GENE MAPPING PROJECT TUTORIAL

This program is written on behalf of Heliks R&D and Biotechnology Inc. using the Python software language as an alternative to gene mapping programs that are commercially available.

In **Step 1:Upload Data** tab, fsa file of the sample, bin file of size ranges in which the alleles are expected to be located, the name of the panel and size ladder where the sample is run, project name and threshold value for size ladder are entered and the "**Submit**" button is pressed. Threshold value for size ladder is by default 250, but user can increase or decrease the value. This threshold value describes the minimum peak height for size ladder to be considered and peak heights which are smaller than the threshold value will be deleted from the size matching graph.

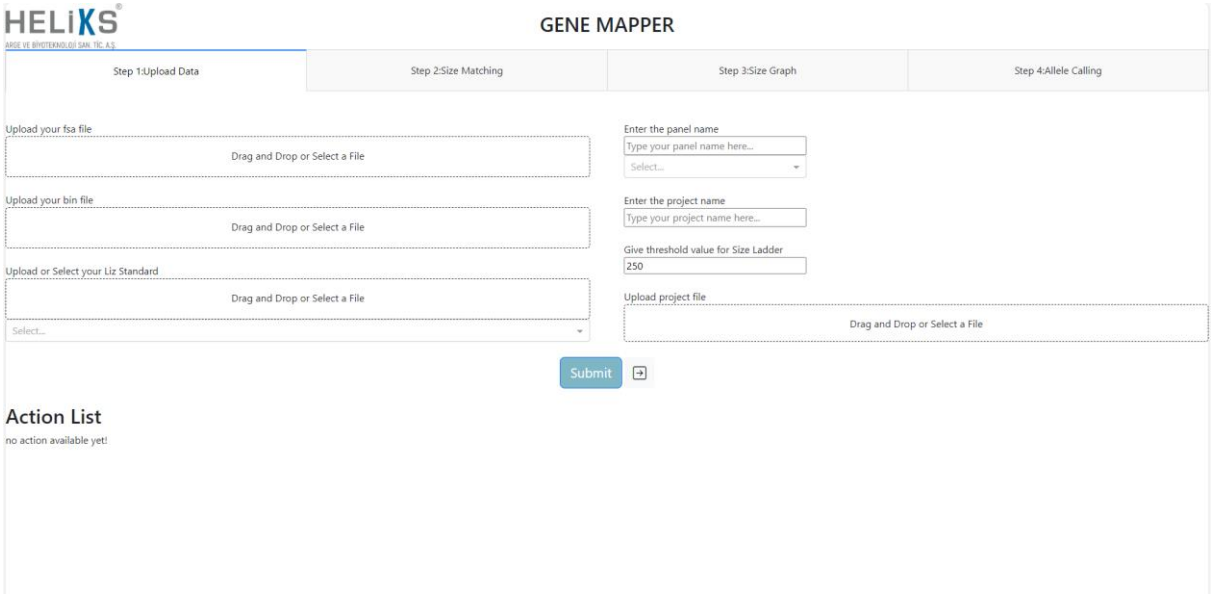


Figure 1. First page of the bioinformatics tool designed for gene mapping.

When the **Submit** button is clicked, the data and information provided are listed in the lower left corner. User can change the information entered at any time. It is enough to click the **Submit** button to update the information.

