A novel PolI DNA polymerase in *Thermus* bacteria

Peter Bischofberger

Faculty of Life and Environmental sciences
University of Iceland

2010

A novel PolI DNA polymerase in *Thermus* bacteria

Peter Bischofberger

30 ECTS thesis
Research project in biology for foreign students, LÍF014M.

Supervisors
Guðmundur Óli Hreggviðsson
Snædís Björnsdóttir
Ólafur H. Friðjónsson

Faculty Representative
Guðmundur Óli Hreggviðsson

Faculty of Life and Environmental sciences
School of Engineering and Natural Sciences
University of Iceland

Division of Biomolecules and Biotechnology
Matis ohf.

Reykjavik, December 2010

Title: A novel PolI DNA polymerase in *Thermus* bacteria

Research project in biology for foreign students, LÍF014M

Faculty of Life and Environmental sciences
School of Engineering and Natural Sciences
University of Iceland
Askja, Sturlugata 7
101, Reykjavik
Iceland
Telephone: 525 4000

Division of Biomolecules and Biotechnology
Matis ohf.
Vinlandsleið 12
113 Reykjavik
Iceland
Telephone 422 5000

Bibliographic information:
Peter Bischofberger, 2010. A novel PolI DNA polymerase in *Thermus* bacteria.
School of Engineering and Natural Sciences University of Iceland, pp. 28.

Reykjavik, Iceland, 22 December 2010

Abstract

In former investigations (unpublished) a new type I polymerase (Family A), different to the prevalent one of *Thermus*, was found in a *Thermus antranikianii*. It has conserved domain structures, related to the distinct polymerase I subgroup of *Aquificales* and seems to feature strand displacement activity (unpublished) similar to the polymerase from phage phi29. The distribution of the gene in the *Thermus* genus was investigated with a PCR-screening. The gene was found to be present in some but not all strains of the species *T. thermophilis* (21% of strains examined), *T. scotoductus* (15% of strain examined), *T. igniterrae* (28% of strains examined) and *T. brockianus* (35% of strains examined), but not in *T. oshimai*. Prior sequence analysis of the gene flanking regions already had revealed that the gene is located in close proximity to a transposase gene, thus linked to it in *Thermus antranikianii*. The same linkage was observed in other *Thermus* strains, where the polymerase was present, by PCR-screening. The aquificae like PolI polymerase has sporadic distribution in other bacterial phyla. Only in the *Aquificae* phylum is it found in every species/strain examined. This is a very unusual species-distribution indicating later gene transfer. The distribution pattern in *Thermus* is even a stronger support for this hypothesis. The linkage to the transposase and other sequence features indicates also that the polymerase is transported by the means of a transposon that can move between strains and across species boundaries.

Table of contents

List of Figures

List of Tables

Abbreviations

SSU	Small subunit
Pol	Polymerase
IS	Insertion sequence
Blast	Basic Alignment Search Tool
bp	Base pairs
rRNA	ribosomal RNA
PCR	Polymerase chain Reaction
Tris	2-Amino-2Hydroxymethyl-propane-1,3-diol

Acknowledgements

I would like to thank my supervisor Guðmundur Óli Hreggviðsson for his supervision, inspirations and giving me the opportunity to do research for Matis during my exchange studies in Iceland. Furthermore I have to thank Ólafur H. Friðjónsson for his supervision and hints concerning bioinformatics, Sólveig Katrín Ólafsdóttir for all the help in the laboratory and Snædís H. Björnsóttir, who lent me a hand as supervisor in every situation. My studies in Iceland were supported and made possible by the Erasmus exchange program. The funding for the study is from Rannsóknasjóður Íslands and the study was a part of the Pan-Thermus project.

1. Introduction

An interest in thermophilic organisms followed the biotechnological revolution in the late 20th century. Defined as organisms, able to grow and reproduce at high temperatures, thermophiles possess proteins and cellular mechanics stable against heat denaturation. Many biotechnological processes require high temperatures (Haki et al., 2003) in order to function and/or to produce adequate yields. To name one application, polymerase chain reaction (PCR) is the best known and a revolutionary example of the use of thermophilic 'technology'.

1.1 Thermophiles

Definitions of thermophilic organisms (from ancient greek θερμός *thermós* „warm" as well as φίλος *phílos* „loving") vary in the literature, but they have been described as those which are able to grow at temperatures above 60°C (Rothschild et al., 2001). This definition excludes eukaryotes, as their highest found growth-temperate is 60°C. Therefore, the capacity is observed only in the prokaryotic domains. Thermophilic archaea can even grow above the boiling point of water and the highest known growth temperature is 121° C (Kashefi et al., 2003). Habitats with such high temperatures are rare and found in most cases in geothermal areas below sea level where the pressure is high enough. Thermophilic organisms may also represent the most ancient mode of life as they cluster around the root of the universal tree of life. The high selective pressure under extreme thermophilic conditions might have caused the 'evolutionary clock' to tick more slowly than in mesophilic counterparts (Stetter, 1996).

1.2 *Thermus*-species

According to the current taxonomy *Thermus* belongs to the *Deinococcus-Thermus* phylum. They are nonsporulating, with an outer membrane (Gram-negative), heterotrophic, rod shaped and obligate aerobes. They grow at near neutral, alkaline pH and as the name *Thermus* indicates the bacteria belonging to this genus are thermophilic growing at temperatures between 55°C to 85°C, with growth optima 65 and 70°C.

The first isolates were obtained in the late 1960s by Brock et al., (1969) in the Yellowstone National Park, a highly active geothermal area. The discovered species *T. aquaticus* is the source of the Taq-polymerase, first discovered in 1976 (Chien et al., 1976). It has become essential for molecular biology and biotechnology. Since then, many other thermophilic polymerases have been cloned, expressed, developed and found uses in various applications. Still, Taq polymerase is the favorite enzyme for PCR (Gibbs et al., 2009).

1.3 DNA-polymerase

The thermophilic Taq DNA-polymerase is of fundamental importance for molecular biology research applications such as PCR gene amplifications and DNA sequencing. Thermo-stability is required due to repeated denaturation steps during the PCR.

DNA directed DNA polymerases are the central enzymes in DNA replication and DNA repair processes. Currently they are categorized into six DNA polymerase families, based on their biochemical and sequence relationship. This polymerase subgroups are Family A, B, C, D, X, and Y. The prevalent families in bacteria are Family A (PolI), B (PolII), and C (PolIII). Polymerase III is part of the replication complex (replisome), which replicates the chromosomal DNA before cell divisions. Polymerase I and II only assist in strand synthesis, especially on the discontinuous 3' to 5' direction. They have repair and gap-filling functions.

A novel polymerase called Thermophi in this report (derived from thermophilic) and belonging to the PolI polymerase family was recently discovered in various Icelandic *Thermus* strains (Hjörleifsdóttir, Hreggviðsson and Friðjónsson, 2010; unpublished). The gene of one particular *Thermus antranikianii* strain (2120) was cloned and sequenced. The obtained sequence was compared with other PolI polymerase sequences and found to belong to the Aquificae like PolI polymerase group.

The family A (PolI) polymerases can be divided into different subgroups based on presence or absence of conserved domains with different functions. The domain structure of the Thermophi polymerase can be seen in Figure 1 and comparisons to other PolI genes can be seen in Figure 2. Previously it had been shown that the well-known Taq polymerase from *Thermus aquaticus* belongs to the PolI family *E. coli* like PolI C. This type of PolI polymerase has been detected in all *Thermus* species and strains investigated. It functions as the PolI C from *E. coli* in DNA repair. On the other hand the domain structure of the Thermophi polymerase clearly indicates that it belongs to the PolA (*Aquificae* like) subgroup (Fig. 2). Polymerases I C (*E.coli*-like) have three domains, the characteristic polymerization domain, the 5'-3'-exonuclease domain and the proofreading 3'-5'-exonuclease domain. In contrast the Thermophi polymerase only has the proofreading 3'-5'-exonuclease domain and the polymerization domain. The Taq polymerase has the three PolI C (*E. coli* like) domains but the 3'-5'-exonuclease domain is smaller and non-functional.

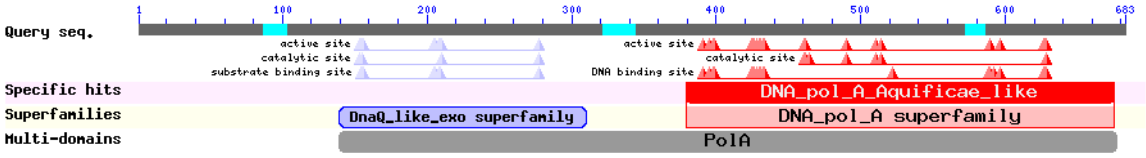

Figure 1. Conserved domains within the Thermophi amino acid sequence. They were identified using the NCBI search engine, based on the Pfam-database (Pfam – Finn et al. 2010). Sequence similarities mark the belonging to the family A polymerases and their subgroup PolA (*Aquificae* like). Furthermore it possesses a 3'-5'-exonuclease domain for proofreading (blue marked).

