

RESEARCH ARTICLE

Open Access



Resequencing of *Capsicum annuum* parental lines (YCM334 and Taeon) for the genetic analysis of bacterial wilt resistance

Yang Jae Kang^{1†}, Yul-Kyun Ahn^{2†}, Ki-Taek Kim³ and Tae-Hwan Jun^{4*}

Abstract

Background: Bacterial wilt (BW) is a widespread plant disease that affects a broad range of dicot and monocot hosts and is particularly harmful for solanaceous plants, such as pepper, tomato, and eggplant. The pathogen responsible for BW is the soil-borne bacterium, *Ralstonia solanacearum*, which can adapt to diverse temperature conditions and is found in climates ranging from tropical to temperate. Resistance to BW has been detected in some pepper plant lines; however, the genomic loci and alleles that mediate this are poorly studied in this species.

Results: We resequenced the pepper cultivars YCM344 and Taeon, which are parental recombinant inbred lines (RIL) that display differential resistance phenotypes against BW, with YCM344 being highly resistant to infection with this pathogen. We identified novel single nucleotide polymorphisms (SNPs) and insertions/deletions (Indels) that are only present in both parental lines, as compared to the reference genome and further determined variations that distinguish these two cultivars from one another. We then identified potentially informative SNPs that were found in genes related to those that have been previously associated with disease resistance, such as the R genes and stress response genes. Moreover, via comparative analysis, we identified SNPs located in genomic regions that have homology to known resistance genes in the tomato genomes.

Conclusions: From our SNP profiling in both parental lines, we could identify SNPs that are potentially responsible for BW resistance, and practically, these may be used as markers for assisted breeding schemes using these populations. We predict that our analyses will be valuable for both better understanding the YCM334/Taeon-derived populations, as well as for enhancing our knowledge of critical SNPs present in the pepper genome.

Keywords: Pepper, Bacterial wilt, Resequencing, SNP, YCM334, Taeon

Background

Bacterial wilt (BW) is a common plant disease that affects a wide array of diverse hosts, ranging from dicots to monocots. It is especially harmful for a number of solanaceous crops, including peppers, tomatoes, and eggplants. BW is caused by the bacterial pathogen, *Ralstonia solanacearum*, which can adapt to diverse temperature conditions and is commonly found in soil from a broad distribution of tropical to temperate climate regions [1]. *R. solanacearum* infects plants through cracks, such as wounds, root tips, and lateral root

emergence sites, and eventually colonizes the root cortex. After invading the xylem vessels transporting water and soluble mineral nutrients from root throughout the plant, the bacterial pathogen can rapidly multiply, filling up and blocking the xylem. Eventually, infection with *R. solanacearum* leads to host wilting and quickly results in plant death. Because of these destructive symptoms, this bacterium is ranked second out of the top 10 pathogens that have importance with regards to economic and scientific consequences [2].

In tomato, the genomic regions that confer resistance against BW have been characterized; the Bwr-12 region is known to confer strong resistance against BW and is specific to phylotype I (Asian) strains. The Bwr-6 region confers weaker resistance than Bwr-12, and this quantitative traits loci (QTL) is specific to both phylotype I

* Correspondence: thjun76@pusan.ac.kr

†Equal contributors

⁴Department of Plant Bioscience, Pusan National University, Miryang, Republic of Korea

Full list of author information is available at the end of the article



and II strains [3, 4]. The Bwr-6 region in particular is also known to mediate broader resistance against other diseases, including root-knot nematodes, potato aphids, *Cladosporium fulvum*, *Oidium lycopersicon*, Tomato yellow leaf curl virus, and Alfalfa mosaic virus [5]. Resistance to BW has also been detected in pepper plants; however, the genomic loci and alleles that mediate resistance responses against BW are poorly understood in this species.

Pepper (*Capsicum annuum*) belongs to Solanaceae family and is one of the most prevalent and economically important crops in the world. The pepper genome has 12 chromosomes and is estimated to be 3.48 Gb [6]. As is the case for other solanaceous species, *R. solanacearum* has been isolated from wilting field-grown pepper in south Florida and has also been observed in Japan [7, 8]. Considering the wide host range and adaptability of *R. solanacearum*, we predict it will be necessary to utilize the collection of the pepper germplasms resistant to BW, in order to breed elite cultivars that can counteract the destructive effects of this disease. However, there have been few efforts to select resistant donor accessions from the pepper germplasm collection, and the biological knowledge required to carry out molecular breeding in this population is limited.

