

Чем будем заниматься?

Информатика и биоинформатика
для поступивших на ФББ в 2025 г.

Сергей Александрович Спирин

План

Первый семестр (осень I курса):

практическая информатика

Второй семестр (весна I курса):

биоинформатика (структуры и последовательности белков)

Третий семестр (осень II курса):

параллельно:

- биоинформатика (структуры и последовательности НК)
- язык R

Четвёртый семестр (весна II курса):

параллельно:

- биоинформатика (эволюция, семейства белков, ...)
- программирование на Python
- **подготовка и защита курсовой работы (!)**

Преподаватели

- Андрей Владимирович Алексеевский aba@belozersky.msu.ru
 - Сергей Александрович Спирин sas@fbb.msu.ru
 - Иван Сергеевич Русинов isrusinov@gmail.com , @isrusin (в ТГ)
 - Дарья Владимировна Диброва udavdasha@gmail.com
 - Дмитрий Дмитриевич Пензар dmitry_penzar_1996@mail.ru
 - Анастасия Александровна Жарикова azharikova@gmail.com
 - Екатерина Рюмина ryu-ekaterina@yandex.ru
 - Юлия Александровна Алёшина vjulia94@gmail.com
 - Иван Ильницкий @Ilnitsky (в ТГ)
 - Арсений Зинкевич @arsen_z (в ТГ)
- ... а также аспиранты и студенты старших курсов

Что такое биоинформатика?

Что такое биоинформатика?

Раздел парапсихологии, посвящённый информационным полям

Что такое биоинформатика?

~~Раздел парапсихологии, посвящённый информационным полям~~

Наука, изучающая потоки информации между живыми организмами и внутри них

Что такое биоинформатика?

~~Раздел парапсихологии, посвящённый информационным полям~~

~~Наука, изучающая потоки информации между живыми организмами и внутри них~~

Компьютерные методы в биологии

Что такое биоинформатика?

~~Раздел парапсихологии, посвящённый информационным полям~~

~~Наука, изучающая потоки информации между живыми организмами и внутри них~~

Компьютерные методы в биологии

*Биоинформатика — не отдельная наука, а совокупность методов!
Наука называется «биология»*

Что такое биоинформатика?

На самом деле по факту уже:

Биоинформатикой принято называть применение компьютерных наук в задачах молекулярной биологии

Что такое биоинформатика?

На самом деле по факту уже:

Биоинформатикой принято называть применение компьютерных наук в задачах **молекулярной** биологии

Причина этому объективная: именно в молекулярной биологии накоплено огромное количество экспериментальных данных, требующих обработки и интерпретации. Прежде всего это последовательности нуклеиновых кислот

Примеры задач биоинформатики

- Предсказание последовательностей белков по последовательности ДНК
- Предсказание пространственной структуры белка по его последовательности
- Предсказание функции белка по его последовательности (или по пространственной структуре)
- Реконструкция эволюции
- ... и очень много более частных (но не менее важных) задач

Видите последовательность, кодирующую белок?

```
gatctaccactgctaaggggaagtcactctaaattcttttgttaaatgtaccatccttcca  
gatacaagtaggaaaagtcgscagaagacaagagctgtagggaaaaccaccaaccctatc  
Ttcaaccacactatggatgagcaagtctgttccagcatttctccaagatgaggtgagtgg  
cagtgtgatgagtgtttatagtggagactttggcaatctggaagttaaaggaaatattca  
gtttgcaattgaatatgtggagtcactgaaggagttgcatgtttttgtggcccagtgtaa  
ggacttagcagcagcggatgtaaaaaaacagcgttcagaccatatagtaaaggcctattt  
gctaccagacaaaggcaaaatgggcaagaagaaaacactcgtagtgaagaaaaccttgaa  
tcctgtgtataacgaaatactgcggtataaaattgaaaaacaaatcttaaagacacagaa  
attgaacctgtccatttggcatcgggatacatttaagcgcataatagtttcctaggggaggt  
ggaacttgatttggaaacatgggactgggataacaacagataaacaattgagatggtac  
cctctgaagcggaaagacagcaccagttgcccttgaagcagaaaaacagaggtgaaatgaaa  
ctagctctccagtatgtcccagagccagtccttggtaaaaagcttcctacaactggagaa  
gtgcacatctgggtgaaggaatgcctttgatgtatgatgggttcaggcctgaagatctga  
tggaagcctgtgtagagcttactgtctgggaccattacaaattaaccaaccaattttgg  
gaggtcttcgtattggctttggaacaggtaaaagttatgggactga
```

