Supplemental Figures
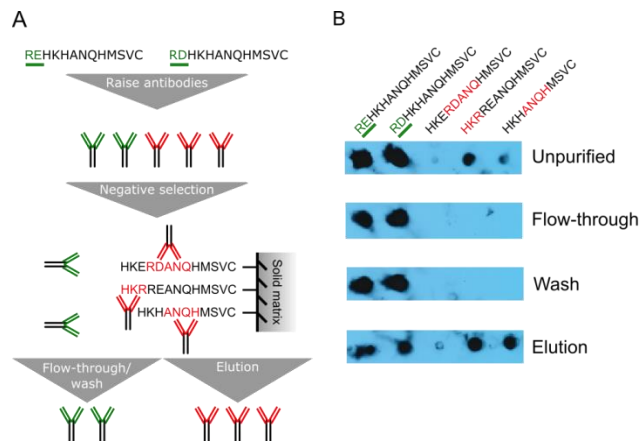

Fig. S1. Schematic overview depicting the negative selection of the initial composition of antibodies against synthetic peptides bearing N-terminal arginine. (A). The sequences used to raise the antibodies (green) as well as the "scrambled" sequences (red) were chosen according to (1). (B) Dot-blot analysis of the purification efficiency of the negative selection. Antibodies from the flow-through and the wash fractions were highly selective for N-terminal arginine. These were used for subsequent immunoprecipitations.
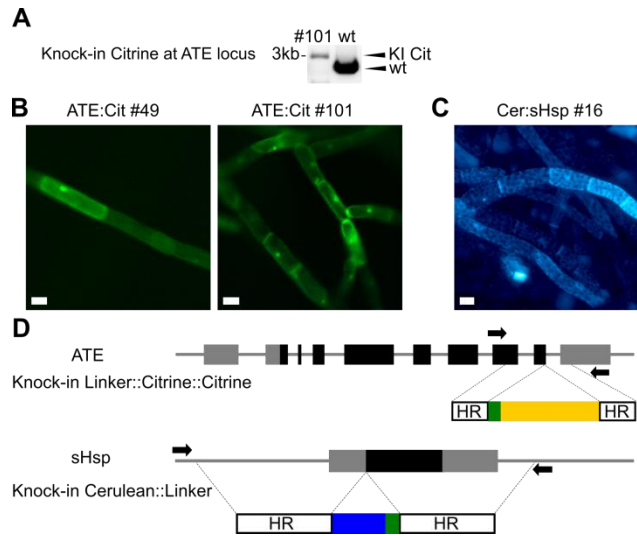
Fig. S2. Knock-in constructs and resulting transgenic moss reporter lines. (A) PCR using genomic DNA, validating the single integration of the linker-Citrine-Citrine construct at the ATE locus for ATE:Citrine#101. (B) Epifluorescence images of protonema from transgenic knock-in lines carrying a single fluorophore. ATE:Citrine fluorescence is present at a low level in many cells, with a higher intensity in many nuclei. (C) Cerulean:sHsp17.2a has a similar subcellular localization pattern in nuclei and cytosol, but is highly abundant in selected cells. Scale bars: 20 μm. (D) Schematic representation of constructs for the generation of fluorophore knock-ins at the endogenous genomic loci of ATE and sHsp17.2a: coding sequence (black), UTRs (grey), homologous regions (HR), Citrine (yellow box), Cerulean (blue box), flexible linker (2) (green box), and primers (arrows) annealing to genomic regions upstream and downstream of the homologous regions used in the knock-in constructs, thus amplifying the full target locus for the validation of integration are shown. For primer sequences, see Material and Methods.

2
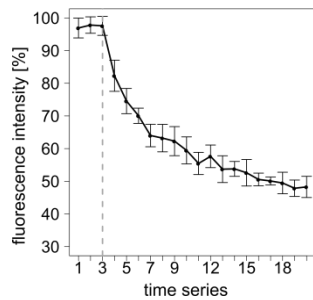
Fig. S3. Bleaching tests for Citrine in the ATE:Citrine knock-in line #49 (background line for FRET lines). The graph displays average values of fluorescence intensities in the Citrine channel after linear unmixing (with standard deviation for n = 4 independent experiments). The grey dashed line indicates start of bleaching (50% laser intensity 514 nm, 50 iterations), bleaching was repeated after every step for 17 imaging cycles. Residual non-bleachable fluorescence signal remained in the Citrine channel, probably due to plant cell autofluorescence. Based on these data, 10 imaging and bleaching cycles were chosen for the FRET experiments.
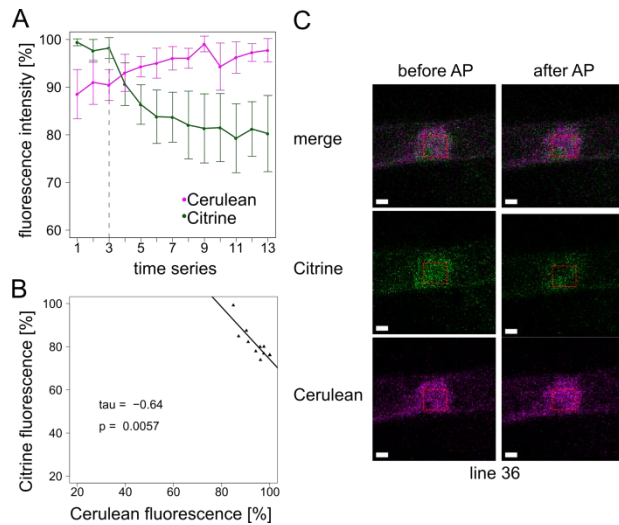
Fig. S4. FRET between Cerulean:sHSP17.2a and ATE:Citrine in stable transgenic moss (*Physcomitrella patens*) lines. FRET was analyzed using acceptor (Citrine) photobleaching and monitoring of the increase in donor (Cerulean) fluorescence. (A) Fluorescence intensities of Citrine and Cerulean were monitored using 458 nm excitation before and after acceptor photobleaching (AP) with 514 nm excitation (the grey dashed line indicates start of bleaching, bleaching was repeated after every step). The graph shows average values (+/- standard deviation) from n = 6 independent experiments. (B) Negative correlation between the decrease of the Citrine fluorescence and the increase of the Cerulean fluorescence is strong (Kendall´s tau = -0.64) and statistically significant (p-value = 0.0057), indicating FRET. (C) Exemplary Confocal microscope images showing start (before AP) and end (after AP) of a bleaching experiment in the nucleus of the stable double transgenic Cerulean:sHSP17.2a/ATE:Citrine knock-in line #36. Scale bars: 2 µm.
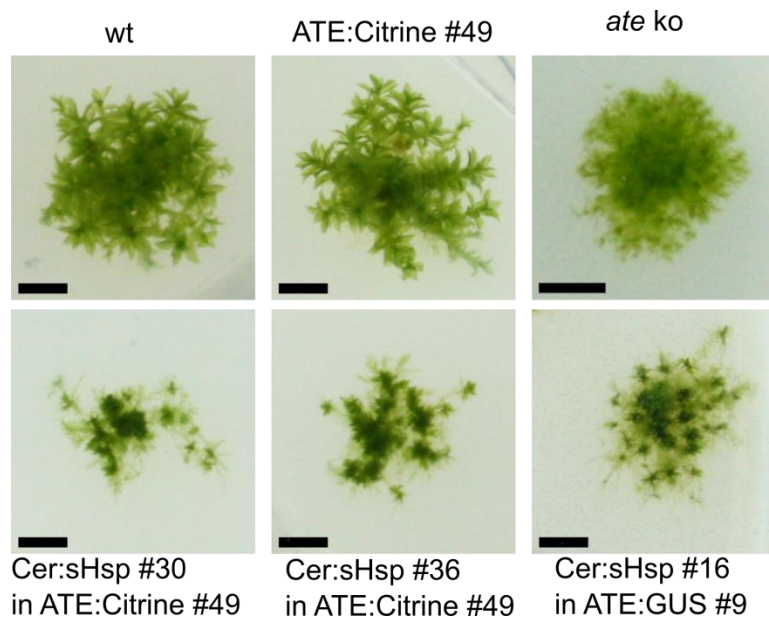
Fig. S5. Phenotype of transgenic moss (*Physcomitrella patens*) lines. Moss plants of the different transgenic lines were grown for 42 d on Knop-Agar with an 18 h light/6 h dark light regime. Whereas the ATE:Citrine knock-in line grew like wildtype (WT), indicating complete functionality of the tagged ATE protein, *ate* plants (3) show no formation of gametophores and are arrested in the protonema state. Transgenic lines carrying the tagged sHsp17.2a protein (Cer:sHSP) show retarded formation of gametophores, pointing to a limited functionality of the tagged protein concerning gametophore formation. This data additionally hints to an involvement of sHsp17.2a in moss development.

