

1    Genome-wide Comparative Analysis Reveals Possible Common Ancestors of NBS Domain  
2    Containing Genes in Hybrid *Citrus sinensis* Genome and Original *Citrus clementina* Genome

3    Yunsheng Wang<sup>1,2</sup>, Email: yunshew@clemson.edu  
4    Lijuan Zhou<sup>3</sup>, Email: Lijuan.Zhou@ARS.USDA.GOV  
5    Dazhi Li<sup>4</sup>, Email: ldazhi@163.com  
6    Amy Lawton-Rauh<sup>5</sup>, Email: AMYLR@clemson.edu  
7    Pradip K. Srimani<sup>2</sup>, Email: psriman@clemson.edu  
8    Liangying Dai<sup>1,\*</sup>, Email: daily@hunau.net  
9    Yongping Duan<sup>3</sup>, Email: YongPing.Duan@ARS.USDA.GOV  
10   Feng Luo<sup>2,\*</sup>, Email: luofeng@clemson.edu

11

12   <sup>1</sup>College of Plant Protection, Hunan Agricultural University, Changsha, China

13   <sup>2</sup>School of Computing, Clemson University, Clemson, USA

14   <sup>3</sup>Agricultural Research Service, U.S. Horticultural Research Laboratory, Fort Pierce, USA

15   <sup>4</sup>National Center for Citrus Improvement, Hunan Agricultural University, Changsha, China

16   <sup>5</sup>Department of Genetics and Biochemistry, Clemson University, Clemson, USA

17   \*Corresponding authors:

18   Liangying Dai: College of Plant Protection, Hunan Agricultural University, Changsha, 410128, China

19   Feng Luo: 310 McAdams Hall, Clemson University, Clemson, SC 29634-0974, USA

20

21

22

23

1    **Abstract**

2    **Background**

3    Recently available whole genome sequences of three citrus species: one *Citrus clementina* and two *Citrus*  
4    *sinensis* genomes have made it possible to understand the features of candidate disease resistance genes  
5    with nucleotide-binding sites (NBS) domain in Citrus and how NBS genes differ between hybrid and  
6    original Citrus species.

7    **Result**

8    We identified and re-annotated NBS genes from three citrus genomes and found similar numbers of NBS  
9    genes in those citrus genomes. Phylogenetic analysis of all citrus NBS genes across three genomes  
10   showed that there are three approximately evenly numbered groups: one group contains the Toll-  
11   Interleukin receptor (TIR) domain and two different groups that contain the Coiled Coil (CC) domain.  
12   Motif analysis confirmed that the two groups of CC-containing NBS genes are from different  
13   evolutionary origins. We partitioned NBS genes into clades using NBS domain sequence distances and  
14   found most clades include NBS genes from all three citrus genomes. This suggests that NBS genes in  
15   three citrus genomes may come from shared ancestral origins. We also mapped the re-sequenced reads of  
16   three pomelo and three Mandarin orange genomes onto the *Citrus sinensis* genome. We found that most  
17   NBS genes of the hybrid *C. sinensis* genome have corresponding homologous genes in both pomelo and  
18   mandarin genome. The homologous NBS genes in pomelo and mandarin may explain why the NBS genes  
19   in their hybrid *Citrus sinensis* are similar to those in *Citrus clementina* in this study. Furthermore,  
20   sequence variation amongst citrus NBS genes were shaped by multiple independent and shared  
21   accelerated mutation accumulation events among different groups of NBS genes and in different citrus  
22   genomes.

23   **Conclusion**

1 Our comparative analyses yield valuable insight into the understanding of the structure, evolution and  
2 organization of NBS genes in *Citrus* genomes. There are significantly more NBS genes in *Citrus*  
3 genomes compared to other plant species. NBS genes in hybrid *C. sinensis* genomes are very similar to  
4 those in progenitor *C. clementina* genome and they may be derived from possible common ancestral gene  
5 copies. Furthermore, our comprehensive analysis showed that there are three groups of plant NBS genes  
6 while CC-containing NBS genes can be divided into two groups.

7

8 **Keywords:**

9 NBS genes, Comparative Genomics, Citrus Genomics

10  
11 **Background**

12 Disease resistance genes (R genes) are essential components of plant immune systems. Amongst five  
13 diverse classes of disease resistance genes [1], the largest class is those proteins that have nucleotide-  
14 binding site (NBS) domain and Leucine-Rich Repeat (LRR) domain. The NBS domain is evolutionarily  
15 conserved and is typically used to identify and characterize plant R genes. The LRR domains in R genes  
16 that mediate direct or indirect interactions [2, 3] with pathogen molecules are usually rapidly evolving to  
17 adapt to the change of pathogen ligands [4]. NBS domain-containing R genes were classified into two  
18 major types based on their domain structures in N terminus: proteins with a Toll-Interleukin receptor  
19 (TIR) domain and proteins with a Coiled Coil (CC) domain. While the TIR domain is usually well  
20 defined [5], the CC domain has higher variation and is less well characterized.

21 Genome wide analyses of NBS genes in many genomes have showed that NBS R genes are  
22 diverse in number, structure and organization [5-12]. For example, grass genomes studied to date do not  
23 have TIR NBS genes; on the other hand, dicot genomes usually contain more TIR NBS genes than CC  
24 NBS genes [13]. Meanwhile, comparative analyses of NBS genes have facilitated the understanding of  
25 the organization, classification and evolution of NBS R genes. Comparison of R genes in multiple

1 genomes indicated that R genes are shaped by dynamic birth-and-death processes [14]. Comparative  
2 analyses of NBS genes in two *Arabidopsis* genomes (*A. thaliana* and *A. lyrata*) indicate that mating  
3 system shift from outcrossing to inbreeding has had a limited impact on the numbers of NBS genes, at  
4 least in short term[15]. Comparative analyses of NBS genes of diploid *Phaseolus* and tetraploid *Glycine*  
5 species concluded that whole genome duplication did not result in NBS R gene number increases.  
6 Recently, comparison of R genes in diverse grass genomes by Yang *et al.* showed that rapid evolution in  
7 R genes in *Zea*, *Sorghum*, and *Brachypodium* is associated with rice blast disease resistance [16].

8 Citrus species are amongst the most important fruit trees and have been cultivated for more than  
9 4000 years [17, 18]. Phylogenetic analyses using molecular markers showed that other cultivated citrus  
10 species, such as, sweet orange, grapefruit, and lemon, are derived from three original cultivated *Citrus*  
11 species: *Citrus medica* (citron), *Citrus reticulata* (Mandarin orange) and *Citrus maxima* (pomelo) [19, 20].  
12 For example, the *Citrus sinensis* (sweet orange) is suggested to be the backcross hybrid of *C. maxima*  
13 (pomelo) and *Citrus reticulata* (Mandarin orange) [19-21]. Phylogenetic analyses of partial NBS genes  
14 of *Poncirus trifoliata* (trifoliate orange), *Citrus reticulata* (tangerine) and their F1 progeny showed that  
15 NBS genes of *Poncirus trifoliata* (trifoliate orange), *Citrus reticulata* formed genus-specific clades.  
16 Additionally, NBS genes of their F1 progeny had sister relationships to only one of the parents [22]. This  
17 suggests that NBS genes in crossing hybrid citrus species are also different from those in original citrus  
18 species.

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

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

3

4 **Results**

5 **Identification and classification of *Citrus* NBS Genes**

6 We searched the *Citrus clementina*, *Citrus sinensis* USA and *Citrus sinensis* China genomes for genes  
7 containing the NBS domain using hmmsearch. Then, the NBS-containing genes were confirmed through  
8 homology searches against the Swiss protein database (see “Materials and Methods”). We identified  
9 similar numbers of NBS domain-containing genes amongst these three genomes. We found 618, 650 and  
10 508 NBS genes from *Citrus clementina*, *Citrus sinensis* China and *Citrus sinensis* USA genomes  
11 respectively (Table 1). Among them, 413, 499 and 484 NBS genes were predicted in the original gene  
12 annotations and 205, 151 and 24 NBS genes in *Citrus clementina*, *Citrus sinensis* China and *Citrus*  
13 *sinensis* USA respectively were newly predicted in this project (Table S1).

14 NBS genes could be classified into different classes based on their domain structures [23]. We  
15 searched the Toll-Interleukin receptor (TIR) and Coiled Coil (CC) domains in the N-terminal region and  
16 the Leucine-Rich Repeat (LRR) domains in the C-terminal of NBS genes. We identified 117 NBS genes  
17 with CC-NBS-LRR (CNL) domains, 82 NBS genes with TIR-NBS-LRR (TNL) domains, 68 NBS genes  
18 with NBS-LRR (NL) and 351 others NBS genes without LRR domains (N, CN and TN) from *Citrus*  
19 *clementina* (Table 1). We also identified 113 CNL, 77 TNL, 68 NL and 398 others NBS genes from  
20 *Citrus sinensis* China and 60 CNL, 30 TNL, 85 NL and 333 others NBS genes from *Citrus sinensis* USA  
21 (Table 1). In comparison to other genomes, there are many more *Citrus* NBS genes without the LRR  
22 domain. There are only 43.2%, 38.8% and 34.4% NBS genes with existing LRR domains in *Citrus*  
23 *clementina*, *Citrus sinensis* China and *Citrus sinensis* USA, respectively, while 72% NBS genes in  
24 *Arabidopsis* [23] and 78% NBS genes in *Populus trichocarpa* [24] have LRR domain.

1       The structures of TNL and CNL NBS genes are significantly different. The TNL NBS genes tend  
2       to have more introns than that of CNL NBS genes as previously found in *Arabidopsis* [23] and *Populus*  
3       [24]. The average numbers of introns in TNL NBS genes are 4.39, 4.70 and 4.47, while the numbers of  
4       introns in CNL NBS genes are 0.89, 1.15 and 1.38 in *Citrus clementina*, *Citrus sinensis* China and *Citrus*  
5       *sinensis* USA respectively (Figure S1). The median numbers of introns of different types of NBS genes in  
6       the three *Citrus* genomes are similar: four in TNL NBS genes and one in CNL NBS genes.

7       The NBS genes were unevenly distributed in the *Citrus* scaffolds/chromosomes. There were 217,  
8       158 and 110 NBS genes, totaling 78.3% of 618 NGS genes, distributed in scaffolds 5, 3, and 7 of *C.*  
9       *clementina*, respectively. There were 125, 121 and 75 NBS genes distributed in chromosomes 1, 3 and 5  
10      of *C. sinensis* China. The majority (95 out of 107) of NBS genes with TIR domain in *C. clementina* was  
11      distributed in scaffold 3. There were 33 and 51 (out of 102) NBS genes with TIR domain in *C. sinensis*  
12      China distributed on chromosome 5 and chromosome unknown, respectively.

### 13      **Phylogenetic and Clade Analysis of the *Citrus* NBS Genes**

14      We used the protein sequences of NBS domains, which are the most conserved part of NBS genes, to  
15      construct a phylogenetic tree of NBS genes. We only selected NBS domain sequences longer than 200  
16      amino acid residues and contain both P-loop and MHDV motifs. Finally, 442 *Citrus clementina*, 393  
17      *Citrus sinensis* China, and 264 *Citrus sinensis* USA NBS domain sequences were used for phylogenetic  
18      analysis (Table S1). A Maximum Likelihood (ML) phylogenetic tree of these 1,099 NBS genes was then  
19      constructed using FastTree [25]. As shown in Figure 1, the un-rooted phylogenetic tree can be divided  
20      into three main groups. Most NBS genes containing the TIR domain are in one branch and most NBS  
21      genes containing a CC domain comprise the other two branches (Figure 1). Therefore, we denoted the  
22      three branches as TIR, CC1 and CC2.

23      There were 452, 382 and 265 *Citrus* NBS genes in CC1, CC2 and TIR groups, respectively.  
24      Table S2 lists the group classification of 1,099 *Citrus* NBS genes. For each sub-group, we constructed a  
25      new ML phylogenetic tree rooted by *Streptomyces coelicolor* protein P25941 (Figure S2). The NBS

1 domains in CC groups, especially in the CC1 group, were relatively more diverged in sequence versus  
2 those in TIR group. The average Poisson corrected distances between sequence pairs within each group  
3 were 0.947, 0.827 and 0.663 for CC1, CC2 and TIR groups, respectively. We aligned the NBS domains  
4 sequences in each group using mafft with default parameters (--auto). The average percentages of  
5 identities for NBS domain sequences were 40%, 44% and 52% for CC1, CC2 and TIR groups,  
6 respectively.

7 The number of LRR domains varied amongst the three classes of *Citrus* NBS genes. The number  
8 of LRR domains in the TIR group is significantly higher than the numbers of LRR domains in the CC1  
9 and CC2 groups (Figure S3). The majority of CC1 NBS genes have only one LRR domain but the  
10 majority of TIR NBS genes have 2 LRR domains. The most frequent type of LRR domain in the CC1  
11 group is LRR\_8 (PF13855), but LRR\_1 (PF00560) is the most frequent domain in both the CC2 group  
12 and the TIR group (Table S3).

13 Previous studies showed that NBS genes from different species fall into separate phylogenetic  
14 groups [26]. In contrast, NBS genes in the three *Citrus* genomes we investigated are mixed amongst  
15 branches of phylogenetic tree as shown in the Figure 1. We partitioned the NBS gene tree into clades  
16 based on NBS domain sequence distance using PhyloPart with distance threshold of 0.025. This resulted  
17 in 114 clades (with more than two genes in each clade) and eight orphan genes (Table S2). The numbers  
18 of clades in each NBS gene group are similar. There are 39, 42, 32 clades in CC1, CC2 and TIR groups,  
19 respectively. We calculated average sequence similarities between NBS domains in each clade using  
20 alistat in SQUID (<http://selab.janelia.org/software.html>). The minimum average sequence similarities of  
21 NBS domains in each clade was 70%. The largest clade included 90 NBS genes. On average, there were  
22 9.56 genes per clade. There were seven clades with more than 40 genes. Among these clades, four clades  
23 were in CC1; two clades were in CC2 and one clade was in the TIR group. Most of the clades contain  
24 NBS genes from the three *Citrus* genomes (Figure 1, Table 2). There are 89 clades with more than three  
25 NBS genes and 87 of these clades contain members from each of the three *Citrus* genomes. Together,

1 these results imply that all of the NBS genes of the three *Citrus* genomes may have possible common  
2 ancestors.

3 To understand the evolutionary dynamics maintaining functional constraint while maintaining so  
4 many gene family members, we tested for evidence of accelerated sequence evolution (positive selection)  
5 by searching for positively-selected sites within each clade. We detected positively selected sites in 53.5%  
6 (61) of clades. There were 18, 24 and 19 clades with sites under positive selection in CC1, CC2 and TIR  
7 groups, respectively. In total, we detected 541 positively selected sites. Consistent with previous reports  
8 [5], there are more positively selected sites in the C-terminal region of NBS genes (LRR domains) than  
9 those in the N-terminal and NBS domains (Table S4). Approximately 56.4% (305 out of 541) of the  
10 positively selected sites were located in the C-terminal region (LRR domain). While most clades have  
11 positively selected sites in the LRR domain, there were clades with a greater number of positively  
12 selected sites within the NBS domain. For example, there were 46 positively selected sites in Clade\_1260  
13 that are located in NBS domain while there are only two positively selected sites in LRR domain (Figure  
14 S4).

