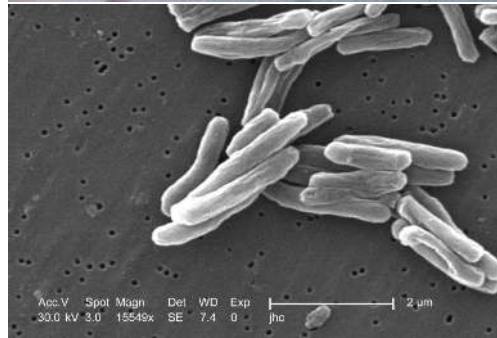


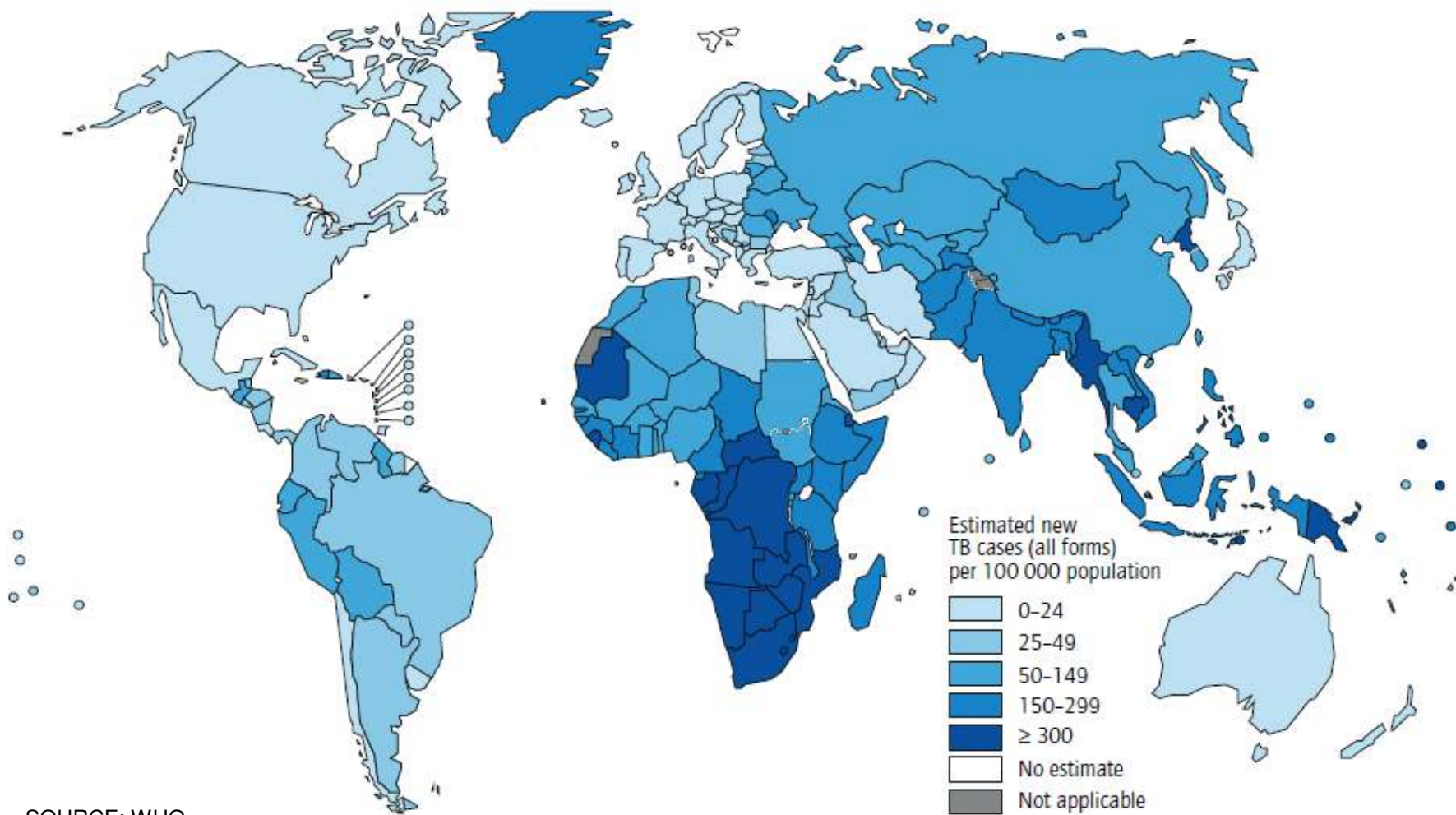
Анализ последовательностей
M. Tuberculosis и подход к организации базы
данных по мутациям лекарственной
устойчивости

