The RAxML v8.2.X Manual

by Alexandros Stamatakis
Heidelberg nstitute !or Theoreti"al Studies
#uly 2$% 2$&'

Stru"ture o! this manual

. About RAxML
. (etting Hel)
. RAxML *eb+servers and (,
-. .o/nloading RAxML
-. 0om)iling RAxML
-. RAxML Likelihood -alues 1 diosyn"rasies
- . Alignment in)ut 2ile 2ormats
- . The RAxML o)tions
X. 3ut)ut 2iles
X. 0om)uting T0 and 0 values
X. Sim)le RAxML Analyses
X . A Sim)le Heterota"hous Model
X . 2re4uently Asked 5uestions

I. About RAxML

RAxML 6Randomi7ed Axelerated Maximum Likelihood8 is a)rogram !or se4uential and)arallel Maximum Likelihood based in!eren"e o! large)hylogeneti" trees. t "an also be used !or)ost+analyses o! sets o!)hylogeneti" trees% analyses o! alignments and% evolutionary)la"ement o! short reads.
t has originally been derived !rom !ast.9Aml /hi"h in turn /as derived !rom #oe 2elsentein:s dnaml /hi"h is)art o! the ;H<L ;)a"kage.

When using RAxML please cite:

A. Stamatakis= >RAxML -ersion 8= A tool !or ;hylogeneti" Analysis and ;ost+Analysis o! Large ;hylogenies>. n *Bioinformatics*% 2$&?% o)en a""ess link= htt)=@@bioin!ormati"s.ox!ordAournals.org@"ontent@early@2$&?@$&@2&@bioin!ormati"s.btu$BB.abstra"tCkeyty)eDre!1iAkeyD-TE4g,#<0."!$k;

II. Getting help

RAxML su))ort is)rovided via the RAxML (oogle grou) at= htt)s=@@grou)s.google."om@!orum@ChlDdeFG!orum@raxml 9ote that% (oogle grou)s have a sear"h !un"tionG

Thus% be!ore)osting to the RAxML google grou)=
&. Sear"h the grou) to see i! your issue has not already been dis"ussedG
2. ! you donHt /ant to get a rude re)ly% read this manual IrstG
B. Read a standard textbook about)hylogeneti"s su"h as Jiheng <angHs ex"ellent *Computational Molecular Evolution*. ! you haven not read one you should rather not be using RAxML.

Do never send emails with RAxML questions to A. Stamatakis directly!

A ste) by ste) tutorial /ith some basi" "ommands is available at= htt)=@@s"o.h+its.org@exelixis@/eb@so!t/are@raxml@handsKon.html

Additional hel)% links% little hel)er s"ri)ts and do"umentation is available at the RAxML so!t/are)age= htt)=@@s"o.h+its.org@exelixis@/eb@so!t/are@raxml@index.html

III. RAxML web-servers and Graphical User Interfaces

*hile there exist several /eb+servers that allo/ you to run RAxML% am dire"tly involved in running three o! them.

&. The 0i)res ;ortal /eb server= htt)=@@/ / /.)hylo.org@subKse"tions@)ortal@
2. The /eb+server at vital T in S/it7erland= htt)=@@embnet.vital+it."h@raxml+bb@
B. A dedi"ated server !or the Evolutionary ;la"ement Algorithm= htt)=@@e)a.h+its.org@raxml

There is no oL"ial gra)hi"al user inter!a"e su))orted by me% but a (, has been develo)ed by resear"hers at the resear"h museum in 2rank!urt% /hi"h is available here= htt)=@@sour"e!orge.net@)roAe"ts@raxmlgui@

9ote that% /ill not)rovide any sort o! su))ort !or the (,% you need to "onta"t the original authors !or this.

IV. Downloading RAxML

RAxML is o)en sour"e under (9, (;L. t is distributed via Alexis github re)ository= htt)s=@@github."om@stamatak@standard+RAxML /here you "an al/ays do/nload the most u) to date version. Make sure to **watch** the github re)ository to remain u) to date regarding "ode "hanges.

*e do not)rovide any su))ort /hatsoever !or)revious versions o! the "ode6

-ersion numbers !ollo/ the notation **x.y.z** /here **x** "hanges /ith maAor "ode reorgani7ations% **y** "hanges /hen ne/ !eatures are added and **z** "hanges /ith bug Ixes.

V. Compiling RAxML

RAxML "omes in a lot o! diMerent Navors.

t "an run se4uentially% in)arallel using either M; 6message)assing inter!a"e8 or ;Threads !or multi+"ore shared memory systems.

t also has a hybrid@"ombined ;Threads O M;)aralleli7ation that uses M; to distribute bootstra) re)li"ates or inde)endent tree sear"hes to diMerent shared memory nodes in a "luster /hile it uses ;Threads to)aralleli7e the likelihood "al"ulations o! single tree sear"hes. *e "all this "oarse grain 6M; 8 and Ine+grain 6;Threads8)arallelism.

Thus% be!ore "om)iling you need to kno/ on /hat kind o! system you intend to exe"ute RAxML. Also note that% the M; version only im)lements@oMers a subset o! the RAxML !un"tionality% it "an only distribute distin"t tree sear"hes to)ro"essors6

Another im)ortant thing to "onsider)rior to "om)iling is /hat your target)ro"essor ar"hite"ture is6 Modern x8')ro"essors are very)o/er!ul be"ause the have so+"alled ve"tor instru"tions6

.e)ending on ho/ ne/ your)ro"essor is it /ill su))ort SSEB ve"tor instru"tions% or% i! ne/er also the !aster A-X or the even !aster A-X2 ve"tor instru"tions. These instru"tions are used by RAxML to substantially a""elerate the likelihood and)arsimony "om)utations it "ondu"ts.

Thus% you should al/ays try to "om)ile the "ode in a /ay that best ex)loits the "a)abilities o! your 0;,6s8. 9ote that% even most modern la)to)s have more than & 0;,@"ore% hen"e you /ill)robably al/ays /ant to "om)ile the ;Threads version o! RAxML.

9o/ letHs have a look at the RAxML sour"e "ode% /hen you do/nload it it /ill be in a Ile "alled=

```
standard-RAxML-8.0.0.tar.gz
```

un"om)ress it by ty)ing=

```
gunzip standard-RAxML-8.0.0.tar.gz
tar xf standard-RAxML-8.0.0.tar
```

and "hange into the dire"tory that "ontains the sour"e Iles and list the "ontents=

```
cd standard-RAxML-8.0.0/
ls
```

There is a subdire"tory "alled `usefulScripts` that "ontains a "ou)le o!)erl s"ri)ts !or various little RAxML tasks.

9ext% letHs list the MakeIles=

```
ls Makefile.*
```

/hi"h /ill generate a listing looking like this=

MakeIles !or se4uential version% hybrid M;@;threads version% se4uential version !or MA0s using the "lang "om)iler% M; version% ;Threads version% ;Threads version !or MA0s using "lang that all rely on the most re"ent A-X2 ve"tor instru"tions=

```
Makefile.AVX2.gcc
Makefile.AVX2.HYBRID.gcc
Makefile.AVX2.mac
Makefile.AVX2.MPI.gcc
Makefile.AVX2.PTHREADS.gcc
Makefile.AVX2.PTHREADS.mac
```

0orres)onding MakeIles using A-X instru"tions=

```
Makefile.AVX.gcc
Makefile.AVX.HYBRID.gcc
Makefile.AVX.mac
Makefile.AVX.MPI.gcc
Makefile.AVX.PTHREADS.gcc
Makefile.AVX.PTHREADS.mac
```

0orres)onding MakeIles not using any ve"tor instru"tions 6only re4uired /hen your 0;, is more than P+' years old68. The Ile "alled= `Makefile.QuartetMPI.gcc` is a dedi"ate MakeIle that im)lements a M;)aralleli7ation o! the 4uartet evaluation !un"tionality% !or details see the se"tion des"ribing the "ommand line arguments6

```
Makefile.gcc
Makefile.HYBRID.gcc
Makefile.MPI.gcc
Makefile.PTHREADS.gcc
Makefile.PTHREADS.mac
Makefile.QuartetMPI.gcc
```

0orres)onding Make I les using SSEB instru"tions=

```
Makefile.SSE3.gcc
Makefile.SSE3.HYBRID.gcc
Makefile.SSE3.mac
Makefile.SSE3.MPI.gcc
Makefile.SSE3.PTHREADS.gcc
Makefile.SSE3.PTHREADS.mac
Makefile.SSE3.QuartetMPI.gcc
```

9o / assume that your "om)uter su)) orts A-X instru"tions% to "om)ile the se4uential version ty)e=

```
make -f Makefile.AVX.gcc
```

this / ill generate a RAxML exe"utable "alled=

```
raxmlHPC-AVX
```

9o /% to also "om)ile the ;Threads version% I rst remove all obAe"t I les generated by "om)iling the se4uential version by ty)ing=

```
rm *.o
```

and then ty)e=

```
make -f Makefile.AVX.PTHREADS.gcc
```

/ hi"h / ill generate an exe"utable "alled

```
raxmlHPC-PTHREADS-AVX
```

0om)iling all other)rogram Navors is analogous% / ith the only di L "ulty that the M; versions need m)i""% the M; "om)iler. A "ommon M; "om)iler distribution is)rovided by 3)enM; and is easy to install on , buntu Linux% !or instan"e.

When to use which Version?

The use o! the se4uential version is intended !or small to medium datasets and !or initial ex)eriments to determine a))ro)riate sear"h)arameters.
The ;Threads version / ill / ork / ell !or very long alignments% but)er!orman"e is extremely hard / are+de)endentG
Ho / ever% even !or relatively short alignments 6&%Q$$ taxa% &%2$$b)% . 9A data8 / e observed s)eedu)s o! around !a"tor '.P on an 8+"ore node.
2or a long alignment 6&2P taxa% 2$%$$$ base+)airs% . 9A8 / e observed signi I "ant su)er+linear s)eedu)s o! around &$+&& on an 8+"ore system.

Warning= Make sure to s)e"i!y the exa"t number o! 0; , s available on your system via the -T o)tion. ! you start more threads than you have "ores available% there / ill be a signi I "ant

)er!orman"e de"reaseG
The M; version is !or exe"uting really large)rodu"tion runs 6i.e. &$$ or &%$$$ bootstra)s8 on a Linux "luster. <ou "an also)er!orm multi)le in!eren"es on larger datasets in)arallel to Ind a best+kno/n ML tree !or your dataset.
2inally% the ra)id RS algorithm and the asso"iated ML sear"h have also been)aralleli7ed /ith M; .

Warning: Redu"ed !un"tionality o! M; versionG The "urrent M; version only /orks)ro)erly i! you s)e"i!y the -# or -N o)tion in the "ommand line% sin"e it has been designed to do multi)le in!eren"es or ra)id@standard RS 6bootstra)8 sear"hes in)arallelG
2or all remaining o)tions% the usage o! this ty)e o! "oarse+grained)arallelism does not make mu"h senseG

Processor Affinity and Thread Pinning with the PThreads Version

An im)ortant as)e"t i! you /ant to use the ;Threads version o! the)rogram is to Ind out ho/ your o)erating system@)lat!orm handles)ro"essor aL nity o! threads. *ithin the shared+memory or multi+"ore "ontext)ro"essor aL nity means that i! you run% !or instan"e% ? threads on a ?+/ay 0;, or ? "ores ea"h individual thread should al/ays run on the same 0;,% that is% `thread0` on `CPU0`% `thread1` on `CPU1` et".
This is im)ortant !or eL "ien"y% sin"e "a"he entries "an be "ontinuously re+used i! a thread% /hi"h /orks on the same)art o! the shared memory s)a"e% remains on the same 0;,. ! threads are moved around% !or instan"e% `thread0` is initially exe"uted on `CPU0` but then on `CPU4` et". the "a"he memory o! the 0;, /ill have to be re+Illed every time a thread is moved. *ith)ro"essor aL nity enabled%)er!orman"e im)rovements o! PS have been measured on suL "iently large and thus memory+intensive datasets.
There is a !un"tion that /ill automati"ally)in threads to 0;,s% that is% en!or"e thread aL nity% under L 9,X@,9 X. Re"ause this !un"tion might o""asionally "ause some error messages during "om)ilation due to)ortability issues it is disabled by de!ault. To enable it% you /ill need to "omment out this Nag=

```
#define _PORTABLE_PTHREADS
```

in the `axml.c` sour"e Ile.

How many Threads shall I use?

t is im)ortant to kno/ that the)arallel eL "ien"y o! the ;Threads version o! RAxML de)ends on the alignment length. 9ormally% you /ould ex)e"t a)arallel)rogram to be"ome !aster as you in"rease the number o! "ores@)ro"essors you are using. This is ho/ever not generally true% be"ause the more)ro"essors you use% the more a""umulated time they s)end /aiting !or the in)ut to be)arsed and "ommuni"ating /ith ea"h other. n "om)uter s"ien"e this)henomenon is kno/ as AmdahlHs la/ 6see htt)=@@en./iki)edia.org@/iki@AmdahlHsKla/8.

Thus% i! you run RAxML /ith B2 instead o! & thread this does not mean that it /ill automati"ally be"ome B2 times !aster% it may a"tually even be"ome slo/er. As already mentioned% the)arallel eL "ien"y% that is% /ith ho/ many threads@"ores you "an still exe"ute it eL "iently in)arallel de)ends on the alignment length% or to be more)re"ise on the number o! distin"t)atterns in your alignment. This number is)rinted by RAxML to the terminal and into the `RAxML_info.runID` Ile and look like this=

```
Alignment has 70 distinct alignment patterns
```

As a rule o! thumb Hd use one "ore@thread)er P$$ .9A site)atterns% i.e.% i! you have less% than itHs)robably better to Aust use the se4uential version. Single+gene .9A alignments /ith around &$$$ sites "an be analy7ed /ith 2 or at most ? threads. Thus% the more)atterns your alignment has% the

more threads@"ores you "an use e L "iently.

Also note that% e L "ien"y varies de)ending on the ty)e o! data% or more)re"isely% the number o! states in your data 6e.g.% ? in . 9A% 2$!or)roteins8. The more states you have% the !e/er site)atterns you need)er thread@"ore !or RAxML to exe"ute e L "iently in)arallel. This is be"ause there is more "om)utational /ork 6more mathemati"al o)erations8 to be done)er site)attern as the number o! states in"reases. *ith)rotein data you thus re4uire less sites)er thread !or RAxML to run e L "iently. Thus% an MSA /ith &$$$)rotein site)atterns may still run e L "iently /hen using &' "ores@threads.
2inally%)arallel e L "ien"y also de)ends on the rate heterogeneity model. *ith the (AMMA model that entails more "om)utations you "an thus ty)i"ally use more threads than /ith the 0AT model that only exe"utes a))roximately T o! the "om)utations the (AMMA model re4uires.
9ote that% these are Aust very rough rules o! thumb% you need to test /hat the o)timal setting is !or your datasetG

VI. RAxML Likelihood Values & Idiosyncrasies

t is very im)ortant to note that the likelihood values)rodu"ed by RAxML "an not be dire"tly "om)ared to likelihood values o! other ML)rograms. Ho/ever% the likelihood values o! the "urrent version are very similar to those obtained by other)rograms /ith res)e"t to)revious releases o! RAxML 6usually bet/een U@V &.$ log likelihood units o! those obtained e.g. by ;H<ML or (ARL 8.

The above o! "ourse re!ers to evaluating the likelihood on identi"al trees% in terms o! tree sear"hes the)rograms /ill yield diMerent tree to)ologies and hen"e diMerent likelihoods in most "ases any/ay.

9ote% that the deviations bet/een ;H<ML@RAxML and (ARL likelihood values "an sometimes be larger be"ause (ARL uses a slightly diMerent)ro"edure to "om)ute em)iri"al base !re4uen"ies 6.erri"k J/i"kl%)ersonal "ommuni"ation% many years ago8 /hile the method in RAxML is exa"tly the same as im)lemented in ;H<ML.

These deviations bet/een RAxML@;H<ML on the one side and (ARL on the other side a))ear to be larger on long multi+gene alignments. Also note that% even likelihood values obtained by diMerent RAxML versions% es)e"ially should not be dire"tly "om)ared /ith ea"h other either.

This is due to !re4uent "ode and data stru"ture "hanges in the likelihood !un"tion im)lementation and model)arameter o)timi7ation)ro"eduresG
Thus% i! you /ant to "om)are to)ologies obtained by distin"t ML)rograms /ith res)e"t to their likelihood% make sure that you o)timi7e bran"h lengths and model)arameters o! Inal to)ologies /ith one and the same)rogram.
This "an be done by either using the res)e"tive RAxML o)tion 6-f e8 or% e.g.% the "orres)onding o)tion in ;H<ML 6see htt)=@@/ / /.atg"+mont)ellier.!r@)hyml@8.

Differences in Likelihood scores

n theory all ML)rograms im)lement the same mathemati"al !un"tion and should thus yield the same likelihood s"ore !or a Ixed model and a given tree to)ology.
Ho/ever% i! /e try to im)lement a numeri"al !un"tion on a Inite ma"hine /e /ill unavoidably obtain rounding errors. Even i! /e "hange the se4uen"e 6or i! it is "hanged by the "om)iler% /hi"h it usually is8 o! some o)erations a))lied to Noating)oint or double)re"ision arithmeti"s in our "om)uter /e /ill)robably get diMerent results

n my ex)eriments have observed diMeren"es among Inal likelihood values bet/een (ARL % 5;99 % ;H<ML% RAxML 6every)rogram sho/ed a diMerent value8.

<ou "an also ex)eriment /ith this by removing the g"" o)timi7ation Nag –O2 in one o! the RAxML MakeIles. This /ill yield mu"h slo/er "ode% that is in theory mathemati"ally e4uivalent to the o)timi7ed "ode% but /ill yield slightly diMerent likelihood s"ores% due to re+ordered Noating)oint o)erations.

My)ersonal o)inion is that the to)ologi"al sear"h 6number o! to)ologies analy7ed8 is mu"h more im)ortant than exa"t likelihood s"ores to obtain *good* Inal ML trees.
Es)e"ially on large trees /ith more than &%$$$ se4uen"es the diMeren"es in likelihood s"ores indu"ed by the to)ology are usually so large% that a very rough)arameter o)timi7ation /ith an e)silon 6RAxML –e o)tion8 o! & log likelihood unit% i.e.% i! the diMeren"e e)silon bet/een t/o su""essive model)arameter o)timi7ation iterations is smaller than &.$ /e sto) the o)timi7ation% /ill already "learly sho/ the diMeren"es.

9ote that% i! you)er!orm a bootstra) analysis you don:t need to /orry too mu"h about likelihood values any/ay% sin"e usually you are only interested in the bootstra))ed to)ologies.

The CAT model of rate heterogeneity

The name o! this model has "aused a lot o! "on!usion be"ause there is a 0AT model also im)lemented in ;hyloRayes 6see htt)=@@megasun.b"h.umontreal."a@;eo)le@lartillot@/ / /@index.htm8

,n!ortunately% /as not a/are o! this /hen introdu"ed the 0AT model in RAxML% the 0AT model in RAxML is something "om)letely diMerentG Ho/ever% de"ided not to "hange the name !or ba"k/ard "om)atibility su"h that ne/ RAxML version kee) /orking /ith old /ra))er s"ri)ts.

Warning: t is not a good idea to use the 0AT a))roximation o! rate heterogeneity on datasets /ith less than P$ taxa. n general there /ill not be enough data)er alignment "olumn available to reliably estimate the)er+site rate)arameters.

0AT has been designed to a""elerate the "om)utations on large datasets /ith many taxaG ;lease read the res)e"tive)a)er
htt)=@@ieeex)lore.ieee.org@x)l@login.As)Ct)D1arnumberD&'BQPBP1urlDhtt)SBAS22S22ieeex)lore.ieee.orgS22x)lsS22absKall.As)SB2arnumberSB.&'BQPBP
to understand ho/ 0AT /orks% /hat the rate "ategories are 6they are "on"e)tually diMerent !rom the dis"rete rate "ategories o! the (AMMA model8% and /hat the limitations o! this method are.

The GTRCAT a))roximation is a "om)utational /orkWaround !or the /idely used (eneral Time Reversible model o! nu"leotide substitution under the (amma model o! rate heterogeneity. 0AT servers the analogous)ur)ose% that is% to a""ommodate sear"hes that in"or)orate rate heterogeneity.
The a!orementioned)a)er des"ribes /hat GTRCAT is and /hy don:t like GTRGAMMA des)ite the !a"t that it is a beauti!ul (reek letter.

The main idea behind GTRCAT is to allo/ !or integration o! rate heterogeneity into)hylogeneti" analyses at a signiI"antly lo/er "om)utational "ost 6about ? times !aster8 and memory "onsum)tion 6? times lo/er8.
Essentially% GTRCAT re)resents a rather un+mathemati"al *quick & dirty* a))roa"h to ra)idly navigate into)ortions o! the tree s)a"e% /here the trees s"ore /ell under GTRGAMMA.
Ho/ever% due to the /ay individual rates are o)timi7ed and assigned to rate "ategories in GTRCAT% the likelihood values "om)uted under GTRCAT are "om)letely meaningless.

