EMbaRC

European Consortium of Microbial Resource Centres

Grant agreement number: 228310

Seventh Framework Programme

Capacities

Research Infrastructures

Combination of Collaborative Project and Coordination and Support Actions

Deliverable D15.19 (formerly D.JRA2.1.1)

Title: Set of new genes to use in addition or replacement of ribosomal sequences

Due date of deliverable: M24

Actual date of submission: M39

Start date of the project: 1st February 2009 **Duration:** 44 months

Organisation name of the lead beneficiary: KNAW-CBS

Version of this document: V0.1

Dissemination level:

PU	**Public**	**X**
PP	Restricted to **other programme participants** (including the Commission)	
RE	Restricted to a **group defined by the Consortium** (including the Commission)	

EMbaRC is financially supported by the Seventh Framework Programme (2007-2013) of the European Communities, Research Infrastructures action

Document properties	
Project	EMbaRC
Workpackage	JRA 2
Deliverable	Deliverable D15.19 (formerly D.JRA2.1.1)
Title	Set of new genes to use in addition or replacement of ribosomal sequences
Version number	V0.1
Authors	Benjamin Stielow, Ursula Eberhardt, Vincent Robert, Gerard Verkley, Joost Stalpers (KNAW-CBS), Laurence Lesage-Meessen (INRA-CIRM)
Abstract	Purpose of this part of the work in JRA 2.1 was to identify alternative candidate marker gene targets for the fungal kingdom, from available fully sequenced genomes using a taxonomy-aware processing *in silico* pipeline, followed by wet testing to evaluate the practical suitability of the targeted regions. A number of different *in silico* approaches were tested and already reported on in published papers of Robert *et al.* (2011) and Lewis *et al.* (2011). For *in silico* analysis, genomes of 74 fungi were downloaded from the several institutions such as *Broad* or *Genoscope*. A total of 54 alignments from the used genomes flanked by suitable primer pairs was used as input for the secondary primer design as described below. The wet work was done in two trial phases. *Trial phase I* was conducted with a set of 15 DNA extracts from the main clades of the fungal kingdom, represented by genome tested strains and species as far as available. Aim of *trial phase I* was to primarily identify well performing primer sets and *secondly*, at the same time, assessing their versatility, i.e. their ability to amplify the targeted gene from various fungal genera. *Trial phase II* was conducted with a larger set of DNA extracts (48) to decrease the taxon bias and to increase the reliability and consistency of the entire experiment. Some of the gene regions that were retrieved during the *in silico* work are indeed novel, i.e. have not been formerly used in fungal systematics or phylogeny, and lab results here obtained show that in particular the regions of the Gamma actin gene (alignment group 2), 60S ribosomal protein L10a (52) and Elongation factor 1-alpha (33) are accessible with reasonably universal primers delivering a single PCR-product. In addition, we demonstrate that it is possible to identify novel primer pairs that have similar amplification efficiency like ribosomal non-protein coding genes, such as ITS, but most likely provide a superior phylogenetic resolution based on our *in silico* based estimations
Validation process	Document prepared by KNAW-CBS in collaboration with INRA-CIRM and submitted to the Executive Committee for agreement.

Contents

Abbreviation key

BPPM	Best Pair of Primers Method
BRC	Biological Resource Centre
ILM	Ideal Locus Method
ITS	Internal Transcribed Spacer (region of the ribosomal RNA gene)
PCR	Polymerase Chain Reaction
T*a*	Annealing temperature

1 Background and Objectives

Identification and classification of eukaryotic organisms increasingly depends on DNA sequences retrieved from standardized genetic markers. **We attempted to identify alternative candidate marker gene targets for the fungal kingdom, from available fully sequenced genomes using a taxonomy-aware processing *in silico* pipeline**. In tandem with an international research project funded by the Sloan foundation, a number of different *in silico* approaches were tested (Lewis *et al.* 2011, Robert *et al.* 2011). Below we report on the results of the EMbaRC-funded approach which resulted in a higher number of primer and gene region candidate suggestions.

The choice of the ideal gene to sequence was based on five major criteria: its presence in all organisms to be studied, ease of PCR amplification and sequencing, its supposed evolutionary rate, and the absence of pseudogenes, paralogs or orthologs that could complicate amplification and analysis, and finally its relevancy to phylogeny and species identification. Putative degenerated primer pairs were designed on the basis of alignments less than 1kb in size, to retrieve a broad overlap across the entire fungal kingdom and to confirm the *in silico* analysis results of Robert *et al.* (2011) in the laboratory. As it was reported by Robert *et al*. (2011), the target genes identified in the *in silico* analyses were analyzed and shown to have a better resolution potential to differentiate species than most presently used ribosomal and protein coding markers.

2 Material and methods

2.1 Selection of new marker genes by in silico analyses of available eukaryote genomes (ML.JRA2.1.1 - work published in Robert et al. 2011)

For this analysis, genomes of 74 fungi were downloaded from the several institutions such as *Broad* or *Genoscope*. The initial approach to find ideal genes was the Ideal *Locus* Method (ILM). Such a *locus* would provide a phylogeny as close as possible to the whole genome phylogeny and would distinguish distantly and closely related species. From our results obtained with ILM, it was clear that finding very good *loci* was possible but that PCR amplification would be extremely problematic to implement. The second approach, called *Best Pair of Primers Method* (BPPM), was devised to identify short conserved regions at a suitable distance from each other that could be used as initials to construct forward and reverse primers for any fungi, or a selected taxonomic group of fungi. The

regions between any two primer initials were retrieved, aligned and their suitability as reliable phylogenetic representatives and/or as potential barcode candidates for identification was assessed.

PCR primers are normally 18-24 bp in length, but in practice only the fit of the 5' end of a primer is decisive for the PCR success. In addition, the most important consideration in primer design is to have a very well conserved section on the 3' end of the primer while the 5' end can show some variability without necessarily hampering the amplification ability of the primer. Hence, an alternative method called the "Short Primer Method" was developed. In summary, the following six steps were executed:
1. Assembling a list of all possible primers of 12 nucleotides. *2.* Searching for the presence of these 12mer primers in all genomes. *3.* Keeping primers that have a good quality (in terms of PCR applicability) and are present in most species.. *4.* Using these best primers to search for possible primer pairs. *5.* Using the best primer pairs to extract the intermediate DNA sequence. The target length of the intermediate sequence was set to 200 to 1000 nucleotides. These numbers were chosen to ensure a minimum variability and to allow for ease of amplification and Sanger sequencing. *6.* Compare the amplified DNA of the different species with each other and build distance matrices and trees for comparison with a reference matrix and tree. This last step is to ensure that the selected DNA region would produce a relatively coherent phylogeny but would also be suitable for discriminating closely related species.
Finding primer candidate pairs with the potential to amplify DNA for species belonging to completely distantly related groups of organisms and providing high phylogenetic information content seems very difficult with limited computation resources. Reducing the species number or reducing the diversity of the species involved, increase significantly the coverage of the primer pairs. In the global search for all true fungi, only seven primer pairs were covering 72 species or more out of the 74 species studied. For the ascomycetes, for example, six primer pairs were found and compatible with all 57 species studied and 186 primer pairs with 56 species via the *in silico* approach.
A total of 54 alignments from the used genomes flanked by suitable primer pairs was used as input for the secondary primer design as descibed below.

2.2 Methodical approach to secondary primer design and laboratory work

The extracted gene regions corresponding to a total of 54 alignments were manually re-

annotated by using *n-* and *xBLAST*. While most ‘alignments’ were re-annotated as elongation factor 1 alpha, only a small set (8 of 54) could be identified as putatively novel candidate regions previously unknown to phylogeny and DNA barcoding of fungi. Since the entire elongation factor 1 locus is already well known to give superior subordinate taxon resolution and contains sufficient primer coverage, priority was given to those gene regions that have no track record in fungal phylogeny to date.

Secondary primer design. To construct PCR primers, the alignments were imported into *Bioedit 7.0.52* and manually screened for the most conserved sites in the available set of taxa to identify suitable primer binding sites. We gave high priority to stable 3’ primed ends with at least eight nucleotide binding sites and the least possible quantity of degeneracy. All primers developed in the secondary primer designed are listed in Table 1. For a number of these primers we replaced all *N*s (corresponding either to G, A, T or C) with *inosit* to improve the binding stability, also included in Table 1 here.

However, the number of conserved sites between any significant number of taxa was relatively low but still resulted in a reasonable number of primers (pairs), often located adjacent to each other with only a few bases difference shifted to either the 5’ or 3’ primed end. To allow aspecific binding, annealing temperatures during PCR (*trial phase I* and *II*) were set below 50 C° to identify the most universal set of primers and those which would result in none or (too many) multiple fragments.

Wet testing of designed primers. The wet work was done in two trial phases. *Trial phase I* was conducted with a small set of DNA extracts covering *very roughly* the entire fungal kingdom, represented by genome tested strains and species as far as available (Annexes, table 4). Aim of *trial phase I* was to primarily identify well performing primer sets and *secondly*, at the same time, assessing their versatility, i.e. their ability to amplify the targeted gene from various fungal genera*Trial phase II* was conducted with a 48 DNA extracts (Annexes table 7) to decrease the taxon bias and to increase the reliability and consistency of the entire experiment.

Reagents used were standardized, using enzymes and dNTPs from Takara (Japan), oligonucleotides synthesized by Integrated DNA Technologies (Netherlands) and PCR reactions were run on gold coated thermo cyclers, SensoQuest (Germany).

