MANUAL

MorphoCatcher: a multiple-alignment based web tool for target selection and designing taxon-specific primers in the loop-mediated isothermal amplification method

Fedor V. Shirshikov[1,*], Yuri A. Pekov[2] and Konstantin A. Miroshnikov[1]

[1]Shemyakin & Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow, Russia
[2]Lomonosov Moscow State University, Moscow, Russia

[*]To whom correspondence should be addressed.

Contact: shrshkv@ya.ru

Introduction

In this short manual we review two case studies of using the MorphoCatcher service for development of taxon-specific or group-specific primer sets and detection of the following potato pathogens by the loop-mediated isothermal amplification (LAMP) method:

- Bacterial potato pathogens from the *Dickeya* genus using an orthologous gene;
- Potato virus Y (PVY) using a whole-genome sequence alignment.

The MorphoCatcher protocol deals with well-established standards of the sequence file formatting. The user can handle FASTA-files with small modifications in the header. The same modifications are needed during formatting of multiple sequence alignment (MSA) using the Clustal Omega web service, because of most deposited nucleotide sequences contain the FASTA-header with spaces that cause trimming of the header after alignment.

It is necessary to understand that the case studies mostly need to illustrate main opportunities of the software developed. We agree that there are more suitable targets for diagnostics of the target taxon. Here, the correct formatting of text files for data processing also shown for each step of the protocol.

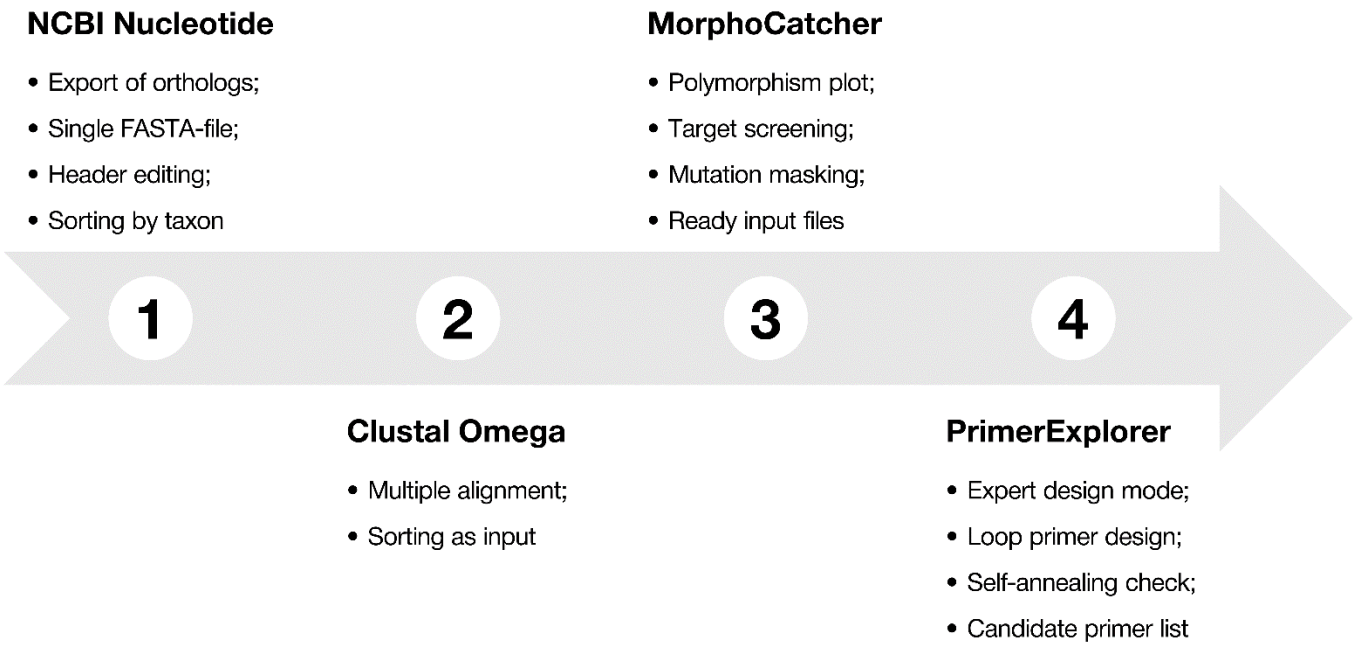
There is no efficient bioinformatics tool for prediction of good primer set. Therefore, we recommend to test different primer sets in laboratory, because at present this is a best way to validate specificity and sensitivity of your LAMP assay. In some cases, the column purification of inner primers is needed to enhance sensitivity, rapidity, and reproducibility of the amplification.

We hope that the MorphoCatcher service will help to our colleagues to create new diagnostic assays not only for the LAMP method, but also for other perspective amplification methods. Please, feel free to contact us by e-mail with any questions and feedback. Together we can make this web tool better for the molecular diagnostics community.

A command line version of the MorphoCatcher tool and set of tutorial files are available at the GitHub public repository: https://github.com/shrshkv/MorphoCatcher.

Case study 1: the bacterial pathogens and orthologous target gene

To perform the primer design correctly, we recommend to follow these steps:



Step 1. Defining closely related bacterial species.

One of the most widespread and dangerous bacterial pathogens of potato belong to the *Pectobacteriaceae* family [1]. The same potato tubers can be contaminated by several species simultaneously, for example by *Dickeya* and *Pectobacterium* spp. Therefore, we should to include some species from the abovementioned genera in our set, for example:

- *Dickeya solani*, which will be a target species for diagnostics;
- *Pectobacterium atrosepticum*;
- *Pectobacterium carotovorum*.

It is necessary to point out that the *Dickeya* genus contains other species, which are also potato pathogens. However, the species have very little differences in their high-conservative genes and symptomatology, therefore, their distinction is not necessary.

Step 2. Selecting target nucleotide sequences.

Potential target nucleotide sequence can be found using the NCBI Nucleotide database (https://www.ncbi.nlm.nih.gov/nuccore), the NCBI GenBank genome annotation, or the NCBI BLAST service (https://blast.ncbi.nlm.nih.gov). To develop specific primer set for detection of *Dickeya* pathogens, we choose a strategy with primer design toward high-conservative housekeeping gene of histidine metabolism, called the *hisC* gene.

Two alternative types of the FASTA-header formatting are compatible with further protocol:

- `>[Taxon ID]|[Target gene ID]|[Strain ID]`
- `>[Taxon ID]|[Strain ID]`

For example, the user can insert instead of `[Taxon ID]` any text that will be associate with target or closely related species. Thus, for the type strain of *D. solani* is IPO2222^T the corresponding FASTA-header are the following:

- `>DSO|hisC|IPO2222`
- `>DSO|IPO2222`

The user is free to use any type of the abovementioned FASTA-header, but for high-performing screening of the target sequence might be useful a long version with gene ID. It allows to differentiate various plots and compare the data more effective. Here, we select the well-established 3-letter species abbreviation and 4-letter gene abbreviation, but any other IDs or custom text and digits can be used as well as NCBI taxon identifier or GenBank accession number. It is necessary to point out that the ID should not to contain spaces, because of the spaces cause the header gap during the alignment by Clustal Omega service.

We recommend to extract target sequence from the type strain with complete genome status. For example, the type strain of *D. solani* is IPO2222^T (GenBank accession no.: NZ_CP015137). In **Example 1.1** you can see required formatting of the FASTA-header in text (*.txt) file with nucleotide sequence of interest.

```
>DSO|hisC|IPO2222
ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCCGCTCGCCGCCTGGGCGGCAACGGTGACGTCTGGCTGAACGCCAACGAG
TACCCGCAGGCGCCGCAGTACCAACTGACCTTGCAGACGCTGAACCGTTACCCGGAATGTCAGCCCGTGCAGGTGATTGAACGCTATGCCGCCTATGCCGGGTTG
AAACCGGAGCAGGTGCTGGTGAGCCGTGGCGCAGACGAAGGCATTGAACTGCTGATTCGCGCGTTCTGCGAACCGGGTAAAGACGCCATTCTGTTCTGTCCACCG
ACCTACGGCATGTACGCCGTCAGCGCTGAAACGTTCGGCGTAGAGCGCCGCGTAGTGCCCTCGACCGCCGACTGGCAACTGGATCTGGCCGCGATCGAAGGCGCG
CTGGACAACGTTAAAGTGATTTATGTTTGTTCGCCCAACAACCCGACCGGAAACCTGATCAACCCAGAGGATCTGCGCCGTCTGCTGGAGTTGGCGCGCGGTCGG
GCCATCGTCGCTATCGACGAAGCCTATATTGAGTTTTGCCCGCAGGCGACCACGGTCGGCTGGCTGGCGGAATACCCGCATCAGGTCGTATTGCGCACGCTGTCC
AAAGCCTTCGCGCTGGCCGGCCTGCGTTGTGGTTTCACCCTTGCCAGCGCAGGGGTGATTCAACTGCTGCTCAAAGTGATCGCCCCCTACCCACTGGCGTTGCCG
GTAGCGGATATCGCCGCGCAGGCGCTGAGCGCTGAGGGTCTGGCGGACATGCGCCGCAATGTGGACGAGGTGCGGGAAAACCGCCGCTGGCTGCGTGAGTCGCTG
AAAGCGTTGTCGAGCATAGAAACGGTATACGCCAGCGAAAGCAATTATCTACTGGTGCGCTTCGCCGATTCGCCAACCGTCTTCAAAACCTTGTGGGATCAGGGC
ATTATTTTACGCGATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTACGAGTGCGAGCGCGTCATCAGCGCATTGCAGGCG
TTATCCGGCAAAACGGCATAA
```

Example 1.1. Required formatting of the target sequence with modified FASTA-header.

