

## **User manual**



**Version 1.15**

## Table of Contents

1. Introduction.....	3
1.1 What is Genome ARTIST ?.....	3
1.2 Reporting Bugs.....	4
2. Installing and Running.....	4
2.1 Requirements.....	4
2.2 Installation.....	4
2.3 Running.....	4
3. Loading data.....	5
3.1. Acquiring genomes.....	5
3.2. Ready-for-use genome files.....	6
3.3. Transposon files.....	8
5. Reading the results.....	11
5.1 Query Info .....	12
5.2 Best result .....	13
5.3 Results table .....	13
5.4 Result panel .....	14
5.5 Gene Map.....	15
6. Saving and loading.....	16
7. Advanced Settings.....	16
7.1 The search algorithm .....	16
7.2 Parameters .....	17
8. License Terms.....	18
9. Authors.....	18
10. Acknowledgments.....	18

# 1. Introduction

## 1.1 What is Genome ARTIST ?

**Genome ARTIST** (ARtificial Transposon Insertion Site Tracker) is a new bioinformatics tool originally developed in order to allow a rapid detection of insertional mutations generated in the genome of *Drosophila melanogaster* by means of artificial *P* element derivatives. Aside from the large gene disruption projects (FlyBase, [www.flybase.org](http://www.flybase.org)), many fly laboratories run small scale transposon mutagenesis screenings. Basically, mobilization with a transposase source of artificial molecular constructs (derived from a natural *P* mobile element or from other transposons) induces insertional mutations in the germline. Many different mutant strains are derived from affected parents using classical genetic crosses and, in the end, their putative useful mutations are analyzed by *inverse PCR* and sequencing. The sequencing product is a mixture of information, where part of it pertains to the fruit fly canonical genome and the rest of it belongs to a specific artificial element. The most critical aspect of sequence analysis is to detect the exact border between the genomic and transposon DNA, equivalent with identification of the insertion site at the nucleotide level. Sequencing products are not always perfect and a few artifact bases mismatches may impair a fluent insertion mapping. Most commonly, the sequences of interest are aligned with BLAST (<http://blast.ncbi.nlm.nih.gov>) or BLAT (<http://www.genome.ucsc.edu>) against *D. melanogaster* official genome, which do not contain neither natural nor artificial *P* transposons. Often, additional manual sequence annotation is needed in order to finish an accurate insertion mapping and here is when **Genome ARTIST** enters the scene and offers a bit of help. The query sequence is simultaneously compared offline against both the *D. melanogaster* genome and the specific transposon sequence, partial sequence alignments are matched to each other, relative scores of alignments are calculated and the best mixt sequence with the genomic and transposon coordinates is offered to the user. Different colors are used for genomic versus transposon fragments, and an intuitive list of results and details is also depicted. One may easy observe the absolute site of insertion according to Release 5.44 (available on *FlyBase* at [www.flybase.org](http://www.flybase.org)), the gene affected by the transposon insertion, and also the genes located in the close vicinity of the insertion. Special biological conditions occurring during mutagenesis experiments, as transposon reinsertions into the original mobile element copy, are not usually detected with other searching algorithms, therefore **Genome ARTIST** is designed to reveal and to interpret such events.

To some extent, **Genome ARTIST** is an alternative for the classical alignment algorithms and may be exploited for checking the specificity of short sequences as primers or probes. Last but not the least, aficionados of different model organisms may use the abilities of **Genome ARTIST** by loading other genomes and/or specific transposons. The performances of **Genome ARTIST** were tested on the genomes of *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila pseudoobscura*, *Ciona intestinalis*, *Danio rerio*, *Gallus gallus* and *Anolis carolinensis* and offers similar results to those obtained on *D. melanogaster*. Additionally, pairwise comparative alignments may be performed among sequences pertaining to various species aside from *D. melanogaster*, allowing the identification of structural orthologous.

## 1.2 Reporting Bugs

So... the inherent bugs... We would be happy to receive any feed-back about such annoying things in order to fix them as soon as possible. Please send any comments at [alexandruecovoio@yahoo.com](mailto:alexandruecovoio@yahoo.com). Thank you in advance!

## 2. Installing and Running

### 2.1 Requirements

#### a. Hardware

##### RAM memory

- 512 MB should be available in order to use only the *D. melanogaster* genome
- For any additional genome files loaded, the RAM requirements will increase with about the size of the loaded file + 50% of its size

##### DISK memory

- 700 MB if you plan to use only the *D. melanogaster* genome
- For any additional genome files loaded, the space requirements will increase with approximately (3 \* SIZE\_OF\_LOADED\_GENOME\_FILE)

##### Processor

- The lowest performance processor the program has been tested with is an Intel Atom 1 GHz

GA was designed for a 32-bit architecture but users of a 64-bit OS version may employ the software by installing the package *ia32-libs*. One can install this package by running the command `sudo apt-get install ia32-libs`.

#### b. Software

##### Operation system

- Linux
- It has been tested on Ubuntu with either *Gnome* or *Unity* desktop environment and also on Linux Mint with similar performances.

##### Other dependencies

- JAVA JRE v1.6 (sometimes called v6)

### 2.2 Installation

The program comes as a .zip archive that you must extract in any folder easy to be accessed. Aside the extraction, there is no other necessary action to be done.

### 2.3 Running

To start the program, you must run the script "Genome-ARTIST.sh" (which should be marked as an executable when imported from an external storage device if Ubuntu versions higher than 10.04 are used as OS).

## 3. Loading data

### 3.1. Acquiring genomes

To load a chromosome, the application needs a .raw file containing only the raw sequence and a Fasta file containing the annotations. On the project's website ([www.genomeartist.ro](http://www.genomeartist.ro)) you can find bundles containing *Drosophila melanogaster* genome or the genomes of some well know model organisms. These genomes can be extracted and loaded using the „Add folder” feature of our program.

However if the genomes you find on our website are not relevant for you, you can forge your own **Genome ARTIST** friendly genomes. We support loading genomes of *D. melanogaster* and *Drosophila pseudoobscura* from *FlyBase* and also converting genomes from *Ensembl* ([www.ensembl.org](http://www.ensembl.org)).

