Выравнивание геномов

Сходство и гомология

Сходство – характеристика последовательностей в соответствии с некими критериями

Гомология – сходство последовательностей, вызванное их общим происхождением.

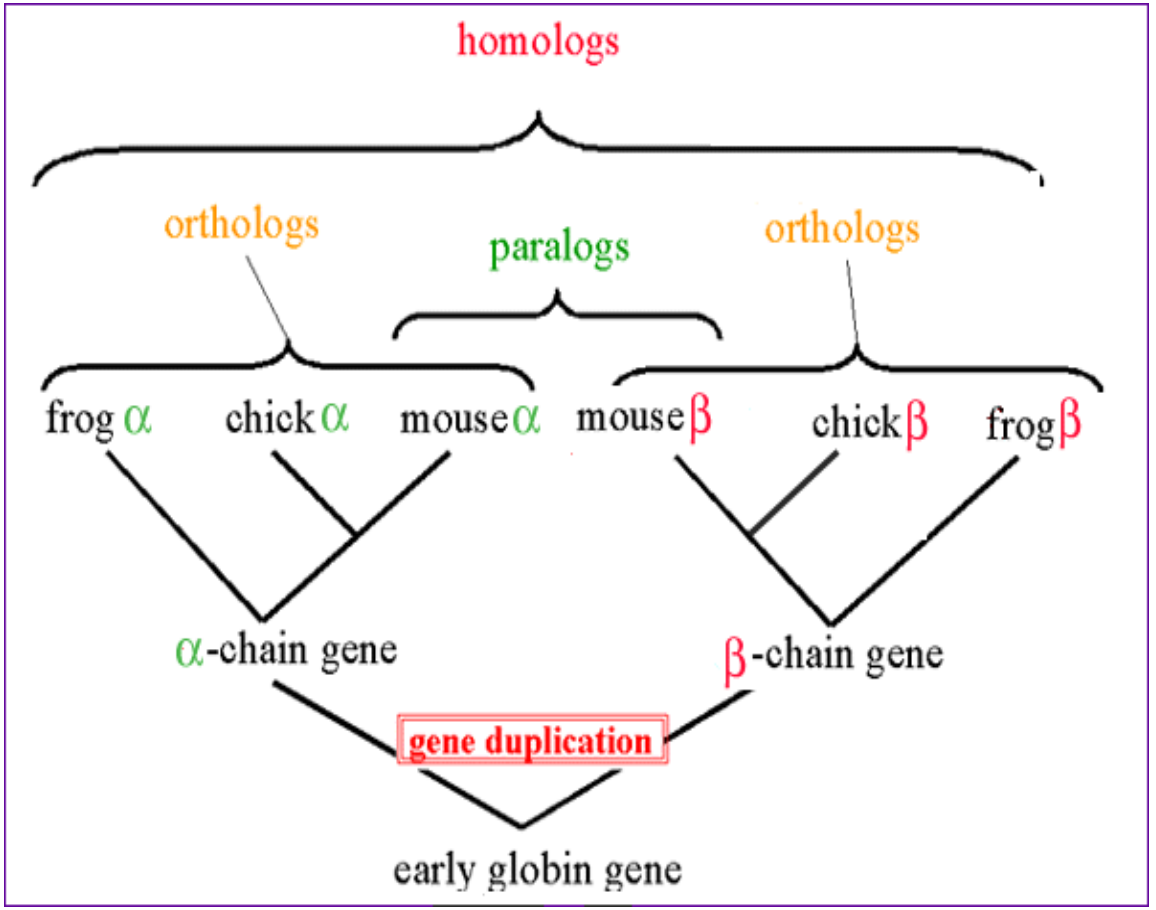
Сходство и гомология часто употребляются, как синонимы, но ими не являются.

Сходство и функция

Консервативные участки соответствуют функционально важным участкам белка

Менее функционально значимые участки легко накапливают случайные мутации

Гомологи глобина



Эволюционные события

Локальная эволюция - точечные замены, вставки, делеции

Глобальная эволюция – одномоментное изменение больших фрагментов генома (вставки, делеции, дупликации, инверсии, транслокации, слияние-разделение хромосом, горизонтальный перенос генов)

Парное выравнивание

Задача: найти ортологичные участки

Вход: две последовательности

Результат:

-набор пар гомологичных участков и их выравниваний

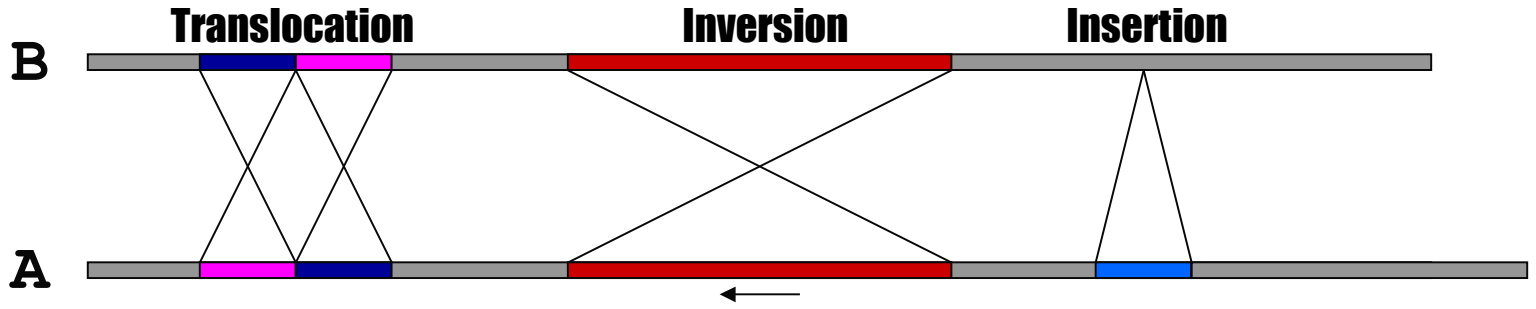
-визуализация

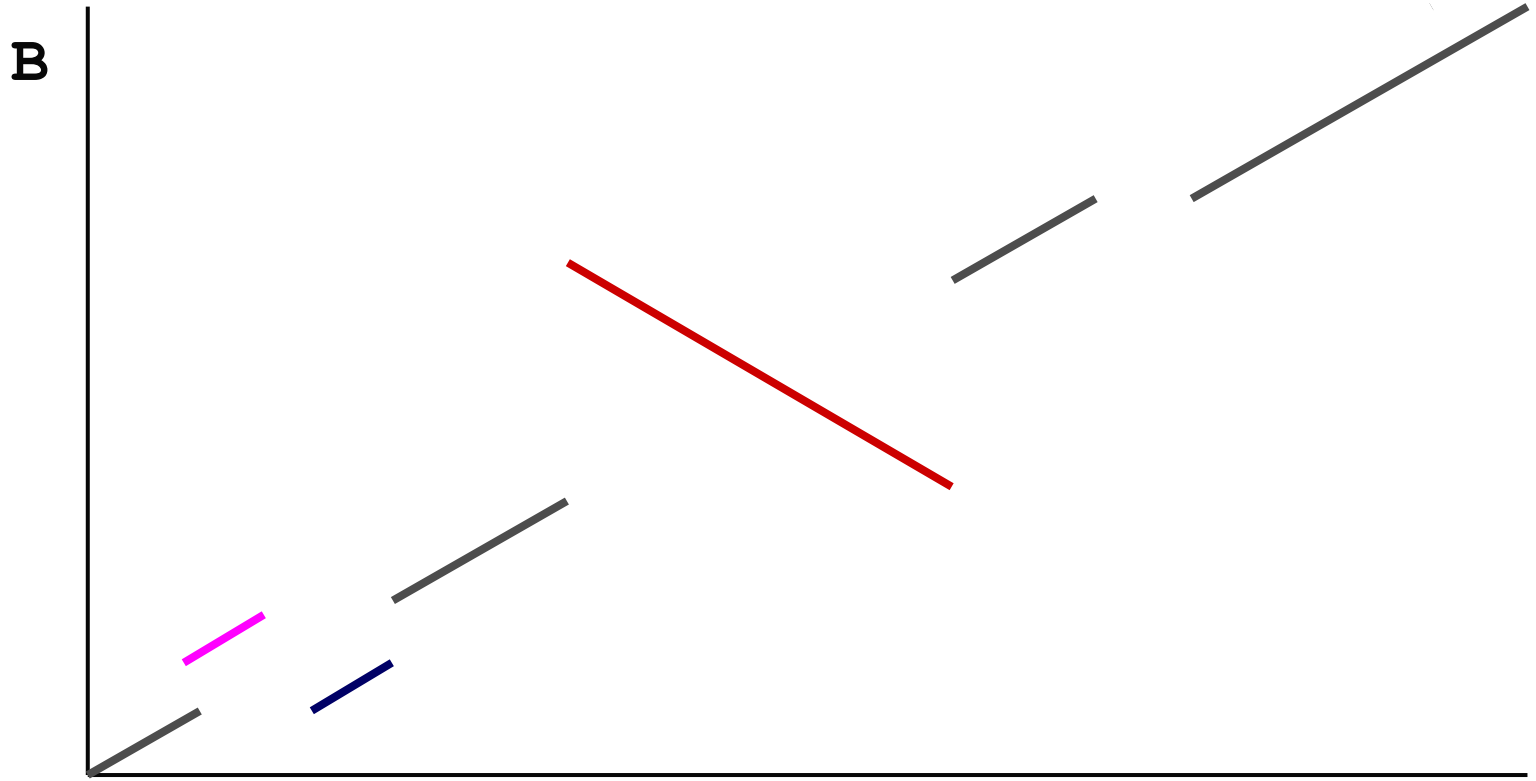
Пример: программа blast2seq и карта локального сходства

blast2seq

База – одна последовательность
Вход – другая последовательность
Результат – набор находок, т.е. пар фрагментов, похожих друг на друга
Алгоритм – тот же blast (обычно blastn)
Визуализация – карта локального сходства:
Каждая находка изображается отрезком

