Identification of Immunity Related Genes to Study the Physalis peruviana – Fusarium oxysporum Pathosystem

Felix E. Enciso-Rodríguez1, Carolina González2, Edwin A. Rodríguez2, Camilo E. López3, David Landsman4, Luz Stella Barrero1,5, Leonardo Mariño-Ramírez1,4,*

1 Plant Molecular Genetics Laboratory, Center for Biotechnology and Bioindustry (CBB), Colombian Corporation for Agricultural Research (CORPOICA), Bogotá, Colombia, 2 Molecular Microbiology Laboratory, Center for Biotechnology and Bioindustry (CBB), Colombian Corporation for Agricultural Research (CORPOICA), Bogotá, Colombia, 3 Laboratorio de Fitopatología Molecular, Departamento de Biología, Universidad Nacional de Colombia, Bogotá, Colombia, 4 Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA, 5 PanAmerican Bioinformatics Institute, Santa Marta, Magdalena, Colombia

Abstract

The Cape gooseberry (Physalis peruviana L.) is an Andean exotic fruit with high nutritional value and appealing medicinal properties. However, its cultivation faces important phytosanitary problems mainly due to pathogens like Fusarium oxysporum, Cercospora physalidis and Alternaria spp. Here we used the Cape gooseberry foliar transcriptome to search for proteins that encode conserved domains related to plant immunity including: NBS (Nucleotide Binding Site), CC (Coiled-Coil), TIR (Toll/Interleukin-1 Receptor). We identified 74 immunity related gene candidates in P. peruviana which have the typical resistance gene (R-gene) architecture, 17 Receptor like kinase (RLKs) candidates related to PAMP-Triggered Immunity (PTI), eight (TIR-NBS-LRR, or TNL) and nine (CC–NBS-LRR, or CNL) candidates related to Effector-Triggered Immunity (ETI) genes among others. These candidate genes were categorized by molecular function (98%), biological process (85%) and cellular component (79%) using gene ontology. Some of the most interesting predicted roles were those associated with binding and transferase activity. We designed 94 primers pairs from the 74 immunity-related genes (IRGs) to amplify the corresponding genomic regions on six genotypes that included resistant and susceptible materials. From these, we selected 17 single band amplicons and sequenced them in 14 F. oxysporum resistant and susceptible genotypes. Sequence polymorphisms were analyzed through preliminary candidate gene association, which allowed the detection of one SNP at the PpIRG-63 marker revealing a nonsynonymous mutation in the predicted LRR domain suggesting functional roles for resistance.

Citation: Enciso-Rodríguez FE, González C, Rodríguez EA, López CE, Landsman D, et al. (2013) Identification of Immunity Related Genes to Study the Physalis peruviana – Fusarium oxysporum Pathosystem. PLoS ONE 8(7): e68500. doi:10.1371/journal.pone.0068500

Editor: I. King Jordan, Georgia Institute of Technology, United States of America

Received November 07, 2012; Accepted May 30, 2013; Published July 3, 2013

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: Support for this research was provided by a grant from the Colombian Ministry of Agriculture Contract Nos. 054/08072-2008L4787-3281 to Luz Stella Barrero and 054/08190-2008L7522-3322 to Victor Manuel Nuñez Zarantes. This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Library of Medicine, and National Center for Biotechnology Information. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: marino@ncbi.nlm.nih.gov

Introduction

The Cape gooseberry (Physalis peruviana L.) is an exotic fruit from the Solanaceae family native to the Andean region which has spread to other parts of the world including Africa and India [1,2]. In addition to its high contents of vitamin A, C, B-complex [3] and minerals like iron and phosphorous [4,5], this tropical fruit is also known for its antioxidant [6], anticancer [3,7], anti-inflammatory [8,9], as well as diabetes and hypertension control properties [10]. Therefore the Cape gooseberry provides an enormous potential for biomedical research and commercial purposes [11]. Colombia is the first Cape gooseberry world producer followed by Zimbabwe, Malaysia, China, Kenya and South Africa [12]. However, Cape gooseberry production in the Andean region faces important phytosanitary problems due mainly to fungal diseases causing great crop losses and consequently a significant reduction in yield and quality. In Colombia, the most important fungal disease is the vascular wilt caused by Fusarium oxysporum, which may generate total crop losses [13]. Although the molecular defense mechanisms underlying plant–pathogen interactions have been studied extensively in a variety of pathosystems [14], little is known about the mechanisms underlying the P. peruviana – F. oxysporum interaction.

The current molecular model for plant immunity indicates the presence of two lines of defense: in the first one, plants sense
conserved microbial molecular signatures called Pathogen-Associated Molecular Patterns (PAMPs) or Microbial-Associated Molecular Patterns (MAMPs) [15]. The recognition of these molecules is mediated by Pattern Recognition Receptors (PRRs), which are cell surface-located transmembrane receptors [15,16]. PRRs consist of extracellular Leucine-Rich Repeats (LRR) or Lysine Motif (LysM) domains. In some cases an intracellular Ser/Thr kinase domain is present [17,18,19]. This first level of recognition is known as PAMP-Triggered Immunity or PTI [20] and it can confer broad-spectrum resistance and durable resistance against different types of non-host or non adapted pathogens [21]. However, successful pathogens evade or suppress the PTI through the translocation of effector proteins into the plant cell [16]. In turn, plants have evolved to directly or indirectly recognize such effectors through additional receptors called resistance (R) proteins [22]. This second layer of defense is known as Effector-Triggered Immunity (ETI), and is often accompanied by Hypersensitive Response (HR), blocking pathogen spread [23]. The majority of R proteins have a modular structure characterized by a conserved central Nucleotide Binding Site (NBS) and a more variable C terminal LRR domain [24,25]. The amino-terminal region present in NBS-LRR proteins has allowed their classification into two major classes defined by the presence of Toll/Interleukin-1 Receptor (TIR) or Coiled Coil (CC) domains [25,26,27]. Sets of candidate disease resistance genes have been identified in model species with complete genomes using the domains described above. Accordingly, 149 NBS-LRRs genes have been identified in Arabidopsis, 875 in Rice and 400 in Populus [28,29,30]. Additionally, the presence of the NBS domain in non-model species has become critical for the association of genomic sequences with disease resistance function [31,32,33,34,35].