Warning: never "om)are alternative tree to)ologies using their 0AT+based likelihood s"oresG

<ou /ill)robably obtain a biased assessment o! trees. This is the reason /hy GTRCAT is "alled

a))roximation instead o! model. The same a))lies to the 0AT a))roximation /hen used /ith AA data or any other data ty)e that RAxML su))orts.

n general% the 0AT a))roximation o! rate heterogeneity /orks very /ell on datasets /ith more than P$ taxa !or "ondu"ting tree sear"hes under a model that a""ommodates rate heterogeneity among sites. n other /ords% i! you s"ore the trees obtained under 0AT using (AMMA you /ill usually obtain e4ually good trees as /ith !ull sear"hes under (AMMA at a substantially lo/er "om)utational "ost. Also% (AMMA may not /ork !or numeri"al reasons on very large trees /ith more than &$%$$$ taxa 6see htt)=@@/ / /.biomed"entral."om@&?X&+2&$P@&2@?X$8.

Another detailed study o! a 0AT+like model is "ondu"ted in the 2astTree+2)a)er% see htt)=@@/ / /.)losone.org@arti"le@in!o=doiS22&$.&BX&S22Aournal.)one.$$$Q?Q$

VII. Alignment input File Formats

Alignment & Tree Input Formats

Alignments: The in)ut alignment !ormat o! RAxML is relaxed interleaved or se4uential ;H<L ; or 2ASTA. *Relaxed* means that se4uen"e names "an be o! variable length bet/een & u) to 2P' "hara"ters.
! you need longer taxon names you "an ada)t the "onstant `#define nmlngth 256` in the sour"e !le `axml.h` a))ro)riately.
Moreover% RAxML is less sensitive /ith res)e"t to the ;H<L ; !ormatting 6tabs% insets% et"8 o! interleaved ;H<L ; !les.

Trees: The in)ut tree !ormat is 9e/i"k% the RAxML in)ut trees must not al/ays be "om)rehensive% i.e.% need not "ontain all taxa o! the alignment.
See htt)=@@evolution.geneti"s./ashington.edu@)hyli)@ne/i"ktree.html !or details on the 9e/i"k !ormat.

Alignment Error Checking

Many alignments need to be "he"ked !or the !ollo/ing errors@insu L"ien"ies be!ore running an analysis /ith RAxML or any other)hylogeneti" in!eren"e)rogram.

RAxML automati"ally analy7es the alignment and "he"ks !or the !ollo/ing errors=

&. denti"al Se4uen"e name6s8 a))earing multi)le times in an alignment% this "an easily ha))en /hen you ex)ort a standard ;H<L ; !le !rom some tool /hi"h trun"ates the se4uen"e names to 8 or &$ "hara"ters.
2. denti"al Se4uen"e6s8 that have diMerent names but are exa"tly identi"al. This mostly ha))ens /hen you ex"luded some hard+to+align alignment regions !rom your alignment and does not make sense to use.
B. ,ndetermined 0olumn6s8 that "ontain only ambiguous "hara"ters that /ill be treated as missing data% i.e. "olumns that entirely "onsist o! `X, ?, *, –` !or AA data and `N, O, X, ?, –` !or .9A data 6analogous !or other data ty)es8
?. ,ndetermined Se4uen"e6s8 that "ontain only ambiguous "hara"ters 6see above8 that /ill be treated as missing data.

;rohibited 0hara"ter6s8 in taxon names are names that "ontain any !orm o! /hites)a"e "hara"ter% like blanks% tabulators% and "arriage returns% as /ell as one o! the !ollo/ing)rohibited "hara"ters= `:` or `()` or `[]`.

n "ase that RAxML dete"ts identi"al se4uen"es and@or undetermined "olumns and /as exe"uted% e.g.% /ith -n alignmentName it /ill automati"ally /rite an alignment Ile "alled alignmentName.reduced /ith identi"al se4uen"es and@or undetermined "olumns removed.

! this is dete"ted !or a)artitioned model analysis a res)e"tive model Ile modelFileName.reduced /ill also be /ritten. n "ase RAxML en"ounters identi"al se4uen"e names or undetermined se4uen"es or illegal "hara"ters in taxon names it /ill exit /ith an error and you /ill have to Ix your alignment.

VIII. The RAxML options

The single by !ar most im)ortant "ommand is the RAxML hel) o)tion that dis)lays all o)tions. also !re4uently use it be"ause "an not remember them all. n the !ollo/ing /ill assume that /e are Aust using the se4uential unve"tori7ed "ode% that is% the raxmlH;0 exe"utable.

/ill dis"uss ea"h o)tion and)rovide a sim)le usage exam)le !or it.

Thus% to get on+line hel) ty)e=

```
raxmlHPC -h
```

and you /ill get the !ollo/ing% very long listing% that /ill be dis"ussed at length belo/=

```
raxmlHPC    -s sequenceFileName -n outputFileName -m substitutionModel
            [-a weightFileName] [-A secondaryStructureSubstModel]
            [-b bootstrapRandomNumberSeed] [-B wcCriterionThreshold]
            [-c numberOfCategories] [-C] [-d] [-D]
            [-e likelihoodEpsilon] [-E excludeFileName]
            [-f a|A|b|B|c|C|d|D|e|E|F|g|G|h|H|i|I|j|J|k|m|n|N|o|p|q|r|R|s|S|t|T|u|
            v|V|w|W|x|y]
            [-F]
            [-g groupingFileName] [-G placementThreshold] [-h] [-H]
            [-i initialRearrangementSetting]
            [-I autoFC|autoMR|autoMRE|autoMRE_IGN]
            [-j] [-J MR|MR_DROP|MRE|STRICT|STRICT_DROP|T_<PERCENT>] [-k] [-K]
            [-L MR|MRE|T_<PERCENT>] [-M]
            [-o outGroupName1[,outGroupName2[,...]]][-O]
            [-p parsimonyRandomSeed] [-P proteinModel]
            [-q multipleModelFileName] [-r binaryConstraintTree]
            [-R binaryModelParamFile] [-S secondaryStructureFile]
            [-t userStartingTree]
            [-T numberOfThreads] [-u] [-U] [-v] [-V] [-w outputDirectory]
            [-W slidingWindowSize]
            [-x rapidBootstrapRandomNumberSeed][-X][-y]
            [-Y quartetGroupingFileName|ancestralSequenceCandidatesFileName]
            [-z multipleTreesFile]
            [-#|-N numberOfRuns|autoFC|autoMR|autoMRE|autoMRE_IGN]
            [--mesquite][--silent][--no-seq-check][--no-bfgs]
            [--asc-corr=stamatakis|felsenstein|lewis]
            [--flag-check]
            [--auto-prot=ml|bic|aic|aicc ]
            [--epa-keep-placements=number]
            [--epa-accumulated-threshold=threshold]
            [--epa-prob-threshold=threshold]
```

```
        [--JC69][--K80]
        [--set-thread-affinity]
        [--bootstop-perms=number]
        [--quartets-without-replacement]
        [--print-identical-sequences]
```

-a Specify a column weight file name to assign individual weights to each column of the alignment. Those weights must be integers separated by any type and number of white-spaces within a separate file.

n addition% there must o! "ourse be as many /eights as there are "olumns in your alignment. The "ontents o! an exam)le /eight Ile "ould look like this=

```
5 1 1 2 1 1 1 1 1 1 1 2 1 1 3 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 4 1 1 1 4 1 1
```

Example: raxmlHPC -a wgtFile -s alg.phy -p 12345 -m GTRCAT -n TEST

-A Specify one of the secondary structure substitution models implemented in RAxML. The same nomenclature as in the PHASE manual is used, available models: S6A, S6B, S6C, S6D, S6E, S7A, S7B, S7C, S7D, S7E, S7F, S16, S16A, S16B

DEFAULT: 16-state GTR model (S16)

9ote that%)artitioning does not /ork /ith se"ondary stru"ture models. That is a se"ondary stru"ture "an only be su)erim)osed to a single)artition. Also note that% you need o! "ourse to also s)e"i!y a Ile de Ining the se"ondary stru"ture via the -s o)tion.

Example: raxmlHPC -S secondaryStructureFile -s alg.phy -A S7D -p 12345 -m GTRGAMMA -n TEST

-b Specify an integer number (random seed) and turn on bootstrapping

DEFAULT: OFF

This o)tion allo/s you to turn on non+)arametri" bootstra))ing. To allo/ !or re)rodu"ibility o! runs in the se4uential)rogram% you have to s)e"i!y a random number seed% e.g.% -b 123476.
9ote ho/ever% that)arallel bootstra)s /ith the)arallel version raxmlH;0+M; are not re)rodu"ible des)ite the !a"t that you s)e"i!y a random number seed.

Example: raxmlHPC -b 12345 -p 12345 -# 100 -s alg -m GTRCAT -n TEST

-B specify a floating point number between 0.0 and 1.0 that will be used as cutoff threshold for the MR-based bootstopping criteria. The recommended setting is 0.03.

DEFAULT: 0.03 (recommended empirically determined setting)

This setting allo/s to s)e"i!y a threshold !or the so+"alled bootsto))ing "riteria that /ill automati"ally determine i! you have "ondu"ted enough bootstra) re)li"ate sear"hes !or obtaining stable su))ort values. 9ote that% this only has an eMe"t i! you use the bootsto))ing "riteria that rely on building maAority rule "onsensus trees !or determining "onvergen"e. The o)tion /ill not have an eMe"t /hen the W not re"ommended W !re4uen"y sto))ing "riterion is being used. ;lease also read and "ite the "orres)onding)a)er=

htt)=@@link.s)ringer."om@"ha)ter@&$.&$$X@QX8+B+'?2+$2$$8+XK&BF)age+&

Example: raxmlHPC -B 0.02 -b 12345 -p 12345 -# AUTOMR -s alg -m GTRCAT -n TEST

-c Specify number of distinct rate categories for RAxML when model of rate heterogeneity is set to CAT.
Individual per-site rates are categorized into numberOfCategories rate categories to accelerate computations.

DEFAULT: 25

Example: raxmlHPC -c 40 -p 12345 -s alg -m GTRCAT -n TEST

Warning: Note that the setting of -c has no effect whatsoever on the number of discrete rate categories that are used to approximate the Gamma distribution of rate heterogeneity! RAxML always uses 4 discrete rate categories to approximate Gamma!

-C Enable verbose output for the "-L" and "-f i" options. This will produce more, as well as more verbose output files

DEFAULT: OFF

The above o)tion /ill generate verbose out)ut and additional out)ut Iles /hen RAxML is used to "om)ute the T0 and 0 measures as introdu"ed by Sali"hos and Rokas=
htt)=@@/ / /.nature."om@nature@Aournal@v?QX@nX??Q@!ull@nature&2&B$.html
The method is des"ribed in more detail in the !ollo/ing)a)er /rote /ith Sali"hos and Rokas=
htt)=@@mbe.ox!ordAournals.org@"ontent@early@2$&?@$2@$X@molbev.msu$'&.abstra"tkeyty)eDre!1iAkeyD 'P2u(9x$H7R23/

Example: raxmlHPC -L -m GTRCAT -L MRE -z treeSet -n TEST

-d start ML optimization from random starting tree

DEFAULT: OFF

This o)tion allo/s you to start the RAxML sear"h /ith a "om)lete random starting tree instead o! the de!ault randomi7ed ste) /ise addition Maximum ;arsimony starting tree. 3n smaller datasets 6around &$$W2$$ taxa8 it has been observed that this might sometimes yield to)ologies o! distin"t lo"al likelihood maxima /hi"h better "orres)ond to em)iri"al ex)e"tations.
t has also been observed that% this sometimes yield better W more diverse W starting trees !or the analysis o! broad)hylogenomi" alignments that have a very strong)hylogeneti" signalG

Example: raxmlHPC -d -p 12345 -s alg -m GTRGAMMA -n TEST

-D ML search convergence criterion. This will break off ML searches if the relative Robinson-Foulds distance between the trees obtained from two consecutive lazy SPR cycles is smaller or equal to 1%. Usage recommended for very large datasets in terms of taxa.
On trees with more than 500 taxa this will yield execution time improvements of approximately 50% while yielding only slightly worse trees.

DEFAULT: OFF

* hen enabling this o)tion% RAxML / ill sto) ML and standard% slo / bootstra) sear"hes early% / hen the R2 distan"e bet / een the trees generated by t / o "onse"utive "y"les o! S;R 6Subtree ;runing Re+(ra!ting8 moves is smaller than &S. This leads to substantial s)eed im)rovements / hile the de"rease in log likelihood s"ores is only very small. The o)tion has been tested on several datasets and the results have been in"luded in the !ollo / ing book "ha)ter=
A. Stamatakis= >;hylogeneti" Sear"h Algorithms !or Maximum Likelihood>. n M. Elloumi% A.< Jomaya% editors. Algorithms in 0om)utational Riology= te"hni4ues% A))roa"hes and A))li"ations% P?X+PXX% #ohn * iley and Sons% 2$&&.

Example: `raxmlHPC -D -p 12345 -s alg -m GTRCAT -n TEST`

`-e` set model optimization precision in log likelihood units for final optimization of tree topology

`DEFAULT: 0.1` for models not using proportion of invariant sites estimate
`0.001` for models using proportion of invariant sites estimate

This allo / s you to s)e"i!y u) to / hi"h likelihood diMeren"e% the model)arameters / ill be o)timi7ed / hen RAxML uses the (AMMA model o! rate or / hen you Aust evaluate a trees / ith the `-f e` o)tion or a bun"h o! trees / ith the `-f n` o)tion or similar o)tions. This has sho / n to be use!ul to 4ui"kly evaluate the likelihood o! a bun"h o! large Inal trees o! more than &%$$$ taxa be"ause it / ill run mu"h !aster.
ty)i"ally use e.g. `-e 1.0` or `-e 2.0` in order to ra)idly "om)are distin"t Inal tree to)ologies based on their likelihood values.
9ote that% to)ology+de)endent likelihood+diMeren"es are ty)i"ally !ar larger than &.$ or 2.$ log likelihood units. The de!ault setting is $.& log likelihood units / hi"h)roves to be su L "ient in most)ra"ti"al "ases.

Example: `raxmlHPC -f e -e 0.00001 -s alg -t tree -m GTRGAMMA -n TEST`

`-E` specify an exclude file name, that contains a specification of alignment positions you wish to exclude.
Format is similar to Nexus, the file shall contain entries like "100-200 300-400", to exclude a single column write, e.g., "100-100", if you specify a partition file via “-q”, an appropriately adapted partition file will also be written.

This o)tion / ill Aust make RAxML / rite a redu"ed alignment Ile / ithout the ex"luded "olumns that "an subse4uently be used !or the analysis you a"tually / ant to "ondu"t.
! you use a)artitioned model% an a))ro)riately ada)ted model Ile / ill also be / ritten. 9ote that% ex"luding sites / ith RAxML is a **two-step**)ro"edure. <ou / ill Irst have to invoke RAxML on"e using the `-E` o)tion to generate a ;H<L ; !ormatted Ile / ith the desired sites ex"luded. Then you / ill have to invoke RAxML again on the alignment Ile that / as / ritten 6"alled `alg.excludeFile` in the exam)le belo / 8 to do an a"tual tree sear"h% !or instan"e.

Example: `raxmlHPC -E excludeFile -s alg -m GTRCAT -q part -n TEST`

n this "ase the alignment Ile / ith "olumns ex"luded / ill be named `alg.excludeFile` and the)artition Ile / ith the s)e"iIed "olumns ex"luded is "alled `part.excludeFile`

! you / ant to ex"lude sites &$$+&QQ 6note that% the borders% that is% sites &$$ and &QQ are in"luded and / ill be removed8 and sites 2$$+2QQ the "orres)onding ex"lude Ile W a)lain

text Ile W /ould "ontain the t/o lines belo/=

```
100-199
200-299
```

-f select algorithm:

The +! o)tion is a very)o/er!ul o)tion be"ause% in many "ases% it allo/s you to sele"t /hat kind o! algorithm RAxML shall exe"ute. ! you donHt s)e"i!y -f at all RAxML /ill exe"ute the standard hill "limbing algorithm 6-f d o)tion8The individual o)tions are listed belo/.

-f a rapid Bootstrap analysis and search for best-scoring ML tree in one program run

Tell RAxML to "ondu"t a ra)id Rootstra) analysis and sear"h !or the best+s"oring ML tree in one single)rogram run.

Example: raxmlHPC -f a -p 12345 -s alg -x 12345 -# 100 -m GTRCAT -n TEST

-f A compute marginal ancestral states on a ROOTED reference tree provided via -t

This o)tion allo/s you to "om)ute marginal an"estral states@se4uen"es on a given% Ixed% and rooted re!eren"e tree. ! you donHt kno/ /hat marginal an"estral states are)lease read Jiheng <angHs book on 0om)utational Mole"ular Evolution.

Example: raxmlHPC -f A -t testTree -s testData -m GTRGAMMA -n TEST

A small)hyli)+!ormatted test alignment /ould look like this=

```
4 4
t1 ACGT
t2 AATT
t3 ATAT
t4 CCGT
```

and a test tree=

```
((t1,t2),(t3,t4));
```

RAxML /ill then out)ut the rooted binary tree again% but /ith inner node labels that must be used to asso"iate an"estral se4uen"es /ith their "orres)onding)ositions in the tree. t also /rites t/o Iles% one "ontaining the marginal)robabilities !or ea"h inner node label as /ell as one "ontaining guesses 6by taking the maximum)robability8 !or the a"tual an"estral se4uen"es.

-f b draw bipartition information on a tree provided with -t (typically the best-known ML tree) based on multiple trees (e.g., from a bootstrap) in a file specified by -z

Example: raxmlHPC -f b -t ref -z trees -m GTRCAT -n TEST

-f B optimize br-len scaler and other model parameters (GTR, alpha, etc.) on a tree provided with -t. The input tree passed via -t needs to contain branch

lengths. The branch lengths will not be optimized, just scaled by a single common value.

This is a slightly idiosyn"rati" o)tion that /as mainly develo)ed to /ork on "onAun"tion /ith the ;artition2inder tool 6see= htt)=@@/ / /.robertlan!ear."om@)artitionInder@8.
The "on"e)t o! bran"h length s"alers is similar to the one im)lemented in MrRayes. That is% /e assume that all)artitions have the same bran"h lengths)ro)ortionally to ea"h other. Ho/ever% !or ea"h)artition *p* /e do a maximum likelihood estimate o! a single s"aling !a"tor *s[p]* by /hi"h /e multi)ly the bran"h lengths. Essentially this re)resents a)arametri7ation trade+oM bet/een estimating a se)arate set o! bran"h lengths !or ea"h)artition 6heavy in"rease in the number o!)arameters Y danger o! over+)arametri7ation8 and doing one "ommon bran"h length estimate a"ross all)artitions 6)ossible under+)arametri7ation8. 9ote that% this /orks only !or evaluating a given% Ixed tree /ith bran"h lengths and not !or tree sear"hesG

Example: raxmlHPC -s alg.phy -t tree -m GTRGAMMA -f B -q part -n TEST

-f c check if the alignment can be properly read by RAxML

Example: raxmlHPC -f c -m GTRCAT -s alg -n TEST

-f C ancestral sequence test for my PhD student Jiajie Zhang, users will also need to provide a list of taxon names via -Y separated by white-spaces

This is an algorithm /e tried to im)lement !or)ost+analy7ing viral)hylogenies. The idea /as to build a test that "an determine i! "ertain se4uen"es in the viral tree are truly an"estral% that is% i! they should essentially be lo"ated at or right next to an inner node o! the tree.

2or this% /e built the !ollo/ing test@algorithm=

(iven a Ixed)hylogeneti" tree 6usually the best+kno/n ML tree8 !or a set o! se4uen"es and% given a list o! "andidate an"estral se4uen"es via -Y% !or ea"h o! these "andidate se4uen"es /e "om)are the likelihood o! the Ixed tree and the given 6o)timal8 bran"h length !or that)utative an"estral se4uen"e /ith the likelihood o! the Ixed tree and an essentially 7ero bran"h length !or the an"estral "andidate.
n !a"t% /e a))ly the test to all three bran"hes the "andidate an"estral se4uen"e is atta"hed to% that is% the bran"h it is dire"tly atta"hed to% as /ell as the t/o bran"hes to the right and le!t o! the node to /hi"h the "andidate an"estral se4uen"e is atta"hed to.

The likelihoods o! the diMerent bran"h length "onIgurations 6the three almost+7ero bran"h lengths8 are "om)ared using the Shimodaira+Haseg/a test to the original 6o)timal8 likelihood. The Shimodaira+Hasega/a test might not be the most modern test available but /as suL "ient !or "ondu"ting initial tests. t turned out that the test /orked very /ell on simulated data% but not at all !or real data. Thus% /e de"ided to abandon !urther testing% but nonetheless kee) the o)tion in RAxML.

Example: raxmlHPC -f C -Y candidateTaxonNames -m GTRGAMMA -t tree -s alg -n TEST

-f d new rapid hill-climbing

DEFAULT: ON

This is the de!ault RAxML tree sear"h algorithm and is substantially !aster than the original

sear"h algorithm. t takes some short"uts% but yields trees that are almost as good as the ones obtained !rom the !ull sear"h algorithm. The algorithm is des"ribed and assessed in this)a)er here= htt)=@@link.s)ringer."om@arti"le@&$.&$$X@s&&2'P+$$X+$$'X+?F)age+&

Example: raxmlHPC -f d -m GTRCAT -p 12345 -s alg -n TEST

-f D execute one or more rapid hill-climbing searches that will also generate RELL bootstraps

This is an im)lementation o! the te"hni4ues introdu"ed in the !ollo/ing)a)er= htt)=@@/ /.n"bi.nlm.nih.gov@)ubmed@2B?&8BQX Some initial tests have sho/n that the RAxML im)lementation yields results that are /ell "orrelated /ith standard bootstra) values. ,sing this o)tion is re"ommended /hen you have very large trees and don#t have the resour"es@time !or "om)uting bootstra) re)li"ates.