Primer name	Primer sequence	Orientation
Al2_14_31_F1	TAYCCHATYGAGCACGG	Forward
Al2_41_63_f2	AACTGGGAYGACATGGAGAAGAT	Forward
Al2_168_184_f3	GTNGCNCCHGAGGAGCA	Forward
Al2_168_184_f3i	GTIGCICCHGAGGAGCA	Forward
Al2_600_583_R1	AGCTTCTCCTTGATGTC	Reverse
Al2_553_536_R2	GANAGNGTANCCDCGCTC	Reverse
Al2_553_536_R2i	GAIAGIGTAICCDCGCTC	Reverse
Al4_47_63_F1	AYYTCRWAGTACTTGTAGGT	Forward
Al4_73_91_F2	GTTRAYCCAGTANGAGTT	Forward
Al4_73_91_F2i	GTTRAYCCAGTAIGAGTT	Forward
Al4_424_406_R1	TVGGVTACAAGGCCAAGCA	Reverse
Al4_440_442_R2_tail	GCTATCATCACAATGGACGAYAAGGCT	Reverse
AlGr5_412-433_F1	CTTVAVYTGGAACTTGATGGT	Forward
Algr5_408_420_F_tail	GCTATCATCACAATGGACCCTTCTTVAVYTGG	Forward
Algr5_1102_1084_R1	GHGACAAGCGTTTCTCNGG	Reverse
Algr5_742_730_r_tail	GCTATCATCACAATGGACAAYAAGAACAAGAA	Reverse
Al13_49_67_F1	CGHCTNCACTTCTTCATGG	Forward
Al13_49_67_F1i	CGHCTICACTTCTTCATGG	Forward
Algr13_552-545_JZHtail_R	GCTATCATCACAATGGACGTRTACCA	Reverse
Al14_64_80_F1	TYNGACATATGCTTGTC	Forward
Al14_64_80_F1i	TYIGACATATGCTTGTC	Forward
Al14_69_80_F_tail	GCTATCATCACAATGGACCATATGCTTGTC	Forward
Al14_429_414_R1	YAGATGACGAGTCHKA	Reverse
Al14_369_360_R_tail	GCTATCATCACAATGGACCCATTCATGC	Reverse
Al33_54_73_F1	GAYTTCATCAAGAACATGAT	Forward
Al33_129_148_F2	GARTTYGARGCYGGTATCTC	Forward
Al33_890_871_R1	GAVACRTTCTTGACGTTGAA	Reverse
Al33_879_859_R2	GACGTTGAADCCRACRTTGTC	Reverse
Al46_169-186_F1a	CATCTCGTCCATDCCCTC	Forward
Al46_166-187_f1b	GAAATTCAGACCCACCCCTC	Forward
Al46_191-198_JZHtail_f	GCTATCATCACAATGGACGTGTACCA	Forward
Al46_693-675_R1	CGHYTNCACTTCTTCATG	Reverse
Al46_693-675_R1i	CGHYTICACTTCTTCATG	Reverse
Al46_687_677_jzhtail_R	GCTATCATCACAATGGACCACTTCTTCAT	Reverse
Al49_ 44_63_F1	CYGGHGGHTGGAAGATGAAG	Forward
Al49_112_129_F2	GAYGARCCNACYAACCA	Forward
Al49_112_129_F2i	GAYGARCCIACYAACCA	Forward
Al49_846_829_r1	TCRTAVSWGTTCTTGAAC	Reverse
Al49_838-827_jzhtail_r	GCTATCATCACAATGGACGTTCTTGAACTT	Reverse
AlGr52_412-433_f1	CTTVAVYTGGAACTTGATGGT	Forward
Algr52_422_412_jzhtail_F	GCTATCATCACAATGGACAACTTGATGGT	Forward
Algr52_1102_1084_R1	GHGACAAGCGTTTCTCNGG	Reverse
Algr52_1102_1084_R1i	GHGACAAGCGTTTCTCIGG	Reverse
Algr52_745_737_jzhtail_r	GCTATCATCACAATGGACAAGAACAAG	Reverse
Algr5_1102_1084_R1i	GHGACAAGCGTTTCTCIGG	Reverse
JZH	GCTATCATCACAATGGAC	Forward/reverse

Table 1: Designed primers for wet testing.

3 Results

In the wet work we tested a set of 51 primers corresponding to 8 distinct regions listed in Table 2 for their ability to amplify easily, reliable and consistent a short standardized region from a predefined gene region resulting in a single PCR fragment.

Trial phase I

Within *trial phase I* we were able to identify from multiple primer sets, a single set each corresponding to one of the regions highlighted in yellow (Table 1) that directly gave single PCR products with a low annealing temperature (T*a*). Thus, PCR conditions could be eventually further optimized. Since PCR reactions with a number of primer pairs containing an *inosit,* instead of all four bases (= N), PCR resulted in almost no signal at all, and were contrasted by far better results using exactly the same sequence but a substitution by an *N* (= A,T,C,G), we decided to continue during *trial phase I* only with *N* multiplex primers.

Trial phase II

During *trial phase II*, the T*a* was not changed but instead the set of extracts expanded, and a subset of seven primers selected that gave the most optimal results during *trial phase I*. These primer sets were identified to perform reliable and consistent for a broad range of fungal species, many of them amplifying the desired single fragment. In addition, some primer sets amplified multiple fragments, likely due to low *Ta*. Roughly, the best working primer sets for alignment group 2 (Gamma actin), 33 (Elongation factor 1 alpha) and 52 (60S ribosomal protein L10a family) directly gave a strong single signal with almost no secondary fragments (likely due to low *Ta*). Those primers designed for group 46 (Beta tubulin 2) and 49 (Elongation factor 3) amplified multiple fragments, but gave a strong primary signal. Most problematic appeared to be the amplification of a gene corresponding to a small RNA processing protein, a putative S6 kinase, which resulted in a number of intense signals. The best performing gene regions are listed in Table 2. The approach of tailed primer design used in this study proved successful in retrieval in PCR of even notriously difficult gene candidates such as Elongation factor 1, which has thus far not been amplifiable via known standard primers with the same efficiency across the whole fungal kingdom.

Alignment group	Gene	Performance in wet testing
2	Gamma actin gene	strong single signal, very weak secondary signal
33	Elongation factor 1-alpha	
52	60S ribosomal protein L10a	
46	Beta-tubulin 2	intense primary signal, but weak secondary signal
49	Elongation factor 3	
14	S6 kinase (RNA processing prot.)	multiple signals of strong intensity
4	60S ribosomal protein L15	no or very poor signal
13	Beta tubulin 2	

Table 2: Best performing gene regions. The numbers of the alignment groups correspond to Table 9 in Robert et al. (2011).

4 Conclusions

The team could not keep up with time-line set in the DOW (Mo 12 from start of EMbaRC) because there was a delay of 6 months in the start of programming required for the *in silico* analyses, and unforeseen problems with computing methodologies for retrieving the target regions. Nonetheless, the planned milestone (ML.JRA2.1.1) was achieved and the results of that were published already in 2011 (Robert *et al.* 2011). The *in silico* approach also met its goal to provide a workable basis for the wet lab-testing phase designed in JRA2.1, which then started with about a 12 mo delay. On this wet work is reported in the current deliverable for the first time, and a scientific manuscript is expected to be written before the end of 2012. Some of the gene regions that were retrieved during the *in silico* work are indeed novel, i.e. have not been formerly used in fungal systematics or phylogeny, and lab results obtained show that in particular the regions of the Gamma actin gene (alignment group 2), 60S ribosomal protein L10a (52) and Elongation factor 1-alpha (33) are accessible with reasonably universal primers delivering a single PCR-product. In addition, we demonstrate that it is possible to identify novel primer pairs that have similar amplification efficiency like ribosomal non-protein coding genes, such as the recently ratified universal fungal barcode marker ITS (Schoch *et al.* 2012), but most likely provide a superior phylogenetic resolution based on our in silico based estimations. Whereas universality is a desirable property in fungal phylogeny, it is one of the basic requirements for any DNA barcoding locus.Efforts in the mycological community to select secondary loci to supplement the ITS barcode for fungi have not been concluded. The results of EMbaRC JRA2.1, are a timely contribution to the discussion.

5 Outlook

In the era of advanced massive parallel genome sequencing, the quest to identify a reliable barcode for the entire fungal kingdom is closer than ever in recent years to succeed. The question, if a single gene will fulfill all expectations, in other words will: *i)* be easy amplifiable *ii)* provides sufficient phylogenetic resolution for various taxonomic hierarchical ranks (in particular on species level) *iii)* provides sufficient bootstrap support on both, basal and deeper internal branches and *iv)* fits most accurately a topology based on all overlapping orthologous genes (whole genome phylogeny) still needs to be evaluated. However, our approach demonstrated that a sophisticated *in silico* analysis, based on available fungal genomes is capable in identifying such putative candidate genes, potentially useful to test this hypothesis. The essential physiological functions within eukaryotic cells of the targeted genes already suggest a ubiquitous distribution across the fungal kingdom, and with that, a likely (partially) conserved nucleotide sequence across various fungal taxa; a basis for designing universal barcode primers. Based on the work conducted within the EMbaRC JRA2 workpackage, we are considering next steps for continued research, in collaboration with interested parties including EMbaRC partners.

These steps are:

- Verification of all PCR products by Sanger sequencing.
- Further optimizing PCR conditions (Ta gradients).
- Redesigning a small number of primers resulting in secondary products based on sequenced PCR products.
- A more detailed evaluation of the potential of each targeted (partial) gene, compared to currently widely used loci for its potential as a discriminative reliable fungal barcode on various taxonomic ranks.
- Continued testing of a subset of the best performing barcode primers (based on the previous step) within various fungal taxa, in the ideal case on multiple diverse sets closely related species where most other conventional loci are already sequenced.
- Evaluating the full potential phylogenetic resolution by testing more taxa.
- Evaluating inter- and intraspecific variation for selected taxa.

References

Robert V, Szöke S, Eberhardt U, Cardinali G, Meyer W, Seifert K, Lévesque A, Lewis CT (2011). The quest for a general and reliable fungal DNA barcode. The Open Applied Informatics Journal, 5: 45-61.

Lewis CT, Bilkhu S, Robert V, Eberhardt U, Szöke S, Seifert K, Lévesque A (2011). Identification of fungal DNA barcode targets and pcr primers based on Pfam protein families and taxonomic hierarchy. The Open Applied Informatics Journal, 5: 30-44.

Schoch *et al*. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. PNAS Early edition DOI: 10.1073/pnas.1117018109.

Annexes

Primer sets

Table 3: Primer combinations for eight selected alignments (See Material and Methods).

Aligned group 2 (Gamma actin gene)	**Aligned group 4 (60S L 15b)**	**Aligned group 13 (beta tubulin 2)**	**Aligned group 14 (S6 kinase)**
1: Al2_14_31_F1 + Al2_600_583_R1	13: Al4_47_63_F1 + Al4_424_406_R1	25: Al13_49_67_F1 + Algr13_552	27: Al14_64_80_F1 + Al14_429_414_R1
2: Al2_14_31_F1 + Al2_553_536_R2	14: Al4_47_63_F1 + Al4_440_442_R2_tail	26: Al13_49_67_F1i + Algr13_552	28: Al14_64_80_F1 + Al14_369_360_R_tail
3: Al2_14_31_F1 + Al2_553_536_R2i	15: Al4_73_91_F2 + Al4_424_406_R1		29: Al14_64_80_F1i + Al14_429_414_R1
4: Al2_41_63_f2 + Al2_600_583_R1	16: Al4_73_91_F2 + Al4_440_442_R2_tail		30: Al14_64_80_F1i + Al14_369_360_R_tail
5: Al2_41_63_f2 + Al2_553_536_R2	17: Al4_73_91_F2i + Al4_424_406_R1		31: Al14_69_80_F_tail + Al14_429_414_R1
6: Al2_41_63_f2 + Al2_553_536_R2i	18: Al4_73_91_F2i + Al4_440_442_R2_tail		32: Al14_69_80_F_tail + Al14_369_360_R_tail
7: Al2_168_184_f3 + Al2_600_583_R1			
8: Al2_168_184_f3 + Al2_553_536_R2			
9: Al2_168_184_f3 + Al2_553_536_R2i			
10: Al2_168_184_f3i + Al2_600_583_R1			
11: Al2_168_184_f3i + Al2_553_536_R2			
12: Al2_168_184_f3i + Al2_553_536_R2i			

Table 3 (continued): Primer combinations for eight selected alignments (See Material and Methods).