5
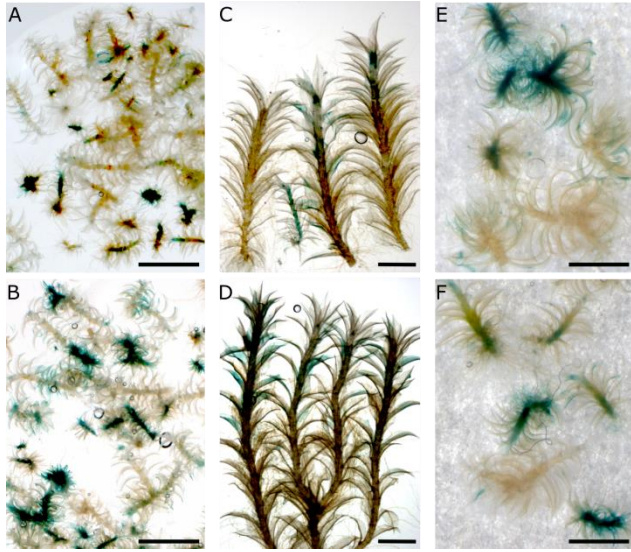
Fig. S6. Histochemical GUS-staining of ATE:GUS gametophores used for IP of arginylated proteins. (A) Untreated gametophores from liquid culture. (B) Supplementation with 1% glucose for one day led to an increase of staining compared to the control (A). (C) Untreated gametophores cultivated on hydroponic ring cultures. (D) Cultivation for seven days in darkness led to a slight increase of the GUS staining compared to the control (C). (E) Untreated gametophores from liquid culture. (F) Treatment with 100 µM MG132 did not change the GUS staining compared to the control (E).
