

Выравнивание геномов

А.Ал., 2021

Три понимания «правильного» выравнивания

1 **Оптимальное выравнивание:** наилучшее по весу

Его ищут программы.

Оптимальное выравнивание существует для любого набора последовательностей, даже негомологичных!

2 **Эволюционное выравнивание:** запись, отражающая ход эволюции

Не поддается достоверной реконструкции в большинстве реальных случаев; может отличаться от оптимального выравнивания.

Алгоритм вычисления веса стараются выбрать так, чтобы можно было ожидать, что эволюционное выравнивание будет среди нескольких оптимальных.

Для негомологичных последовательностей эволюционного выравнивания не существует!

3 **Функциональное выравнивание:** сопоставление функционально идентичных частей белков или нуклеиновых кислот

Объясняет сохранение в эволюции одних частей белка и варьирование других. Поскольку функция и 3D-структура белка очень тесно связаны, функционально выровненные аминокислотные остатки должны иметь примерно одинаковое расположение в пространстве.

1. Биологический смысл эволюционного выравнивания

- Локальные изменения генома – замены, маленькие делеции и вставки – в геноме потомка по сравнению с геномом предка – теоретически позволяют определить предковый нуклеотид у почти каждого нуклеотида потомка.
- Через геном общего предка можно установить гомологию нуклеотидов нескольких современных гомологичных последовательностей
- Примеры:
 - Реконструкция происхождения рукописей от разных переписчиков.
 - Реконструкция путей распространения SARS-CoV-2 по сходству геномов.
- Цель программы выравнивания ГОМОЛОГИЧНЫХ последовательностей поставить гомологичные нуклеотиды в одну колонку (эволюционное выравнивание).
- Программы ориентируются на сходство последовательностей
- Для геномов целиком обычные программы выравнивания не применимы потому, что в геноме потомка по сравнению с геномом предка могут произойти не локальные изменения

2. Проблемы эволюционного выравнивания геномов

Локальное или глобальное?

Локальное выдаст одно оптимальное выравнивание фрагментов, по одному из каждого генома

Глобальное, но геномы могут сильно отличаться в каких-то местах

.... не только это!

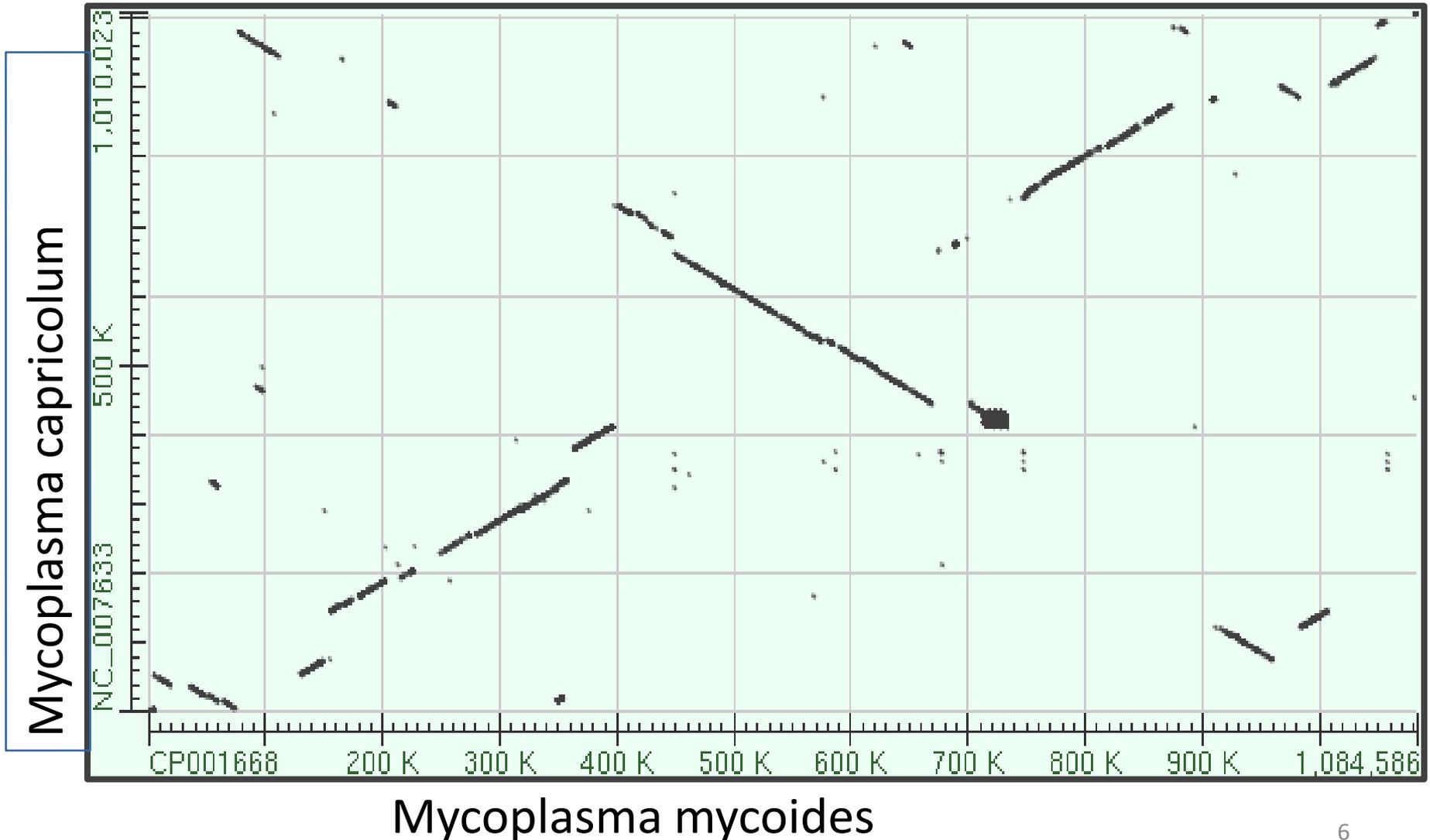
Идея: глобально-локальное выравнивание геномов! Т.е. **набор достоверных локальных выравниваний**

3. Парное выравнивание двух геномов: BLAST two sequences (blast2seq)

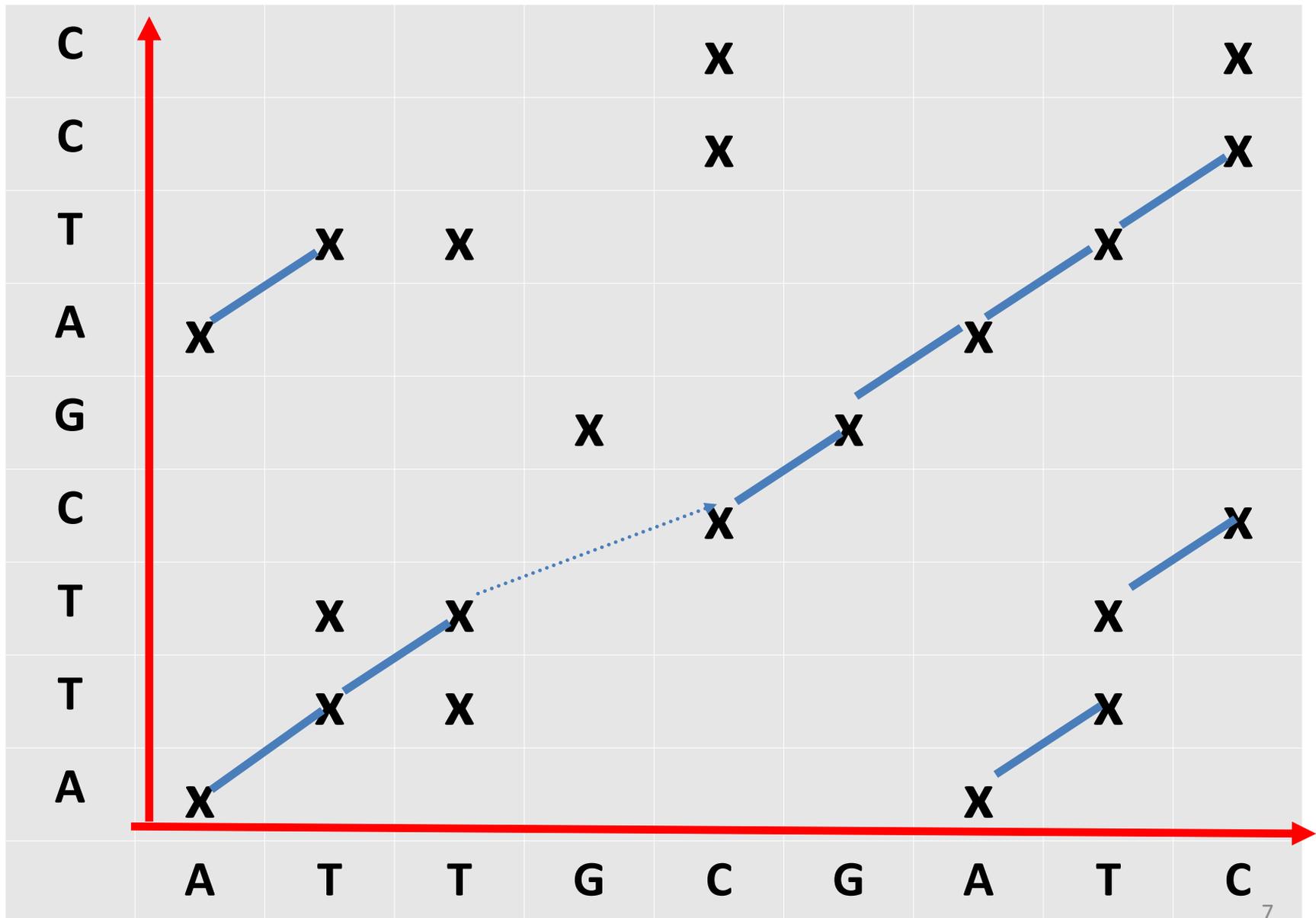
- Первая последовательность (subject) считается базой и индексируется на лету
- Вторая последовательность считается входной последовательностью (query)
- Применяется стандартный алгоритм Blast, стандартный интерфейс, стандартные параметры и стандартный формат выдачи
- Дополнительно выдается Карта локального сходства (Dot Plot).

