

RESEARCH ARTICLE

The thionin family of antimicrobial peptides

Katharina Höng¹, Tina Austerlitz¹, Timo Bohlmann¹, Holger Bohlmann¹*

Department of Crop Sciences, Institute of Plant Protection, University of Natural Resources and Life Sciences, Vienna, Austria

* These authors contributed equally to this work.

* holger.bohlmann@boku.ac.at

Abstract

Thionins are antimicrobial peptides found only in plants. They are first produced as preproteins and then processed to yield the usually 5 kDa, basic thionin peptide with three or four disulfide bridges. So far, thionins had only been found in some plant families of angiosperms. The One Thousand Plant Transcriptomes Initiative (1KP project) has sequenced the transcriptomes of more than 1000 plant species. We have used these data to search for new thionin sequences which gave 225 hits. After removing doublets these resulted in 133 new thionins. No sequences were found in algae and mosses. The phylogenetically earliest hits were from *Selaginella* species and from conifers. Many hits were from angiosperm plant families which were previously not known to contain thionins. A large gene family for thionins was found in *Papaver*. We isolated a genomic clone from *Papaver somniferum* which confirmed the general genomic structure with two small introns within the acidic domain. We also expressed the thionin encoded by the genomic clone and found that it had antimicrobial activity *in vitro*, especially against fungi. Previously, we had grouped thionins into four classes. The new data reported here led us to revise this classification. We now recognize only class 1 thionins with eight cysteine residues and class 2 thionins with six cysteine residues. The different variants that we found (and also previously known variants) can all be traced back to one of these two classes. Some of the variants had an uneven number of cysteine residues and it is not clear at the moment what that means for their threedimensional structure.

OPEN ACCESS

Citation: Höng K, Austerlitz T, Bohlmann T, Bohlmann H (2021) The thionin family of antimicrobial peptides. PLoS ONE 16(7): e0254549. <https://doi.org/10.1371/journal.pone.0254549>

Editor: Rana Muhammad Atif, University of Agriculture Faisalabad, PAKISTAN

Received: May 3, 2021

Accepted: June 28, 2021

Published: July 14, 2021

Copyright: © 2021 Höng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting information files](#).

Funding: This work has been supported by grant P28984_B22 of the Austrian Science Fund (FWF) <https://www.fwf.ac.at/de/> to Holger Bohlmann. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

A large variety of relatively small, basic, and often cysteine-rich polypeptides have been isolated from different organisms and shown to have antimicrobial activities *in vitro* [1–3]. Some prominent examples from animals include magainins from the skin of *Xenopus laevis* and mammalian defensins and β -defensins. Antimicrobial peptides (AMPs) seem to be distributed ubiquitously in multicellular organisms. Plants have also been shown to contain a variety of different AMPs, including thionins [4,5] and plant defensins [6,7]. It is thought that the molecular targets of the majority of the usually basic AMPs are acidic phospholipids in biomembranes [8]. However, at least for some AMPs, biomembranes may just be a barrier hindering access to their primary target inside the cell [9–11].

Thionins (Fig 1) are a group of AMPs whose toxic activity was apparently first described by Jago and Jago in 1885 [12] who reported that wheat flour contains a substance that was toxic to yeast. Balls and Hale [13] found that the polypeptide responsible for the toxicity could be extracted with petroleum ether and named it purothionin. It was later shown to consist of three different isoforms [14–16] whose sequences have been determined. Purothionins, as the majority of thionins, are basic and cysteine-rich polypeptides with a molecular weight of about 5 kDa. Related polypeptides, all with eight conserved cysteine residues, have been isolated and sequenced from the endosperm of oat and barley and are also present in other species of the *Poaceae* [17].

Cultivated barley (*Hordeum vulgare* L.) and also many wild *Hordeum* species contain a large multigene family of closely related genes for leaf thionins [20,28,29] in addition to the small gene family for endosperm specific hordothionins [19]. Mature leaf thionins have been found in the vacuole [30] and in the cell wall [28] of barley leaves. A prominent feature of leaf thionin genes in barley is their light responsiveness. Etiolated barley seedlings contain a very high level of leaf thionin transcripts which drops drastically after illumination, probably mediated by two photoreceptors, phytochrome and a blue-light-absorbing photoreceptor [31]. Barley leaf thionins and hordothionins have repeatedly been shown to have antimicrobial activities *in vitro* against different phytopathogenic bacteria and fungi [28,32,33]. This *in vitro* toxicity and the cellular distribution of barley leaf thionins suggested a role in plant defense.

Thionins are also present in dicot species. One large group of thionins (Fig 1) has been found in different species of mistletoes [17]. The toxic effects of viscotoxins towards mammals have been investigated in detail, for instance by Rosell and Samuelsson [34] because of the medical use of mistletoe extracts in the treatment of cancer [35]. Several thionins with six cysteine residues have been described from species of the family *Brassicaceae* (Fig 1). These include crambin with six cysteine residues from the seeds of *Crambe abyssinica* [36]. Crambin has a pI of 5.73 and has no known toxic activities [37]. Schrader-Fischer and Apel found in addition to crambin several novel and highly variable thionins in *Crambe abyssinica* using a PCR based approach [38]. *Arabidopsis* contains four thionin genes, of which two have been studied in detail [26,39–42], giving strong support for a role in plant defense.

Thionins are characterized by two consecutive cysteine residues at position 3 and 4. Some variants are known that are missing one of these cysteine residues. Accordingly, thionins have been classified [18] as class 1 with eight cysteine residues, class two with six cysteine residues and variants of class 3 and 4 (Fig 1).

The seed specific crambin has been used intensively for threedimensional structure analysis [43]. Structures have also been determined for several other thionins. The general threedimensional structure of thionins shows a compact molecule consisting of two alpha helices and two beta sheets (reviewed by [44]). The threedimensional structure is related to the toxicity and antimicrobial activity of thionins. Many studies have shown that thionins can induce cell membrane permeability [5]. A phospholipid binding site, which includes the tyrosine at position 13 (Fig 1) which has been shown to be indispensable for toxicity, is present in all known structures of thionins with antimicrobial activity and it has been proposed that thionins interact with negatively charged phospholipids present in the cell membrane. This binding and the withdrawal of phospholipids disturbs the fluidity of the membrane and finally results in membrane lysis [44]. Another possible mechanism has been reported, claiming that thionins which are inserted into the membrane act as water channels. Water is delivered through these channels to the lipophilic center of the biomembrane, leading to local membrane disruption [45].

Thionins have been isolated as the 5 kDa mature peptides (Fig 1) from different plant species but are synthesized as much larger precursors [46]. Cloning of different thionin cDNAs and genomic clones confirmed the original observation and revealed that these mature

Class1 thionins (8 Cys)

HvTHI1.1¹ KSCCRSTLGRNCYNLCRVRGAQKLCAGVCRCKLTSSGKCP TGF PK
 HvTHI1.3² KSCCKDTLARNCYNTCHFAAGG-SRPVCAGACRCKIISGPKCPSDYPK
 PpTHI1.1³ KSCCRNTWARNCYNVCRLPGTISR EICAKKCDCKIISGTTCP SDYPK

Class2 thionins (6 Cys)

VaTHI2.1⁴ KSCCPNTTGRNIYNACRLTGAPRPTCAKLSGCKIISGSTCPSDYPK
 CaTHI2.1⁵ TTCCPSIVARSNFNVCRRLPGTSEAI C ATYTGCIIIPGATCPGDYAN
 AtTHI2.1⁶ KICCP SNQARNGYSVCRIRFSKGR-CMQVSGCQNS--DTCPRGWVN

Class3 thionins (6 Cys; without Cys2 and Cys8)

TaTHI3.1⁷ VDCGANPFK VACFNSCLLGPSTVFQCA DFCACRLPAG

Class4 thionins (8 Cys; without Cys2, new Cys between Cys4 and Cys5)

TgTHI4.1⁸ KSCFPSTA AKYCYNACRLPGCRPETIC AARC GCKIISSGNCP PGYEN

blue, basic amino acids; **red**, acidic amino acids; **yellow**, cysteine; **pink**, tyrosin; **green**, phenylalanin

Fig 1. Examples of thionins including previous classification of thionins [18]. ¹HTH1, Hordothionin α [19]; ²Barley leaf-thionin DB4 [20]; ³Pyricularia thionin [21]; ⁴Viscotoxin A3 [22,23]; ⁵Crambin 2 [24,25]; ⁶*Arabidopsis thaliana* thionin 2.1 [26]; ⁷pTTH20 Neutral wheat thionin [27]; ⁸*Tulipa gesneriana* thionin4.1 (Genbank, submitted by Luyten et al. 1994).

<https://doi.org/10.1371/journal.pone.0254549.g001>

thionins are first synthesized as preproteins (S1 and S2 Figs). All precursors have a typical N-terminal signal peptide with a conserved signal peptide cleavage site that directs the proprotein into the endoplasmic reticulum. The proprotein consists of the thionin itself and a C-terminal acidic domain. In those cases where genomic DNA sequences are known [19,42,47], there are two small introns in the acidic domain (S2 Fig).

The function of the acidic domain is still not clear. Acidic domains have six cysteine residues (S1 Fig) which are highly conserved as are the cysteine residues in the thionin and are usually acidic (as the name says). An exception are the acidic domains deduced from the *Tulipa* thionin precursors that have only three cysteine residues (sequences submitted to GenBank by Luyten et al. 1994). In the case of the acidic crambin, the acidic domain too is slightly acidic (S1 Table). The acidic domain of the *Arabidopsis* THI2.3 proprotein is actually basic (S1 Table). A protein corresponding to the acidic domain has, to our knowledge, never been isolated from plants. To date, there is no experimental information available about possible functions of the acidic domain. But it is clearly not dispensable as shown by the high conservation of the cysteine residues, even in the case of viscotoxin precursors which have several deletions in the acidic domain [23]. Furthermore, Florack et al. [48] found that expression of α -hordothionin in transgenic tobacco plants without the acidic domain resulted in significantly lower levels of the mature thionin.

One possible function might be that the acidic domain contains information to guide the thionin through the secretory pathway to its final destination (vacuoles, cell walls, protein bodies). Another function of the acidic domain might be to neutralise the basic thionin and thereby protect the cell against its own toxin. A toxic effect of barley leaf thionins has been demonstrated against tobacco protoplasts and against barley protoplasts [49].

We have recently purified a thionin precursor processing enzyme (TPPE) from etiolated barley seedlings which contain large amounts of leaf-specific thionins [31]. We used a fluorescently labelled peptide that incorporated the flanking sequence between barley leaf-specific thionins and their acidic domains to identify a TPPE protein as the subtilase BAJ93208. The barley TPPE was produced as a *strep*-tagged protein in *E. coli* and used to study the processing of a recombinant leaf-thionin precursor *in vitro*. It could be shown that this protease produced thionins with a correct C-terminus while the acidic domain was cleaved several times. This explains why a peptide corresponding to the acidic domain has never been isolated from plants. Furthermore, it could be shown that the thionin domain was protected by its three-dimensional structure against cleavage by the TPPE [50].

While plant defensins are ubiquitously distributed in plants, thionins were only found in a limited number of plant families. The One Thousand Plant Transcriptomes Initiative (1KP project) which sequenced the transcriptomes of more than 1000 plant species [51] was therefore a good opportunity to study the distribution of thionins in the plant kingdom. We searched the data obtained in that project for thionin sequences. We found thionin sequences from several families which had previously no reported thionins, including a large gene family in the genus *Papaver*. However, a large number of plants was without thionin sequences. We also found a number of sequences that would code for unknown thionin variants which prompted us to propose a new classification system for thionins. A genomic clone was isolated from *P. somniferum* and the encoded thionin was expressed in *E. coli* and shown to have antimicrobial activity.

