

# Multifaktoriální dědičnost komplexních vlastností a genetická onemocnění

Tomáš Urban  
MENDELU



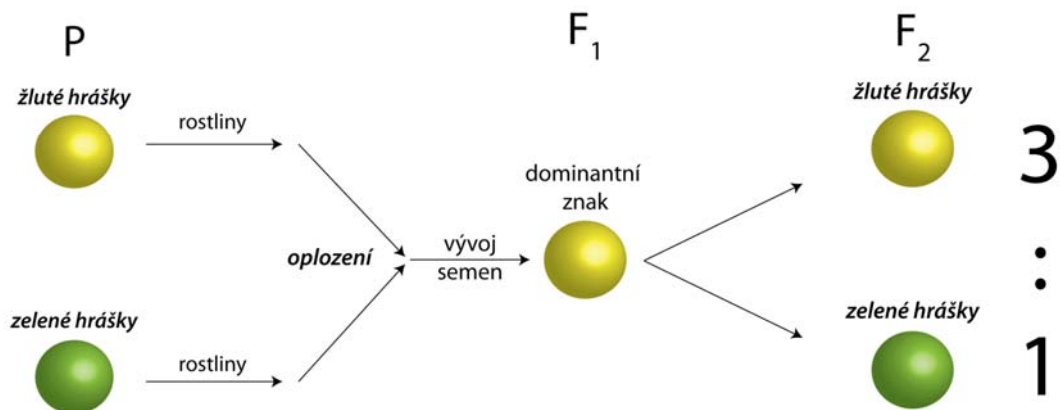
Agronomická  
fakulta

Mendelova  
univerzita  
v Brně



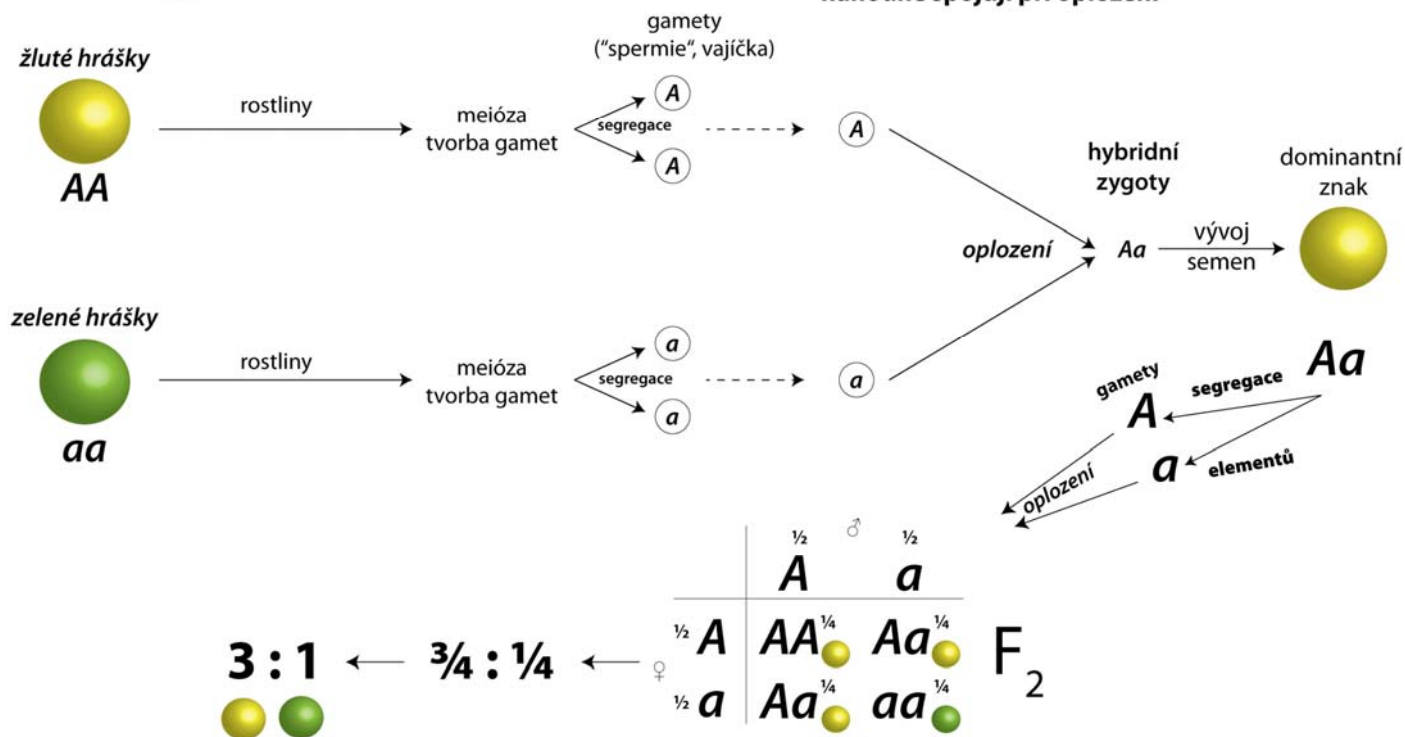
## Genetika a dědičnost

- Genetika – věda zabývající se dědičností a variabilitou
- Dědičnost – přenos genů mezi generacemi
- Variabilita – rozdílnost mezi jedinci v populaci (fenotypová < genetická)

