Epistatic selection and coadaptation in the Prf resistance complex of wild tomato

Lukasz Grzeskowiak [a,b], Wolfgang Stephan [a], Laura E. Rose [a,c,*]

[a] Section of Evolutionary Biology, Ludwig Maximilian University Munich, Grosshadernerstr. 2, 82152 Planegg, Germany
[b] Center for Research and Innovation, Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele all'Adige, Italy
[c] Institute of Population Genetics, Heinrich Heine University Duesseldorf, Universitaetsstr. 1, 40225 Duesseldorf, Germany

ABSTRACT

Natural selection imposed by pathogens is a strong and pervasive evolutionary force structuring genetic diversity within their hosts' genomes and populations. As a model system for understanding the genomic impact of host–parasite coevolution, we have been studying the evolutionary dynamics of disease resistance genes in wild relatives of the cultivated tomato species. In this study, we investigated the sequence variation and evolutionary history of three linked genes involved in pathogen resistance in populations of *Solanum peruvianum* (*Pto*, *Fen*, and *Prf*). These genes encode proteins, which form a multimeric complex and together activate defense responses. We used standard linkage disequilibrium, as well as partitioning of linkage disequilibrium components across populations and correlated substitution analysis to identify amino acid positions that are candidates for coevolving sites between Pto/Fen and Prf. These candidates were mapped onto known and predicted structures of Pto, Fen and Prf to visualize putative coevolving regions between proteins. We discuss the functional significance of these coevolving pairs in the context of what is known from previous structure–function studies of Pto, Fen and Prf.

1. Introduction

Protein sequence variation at disease resistance loci in natural plant populations is a well-documented observation. The presence of this polymorphism, maintained by on-going coevolutionary interactions with pathogens, should allow for effective crop protection and has contributed to many successful breeding programs. However, due to the lengthy and costly process of generating cultivars with novel resistance phenotypes, great effort is placed on extending the lifetime of the plant cultivars currently available. To what degree this is possible and how this can be achieved is the focus of this special issue. At the heart of this challenge is to predict and/or identify which resistance genes pose an insurmountable hurdle to the pathogen and therefore will provide effective protection over multiple growing seasons. If such genes exist (which is perhaps unlikely due to the effectiveness of natural selection and evolutionary potential in nearly all pathogen populations), do these genes share common features that could act as targets in future breeding programs? A parallel line of investigation, which

goes hand-in-hand with the first objective, is to determine the best deployment strategies to extend the lifetime of the resistance genes already available to breeders. Our work on this question emerges from prior observations about the genetic and phenotypic variation for disease resistance present in natural plant populations (Rose et al., 2004, 2005, 2007, 2011, 2012; Hoerger et al., 2012). Natural selection imposed by pathogens is exquisitely tuned to differences among plants in disease resistance and therefore, observations in natural populations may help to inspire more successful plant protection strategies.

In this paper, we describe the sequence variation and evolutionary history of three linked genes involved in pathogen resistance in wild tomato (*Pto*, *Fen*, and *Prf*). These genes encode proteins, which form a multimeric complex and together activate defense responses (Gutierrez et al., 2010; Ntoukakis et al., 2013). We are interested how the genetic linkage and the functional interaction combine to influence the sequence evolution of these genes. Since these genes reside in a small region of the tomato genome (60 kb), nearly all aspects of their evolutionary history are shared. A focus of this research is to determine to what degree these genes are coadapted to one another and how their roles in pathogen recognition mutually influence the evolution of neighboring and functionally linked genes. The presence and strength of epistatic selection in resistance gene complexes has broader implications because

* Corresponding author. Address: Institute of Population Genetics, Heinrich Heine University – Duesseldorf, Universitaetsstr. 1, 40225 Duesseldorf, Germany. Tel.: +49 211 81 13406.
  E-mail address: Laura.Rose@hhu.de (L.E. Rose).

strong epistatic interactions can serve as a road-block to breeding programs, leading to genetic incompatibilities due to the disruption of R-gene functional units (reviewed in Bomblies, 2009).

The *Pto* resistance gene belongs to a small multigene family of five to six family members in the *Lycopersicon* clade (Martin et al., 1993), however functions have not been ascribed to all of these genes. The entire 60 kb region of chromosome 5 containing the *Pto* gene family has been sequenced from a susceptible *Solanum lycopersicum* cultivar and a resistant cultivar containing the *Pto* locus introgressed from the sister species *Solanum pimpinellifolium* (GenBank accessions AF220602 and AF220603). The two haplotypes share five orthologous, clustered genes (*Fen, Pth2, Pth3, Pth4* and *Pth5*).

Pto confers resistance to strains of *Pseudomonas syringae pv. tomato* expressing either AvrPto or AvrPtoB. It was the first race-specific R-gene to be isolated (Martin et al., 1993). This small gene without introns and the open reading frame (ORF) of 963 nucleotides encodes a functional serine/threonine kinase capable of autophosphorylation (Loh and Martin, 1995). The current model for Pto activation involves Pto binding to the pathogen ligand in the plant cell and a change in protein conformation, induced through this physical interaction. The stabilization of the Pto molecule in the proper conformation is dependent on Pto kinase activity. Next, the activated Pto protein transduces the signal, which is sensed by Prf to activate downstream plant immune responses. This includes the synthesis of antimicrobial compounds and results in programmed cell death at the site of infection (Rathjen et al., 1999; Sessa and Martin, 2000; Wu et al., 2004; Mucyn et al., 2006; Xing et al., 2007).

Fen, one of the *Pto* family members, is a functional serine/threonine kinase and confers sensitivity to the insecticide fenthion (Martin et al., 1994; Chang et al., 2002). The Fen protein shares 80% sequence identity with Pto, but does not confer AvrPto-dependent resistance (Scofield et al., 1996; Jia et al., 1997; Frederick et al., 1998). However, this paralog can recognize and activate defense responses to variants of AvrPtoB effector lacking E3 ubiquitin ligase activity (Rosebrock et al., 2007). Nonetheless, the wild type form of AvrPtoB ubiquitinates certain Fen alleles, which leads to their degradation in the plant cell. This suggests that the genes in the *Pto* cluster paralogs have experienced a complex history of host-pathogen coevolution.

Both Pto and Fen proteins do not act alone, but require a second protein, Prf, for the activation of disease resistance. *Prf* is a large gene embedded within the *Pto* gene cluster, although it is phylogenetically unrelated to *Pto* and its paralogs. In *S. pimpinellifolium* Rio Grande 76R, the 3′ end of this gene is located about 500 bp from the ORF of *Fen* and 24 kb from the ORF of *Pto*. The complete transcribed region of *Prf* is almost 11 kb and contains five introns. The protein coding region is 5.5 kb long. The resultant Prf protein is a large molecule (209.7 kDa) and contains NBS-LRR motifs, common to many other plant R-proteins (Salmeron et al., 1996).

Both of the kinases, Pto and Fen, functionally interact with the same N-terminal portion of Prf (Mucyn et al., 2006, 2009; Ntoukakis et al., 2009). Silencing of Prf prevents signaling by Fen or Pto, indicating that Prf acts epistatically to Fen and Pto. Recognition of *avrPtoB* by alleles of Pto from tomato expressed in *Nicotiana benthamiana* was only possible if the tomato allele of Prf was also co-expressed. The *N. benthamiana* allele of Prf could not complement this phenotype. This indicates that specific pairs of interacting partners are required for the full range of resistance (Balmuth and Rathjen, 2007; Mucyn et al., 2009). Thus, not only are *Pto* and *Fen* physically linked with *Prf*, which may indirectly affect their evolutionary history, but the functional protein interaction may require coadaptation between these molecules.

The molecular set up of this resistance cluster is analogous to a socket-wrench set, with Prf serving as the wrench handle or drive and Pto and its paralogs representing the individual sockets, adapted to recognizing different pathogen ligands. The physical linkage of Prf and the Pto cluster means that the entire set can be bred as a unit. This has advantages because the coadapted unit is not disrupted during the breeding process; e.g. the functional components do not become separated. As demonstrated in different studies, disrupting an R-gene functional unit can lead to unregulated cell death and hybrid incompatibilities (reviewed in Bomblies, 2009). As such, this small cluster of genes resembles the genetic cassettes constructed by modern breeders. The main difference is that the combination of specificities in a single plant genotype is guided by natural evolutionary history – including non-selective influences such as drift and demography – rather than by breeders and molecular geneticists. As a consequence, the number and breadth of specificities (and hence the range of pathogen protection) may be more limited than what is desired from a breeder's perspective, but the advantage of this "evolutionary genetic cassette" is that the individual components are guaranteed to function with one another.

What is the likelihood that these three genes represent a coadapted gene complex? Previous studies have shown that activity of tomato Fen is suppressed when combined with the allele of Prf from *N. benthamiana* (Mucyn et al., 2009). Likewise, *Pto* is subject to a mixture of balancing and purifying selection (Rose et al., 2007, 2011). This suggests that some types of *Pto* and *Fen* may function best with certain types of *Prf*. In this study we addressed these following questions: (1) Do the tightly linked genes *Pto*, *Fen* and *Prf* evolve in a correlated fashion? (2) Does epistatic selection operate in the Pto signaling pathway? (3) Does the maintenance of allelic variation at the *Pto* and *Fen* genes lead to the maintenance of allelic diversity at the *Prf* locus? We identified amino acid positions that are candidates for coevolving sites between Pto/Fen and Prf using standard linkage disequilibrium, as well as partitioning of linkage disequilibrium components across populations and correlated substitution analysis. These candidates were mapped onto known and predicted structures of Pto, Fen and Prf to visualize putative coevolving regions between proteins. Functional significance of these coevolving pairs is discussed in the context of what is known from previous structure–function studies of Pto, Fen and Prf.

2. Materials and methods

2.1. Plant materials

For this study, we collected multiple individuals of *Solanum peruvianum*, a species closely related to the cultivated tomato and endemic to the western coast of South America. This species is widespread and often occurs in large stands in central and southern Peru and northern Chile (reviewed in Chetelat et al., 2009). Individuals of *S. peruvianum* are diploid, obligate outcrossing, short-lived perennials and census population sizes range from a few plants to several hundred (tgrc.ucdavis.edu).

We sampled this species from three different geographical locations: (1) Canta (Central Peru, 11°31′S, 76°41′W; 2050 m altitude) (2) Nazca (Southern Peru, 14°51′S, 74°44′W; 2130 m altitude), and (3) Tarapaca (Northern Chile, 18°33′S, 70°09′W: 400 m altitude). The samples from Nazca and Canta, gathered in May 2004 by T. Staedler and T. Marczewski, are described in Staedler et al. (2008). Six plants were collected per population. DNA was extracted from leaf material using the DNeasy Plant Mini Kit (Qiagen GmbH, Hilden, Germany). Seeds from the Tarapaca population were collected by C. Rick in April 1986 and stored at the Tomato Genetics Resource Center (TGRC) at the University of California, Davis (tgrc.ucdavis.edu; accession LA2744). This accession,

exceedingly variable for many traits, belongs to the TGRC collection. In 1996, seeds from ten different plants were germinated and grown under standard greenhouse conditions in Davis, CA. Genomic DNA was isolated using the CTAB method (Doyle and Doyle, 1987) from 2 g of leaf tissue collected from each plant. The DNA was resuspended in 300–1000 μl TE, depending on yield. An individual of *Solanum habrochaites* from Ancash, Peru (TGRC accession LA 1775) was used for outgroup comparisons. Plant growth conditions and DNA extraction were identical as used for the Tarapaca population.

2.2. Gene amplification and sequencing

We followed similar methods as described in Rose et al. (2011) to amplify and sequence alleles of the *Pto*, *Prf* and *Fen* genes. For each gene, the entire coding region was amplified using a proof-reading polymerase, either Pfu (Stratagene, La Jolla, CA) or Phusion (Finnzymes, Espoo, Finland). PCR fragments were cloned into pCR-Blunt or Zero Blunt TOPO (Invitrogen, Carlsbad, CA). Direct sequencing of PCR products and sequencing of minipreped plasmid DNA from clones were conducted in parallel for each gene on an ABI 3730 DNA analyzer (Applied Biosystems/Hitachi, Foster City, CA). Multiple clones per gene, per individual were sequenced and ambiguous positions were compared to the direct sequences from the original PCR products. When necessary, independent rounds of PCRs, cloning and sequencing were conducted to resolve ambiguities.