... а она здесь есть!

```
gatctaccactgctaaggggaagtcatactaaattcttttgttaaatgtaccatccttcca  
gatacaagtaggaaaagtcgccagaagacaagagctgtagggaaaaccaccaaccctatc  
Ttcaaccacactatggatgagcaagtctgttccagcatttctccaagatgaggtgagtgg  
cagtgtgatgagtgtttatagtggagactttggcaatctggaagttaaaggaaatattca  
gtttgcaattgaatatgtggagtcactgaaggagttgcatgtttttgtggcccagtgtaa  
ggacttagcagcagcggatgtaaaaaaacagcgttcagacccatatgtaaaggcctat  
gctaccagacaaaggcaaaatgggcaagaagaaaacactcgtagtgaagaaaacctgaa  
tcctgtgtataacgaaatactgcggtataaaattgaaaaacaaatcttaaagacacagaa  
attgaacctgtccatttggcatcgggatacatttaagcgcfaatagtttcctaggggaggt  
ggaacttgatttggaaacatgggactgggataacaacagataaacaattgagatggta  
cctctgaagcggaaagacagcaccagttgcccttgaagcagaaaacagaggtgaaatgaaa  
ctagctctccagtatgtcccagagccagtccttggtaaaaagcttcctacaactggagaa  
gtgcacatctgggtgaaggaatgcctttgagtgtatgatgggttcaggcctgaagatctga  
tggaagcctgtgtagagcttactgtctgggaccattacaaattaaccaaccaatttttgg  
gaggtcttcgtattggcctttggaacaggtaaaagtattatgggactga
```

Чем занимаются биоинформатики?

Да кто чем...

Одни, сидя за компьютером, изучают содержимое публичных банков данных и результаты знакомых экспериментаторов.

При этом, разумеется, используют самые разные компьютерные программы.

Другие пишут программы для первых.

Третьи придумывают алгоритмы для таких программ.

Многие занимаются всем перечисленным попеременно.

И все всё время пишут статьи.

План первого семестра

Блок 1: введение

Нужно завести учётные записи, выбрать бактерию, на примере генома которой вы будете осваивать методы и т.п.

Блок 2: работа в командной строке Linux

Совершенно необходимый навык!

Блок 3: электронные таблицы

Иногда очень удобный инструмент, надо владеть

Блок 4: программирование на Python

Самые основы, более глубоко будете изучать в 4-ом семестре

Блок 5: мини-обзор генома бактерии

С самого начала нужно будет применять изучаемые методы к данным о геноме выбранной бактерии или археи. К концу семестра нужно будет подготовить и «защитить» описание полученных результатов

Другие семестры

Второй семестр:

- язык HTML, создание собственного веб-сайта
- визуализация пространственных структур белков
- банк последовательностей белков Uniprot
- выравнивание последовательностей гомологичных белков и поиск в банках белков, гомологичных данному

Третий семестр:

- пространственная структура нуклеиновых кислот
- банки последовательностей нуклеиновых кислот
- работа с результатами секвенирования