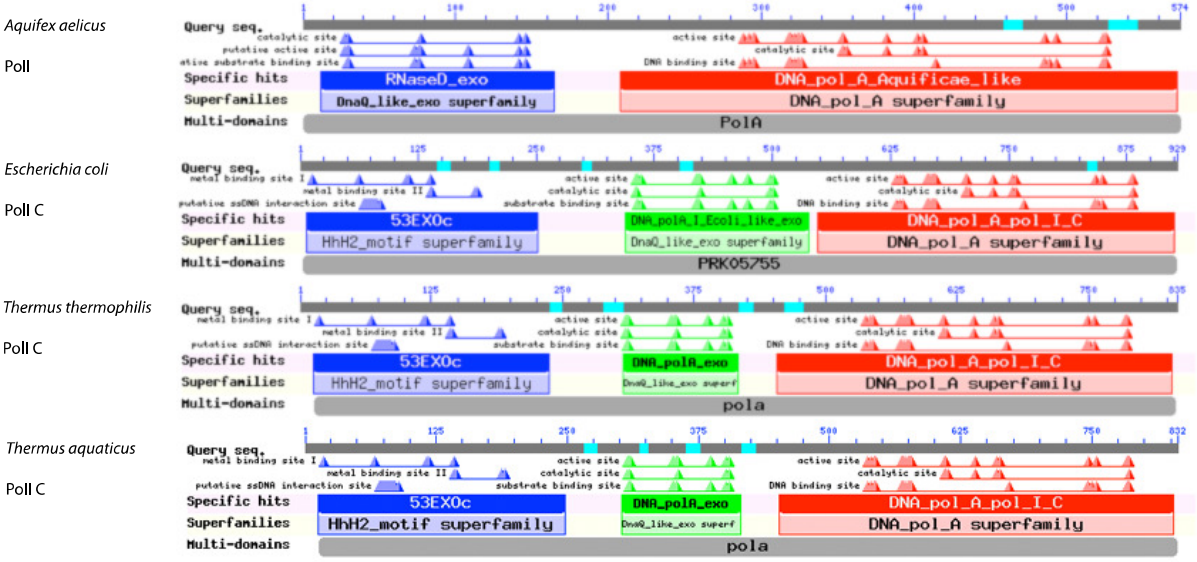

Figure 2. Polymerase I domain structures of *Aquifex aelicus* (GenBank O67779.1), *Escherichia coli* (AAA24402.1), *Thermus thermophilis* (YP_144320.1) and *Thermus aquaticus* (1TAQ_A). Polymerases I C (*E.coli* like) have the characteristic polymerization domain (red marked), the 3'-5'-exonuclease domain for proofreading (green marked for PolI C; blue marked for *A. aelicus*) and in

contrast to PolI in *Aquifex,* there is an additional 5'-3'-exonuclease domain (blue marked). However, in both *Thermus* species the 3'-5'-exonuclease domain is smaller and non-functional.

The Thermophi-polymerase appears to have unique properties, which make it interesting for further investigation and biotechnological utilization. Investigations carried out at Matis indicate that the Thermophi polymerase has strand displacement activity, similar to the polymerase that originates from the phage phi29 (unpublished). Conventional polymerases used in PCR have problems with amplifying long sequences and areas with a high G/C – content, because they have no 'helicase'-function for splitting the DNA double strand and stop the replication reaction if a strand is not denatured completely.

1.4 Transposons

Bioinformatical analyses of the Thermophi gene sequence and its flanking regions from two strains of *Thermus* revealed that the gene is located in close proximity to a transposase gene. This fact and other sequence features indicated that the polymerase gene might be a part of a transposon. Consequently, the gene could have a sporadic distribution within the genus, either species specific or strain specific and independent of species boundaries.

Transposons have the ability to move from one site in the genome to another, leading to a rearrangement of the genome. Insertions within a gene can cause mutations and incorporation of stop codons or termination sequences and can affect the expression of most genes, as observed by Dyson et al., 1999.

Both ends of transposons are flanked by inverted repeats, which can vary in length from 9 to 41bp, being closely related to each other, but not identical. During a transposition a small copy of the target host-DNA sequence is created due to staggered cutting and later filled on complementary. This leads to characteristic similar direct-repeats on both sides of the sequence where the transposon was inserted to (Mahillon et al., 1998). Though these areas can not pointed out very clearly on the *T. antranikianii* genome yet, there is evidence that these sequences exist and are located close to the Thermophi gene.

However, a transposon that includes a polymerase has not yet been reported in bacteria. The specific usage and possible advantage of the Thermophi polymerase for a species is unknown. Here we find evidence that the gene is linked to a transposase and can be found in many strains of several *Thermus* species. This was demonstrated by PCR screening using different primer locations within the gene and transposase, respectively. There is no specific pattern, regarding the appearance of the transposon in distinct geothermal areas. Therefore, its role in evolution might be to keep genetic variability, as proposed by Nevers *et al.* (1977).

1.5 Aim of the study

The aim of this study is to investigate the distribution of the Thermophi gene in the *Thermus* genus. The hypothesis is that Thermophi polymerase is a transposon-encoded polymerase and can move between strains and species by lateral gene transfer.

2. Materials and Methods

2.1 Bacterial strains

For this work, 96 *Thermus* strains (Tab. 1) were screened for the frequency of the Thermophi gene. The strains were previously isolated from the geothermal areas around Iceland shown in Figure 3. Most of the strains had been classified to species using isoenzyme analysis (Hreggvidsson et al., 2006).

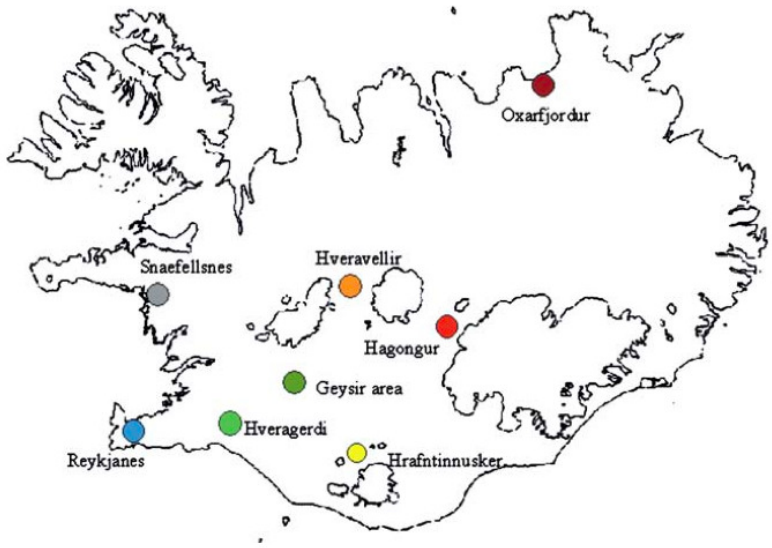

Figure 3. Geothermal areas of Iceland from which the *Thermus* strains, used in this work, were isolated from (Hreggvidsson et al., 2006).

Table 1. *Thermus* strains used in this work, their growth conditions and former classification by isoenzyme analysis according to Hreggvidsson et al. (2006).

Strain	Growth temperature [°C]	Medium	Classification in Hreggvidsson et al., 2006	Strain	Growth temperature [°C]	Medium	Classification in Hreggvidsson et al., 2006
51	65	162	*T. oshimai*	**2111**	72	166	*T. brockianus*
52	65	162	*T. scotoductus*	**2115**	72	166	*T. brockianus*
74	65	166	-	**2117**	65	R_2A	*T. igniterrae*
77	65	162	*T. oshimai*	**2118**	65	166	*T. scotoductus*
79	65	162	*T. brockianus*	**2120**	72	R_2A	*T. antranikianii*
80	65	162	*T. oshimai*	**2121**	65	R_2A	*T. scotoductus*
129	72	162	*T. brockianus*	**2122**	65	R_2A	*T. scotoductus*
140	72	162	*T. brockianus*	**2123**	65	166	*T. sp.*
154	65	162	*T. scotoductus*	**2124**	65	R_2A	*T. sp.*
165	65	162	*T. igniterrae*	**2125**	72	R_2A	*T. scotoductus*
206	65	162	*T. brockianus*	**2126**	65	166	*T. scotoductus*
210	65	162	*T. brockianus*	**2127**	60	R_2A	*T. scotoductus*
211	65	162	*T. brockianus*	**2128**	65	R_2A	*T. brockianus*
220	65	162	*T. oshimai*	**2129**	65	R_2A	*T. brockianus*
252	65	166	-	**2130**	65	166	*T. igniterrae*

