

RESEARCH ARTICLE

Genome-Wide Comparative Analysis Reveals Similar Types of NBS Genes in Hybrid *Citrus sinensis* Genome and Original *Citrus clementina* Genome and Provides New Insights into Non-TIR NBS Genes

Yunsheng Wang^{1,2}, Lijuan Zhou³, Dazhi Li⁴, Liangying Dai^{1*}, Amy Lawton-Rauh⁵, Pradip K. Srimani², Yongping Duan³, Feng Luo^{2*}

1 College of Plant Protection, Hunan Agricultural University, Changsha, China, **2** School of Computing, Clemson University, Clemson, South Carolina, United States of America, **3** Agricultural Research Service, U. S. Horticultural Research Laboratory, Fort Pierce, Florida, United States of America, **4** National Center for Citrus Improvement, Hunan Agricultural University, Changsha, China, **5** Department of Genetics and Biochemistry, Clemson University, Clemson, South Carolina, United States of America

* daily@hunau.net (LD); luofeng@clemson.edu (FL)



OPEN ACCESS

Citation: Wang Y, Zhou L, Li D, Dai L, Lawton-Rauh A, Srimani PK, et al. (2015) Genome-Wide Comparative Analysis Reveals Similar Types of NBS Genes in Hybrid *Citrus sinensis* Genome and Original *Citrus clementina* Genome and Provides New Insights into Non-TIR NBS Genes. PLoS ONE 10(3): e0121893. doi:10.1371/journal.pone.0121893

Academic Editor: Ji-Hong Liu, Key Laboratory of Horticultural Plant Biology (MOE), CHINA

Received: October 28, 2014

Accepted: February 4, 2015

Published: March 26, 2015

Copyright: 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](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported in part by the Citrus Research and Development Foundation, Inc. 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.

Abstract

In this study, we identified and compared nucleotide-binding site (NBS) domain-containing genes from three *Citrus* genomes (*C. clementina*, *C. sinensis* from USA and *C. sinensis* from China). Phylogenetic analysis of all *Citrus* NBS genes across these three genomes revealed that there are three approximately evenly numbered groups: one group contains the Toll-Interleukin receptor (TIR) domain and two different Non-TIR groups in which most of proteins contain the Coiled Coil (CC) domain. Motif analysis confirmed that the two groups of CC-containing NBS genes are from different evolutionary origins. We partitioned NBS genes into clades using NBS domain sequence distances and found most clades include NBS genes from all three *Citrus* genomes. This suggests that three *Citrus* genomes have similar numbers and types of NBS genes. We also mapped the re-sequenced reads of three pomelo and three mandarin genomes onto the *C. sinensis* genome. We found that most NBS genes of the hybrid *C. sinensis* genome have corresponding homologous genes in both pomelo and mandarin genomes. The homologous NBS genes in pomelo and mandarin suggest that the parental species of *C. sinensis* may contain similar types of NBS genes. This explains why the hybrid *C. sinensis* and original *C. clementina* have similar types of NBS genes in this study. Furthermore, we found that sequence variation amongst *Citrus* NBS genes were shaped by multiple independent and shared accelerated mutation accumulation events among different groups of NBS genes and in different *Citrus* genomes. Our comparative analyses yield valuable insight into the structure, organization and evolution of NBS genes in *Citrus* genomes. Furthermore, our comprehensive analysis showed that the non-TIR NBS genes can be divided into two

groups that come from different evolutionary origins. This provides new insights into non-TIR genes, which have not received much attention.

Introduction

Disease resistance genes (*R* genes) are essential components of plant immune systems. Amongst five diverse classes of disease resistance genes [1], the largest class of *R* genes includes genes that encode proteins with a nucleotide-binding site (NBS) domain and a Leucine-Rich Repeat (LRR) domain. The NBS domain is highly sequence constrained and is typically used to identify and characterize plant *R* genes. The LRR domains in *R* genes that mediate direct or indirect interactions [2, 3] with pathogen molecules are usually rapidly evolving to adapt to the change of pathogen ligands [4]. NBS domain-containing *R* genes were classified into two major types based on their domain structures in N terminus: proteins with a Toll-Interleukin receptor (TIR) domain and proteins without TIR domains, which usually contain a Coiled Coil (CC) domain. While the TIR domain is usually well defined, the CC domain has higher sequence variation and is less well characterized.

Genome-wide analyses of NBS genes in many genomes have shown that NBS *R* genes are diverse in number, structure and organization [5–10]. For example, grass genomes studied to date do not have TIR NBS genes; on the other hand, dicot genomes usually contain more TIR NBS genes than non-TIR NBS genes [11]. Comparison of NBS *R* genes in multiple genomes indicated that *R* genes are shaped by dynamic birth-and-death processes [12]. Comparative analyses of NBS genes in two *Arabidopsis* genomes (*A. thaliana* and *A. lyrata*) indicate that mating system shift from outcrossing to inbreeding has had a limited impact on the numbers of NBS genes, at least in the time since *A. lyrata* and *A. thaliana* diverged from their common ancestor approximately 5 million years ago [13]. Comparative analyses of NBS genes of diploid *Phaseolus* and tetraploid *Glycine* species concluded that whole genome duplication did not result in NBS *R* gene number increasing. Recently, comparison of *R* genes in diverse grass genomes by Yang *et al.* showed that rapid evolution in *R* genes in *Zea*, *Sorghum*, and *Brachypodium* is associated with rice blast disease resistance [14].

Citrus species are amongst the most important fruit trees and have been cultivated for more than 4000 years [15, 16]. Phylogenetic analyses using molecular markers showed that cultivated *Citrus* species (sweet orange, grapefruit, and lemon) are derived from three original cultivated *Citrus* species: *C. medica* (citron), *C. reticulata* (mandarin) and *C. maxima* (pomelo) [17, 18]. For example, the *C. sinensis* (sweet orange) is suggested to be the backcross hybrid of *C. maxima* (pomelo) and *C. reticulata* (mandarin) [17–19]. Phylogenetic analyses of partial NBS genes of *Poncirus trifoliata* (trifoliolate orange), *C. reticulata* (tangerine) and their F1 progeny showed that NBS genes of *Poncirus trifoliata* (trifoliolate orange), *C. reticulata* formed genus-specific clades. Additionally, NBS genes of their F1 progeny had sister relationships to only one of the parents [20]. This suggests that NBS genes in crossing hybrid *Citrus* species are also different from those in original *Citrus* species.

Recently available draft whole genome sequences of three *Citrus* species: *C. clementina* [21], *C. sinensis* from USA and *C. sinensis* from China [19], have made it possible to scan and identify all NBS genes in those genomes. In this study, we performed a genome-wide comparative analysis of NBS genes in three *Citrus* genomes (*C. clementina*, *C. sinensis* from China and *C. sinensis* from USA) to address the following questions: (1) What are the features of NBS genes in *Citrus*, such as numbers, physical locations, within-gene domain structures, and

evolutionary dynamics of NBS genes?; (2) Do NBS genes differ between hybrid and non-hybrid *Citrus* species?; and (3) Do *Citrus* NBS genes differ from NBS genes in other plant genomes?

Materials and Methods

Sequences Used

We downloaded the draft genome sequences and the original gene annotations of *C. clementina* (clementine) [21] and *C. sinensis* USA (sweet orange from USA) from the *Citrus* Genome Database (<http://www.citrusgenomedb.org/>) and those of *C. sinensis* China (Chinese sweet orange [19]) from the *C. sinensis* annotation project (<http://citrus.hzau.edu.cn/orange/>). The sizes of assembled genomes of *C. clementina*, *C. sinensis* China and *C. sinensis* USA are 301, 328 and 319 million base pairs, respectively. The *C. clementina* genomes assembled into nine major scaffolds and 95.8% of the sequences were assigned to those nine scaffolds. About 72.9% of the *C. sinensis* China genome assigned to nine chromosomes. The genome of *C. sinensis* USA only assembled to 12,574 scaffolds and the N50 of scaffolds is 250 kb. The original gene annotations of *C. clementina*, *C. sinensis* China and *C. sinensis* USA have 24,533, 29,385 and 25,397 genes respectively. We also downloaded the re-sequence data of three mandarin and three pomelo genomes from the *C. sinensis* annotation project (<http://citrus.hzau.edu.cn/orange/download/data.php>). The re-sequence data is single end Illumina reads with length of 101 bp.

Identification of NBS Genes in *Citrus* Genomes

We first screened the original predicted *Citrus* open reading frames (ORFs) using *hmmsearch* [22] with the hidden Markov models (HMM) of Pfam [23] for NBS domain presence (NB-ARC, PF00931) using an e-value cut-off of 0.1 for the *hmmsearch*. Then, the proteins selected by the HMM were searched against the Swiss protein database [24] using BLASTP [25]. Only the proteins that have a significant match (e-value < 1E-5 in BLASTP search) with the NBS proteins or resistant proteins in the Swiss protein database were identified as potential NBS-containing proteins. To recover possible NBS genes that may be missed in the original gene annotations, we mapped the identified NBS genes to the draft genome using TBLASTN. The matched sequences with e-value < 1E-5 were then predicted using Genewise [26]. The new potential NBS genes predicted by Genewise were also confirmed by BLASTP searching through the Swiss protein database. Finally, we scanned all potential NBS genes for NBS domain using *hmmsearch* [22] and only keep the NBS genes with e value lower than more stringent cutoff of 1E-5.

Identification of Orthologous NBS genes in *C. clementina* and *C. sinensis*

Orthologous NBS genes of *C. clementina* and *C. sinensis* were identified using the reciprocal best blast method [27]. We used NBS genes of *C. clementina* as query sequences to search against the NBS genes of *C. sinensis* and *vice versa*. Protein pairs with reciprocal best hits of e-value < 1E-20 were defined as orthologs.

To calculate the rates of nonsynonymous, synonymous and their rate ratio (dN, dS and dN/dS) of orthologous pairs, we first aligned orthologous protein sequence pairs using *mafft*, and then converted the protein alignments to codon-based alignments using PAL2NAL [28]. We calculated the dN, dS and dN/dS rate ratios using the *codeml* program in PAML version 4.7 [29].

Phylogenetic Analysis

We constructed the best fit phylogenetic tree of NBS genes of the three *Citrus* genomes using only the conserved NBS domain. Only 1,099 NBS domain sequences that have both the P-loop and the MHDV motifs and were longer than 70% (200 amino acids) of the full-length NBS domain were included. Next, NBS domain sequences were aligned in **mafft** [30] using an auto alignment model and a best fit maximum likelihood phylogenetic tree of NBS genes was constructed using **FastTree** [31] with default parameters (JTT+CAT). The resultant best fit phylogenetic tree was divided into 3 main groups based upon clade support. For each group, we constructed a new ML tree with *Streptomyces coelicolor* protein P25941 as the outgroup using **FastTree** [31]. The average identity of each group was calculated using **alistat** implemented in **SQUID** (<http://selab.janelia.org/software.html>).

To evaluate the confidence level of the phylogenetic tree, we resampled the alignments 1000 times using **SEQBOOT** [32]. Then, we constructed the phylogenetic tree from each resampled alignments using **FastTree** [31]. The bootstrap values were calculated by comparing resampled phylogenetic trees to the original tree using **CompareToBootstrap** program in the **FastTree** [31].

We further partitioned the NBS gene tree into clades using the depth-first phylogeny partition method in **PhyloPart** [33] with distance threshold 0.025. This clustered the NBS genes into 114 clades, which contain more than 2 genes, and 8 orphan genes.