Expressed Sequenced Tags (ESTs) analyses are critical for the discovery of novel genes like those involved in PTI and ETI, particularly in non-model plants for which a complete genome is currently not available as in the case of the Cape gooseberry. We recently characterized the Cape gooseberry foliar transcriptome and have developed microsatellite markers for plant improvement in this species [36,37]. This manuscript provides the first description of a computational strategy to identify putative immunity related genes (IRGs) encoded in the P. peruviana foliar transcriptome used to design primers to PCR amplify a number of IRG fragments in a pool of F. oxysporum resistant and susceptible genotypes. We were able to identify a set of polymorphisms (SNPs) initially related to resistance, this finding is the basis for further experimentation that will include larger populations and markers to contribute future marker assisted selection in Cape gooseberry.

Materials and Methods

1: Plant material and DNA isolation

Cape gooseberry and related taxa genotypes (Table 1) were selected from the Colombian Corporation for Agricultural Research (CORPOICA) germ plasm collection, based on resistance and susceptibility responses against F. oxysporum (Table S1). Plants were propagated in vitro and maintained in growing chambers at 20°C. Genomic DNA was isolated from young leaves, following a previously described methodology [38] with some modifications. Briefly, approximately 500 mg young leaves were ground using a mortar and pestle with liquid nitrogen and incubated at 65°C for 30 min in extraction buffer (0.15M Tris-HCl pH 8, 0.01M EDTA pH 8, 1% CTAB, 1% PVP 360000 and 5% β-mercaptoethanol). Samples were centrifuged at 18,000 g and the supernatant was extracted twice with one volume of phenol-chloroform-isooamyl alcohol (25:24:1). Next, the supernatant was precipitated with isopropyl alcohol and washed with 70% ethanol. Pellets were dissolved in 1X TE buffer, treated with RNase A (10 mg/mL) at 37°C for 20 min and stored at -20°C until use. The quantity and quality of DNA was checked using a NanoDrop® ND-1000 Spectrophotometer and on 1% (w/v) agarose gels stained with ethidium bromide (0.5 μg/mL).

2: Evaluation of the resistance phenotype

Fifteen genotypes from P. peruviana and related taxa (Table 1) were challenged with a F. oxysporum pathogenic strain (Map 5) isolated from P. peruviana and supplied by the Fusarium Collection at Corpoica’s Molecular Microbiology Laboratory. For inoculum production, the monosporic strain Map 5 was reactivated in Potato Dextrose Agar (PDA) medium (BD Franklin Lakes, NJ) for 15 days at 28°C. Then, it was grown in liquid Potato Dextrose Broth (PDB) (BD Franklin Lakes, NJ) for ten days at 28°C in constant agitation (140 rpm); the inoculum was prepared according to Namiki et al. [39] and adjusted to desired final concentration. Once P. peruviana seedlings had a pair of true leaves and were 5 to 7 cm tall, 12 seedlings of each genotype were transplanted individually into plastic pots with 255 g sterilized soil-rice husk substrate 3:1 ratio and were then sowed in a completely randomized design under field conditions in the year 2011 in Mosquera, Cundinamarca, Colombia located at 2,516 meters above sea level. Plants were inoculated with a conidial suspension using one as a control in sterile water. The inoculation was done by the root dip method [39]. Briefly, the roots were dipped in 75 mL of spore suspension (1x10⁶ CFU/mL) for three minutes and were transplanted into the same pots. External symptoms were scored 2 weeks after inoculation during 45 days. The severity degree of the disease was registered daily using a scale of symptoms proposed for the pathosystem Physalis peruviana-Fusarium oxysporum (Supplementary Table S1). The scale of symptoms was based in other scales [39,40,41,42].

3: Computational identification of candidate resistance genes (Immunity Related Genes-IRGs)

A complete set of plant protein sequences was obtained from the NCBI Protein (http://www.ncbi.nlm.nih.gov/protein) and the SOL Genomics Network (SGN) [43] databases. All proteins retrieved were clustered using Blastclust (ftp://ftp.ncbi.nih.gov/blast/) to eliminate redundancy. We used the Pfam database [44] to retrieve a HMM profiles from NB-ARC (PF00931), TIR (PF01582), LysM (PF01476) and the Pkinase domain (PF00069). These profiles were used to search all plant proteins using HMMER version 3.0 [45], and the predicted...
The collection of the Cape gooseberry and related taxa was used in this study to identify possible intron and exon regions using NCBI’s Spidey (http://www.ncbi.nlm.nih.gov/spidey) for primer design at intron-exon boundaries in all immunity related transcripts. Primers were designed using Primer3 [52] and subsequently checked for self-complementarity, hairpins and dimers. PCR amplification was carried out in an i-Cycler thermal Cycler (Bio-Rad, Hercules, CA, USA) as follows: one cycle of initial denaturation for 5 min at 94°C, followed by 35 cycles for 30 s at 94°C, 1 min at 54°C and 2 min at 72°C, followed by a final extension of 10 min at 72°C, and preserved at 4°C until further analysis. To verify PCR amplification success of the designed primers, we tested them on six Cape gooseberry and related taxa genotypes with variable resistance responses (Table 1). Amplification products were separated by gel electrophoresis on 2% (w/v) agarose in a 1X TAE buffer (40 mM Tris-acetate and 1 mM EDTA), and then stained with ethidium bromide (0.5 μg/mL).