## 15 **Mapping Re-sequencing Reads Showed the Conservation of NBS Genes in *Citrus* Genomes.**

16 To verify that *Citrus sinensis* (sweet orange) is a backcross hybrid of *C. maxima* (pomelo) and *C.*  
17 *reticulata* (Mandarin orange) [21], Xu et al. re-sequenced three *C. maxima* cultivars and three *C.*  
18 *reticulata* cultivars and showed that SSR and SNP markers in the *C. sinensis* genome derive from the *C.*  
19 *maxima* and *C. reticulata* genomes at an approximately ratio of 1:3 [21]. Since our comparison of *C.*  
20 *maxima* and *C. reticulata* genomes showed that most of their NBS genes are very similar in sequence, we  
21 investigated the relationships amongst the NBS genes of *C. sinensis* (sweet orange), *C. maxima* (pomelo),  
22 and *C. reticulata* (Mandarin orange) to see if the evolutionary dynamics of NBS genes is informative.

23 We mapped the re-sequenced reads of three *Citrus maxima* (pomelo) genomes and three *C.*  
24 *reticulata* (Mandarin orange) genomes to hybridized genome of *Citrus sinensis* (sweet orange) China  
25 using BWA with mem method [27] and calculated the coverage of each NBS gene in *Citrus sinensis*

1 China. We detected 524 of 650 *Citrus sinensis* China NBS genes with more than 50% coverage in all of  
2 the six resequenced genomes. Additionally, 249 of 650 *Citrus sinensis* China NBS genes with more than  
3 85% coverage in all six resequenced genomes and 592 (91%) of *Citrus sinensis* China NBS genes with  
4 coverage over 85% in at least one of the resequenced *Citrus* genomes. Using a cutoff of 15% coverage as  
5 missing calls over the entire length of the gene, only 25 (<4%) of *Citrus sinensis* China NBS genes seem  
6 to be deleted in at least one of the six resequenced genomes. No NBS genes were lost amongst all of the  
7 six resequenced genomes.

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

21 **Motif Patterns in *Citrus* NBS Genes**

22 To further examine the gene structures of three *Citrus* NBS gene groups, we searched the motifs in the  
23 sequences of each group using MEME. Figure 2 lists the top 20 motifs identified by MEME from the  
24 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, GSQEIDKLCGGYCSKNCKSSYKFGKKVA, was present in 189 of 452 CC1 NBS genes. Both motifs have high level of sequence similarities with the CC NBS genes of *Arabidopsis* (Figure S6 (A)). The MEME found three motifs from CC2 group. Two of the motifs, KKLTNMLEMIKAVLDDAEEKQ and AVKLWLGKLDAAYDVEDVLDEFQTEALR were identified in 325 and 340 of 382 CC2 NBS genes, respectively. The motif AVKLWLGKLDAAYDVEDVLDEFQTEALR has high levels of similarity with the motif identified from the CC NBS genes of *Oryza sativa japonica* (Japonica rice) [28] (Figure S6 (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, AIEASAISVIIFSEGYASSRWCLDELVKI, GQIVIPVFYRVDPSDVRKQTG, and ENPEKVQKWRDALKEA were very similar to the TIR1-4 motifs identified from TIR NBS genes in *Arabidopsis* [23] and *Populus trichocarpa* [24] (Figure S6 (C)).

The MEME algorithm identified eight motifs from NBS domains of CC2 and TIR groups, which are similar to the motif structures of NBS domains in *Arabidopsis* [23] and *Populus trichocarpa* NBS genes [24]. MEME results also showed that the motif structure of NBS domain of CC1 groups is slightly different from those of CC2 and TIR groups. Ten motifs were identified from NBS domains of CC1 group with two extra RNBS motifs (RNBS-F, RNBS-G) between GPL and MHDV motifs. Five of eight motifs: P-loop, Kinase-2, RNBS-B, GPL and MHDV, showed high levels of sequence similarity amongst the three NBS groups, which were also similar to those motifs from *Arabidopsis* [23] and *Populus trichocarpa* NBS genes [24]. The MHDV motif in *Citrus* was often slightly modified to MHDL, as found in *Arabidopsis* and *Populus* previously. Meanwhile, the motifs RNBS-A, RNBS-C, and RNBS-D were quite dissimilar to each other amongst three groups.

1       The LRR domains in the C-terminal of NBS genes usually have high sequence diversity as they  
2       play a role in recognizing pathogen virulence proteins. The MEME algorithm identified five, six, and four  
3       motifs from LRR domains of CC1, CC2 and TIR NBS genes. Most of the motifs contained LxxL repeats.  
4       Motifs from different groups were highly variable in sequence, which implied that the three groups of  
5       NBS genes have different levels of sequence, and thus functional, constraint and play different roles in the  
6       *Citrus* immune system. Some motifs had repeated several times in the same gene. For examples, motif  
7       APNLKSLEVSSCxMEEIISV was found in 367 of 452 CC1 group with an average 4.8 motifs per gene;  
8       motif IKTLPESVCELYNLQTLLEGCRRLKKLP was found in 332 of 382 CC2 group with an average  
9       2.8 times per gene and motifs CKRLKSLPSSLCKLKGxLNLSGCSNLE and  
10      GNISELFLDGTAIEELPSSIE were found in 208 and 209 of 265 TIR group with 2.5 and 2.1 times per  
11      gene, respectively.

## 12     **Analysis of NBS Gene Clusters in *Citrus* Genomes**

13     The majority of *Citrus* NBS genes were physically clustered in genome (Table 3). 525 of 618 NBS genes  
14     in *C. clementina* were found in 108 clusters and 500 of 650 NBS genes in *C. sinensis* China were found in  
15     126 clusters (Table S2). Although the assembly of *C. sinensis* USA is more fragmentary, there still are  
16     207 of 508 NBS genes present in 72 clusters. The largest number of gene clusters in *C. clementina*, *C.*  
17     *sinensis* China and *C. sinensis* USA contain 55, 37 and 13 NBS genes, respectively.

18       Most clusters contain NBS genes from the same group. In *C. clementina*, there were 38 clusters  
19       with NBS genes of CC1 group, 40 clusters with NBS genes of CC3 group and 23 clusters with NBS  
20       genes of TIR group. Only seven out of 108 clusters contain the NBS genes from two or three groups. In  
21       *C. sinensis* China, there were 41 clusters with NBS genes of CC1 group, 44 clusters with NBS genes of  
22       CC2 group and 26 clusters with NBS genes of TIR group. There were 15 clusters containing NBS genes  
23       from two or three groups in *C. sinensis* China. In *C. sinensis* USA, there were 27 clusters with NBS genes  
24       of CC1 group, 25 clusters with NBS genes of CC2 group and 15 clusters with NBS genes of TIR group.

1 Only five clusters contain NBS genes from different groups in *C. sinensis* USA. The lower number of  
2 clusters in *C. sinensis* USA may due to its fragmental assembly.

3 The sequences of NBS genes within clusters are much more similar to each other than those  
4 between clusters (T-test p value < 2.2e-16). The mean identities between genes within and between  
5 clusters were 0.628 and 0.393 respectively (Figure 3). Furthermore, NBS genes in the same cluster tend to  
6 be in the same strand, which indicates that the NBS genes in the clusters are due to tandem duplication.  
7 We detected 204 pairs of tandem duplications in *C. clementina* and 217 pairs of tandem duplications in *C.*  
8 *sinensis* China using MCScanX [29]. The numbers of tandem duplications within CC1 and CC2 groups  
9 are much greater than that within the TIR group. Among 204 pairs in *C. clementina*, 90 and 85 pairs were  
10 present in the CC1 and CC2 groups and only 29 pairs from TIR group. Among 214 tandem gene pairs in  
11 *C. sinensis* China, 88 and 87 pairs were present in the CC1 and CC2 groups and only 42 pairs in the TIR  
12 group. The fewer tandem duplications of NBS genes in the TIR group may be the reason that there are  
13 fewer clusters of TIR groups in *Citrus* genomes.

14 We identified 254 and 246 gene conversion events from 76 NBS gene clusters in *C. clementina*  
15 and 75 NBS gene clusters in *C. sinensis* China, respectively (Table 4). The gene conversion events in *C.*  
16 *sinensis* USA is much less due to fragmental assembly. It is interesting that most of the conversion events  
17 (483 out of 520) were identified from the relatively small clusters with less than 10 NBS genes. Among  
18 these conversions, 119 events located in N-terminal, 178 in NBS domains and 223 in C-terminal, which  
19 indicating that there was no significantly bias in the location of conversion. Most gene conversions were  
20 between genes from the same group. We identified 101, 184 and 229 conversion events from NBS genes  
21 in TIR, CC1 and CC2 groups respectively and only four conversion events were identified between NBS  
22 genes of TIR and CC2 groups. While we could identify similar total numbers of conversion events in *C.*  
23 *clementina* and *C. sinensis* China, we found almost double such events in *C. sinensis* China TIR clusters  
24 versus in *C. clementina* TIR clusters. Meanwhile, we also found many more conversion events in *C.*  
25 *clementina* CC2 clusters than that in *C. sinensis* China CC2 clusters (Table 4).

1    **Analysis of NBS Orthologs**

2    We identified 719 *Citrus* NBS gene pairs of orthologs amongst *Citrus clementina*, *C. sinensis* China and  
3    *C. sinensis* USA. 270 orthologs were shared between *C. clementina* and *C. sinensis* China (Figure S7);  
4    227 orthologous gene pairs were shared between *C. clementina* and *C. sinensis* USA and 222 orthologous  
5    gene pairs were shared between *C. sinensis* China and *C. sinensis* USA. The percentages of identities  
6    between orthologous genes range from 53.8% to 100% (mean of 93.6%, median of 96.8%). The  
7    percentages of identities between orthologous genes from TIR group (TNL and TN,  $88.42 \pm 10.04$ ) were  
8    significantly (T-test:  $P < 2.6e-6$ ) lower than those from CC groups (CNL and CN,  $92.92 \pm 7.03$ ). The  
9    nonsynonymous divergence (dN) values of the 719 orthologs ranged from  $1.46e-6$  to 0.31 (mean of 0.03,  
10   median of 0.08) and the synonymous divergence (dS) values ranged from  $1.14e-5$  to 0.75 (mean of 0.05,  
11   median of 0.019) (Figure 4). The dN/dS ratios ranged from 0.001 to 50 with a median of 0.62. Similar to  
12   *Arabidopsis*, the dN and dS of orthologs from TIR group (TNL and TN) were higher than those from CC  
13   groups (CNL and CN). The dN values of orthologs from TIR group ranged from 0.001 to 0.24 (mean of  
14   0.038, median of 0.017) while those of orthologs from CC group ranged from  $1.4e-6$  to 0.23 (mean of  
15   0.025, median of 0.007). The dS of orthologs from TIR group ranged from  $3.27e-5$  to 0.76 (mean of  
16   0.078, median of 0.029) while those of orthologs from CC group ranged from  $1e-5$  to 49 (mean of 0.04,  
17   median of 0.016). However, the dN/dS ratios of orthologs from TIR group were relatively lower than that  
18   of orthologs from CC group (Figure 4(A)).

19       The dN and dS rates as well as dN/dS rate ratios of ortholog residues in clusters are generally  
20   greater than those of orthologous singletons (Figure 4(B)). The median dN/dS ratio of 148 orthologous  
21   singletons was 0.55 while that of 361 orthologs in clusters was 0.63. There were 124 orthologs with  
22   dN/dS ratios above 1. These orthologs could undergo positive selection pressure. Most of the positive  
23   selected orthologs (87.9%, 109 out of 124) belonged to CC groups (49 orthologs from CC1 group and 50  
24   from CC2 group).

1 We also detected 38 NBS gene syntenic blocks between *C. clementina* and *C. sinensis* China  
2 using MCScanX [29] (Figure S8). On average, there are 11 genes per block. The biggest block contains  
3 45 NBS genes and was located in scaffold\_5 of *C. clementina* and chromosome 3 of *C. sinensis* China.  
4 There were 80, 77 and 71 syntenic NBS genes in scaffold\_5 of *C. clementina*/chromosome 5 of *C.*  
5 *sinensis* China, scaffold\_3 of *C. clementina*/chromosome 5 of *C. sinensis* China and scaffold\_7 of *C.*  
6 *clementina*/chromosome 1 of *C. sinensis* China. Furthermore, there were 68, 66 and 8 syntenic NBS  
7 genes in the unknown chromosome of *C. sinensis* China corresponding to NBS genes in scaffold three,  
8 five and seven of *C. clementina*, respectively.

9 **Analysis of Conserved NBS Gene Clusters**

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

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

1 and one NBS gene was lost in *C. sinensis*. Furthermore, there was a recombination event within *C.*  
2 *clementina*.

3 **Mutations and Transposons in *Citrus* NBS Genes**

4 Plant NBS genes are continuously evolving. Sequence variation and structural constraints are shaped by  
5 gene birth-and-death processes [30, 31]. Besides possible interallelic recombination and gene conversion,  
6 gene mutations and transposable elements appear to play important roles in NBS gene evolution.

7 We compared the DNA sequences of NBS genes from different *Citrus* genomes to identify  
8 mutations. Approximately half of *Citrus* NBS genes have mutations maintained amongst corresponding  
9 orthologs. When a mutation took place in an exon and resulted in stop-codon gaining or frame-shift, the  
10 target gene often became a pseudogene. For example, Cs1g18610.1\_cc\_32, Cs1g18610.1 and  
11 orange1.1g003367m are orthologs of *C. clementina*, *C. sinensis* China and *C. sinensis* USA. There were 2  
12 stop-codon gaining mutations in Cs1g18610.1, which means that Cs1g18610.1 might become a pseudo  
13 (nonfunctional) gene after gaining these mutations.

14 LTR retrotransposons are widespread in eukaryotic genomes, especially plant genomes. We  
15 predicted 19014, 6296 and 1479 LTR retrotransposons with typical LTR characters in the draft genomes  
16 of *C. clementina*, *C. sinensis* China and *C. sinensis* USA, respectively. Then, we filtered out LTR  
17 retrotransposons with low similarity with known TE proteins using BLASTX with e-value greater than  
18 1e-5. Finally, 4920, 3726 and 1240 LTR retrotransposons remained in *C. clementina*, *C. sinensis* China  
19 and *C. sinensis* USA, respectively. We identified 33, 32 and 4 NBS genes that were inserted with LTR  
20 retrotransposons in *C. clementina*, *C. sinensis* China and *C. sinensis* USA respectively (Table S6). Most  
21 of these genes will likely become pseudogenes due to these insertions. For an instance,  
22 orange1.1g043039m, orange1.1g043039m\_cc\_116 and orange1.1g043039m\_csc\_123 were orthologs of  
23 *C. sinensis* USA, *C. clementina* and *C. sinensis* China respectively (Figure S10 (A)). The results of gene  
24 structure analysis showed that structure in orange1.1g043039m in *C. sinensis* USA seems relatively well  
25 maintained. But there is an about 10 kb fragment of LTR-retrotransposons in the corresponding

1 homologous gene in *C. clementina* and *C. sinensis* China (Figure S10 (B)). We identified both  
2 orange1.1g043039m\_cc\_116 and orange1.1g043039m\_csc\_123 as pseudogenes using PseudoPipe [32].

