PoSE: Pattern of Sequence Evolution

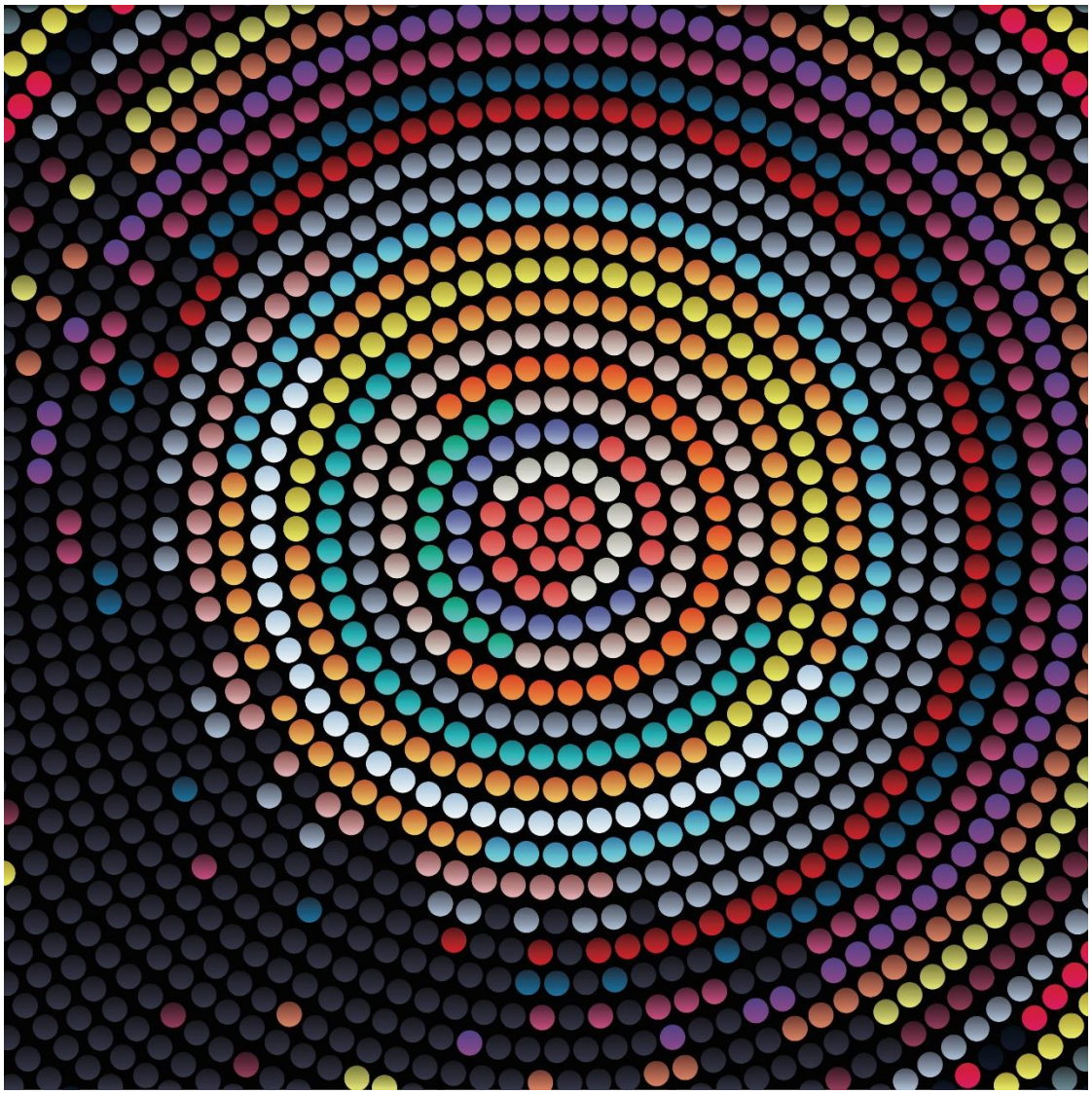


Table of Contents

1 Introduction

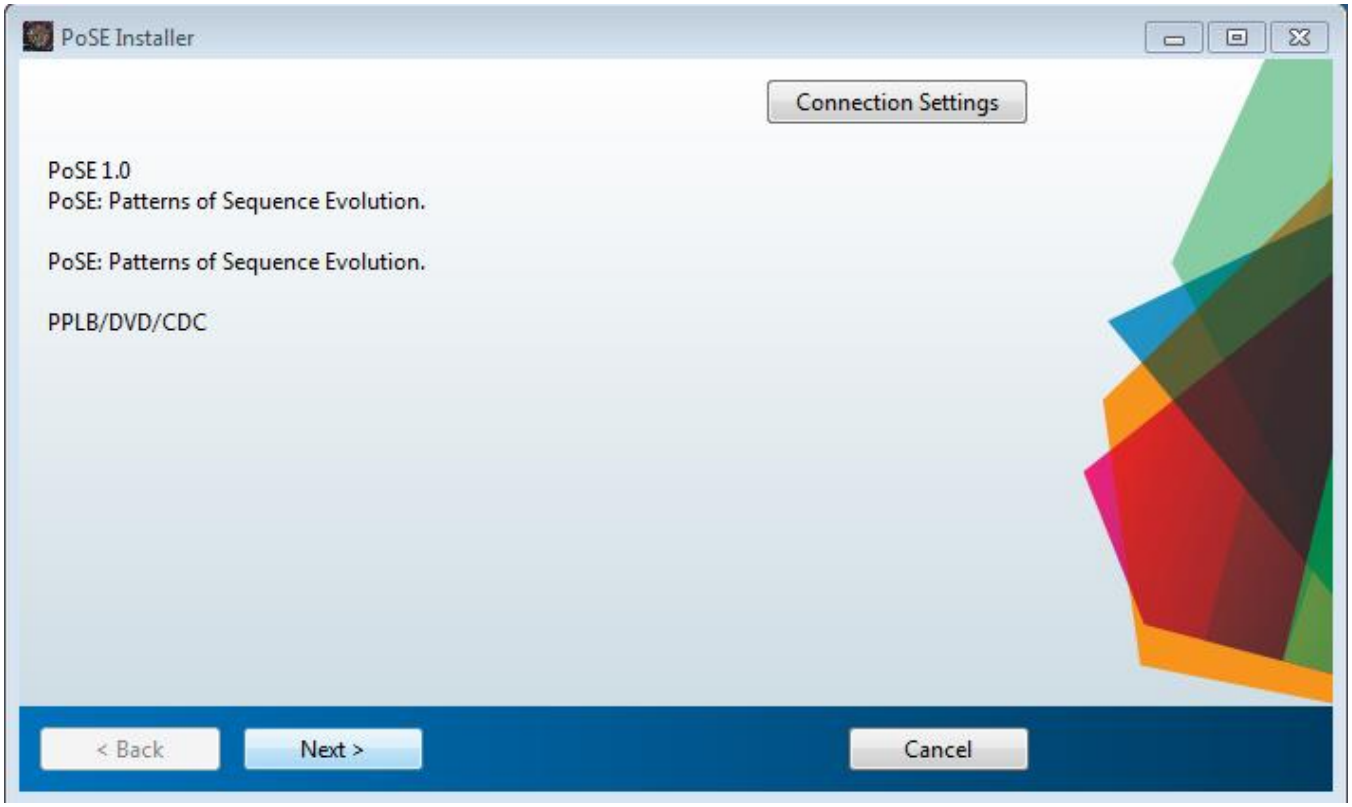
PoSE: (Pattern of Sequence Evolution) provides visualization and annotation of amino acid substitutions to help determine major patterns during sequence evolution of protein-coding sequences, hypervariable regions, or changes in dN/dS ratios.

PoSE is publicly available at https://github.com/CDCgov/PoSE