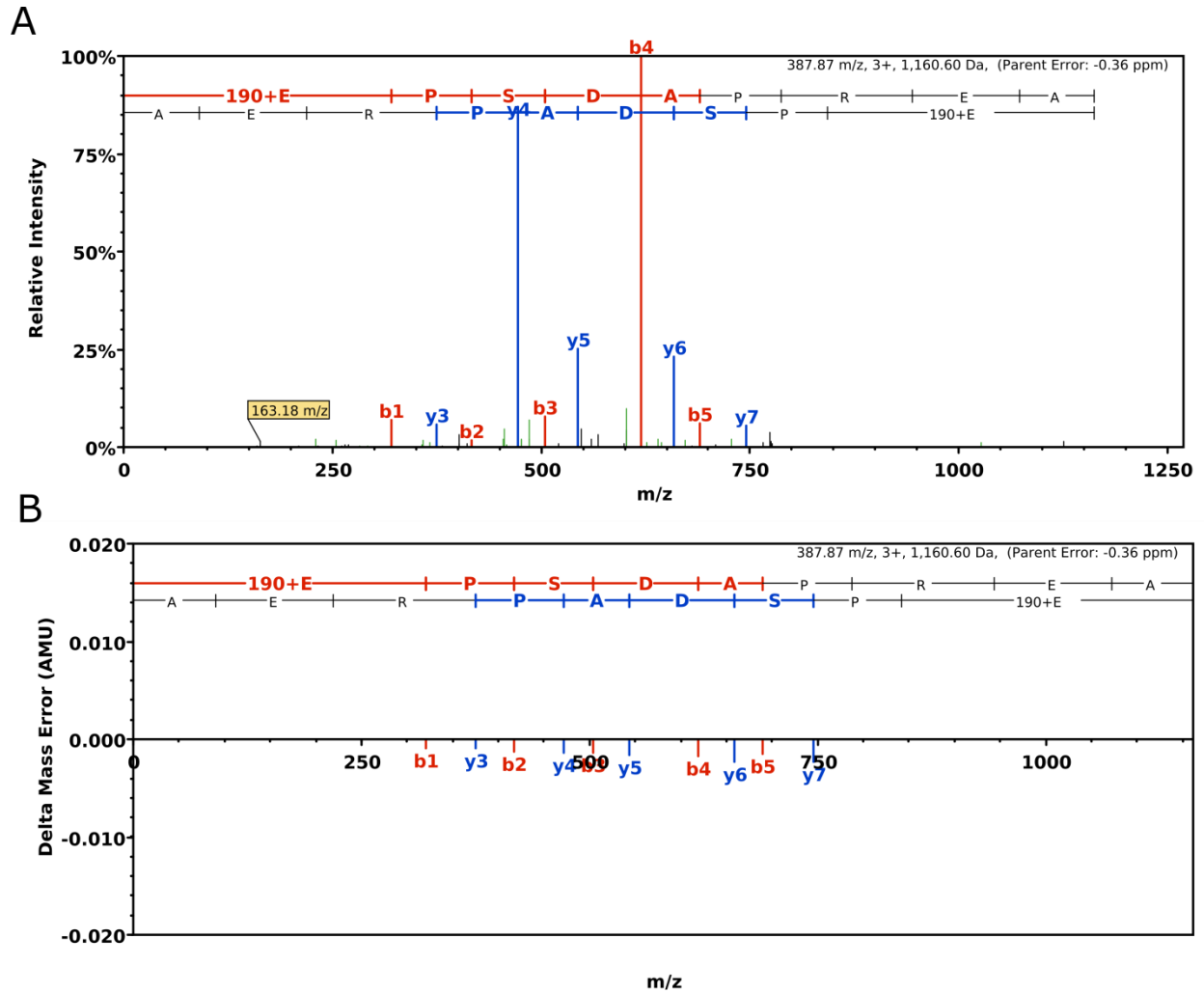
Fig. S7. **False positive identification of an N-terminally arginylated protein.** (A) HCD fragment ion spectrum of a peptide that is falsely annotated to be arginylated as the N-terminal arginine is derived from the genomic sequence (Pp1s619_3V6.1, R[304].EPSDAPREA.T[314]). The immonium ion of the N-terminal dimethylated arginine (163.18) is highlighted. (B) The fragment errors of the b- and y-ion series are homogeneously distributed.
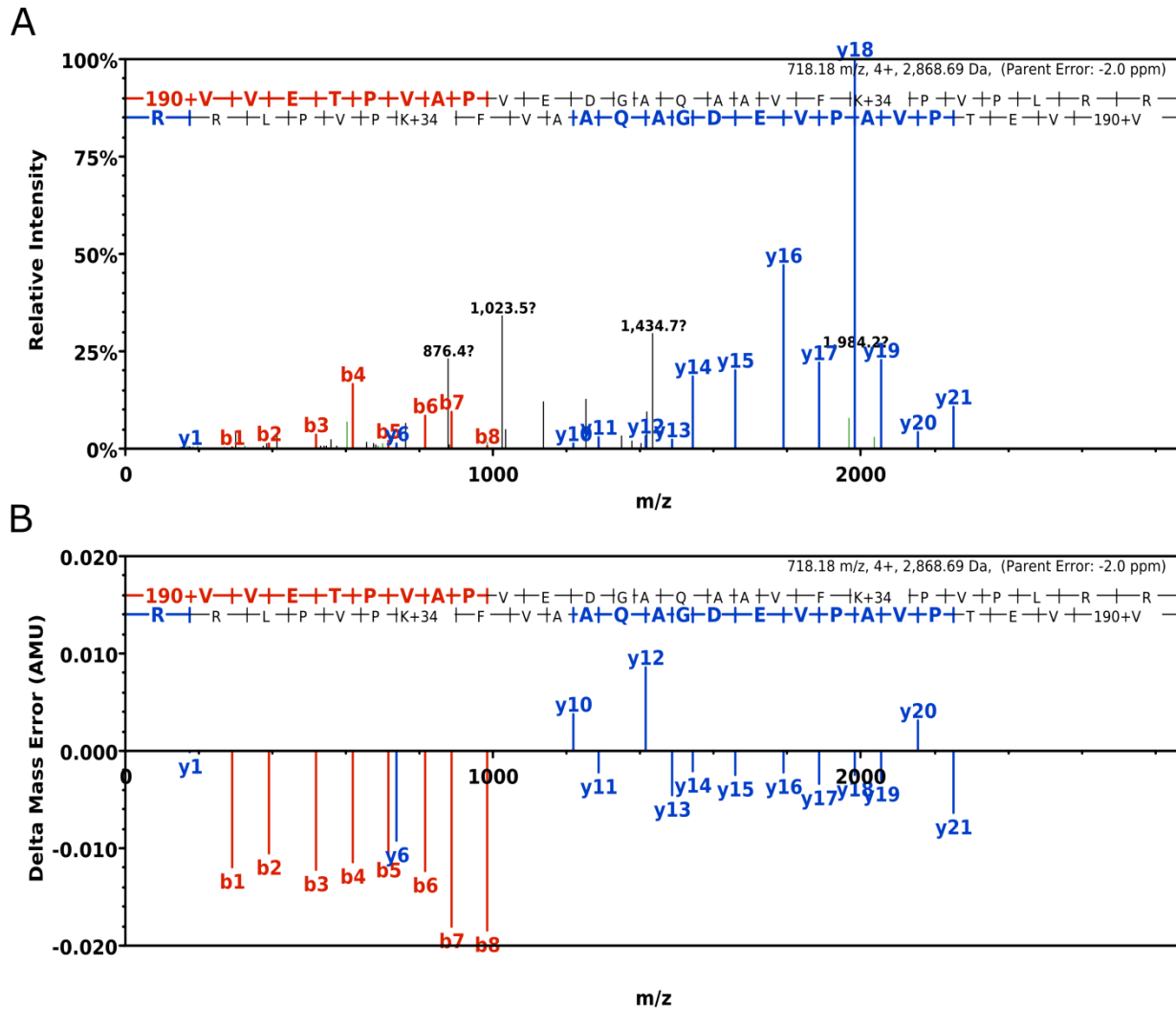
