

LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell

Jana Sperschneider^{1,*}, Ann-Maree Catanzariti², Kathleen DeBoer³, Benjamin Petre⁴, Donald M. Gardiner⁵, Karam B Singh¹, Peter N. Dodds⁶, Jennifer M. Taylor⁶

¹Centre for Environment and Life Sciences, CSIRO Agriculture and Food, Perth, WA, Australia

²Division of Plant Sciences, Research School of Biology, The Australian National University, Acton, ACT, Australia

³The UWA Institute of Agriculture, The University of Western Australia, Crawley, WA, Australia

⁴The Sainsbury Laboratory, Norwich Research Park, Norwich, United Kingdom

⁵Queensland Bioscience Precinct, CSIRO Agriculture and Food, Brisbane, QLD, Australia

⁶Black Mountain Laboratories, CSIRO Agriculture and Food, Canberra, ACT, Australia

* corresponding author: Jana Sperschneider, phone: +61 8 9333 6205, email: jana.sperschneider@csiro.au

Abstract

Pathogens are able to deliver effector proteins into plant cells to enable infection. Some effectors have been found to enter subcellular compartments by mimicking host targeting sequences. Although many computational methods exist to predict plant protein subcellular localization, they perform poorly for effectors. We introduce LOCALIZER for predicting plant and effector protein localization to chloroplasts, mitochondria, and nuclei. LOCALIZER shows greater prediction accuracy for chloroplast and mitochondrial targeting compared to other methods for 652 plant proteins. For 108 eukaryotic effectors, LOCALIZER outperforms other methods and predicts a previously unrecognized chloroplast transit peptide for the ToxA effector, which we show translocates into tobacco chloroplasts. Secretome-wide predictions and confocal microscopy reveal that rust fungi might have evolved multiple effectors that target chloroplasts or nuclei. LOCALIZER is the first method for predicting effector localisation in plants and is a valuable tool for prioritizing effector candidates for functional investigations. LOCALIZER is available at <http://localizer.csiro.au/>.

Keywords: effector, subcellular localization, plant cell, chloroplast, plant pathogen, ToxA

Introduction

Plant cells feature subcellular compartments such as mitochondria or chloroplasts which contain distinct suites of proteins related to their specialised biological functions. Plant proteins are translocated from the cytosol into specific organelles by means of N-terminal transit peptides in the case of chloroplasts and mitochondria^{1,2}, or nuclear localization signals (NLSs) in the case of nuclei. Transit peptides rarely share sequence conservation and vary in length^{1,3}.

Eukaryotic filamentous plant pathogens deliver cytoplasmic effectors into host tissues to subvert plant functions to their advantage⁴. Whilst there are many bacterial effectors targeting specific compartments, the extent to which eukaryotic effectors enter plant organelles is less understood⁵⁻¹¹. Recently, several rust effector candidates were shown to mimic transit peptides to translocate into chloroplasts¹². Determining effector subcellular localization in plant cells provides important clues about their virulence function. However, experimental methods are labour-intensive, prone to artefacts, and not appropriate for high-throughput screening of the large effector repertoires predicted in fungi and oomycetes. No method for effector subcellular localization prediction is available thus far to prioritize effector candidates for experimental investigations.

In principle, a plant-trained classifier should work on both host and pathogen, as effectors might exploit the plant machinery to enter organelles. However, signal peptides, pro-domains and the rapid evolution of effectors pose challenges (Fig. 1). Firstly, if effectors carry transit peptides these can be separated from N-terminal signal peptides by a pro-domain of varying length¹³. This poses a challenge to plant predictors such as TargetP¹⁴ or ChloroP¹⁵, which analyse the N-terminal amino acid composition and assume that potential transit peptides start at the first residue after the signal peptide. For example, the ToxA effector localizes to chloroplasts, interacts with the chloroplast-localized protein ToxABP1¹⁶ and contains a signal peptide followed by a pro-domain that are cleaved during secretion¹⁷. Thus, if ToxA carries a transit peptide it is likely to occur after the pro-domain. The second challenge lies in the lack of sequence homology between effector proteins and non-effector proteins. Methods such as WoLF PSORT¹⁸ include homology-based information from proteins with experimentally validated subcellular localizations. However, effectors rarely share sequence similarity with other proteins¹⁹.

To address these challenges, we introduce a machine learning prediction tool called LOCALIZER, which can be run in two modes:

1) Plant mode: LOCALIZER predicts chloroplast/mitochondrial transit peptides and/or NLSs in plant proteins.

2) Effector mode: LOCALIZER uses a sliding window approach to predict chloroplast/mitochondrial transit peptides and/or NLSs in effectors.

In both modes, LOCALIZER can predict if a plant or effector protein can localize to multiple compartments. This is important because dual targeting of plant proteins is common²⁰⁻²², but only WoLF PSORT and YLoc+ are capable of predicting dual localization²³.

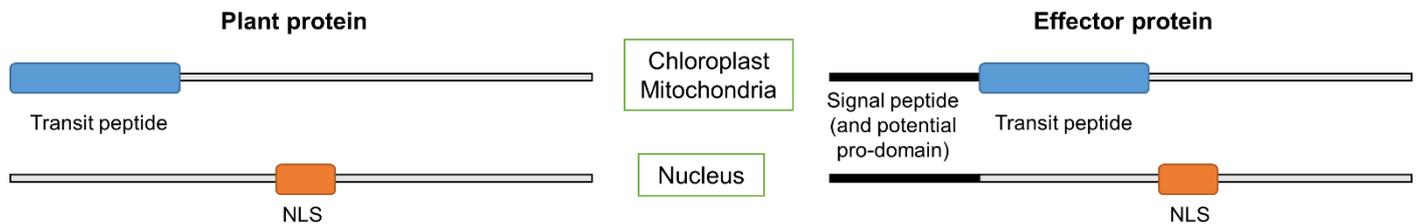


Fig. 1: Plant proteins can carry transit peptides at their N-terminus that guide them to chloroplast or mitochondria, or nucleus localization signals (NLSs) that guide them to the nucleus. Effector proteins can also target plant chloroplast, mitochondria, or nuclei through mimicry of transit peptides or NLSs occurring after their signal peptides. Some effectors might also have pro-domains after the signal peptide that are cleaved off during or after secretion.

Results

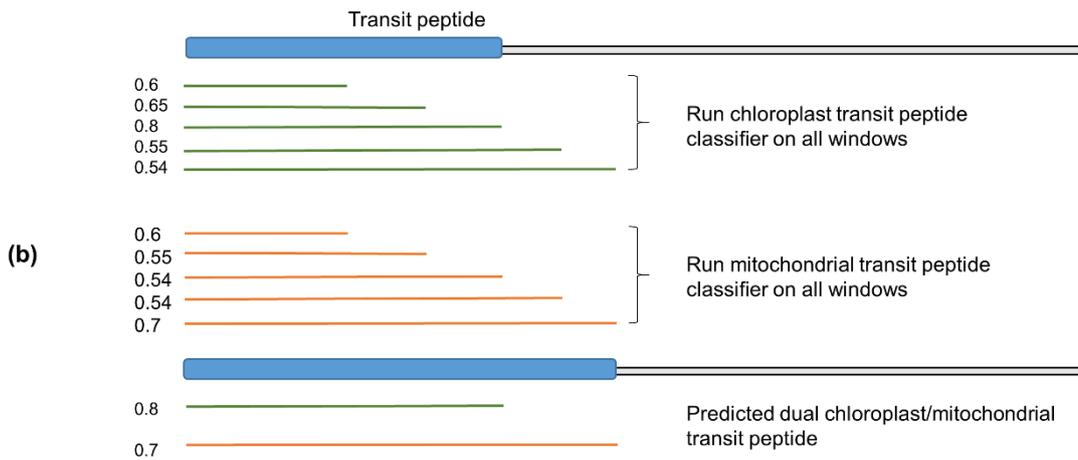
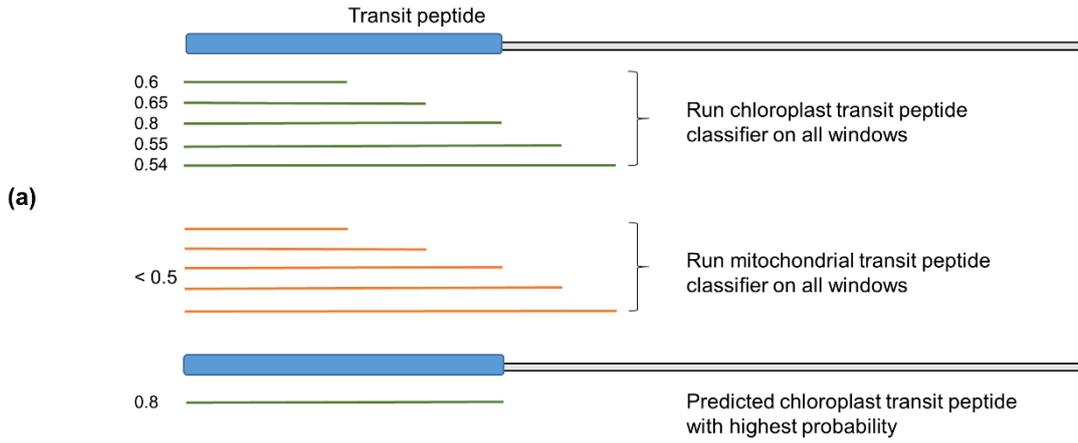
LOCALIZER: a method for predicting plant and effector protein targeting to chloroplasts, mitochondria or nuclei

To predict if a plant or an effector protein localizes to chloroplasts or mitochondria, we applied a machine learning approach trained on plant proteins with experimentally verified localization data. LOCALIZER uses two classifiers trained to recognize chloroplast or mitochondrial transit peptides and a NLS search for predicting nucleus localization (Supplementary Table S1).