2.2.1. Specific amplification and cloning strategy for Pto

The primers SSP17 and JCP32 were used to initially amplify alleles of *Pto* (Supplemental Table 1). These primers also amplify to a lesser degree two paralogs of *Pto*, namely *Pth3* and *Pth5*. The restriction enzyme BstXI was used to differentiate between clones containing *Pto* versus its paralogs. This enzyme specifically digests alleles of *Pth3* and *Pth5*, but not *Pto*. To circumvent non-specific amplification of *Pto* alleles and to facilitate direct sequencing of *Pto* for confirmation of homozygosity/heterozygosity respectively, two *Pto*-specific primers in the upstream region of *Pto* were developed. These primers, FromPth5A and FromPth5B, were used in combination with the JCP32 primer, which anneals at the 3′ end of *Pto*.

2.2.2. Specific amplification and cloning strategy for Fen

A similar strategy as used for *Pto* was employed for sequencing of *Fen* alleles (Supplemental Table 1). The primers SSP17 and SSP19 were used initially to amplify alleles of *Fen*. Cloning of these PCR products revealed that these primers did not specifically amplify alleles of *Fen*. Ultimately two additional *Fen*-specific primers were designed, one upstream of *Fen* and one downstream of *Fen*, based upon the GenBank sequence AF220602 of this region from the *S. lycopersicum* cv. Rio Grande 76R haplotype. These two intergenic primers, FenFor and FenRev, were used in combination and with SSP19 or SSP17, respectively.

2.2.3. Specific amplification and cloning strategy for Prf

Prf is a large gene (5587 bp from start to stop codon), therefore it was divided into two overlapping parts for PCR and these were sequenced separately. The first part of *Prf* is well-known for being difficult to clone, so a direct sequencing strategy involving allele-specific sequencing primers was used to resolve phase. Both direct sequencing of PCR products and cloning were employed to generate the data for the second part of the gene (approximately 58% of *Prf*). A large number of primers were designed for sequencing and allele-specific amplification (Supplemental Table 1). For individuals from Nazca and Canta, the first 1701 bp of *Prf* was amplified. These PCR products were sequenced and phase was inferred using

the ELB algorithm implemented in Arlequin (Excoffier et al., 2003; Excoffier and Lischer, 2010).

2.2.4. Specific amplification and cloning strategy for reference genes

The sequences of eight nuclear loci served as the reference gene set (Arunyawat et al., 2007). These loci and the three R-genes were sequenced from the same individuals (Supplementary Table 2). The reference loci were developed from cDNA markers used in the genetic map of tomato (Tanksley et al., 1992).

2.3. DNA sequence analyses

The standard summary statistics (including π, haplotype diversity, Tajima's D, Z_{nS}, F_{ST}) were calculated using DnaSP (Librado and Rozas, 2009). The population recombination parameter ρ was estimated using composite likelihood method of Hudson (2001), implemented in the LDhat package (McVean et al., 2002). The expected decay of linkage disequilibrium within resistance genes was modeled using the equation given by Hill and Weir (1988) and fitted to the data in R statistical package (r-project.org). LD between pairs of sites of *Pto* and *Prf* or *Fen* and *Prf* was calculated using the composite-disequilibrium R^2 statistic (Zaykin et al., 2008). This method allows for greater than two alleles per site and can be applied to genotypic data (i.e. unphased data). This composite LD can be interpreted as the total correlation between a pair of loci (Cockerham and Weir, 1977; Weir, 1979, 1996). It is estimated directly from genotypic counts and is not biased by inbreeding or higher order departures from random assortment (i.e. Hardy–Weinberg equilibrium). The program MCLD was used to calculate both the approximate and exact (permutational, based on 50,000 permutations) p-values for R^2 tests (Zaykin et al., 2008).

Individuals carrying pseudogenes were excluded from these analyses. Only two pseudogenes were observed among the 44 alleles sequenced from these three genes. Both pseudogenes were found in the Tarapaca population – one in *Pto* from individual 7232 and one in *Fen* from individual 7236. Therefore, a total of nine genotypes for each gene combination (*Pto* versus *Prf* or *Fen* versus *Prf*) were analyzed from Tarapaca, six genotypes from Nazca and six genotypes from Canta. Singleton polymorphisms were excluded from association analyses.

Departures from linkage equilibrium may be caused by natural selection or stochastic (neutral) processes. We applied a method proposed by Ohta to partition the variance components of linkage disequilibrium to determine what fraction of the observed associations could be attributed to epistatic selection between these genes (Ohta, 1982a,b). This partitioning is similar to Wright's F-statistics describing the partitioning of deviations from Hardy–Weinberg equilibrium frequencies into F_{ST} (the average deviation attributable to differences in allele frequency among populations) and F_{IS} (the average deviation within populations) (Wright, 1940).

Ohta's D-statistics consists of the within, the between subpopulations and total components of LD in a subdivided population: D_{IS}, D_{ST}, D'_{IS}, D'_{ST}, D_{IT}. The subscript "IS" stands for "individuals within subpopulations", the "ST" for "subpopulations within the total population" and the "IT" for "individuals within the total population". Thus, D_{IS} is the average LD measured within individuals within subpopulations, D_{ST} is the contribution to the overall LD caused by differences in allele frequencies among subpopulations, D'_{IS} is the variance in the observed frequency that a certain nucleotide combination appears within individuals within subpopulations, D'_{ST} is the variance of LD in the total population, while D_{IT} is the same measure made within all individuals irrespective of the subpopulation they come from.

Ohta's D-statistics discriminate between different sources of LD (Ohta, 1982a,b). The use of two inequalities based on LD variance components allows us to characterize three patterns (Whittam

et al., 1983; Black and Krafsur, 1985). Three patterns correspond to three different types of LD: (1) If $D_{IS} < D_{ST}$ and $D'_{IS} > D'_{ST}$, LD is considered to be non-systematic (i.e. LD is caused by random genetic drift and limited migration among subpopulations); (2) if $D_{IS} > D_{ST}$ and $D'_{IS} < D'_{ST}$, LD is considered to be systematic (epistatic selection) and (3) if $D_{IS} > D_{ST}$ and $D'_{IS} > D'_{ST}$, LD is considered to be unequal systematic (e.g. when epistatic selection does not operate across all subpopulations). Ohta's D-statistics were calculated using the Linkdos program (Black and Krafsur, 1985; Garnier-Gere and Dillmann, 1992).

While linkage disequilibrium analyses from pairs of genes across multiple populations allow us to determine the degree to which epistatic selection has shaped the evolution of these genes, F_{ST} analysis of these same genes across populations allows us to identify signatures of local adaptation or balancing selection operating at these loci individually. Loci showing significantly greater (or lesser) allelic differentiation than the genome wide average can be identified using the method of Beaumont and Nichols (1996). These loci are candidates for sites experiencing either strong directional selection (e.g. local adaptation, observed $F_{ST} >$ expected) or balancing selection (observed $F_{ST} <$ expected). We implemented this method in the program FDIST2 (Beaumont and Nichols, 1996; Flint et al., 1999), which calculates the F_{ST} estimator of Weir and Cockerham (1984) for each gene in the sample. Coalescent simulations were then performed to generate data sets with a distribution of F_{ST} close to the empirical distribution. Based on this simulated distribution it is possible to calculate quantiles for outlier SNP loci. First, we analyzed the eight reference genes from these populations to determine the appropriate mean F_{ST} for creating the expected distribution of F_{ST} and heterozygosity against which to test our resistance genes (Pto, Fen and Prf). Following this first pass, SNP loci falling outside of the 95% confidence intervals were discarded and the analysis was run again to calculate the mean "neutral" F_{ST}. This procedure is recommended, since it lowers the bias on the estimation of the mean neutral F_{ST} by removing the most extreme loci from the estimation (Beaumont and Nichols, 1996). Simulations were then run using 30,000 iterations, assuming 100 populations, 12 alleles per sample and an infinite-site mutation model. Simulated F_{ST} values were plotted against heterozygosity to yield a distribution for F_{ST} under a neutral model. Polymorphic sites with F_{ST} values for a given level of heterozygosity that fell outside the 0.95 quantile were considered candidates for directional positive selection. Conversely, loci with F_{ST} values that fell below the 0.05 quantile of the distribution were considered candidates for balancing selection. In addition, to confirm the robustness of the above frequentist method of moments, we used the Bayesian-based F_{ST} outlier detection method of Foll and Gaggiotti (2008), implemented in the BayeScan software. This method calculates the locus-population-specific F_{ST} coefficients (which are different from observed F_{ST} values in FDIST2) and the posterior probability that a locus is subject to selection as measured by the decimal logarithm of the Bayes factor. The Bayes factor provides a scale of evidence in favor of a selection model versus the standard neutral model. To calculate these values we used in total 600 non-singleton SNP loci, both non-synonymous and synonymous, from three R-genes and eight reference genes.

2.4. Protein sequence analyses

We used the method developed to identify coevolving residues between protein domains to determine which residues in Pto or Fen were likely to be coevolving with Prf. The method, called ELSC (explicit likelihood of subset covariation), is based on alignment perturbation and also evaluates the correlation between sites (Dekker et al., 2004). Here however, the full joint alignment of the two proteins is broken into subalignments based on a per-site

inspection. For example, a given site polymorphic in Pto (denoted here as site A) is chosen and the alignment is broken into two sub-alignments – the one subalignment containing all haplotypes linked to the major allele (the most prevalent amino acid polymorphism) at site A and the other subalignment containing the haplotypes associated with the minor allele or alleles at this site. Then the distribution of amino acids at a polymorphic site in Prf (denoted here as site B) in the subalignment containing the major allele of Pto at site A is compared to the distribution of all amino acids at site B. A normalized statistic that gives the probability of drawing at random the composition observed in the subalignment relative to the probability of drawing the most likely composition is then calculated. The final score is the negative natural log of this ratio of likelihoods. High values (>3) are indicative of sites that show correlated evolution. The algorithm was executed in the package provided from www.afodor.net. For both analyses, a multiple sequence alignment of Pto, Fen and Prf from the three populations of S. peruvianum was used. The two pseudogene sequences of Pto and Fen were excluded. Gametic phase between Pto and Prf or Fen and Prf was inferred using the ELB algorithm implemented in Arlequin (Excoffier et al., 2003; Excoffier and Lischer, 2010).

2.5. Tertiary structures of Pto, Fen and Prf

The Pto crystal structure was determined by Xing et al. (2007; PDB 2qkw), but the native tertiary structures of Fen and Prf have not yet been experimentally determined. Therefore, for these two proteins, we used I-TASSER (iterative threading assembly refinement algorithm; Zhang, 2008) to predict the structures of Fen and Prf. This method first searches for template proteins of similar folds from the PDB (protein database) library. Then the continuous fragments from the PDB templates are reassembled into full-length models by replica-exchange Monte Carlo simulations and the unaligned regions (mainly loops) are built by ab initio modeling. When no appropriate template is identified, I-TASSER builds the entire structure by ab initio modeling. Subsequently, fragment assembly simulation is performed to refine the global topology of the protein structure. Final full-atomic models are obtained by optimization of the hydrogen-bonding network. Due its high sequence similarity and evolutionary relatedness, Fen (GenBank accession AAF76307) was modeled by threading onto the crystal structure of Pto. For Prf (GenBank accession AAF76312), only the first 1500 residues were analyzed, including the region, which has been described to interact with Pto and Fen. In the modeling process several parent proteins with functions essential in disease resistance were used: (1) a protein phosphatase – scaffold protein from human (PDB 1b3u:A), (2) oxidoreductase from Neurospora crassa involved in response to oxidative stress (PDB 1sy7:A), (3) clathrin adaptor protein core from mouse (PDB 1w63:A) involved in binding and intracellular protein transport, (4) importin beta subunit from human (PDB 1qgr:A), which transfers proteins into nucleus, (5) beta-catenin from human (PDB 1jdh:A) that functions in the transcription process, (6) TIP20 protein from human (TATA binding protein that enhances transcription, part of multisubunit cullin-dependent ubiquitin ligase), which is involved in protein ubiquitination, negative regulation of catalytic activity and positive regulation of transcriptional complex assembly (PDB 1u6g:C), (7) apoptosis regulatory complex CED-4/CED-9 from nematode (PDB 2a5y) and (8) apoptotic protease activating factor Apaf-1 from human (PDB 1z6t:A). Residues identified as either coevolving between these proteins or under natural selection are highlighted on these protein structures using the program PyMOL (DeLano, 2008). Amino acid positions are numbered according to a reference protein sequence from S. pimpinellifolium (Pto, GenBank accession AAF76306; Fen, GenBank accession AAF76307) and S. lycopersicum (Prf, GenBank accession AAF76312).