Fig. S8. False positive identification of an N-terminally arginylated protein. (A) HCD fragment ion spectrum of a peptide that is falsely annotated to be arginylated. The spectrum matches to Pp1s212_26V6.1. The preceding amino acids of the peptide G^3.VVETPVAPVEDGAQAAVFKPVPLRR.V^{29} are V^2G^3. The monoisotopic mass of VG is 156.0899, the monoisotopic mass of an arginine residue is 156.1011. This mass shift cannot be distinguished at the precursor level (-2.0 ppm). (B) The error distribution of the b-ion series (red) differs remarkably from that of the y-ion series (blue).
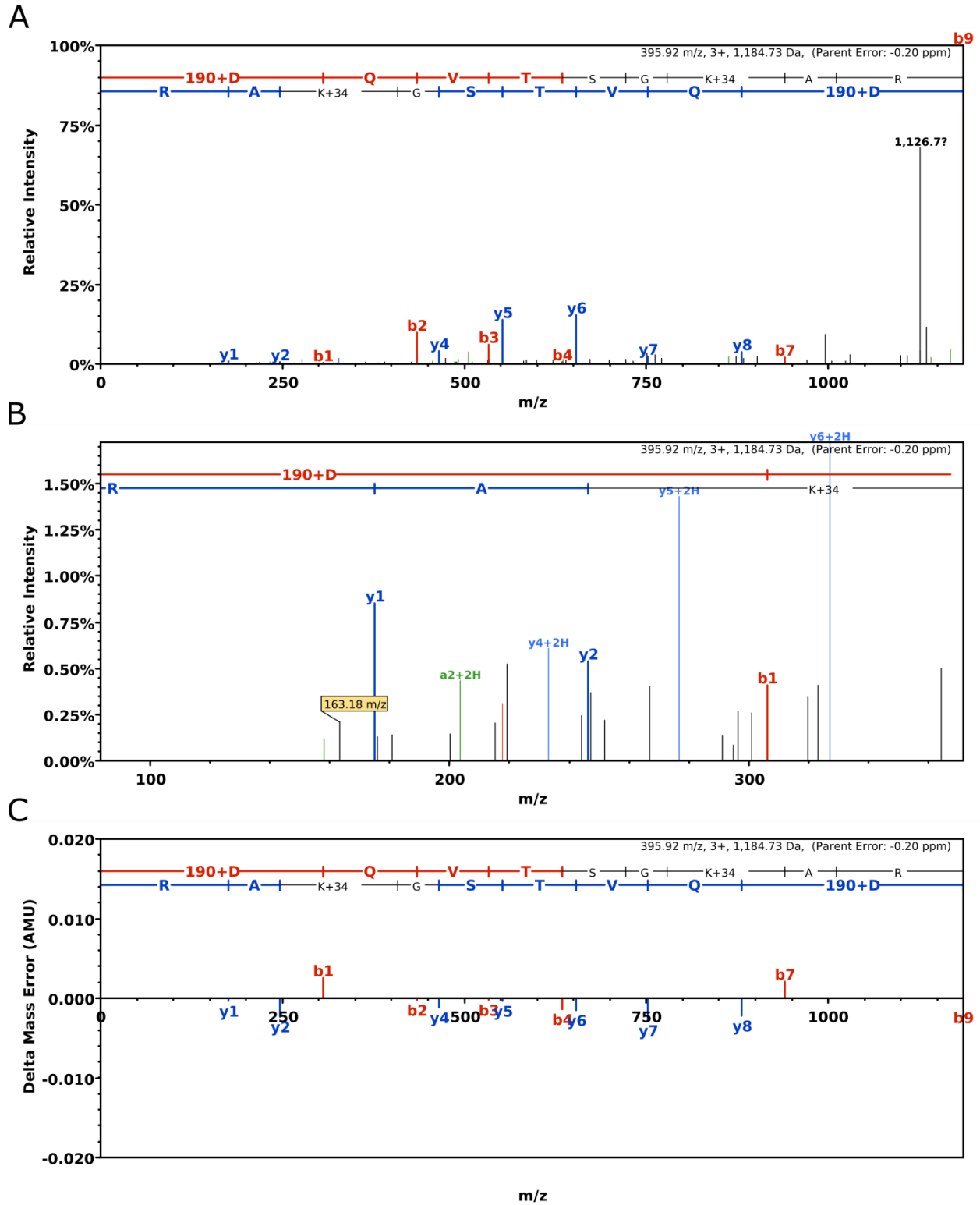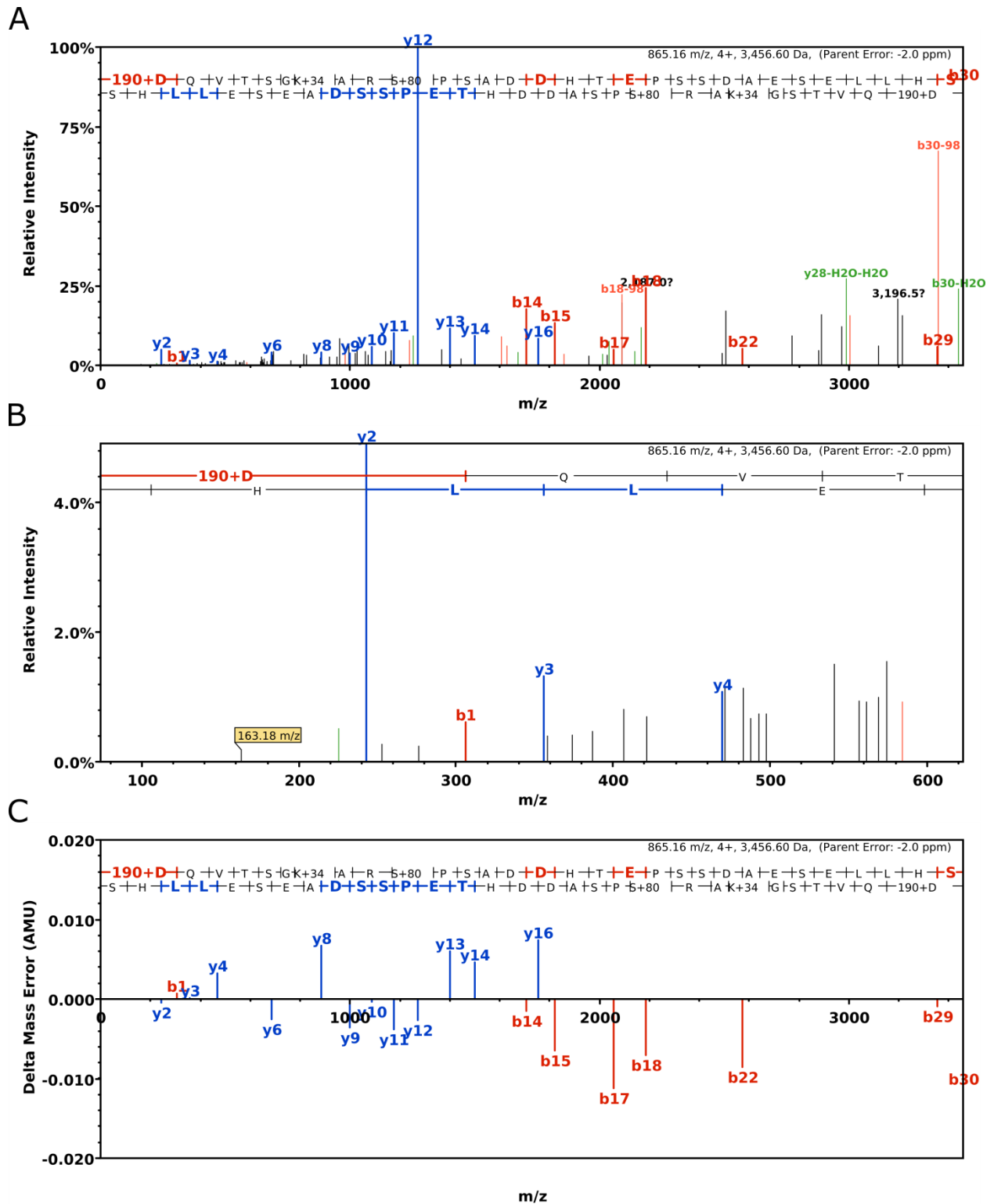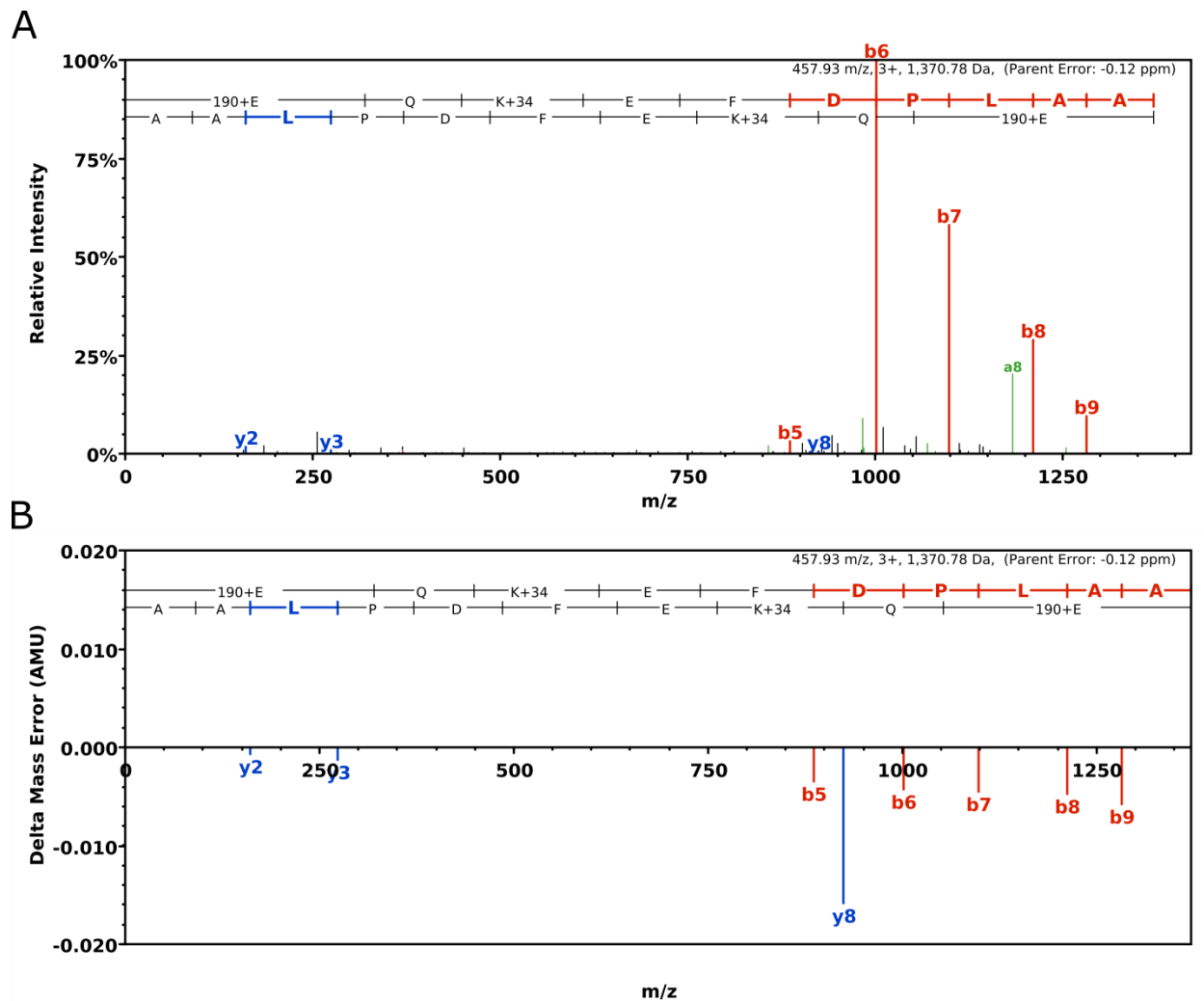
8