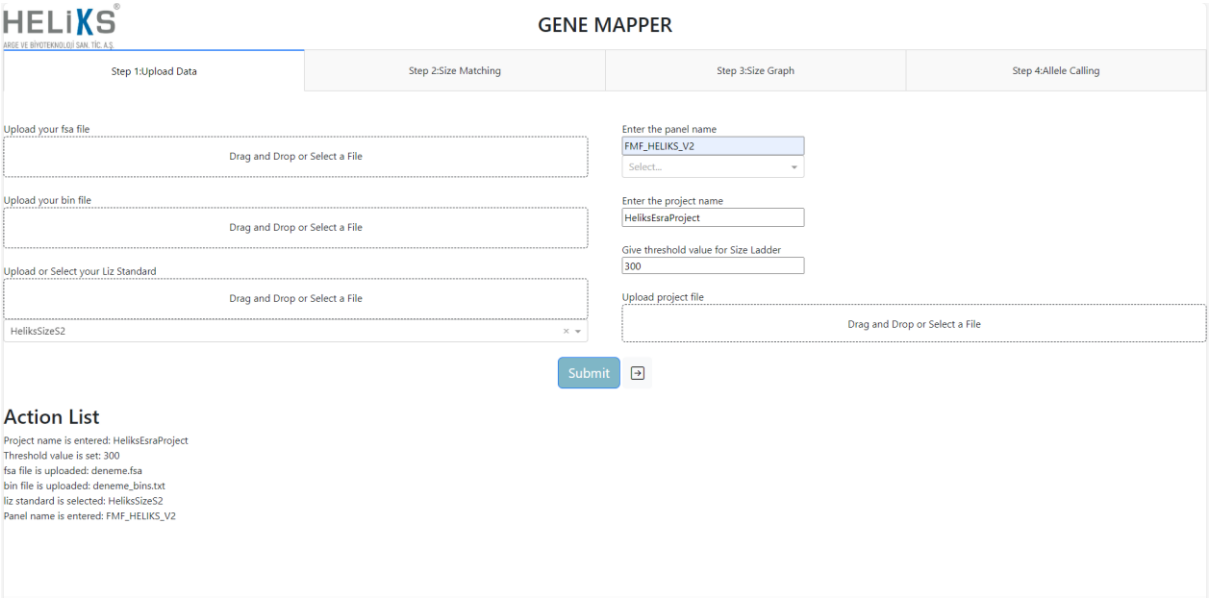


Figure 2. List of the supplied informations are seen on the page after Submit button is pressed.

When the arrow button next to the “Submit” button is pressed, the tool directs the user to **Step2:Size Matching** tab (see Figure 3). Size matching is performed using the size ladder information and fsa file of sample given in Step 1. In this step, we expect to see the same amount of size peaks of sample with sizes designed in the size ladder. But sometimes unfavorable peaks can be seen due to the contamination, or if the expected peaks are weakly reflected, they may not be visible due to the size ladder threshold determined in the first step. For this reason, program applies an estimated matching by taking the distances between the previously designed size ladder into account and looking for the same pattern in the size peaks and peak distances it obtains. If the program evaluates peak as contamination, it marks peak with "**no match**" (Figure 3). Also, the tolerance value used for the prediction is indicated on the top left of the size matching chart. How much the tolerance value is smaller than 1, the less reliable the estimate is. If the program is unable to provide an estimate with the current tolerance value, it lowers the tolerance value and try to make an estimate again with a lower tolerance value. In this loop, the smallest possible tolerance value is 0.1.

If user thinks that the estimated size matching is not correct, he/she can also make the match manually through the table.