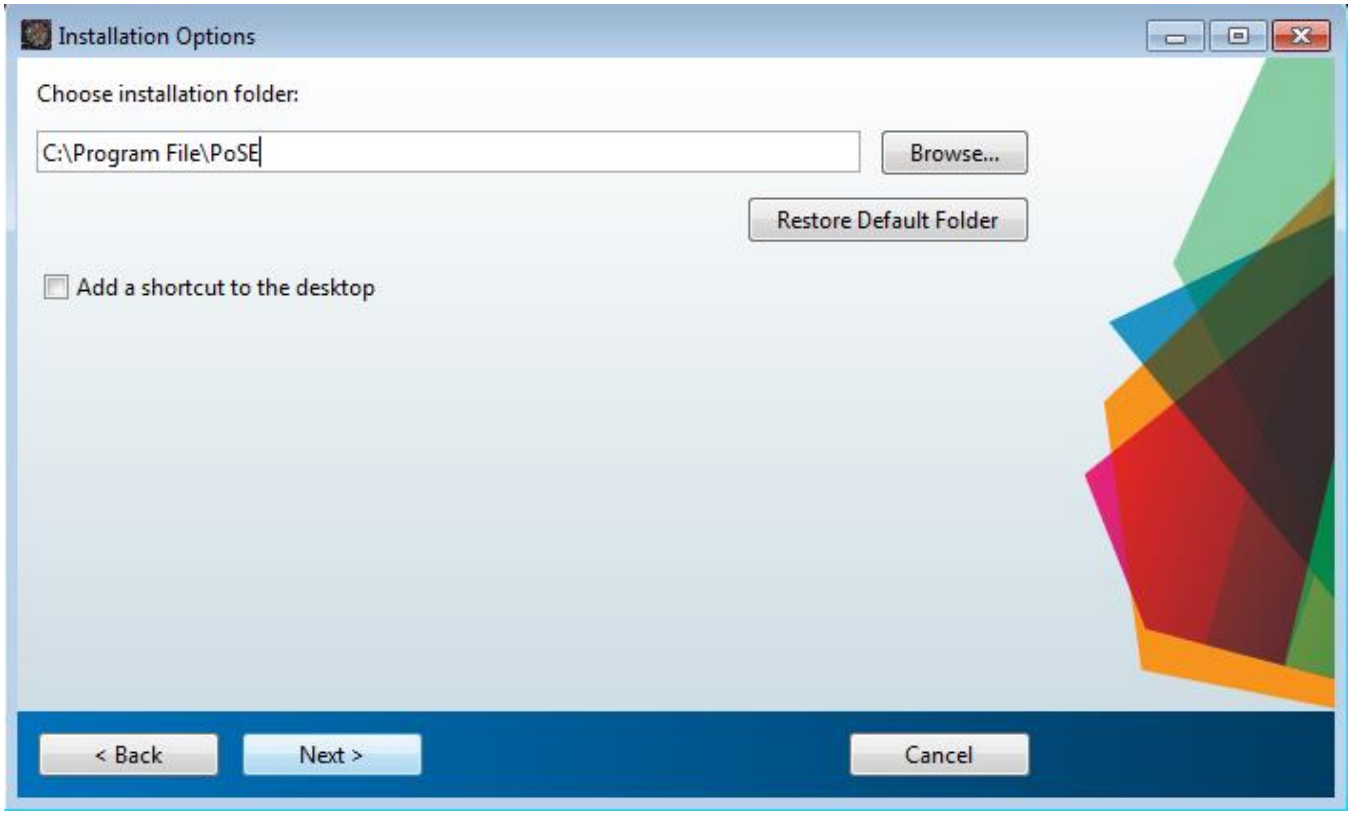
2 Software Requirements and Installation

User needs admin right to perform installation and installation should be straightforward. Please using setup default for installation. No changes are necessary and see following print screens as reference. Please note: installing Matlab runtime requires a substantial download (~ 800Mb) and installation of PoSE may take some time (depends on configuration of individual computer).

