

Sequence analysis

The WHY domain mediates the response to desiccation in plants and bacteria

Francesca D. Ciccarelli* and Peer Bork

European Molecular Biology Laboratory, 69012 Heidelberg, Meyerhofstrasse 1, Germany

Received on November 4, 2004; revised on November 28, 2004; accepted on December 12, 2004

Advance Access publication December 14, 2004

ABSTRACT

Motivation: The hypersensitive response (HR) is a process activated by plants after microbial infection. Its main phenotypic effects are both a programmed death of the plant cells near the infection site and a reduction of the microbial proliferation. Although many resistance genes (*R* genes) associated to HR have been identified, very little is known about the molecular mechanisms activated after their expression.

Results: The analysis of the product of one of the *R* genes, the Hin1 protein, led to the identification of a novel domain, which we named WHY because it is detectable in proteins involved in Water stress and Hypersensitive response. The expression of this domain during both biotic infection and response to desiccation points to a molecular machinery common to these two stress conditions. Moreover, its presence in a restricted number of bacteria suggests a possible use for marking plant pathogenicity.

Contact: francesca.ciccarelli@embl.de

Supplementary information: Supplementary data (Figures S1 and S2 and Table S1) and the alignment in clustal format are available at http://www.bork.embl.de/~ciccarelli/WHY_add_data.html

INTRODUCTION

Plants react to microbial infections through a process known as hypersensitive response (HR). The main effects of its activation are a rapid programmed cell death (PCD) of the cells near the infection site, and a reduction of bacterial growth and spread. The molecular machinery through which the HR acts is still largely unknown. However, it has been shown that this defense response is initiated by the secretion of bacterial molecules which are able to induce the expression of plant resistance genes (*R* genes) in a process called 'gene-for-gene response' (Hueck, 1998; Cornelis and Van Gijsegem, 2000). According to the currently accepted view, bacteria start their aggression translocating specific proteins into the plant cell via the Hrp pathway (reviewed in Cornelis and Van Gijsegem, 2000). The presence of these proteins in the plant cytosol stimulates the expression of the plant *R* genes, which in turn initiate the HR. Several families of *R* gene products have been isolated, some of which share common sequence features (Hammond-Kosack and Jones, 1997). The largest group bears a nucleotide binding motif (NBS) and a leucine-rich repeat (LRR), and its members are consequently known as NBS-LRR proteins (Ellis and Jones, 1998).

Other components of HR are poorly characterized, among them the Hin1 genes, which have been isolated during a screen of genes induced after bacterial infection (Gopalan *et al.*, 1996). Hin1 seems to take part in a non-specific pathway activated as a general response of plants to pathogenic infection (Gopalan *et al.*, 1996). Differently from other *R* genes, Hin1 does not share significant sequence similarity to any functionally characterized eukaryotic protein family and its function is unknown.

Here we report the identification of a novel domain that links the Hin1 proteins to the plant family LEA-14, which is expressed under water stress conditions and during late embryogenesis and to a variety of uncharacterized bacterial and archaeal proteins. We named this module WHY for Water stress and Hypersensitive response domain.

The detection of a common domain in two families of proteins both expressed during the plant response to external stresses indicates a common molecular mechanism, involving the activation of similar proteins. Interestingly, the domain is also found in bacteria and archaea, which suggests a possible similar pathway for the stress response, maybe acquired by horizontal transfer.