3. Results

3.1. Nucleotide diversity

The three R-genes *Pto*, *Fen* and *Prf* are well known molecules in the Pto signaling pathway and physical interaction between the Pto/Fen kinases and Prf has been extensively studied (Mucyn et al., 2006, 2009; Chen et al., 2008). To analyze the coevolutionary relationship between these molecules, we sequenced in total 44 alleles of *Pto*, *Fen* and *Prf* from 22 individuals across three populations of *S. peruvianum* (Supplementary Fig. 1). Average sequence polymorphism across these three populations at synonymous sites in these three R-genes from the *Pto* cluster is half that observed at the eight reference loci from the same individuals (1.56% at *Pto*, *Fen* and *Prf* versus 2.95% at the reference loci; Table 1). In contrast, non-synonymous polymorphism is more than three and half times higher at the resistance gene loci as compared to the eight reference loci (1.04% at *Pto*, *Fen* and *Prf* versus 0.29% at the reference loci). As a result, the ratio of non-synonymous to synonymous polymorphism is more than six times higher for the resistance genes compared to the reference loci. The sequence variation of R-gene *Pto* and the functional consequences of this variation within and between populations of seven tomato species were previously characterized (Rose et al., 2005, 2007). Those studies also reported a significantly higher level of non-synonymous variation at *Pto* in *S. peruvianum* compared to a set of reference genes.

Table 1
Summary statistics for R-genes *Pto, Fen, Prf* and eight reference genes within and across three populations of *S. peruvianum*.

Locus	Population	Length[a]	n[b]	S[c]	Hd[d]	π_{syn}[e]	π_{non}[f]	π_{non}/π_{syn}	D[g]	Z_{nS}[h]	ρ[i]
Pto	Total	960	43	68	0.982	1.808	1.437	0.79	−0.262	0.097	0.070
	Tarapaca		19	43	0.912	2.038	1.278	0.63	0.350	0.235	0.027
	Nazca		12	50	1.000	1.971	1.784	0.91	0.257	0.255	0.035
	Canta		12	34	0.985	1.021	1.124	1.10	−0.293	0.195	0.046
Fen	Total	957	43	76	0.984	2.018	0.919	0.46	−1.320	0.081	0.060
	Tarapaca		19	34	0.942	1.560	0.676	0.43	−0.476	0.173	0.031
	Nazca		12	25	1.000	1.558	0.506	0.32	−0.631	0.178	0.047
	Canta		12	53	0.955	2.277	1.545	0.68	−0.327	0.345	0.027
Prf	Total	1701	44	75	0.979	0.842	0.764	0.91	−0.818	0.062	0.017
	Tarapaca		20	49	0.947	1.114	0.705	0.63	−0.071	0.132	0.011
	Nazca		12	32	1.000	0.421	0.702	1.67	0.114	0.194	0.021
	Canta		12	26	0.848	0.331	0.487	1.47	−0.481	0.410	0.002
CT066	Total	1346	34	66	0.966	3.392	0.217	0.06	−0.616	0.091	0.024
	Tarapaca		10	40	0.933	3.369	0.204	0.06	−0.307	0.281	0.010
	Nazca		12	25	0.773	2.431	0.141	0.06	0.650	0.453	0.001
	Canta		12	43	0.985	2.880	0.177	0.06	−0.930	0.145	0.043
CT093	Total	1389	34	60	0.991	1.473	0.195	0.13	−1.708	0.105	0.012
	Tarapaca		10	23	0.956	1.763	0.105	0.06	−0.141	0.242	0.013
	Nazca		12	24	0.955	1.208	0.169	0.14	−0.598	0.371	0.004
	Canta		12	31	1.000	1.111	0.241	0.22	−1.001	0.187	0.023
CT166	Total	1265	32	114	0.986	0.894	0.069	0.08	−1.622	0.097	0.013
	Tarapaca		8	42	0.893	0.548	0.000	0.00	−0.514	0.475	0.002
	Nazca		12	45	0.970	1.019	0.000	0.00	−1.067	0.238	0.014
	Canta		12	75	0.970	1.131	0.164	0.15	−0.753	0.264	0.005
CT179	Total	899	34	91	0.991	4.355	0.082	0.02	−1.112	0.057	0.097
	Tarapaca		10	29	0.911	3.456	0.000	0.00	−0.003	0.284	0.011
	Nazca		12	49	1.000	3.751	0.117	0.03	−0.368	0.137	0.102
	Canta		12	56	0.985	4.604	0.117	0.03	−0.400	0.156	0.055
CT198	Total	693	30	101	0.986	5.439	0.364	0.07	−0.732	0.088	0.060
	Tarapaca		10	62	0.911	5.648	0.182	0.03	0.070	0.342	0.010
	Nazca		10	50	0.978	4.312	0.364	0.08	−0.050	0.223	0.041
	Canta		10	57	0.978	5.891	0.519	0.09	−0.050	0.294	0.015
CT208	Total	1069	32	83	0.938	1.720	0.018	0.01	−0.951	0.161	0.002
	Tarapaca		8	41	0.893	0.993	0.074	0.07	−0.227	0.745	0.000
	Nazca		12	47	0.803	1.419	0.000	0.00	−0.540	0.408	0.000
	Canta		12	43	0.773	1.784	0.000	0.00	−0.509	0.552	0.000
CT251	Total	1672	32	127	0.990	2.978	0.811	0.27	−0.877	0.092	0.031
	Tarapaca		10	70	0.933	3.443	0.721	0.21	−0.140	0.383	0.005
	Nazca		12	55	0.970	2.198	0.719	0.33	0.280	0.227	0.015
	Canta		10	70	1.000	2.598	0.643	0.25	−0.350	0.197	0.029
CT268	Total	1881	34	128	1.000	3.360	0.569	0.17	−1.019	0.054	0.080
	Tarapaca		10	56	1.000	2.586	0.446	0.17	−0.510	0.246	0.031
	Nazca		12	70	1.000	3.451	0.615	0.18	0.150	0.173	0.042
	Canta		12	68	1.000	3.061	0.516	0.17	−0.350	0.119	0.077

Total = pooled sample, treated as a single population;
[a] Excluding indels.
[b] Number of alleles analyzed.
[c] Segregating sites.
[d] Haplotype diversity.
[e] Percent, nucleotide diversity per synonymous site.
[f] Percent, nucleotide diversity per non-synonymous site.
[g] Tajima's D for all sites (Tajima, 1989).
[h] Intralocus linkage disequilibrium, average of R^2 across all pairwise comparisons of polymorphic sites (Kelly, 1997).
[i] Population recombination rate per site (Hudson, 2001).

Evidence for elevated levels of amino acid polymorphism is consistent with balancing selection at this locus.

3.2. Population differentiation in Pto, Fen and Prf

The level of genetic differentiation between populations can be influenced by both demographic history and natural selection. Large differences in the amount of population differentiation between loci can point to individual loci that have been the targets of selection. We compared the levels of genetic differentiation between these three resistance genes and eight reference genes. F_{ST} ranged from 0.08 at Pto to 0.22 at Prf. These values were within the range of variation we observed at other loci from these same individuals (Supplementary Table 3). Therefore, on an individual gene basis, we did not detect deviations among these genes in the degree of population differentiation.

Recent methods have been developed to evaluate whether individual nucleotide positions within a gene show greater or lesser differentiation than expected based on population differentiation at an independent set of reference loci. Using these methods, we identified candidates for either balancing selection or directional selection in the R-genes (Fig. 1 and Fig. 2). In general, given the amount of population differentiation estimated from the reference loci and used for generating the 95% confidence intervals for the FDIST2 test of Beaumont and Nichols (1996) (mean "neutral" F_{ST} = 0.156), we had limited power to detect selection at these positions. Also using the Bayesian method of Foll and Gaggiotti (2008) with our data set, we obtained low Bayes factor values. Thus, we considered results as significant for all BF > 1.8, which is in agreement with results from the FDIST2 test. This corresponds to the posterior probability values between 0.64 and 0.8. According to the scale of evidence described by Jeffreys (1961), it represents weak to substantial evidence for the model assuming that the SNP loci are subject to selection. Collectively, using both methods, we detected two polymorphic positions as candidates for directional selection and six polymorphic positions as candidates for balancing selection.

These three resistance genes behave similarly to one another based on their patterns of nucleotide diversity and population differentiation. We were particularly interested how physical linkage and/or epistatic selection may have affected the evolutionary history of these genes. Pto and Fen both form complexes with Prf and are physically linked to Prf. Genome sequencing of this region in S. pimpinellifolium indicates that the Fen gene is only 2 kb from the coding region of Prf, while Pto is over 25 kb away. Recombination rates, which determine how quickly linkage associations break down, vary substantially across the tomato genome (see Table 1). The weighted average estimate of ρ across the set of reference loci in S. peruvianum is 0.0234 and LD decays rapidly in this outcrossing species (Arunyawat et al., 2007). Pto and Fen show relatively high levels of recombination (ρ = 0.07 and ρ = 0.06, respectively), while Prf shows a more moderate amount of recombination (ρ = 0.017). LD decays quite rapidly within these genes and the expectation of R^2 drops below 0.05 in less than 0.4 kb (Supplementary Fig. 2).

3.3. Linkage disequilibrium between Pto/Fen and Prf

Since LD decays on average relatively rapidly within these three R-genes, associations through chromosomal linkage may only play a minor role in correlated evolutionary patterns between these genes. However, if natural selection favors particular combinations of alleles at these loci, epistatic selection may still contribute to correlated evolutionary histories. To test for coevolution between these genes, we estimated LD at pairs of polymorphic non-synonymous positions between genes. This estimate of LD is based on observed genotypes and does not require the data to be phased.

We analyzed associations between loci for each population separately, since pooling alleles across populations could lead to spurious associations. In the Tarapaca population, we discovered 29 pairs of sites that were in LD between Pto and Prf (Supplementary Table 4). For Fen, we discovered 14 pairs of sites in LD with Prf (Supplementary Table 5). For this population, since we sequenced the entire Prf coding region, we could detect LD not only with the N-terminal region presumed to bind Pto and Fen, but also with other regions of Prf. Statistical significance of these associations was evaluated based on two kinds of p-values. Approximate p-values are derived from the composite disequilibrium coefficient using a chi-square approximation, while permutation based p-values correspond to the proportion of times the R^2 test statistic computed from randomly sampled data was found to be as extreme or more extreme than the statistical value of the original data. For the Tarapaca population, we report all pairs of sites for which these two p-values fell below the 0.05 level. For the other two populations, since fewer alleles were sampled, we had less power to detect associations. For these two populations, we report the LD between Pto/Fen and Prf that had approximate p-values lower than 0.05. For none of these pairs of sites, however, did the permutation based p-values fall below 0.05. Consequently, in Nazca we detected 18 pairs of sites that showed LD between Pto and Prf and three pairs of sites that showed LD between Fen and Prf (Supplementary Table 6). In turn, in Canta, we detected 7 pairs of sites that showed LD between Pto and Prf and 17 pairs of sites that showed LD between Fen and Prf (Supplementary Table 7).

3.4. Partitioning of LD variance components

We used a method developed by Ohta (1982a,b) to determine the relative contribution of epistatic selection to overall LD observed between these genes. We found that a significant portion of the LD observed between these genes could be attributed to neutral causes (genetic drift, population subdivision and limited migration; Table 2). Only a small fraction of the sites had a signature of what is considered unequal systematic disequilibrium. Unequal systematic disequilibrium can arise when epistasis operates in some, but not all subpopulations. Between Pto and Prf, 26 pairs of sites were identified for which LD was considered to be unequal, systematic (Supplementary Table 8). Between Fen and Prf, 42 pairs of sites were identified for which LD was considered to be unequal, systematic (Supplementary Table 9). The six Fen-Prf SNP pairs found to be candidates for epistasis/systematic disequilibrium between Fen and Prf included synonymous sites or doubletons and therefore were not considered further.

3.5. Correlated substitutions in proteins

We applied the method that does not rely explicitly on LD to determine if sites between Pto/Fen and Prf proteins are coevolving. This method, called ELSC, identifies putatively coevolving sites by evaluating how the distribution of amino acid residues at one site is dependent on the distribution of amino acid residues at a second site. This method does not take into account biochemical characteristics of the residues, but considers how the distribution of amino acid residues at different sites in a protein changes in sub-alignments, conditioned on a single site of the protein. Using this method, we identified eight pairs of sites between Pto and Prf and 14 pairs of sites between Fen and Prf that were putatively coevolving (Supplementary Table 10).

