

Использование карт для сборки геномов

Алекс Макунин

Центр Геномной Биоинформатики СПбГУ
Лаборатория Сравнительной Геномики,
ИМКБ СО РАН, Новосибирск

Конечная цель сборки генома

- Финишированный геном – полная последовательность всех хромосом без гэпов («дырок»)

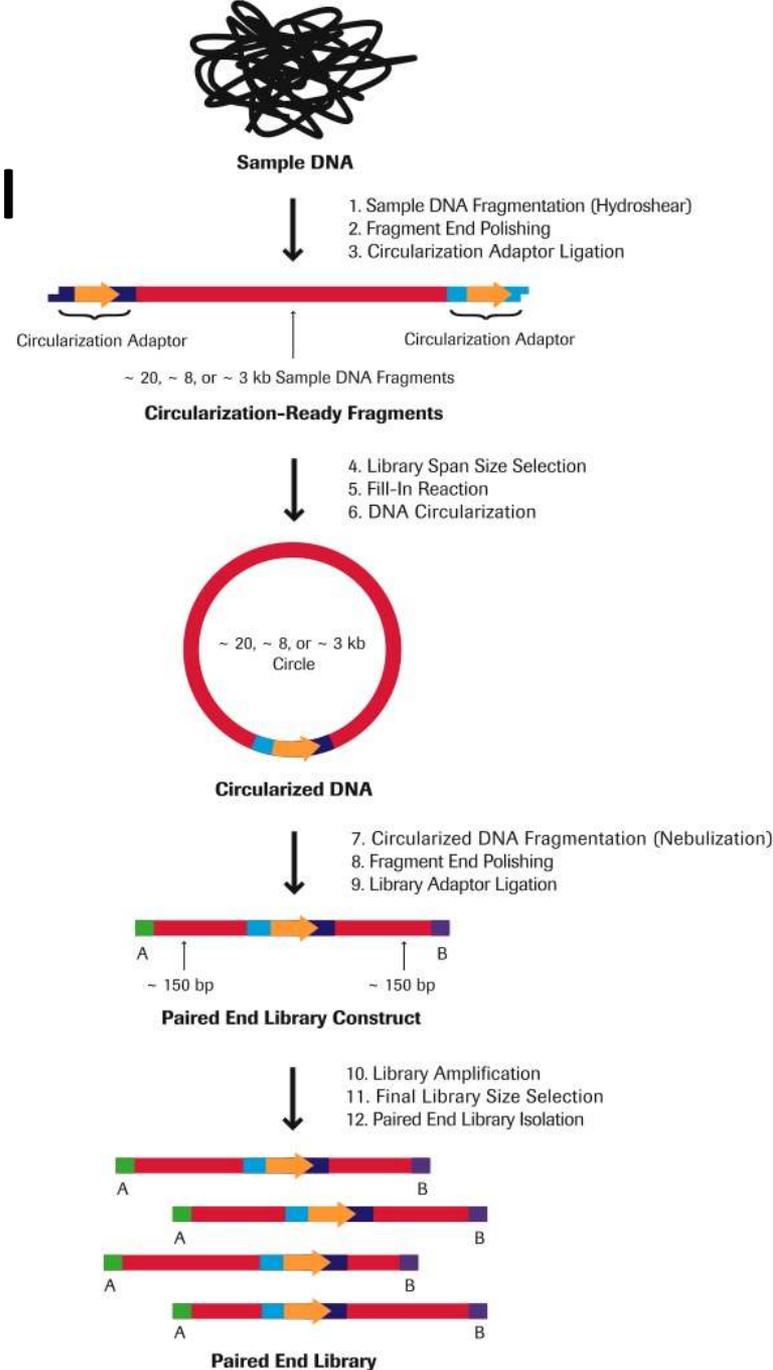
Что мешает финишировать большие геномы?