3 **Experimental Validation of One NBS Gene in *Citrus* Genomes**

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

17

18 **Discussion**

19 **Possible Common Ancestor of NBS Genes in Hybrid *Citrus sinensis* and Original *Citrus clementina***  
20 **Genome**

21 After carefully reannotating the *Citrus* genome sequences, we found similar numbers of NBS genes in  
22 *Citrus clementina* and *Citrus sinensis* China. There are slightly fewer NBS genes in *Citrus sinensis* USA,  
23 possibly due to the more fragmental assembly of this genome. In phylogenetic tree using NBS domains,  
24 the NBS genes from three different *Citrus* genomes are mixed together. After partitioning NBS genes on  
25 the tree into clades, we found that 97.7% of the clades containing more than three genes had members

1 from each of three *Citrus* genomes. This pattern suggests that these three *Citrus* genomes have similar  
2 NBS genes derived from common ancestors. Because *Citrus sinensis* is the hybrid of *C. reticulata*  
3 (Mandarin orange) and *C. maxima* (pomelo), it should be heterozygous and some of NBS genes in *Citrus*  
4 *sinensis* are expected have different genetic distances from NBS genes in *Citrus clementine*. This would  
5 support previous observations of NBS genes in F1 progeny of *Poncirus trifoliata* (trifoliate orange) and  
6 *Citrus reticulata* (tangerine). However, this is clearly not the case.

7 Furthermore, we mapped the resequenced reads of three *C. maxima* (pomelo) genomes and three  
8 *C. reticulata* (Mandarin orange) genomes onto the genome of *Citrus sinensis* China. In this case, 62% of  
9 *C. sinensis* China NBS genes have a copy present in all six resequenced genomes and 99% of *C. sinensis*  
10 China NBS genes have a copy in at least one of the resequenced genomes. The mapping results confirmed  
11 that a significant percentage of NBS genes of hybrid *C. sinensis* genomes have corresponding  
12 homologous genes in both the *C. maxima* and the *C. reticulata* genomes. Because the reference genome  
13 sequence of *C. maxima* is not yet available, the total number of NBS genes in *C. maxima* genomes is still  
14 not known. However, we can at least conclude from the mapping of resequenced genomes that the *C.*  
15 *maxima* genome has homologous copies of NBS genes in *C. reticulata* and *C. sinensis* genomes. The  
16 homologous NBS genes in *C. maxima* and *C. reticulata* may be the reason that NBS genes in their hybrid  
17 *Citrus sinensis* are similar to those in *C. reticulata* in this study.

18 **Three Groups of *Citrus* NBS Genes**

19 We identified 442, 393 and 264 genes with full length NBS domains from *Citrus clementina*, *Citrus*  
20 *sinensis* China and *Citrus sinensis* USA reference genomes, respectively. There are also many genes with  
21 short NBS domains in three *Citrus* genomes. The *Citrus* NBS genes can be divided into three groups  
22 according to the phylogenetic tree of NBS domains: two of them contain CC domain and the other group  
23 contains TIR domain. The number of CC NBS genes is three times of the number of TIR NBS genes. In  
24 most of the TIR NBS genes, we can find the LRR domains defined in Pfam database [33]. We only can  
25 identify LRR domains in small part of CC NBS genes using Pfam LRR domain definition. However, we

1 can find the LxxLs repeats in most of the CC NBS genes as shown in motifs from MEME. This implied  
2 that there may be other types of LRR domain in *Citrus* CC NBS genes. Our motif analyses also showed  
3 that motifs of TIR domains in *Citrus* CC2 and TIR NBS genes are similar to those of TIR domains in  
4 *Arabidopsis* [23] and *Populus trichocarpa* [24] TIR NBS genes. Furthermore, the motifs of *Citrus* CC  
5 domains in CC1 NBS genes are similar to those motifs of CC domains in Arabidopsis CC NBS genes and  
6 the motifs of *Citrus* CC domains in *Citrus* CC2 NBS genes are similar to those motifs of CC domains in  
7 japonica rice CC NBS genes [28]. The different structure of motifs in NBS domain and CC domain  
8 between *Citrus* CC1 NBS genes and *Citrus* CC2 NBS genes implied that they are from different  
9 evolutionary origin.

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

### 23 **Highly Clustering of *Citrus* NBS Genes**

24 The *Citrus* NBS genes are highly clustered in the genome. 84.9% of NBS genes in *Citrus clementina* and  
25 76.9% of NBS genes in *Citrus sinensis* China were found in clusters. Previous studies showed 76% of

1 rice [34], 64% of *A. lyrata* and 71% *A. thaliana* [15], 83.2% of grapevine and 67.5% of poplar [11] NBS  
2 genes are found in clusters. The percentage of *Citrus* NBS genes in clusters is in the high level comparing  
3 to other genomes. The average number of genes per clusters is 4.86 in *Citrus clementina* and 3.97 in  
4 *Citrus sinensis* China. These numbers are similar to those in other genomes [11]. Furthermore, most  
5 *Citrus* NBS genes in the same cluster belong to the same phylogenetic group, suggesting that tandem  
6 duplication is the primary mechanism for the expansion of NBS genes in the *Citrus* genus.

7 **Molecular Evolution of *Citrus* NBS Genes**

8 Similar to NBS genes in other genomes, *Citrus* NBS genes are highly dynamic and are shaped by several  
9 evolutionary processes leading to several differences amongst NBS genes, including domain presence and  
10 mutation constraints and genome organization. Our results revealed multiple molecular evolution events  
11 amongst *Citrus* NBS genes including gene duplications, gene conversions, mutation constraint changes,  
12 recombination and transposable element insertions. Likely these events support a birth-and-death process  
13 leading to the origins of new NBS genes as well as malfunction and loss of other NBS genes [14]. We  
14 found more than 200 tandem duplications in both *C. clementina* and *C. sinensis* China genomes alone.  
15 We also found that NBS genes become pseudogenes following original frame-shift mutations leading to  
16 mutation accumulations plus disruption of gene constraint leading to loss of function through transposable  
17 element insertions.

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

25

1    **Conclusions**

2    Our comparative analyses yield valuable insight into the understanding of the structure, evolution and  
3    organization of NBS genes in *Citrus* genomes. There are significantly more NBS genes in *Citrus*  
4    genomes compared to other plant species. *Citrus* NBS genes are structurally highly clustered. NBS genes  
5    in hybrid *C. sinensis* genomes are very similar to those in progenitor *C. clementina* genomes and their  
6    NBS genes may be derived from possible common ancestral gene copies. Furthermore, our  
7    comprehensive analysis also showed that there are three groups of plant NBS genes while NBS genes  
8    containing CC domains can be divided into two groups.

9

10    **Materials and Methods**

11    **Sequences Used**

12    We downloaded the draft genome sequences and the original gene annotations of *Citrus clementina*  
13    (clementine) and *Citrus sinensis* USA (sweet orange from USA) from the Citrus Genome Database  
14    (<http://www.citrusgenomedb.org/>) and those of *Citrus sinensis* China (Chinese sweet orange [21]) from  
15    the *Citrus sinensis* annotation project (<http://citrus.hzau.edu.cn/orange/>). The sizes of assembled genomes  
16    of *Citrus clementina*, *Citrus sinensis* China and *Citrus sinensis* USA are 301, 328 and 319 million  
17    basepairs, respectively. The *C. clementina* genomes assembled into nine major scaffolds and 95.8% of the  
18    sequences were assigned to those nine scaffolds. About 72.9% of the *C. sinensis* China genome assigned  
19    to nine chromosomes. The genome of *C. sinensis* USA only assembled to 12,574 scaffolds and the N50 of  
20    scaffolds is 250 kb. The original gene annotations of *Citrus clementina*, *Citrus sinensis* China and *Citrus*  
21    *sinensis* USA have 24,533, 29,385 and 25,397 genes respectively. We also downloaded the resequence  
22    data of three *Citrus clementina* (clementine) and three *Citrus maxima* (pomelo) genomes from the *Citrus*  
23    *sinensis* annotation project (<http://citrus.hzau.edu.cn/orange/>).

24    **Identification of NBS Genes in *Citrus* Genomes**

1 We first screened the original predicted citrus open reading frames (ORFs) using hmmsearch [35] with  
2 the hidden Markov models (HMM) of Pfam [36] for NBS domain presence (NB-ARC, PF00931) using  
3 an e-value cut-off of 0.1 for the hmmsearch. Then, the proteins selected by the HMM were searched  
4 against the Swiss protein database [37] to confirm the annotation using BLASTP [38]. Only the proteins  
5 that have a significant match (e-value < 1E-5 in BLASTP search) with the NBS proteins or resistant  
6 proteins in the Swiss protein database were identified as NBS-containing proteins. To recover possible  
7 NBS genes that may be missed in the original gene annotations, we mapped the identified NBS genes to  
8 the draft genome using TBLASTN. The matched sequences with e-value < 1E-5 were then predicted  
9 using Genewise [39]. The new genes predicted by Genewise were also confirmed using NBS domain  
10 HMM screening and by BLASTP searching through the Swiss protein database.

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

12 Orthologous NBS genes of *C. clementina* and *C. sinensis* were identified using the reciprocal best blast  
13 method [40]. We used NBS genes of *C. clementina* as query sequences to search against the NBS genes  
14 of *C. sinensis* and *vice versa*. Protein pairs with reciprocal best hits of e-value < 1E-20 were defined as  
15 orthologs.

16 To calculate the rates of nonsynonymous, synonymous and their rate ratio (dN, dS and dN/dS) of  
17 orthologous pairs, we first aligned orthologous protein sequence pairs using mafft, and then converted the  
18 protein alignments to codon-based alignments using PAL2NAL [41]. We calculated the dN, dS and  
19 dN/dS ratios using the codeml program in PAML version 4.7 [42].

20 **Phylogenetic Analysis**

21 We constructed the phylogenetic tree of NBS genes of the three citrus genomes using only the conserved  
22 NBS domain. Only 1,099 NBS domain sequences that have both the P-loop and the MHDV motifs and  
23 were longer than 70% (200 amino acids) of the full-length NBS domain were included. Next, NBS  
24 domain sequences were aligned in mafft [43] using an auto alignment model and a best fit maximum  
25 likelihood phylogenetic tree of NBS genes was constructed using FastTree [25] with default parameters

1 (JTT+CAT). The resultant best fit phylogenetic tree was divided into 3 main groups based upon clade  
2 support. For each group, we constructed a new ML tree with *Streptomyces coelicolor* protein P25941 as  
3 the outgroup using FastTree [25]. The average identity of each group was calculated using a liststat  
4 implemented in SQUID (<http://selab.janelia.org/software.html>). We further partitioned the NBS gene tree  
5 into clades using the depth-first phylogeny partition method in PhyloPart [44] with distance threshold  
6 0.025. This clustered the NBS genes into 114 clades, which contain more than 2 genes, and 8 orphan  
7 genes.

8 **Domain and Motif Annotation**

9 The Toll-Interleukin receptor (TIR) domains in *Citrus* NBS-containing proteins were identified using  
10 hmmsearch [35] with the HMM model of Pfam domain PF01582 and a 0.1 e-value cut-off. The Leucine-  
11 Rich Repeat (LRR) domains in *Citrus* NBS containing proteins were identified using the HMM models of  
12 Pfam LRR domains with e-value cut-off of 0.1. As long as there is a significant hit to one of LRR  
13 domain models (e-value < 0.1), it was defined as an LRR-containing protein. We used MARCOIL [45]  
14 with a threshold probability of 90 and COILS [46] with a threshold of 0.9 to search for Coiled Coil (CC)  
15 domains in the N-terminal region of *Citrus* NBS-containing protein. We considered a protein as a CC-  
16 containing protein if either MARCOIL or COILS reported a CC domain in it.

17 We identified 20 motifs amongst the NBS genes in each of the three main phylogenetic groups  
18 separately using MEME SUITE [47]. The motif width was set to between 6 and 50 for MEME. Then, we  
19 searched the motif structure of all genes in each group using MAST with default parameters (-ev 10 -mt  
20 0.0001).

21 **Pseudogene Identification**

22 We identified possible pseudogenes using PseudoPipe [32] with default parameters (-e 0.1). The  
23 PseudoPipe algorithm identifies pseudo genes by integrating sequence similarity, intron-exon structure,  
24 plus presence of stop codons and frame-shifts. We used all *Citrus* NBS genes to search *Citrus* genomes  
25 for potential NGS pseudo genes.

1    **Transposon Identification**

2    All long terminal repeat (LTR) retrotransposons in each citrus genome were identified using LTR finder  
3    [48] with default parameters (-o 3 -t 1 -e 1 -m 2 -u -2). Then, we used a script program to match the  
4    location of the LTR transposons to the NBS-LRR genes in each citrus genome.

5    **NBS Gene Synteny Identification**

6    Gene synteny was identified and defined using MCScanX [29] with default parameters (-A -u 5000). First,  
7    NBS genes between two genomes were aligned using BLASTP and the matches with E-value < 1e-5 were  
8    sorted according to their chromosome positions. Synteny scores were then calculated for each block based  
9    upon gene position. Two genes were considered in the same block if there were fewer than 25 genes  
10   separating them. MCScanX reported blocks with at least 5 collinear gene pairs.

11   **Gene Cluster Analysis**

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

17   **Gene Conversion Detection**

18   We first aligned the sequences of NBS genes in the same cluster using mafft [43]. Then, we used  
19   GENECONV [49] version 1.81a with default settings (N=10,000) to detect gene conversions.  
20   GENECONV identifies gene conversions by finding identical fragments between pairs of sequences in a  
21   nucleotide alignment. A global *P* value  $\leq 0.05$  was used to assess the statistical significance of the  
22   observed conversions. GENECONV requires at least three sequences for analyses in order to account for  
23   shared ancestral states. Thus, we only detected conversions in clusters containing three or more genes.

24   **Tests for Sites under Positive Selection**

1 The amino acid sequences of NBS genes from the same clade were aligned with mafft [43]. Then we  
2 converted the protein alignments to codon-based alignment using PAL2NAL [41]. The positively selected  
3 sites were statistically identified using the Bayesian approach implemented in codeml within PAML [42].  
4 We also further examined sites in the  $\omega > 1$  class with >90% posterior probability.

5 **Mapping of Re-sequencing Data of *Citrus clementina* and *Citrus maxima***

6 We mapped the raw reads of each re-sequenced sample to the draft genome of *Citrus sinensis* China using  
7 BWA [27] with default parameters (-k 19 -d 100 -A 1 -B 4 -O 6). Then, the mapped reads of NBS genes  
8 regions were extracted using BEDtools [50]. Two types of coverage of each NBS gene in *Citrus sinensis*  
9 were calculated. One coverage type divides the length of mapped sequences by the whole gene sequence  
10 length and the other type divides the length of mapped sequences by the length of exon sequence only.