Роман Сергеев



- **Туберкулёз** (от лат. *tuberculum* — бугорок) — широко распространённое в мире инфекционное заболевание человека и животных, вызываемое различными видами микобактерий
- Передаётся воздушно-капельным путём при разговоре, кашле и чихании больного
- Существует мнение, что *M. tuberculosis* инфицирована примерно 1/3 населения Земли
- Примерно 5-10% случаев скрытой инфекции переходит в активную форму

Estimated TB incidence rate, 2011

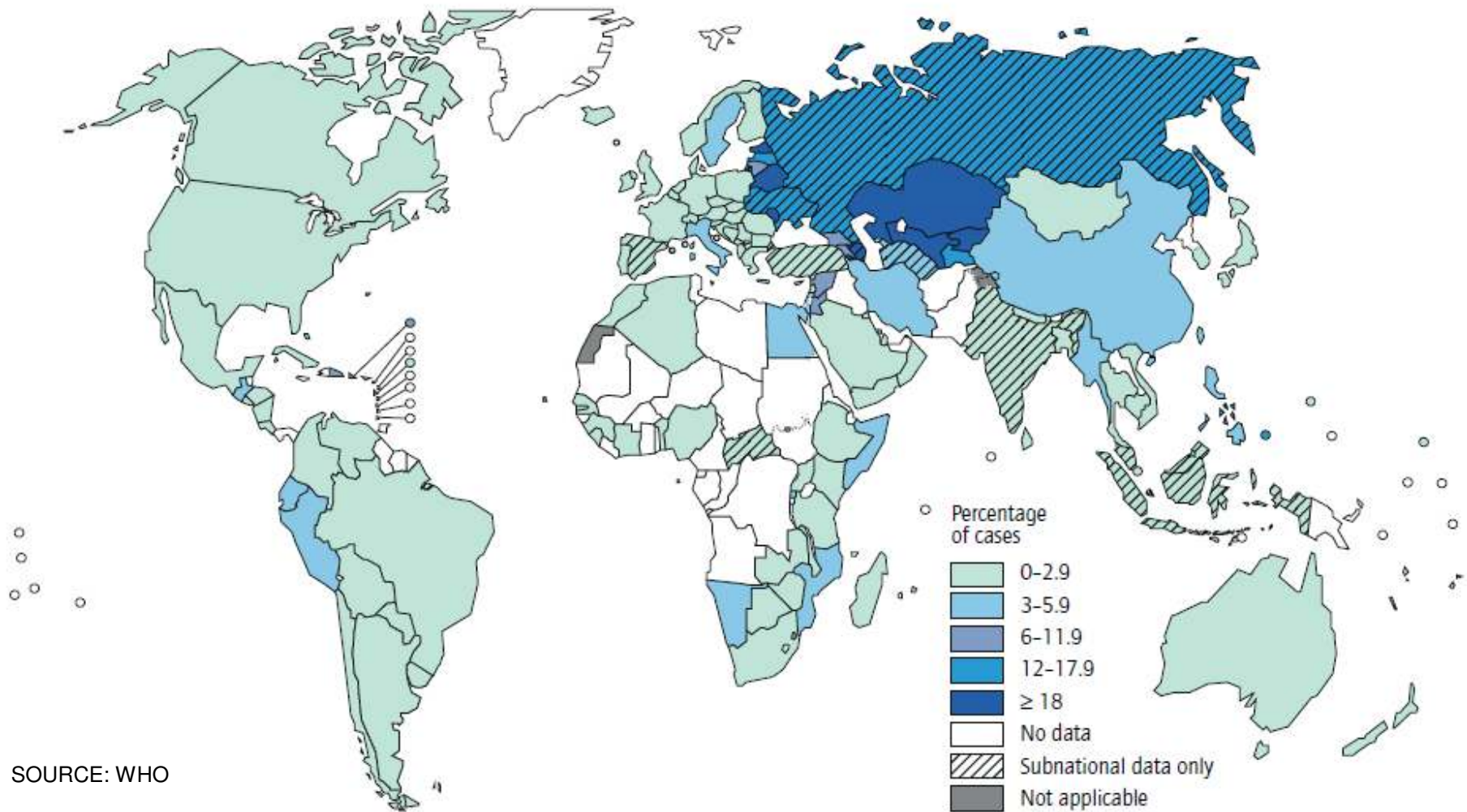


SOURCE: WHO

В 2011г. в мире **8,7** млн человек заболели и **1,4** млн человек умерли от ТБ

- **Монорезистентность** возникает в результате хромосомных мутаций среди небольшой части бактерий дикого штамма в условиях монотерапии. Вероятность возникновения 1 мутации в популяции $\sim 10^{-6} - 10^{-8}$.
- **Мультирезистентный туберкулез (MDR TB)**, мультиустойчивый к лекарственным средствам ТБ) вызывается штаммами, устойчивыми к препаратам основного ряда - изониазид и рифампицину.
- **Экстремально резистентный туберкулёз (XDR TB)** - MDR-TB, устойчивый также к фторхинолонам и препаратам резервного ряда (капреомицину, канамицину или амикацину). Вызывает особо тяжелую форму заболевания, практически не поддается лечению.

