Szakmai zárójelentés: OTKA K115403 "Az *Rcg1* gyökérgolyva rezisztencia gén azonosítása szőlőben" című kutatás, 2016-02-01 - 2021-04-30

Project: OTKA K115403 – Identification of crown gall resistance gene *Rcg1* in grapevine

Most agrobacteria are efficient plant pathogens. During their infection, a bacterial DNA segment (T-DNA) is transferred into the genome of plant cells where the T-DNA oncogenes direct an unregulated production of plant hormones. This leads to abnormal cell proliferation and to the development of crown gall or hairy root. By these symptoms agrobacteria may cause serious damage in the plantations of woody plants such as grapevines, apple tree and stone fruits. Cultivated European grapevines (*Vitis vinifera*) are susceptible to this disease but some wild *Vitis* species, including *V. amurensis*, have resistant genotypes. This natural resistance was introgressed to *V. vinifera* where it was inherited as a single and dominant Mendelian trait. We have shown recently that this resistance gene (*Rcg1*) is located on chromosome 15 between 7,87 Mb and 9,31 Mb (Theor Appl Genet 125:1565-1574).

In the first year of the project we have generated new DNA markers to establish a detailed genetic map for the map-based cloning of the *Rcg1* locus and characterized new progeny derived from self pollination of heterozygous plants to find siblings that are homozygous for the *Rcg1* locus.

1) *Rcg1* coupled new DNA markers

Some recent mapping data showed that the *Rcg1* locus is located between the OPQ15 and the VVS16 microsatellite markers that are at 7,47 Mb and 9,57 Mb positions on the sequence of chromosome 15, respectively. According to our mapping data the *Rcg1* locus was 5,5 cM far from both markers, therefore we examined the predicted genes around 8,5 Mb. At the position of 8,70 Mb we found a gene (LOC100255228) encoding a homologue of the EDR-2 (enhanced disease resistance) gene of *Arabidopsis*. The sequence of this gene was established both from the resistant and the susceptible parents, and a SCAR marker (g11-5) was designed for a polymorphism that is coupled to the *Rcg1* locus. Unexpectedly, when we tested the presence of this new locus on the whole mapping population the genetic distance between g11-5 and *Rcg1* was calculated for 6 cM and the *Rcg1* locus was located between markers g11-5 and 9m3-3 (from 8,71Mb to 9,31 Mb of the Pinot Noir sequence) see Fig. 1.

Using the genomic sequence several PCR primer pairs were designed for this chromosomal region to amplify possible microsatellite sequences and other intergenic sequences. Some of these primers that could detect polymorphism between the resistant and susceptible parents were tested on the whole mapping population (272 progeny) and the linkage of the polymorphic region to the resistance gene was verified. Four new DNA markers were developed that are located at 9,02 Mb (9m0-101), 9,27 Mb (9m2-103), 9,31 Mb (Dcc1-Kb), and 9,36 Mb (9m3-6) according to the sequence of chromosome 15 of Pinot Noir (Fig.1).

Our mapping data suggests unusually high recombination frequency in the *Rcg1* region. According to the integrated genetic map of grapevine (Theor Appl Genet 117:499-511) 1 cM genetic distance corresponds to 300 to 400 kb. We measured a similar frequency (260 kb/cM) between VVS16 and VVIV67 microsatellite markers that located outside of but close to the *Rcg1* region on chromosome 15 (at 9,57 Mb and 10,89 Mb, respectively). The genetic distance of VVS16 and 9m3-6 was measured for 6,7 cM although the physical distance according to the Pinot Noir sequence is 212 kb. This corresponds to 30 kb/cM recombination frequency. All of

the mapping data calculated for the new markers show unusually high recombination in this region (Fig.1). Some of our data suggested that this region may contain repetitive sequences that disturb the recombinational mapping.

