A Web Application for Efficient Analysis of Peptide Libraries

Eric Riemer Hare

February 2, 2014

Abstract

Peptide libraries have important theoretical and practical applications in the fields of biology and medicine. Despite their importance, little analysis has been done to assess the statistical properties of different libraries. This paper introduces a web application called PeLiCa, or Peptide Library Calculator, built upon the Shiny web-application framework. PeLiCa provides an easy-to-use and powerful front-end to the R package peptider, allowing users to conduct a statistical analysis of a set of pre-defined peptide library schemes, or a custom-defined scheme of their own choosing. Results are instantly displayed, to help the user make a more informed decision regarding what specific peptide library will be most useful for their particular application or research.

1 Introduction

With the introduction of the R package *peptider*, analysis of the statistical properties of various peptide libraries is now possible. However, use of this package still requires familiarity with the R language, and more general programming concepts. In this paper, I introduce a new web interface called *PeLiCa* (or Peptide Library Calculator) which provides a useable and flexible front-end to peptider. PeLiCa is designed for biologists and others working with peptide libraries, and does not require programming knowledge to conduct an analysis. PeLiCa can be accessed from the url `http://www.pelica.org`.

2 Structure

PeLiCa is a Shiny application. Shiny is a framework for writing web-applications in the R language, requiring little-to-no javascript programming knowledge. PeLiCa uses this framework to provide interactivity. For instance, when the user of PeLiCa changes a property of the peptide library, such as the encoding scheme, the results, tables, and plots will instantly update to reflect the new library. PeLiCa is currently hosted on the Glimmer server provided by the RStudio team.

3 User Interface

Similar to other Shiny applications, PeLiCa consists of three primary UI components, the Configuration Panel, the Tab Panel, and the Results Panel, each illustrated in Figure 1. The Configuration Panel is located along the left-hand column. This panel allows for various parameters of peptide libraries to be adjusted, and some configuration options relating to PeLiCa to be changed depending on user preference. The top panel is the Tab Panel, which contains various tabs corresponding to different properties of peptide libraries that can be investigated. The bottom panel is the Results panel, which will contain the results of the analysis depending on the tab and configuration options selected.

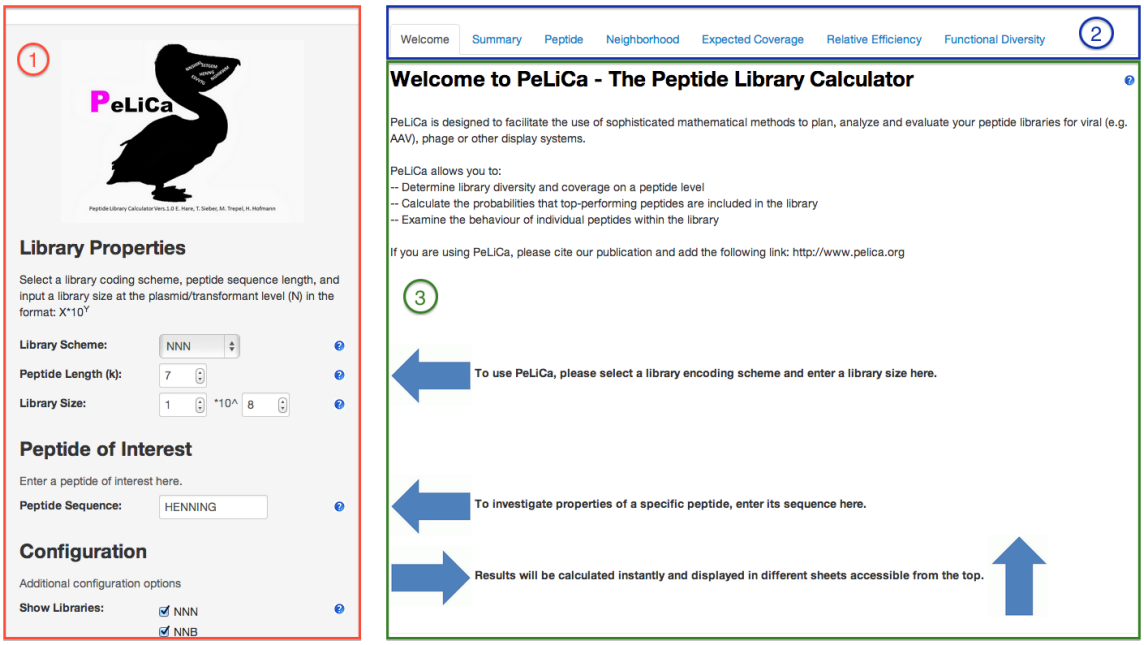

Figure 1: Screenshot of PeLiCa indicating the three primary UI components. (1) The Configuration Panel (2) The Tab Panel (3) The Results Panel.

