approach. Can. J. Fish. Aquat. Sci. 57: 687-698.

Hou, Z. Fu, F. Li, S. 2007. A molecular phylogeny of the genus Gammarus based on mitochondrial and nuclear gene sequences. Mol. Phy. and Evol. 45. 596-611.