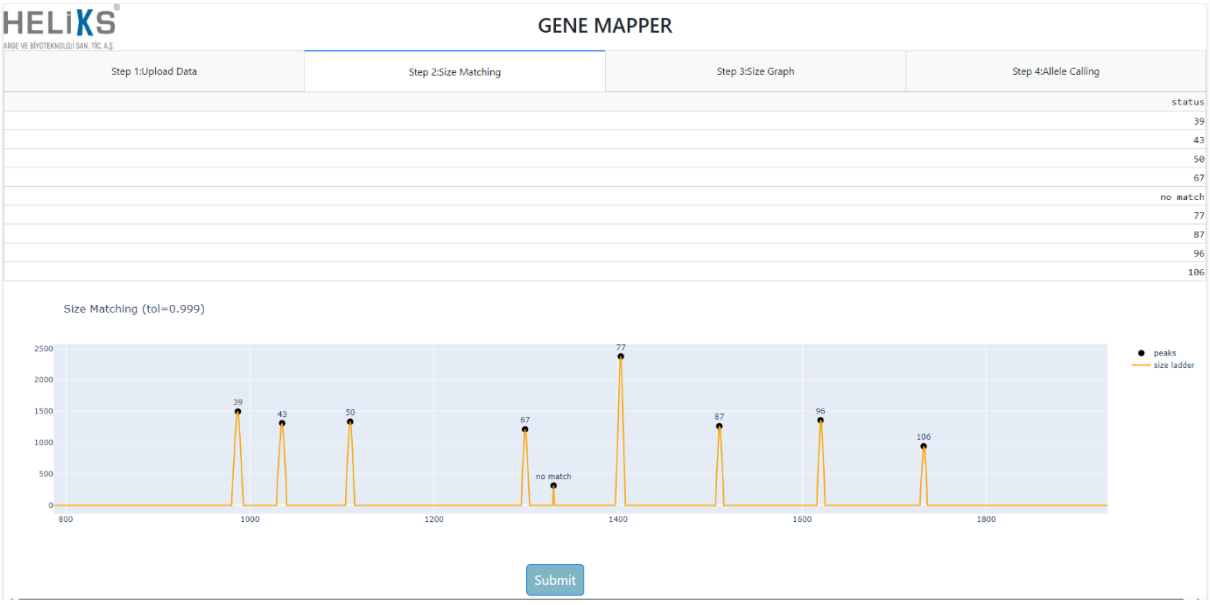


Figure 3. Size Matching step of the gene mapping process.

After matching the size peaks of sample which is given as datapoints and the sizes of the designed size ladder, the resulting datapoint-size match was used to train the regression model to be used and the base pair equivalent of the patient's track data was estimated according to this model. Track data is about the allele informations carried by sample. When the user clicks the submit button, tool automatically takes them to **Step 3: Size Graph** which shows the track data of sample in base pair.

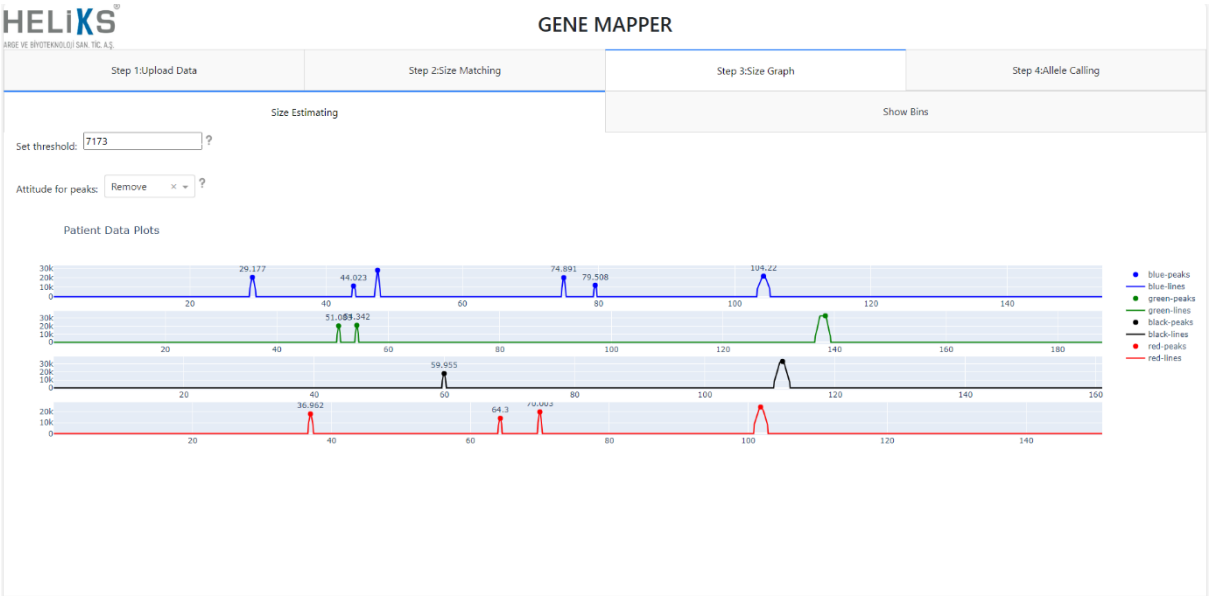


Figure 4. Size Estimating subtab of Step 3:Size Graph. Graphical representation of track data of sample in different colors (blue, green, black and red) is shown. Threshold value for minimum peak height is estimated as 7173 for this sample by taking peak heights into account. According to the "Attitude for peaks" selection, peaks under the threshold value is removed.

