

Visualizing and exploring genomic information for non-protein-coding RNA identification using ApolloRNA

Marie-Josée Cros

INRA, France

Erika Sallet

INRA, France

Annick Moisan

INRA, France

Christine Cierco-Ayrolles

INRA, France

Christine Gaspin

INRA, France

Method Article

Keywords: ncRNA, visualization, annotation

Posted Date: July 6th, 2007

DOI: <https://doi.org/10.1038/nprot.2007.285>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Introduction

This protocol describes installation and use of the ApolloRNA software to visualize and explore various genomic information for non-protein-coding RNA (ncRNA) identification in prokaryote. The first part details how to install Apollo¹ and the extension ApolloRNA on a computer with Windows, Linux or MAC OS X. The second part describes how to collect and visualize annotations and some information of an organism. This protocol allows to bring together information on an organism to visualize and analyze them locally on a computer in order to define and select valuable ncRNA candidates to be experimentally tested.

Equipment

A modern personal computer with Windows (Microsoft), Linux (e.g. debian or Red Hat) or MacOS X (Apple) operating system is required as well as an Internet connection and any graphical user interface-based web-browser.

Procedure

Part 1. Apollo and ApolloRNA installation. TIMING 35 min (Windows), 40 min (Linux) or 45 min (MAC)

1 | Download and install Apollo using option A, B or C depending on the operating system.

(A) For Windows

- (i) In the directory 'My Documents', create the subdirectory 'Applications'.
- (ii) In the directory 'Applications', create the subdirectory 'Apollo'.
- (iii) Access the download Apollo page at "<http://www.fruitfly.org/annot/apollo/install.html>".
- (iv) In the line 'Windows', in the column 'includes Java VM', right-click the link 'Download'. In the menu, click 'Save link as...' and save the file in the directory 'Apollo'.
- (v) Double-click the downloaded file 'install.exe' and follow instructions choosing the 'Apollo' directory as installation directory.

(B) For Linux

- (i) Create the directory '~/.Applications'.
- (ii) In the directory 'Applications', create the subdirectory 'Apollo'.
- (iii) Access the download Apollo page at "<http://www.fruitfly.org/annot/apollo/install.html>".
- (iv) In the line 'Linux', in the column 'includes Java VM', right-click the link 'Download'. In the menu, click 'Save link as...' and save the file in the directory 'Apollo'.
- (v) Open a terminal window and move (with the command 'cd') to the directory 'Apollo'. Type: `sh ./apolloinst.bin` and follow instructions choosing 'Apollo' for installation directory.
- (vi) Configure Apollo for Java. Open a terminal window, move to the directory 'Apollo' and type: `pwd`. The displayed line is the Apollo path, for instance '/home/dupont/Applications/Apollo'. In the subdirectory 'conf' of the directory 'Apollo', open the file 'apollo.cfg'. Add a line defining the JavaPath variable as the Apollo path extended with '/jre', for instance add the line: `JavaPath == "/home/dupont/Applications/Apollo/jre" ==`

(C) For Mac

- (i) In your user directory, if the directory 'Applications' doesn't exist, create it.
- (ii) In the directory 'Applications', create the subdirectory 'Apollo'.
- (iii) Access the download Apollo page at "<http://www.fruitfly.org/annot/apollo/install.html>".
- (iv) In the line 'Mac OS X', in the column 'without Java VM', right-click the link 'Download'. In the menu, click 'Save link as...' and save the file in the directory 'Apollo'.
- (v) Double-click the downloaded file 'install', and follow instructions choosing the 'Apollo' directory as installation directory.
- (vi) To be sure Java 1.4 is on your computer, open a terminal window and type: `java -version`. If the displayed message starts with 'java version 1.4', go directly to the step 2.
- (vii) Access the download Java page at "<http://www.apple.com/support/downloads/javaupdate142.html>". Click the arrow on the Download line and save the file 'Java142Update2.dmg' in your directory 'Applications'. Double-click 'Java142Update2.dmg' and follow instructions.
- (viii) Configure Apollo for Java. Open a terminal window and type: `which java`. The displayed line is the java executable location, for instance '/usr/bin/java'. In the subdirectory 'conf' of the directory 'Apollo', open the file 'apollo.cfg'. Add a line defining the 'JavaPath' variable to the path before 'bin/java', for example: `JavaPath == "/usr/" ==`

2 | Download and install RNAfold (predict minimum energy secondary structures and pair probabilities) and RNAeval (evaluate energy of RNA secondary structures) of the Vienna RNA package² (programs for RNA Secondary Structure prediction and comparison) using option A or B depending on the operating system.