Aligned group 33 (elongation factor 1-alpha)	**Aligned group 46 (beta tubulin2)**	**Aligned group 49 (Elongation factor 3)**	**Aligned group 52 (60S L10a)**
33: Al33_54_73_F1 + Al33_890_871_R1	37: Al46_169-186_F1a + Al46_693-675_R1	46: Al49_ 44_63_F1 + Al49_846_829_r1	52: AlGr52_412-433_f1 + Algr52_1102_1084_R1
34: Al33_54_73_F1 + Al33_879_859_R2	38: Al46_169-186_F1a + Al46_693-675_R1i	47: Al49_ 44_63_F1 + Al49_838-827_jzhtail_r	53: AlGr52_412-433_f1 + Algr52_1102_1084_R1i
35: Al33_54_73_F2 + Al33_890_871_R1	39: Al46_169-186_F1a + Al46_687_677_jzhtail_R	48: Al49_112_129_F2 + Al49_846_829_r1	54: AlGr52_412-433_f1 + Algr52_745_737_jzhtail_r
36: Al33_54_73_F2 + Al33_879_859_R2	40: Al46_166-187_f1b + Al46_693-675_R1	49: Al49_112_129_F2 + Al49_838-827_jzhtail_r	55: Algr52_422_412_jzhtail_F + Algr52_1102_1084_R1
	41: Al46_166-187_f1b + Al46_693-675_R1i	50: Al49_112_129_F2i + Al49_846_829_r1	56: Algr52_422_412_jzhtail_F + Algr52_1102_1084_R1i
	42: Al46_166-187_f1b + Al46_687_677_jzhtail_R	51: Al49_112_129_F2i + Al49_838-827_jzhtail_r	57: Algr52_422_412_jzhtail_F + Algr52_745_737_jzhtail_r
	43: Al46_191-198_JZHtail_f + Al46_693-675_R1		
	44: Al46_191-198_JZHtail_f + Al46_693-675_R1i		
	45: Al46_191-198_JZHtail_f + Al46_687_677_jzhtail_R		

Trial I

Initial test phase with 15 different fungal species (extracts).

Selected species:

Table 4: Selected taxa for trial I.

CBS No.	Extract	Taxon
CBS 109036	32190	*Trichophyton equinum*
CBS 115846	32193	*Cryphonectria parasitica*
CBS 115943	32194	*Mycosphaerella graminicola*
CBS 123668	32198	*Fusarium oxysporum f.sp. lycopersici*
CBS 127170	32201	*Verticillium dahliae*
CBS 142.95	32204	*Trichoderma atroviride*
CBS 1954	32206	*Candida parapsilosis var. parapsilosis*
CBS 405.96	32213	*Schizophyllum commune*
CBS 445.79	32214	*Laccaria bicolor*
CBS 513.88	32216	*Aspergillus niger*
CBS 708.71	32222	*Neurospora crassa*
CBS 818.72	32223	*Aspergillus oryzae var. brunneus*
CBS 277.49	32226	*Mucor circinelloides f. Lusitanicus*
FGSC 9596	32233	*Nectria haematococca*
FGSC 10004	32235	*Phycomyces blakesleeanus*
Negative control	negative control	water

Selected alignment groups for Trial I

Alignment group 2 (Gamma actin gene)

Technical notes:

- Samples were diluted 1:10 prior using DNA sample for PCR.

PCR cycle

7 min 95°

1 min 95°

1 min 42° → 52°C

(+0.5° C/cycle)

2 min 72°

40 X

10 min° 72

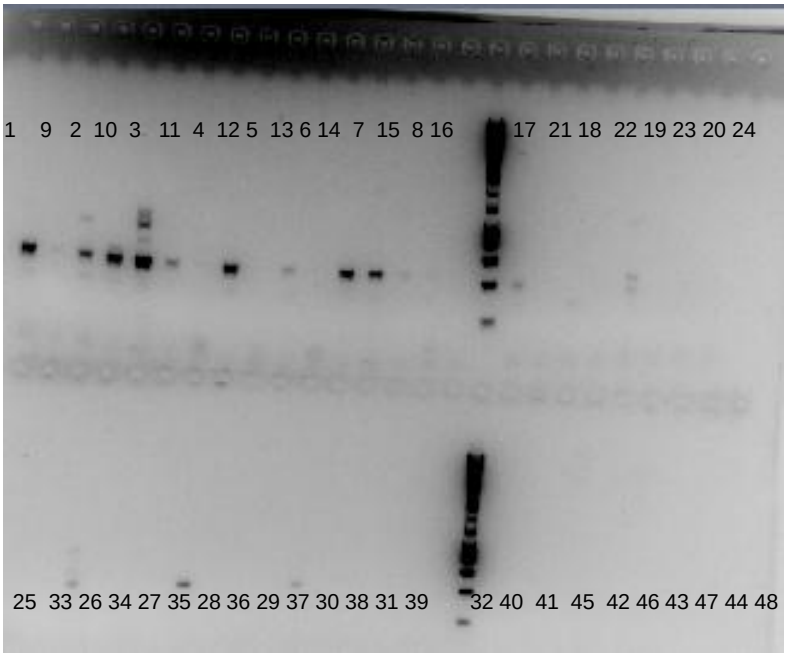


Figure 1: Results for Gamma actin (alignment group 2).

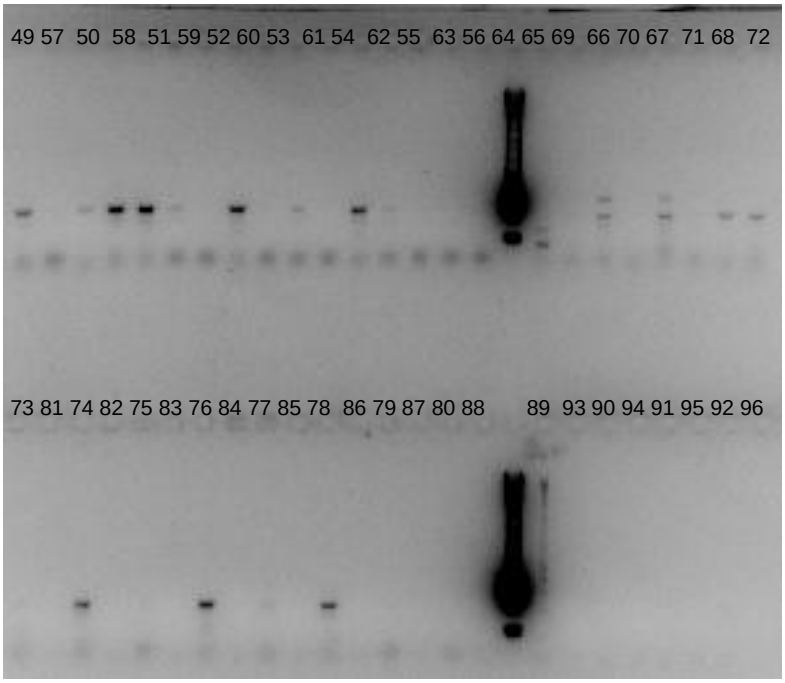


Figure 2: Results for Gamma actin (alignment group 2); continued.

Sample 1-16: Primer set 1 (Table 3, set 1)

Sample 17-32: Primer set 2 (Table 3, set 2)

Sample 33-48: Primer set 3 (Table 3, set 4)

Sample 49-64: Primer set 4 (Table 3, set 5)

Sample 65-80: Primer set 5 (Table 3, set 7)

Sample 81-96: Primer Set 6 (Table 3, set 8)

- Degenerated primers substituted with *inosit* instead of N (= all four bases) did result in no PCR products and were subsequently neglected. (It could be potentially interesting to further test *inosit* based primers with an improved very slow ramp rate).

*Table 5: Results gamma actin amplification. Columns indicate primer combinations (**Page 19**), rows indicate taxa (**Table 4**). Symbols: + indicates strong signal, +- faint signal, - poor or no signal.*

Samples	1	2	3	4	5	6
1	+	+	-	+	+	-
2	+	-	-	+/-	+	-
3	+	+	-	+	+	-
4	-	-	-	-	+	-
5	-	-	-	-	-	-
6	-	-	-	-	-	-
7	+	-	-	-	-	-
8	-	-	-	-	+	-
9	-	-	-	-	+/-	-
10	+	+/-	-	+	+	-
11	+	-	-	+/-	-	-
12	+	+	-	+	+	-
13	+/-	-	-	+/-	-	-
14	+	+	-	+	+	-
15	-	-	-	-	-	-
16	-	-	-	-	-	-

Alignment group 4 (60S L15b)

Technical notes:

Samples 1-16 PCR cycle LOW

Samples 17-32 PCR cycle HIGH

Samples 33-48 PCR cycle LOW

Samples 49-64 PCR cycle LOW

Samples 65-80 PCR cycle LOW+HIGH

PCR cycles:

LOW:

PCR cycle

7 min 95°

1 min 95°

1 min 37°

2 min 72°

40 X

10 min° 72

HIGH:

PCR cycle

7 min 95°

1 min 95°

1 min 45°

2 min 72°

40 X

10 min° 72

LOW+HIGH:

PCR cycle

7 min 95°

1 min 95°
1 min 37°
2 min 72°
} 5 X

1 min 95°
1 min 45°
2 min 72°
} 35 X

10 min° 72

Technical notes:

Results with primers designed for alignment group 4 were not convincing, since only a few positive signals were detected; primers for alignment group 4 have to be revaluated. Combinations of low and high PCR cycles were not used in all subsequent experiments.

Alignment groups 13, 14, 33, 46, 49, 52

Technical notes:

Primers designed for alignment groups 13, 14, 33, 46, 49 and 52 were evaluated on a *Ta* basis, meaning primer sets with similar *Ta* were combined on 96 well PCR plates. PCR conditions were equal to those given on page 18, except for the substituted annealing temperature given in Table 6.

Table 6: Combined primers sets and selected annealing temperature. Rows left and right beside the images (Figure A-D) are corresponding to the numbers given below (primer pairs).