Another important component of the user interface are the help tooltips. Throughout the application, blue question marks will be available. When the user moves their mouse cursor over these icons, helpful tooltips will appear to instruct the user on how to proceed, or provide more information about the library property being investigated.

4 Features

4.1 Configuration Panel

The first set of features available in PeLiCa involve the specification of properties of the library of interest. These features are displayed in Figure 2. Users can first select a library scheme. PeLiCa provides several built-in library schemes, the first of which is the NNN scheme, in which all four bases (Adenine, Guanine, Cytosine, and Thymine) can occur at all three positions in a particular codon. The second is the NNB scheme, where the first two positions are unrestricted, but the third position can only be three bases. The third is NNK/S, which covers both NNK and NNS schemes, with the third position restricted to two bases. Both NNK and NNS have identical statistical properties in the analysis available in PeLiCa[REF]. Finally, there are trimer-based libraries in which the codons are pre-defined. PeLiCa also includes variations of these four library types in which Cysteine is treated as a non-viable amino acid.

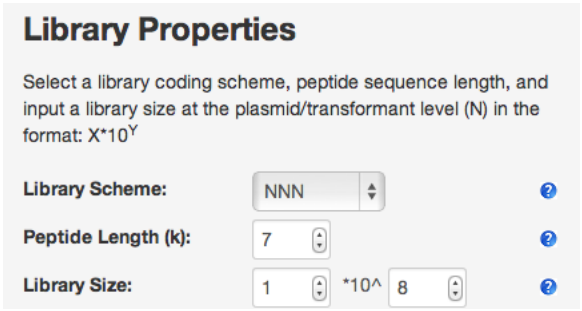

Figure 2: The Library Properties section of the Configuration Panel.

PeLiCa also has support for user-defined library schemes. If the user selects "Custom Scheme" for the library scheme, they will be presented with an upload dialog, along with a set of instructions for uploading a custom scheme. Using a custom scheme with PeLiCa requires a minimal use of programming and may be less suitable for those unfamiliar with R.

Users can also specify the peptide length and the library size. Peptide lengths can range from six to ten amino acids, with support for larger peptides coming soon. The library size is specified in scientific notation of the form $x \times 10^y$. The user will specify values for x and y in this equation. x can range from 1.0 to 9.9 in increments of 0.1, and y can range from six to 14 in increments of one, yielding a supported range of library sizes between 1.0×10^6 and 9.9×10^{14}. Support for large library sizes is also in progress.

Users can then specify a peptide of interest, as shown in Figure 3, and configure which other libraries to display in the Results Panel in Figure 4.

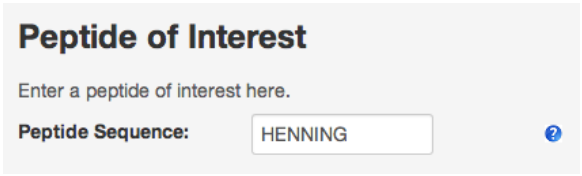

Figure 3: The Peptide of Interest section of the Configuration Panel.

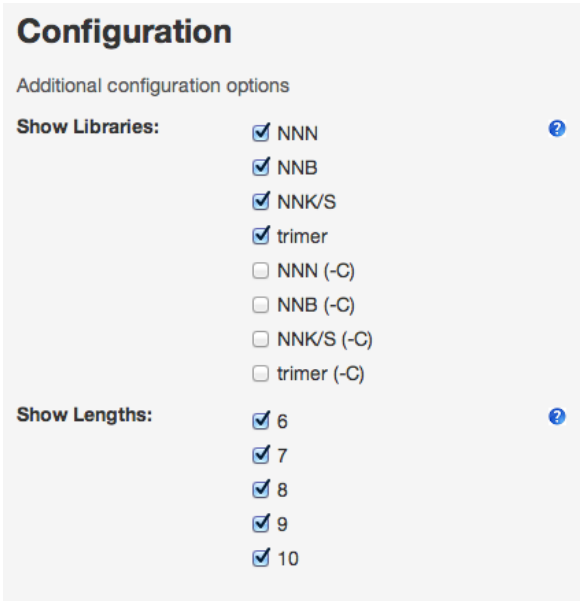

Figure 4: The Configuration section of the Configuration Panel.