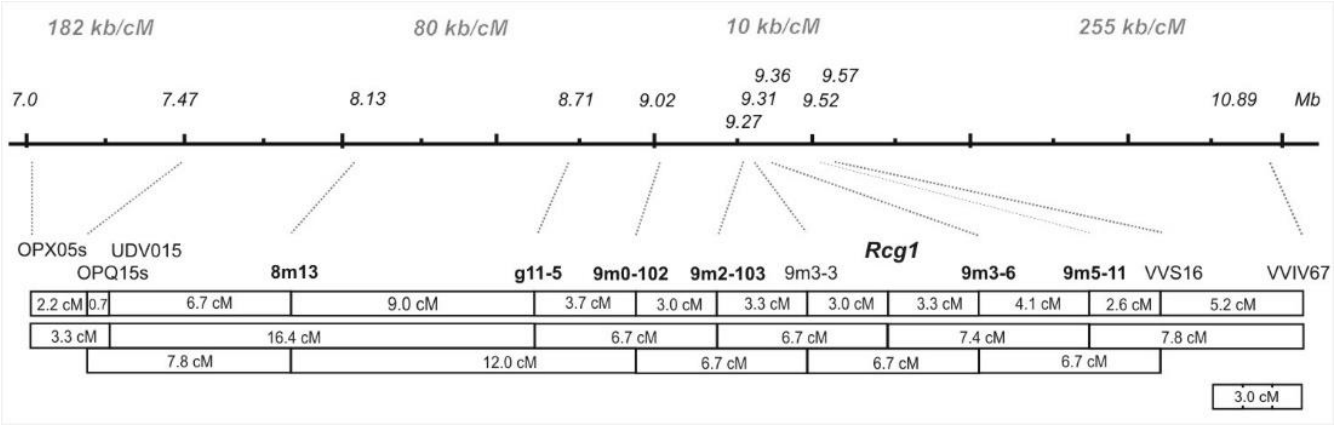

Fig. 1 Comparison of the genetic map of the *Rcg1* locus and the corresponding physical map of Pinot Noir. Solid line represents a segment of chromosome 15 of Pinot Noir (coordinates are in Mb). Below the line DNA markers coupled to the *Rcg1* locus are shown. Dotted lines show their positions in the Pinot Noir sequences. Numbers on the top of the figure show the region specific recombination frequencies (kb/cM) estimated on the basis of the Pinot Noir sequence (physical map).

2) The Rcg1 region carries multiple copies of a WAX2-like gene

When we look through the annotations of this chromosomal region we found seven copies of a WAX2-like (ECERIFERUM) gene between coordinates 9.040.000 and 9.360.000 bps. This gene may be involved in plant cuticular-wax production, as it was shown in *Arabidopsis* (CER1 gene; Plant Cell 7:2115-2127). In contrast to grapevine there is only one copy of CER1 gene in *Arabidopsis*. We carried out bioinformatic analyses (Pairwise and multiple alignments and BLAST searches) to compare the WAX2-like genes of Pinot Noir. We referred these genes as WAX2-1 to WAX2-7. The complete coding sequence of this gene consists of then exons and cover more than 7 kb. WAX2-1, WAX2-2 and WAX-2-3 genes are oriented from right to left while the other four copies are directed from left to right. WAX2-1, WAX2-2, and WAX2-6 are truncated genes lacking some exon sequences while WAX2-3 harbors a retrotransposon like sequences in intron 4. Interestingly, WAX2-5 and WAX2-7 share 93% identity, even their intron sequences are very similar to each other.

When we ran a BLAST analysis on the Pino Noir genome sequence we discovered two additional copies of WAX2 homologues between 7,343 Mb and 7,549 Mb on chromosome 15. Thus there are 9 WAX2-like genes in the genome on the same chromosome. Seven of these WAX2-like genes form a repeated gene cluster close to or around the *Rcg1* resistance locus. This chromosomal structure may enhance recombination frequency resulting in several nonallelic crossing over and chromosomal rearrangements in this region.

All of our results suggested that the resistant parent Kunbarát and its resistant progeny except the recombinants carry the whole chromosome 15 from *V. amurensis*. All of our DNA markers coupled to the resistance gene can be found in *V. amurensis* and they were not detected in the other descendants of the mapping population. Therefore it is very likely that the *Rcg1* region harbors a WAX2-like gene cluster that is structurally different from that of the Pinot Noir. One

of these gene variants may be responsible for the crown gall resistance or within this WAX gene cluster there may be a unique gene that is not present in the cultivated grapevine derivatives.

3) Establishing homozygous resistant and susceptible progeny

The identification of the Rcg1 locus by map-based cloning would be greatly enhanced by plants homozygous either for the dominant (resistant) or for the recessive allele. In order to establish homozygous progeny the heterozygous parent Kunbarát and a heterozygous progeny (No189) from the mapping population were self pollinated. Seeds were germinated and siblings were tested for the presence of resistance coupled (*Rcg1* locus) and some susceptibility coupled (*rcg1* locus) markers. Microsatellite primers 9m2-103, 9m0-102 and OPX05 multiply two DNA fragments with different length in heterozygous samples. Homozygous plants carry only one of them. When only the *Rcg1* coupled fragment was detected the progeny was considered as homozygous. According to the DNA markers out of 20 siblings from the Kunbarát line 4 progeny was supposed to be homozygous for the *Rcg1* locus and 7 were considered as homozygous recessive. From the No189 self pollination 15 progeny were predicted to be homozygous dominant, 6 progeny were homozygous recessive and 19 were heterozygote. All of these progeny were vegetatively propagated and were tested in *Agrobacterium* infection tests. At the site of inoculation crown gall developed in all of the progeny that were predicted to carry only the *rcg1* allele while in all other lines crown galls were not formed. These plants will promote the identification of additional resistance coupled DNA markers and the isolation of the resistance gene from a BAC library prepared from a homozygous resistant plant DNA.