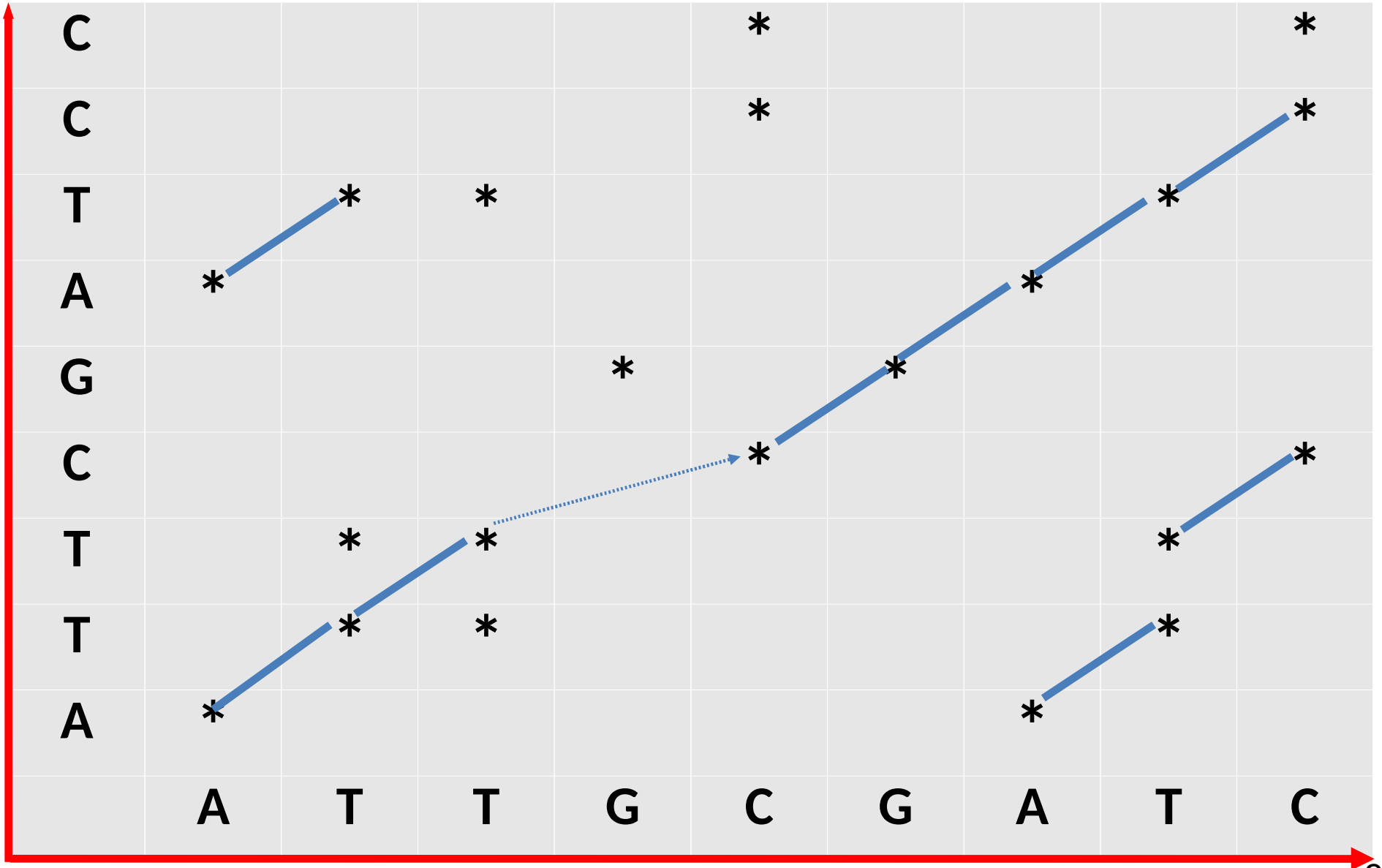
Карта локального сходства



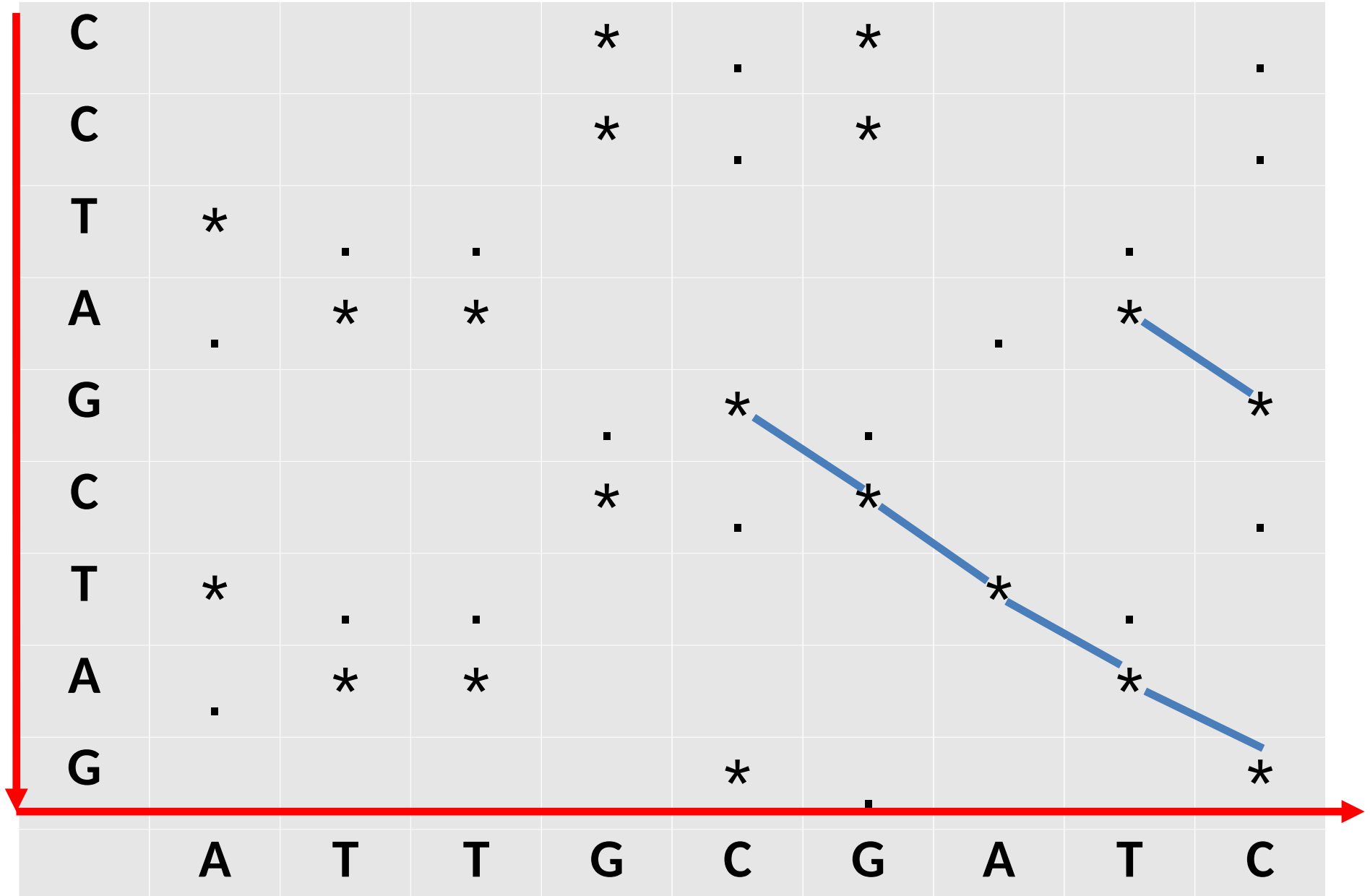


http://mummer.sourceforge.net/manual/AlignmentTypes.pdf