4.2 Results Panel

The primary functionality for PeLiCa is available in the Results Panel, which are accessible through each of the tabs in the Tab Panel.

4.2.1 Welcome

PeLiCa begins on the Welcome tab. The Welcome tab provides information on the functionality of PeLiCa, and a quick guide for its use. The tooltip on this tab illustrates the system requirements.

4.2.2 Summary

The Summary tab contains most of the key information from the other tabs, condensed into an easy-to-digest format. First, information on your library is displayed. Some of this information includes the coverage, the peptide diversity, and the probabilities of peptide inclusion. Information on your selected peptide is displayed below this, summarizing the inclusion probability and the number of different DNA encodings of this particular peptide. Finally, a table displaying a randomly generated sample of peptides is shown at the bottom. This table includes the amino acids comprosing the peptide, the peptide class under the chosen encoding scheme, the number of DNA encodings, and the probability of inclusion in your library.

4.2.3 Peptide

The peptide tab includes information allowing investigation of the peptide selected in Figure 3. PeLiCa will display the inclusion probability of this peptide, as well as a set of boxplots illustrating the inclusion probability of all peptides. The boxplots allow for a comparison of the relative differences between different schemes and library sizes. An example of this plot is given in Figure 5. In this case, the selected library is an NNN scheme library with peptide length seven, and a library size of 10^8. The probability of inclusion is given across different library sizes, with the particular peptide of interest displayed as a circle on the plot.

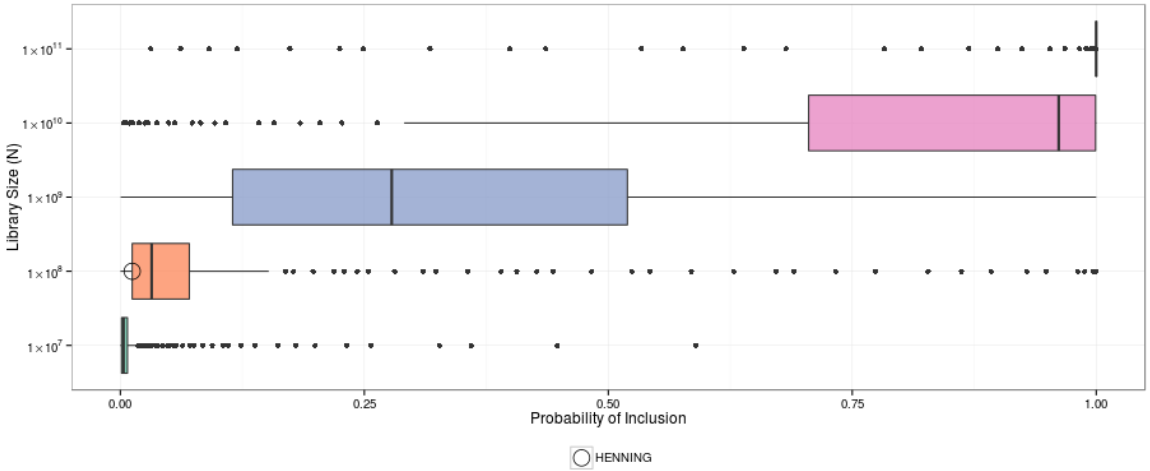

Figure 5: Boxplots displayed in PeLiCa representing the inclusion probabilities for an NNN library with peptide length seven and a library size of 1×10^8. The inclusion probability of HENNING is displayed in the plot as a circle.