	Annealing temperature = 42	Annealing temperature = 45
Figure A	Primer sets: 1, 2, 3, 4, 5, 11	
Figure B	Primer sets: 13, 15, 17, 19, 21, 22	
Figure C		Primer sets: 6, 7, 8, 9, 10, 12
Figure D		Primer sets: 14, 16, 18, 20

Primer sets:

1: Al13_49_67_F1 + Algr13_552R

2: Al14_64_80_F1 + Al14_429_414_R1

3: Al14_64_80_F1 + Al14_369_360_R_tail

4: Al14_69_80_F_tail + Al14_429_414_R1

5: Al14_69_80_F_tail + Al14_369_360_R_tail

6: Al33_54_73_F1 + Al33_890_871_R1

7: Al33_54_73_F1 + Al33_879_859_R2

8: Al33_54_73_F2 + Al33_890_871_R1

9: Al33_54_73_F2 + Al33_879_859_R2

10: Al46_169-186_F1a + Al46_693-675_R1

11: Al46_169-186_F1a + Al46_687_677_jzhtail_R

12: Al46_166-187_f1b + Al46_693-675_R1

13: Al46_166-187_f1b + Al46_687_677_jzhtail_R

14: Al46_191-198_JZHtail_f + Al46_693-675_R1

15: Al46_191-198_JZHtail_f + Al46_687_677_jzhtail_R

16: Al49_ 44_63_F1 + Al49_846_829_r1

17: Al49_ 44_63_F1 + Al49_838-827_jzhtail_r

18: Al49_112_129_F2 + Al49_846_829_r1

19: Al49_112_129_F2 + Al49_838-827_jzhtail_r

20: AlGr52_412-433_f1 + Algr52_1102_1084_R1

21: AlGr52_412-433_f1 + Algr52_745_737_jzhtail_r

22: Algr52_422_412_jzhtail_F + Algr52_1102_1084_R1

Figure A

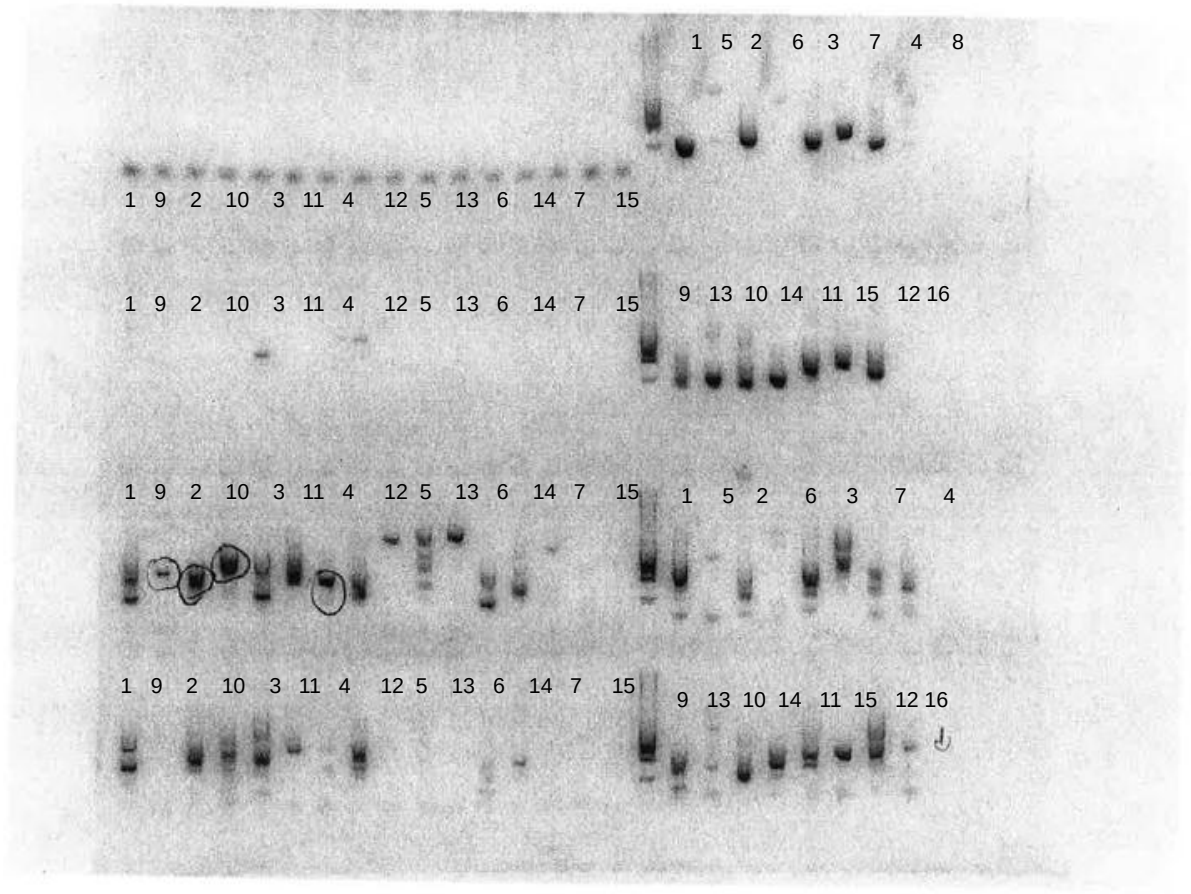
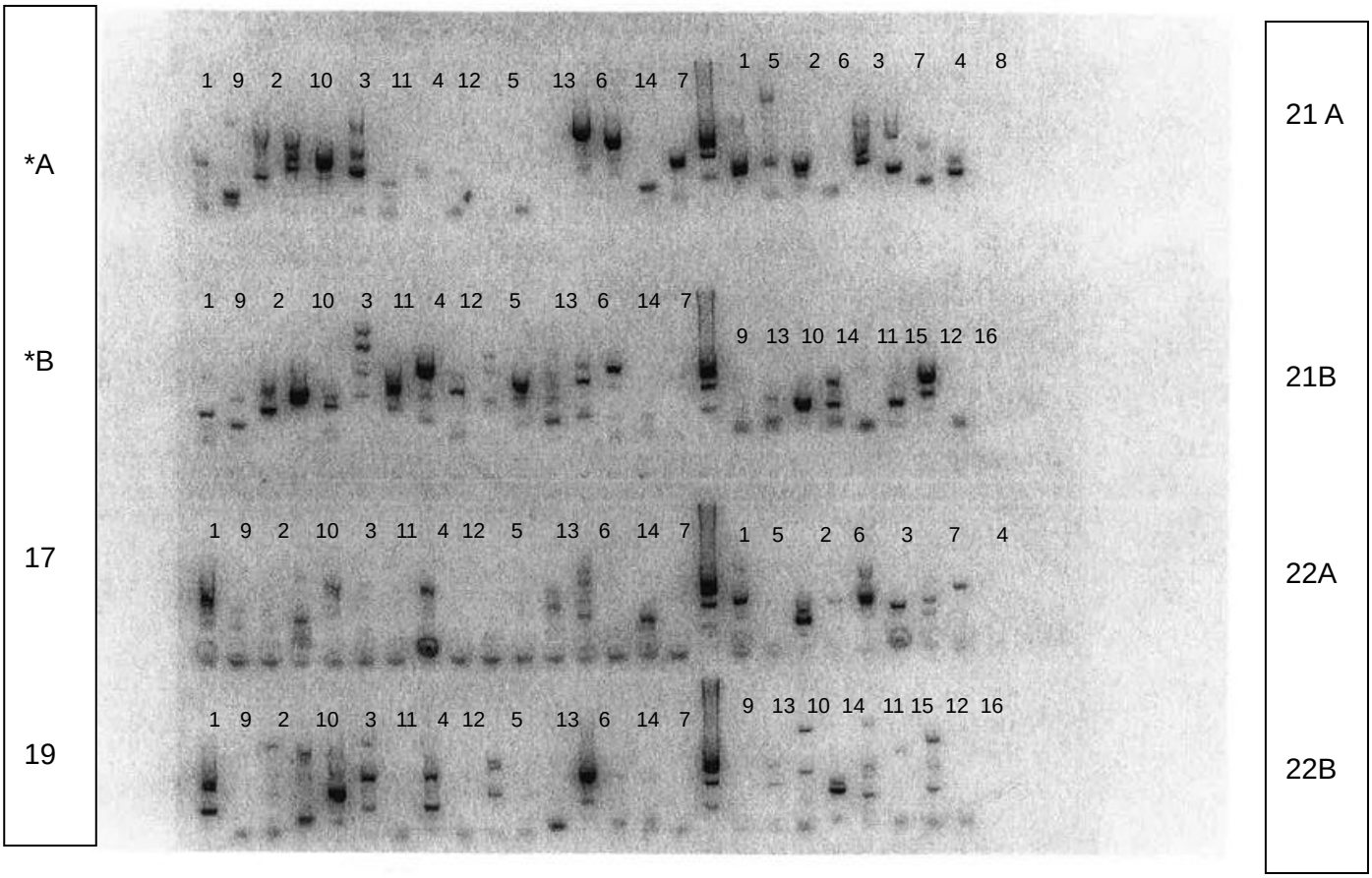

1
2
3
4
1 9 2 10 3 11 4 12 5 13 6 14 7 15
1 5 2 6 3 7 4 8
1 9 2 10 3 11 4 12 5 13 6 14 7 15
9 13 10 14 11 15 12 16
1 9 2 10 3 11 4 12 5 13 6 14 7 15
1 5 2 6 3 7 4
1 9 2 10 3 11 4 12 5 13 6 14 7 15
9 13 10 14 11 15 12 16
5a
5b
11a
11b

