

How to read xmls

Nicola F. Müller

e-mail: Nicola.mueller@ucsf.edu

Why learn about BEAST xmls?

- It is easier to change small parts of an analysis in the xml directly (change chain length or a prior)
- Some of the more complex analyses are only possible in the xml directly (e.g. linking parameters in different models).
- You can check someone else's analysis (i.e. see what was actually done)

The xml contains all the data

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?><beast beautitemplate='Standard' beautistatus='' namespace="beast.core:beast.evolution.alignment:beast.evolution.tree.coalesc

  <data
id="h1n1pdm_HA"
spec="Alignment"
name="alignment">

  <sequence id="seq_A/Alaska/07/2017|2017-02-13" spec="Sequence" taxon="A/Alaska/07/2017|2017-02-13" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTACAACCGCAAATGC
  <sequence id="seq_A/Arizona/32/2015|2015-12-29" spec="Sequence" taxon="A/Arizona/32/2015|2015-12-29" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACCGCAAAT
  <sequence id="seq_A/Arizona/33/2017|2017-05-07" spec="Sequence" taxon="A/Arizona/33/2017|2017-05-07" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACCGCAAAT
  <sequence id="seq_A/Bangkok/INS3_681/2012|2012-09-21" spec="Sequence" taxon="A/Bangkok/INS3_681/2012|2012-09-21" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTT
  <sequence id="seq_A/California/19/2017|2017-01-29" spec="Sequence" taxon="A/California/19/2017|2017-01-29" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACC
  <sequence id="seq_A/California/45/2016|2016-02-20" spec="Sequence" taxon="A/California/45/2016|2016-02-20" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACC
  <sequence id="seq_A/Durham/INS3_648/2012|2012-03-08" spec="Sequence" taxon="A/Durham/INS3_648/2012|2012-03-08" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTGCG
  <sequence id="seq_A/Finland/75/2014|2014-02-10" spec="Sequence" taxon="A/Finland/75/2014|2014-02-10" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTGCAACCGCAAAT
  <sequence id="seq_A/Finland/87/2014|2014-02-14" spec="Sequence" taxon="A/Finland/87/2014|2014-02-14" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTGCAACCGCAAAT
  <sequence id="seq_A/Florida/100/2015|2015-12-20" spec="Sequence" taxon="A/Florida/100/2015|2015-12-20" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACCGCAA#
  <sequence id="seq_A/Hawaii/67/2014|2014-10-12" spec="Sequence" taxon="A/Hawaii/67/2014|2014-10-12" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATACACATTTGCAACCGCAAATGC
  <sequence id="seq_A/Helsinki/473N/2014|2014-02-10" spec="Sequence" taxon="A/Helsinki/473N/2014|2014-02-10" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTGCAACCG
  <sequence id="seq_A/Illinois/26/2017|2017-03-16" spec="Sequence" taxon="A/Illinois/26/2017|2017-03-16" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACRCRA#
  <sequence id="seq_A/Kansas/14/2016|2016-03-24" spec="Sequence" taxon="A/Kansas/14/2016|2016-03-24" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACCGCAAATGC
  <sequence id="seq_A/Nepal/VIROAF5/2012|2012-08-26" spec="Sequence" taxon="A/Nepal/VIROAF5/2012|2012-08-26" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTGCAACCG
  <sequence id="seq_A/New_Mexico/19/2016|2016-02-28" spec="Sequence" taxon="A/New_Mexico/19/2016|2016-02-28" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACCG
  <sequence id="seq_A/New_York/WC_LVD_14_021/2014|2014-01-30" spec="Sequence" taxon="A/New_York/WC_LVD_14_021/2014|2014-01-30" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCG
  <sequence id="seq_A/New_York/WC_LVD_14_063/2014|2014-02-09" spec="Sequence" taxon="A/New_York/WC_LVD_14_063/2014|2014-02-09" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCG
  <sequence id="seq_A/Nicaragua/6322_06/2015|2015-12-08" spec="Sequence" taxon="A/Nicaragua/6322_06/2015|2015-12-08" totalcount="4" value="ATGAAGGCAATACTAGCAGTTCTGCTATATACAT
  <sequence id="seq_A/Nizhnii_Novgorod/CRIE_BLM/2011|2011-01-26" spec="Sequence" taxon="A/Nizhnii_Novgorod/CRIE_BLM/2011|2011-01-26" totalcount="4" value="ATGAAGGCAATACTAGT#
  <sequence id="seq_A/North_Dakota/15/2017|2017-03-19" spec="Sequence" taxon="A/North_Dakota/15/2017|2017-03-19" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTAC
  <sequence id="seq_A/Seoul/224/2016|2016-02-03" spec="Sequence" taxon="A/Seoul/224/2016|2016-02-03" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACCGCAAATGC
  <sequence id="seq_A/South_Dakota/13/2017|2017-02-18" spec="Sequence" taxon="A/South_Dakota/13/2017|2017-02-18" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTAC
  <sequence id="seq_A/Utah/35/2016|2016-04-22" spec="Sequence" taxon="A/Utah/35/2016|2016-04-22" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACCGCAAATGCAGAC
  <sequence id="seq_A/Washington/01/2017|2017-01-06" spec="Sequence" taxon="A/Washington/01/2017|2017-01-06" totalcount="4" value="ATGAAGGCAATACTAGTAGTTCGCTATATACATTTACAACC