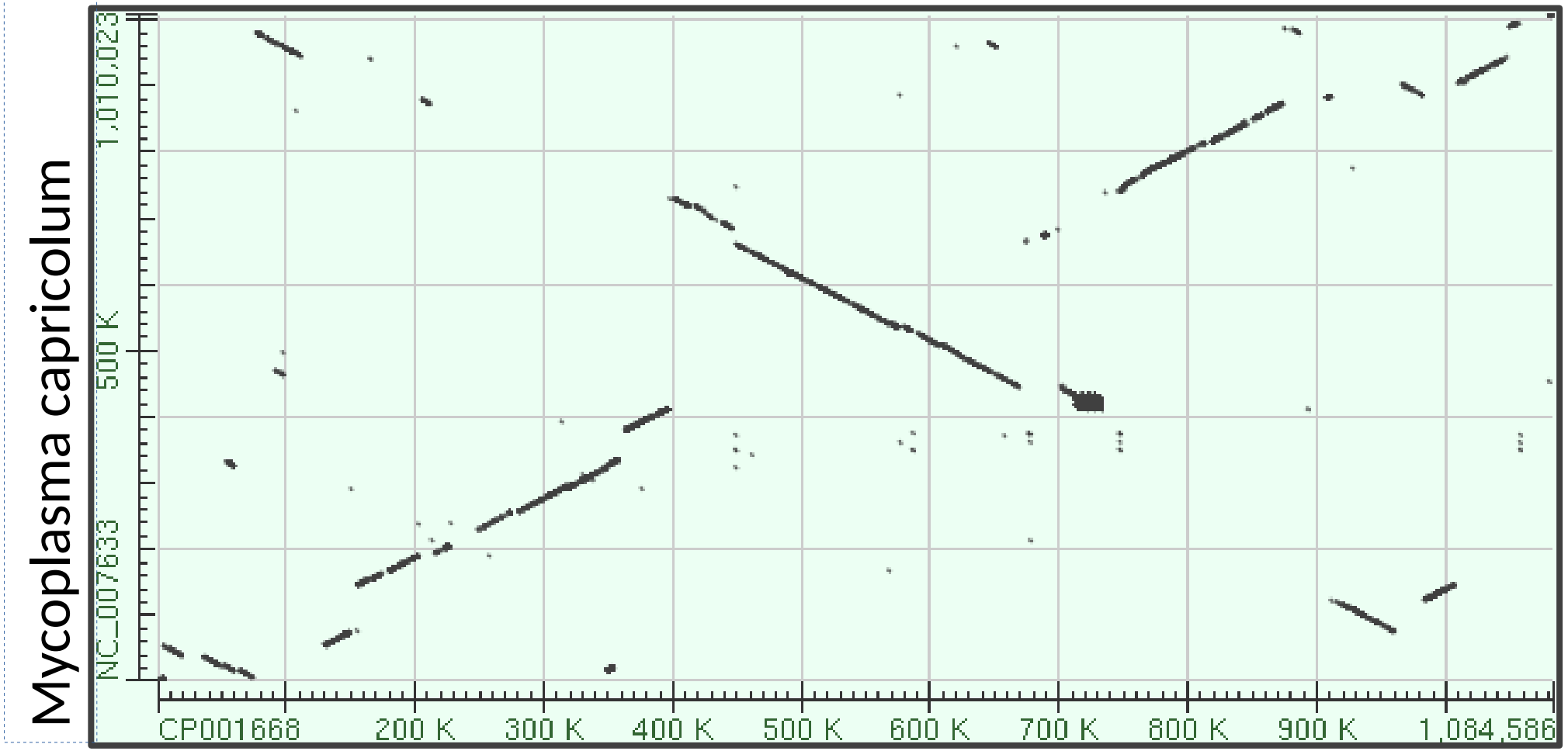
Карта сходства 2х последовательностей



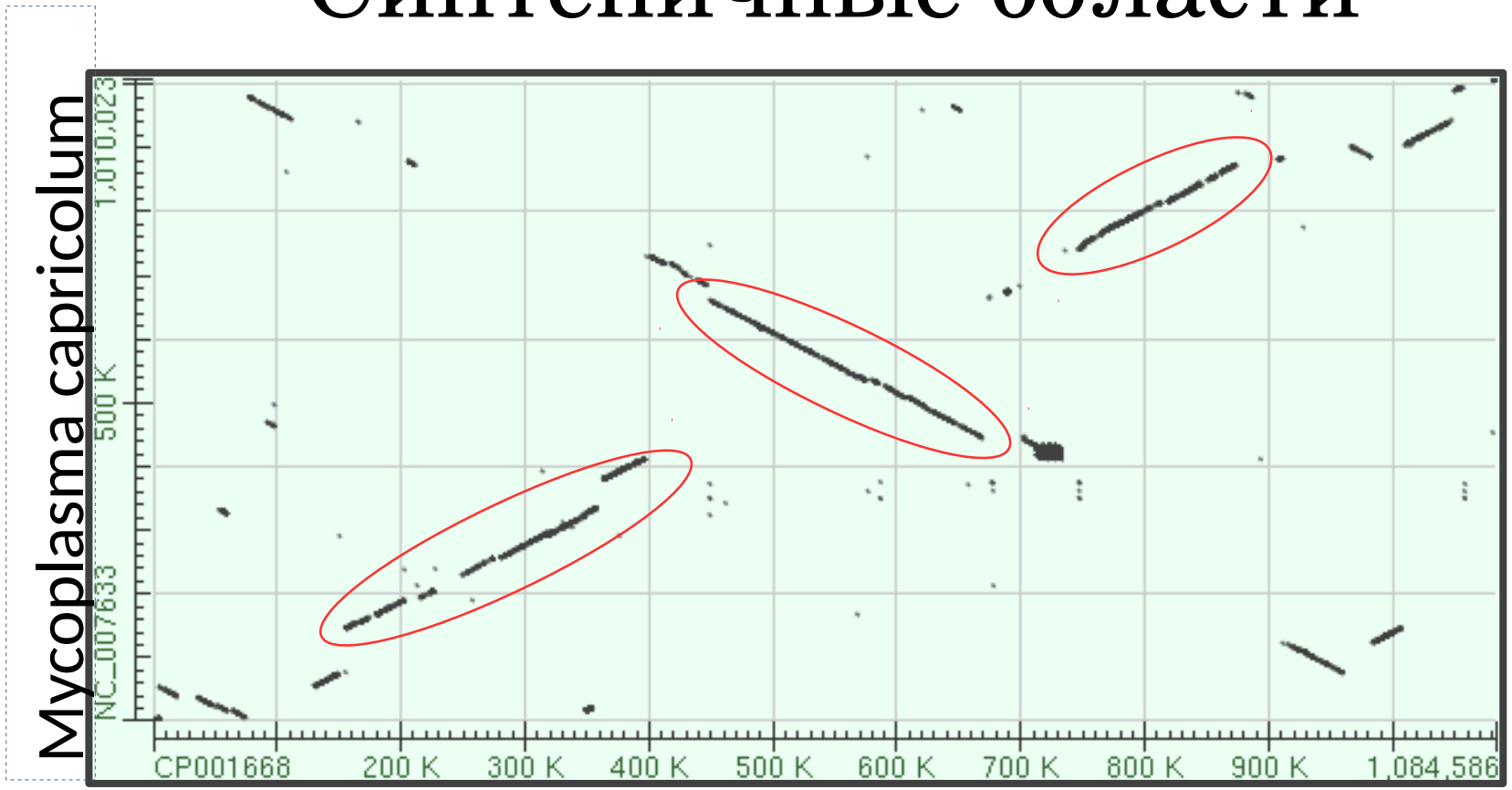
Карта сходства 2х последовательностей с учетом комплементарной цепочки