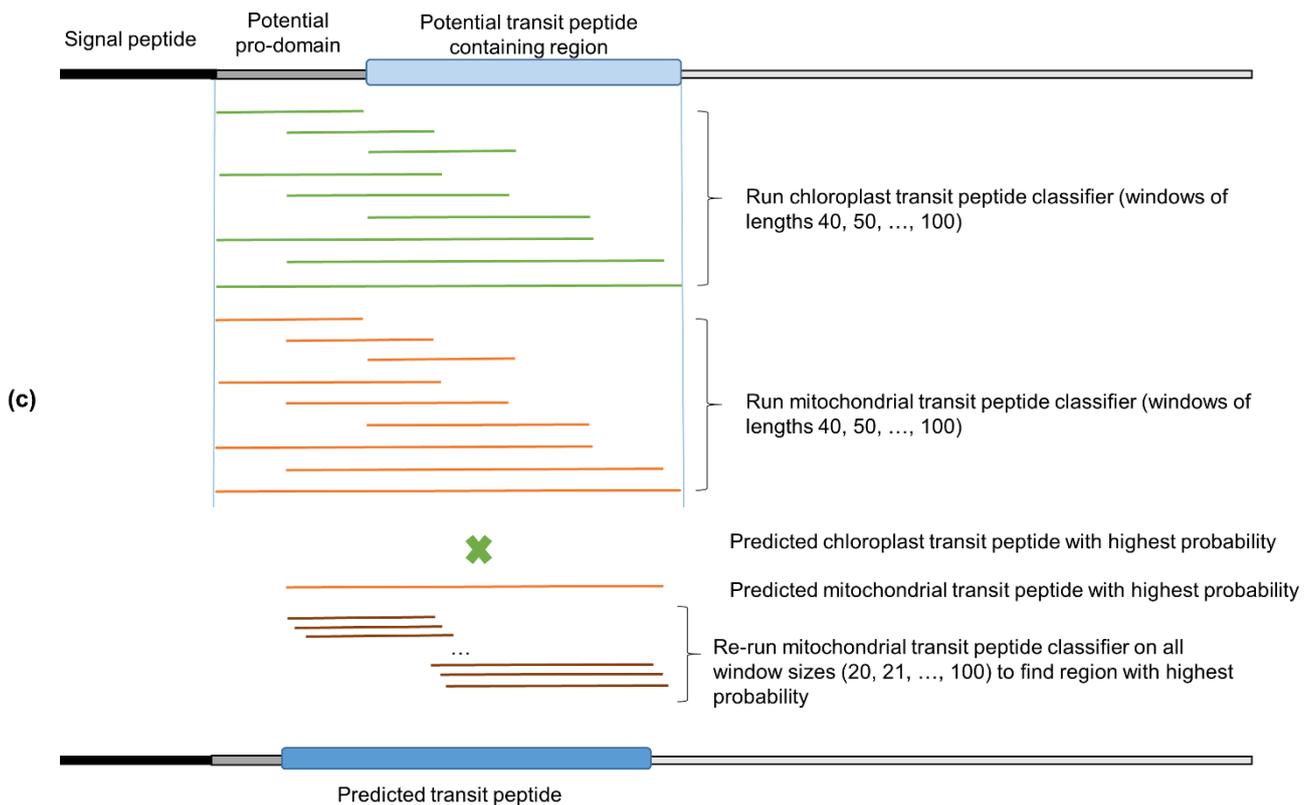
LOCALIZER has been trained on transit peptides and therefore, a window of sequence harbouring a potential transit peptide should be presented to the classifier. To search for transit peptides in plant proteins, sequence windows of varying lengths starting at the first position in the sequence were used. For each of the windows, the chloroplast and mitochondrial classifiers were called. If only one of the classifiers returns predicted transit peptides, the transit peptide with the highest probability is returned as the result (Fig. 2a). If both classifiers return predicted transit peptides, the one with the highest probability is returned as the main localization, e.g. chloroplast, and the weaker classification from the other classifier is reported as a possible dual-localization, e.g. chloroplast and possible mitochondrial (Fig. 2b).

Unlike in plant proteins, where the transit peptides starts at the N-terminus, the transit peptide start in effectors is more variable due to signal peptide lengths and the potential presence of pro-domains. To address this, LOCALIZER uses a sliding window approach to scan for the start of potential transit peptides (Fig. 2c). In the following, we first benchmark LOCALIZER on plant proteins and then demonstrate that LOCALIZER also accurately predicts effector localization in the plant cell.

Plant protein



Effector protein



87 **Fig. 2: (a)** An example of a predicted chloroplast-targeting plant protein with probability 0.8. Only the chloroplast
88 classifier returns predicted transit peptides, whereas the mitochondrial classifier does not return positive predictions
89 (probabilities < 0.5). **(b)** An example of a predicted dual-targeting plant protein with a predicted chloroplast transit
90 peptide (probability 0.8) and a possible mitochondrial transit peptide (probability 0.7). **(c)** An example of the sliding
91 window approach for effector proteins. Sliding windows of varying lengths are moved along the mature effector
92 sequences to find potential transit peptide start positions. Both chloroplast and mitochondrial transit peptides are
93 predicted for each window using the two classifier. In this example, the effector has a predicted mitochondrial
94 transit peptide, but no predicted chloroplast transit peptide (indicated by a cross).

96 LOCALIZER has the highest accuracy for predicting chloroplast and mitochondrial localization in plant proteins

97 Due to the relatively small number of plant proteins with experimentally validated transit peptides in the UniProt
98 database, we could not use parts of our training set as independent test sets. Therefore, we took the manually
99 curated cropPAL database which features the subcellular localization of rice, wheat, maize and barley proteins²⁴. We
100 used only those proteins that are supported by GFP studies and localize to one subcellular compartment, resulting in
101 a test set of 530 sequences. The overlap with our UniProt training set is minimal, out of 100 cropPAL proteins that
102 are targeted to plastids only four are part of our chloroplast training set and out of 61 cropPAL proteins that are
103 targeted to mitochondria only four are part of our mitochondria training set. The test set is of high quality due to the
104 manual curation and is also very valuable because it exclusively features crop species, which are host species for the
105 most intensively studied plant pathogens. As a second test set, we used plant proteins with experimentally validated
106 localization that were entered into UniProt after the training of LOCALIZER. This independent test set contains 122
107 plant proteins.

108 On the test set of 652 plant proteins, LOCALIZER achieves the highest MCC of 0.71 and highest accuracy of 91.4% for
109 chloroplast proteins and the highest MCC of 0.54 and highest accuracy of 91.7% for mitochondrial proteins
110 compared to other programs (Table 1). However, the homology- and annotation-based methods YLoc and WoLF
111 PSORT achieve higher accuracy of 79.8% and BaCelLo achieves higher MCC of 0.42 on the nuclear localized protein
112 set than LOCALIZER (accuracy 73%, MCC 0.4). The protein property based predictors BaCelLo, NLStradamus and
113 PredictNLS²⁵ have lower accuracies than LOCALIZER. PredictNLS in particular correctly identifies only a small
114 proportion (8.9%) as nuclear-localized, suggesting that its default collection of NLSs is incomplete for plants. Note
115 that for LOCALIZER's evaluation we only considered proteins predicted to localize to the nucleus exclusively in this
116 benchmark. When counting all proteins that are predicted to carry a NLS as nuclear-localized regardless of an
117 additional transit peptide prediction, LOCALIZER's sensitivity increases, however specificity also decreases. Taken
118 together, this evaluation might suggest that homology-based methods are advantageous for nuclear localization
119 prediction in plants, as a NLS motif search in plant proteomes has a higher false positive rate. Homology-based
120 methods could be combined with a NLS motif search used by LOCALIZER for increased sensitivity in nuclear
121 localization prediction for plant proteins.

122 Of the 120 chloroplast-localized proteins, our classifier identifies 87 as carrying a chloroplast transit peptide. Of
123 these 87 proteins, 8 are predicted to also carry a potential mitochondrial transit peptide, 33 are predicted to carry an
124 additional NLS and five are predicted to carry a potential mitochondrial transit peptide as well as an NLS. We used
125 the cropPAL database to build a set of plant proteins that have been shown to localize to two or more compartments
126 based on GFP studies. Out of 34 proteins that have been experimentally shown to localize to both chloroplasts and
127 mitochondria, LOCALIZER correctly predicts dual localization for 13 (38.2%). In contrast, YLoc+ predicts only 6 as
128 dual-targeted to chloroplasts and mitochondria, whereas WoLF PSORT does not assign any of the 34 proteins to the
129 dual compartment chloroplast-mitochondria. Other examples from the literature are the dual-localized maize CFM6
130 protein²⁶, for which LOCALIZER correctly predicts dual localization to nucleus and mitochondria and the WHIRLY1
131 protein²⁷, for which LOCALIZER correctly predicts dual localization to nucleus and chloroplasts. Overall, the
132 independent validation on the test set shows that LOCALIZER outperforms other methods for predicting chloroplast
133 and mitochondrial transit peptides in plant proteins as well as dual-targeting of plant proteins, whereas YLoc and
134 WoLF PSORT are the most accurate tools for plant nuclear localized proteins.

137
138

Table 1: Performance evaluation on the combined test set of cropPAL crop plant proteins and UniProt plant proteins. MCC stands for Matthews Correlation Coefficient and PPV for positive predictive value.