</data>

<map name="Uniform" >beast.base.inference.distribution.Uniform</map>

<map name="Exponential" >beast.base.inference.distribution.Exponential</map>

<map name="LogNormal" >beast.base.inference.distribution.LogNormalDistributionModel</map>

<map name="Normal" >beast.base.inference.distribution.Normal</map>

<map name="Beta" >beast.base.inference.distribution.Beta</map>

<map name="Gamma" >beast.base.inference.distribution.Gamma</map>

<map name="LaplaceDistribution" >beast.base.inference.distribution.LaplaceDistribution</map>
```

The settings and specifications of the MCMC

```
<run id="mcmc" spec="MCMC" chainLength="10000000">
  <state id="state" spec="State" storeEvery="5000">
    <tree id="Tree.t:h1n1pdm_HA" spec="beast.base.evolution.tree.Tree" name="stateNode">
      <taxonset id="TaxonSet.h1n1pdm_HA" spec="TaxonSet">
        <alignment idref="h1n1pdm_HA"/>
      </taxonset>
    </tree>
    <parameter id="birthRate.t:h1n1pdm_HA" spec="parameter.RealParameter" lower="0.0" name="stateNode">1.0</parameter>
  </state>
  <init id="RandomTree.t:h1n1pdm_HA" spec="RandomTree" estimate="false" initial="@Tree.t:h1n1pdm_HA" taxa="@h1n1pdm_HA">
    <populationModel id="ConstantPopulation0.t:h1n1pdm_HA" spec="ConstantPopulation">
      <parameter id="randomPopSize.t:h1n1pdm_HA" spec="parameter.RealParameter" name="popSize">1.0</parameter>
    </populationModel>
  </init>
  <distribution id="posterior" spec="CompoundDistribution">
    <distribution id="prior" spec="CompoundDistribution">
      <distribution id="YuleModel.t:h1n1pdm_HA" spec="beast.base.evolution.speciation.YuleModel" birthDiffRate="@birthRate.t:h1n1pdm_HA" tree="@Tree.t:h1n1pdm_HA"/>
      <prior id="YuleBirthRatePrior.t:h1n1pdm_HA" name="distribution" x="@birthRate.t:h1n1pdm_HA">
        <Uniform id="Uniform.1" name="distr" upper="Infinity"/>
      </prior>
    </distribution>
    <distribution id="likelihood" spec="CompoundDistribution" useThreads="true">
      <distribution id="treeLikelihood.h1n1pdm_HA" spec="ThreadedTreeLikelihood" data="@h1n1pdm_HA" tree="@Tree.t:h1n1pdm_HA">
        <siteModel id="SiteModel.s:h1n1pdm_HA" spec="SiteModel">
          <parameter id="mutationRate.s:h1n1pdm_HA" spec="parameter.RealParameter" estimate="false" lower="0.0" name="mutationRate">1.0</parameter>
          <parameter id="gammaShape.s:h1n1pdm_HA" spec="parameter.RealParameter" estimate="false" lower="0.1" name="shape">1.0</parameter>
          <parameter id="proportionInvariant.s:h1n1pdm_HA" spec="parameter.RealParameter" estimate="false" lower="0.0" name="proportionInvariant" upper="1.0">0.0</parameter>
          <substModel id="JC69.s:h1n1pdm_HA" spec="JukesCantor"/>
        </siteModel>
        <branchRateModel id="StrictClock.c:h1n1pdm_HA" spec="beast.base.evolution.branchratemodel.StrictClockModel">
          <parameter id="clockRate.c:h1n1pdm_HA" spec="parameter.RealParameter" estimate="false" lower="0.0" name="clock.rate">1.0</parameter>
        </branchRateModel>
      </distribution>
    </distribution>
  </distribution>
</distribution>
```

And what is logged

