CSIRO PUBLISHING
www.publish.csiro.au/journals/fpb *Functional Plant Biology*, 2010, **37**, 513–523

Isolation, phylogeny and evolutionary analysis of *Pto*-type disease resistance gene analogues from a *Cucumis hystrix* introgression line of cucumber (*C. sativus*)

Hongjian Wan[A,B], *Chuntao Qian*[A], *Ahmed Abbas Malik*[A], *Zhenguo Zhao*[A] *and Jinfeng Chen*[A,C]

[A]State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, People's Republic of China.
[B]Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, People's Republic of China.
[C]Corresponding author. Email: jfchen@njau.edu.cn

Abstract. Multiple resistances to bacterial and fungal pathogens due to *Pto*-type serine/threonine kinase (STK) has made this protein a promising candidate for resistance breeding and provided motivation to identify functional *Pto*-type genes from other plant species. In our present paper, eight classes of STKs were isolated from the cucumber (*Cucumis sativus* L.) introgressed line 5211S that shows downy mildew resistance. Phylogenetic analysis revealed six classes of cucumber *Pto*-type resistance gene analogues (RGAs). Multiple sequence alignment of deduced amino acid sequences from these *Pto*-type RGAs and *Pto* showed the typical structure of a *Pto* protein. Non-synonymous to synonymous nucleotide substitution (K_a : K_s) ratios for these *Pto*-type RGAs were less than 1, indicating that purifying selection played an important role in the evolutionary process. The *Pto*-type RGAs identified in this study will permit further genetic characterisation that may lead to the development of specific or even broad-spectrum pathogen resistance in 5211S.

Additional keywords: non-synonymous nucleotide substitution, *Pto*-RGA, synonymous nucleotide substitution.

Introduction

Over 40 plant disease resistance genes, often members of closely linked multigene families, have now been isolated from a wide variety of plant species by map-based cloning or transposon tagging. These proteins provide resistance to diverse pathogens such as insects, nematodes, fungi, oomycetes, bacteria and viruses (Xu *et al*. 2005). However, the required infrastructure for using these cloning techniques is not available for many crop plants. In this context, resistance gene analogues (RGAs) can now be easily identified by sequence similarity because certain functional domains are highly conserved, even among distantly related resistance genes. The occurrence of conserved structural features in plant disease resistance genes provides great potential for the isolation of novel sequences analogous to disease resistance genes. Therefore, RGAs can be cloned by PCR-based approaches in plants in which insufficient understanding of genetic information prevents the use of conventional molecular techniques (McDowell and Woffenden 2003).

Compared with nucleotide binding site-type RGAs (the largest class of disease resistance genes with a nucleotide binding site and a leucine-rich repeat region), very few *Pto*-type RGAs have been isolated from crops. *Pto* homologues have only been analysed from the *Solanum* genus (Vleeshouwers *et al*. 2001), common bean (*Phaseolus vulgaris* L.; Vallad *et al*. 2001), *Citrus* sp. (Deng and Gmitter 2003), grapevine (*Vitis vinifera* L.; Di Gaspero and Cipriani 2003), banana (*Musa acuminate* Colla; Peraza-Echeverría *et al*. 2007) and strawberry (*Fragaria ananassa* Duch.; Martínez Zamora *et al*. 2008). The *Pto* resistance gene was the first plant resistance (R) gene cloned that functions in a gene-for-gene manner (Martin *et al*. 1993). *Pto* resistance involves recognition of the bacterial protein AvrPto by two host proteins, *Pto* and *Prf*. The *Pto* gene encodes a serine/threonine kinase (STK) that interacts physically with the avirulence proteins AvrPto and AvrPtoB from *Psedomonas syringae* pv tomato, conferring hypersensitive response (HR)-mediated resistance. The STK can also interact with and phosphorylate a second STK, *Pti1*, and can bind to defence-related transcription factors *Pti4*, *Pti5* and *Pti6* (Cohn *et al*. 2001). Plants overexpressing *Pto* show resistance not only to *P. syringae* but also to *Xanthomonas campestris* pv vesicatoria and to the fungal pathogen *Cladosporium fulvum* (Tang *et al*. 1999). These findings make *Pto* an interesting candidate for engineering broad-spectrum pathogen resistance in agriculture and encourage the search for functional *Pto*-type genes in other plant species.

Despite tremendous advances in the structural molecular genetics of *R* genes, the evolution of *R* gene specificities remains poorly understood. Clarifying the evolutionary history of *R* genes is crucial for understanding how plants maintain and

 10.1071/FP09205 1445-4408/10/060513

adapt their defences to pathogens. Generally, gene duplication, which may occur through unequal recombination, and subsequent diversification are the usual mechanisms in plant *R* gene evolution (Richter and Ronald 2000). In tomato (*Lycopersicon esculentum* Mill.), the *Pto* gene belongs to a small gene family, consisting of six members (*Pto* and five *Pto* paralogues) (Martin *et al*. 1993). *Pto* paralogues share 78–91% nucleotide identity with *Pto*, and although most of these paralogues are functional protein kinases (PKs), none of them encode for recognition of AvrPto and AvrPtoB (Chang *et al*. 2002; Kim *et al*. 2002). It is presumed that duplication and diversification of the *Pto* gene family have led to the generation of alternative recognition specificities. The *Fen* gene, one of these five *Pto* paralogues, is 87% identical to *Pto* and participates in the same signal transduction pathway leading to HR, but is activated by a different signal (Martin *et al*. 1994; Loh and Martin 1995).

On the other hand, characterising nucleotide substitution patterns can provide insight into the evolution of resistance gene families. In most protein-coding genes, the $K_a : K_s$ ratio (ω) is less than 1; this observation is consistent with a functional constraint against amino acid replacements. Diversifying selection plays an important role in *R* gene evolution. Conversely, if ω is greater than 1, this indicates that adaptive selection events have fuelled divergence between genes (Richter and Ronald 2000).

Evidence for adaptive selection is rare, but it appears to be most commonly found in gene regions that function in host and pathogen recognition. Previous researchers studied sequence variation among the *Pto* alleles of seven species of wild tomatoes (*Lycopersicon* spp.). In comparison to the variation observed at 14 non-*R* gene loci, higher levels of non-synonymous and similar levels of synonymous polymorphisms have been found (Rose *et al*. 2007). These changes are thought to be subject to strong adaptive selection. A larger proportion of the *Pto* protein can tolerate more amino acid variation than the non-*R* genes, which shows that diversity is possibly neutral to *Pto* protein function. This also suggests that both purifying and balancing selection have influenced the evolution of the *Pto* locus (Rose *et al*. 2007).

We have been focusing on interspecific hybridisation between *Cucumis hystrix* Chakr. and *C. sativus* L. (Chen *et al*. 1997). Our objective is to improve current cucumber (*C. sativus*) varieties using the excellent characteristics of its wild relative, *C. hystrix*. At present, massive interspecific hybridisation introgressed lines (Zhou *et al.* 2009) and some special germplasm resources, such as allotetraploid (Chen and Kirkbride 2000), allotriploid (Chen *et al*. 2003) and monosomic alien addition lines (Chen *et al*. 2004), have been obtained by back- and self-crossing between interspecific hybrids and cultivated cucumbers whose phenotypic traits showed extensive segregation. Recently, we also identified an introgressed line ($2n = 14$) from the progenies of interspecific hybridisation of the *Cucumis* species that shows resistance to downy mildew (H. Wan and J. Chen, unpubl. data).