	Sensitivity	Specificity	PPV	MCC	Accuracy
Chloroplast (120 proteins)					
LOCALIZER	72.5%	95.7%	79.1%	0.71	91.4%
TargetP	75%	92.9%	70.3%	0.66	89.6%
ChloroP	72.5%	86.7%	55.1%	0.53	84%
WoLF PSORT	80.8%	73.1%	40.4%	0.43	74.5%
Predotar	67.5%	95.9%	78.6%	0.67	90.6%
YLoc	77.5%	81.2%	48.2%	0.5	80.5%
BaCelLo	75%	80%	46.6%	0.47	79.1%
Mitochondria (65 proteins)					
LOCALIZER	60%	95.2%	58.2%	0.54	91.7%
TargetP	64.6%	89.1%	39.6%	0.44	86.7%
WoLF PSORT	21.5%	95.7%	35.9%	0.22	88.3%
Predotar	60%	94.4%	54.2%	0.52	91%
YLoc	46.2%	94%	46.2%	0.4	89.3%
BaCelLo	6.2%	99.7%	66.7%	0.18	90.3%
Nucleus (225 proteins)					
LOCALIZER	60%	79.9%	61.1%	0.4	73%
WoLF PSORT	55.1%	92.7%	80%	0.53	79.8%
YLoc	61.8%	89.2%	75.1%	0.54	79.8%
BaCelLo	82.2%	62.1%	53.3%	0.42	69%
NLStradamus	32.9%	86.9%	56.9%	0.24	68.3%
PredictNLS	8.9%	98.6%	76.9%	0.18	67.6%

139

140

LOCALIZER in effector mode: high accuracy on fungal and oomycete effector localization

141

142

143

144

145

146

147

148

149

150

151

152

153

154

We used data sets from the literature to compare the performance of LOCALIZER to other methods on 108 GFP-tagged fungal and oomycete effectors/effector candidates (Supplementary Table S2, results given in Supplementary File 2, bacterial effector prediction described in Supplementary File 1). Out of the 108 effectors, seven effectors have been shown to localize to chloroplasts and 51 effectors have been shown to localize to the plant nucleus. To the best of our knowledge, no localization to mitochondria exclusively has been shown for eukaryotic effectors from plant pathogens. We tested LOCALIZER in effector mode and compared it to the performance of plant subcellular localization methods applied to mature effector sequences. Overall, LOCALIZER had the best performance when predicting chloroplast- or nucleus-localized effectors, whereas plant subcellular localization methods showed low sensitivity, MCCs close to zero and high false positive rates (Table 2) and are therefore not suitable for predicting effector localization. Notably, the chloroplast-targeting effector set demonstrates that plant subcellular localization prediction methods should not be used for predicting the chloroplast-targeting of mature effector sequences due to poor performance. LOCALIZER predicts chloroplast transit peptides for 5 out of 7 effectors (Table 3), whereas other methods rarely predict chloroplast localization in these effectors and feature higher false positive rates of 5% to 21.8% (Table 2).

155

156

157

158

159

160

161

162

163

164

165

We then tested LOCALIZER on the set of 51 effectors that have been experimentally shown to localize to the plant nucleus. On this set, LOCALIZER had the highest MCC of 0.42 and highest accuracy of 70.4% (Table 2). Despite their superior performance on plant proteins, BaCelLo, WoLF PSORT and YLoc have lower MCCs and accuracies for nucleus-localized effectors. In particular, the majority of *Hyaloperonospora arabidopsis* RxLR (HaRxLR) effector candidates in Caillaud, et al. ⁶ were experimentally found to localize to the nucleus, whereas none have been found to enter chloroplasts and mitochondria. LOCALIZER predicts NLSs in 7 out of 13 HaRxLR effectors (53.8%) that exclusively localize to the nucleus, but only in 3 out of 15 HaRxLR effectors (20%) that have been shown to localize to the nucleus and cytoplasm, which might indicate that their localization to nuclei is inconclusive. For the set of 45 HaRxLR effector candidates, LOCALIZER returns seven false positive predictions of chloroplast or mitochondrial targeting (15.6%). Interestingly, when providing LOCALIZER with HaRxLR mature effector sequences that are cleaved after the leucine in the RxLR motif, this changes to two false positives (4.4%). Our negative set also includes 33

166 candidate rust effectors that show no specific localization to mitochondria, chloroplast or nuclei in experiments
 167 ^{12,28,29}. For these 33 effector candidates, LOCALIZER only predicts three as nuclear-localized and one as chloroplast-
 168 localized, whereas ChloroP predicts five as chloroplast-localized and TargetP returns 11 false positives. Overall, these
 169 results show that plant subcellular localization methods are not suitable for predicting effector localization and that
 170 LOCALIZER greatly improves prediction accuracy for effectors targeting plant chloroplasts and nuclei.

171 **Table 2:** Evaluation of effector localization using LOCALIZER and other plant-trained subcellular localization methods.

Chloroplast effectors		Sensitivity	Specificity	PPV	MCC	Accuracy	False positive rate
	LOCALIZER	85.7%	96%	60%	0.69	95.4%	4%
	TargetP	0%	83.2%	0%	-0.11	77.8%	16.8%
	ChloroP	14.3%	92.1%	11.1%	0.06	87%	7.9%
	WoLF PSORT	42.9%	78.2%	12%	0.12	75.9%	21.8%
	BaCelLo	0%	88.1%	0%	-0.09	82.4%	11.9%
	YLoc	14.3%	92.1%	11.1	0.06	87%	7.9%
	Predotar	0%	95%	0%	-0.06	88.9%	5%
Nucleus effectors							
	LOCALIZER	51%	87.7%	78.8%	0.42	70.4%	12.3%
	WoLF PSORT	47.1%	75%	63.2%	0.22	61.1%	25%
	BaCelLo	52.9%	62.5%	56.3%	0.15	57.4%	37.5%
	YLoc	58.8%	75%	68.2%	0.33	66.7%	25%
	NLStradamus	23.7%	94.7%	82.4%	0.3	63%	5.3%
	PredictNLS	11.8%	100%	100%	0.15	54.6%	0%

172

173 **Table 3:** Test set of effectors and effector candidates that have been experimentally shown to localize to the plant
 174 chloroplast and the predictions by several classifiers. All methods except LOCALIZER were supplied with the mature
 175 effector sequences.

Effector	CTP1	CTP2	CTP3	MLP124111	PST03196	PST18220	ToxA
Reference	Petre <i>et al.</i> (2015b), Petre <i>et al.</i> (2015a)				Petre <i>et al.</i> (2016)		Manning and Ciuffetti (2005)
Species	<i>Melampsora larici-populina</i>		<i>M. lini</i>	<i>M larici-populina</i>	<i>Puccinia striiformis</i> f. sp. <i>tritici</i>		<i>Pyrenophora tritici-repentis</i>
Comment	Also in mitochondria	-	-	Also in cytosolic aggregates		Also in nucleus	Also in cytoplasm
Subcellular localization prediction							
LOCALIZER	Chloroplast	Chloroplast	Chloroplast	-	Chloroplast, Mitochondria, Nucleus	-	Chloroplast
ChloroP	Chloroplast	-	-	-	-	-	-
TargetP	Mitochondria	Mitochondria	-	Mitochondria	-	-	-
WoLF PSORT	Chloroplast	Chloroplast	Chloroplast	Mitochondria	Cytosol	Nucleus	Cytoskeleton
YLOC	Secreted	Secreted	Secreted	Secreted	Chloroplast	Nucleus	Cytoplasm
Predotar	-	-	-	-	-	-	-
BaCelLo	Secreted	Secreted	Secreted	Secreted	Nucleus	Nucleus	Nucleus

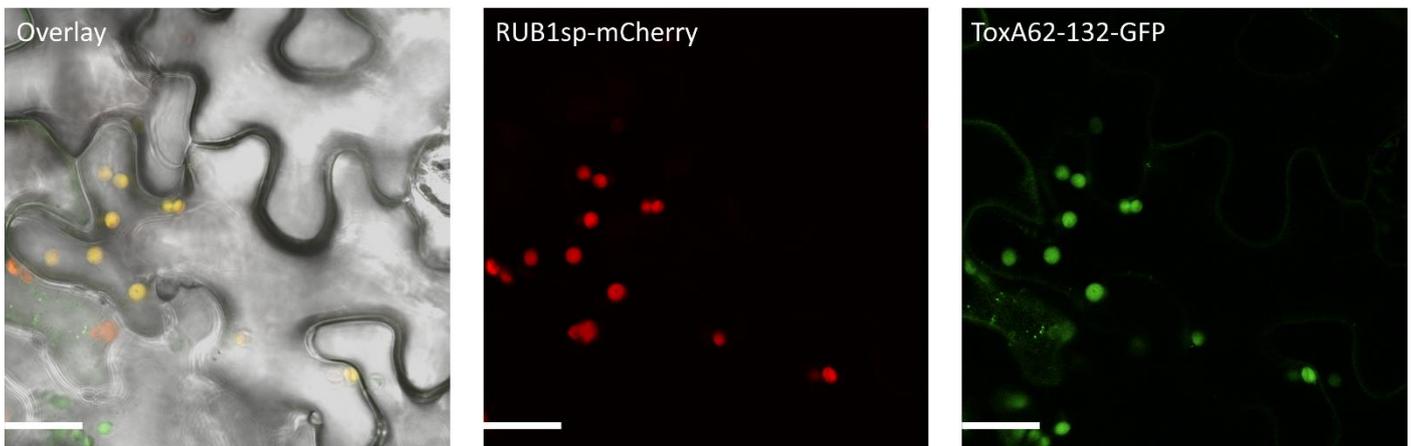
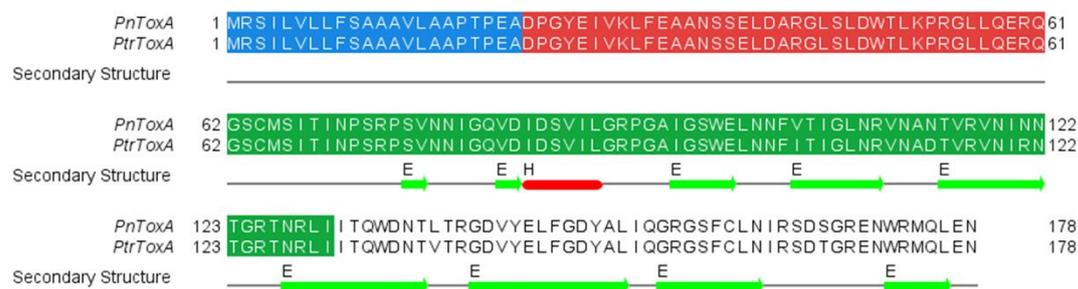
176

177 **LOCALIZER predicts a chloroplast transit peptide for ToxA and suggests that many RxLR effectors target nuclei**

178 We collected a set of experimentally validated 69 fungal effectors and 51 oomycete effectors and ran LOCALIZER to
 179 predict their subcellular localization. For the 69 fungal effectors, only three are predicted to carry a chloroplast
 180 transit peptide, i.e. Six6 (*Fusarium oxysporum* f. sp. *lycopersici*), Ave1 (*Verticillium dahliae*) and ToxA (*Pyrenophora*
 181 *tritici-repentis* and *Parastagonospora nodorum*). Both Six6 and Ave1 are likely false positives, Six6-GFP has been
 182 shown to localize to the cytoplasm and cell nucleus in tobacco ³⁰ and Ave1 is recognized on the cell surface.
 183 Interestingly, Ave1 has been suggested to be horizontally transferred from plants ³¹. Only the Avra10 effector from

184 *Blumeria graminis* f.sp. *hordei* is predicted to target mitochondria and also has a predicted NLS. Nine effectors are
 185 predicted to carry a NLS (RTP1, Avr4, Avr2, Six1, Bas107, Avr-Pik, UhAvr1, Pit2, SP7; 13% of effectors). Of these,
 186 RTP1, SP7 and Bas107 have thus far been shown to localize to the plant nucleus³²⁻³⁴. We could not predict nuclear
 187 localization for the fungal effectors MISSP7³⁵, See1³⁶ and Six3³⁷, suggesting that these might not rely on NLSs to
 188 enter plant nuclei.

