Outline

- Brief Primer
  - Current state of knowledge
- Main Question
  - Motivation
- Methods
  - Multiple sequence alignment
  - Tree construction using maximum likelihood, and bayesian analysis
  - Supertree construction
- Results
- Conclusions

What do we know?

- Work in this field by Crespi and Fulton in 2003 had gathered up data to infer the phylogenies of 30 species of salmonids
- This work posited certain relationships between organisms in the Salmonidae family
  - *Specifically, it claimed that Oncorhynchus masou* forms a monophyletic group with *Oncorhynchus mykiss* and *Oncorhynchus clarki*
  - This suggests that *Oncorhynchus masou* is a **trout**
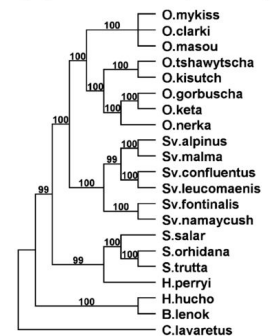

O. clarki
(Cutthroat Trout)

O. masou
(Cherry Trout?)

O. mykiss
(Rainbow Trout)

(B) BAY: nDNA (ex. MHC)

Is what we know correct?

- Do we really know that these relationships are real?
- A more recent paper by Crete-Lafreniere et. al. in 2012 which, incorporated a larger data set and more species within the family Salmonidae, challenged the old belief that *O. masou* is actually a trout
    - Instead the paper posited that *Oncorhynchus* is actually a paraphyletic group and placed *Oncorhynchus masou* along with other pacific **salmon**
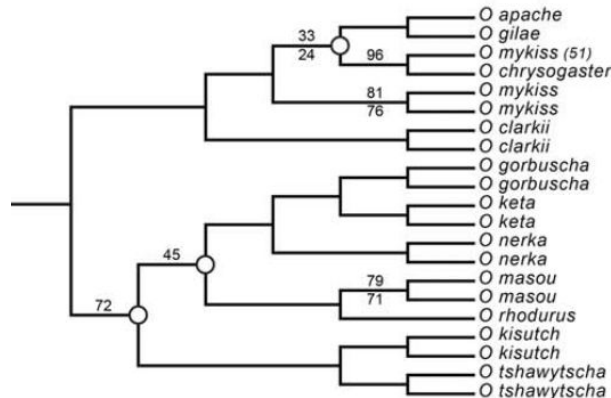
O. masou
(Cherry Salmon?)

"Is Oncorhyncus masou
a trout or Salmon?"

Why do we care?

- Salmonids are important to both humans and to both terrestrial and marine ecosystems
- Salmonid fishes are becoming increasingly important as model systems for answering many evolutionary and ecological questions
- A robust phylogeny of this family is important for multiple fields including
  - Adaptation analysis
  - Comparative genomics
  - Analysis of ancestral states
  - Evaluation of conservation priorities

Comparison of Methodologies

Previous Methods

- Multiple sequence alignments were done with ClustalW
- Tree construction was accomplished using:
  - PAUP (**P**hylogenetic **A**nalysis **U**sing **P**arsimony) for maximum parsimony
  - RAxML (**R**andomized **Ax**elerated **M**aximum **L**ikelihood) for maximum likelihood
 - MrBayes for bayesian

Current Methods

- Multiple sequence alignments were attempted with MAFFT (**M**ultiple **A**lignment using **F**ast **F**ourier **T**ransform) and MUSCLE (**Mu**ltiple **S**equence **C**omparison by **L**og- **E**xpectation)
  - Sequence alignments with MUSCLE were used for tree construction as they yielded higher log-likelihood values
- Tree construction was accomplished using:
  - RAxML as well as MEGA7 for maximum likelihood
  - BEAST for bayesian
- Supertrees were constructed using ASTRAL (**A**ccurate **S**pecies **T**ree **AL**gorithm)

Multiple Sequence Alignment

- MSA's were attempted with MAFFT and MUSCLE
- MUSCLE Parameters:
  - First Run:
    - Gap Opening: -400
    - Gap Extension: 0
  - Second Run:
   - Gap Opening: -15
    - Gap Extension: -7