(A) For Windows

- (i) Access the download ViennaRNA package page at "<http://www.tbi.univie.ac.at/ivo/RNA/windoze/>".
- (ii) Below 'Download', right-click the 'RNAfold.exe' link and save the file in the directory 'Apollo'.
- (iii) Repeat the step (ii) to save the file 'RNAeval.exe'.
- (iv) In the taskbar, click 'Start' then chose 'Control Panel'. In the new window, double click on 'System' or if it does

not exist on 'Performance and Maintenance' then on 'System'. In the 'System properties' window, click the 'Advanced' tab, click the 'Environment Variables' button to open the environment variables window (**Figure 1a**). (v) If the user variable 'Path' exists, select the line of the 'Path' variable, click the 'Edit...' button and add ';' at the end of the value followed by the path of the 'Apollo' directory. For instance, if the path of the 'Apollo' directory is 'C:\dupont\Apollo', add text: **;C:\dupont\Apollo**. Else, if the user variable 'Path' does not exist, click the 'New...' button. A new window (**Figure 1b**) appears that allows to define a new variable named 'Path' with the path of the 'Apollo' directory for value. For instance, if the path of the 'Apollo' directory is 'C:\dupont\Apollo', the value to be typed is: **C:\dupont\Apollo**.

(B) For Linux or MAC. CRITICAL STEP with administrator privileges.

(i) Access the Vienna RNA package download page at "<http://www.tbi.univie.ac.at/ivo/RNA/>":<http://www.tbi.univie.ac.at/ivo/RNA/>. Click 'source distribution here' link and save the 'ViennaRNA.gz' file in the directory 'Applications'. (ii) Open a terminal window and move (with the command 'cd') to the directory 'Applications'. To install the Vienna RNA package, type: **tar xzvf ViennaRNA-1.6.2.tar.gz** **cd ViennaRNA-1.6.2** **./configure** **make** **make install**

3| Download and install ImageMagick (a software suite to create, edit, and compose bitmap images) using option A or B depending on the operating system. CRITICAL STEP with administrator privileges.

(A) For Windows (i) Access the download ImageMagick page at "<http://www.imagemagick.org/script/binary-releases.php#windows>":<http://www.imagemagick.org/script/binary-releases.php#windows>. (ii) In the first line, in the 'HTTP' column, right-click the link 'download'. In the menu, click 'Save link as...' and save the file in the directory 'Applications'. (iii) Double-click the downloaded file 'ImageMagick-6.3.3-10-Q16-windows-dll.exe' and follow instructions.

(B) For Linux or MAC (i) Access the download ImageMagick page at "<http://www.imagemagick.org/script/install-source.php>":<http://www.imagemagick.org/script/install-source.php>. (ii) Right-click the link 'ImageMagick.tar.gz'. In the menu, click 'Save link as...' and save the file in the directory 'Applications'. (iii) Open a terminal window and move to the directory 'Applications'. To install ImageMagick, type: **tar xzvf ImageMagick.tar.gz** **cd ImageMagick-6.3.1** **./configure** **make** **make install**

4| Download and install GhostScript (an interpreter for the PostScript language and for PDF) using option A or B depending on the operating system. CRITICAL STEP with administrator privileges.

(A) For Windows (i) In the taskbar, click 'Start' then choose 'Control Panel'. In the new window, click 'Add or Remove Programs'. Look for the program 'Ghostscript 8.56' in the list of programs. If the program belongs to the list, go directly to the step 5. (ii) Access the download GhostScript page at "<http://www.ghostscript.com/awki>":<http://www.ghostscript.com/awki>. Click the link 'sourceforge.net'. (iii) Right-click the link 'gs856w32.exe'. In the menu, click 'Save link as...' and save the file in the directory 'Applications'. (iv) Double-click the downloaded file 'gs856w32.exe' and follow the instructions.