189 ToxA has been implicated to localize to chloroplasts, but the presence of a transit peptide has thus far not been
 190 confirmed using existing prediction methods^{16,38}. However, LOCALIZER predicts a chloroplast transit peptide at
 191 position 62 to 130 with probability 0.885 for the *Parastagonospora nodorum* ToxA. Mapping the chloroplast transit
 192 peptide predicted by LOCALIZER onto the ToxA structure reveals that it is predicted to start immediately after the
 193 pro-domain (Fig. 3). The pro-domain of ToxA has been suggested to finish at amino acid 60 and to be important for
 194 folding, but not necessary for toxic activity³⁹. We found that ToxA₆₂₋₁₃₂-GFP accumulates in *N. benthamiana*
 195 chloroplasts (Fig. 3). A shorter version of the predicted transit peptide (ToxA₆₂₋₉₃-GFP), where amino acid position 93
 196 is the start of the first beta sheet in the three-dimensional structure⁴⁰ did not localize to tobacco chloroplasts (Fig.
 197 3). Furthermore, the full-length ToxA protein (with and without the pro-domain) did not localize to tobacco
 198 chloroplasts (data not shown). Taken together, this suggests that the N-terminal region of ToxA (ToxA₆₂₋₁₃₂) has the
 199 ability to enter chloroplasts, however it likely needs to be unstructured in order to do so, which might be achieved
 200 through the pro-domain or post-translational modifications.



201

202 **Fig. 3:** The sequence that was tested for chloroplast localization for PnToxA (*Parastagonospora nodorum*) is shown in
 203 green in the alignment. The signal peptide and pro-domain are shown in blue and red, respectively. The secondary
 204 structure from PtrToxA (PDB entry 1zle) is also shown. Confocal images demonstrating the chloroplast localisation of
 205 ToxA₆₂₋₁₃₂-GFP when transiently expressed in the pavement cells of *N. benthamiana* leaves. The left panel displays
 206 the transmitted light image overlaid with the ToxA₆₂₋₁₃₂-GFP and RUB1sp-mCherry fluorescence. The middle panel
 207 displays the RUB1sp-mCherry fluorescence (chloroplast marker) whilst the right panel displays the ToxA₆₂₋₁₃₂-GFP
 208 fluorescence. Scale bar = 20 μm

209 For predictions on the oomycete effector set we used RxLR mature effector sequences cleaved after the leucine in
 210 the RxLR motif. Only the *Phytophthora sojae* PsIscl1 effector is predicted as carrying a chloroplast transit peptide,
 211 which may be consistent with its function in suppressing salicylate-mediated immunity by degrading isochlorismate
 212 (produced in the chloroplast)⁴¹. The RxLR effector Avh241 is predicted to target mitochondria but has been shown to
 213 localize to plasma membranes⁴². A large proportion of oomycete effectors are predicted to carry NLSs (ATR13,
 214 Avrblb2, CRN1, CRN2, CRN8, CRN15, CRN16, CRN63, CRN115, Avh18a1, PiAvr2, PiAvrVnt1, PsAvr3b, PsAvr4/6; 27.5%
 215 of effectors). Several Crinkler (CRN) effector proteins have been shown to localize to the plant nucleus and to require

216 nuclear accumulation to induce plant cell death¹⁰. LOCALIZER predicts all seven Crinkler effectors as nuclear-
 217 localized, whereas YLoc predicts three, WoLF PSORT and BaCelLo predict two and NLStradamus and PredictNLS
 218 predict only one as nuclear-localized. On a set of 358 RxLR effector candidates determined using the Hidden Markov
 219 model⁴³ and cleaved after the leucine in the RxLR motif, LOCALIZER predicts 0.9% predicted as chloroplast-targeting,
 220 6.1% predicted as mitochondrial-targeting and 22.3% predicted as nucleus-targeting. This suggests that oomycete
 221 effectors may not target chloroplasts, but given that a large proportion carry predicted NLSs, may preferably target
 222 the plant nucleus.

223 Localization predictions of effector candidates reveals expanded chloroplast- and nuclear-targeting in rust 224 pathogens

225 To further investigate the extent to which fungal effectors are targeted towards specific plant compartments, we
 226 first predicted secretomes from 61 fungal species including pathogens and saprophytes using a sensitive approach
 227 tailored to effector finding described in Sperschneider, et al.⁴⁴. Importantly, the presence of a predicted signal
 228 peptide does not necessarily guarantee the full secretion of a protein from the pathogen. Firstly, prediction tools can
 229 return false positives. On eukaryotic data, SignalP 3 (which is used here for secretome prediction) has been
 230 estimated to have sensitivity of 98.8% with a false positive rate of 0.8% to 11.7% depending on the presence of
 231 transmembrane domains (<http://www.cbs.dtu.dk/services/SignalP/performance.php>). Secondly, proteins with a
 232 signal peptide can have retention signals such as the consensus sequences KDEL/HDEL that can keep them in the ER
 233 or Golgi² or proteins with a signal peptide can be anchored to the pathogen cell wall. Therefore, we also ran
 234 EffectorP⁴⁵ on the secretomes to limit the subcellular localization predictions to likely effector candidates.

235 Saprophytes are unlikely to produce effectors targeted to plant subcellular compartments, and indeed only very low
 236 numbers of predicted effectors were predicted to show chloroplast (1.24%), mitochondrion (0.73%) or nuclear
 237 (1.78%) localisation from these fungi (Table 4, Supplementary Figs. S1-S3). The proportions of predicted chloroplast,
 238 mitochondria and nuclear targeted proteins were slightly higher amongst effector candidates predicted from
 239 pathogen/symbiont secretomes, with a total of 2.18%, 1.03% and 2.62% respectively. Interestingly, the haustoria-
 240 forming pathogens (rusts and powdery mildews) showed even higher proportions of predicted nuclear-targeted
 241 effectors (~4.5%). The highest proportions (> 5%) of effector candidates that were also predicted to target nuclei
 242 were found in *Puccinia graminis* f. sp. *tritici* (5.61%), *P. triticina* (5.51%), *Magnaporthe oryzae* (5.46%) and *Blumeria*
 243 *graminis* f. sp. *hordei* (5.1%). Likewise, the rust pathogens had a higher proportion of predicted chloroplast-targeting
 244 effectors (3.35%), while this was not observed for *Blumeria* species (1.03%), which infect only cereal epidermal cells
 245 deficient in chloroplast. The apoplastic pathogen *Cladosporium fulvum* showed only low levels of targeted effector
 246 candidates similar to the observations for saprophytic fungi and the animal pathogen *Batrachochytrium*
 247 *dendrobatidis* also has a low percentage of predicted chloroplast-targeting effectors of 1.33% (Table 4).

248 **Table 4:** Percentages of secreted proteins that are EffectorP effector candidates and are also predicted to target
 249 chloroplast, mitochondria or nuclei by LOCALIZER across groups of fungal genomes.

	Number of genomes	Average % of secreted proteins that are EffectorP effector candidates and are also predicted to target		
		Chloroplast	Mitochondria	Nucleus
Plant pathogens/symbionts	36	2.18%	1.03%	2.62%
Saprophytes	19	1.24%	0.73%	1.78%
Rust pathogens	5	3.35%	0.93%	4.5%
Blumeria pathogens	2	1.03%	1.74%	4.37%
<i>Cladosporium fulvum</i>	1	1.19%	0.37%	2.2%
<i>Magnaporthe oryzae</i>	1	2.58%	0.8%	5.46%
<i>Batrachochytrium dendrobatidis</i>	1	1.33%	0.48%	3.03%