Figure B



* Row A samples 1-8 = 13 samples: 1-8; row A samples 9-16 = 15 samples 1-8

* Row B samples 1-8 = 13 samples: 9-16; row B samples 9-16 = 15 samples 9-16

Figure C

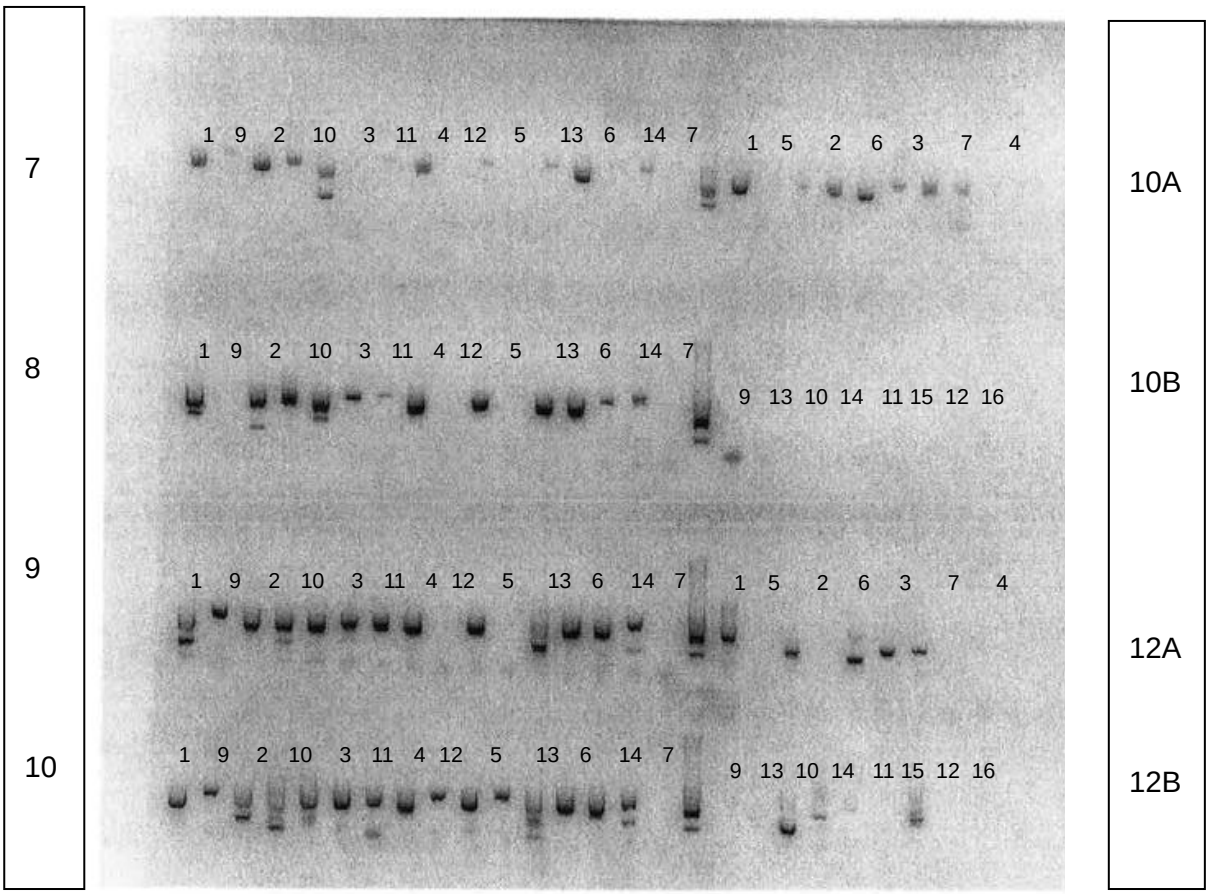
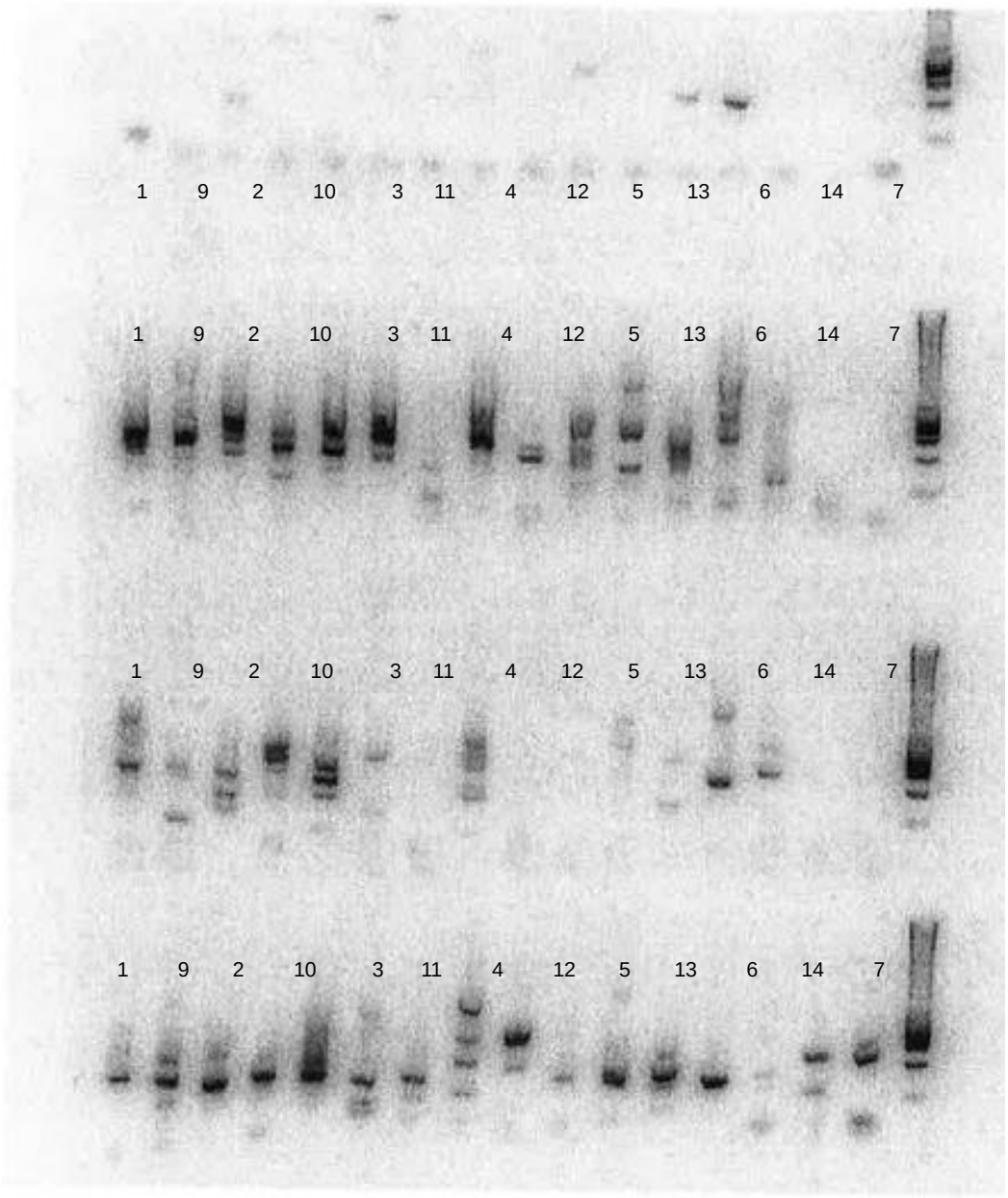

7
8
9
10
10A
10B
12A
12B
1 9 2 10 3 11 4 12 5 13 6 14 7
1 5 2 6 3 7 4
1 9 2 10 3 11 4 12 5 13 6 14 7
9 13 10 14 11 15 12 16
1 9 2 10 3 11 4 12 5 13 6 14 7
1 5 2 6 3 7 4
1 9 2 10 3 11 4 12 5 13 6 14 7
9 13 10 14 11 15 12 16

Figure D


14
16
18
20
1 9 2 10 3 11 4 12 5 13 6 14 7
1 9 2 10 3 11 4 12 5 13 6 14 7
1 9 2 10 3 11 4 12 5 13 6 14 7
1 9 2 10 3 11 4 12 5 13 6 14 7

Results (Tables)

*Columns indicate primer combinations (**Table 3**), rows indicate taxa (**Table 4**). Symbols: + indicates strong signal, +- faint signal, - poor or no signal.*

Alignment group 14

Sample	Primer set 27	Primer set 28	Primer set 31	Primer set 32
1	-	+	+	+
2	-	+	+	+
3	+	+/-	+	+
4	-	+	-	+
5	-	+ (largefragment)	-	+/--
6	-	+ (large fragment)	-	-
7	-	+	+	+
8	-	-	-	-
9	-	+ (large fragment)	-	+
10	-	+ (large fragment)	+	+
11	-	+	+	+
12	+/-	+/-	+	+
13	-	+	-	+
14	-	+	+/-	+
15	-	+ (large fragment)	-	+
16	-	-	-	-

Alignment group 33

Sample	Primer set 33	Primer set 34	Primer set 35	Primer set 36
1	+	+	+/- (double band)	+
2	+	+	+	+ (double band)
3	+	+	+	+
4	+/-	+/-	+	+
5	-	-	-	+ (large band)
6	-	-	-	+ (large band)
7	+	+	+	+
8	+/-	+	+ (large band)	+ (double band)
9	+/-	-	+ (large band)	+ (large band)
10	+	+	+	+ (small band)
11	-	+	+	+
12	+	+	+	+
13	+/-	+	+	+
14	+/-	+	+ (small band)	+/- triple band
15	+/--	+	+	+
16	-	-	-	-

Alignment group 46

Sample	Primer set 37	Primer set 39	Primer set 40
1	+	+	+
2	+/--	+/- (small band)	+ (small band)
3	+	+/- (double band)	+ (small band)
4	+/-	+/- (double band)	+ (small band)

5	-	+	-
6	+/-	-	-
7	+/-	+/- (double band)	+ (small band)
8	+/-	+/- (small band)	-
9	-	+/- (small band)	-
10	-	+ (small band)	+ (small band)
11	-	+ (double band)	+/-
12	-	+	+/- (double band)
13	-	+/-	-
14	-	+	+/-
15	-	+	-
16	-	+/- (?? N.C.)	-

Sample	Primer set 42	Primer set 43	Primer set 45
1	+/-	-	+ (small band)
2	+/- (double band)	+/-	+ (double band)
3	+	-	+ (double band)
4	-	-	+ (double band)
5	-	-	+
6	-	-	+
7	+	+	+ (double band)
8	-	-	-
9	-	-	+/- (double band)
10	+ (triple band)	-	+
11	+ (triple band)	-	+ (double band)
12	-	-	+
13	-	-	+/- (double band)
14	-	+/-	+/- (double band)

15	+ (large band)	-	+
16	+ (?? N.C)	-	-

Alignment group 49

Sample	primer set 46	Primer set 47	Primer set 48	Primer set 49
1	+ (double band)	+ (double band)	+ (double band)	+ (double band)
2	+ (double band)	-	+ (double band)	+/- (double band)
3	+ (double band)	+/- (double band)	+ (double band)	+ (double band)
4	-	-	-	-
5	+ (small band)	-	-	-
6	+ (double band)	-	+/- (large band)	-
7	+ (double band)	+/- (double band)	+ (double band)	+ (double band)
8	-	+ (small band)	-	+/--
9	+ (double band)	+/- (double band)	+/-	-
10	+ (small band)	+/- (double band)	+ (double band)	+/-- (double band)
11	+ (double band)	+/- (double band)	+/-	+ (double band)
12	+ (double band)	+	+/- (double band)	+ (double band)
13	+ (double band)	-	-	+ (double band)
14	+ (double band)	+/- (double band)	+/-- (double band)	-
15	+/- (double band)	-	+ (double band)	+/--
16	-	-	-	-

Alignment group 52

Sample	Primer set 52	Primer set 54	Primer set 55
1	+ (double band)	+	+

2	+ (double band)	+	+ (double band)
3	+ (double band)	+ (double band)	+ (double band)
4	+	+ (double band)	+/- (double band)
5	+ (double band)	+/-	-
6	+	-	+/-
7	+	+ (double band)	+
8	+ (double band)	+ (double band)	+
9	+ (double band)	-	-
10	+	+	+ (double band)
11	+ (double band)	-	+/- (double band)
12	+ (double band)	+ (double band)	+ (double band)
13	+	+/- (double band)	+/- (double band)
14	+ (double band)	+ (double band)	+ (double band)
15	+/--	+	-
16	+ (?? N.C.)	-	-

Trial II

Technical notes:

The following primer sets appear to be most promising and are used for further testing and optimization. All sets were tested with 48 samples (including negative control). PCR conditions were equal to those given on page 18, at either 42 or 45 °C.