4) Identification of WAX2 homologues coupled to the *Rcg1* resistance locus

At the beginning of the work we had very limited amount of plant material from the homozygous lines therefore first we identified DNA sequences of the region by PCR. We carried out pairwise and multiple alignments to compare the WAX2-like sequences of Pinot Noir and designed several primer pairs for conserved regions. Using these primers we successfully cloned several different WAX2-like sequences from a homozygous resistant progeny. We have determined the DNA sequences of the cloned fragments and designed individual primers to isolate longer portions of the unique WAX-2 homologues by inverse PCR. The DNA sequences of the new clones were also determined and the subsequences were assembled to get longer sequence of the individual WAX2 copies. As a result of several PCR and inverse PCR cloning and sequencing experiments we identified at least 11 full length or partial WAX2-like genes from the homozygous resistant DNA. The complete coding sequence of a WAX2-like gene consists of then exons and cover about 7 kb. The majority of the newly identified sequences showed high similarity to the WAX2-3 gene of Pinot Noir. Using multiple alignments we compared these sequences and designed primers that may multiple individual loci coupled to the *Rcg1* resistance. In this way six DNA (SCAR) markers were developed and were tested on the mapping population (272 progeny). All of these markers showed high linkage to the resistance locus but the linkage was never 100%, therefore these new markers are still not optimal for the map based cloning of the resistance gene.

5) Preparation and screening of a genomic library

Preparation of high molecular weight nuclear DNA must be the most difficult steps in making a plant BAC library. We tried to prepare plant DNA from homozygous resistant plants by three different methods but the quality of these DNA preparations were not satisfactory for the requirements of a BAC library. Unfortunately, after isolation of cell nuclei the DNA were

slightly degraded. In addition, we have a limited amount of plant material, since we have only a few plants from each progeny, after vegetative propagation. The third reason we gave up the preparation of BAC library is the repetitive nature of the DNA region around the resistance locus. More repeated genes in a longer insert may results in a less stable genomic clone which makes the isolation more difficult. Therefore we decided to prepare a cosmid genomic library instead of a BAC library.

The CopyRight v2.0 pSMART FOS vector (Lucigen Corporation, Middleton, USA) has several features that make it an ideal vector for cosmid library. It has a single-copy replication origin and an inducible medium-copy replication origin, in addition transcriptional terminators border the cloning site to stabilize recombinant clones. The size of the vector is 7,5 kb and it can carry a 35-45 kb large insert DNA after in vitro packaging. For the library we have prepared nuclear DNA and shared it mechanically. After an end repair reaction the 40-50 kb fraction was isolated on a pulsed field apparatus (CHEF-DR II). After ligation and *in vitro* packaging we have get over 200.000 phage particles representing the library. The *V. vinifera* Pinot Noir genome size is 486,2 Mbs. Calculating with 40 kb average insert size 64,398 clones would represent the whole genome at P=99,5% probability. To store a library with this size as single clones would need 700 microtiter plates with 96-wells. In addition, to screen this huge amount of clones is almost impossible with our technical background.

We carried out a pilot experiment to test whether one positive clone from 1.000 or 10.000 clones can be detected by specific PCR reactions using bacterial suspensions in freezing medium. The test was successful therefore we decided to store the library by 20 clones/well in a 96-well microtiter plates. To store the library after transfection chloramphenicol resistant colonies were pick up and grow in small petri dishes 20 by 20. The content of each plate were suspended in freezing medium and were transferred into the wells of two parallel microtiter plates. In addition we collected the suspensions of 12 wells (one row in a microtiter plate) in Eppendorf tubes for the first screen. In this way in eight PCR reaction we can determine whether a row of a plate contains a positive clone or not. Once a positive row is identified additional 12 PCR reactions are necessary to identify the positive well in this row. Finally, the positive clone can be identified in colony PCR experiments after the dilution of the bacteria of the well into single colonies. Altogether 76,800 clones were picked up and stored in two series of 40 microtiter plates. (One copy is stored at -80 ^{o}C for longer storage and one is kept at -20 ^{o}C for the daily screens.)