Strain	Growth temperature [°C]	Medium	Classification in Hreggvidsson et al., 2006	Strain	Growth temperature [°C]	Medium	Classification in Hreggvidsson et al., 2006
319	60	162	*T. brockianus*	**2131**	65	166	*T. oshimai*
338	72	162	*T. brockianus*	**2132**	65	166	*T. igniterrae*
339	72	162	*T. brockianus*	**2133**	65	166	*T. igniterrae*
340	72	162	*T. brockianus*	**2134**	65	166	*T. igniterrae*
346	72	162	*T. scotoductus*	**2135**	65	R_2A	*T. oshimai*
360	72	162	*T. brockianus*	**2137**	65	R_2A	*T. igniterrae*
761	65	R_2A	*T. thermophilus*	**2139**	72	166	*T. sp.*
781	65	R_2A	*T. thermophilus*	**2141**	70	166	-
791	72	R_2A	*T. thermophilus*	**2142**	70	166	-
797	65	R_2A	*T. thermophilus*	**2143**	70	166	-
862	65	166	-	**2144**	65	R_2A	*T. igniterrae*
872	65	R_2A	*T. thermophilus*	**2145**	70	166	-
945	65	160	-	**2146**	70	166	-
1003	65	166	-	**2147**	70	166	-
1087	65	166	-	**2148**	70	166	-
1251	72	R_2A	*T. thermophilus*	**2149**	70	166	-
1262	72	166	*T. thermophilus*	**2150**	70	166	-
1270	65	166	*T. thermophilus*	**2151**	70	166	-
1285	72	166	*T. thermophilus*	**2152**	70	166	-
1318	65	166	*T. thermophilus*	**2153**	72	166	*T. brockianus*
1340	65	R_2A	*T. thermophilus*	**2154**	65	166	*T. scotoductus*
1373	72	166	*T. thermophilus*	**2214**	72	166	-
2100	65	166	*T. scotoductus*	**2443**	65	R_2A	*T. igniterrae*
2101	72	R_2A	*T. scotoductus*	**2631**	72	166	*T. scotoductus*
2102	72	166	*T. scotoductus*	**2788**	65	166	*T. brockianus*
2103	65	R_2A	*T. brockianus*	**2789**	72	R_2A	*T. sp.*
2104	65	166	*T. igniterrae*	**2790**	72	R_2A	*T. igniterrae*
2105	65	166	*T. igniterrae*	**2791**	65	R_2A	*T. igniterrae*
2106	72	166	*T. brockianus*	**2792**	65	R_2A	*T. igniterrae*
2107	72	166	*T. igniterrae*	**2793**	72	R_2A	*T. igniterrae*
2108	72	166	*T. scotoductus*	**2795**	72	166	*T. sp.*
2109	72	166	*T. scotoductus*	**2797**	72	R_2A	*T. brockianus*
2110	72	166	*T. igniterrae*	**2811**	65	R_2A	*T. scotoductus*

2.2 Growth media

Purified strains (from Matis/Iceland) were streaked onto agar plates as listed in Table 1, in order to get single colonies and freshly growing cells. Four different nutrient media were used for the cultivation: Medium 162 developed by Degryse et al. (1978). Medium 160, which is the same as 162 but with only 1/10 of the phosphate buffer, containing per liter 2.5 g yeast extract, 2.5 g tryptone, 1.0 g nitrilotriacetic acid, 0.4 g $CaSO_4 \cdot 2H_2O$, 2.0 g $MgSO_4 \cdot 7H_2O$, 1.5 ml 0.2 M $Na_2HPO_4 \cdot 12H_2O$, 1.0 ml 0.2 M KH_2PO_4, 0.5 ml 0.01 M Fe(III) citrate $\cdot 5H_2O$ and 5.0 ml of trace element solution, pH 7.5. Medium 166 contains per liter geothermal tap water, 0.3 g of K_2HPO_4, 1 g of yeast extract, 1 g of peptone, 1 g of tryptone, 0.5 g of glucose, 0.5 g of starch, 0.6 g of pyruvic acid, 0.3 g of proline and 0.18 g of Na_2CO_3, 0.5 ml of 0.01 M Fe(III) citrate $\cdot 5 H_2O$ and 5.0 ml of trace element solution, pH 7.5. And for 1 l medium R_2A (Difco) 0.5 g yeast extract, 0.5 g proteose peptone, 0.5 g casamino acids, 0.5 g glucose, 0.5 g soluble starch, 0.3 Na-pyrauvate, 0.3 g K_2HPO_4, 0.03 g $MgSO_4 \cdot 7H_2O$ is needed. The trace element solution contains (per liter): 0.22 g $MnSO_4 \cdot H_2O$, 0.05 g $ZnSO_4 \cdot 7H_2O$, 0.05 g H_3BO_3, 0.0025 g $CuSO_4 \cdot 5H_2O$, 0.0025 g $Na_2MoO_4 \cdot 2H_2O$ and 0.0046 g $CoCl_2 \cdot 6H_2O$.

2.3 DNA Extraction

DNA was isolated from cells grown on agar plates overnight using the MasterPure™ DNA Purification Kit (Epicentre, Madison, WI, USA) according to the manual of the manufacturer. All extractions were examined for quality and quantity on agarose gels, before being used as templates for PCR.

2.4 PCR – screening

PCR amplifications with the Matis-produced T_{eg}-polymerase were performed with initial denaturation at 94°C for 4 min, 30 amplification cycles of 94°C for 50 s, 57 or 70°C for 50s and 72°C for 1min per 1 kbp of the predicted product size, and a final extension step for 7 min at 72°C. Three primer pairs were used for the screening and were designed according to the reference sequence of *Thermus antranikianii* (strain 2120).

As shown in Figure 4, the primers Pol11F/Pol11R (5'- GGAGGGGTTTGAACTCCACTAC-3'; 5'- TCATGCCTCCTCCCACGG-3'), T_A-57°C; Pol320F/Pol1170R (5'-ACTGGCCCACCAGGTGCTTC-3'; 5'- TCCTCCTTGCCCACCTCTTC-3'), T_A-57°C; both pairs were used for the detection of the Thermophi gene. The primers Tphi_tspF1/Tphi_tspF1 (5'- GGCCACGCCGTGGGAGGAG-3'; 5'- CCCTTGCCCTCGTGGTAGAAGAC-3'), T_A-70°C, were used to amplify the region between the Thermophi-gene and the transposase gene.

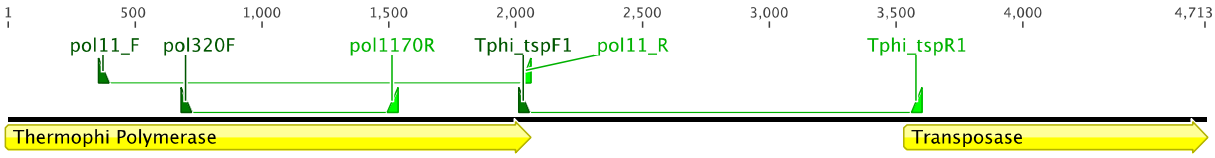

Figure 4. Primer design, based on the reference-sequence of *T. antranikianii* strain 2120. The primer-pairs and the predicted product sizes Pol11F/Pol11R (**Pair 1** - 1681 bp) and Pol320F/Pol1170R (**Pair 2** - 850 bp) are located within the **Thermophi**-gene. The third primer-pair Tphi_tspF1/Tphi_tspR1 (**Pair 3** - 1566 bp) indicates the linkage of both genes, closely located on the chromosome.

2.5 Sequence analysis

Nucleotide sequences were determined with an Applied Biosystems 3730 DNA analyzer and the BigDye terminator cycle sequencing kit. PCR amplifications of the genes were performed with T_{eg}-polymerase (Matis). PCR products were treated with Exo-Sap-It™ (Amersham Biosciences) prior to sequencing. Thermophi PCR-screening products were sequenced using Pol11R for the first primer-pair and Pol1170R for the second. Sequence analysis was performed using Geneious 4.8 (Drummond et al., 2010).

2.6 16S rRNA based phylogenetic analysis

Bacteria specific primers were used for the SSU rRNA gene and amplified with T_{eq}-polymerase (Matis). The positions on the gene are based on the *E. coli* SSU rRNA sequence number, as described in Skirnisdottir et al. (2000): F9 (5'-GAGTTTGATCCTGGCTCAG-3'; *E. coli* positions 9 to 27) and R805 (5'-GACTACCCGGGTATCTAATCC-3'; 805 to 785). The primer for the sequencing reaction was R805.