Fig. S9. True positive identification of an N-terminally arginylated protein. (A) HCD

fragment ion spectrum of an arginylated peptide. (B) Zoomed range of the spectrum showing the representative immonium ion of an N-terminally dimethylated arginine (163.18). The peptide (M^1.DQVTSGKAR.S^{11}) matches to Pp1s619_3V6.1. (C) The fragment errors of the b- and y-ion series are homogeneously distributed.

10

Fig. S10. True positive identification of an N-terminally arginylated protein. (A) HCD

fragment ion spectrum of an arginylated peptide. (B) Zoomed range of the spectrum showing the

11

representative immonium ion of an N-terminally dimethylated arginine (163.18). The peptide (M[1].DQVTSGKARSPSADDHTEPSSDAESELLHS.L[32]) matches to Pp1s619_3V6.1. (C) The fragment errors of the b- and y-ion series are homogeneously distributed.
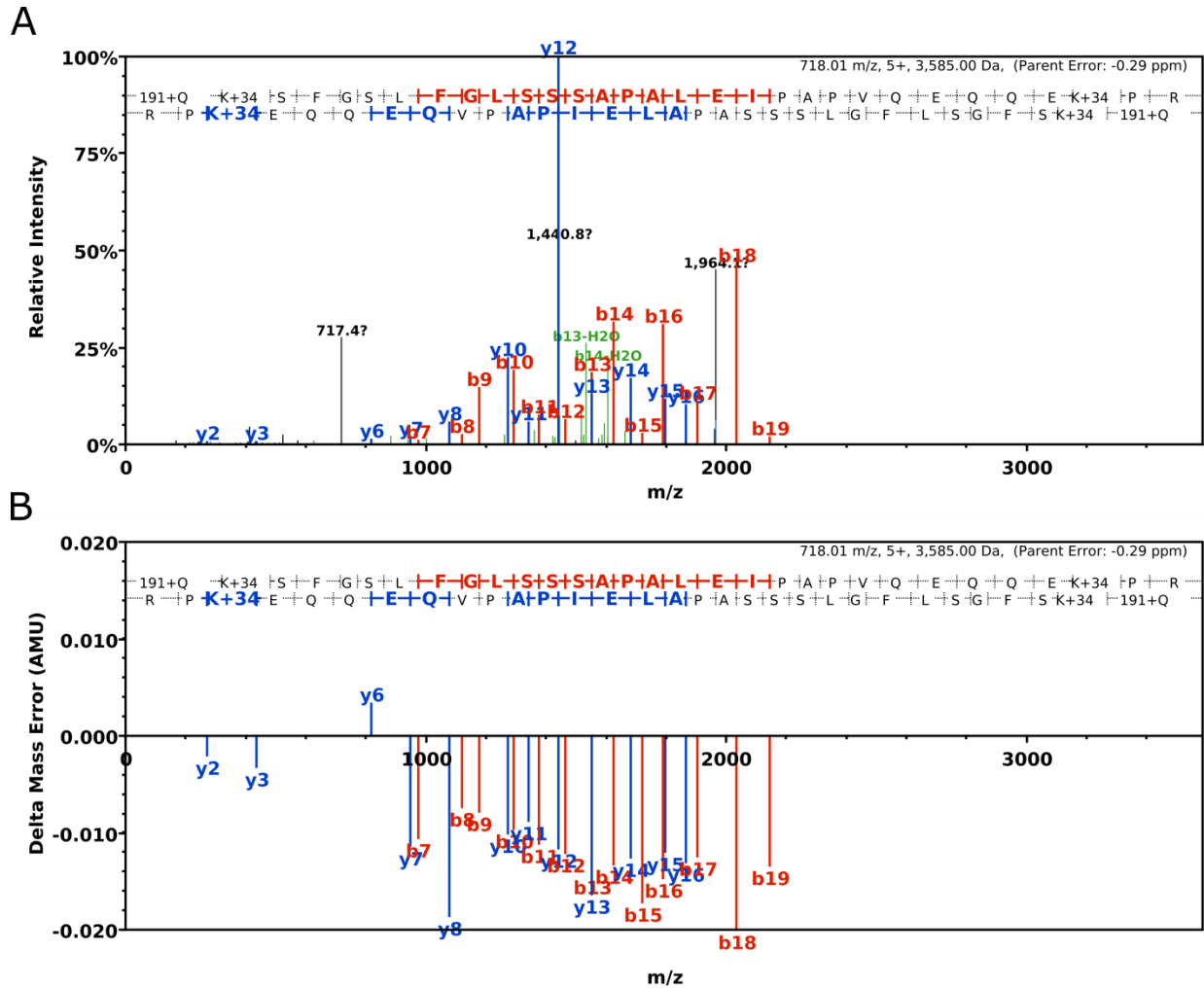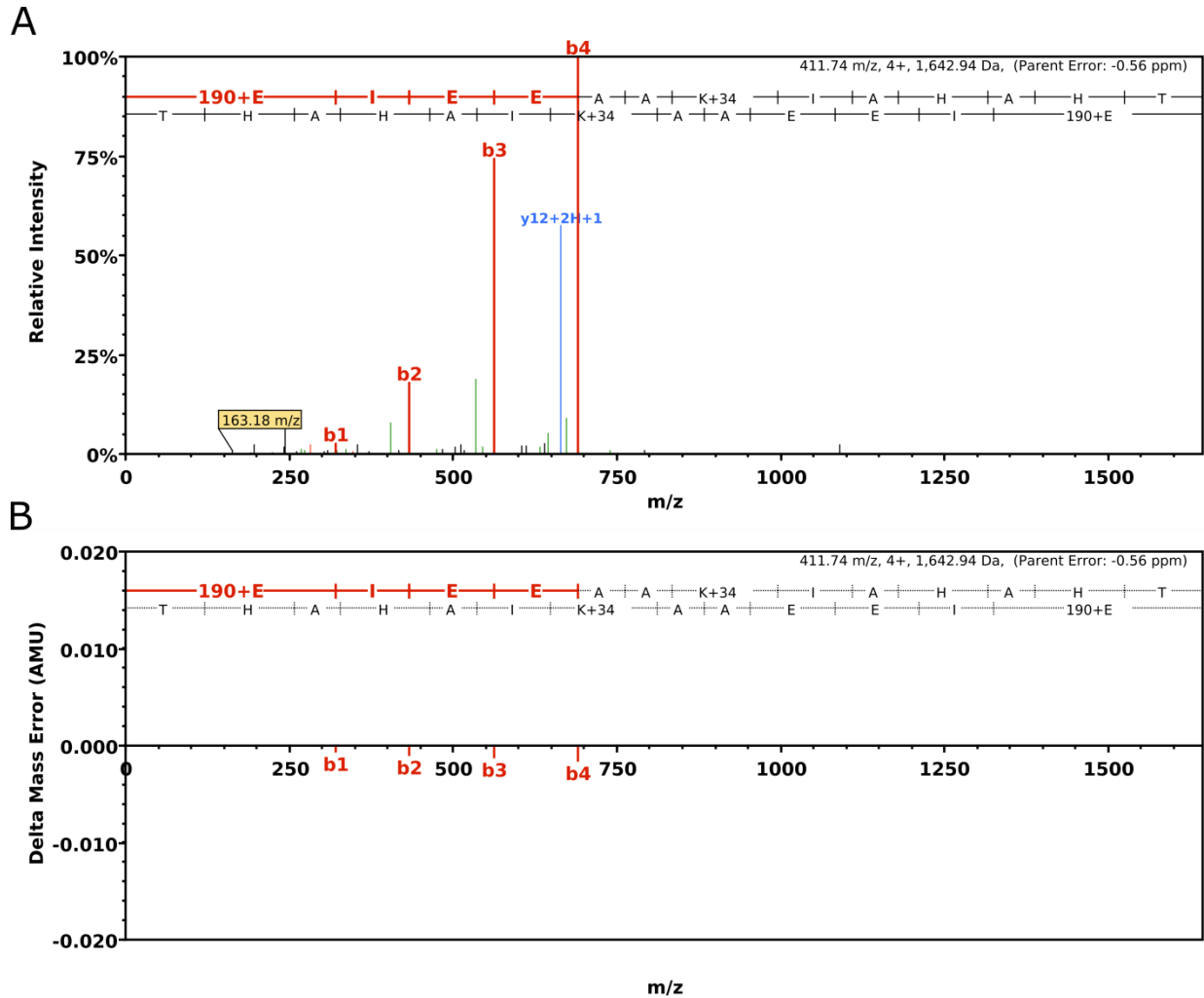
Fig. S11. True positive identification of an N-terminally arginylated protein. (A) HCD fragment ion spectrum of an arginylated peptide. The representative immonium ion of an N-terminally dimethylated arginine (163.18) is not present. The peptide (G^2.EQKEFDPLAA.K^{13}) matches to Pp1s68_62V6.1. (B) The fragment errors of the b- and y-ion series are homogeneously distributed.

Fig. S12. True positive identification of an N-terminally arginylated protein. (A) HCD fragment ion spectrum of an arginylated peptide. Prior to arginylation, the N-terminal glutamine residue was deamidated to glutamate. The representative immonium ion