- Тандемные повторы (центромера, сателлиты)
- Диспергированные повторы
- Сегментные дубликации
- Полногеномная дубликация

Как преодолеть недлинные повторы
или скэффолдинг

Методы: парные риды

- Фрагментация генома
- Отбор фрагментов нужного размера
- Циркуляризация-фрагментация
- Прикрепление адаптеров
- Секвенирование



Методы: парные ряды

+

- Высокая производительность
- Автоматизация

-

- Ограниченная длина (10-20 тпн)
- Образование химер
- Неизвестна точная длина вставки

Методы: длинные ряды

+

- Довольно высокая производительность
- Автоматизация
- Известная длина

-

- Ограниченная длина (ок 5 тпн для РасВіо)
- Много ошибок чтения

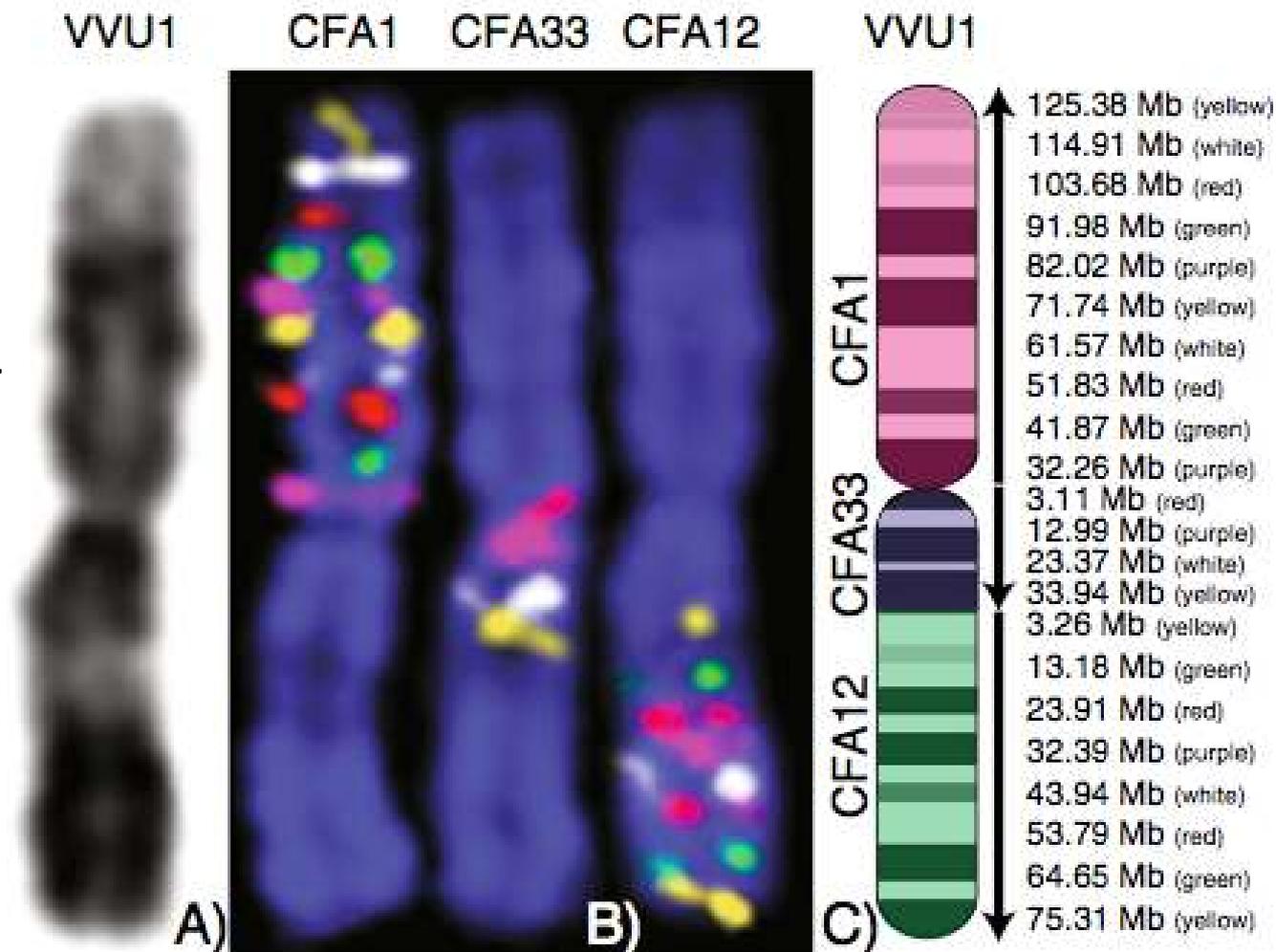
Как взглянуть на хромосому
целиком или карты генома