Genes	Current	Muscle	MAFFT	MUSCLE	MUSCLE (W/O 3rd)
1-1141	-13463.83	-13463.83	-13461.24	-11169.6	-7564.36
1-11413	-71407.52	-71631.24	-71464.65	-68347.48	-40151.49
1-15597	-83466.01	-85263.61	-84944.74		
1142-2692	-13847.75	-13953.48	-13918.05	-12210.38	-7504.05
2693-3378	-2292.06	-2292.06	-2292.06	-2119.78	-1482.83
3379-4164	-3093.56	-3093.56	-3095.44	-2774.02	-1819.22
4165-4848	-5302.98	-5302.98	-5302.98	-4465.47	-2894.82
4849-5016	-407.4	-407.4	-407.4	-375.86	-242.9
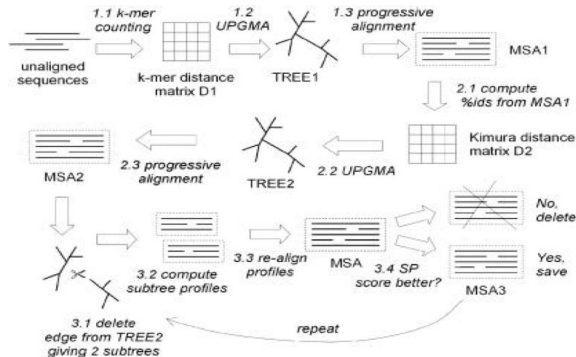
Tree Construction

- We built phylogenetic trees for:
  - Each individual gene
  - The full set of nuclear genes
  - The full set of mitochondrial genes
  - For all the genes combined
- We did this for both for the original alignments that were already done for us and the MUSCLE alignments which we did ourselves
- RAxML was run with the following parameters:
  - Bootstrap Value: 1000
  - Substitution Model: GTRGAMMA
  - No outgroups were used
    - Tree produced was not rooted
- BEAST was run with the following parameters:
  - Molecular Clock: Relaxed Clock Log Normal
  - Substitution Model: GTR
  - Prior Distribution: YULE Model
    - Birth rate: Uniform
  - No outgroups were used

RAxML Schematic
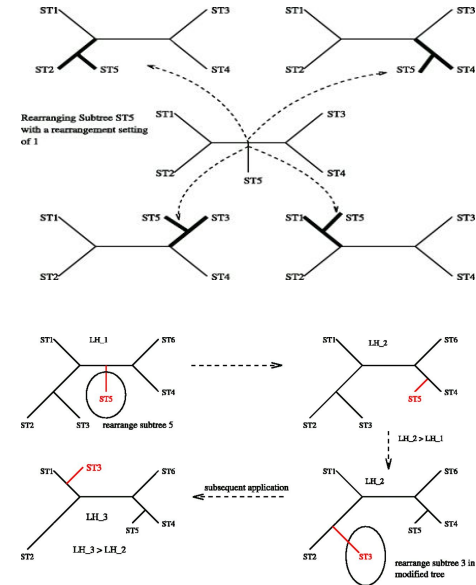
Supertree Construction

- We constructed supertrees using ASTRAL
- ASTRAL constructs the supertrees using the inputted gene trees, maximizing the Quartet Score (QS)
- QS is computed for a clade structure as ratio of the gene trees containing that particular clade to the total number of gene trees
- These supertrees are used as an effective species trees needed to perform introgression analysis
- We are in the process of performing reconciliation and introgression analysis.
  - We use the supertrees as our reference and compare the gene trees to the respective species tree
 - Also compare the tree created from all available data to the reference supertrees