4. Blast двух геномов бактерий

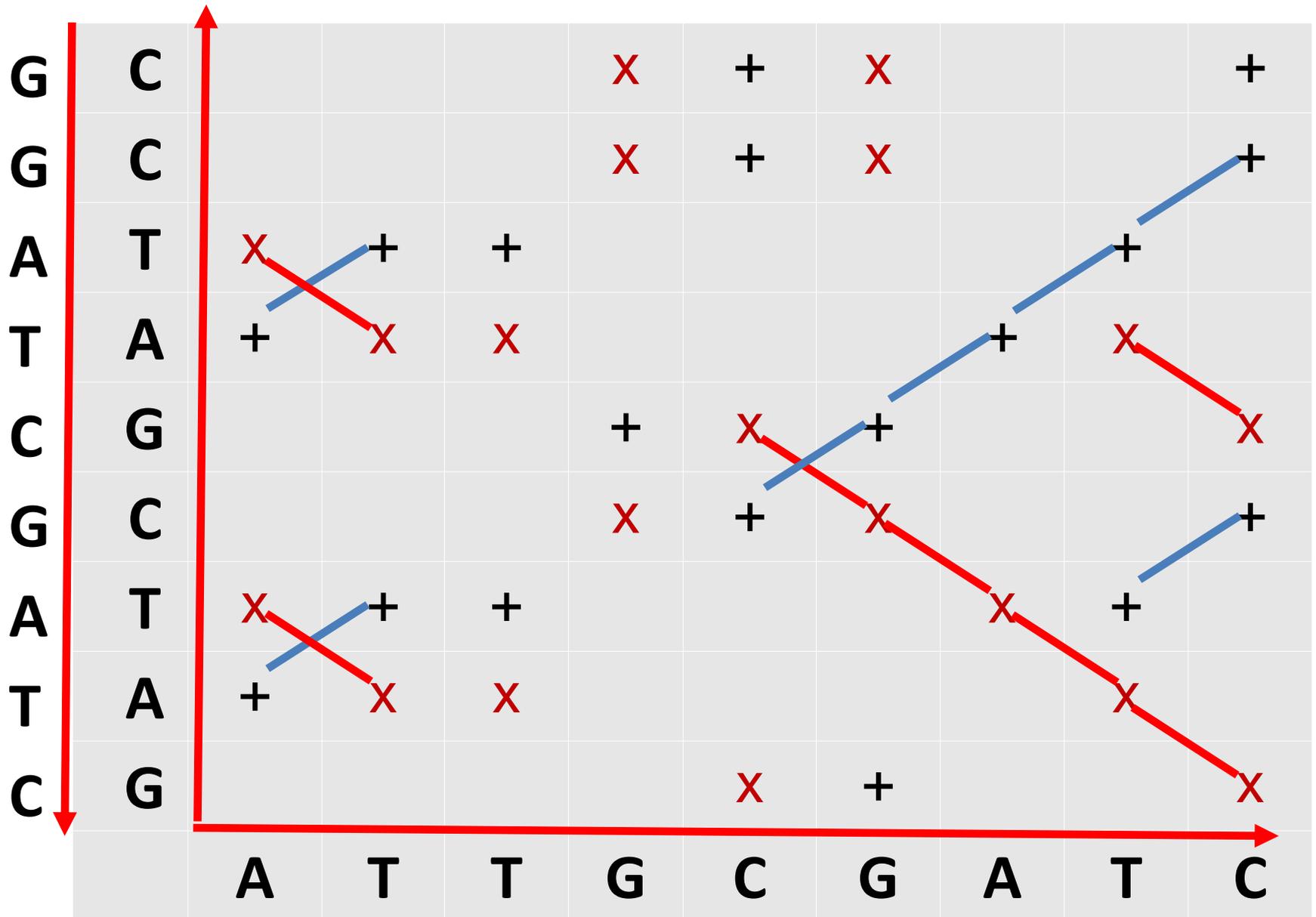
Карта локального сходства геномов *M. capricolum* и *M. mycoides*



Напоминалка про Карту локального сходства



Карта сходства с учетом комплементарной цепочки



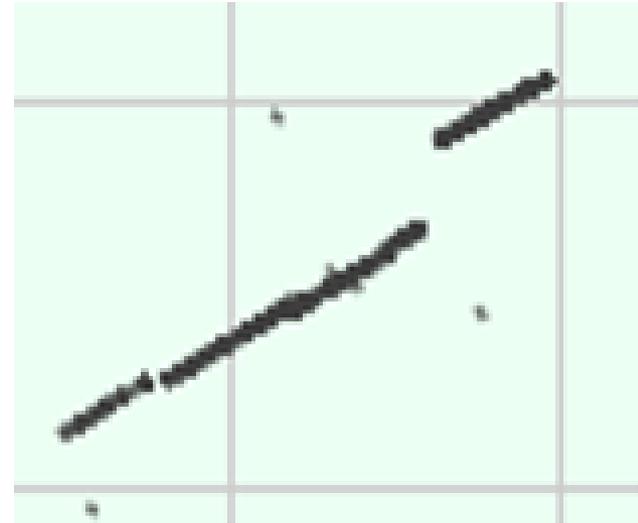
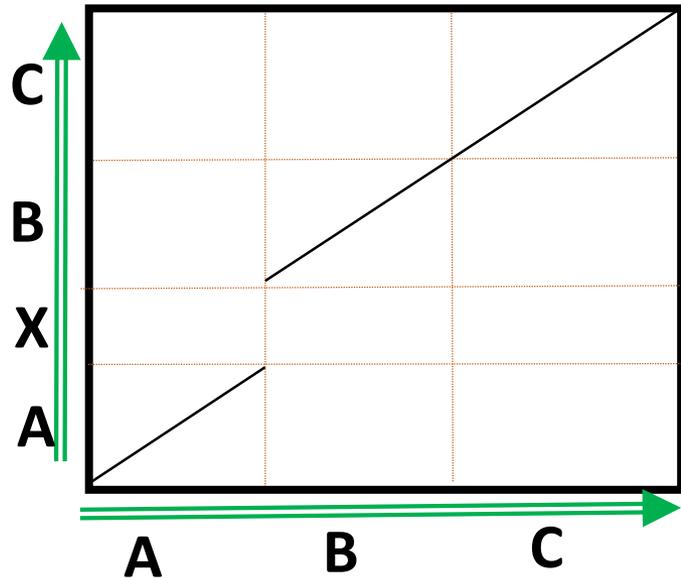
На карте blast2seq

- Отдельные буквы заменены на находки Blast
- Каждая находка изображена отрезком диагонали
- Диагональ
 - “/” - сходство subject (по оси X) с прямой цепью query
 - “\” – с комплементарной цепью query

Выдача blast2seq

- Состоит из множества парных локальных выравниваний: фрагмента 2го генома (query) с фрагментом 1го генома (subject)
Такие выравнивания называются блоками (парные в blast2seq)
- Можно скачать все такие выравнивания (в формате text ☹)
- Можно скачать таблицу блоков с указаниями координат фрагментов и параметров выравнивания

4. Крупные события в эволюции геномов

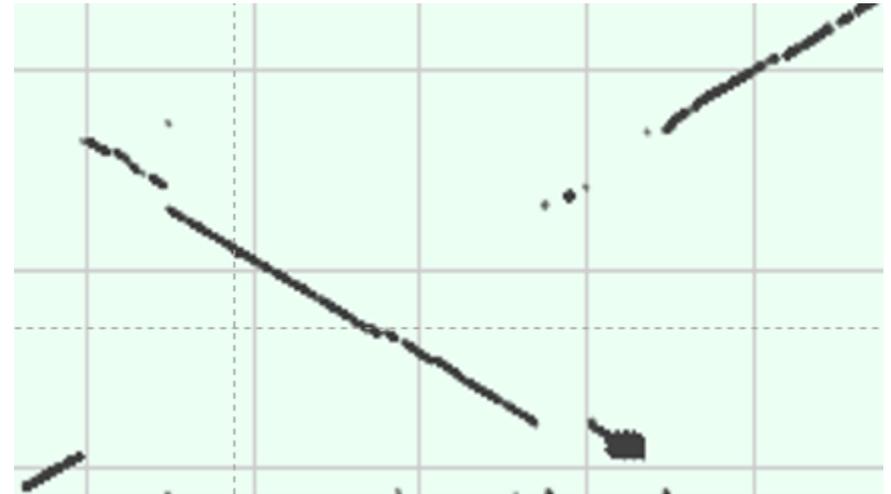
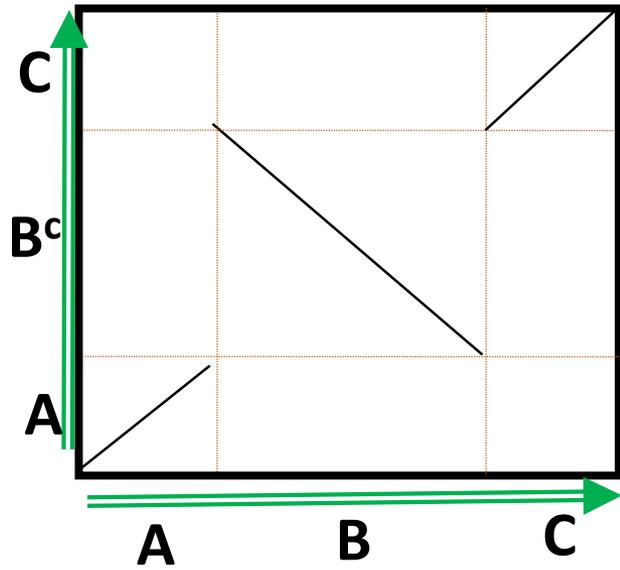


Индель – делеция в subject или вставка в query ?