11 ***Citrus* DNA Extraction and PCR Amplification**

12 *Citrus* leaf samples were collected from the six *Citrus* plants in USHRL's (USDA Horticultural Research  
13 Laboratory, Fort Pierce, Florida): *Citrus sinensis* (sweet, Navel orange), *Citrus aurantium* (Kumquat, jamur,  
14 sour orange), *Citrus reticulata* (Mandarin orange), *Citrus clementina* (Clementine), *Citrus aurantiifolia*  
15 (sweet lime), *Citrus japonica* (Yuzu, kumquat), and *Citrus maxima* (pomelo). Total DNA was extracted  
16 from leaf midribs following the DNeasy® Plant Mini Kit standard protocol (Qiagen Inc., Valencia, CA),  
17 followed by DNA quantity and quality evaluation with Nanodrop. We chose the NBS gene, Cs1g09350,  
18 which was conserved in *C. clementina* and *C. sinensis* for validating the conservation of NBS gene  
19 among different *Citrus* genomes. Primers used in this study were designed using Oligo 7.23 (Molecular  
20 Biology Insights, Inc., Cascade, CO, USA). High Fidelity Platinum® Taq DNA Polymerase (Invitrogen,  
21 Carlsbad, CA, USA) was used to amplify the NBS-LRR genes from *Citrus* DNA. For PCR, 20 µL  
22 reactions using standard conditions provided by the manufacturer for High Fidelity Platinum® Taq DNA  
23 Polymerase. PCR was performed using an initial denaturation at 95°C for 3 minutes, 35 cycles of 94°C  
24 for 20 seconds, 50-52°C for 20 seconds (specified by different primer sets) and 68°C for 3 minutes, follow  
25 by final extension at 68°C for 10 minutes in a C1000™ Thermal Cycler (Bio-Rad, Hercules, CA). The

1 cloning and sequencing analysis of amplified PCR products were conducted as previously  
2 described [51]

3

4 **Competing interests**

5 The authors declare no conflicts of interest.

6

7 **Authors' Contribution**

8 FL, YD, PS and LD designed the research. YW performed the bioinformatics studies. LZ performed the  
9 experimental verification. YW, DL, ALR and FL performed the genetic and evolution analyses. ALR, PS,  
10 LD, YD and FL wrote the manuscript. All authors read and approved the final manuscript

11

12 **Acknowledgement**

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

14

15 **References:**

- 16 1. Dangl JL, Jones JD: **Plant pathogens and integrated defence responses to infection.** *Nature*  
17 2001, **411**(6839):826-833.
- 18 2. Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B: **Direct interaction of resistance gene**  
19 **and avirulence gene products confers rice blast resistance.** *The EMBO Journal* 2000,  
20 **19**(15):4004-4014.
- 21 3. Ellis JG, Lawrence GJ, Luck JE, Dodds PN: **Identification of regions in alleles of the flax rust**  
22 **resistance gene L that determine differences in gene-for-gene specificity.** *The Plant Cell*  
23 *Online* 1999, **11**(3):495-506.
- 24 4. Ellis J, Dodds P, Pryor T: **The generation of plant disease resistance gene specificities.** *Trends*  
25 *in plant science* 2000, **5**(9):373-379.

- 1      5. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW: **Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis**. *The Plant Cell Online* 2003, **15**(4):809-834.
- 2
- 3      6. Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W, Martin F: **Genome-wide identification of NBS resistance genes in Populus trichocarpa**. *Plant molecular biology* 2008, **66**(6):619-636.
- 4
- 5
- 6      7. 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):617-631.
- 7
- 8      8. 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.
- 9
- 10
- 11     9. Sun X, Wang G-L: **Genome-wide identification, characterization and phylogenetic analysis of the rice LRR-kinases**. *PloS one* 2011, **6**(3):e16079.
- 12
- 13     10. Zhou T, Wang Y, Chen J-Q, Araki H, Jing Z, Jiang K, Shen J, Tian D: **Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes**. *Molecular Genetics and Genomics* 2004, **271**(4):402-415.
- 14
- 15
- 16     11. 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):187-198.
- 17
- 18
- 19     12. Ameline-Torregrosa C, Wang B-B, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB: **Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant Medicago truncatula**. *Plant physiology* 2008, **146**(1):5-21.
- 20
- 21
- 22     13. 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):203-213.
- 23
- 24     14. 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):1113-1130.
- 25
- 26     15. 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):757-769.
- 27
- 28

- 1      16. Yang S, Li J, Zhang X, Zhang Q, Huang J, Chen J-Q, Hartl DL, Tian D: **Rapidly evolving R**  
2      **genes in diverse grass species confer resistance to rice blast disease.** *Proceedings of the*  
3      *National Academy of Sciences* 2013.
- 4      17. Barrett H, Rhodes A: **A numerical taxonomic study of affinity relationships in cultivated**  
5      **Citrus and its close relatives.** *Systematic Botany* 1976;105-136.
- 6      18. Scora RW: **On the history and origin of Citrus.** *Bulletin of the Torrey Botanical Club*  
7      1975;369-375.
- 8      19. Nicolosi E, Deng Z, Gentile A, La Malfa S, Continella G, Tribulato E: **Citrus phylogeny and**  
9      **genetic origin of important species as investigated by molecular markers.** *Theoretical and*  
10     *Applied Genetics* 2000, **100**(8):1155-1166.
- 11     20. Li X, Xie R, Lu Z, Zhou Z: **The origin of cultivated citrus as inferred from internal**  
12     **transcribed spacer and chloroplast DNA sequence and amplified fragment length**  
13     **polymorphism fingerprints.** *Journal of the American Society for Horticultural Science* 2010,  
14     **135**(4):341-350.
- 15     21. Xu Q, Chen LL, Ruan X, Chen D, Zhu A, Chen C, Bertrand D, Jiao WB, Hao BH, Lyon MP *et*  
16     *al:* **The draft genome of sweet orange (*Citrus sinensis*).** *Nature genetics* 2013, **45**(1):59-66.
- 17     22. Xu Q, Biswas MK, Lan H, Zeng W, Liu C, Xu J, Deng X: **Phylogenetic and evolutionary**  
18     **analysis of NBS-encoding genes in Rutaceae fruit crops.** *Molecular Genetics and Genomics*  
19     2011, **285**(2):151-161.
- 20     23. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW: **Genome-wide analysis of NBS-**  
21     **LRR-encoding genes in *Arabidopsis*.** *The Plant cell* 2003, **15**(4):809-834.
- 22     24. Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W,  
23     Martin F: **Genome-wide identification of NBS resistance genes in *Populus trichocarpa*.** *Plant*  
24     *molecular biology* 2008, **66**(6):619-636.
- 25     25. Price MN, Dehal PS, Arkin AP: **FastTree: computing large minimum evolution trees with**  
26     **profiles instead of a distance matrix.** *Molecular biology and evolution* 2009, **26**(7):1641-1650.
- 27     26. Plocik A, Layden J, Kesseli R: **Comparative analysis of NBS domain sequences of NBS-LRR**  
28     **disease resistance genes from sunflower, lettuce, and chicory.** *Molecular phylogenetics and*  
29     *evolution* 2004, **31**(1):153-163.

- 1    27. Li H: Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.  
2    *arXiv:1303.3997v2* 2013.
- 3    28. Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D: Genome-wide  
4    identification of NBS genes in japonica rice reveals significant expansion of divergent non-  
5    TIR NBS-LRR genes. *Molecular genetics and genomics : MGG* 2004, **271**(4):402-415.
- 6    29. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al*:  
7    **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and**  
8    **collinearity.** *Nucleic acids research* 2012, **40**(7):e49.
- 9    30. Michelmore RW, Meyers BC: Clusters of resistance genes in plants evolve by divergent  
10   selection and a birth-and-death process. *Genome research* 1998, **8**(11):1113-1130.
- 11   31. Tan S, Wu S: Genome Wide Analysis of Nucleotide-Binding Site Disease Resistance Genes in  
12   **Brachypodium distachyon.** *Comparative and functional genomics* 2012, **2012**:418208.
- 13   32. Zhang Z, Carriero N, Zheng D, Karro J, Harrison PM, Gerstein M: **PseudoPipe: an automated**  
14   **pseudogene identification pipeline.** *Bioinformatics* 2006, **22**(12):1437-1439.
- 15   33. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G,  
16   Clements J *et al*: **The Pfam protein families database.** *Nucleic Acids Research* 2012,  
17   **40**(D1):D290-D301.
- 18   34. Yang S, Feng Z, Zhang X, Jiang K, Jin X, Hang Y, Chen J-Q, Tian D: **Genome-wide**  
19   **investigation on the genetic variations of rice disease resistance genes.** *Plant molecular*  
20   *biology* 2006, **62**(1-2):181-193.
- 21   35. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity**  
22   **searching.** *Nucleic acids research* 2011, **39**(Web Server issue):W29-37.
- 23   36. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G,  
24   Clements J *et al*: **The Pfam protein families database.** *Nucleic acids research* 2012,  
25   **40**(Database issue):D290-301.
- 26   37. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ,  
27   Michoud K, O'Donovan C, Phan I *et al*: **The SWISS-PROT protein knowledgebase and its**  
28   **supplement TrEMBL in 2003.** *Nucleic acids research* 2003, **31**(1):365-370.

- 1    38. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped**  
2    **BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic*  
3    *acids research* 1997, **25**(17):3389-3402.
- 4    39. Birney E, Clamp M, Durbin R: **GeneWise and Genomewise.** *Genome research* 2004, **14**(5):988-  
5    995.
- 6    40. Moreno-Hagelsieb G, Latimer K: **Choosing BLAST options for better detection of orthologs**  
7    **as reciprocal best hits.** *Bioinformatics* 2008, **24**(3):319-324.
- 8    41. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence**  
9    **alignments into the corresponding codon alignments.** *Nucleic acids research* 2006, **34**(Web  
10    Server issue):W609-612.
- 11    42. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Molecular biology and*  
12    *evolution* 2007, **24**(8):1586-1591.
- 13    43. Katoh K, Standley DM: **MAFFT Multiple Sequence Alignment Software Version 7:**  
14    **Improvements in Performance and Usability.** *Molecular biology and evolution* 2013,  
15    **30**(4):772-780.
- 16    44. Prosperi MC, Ciccozzi M, Fanti I, Saladini F, Pecorari M, Borghi V, Di Giambenedetto S,  
17    Bruzzone B, Capetti A, Vivarelli A *et al*: **A novel methodology for large-scale phylogeny**  
18    **partition.** *Nature communications* 2011, **2**:321.
- 19    45. Delorenzi M, Speed T: **An HMM model for coiled-coil domains and a comparison with**  
20    **PSSM-based predictions.** *Bioinformatics* 2002, **18**(4):617-625.
- 21    46. Lupas A: **Prediction and analysis of coiled-coil structures.** *Methods in enzymology* 1996,  
22    **266**:513-525.
- 23    47. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS:  
24    **MEME SUITE: tools for motif discovery and searching.** *Nucleic acids research* 2009, **37**(Web  
25    Server issue):W202-208.
- 26    48. Xu Z, Wang H: **LTR\_FINDER: an efficient tool for the prediction of full-length LTR**  
27    **retrotransposons.** *Nucleic acids research* 2007, **35**(Web Server issue):W265-268.
- 28    49. Sawyer S: **Statistical tests for detecting gene conversion.** *Molecular biology and evolution*  
29    1989, **6**(5):526-538.

- 1    50. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.**  
2    *Bioinformatics* 2010, **26**(6):841-842.
- 3    51. Zhou L, Powell CA, Hoffman MT, Li W, Fan G, Liu B, Lin H, Duan Y: **Diversity and plasticity**  
4    **of the intracellular plant pathogen and insect symbiont "Candidatus Liberibacter**  
5    **asiaticus" as revealed by hypervariable prophage genes with intragenic tandem repeats.**  
6    *Applied and environmental microbiology* 2011, **77**(18):6663-6673.
- 7  
8

9

10

11 **Figure Legends:**

12 **Figure 1. Maximum likelihood phylogenetic tree of *Citrus* NBS-LRR genes constructed from**  
13 **multiple sequences alignment of NBS domain.**

14 There are three groups in the phylogenetic tree: CC1, CC2 and TIR. Clades were classified using  
15 PhyloPart as shown in alternating color. The outer circle shows species with *Citrus clementina* (Cc) in  
16 red, *Citrus sinensis* China (CsCN) in green and *Citrus sinensis* USA (CsUSA) in blue.

17

18 **Figure 2. Architecture of NBS gene motifs in *Citrus*.**

19 Top 20 motifs identified by MEME for each group of citrus NBS genes are listed. The motifs underlined  
20 in green were repeated in C-terminal (LRR domain).

21

22 **Figure 3. Percentage identity distribution of NBS-encoding genes in *C. clementina* (Cc) and *C.***  
23 ***sinensis* China (Csc).**

24 Green lines indicate pairwise identity distribution of inter-cluster NBS-encoding genes. Orange lines  
25 indicate intro-cluster NBS-encoding genes.

26

1    **Figure 4. Divergence of orthologous NBS-encoding genes amongst *Citrus* species.**

2    A) Divergence of different classes. B. Divergence of singletons and cluster members.

3

4    **Figure 5. Duplication histories of one of the conserved NBS-encoding genes cluster in *C. clementina***  
5    **and *C. sinensis*. A.**

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

10

11    **Figure 6. Phylogenetic analysis of NBS genes of *Citrus clementina*, *Arabidopsis thaliana*, *Populus***  
12    ***trichocarpa*, *Oryza sativa* and *Vitis* (grape).**

13    Species are indicated by the color of the outer circle and branches.

14

15    **Tables:**

16    Table 1. Classification of citrus NBS genes

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

17

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

23

1 Table 2. Type of citrus NBS gene clades

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

2

3

4 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>Citrus clementina</i>	618	525	108	84.9
<i>Citrus sinensis</i> China	650	500	126	76.9
<i>Citrus sinensis</i> USA	508	207	72	40.7

5

6 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>Citrus clementina</i>	TIR	27	35	32	111
	CC1	43	86	67	137
	CC2	46	133	83	146
<i>Citrus sinensis</i> China	TIR	36	69	42	137
	CC1	51	89	48	162
	CC2	57	92	75	116

7

8

1    **Support Information**

2    **Supplemental Figures**

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

5

6    **Figure S2.** Phylogenetic trees of three groups of citrus NBS gene with P25941 as outgroup. A, CC1  
7    group; B, CC2 group and C, TIR group.

8

9    **Figure S3.** Structure of domains of citrus NBS genes. The red rectangles indicated the NBS domains, the  
10   green ellipses indicated CC domains, the left pointing pentagrams indicated TIR domains and the left  
11   pointing triangle indicated LRR domains.

12

13   **Figure S4.** Positive selection sites of Clade\_1260. The grey box indicated the NBS domain.

14

15   **Figure S5.** The exon coverage of NBS-encoding genes in the resequences of citrus genome (3 mandarin  
16   and 3 pomelo).

17

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

19

20   **Figure S7.** Orthologs of citrus NBS-encoding genes of Cc and CsCN and their genome locations. The  
21   outer blue circles indicated 9 largest scaffolds of Cc and orange ones indicated 9 chromosomes plus an  
22   un-located pseudo-chromosome. The NBS-encoding genes were arranged due to their locations on the  
23   chromosomes. The NBS genes in the TIR group were indicated with orange links, CC1 in blue and CC2

1 in light blue ones. The NBS-encoding orthologs of *C. clementina* and *C. sinensis* were indicated with the  
2 link lines.

3

4 **Figure S8.** Dot plotting of syntenic orthologs of the NBS genes between *C. clementina* (Cc) and *C.*  
5 *sinensis* China (CsCN) produced by MCScanX. The numbers in the grids indicated the number of  
6 syntenic ortholog pairs in the corresponding scaffolds. The scaffolds of *C. clementina* (Cc) were arranged  
7 in x-axis and that of *C. sinensis* China (CsCN) were arranged in y-axis.

8

9 **Figure S9.** Diagram of conserved cluster between *C. clementina* and *C. sinensis*. The nodes with orange  
10 color indicated the clusters of *C. clementina* and the blue ones indicated clusters of *C. sinensis* and size of  
11 the nodes were relative to cluster size. The numbers on the edges indicated the number of orthologs could  
12 be found from the other species.

