

A stepwise tutorial for CoRNeA

Software required

❑ Offline tools

- **Jalview**– for alignment visualization and editing
- **R and Rstudio**– for feature generation
- **Anaconda-navigator**– for utilizing trained model for predictions

❑ Online tools

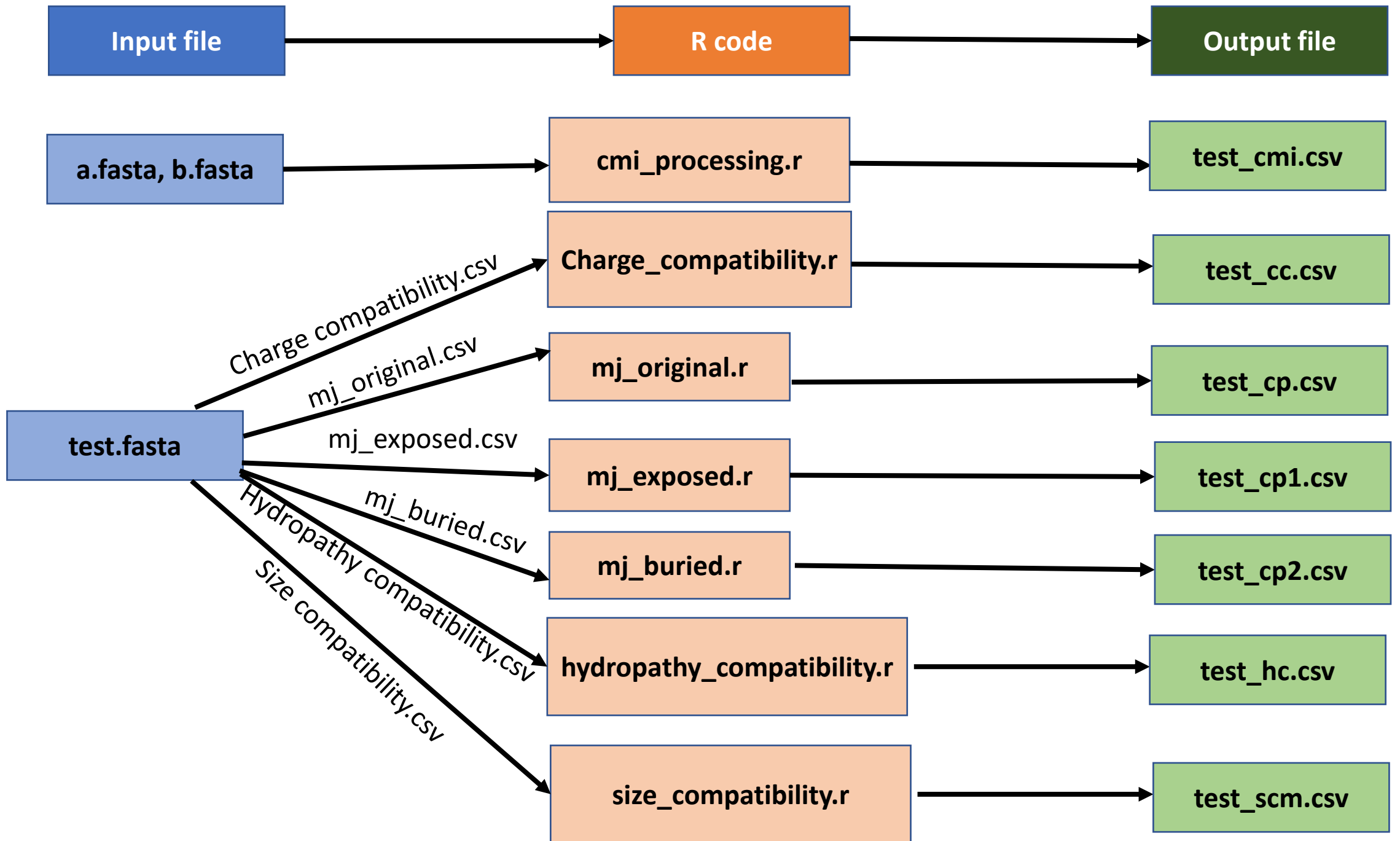
- **FaBox**– for creating fasta files for alignment to equalize species
- **PSIPRED**– for secondary structure predictions
- **SPIDER2/3**– for relative solvent accessibility predictions