#### a. Loading data from *FlyBase*

[1] <ftp://ftp.flybase.net/genomes/>

- From *FlyBase* FTP genome repository [1] you should download the .raw files and the .fasta files for the genome of your choice
- Place all the downloaded file into one single folder, and extract all the archives
- Rename the .raw files (example 2L.raw)
- Rename the annotation fasta file using the following syntax: <chromosome name>\_gene.fasta (example 2L\_gene.fasta). The term “\_gene.fasta” is mandatory
- Load the genome into **Genome ARTIST** using the “Settings > Add Folder” path and press OK to finish the uploading. Completion of the task will take between few minutes for invertebrates and a few tens of minutes for vertebrates, pending also on the hardware performances.

#### b. Loading data from *Ensembl*

[2] <http://www.ensembl.org/info/data/ftp/index.html>

- From *Ensembl* FTP repository [2] you should download the following files for a given genome:
  - From “DNA sequence > FASTA” you should download all the chromosome's sequences from the beginning of the list till the folder dna.toplevel.fa.gz which should not be uploaded. For example, for *D. melanogaster* all the folders starting with *Drosophila\_melanogaster.BDGP.66.dna.chromosome.2L.fa.gz* and ending with *Drosophila\_melanogaster.BDGP.66.dna.chromosome.dmel\_mitochondrion\_genome.fa.gz* should be downloaded. If one is interested in the masked version of the genome, then the folders from *Drosophila\_melanogaster.BDGP.66.dna\_rm.chromosome.2L.fa.gz* and ending with *Drosophila\_melanogaster.BDGP.66.dna\_rm.chromosome.dmel\_mitochondrion\_genome.fa.gz* should be downloaded. Enough time should be allowed for the output folder to be completed.
  - From “Annotated sequence -> EMBL” you should download all the archives
- Place all downloaded files into a single folder
- Extract all archives

- From **Genome ARTIST** install folder, copy the two script files found in ./scripts/Ensembl/ into the folder containing the downloaded files
- Ensure the two script files are marked as executable (chmod +x)
- Run the “parse\_ensemble.sh”
- After the script is executed, a folder named “output” will be generated. This folder contains all the required files for loading the genome into **Genome ARTIST**
- Load the genome into **Genome ARTIST** using the “Settings > Add Folder” option then press OK to finish uploading

### c. Loading data from NCBI

[3] <ftp://ftp.ncbi.nlm.nih.gov/genomes/>

- From NCBI FTP repository [3] one should download the following files for a genome of interest:
  - if the genome has more than one chromosomes, then for each chromosome one should download the file ending with extension “.gbk”.
  - if the genome of interest contains only one chromosome (as in bacteria), the one should download the file ending with “.gbk”
- Place all downloaded files into a single folder
- From **Genome ARTIST** install folder, copy the three script files found in ./scripts/NCBI/ into the folder containing the downloaded chromosome files
- Ensure the three script files are marked as executable (chmod +x)
- Run the “parse\_ncbi.sh” file
- After the script is executed, a folder named “output” will be generated. This folder contains all the required files for loading the genome and the annotations into **Genome ARTIST**
- Load the genome the genome of interest into **Genome ARTIST** using the “Settings > Add Folder” option then press OK to finish uploading

**NOTE: The scripts described above work only for some of the genomes available in the FTP repository of NCBI.**

## 3.2. Ready-for-use genome files

### a. Loading a whole genome

At [www.genomeartist.ro](http://www.genomeartist.ro) you may find the ready-for-use *Ensembl*-derived genome files for *D. melanogaster*, *D. pseudoobscura*, *S. cerevisiae*, *C. elegans*, *C. intestinalis*, *D. rerio*, *A. carolinensis* and *G. gallus* experimental models. You just have to extract the archive of interest in a folder of your choice. Then open the Settings Panel in Genome ARTIST and under Genome files you will find the button „Add folder”. You will be prompted for a name and a folder. This name will be used as a suffix for naming the loaded chromosomes (example dmel\_2L, dmel\_3L, etc.). You then need to select the folder that appeared after extracting the genome archive. After choosing the folder and the name, just press OK and the loading of the genome into the application should start. Warning: This will take some time depending on the size of the genome and the computer's performances. After loading a genome you may co-load a

different genome if you are interested in comparative sequence analysis, following the same steps as above. We offer a package containing the genomes of *D. melanogaster* and *D. pseudoobscura* and a package containing the genomes of *S. cerevisiae*, *C. elegans*, *D. melanogaster* and *C. intestinalis*. For running such packages 1 GB of RAM are enough, but for genomes of the mentioned vertebrates at least 3 GB of RAM are required.

## b. Loading a chromosome (genome file)

Instead of using “Add folder” option in order to load a whole genome, one may load only one or a few chromosomes (here regarded as genome files) step by step. Loading (even an entire genome) file by file is advantageous since the right association between a sequence and its annotation is manually performed, therefore is not necessary to adjust the names of the downloaded .raw and .fasta files. The names may be kept unchanged, as in the host database.

To load a genome file, access the “Settings” menu and select the “Genome files” tab. Once there, press the “Add file” button and the necessary information will be prompted:

- “Name”: enter the name for the new genome data source
- “Sequence location (.raw)”: press the browse button on the right of the field and select the file containing the raw data (a continuous sequence of nucleotides); only A, a, C, c, G, g, T, t characters are permitted inside the file since any other characters (like the space character, or new line) will cause the file to be processed in a wrong manner
- “Genes location (.fasta)”: press the browse button on the right of the field and select the file containing the genes information corresponding to the .raw file, written in the .fasta format.
- The first two fields are required, and the third field is optional.

Press the “Ok” button (see Fig. 1) and the program will process your input to create the new genome data source. WARNING: creating the new data source may take around 30 seconds (for a normal *D. melanogaster* genome file), as it requires a lot of processing.