13

14 **Figure S10.** Pseudogenization of NBS-encoding genes of citrus due to mutation (A) and retrotransposon  
15 insertion (B). A: Sequences alignment of Cs1g18610.1, Cs1g18610.1\_cc\_32 and orange1.1g003367m.  
16 The bases indicated the variation bases in each citrus species. The star in red color represented stop-codon  
17 gaining mutation and may result in pseudogene of Cs1g18610.1. B: LTR retrotransposon insertion in  
18 NBS-encoding genes orange1.1g043039m\_cc\_116 and orange1.1g043039m\_csc\_123 from Cc and CsCN  
19 respectively, and the multiple sequences alignment of the orthologs. LTR, Long terminal repeats; TSR,  
20 Target site repeats.

21

22 **Figure S11.** Validation and phylogenetic analysis of NBS-LRR gene, Cs1g09350.1, in different citrus  
23 species. A) multiple sequences alignment, the blue bars indicate the exons and the thin lines indicate the  
24 introns. The vertical lines in each alignment blocks indicate mutations comparing with the reference

1 sequences from *C. sinensis*. B) neighbor-joining phylogenetic tree of the orthologs of Cs1g09350.1 in  
2 different citrus species and the bootstrapping values displayed on the branches.

3

4 **Supplemental Tables**

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

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

7 **Table S3.** LRR domain distribution of citrus NBS gene

8 **Table S4.** Positive selection sites distribution


9 **Table S5.** The exome coverage of NBS genes from the 3 clementine and 3 pummelo resequencing  
10 genomes

11 **Table S6.** Citrus NBS gene interupted with LTR transposons

12

Figure 1

*C. clementina*  
 *C. sinensis* China  
 *C. sinensis* USA



# Figure 2

## CC1

EEQQQMRLNQVQGWLSRVEA GSQEIDKLCLGGYCSKNCKSSYKFGKKVA xEENVGIIGLY**MGGVGKTTLLQINNKF** FDVVIWVVVSQDLDL  
P-loop RNBS-A  
LSEKKFLL**I**DDIWER GSKIVFTTRsxEVCG NFKVECLSEEAEWELFKKKVG HPDIPELAQTVAKECG**GPL**ALITIGRAM KKTPEEWKDAIEVLRRSASKF  
Kinase-2 RNBS-B RNBS-C GPL RNBS-D  
VYSRLKFSYDSLPSD SCFLYCSLFPEDYEIPKEELIDY EARNRGYTIIGxLKHACLLEE EGExVKM**HDV**IRDIMALWIAS LPEIPTCPHLLTLFL  
RNBS-E RNBS-F RNBS-G MHDV  
IPDNFFQFMPSLKVL DLSGSDIEELPEELKALTNLKLLNLEYC LKVIPPNVISNLSRLEELRMF EALLEELLGLKHLTVLELTLRSSHAL  
APNLKSLEVSSCxMEEIISV PLPFPHLKEISVSGCPKLKKLPLDSNSAKERK

## CC2

KKLTNMLEMIAVLDAAEKQ AVKLWLGLKDAAYDVEDVLDEFQTEALR GNGDPAAQDQPSSRTTSKFRKLIPTCCTFTPQSIQFDYALMSKIKE  
VIPIV**MGGLGKTT**LAQLVYNDKRVQDHF AWCVSDDFDVIRLKAILES **GKKFLLVLDWVNENYNDWEP** LKRPLKAGAPGS**KIVTTR**NQEVASIMGT  
P-loop RNBS-A Kinase-2 RNBS-B  
AYQLKKLSEEDCWSLFAQHAF LEEIGKKIVxKCK**GPL**AAKTLGGLLRGK LKQCFAYCSLFPKDYEFEELILLWMAQGFLDH  
RNBS-C GPL RNBS-D  
MEDLGREYFQELASRSFFQQS **FKMHDLVHDLAQWVAGEECFTLEYNSEVN** EMPVGIGKLTCLQTLSNFVVG LDMLQPHKNLEQLCICGYGGT  
RNBS-E MHDV  
LEDCGMCTALPSVGQLPSLKHTVRGMSRVKRLGSEFYG PIPFPCLETRLFEDLQEWEWDWIPHGSQGVEGFPKLRELHILRCSKLQGT  
GCKVVWRSATDHLGSQNSVCRDTSNQVFLAGPLKPRIPK ETYIWKSHNELLQDICSLKRLTIDSCPQLQSLVAEEKDQQQQLCELSCR  
IKLPESVCELYNLQTLDEGCRLKKLP HIRGNMEIWKSMIERGRGFHRFSSLRHLTISGCDDDMVSFP

## TIR

YDVFLSFRGEDTRDNFTSHLY AIEASAISVIIFSEGYASSRWCLDELVKI GQIVIPVFYRVDPDSVORKQTG KENPEKVQKWRDALKEAANLSGF  
NKNGLVGVESRIEIESLLGVGSKDVYAL **GIW****MGGIGKTT**IARAIFDKI SREFEGSCFLANVREESEKGG INLIFRRLSRMKV**LIVFDDVT**  
P-loop RNBS-A Kinase-2  
QLESLIGSLDWFGPGSRI**ITRD**KQVLR GVRKIYEMKALEYDEALELFSRHAKQNHLSSKVVKYAQ**GPL**ALKVLSFL ISYDGLDDKEKNIFLDIACFF  
RNBS-B RNBS-C GPL RNBS-D  
GEDRDLVMKFLDACGFYPEIGISVLVDKS NKIT**MHDLLQEMGREIVRQES** WHHEDIYEVLTKNTGTEKIEGICLD MSKVKEIHLNPNTFTKMPKLRFLKFY  
RNBS-E MHDV  
LFAELRYLHWGYPLKSLPSN CKRLKSLPSSLCKLKGxLNLSGCSNLE GNISELFLDGTAIEELPSSIE NLSKNNFERLPESIQLSKLRYLLSYCERLQSLPE

Figure 3

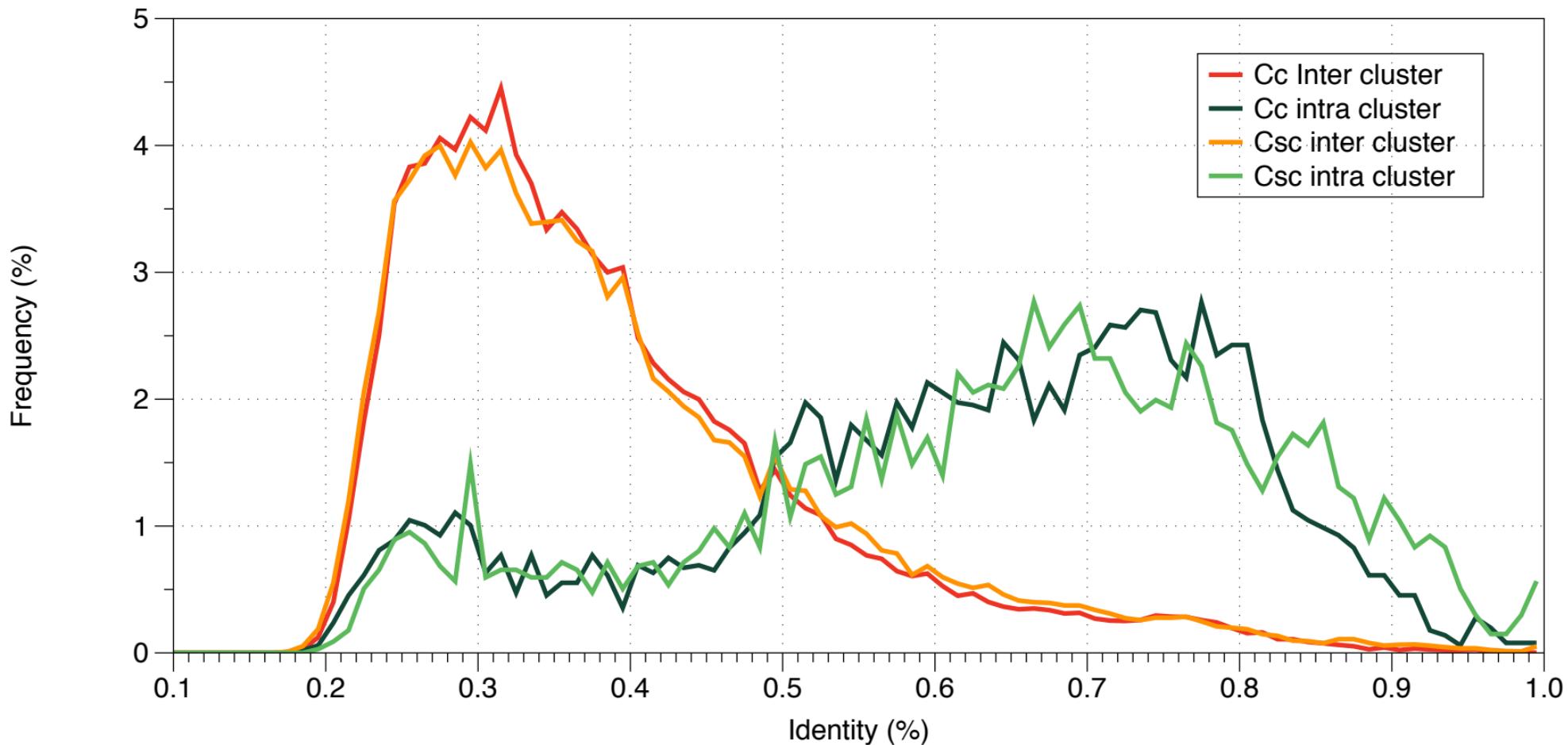


Figure 4

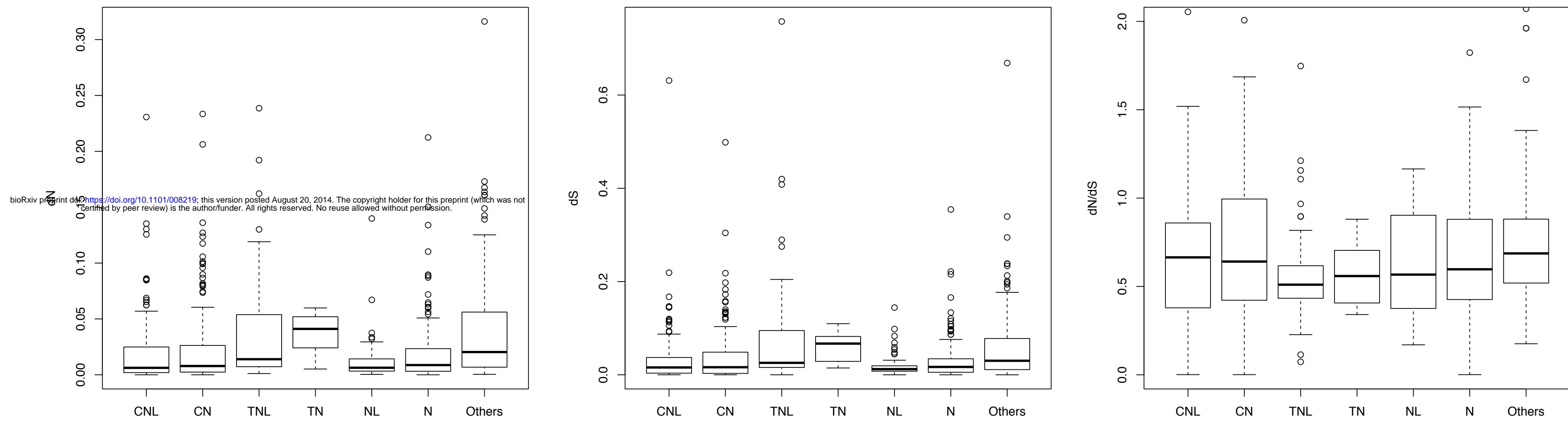
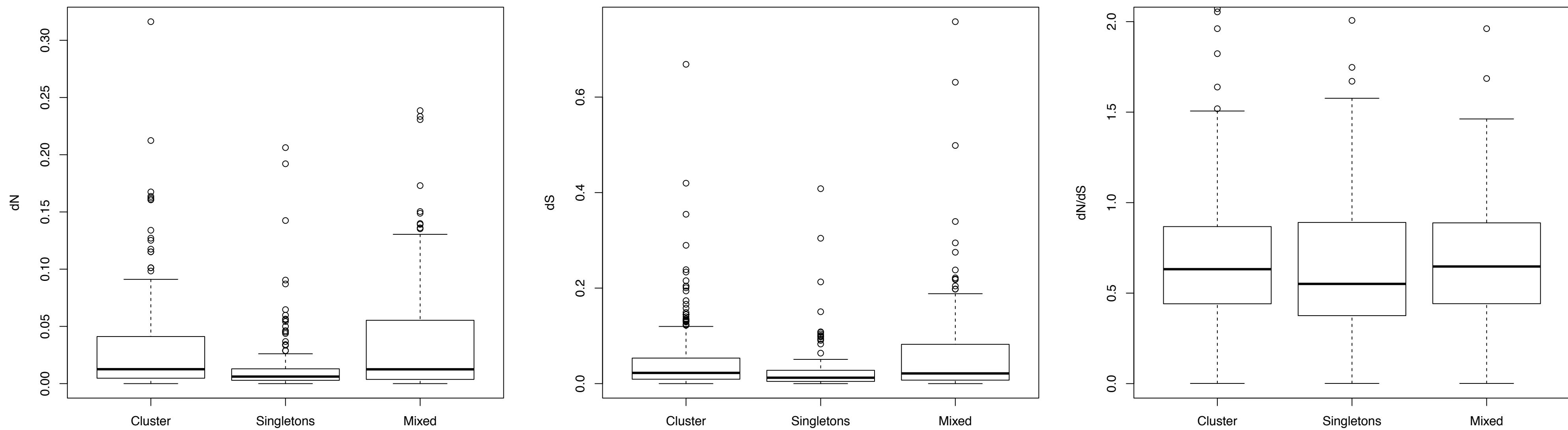
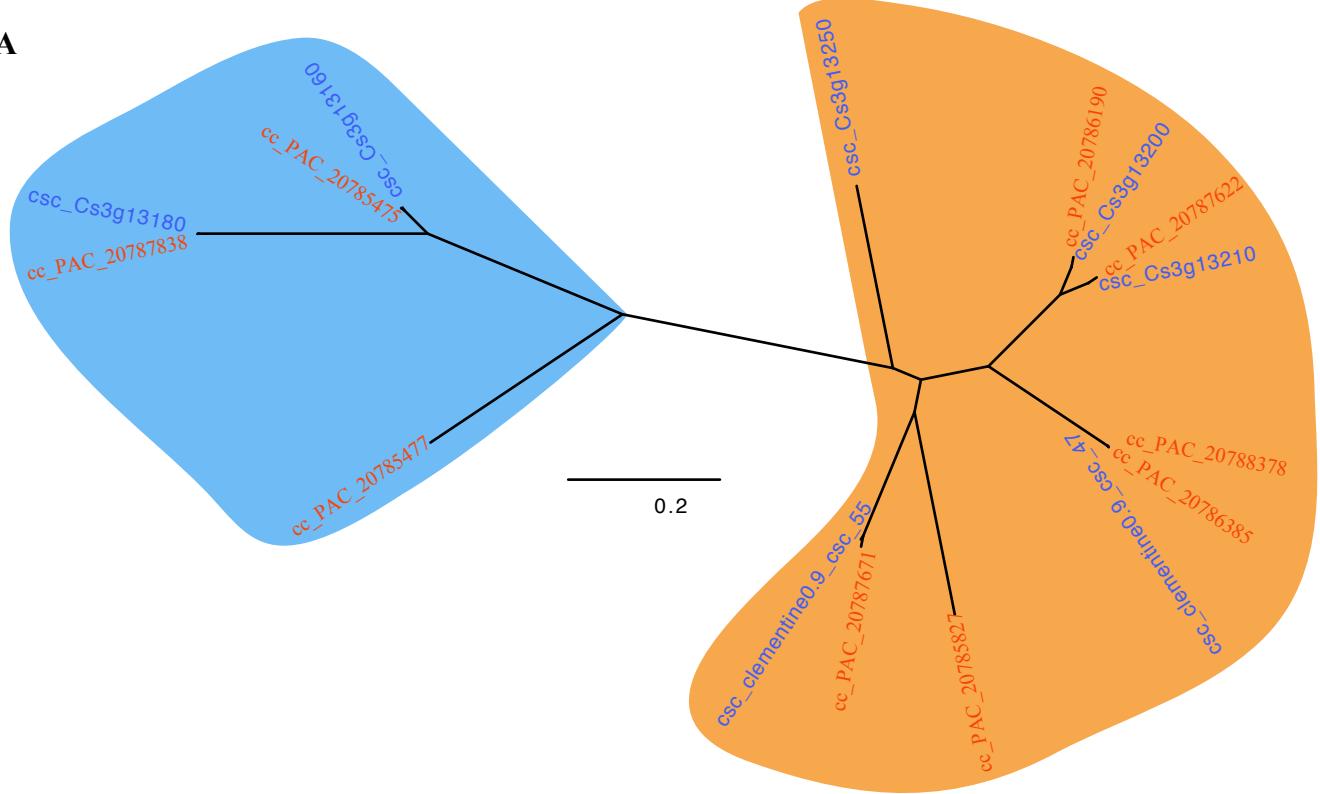
**A****B**