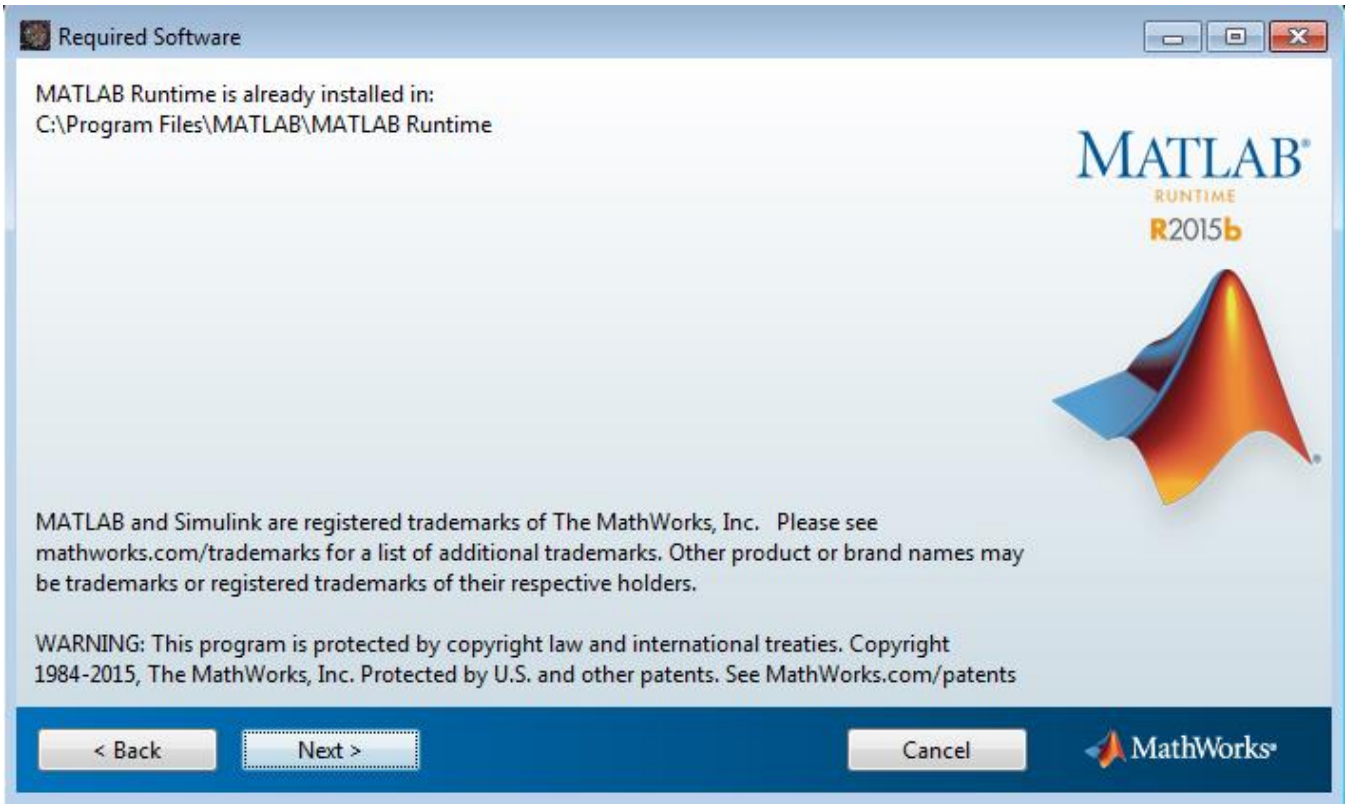
1.



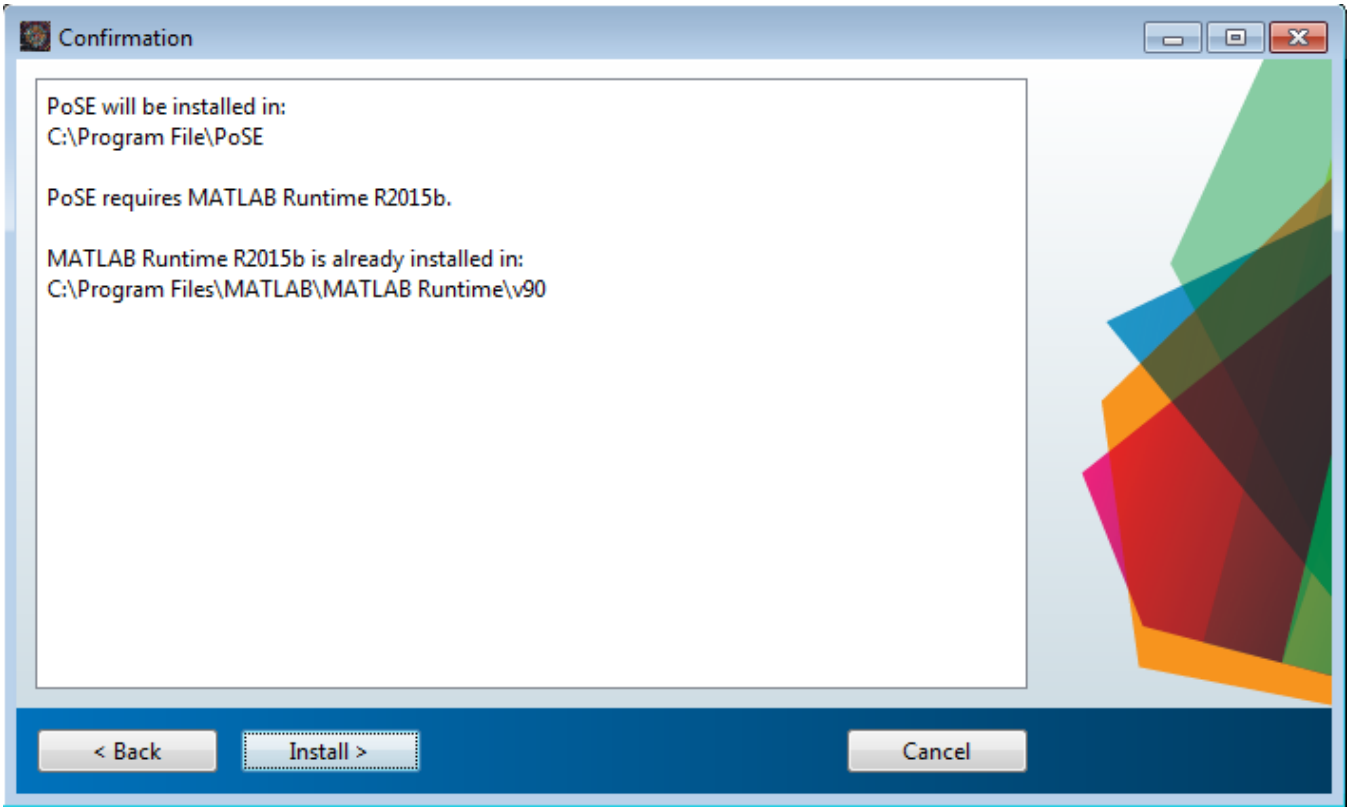
2.