6: PCR sequencing and analysis

Seventeen PCR products from IRGs were selected based on single band amplification on the six genotypes tested, and were sequenced initially by single pass, using forward and reverse primers for each amplicon in two of the six genotypes (09U47-1 and 09U89-1) using the same PCR conditions as described above with commercial Taq DNA polymerase (Invitrogen, San Diego, CA). Each read was processed with a variant call pipeline [53] that uses Phred/Phrap/Consed for base calling and assembly [54,55].

Then, we amplified the 17 IRGs on the set of 15 genotypes from 14 accessions (Table 1), and sequenced them using the 454 GS-FLX Titanium platform (Roche Diagnostics Corporation). The raw reads were deposited in the NCBI Sequence Read Archive (accession number SRX216233). Reads for all genotypes and markers were trimmed using Mothur [56], aligned with BWA [57], using as consensus the sequence produced by Sanger sequencing. SNP calling was achieved using Samtools [58] with default parameters. We performed a General Lineal Model (GLM) procedure using the Tassel version 3.0.146 [59] with a Minimum Allele Frequency (MAF) of 0.1 and removing minor SNP states in order to avoid false positives. The analyses took into account the structure and ancestry of the sample population by using Q matrix for each genotype generated by Structure 2.3.3 [60]. The K parameters ranged from K = 1 to K = 10 across ten runs with 106 iterations and a burning period of 500,000. Bonferroni

4: Annotation of candidate resistance genes in Cape gooseberry

The Cape gooseberry resistance and immunity related transcripts identified above were compared with the UniProtKB/Swiss-Prot database [48], using BLASTX [49]. The top five protein hits for each query were submitted for functional classification using Blast2GO [50]. GO terms were assigned employing three Gene Ontology categories: cellular component, molecular function and biological process [51].

5: Primer design and PCR amplification

We used the Solanum lycopersicum genome and the Cape gooseberry mRNA to identify possible intron and exon regions using NCBI’s Spidey (http://www.ncbi.nlm.nih.gov/spidey) for primer design at intron-exon boundaries in all immunity related transcripts. Primers were designed using Primer3 [52] and subsequently checked for self-complementarity, hairpins and dimers. PCR amplification was carried out in an i-Cycler thermal Cycler (Bio-Rad, Hercules, CA, USA) as follows: one cycle of initial denaturation for 5 min at 94°C, followed by 35 cycles for 30 s at 94°C, 1 min at 54°C and 2 min at 72°C, followed by a final extension of 10 min at 72°C, and preserved at 4°C until further analysis. To verify PCR amplification success of the designed primers, we tested them on six Cape gooseberry and related taxa genotypes with variable resistance responses (Table 1). Amplification products were separated by gel electrophoresis on 2% (w/v) agarose in a 1X TAE buffer (40 mM Tris-acetate and 1 mM EDTA), and then stained with ethidium bromide (0.5 μg/mL).

6: PCR sequencing and analysis

Seventeen PCR products from IRGs were selected based on single band amplification on the six genotypes tested, and were sequenced initially by single pass, using forward and reverse primers for each amplicon in two of the six genotypes (09U47-1 and 09U89-1) using the same PCR conditions as described above with commercial Taq DNA polymerase (Invitrogen, San Diego, CA). Each read was processed with a variant call pipeline [53] that uses Phred/Phrap/Consed for base calling and assembly [54,55].

Then, we amplified the 17 IRGs on the set of 15 genotypes from 14 accessions (Table 1), and sequenced them using the 454 GS-FLX Titanium platform (Roche Diagnostics Corporation). The raw reads were deposited in the NCBI Sequence Read Archive (accession number SRX216233). Reads for all genotypes and markers were trimmed using Mothur [56], aligned with BWA [57], using as consensus the sequence produced by Sanger sequencing. SNP calling was achieved using Samtools [58] with default parameters. We performed a General Lineal Model (GLM) procedure using the Tassel version 3.0.146 [59] with a Minimum Allele Frequency (MAF) of 0.1 and removing minor SNP states in order to avoid false positives. The analyses took into account the structure and ancestry of the sample population by using Q matrix for each genotype generated by Structure 2.3.3 [60]. The K parameters ranged from K = 1 to K = 10 across ten runs with 106 iterations and a burning period of 500,000. Bonferroni
corrections were performed to establish \( P \) value cutoffs of \( \alpha = 0.01 \) and \( \alpha = 0.05 \).

**Results**

**Generation of a plant immunity related protein database**

We generated a curated plant immunity database with a total of 3,691 proteins using a methodology described in Figure 1. These proteins were identified mainly in model species like *Populus trichocarpa*, *Vitis vinifera* and *Oryza sativa* as well as other non-model species like *Solanum demissum*, *Solanum bulbocastanum* (Figure 2). Nearly all proteins contain NBS domains (3,456), which are associated with TIR (590), CC (959) or Pkinase (61) domains, and others were associated with LysM and LRR (62 and 1017 respectively, Table 2). Notice that one of the most abundant architectures presented here is CC–NBS-LRR, known to be associated with Effector-Triggered Immunity (ETI).

