

Отчёт по практической работе «Реконструкция филогении»

Пуныко Анастасия

Введение

Для реконструкции филогении была выбрана статья Chen et al (2022) «Characterization of the complete mitochondrial genome and phylogenetic analysis of *Opsarius pulchellus* (Cypriniformes, Danionidae, Chedrinae)». Целью исследования Chen et al. было секвенирование и аннотирование полной последовательности ДНК митохондриального генома *Opsarius pulchellus*, а также проведение филогенетического анализа.

Opsarius pulchellus принадлежит к подсемейству Chedrinae (семейство Danionidae). Деградация среды обитания, вызванная вырубкой лесов и сельскохозяйственным загрязнением, является основной потенциальной угрозой для *O. pulchellus*. Авторы исследования впервые сообщают о секвенировании полного митохондриального генома *O. pulchellus*, который будет полезен для будущих исследований в области популяционной генетики и эволюции подсемейства Chedrinae.

Для филогенетического анализа авторы использовали 13 аминокислотных последовательностей белок-кодирующих генов (ND1, ND2, COX1, COX2, ATP8, ATP6, COX3, ND3, ND4L, ND4, ND5, ND6, CYTB) *O. pulchellus* (MZ853157.1) и еще 16 рыб из подсемейств Chedrinae и Danioninae. В качестве внешней группы использовались *Amblypharyngodon mola* и *Rasbora tornieri* из семейства Cyprinidae.

Филогенетические отношения были реконструированы с использованием метода наибольшего правдоподобия (ML) с 1000 бутстреп-репликами. Эволюционный анализ проводился в MEGA X [1], а в качестве лучшей модели использовалась General Reversible Mitochondrial (mtREV) и гамма распределение с инвариантными сайтами (G+I) [2].

Результат филогенетического анализа показал, что вид *O. pulchellus* близок к *Barilius bendelisis* (Рисунок 1).