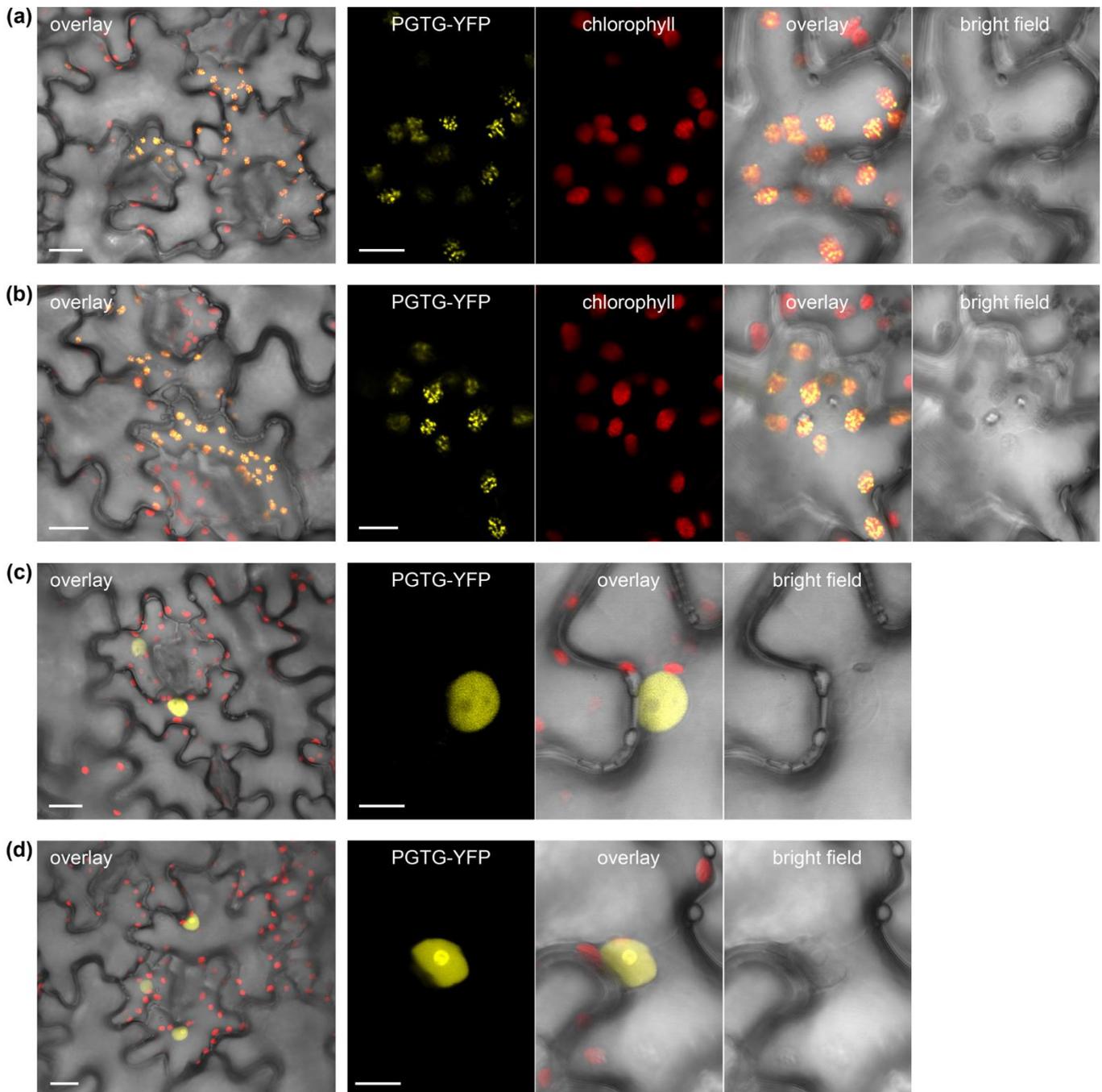
250

251 Subcellular localization prediction of effector candidates relies strongly on accurate signal peptide prediction. Some
 252 fungal proteins might be retained in the fungal cell despite a predicted signal peptide and localize to fungal
 253 mitochondria or nuclei. The use of sensitive signal peptide prediction methods is especially advantageous for

254 effector finding ⁴⁴, however it might lead to a higher number of false positives. For example, we found that the
255 *Saccharomyces cerevisiae* NAD-dependent protein deacetylase HST1 is predicted as secreted and LOCALIZER predicts
256 it to have an NLS, but is annotated in UniProt as non-secreted and nuclear. Another example is the *S. cerevisiae*
257 fumarate reductase 2 protein which has been shown to localize to mitochondria ⁴⁶, but is predicted to be secreted by
258 SignalP 3.0. In effector mode, LOCALIZER predicts chloroplast localization for the mature protein with probability
259 0.98 whereas in plant mode, LOCALIZER correctly identifies a mitochondrial transit peptide at the start of the protein
260 with probability 0.992. Therefore, the combination of effector candidate prediction with subcellular localization
261 prediction might be advantageous to arrive at a confident set of likely effectors that can target plant cell
262 compartments.

263 We selected four secreted proteins from the wheat stem rust fungus *P. graminis* f. sp. *tritici* and tested them for
264 their subcellular localization in tobacco. The four candidates were selected based on their differential expression in
265 haustoria compared to spores ⁴⁷, their sequence homology to rust pathogens only and their predicted subcellular
266 localization to chloroplast, mitochondria or nucleus based on LOCALIZER and other tools. We ran SignalP 4.1 to
267 determine the signal peptide cleavage site and ran localization predictions on the mature sequences (Supplementary
268 Table S3). The localization experiments showed that the mature sequences of PGTG_00164 and PGTG_06076
269 accumulate in chloroplasts, whereas PGTG_13278 and PGTG_15899 localize to nuclei (Fig. 4).

270 PGTG_00164 is predicted as chloroplast-targeted by YLoc, Predotar and BaCelLo, whereas LOCALIZER detects only a
271 bipartite NLS in the sequence and also in two of its homologs (Supplementary Fig. S5). PGTG_06076 is a larger
272 effector candidate protein of 336 aas and is correctly predicted to target chloroplasts by all tested methods including
273 LOCALIZER. LOCALIZER also predicts chloroplast targeting for the homologs PST130_11623 and PTTG_29922, but not
274 for PGTG_07697 (Supplementary Fig. S6). PGTG_13278 has sequence homology to a large number of proteins in
275 other rust pathogens and is correctly predicted to localize to nuclei by all predictions tools that can predict nucleus
276 localization (Supplementary Table S4). PGTG_15899 localizes to the nucleus and, to a lesser extent, the cytoplasm,
277 but is predicted to target chloroplasts and/or mitochondria by ChloroP, TargetP, WoLF PSORT, YLoc and BaCelLo.
278 LOCALIZER predicts a mitochondrial transit peptide in the mature PGTG_15899 sequence, but also identifies a
279 bipartite NLS (Supplementary Fig. S7).



280

281

282

283

284

285

286

287

288

289

290

291

Discussion

292

293

294

Plants face attack by pathogens that secrete effector proteins into plant cells to subvert plant defences or manipulate physiology by targeting host proteins. These effectors would be expected to co-localize in subcellular compartments with their respective host targets. For example, plant chloroplasts have emerged as a target of

295 microbial effector proteins^{48,49} and bacterial effectors have been found to directly target mitochondria⁵⁰. The plant
296 nucleus is another prime target for pathogen effectors both from bacteria and eukaryotes^{10,11,51}. Furthermore,
297 chloroplast-nucleus communication via stromules has been suggested to play a role in the plant immune response⁵².
298 Accurate predictions of subcellular localization for both plant and pathogen proteins is therefore essential for
299 understanding pathogen-targeted host components and compartments. Many computational subcellular localization
300 prediction tools have been developed for plant proteins, however no dedicated methods are available for predicting
301 effector localization in the plant cell and we found that plant subcellular localization methods on mature effector
302 sequences give unreliable results. Thus, we developed LOCALIZER as the first computational prediction method
303 capable of accurately predicting both the localization of pathogen effector proteins as well as plant proteins to
304 chloroplasts, mitochondria or nuclei.

305 For nuclear localization prediction, LOCALIZER uses a simple search for eukaryotic NLSs. On the set of plant proteins,
306 LOCALIZER is outperformed by YLoc and WoLF PSORT that include homology-based information due to the low
307 specificity of the NLS search. However, on secreted effector proteins a NLS search employed by LOCALIZER shows
308 the highest accuracy. For effector proteins in particular, using homology to known nuclear-localized proteins can be
309 a misleading strategy as gene families might have members that have evolved different functions or subcellular
310 localization. For example, the rust effector RTP1 from *Uromyces fabae* (Uf-RTP1) has been shown to localize to the
311 plant nucleus, whereas its homolog from *U. striatus* (Us-RTP1) has been found only in the host cytoplasm and is
312 barely visible in nuclei³². LOCALIZER predicts nuclear localization for Uf-RTP1, but not for Us-RTP1, whereas YLoc
313 predicts nuclear localization for Us-RTP1. Furthermore, effector proteins rarely share significant similarity to know
314 proteins and thus, homology-based methods underperform on the set of nuclear-localized effectors whereas
315 LOCALIZER delivers the highest accuracy.

316 On sets of experimentally validated fungal and oomycete effectors, we found that the mimicry of chloroplast and
317 mitochondrial transit peptides seems rare. However, the ability to target chloroplasts can be significant for effector
318 function, as exemplified by the ToxA effector and its interaction with a chloroplast-localized plant protein as well as
319 light-dependency^{16,38}. LOCALIZER is the first method which predicts a previously undetected chloroplast transit
320 peptide in the ToxA effector, and we demonstrate that the predicted transit peptide indeed has the ability to enter
321 tobacco chloroplasts. Many experimentally validated effectors, especially from oomycetes, were predicted to target
322 host nuclei. When scanning predicted fungal secretomes for subcellular localization, we found that rust pathogens,
323 *Blumeria* pathogens and *Magnaporthe oryzae* seem to be enriched for predicted nucleus-targeting effector
324 candidates relative to other fungi. Rust pathogens also showed an enrichment for predicted chloroplast-targeting
325 effector candidates relative to other fungi. Using confocal microscopy and transient expression, we confirmed
326 chloroplast-targeting and nucleus-targeting for four rust effector candidates.

327 We have demonstrated that LOCALIZER will facilitate both functional plant protein and effector studies and improve
328 our understanding of plant-pathogen interactions. Future developments of LOCALIZER will include the capability to
329 predict the localization to other compartments in the plant cell such as the apoplast, peroxisome, cytoplasm and
330 plasma membrane.