(B) For Linux or MAC (i) Open a terminal window and type: **gs -h**. If the message 'Command not found' is not displayed, go directly to the step 5. (ii) Access the download GhostScript page at "<http://www.ghostscript.com/awki>":<http://www.ghostscript.com/awki>. Click the link 'sourceforge.net'. (iii) Right-click the link 'ghostscript-8.56.tar.gz'. In the menu, click 'Save link as...' and save the file in the directory 'Applications'. (iv) Open a terminal window and move to the directory 'Applications'. To install GhostScript, type: **tar xzvf ghostscript-8.56.tar.gz** **cd ghostscript-8.56** **./configure** **make** **make install**

5| Download and install ApolloRNA (i) In the directory 'Apollo', rename 'apollo.jar' to 'apollo.jar.save'. (ii) Access the ApolloRNA project page at "http://mulcyber.toulouse.inra.fr/frs/?group_id=47":http://mulcyber.toulouse.inra.fr/frs/?group_id=47. Right-click the link 'apollo.jar'. In the menu, click 'Save link as...' and save the file in the directory 'Apollo'. **[Alert!]** See author comment below. The correct link is: "<http://mulcyber.toulouse.inra.fr/gf/project/apollorna/frs/>":<http://mulcyber.toulouse.inra.fr/gf/project/apollorna/frs/> **Part 2**. Load and visualize data about *Pyrococcus abyssi*_GE5 genome. TIMING 30 min

6| Download *Pyrococcus abyssi*_GE5 GenBank annotations in GAME XML format (i) Access the ftp NCBI page at "ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Pyrococcus_abyssi":ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Pyrococcus_abyssi. Click the link 'NC_000868.gb' and copy the whole text. (ii) Access to the ApolloRNA convert data page at "<http://carlit.toulouse.inra.fr/ApolloRNA/convert.html>":<http://carlit.toulouse.inra.fr/ApolloRNA/convert.html> and paste the data in the text area. In the 'Input format' list, select 'Genbank' and in the 'Output format' list, select 'Game XML'. Click the 'Convert' button. (iii) Copy the resulting text and paste it in a new file called 'NC_000868.xml'.

7| Download predicted transcription terminator in GAME XML format (i) Access to the TransTermHP³ page at "<http://transterm.cbcb.umd.edu/cgi-bin/transterm/predictions.pl>":<http://transterm.cbcb.umd.edu/cgi-bin/transterm/predictions.pl> and click 'Pyrococcus abyssi GE5'. Copy all the displaying text, except the last part (terminators of the plasmid) starting by the line: SEQUENCE [gil10954552\[ref|NC_001773.1\]](http://www.ncbi.nlm.nih.gov/nuccore/gil10954552?from=ref&from_pos=1) *Pyrococcus abyssi* GE5 plasmid pGT5, complete sequence (ii) Access to the ApolloRNA convert data page and paste the data in the text area. In the 'Input format' list, select 'TransTermHP', and in the 'Output format' list, select 'Game XML'. In the 'Related genomic sequence coordinates' fields, enter **1** for the beginning and **1765118** for the end. Click