ВАС-клоны

- ВАС – bacterial artificial chromosome – может нести фрагмент длиной до 350 тпн
- Последовательность фрагментов (клонов) определяем одновременной гибридизацией (FISH) нескольких близлежащих на хромосому.
- Положение на скэффолдах определяем секвенированием концов клонов или определением полной последовательности клона

ВАС-клоны. Пример

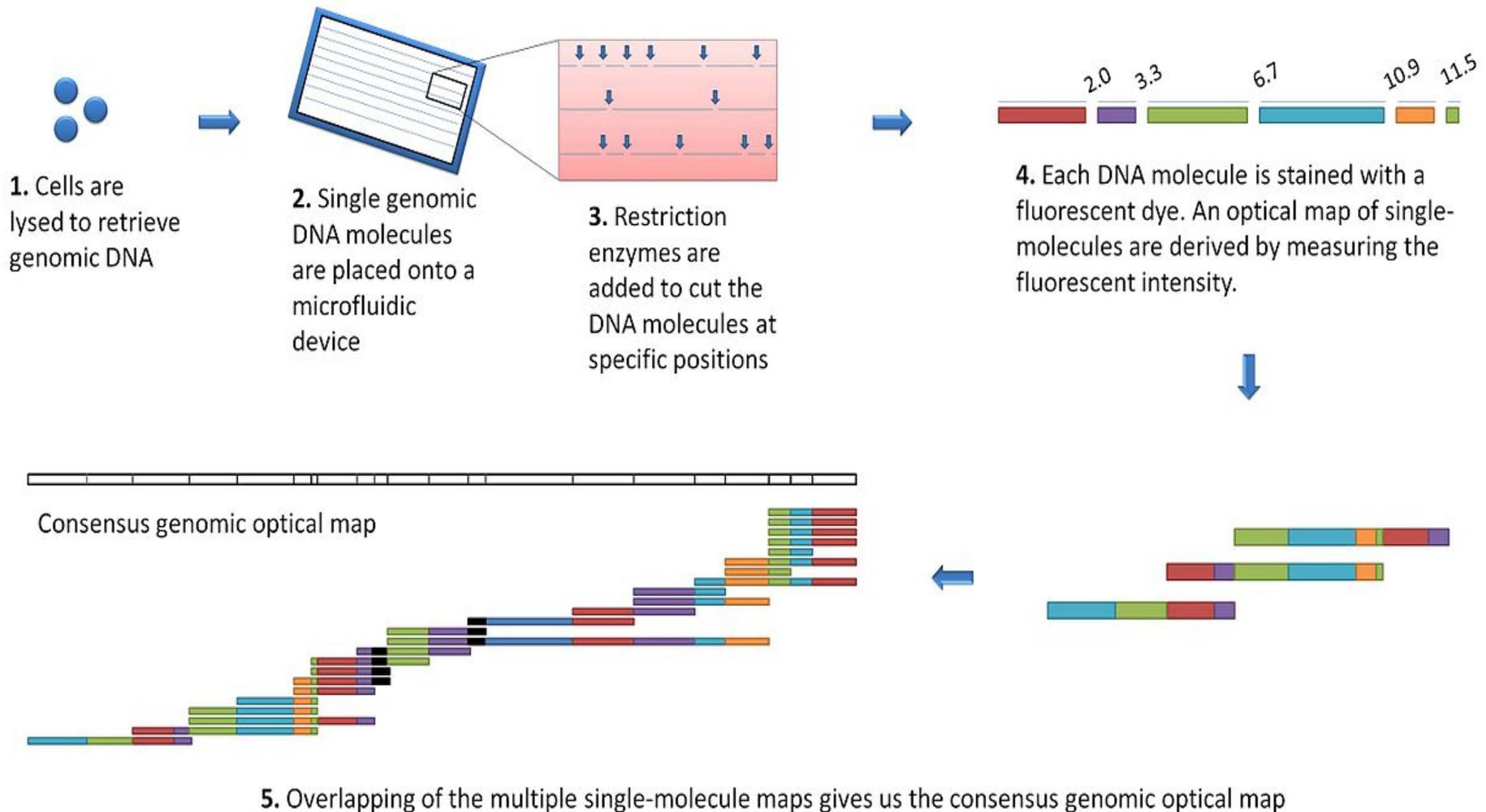
275 ВАС-клонов
собаки (разрешение
10мпн)
локализованы на
хромосомах лисицы.