**Identification of transcripts related to plant immunity in Cape gooseberry**

Similarity TBLASTN searches using the plant immunity related protein database described here were carried out on the Cape gooseberry transcriptome with an e-value cutoff of 1e-4 [49], obtaining a total of 74 isotigs with hits (Table 3). We successfully identified immunity related transcripts similar to others reported in model species even though the tissue used was not inoculated with a pathogen. A total of 42 isotigs were associated with the NBS domain, 11 with TIR, 15 with CC, 48 with LRR, 21 with Pkinase and 5 with LysM. Therefore, we report 19 candidate transcripts encoding typical domains related with the first layer of defense (PTI). Among the PTI candidate transcripts, 17 of them (14 LPk and three LysMPk) were identified as RLKs (Receptor like kinase), and only two (LysM) as RLP (Receptor like protein), which have been reported as an important PRR to recognize fungal pathogens [61] like *F. oxysporum*. Regarding the second layer of immunity or ETI, we identified a total of 45 isotigs associated mainly with the NBS domain. Among these genes, 10 were identified as NL (NBS-LRR), eight as TNLs (TIR-NBS-LRR), and nine as CNLs (CC-NBS-LRR). In addition, 18 candidate transcripts that lack a...
canonical architecture typical of resistance genes were classified as N with 10 candidates, T with two, C with one, CN with four as well as the non-common architecture TCNL with one transcript (Table 3).

Functional annotation of resistance related transcripts

We assigned functional gene ontology categories [51] to all 74 isotigs with sequence similarity to immunity related proteins based on the presence of conserved domains. Each isotig was compared to the UniProtKB/Swiss-Prot database and GO terms were assigned to each query using Blast2GO [50]. A total of 1047 ontologies were associated with 68 sequences annotated and assigned to the three principal categories. For Biological process, 85% of genes were associated mainly with response to stress and regulation of biological process. For Molecular function and Cellular component, 98% and 79% genes were associated mainly with transferase activity and small molecule binding for the first one, and cell organelle and cytoplasm for the second one (Figure 3). The remaining six sequences did not show similarity to known proteins and were not assigned to any GO categories.

Identification of polymorphisms at IRGs and their relationship to *F. oxysporum* resistance

We designed 94 primer pairs from 74 isotigs with sequence similarity to immunity related proteins. In some cases we designed additional primer pairs per isotig in which more than one intron-exon boundaries were predicted. A total of 85 primer pairs (90%) produced amplification products on six genotypes (Supplementary Table S2).

From the 17 454-sequenced IRGs on 15 genotypes, we could process 14 of them, since the genotype 09U178-4 had more than 60% missing data probably due to the fact that it belongs to another genus (*Solanum*), therefore, it was eliminated from further analyses. Read coverage for each amplicon ranged from 0 to 401 reads per IRG (Supplementary Table S3). We obtained a multi-fasta alignment with an average 7804 nt per genotype and 109,256 nt over all sequences, identifying a total of 213 SNPs filtered with a MAF = 0.1.

To reduce spurious associations between the SNPs from IRGs and the resistance trait, we performed structure analyses on the tested population sample. The analyses revealed two subpopulations using the ΔK method [62]; the first was formed...
by related taxa (P. floridana and P. angulata genotypes) and the second clustered all the P. peruviana genotypes. The clusters formed were not related to resistance against F. oxysporum nor to the place of origin or state of cultivation, they were only related at the species level. Next, preliminary candidate gene association analyses were performed using GLM from structure results as well as the SNP and resistance phenotype data. After Bonferroni correction, significant association was found in one SNP (SNP_483), with a marker \( p = 1.72E-6 \) at \( \alpha = 0.01 \) level (\( p \) value cutoff = 4.69E-5) and \( \alpha = 0.05 \) level (\( p \) value cutoff = 2.34E-4). This SNP is located at the PpIRG-63 corresponding to a CC–NB-LRR predicted protein RGA3-like (Figure 4).

### Discussion

Cape gooseberry (\( P. \) peruviana \( L. \)) is an important commercial crop with high nutritional value and interesting medicinal properties. This study constitutes an initial effort for future improvement of Cape gooseberry cultivars against pathogens like \( F. \) oxysporum. Molecular breeding to improve Cape gooseberry materials for resistance or tolerance against different diseases require the identification and isolation of immunity genes for incorporation into susceptible commercial cultivars using different strategies such as marker assisted selection, transgenesis or directed mutagenesis, which might advance towards genome wide selection [63], as more resistant phenotype and marker data becomes available for the species. These strategies are still under development for this orphan species, however existing transcriptome data available [37] is being used effectively to identify markers and immunity related genes for future development efforts.

Immunity related genes, which confer resistance to diverse pathogens like fungi, bacteria and viruses have conserved domains where the most common are LRR, LysM and kinase domains present in the PRR proteins involved in PTI. The R proteins are characterized by the presence of LRR and NBS domains that can be accompanied by a TIR or a CC domain located at their N-termini. The presence of conserved domains, particularly NBS have allowed development of strategies based on degenerate primers to amplify genes related to plant immunity in different species [31,32,34]. With the expansion of complete genome sequences or transcriptomes it is possible to identify genes coding for proteins with these conserved domains [64]. The number of these types of genes is variable in different species. In \( Arabidopsis \) for example there are 149 NBS-LRR genes, 875 in rice, 738 in potato, 400 in \( Populus \), 630 in sunflower, 92 in \( Brassica \) among others [29,30,65,66,67,68].