METHODS

The protein sequence of *Nicotiana tabacum* Hin1 (P93353) was used as a query for standalone PSI-Blast (Altschul *et al.*, 1997) against the protein nr-database. After three iterations, which retrieved only homologs of the Hin1 family in several plants, the first protein from the bacterium *Ralstonia metallidurans* was detected (ZP_00273651, $E = 0.0003$). This is an uncharacterized protein whose homology with the plant Hin1 almost covers the entire sequence. The search converged at the 6th iteration, although the grey zone just after the E -value threshold of 0.001 was clearly enriched in false negatives, such as some plant sequences (Q9C573, Q7X8V0, $E = 0.006$) and one protein from the bacterium *Burkholderia pseudomallei* (CAH37099, $E = 0.009$). Indeed, using these sequences as queries for PSI-Blast, both the Hin1 family and the bacterial sequences were retrieved. The three sequences were then manually added to the alignment and the resulting profile was used to continue the PSI-Blast search. More bacterial sequences were retrieved until the search converged again at the 9th iteration. A hidden Markov model (HMM) profile was built using non-redundant representatives of all the sequences detected with PSI-Blast (seq.id. <80%). The borders of the shared region were defined according to the PSI-Blast pairwise alignments. With the resulting profile a HMM search (Eddy, 1998) against the nr-database was run. This search added two new families to the alignment: the family of plant LEA-14 (Late Embryogenesis Abundant) proteins ($E = 1.2 \times 10^{-13}$ at the 2nd iteration) and some archaeal sequences with a duplication of the domain ($E = 3.2 \times 10^{-5}$ at the 3rd iteration). The HMM search converged at the 4th iteration. Two distinct Pfam family domains are detectable in the Hin1 and the LEA-14 proteins (Hin1 and LEA_2, respectively). In both cases the

*To whom the correspondence should be addressed.