Domain and Motif Annotation

The Toll-Interleukin receptor (TIR) domains in *Citrus* NBS-containing proteins were identified using **hmmsearch** [22] with the HMM model of Pfam domain PF01582 and a 0.1 e-value cut-off. The Leucine-Rich Repeat (LRR) domains in *Citrus* NBS containing proteins were identified using the HMM models of Pfam LRR domains with e-value cut-off of 0.1. As long as there is a significant hit to one of LRR domain models (e-value < 0.1), it was define as an LRR-containing protein. We used **MARCOIL** [34] with a threshold probability of 90 and **COILS** [35] with a threshold of 0.9 to search for Coiled Coil (CC) domains in the N-terminal region of *Citrus* NBS-containing protein. We considered a protein as a CC-containing protein if either **MARCOIL** or **COILS** reported a CC domain in it.

We identified 20 motifs amongst the NBS genes in each of the three main phylogenetic groups separately using **MEME SUITE** [36]. The motif width was set to between 6 and 50 for **MEME**. Then, we searched the motif structure of all genes in each group using **MAST** with default parameters (-ev 10—mt 0.0001).

Pseudogene Identification

We identified possible pseudogenes using **PseudoPipe** [37] with default parameters (-e 0.1). The **PseudoPipe** algorithm identifies pseudo genes by integrating sequence similarity, intron-exon structure, plus presence of stop codons and frame-shifts. We used all *Citrus* NBS genes to search *Citrus* genomes for potential NGS pseudo genes.

Transposon Identification

All long terminal repeat (LTR) retrotransposons in each *Citrus* genome were identified using LTR finder [38] with default parameters (-o 3 -t 1 —e 1 -m 2 -u -2). Then, we used a script program to match the location of the LTR transposons to the NBS-LRR genes in each *Citrus* genome.

NBS Gene Synteny Identification

Gene synteny was identified and defined using **MCSanX** [39] with default parameters (-A-u 5000). First, NBS genes between two genomes were aligned using **BLASTP** and matches with E-value < 1e-5 were sorted according to their chromosome positions. Synteny scores were then calculated for each block based upon gene position. Two genes were considered in the same block if there were fewer than 25 genes separating them. **MCSanX** reported blocks with at least 5 collinear gene pairs.

Gene Cluster Analysis

We grouped the NBS genes in each *Citrus* genome into the same cluster if the genome location between two genes was within 200 kb. We also identified the conserved gene clusters between the *C. clementina* and *C. sinensis*. If all genes in a cluster of *C. clementina* have orthologs in the corresponding cluster of *C. sinensis* and vice versa, then these two clusters were called completely conserved clusters. If only part of genes in the clusters have orthologs, we called these two clusters partially conserved.

Gene Conversion Detection

We first aligned the sequences of NBS genes in the same cluster using **mafft** [30]. Then, we used **GENECONV** [40] version 1.81a with default settings (N = 10,000) to detect gene conversions. **GENECONV** identifies gene conversions by finding identical fragments between pairs of sequences in a nucleotide alignment. A global *P* value ≤ 0.05 was used to assess the statistical significance of the observed conversions. **GENECONV** requires at least three sequences for analyses in order to account for shared ancestral states. Thus, we only detected conversions in clusters containing three or more genes.

Tests for Sites under Positive Selection

The amino acid sequences of NBS genes from the same clade were aligned with **mafft** [30]. Then we converted the protein alignments to codon-based alignment using **PAL2NAL** [28]. The positively selected sites were statistically identified using the Bayesian approach implemented in **codeml** within **PAML** [29]. We also further examined sites in the $\omega > 1$ class with >90% posterior probability.

Mapping of Re-sequencing Data of Mandarin and Pomelo

We mapped the raw reads of each re-sequenced sample to the draft genome of *C. sinensis* China using **BWA** [41] with default parameters (-k 19-d 100-A 1-B 4-O 6). Then, the mapped reads of NBS genes regions were extracted using **BEDtools** [42]. Two types of coverage of each NBS gene in *C. sinensis* were calculated. One coverage type divides the length of mapped sequences by the whole gene sequence length and the other type divides the length of mapped sequences by the length of exon sequence only.

Citrus DNA Extraction and PCR Amplification

Citrus leaf samples were collected from the six *Citrus* plants in USHRL's (USDA Horticultural Research Laboratory, Fort Pierce, Florida): *C. sinensis* (sweet, navel orange), *C. aurantium* (karum jamir, sour orange), *C. reticulata* (mandarin orange), *C. clementina* (clementina), *C. aurantiifolia* (sweet lime), *C. japonica* (Yuzu, kumquat), and *C. maxima* (pomelo). Total DNA was extracted from leaf midribs following the Plant Mini Kit standard protocol from Qiagen Inc. (Valencia, CA), followed by DNA quantity and quality evaluation with Nanodrop. We

chose the NBS gene, Cs1g09350, which was conserved in *C. clementina* and *C. sinensis* for validating the conservation of NBS gene among different *Citrus* genomes. Primers used in this study were designed using Oligo 7.23 (Molecular Biology Insights, Inc., Cascade, CO, USA). DNA Polymerase (Invitrogen, Carlsbad, CA, USA) was used to amplify the NBS-LRR genes from *Citrus* DNA. For PCR, 20 µL reactions using standard conditions provided by the manufacturer for DNA Polymerase. PCR was performed using an initial denaturation at 95°C for 3 minutes, 35 cycles of 94°C for 20 seconds, 50–52°C for 20 seconds (specified by different primer sets) and 68°C for 3 minutes, follow by final extension at 68°C for 10 minutes. The cloning and sequencing analysis of amplified PCR products were conducted as previously described [43]

Results

Identification and classification of *Citrus* NBS Genes

We searched the *C. clementina*, *C. sinensis* China and *C. sinensis* USA genomes for genes containing the NBS domain using **hmmsearch**. Then, the NBS-containing genes were confirmed through homology searches against the Swiss protein database (see “[Materials and Methods](#)”). We identified similar numbers of NBS domain-containing genes amongst these three genomes. We found 618, 650 and 508 NBS genes from *C. clementina*, *C. sinensis* China and *C. sinensis* USA genomes respectively ([Table 1](#)). Among them, 413, 499 and 484 NBS genes were predicted in the original gene annotations and 205, 151 and 24 NBS genes in *C. clementina*, *C. sinensis* China and *C. sinensis* USA respectively were newly predicted in this project ([S1 Table](#)).

NBS genes could be classified into different classes based on their domain structures [6]. We searched the Toll-Interleukin receptor (TIR) and Coiled Coil (CC) domains in the N-terminal region and the Leucine-Rich Repeat (LRR) domains in the C-terminal of NBS genes. We identified 117 NBS genes with CC-NBS-LRR (CNL) domains, 82 NBS genes with TIR-NBS-LRR (TNL) domains, 68 NBS genes with NBS-LRR (NL) and 351 others NBS genes without LRR domains (N, CN and TN) from *C. clementina* ([Table 1](#)). We also identified 113 CNL, 77 TNL, 62 NL and 398 others NBS genes from *C. sinensis* China and 60 CNL, 30 TNL, 85 NL and 333 others NBS genes from *C. sinensis* USA ([Table 1](#)). In comparison to other genomes, there are many more *Citrus* NBS genes without the LRR domain. There are only 43.2%, 38.8% and 34.4% NBS genes with existing LRR domains in *C. clementina*, *C. sinensis* China and *C. sinensis* USA, respectively, while 72% NBS genes in *Arabidopsis thaliana* [6] and 78% NBS genes in *Populus trichocarpa* [44] have LRR domain.

Table 1. Classification of *Citrus* NBS genes.

Class	<i>C. clementina</i>	<i>C. sinensis</i> China	<i>C. sinensis</i> USA
CNL	117	113	60
TNL	82	77	30
NL	68	62	85
Others	351	398	333
Total	618	650	508

The CNL class of NBS genes contains three domains: Coiled Coil (CC), nucleotide-binding sites (NBS) and Leucine-Rich Repeat (LRR). The TNL class of NBS genes contains three domains: Toll-Interleukin receptor (TIR), nucleotide-binding sites (NBS) and Leucine-Rich Repeat (LRR). The NL class of NBS genes contains two domains: nucleotide-binding sites (NBS) and Leucine-Rich Repeat (LRR). The other NBS genes contain no LRR domain.

doi:10.1371/journal.pone.0121893.t001

The structures of TNL and CNL NBS genes are significantly different. The TNL NBS genes tend to have more introns than that of CNL NBS genes as previously found in *Arabidopsis* [6] and *Populus* [44]. The average numbers of introns in TNL NBS genes are 4.39, 4.70 and 4.47, while the numbers of introns in CNL NBS genes are 0.89, 1.15 and 1.38 in *C. clementina*, *C. sinensis* China and *C. sinensis* USA respectively (S1 Fig.). The median numbers of introns of different types of NBS genes in the three *Citrus* genomes are similar: four in TNL NBS genes and one in CNL NBS genes.

The NBS genes were unevenly distributed in the *Citrus* scaffolds/chromosomes. There were 158, 217 and 110 NBS genes, totaling 78.3% of 618 NGS genes, distributed in scaffolds_3, _5 and _7 of *C. clementina*, respectively. There were 125, 121 and 75 NBS genes distributed in chromosomes 1, 3 and 5 of *C. sinensis* China. The majority (95 out of 107) of NBS genes with TIR domain in *C. clementina* was distributed in scaffold 3. There were 33 and 51 (out of 102) NBS genes with TIR domain in *C. sinensis* China distributed on chromosome 5 and chromosome unknown, respectively.

Phylogenetic and Clade Analysis of the *Citrus* NBS Genes

To avoid mutation saturation effects in nucleotide sequences over time which can lead to an underestimation of the number of mutation events due to higher substitution rates versus proteins, we used the protein sequences of NBS domains, which are the most conserved part of NBS genes, to construct a phylogenetic tree of NBS genes. We only selected NBS domain sequences longer than 200 amino acid residues and contain both P-loop and MHDV motifs. Finally, 442 *C. clementina*, 393 *C. sinensis* China, and 264 *C. sinensis* USA NBS domain sequences were used for phylogenetic analysis (S2 Table). A Maximum Likelihood (ML) phylogenetic tree of these 1,099 NBS genes was then constructed using **FastTree** [31]. As shown in Fig. 1, the un-rooted phylogenetic tree can be divided into three main groups. Most NBS genes containing the TIR domain are in one branch and the non-TIR NBS genes comprise the other two branches (Fig. 1), in which most of them contain a CC domain. Therefore, we denoted the three branches as TIR, CC1 and CC2.

There were 452, 382 and 265 *Citrus* NBS genes in CC1, CC2 and TIR groups, respectively. S2 Table lists the group classification of 1,099 *Citrus* NBS genes. For each sub-group, we constructed a new ML phylogenetic tree rooted by *Streptomyces coelicolor* protein P25941 (S2 Fig.). The NBS domains in CC groups, especially in the CC1 group, were relatively more diverged in sequence versus those in TIR group. The average Poisson corrected distances between sequence pairs within each group were 0.947, 0.827 and 0.663 for CC1, CC2 and TIR groups, respectively. We aligned the NBS domains sequences in each group using mafft with default parameters (—auto). The average percentages of identities for NBS domain sequences were 40%, 44% and 52% for CC1, CC2 and TIR groups, respectively.

The number of LRR domains varied amongst the three classes of *Citrus* NBS genes. The number of LRR domains in the TIR group is significantly higher than the numbers of LRR domains in the CC1 and CC2 groups (S3 Fig.). The majority of CC1 NBS genes have only one LRR domain. On other hand, the TIR NBS genes have 2.4 LRR domains on average. The most frequent type of LRR domain in the CC1 group is LRR_8 (PF13855), but LRR_1 (PF00560) is the most frequent domain in both the CC2 group and the TIR group (S3 Table).