Step 3. Exporting orthologous sequences.

Using the NCBI BLAST service and the target nucleotide sequence as a reference, the user should to export some orthologous sequences from closely related bacterial species. In present example, the *hisC* sequences were extracted from three bacterial species, each of them was represented by five different strains, including type strain. All sequences with modified FASTA-header should be merged in one text (*.txt) file as shown below in **Example 1.2**.

```
>DSO|hisC|IPO2222
ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCCGCTCGCCGCCTGGGCGGCAACGGTGACGTCTGGCTGAACGCCAACGAG
TACCCGCAGGCGCCGCAGTACCAACTGACCTTGCAGACGCTGAACCGTTACCCGGAATGTCAGCCCGTGCAGGTGATTGAACGCTATGCCGCCTATGCCGGGTTG
AAACCGGAGCAGGTGCTGGTGAGCCGTGGCGCAGACGAAGGCATTGAACTGCTGATTCGCGCGTTCTGCGAACCGGGTAAAGACGCCATTCTGTTCTGTCCACCG
ACCTACGGCATGTACGCCGTCAGCGCTGAAACGTTCGGCGTAGAGCGCCGCGTAGTGCCCTCGACCGCCGACTGGCAACTGGATCTGGCCGCGATCGAAGGCGCG
CTGGACAACGTTAAAGTGATTTATGTTTGTTCGCCCAACAACCCGACCGGAAACCTGATCAACCCAGAGGATCTGCGCCGTCTGCTGGAGTTGGCGCGCGGTCGG
GCCATCGTCGCTATCGACGAAGCCTATATTGAGTTTTGCCCGCAGGCGACCACGGTCGGCTGGCTGGCGGAATACCCGCATCAGGTCGTATTGCGCACGCTGTCC
AAAGCCTTCGCGCTGGCCGGCCTGCGTTGTGGTTTCACCCTTGCCAGCGCAGGGGTGATTCAACTGCTGCTCAAAGTGATCGCCCCCTACCCACTGGCGTTGCCG
GTAGCGGATATCGCCGCGCAGGCGCTGAGCGCTGAGGGTCTGGCGGACATGCGCCGCAATGTGGACGAGGTGCGGGAAAACCGCCGCTGGCTGCGTGAGTCGCTG
AAAGCGTTGTCGAGCATAGAAACGGTATACGCCAGCGAAAGCAATTATCTACTGGTGCGCTTCGCCGATTCGCCAACCGTCTTCAAAACCTTGTGGGATCAGGGC
ATTATTTTACGCGATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTACGAGTGCGAGCGCGTCATCAGCGCATTGCAGGCG
TTATCCGGCAAAACGGCATAA
>DSO|hisC|F012
ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCCGCTCGCCGCCTGGGCGGCAACGGTGACGTCTGGCTGAACGCCAACGAG
TACCCGCAGGCGCCGCAGTACCAACTGACCTTGCAGACGCTGAACCGTTACCCGGAATGTCAGCCCGTGCAGGTGATTGAACGCTATGCCGCCTATGCCGGGTTG
AAACCGGAGCAGGTGCTGGTGAGCCGTGGCGCAGACGAAGGCATTGAACTGCTGATTCGCGCGTTCTGCGAACCGGGTAAAGACGCCATTCTGTTCTGTCCACCG
ACCTACGGCATGTACGCCGTCAGCGCTGAAACGTTCGGCGTAGAGCGCCGCGTAGTGCCCTCGACCGCCGACTGGCAACTGGATCTGGCCGCGATCGAAGGCGCG
CTGGACAACGTTAAAGTGATTTATGTTTGTTCGCCCAACAACCCGACCGGAAACCTGATCAACCCAGAGGATCTGCGCCGTCTGCTGGAGTTGGCGCGCGGTCGG
GCCATCGTCGCTATCGACGAAGCCTATATTGAGTTTTGCCCGCAGGCGACCACGGTCGGCTGGCTGGCGGAATACCCGCATCAGGTCGTATTGCGCACGCTGTCC
AAAGCCTTCGCGCTGGCCGGCCTGCGTTGTGGTTTCACCCTTGCCAGCGCAGGGGTGATTCAACTGCTGCTCAAAGTGATCGCCCCCTACCCACTGGCGTTGCCG
GTAGCGGATATCGCCGCGCAGGCGCTGAGCGCTGAGGGTCTGGCGGACATGCGCCGCAATGTGGACGAGGTGCGGGAAAACCGCCGCTGGCTGCGTGAGTCGCTG
AAAGCGTTGTCGAGCATAGAAACGGTATACGCCAGCGAAAGCAATTATCTACTGGTGCGCTTCGCCGATTCGCCAACCGTCTTCAAAACCTTGTGGGATCAGGGC
ATTATTTTACGCGATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTACGAGTGCGAGCGCGTCATCAGCGCATTGCAGGCG
TTATCCGGCAAAACGGCATAA

...

>PCA|hisC|YCD57
ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCGTTGACCCCCTACCAATCTGCCCGTCGTCTGGGCGGTAACGGCGATGTCTGGTTGAATGCGAAT
GAATATCCTGAAGCACCAGAGTTTCAGTTAACGCTGCAAACGCTGAACCGCTACCCGGAGTGTCAGCCGGTCACGGTGATCAACCGTTACGCTGAATATGCTGGC
GTAACGCCGGAGCAGGTTCTGGTCAGCCGCGGTGCCGACGAAGGTATTGAGCTGCTGATTCGCGCGTTCTGCGAACCGGGGAAAGATGCCATCCTGTTCTGTCCG
CCAACCTACGGCATGTACGCCGTCAGCGCTGAGACGTTTGGCGTCGAACGTCGCACCGTAGCCAGTAAATCAGACTGGCAGTTGGATCTCGATGCGATTGAAGCA
CAACTGGACGGCACCAAAGTCATTTACGTATGCAGCCCAAATAACCCAACCGGCAACCTGATTGCACGGGAGGATTTACGCCAGTTGCTCACAATGGCACAGGGC
AAAGCGTTGGTCGTTATCGATGAGGCCTACATCGAGTTCTGTCCGCAGGCGTCCACCGCAGTCTGGCTAAGCGAGTTTCCTCATTTGGTGATTCTGCGTACCCTG
TCCAAAGCCTTTTCACTCGCGGGGCTGCGCTGTGGCTTCACGCTGGCGAACCCTGAAGTGATTCAGCTTCTGCTGAAGGTGATTGCCCCTTATCCGCTGTCTACT
CCGGTGGCAGATATCGCCGCGCAGGCGCTGAGTCACGAAGGGATCGCCAAAATGAAGGCGAATGTTGAGGAAATTACCTCCGCTCGCCGCTGGCTGAGCGATGCC
TTAAAAGACATTCCCTGCGTTGAAGAAATCTTCCCCAGCGAGAGCAACTATTTGCTGGTGCGCTTTACCGCTTCCCCTTCTGTATTTAAAACGCTGTGGGATCAA
GGCATTATTCTGCGTGACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTACGAATGTGAACGCGTTGTTGCTGCATTACAA
TCATTGCCGGGCATCAACGCTTAA
```

Example 1.2. Single FASTA-file of orthologous sequences with modified headers.

Step 4. Making multiple sequence alignment.

To perform multiple alignment of the orthologs, we recommend to use the Clustal Omega service (https://www.ebi.ac.uk/Tools/msa/clustalo) with parameters that shown in **Figure 1.1**. It is necessary to select an "Input" value of the "Order" menu to save the species order in single FASTA-file. This point is important because the MorphoCatcher algorithm is based to analyze the strains of different species as separate groups and find their conservative species-specific mutations. Then the user should to save an alignment text (*.txt) file as shown in **Example 1.3**.

Figure 1.1. Screenshot of the Clustal Omega service with required options for alignment. The "Input" option of "Order" menu should be selected.

```
CLUSTAL O(1.2.4) multiple sequence alignment


DSO|hisC|IPO2222            ---ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC       57
DSO|hisC|F012               ---ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC       57
DSO|hisC|MK10               ---ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC       57
DSO|hisC|MK16               ---ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC       57
DSO|hisC|RNS08.23.3.1.A     ---ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC       57
PAT|hisC|NCPPB549           ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCATTGACACCTTATCAGTCT       60
PAT|hisC|JG10-08            ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCATTGACACCTTATCAGTCT       60
PAT|hisC|NCPPB3404          ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCATTGACACCTTATCAGTCT       60
PAT|hisC|SCRI1043           ATGAGCAGCATTGAAGAACTAGCACGCGCCAACGTCCGTGCATTGACACCGTATCAGTCT       60
PAT|hisC|21A                ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCATTGACACCTTATCAGTCT       60
PCA|hisC|NCPPB312           ATGAGCAGCATTGAAGAACTGGCACGCGCCAATGTCCGTGCGCTGACGCCGTACCAATCT       60
PCA|hisC|BCT2               ATGAGCAGTATTGAAGAACTGGCACGCGCCAATGTCCGTGCGCTGACGCCGTACCAATCT       60
PCA|hisC|BCT5               ATGAGCAGCATTGAAGAACTGGCACGCGCCAATGTCCGTGCGCTGACGCCGTACCAATCT       60
PCA|hisC|WPP14              ATGAGCAGCATTGAAGAACTGGCACGCGCCAATGTCCGTGCGCTGACGCCGTACCAATCT       60
PCA|hisC|YCD57              ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCGTTGACCCCCTACCAATCT       60
                               *  ** *****  * ** ** ** * *** ** **     **** ** ** ** **
...

DSO|hisC|IPO2222            GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC       1017
DSO|hisC|F012               GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC       1017
DSO|hisC|MK10               GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC       1017
DSO|hisC|MK16               GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC       1017
DSO|hisC|RNS08.23.3.1.A     GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC       1017
PAT|hisC|NCPPB549           GACCAAAATAAGCAACCCAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC       1020
PAT|hisC|JG10-08            GACCAAAATAAGCAACCCAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC       1020
PAT|hisC|NCPPB3404          GACCAAAATAAGCAACCCAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC       1020
PAT|hisC|SCRI1043           GACCAAAATAAGCAACCCAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC       1020
PAT|hisC|21A                GACCAAAATAAGCAACCCAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC       1020
PCA|hisC|NCPPB312           GACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC       1020
PCA|hisC|BCT2               GACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC       1020
PCA|hisC|BCT5               GACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTACGCATTACTATCGGCAACCGCTAC       1020
PCA|hisC|WPP14              GACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC       1019
PCA|hisC|YCD57              GACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC       1020
                            ** ** ** ** ** **  * ** ** ** *** ********** ***************

DSO|hisC|IPO2222            GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA      1071
DSO|hisC|F012               GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA      1071
DSO|hisC|MK10               GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA      1071
DSO|hisC|MK16               GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA      1071
DSO|hisC|RNS08.23.3.1.A     GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA      1071
PAT|hisC|NCPPB549           GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA      1074
PAT|hisC|JG10-08            GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA      1074
PAT|hisC|NCPPB3404          GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA      1074
PAT|hisC|SCRI1043           GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCAGGCATCAACGCTTAA      1074
PAT|hisC|21A                GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA      1074
PCA|hisC|NCPPB312           GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA      1074
PCA|hisC|BCT2               GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA      1074
PCA|hisC|BCT5               GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA      1074
PCA|hisC|WPP14              GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA      1073
PCA|hisC|YCD57              GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA      1074
                            ** ** ** *****  *    ***** **  * **  * ****  *  ** ***
```