The objective of the present study was to isolate putative *Pto*-type sequences from the introgressed line 5211S and to study their molecular features, phylogenetic characteristics and evolutionary relationships.

Materials and methods

Plant material and DNA extraction

To identify *Pto*-type RGAs in cucumber, a downy mildew resistant cucumber introgressed line, 5211S, derived from interspecific hybridisation between *C. hystrix* and *C. sativus*, was used as the template for PCR amplification. Young leaves from 5211S were harvested for nucleic acid extraction. Genomic DNA was isolated using a commercial plant DNA extraction kit (Bioteke, Beijing, China).

PCR amplification

To amplify the region between subdomain I and IX of STKs in 5211S, a pair of degenerate primers designed by Vallad *et al*. (2001), which were based on conserved regions of amino acid identity between the STKs *Pto*, *Fen* and *Pti1* from tomato, and *MHK* and *APK1* from *Arabidopsis thaliana* L., were selected. Primer sequences were P3 (subdomain I): 5′-TNGGNSANG GNGKNTTYGG-3′ and P2R (subdomain IX): 5′-ACNCCR AANGARTANACRTC-3′. A 50 µl PCR reaction contained 2U Taq DNA polymerase (TaKaRa, Kyoto, Japan), 1× PCR buffer, 1.5 mM $MgCl_2$, 0.5 µM each of forward and reverse primers, 0.4 mM dNTP, and 100 ng of template DNA. PCR was carried out in a PTC-200 thermal cycler (MJ Research, Inc., Waltham, MA, USA) programmed for an initial denaturation at 94°C for 3 min, followed by 40 cycles at 94°C for 30 s, 50°C for 45 s, 72°C for 1 min and a final extension step of 7 min at 72°C; the mixture was then held at 4°C.

Cloning and sequencing of PCR products

DNA fragments from PCR were separated on 1.0% agarose gels. Fragments of the expected size were excised and reclaimed from the gel and purified with a MinElute purification kit (Qiagen, Hilden, Germany). The obtained DNA fragments were cloned into a pGEM-T vector (Promega, Madison, WI, USA). Blue versus white screening was employed to select recombinant clones. All of the white colonies were collected and subjected to re-amplification with M13 universal primers (Bioteke, Beijing, China). The inserts were selected for sequencing with the ABI 3730 sequencer (Applied Biosystems, Foster City, CA, USA).

Sequence and similarity analysis and multiple sequence alignment

Each of the subclones was trimmed of vector sequence contamination using VecScreen at NCBI (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html, accessed 16 December 2008). The sequence data were subjected to GenBank searches with BLASTX algorithm (Altschul *et al*. 1997). Searches for open reading frames (ORF) were done using ORF Finder at the NCBI server (http://www.ncbi.nlm.nih.gov/projects/gorf/, accessed 16 December 2008). Assembly of DNA sequences and translation to the predicted amino acid sequence were carried out using the DNAMAN software (Version 4.03). Multiple alignments of amino acid sequences were performed using the CLUSTAL_X program from BioEdit software (Thompson *et al*. 1997).

Phylogenetic and evolutionary analysis

A total of 18 *Pto*-type RGAs were used to construct phylogenetic trees based on the bootstrap neighbour-joining (NJ) method with a Kimura 2-parameter model by MEGA v4.0 (Tamura *et al*. 2007). The stability of internal nodes was assessed by bootstrap analysis with 1000 replicates. The number of non-synonymous substitutions per non-synonymous site and the number of synonymous substitutions per synonymous site were denoted by K_a and K_s, respectively. The ratios of non-synonymous (K_a) versus synonymous (K_s) substitution values were calculated among the *Pto*-type RGAs of each of the phylogenetic classes using the protocol of Nei and Gojobori (Nei and Gojobori 1986) with the K-Estimator 6.0 software package (Comeron 1995, 1999).

Results

Isolation of Pto*-type resistance gene analogues and other serine/threonine kinase-like sequences in 5211S*

Four major DNA bands of ~450 bp, 550 bp, 650 bp and 950 bp were observed after PCR amplification from 5211S with a pair of degenerate primers, P3 and P2R, which were designed by Vallad *et al*. (2001) (Fig. 1). Among these bands, only the band of ~550 bp corresponded to the expected size of the PCR amplification product of the tomato *Pto* resistance gene. Therefore, only this 550 bp band was cloned for further analysis.

This 550 bp band was recovered from the agarose gel, purified and cloned. Seventy clones were selected for DNA sequencing. Of these, 31 clones showed a high level of identity to the corresponding STK gene sequences and *Pto*-type RGAs from other species such as *Brassica oleracea* L., *A. thaliana* and *Oryza sativa* L. in GenBank. Sequence analysis revealed the presence of stop codons or frame shift mutations in 13 of the 31 RGA genomic clones. These 13 sequences appeared to be non-functional genes or pseudogenes, and they were excluded from further investigation. The remaining 18 sequences (designated CSPTO1 to CSPTO18) contained uninterrupted ORFs. BLASTX searches of 18 sequences revealed the presence of the conserved subdomains I and IX (FK/F/L/SVYK/RG, DY/IYSF/YG), which were used to design degenerate primers for isolating *Pto*-type RGAs from other crops.

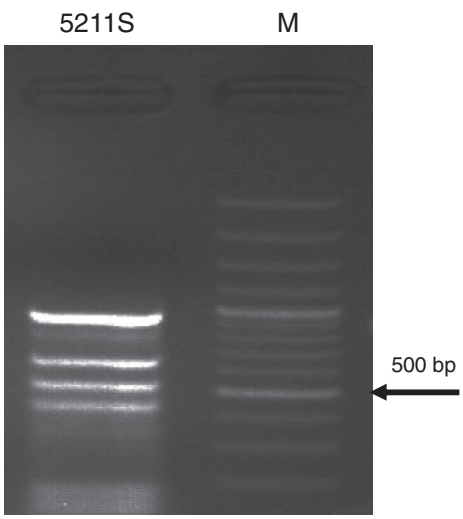


Fig. 1. PCR amplification products generated by a pair of degenerate primers (P3 and P2R) in 5211S. M: marker3000, 5211S: the introgressed line of *C. hystrix*/*C. sativus* with high resistance to downy mildew.

Phylogenetic analysis showed that 18 STK amino acid sequences of 5211S separated into eight different groups. (Fig. 2). Among these eight groups, six were designated as cucumber *Pto* resistance gene analogues (*Pto*-RGAs) based on their significant similarity with the disease resistance protein *Pto* from tomato. The remaining two groups were designated as cucumber STKs, since they showed significant similarity to other types of STKs, including wall-associated receptor-like kinase and hypothetical proteins (Table 1). Each group was named after a single clone representative of the group and used for further analysis in this study.

