

Preference kodonů

Přednáška 3

ABIN – 2018/19

Kodony

- Kodon = trojice nukleotidů v kódující sekvenci DNA (exonu).
- 64 různých kodonů kóduje 20 aminokyselin → většina aminokyselin je kódována více kodony = degenerovaný (redundantní) genetický kód.
- **Synonymní kodony** kódují stejnou aminokyselinu a většinou se liší třetím nukleotidem.

Standardní genetický kód

		2. nukleotid									
		T		C		A		G			
1. nukleotid	T	TTT	Phe (F)	TCT	Ser (S)	TAT	Tyr (Y)	TGT	Cys (C)	T	3. nukleotid
		TTC		TCC		TAC		TGC		C	
		TTA	Leu (L)	TCA		TAA	STOP	TGA	STOP	A	
		TTG		TCG		TAG		TGG	Trp (W)	G	
	C	CTT	Leu (L)	CCT	Pro (P)	CAT	His (H)	CGT	Arg (R)	T	
		CTC		CCC		CAC		CGC		C	
		CTA		CCA		CAA	Gln (Q)	CGA		A	
		CTG		CCG		CAG		CGG		G	
	A	ATT	Ileu (I)	ACT	Thr (T)	AAT	Asn (N)	AGT	Ser (S)	T	
		ATC		ACC		AAC		AGC		C	
		ATA		ACA		AAA	Lys (K)	AGA	Arg (R)	A	
		ATG	Met (M)	ACG		AAG		AGG		G	
	G	GTT	Val (V)	GCT	Ala (A)	GAT	Asp (D)	GGT	Gly (G)	T	
		GTC		GCC		GAC		GGC		C	
		GTA		GCA		GAA	Glu (E)	GGA		A	
		GTG		GCG		GAG		GGG		G	

Mitochondriální genetický kód

		2. nukleotid									
		T		C		A		G			
1. nukleotid	T	TTT	Phe (F)	TCT	Ser (S)	TAT	Tyr (Y)	TGT	Cys (C)	T	3. nukleotid
		TTC		TCC		TAC		TGC		C	
		TTA	Leu (L)	TCA		TAA	STOP	TGA	Trp (W)	A	
		TTG		TCG		TAG		TGG		G	
	C	CTT	Leu (L)	CCT	Pro (P)	CAT	His (H)	CGT	Arg (R)	T	
		CTC		CCC		CAC		CGC		C	
		CTA		CCA		CAA	Gln (Q)	CGA		A	
		CTG		CCG		CAG		CGG		G	
	A	ATT	Ileu (I)	ACT	Thr (T)	AAT	Asn (N)	AGT	Ser (S)	T	
		ATC		ACC		AAC		AGC		C	
		ATA	Met (M)	ACA		AAA	Lys (K)	AGA	STOP	A	
		ATG		ACG		AAG		AGG		G	
	G	GTT	Val (V)	GCT	Ala (A)	GAT	Asp (D)	GGT	Gly (G)	T	
		GTC		GCC		GAC		GGC		C	
		GTA		GCA		GAA	Glu (E)	GGA		A	
		GTG		GCG		GAG		GGG		G	

Bakteriální genetický kód

		2. nukleotid									
		T		C		A		G			
1. nukleotid	T	TTT	Phe (F)	TCT	Ser (S)	TAT	Tyr (Y)	TGT	Cys (C)	T	3. nukleotid
		TTC		TCC		TAC		TGC		C	
		TTA	Leu (L)	TCA		TAA	STOP	TGA	STOP	A	
		TTG		TCG		TAG		TGG		Trp (W)	
	C	CTT	Leu (L)	CCT	Pro (P)	CAT	His (H)	CGT	Arg (R)	T	
		CTC		CCC		CAC		CGC		C	
		CTA		CCA		CAA	Gln (Q)	CGA		A	
		CTG		CCG		CAG		CGG		G	
	A	ATT	Ileu (I)	ACT	Thr (T)	AAT	Asn (N)	AGT	Ser (S)	T	
		ATC		ACC		AAC		AGC		C	
		ATA		ACA		AAA	Lys (K)	AGA	Arg (R)	A	
		ATG	Met (M)	ACG		AAG		AGG		G	
	G	GTT	Val (V)	GCT	Ala (A)	GAT	Asp (D)	GGT	Gly (G)	T	
		GTC		GCC		GAC		GGC		C	
		GTA		GCA		GAA	Glu (E)	GGA		A	
		GTG		GCG		GAG		GGG		G	

Mutace a poškození DNA

- **Poškození** = změna v obsahu DNA, která se nepřenáší do další generace.
- **Mutace** = změna v obsahu DNA přenesená do další generace.
- Fyzikální a chemické faktory zapříčiňují změny v DNA.
- Při dělení buňky dochází k replikaci DNA, která není bezchybná.