The automatically calculated threshold value for sample track data can be changed by the user. If peaks below the threshold value are to be removed, **Attitute for peaks** can be set to Remove (see Figure 4). If peaks below the threshold value are not desired to be deleted, **Attitute for peaks** can be

set to Ignore (see Figure 5).

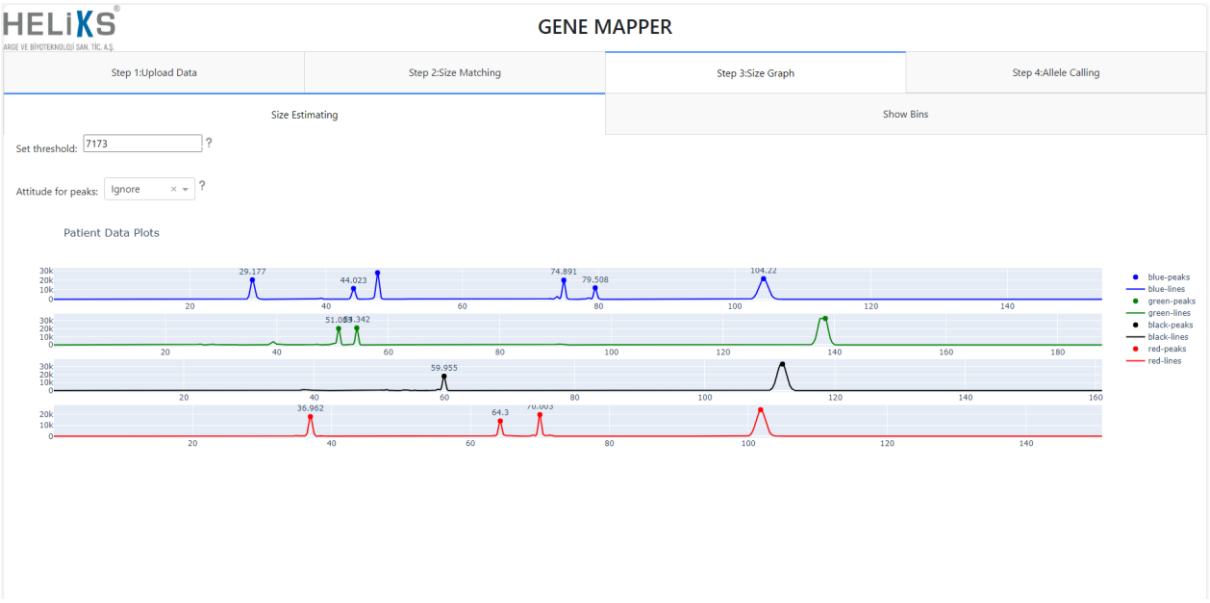


Figure 5. Size Estimating subtab of Step 3:Size Graph. Graphical representation of track data of sample in different colors (blue, green, black and red) is shown. According to the "Attitude for peaks" selection, peaks under the threshold value is just ignored without deleting completely.