Materials and methods

Database searches

We used BLAST to search the translated 1KP data with thionin amino acid sequences. As bait sequences we used a range of well-known thionins, for instance THI2.1 from Arabidopsis, VaTHI2.1 (Viscotoxin A3), CaTHI2.1 (Crambin), HvTHI1.3 (leaf thionin DB4 from Barley) and PpTHI1.1 (Pyricularia thionin) as well as some sequences that were found using the former baits (S2 Table). Searching was done with each of these bait sequences against all samples separately. From these results we removed the sequences which were repeatedly found by the different bait sequences. This resulted in a list of 2130 primary hits. All hits were then curated manually to remove the non-thionin sequences. Typically, thionins would have a double cysteine motif at position 3 and 4 as well as an additional four or six cysteines and an aromatic amino acid at position 13. Furthermore, we also looked for a possible acidic domain extension with its typical three and three cysteines. Manual inspection also allowed us to discover several thionin variants with untypical cysteine motifs. This produced a final list of 225 hits of which some resulted in the same sequences because the 1KP project often sequenced several different tissues for one species. Sometimes these different tissues were in addition pooled and sequenced, also resulting in redundancy. Signal peptides and acidic domain sequences were removed because they were often truncated. In some cases where the end of the signal peptide was not obvious (in case of typical thionin sequences with a basic amino acid at position 1) SignalP 5.0 (<http://www.cbs.dtu.dk/services/SignalP/>) was used to predict the signal peptide

cleavage site [52]. The cleavage site between thionin domain and acidic domain was predicted according to known thionin sequences being six amino acids after the last cysteine.

Mw and pI were computed at https://web.expasy.org/compute_pi/. Pairwise alignment was performed by using emboss needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) [53] and clustal omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) was used for multiple sequence alignments [53]. Sequence logos were created at <http://weblogo.threeplusone.com/create.cgi> [54]. For creation of the logo two truncated class 1 sequences were removed. For the class 2 logo we removed a truncated sequence and two sequences (ChTHI2.1 and MfTHI2-1) with insertions in the middle for easier comparison with class 1.

Cloning

All primers used for cloning can be found in [S3 Table](#). Genomic DNA was prepared from *P. somniferum* seedlings according to [55]. A genomic clone was amplified using primers PsoThi1.5forNco and PsoThi1.5revBam. A fragment of approximately 730 bp was obtained. It was digested with NcoI and BamHI and ligated to the vector fragment of pMAAred [56] digested with the same enzymes, thus replacing the GUS sequence. The insert was sequenced and shown to be a thionin preproprotein with two introns.

For expression of the thionin encoded by the genomic clone we constructed an expression vector for *E. coli* which was based on a modified pETtrx1a vector [57] with an NdeI site at the start codon and a BamHI site behind the stop codon. The vector contained a His-tag in front of the thioredoxin sequence followed by a TEV site. We amplified the thioredoxin part of the vector by PCR using primers pETtrxfor1 and pETtrxTEVrev. The thionin coding sequence of the PsoTHI1.7 genomic clone was also amplified by PCR using the primers TEV1PsoTHI1.7for and PsoTHI1.7Bamrev. The thioredoxin fusion fragment and the thionin fragment were combined and amplified with primers pETtrxfor1 and PsoTHI1.7Bamrev. The final PCR fragment was digested with NdeI and BamHI and cloned into the modified pETtrx1a vector previously digested with the same enzymes. The final expression vector was verified by sequencing.

Expression and purification of PsoTHI1.7

The recombinant pETtrx::PsoTHI1.7 vector was transformed into the *E. coli* strain C3030 Shuffle [58]. A 100 ml TB medium pre-culture (containing 50 mg/ml kanamycin) was inoculated with a single colony and incubated overnight at 37°C and 180 rpm. The next day, 2 L TB medium (containing 50 mg/ml kanamycin) were inoculated with 3% v/v pre-culture and grown at 30°C and 140 rpm until the OD₆₀₀ reached 1. Expression was induced by addition of 0.75 mM IPTG and incubated overnight at 20°C and 140 rpm. Cells were harvested by centrifugation for 30 min at 4°C (Sorvall RC6+ centrifuge, Thermo Scientific) and resuspended in 30 ml of ice-cold buffer (20 mM NaH₂PO₄, 0.5 M NaCl, 50 mM imidazole, pH 7.4). Cells were lysed by sonication (Sonifier W-250D, Branson Ultrasonics) in an ice-water bath. The lysate was clarified by centrifugation at 18,000 rpm for 30 min at 4°C.

The supernatant was used to isolate the fusion protein by immobilized metal ion affinity chromatography on an ÄKTA purifier10 FPLC system (GE healthcare) with a HisTrap HP 5ml column (GE Healthcare), equilibrated with 10 CV binding buffer (20 mM NaH₂PO₄, 0.5 M NaCl, 50 mM imidazole, pH 7.4). The sample was loaded with the sample pump at a flow rate of 5 ml/min and the column was washed with binding buffer until the 280 nm absorbance reached a steady baseline. The fusion protein was eluted with 2 CV of Elution buffer (20mM NaH₂PO₄, 0.5 M NaCl, 500 mM imidazole, pH 7.4) and collected with a fraction collector. Flow-through and the eluted fractions were analyzed by SDS-PAGE [59] using a Mini PROTEAN Tetra Cell Imm (BioRAD). Fractions containing fusion protein were pooled and

dialysed (SpectraPor RC tubing, MWCO 3.5 kDa Spectrumlabs) to a buffer containing 25 mM Tris-HCl pH 8, 0.5 mM EDTA, and 0.25 mM DTT was added directly to the sample before proteolytic digest. The fusion protein was cleaved using TEV protease at a ratio of 1:10 and cleavage was confirmed by PAGE (S3A Fig).

The cleaved fusion protein was dialysed to a binding buffer (20 mM NaH₂PO₄, 0.5 M NaCl, 5 mM imidazole, pH 7.4) for negative His-tag purification. Binding and loading to a column were performed as described before. The flow through, which contained PsoTHI1.7, was collected and the washing step was omitted. (His)6-TRX, remaining undigested fusion protein and the TEV protease were eluted (20 mM NaH₂PO₄, 0.5 M NaCl, 500 mM imidazole, pH 7.4) and discarded. The volume of the flowthrough was reduced to 500 µl using an Amicon Ultra centrifugal filter unit (MWCO 3 kDa). The sample was applied to a Superdex 75 10/300 GL column on an ÄKTA purifier10 (GE Healthcare) using the 100 µl sample loop with a flow of 0.5 ml/min. The running buffer was 100 mM TRIS-HCl pH 8, 500 mM NaCl. The purified peptide was run on a PAGE gel (S3B Fig) and silver stained according to [60]. For protein quantification, Thermo Scientific™ Pierce™ BCA Protein Assay kit was used. Page Ruler Unstained (Thermofisher 26614) and Page Ruler Unstained Low range (Thermofisher 26632) were used as PAGE marker.

Antimicrobial assays

Antimicrobial assays were done according to [61]. Fungal test organisms were *Botrytis cinerea* 05.10 and *Fusarium oxysporum* f.sp. *matthiolarum*. Both fungi were grown on PDA for two weeks at room temperature. *B. cinerea* was grown in the dark while *F. oxysporum* was grown in the light on the lab bench. Spores were harvested by flooding the plate with 10 ml of sterile water and filtering the spore suspension through two layers of cheesecloth. The spores were counted in a Thoma hemocytometer and the number of spores per µl was calculated. Spore concentration was adjusted to 2 x 10⁴ spores/ml in ½ concentrated PDB. 25 µl of the test peptide were pipetted into the wells of a clear 96 well cell culture microplate (PS,TC, F-bottom with a lid, Greiner Bio-One). Then 75 µl of the spore suspension was added. Final peptide concentrations were 100 µg/ml, 50 µg/ml, 25 µg/ml, 12.5 µg/ml, 6.25 µg/ml, 3.125 µg/ml, 1.56 µg/ml and 0 µg/ml. Plates were incubated in the dark at room temperature. The OD₆₀₀ was measured on a Fluostar Omega Plate reader (BMG Labtech) using the well scan setting with a diameter of 3 mm and nine measuring points and double orbital shaking for 30 sec. before the measurement. OD was measured at the beginning of the experiment and again after 24 h and 48 h. The percentage of growth inhibition was calculated, using the median of all measuring points per well, as follows:

$$\% \text{ growth inhibition} = 100 \times (\text{absorbance of control} - \text{absorbance of test}) / \text{absorbance of control}$$

Bacteria were tested in a similar way. Test organisms were *Agrobacterium tumefaciens* GV3101 and *Pseudomonas syringae* pv. *tomato* DC3000. Bacteria were grown in 5 ml LB medium over night at 30°C. The cultures were centrifuged for 10 min at 7000 rpm and resuspended in 1% tryptone. Bacteria were diluted to an OD₆₀₀ of 0.05 in 1% tryptone before using them for the test. The peptide was pipetted into a 96 well plate as described above and 75 µl of bacterial culture was added. OD₆₀₀ was measured in a Fluostar Omega plate reader as described above. Plates were incubated in the dark at 30°C and measured again after 24h. Percentage of growth inhibition was calculated as described above.

Results

New thionins discovered by the 1KP project

The 1KP project sequenced the transcriptomes of more than 1000 plant species including algae, mosses, ferns, gymnosperms and angiosperms [51,62,63]. We used these data to search for thionin sequences in the translated protein sequences. Our search produced 225 hits (Table 1) of which some resulted in the same sequences because the 1KP project often sequenced different tissues from the same species. We thus ended up in a total of 133 thionin sequences (Table 2).

No thionin sequences were found in the transcriptomes of algae and mosses. The phylogenetically oldest hits were a class 1 thionin sequence from *Selaginella wallacei* and from *Selaginella stauntoniana* (phylum *Lycophyta*) and two class 2 thionin sequences from the fern *Cystopteris fragilis*, indicating that thionins have perhaps evolved in the first vascular plants. Furthermore, class 1 thionin sequences were found in the conifers *Manoao colensoi*, *Sequoia dendron giganteum* and *Athrotaxis cupressoides* (two hits). Already these phylogenetically oldest thionins were produced as preproteins including a typical pro-domain (“acidic” domain).

The 1KP project had reported that some data have to be treated with caution (<https://cyverse.atlassian.net/wiki/spaces/iptol/pages/242170988/Sample+source+and+purity>), including the *Cystopteris fragilis* XXHP sample that was flagged as “Worrisome Contamination”. The two hits for *Cystopteris fragilis* were from fronds (sample XXHP). A separate tissue sample from young *Cystopteris fragilis* leaves (not reported to be problematic) contained no hits. The 1KP project had also sequenced the transcriptomes of three other *Cystopteris* species and these contained also no hits for thionins. We tried to confirm the thionin sequences in the genome of *Cystopteris fragilis* by PCR but were unable to find these sequences. We must therefore conclude that the thionin sequences from *Cystopteris fragilis* were probably a contamination. In case of the *Selaginella* thionins, the hits were from two different species (in total, transcriptomes of 8 species were sequenced). In this case we had no plant material available to confirm the sequence by PCR.

In total, 53 different class 1 sequences and 62 class 2 sequences were found, many of them in plant families of angiosperms from which previously no thionins had been reported (Fig 2). These class 1 and class 2 thionins had an average molecular weight of 4939.41 and 4959.49 Da, respectively. In general, there was more variability within the class 2 thionins than within the class 1 thionins (Table 2 and Fig 3). Of the 50 class 1 sequences that were not truncated at the N-terminus, four started with arginin as the first amino acid and one with isoleucine. All others had lysin as the first amino acid. In case of the class 2 thionins only 40 of 59 N-terminally untruncated sequences had lysin as first amino acid. In addition, seven sequences had an additional amino acid at the N-terminus, thus the double cysteine was at position 4/5 instead of 3/4 as determined by SignalP. One thionin had a N-terminal sequence of KSKKACC.

The average pI was 9.18 and 8.47 for class 1 and class 2 thionins, respectively. Thus, the class 1 thionins are more basic with no member having a pI below 7 while there were eight class two members with an acidic pI. The lowest pI of 4.04 was found for AaTHI2.3 from *Arabis alpina*. There were two more thionins from this species with an acidic pI while three had a basic pI. All *A. alpina* hits were from young leaves. The most basic thionin with a pI of 11.17 was also a class 2 thionin (DsTHI2.1 from *Draba sachalinensis*) with nine basic residues and no acidic residue. In addition to class 1 and class 2 thionins, variants were discovered that were previously unknown and which could have been evolved from class 1 (seven) or class 2 [11] thionins by deletion or insertion of cysteine residues. Some of these variants had an uneven number of cysteine residues (Table 2).