Example: raxmlHPC -f D -m GTRCAT -p 12345 -s alg -n TEST

-f e optimize model parameters+branch lengths for given input tree

<ou need to)rovide an in)ut tree via the -t o)tion.

Example: raxmlHPC -f e -t ref -m GTRGAMMA -s alg -n TEST

-f E execute very fast experimental tree search, at present only for testing

This o)tion /ill exe"ute a very !ast tree sear"h algorithm that /ill not try as hard to o)timi7e the likelihood. t is intended !or very large trees and !ollo/s a similar logi" as the 2astTree)rogram= htt)=@@meta.mi"robesonline.org@!asttree@

Example: raxmlHPC -f E -m GTRCAT -p 12345 -s alg -n TEST

-f F execute fast experimental tree search, at present only for testing

This o)tion /ill exe"ute a !ast tree sear"h algorithm that /ill not try as hard to o)timi7e the likelihood. t is intended !or very large trees and !ollo/s a similar logi" as the 2astTree)rogram= htt)=@@meta.mi"robesonline.org@!asttree@

Example: raxmlHPC -f F -m GTRCAT -p 12345 -s alg -n TEST

-f g compute per site log Likelihoods for one ore more trees passed via -z and write them to a file in treepuzzle format that can be read by CONSEL
The model parameters will be estimated on the first tree only!

Example: raxmlHPC -f g -s alg -m GTRGAMMA -z trees -n TEST

-f G compute per site log Likelihoods for one ore more trees passed via -z and write them to a file in treepuzzle format that can be read by CONSEL.
The model parameters will be re-estimated for each tree!

Example: raxmlHPC -f G -s alg -m GTRGAMMA -z trees -n TEST

-f h compute log likelihood test (SH-test) between best tree passed via -t and a bunch of other trees passed via -z
The model parameters will be estimated on the first tree only!

Example: raxmlHPC -f h -t ref -z trees -s alg -m GTRGAMMA -n TEST

-f H compute log likelihood test (SH-test) between best tree passed via -t and a bunch of other trees passed via -z
The model parameters will be re-estimated for each tree !

Example: raxmlHPC -f H -t ref -z trees -s alg -m GTRGAMMA -n TEST

-f i calculate IC and TC scores (Salichos and Rokas 2013 htt)=@@/ / /.nature."om@nature@Aournal@v?QX@nX??Q@!ull@nature&2&B$.html) on a reference tree provided with -t based on multiple trees (e.g., from a bootstrap) in a file specified by -z

The method is des"ribed in more detail in the !ollo/ing)a)er /rote /ith Sali"hos and Rokas=

htt)=@@mbe.ox!ordAournals.org@"ontent@early@2$&?@$2@$X@molbev.msu$'&.abstra"tkeyty)eDre!1iAkeyD 'P2u(9x$H7R23/

The method is then extended here htt)=@@dx.doi.org@&$.&&$&@22PB !or)artial gene trees.

Warning: The de!ault T0@ 0 "al"ulations are not done exa"tly as des"ribed in our MRE)a)er. To have RAxML do exa"tly /hat is des"ribed in the)a)er%)leased edit Ile bipartitionList.c by "ommenting out or removing the line
#define BIP_FILTER

Example: raxmlHPC -f i -m GTRCAT -t referenceTree -z bootstrapTrees -n TEST

-f I a simple tree rooting algorithm for unrooted trees. It roots the tree by rooting it at the branch that best balances the subtree lengths (sum over branches in the subtrees) of the left and right subtree. A branch with an optimal balance does not always exist! You need to specify the tree you want to root via -t.

Example: raxmlHPC -f I -m GTRCAT -t unrootedTree -n TEST

-f j generate a bunch of bootstrapped alignment files from an original alignment file. You need to specify a seed with -b and the number of replicates with -#

Example: raxmlHPC -f j -b 12345 -# 100 -s alg -m GTRCAT -n TEST

Warning: 9ote that% i! you are generating those RS re)li"ate alignments !or a dataset !or /hi"h you subse4uently intend to "ondu"t a)artitioned analysis you /ill also need to s)e"i!y the)artition Ile here via -q be"ause RAxML re+sam)les sites on a)er+)artition basisG

-f J Compute SH-like support values on a given tree passed via -t.

This o)tion /ill "om)ute sh+like su))ort values as des"ribed here htt)=@@/ / /.n"bi.nlm.nih.gov@)ubmed@2$P2P'B8 on a given tree. The in)ut tree is ty)i"ally the best+kno/n ML tree !ound by a RAxML analysis. Zee) in mind that !or a))lying the SH+ like test the tree needs to be 99 69earest 9eighbor nter"hange8 o)timal. Thus% RAxML /ill initially try to a))ly 99 moves to !urther im)rove the tree and then "om)ute the SH test !or ea"h inner bran"h o! the tree.

Example: raxmlHPC -f J -p 12345 -m GTRGAMMA -s alg -t tree -n TEST

-f k Fix long branch lengths in partitioned data sets with missing data using the branch length stealing algorithm.
This option only works in conjunction with "-t", "-M", and "-q".
It will print out a tree with shorter branch lengths, but having the same likelihood score.

This o)tion tries to Ix the issue o! very long bran"h lengths in)artitioned datasets /ith missing data. 2or ea"h)artition and ea"h bran"h o! a given tree it "he"ks i! there is data available !or that)artition on both sides o! the tree as deIned by the bran"h. ! this is not the "ase this means that there is only missing data on one side. n su"h a "ase /e usually have a very long bran"h length. The algorithm Ixes this issue by stealing bran"h lengths !rom those)artitions that have data on both sides o! the bran"h under "onsideration. ! there are several other)artitions that have data it "om)utes a /eighted average !or the stolen bran"h length based on the site "ounts in ea"h)artition !rom /hi"h it stole a bran"h length. The ni"e)ro)erty o! this algorithm is that by "hanging the bran"h lengths in this /ay% the likelihood o! the tree is not "hanged.

Example: raxmlHPC -f k -m GTRGAMMA -s alg -q part -M -t tree -n TEST

A 9e/i"k tree Ile "ontaining a shorter tree /ith stolen bran"h lengths /ill be /ritten to a Ile "alled= RaxML_stolenBranchLengths.TEST

-f m compare bipartitions between two bunches of trees passed via "-t" and "-z" respectively. This will return the Pearson correlation between all bipartitions found in the two tree files. A file called RAxML_bipartitionFrequencies.outpuFileName will be printed that contains the pair-wise bipartition frequencies of the two sets

Example: raxmlHPC -f m -t trees1 -z trees2 -s alg -m GTRCAT -n TEST

-f n compute the log likelihood score of all trees contained in a tree file provided by -z
The model parameters will be estimated on the first tree only!

Example: raxmlHPC -f n -z trees -s alg -m GTRGAMMA -n TEST

-f N compute the log likelihood score of all trees contained in a tree file provided by -z
The model parameters will be re-estimated for each tree!

Example: raxmlHPC -f N -z trees -s alg -m GTRGAMMA -n TEST

-f o old and slower rapid hill-climbing without the heuristic cutoff described in htt)=@@link.s)ringer."om@arti"le@&$.&$$X@s&&2'P+$$X+$$'X+?F)age+&

! you use this o)tion you /ill ty)i"ally get slightly better likelihood s"ores /hile the run times are ex)e"ted to in"rease by !a"tor 2 to B.

Example: `raxmlHPC -f o -p 12345 -m GTRCAT -s alg -n TEST`

`-f p` perform pure stepwise MP addition of new sequences to an incomplete starting tree and exit

Example: `raxmlHPC -f p -t ref -p 12345 -s alg -m GTRCAT -n TEST`

`-f q` fast quartet calculator

This o)tion /ill "al"ulate the likelihood s"ores o! all 4uartets !or a given in)ut alignment. nitially RAxML /ill "onstru"t a randomi7ed ste) /ise addition order)arsimony starting tree or you "an also)ass a given tree 6e.g.% the best+kno/n ML tree8 to RAxML via `-t`. This tree is used to estimate model)arameters that /ill then remain Ixed during the 4uartet "al"ulations. 2or ea"h 4uartet that is evaluated% only the bran"h lengths /ill be o)timi7ed.

The algorithm itsel! has three Navors= 6i8 it "an randomly evaluate a s)e"iI" number o! 4uartets% 6ii8 it "an evaluate all 4uartets% or 6iii8 it "an evaluate 4uartets !rom a)re+s)e"iIed 4uartet "onstraint tree Ile that must "ontain exa"tly ? mono)hyleti" and multi+!ur"ating grou)s.

2or details on the in)ut !ormat !or the 4uartet "onstraint Ile% see the `-Y` o)tionG

To not evaluate all)ossible 4uartets 6kee) in mind that this number gro/s 4ui"kly as a !un"tion o! the taxa in the alignment@tree8 you have to use either `-#` or `-N` to s)e"i!y ho/ many 4uartets you /ant to randomly evaluate.

Example 1: `raxmlHPC -m GTRGAMMA -t tree -s alg -f q -p 12345 -n TEST`

The above exam)le /ill evaluate all)ossible 4uartets.

Example 2: `raxmlHPC -m GTRGAMMA -t tree -s alg -f q -p 12345 -N 100 -n TEST`

The above exam)le /ill evaluate &$$ randomly "hosen 4uartets.

The out)ut is /ritten to a Ile "alled= `RaxML_quartets.TEST`

t looks like this=

```
Taxon names and indexes:

Cow 1
Carp 2
Chicken 3
Human 4
Loach 5
Mouse 6
Rat 7
Seal 8
Whale 9
Frog 10
```

```
1 2 | 3 4: -2640.417355
1 3 | 2 4: -2640.810748
1 4 | 2 3: -2638.359608
1 2 | 3 5: -2552.181311
1 3 | 2 5: -2543.674965
...
```

2irst% ea"h taxon name is assigned a number% then the 4uartets are re)resented as bi)artitions% !or instan"e% `1 2 | 3 4: -2640.417355` is the 4uartet `((Cow,Carp), (Chicken, Human))` and has a likelihood o! `-2640.417355` .

The M; version !or this s)e"iI" o)tion that "an be "om)iled /ith the "orres)onding MakeIles /ill distribute 4uartet evaluations a"ross)ro"essors /ith M; .

`-f r` compute pairwise Robinson-Foulds (RF) distances between all pairs of trees in a tree file passed via -z. If the trees have node labels represented as integer support values the program will also compute two flavors of the weighted Robinson-Foulds (WRF) distance

Example: `raxmlHPC -m GTRCAT -z trees -f r -n TEST`

The t/o Navors o! the *eighted R2 distan"e that are "om)ute are=

&. Aust add the su))ort o! those bi)artitions "ontained in one tree% but not the other
2. add the su))ort o! those bi)artitions "ontained in one tree% but not the other and also add the diMeren"e in su))ort !or ea"h shared bi)artition

The out)ut looks as !ollo/s 6not in"luding /eighted R2 distan"es8=

```
0 1: 8 0.125000
```

/here `0` and `1` denote that this is the distan"e bet/een tree `0` and `1` in the tree Ile 6the Irst and se"ond tree in there8% /here `8` is the)lain R2 distan"e and `0.12500` is the normali7ed R2 distan"e "orres)onding to &2.PS.

The normali7ed@relative R2 distan"e is "al"ulated by dividing the)lain R2 by *2(n-3)*% /here *n* is the number o! taxa.
The value o! the normali7ed R2 distan"e ranges bet/een $.$ and &.$ and is inter)reted as the)er"entage o! s)lits 6bi)artitions8 that are uni4ue to one o! the t/o trees. n the above exam)le% &2.PS o! the s)lits are uni4ue to either tree $ or tree &.

`-f R` compute all pairwise Robinson-Foulds (RF) distances between a large reference tree passed via -t and many smaller trees (that must have a subset of the taxa of the large tree) passed via -z.
This option is intended for checking the plausibility of very large phylogenies that can not be inspected visually any more.

Example: `raxmlHPC -f R -m GTRCAT -t hugeTree -z manySmallTrees`

`-f s` split up a multi-gene partitioned alignment into the respective subalignments

Example: `raxmlHPC -f s -q part -s alg -m GTRCAT -n TEST`

Warning: 9ote that the sub+alignments /ill be named by the names you have given to

the individual)artitions. ! your)artition Ile "ontains t/o)artitions "alled part1 and part2% this "ommand /ill generate t/o alignment Iles "alled part1.phy and part2.phy

-f S compute site-specific placement bias using a leave one out test inspired by the evolutionary placement algorithm

The leave+one+out a))roa"h !or assessing site+s)e"il" "ongruen"e@in"ongruen"e /ith the underlying tree is based on the evolutionary)la"ement algorithm% see htt)=@@sysbio.ox!ordAournals.org@"ontent@'$@B@2Q&.short

,sing a given re!eren"e tree 6ty)i"ally the best+kno/n ML tree8 a leave+one+out test is "ondu"ted by)runing a single taxon *t* at a time% "arrying out site+s)e"il" "om)utations !or *t*% and subse4uently re+inserting it again into its original)osition.
2or ea"h taxon *t* /e use a sliding /indo/ o! si7e *w* 6in the de!ault "ase *w:=100*8 o! "onse"utive sites in the alignment and "om)ute the maximum likelihood)la"ement in the tree /ith res)e"t to those *w* sites only.

This is done re)eatedly !or ea"h alignment site by moving the /indo/ over the alignment on a site by site basis.

3n"e /e have obtained the best)la"ements !or ea"h sliding /indo/ starting)osition !or a taxon *t*% /e "om)ute a s"ore *s[i][t]* !or ea"h site.

The s"ore *s[i][t]* is the mean distan"e 6in terms o! number o! nodes in the tree8 bet/een the res)e"tive best)la"ements !or all sliding /indo/s that "om)rise site *i* and the original)la"ement o! taxon *t*.

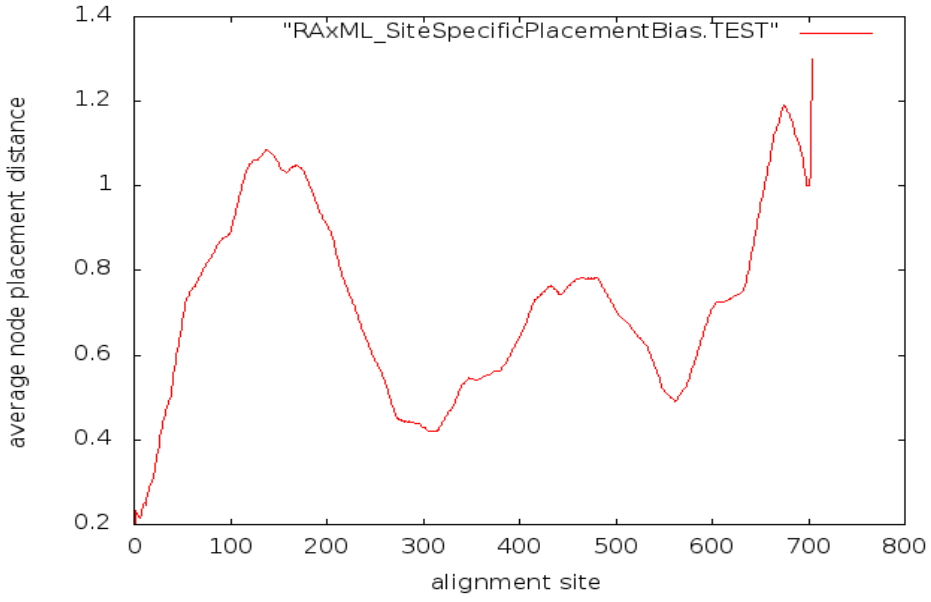
2inally% /e "al"ulate a global *S[i]* s"ore !or ea"h site by "om)uting the mean o! the *s[i][t]* "om)uted !or ea"h taxon *t*.

Thus% RAxML /ill return an list o! *n S[i]* average node distan"es% /here *n* is the number o! sites in the alignment.

Example: `raxmlHPC -f S -s alg -t tree -m GTRGAMMA -n TEST`

The main idea behind this algorithm is to)rovide a means !or assessing the)hylogeneti" variability o! diMerent areas@)arts o! a gene in order to de"ide /hi"h)art o! the gene 6e.g.% a &'S gene8 to am)li!y and se4uen"e using 9(S.

The o)tion above /ill generate an out)ut Ile "alled= `RaxML_SiteSpecificPlacementBias.TEST` /hi"h you "an then)lot using a tool like gnu)lot% to obtain the !ollo/ing)lot=



The)lot indi"ates that the alignment site bet/een)osition 2$$ and '$$ /ill yield the most "ongruent signal /ith the tree "al"ulated on the entire alignment. 3verall% the mean node)la"ement distan"es are relatively lo/% meaning that this alignment has a relatively stable)hylogeneti" signal over its entire length.

-f t do randomized tree searches on one fixed starting tree

This "ommand /ill exe"ute a s)e"iIed number 6by -N or -#8 o! randomi7ed tree sear"hes. The diMeren"e to the standard sear"h algorithm is% that% given a starting tree it /ill a))ly S;R 6subtree)runing re+gra!ting8 moves in a randomi7ed order and thus may Iend better trees.

Example: raxmlHPC -f t -t -m GTRCAT -s alg -p 12345 -n TEST

-f T do final thorough optimization of ML tree from rapid bootstrap search in stand-alone mode

This o)tion allo/s to do a more thorough tree sear"h that uses less la7y% i.e.% more exhaustive S;R moves% in stand+alone mode. This algorithm is ty)i"ally exe"uted in the very end o! a sear"h done via -f a.

Example: raxmlHPC -p 12345 -m GTRGAMMA -s alg -f T -t tree -n TEST

-f u execute morphological weight calibration using maximum likelihood, this will return a weight vector. you need to provide a morphological alignment and a reference tree via -t

This o)tion /ill determine to /hi"h degree the sites o! a mor)hologi"al alignment are "ongruent /ith a given mole"ular re!eren"e tree. As out)ut it /ill generate a RAxML /eight Ile that reNe"ts the degree o! "ongruen"e and that "an be subse4uently read via the -a o)tion to "ondu"t a more in!ormed evolutionary)la"ement o! !ossils using the evolutionary)la"ement algorithm 6-f v o)tion8. 2or more details% see the "orres)onding)a)ers= Evolutionary)la"ement htt)=@@sysbio.ox!ordAournals.org@"ontent@'$@B@2Q&.short 2ossil "alibration htt)=@@ieeex)lore.ieee.org@x)l@login.As)Ct)D1arnumberDPP8'QBQ1urlDhtt)SBAS2S22ieeex)lore.ieee.orgS22x)lsS22absKall.As)SB2arnumberSB.PP8'QBQ

Example: raxmlHPC -f u -m BINGAMMA -t molecularTree -s morphologicalAlignment

-f v classify a bunch of environmental sequences into a reference tree using thorough read insertions you will need to start RAxML with a non-comprehensive reference tree and an alignment containing all sequences (reference + query)

2or details on the design and)er!orman"e o! this algorithm% see=
htt)=@@sysbio.ox!ordAournals.org@"ontent@'$@B@2Q&.short
RAxML /ill)rodu"e a "ou)le o! idiosyn"rati" out)ut Iles !or the)la"ements% but also an out)ut Ile a""ording to the "ommon Ile standard deIned by Eri"k Matsen !or his pplacer 6see htt)=@@matsen.!h"r".org@))la"er@8)rogram that is similar to the E;A 6Evolutionary ;la"ement Algorithm8 and mysel!.
The "ommon Ile !ormat is des"ribed in this)a)er here=
htt)=@@//.)losone.org@arti"le@in!o SBAdoiS22&$.&BX&S22Aournal.)one.$$B&$$Q

Example: `raxmlHPC -f v -s alg -t referenceTree -m GTRCAT -n TEST`

! you !re4uently /ant to insert diMerent se4uen"e into the same re!eren"e tree% there is a short"ut to make this more eL"ient% be"ause the model)arameters and bran"h lengths are estimated !rom s"rat"h !or the re!eren"e tree ea"h time /hen you invoke the)rogram. ! you /ant to avoid this you)ro"eed as !ollo/s.

(enerate a binary Ile "ontaining the model)arameters !or the re!eren"e tree

Example 1: `raxmlHPC -f e -m GTRGAMMA -s referenceAlignment -t referenceTree -n PARAMS`

This /ill generate a Ile "alled= `RAxML_binaryModelParameters.PARAMS` /hi"h "an then be read by the E;A to avoid re+o)timi7ing)arameters=

Example 2: `raxmlHPC -f v -R RAxML_binaryModelParameters.PARAMS -t RAxML_result.PARAMS -s alg -m GTRGAMMA -n TEST2`

Warning: 9ote that% /hen using binary model Iles= The models o! rate heterogeneity in exam)les & 1 2 6e.g.% `GTRGAMMA` or `GTRCAT`8 must be the sameG Se"ondly% /hen using `GTRCAT` you must disable)attern "om)ression via the `-H` Nag in both runsG Thirdly% the re!eren"e tree read in via `-t` in exam)le 2 must have bran"h lengths a""ording to the model s)e"iIed via `-m`. The bran"h lengths o! the re!eren"e tree are not re+o)timi7ed by the "all in exam)le 2. t is very easy to make a mistake and read in a tree /hose bran"h lengths /ere estimated under `GTRGAMMA` /hile you are using `GTRCAT`. All o! the above only a))lies /hen using binary model Iles through the `-R` o)tion.