the 'Convert' button. \(\text{iii}\) Copy the resulting text and paste it in a new file called 'NC_000868_TransTermHP.xml'. ****8| Visualizing both information**** \(\text{i}\) Launch Apollo using option A or B depending on the operating system. \(\text{A}\) For Windows: In the directory 'Apollo', click the Apollo icon. \(\text{B}\) For Linux or MAC: Open a terminal window, move to the directory 'Apollo' and type: ****bin/Apollo****. The Apollo load data window \(\text{**Figure 1c**}\) appears. \(\text{ii}\) In the list 'Choose data source', choose 'GAME XML format'. Click the 'Browse...' button: a window appears. Select the file 'NC_000868.xml' and click the 'Open' button. In the Apollo load data window, click the 'OK' button. When Apollo has loaded data, the Apollo main window appears. \(\text{iii}\) Open the 'File' menu and click the 'Layer more results or annotations...' item. In the 'Choose data source' list, choose 'GAME XML format'. Click the 'Browse...' button. Select the file 'NC_000868_TransTermHP.xml' and click the 'Open' button. In the Apollo load data window, click the 'OK' button. ****9| Visualizing a predicted secondary structure, example for a tRNA**** \(\text{i}\) Click twice on the 'X10' zoom button. The main Apollo window displays both GenBank annotations \(\text{in the blue parts}\) and TransTermHP data \(\text{in the black parts}\) of *_Pyrococcus abyssi_* genome. Click the inorganic polyphosphate/ATP-NAD kinase protein annotation to display \(\text{**Figure 1d**}\) more information about the annotation. \(\text{ii}\) Open the 'Edit' menu and click the 'Find...' item. A new 'Find' window opens. In the field 'Name', type: ****PABt04****. Click the 'Find' button on the line. The Apollo main view zooms on the tRNA PABt04. Click the 'X.1' zoom button to change scale. \(\text{iii}\) Open the 'RNA' menu and click the 'Secondary structure compute...' item. A new window \(\text{**Figure 2a**}\) opens that displays the predicted secondary structure of the tRNA and the associated free energy. Click the 'Close' button. ****10| Visualizing possible RNA/RNA interactions, example for the interaction between a sno C/D RNA and the 16S rRNA**** \(\text{i}\) In the opened 'Find' window, in the field 'Name', type now: ****PABsnRNA29****. Click the 'Find' button on the line. The Apollo main view zooms on the sno C/D box RNA called 'PABsnRNA29'. \(\text{ii}\) Now select the subsequence from 1403702 to 1403711 \(\text{a subpart of the snoRNA}\) that is known⁴ to interact with the 16S rRNA. That is, press the middle mouse button at the position 1403702 immediately above the axis, and drag the mouse to 1403711 surrounding the sequence 'TGTAGCGCGC' \(\text{**Figure 2b**}\). A window opens that shows the selected sequence. Close it. \(\text{iii}\) Open the 'RNA' menu and click the 'Search RNA/RNA interaction' item. A window \(\text{**Figure 2c**}\) opens for configuring the interaction search. \(\text{iv}\) In the top part of the window, type the genomic positions: ****205039**** to ****206541**** \(\text{the 16S rRNA positions}\). In the strand list, choose the strand +. \(\text{v}\) In the bottom part, click the 'Search' button to launch the search of targets in the specified regions. After a few seconds, a new window \(\text{**Figure 2d**}\) opens that lists the predicted targets of the previously selected sequence. The first on the list is the good⁴ one. \(\text{vi}\) Click the first interaction picture to visualize the interaction in a new window \(\text{**Figure 2e**}\). Click the 'Close' button and then click the 'Close' button on the previous window. ****11| Visualizing graphs**** \(\text{i}\) Go to the ApolloRNA page "<http://carlit.toulouse.inra.fr/ApolloRNA/apollornauserguide.html>":<http://carlit.toulouse.inra.fr/ApolloRNA/apollornauserguide.html>, click the link 'NC_000868_mfe150.txt' and save the file on your computer. This file contains the minimum free energy computed on a sliding window of 150 nucleotides. \(\text{ii}\) Open the 'Analysis' menu and click the 'Show quantitative variable...' item. A new window \(\text{**Figure 3a**}\) opens. \(\text{iii}\) In the first table of the window, in the first line, click the 'Load...' button. Select the file 'NC_000868_mfe150.txt' and click the 'Open' button. In the column 'Color', click the black rectangle, choose the red color and click the 'OK' button. \(\text{iv}\) In the second table of the window, in the third line 'ComputeGC%': click the 'Show' checkbox, in the column 'Color', click the black rectangle and choose the blue color. Click the 'OK' button and then click the 'Close' button. \(\text{v}\) In the yet opened 'Find' window, in the field 'Name', type now: ****16S****. Click the 'Find' button on the line. Click the 'X.1' zoom button to change scale. Move down the superior vertical scrollbar to see RNA annotations on the direct strand. The Apollo main window \(\text{**Figure 3b**}\) displays now two additional graphs that vary on the ribosomal RNAs region.

Timing

The first part can be performed in less than 1h. The second part can be performed in 30mn.

Critical Steps

Steps that require administrator privileges on the computer are emphasized as critical.