The sequence alignments of the TIR-NBS domain were used to design degenerate primers to identify genes coding for the TIR-LRR and TIR-NBS domains. In this study we identified 74 transcripts related to plant immunity where the majority are represented by NBS and LRR domains (58.1% respectively). It is possible that some transcripts do not represent the full-length mRNA or the complete plant resistance gene architecture. This statement is raised because of the fact that most of the immunity related isoligs reported here had a size less than 4,640 bp (Figure 5). Thus, in most cases we were only able to identify incomplete proteins with classic TNLs or CNLs domain architectures. Nonetheless, these gene fragments were still useful for studying function \textit{in silico} and for preliminarily identifying polymorphisms related to resistance.

Most immunity genes, notably those coding for NBS-LRR proteins, are constitutively expressed [69]. Although we used foliar tissue that is not usually infected by \( F. \) oxysporum, we successfully identified candidates with typical plant resistance gene architectures that are useful for initial screenings of different resistant and susceptible sources in germ plasm collections. Some of the genes related to plant immunity

### Table 2. Protein architecture of plant resistance genes identified in a plant protein database created from databases at NCBI and SGN.

| Protein Architecture       | Number of proteins |
|----------------------------|--------------------|
| LRR                        | 128                |
| LRR-Pkinase                | 11                 |
| LysM                       | 29                 |
| LysM-Pkinase               | 33                 |
| NBS                        | 1596               |
| NBS-Pkinase                | 6                  |
| TIR                        | 33                 |
| TIR-NBS                    | 115                |
| NBS-LRR                    | 392                |
| NBS-LRR-Pkinase            | 3                  |
| TIR-NBS-LRR                | 384                |
| TIR-CC-NBS                 | 17                 |
| TIR-CC-NBS-LRR             | 40                 |
| TIR-NBS-LRR-Pkinase        | 1                  |
| CC–NBS-LRR                 | 452                |
| CC–NBS-Pkinase             | 4                  |
| CC–NBS                     | 446                |
| Pkinase                    | 3                  |
| **Total**                  | **3,691**          |

### Table 3. Predicted domain architecture of resistance gene candidates in \( P. \) peruviana related to the first and second layer of defense.

| Predicted Protein Domains | Letter code | Isotigs with Hits | Defense layer |
|---------------------------|-------------|-------------------|---------------|
| LRR-Pkinase               | Lpk         | 14                | PTI           |
| LysM                      | LysM        | 2                 |               |
| LysM-Pkinase              | LysMPk      | 3                 |               |
| NBS                       | N           | 10                | ETI           |
| NBS-LRR                   | NL          | 10                |               |
| TIR                       | T           | 2                 |               |
| TIR-NBS-LRR               | TNL         | 8                 |               |
| TIR-CC-NBS-LRR            | TCNL        | 1                 |               |
| CC                        | C           | 1                 |               |
| CC–NBS                    | CN          | 4                 |               |
| CC–NBS-LRR                | CNL         | 9                 |               |
| **Other***                |             | 1                 |               |
| **Total**                 |             | 74                |               |

* Additional domain LRR (six isoligs) and Pkinase (four isoligs) that might be implicated in processes other than defense.
reported here, are widely recognized for their response against plant pathogens in model plants (Table 4). Genes like Prf, Tm2 and BS4 had the lowest TBLASTN E-values indicating high similarity to their Cape gooseberry counterparts.

In tomato, the I2 gene confers full resistance against *F. oxysporum lycopersici* race 2 [70]. I2 is present as a monophyletic group in the Solanaceae family, leading to the conservation of binding and hydrolysis functions at the NBS domain, suggesting that its function is maintained in many members of the I2 family [71]. Besides, the relatively slow rate of birth-and-death in the I2 family and because of the fact that new I2 duplicates acquire novel functions or become

---

Figure 3. Functional distribution of resistance related isotigs based on gene ontologies: molecular function, biological processes and cellular component in the Cape gooseberry transcriptome. Only major hits are shown (E-value ≤ 1E-4) for each GO category.

doi: 10.1371/journal.pone.0068500.g003

Figure 4. Single nucleotide polymorphisms (SNPs) detected in the PpIRG-63 marker located on the predicted LRR domain. Polymorphisms are shown in red.

doi: 10.1371/journal.pone.0068500.g004
specialized representing functional copies on distant phylogenetic species from potato to tobacco [71]; we might expect that the homologous I2 gene found in Cape gooseberry could be functional and possibly detect effectors secreted by *F. oxysporum* on *P. peruviana*. Similarly, we found homologs for Prf, BS4, and N identified in Solanaceae species (Prf and BS4 in *S. lycopersicum* and N in *Nicotiana glutinosa*, Table 4), therefore we might expect as with I2, their conservation in Cape gooseberry as functional PRR, R genes and defense proteins like the ones in tomato and other Solanaceae species since this family is characterized by a high conservation of gene content and order and affected by many of the same pathogens [72].

On the other hand, our results are consistent with the role and localization of the proteins involved in plant resistance, as evidenced by the functional annotation using Blas2GO. For example, the NBS domain, characteristic of the R proteins is involved in plant defense through direct or indirect recognition of virulence factors [24]. This domain is required for ATP and GTP binding [73], as reported for I2 and Mi-1 in tomato modulating the binding and hydrolysis of ATP in the signaling cascade of resistance [74]. Additionally, the NBS domain has been related to nucleotide phosphatase activity in a subset of R genes and seems that its biochemical activity co-evolved with the plant resistance signaling pathways [75]. On the other hand, RLKs are related to transmembrane signaling receptor activity, through PAMP/MAMPs recognition trigger defense responses [15].