`-f V` classify a bunch of environmental sequences into a reference tree using thorough read insertions you will need to start RAxML with a non-comprehensive reference tree and an alignment containing all sequences (reference + query)

This is an ex)erimental extension o! the E;A !or)la"ing short reads in multi+gene datasets. t uses some "om)utational short"uts to make the)la"ement !aster. t has been used !or the !ollo/ing)a)er= htt)=@@//.nature."om@nmeth@Aournal@v&$@n&2@!ull@nmeth.2'QB.html

Example: `raxmlHPC -f V -q part -s alg -t referenceTree -m GTRCAT -n TEST`

Warning: this is a test im)lementation !or more eL"ient handling o! multi+gene 1 /hole+ genome datasetsG

-f w compute ELW test on a bunch of trees passed via -z
The model parameters will be estimated on the first tree only!

0om)utes the Extended Likelihood *eights test by Strimmer and Rambaut% see htt)=@@/ / /.n"bi.nlm.nih.gov@)ubmed@&&XQ8?28% on a set o! trees.

Example: raxmlHPC -f w -m GTRGAMMA -s alg -z trees -n TEST

-f W compute ELW test on a bunch of trees passed via -z
The model parameters will be re-estimated for each tree

Example: raxmlHPC -f W -m GTRGAMMA -s alg -z trees -n TEST

-f x compute pair-wise ML distances, ML model parameters will be estimated on an MP starting tree or a user-defined tree passed via -t, only allowed for GAMMA-based models of rate heterogeneity

This o)tion /ill Aust "om)ute the)air+/ise distan"es 6bran"h lengths8 bet/een the se4uen"es o! an alignment using a maximum likelihood estimate. To obtain a model)arameter estimate 6(TR% al)ha sha)e)arameter et".8 it /ill initially either read in a user s)e"i l ed tree 6e.g.% the best+kno/n ML tree8 or generate a randomi7ed ste) /ise addition)arsimony starting tree i! no in)ut tree is)rovided. t then o)timi7es model)arameters on this tree and then "om)utes all)air+/ises distan"es.

Example: raxmlHPC -f x -p 12345 -s alg -m GTRGAMMA -n TEST

-f y classify a bunch of environmental sequences into a reference tree using parsimony you will need to start RAxML with a non-comprehensive reference tree and an alignment containing all sequences (reference + query)

This "ommand is analogous to the -f v "ommand /ith the only diMeren"e that it uses)arsimony !or)la"ing reads into the re!eren"e tree. Re"ause it uses)arsimony it is orders o! magnitude !aster than the likelihood+based)la"ement. <ou "an use it !or)la"ing millions o! reads into extremely large re!eren"e trees.

Example: raxmlHPC -f y -m GTRCAT -s alg -t referenceTree -n TEST

Warning: This algorithm is based on)arsimony. Thus% a read may be)la"ed into more than one% e4ually)arsimonious)osition@bran"h o! the re!eren"e tree. Re "are!ul to take all e4ually)arsimonious)la"ements into a""ount /hen analy7ing the results6

-F enable ML tree searches under CAT model for very large trees without switching to GAMMA in the end (saves memory).
This option can also be used with the GAMMA models in order to avoid the thorough optimization of the best-scoring ML tree in the end.

DEFAULT: OFF

*hen "ondu"ting analyses under the 0AT model o! rate heterogeneity in RAxML% the)rogram /ill in most "ases try to evaluate the (AMMA+based s"ore and do some additional o)timi7ation o! the tree under (AMMA in the very end o! the sear"h. ! you /ant to avoid this be"ause o! time and@or memory "onstraints 6(AMMA needs !our times more memory than 0AT8 you "an use this o)tion.

Example: raxmlHPC -F -m GTRCAT -p 12345 -s alg -N 10 -n TEST

-g specify the file name of a multifurcating constraint tree this tree does not need to be comprehensive, i.e. must not contain all taxa

This o)tion !re4uently "auses "on!usion among users !or t/o main reasons= 6i8 i! the "onstraint tree you)rovide is not "om)rehensive% that is% does not "ontain all taxa in the alignment% the taxa not in"luded in the "onstraint are !ree% that is% they are allo/s to !all into any)art o! the treeG 6ii8 users o!ten state that the RAxML tree does not "orres)ond to their "onstraint tree. This is mostly due to the tree visuali7ation tools that are used% de)ending at /hi"h to) level tri!ur"ation 6remember that maximum likelihood trees are al/ays unrootedG8 the tree is rooted it may look as i! it does not "om)ly /ith the "onstraint. ;lease double+"he"k t/i"e and thri"e be!ore)osting to the RAxML google grou)G
2inally also note that% any multi+!ur"ations in the in)ut tree /ill be resolved via a maximum likelihood sear"h.

Example: raxmlHPC -g constraintTree -m GTRGAMMA -p 12345 -s alg -n TEST

-G enable the ML-based evolutionary placement algorithm heuristics by specifying a threshold value (fraction of insertion branches to be evaluated using thorough insertions under ML).

This o)tion "an be used in "onAun"tion /ith the +! v and +! - o)tions to s)eed u) evolutionary)la"ements o! short reads. A setting o! $.& 6&$S o! bran"hes "onsidered !or thorough insertions8 yields a good a""ura"y@s)eed trade+oM. 2or more details you may a"tually "onsider to read the)a)er= htt)=@@sysbio.ox!ordAournals.org@"ontent@'$@B@2Q&.short

Example: raxmlHPC -f v -G 0.1 -s alg -t referenceTree -m GTRCAT -n TEST

-h Display this help message.

Example: raxmlHPC -h

-H Disable pattern compression.

DEFAULT: ON

This o)tion allo/s to disable alignment site "om)ression in the in)ut alignment. ! alignments "ontain identi"al sites they "an be "om)ressed to be"ome shorter and thus s)eed u) the likelihood "al"ulations. This o)tion has been develo)ed mostly !or internal use in my lab.

Example: raxmlHPC -H -s alg -p 12345 -m GTRCAT -n TEST

Warning: using this o)tion may "onsiderably slo/ do/n RAxML

-i Initial rearrangement setting for the subsequent application of topological changes phase

This o)tion allo/s to s)e"i!y the radius 6number o! nodes a/ay !rom the original)runing)osition8 u) to /hi"h)runed subtrees /ill be re+inserted in the "ourse o! S;R 6subtree

)runing re+gra!ting moves8 during the tree sear"h. Ry de!ault this setting /ill be determined automati"ally by RAxML. The exam)le belo/ s)e"i!es a radius o! &$% that is% subtrees /ill be inserted into all bran"hes that are bet/een & and &$ nodes a/ay !rom their)runing)osition.

Example: raxmlHPC -i 10 -s alg -m GTRCAT -p 12345 -n TEST

-I a posteriori bootstopping analysis. Use:

-I autoFC for the frequency-based criterion
-I autoMR for the majority-rule consensus tree criterion
-I autoMRE for the extended majority-rule consensus tree criterion
-I autoMRE_IGN for metrics similar to MRE, but include bipartitions under the threshold whether they are compatible or not. This emulates MRE but is faster to compute.

You also need to pass a tree file containing several bootstrap replicates via -z

This o)tion allo/s to "arry out the bootstra) "onvergen"e test% that is% the test that determines i! you have "om)uted su!"ient RS re)li"ates !or getting stable su))ort values% a)osteriori% that is a!ter you have already "om)uted% !or instan"e% &$$ bootstra) re)li"ates. This o)tion is)arti"ularly use!ul /hen you are doing large+s"ale tree sear"hes on su)er"om)uters. Here you /ould initially run P$ re)li"ates and "he"k i! they are su!"ient. Then you8d run another P$ and asses i! the &$$ trees you have no/ are su!"ient% et". My !avorite Navors in terms o! "onvergen"e "riteria are= autoMRE and i! the tree is very large in terms o! number o! taxa 6more than &$$$8 autoMRE_IGN to save time. Re!ore using this o)tion you should read the "orres)onding)a)er= htt)=@@link.s)ringer."om@"ha)ter@&$.&$$X@QX8+B+'?2+$2$$8+XK&BF)age+&

Example: raxmlHPC -m GTRCAT -z RAxML_bootstrap.BS -I autoMRE -n TEST

9ote that% i! you have several !les "ontaining bootstra) trees they "an easily be "on"atenated by using the Linux cat "ommand.

Example: cat RAxML_bootstrap.BS_1 RAxML_bootstrap.BS_2 > allBootstraps

-j Specifies that intermediate tree files shall be written to file during the standard ML and BS tree searches.

DEFAULT: OFF

This /ill sim)ly)rint out a "ou)le o! intermediate trees during the tree sear"h and not the !nal tree only. The intermediate trees are /ritten to !les "alled= RAxML_checkpoint.TEST.0, RAxML_checkpoint.TEST.1, et".

Example: raxmlHPC -j -m GTRGAMMA -s alg -p 12345 -n TEST

-J Compute majority rule consensus tree with "-J MR" or extended majority rule consensus tree with "-J MRE" or strict consensus tree with "-J STRICT". For a custom consensus threshold >= 50%, specify T_<NUM>, where 100 >= NUM >= 50.
Options "-J STRICT_DROP" and "-J MR_DROP" will execute an algorithm that identifies dropsets which contain rogue taxa as proposed by Pattengale et al. in the paper "Uncovering hidden phylogenetic consensus".
You will also need to provide a tree file containing several UNROOTED trees

via "-z"

This option actually triggers two different sets of algorithms, one set for building consensus trees and one set for finding rogue taxa. Also note that, these algorithms have been parallelized, that is, if you use the -Threads version they will run faster.

You should read and cite the following papers when using these options:

For consensus trees:
http://www.exelixis-lab.org/software.html

For identifying rogue taxa:
http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6375971&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6375971

The two examples below compute an extended majority rule consensus tree and a MRE majority rule consensus tree.

Example: raxmlHPC -J MRE -z trees -m GTRCAT -n TEST

Example: raxmlHPC -J T_75 -z trees -m GTRCAT -n TEST

The example below identifies rogues using the aforementioned drop-set method and using a majority rule threshold. For details, please read the paper.

Example: raxmlHPC -J MR_DROP -z trees -m GTRCAT -n TEST

-k Specifies that bootstrapped trees should be printed with branch lengths. The bootstraps will run a bit longer, because model parameters will be optimized at the end of each replicate under GAMMA or GAMMA+P-Invar respectively.

DEFAULT: OFF

Example: raxmlHPC -b 12345 -p 12345 -# 100 -k -s alg -m GTRGAMMA -n TEST

-K Specify one of the multi-state substitution models (max 32 states) implemented in RAxML.
Available models are: ORDERED, MK, GTR

DEFAULT: GTR model

Evidently, for this option to have an effect you need to have an alignment containing multi-state characters. If you have several partitions that consist of multi-state characters the model specified via -Z will be applied to all models. Thus, it is not possible to assign different models to distinct multi-state partitions.

Example: raxmlHPC -p 12345 -m MULTIGAMMA -s multiStateAlignment -n TEST

-L Compute consensus trees labeled by IC supports and the overall TC value as proposed in Salichos and Rokas 2013.
Compute a majority rule consensus tree with -L MR or an extended majority rule consensus tree with -L MRE.
For a custom consensus threshold >= 50%, specify -L T_<NUM>, where 100 >=

NUM >= 50.
You will of course also need to provide a tree file containing several UNROOTED trees via -z!

This o)tion allo/s to "om)ute the T0 and 0 metri"s on a "onsensus tree% as des"ribed by Sali"hos and Rokas in this)a)er here=
htt)=@@/ / /.nature."om@nature@Aournal@v?QX@nX??Q@!ull@nature&2&B$.html

The method is des"ribed in more detail in the !ollo/ing)a)er /rote /ith Sali"hos and Rokas=
htt)=@@mbe.ox!ordAournals.org@"ontent@early@2$&?@$2@$X@molbev.msu$'&.abstra"tkeyty)eDre!1iAkeyD 'P2u(9x$H7R23/

Warning: The de!ault T0@0 "al"ulations are not done exa"tly as des"ribed in our MRE)a)er. To have RAxML do exa"tly /hat is des"ribed in the)a)er%)leased edit Ile bipartitionList.c by "ommenting out or removing the line
#define BIP_FILTER

There is a dedi"ated se"tion !urther do/n in this manual /ith detailed instru"tions.

Example: raxmlHPC -m GTRCAT -L MRE -z trees -n TEST

-m Model of Binary (Morphological), Nucleotide, Multi-State, or Amino Acid Substitution:

This is)robably the most "om)lex and "on!using "ommand line o)tion be"ause it "an be used to s)e"i!y a lot o! things. A very im)ortant thing to be a/are o! is that% in "ase you use a)artitioned alignment the data ty)e in the)artitions and the s)e"iI" models o! substitution to be used are a"tually s)e"iIed in the)artition Ile. Thus% i! you use the -q o)tion the only in!ormation that /ill be extra"ted !rom this string here that is)assed via -m is /hi"h model o! rate heterogeneity is going to be used. 9ote that% the rate heterogeneity model is al/ays a))lied to all data)artitions% hen"e you "an not have one)artition be analy7ed under CAT and another one be analy7ed under GAMMA.

n general% the term CAT in the strings al/ays indi"ates that you /ant to use the CAT model o! rate heterogeneity. The string CATI indi"ates that a!ter a tree sear"h under CAT% you /ant to evaluate the Inal trees under a GAMMA)lus)ro)ortion o! invariable sites estimate instead o! the de!ault)ure GAMMA.

Ry a))ending X to the model strings you indi"ate that you /ant to use a maximum likelihood estimate !or the base !re4uen"ies.

Ry)re)ending the string ASC_ you indi"ate that you /ant to a))ly an as"ertainment bias "orre"tion to the likelihood "al"ulations. This is use!ul !or binary@mor)hologi"al datasets that only "ontain variable sites 6the identi"al mor)hologi"al !eatures are usually not in"luded in the alignments% hen"e you need to "orre"t !or this% see% e.g.% htt)=@@sysbio.ox!ordAournals.org@"ontent@P$@'@Q&B.short8.

2or .9A data this o)tion might be use!ul /hen you analy7e alignments o! S9;s that also donHt "ontain "onstant sites. 9ote that% !or mathemati"al and numeri"al reasons you "an not a))ly an as"ertainment bias "orre"tion to datasets or)artitions that "ontain "onstant sites. n this "ase% RAxML /ill exit /ith an error.

Also note ho/ ambiguous sites are deIned in RAxML. ! you have a site looking like this=

AAAA----MR

this site is still "onsidered to be an invariable site sin"e - are "onsidered 6as in most likelihood+based "odes8 as "om)letely undetermined "hara"ters and thus -% M as /ell as R "ould be A. n other /ords% RAxML /ill not allo/ you to "ondu"t an analysis /ith an as"ertainment bias "orre"tion i! it en"ounters su"h a site in your MSA.

*e also observed that some S9; datasets do not re4uire a model o! rate heterogeneity a""ording to the standard model tests. *hen this is the "ase you "an use% !or instan"e% -m ASC_GTRCAT in "ombination /ith the -V o)tion. This /ill make RAxML exe"ute an in!eren"e under a)lain (TR model /ithout any "orre"tion !or rate heterogeneity.

Also note that% /hile do oMer the)er+site rate model o! rate heterogeneity 6the 0AT model8 /ith an as"ertainment bias "orre"tion% re"ommend that you either use 0AT in "ombination /ith -V 6no rate heterogeneity8 or the "orres)onding (amma model o! rate heterogeneity /ith as"ertainment bias "orre"tion 6-m ASC_GTRGAMMA et"8.

2inally% note that% as o! version 8.&.X o! RAxML you /ill need to s)e"i!y the ty)e o! as"ertainment bias "orre"tion you /ant to use% /e kno/ oMer three distin"t "orre"tion ty)es. 2or details%)lease re!er to the ++asc-corr o)tion.

BINARY:

-m BINCAT[X]	Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency. Final tree might be evaluated automatically under BINGAMMA, depending on the tree search option. With the optional "X" appendix you can specify a ML estimate of base frequencies.
-m BINCATI[X]	Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency. Final tree might be evaluated automatically under BINGAMMAI, depending on the tree search option. With the optional "X" appendix you can specify a ML estimate of base frequencies.
-m ASC_BINCAT[X]	Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency. Final tree might be evaluated automatically under BINGAMMA, depending on the tree search option. With the optional "X" appendix you can specify a ML estimate of base frequencies. The ASC prefix will correct the likelihood for ascertainment bias. You will also need to specify the correction type via --asc-corr!
-m BINGAMMA[X]	GAMMA model of rate heterogeneity (alpha parameter will be estimated). With the optional "X" appendix you can specify a ML estimate of base frequencies.
-m ASC_BINGAMMA[X]	GAMMA model of rate heterogeneity (alpha parameter will be

estimated). The ASC prefix will correct the likelihood for ascertainment bias.
With the optional "X" appendix you can specify a ML estimate of base frequencies. You will also need to specify the correction type via --asc-corr!

-m BINGAMMAI[X] Same as BINGAMMA, but with estimate of proportion of invariable sites. With the optional "X" appendix you can specify a ML estimate of base frequencies.

NUCLEOTIDES:

-m GTRCAT[X] GTR + Optimization of substitution rates + Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency. Final tree might be evaluated under GTRGAMMA, depending on the tree search option.
With the optional "X" appendix you can specify a ML estimate of base frequencies.

-m GTRCATI[X] GTR + Optimization of substitution rates + Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency. Final tree might be evaluated under GTRGAMMAI, depending on the tree search option.
With the optional "X" appendix you can specify a ML estimate of base frequencies.

-m ASC_GTRCAT[X] GTR + Optimization of substitution rates + Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency. Final tree might be evaluated under GTRGAMMA, depending on the tree search option.
With the optional "X" appendix you can specify a ML estimate of base frequencies.
The ASC prefix will correct the likelihood for ascertainment bias. You will also need to specify the correction type via --asc-corr!

-m GTRGAMMA[X] GTR + Optimization of substitution rates + GAMMA model of rate heterogeneity (alpha parameter will be estimated).
With the optional "X" appendix you can specify a ML estimate of base frequencies.

-m ASC_GTRGAMMA[X] GTR + Optimization of substitution rates + GAMMA model of rate heterogeneity (alpha parameter will be estimated).
The ASC prefix will correct the likelihood for ascertainment bias.
With the optional "X" appendix you can specify a ML estimate of base frequencies. You will also need to specify the correction type via --asc-corr!

-m GTRGAMMAI[X] Same as GTRGAMMA, but with estimate of proportion of invariable sites. With the optional "X" appendix you can specify a ML estimate of base frequencies.

MULTI-STATE:

-m MULTICAT[X] Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency. Final tree might be evaluated automatically under MULTIGAMMA, depending on the tree search option.

With the optional "X" appendix you can specify a ML estimate of base frequencies.

-m MULTICATI[X] Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency. Final tree might be evaluated automatically under MULTIGAMMAI, depending on the tree search option.
With the optional "X" appendix you can specify a ML estimate of base frequencies.

-m ASC_MULTICAT[X] Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency. Final tree might be evaluated automatically under MULTIGAMMA, depending on the tree search option.
With the optional "X" appendix you can specify a ML estimate of base frequencies. The ASC prefix will correct the likelihood for ascertainment bias. You will also need to specify the correction type via --asc-corr!

-m MULTIGAMMA[X] GAMMA model of rate heterogeneity (alpha parameter will be estimated). With the optional "X" appendix you can specify a ML estimate of base frequencies.

-m ASC_MULTIGAMMA[X] GAMMA model of rate heterogeneity (alpha parameter will be estimated). The ASC prefix willl correct the likelihood for ascertainment bias. With the optional "X" appendix you can specify a ML estimate of base frequencies. You will also need to specify the correction type via --asc-corr!

-m MULTIGAMMAI[X] Same as MULTIGAMMA, but with estimate of proportion of invariable sites. With the optional "X" appendix you can specify a ML estimate of base frequencies.
You can use up to 32 distinct character states to encode multi-state regions, they must be used in the following order:
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V
i.e., if you have 6 distinct character states you would use 0, 1, 2, 3, 4, 5 to encode these.
The substitution model for the multi-state regions can be selected via the "-K" option

AMINO ACIDS:

-m PROTCATmatrixName[F|X] specified AA matrix + Optimization of substitution rates + Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate

categories for greater computational efficiency. Final tree might be evaluated automatically under PROTGAMMAmatrixName[F|X], depending on the tree search option.
With the optional "X" appendix you can specify a ML estimate of base frequencies.

-m PROTCATImatrixName[F|X] specified AA matrix + Optimization of substitution rates + Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency.
Final tree might be evaluated automatically under PROTGAMMAImatrixName[F|X], depending on the tree search option. With the optional "X" appendix you can specify a ML estimate of base frequencies.

-m ASC_PROTCATmatrixName[F|X] specified AA matrix + Optimization of substitution rates + Optimization of site-specific evolutionary rates which are categorized into numberOfCategories distinct rate categories for greater computational efficiency.
Final tree might be evaluated automatically under PROTGAMMAmatrixName[F|X], depending on the tree search option.
With the optional "X" appendix you can specify a ML estimate of base frequencies. The ASC prefix will correct the likelihood for ascertainment bias. You will also need to specify the correction type via --asc-corr!

-m PROTGAMMAmatrixName[F|X] specified AA matrix + Optimization of substitution rates + GAMMA model of rate heterogeneity (alpha parameter will be estimated).
With the optional "X" appendix you can specify a ML estimate of base frequencies.

-m ASC_PROTGAMMAmatrixName[F|X] specified AA matrix + Optimization of substitution rates + GAMMA model of rate heterogeneity (alpha parameter will be estimated). The ASC prefix will correct the likelihood for ascertainment bias.
With the optional "X" appendix you can specify a ML estimate of base frequencies. You will also need to specify the correction type via --asc-corr!

-m PROTGAMMAImatrixName[F|X] Same as PROTGAMMAmatrixName[F|X], but with estimate of proportion of invariable sites.
With the optional "X" appendix you can specify a ML estimate of base frequencies.