Figure 5

A



B

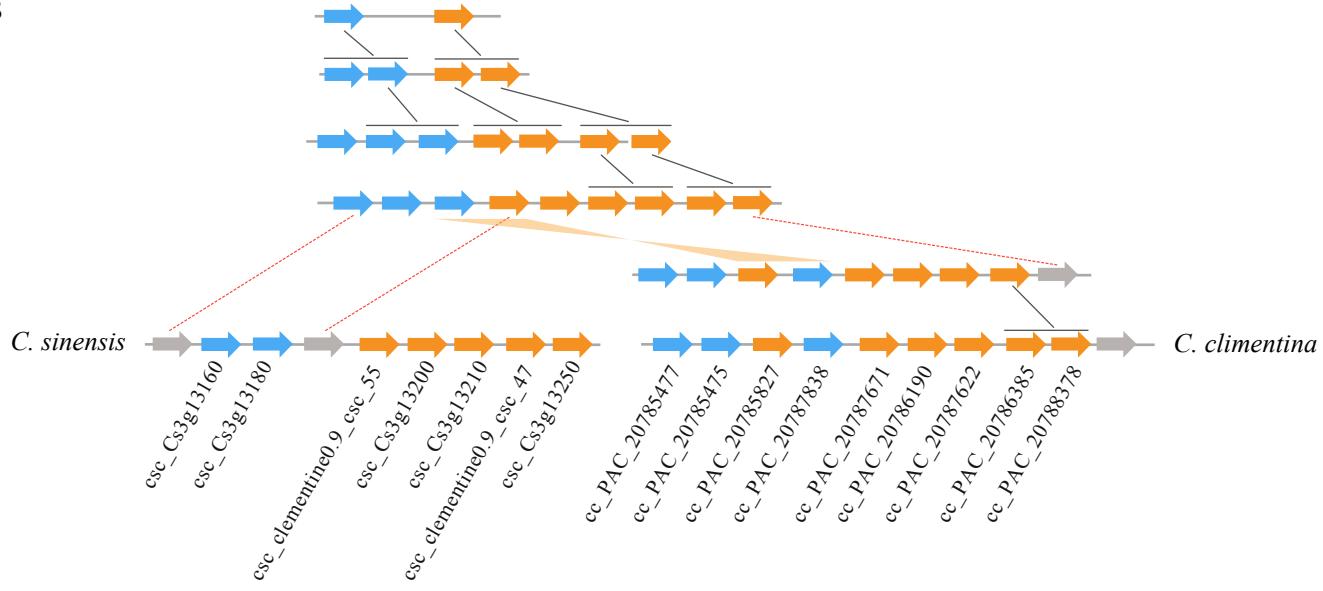


Figure 6

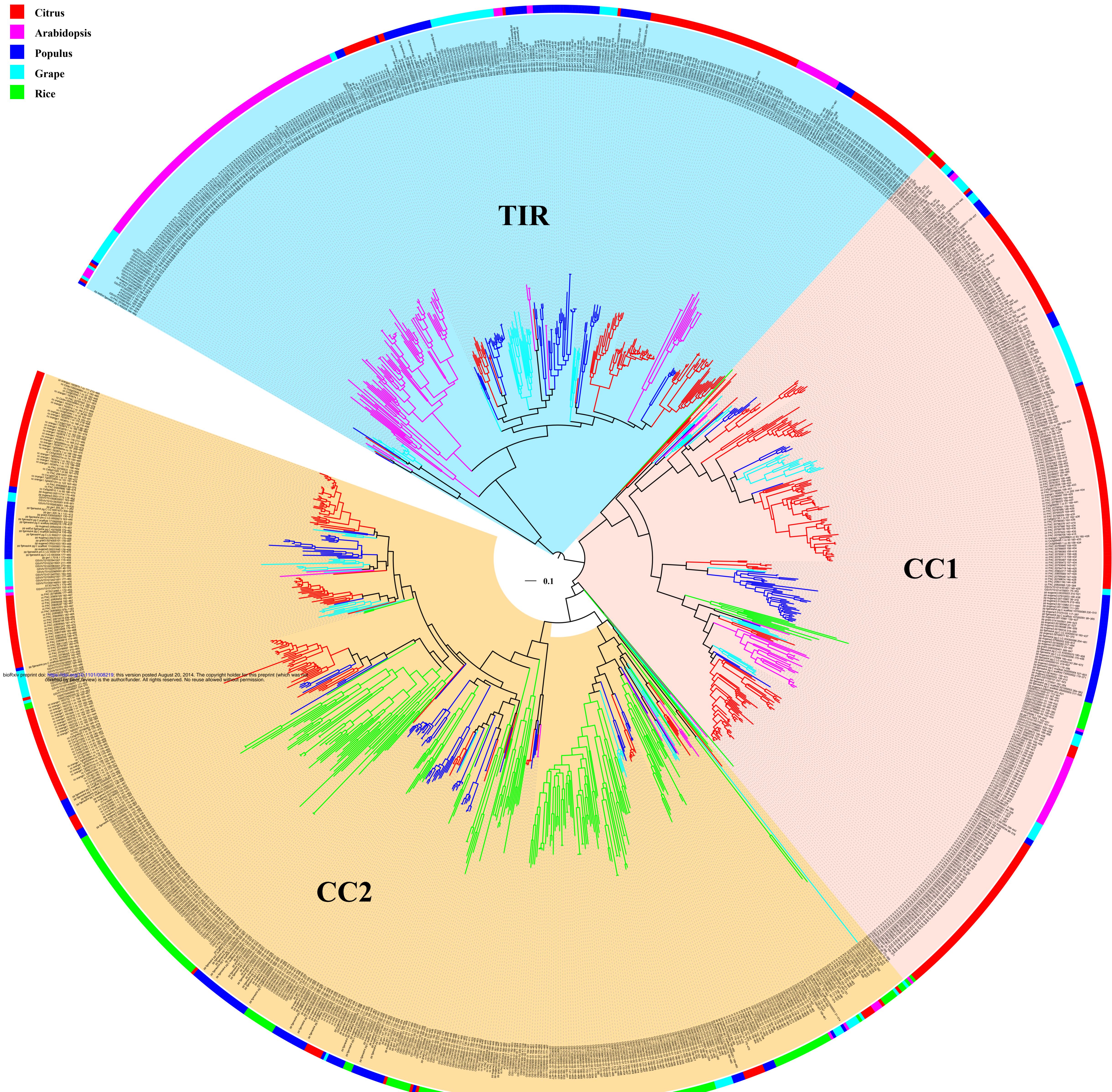


Figure S1

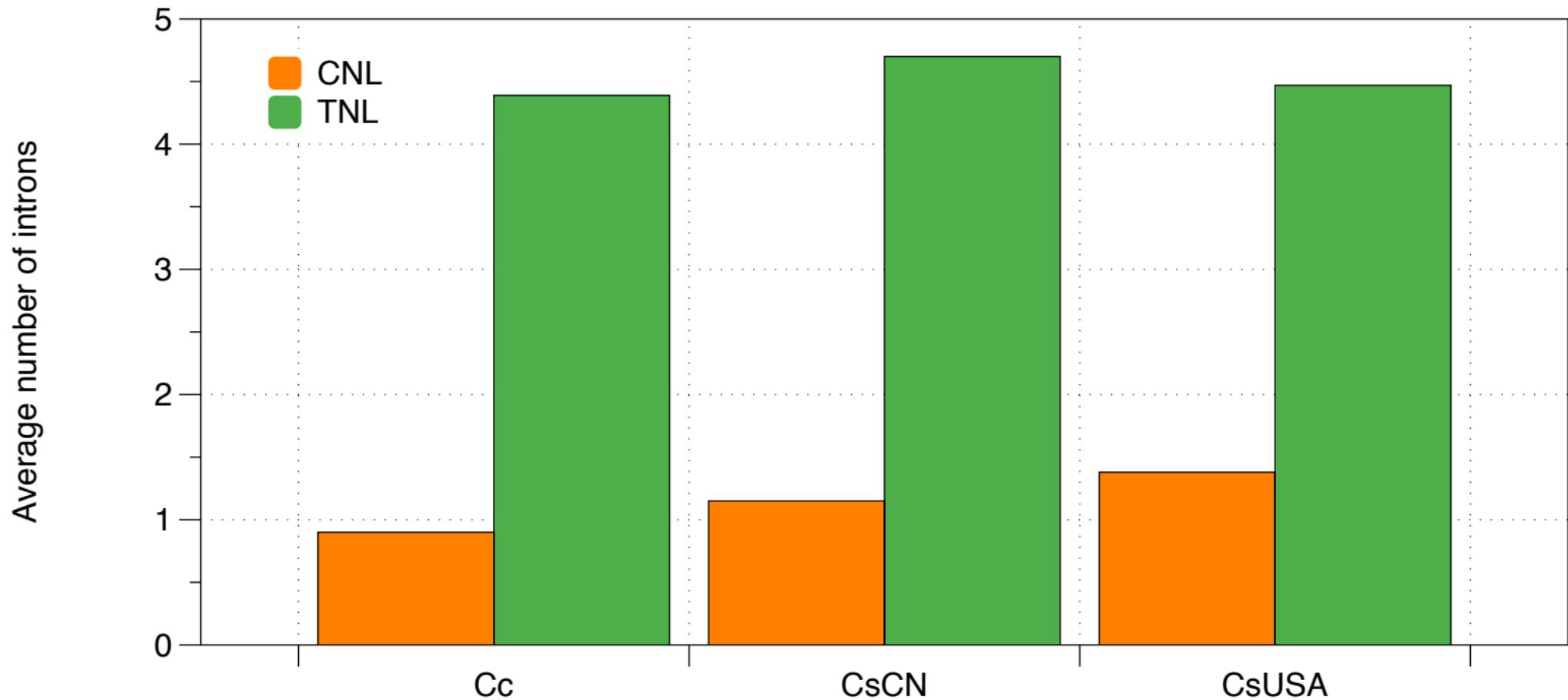
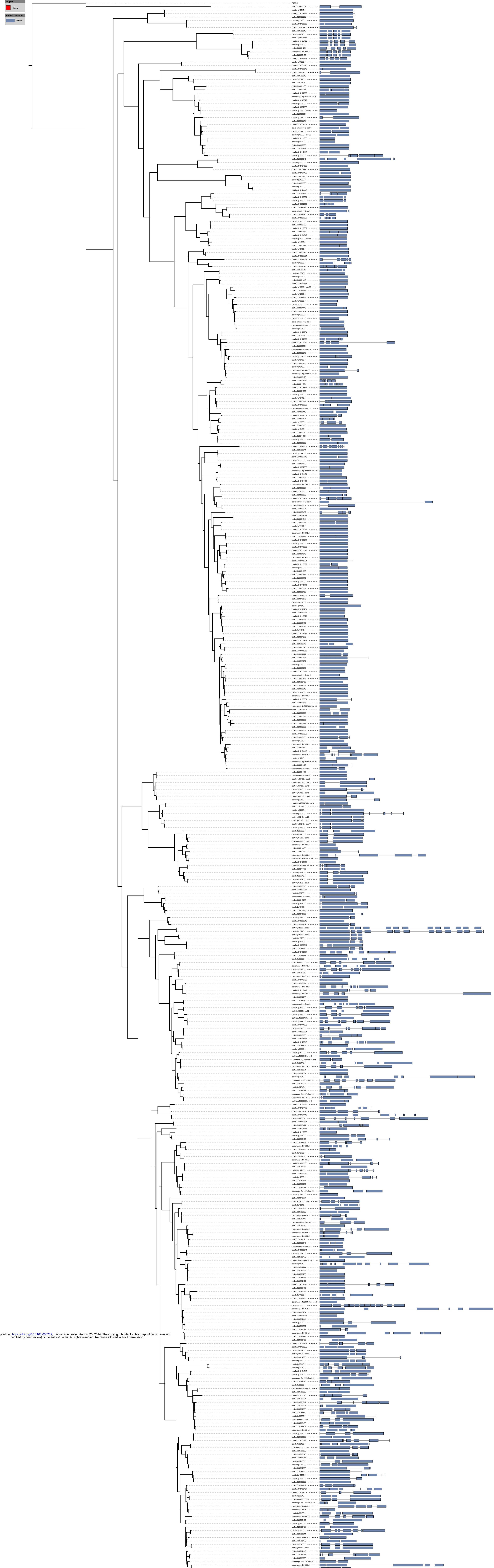


