



Секвенирование генома

(с) Г.Е. Генихович "Геном и эмбриональное развитие"

«Геном и эмбриональное развитие»

Лекция 2

Большая просьба

- Выключите, пожалуйста, звук у Ваших телефонов
- Если что-то непонятно, сразу поднимайте руку и останавливайте меня. А то потом забудем, что было неясно. Если вопрос «длинный» - подходите в перерыве или после лекции.

(с) Г.Е. Генихович "Геном и эмбриональное развитие"

Больших успехов достигла американская фирма «Celera Genomics», секвенировавшая геном человека. Оказалось, он состоит всего из четырёх нуклеотидов: А, С, G и Т. Теперь осталось только установить, в каком порядке они следуют.

Новости на НТВ (неточная цитата)

(с) Г.Е. Генихович "Геном и эмбриональное развитие"

Больших успехов достигла американская фирма «Celera Genomics», секвенировавшая геном человека. Оказалось, он состоит всего из четырёх нуклеотидов: А, С, G и Т. Теперь осталось только установить, в каком порядке они следуют.

Новости на НТВ (неточная цитата)

(с) Г.Е. Генихович "Геном и эмбриональное развитие"



Выбор организма для секвенирования

- Модельный объект (имеющиеся ресурсы, EST-проект)
- Важность для здравоохранения и/или экономики
- Филогенетически важное положение
(с) Г.Е. Генихович "Геном и эмбриональное развитие"
- Метагеномика – секвенирование групп организмов из различных биотопов (почва, нефть и т.п.)

Сэнгеровский метод на нуклеотидах-терминаторах

dATP, dCTP, dGTP, dTTP + ddATP, ddCTP, ddGTP, ddTTP



ПЦР с одним праймером



Электрофорез

(с) Г.Е. Генихович "Геном и эмбриональное развитие"

Т А Т С С G G Т А А С Т А Т С G Т С Т Т G А G Т С С А А

— А*

_____ А*

_____ А*

_____ А*

_____ А*

_____ А*

_____ А*

Сэнгеровский метод на нуклеотидах-терминаторах

dATP, dCTP, dGTP, dTTP + ddATP, ddCTP, ddGTP, ddTTP

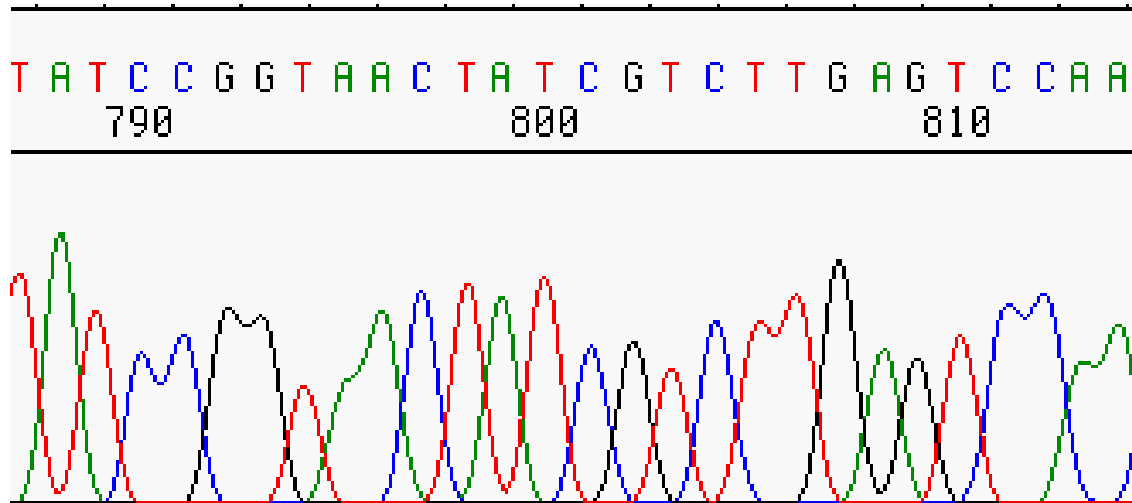


ПЦР с одним праймером



Электрофорез

(с) Г.Е. Генихович "Геном и эмбриональное развитие"



EST – expressed sequence tags

Выделение мРНК, синтез двуцепочечной кДНК



(с) Г.Е. Генихович "Геном и эмбриональное развитие"
кДНК-библиотека



Секвенирование вставок с двух сторон

Shotgun genome sequencing

Фрагментирование геномной ДНК, создание библиотеки



Секвенирование множества перекрывающихся кусочков (reads), <1kb



(с) Г.Е. Генихович "Геном и эмбриональное развитие"
Восстановление непрерывных последовательностей (contigs)
путём складывания перекрывающихся кусочков вместе



Определение более длинных последовательностей (scaffolds)
на основе сиквенса концов ВАС-библиотек или фосмид



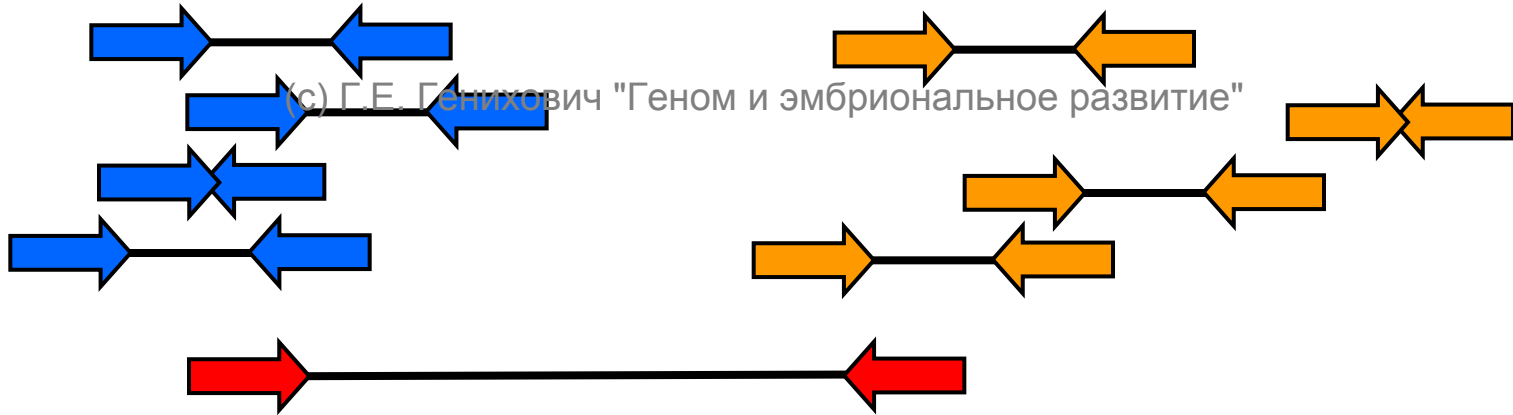
Физическое картирование генома

Shotgun genome sequencing

Scaffold

Contig1

Contig2



Read



Фрагмент



Несеквенированная последовательность примерно известной длины

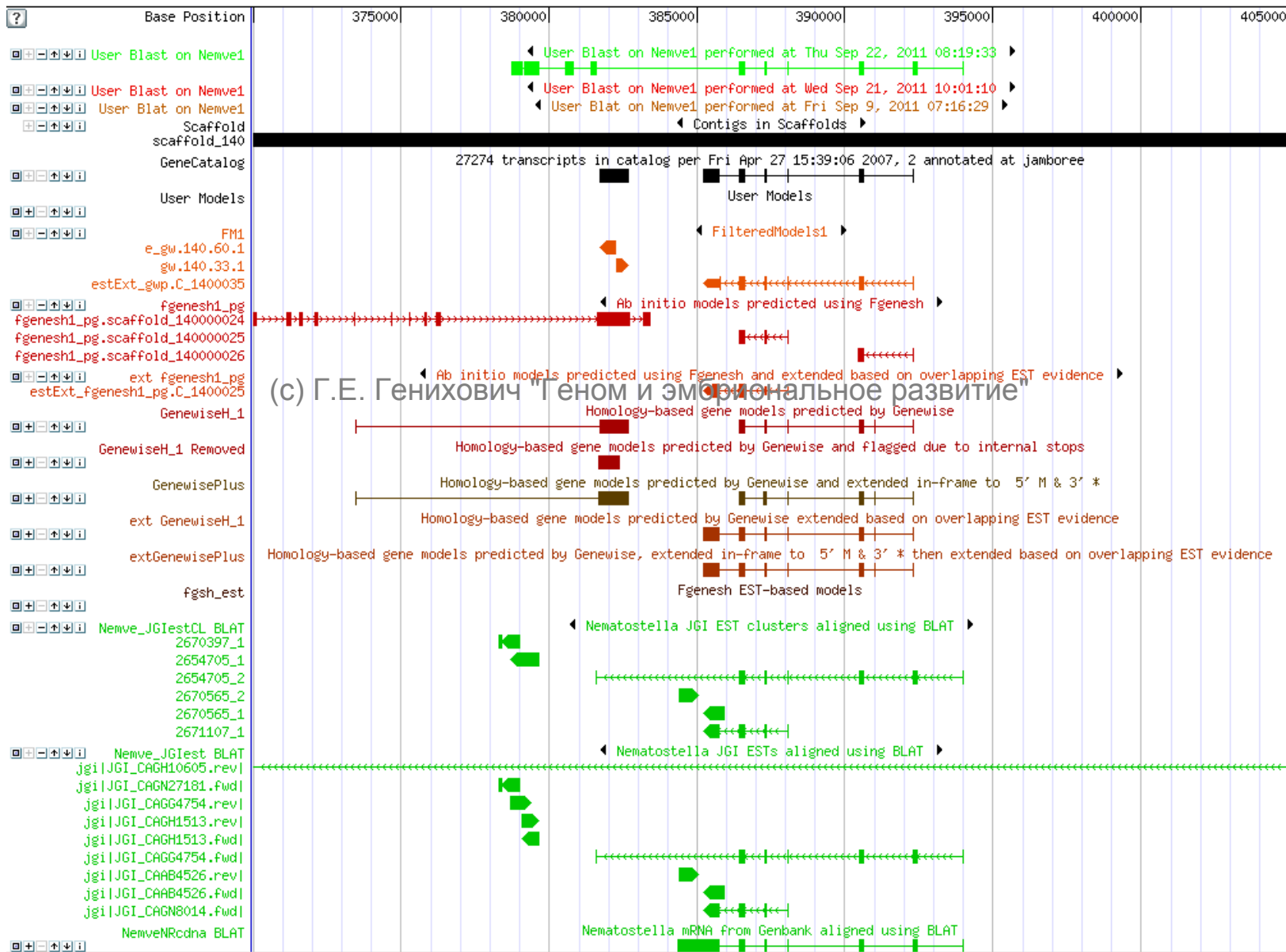
Position

scaffold_140:370000-405000

Zoom

General

Size: 35001 Feature: 385206-388121 (-) No hit



(с) Г.Е. Генихович "Геном и эмбриональное развитие"

Новое поколение секвенаторов: sequencing by synthesis

Solexa/Illumina:

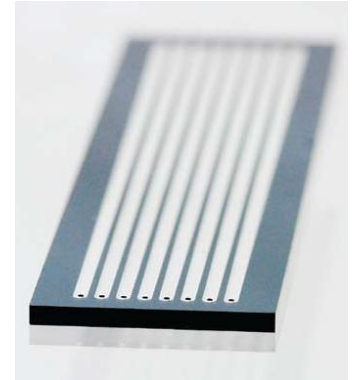
Десятки миллионов секвенсов за один заход.

Длина - 25-100 bp.

+ Много данных, отлично подходит для ChIP-Seq и RNA-Seq (если секвенирован геном)

- Сложно складывать маленькие кусочки

(с) Г.Е. Генихович "Геном и эмбриональное развитие"

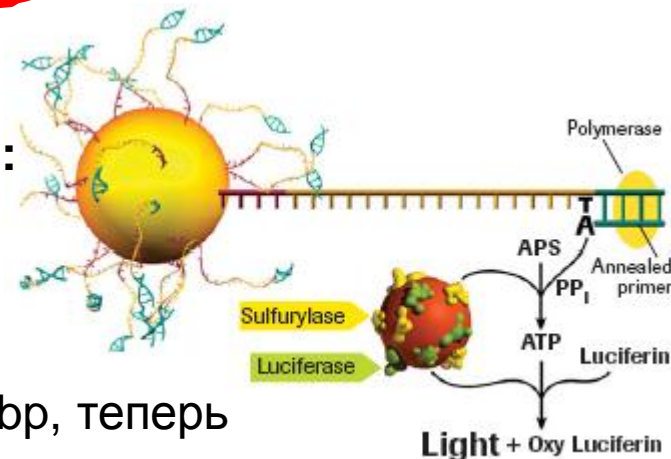


Пиросеквенирование – 454 Life Sciences/Roche:

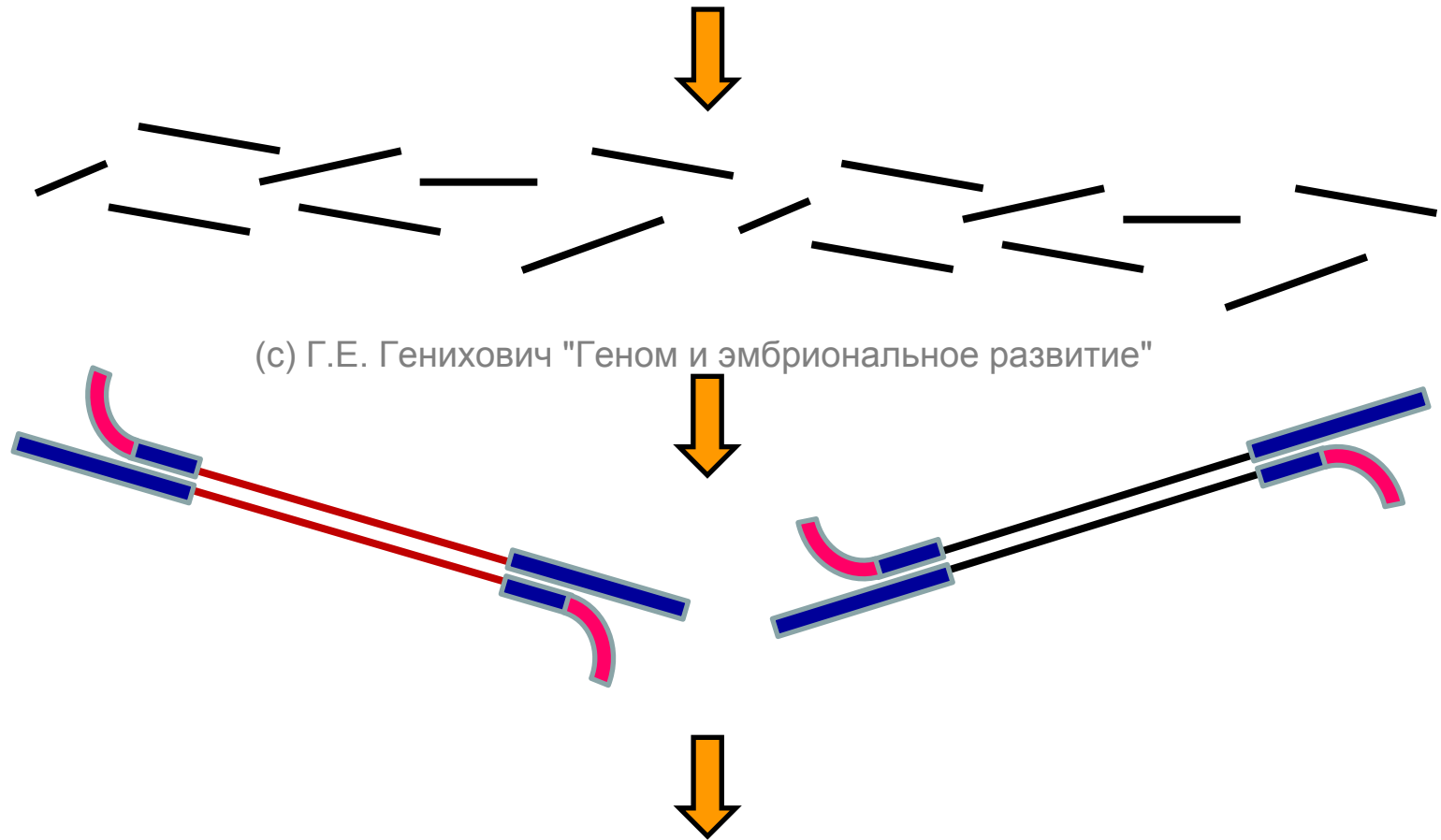
1.800.000 reads, 400-700 bp

- Меньше данных

+ Куски значительно длиннее. Сначала – 200-300bp, теперь 400-700 bp, обещают 800-1000 bp.