Example 1.3. Multiple sequence alignment of the *hisC* orthologous sequences. Conservative nucleotide positions are highlighted by "*" symbols.

Step 5. Masking species-specific mutations.

Here, we suggest to use the MorphoCatcher service (http://morphocatcher.ru) for processing of the MSA. The MorphoCatcher allows the user to upload the alignment and select the output file format, which depends on computer operating system (e.g. Windows or UNIX). The default parameter set is "Windows" output file format.

After alignment uploading the user should to click on the "Catch" button, which starts processing of data to detect species-specific mutations (e.g. single nucleotide polymorphisms, insertions, and deletions). Then, the new tab with results will shows a graphic representation of the mutation density using a sliding-window function (**Figure 1.2**).

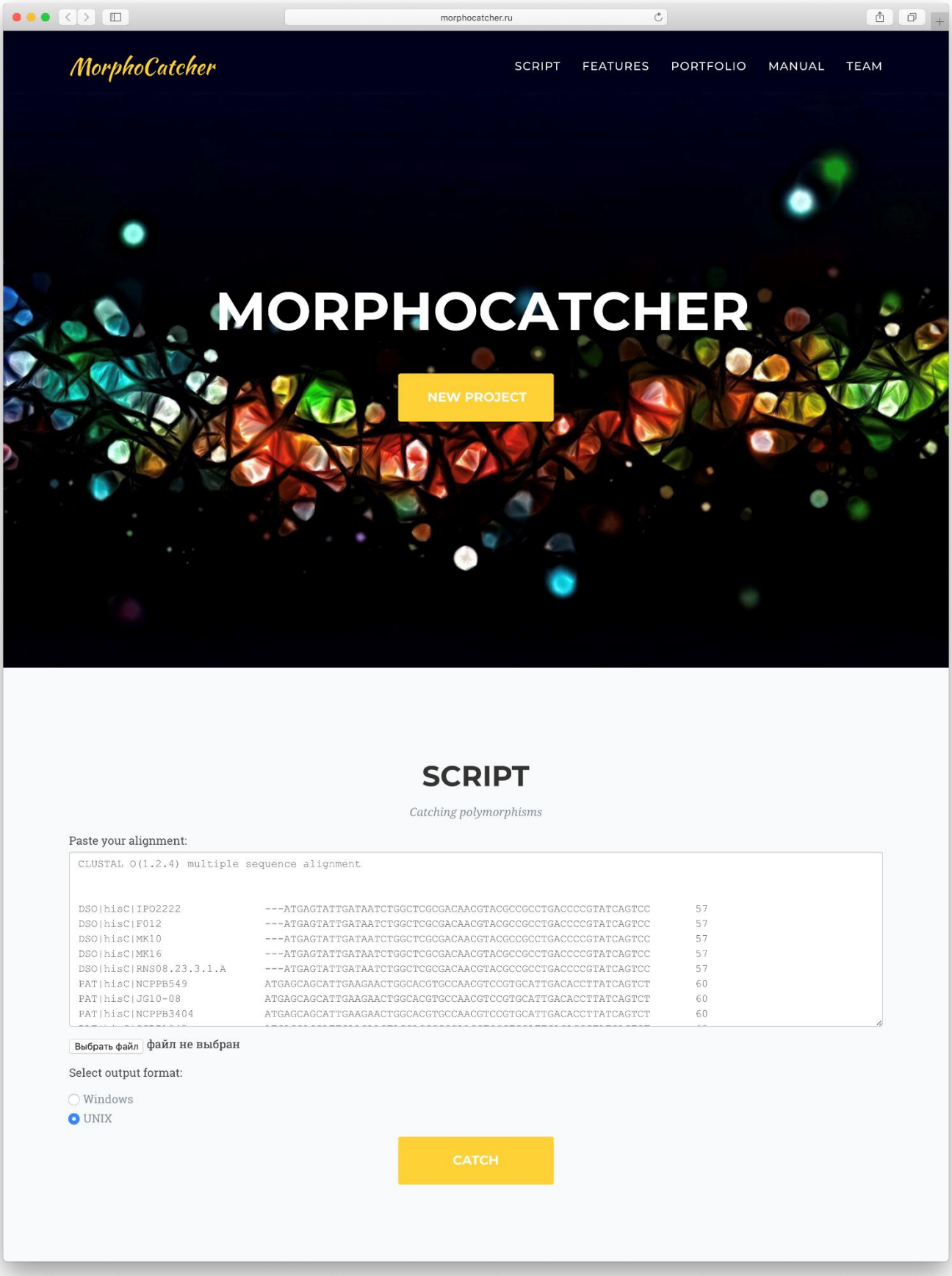


Figure 1.2. Screenshot of the MorphoCatcher service with option of the output format.

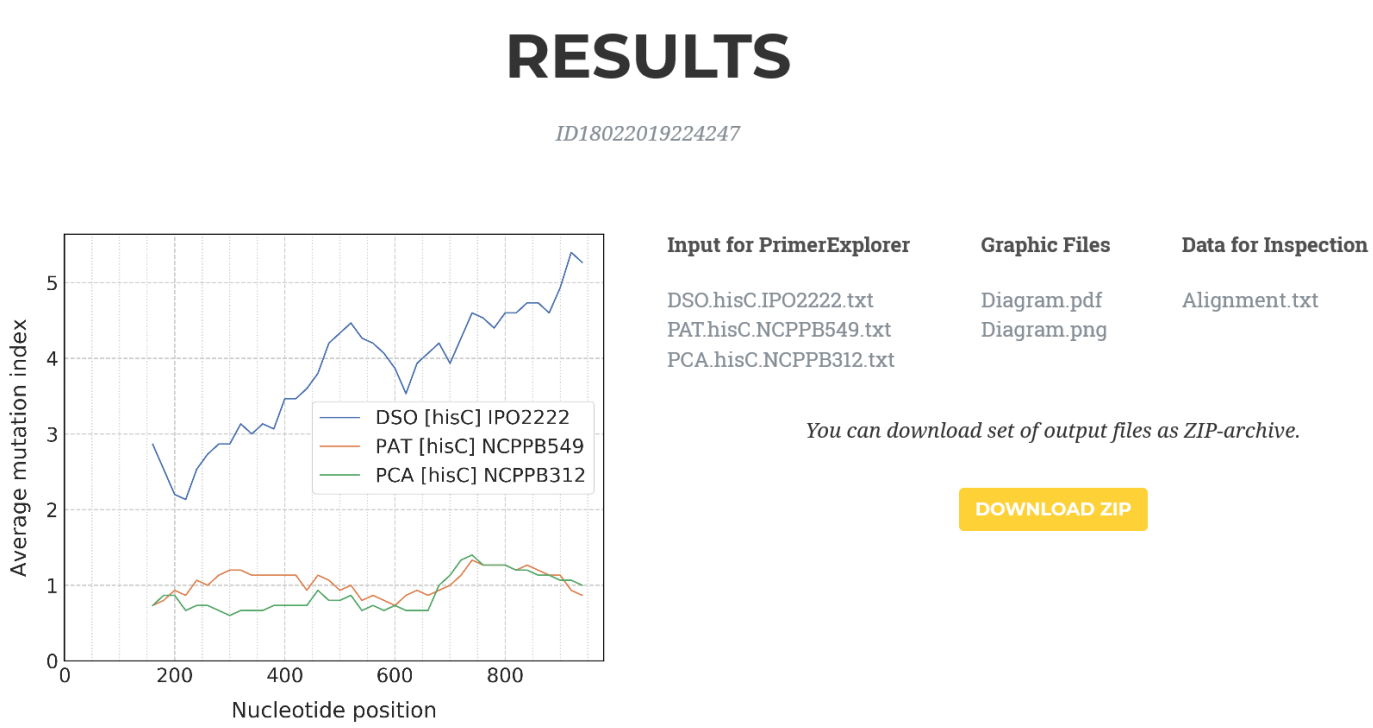


Figure 1.3. Screenshot of tab with results after processing of the multiple sequence alignment.