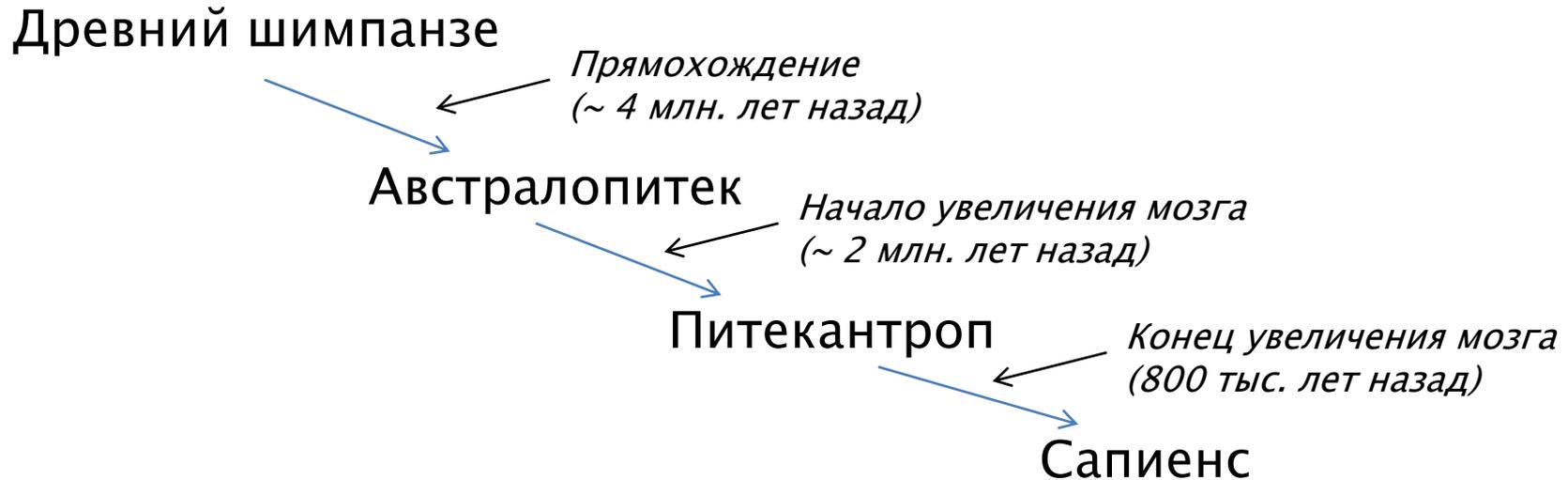
Четвёртый семестр:

- реконструкция филогении
- семейства белков и связанные задачи

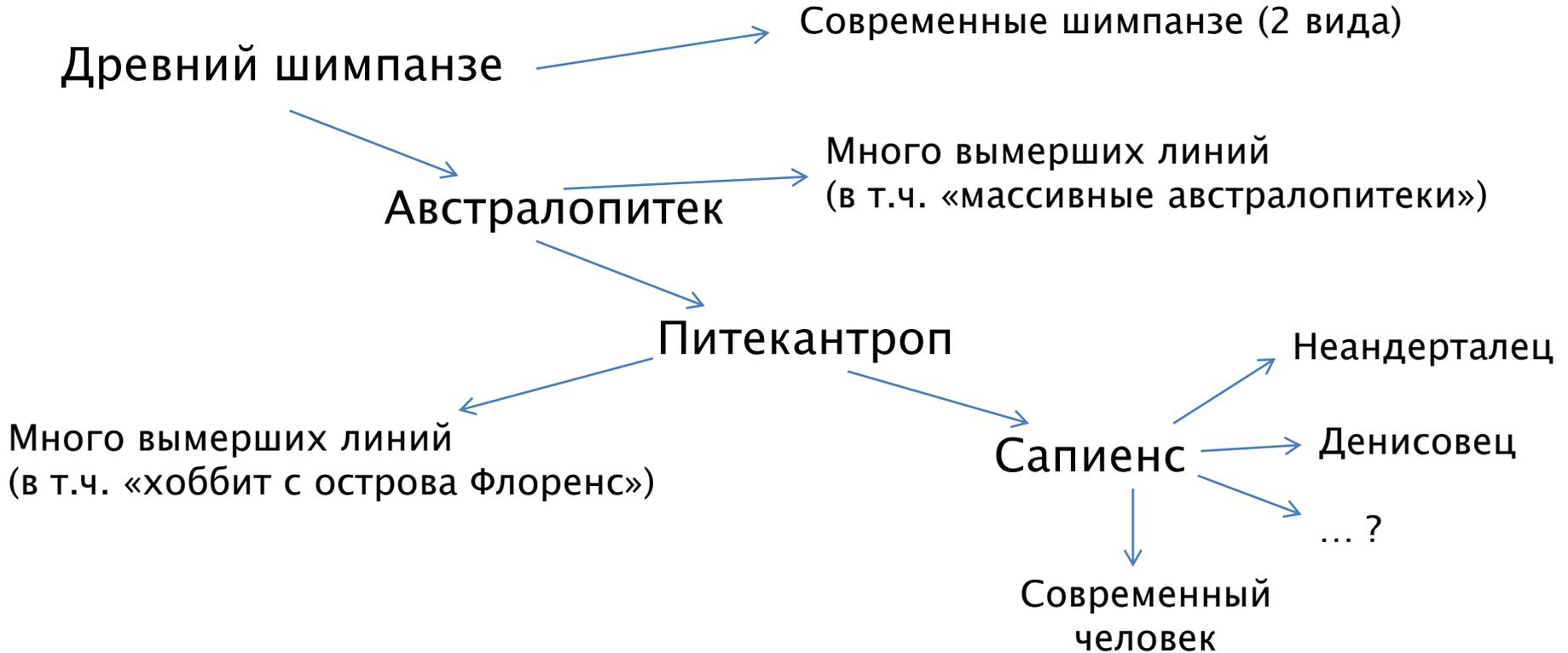
Немного об одной задаче

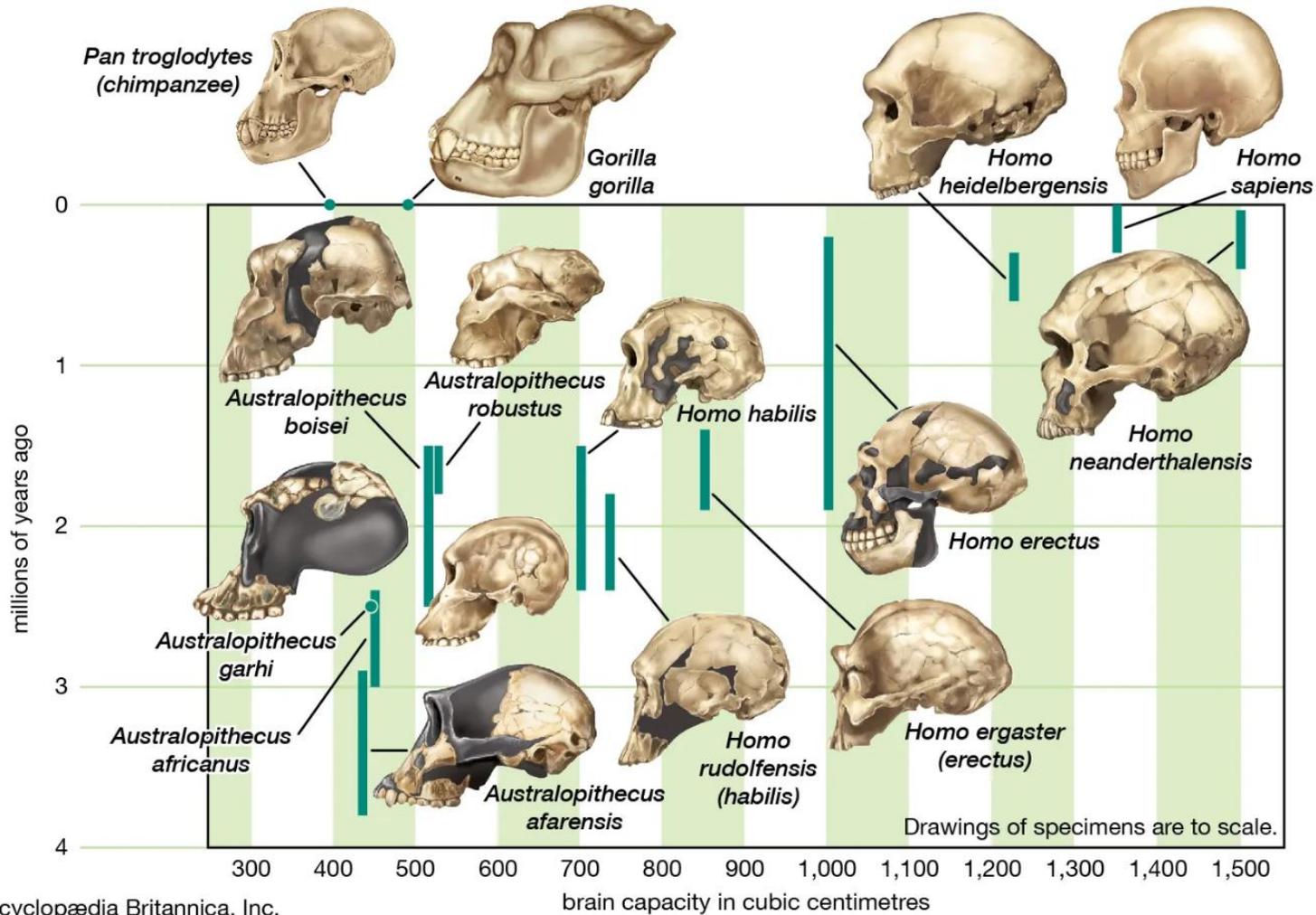
ПРО ГЕНОМЫ ДРЕВНИХ ЛЮДЕЙ

Происхождение человека (вкратце)



Разумеется, было много боковых ветвей...