3.6. Candidate sites in Pto

Previous molecular and biochemical studies have identified many residues in Pto that are important for interaction with AvrPto
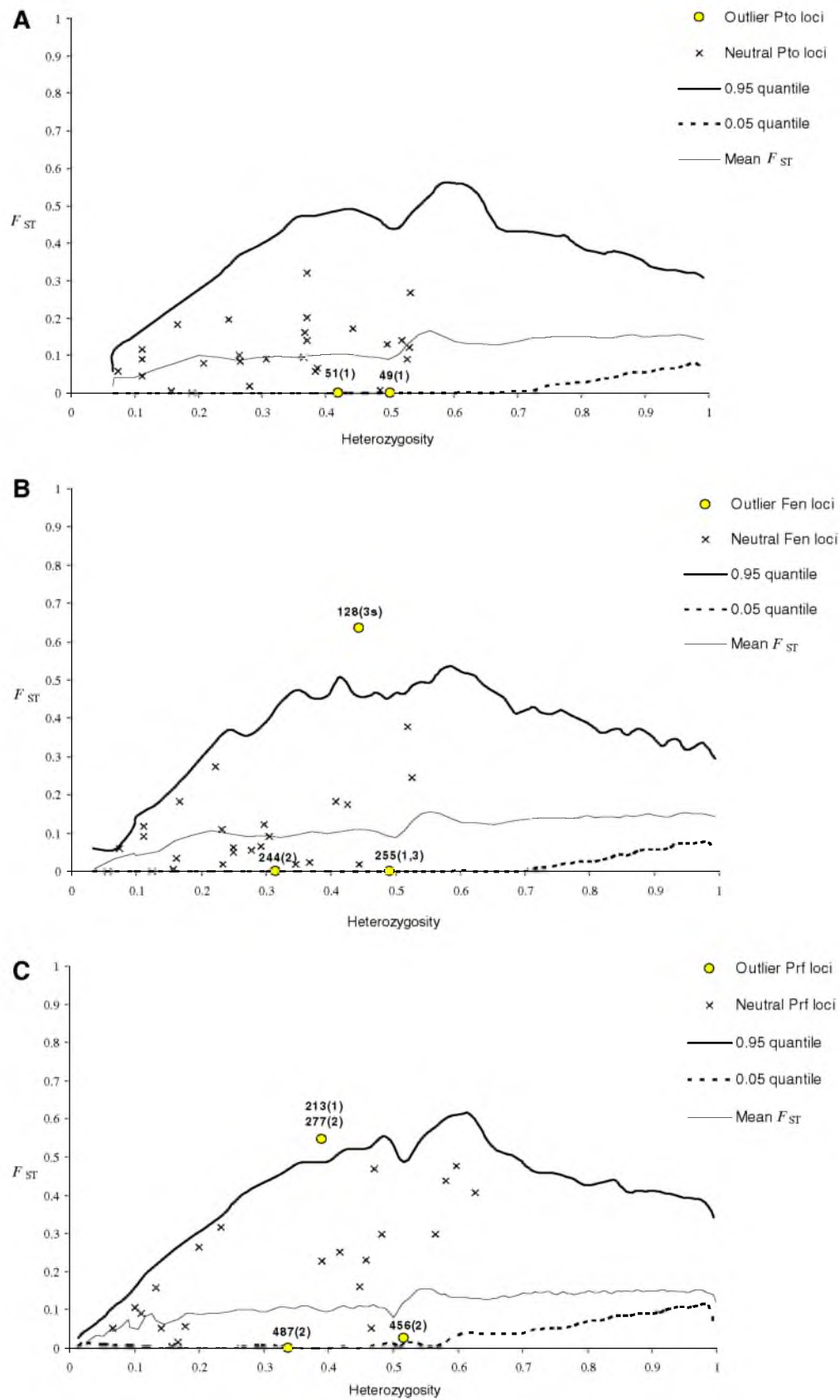
Fig. 1. Outlier SNP loci in (A) Pto, (B) Fen, (C) Prf based on the method of Beaumont and Nichols (1996). Each data point is a SNP locus. Loci with an F_{ST} value in 95% confidence interval were considered to be outlier loci (below 0.05 quantile – candidates for balancing selection, above 0.95 quantile – candidates for directional positive selection/local adaptation). Numbers indicate encoded amino acid and base position in codon (in parentheses; s – synonymous site).
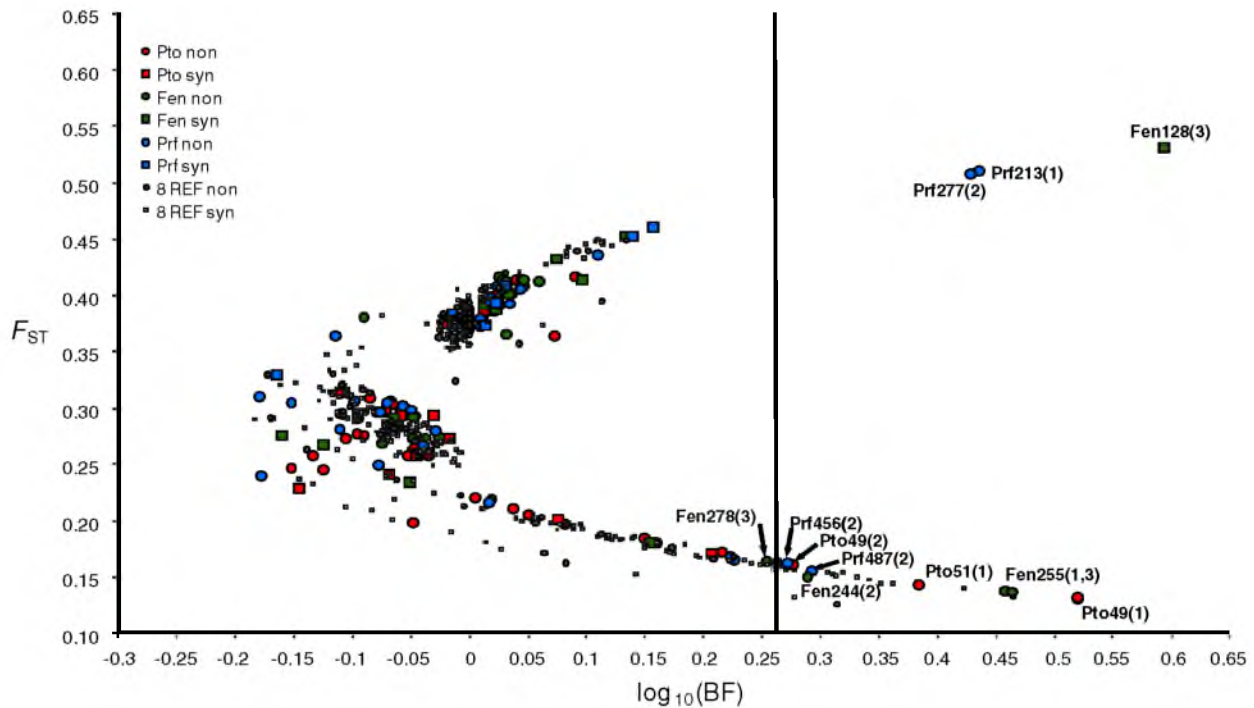
Fig. 2. Outlier SNP loci from Pto, Fen and Prf based on the method of Foll and Gaggiotti (2008). syn – synonymous sites, non – non-synonymous sites, 8 REF – eight reference genes. Numbers denote encoded amino acid and base position in codon (in parentheses). The line is a threshold indicating candidate sites consistent with results of the method by Beaumont and Nichols (1996).

Table 2
Summary of Ohta's LD coefficients between: 3420 SNP pairs of *Pto* and *Prf*,   2640 SNP pairs of *Fen* and *Prf*.

	Dual relationship	Number of *Pto-Prf* SNP pairs	Average values of Ohta's LD coefficients across SNP pairs				
			D_{IS}	D'_{IS}	D_{ST}	D'_{ST}	D_{IT}
1.	$D_{IS} < D_{ST}$ and $D'_{IS} > D'_{ST}$	3309 (96.75%)	0.005081	0.222157	0.056112	0.004625	0.226781
2.	$D_{IS} > D_{ST}$ and $D'_{IS} < D'_{ST}$	0	–	–	–	–	–
3.	$D_{IS} > D_{ST}$ and $D'_{IS} > D'_{ST}$	111 (3.25%)	0.030078	0.073305	0.022024	0.015795	0.089097

	Dual relationship	Number of *Fen-Prf* SNP pairs	Average values of Ohta's LD coefficients across SNP pairs				
			D_{IS}	D'_{IS}	D_{ST}	D'_{ST}	D_{IT}
1.	$D_{IS} < D_{ST}$ and $D'_{IS} > D'_{ST}$	2543 (96.32%)	0.004655	0.232894	0.059046	0.00409	0.236982
2.	$D_{IS} > D_{ST}$ and $D'_{IS} < D'_{ST}$	6 (0.23%)	0.066890	0.026460	0.012030	0.03128	0.057740
3.	$D_{IS} > D_{ST}$ and $D'_{IS} > D'_{ST}$	91 (3.45%)	0.043789	0.107362	0.024747	0.017258	0.124619

Interpretation:
1. Nonsystematic disequilibrium: restricted migration, genetic drift.
2. Systematic disequilibrium: epistatic selection.
3. Unequal systematic disequilibrium: partial epistatic selection.

and AvrPtoB and downstream signaling. Here we describe the functional context of the 18 sites in Pto that were recognized as candidates for natural selection and coevolution with Prf, using the methods explained above (Fig. 3A; Supplementary Fig. 3A and Supplementary Fig. 4A and Supplementary Fig. 5; Supplementary Table 11).

3.6.1. Domain I

The protein polymorphism in Pto between sites 43 and 88 is structured into two major haplotypes and many of the variable sites in this region show a pattern of LD that is consistent with unequal systematic epistasis (Supplementary Fig. 1; Supplementary Table 8). Sites 43 and 46 are associated with each other and are polymorphic in the Tarapaca and Nazca populations of *S. peruvianum*. This region of Pto was also identified in a DNA shuffling

study as important for AvrPto and AvrPtoB binding (Bernal et al., 2005). Sites 49 and 51 in Pto form hydrophobic contacts with Avr-Pto molecule and are described as one interface with AvrPto (Xing et al., 2007). F_{ST} analyses identified these sites as candidates for balancing selection. LD-based analyses pinpointed these sites as associated with Prf. Three alleles (H, E, A) segregate at site 49 in these populations. These segregating amino acid residues have very different biochemical properties (i.e. H is polar, basic and large, E is polar, acidic and small, while A is non-polar and small). Three alleles (V, L and G) also segregate at site 51 in these populations, although these amino acid differences are conservative. Site directed mutagenesis at sites 49 and 51 in Pto showed that the joint replacement of the H49E and V51G/D resulted in significantly reduced interactions with AvrPto, but not AvrPtoB (Xing et al., 2007; Dong et al., 2009). Many of our alleles also have the
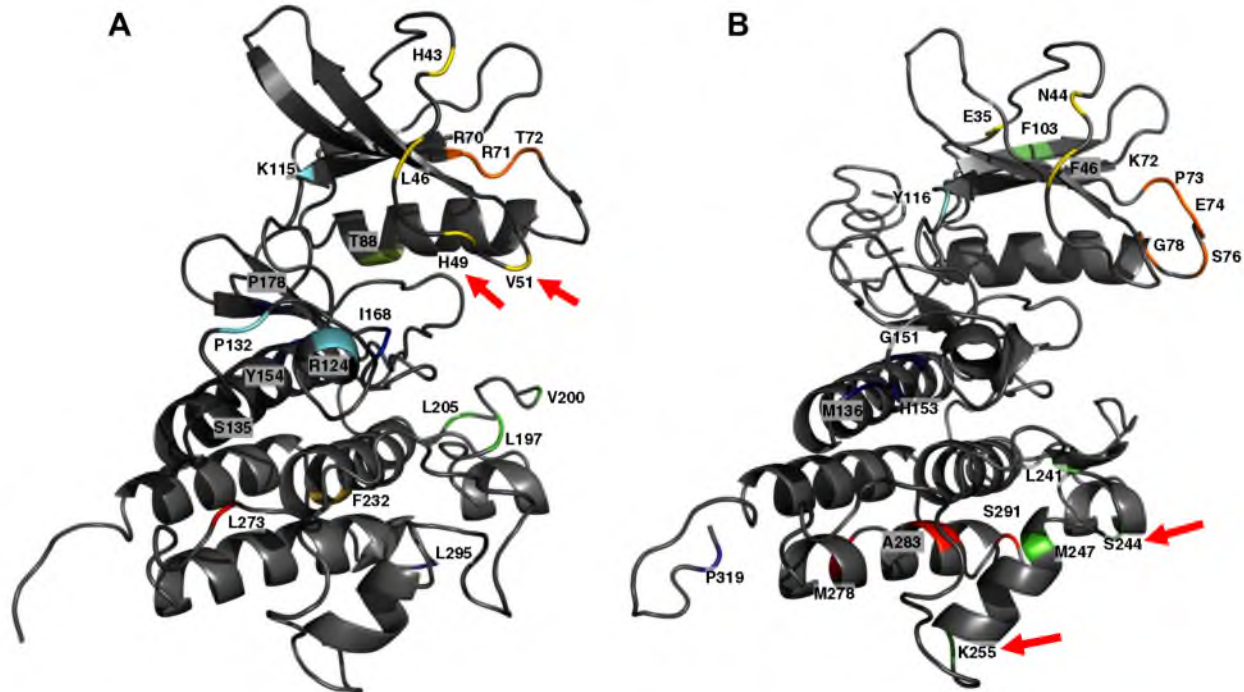
Fig. 3. Ribbon diagram of the (A) Pto crystal structure (*S. pimpinellifolium*, PDB 2qkw), (B) Fen protein model based on the allele from *S. pimpinellifolium*. Red arrows indicate the two polymorphic residues that are candidates for balancing selection. Other important polymorphisms are indicated by boldface letter and position.

combination of amino acid E49 with G51, as does the paralog Fen. Functional studies of other Pto alleles that contained E49/G51 from wild tomato species were able to activate an AvrPto specific resistance response (Rose et al., 2005). However, these alleles (parv94, chm115, peru567) differ not only at these two positions and therefore variation at other amino acid positions may have contributed to AvrPto binding and activation of disease resistance. Since functional versions of Pto and Fen have the combination of E49 and G51 and the proteins successfully signal through Prf, these substitutions do not appear to compromise Prf signaling. The observed correlation of above-mentioned Pto substitutions with Prf may be driven in part because this portion of Pto forms an exposed interface, perhaps not only for the pathogen ligand AvrPto, but for other interacting molecular partners such as Prf.