Troubleshooting

If troubleshooting was encountered in installing Apollo, notes of the page 'Download and Install Apollo' \(\text{ "<http://www.fruitfly.org/annot/apollo/install.html>":<http://www.fruitfly.org/annot/apollo/install.html>) could help. If the genome is

long or information files voluminous, loading time could be long (depending on the characteristics of the computer).

Anticipated Results

In the protocol, the page

"<http://carlit.toulouse.inra.fr/ApolloRNA/convert.html>":<http://carlit.toulouse.inra.fr/ApolloRNA/convert.html> was mentioned to convert GenBank annotation data or TransTermHP predictions of transcription terminator in an Apollo suitable GAME XML format. This page could also be useful for converting data as CRISPRfinder⁵ predictions of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), MOSAIC⁶ comparative data of closely related bacterial genomes, PatScan⁷ pattern matching data, DARN⁸ non-coding RNA detection data. On request, convert capabilities could be extended. Beside the individual interest of the environment, we found useful to share in a project a common representation of information. That is achieved by defining how information is displayed (color, shape, individual or shared track...). This could then be stored in a type preference file which could be duplicated. Another interesting feature is the ability for several persons to walk along the genome on an inflated video projection of the representation and discuss on particular regions. Finally the software could support interdisciplinary work allowing a common representation. This is put into practice on three genomes: *Pyrococcus abyssi*, *Staphylococcus aureus* and *Ralstonia solanacearum*. For *S. aureus*, the approach for RNA/RNA interaction prediction was successfully used to predict targets⁹ of the RNA III and some small ncRNAs candidates have yet been experimentally validated (to be published).

References

1. Lewis, S.E. et al. Apollo : a sequence annotation editor. *Genome Biol.* **3**, R12, (2002).
2. Hofacker, I.L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429-3431, (2003).
3. Kingsford, C., Ayanbule, K., Salzberg, S.L. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.* **8**, R22, (2007).
4. Gaspin, C. et al. Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J. Mol. Biol.* **297**, 895-906, (2000).
5. Grissa, I., Vergnaud, G., Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* published online on May 31, (2007).
6. Chiapello, H. et al. Systematic determination of the mosaic structure of bacterial genomes: species backbone versus strain-specific loops. *BMC Bioinformatics* **6**, 171, (2005).
7. Dsouza, M., Larsen, N., Overbeek, R. Searching for patterns in genomic data. *Trends Genet.* **13**, 497-498, (1997).
8. Zytnicki, M., Gaspin, C., Schiex, T. DARN! A soft constraint solver for RNA motif localization. To be published in *Constraints*, (2007).
9. Boisset, S. et al. *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev.* **21**, 1353-1366, (2007).

Figures



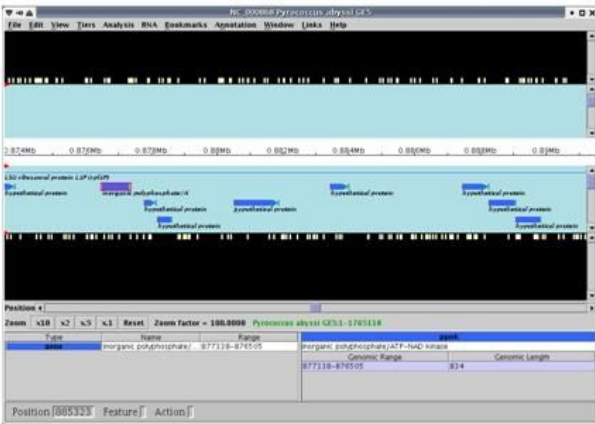
(a) An 'Environment Variables' window displaying both user and system defined variables on Windows operating system.



(b) A 'New User Variable' window allowing to define a user variable (name and value) on Windows operating system.



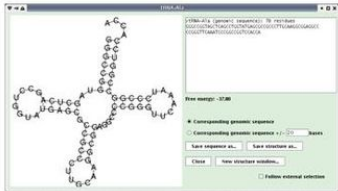
(c) An Apollo load data window that allows to specify the format of input data and the file to load.



(d) An Apollo main window that displays both GenBank annotations (in the blue parts) and TransTermHP data (in the black parts) of the *Pyrococcus abyssi* GE5 genome. The inorganic polyphosphate/ATP-NAD kinase protein annotation is selected (surrounded in red). So the information associated to this annotation is displayed in the lower part of the window.