© Encyclopædia Britannica, Inc.

<https://www.britannica.com/animal/Hominidae>

Древняя ДНК

- фрагментированная
- загрязнённая современной ДНК
- испорченная (химическими изменениями в основаниях)

Чем холоднее климат, тем лучше сохраняется ДНК.

Недавно научились секвенировать фрагменты древней ДНК. Самая древняя секвенированная ДНК— из гренландских доледниковых отложений возрастом 2 млн. лет (2022).

Самая древняя секвенированная ДНК из конкретного организма — из мамонта, замёрзшего в вечной мерзлоте 1 млн. лет назад (2021)

ДНК древних людей

- Митохондриальная ДНК неандертальца (1997)
- Митохондриальная ДНК из кости, найденной в Денисовой пещере в слое с отложениями примерно 40 000-летней давности (2010).
Анализ последовательностей выявил, что кость принадлежала представителю нового (под)вида человека, которого назвали денисовцем
- Ядерный геном неандертальца (2010)
- Ядерный геном денисовца (2012)
- Сейчас много кто ещё (самый древний — денисовец, живший ~200 000 лет назад)

Результат обработки

После очистки от загрязнений, исправления испорченных букв и т.д. прочтения сравниваются с референсным геномом современного человека и генерируется файл примерно такого вида:

#CHROM	POS	ID	REF	ALT
1	834832	.	G	C
1	846864	.	G	C
1	851204	.	G	C
1	852443	.	A	C
1	853267	.	G	T
1	853954	.	C	A
1	855075	.	C	G
1	856099	.	T	G
1	856329	.	C	G
1	857653	.	C	G
1	864726	.	T	A
1	868329	.	A	C
1	872091	.	T	G
1	872352	.	G	C
1	897325	.	G	C
1	901337	.	T	G
1	901806	.	G	C
1	909238	.	G	C
1	917060	.	C	G
1	918238	.	C	G
1	936210	.	C	A
1	961464	.	G	T
1	968226	.	C	A

Вставки ДНК древних людей в геномах современных

Авторы статьи 2010 г. о ядерной ДНК неандертальца сравнили её последовательность с геномами пяти современных людей и выяснили, что у трёх из них (из Франции, Китая и Новой Гвинеи) имеются участки геномов, унаследованные от неандертальцев, а у двух других (из Африки) таких участков нет.

Это можно было бы объяснить двояко: (1) среди предков секвенированных неандертальцев попадались предки современных евразийцев; (2) среди предков современных евразийцев попадались неандертальцы.

Авторы показали, что правильное объяснение — второе

A Draft Sequence of the Neandertal Genome

RICHARD E. GREEN, JOHANNES KRAUSE, ADRIAN W. BRIGGS, TOMISLAV MARICIC, UDO STENZEL, MARTIN KIRCHER, NICK PATTERSON, HENG LI, WEIWEI ZHAI, [...], AND

SVANTE PÄÄBO

+46 authors

[Authors Info & Affiliations](#)

SCIENCE · 7 May 2010 · Vol 328, Issue 5979 · pp. 710-722 · DOI: 10.1126/science.1188021

Direction of gene flow. A parsimonious explanation for these observations is that Neandertals exchanged genes with the ancestors of non-Africans. To determine the direction of gene flow consistent with the data, we took advantage of the fact that non-Africans are more distantly related to San than to Yoruba (73–75) (Table 4). This is reflected in the fact that $D(P, San, Q, chimpanzee)$ is 1.47 to 1.68 times greater than $D(P, Yoruba, Q, chimpanzee)$, where P and Q are non-Africans (SOM Text 15). Under the hypothesis of modern human to Neandertal gene flow, $D(P, San, Neandertal, chimpanzee)$ should be greater than $D(P, Yoruba, Neandertal, chimpanzee)$ by the same amount, because the deviation of the D statistics is due to Neandertals inheriting a proportion of ancestry from a non-African-like population Q. Empirically, however, the ratio is significantly smaller (1.00 to 1.03, $P \ll 0.0002$) (SOM Text 15). Thus, all or almost all of the gene flow detected was from Neandertals into modern humans.