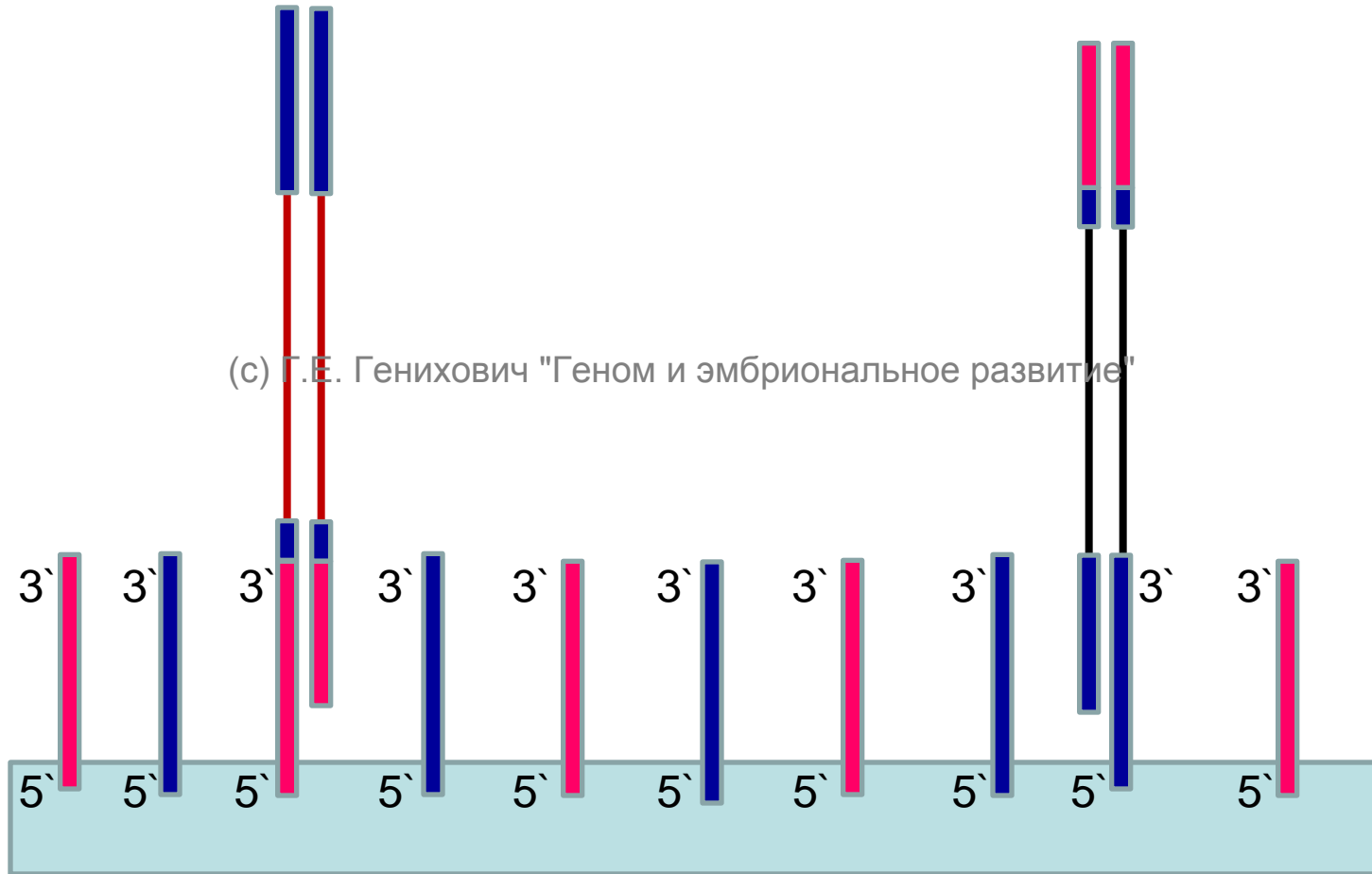
Solexa/Illumina – приготовление библиотеки



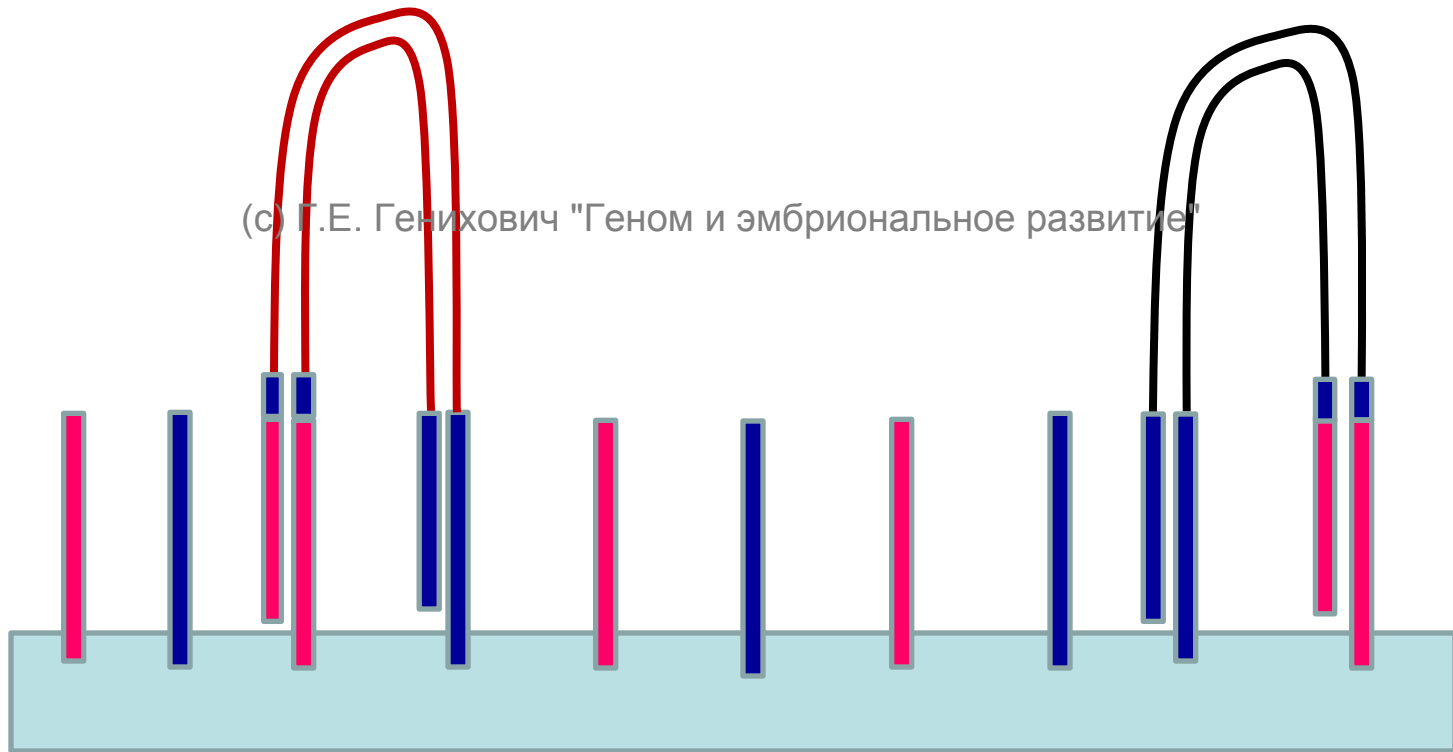
(с) Г.Е. Генихович "Геном и эмбриональное развитие"

Селекция по размеру – ~200 bp

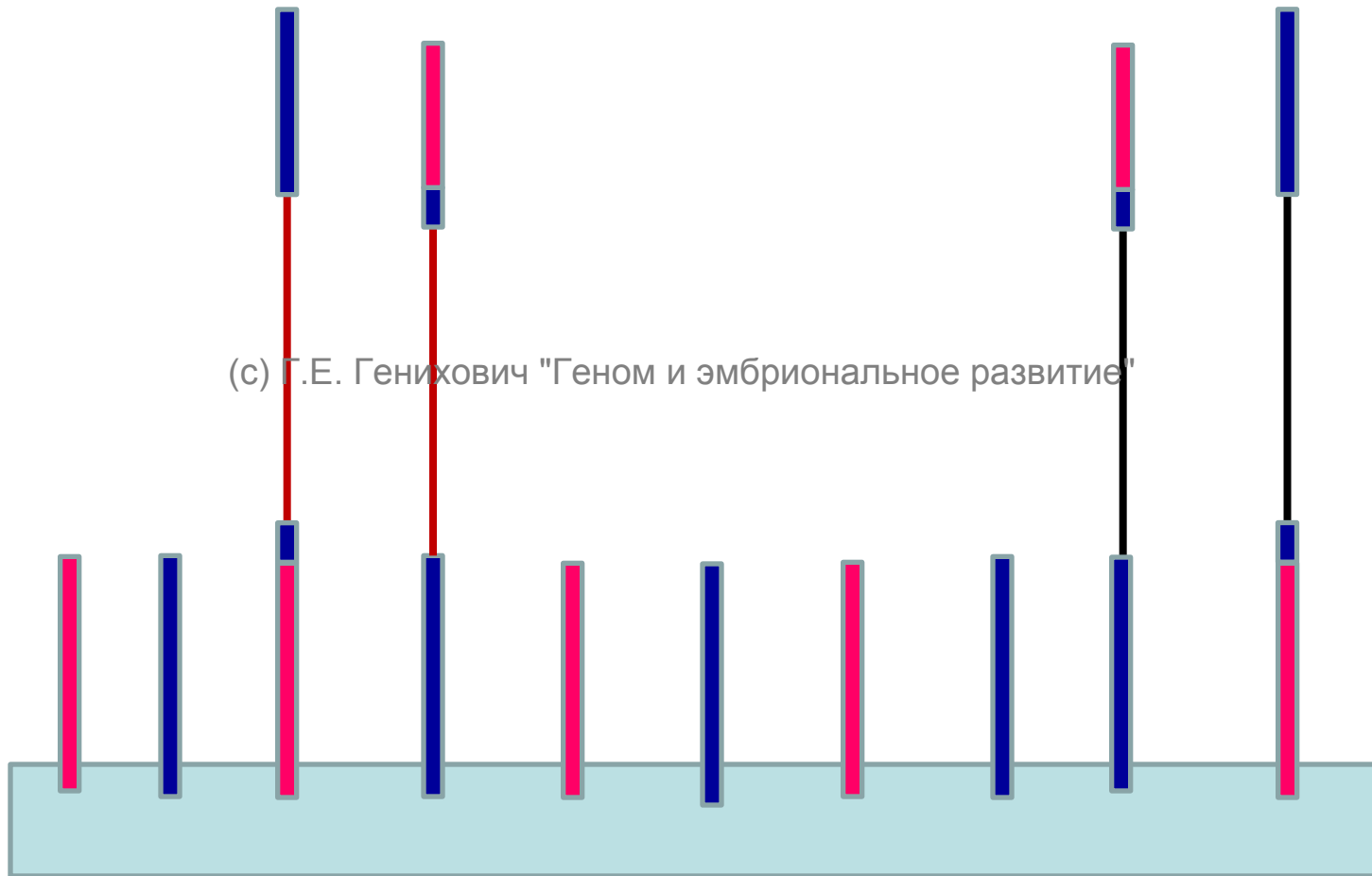
Solexa/Illumina – гибридизация с праймерами на Flow Cell, полимеризация комплементарной цепи



Solexa/Illumina – ПЦР, образование кластеров

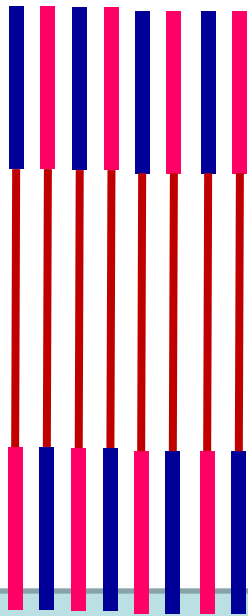


Solexa/Illumina – ПЦР, образование кластеров

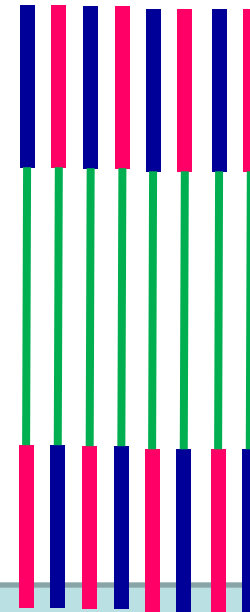
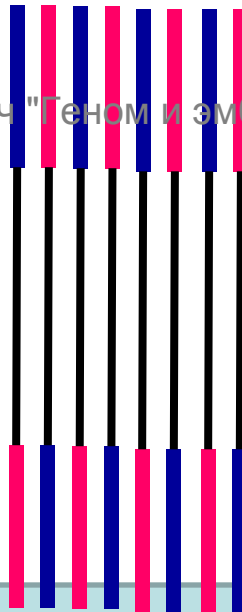


Solexa/Illumina – ПЦР, образование кластеров

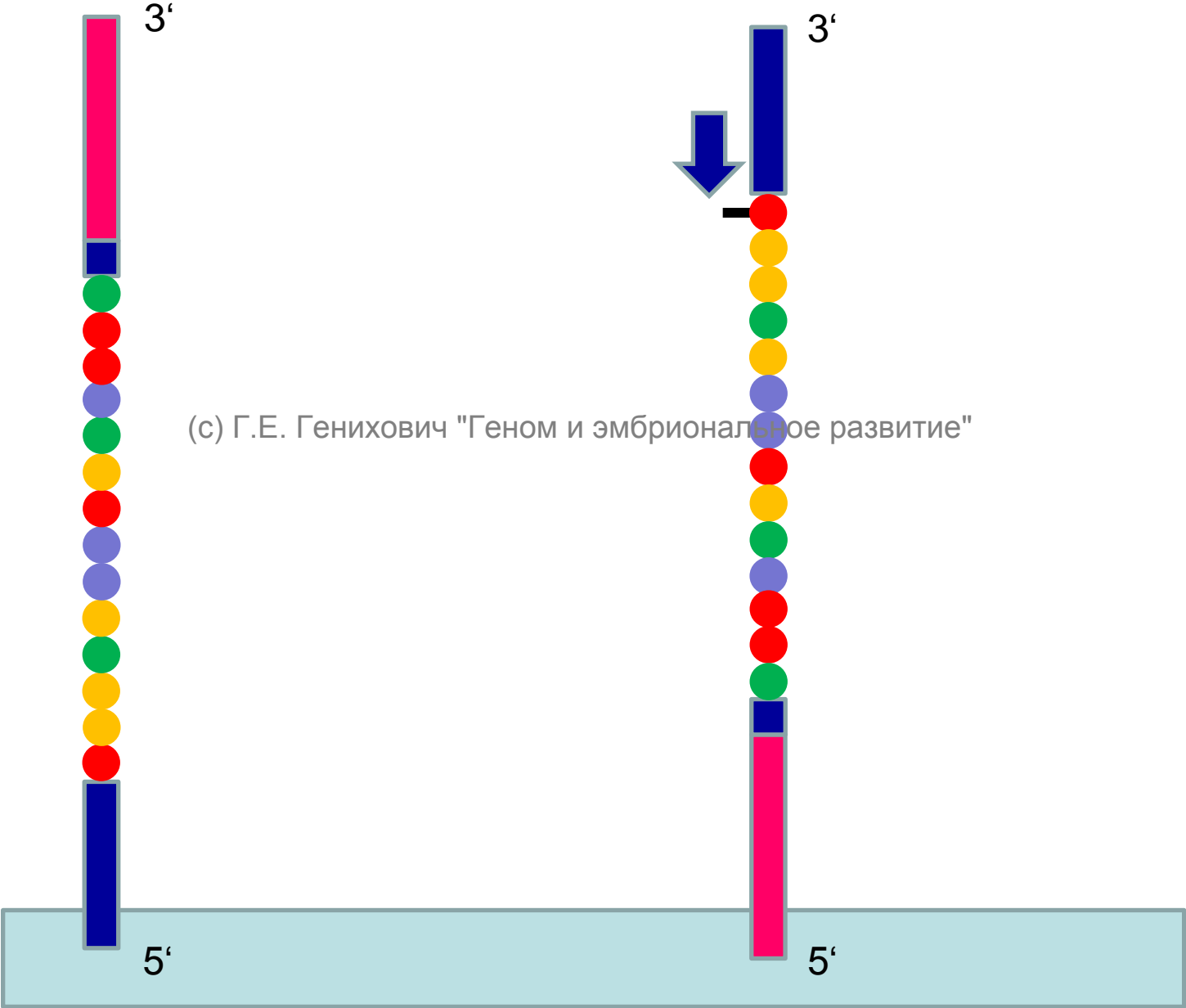
все эти молекулы закреплены на Flow cell своими 5'-концами и имеют свободный 3'-конец



(с) Г.Е. Генихович "Геном и эмбриональное развитие"

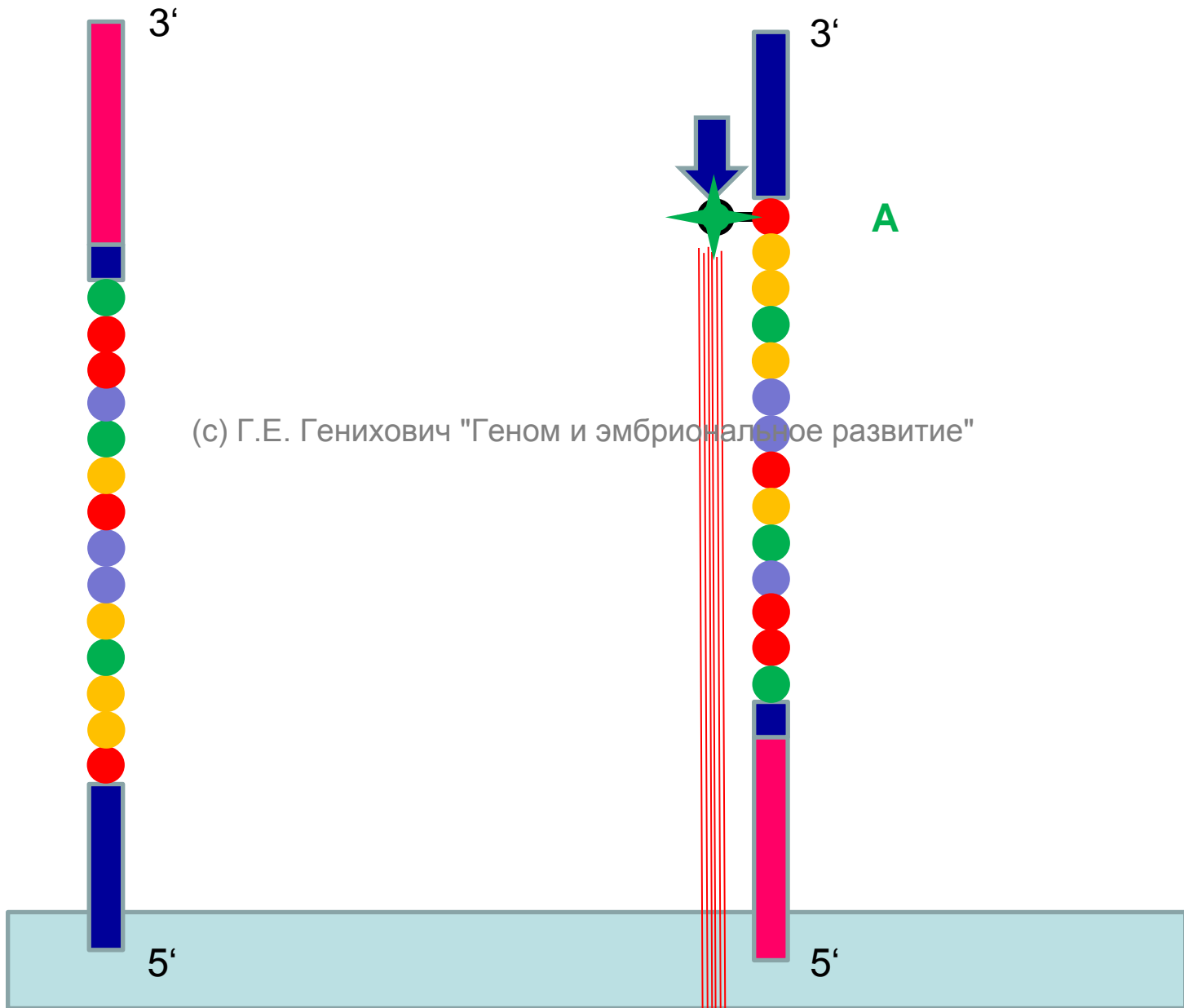


Solexa/Illumina – секвенирование

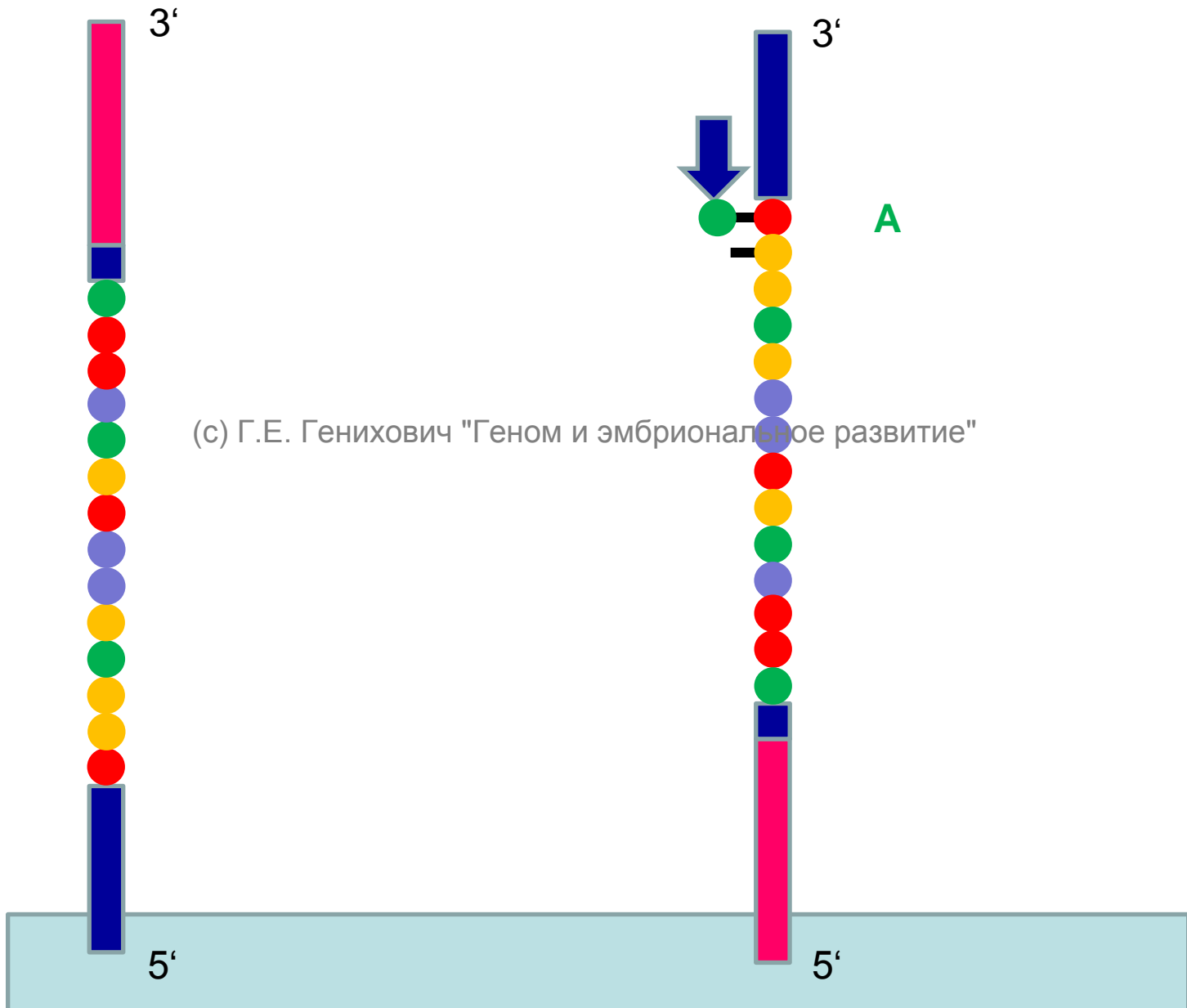


(с) Г.Е. Генихович "Геном и эмбриональное развитие"