Available AA substitution models:

DAYHOFF, DCMUT, JTT, MTREV, WAG, RTREV, CPREV, VT, BLOSUM62, MTMAM, LG, MTART, MTZOA, PMB, HIVB, HIVW, JTTDCMUT, FLU, STMTREV, DUMMY, DUMMY2, AUTO, LG4M, LG4X, PROT_FILE, GTR_UNLINKED, GTR

With the optional "F" appendix you can specify if you want to use empirical base frequencies. AUTOF and AUTOX are not supported any more, if you specify AUTO it will test prot subst. models with and without empirical

base frequencies now!
Please note that for partitioned models you can in addition specify the per-gene AA model in the partition file. Also note that if you estimate AA GTR parameters on a partitioned dataset, they will be linked (estimated jointly) across all partitions to avoid over-parametrization

Warning: * hen using the **LG4X**)rotein substitution model%)lease kee) in mind that% this model a"tually has more !ree)arameters than a normal% Ixed)rotein substitution model. Thus% !or ea"h)artition that evolves under L(?X there are ' additional !ree)arameters 6B !or the /eights and B !or the rates8. 9ote that it is only B /eight and B rate)arameters be"ause the /eights need to sum to &.$ 6the ?th)arameter is thus determined by the other B8 and the sum o! the)rodu"t o! ea"h /eight /ith the res)e"tive rate also needs to be &.$ 6see)age 2Q2? in the res)e"tive)a)er=
htt)=@@mbe.ox!ord^ournals.org@"ontent@2Q@&$@2Q2&.!ull.)d!8.
Also note that% in RAxML% L(?X "an be used /ith em)iri"al base !re4uen"ies% dra/n !rom the data 6L(?X2% L(?M28 and /ith a ML estimate o! base !re4uen"ies 6L(?XX% L(?MX8.
,nlike in the)ublished L(?X model% the base !re4uen"ies are the same 6**shared/linked**8 !or all ? substitution matri"es o! the L(? models and there are &Q additional)arameters in these models.
* hen using this modiIed !orm o! L(?X)lease make sure to state it ex)li"itly /hen you)ublish results and kee) in mind that they have more)arameters /hen you do A 0 1 R 0 test et". Also% these o)tions have been im)lemented !or "onvenien"e in RAxML% but do not "orres)ond to the s)irit o! the above)a)er /here ea"h o! the ? *Q* matri"es has its o/n set o! base !re4uen"ies.
Olivier Gascuel asked me to explicitly state that he does not like the following models: LG4XF, LG4MF, LG4XX, LG4MX to be used and I think he has a point!

The exam)le belo/ /ill exe"ute a sim)le sear"h under the 0AT a))roximation o! rate heterogeneity on a .9A dataset and evaluate the Inal tree under (AMMA.

Example: `raxmlHPC -m GTRCAT -s alg -p 12345 -n TEST`

The exam)le belo/ /ill exe"ute a sim)le sear"h under the 0AT a))roximation o! rate heterogeneity on a .9A dataset% but evaluate the Inal tree under (AMMAU;+nvar.

Example: `raxmlHPC -m GTRCATI -s alg -p 12345 -n TEST`

The exam)le belo/ /ill exe"ute a sim)le sear"h under the 0AT a))roximation o! rate heterogeneity on a .9A dataset and do a maximum likelihood estimate o! the base !re4uen"ies% instead o! using em)iri"al !re4uen"ies.

Example: `raxmlHPC -m GTRCATX -s alg -p 12345 -n TEST`

The exam)le belo/ /ill exe"ute a sim)le sear"h under the (AMMA model o! rate heterogeneity on a .9A dataset.

Example: `raxmlHPC -m GTRGAMMA -s alg -p 12345 -n TEST`

The exam)le belo/ /ill exe"ute a sim)le sear"h under the (AMMA model o! rate heterogeneity and "orre"t !or as"ertainment bias on a .9A dataset.

Example: `raxmlHPC -m ASC_GTRGAMMA -s alg -p 12345 -n TEST`

The exam)le belo/ / ill exe"ute a sim)le sear"h under a)lain (TR model 6no rate heterogeneity8 and "orre"t !or as"ertainment bias on a . 9A dataset.

Example: `raxmlHPC –m ASC_GTRCAT –V –s alg –p 12345 –n TEST`

The exam)le belo/ / ill exe"ute a sim)le sear"h under the (AMMA model o! rate heterogeneity on a binary dataset.

Example: `raxmlHPC –m BINGAMMA –s alg –p 12345 –n TEST`

PROTGAMMAILGX –s alg –p 12345 –n TEST

The exam)le belo/ / ill exe"ute a sim)le sear"h under the (AMMA model o! rate heterogeneity on a)rotein dataset using the substitution matrix and the base !re4uen"ies that "ome / ith the *A(model.

Example: `raxmlHPC –m PROTGAMMAWAG –s alg –p 12345 –n TEST`

The exam)le belo/ / ill exe"ute a sim)le sear"h under the (AMMA model o! rate heterogeneity on a)rotein dataset using em)iri"al base !re4uen"ies and the L(substitution model.

Example: `raxmlHPC –m PROTGAMMALGF –s alg –p 12345 –n TEST`

The exam)le belo/ / ill exe"ute a sim)le sear"h under the (AMMAU;+ nvar model o! rate heterogeneity on a)rotein dataset using a maximum likelihood estimate o! the base !re4uen"ies and the L(substitution model.

Example: `raxmlHPC –m PROTGAMMAILGX –s alg –p 12345 –n TEST`

The exam)le belo/ / ill exe"ute a sim)le sear"h under the (AMMA model o! rate heterogeneity on a)rotein dataset using em)iri"al base !re4uen"ies and estimating a (TR model o! amino a"id substitution.

Warning: This may be very slo/ and you might over+)arametri7e the model sin"e / e need to do a maximum likelihood estimate o! &8Q rate)arameters in the (TR matrix.

Example: `raxmlHPC –m PROTGTRGAMMA –s alg –p 12345 –n TEST`

The exam)le belo/ / ill automati"ally determine / hi"h is the best 6the one / ith the highest likelihood s"ore on the)arsimony starting tree8)rotein substitution model !or your dataset using the base !re4uen"ies that "ome / ith the models. t / ill "hose among the !ollo / ing models= `DAYHOFF, DCMUT, JTT, MTREV, WAG, RTREV, CPREV, VT, BLOSUM62, MTMAM, LG, MTART, MTZOA, PMB, HIVB, HIVW, JTTDCMUT, FLU, DUMMY, DUMMY2.` These models / ill not be "onsidered= `LG4M, LG4X, PROT_FILE,GTR_UNLINKED, GTR!`

RAxML / ill no/ also automati"ally test models / ith and / ithout em)iri"al base !re4uen"ies. The "riterion !or making this "hoi"e "an no/ be sele"ted via the

files
A weighted average of the branch lengths is computed by using the respective partition lengths

DEFAULT: OFF

This will enable an independent repartition estimate of branch lengths. Note that this increases dramatically the number of free parameters in your model. An unrooted binary tree has *2n-3* branch lengths where n is the number of taxa. If you use this option instead of *2n-3* parameters you will get *p(2n-3)* parameters where *p* is the number of partitions. RAxML will also run notably slower if you use this option.
Apart from the normal output tree, RAxML will also generate output files for each partition individually that allow to recover the repartition branch lengths if needed.

Example: raxmlHPC -p 12345 -M -m GTRGAMMA -s alg -q part -n TEST

-n Specifies the name of the output file.

This option has to be always specified. The arbitrary name passed via -n will be appended to all RAxML output files such that you know which files have been generated by which invocation.
If you intend to do two runs that write files into the same directory with the same name specified by -n the program will exit with an error to prevent you overwriting output files from a previous run.

Example: raxmlHPC -p 12345 -m GTRGAMMA -s alg
-n MySuperDuperRAxMLOutputFileName

-o Specify the name of a single outgroup or a comma-separated list of outgroups, e.g., -o Rat or -o Rat,Mouse, in case that multiple outgroups are not monophyletic the first name in the list will be selected as outgroup, don't leave spaces between taxon names!

If there is more than one outgroup a check for monophyly of the outgroups in the respective output@Inal tree(s) will be performed. If the outgroups is not monophyletic the tree will be rooted at the Irst outgroup in the list and a respective
warning will be printed.
Keep in mind that outgroups are Aust a tree drawing option, the trees remain essentially unrooted trees.

Example: raxmlHPC -s alg -p 12345 -m GTRGAMMA -o Rat,Mouse -n TEST

A comment on using outgroups: How I would do it.

In general Hd avoid using outgroups in the initial ML and bootstrap analyses.
Hd proceed as follows and provide the rationale for this below

1. Build a tree (MLURS search) only on the ingroup
2. Then, use E;A (see= http)=@@sysbio.oxfordAournals.org@"ontent@'$@B@2Q&.short) to place the outgroups onto the tree a posteriori.

This has several advantages=

&. <ou can use@test diMerent outgroups
2. <ou donHt have to re+run the entire analysis if the outgroup somehow perturbed the

analysis o! the ingrou). There have been)a)ers on this issue.
B. <ou avoid the outgrou) aMe"ting the bran"h lengths o! the ingrou) be"ause o! the /ay the evolutionary)la"ement algorithm has been built.
?. <ou get)la"ement)robabilities 6likelihodo /eights8 !or ea"h outgrou)% e.g.% ho/ likely the outgrou) is to !all into one bran"h or another. This also tells you i! your outgrou) is good 6high)robability !or being)la"ed into one s)e"i!" bran"h8 or i! it is bad 6s"atters a"ross the tree8

-O Disable check for completely undetermined sequence in alignment. The program will not exit with an error message when "-O" is specified.

DEFAULT: check enabled

n general it does not make sense to analy7e alignments /here one or more se4uen"es "onsist o! "om)letely undetermined "hara"ters 6e.g.% -,N,?,X et".8 be"ause you don#t have any in!ormation about these se4uen"es and they /ill sim)ly randomly s"atter throughout the tree. 9ormally RAxML /ill exit /ith an error message /hen it dete"ts su"h se4uen"es. This o)tion here "an disable this behavior% but be /arned that you really need to kno/ /hat you are doing /hen using this o)tion.

Example: raxmlHPC -s alg -p 12345 -m GTRGAMMA -O -n TEST

-p Specify a random number seed for the parsimony inferences. This allows you to reproduce your results and will help me debug the program.

2or all o)tions@algorithms in RAxML that re4uire some sort o! randomi7ation% this o)tion must be s)e"i!ied. Make sure to)ass diMerent random number seeds to RAxML and not only &2B?P as have done in the exam)les. * hen not s)e"i!ying +) /hen it is re4uired by RAxML% the)rogram /ill exit /ith a res)e"tive error message.
n the exam)le belo/ 6a sim)le tree sear"h !or the ML tree8 the random number seed is re4uired !or randomi7ed ste) /ise addition order)arsimony starting tree that is "om)uted)rior to the a"tual ML o)timi7ation.

Example: raxmlHPC -s alg -p 2352890 -m GTRGAMMA -n TEST

-P Specify the file name of a user-defined AA (Protein) substitution model. This file must contain 420 entries, the first 400 being the AA substitution rates (this must be a symmetric matrix) and the last 20 are the empirical base frequencies

9ote that% the substitution model you s)e"i!y via -m PROTGAMMAWAG in the exam)le belo/ /ill be ignored here. RAxML /ill only extra"t the in!ormation that this is a)rotein data alignment and that you /ant to use the (AMMA model o! rate heterogeneity !rom the string and /ill ignore WAG.
! you /ant to use your o/n)rotein substitution models !or)artitioned datasets% the substitution model !iles that have the same !ormat as des"ribed here need to be s)e"i!ied in the)artition !ile.

Example: raxmlHPC -s alg -p 12345 -P myProteinModel -m PROTGAMMAWAG -n TEST

-q Specify the file name which contains the assignment of models to alignment partitions for multiple models of substitution. For the syntax of this file please consult the manual.

This option allows you to specify the regions of your alignment for which an individual model of nucleotide substitution should be estimated. They can also contain different types of data.
This will typically be useful to infer trees for long (in terms of base)pairs) multi-gene alignments. (e.g.) -m (TRGAMMA is used) individual alpha-shape parameters) (TR rates) and empirical base frequencies will be estimated and optimized for each partition.
Since RAxML can handle alignments consisting of different data types (binary data) DNA data) protein data etc.) you must specify the type of data for each partition in the partition file) before the partition name.
For DNA data this must means that you have to add DNA to each line in the partition file) for AA data this is done by specifying the respective AA substitution matrix (e.g.) WAG or LG) you want to use for a partition. For binary data you'd specify BIN and for multi-state data MULTI.
Now) if you want to analyze a specific partition using an ascertainment bias correction you need to prepend ASC_ to the data type) e.g.) ASC_WAG or ASC_DNA.
For a specific partition you want to use a maximum likelihood estimate for the base frequencies) you'd append X to the data type name) e.g.) DNAX or LGX.
If you want to use empirical base frequencies (instead of the default) pre-defined ones that ship with the models) you'd write) e.g.) WAGF or JTTF.
If you want to use a protein substitution model of your own for a specific partition) it must be in the same format as specified for the -P option. Instead of specifying the data type with WAG for instance) you will need to specify the protein substitution model file name in square brackets) e.g.) [myProtenSubstitutionModelFileName].

If you want to do a partitioned analysis of concatenated AA and DNA partitions you can either specify -m GTRGAMMA or) e.g.) -m PROTGAMMAWAG. The only thing that will be extracted from the string passed via -m is the model of rate heterogeneity you want to use.

If you have a pure DNA alignment with 1000 bases (from two genes gene1 (positions 1-500) and gene2 (positions 501-1000) the information in the partition file should look as follows:

```
DNA, gene1 = 1-500
DNA, gene2 = 501-1000
```

If you want an ML estimate of frequencies for the first partition it will look like this:

```
DNAX, gene1 = 1-500
DNA, gene2 = 501-1000
```

To analyze the first partition with ascertainment bias correction you need to write:

```
ASC_DNA, gene1 = 1-500
DNA, gene2 = 501-1000
```

If gene1 is scattered through the alignment) e.g. positions 1-200) and 800-1,000 you specify this with:

```
DNA, gene1 = 1-200, 800-1,000
DNA, gene2 = 201-799
```

You can also assign distinct models to the codon positions) i.e. if you want a distinct model to be estimated for each codon position in gene1 you can specify:

```
DNA, gene1codon1 = 1-500\3
DNA, gene1codon2 = 2-500\3
DNA, gene1codon3 = 3-500\3
DNA, gene2 = 501-1000
```

! you only need a distin"t model !or the Brd "odon)osition you "an / rite=

```
DNA, gene1codon1andcodon2 = 1-500\3, 2-500\3
DNA, gene1codon3 = 3-500\3
DNA, gene2 = 501-1000
```

As already mentioned% !or AA data you must s)e"i!y the transition matri"es !or ea"h)artition=

```
JTT, gene1 = 1-500
WAGF, gene2 = 501-800
WAG, gene3 = 801-1000
```

The AA substitution model must be the Irst entry in ea"h line and must be se)arated by a "omma !rom the gene@)artition name% Aust like the . 9A token above. <ou "an not assign diMerent models o! rate heterogeneity to diMerent)artitions% i.e.% it / ill be either 0AT % (AMMA % (AMMA et". !or all)artitions% as s)e"iIed / ith -m .

! you / ant to use a)rotein model o! your o / n% !or)artition one% youHd / rite=

```
[prot~myProtenSubstitutionModelFileName], gene1 = 1-500
WAGF, gene2 = 501-800
WAG, gene3 = 801-1000
```

To use a maximum likelihood estimate o! the base !re4uen"ies !or all)artitions youHd / rite=

```
JTTX, gene1 = 1-500
WAGX, gene2 = 501-800
WAGX, gene3 = 801-1000
```

2inally% i! you have a "on"atenated . 9A and AA alignment% / ith . 9A data at)ositions &W P$$ and AA data at P$&+&%$$$ / ith the *A(model the)artition Ile should look as !ollo / s=

```
DNA, gene1 = 1-500
WAG, gene2 = 501-1000
```

A)artition Ile !or binary data / ould look like this=

```
BIN, gene1 = 1-500
BIN, gene2 = 501-1000
```

and !or multi+state data

```
MULTI, gene1 = 1-500
MULTI, gene2 = 501-1000
```

Example: `raxmlHPC -s alg -m GTRGAMMA -q part -p 12345 -n TEST`

-r Specify the file name of a binary constraint tree.
This tree does not need to be comprehensive, i.e. must not contain all taxa

This o)tion allo/s you to)ass a binary@bi!ur"ating "onstraint@ba"kbone tree in 9E* 0Z !ormat to RAxML.
9ote% that using this o)tion only makes sense i! this tree "ontains less taxa than the in)ut alignment. The remaining taxa /ill initially be added by using)arsimony "riterion. 3n"e a "om)rehensive tree /ith all taxa has been obtained it /ill be o)timi7ed under ML res)e"ting the restri"tions o! the binary "onstraint tree.

Thus% o)tion /ill not "hange the stru"ture o! the binary ba"kbone tree you)rovide as in)ut at all. * hat it /ill do is to Aust add the taxa not "ontained in the binary ba"kbone tree to the best)ositions based on their likelihood. The result is a !ully resolved binary tree.

Example: raxmlHPC -s alg -m GTRGAMMA -r constr -p 12345 -n TEST

-R Specify the file name of a binary model parameter file that has previously been generated with RAxML using the -f e tree evaluation option. The file name should be: RAxML_binaryModelParameters.runID

This o)tion "an be used to save time !or some other RAxML o)tions% by not re+estimating the model)arameters o! a Ixed tree again% but only doing this on"e. The binary model)arameter in)ut Ile is al/ays automati"ally generated by the -f e tree evaluation !un"tion.

t is use!ul /ith the -f v o)tion /hen you /ant to)la"e diMerent reads into the same re!eren"e tree and /ith the 4uartet evaluation !un"tion -f q% in)arti"ular the)arallel version o! itG

Example: raxmlHPC -f v -R RAxML_binaryModelParameters.PARAMS
-t RAxML_result.PARAMS -s alg -m GTRCAT -n TEST2

-s Specify the name of the alignment data file in PHYLIP or FASTA format

S)e"i!y the name o! the alignment data Ile /hi"h "an be in relaxed ;H<L ; !ormat. Relaxed means that you don:t have to /orry i! the se4uen"e Ile is interleaved or se4uential and that the taxon names are too long. * hat you do need to /orry about though is that there al/ays needs to be a s)a"e bet/een the taxon name and the se4uen"e.
Se4uen"e names "an be o! variable length bet/een & u) to 2P' "hara"ters. ! you need longer taxon names you "an ada)t the "onstant #define nmlngth 256 in Ile axml.h a))ro)riately. Moreover% RAxML is not sensitive /ith res)e"t to the !ormatting 6tabs% insets% et"8 o! interleaved ;H<L ; Iles.

RAxML "an no/ also)arse 2ASTA !ormat. ! RAxML noti"es that it "anHt)arse a)hyli) !ormat it /ill try to)arse the alignment Ile as 2ASTA !ormat. So !ar it has been able to)arse all 2ASTA Iles.

A)lethora o! small exam)le in)ut Iles !or diMerent data ty)es "an be !ound in the ste)+by+ste) on+line tutorial= htt)=@@s"o.h+its.org@exelixis@/eb@so!t/are@raxml@handsKon.html

Example: raxmlHPC -s alignment.fasta -m GTRGAMMA -p 12345 -n TEST

-S Specify the name of a secondary structure file. The file can contain "." for alignment columns that do not form part of a stem and characters "()<>[]{}" to define stem regions and pseudoknots

S)e"i!ying se"ondary stru"ture models !or an R9A alignment /orks slightly diMerently than)assing other data+ty)es to RAxML be"ause /e read in a)lain R9A alignment and then need to tell RAxML by an additional text Ile that is)assed via -s /hi"h R9A alignment sites need to be grou)ed@evolve together.
*e do this in a standard bra"ket notation /ritten into a)lain text Ile% e.g.% our .9A test alignment has '$ sites% thus our se"ondary stru"ture Ile needs to "ontain a string o! '$ "hara"ters like this one=

```
..................((.......)).............(.........).........
```

The H.H symbol indi"ates that this is Aust a normal R9A site /hile the bra"kets indi"ate stems. Evidently% the number o! o)ening and "losing bra"kets mus mat"h. n addition% it is also)ossible to s)e"i!y)seudo knots /ith additional symbols= <>[]{} !or instan"e=

```
..................((.......)).......{....(....}....).........
```

n terms o! models there are '+state% X+state and &'+state models !or a""ommodating se"ondary stru"ture that are s)e"iIed via -A.
Available models are= S6A, S6B, S6C, S6D, S6E, S7A, S7B, S7C, S7D, S7E, S7F, S16, S16A, S16B. The de!ault is the (TR &'+state model 6-A S168. n RAxML the same nomen"lature as in ;HASE is used% so)lease "onsult the)hase manual at http://intranet.cs.man.ac.uk/ai//Software/phase/phase-2.0-manual.pdf !or a ni"e and detailed des"ri)tion o! these models.

Example: `raxmlHPC -m GTRGAMMA -p 12345 -S secondaryStructure.txt -s rna.phy -n TEST`

A "ommon 4uestion is /hether se"ondary stru"ture models "an also be)artitioned. This is)resently not)ossible. Ho/ever% you "an)artition the underlying R9A data% e.g.% use t/o)artitions !or our .9A dataset as be!ore. *hat RAxML /ill do internally though is to generate a third)artition !or se"ondary stru"ture that does not take into a""ount that distin"t se"ondary stru"ture site)airs may be)art o! diMerent)artitions o! the alignment.