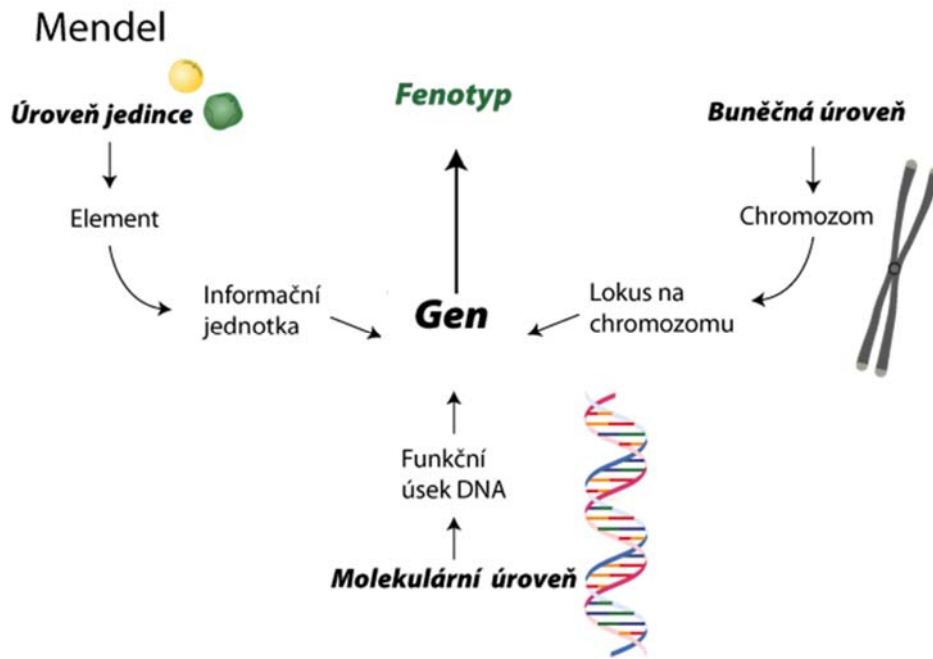


Dvě alely pro každou vlastnost se oddělují během tvorby gamet (anafáze I)

Dvě gamety, každá od jednoho rodiče, se náhodně spojují při oplození

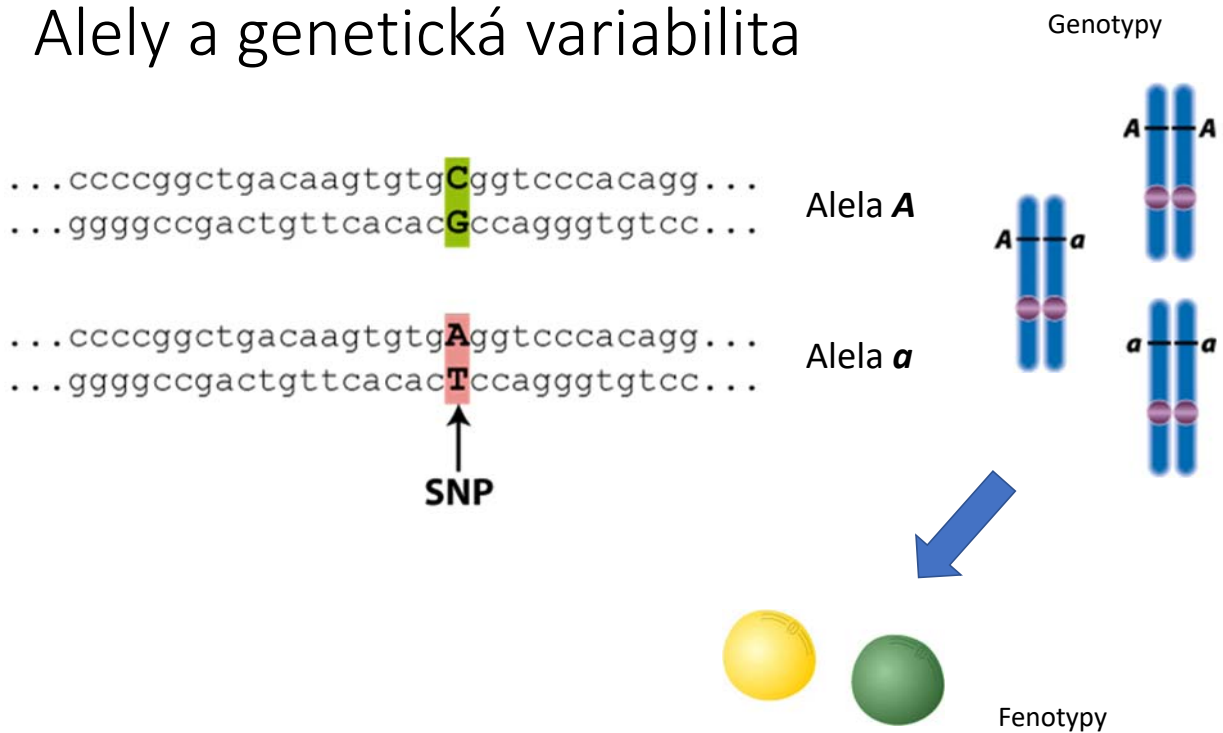


# Mendel, genetika a gen

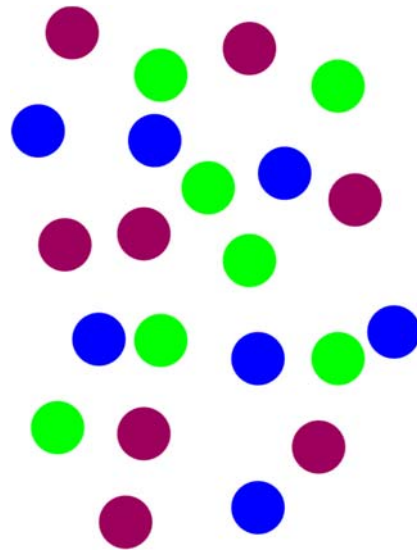


5

# Alely a genetická variabilita



Genetická variabilita jako podmínka evoluce,  
možnost se adaptovat na měnící se prostředí