Next generation sequencing (NGS) technologies have significantly advanced genomic studies, enhancing both the amount and accuracy of sequencing data that can be affordably obtained. These techniques have almost completely replaced laborious and time-consuming gel-based genotyping procedures, at least for marker development, and consequently, the majority of beneficial crop species have been sequenced and assembled into draft reference genomes, after which, the genomic resources for a given crop species are often enriched using resequencing strategies [9]. For example, after completion of the pepper (*C. annuum*) reference genome, which covers 87.9 % of the estimated genome size, two pepper cultivars (Perennial and Dempsey) and a wild species of pepper (*C. chinense* PI159236) were resequenced, revealing millions of single nucleotide polymorphisms (SNPs) that may discriminate between cultivars or between species [6]. Moreover, 18 accessions of pepper cultivar and two semi-wild accessions were resequenced to investigate how artificial selection traces present in the pepper genome correlate with pepper breeding history [10]. Parental lines of breeding populations were also resequenced to identify the causal regions conferring resistance against the Potato virus Y [11]. The availability of NGS technology and the *C. annuum* reference sequence provides us with the opportunity to employ this resequencing strategy to address more specific and practical questions in genome-assisted breeding schemes to cope with BW.

Here, we resequenced *C. annuum* YCM344 and Taeon, which are parental recombinant inbred lines (RIL) that

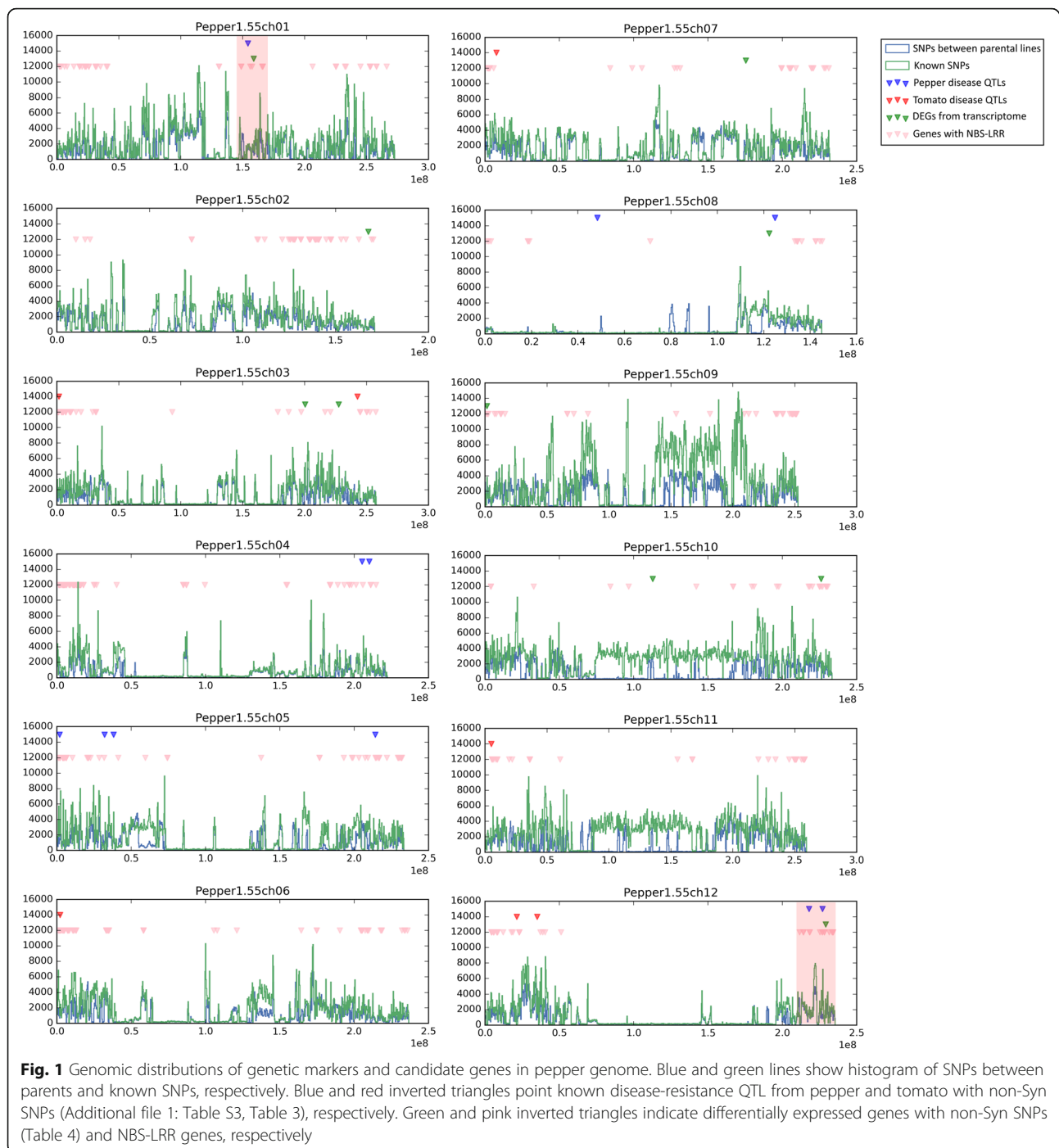
are distinguished by differing resistance against BW. Compared to the previously known SNPs, we identified novel variations existing only in both parental lines, as well as those that distinguish these cultivars from one another. We further annotated informative SNPs by identifying those variants found in genes related to known disease resistance genes, such as the R genes and stress response genes. Moreover, via comparative analysis, we identified SNPs located in genomic regions that are homologous to known BW resistance genes in the tomato genomes. Using these SNP profiling data, we can narrow down the list of informative SNPs to identify those likely to be involved in BW resistance in pepper, and they can be practically used for marker-assisted breeding schemes with these populations.