3.6.2. Domain II

The next sites in Pto strongly correlated with positions in Prf are sites 70, 71, and 72. Variation at these sites is structured into two distinct protein haplotypes (RRQ and SCK). These positions are variable in the Tarapaca and Nazca populations and all are non-conservative. Close to these sites is site K69, which is invariant in protein kinases and is required for ATP binding. Mutations at K69 abolish Pto kinase activity and the ability of Pto to interact with AvrPto (Scofield et al., 1996; Tang et al., 1996). Along with sites P73, E74, S76, G78, this region is necessary for binding of AvrPtoB, but not AvrPto (Bernal et al., 2005).

3.6.3. Domain III/Domain V

Polymorphisms at site 88, in kinase domain III, are also associated with polymorphisms in Prf and this site was identified using the ELSC and the LD-based methods. The T/I polymorphism in populations of Tarapaca and Nazca is rather conservative. This region is involved in anchoring and orienting the ATP molecule and it is generally strongly conserved in protein kinases (Hanks and Hunter, 1995). Sites 115 and 124 in domain V were also identified using these analytical methods. Four amino acids segregate at site

115 in these populations of *S. peruvianum* and this site was identified using the ELSC method. The K allele is found in all three populations, and the minor alleles Q, D and E are found in Tarapaca, Nazca and Canta, respectively. Although these substitutions are radical relative to one another, site-directed substitutions of K115E, and K115D in Pto did not affect the ability to bind to AvrPto and AvrPtoB (Bernal et al., 2005), indicating that even radical changes at this position may not negatively affect downstream signaling through Prf. Site 124 shows an S/R polymorphism in each population and was identified as a candidate for partial epistasis with Prf using Ohta's LD partitioning method.

3.6.4. Domain VIa/domain VIb

Sites 132 and 135 occur at the junction between domains V and VI. These sites are polymorphic in Nazca and form two haplotypes: P132/S135 and L132/F135. The major allele P132/S135 is conserved across most alleles of Pto in other wild tomato species, as well as in Pth2, Pth3, Pth5 and Fen in *S. pimpinellifolium* (Rose et al., 2005). The P to L substitution at 132 is conservative (both amino acids are small and non-polar), while the S to F substitution at 135 is non-conservative (S is polar, neutral and small, while F is non-polar and large). Domain VIa normally forms an extensive hydrophobic α-helix that stretches through the large lobe of the protein kinase. A polymorphism at site 154, towards the end of the α-loop, is correlated with variation in the Prf gene. This site is polymorphic in all three *S. peruvianum* populations and the two amino acid residues, F and Y, are at nearly equal frequency in these populations. The functional effect of this substitution has not been explicitly tested and this site was not polymorphic among the chimeras tested for AvrPto and AvrPtoB recognition by Bernal and colleagues (2005).

Domain VIb contains two β-strands with an intervening loop. The loop is known as the catalytic loop because it helps mediate phosphoryl transfer. In protein kinases, this loop is formed by the sequence HRD(L/V)KxxN. Across our alleles, no polymorphism is present in this loop except at position 168. This corresponds to

the first small "x" of the consensus sequence. This site is polymorphic in all three populations and is associated with Prf. Three different amino acid segregate (S, I and T), with S being the minor allele. Substitutions of I and T are conservative, while a substitution of S is non-conservative. This residue, along with 169, is predicted to be surface-exposed and to control Pto signaling. However, variation at 168 did not directly affect AvrPto and AvrPtoB recognition (Wu et al., 2004).

3.6.5. Domain VIII

Domain VIII comprises the P + 1 loop and plays a major role in ligand recognition. Two sites emerged as interesting candidates in this domain. Site 200 is polymorphic for I and V and both variants segregate in all three populations. This site was identified through the ELSC method as coevolving with Prf. Residue 205 was identified as a candidate for partial epistasis with Prf. Site 205 is polymorphic in all three populations and the L/F polymorphism is rather conservative (both are non-polar, small → large). Together with sites T204, I208, F213, site 205 forms the second interface for binding AvrPto (Xing et al., 2007) and with residues F213, V242, V250, N251, the first interface for binding AvrPtoB. Site directed mutagenesis at sites 205 and 213 in Pto showed that the joint replacement of the L205A and F213A disrupted the interaction of Pto with AvrPtoB, but not AvrPto (Dong et al., 2009). Furthermore, this residue along with I214 and N251 forms a negative regulatory patch (NRP), which controls many aspects of signaling, including a negative regulation of signaling through Prf (Wu et al., 2004).

3.6.6. Domain X/domain XI

One site is found to be a candidate for coevolution in domain X. Site 273 emerged as a candidate from both Ohta's LD analysis and the ELSC method. This site is polymorphic for I/L in all populations. Little is known about the potential functional effects of variation at this position; however residues in the domain X are required for interaction with the pathogen effectors and downstream signaling (Bernal et al., 2005).

One site in domain XI was identified as a candidate for partial epistasis with Prf. Site 295 is polymorphic in Nazca (L/S), but not in the other two populations. This polymorphism results in a non-conservative change. Mutational analysis of this site showed that the non-conservative substitution of L295D behaved as wild type and was able to induce AvrPto-dependent cell death (Mucyn et al., 2009). This may indicate that functional differences between the L and S alleles may not be evident in AvrPto-based detection assays.

3.7. Candidate sites in Fen

Functional information is also available on the Fen protein kinase. Here we describe 11 sites identified in Fen as candidates for natural selection and coevolution with Prf (Fig. 3B, Supplementary Fig. 3B, Supplementary Fig. 4B, Supplementary Fig. 6; Supplementary Table 11).

3.7.1. Domain I/domain II

Site 46 of Fen is polymorphic in only a single population of *S. peruvianum*, namely Nazca. The major allele at this locus encodes a phenylalanine, however most individuals in Nazca are heterozygous for F and L. ELSC identified this site as potentially coevolving with Prf. Site 76 in domain II was identified in all three methods as coevolving with sites in Prf. The homologous position in Pto has been shown to be critical for AvrPtoB binding and this site is monomorphic for serine in our collection of 54 Pto alleles across seven tomato species. In contrast, at Fen four different amino acid residues are present: R, K, S, and G, and the major allele is R. Alleles

of Fen from *S. lycopersicum* and *S. pimpinellifolium* bind AvrPtoB and encode S at this site, as does Pto. This raises the question whether the Fen alleles from *S. peruvianum* are also able to bind AvrPtoB in a similar way and what role this site may play in protein interactions with Prf.

3.7.2. Domain VIa

Domain VIa typically forms a large α-helix away from the active site of the protein and may serve as a structural support of the kinase (Hanks and Hunter, 1995). Site 136 within this region was polymorphic in Nazca and Canta for I and M, but fixed for I in the Tarapaca population. This site was identified as coevolving with five sites in Prf. Site 151, polymorphic only in the Tarapaca population, also was identified as coevolving with sites in Prf. Three coevolving sites in Prf were consistently identified across three methods (LD, Ohta's LD partitioning and ELSC) and these Prf sites were different from those identified as coevolving with site Fen136. Site 153 in this domain was polymorphic in the Nazca and Canta populations and was identified as coevolving with Prf using Ohta's LD partitioning method.

3.7.3. Domain X/domain XI

Three sites in domain X were identified as coevolving with Prf or under balancing selection. Site 241 was polymorphic in Tarapaca and identified as coevolving with a site in Prf. Sites 244 and 255 were polymorphic in all three populations and were identified as both coevolving with Prf and experiencing balancing selection. There is an overlap in the coevolving partners identified in Prf for these two Fen polymorphisms. It was shown previously that the region between residues 243 and 258 in Pto is important either for correct protein folding or binding to the Avr proteins and downstream components (Bernal et al., 2005). Three sites in domain XI were identified as coevolving with Prf. Site 278 is polymorphic in all three populations and identified as experiencing balancing selection by one of the F_{ST}-based methods, while site 283 and site 291 are polymorphic in Tarapaca and Canta, but not in Nazca. The coevolving sites in Prf identified for these three sites in Fen are located towards the distal region of the Prf N-terminus.

3.8. Candidate sites in Prf

Prf is a large protein with five domains (Fig. 4). The N-terminal domain of Prf functionally interacts with Pto and Fen (Mucyn et al., 2006, 2009; Chen et al., 2008) and shows an excess of non-synonymous variation, compared with other domains in this protein (Rose et al., 2011). Twenty-one amino acid positions were identified as candidates for natural selection and coevolution with Pto and Fen. Three regions in the N-terminal domain of Prf can be recognized: (1) proximal, amino acid sites 23–120, (2) middle, 135–277 and (3) distal 397–536 (Fig. 4A; Supplementary Fig. 3C; Supplementary Table 11).

3.8.1. The distal region of Prf N-terminus

In this region many residues show partial epistasis with Fen and Pto. Site 397 is polymorphic in the Tarapaca population for Q and L. The L allele is the minor allele in this population and appears also in Prf from *S. habrochaites*. This residue is in LD with Pto154 in Tarapaca and is a candidate for epistatic selection with many sites in Pto and Fen, namely Pto49, 124, 154 and Fen76, 255, 283, 291. In addition, site Fen76 was identified as coevolving with Prf397 by the ELSC method. Site 456 is polymorphic in all three populations, with C as the major allele in Tarapaca, Y predominating in Canta, and with both alleles in equal frequency in Nazca. This site shows unequal systematic disequilibrium with residues Pto273 and Pto295, as well as sites 151, 153, 244, 255 and 278 in Fen. Furthermore, this locus is a candidate for balancing selection detected by
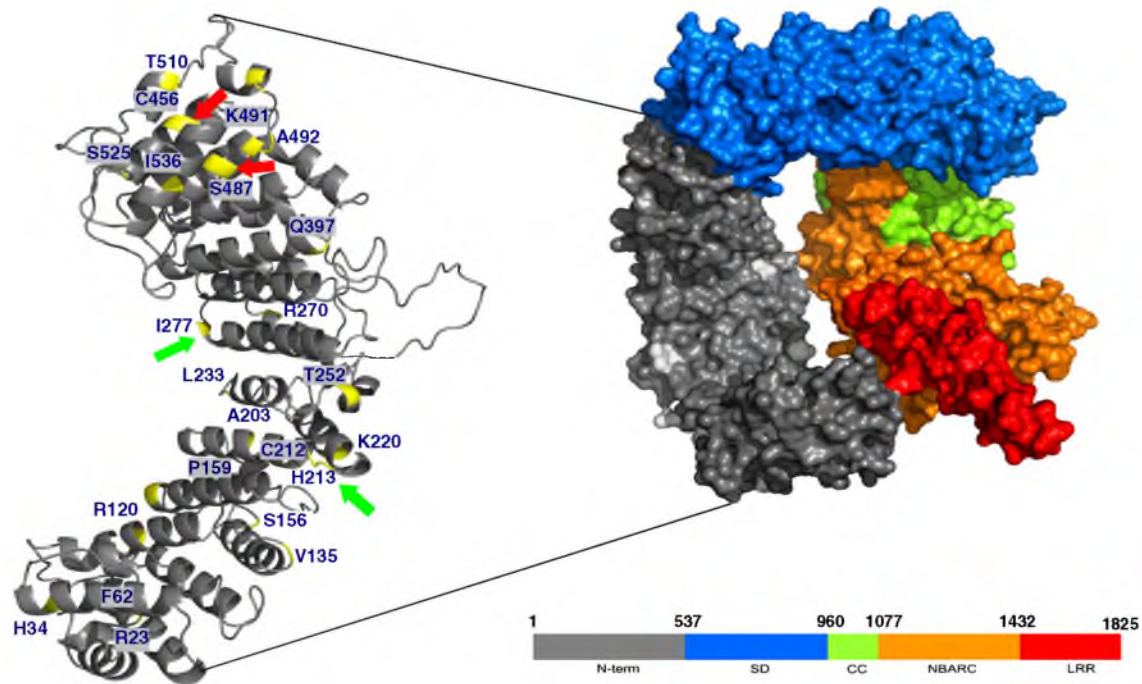
Fig. 4. Model of the Prf protein. Molecular surface representation of Prf (on the right) with a domain diagram (below) and the N-terminal domain shown as a ribbon diagram with polymorphic sites in yellow (on the left); arrows indicate candidate sites for balancing selection (red) and directional selection (green); numbers indicate amino acid positions; N-term, N-terminus domain; SD, Solanaceous domain; CC, coiled-coil domain; NBARC, nucleotide-binding domain shared by Apaf-1, certain R proteins, and CED-4 ATPase domain; LRR, leucine-rich repeat domain.