Figure S2-A



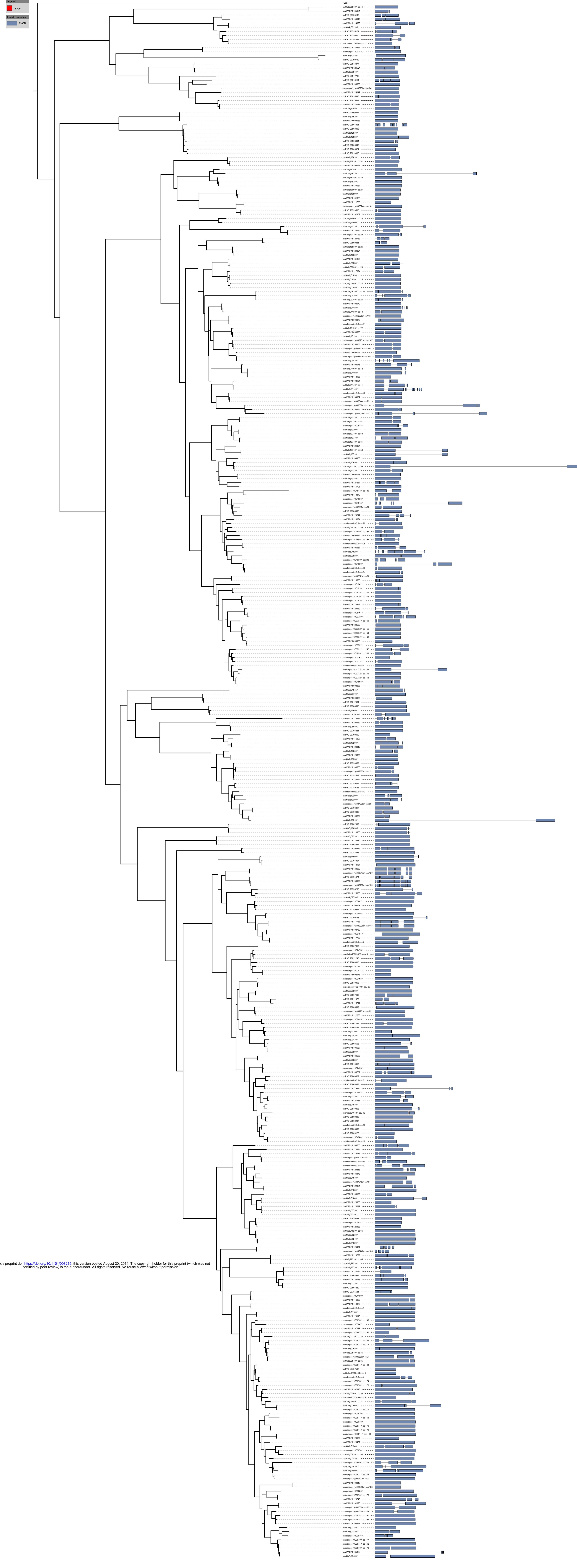




Figure S2-C

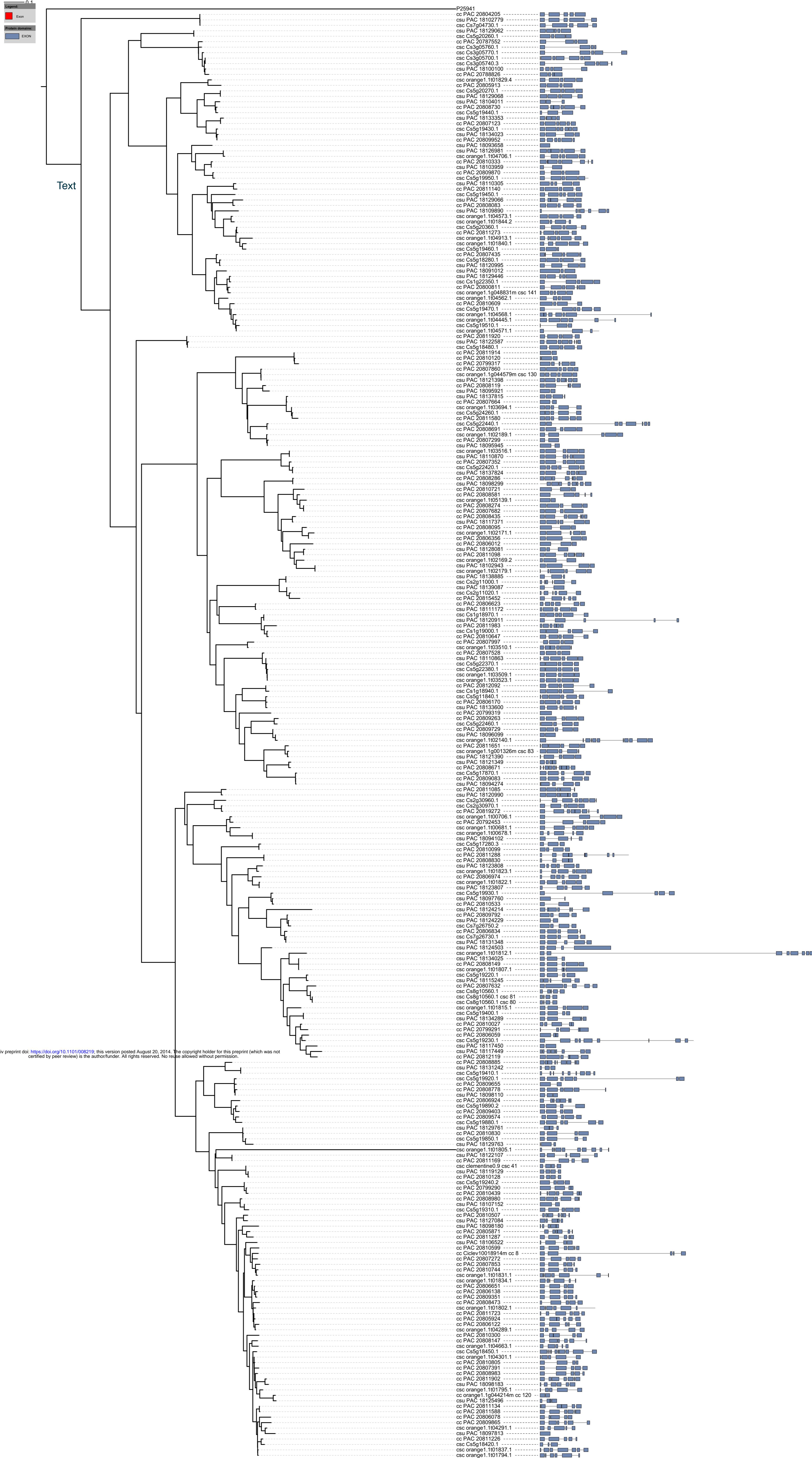
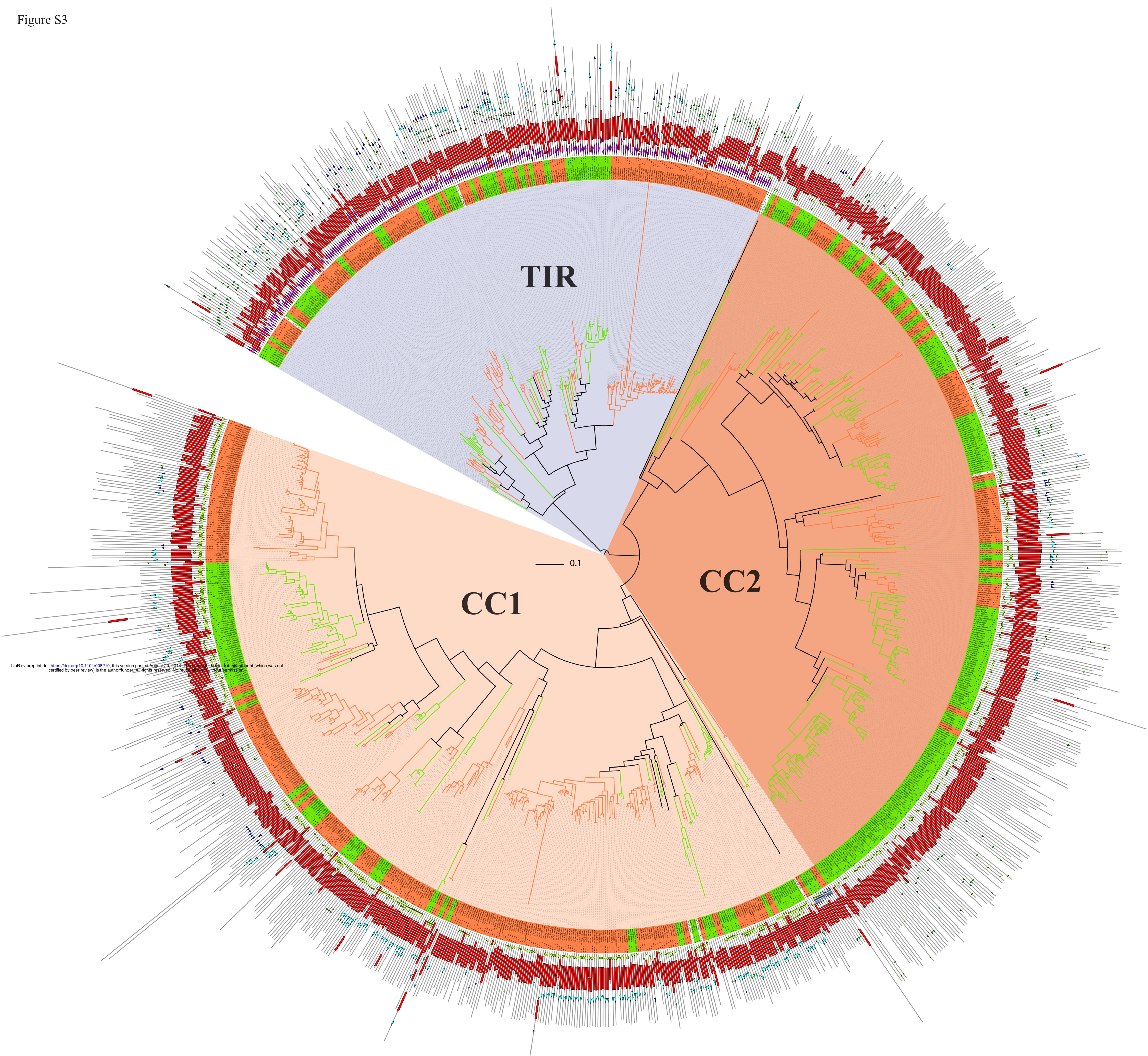


Figure S3



# Figure S4

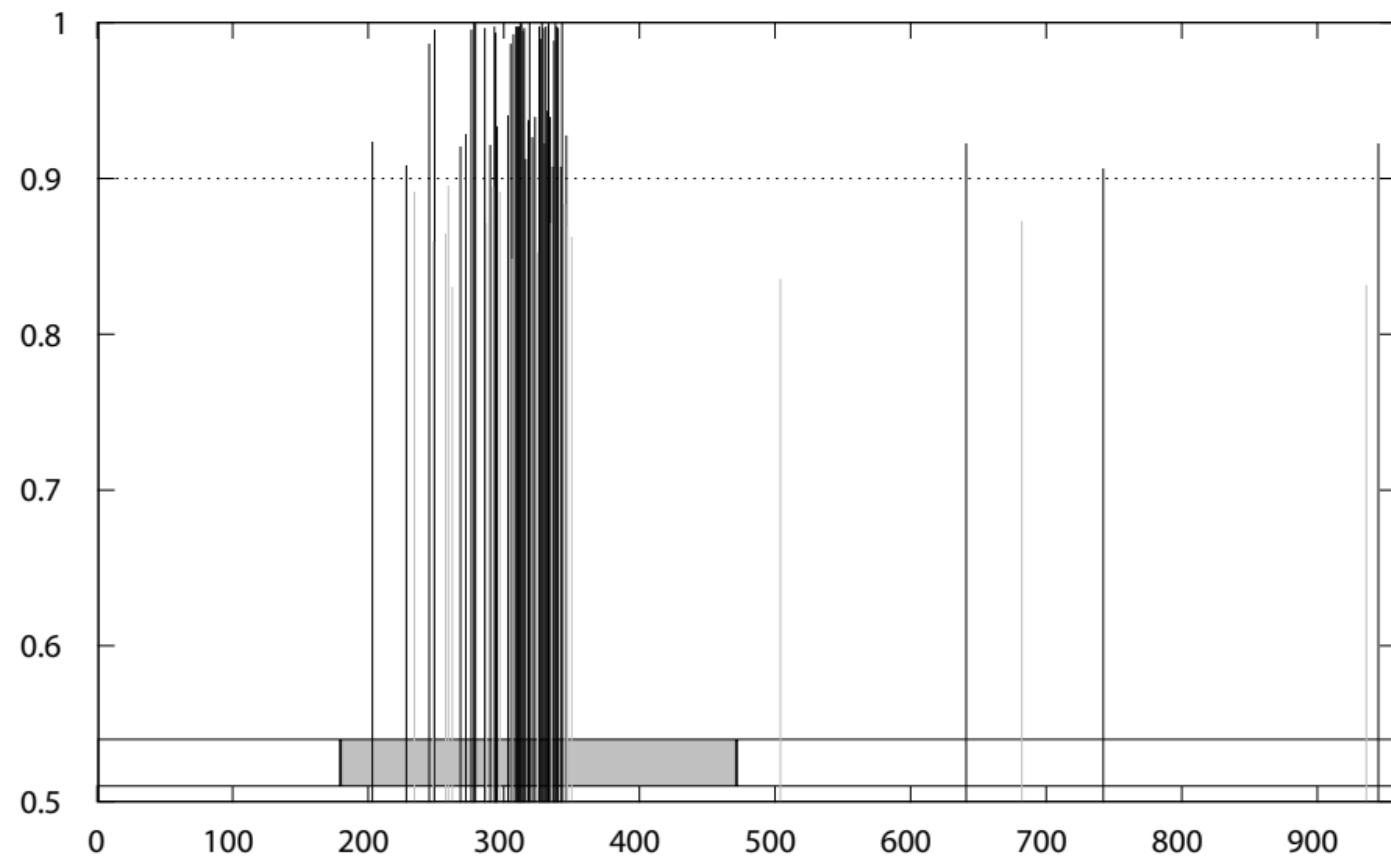


Figure S5

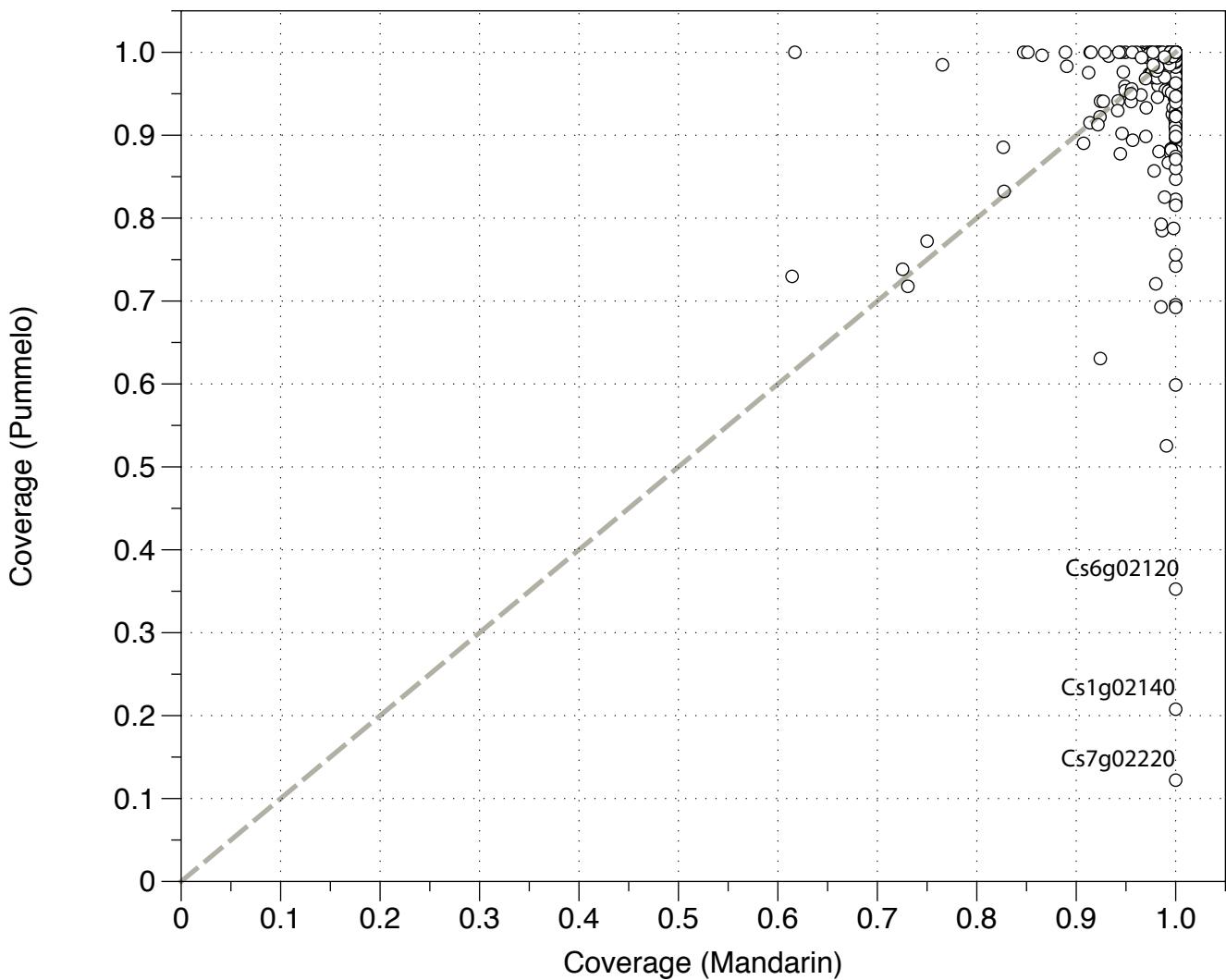


Figure S6

bioRxiv preprint doi: <https://doi.org/10.1101/008219>; this version posted August 20, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A