Installation Options
Choose installation folder:
C:\Program File\PoSE
Browse...
Restore Default Folder
Add a shortcut to the desktop
< Back
Next >
Cancel

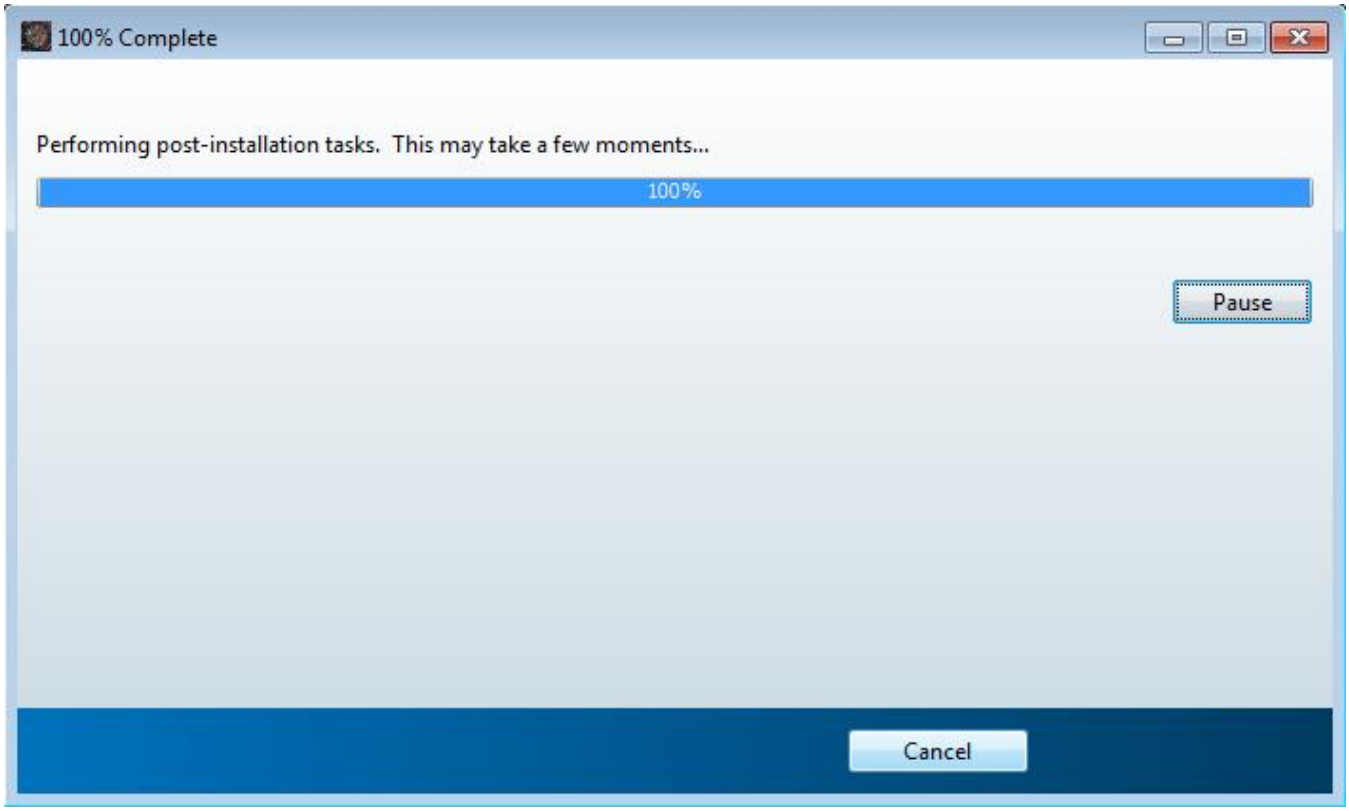
3.


Required Software
MATLAB Runtime is already installed in:
C:\Program Files\MATLAB\MATLAB Runtime
MATLAB
RUNTIME
R2015b
MATLAB and Simulink are registered trademarks of The MathWorks, Inc. Please see mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
WARNING: This program is protected by copyright law and international treaties. Copyright 1984-2015, The MathWorks, Inc. Protected by U.S. and other patents. See MathWorks.com/patents
< Back
Next >
Cancel
MathWorks

4.