## Složité, komplexní vlastnosti a dědičnost

- Velikost, obezita, dlouhověkost jsou velmi variabilní mezi jedinci a mají kontinuálně distribuované fenotypy, které nevykazují jednoduchou mendelistickou dědičnost ani diskrétní variabilitu
- **Kvantitativní genetika, genetika kvantitativních vlastností, genetika komplexních vlastností**
- Geny pro kvantitativní vlastnosti jsou „děděny“, tj. přenášeny z generace na generaci, platí pro ně *princip segregace a volné kombinovatelnosti* (pokud nejsou ve vazbě)
  - Jsou to diskrétní informační jednotky, mají svoji variabilitu...
- Hodnocení je rovněž komplexní – složité matematicko statistické přístupy hodnocení variability v populacích

## Která Mendelova pravidla jsou platná i pro dědičnost komplexních kvantitativních vlastností (QV)?

- Segregace – každý gen pro QV je v každém  $2n$  jedinci  $2\times$  (tvoří genotyp pro daný lokus) -> při **meióze I segregují** až do jednotlivých gamet
- Alely různých genů se mohou **nezávisle kombinovat** (pokud nejsou geny ve vazbě) – polygennost a volná kombinovatelnost
- NELZE však „jednoduše“ určovat efekt jednotlivých genů na fenotypovou variabilitu -> **black box**
- -> celkový fenotypový projev je dán efekty **mnoha genů (polygenní dědičnost)**, interakcí, epigenetických efektů... a **prostředí** a jejich interakcemi

## Co je „genetická choroba“

**Dědičné poruchy zdraví jsou geneticky zapříčiněné fyzické nebo funkční odchylky od normálního stavu s negativním vlivem na vitalitu.**

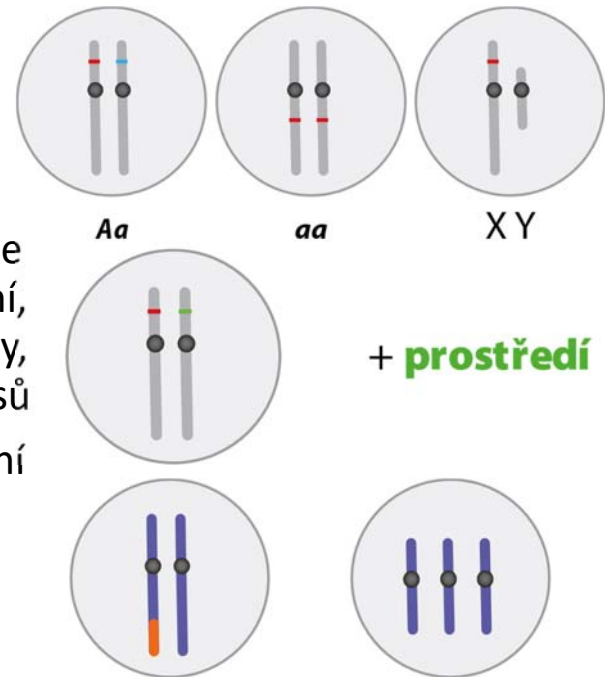
### Příklad u zvířat: Vznik dvojitého osvalení skotu

- Gen MYOSTATIN (*MSTN* – **BTA 2**) – protein (člen velké rodiny molekul zvaných transforming growth factors beta (TGF- $\beta$ ), někdy označovaný jako growth and differentiation factor-8 (**GDF-8**)
  - blokuje vliv myogeninu a tím zamezuje tvorbě dalších myoblastů, myofibril a svalových vláken
- **MUTACE MSTN = ztráta funkce blokády = větší množství svalových buněk, fibril a vláken**
  - Vznik **DVOJITÉHO OSVALENÍ** u skotu Belgického modro-bílého plemene, jeho systematické šlechtění od roku 1807, podmíněné genotypy recesivních homozygotů
- **20 – 25% nárůst svalové hmoty, 50% redukce lipidů (IMT), lepší konverze krmiva**
- **snížení velikosti vnitřních orgánů, snížení plodnosti u samic, oddálení pohlavní dospělosti, nižší životaschopnost potomků**

# Příčiny zodpovědné za genetické choroby

Genetická choroba vzniká v důsledku mutace a jejím přenosu (~ zdědění) na potomstvo.

- 1. Monogenní choroby** -> mutace v jednom genu (na jednom nebo obou chromozomech) – např. cystická fibróza, srpkovitá anémie
- 2. Multifaktoriálně dědičné choroby** -> kombinace variability mnoha genů a prostředí (kvantitativní, komplexní choroby) – choroby srdeční, cukrovky, většina rakovin, dysplazie kyčelního kloubu u psů
- 3. Chromozomální choroby** -> zvýšení nebo snížení počtu genů změnou počtu chromozomů, nebo strukturální změny v chromozomech: Downův syndrom, ...



## Monogenní onemocnění, modely dědičnosti a geny

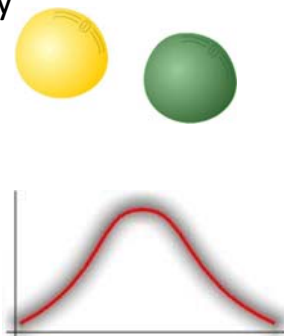
Disease	Type of Inheritance	Gene Responsible
Phenylketonuria (PKU)	Autosomal recessive	Phenylalanine hydroxylase ( <i>PAH</i> )
Cystic fibrosis	Autosomal recessive	Cystic fibrosis conductance transmembrane regulator ( <i>CFTR</i> )
Sickle-cell anemia	Autosomal recessive	Beta hemoglobin ( <i>HBB</i> )
Albinism, oculocutaneous, type II	Autosomal recessive	Oculocutaneous albinism II ( <i>OCA2</i> )
Huntington's disease	Autosomal dominant	Huntington ( <i>HTT</i> )
Myotonic dystrophy type 1	Autosomal dominant	Dystrophia myotonica-protein kinase ( <i>DMPK</i> )
Hypercholesterolemia, autosomal dominant, type B	Autosomal dominant	Low-density lipoprotein receptor ( <i>LDLR</i> ); apolipoprotein B ( <i>APOB</i> )
Neurofibromatosis, type 1	Autosomal dominant	Neurofibromin 1 ( <i>NF1</i> )
Polycystic kidney disease 1 and 2	Autosomal dominant	Polycystic kidney disease 1 ( <i>PKD1</i> ) and polycystic kidney disease 2 ( <i>PKD2</i> )
Hemophilia A	X-linked recessive	Coagulation factor VIII ( <i>F8</i> )
Muscular dystrophy, Duchenne type	X-linked recessive	Dystrophin ( <i>DMD</i> )
Hypophosphatemic rickets, X-linked dominant	X-linked dominant	Phosphate-regulating endopeptidase homologue, X-linked ( <i>PHEX</i> )
Rett's syndrome	X-linked dominant	Methyl-CpG-binding protein 2 ( <i>MECP2</i> )
Spermatogenic failure, nonobstructive, Y-linked	Y-linked	Ubiquitin-specific peptidase 9Y, Y-linked ( <i>USP9Y</i> )