Можно ли определить делеция или вставка, т.е. горизонтальный перенос из генома далекого организма (у прокариот и вирусов) ?

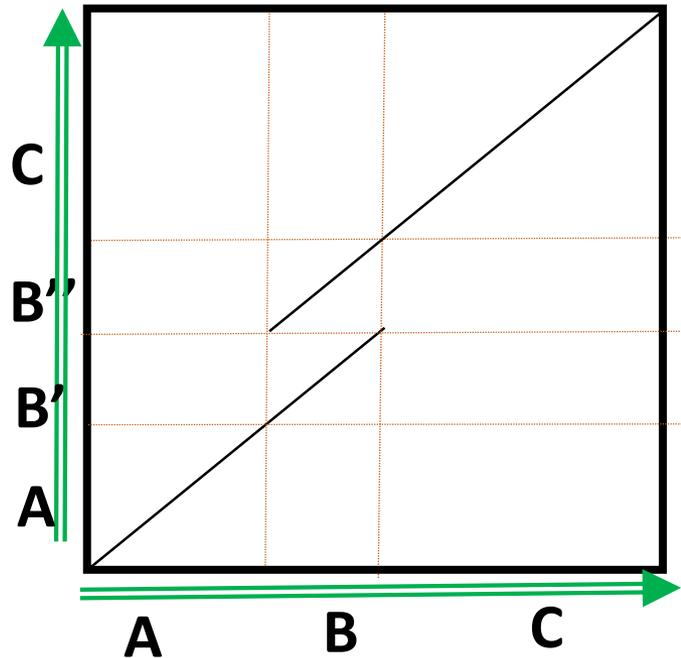
Да, если построить множественное выравнивание геномов

4. Крупные события в эволюции геномов

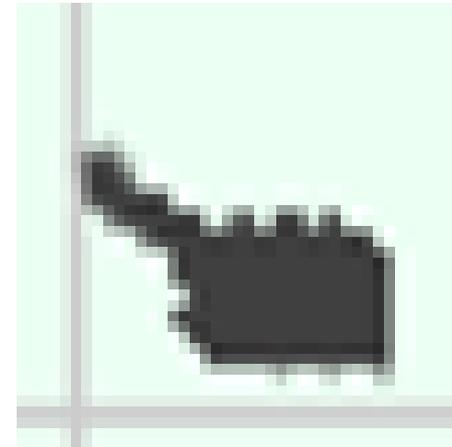


Инверсия

4. Крупные события в эволюции геномов

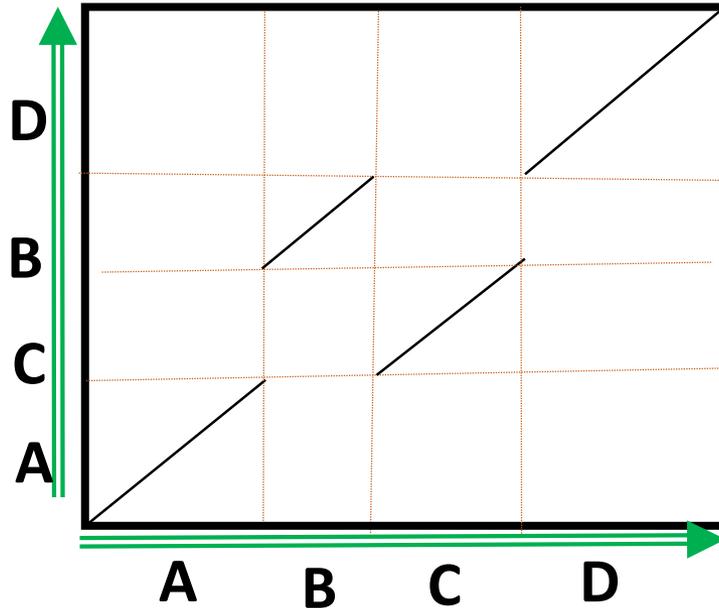


Дупликация



Последовательности малой сложности – результат множественных дупликация коротких фрагментов ДНК

4. Крупные события в эволюции геномов



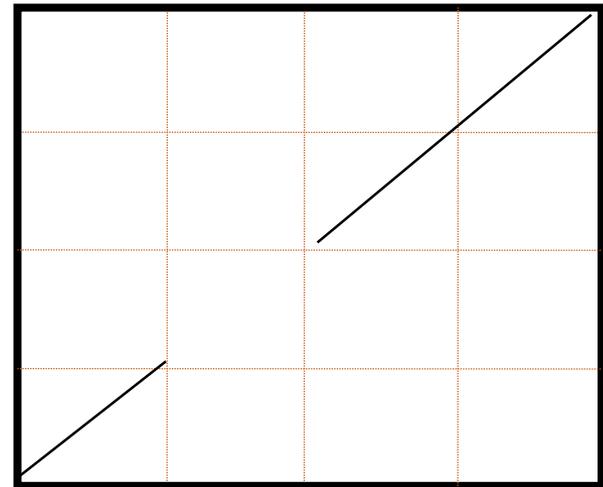
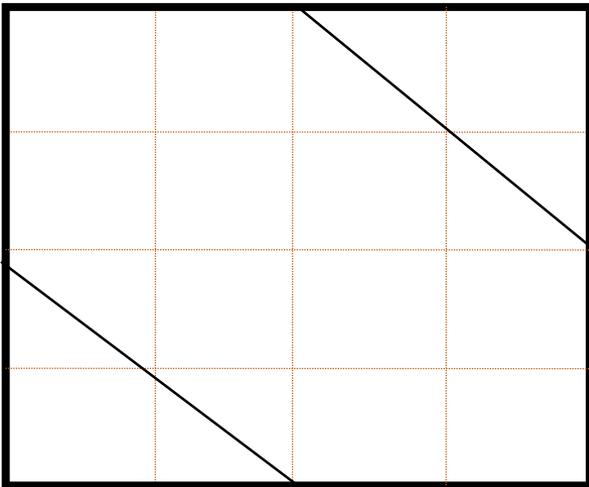
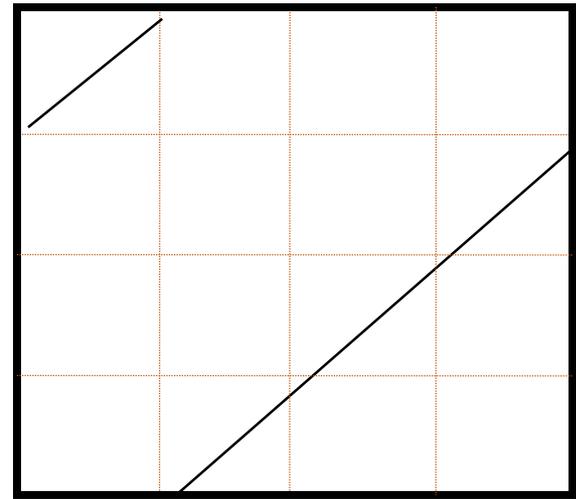
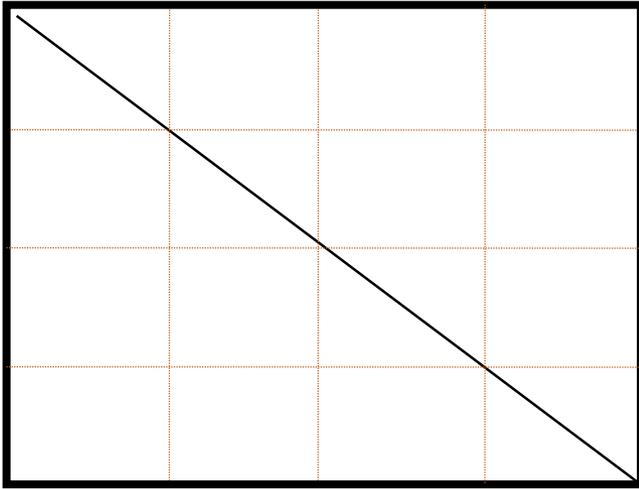
Транслокация

Транслокация может быть результатом нескольких инверсий:

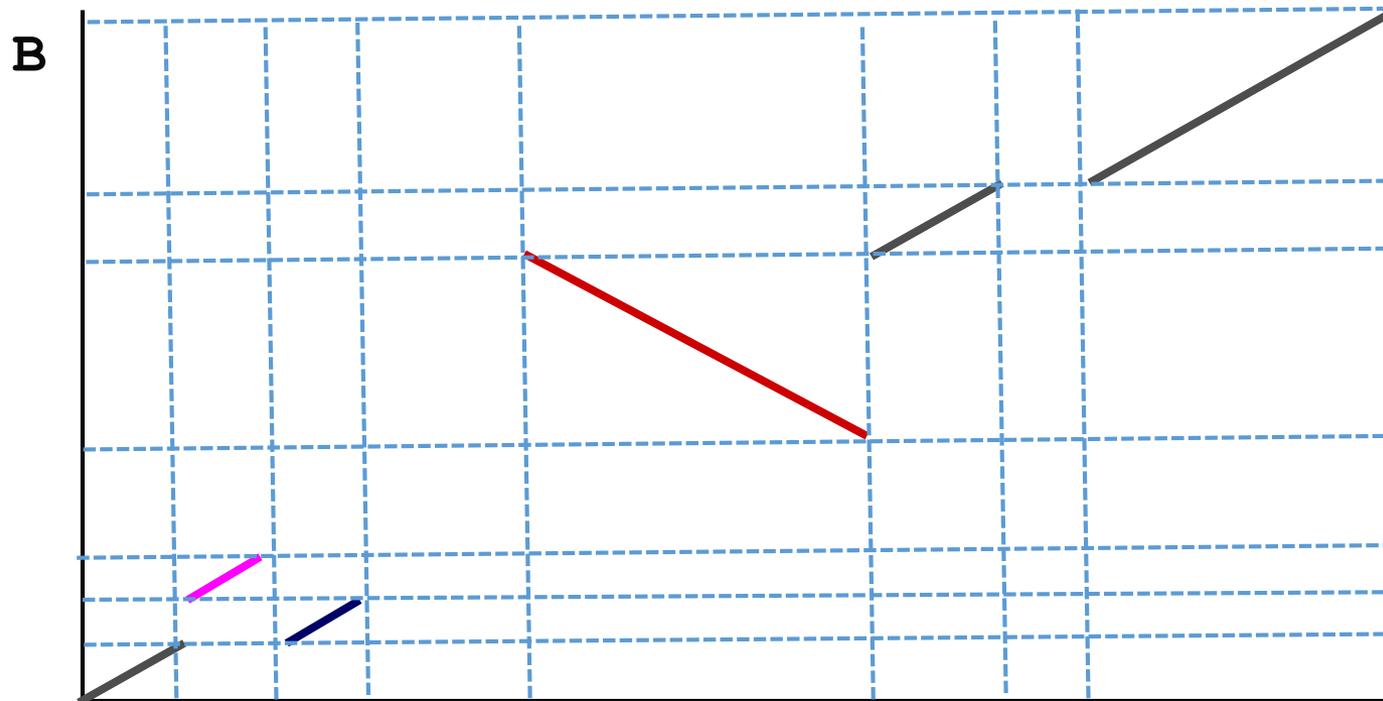
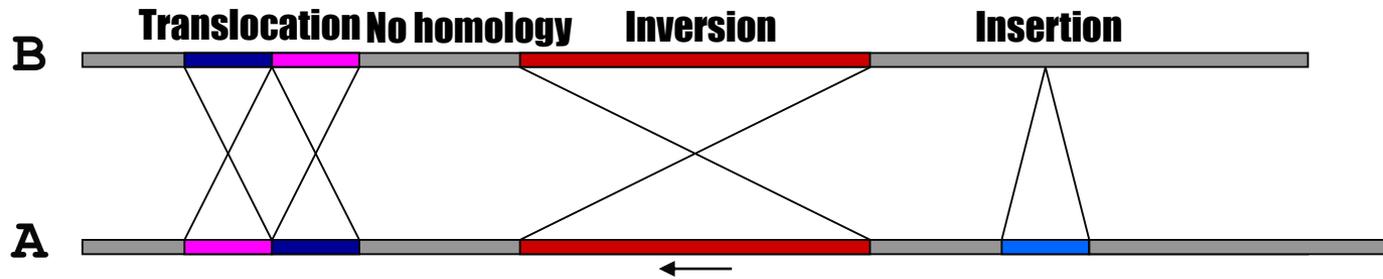
$\dots AB \dots \Rightarrow \text{inv}(AB) \Rightarrow \dots B^c A^c \dots \Rightarrow \text{inv}(B^c) \Rightarrow \dots B A^c \dots \Rightarrow \text{inv}(A^c) \Rightarrow \dots BA \dots$

Транслокации между ДНК одного генома – хромосомами, плазмидами.

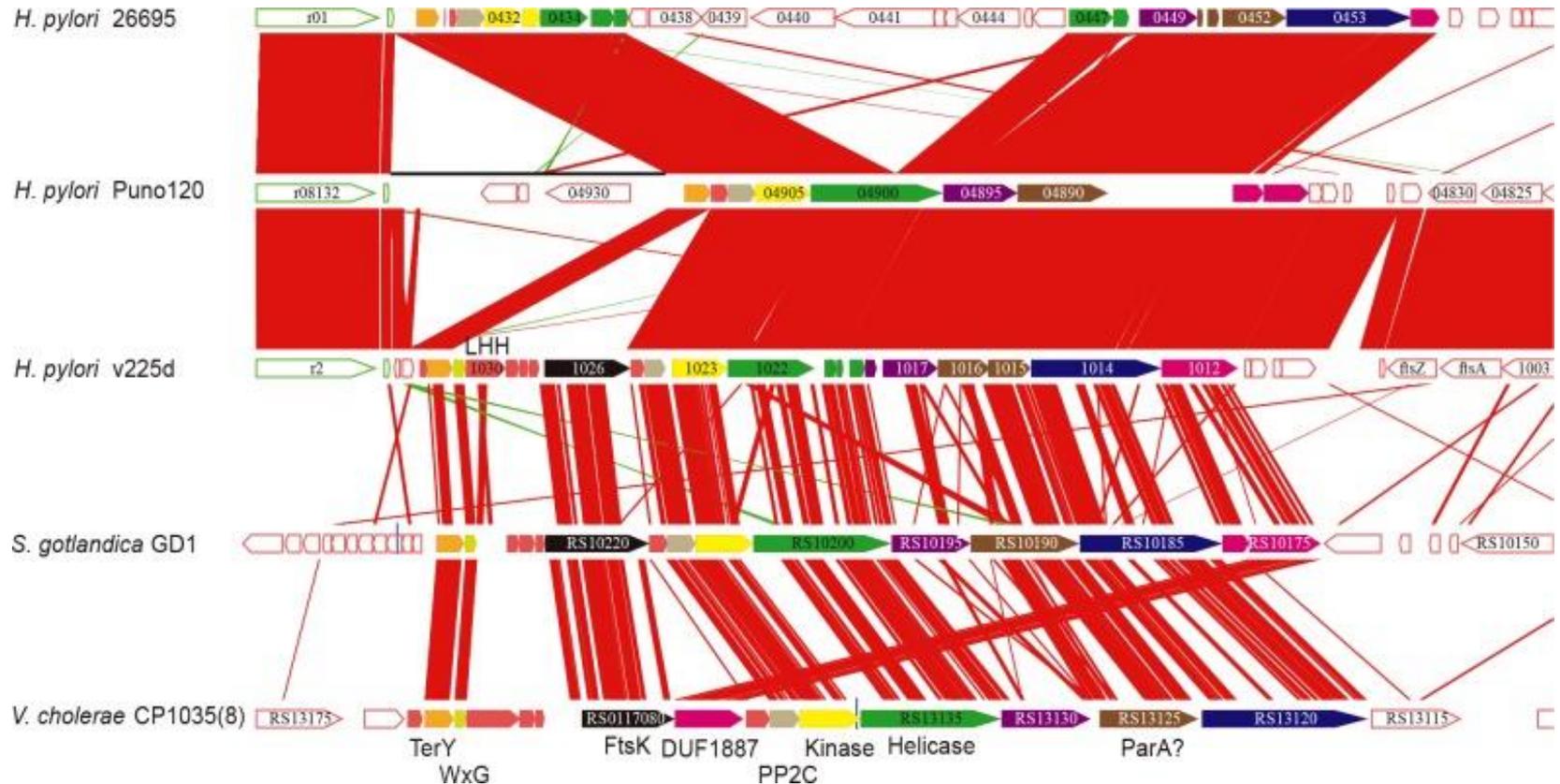
4. Крупные события в эволюции геномов?



5. Карту можно изобразить и так



Или так для нескольких геномов:
... не слишком понятно ☹️



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.

6. Множественное выравнивание геномов близкородственных организмов.

Нуклеотидный пангеном (НПГ)

- Какие организмы считать близкородственными?
Те, для которых гомологию фрагментов ДНК можно достоверно установить на основании сходства последовательностей. Ориентировочно, идентичность более 80 – 90%
- Для прокариот это примерно штаммы одного вида или, иногда, одного рода
- Почему только близкородственные геномы?
Потому то понятнее и задание будет по программе, предназначенной для такого случая.
Потому, что методы сравнения геномов более удаленных организмов дают менее точный ответ – находят **синтеничные участки**, на основании ортологичных генов белков; выравнивания геномов – не строят (?)