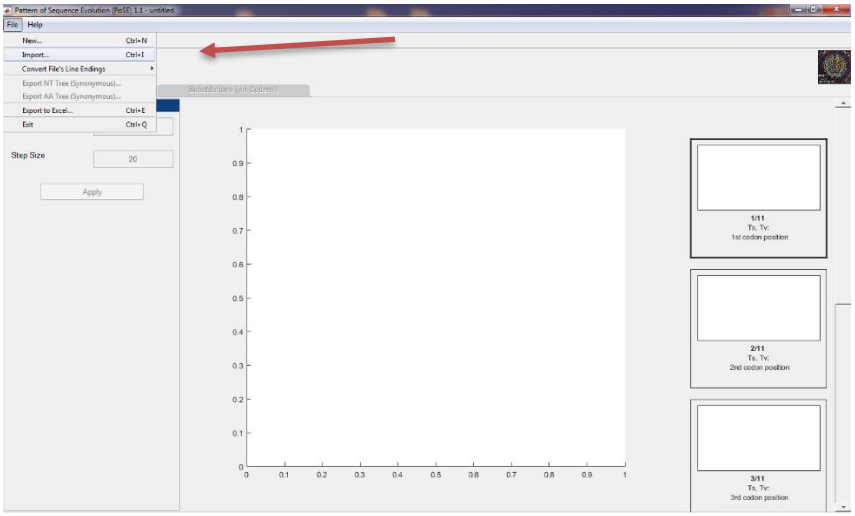

Confirmation
PoSE will be installed in:
C:\Program File\PoSE
PoSE requires MATLAB Runtime R2015b.
MATLAB Runtime R2015b is already installed in:
C:\Program Files\MATLAB\MATLAB Runtime\v90
< Back
Install >
Cancel

5.

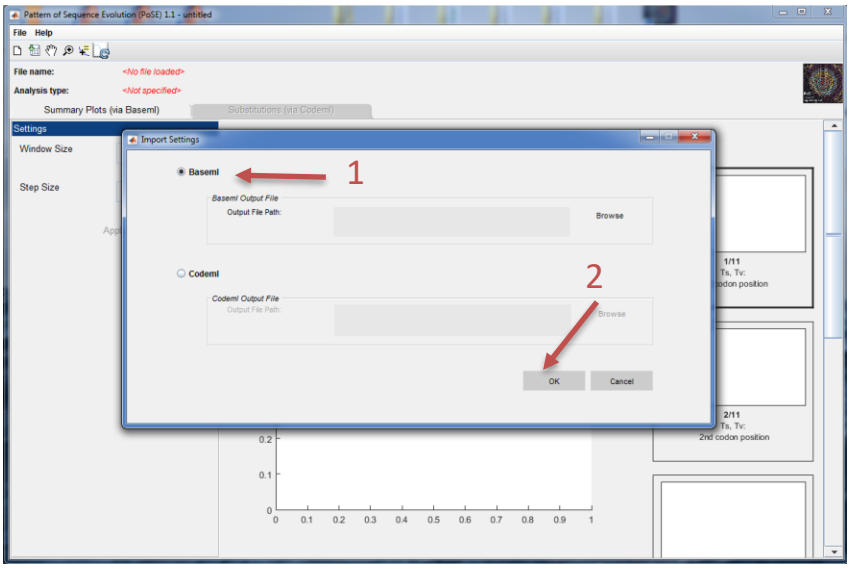


3 Usage Examples

After installation, run PoSE to open the graphic user interface. Then, click File -> Import to open the Input menu. The sample data for these tutorials can be found under the “Example_PoSE_File” folder: “baseml_rst.txt” and “codeml_rst.txt” are used for direct input into PoSE’s functionality of “Baseml” and “Codeml”, respectively. Files “baseml.ctl” and “codeml.ctl”, in the PAML_Files folder, are used for a complete PAML run. [Note: PAML should first be run to obtain the correct input file (rst file). See Tips and Tricks for running PAML.]

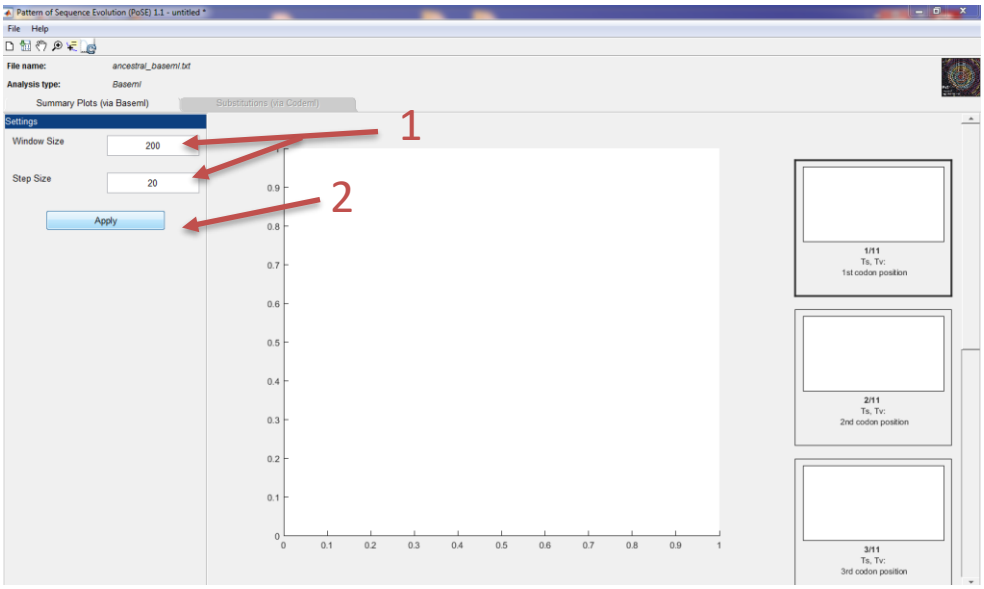


First, select either Baseml or Codeml (nr.1) and click OK at the bottom (nr.2).