Table 1. Total hits obtained.

species/genus	family	order	Class I/II hits		
			I	II	both
<i>Selaginella wallacei</i>	Selaginellaceae	Selaginellales	1		1
<i>Selaginella stauntoniana</i>	Selaginellaceae	Selaginellales	1		1
<i>Cystopteris fragilis</i>	Cystopteridaceae	Polypodiales		2	2
<i>Manoao colensoi</i>	Podocarpaceae	Coniferales	1		1
<i>Sequoiadendron giganteum</i>	Cupressaceae	Coniferales	1		1
<i>Athrotaxis cupressoides</i>	Cupressaceae	Coniferales	2		2
<i>Papaver</i>	Papaveraceae	Ranunculales	72	21	93
<i>Thalictrum thalictroides</i>	Ranunculaceae	Ranunculales	1	2	3
<i>Hakea prostrata</i>	Proteaceae	Proteales		2	2
<i>Oresitrophe rupifraga</i>	Saxifragaceae	Saxifragales	3		3
<i>Ribes aff. giraldii</i>	Grossulariaceae	Saxifragales	1		1
<i>Sesuvium</i>	Aizoaceae	Caryophyllales		10	10
<i>Cypselea humifusum</i>	Aizoaceae	Caryophyllales		3	3
<i>Trianthema portulacastra</i>	Aizoaceae	Caryophyllales		6	6
<i>Zaleya pentandra</i>	Aizoaceae	Caryophyllales		1	1
<i>Aerva</i>	Amaranthaceae	Caryophyllales		8	8
<i>Alternanthera caracasana</i>	Amaranthaceae	Caryophyllales		2	2
<i>Polygonum convolvulus</i>	Polygonaceae	Caryophyllales	1		1
<i>Polycarpha repens</i>	Caryophyllaceae	Caryophyllales		3	3
<i>Portulaca mauii</i>	Portulacaceae	Caryophyllales		2	2
<i>Mirabilis jalapa</i>	Nyctaginaceae	Caryophyllales		1	1
<i>Sassafras albidum</i>	Lauraceae	Laurales	1		1
<i>Lindera benzoin</i>	Lauraceae	Laurales	1		1
<i>Cannabis sativa</i>	Cannabaceae	Rosales		7	7
<i>Humulus lupulus</i>	Cannabaceae	Rosales		4	4
<i>Urtica dioica</i>	Urticaceae	Rosales		2	2
<i>Linum perenne</i>	Linaceae	Malpighiales		2	2
<i>Chrysobalanus icaco</i>	Chrysobalanaceae	Malpighiales	1		1
<i>Viola tricolor</i>	Violaceae	Malpighiales	1		1
<i>Brassica nigra</i>	Brassicaceae	Brassicales		3	3
<i>Arabis alpina</i>	Brassicaceae	Brassicales		6	6
<i>Sinapis alba</i>	Brassicaceae	Brassicales		3	3
<i>Draba</i>	Brassicaceae	Brassicales	3	6	9
<i>Cochlearia officinalis</i>	Brassicaceae	Brassicales		1	1
<i>Daphne geraldii</i>	Thymelaeaceae	Malvales	2		2
<i>Sarcodes sanguinea</i>	Ericaceae	Ericales	1		1
<i>Eleusine coracana</i>	Poaceae	Poales	4		4
<i>Panicum miliaceum</i>	Poaceae	Poales	1		1
<i>Neurachne</i>	Poaceae	Poales	3	3	6
<i>Thyridolepis</i>	Poaceae	Poales	2	2	4
<i>Lepidosperma gibsonii</i>	Cyperaceae	Poales		1	1
<i>Polyscias fruticosa</i>	Araliaceae	Apiales	1		1
<i>Psychotria ipecacuanha</i>	Rubiaceae	Gentianales		2	2
<i>Myristica fragrans</i>	Myristicaceae	Magnoliales		3	3
<i>Hydrocotyle umbellata</i>	Araliaceae	Apiales		1	1
<i>Prunella vulgaris</i>	Lamiaceae	Lamiales		1	1

(Continued)

Table 1. (Continued)

species/genus	family	order	Class I/II hits		
			I	II	both
<i>Teucrium chamaedrys</i>	Lamiaceae	Lamiales		1	1
<i>Orobanche fasciculata</i>	Orobanchaceae	Lamiales	3		3
<i>Schlegelia parasitica</i>	Schlegeliaceae	Lamiales	2		2
<i>Heliotropium texanum</i>	Boraginaceae	Boraginales (Lamiales)		2	2
<i>Phoradendron serotinum</i>	Santalaceae	Santalales		2	2
			110	115	
Total hits					225

<https://doi.org/10.1371/journal.pone.0254549.t001>

A large gene family for thionins in *Papaver* species

Our results revealed a large gene family for thionins in *Papaver* species. Different tissues of several *Papaver* species resulted in a total of 93 hits. We found 11 class 1 thionins for *P. rhoeas*, six for *P. somniferum*, three for *P. setigerum* and seven for *P. bracteatum* (Table 2). In addition to class 1 thionins, all four *Papaver* species contained a few class 2 thionins, two in *P. rhoeas* and one each in the other three species. While the majority of *Papaver* thionins had the usual acidic domain, there were a few class 1 thionins from *P. somniferum* and *P. setigerum* which had an unusual pro domain (Fig 4). We found three hits each in both species. These were from developing fruits, leaves and stems in case of *P. setigerum* while the *P. somniferum* hits were from flower buds, leaves and “five samples combined”. Following the thionin domain (pI 8.63) was a sequence with six cysteine residues and a pI of 8.89. Up to there all sequences were identical, which holds also for the DNA sequences. Following this were extensions with several cysteine residues which were rich in proline and serine residues. However, they were all truncated with the exception of EPRK-20112 from *P. setigerum* with a total of 496 amino acids. This extension contained 21 cysteine residues and was also basic with a pI of 8.53.

All *Papaver* species contained in addition variants (Table 2) of class 1 thionins with most of them having an uneven number of cysteine residues. We found variants with seven cysteines in *P. rhoeas*, *P. somniferum* and *P. bracteatum*. *P. setigerum* had one variant with nine cysteines. One *P. setigerum* thionin contained six cysteine residues, thus, seemingly being a class 2 member. However, closer inspection revealed a deletion of two cysteines from a class 1 thionin.

From other thionin genes it is known that the acidic domain contains two small introns (S3 Fig). The transcriptome data from the IKP project give of course no indication about introns. We have therefore cloned a genomic sequence from *Papaver somniferum* as described in Materials and methods. We derived the primer sequences from the DNA sequence of the *P. somniferum* clone BMRX-2017616 which encodes PsoTHI1.5. The genomic clone that we obtained coded for a typical thionin preprothionin and contained two small introns within the region of the acidic domain (Fig 5). The thionin sequence had two mismatches with PsoTHI1.5 and one with PsoTHI1.6 (Table 2) and was named PsoTHI1.7. It has a pI of 9.54 and a Mw of 4859.75 while the acidic domain has a pI of 4.26 and a Mw of 7724.47. The PsoTHI1.7 pI is similar to that of the other class 1 *Papaver* thionins (Table 2).

Basic thionins have repeatedly been reported to have antimicrobial activity. Having cloned the PsoTHI1.7 we used this sequence to produce an expression vector for *E. coli*. We produced PsoTHI1.7 as a fusion protein with thioredoxin, purified the fusion protein and cleaved it with TEV protease. Finally, PsoTHI1.7 was purified as described in Methods. MALDI-TOF mass spectroscopy showed that the size of the purified peptide corresponded to an oxidized peptide

Table 2. All thionin sequences from the 1 KP project.

name	Species	class	sequence	hits	pI	Mw (Da)
HvTHII.1 ¹	<i>Hordeum vulgare</i>	1	KSCCRSTLGRNRYNLCRVRGAQKLCAGVCRCKLTSSTGKCPGTFPK	-	9.75	4855.81
CaTHII.2.1 ²	<i>Crambe abyssinica</i>	2	TTCCPSIVARSNFNVRLPPTSEAIICATYTGIIIPGATCPGDVAN	-	5.73	4726.42
VaTHII.2.1 ³	<i>Viscum album</i>	2	KSCCPNTTGRNRYNACRLTGAPRPTCAKLSGCKIISGTCFSDYPK	-	9.30	4835.60
AtTHII.2.1	<i>Arabidopsis thaliana</i>	2	KICCPNQARNYSVCRIRFSKGRCMQVSGCQNSDTCPRGMVN	-	9.43	4811.52
TaTHII.3.1 ⁴	<i>Triticum aestivum</i>	3	VDCGANPFKAVFNSLLGPFSTVFQCADFCARLPAG	-	5.90	3826.47
TgTHII.4.1	<i>Tulipa gesneriana</i>	4	KSCFPSTAAKYCNACRLPGCRPETIICAAACGCKIISGNCPPGYDY	-	8.71	5040.85
SwTHII.1 ⁹	<i>Selaginella wallacei</i>	1	KSCCPSTAAARNYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	8.21	4769.40
SsTHII.1 ⁶	<i>Selaginella stauntoniana</i>	1	KSCCPSTAAARNYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	-	-
McTHII.1 ¹⁰	<i>Manoao colensoi</i>	1	KSCCPSTAAARNYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	8.51	4810.50
SgTHII.1.1	<i>Sequoiadendron giganteum</i>	1	KSCCKNTTGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	9.25	4860.65
AcTHII.1.1 ¹⁰	<i>Athrotaxis cupressoides</i>	1	KSCCPSTAAARNYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	2	8.51	4810.50
OrTHII.1.1 ¹¹	<i>Oreolithrophe rupifraga</i>	1	KSCCVNTTARNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	8.21	4857.60
OrTHII.2.6	<i>Oreolithrophe rupifraga</i>	1	XSCCPSTAAARNYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	-	-
OrTHII.3.6	<i>Oreolithrophe rupifraga</i>	1	ARNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	-	-
VtTHII.1 ⁹	<i>Viola tricolor</i>	1	KSCCPSTAAARNYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	8.21	4769.40
SpTHII.1.1 ^{6,9}	<i>Schlegelia parasitica</i>	1	KSCCPSTAAARNYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	-	-
RgTHII.1.1 ⁶	<i>Ribes giraldii</i>	1	TTARNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	2	-	-
DgTHII.1.1	<i>Daphne gerardii</i>	1	KSCCRNTLGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	-	-
SaSTHII.1.1 ⁶	<i>Sarcodes sanguinea</i>	1	KSCCKNTTGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	2	8.76	4976.68
PrTHII.1.1	<i>Papaver rhoeas</i>	1	RSCCDSKAGRNYQACIRRTGATQLCARSCGGFTRENRCFSSHPW	1	-	-
PrTHII.2.2	<i>Papaver rhoeas</i>	1	KSCNTTIKRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	2	9.37	5197.91
PrTHII.3.3	<i>Papaver rhoeas</i>	1	KSCCKTTIKRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	9.08	5003.87
PrTHII.4.4	<i>Papaver rhoeas</i>	1	KSCCKTTIKRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	2	9.35	5091.00
PrTHII.5.5	<i>Papaver rhoeas</i>	1	KSCCRNTTGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	2	9.06	5041.88
PrTHII.6.6	<i>Papaver rhoeas</i>	1	RICCKDSVARSFNSQPGTPRVSVCATTCRCRTISGLCPSSYPS	6	8.97	4703.43
PrTHII.7.7	<i>Papaver rhoeas</i>	1	KICCKSTTARSQFACRIRLSRETCASTCSCKIILGNCPSDYK	1	9.25	4835.72
PrTHII.8.8	<i>Papaver rhoeas</i>	1	KPCCTYVNGKKNRWDGNGKDSKADMGCKSCPRIEVG	1	8.72	4503.17
PrTHII.9.9	<i>Papaver rhoeas</i>	1	KSCCKSTVGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	6	9.54	4859.75
PrTHII.10.10	<i>Papaver rhoeas</i>	1	KSCCKSTLGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	9.74	4905.86
PrTHII.11.11	<i>Papaver rhoeas</i>	1	KSCCKSTLGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	9.80	4970.94
PsoTHI.1.1	<i>Papaver somniferum</i>	1	KSCCKSTVGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	3	9.70	4918.82
PsoTHI.2.2	<i>Papaver somniferum</i>	1	RSCCGSKAGRNYQACVRRGTGATQLCARSCGGFTRENRCFSSHPW	2	9.60	5125.85
PsoTHI.3.3	<i>Papaver somniferum</i>	1	KSCCKTTAARNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	5	9.38	4934.77
PsoTHI.4.4	<i>Papaver somniferum</i>	1	KNCKTAFGRHRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	9.30	5295.15
PsoTHI.5.5	<i>Papaver somniferum</i>	1	KSCCKSTVGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	1	9.75	4946.83
PsoTHI.6.6 ¹²	<i>Papaver somniferum</i>	1	KSCCKSTLGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	4	9.58	4901.79
PsoTHI.7.7 ¹³	<i>Papaver somniferum</i>	1	KSCCKSTVGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	6	9.58	4887.76
PseTHI.1.1	<i>Papaver setigerum</i>	1	KSCCKSTVGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	2	9.54	4859.75
PseTHI.2.2 ¹²	<i>Papaver setigerum</i>	1	KSCCKSTVGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	2	9.75	4946.83
PseTHI.3.3	<i>Papaver setigerum</i>	1	KSCCKSTVGRNRYNACRLVETSQTTCASLGGCIHVDGNTCPPNYP	2	9.58	4887.76
PbTHI.1.1	<i>Papaver bracteatum</i>	1	RSCCDSKAGRNYQACVRRGTGATQLCARSCGGFTRENRCFSSHPW	2	9.86	5008.90
				2	9.33	5128.84