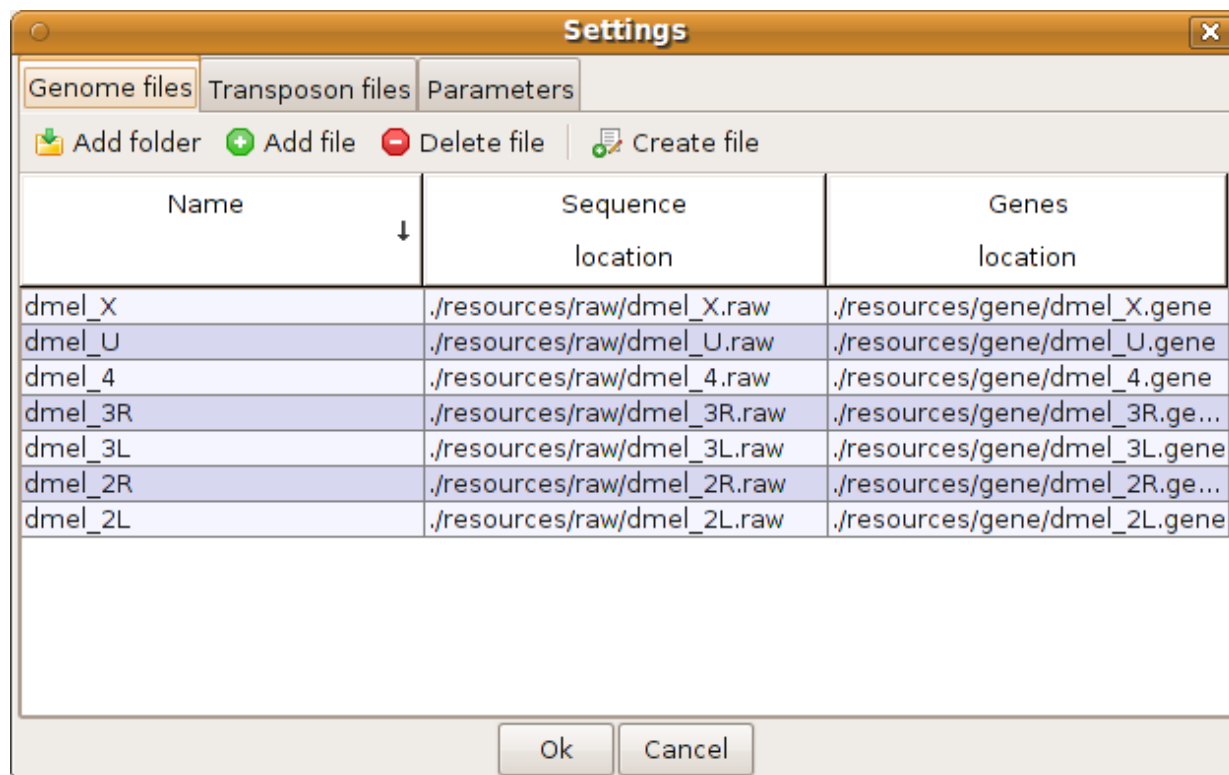


Fig. 1 Pressing Ok will load all the genome files (chromosomes) in RAM memory.

### c. Creating a new genome data source directly inside the program:

To create a new genome data source, you must access the “Settings” menu, and select the “Genome files” tab. Once there, press the “Create file” button and you will be prompted for the necessary information:

- “Name”: enter the name for the new genome data source
- “Sequence”: enter the nucleotide sequence; here you can enter any character, but only A, a, C, c, G, g, T, t, N, n characters will be kept after processing (so you can copy/paste from another location that contains spaces, new lines, or any other unnecessary character)

Press the “Ok” button and the program will process your input to create the new genome data source.

### d. Deleting a genome data source

To delete a genome data source, you must access the “Settings” menu, and select the “Genome files” tab. Once there, select the source you want to delete and press the “Delete file” button. You can also select a single chromosome from a loaded genome then press “Delete file”.

## 3.3. Transposon files

### a. Loading an existing transposon file

To load a new transposon file, you must access the “Settings” menu, and select the “Transposon files” tab. Once there, press the “Add file” button and you will be prompted for the necessary information:

- “Name”: enter the name for the new transposon data source
- “Sequence location (.raw)”: press the browse button on the right of the field and select the file containing the raw data (a continuous sequence of nucleotides); only A, a, C, c, G, g, T, t characters are permitted inside the file; any other characters (like the space character, or new line) will cause the file to be processed in the wrong way
- “Genes location” (.fasta): press the browse button on the right of the field and select the file containing the genes information for the .raw file, written in the .fasta format.
- The first two fields are required, and the third field is optional.

Press the “Ok” button (Fig. 2) and the program will process your input to create the new transposon data source.

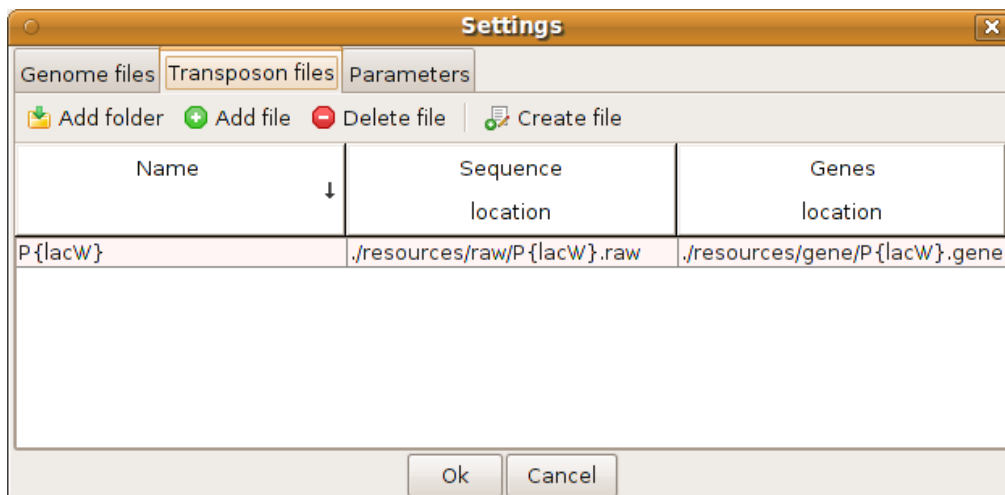


Fig. 2 Loading a transposon file.