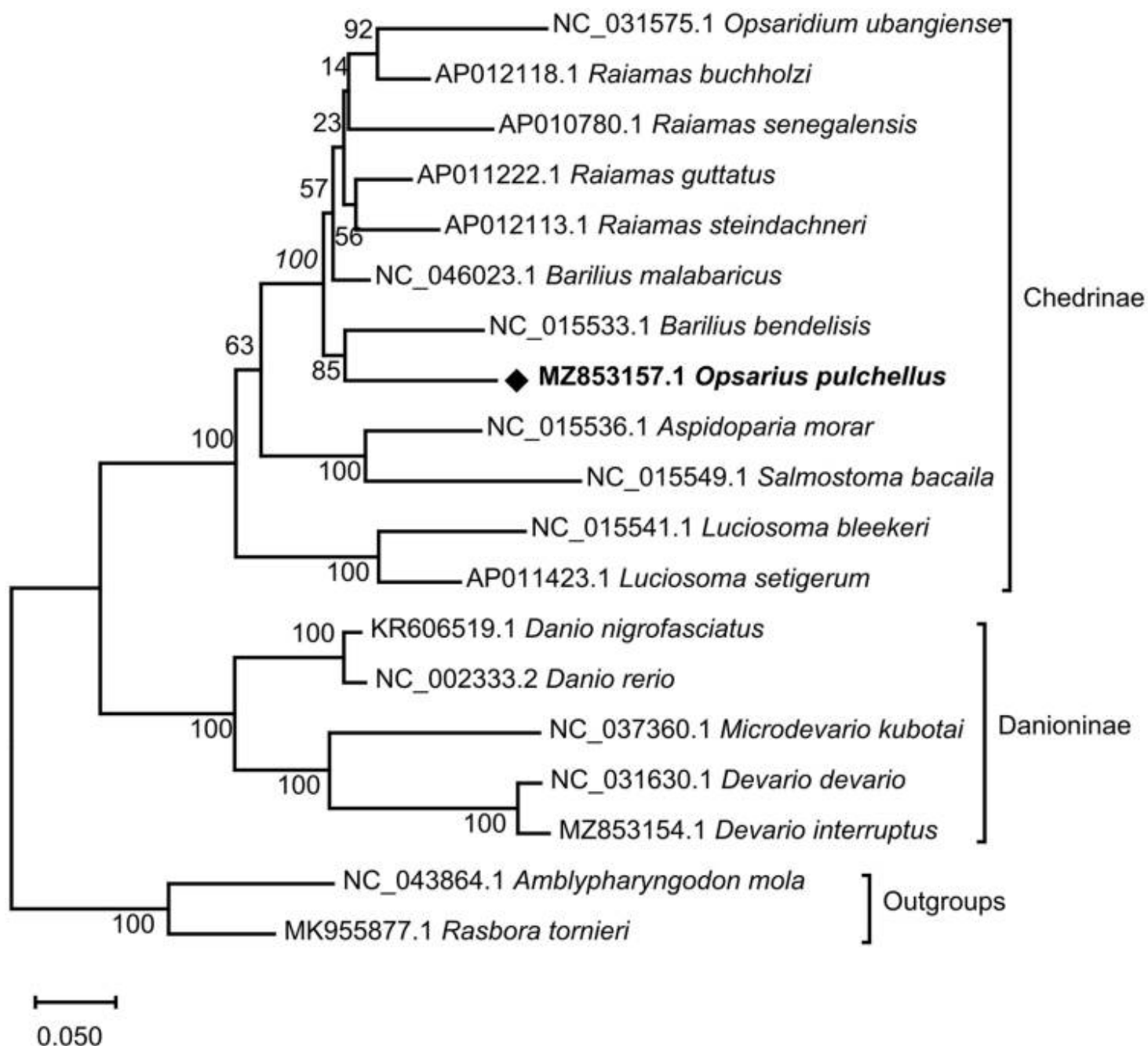


Рисунок 1. Филогенетическое дерево реконструированное с использованием 13 аминокислотных последовательностей белок-кодирующих генов *O. pulchellus* вместе с другими 18 рыбами.

Материалы и методы

13 аминокислотных последовательностей белок-кодирующих генов (ND1, ND2, COX1, COX2, ATP8, ATP6, COX3, ND3, ND4L, ND4, ND5, ND6, CYTB) 19-ти рыб были взяты из GenBank [3] по их идентификаторам согласно рисунку 1. В прикрепленном файле №1 в формате fasta находятся последовательности, используемые в исследовании.

Множественное выравнивание осуществлялось с помощью алгоритма MUSCLE [4]. Для дальнейшего переформатирования выравнивания в формат “phylip-relaxed” использовался BioPython (модуль AlignIO) [5]. В прикрепленном файле №2 находится выравнивание, используемое для реконструкции филогении. Для построения филогенетического дерева в MEGA, выравнивание с помощью алгоритма MUSCLE осуществлялось также в MEGA.

Для воспроизведения филогенетического дерева из статьи Chen et al. использовался софт MEGA, метод наибольшего правдоподобия (ML) с 1000 бутстреп-репликами, модель General Reversible Mitochondrial (mtREV) и гамма распределение с инвариантными сайтами (G+I).

Для дальнейшего построения филогенетических деревьев использовалась программа FastME [6], метод минимальной эволюции (ME), реализация двух вариантов: OLS и BME, бутстреп 100, модели MtREV и JTT с гамма распределением. Визуализация осуществлялась в iTOL [7].

Результаты

Были построены деревья в MEGA и с помощью программы FastME. В MEGA с настройками аналогичными в статье, а в FastME в трех вариантах:

- 1- модель mtREV, алгоритм BME
- 2- модель mtREV, алгоритм OLS
- 3- модель JTT, алгоритм OLS

В построенных деревьях от дерева в статье отличались только ветви с низкой ‘поддержкой’.

Построенное дерево в MEGA и в FastME (модель mtREV, алгоритм BME) не отличались, также не отличались между собой деревья в FastME с моделью mtREV + алгоритм OLS и с моделью JTT + алгоритм OLS.

Таким образом на реконструкции филогении повлиял выбор варианта минимальной эволюции: OLS или BME. А также, вероятно, параметры, которые не были обозначены в статье, вследствие чего не получилось воспроизвести дерево из статьи.