Primer sets trial II:

PSS1: Al14_64_80_F1 + Al14_369_360_R_tail

PSS2: Al14_69_80_F_tail + Al14_369_360_R_tail

PSS3: Al33_54_73_F1 + Al33_879_859_R2

PSS4: Al33_129_148_F2 + Al33_879_859_R2

PSS5: Al46_191-198_JZHtail_f + Al46_687_677_jzhtail_R

PSS6: Al49_ 44_63_F1 + Al49_846_829_r1

PSS7: AlGr52_412-433_f1 + Algr52_1102_1084_R1

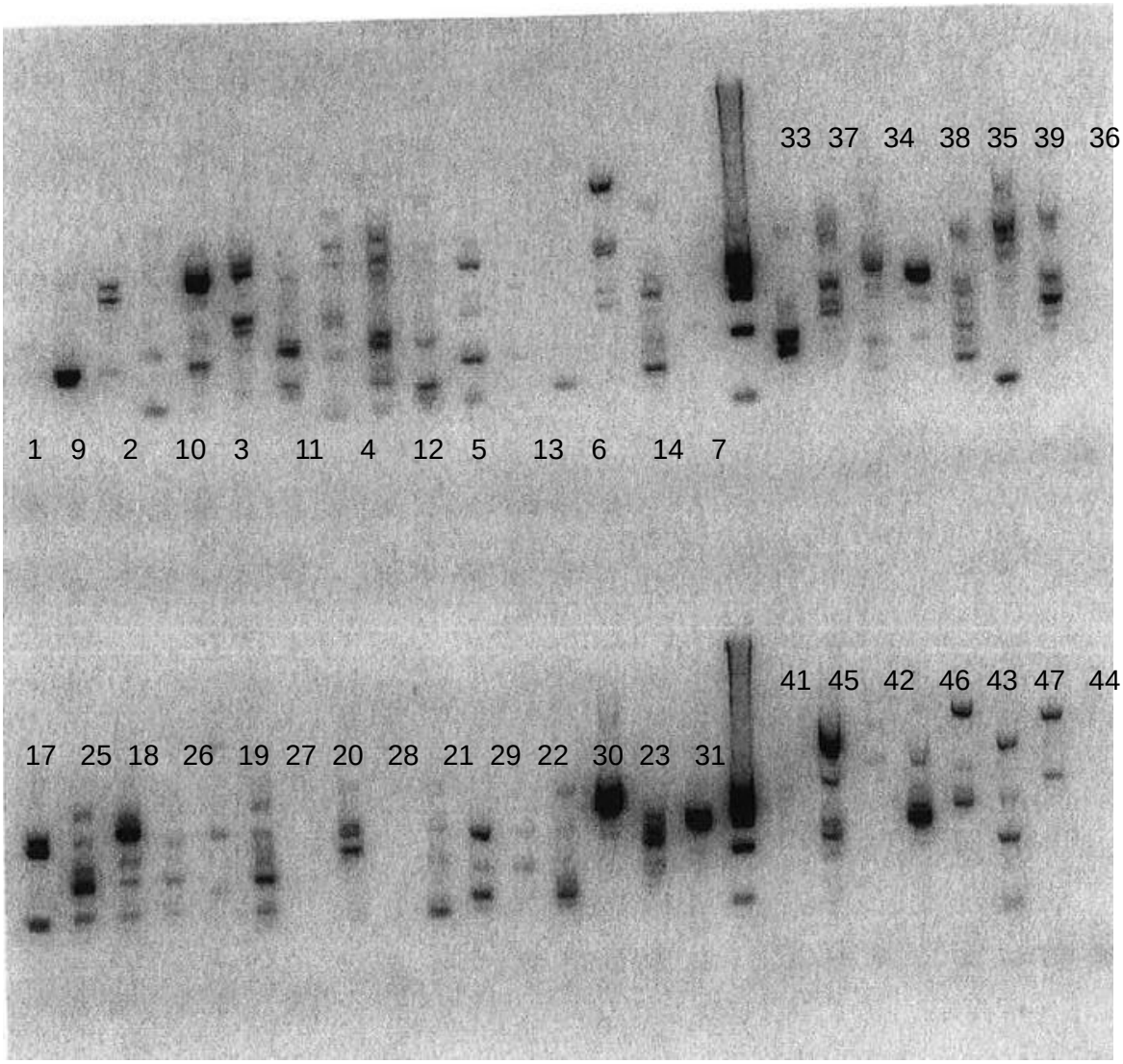
Table 7: Selected taxa for trial II

Sample	CBS No.	Extract	Taxon
1	CBS 113480	32191	*Microsporum canis*
2	CBS 113850	32192	*Coccidioides immitis*
3	CBS 117146	32195	*Pyrenophora tritici-repentis*
4	CBS 118699	32196	*Alternaria brassicicola*
5	CBS 120258	32197	*Mycosphaerella fijiensis*
6	CBS 123670	32199	*Fusarium verticillioides*
7	CBS 124811	32200	*Schizophyllum commune*
8	CBS 128304	32202	*Pyricularia grisea*
9	CBS 136.29	32203	*Bipolaris maydis*
10	CBS 180.27	32205	*Neurospora tetrasperma*
11	CBS 208.27	32207	*Sporobolomyces roseus*
12	CBS 223.38	32208	*Neurospora tetrasperma*
13	CBS 2605	32209	*Lodderomyces elongisporus*
14	CBS 287.54	32210	*Histoplasma capsulatum*
15	CBS 372.73	32211	*Paracoccidioides brasiliensis*
16	CBS 392.92	32212	*Trichoderma reesei*
17	CBS 484	32215	*Sporidiobolus pararoseus*
18	CBS 658.66	32219	*Pyricularia grisea*
19	CBS 674.68	32220	*Ajellomyces dermatitidis*
20	CBS 693.94	32221	*Trichoderma atroviride*
21	CBS 8710	32224	*Filobasidiella neoformans*
22	CBS 8758	32225	*Candida albicans var. albicans*
23	FGSC 1144	32228	*Aspergillus niger*
24	FGSC 1156	32229	*Aspergillus terreus*
25	FGSC 1089	32230	*Gibberella fujikuroi*
26	FGSC 9002	32231	*Phanerochaete chrysosporium*
27	FGSC 9075	32232	*Fusarium graminarum*
28	FGSC 9935	32234	*Fusarium oxysporum fsp lycopersici*
29	CBS 101191	32237	*Neurospora tetrasperma*
30	CBS 127.97	12013	*Trichophyton equinum var. equinum*
31	CBS 126970	13550	*Coprinus cinereus*
32	CBS 126971	13551	*Rhizopus orysae*
33	CBS 126972	13552	*Aspergillus nidulans*
34	CBS 126969	13549	*Podospora pauciseta*
35	CBS 127169	13633	*Verticillium alboatrum*
36	CBS 127171	13635	*Stagonospora nodorum*
37	CBS 127172	13636	*Ustilago maydis*
38	CBS 113843	22379	*Coccidioides posadasii*

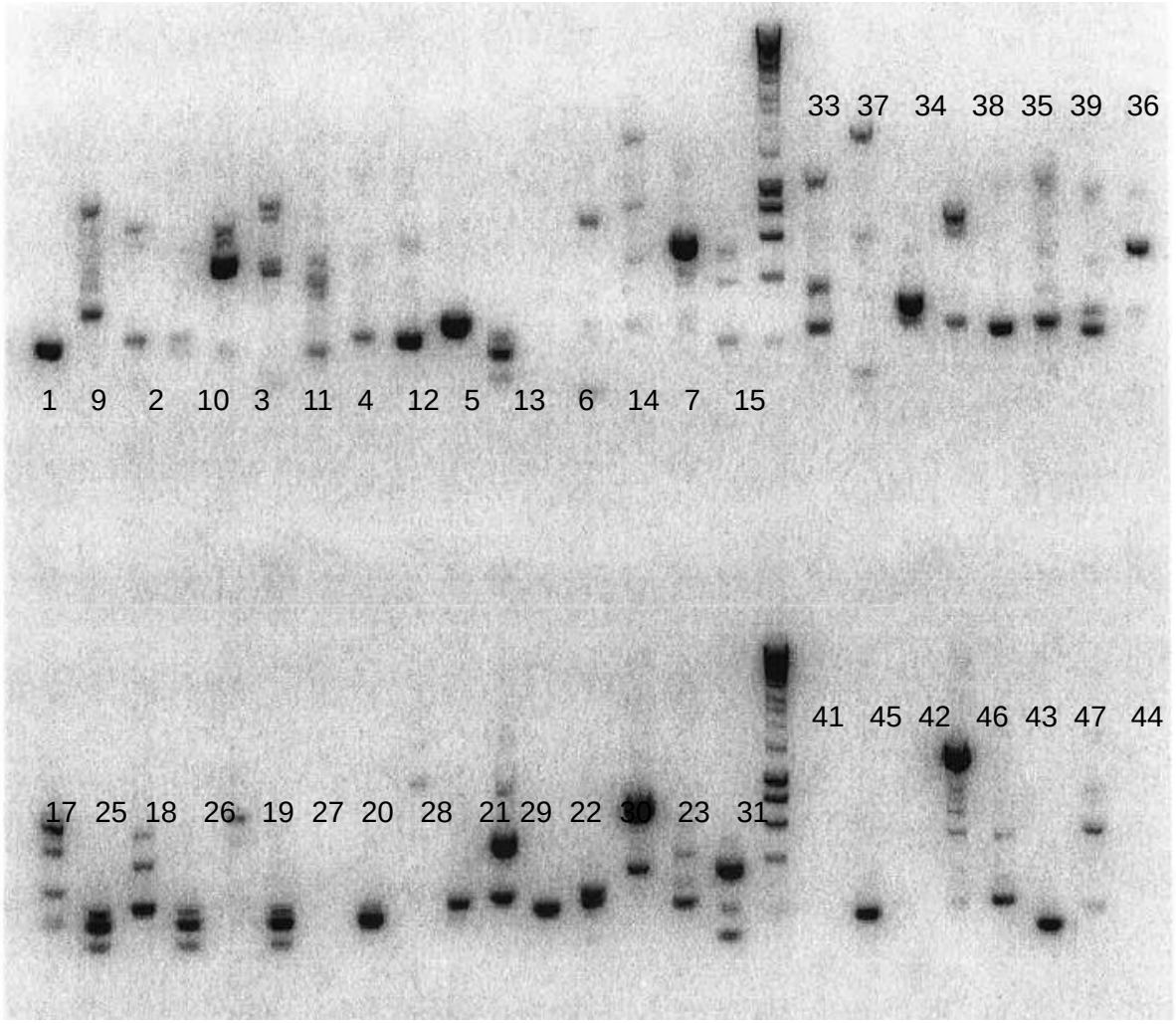
39	CBS 375.48	D00744	*Talaromyces stipitatus*
40	CBS 131.61	D01707	*Aspergillus flavus var. Flavus*
41	CBS 114389	D01728	*Blasomyces dermatitidis*
42	CBS 383.78	D01823	*Trichoderma reesei*
43	CBS 668.78	D01837	*Uncinocarpus reesii*
44	CBS 7116	10029	*Schizosaccharomyces japonicus*
45	CBS 599	10136	*Yarrowia lipolytica*
46	CBS 6054	20425	*Pichia stipitis*
47	CBS 732	21072	*Zygosaccharomyces rouxii*
48	water		*negative control*