Replikační chyby

- Při replikaci dochází k občasným **mismatch** chybám (špatné párování bází) nebo k vypadnutí/přidání nukleotidů.
- Chybovost DNA polymerázy je 1 chyba na 100 000 nt.
- Diploidní lidská buňka má přes 6,5 miliard nt → 65 000 chyb při jednom dělení buňky.
- Drtivá **většina** replikačních **chyb** je různými procesy **opravena**.
- V bakteriích vzniká 1 mutace na 10^8 až 10^9 nt (délka *E. coli* je $4,6 \cdot 10^6$ nt)
- U eukaryot je chybovost větší, ale buňky se dělí méně často.

Bodové mutace v DNA

- Bodové mutace se týkají **jednotlivých nukleotidů**.
- Rozlišujeme 3 typy:
 - substituce,
 - inzerce,
 - delece.
- **Substituce** je změna nukleotidu na jiný nukleotid.
- **Inzerce** je vložení nového nukleotidu.
- **Delece** je odstranění nukleotidu.

Substituce v DNA

- Substituce **nemění čtecí rámec**.
- Typy substitucí:
 - tranzice,
 - transverze.
- **Tranzice** je záměna purinový \leftrightarrow purinový nukleotid nebo pyrimidinový \leftrightarrow pyrimidinový nukleotid.
- **Transverze** je záměna purinový \leftrightarrow pyrimidinový nukleotid a naopak.
- Tranzice jsou častější než transverze.

Substituce v DNA

- Substituce mohou změnit význam daného kodonu.
- **Synonymní** substituce vytvoří kodon, který kóduje stejnou aminokyselinu jako původní kodon.
- **Nesynonymní** substituce vytvoří kodon, který kóduje jinou aminokyselinu než původní kodon.
- Pouze nesynonymní nukleotidové substituce mění kódované aminokyseliny → **aminokyselinové substituce**.

Indel

- Inzerce a **delece** posouvají čtecí rámec a za nimi dochází ke **změně kódovaných aminokyselin** (často vznikají předčasné stop kodony). Tyto změny jsou radikální a výsledný kódovaný protein je často nefunkční.
- Současný indel tří nukleotidů čtecí rámec neposouvá, pouze vypadne jeden kodon = výsledný protein je o 1 amk kratší nebo delší a nemusí být ovlivněna funkce proteinu.

Preference kodonů

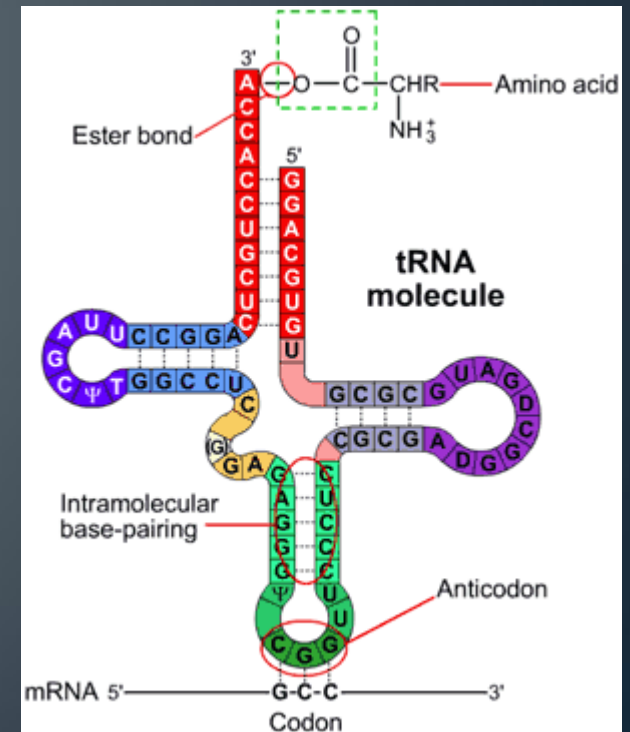
- Pokud by synonymní mutace byly striktně neutrální, pak by mělo být používání kodonů náhodné.
- Avšak **některé kodony jsou více preferované** (codon-bias) = používají se častěji než jiné pro tu stejnou amk.
- Preference se liší v rámci druhů i v rámci jednotlivých genů.
- Některé organismy mají silnou preferenci kodonů.
- Spekuluje se o vlivu preference kodonů na rychlost a přesnost translace.