Results

Whole genome resequencing of parental lines for BW resistance breeding

The parental lines, YCM334 and Taeon, were selected based on their differing resistance to BW disease; YCM334 displays high levels of resistance and Taeon is susceptible. These were resequenced using the Illumina HiSeq2000 platform, producing reads totalling 36.88 Gb and 35.95 Gb, respectively, which provide approximately 10× coverage of the pepper genome (estimated size of 3.48 Gb) (Additional file 1: Table S1) [6]. For the precise call of sequence variations, we trimmed the reads based on quality, using the SolexaQA package (Additional file 1: Table S1) [12]. The processed reads from both YCM334 and Taeon were then successfully mapped to the pepper reference genome sequence (version 1.55) with mapping rates of 93.56 % and 93.55 %, respectively. Using the SAMtools software package [13], we identified genomic variations between the reference genome and each cultivar, including SNPs and insertions/deletions (Indels). A total of 7,002,670 and 6,779,745 SNPs were found, with frequencies of 2.01 SNPs/kb and 1.95 SNPs/kb for YCM334 and Taeon, respectively (Fig. 1 and Table 1). Of these, around 95 % were identified as being homozygous, suggesting our well-developed inbred lines can be used as parental lines for a breeding population, for example, to produce RILs.

When compared to the previous resequencing efforts of Kim et al. [6], which found that the pepper cultivars Perennial and Dempsey contain 10.9 and 11.9 million SNPs, respectively, our resequencing effort revealed an additional 2,748,164 SNPs that are specific for our parental lines (Additional file 1: Table S2). Further, a total 177,148 and 165,875 homologous Indels were identified in YCM334 and Taeon, respectively, as compared to the reference genome. Of these, 683 and 698, respectively, were located within the coding sequence (CDS). The reliability of SNP calling was confirmed using the Sanger sequencing method on gene CA04g03400



(Additional file 1: Figure S1). This result suggests that there were no false positive SNP callings, although two false negatives were found on chromosome 4:10285038 [T/A] and 10285121 [A/G]. Analysis with the Bowtie2 [14] SNP calling pipeline decreased the false negatives, while also adding false positives (Additional file 1: Figure S1). Hence, we selected conservative BWA-based pipeline to have more confident genotypes for further analysis.

Comparison between YCM334 and Taeon

To identify the most informative alleles, in regards to BW resistance, present in YCM334 and Taeon, we compared the genotypes to one another and identified the variations. A total of 5,681,208 SNPs were found that differ between the two cultivars (Fig. 1), and based on the Indel calls from the resequencing results, we found 149,223 polymorphic Indels differing between them as well. We then designed 678,998 high-resolution melting

Table 1 Summary of SNPs from YCM334 and Taean against reference genome

Cultivar	Region	Number of SNPs				Number of Indels			
		Number of total SNP	Homozygous SNP	Heterozygous SNP	Ambiguous SNP ^a	Number of total Indel	Homozygous Indel	Heterozygous Indel	Ambiguous Indel
YCM334	intergenic-region	6,885,942	6,592,983	109,558	183,401	347,901	172,360	25,321	150,220
	genic-region	116,728	106,531	3,735	6,462	11,199	4,788	1,071	5,340
	coding sequence (CDS)	42,786	36,950	2,146	3,690	1,918	683	224	1,011
	Intron	73,942	69,581	1,589	2,772	9,281	4,105	847	4,329
	Total	7,002,670	6,699,514	113,293	194,563	359,100	177,148	26,392	155,560
Taean	intergenic-region	6,659,677	6,344,381	120,733	120,733	329,910	160,997	24,753	144,160
	genic-region	120,068	109,622	3,797	6,649	11,320	4,878	1,014	5,428
	coding sequence (CDS)	44,392	38,238	2,186	3,968	1,880	698	202	980
	Intron	75,676	71,384	1,611	2,681	9,440	4,180	812	4,448
	Total	6,779,745	6,454,003	124,530	127,382	341,230	165,875	25,767	149,588

^aSNPs that have not enough depth coverage to determine whether homo/hetero they are

analysis primers for high-throughput genotyping and identified 12,062 possible Cleaved Amplified Polymorphic Sequences (CAPS) marker sites in 5647 genes, based on the polymorphic information (Additional files 2 and 3). Based on the Indels, an additional six possible CAPS marker sites were identified (Additional file 4). These genetic markers can be applied for high-throughput genotyping on the breeding populations to map segregating traits, such as BW tolerance.