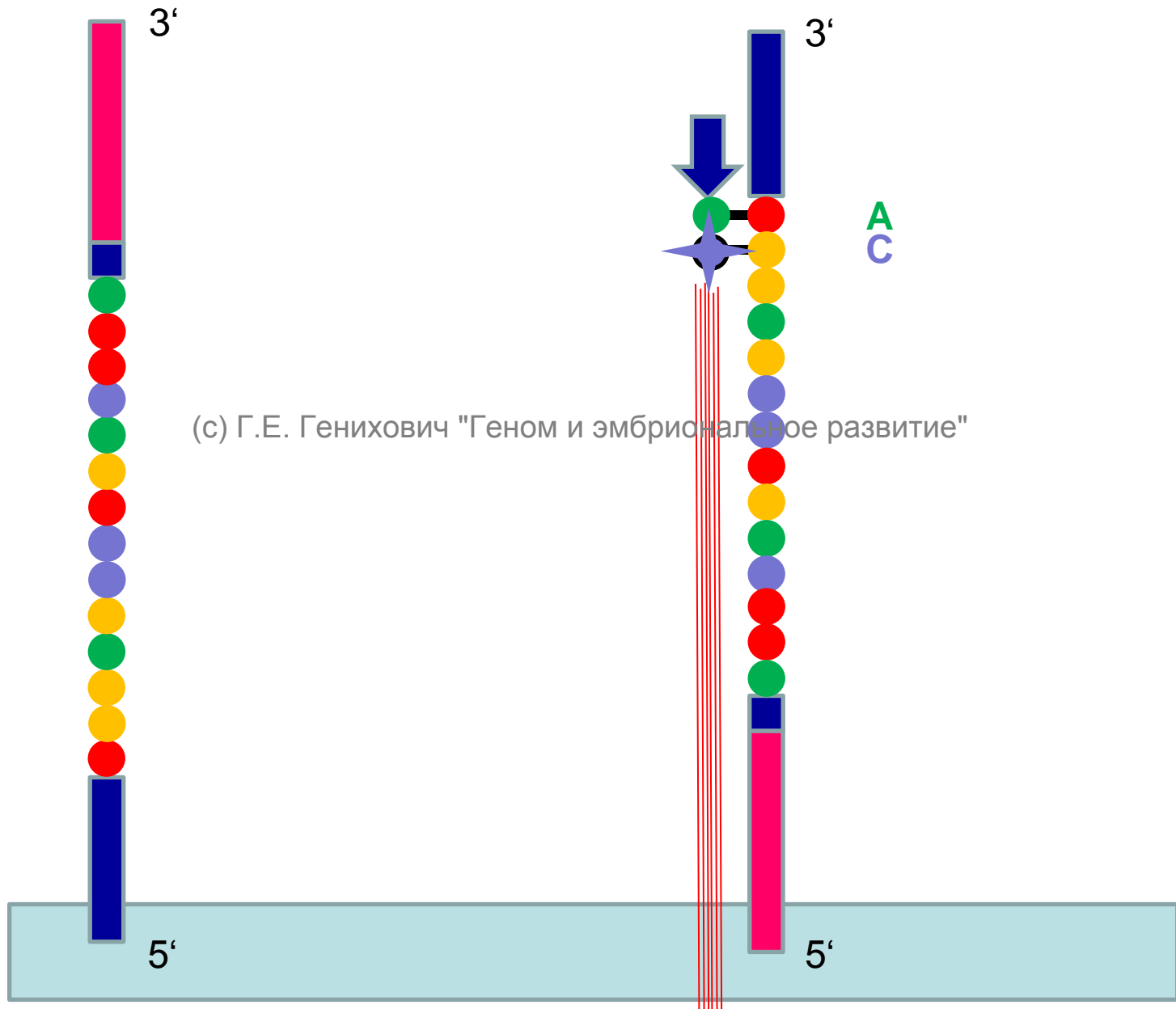
Solexa/Illumina – секвенирование



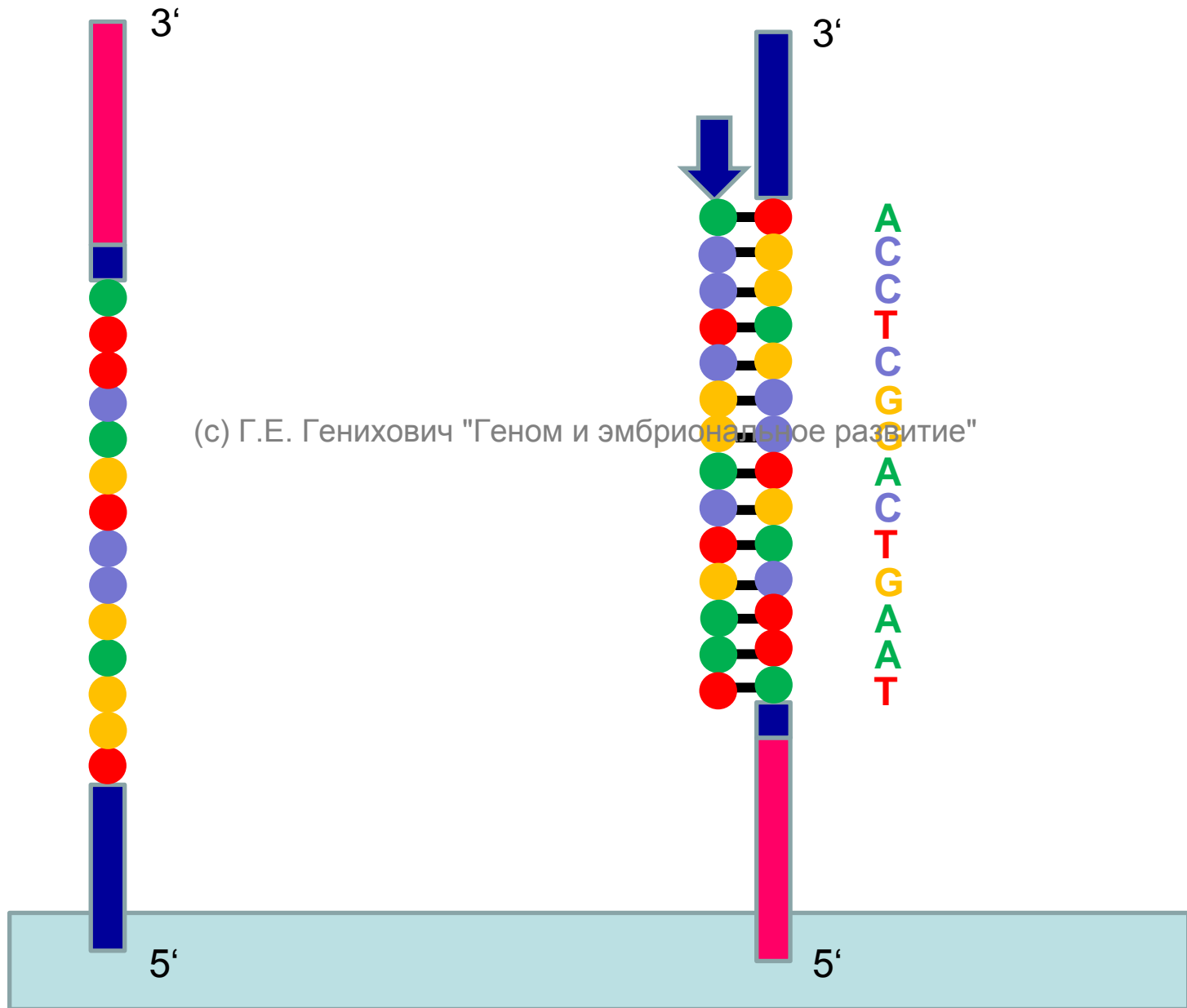
Solexa/Illumina – секвенирование



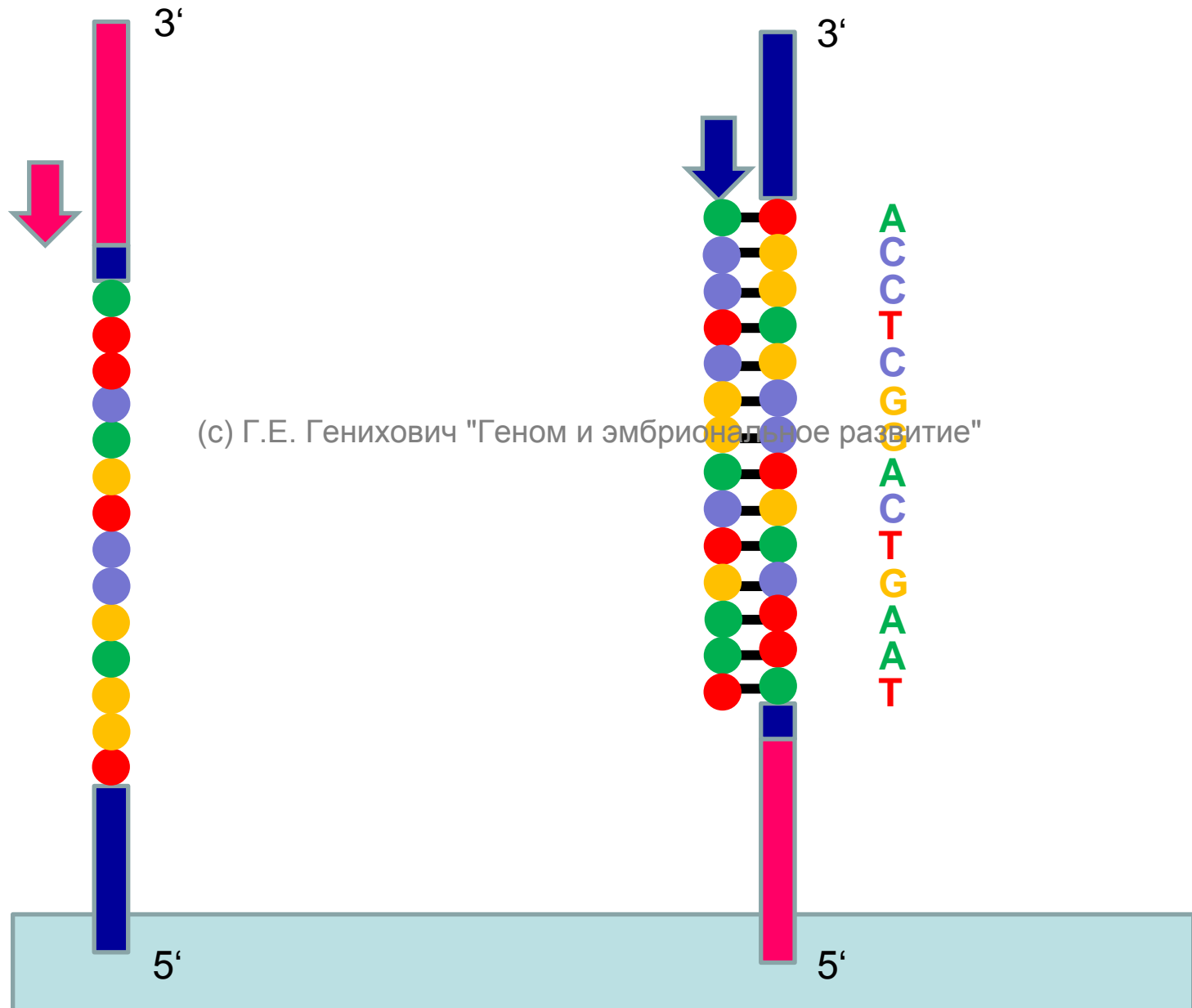
Solexa/Illumina – секвенирование



Solexa/Illumina – секвенирование



Solexa/Illumina – секвенирование с обратной стороны



Solexa/Illumina – выравнивание (alignment)



(с) Г.Е. Генихович "Геном и эмбриональное развитие"

Удобно, если есть известная контрольная последовательность, например, секвенированный геном, к которому можно «приложить» короткие просеквенированные кусочки. Но компьютеры теперь умеют складывать из коротких кусков целые геномы и de novo.

Новое поколение секвенаторов: sequencing by synthesis

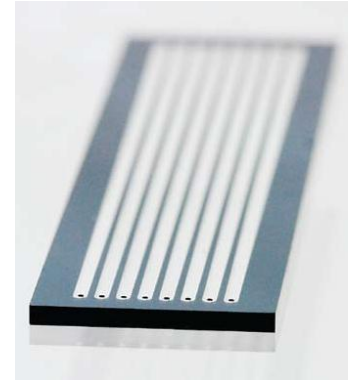
Solexa/Illumina:

Десятки миллионов сиквенсов за один заход.

Длина - 25-100 bp.

+ Много данных, отлично подходит для ChIP-Seq и RNA-Seq (если секвенирован геном)

- Сложно складывать маленькие кусочки

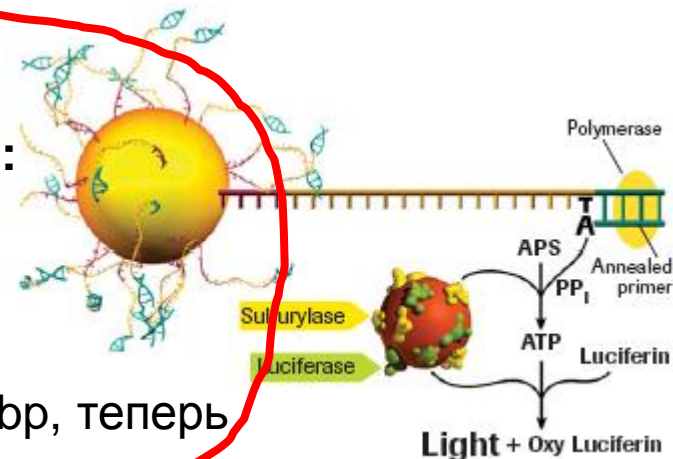


Пиросеквенирование – 454 Life Sciences/Roche:

1.800.000 reads, 400-700 bp

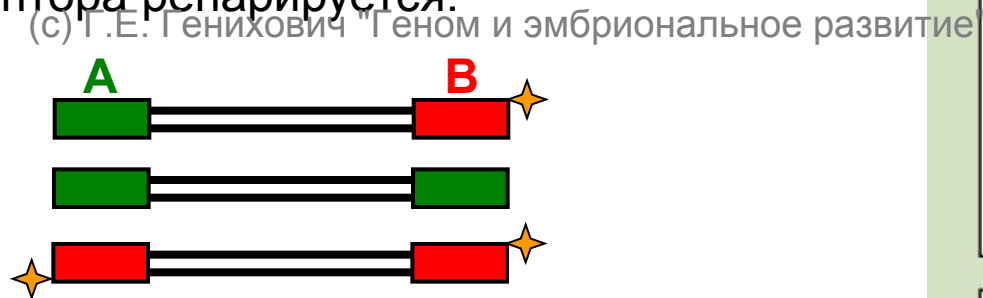
- Меньше данных

+ Куски значительно длиннее. Сначала – 200-300bp, теперь 400-700 bp, обещают 800-1000 bp.