(Continued)

Table 2. (Continued)

name	Species	class	sequence	hits	pI	Mw (Da)
PbThi1.2	<i>Papaver bracteatum</i>	1	KSCQNTIARNYVCRVAFAGSREACAKACNCKIIITETDCPSDYK	2	8.50	5037.76
PbThi1.3	<i>Papaver bracteatum</i>	1	KSCCKSASGRRCYVCRVFRQNCGAICNCAIHWASNTCFYFAPY	1	9.27	5270.14
PbThi1.4	<i>Papaver bracteatum</i>	1	KSCCKSTFRNCYNACRLKFKPKTCASTCSCKIVGNRCPFGYK	2	9.65	4935.85
PbThi1.5	<i>Papaver bracteatum</i>	1	KSCCKSTLGRNCYNACRLKFKPKTCASTCNCKILKGNKCFSGYK	1	9.70	4975.95
PbThi1.6	<i>Papaver bracteatum</i>	1	KSCCKSTYNACRLKFKPKTCASTCNCKILKGNKCFSGYK	1	9.57	4535.45
PbThi1.7	<i>Papaver bracteatum</i>	1	KSCCKSTAAECYNACHSAGAPRYVYICKLIIISGKCPAYRY	7	9.03	5051.96
SaThi1.1	<i>Sassafras albidum</i>	1	KSCCRSTTARNYVCRVLSGSSRPTCASLDCCKIITGTTCPSDYK	1	8.92	4952.69
LbThi1.1	<i>Lindera benzoin</i>	1	KSCCRSTTARNYVCRVLAETPRETCAKLCDCIITGTTCPSGYK	1	8.92	4960.80
ClThi1.1 ¹¹	<i>Chrysobalanus icaco</i>	1	KSCCVNTTARNYVCRVLTQAFCANLGCCHIDGTTCPDPYK	1	8.21	4857.60
PnThi1.1	<i>Panicum miliaceum</i>	1	KSCCKSTIARNYVCRVLRGARSVCATTCGCKIIGKTKCPFGYK	1	9.65	4840.83
PfThi1.1 ¹¹	<i>Polyscias fruticosa</i>	1	KSCCVNTTARNYVCRVLTQAFCANLGCCHIDGTTCPDPYK	1	8.21	4857.60
TlThi1.1	<i>Thalictrum thalictroides</i>	1	KSCCPGTIQRNCYNLCRVGGKITSETCAKTCGCKHVGVYCPGQWS	1	9.10	5030.90
DsThi1.1	<i>Draba sachalinensis</i>	1	KSCCPSTIARNYVCRVLTQAFCANLGCCHIDGTTCPDPYK	3	8.95	4760.48
EcThi1.1	<i>Eiulus coracana</i>	1	ISCCPDTTARNYVCRVLSGSSRPTCASLDCCKIISGNTCPGFPFS	4	9.05	5053.01
NaThi1.1	<i>Neurachne alopecuroidea</i>	1	KSCCRSTTARNYVCRVLRGARSVCATTCGCKIIGKTKCPFGYK	1	9.63	5038.97
NmThi1.1	<i>Neurachne munroi</i>	1	KSCCKNTAGRNYNICRRAGGSQVARRCGCIITGNRCPNYPK	1	9.63	5035.89
TmuThi1.1 ⁷	<i>Thyridolepis multiculmis</i>	1	KSCCKSTWARNYNICRFKGPLVCAQMGCKIIGQKCFSDFFK	1	9.38	4976.03
TmiThi1.1 ⁷	<i>Thyridolepis mitchelliana</i>	1	KSCCKSTWARNYNICRFKGPLVCAQMGCKIIGQKCFSDFFK	1	9.38	4976.03
OfThi2.1	<i>Orobancha fasciculata</i>	1	KSCCEDTARYVCRVLTQAFCANLGCCHIDGTTCPDPYK	3	8.72	5052.85
		53			9.18	4939.41
PiThi2.1	<i>Psychotria ibecacuanha</i>	2	TTSCCPSAYARSTYNLCSLYKSOIICARLSGCILIDGTSFPSNYPK	2	8.62	4954.71
MfThi2.1	<i>Myristica fragrans</i>	2	ESCCPSAKALNYVCRVQSDPHYFTKSFANLACGLADGKCEPPYDH	2	8.31	5717.46
CfThi2.1	<i>Cystopteris fragilis</i>	2	KSCCPTIIVARNQSVCRVAFAGSREACAKLGLGCKIVDGECPGGYNR	1	8.90	4796.53
CfThi2.2	<i>Cystopteris fragilis</i>	2	KTCCPSTARSITRTCRFGGSGTQCAQISGCKIVSCEGCPGGYK	1	9.13	4608.25
HlThi2.1 ⁶	<i>Heliotropium texanum</i>	2	KSCCPTTARNYVCRVLAETPREVFCASISGCKIITGKCPKGY	2	-	-
MjThi2.1	<i>Mirabilis jalapa</i>	2	KSCCPTTARNYVCRVLAETPREVFCASISGCKIITGKCPKGYEK	1	9.46	4863.67
HpThi2.1 ¹⁴	<i>Hakea prostrata</i>	2	KVACCPSTAAASNYISICRLYGASGPKAKIEDCKIIVDGECPGSGTYP	2	6.24	4962.69
LgThi2.1 ¹⁴	<i>Lepidosperma gibsonii</i>	2	KVACCPSTAAASNYISICRLYGASGPKAKIEDCKIIVDGECPGSGTYP	1	6.24	4962.69
TcThi2.1	<i>Teucrium chamaedrys</i>	2	KSCCPTTARNYVCRVLAETPREVFCASISGCKIIVDGECPGSGTYP	1	9.13	4754.49
PrThi2.1	<i>Papaver rhoeas</i>	2	TSCCPSAYARSTYNLCSLYKSOIICARLSGCILIDGTSFPSNYPK	7	8.62	4853.61
PrThi2.2	<i>Papaver rhoeas</i>	2	KICOMDTRRNRYKDCINTGASVTSAGVSGCLIVSGLCPNYPY	4	8.65	4932.69
PsoThi2.1 ⁵	<i>Papaver somniferum</i>	2	TSCCESTKARNYSVCRVLRGASKNCAKLTGCIIDGTSFPSDYPI	3	8.63	4887.63
PseThi2.1 ⁵	<i>Papaver setigerum</i>	2	TSCCESTKARNYSVCRVLRGASKNCAKLTGCIIDGTSFPSDYPI	3	8.63	4887.63
PbThi2.1	<i>Papaver bracteatum</i>	4	KICOMDTRRNRYEVCINTGASVTSAGVSGCLIVSGLCPNYPYH	4	8.34	4861.61
CoThi2.1	<i>Cochlearea officinalis</i>	2	TLCCPNKKTADIYACTRTSGVSKYMCERLSGCKNVSGTDCPDILQN	1	8.62	4874.65
HuThi2.1	<i>Hydrocotyle umbellata</i>	2	KSCCPTTARNYVCRVLRGASVTSAGVSGCLIVSGLCPNYPYH	1	8.96	4818.48
PvThi2.1	<i>Prunella vulgaris</i>	2	KSCCPTTARNYVCRVLRGASVTSAGVSGCLIVSGLCPNYPYH	1	8.66	4865.63
TlThi2.1	<i>Thalictrum thalictroides</i>	2	KSCCPTTARNYVCRVLAETPRETCAKLSGCKIIGKTKCPFGYK	1	9.06	4985.78
TlThi2.2	<i>Thalictrum thalictroides</i>	2	YVVCCKNIQARNYVCRVLAETPRETCAKLSGCKIIGKTKCPFGYK	1	9.15	5326.08
SvThi2.1	<i>Sesuvium ventricosum</i>	2	KSCCPTTARNYVCRVLAETPRETCAKLSGCKIIGKTKCPFGYK	1	9.46	4897.77
ChThi2.1	<i>Cypseleia humifusum</i>	2	KSCCPTTARNYVCRVLAETPRETCAKLSGCKIIGKTKCPFGYK	1	9.64	5587.53

(Continued)

Table 2. (Continued)