Supertrees constructed using gene trees from maximum likelihood methods

QS scores for supertrees constructed using output from MEGA and raxML

Normalized quartet scores (in %)	MEGA	MEGA (without 3rd codon)	raxML
Nuclear genes	89.15	86.24	86.19
Mitochondrial genes	94.12	92.32	94.10
Entire set of genes	93.99	92.18	93.99
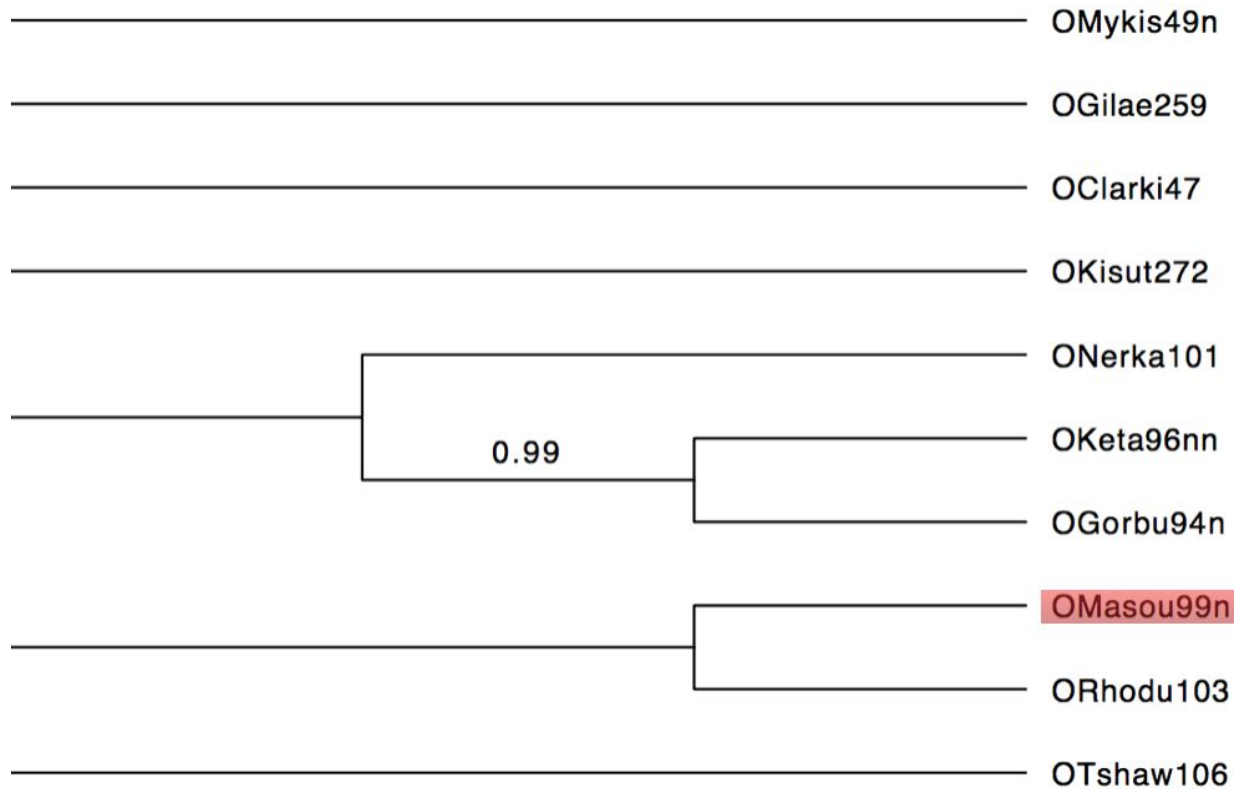
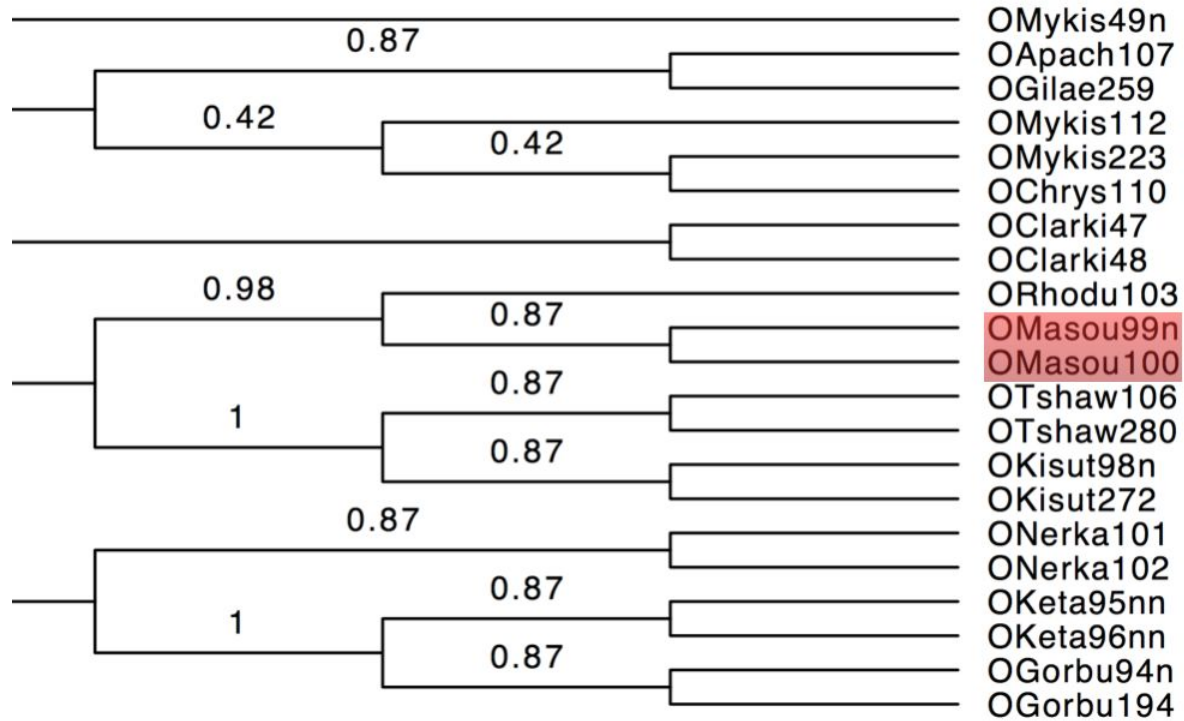Phylogenetic Trees - Full Nuclear Data Set (3rd codon removed)
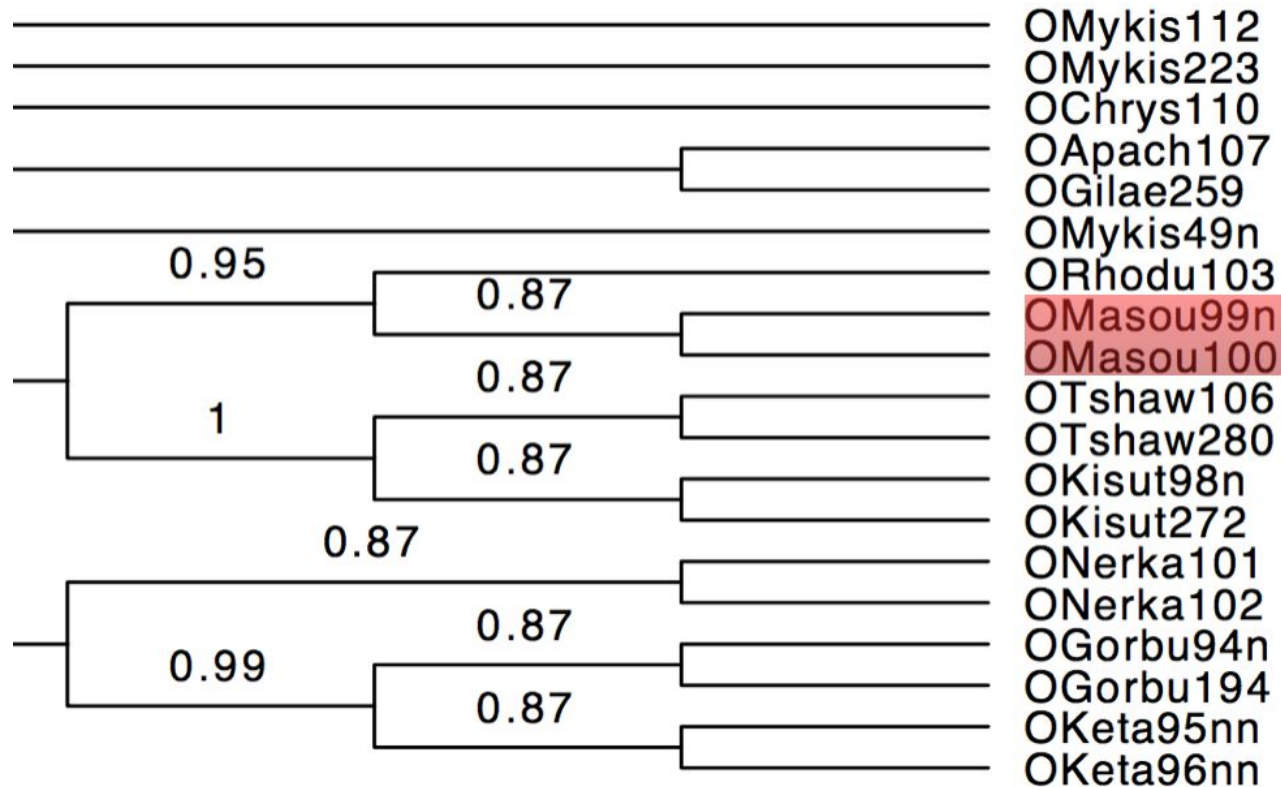
Phylogenetic Trees - Full Nuclear Data Set (3rd codon removed)