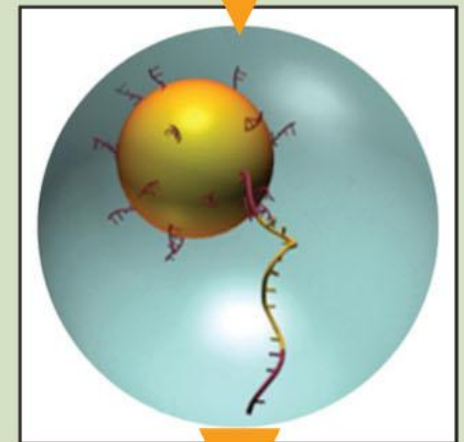
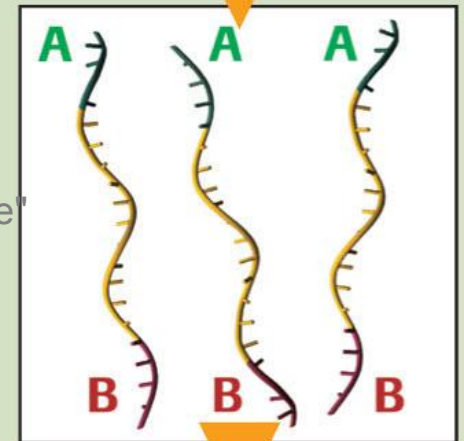


454

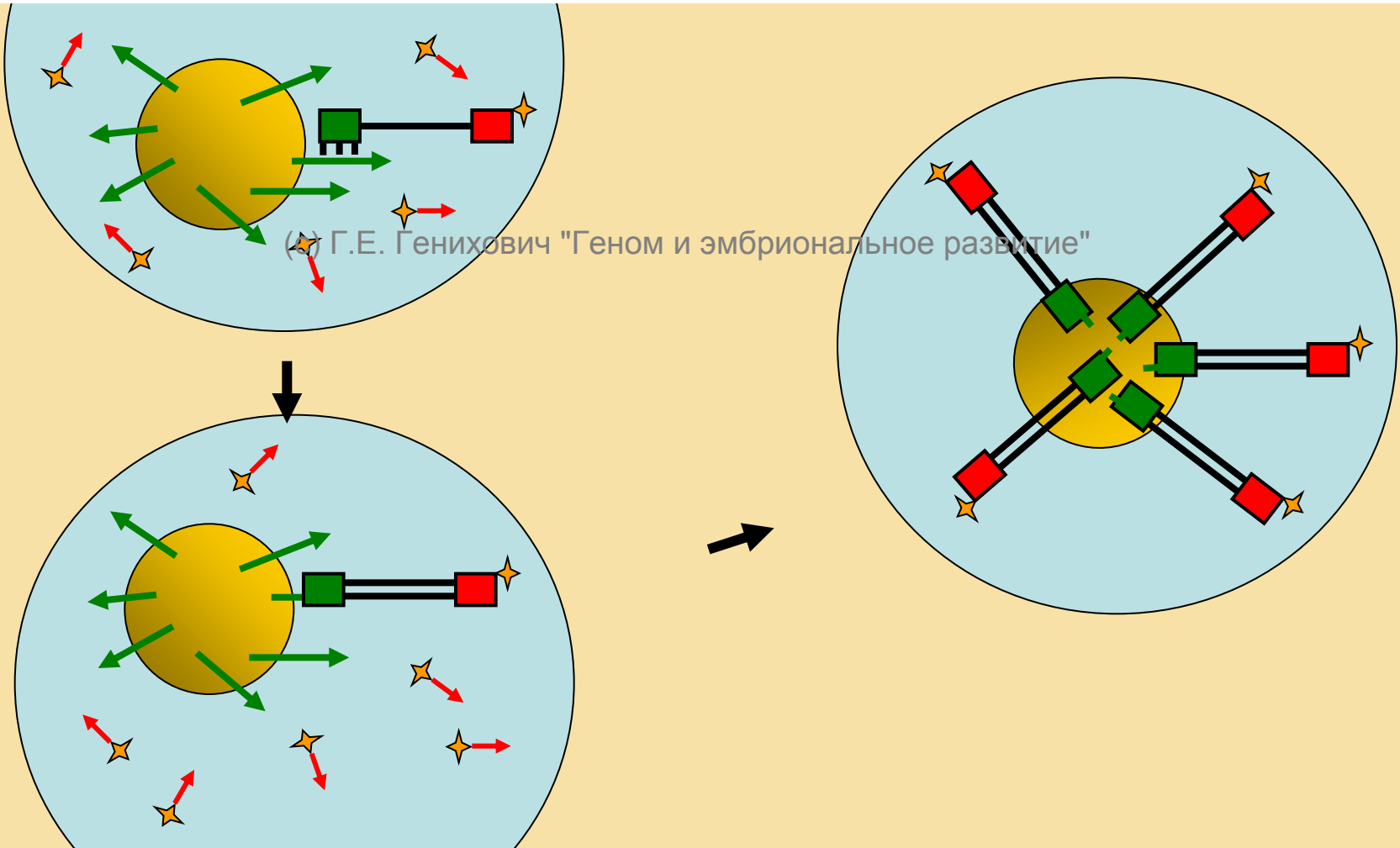
- Фрагментирование ДНК на 300-1000 bp, очистка, фосфорилирование, лигирование дефосфорилированных адапторов А и В (44 bp, 20 bp PCR primer, 20 bp sequencing primer, В – биотинилирован, и «ключевую» последовательность TCAG для калибровки прибора). Фрагменты с одним или двумя адапторами В сажаются на магнитный покрытый стрептавидином шарик. Затем дырка перед 5` концом адаптора репарируется.



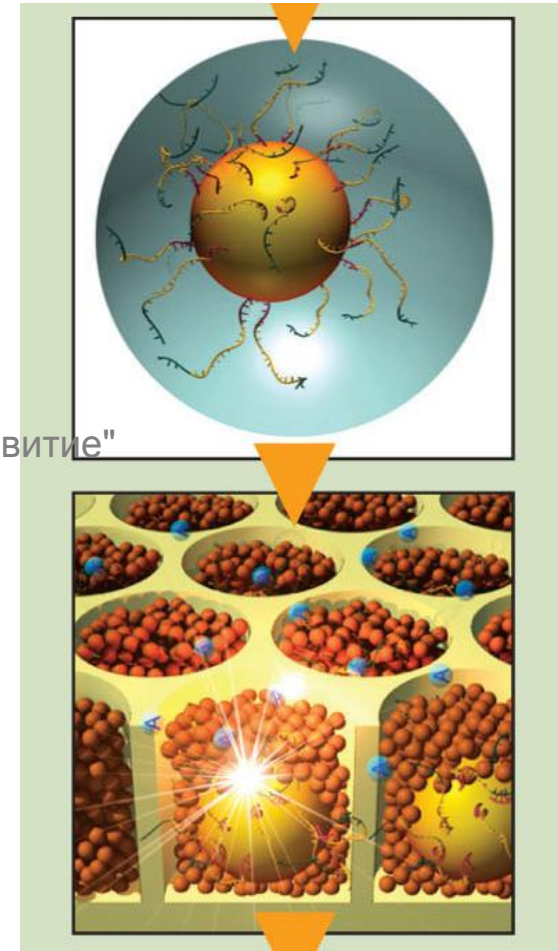
- ДНК денатурируется щелочью, и небютинированные цепи уплывают (single strand template library).
- Неуплывшая ДНК гибридизуется с покрытым праймером к адаптору А шариком.

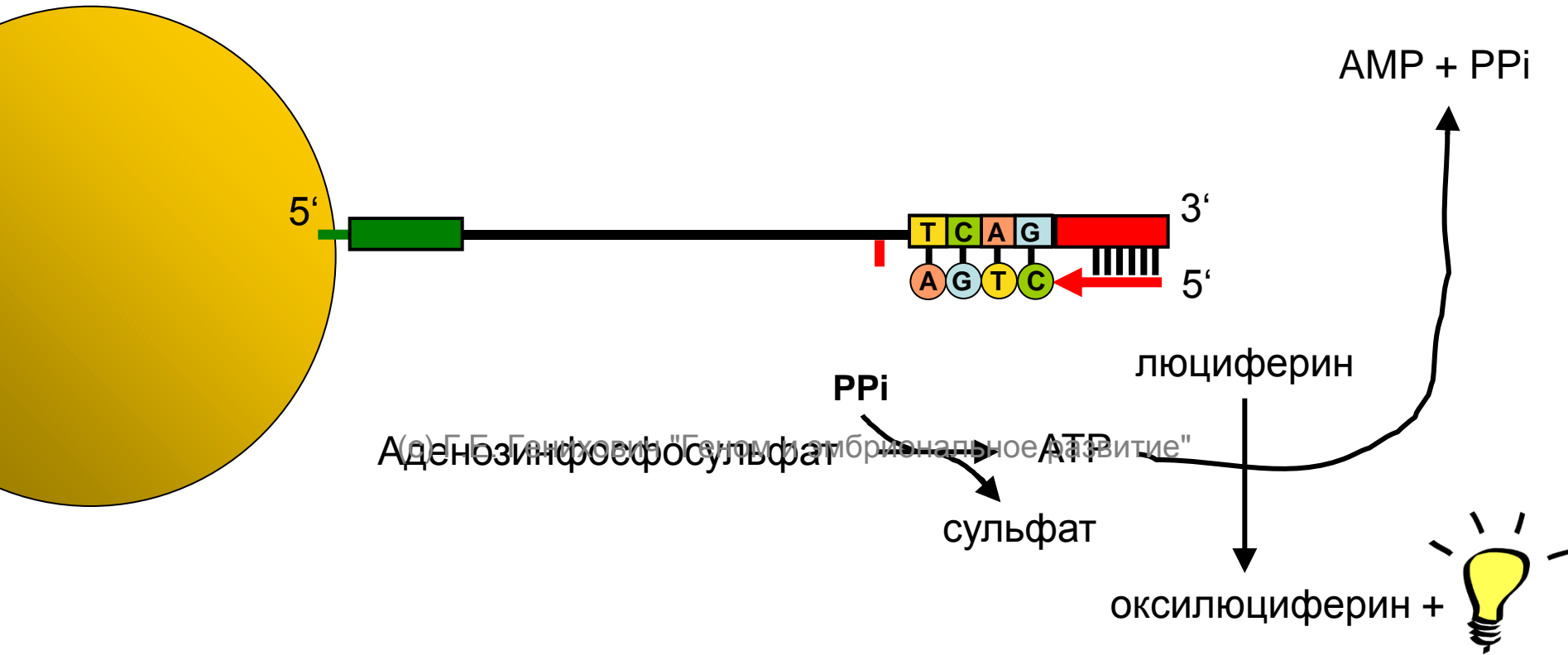


- Шарики с гибридованными с праймером молекулами оцДНК помещаются в эмульсию «вода в масле». В каждом пузырьке воды находится 1 шарик и все компоненты ПЦР. Праймер В биотинилирован. В результате на каждом шарике вырастает клон молекул.

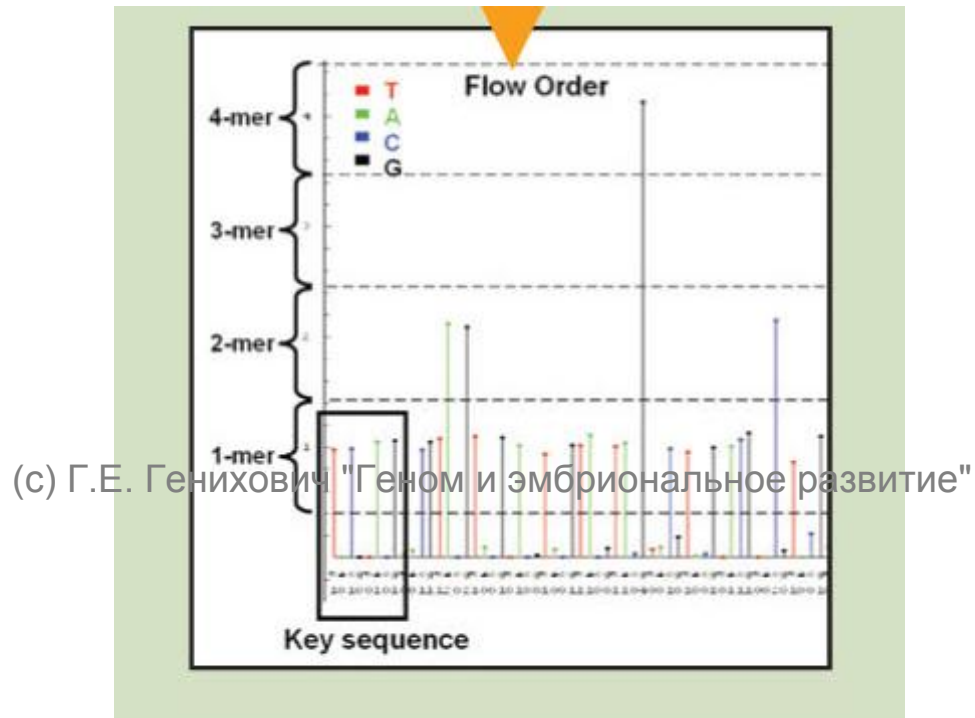


- Далее эмульсию разрушают и обрабатывают магнитными стрептавидиновыми шариками. Шарики с клонами ДНК сажаются на магнит, а пустые – смываются. Потом отделяются магнитные шарики.
- ДНК денатурируется щёлочью, не связанные 5' концом с шариком цепи смываются. Цепи с шариком смываются эмбриональное развитие"
- Шарики с клонами оцДНК рассаживаются по одному в 40 μm лунки оптоволоконной пикотитровальной платы вместе с «ферментативными» и «упаковочными» шариками. Ферментативные шарики содержат сульфурилазу и люциферазу.





- Через лунки пропускают ПЦР-буфер и нуклеотиды по одному. Если нуклеотид комплементарный, он включается в синтезируемую цепь. При этом освобождается пиррофосфат. Сульфурилаза превращает аденозинфосфосульфат и PPi в ATP. С помощью ATP люцифераза окисляет люциферин и при этом выделяется свет в количестве, пропорциональном количеству включенных нуклеотидов. Испускаемый свет измеряется, и определяется число включенных нуклеотидов.



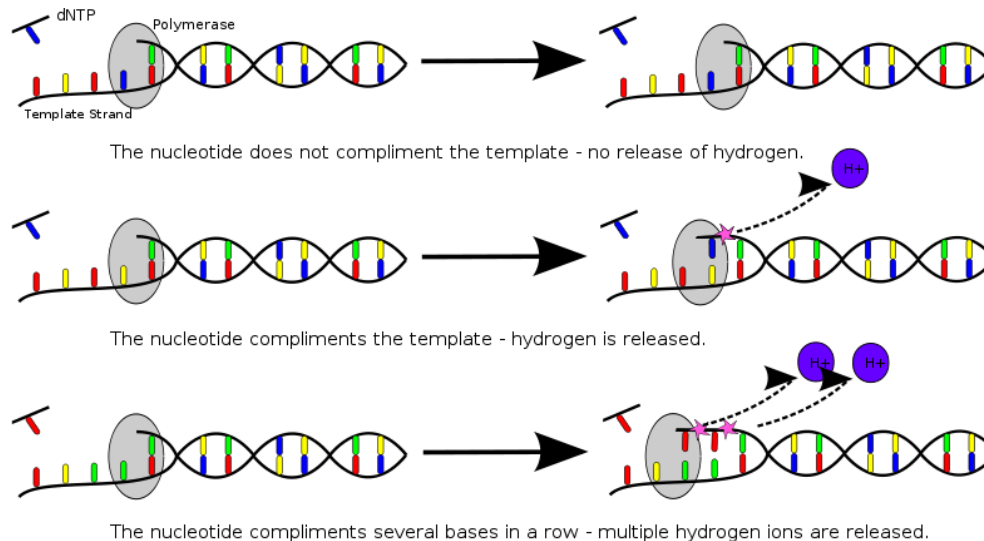
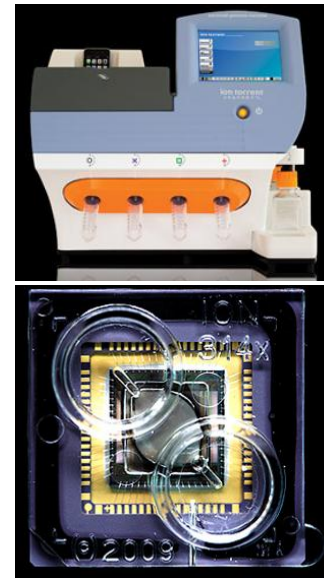
- Секвенирование происходит одновременно во всех лунках. Затем полученные последовательности складываются в контиги.

Технологии ближайшего будущего

Полупроводниковое секвенирование (Ion torrent; Life Technologies):

- 1,2-11 миллионов сиквенсов за один заход
- средняя длина 100 bp, нуклеотиды добавляются по-очереди
- детектируется не свет, а изменение pH на 0.02, возникающее при отделении протона, когда образуется одна новая фосфодиэфирная связь
- Очень дешево по сравнению с машинами, детектирующим свет

(с) Г.Е. Генихович "Геном и эмбриональное развитие"



Технологии ближайшего будущего

Single molecule real time DNA sequencing (SMRT)

- секвенируется одиночная молекула ДНК, прикрепленная ко дну лунки диаметром ~ 70 nm
- измеряется длительное (миллисекунды) свечение одного цвета при присоединении полимеразой «цветного» нуклеотида на фоне быстро диффундирующих (микросекунды) четырёх
- флуорофор сидит не на нуклеотиде, а на последнем фосфате и отделяется вместе с PP_i , оставляя нормальный нуклеотид без «довесков»
- средняя длина 1000 bp, максимальная >10.000 bp
- очень быстрый процесс – 30 минут на всё про всё
- взяв вместо ДНК-полимеразы обратную транскриптазу, можно секвенировать РНК



это – не я

**Как выглядит геном после shotgun-
секвенирования и как можно его
улучшить?**

(с) Г.Е. Генихович "Геном и эмбриональное развитие"

«Глубокое» секвенирование РНК (RNA-seq)

Выделение polyA+РНК, фрагментация РНК ультразвуком



Синтез двухцепочечной кДНК с использованием случайных праймеров



(с) Г.Е. Генихович "Геном и эмбриональное развитие"

Лигирование адаптеров для секвенирования на приборе Illumina



Секвенирование



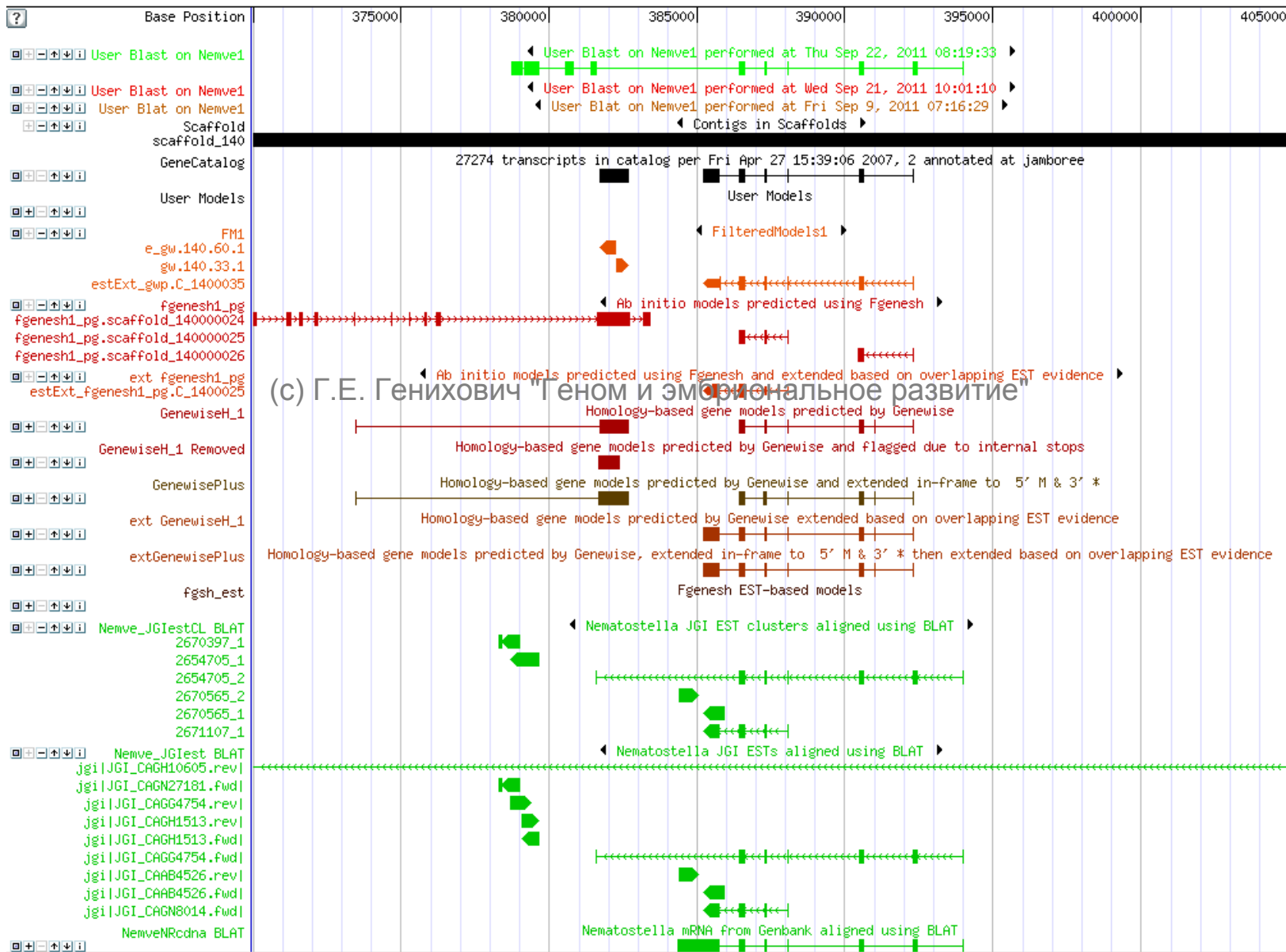
Alignment кусочков к геному

Position
 <<< << < > >> >>> << Scaffold >>
 scaffold_140:370000-405000 Apply

Zoom
 1.5x 3x 10x Base
 1.5x 3x 10x AA Fit

General
 DNA Permalink Add custom tracks
 + Strand Flip Open / Close Toolbar

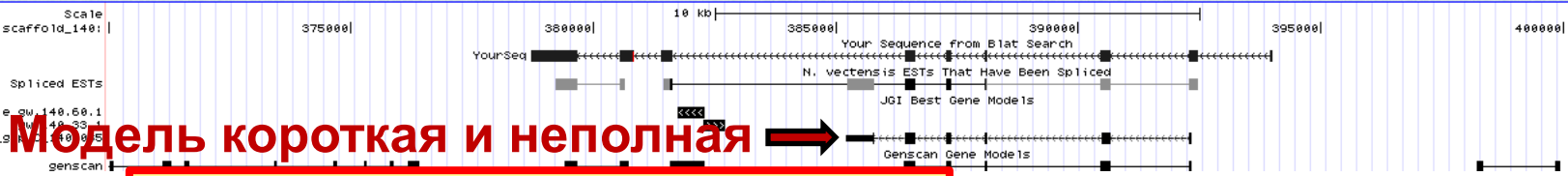
Size: 35001 Feature: 385206-388121 (-) No hit



UCSC Genome Browser on *N. vectensis* June 2007 Assembly (nemVec1)

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search scaffold_140:370,000-405,000 jump clear size 35,001 bp. configure



Модель короткая и неполная →

RNA-seq позволяет:

- уточнять модели генов
- исследовать альтернативный сплайсинг
- определять уровень экспрессии генов на разных стадиях развития или в разных типах клеток
- искать некодирующие транскрипты и малые РНК...