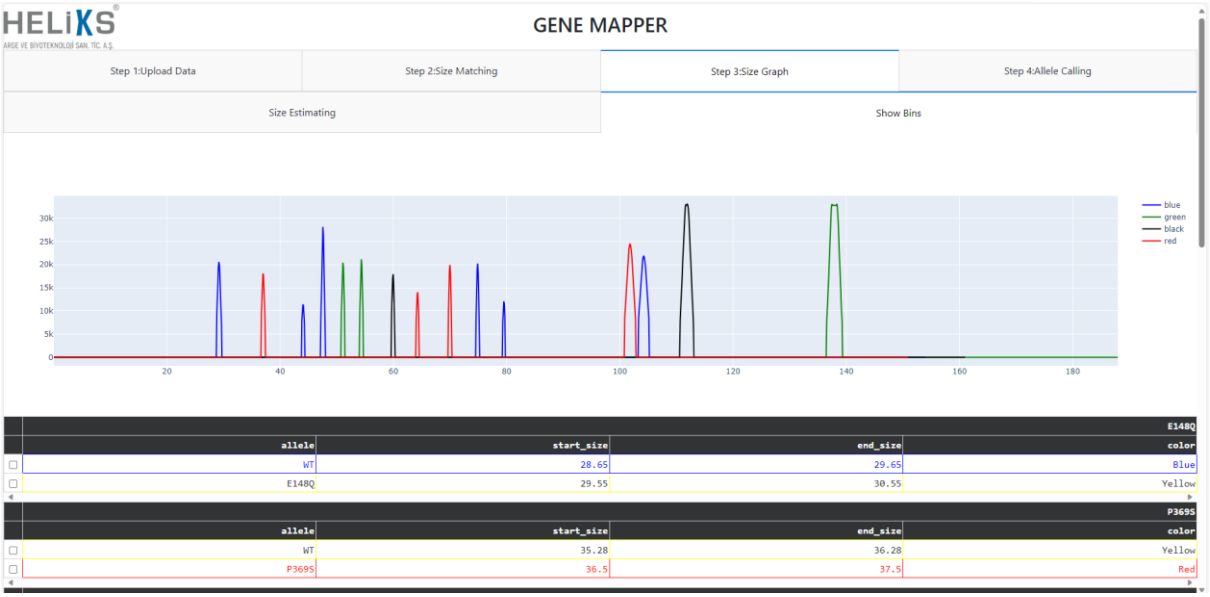


Figure 6. Representation of track data and bin tables.

In the **Show Bins** subtab of Step 3:Size Graph, the track data of the patient is plotted on top of each other and tables are created as much as the number of markers of the panel used (see Figure 6).

Each track data is expressed with a color and is associated with the allele expressed with the same color. Alleles are separated into tables according to marker names, and the rows in the tables are colored by allele color.

In the tables, the start_size and end_size columns next to the allele name refer to the base pair range in which the allele will be represented, and if that allele is represented in that form in the sample, we would expect to see a peak with the same color as the allele color within the scope of this base pair

range, called bin. if the default bin ranges do not match the program's size pair calculation, bin range can be changed on the table (see Figure 8).