Как так получилось?

- Гибридизация
- Потомки гибрида скрещиваются с людьми современного типа и постепенно почти перестают от них отличаться
- Кроссинговер приводит к тому, что отдельные неандертальские участки присутствуют в геномах потомков
Причём у разных потомков эти участки разные!

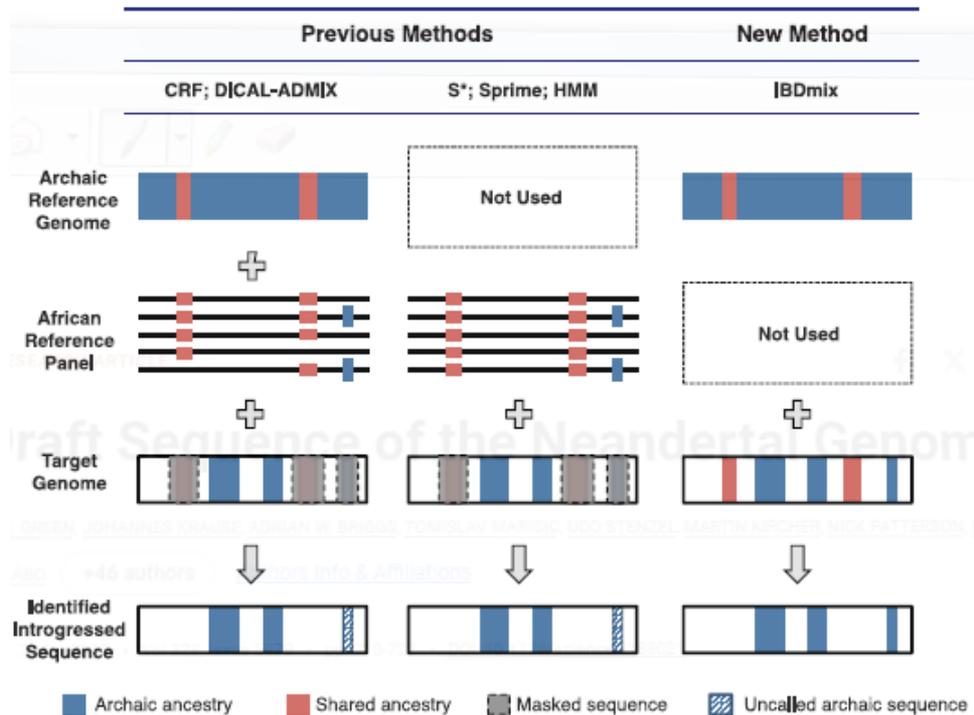
Что выяснилось за это время

Во-первых, появились гораздо более качественные результаты секвенирования многих образцов неандертальцев, денисовцев, и даже гибридов неандертальцев с денисовцами и неандертальцев с людьми современного типа