Among output files (**Figure 1.3**) the user can find the following files with useful data:

- Processed MSA with masked species-specific mutations;
- Sliding-window plot of gene polymorphisms for each species (**Figure 1.4**);
- Ready-to-use files for specific primer design using the PrimerExplorer service.

All these files can be download as ZIP-archive with corresponding identification number of the project. Graphic files have a high resolution for publications and available with different extensions (*.eps, *.pdf, *.png, *.svg, and *.tiff). Some large files (*.tiff, *.eps, and *.svg) are included only in ZIP-archive. Other graphic (*.pdf and *.png) and text (*.txt) files are available for download and preview in new tab of your browser.

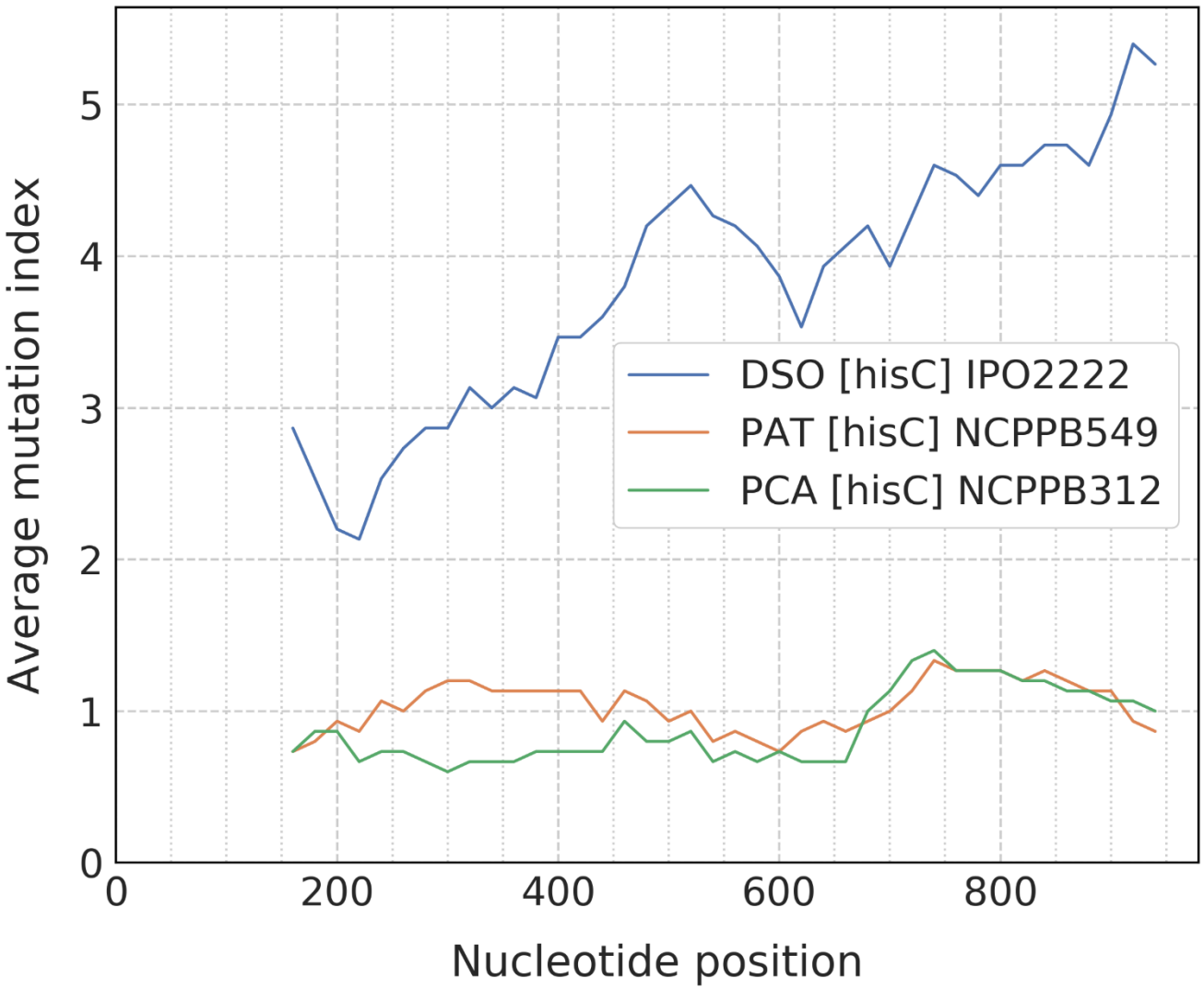


Figure 1.4. Sliding-window plot of the *hisC* gene polymorphism in closely related species. Peaks correspond to the high density of species-specific mutations.

The sequence regions with maximal average mutation index correspond to the high density of species-specific mutations. To select most suitable target region, we recommend the average mutation index values from 2 and higher. If the target selected will be compatible with the design algorithm of PrimerExplorer, the most of species-specific mutations can be included at the primer ends for better specificity.

The user can also inspect positions of the species-specific mutations manually by using processed alignment text file from the output. For this procedure the processed alignment file contains number of species-specific mutations for each 20 nucleotides (**Example 1.4**).

```
DSO|hisC|IPO2222            ---ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC   57
DSO|hisC|F012               ---ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC   57
DSO|hisC|MK10               ---ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC   57
DSO|hisC|MK16               ---ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC   57
DSO|hisC|RNS08.23.3.1.A     ---ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC   57
                            >>>*--*********--*-*****-****-******-**----****************-
                            5                   5                   3
PAT|hisC|NCPPB549           ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCATTGACACCTTATCAGTCT   60
PAT|hisC|JG10-08            ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCATTGACACCTTATCAGTCT   60
PAT|hisC|NCPPB3404          ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCATTGACACCTTATCAGTCT   60
PAT|hisC|SCRI1043           ATGAGCAGCATTGAAGAACTAGCACGCGCCAACGTCCGTGCATTGACACCGTATCAGTCT   60
PAT|hisC|21A                ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCATTGACACCTTATCAGTCT   60
                            ****************************************-*****************
                            0                   0                   1
PCA|hisC|NCPPB312           ATGAGCAGCATTGAAGAACTGGCACGCGCCAATGTCCGTGCGCTGACGCCGTACCAATCT   60
PCA|hisC|BCT2               ATGAGCAGTATTGAAGAACTGGCACGCGCCAATGTCCGTGCGCTGACGCCGTACCAATCT   60
PCA|hisC|BCT5               ATGAGCAGCATTGAAGAACTGGCACGCGCCAATGTCCGTGCGCTGACGCCGTACCAATCT   60
PCA|hisC|WPP14              ATGAGCAGCATTGAAGAACTGGCACGCGCCAATGTCCGTGCGCTGACGCCGTACCAATCT   60
PCA|hisC|YCD57              ATGAGCAGCATTGAAGAACTGGCACGTGCCAACGTCCGTGCGTTGACCCCCTACCAATCT   60
                            ****************************************-**********-**-***
                            0                   0                   3
...


DSO|hisC|IPO2222            GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC   1017
DSO|hisC|F012               GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC   1017
DSO|hisC|MK10               GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC   1017
DSO|hisC|MK16               GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC   1017
DSO|hisC|RNS08.23.3.1.A     GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC   1017
                            **-**-**-**-**-***-*-**-**-**-***-**********-***************
                            6                   5                   1
PAT|hisC|NCPPB549           GACCAAAATAAGCAACCCAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC   1020
PAT|hisC|JG10-08            GACCAAAATAAGCAACCCAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC   1020
PAT|hisC|NCPPB3404          GACCAAAATAAGCAACCCAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC   1020
PAT|hisC|SCRI1043           GACCAAAATAAGCAACCCAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC   1020
PAT|hisC|21A                GACCAAAATAAGCAACCCAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC   1020
                            *****************-*********************************************
                            1                   0                   0
PCA|hisC|NCPPB312           GACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC   1020
PCA|hisC|BCT2               GACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC   1020
PCA|hisC|BCT5               GACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC   1020
PCA|hisC|WPP14              GACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC   1019
PCA|hisC|YCD57              GACCAAAATAAGCAACCGAGCCTCGCAGGCTGTTTGCGCATTACTATCGGCAACCGCTAC   1020
                            ************************************************************
                            0                   0                   0


DSO|hisC|IPO2222            GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA   1071
DSO|hisC|F012               GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA   1071
DSO|hisC|MK10               GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA   1071
DSO|hisC|MK16               GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA   1071
DSO|hisC|RNS08.23.3.1.A     GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA   1071
                            **-**-**-*******--*----******-**--*-**********--*--**-***
                            8                   7                   5
PAT|hisC|NCPPB549           GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA   1074
PAT|hisC|JG10-08            GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA   1074
PAT|hisC|NCPPB3404          GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA   1074
PAT|hisC|SCRI1043           GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCAGGCATCAACGCTTAA   1074
PAT|hisC|21A                GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA   1074
                            ********************************************************
                            0                   0                   0
PCA|hisC|NCPPB312           GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA   1074
PCA|hisC|BCT2               GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA   1074
PCA|hisC|BCT5               GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA   1074
PCA|hisC|WPP14              GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA   1073
PCA|hisC|YCD57              GAATGTGAACGCGTTGTTGCTGCATTACAATCATTGCCGGGCATCAACGCTTAA   1074
                            ********************************************************
                            0                   0                   0
```

Example 1.4. Processed alignment of the *hisC* sequences with masked species-specific mutations. Symbols: "*" — conservative nucleotide position; "-" — species-specific nucleotide position; and ">" — presence of deletion in gene. When some positions require the user decision, such nucleotides can be highlighted by "?" symbol.