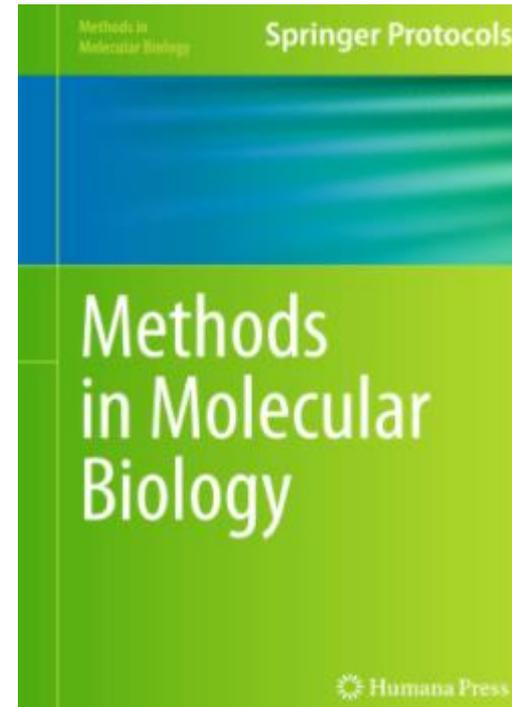
Множественное выравнивание геномов – нуклеотидный пангеном

- Вход:
 - несколько геномов
 - каждый геном представлен одной или несколькими хромосомами и плазмидами
 - вариант: набор контигов или скэффолдов
- Результат:
 - Набор блоков, т.е. выравниваний гомологичных участков ДНК из всех или части геномов
 - Каждая ДНК представлена последовательностью имен блоков

Определение выравнивания геномов.

Colin N. Dewey, 2012, переиздание 2019. Methods in Molecular Biology book series (MIMB, volume 1910) (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



Part II. Chapter 4 **Whole-Genome Alignment**

Colin N. Dewey

Иллюстрация геномного выравнивания

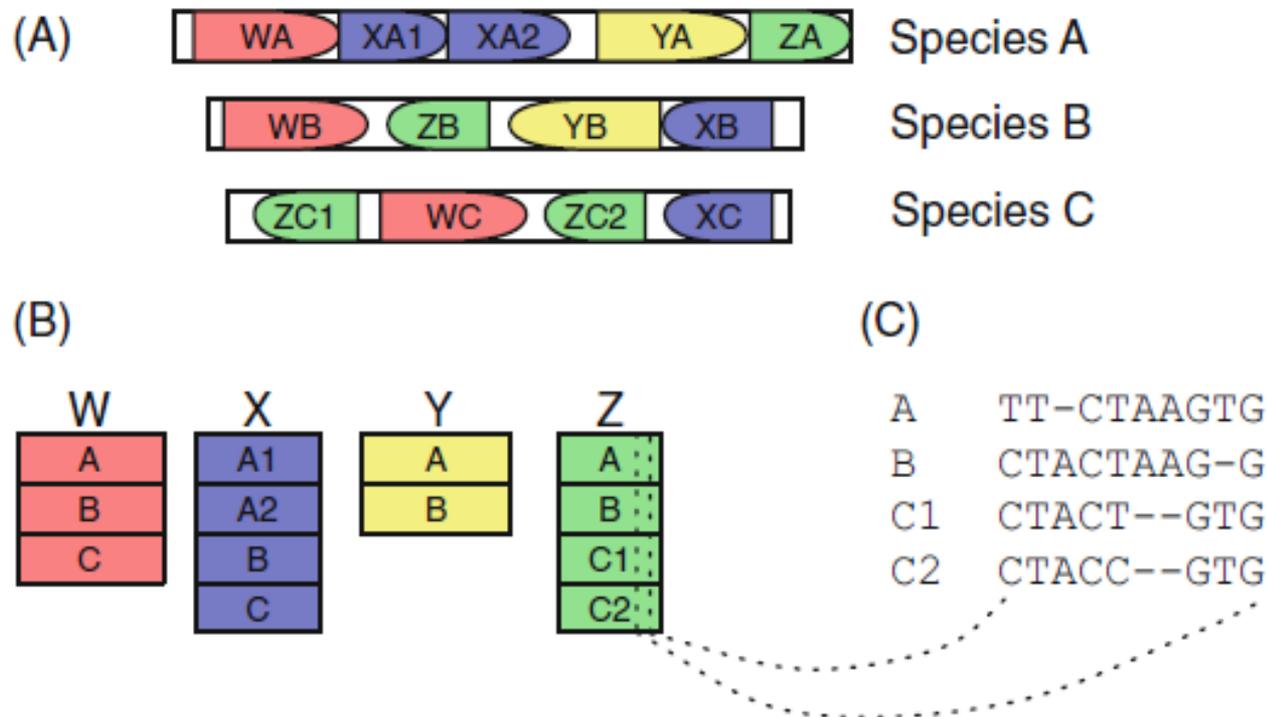


Fig. 2 An example WGA of three genomes represented as a set of alignment blocks. (a) The positions of the genomic segments that are in the alignment blocks are shown as shaded bullet-like shapes (the direction of the bullet indicates the orientation of the segment). In this example, not all genomic segments belong to a block (note the unshaded intervals). (b) The alignment blocks of the WGA. Note that blocks do not need to contain a segment from all genomes (e.g., block Y) and that some blocks can contain multiple segments from the same genome (e.g., blocks X and Z). (c) A slice of alignment block Z, which is a nucleotide-level alignment

Классификация блоков в программе NPG-explorer (Б.Нагаев, А.Алексеевский)

- Стабильные блоки (s-блоки): по одному фрагменту из каждого генома
- Полустабильные (h-блоки): по одному фрагменту из двух или более геномов, но не из всех геномов
- Уникальные последовательности (u-блоки): одна последовательность из одного генома, не имеющая ни одной похожей последовательности во входных данных
- Блоки с повторами (r-блоки): включают более одного фрагмента хотя бы из одного генома
Блоки выше называются мажорными
- Минорные блоки (m-блоки): короткие блоки со сходством менее порога, лежащие между двумя мажорными блоками (т.е. s, h, u, r блоками). Серая зона НПГ.

Все эти блоки вместе называются нормальными

Критерии нуклеотидного пангенома

- Параметры НПГ:
 - порог процент абсолютно консервативных (идентичных) позиций в каждом мажорном блоке (по умолчанию $\geq 90\%$)
 - порог длины (числа позиций) мажорного блока (по умолчанию ≥ 100 поз.)
- Все блоки в совокупности перекрывают полностью все входные последовательности без перекрытий
Значит, каждый геном можно представить как последовательность названий блоков
- В каждом мажорном блоке %идентичных позиций в скользящем окне фиксированной длины (те же 100 поз.) превышает порог (те же 90%)
- Результат blastn поиска консенсусов мажорных блоков “все против всех” не содержит находок, удовлетворяющих критериям мажорных блоков, приведенным выше

Визуализация НПГ в NPG-explorer

only blocks of >= 2 fragments Search: Clear Global blocks Normal blocks blockset alignment: g29x453820

	fr: ^	colur	% id	% GC	gene	split	low s	% lo
s29x3446	29	3446	96...	42...	0	0	2	5.45
s29x3443	29	3443	96...	38...	0	0	2	9.61
s29x3419	29	3419	90...	42...	0	0	4	23...
s29x3415	29	3415	91...	45...	0	0	9	39...
s29x3350	29	3350	90...	44...	0	0	5	54...
s29x3344	29	3344	98...	40...	0	0	1	0.53
s29x3337	29	3337	96...	32...	0	0	2	5.27
s29x3244	29	3244	93...	40...	0	0	1	9.86
s29x3176	29	3176	92...	43.5	0	0	4	45...
s29x3166	29	3166	93...	44...	0	0	5	47...
s29x3139	29	3139	92...	38...	0	0	2	29...
s29x3077	29	3077	97...	40...	0	0	1	2.59
s29x3066	29	3066	98...	34...	0	0	1	0.91
s29x3035	29	3035	93...	35...	0	0	2	48...
s29x3008	29	3008	96...	38...	0	0	3	5.48

	1	2	3	4	5	6	7	8
+SpnCGSP14&chr1&c	s29x3446 >	r60x101 >	m10x21 >	r16x101 >	m6x63 >	-	-	-
+SpnD39&chr1&c	s29x3446 >	-	-	-	-	-	-	-
+SpnG54&chr1&c	s29x3446 >	-	-	-	-	-	-	-
+SpnHungary19A6&chr1&c	s29x3446 >	r60x101 >	m10x21 >	r16x101 >	m6x63 >	-	-	-
+SpnINV104&chr1&c	s29x3446 >	-	m10x21 >	r16x101 >	-	m3x3 >	h10x105 >	-
+SpnINV200&chr1&c	s29x3446 >	r60x101 >	m10x21 >	r16x101 >	m6x63 >	-	-	-
+SpnJJA&chr1&c	s29x3446 >	-	m10x21 >	r16x101 >	m6x63 >	-	-	-
-SpnNCTC7465&chr1&c	s29x3446 >	-	-	-	-	-	-	-
+SpnNT11058&chr1&c	s29x3446 >	-	-	-	-	-	-	-
+SpnOXC141&chr1&c	s29x3446 >	-	m10x21 >	r16x101 >	m6x63 >	-	-	-
+SpnP1031&chr1&c	s29x3446 >	m6x95 >	-	-	-	-	-	-
+SpnR6&chr1&c	s29x3446 >	-	-	-	-	-	-	-
+SpnSNP034156&chr1&c	s29x3446 >	-	m10x21 >	r16x101 >	m1x28 >	r15x111 >	m1x15 >	-
+SpnSNP034183&chr1&c	s29x3446 >	-	m10x21 >	r16x101 >	m6x63 >	-	-	-
+SpnSNP994038&chr1&c	s29x3446 >	r60x101 >	m10x21 >	r16x101 >	-	m3x3 >	h10x105 >	-
+SpnSNP994039&chr1&c	s29x3446 >	-	m10x21 >	r16x101 >	-	m3x3 >	h10x105 >	-
+SpnSPN032672&chr1&c	s29x3446 >	-	-	-	-	-	-	-