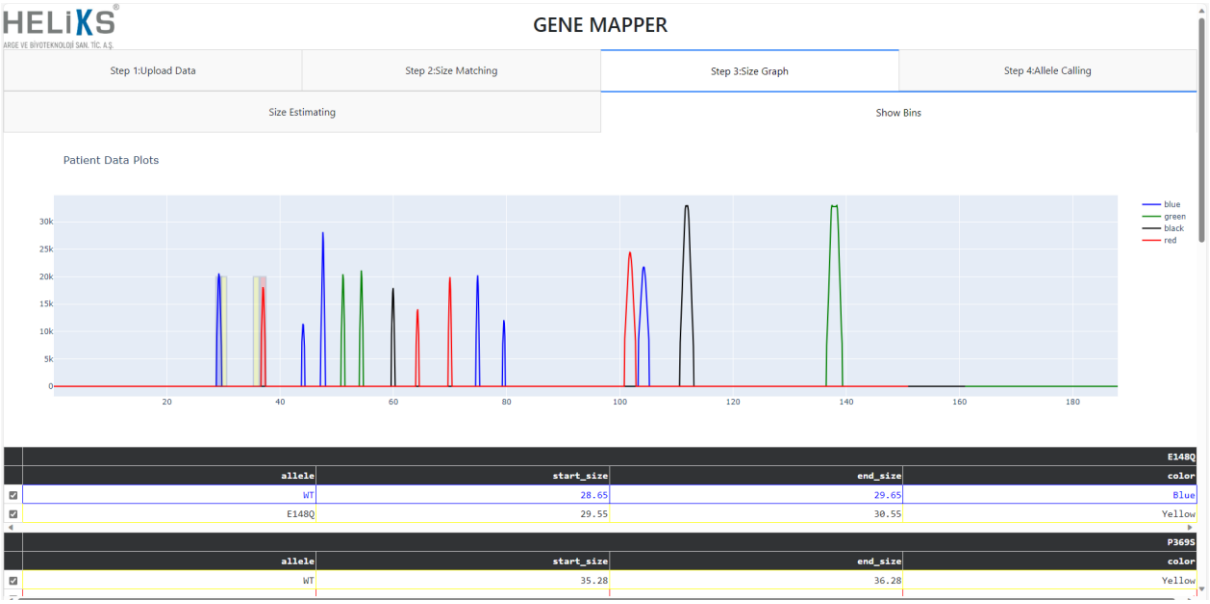


Figure 7. Representation of the alleles on graph which are represented by bin intervals on the table.

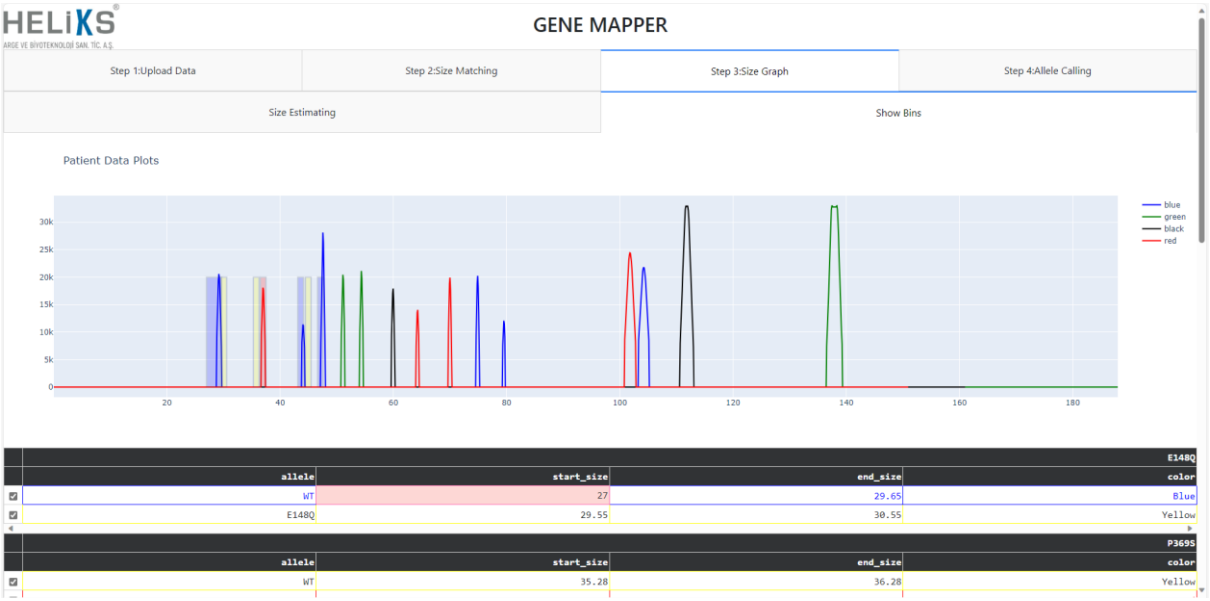


Figure 8. Demonstration that the interval of scope of alleles can be changed in the table.