### **b. Creating a new transposon data source directly inside the program:**

To create a new transposon data source, you must access the “Settings” menu, and select the “Transposon files” tab. Once there, press the “Create file” button and you will be prompted for the necessary information:

- “Name”: enter the name for the new transposon data source
- “Sequence”: enter the nucleotide sequence; here you can enter any character, but only A, a, C, c, G, g, T, t, N, n characters will be kept after processing (so you can copy/paste from another location that contains spaces, new lines, or any other unnecessary character)

Press the “Ok” button and the program will process your input to create the new transposon data source.

### **c. Deleting a transposon data source**

To delete a transposon data source, you must access the “Settings” menu, and select the “Transposon files” tab. Once there, select the source you want to delete and press the “Delete file” button.

## **4. Launching a query**

To launch a query, press the “New Search” button on the main interface, and you will be prompted for the necessary information:

- “Name”: enter the name for the new query
- “Query”: enter the query you want to run;

Accepted input:

- You can simply enter the continuous nucleotide sequence
- You can enter the query in the *GenBank* or in Fasta format
- **Actually, any format, regardless of fonts and colours may be employed.**

We present here a real case obtained in our laboratory of a *P{lacW}* insertion close to *pyd* gene from *D. melanogaster*. In Fig. 3 is depicted the sequence obtained, consecutive to inverse PCR, with the primer *Sp1* (<http://www.fruitfly.org/about/methods/inverse.pcr.html>) and in Fig. 4 the reverse complement sequence (simply obtained by checking the dedicated button) is shown. Sometimes reverse complement sequence is more intuitive to analyze relative to the reference strand of the genome.

**New Search**

Query

Query name: query\_001

Query content:

```

ATTTAAGTGTATACTTCGGTAAGCTTCGGCTATCGACGGGACCACCTTATGTT
ATTTTCATCATGCTCAGTCGGTTCAGATTATCGCGCTTGTGCGGTTGTGCGGAG
CGGACGAGCTGAAGTGGCCGAGTCGTGAACTTGAAATCTATACAGGCGTTTT
AAACATAAAACAAACAAATACGAATGCGAAAGAGCCGGTAAAAGTTTAAAT
GTTTGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGAT
GTGCTGCAAGGCGATTAAGTTGGGTAACCGCCAGGGTTTCCCCAGTCACGTA
CGTTGTAAAACGACGGCCAGTGCCAAGCTCTGGCTGCTCTAAACTACGCATTT
CGTACTCCAAGTACGAATTTTTCCTCACGCCTCTTATNCATTAAACATGAAC
GGACCCTACCGCACAGTAG

```

☐ Reverse complement query

Search Cancel

Fig. 3 The original sequence obtained with the primer Spl.

**New Search**

Query

Query name: query\_001

Query content:

```

CTACTGTGCGGTAGGGTCCAGTTCATGTTAATGNATAAGAGGCGTGAGGAAAAATTC
GTAAGTTGGAGTACGAAATGCGTAGTTTAGAGCAGCCAGAGCTTGGCACTGGCCGTCG
TTTTACAACGTACGTGACTGGGGAAACCCTGGCGGTTACCCAACTTAATCGCCTTGCA
GCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCAAACA
TTTAACTTTTACCGGCTCTTTCGCATTTCGATTTGTTTGTATGTTTTAAACGCCTGTA
TAGATTTCAAGTTCACGACTCGGCCACTTCAGCTCGTCCGCTCCGCACAACCGCACAA
GCGCGATAATCTGAACCGACTGAGCATGATGAAATAACATAAGGTGGTCCCGTCGAT
AGCCGAAGCTTACCGAAGTATACACTTAAAT

```

☒ Reverse complement query

Search Cancel

Fig. 4 The reverse complement sequence of pyd associated insertion.

The program will strip away any unnecessary information (like spaces, new lines, GenBank or FASTA additional information), and keep only the relevant characters: A, a, C, c, G, g, T, t, N, n.

After you have entered all the necessary information, press “Search” button and the program will launch the query. While the query is processed, a message will inform you that the query is running and that you must wait. After the program computes the results, they will be show in a new tab, in the main interface.

WARNING: a normal query (100-500 nucleotides) will take up to 1 minute to compute, on a system with a Dual Core 2 GHz processor (the system must also correspond to the other hardware requirements).

## 5. Reading the results

The results window has 3 sections as presented in Fig. 5:

- Query Info
- Best result
- Results table

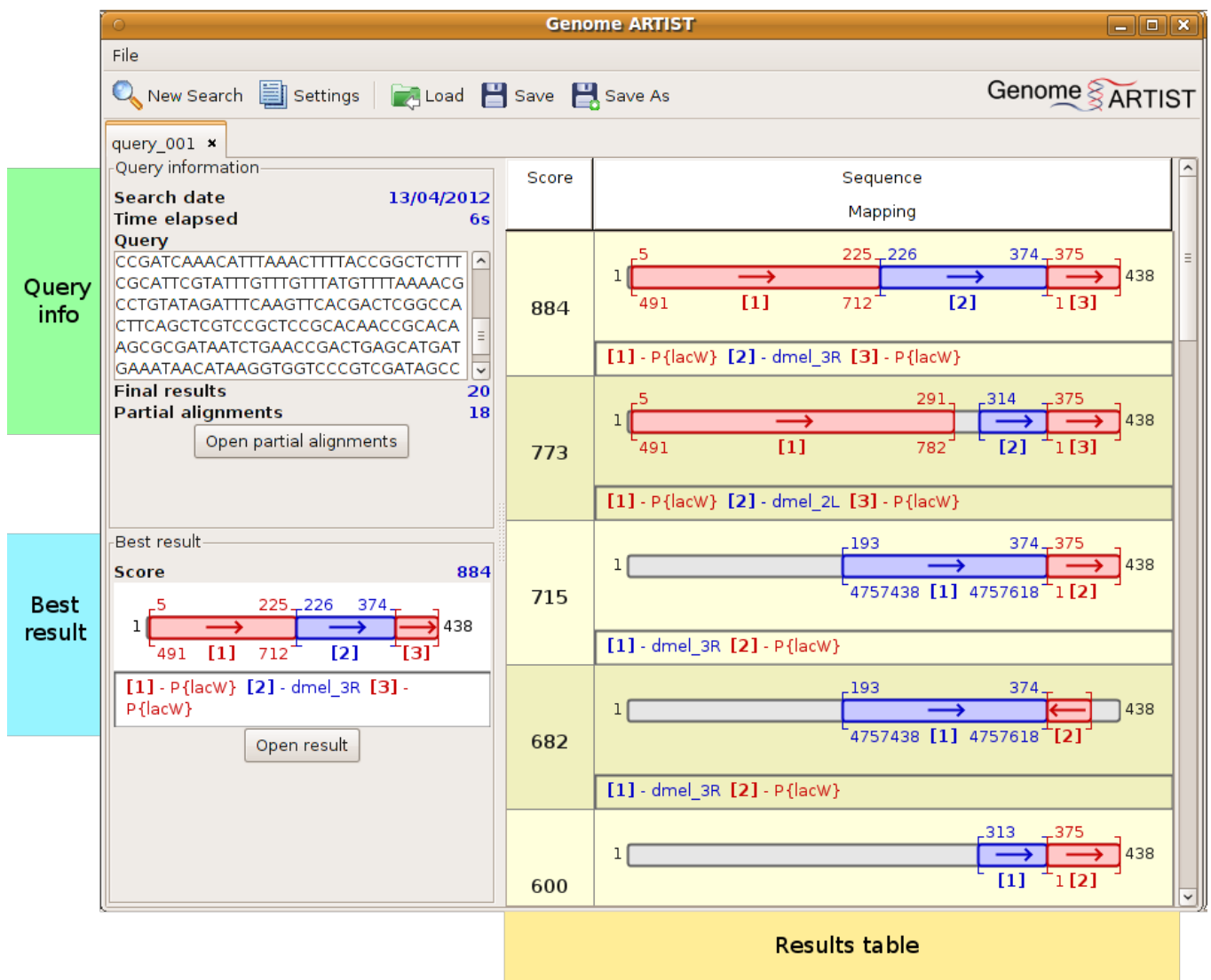


Fig. 5 Details of the results window.

### 5.1 Query Info

At the top left you can find summary information about the query. This shows the context in which the query has ran.

The summary fields are:

- Search Date - the date the search was first launched
- Time elapsed - the time required for the search to take completion
- Query - The sequence that was processed during the search
- Results found - The number of results that have been found for this search

NOTE: the maximum number of result can be changed from Settings

- Partial alignments found - The number of partial alignments which were used to assemble the final results may be seen by pressing button “Open partial alignments” (Fig. 6)

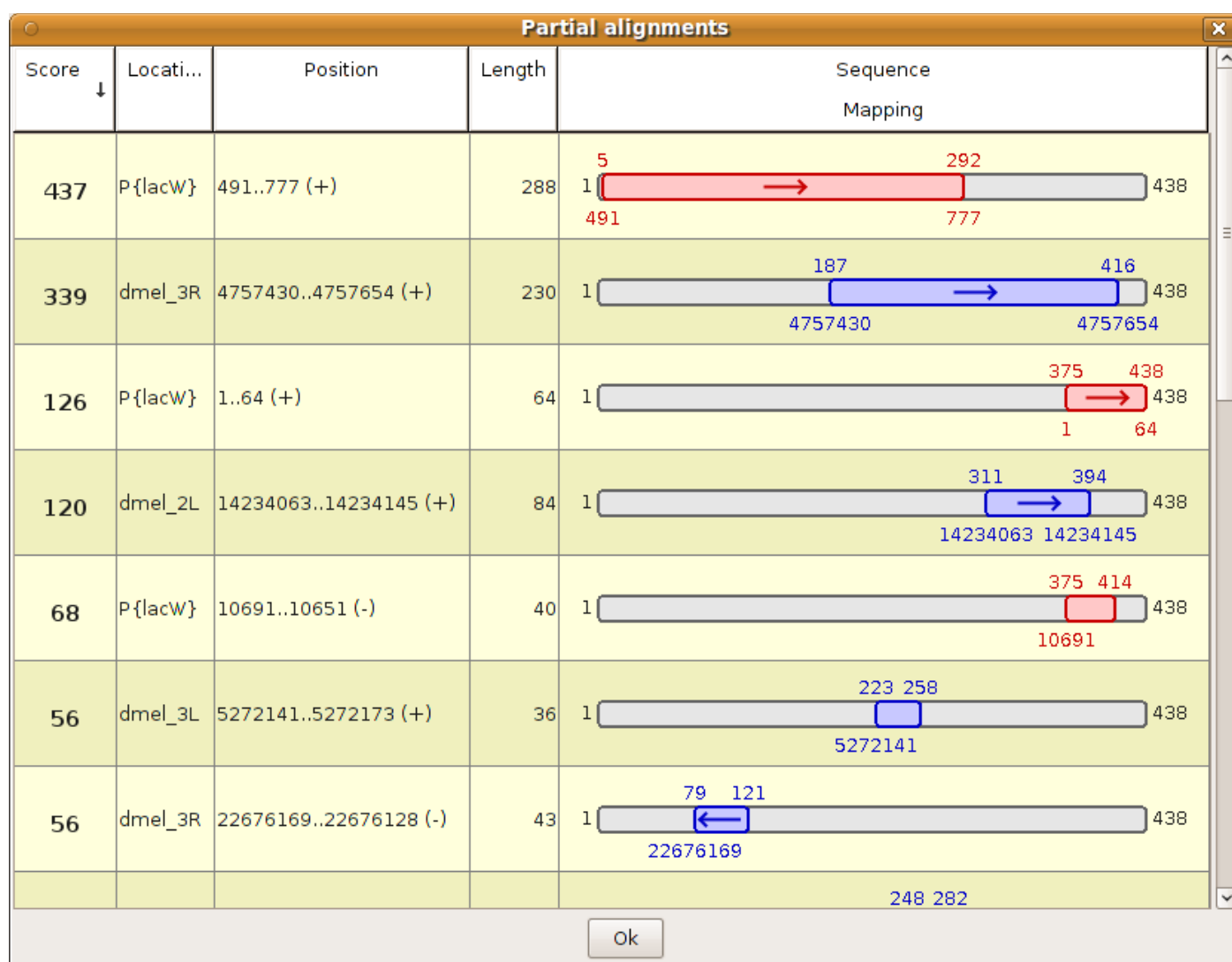


Fig. 6 A list of partial alignments of the query with either genomic (blue) or transposon (magenta) sequences.