We further analysed the SNPs present within gene regions, which may mediate functional variations. Among the polymorphic SNPs, 106,585 were present within gene regions, and 36,678 of these were in the CDS region. Among the CDS SNPs, 23,396 showed non-synonymous (non-Syn) protein changes in 9102 genes (Additional file 5). We then identified the top 10 genes that are highly polymorphic between two cultivars based on the non-Syn SNPs (Table 2). Interestingly, the most polymorphic gene, with 39 non-Syn SNPs, was CA10g15480, which is annotated as a “Putative disease resistance protein”. This gene was assigned to the “Late blight resistance protein R1” gene family (IPR021929) by Interproscan [15]. Additionally, CA12g20430, a highly polymorphic gene with 29 non-Syn SNP, was characterized as belonging to the “Late blight resistance protein R1” gene family as well, suggesting that polymorphism of this gene family can be important for the different disease responses in two cultivars (Additional file 1: Figure S2). Other genes with a large number of non-Syn SNPs include, polyprotein, LRR like receptor kinase, N-like protein, CC-NBS-LRR, and putative phosphatidylinositol 4-kinase. These were particularly prevalent in the nucleotide-binding site-leucine-rich repeat (NBS-LRR) regions that are well-known in R genes to provide resistance against pathogens [16]. A total of 286

NBS-LRR genes showed non-Syn SNP changes between the two cultivars (Additional file 6).

SNP annotation utilizing homologous pepper and tomato genes involved in pathogen resistance

The numerous high-quality SNPs that were identified using NGS can be utilized to better understand the genomic variation between two cultivars. However, it would also be informative to have annotation for certain SNPs that have possible linkage or overlap to known loci of interest. Therefore, to assign annotation for the SNP, we surveyed the literature for previous knowledge of gene function, particularly disease-related, QTL, or trait-associated markers. For this purpose, we first utilized the well-studied related species, tomato (*Solanum lycopersicum*), which is model plant from the Solanaceae family. We found tomato genes that have been associated with several bacterial, fungal, nematode, and virus diseases, and then compiled a list of the pepper genes that are highly homologous to those genes. Among them, a total of seven genes showed non-Syn changes between YCM334 and Taean, which may result in functional differences between the cultivars (Table 3). These seven genes represent strong candidate loci that in YCM334 are likely to contribute to the resistance phenotype against BW disease.

To further annotate our SNPs, we also took advantage of a previous transcriptome analysis of resistance and susceptible pepper lines, which identified differentially expressed genes (DEGs) in these cultivars using *Arabidopsis* gene chip analysis [17]. The corresponding direct orthologs with the *Arabidopsis* gene ID were regarded as candidate DEGs in pepper gene model, and we identified those with non-Syn SNPs between YCM334 and Taean (Table 4). One of these, beta-galactosidase 4 (CA03g1

Table 2 Top 10 genes that are highly polymorphic between YCM334 and Taaen by non-syn SNPs

Genename	Annotation	Genic SNP	CDS SNP	Non-syn SNP
CA10g15480	Putative disease resistance protein	46	46	39
CA02g09500	Polyprotein	39	39	34
CA06g11250	Detected protein of unknown function	55	44	33
CA04g02500	Detected protein of unknown function	69	44	30
CA12g20430	Putative disease resistance protein	36	34	29
CA05g02730	PREDICTED: probable LRR receptor-like serine/threonine-protein kinase At3g47570-like [<i>Solanum tuberosum</i>]	75	43	26
CA12g02380	N-like protein	163	28	24
CA04g00830	Phosphatidylinositol 4-kinase, putative	195	38	24
CA03g04340	Putative CC-NBS-LRR resistance protein	30	29	24
CA12g02650	PREDICTED: probable LRR receptor-like serine/threonine-protein kinase At4g36180-like [<i>Solanum lycopersicum</i>]	93	31	23

7620) contained 15 SNPs resulting in non-synonymous protein changes, and the CC-NBS-LRR family gene (CA12g19770) contained seven non-synonymous SNPs.

We also surveyed the pepper disease resistance QTLs, including those involved in resistance to *Phytophthora capsici*, *Colletotrichum acutatum*, and *Ralstonia solanacearum* (Additional file 1: Table S3). We found 11 reported genetic markers from the literature and a Korean patent (<http://patent.ndsl.kr/>). We then identified two SNP regions that are proximal to pepper disease QTL, as well as to DEGs, NBS-LRR clusters (Fig. 1). These highly overlapping regions with several annotations would also be candidate regions for mediating BW resistance.

Discussion

The selection of parental lines with specific characteristics is critical for effective crop breeding schemes, which are highly dependent on phenotypic selection after the development of breeding populations. Once the parental lines are determined based on a target phenotype, genotypic features are also informative for developing polymorphic molecular markers that distinguish between target parental lines and can be used to trace down loci responsible for observed genetic variation. The trait mapping resolution increases along with the number of molecular markers that are applied to genotyping; however, this also increases the cost of genotyping. With the development of NGS resequencing technology, we can identify all possible polymorphisms between target parental lines and select highly informative variations based on previous knowledge, such as known gene function and QTL of the corresponding species. We can also take advantage of related model species using comparative genomics approaches [9]. These technological and analytical advances can, in fact, reduce the number of molecular markers required for genotyping and increase the efficiency of marker-assisted breeding schemes, by allowing us to assign

priority on each possible molecular marker. In wheat, for example, selected molecular markers that are tightly linked to phenotype were reliably genotyped in a cost effective and high-throughput manner by a multiplexing amplicon NGS sequencing strategy [18].