331 **Methods**

332 **Training set selection and support vector machine training**

333 The UniProt data base was accessed and plant proteins (taxonomy:"Viridiplantae [33090]") were retrieved that
334 localize to the following compartments supported by experimental evidence: chloroplast ("Plastid [SL-0209]",
335 "Chloroplast [SL-0209]"); mitochondria ("Mitochondrion [SL-0173]"); nucleus ("Nucleus [SL-0191]"); cytoplasm
336 ("Cytoplasm [SL-0086]"); membranes ("membrane"); secreted ("Secreted [SL-0243]"). Sequences that were shorter
337 than or equal to 50 amino acids (aas) and those that did not start with an 'M' were removed. From the set of 1,600
338 chloroplast localized proteins, 1,279 annotated transit peptides were extracted and saved as a FASTA file. The
339 average length of transit peptides was 53 (min:15 max:113). After homology reduction by excluding transit peptides
340 that share sequence similarity with another one at E-value < 0.00001 using phmmer⁵³, 639 transit peptides
341 remained in the training set. From the set of 326 mitochondrial localized proteins, 201 annotated transit peptides
342 were extracted and saved as a FASTA file. The average length of transit peptides was 41 (min:12 max:115). After
343 homology reduction, 194 transit peptides remained in the training set.

344 We also retrieved nuclear-localized eukaryotic proteins from UniProt (taxonomy:"Eukaryota [2759]", "Nucleus [SL-
345 0191]", evidence:experimental). The annotated NLSs were extracted and saved as a list of motifs. These were
346 supplemented with a list of regular expressions that describe known NLSs given in Kosugi, et al.⁵⁴ and the bipartite
347 NLS defined as follows: two adjacent basic amino acids (R/K) followed by a spacer region of 8-12 residues followed
348 by at least three basic amino acids (R/K) in the five positions after the spacer region. Additionally, we used
349 NLStradamus, a NLS predictor using Hidden Markov Models which has been trained on yeast sequences⁵⁵. A protein
350 is labelled as containing a NLS if at least one of the two methods are positive: 1) NLS sequence motif search, or 2)
351 NLS prediction by NLStradamus. For each protein, all predicted NLSs are returned and if one motif is contained in
352 another one, the longer NLS is returned as the result. All training sets and the NLS list are available at
353 <http://localizer.csiro.au/data.html>.

354 The chloroplast classifier was trained to distinguish chloroplast transit peptides from N-termini of non-chloroplast
355 proteins. The set of 639 chloroplast transit peptides was used as the positive set and the set of homology-reduced
356 non-chloroplast plant proteins was used as the negative set (1,597 proteins). For each protein from the negative set,
357 a random choice of the first x amino acids was used, with x ranging from 40 to 120. The mitochondria classifier was
358 trained to distinguish mitochondrial transit peptides from N-termini of non-mitochondrial proteins. The set of 194
359 mitochondrial transit peptides was used as the positive set and the set of homology-reduced non-mitochondrial
360 plant proteins was used as the negative set (1,878 proteins). As the negative set is much larger than the positive set,
361 a randomly chosen subset of 626 proteins was used as the negative set. For each protein from the negative set, a
362 random choice of the first x amino acids was used, with x ranging from 40 to 120.

363 The feature vector for each protein consists of 58 features given in Supplementary Table S1. Both classifiers were
364 trained using the SMO support vector machine classifier with the RBF Kernel and the option -M for fitting logistic
365 models using the Weka software (weka.classifiers.functions.SMO)⁵⁶. To optimize the complexity parameter C and
366 gamma parameter G of the support vector machine, the Weka grid search was used to find the best parameters. For
367 the chloroplast classifier, the best parameters found were $C = 2.0$ and $G = 1.0$ and for the mitochondrial classifier,
368 the best parameters found were $C = 3.0$ and $G = 0.0625$.

369 Test sets and performance evaluation

370 We downloaded the set of crop plant proteins (barley, wheat, rice, maize) from the cropPal database²⁴ and chose
371 those that have a subcellular localization of either 'plastid' (100 proteins), 'mitochondrion' (61 proteins), 'nucleus'
372 (165 proteins), 'peroxisome' (11 proteins), 'vacuole' (18 proteins), 'plasma membrane' (84 proteins), 'endoplasmic
373 reticulum' (43 proteins) and 'cytosol' (48 proteins) determined by GFP-tagging. We only kept those sequences that
374 started with an 'M'. For the UniProt test set, we downloaded plant proteins (taxonomy:"Viridiplantae [33090]") that
375 were entered after our training sets were compiled (created:[20160301 TO 20160902]) for several compartments
376 supported by experimental evidence ("Nucleus [SL-0191]"; "Mitochondrion [SL-0173]", "Chloroplast [SL-0049]",
377 "Peroxisome [SL-0204]", "Vacuole", "Secreted", "Endoplasmic reticulum", "Cytoplasm"). We manually removed
378 those entries that localize to multiple compartments, except for the category nucleus for which we also allowed an
379 additional cytoplasmic localization. All plant and effector test sets are available at <http://localizer.csiro.au/data.html>.

380 When evaluating performance, the number of true positives (TPs), true negatives (TNs), false positives (FPs) and
381 false negatives (FNs) were used. Sensitivity ($\frac{TP}{TP+FN}$) is defined as the proportion of positives that are correctly
382 identified whereas specificity ($\frac{TN}{TN+FP}$) is the proportion of negatives that are correctly identified. Precision (positive
383 predictive value, PPV, $\frac{TP}{TP+FP}$) is a measure which captures the proportion of positive predictions that are true. Both
384 accuracy ($\frac{TP+TN}{TP+FP+FN+TN}$) and the Matthews Correlation Coefficient (MCC, $\frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FN)(TP+FP)(TN+FP)(TN+FN)}}$) can be
385 used to evaluate the overall performance of a method. The MCC ranges from -1 to 1, with scores of -1 corresponding
386 to predictions in total disagreement with the observations, 0.5 to random predictions and 1 to predictions in perfect
387 agreement with the observations. For our classifier, we count LOCALIZER predictions that are 'chloroplast',
388 'chloroplast and possible mitochondrial', 'chloroplast and nucleus' and 'chloroplast & possible mitochondrial and
389 nucleus' as chloroplast predictions (same strategy for mitochondrial predictions). A protein that carries a predicted
390 transit peptide with an additional predicted NLS might have experimental evidence only for one of those locations
391 due to the technical hurdles of recognizing dual targeting²⁰ and should thus not necessarily be counted as a false
392 positive prediction. A protein is counted as a nucleus prediction only if it has the category 'nucleus' to avoid
393 assigning a protein to multiple predictions in the evaluation. Many plant subcellular localization methods have been

published, however only a small number are available as standalone software or have the option of submitting large batch sequence files to a web server. This makes it prohibitive for researchers to use them routinely for data analysis and thus, our benchmark only includes methods that can be locally installed with ease or have a web server with a batch file submission option (Supplementary Table S2).

Detection of transit peptides in plant proteins and effectors

In plant mode, windows of varying lengths (40, 50, 60, 70, 80, 90, 100) starting from the first position in the sequence are used to search for chloroplast and mitochondrial transit peptides in plant proteins. In effector mode, sliding windows of varying lengths (40, 50, 60, 70, 80, 90, 100) are moved with step size of one amino acid along the mature effector sequences to find potential transit peptide start positions under the following criteria: 1) to allow for potential pro-domains, the transit peptide can have a start position ranging from residue 1 to 50 in the mature sequence, and 2) at least 40 aas must remain in the C-terminal to allow for the effector domain after the potential transit peptide region. Unless mature effector sequences are provided by the user, LOCALIZER removes the signal peptide by deleting the first 20 aas. In the evaluation, we use the default method of deleting the first 20 aas in the effector sequences unless otherwise specified.

For each of the windows, the chloroplast and mitochondrial classifiers are called and the support vector machine (SVM) probability for a positive classification is returned. To reduce false positive predictions, proteins that have five or less positive predictions in all sliding windows are discarded as weak predictions in effector mode. In both plant and effector mode, a transit peptide prediction is only made if the probability is larger than 0.6. If only one of the classifiers returns predicted transit peptides, the transit peptide with the highest probability is returned as the result. If both classifiers return predicted transit peptides, the one with the highest probability is returned as the main localization and the weaker classification from the other classifier is reported as a possible dual-localization. Up to this point, only certain windows of sizes (40, 50, 60, 70, 80, 90, 100) were considered due to computational runtime and thus only approximate locations of transit peptides are returned. Let x be the length of the predicted transit peptide window with highest probability. To identify the location of the predicted transit peptide more accurately, a second round of sliding window predictions (windows of varying lengths from size 20, 21, 22, ..., x ; $x < 100$; step size of one amino acid) are run on the predicted transit peptide window with the highest probability. If one of the refined windows return a transit peptide with a higher probability, this is reported in the final prediction.

If a protein contains more than 10% unknown bases (B, Z, X) in its sequence, it is not used for prediction of localization. If it contains less than 10% unknown bases, these are randomly replaced with the respective amino acid (B replaced with D or N; Z replaced with E or Q, X replaced with any amino acid). In plant and effector mode, only proteins longer than 40 aas will be used for prediction of localization.

Prediction of fungal pathogen secretomes

The set of secreted proteins was predicted in various fungal genomes (Supplementary File 3) using a pipeline described in Sperschneider, et al.⁴⁴, which assigns a protein as secreted if it is predicted to be secreted by the neural network predictors of SignalP 3 and by TargetP and if it has no predicted transmembrane domain outside the first 60 aas using TMHMM and no predicted transmembrane domain using Phobius^{14,57-59}. EffectorP 1.0 was run with default parameters to predict effector candidates from the fungal secretomes⁴⁵.