## 5.2 Best result

This panel shows the supposed best result. If by examining the results you find that another result is biologically the best, you can mark it accordingly as “Best result” and it will be shown in the panel.

### **5.3 Results table**

The results table shows the best matches of the given query to the files stored in database. A result contains “Score” and “Sequence mapping”. The score is proportional to the number of nucleotide that had perfectly matched the query.

The “Sequence Mapping” represents how the genome and transposon sequences matched the query. Genome files and transposon files are differentiated by color (genome sequences are depicted in blue and transposon sequences are shown in magenta). The numbers above the mapping represent the position in the query, while the number below represents the position in the mapped files. Under the mapping there is a legend specifying the files which were mapped and the number associated with it. Numbers are used to visually identify the sequences. The strand with which the query matched it is also indicated. This is shown by an arrow on the given query sequence. If the arrow is pointing to the right, the sequence matches the forward (or reference) genomic strand. If the arrow is pointing to the left, then the sequence matches the reverse (reverse-complement) genomic strand.

This panel shows the minimum amount of information for each query. To find more detailed information about one particular result you can double-click on it and a new window will appear.

### **5.4 Result panel**

This panel offers several details about a particular result.

It is composed from 3 sections:

- Sequence mapping
- Sequence alignment
- Annotations

The “Sequence mapping” is the same as in the results table and shows how specific subsequences have matched the given query. The Sequence alignment offers details at a nucleotide level.

In the “Annotations” section, the alignments are described along with the nucleotide coordinates pertaining to the genome/transposon. If more genes are in the area, only one is shown. If you need more information about a specific gene or there are more genes that intersect the sequence, you can access the “Gene map” by double-clicking the “Annotation” field which you want to visualize.

From this panel you are given the option to export or print the result. Another option is to set the any particular result as “Best result”. The result will be shown in the “Best result” panel in the main query window. Using this panel, you can locate the exact position of the transposon insertion (Fig. 7).

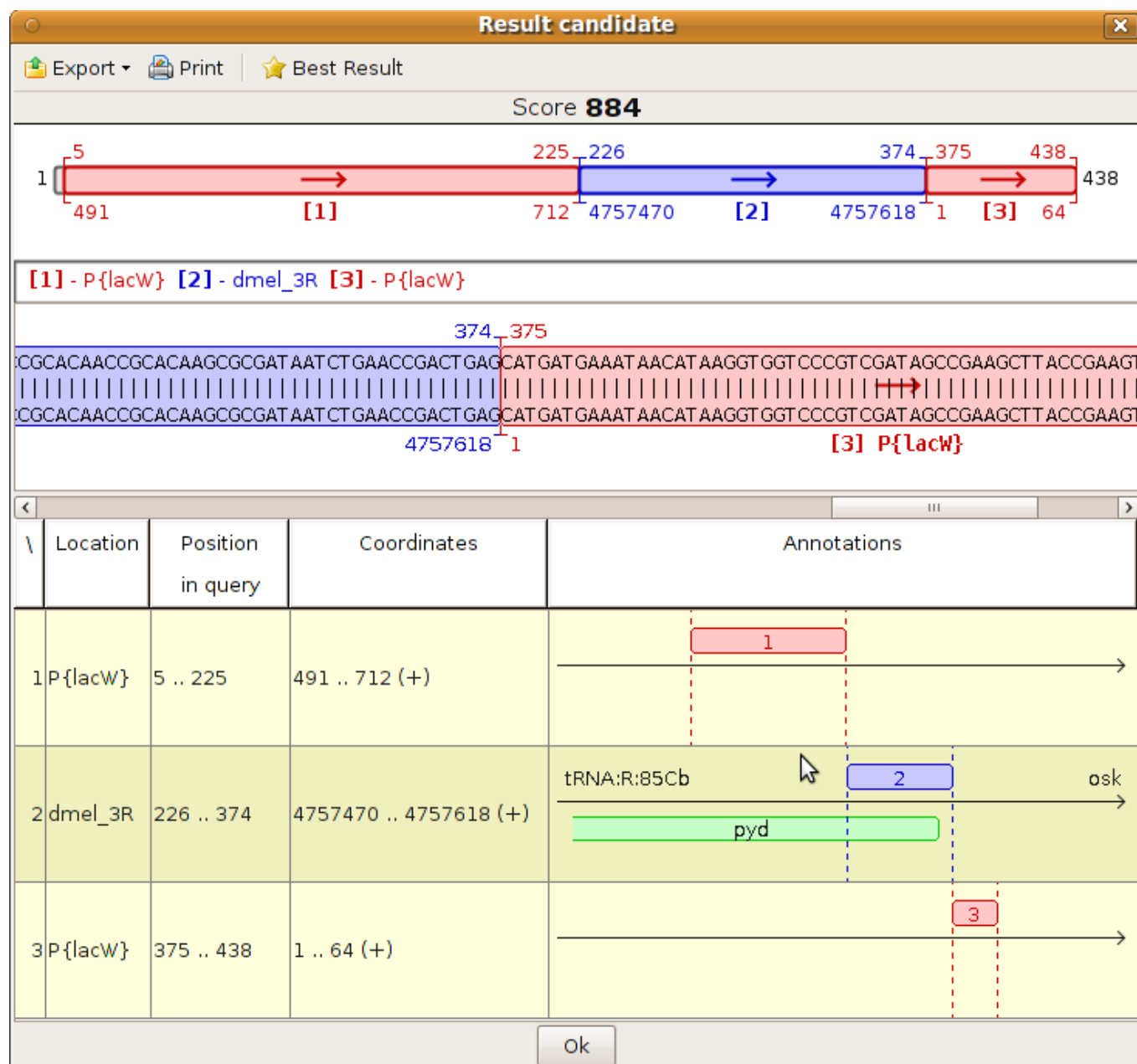


Fig. 7 In the result table panel one may notice that P{lacW} insertion is located close to pyd gene, at nucleotide 4757618 (release 5.44); nucleotide 1 of P{lacW} artificial transposon (www.flybase.org) coincides with nucleotide 375 of the query.