name	Species	class	sequence	hits	pI	Mw (Da)
ChTHI2.2 ⁶	<i>Cypseleia humifusum</i>	2	LSCCPSTTSVRHVFSSCRLAGGSRVSCASLSGCKIVSGATCPRAYDK	1	9.18	4810.55
TpTHI2.1	<i>Trianthema portulacastrum</i>	2	KSCCPSTTARNYVYVRLAGTPRFVCAISGKIIITGKCPKGYEK	1	9.46	4865.71
TpTHI2.2	<i>Trianthema portulacastrum</i>	2	KSCCPNTTARNIYNTCRAGGSRFQCSLSGCIISGNKCPSTHPK	1	9.38	4792.45
TpTHI2.3	<i>Trianthema portulacastrum</i>	2	KSCCRSTTARNIYVCRFAGGTQSCATVSGKIIISGNKCPSDYPK	2	9.30	4863.53
TpTHI2.4	<i>Trianthema portulacastrum</i>	2	KSCCRSTTARNIYVCRFAGGQQCATVSGKIIISGNKCPSDYPK	1	9.30	4819.47
ApTHI2.1	<i>Aerva persica</i>	2	KSCCPSTSARNIYNTCRFTGGSRMCAVSGKIIKGTCKPPGYEK	2	9.27	4935.74
AcTHI2.1 ⁶	<i>Alternanthera caracasana</i>	2	XSCCPSTSARNIYVCRLLGGASREVCAKLSGCKHMDACKPPYIH	2	--	--
PrTHI2.1	<i>Polycarpaea repens</i>	2	KSCCPNTTARNIYNAARLTGASREFCANLSGKIIIDVTTCPSDYPS	1	8.33	4946.61
PrTHI2.2	<i>Polycarpaea repens</i>	2	KSCCPTEKARISYMFCLKSEVNDQALCASASGCKIAKDRICPPGYDH	1	8.30	5139.96
CsTHI2.1	<i>Cannabis sativa</i>	2	KSCCPSTTARNIYNTCRFGGSRRETCAISGRIITGNTCPGGWTN	5	8.96	4803.39
CsTHI2.2	<i>Cannabis sativa</i>	2	KSCCPSTTARNIYNTCRFGGSRRETCAISGRIITGNTCPGGNDH	1	8.69	4768.30
CsTHI2.3	<i>Cannabis sativa</i>	2	KSCCPSTTARNIYNAARLTGDSRKFCAISGKIIISGNTCPDGYNH	1	8.90	4929.59
UtTHI2.1	<i>Urtica dioica</i>	2	KSCCPNTYARTSYDFCRSNHGPEIRCSLESGCLIVDNGARCPDYPYI	1	6.71	5210.85
UtTHI2.2	<i>Urtica dioica</i>	2	XSCCPSTTARKIYNTCRFAGGTREFFCASVSGKIVTGKPPHYDK	1	--	--
HiTHI2.1 ^{6,8}	<i>Humulus lupulus</i>	2	KSCCPNSARNIYVCRLLGGIIPQSTCAQLTGQIIISGS	2	--	--
HiTHI2.2 ⁸	<i>Humulus lupulus</i>	2	KSCCPNNSARTTAVVCRFGGASQICASLSGQQLISGTCPSGFTH	2	8.70	4687.26
LpTHI2.1	<i>Linum perenne</i>	2	KSCCPSTTARNIYDAGRAGGSRREFCAKLSGKIIISGNTCPDGYNH	2	8.65	4888.54
BnTHI2.1	<i>Brassica nigra</i>	2	TSCCPSSSARNIYVCRLLPPTPRPICASISGKIVSGTCCPPGYAN	1	8.93	4591.26
BnTHI2.2	<i>Brassica nigra</i>	2	MICCPSTARARYLFDVCSRRRIISIIICSGTGTGCVPIFGDTCPPDYPG	1	7.56	4907.76
AaTHI2.1	<i>Arabis alpina</i>	2	KSCCPSTNARNIYVCRLLVQHPRSVCAKISQCKIVQENTCPCPYPK	1	9.13	5120.99
AaTHI2.2	<i>Arabis alpina</i>	2	FACCPDAAARDIFSECTYSGKVTFCASLSGQQLNRYHNCPPGYTH	1	6.87	5122.75
AaTHI2.3	<i>Arabis alpina</i>	2	WTCCPDAARDMFECTYSGKDTDCALLSDCKLDIWDCCPPGYDH	1	4.04	5180.75
AaTHI2.4	<i>Arabis alpina</i>	2	ATICPPTTKARTAYNFCRLRGTSLLICAYRNGCKINEEDECPPDFPK	1	8.33	5259.03
AaTHI2.5	<i>Arabis alpina</i>	2	IQACCPTRAHKIFDKICGGYKGYRCAVSVGCLIIENNCPPGFTH	1	7.78	5144.97
AaTHI2.6	<i>Arabis alpina</i>	2	AVCCPDRYARNFYNFLQQLGEIYLPYTCISLTKCCQIIFMCPKDTDD	1	4.78	5154.97
SaTHI2.1	<i>Sinapis alba</i>	2	KSCCPSTIARNIYVCRFTGSSRQTCARLSGCTIVSGTTCPPQGYPH	1	9.18	4914.57
SaTHI2.2	<i>Sinapis alba</i>	2	RVCCPNNDLDRKFDLCSATGFPKSAKARLSECKIIFGNTCPSDFPA	1	8.66	5009.80
SaTHI2.3	<i>Sinapis alba</i>	2	KTCPPSTTARNIYDRCLAIGIPMIICLDITGCVIISTRPTCPNPPFA	1	8.35	4929.90
DsTHI2.1	<i>Draba sachalinensis</i>	2	KTCCTRIVARNIYVCRFTGSSIVFCVQSSRGLCRIFPGRPCRPFFSF	1	11.17	5496.48
DhTHI2.1	<i>Draba hispida</i>	2	RTCCLTVADRDFISECTYSGKIMTFCASLSGCKLMPRYSCPPGYIH	1	8.32	5182.04
DoTHI2.1	<i>Draba oligosperma</i>	2	KYCCVSTNARKNFNTCLSKGIPSTTCAYDSGCKFTYNTCPGGYAD	1	8.62	4996.62
DaTHI2.1	<i>Draba aizoides</i>	2	KSCCPSTSARNIYNTCRFAGGSRRETCAVSGKIVSGTCCPPGFTH	1	9.18	4611.26
DmTHI2.2 ^{6,8}	<i>Draba magellanica</i>	2	KSKKACCPSYDARNYVCRIRLQNPDCASLSECLIIDNGNCPPAG	2	--	--
NiTHI2.1	<i>Neurachne lanigera</i>	2	NVCCPSKLTTRVYVCRVYVSGTECAEWTTCKVVDGECKPPYDH	1	6.72	5155.86
NaTHI2.1 ¹⁴	<i>Neurachne alopecuroidea</i>	2	KVACCPSTTARNIYVCRVYVSGTECAEWTTCKVVDGECKPPYDH	1	6.24	4962.69
TmiTHI2.1	<i>Thyridolepis mitchelliana</i>	2	KSCCPSTTARNIYVCRVYVSGTECAEWTTCKVVDGECKPPYDH	1	9.10	5046.87
TmuTHI2.1	<i>Thyridolepis multicaulis</i>	2	KSCCPSTTARNIYVCRVYVSGTECAEWTTCKVVDGECKPPYDH	1	9.13	4955.71
MfTHI2.2	<i>Myristica fragrans</i>	2	KSCCPSTTARNIYVCRVYVSGTECAEWTTCKVVDGECKPPYDH	1	8.89	5011.82
PsTHI2.1	<i>Phoradendron serotinum</i>	2	KVCCPSTTARNIYVCRVYVSGTECAEWTTCKVVDGECKPPYDH	1	8.94	4661.51
PsTHI2.2	<i>Phoradendron serotinum</i>	2	KVCCPSTTARNIYVCRVYVSGTECAEWTTCKVVDGECKPPYDH	1	8.68	4452.16
PrTHI2.3 ⁶	<i>Polycarpaea repens</i>	2	XXXXXXXXXXRNINACRLAGGSRERCAVSGKIVSGTTCPSLFPN	1	--	--

(Continued)

Table 2. (Continued)

name	Species	class	sequence	hits	pI	Mw (Da)
		62			8.47	4959.49
PpTHI-BEKN-18329*	<i>Papaver rhoeas</i>	1X	KSCCKSTVGRNCYNACRLKFAKTCCKIVGGNKCPRGYPK	1	9.78	4402.25
PsoTHI-RQNK-5714*	<i>Papaver somniferum</i>	1X	KSCCYNACRLRFARKVCASTCCKIVGGNRCPRGYPK	1	9.73	4100.91
PseTHI-JSYC-2507	<i>Papaver setigerum</i>	1X	KSCCKSTVGRNVGYARQVCASTCCKIVGGNRCPRGYPK	1	9.70	4138.85
PseTHI-QCOU-26075*	<i>Papaver setigerum</i>	1X	KSCCKSTVGRNCYNACRLRFARQVCASTCCKIVGGNRCPRGYPK	1	9.88	5774.83
PbTHI-SSDU-32671*	<i>Papaver bracteatum</i>	1X	KSCCKSTACRLKFPKRTKASTCNCCKILKGNKCPGYPK	1	9.72	4155.03
PcTHI-FYSJ-17384* ⁸	<i>Polygonum convolvulus</i>	1X	TVCCPSTATRNAYTVCVSGGLSSDDCALVCGCITQTSSTCESGYDY	1	4.03	4666.16
NaTHI-ZENX-85534*	<i>Neurachne alopecuroidea</i>	1X	KVCCCKTEIDMYCYDVCHLKRPELCAFLCNCITSIDICPPDHPK	1	5.40	5017.93
		7			8.32	4608.0
SVTHI-EDIT-887*	<i>Sesuvium ventricosum</i>	2X	RICRTPAKTIIYNTCRNAGGSRQMCALSLGCHI VVGPCFVGCNHNLSLQ	3	9.27	5206.14
SVTHI-HZTS-5985*	<i>Sesuvium portulacastrum</i>	2X	KSCCFAGGSRPWCASLSGCKIVNGNKCPGYPH	1	9.04	3468.08
SVTHI-OPZX-76	<i>Sesuvium ventricosum</i>	2X	SISCPFSARNIYNVCRLLGGASKNTCAQLSGCTVTSVS	4	9.30	3892.41
SVTHI-ZJDK-458*	<i>Sesuvium ventricosum</i>	2X	KSCCPTITARNIYNACRVPTPRFVCA SLNEYCKLGCVSSVCGAMVTL	1	8.82	5052.00
ChTHI-GJNX-209	<i>Cypseleia humifusum</i>	2X	VLI SLMMSQFVEGIGISCPASAKNIYIICRVAGGSESACSKLSGCTLTIGK	1	8.50	5183.13
ZpTHI-BERS-225	<i>Zaleya pentandra</i>	2X	QIQVEANIYNTCRFAGGSRDQCAKLSGCKHMDTCVPPYTK	1	8.49	4435.06
AlTHI-EMIG-63* ^{8,15}	<i>Aerva lanata</i>	2X	RTKCCCPDYIATNKYKVCRFAGASRAQACASLTGCIHITESTCPSNNPI	6	8.57	4913.69
BYNZ-2059306* ^{8,15}	<i>Portulaca mauii</i>	2X	RTKCCCPDYIATNKYKVCRFAGASRAQACASLTGCIHITESTCPSNNPI	2	8.81	5170.98
BLWH-2085188*						
NaTHI-ZENX-85558	<i>Neurachne alopecuroidea</i>	2X	KSCCCSTTARNIYNGCRAAGFSREKCA SVSGCKIVGKCKPCCDRFSDS	1	8.91	5225.99
BrTHI-IPWB-14968*	<i>Brassica nigra</i>	2X	RVCPSLSASNVFMVCCI QFPRSACL SVSGCIQVSGNVCFSGYSN	1	8.28	4675.46
TpTHI-QNIK-96290	<i>Trianthema portulacastrum</i>	2X	KSCCPSITARNIYNSCRFTGSSRDMCASLSGCKIVNGSKPCSCPKLAID	1	8.92	5236.08
		11		225	8.81	4769.0
		133				

For comparison, some typical previously known thionins have been included. Shown are the thionin name, plant species, class affiliation, sequence, number of hits, pI and molecular weight (Mw). Cysteine yellow, basic amino acids blue, acidic amino acids red, tyrosine pink, phenylalanine green.

* uneven number of cysteine residues,

¹HTHI,

²Crabbin,

³ViscotoxinA3,

⁴pTTHZ0,

⁵Same sequence, unusual pro-domains!,

⁶Not complete,

⁷Same sequence,

⁸Signal peptide cleavage site unclear,

⁹Same sequence,

¹⁰Same sequence,

¹¹Same sequence,

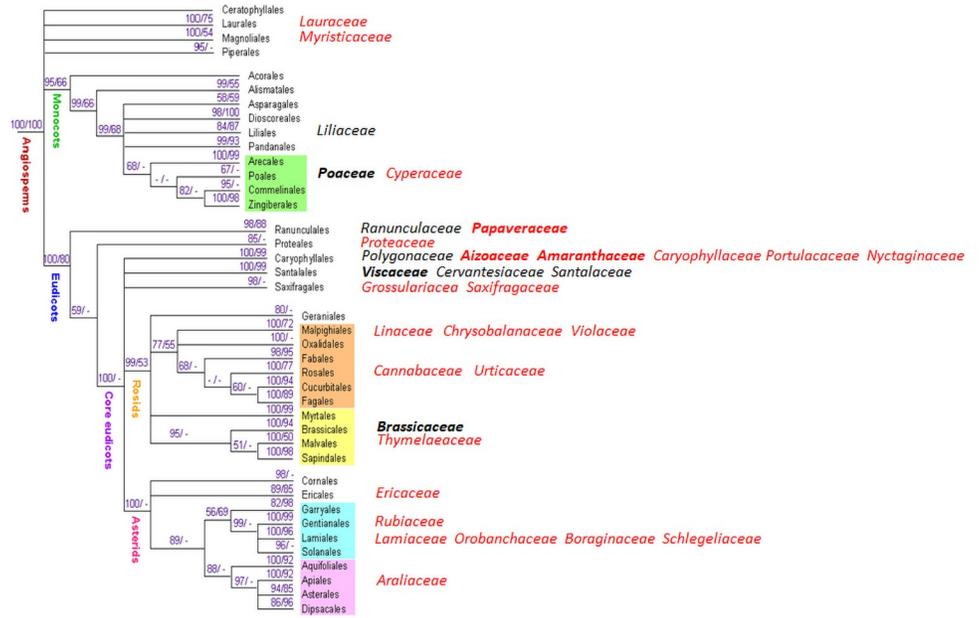
¹²Same sequence,

¹³Genomic sequence,

¹⁴Same sequence,

¹⁵Same sequence.