Figure 1

Windows for configuring Windows operating system: (a) and (b) and displayed windows calling Apollo (c) and (d).



(a) An ApolloRNA window that displays the predicted secondary structure of the current selected sequence: tRNA-Ala.

```

T G A C C G T A G C C G C C T C T G A
1403702
A C T G A C A T C G C G A G A C T
  
```

(b) A portion of an Apollo main window showing the selection (surrounded by a red line) of a subsequence from 1403702 to 1403711 in the direct strand using the middle mouse button to surround the corresponding nucleotides.

(c) An ApolloRNA window to parameterize and launch RNA/RNA interactions search for the currently selected sequence.

Position	Strand	Interaction	Gene	Start	Stop	Score
216241	+	5'-UUUUUUU-3'	PAB0133	PAB0137	-18.2	-18.7
216240	+	5'-UUUUUUU-3'	PAB0133	PAB0137	-18.4	-19.2
216242	+	5'-UUUUUUU-3'	PAB0133	PAB0137	-18.9	-19.4
216243	+	5'-UUUUUUU-3'	PAB0133	PAB0137	-19.4	-19.3
216219	+	5'-UUUUUUU-3'	PAB0133	PAB0137	-19.4	-19.3
216221	+	5'-UUUUUUU-3'	PAB0133	PAB0137	-19.4	-19.6
216220	+	5'-UUUUUUU-3'	PAB0133	PAB0137	-19.6	-19.2
216220	+	5'-UUUUUUU-3'	PAB0133	PAB0137	-19.7	-19.1
216220	+	5'-UUUUUUU-3'	PAB0133	PAB0137	-19.8	-19.4

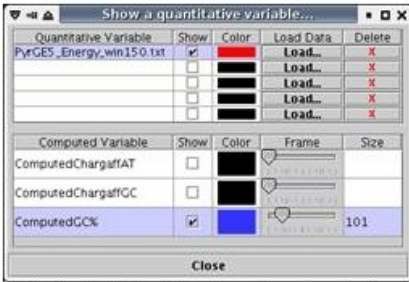
(d) An ApolloRNA window that lists predicted RNA/RNA interactions for the currently selected sequence



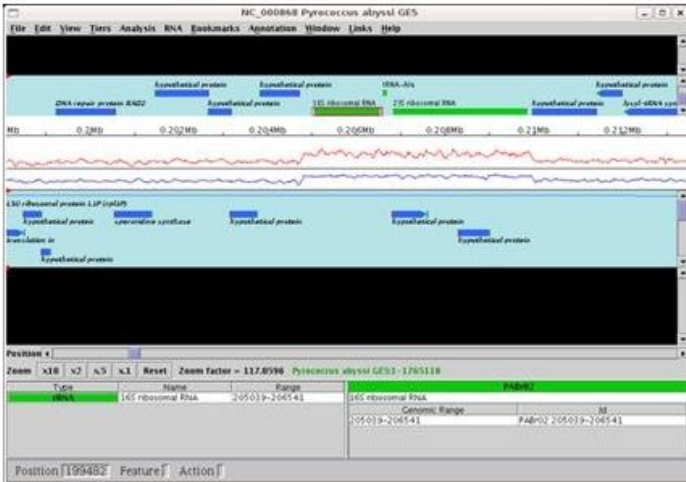
(e) An ApolloRNA window that displays a predicted interaction between two sequences. The currently selected sequence is written in blue and the predicted target in red.

Figure 2

Displayed windows looking for a secondary structure: (a) and searching RNA/RNA interactions: (b), (c), (d) and (e).



(a) An ApolloRNA window to configure the display of quantitative variables.



(b) A Apollo main window displaying two additional graphs: in blue, the GC% computed on a 101 nucleotides sliding window graph and in red, the minimum free energy computed on a 150 nucleotides sliding window. In the bottom of the window, some information are displayed on the selected 16S ribosomal RNA of the *Pyrococcus abyssi* GE5 genome.

Figure 3

Displayed windows visualizing quantitative variables: (a) and (b).



Figure 4

This word file contains larger versions of all the figures associated with this protocol.