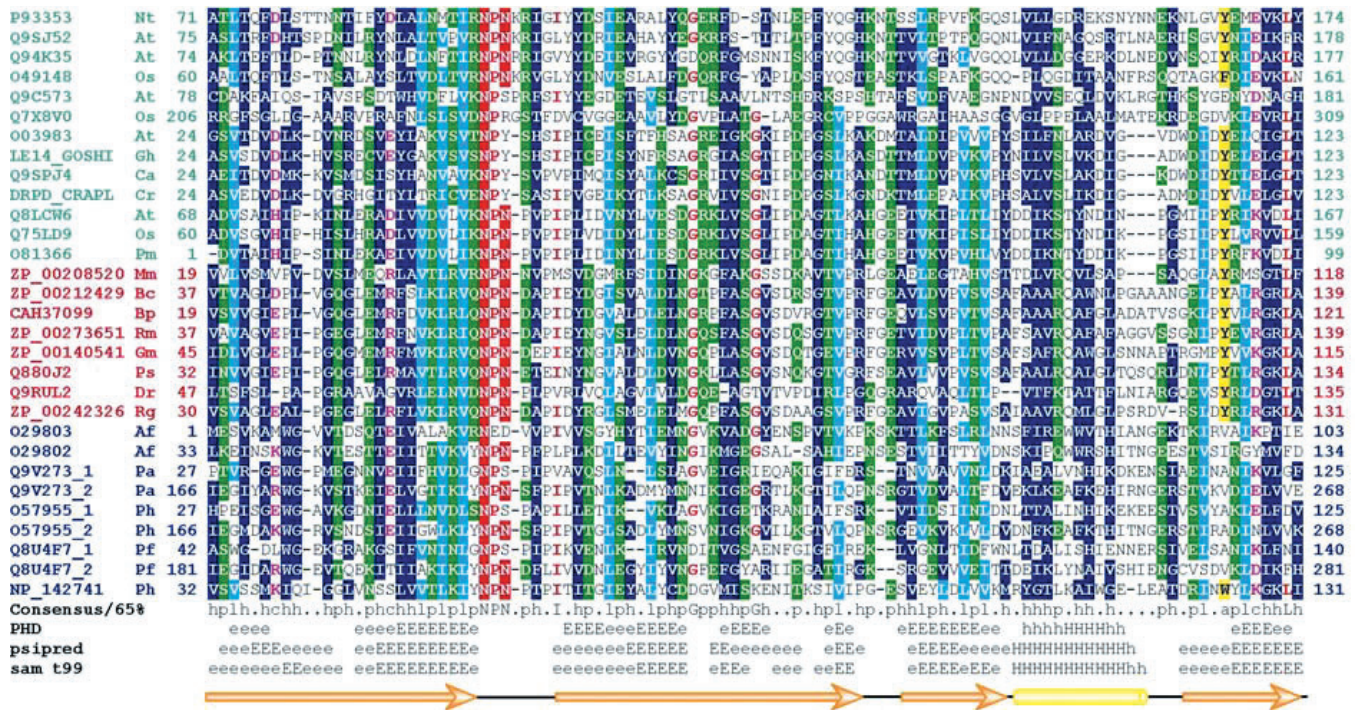


Fig. 1. Multiple sequence alignment of representatives of Why domain. The multiple alignment was built using T-coffee (Notredame *et al.*, 2000) and refined manually. The sequences are indicated using the database accession number followed by the species. The starting and the ending residues are reported before and after the corresponding sequence. The plant sequences are depicted in green, the bacteria in red and the archaea in blue. The consensus in 65% of the sequences is reported below the alignment; h, l, p, c and a indicate hydrophobic, aliphatic, polar, charged and aromatic residues, respectively. Capital letters denote conserved amino acids. Hydrophobic residues are highlighted in dark blue, aliphatic residues in cyan, polar residues in green, charged in pink, aromatic in yellow and other conserved residues in red. Secondary structure predictions using three different methods are reported: for PHD (Rost, 1996) the entire alignment was used as a seed for the prediction, with upper cases indicating predictions with expected average accuracy >82% and lower cases predictions with average accuracy <82%. For Psipred (McGuffin *et al.*, 2000) and sam t99 (Karplus *et al.*, 1998), independent predictions have been performed using representatives of the four families (Hin1; LEA-14; bacterial and archaea proteins) and the consensus is reported as upper cases. Abbreviations: E, strands; H, helices; Af, *Archaeoglobus fulgidus*; At, *Arabidopsis thaliana*; Bc, *Burkholderia cepacia*; Bp, *Burkholderia pseudomallei*; Ca, *Capsicum annuum*; Cr, *Craterostigma plantagineum*; Dr, *Deinococcus radiodurans*; Gh, *Gossypium hirsutum*; Gm, *Geobacter metallireducens*; Mm, *Magnetospirillum magnetotacticus*; Nt, *Nicotiana tabacum*; Os, *Oryza sativa*; Pa, *Pyrococcus abyssis*; Pf, *Pyrococcus furiosus*; Ph, *Pyrococcus horikoshii*; Pm, *Prunus armeniaca*; Pp, *Pseudomonas putida*; Ps, *Pseudomonas syringae*; Rg, *Rubrivivax gelatinosus*; Rm, *Ralstonia metallidurans*.

domains span the whole protein length and extend, as their names suggest, only to members of the two families.

The secondary structure predictions were assigned using PHD (Rost, 1996), PsiPred (McGuffin *et al.*, 2000) and SAM-T99 (Karplus *et al.*, 1998).

The phylogeny was reconstructed using the non-redundant alignment of 44 protein sequences and both minimum evolution (ME) and Bayesian inferences implemented in MEGA (Kumar *et al.*, 2001) and MrBayes (Huelsenbeck and Ronquist, 2001), respectively. For the ME estimation, the bootstrap test of phylogeny was set to 1000 replicates, and the initial tree was calculated using the Neighbor-Joining algorithm. In the Bayesian analysis, the TITJ model for amino acid substitution (Jones *et al.*, 1992) was used and each run was set for 10⁵ generations, with three heated and one cold chains. Convergence was assessed by repeating the analysis for four runs. Comparable tree topologies were obtained using both methods.

RESULTS AND DISCUSSION

The Why domain is an ~100 amino acid long module, with an alternation of hydrophilic and hydrophobic residues and an almost invariable NPN motif at its N-terminus (Fig. 1). The predicted secondary structure is mostly composed of beta strands with a C-terminal α -helix (Fig. 1). No similarity with any of the known

structural folds could be assigned using the metasever 3D-Jury (Ginalski *et al.*, 2003). The modular nature of the Why domain is supported by the fact that it is present in several protein families with different domain architectures and by its duplication in archaea (Fig. 2A).