ВАС-клоны

- Разрешение зависит от числа клонов
- Прямая визуализация
- Можно найти крупную дубликацию
- Локализация – медленно и вручную
- Не все районы вставляются в ВАС

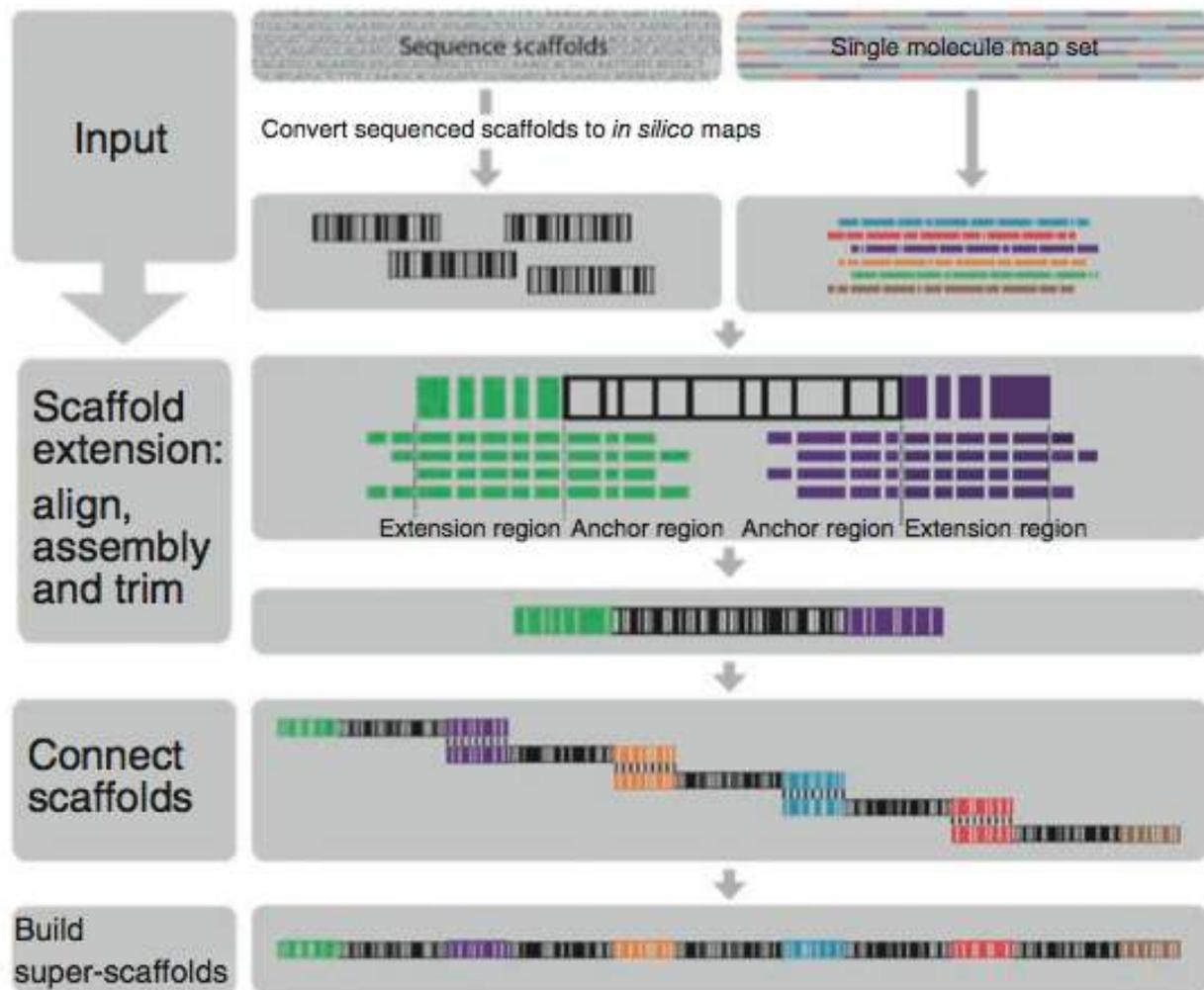
Оптическое картирование



Оптическое картирование

Сборка карты с использованием скэффолдов:

- Поиск сайтов рестрикции на скэффолдах
- Наложение карт на скэффолды
- Объединение карт внутри скэффолда
- Объединение скэффолдов в суперскэффолды



Оптическое картирование. Пример

- Исследовали геном козы
- Получили 3.5 млн карт
- Из 2090 скэффолдов с помощью этих карт получили 315 суперскэффолдов
- Суперскэффолды и скэффолды собраны в псевдохромосомы по локализации 110 тысяч зондов SNP коровы

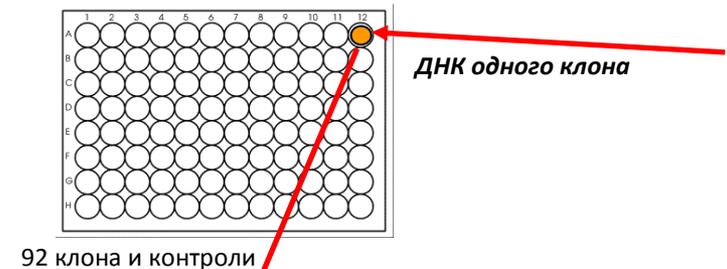
Оптическое картирование

- Разрешение определяется подбором ферментов рестрикции (0.2-1Mbp)
- Патентованная высокопроизводительная технология (прибор MapCard + софт Genome-Builder)

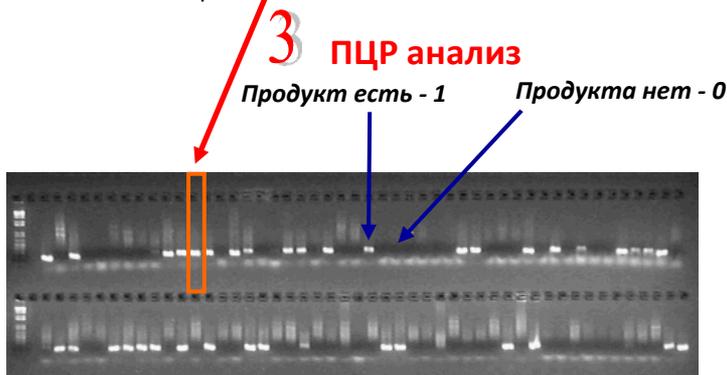
Радиационные гибриды



Панель радиационных гибридов



Одни фрагменты хромосом попадают в одни клоны, другие в другие и таким образом создается разнообразие

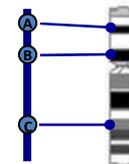


4 Статистический анализ

A=1010000001111011...
 B=1010000001011011...
 C=0001001000101011...

Для каждого маркера строится вектор, отражающий присутствие/отсутствие маркера в каждом гибридном клоне