Here we report one SNP mapped to the PpIRG-63 marker, preliminary associated with the resistance response against *F. oxysporum* (Figure 4). This gene is involved in the second layer of plant resistance and has a typical R gene architecture CC–NB-LRR (Supplementary Table S2), representing a potential plant immunity gene in Cape gooseberry. The SNP_438 is present at the exonic region in the predicted LRR domain, which is believed to be the major determinant of effector recognition [77]. In several plant–pathogen systems, the sequence variation in the LRR domain, particularly in the β stand/β turn motif (xxLxxLxx motif) has been

---

**Table 4. Resistance genes reported in model organisms with significant hits in the Cape gooseberry transcriptome.**

| Protein ID  | Domains     | Isotig   | Gene/Protein Name | Species name | BLAST E-value |
|------------|-------------|----------|-------------------|--------------|---------------|
| 4689223    | CC–NBS-LRR  | JO142447 | I2                | *Solanum lycopersicum* | 1E-102        |
| 8547237    | CC–NBS      | JO133481 | Prf               | *Solanum lycopersicum* | 0             |
| 38459219   | TIR-NBS     | JO132049 | BS4               | *Solanum lycopersicum* | 2E-142        |
| 558887     | TIR-NBS-LRR | JO129083 | N                 | *Nicotiana glutinosa* | 8E-7          |
| 56406364   | CC–NBS      | JO138325 | Tm-2 ToMV resistant protein | *Solanum lycopersicum* | 2E-163        |

§. GenBank protein ID.

---

**Figure 5.** Isotig size distribution for immunity related genes in Cape gooseberry.

doi: 10.1371/journal.pone.0068500.g005
shown to be responsible for different recognition or resistance specificities [78,79]. As demonstrated in rice, the knowledge of the variation patterns of R-genes through the study of allelic diversity in NBS-LRR genes, is of fundamental importance for plant breeders attempting to preserve resistant germplasm [80]. These observations bolster the possibility that the SNP polymorphism present in the PpiRG-63 related with plant immunity could be associated with pathogen recognition in P. peruviana.

Interestingly, the SNP described here is a non-synonymous substitution (valine to glycine) and both encode nonpolar amino acids. This kind of substitution is frequently present in NBS-LRR proteins maintained by diversifying selection [25] as a possible response to new variants of the pathogen. Whether this polymorphism is causal to resistance against F. oxysporum awaits further investigation through association studies with larger populations and markers or functional approaches.

Conclusion

This study is the first report on the discovery of genes putatively related to plant immunity in Cape gooseberry. We reported 74 genes related to the first and second layer of plant pathogen recognition found in the P. peruviana foliar transcriptome. We identified genes with the typical R architecture and found 17 RLKs candidates related with PTI, eight TNLs and nine CNLs candidates related with ETI. Functional annotation using gene ontologies predict their roles in resistance against plant pathogens. Ninety-four primers were designed from these candidate genes, but no InDels (>50 bp) were found between resistant and susceptible plant genotypes. We sequenced, 17 IRGs to perform preliminary candidate gene association analyses on a set of 14 genotypes. One marker (PpiRG-63) revealed a non-synonymous SNP polymorphism (SNP_438) in the LRR predicted domain and was preliminary associated with resistance to F. oxysporum, representing the first pathogen resistance candidate gene known in this pathosystem. Further association analyses using larger population sizes would be needed to determine its function as molecular marker for breeding and phytosanitary programs in P. peruviana.

References

1. Popeneo H, King SR, Leon J, Kalinowski LS (1990) Goldenberry (cape gooseberry). In: NR Council. Lost Crops of The Incas: Little-known Plants of the Andes with Promise for Worldwide Cultivation. Washington DC: National Academy Press. pp. 241–252.
2. Menzel MY (1951) The Cytotaxonomy and Genetics of Physalis. Proc Am Philos Soc 95: 132-183.
3. Yen CY, Chiu CC, Chang FR, Chen JY, Hwang CC et al. (2010) 4beta-Hydroxywithanolide E from Physalis peruviana (golden berry) inhibits growth of human lung cancer cells through DNA damage, apoptosis and G2/M arrest. BMC Cancer 10: 46. doi:10.1186/1471-2407-10-46. PubMed: 20167063.
4. Rodrigues E, Rockenbach II, Cataneo C, Gonzaga LV, Chaves ES et al. (2006) Minerals and essential fatty acids of the exotic fruit Physalis peruviana L. Cienc Tecnologia Alimentos 29: 642-645. doi:10.1590/S0101-20612006000300029.
5. Martínez F, Sarmiento J, Fischer G, Jiménez F (2008) Effect of N, P, K, Ca, Mg and B deficiency on production and quality components of cape gooseberry (Physalis peruviana L.). Agron Colomb 26: 389-398.
6. Wu SJ, Ng LT, Huang YM, Lin DL, Wang SS et al. (2005) Antioxidant activities of Physalis peruviana. Biol Pharm Bull 28: 963-966. doi:10.1248/bpb.28.963. PubMed: 15930727.
7. Wu SJ, Chang SP, Lin DL, Wang SS, Hou FF et al. (2009) Supercritical carbon dioxide extract of Physalis peruviana induced cell cycle arrest and apoptosis in human lung cancer H661 cells. Food Chem Toxicol 47: 1132-1138. doi:10.1016/j.fct.2009.01.044. PubMed: 19425186.
8. Franco LA, Maitz GE, Calle J, Pinzón R, Ospina LF (2007) [Antinflammatory activity of extracts and fractions obtained from Physalis peruviana L. calyces]. Biomedica 27: 110-115. PubMed: 17546228.
9. Martínez W, Ospina LF, Granados D, Delgado G (2010) In vitro studies on the relationship between the anti-inflammatory activity of Physalis peruviana extracts and the phagocytic process. Immunopharmacol Immunotox 32: 63-73. doi:10.3109/08923970903143957. PubMed: 19678736.
10. Pinto Mda S, Ranilla LG, Apostolidis E, Lajolo FM, Genovese MI et al. (2009) Evaluation of antihyperglycemia and antihyperpertension potential of native Peruvian fruits using in vitro models. J Med Food 12: 278-291. doi:10.1089/jmf.2008.0113. PubMed: 19497277.
11. Ramadan MF (2011) Bioactive phytochemicals, nutritional value, and functional properties of cape gooseberry (Physalis peruviana): An overview. Food Res Int 44: 1830-1836. doi:10.1016/j.foodres.2010.12.042.
12. Bonilla M, Hersilia, Arias P, Antonio, Landínez L, Marcela, Moreno J, Manue, Cardozo F, et al (2009) Agenda prospectiva de investigación y desarrollo tecnológico para la cadena productiva de la uchuva en freco para exportación en Colombia- Ministerio De Agricultura y Desarrollo Rural; Proyecto Transición De La Agricultura; Universidad Nacional De Colombia CDDIAc, editor BOGOTÁ D.C., Colombia: Ministerio De Agricultura y Desarrollo Rural