# Kvantitativně založené choroby u lidí

- Vysoký krevní tlak
- Rakovina...
- Crohnova choroba a ulcerózní kolitida (ulcerative colitis)
- Chronické onemocnění ledvin (Chronic kidney disease)
- kardiovaskulární choroby
- obezita a diabetes typu 1 a 2
- ...

## Diskrétní x kontinuální variabilita

- Základní podmínkou evoluce a možnost se adaptovat ~ vývoje populací je VARIABILITA
- **Diskrétní variabilita** – buď a nebo – jednoduché znaky
  - Podmíněny málo geny
  - $P = G$  - jedinec
  - $V_P = V_G$  - populace
- **Kontinuální variabilita** – komplexní znaky
  - Polygeny
  - Prostředí
  - $P = G + E$  - jedinec
  - $V_P = V_G + V_E$  - populace



# Kvantitativní – komplexní - polyfaktoriální vlastnosti

## Zásadní problém hodnocení dědičnosti kvantitativních znaků!!!

- Jeden genotyp může v různém prostředí nabývat různých fenotypových hodnot!!!

## Genetika populací

- Popis genetické struktury populací a zákonitosti jejich změn
- Genetika populací
  - kvalitativních znaků
  - kvantitativních znaků (Genetika kvantitativních znaků)
- Jsou založeny na principech
  - Mendelistické genetiky
  - Matematicko-statistické metody



# Genetické parametry pro hodnocení populací

- Genetická variance
- Koeficient heritability (dědivosti)
- Koeficient opakovatelnosti
- Genetické korelace

## Celkový rozklad fenotypové variability

Jedinec

$$P = G + E$$

Výška Karla: 176 cm

Populace

$$V_P = V_G + V_E$$

Fenotypová **variance**  
výšky populace: 2896 cm<sup>2</sup>

$$V_X = \frac{\sum(X_i - \bar{X})^2}{n}$$

$$V_P = V_G + V_E + V_{GE} + 2cov_{GE}$$

$$V_P = V_A + V_D + V_I + V_E + V_{GE} + 2cov_{GE}$$

$$V_P = V_A + V_D + V_I + V_{EP} + V_{ET} + V_{GE} + 2cov_{GE}$$

CZ **Dědičnost**

ENG Heredity  
Inheritance  
GER Vererbung

×

**Dědivost**

Heritability  
Heritabilität  
Erblichkeit

$$h^2 = \frac{V_G}{V_P} = \frac{V_G}{V_G + V_E}$$

Geneticky  
kontrolovaná  
variance

- » proces přenosu genetické informace z rodiče na potomka
- » jev, kde biologická vlastnost se jeví že je přenášena z jedné generace do druhé; ve skutečnosti se přenáší dědičné informace o těchto vlastnostech (geny)
- » po přenosu se následně geny mohou realizovat, tj. projevovat (exprimovat) do fenotypu jedince
- » je o DNA jako takové

- » charakteristika ( **statistický parametr** ) vlastnosti populace, která se vyjadřuje jako část celkové fenotypové variability způsobená genetickou variabilitou (tj. na kolik jsou fenotypové rozdíly mezi jedinci v populaci zapříčiněny genetickými rozdíly)
- »  $h^2 = \frac{V_G}{V_P} = \frac{V_G}{V_G + V_E}$
- » kvantifikuje vliv genetické variability (množství rozdílných genotypů mezi jedinci v populaci) na fenotypovou variabilitu populace
- » je o rozdílnosti DNA v mnoha genech u mnoha jedinců a jejím vlivu na variabilní fenotypy v populaci
- » lze ji využít pro predikci odpovědi na selekci (genetického zisku)

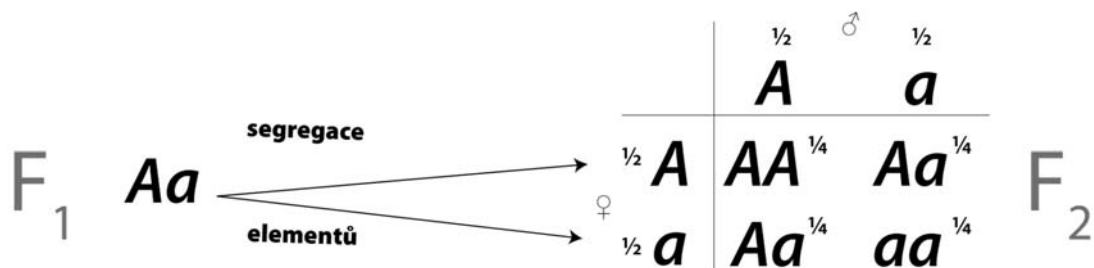
Dědičnost	
Úroveň	Proces
Gen	~ replikace DNA ~ transkripce a translace
Buňka	~ dělení buňky (mitóza, meióza), segregace, volná kombinovatelnost, rekombinace
Jedinec	~ projevený fenotyp

Populace	~ rozdílné varianty mnoha genů podmiňující jednu vlastnost
Dědivost	~ vztah genetické a prostředí variability k fenotypové variabilitě

$$h^2 = \frac{V_G}{V_P}$$

Populační genetický parametr – dědivost/heritabilita

# Dědivost není dědičnost!



## Heritabilita v širším smyslu

$$H^2 = \frac{V_G}{V_P}$$

- Množství variance přisouzené genetickým faktorům, pokud geneticky identičtí jedinci (klony) mohou být vystaveni různorodému prostředí (tj. množství variance vysvětlené klony ve všech prostředích). Ve studiích lidí tomu odpovídá studie MZ dvojčat vychovávaných odděleně.