Карта локального сходства геномов *M.capricolum* и *M.mycoides*



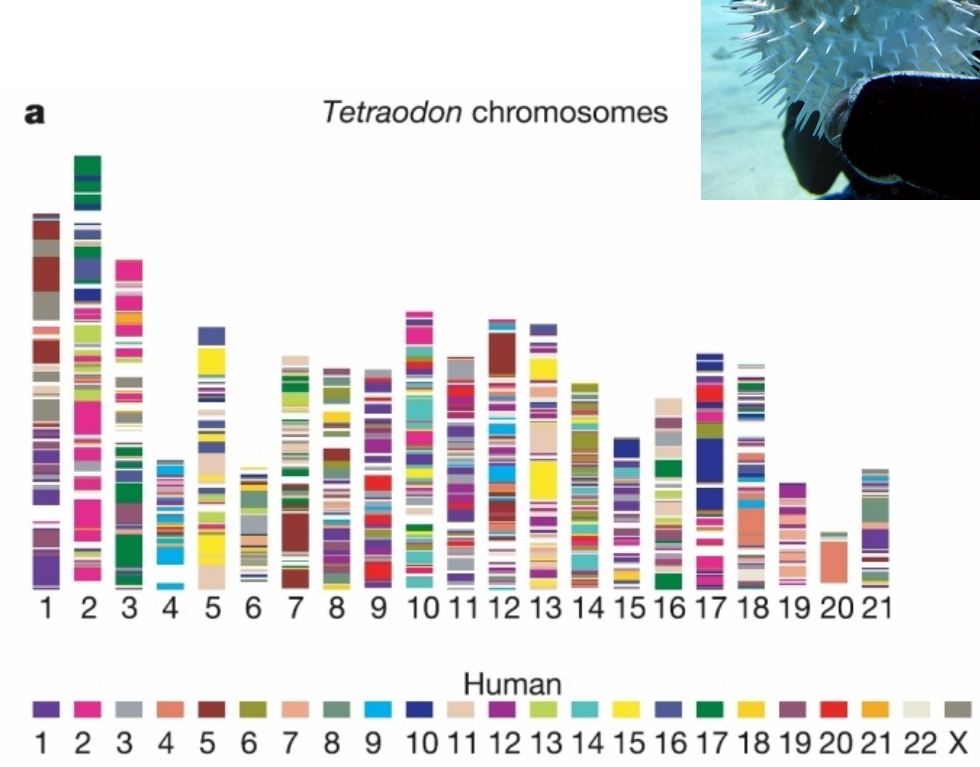
Синтеничные области

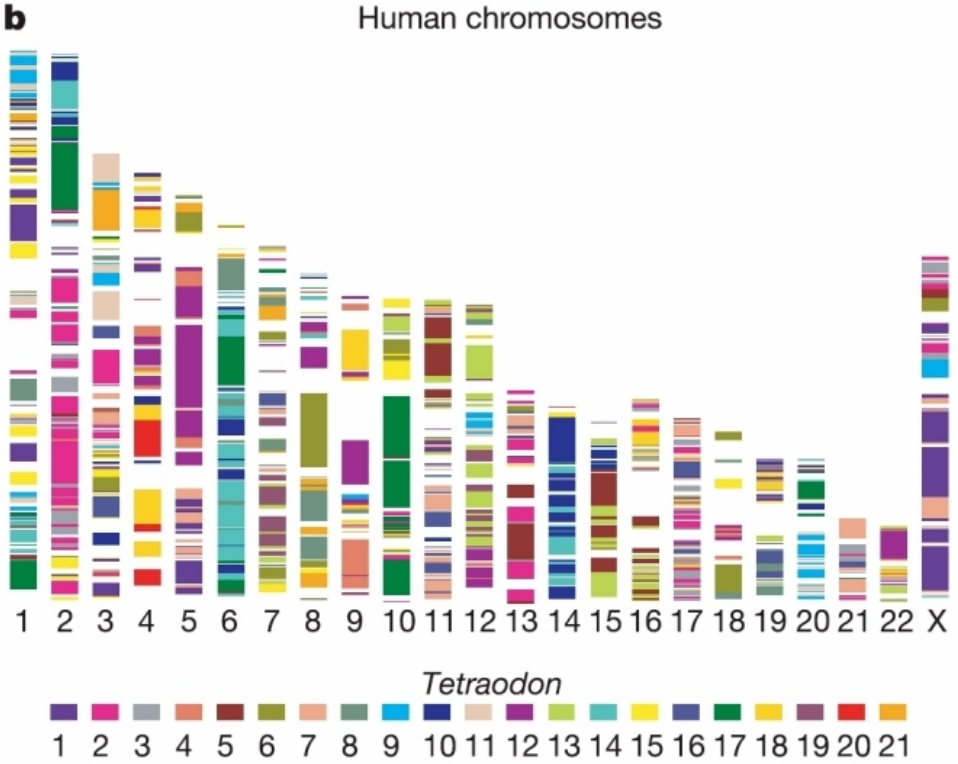


Mycoplasma mycoides

Синтеничные области - участки геномов, состоящие из ортологичных областей с сохранением их порядка на хромосоме для сравниваемых геномов

Синтеничные фрагменты на хромосомах человека и скалозуба





Jaillon, O., et al. Nature 431, 946–957 (2004) doi:10.1038/nature03025

Множественное выравнивание

- Вход:
  - несколько геномов
  - каждый геном представлен одной или несколькими хромосомами и плазмидами
  - вариант: набор контигов или скэффолдов
- Результат:
  - Наборы ортологичных участков из всех или части геномов и их выравнивание (блок)
  - последовательность наборов в каждой ДНК