13. Bernal A, Cotes AM, Navas A, Rodríguez A, López C et al. (2013) Generación de valor para el desarrollo competitivo del cultivo de la uchuva como recurso de biodiversidad de frutas en Colombia. In: AM CotesLS Barrerof RodríguezMV ZuluagaH Arevalo, Bioprospección para el desarrollo del sector agropecuario de Colombia. Bogotá, D.C. Produmedios.

14. Schuler C, Collmer A (2010) Studying plant-pathogen interactions in the genomes era: beyond molecular Koch’s postulates to systems biology. Annu Rev Phytopathol 48: 457-479. doi:10.1146/annurev-phyto-073009-114411. PubMed: 20687334.

15. Ziegel C (2008) Pattern-recognition receptors in plant innate immunity. Curr Opin Immunol 20: 10-16. doi:10.1016/j.coi.2007.11.003. PubMed: 18263630.

16. Rafiqi M, Bernoux M, Ellis KG, Dodds PN (2009) in the trenches of plant pathogen recognition: Role of NB-LRR proteins. Semin Cell Dev Biol 20: 1017-1024. doi:10.1016/j.semcdb.2009.04.010. PubMed: 19398031.

17. Gómez-Gómez L, Boller T (2000) FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol Cell 5: 1003-1011. doi:10.1016/S1097-2765(00)00265-8. PubMed: 10911994.

18. Wan J, Zhang XC, Neece D, Ramonnell KM, Clough S et al. (2008) A LyA receptor-like kinase plays a critical role in chitin signaling and fungal resistance in Arabidopsis. Plant Cell 20: 471-481. doi:10.1105/tpc.107.056754. PubMed: 18263776.

19. Postel S, Kammerling B (2009) Plants systems for recognition of pathogen-associated molecular patterns. Semin Cell Dev Biol 20: 1005-1031. doi:10.1016/j.semcdb.2009.06.002. PubMed: 19540353.

20. Jones JD, Dangl JL (2006) The plant immune system. Nature 444: 424-432. doi:10.1038/nature05286. PubMed: 17108957.

21. Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ (2009) Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts disease caused by Fusarium oxysporum f. sp. sesami. Tunisian J Plants Protect 4: 29-39.

22. Estupiñan R, Ossa A (2007) Efecto del agente causal de la marchitez vascular de la uchuva (Physalis peruviana) en el hongo Fusarium oxysporum Schltldl, sobre algunas solanáceas y otras especies cultivadas afectadas por formas especiales del microorganismo. Pontificia Universidad Javeriana. p. 89.

23. BLAST (1997) Sistema exportado para la evaluación de gernoplasma de frjol; PM Schoohoven Av. Cali.

24. Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R et al. (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. Plant Physiol 138: 1310-1317. doi:10.1104/pp.105.060707. PubMed: 16010005.

25. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J et al. (2012) The Pfam protein families database. Nucleic Acids Res 40: D290-D301. doi:10.1093/nar/gkr1065. PubMed: 22127870.

26. Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. Genome Inform Int Conference Genome Informatics 23: 205-211.

27. Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252: 1162-1164. doi:10.1126/science.252.5009.1162. PubMed: 2301185.

28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410. doi:10.1016/0022-2836(90)90570-k. PubMed: 22217972.

29. UniProt (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt), Nucleic Acids Res 40: D71-D75. doi:10.1093/nar/gkr060. PubMed: 22102690.

30. Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. Genome Inform Int Conference Genome Informatics 23: 205-211.

31. Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252: 1162-1164. doi:10.1126/science.252.5009.1162. PubMed: 2301185.

32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410. doi:10.1016/0022-2836(90)90570-k. PubMed: 22217972.

33. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H et al. (2000) Gene Ontology: tool for the unification of biology. Gene Ontology Consortium Nat Genet 25: 25-29.

34. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365-386. PubMed: 10420757.

35. Bedoya-Reina GC, Barrero LS (2010) Preliminary assessment of COSII gene diversity in lulo and a relative species: initial identification of genes potentially associated with domestication. Gene 458: 27-36. doi:10.1016/j.gene.2010.03.005. PubMed: 20302928.