Subsequent classifications with the obtained 16S rRNA sequences were performed using the NCBI-Blast (Sayers et al., 2010) and aligned together with other SSU rRNA sequences of the *Thermus* group, obtained from NCBI-database. The GenBank accession numbers are as

follows: *T. scotoductus* IT252 (AF032127), *T. islandicus* PRI-2268 (EU753248), *T. aquaticus* YT-1 (TTHYT1), *T. thermophilus* HB8 (X07998), *T. igniterrae* RF-4T (Y18406), *T. brockianus* 15038T (Y18409), *T. antranikianii* HE-5 (Y18412), *T. oshimai* SPS-17T (Y18416). Evolutionary distances were computed from pairwise identity using the Tamura Nei correction. The software package Geneious was utilized to construct the phylogenetic tree by the neighbor-joining algorithm.

2.7 Phylogenetic studies of PolI polymerases

Three different subgroups of PolI (Family A) polymerases were compared; PolI A, PolI C and PolI *Aquificae* like. The amino sequences were obtained from the NCBI-database of different organisms with the following GenBank accession numbers: *Clostridium bolteae* (ZP_02085015), *Thermotoga neapolitana* (ACM23015), *Escherichia coli* (AAA24402), *Deinococcus radiodurans* (1922368A), *Thermus aquaticus* (1TAQ_A), *Thermus thermophilus* (YP_144320), *Lyngbya sp.* (ZP_01624319), *Plasmodium falciparum* (XP_001348285), *Toxoplasma gondii* (ACN59873), *Theileria annulata* (XP_954352), *Babesia bovis* (XP_001610510), *Clostridium sp.* (ZP_05130182), *Hydrogenobaculum sp.* (YP_002121207), *Aquifex aeolicus* (DPO1_AQUAE), *Sulfurihydrogenibium sp.* (YP_001930793), *Hydrogenivirga sp.* (ZP_02179064), *Rubrobacter xylanophilus* (YP_645884), *Gemmata obscuriglobus* (ZP_02730163), *Meiothermus silvanus* (ZP_04036592), *Desulfitobacterium hafniense* (YP_520594), *Clostridium hathewayi* (ZP_06409803), *Cyanothece sp.* (YP_001806245), *Cyanothece sp.* (ZP_03156829). The phylogenetic tree is based on the neighbor-joining algorithm with the Tamura Nei correction.

3. Results

3.1 PCR-screening for Thermophi gene in *Thermus*

Two primer pairs were used in the screening for the presence of Thermophi in different Thermus strains. One primer pair (Pair 1) was based on Thermophi regions that were located in conserved regions in all PolI sequences. The other pair (Pair 2) was based on non conserved regions but unique to Thermophi sequences.

25 of 96 strains were positive (Tab. 2) for the Thermophi gene as judged by the primer specific PCR-product bands on agarose gels (Fig. 5). The Pair 2 (Pol320F/Pol1170R) showed 50% more positives than Pair 1 (Pol11F/Pol11R). The assumption was that positives using primer Pair 2 resulted in unspecific annealing. This could be demonstrated by sequencing the PCR products.

3.2 PCR-screening for Thermophi polymerase, transposase linkage

In order to demonstrate that the Thermophi gene might be part of a transposon, the primer pair (Pair 3) was designed to evidence the close proximity to the transposase gene. Almost all the same strains, that seem to contain the Thermophi gene, display the transposase linkage as shown Figure 5. However, evidence was not found for the strains 872, 2102 and 2103, that seem to lack the linkage or cannot be screened with the same primers (Tab. 2). At least five strains (140, 206, 210, 211, 2130; all *T. brockianus*) also indicated, remarkably, a shorter region between the genes, what might indicate a different genotype or structure of the transposon perhaps with fewer direct-repeats.

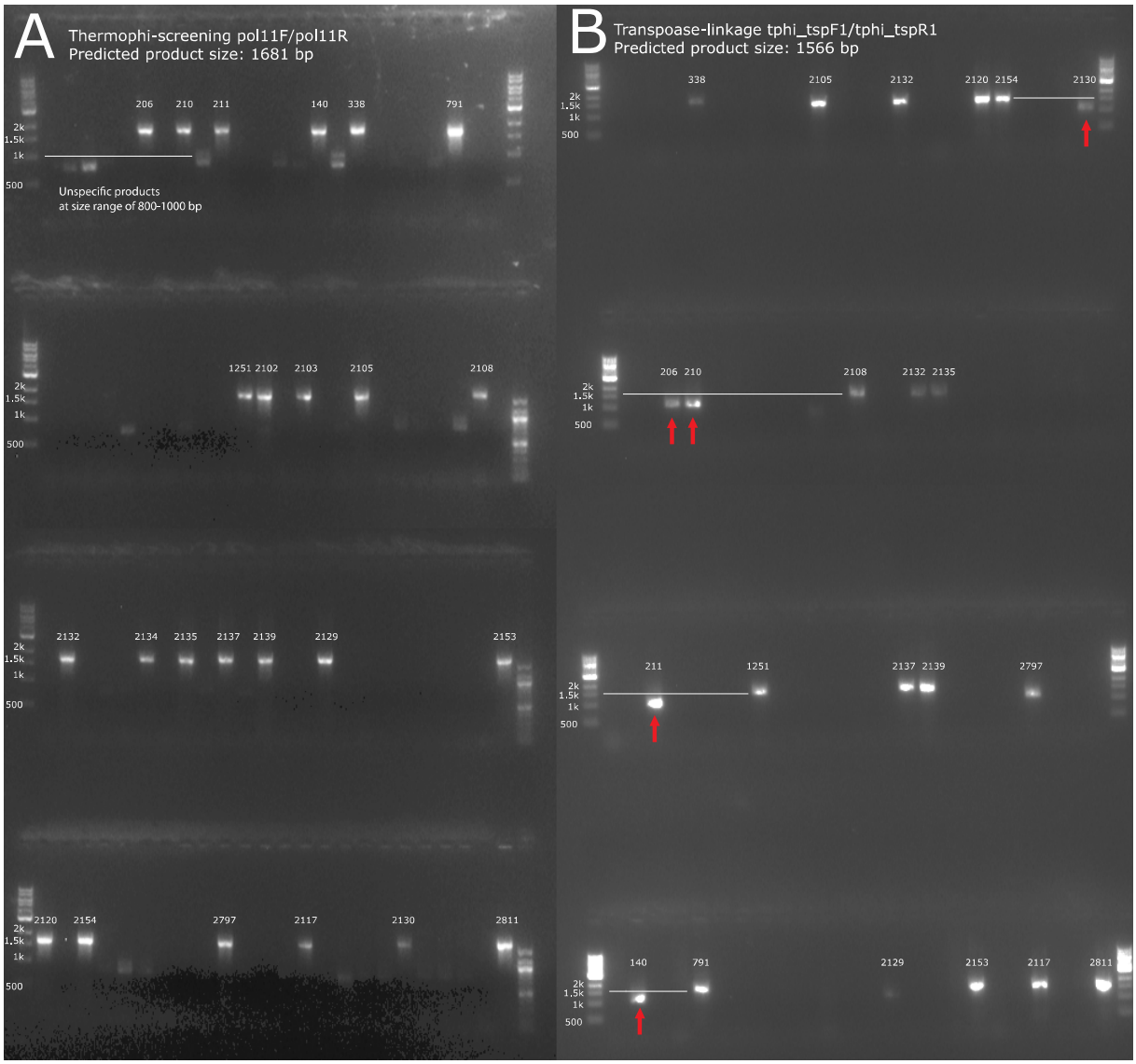

Figure 5. Agarose gel images that show the results of PCR-screenings. On the left (**A**) positive strains for Thermophi-gene screening is displayed. Few negative strains show small unspecific products, but varying clearly in length from the predicted product. The right side (**B**) displays the Thermophi-transposase linkage screening results with the predicted band sizes. But there are positive strains, whose amplification products seem to differ (red arrows) and are shorter than predicted.

Table 2. List of the positive strains, which showed the expected PCR product.

	Positive strains												
Thermophi-gene	140	206	210	211	338	791	872	1251	2102	2103	2105	2108	140
	2120	2129	2130	2132	2134	2135	2137	2139	2153	2154	2797	2811	2120
Transposase-linkage	140	206	210	211	338	791	-	1251	-	-	2105	2108	2117
	2120	2129	2130	2132	2134	2135	2137	2139	2153	2154	2797	2811	

3.3 Sequence analysis of PCR products

The obtained product sequences from both tested primer pairs (Pair 1, Pair 2) were aligned with the Thermophi-gene reference sequence of strain 2120 (*T. antranikianii*), shown in Figure 6. For Pair 1 the sequence reaction starting-primer was Pol11.R, for Pair 2 it was Pol1170.R in order to get an overlapping region of both sequences. All strains with positive PCR result (Pair 1) had genes highly similar to the reference gene, whereas this was not true for sequence products obtained using Pair 2.