In this study, we resequenced the parental lines YCM334 and Taaen that display distinct BW resistance phenotypes. Our data allowed us to develop genetic markers covering the whole pepper genome that are highly informative for quantitative trait loci mapping of BW disease resistance and may be utilized in a breeding scheme to develop a resistant elite cultivar. We identified the genetic variations differing between these two cultivars and further annotated them based on previous functional knowledge, both in pepper, as well as in the related model crop, tomato. We further took advantage of the gene annotation of loci in the NBS-LRR, which are known to have disease-related functions. Although we could not determine which variations from the analysis are clearly responsible to our target trait, the SNPs and Indels identified in this study, as well as their annotation-based priority, will be valuable for genotyping RILs and near isogenic lines originating from a combination of YCM334 and Taaen. Further, the overlap between the variations and our previous knowledge of their likely function also provide evidence that this breeding combination contains allele resources that would show segregation on our target trait. With 169 RILs from a cross between the parent lines, a single factor ANOVA test on quantitative resistance responses of groups classified by the genotype of one selected candidate gene showed significance ($P < 0.05$) (Additional file 7). Furthermore, the parental genotype information would be highly useful for the genotype imputation and curation for ambiguous or missing data, especially from low-coverage resequencing or genotype by sequencing (GBS) data from large numbers of individuals from breeding population, allowing us impute missing alleles in linkage with known

Table 3 Non-syn SNPs in the homologs between pepper and tomato genes where disease related QTLs have been mapped

Disease	Gene list	Donor species	tomato ID	Pubmed ID	Top hit to Pepper (based on blast score)	SNP context
Bacterial speck	Prf	<i>S. pimpinellifolium</i>	Solyc05g013280.2	11952131	CA11g02030	ATGTCAAGGGTTATAGACCC(T/G)CTTGGTATTA CATGTTGTAT CCTCTTGGTATTACATGTTG(T/C)ATCTCTCTGAT GTTGAGAAA TCTCATCCACTCTGGTACAA(A/C)ATTCITTTGGAT TTCTGAAGT GCATTAGGCTATTCAGAGAA(T/A)GTGAAGGGA CGGTGTGTTCT TCAAATACTTAGAATTGGAC(A/G)ACCTCAATAT TTCACAGTGG
Bacterial spot	Bs4	<i>S. pennellii</i>	Solyc05g007850.1	14675431	CA12g06200	TGAAAATTGGTATGTAGGTG(C/A)TAACTTCTTGG GATTTTTCTG TATTTTTCGGAAGAATTGAA(G/C)GAGTTTGGAC TTCGTTTTGTT TGTATAAAGATGAACCAACA(G/A)AACATGATG ATGAAGTCCGT
Alternaria stem canker	Asc-1	<i>S. lycopersicum</i>	Solyc03g114600.2	10781105	CA03g29040	ATGCTAGGCATTGGCTAAGC(G/T)AATGATTTTT GGAGAGAAGG AAGAGTCGGCATGGAAGTTT(G/A)TGACTTTCT ATCTGCTGAG
Leaf mold	Cf-4, Cf-9B (Hcr9-9B)	<i>S. habrochaites</i> , <i>S. pimpinellifolium</i>	Solyc01g006550.2	9413991	CA12g07610	ATCCCATGAACAGCAATCCG(C/T)GCTCCTATTC CACGAAAGAG TGGTGAATCTTCTTCTT(C/T)TTCTTCGTCACG CTCAACTG
Fungal disease	LeEIX1, LeEIX2	<i>S. lycopersicum</i>	Solyc07g008620.1, Solyc07g008630.1	15155877	CA07g01930	AAGGCCTCTTTTGAACCTCAA(C/T)AAGGGCAGCT CTCTCCTTTT TTCTTTTTATCTTCTTCAT(A/C)ACCCCATGTTGA TAAACGAC GATGGGAACCTCCTCCTC(C/A)TCATCCTCATC ATCATCATC AAGAATCCCCAAAATGCGAC(C/G)AAGAAAACCT AGCACCATCGA
Tobacco mosaic virus	Tm-2a, Tm-2	<i>S. peruvianum</i>	Solyc09g018220.1	17246482, 16172136	CA03g00810	CTTAAGGCAACAACAGATTG(C/A)GCCTTTGCAC TTGTTGGATT TGTTCCCAAATATATTCGGG(T/G)TAGTGACTCTT TGATTAATAA AAAAGAACCCAAAATATTCT(C/A)TATGCAATCC GTAATGAAGA TGTTTGGGCAAAATTGGCTT(A/C)TATTAGAAAAG AACCCAAA TATTTGCTCTTTCGTGCGAC(C/A)GGAAAAATAGA GCTTCAGAA AGACGACTATCTATCTTAAG(C/A)TCGACAAGAG TAAGCTTGAC
Nematode (root knot)	Mi1.2	<i>S. peruvianum</i>	Solyc06g008450.2	9707547	CA06g00990	TTAACCAAGTACC GGCTCG(G/A)ATTGAAGTTC AGTGAGGAT

alleles for the majority of genomic regions [19]. Thus, we predict that our analyses will be valuable, not only for the fundamental analysis of YCM334/Taeon-derived populations, but also for enhancing our general knowledge of variation in the pepper genome.