... и все это не с отдельными генами, а со всеми генами сразу

(с) Г.Е. Генихович "Геном и эмбриональное развитие"



- ChIP-seq с антителами против характерным образом модифицированных гистонов позволяет находить промоторы и энхансеры генов
- ... и, опять-таки, все это не с отдельными генами, а со всеми генами сразу

Два альтернативных промотора

Д. В. Савицкий, 2010, "Эволюция и эмбриональное развитие"

Геном секвенирован – что дальше?

- Молекулярная филогенетика
- Генный репертуар, эволюция семейств генов
- Исследования функций и регуляции отдельных генов
- Выявление мутаций, вызывающих известный фенотип – в первую очередь исследование генетических заболеваний.

(с) Г.Е. Генихович. Геном и эмбриональное развитие

Evo/Devo: два важных для понимания эволюции генома



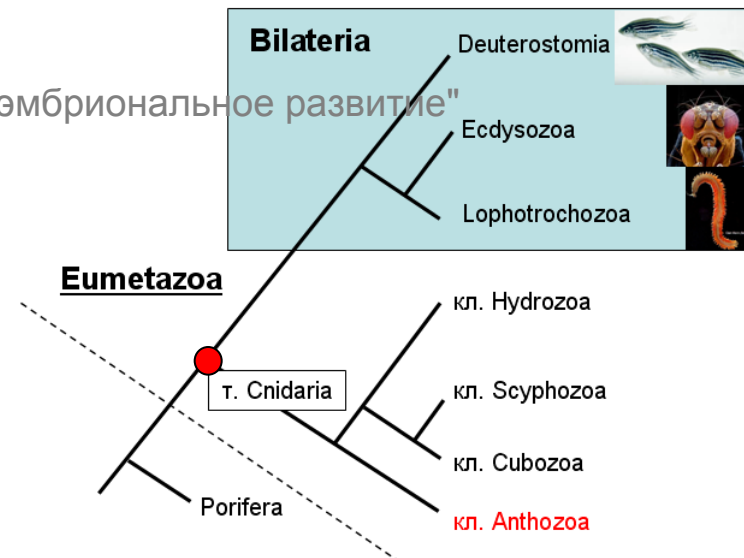
(с) Г.Е. Генихович "Геном и эмбриональное развитие"



Nematostella vectensis Stephenson 1935

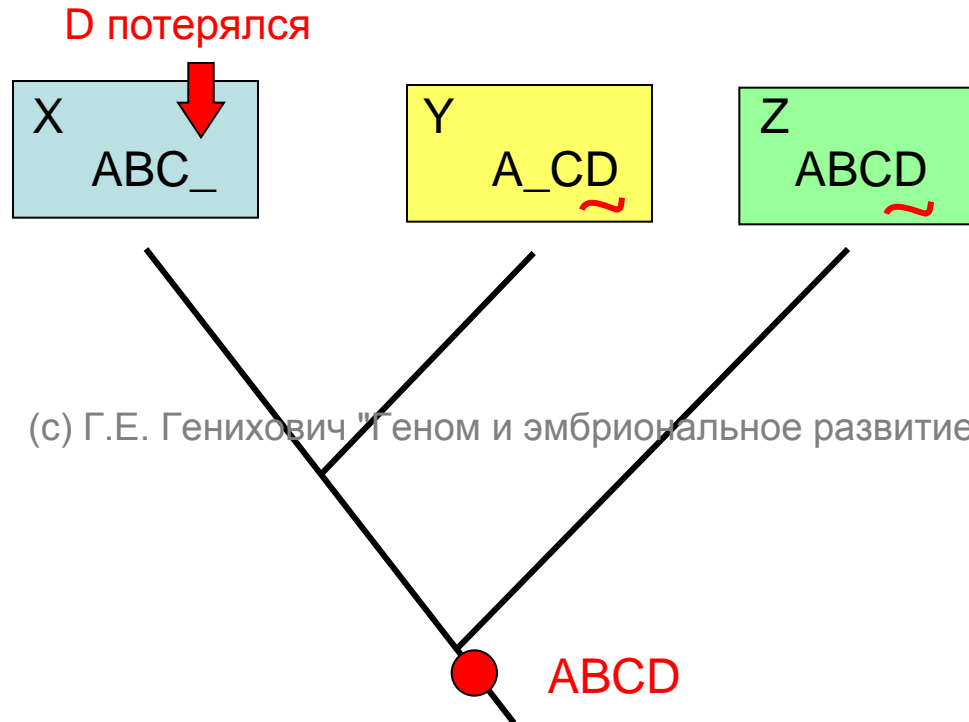


- Солоноватоводная актиния
- Копаются в иле
- Важное филогенетическое положение



- Модельный объект

Отступление: про потери и находки



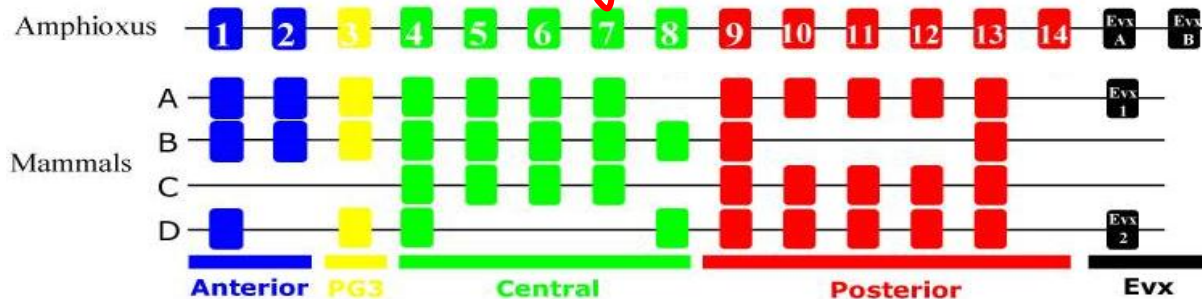
Если бы не было информации, что D есть в группе животных "Z", он бы считался «изобретением» группы "Y", а на самом деле он был уже у общего предка "X", "Y" и "Z". Та же ситуация с геном B и его потерей группой "Y".

Nematostella vectensis

Покрытие – показатель того, сколько раз один и тот же нуклеотид встречается среди просеквенированных последовательностей

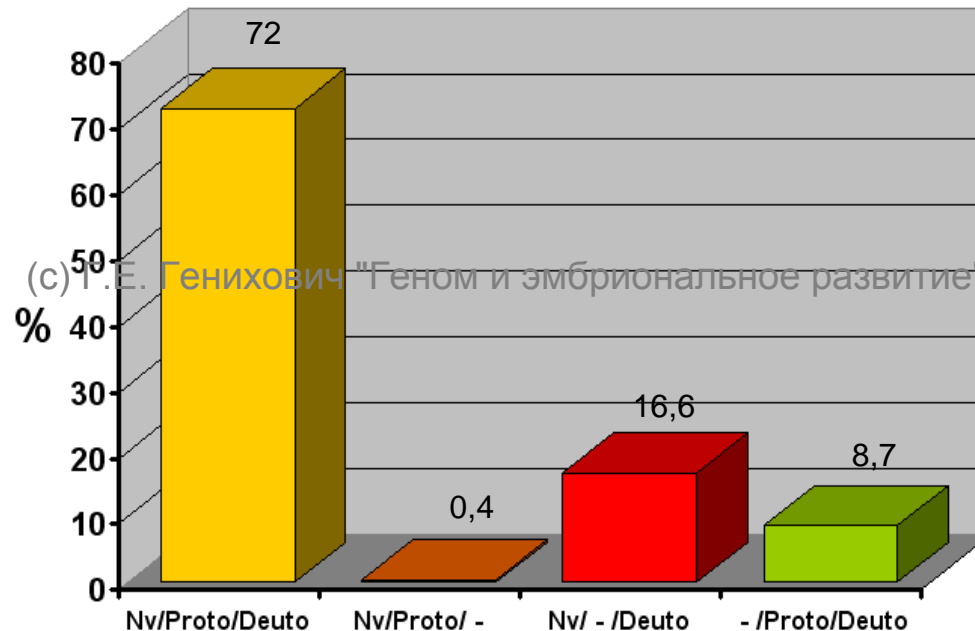
- Shotgun sequencing, 6.5x покрытие
- Размер собранного генома 357 Мбр, половина – в скаффолдах >480 kb, 95% имеющихся EST (> 146.000) выравниваются с геномом
- Примерно треть последовательностей – минисателлиты > 1 kb. Общий размер собранного генома + повторов
- Число генов, кодирующих белки оценено как 18 000
- Сравнение с *Drosophila*, *Caenorhabditis*, *Fugu*, *Xenopus* позволило определить 7766 семейств генов, имеющих предка многоклеточных животных. В это число не входят гены.

Семейство генов – все гены, которые происходят от одного общего гена-предка у последнего общего предка двух групп животных.



Белки Cnidaria больше похожи на белки человека, чем на белки мухи

Процент генов, имеющих ортологов



(с) Г.Е. Генихович "Геном и эмбриональное развитие"



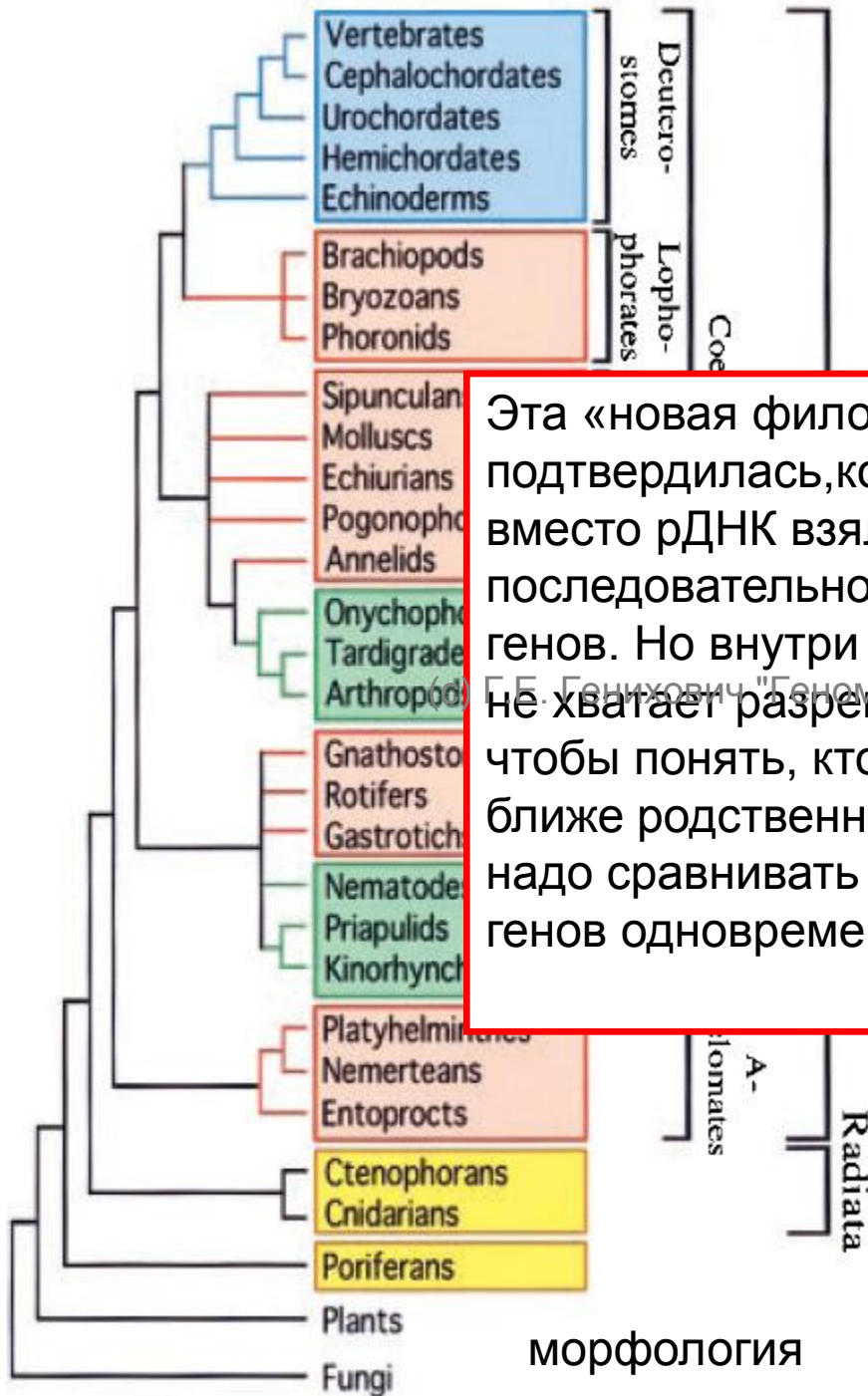
Возможные «приобретения» Bilateria

Секвенирование *Nematostella* показало, что *Drosophila* и *Caenorhabditis* потеряли в ходе эволюции множество генов. Незнание этого вызвало ложные представления о множестве генных «нововведений» у позвоночных.

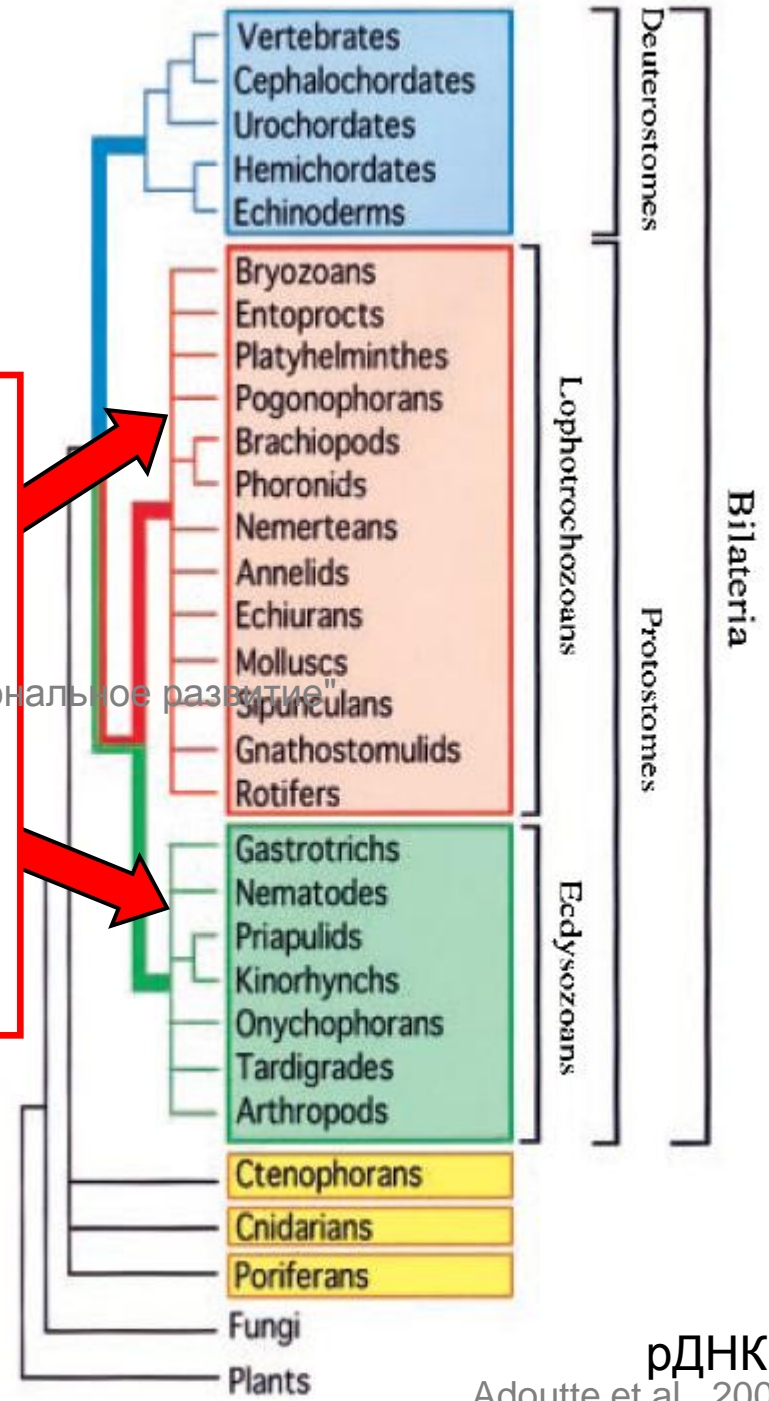
Пример:

(с) Г.Е. Генихович "Геном и эмбриональное развитие"

У *Nematostella* найдены представители 11 из 12 семейств генов группы *Wnt*, имеющих у позвоночных, тогда как *Drosophila* имеет только 8 из 12 – остальные утрачены.