`-t` Specify a user starting tree file name in Newick format

S)e"iIes a user starting tree Ile name /hi"h must be in 9e/i"k !ormat. The bran"h lengths o! that tree /ill generally be ignored !or most o)tions and re+estimated instead by RAxML. 9ote% that you "an also s)e"i!y a non+"om)rehensive 6not "ontaining all taxa in the alignment8 starting tree. This might be use!ul i! ne/ly aligned@se4uen"ed taxa have been added to your alignment and you /ant to extend the "urrent tree.
nitially% taxa /ill be added to the tree using)arsimony. The "om)rehensive tree /ill then be o)timi7ed under ML.

Example: `raxmlHPC -s alg -m GTRGAMMA -t tree -p 12345 -n TEST`

`-T` PTHREADS VERSION ONLY! Specify the number of threads you want to run. Make sure to set "-T" to at most the number of CPUs you have on your machine, otherwise, there will be a huge performance decrease!

T is set to $ by de!ault% the ;Threads version /ill)rodu"e an error i! you do not set -T to at least 2.

Example: `raxmlHPC-PTHREADS -T 4 -s alg -m GTRGAMMA -p 12345 -n TEST`

-u use the median for the discrete approximation of the GAMMA model of rate heterogeneity

DEFAULT: OFF

! you s)e"i!y this o)tion% RAxML /ill use the median instead o! the mean !or the dis"rete (amma model o! rate heterogeneity. Ty)i"ally% using the median /ill yield slightly better likelihood s"ores.

Example: raxmlHPC -p 12345 -u -s alg -m GTRGAMMA -n TEST

-U Try to save memory by using SEV-based implementation for gap columns on large gappy alignments. The technique is described here: http://www.biomedcentral.com/1471-2105/12/470
This will only work for DNA and/or PROTEIN data and only with the SSE3 or AVX-vextorized version of the code.

This o)tion "an hel) you to save memory and)otentially also time on large ga))y multi+gene alignments /ith missing data% in)arti"ular in "ases /here you have a ty)i"al ga))y)artitioned alignment 6data !or "ertain taxa missing !or "ertain genes@)artitions8. The amount o! saved memory is roughly)ro)ortional to the)ro)ortion o! missing data.

Example: raxmlHPC -p 12345 -U -s alg -m GTRGAMMA -n TEST

-v Display version information

. is)lays the RAxML version you are using.

Example: raxmlHPC -v

-V Disable rate heterogeneity among sites model and use one without rate heterogeneity instead. Only works if you specify the CAT model of rate heterogeneity.

DEFAULT: use rate heterogeneity

This o)tion "an be used i! your dataset better Its a model /ithout rate heterogeneity. This is rare% but su"h datasets do exist.

Example: raxmlHPC -m GTRCAT -s alg -p 12345 -V -n TEST

-w FULL (!) path to the directory into which RAxML shall write its output files

DEFAULT: current directory

9ame o! the /orking dire"tory /here RAxML shall /rite its out)ut Iles to. 9ote that you need to s)e"i!y the !ull)ath and not the relative)ath6

Example: raxmlHPC -m GTRCAT -p 12345 -s alg
-w /home/stamatak/Desktop/myAnalysys -n TEST

-W Sliding window size for leave-one-out site-specific placement bias algorithm only effective when used in combination with -f S

DEFAULT: 100 sites

.elnes the sliding /indo/ si7e !or the +! S o)tion% &$$ usually yields relatively smooth)lots.

Example: raxmlHPC -f S -s alg -t tree -m GTRGAMMA -N 100 -W 200 -n TEST

-x Specify an integer number (random seed) and turn on rapid bootstrapping
CAUTION: unlike in previous versions of RAxML will conduct rapid BS replicates under the model of rate heterogeneity you specified via -m and not by default under CAT

This /ill invoke the ra)id bootstra))ing algorithm des"ribed in
htt)=@@sysbio.ox!ord^ournals.org@"ontent@PX@P@XP8.short

Example: raxmlHPC -x 12345 -p 12345 -# 100 -m GTRCAT -s alg -n TEST

<ou "an also "ombine this /ith the bootsto))ing o)tion 6bootstra) "onvergen"e "riterion8% e.g.=

Example: raxmlHPC -x 12345 -p 12345 -# AUTO_MRE -m GTRCAT -s alg -n TEST

or also have RAxML sear"h !or the best+s"oring ML tree a!ter the bootstra) sear"hes using the -f a o)tion=

Example: raxmlHPC -x 12345 -p 12345 -# AUTO_MRE -m GTRCAT -s alg -f a -n TEST

-X Same as the "-y" option below, however the parsimony search is more superficial.
RAxML will only do a randomized stepwise addition order parsimony tree reconstruction without performing any additional SPRs.
This may be helpful for very broad whole-genome datasets, since this can generate topologically more different starting trees.

DEFAULT: OFF

Example: raxmlHPC -X -m GTRCAT -s alg -p 12345 -n TEST

-y If you want to only compute a parsimony starting tree with RAxML specify -y, the program will exit after computation of the starting tree

DEFAULT: OFF

! you /ant to only "om)ute a randomi7ed)arsimony starting tree /ith RAxML and not exe"ute an ML analysis o! the tree s)e"i!y -y. The)rogram /ill exit a!ter "om)utation o! the starting tree.

Example: raxmlHPC -y -m GTRCAT -s alg -p 12345 -n TEST

-Y Pass a quartet grouping file name defining four groups from which to draw quartets
The file input format must contain 4 groups in the following form:
(Chicken, Human, Loach), (Cow, Carp), (Mouse, Rat, Seal), (Whale, Frog);

Only works in combination with -f q !

The exam)le belo/ % /ill randomly dra/ &$$ 4uartets !rom the)redi1ned 4uartet grou)ing above and evaluate their likelihood.

Example: `raxmlHPC -f q -m GTRGAMMA -t tree -Y quartetFile -n 100 -n TEST`

`-z` Specify the file name of a file containing multiple trees e.g. from a bootstrap that shall be used to draw bipartition values onto a tree provided with -t.
It can also be used, for instance to compute per site log likelihoods in combination with -f g and to read a bunch of trees for a couple of other options (-f h, -f m, -f n).

This o)tion is re4uired in "ombination /ith a lot o! other RAxML o)tions% in essen"e every time you need to read in a bun"h o! trees. Relo/ is an exam)le /here you use the o)tion to dra/ RS su))ort values on a best+kno/n ML tree.

Example: `raxmlHPC -f b -m GTRCAT -t mlTree -z bootstrapTrees -n TEST`

`-#|-N` Specify the number of alternative runs on distinct starting trees
In combination with the "-b" option, this will invoke a multiple bootstrap analysis
Note that "-N" has been added as an alternative since -# sometimes caused problems with certain MPI job submission systems, since -# is often used to start comments.
If you want to use the bootstopping criteria specify -# autoMR or -# autoMRE or -# autoMRE_IGN for the majority-rule tree based criteria (see -I option) or -# autoFC for the frequency-based criterion.
Bootstopping will only work in combination with -x or -b

DEFAULT: 1 single analysis

S)e"i1es the number o! alternative runs on distin"t starting trees% e.g.% i! -# 10 or -N 10 is s)e"i1ed% RAxML /ill "om)ute &$ distin"t ML trees starting !rom &$ distin"t randomi7ed maximum)arsimony starting trees.
n "ombination /ith the -b o)tion% this /ill invoke a multi)le bootstra) analysis. n "ombination /ith -x this /ill invoke a ra)id RS analysis and "ombined /ith -f a -x a ra)id RS sear"h and therea!ter a thorough ML sear"h on the original alignment.
*e introdu"ed -N as an alternative to -# sin"e the s)e"ial "hara"ter # seems to sometimes "ause)roblems /ith "ertain bat"h Aob submission systems.
n "ombination /ith -f j this /ill generate number3!Runs bootstra))ed alignment 1les.

9ote that% several other)rogram o)tions su"h as the 4uartet evaluation algorithm 6-f q o)tion8 or the site+s)e"i1")la"ement bias algorithm 6-f S o)tion8 re4uire this o)tion to be s)e"i1ed as /ell. ! you !orgot to s)e"i!y this o)tion% RAxML /ill tell you that you need to do so.

The exam)le belo/ /ill do 2$ inde)endent ML tree sear"hes on 2$ randomi7ed ste)/ise addition order)arsimony tree.

Example: `raxmlHPC -p 12345 -s alg -n TEST -m GTRGAMMA -# 20`

`--mesquite` Print output files that can be parsed by Mesquite.

DEFAULT: Off

Example: raxmlHPC -p 12345 -s alg --mesquite -n TEST -m GTRGAMMA

--silent Disables printout of warnings related to identical sequences and entirely undetermined sites in the alignment . The program might run faster when this is enabled.

DEFAULT: Off

Example: raxmlHPC -p 12345 --silent -s alg -n TEST -m GTRGAMMA

--no-seq-check Disables checking the input MSA for identical sequences and entirely undetermined sites. Enabling this option may save time, in particular for large phylogenomic alignments.
Before using this, make sure to check the alignment using the -f c option!

DEFAULT: Off

Example: raxmlHPC -p 12345 -s alg --no-seq-check -n TEST -m GTRGAMMA

--no-bfgs Disables automatic usage of BFGS method to optimize GTR rates on unpartitioned DNA datasets . Using BFGS can improve speeds for model optimization by up to 30%. It's enabled by default when analyzing single-partition DNA datasets.

DEFAULT: BFGS on

Example: raxmlHPC -p 12345 -s alg --no-bfgs -n TEST -m GTRGAMMA

--asc-corr Allows to specify the type of ascertainment bias correction you wish to use. There are three types available:

--asc-corr=lewis: the standard correction by Paul Lewis
--asc-corr=felsenstein: a correction introduced by Joe Felsenstein that allows to explicitely specify the number of invariable sites (if known) one wants to correct for.
--asc-corr=stamatakis: a correction introduced by myself that allows to explicitely specify the number of invariable sites for each character (if known) one wants to correct for.

Let's start with the Lewis correction that is easy to handle:

Example: raxmlHPC -p 12345 -s alg -n TEST -m ASC_GTRGAMMA –asc-corr=lewis

The above will run a standard tree search with the ascertainment bias correction by Paul Lewis.

The two other options are a bit more tricky to use. They have been designed for cases where the exact number of invariable sites for the dataset is actually known (felsestein option) or the exact frequencies of each invariable character state is known (stamatakis option). Note that in such cases the number of invariable sites can be larger than the

number o! variable sites in the alignment. This is not the "ase !or the Le / is "orre"tion% be"ause the maths break do / n.

2or those t / o "orre"tions / e thus need to tell RAxML / hat the number o! invraiable sites !or ea"h)artition 6felsenstein8 is and / hat the numbers o! invariable sites)er state !or ea"h)artition is 6stamatakis8. This is best handled via the)artition Ile% hen"e even i! you only have one)artition% you nonetheless need to s)e"i!y a)artition Ile via -q to)ass this in!ormation to RAxML.

Example 2:
```
raxmlHPC -p 12345 -s alg -n TEST -m ASC_GTRGAMMA
                --asc-corr=stamatakis -q part
```

Example 3:
```
raxmlHPC -p 12345 -s alg -n TEST -m ASC_GTRGAMMA
                --asc-corr=felsenstein -q part
```

Relo / % / e give an exam)le o! su"h a)artition Ile=

```
[asc~p1.txt], ASC_DNA, p4=1-1000
[asc~p2.txt], ASC_DNA, p5=1001-1965
```

Here% the Irst entries in ea"h line%)rovide the name o! the Ile "ontaining the invariable site "ounts or !re4uen"ies !or ea"h)artition that are stored in)lain text Iles "alled p1.txt and p2.txt. 9ote that% the res)e"tive Ile names must be)re"eded by the asc key / ord and the ~ se)arator symbolG
Thus% ea"h)artition is asso"iated / ith a Ile "ontaining these "ounts.

2or the 2elsenstein "orre"tion% these Iles "ould look like this=

```
)&.txt=
1000

)2.txt=
2000
```

* ith that / e tell RAxML% that the likelihood Irst)artition shall be "orre"ted !or 1000 invariable sites% / hile the likelihood o! the se"ond)artition shall be "orre"ted !or 2000 invariable sites. 9ote that% ea"h Ile 6p1.txt and p2.txt8 must only "ontain a single line "om)rising integer valuesG

2or my "orre"tion% the Iles "ould look as !ollo / s !or . 9A data=

```
)&.txt=
250 300 400 100

)2.txt=
500 300 200 200
```

* ith that / e tell RAxML that the likelihood o! the Irst)artition shall be "orre"ted !or 250 sites "onsisting o! As% 300 sites "onsisting o! 0s% 400 sites o! (s and 100 sites o! Ts. 2or the se"ond)artition the likelihood is "orre"ted !or 500 invariable sites o! As% 300 0s% 200 (s and 200 Ts.
The diMeren"e among the t / o "orre"tions is that / ith #oe 2elsensteinHs "orre"tion / e kno / ho / many invariable sites there are% but not their "om)osition% so / e "orre"t !or the absen"e o! &$$$ invariable sites in the Irst)artition that "ould "onsists o! As or 0s or (s or

Ts. My "orre"tion "an be used /hen /e do kno/ the exa"t !re4uen"ies o! invariable site)atterns.

WARNING: *hen generating these as"ertainment bias "orre"tion Iles 6e.g.% `p1.txt` and `p2.txt`8 !or my "orre"tion)ay)arti"ular attention to the order o! states that "orres)onds to the order o! !re4uen"ies o! invariable sites)er state=

```
BINARY:      0,1
DNA:         A, C, G, T
PROTEIN:     A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V
```

`--flag-check` When using this option, RAxML will only check if all command line flags specifed are available and then exit with a message listing all invalid command line flags or with a message stating that all flags are valid.

This o)tion is intended !or B[rd])arty /ra))er so!t/are that invokes RAxML. t allo/s !or "he"king be!orehand i! "ertain o)tions are available in the RAxML version being used or not. The o)tion /ill Aust analy7e the "ommand line !or valid@invalid Nags and then exit /ith an a))ro)riate message.

Example 1:
```
raxmlHPC --flag-check -f a -s alg -p 12345 -N 20 -x 12345
      -m GTRGAMMA -n T1
```

Output: `All options supported`

Example 2:
```
raxmlHPC --flag-check -f a -s alg -p 12345 -N 20 -x 12345
      -m GTRGAMMA --nonsense-option -n T1
```

Output: `Option --nonsense-option not supported`

`--auto-prot=ml|bic|aic|aicc` When using automatic protein model selection you can chose the criterion for selecting these models. RAxML will test all available prot subst. models except for LG4M, LG4X, and GTR-based models with and without empirical base frequencies.
You can chose between ML score based selection and the BIC, AIC, and AICc criteria.

DEFAULT: ml

Example:
```
raxmlHPC -s alg -p 12345 -m PROTGAMMAAUTO --auto-prot=bic -n T1
```

The above invo"ation /ill sele"t bet/een)rotein models using the Rayesian n!ormation 0riterion.

2or understanding the !ollo/ing three o)tions that all re!er to the standard)la"ement out)ut !ormat `6.jplace` out)ut Ile o! the E;A8 /e have de Ined% it /ill be very hel)!ul i! you read the "orres)onding)a)er=
htt)=@@/ / /.)losone.org@arti"le@in!o SBAdoi S22&$.&BX& S22Aournal.)one.$$B&$$Q

`--epa-keep-placements=number` specify the number of potential placements you want to keep for each read in the EPA algorithm .
Note that, the actual values printed will also depend on the settings for

--epa-prob-threshold=threshold !

DEFAULT: 7

Example: raxmlHPC -f v --epa-keep-placements=100 -t tree -m GTRCAT -s alg -n T1

Here% RAxML / ill)rint at most &$$)la"ements to the .jplace)la"ement Ile% de)ending on the setting o! --epa-prob-threshold.

--epa-prob-threshold=threshold specify a percent threshold for including potential placements of a read depending on the maximum placement weight for this read. If you set this value to 0.01 placements that have a placement weight of 1 per cent of the maximum placement will still be printed to file if the setting of --epa-keep-placements allows for it

DEFAULT: 0.01

Example: raxmlHPC -f v –epa-prob-threshold=0.05 -t tree -m GTRCAT -s alg -n T1

Here% RAxML / ill)rint those)la"ements to the .jplace)la"ement Ile that have a likelihood / eight o! at least PS o! that o! the maximum)la"ement / eight !or the s)e"iI" read and i! the number o!)la"ements to)rint as s)e"iIed by Wepa-keep-placements has not been ex"eeded.

--epa-accumulated-threshold=threshold specify an accumulated likelihood weight threshold for which different placements of read are printed to file. Placements for a read will be printed until the sum of their placement weights has reached the threshold value.
Note that, this option can neither be used in combination with --epa-prob-threshold nor with --epa-keep-placements!

Example: raxmlHPC -f v --epa-accumulated-threshold=0.95 -t tree -m GTRCAT -s alg -n T1

Here% RAxML / ill)rint those)la"ements o! a read to the .jplace)la"ement Ile until their a""umulated likelihood / eights has rea"hed a value o! $.QP 6out o! a total o! &.$8.

--JC69 specify that all DNA partitions will evolve under the Jukes-Cantor model, this overrides all other model specifications for DNA partitions.

DEFAULT: Off

Example: raxmlHPC -p 12345 -m GTRGAMMA -s alg --JC69 -n T1

--K80 specify that all DNA partitions will evolve under the K80 model, this overrides all other model specifications for DNA partitions.

DEFAULT: Off

9ote that% the out)ut o! the)rogram might look a bit / eird% sin"e unlike in the deInition o! the model% RAxML a"tually estimates the rates !rom A O (and 0 O T / hile all other rates are set to &.$. 9ote that% this does not matter% sin"e the rates in the rate matrix are relative

rates[the results 6likelihoods8 / ill be the same.

Example: raxmlHPC -p 12345 -m GTRGAMMA -s alg --K80 -n T1

--HKY85 specify that all DNA partitions will evolve under the HKY85 model, this overrides all other model specifications for DNA partitions.

DEFAULT: Off

9ote that% the out)ut o! the)rogram might look a bit / eird% sin"e unlike in the de I nition o! the model% RAxML a"tually estimates the rates !rom A O (and 0 O T / hile all other rates are set to &.$. 9ote that% this does not matter% sin"e the rates in the rate matrix are relative rates[the results 6likelihoods8 / ill be the same.

Example: raxmlHPC -p 12345 -m GTRGAMMA -s alg --HKY85 -n T1

--set-thread-affinity specify that thread-to-core affinity shall be set by RAxML for the hybrid MPI-PThreads version

DEFAULT: Off

<ou might have to use this o)tion !or the hybrid M; +;Threads version o! RAxML on some "luster installations su"h that the "ode intera"ts)ro)erly / ith the s"heduling system. 3n other systems ho / ever% enabling this o)tion may yield)roblemsG tHs un!ortunately a matter o! trial and error that is due to ho / diMerent s"hedulers handle hybrid M; +;Threads "odesG

Example: raxmlHPC-HYBRID -m GAMMA -s alg -p 12345 -N 20 --set-thread-affinity -n T1

--bootstop-perms=number specify the number of permutations to be conducted for the bootstopping/bootstrap convergence test. The allowed minimum number is 100!

DEFAULT: 100

<ou might / ant to set this to &$$$ or higher to redu"e the variation in the number o! bootstra) trees you "al"ulate on noisy@di L "ult datasets. The varian"e o! the number o! bootstra) trees that bootstra) runs / ith diMerent random number seeds / ill generate / ill de"rease by in"reasing the number o!)ermutations.

Example: raxmlHPC -B 0.02 --bootstop-perms=1000 -b 12345 -p 12345 -# AUTOMR -s alg -m GTRCAT -n TEST

--quartets-without-replacement specify that quartets are randomly subsampled, but without replacement.

DEFAULT: random sampling with replacements

Example: raxmlHPC -s alg -p 12345 -m GTRGAMMA -f q -N 50 --quartets-without-replacement -n TEST

--print-identical-sequences specify that RAxML shall automatically generate a .reduced alignment with all undetermined columns removed, but without removing exactly identical sequences

DEFAULT: identical sequences will also be removed in the .reduced file

Example: raxmlHPC -s alg -p 12345 -m GTRGAMMA --print-identical-sequences -n TEST

IX. Output Files

.e)ending on the sear"h)arameter settings RAxML /ill /rite a number o! out)ut Iles.
The most im)ortant Iles% a run named -n exampleRun /ill /rite% are listed belo/.
Sin"e the number and names o! out)ut Iles vary de)ending on the RAxML o)tions you used% the easiest /ay to list all out)ut Iles o! the run is to ty)e the !ollo/ing Linux "ommand=

ls *exampleRun*

RaxML_info.exampleRun: "ontains in!ormation about the model and algorithm used and ho/ RAxML /as "alled. The Inal (AMMA+based likelihood6s8 as /ell as the al)ha sha)e)arameter6s8 are)rinted to this Ile. n addition% i! the rearrangement setting /as determined automati"ally 6-i has not been used8 the rearrangement setting !ound by the)rogram /ill be indi"ated.
This is the most important output file be"ause it tells you /hat RAxML did and is al/ays /ritten irres)e"tive o! the "ommand line o)tion. n addition it "ontains in!ormation about all other out)ut Iles that /ere /ritten by your run.