Genetic diversity analysis of Pto*-type sequences and serine/threonine kinases in 5211S*

The six classes of *Pto*-type RGAs were highly divergent. Sequence similarity among the six classes ranged from 61.1% (between CSPTO1 and CSPTO14) to 81.0% (between CSPTO3 and CSPTO5) at the nucleotide level and from 57.5% (between CSPTO1 and CSPTO5) to 86.7% (between CSPTO9 and CSPTO5) at the amino acid level. Amino acid identity between the predicted amino acid sequence of cucumber *Pto*-RGAs and the tomato *Pto* protein ranged from 49.7% (CSPTO14) to 58.8% (CSPTO2) (Table 2). The percent amino acid similarity values between the classes were less than 85%, except for that between CSPTO2 and CSPTO10, which shared more than 86.7% amino acid similarity (Table 2).

In addition, BLASTX searches revealed that other STK sequences (non-*Pto*-type RGAs) in 5211S showed the highest BLAST scores when compared with the STKs from other plant species, as reported in Table 1. Additionally, BLASTX searches also found that one *Pto*-type RGA in 5211S, CSPTO10, was highly similar to *Mangifera indica* L. sequences present in the GenBank database as accession number AAT94933 from *M. indica*. They shared ~93% amino acid similarity and 92% identity, whereas other sequence similarities ranged from 74% to 96% and identities ranged from 66% to 92%.

Phylogenetic analysis of the serine/threonine kinase family in 5211S

A phylogenetic analysis was conducted to determine the evolutionary relationships among eight classes of STK sequences from 5211S, the tomato *Pto* R protein, and 38 STKs from *A. thaliana* (Hardie 1999; Hartwell *et al*. 1999) (Fig. 3). Six classes of 5211S STKs (1st–6th) that had previously been identified as *Pto*-type RGAs and the *Pto* R protein from tomato were classified into one group based on these sequence alignments, indicating that the 5211S STKs have the greatest similarity to the disease resistance protein *Pto* from tomato. This further supported the designation of these sequences as *Pto*-type STKs. In contrast, the two remaining 'non-*Pto* type RGAs' revealed a high degree of identity and similarity to serine/threonine kinases.

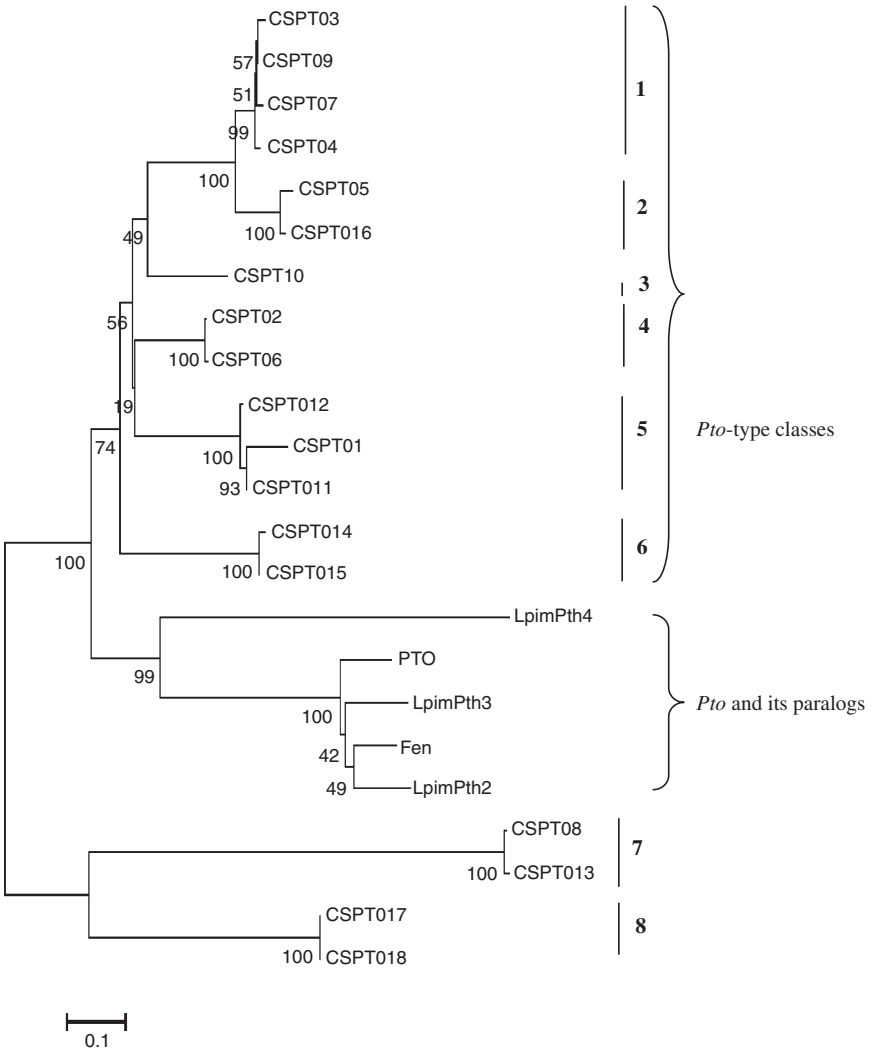


Fig. 2. Neighbour-joining phylogenetic tree based on the CLUSTALX alignment of tomato *Pto* R protein and its paralogues (GenBank accession number AF220602), and the deduced amino acid sequences of 5211S STKs (GenBank accession numbers GQ423631 to GQ423648). Amino acids from the end of subdomain I to the start of subdomain IX of STKs were used for the analysis. Branches corresponding to cucumber STK classes are labelled 1 to 8.

Table 1. Best BLASTX alignment hits among serine/threonine kinases (STKs) from 5211S and other species

STK-type sequences (CSPTO14, CSPTO8, CSPTO5, CSPTO10, CSPTO2, CSPTO11, CSPTO9 and CSPTO18) GenBank accession numbers GQ423644, GQ423638, GQ423635, GQ423640, GQ423632, GQ423641, GQ423639 and GQ423648, respectively

STK-type sequences	GenBank accession number	Plant	Similar to	Length of BLASTX alignment (in amino acids)	Identity %	Similarity %	Expected (E) value
CSPTO14 (555 bp)	ACA05215	*Fragaria ananassa*	*Pto*-type serine/threonine kinase	185	78	86	3e-80
CSPTO8 (546 bp)	AAF81356	*Arabidopsis thaliana*	Wall-associated receptor-like kinase	182	70	84	4e-71
CSPTO5 (543 bp)	ACA05210	*Fragaria chiloensis*	*Pto*-type serine/threonine kinase	181	66	74	8e-47
CSPTO10 (546 bp)	AAT94933	*Mangifera indica*	*Pto*-type serine/threonine kinase	182	92	93	4e-80
CSPTO2 (546 bp)	AAT28295	*Rosa roxburghii*	*Pto*-type serine/threonine kinase	182	91	96	1e-93
CSPTO11 (546 bp)	AAQ82660	*Capsicum chinense*	*Pto*-type serine/threonine kinase	182	82	91	5e-85
CSPTO9 (543 bp)	AAT28293	*Rosa roxburghii*	*Pto*-type serine/threonine kinase	181	76	86	6e-75
CSPTO18 (627 bp)	YP_002275845	*Vitis vinifera*	Hypothetical protein	209	70	79	1e-50

Table 2. Percent amino acid identity between the 5211S *Pto*-type RGAs and the tomato *Pto* disease resistance protein

Pto RGA[A]	CSPTO9	CSPTO10	CSPTO11	CSPTO14	CSPTO5	*Pto*
CSPTO2	72.5	77.6	74.3	71.0	67.4	58.8
CSPTO9	–	73.6	68.1	65.9	86.7	51.7
CSPTO10	–	–	73.2	66.7	68.5	55.4
CSPTO11	–	–	–	67.2	62.4	55.9
CSPTO14	–	–	–	–	60.8	49.7
CSPTO5	–	–	–	–	–	51.4

[A]Amino acids from the end of subdomain I to the start of subdomain IX of STKs were used for the analysis.