	10	20	30	40	50	60	70	80	
Son6706B&chr1&c 84929 88366	A A C T C T A T C A G G A A A G T C A A A T T A A T T T A T A G A A A T A T T T T A G C A G T C A A G G T G T A C T A T T A T A G A T T C A A T A T A C T A T A T T A T	A A C T C T A T C A G G A A A G T C A A A T T A A T T T A T A G A A A T A T T T T A G C A G T C A A G G T G T A C T A T T A T A G A T T C A A T A T A C T A T A T T A T	A A C T C T A T C A G A A A A G T C A A A T T A A T T T A T A G A A A T A T T T T A G C A G T C A A G A T G T A C T A T T A T A G A T T C A A T A T G C T A T A T T A T	A A C T C T A T C A G G A A A G T C A A A T T A A T T T A T A G A A A T A T T T T A G C A G T C A A G G T G T A C T A T T A T A G A T T C A A T A T A C T A T A T T A T	A A C T C T A T C A G G A A A G T C A A A T T A A T T T A T A G A A A T A T T T T A G C A G T C A A G G T G T A C T A T T A T A G A T T C A A T A T A C T A T A T T A T	A A C T C T A T C A G G A A A G T C A A A T T A A T T T A T A G A A A T A T T T T A G C A G T C A A G G T G T A C T A T T A T A G A T T C A A T A T A C T A T A T T A T	A A C T C T A T C A G G A A A G T C A A A T T A A T T T A T A G A A A T A T T T T A G C A G T C A A G G T G T A C T A T T A T A G A T T C A A T A T A C T A T A T T A T	A A C T C T A T C A G G A A A G T C A A A T T A A T T T A T A G A A A T A T T T T A G C A G T C A A G G T G T A C T A T T A T A G A T T C A A T A T A C T A T A T T A T	A A C T C T A T C A G G A A A G T C A A A T T A A T T T A T A G A A A T A T T T T A G C A G T C A A G G T G T A C T A T T A T A G A T T C A A T A T A C T A T A T T A T

Здесь есть гены

NPG-explorer

only blocks of >= 2 fragments Search: s10x5

Global blocks Normal blocks blockset alignment: i3x37174

fragment	columns	% identit	% GC	genes	split part	low simil	% low sir
s10x57079	10	57079	99.32	34.91	676	0	0
s10x5358	10	5358	98.71	34.22	86	0	1
s10x5561	10	5561	97.49	37.06	73	0	1
s10x5767	10	5767	98.76	34.6	72	0	0
s10x5289n1	10	5289	98.26	35.32	72	0	1
s10x5289	10	5289	94.26	38.1	59	0	8
s10x5500	10	5500	98.3	37.29	49	0	0
s10x5478	10	5478	98.46	34.71	76	0	1
s10x5820	10	5820	98.13	37.28	58	0	1
s10x5028	10	5028	96.38	37.59	48	0	2

show genes

	50	51	52
-SicHVF10&chr&c	-	-	-
+SicI Δ1141&chr&c	-	-	-
+SicRFV15Δ&chr&c	r5x106 >	r66x110 <	m36x1 >

	11570	11580	11590	11600	11610	11620	11630
	GAGGCCATGCTAGGAGTTAA	CATTTAATTATT	GCTAATTCGCTTCAGTAAAT	ATGATTAACCGTTT			
SicHVF10&chr&c 1248336 1305402	GAGGCCATGCTAGGAGTTAA	CATTTAATTATT	GCTAATTCGCTTCAGTAAAT	ATGATTAACCGTTT			
SicI Δ1141&chr&c 1104881 1161951	GAGGCCATGCTAGGAGTTAA	CATTTAATTATT	GCTAATTCGCTTCAGTAAAT	ATGATTAACCGTTT			
SicI10&chr&c 1242660 1299771	GAGGCCATGCTAGGAGTTAA	CATTTAATTATT	GCTAATTCGCTTCAGTAAAT	ATGATTAACCGTTT			
SicI Δ215&chr&c 1300967 1358034	GAGGCCATGCTAGGAGTTAA	CATTTAATTATT	GCTAATTCGCTTCAGTAAAT	ATGATTAACCGTTT			
SicM1425&chr&c 1231457 1288522	GAGGCCATGCTAGGAGTTAA	CATTTAATTATT	GCTAATTCGCTTCAGTAAAT	ATGATTAACCGTTT			
SicM1627&chr&c 1311541 1368604	GAGGCCATGCTAGGAGTTAA	CATTTAATTATT	GCTAATTCGCTTCAGTAAAT	ATGATTAACCGTTT			
SicM164&chr&c 1272229 1329294	GAGGCCATGCTAGGAGTTAA	CATTTAATTATT	GCTAATTCGCTTCAGTAAAT	ATGATTAACCGTTT			
SicRFV15Δ&chr&c 1117131 1174197	GAGGCCATGCTAGGAGTTAA	CATTTAATTATT	GCTAATTCGCTTCAGTAAAT	ATGATTAACCGTTT			
SicV(5/14)&chr&c 1201137 1258205	GAGGCCATGCTAGGAGTTAA	CATTTAATTATT	GCTAATTCGCTTCAGTAAAT	ATGATTAACCGTTT			

Визуализация НПГ (qnrge)

- (Окно ВерхЛев = ВЛ) Таблица всех блоков

Пользователь задает сопоставление ДНК из разных геномов. Например, ДНК названные chr1, плазмиды названные plasmidA и т.п.

- (Окно ВПм, маленькое) указание имени ДНК для визуализации
- (Окно ВП, под ВПм) Для указанной ДНК: представление ДНК каждого генома в виде последовательности имен блоков и выравнивание этих последовательностей
- (Окно Нижнее) Выравнивание последовательностей выделенного блока
 - Выделение аннотированных генов в этом выравнивании (белым шрифтом, подчеркивание указывает, что ген на обратной цепи)
- (Узкое окно между Верхн и Нижн): название выделенного гена
 - Переходы по генам в строке Нижнего окна: Ctrl → и Ctrl ←

Синтеничные участки по генам белков

- Синтения – неологизм (*John Renwick, 1971*).
 - Greek: *σύν* – вместе x *ταινία, tainiā* - лента
 - Изобразим гены в геноме стрелочками. Ортологичные гены пометим одинаковым цветом или еще как-нибудь (одинаковым номером)
 - Предположим, в геноме 1 и 2 найдутся одинаковые последовательности генов. Тогда соответствующие участки геномов называются синтеничными. Конечно, берутся максимальные такие последовательности.
- Предполагается, что синтеничные участки эволюционировали локально, без крупных эволюционных событий.
 - Скорости эволюции генов и межгенников различны.

Синтении в НПГ, построенном с помощью NPG-explorer, g-блоки

- g-блоком называется максимальный по ширине блок, начинающийся с s-блока и заканчивающийся s-блоком, такой, что последовательности s-блоков внутри него идут в одинаковом порядке во всех геномах. Внутри g-блока разрешены блоки других типов.
- Участки между g-блоками объединяются в промежуточные (intermediate) блоки (i-блоки). Так же, как минорные блоки между мажорными блоками.
- В окне qnrge для указания ДНК можно выбрать g-блок

Пакет NPG-explorer

- Начало работы
 - Скачать и установить пакет
 - Создать директорию для НПГ
 - В ней создать файл `genomes.tsv` с информацией о геномах
- Можно запускать программы пакета
 - Открыть командную строку
 - `npge Prepare`
 - `npge Examine`
 - [изменить параметры]
 - `npge MakePangenome`
 - [`npge CheckPangenome`]
 - `npge PostProcessing`

genomes.tsv для 10 геномов археи *Sulfolobus islandicus*

all:embl:CP002426.1	SisHVE10	chr	c	Sulfolobus islandicus HVE10/4
all:embl:CP001731.1	SisLD85	chr	c	Sulfolobus islandicus L.D.8.5
all:embl:CP001732.1	SisLD85	pl	c	Sulfolobus islandicus L.D.8.5
all:embl:CP001399.1	SisLS215	chr	c	Sulfolobus islandicus L.S.2.15
all:embl:CP003928.1	SisLAL141	chr	c	Sulfolobus islandicus LAL14/1
all:embl:CP001400.1	SisM1425	chr	c	Sulfolobus islandicus M.14.25
all:embl:CP001401.1	SisM1627	chr	c	Sulfolobus islandicus M.16.27
all:embl:CP001402.1	SisM164	chr	c	Sulfolobus islandicus M.16.4
all:embl:CP002425.1	SisREY15A	chr	c	Sulfolobus islandicus REY15A
all:embl:CP001403.1	SisYG5714	chr	c	Sulfolobus islandicus Y.G.57.14
all:embl:CP001404.1	SisYN1551	chr	c	Sulfolobus islandicus Y.N.15.51
all:embl:CP001405.1	SisYN1551	pl	c	Sulfolobus islandicus Y.N.15.51

Выходные файлы в создаваемых NPG-explorer директориях

..		Up
genes		Folder
mutations		Folder
global-blocks		Folder
extra-blocks		Folder
trees		Folder
pangenome		Folder
check		Folder
Sulfolobus_islandicus_10_12	log	21198
genomes-renamed	fasta	26 M
genomes	tsv	799

pangenome/pangenome.info

про блоки всех типов

- **Stable blocks** (represented in all genomes once, but not minor):

fragments: 3640

blocks: 364

fragments / blocks: 10

Block identity:

min: 0.896

median: 0.9749

avg: 0.9686

max: 1

identity of joined blocks: 0.982929

Block sizes: 10

.....