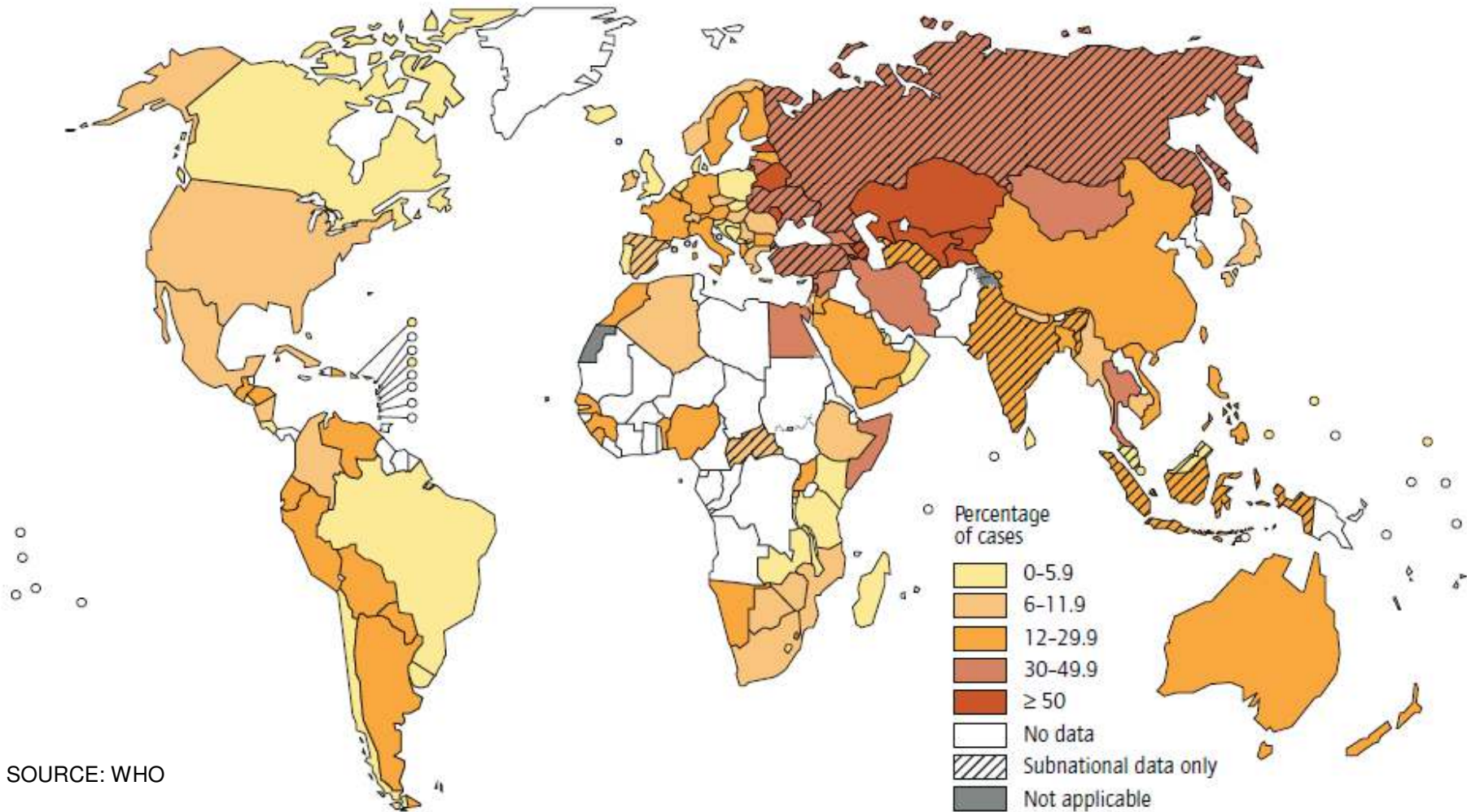
Percentage of new TB cases with MDR-TB



SOURCE: WHO

^a Figures are based on the most recent year for which data have been reported, which varies among countries.

Percentage of previously treated TB cases with MDR-TB



^a Figures are based on the most recent year for which data have been reported, which varies among countries.

Molecular mechanisms of drug resistance in *Mycobacterium tuberculosis*

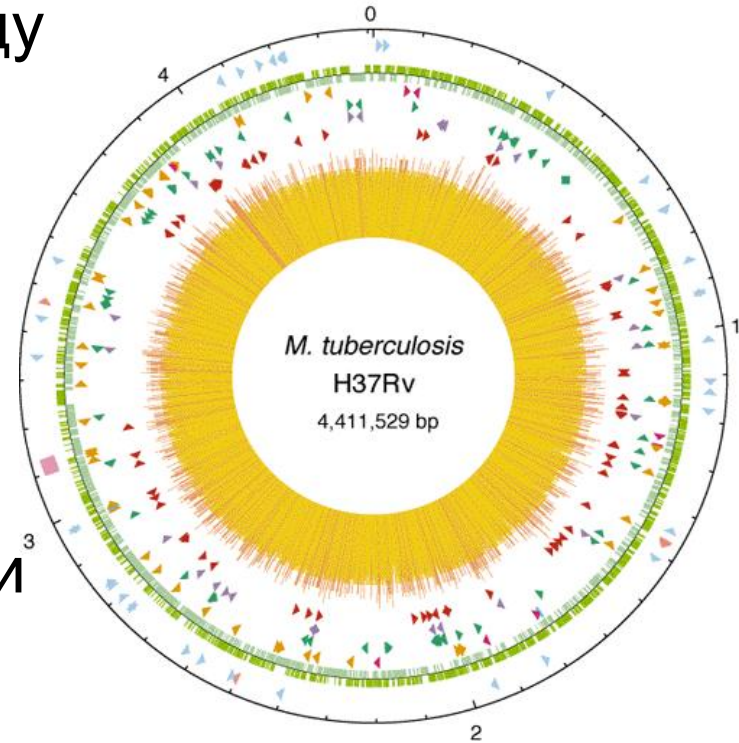
Drug	Genes	Target	Mechanism
Isoniazid	katG; inhA; ndh; ahpC	Catalase-peroxidase; Enoyl ACP reductase; NADH dehydrogenase II; Alkyl hydroperoxidase	Inhibition of mycolic acid biosynthesis and other effects on metabolism of lipids
Rifampicin	rpoB	RNA polymerase	Inhibition of transcription
Pyrazinamide	pncA	Nicotinamidase/pyrazinamidase	Acidification of cytoplasm and deenergizing membrane
Ethambutol	embCAB	arabinosyl transferase	Inhibition of arabinogalactan synthesis Inhibition of protein synthesis
Streptomycin	rpsL; rrs	S12 ribosomal protein; 16S rRNA	Inhibition of protein synthesis
Amikacin	rrs	16S rRNA	Inhibition of protein synthesis
Capreomycin	tlyA	16S rRNA	Inhibition of protein synthesis
Fluoroquinolones	gyrA; gyrB	DNA gyrase subunit A; DNA gyrase subunit B	Inhibition of DNA gyrase
Ethionamide	etaA/ethA; inhA	Flavin mono-oxygenase	Inhibition of mycolic acid biosynthesis

