



# YaccConstructor

## Задачи на весенний семестр 2018

**Автор: Семён Григорьев**

Лаборатория языковых инструментов JetBrains  
Санкт-Петербургский государственный университет  
Математико-механический факультет

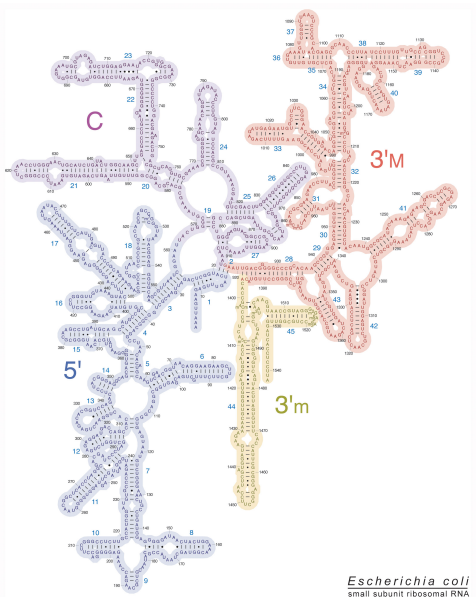
12 февраля 2018г.

- Исследования в области формальных языков
- Открытый исходный код
  - ▶ <https://github.com/YaccConstructor>
- Основной язык разработки — F#

# Классификация 16s на основе вторичной структуры

- Задача
  - ▶ Распознавание и классификация микроорганизмов
- Возможное решение
  - ▶ Использование особенностей вторичной структуры маркерных генов, которая более консервативна, чем первичная

# Вторичная структура

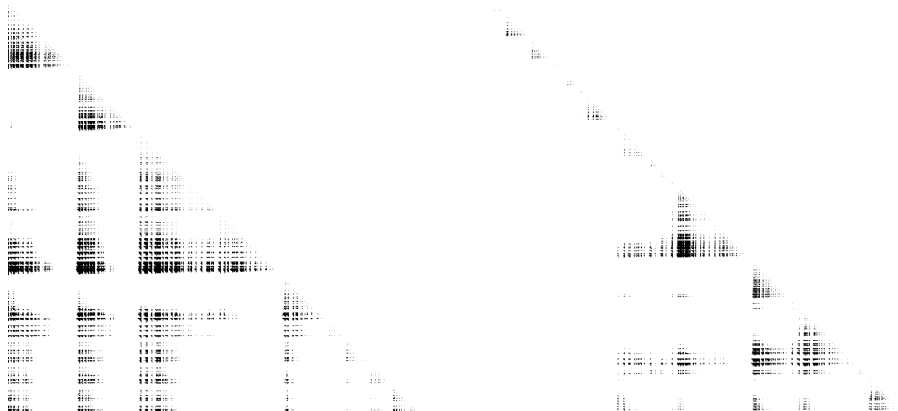


*Escherichia coli*  
small subunit ribosomal RNA

# Вторичная структура

- С некоторой точностью может быть описана контекстно-свободной грамматикой
- Извлечение особенностей вторичной структуры — синтаксический анализ
- Результат может быть представлен в виде битовой матрицы  $M$ : для фиксированного нетерминала  $S$   $M.[i,j] = 1$  если  $S \rightarrow^* \omega.[i,j]$

# Примеры матриц



# Задачи

- Обработка химер
- Интеграция с базой данных Greengenes
- Извлечение ключевых, с точки зрения нейронной сети, особенностей вторичной структуры
- Описание задач: <https://goo.gl/3jBUKF>
- Тестовая задача: <https://goo.gl/DmGNYs>
- “Направления развития”: курсовая, диплом, публикации

# Требования к знаниям и навыкам

- Методы и алгоритмы обработки данных, статистические методы, машинное обучение, нейронные сети (теория и практика)
- Умение читать и понимать научные статьи
- Умение читать и понимать чужой код



- Почта: `rsdpisuy@gmail.com`
- Исходный код YaccConstructor:  
`https://github.com/YaccConstructor`