<https://doi.org/10.1371/journal.pone.0254549.t002>



<https://comenius.susqu.edu/biol/202/archaeplastida/viridiplantae/apg-phylogeny.htm>

Fig 2. Phylogeny of angiosperms showing plant families with species having thionins. Black, families containing previously known thionins; red, families with newly found thionins from the IKP project. Bold, families containing many thionins. Phylogeny modified from <https://comenius.susqu.edu/biol/202/archaeplastida/viridiplantae/apg-phylogeny.htm>.

<https://doi.org/10.1371/journal.pone.0254549.g002>

with all four disulfide bonds. We tested the antimicrobial activity against *Fusarium oxysporum* and *Botrytis cinerea* (Figs 6 and 7). The IC₅₀ was approximately 1 µg/ml for *F. oxysporum* and 6 µg/ml for *B. cinerea* after 48 hours. Microscopic inspection of the hyphae revealed no morphological alterations such as hyperbranching or swelling of hyphal tips (Fig 7). The activity

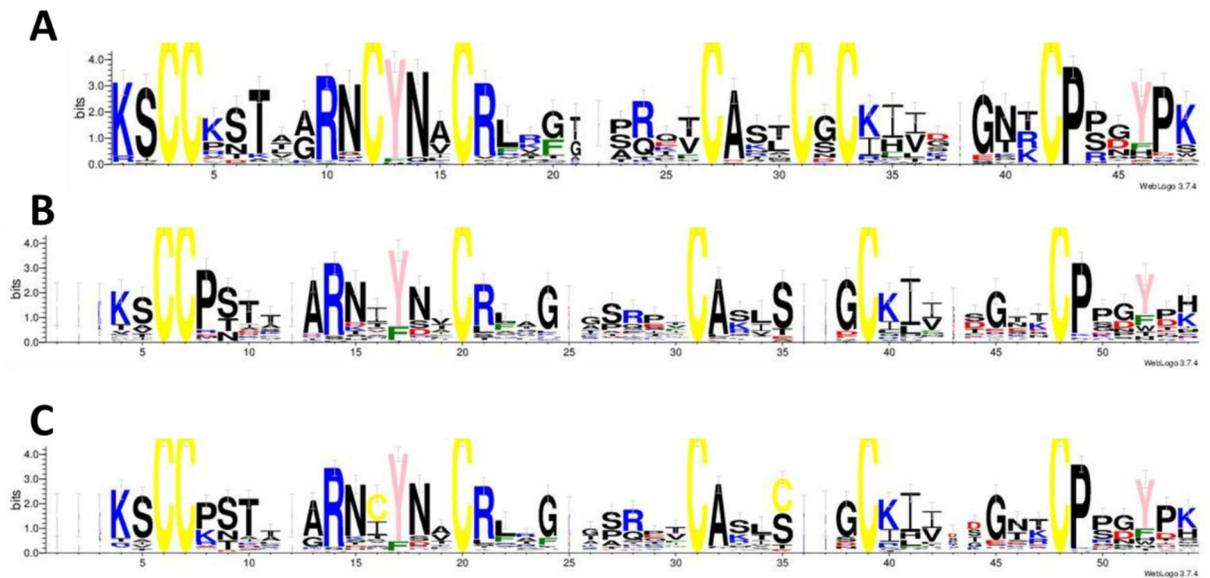


Fig 3. Sequence logos for thionins discovered by the IKP project. A, class 1 thionins; B, class 2 thionins; C, class 1 and class 2 thionins. Cysteine yellow, basic amino acids blue, acidic amino acids red, tyrosine pink, phenylalanine green.

<https://doi.org/10.1371/journal.pone.0254549.g003>

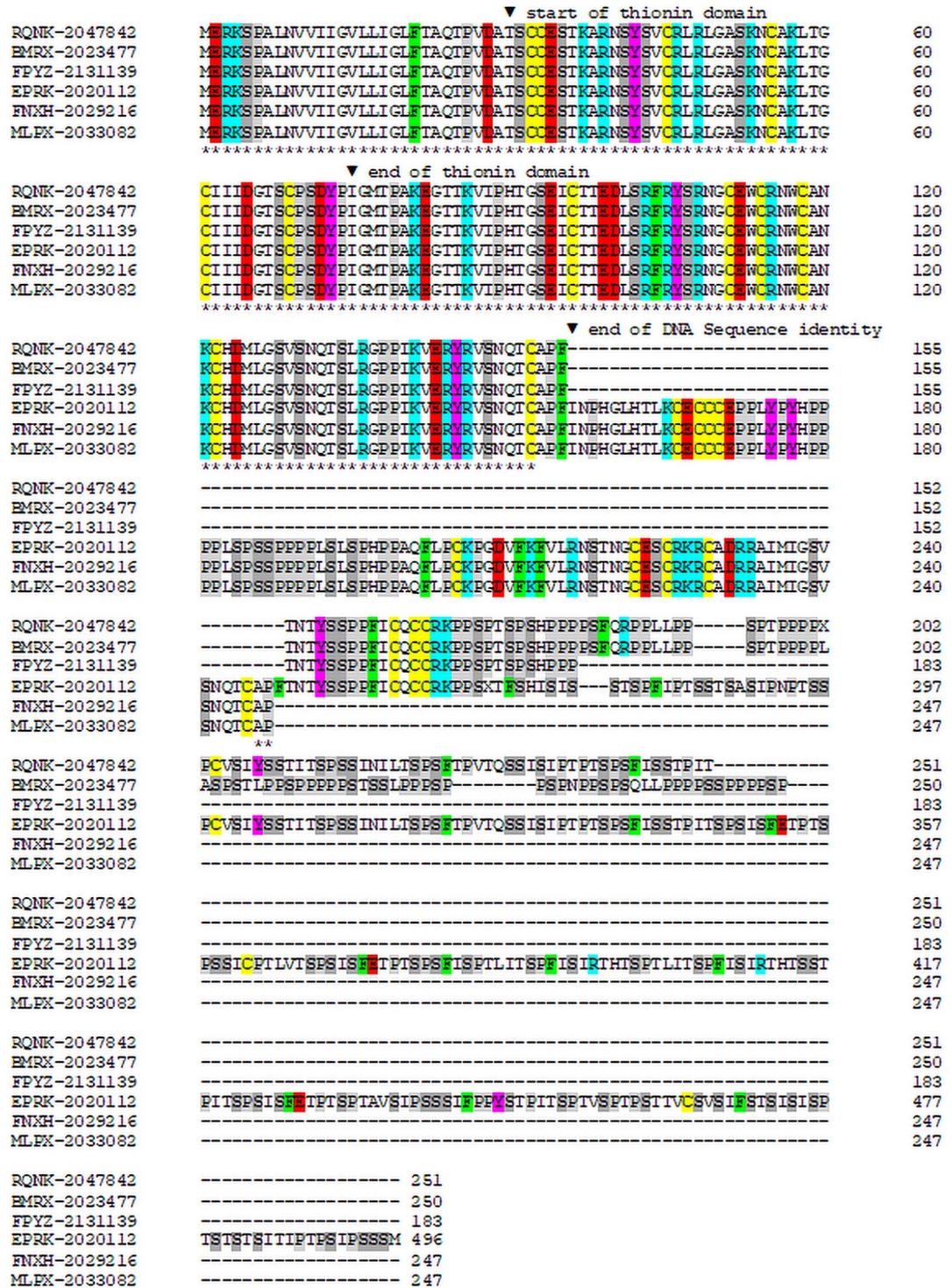


Fig 4. Alignment of *Papaver* thionin preproteins with unusual pro domains. RQNK-2047842, BMRX-2023477 and FPYZ-2131139 are from *P. somniferum* (PsoTHI2.1) while EPRK-2020112, FNXH-2029216 and MLPX-2033082 are from *P. setigerum* (PseTHI2.1). Cysteine yellow, basic amino acids blue, acidic amino acids red, tyrosine pink, phenylalanine green, proline grey, serine dark grey.

<https://doi.org/10.1371/journal.pone.0254549.g004>

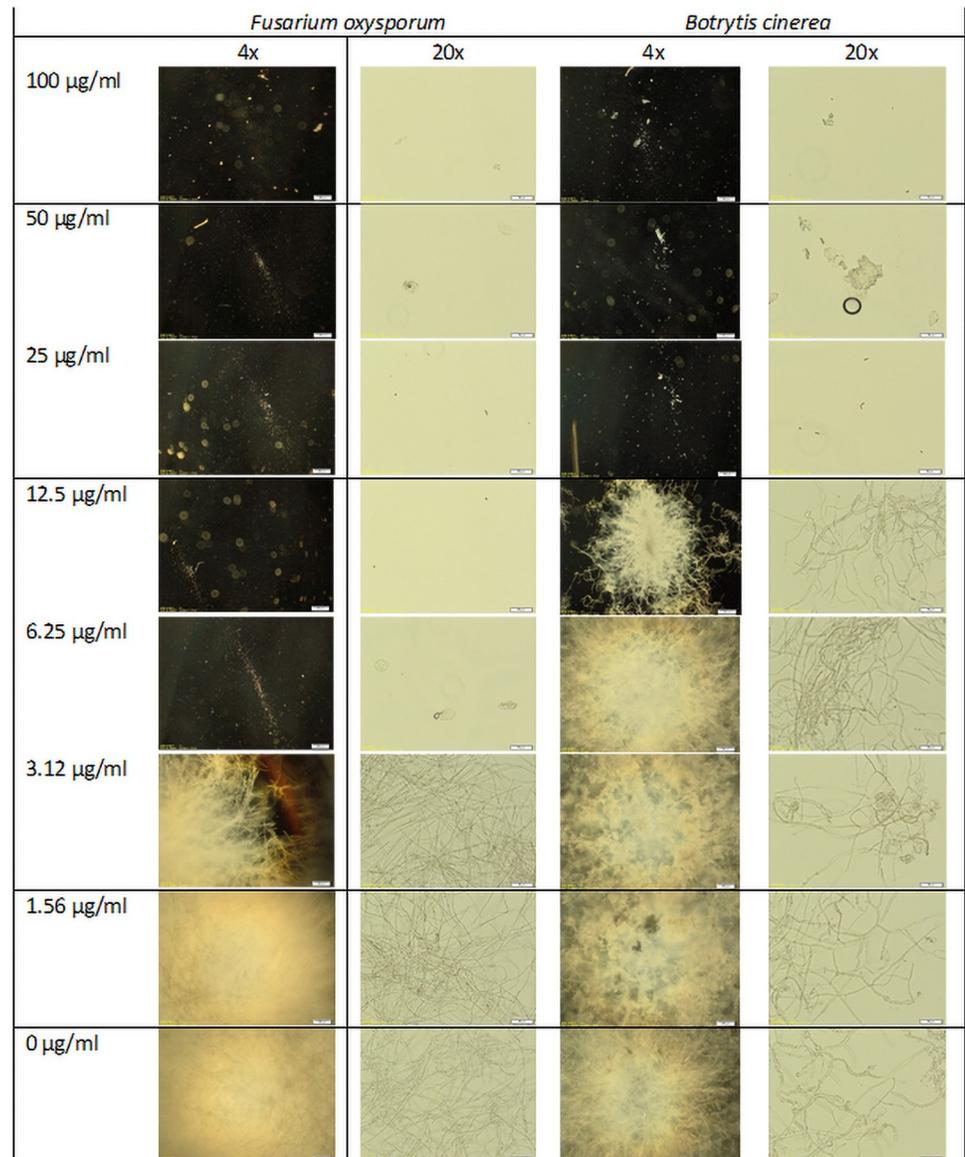


Fig 7. Representative pictures taken from the microtiter plate wells of antimicrobial assays for *B. cinerea* and *F. oxysporum*.

<https://doi.org/10.1371/journal.pone.0254549.g007>

order *Caryophyllales*) had a typical class 2 thionin and in addition three different class 2 variants with a typical acidic domain. Two of them had an uneven number of cysteines resulting from an additional residue. The other one had only four cysteine residues, missing cysteine number one plus another cysteine residue which was hard to define. This thionin was rather small with only 28 amino acids and was basic with a typical acidic domain. A second *Sesuvium* species, *S. portulacastrum*, contained another class 2 variant with a total of only 34 amino acids, this time with five cysteine residues but also with a typical acidic domain. The class 2 variant from *Cypselea humifusum* (Fam. *Aizoaceae*) had a typical acidic domain but was probably truncated at the N-terminus as SignalP did not find a signal peptide.