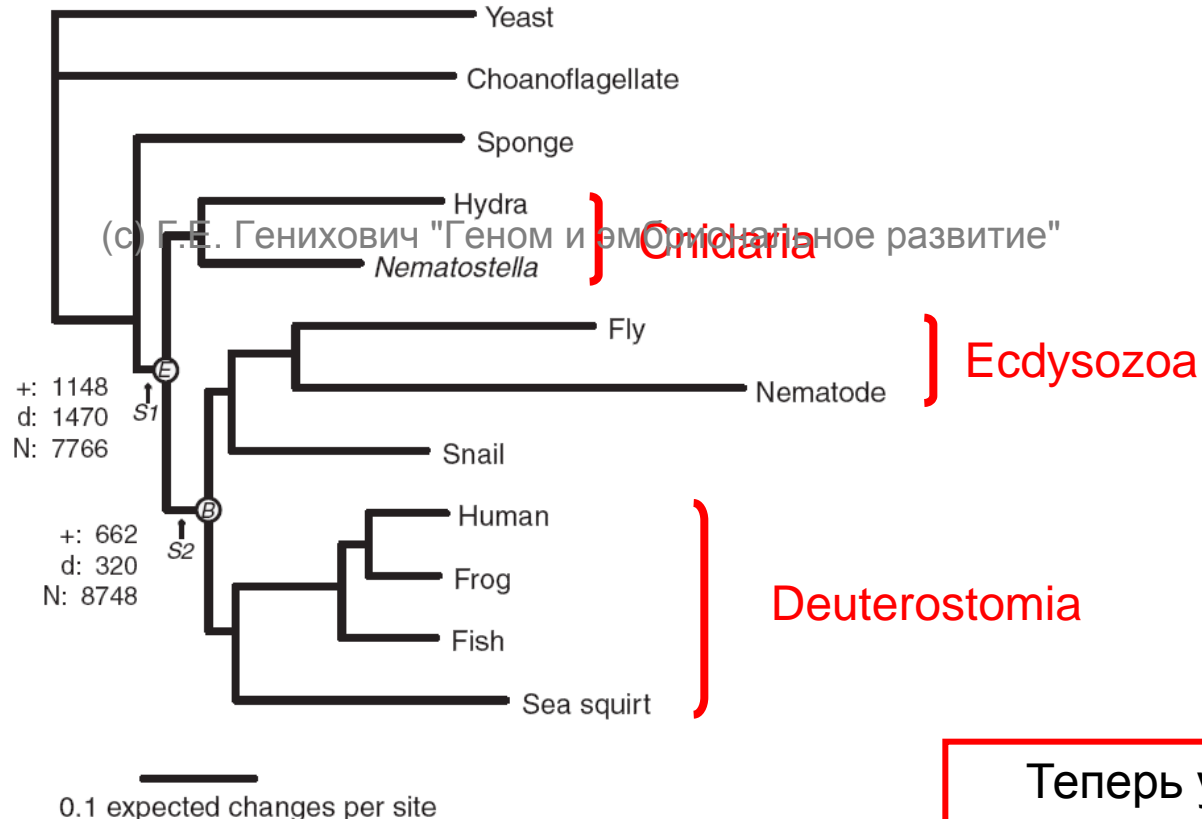


Эта «новая филогения» подтвердилась, когда вместо рДНК взяли последовательности Нох генов. Но внутри групп не хватает разрешения, чтобы понять, кто кому ближе родственник => надо сравнивать больше генов одновременно.



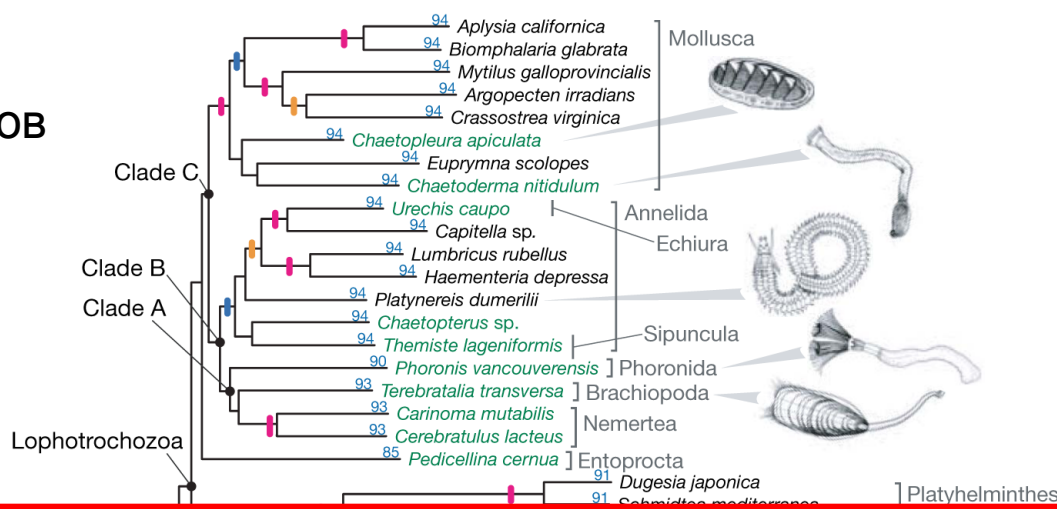
Nematostella vectensis

- Взяли 337 групп генов, имеющих единственную копию в геномах всех рассматриваемых животных, и сделали дерево (Байезова статистика).



Теперь у дерева
стало мало веток.

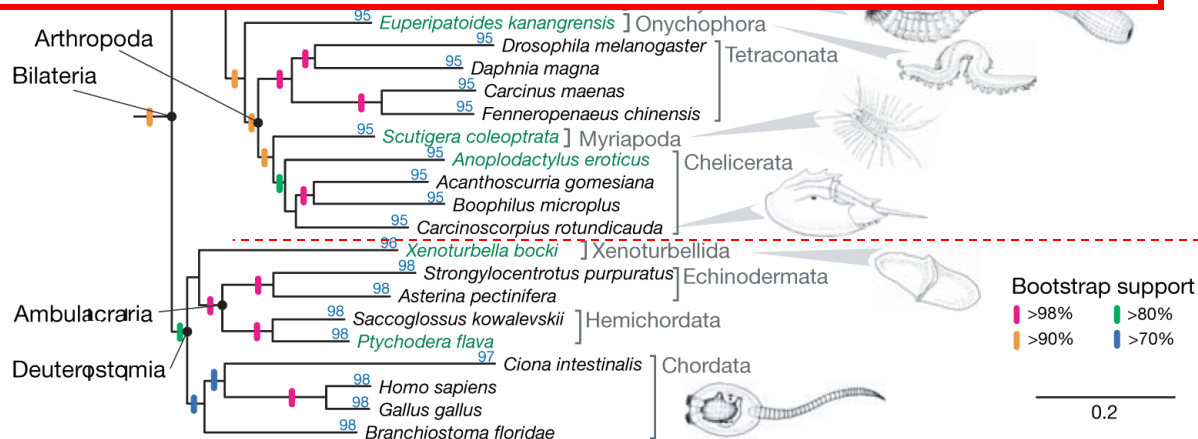
150 генов



Lophotrochozoa

«Новая филогения» подтверждается, хотя спорные группы ещё есть (губки, Trichoplax, взаимное положение Cnidaria и Stenophora, положение Chaetognata и так далее).

(с) Г.Е. Генихович "Геном и эмбриональное развитие"



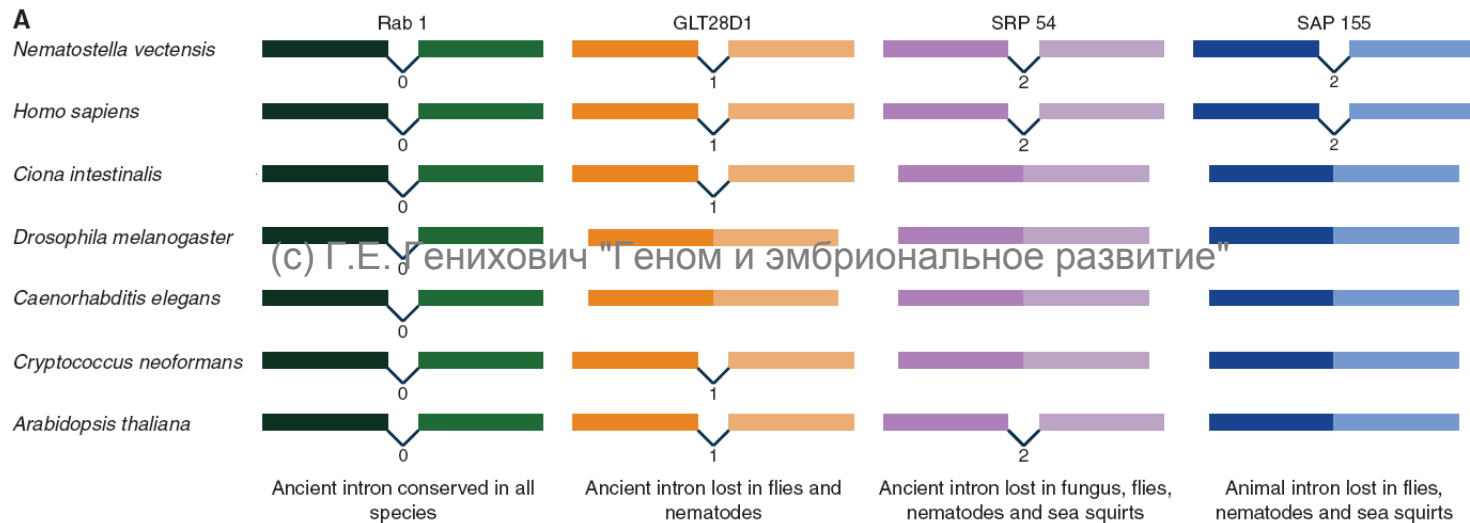
Ecdysozoa

Deuterostomia

Dunn et al., 2008

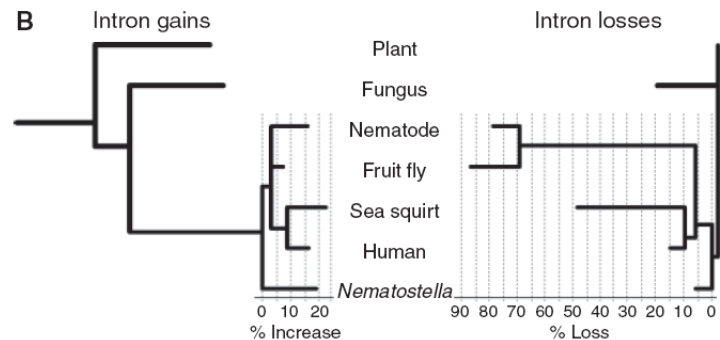
Nematostella vectensis

- Консервативные интроны Bilateria и *Nematostella* (число, положение, фаза) показывают, что гены общего предка Eumetazoa имели много интронов. При рассмотрении генов, имеющих гомологов у *Nematostella* и человека, оказалось, что 81% интронов консервативны. 26% - древние эукариотические интроны, имеющиеся и у *Arabidopsis*



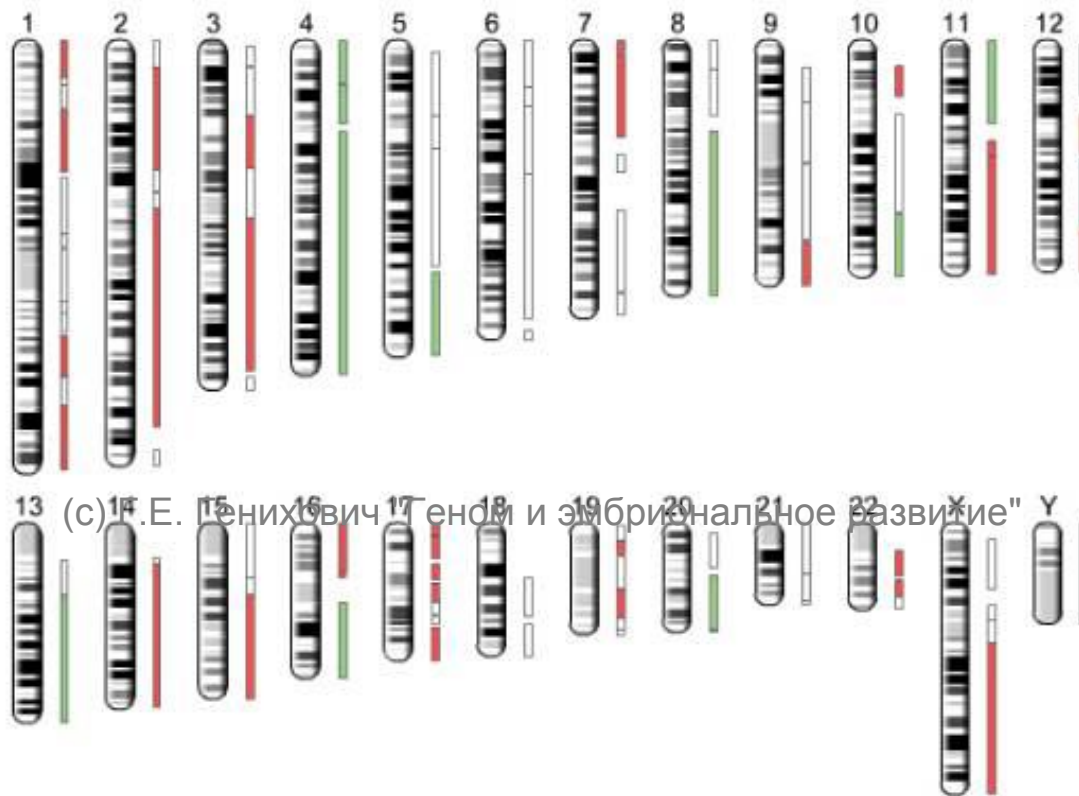
(с) Г.Е. Яенихович "Геном и эмбриональное развитие"

Fig. 2. Patterns of intron evolution in eukaryotes. **(A)** Examples of different patterns of intron gain and loss. Bars of the same color represent conserved regions across all species. Chevrons indicate introns and the number below the chevron shows the phase of the intron. **(B)** Branch lengths proportional to the number of inferred intron gains (left), and intron losses (right) under the Dollo parsimony assumption that introns with conserved position and phase were gained only once in evolution. The bottom scale indicates the change in intron number for gains (left) and losses (right), relative to the inferred introns of the eumetazoan ancestor. Based on a sample of 5175 introns at highly conserved protein sequence positions from *Arabidopsis thaliana* (plant), *Cryptococcus neoformans* (fungus), *C. elegans* (nematode), *D. melanogaster* (fly), *C. intestinalis* (sea squirt), *Homo sapiens* (human), and *Nematostella* (32).



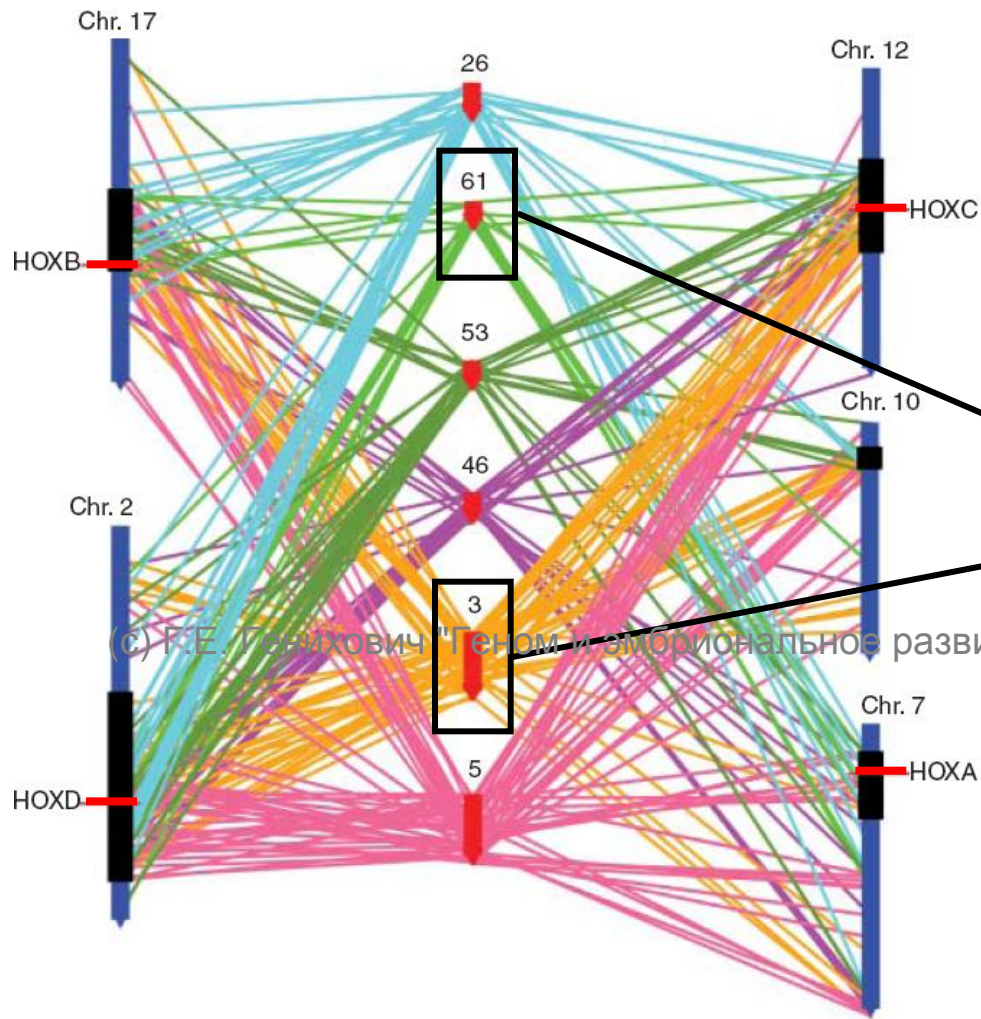
Nematostella vectensis

- В ходе эволюции порядок и ориентация генов сильно менялись из-за внутрихромосомных инверсий и транспозиций, однако общий набор генов на том или ином сегменте хромосомы мог в значительной степени сохраниться.
- Сравнение геномов позвоночных выявило 98 сегментов, в которых в последнее время не происходили хромосомные перестройки. Эти сегменты покрывают 81% генома человека. Если сравнить набор (но не взаимное расположение и ориентацию) генов в их пределах с положением их гомологов у *Nematostella*, то окажется, что существуют крупные участки синтении => PAL (putative ancestral eumetazoan linkage group)



(с) Е. Е. Пеникович "Геном и эмбриональное развитие"

Красные сегменты – 12 компактных PAL, зеленые сегменты – тринадцатый диффузный PAL, белые сегменты – нет синтении с *Nematostella*. Не менее 30% генома нематостеллы входит в PAL.



© F.E. Генихович "Геном и эмбриональное развитие"

Скаффолды, в которых *Нох* гены нематостеллы находятся среди характерных для *Нох*-ов генов-соседей.