Препараты
первого
ряда

Препараты
второго
ряда

Геном *M. Tuberculosis*

- Ссылочная последовательность **H37Rv** секвенирована в 1998 году
- Размер генома **4,411,529** b.p. (уступает лишь *E.coli*)
- ~**4000** генов, распределены равномерно между двумя нитями
- **6** псевдогенов
- **3-4%** генома состоит из инсерционных последовательностей (IS) и пробактериофагов (*phiRv1*, *phiRv2*)



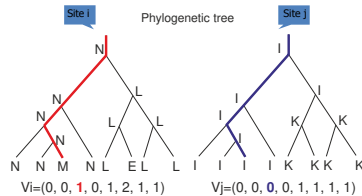
Задачи

- 1) Секвенировать последовательности *M. Tuberculosis* от пациентов с MDR и XDR-TB
- 2) Проанализировать последовательности, выявить мутации, сравнить с имеющимися в *dbSNP*
- 3) Как изменения в геноме связаны с развитием лекарственной устойчивости?

Как изменения в геноме связаны с развитием лекарственной устойчивости?

Ковариации в сиквенсах

```
TGCCTGTTTACGCGCCGATTGTTGC  
CAGGTGCCCGATGCGAGGTTGTTG  
CTAAAGTTCCTCACCCCGCACCAT
```