Phylogenetic analysis also revealed that each of the six classes of *Pto*-type RGAs contained between two and five subgroups, with the exception of the fourth class, which included only one member (Fig. 2).

The deduced amino acid sequences of the six *Pto*-type RGAs and the corresponding region of the disease resistance *Pto* gene from tomato were aligned using CLUSTALX (version 1.83) software. The alignment revealed that nine conserved subdomains (class I to class IX) present in the disease resistance *Pto* protein from tomato were also highly conserved in the 5211s *Pto*-type RGAs. The activation domain between subdomains VII and VIII (the region between amino acids 182 and 211 of *Pto*) and its internal P+1 loop site were also present in 5211S *Pto*-type RGAs (Fig. 4).

Multiple amino acid sequence alignment of Pto*-type sequences in 5211S*

In the *Pto*-type sequences from 5211S, 14 amino acid residues of the protein kinase consensus were highly conserved, with the exception of two variant amino acids distributed throughout the sequences, such as K166 in CSPTO3 and 55S in CSPTO1 (Table 3). Furthermore, two amino acid deletions (T133 and M134) and three amino acid insertions (positions 160 to 162) were observed in the *Pto*-type sequences from 5211S corresponding to the conserved region between subdomain I and subdomain IX from the tomato disease resistance protein *Pto*. Among these, 'AQG' was inserted in the third and fourth *Pto*-type RGAs. 'AGG', 'SSG', 'TKA' and 'NAT' were inserted in the first, second, fifth and sixth *Pto*-type RGAs. At the same time, in this region, four of the five autophosphorylation sites (T133, T190, T195, S198 and T199) present in the disease resistance *Pto* protein from tomato were observed in the *Pto*-type sequences from 5211S that were inside the activation domain. Among the four autophosphorylation sites, T195 and T199 were fully conserved. T190 was substituted for P190 in the first, third, fourth and fifth class, and for S190 and A190 in the second and sixth class of *Pto*-type RGAs. S198 was conserved in the majority of cucumber *Pto*-type RGAs. The exception was CSPTO7 from the sixth class, in which the serine was replaced by asparagine.

Previous studies have shown that some amino acid residues that proved to be involved in *Pto*–AvrPto recognition or binding are essential for understanding *Pto* function. For example, residues S76 and G78 were important for the recognition specificity (Bernal *et al*. 2005), but their degrees of conservation were different. G78 was fully conserved, whereas S76 showed various degrees of conservation. Similar results were also observed for the 5211S *Pto*-type RGAs. Residues V55 and H94, which are involved in the interaction with AvrPto, were also conserved in 5211S *Pto*-type RGAs (Scofield *et al*. 1996).

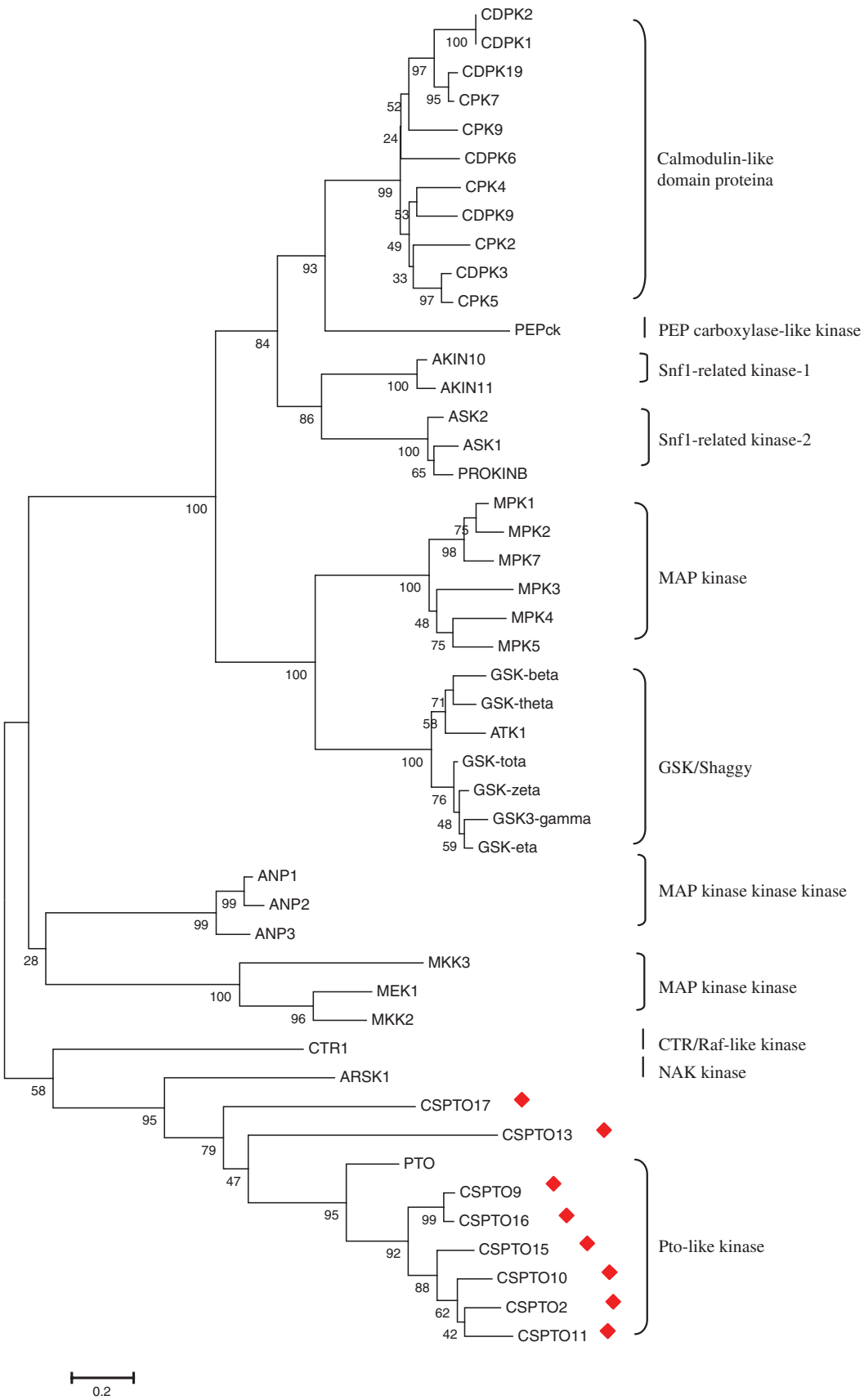
Five residues (K202, T204, L205, Y207 and D209) are essential for *Pto* function (Frederick *et al*. 1998; Rathjen *et al*. 1999). In the present study, three of these residues, K202, Y207 and D209, which influence binding properties between *Pto* and AvrPto in the P+1 loop, are highly fully conserved in all the *Pto*-type RGAs. The residue T204, required for recognition specificity, is conserved in two classes of *Pto*-type RGAs (the first and second), whereas in the remaining four classes (the third to sixth), it is replaced by a serine. The Y205 residue, which plays a subsidiary role in recognition, is replaced by a phenylalanine in all the *Pto*-type RGAs, with the exception of the second class, which contains an isoleucine at this site.