PAL A (черные прямоугольники) человека и соответствующие скаффолды нематостеллы

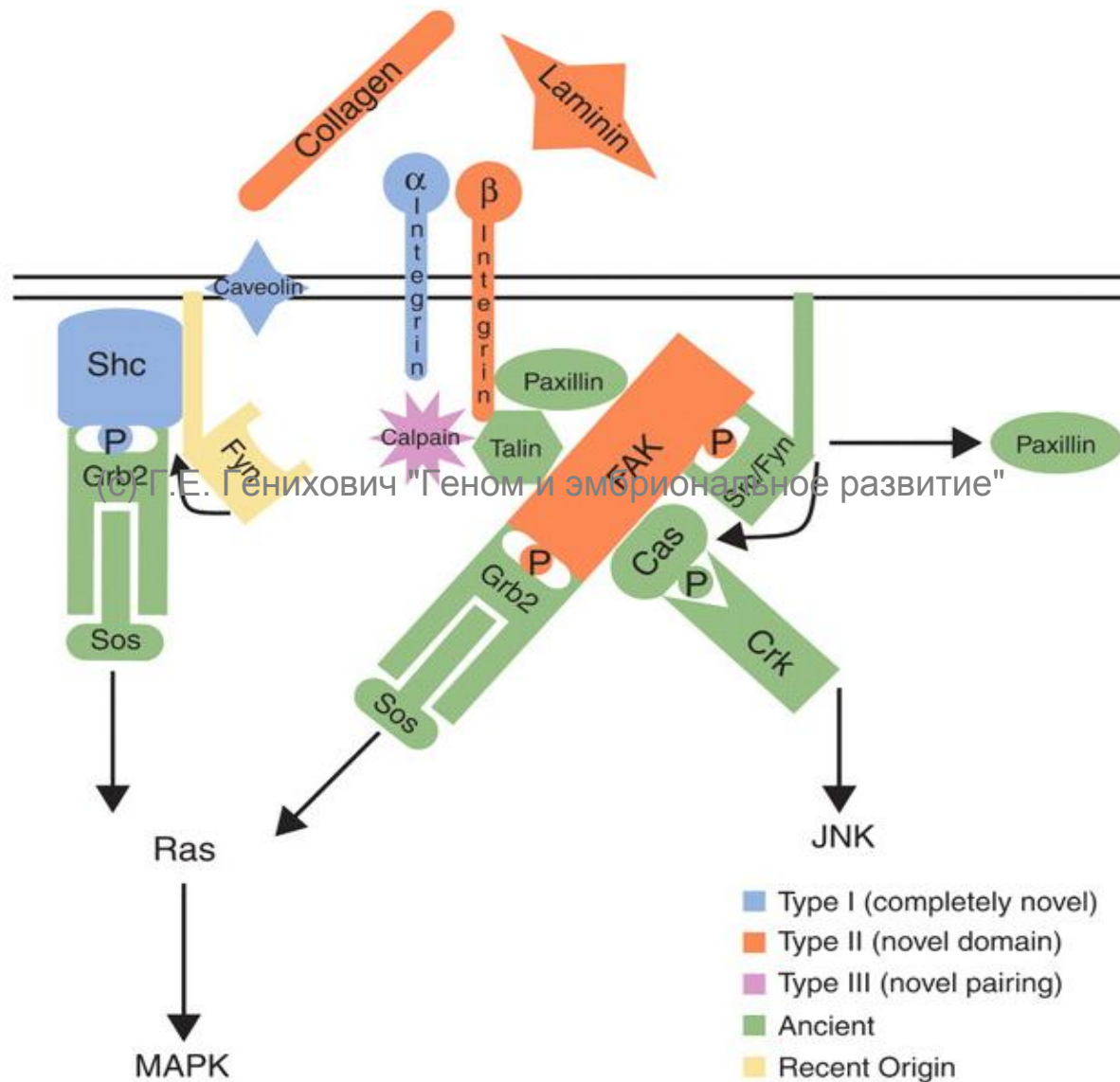
Геном нематостеллы позволяет оценить, как эволюционировали гены многоклеточных животных



(с) Г.Е. Генихович "Геном и эмбриональное развитие"

- Type I** Новые белковые домены: Wnt, SMAD, FGF, T-box
- Type II** Новые + древние домены: Notch = Notch + EGF
- Type III** Перетасованные древние домены: Islet = LIM + Homeobox
- Ancient** Гомологи белков, имеющих у растений, грибов и протистов

«Новые» и «древние» белки вовлекаются в одни и те же каскады



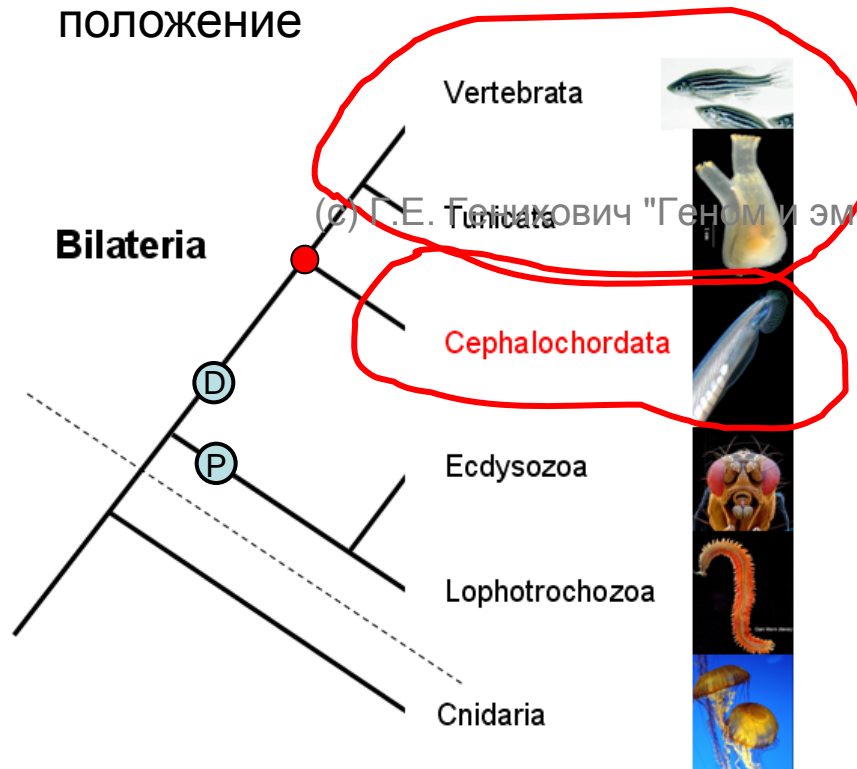
Г.Е. Генихович "Геном и эмбриональное развитие"

Evo/Devo: два важных для понимания эволюции генома



Ланцетник

- Морской житель, «идеальное хордовое»
- Копается в песке
- Важное филогенетическое положение



- Модельный объект

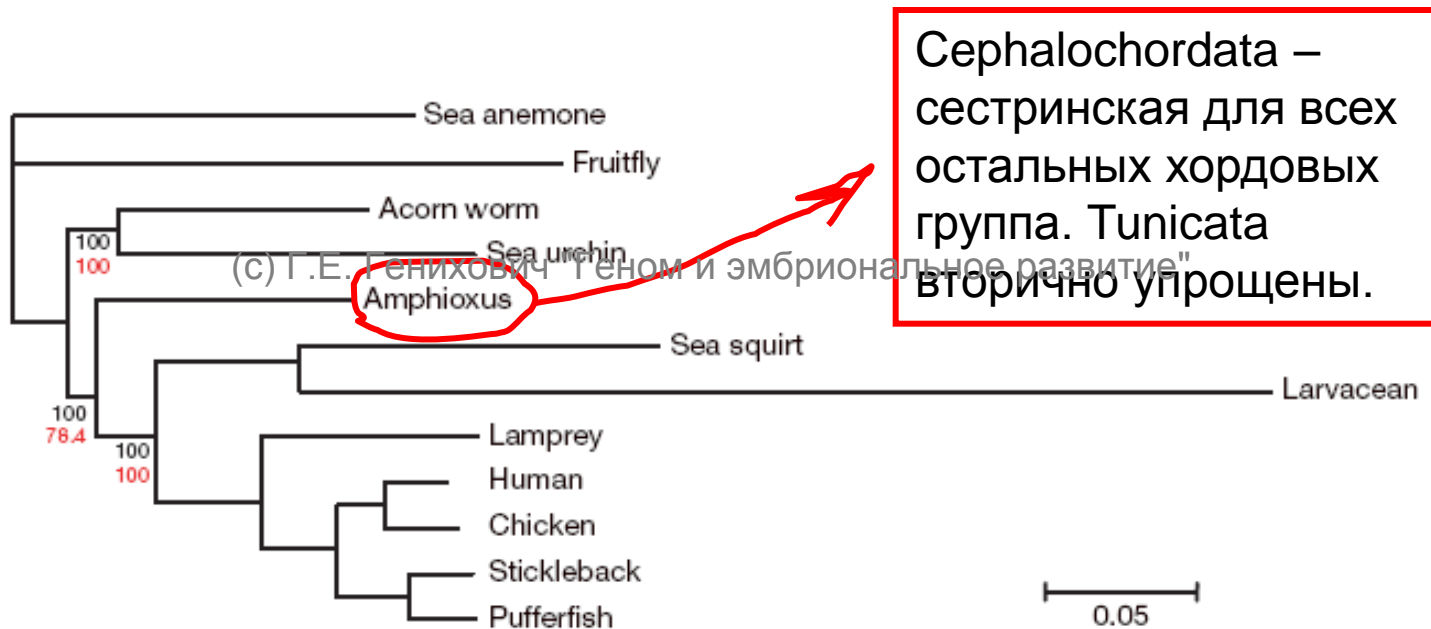


Branchiostoma floridae

- Shotgun sequencing, 11.5x покрытие
- Размер собранного генома 520 Mbp, половина – в скаффолдах >2.5 Mb, 95% имеющихся EST (> 480.000) выравниваются с геномом.
- Примерно треть последовательностей – минисателлиты длиной >100 kb. Общий размер собранного генома + повторов примерно 450 Mbp
- Число генов, кодирующих белки оценено как 21.900. Сравнение с другими хордовыми позволило определить 8437 семейств генов, имевшихся у общего предка Chordata. На эти семейства приходится 13.610 генов ланцетника, 13.401 генов человека и 7.216 генов *Ciona intestinalis*. *C. intestinalis*.
- 8 семейств генов, видимо, потерялось на ветке, ведущей к хордовым (есть у нематостеллы или дрозофилы и у морского ежа, но не у хордовых)
- 239 групп генов являются, вероятно, новшествами хордовых (есть у ланцетника и асцидий или позвоночных). 137 из них имеют характерные только для хордовых белковые домены, 10 – смесь более древних доменов и доменов, характерных для хордовых, 92 – перетасованные более древние домены.

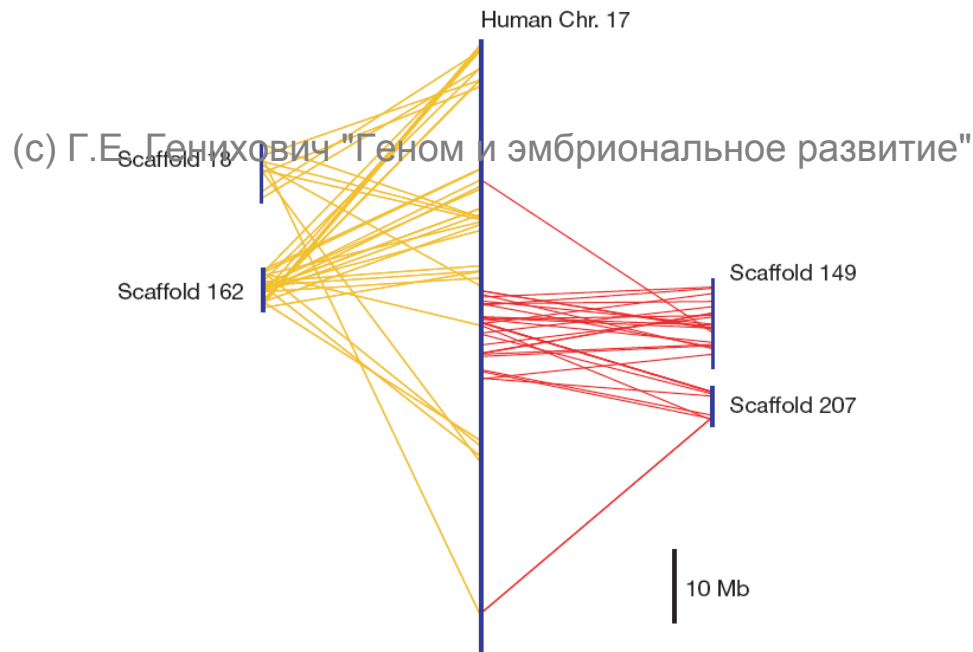
Branchiostoma floridae

- Взяли 1090 генов, имеющих как минимум одну копию в геномах всех рассматриваемых животных, и сделали дерево (Байезова статистика).



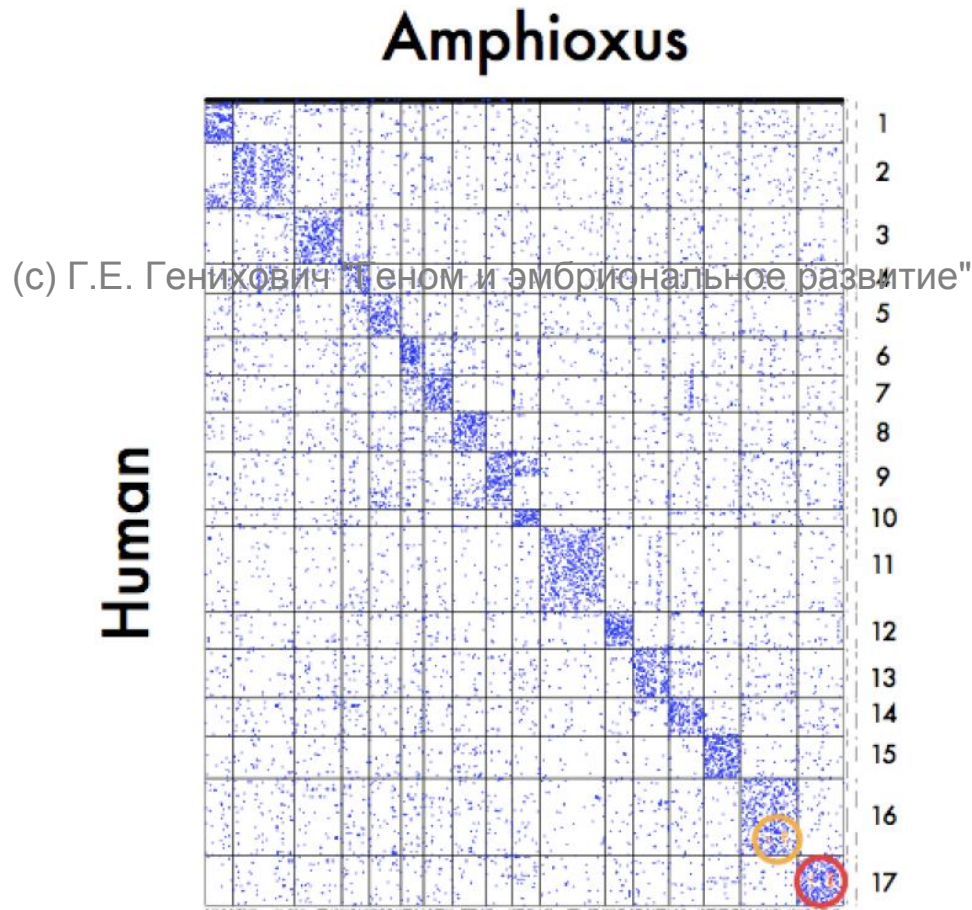
Branchiostoma floridae

- Анализ синтении показал наличие на хромосомах человека участков, соответствующих определенным скаффолдам ланцетника => человеческий геном был разделен на 135 сегментов, сохранивших древнюю синтению.



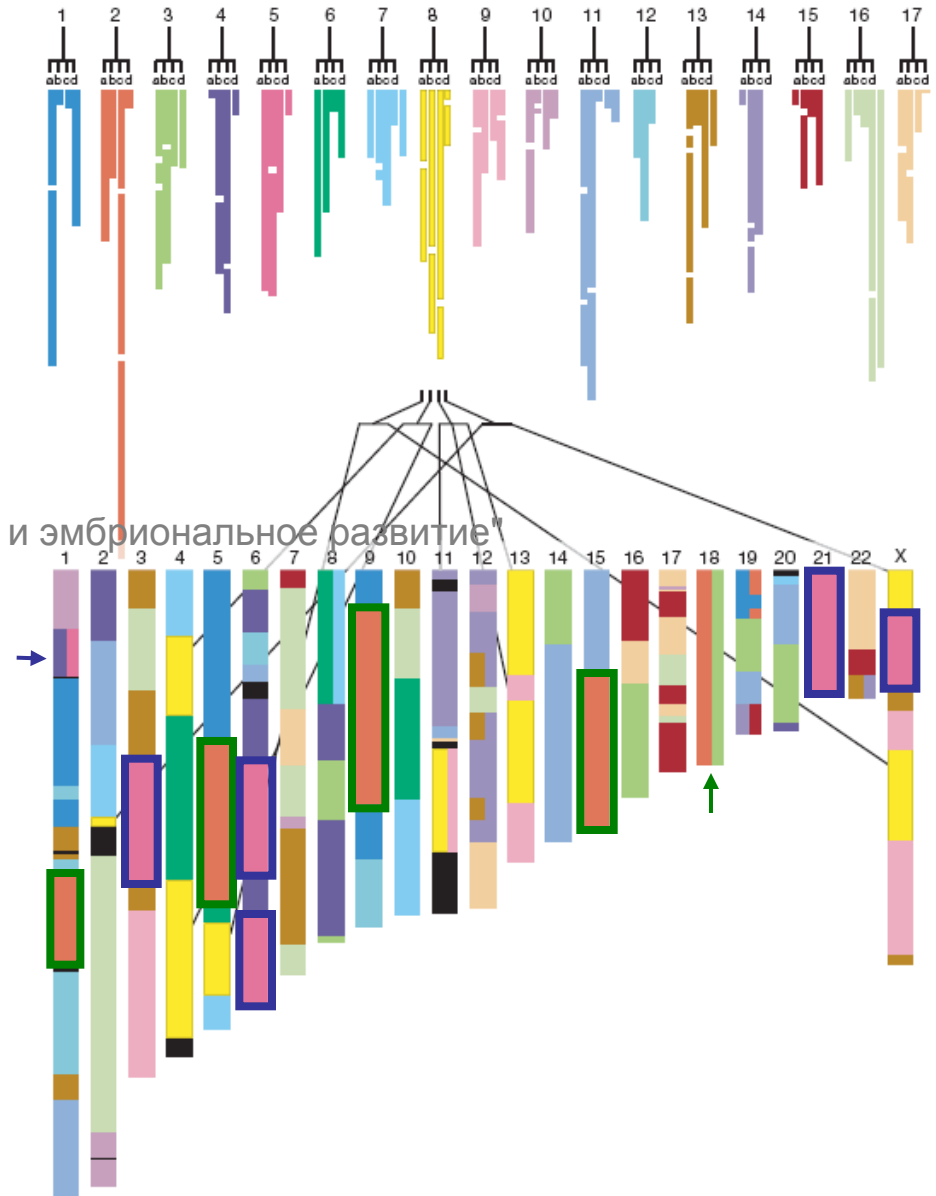
Branchiostoma floridae

- Когда сравнили, в каких позициях у человека и ланцетника находятся ортологичные гены, оказалось, что существует 17 крупных блоков, где гены имеют одинаковых соседей – 17 групп сцепления (CLG = chordate linkage group = предполагаемые 17 хромосом общего предка хордовых).



