

# scientific reports



OPEN

## Genome-wide identification of Argonautes in Solanaceae with emphasis on potato

Zhen Liao<sup>1,3</sup>, Kristian Persson Hodén<sup>1,3</sup>, Ravi Kumar Singh<sup>1,2</sup> & Christina Dixelius<sup>1</sup>✉

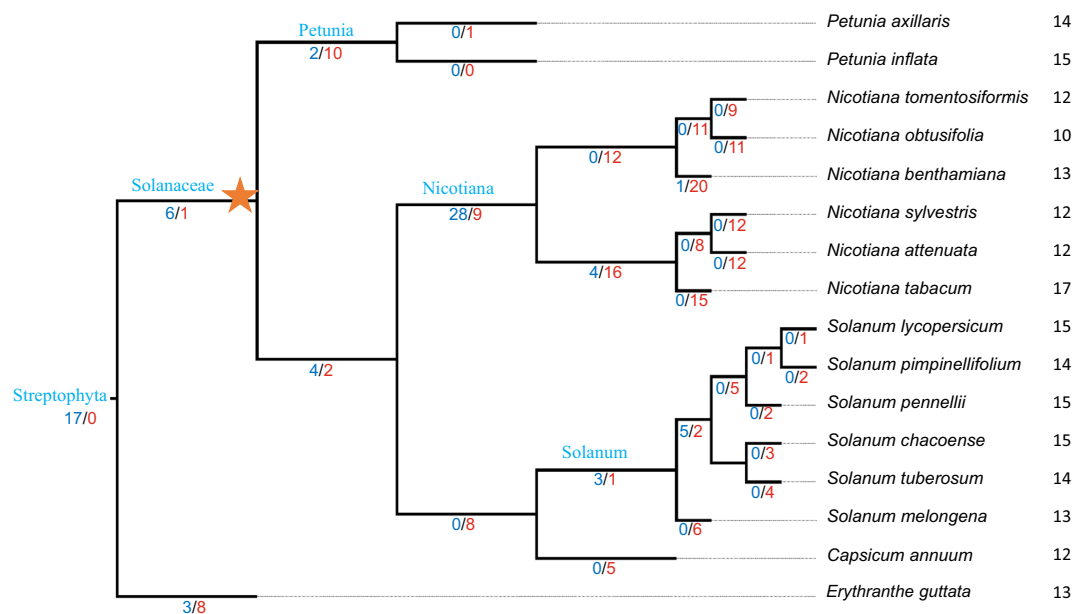
Regulatory small RNAs (sRNAs) play important roles in many fundamental processes in plant biology such as development, fertilization and stress responses. The AGO protein family has here a central importance in gene regulation based on their capacity to associate with sRNAs followed by mRNA targeting in a sequence-complementary manner. The present study explored Argonautes (AGOs) in the *Solanaceae* family, with emphasis on potato, *Solanum tuberosum* (St). A genome-wide monitoring was performed to provide a deeper insight into gene families, genomic localization, gene structure and expression profile against the potato late blight pathogen *Phytophthora infestans*. Among 15 species in the *Solanaceae* family we found a variation from ten AGOs in *Nicotiana obtusifolia* to 17 in *N. tabacum*. Comprehensive analyses of AGO phylogeny revealed duplication of *AGO1*, *AGO10* and *AGO4* paralogs during early radiation of Solanaceae. Fourteen AGOs were identified in potato. Orthologs of *AGO8* and *AGO9* were missing in the potato genome. However, *AGO15* earlier annotated in tomato was identified. StAGO15 differs from the other paralogs having residues of different physico-chemical properties at functionally important amino acid positions. Upon pathogen challenge StAGO15 was significantly activated and hence may play a prominent role in sRNA-based regulation of potato defense.

Non-coding small RNAs (sRNAs) are ubiquitous components of eukaryotic and prokaryotic gene regulatory processes<sup>1</sup>. Although variation in lengths, biogenesis, functions and targets not least in plants, the unified outcome is restricting the action of target molecules either on transcriptional or post-transcriptional levels. The core part required for this regulatory RNA-based process involves canonical ribonucleases that participate in initiator and effector steps. They are Dicer or DICER-LIKE (DCL) proteins that cleave double-stranded RNA and Argonautes (AGOs), important for the small RNA association and formation of the RNA-induced RNA silencing complexes (RISCs)<sup>2,3</sup>. RNA-dependent RNA polymerase (RDR) is a third enzyme that takes part in amplifying and maintaining the silencing signal in many organisms<sup>4</sup>.

Eukaryotic AGO proteins are characterized by four domains of importance to bind diverse sRNA classes. They are: PAZ (Piwi-Argonaute-Zwille), MID, PIWI, and N-terminal sequences<sup>5</sup>. The four domains have distinct functions to facilitate sRNA loading and activity of RISC<sup>6</sup>. In eukaryotes, the sRNA guide strand is anchored at the PAZ domain via its 3' end<sup>7</sup>. At the opposite side, MID-PIWI sequences form a binding pocket for the 5' end<sup>8,9</sup>. In plants, each class of sRNAs display a bias towards certain 5' nucleotides (nt)<sup>10</sup>, and modulation of RISC turnover rate and possible recycling of AGOs, is channeled via the 3' sequence attached to the PAZ domain<sup>11</sup>. Apart from Dicer-dependent pathways, AGOs also are involved in Dicer-independent events where for example short hairpin RNA rely on AGO2 for maturation<sup>12</sup>.

There is a wide variation in numbers of DCL and AGO genes in different organisms<sup>5,13</sup>. *Arabidopsis thaliana* and rice have four and five DCLs, respectively, whereas *Drosophila* has two and *C. elegans* only one DCL gene<sup>14</sup>. Similarly, the numbers of AGOs vary greatly in different species. For instance, there is only one AGO gene in the fission yeast *S. pombe* but as many as 27 are found in *C. elegans*<sup>15</sup>. Species in the green clade are no exceptions. In the evolution of Viridiplantae, the AGO gene family has expanded from three members in green algae<sup>16</sup> to six in moss, ten in *Arabidopsis*, 19 in rice and 22 in soybean<sup>17,18</sup>. Plant AGO proteins are grouped into three major clades named after phylogenetic analysis and comparison with *Arabidopsis*: AGO1/5/10, AGO2/3/7, and AGO4/6/8/9<sup>19</sup>. These three AGO clades are dated back to the most recent common ancestors of land plants. Thus, the diversification of the AGO family in Viridiplantae is an ancient and most likely a continuous process<sup>20</sup>.