## Heritabilita v užším smyslu

$$h^2 = \frac{V_A}{V_P}$$

Additivně genetické hodnoty A  
Součet průměrných efektů alel jedinců, které jsou přeneseny na potomstvo a v kontextu onemocnění jako průměrné riziko onemocnění jedince, který jej přenáší na potomstvo.

$V_A$  – aditivně genetická variance

## Základy hodnocení genetické variability

- Statistické metody byly zavedeny již R. Fisherem (1918) a S. Wrightem (1921) – **analýza variance** (pro rozdělení celkové fenotypové variance a odhad genetické variance), **úsekové koeficienty** (pro popis podobnosti mezi příbuznými jedinci) - > **kvantitativní genetik**
- **CO musíme mít k dispozici?**
  - Fenotypová data
  - Dostatečně početné příbuzenské skupiny
  - Vhodné statistické metody

	Typ podobnosti rodiny								
	Žádná			Střední			Vysoká		
Otec	1	2	3	1	2	3	1	2	3
Hodnota jedinců	1	1	1	3	2	1	1	2	3
	2	2	2	2	1	3	1	2	3
	3	3	3	1	2	3	1	2	3
Var mezi rodinami	Nízká			Střední			Velká		
Var v rodinách	Velká			Střední			Nízká		
<b>h<sup>2</sup></b>	<b>nízká</b>			střední			<b>vysoká</b>		

Celková var = var mezi rodinami + var v rodinách

Stená jako cov v rodinách

## Př. Otcovský (Sire) model – 1 f ANOVA

- Odhad korelace polosourozenců

$$y_{ij} = \mu + a_i + e_{ij}$$

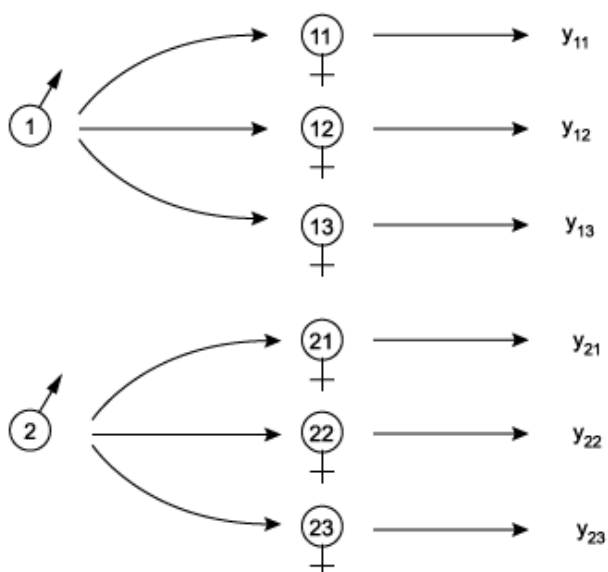
$$V_S = \sigma_S^2 = \frac{1}{4} \sigma_A^2$$

$$V_e = \sigma_e^2 = \frac{3}{4} \sigma_A^2 + \sigma_E^2$$

$$\sigma_y^2 = \sigma_S^2 + \sigma_e^2$$

- předpoklad, že otcové (sire) a matky (dam) jsou nepříbuzní, náhodně páření, bez selekce
- balancovaný design:  $p$  otců (sire) pářeno s  $n$  matkami (dam)  $\Rightarrow$  1 potomka

### Design pokusu pro Sire model



Variance mezi skupinami polosourozenců = kovarianci mezi polosourozenci ve skupině

$$\text{COV}_{(\text{polos.})} = \text{COV}(y_{ijr}, y_{ik}) = \sigma_S^2$$

$$4\sigma_S^2 = \sigma_A^2$$

To lze pomocí ANOVA odhadnout

### Sire model – tabulka 1 f ANOVA

Zdroj proměnlivosti	df	SS	MS / Var	E(MS)
Mezi rodinami (mezi otci)	$p - 1$	$SS_s = \sum_{i=1}^p n_i (\bar{y}_i - \bar{y})^2$	$MS_s = \frac{SS_s}{(p-1)}$	$\sigma_e^2 + n_0 \sigma_g^2$
V rodinách (reziduální)	$n - p$	$SS_e = \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	$MS_e = \frac{SS_e}{(n-p)}$	$\sigma_e^2$
Celkem	$n - 1$	$SS_c = \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$	$MS_c = \frac{SS_c}{(n-1)}$	

$$MS_e \doteq \sigma_e^2$$

$$n_0 = \frac{n - (\sum n_i^2 / n)}{n - 1}$$

$$MS_a = \sigma_e^2 + n_0 \sigma_g^2 = MS_e + n_0 \sigma_g^2$$

$$\sigma_g^2 = \frac{MS_a - MS_e}{n_0}$$

Intraklasní korelační koeficient

$$\rho = r_i = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

$$h^2 = 4\rho = 4 \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 4 \frac{\sigma_g^2}{\sigma_p^2}$$