## 5.5 Gene map

The “Gene map” panel shows how an aligned sequence is located relative to the genes loaded for the corresponding file. The arrow shows the relative orientation of the query sequence. At the ends of the arrow the names of the closest upstream and downstream genes are depicted. Under the graphical view there is a table that briefly describes the genes, their cytological map, the absolute genomic coordinates and location of their sense strands relative to the reference strand of the genome (Fig. 8).

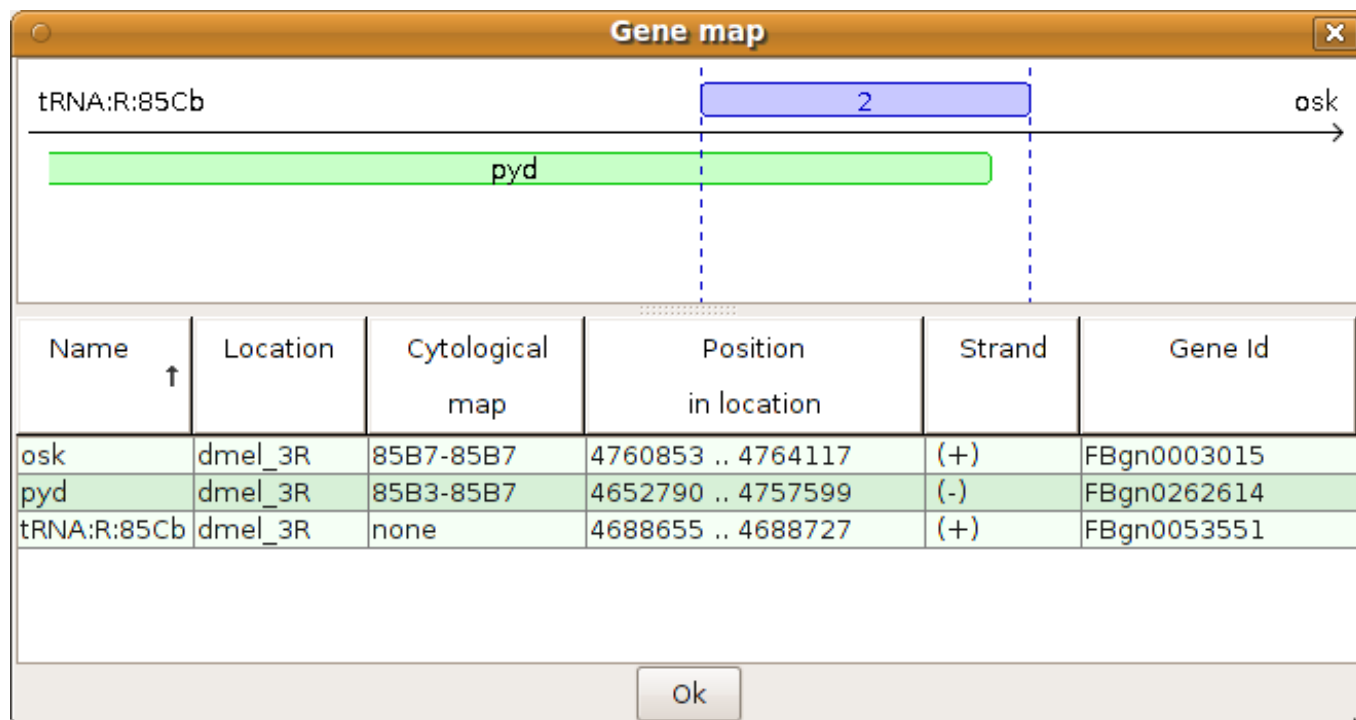


Fig. 8 In the “Gene map” panel it may be noticed that the  $P\{lacW\}$  insertion is located upstream to *pyd* gene as the sense strand of *pyd* is on the reverse strand (-) of *D. melanogaster* genome.

## 6. Saving and loading

At any time you can save the whole result set that has been found for a given query by using the “Save” and “Save As” button. The result is saved in a **Genome ARTIST** format and will have the extension .GA. The files may be stored for a further analysis. The “Load” button only works with files having .GA extension and it would load a complete result into the interface.

## 7. Advanced Settings

The settings for the search algorithm are found in Settings->Parameters.

WARNING: To understand the meaning of the parameters you must understand how the search algorithm works.

### 7.1 The search algorithm

Overlapped intervals of 10 nucleotides (the *k-mers* or decamers) further referred as *basic intervals*, are used both for indexing the reference sequences and for spanning the query sequence. When loaded into the built-in database of GA, each genome or transposon sequence is hashed and an index of addresses is generated for all of the theoretical 1.048.576 ( $4^{10}$ ) distinct decamers. **When a comparative search is started, the query sequence is scanned for all its overlapping k-mers then the appropriate index matches (hits) are retrieved.** Many initial alignments occur, then the overlapped and/or adjacent basic intervals are fused and *merged continuous intervals (MCIs)* are generated. At this stage each of the *MCIs* from query perfectly aligns with a *MCI* from the reference sequence. The *MCIs* along with some selected basic intervals are gathered in a pool which is considered for the next step of the algorithm. Then, an

extension step is performed to surpass the alignment imperfections. An initial score of alignment is defined for each of the selected intervals and a window of 4 nucleotides (equivalent to a byte) is used to check the alignments in the very vicinity of the borders. An implementation of Smith-Waterman algorithm combined with an original formula that penalizes mismatches is used for computing the score of extension. Any extension stops when the sum of the initial alignment score of an interval and the score of extension drops below zero for each of its borders. Intrinsic to the algorithm, a few mismatches are still incorporated in the *extended intervals (EIs)* which are generated this way. When *EIs* are overlapped or adjacent to each other, they are coalesced into *merged extended intervals (MEIs)*.