<sup>1</sup>Department of Plant Biology, Uppsala BioCenter, Linnéan Center for Plant Biology, Swedish University of Agricultural Sciences, P.O. Box 7080, 75007 Uppsala, Sweden. <sup>2</sup>Present address: Genomics and Bioinformatics Laboratory, University Department of Botany, Magadh University, Bodh Gaya 824234, India. <sup>3</sup>These authors contributed equally: Zhen Liao and Kristian Persson Hodén. ✉email: Christina.Dixelius@slu.se



**Figure 1.** Evolutionary events of the Argonaute (AGO) family in Solanaceous genomes. The values on branches correspond to the number of gene gains (blue) and losses (red). Right panel showing the total number of AGO genes in the genome of each species. ★ indicates a whole genome triplication event.

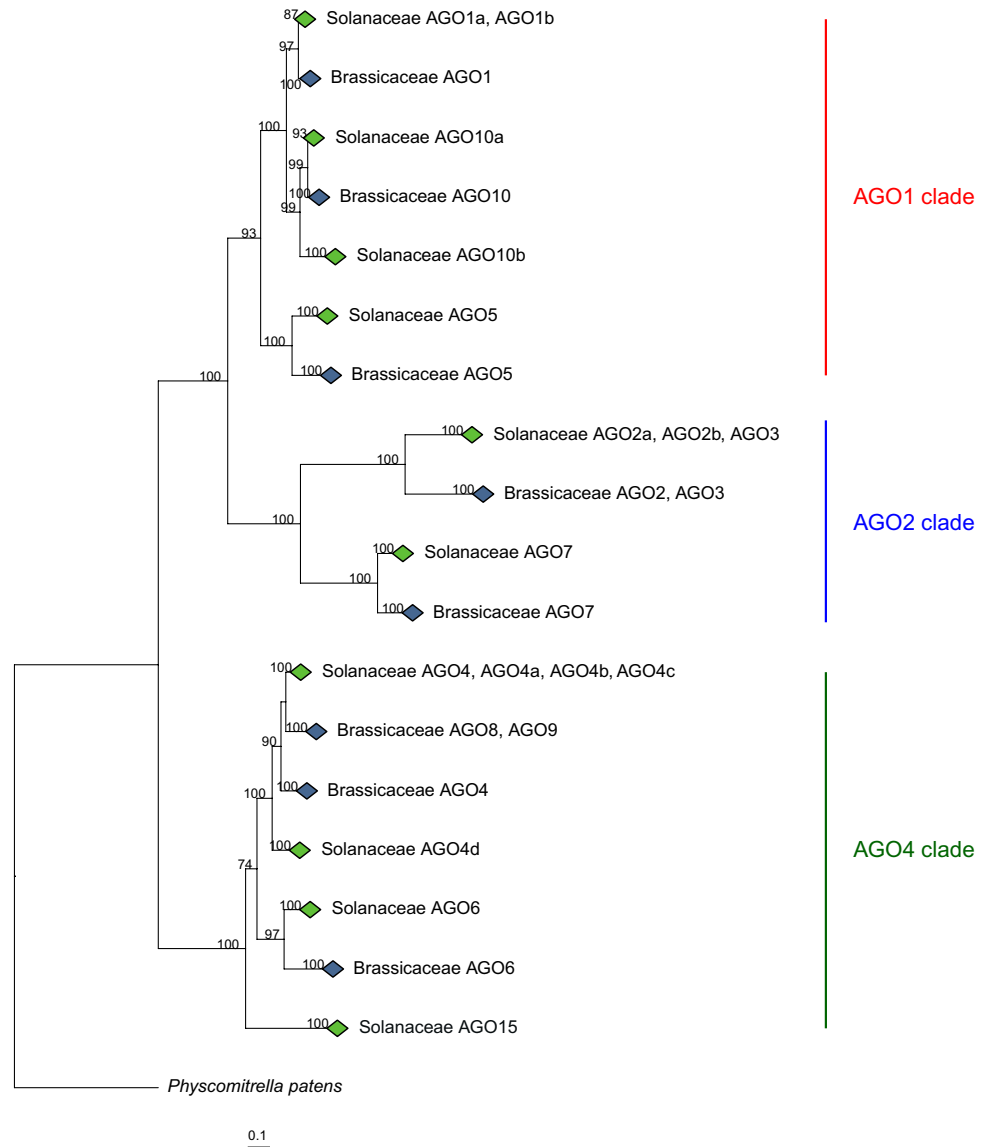
High-throughput sequencing followed by comparative genomics has revealed several gains and losses of AGO encoded genes. In addition, genes not previously annotated are still a source of new information such as reported for AGOs in *Nicotiana attenuata*<sup>21</sup>. Species in the *Solanaceae* family have experienced a specific ploidy event after the split with Asterids about 49 million years ago followed by further species divergence<sup>22</sup>. The large genus *Solanum* diverged from peppers (*Capsicum*) c. 19 Million years ago (Mya) whereas the *Solanum* crops, potato and tomato split rather recently c. 8 Mya<sup>23</sup>. Overall, the *Solanaceae* gene families vary in size due to duplications and different gene evolutionary events<sup>24</sup>.

In this study we used all public genomic and transcriptomic resources to clarify the number of AGO encoding genes, and their divergence in *Solanaceae* followed by tests of expression upon filamentous pathogen infection. Extensive gains and losses of AGOs have occurred in *Nicotiana* species compared to *Solanum* species. The AGO4 clade which harbors a specific *Solanaceae* sub-clade, AGO15, with novel sequence and structural features received our attention. StAGO15 was activated by filamentous pathogen challenge suggesting an important role in the immune responses of potato. Details of its sRNA binding and function under biotic stress remains to be elucidated.

## Results

**Extensive gains and losses of AGOs in Solanaceous genomes.** We searched full length AGO sequences in potato and related genomes of *Solanaceae* species to generate an overview picture of the number of AGOs in the *Solanaceae* family and their evolutionary history. To generate confidence over number of gene gain and loss events, the reconciliation of the gene trees was repeated three times using three different species as out-groups *Arabidopsis*, *Vitis vinifera* and *Erythranthe guttata*. We found six AGO duplication events prior the split between *Petunia* and the other *Solanaceae* lineages (Fig. 1; Supplementary Fig. S1). After the divergence of *Petunia*, the ancestral line experienced four duplication and two loss events followed by speciation processes leading to *Nicotiana* and *Solanum* lineages. The AGO family in the ancestral *Nicotiana* lineage has experienced extensive changes. Based on the six *Nicotiana* species analyzed, 33 duplication and 59 loss events had occurred prior further speciation. Processes involving gene losses have continued within each species. Maximum losses have occurred in *N. benthamiana* (20) followed by *N. tabacum* (15). Compared to *Nicotiana*, the expansion of the AGO family is less variable in the *Solanum* genus. Overall, the number of AGOs in the *Solanaceae* species analyzed varies from ten in *N. obtusifolia* to 17 in *N. tabacum*.