For two of the families bearing the Why domain, namely the Hin1 and the LEA-14 proteins, some functional information is available. The Hin1 proteins are expressed during the response to bacterial infection and play an uncharacterized role during HR. The LEA-14 proteins are expressed in late embryogenesis but also as part of the plant response to desiccation (Galau *et al.*, 1993). These proteins do not share any sequence similarity with other members of the large LEA family (Maitra and Cushman, 1994). The identification of a common domain between proteins expressed in HR and during desiccation suggests a shared mechanism in plant response to these two stress conditions. Many different observations support this hypothesis. First, genes involved in the response to both biotic and abiotic stresses partly share the same signal transduction pathway (Xing *et al.*, 2001). Second, a direct link between HR and water stress has been recently postulated (Wright and Beattie, 2004). It has been

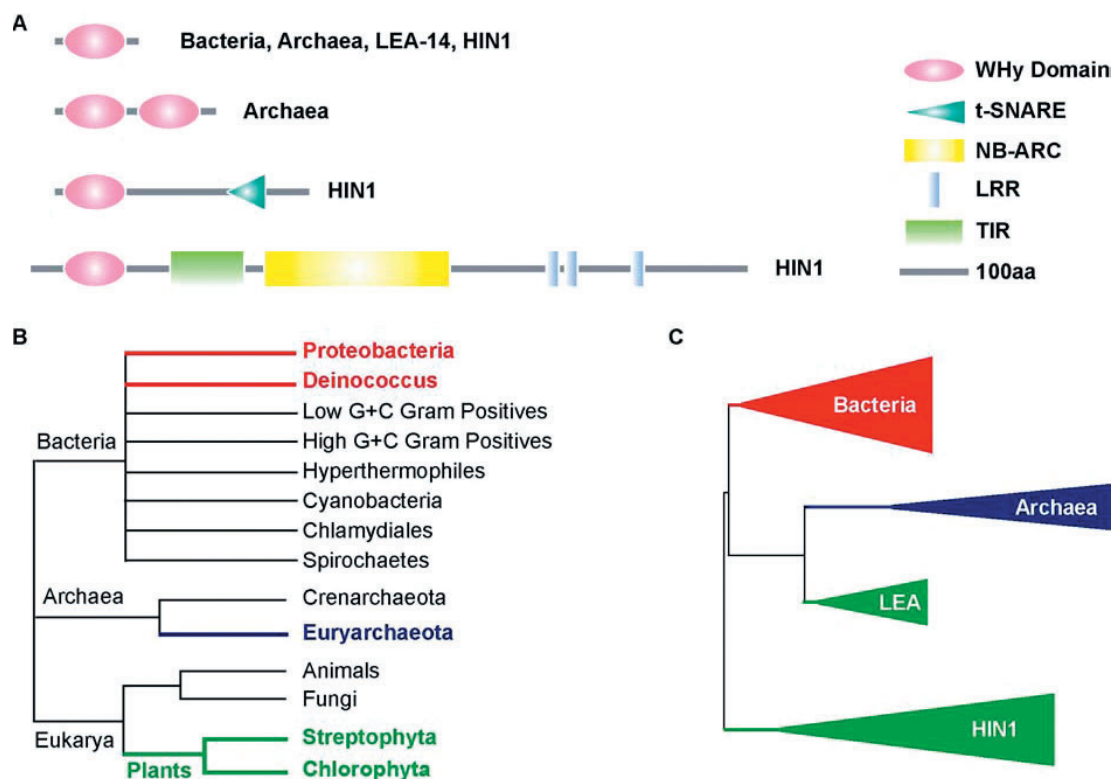


Fig. 2. (A) Domain architecture of the Why containing proteins. The domain compositions and organizations were obtained using the SMART (Letunic *et al.*, 2004) and Pfam (Bateman *et al.*, 2004) resources. NB-ARC, LRR and TIR are domains frequently found in proteins expressed during HR. (C) Phylogenetic relationships between the different Why domains. The tree reports the common topology obtained using Minimum Evolution and Bayesian algorithms. For the actual trees see Supplementary Figure S1. (B) Species distribution of the Why domain. The coloured lines refer to branches bearing the domain (red, bacteria; blue, archaea; green, plants). The black lines indicate absence of the domain in the corresponding branch. The species distribution was obtained by blasting one representative for each family against the non-redundant protein database. For both archaea and bacteria, additional blast searches were performed against the NCBI set of prokaryotic genomes. Note that within the proteobacteria only some species contain the domain, mainly plant pathogens or symbionts. For a complete list of species containing the Why domain see Supplementary Table S1.