w

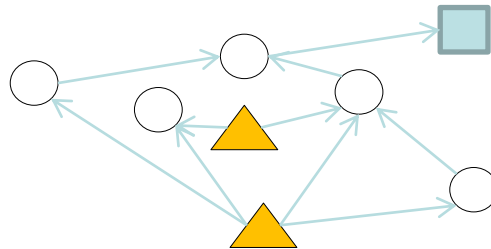
Ассоциации: генотип, фенотип, терапия

$$Y = X\beta + Zu + \varepsilon$$

w

w

Вероятностные сети зависимостей



Ковариации в сайтах выравнивания

Способ #1: Ассоциации из эволюционной информации

1 5 10 16

IRIIRIK**S****E****N**VVIQIK

-----E-----L-K-

-E---R---M-----

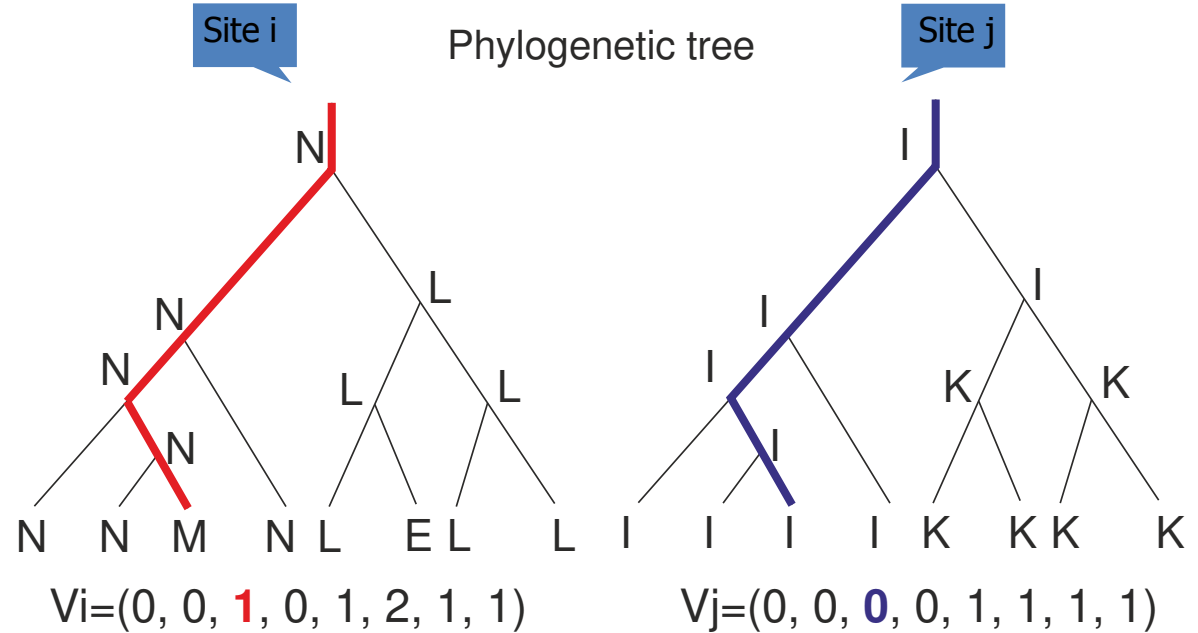
-----R-----

KI---R--**K****L**---S-

KE-----E--E--

K---L---L-----

K-----L---RR



Пример: Замена **I**->**K** в позиции 1 коррелирует с **N**->**L** в позиции 10

$$\rho_{ij} = \frac{\text{cov}(V_i, V_j)}{\sigma_{V_i} \cdot \sigma_{V_j}}$$

Ковариации в сайтах выравнивания

Способ #2: Ассоциации из выравнивания

Мера, основанная на взаимной информации:

$$MI_{ij} = H_i + H_j - H_{ij}$$

$$H_i = - \sum_{x_i \in S} P(x_i) \log P(x_i) \text{ - энтропия по Шеннону}$$

$$H_{ij} = - \sum_{x_i, y_j \in S} P(x_i, y_j) \log P(x_i, y_j) \text{ - взаимная энтропия}$$

Ассоциации с терапией

1) Точный тест Фишера для таблиц сопряженности:

Y\X	Yes	No	Σ
Yes	n_{11}	n_{12}	n_{1*}
No	n_{21}	n_{22}	n_{2*}
Σ	n_{*1}	n_{*2}	n

X - терапия проводилась {Y/N}
 Y - мутация произошла {Y/N}
 n_{ij} - число последовательностей,
где $X = i, Y = j$

2) Статистический тест на основе анализа частот несинонимичных/синонимичных замен (пакет PAML) – положительный отбор при $w > 1$

$$w = d_N / d_S$$

Ассоциации SNPs с фенотипом

3) Смешанная линейная модель (иерархическая регрессия):

$$Y = X\beta + Zu + \varepsilon$$

Y – вектор фенотипов

X – матрица SNPs (fixed effects)