**Potato has 14 Argonaute encoding genes.** The search for homologous AGO sequences in the *Solanaceae* family generated fourteen full length AGO genes recovered from the *S. tuberosum* (potato) genome (StAGOs). PAZ and PIWI domains were found in all 14 AGO sequences whereas presence of other conserved parts such as MID, N terminal and linker domains were not predicted in all candidates when applying default settings (Supplementary Fig. S2a). Based on sequence identity and phylogenetic clustering, the orthologs of *Arabidopsis* AGO1, AGO2, AGO3, AGO4, AGO5, AGO6, AGO7 and AGO10 were all discovered in the potato genome (Supplementary Fig. S2). In addition, two orthologs each for AGO1, AGO2, AGO10 and three for AGO4 were identified. AGO6 was not found in the reference annotation of the potato genome (PGSC) but later identified



**Figure 2.** Maximum likelihood phylogeny (RAxML, model JTT +  $\Gamma$ , 100 replicates) of Argonaute (AGO) family in Solanaceae and Brassicaceae. Blue and green diamonds represent collapsed Solanaceae and Brassicaceae clades, respectively. Bootstrap values > 70% are indicated. Bar = number of substitutions per site. Outgroup = *Physcomitrella patens*.

in the annotation by the International Tomato Annotation Group (ITAG). Two candidates, *AGO8* and *AGO9* were however not found in any dataset. Phylogenetic clustering conferred five homologs of *StAGOs* in each clade of *AGO1* and *AGO4*, while four homologs grouped in the *AGO2* clade based on the *AGO* classification in *Arabidopsis*<sup>19</sup>. The *StAGO* paralogs showed varied degree of identity and genetic distances among each other (Supplementary Fig. S2b). Two candidates, *StAGO15* and *StAGO7* were the most divergent *AGOs* compared to the other members.

Next, the *StAGOs* were mapped on the *S. tuberosum* chromosomes (Supplementary Fig. S3). The close positions of *StAGO2a*, *StAGO2b* and *StAGO3* on chromosome 2, and *StAGO1a*, *StAGO4a* and *StAGO10c* on chromosome 6, together with sequence similarities, suggest that they have experienced gene duplications. Similar tandem gene duplications are observed on chromosome 2 and 6 in tomato<sup>25</sup>. In tomato and potato, no *AGOs* are found on chromosome 4, 5, 8, 10 and 11.

A maximum likelihood phylogeny was reconstructed by using a total of 203 *AGO* homologs from the sampled Solanaceae lineages. To get confidence over the topology and partitions, 99 *AGO* homologs were added from nine sequenced Brassicaceae species (Supplementary Fig. S4). In line with earlier clustering (Supplementary Fig. S2b) three major clades (*AGO1*, *AGO2*, *AGO4*) were formed (Fig. 2). Homologs from almost all species were observed in all three clades including duplications in Solanaceae. The clustering pattern and topology indicates that a duplication of *AGO10* has occurred in an early ancestor prior the divergence of Solanaceae and

Brassicaceae, but the duplicated ortholog has been lost in Brassicaceae. The duplication of the *AGO1* gene, on the other hand, has most likely occurred early at the base of Solanaceae after the split with Asterids. Two gene copies of *AGO1* and *AGO10* were found in *Nicotiana benthamiana*, and two copies for *AGO5* in the four species *Petunia inflata*, *N. tabacum*, *N. tomentosiformis* and *N. benthamiana*. Orthologs in the *AGO5* sub-clade are more dissimilar than those in the *AGO1* and *AGO10* sub-clades, a situation which also is reflected in variations among the branch lengths. Orthologs of *AGO7* were observed in almost all Solanaceae lineages and a separate partition with the Brassicaceae orthologs formed the *AGO7* sub-clade. Almost the same branch length indicates low variation among the orthologs. Further, *AGO2* and *AGO3* are sister orthologs that grouped together forming a separate sub-clade in which Solanaceae lineages showed varied number of paralogs (Fig. 2). Only one ortholog for *AGO2* and *AGO3* was detected from *Nicotiana*, *Petunia* and *Capsicum* lineages, while the *Solanum* species analyzed had three orthologs, except *S. chacoense* that had two gene copies.

The *AGO4* clade was most likely formed by orthologs of *AGO4* exhibiting clear partition with *AGO8* and *AGO9* sequences from Solanaceae (Fig. 2). Two groups of *AGO4* orthologs were found in Solanaceae. Based on identity with the corresponding members in *Arabidopsis*, three paralogs in for example potato were found (*StAGO4*, *4a* and *4d*). Orthologs of *AGO6* from the two plant families were also identified.

**StAGO15 has novel domain architecture.** A well supported sub-clade containing a single member from each Solanaceae species, except from *N. attenuata* and *N. benthamiana*, was found at the base of the *AGO4* clade, earlier annotated as *AGO15* in tomato<sup>25</sup> (Fig. 2; Supplementary Fig. S4). The relative position of this node in the phylogeny and branch lengths suggests that the evolution of *AGO15* occurred early in Solanaceae. We searched databases and compared *AGO15* sequences from Solanaceae and Poaceae in order to detect potential similarities in sequence and function. The phylogenetic tree generated two monophyletic clades with representatives from each plant family (Supplementary Fig. S5). The topology of the tree coincides with estimated divergence time between Monocots and Eudicots<sup>26</sup>. In rice, the *AGO4* clade comprise of four members, *OsAGO4a*, *OsAGO4b*, *OsAGO15* and *OsAGO16* (Fig. 3). *StAGO15* does not cluster with any of the potato or rice *AGOs* in this clade, suggesting independent evolution.

We aligned the *AGO15* and the *AGO1* clade sequences and found clear divergence in the PIWI domain among the Solanaceae species, particularly among the amino acids corresponding to the catalytic “slicing” residues D-E-D-H/D (Supplementary Fig. S6). In potato, *StAGO15* has replaced the catalytic tetrad ‘D-E-D-H/D’ with a G-E-Q-R motif with unknown slicer function. Likewise, tomato has a G-Q-R/P motif at the catalytic site of *SlAGO15*<sup>25</sup>.