of an N-terminally dimethylated arginine (163.18) is not detectable. The peptide (T[29].QKSFGSLFGLSSSAPALEIPAPVQEQQEKPR.N[61]) matches to Pp1s106_174V6.1. (B) The fragment errors of the b- and y-ion series are homogeneously distributed.

14

Fig. S13. True positive identification of an N-terminally arginylated protein. (A) HCD fragment ion spectrum of an arginylated peptide. (B) Zoomed range of the spectrum showing the representative immonium ion of an N-terminally dimethylated arginine (163.18). The peptide (D^{528}.EIEEAAKIAHAHT.F^{542}) matches to Pp1s29_108V6.1. (C) The fragment errors of the b-series are homogeneously distributed. Fragment errors for the y-ion series are not present.
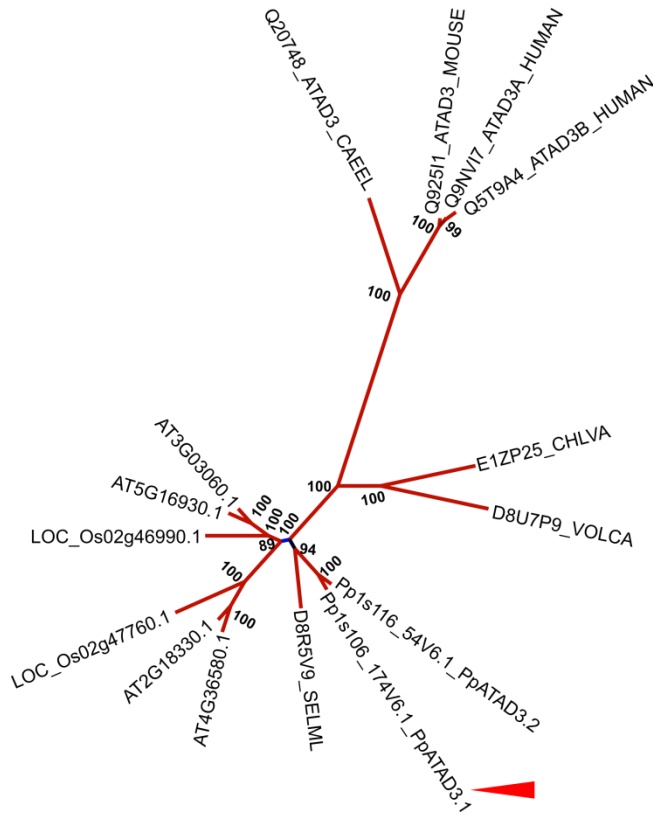
Fig. S14. Phylogenetic tree of homologous proteins of PpATAD3.1. Unrooted phylogenetic tree of homologous putative AAA-type ATPases of the identified N-terminally arginylated PpATAD3.1 (Pp1s106_174V6.1, red arrow) including proteins with a combination of an N-terminal DUF domain (Pfam DUF3523) and a C-terminal AAA domain (Pfam PF00004) from *P. patens* (Pp1s106_174V6.1_PpATAD3.1, Pp1s116_54V6.1_PpATAD3.2)*, A. thaliana* (AT3G03060.1, AT5G16930.1, AT2G18330.1, AT4G36580.1), *O. sativa* (LOC_Os02g47760.1, LOC_Os02g46990.1), *S. moellendorffii* (D8R5V9_SELML), *C. variabilis* (E1ZP25_CHLVA), *V. carteri* (D8U7P9_VOLCA), *C. elegans* (Q20748_ATAD3_CAEEL), *M. musculus* (Q925I1_ATAD3_MOUSE) and *H. sapiens* (Q9NVI7_ATAD3A_HUMAN, Q5T9AV_ATAD3B_HUMAN). Node colors and values represent posterior probabilities.