observed that the water potential in the bacterial cell reduces during HR, becoming sufficiently low to prevent cell division in many cells of the bacterial colony (Wright and Beattie, 2004). Although experimental proofs to this hypothesis are required, such a link suggests a role of the Hin1 family in the desiccation-like process taking place during the HR and leading to the microbial death.

The Why domain is widespread among plants, but it is absent in fungi and animals. Interestingly, it is also detectable in prokaryotes such as some proteobacteria, *Deinococcus radiodurans* and few euryarchaeota (Fig. 2B and Supplementary Table S1). This patchy distribution can be explained either by a selective retention of the domain, and the function it accomplishes during evolution or by its lateral transfer from plant to prokaryotes. This latter case would be an example of capture of plant defense genes by bacteria either to integrate them in their own defense system or to interfere with the host signal pathway. The tree topology obtained most frequently seems to support this interpretation, pointing to an early appearance of the Why module in plants followed by horizontal transfers from plant to bacteria and archaea (Fig. 2C). A first lateral transfer probably took place between plants and proteobacteria and a second one between plants and archaea after the duplication of the plant domain and the divergence between Hin1 and LEA (Fig. 2C). Several observations

support this scenario. First, the HR process had an early origin in plants (Meyers *et al.*, 1999). Indeed, we could find a hit for Hin1 in the ancient green alga *Chlamydomonas reinhardtii* (AV395132), but no hits for LEA. The fact that the *C.reinhardtii* genome is not complete does not allow for a final statement, but this result at least confirms the early origin of Hin1 in the plant kingdom. Second, the domain distribution among bacteria seems to be restricted to plant pathogens or at least symbionts (Supplementary Table S1). The only bacterial species not belonging to proteobacteria is *D.radiodurans*, which is known to be prone to horizontal transfers (Makarova *et al.*, 2001). Indeed, the position of the *D.radiodurans* sequence in the Why domain tree does not reflect the usual phylogenetic relationships within bacteria (Supplementary Figure S2). Finally, horizontal transfer among domains of life as a mechanism for the acquisition of novel functions has been reported in several cases (Koonin *et al.*, 2001; Ponting *et al.*, 1999; Schultz, 2004). While the horizontal transfer of Why from plants to symbiotic or pathogenic bacteria can be seen as the acquisition of a function putatively related to the cell defense, the lateral acquisition by euryarchaeota is more difficult to explain. In general, horizontal transfer of domains from eukaryotes to archaea does not seem to be a frequent event (Ponting *et al.*, 1999). In addition to that, the lifestyle of hyperthermophilic euryarchaeota does not allow an easy

interpretation of the domain acquisition, except in the context of a general response to stress conditions. Indeed, these organisms live in extreme environments which make the contact with eukaryotes difficult. Nevertheless, examples have been reported of horizontal transfer from eukaryotes to hyperthermophilic microbes, as in the case of the TrpRS gene (Wolf *et al.*, 1999).

Although horizontal transfer seems the most suitable explanation, we cannot exclude an alternative interpretation for the evolution of the Why domain. It implies its appearance in the bacterial kingdom and a selective retention only in plants, some proteobacteria and few euryarchaeota. This scenario requires several gene losses in bacteria and in the common ancestor between animal and fungi (Fig. 2B), and one horizontal gene transfer from archaea to plants (Fig. 2C). In this case, the patchy distribution of the Why domain would be the result of a selective retention of defense mechanisms during evolution.

In either scenario, the actual presence of the domain in a restricted number of bacteria makes it suitable as a marker for detecting plant pathogenicity and to eventually develop new bactericidal strategies.