β – вектор коэффициентов для fixed effects

u – вектор random effects (вектор СВ,

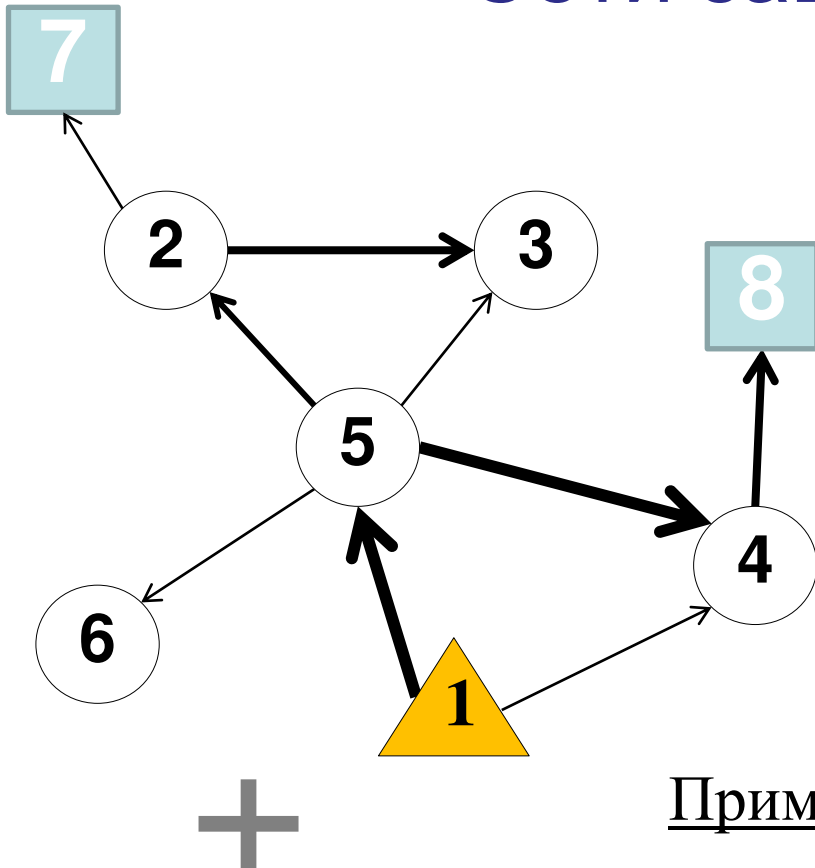
для которого $\text{Var}(u) = \sigma^2 K$)

K – kinship matrix (на основании матрицы сходства последовательностей)

ε – вектор ошибок измерений

Сети зависимостей

Методы из байесовских сетей

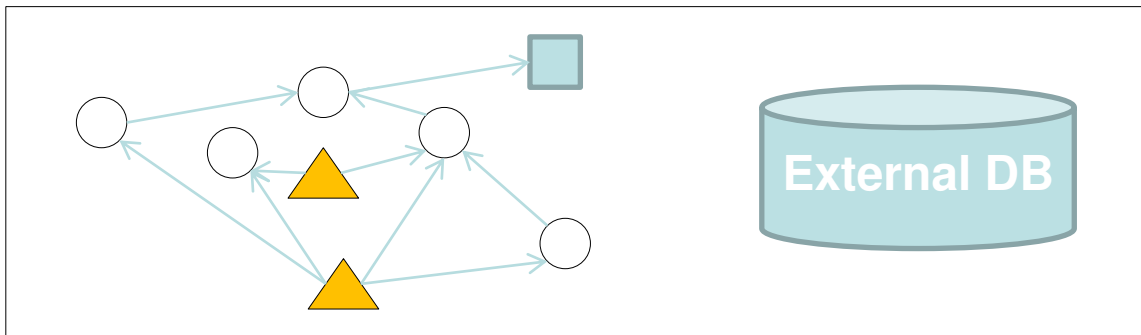


Evidence: $E = e$

Query: подмножество Y

Task: найти $P(Y | E=e)$

Пример: ? $P(Y8 | Y1 = y1, Y3=y3, Y2=y2)$



■ - resistant phenotype

○ - mutation

▲ - drug

Спасибо за внимание!

Роман Сергеев,

аспирант,

Объединенный институт проблем информатики,

Минск, Беларусь

roma.sergeev@gmail.com