Transient *in planta* expression and confocal microscopy of ToxA transit peptide

The ToxA transit peptide was codon optimised for expression in *Nicotiana benthamiana* using the Integrated DNA Technologies (IDT) codon optimisation tool (<https://sg.idtdna.com/CodonOpt>). The codon optimised ToxA transit peptide was fused to the N-terminus of GFP synthetically as a gBlocks gene fragment (IDT), and was subsequently cloned into the pAGM4723 binary vector, behind the CaMV 35S promoter, using the Golden Gate system⁶⁰. For chloroplast co-localisation, we used the pCMU-PLAr plasmid containing the *AtUBQ10* promoter: RUB1sp-mCherry expression cassette, as described in Ivanov and Harrison⁶¹. Localisation experiments were undertaken essentially as described in Sparkes, et al.⁶². The constructs were transformed into *Agrobacterium tumefaciens* (AGL1) using electroporation, and grown at 28 degrees for 24-48 hours prior to being resuspended in the infiltration buffer (50 mM MES, 0.5% (w/v) glucose, 100 μ M acetosyringone; pH 5.6) to a final OD₆₀₀ of 0.1. The *Agrobacterium* was infiltrated into the upper leaves of 3-4 week old *N. benthamiana* plants which had been grown in a constant temperature room at 22°C with a 16-h light/8-h dark cycle. After 3-4 days the infiltrated leaves were examined using a Nikon A1Si confocal microscope (Nikon Plan Apo VC 60x NA1.2 water-immersion objective). The 488nm laser was

444 used for excitation of GFP with the 521/50 nm band pass filter for fluorescence detection. The 561 nm laser was
445 used for excitation of mCherry with the 595/50 nm band pass filter used to detect emitted fluorescence. Images
446 were processed using the Nikon NIS elements software package.

447 **Transient *in planta* expression and confocal microscopy of rust effector candidates**

448 Gene coding sequences without the predicted signal peptide sequence (SignalP 4.1) and stop codon were
449 synthesised as single gBlock gene fragments (Integrated DNA Technologies, Coralville, Iowa, USA). Each gBlock
450 included CACCATG at the 5' end for cloning into the pENTR/D-TOPO Gateway entry vector (Invitrogen, Carlsbad, CA,
451 USA) and to provide a start codon to initiate translation. The DNA was cloned directly into the entry vector, then
452 transferred into pB7YWG2 containing the cauliflower mosaic virus (CaMV) 35S promoter, C-terminal EYFP gene and
453 35S terminator⁶³ using LR Clonase II (Invitrogen). The final constructs were verified by DNA sequencing and then
454 transferred by electroporation into the *Agrobacterium* strain GV3101 (pMP90). For *in planta* expression,
455 *Agrobacterium* cultures were prepared to an optical density at 600 nm of 1.0 in 10 mM MES (pH 5.6) buffer with 10
456 mM MgCl₂ and 200 μM acetosyringone, and then infiltrated into *N. tabacum* leaves. Infiltrated plants were kept in a
457 25°C growth room with a 16-hour day length. Tobacco leaf sections were imaged 2 days after agroinfiltration on a
458 Zeiss confocal LSM 780 microscope using a 40× water immersion objective (LD C-Apochromat 40×/1.1W Korr M27)
459 with excitation at 514 nm. EYFP fluorescence was detected at 520–600 nm and autofluorescence of the chloroplasts
460 was detected at 605–720 nm or 650–720 nm. Zen 2012 digital imaging software (Zeiss, Oberkochen, Germany) was
461 used during image acquisition.

462 **Immunoblot analysis**

463 Two days after agroinfiltration, tobacco leaf tissue was frozen in liquid nitrogen and ground in 2× Laemmli buffer
464 with 0.2 M dithiothreitol. The protein samples were boiled for 5 min, centrifuged to pellet leaf debris, then
465 separated by sodium dodecylsulphate–polyacrylamide gel electrophoresis and transferred by electroblotting to a
466 nitrocellulose membrane. The membrane was blocked in 5% skim milk then probed with anti-GFP (7.1 and 13.1;
467 Roche, Penzberg, Germany), followed by sheep anti-mouse IgG antibody conjugated to horseradish peroxidase.
468 Labelling was detected using the Amersham ECL Western Blotting Detection Reagent and the ImageQuant LAS 4000
469 CCD camera system (GE Healthcare Life Sciences, Little Chalfont, UK).

470 **Acknowledgements**

471 We thank Sophien Kamoun and his group for useful discussions. We thank Brendan Kidd and Laura Davies for their
472 valuable feedback. JS is supported by an OCE Postdoctoral Fellowship.

473 **Author Contribution**

474 J.S., P.D., K.S. and J.T. planned and designed the research. J.S. developed and implemented the machine learning
475 method. A-M.C., K.D, B.P and D.G performed localization experiments. All authors analysed and interpreted the data
476 and results and wrote the manuscript.

477 **Competing financial interests**

478 The authors declare no competing financial interests.

480 **References**

- 481 **1** Jarvis, P. Targeting of nucleus-encoded proteins to chloroplasts in plants. *The New phytologist* **179**, 257-
482 **285** (2008).
- 483 **2** Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP,
484 SignalP and related tools. *Nature protocols* **2**, 953-971, doi:10.1038/nprot.2007.131 (2007).
- 485 **3** Holbrook, K. *et al.* Functional Analysis of Semi-conserved Transit Peptide Motifs and Mechanistic
486 Implications in Precursor Targeting and Recognition. *Molecular plant* **9**, 1286-1301,
487 doi:10.1016/j.molp.2016.06.004 (2016).
- 488 **4** Kamoun, S. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annual review of*
489 *phytopathology* **44**, 41-60 (2006).

- 490 5 Hicks, S. W. & Galan, J. E. Exploitation of eukaryotic subcellular targeting mechanisms by bacterial
491 effectors. *Nature reviews. Microbiology* 11, 316-326, doi:10.1038/nrmicro3009 (2013).
- 492 6 Caillaud, M. C. *et al.* Subcellular localization of the Hpa RxLR effector repertoire identifies a tonoplast-
493 associated protein HaRxL17 that confers enhanced plant susceptibility. *The Plant journal : for cell and*
494 *molecular biology* 69, 252-265, doi:10.1111/j.1365-313X.2011.04787.x (2012).
- 495 7 McLellan, H. *et al.* An RxLR effector from *Phytophthora infestans* prevents re-localisation of two plant NAC
496 transcription factors from the endoplasmic reticulum to the nucleus. *PLoS pathogens* 9, e1003670,
497 doi:10.1371/journal.ppat.1003670 (2013).
- 498 8 Bozkurt, T. O. *et al.* *Phytophthora infestans* effector AVRblb2 prevents secretion of a plant immune
499 protease at the haustorial interface. *Proceedings of the National Academy of Sciences of the United States*
500 *of America* 108, 20832-20837, doi:10.1073/pnas.1112708109 (2011).
- 501 9 Bozkurt, T. O., Schornack, S., Banfield, M. J. & Kamoun, S. Oomycetes, effectors, and all that jazz. *Current*
502 *opinion in plant biology* 15, 483-492, doi:10.1016/j.pbi.2012.03.008 (2012).
- 503 10 Schornack, S. *et al.* Ancient class of translocated oomycete effectors targets the host nucleus. *Proceedings*
504 *of the National Academy of Sciences of the United States of America* 107, 17421-17426,
505 doi:10.1073/pnas.1008491107 (2010).
- 506 11 Rivas, S. & Genin, S. A plethora of virulence strategies hidden behind nuclear targeting of microbial
507 effectors. *Front Plant Sci* 2, 104, doi:10.3389/fpls.2011.00104 (2011).
- 508 12 Petre, B. *et al.* Rust fungal effectors mimic host transit peptides to translocate into chloroplasts. *Cellular*
509 *microbiology*, doi:10.1111/cmi.12530 (2015).
- 510 13 Rep, M. Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS microbiology*
511 *letters* 253, 19-27, doi:10.1016/j.femsle.2005.09.014 (2005).
- 512 14 Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins
513 based on their N-terminal amino acid sequence. *Journal of molecular biology* 300, 1005-1016,
514 doi:10.1006/jmbi.2000.3903 (2000).
- 515 15 Emanuelsson, O., Nielsen, H. & von Heijne, G. ChloroP, a neural network-based method for predicting
516 chloroplast transit peptides and their cleavage sites. *Protein science : a publication of the Protein Society*
517 8, 978-984, doi:10.1110/ps.8.5.978 (1999).
- 518 16 Manning, V. A., Hardison, L. K. & Ciuffetti, L. M. Ptr ToxA interacts with a chloroplast-localized protein.
519 *Molecular plant-microbe interactions : MPMI* 20, 168-177, doi:10.1094/MPMI-20-2-0168 (2007).
- 520 17 Ciuffetti, L. M., Manning, V. A., Pandelova, I., Betts, M. F. & Martinez, J. P. Host-selective toxins, Ptr ToxA
521 and Ptr ToxB, as necrotrophic effectors in the *Pyrenophora tritici-repentis*-wheat interaction. *The New*
522 *phytologist* 187, 911-919, doi:10.1111/j.1469-8137.2010.03362.x (2010).
- 523 18 Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic acids research* 35, W585-587,
524 doi:10.1093/nar/gkm259 (2007).
- 525 19 Sperschneider, J. *et al.* Advances and challenges in computational prediction of effectors from plant
526 pathogenic fungi. *PLoS pathogens* 11, e1004806, doi:10.1371/journal.ppat.1004806 (2015).
- 527 20 Carrie, C. & Whelan, J. Widespread dual targeting of proteins in land plants: when, where, how and why.
528 *Plant signaling & behavior* 8, doi:10.4161/psb.25034 (2013).
- 529 21 Kisslov, I., Naamati, A., Shakarchy, N. & Pines, O. Dual-targeted proteins tend to be more evolutionarily
530 conserved. *Molecular biology and evolution* 31, 2770-2779, doi:10.1093/molbev/msu221 (2014).
- 531 22 Carrie, C. & Small, I. A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts.
532 *Biochimica et biophysica acta* 1833, 253-259, doi:10.1016/j.bbamcr.2012.05.029 (2013).
- 533 23 Briesemeister, S., Rahnenfuhrer, J. & Kohlbacher, O. YLoc--an interpretable web server for predicting
534 subcellular localization. *Nucleic acids research* 38, W497-502, doi:10.1093/nar/gkq477 (2010).
- 535 24 Hooper, C. M., Castleden, I. R., Aryamanesh, N., Jacoby, R. P. & Millar, A. H. Finding the Subcellular
536 Location of Barley, Wheat, Rice and Maize Proteins: The Compendium of Crop Proteins with Annotated
537 Locations (cropPAL). *Plant & cell physiology* 57, e9, doi:10.1093/pcp/pcv170 (2016).
- 538 25 Cokol, M., Nair, R. & Rost, B. Finding nuclear localization signals. *EMBO reports* 1, 411-415,
539 doi:10.1093/embo-reports/kvd092 (2000).
- 540 26 Barkan, A. *et al.* The CRM domain: an RNA binding module derived from an ancient ribosome-associated
541 protein. *RNA* 13, 55-64, doi:10.1261/rna.139607 (2007).
- 542 27 Isemer, R. *et al.* Recombinant Whirly1 translocates from transplastomic chloroplasts to the nucleus. *FEBS*
543 *letters* 586, 85-88, doi:10.1016/j.febslet.2011.11.029 (2012).