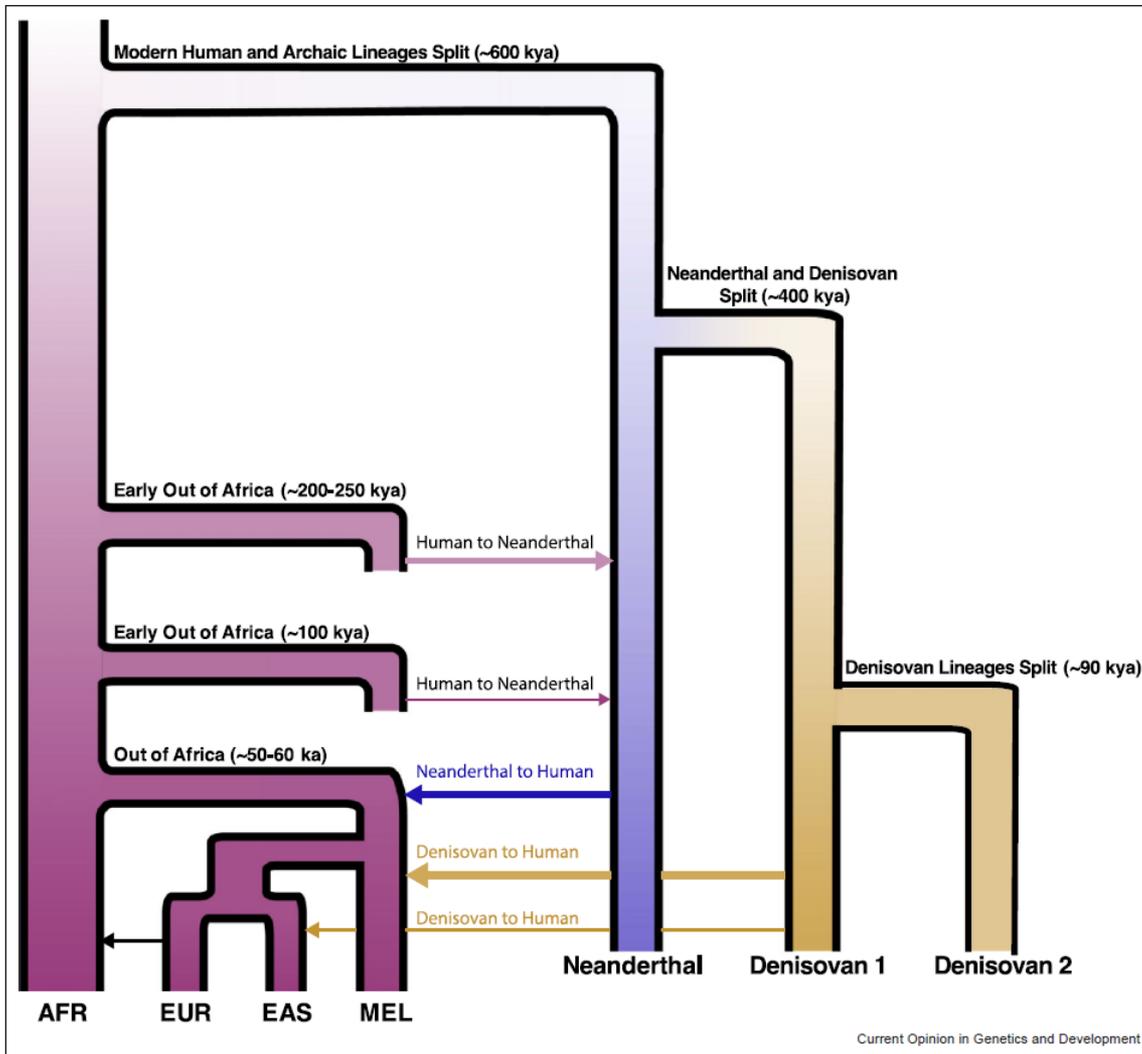
Во-вторых, были разработаны более совершенные методы обнаружения вставок

В-третьих, сейчас можно анализировать не пять, а тысячи геномов современных людей

Из статьи про программу IBDmix Chen et al., 2020, Cell 180, 677-687



Ныне это самая популярная программа для обнаружения интрогрессий в геномах (процитирована 248 раз)



D. Tagore & J.M. Akey
Archaic hominin admixture and its consequences for modern humans
 Current Opinion in Genetics & Development 2025, 90:102280

Судьба неандертальских вставок

Некоторые участки генома неандертальца оказались вредны для человека современного типа: они у современных людей не встречаются, что означает их элиминацию естественным отбором

Есть участки, оказавшиеся полезными, и поэтому закрепившиеся почти у всех евразийцев (некоторые такие участки содержат гены, связанные с цветом кожи)

Большинство участков нейтральны и встречаются у какой-то части современных евразийцев

Нобелевская премия 2022

Svante Pääbo

Facts



© Nobel Prize Outreach.
Photo: Stefan Bladh

Svante Pääbo

Nobel Prize in Physiology or Medicine 2022

Born: 20 April 1955, Stockholm, Sweden

Affiliation at the time of the award: Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; Okinawa Institute of Science and Technology, Okinawa, Japan; Leipzig University, Leipzig, Germany

Prize motivation: “for his discoveries concerning the genomes of extinct hominins and human evolution”

Prize share: 1/1