If the region with maximal mutation density is suitable for the taxon of interest, then the user can submit text files from the "Input for PrimerExplorer" column (**Example 1.5**) directly to the PrimerExplorer service. It is important that the PrimerExplorer algorithm has the limitation of uploaded target sequence and, therefore, the target sequence region should be less than 2000 nucleotides. If the length greater than the required limitation, we should to trim the output sequence with consensus string manually. We recommend to use the target region with maximal average mutation index as the PrimerExplorer input file for better primer specificity.

```
ATGAGTATTGATAATCTGGCTCGCGACAACGTACGCCGCCTGACCCCGTATCAGTCC
*--**********--*-*****-****-******-**----****************-

GCTCGCCGCCTGGGCGGCAACGGTGACGTCTGGCTGAACGCCAACGAGTACCCGCAGGCG
*****-*****************-**-**********-*****-**-*********-***

CCGCAGTACCAACTGACCTTGCAGACGCTGAACCGTTACCCGGAATGTCAGCCCGTGCAG
***-***-----*-*-**-****************-*********************---*

...

CGCTTCGCCGATTCGCCAACCGTCTTCAAAACCTTGTGGGATCAGGGCATTATTTTACGC
*****--***-****-*****-**-**-******--**********-**********-*-**-

GATCAGAACAAACAGCCGGGGCTGGCGGGTTGTCTGCGCATTACCATCGGCAACCGCTAC
**-**-**-**-**-***-*-**-**-**-**-**********-***************

GAGTGCGAGCGCGTCATCAGCGCATTGCAGGCGTTATCCGGCAAAACGGCATAA
**-**-**-******--*----*****-*---*-**--*******--**-***
```

Example 1.5. Ready-to-use input file for specific primer design using the PrimerExplorer software. Symbols: "*" — conservative nucleotides; "-" — target species-specific mutations.

Step 6. Designing specific primers.

We have developed the MorphoCatcher to automatize some steps of screening for target nucleotide sequence in genome of any biological taxa and enhance specific primer designing option of the PrimerExplorer service, version 5.0 (http://primerexplorer.jp/e). Now, molecular biologists have very useful tool to screen group-specific mutations for taxa of interest and introduce them at primer ends to increase the specificity of new LAMP-based diagnostic assay.

To start the primer design for species of interest, the user should to upload the text file with target nucleotide sequence (see **Example 1.5**) from the "Input for PrimerExplorer" column directly to the PrimerExplorer service using the "Browse" button. Then the user should to click on the "Primer design" button. After these actions, the user will be redirected to the page with primer design options (**Figure 1.5**). The user can read an official manual for the PrimerExplorer service from the abovementioned web site for any details.

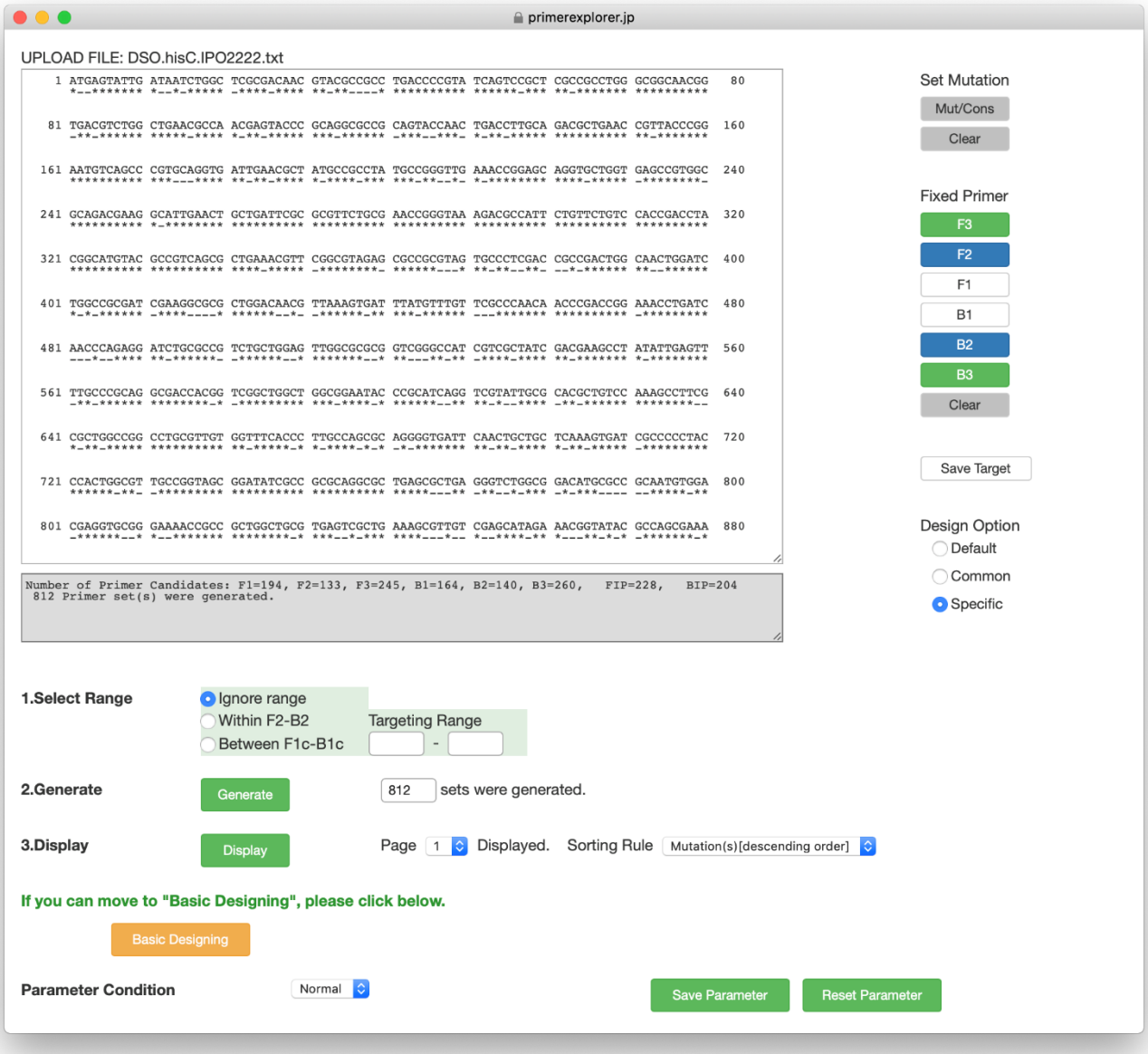


Figure 1.5. Interface of the PrimerExplorer service after uploading target nucleotide sequence. Any specific mutation that can be covered by primer ends is highlighted by "–" symbol.

To design primer sets with species-specific mutations at 5'- or 3'-ends, the user should to configure these important primer parameters of the PrimerExplorer service:

- To select "Specific" mode of the "Primer option" menu;
- To click on the "Detail settings" button to turn on an expert mode;
- To select a required GC-content using the "Parameter condition" menu;
- To click on the "Generate" button;
- To set "Descending order" of primer sets in the "Sorting rule" menu;
- To click on the "Display" button.

After that the user will be redirected to the page with various combinations of primer annealing sites, which can cover various species-specific mutations (**Figure 1.6**). Some of the mutations are localized at primer ends (**Figure 1.7**) that able to increase specificity of the amplification.

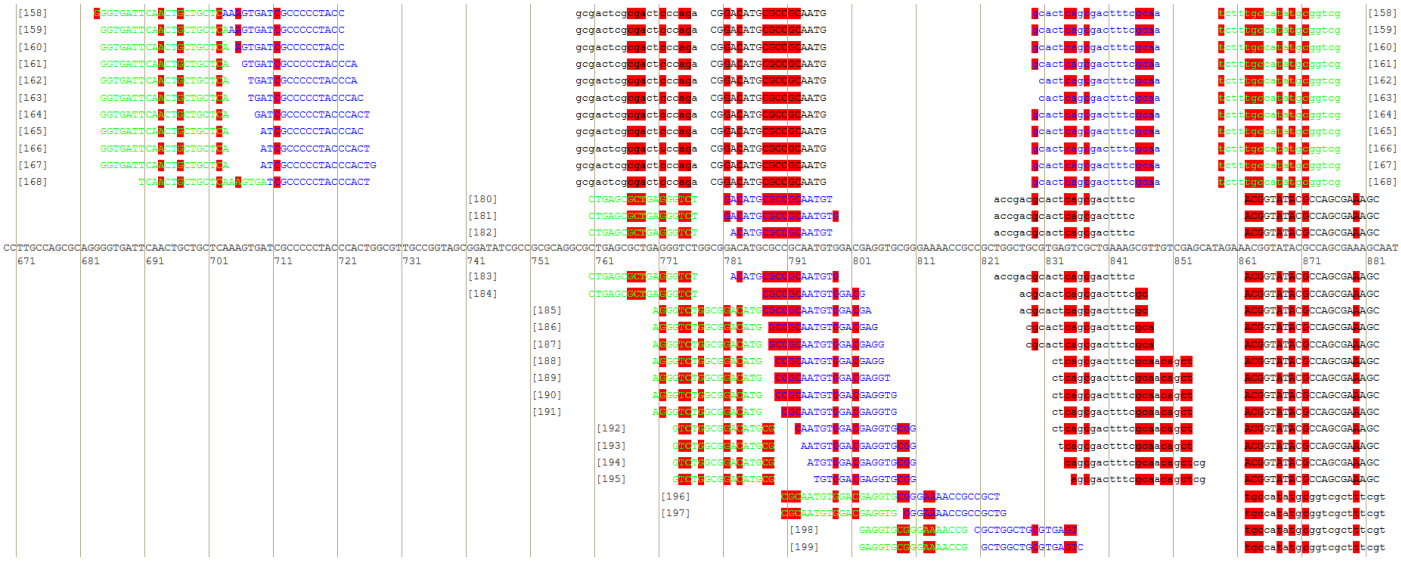


Figure 1.6. The PrimerExplorer output page with full view of designed primer sets. Detected species-specific mutations for strains of *D. solani* are highlighted by red color. The same mutations can be covered by oligos from different primer sets.