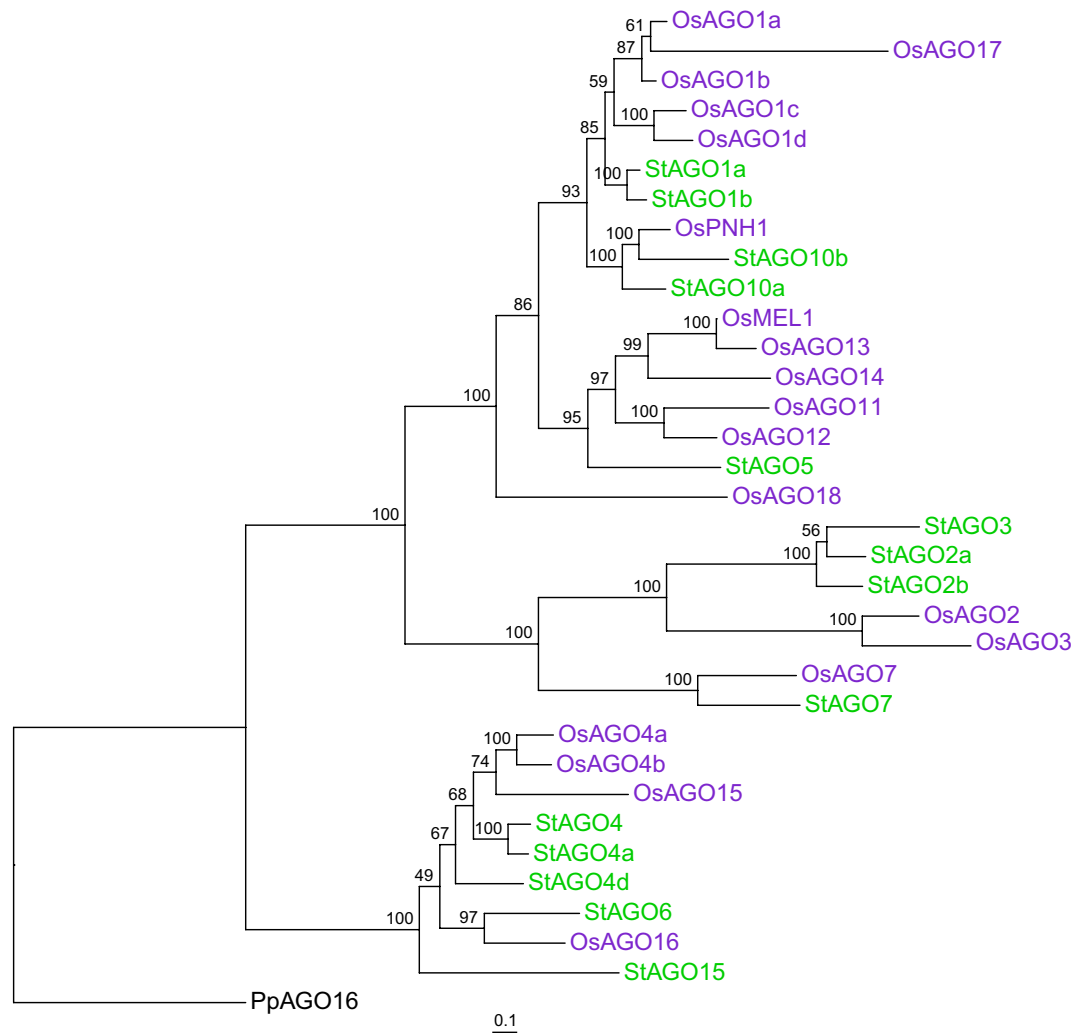
We closer examined the MID domains that did not fulfill the default sequences in comparison with *AtAGO1* (Supplementary Fig. S7a,b). Of particular interest was the nucleotide specificity loop (NSL) which in *Arabidopsis* is known to regulate 5' specificity (C, U or A)<sup>9,10,27</sup>. The Solanaceae *AGO15* protein sequences deviate substantially from the *AGO1* clade in the NSL positions, where for example *StAGO15* has AFY as 5' end recognition sequence (Supplementary Fig. S7b).

A three-dimensional protein structure comparison was performed by first model the human *AGO2* protein to visualize the different main domains together with their interaction with miR20a<sup>28</sup> (Fig. 4a). Next, models of *AtAGO1* and *StAGO15* were constructed to facilitate identification of divergent units (Fig. 4b,c). Merged protein structures of *AtAGO1* and *StAGO15* showed large similarities (Fig. 4d). However, the *StAGO15* protein appeared somewhat “bulky”. This feature is explained by three single coils, located either at the N-terminal, in the L1 domain or at the opening of the central pocket. In comparison with *AtAGO1*, the NSL sequence of *StAGO15* has a hydrophobic residue (Phe583) replacing Asn 687 in *AtAGO1* (Fig. 4e). This residue is of importance for the 5' nucleotide selection in *Arabidopsis*<sup>9</sup>. Further, the D-E-D-H catalytic pocket observed in *AtAGO1* is replaced by a G-E-Q-R motif in *StAGO15* (Fig. 4f). The D-E-D-H and G-E-Q-R motifs resemble each other, sharing the glutamic acid (Glu 708 vs. Glu 803) as 2nd motif residue with negative charge and a 4th positive residue (Arg 882 vs. His 988). The 3rd motif residue, being negative in *AtAGO1* (Asp 848) and polar in *StAGO15* (Gln 750) is in both cases capable of binding positive residues, hence this substitute may not affect the function of *StAGO15*. The charge of Gly 667 in G-E-Q-R is pH dependent, a feature whose impact is unknown particularly under stress condition. The divergent recognition and binding motifs compared to *AGO1* clade may suggest specific function(s).

**StAGO15 is elevated upon pathogen infection.** Based on generated RNAseq data on potato challenged by the *Phytophthora infestans* virulent strain 11388 *StAGO4c*, *StAGO10b*, and *StAGO15* were found up-regulated (Supplementary Fig. S8). Quantitative real-time PCR supported the activation of *StAGO15* (Fig. 5a). In a time-course experiment the gene activation was clearly observed 4 to 5 days post inoculation (Fig. 5b, Supplementary Fig. S9) when *P. infestans* has switched from biotrophic to necrotrophic stage<sup>29</sup>. To clarify if this elevated *AGO15* activity was specific for *P. infestans*, the early blight fungus *Alternaria solani* was used for potato infection in parallel experiments. Again, *StAGO15* was up-regulated but not as much as seen in the *P. infestans* response (Fig. 5a).

## Discussion

The *Solanaceae* plant family comprises many important crop species with variable genome and gene family sizes, reflecting their history of genome duplications and variable selective constraints<sup>22,30</sup>. In an analysis of twelve *Solanaceae* species, gene duplication rate, strength of selection, and gene function was shown to vary extensively together impacting the gene family sizes<sup>24</sup>. Genes were detected enriched in the genomes either by whole genome duplication or by tandem duplication. Members in gene families with low domain variability displayed a tendency of housekeeping functions. Aforesaid genes appeared to have duplicated by whole genome duplication, in contrast to the tandem duplicated category that showed higher variability. In our analysis of *AGO* genes in 15 *Solanaceae* species a rather extensive variation in gene numbers were detected, particularly when comparing