Vliv používání kodonů

- Znalost preference kodonů pomáhá při **identifikaci genů** => pokud úsek predikovaný jako exon obsahuje mnoho vzácně používaných kodonů, pak pravděpodobně nejde o exon.
- Méně často používané kodony bývají častěji na začátku transkriptu (mRNA) a **zpomalují translaci**.
- Používání kodonů ovlivňuje sekundární strukturu mRNA (prostorové uspořádání mRNA ovlivňuje rychlost translace).

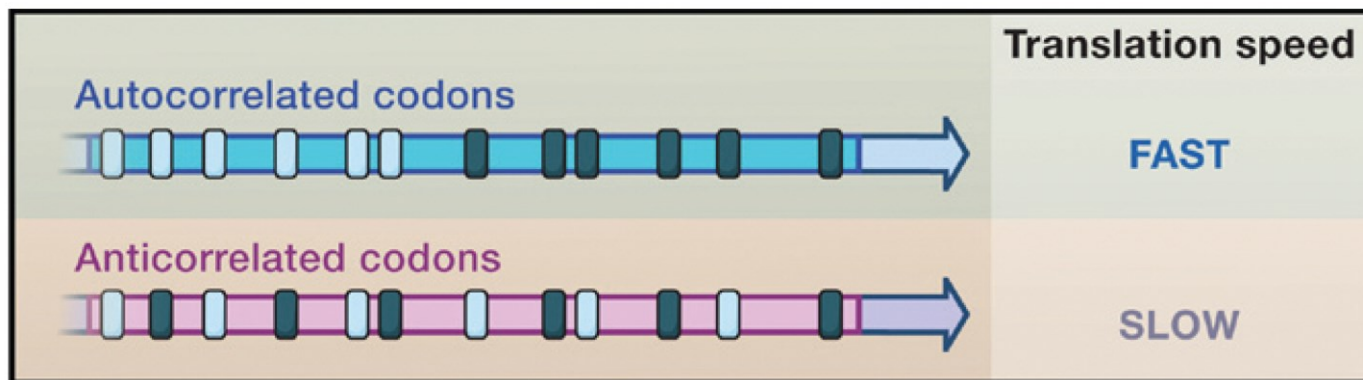
Vliv tRNA

- Při translaci se tRNA váže na kodon prostřednictvím antikodonu.
- Každému kodonu odpovídá jedna tRNA.
- Každá tRNA má jiný počet kopií v DNA, různý level exprese a různou afinitu k cílovému kodonu.