Products of Pair 2 were obtained from five strains could not be aligned to the reference sequence. Those five strains, however, gave also no amplification products with the other primer pair (Pair 1). Sequencing and subsequent analysis revealed that the same sequence, different to Thermophi, was amplified in these five strains. Using the Blast, a high similarity with a transport-permease protein (not shown) in the *T. thermophilus* genome (strains HB8, HB27) was found

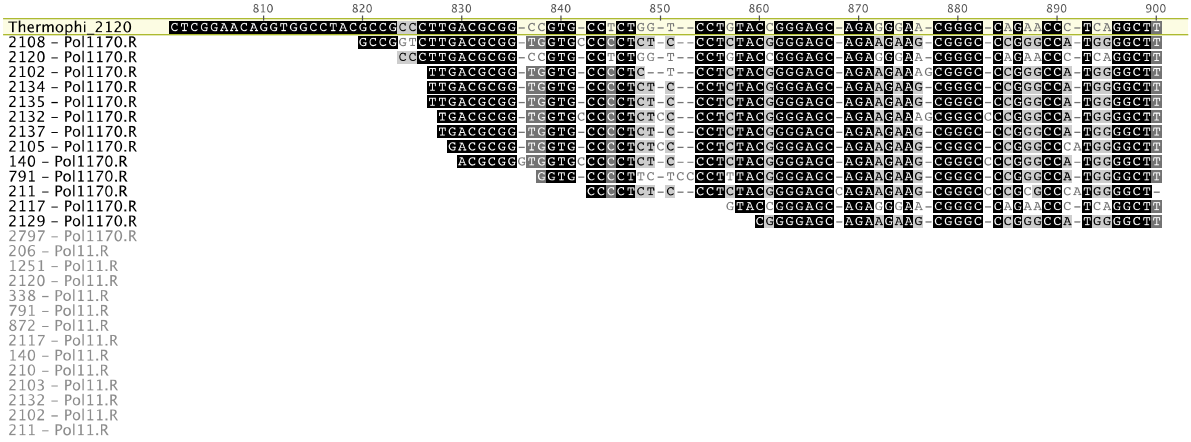

Figure 6. Alignment of the sequenced PCR products obtained with the primers Pol11.R (for Pair 1) and Pol1170.R (for Pair 2) with the Thermophi-gene reference sequence (*T. antranikianii* 2120). It gives proof for the presence of the gene within the strains. Black block and white letter means 100% similarity between the sequences; dark-grey block and white character 80 to 100%; grey block and black character 60 to 80%; white block and grey character less than 60%. There are more ambiguities at both ends of the obtained sequences due to weak sequencing signals. The alignment is continued on page 10, 11 and 12.

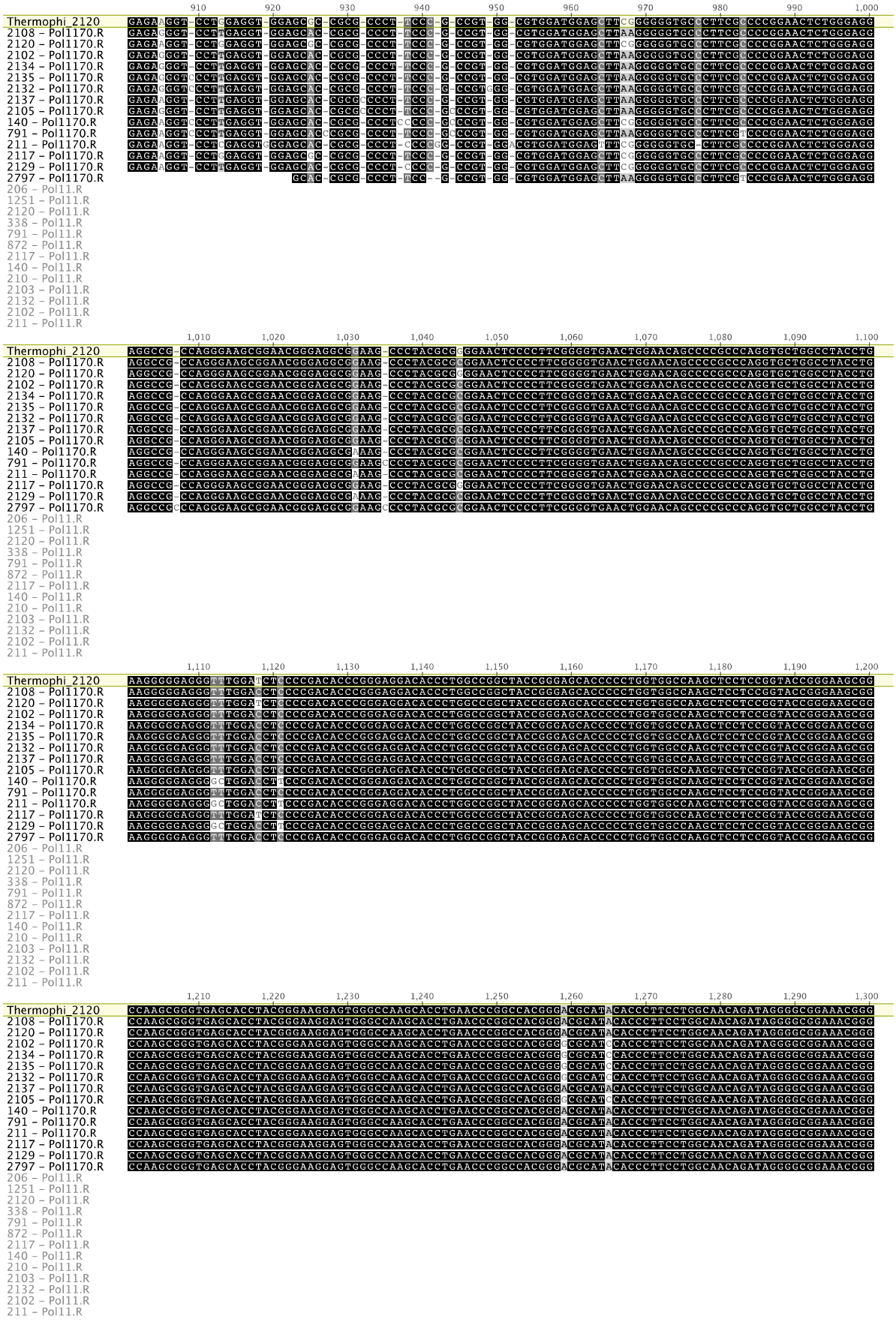

Figure 6. Continued.