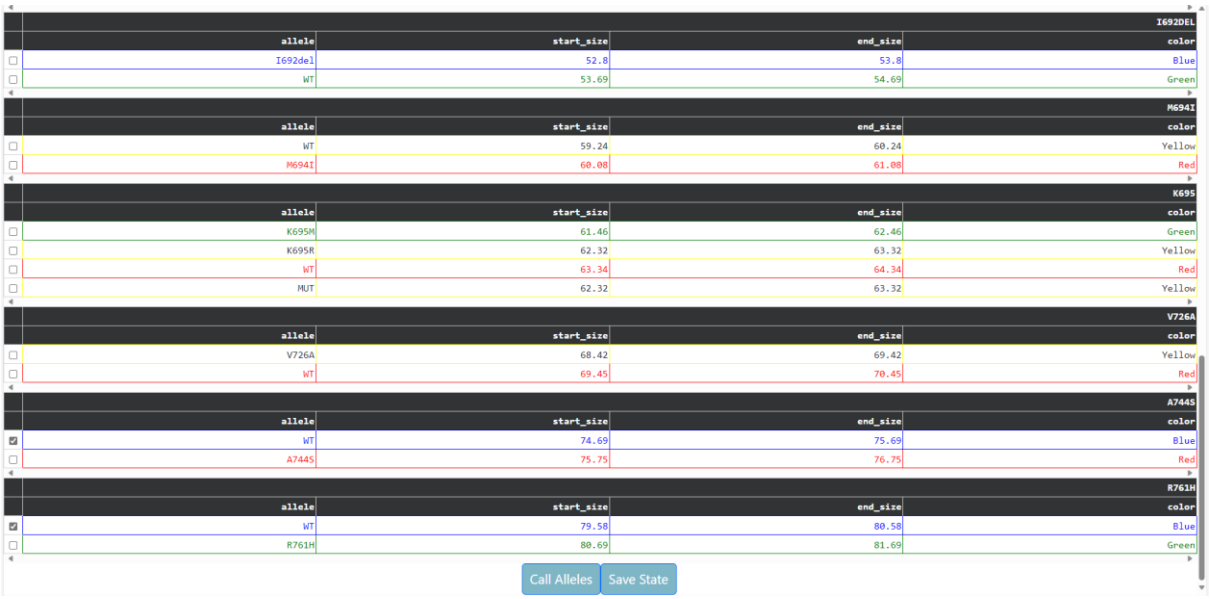


Figure 9. Representation of the intervals of alleles contained in tables separated by marker names.