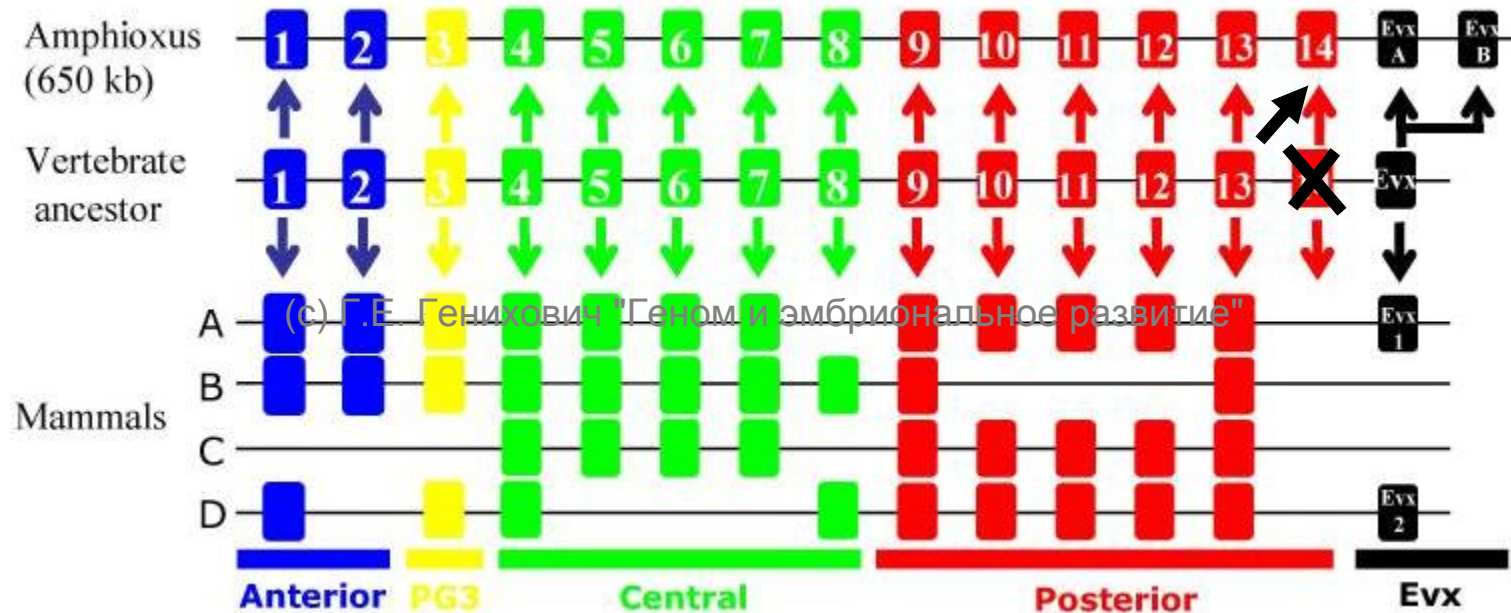
Branchiostoma floridae

- Для проверки взяли одну из таких гипотетических хромосом за No 15 и сделали FISH на хромосомах ланцетника со всеми 16-ю из скаффолдов, составляющих CLG15. 15 из 16 проб покрасили одну и ту же хромосому ланцетника, а шестнадцатая проба дала сильный фон.
- Распределение консервативных сегментов в хромосомах человека согласуется с представлениями о двух раундах дупликации генома у позвоночных.



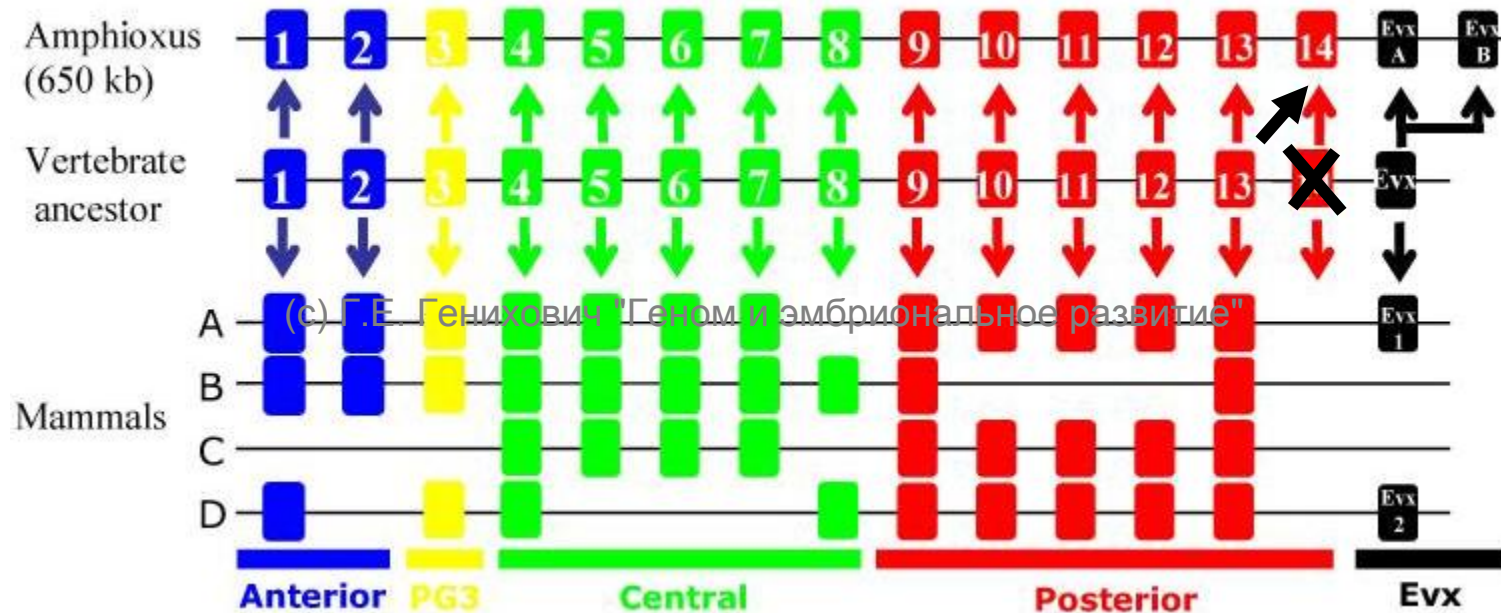
Пример про 2R: *Нох* гены

- 4 кластера *Нох* генов позвоночных – классический пример того, что происходит при дупликации генома.



Пример про 2R: *Нох* гены

- 4 кластера *Нох* генов позвоночных – классический пример того, что происходит при дупликации генома.



International Journal of Biological Sciences

ISSN 1449-2288 www.biolsci.org 2005 1:19-23

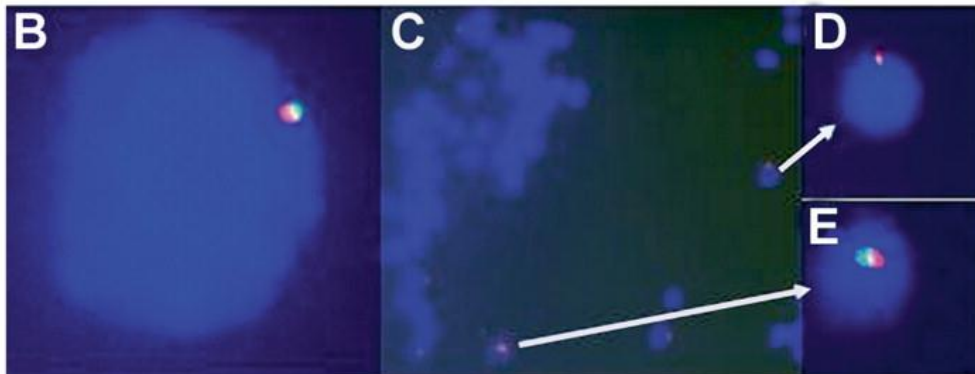
©2005 Ivyspring International Publisher. All rights reserved

No more than 14: the end of the amphioxus Hox cluster

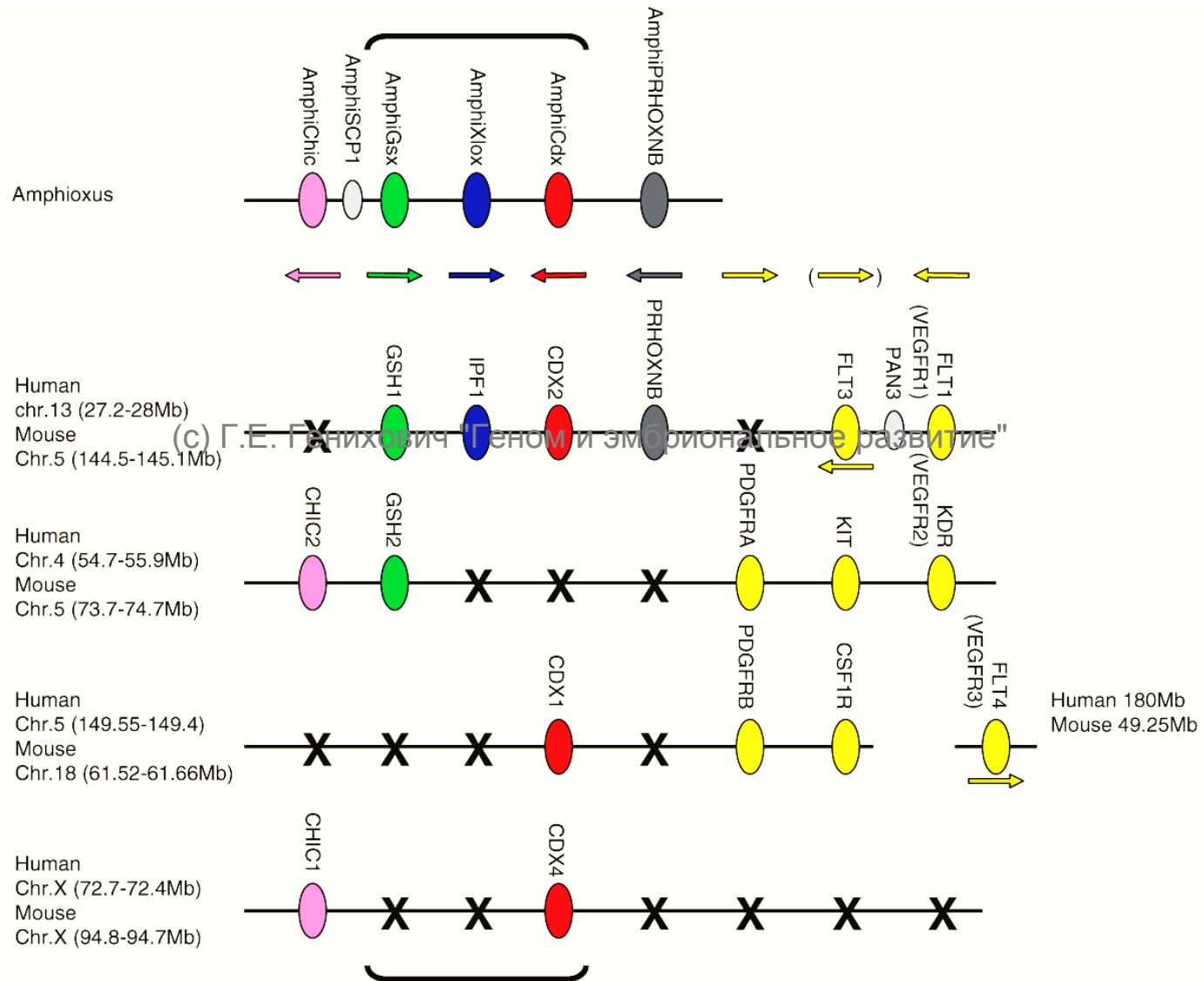
Carolina Minguillón^{1,4}, Josep Gardenyes¹, Elisa Serra¹, L. Filipe C. Castro², Alicia Hill-Force³, Peter W.H. Holland², Chris T. Amemiya³ and Jordi Garcia-Fernàndez¹

Пример про 2R: *Нох* гены

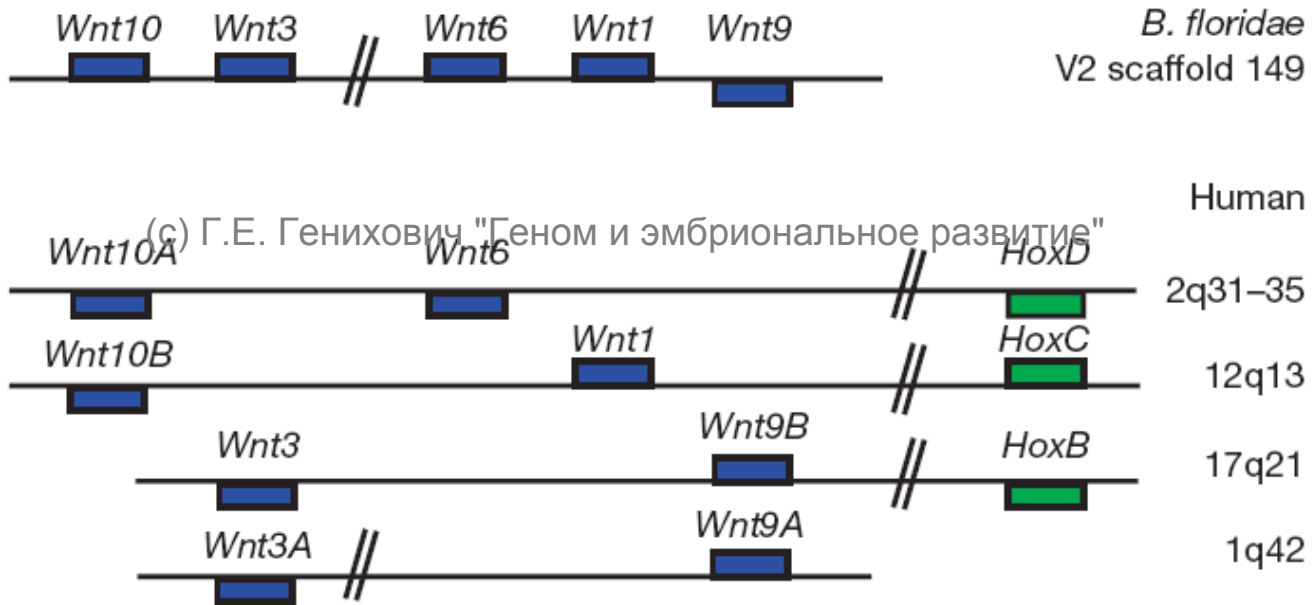
- При анализе генома у ланцетника обнаружили ранее незамеченный пятнадцатый *Нох* в кластере



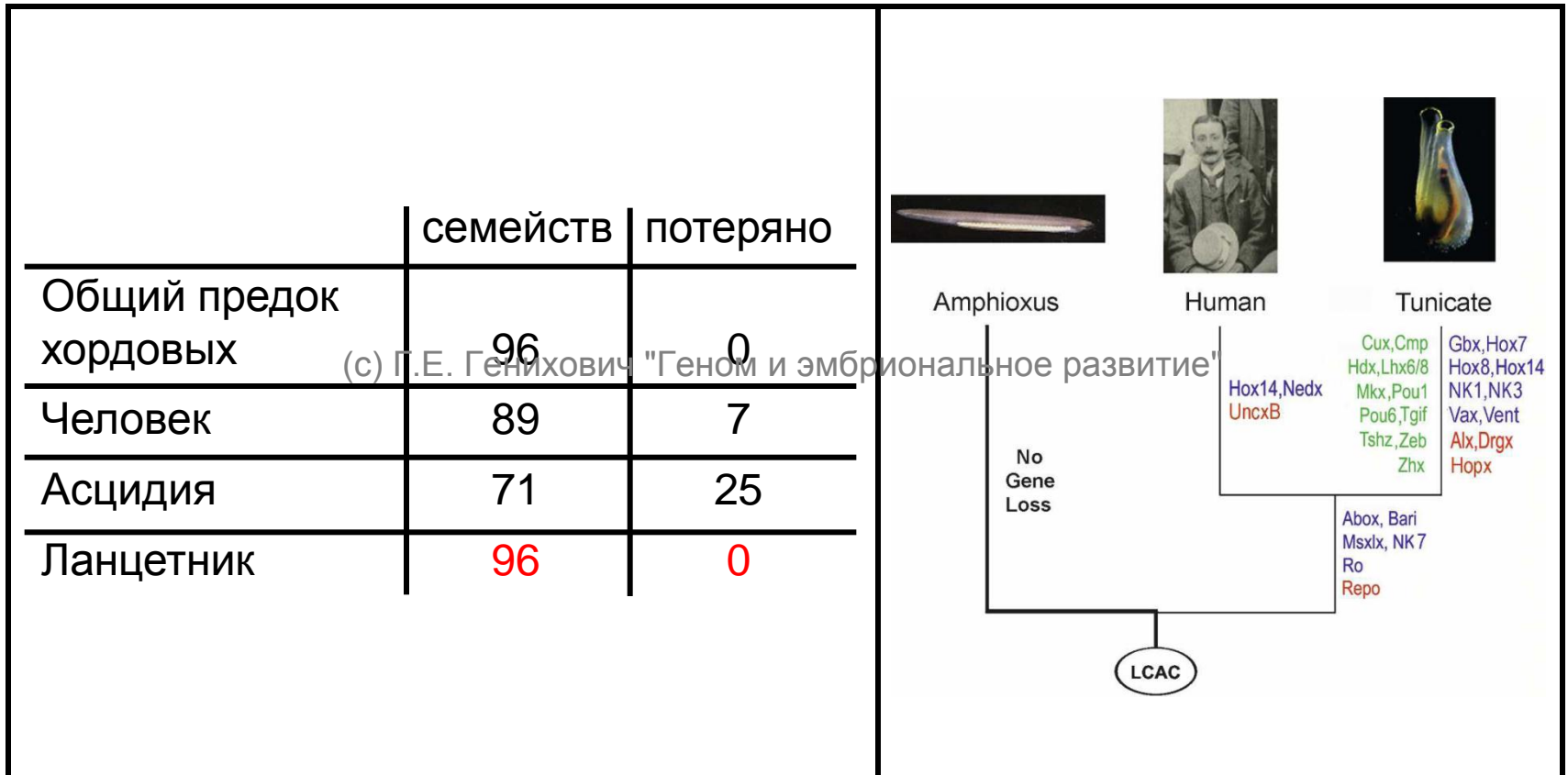
Пример про 2R: *ParaNox* гены



Пример про 2R: *Wnt* гены



Ланцетник практически не потерял гены



- Гомеобоксные гены не потерялись вообще

Заключительные соображения

- Секвенирование генома – мощный инструмент, дающий массу возможностей для специалистов из любых областей биологии. Чем дальше, тем более быстрым и недорогим станет секвенирование.
- Среди получаемых «бонусов» - бесценная эволюционная информация.
(с) Г.Е. Генихович "Геном и эмбриональное развитие"
- Знание последовательности нуклеотидов – лишь начало изучения особенностей работы генов вида. Самое важное – то, как гены регулируются, – из первичной последовательности не поймешь. Для этого существует множество дополнительных методов.

КОНЕЦ

(с) Г.Е. Генихович "Геном и эмбриональное развитие"



(с) Г.Е. Генихович "Геном и эмбриональное развитие"