NBS genes in the three *Citrus* genomes are polyphyletic amongst branches of phylogenetic tree as shown in the Fig. 1. We partitioned the NBS gene tree into clades based on NBS domain sequence distance using PhyloPart with distance threshold of 0.025. This resulted in 114 clades (with at least two genes) and eight orphan genes (S2 Table). The numbers of clades in each NBS gene group are similar. There are 39, 42, 32 clades in CC1, CC2 and TIR groups,

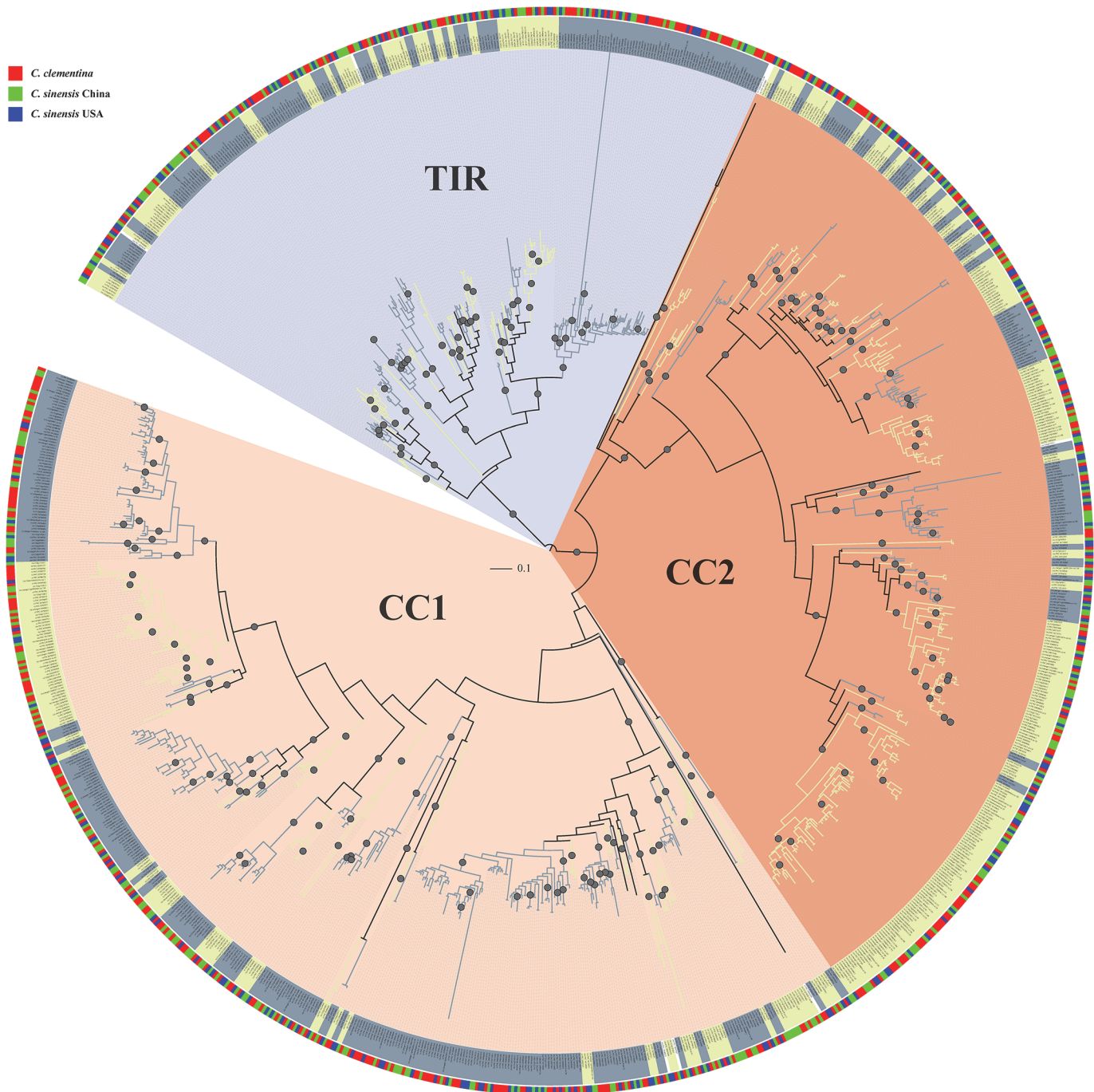


Fig 1. Maximum likelihood phylogenetic tree of Citrus NBS-LRR genes constructed from multiple sequences alignment of NBS domain. There are three gene groups in the phylogenetic tree: CC1 (pink fan-shape), CC2 (orange fan-shape) and TIR (light grey fan-shape). Clades were classified using PhyloPart as shown in alternating color and represented by alternative light yellow and grey colors in the middle circle. The outer circle shows species with *C. clementina* in red, *C. sinensis* China in green and *C. sinensis* USA in blue. Nodes with bootstrap support of 100 are indicated with grey circles on the tree.

doi:10.1371/journal.pone.0121893.g001

respectively. We calculated average sequence similarities between NBS domains in each clade using *alistat* in *SQUID* (<http://selab.janelia.org/software.html>). The minimum average sequence similarity of NBS domains in each clade was 70%. The largest clade included 90 NBS

Table 2. Types of Citrus NBS gene clades.

NBS genes from Citrus genomes	Number of Clade
<i>C. clementina</i> , <i>C. sinensis</i> China, <i>C. sinensis</i> USA	87
<i>C. clementina</i> , <i>C. sinensis</i> China	13
<i>C. clementina</i> , <i>C. sinensis</i> USA	7
<i>C. sinensis</i> China, <i>C. sinensis</i> USA	5
<i>C. clementina</i> ,	1
<i>C. sinensis</i> China,	0
<i>C. sinensis</i> USA	1

doi:10.1371/journal.pone.0121893.t002

genes. On average, there were 9.56 genes per clade. There were seven clades with more than 40 genes. Among these clades, four clades were in CC1; two clades were in CC2 and one clade was in the TIR group. Most of the clades contain NBS genes from the three *Citrus* genomes (Fig. 1, Table 2). There are 89 clades with at least three NBS genes and 87 of these clades contain members from each of the three *Citrus* genomes. Together, these results imply that three *Citrus* genomes may have similar types of NBS genes.

To understand the evolutionary dynamics maintaining functional constraint while maintaining so many gene family members, we tested for evidence of accelerated sequence evolution (positive selection) by searching for positively-selected sites within each clade. We detected positively selected sites in 53.5% (61) of clades. There were 18, 24 and 19 clades with sites under positive selection in CC1, CC2 and TIR groups, respectively. In total, we detected 541 positively selected sites. Consistent with previous reports [6], there are more positively selected sites in the C-terminal region of NBS genes (LRR domains) than those in the N-terminal and NBS domains (S4 Table). Approximately 56.4% (305 out of 541) of the positively selected sites were located in the C-terminal region (LRR domain). While most clades have positively selected sites in the LRR domain, there were clades with a greater number of positively selected sites within the NBS domain. For example, there were 46 positively selected sites in Clade_1260 that are located in NBS domain while there are only two positively selected sites in LRR domain (S4 Fig.).

Mapping Re-sequencing Reads Showed the Conservation of NBS Genes in Citrus Genomes

To verify that *C. sinensis* (sweet orange) is a backcross hybrid of *C. maxima* (pomelo) and *C. reticulata* (mandarin) [19], Xu *et al.* previously re-sequenced three pomelo cultivars and three mandarin cultivars and showed that SSR and SNP markers in the *C. sinensis* genome derive from the pomelo and mandarin genomes at an approximately ratio of 1:3 [19]. Since our comparison of *C. sinensis* and *C. clementina* genomes showed that most of their NBS genes are very similar in sequence, we investigated the relationships amongst the NBS genes of *C. sinensis* (sweet orange), *C. maxima* (pomelo), and *C. reticulata* (mandarin) to compare and contrast how mutations accumulated and determine if there are differences in sequence, and hence potentially functional, constraints amongst NBS gene types in *Citrus* genomes.

We mapped the re-sequenced reads of three *C. maxima* (pomelo) genomes and three *C. reticulata* (mandarin) genomes to hybridized genome of *C. sinensis* (sweet orange) China using BWA with mem method [41] and calculated the coverage of each NBS gene in *C. sinensis* China. We detected 524 of 650 *C. sinensis* China NBS genes with more than 50% coverage in all of the six re-sequenced genomes. Additionally, 249 of 650 *C. sinensis* China NBS genes with more than 85% coverage in all six resequenced genomes and 592 (91%) of *C. sinensis* China

NBS genes with coverage over 85% in at least one of the re-sequenced *Citrus* genomes. Using a cutoff of 15% coverage as missing calls over the entire length of the gene, only 25 (<4%) of *C. sinensis* China NBS genes seem to be deleted in at least one of the six re-sequenced genomes. No NBS genes were lost amongst all of the six re-sequenced genomes.

Re-sequenced reads tended to map to exons rather than to introns (S5 Table). For example, the entire coding region of NBS gene Cs9g13310 was 3,693 bp in length and covered by all reads from the re-sequenced samples. However, the 16 kb introns were not consistently present in all of the re-sequenced reads. When we used exon presence instead of whole gene coverage, 403 out of 650 (62%) *C. sinensis* China NBS genes have more than 85% coverage in all six re-sequenced *Citrus* genomes and 645 (99%) *C. sinensis* China NBS genes have more than 85% coverage in at least one of the resequenced *Citrus* genomes. Most *C. sinensis* China NBS genes have high exon coverage on both the *C. maxima* (pomelo) and *C. reticulata* (mandarin) genomes (S5 Fig.). There were only three *C. sinensis* China NBS genes, Cs1g02140.1, Cs6g02120.1 and Cs7g02220.1, with low (<40%) exon coverage in all three *C. maxima* genomes, but high coverage was achieved in at least two of the three *C. reticulata* genomes (S5 Fig.). Our mapping results showed that most NBS genes of *C. sinensis* (sweet orange) have a corresponding copy in the re-sequenced *C. maxima* (pomelo) and *C. reticulata* (mandarin) genomes.

Motif Patterns in *Citrus* NBS Genes

To further examine the gene structures of three *Citrus* NBS gene groups, we searched the motifs in the sequences of each group using MEME. Fig. 2 lists consensus sequences of the top 20 motifs identified by MEME from the CC1, CC2 and TIR *Citrus* NBS gene groups.

The motifs in the N-terminal of CC1 and CC2 groups showed little similarity to each other. The MEME found two motifs from CC1 group. One of them, EEQQQMRRRLNQVQGWLSRVEA, was present in 356 of 452 CC1 NBS genes. The other motif, GSQEIDKLCLGGYCSKNCKSSYKF GKKVA, was present in 189 of 452 CC1 NBS genes. Both motifs have high level of sequence similarities with the CC NBS genes of *Arabidopsis* (S6 Fig. A). The MEME found three motifs from CC2 group. Two of the motifs, KKLTNMLEMIKAVLDDAEKQ and AVKLWLGKLDAAAYDVEDVLDEFQTEALR were identified in 325 and 340 of 382 CC2 NBS genes, respectively. The motif AVKLWLGKLDAAAYDVEDVLDEFQTEALR has high levels of similarity with the motif identified from the CC NBS genes of *Oryza sativa japonica* (Japonica rice) [45] (S6 Fig. B). The different motifs in CC1 and CC2 NBS genes implied that they may be from different evolutionary origins.