```
<operator id="YuleBirthRateScaler.t:h1n1pdm_HA" spec="kernel.BactrianScaleOperator" parameter="@birthRate.t:h1n1pdm_HA" upper="10.0" weight="3.0"/>
<operator id="YuleModelTreeRootScaler.t:h1n1pdm_HA" spec="kernel.BactrianScaleOperator" rootOnly="true" scaleFactor="0.5" tree="@Tree.t:h1n1pdm_HA" upper="10.0" weight="3.0"/>
<operator id="YuleModelUniformOperator.t:h1n1pdm_HA" spec="kernel.BactrianNodeOperator" tree="@Tree.t:h1n1pdm_HA" weight="30.0"/>
<operator id="YuleModelSubtreeSlide.t:h1n1pdm_HA" spec="kernel.BactrianSubtreeSlide" tree="@Tree.t:h1n1pdm_HA" weight="15.0"/>
<operator id="YuleModelNarrow.t:h1n1pdm_HA" spec="Exchange" tree="@Tree.t:h1n1pdm_HA" weight="15.0"/>
<operator id="YuleModelWide.t:h1n1pdm_HA" spec="Exchange" isNarrow="false" tree="@Tree.t:h1n1pdm_HA" weight="3.0"/>
<operator id="YuleModelWilsonBalding.t:h1n1pdm_HA" spec="WilsonBalding" tree="@Tree.t:h1n1pdm_HA" weight="3.0"/>
<operator id="YuleModelBICEPSEpochTop.t:h1n1pdm_HA" spec="EpochFlexOperator" scaleFactor="0.1" tree="@Tree.t:h1n1pdm_HA" weight="2.0"/>
<operator id="YuleModelBICEPSEpochAll.t:h1n1pdm_HA" spec="EpochFlexOperator" fromOldestTipOnly="false" scaleFactor="0.1" tree="@Tree.t:h1n1pdm_HA" weight="2.0"/>
<operator id="YuleModelBICEPSTreeFlex.t:h1n1pdm_HA" spec="TreeStretchOperator" scaleFactor="0.01" tree="@Tree.t:h1n1pdm_HA" weight="2.0"/>
<logger id="tracelog" spec="Logger" fileName="$(filebase).log" logEvery="1000" model="@posterior" sanitiseHeaders="true" sort="smart">
  <log idref="posterior"/>
  <log idref="likelihood"/>
  <log idref="prior"/>
  <log idref="treeLikelihood.h1n1pdm_HA"/>
  <log id="TreeHeight.t:h1n1pdm_HA" spec="beast.base.evolution.tree.TreeStatLogger" tree="@Tree.t:h1n1pdm_HA"/>
  <log idref="YuleModel.t:h1n1pdm_HA"/>
  <log idref="birthRate.t:h1n1pdm_HA"/>
</logger>
<logger id="screenlog" spec="Logger" logEvery="1000">
  <log idref="posterior"/>
  <log idref="likelihood"/>
  <log idref="prior"/>
</logger>
<logger id="treelog.t:h1n1pdm_HA" spec="Logger" fileName="$(filebase)-$(tree).trees" logEvery="1000" mode="tree">
  <log id="TreeWithMetaDataSetter.t:h1n1pdm_HA" spec="beast.base.evolution.TreeWithMetaDataSetter" tree="@Tree.t:h1n1pdm_HA"/>
</logger>
<operatorschedule id="OperatorSchedule" spec="OperatorSchedule"/>
</run>
</beast>
```

Organization

- Data -> Alignments
- Run -> MCMC
 - State -> Initial states + definitions of what can change
 - Initialization -> Tree initialization
 - Posterior -> Everything that goes into the posterior
 - Prior
 - Tree Prior
 - Parameter Priors
 - Likelihood
 - Operators
 - Loggers

Xml is a hierarchical structure

```
<distribution id="CoalescentConstant.t:h1n1pdm_HA" spec="Coalescent">  
  <populationModel id="ConstantPopulation.t:h1n1pdm_HA" spec="ConstantPopulation"  
    popSize="@popSize.t:h1n1pdm_HA"/>  
  <treeIntervals id="TreeIntervals.t:h1n1pdm_HA" spec="beast.base.evolution.tree.TreeIntervals"  
    tree="@Tree.t:h1n1pdm_HA"/>  
</distribution>
```