Nucleotides in blocks: 19533363

The percentage of input length: 73.5%

/rangenome.bi список всех блоков с информацией о них

block	fragments	cols	ident	GC	SisHVE10	SisLAL141	SisLD85	SisLS215	SisM1425	SisM1627	SisM164	SisREY15A	SisYG5714	SisYN1551
r66x110	66	110	0.9181	0.4161	8	3	14	13	0	0	0	19	5	4
s10x85594	10	85594	0.9907	0.3439	1	1	1	1	1	1	1	1	1	1
s10x124n2	10	124	0.9919	0.425	1	1	1	1	1	1	1	1	1	1
s10x210	10	210	0.9476	0.3314	1	1	1	1	1	1	1	1	1	1
m10x3	10	3	0.6666	0.4333	1	1	1	1	1	1	1	1	1	1
h9x10253	9	10253	0.9801	0.3574	1	1	1	1	1	1	1	1	0	1
h8x648	8	648	0.9197	0.3878	1	1	1	1	1	0	1	1	0	1
r8x132	8	132	0.9128	0.4182	0	0	3	1	0	0	0	0	1	3
u1x28840	1	28840	10.3411		1	0	0	0	0	0	0	0	0	0
u1x131	1	131	10.4274		0	0	0	0	0	0	1	0	0	0

Что есть в выходных данных NPG-explorer

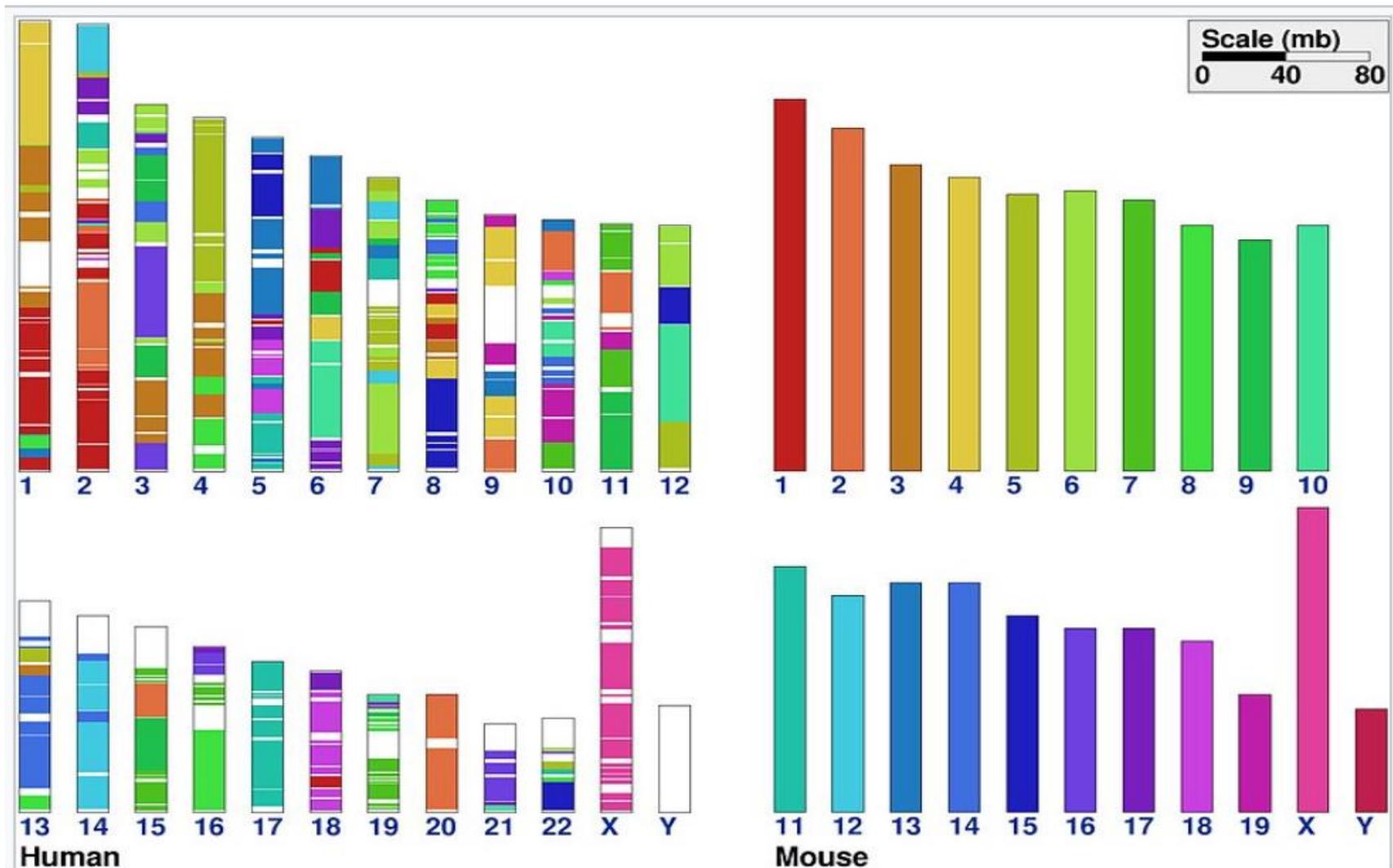
<code>pangenome/ pangenome.blocks</code>	G-блоки и I-блоки как последовательности нормальных блоков
<code>global-blocks/ blocks.blocks</code>	Хромосомы как последовательности g-блоков и i-блоков
<code>pangenome/ fragments.tsv</code>	Координаты в геноме всех фрагментов всех нормальных блоков
<code>global-blocks/ global-fragments.tsv</code>	Координаты в геноме всех фрагментов всех g-блоков и i-блоков
<code>pangenome/ pangenome.bs</code>	Выравнивания всех блоков
<code>genes/</code>	Координаты всех генов из всех геномов в блоках и в геномах
<code>trees/</code>	nj-дерево геномов, построенное по объединенному выравниванию s-блоков
<code>mutations/</code>	Файл с консенсусами всех блоков и файл с описанием всех мутаций во всех блоках

Как найти с помощью NPG-explorer крупные эволюционные события в эволюции геномов?

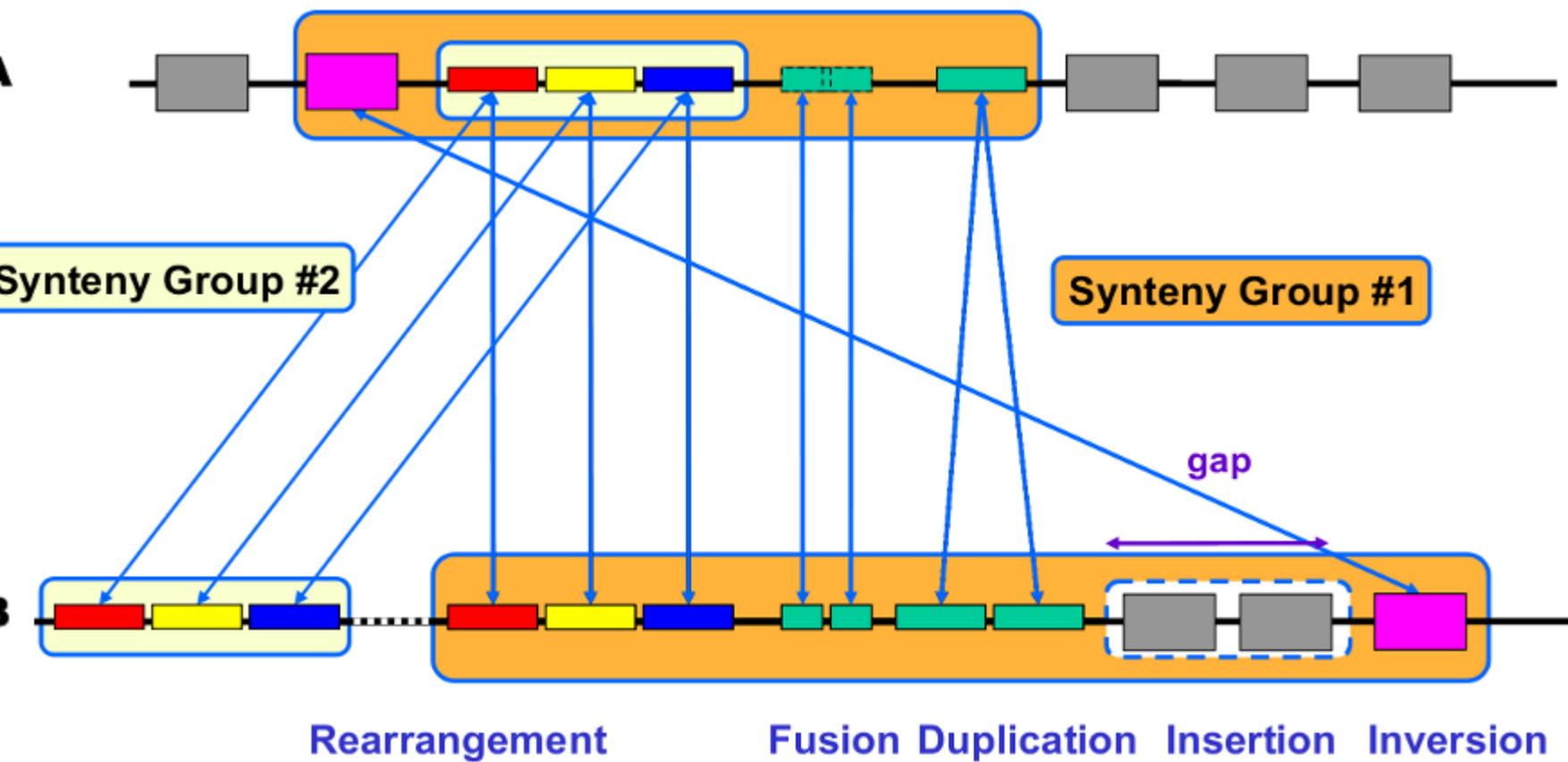
- Делеция большого участка ДНК
- Вставка участка ДНК за счёт
 - либо горизонтального переноса
 - либо дубликации участка собственной ДНК
- Инверсия участка ДНК
- Транслокация участка ДНК с инверсией или без неё
- Перестановки синтеничных участков

