Understanding the content of JSON output files

Most standard analyses in HyPhy output results in JSON format, essentially a nested dictionary. This page describes the contents of each method's JSON output.

We note that these files are easily parsed for downstream use with standard scripting languages, e.g. using the `json` package in Python or `jsonlite` package in R.

Bookkeeping fields

There are several fields which appear in JSONs that are used strictly for displaying results in HyPhy Vision and have no scientific meaning. These fields, which you can safely ignore, include the following: + **`display order`**, in all analyses +

Shared fields

All standard selection analyses will have the following top-level fields:

`analysis`

This field contains information about the analysis method of interest and is comprised of the following keys:

- **`info`**, Gives method name a brief statement of its intended use
- **`version`**, Method version
- **`citation`**, Method reference
- **`authors`**, names of primary authors of method
- **`contact`**, primary author to contact
- **`requirements`**, Data required in order to execute method
- **`input`** will contain information about the inputted dataset being analyzed and is comprised of the following keys:
  - **`file name`**, Full path to the name of user-inputted alignment file
  - **`number of sequences`**, The number of sequences in the user-inputted alignment
  - **`number of sites`**, The number of sites in the user-inputted alignment
  - **`partition count`**, The number of user-specified data partitions to be used in the given analysis
  - **`trees`**, The inputted trees (with user-supplied branch lengths, if applicable). This field is itself a dictionary containing all provided trees in newick format. The key for each tree corresponds to the data partition to which it belongs, starting from 0. For example, if there is one specified partition, this field might look like the following:

    ```
    trees":{
        "0":"(t1:0.140821,(((t4:0.0918075,t9:0.0712801)Node5:0.131355,(t10:0.416582,t7:0.0476408)No
        }
    ```

    Alternatively, for two specified partitions, this field might look like the following:

    ```
       "trees":{
     "0":"(T1:0.140821,(((T4:0.0918075,T9:0.0712801)Node5:0.131355,(T10:0.416582,T7:0.0476408)Node8
     "1":"(((T3:0.0197597,T10:0.16236)Node3:0.288468,T8:0.697417)Node2:0.235357,(((T11:0.178972,T5
        }
    ```

- **fits** will contain information about each fitted model in the given analysis, and as such each method will return different fits. Most fit fields will contain the following fields:
    - **Log Likelihood**, the log likelihood of the fitted model
    - **estimated parameters**, the number of parameters estimated during likelihood optimization (i.e., not including empirically-determined parameters)
    - **AIC-c**, the small-sample AIC for the fitted model
    - **Rate distributions**, the inferred rate distribution under the fitted model. Depending on the model, this field can refer to different rates. See each fit description for an explanation of the rates provided.

Within **fits**, all methods will contain a field **Nucleotide GTR**, and most will contain a field for **Global MG94xREV** (note that this is termed **MG94xREV with separate rates for branch sets** in methods RELAX and BUSTED, and termed **Baseline MG94xREV** in aBSREL). Each of these fields will contain the following additional fields: + **Nucleotide GTR** + **Equilibrium frequencies**, a vector of nucleotide frequencies obtained empirically, in alphabetical order (A, C, G, T) + **Rate distributions**, a dictionary of inferred nucleotide substitution rates under the GTR model. Note that, for all inferences, the rate `A->G` (`G->A`) is constrained to equal 1. + **Global MG94xREV** (or analogous field in RELAX, BUSTED, aBSREL) + **Equilibrium frequencies**, a vector of codon frequencies obtained using the CF3x4 estimator, in alphabetical order (AAA, AAC, AAG,. . . , TTT) + **Rate distributions**, inferred ω rates under the fitted model. Content in this field is method-specific.

- **data partitions**