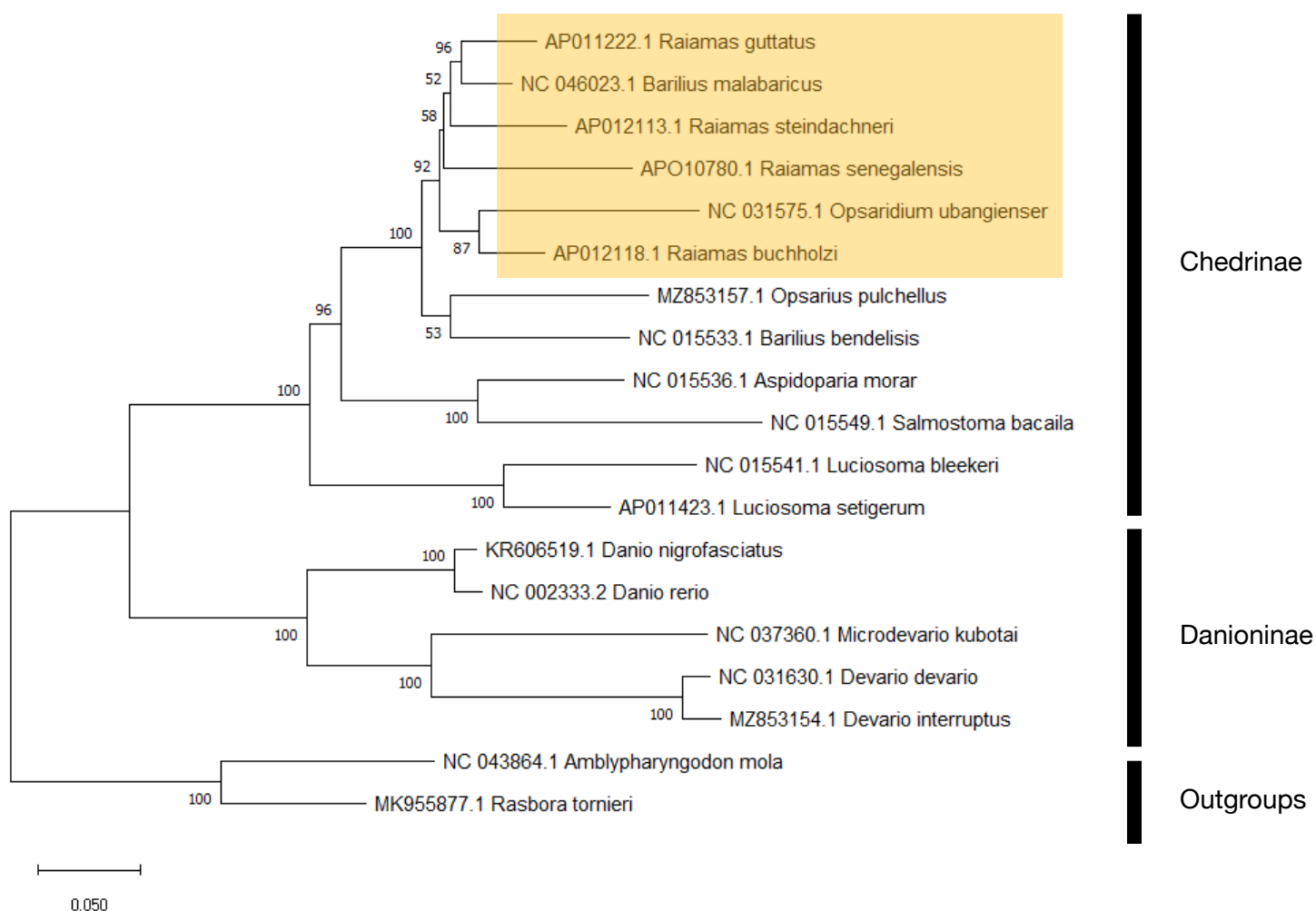


Рисунок 2. Филогенетическое дерево из MEGA (метод ML, бутстреп 1000, модель mtREV и гамма распределение с инвариантными сайтами (G+I)).

На рисунке 2 изображено филогенетическое дерево, построенное в MEGA. Метод наибольшего правдоподобия, бутстреп 1000, модель mtREV и гамма распределение с инвариантными сайтами (G+I). Укоренение с помощью внешней группы. На ветвях отображены значения бутстрепа.