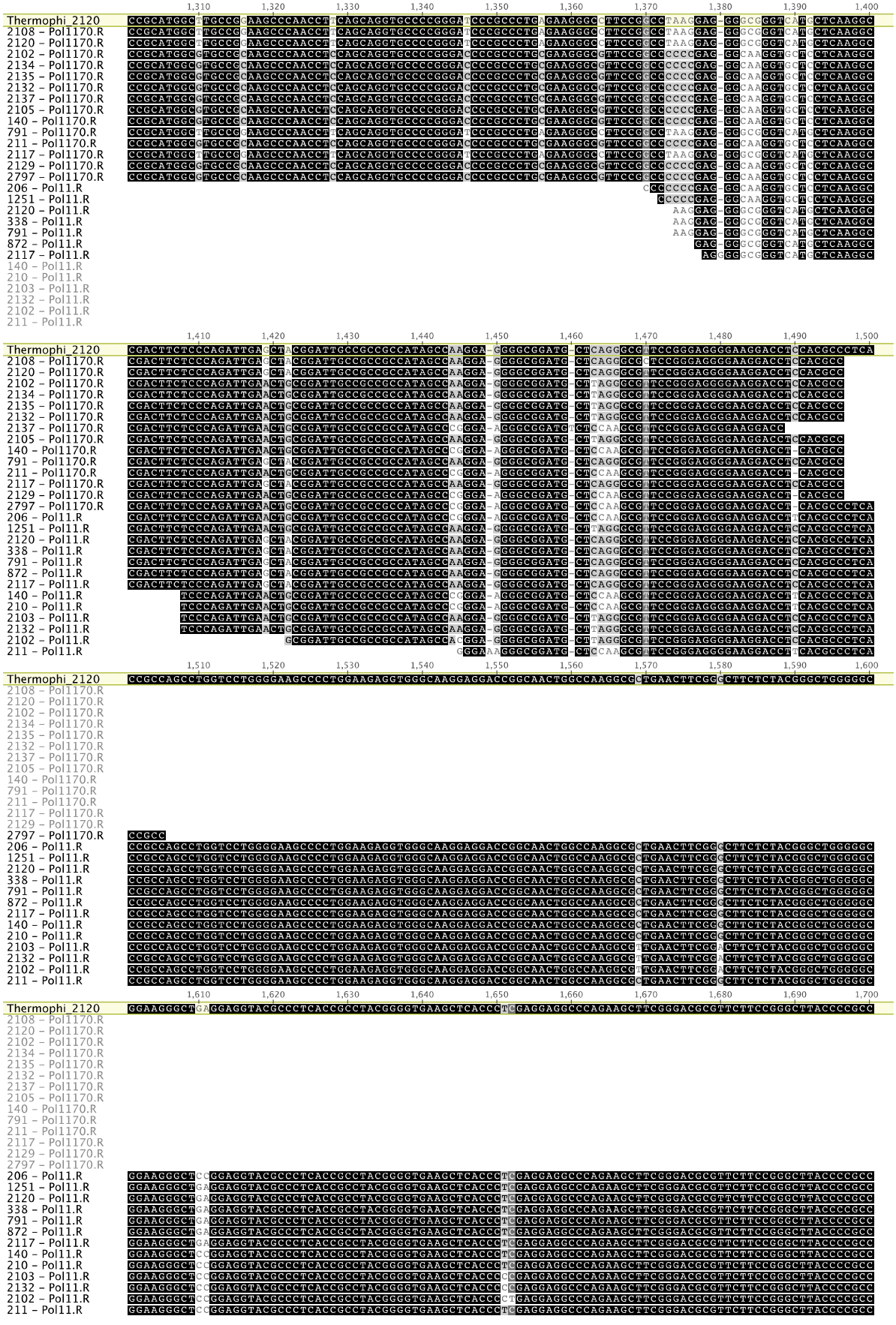

Figure 6. Continued.

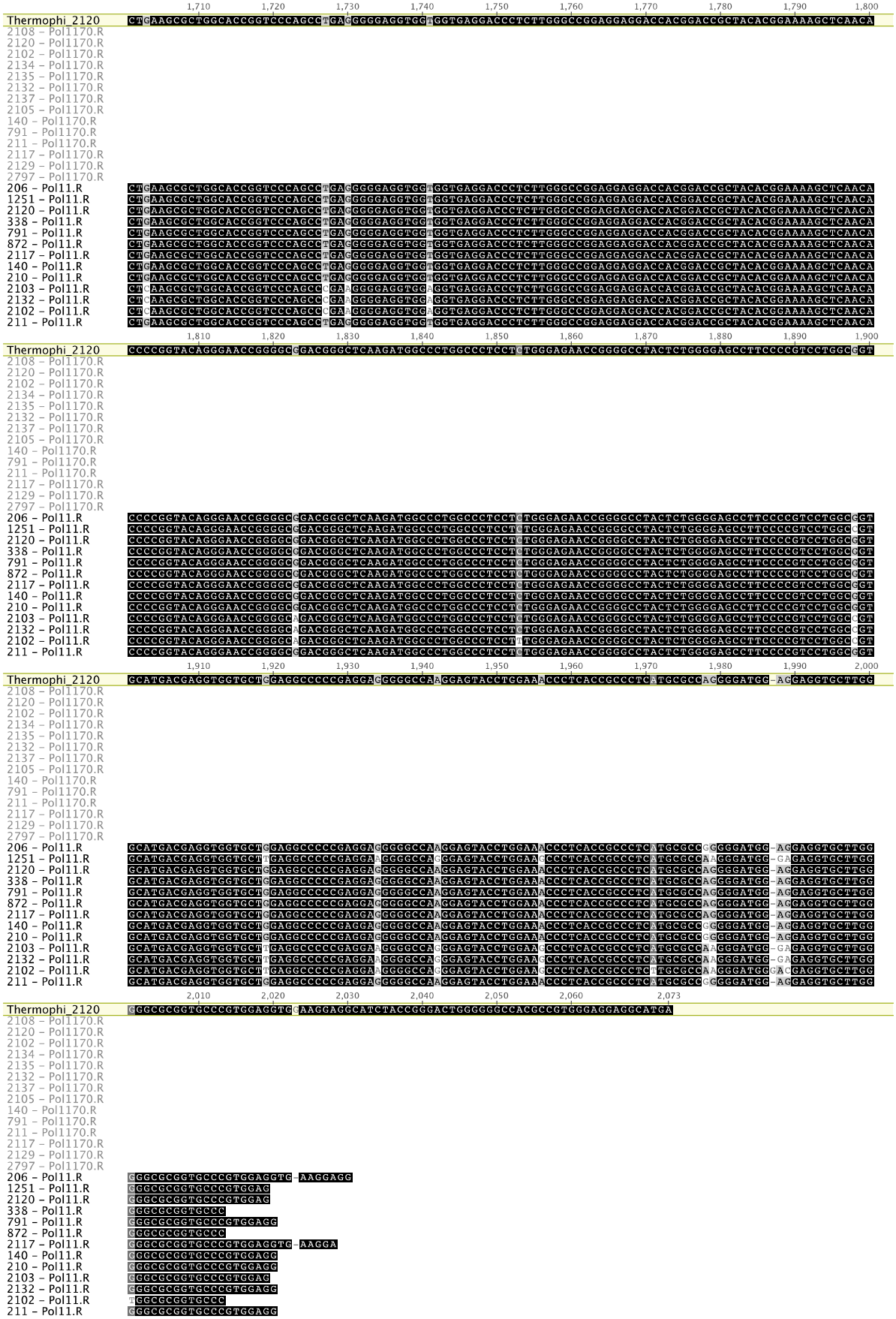

Figure 6. Continued.

Table 3. Phylogenetic classification based on the 16S ribosomal SSU gene-sequences, compared to the former classification of Hreggvidsson et al. (2006). Few formerly unclassified strains became classified, the *T. sp.* and some other species classifications changed mostly to *T. brockianus*.

Strain	Classification in Hreggvidsson et al., 2006	Classification by 16S-analysis	Strain	Classification in Hreggvidsson et al., 2006	Classification by 16S-analysis
51	*T. oshimai*	*T. oshimai*	**2111**	*T. brockianus*	*T. brockianus*
52	*T. scotoductus*	*T. scotoductus*	**2115**	*T. brockianus*	*T. brockianus*
74	-	*T. aquaticus*	**2117**	*T. igniterrae*	*T. antranikianii*
77	*T. oshimai*	*T. oshimai*	**2118**	*T. scotoductus*	*T. scotoductus*
79	*T. brockianus*	*T. brockianus*	**2120**	*T. antranikianii*	*T. antranikianii*
80	*T. oshimai*	*T. oshimai*	**2121**	*T. scotoductus*	*T. scotoductus*
129	*T. brockianus*	*T. brockianus*	**2122**	*T. scotoductus*	*T. brockianus*
140	*T. brockianus*	*T. brockianus*	**2123**	*T. sp.*	*T. brockianus*
154	*T. scotoductus*	*T. scotoductus*	**2124**	*T. sp.*	*T. brockianus*
165	*T. igniterrae*	*T. igniterrae*	**2125**	*T. scotoductus*	*T. igniterrae*
206	*T. brockianus*	*T. brockianus*	**2126**	*T. scotoductus*	*T. scotoductus*
210	*T. brockianus*	*T. brockianus*	**2127**	*T. scotoductus*	*T. brockianus*
211	*T. brockianus*	*T. brockianus*	**2128**	*T. brockianus*	*T. scotoductus*
220	*T. oshimai*	*T. oshimai*	**2129**	*T. brockianus*	*T. brockianus*
252	-	*T. scotoductus*	**2130**	*T. igniterrae*	*T. igniterrae*
319	*T. brockianus*	*T. brockianus*	**2131**	*T. oshimai*	*T. igniterrae*
338	*T. brockianus*	*T. brockianus*	**2132**	*T. igniterrae*	*T. igniterrae*
339	*T. brockianus*	*T. brockianus*	**2133**	*T. igniterrae*	*T. scotoductus*
340	*T. brockianus*	*T. brockianus*	**2134**	*T. igniterrae*	*T. igniterrae*
346	*T. scotoductus*	*T. scotoductus*	**2135**	*T. oshimai*	*T. igniterrae*
360	*T. brockianus*	*T. brockianus*	**2137**	*T. igniterrae*	*T. igniterrae*
761	*T. thermophilus*	*T. thermophilus*	**2139**	*T. sp.*	*T. brockianus*
781	*T. thermophilus*	*T. thermophilus*	**2141**	-	*T. scotoductus*
791	*T. thermophilus*	*T. thermophilus*	**2142**	-	*T. scotoductus*
797	*T. thermophilus*	*T. thermophilus*	**2143**	-	*T. scotoductus*
862	-	*T. scotoductus*	**2144**	*T. igniterrae*	*T. scotoductus*
872	*T. thermophilus*	*T. thermophilus*	**2145**	-	-
945	-	*T. thermophilus*	**2146**	-	*T. scotoductus*
1003	-	*T. brockianus*	**2147**	-	*T. igniterrae*
1087	-	*T. thermophilus*	**2148**	-	*T. igniterrae*
1251	*T. thermophilus*	*T. thermophilus*	**2149**	-	*T. igniterrae*
1262	*T. thermophilus*	*T. thermophilus*	**2150**	-	*T. scotoductus*
1270	*T. thermophilus*	*T. thermophilus*	**2151**	-	*T. scotoductus*
1285	*T. thermophilus*	*T. thermophilus*	**2152**	-	*T. scotoductus*
1318	*T. thermophilus*	*T. thermophilus*	**2153**	*T. brockianus*	*T. brockianus*
1340	*T. thermophilus*	*T. thermophilus*	**2154**	*T. scotoductus*	*T. scotoductus*
1373	*T. thermophilus*	*T. thermophilus*	**2214**	-	*T. brockianus*
2100	*T. scotoductus*	*T. scotoductus*	**2443**	*T. igniterrae*	*T. igniterrae*
2101	*T. scotoductus*	*T. scotoductus*	**2631**	*T. scotoductus*	*T. scotoductus*
2102	*T. scotoductus*	*T. scotoductus*	**2788**	*T. brockianus*	*T. brockianus*
2103	*T. brockianus*	*T. brockianus*	**2789**	*T. sp.*	*T. brockianus*
2104	*T. igniterrae*	*T. igniterrae*	**2790**	*T. igniterrae*	*T. igniterrae*
2105	*T. igniterrae*	*T. igniterrae*	**2791**	*T. igniterrae*	*T. brockianus*
2106	*T. brockianus*	*T. brockianus*	**2792**	*T. igniterrae*	*T. igniterrae*
2107	*T. igniterrae*	*T. igniterrae*	**2793**	*T. igniterrae*	*T. igniterrae*
2108	*T. scotoductus*	*T. scotoductus*	**2795**	*T. sp.*	*T. brockianus*
2109	*T. scotoductus*	*T. scotoductus*	**2797**	*T. brockianus*	*T. brockianus*
2110	*T. igniterrae*	*T. igniterrae*	**2811**	*T. scotoductus*	*T. scotoductus*