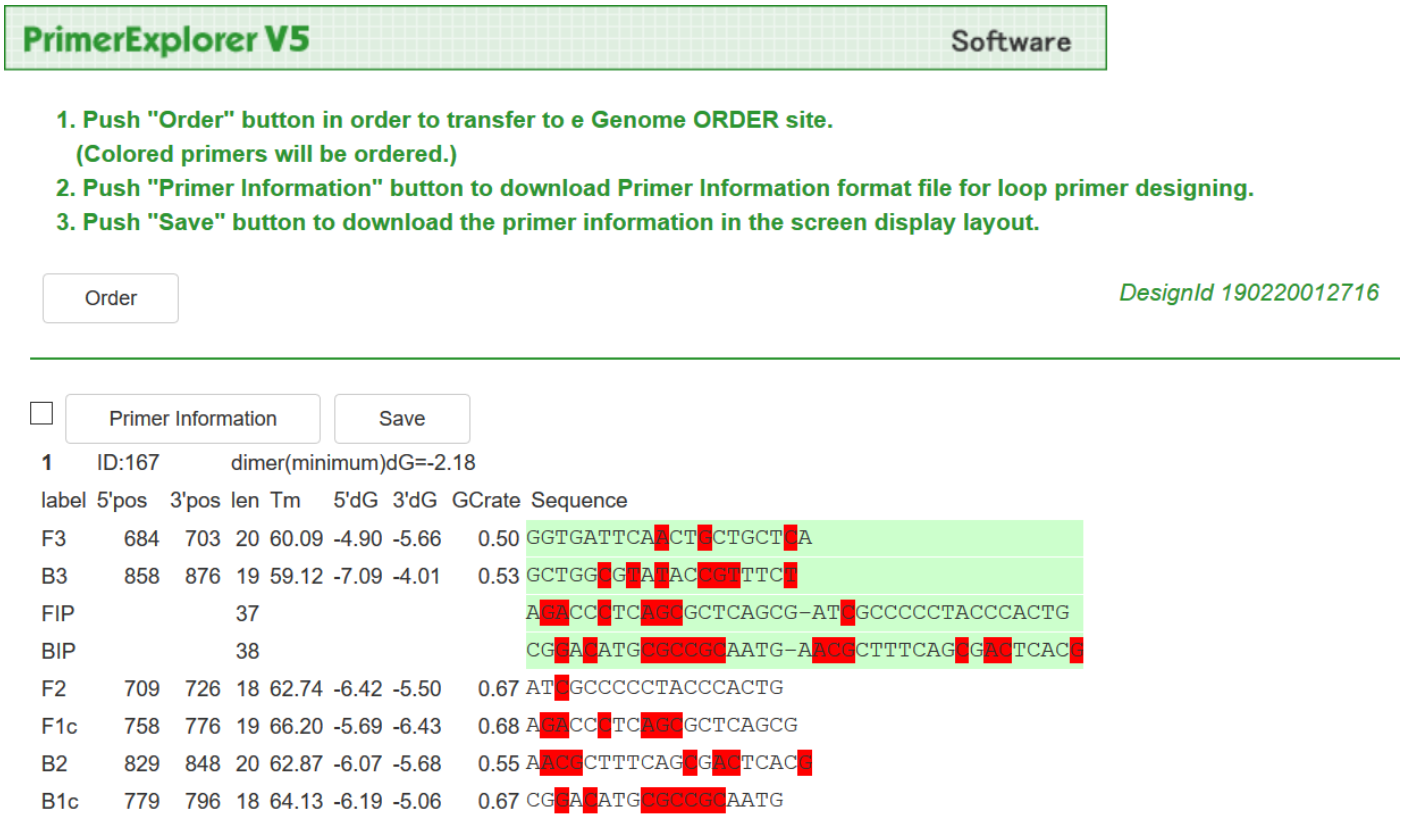


Figure 1.7. Detailed view of the primer set sequences with different thermodynamic parameters. Species-specific mutations that covered by primer ends are highlighted by red color.

Case study 2: the viral targets and whole-genome multiple alignment

Step 1. Defining strains from different serotypes.

Well-known viral potato pathogen is potato virus Y (PVY) that causes mosaic and leaf drop. Genome of PVY is the single-strand (+) RNA molecule and the length is approximately 9.7 kb. Most potato infections are caused by two main strains that have similar genome organization:

- PVY^O — the ordinary strain;
- PVY^N — the necrotic strain.

There are also various new recombinants derived from the abovementioned strains [2]. All these strains and recombinants usually belong to two serotypes, called O- and N-serotype. Due to different symptoms caused by the PVY serotypes, it is necessary to distinguish them for the potato seed certification purposes. In this case study, we attempt to use whole genome sequences for designing LAMP primer sets for the detection of different PVY serotypes. The genome size is applicable for fast multiple alignment of different PVY strains. Our main goal was to decide which genome region can be used for the developing of serotype-specific assays.

Step 2. Selecting target genome sequences.

Here, we attempt to distinguish the PVY serotypes using the MSA of original PVY strains and their recombinants and try to apply the strategy of serotype-specific polymorphisms search.

Step 3. Exporting genome sequences.

To obtain whole-genome sequences of PVY strains, we collect data from NCBI Nucleotide database (https://www.ncbi.nlm.nih.gov/nuccore) and scientific literature [3] by entering corresponding accession numbers in the service search bar. The PVY genomes with corresponding accession numbers can be easily accessed at the link:

https://www.ncbi.nlm.nih.gov/nuccore/AJ585195,AJ585196,AJ890349,DQ157178,EF558545,HQ912864,HQ912865,HQ912888,U09509,X12456,AJ585197,AJ585342,AJ889866,AY166866,AY884982,AY884983,FJ666337,HQ912869,M95491,X97895

The user can to download all these genome sequences as single FASTA-file using corresponding menu of NCBI Nucleotide web page (see **Figure 2.1**) and after that to change the *.fasta extension to *.txt by renaming of the file for comfort data processing. For correct protocol execution the user should to modify the FASTA-header for each genome of interest. The user is free to include any additional service information to the FASTA-header. We use the accession number, because of the number is already part of the standard FASTA-header of each genome at the NCBI Nucleotide database. Thus, we have two main groups that correspond to O- and N-serotype, therefore, the FASTA-headers for each PVY genome can be modified using the following mask:

- `>PVYO|[Accession]`
- `>PVYN|[Accession]`

The final view of PVY genome sequences is shown in **Example 2.1**.