Evolutionary analysis of Pto*-type sequences in 5211S*

The $K_a : K_s$ ratio (ω) is an indicator of evolutionary pressures in resistance genes (Michelmore and Meyers 1998). $\omega < 1$ may result from the elimination of most non-synonymous substitutions through purifying selection. $\omega > 1$ indicates diversifying selection, whereas $\omega = 1$ shows neutral selection (Martin *et al*. 1993). In this study, frequencies of non-synonymous (K_a), synonymous (K_s) nucleotide substitution and ω were calculated among *Pto*-type RGAs within a maximum nucleotide sequence identity of 99%, and between members of clusters 1, 2, 3, 4, 5 and 6 that belonged to 5211S *Pto*-type RGAs (Table 4). All values of ω were significantly less than 1 within pairwise comparisons, suggesting that all the nucleotide variations between *Pto*-type RGAs represented synonymous substitutions. The ratios for all of the pairwise comparisons were significantly <1, indicating that purifying selection plays an important role in the evolutionary processes of the *Pto*-type resistance gene.

Discussion

Eighteen 5211S STKs, clustered within eight classes, were successfully isolated using a PCR cloning strategy with degenerate primers P3 and P2R, which were designed based on conserved regions of subdomain I and IX of the serine/threonine domain of tomato and *A. thaliana* by Vallad *et al*. (2001). Among these, six classes were *Pto*-type RGAs, whereas the remaining two classes were other STK types (non-*Pto* type RGAs), based on phylogenetic tree analysis. All six amplified classes of *Pto*-type RGAs showed nine fully conserved serine/threonine kinase motifs and plant STK consensus sequences,


CDPK2
CDPK1
CDPK19
CPK7
CPK9
CDPK6
CPK4
CDPK9
CPK2
CDPK3
CPK5
Calmodulin-like domain proteina
PEPck
PEP carboxylase-like kinase
AKIN10
AKIN11
Snf1-related kinase-1
ASK2
ASK1
PROKINB
Snf1-related kinase-2
MPK1
MPK2
MPK7
MPK3
MPK4
MPK5
MAP kinase
GSK-beta
GSK-theta
ATK1
GSK-tota
GSK-zeta
GSK3-gamma
GSK-eta
GSK/Shaggy
ANP1
ANP2
ANP3
MAP kinase kinase kinase
MKK3
MEK1
MKK2
MAP kinase kinase
CTR1
CTR/Raf-like kinase
ARSK1
NAK kinase
CSPTO17
CSPTO13
PTO
CSPTO9
CSPTO16
CSPTO15
CSPTO10
CSPTO2
CSPTO11
Pto-like kinase
0.2

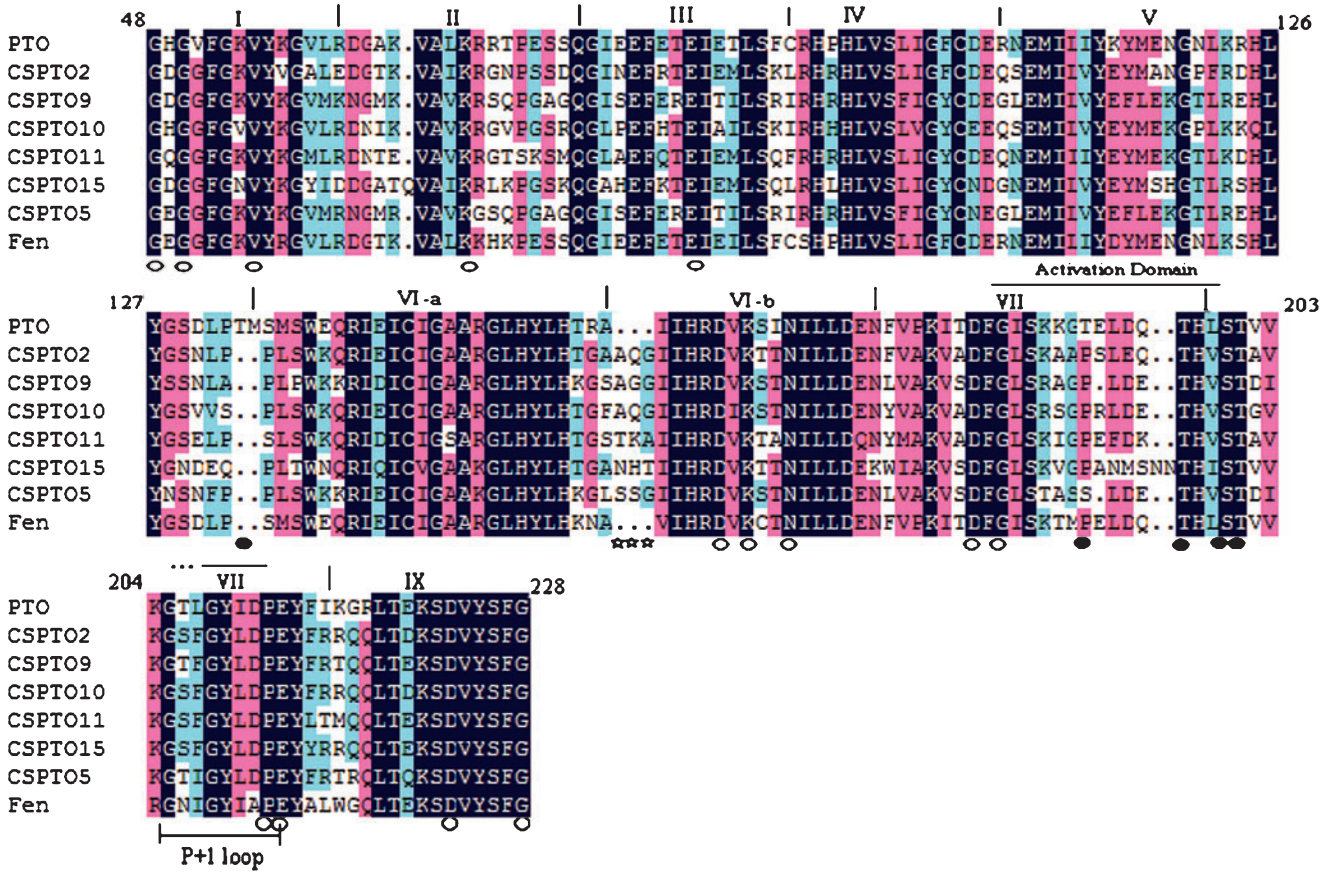


Fig. 4. CLUSTALX multiple-amino acid sequence alignment of 5211S *Pto*-like sequences with the corresponding region of the tomato R proteins *Pto* and *Fen*. Conserved subdomains (labelled with Roman numbers), the activation domain and the P+1 loop are indicated. Kinase consensus sequences in *Pto* at positions 48, 50, 55, 69, 84, 164, 166, 169, 182, 184, 210, 211, 223 and 228 are indicated by ellipses. *Pto* autophosphorylation sites are indicated by filled ellipses. *Pto* insertion and deletion sites are indicated by pentacles.

suggesting that the genes revealed here are likely to encode active kinases and are members of a large family of *Pto*-type kinases.