Пангеном

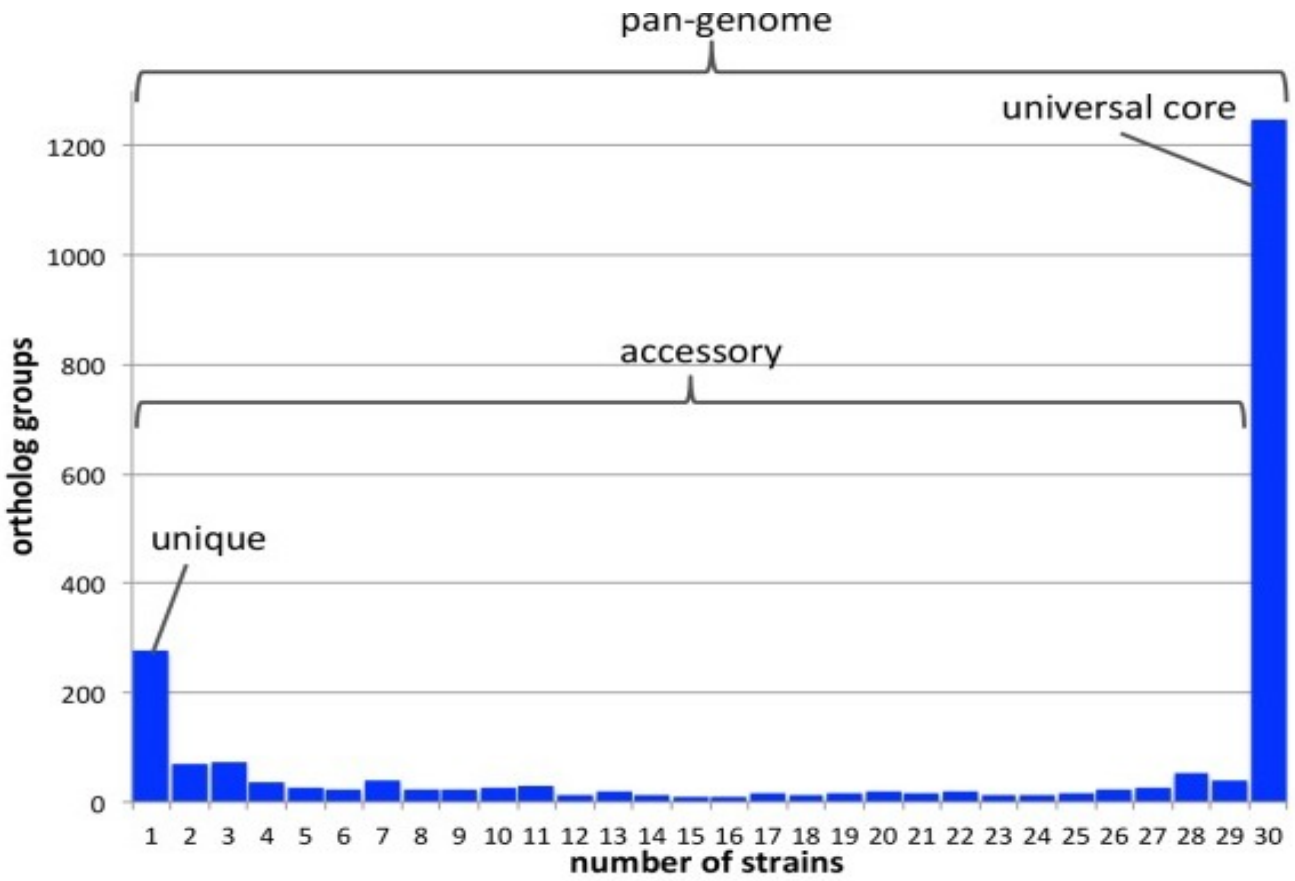
Пангеном может быть построен на основе ортологичных генов или ортологичных участков генома (нуклеотидный пангеном)

Кор — блоки, которые включают участки, встречающиеся во всех геномах

Дополнительные участки — встречаются не во всех геномах

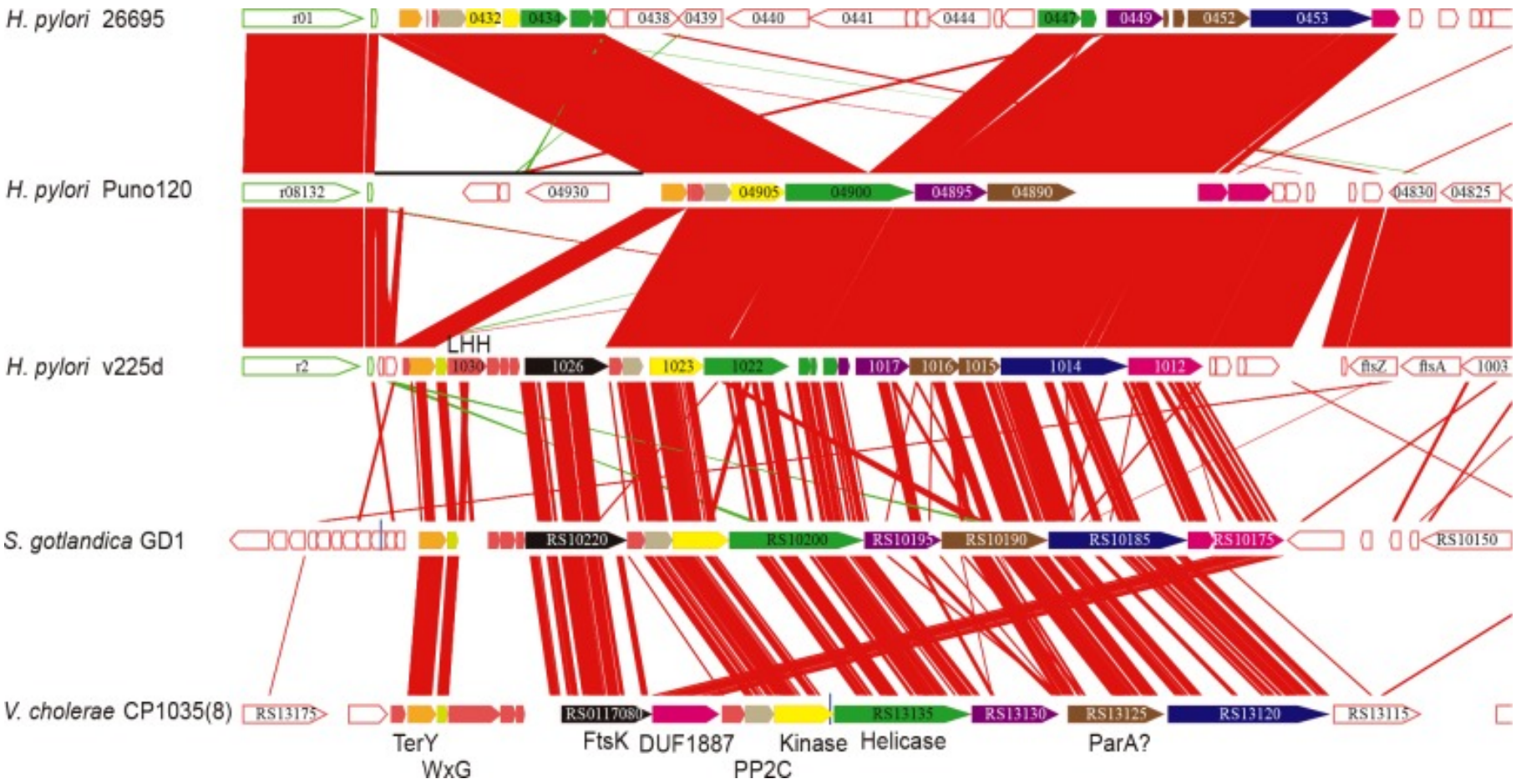
Уникальные участки — встречаются только в одном геноме