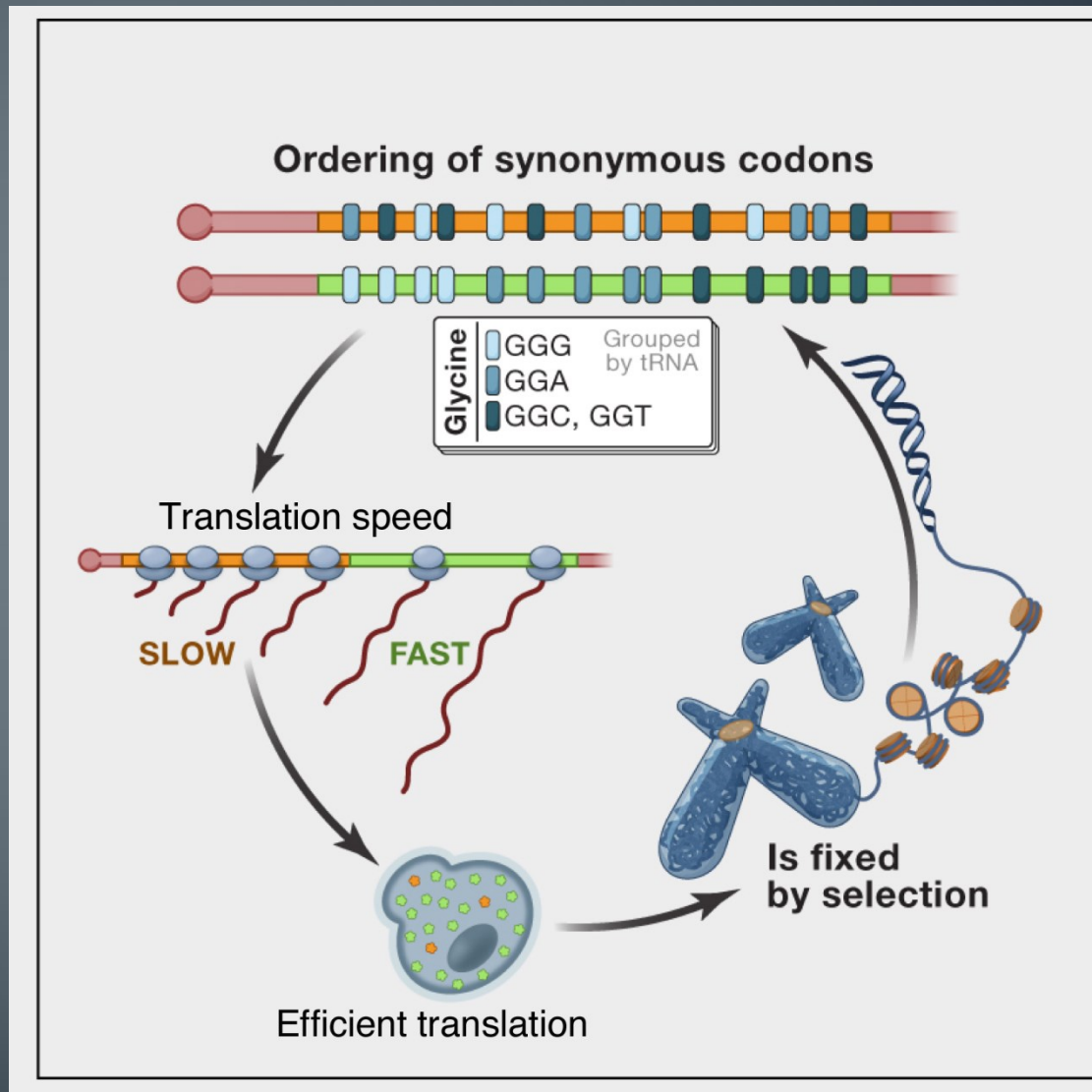


Ovlivnění translace

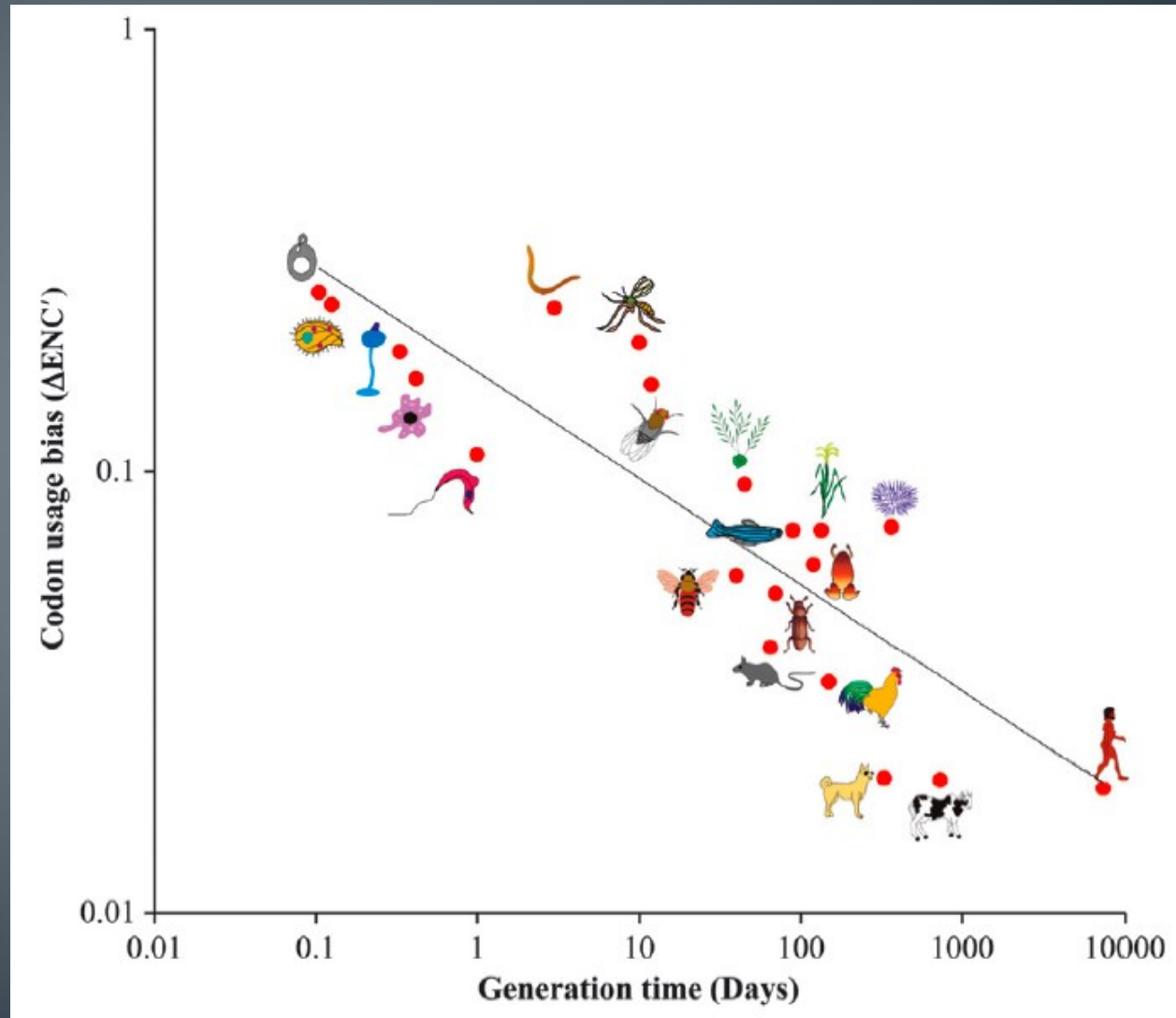
- Efektivita translace je lepší při používání stejného kodonu.
- Translace je pomalá v oblastech rozdílného používání kodonů.



