

Основные сведения о дереве кофе

Полина Дроздова

4 октября 2014 г.

Кофе (род *Coffea*) — это растение семейства Маревые (Rubiaceae), порядка Горечавкоцветные (Gentianales), клады Asterids. Для производства любимых нами всеми зёрен используют в основном два вида: *C. arabica* и *C. canephora* [<http://en.wikipedia.org/wiki/Coffea>]

Недавно (статья в Science вышла 5 сентября 2014 года) был секвенирован геном *C. canephora*. 11 пар хромосом, 710 Mbp. [<http://www.sciencemag.org/content/345/6201/1181.full>]. Авторы не пишут прямо, но сильно подозреваю, что этот вид предпочли потому, что он диплоид, а *C. arabica* — тетраплоид. Чем больше геном, тем сложнее собирать.

Ещё некоторые полезные ссылки:

<https://genomeevolution.org/wiki/index.php/File:Totalsequencedtree.png>

<http://www.coffeegenome.org/>

<http://coffee-genome.org/>

Я утверждаю, что ближайший к кофе секвенированный геном — это либо помидор (*Solanum lycopersicum*), либо картофель (*S. tuberosum*), двух представителей сем. Паслёновые (Solanaceae), входящего в ту же кладу Asterids. Всего десяток лет назад кофе относили к тому же семейству Паслёновые. Проекты по секвенированию генома картофеля и томата живут здесь:

<http://www.sgn.cornell.edu/>.

Ещё можно попробовать поработать с генами классического модельного объекта генетики растений *Arabidopsis thaliana*. Насколько я знаю, основной кладёзь информации про этот геном здесь:

<http://www.arabidopsis.org/>.

Арабидопсис намного дальше от кофе, чем томат и картофель, зато изучен несравнимо лучше.

Я вот нашла статью о том, что у кофе и помидора схожие наборы генов, а на арабидопсисные они как раз не очень похожи: [<http://link.springer.com/article/10.1007%2Fs00122-005-0112-2>].

А это самая главная страничка, с которой и стоит скачивать данные:

[<http://coffee-genome.org/download>] Паша Добрынин: сказал, что мы можем работать с файлами .fasta с каждой хромосомой. Разбираться со скэффолдами отдельно веселее, но сложнее.

Файл .gff3 в самом начале — это и есть цель нашей работы, поэтому заглядывать в него пока не стоит.

В общем, пока что я предлагаю работать с генами *Arabidopsis thaliana*, *Solanum lycopersicum* и *Solanum tuberosum* и посмотреть, что лучше получится.

Код

```
tblastn -query ./raw_data/sample-proteins.fa -db nr -remote -out ./blastx.out  
exonerate --model protein2genome sample-proteins.fa coffeechrom.fna >../exonerate_sam
```