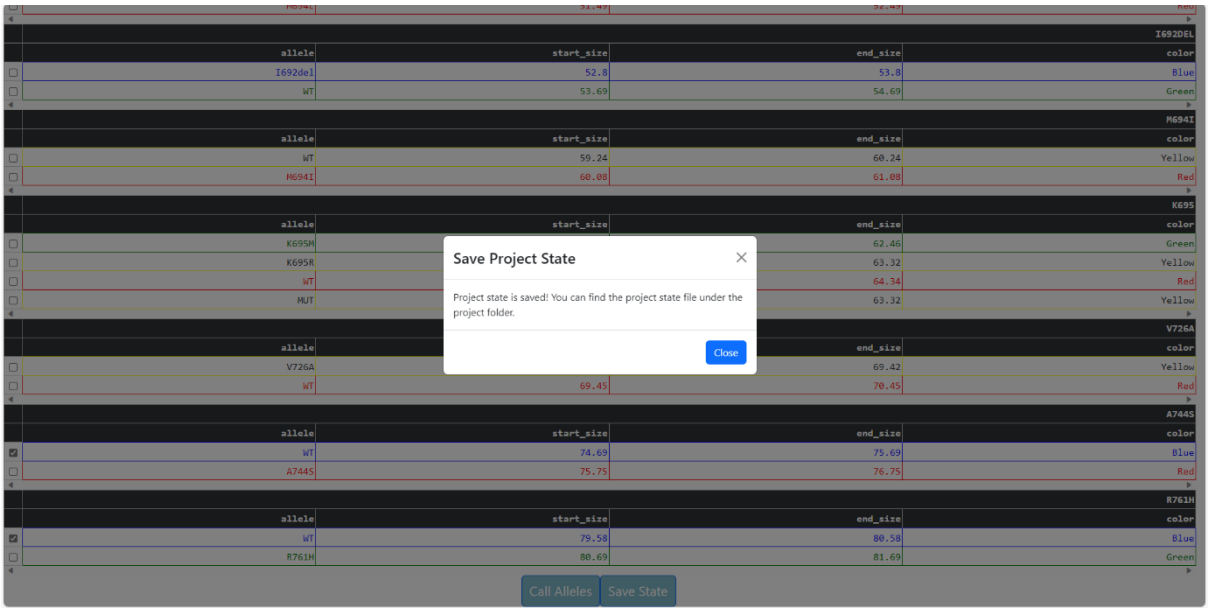
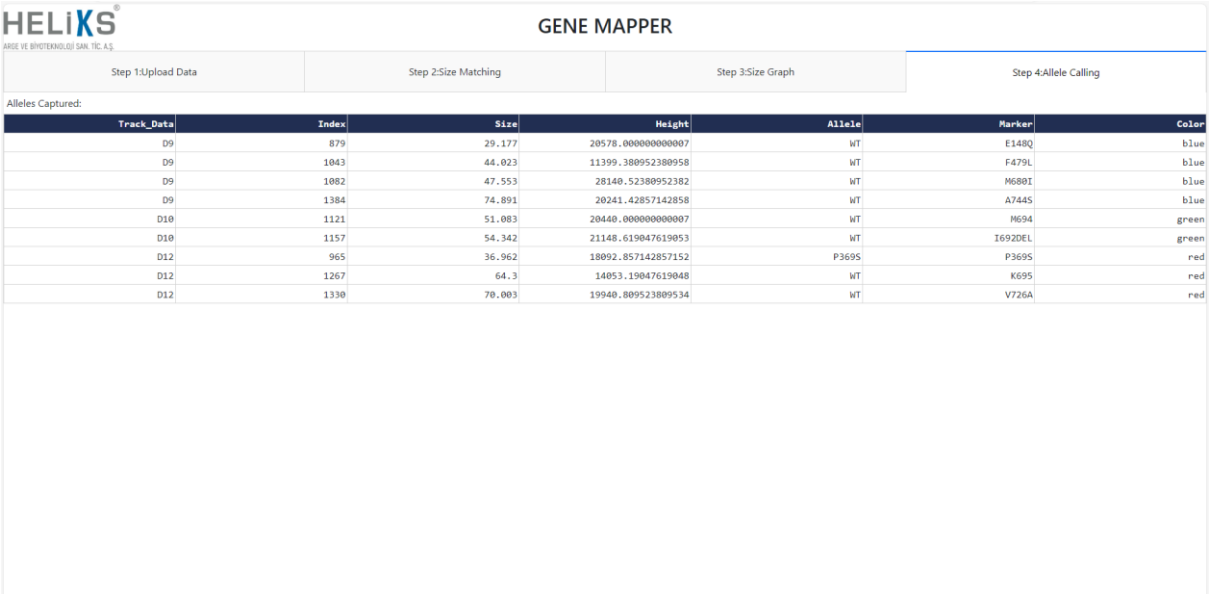


Figure 10. The Save State button saves the selections and changes made in the project to a file.

The Call Alleles button redirects the user to the Step4:Allele Calling tab where a list of alleles found in the sample is presented. With the Save State button, selected files, selections and every change made to tables are saved in a file (Figure 9, Figure 10).

In Figure 11, available alleles are listed after clicking on Call Alleles button.



Track_Data	Index	Size	Height	Allele	Marker	Color
D9	879	29.177	20578.000000000007	WT	E148Q	blue
D9	1043	44.023	11399.380952380958	WT	F479L	blue
D9	1082	47.553	28140.52380952382	WT	M680I	blue
D9	1384	74.891	20241.42857142858	WT	A744S	blue
D10	1121	51.083	20440.000000000007	WT	M694	green
D10	1157	54.342	21148.619047619053	WT	I692DEL	green
D12	965	36.962	18092.857142857152	P369S	P369S	red
D12	1267	64.3	14053.19047619048	WT	K695	red
D12	1330	70.003	19940.809523809534	WT	V726A	red

Figure 11. The Call Alleles button in Show Bins subtab redirects the user to the Step4:Allele Calling tab where a list of alleles found in the sample is presented.