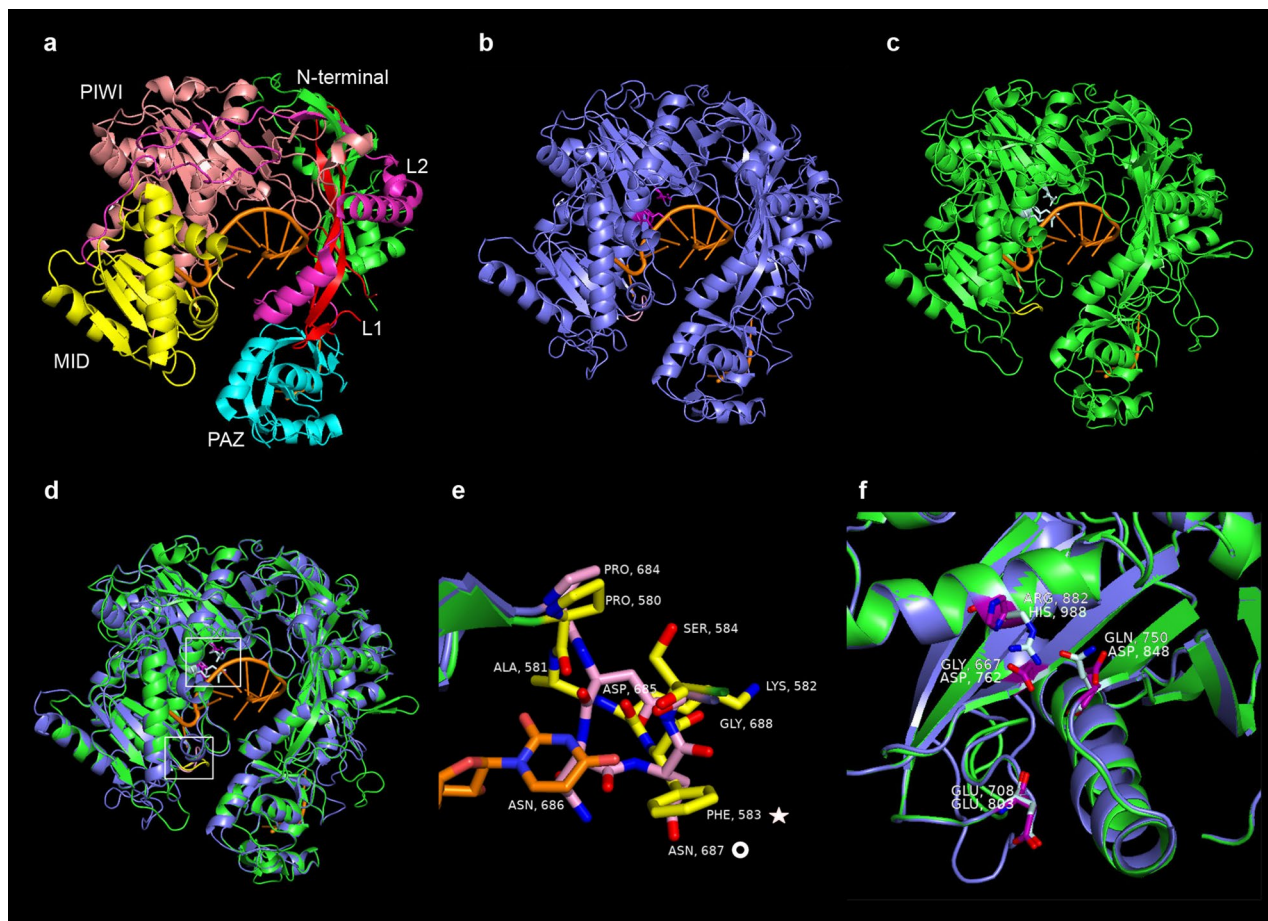
**Figure 3.** Maximum likelihood phylogeny (RAxML, model JTT +  $\Gamma$ , 150 replicates) of the Argonaute (AGO) family in potato (green) and rice (purple). Bootstrap values > 70% are indicated. Outgroup = *Physcomitrella patens* AGO16. Bar = number of substitutions per site.

species in the *Nicotiana* genus with numbers in the genera *Solanum* and *Petunia*. In potato, remnants of AGO gene duplications can be observed on chromosome 2 and 6 and in tomato on chromosome 1, 2, 3, and 6<sup>31</sup>. In potato, we discovered three *StAGO4* (*4*, *4a*, *4d*) genes whereas four was found in tomato (*SlAGO4a*, *4b*, *4c*, *4d*). When considering branch lengths, AGO4 is closer to AGO4b compared to AGO4c (Supplementary Fig. S3). In this case it is still not clear whether an incomplete gene duplication or a gene loss has occurred.

The split between *Nicotiana* and *Solanum* species is estimated to be rather recent *c.* 24 Myr<sup>22</sup>, hence gene family expansion and gene turnover rates should not deviate much between the two genera as found in our analysis. We can only speculate that human selection and clonal propagation could have had a major impact on gene content beside different duplication mechanisms as earlier suggested<sup>31</sup>.

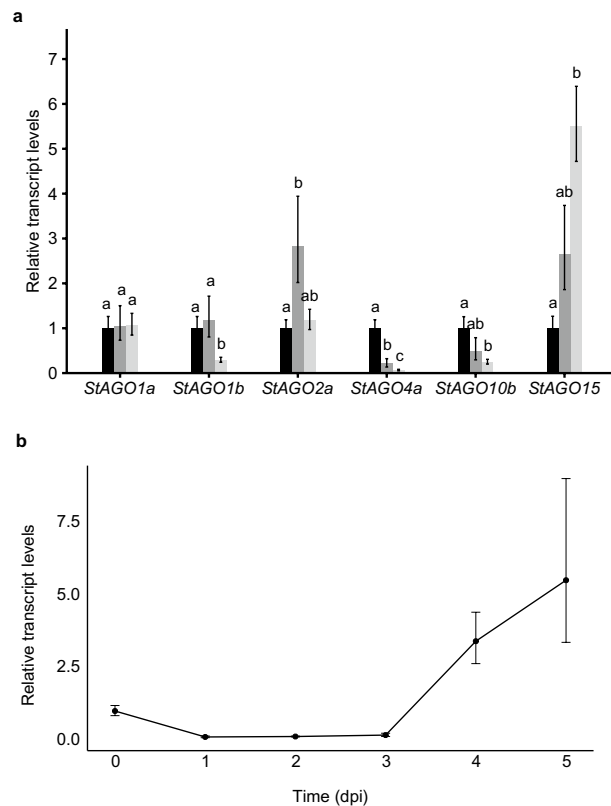
It is believed that a functional RNAi pathway was present in the last common ancestor for eukaryotes as a defense system against viruses and transposons, a system that has expanded to comprise regulation of endogenous RNAs<sup>32</sup>. Members of AGO super-proteins can be found in a majority of eukaryotic super-groups, where AGOs act as partners in a multi-protein regulatory system impacting an array of processes<sup>15</sup>. This multi-function feature also applies on plants, including defense. For example, AGO1, AGO2, AGO4, AGO5, AGO7, and AGO10 in *Arabidopsis* are known to participate in mechanisms involving defense responses towards different types of viruses<sup>33,34</sup>. More precisely antiviral AGOs associate with virus-derived small RNAs to repress complementary viral RNAs or DNAs, or with endogenous small RNAs to regulate host gene expression and promote antiviral defense. In infected *N. benthamiana* plants, 21 and 22 nt sRNA from the potato spindle tuber viroid associate with AGO1, AGO2 and AGO3, while 24 nt viroid sRNA bind to AGO4, AGO5, and AGO6<sup>35</sup>. Similar events are also reported from rice where OsAGO1 and OsAGO18 act against *Rice stripe tenuivirus* and *Rice dwarf phyto-reovirus*<sup>36</sup>. Not much is known about function of OsAGO15. It is believed to have evolved by duplication events followed by differentiation within the AGO4 clade<sup>37</sup>. OsAGO15 is expressed in reproductive tissue and harbor a D-D-H/P catalytic motif.