F_{ST}-based methods. Site 487 is also segregating in all three populations (S/F) with the major allele S. The substitution S487F is a radical change (polar, small → non-polar, large) and the F allele is present in *S. habrochaites*. This site is not only a candidate for epistatic relationship with multiple sites in Pto and Fen (namely Pto43, 46, 49, 51, 70, 71, 88, 154, 205 and Fen76, 153, 244, 255, 278, 283, 291), but also indicated as experiencing balancing selection in F_{ST}-based tests. Of these Pto and Fen residues, Pto49, 51 and Fen244, 255, 278 are also candidates for balancing selection. Sites 491 and 492 are polymorphic only in Tarapaca, where they form haplotypes K491/A492 and N491/S492. The former is the major allele and the replacement K491N is a radical change (basic, large → neutral, small), whereas the replacement A492S is rather conservative. The residues Prf491 and Prf492 in the Tarapaca population are in LD with sites Pto49 and Fen151, and are candidates for epistatic selection with these loci. The site Fen151 was identified by the ELSC method as coevolving with these Prf loci. Two additional sites in Fen (Fen255 and Fen278) showed significant LD with these Prf sites. The next two residues were identified in the ELSC method as coevolving partners of Fen only. Prf site 510 is putatively coevolving with Fen136. This locus is segregating for S and T in Nazca and Canta, and is fixed for T in Tarapaca. This replacement is conservative, however, the S allele is a major allele in Canta and both alleles are present in nearly equal frequency in Nazca. Site 536 is polymorphic only in Tarapaca for I and M with I as the major allele. This locus was identified by the ELSC method as putatively coevolving with Fen241.

3.8.2. The middle region of Prf N-terminus

Sites in this region are identified by different methods as coevolving with Fen only. Site 135 is polymorphic in Nazca and Canta with a conservative change from V to L. The transition from the fixed allele V in Tarapaca to the predominating allele L in Canta is similar to that observed at site 120 in the proximal region of

Prf N-terminus. This position is associated with Fen136 as predicted by the ELSC method. Sites 156 and 159 in Prf are putatively coevolving with amino acid Fen46, as was shown by the ELSC analysis. Both sites are polymorphic and represented by two distinct haplotypes in these populations (S156/P159, R156/S159). The replacement S156R is a radical change (serine is neutral and small, arginine is basic and large), while the substitution P159S is rather conservative (both are small, non-polar → polar neutral). The combination S156/P159 is the major allele, present also in *S. habrochaites* and *S. lycopersicum* (GenBank AAF76312). The allele R156/S159 is segregating only in Nazca with one case observed in Canta. Sites 213 and 277 are polymorphic only in the Canta population with 213D/277T as the major allele. Sequences from *S. habrochaites* and *S. lycopersicum* have 213H/277I allele, which is fixed in Nazca and Tarapaca. The change H213D is a radical change (histidine is basic and large, whereas asparagine is acidic and small). In contrast, the replacement I277T is rather conservative (both isoleucine and threonine are small, non-polar → polar neutral). Moreover, residues Prf213 and Prf277 are shown by F_{ST}-based methods as significant candidates for directional selection. The method ELSC indicated both these loci as coevolving with Fen136. Site 220 is a candidate for experiencing epistatic selection together with Fen255. Prf220 in the Tarapaca population is segregating for K and I. The I allele is the major allele, present also in Prf from *S. habrochaites*, while K appears in Prf from *S. lycopersicum*. The substitution I220K is a radical change – isoleucine is non-polar and small, whereas lysine is polar, basic and large. Other sites in this region (203, 212, 233 and 252) were candidates for coadaptation with both Pto and Fen. Site 203 is polymorphic across three populations of *S. peruvianum*. In the Tarapaca population, A is the major allele, while T is the major allele in Nazca and Canta. Replacement from A to T is rather a conservative change. The residue 203 was predicted by the ELSC method as a candidate for correlated evolution with Pto273 and Fen136. Site 212 is segregating in three populations
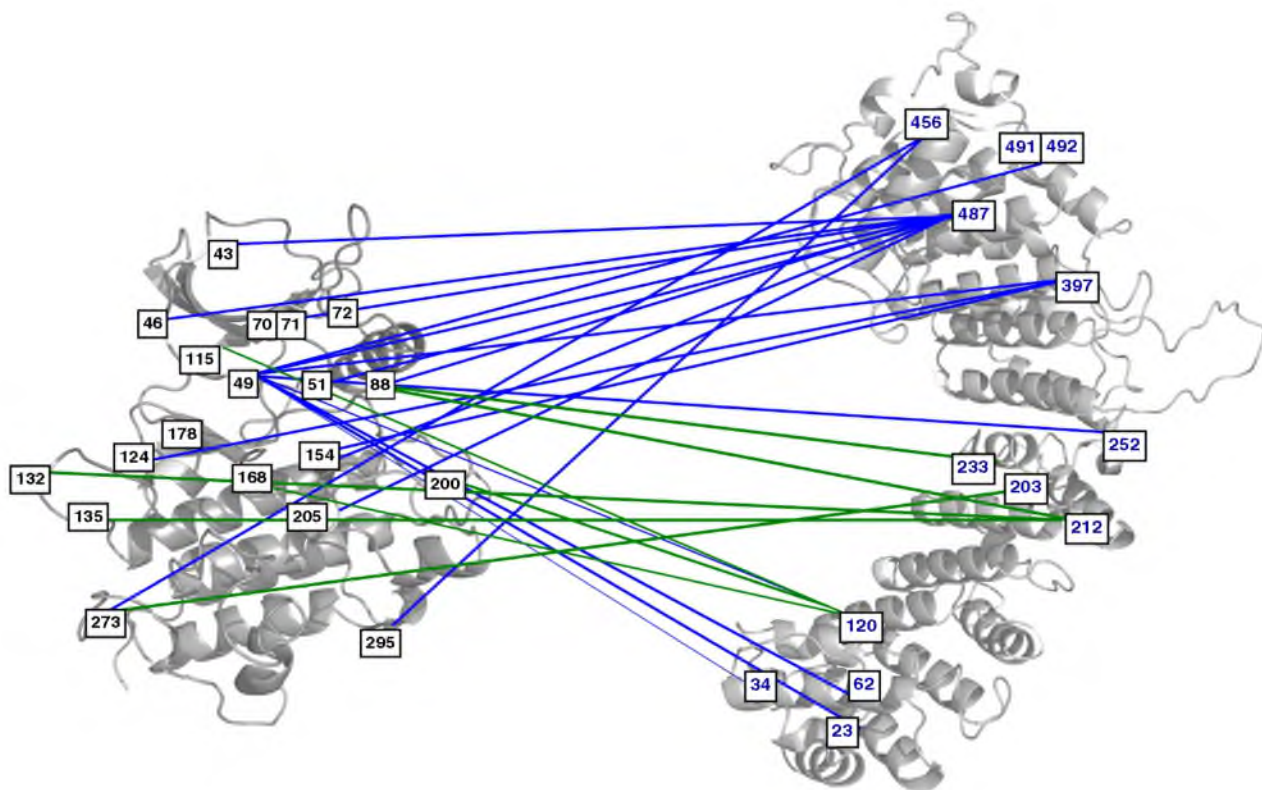
Fig. 5. Combined picture of regions coevolving between Pto and Prf detected by Ohta's partitioning of LD (blue lines) and ELSC (green lines).
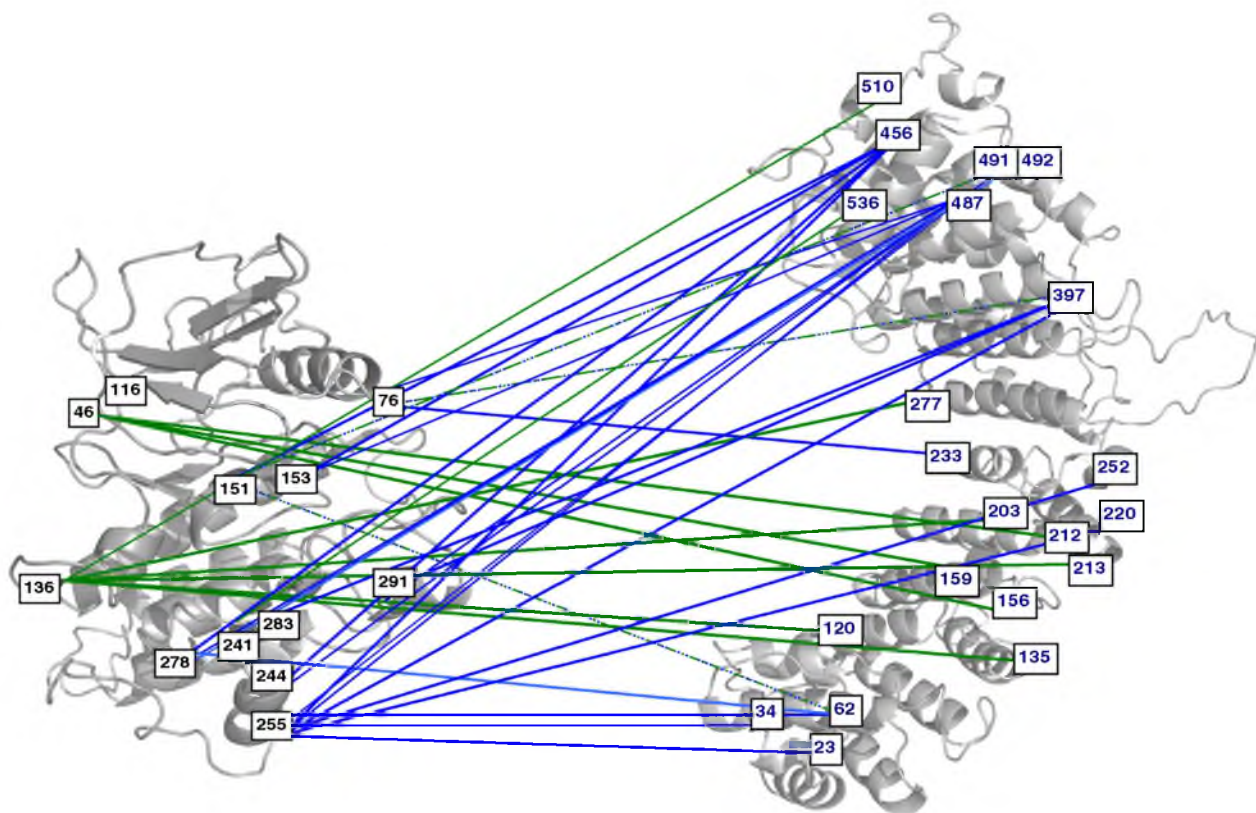

Fig. 6. Combined picture of regions coevolving between Fen and Prf detected by Ohta's partitioning of LD (blue lines) and ELSC (green lines).

with C as the major allele in Tarapaca and Canta, present also in *S. lycopersicum*. In Nazca, F predominates and appears also in *S. habrochaites*. A third allele Y is present in Nazca and Canta. This site was identified by the ELSC method as coevolving with Pto sites 88/132/135 and site 46 in Fen. Site 233 is polymorphic for L and M in these three populations. The allele L is the major in Tarapaca and Canta, whereas in Nazca the two alleles are present in equal frequency. The methionine is found in Prf sequence from *S. habrochaites* and the substitution L233M is conservative. The ELSC method indicates this site as putatively coevolving with Pto88, while the partitioning of LD suggests that this site is in epistatic relationship with Fen76. Site 252 in Prf, along with site 220 in this region, is another candidate for partial epistatic selection not only with Fen255, but also Pto49. This site segregates in three populations of *S. peruvianum* with alleles K and T. The K allele is the major allele in Tarapaca, while T predominates in Nazca and Canta. The change from K to T is radical – lysine is polar and large and threonine is neutral and small.