ACKNOWLEDGEMENTS

The authors are grateful to David Torrents and Christian von Mering for their critical reading of the manuscript.

REFERENCES

- Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths-Jones,S., Khanna,A., Marshall,M., Moxon,S., Sonnhammer,E.L.L. *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–D141.
- Cornelis,G.R. and Van Gijsegem,F. (2000) Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.*, **54**, 735–774.
- Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
- Ellis,J. and Jones,D. (1998) Structure and function of proteins controlling strain-specific pathogen resistance in plants. *Curr. Opin. Plant Biol.*, **1**, 288–293.
- Galau,G.A., Wang,H.Y. and Hughes,D.W. (1993) Cotton *Lea5* and *Lea14* encode atypical late embryogenesis-abundant proteins. *Plant Physiol.*, **101**, 695–696.
- Ginalski,K., Eloffson,A., Fischer,D. and Rychlewski,L. (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*, **19**, 1015–1018.
- Gopalan,S., Wei,W. and He,S.Y. (1996) *hrp* gene-dependent induction of *hin1*: a plant gene activated rapidly by both harpins and the *avrPto* gene-mediated signal. *Plant J.*, **10**, 591–600.
- Hammond-Kosack,K.E. and Jones,J.D. (1997) Plant disease resistance genes. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **48**, 575–607.
- Hueck,C.J. (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.*, **62**, 379–433.
- Huelsenbeck,J.P. and Ronquist,F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.
- Jones,D.T., Taylor,W.R. and Thornton,J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.*, **8**, 275–282.
- Karplus,K., Barrett,C. and Hughey,R. (1998) Hidden Markov models for detecting remote protein homologies. *Bioinformatics*, **14**, 846–856.
- Koonin,E.V., Makarova,K.S. and Aravind,L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.*, **55**, 709–742.
- Kumar,S., Tamura,K., Jakobsen,I.B. and Nei,M. (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*, **17**, 1244–1245.
- Letunic,I., Copley,R.R., Schmidt,S., Ciccarelli,F.D., Doerks,T., Schultz,J., Ponting,C.P. and Bork,P. (2004) SMART 4.0: towards genomic data integration. *Nucleic Acids Res.*, **32** (Database issue), D142–D144.
- Maitra,N. and Cushman,J.C. (1994) Isolation and characterization of a drought-induced soybean cDNA encoding a D95 family late-embryogenesis-abundant protein. *Plant Physiol.*, **106**, 805–806.
- Makarova,K.S., Aravind,L., Wolf,Y.I., Tatusov,R.L., Minton,K.W., Koonin,E.V. and Daly,M.J. (2001) Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. *Microbiol. Mol. Biol. Rev.*, **65**, 44–79.
- McGuffin,L.J., Bryson,K. and Jones,D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404–405.
- Meyers,B.C., Dickerman,A.W., Michelmore,R.W., Sivaramakrishnan,S., Sobral,B.W. and Young,N.D. (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.*, **20**, 317–332.
- Notredame,C., Higgins,D.G., Heringa,J. (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.
- Ponting,C.P., Aravind,L., Schultz,J., Bork,P. and Koonin,E.V. (1999) Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J. Mol. Biol.*, **289**, 729–745.
- Rost,B. (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol.*, **266**, 525–539.
- Schultz,J. (2004) HTTM, a horizontally transferred transmembrane domain. *Trends Biochem. Sci.*, **29**, 4–7.
- Wolf,Y.I., Aravind,L., Grishin,N.V. and Koonin,E.V. (1999) Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res.*, **9**, 689–710.
- Wright,C.A. and Beattie,G.A. (2004) *Pseudomonas syringae* pv. tomato cells encounter inhibitory levels of water stress during the hypersensitive response of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci USA*, **101**, 3269–3274.
- Xing,T., Malik,K., Martin,T. and Miki,B.L. (2001) Activation of tomato PR and wound-related genes by a mutagenized tomato MAP kinase through divergent pathways. *Plant Mol. Biol.*, **46**, 109–120.