Conclusions

Resequencing of the parental lines, YCM334 and Taeon, has allowed for the identification of genetic markers, such as SNPs and Indels, which distinguish these cultivars, both from the reference genome and from one

Table 4 Non-syn SNPs in the homologs between pepper and tomato genes that are reported at differentially expressed genes (DEG)

A.th. Gene	Description	PLAZA ^a tomato Ortholog	Blast top hit btw. tomato and pepper	SNP context
S response				
AT5G59120	subtilase 4.13	SL09G009750	CA09g00620	TAGACCATACAAGCTAGTGC(C/T)CGGGACGGGTATCCCACTGA
		SL01G091930	CA01g19460	TACTTTCAATCTCTTTTAG(T/C)TGTTGCAGGCAAACCAATAA
AT5G56870	beta-galactosidase 4	SL03G019890	CA03g23820	ATTAGTACCTTGAACCTCT(T/C)GAGTATCTGCGCTCTTGGCT
		SL07G042220	CA02g28670	CCTGTCATTTTTATCATCCA(A/T)ATTGGCAAGAAAGGCAGCAC
		SL09G092160	CA03g17620	ACTCGCCGGCGATGATCGTC(A/C)AGTGTATACTTAACGCCGTT
AT1G53350	Disease resistance protein (CC-NBS-LRR class) family	SL12G096920	CA12g19770	TTTCAGCAGCCCATTTTCATA(T/C)ATATCTTCCCTTGGGACCG
				ATCATAATCATAGCTAGTGA(T/G)TTGAGCTGGCCTCCAGCAG
AT5G20080	FAD/NAD(P)-binding oxidoreductase	SL05G018520	CA10g18310	ATAACTGACCTCTTGTGTTG(G/T)TCCAGTTAATATACTGAG
				TTTAATTGAAGTTTGTGCGG(T/C)CACCTGTTAAGGAATGGAA
AT3G61220	NAD(P)-binding Rossmann-fold superfamily protein	SL01G094220	CA08g06650	GATTGTGAAATACTTTGAG(A/C)TTGCTTTTAGTGCTTAATTG
				CAAAATCGCGCATGCTATCA(A/C)TATCAATCGTTGGACTAACA
AT2G38540	lipid transfer protein 1	SL10G075100	CA10g08470	TGATCCATTTGCCTTTCACA(C/A)TACCTTCATTTGCAACAGAC
		SL07G049280	CA07g10490	ATCCAGGGCAACGGGATCTG(T/C)TCTGCTGGGGCATCAA
				TGACCTTTTCCCATGCTACT(T/A)AAATCCAGGGCAACGGGATC
				TGCAACCAAAGTCCAATATC(T/A)TCCTATATGGTGACCATTAA
				AGTTTGTCTACATTTATCA(A/G)AGCCGTAAGCACCACGATAA
				TGCTTTATCCGTCAGAGAAA(T/G)TTTCCCGTCAACTCTGAGT
				CCTGACAATGTGTCGACATG(C/G)ATCTCCAACACTACGGCATTGG
				AACAGATTAACAAGATGAGG(A/G)ATGACGAGCTCGTTAAGGCA
				GCGTCTGGCTTTAAGTTAT(G/T)ATGATTTACCGTATCAGCTT
				GTGTTTTCTGTACTTGGGCA(G/A)CTTCCGGAGGGTGAAAAGA
				GGGCTGCTGAAGAAATTATA(G/C)CATTGGAAGGTAACCAAGGA
				ATGGTTCAGGTGCAACTAGA(C/G)GAAACAATCGGAAGGATCAA
				CCTGGAGGGTCGAGACAGGC(A/G)CCATGCCTAATCTAGTTCAT
				CCAAGATTAATCCAGAATG(T/A)TATTTACAGTACTCAGTCA
				TAAAATAAATGGGGCTGACG(G/A)CCAATATCGTTCATCACCCA
				ATACGAGCAAGAGCAATATC(T/C)ATATCATTCTTAAGTCGGGT
				ACAATGCAGGAGTTGGTGGA(G/T)TCACTGCAGATGCTGATGCC
R response				
AT2G38540	lipid transfer protein 1	SL10G075100	CA10g08470	AGGGAGAGCAGCAGCTTGG(C/G)CATGTCAATGCCCTTGATTG
		SL07G049280	CA07g10490	AATTACTCCTACAATGTAAT(C/T)TGTAACACTTTTAAGTGTTT
				ATTCCTTAGAGCCCACTCTT(T/C)TCTCAAGGAAGCAAATTTTT

^a[22]

another. The downstream analyses of these variations, focusing on those in gene coding regions, and comparing to previously identified genomic regions responsible for resistance, such as QTL, and functional markers, has

allowed us to generate a list of highly informative genetic markers that can facilitate genetic analysis using high generation populations, such as RIL. Our results are likely to provide a valuable resource, not only for the

study of pepper BW, but also for the other pepper diseases against which YCM334 displays resistance.