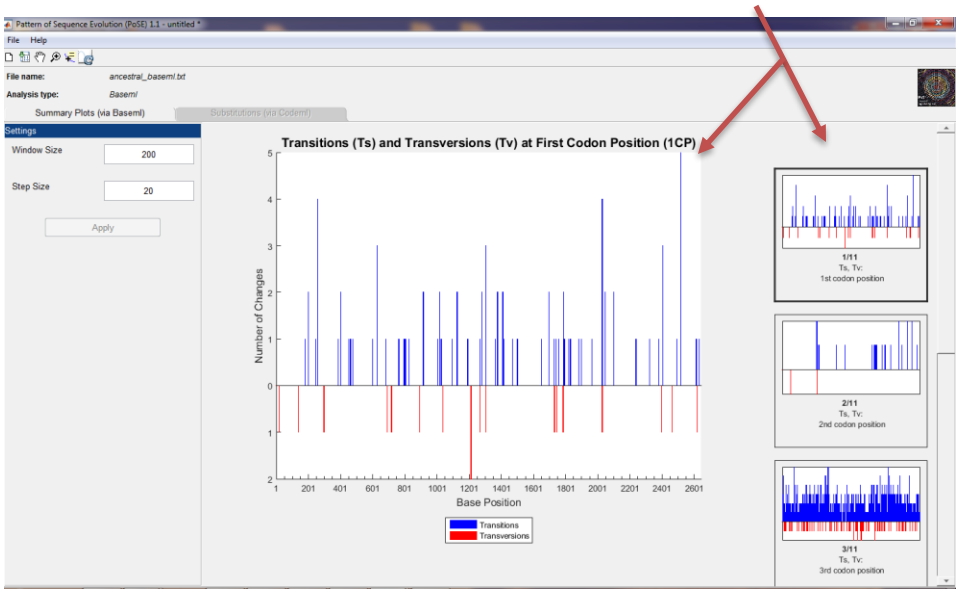


3.1 Baseml

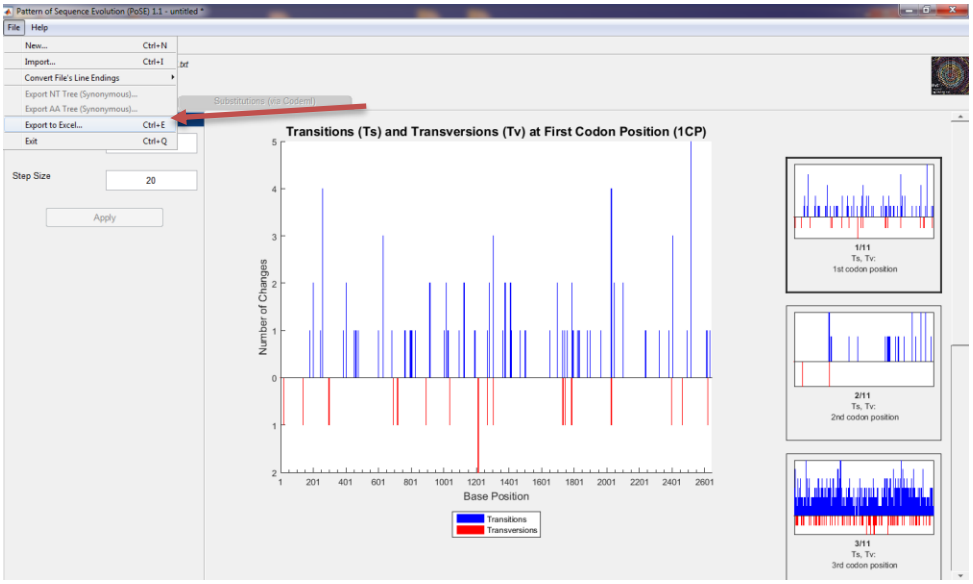
1. Import “Example_PoSE_File / baseml_rst.txt” into PoSE through Baseml -> Import Output File.
2. Set the Window Size and Step Size (nr. 1), then select Apply (nr. 2). Window size defines the length of a sliding window, given a certain step size.



3. After import, the follow graphs should now be visible:
    a. Transitions (Ts) and Transversions (Tv) at First Codon Position (1CP)
    b. Transitions (Ts) and Transversions (Tv) at Second Codon Position (2CP)
    c. Transitions (Ts) and Transversions (Tv) at Third Codon Position (3CP)
    d. Transitions (Ts) and Transversions (Tv) at All Codon Positions
    e. Transitions (Ts) and Transversions (Tv) by Codon Position
    f. Base Pair Changes by Codon Position
    g. Synonymous and Nonsynonymous Substitutions (2 windows)
    h. Synonymous and Nonsynonymous Substitutions Sliding Window
    i. Cumulative Synonymous and Nonsynonymous Substitutions
    j. Cumulative Nonsynonymous and All Substitutions