For the TIR group, the MEME identified five motifs in the N-terminal. They can be found in 76% to 90% of 265 TIR *Citrus* NBS genes. The first four motifs YDVFLSFRGEDTRDNFTSHLY, AIEASAI SVIIFSEGYASSRWCLDELVKI, GQIVIPVFYRVDPSDVRKQTG, and ENPEKVQK WRDALKEA were very similar to the TIR1-4 motifs identified from TIR NBS genes in *Arabidopsis* [6] and *Populus trichocarpa* [44] (S6 Fig. C).

The MEME algorithm identified nine motifs from NBS domains of CC2 and TIR groups, which are similar to the motif structures of NBS domains in *Arabidopsis thaliana* [6] and *Populus trichocarpa* NBS genes [44]. MEME results also showed that the motif structure of NBS domain of CC1 groups is slightly different from those of CC2 and TIR groups. Eleven motifs were identified from NBS domains of CC1 group with two extra RNBS motifs (RNBS-F, RNBS-G) between GLPL and MHDV motifs. Five motifs: P-loop, Kinase-2, RNBS-B, GLPL and MHDV, showed high levels of sequence similarity amongst NBS genes of three groups, which were also similar to those motifs from *Arabidopsis* [6] and *Populus trichocarpa* NBS genes [44]. The MHDV motif in *Citrus* was often slightly modified to MHDL, as found in *Arabidopsis* and

A. (CC1)

EEQQMRRRLNQVQGWLSRVEA GSQEIDKLCGGYCSKNCKSSYKFGKKVA xEENVGIIGLY**GMGGVGKTT**LLKQINNKF FDVVIWVVSQDLDL
P-loop RNBS-A
LSEKFL**LILDDI**WER GSK**IVFTTR**SxEVCG NFKVECLSEEEAWELFKKVG HPDIPELAQTVAKECG**GLPL**ALITIGRAM KKTPEEWKDAIEVLRRSASKF
Kinase-2 RNBS-B RNBS-C GLPL RNBS-D
VYSRLKFSYDSLPSD SCFLYCSLFPEDYEIPKEELIDY EARNRGYTIIGxLKHACLEEE EGEDxV**MHDV**IRDMAWIAS LPEIPTCPHLLTLFL
RNBS-E RNBS-F RNBS-G MHDV
IPDNFFQFMPSLKVL DLSGSDIEELPEELKALTNLKLNLLEYC LKVIPPNVISNLSRLEELRMF EALLEELLGLKHLTVLELTRSSHAL
APNLKSLEVSSCxxMEEIISV PLPFPHLKEISVSGCPKLLKPLDSNAKERK

B. (CC2)

KKLTNMLEMIKAVLDDAAEQ AVKLWLGLKDAAYDVEDVLEDFQTEALR GNGDPAQAQDPSSSRTRTSKFRKIPTCTTFTPQSIQFDYALMSKIKE
VIPIV**GMGLGKTT**LAQLVYNDKRVQDHF AWVCVDDDFVIRLTKAILES GKKFL**LVLD**DVWNNENYNDWEP LKRPLKAGAPGSKI**IVTTR**NQEVASIMGT
P-loop RNBS-A Kinase-2 RNBS-B
AYQLKLLSEEDCWSLFAQHAF LEEIGKKIVxKCK**GLPL**AAKTGLGLLRGK LKQCFAYCSLFPKDYEFEEEEILLWMAQGFLDH
RNBS-C GLPL RNBS-D
MEDLGREYFQELASRSFFQOS FK**MHDL**VHDLAQWVAGEECFTLEYNSEVN EMPVGIGKLTCLQTLNMFVVG LDMLQPHKNLEQLCICGYGGT
RNBS-E MHDV
LEDGCMCTALPSVQQLPSLKHVTVRGMSRVKRLGSEFYG PIPFPCLETFRFEDLQEWEDWIPHGSSQGVGFPKRELHILRCSKLGQT
GCKKVVWRSATDHLGSONSVVCRDTSNQVFLAGPLKPRIPK ETYIWKSHNELLQDICSCLKRLTIDSCPKLQSLVAEEEEKQQQQLCELSCR
IKTLPEVCELYNLQTLDLEGCRRLKPLP HIRGNMEIWKSMIERGRGFHRFSSLRHLTISGCDDDMVSFP

C. (TIR)

YDVFLSFRGEDTRDNFTSHLY AIEASAVISVIFSEGYASSRWCLDELVKI GQIVIPFYRVDPDVRKQTG KENPEKVQKWRDALKEAANLSGF
NKNGLVGVESRIEIESLLGVGSKDVYAL GIW**GMGGIGKTTI**ARAIFDKI SREFEGSCFLANVREESEKGG INLIFRRLSRMKV**LIVFDDVT**
P-loop RNBS-A Kinase-2
QLESIGSLDWFGPGSR**IITTR**DKQVLR GVRKIYEMKALEYDEALELFSRHAFKQNH LSSKVVKYAA**QVPL**ALKVLSGFL ISYDGLDDKEKNIFLDIACFF
RNBS-B RNBS-C GLPL RNBS-D
GEDRDLVMKFLDACGFYPEIGISVLVDKS NKIT**MHDL**LQEMGREIVRQES WHHEDIYEVLTNTGTGTEKIEGICLD MSKVKEIHLNPNTFTKMPKLRFLKFY
RNBS-E MHDV
LFAELRYLHWGHYPLKSLPSN CKRLKSLPSSLCKLKSGLxLNLGCSNLE GNISELFLDGTAEIELPSSIE NLSKNNFERLPESIIQLSKLRYLLLSYERLQSLPE

Fig 2. Architecture of NBS gene motifs in Citrus. The consensus sequences of the top 20 motifs identified by MEME for each group of Citrus NBS genes are listed (A: CC1 group, B: CC2 group, C: TIR group). The motifs in blue are identified from N-terminal (CC or TIR domain). The motifs in red are identified from NBS domain. The motifs in green are identified from C-terminal (LRR domain). The motifs in black are identified from other region of NBS genes. The bold letters highlight the conserved motifs in Arabidopsis. The motifs underlined in green were repeated in C-terminal (LRR domain).

doi:10.1371/journal.pone.0121893.g002

Populus previously. Meanwhile, the motifs RNBS-A, RNBS-C, and RNBS-D were quite dissimilar to each other amongst three groups.

The LRR domains in the C-terminal of NBS genes usually have high sequence diversity as they play a role in recognizing pathogen virulence proteins. The MEME algorithm identified five, six, and four motifs from LRR domains of CC1, CC2 and TIR NBS genes, respectively. Most of the motifs contained LxxL repeats. The LRR motifs from different groups were highly variable in sequence, which implied that the three groups of NBS genes play different roles in the Citrus immune system. Some LRR motifs had repeated several times in the same gene. Examples include: motif APNLKSLEVSSCxxMEEIISV found in 367 of 452 CC1 group with an average 4.8 motifs per gene; motif IKTLPEVCELYNLQTLDLEGCRRLKPLP found in 332 of 382 CC2 group with an average 2.8 times per gene and motifs CKRLKSLPSSLCKLKSGLxLNLGCSNLE and GNISELFLDGTAEIELPSSIE found in 208 and 209 of 265 TIR group with 2.5 and 2.1 times per gene, respectively.

Table 3. Citrus NBS genes in clusters.

Citrus species	Total NBS genes	Number of NBS genes in cluster	Number of clusters	Percent of NBS genes in cluster (%)
<i>C. clementina</i>	618	525	108	84.9
<i>C. sinensis</i> China	650	500	126	76.9
<i>C. sinensis</i> USA	508	207	72	40.7

doi:10.1371/journal.pone.0121893.t003

Analysis of NBS Gene Clusters in Citrus Genomes

The majority of *Citrus* NBS genes were physically clustered in genome (Table 3). 525 of 618 NBS genes in *C. clementina* were found in 108 clusters and 500 of 650 NBS genes in *C. sinensis* China were found in 126 clusters (S2 Table). Although the assembly of *C. sinensis* USA is more fragmented, there still are 207 of 508 NBS genes present in 72 clusters. The largest number of gene clusters in *C. clementina*, *C. sinensis* China and *C. sinensis* USA contain 55, 37 and 13 NBS genes, respectively.

Most clusters contain NBS genes from the same group. In *C. clementina*, there were 38 clusters with NBS genes of CC1 group, 40 clusters with NBS genes of CC2 group and 23 clusters with NBS genes of TIR group. Only seven out of 108 clusters contain the NBS genes from two or three groups. In *C. sinensis* China, there were 41 clusters with NBS genes of CC1 group, 44 clusters with NBS genes of CC2 group and 26 clusters with NBS genes of TIR group. There were 15 clusters containing NBS genes from two or three groups in *C. sinensis* China. In *C. sinensis* USA, there were 27 clusters with NBS genes of CC1 group, 25 clusters with NBS genes of CC2 group and 15 clusters with NBS genes of TIR group. Only five clusters contain NBS genes from different groups in *C. sinensis* USA. The lower number of clusters in *C. sinensis* USA may due to its fragmented assembly.

The sequences of NBS genes within clusters are much more similar to each other than those between clusters (T-test p value < 2.2e-16). The mean identities between genes within and between clusters were 0.628 and 0.393 respectively (Fig. 3). Furthermore, NBS genes in the same

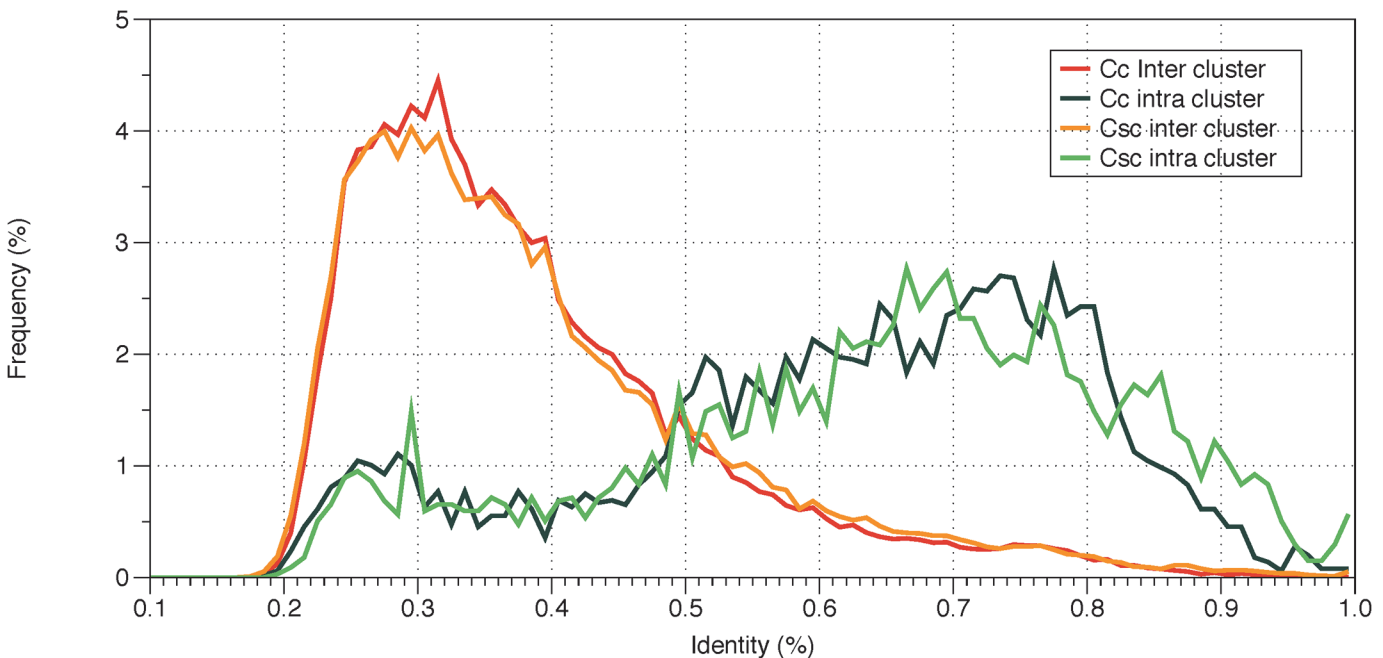


Fig 3. Percentage identity distribution of NBS-encoding genes in *C. clementina* (Cc) and *C. sinensis* China (Csc). Green lines indicate pairwise identity distribution of inter-cluster NBS-encoding genes. Orange lines indicate intro-cluster NBS-encoding genes.

doi:10.1371/journal.pone.0121893.g003

Table 4. Gene conversion events found in *C. clementina* and *C. sinensis* China.