Sequence alignment showed that six classes of cucumber *Pto*-type RGAs contained the same nine conserved domains present in the disease resistance gene *Pto* from tomato. In addition, the region of subdomain IX was more highly conserved than that of subdomain I, which was used to design degenerate primers, suggesting that the extent of the different conserved regions is distinct. Two amino acid deletions (position 133 and 134) and three amino acid insertions (positions 160, 161 and 162) were also observed in the corresponding region of all 5211S *Pto*-type RGAs, the reason for this is unclear. This phenomenon was also clearly observed in *Pto*-type RGAs from other species such as the monocot *Musa acuminata* (GenBank accession number EF492528) and the dicot *A. thaliana* (GenBank accession number NP197789). However, three amino acid insertions were absent in *Pto*-type RGAs from the *Solanaceae* (Vleeshouwers *et al*. 2001).

The invariant kinase consensus positions in all *Pto*-type RGAs cloned from 5211S were also conserved through sequence alignment, suggesting that these are likely to encode active kinase. Additionally, in all 5211S *Pto*-type RGAs, the autophosphorylation site T133 is absent, although in the *Pto*-type RGAs from *Solanum*, the threonine 133 was often replaced by a serine, thereby providing an alternative phosphorylation site. The phosphorylation sites T195, S198 and T199 were conserved, whereas T190 was variable in the *Pto*-type RGAs of 5211S. Therefore, it is conjectured that the 'polymorphism' of these phosphorylation sites probably occurred before the divergence of solanaceous and other species.

The findings from Scofield *et al*. (1996) showed that mutations at V55 and H94 positions disrupted the interaction between *Pto* and AvrPto in yeast and abrogated *Pto*-mediated

Fig. 3. Plant protein kinase tree derived from comparing the deduced amino acid sequences of eight classes of 5211S STK sequences (CSPTO17, CSPTO13, CSPTO9, CSPTO16, CSPTO15, CSPTO10, CSPTO2, and CSPTO11; GenBank accession number GQ423647, GQ423643, GQ423639, GQ423646, GQ423645, GQ423640, GQ423632, and GQ423641, respectively) and the other 10 classes of STKs protein kinase and serine/threonine kinase from *A. thaliana* (Hardie 1999; Hartwell *et al*. 1999) by using CLUSTALX alignment. *Pto*-type sequences and the other STKs sequence of 5211S (filled rhombus) are indicated.

Table 3. Comparison of cucumber *Pto*-type sequences with *Pto* and four of its paralogues from tomato

Class	*Pto* and *Pto*-type sequence[A]	Deletion	Insertion	A[B]		B[C]		Kinase consensus[D]	C[E]					Autophosphorylation sites[F]				
		133	159	76	78	55	94		202	204	205	207	209	133	190	195	198	199
	Pto	TM		S	G	V	H	GGVKEDKNDGPEDG	K	T	L	Y	D	T	T	T	S	T
1	CSPTO3		AGG	G	G	V	H	GGVKEDQNDGPEDG	K	T	F	Y	D		P	T	S	T
	CSPTO4		AGG	A	G	V	H	GGVKEDKNDGPEDG	K	T	F	Y	D		P	T	S	T
	CSPTO7		AGG	G	G	V	H	GGVKEDKNDGPEDG	K	T	F	Y	D		P	T	N	T
	CSPTO9		AGG	G	G	V	H	GGVKEDKNDGPEDG	K	T	F	Y	D		P	T	S	T
2	CSPTO16		SSG	G	G	V	H	GGVKEDKNDGPEDG	K	T	I	Y	D		S	T	S	T
	CSPTO5		SSG	G	G	V	H	GGVKEDKNDGPEDG	K	T	I	Y	D		S	I	S	T
3	CSPTO10		AQG	R	G	V	H	GGVKEDKNDGPEDG	K	S	F	Y	D		P	T	S	T
4	CSPTO2		AQG	D	G	V	H	GGVKEDKNDGPEDG	K	S	F	Y	D		P	T	S	T
	CSPTO6		AQG	D	G	V	H	GGVKEDKNDGPEDG	K	S	F	Y	D		P	T	S	T
5	CSPTO1		TKA	M	G	S	H	GGSKEDKNDGPEDG	K	S	F	Y	D		P	T	S	T
	CSPTO11		TKA	M	G	V	H	GGVKEDKNDGPEDG	K	S	F	Y	D		P	T	S	T
	CSPTO12		TKA	M	G	V	H	GGVKEDKNDGPEDG	K	S	F	Y	D		P	T	S	T
6	CSPTO14		NAT	K	G	V	H	GGVEEDKNDGPEDG	K	S	F	Y	D		A	T	S	T
	CSPTO15		NAT	K	G	V	H	GGVKEDKNDGPEDG	K	S	F	Y	D		A	T	S	T

[A]See Fig. 2.

[B]Residues S76 and G78, identified by the DNA-shuffling approach, provide specificity in binding AvrPtoB but do not affect binding to AvrPto (Bernal *et al.* 2005).

[C]Residues V55 and H94, identified as important for resistance by ethyl methane sulfonate mutagenesis (Scofield *et al.* 1996).

[D]Kinase consensus sequence at positions 48, 50, 55, 69, 84, 164, 166, 169, 182, 184, 210, 211, 223 and 228 (Hanks and Quinn 1991).

[E]Positions inside the P+1 loop that are important for AvrPto binding specificity, identified by swaps between *Pto* and *Fen* together with site-directed mutagenesis (Scofield *et al.* 1996; Frederick *et al.* 1998), and by alanine substitution analysis (Rathjen *et al.* 1999).

[F]Pto autophosphorylation sites (positions133, 195, 198 and 199) between subdomain I to IX of Pto-type sequences from cucumber (Sessa *et al.* 2000).