36. Gordon D (2003) Reviewing and editing assembled sequences using Consed. Curr Protoc Bioinforma Chapter 11: Unit11 12. PubMed: 1842695.
55. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy Assess Genome Res 8: 175-185.

56. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75: 7537-7541. doi:10.1128/AEM.01541-09. PubMed: 19801484.

57. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589-595. doi:10.1093/bioinformatics/btp698. PubMed: 20080505.

58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079. doi:10.1093/bioinformatics/btp352. PubMed: 19505943.

59. Bradbury PJ, Zhang Z, Kroon DE, Casas-evertens TM, Ramdoss Y et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635. doi:10.1093/bioinformatics/btm308. PubMed: 17586929.

60. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959. PubMed: 10835412.

61. Miy a A, Albert P, Shinya T, Desaki Y, Ichimura K et al. (2007) CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. Proc Natl Acad Sci U S A 104: 19613-19618. doi:10.1073/pnas.0705147104. PubMed: 18042724.

62. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x. PubMed: 15969739.

63. Jauhar PP (2006) Modern Biotechnology as an Integral Supplement to Conventional Plant Breeding: The Prospects and Challenges Mention Immunity Related Genes in Physalis peruviana

64. Liu J, Liu X, Dai L, Wang G (2007) Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. J Genet Genomics 34: 765-776. doi:10.1016/S1673-8527(07)80087-3. PubMed: 17110940.

65. Bakker E, Born T, Prins P, van der Vossen E, Uenk G et al. (2011) A genome-wide genetic map of NB-LRR disease resistance loci in potato. TAG Theoretical And Appl Genet Theoretische Angewandte Genetik 123: 493-508. doi:10.1007/s00122-011-1602-2.

66. Radwan O, Gandi S, Heesacker A, Whitaker B, Taylor C et al. (2008) Genetic diversity and genomic distribution of homologs encoding NBS-LRR disease resistance proteins in sunflower. Mol Genet Genomics 280: 111-125. doi:10.1007/s00438-008-0346-1. PubMed: 18553108.

67. Mun JH, Yu HJ, Park S, Park BS (2009) Genome-wide identification of NBS-encoding resistance genes in Brassica rapa. Mol Genet Genomics 282: 617-631. doi:10.1007/s00438-009-0492-0. PubMed: 19838736.

68. Gu L, Guo R (2007) Genomewide detection and analysis of alternative splicing for nucleotide binding site-leucine-rich repeat sequences in rice. J Genet Genomics = Yi Chuan Xue Bao 34: 247-257. doi:10.1016/S1673-8527(07)60026-2. PubMed: 17077156.

69. Li Y, Yang S, Yang H, Hua J (2007) The TIR-NB-LRR gene SNC1 is regulated at the transcript level by multiple factors. Mol Plant Microbe Interact 20: 1449-1456. doi:10.1094/MPMI-20-11-1449. PubMed: 17977156.

70. Simons G, Groenenendijk W, Wijbrants J, Reijans M, Groenen J et al. (1998) Dissection of the Fusarium I2 gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell 10: 1055-1068. doi:10.2307/3870690. PubMed: 9634592.

71. Couch BC, Spangler R, Rams C, May G (2006) Pervasive purifying selection characterizes the evolution of I2 homologs. Mol Plant Microbe Interact 19: 288-303. doi:10.1094/MPMI-19-0288. PubMed: 16570659.

72. Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J et al. (2008) Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae. Genetics 180: 391-408. doi:10.1534/genetics.108.087981. PubMed: 18723883.

73. Saraste M, Sibblard PR, Wittringhofer A (1990) The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci 15: 430-434. doi:10.1016/0968-0004(90)90218-F. PubMed: 2126155.

74. Tameling WI, Elzinga SD, Darmin PS, Vossen JH, Taklen FL et al. (2002) The tomato R gene products I-2 and M1-1 are functional ATP binding proteins with ATPase activity. Plant Cell 14: 2929-2939. doi:10.1105/tpc.005793. PubMed: 12417711.

75. Fenyk S, Campillo Ade S, Pohl E, Hussey PJ, Cann MJ (2012) A nucleotide phosphatase activity in the nucleotide binding domain of an orphan resistance protein from rice. J Biol Chem 287: 4023-4032. doi:10.1074/jbc.M111.314450. PubMed: 22157756.

76. Jones DA, Jones JDG, Andrews JH. ICT, Callow JA (1997) The Role of Leucine-Rich Repeat Proteins in Plant Defences. Advances in Botanical Research: Academic Press. pp. 89-167

77. DeYoung BJ, Innes RW (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. Nat Immunol 7: 1243-1249. doi:10.1038/ni1410. PubMed: 17110940.

78. Ellis LG, Lawrence GJ, Luck JE, Dodds PN (1999) Identification of Regions in Alleles of the Flax Rust Resistance Gene L That Determine Differences in Gene-for-Genec Specificity. Plant Cell Available: 11: 495-506

79. Jiang H, Wang C, Ding L, Tian D, Yang S (2007) Pattern of LRR nucleotide variation in plant resistance genes. Plant Cell 17: 253-261. doi:10.1105/tpc.005793. PubMed: 17977156.

80. Yang S, Gu T, Pan C, Feng Z, Ding J et al. (2008) Genetic variation of NBS-LRR class resistance genes in rice lines. Theor Appl Genet 116: 165-177. doi:10.1007/s00122-007-0656-4. PubMed: 17932646.