Citrus species	Cluster type	Number of clusters (with conversion events)	Number of conversion events	Affected genes	Mean size of conversion tracts
<i>C. clementina</i>	TIR	27	35	32	111
	CC1	43	86	67	137
	CC2	46	133	83	146
<i>C. sinensis</i> China	TIR	36	69	42	137
	CC1	51	89	48	162
	CC2	57	92	75	116

doi:10.1371/journal.pone.0121893.t004

cluster tend to be in the same strand, which indicates that the NBS genes in the clusters are due to tandem duplication. We detected 204 pairs of tandem duplications in *C. clementina* and 217 pairs of tandem duplications in *C. sinensis* China using MCScanX [39]. The numbers of tandem duplications within CC1 and CC2 groups are much greater than that within the TIR group. Among 204 pairs in *C. clementina*, 90 and 85 pairs were present in the CC1 and CC2 groups and only 29 pairs from TIR group. Among 217 tandem gene pairs in *C. sinensis* China, 88 and 87 pairs were present in the CC1 and CC2 groups and only 42 pairs in the TIR group. The fewer tandem duplications of NBS genes in the TIR group may be the reason that there are fewer clusters of TIR groups in *Citrus* genomes.

We identified 254 and 246 gene conversion events from 116 NBS gene clusters in *C. clementina* and 144 NBS gene clusters in *C. sinensis* China, respectively (Table 4). The gene conversion events in *C. sinensis* USA is much less due to fragmented assembly. It is interesting that most of the conversion events (483 out of 520) were identified from the relatively small clusters with less than 10 NBS genes. Among these conversions, 119 events located in N-terminal, 178 in NBS domains and 223 in C-terminal, which indicating that there was no significantly bias in the location of conversion. Most gene conversions were between genes from the same group. We identified 101, 184 and 229 conversion events from NBS genes in TIR, CC1 and CC2 groups respectively and only four conversion events were identified between NBS genes of TIR and CC2 groups. While we could identify similar total numbers of conversion events in *C. clementina* and *C. sinensis* China, we found almost double such events in *C. sinensis* China TIR clusters versus in *C. clementina* TIR clusters. Meanwhile, we also found many more conversion events in *C. clementina* CC2 clusters than that in *C. sinensis* China CC2 clusters (Table 4).

Analysis of NBS Orthologs

We identified 719 *Citrus* NBS gene pairs of orthologs amongst *C. clementina*, *C. sinensis* China and *C. sinensis* USA. 270 orthologs were shared between *C. clementina* and *C. sinensis* China (S7 Fig.); 227 orthologous gene pairs were shared between *C. clementina* and *C. sinensis* USA and 222 orthologous gene pairs were shared between *C. sinensis* China and *C. sinensis* USA. The percentages of identities between orthologous genes range from 53.8% to 100% (mean of 93.6%, median of 96.8%). The percentages of identities between orthologous genes from TIR group (TNL and TN, 88.42 ± 10.04) were significantly (T-test: $P < 2.6e-6$) lower than those from CC groups (CNL and CN, 92.92 ± 7.03). The nonsynonymous divergence (dN) values of the 719 orthologs ranged from $1.46e-6$ to 0.31 (mean of 0.03, median of 0.08) and the synonymous divergence (dS) values ranged from $1.14e-5$ to 0.75 (mean of 0.05, median of 0.019) (Fig. 4A, B). The dN/dS rate ratios ranged from 0.001 to 50 with a median of 0.62. Similar to *Arabidopsis*, the dN and dS of orthologs from TIR group (TNL and TN) were higher than those from CC groups (CNL and CN). The dN values of orthologs from TIR group ranged

from 0.001 to 0.24 (mean of 0.038, median of 0.017) while those of orthologs from CC group ranged from 1.4e-6 to 0.23 (mean of 0.025, median of 0.007). The dS rate of orthologs from TIR group ranged from 3.27e-5 to 0.76 (mean of 0.078, median of 0.029) while those of orthologs from CC group ranged from 1e-5 to 0.67 (mean of 0.04, median of 0.016). However, the dN/dS rate ratios of orthologs from TIR group were relatively lower than that of orthologs from CC group (Fig. 4C).

The dN and dS rates as well as dN/dS rate ratios of ortholog residues in clusters are generally greater than those of orthologous singletons (Fig. 4D-F). The median dN/dS rate ratio of 148 orthologous singletons was 0.55 while that of 361 orthologs in clusters was 0.63. There were 124 orthologs with dN/dS rate ratios above 1. Variation in these orthologs may be shaped by positive selection pressure. Most of the positive selected orthologs (87.9%, 109 out of 124) belonged to CC groups (49 orthologs from CC1 group and 50 from CC2 group).

We also detected 38 NBS gene syntenic blocks between *C. clementina* and *C. sinensis* China using MCSanX [39]. Totally, there are 416 syntenic orthologs of NBS genes between *C. clementina* and *C. sinensis* China (S8 Fig.). On average, there are 11 genes per block. The biggest block contains 45 NBS genes and was located in scaffold_5 of *C. clementina* and chromosome

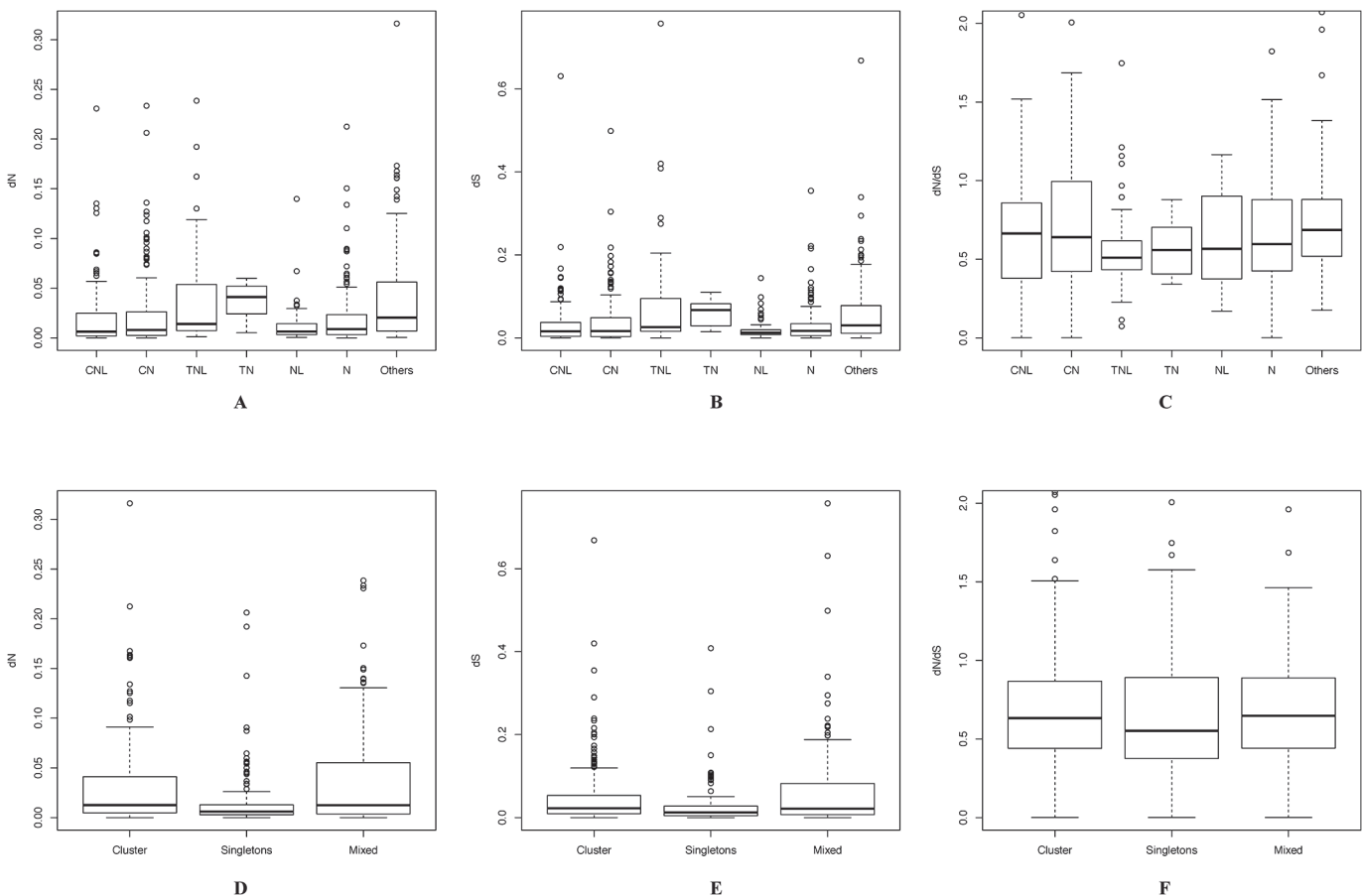


Fig 4. Divergence of orthologous NBS-encoding genes amongst *Citrus* species. A-C. Divergence of different classes. D-F. Divergence of singletons and cluster members. dN indicates the nonsynonymous mutation rate (rate of nonsynonymous mutations per nonsynonymous site). dS indicates the synonymous mutation rate. CNL, CN, TNL, TN, NL and N indicate orthologs between two proteins belonging to same cluster with CC-NBS-LRR domains (CNL), CC-NBS domains (CN), TIR-NBS-LRR domains (TNL), TIR-NBS domains (TN), NBS-LRR domains (NL), NBS domain (N), respectively. The others indicate orthologs between different clusters.

doi:10.1371/journal.pone.0121893.g004

3 of *C. sinensis* China. There were 80, 77 and 71 syntenic NBS genes in scaffold_5 of *C. clementina*/chromosome 5 of *C. sinensis* China, scaffold_3 of *C. clementina*/chromosome 5 of *C. sinensis* China and scaffold_7 of *C. clementina*/chromosome 1 of *C. sinensis* China. Furthermore, there were 68, 66 and 8 syntenic NBS genes in the unknown chromosome of *C. sinensis* China corresponding to NBS genes in scaffold_3, _5 and _7 of *C. clementina*, respectively.

Analysis of Conserved NBS Gene Clusters

We identified 118 pairs of conserved NBS gene clusters between *C. clementina* and *C. sinensis* China (S9 Fig.). There were 19 NBS gene clusters completely conserved in *C. clementina* and *C. sinensis* China. For example, all 14 NBS genes in cluster CL225 of *C. sinensis* China have orthologs in the cluster CL168 *C. clementina* which contains nine NBS genes and vice versa. Some clusters (in both *C. clementina* and *C. sinensis* China) have several corresponding conserved clusters. This may be due to either the genome arrangement or the incomplete assembly of genomes. Furthermore, there were 19 clusters of *C. clementina* with no conserved clusters in *C. sinensis* China and 18 clusters of *C. sinensis* China with no conserved clusters in *C. clementina*.