Ovlivnění translace



Preference kodonů u organismů



Použití kodonů

Lidský mitochondriální gen ND1- 954 bp, 318 kodonů.

Leu

CTT

CTC

CTA

CTG

TTA

TTG

AAA - 6	AAC - 13	AAG - 1	AAT - 0
ACA - 14	ACC - 20	ACG - 0	ACT - 1
AGA - 0	AGC - 2	AGG - 0	AGT - 1
ATA - 13	ATC - 13	ATG - 3	ATT - 10
CAA - 6	CAC - 2	CAG - 0	CAT - 0
CCA - 2	CCC - 17	CCG - 1	CCT - 2
CGA - 5	CGC - 2	CGG - 0	CGT - 0
CTA - 27	CTC - 19	CTG - 5	CTT - 6
GAA - 8	GAC - 4	GAG - 3	GAT - 0
GCA - 10	GCC - 14	GCG - 0	GCT - 3
GGA - 1	GGC - 8	GGG - 3	GGT - 0
GTA - 3	GTC - 5	GTG - 0	GTT - 2
TAA - 0	TAC - 9	TAG - 0	TAT - 5
TCA - 8	TCC - 8	TCG - 0	TCT - 2
TGA - 9	TGC - 0	TGG - 0	TGT - 0
TTA - 5	TTC - 12	TTG - 1	TTT - 4

Thr

ACT

ACC

ACA

ACG

Pro

CCT

CCC

CCA

CCG

Použití kodonů

13 lidských mitochondriálních genů, celkem 3796 kodonů.

Leu
Thr
Pro

AAA - 85	AAC - 132	AAG - 10	AAT - 32
ACA - 134	ACC - 155	ACG - 10	ACT - 52
AGA - 1	AGC - 39	AGG - 1	AGT - 14
ATA - 167	ATC - 196	ATG - 40	ATT - 124
CAA - 82	CAC - 79	CAG - 8	CAT - 18
CCA - 52	CCC - 119	CCG - 7	CCT - 41
CGA - 28	CGC - 26	CGG - 2	CGT - 7
CTA - 276	CTC - 167	CTG - 45	CTT - 65
GAA - 64	GAC - 51	GAG - 24	GAT - 15
GCA - 80	GCC - 124	GCG - 8	GCT - 43
GGA - 67	GGC - 87	GGG - 34	GGT - 24
GTA - 70	GTC - 48	GTG - 18	GTT - 31
TAA - 3	TAC - 89	TAG - 2	TAT - 46
TCA - 83	TCC - 99	TCG - 7	TCT - 32
TGA - 93	TGC - 17	TGG - 11	TGT - 5
TTA - 73	TTC - 139	TTG - 18	TTT - 77