4.2.4 Neighborhood

The neighborhood tab provides information on the inclusion probabilities of peptides in the degree-one and degree-two neighborhoods of the selected peptide. It also provides a range of inclusion probabilities for degree-one and degree-two neighborhoods of all peptides. As for the Peptide tab, plots of the inclusion probabilities across different library sizes and schemes is displayed to allow for a quick comparison.

4.2.5 Expected Coverage

The Expected Coverage tab displays a numerical value for the expected coverage of your library, and a plot of coverages for different schemes and library sizes. An example of the plot, using an NNN library with peptide length seven and library size 10^8, is shown in Figure 6.

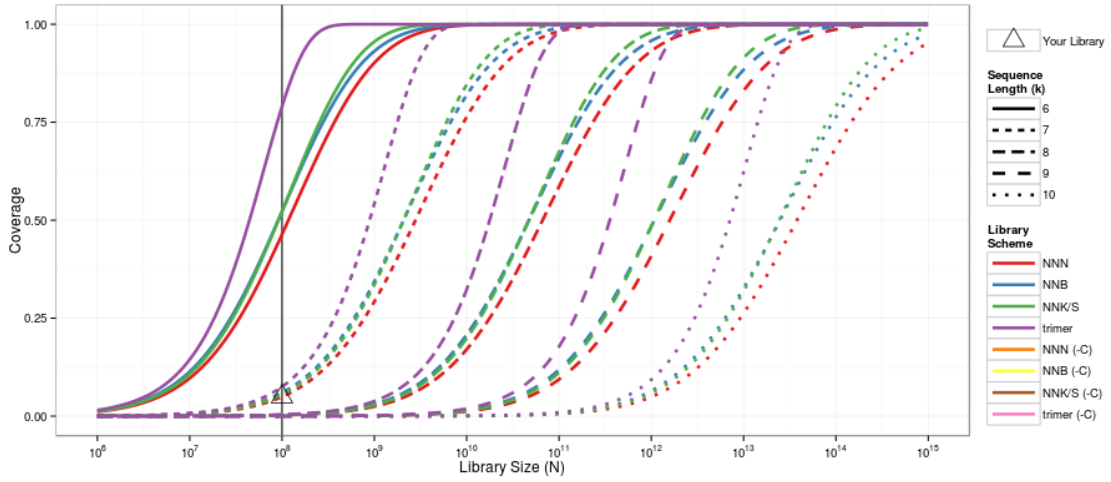

Figure 6: Plots of expected coverage for an NNN library with peptide length seven and a library size of 1×10^8

4.2.6 Relative Efficiency

Similar the the Expected Coverage tab, the Relative Efficiency tab provides a numerical value for the relative efficiency, and a comparison of the selected library to other library schemes and sizes. In conjunction with expected coverage, this allows an optimization of the cost-benefit properties of the library. As relative efficiency decreases with the library size, a library which optimizes the relative efficiency while still maintaining a desired coverage level will set a bound on the size of the library. This will help to identify a library size and scheme that has the diversity properties desired and is not prohibitively expensive.

4.2.7 Functional Diversity

The final tab in PeLiCa is the Functional Diversity tab, which displays both the functional diversity of your library, and a table of the functional diversity for other schemes and peptide lengths (Note that the functional diversity does not depend on the size of the library).

5 Further Work

A new version of PeLiCa is in currently in progress. The new version supports lower resolution monitors, includes a new framework for tooltips, and supports a wider range of peptide lengths and library sizes. This new version is currently deployed on ShinyApps at `http://erichare.shinyapps.io/pelica`.

6 Conclusion

By utilizing the R package *peptider* and the web application framework Shiny, PeLiCa allows for a powerful statistical analysis of peptide libraries. It is flexible enough to allow investigation of a wide variety of different library schemes, peptide lengths, and library sizes. Still, the application is web-based and easy to use, making the barrier of entry for those outside the field of statistics very low.

References

[1] Andreas Buja. *discreteRV: Functions to create and manipulate discrete random variables*. R package version 1.0.2. 2013.

[2] Heike Hofmann, Eric Hare, and ggobi Foundation. *peptider: Evaluation of diversity in nucleotide libraries*. R package version 0.1.2. 2013.

[3] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2013. URL: `http://www.R-project.org/`.

[4] Timo Sieber et al. "Biomathematical Description of Peptide Library Properties". In: *Bioinformatics* (2014).