The conserved clusters provided additional insights into NBS gene evolution within and between *Citrus* genomes. For example, cluster CL142 (9 NBS genes) in *C. clementina* and cluster CL282 (7 NBS genes) in *C. sinensis* China were highly conserved. The phylogenetic tree of these 16 genes suggested division into two subgroups (depicted in blue and orange color in Fig. 5A). This division suggests two ancestral genes for this conserved cluster. Using the phylogenetic tree as a framework, we reconstructed the evolutionary history of these two clusters. There were several tandem duplication events in the evolutionary history of the conserved clusters, and extra tandem duplication was observed in *C. clementina* after it separated from *C. sinensis* (Fig. 5B). Two NBS genes were lost in *C. sinensis* and one NBS gene was lost in *C. clementina*. Furthermore, there was a recombination event within *C. clementina*.

Mutations and Transposons in *Citrus* NBS Genes

Plant NBS genes are continuously evolving. Sequence variation and structural constraints are shaped by gene birth-and-death processes [46]. Besides possible interallelic recombination and gene conversion, gene mutations and transposable elements appear to play important roles in NBS gene evolution.

We compared the DNA sequences of NBS genes from different *Citrus* genomes to identify mutations. Approximately half of *Citrus* NBS genes have mutations maintained amongst corresponding orthologs. When a mutation took place in an exon and resulted in stop-codon gaining or frame-shift, the target gene often became a pseudogene. For example, Cs1g18610.1_cc_32, Cs1g18610.1 and orange1.1g003367m are orthologs of *C. clementina*, *C. sinensis* China and *C. sinensis* USA. There were 2 stop-codon gaining mutations in Cs1g18610.1 (S10 Fig. A), which means that Cs1g18610.1 might become a pseudo (nonfunctional) gene after gaining these mutations.

LTR retrotransposons are widespread in eukaryotic genomes, especially plant genomes. We predicted 19014, 6296 and 1479 LTR retrotransposons with typical LTR characters in the draft genomes of *C. clementina*, *C. sinensis* China and *C. sinensis* USA, respectively. Then, we filtered out LTR retrotransposons with low similarity with known TE proteins using BLASTX with e-value greater than 1e-5. Finally, 4920, 3726 and 1240 LTR retrotransposons remained in *C. clementina*, *C. sinensis* China and *C. sinensis* USA, respectively. We identified 33, 32 and 4 NBS genes that were inserted with LTR retrotransposons in *C. clementina*, *C. sinensis* China and *C. sinensis* USA respectively (S6 Table). Most of these genes will likely become pseudogenes due to these insertions. For an instance, orange1.1g043039m,

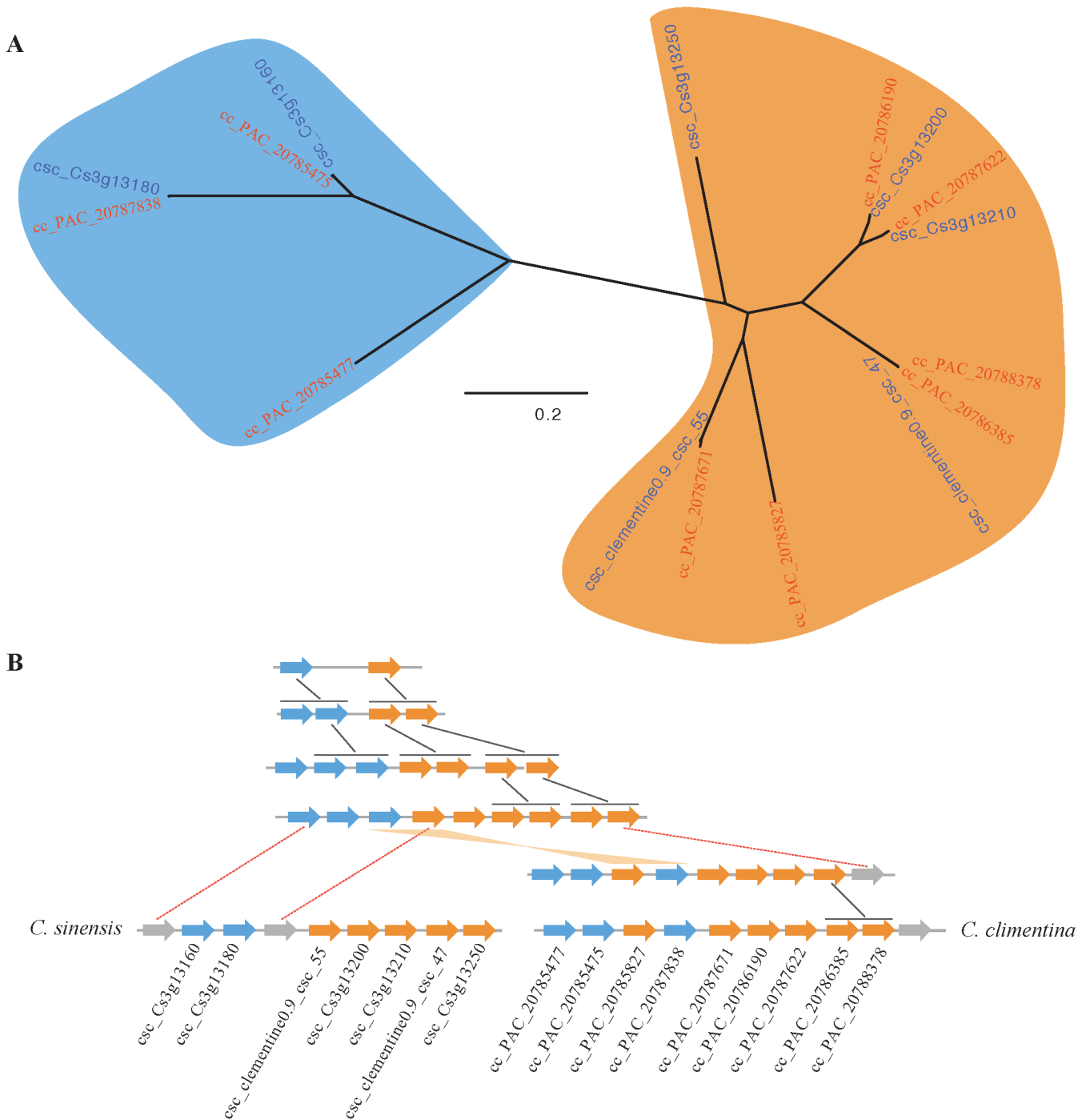


Fig 5. Duplication histories of one of the conserved NBS-encoding genes cluster in *C. clementina* and *C. sinensis*. **A.** Phylogenetic tree of the NBS-encoding genes; **B.** Orthologs in the conserved cluster of *C. clementina* and *C. sinensis*. Black connectors indicate tandem duplications and red dotted connectors indicate gene loss. A gene recombination event occurred in the ancestor of *C. clementina*, as indicated with an orange connector.

doi:10.1371/journal.pone.0121893.g005

orange1.1g043039m_cc_116 and orange1.1g043039m_csc_123 were orthologs of *C. sinensis* USA, *C. clementina* and *C. sinensis* China respectively (S10 Fig. B). The results of gene structure analysis showed that structure in orange1.1g043039m in *C. sinensis* USA seems relatively well

maintained. But there is an about 10 kb fragment of LTR-retrotransposons in the corresponding homologous gene in *C. clementina* and *C. sinensis* China (S10 Fig. B). We identified both orange1.1g043039m_cc_116 and orange1.1g043039m_csc_123 as pseudogenes using PseudoPipe [37].

Experimental Validation of One NBS Gene in *Citrus* Genomes

We validated the orthologs of a conserved NB gene, Cs1g09350.1, in a wide range of *Citrus* species. Cs1g09350.1 conserved in the 3 sequenced *Citrus* genomes. It is a CNL NBS gene and has 5 exons with about 4 kb in length. We designed the primers targeting about 3.5 kb fragment and could amplify a 3.5 kb fragment using PCR in different *Citrus* species, including *C. sinensis* (sweet orange), *C. clementina* (clementine), *C. japonica* (kumquat), *C. sinensis* Navelina (navel orange), *C. maxima* (pomelo), *C. aurantiifolia* (lime), and *C. aurantium* (sour orange). The orthologs from *C. sinensis* and *C. clementina* were identical as expected. There are a few mutations in the orthologs of *C. japonica*, *C. sinensis* Navelina, and *C. maxima*, but the orthologs of *C. aurantiifolia* and *C. aurantium* should be undergoing a pseudogenization process. There was a deletion plus several mutations in the NBS domain in *C. aurantiifolia*. There was an eight-base deletion in the LRR domain in *C. aurantiifolia* and *C. aurantium*. Most *C. aurantiifolia* mutations were shared with *C. aurantium* (S11 Fig.), and these mutations were not shared with other species. The results suggest that in *C. aurantiifolia* and *C. aurantium*, this gene is likely derived from the same common ancestor and was inherited as a pseudogene.

Discussion

Similar NBS Genes in Hybrid *Citrus sinensis* and Original *Citrus clementina* Genome

After carefully reannotating the *Citrus* genome sequences, we found similar numbers of NBS genes in *C. clementina* and *C. sinensis* China. There are slightly fewer NBS genes in *C. sinensis* USA, possibly due to the more fragmented assembly of this genome. In phylogenetic tree using NBS domains, the NBS genes from three different *Citrus* genomes are spread across clades. After partitioning NBS genes on the tree into clades, we found that 97.7% of the clades containing at least three genes had members from each of three *Citrus* genomes. This pattern suggests that these three *Citrus* genomes have similar types of NBS genes derived from common ancestors. Because *C. sinensis* is the hybrid of *C. reticulata* (mandarin) and *C. maxima* (pomelo), it should be heterozygous and some of NBS genes in *C. sinensis* are expected have different genetic distances from NBS genes in *C. clementina*. This would support previous observations of NBS genes in F1 progeny of *Poncirus trifoliata* (trifoliolate orange) and *Citrus reticulata* (tangerine). However, this is clearly not the case.

Furthermore, we mapped the re-sequenced reads of three *C. maxima* (pomelo) genomes and three *C. reticulata* (Mandarin orange) genomes onto the genome of *C. sinensis* China. In this case, 62% of *C. sinensis* China NBS genes have a copy present in all six resequenced genomes and 99% of *C. sinensis* China NBS genes have a copy in at least one of the re-sequenced genomes. The mapping results confirmed that a significant percentage of NBS genes of hybrid *C. sinensis* genomes have corresponding homologous genes in both the *C. maxima* and the *C. reticulata* genomes. Because the reference genome sequence of *C. maxima* is not yet available, the total number of NBS genes in *C. maxima* genomes is still unknown. However, we can at least conclude from the mapping of re-sequenced genomes that the *C. maxima* genome has homologous copies of NBS genes in *C. reticulata* and *C. sinensis* genomes. The homologous NBS genes

in *C. maxima* and *C. reticulata* may be the reason that NBS genes in their hybrid *C. sinensis* are similar to those in *C. clementina* in this study.

Three Groups of *Citrus* NBS Genes

We identified 442, 393 and 264 genes with full length NBS domains from *C. clementina*, *C. sinensis* China and *C. sinensis* USA reference genomes, respectively. There are also many genes with short NBS domains in three *Citrus* genomes. The *Citrus* NBS genes can be divided into three groups according to the phylogenetic tree of NBS domains: two of them without TIR domain (CC1 and CC2 groups) and the other group with TIR domain (TIR group). The number of non-TIR NBS genes is three times of the number of TIR NBS genes. In most of the TIR NBS genes, we can find the LRR domains defined in Pfam database [47]. We only can identify LRR domains in small part of non-TIR NBS genes using Pfam LRR domain definition. However, we can find the LxxLs repeats in most of the non-TIR NBS genes as shown in motifs from MEME. This implied that there may be other types of LRR domain in *Citrus* non-TIR NBS genes.