Several other species of the *Caryophyllales* contained also class 2 variants. *Aerva lanata* and *Portulaca mauii* both contained a variant with three consecutive cysteines at position 4–6

resulting in an uneven number of cysteines. Another class 2 variant with three consecutive cysteines, this time at position 3–5, was found in *Neurachne alopecuroidea*, a perennial grass. This variant has another additional cysteine residue, resulting in an even number of eight cysteines. *N. alopecuroidea* is a species with class 1 and 2 thionins and which contains in addition variants of class 1 (with an uneven number of cysteines due to an additional residue) and the already mentioned variant of class 2.

Discussion

Thionins have originally been isolated as peptides from a number of plant species, especially mistletoes and cereals. Later on, DNA sequences have revealed thionin sequences in various other plant species. However, compared to defensins, there were many plant families for which no thionins had been reported. The IKP project was thus an opportunity to extend the knowledge about the distribution of thionins within plants. Since the IKP project also sequenced a number of algae, mosses, ferns and conifers it was also possible to test if thionins might be found in plant species other than angiosperms. Using various bait sequences we found 225 thionins. The IKP project sequenced cDNAs from specific plant organs but also pooled samples which resulted in multiple hits for some thionins. In addition, some thionins were found in different samples. This reduced the number to a total of 133 different thionins.

We did not find thionin sequences in algae and mosses. Shelenkov et al. [64] also recently mined the IKP data for AMP sequences, including thionins. They reported that they found thionin sequences in algae with 36 thionins in *Chlamydomonadaceae* and 10 in *Desmidiaceae*. We had 31 primary hits in each of these families. However, thorough inspection showed that all these hits did not contain thionins. Since Shelenkov et al. [64] reported only numbers without sequences it is at the moment not possible to resolve these differences.

The phylogenetically earliest hits were from *Selaginella wallacei* and *S. stauntoniana* which belonged to class 1. We had no plant material available to confirm the sequences by PCR. We found four class 1 sequences from conifers, one from *Manoao colensoi* (Fam. *Podocarpaceae*) and the same sequence with two hits from *Athrotaxis cupressoides* (Fam. *Cupressaceae*). Furthermore, another species from the family *Cupressaceae*, *Sequoiadendron giganteum*, also had one hit. All other sequences were from angiosperms. Up to the IKP project there were eight plant families of the angiosperma which were known to contain thionins. The IKP project identified 25 plant families which contained previously no known thionins. However, there are still many plant families which do not seem to contain thionins, although the negative results from the IKP project have to be treated with care. All these data are from transcriptomes, sometimes only from one tissue. It is therefore possible that genes which are not expressed in that tissue or only expressed at a low level are missed. In those cases the genome sequence is needed.

According to the data from the IKP project, thionins emerged first in lycophytes as class 1 thionins and class 1 thionins are also found in conifers. Class 2 thionins then came up within the angiosperms which now contain both classes. Most plant families contain only one class of thionins. However, there are also plant species which contain both classes of thionins. *Poaceae* have mainly class 1 thionins while *Brassicaceae* have mainly class 2 thionins but there are some exceptions. *N. alopecuroidea*, family *Poaceae*, had a class 1 and a class 2 thionin. The majority of thionins in the *Papaver* species are class 1 but all four species that were included in the IKP project also contain one class 2 thionin (two in the case of *P. rhoeas*).

There was a slightly higher number of class 2 thionins than class 1 and half of the class 1 thionins were from *Papaver* species. This might indicate that class 2 thionins are perhaps more

widely distributed than class 1 thionins. In line with this, the variability among class 2 thionins was higher and acidic thionins were only found within class 2.

Large gene family for thionins in *Papaver* species

The 1KP project revealed a large number of thionin sequences in all four *Papaver* species that were sequenced. It was previously not known that *Papaver* species contain thionins, supporting the importance of large scale transcriptome and genome projects. The *Papaver* thionins included typical class 1 and class 2 thionins as well as several variants (see below). In addition, class 1 thionins from *P. somniferum* and *P. setigerum* were discovered which had an unusual pro domain. Most of the clones were truncated but one seemed to be complete with a total of 496 amino acids. It is unlikely that these clones were artifacts because they were found in 3 samples each of 2 different *Papaver* species. However, it would still be important to confirm these with genomic sequences. Furthermore, this unusual structure also raises the question how the pro domain is processed. Could the processing give rise to not only a thionin but also additional protein/s with specific functions? This should be investigated experimentally.

We isolated a genomic clone from *P. somniferum* which allowed us to check for introns because this information is of course not evident from transcriptomics. We found two small introns within the region encoding the acidic domain. These two small introns seem to be generally conserved in thionin genes [19,42,47]. It would be interesting to see if these are already found in the phylogenetically earliest thionin genes.

We expressed the *Papaver* thionin PsoTHI1.7 in *E. coli*. It had significant antimicrobial activity *in vitro* against the fungi we tested but only low activity against *P. syringae*. This is in line with other thionins that have been tested *in vitro*, for instance TaTHI1.3, α -purothionin [65]. Molina et al. [33] also tested wheat purothionins against different pathogens. For three different *B. cinerea* strains they found IC₅₀ values between 5 and 12 μ g/ml which is comparable to the IC₅₀ value of 6 μ g/ml for the *B. cinerea* strain that we tested. All *Papaver* class 1 thionins have a basic pI and contain lysin as first amino acid. Most of them also have tyrosin at position 13, indicating that these have most likely antimicrobial or toxic activities [4]. If we compare PsoTHI1.7 with the other *Papaver* class 1 thionins it is safe to conclude that also the other class 1 thionins will have antimicrobial activity, maybe with the exception of those that do not have tyrosin at position 13. With all the data available, it is very likely that the *Papaver* thionin gene family is involved in resistance against pathogens.

Variants and a new classification for thionins

Thionins were previously grouped into four classes [18]. With the data obtained by the 1KP project we will now recognize only two classes and variants thereof. The reason for this is that all the variants that we found can be traced back to either class 1 (eight cysteines) or class 2 (six cysteines). With all the different variants (a total of 18 variants) uncovered by the 1KP project we would have to extend the number of classes substantially, making the system confusing. Accordingly, TaTHI3.1 and TgTHI4.1, formerly class 3 and class 4 thionins [18], respectively, are both variants of class 1. Many of the variants that we found here, six of the seven class 1 variants and six of the 11 class 2 variants, have an uneven number of cysteine residues. This raises the question about the structure of the peptides. We can assume that most of these cysteines form disulfide bridges, as has repeatedly been shown for many class 1 and class 2 thionins, leaving probably one single cysteine. It could be possible that the single cysteines engage in forming dimers. It has been shown before that thionins without single cysteines form dimers [44]. However, formation of intermolecular disulfide bridges has to be determined experimentally by isolating the peptides from the plants or by expressing them in expression hosts.

Supporting information

S1 Fig. Sequences of different thionin proproteins.

(DOCX)

S2 Fig. General structure of thionin preproteins and position of the introns.

(DOCX)

S3 Fig. Expression and purification of PsoTHI1.7.

(DOCX)

S1 Table. pI of some thionins, acidic domains and proproteins.

(DOCX)

S2 Table. Bait sequences that were used to search the 1KP data.

(DOCX)

S3 Table. Primers used in this work.

(DOCX)

Acknowledgments

We thank Gane Ka-Shu Wong and Michael K. Deyholos for access to data from the 1KP project. We appreciate Clemens Grünwald-Gruber for performing the mass spectroscopy.

Author Contributions

Conceptualization: Holger Bohlmann.

Data curation: Holger Bohlmann.

Investigation: Katharina Höng, Tina Austerlitz, Timo Bohlmann.

Supervision: Holger Bohlmann.

Visualization: Holger Bohlmann.

Writing – original draft: Katharina Höng, Tina Austerlitz, Holger Bohlmann.

Writing – review & editing: Holger Bohlmann.

References

1. Broekaert WF, Cammue BPA, DeBolle MFC, Thevissen K, DeSamblanx GW, Osborn RW. Antimicrobial peptides from plants. *Critical Reviews in Plant Sciences*. 1997; 16(3):297–323.
2. Jenssen H, Hamill P, Hancock RE. Peptide antimicrobial agents. *Clin Microbiol Rev*. 2006; 19(3):491–511. <https://doi.org/10.1128/CMR.00056-05> PMID: 16847082
3. Yeaman MR, Yount NY. Mechanisms of antimicrobial peptide action and resistance. *Pharmacol Rev*. 2003; 55(1):27–55. <https://doi.org/10.1124/pr.55.1.2> PMID: 12615953
4. Bohlmann H. The Role of Thionins in Plant-Protection. *Critical Reviews in Plant Sciences*. 1994; 13(1):1–16.
5. Bohlmann H, Apel K. Thionins. *AnnuRevPhysiolPlant MolBiol*. 1991; 42:227–40.
6. Parisi K, Shafee TMA, Quimbar P, van der Weerden NL, Bleackley MR, Anderson MA. The evolution, function and mechanisms of action for plant defensins. *Semin Cell Dev Biol*. 2019; 88:107–18. <https://doi.org/10.1016/j.semcdb.2018.02.004> PMID: 29432955
7. Thomma BP, Cammue BP, Thevissen K. Plant defensins. *Planta*. 2002; 216(2):193–202. <https://doi.org/10.1007/s00425-002-0902-6> PMID: 12447532
8. Zasloff M. Antimicrobial peptides of multicellular organisms. *Nature*. 2002; 415(6870):389–95. <https://doi.org/10.1038/415389a> PMID: 11807545