16
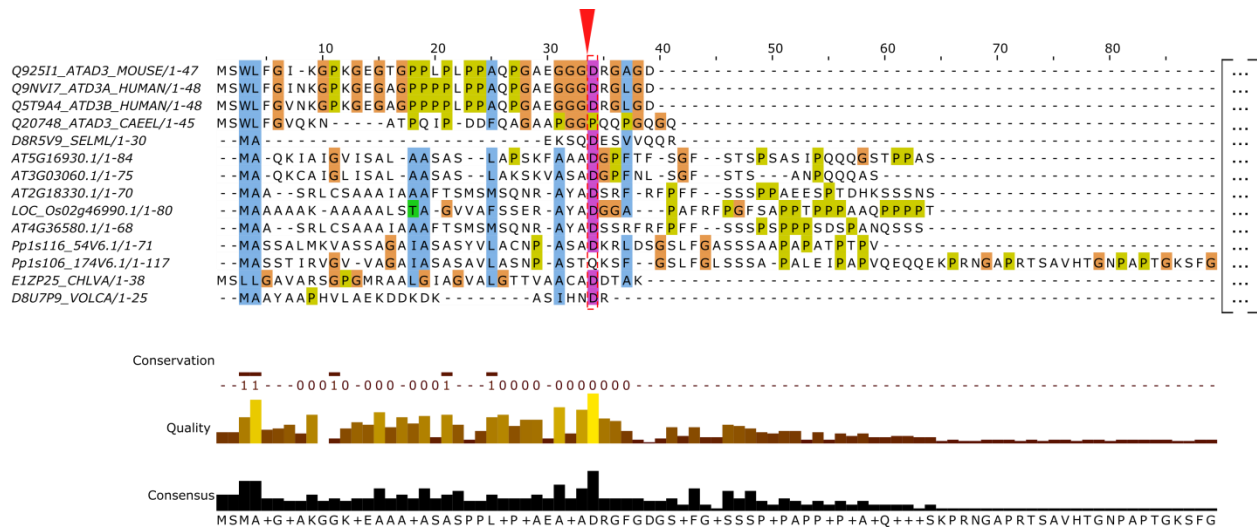
Fig. S15. Section from a multiple sequence alignment of homologous proteins of PpATAD3.1 (Pp1s106_174V6.1). The same homologous proteins sequences as used for the phylogeny were used (Fig S. 14, see also Material and Methods). Protein sequences were aligned with Jalview version 2.9.0b2 (5) using the Muscle algorithm (default settings) and further adapted manually. The coloring is according to the *ClustalX* color scheme. The red arrow indicates the observed cleavage site in PpATAD3.1 (Pp106_174V6.1). The red dashed box encompasses the identified arginylated residue Q^{30} (deamidated to E^{30}) of PpATAD3.1 and the aligned residues from the homologous proteins. The red arrow indicates the cleavage site. As N-termini differ in sequence similarity and length, a conservation of the arginylated residue is clear for plants and possible for animals.