Our motif analyses also showed that motifs of TIR domains in *Citrus* TIR NBS genes are similar to those of TIR domains in *Arabidopsis* [6] and *Populus trichocarpa* [44] TIR NBS genes. Furthermore, the motifs of *Citrus* CC domains in CC1 NBS genes are similar to those motifs of CC domains in *Arabidopsis* CC NBS genes and the motifs of *Citrus* CC domains in *Citrus* CC2 NBS genes are similar to those motifs of CC domains in japonica rice CC NBS genes [45]. The different structure of motifs in NBS domain and CC domain between *Citrus* CC1 NBS genes and *Citrus* CC2 NBS genes suggested that they are likely from separate evolutionary origins.

To further confirm the three groups of NBS genes, we identified the NBS genes from *Arabidopsis*, *Populus*, *Oryza sativa* and grape. We used the same criteria to select the NBS domains from NBS genes of these genomes. Finally, we selected 152, 216, 209, 126 NBS domains from *Arabidopsis*, *Populus*, *Oryza sativa* and grape, respectively. Then, together with 442 NBS domains selected from *C. clementina*, we constructed a phylogenetic tree of those 1145 NBS domain sequences from five genomes. As shown in Fig. 6, the un-rooted phylogenetic tree was divided into three main branches. The *Populus*, grape and *Citrus* genomes have significant amount of NBS genes in all three branches. However, the NBS genes of *Oryza sativa* dominated in CC2 branch and most of *Arabidopsis* NBS genes located in CC1 and TIR branches. Our study showed that the NBS genes can be divided into three major groups as the non-TIR NBS genes are separated into two groups. The three groups of NBS genes underwent divergent evolution in different genomes. Further comparison of NBS genes of more genomes may help to understand the evolution of NBS genes and will help elucidate how plants maintain and adapt their defense system against pathogens.

Highly Clustering of *Citrus* NBS Genes

The *Citrus* NBS genes are highly clustered in the genome. 84.9% of NBS genes in *C. clementina* and 76.9% of NBS genes in *C. sinensis* China were found in clusters. Previous studies showed 76% of rice [48], 64% of *A. lyrata* and 71% *A. thaliana* [13], 83.2% of grapevine and 67.5% of poplar [9] NBS genes are found in clusters. The percentage of *Citrus* NBS genes in clusters is in the high level comparing to other genomes. The average number of genes per clusters is 4.86 in *C. clementina* and 3.97 in *C. sinensis* China. These numbers are similar to those in other genomes [9]. Furthermore, most *Citrus* NBS genes in the same cluster belong to the same phylogenetic group, suggesting that tandem duplication is the primary mechanism for the expansion of NBS genes in the *Citrus* genus.

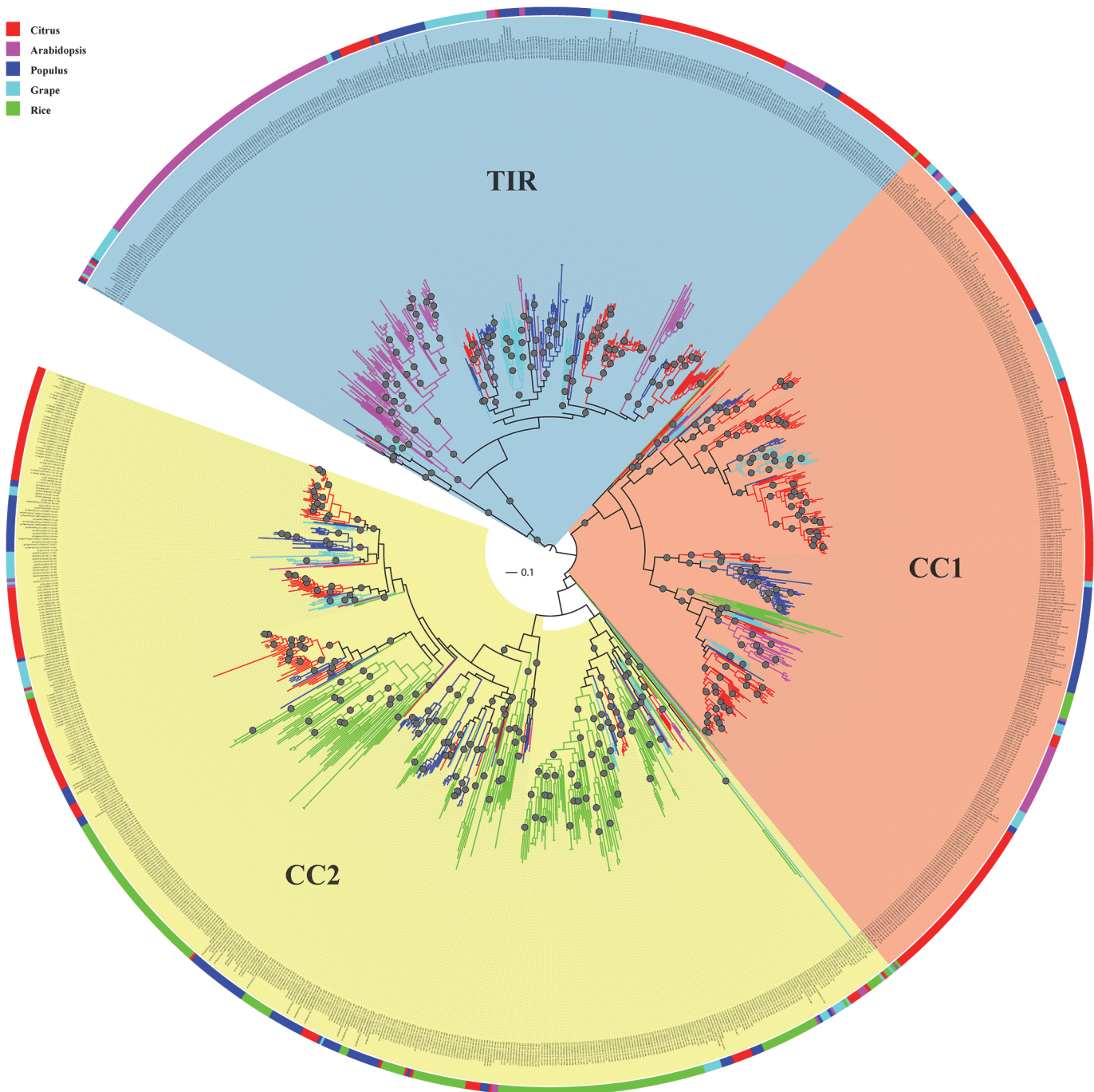


Fig 6. The maximum likelihood phylogenetic tree of NBS genes of *C. clementina*, *Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa* and *V. vinifera* (grape). Species are indicated by the color of the outer circle and branches. Nodes with bootstrap support of 100 are indicated with grey circles on the tree.

doi:10.1371/journal.pone.0121893.g006

Molecular Evolution of *Citrus* NBS Genes

Similar to NBS genes in other genomes, *Citrus* NBS genes are highly dynamic and are shaped by several evolutionary processes leading to several differences amongst NBS genes, including

domain presence and mutation constraints and genome organization. Our results revealed multiple molecular evolution events amongst *Citrus* NBS genes including gene duplications, gene conversions, mutation constraint changes, recombination and transposable element insertions. Likely these events support a birth-and-death process leading to the origins of new NBS genes as well as nonfunctionalization and total loss of other NBS genes [12]. We found more than 200 tandem duplications in both *C. clementina* and *C. sinensis* China genomes alone. We also found that NBS genes become pseudogenes following original frame-shift mutations leading to mutation accumulations plus disruption of gene constraint leading to loss of function through transposable element insertions.

Most molecular evolutionary events occurred within NBS gene groups, suggesting that gene birth-death processes have been occurring since divergence from a common ancestral gene copy. Interestingly, the molecular evolution processes occurred differently amongst NBS gene groups and differently within each *Citrus* genome. For example, the number of gene conversion events in *C. sinensis* China TIR clusters is almost double than that in *C. clementina* TIR clusters, while there are many more conversion events in *C. clementina* CC2 clusters than in *C. sinensis* China CC2 clusters (Table 4). Numbers of tandem duplications in CC1 and CC2 groups are much greater than that from the TIR group.

Conclusions

Our comparative analyses yield valuable insight into the understanding of the structure, evolution and organization of NBS genes in *Citrus* genomes. *Citrus* NBS genes are structurally highly clustered. The hybrid *C. sinensis* genomes have similar types of NBS genes as those progenitor *C. clementina* genomes have. Furthermore, our comprehensive analysis also showed that there are three groups of plant NBS genes while non-TIR NBS genes can be divided into two groups. The distributions of three groups of NBS genes among genomes are different implied that the three groups of NBS genes underwent divergent evolution in different genomes.

Supporting Information

S1 Fig. Average intron number of CNL and TNL. Cc: *C. clementina*, CsCN: *C. sinensis* China and CsUSA: *C. sinensis* USA.

(PDF)

S2 Fig. Phylogenetic trees of three groups of *Citrus* NBS gene with P25941 as outgroup. A. CC1 group; B. CC2 group; and C. TIR group. Nodes with bootstrap support of 100 are indicated with grey circles on the tree.

(PDF)

S3 Fig. Structure of *Citrus* NBS gene domains. The red rectangles indicate NBS domains, the green ellipses indicate CC domains, the left pointing pentagons indicate TIR domains and the left pointing triangles indicate LRR domains.

(PDF)

S4 Fig. Positively selection sites of Clade_1260. The grey box indicate the NBS domain.

(PDF)

S5 Fig. The exon coverage of NBS-encoding genes in the resequences of *Citrus* genome (3 mandarin and 3 pomelo).

(PDF)

S6 Fig. Multiple alignments of motifs from CC1 domain (A), CC2 domain (B) and TIR domain (C).

(PDF)

S7 Fig. Orthologs of *Citrus* NBS-encoding genes of *C. clementina* and *C. sinensis* and their genome locations. The outer blue circles indicate the nine largest scaffolds of *Cc* and orange circles indicate nine chromosomes plus a non-localized pseudo-chromosome. The NBS-encoding genes were arranged by their chromosomal positions. The NBS genes in the TIR group are indicated with orange links, CC1 in blue links and CC2 in light blue links. The NBS-encoding orthologs of *C. clementina* and *C. sinensis* are indicated by link lines.

(PDF)

S8 Fig. Dot plots of syntenic orthologs of the NBS genes between *C. clementina* (*Cc*) and *C. sinensis* China (*CsCN*) produced by MCScanX. The numbers in the grids indicate the number of syntenic ortholog pairs in the corresponding scaffolds. The scaffolds of *C. clementina* (*Cc*) are arranged along the x-axis and that of *C. sinensis* China (*CsCN*) are arranged on the y-axis.

(PDF)

S9 Fig. Diagram of conserved cluster between *C. clementina* and *C. sinensis*. The nodes in orange color indicate clusters of *C. clementina*, the blue nodes indicate clusters of *C. sinensis*, and the sizes of the nodes are shown relative to cluster size. The numbers on the edges indicate the number of orthologs identified in other species.

(PDF)

S10 Fig. Pseudogenization of NBS-encoding genes of *Citrus* due to mutation (A) and retrotransposon insertion (B). A: Sequences alignment of *Cs1g18610.1*, *Cs1g18610.1_cc_32* and *orange1.1g003367m*. The bases indicated the variation bases in each *Citrus* species. The star in red color represented stop-codon gaining mutation and may result in pseudogene of *Cs1g18610.1*. B: LTR retrotransposon insertion in NBS-encoding genes *orange1.1g043039m_cc_116* and *orange1.1g043039m_csc_123* from *C. clementina* and *C. sinensis* respectively, and the multiple sequences alignment of the orthologs. LTR, Long terminal repeats; TSR, Target site repeats.