Table 7 (continued): Selected taxa for trial II

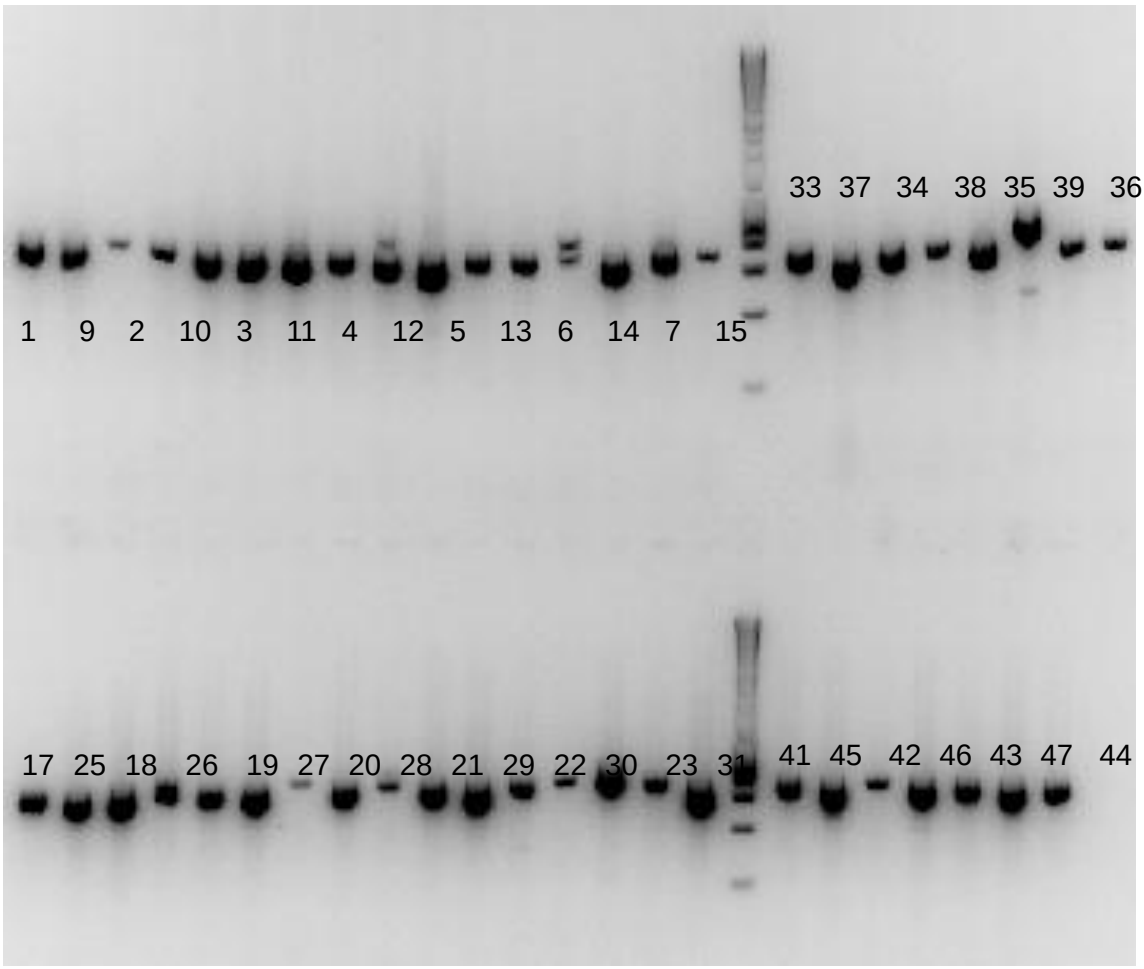
PSS1 (expected size = ~ 300 nt)



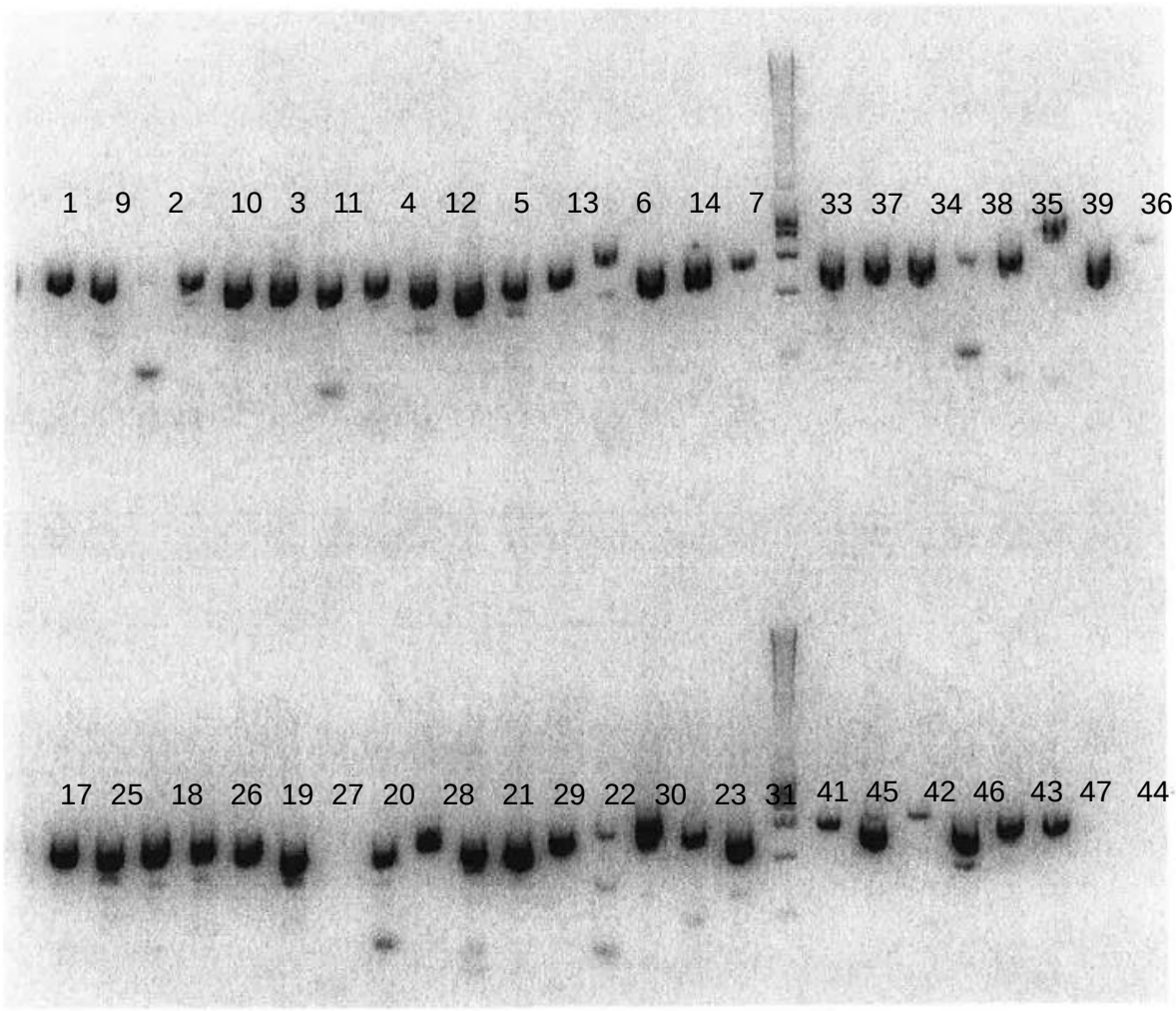
PSS2 (expected size = ~300)