Citrus_CC1-1	GSQEIDKLCLGGYCSKNCKSSYKFGKQVA-----	29
Arabidopsis_CC1-1	---EIQRLCLCGYCSKNCCSSYRYGKRVFLMLEEVEKLKSQGFF	41
	*** : : *** * ***** * *** : : * : *	
Citrus_CC1-2	EEQQQMRRRLNQVQGWLSRVEA-----	21
Arabidopsis_CC1-2	LAQVQVWLRSRVQTIEENQFNDL	21
	* *** * ***** : :	

B

Citrus_CC2	-----AVKHWLGKLQDAAYDVEDVLDEWQTEAFR	29
Rice_CC2	-----KXWLXELRELAYDAEDCIDEF-----	21
Arabidopsis_CC2	LKTELTCIHCYLKDAEARQREDEVVKHWVAGIRDAAYDAEDILDTYFLKA--	50
Populus_CC2	-----DAEEKQWTNEAVKDWLDDLKDAAAYDADDV-----	29
	* * : : : : * *** . : *	

C

Citrus_TIR1	YDVFLSFRGEDTRDNFTSHLY-----	21
Populus_TIR1	YDVFLSFRGEDTRNNFTDHLYTALCQAGIHTFRDD	35
Arabidopsis_TIR1	-DVFPSFRGEDVRKTFLSHLLKEF-----	23
	*** * ***** . * . * . **	
Citrus_TIR2	-----AIEASAISVIIFSEGYASSRWCLDELVKI---	29
Populus_TIR2	ELPRGEEISPHLWKAIQESRISIIVFSKDYASP-WCLDELVKI---	42
Arabidopsis_TIR2	-----IGPELIQAIRESRIAIVVLSKNYASSSWCLDELVEIMKC	39
	** . * * : : : : * : . *** . * ***** : *	
Citrus_TIR3	-----GQIVIPVFYRVDPSDVRKQTG---	21
Populus_TIR3	CKKXTGQIVLPVFYDVDPSDVRKQTGSFA	29
Arabidopsis_TIR3	---ELGQIVMPIFYGVDPSDVRKQ-----	21
	***** : * : * * *****	
Citrus_TIR4	KENPEKVQKWRDALKEAANLSGF--	23
Populus_TIR4	-----VQRWRDALTEAANLSGWD-	18
Arabidopsis_TIR4	-----WRKALTDVANIAGEHS	16
	** . * * . : . * : : *	

# Figure S7




Figure S8

bioRxiv preprint doi: <https://doi.org/10.1101/008219>; this version posted August 20, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

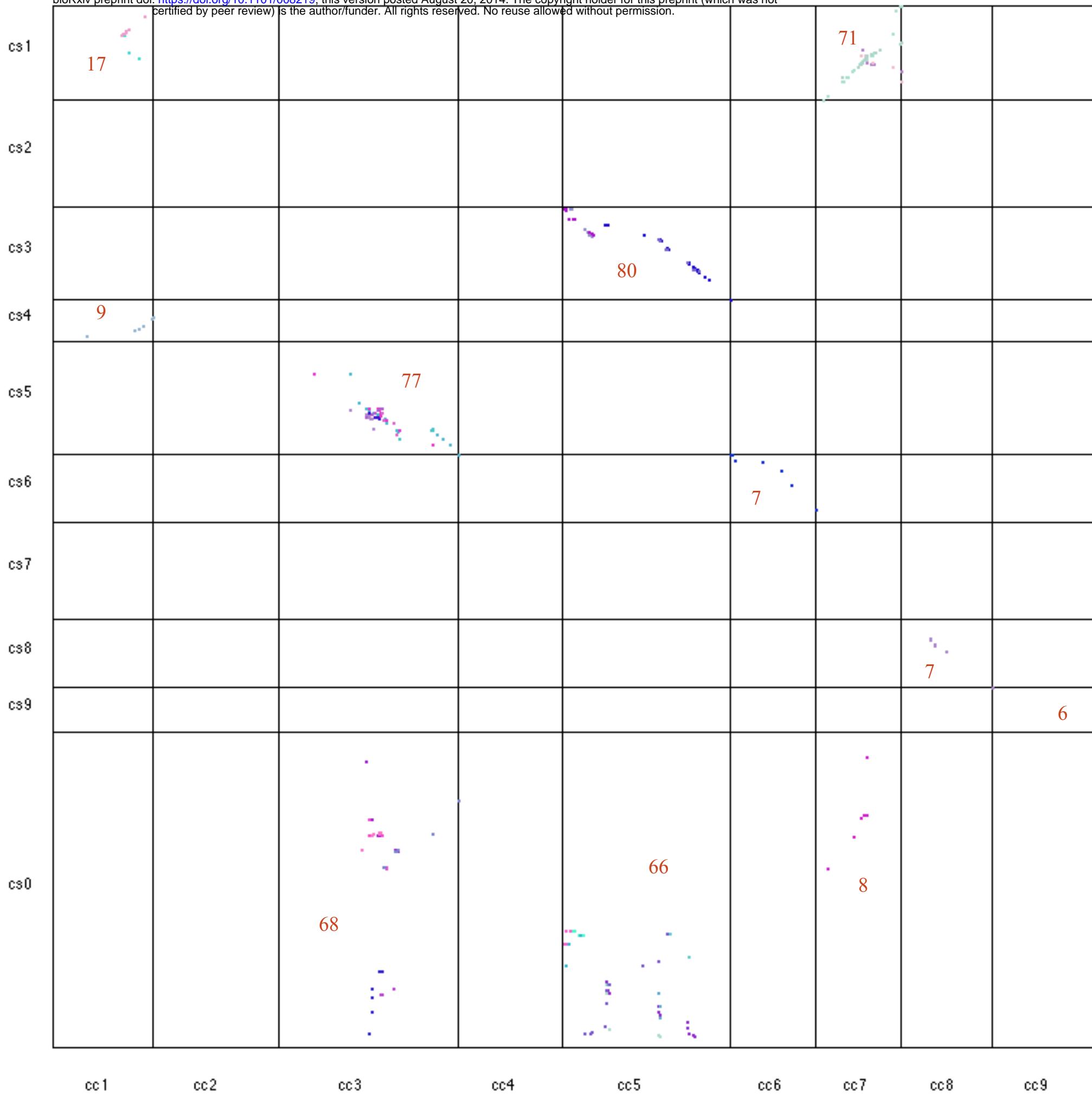


Figure S9

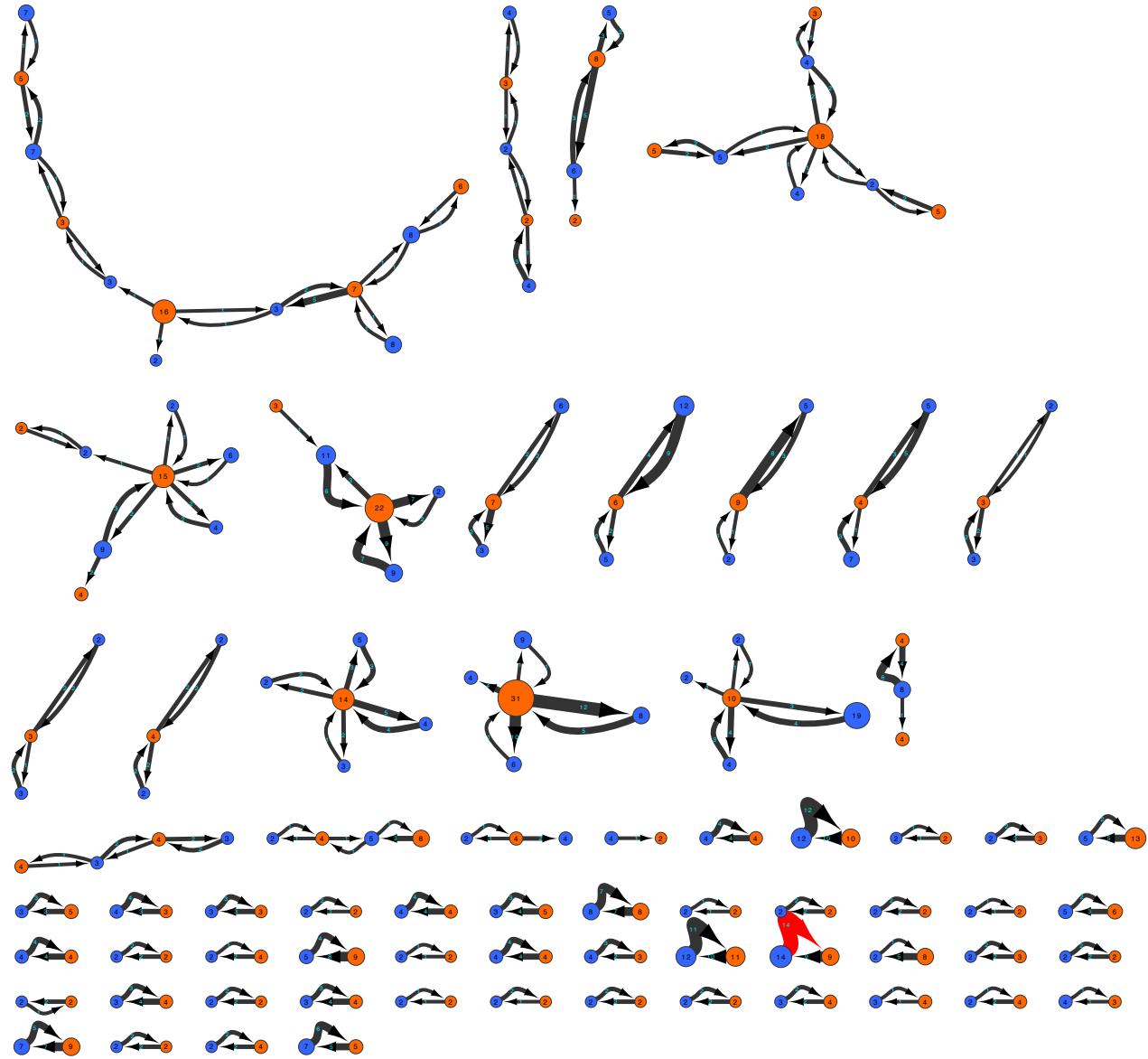
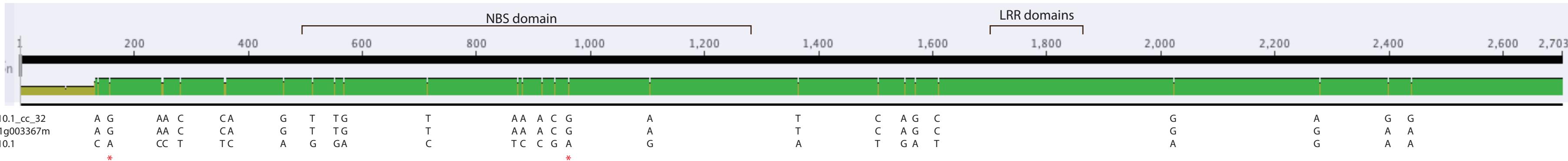
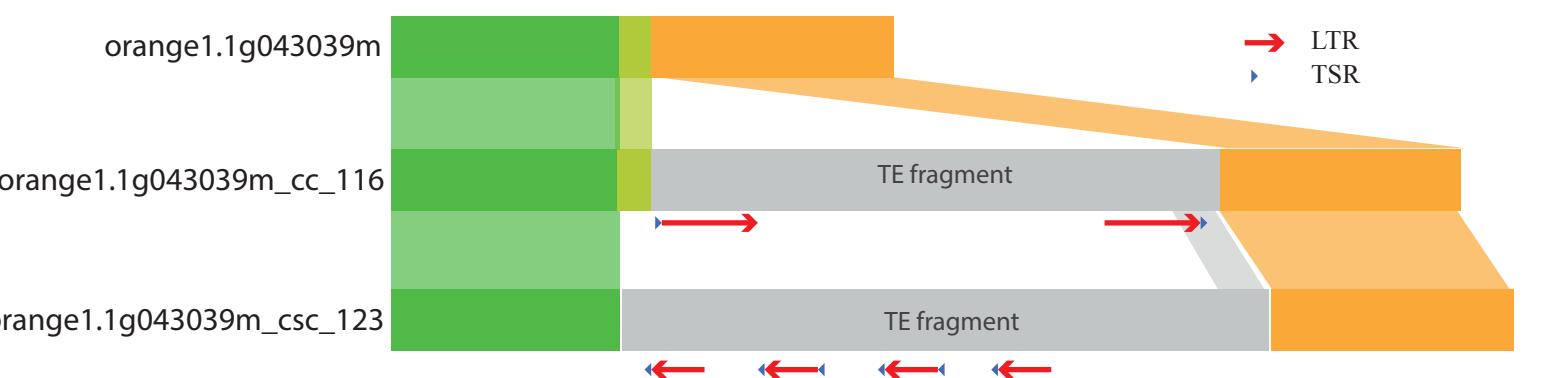


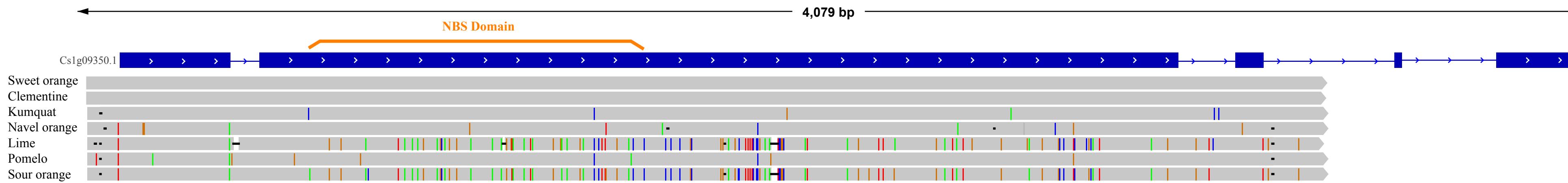
Figure S10

**A****B**

# Figure S11

bioRxiv preprint doi: <https://doi.org/10.1101/008219>; this version posted August 20, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A



B