(PDF)

S11 Fig. Validation and phylogenetic analysis of NBS-LRR gene, *Cs1g09350.1*, in different *Citrus* species. A) multiple sequence alignment, with blue bars along exons and thin lines along introns. The vertical lines in each alignment blocks indicate mutations that differ from the reference sequences of *C. sinensis*. B) neighbor-joining phylogenetic tree of the orthologs of *Cs1g09350.1* in different *Citrus* species and bootstrap support displayed on the branches.

(PDF)

S1 Table. The NBS genes in three *Citrus* genomes.

(DOCX)

S2 Table. Classification, cluster and clade of *Citrus* NBS genes.

(XLSX)

S3 Table. LRR domain distribution of *Citrus* NBS gene.

(XLSX)

S4 Table. Positively selection sites distribution.

(XLSX)

S5 Table. The exome coverage of NBS genes from the 3 mandarin and 3 pomelo resequenced genomes.

(XLSX)

S6 Table. *Citrus* NBS gene interrupted with LTR transposons.

(XLSX)

Author Contributions

Conceived and designed the experiments: FL YD PS LD. Performed the experiments: LZ. Analyzed the data: YW DL ALR FL. Wrote the paper: ALR PS LD YD FL.

References

1. Dangl JL, Jones JD. Plant pathogens and integrated defence responses to infection. *Nature*. 2001; 411(6839):826–33. PMID: [11459065](#)
2. Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *The EMBO Journal*. 2000; 19(15):4004–14. PMID: [10921881](#)
3. Ellis JG, Lawrence GJ, Luck JE, Dodds PN. Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. *The Plant Cell Online*. 1999; 11(3):495–506. PMID: [10072407](#)
4. Ellis J, Dodds P, Pryor T. The generation of plant disease resistance gene specificities. *Trends in plant science*. 2000; 5(9):373–9. PMID: [10973092](#)
5. Mun J-H, Yu H-J, Park S, Park B-S. Genome-wide identification of NBS-encoding resistance genes in *Brassica rapa*. *Molecular Genetics and Genomics*. 2009; 282(6):617–31. doi: [10.1007/s00438-009-0492-0](#) PMID: [19838736](#)
6. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *The Plant cell*. 2003; 15(4):809–34. PMID: [12671079](#)
7. Wan H, Yuan W, Bo K, Shen J, Pang X, Chen J. Genome-wide analysis of NBS-encoding disease resistance genes in *Cucumis sativus* and phylogenetic study of NBS-encoding genes in Cucurbitaceae crops. *BMC genomics*. 2013; 14(1):109.
8. Sun X, Wang G-L. Genome-wide identification, characterization and phylogenetic analysis of the rice LRR-kinases. *PloS one*. 2011; 6(3):e16079. doi: [10.1371/journal.pone.0016079](#) PMID: [21408199](#)
9. Yang S, Zhang X, Yue J-X, Tian D, Chen J-Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Molecular Genetics and Genomics*. 2008; 280(3):187–98. doi: [10.1007/s00438-008-0355-0](#) PMID: [18563445](#)
10. Ameline-Torregrosa C, Wang B-B, O'Bleness MS, Deshpande S, Zhu H, Roe B, et al. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant physiology*. 2008; 146(1):5–21. PMID: [17981990](#)
11. Pan Q, Wendel J, Fluhr R. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *Journal of Molecular Evolution*. 2000; 50(3):203–13. PMID: [10754062](#)
12. Michelmore RW, Meyers BC. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research*. 1998; 8(11):1113–30. PMID: [9847076](#)
13. Guo Y-L, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D. Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. *Plant physiology*. 2011; 157(2):757–69. doi: [10.1104/pp.111.181990](#) PMID: [21810963](#)
14. Yang S, Li J, Zhang X, Zhang Q, Huang J, Chen J-Q, et al. Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. *Proceedings of the National Academy of Sciences*. 2013. doi: [10.1073/pnas.1318211110](#)
15. Barrett H, Rhodes A. A numerical taxonomic study of affinity relationships in cultivated Citrus and its close relatives. *Systematic Botany*. 1976:105–36.
16. Scora RW. On the history and origin of Citrus. *Bulletin of the Torrey Botanical Club*. 1975:369–75.
17. Nicolosi E, Deng Z, Gentile A, La Malfa S, Continella G, Tribulato E. Citrus phylogeny and genetic origin of important species as investigated by molecular markers. *Theoretical and Applied Genetics*. 2000; 100(8):1155–66.

18. Li X, Xie R, Lu Z, Zhou Z. The origin of cultivated citrus as inferred from internal transcribed spacer and chloroplast DNA sequence and amplified fragment length polymorphism fingerprints. *Journal of the American Society for Horticultural Science*. 2010; 135(4):341–50.
19. Xu Q, Chen LL, Ruan X, Chen D, Zhu A, Chen C, et al. The draft genome of sweet orange (*Citrus sinensis*). *Nature genetics*. 2013; 45(1):59–66. doi: [10.1038/ng.2472](https://doi.org/10.1038/ng.2472) PMID: [23179022](https://pubmed.ncbi.nlm.nih.gov/23179022/)
20. Xu Q, Biswas MK, Lan H, Zeng W, Liu C, Xu J, et al. Phylogenetic and evolutionary analysis of NBS-encoding genes in Rutaceae fruit crops. *Molecular Genetics and Genomics*. 2011; 285(2):151–61. doi: [10.1007/s00438-010-0593-9](https://doi.org/10.1007/s00438-010-0593-9) PMID: [21153735](https://pubmed.ncbi.nlm.nih.gov/21153735/)
21. Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat Biotech*. 2014; 32(7):656–62. doi: [10.1038/nbt.2906](https://doi.org/10.1038/nbt.2906) PMID: [24908277](https://pubmed.ncbi.nlm.nih.gov/24908277/)
22. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic acids research*. 2011; 39(Web Server issue):W29–37. doi: [10.1093/nar/gkr367](https://doi.org/10.1093/nar/gkr367) PMID: [21593126](https://pubmed.ncbi.nlm.nih.gov/21593126/)
23. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic acids research*. 2012; 40(Database issue):D290–301. doi: [10.1093/nar/gkr1065](https://doi.org/10.1093/nar/gkr1065) PMID: [22127870](https://pubmed.ncbi.nlm.nih.gov/22127870/)
24. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic acids research*. 2003; 31(1):365–70. PMID: [12520024](https://pubmed.ncbi.nlm.nih.gov/12520024/)
25. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*. 1997; 25(17):3389–402. PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/)
26. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome research*. 2004; 14(5):988–95. PMID: [15123596](https://pubmed.ncbi.nlm.nih.gov/15123596/)
27. Moreno-Hagelsieb G, Latimer K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*. 2008; 24(3):319–24. PMID: [18042555](https://pubmed.ncbi.nlm.nih.gov/18042555/)
28. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research*. 2006; 34(Web Server issue):W609–12. PMID: [16845082](https://pubmed.ncbi.nlm.nih.gov/16845082/)
29. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*. 2007; 24(8):1586–91. PMID: [17483113](https://pubmed.ncbi.nlm.nih.gov/17483113/)
30. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular biology and evolution*. 2013; 30(4):772–80. doi: [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010) PMID: [23329690](https://pubmed.ncbi.nlm.nih.gov/23329690/)
31. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution*. 2009; 26(7):1641–50. doi: [10.1093/molbev/msp077](https://doi.org/10.1093/molbev/msp077) PMID: [19377059](https://pubmed.ncbi.nlm.nih.gov/19377059/)
32. Retief JD. *Phylogenetic analysis using PHYLIP*. *Bioinformatics methods and protocols*: Springer; 1999. p. 243–58.
33. Prosperi MC, Ciccozzi M, Fanti I, Saladini F, Pecorari M, Borghi V, et al. A novel methodology for large-scale phylogeny partition. *Nature communications*. 2011; 2:321. doi: [10.1038/ncomms1325](https://doi.org/10.1038/ncomms1325) PMID: [21610724](https://pubmed.ncbi.nlm.nih.gov/21610724/)
34. Delorenzi M, Speed T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics*. 2002; 18(4):617–25. PMID: [12016059](https://pubmed.ncbi.nlm.nih.gov/12016059/)
35. Lupas A. Prediction and analysis of coiled-coil structures. *Methods in enzymology*. 1996; 266:513–25. PMID: [8743703](https://pubmed.ncbi.nlm.nih.gov/8743703/)
36. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*. 2009; 37(Web Server issue):W202–8. doi: [10.1093/nar/gkp335](https://doi.org/10.1093/nar/gkp335) PMID: [19458158](https://pubmed.ncbi.nlm.nih.gov/19458158/)
37. Zhang Z, Carriero N, Zheng D, Karro J, Harrison PM, Gerstein M. PseudoPipe: an automated pseudo-gene identification pipeline. *Bioinformatics*. 2006; 22(12):1437–9. PMID: [16574694](https://pubmed.ncbi.nlm.nih.gov/16574694/)
38. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic acids research*. 2007; 35(Web Server issue):W265–8. PMID: [17485477](https://pubmed.ncbi.nlm.nih.gov/17485477/)
39. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic acids research*. 2012; 40(7):e49. doi: [10.1093/nar/gkr1293](https://doi.org/10.1093/nar/gkr1293) PMID: [22217600](https://pubmed.ncbi.nlm.nih.gov/22217600/)
40. Sawyer S. Statistical tests for detecting gene conversion. *Molecular biology and evolution*. 1989; 6(5):526–38. PMID: [2677599](https://pubmed.ncbi.nlm.nih.gov/2677599/)

41. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997v2. 2013.
42. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26(6):841–2. doi: [10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033) PMID: [20110278](https://pubmed.ncbi.nlm.nih.gov/20110278/)
43. Zhou L, Powell CA, Hoffman MT, Li W, Fan G, Liu B, et al. Diversity and plasticity of the intracellular plant pathogen and insect symbiont "Candidatus Liberibacter asiaticus" as revealed by hypervariable prophage genes with intragenic tandem repeats. *Applied and environmental microbiology*. 2011; 77(18):6663–73. doi: [10.1128/AEM.05111-11](https://doi.org/10.1128/AEM.05111-11) PMID: [21784907](https://pubmed.ncbi.nlm.nih.gov/21784907/)
44. Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, et al. Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant molecular biology*. 2008; 66(6):619–36. doi: [10.1007/s11103-008-9293-9](https://doi.org/10.1007/s11103-008-9293-9) PMID: [18247136](https://pubmed.ncbi.nlm.nih.gov/18247136/)
45. Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, et al. Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular genetics and genomics: MGG*. 2004; 271(4):402–15. PMID: [15014983](https://pubmed.ncbi.nlm.nih.gov/15014983/)
46. Tan S, Wu S. Genome Wide Analysis of Nucleotide-Binding Site Disease Resistance Genes in *Brachypodium distachyon*. *Comparative and functional genomics*. 2012; 2012:418208. doi: [10.1155/2012/418208](https://doi.org/10.1155/2012/418208) PMID: [22693425](https://pubmed.ncbi.nlm.nih.gov/22693425/)
47. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic Acids Research*. 2012; 40(D1):D290–D301.
48. Yang S, Feng Z, Zhang X, Jiang K, Jin X, Hang Y, et al. Genome-wide investigation on the genetic variations of rice disease resistance genes. *Plant molecular biology*. 2006; 62(1–2):181–93.