3.4 Phylogenetic classification

16S rRNA sequences were amplified from all the *Thermus* strains used in the study. As expected, all the examined strains belong to *Thermus* according to the SSU sequence analysis. Most of the Blast results (≥ 98 % identity to the database reference) correlate with the former categorization of Hreggvidsson et al. (2006) and listed in Table 3. *T. thermophilus*, *T. scotoductus*, *T. igniterrae* and *T. brockianus* seem to be subdivided into distinct genetic clusters, as already mentioned in Hreggvidsson et al. (2006). The *T. sp.* cluster can be classified as *T. brockianus* with a Blast results identity of >98 %, which is in accord with the definition of species as proposed by Stackebrandt et al. (2006).

3.5 The distribution of the Thermophi gene

Evidence suggests that the gene for the Thermophi polymerase is present in some strains of all examined *Thermus* species except *T. aquaticus* and *T. oshimai*. The observed ratio of positive to negative strains within a genus is on average □=25 % (*T. thermophilus* 21 %, *T. scotoductus* 15 %, (*T. antranikianus* 100 %, but not statistical significant), *T. igniterrae* 28 %, *T. brockianus* 35 %).

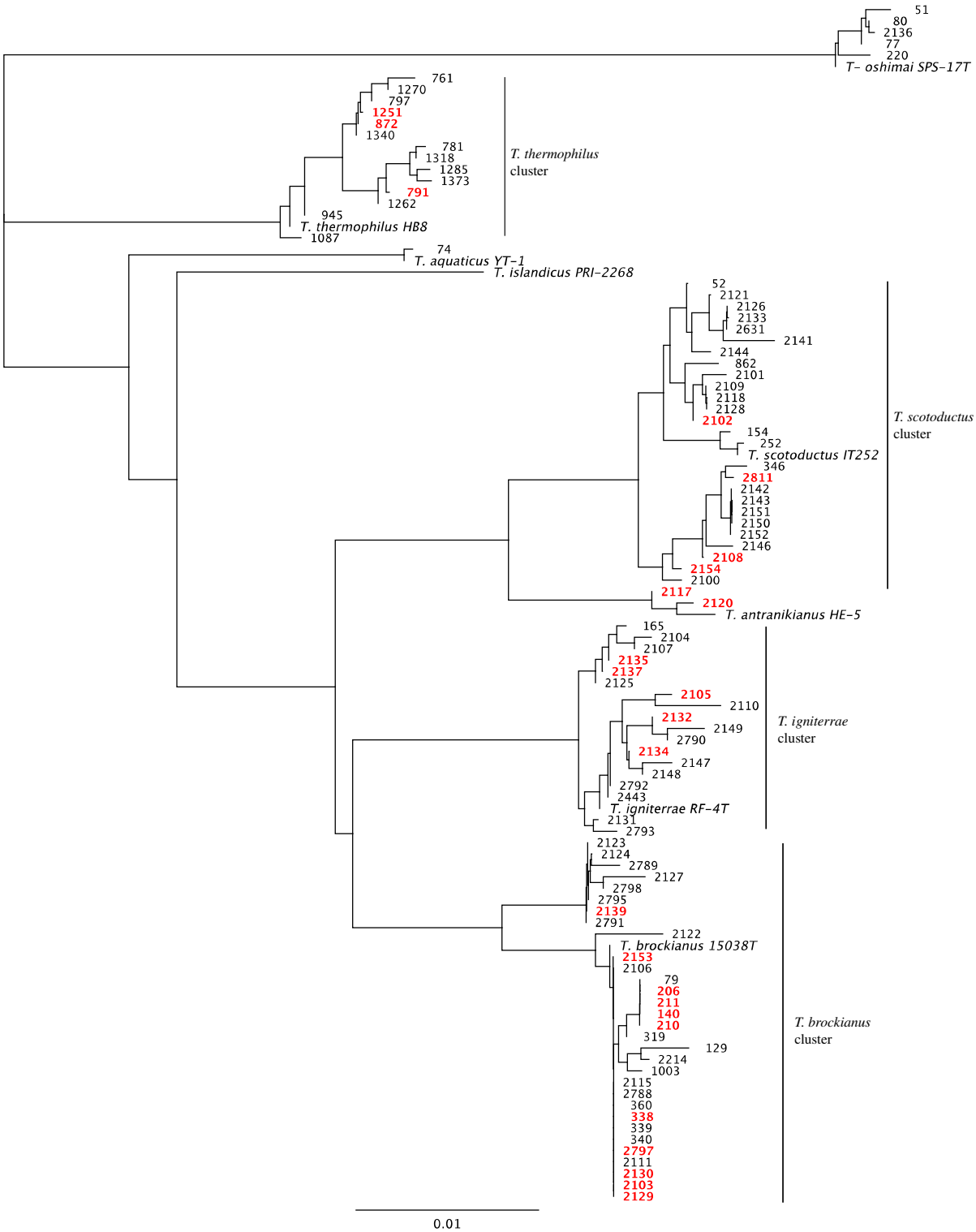

Figure 7. Phylogenetic relationship of the examined strains. The strains marked red contain the Thermophi polymerase gene according to the PCR screening.

4. Discussion

4.1 Thermophi-gene

The Thermophi PCR screening results mapped on the phylogenetic distance tree indicates a widespread presence of the Thermophi-gene within the *Thermus* genus. However, the results show that the distribution is clearly discontinuous, as it is not found in all strains within any of the *Thermus* species. The sequence analysis of selected PCR-products confirms the presence of the *Aquificae* like polymerase in the strains, showed to be positive by PCR screening. This does not exclude the possibility that the gene is present in strains shown to be negative by PCR screening. Additional evidence for the sporadic distribution is best acquired with southern blot studies.

4.2 Transposon

In the study almost all strains (88%), which were believed to possess the Thermophi polymerase shown by PCR amplification, also contain the transposase gene in close proximity. This indicates, but does not prove that the gene is a transposon-encoded polymerase. The size of the obtained PCR products, the region between the polymerase and the transposase, was not uniformly 1,6 kbp in length. Five strains seemed to have smaller PCR products (140, 206, 210, 211, 2130). Interestingly, they all belong to *T. brockianus* what might indicate a structural difference and possibly a species originating from a different geographical region than the bulk of the Icelandic *Thermus* species/strains. It could be reflected in fewer repeats and other sequence differences.

4.3 Phylum distribution of *Aquificae* like PolI polymerases

Polymerase amino acid sequence relation analysis in Figure 8 indicates that the Thermophi polymerase may have a genetic origin outside of the phylum and might have been acquired by lateral gene transfer. It shows that the Thermophi polymerase clusters with the distinct group of *Aquificae* like polymerases (Family A), the main PolI polymerase in the phylum of *Aquificae*, together with similar polymerases that have sporadic distribution in other bacterial phyla. They are even found in cell organelles of some eukaryotic protozoa species. The *E. coli* like PolI C sequences more or less reflect genetic relationships obtained by 16S rRNA analysis. Thus the primary PolI C of *Thermus* groups together with the PolI from related species, as expected. The *Aquificae* like polymerases are not as universal as the PolI C polymerases and the phylogenetic relationships disagree with phylogenetic relationships based on 16S rRNA sequences between the host species.