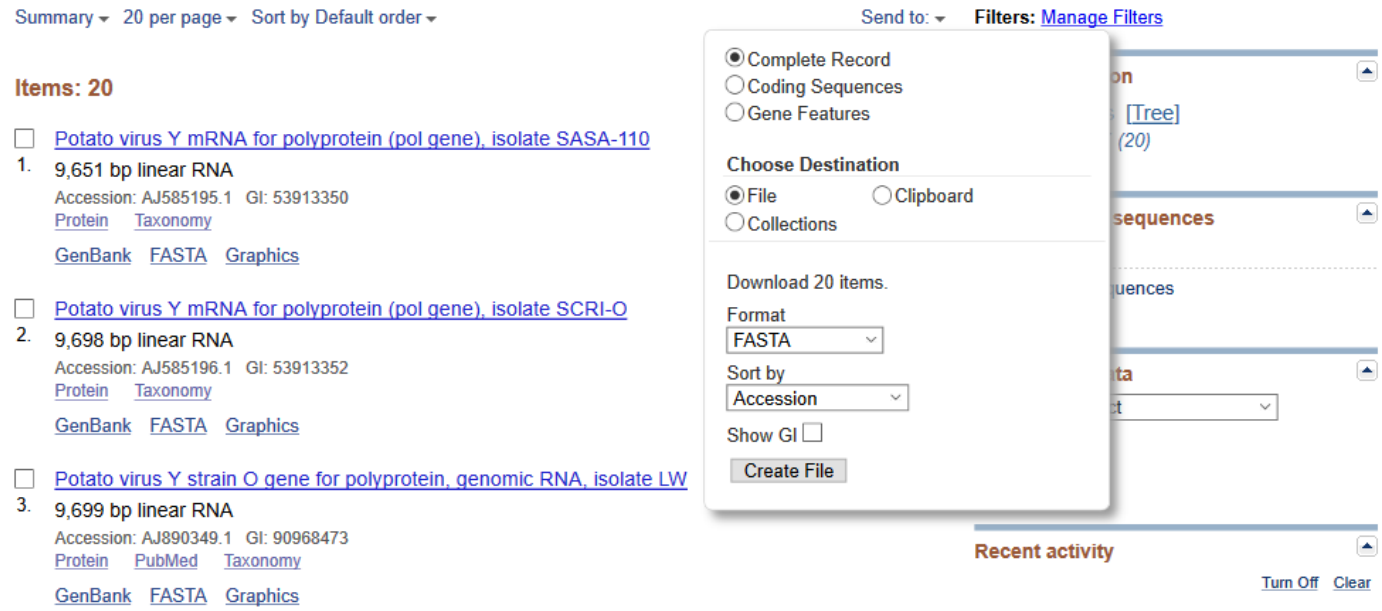


Figure 2.1. The NCBI Nucleotide web page with list of well-known PVY genomes. To download all sequences, the user should to click the "Send to:" menu, then to select "Complete record" and "File" destination, and choose the "FASTA" format with sorting by accession number. To save a text file, click on "Create file" button to save a text file.

```
>PVYO|AJ585195
AAACCAACGCAAAAACACTTATAAACGCTTATTCTCACTCAAGCAACTTACTAAGTTTCAGTTTAAATCA
TTTCCTTGCAATTCTCTTAAACAATATTGGAAACTATTTCAACTCAACAAGCAATTTCATCACTTCCAAC
CAATTTTAGATCCTCAATGGCAACTTACATGTCAACAATCTGTTTCGGTTCGTTTGAATGCAAGCTACCA
TACTCACCCGCCTGTTGCGGGCATATTGTGAAGGAACGAGAAGTGCTGGCTTCCGTTGATCCTTTCGCAG
ATCTGGAAACACAACTTAGTGCACGATTGCTCAAGCAAAAATATGCTACTGTTCGTGTGCTCAAGAACGG
TACTCTTACGTACCGATACAAGAATGATGCCCAGATAACGCGCATCCAGAAGAAACTGGAAAGGAAGGAT
AGGGAAGAATATCACTTCCAGATGGCAGCTCCTAGTATTGTGTCAAAAATTACTATAGCTGGTGGAGATC
CTCCATCAAAGTCAGAGTCACAAGCACCAAGAGGTGTCATTCATACAACTCCAAGGGTGCGTAAAGTCAA
GACACGCCCCATAACAAAGTTGACAGAAGGCCAGATGAATCATCTCATTAAGCAGGTGAAGCAGATTATG
TCGGAGAAGAGAGGGTCCGTCCACTTAATTAGTAAGAAGACCACTCATGTTCAATATAAGGAGATACTTG

...

TGAGTACCCGTTGAAACCAATTGTTGAGAATGCAAAACCAACCCTTAGGCAAATCATGGCACATTTCTCA
GATGTTGCAGAAGCGTATATAGAAATGCGCAACAAAAAGGAACCATATATGCCACGATATGGTTTAATTC
GAAATCTGCGGGATGTGGGTTTAGCGCGCTATGCCTTTGACTTCTACGAGGTCACATCACGAACACCAGT
GAGGGCTAGGGAAGCGCACATTCAAATGAAGGCCGCAGCATTGAAATCAGCCCAACCTCGACTTTTCGGG
TTGGACGGTGGCATCAGTACACAAGAGGAGAACACAGAGAGGCACACCACCGAGGATGTCTCTCCAAGTA
TGCATACTCTACTTGGAGTCAAGAACATGTGATGTAGTGTCTCTCCGGACGATATATAAGTATTTACATA
TGCAGTAAGTATTTTGGCTTTTCCTGTACTACTTTTATCATAATTAATAATCAGTTTGAATATTACTAAT
AGATAGAGGTGGCAGGGTGATTTCGTCATTGTGGTGACTCTATCTGTTAATTTCGCATTATTAAGTCTTA
GATAAAAGTGCCGGGTTGTCGTTGTTGTGGATGATTCATCGATTAGGTGATGTTGCGATTCTGTCGTAGC
AGTGACTATGTCTGAATCTATCTGCTTGGGTGGTGTTGTGATTTCGTCATGACAGTGACTG
```

Example 2.1. The single FASTA-file of PVY genome sequences with modified headers.

Step 4. Making multiple sequence alignment.

Due to the PVY genome size, it is possible to obtain the MSA of PVY genomes by using the Clustal Omega service (https://www.ebi.ac.uk/Tools/msa/clustalo) with the same parameters that shown in **Figure 1.1**.

```
CLUSTAL O(1.2.4) multiple sequence alignment


PVYO|AJ585195      ------------------------------AAACCAACGCAAAAACACTTATAAACGCTTA  31
PVYO|AJ585196      -AATTAAAACAACTCAATACAACATAAGAAAAACAACGCAAAAACACTCATAAACGCTTA  59
PVYO|AJ890349      AAATTAAAACAACTCAATACAACATAAGAAAAACAACGCAAAAACACTCATAAACGCT--  58
PVYO|DQ157178      ---------------------------------CAACGCAAAAACACTCATAAACGCTTT  27
PVYO|EF558545      -AATTAAAACAACTCAATACAACATAAGAAAAACAACGCAAAAACACTCATAAACGCTCA  59
PVYO|HQ912864      -AATTAAAACAACTCAATACAACATAAGAAAAACAACGCAAAAACACTCATAAACGCTTA  59
PVYO|HQ912865      -AATTAAAACAACTCAATACAACATAAGAAAAACAACGCAAAAACACTCATAAACGCTTA  59
PVYO|HQ912888      --------------------------------TCAACGCAAAAACACTCATAAAAGCTTA  28
PVYO|U09509        -AATTAAAACAACTCAATACAACATAAGAAAAACAGCGCAAAAACACTCATAAACGCTTA  59
PVYO|X12456        -AATTAAAACAACTCAATACAACATAAGAAAAACAACGCAAAAACACTCATAAACGCTCA  59
PVYN|AJ585197      ------------------------------AAATCAACGCAAAAACACTCACAAAAGCTTT  31
PVYN|AJ585342      -------------------------------------GCAAAAACACTCACAAAAGCTTT  23
PVYN|AJ889866      AAATTAAAACAACTCAATACAACATAAGAAAAACAACGCAAAAACACTCATAAACGCTTA  60
PVYN|AY166866      -AATTAAAACAACTCAATACAACATAAGAAAATCAACGCAAAAACACTCACAAAAGCTTT  59
PVYN|AY884982      -AATTAAAACAACTCAATACAACATAAGAAAAACAACGCAAAAACACTCACAAAAGCTTT  59
PVYN|AY884983      -AATTAAAACAACTCAATACAACATAAGAAAAACAACGCAAAAACACTCACAAAAGCTTT  59
PVYN|FJ666337      ------------------------------------------------------------  0
PVYN|HQ912869      --------------------------------TCAACGCAAAAACACTCATAAAAGCTTT  28
PVYN|M95491        AAATTAAAACAACTCAATACAACATAAGAAAATCAACGCAAAAACACTCACAAAAGCTTT  60
PVYN|X97895        AAATTAAAACAACTCAATACAACATAAGAAAATCAACGCAAAAACACTCACAAAAGCTTT  60

...

PVYO|AJ585195      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGAATCTATCTGCTT  9617
PVYO|AJ585196      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTACTT  9645
PVYO|AJ890349      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTACTT  9646
PVYO|DQ157178      ATCGATTAGGTGGTGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCCATCTGCTT  9617
PVYO|EF558545      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTACTT  9645
PVYO|HQ912864      ATCGATTAGGTGATGTCGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTGCTT  9645
PVYO|HQ912865      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTGCTT  9645
PVYO|HQ912888      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTGCTT  9614
PVYO|U09509        ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTGCTT  9645
PVYO|X12456        ATCGATTAGGTGATGCTGTGATTCTGTCATAGCAGTGACTATGTCTGGATTTAGTTACTT  9651
PVYN|AJ585197      ATCGATTAGGTGATGTTGCGA-TTTGTCGTAGCAGTGACTATGTCTGGATTTAGTTACTT  9619
PVYN|AJ585342      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTGCTT  9613
PVYN|AJ889866      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTACTT  9646
PVYN|AY166866      ATCGATTGGGTGATGTTGCGA-TTCGTCGTAGCAGTGACCATGTCTGGATTTAGTTACTT  9647
PVYN|AY884982      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTGCTT  9649
PVYN|AY884983      ATCGATTAGGTGATGTTGCGA-TTTGTCGTAGCAGTGACTATGTCTGGATTTAGTTACTT  9647
PVYN|FJ666337      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTACTT  9558
PVYN|HQ912869      ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTGCTT  9618
PVYN|M95491        ATCGATTAGGTGATGTTGCGATTCTGTCGTAGCAGTGACTATGTCTGGATCTATCTGCTT  9650
PVYN|X97895        ATCGATTAGGTGATGTTGCGA-TTTGTCGTAGCAGTGACTATGTCTGGATTTAGTTACTT  9648
                   ******* **** **  * ** *  *** ********** ******* **  *  * ***

PVYO|AJ585195      GGGTGGTGTTGTGATTTCGTCATGACAGTGACTG-------------------  9651
PVYO|AJ585196      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9698
PVYO|AJ890349      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9699
PVYO|DQ157178      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9670
PVYO|EF558545      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9698
PVYO|HQ912864      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9698
PVYO|HQ912865      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9698
PVYO|HQ912888      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9667
PVYO|U09509        GGGTGATGTTGTGATTTTGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9698
PVYO|X12456        GGGTGATGCTGTGATTCTGTCATAGCAGTGACTGTAAACTTCAATCAGGAGAC  9704
PVYN|AJ585197      GGGTGATGCTGTGATTCTGTCATAGCAGTGACTG-------------------  9653
PVYN|AJ585342      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTG-------------------  9647
PVYN|AJ889866      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAG--  9697
PVYN|AY166866      GGGTGATGCTGTGATTCTGTCATAGCAGTGGCTGTAAACTTCAATCAGGAGAC  9700
PVYN|AY884982      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9702
PVYN|AY884983      GGGTGATGTTGTGATTCTGTCATAGCAGTGACTGTAAACTTCAATCAGGAGAC  9700
PVYN|FJ666337      GGGTGGTGTTGTGATTTCGTCATAA---------------------------  9583
PVYN|HQ912869      GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9671
PVYN|M95491        GGGTGGTGTTGTGATTTCGTCATAACAGTGACTGTAAACTTCAATCAGGAGAC  9703
PVYN|X97895        GGGTGATGCTGTGATTCTGTCATAGCAGTGACTGTAAACTTCAATCAGGAGAC  9701
                   ***** ** *******  *****
```

Example 2.2. Multiple sequence alignment of PVY strains that belong to different serotypes.