Altogether 24 region specific clones have been identified and sequenced in three consecutive screening experiments (see Fig. 2). Unfortunately, the isolation of some of the positive clones were failed in these screens. We lost these clones in the course of isolation as a single colony. It is probably, that they replicate poorly in *E. coli* and diluted out under propagation. This makes the assembly of the sequence of the whole *Rcg1* region difficult, therefore we decided to use nanopore based sequencing to promote the sequence determination of the region. This was not involved in the original work plan (see chapter 7).

6) Preliminary studies for gene expression experiments

Expression studies with the resistance gene will be important to provide evidence for its role in crown gall resistance. For this purpose semi-quantitative PCR and real time qPCR will be used. To these experiments as an internal control constitutively expressed housekeeping genes are used. We designed new sets of primers for the *V. vinifera* phosphoenolpyruvate carboxylase, actin, tubulin and elongation factor 1-α genes which encompass one to three introns. Using

these primers the expected sequences were amplified from genomic DNAs of 24 grapevine cultivars. PCR products amplified from the contaminating genomic DNA of RNA samples and from cDNAs can be clearly distinguished since cDNA-derived products are smaller in size compared to those amplified from genomic DNA.

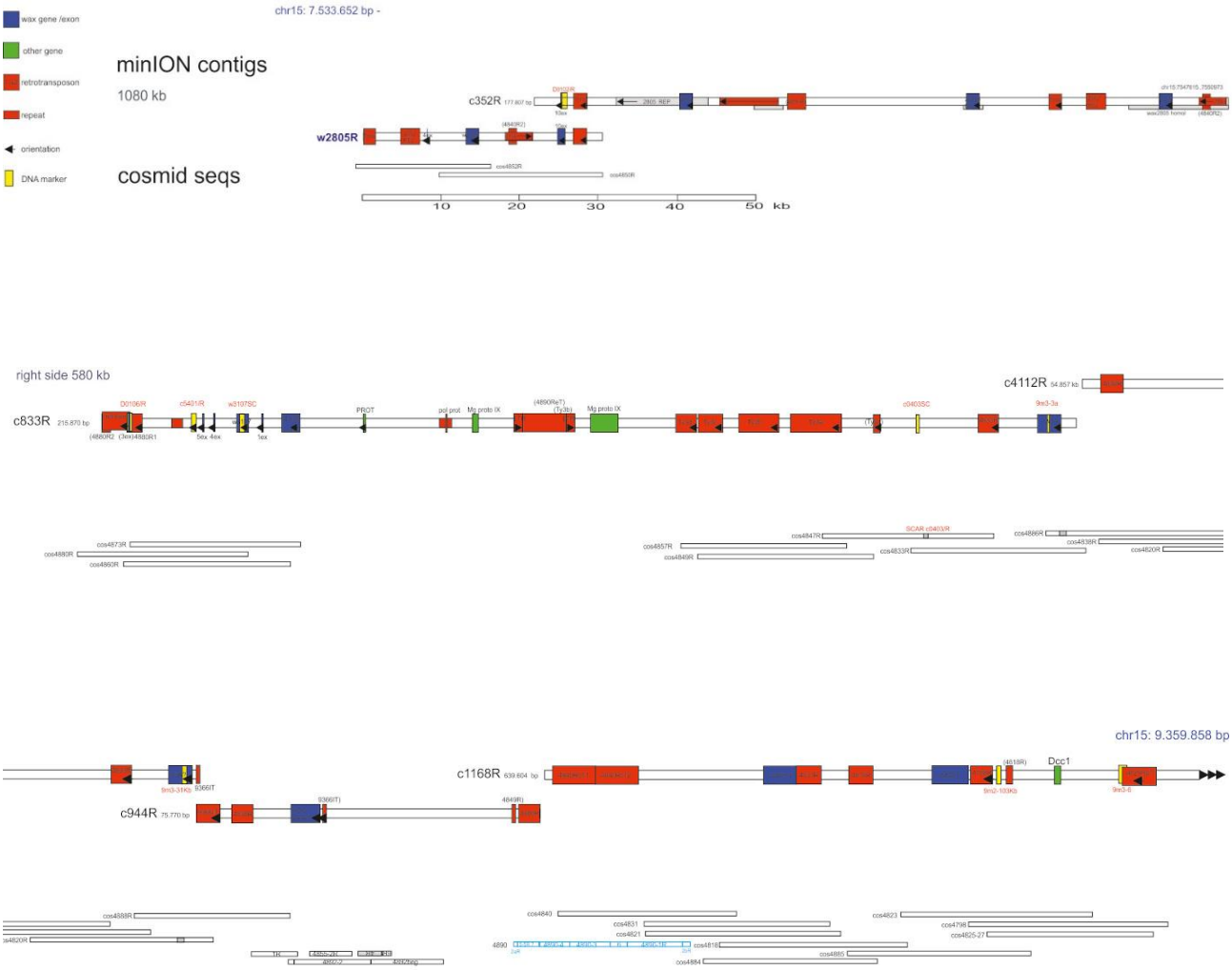

Fig. 2 Physical and genetic map of the *Rcg1* region
Summary of cosmid clones and MiniION contigs in the region. Lines with color boxes represent the MinION contigs: blue boxes shows wax2-like gene exons, green boxes: other genes; red boxes: retrotransposon and repeat sequences; yellow boxes: DNA markers identified and genetically mapped in the region. The right side of the sequenced region is 580 kb. White boxes represent the cosmid sequences identified in the region.

As shown by conventional semi-quantitative PCR these genes were uniformly expressed in cambial tissues of dormant canes, leaf laminas, petioles and in vitro leaves of 12 grapevine cultivars. Thus these primers may be useful reference markers in PCR-based gene expression assays for the validation of cDNA synthesis (Eur J Plant Pathol 149:765–770).

7) Nanopore based sequencing of a homogenous resistant genome

Nanopore based sequencing (Oxford Nanopore, MinION device) is a simple, portable and effective approach developed recently and allows the direct sequencing of an isolated DNA population without cloning and amplification. In this way probably we can determine the sequence of those regions that replicate poorly in *E. coli*. Additional advantage of this technique compared with other sequencing technology is that the read length is not limited, it is depending on the quality of the DNA isolates. This is also very important feature when a long repeated region is the subject of the sequencing. On the other hand a serious disadvantage of this method is in its accuracy, the error rate is around or more than 12%. This makes the assembly and gene prediction more difficult.