От дерева в статье отличаются только ветви с низкой 'поддержкой'.

Вот так часть дерева с низкой поддержкой выглядела в статье:
 (((*Opsaridium ubangienser*, *Raiamas buchholzi*), *Raiamas senegalensis*),
 (*Raiamas guttatus*, *Raiamas steindachneri*)), *Barilius malabaricus*);

А так построенная в MEGA:

(((*Raiamas guttatus*, *Barilius malabaricus*), *Raiamas steindachneri*),
Raiamas senegalensis), (*Opsaridium ubangienser*, *Raiamas buchholzi*));

Tree scale: 0.1

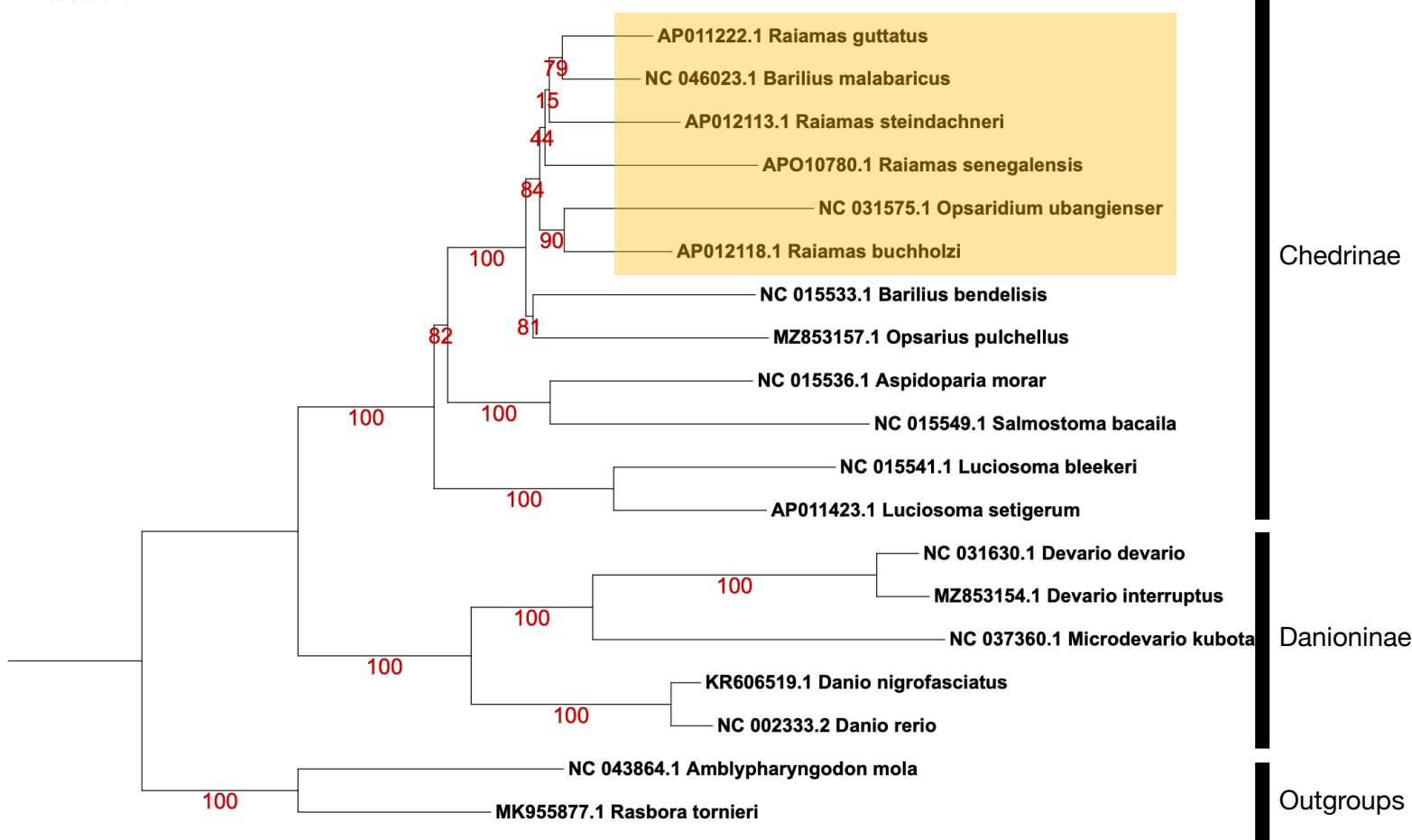


Рисунок 3. Филогенетическое дерево из FastME (метод ME, BME, бутстреп 100, модель mtREV с гамма распределением).

