

Использование КС-грамматика для распознавания центрального домена 16s

Semyon Grogorev

13 июня 2017 г.

1 Введение

Вторичная структура достаточно богата. Более того, известно, что некоторые участки обладают достаточно консервативной вторичной структурой.

Грамматика позволяет минимизировать знания о первичной структуре. Поиск структурного шаблона. Грамматика задаётся вручную, но возможен и вывод грамматика, но это тема для отдельного исследования.

2 Грамматика

Язык YARD. Примеры конструкций, которые используются при описании грамматики 16s.

Используемая грамматика.

Огромный листинг с грамматикой.

3 Эксперименты

Два эксперимента: обработка баз известных 16s, обработка полноразмерных геномов.

Базы 16s: все нитки должны быть распознаны.

Базы размеченных полноразмерных геномов с информацией о 16s: оценить точность, полноту и т.д. (сколько из отмеченных найдено, сколько из отмеченных не найдено, сколько найдено неотмеченных). Проанализировать ложные срабатывания и пропущенных кандидатов.