Table 4. K_a : K_s ratios for pairwise comparisons of *Pto*-type RGAs from 5211S
Significantly different ($P < 0.05$) from a ω value of 1

Pto-type pair[A]	Degree of nucleotide acid sequence identity	Degree of amino acid sequence identity	K_a	K_s	ω[B]	χ^{2C}	Probability
CSPTO2/CSPTO6	98.7	99.4	0.00226	0.00987	0.229	0.050	0.82266
CSPTO3/CSPTO4	98.7	97.8	0.01008	0.02189	0.460	0.040	0.84123
CSPTO3/CSPTO7	98.4	97.8	0.00955	0.04308	0.221	0.062	0.80334
CSPTO3/CSPTO9	98.5	98.9	0.00451	0.04906	0.091	0.692	0.40541
CSPTO4/CSPTO7	98.5	97.8	0.01059	0.03234	0.327	0.00013	0.99085
CSPTO4/CSPTO9	98.3	98.9	0.00555	0.04978	0.111	0.490	0.48398
CSPTO7/CSPTO9	98.5	98.9	0.00501	0.05405	0.092	0.684	0.40809
CSPTO5/CSPTO16	81.0	97.2	0.06391	0.93330	0.068	1.150	0.28346
CSPTO1/CSPTO11	98.1	93.9	0.01267	0.01684	0.752	0.264	0.60770
CSPTO2/CSPTO11	68.3	74.2	0.20136	2.09964	0.096	0.661	0.41615
CSPTO1/CSPTO2	67.8	68.8	0.21673	1.95046	0.111	0.500	0.47964
CSPTO14/CSPTO15	98.2	98.9	0.00487	0.06089	0.080	0.897	0.34364

[A]ω was calculated for *Pto*-type pairs with a maximum nucleotide sequence identity of 99% between members of clusters 1, 2, 3, 4, 5 and 6.
[B]Ratio of non-synonymous to synonymous substitutions.
[C]χ^2 values were calculated with Yates' adjustment using a 2 × 2 contingency table. χ^2 values: $\chi^2_{(0.05,\,1)} = 3.84$, $\chi^2_{(0.01,\,1)} = 6.63$; $^*0.01 < P < 0.05$, $^{**}0.001 < P < 0.01$, $^{***}P < 0.001$.

resistance, suggesting that these two residues played an important role in the interaction of *Pto* and AvrPto in tomato. In the present study, these two residues were highly conserved in all of the 5211S *Pto*-type RGAs. Bernal *et al*. (2005) identified two amino acid positions, S76 and G78, as critical for recognition specificity and found, via PAML analysis of *Pto* orthologues and paralogues, that the position S76 may have been subject to significant divergent selection. In *Pto*-type RGAs from 5211S, residue G78 is highly conserved, whereas the other residue, S76, is divergent, suggesting that these residues may have similar roles to G78 and S76 in the *Pto* resistance gene from tomato.

To characterise the interrelation between cucumber STKs (*Pto*-type RGAs and non-*Pto* type RGAs), the disease resistance gene protein *Pto* from tomato and STKs from other species, a phylogenetic tree was constructed using CLUSTALX software (version 1.83) based on amino acid alignment. All of the *Pto*-type RGAs and *Pto* from tomato clustered in one class, suggesting these *Pto*-type RGAs and *Pto* from tomato may share a common evolutionary origin and are likely to have a similar function in disease resistance. Furthermore, according to Peraza-Echeverría's preliminary inference principle (Peraza-Echeverría *et al*. 2007), the six classes of 5211S *Pto*-type RGAs clustered in this study represent a significant proportion of all the *Pto*-type RGAs from a cucumber introgressed line with resistance to downy mildew.

The percentage of identity of the *Pto*-type RGAs from 5211S with the disease resistance protein *Pto* from tomato ranged from 49.7% to 58.8% (Table 3), which is similar to that from banana (Peraza-Echeverría *et al*. 2007), strawberry (Martínez Zamora *et al*. 2008) and common bean (Vallad *et al*. 2001). In addition, 18 *Pto*-type RGAs cloned from 5211S in this study shared 62.8–98.7% identity at the nucleotide level, with the highest value appearing for CSPTO1 and CSPTO11 (98.7%), and the lowest value appearing for CSPTO1 and CSPTO14 (62.8%), suggesting that cucumber *Pto*-type RGAs are highly divergent. An assumption can be made that CSPTO1 and CSPTO11 might have arisen from a recent duplication of a common ancestral gene, whereas CSPTO1 and CSPTO14 represent divergent origins.

Published models for the evolution of resistance genes propose recombination, gene conversion, point mutations and unequal crossing over as the primary mechanisms (Michelmore and Meyers 1998). Recombination between alleles and paralogues could result in novel binding properties (Meyers *et al*. 1998). At the same time, the importance of single base changes in the evolution of plant disease resistance genes has also been recognised (Parniske *et al*. 1997; Meyers *et al*. 1998; Wang *et al*. 1998). The relative importance of recombination versus single base changes depends on the rate at which each of these events occurs. In this study, the high degree of sequence diversity for cucumber *Pto*-type RGAs indicated that recombination and gene conversion had not homogenised these sequences and therefore were infrequent events. Novel amino acid substitutions may therefore be more important than intergenic recombination and gene conversion in the rapid evolution of novel specificities. Single nucleotide differences, insertions and deletions were all found in *Pto*-type RGAs from 5211S.

Calculation of nucleotide base substitution (non-synonymous and synonymous substitution) can be provided by ω analysis (Bergelson *et al*. 2001). In this study, the average ω in each class of cucumber *Pto*-type RGA was significantly less than 1, suggesting that purifying rather than diversifying selection (Table 4) had occurred. However, since the full length of these *Pto*-type RGAs has not been obtained, we cannot exclude the possibility that positive selection might have occurred.

In summary, we have identified 18 *Pto*-type RGAs from 5211S and have analysed their phylogenetic relationships with known *R* genes. This enabled us to gain new insight into their structure and evolution, and to lay the foundation for ultimately isolating the functional *R* gene.

Acknowledgements

This research was partially supported by the Key Program (30830079), the General Programs 30671419 and 30700541 from the National Natural Science Foundation of China, and the '863' Programs 2008AA10Z150, 2006AA10Z1A8 and 2006AA100108.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402. doi:10.1093/nar/25.17.3389

Bergelson J, Kreitman M, Stah EA, Tian D (2001) Evolutionary dynamics of plant *R*-genes. *Science* **292**, 2281–2285. doi:10.1126/science.1061337

Bernal A, Pollack J, Pan Q, Rose L, Kozik A (2005) Functional analysis of the plant resistance gene *Pto* using DNA shuffling. *The Journal of Biological Chemistry* **280**, 23 073–23 083. doi:10.1074/jbc.M500992200

Chang JH, Tai YS, Bernal AJ, Lavelle DT, Staskawicz BJ (2002) Functional analyses of the *Pto* resistance gene family in tomato and the identification of a minor resistance determinant in a susceptible haplotype. *Molecular Plant–Microbe Interactions* **15**, 281–291. doi:10.1094/MPMI.2002.15.3.281

Chen JF, Kirkbride JH (2000) A new synthetic species of *Cucumis* (Cucurbitaceae) from interspecific hybridization and chromosome doubling. *Bittonia* **52**(4), 315–319. doi:10.2307/2666583

Chen JF, Staub JE, Tashiro Y, Isshiki S, Miyazaki S (1997) Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr. *Euphytica* **96**, 413–419. doi:10.1023/A:1003017702385

Chen JF, Staub JE, Qian CT, Jiang J, Luo XD, Zhuang FY (2003) Reproduction and cyogenetic characterization of interspecific hybrids derived from *Cucumis hystrix* Chakr. × *C. sativus* L. *Theoretical and Applied Genetics* **106**, 688–695.