5 Карта радиационных гибридов

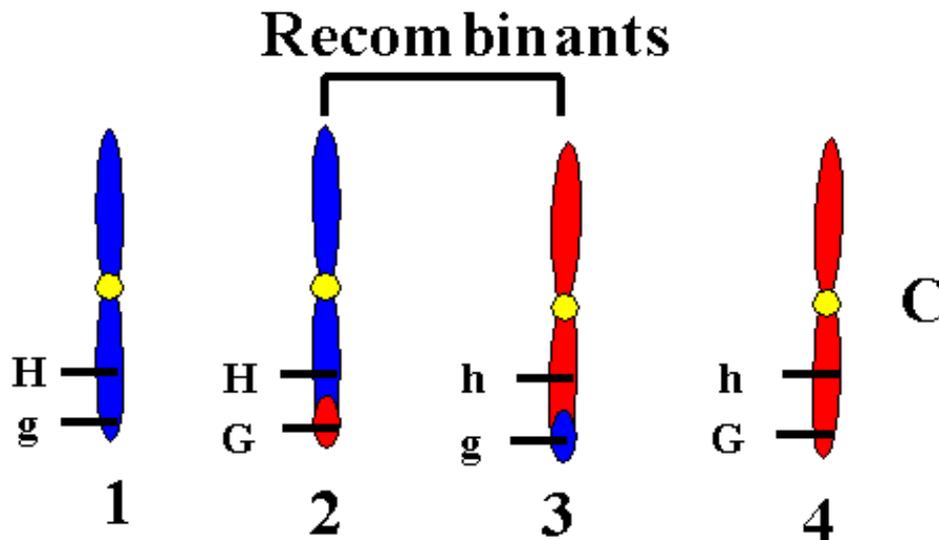
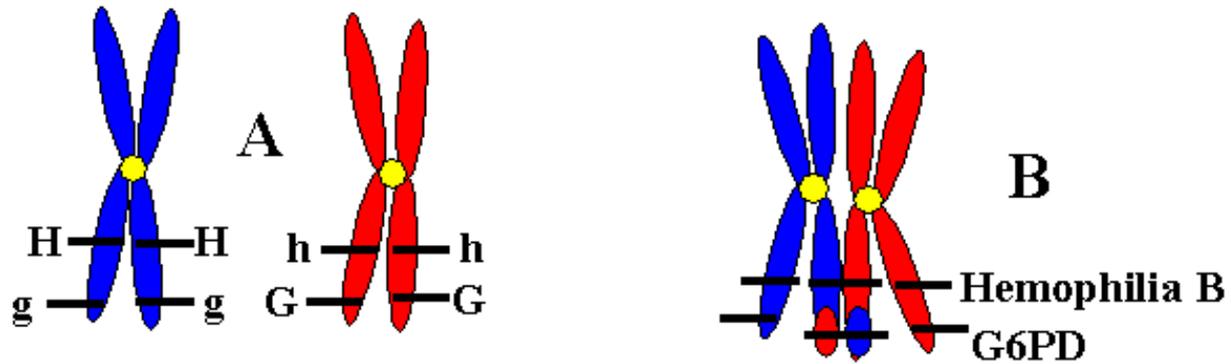


Статистически обобщается различия между векторами, отражающая расстояние между маркерами.
 У маркеров А и В похожее распределение 1 и 0, что означает, что данные маркеры расположены довольно близко на одной хромосоме.

Радиационные гибриды

- Разрешение карты зависит от фрагментированности хромосом
- Панель радиационных гибридов – дорого
- Маркеры – панель зондов вместо ПЦР
- Сборка карт – программами CarthaGene, rh_tsp_map

Генетические карты. Рекомбинация



Генетические карты

- Популяция – семья из трех поколений (есть другие варианты)
- Маркеры – ПЦР-фрагменты, SNP панели.
- Сборка карты – программы CriMap, MultiMap, JoinMap.

Генетическая карта и карта радиационных гибридов для сборки генома кошки (ок. 2.5Gbp)

	GL	RH
Маркеров использовано	62,897 SNP	2,674 STR
Длина карты	3970 cM	2540 cR
Уникальные позиции карты	1,918	1,453
Маркеров в уникальных позициях	11,742	1,746
Скэффолды с маркерами	240	199
Химерные скэффолды	34	66
Конечное число фрагментов	274	265