- 544 28 Petre, B. *et al.* Heterologous Expression Screens in *Nicotiana benthamiana* Identify a Candidate Effector of
545 the Wheat Yellow Rust Pathogen that Associates with Processing Bodies. *PLoS one* 11, e0149035,
546 doi:10.1371/journal.pone.0149035 (2016).
- 547 29 Petre, B. *et al.* Candidate Effector Proteins of the Rust Pathogen *Melampsora larici-populina* Target
548 Diverse Plant Cell Compartments. *Molecular plant-microbe interactions : MPMI* 28, 689-700,
549 doi:10.1094/MPMI-01-15-0003-R (2015).
- 550 30 Gawehns, F. *et al.* The *Fusarium oxysporum* effector Six6 contributes to virulence and suppresses I-2-
551 mediated cell death. *Molecular plant-microbe interactions : MPMI* 27, 336-348, doi:10.1094/MPMI-11-13-
552 0330-R (2014).
- 553 31 de Jonge, R. *et al.* Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens
554 uncovered by genome and RNA sequencing. *Proceedings of the National Academy of Sciences of the*
555 *United States of America* 109, 5110-5115, doi:10.1073/pnas.1119623109 (2012).
- 556 32 Kemen, E. *et al.* Identification of a protein from rust fungi transferred from haustoria into infected plant
557 cells. *Molecular plant-microbe interactions : MPMI* 18, 1130-1139, doi:10.1094/MPMI-18-1130 (2005).
- 558 33 Giraldo, M. C. & Valent, B. Filamentous plant pathogen effectors in action. *Nature reviews. Microbiology*
559 11, 800-814, doi:10.1038/nrmicro3119 (2013).
- 560 34 Klopplolz, S., Kuhn, H. & Requena, N. A secreted fungal effector of *Glomus intraradices* promotes
561 symbiotic biotrophy. *Current biology : CB* 21, 1204-1209, doi:10.1016/j.cub.2011.06.044 (2011).
- 562 35 Plett, J. M. *et al.* A secreted effector protein of *Laccaria bicolor* is required for symbiosis development.
563 *Current biology : CB* 21, 1197-1203, doi:10.1016/j.cub.2011.05.033 (2011).
- 564 36 Redkar, A., Villajuana-Bonequi, M. & Doehlemann, G. Conservation of the *Ustilago maydis* effector See1 in
565 related smuts. *Plant signaling & behavior* 10, e1086855, doi:10.1080/15592324.2015.1086855 (2015).
- 566 37 Ma, L., Cornelissen, B. J. & Takken, F. L. A nuclear localization for Avr2 from *Fusarium oxysporum* is
567 required to activate the tomato resistance protein I-2. *Front Plant Sci* 4, 94, doi:10.3389/fpls.2013.00094
568 (2013).
- 569 38 Manning, V. A. & Ciuffetti, L. M. Localization of Ptr ToxA Produced by *Pyrenophora tritici-repentis* Reveals
570 Protein Import into Wheat Mesophyll Cells. *The Plant cell* 17, 3203-3212, doi:10.1105/tpc.105.035063
571 (2005).
- 572 39 Tuori, R. P., Wolpert, T. J. & Ciuffetti, L. M. Heterologous expression of functional Ptr ToxA. *Molecular*
573 *plant-microbe interactions : MPMI* 13, 456-464, doi:10.1094/MPMI.2000.13.4.456 (2000).
- 574 40 Sarma, G. N., Manning, V. A., Ciuffetti, L. M. & Karplus, P. A. Structure of Ptr ToxA: an RGD-containing
575 host-selective toxin from *Pyrenophora tritici-repentis*. *The Plant cell* 17, 3190-3202,
576 doi:10.1105/tpc.105.034918 (2005).
- 577 41 Liu, T. *et al.* Unconventionally secreted effectors of two filamentous pathogens target plant salicylate
578 biosynthesis. *Nature communications* 5, 4686, doi:10.1038/ncomms5686 (2014).
- 579 42 Yu, X. *et al.* The RxLR effector Avh241 from *Phytophthora sojae* requires plasma membrane localization to
580 induce plant cell death. *The New phytologist* 196, 247-260, doi:10.1111/j.1469-8137.2012.04241.x (2012).
- 581 43 Win, J. *et al.* Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant
582 pathogenic oomycetes. *The Plant cell* 19, 2349-2369, doi:10.1105/tpc.107.051037 (2007).
- 583 44 Sperschneider, J., Williams, A. H., Hane, J. K., Singh, K. B. & Taylor, J. M. Evaluation of Secretion Prediction
584 Highlights Differing Approaches Needed for Oomycete and Fungal Effectors. *Front Plant Sci* 6, 1168,
585 doi:10.3389/fpls.2015.01168 (2015).
- 586 45 Sperschneider, J. *et al.* EffectorP: predicting fungal effector proteins from secretomes using machine
587 learning. *The New phytologist*, doi:10.1111/nph.13794 (2015).
- 588 46 Muratsubaki, H. & Enomoto, K. One of the fumarate reductase isoenzymes from *Saccharomyces cerevisiae*
589 is encoded by the OSM1 gene. *Archives of biochemistry and biophysics* 352, 175-181,
590 doi:10.1006/abbi.1998.0583 (1998).
- 591 47 Upadhyaya, N. M. *et al.* Comparative genomics of Australian isolates of the wheat stem rust pathogen
592 *Puccinia graminis* f. sp. *tritici* reveals extensive polymorphism in candidate effector genes. *Front Plant Sci*
593 5, 759, doi:10.3389/fpls.2014.00759 (2014).
- 594 48 Serrano, I., Audran, C. & Rivas, S. Chloroplasts at work during plant innate immunity. *Journal of*
595 *experimental botany* 67, 3845-3854, doi:10.1093/jxb/erw088 (2016).
- 596 49 de Torres Zabala, M. *et al.* Chloroplasts play a central role in plant defence and are targeted by pathogen
597 effectors. *Nature plants* 1, 15074, doi:10.1038/nplants.2015.74 (2015).

- 598 **50** Block, A. *et al.* The *Pseudomonas syringae* type III effector HopG1 targets mitochondria, alters plant
599 development and suppresses plant innate immunity. *Cellular microbiology* 12, 318-330,
600 doi:10.1111/j.1462-5822.2009.01396.x (2010).
- 601 **51** Canonne, J. & Rivas, S. Bacterial effectors target the plant cell nucleus to subvert host transcription. *Plant*
602 *signaling & behavior* 7, 217-221, doi:10.4161/psb.18885 (2012).
- 603 **52** Caplan, J. L. *et al.* Chloroplast Stromules Function during Innate Immunity. *Developmental cell* 34, 45-57,
604 doi:10.1016/j.devcel.2015.05.011 (2015).
- 605 **53** Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching.
606 *Nucleic acids research* 39, W29-37, doi:10.1093/nar/gkr367 (2011).
- 607 **54** Kosugi, S. *et al.* Six classes of nuclear localization signals specific to different binding grooves of importin
608 alpha. *The Journal of biological chemistry* 284, 478-485, doi:10.1074/jbc.M807017200 (2009).
- 609 **55** Nguyen Ba, A. N., Pogoutse, A., Provard, N. & Moses, A. M. NLStradamus: a simple Hidden Markov Model
610 for nuclear localization signal prediction. *BMC bioinformatics* 10, 202, doi:10.1186/1471-2105-10-202
611 (2009).
- 612 **56** Hall, M. *et al.* The WEKA data mining software: an update. *SIGKDD Explor. Newsl.* 11, 10-18,
613 doi:10.1145/1656274.1656278 (2009).
- 614 **57** Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0.
615 *Journal of molecular biology* 340, 783-795, doi:10.1016/j.jmb.2004.05.028 (2004).
- 616 **58** Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology
617 with a hidden Markov model: application to complete genomes. *Journal of molecular biology* 305, 567-
618 580, doi:10.1006/jmbi.2000.4315 (2001).
- 619 **59** Kall, L., Krogh, A. & Sonnhammer, E. L. A combined transmembrane topology and signal peptide prediction
620 method. *Journal of molecular biology* 338, 1027-1036, doi:10.1016/j.jmb.2004.03.016 (2004).
- 621 **60** Engler, C. *et al.* A golden gate modular cloning toolbox for plants. *ACS synthetic biology* 3, 839-843,
622 doi:10.1021/sb4001504 (2014).
- 623 **61** Ivanov, S. & Harrison, M. J. A set of fluorescent protein-based markers expressed from constitutive and
624 arbuscular mycorrhiza-inducible promoters to label organelles, membranes and cytoskeletal elements in
625 *Medicago truncatula*. *The Plant journal : for cell and molecular biology* 80, 1151-1163,
626 doi:10.1111/tpj.12706 (2014).
- 627 **62** Sparkes, I. A., Runions, J., Kearns, A. & Hawes, C. Rapid, transient expression of fluorescent fusion proteins
628 in tobacco plants and generation of stably transformed plants. *Nature protocols* 1, 2019-2025,
629 doi:10.1038/nprot.2006.286 (2006).
- 630 **63** Karimi, M., De Meyer, B. & Hilson, P. Modular cloning in plant cells. *Trends in plant science* 10, 103-105,
631 doi:10.1016/j.tplants.2005.01.008 (2005).

632

633