## Hodnocení genetické variability u kvantitativních vlastností

- **Klasický přístup** – změřené fenotypy jedinců se známými příbuzenskými vztahy; genetické a prostředové zdroje fenotypové variance jsou určeny statistickými metodami – lineární modely (ANOVA, BLUP, ML,...)
  - Příbuzní jedinci sdílejí určitý podíl společných genů (to je kvantifikováno koeficientem příbuznosti  $r_{ij}$ )
  - $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$
- **Molekulárně statistický přístup** – nejnovější technika mapování lokusů kvantitativních vlastností (**QTL** – quantitative trait loci) na základě celogenomových markerů SNP (sekvenování genomů), GWAS a detekování oblastí QTL až identifikace konkrétních genů
  - Hledají se sekvence DNA pomocí genetických markerů po celém genomu a jejich statistické asociace k fenotypové variabilitě, popř. expresi ([opět má slovo statistika](#) – **vyhodnocení pomocí statistických modelů**)

# Typy kvantitativních chorob

- Lidské komplexní nemoci jsou vysoce heterogenní
- Mnohé nemoci jsou **dichotomické** a jsou měřeny v binární nebo 0/1 měřítku - to znamená, že onemocnění buď chybí (0), nebo je přítomno (1)
- Některá onemocnění však vykazují **různé úrovně závažnosti** onemocnění (například strukturování a zánětlivá ulcerózní kolitida), což umožňuje zařazení do většího počtu fenotypových tříd
- Onemocnění může být také definováno tím, že se stanovuje **prahová hodnota** pro kontinuální vlastnost, aby se vytvořila diskontinuita (například osteoporóza nebo hypertenze)

## Heritabilita pro dichotomické nemoci

- Fenotypová variance závisí na průměru vlastnosti (tj. rozšíření choroby) a metody k odhadu heritability pro kontinuální kvantitativní vlastnosti nejsou přímo využitelné
- Proto se využívá odhad **heritability pro normálně distribuovanou náchylnost k nemoci** (liability) ( $h^2_x$ ) - odlišná od heritability nemocí v pozorovaném rozsahu 0-1
- 4 hlavní přístupy odhadu heritability pro dichotomické nemoci
  - Korelace stavu onemocnění v párech příbuzných jedinců v náhodném vzorku populace (**tetrachorické korelace**)
  - **Metoda dvojčat**, porovnává podobnost mezi monozygotními (MZ) dvojčaty a dizygotními (DZ) dvojčaty
  - **Falconerova metoda**, stanovení korelace náchylnosti k nemoci mezi příbuznými z výskytu nemoci v celkové populaci a mezi příbuzným s postiženými jedinci
  - **zobecněné smíšené lineární modely**; velká flexibilita. Používají se Bayesian metody nebo metody maximum likelihood k odhadu komponent genetické (a jiné) variance v rodinách nebo rodokmenech

# Velké odchytky odhadů heritability závisí...

- Data
  - Strategie výběru dat (náhodný výběr vs. rodiny s nemocným jedincem)
  - Metody
- Smith *et al.* (1972) odhadovali heritabilitu diabetu s použitím prvního, druhého a třetího stupně příbuznosti a získali odhady v rozpětí 0,19 až 1,02. „*The range of heritability estimates from the different kinds of relatives raises doubts about both the data collected and about the methods of analysis used“*“

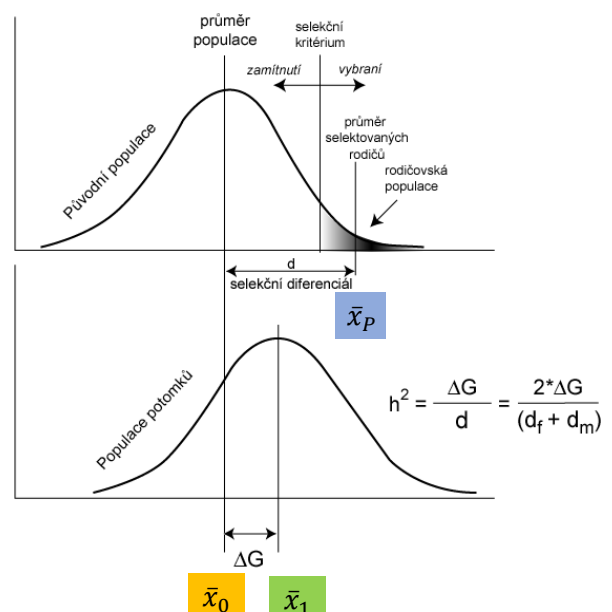
## Využití hodnoty koeficientu heritability

- Predikce průměrné hodnoty vlastnosti u potomků po vybraných rodičích ze základní populace
  - Efekt SELEKCE u zvířat a rostlin
  - Predikce fenotypu u potomků lidí ~ např. náchylnost k nemoci a riziko u příbuzných

Šlechtitelská rovnice  $\Delta G = \Delta\mu = h^2 d$

$$(\bar{x}_1 - \bar{x}_0) = h^2(\bar{x}_P - \bar{x}_0)$$

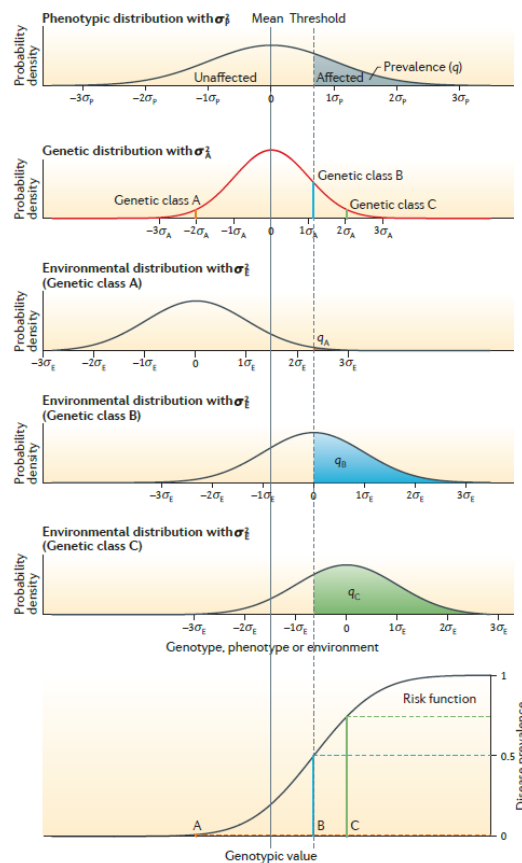
$$\bar{x}_1 = \bar{x}_0 + h^2(\bar{x}_P - \bar{x}_0)$$