RAxML_log.exampleRun: A Ile that)rints out the time% likelihood value o! the "urrent tree and number o! the "he"k)oint Ile 6i! the use o! "he"k)oints has been s)e"iIed8 a!ter ea"h iteration o! the sear"h algorithm. n the last line it also "ontains the Inal likelihood value o! the Inal tree to)ology. This Ile is not /ritten i! multi)le bootstra)s are exe"uted% i.e. +# and -b have been s)e"iIed. n "ase o! a multi)le in!eren"e on the original alignment 6-# o)tion8 the Log+2iles are numbered a""ordingly.

RAxML_result.exampleRun: 0ontains the Inal tree to)ology o! the "urrent run. This Ile is also /ritten a!ter ea"h iteration o! the sear"h algorithm% su"h that you "an restart your run /ith -t in "ase your "om)uter "rashed. This Ile is not /ritten i! multi)le bootstra)s are exe"uted% i.e. -# and -b have been s)e"iIed.

RAxML_parsimonyTree.exampleRun: "ontains the randomi7ed)arsimony starting tree i! the)rogram has not been)rovided a starting tree by -t. Ho/ever% this Ile /ill not be /ritten i! a multi)le bootstra) is exe"uted using the -# and -b o)tions.

RAxML_randomTree.exampleRun: "ontains the "om)letely random starting tree i! the)rogram /as exe"uted /ith -d.

RAxML_checkpoint.exampleRun.checkpointNumber: ;rinted i! you s)e"iIed by -j that "he"k)oints shall be /ritten. 0he"k)oints are numbered !rom $ to n /here n is the number o! iterations o! the sear"h algorithm. Moreover% the "he"k)oint Iles are additionally numbered i! a multi)le in!eren"e on the original alignment has been s)e"iIed using -#. *riting o! "he"k)oint Iles is disabled /hen a multi)le bootstra) is exe"uted.

RAxML_bootstrap.exampleRun: ! a multi)le bootstra) is exe"uted by -# and -b or -x all Inal

bootstra))ed trees /ill be /ritten to this one% single Ile.

`RAxML_bipartitions.exampleRun:` ! you used the `-f b` o)tion% this Ile /ill "ontain the in)ut tree /ith "onIden"e values !rom $ to &$$ dra/n on its nodesG t is also)rinted /hen `+f a -x` have been s)e"iIed% at the end o! the analysis the)rogram /ill dra/ the RS su))ort values on the best tree !ound during the ML sear"h.

`RAxML_bipartitionsBranchLabels.exampleRun:` 0ontains the same in!ormation as the Ile above% but su))ort values are "orre"tly dis)layed as 9e/i"k bran"h labels and not node labelsG **Support values always refer to branches/splits of trees and never to nodes of the tree.** 9ote that% some tree vie/ers have)roblems dis)laying bran"h labels% they are ho/ever)art o! the standard 9e/i"k !ormat.

`RAxML_bipartitionFrequencies.exampleRun:` 0ontains the)air+/ise bi)artition !re4uen"ies o! all trees "ontained in Iles)assed via `-t` and `-z` /hen the `-f m` o)tion has been used.

`RAxML_perSiteLLs.exampleRun:` 0ontains the)erWsite log likelihood s"ores in Tree)u77le !ormat !or usage /ith 039SEL 6htt)=@@/ / /.is.tite"h.a".A)@\shimo@)rog@"onsel@8. This Ile is only)rinted /hen `-f g` is s)e"iled.

`RAxML_bestTree.exampleRun:` 0ontains the best+s"oring ML tree o! a thorough ML analysis.

`RAxML_distances.exampleRun:` 0ontains the)air+/ise ML+based distan"es bet/een all taxon+)airs in the alignment. This Ile is only)rinted /hen the `-f x` o)tion is used.

X. Computing TC and IC values

n the !ollo/ing)rovide some more detailed exam)les !or "om)uting the 0 and T0 metri"s !rom Sali"hos and Rokas 2$&B htt)=@@/ / /.n"bi.nlm.nih.gov@)ubmed@2B'PX2P8

The method is des"ribed in more detail in the !ollo/ing)a)er /rote /ith Sali"hos and Rokas= htt)=@@mbe.ox!ordAournals.org@"ontent@early@2$&?@$2@$X@molbev.msu$'&.abstra"tkeyty)eDre!1iAkeyD 'P2u(9x$H7R23/

Also note that% as o! version 8.2.$ *e "an "ondu"t T0@0 "al"ulations on "olle"tions o!)artial gene trees. The "orre"tions /e a))ly !or)artial gene trees are des"ribed in this)a)er here= htt)=@@dx.doi.org@&$.&&$&@22PB

> **Warning:** The de!ault T0@0 "al"ulations are not done exa"tly as des"ribed in our MRE)a)er. To have RAxML do exa"tly /hat is des"ribed in the)a)er%)leased edit Ile `bipartitionList.c` by "ommenting out or removing the line `#define BIP_FILTER`
>
> The /ay it is im)lemented in RAxML by de!ault and /here it diMers !rom the version des"ribed in the)a)er regards the "al"ulation o! the 0A and T0A s"ores. *hen there are several "onNi"ting bi)artitions% the standard method takes all o! them into a""ount !or "om)uting the 0A s"ore% as long as their !re4uen"y ex"eeds a !re4uen"y threshold o! PS. t does ho/ever not a""ount !or the !a"t that some o! these "onNi"ting bi)artitions% /ith res)e"t to the re!eren"e bi)artition do not "onNi"t /ith ea"h other% but are in !a"t "om)atible /ith ea"h other. Thus% by de!ault RAxML only uses those bi)artitions that "onNi"t /ith the re!eren"e bi)artition that are also mutually in"om)atible /ith ea"h other to "al"ulate the 0A s"ore. Thus% less bi)artitions /ill ty)i"ally be used to "al"ulate the 0A and "onse4uently the T0A s"ores. My)ersonal o)inion is that this is more reasonable% sin"e /e are not "ounting "onNi"ts several times in "ase these "onNi"ts emerge !rom "om)atible bi)artitions.

(iven a set o! gene trees% RAxML "an dire"tly "al"ulate a maAority rule "onsensus 6MR in RAxML terminology8 as /ell as an extended maAority rule "onsensus tree 6MRE in RAxML terminology8 on this set that has every internode 6that is% internal bran"h8 annotated by their res)e"tive 0 and 0A s"ores.

2or instan"e% to "om)ute the 0% 0A% T0% and T0A s"ores !or a given set o! gene trees on a MR0 tree you /ould ty)e=

```
raxmlHPC -L MR -z 1070_yeast_genetrees.tre -m GTRCAT -n T1
```

/here

- `-L MR` s)e"iIes that the s"ores /ill be dis)layed on a MR tree that is "om)uted by RAxML
- `-z 1070_yeast_genetrees.tre` s)e"iIes the IIename that "ontains the set o! gene trees 6/hi"h are the maximum likelihood trees !rom the &%X yeast genes analy7ed by Sali"hos% and Rokas 2$&B% and /hi"h are)rovided as su))lementary data to this manus"ri)t8
- `-m GTRCAT` is an arbitrary substitution model 6this /ill have no eMe"t /hatsoever% but is re4uired as in)ut to RAxML8
- `-n T1` is the run . that is a))ended to out)ut IIes.

RAxML /ill automati"ally build the MR tree% annotate it /ith the 0 and 0A s"ores% and re)ort both in an out)ut IIe named `RAxML_MajorityRuleConsensusTree_IC.T1`% /hi"h /ill look like this=

```
(Scer,Spar,(Smik,(Skud,(Sbay,(Scas,(Cgla,(Kpol,(Zrou,((Clus,((Psti,((Ctro,
(Calb,Cdub):1.0[0.95,0.95]):1.0[0.77,0.77],
(Cpar,Lelo):1.0[0.76,0.76]):1.0[0.75,0.75]):1.0[0.11,0.11],
(Cgui,Dhan):1.0[0.02,0.07]):1.0[0.02,0.08]):1.0[0.97,0.97],((Sklu,
(Kwal,Kthe):1.0[0.97,0.97]):1.0[0.32,0.23],
(Agos,Klac):1.0[0.08,0.08]):1.0[0.04,0.10]):1.0[0.59,0.47]):1.0[0.02,0.02]):1.0[0.1
1,0.11]):1.0[0.02,0.02]):1.0[0.97,0.97]):1.0[0.05,0.14]):1.0[0.30,0.27]):1.0[0.54,0
.54]);
```

2or ea"h internode or internal bran"h o! the "onstru"ted MR tree% RAxML /ill assign an `length[x,y]` bran"h label% /here `length` "orres)onds to the bran"h length 6be"ause this is a MRE tree% all internal bran"h lengths have been arbitrarily set to `1.0` by de!ault8% `x` "orres)onds to the 0 s"ore and `y` to the 0A s"ore.

RAxML /ill also "al"ulate the T0 and T0A s"ores !or the MR tree% as /ell as the relative T0 and T0A s"ores that are normali7ed by the maximum)ossible T0 and T0A s"ores !or a !ully bi!ur"ating tree !rom the same number o! taxa.

The s"ores are dis)layed in the terminal out)ut and in the `RAxML_info.runID` standard out)ut IIe asso"iated /ith the run 6in this "ase `RAxML_info.T1`8 and /ill look like this=

Tree certainty for this tree: 7.642240
Relative tree certainty for this tree: 0.382112

```
Tree certainty including all conflicting bipartitions (TCA) for this tree: 7.580023
Relative tree certainty including all conflicting bipartitions (TCA) for this tree: 0.379001
```

(iven a set o! gene trees% RAxML "an also dire"tly "al"ulate an extended MR tree on this set that has every internode 6that is% internal bran"h8 annotated by their res)e"tive 0 and 0A s"ores.
The)arti"ularly "om)ute+intensive in!eren"e o! extended MR0 trees 6Inding the o)timal extended MR0 tree is% in !a"t% 9;+hard[;hilli)s% and *arno/ &QQ'8 relies on RAxML:s !ast)arallel im)lementation. Thus i! you use the ;Threads version o! RAxML% this)art /ill run in)arallel.

To "om)ute 0% 0A% T0% and T0A s"ores on an extended MR tree you /ould ty)e=

```
raxmlHPC -L MRE -z 1070_yeast_genetrees.tre -m GTRCAT -n T2
```

RAxML "an "om)ute MR and extended MR trees% using both !ully bi!ur"ating and)artially resolved@multi!ur"ating trees as an in)ut. RAxML "an also "om)ute stri"ter MR trees /ith arbitrary threshold settings that range bet/een P& and &$$S. 2or instan"e% by ty)ing

```
raxmlHPC -L T_75 -z 1070_yeast_genetrees.tre -m GTRCAT -n T3
```

RAxML /ill dis)lay 0% 0A% T0 and T0A s"ores on a MR tree that only in"ludes those bi)artitions that have] XPS su))ort.

*e have also im)lemented an o)tion 6-f i8 that allo/s the user to "al"ulate and dis)lay 0% 0A% T0% and T0A s"ores onto a given% stri"tly bi!ur"ating re!eren"e tree 6!or exam)le% the best+kno/n ML tree8.
This is analogous to the standard -f b o)tion in RAxML that dra/s bootstra) su))ort values !rom a set o! bootstra) trees onto a re!eren"e)hylogeny. The o)tion "an be invoked by ty)ing

```
raxmlHPC -f i -t yeast_concatenationtree.tre -z 1070_yeast_genetrees.tre -m GTRCAT -n T4
```

9ote that% the tree "ontained in !ile yeast_concatenationtree.tre needs to be stri"tly bi!ur"ating and "ontain bran"h lengths. n this exam)le% the yeast_concatenationtree.tre !ile is the best+kno/n maximum likelihood tree re"overed by "on"atenation analysis o! the &%X yeast genes !rom the)a)er.
,sing this "ommand% RAxML /ill annotate the tree in yeast_concatenationtree.tre /ith the 0 and 0A s"ores% and re)ort both in an out)ut !ile named RAxML_IC_Score_BranchLabels.T4% /hi"h /ill look like this=

```
(((((((Clus:0.47168135428609103688((((Lelo:0.30356174702769450624,Cpar:0.2549087423
9480920682):0.13023178275857649755[0.76,0.76],
(Ctro:0.18383414558272206940(Calb:0.04124660275465741321,Cdub:0.0429080158839683228
9):0.14526604486383792869[0.95,0.95]):0.12355825028654655873[0.77,0.77]):0.17335821
030783615804[0.75,0.75],Psti:0.42255112174261910685):0.07862888222310976461[0.11,0.
11],
(Cgui:0.45961028886034632768,Dhan:0.28259245937168109286):0.05586015476156453580[0.
02,0.07]):0.08116340505230199009[0.02,0.08]):1.03598510402913923656[0.97,0.97],
((Agos:0.53332956655591512440,Klac:0.47072785596320687596):0.08132006357704427146[0
.08,0.08],
((Kthe:0.17123899487739652203,Kwal:0.17320923240031221857):0.25620117495110567019[0
.97,0.97],Sklu:0.24833228915799765435):0.05646992617871094550[0.32,0.23]):0.0523630
6187235122145[0.04,0.10]):0.10686517691208799463[0.59,0.47],Zrou:0.4130783368556378
2877):0.03792570537296727218[0.02,0.02],Kpol:0.43287284049576529865):0.045603416931
36910068[0.11,0.11],Cgla:0.49584136365135367264):0.04363310339731014259[0.02,0.02],
Scas:0.37212829744050218705):0.29362133996280515014[0.97,0.97],
(Skud:0.06926467973344750673,(Smik:0.06535810850036427588,
(Scer:0.04285848856634000975,Spar:0.03030513540244994877):0.02506719066056842596[0.
54,0.54]):0.02459323291555862850[0.30,0.27]):0.02524223867026276907[0.05,0.14],Sbay
:0.06506923220637816918);
```

2or ea"h internode or internal bran"h o! this out)ut tree RAxML /ill assign a length[x,y] bran"h label% /here length "orres)onds to the original bran"h length in the in)ut tree)assed via -t% x "orres)onds to the 0 s"ore and y to the 0A s"ore. RAxML /ill also dis)lay the T0 and T0A s"ores o! this tree both in the terminal out)ut and in the RAxML_info.T4 out)ut !ile asso"iated /ith the run.

t should !urther be noted that the 0 and 0A s"ores are re)resented as bran"h labels% sin"e% as is the "ase !or bootstra) su))ort values% in!ormation asso"iated to internodes@s)lits@bi)artitions o! a tree al/ays re!ers to bran"hes and not nodes.
Ea"h tree vie/er 6e.g.% .endros"o)e htt)=@@ab.in!.uni+tuebingen.de@so!t/are@dendros"o)e@8 that "an)ro)erly)arse the 9e/i"k tree !ormat is able to dis)lay these bran"h labels.
The rationale !or not)roviding 0 and 0A s"ores as node labels is that% some vie/ers may not)ro)erly rotate the node labels /hen the tree is re+rooted by the user% /hi"h /ill lead to an erroneous bran"h+to+ 0 and bran"h+to+ 0A s"ore asso"iation.

*hen "al"ulating 0 and 0A s"ores on extended MR trees or /hen dra/ing 0 and 0A s"ores onto a given re!eren"e tree it may o""ur that the bi)artition that has been in"luded in the tree has lo/er su))ort than one or more "onNi"ting bi)artitions. n this "ase% RAxML /ill re)ort 0 and 0A s"ores on the in!erred tree /ith negative signs.

2inally% /e have im)lemented a verbose out)ut o)tion that allo/s users to !urther s"rutini7e)arti"ularly interesting "onNi"ting bi)artitions.
-erbose mode is a"tivated by adding the –C "ommand line s/it"h to any o! the above exam)les. n verbose mode RAxML /ill generate t/o additional ty)es o! out)ut Iles= 3ne set o! Iles "ontaining one in"luded bi)artition and the "orres)onding "onNi"ting bi)artitions in 9e/i"k !ormat 6"alled `RAxML_verboseIC.runID.0` … `RAxML_verboseIC.runID.N-1`% /here `N` is the number o! bi)artitions in the tree8 and an out)ut Ile that lists all bi)artitions 6in"luded and "onNi"ting8 in a ;H<L ;+like !ormat 6"alled `RAxML_verboseSplits.runID`8.

2or exam)le% by adding –C to the)revious "ommand

```
raxmlHPC -f i -t yeast_concatenationtree.tre -z 1070_yeast_genetrees.tre -m GTRCAT -n T5 -C
```

/ill)rodu"e 2$ Iles 6one !or ea"h o! the 2$ bi)artitions)resent in the `yeast_concatenationtree.tre`8 named `RAxML_verboseIC.T5.0`, `RAxML_verboseIC.T5.1`, …, `RAxML_verboseIC.T5.19`

2or exam)le% the `RAxML_verboseIC.T5.0` Ile /ill look like this=

```
((Cpar, Lelo),(Scer, Smik, Skud, Cgla, Kpol, Zrou, Kwal, Kthe, Agos, Klac, Clus,
Cgui, Psti, Ctro, Calb, Cdub, Dhan, Sklu, Scas, Sbay, Spar));
((Cpar, Ctro, Calb, Cdub),(Scer, Smik, Skud, Cgla, Kpol, Zrou, Kwal, Kthe, Agos,
Klac, Clus, Cgui, Psti, Lelo, Dhan, Sklu, Scas, Sbay, Spar));
```

/here the Irst 9e/i"k string re)resents the bi)artition that /as in"luded in the `yeast_concatenationtree.tre` and all !ollo/ing 9e/i"k strings re)resent the "orres)onding "onNi"ting bi)artitions in des"ending order o! their !re4uen"y o! o""urren"e n the "ase o! the `RAxML_verboseIC.T5.0` Ile the Irst bi)artition% /hi"h is in"luded in the `yeast_concatenationtree.tre` "onNi"ts /ith only one other bi)artition% /hi"h is listed as the se"ond bi)artition.

Analogously% the out)ut Ile that lists all bi)artitions 6in"luded and "onNi"ting8 in a ;H<L ;+like !ormat 6`RAxML_verboseSplits.T5`8% looks like this=

```
1. Scer
2. Smik
3. Skud
4. Cgla
5. Kpol
6. Zrou
```

7. Kwal
8. Kthe
9. Agos
10. Klac
11. Clus
12. Cgui
13. Psti
14. Cpar
15. Lelo
16. Ctro
17. Calb
18. Cdub
19. Dhan
20. Sklu
21. Scas
22. Sbay
23. Spar

partition:
----- ----- ---** ----- --- 956/89.345794/0.761406
----- ----- ---*- ***-- --- 39/3.644860/0.761406

partition:
----- ----- ----- -**-- --- 1051/98.224299/0.949483
----- ----- ----- **--- --- 6/0.560748/0.949483

.
.
.
partition:
--*** ***** ***** ***** **- 641/59.906542/0.303620
-**-- ----- ----- ----- -*- 148/13.831776/0.303620
-*-** ***** ***** ***** *-- 114/10.654206/0.303620

partition:
-**** ***** ***** ***** **- 825/77.102804/0.545775
-**-- ----- ----- ----- -** 87/8.130841/0.545775

bestWs"oring ML tree on the original alignment.

To Aust do a RS sear"h you /ould ty)e=

```
raxmlHPC -x 12345 -p 12345 -# 100 -m GTRGAMMA -s ex_al -n TEST
```
or using the bootstra) "onvergen"e "riterion=

```
raxmlHPC -x 12345 -p 12345 -# autoMRE -m GTRGAMMA -s ex_al -n TEST
```

This /ill "ondu"t ra)id bootstra))ing and and an ML sear"h under the (AMMA model o! rate heterogeneity.

9o/% i! you /ant to run a !ull analysis% i.e.% RS and ML sear"h ty)e=

```
raxmlHPC -f a -x 12345 -p 12345 -# 100 -m GTRGAMMA -s ex_al -n TEST
```

This /ill Irst "ondu"t a RS sear"h and on"e that is done a sear"h !or the bestWs"oring ML tree. Su"h a)rogram run /ill return the bootstra))ed trees 6RAxML_bootstrap.TEST8% the best s"oring ML tree 6RAxML_bestTree.TEST8 and the RS su))ort values dra/n on the best+s"oring tree as node labels 6RAxML_bipartitions.TEST8 as /ell as% more "orre"tly sin"e su))ort values re!er to bran"hes as bran"h labels 6RAxML_bipartitionsBranchLabels.TEST8.

2inally% note that% by in"reasing the number o! RS re)li"ates via -# you /ill also make the ML sear"h more thorough% sin"e !or ML o)timi7ation every Pth RS tree is used as a starting)oint to sear"h !or ML trees.

*hen -# autoMRE is s)e"iIed RAxML /ill exe"ute a maximum o! &$$$ RS re)li"ate sear"hes% but it may% o! "ourse "onverge earlier.

2rom /hat have observed so !ar% this ne/ ML sear"h algorithm yielded better trees than /hat is obtained via 2$ standard ML sear"hes on distin"t starting trees !or all datasets /ith ^ &%$$$ se4uen"es.

2or larger datasets it might be /orth/hile to "ondu"t an additional ML sear"h as des"ribed belo/% Aust to be sure.

Warning: note that the ra)id RS sear"h /ill "urrently ignore "ommands asso"iated to user tree Iles)assed via -t or -z.

Ho/ever% the "onstraint and ba"kbone tree o)tions 6-g and -r8 do /ork /ith ra)id RS.

The Hard & Slow Way

.es)ite the observation that the de!ault)arameters and the ra)id RS and ML algorithm des"ribed above /ork /ell in most)ra"ti"al "ases% a good thing to do is to ada)t the)rogram)arameters to your alignment.
This re!ers to a *good* setting !or the rate "ategories o! -m GTRCAT and the initial rearrangement setting.