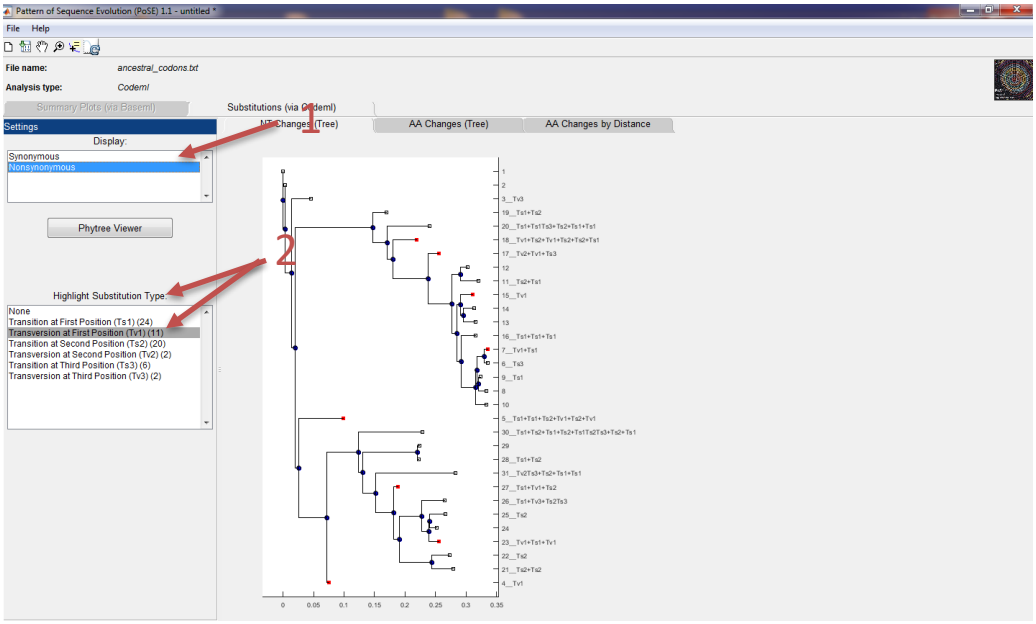
4. To export click File -> Export to Excel. The spreadsheet contains two tabs: a list of all substitutions and a relative frequency matrix for the nucleotides.



3.2 Codeml

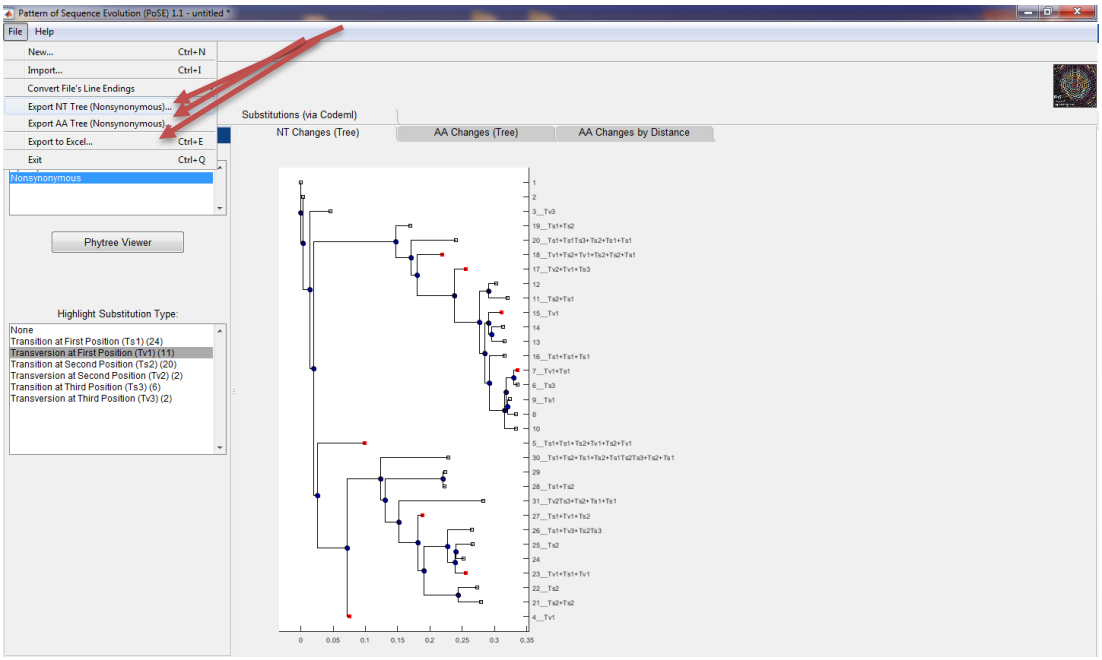
1. Import codeml_rst.txt into PoSE through Codeml -> Import Output File.

2. Under Settings, select the display option of synonymous or nonsynonymous (nr. 1). The selection is shown to the right of the phylogenetic tree. The transitions and transversion can be highlighted on the tree by selecting from the Highlight Substitution Type menu (nr. 2). The numbers following the options are the substitution counts for each (nr. 2).



3. There are two tabs above the display screen which show: the nucleotide changes to the right of the tree and the amino acid changes to the right of the tree.