Methods

Plant materials

The *C. annuum* YCM334 and Taeon germplasms were provided by the National Institute of Horticultural & Herbal Science, Rural Development Administration, in the Republic of Korea. YCM334 was originally collected from AVRDC (World Vegetable Center) and is a recombinant inbred line derived from a cross between cv. Yolo Wonder and CM334. According to our observation, YCM334 showed resistance against *R. solanacearum* and is also known to have high resistance against *P. capsici* infection, whereas Taeon is susceptible to *R. solanacearum* [20].

Analysis of NGS results

The raw sequences produced from the Illumina HiSeq2000 were processed by the SolexaQA package [12], and low-quality bases with a phred score <20 were removed using DynamicTrim which is part of SolexaQA package. After trimming, read lengths below 25 bp were removed by LengthSort function of the package prior to mapping analysis. The processed reads were then mapped to the reference sequences using BWA software [21] with the following options: maximum number of gap extensions (-e) = 50, seed length (-l) = 30, maximum differences in the seed (-k) = 1, number of threads (-t) = 16, mismatch penalty (-M) = 6, gap open penalty (-O) = 15, and gap extension penalty (-E) = 8. The variations in samples were extracted by SAMtools software with following options: minimum mapping quality for SNPs (-Q) = 30, minimum mapping quality for gaps (-q) = 15, minimum read depth (-d) = 3, maximum read depth (-D) = 89, min Indel score for nearby SNP filtering (-G) = 30, SNP within INT bp around a gap to be filtered (-w) = 15, and window size for filtering dense SNPs (-W) = 15 [13]. For comparison of SNP calling sensitivity, we tested different pipelines for read mapping using the Bowtie2 aligner with default parameters.

Orthologs retrieval between pepper and Arabidopsis genes

To take advantage of a published transcriptome analysis comparing YCM334 and Taeon and performed using the *Arabidopsis* gene chip [17], we attempted to identify pepper genes orthologous to those *Arabidopsis* gene IDs previously identified as DEGs in this study. For this analysis we used the PLAZA 3.0 dicot database [22]. However, because this database does not currently cover the pepper genome, we first retrieved tomato gene IDs directly orthologous to the *Arabidopsis* gene IDs. Top pepper genes closely matching these tomato genes were then identified by BLASTP protein sequence alignment and were regarded as orthologs in the pepper gene model.

Additional files

Additional file 1: Table S1. Summary of raw read quality control. **Table S2.** Summary of SNP identification from current and previous researches. **Table S3.** List of pepper QTLs against various pathogens and the corresponding literatures. **Figure S1.** Comparison between NGS and Sanger sequencing result. Red colored bases are shared SNP calling from both Bowtie2 and BWA pipelines and green colored bases are additional SNPs only from Bowtie2 pipeline. Comparing with Sanger result, BWA pipeline showed no false positives and two false negative (10285038(T/A), 10285121(A/G)) while bowtie2 pipelines showed one false positive (10285103 (T/A)) and one false negative (10285038 (T/A)). **Figure S2.** Interproscan annotation of CA10g15480 and CA12G20430. (DOCX 442 kb)

Additional file 2: HRM primer list based on the polymorphic SNPs between YCM334 and Taeon. (XLSX 63194 kb)

Additional file 3: CAPS primer set and restriction enzyme for detecting polymorphism on genic regions. (XLSX 6953 kb)

Additional file 4: Indel based CAPS primers. (XLSX 15 kb)

Additional file 5: The list of gene-associated SNPs with type, positions and sequence context. (XLSX 11974 kb)

Additional file 6: SNPs exerting the nonsynonymous protein changes on NBS-LRR genes. (XLSX 55 kb)

Additional file 7: CAPS genotyping results in 156 RIL population from a cross of YCM334 and Taeon using selected SNP marker (CA04G03400 SNP1, Additional file 1: Figure S1) and BW resistance phenotype scored from 1 (most resistant) to 5 (most susceptible). (XLSX 13 kb)

Abbreviations

BW: Bacterial wilt; CAPS: Cleaved amplified polymorphic sequences; CDS: Coding sequence; DEG: Differentially expressed genes; GBS: Genotype by sequencing; HRM: Resolution melting analysis; Indel: Insertion and deletion; NBS-LRR: Nucleotide-binding site-leucine-rich repeat; NGS: Next generation sequencing; Non-Syn: Nonsynonymous; QTL: Quantitative traits loci; RDA: Rural development administration; RIL: Recombinant inbred line; SNP: Single nucleotide polymorphism

Acknowledgements

We thank Daewoong Lee and Hyunhee Kim for their technical help in this study. We also thank the employees at the Seeders for their help in Sequencing.