КОНЕЦ

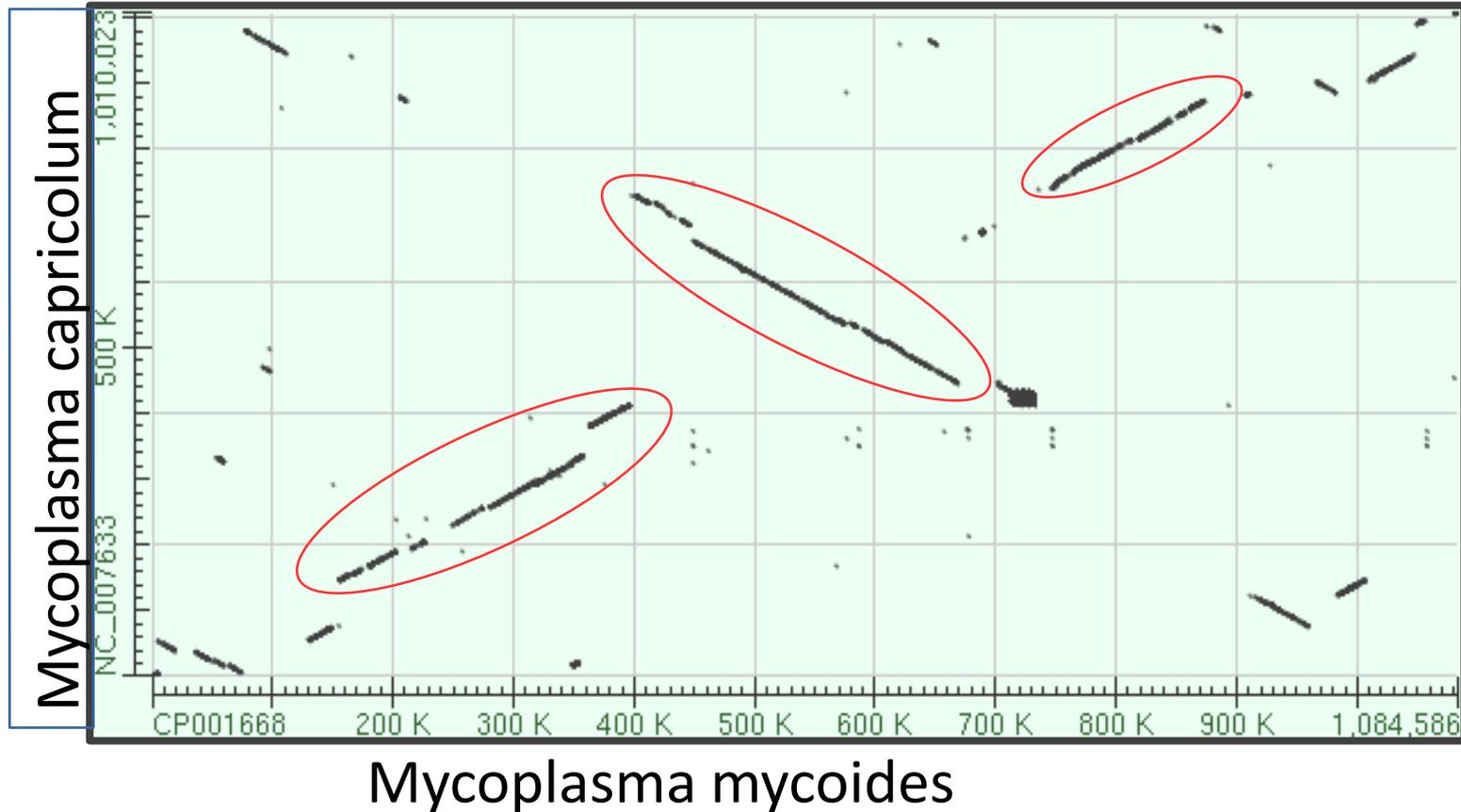
Синтении между геномами человека и мыши



Synteny between human and mouse chromosomes. Colors indicate homologous regions. For instance, sequences homologous to mouse chromosome 1 are primarily on human chromosomes 1 and 2, but also 6, 8, and 18. The X chromosome is almost completely syntenic in both species.^[1]



Синтении можно найти в НПГ



Синтения - блоки НПГ, идущие подряд во всех геномах.

Какие-то перестройки внутри синтеничных последовательностей могут наблюдаться.

Но в целом структура генома внутри синтеничных участков сохраняется.

8. Пангеном (различать с НПГ!)

Пангеномом называют совокупность всех генов белков, закодированных в наборе геномов данного таксона, разделенных на группы ортологичных белков.

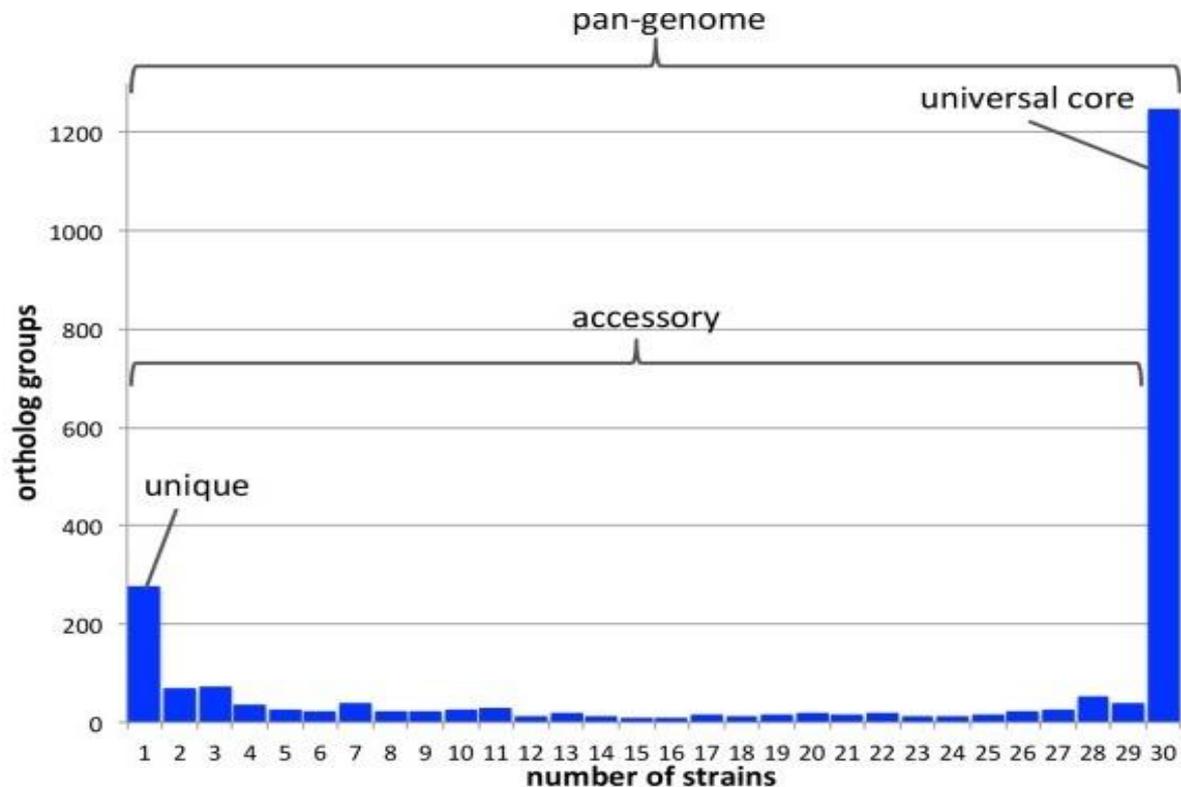
Белки пангенома делят на

- универсальные, встречающиеся во всех гномах
- специфические для штамма или вида, встречающиеся в одном виде
- периферия, белки, встречающиеся во многих геномах но не во всех

Будет, я думаю, в курсе М.Гельфанда (8й семестр, пока)

Концепция пангенома позволяет глубже погрузиться в эволюцию, чем НПГ, так как последовательности белков консервативнее нуклеотидных последовательностей

Пангеном по белкам 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.