Генетические карты

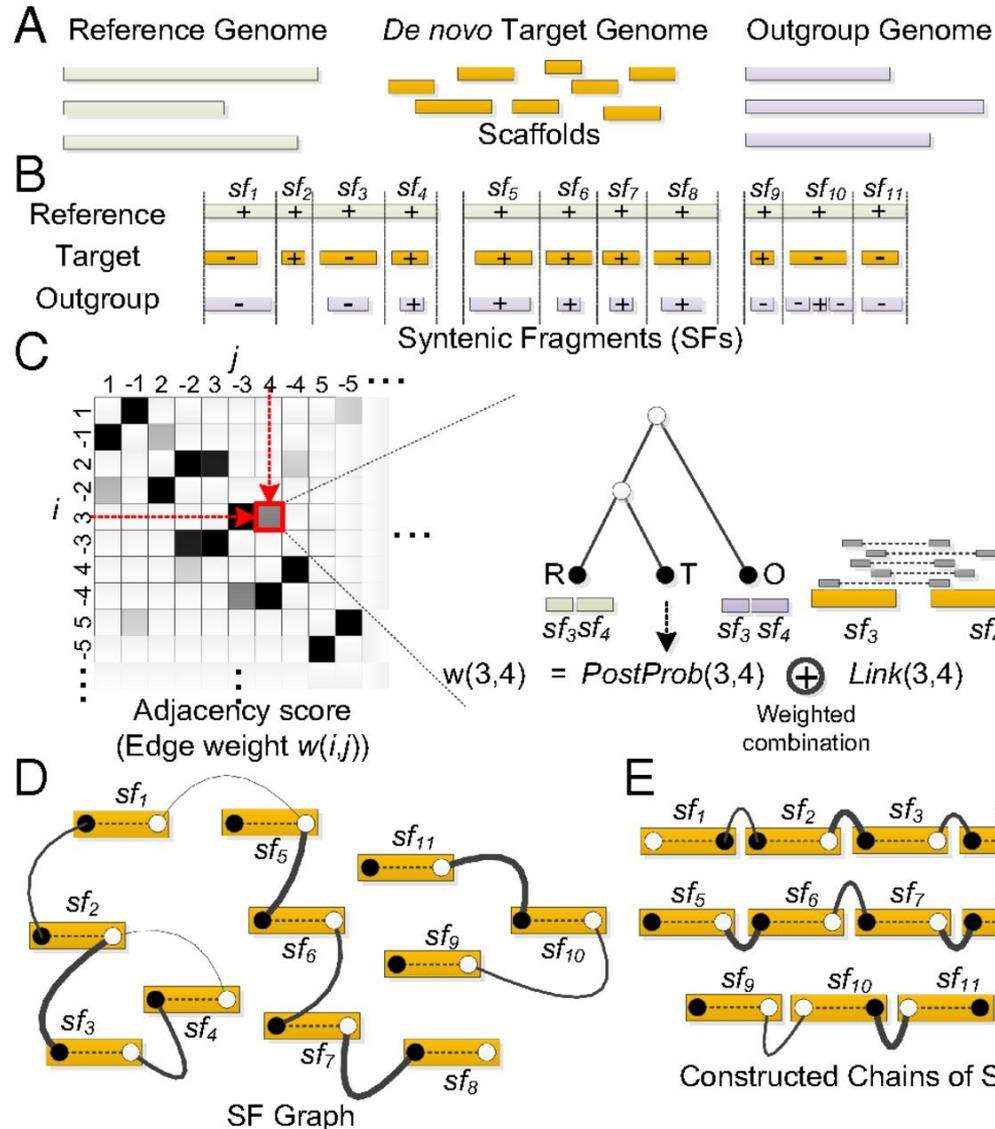
- Алгоритм чувствителен к числу маркеров
- Разрешение зависит от числа родственных особей
- Нужна панель SNP

Картирование секвенированием отдельных сперматозоидов – решение последних двух проблем.

Сравнительный подход

- Если районы соседствуют у родственных видов, вероятно они будут соседями у нашего вида (исключая случаи хромосомных перестроек)
- Можно сравнивать как полную последовательность (multiple genome alignment), так и отдельные маркеры (SNP, STR)

Reference-assisted chromosome assembly



Kim J et al. PNAS 2013;110:1785-1790

Reference-assisted chromosome assembly

- Геном тибетской антилопы (опыт), коровы (референс) и человека (внешняя группа)
- 1,434 (из 15,996) скэффолдов тибетской антилопы разбито на 1,597 синтенных фрагментов.
- Из синтенных фрагментов реконструированы 60 фрагментов хромосом тибетской антилопы ($2n=60$). 16 из них гомологичны полным хромосомам коровы

Заключение

- Построение карты – последовательности точек с привязкой к сиквенсу
- Построение порядка скэффолдов в соответствии с картой
- Сравнение или объединение данных разных карт

Заключение

- ВАС-клоны (BAC-clone mapping)
- Оптическое картирование (optical mapping)
- Радиационные гибриды (radiation hybrid)
- Генетическое картирование (genetic linkage mapping)
- Сравнительные методы (comparative approaches)