На рисунке 3 изображено филогенетическое дерево, построенное в FastME. Метод минимальной эволюции (ME), вариант BME, бутстреп 100, модели MtREV с гамма распределением. Укоренение с помощью внешней группы. На ветвях отображены значения бутстрепа.

От дерева в статье отличаются только ветви с низкой 'поддержкой'. Вот так часть дерева с низкой поддержкой выглядела в статье: (((Opsaridium_ubangienser, Raiamas_buchholzi), Raiamas_senegalensis), (Raiamas_guttatus, Raiamas_steindachneri)), Barilius_malabaricus); А так построенная в FastME (BME, mtREV): (((Raiamas_guttatus, Barilius_malabaricus), Raiamas_steindachneri), Raiamas_senegalensis), (Opsaridium_ubangienser, Raiamas_buchholzi));

От дерева, построенного мной в MEGA, данное дерево не отличается.

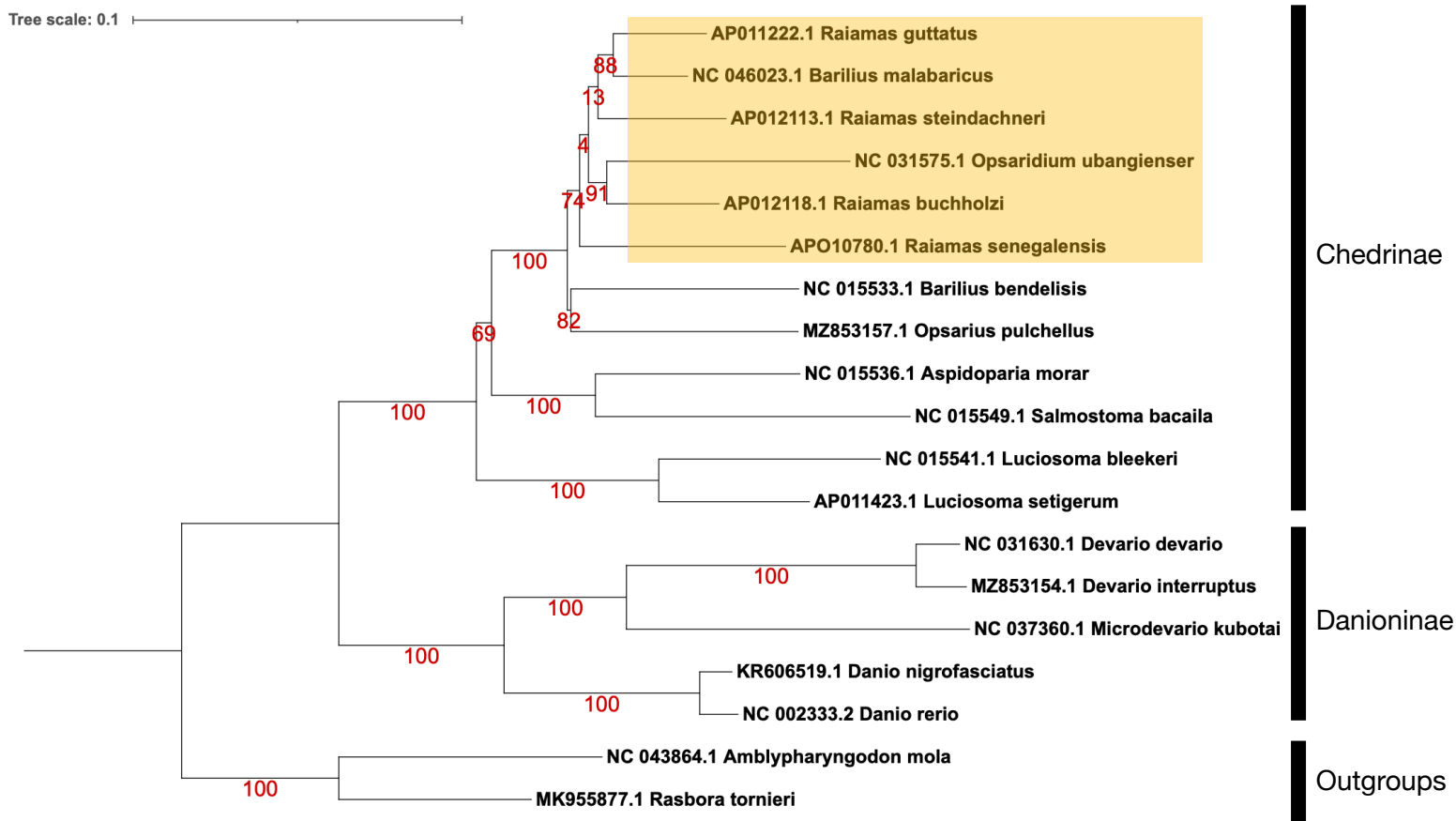


Рисунок 4. Филогенетическое дерево из FastME (метод ME, OLS, бутстреп 100, модель mtREV с гамма распределением).

На рисунке 4 изображено филогенетическое дерево, построенное в FastME. Метод минимальной эволюции (ME), вариант OLS, бутстреп 100, модели MtREV с гамма распределением. Укоренение с помощью внешней группы. На ветвях отображены значения бутстрапа.

От дерева в статье отличаются только ветви с низкой 'поддержкой'. Вот так часть дерева с низкой поддержкой выглядела в статье: (((*Opsaridium ubangienser*, *Raiamas buchholzi*), *Raiamas senegalensis*), (*Raiamas guttatus*, *Raiamas steindachneri*)), *Barilius malabaricus*); А так построенная в FastME (OLS, mtREV): (((*Raiamas guttatus*, *Barilius malabaricus*), *Raiamas steindachneri*), (*Opsaridium ubangienser*, *Raiamas buchholzi*), *Raiamas senegalensis*);

Деревья с реализацией разных вариантов минимальной эволюции (OLS и BME) отличаются друг от друга.