The processing of such large nucleotide sequences like viral genomes is required long computation time, therefore, the Clustal Omega user can to obtain the alignment by email after few minutes or hours. The computation time usually depends from number of strains and their sequence length. The user should to save the alignment as text file with *.txt extension.

Step 5. Masking serotype-specific mutations.

Here, we again suggest to use the MorphoCatcher service (http://morphocatcher.ru) for the MSA processing. To visualize genome region with high density of serotype-specific mutations, the sliding-window plot of average mutation index should be obtained (**Figure 2.2**).

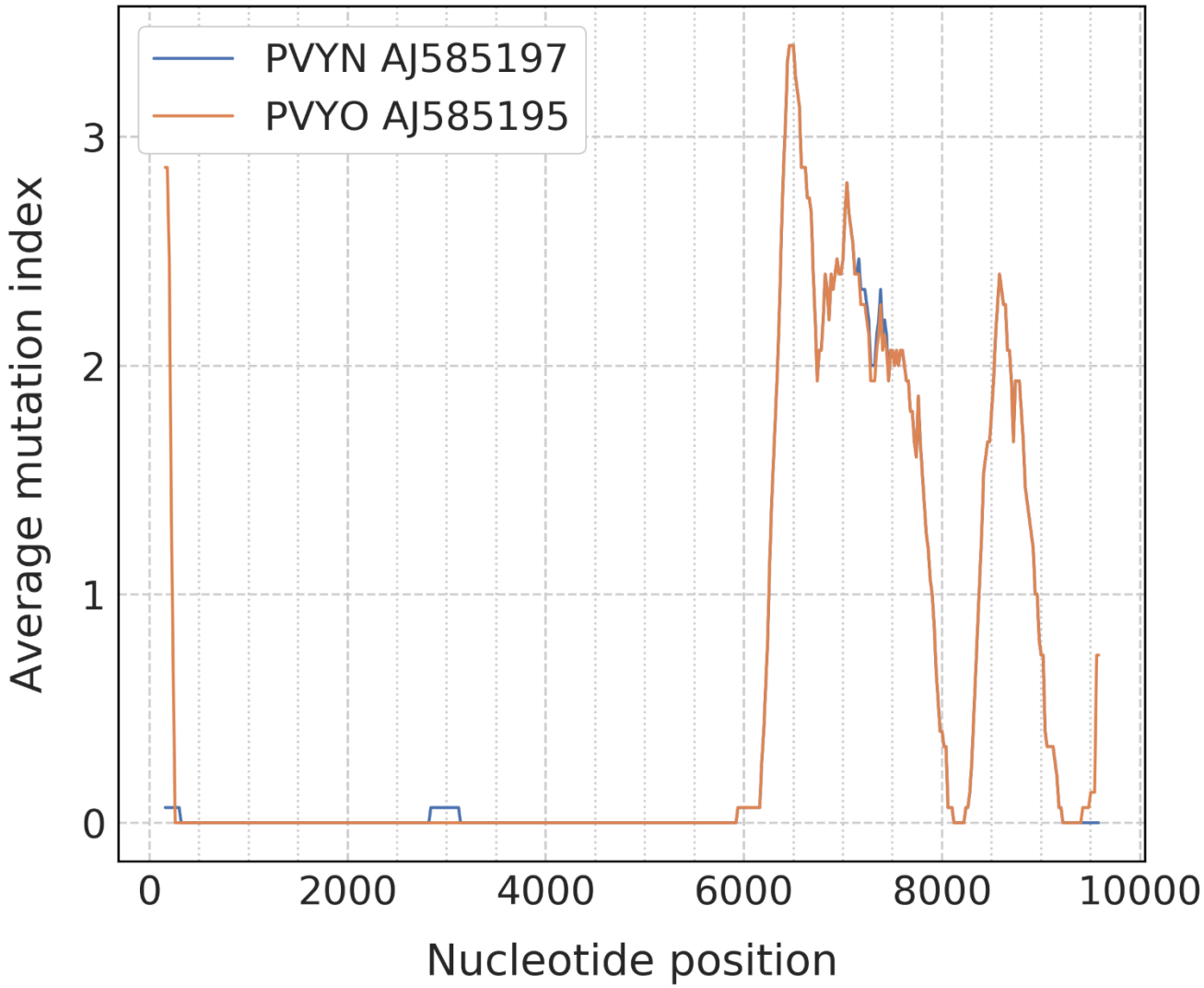


Figure 2.2. The alignment-based average mutation index plot for two PVY serotypes.

Step 6. Designing specific primers.

Finally, the user should to upload corresponding text file from the MorphoCatcher output (see files from the "Input for PrimerExplorer" column) as the input file to the PrimerExplorer service (http://primerexplorer.jp/e) for designing of serotype-specific primer set.

Here, the fragment of primer set list is shown (**Figure 2.3**) to demonstrate how different serotype-specific mutations can be covered by the primer ends of various primer sets.

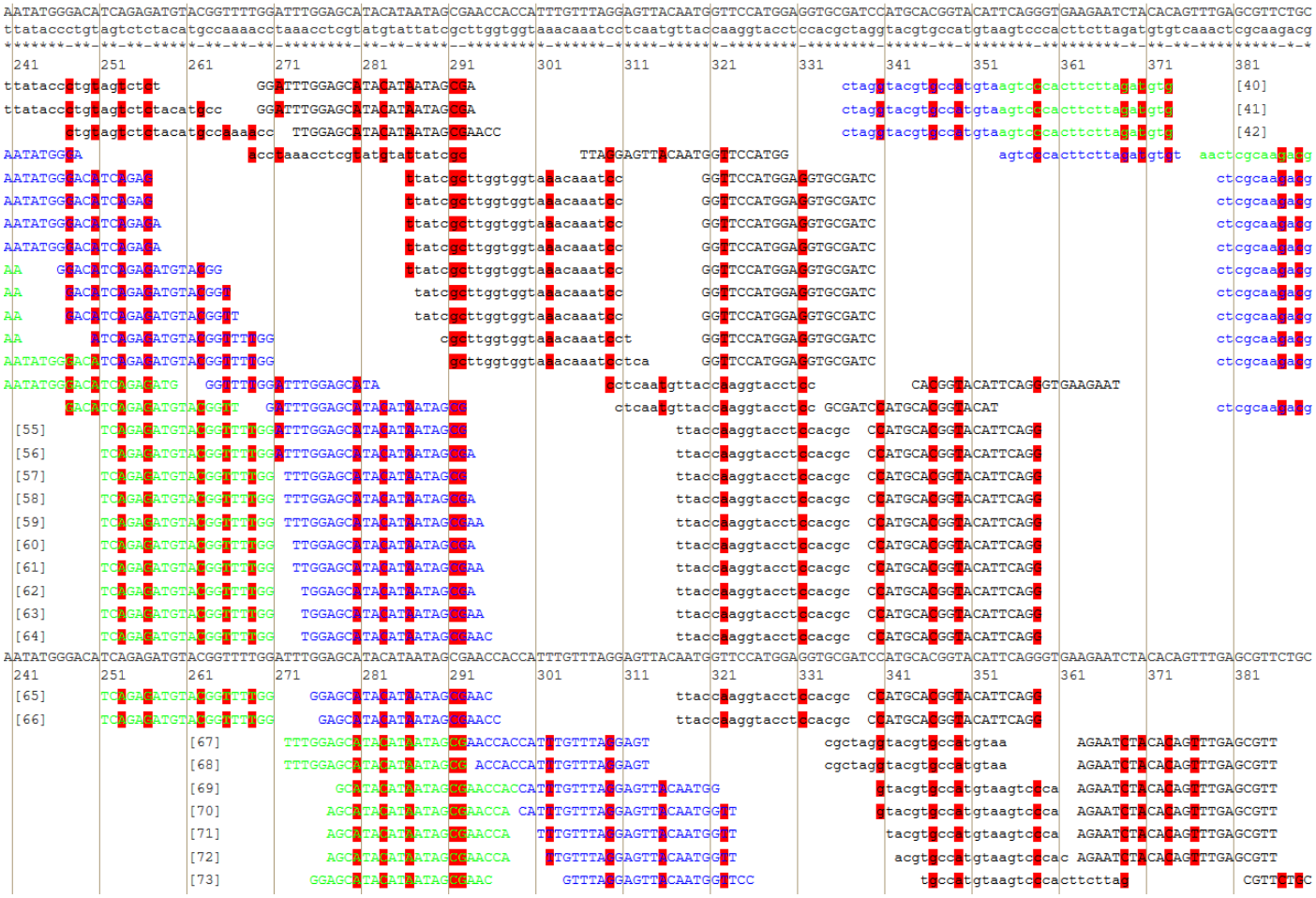


Figure 2.3. The PrimerExplorer output page with full view of designed primer sets. Specific mutations for PVY strains of O-serotype are highlighted by red color. The same mutations can be covered by oligos from different primer sets.

Credits

We are grateful to the "Start Bootstrap" service (https://startbootstrap.com) for providing a free "Agency" template of HTML5 page. Some graphics for web design was provided by contributors of the "Noun Project" collection of icons (https://thenounproject.com) — we thank Danil Polshin ("Abstract" icon), Wes Breazell ("DNA" icon), Alexander Blagochevsky ("Pedigree Chart" icon), Ervin Bolat ("Text Editor" icon), Cris Dobbins ("Sliders" icon), and ProSymbols ("Idea" icon).

References

1. PubMed ID: 25684775;
2. PubMed ID: 29119360;
3. PubMed ID: 21675922.