**Figure 4.** Three-dimensional AGO protein structure predictions. **(a)** Crystal structure of the human AGO2-miR-20a complex. N-terminal = green, L1 = red, PAZ = cyan, L2 = magenta, MID = yellow, PIWI = salmon. **(b)** AtAGO1 (steel blue) and Hs-miR20 (orange, PDB ID 4F3T:R). Residues of the D-E-D-H motif are in purple. The nucleotide specificity loop (NSL) in pink. **(c)** StAGO15 (green) and Hs-miR20 (orange). Residues of the G-E-Q-R motif in light cyan. The NSL in yellow. **(d)** Figure b and c merged. **(e)** Close view of the NSL in figure d (upper square). Nitrogen (red) and oxygen atoms (blue). The replacement of Asparagine (ASN, 687) in AtAGO1 to Phenylalanine (PHE, 583) in the StAGO15 NSL is highlighted with a white star (PHE, 583) and a white ring (ASN, 687). **(f)** Close view of the D-E-D-H/G-E-Q-R motives in figure d (lower square, Hs-miR20 removed). Nitrogen (red) and oxygen atoms (blue). StAGO15 residue labels are placed above the AtAGO1 labels.

We checked for the presence of nuclear export signal (NES) and the nuclear localization signal (NLS) domains in the 14 potato AGO sequences. StAGO1a and 1b were the only AGOs containing both NES and NLS domains, known to be of importance for nuclear-cytoplasmic shuttling<sup>38</sup>. High scores of NLS were only detected for StAGO15, suggesting nuclear localization. For nuclear to cytoplasm transport, potato has five members in the exportin family. However, details on translocation from the nucleus, including loading partners and associated processes reported in *Arabidopsis*, are not known in potato. The protein sequence of StAGO15 differs at the NSL and the catalytic tetrad sequences compared to AtAGO1. The G-E-Q-R motif is so far only observed in the Solanaceae AGO15 proteins. Uracil is the most preferred 5' nucleotide of AtAGO1 bound sRNA<sup>10</sup>, however adenine is the most hydrophobic nucleotide. The change from the polar Asn 687 in AtAGO1 to the hydrophobic Phe 583 in StAGO15, could indicate a preference for adenine as the 5'-nucleotide of sRNA binding. In *Arabidopsis* 5' A is a signature for a loading preference of 24 nt sRNAs<sup>39</sup>. These features open up several functional possibilities, including induction of 24 nt phasiRNAs upon pathogen infection. Overlapping functions cannot be excluded at this stage. Resistance genes can become negatively regulated by host miRNAs upon pathogen attack as a self-defense response. In tomato, particularly the miR482/2118 family are active and *R* gene mRNA can be targeted both by these miRNAs and by self-generated secondary sRNAs<sup>40</sup>. There are many *R* genes in individual plant species, not least in potato, and it is thought that self-suppression by RNA silencing is a strategy to balance costs and benefits under pathogen attack. However, there is a complex co-evolutionary relationship between sequence diversity in *R* genes and interactions of evolving miRNA where much remains to be clarified<sup>41</sup>. In this context adds the potential impact of miR8788 from *P. infestans* on susceptibility in potato during infection another level of complexity<sup>42</sup>.



**Figure 5.** Argonaute expression in potato cv. Desirée during pathogen challenge. **(a)** Relative transcript levels of *StAGO* genes during pathogen challenge five days post infection (dpi). Black = H<sub>2</sub>O, dark grey = *A. solani*, light grey = *P. infestans* NL 11388 strain. Error bars indicate mean  $\pm$  standard error of the mean (n = 4). Letters in the bar charts (a–c) represent significant differences (one-way ANOVA and Tukey’s HSD test: P < 0.05). **(b)** Relative transcript levels of *StAGO15* in a time-course from 0 to 5 dpi when infected by the *P. infestans* NL 11388 strain. Error bars indicate mean  $\pm$  standard error of the mean (n = 4).

## Methods

**Dataset assembly.** Genome sequences with annotated gene models present in Solanaceae databases (<http://solanaceae.plantbiology.msu.edu/index.shtml>, <https://solgenomics.net>) were searched to identify AGO homologs. A HMM-profile search using ‘HMMER’<sup>43</sup> with *e* value of 0.0001 was applied on the translated version of in house transcriptome data, generated from potato cv. Sarpo Mira inoculated with *Phytophthora infestans*<sup>44</sup>. Only full length, characterized plant AGO sequences (either at amino acid or transcript level), containing at least three characteristic domains, were downloaded from UniProt<sup>45</sup>. *Arabidopsis* and rice AGO homologs were retrieved from TAIR; [www.arabidopsis.org](http://www.arabidopsis.org) and RGAP; <http://rice.plantbiology.msu.edu/>. In total, 84 AGO protein sequences from 15 genomes and ten transcriptomes were used to construct the HMM profile. AGO sequences from *Arabidopsis* and tomato were used as queries for homology searches using tBLASTn<sup>46</sup> with *e* value 0.001 to mine the genome sequences. The above methods were also used for the mining of AGO homologs in other species in the *Solanaceae* family. Next, potential homologs were further confirmed by the presence or absence of three characteristic domains: N terminal, PAZ and PIWI using Pfam<sup>47</sup>. Domains were predicted using HMMscan (HmmerWeb version 2.41.1). Default parameters (cut-off values: > 25 sequence and 22 hit bit scores) were applied<sup>48</sup>. Tandem duplicated gene pairs were identified if gene pairs were located within a distance of 100 kb on the genome or if the gene pairs were separated with four genes from each other.