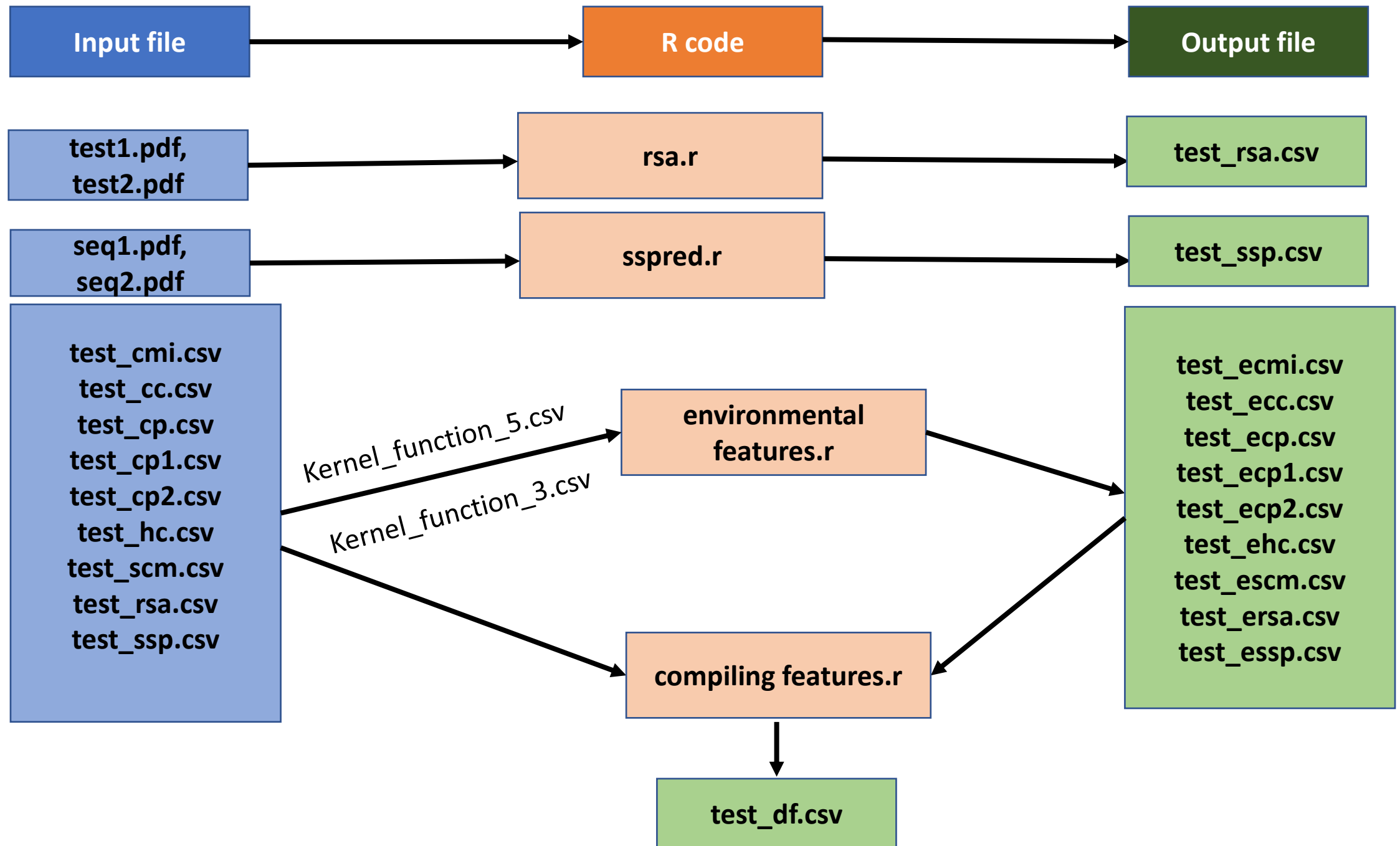
Files required for feature generation

- **a.fasta** : alignment file for Protein A
 - **b.fasta**: alignment file for Protein B
- Co-evolution feature (CMI) and intra contacts
- **test.fasta** : sequence of both Protein A and Protein B
(>1
MLK....
>2
MAL....)
- Matrix based features: CP, CP1, CP2, CCM, HCM, SCM
- **seq1.pdf, seq2.pdf**: PSIPRED files for **SSP** feature calculation for both Protein A and Protein B
 - **test1.pdf, test2.pdf**: SPIDER files for **RSA** feature calculation for both Protein A and Protein B

Feature Generation

- Create a **single folder** with all the input files and codes for feature generation (**Github**)
- Run the “.r” files in Rstudio (all dependencies are mentioned in the respective codes)
- If the dependencies are not installed, can be installed in Rstudio easily using Install plugin
- List of dependencies: **seqinr, reshape2, scales, gplots, RColorBrewer, OpenImageR, dplyr, igraph, tm, pdftools**





Prediction using random forest models

- ❑ Launch Anaconda-navigator and then Jupiter notebook (opens in web browser)
- ❑ Open file “**testing.ipynb**” from the list of files in Jupiter notebook
- ❑ Input file: **test_df.csv** (environmental features can be either from kernel_function_3 or kernel_function_5)
- ❑ Run all commands in the Jupiter notebook after changing the name of the test file
- ❑ Output file: **prediction_test.csv**

Post processing for prediction results obtained from random forest