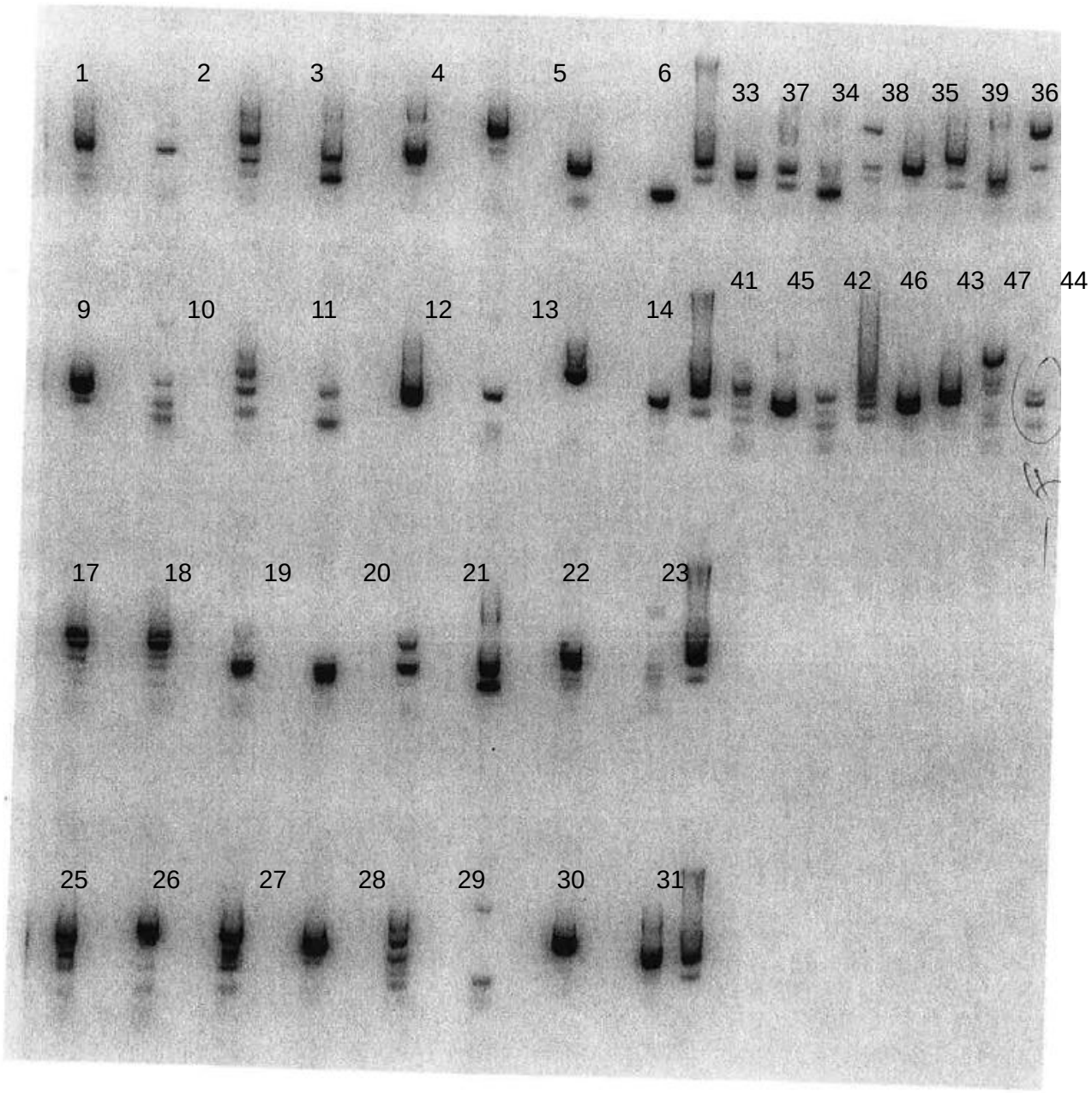
PSS3 (Expected size = ~750 nt)



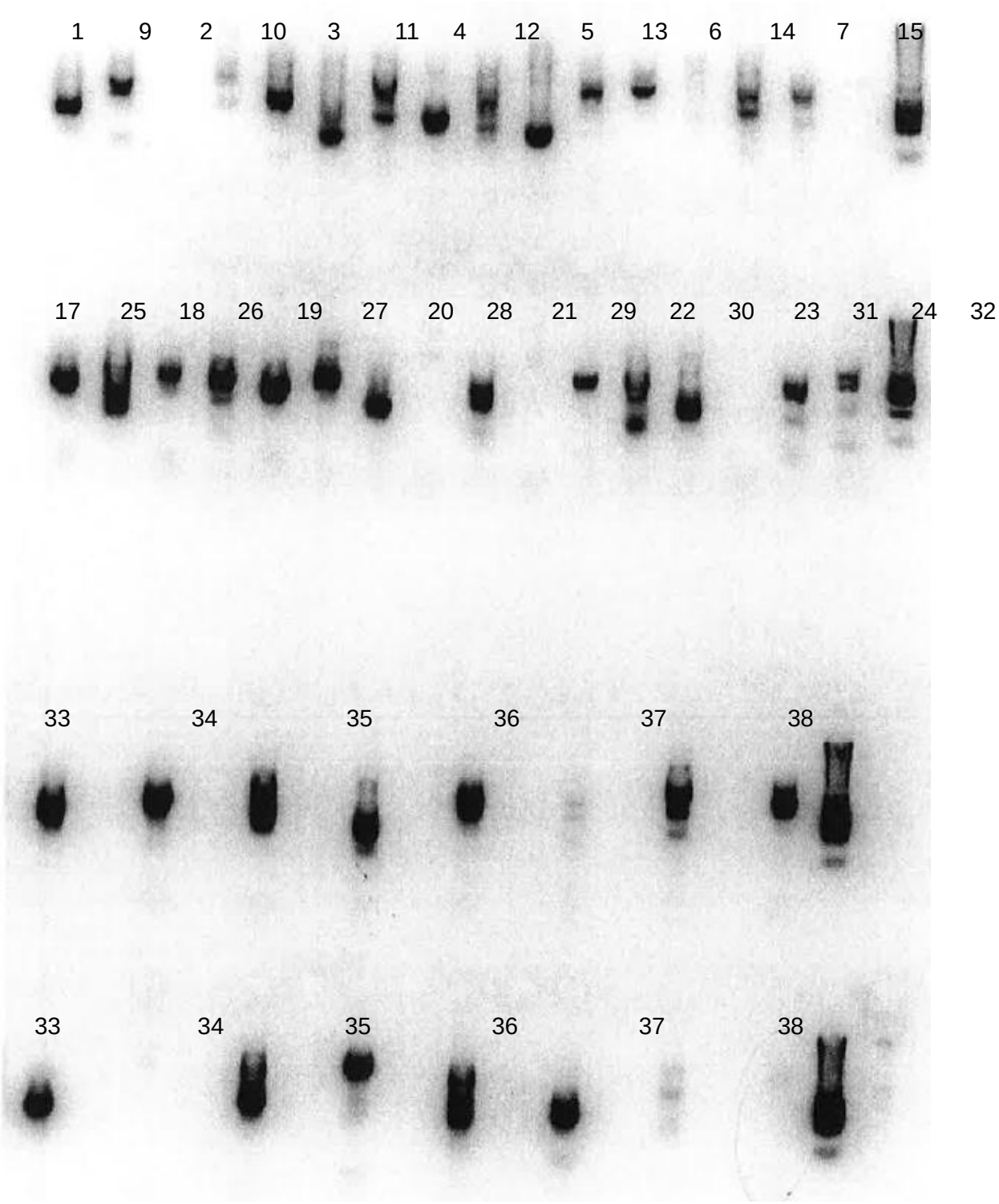
PSS4 (expected size = ~650 nt)



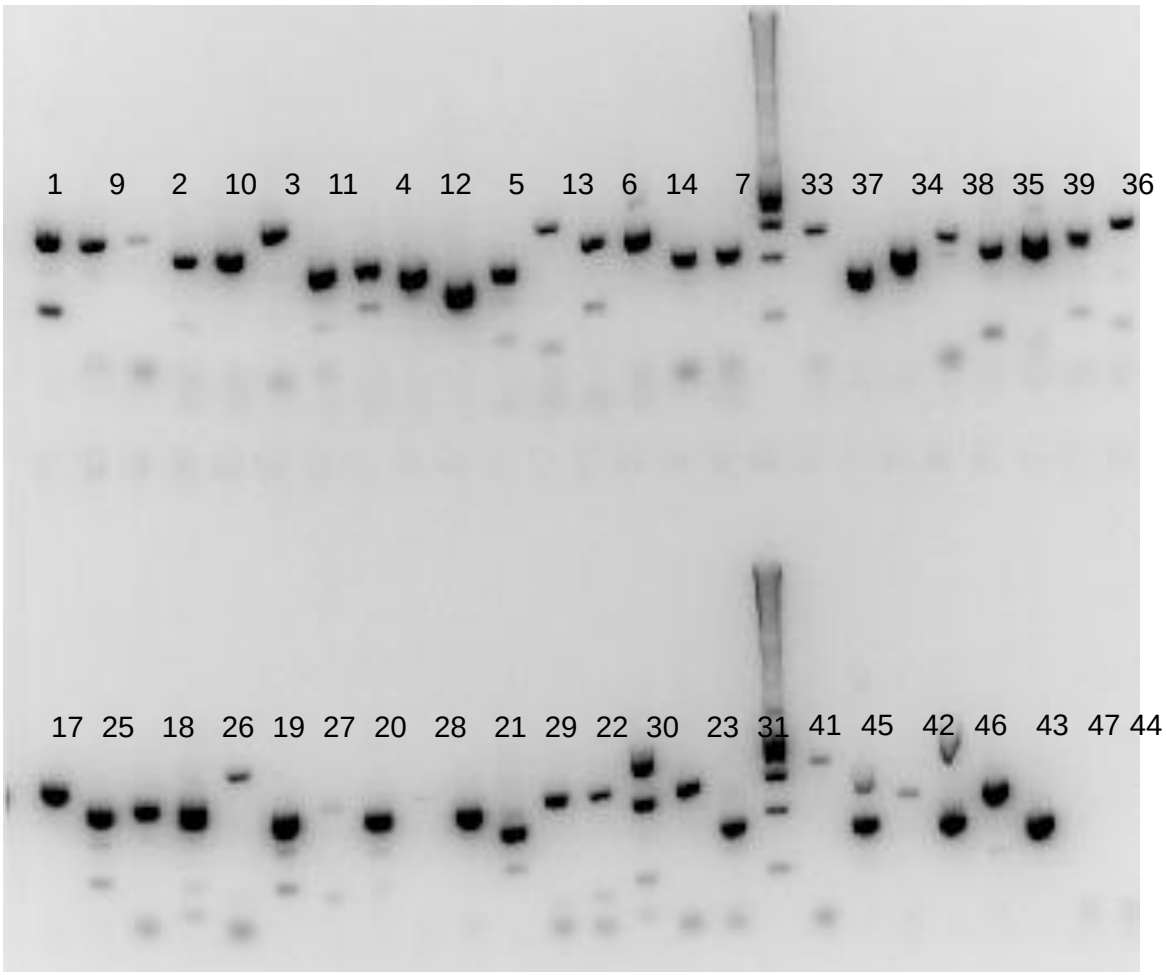
PSS5 (expected size = 500 nt)



PSS6 (expected size = ~750 nt)



PSS7 (Expected size = ~650)



Significance of this deliverable

The use of this *in silico* approach to target new genes from whole genomes to identify fungi seems promising. Some of the gene regions proposed here have not been formerly used in fungal systematics or phylogeny. They have a better resolution potential to differentiate species than most presently used ribosomal and protein coding markers. These promising results should now be extended to more fungi taxa.