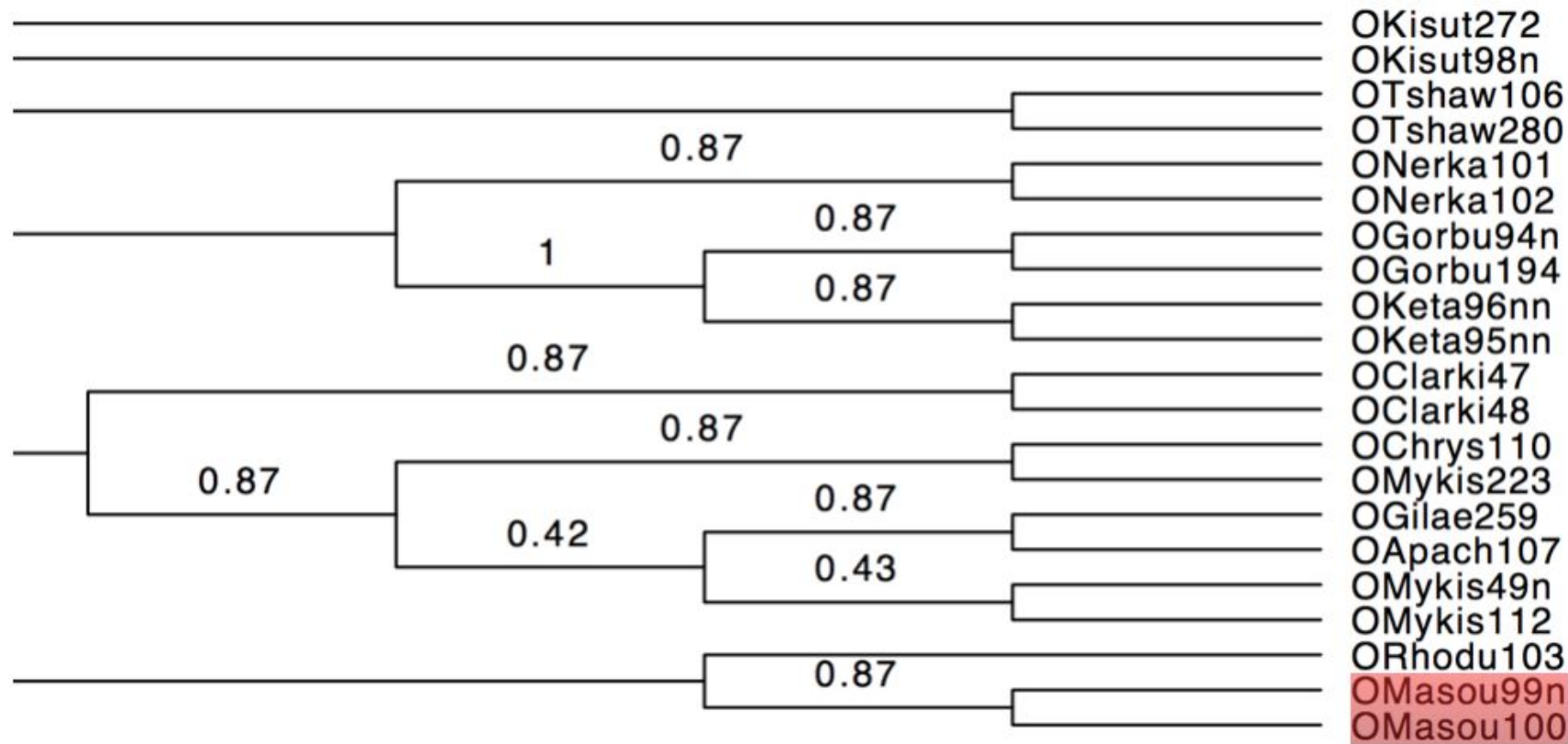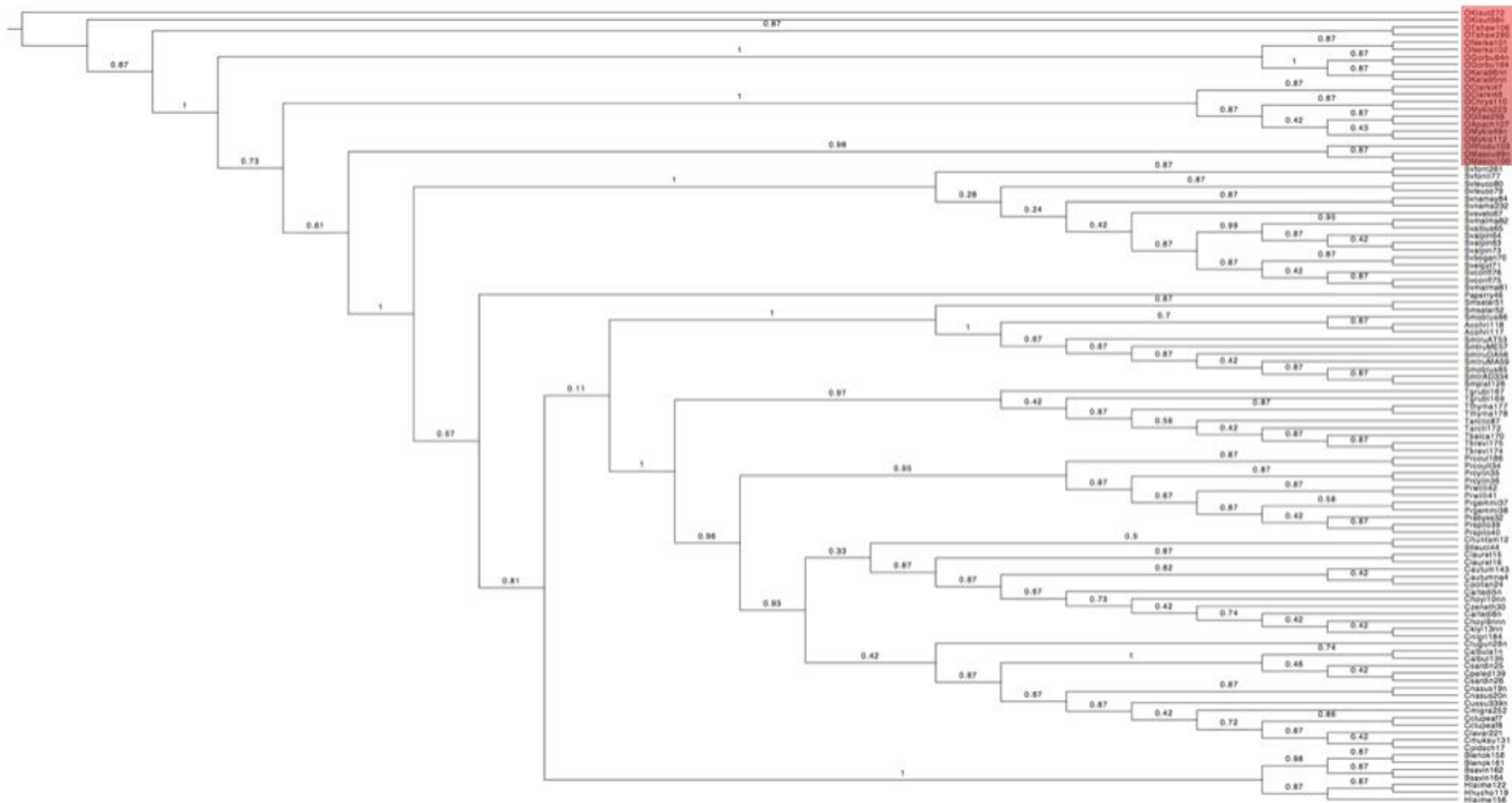Phylogenetic Trees - All mitochondrial gene data (3rd codon removed)

Phylogenetic Trees - All mitochondrial gene data (3rd codon removed)

Phylogenetic Trees - All data (3rd codon removed)

Phylogenetic Trees - All data (3rd codon removed)

Phylogenetic Trees - All Data

Phylogenetic Trees - All Data

Conclusions

- We analyzed a given data set of a large number of species within the family Salmonidae in order to establish relationships between the species
- We utilized new, improved, tools/software that were previously unavailable in our analysis to test whether the conclusions from previous studies were correct
- The analysis that we have conducted finds evidence that, in contrast to older findings, and in agreement with more recent results, Oncorhyncus masou is indeed a salmon and not a trout.