9. Cardoso MH, Meneguetti BT, Costa BO, Buccini DF, Oshiro KGN, Preza SLE, et al. Non-Lytic Antibacterial Peptides That Translocate Through Bacterial Membranes to Act on Intracellular Targets. *Int J Mol Sci*. 2019; 20(19). <https://doi.org/10.3390/ijms20194877> PMID: 31581426
10. Hale JD, Hancock RE. Alternative mechanisms of action of cationic antimicrobial peptides on bacteria. *Expert Rev Anti Infect Ther*. 2007; 5(6):951–9. <https://doi.org/10.1586/14787210.5.6.951> PMID: 18039080
11. Otvos L Jr. Antibacterial peptides and proteins with multiple cellular targets. *J Pept Sci*. 2005; 11(11):697–706. <https://doi.org/10.1002/psc.698> PMID: 16059966
12. Ohtani S, Okada T, Yoshizumi H, Kagamiyama H. Complete primary structures of two subunits of purothionin A, a lethal protein for brewer's yeast from wheat flour. *J Biochem*. 1977; 82(3):753–67. <https://doi.org/10.1093/oxfordjournals.jbchem.a131752> PMID: 914810
13. Balls AK, Hale WS, Harris TH. A crystalline protein obtained from a lipoprotein of wheat flour. *Cereal Chemistry*. 1942; 19:279–88.
14. Nimmo CC, Kasarda DD, Lew EJ. Physical characterization of the wheat protein purothionin. *J Sci Food Agric*. 1974; 25(6):607–17. <https://doi.org/10.1002/jsfa.2740250603> PMID: 4845102
15. Mak AS, Jones BL. The amino acid sequence of wheat beta-purothionin. *Can J Biochem*. 1976; 54(10):835–42. <https://doi.org/10.1139/o76-120> PMID: 990986
16. Fisher N, Redman DG, Elton GAH. Fractionation and characterization of purothionin. *Cereal Chemistry*. 1968; 45:48–57.
17. Bohlman H, Apel K. Thionins. *Annual Review of Plant Physiology and Plant Molecular Biology*. 1991; 42:227–40.
18. Bohlmann H, Apel K, Garcia-Olmedo F. Thionins. *Plant Molecular Biology Reporter*. 1994; 12(2):S75–S.
19. Rodriguez-Palenzuela P, Pintor-Toro JA, Carbonero P, Garcia-Olmedo F. Nucleotide sequence and endosperm-specific expression of the structural gene for the toxin alpha-hordothionin in barley (*Hordeum vulgare* L.). *Gene*. 1988; 70(2):271–81. [https://doi.org/10.1016/0378-1119\(88\)90199-0](https://doi.org/10.1016/0378-1119(88)90199-0) PMID: 2850969
20. Bohlmann H, Apel K. Isolation and characterization of cDNAs coding for leaf-specific thionins closely related to the endosperm-specific hordothionin of barley (*Hordeum vulgare* L.). *Molecular General Genetics*. 1987; 207:446–54.
21. Vernon LP, Evett GE, Zeikus RD, Gray WR. A toxic thionin from *Pyricularia pubera*: purification, properties, and amino acid sequence. *Arch Biochem Biophys*. 1985; 238(1):18–29. [https://doi.org/10.1016/0003-9861\(85\)90136-5](https://doi.org/10.1016/0003-9861(85)90136-5) PMID: 3985614
22. Samuelsson G, Seger L, Olson T. The amino acid sequence of oxidized viscotoxin A3 from the European mistletoe (*Viscum album* L., Loranthaceae). *Acta Chem Scand*. 1968; 22(8):2624–42. <https://doi.org/10.3891/acta.chem.scand.22-2624> PMID: 5719166
23. Schrader G, Apel K. Isolation and characterization of cDNAs encoding viscotoxins of mistletoe (*Viscum album*). *Eur J Biochem*. 1991; 198(3):549–53. <https://doi.org/10.1111/j.1432-1033.1991.tb16049.x> PMID: 1710983
24. Teeter MM, Mazer JA, L'Italien JJ. Primary structure of the hydrophobic plant protein crambin. *Biochemistry*. 1981; 20(19):5437–43. <https://doi.org/10.1021/bi00522a013> PMID: 6895315
25. Vermeulen JAWH, Lamerichs RMJN, Berliner LJ, De Marco A, Llinas M, Boelens R, et al. 1H-NMR characterization of two crambin species. *FEBS Lett*. 1987; 219:426–30.
26. Epple P, Apel K, Bohlmann H. An Arabidopsis thaliana thionin gene is inducible via a signal transduction pathway different from that for pathogenesis-related proteins. *Plant Physiol*. 1995; 109:813–20. <https://doi.org/10.1104/pp.109.3.813> PMID: 8552715
27. Castagnaro A, Marana C, Carbonero P, Garcia-Olmedo F. Extreme divergence of a novel wheat thionin generated by a mutational burst specifically affecting the mature protein domain of the precursor. *J Mol Biol*. 1992; 224(4):1003–9. [https://doi.org/10.1016/0022-2836\(92\)90465-v](https://doi.org/10.1016/0022-2836(92)90465-v) PMID: 1569564
28. Bohlmann H, Clausen S, Behnke S, Giese H, Hiller C, Reimann-Philipp U, et al. Leaf-specific thionins of barley—a novel class of cell wall proteins toxic to plant-pathogenic fungi and possibly involved in the defence mechanism of plants. *EMBO J*. 1988; 7(6):1559–65. PMID: 16453847
29. Gausing K. Thionin genes specifically expressed in barley leaves. *Planta*. 1987; 171(2):241–6. <https://doi.org/10.1007/BF00391100> PMID: 24227332
30. Reimann-Philipp U, Schrader G, Martinoia E, Barkholt V, Apel K. Intracellular thionins of barley. A second group of leaf thionins closely related to but distinct from cell wall-bound thionins. *J Biol Chem*. 1989; 264(15):8978–84. PMID: 2722812
31. Reimann-Philipp U, Behnke S, Batschauer A, Schafer E, Apel K. The effect of light on the biosynthesis of leaf-specific thionins in barley, *Hordeum vulgare*. *Eur J Biochem*. 1989; 182(2):283–9. <https://doi.org/10.1111/j.1432-1033.1989.tb14828.x> PMID: 2737201

32. Florack DE, Visser B, De Vries PM, Van Vuurde JWL, Stiekema W. Analysis of the toxicity of purothionins and hordothionins for plant pathogenic bacteria. *Netherlands Journal of Plant Pathology*. 1993; 99:259–68.
33. Molina A, Ahl-Goy P, Fraile A, Sanchez-Monge R, Garcia-Olmedo F. Inhibition of bacterial and fungal plant pathogens by thionins of types I and II. *Plant Science*. 1993; 92:169–77.
34. Rosell S, Samuelsson G. Effect of mistletoe viscotoxin and phoratoxin on blood circulation. *Toxicol*. 1966; 4(2):107–10. [https://doi.org/10.1016/0041-0101\(66\)90005-5](https://doi.org/10.1016/0041-0101(66)90005-5) PMID: 6005043
35. Selawry OS, Vester F, Mai W, Schwartz MR. On the identification of the constituents of *Viscum album*. II. Tumor inhibiting substances. *Hoppe Seylers Z Physiol Chem*. 1961; 324:262–81. <https://doi.org/10.1515/bchm2.1961.324.1.262> PMID: 13749858
36. Vanetten CH, Nielsen HC, Peters JE. A crystalline polypeptide from the seed of *Crambe abyssinica*. *Phytochemistry*. 1965; 4(3):467–73.
37. Van Etten CH, Gagne WE, Robbins DJ, Booth AN, Daxenbichler ME, Wolff IA. Biological evaluation of *Crambe* seed meals and derived products by rat feeding. *Cereal Chemistry*. 1969; 46:145–55.
38. Schrader-Fischer G, Apel K. Organ-specific expression of highly divergent thionin variants that are distinct from the seed-specific crambin in the crucifer *Crambe abyssinica*. *Mol Gen Genet*. 1994; 245(3):380–9. <https://doi.org/10.1007/BF00290119> PMID: 7816048
39. Bohlmann H, Vignutelli A, Hilpert B, Miersch O, Wasternack C, Apel K. Wounding and chemicals induce expression of the *Arabidopsis thaliana* gene Thi2.1, encoding a fungal defense thionin, via the octadecanoid pathway. *FEBS Lett*. 1998; 437(3):281–6. [https://doi.org/10.1016/S0014-5793\(98\)01251-4](https://doi.org/10.1016/S0014-5793(98)01251-4) PMID: 9824308
40. Epple P, Apel K, Bohlmann H. Overexpression of an endogenous thionin enhances resistance of *Arabidopsis* against *Fusarium oxysporum* The Plant Cell. 1997; 9:509–20. <https://doi.org/10.1105/tpc.9.4.509> PMID: 9144959
41. Epple P, Vignutelli A, Apel K, Bohlmann H. Differential induction of the *Arabidopsis thaliana* Thi2.1 gene by *Fusarium oxysporum* f. sp. *matthioli*. *Mol Plant Microbe Interact*. 1998; 11(6):523–9. <https://doi.org/10.1094/MPMI.1998.11.6.523> PMID: 9612950
42. Vignutelli A, Wasternack C, Apel K, Bohlmann H. Systemic and local induction of an *Arabidopsis* thionin gene by wounding and pathogens. *The Plant Journal*. 1998; 14(3):285–95. <https://doi.org/10.1046/j.1365-3113x.1998.00117.x> PMID: 9628023
43. Stec B. Plant thionins—the structural perspective. *Cell Mol Life Sci*. 2006; 63(12):1370–85. <https://doi.org/10.1007/s00018-005-5574-5> PMID: 16715411
44. Stec B, Markman O, Rao U, Heffron G, Henderson S, Vernon LP, et al. Proposal for molecular mechanism of thionins deduced from physico-chemical studies of plant toxins. *J Pept Res*. 2004; 64(6):210–24. <https://doi.org/10.1111/j.1399-3011.2004.00187.x> PMID: 15613085
45. Oard SV. Deciphering a mechanism of membrane permeabilization by alpha-hordothionin peptide. *Biochim Biophys Acta*. 2011; 1808(6):1737–45. <https://doi.org/10.1016/j.bbame.2011.02.003> PMID: 21315063
46. Ponz F, Paz-Ares J, Hernandez-Lucas C, Garcia-Olmedo F, Carbonero P. Cloning and nucleotide sequence of a cDNA encoding the precursor of the barley toxin alpha-hordothionin. *Eur J Biochem*. 1986; 156(1):131–5. <https://doi.org/10.1111/j.1432-1033.1986.tb09557.x> PMID: 3082629
47. Holtorf S, Apel K, Bohlmann H. Specific and different expression patterns of two members of the leaf thionin multigene family of barley in transgenic tobacco. *Plant Science*. 1995; 111(1):27–37.
48. Florack DE, Dirkse WG, Visser B, Heidekamp F, Stiekema WJ. Expression of biologically active hordothionins in tobacco. Effects of pre- and pro-sequences at the amino and carboxyl termini of the hordothionin precursor on mature protein expression and sorting. *Plant Mol Biol*. 1994; 24(1):83–96. <https://doi.org/10.1007/BF00040576> PMID: 8111029
49. Lee D-W, Apel K, Binding H. The Formation of Leaf Thionins and their Effect on the Viability of Isolated Protoplasts of Barley (*Hordeum vulgare* L.). *Journal of Plant Physiology*. 1995; 145(1–2):131–6.
50. Plattner S, Gruber C, Stadlmann J, Widmann S, Gruber CW, Altmann F, et al. Isolation and Characterization of a Thionin Proprotein-processing Enzyme from Barley. *J Biol Chem*. 2015; 290(29):18056–67. <https://doi.org/10.1074/jbc.M115.647859> PMID: 26013828
51. Leebens-Mack JH. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. 2019; 574(7780):679–85. <https://doi.org/10.1038/s41586-019-1693-2> PMID: 31645766
52. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019; 37(4):420–3. <https://doi.org/10.1038/s41587-019-0036-z> PMID: 30778233
53. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019; 47(W1):W636–W41. <https://doi.org/10.1093/nar/gkz268> PMID: 30976793

54. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14(6):1188–90. <https://doi.org/10.1101/gr.849004> PMID: 15173120
55. Edwards K, Johnstone C, Thompson C. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research.* 1991; 19(6):1349. <https://doi.org/10.1093/nar/19.6.1349> PMID: 2030957
56. Ali MA, Shah KH, Bohlmann H. pMAA-Red: a new pZP-derived vector for fast visual screening of transgenic Arabidopsis plants at the seed stage. *BMC Biotechnol.* 2012; 12:37. <https://doi.org/10.1186/1472-6750-12-37> PMID: 22747516
57. Bogomolovas J, Simon B, Sattler M, Stier G. Screening of fusion partners for high yield expression and purification of bioactive viscotoxins. *Protein Expr Purif.* 2009; 64(1):16–23. <https://doi.org/10.1016/j.pep.2008.10.003> PMID: 18983922
58. Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb Cell Fact.* 2012; 11:56. <https://doi.org/10.1186/1475-2859-11-56> PMID: 22569138
59. Schägger H. Tricine-SDS-PAGE. *Nat Protoc.* 2006; 1(1):16–22. <https://doi.org/10.1038/nprot.2006.4> PMID: 17406207
60. Gharahdaghi F, Weinberg CR, Meagher DA, Imai BS, Mische SM. Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: a method for the removal of silver ions to enhance sensitivity. *Electrophoresis.* 1999; 20(3):601–5. PMID: 10217175
61. Broekaert WF, Terras FRG, Cammue BPA, Vanderleyden J. An automated quantitative assay for fungal growth inhibition. *FEMS Microbiol Lett.* 1990; 69:55–60.
62. Wong GK, Soltis DE, Leebens-Mack J, Wickett NJ, Barker MS, Van de Peer Y, et al. Sequencing and Analyzing the Transcriptomes of a Thousand Species Across the Tree of Life for Green Plants. *Annu Rev Plant Biol.* 2020; 71:741–65. <https://doi.org/10.1146/annurev-arplant-042916-041040> PMID: 31851546
63. Carpenter EJ, Matasci N, Ayyampalayam S, Wu S, Sun J, Yu J, et al. Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *Gigascience.* 2019; 8(10).
64. Shelenkov A, Slavokhotova A, Odintsova T. Predicting Antimicrobial and Other Cysteine-Rich Peptides in 1267 Plant Transcriptomes. *Antibiotics (Basel, Switzerland).* 2020; 9(2). <https://doi.org/10.3390/antibiotics9020060> PMID: 32032999
65. Terras F, Schoofs H, Thevissen K, Osborn RW, Vanderleyden J, Cammue B, et al. Synergistic Enhancement of the Antifungal Activity of Wheat and Barley Thionins by Radish and Oilseed Rape 2S Albumins and by Barley Trypsin Inhibitors. *Plant Physiol.* 1993; 103(4):1311–9. <https://doi.org/10.1104/pp.103.4.1311> PMID: 12232024