3.8.3. The proximal region of Prf N-terminus

Sites 23 and 34 are polymorphic in Tarapaca and Nazca and form two haplotypes: R23/H34 and W23/Y34. The haplotype W23/Y34 is fixed in Canta, R23/H34 is present in the Prf allele from *S. lycopersicum* and the combination R23/Y34 in the *S. habrochaites* outgroup allele. Both substitutions are rather conservative (R23 W – basic, large, polar → non-polar, and H34Y – polar, large, basic → neutral). Both sites are associated with candidates for balancing selection (Pto49 and Fen255) and this is consistent with partial epistatic selection as detected by partitioning of LD.

Site 62 is polymorphic only in Tarapaca and a substitution at this site from F to Y is rather conservative. This site showed significant associations with putative balanced polymorphisms at residues Pto49, Fen255 and Fen278. In addition, the ELSC method indicated that this site is coevolving with Fen151, which was supported by LD analysis. Site 120 is polymorphic in all three populations with the minor allele L only in Tarapaca and the major allele changing from R in Tarapaca to Q in Canta with a transitory state R/Q in Nazca. The R allele is present in Prf from *S. lycopersicum* and the Q allele in *S. habrochaites*. A displacement of the polar, basic, large R to polar, neutral, small Q and non-polar, small L is a radical change. The ELSC method detects associations with a similar region in Pto (sites 115, 168, 200) and site 136 in Fen.

In summary, polymorphisms in the middle region of the N-terminus of Prf were identified as coevolving with Pto or Fen based on the ELSC method. Some of these same sites were identified by F_{ST}-based methods as candidates for directional positive selection. Several sites in the distal region of the Prf N-terminus domain were identified as candidates for unequal systematic disequilibrium (i.e. partial epistasis) with Pto or Fen using Ohta's partitioning of LD method. Some of these same sites were identified by F_{ST}-based methods as candidates for balancing selection. The partner sites in Pto or Fen were also identified as candidates for balancing selection (Figs. 5 and 6; Supplementary Table 11).

4. Discussion

4.1. Detecting epistatic selection between interacting proteins

In this work, we implemented population genetic and bioinformatics methods to analyze selective constraints and to determine if polymorphisms in interacting proteins were significantly associated with one another. In general, our results based on partitioning of LD variance and correlated substitutions analysis did not overlap greatly. The differences are not surprising, since the underlying assumptions of these methods are quite different.

One of the approaches used was Ohta's method to partition the total variance of linkage disequilibrium into within and between population components (Ohta, 1982a,b). This method was developed to discriminate between epistatic natural selection and stochastic processes as the main cause of the observed LD. Systematic associations among alleles in isolated populations of a species may be taken as evidence of the direct action of natural selection on the loci involved (Lewontin, 1974). For systematic associations, there is a relatively large within-population component and a relatively small between-population component, because LD is in the same direction in each population. In contrast, a large between-population component of LD is most readily attributable to non-selective effects of population subdivision or founder effects (Brown and Feldman, 1981; Ohta, 1982a,b). In this study, results of LD partitioning also suggest that restricted migration and genetic drift are the main causes of observed associations between genes from the *Pto* cluster.

Approximately 3% of the pairs of polymorphic residues of Pto and Prf or Fen and Prf meet the criteria for unequal systematic disequilibrium. This type of LD is an intermediate between systematic and non-systematic disequilibrium and is equivalent to partial epistasis. This means that epistatic selection occurs in only a few subpopulations or might be also interpreted as interaction of genetic drift and epistatic selection. In our study, sites identified as targets of balancing selection based on F_{ST}-outlier methods also show epistatic associations between proteins (e.g. domain I in Pto, domain X in Fen and the distal region of the Prf N-terminus; Fig. 1, Fig. 2, Fig. 5 and Fig. 6; Supplementary Table 11).

Coevolution with pathogens in a local population may lead to the maintenance of different allelic variants at a particular locus on a small geographic scale. Since host resistance proteins must interact with other proteins in the plant cell to activate resistance, compensatory changes of the partner proteins may be necessary, and these evolutionary changes may even take place in isolated populations. Selection against particular allelic combinations between the interacting proteins that fail to activate resistance properly will result in an under-representation of those particular combinations and an over-representation of the protein variant pairs, which activate resistance properly. This enrichment of "coadapted" variant pairs can leave a signature of partial epistasis, if pathogen selection pressure is variable across host populations. In our study, the polymorphic sites that showed evidence of balancing selection on an individual locus level coincided with signatures of partial epistasis.

According to an independent method to detect associations between protein partners (ELSC; Dekker et al., 2004), most intermolecular associations involved sites from the middle region of the Prf N-terminus. Based on Ohta's method, stochastic processes were the likely source of linkage associations with this region of Prf. Results of F_{ST}-based outlier detection methods suggest that some sites in this region of Prf are targets of directional selection. In addition, the interacting sites from Pto and Fen are not polymorphic in all subpopulations. Taken together the set of significant Pto-Prf and Fen-Prf pairs detected by the ELSC method could reflect sites subject to directional selection in particular local populations.

4.2. Distribution of natural selection across genes in the Prf complex

Many of the sites in Pto associated with Prf are known to form contact interfaces with the bacterial effectors. Based on this study, these same sites are likely to be experiencing balancing selection. Likewise, residues in domain VIII of Pto form a negative regulatory patch (NRP), and when these sites are mutated, Pto fails to interact with its pathogen ligands (Wu et al., 2004; Xing et al., 2007; Mucyn et al., 2009; Dong et al., 2009). Therefore, this overlap between NRP and AvrPto/AvrPtoB interaction sites in Pto, together with the

results in present study, imply that effector binding interferes with inhibitory residues of Pto and disrupts negative regulation to trigger Prf-dependent immune response.

A crystal structure of Fen is not solved, but due its homology to Pto, the functional importance of some of these same sites may be extrapolated. In contrast to Pto, residues in domain X of Fen seem to experience balancing selection. Domain X is highly variable among kinases and seems to be more conserved in subfamilies that share similar functions. The homologous region in Pto is conserved in S. peruvianum and previously it was observed that replacements of sites from 243 to 258 disrupted all phenotypes, suggesting the importance of this region for pathogen ligand biding, downstream signaling or correct protein folding (Bernal et al., 2005).

Lack of a crystal structure of Prf makes it difficult to ascertain functionally important amino acids within this protein. Therefore the structure of the Prf protein was computationally predicted. According to the protein model, the N-terminal domain of Prf forms a large molecular arm jutting out from the one side of the protein (Fig. 4). The distal region of Prf N-term shows most of the associations with Pto and Fen, consistent with partial epistatic selection. In addition, some sites in this region were identified as candidates for balancing selection. In contrast, the middle region of Prf N-term domain shows coevolving candidate sites detected by the ELSC method and a few sites consistent with partial epistatic selection between Prf and Pto/Fen. Many of these residues are predicted to coevolve with Fen, but not with Pto and are not polymorphic in all subpopulations. Moreover, some sites of this region are candidates for directional positive selection. This may suggest that these amino acids experience epistatic adaptation with Pto and Fen in only some subpopulations.

A key question is what role the N-term domain of Prf plays in interaction with Fen and Pto. Although this domain is a novel sequence of unknown function, it seems likely that the Prf N-term–kinase complex could provide a regulatory node, in addition to the NBARC-LRR portion of Prf. Mutational analyses suggest that the NBARC-LRR moiety of Prf acts downstream of the N-term/kinase switch, but these nodes do not act independently during signaling (Mucyn et al., 2009; Du et al., 2012; Kud et al., 2013). Hence, Fen/Pto and Prf may associate to form a recognition complex characterized by multiple regulatory molecular interactions. This suggests that radical replacements within the interacting interfaces of either Prf protein or Pto/Fen kinase could lead to inactivation or incorrect activation of signaling, which may be detrimental to the host. This can explain the genomic colocation and tight coadaptation of Prf with the Pto gene family. Another example of tightly linked genes involved in the same physiological process is the S locus in the Brassica species. The two genes that control self-incompatibility, SRK and SLG, are separated by a maximum distance of 220 kb (Boyes and Nasrallah, 1993). The SRK gene encodes a receptor kinase that determines specificity of the stigma in self-incompatibility recognition reactions and SLG encodes a glycoprotein that can enhance this process. By analogy to the Prf protein complex, SRK and SLG are proposed to interact (Takasaki et al., 2000) and like in the plant defense response, this involves a growth restriction of an invading organism (in this case the pollen tube). The components of these recognition complexes may be somewhat unique in that they are dependent on each other for a specific function and have strong fitness effects.

The high degree of Prf sequence conservation may reflect its ancient origin. Pto and Fen share their common ancestor between 27.9 and 34.0 mya (Rose, 2002). Since both Fen and Pto require Prf, this suggests that Prf evolved to function with a progenitor of the Pto family (Rosebrock et al., 2007). How the structurally unrelated Prf gene became clustered in the Pto gene family is an interesting question. The structural differences between components of these multifunctional loci exclude the possibility that they

arose by duplication and divergence of a single ancestral gene. Another pathway that uses both a protein kinase and an LRR-containing protein is the pathway involved in resistance of rice to bacterial blight. There the kinase and the LRR domain are both encoded by a single gene (Song et al., 1995). It is common that functionally interacting proteins, which are encoded by separate genes in some organisms, are fused in a single polypeptide chain in others. Thus, one evolutionary scenario is that the Prf and Pto family members are derived from an ancestral tomato resistance gene in which these domains were fused (Salmeron et al., 1996). On the other hand, some type of transposition or rearrangement brought the two types of genes close to each other and selection favored this system because of the correlation of the genes for the resistance phenotype. In this case, proximity to Prf and possible simultaneous expression of proteins with distinct kinases, could be more flexible than fusion with only one kinase (e.g. the closely located Fen). This may be required by the host to counteract ongoing pathogen evolution. For instance, Pto homologs may be able to confer resistance to different pathogen isolates (Chang et al., 2002).

Interaction between Prf and Pto/Fen kinase may suggest that Prf residues could contribute to pathogen detection via Avr proteins binding (Mucyn et al., 2006; Balmuth and Rathjen, 2007). In addition to its role in signaling, Prf might serve as a targeting subunit which acts as organizing platform that recruits both the kinase and the effector (or kinase substrate) to the same complex. The role as a scaffolding adaptor protein is suggested by function of parental proteins used in this study for Prf modeling (see Materials and methods). Furthermore, Prf was proposed as an indirect target of bacterial effectors (Ntoukakis et al., 2009), thus some of the associations detected here may emerge from an indirect link with Avr proteins.

In this study, we detected a number of polymorphic sites in Pto and Fen that showed associations with polymorphisms in the Prf gene. These associations with Prf were mostly concentrated in the N-terminus of Prf, which has been proposed to be a binding surface for Pto and Fen based on Co-IP studies. Since pathogens are a variable and ever-changing aspect of an organism's environment, molecular evolutionary studies of host resistance provide insights into the adaptive process in general. The tight physical linkage in this small gene network may itself represent an adaptive strategy, because the tight physical linkage allows for a much faster rate of adaptation than does loose linkage (e.g. with the genes scattered across the genome). Precisely, the fast adaptive response which is necessary for keeping up with the reciprocal adaptation of ones pathogens can be achieved through tight physical linkage, as predicted by models, such as Kimura's compensatory evolution model (Kimura, 1985).

In previous studies, we found evidence for the maintenance of protein polymorphism specifically in Pto and the N-terminus of Prf (Rose et al., 2007, 2011). For future breeding purposes, the protein polymorphism detected in this R-gene complex could be mined for novel resistance specificities. However, care should be taken to avoid a mix and match approach. Precisely, the presence of protein polymorphism in the interacting partners of this multimeric complex, resulting from selective mechanisms to maintain rather than remove amino acid variation in natural populations, makes this R-gene functional unit susceptible to genetic incompatibilities. In the presence of balancing or diversifying selection, the allele frequency of any given protein morph may be low within the population or species, and all that can be assumed is that an allelic form of the protein functions with the other proteins encoded by genes in the same genome. Whether this allele type will function with protein partners from other genotypes cannot be predicted. When the multimeric functional unit can be bred intact, incompatibilities can be avoided, but the creation of a novel genotype from a mixture of non-adapted alleles continues to pose

a challenge to breeders. The presence in natural populations of extensive protein polymorphism, not only in the confirmed pathogen receptors, but also in other components of multimeric resistance complexes, sends a clear message that a long-term strategy for effective crop protection must actively embrace genetic diversity, with both its negative and positive consequences. Protein polymorphism present within a population is not simply a quirky property of wild plant-pathogen systems, but should be recognized as a guide for applied crop protection.