Absolutní a relativní četnost

- **Absolutní četnost** AF_{ac} (Absolute Frequency) pro kodon c aminokyseliny a je celkový výskyt daného kodonu:

$$AF_{ac} = n_{ac}$$

- **Relativní četnost** RF_{ac} (Relative Frequency) pro kodon c aminokyseliny a je celkový výskyt daného kodonu ku sumě celkových výskytů synonymních kodonů:

$$RF_{ac} = \frac{n_{ac}}{\sum_{c=1}^{d_a} n_{ac}}$$

kde n_{ac} je počet kodonů c v sekvenci a d_a je stupeň degenerace pro amk a .

$$1 \leq d_a \leq 6$$

Absolutní četnost kodonů

Mitochondriální gen ND1 - 957 bp, 319 kodonů.



Leu
CTT
CTC
CTA
CTG
TTA
TTG

'A'	'GCT 9'	'GCC 11'	'GCA 9'	'GCG 0'		
'R'	'CGT 1'	'CGC 1'	'CGA 6'	'CGG 0'		
'N'	'AAT 2'	'AAC 13'				
'D'	'GAT 0'	'GAC 4'				
'C'	'TGT 2'	'TGC 0'				
'Q'	'CAA 6'	'CAG 0'				
'E'	'GAA 8'	'GAG 2'				
'G'	'GGT 0'	'GGC 5'	'GGA 6'	'GGG 2'		
'H'	'CAT 2'	'CAC 1'				
'I'	'ATT 12'	'ATC 16'				
'L'	'TTA 8'	'TTG 1'	'CTT 3'	'CTC 11'	'CTA 33'	'CTG 2'
'K'	'AAA 8'	'AAG 0'				
'M'	'ATA 13'	'ATG 1'				
'F'	'TTT 8'	'TTC 15'				
'P'	'CCT 5'	'CCC 5'	'CCA 7'	'CCG 0'		
'S'	'TCT 2'	'TCC 10'	'TCA 13'	'TCG 0'	'AGT 0'	'AGC 0'
'T'	'ACT 2'	'ACC 4'	'ACA 12'	'ACG 0'		
'W'	'TGA 9'	'TGG 0'				
'Y'	'TAT 7'	'TAC 5'				
'V'	'GTT 3'	'GTC 7'	'GTA 6'	'GTG 0'		

Thr
ACT
ACC
ACA
ACG

Pro
CCT
CCC
CCA
CCG

Relativní četnost kodonů

Mitochondriální gen ND1 - 957 bp, 319 kodonů.



'A'	'GCT 0.31'	'GCC 0.38'	'GCA 0.31'	'GCG 0'		
'R'	'CGT 0.13'	'CGC 0.13'	'CGA 0.75'	'CGG 0'		
'N'	'AAT 0.13'	'AAC 0.87'				
'D'	'GAT 0'	'GAC 1'				
'C'	'TGT 1'	'TGC 0'				
'Q'	'CAA 1'	'CAG 0'				
'E'	'GAA 0.8'	'GAG 0.2'				
'G'	'GGT 0'	'GGC 0.38'	'GGA 0.46'	'GGG 0.15'		
'H'	'CAT 0.67'	'CAC 0.33'				
'I'	'ATT 0.43'	'ATC 0.57'				
'L'	'TTA 0.14'	'TTG 0.02'	'CTT 0.05'	'CTC 0.19'	'CTA 0.57'	'CTG 0.03'
'K'	'AAA 1'	'AAG 0'				
'M'	'ATA 0.93'	'ATG 0.07'				
'F'	'TTT 0.35'	'TTC 0.65'				
'P'	'CCT 0.29'	'CCC 0.29'	'CCA 0.41'	'CCG 0'		
'S'	'TCT 0.08'	'TCC 0.4'	'TCA 0.52'	'TCG 0'	'AGT 0'	'AGC 0'
'T'	'ACT 0.11'	'ACC 0.22'	'ACA 0.67'	'ACG 0'		
'W'	'TGA 1'	'TGG 0'				
'Y'	'TAT 0.58'	'TAC 0.42'				
'V'	'GTT 0.19'	'GTC 0.44'	'GTA 0.38'	'GTG 0'		

RSCU

- **Relativní použití synonymních kodonů** (Relative Synonymous Codon Usage)
- Kolikrát se určitý kodon vyskytl v sekvenci vůči počtu kolikrát by se vyskytl, kdyby použití kodonů bylo rovnoměrné.

$$RSCU_{ac} = \frac{n_{ac}}{\frac{1}{d_a} \sum_{c=1}^{d_a} n_{ac}} = d_a RF_{ac}$$

- Pokud by všechny synonymní kodony byly používány stejnoměrně, pak je *RSCU* hodnota všech kodonů rovna 1.