Пангеном 30 штаммов *Helicobacter pylori*



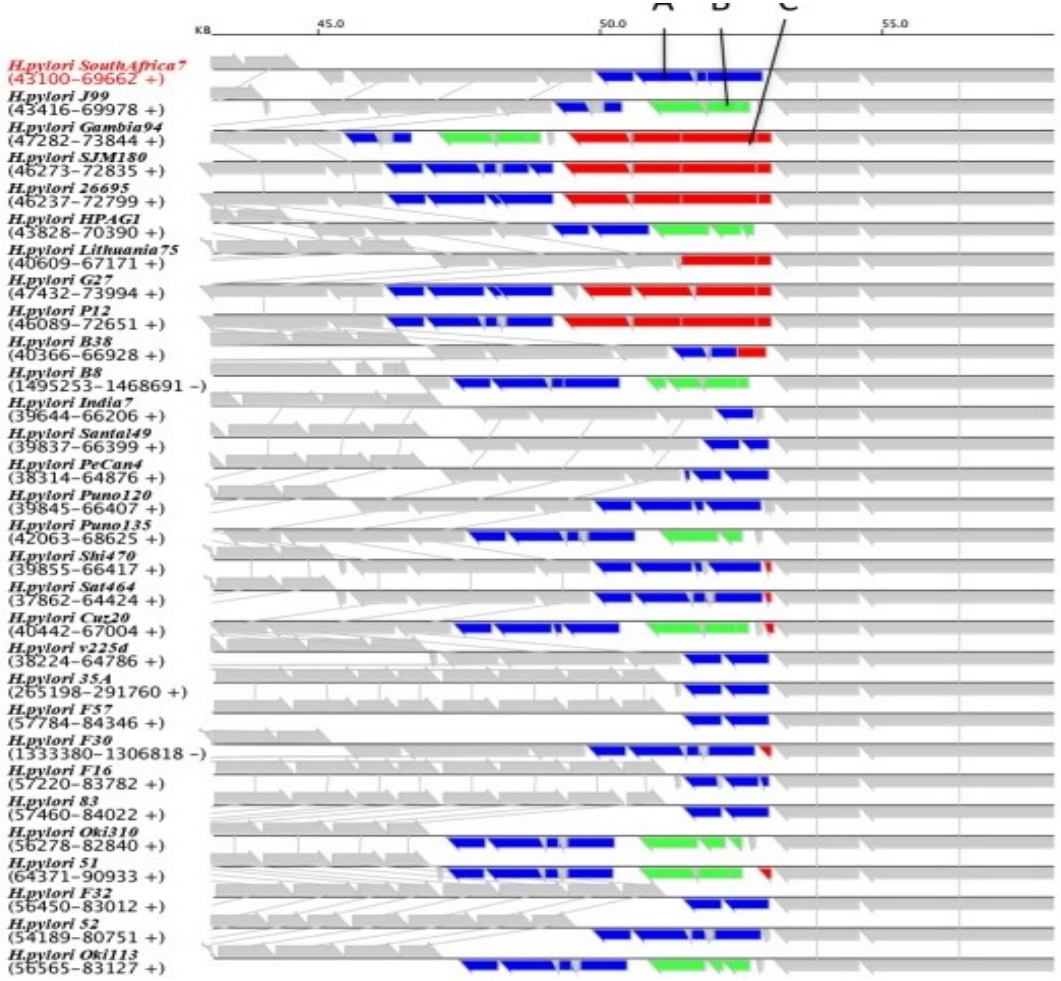
Uchiyama I, Albritton J, Fukuyo M, Kojima KK, Yahara K, Kobayashi I. A Novel Approach to Helicobacter pylori Pan-Genome Analysis for Identification of Genomic Islands. PLoS One. 2016 Aug 9;11(8):e0159419.

Сравнение нескольких геномов: визуализация



Uchiyama I, Albritton J, Fukuyo M, Kojima KK, Yahara K, Kobayashi I. A Novel Approach to Helicobacter pylori Pan-Genome Analysis for Identification of Genomic Islands. PLoS One. 2016 Aug 9;11(8):e0159419.