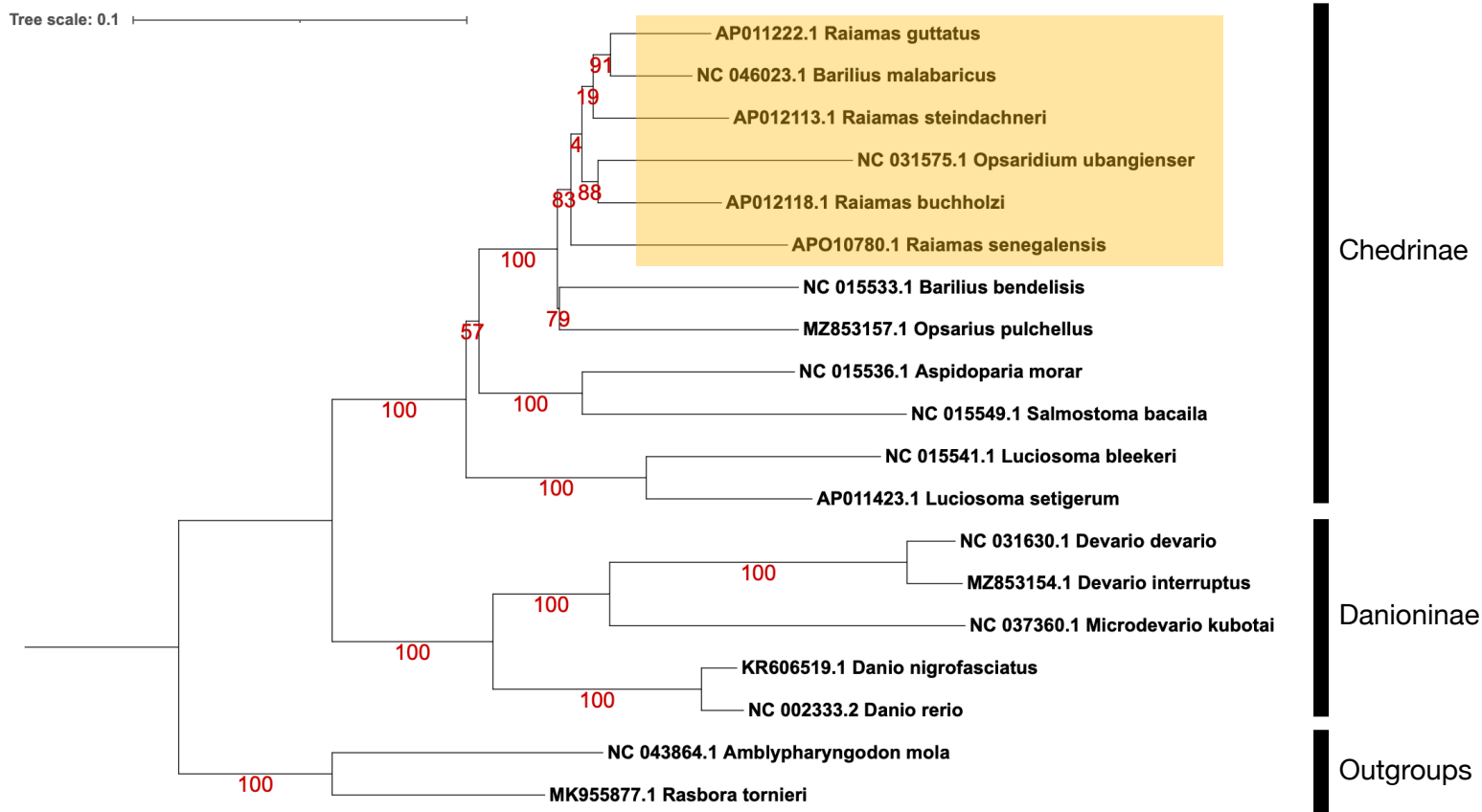


Рисунок 5. Филогенетическое дерево из FastME (метод ME, OLS, бутстреп 100, модель JTT с гамма распределением).

На рисунке 5 изображено филогенетическое дерево, построенное в FastME. Метод минимальной эволюции (ME), вариант OLS, бутстреп 100, модели JTT с гамма распределением. Укоренение с помощью внешней группы. На ветвях отображены значения бутстрепа.

От дерева в статье отличаются только ветви с низкой 'поддержкой'. Вот так часть дерева с низкой поддержкой выглядела в статье:
 (((Opsaridium_ubangiense, Raiamas_buchholzi), Raiamas_senegalensis), (Raiamas_guttatus, Raiamas_steindachneri)), Barilius_malabaricus);
 А так построенная в FastME (OLS, JTT):
 (((Raiamas_guttatus, Barilius_malabaricus), Raiamas_steindachneri), (Opsaridium_ubangiense, Raiamas_buchholzi), Raiamas_senegalensis);

Данное дерево не отличается от дерева, построенного с помощью метода минимальной эволюции, вариант OLS, модель MtREV с гамма распределением.

Список использованных источников

1. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018 Jun 1;35(6):1547-1549. doi: 10.1093/molbev/msy096. PMID: 29722887; PMCID: PMC5967553.
2. Adachi J, Hasegawa M. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J Mol Evol*. 1996 Apr;42(4):459-68. doi: 10.1007/BF02498640. PMID: 8642615.
3. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res*. 2013 Jan;41(Database issue):D36-42. doi: 10.1093/nar/gks1195. Epub 2012 Nov 27. PMID: 23193287; PMCID: PMC3531190.
4. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004 Mar 19;32(5):1792-7. doi: 10.1093/nar/gkh340. PMID: 15034147; PMCID: PMC390337.
5. <https://biopython.org/wiki/AlignIO>
6. Vincent Lefort, Richard Desper, Olivier Gascuel, FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program, *Molecular Biology and Evolution*, Volume 32, Issue 10, October 2015, Pages 2798–2800, <https://doi.org/10.1093/molbev/msv150>
7. Ivica Letunic, Peer Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Research*, Volume 49, Issue W1, 2 July 2021, Pages W293–W296, <https://doi.org/10.1093/nar/gkab301>