RSCU

Mitochondriální gen ND1 - 957 bp, 319 kodonů.



'A'	'GCT 1.24'	'GCC 1.52'	'GCA 1.24'	'GCG 0'		
'R'	'CGT 0.5'	'CGC 0.5'	'CGA 3'	'CGG 0'		
'N'	'AAT 0.27'	'AAC 1.73'				
'D'	'GAT 0'	'GAC 2'				
'C'	'TGT 2'	'TGC 0'				
'Q'	'CAA 2'	'CAG 0'				
'E'	'GAA 1.6'	'GAG 0.4'				
'G'	'GGT 0'	'GGC 1.54'	'GGA 1.85'	'GGG 0.62'		
'H'	'CAT 1.33'	'CAC 0.67'				
'I'	'ATT 0.86'	'ATC 1.14'				
'L'	'TTA 0.83'	'TTG 0.1'	'CTT 0.31'	'CTC 1.14'	'CTA 3.41'	'CTG 0.21'
'K'	'AAA 2'	'AAG 0'				
'M'	'ATA 1.86'	'ATG 0.14'				
'F'	'TTT 0.7'	'TTC 1.3'				
'P'	'CCT 1.18'	'CCC 1.18'	'CCA 1.65'	'CCG 0'		
'S'	'TCT 0.48'	'TCC 2.4'	'TCA 3.12'	'TCG 0'	'AGT 0'	'AGC 0'
'T'	'ACT 0.44'	'ACC 0.89'	'ACA 2.67'	'ACG 0'		
'W'	'TGA 2'	'TGG 0'				
'Y'	'TAT 1.17'	'TAC 0.83'				
'V'	'GTT 0.75'	'GTC 1.75'	'GTA 1.5'	'GTG 0'		

Relativní adaptační index

- Nejprve se pro **hodně exprimované geny** vypočítají hodnoty *RSCU*. Z tabulky *RSCU* hodnot je vidět, který kodon je pro určitou aminokyselinu preferován.
- **Relativní adaptační index** (relative adaptiveness) kodonu *c*:

$$w_{ac} = \frac{RSCU_{ac}}{RSCU_{amax}}$$

kde $RSCU_{amax}$ je hodnota pro nejčastěji používaného kodonu pro aminokyselinu *a*.

Relativní adaptační index

Mitochondriální gen ND1 - 957 bp, 319 kodonů.



'A'	'GCT 0.82'	'GCC 1'	'GCA 0.82'	'GCG 0'		
'R'	'CGT 0.17'	'CGC 0.17'	'CGA 1'	'CGG 0'		
'N'	'AAT 0.15'	'AAC 1'				
'D'	'GAT 0'	'GAC 1'				
'C'	'TGT 1'	'TGC 0'				
'Q'	'CAA 1'	'CAG 0'				
'E'	'GAA 1'	'GAG 0.25'				
'G'	'GGT 0'	'GGC 0.83'	'GGA 1'	'GGG 0.33'		
'H'	'CAT 1'	'CAC 0.49'				
'I'	'ATT 0.75'	'ATC 1'				
'L'	'TTA 0.25'	'TTG 0.04'	'CTT 0.09'	'CTC 0.33'	'CTA 1'	'CTG 0.05'
'K'	'AAA 1'	'AAG 0'				
'M'	'ATA 1'	'ATG 0.08'				
'F'	'TTT 0.54'	'TTC 1'				
'P'	'CCT 0.71'	'CCC 0.71'	'CCA 1'	'CCG 0'		
'S'	'TCT 0.15'	'TCC 0.77'	'TCA 1'	'TCG 0'	'AGT 0'	'AGC 0'
'T'	'ACT 0.16'	'ACC 0.33'	'ACA 1'	'ACG 0'		
'W'	'TGA 1'	'TGG 0'				
'Y'	'TAT 1'	'TAC 0.72'				
'V'	'GTT 0.43'	'GTC 1'	'GTA 0.86'	'GTG 0'		

Kodonový adaptační index

- **Kodonový adaptační index** (Codon Adaptation Index) udává stupeň preference kodonů v genu.
- Hodnota *CAI* pro určitý gen je geometrickým průměrem w_{ac} nenulových hodnot všech kodonů používaných v daném genu.

$$CAI = \left(\prod_{c=1}^L w_c \right)^{\frac{1}{L}} = \exp \left(\frac{1}{L} \sum_{c=1}^L \ln w_c \right)$$

kde L je počet používaných kodonů v sekvenci (počet nenulových w).