This may support the hypothesis that this type of PolI polymerases moves by lateral transfer across species and phylum boundaries. It would be interesting to study the flanking regions of these polymerases for the evidence of transposase genes or transposon like sequence features.

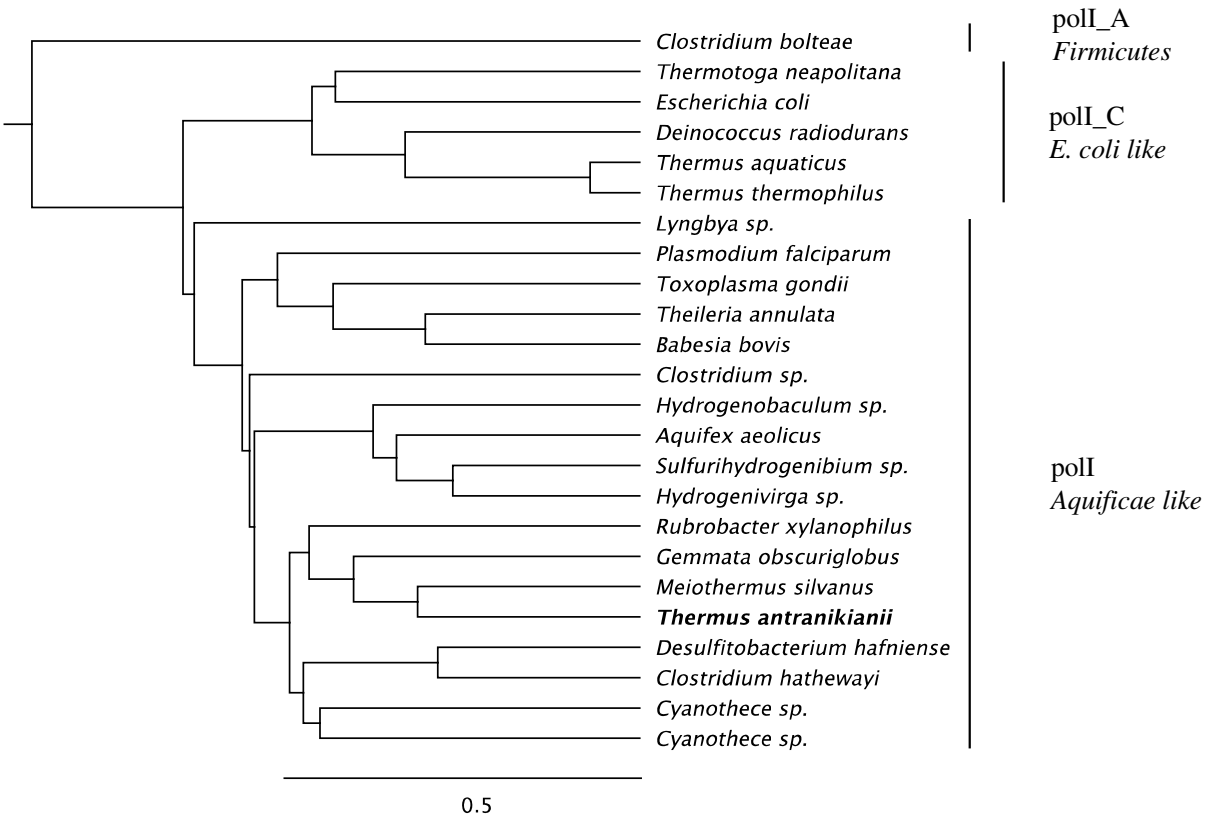

Figure 8. Genetic distance tree of type I polymerases (Family A). All three different types form an own cluster. The PolI A cluster represents the polymerase I of *Firmicutes*, which was used here as a distinct out-group. PolI C is the most common Family A polymerase, including the *Deinococcus-Thermus* phylum. The Thermophi polymerase is located within the PolI *Aquificae* like cluster.

4.4 Conclusions

The hypothesis for this study was that the Thermophi polymerase is a transposon encoded polymerase and can move between strains and species by lateral gene transfer. This investigation strongly supports the hypothesis. The gene has a sporadic distribution within the *Thermus* genus and is associated with transposon like sequences including a transposase gene. It also shows conserved sequence features of a special type of PolI polymerase that appears to be the main PolI polymerase in the phylum *Aquificae*. In addition this polymerase shows sporadic distribution in other bacterial phyla.

Thermophilic organisms are restricted to small geothermal habitats on a global scale. Surrounding physicochemical barriers hinder their distribution and may delay migration to distant habitats. For instance, *T. aquaticus* has only been found on the North American continent whereas *T. igniterrae* seems to be present only in Iceland and Australia. *T. brockianus* has a cosmopolitan distribution. In this case it is noteworthy that the genotypic variability of *T. brockianus* isolates is very low compared to other Icelandic species, perhaps indicating a recent migration to Iceland (Hreggvidsson et al. 2006). In this context it is interesting to note that the putative Thermophi-transposon sequence might have a different structure from the corresponding sequence found in other Icelandic *Thermus* species. It would be also interesting in this context to analyze the presence/absence pattern of this sequence-region and its structure in North American *T. brockianus* isolates. Similar, further studies would benefit from including a number of *T. aquaticus* strains from North America in order to

investigate possible geographic influences on the presence/absence pattern of this transposon and its structure.

However, there are still other *Thermus* species, from even more distant regions, like *T. filiformis* (only found in New Zealand) and *T. yunnanensis* (only found in China). It is clearly of interest to investigate also these apparently geographically isolated species in relation to the presence of the Thermophi gene.

5. References

Berezovsky, I., & Shakhnovich, E. (2005). Physics and evolution of thermophilic adaptation. *Proceedings of the National Academy of Sciences of the United States of America , 102* (36), 12742-7.

Brock, T., & Freeze, H. (1969). Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile. *Journal of bacteriology , 98* (1), 289-97.

Chien, A., Edgar, D., & Trela, J. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *Journal of bacteriology , 127* (3), 1550-7.

Degryse, E., Glansdorff, N., & Piérard, A. (1978). A comparative analysis of extreme thermophilic bacteria belonging to the genus Thermus. *Archives of microbiology , 117* (2), 189-96.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2010) Geneious v5.1, Available from http://www.geneious.com

Dyson (1999). Isolation and development of transposons. *Method Microbiol* vol. 29 pp. 133-167

Finn, R., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J., et al. (2010). The Pfam protein families database. *Nucleic acids research , 38* (Database issue), D211-22.

Gibbs, M., Reeves, R., Mandelman, D., Mi, Q., Lee, J., & Bergquist, P. (2009). Molecular diversity and catalytic activity of Thermus DNA polymerases. *Extremophiles , 13* (5), 817-26.

Haki, G., & Rakshit, S. (2003). Developments in industrially important thermostable enzymes: a review. *Bioresource technology , 89* (1), 17-34.

Hreggvidsson, G., Skirnisdottir, S., Smit, B., Hjorleifsdottir, S., Marteinsson, V., Petursdottir, S., et al. (2006). Polyphasic analysis of Thermus isolates from geothermal areas in Iceland. *Extremophiles , 10* (6), 563-75.

Kashefi, K., & Lovley, D. (2003). Extending the upper temperature limit for life. *Science (New York, NY) , 301* (5635), 934.

Mahillon, J., & Chandler, M. (1998). Insertion sequences. *Microbiology and molecular biology reviews : MMBR , 62* (3), 725-74.

Nevers, P., & Saedler, H. (1977). Transposable genetic elements as agents of gene instability and chromosomal rearrangements. *Nature , 268* (5616), 109-15.

Rothschild, L., & Mancinelli, R. (2001). Life in extreme environments. *Nature , 409* (6823), 1092-101.

Sayers, E., Barrett, T., Benson, D., Bolton, E., Bryant, S., Canese, K., et al. (2010). Database resources of the National Center for Biotechnology Information. *Nucleic acids research , 38* (Database issue), D5-16.

Skirnisdottir, S., Hreggvidsson, G., Hjörleifsdottir, S., Marteinsson, V., Petursdottir, S., Holst, O., et al. (2000). Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. *Applied and environmental microbiology , 66* (7), 2835-41.

Stackebrandt, E. & Ebers, J. (2006). Taxonomic parameters revisited: tarnished gold standards. *Microbiol. Today* 33: 152–5.

Stetter, K. (1996). Hyperthermophiles in the history of life. *Ciba Foundation symposium , 202*, 1-10; discussion 11-8.