- Hodnota koeficientu heritability je specifická pro každou komplexní chorobu, populaci a dané podmínky prostředí
- Nelze ji přenášet mezi populacemi

$$h^2 = \frac{V_G}{V_P} = \frac{V_G}{V_G + V_E}$$



# Lineární modely

## Smíšené lineární modely

- Statistické modely pro analýzu skupinových dat – opakovaná měření, korelace uvnitř skupiny, které potřebují být započítány tvorbu správných závěrů. V kontextu odhadu heritability, oddělují pevné efekty (např. pohlaví nebo věk) od náhodných efektů (jedinci).

## Bayesiánské metody

- B. metody inference kombinuje prior názor o hypotéze s informací poskytnutých z dostupných dat k úpravě těchto prior názorů. Čím silnější jsou prior názory, tím více dat je vyžadováno pro jejich modifikaci. B. metody je vhodné používat na méně početné vzorky, kde metody ML mohou chybovat.

## Maximum likelihood metody

- Metody nebo techniky pro statistické závěry. Využívány pro derivování funkcí výběru (tzv. estimators, odhadci), které když jsou aplikovány na konkrétní vzorky dává odhady populačních parametrů. Maximum likelihood odhady neznámých parametrů jsou nejvíce pravděpodobné odhady parametry, generované z pozorovaných dat

# Molekulárně genetický přístup

## Mapování QTL a genů

- Cílem mnoha generací genetiků je nalézat příčinné geny zodpovědné za variabilitu kvantitativních vlastností
- Před sekvenováním genomů – využívání několika málo genetických markerů (RFLP, mikrosatelity...)
  - Velmi pomalé a náročné detekce QTL oblastí a následně genů
- Osekvenováním genomů -> nalézány milióny polymorfních míst, zejména SNP -> hustší síť pro identifikaci QTL oblastí a genů, rychlejší a přesnější

-> hodnocení kvantitativních znaků reálnou genetickou variabilitou -> přesnější odhady genetických parametrů

# Genetické markery

**Genetický marker (GM):** polymorfní znak, jehož varianty vykazují mendelistickou dědičnost a mohou být v asociaci s variabilitou znaku důležitou pro šlechtění (příp. mapování genů)

Marker přímý (příčinný gen) a nepřímý (ve vazbě s neunámým příčinným genem)

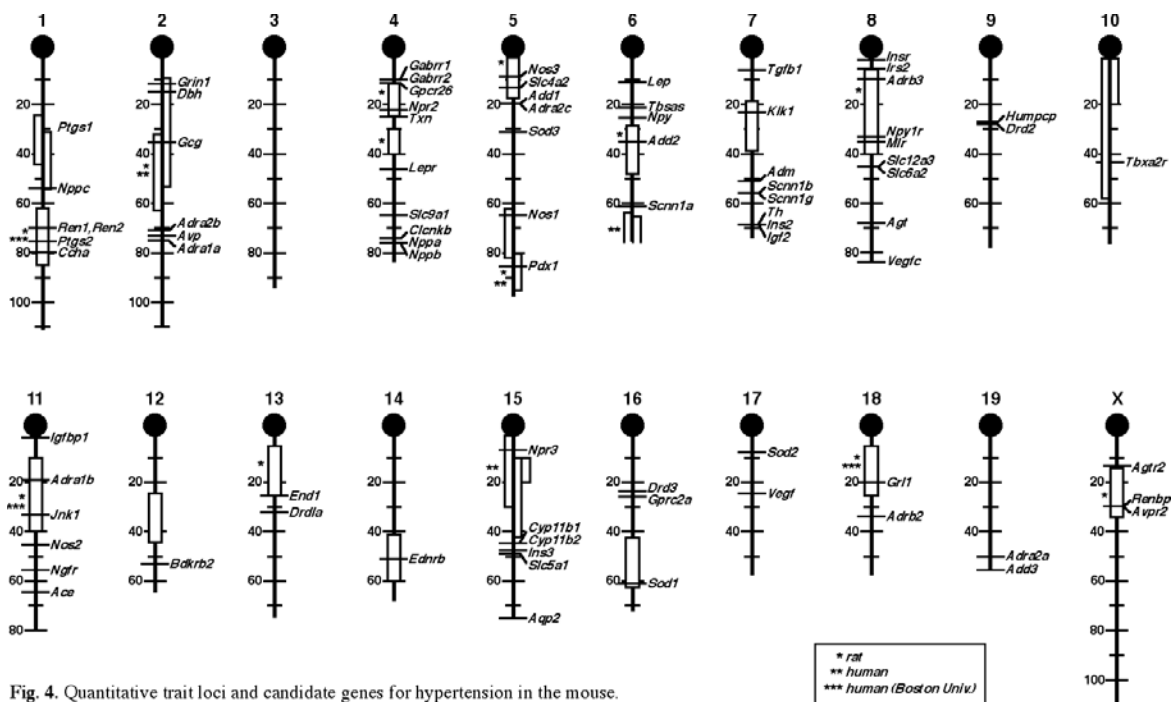
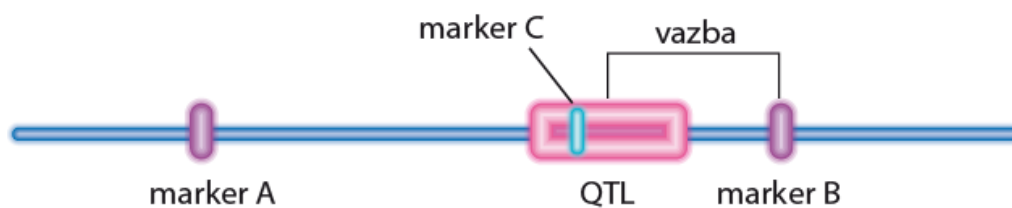
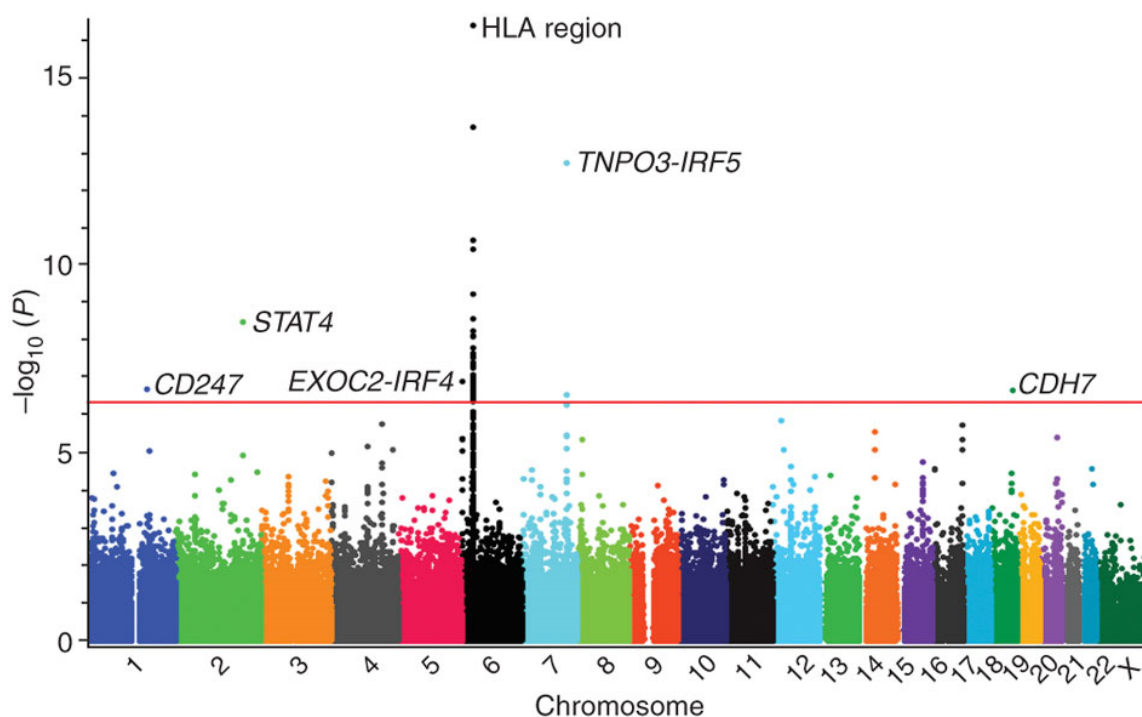