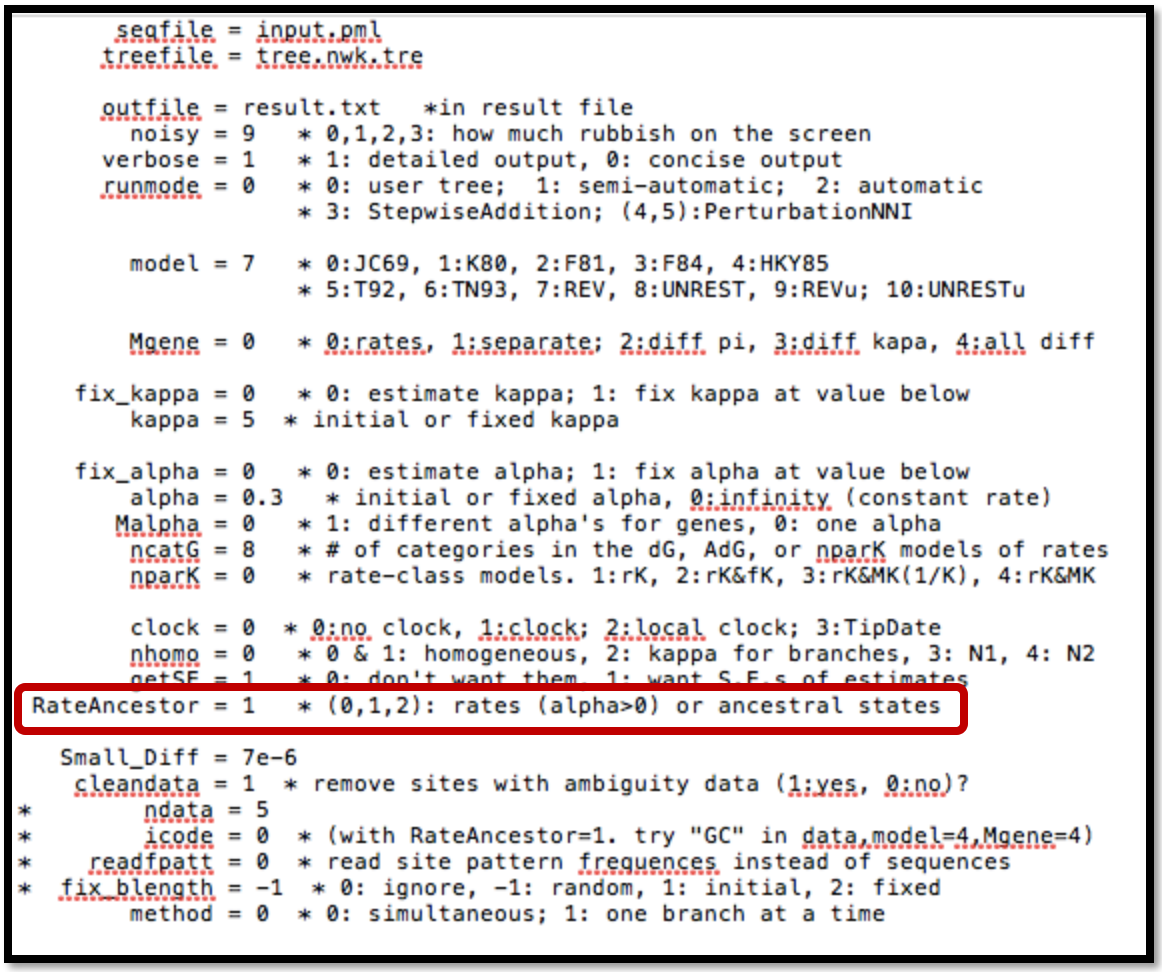
4. There are three separate export options. To export the Excel click File -> Export to Excel. The spreadsheet contains six tabs that have information on amino acid changes with substitution classification and position. The two trees can also be exported.



3.3 Tips and Tricks

PoSE processes one of the output files (the “rst” file) generated after running the executables baseml or codeml in PAML (http://abacus.gene.ucl.ac.uk/software/paml.html). There is a GUI version for PAML called PAML-X (available in the PAML web page). The control file (baseml.ctl, codeml.ctl) contains the specifications of the phylogenetic analysis, including the sequence file name and the tree file name. Baseml and codeml are running in the command-line and the control file needs manual editing.

The options in the control file are shown in the figure below. In order to obtain the correct “rst” file, the option Rate Ancestor needs to be set to 1. This option will produce likelihood-based reconstruction of ancestral states as part of the “rst” file that subsequently will be processed in PoSE. It is highly recommended to follow the instructions contained in the PAML manual and PAML FAQs before executing baseml and codeml.



```
      seqfile = input.pml
     treefile = tree.nwk.tre

      outfile = result.txt   *in result file
        noisy = 9   * 0,1,2,3: how much rubbish on the screen
      verbose = 1   * 1: detailed output, 0: concise output
      runmode = 0   * 0: user tree;  1: semi-automatic;  2: automatic
                    * 3: StepwiseAddition; (4,5):PerturbationNNI

        model = 7   * 0:JC69, 1:K80, 2:F81, 3:F84, 4:HKY85
                    * 5:T92, 6:TN93, 7:REV, 8:UNREST, 9:REVu; 10:UNRESTu

        Mgene = 0   * 0:rates, 1:separate; 2:diff pi, 3:diff kapa, 4:all diff

    fix_kappa = 0   * 0: estimate kappa; 1: fix kappa at value below
        kappa = 5  * initial or fixed kappa

    fix_alpha = 0   * 0: estimate alpha; 1: fix alpha at value below
        alpha = 0.3   * initial or fixed alpha, 0:infinity (constant rate)
       Malpha = 0   * 1: different alpha's for genes, 0: one alpha
        ncatG = 8   * # of categories in the dG, AdG, or nparK models of rates
        nparK = 0   * rate-class models. 1:rK, 2:rK&fK, 3:rK&MK(1/K), 4:rK&MK

        clock = 0  * 0:no clock, 1:clock; 2:local clock; 3:TipDate
        nhomo = 0   * 0 & 1: homogeneous, 2: kappa for branches, 3: N1, 4: N2
        getSE = 1   * 0: don't want them, 1: want S.E.s of estimates
 RateAncestor = 1   * (0,1,2): rates (alpha>0) or ancestral states

   Small_Diff = 7e-6
    cleandata = 1  * remove sites with ambiguity data (1:yes, 0:no)?
*         ndata = 5
*         icode = 0  * (with RateAncestor=1. try "GC" in data,model=4,Mgene=4)
*    readfpatt = 0  * read site pattern frequences instead of sequences
*  fix_blength = -1  * 0: ignore, -1: random, 1: initial, 2: fixed
        method = 0  * 0: simultaneous; 1: one branch at a time
```

4 Citing PoSE

[Paper]

5 Contact

Molecular Epidemiology and Surveillance Laboratory, Polio and Picornavirus Laboratory Branch, G-10, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Rd., N.E., Atlanta, GA 30329.

E-mail: vzt5@cdc.gov (Kun Zhao); ycv0@cdc.gov (Kelley Bullard); JJorba@cdc.gov (Jaume Jorba)