Acknowledgements

References

Arunyawat, U., Stephan, W., Staedler, T., 2007. Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. Mol. Biol. Evol. 24, 2310–2322.

Balmuth, A., Rathjen, J.P., 2007. Genetic and molecular requirements for function of the Pto/Prf effector recognition complex in tomato and *Nicotiana benthamiana*. Plant J. 51, 978–990.

Beaumont, M.A., Nichols, R.A., 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. Lond. B 263, 1619–1626.

Bernal, A.J., Pan, Q.L., Pollack, J., Rose, L., Kozik, A., et al., 2005. Functional analysis of the plant disease resistance gene *Pto* using DNA shuffling. J. Biol. Chem. 280, 23073–23083.

Black, W.C., Krafsur, E.S., 1985. A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. Theor. Appl. Genet. 70, 491–496.

Bomblies, K., 2009. Too much of a good thing? Hybrid necrosis as a by-product of plant immune system diversification. Botany 87, 1013–1022.

Boyes, D.C., Nasrallah, J.B., 1993. Physical linkage of the *SLG* and *SRK* genes at the self-incompatibility locus of *Brassica oleracea*. Mol. Gen. Genet. 236, 369–373.

Brown, A.H.D., Feldman, M.W., 1981. Population structure of multilocus associations. PNAS 78, 5913–5916.

Chang, J.H., Tai, Y.-S., Bernal, A.J., Lavelle, D.T., Staskawicz, B.J., Michelmore, R.W., 2002. Functional analysis of the *Pto* resistance gene family in tomato and the identification of a minor resistance determinant in a susceptible haplotype. Mol. Plant Microbe. Interact. 15, 281–291.

Chen, H., Zou, Y., Shang, Y., Lin, H., Wang, Y., et al., 2008. Firefly luciferase complementation imaging assay for protein–protein interactions in plants. Plant Physiol. 146, 368–376.

Chetelat, R.T., Pertuze, R.A., Faundez, L., Graham, E.B., Jones, C.M., 2009. Distribution, ecology and reproductive biology of wild tomatoes and related nightshades from the Atacama Desert region of northern Chile. Euphytica 167, 77–93.

Cockerham, C.C., Weir, B.S., 1977. Digenic descent measures for finite populations. Genet. Res. 30, 121–147.

Dekker, J.P., Fodor, A., Aldrich, R.W., Yellen, G., 2004. A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. Bioinformatics 20, 1565–1572.

DeLano, W.L., 2008. The PyMOL Molecular Graphics System. DeLano Scientific LLC, Palo Alto, CA.

Dong, J., Xiao, F., Fan, F., Gu, L., Cang, H., et al., 2009. Crystal structure of the complex between *Pseudomonas* effector AvrPtoB and the tomato Pto kinase reveals both a shared and a unique interface compared with AvrPto-Pto. Plant Cell 21, 1846–1859.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. Phytochem. Bull. 19, 11–15.

Du, X., Miao, M., Ma, X., Liu, Y., Kuhl, J.C., Martin, G.B., Xiao, F., 2012. Plant programmed cell death caused by an autoactive form of Prf is suppressed by co-expression of the Prf LRR domain. Mol. Plant 5, 1058–1067.

Excoffier, L., Laval, G., Balding, D., 2003. Gametic phase estimation over large genomic regions using an adaptive window approach. Hum. Genomics 1, 7–19.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Res. 10, 564–567.

Flint, J., Bond, J., Rees, D.C., Boyce, A.J., Roberts-Thompson, J.M., et al., 1999. Minisatellite mutational processes reduce Fst estimates. Hum. Genetics 105, 567–576.

Foll, M., Gaggiotti, O., 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. Genetics 180, 977–993.

Frederick, R.D., Thilmony, R.L., Sessa, G., Martin, G.B., 1998. Recognition specificity for the bacterial avirulence protein AvrPto is determined by Thr-204 in the activation loop of the tomato Pto kinase. Mol. Cell 2, 241–245.

Garnier-Gere, P., Dillmann, C., 1992. A computer program for testing pairwise linkage disequilibria in subdivided populations. J. Hered. 83, 239.

Gutierrez, J.R., Balmuth, A.L., Ntoukakis, V., Mucyn, T.S., Gimenez-Ibanez, S., Jones, A.M., et al., 2010. Prf immune complexes of tomato are oligomeric and contain multiple Pto-like kinases that diversify effector recognition. Plant J. 61, 507–518.

Hanks, S.K., Hunter, T., 1995. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 9, 576–596.

Hill, W.G., Weir, B.S., 1988. Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. 33, 54–78.

Hoerger, A.C., Ilyas, M., Stephan, W., Tellier, A., van der Hoorn, R.A.L., Rose, L.R., 2012. Balancing selection at the tomato RCR3 guardee gene family maintains variation in strength of pathogen defense. PLoS Gene. 8, e1002813.

Hudson, R.R., 2001. Two-locus sampling distributions and their application. Genetics 159, 1805–1817.

Jeffreys, H., 1961. Theory of Probability, third ed. Oxford University Press.

Jia, Y., Loh, Y.T., Zhou, J., Martin, G.B., 1997. Alleles of *Pto* and *Fen* occur in bacterial speck-susceptible and fenthion-insensitive tomato cultivars and encode active protein kinases. Plant Cell 9, 61–73.

Kelly, J.K., 1997. A test of neutrality based on interlocus associations. Genetics 146, 1197–1206.

Kimura, M., 1985. The role of compensatory neutral mutations in molecular evolution. J. Genet. 64, 7–19.

Kud, J., Zhao, Z., Du, X., Liu, Y., Zhao, Y., Xiao, F., 2013. SGT1 interacts with the Prf resistance protein and is required for Prf accumulation and Prf-mediated defense signalling. Biochem. Biophys. Res. Commun. 431, 501–505.

Lewontin, R., 1974. The Genetic Basis of Evolutionary Change. Columbia University Press, NY.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Loh, Y.T., Martin, G.B., 1995. The *Pto* bacterial resistance gene and the *Fen* insecticide sensitivity gene encode functional protein kinases with serine/threonine specificity. Plant Physiol. 108, 1735–1739.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganal, M.W., et al., 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262, 1432–1436.

Martin, G.B., Frary, A., Wu, T.Y., Brommonschenkel, S., Chunwongse, J., et al., 1994. A member of the tomato *Pto* gene family confers sensitivity to fenthion resulting in rapid cell death. Plant Cell 6, 1543–1552.

McVean, G., Awadalla, P., Fearnhead, P., 2002. A coalescent-based method for detecting and estimating recombination rates from gene sequences. Genetics 160, 1231–1241.

Mucyn, T.S., Clemente, A., Andriotis, V.M.E., Balmuth, A.L., Oldroyd, G.E.D., et al., 2006. The tomato NBARC-LRR protein Prf interacts with Pto kinase *in vivo* to regulate specific plant immunity. Plant Cell 18, 2792–2806.

Mucyn, T.S., Wu, A.-J., Balmuth, A.L., Arasteh, J.M., Rathjen, J.P., 2009. Regulation of tomato Prf by Pto-like protein kinases. Mol. Plant Microbe. Interact. 22, 391–401.

Ntoukakis, V., Mucyn, T.S., Gimenez-Ibanez, S., Chapman, H.C., Gutierrez, J.R., et al., 2009. Host inhibition of a bacterial virulence effector triggers immunity to infection. Science 324, 784–787.

Ntoukakis, V., Balmuth, A.L., Mucyn, T.S., Gutierrez, J.R., Jones, A.M.E., et al., 2013. The tomato Prf complex is a molecular trap for bacterial effectors based on Pto transphosphorylation. PLoS Pathog. 9, e1003123.

Ohta, T., 1982a. Linkage disequilibrium due to random genetic drift in finite subdivided populations. PNAS 79, 1940–1944.

Ohta, T., 1982b. Linkage disequilibrium with the island model. Genetics 101, 139–155.

Rathjen, J.P., Chang, J.H., Staskawicz, B.J., Michelmore, R.W., 1999. Constitutively active Pto induces a Prf-dependent hypersensitive response in the absence of avrPto. EMBO 18, 3232–3240.

Rose, L.E., 2002. Dissertation: the population genetics and functional analysis of the *Pto* disease resistance gene in *Lycopersicon* spp. and the *RPP13* gene in *Arabidopsis thaliana*. University of California, Davis.

Rose, L.E., Bittner-Eddy, P.D., Langley, C.H., Holub, E.B., Michelmore, R.W., Beynon, J.L., 2004. The maintenance of extreme amino acid diversity at the disease resistance gene, *RPP13*, in *Arabidopsis thaliana*. Genetics 166, 1517–1527.

Rose, L.E., Langley, C.H., Bernal, A.J., Michelmore, R.W., 2005. Natural variation in the *Pto* pathogen resistance gene within species of wild tomato (*Lycopersicon*). I. Functional analysis of *Pto* alleles. Genetics 171, 345–357.

Rose, L.E., Michelmore, R.W., Langley, C.H., 2007. Natural variation in the *Pto* disease resistance gene within species of wild tomato (*Lycopersicon*). II. Population genetics of *Pto*. Genetics 175, 1307–1319.

Rose, L.E., Grzeskowiak, L., Hoerger, A., Groth, M., Stephan, W., 2011. Targets of selection in a disease resistance network in wild tomatoes. Mol. Plant Pathol. 12, 921–927.

Rose, L.E., Atwell, S., Grant, M., Holub, E.B., 2012. Parallel loss-of-function at the *RPM1* bacterial resistance locus in *Arabidopsis thaliana*. Front. Plant Sci. 3.

Rosebrock, T.C., Zeng, L., Brady, J.J., Abramovitch, R.B., Xiao, F., Martin, G.B., 2007. A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity. Nature 448, 370–375.

Salmeron, J., Oldroyd, G., Rommens, C., Scofield, S., Kim, H.-S., et al., 1996. Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell 86, 123–133.

Scofield, S.R., Tobias, C.M., Rathjen, J.P., Chang, J.H., Lavelle, D.T., et al., 1996. Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. Science 274, 2063–2065.

Sessa, G., Martin, G.B., 2000. Signal recognition and transduction mediated by the tomato Pto kinase: a paradigm of innate immunity in plants. Microbes Infect. 2, 1591–1597.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., et al., 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science 270, 1804–1806.

Staedler, T., Arunyawat, U., Stephan, W., 2008. Population genetics of speciation in two closely related wild tomatoes (Solanum Section Lycopersicon). Genetics 178, 339–350.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A., Hinata, K., 2000. The S receptor kinase determines self-incompatibility in *Brassica stigma*. Nature 403, 913–916.

Tang, X., Frederick, R.D., Zhou, J., Halterman, D.A., Jia, Y., Martin, G.B., 1996. Initiation of plant disease resistance by physical interaction of AvrPto and Pto kinase. Science 274, 2060–2063.

Tanksley, S.D., Ganal, M.W., Prince, J.P., De Vicente, M.C., Bonierbale, M.W., et al., 1992. High density molecular linkage maps of the tomato and potato genomes. Genetics 132, 1141–1160.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370.

Weir, B.S., 1979. Inferences about linkage disequilibrium. Biometrics 35, 235–254.

Weir, B.S., 1996. Genetic Data Analysis II. Sinauer Associates, Sunderland, MA.

Whittam, T.S., Ochman, H., Selander, R.K., 1983. Geographic components of linkage disequilibrium in natural populations of *Escherichia coli*. Mol. Biol. Evol. 1, 67–83.

Wright, S., 1940. Breeding structure of populations in relation to speciation. Am. Nat. 74, 232–248.

Wu, A.-J., Andriotis, V.M.E., Durrant, M.C., Rathjen, J.P., 2004. A patch of surface-exposed residues mediates negative regulation of immune signaling by tomato Pto kinase. Plant Cell 16, 2809–2821.

Xing, W., Zou, Y., Liu, Q., Liu, J., Luo, X., et al., 2007. The structural basis for activation of plant immunity by bacterial effector protein AvrPto. Nature 449, 243–247.

Zaykin, D.V., Pudovkin, A., Weir, B.S., 2008. Correlation-based inference for linkage disequilibrium with multiple alleles. Genetics 180, 533–545.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9, 40.