July 9 2013: qmmraxml version 2 released

What's new in qmmraxml 2.0?
1) implements an amino acid substitution selection model to the standard F+Gamma model or the cF models.
Use -S to invoke a selection model and use -x to specify a file for the amino acid frequencies under no selection.

2) in addition to the original 4 cF profiles (cF4), the new version adds a model of 9 amino acid profiles (cF9) (Sjolander et al. CABIOS 12:327-345, 1996) and a model of 20 amino acid profiles (cF20) (Quang, Gascuel, Lartillot, Bioinformatics 24:2317-2323, 2008) of amino acid site frequencies.

3) qmmraxml 2.0 can also run the empirical protein profile mixture model (CAT-C20) with F81 rate matrix (Quang, Gascuel, Lartillot, Bioinformatics 24:2317-2323, 2008)

4) a bug in searchAlgo.c in qmmraxml 1.0 as well as the original RAxML-VI-HPC was found and fixed. Also some minor changes in the tree search algorithms in searchAlgo.c to be consistent with the updates in RAxML 7.1 over RAxML VI.

5) To print site-wise log likelihood scores, add -L options on the command line.
e.g. -L sitelkh.dat will print log site likelihood to a file named sitelkh.dat.

Fsel: standard F+Gamma under selection model:
./qmmraxmlHPC -m PROTGAMMALGF -s cox2.seq -n cox2.out -S -x aafreq.equalCodons.dat

cF4:
./qmmraxmlHPC -m PROTGAMMALGF -s cox2.seq -n cox2.out -@

cF4sel: cF4 + selection model:
./qmmraxmlHPC -m PROTGAMMALGF -s cox2.seq -n cox2.out -@ -S -x aafreq.equalCodons.dat

cF9:
./qmmraxmlHPC -m PROTGAMMALGF -s cox2.seq -n cox2.out -B

cF9sel: cF9 + selection model:
./qmmraxmlHPC -m PROTGAMMALGF -s cox2.seq -n cox2.out -B -S -x aafreq.equalCodons.dat

cF20:
./qmmraxmlHPC -m PROTGAMMALGF -s cox2.seq -n cox2.out -D

cF20sel: cF20 + selection model:
./qmmraxmlHPC -m PROTGAMMALGF -s cox2.seq -n cox2.out -D -S -x aafreq.equalCodons.dat

CAT-C20: note this model uses F81 as the rate matrix
./qmmraxmlHPC -m PROTGAMMAF81 -s cox2.seq -n cox2.out -C

For running the selection models (Fsel, cF4sel, cF9sel and cF20sel), aafreq.equalCodons.dat, which includes in the release, is the supposed aa-freqs under no selection, which is obtained as the expected amino acid frequencies under equal codon frequency (frequency of all sense codons is 1/61). One may obtain aa-freqs under no selection based on the nuclotide composition at the 3rd codon positions of the dataset of interest if the coding sequence data are available. Or supply your own neutral amino acid frequency data.

For reference please cite:
Huai-Chun Wang, Karen Li, Edward Susko and Andrew J. Roger: A class frequency mixture model that adjusts for site-specific amino acid frequencies and imporves inference of protein phylogeny. BMC Evolutionary Biology 8: 331.

Huai-Chun Wang, Edward Susko and Andrew J. Roger: An amino acid substitution-selection model adjusts residue frequencies to imporve phylogenetic estimation. (in revision)

December 12 2008

qmmRAxML: Q-matrix mixture RAxML

The qmmraxml is based on RAxML-VI-HPC (version 2.2). It implements a class-frequency mixture model for ML inference of protein phylogeny. For the moment, only protein sequences can be used for qmmRAxML.

To compile the program, type:
make

To run the program with the amino acid class frequency mixture model, type:

(for tree search starting from a random tree)
./qmmraxmlHPC -m PROTGAMMAJTTF -s cox2.seq -n cox2.out -@

(for tree search starting from a user input tree)
./qmmraxmlHPC -m PROTGAMMAJTTF -s cox2.seq -t cox2.tre -n cox2.out -@

(for fixed tree)
./qmmraxmlHPC -m PROTGAMMAJTTF -s cox2.seq -t cox2.tre -f e -n cox2.out -@

To run the program with a user-defined rate matrix or user-defined amino acid class frequency models, say, myModels.dat, type:

(for tree search starting from a random tree)
./qmmraxmlHPC -m PROTGAMMAJTTF -s cox2.seq -n cox2.out -u myModels.dat

(for tree search starting from a user input tree)
./qmmraxmlHPC -m PROTGAMMAJTTF -s cox2.seq -t cox2.tre -n cox2.out -u myModels.dat

(for fixed tree)
./qmmraxmlHPC -m PROTGAMMAJTTF -s cox2.seq -t cox2.tre -f e -n cox2.out -u myModels.dat

Please refer to RAxML-VI-HPC (version 2.2.3) User Manual (http://icwww.epfl.ch/~stamatak/index-Dateien/software/RAxML-Manual.2.2.3.pdf) for details about the command line arguments.
-m specifies protein model.

-s specifies input sequence file name.
-n specifies output file name.
-t specifies starting tree.
-f e optimizes model+branch lengths for a given input tree (topology is fixed during optimization) under GTRGAMMA only.

-@ allows doing amino acid class frequency mixture model. Four class frequency vectors will be conbined with the proetin model (specified by the -m option) and the amino acid frequency vector of the whole dataset (specified by the F component of the -m option), as discussed in the Wang, Li, Susko and Roger's paper cited above.

-u specifies user-defined amnio acid substitution matrices and amino acid frequency vectors. For instance, a new rate matrix, LG, was recently reported. If you want to use LG with amino acid class frequency models, you can use myModels.dat, which is included with the release. For your own models, you should modify myModels.dat but keep the same format.

Note: the -@, -u and -L options are specific to qmmraxmlHPC. It is recommendated that the two options should not be used simultaneously.

If without -@ nor -u option, the qmmraxmlHPC simply runs like Alexandros Stamatakis's original raxmlHPC (a single rate matrix model).

An example sequence file (cox2.seq) and a tree file (cox2.tre) are included with the release.

If you have a problem with running qmmraxmlHPC, please contact Huaichun Wang (huaichun.wang@dal.ca).