17
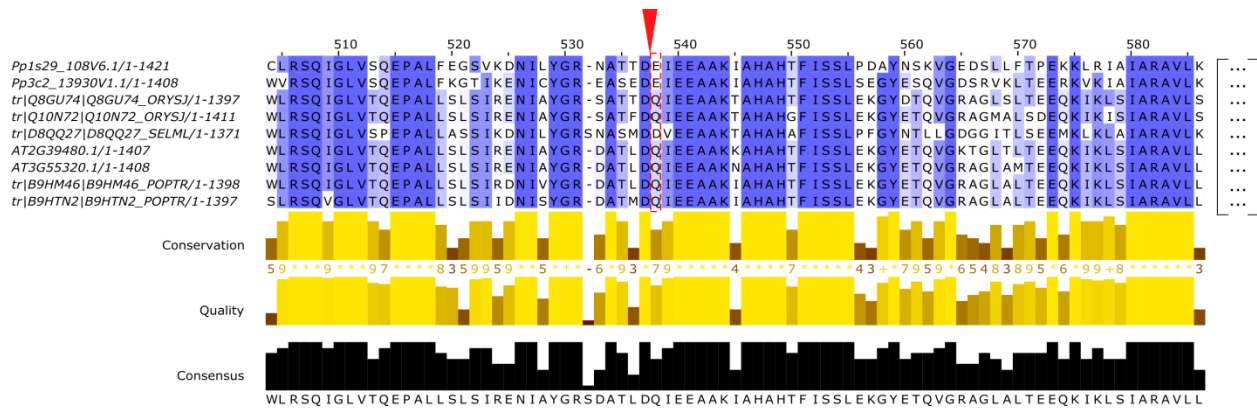
Fig. S16. Section from a multiple sequence alignment of homologous proteins of PpABCB20 (Pp1s29_108V6.1). Homologous proteins of *O. sativa* (Q8GU74, Q10N72), *A. thaliana* (AT2G39480.1, AT3G55320.1), *S. moellendorffii* (D8QQ27) and *P. trichocarpa* (B9HM46, B9HTN2) were chosen according to the previously established protein family definitions (4) as described in (3). Protein sequences were aligned with Jalview version 2.9.0b2 (5) using the Muscle algorithm (default settings). The color indicates percentage of identity. The red arrow indicates the observed cleavage site in PpABCB20 (Pp1s29_108V6.1). The red dashed box encompasses the identified arginylated residue E^{528} of PpABCB20 and the aligned residues from the homologous proteins. Pp3c2_13930V1.1 represents a paralogous protein from *P. patens* which is based on *P. patens* genome version 3 (cosmoss.org). Protein sequences were downloaded from Uniprot (uniprot.org) in the case of *S. moellendorffii*, *P. trichocarpa* and *O. sativa* or TAIR (arabidopsis.org) for *A. thaliana*.
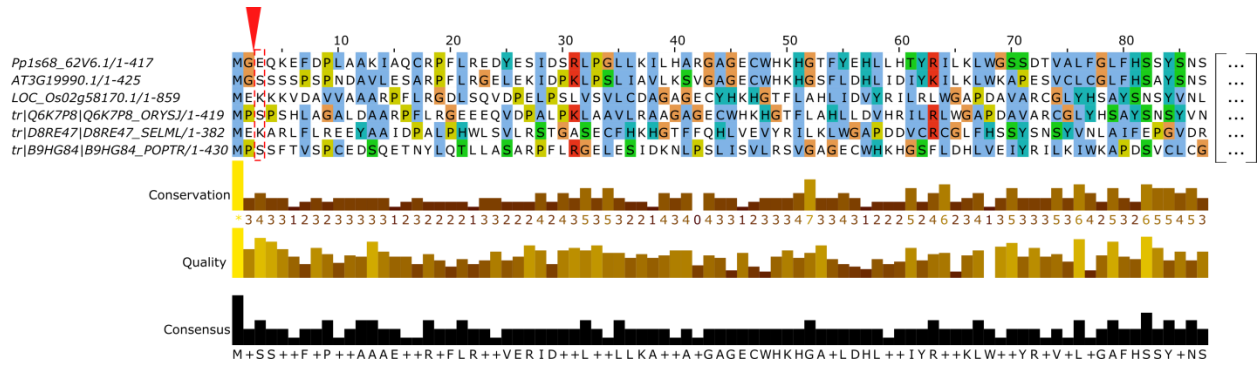
Fig. S17. Section from a multiple sequence alignment of homologous proteins of UP (Pp1s68_62V6.1). Homologous proteins of *O. sativa* (Loc_Os02g58170.1, Q6K7P8), *A. thaliana* (AT3G19990.1), *S. moellendorffii* (D8RE47) and *P. trichocarpa* (B9HG84) were chosen according to the previously established protein family definitions (4) as described in (3). Protein sequences were aligned at the initiator methionine of the corresponding protein model with Jalview version 2.9.0b2 (5). The coloring is according to the *ClustalX* color scheme. The red arrow indicates the observed cleavage site in UP (Pp1s29_108V6.1). The red dashed box encompasses the identified arginylated residue E^3 of UP and the aligned residues from the homologous proteins. Protein sequences were downloaded from Uniprot (uniprot.org) in the case of *S. moellendorffii*, *P. trichocarpa* and *O. sativa* (Q6K7P8), TAIR (arabidopsis.org) for *A. thaliana* or from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu, Loc_Os02g58170.1).
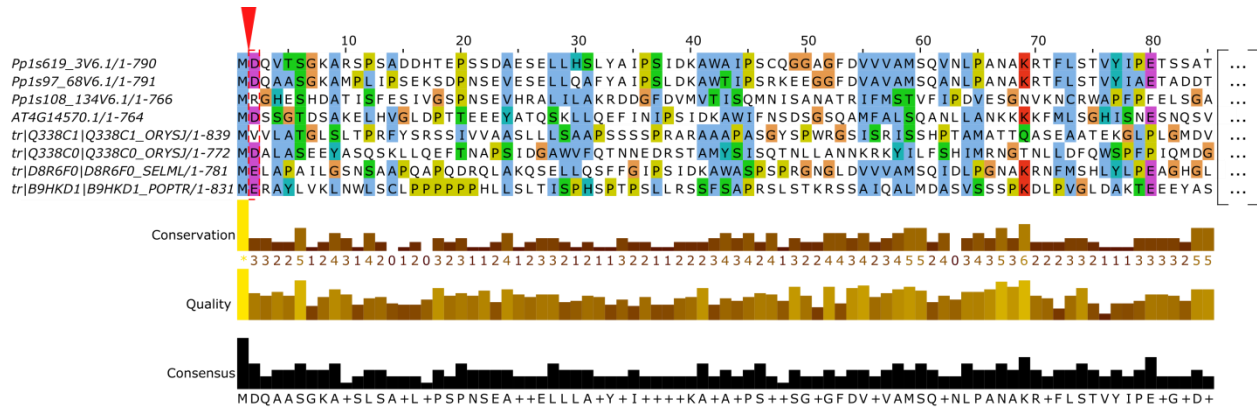
Fig. 18. Section from a multiple sequence alignment of homologous proteins of PpAARE (Pp1s619_3V6.1). Homologous proteins of *O. sativa* (Q338C1, Q338C0), *A. thaliana* (AT4G14750.1)*, S. moellendorffii* (D8R6F0) and *P. trichocarpa* (B9HKD1) were chosen according to the previously established protein family definitions (4) as described in (3). Protein sequences were aligned at the initiator methionine of the corresponding protein model with Jalview version 2.9.0b2 (5). The coloring is according to the *ClustalX* color scheme. The red arrow indicates the observed cleavage site in PpAARE (Pp1s619_3V6.1). The red dashed box encompasses the identified arginylated residue D^2 of PpAARE and the aligned residues from the homologous proteins. Protein sequences were downloaded from Uniprot (uniprot.org) in the case of *S. moellendorffii, P. trichocarpa* and *O. sativa* or TAIR (arabidopsis.org) for *A. thaliana*.

Supplemental table S1

Overview on the results of the MaxQuant analysis of the CoIP.

Supplemental table S2

20

Overview on the distribution of spectral counts for proteins identified in the IP for arginylated proteins.

References

1. Wong, C. C. L., Xu, T., Rai, R., Bailey, A. O., Yates, J. R., Wolf, Y. I., Zebroski, H., and Kashina, A. (2007) Global analysis of posttranslational protein arginylation. *PLoS Biol.* 5, e258

2. Tian, G.-W., Mohanty, A., Chary, S. N., Li, S., Paap, B., Drakakaki, G., Kopec, C. D., Li, J., Ehrhardt, D., Jackson, D., Rhee, S. Y., Raikhel, N. V., and Citovsky, V. (2004) High-throughput fluorescent tagging of full-length *Arabidopsis* gene products in planta. *Plant Physiol.* 135, 25–38

3. Schuessele, C., Hoernstein, S. N. W., Mueller, S. J., Rodriguez-Franco, M., Lorenz, T., Lang, D., Igloi, G. L., and Reski, R. (2016) Spatio-temporal patterning of arginyl-tRNA protein transferase (ATE) contributes to gametophytic development in a moss. *New Phytol.* 209, 1014–1027

4. Zimmer, A. D., Lang, D., Buchta, K., Rombauts, S., Nishiyama, T., Hasebe, M., Van de Peer, Y., Rensing, S. A., and Reski, R. (2013) Reannotation and extended community resources for the genome of the non-seed plant *Physcomitrella patens* provide insights into the evolution of plant gene structures and functions. *BMC Genomics* 14, 498

5. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191