Funding

This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01106802)" Rural Development Administration, Republic of Korea.

Availability of data and materials

All primer information including HRM, CAPS and Indel derived from this study are listed in Additional files 2, 3 and 4. The vcf files of two pepper varieties used in the study have also been deposited into Figshare database (https://figshare.com/articles/Taeahn_YCM334_samtools_raw_vcf/3750561).

Authors' contributions

THJ and YKA conceived and designed the experiments. KTK and YKA developed plant materials. THJ and YKA carried out sequencing analyses and linkage analysis. YJK contributed to the statistical analyses and interpreted the data. All authors participated in writing and approved the final manuscript.

Competing interests

The author(s) declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

¹Plant Systems Biology, School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany. ²Vegetable Research Division, National Institute of Horticultural & Herbal Science, Rural Development Administration, Wanju-gun, Republic of Korea. ³The Foundation of Agricultural Technology Commercialization and Transfer, 441-100 Suwon, Republic of Korea. ⁴Department of Plant Bioscience, Pusan National University, Miryang, Republic of Korea.

Received: 23 August 2016 Accepted: 25 October 2016

Published online: 28 October 2016

References

- Hayward AC. Biology and Epidemiology of Bacterial Wilt Caused by *Pseudomonas-Solanacearum*. *Annu Rev Phytopathol*. 1991;29:65–87.
- Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Verdier V, Beer SV, Machado MA, et al. Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol*. 2012;13(6):614–29.
- Wang J-F, Ho F-I, Truong HTH, Huang S-M, Balatero CH, Dittapongpich V, Hidayati N. Identification of major QTLs associated with stable resistance of tomato cultivar ‘Hawaii 7996’ to *Ralstonia solanacearum*. *Euphytica*. 2013;190(2):241–52.
- Carneille A, Caranta C, Dintinger J, Prior P, Luisetti J, Besse P. Identification of QTLs for *Ralstonia solanacearum* race 3-phyloptype II resistance in tomato. *Theor Appl Genet*. 2006;113(1):110–21.
- Kim B-S, French E, Caldwell D, Harrington EJ, Iyer-Pascuzzi AS. Bacterial wilt disease: Host resistance and pathogen virulence mechanisms. *Physiol Mol Plant Pathol*. 2016;95:37–43.
- Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat Genet*. 2014;46(3):270–8.
- Horita M, Tsuchiya K. Genetic diversity of Japanese strains of *Ralstonia solanacearum*. *Phytopathol*. 2001;91(4):399–407.
- Ji P, Allen C, Sanchez-Perez A, Yao J, Elphinstone JG, Jones JB, Momol MT. New Diversity of *Ralstonia solanacearum* Strains Associated with Vegetable and Ornamental Crops in Florida. *Plant Dis*. 2007;91(2):195–203.
- Kang YJ, Lee T, Lee J, Shim S, Jeong H, Satyawan D, Kim MY, Lee SH. Translational genomics for plant breeding with the genome sequence explosion. *Plant Biotechnol J*. 2016;14(4):1057–69.
- Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y, et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc Natl Acad Sci U S A*. 2014;111(14):5135–40.
- Devran Z, Kahveci E, Ozkaynak E, Studholme DJ, Tor M. Development of molecular markers tightly linked to gene in pepper using next-generation sequencing. *Mol Breed*. 2015;35(4):101.
- Cox MP, Peterson DA, Biggs PJ. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinf*. 2010;11:485.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinf*. 2009;25(16):2078–9.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. *Nucleic Acids Res*. 2005;33:W116–20.
- DeYoung BJ, Innes RW. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat Immunol*. 2006;7(12):1243–9.
- Hwang J, Choi Y, Kang J, Kim S, Cho M, Mihalte L, Park Y. Microarray Analysis of the Transcriptome for Bacterial Wilt Resistance in Pepper (*Capsicum annuum* L.). *Not Bot Horti Agrobo*. 2011;39(2):49–57.
- Bernardo A, Wang S, St Amand P, Bai G. Using Next Generation Sequencing for Multiplexed Trait-Linked Markers in Wheat. *PLoS ONE*. 2015;10(12):e0143890.
- Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 2010;11(7):499–511.
- Lu F-H, Yoon M-Y, Cho Y-I, Chung J-W, Kim K-T, Cho M-C, Cheong S-R, Park Y-J. Transcriptome analysis and SNP/SSR marker information of red pepper variety YCM334 and Taeon. *Sci Hortic-Amsterdam*. 2011;129(1):38–45.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinf*. 2009;25(14):1754–60.
- Proost S, Van Bel M, Vaneechoutte D, Van de Peer Y, Inze D, Mueller-Roeber B, Vandepoele K. PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res*. 2015;43(Database issue):D974–981.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