! you use)artitioned models you should add -q partitionFileName to all o! the !ollo/ing "ommands.

Getting the Initial Rearrangement Setting right

! you don:t s)e"i!y an initial rearrangement setting /ith the –i o)tion the)rogram /ill automati"ally deter+mine a good setting based u)on the randomi7ed M; starting tree. t /ill take the starting tree and a))ly la7y subtree rearrangements /ith a rearrangement setting o! P% &$% &P% 2$% 2P. The minimum setting that yields the best likelihood im)rovement on the starting trees /ill be used as initial rearrangement setting.

This)ro"edure "an have t/o disadvantages=

> 2irstly% the initial setting might be very high 6e.g. 2$ or 2P8 and the)rogram /ill slo/ do/n "onsiderably.
>
> Se"ondly% a rearrangement setting that yields a high im)rovement o! likelihood s"ores on the starting tree might let the)rogram get stu"k earlier in some lo"al maximum 6this behavior "ould already be observed on a real dataset /ith about &%Q$$ taxa8.

There!ore% you should run RAxML a "ou)le o! times 6the more the better8 /ith the automati" determination o! the rearrangement setting and /ith a)re+deIned value o! &$ /hi"h)roved to be suL "iently large and eL "ient in many)ra"ti"al "ases.

n the exam)le belo/ /e /ill do this based on P Ixed starting trees. So let:s Irst generate a "ou)le o! randomi7ed M; starting trees.

9ote that in RAxML you also al/ays have to s)e"i!y a substitution model% regardless o! /hether you only /ant to "om)ute an M; starting tree /ith the –y o)tion. 9ote that% /e have to)ass diMerent random number seeds via –p to obtain distin"t starting trees here6

```
raxmlHPC –y –p 12345 –s ex_al –m GTRCAT –n ST0
...
raxmlHPC –y –p 34556 –s ex_al –m GTRCAT –n ST4
```

Then% in!er the ML trees !or those starting trees using a Ixed setting –i 10

```
raxmlHPC –f d –i 10 –m GTRCAT –s ex_al –t RAxML_parsimonyTree.ST0 –n FI0
...
raxmlHPC –f d –i 10 –m GTRCAT –s ex_al –t RAxML_parsimonyTree.ST4 –n FI4
```

and then using the automati"ally determined setting on the same starting trees=

```
raxmlHPC –f d –m GTRCAT –s ex_al –t RAxML_parsimonyTree.ST0 –n AI0
...
raxmlHPC –f d –m GTRCAT –s ex_al –t RAxML_parsimonyTree.ST4 –n AI4
```

The setting that yields the best Inal likelihood s"ores as automati"ally "om)uted under the (AMMA model o! rate heterogeneity should then be used !or subse4uent analyses.

Getting the Number of Categories right

Another issue is to get the number o! rate "ategories right. .ue to the redu"ed memory !oot)rint and signiI"antly redu"ed in!eren"e times the re"ommended model to use /ith RAxML on large dataset is (TR0AT i! you are doing runs to Ind the best+kno/n ML tree on the original alignment and !or bootstra))ing.

Thus, you should experiment with a couple of -c settings and then look which gives you the best (gamma-based likelihood value.

Suppose that in the previous section you found that automatically determining the rearrangement setting works best for your alignment.

You should then re-run the analyses with distinct "-" settings by increments of e.g. &P rate categories

```
raxmlHPC -f d -c 10 -m GTRCAT -s ex_al -t RAxML_parsimonyTree.ST0 -n C10_0
...
raxmlHPC -f d -c 10 -m GTRCAT -s ex_al -t RAxML_parsimonyTree.ST4 -n C10_4
```

You don't need to run it with the default setting of `-c 25` since you already have that data, such that you
can continue with ...

```
raxmlHPC -f d -c 40 -m GTRCAT -s ex_al -t RAxML_parsimonyTree.ST0 -n C40_0
...
raxmlHPC -f d -c 40 -m GTRCAT -s ex_al -t RAxML_parsimonyTree.ST4 -n C40_4
```

and so on and so forth.

Since the (TR0AT approximation is still a new / "one)t little is known about the appropriate setting for `-c 25`.
However, empirically `-c 25` worked best on &Q real-world alignments. So testing up to `-c 55` should usually be sufficient, except if you notice a tendency for final (TR(AMMA likelihood values to further improve with increasing rate category number.

Thus, the assessment of the *good* `-c` setting should once again be based on the final (TR(AMMA likelihood values.

If you don't have the time or computational power to determine both *good* `-c` and `-i` settings you should rather stick to determining `-i` since it has shown to have a greater impact on the final results.

Also note, that increasing the number of distinct rate categories has a negative impact on execution times. 2inally, if the runs with the automatic determination of the rearrangement settings from the previous Section have yielded the best results you should then use exactly the same rearrangement settings for each series of experiments to determine a *good* `-c` setting.
The automatically determined rearrangement settings can be retrieved from files `RAxML_info.AI_0 ... RAxML_info.AI_4`

Finding the Best-Known Likelihood tree (BKL)

As already mentioned RAxML uses randomi7ed stepwise addition order parsimony starting trees on which it then initiates an ML-based optimi7ation. Those trees are obtained by using a randomi7ed stepwise addition sequence to insert one taxon after the other into the tree. When all sequences have been inserted a complete subtree rearrangements (also called subtree pruning re-grafting) with a fixed rearrangement distance of 2$ are executed to further improve the M; score.
The concept to use randomi7ed M; starting trees in contrast to the 9# (9eighbor #oining) starting trees many other ML programs use, is regarded as an advantage of RAxML. This allows the program to start ML optimi7ations of the topology from a distinct starting point in the immense topological search space each time.
Therefore, RAxML is more likely to find good ML trees if executed several times.

This also allo/s you to build a "onsensus tree out o! the Inal tree to)ologies obtained !rom ea"h individual run on the original alignment. Ry this and by "om)aring the Inal likelihoods you "an get a !eeling on ho/ stable 6)rone to get "aught in lo"al maxima8 the sear"h algorithm is on the original alignment.

<ou "an also use the -f r o)tion to "om)ute)air/ise to)ologi"al Robinson+2oulds distan"es bet/een the ML trees you have !ound.

Thus% i! you have su L "ient "om)uting resour"es available% in addition to bootstra))ing% you should do multi)le in!eren"es 6 exe"uted 2$$ in!eren"es in some re"ent real+/orld analyses /ith Riologists8 /ith RAxML on the original alignment.

3n smaller datasets or large)hylogenomi" datasets it /ill also be /orth/hile to use the -d o)tion !or a "ou)le o! runs to see ho/ the)rogram behaves on "om)letely random starting trees. This is /here the -# o)tion as /ell as the)arallel M; version raxmlH;0+M; "ome into)lay.

So% to exe"ute a multi)le in!eren"e on the original alignment on a single)ro"essor Aust ty)e=

```
raxmlHPC -f d -p 12345 -m GTRCAT -s ex_al -# 10 -n MultipleOriginal
```

! you have a "luster available you /ould s)e"i!y=

```
raxmlHPC-MPI -f d -m GTRCAt -p 12345 -s ex_al -# 100 -n MultipleOriginal
```

)re"eded by the res)e"tive M; run+time "ommands% e.g. `mpiexec` or `mpirun` de)ending on your lo"al installation 6)lease "he"k /ith your lo"al "om)uter s"ientist8.

Bootstrapping with RAxML

To "arry out a multi)le non+)arametri" bootstra) /ith the se4uential version o! RAxML Aust ty)e=

```
raxmlHPC -f d -m GTRCAT -s ex_al -p 12345 -# 100 -b 12345 -n MultipleBootstrap
```

<ou have to s)e"i!y a random number seed a!ter -b !or the random number generator o! the bootstra)s and -p !or the random number generator o! the)arsimony starting trees. This /ill allo/ you to generate re)rodu"ible results.

To do a)arallel bootstra) ty)e=

```
raxmlHPC-MPI -f d -m GTRCAT -s ex_al -# 100 -p 12345 -b 12345 -n MultipleBootstrap
```

on"e again)re"eded by the a))ro)riate M; exe"ution "ommand.

Obtaining Confidence Values

Su))ose that you have exe"uted 2$$ in!eren"es on the original alignment and &%$$$ bootstra) runs. <ou "an no/ use the RAxML -f b o)tion to dra/ the in!ormation !rom the &%$$$ bootstra))ed to)ologies onto some tree and obtain a to)ology /ith su))ort values. 2rom my)oint o! vie/ the most reasonable thing to do is to dra/ them on the best+s"oring ML tree !rom those 2$$ runs. Su))ose% that the best+s"oring tree /as !ound in run number QQ and the res)e"tive tree Ile is "alled `RAxML_result.MultipleOriginal.RUN.99`.
! you have exe"uted more than one bootstra) runs /ith the se4uential version o! RAxML on distin"t "om)uters% i.e. &$ runs /ith &$$ bootstra)s on &$ ma"hines you /ill Irst have to "on"atenate the boot+stra) Iles. ! your bootstra) result Iles are "alled e.g. `RAxML_bootstrap.MultipleBootstrap.0...RAxML_bootstrap.MultipleBootstrap.9` you "an

easily "on"atenate them by using the L 9 , X@ , 9 X `cat` "ommand% e.g.

```
cat RAxML_bootstrap.MultipleBootstrap.* > RAxML_bootstrap.All
```

n order to get a tree /ith bootstra) values on it Aust exe"ute RAxML as indi"ated belo/=

```
raxmlHPC -f b -m GTRCAT -z RaxML_bootstrap.All -t RAxML_result.MultipleOriginal.RUN.99 -n BS_TREE
```

This /ill return the tree)assed via `-t` annotated by su))ort values either as bran"h labels or as node labels. Some tree vie/ers have)roblems /ith dis)laying trees /ith node lables% in)arti"ular i! they are used to re+root the tree% hen"e you should use the bran"h+labeled tree% i!)ossible.

<ou "an no/ also "om)ute a maAority rule "onsensus tree out o! the bootstra) re)li"ates=

```
raxmlHPC -J MR -m GTRCAT  -z RaxML_bootstrap.All -n MR_CONS
```

an extended maAority rule "onsensus tree=

```
raxmlHPC -J MRE -m GTRCAT  -z RaxML_bootstrap.All -n MRE_CONS
```

or a stri"t one=

```
raxmlHPC -J STRICT -m GTRCAT  -z RaxML_bootstrap.All -n STRICT_CONS
```

2or Inding rogue taxa you "an use the `-J STRICT_DROP` or `-J MR_DROP` o)tions. Alternatively you "an use the /eb+server im)lemented by my lab= htt)=@@rnr.h+its.org@rnr

XII. A Simple Heterotachous Model

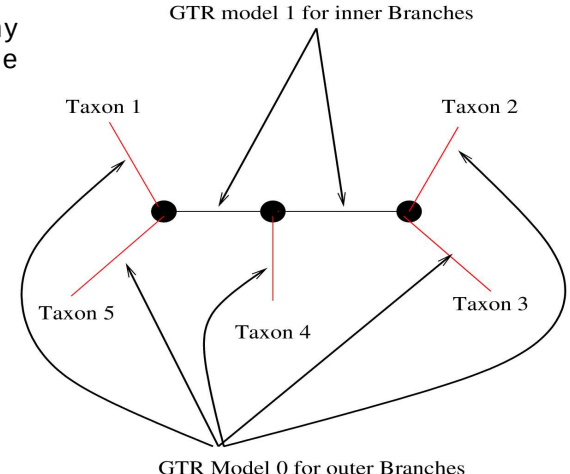
*e im)lemented a sim)le heterota"hous model in RAxML by assigning one (TR model to terminal bran"hes o! the tree and another distin"t (TR model to the inner bran"hes o! the tree. The idea /as to better "a)ture early ra)id radiations /ith this model% but it didnHt /ork /ell.
All other)arameters su"h as the al)ha sha)e)arameter o! the (amma model o! rate heterogeneity and the base !re4uen"ies are shared a"ross the entire tree. The model)arameters o! the t/o (TR models on the tree are o)timi7ed /ith res)e"t to the overall likelihood o! the tree using standard numeri"al o)timi7ation)ro"edures.

This heterota"hous model is "urrently only available in the se4uential version o! RAxML and only !or analy7ing . 9A se4uen"e data under the (amma model o! rate heterogeneityG

There is no "ommand line s/it"h !or enabling this model. nstead% you /ill need to re"om)ile RAxML by adding `-D_HET` to the line starting by `CFLAGS =` in the res)e"tive MakeIle. Then% /hen you s)e"i!y `-m GTRGAMMA` the sear"h and all other likelihood "al"ulations /ill be "ondu"ted under this heterota"hous model.

Zee) in mind that% /hen doing model tests et".% having 2 (TR matri"es)er)artition in"du"es a total o! &$!ree)arameters !or the substitution matrix instead o! P !or a single (TR matrix.

3ur sim)le heterota"hy model is outlined in the Igure belo/.



XIII. Frequently Asked Questions

Q: * hen)er!orming a bootstra) sear"h using a)artitioned model% does RAxML)er!orm a "onserved+ bootstra) re+sam)ling% i.e.% does it re+sam)le / ithin genes so that)artitions are sustainedC

A: That is the "ase. * hen)er!orming Rootstra)s on)artitioned data sets% bootstra))ed alignments / ill be sam)led !rom / ithin)artitions% i.e.% bootstra))ed)artitions are sustained and "ontain exa"tly the same number o! alignment "olumns as the original)artition.

Q: 0an use 9EX,S+style in)ut Iles !or analyses / ith RAxMLC

A: 9ot dire"tly% but my "olleague 2rank ZauM 6!kauM_rhrk.uni+kl.de8 at the ,niversity o! Zaiserslautern has / ritten a "ool bio)ython / ra))er "alled ;<RAXML2. This is a s"ri)t that reads nexus data Iles and)re)ares the ne"essary in)ut Iles and "ommand+line o)tions !or RAxML. <ou "an do / nload it at htt)=@@/ / /.lut7onilab.net@do / nloads@

Q: * hy don:t you like the)ro)ortion o! nvariable 6;+ nvar8 Sites estimate% des)ite the !a"t that you im)lemented itC

A: only im)lemented ;+ nvar in RAxML to make some users ha))y% but still strongly disagree / ith its usage.

Personal opinion: t is un4uestionable that one needs to in"or)orate rate heterogeneity in order to obtain *publishable* results. ;ut aside the *publish-or-perish* argument% there is also strong biologi"al eviden"e !or rate heterogeneity among sites. The rationale !or being ske)ti"al about ;+ nvar in RAxML is that all three alternatives% (TR(AMMA% (TR0AT% and ;+ nvar re)resent distin"t a))roa"hes to in"or)orate rate heterogeneity. Thus% in)rin"i)le they a""ount !or the same)henomenon by diMerent mathemati"al means. Also some un)ublished "on"erns have been raised that the usage o! ;+ nvar in "ombination / ith (amma "an lead to a *ping-pong* eMe"t sin"e a "hange o! ;+ nvar leads to a "hange in (amma and vi"e versa. This essentially means that those t / o)arameters% i.e.% al)ha and ;+ nvar "an not be o)timi7ed inde)endently !rom ea"h other% and might "ause signi I"ant trouble and)roblems during the model)arameter 6everything ex"e)t tree to)ology8 o)timi7ation)ro"ess. n !a"t% already observed this / hen / as im)lementing ;+ nvar in RAxML on a very small AA dataset.
Although this has never been)ro)erly do"umented% several / ell+kno / n resear"hers in)hylogeneti"s share this o)inion. Hll 4uote Jiheng <ang !rom an email in 2$$8 regarding this)art o! the RAxML manual=

I entirely agree with your criticism of the Pinv+Gamma model, even though as you said, it is very commonly used.

Jiheng also addresses the issue in his book on 0om)utational Mole"ular Evolution 63x+ !ord ,niversity ;ress% 2$$'8[4uote !rom)ages &&BW&&?=

The model is known as I+G and has been widely used. This model is somewhat pathological as the gamma distribution with alpha already allows for sites with very low rates; as a result, adding a proportion of invariable sites creates a strong correlation between p0 and alpha, making it impossible to estimate both parameters reliably.

n any "ase% have so !ar not en"ountered any di L "ulties /ith revie/s !or the !e/ real)hylogeneti" analyses)ublished /ith "olleagues !rom Riology% /hen /e used the (TR(AMMA model instead o! the more /idely s)read (TRU`U model.

Q: * hy does RAxML only im)lement (TR+based models o! nu"leotide substitutionC

A: 2or ea"h distin"t model o! nu"leotide substitution RAxML uses a se)arate% highly o)timi7ed set o! likelihood !un"tions. The idea behind this is that (TR is the most "ommon and general model !or real+/orld .9A analysis. Thus% it is better to e L "iently im)lement and o)timi7e this model instead o! oMering a)lethora o! distin"t models /hi"h are only s)e"ial "ases o! (TR but are)rogrammed in a generi" and thus ine L "ient /ay.

Personal opinion: My)ersonal vie/ is that using a sim)ler model than (TR only makes sense /ith res)e"t to the "om)utational "ost% i.e. it is less ex)ensive to "om)ute. ;rograms su"h as `Modeltest`)ro)ose the usage o! a sim)ler model !or a s)e"i l" alignment i! the likelihood o! a I xed to)ology under that sim)ler model is not signi l"antly /orse than that obtained by (TR based on a likelihood ratio test. My ex)erien"e is that (TR al/ays yields a slightly better likelihood than alternative sim)ler models. n addition% sin"e RAxML has been designed !or the in!eren"e o! large datasets the danger o! over+)arameteri7ing su"h an analysis is "om)aratively lo/. ;rovided these arguments the design de"ision /as taken to rather im)lement the most general model e L "iently than to)rovide many ine L "ient generi" im)lementations o! models that are Aust s)e"ial "ases o! (TR.
2inally% the design)hiloso)hy o! RAxML is based u)on the observation that a more thorough to)ologi"al sear"h has a greater im)a"t on I nal tree 4uality than modeling details.
Thus% the e L "ient im)lementation o! a ra)id sear"h me"hanisms is "onsidered to be more im)ortant than model details.

Q: * hy has the)er!orman"e o! RAxML mainly been assessed using real+/orld dataC

A: *Personal opinion*= .es)ite the un4uestionable need !or simulated data and trees to veri!y and test the)er!orman"e o! "urrent ML algorithms the "urrent methods available !or generation o! simulated alignments are not very realisti".
2or exam)le% only !e/ methods exist that in"or)orate the generation o! ga)s in simulated alignments. Sin"e the model a""ording to /hi"h the se4uen"es are generated on the true tree is)re+de I ned /e are a"tually assuming that ML exa"tly models the true evolutionary)ro"ess% /hile in reality /e sim)ly don`t kno/ ho/ se4uen"es evolved.
The above sim)li I "ations lead to *perfect* alignment data /ithout ga)s% that evolved exa"tly a""ording to a)re+de I ned model and thus exhibits a very strong)hylogeneti" signal in "ontrast to real data.
n addition% the given true tree% must not ne"essarily be the Maximum Likelihood tree. This diMeren"e mani!ests itsel! in substantially diMerent behaviors o! sear"h algorithms on real and simulated data. Ty)i"ally% sear"h algorithms exe"ute signi l"antly less 6!a"tor PW &$8 to)ologi"al moves on simulated data until "onvergen"e as o))osed to real data% i.e. the number o! su""ess!ul 9earest 9eighbor nter"hanges 699 s8 or subtree rearrangements is lo/er. Moreover% in several "ases the likelihood o! trees !ound by RAxML on simulated data /as better than that o! the true tree.
Another im)ortant observation is that)rogram)er!orman"e "an be inverted by simulated data. Thus% a)rogram that yields *good* to)ologi"al RobinsonW2ould distan"es on simulated data "an in !a"t)er!orm mu"h /orse on real data than a)rogram that does not)er!orm

/ell on simulated data.
! one is /illing to really a""e)t ML as in!eren"e "riterion on real data one must also be /illing to assume that the tree /ith the best likelihood s"ore is the tree that is "losest to the true tree.
My)ersonal "on"lusion is that there is a strong need to im)rove simulated data generation and methodology. n addition% the)erha)s best /ay to assess the validity o! our tree in!eren"e methods "onsists in an em)iri"al evaluation o! ne/ results and insights obtained by real)hylogeneti" analysis.
This should be based on the)rior kno/ledge o! Riologists about the data and the medi"al and s"ientiI" beneIts attained by the "om)utation o!)hylogenies.

Q: *hy am getting /eird error messages !rom the M; versionC
A: <ou)robably !orgot to s)e"i!y the –# or –N o)tion in the "ommand+line /hi"h must be used !or the M; version to /ork)ro)erly.

Q: *hen using)artitioned models% "an link the model)arameters o! distin"t)artitions to be estimated Aointly% in a similar as /ay MrRayes does itC
A: 0urrently not% but the im)lementation o! su"h an o)tion is)lanned. Ho/ever% /e are still la"king good "riteria and methods that tell us ho/ and /hy to link@unlink "ertain)arameters a"ross "ertain)artitions.

Q: Ho/ does RAxML handle ga)s
A: Like most other likelihood+based)hylogeny)rograms it handles them as undetermined "hara"ters% that is ga)s are treated as Ns. ! this is ne/ to you you may /ant to read #oe 2elsensteinHs textbook *Inferring Phylogenies* /here he ex)lains /hy this is a reasonable /ay to model ga)s.

Q: *hy am getting long bran"h lengths on my)hylogenomi" datasets /ith missing data
A: This is normal and due to the /ay missing data is modeled% the missing data)arts o! the se4uen"es about /hi"h /e donHt kno/ be"ome very distant !or exa"tly this reason.