**Phylogenetic analysis.** The AGO homologs were aligned using ‘MAFFT v7.123b’<sup>49</sup> with ‘ensi’ option. Poorly aligned regions were cleaned using ‘trimAl’<sup>50</sup> and option ‘Automated1’. Phylogenetic trees were reconstructed using Maximum Likelihood (ML) method as implemented in RAxML v 8.2.11<sup>51</sup>. The best substitution model JTT +  $\Gamma$  was applied for all trees. Robustness of the topologies and branches were assessed with 100 or 150 bootstrap replicates. The AGO homolog from *Physcomitrella patens* was used as outgroup for the rooting of the analysis and the R package ggtree for drawing. To infer evolutionary events, the AGO gene family tree was reconciled with the species tree, generated by the NCBI taxonomy browser, using NOTUNG<sup>52</sup>. *Erythranthe guttata* was used as outgroup in the gain and loss gene predictions. Pairwise identities, genetic distances and corresponding Neighbor-Joining tree were computed using MEGA v.7<sup>53</sup>.

**In silico protein analysis and modeling.** Alignment of the MID domain and the D-E-D-H/D motif was displayed with the R package ggmsa and the secondary structure of the MID domain was predicted by the RaptorX-Property tool<sup>54</sup>. The potato MID domains are estimated from aligning them to the MID domain of AtAGO1<sup>27</sup>. Protein sequences of AtAGO1 and StAGO15 were used for three-dimensional structure modeling with SWISS-MODEL, against the AGO2-miR-20a complex (PDB ID 4F3T<sup>28</sup>), being the template with highest Global Model Quality Estimation number (0.61) and providing a miRNA to the model. The PyMOL Molecular Graphics System (PyMOL) was utilized for visualization of the predicted structures<sup>55</sup>. NES were predicted applying NESmapper (<https://sourceforge.net/projects/nestmapper/>) and NLS were predicted with cNLS Mapper ([http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS\\_Mapper\\_form.cgi](http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi)).

**Pathogens and inoculations.** Potato plants (cv. Désirée) were inoculated with *Phytophthora infestans* strain NL11388 as earlier described<sup>44</sup>. Leaf inoculations using 10 µl of a 20 × 10<sup>4</sup> spores/ml water of the fungus *Alternaria solani*, strain 142.2 collected from a field in Nymö, located in southern Sweden<sup>56</sup> were also performed. Sterile water was used as control.

**Quantitative real time PCR.** Total RNA was isolated from infected potato leaves and control samples using RNeasy Plant Mini Kit (Qiagen) according to the manufactory's recommendation. Prior to qRT-PCR analysis, 1 µg dsDNase (Thermo Scientific) treated RNA was reverse-transcribed into cDNA using Maxima H Minus First Strand cDNA Synthesis Kit (Thermo Scientific). At least four biological replicates were examined using iTaq Universal SYBR Green Supermix (Bio-Rad).

Sequences of DNA oligonucleotides are listed in Supplementary Table S1. The *StEF1α* and *StACT101* gene were used as the internal reference<sup>57</sup>.

***P. infestans* DNA quantification.** To evaluate *P. infestans* infection, its genomic DNA (gDNA) was quantified by qPCR essentially as described earlier<sup>42,58</sup>. Genomic DNA was isolated from potato leaves inoculated with *P. infestans*. Concentration of obtained gDNA was determined using Qubit dsDNA BR Assay Kit (Thermo Scientific). qPCR analyses with four biological replicates were carried out using iTaq Universal SYBR Green Supermix (Bio-Rad). 20 ng gDNA was used as template in each qPCR reaction together with primers for *PiO8* or *StACT101*. Primers are listed in Supplementary Table S1. All statistics were calculated as detailed as earlier described<sup>42</sup>.

**Transcriptome sequencing and bioinformatic analysis.** Leaves were collected from potato plants (cv. Sarpo Mira) 5 days post inoculation (dpi) using *P. infestans* isolates 88069 and 11388. Uninfected and water inoculated leaves were used as controls. For each sample at least 3 leaves were pooled. Total RNA was extracted using the RNeasy mini kit (QIAGEN). Thirteen transcript libraries followed by Illumina HiSeq 2500 sequencing were performed at the National Genomics Infrastructure, Illumina sequencing platform (Stockholm). The Illumina adaptor sequences and low-quality reads were removed using Trimmomatic v0.36<sup>59</sup>. The filtered datasets were mapped to *S. tuberosum* v4.04 and the *P. infestans* reference genomes<sup>60,61</sup> using kallisto v0.43.0<sup>62</sup>. Differential expression analysis was performed using the DESeq2 package<sup>63</sup>. All calculations were performed in R v3.2.0 ([www.R-project.org](http://www.R-project.org)).

## Data availability

Transcriptome data are available through the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus, Series accession number GSE159015 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159015>). Additional sequence and phylogeny data can be found in the Treebase repository, <http://purl.org/phylo/treebase/phyloids/study/TB2:S26565>.

Received: 6 July 2020; Accepted: 3 November 2020

Published online: 25 November 2020

## References

- Felden, B. & Paillard, L. When eukaryotes and prokaryotes look alike: the case of regulatory RNAs. *FEMS Microbiol. Rev.* **41**, 624–639 (2017).
- Brodersen, P. & Voinnet, O. The diversity of RNA silencing pathways in plants. *Trends Genet.* **22**, 268–280 (2006).
- Borges, F. & Martienssen, R. A. The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* **16**, 727–741 (2015).
- Bologna, N. G. & Voinnet, O. The diversity, biogenesis, and activities of endogenous silencing small RNAs in *Arabidopsis*. *Annu. Rev. Plant Biol.* **65**, 473–503 (2014).
- Höck, J. & Meister, G. The Argonaute protein family. *Genome Biol.* **9**, 210 (2008).
- Ipsaro, J. J. & Joshua-Tor, L. From guide to target: molecular insights into eukaryotic RNA-interference machinery. *Nat. Struct. Mol. Biol.* **22**, 20–28 (2015).
- Ma, X., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).
- Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E. & Weichenrieder, O. Crystal structure of the MID-PIWI lobe of the eukaryotic Argonaute protein. *Proc. Natl. Acad. Sci. USA* **108**, 10466–11047 (2011).
- Frank, F., Hauver, J., Sonenberg, N. & Nagar, B. *Arabidopsis* Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *EMBO J.* **31**, 3588–3595 (2012).
- Mi, S. *et al.* Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **133**, 116–127 (2008).
- Fátyol, K., Ludman, M. & Burgyán, J. Functional dissection of a plant Argonaute. *Nucleic Acids Res.* **44**, 1384–1397 (2015).