All nanopore sequencing experiments were carried out with the same DNA isolate that were used for cosmid library preparation. The first two sequencing experiments were carried out without any fragmentation to reach longer read length. Library preparation was performed by the Ligation Sequencing Kit and the constructed libraries were loaded into Flow Cell R9.4.1. Both sequencing were run 48 hours with standard basecalling. Altogether 344,920 reads with N50=20,957 bp were generated by the MinION system, that resulted in a 8x coverage based on the *V. vinifera* Pinot Noir reference genome (427,171 Mb).

The resulted MinION database was screened first for WAX2 like sequences by BLASTN search. The position and orientation of the WAX2-like sequences as well as retrotransposon sequences identified on the cosmid clones were established on the selected reads. In this way we characterized 48 MinION long reads. Some of them overlap with other MinION read and/or belong to one of the four contigs assembled from the cosmid sequences. (MinION reads contain many errors, therefore comparison of the “gene patterns” seems to be a useful way to find out overlapping reads.) Out of the selected 48 reads 18 were 40 to 60 kb long. Sequence data derived from the cosmid and MinION sequencing until now are not enough to assemble the Rcg1 region into one contig. According to our sequence data it is very probably that the length of this region is over 1,000 kb that is much longer than the homologous region in the Pinot Noir genome (Fig. 2).

In the second two MinION sequencing DNA fragmentation was applied using G-tube (Covaris), where the targeted read length was 12 kb. These two runs resulted in an 8.34 Gb output in 622.670 reads with N50 = 13.420 bp. With the new sequencing runs, the estimated coverage, based on the *V. vinifera* reference genome, increased to 17x. From the sequencing results, high quality reads were separated and assembled with the Flye assembler program. Five contigs were generated, that harbour WAX2-like sequences and overlap with the contigs derived from the cosmid sequences (Fig. 2). According to our present data the *Rcg1* region exceeds the 1000 kb and consists of two contigs. The gap between the two contigs is probably several hundreds of kilobases. Our mapping data suggests that the *Rcg1* locus is situated in this uncovered region.

In order to complete the sequence in the *Rcg1* region two additional sequencing runs were performed, where short reads were eliminated using of Short Read Eliminator Kit (Circulomics). These new runs gave us 7.12 Gb output in 312.465 reads with N50 = 35.870 bp. Unfortunately, the evaluation of the sequence data in the last period failed due to the COVID-19 pandemic situation. Personnel and computer capacities were concentrated for the bioinformatic analysis of the coronavirus sequences.