Chen JF, Luo XD, Qian CT, Jahn MM, Staub JE, Zhuang FY, Lou QF, Ren G (2004) *Cucumis* monosomic alien addition lines: morphological, cytological, and genotypic analyses. *Theoretical and Applied Genetics* **108**, 1343–1348. doi:10.1007/s00122-003-1546-z

Cohn J, Sessa G, Martin GB (2001) Innate immunity in plants. *Current Opinion in Immunology* **13**, 55–62. doi:10.1016/S0952-7915(00)00182-5

Comeron JM (1995) A method for estimating the numbers of synonymous and non-synonymous substitutions per site. *Journal of Molecular Evolution* **41**, 1152–1159. doi:10.1007/BF00173196

Comeron JM (1999) K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics (Oxford, England)* **15**, 763–764. doi:10.1093/bioinformatics/15.9.763

Deng Z, Gmitter FG (2003) Cloning and characterization of receptor kinase class disease resistance gene candidates in *Citrus*. *Theoretical and Applied Genetics* **108**, 53–61. doi:10.1007/s00122-003-1410-1

Di Gaspero G, Cipriani G (2003) Nucleotide binding site/leucine-rich repeats, *Pto*-like and receptor-like kinases related to disease resistance in grapevine. *Molecular Genetics and Genomics* **269**, 612–623. doi:10.1007/s00438-003-0884-5

Frederick RD, Thilmony RL, Sessa G, Martin GB (1998) Recognition specificity for the bacterial avirulence protein AvrPto is determined by Thr-204 in the activation loop of the tomato *Pto* kinase. *Molecular Cell* **2**, 241–245. doi:10.1016/S1097-2765(00)80134-3

Hanks SK, Quinn AM (1991) Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods in Enzymology* **200**, 38–62. doi:10.1016/0076-6879(91)00126-H

Hardie DG (1999) Plant protein serine/threonine kinases: classifications and functions. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**, 97–131. doi:10.1146/annurev.arplant.50.1.97

Hartwell J, Gill A, Nimmo GA, Wilkins MB, Jenkins GI, Nimmo HG (1999) Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. *The Plant Journal* **20**, 333–342. doi:10.1046/j.1365-313X.1999.t01-1-00609.x

Kim YJ, Lin NC, Martin GB (2002) Two distinct *Pseudomonas* effector proteins interact with the Pto kinase and activate plant immunity. *Cell* **109**, 589–598. doi:10.1016/S0092-8674(02)00743-2

Loh YT, Martin GB (1995) The *Pto* bacterial resistance gene and the *Fen* insecticide sensitivity gene encode functional protein kinases with serine/threonine specificity. *Plant Physiology* **108**, 1735–1739. doi:10.1104/pp.108.4.1735

Martin GB, Brommonschenkel S, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432–1436. doi:10.1126/science.7902614

Martin GB, Frary A, Wu T, Brommonschenkel S, Chunwongse J, Earle ED, Tanksley SD (1994) A member of the tomato *Pto* gene family confers sensitivity to fenthion resulting in rapid cell death. *The Plant Cell* **6**, 1543–1552.

Martínez Zamora MG, Castagnaro JC, Díaz Ricci JC (2008) Genetic diversity of *Pto*-like serine/threonine kinase disease resistance genes in cultivated and wild strawberries. *Journal of Molecular Evolution* **67**, 211–221. doi:10.1007/s00239-008-9134-0

McDowell JM, Woffenden BJ (2003) Plant disease resistance genes: recent insights and potential applications. *Trends in Biotechnology* **21**(4), 178–183. doi:10.1016/S0167-7799(03)00053-2

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in highly duplicated and spans several megabases. *The Plant Cell* **10**, 1817–1832.

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research* **8**, 1113–1130.

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418–426.

Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BBH, Jones JDG (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf4/9 locus of tomato. *Cell* **91**, 821–832. doi:10.1016/S0092-8674(00)80470-5

Peraza-Echeverría S, James-Kay A, Canto-Canché B, Castillo-Castro E (2007) Structural and phylogenetic analysis of *Pto*-type disease resistance gene candidates in banana. *Molecular Genetics and Genomics* **278**, 443–453. doi:10.1007/s00438-007-0262-9

Rathjen JP, Chang JH, Staskawicz BJ, Michelmore RW (1999) Constitutively active *Pto* induces a Prf-dependent hypersensitive response in the absence of AvrPto. *The EMBO Journal* **18**, 3232–3240. doi:10.1093/emboj/18.12.3232

Richter TE, Ronald PC (2000) The evolution of disease resistance genes. *Plant Molecular Biology* **42**, 195–204. doi:10.1023/A:1006388223475

Rose LE, Michelmore RW, Langley CH (2007) Natural variation in the *Pto* pathogen resistance gene within species of wild tomato (*Lycopersicon*): II. Population genetics of *Pto*. *Genetics* **175**, 1307–1319. doi:10.1534/genetics.106.063602

Scofield SR, Tobias CM, Rathjen JP, Chang JH, Lavelle DT, Michelmore RW, Staskawicz BJ (1996) Molecular basis of gene for gene specificity in bacterial speck disease of tomato. *Science* **274**, 2063–2065. doi:10.1126/science.274.5295.2063

Sessa G, D'Ascenzo M, Martin GB (2000) Thr38 and Ser198 are Pto autophosphorylation sites required for the AvrPto-Pto-mediated hypersensitive response. *The EMBO Journal* **19**, 2257–2269. doi:10.1093/emboj/19.10.2257

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596–1599. doi:10.1093/molbev/msm092

Tang X, Xie M, Kim YJ, Zhou J, Klessig DF, Martin GB (1999) Overexpression of *Pto* activates defense responses and confers broad resistance. *The Plant Cell* **11**, 15–29.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882. doi:10.1093/nar/25.24.4876

Vallad G, Rivkin M, Vallejos C, McClean P (2001) Cloning and homology modelling of a *Pto*-like protein kinase family of common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* **103**, 1046–1058. doi:10.1007/s001220100705

Vleeshouwers VGAA, Martens A, van Dooijeweert W, Colon LT, Govers F, Kamoun S (2001) Ancient diversification of the *Pto* kinase family preceded speciation in *Solanum*. *Molecular Plant–Microbe Interactions* **14**, 996–1005. doi:10.1094/MPMI.2001.14.8.996

Wang GL, Ruan DL, Song WY, Sideris S, Chen L, *et al*. (1998) Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. *The Plant Cell* **10**, 765–779.

Xu Q, Wen XP, Deng XX (2005) Isolation of TIR and nonTIR NBS-LRR resistance gene analogues and identification of molecular markers linked to a powdery mildew resistance locus in chestnut rose (*Rosa roxburghii* Tratt). *Theoretical and Applied Genetics* **111**, 819–830. doi:10.1007/s00122-005-0002-7

Zhou XH, Qian CT, Lou QF, Chen JF (2009) Molecular analysis of introgression lines from *Cucumis hystrix* Chakr. to *C. sativus* L. *Scientia Horticulturae* **119**, 232–235. doi:10.1016/j.scienta.2008.08.011

Manuscript received 2 August 2009, accepted 27 February 2010