Fig. 4. Quantitative trait loci and candidate genes for hypertension in the mouse.

# GWAS – genome wide association studies

- Metoda identifikace lokusů (jejich variabilita) zodpovědných za komplexní nemoci pomocí genomických markerů
- Založeno na skenování genomů mnoha jedinců a identifikování genetických variant potenciálně zodpovědné (statisticky odhadnuté) za onemocnění pomocí statistických testů
- High-density microarray technologie pro určení tisíců SNP (single nucleotide polymorphisms)
- Statistické hodnocení!!!





## Diabetes 2 typu

- Studie GWAS odhalily mezi lety 2007-2015 přes 260 genů asociovaných s obezitou a diabetem typu 2 (T2D)
- meta-analýzy 18 GWAS celkově 26 676 T2D případů a 132 532 kontrolních subjektů. Zahrnuly 12,1 milionů SNPs použitím multi-etnického projektu 1000 genomů (1000G) referenčního panelu. 29 SNPs vykazují slibné asociace ( $P < 10^{-5}$ ) s T2D
- 13 nových T2D lokusů uvnitř nebo blízko genů *ACSL1*, *HLA-DQA1*, *SLC35D3*, *MNX1*, *ABO*, *PLEKHA1*, *HSD17B12*, *MAP3K11*, *NRXN3*, *CMIP*, *ZZEF1*, *GLP2R*, a *GIP* dosáhlo průkazné GWAS ( $P < 5 \times 10^{-8}$ ) s T2D
- Budoucí GWAS na milionech vzorků, s daty z celogenomového sekvenování – referenční panel stovek tisíc jedinců, pomůže určit DNA spojenou s T2D v etnicky různorodých skupinách.

## Perspektivy výzkumu v QG

- I přes polemiky, které ji obklopují, heritabilita onemocnění zůstává po více než 60 letech nejužitečnějším souhrnným statistickým údajem o genetickém přínosu pro hromadění komplexních nemocí v rodinách.
- Identifikace zdrojů podobnosti v rodinách a praktického využívání těchto znalostí by měla zůstat zaměřením budoucího výzkumu
- Identifikace nových genů
- **Hodnocení jak genetický make-up a prostředí ovlivňují genovou regulaci pomocí epigenetických mechanismů**
- Vývoj statistických a analytických nástrojů a strategií pro využívání neustále se zvyšující komplexnosti genomických dat
- Začleňování těchto znalostí do **vývoje a vyhodnocování nových léčebných postupů**

## A co genomové editace?

- Genová terapie má za cíl opravu nebo náhradu poškozených genů
- Viry jako nosiče cílené sekvence DNA (vychází z techniky transgenózy)
- NOVĚ – přímé zásahy do sekvencí DNA v živých buňkách jedince -> **genomové editace**
- **EK se vyjádřila zatím být proti - problém bezpečnosti a etiky**

Globální pohled na komplexní vlastnosti jedince