12. Herrera-Carrillo, E. & Berkhout, B. Dicer-independent processing of small RNA duplexes: mechanistic insights and applications. *Nucleic Acids Res.* **45**, 10369–10379 (2017).
13. Mukherjee, K., Campos, H. & Kolaczowski, B. Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. *Mol. Biol. Evol.* **30**, 627–641 (2013).
14. de Jong, D. *et al.* Multiple dicer genes in the early-diverging metazoa. *Mol. Biol. Evol.* **26**, 1333–1340 (2009).
15. Swarts, D. C. *et al.* The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**, 743–753 (2014).
16. Zhao, T. *et al.* A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes & Devel.* **21**, 1190–1203 (2007).
17. Zhang, H., Xia, R., Meyers, B. C. & Walbot, V. Evolution, functions, and mysteries of plant ARGONAUTE proteins. *Curr. Opin. Plant Biol.* **27**, 84–90 (2015).
18. Fang, X. & Qi, Y. RNAi in plants: an Argonaute-centered view. *Plant Cell* **28**, 272–285 (2016).
19. Vaucheret, H. Plant ARGONAUTES. *Trends Plant Sci.* **13**, 350–358 (2008).
20. You, C. *et al.* Conservation and divergence of small RNA pathways and microRNAs in land plants. *Genome Biol.* **18**, 158 (2017).
21. Singh, R. K., Gase, K., Baldwin, I. T. & Pandey, S. P. Molecular evolution and diversification of the Argonaute family of proteins in plants. *BMC Plant Biol.* **15**, 23 (2015).
22. Bombarely, A. *et al.* Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nature Plants* **2**, 16074 (2016).
23. Särkinen, T., Bohs, L., Olmstead, R. G. & Knapp, S. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 100-tip tree. *BMC Evol. Biol.* **13**, 214 (2013).
24. Wang, P. *et al.* Factors influencing gene family size variation among related species in a plant family, Solanaceae. *Genome Biol. Evol.* **10**, 2596–2613 (2018).
25. Bai, M. *et al.* Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in *Solanum lycopersicum*. *Gene* **501**, 52–62 (2012).
26. Zheng, L. *et al.* Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat. Comm.* **5**, 4956 (2014).
27. Zha, X., Xia, Q. & Adam Yuan, Y. Structural insights into small RNA sorting and mRNA target binding by Arabidopsis Argonaute Mid domains. *FEBS Lett.* **586**, 3200–3207 (2012).
28. Elkayam, E. *et al.* The structure of human argonaute-2 in complex with miR-20a. *Cell* **150**, 100–110 (2012).
29. Avrova, A. O. *et al.* A novel *Phytophthora infestans* haustorium-specific membrane protein is required for infection of potato. *Cell. Microbiol.* **10**, 2271–2284 (2008).
30. Barchi, L. *et al.* A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Sci. Rep.* **9**, 11769 (2019).
31. Xian, Z. *et al.* Molecular cloning and characterization of SLAGO family in tomato. *BMC Plant Biol.* **13**, 126 (2013).
32. Muhammad, T., Zhang, F., Zhang, Y. & Liang, Y. RNA interference: a natural immune system of plants to counteract biotic stressors. *Cells* **8**, 38 (2019).
33. Garcia-Ruiz, H. *et al.* Roles and programming of Arabidopsis ARGONAUTE proteins during Turnip mosaic virus infection. *PLoS Pathog.* **11**, e1004755 (2015).
34. Carbonell, A. & Carrington, J. C. Antiviral roles of plant ARGONAUTES. *Curr. Opin. Plant Biol.* **27**, 11–117 (2015).
35. Minoia, S. *et al.* Specific Argonautes selectively bind small RNAs derived from Potato Spindle Tuber Viroid and attenuate viroid accumulation *in vivo*. *J. Virol.* **88**, 11933–11945 (2014).
36. Wu, J. *et al.* Viral-inducible Argonaute18 confers broad-spectrum virus resistance in rice by sequestering a host microRNA. *eLife* **4**, e05733 (2015).
37. Kapoor, M. *et al.* Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genom.* **9**, 451 (2008).
38. Bologna, N. G. *et al.* Nucleo-cytosolic shuttling of ARGONAUTE1 prompts a revised model of the plant microRNA pathway. *Mol. Cell* **69**, 709–719.e5 (2018).
39. Havecker, E. R. *et al.* The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell* **22**, 321–334 (2020).
40. Canto-Pastor, A. *et al.* Enhanced resistance to bacterial and oomycete pathogens by short tandem target mimic RNAs in tomato. *Proc. Nat. Acad. Sci. USA* **116**, 2755–2760 (2019).
41. Zhang, Y., Xia, R., Kuang, H. & Meyers, B. S. The diversification of plant NBS-LRR defense genes directs the evolution of microRNAs that target them. *Mol. Biol. Evol.* **33**, 2692–2705 (2016).
42. Hu, X. *et al.* *Phytophthora infestans* Ago1-bound miRNA promotes potato late blight disease. *BioRxiv* <https://doi.org/10.1101/2020.01.28.924175> (2020).
43. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comp. Biol.* **7**, e1002195 (2011).
44. Jahan, S. *et al.* Plant-mediated gene silencing restricts growth of the potato late blight pathogen *Phytophthora infestans*. *J. Exp. Bot.* **66**, 2785–2794 (2015).
45. Apweiler, R. *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **32**, D115–D119 (2004).
46. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
47. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
48. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).
49. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
50. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
51. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
52. Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **7**, 429–447 (2000).
53. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
54. Wang, S., Li, W., Liu, S. & Xu, J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res.* **44**, 430–435 (2016).
55. Schrödinger, L.L.C. *The PyMOL Molecular Graphics System, Version 1.8.4, Open-Source PyMOL* (2016).
56. Odilbekov, F. *et al.* Within-season changes in *Alternaria solani* populations in potato in response to fungicide application strategies. *Eur. J. Plant Pathol.* **155**, 953–965 (2019).
57. Tang, X., Zhang, N., Si, H. & Calderón-Urrea, A. Selection and validation of reference genes for RT-qPCR analysis in potato under abiotic stress. *Plant Meth.* **13**, 85 (2017).
58. Llorente, B. *et al.* A quantitative real-time PCR method for in planta monitoring of *Phytophthora infestans* growth. *Letts. Appl. Microbiol.* **51**, 603–610 (2010).

59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
60. Haas, B. J. *et al.* Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**, 393–398 (2009).
61. Hardigan, M. A. *et al.* Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. *Plant Cell* **28**, 388–405 (2016).
62. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

### Acknowledgements

The authors want to thank Dr. Erland Liljeroth for providing the *A. solani* isolate. The work was supported by the Carl-Trygger's Foundation (17:122), the Swedish Research Council VR (2015-04259), and the Swedish University of Agricultural Sciences.

### Author contributions

Z.H., K.P.H., R.S. and C.D. planned and designed the research. Z.H. performed experiments. K.P.H. and R.S. analyzed data. Z.H., K.P.H., C.D. wrote the manuscript. All authors reviewed the manuscript.

### Funding

Open Access funding provided by Swedish University of Agricultural Sciences.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-77593-y>.

**Correspondence** and requests for materials should be addressed to C.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020