After this stage, a list of alignments referred also *candidate intervals (CIs)* are created. The list contains the candidate *basic intervals*, *MCIs*, *EIs* and *MEIs* that cover for each position in the query. Using again a Smith-Waterman implementation for a rigorous **realignment** of the *CIs* (where a match = 2 and a mismatch = -1), the best scoring alignments are obtained and presented to the user as a list of *partial alignments (PAs)* in the graphical interface of *GA*. Some of the *PAs* pertains to the genome/chromosome and others to the transposon sequence. The most distinctive property of *GA* consists in its ability to generate in the final step of the algorithm a hierarchy list of high scoring *assembled local alignments* or results, where the site of insertion is represented.

## 7.2 Parameters

The parameters that are found in Parameters tab (Fig. 9.a and Fig. 9.b) are:

**Type of interval expansion** – Tells the algorithm how to expand the initial intervals. Short means that any mismatches during the expansion process will be badly penalized and long means the algorithm is less severe with mismatches

**Zero offset** - [Advanced] Represents the shift of the score considered neutral for the expansion algorithm. Negative values mean greatly punishing bad alignments.

**Score match** - [Expansion] This is a bonus for an exact match in the first position

**Score mismatch** - [Expansion] Penalty for a mismatch on the first position

**Length modifier** - [Expansion] The multiplier for the initial score of the expansion phase

**Picking Depth** - Affect the number of small pieces that are picked as candidates for assembling the final result

**Nucleus Size** - Represent the minimum size of an interval in the final result

**Results number** - The number of final results that will be assembled

**Give bonus to insertion candidates** – A bonus is given if the sequence has a transposon flanking a genomic sequence. The transposon must be inserted with one of its ends.



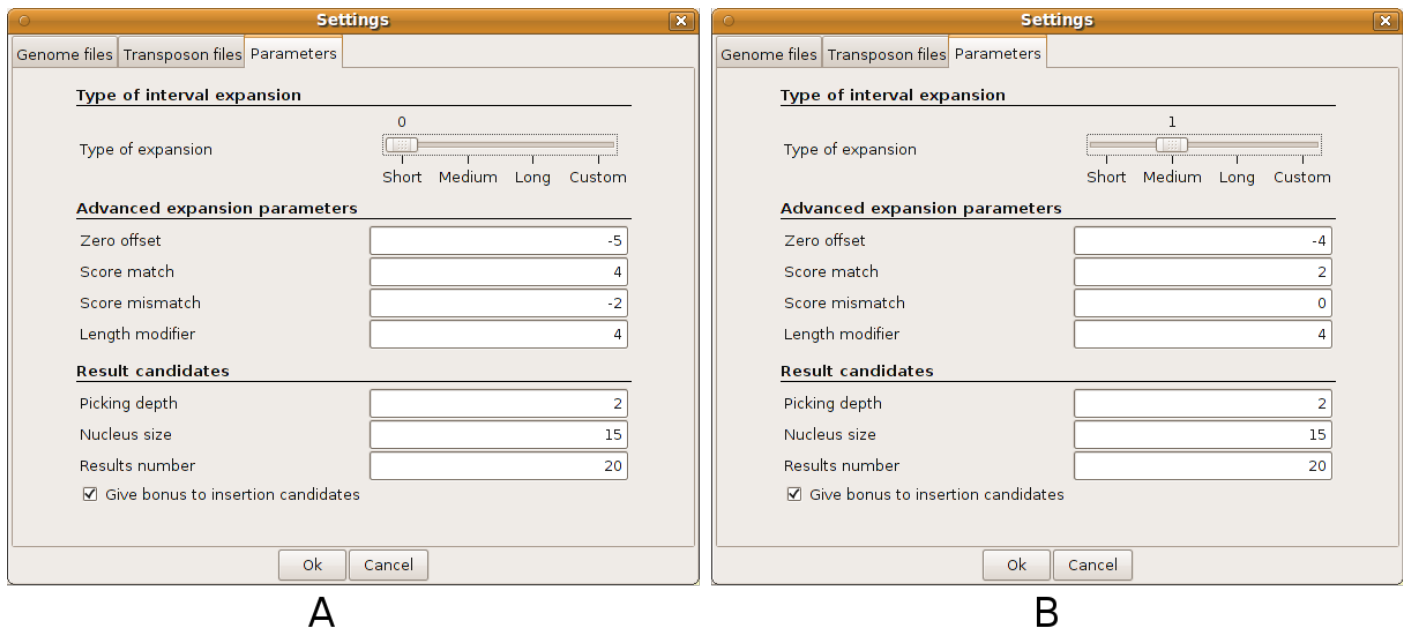


Fig. 9 a) the parameters are set for Short type of expansion; b) the parameters werereset for Medium type of expansion, therefore the advanced expansion parametersshifted to different values.

## 8. License Terms

**Genome ARTIST** is a freeware application for academic use and a proper citation is asked for any published data obtained or derived by using this software. For any commercial purpose, a written consent should be obtained from the authors. We are not responsible for any damages of computers or electronic devices induced during installing or using of **Genome ARTIST**. The direct and indirect effects of using the data obtained or derived with **Genome ARTIST** are entirely the responsibility of the user.

## 9. Authors

The authors of the present form of **Genome ARTIST** software are Alexandru Al. Ecovoiu, Iulian Constantin Ghionoiu, Andrei Mihai Ciuca and Attila Cristian Ratiu and they are the owners of all intellectually property rights derived from.

If you use Genome ARTIST please cite our work as “**Ecovoiu Al. Al., Ghionoiu I. C., Ciuca A. M. and Ratiu A. C., *Genome ARTIST (ARtificial Transposon Insertion Site Tracker), a bioinformatics tool for annotation of insertional mutagenesis data; 2012, www.genomeartist.ro***”

## 10. Acknowledgments

Alexandru Al. Ecovoiu and Attila Ratiu Cristian were partially supported by the PNII research grant **IDEI nr.147/2007, CNCSIS**, which is kindly acknowledged.