Горизонтальный перенос генов



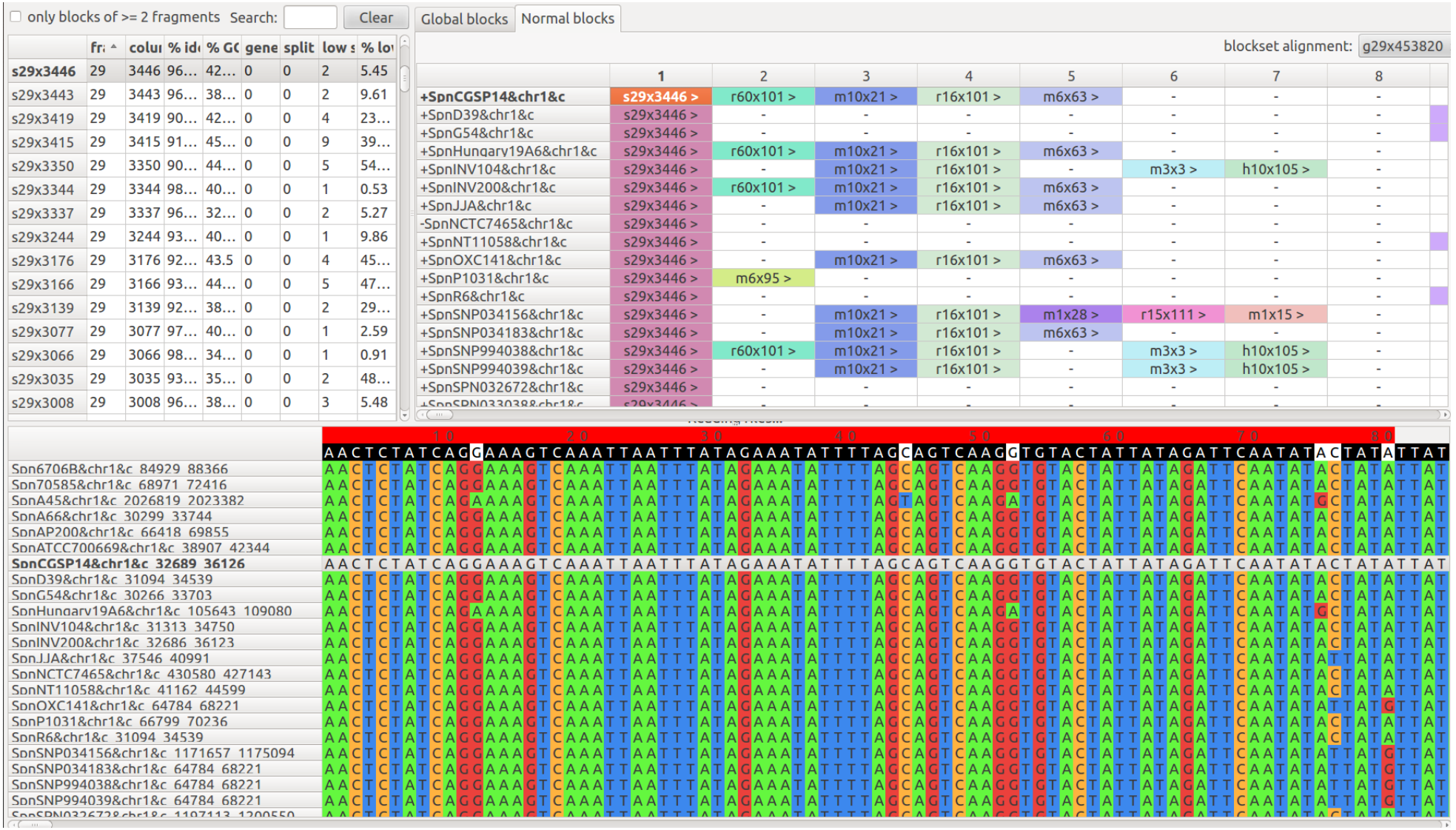
Uchiyama I, Albritton J, Fukuyo M, Kojima KK, Yahara K, Kobayashi I. A Novel Approach to Helicobacter pylori Pan-Genome Analysis for Identification of Genomic Islands. PLoS One. 2016 Aug 9;11(8):e0159419.

NPG Explorer

Программа для поиска и визуализации нуклеотидного пангенома http://mouse.belozersky.msu.ru/tools/npge.html

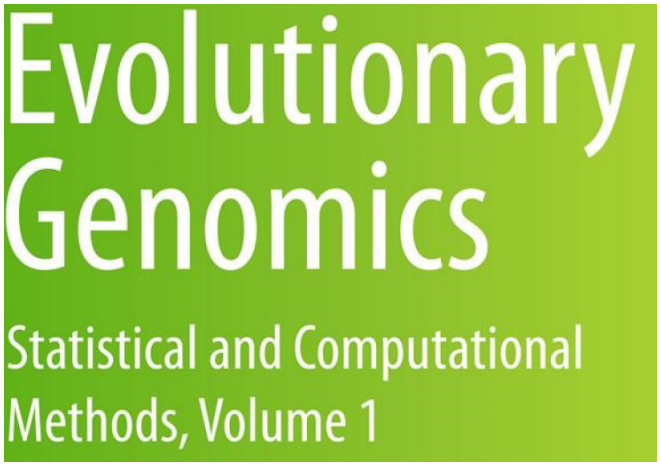
Для построения пангенома NPGE использует данные о гомологичных участках, а не генах
Один блок может включать ортологи и паралоги
Обозначения блоков:
s – такие участки присутствуют по одному у всех изучаемых геномов
h – такие участки есть в части геномов
u – уникальные участки
r – повторы
m - минорные участки, длиной меньше 100 н.п., разделяющие остальные блоки

NPG-explorer



Литература (книга открыта)

Method	Category	Relationships predicted	Pairwise or multiple
BLAST	Local alignment	Homology	Pairwise
LASTZ	Local genomic alignment	Homology	Pairwise
MUMmer	Local genomic alignment	Orthology	Pairwise
CHAOS	Local genomic alignment	Homology	Pairwise
GRIMM-synteny	Orthology mapping	Toporthology	Multiple
DRIMM-synteny	Orthology mapping	Orthology, paralogy	Multiple
Mercator	Orthology mapping	Toporthology	Multiple
Enredo	Orthology mapping	Orthology, paralogy	Multiple
OSfinder	Orthology mapping	Orthology	Multiple
SuperMap	Orthology mapping	Orthology, paralogy	Multiple
progressiveMauve	Hierarchical WGA	Toporthology	Multiple
MUGSY	Hierarchical WGA	Toporthology	Multiple
MAVID	Global genomic alignment	Colinear homology	Multiple
LAGAN/Multi-LAGAN	Global genomic alignment	Colinear homology	Pairwise/multiple
DIALIGN	Global genomic alignment	Colinear homology	Multiple
SeqAn::T-Coffee	Global genomic alignment	Colinear homology	Multiple
FSA	Global genomic alignment	Colinear homology	Multiple
Pecan	Global genomic alignment	Colinear homology	Multiple
NUCmer/PROmer	Local WGA	Orthology	Pairwise
MULTIZ/TBA	Local WGA	Orthology, paralogy	Multiple
AXTCHAIN/CHAINNET	Alignment chaining and filtering	Orthology	Pairwise



Chapter 8

Whole-Genome Alignment

Colin N. Dewey