Kodonový adaptační index



Mitochondriální gen ND1 - 957 bp, 319 kodonů:

$CAI = 0,5110$

$CAI = 1$ u sekvencí, kde se používají pouze nejčastěji používané kodony.

$CAI = 0$ u sekvencí, kde se používají nejméně časté kodony.

Homozygotnost kodonů

- Pro každou aminokyselinu mající alespoň 2 synonymní kodony se počítá **homozygotnost** F :

$$F_a = \frac{\left(N_a \sum_{c=1}^{d_a} f_{ac}^2 \right) - 1}{N_a - 1}$$

$$f_{ac} = \frac{n_{ac}}{N_a}$$

kde N_a je počet dané amk a v genu a f_{ac} je frekvence výskytu kodonu c pro aminokyselinu a .

- Maximální hodnota homozygotnosti je 1, pokud se používá pouze jeden synonymní kodon.

Efektivní počet kodonů pro amk

- Efektivní počet kodonů Nc_a (Effective Number of Codons) pro aminokyselinu a přibližně udává kolik synonymních kodonů je v sekvenci používáno:

$$Nc_a = \frac{1}{F_a}$$

- Např.: pro amk mající 2 synonymní kodony, které jsou používány stejnou mírou, je efektivní počet kodonů roven 2. Čím nižší číslo, tím větší preference některého kodonu. Minimum je 1.

Fa a Nca



Mitochondriální gen ND1 - 957 bp, 319 kodonů

Homozygotnost kodonů

'A'	'R'	'N'	'D'	'C'	'Q'	'E'	'G'	'H'	'I'	'L'	'K'
'0.31'	'0.54'	'0.75'	'1'	'1'	'1'	'0.64'	'0.33'	'0.33'	'0.49'	'0.37'	'1'
'M'	'F'	'P'	'S'	'T'	'W'	'Y'	'V'				
'0.86'	'0.53'	'0.3'	'0.41'	'0.48'	'1'	'0.47'	'0.33'				

Efektivní počet kodonů

'A'	'R'	'N'	'D'	'C'	'Q'	'E'	'G'	'H'	'I'	'L'	'K'
'3.23'	'1.85'	'1.33'	'1'	'1'	'1'	'1.56'	'3.03'	'3.03'	'2.04'	'2.7'	'1'
'M'	'F'	'P'	'S'	'T'	'W'	'Y'	'V'				
'1.16'	'1.89'	'3.33'	'2.44'	'2.08'	'1'	'2.13'	'3.03'				

Celkový efektivní počet kodonů

- **Efektivní počet kodonů** N_c **pro gen** přibližně udává kolik kodonů se v daném genu používá. Nepočítají se stop kodony. Maximální možná hodnota je 61. Geny se silnou preferencí kodonů mají hodnotu N_c např. kolem 20.

Pro standardní a bakteriální genetický kód.

$$N_c = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}$$

kde např. $\overline{F_4}$ je průměrná homozygotnost aminokyselin mající stupeň degenerace 4.

$$\overline{F_4} = \frac{F_{Ala} + F_{Gly} + F_{Pro} + F_{Val} + F_{Thr}}{5}$$

Celkový efektivní počet kodonů

- Pokud geny nemají žádnou kodonovou preferenci, pak:

$$\overline{F_x} = \frac{1}{X} \quad \text{např.:} \quad \overline{F_4} = \frac{1}{4}$$

- Pro geny bez kodonové preference se hodnota efektivního počtu kodonů N_c blíží celkovému počtu kodonů (bez stop kodonů).
- Pro geny se silnou kodonovou preferencí se hodnota efektivního počtu kodonů N_c blíží 20 (20 je minimum).

Nc genu



Mitochondriální gen ND1 - 957 bp, 319 kodonů

$$N_c = \frac{12}{F_2} + \frac{6}{F_4} + \frac{2}{F_6} = 36,72$$

Gen ND1 má silnou kodonovou preferenci a efektivně využívá 36,72 kodonů.