

Mechanizmy wykrywania i zapobiegania przedwczesnym kodonom STOP

Małgorzata Grabińska

Zakład Genomiki, Wydział Biotechnologii
Uniwersytet Wrocławski
malgorzata.grabinska@smorfland.uni.wroc.pl

30 listopada 2015

Przedwczesne kodony STOP (PTCs-premature STOP codons)

- są wynikiem błędów powstałych w wyniku replikacji, translacji a przede wszystkim transkrypcji;
- najczęstszymi błędami są mutacje punktowe i przesunięcia ramek odczytu;
- utworzone PTC może być dziedziczone i prowadzić do różnych chorób genetycznych, np. dystrofia mięśniowa Duchenna;
- PTC może prowadzić także do utraty funkcjonalności sekwencji lub produkcji cytotoksyn.

Przedwczesne kodony STOP (PTCs-premature STOP codons)

- są wynikiem błędów powstałych w wyniku replikacji, translacji a przede wszystkim transkrypcji;
- najczęstszymi błędami są mutacje punktowe i przesunięcia ramek odczytu;
- utworzone PTC może być dziedziczone i prowadzić do różnych chorób genetycznych, np. dystrofia mięśniowa Duchenna;
- PTC może prowadzić także do utraty funkcjonalności sekwencji lub produkcji cytotoksyn.

Przedwczesne kodony STOP (PTCs-premature STOP codons)

- są wynikiem błędów powstałych w wyniku replikacji, translacji a przede wszystkim transkrypcji;
- najczęstszymi błędami są mutacje punktowe i przesunięcia ramek odczytu;
- utworzone PTC może być dziedziczone i prowadzić do różnych chorób genetycznych, np. dystrofia mięśniowa Duchenna;
- PTC może prowadzić także do utraty funkcjonalności sekwencji lub produkcji cytotoksyn.

Przedwczesne kodony STOP (PTCs-premature STOP codons)

- są wynikiem błędów powstałych w wyniku replikacji, translacji a przede wszystkim transkrypcji;
- najczęstszymi błędami są mutacje punktowe i przesunięcia ramek odczytu;
- utworzone PTC może być dziedziczone i prowadzić do różnych chorób genetycznych, np. dystrofia mięśniowa Duchenna;
- PTC może prowadzić także do utraty funkcjonalności sekwencji lub produkcji cytotoksyn.

Przedwczesne kodony STOP (PTCs-premature STOP codons)

- są wynikiem błędów powstałych w wyniku replikacji, translacji a przede wszystkim transkrypcji;
- najczęstszymi błędami są mutacje punktowe i przesunięcia ramek odczytu;
- utworzone PTC może być dziedziczone i prowadzić do różnych chorób genetycznych, np. dystrofia mięśniowa Duchenna;
- PTC może prowadzić także do utraty funkcjonalności sekwencji lub produkcji cytotoksyn.

Mechanizmy wykrywania i zapobiegania PTC

- 1 NMD (nonsense mediated decay) zjawisko występujące u większości eukariotów;
- 2 używalność kodonów bezpiecznych.

cel

Jak organizmy prokariotyczne zapobiegają PTC?

Mechanizmy wykrywania i zapobiegania PTC

- 1 NMD (nonsense mediated decay) zjawisko występujące u większości eukariotów;
- 2 używalność kodonów bezpiecznych.

cel

Jak organizmy prokariotyczne zapobiegają PTC?

Mechanizmy wykrywania i zapobiegania PTC

- 1 NMD (nonsense mediated decay) zjawisko występujące u większości eukariotów;
- 2 używalność kodonów bezpiecznych.

cel

Jak organizmy prokariotyczne zapobiegają PTC?

Mechanizmy wykrywania i zapobiegania PTC

- 1 NMD (nonsense mediated decay) zjawisko występujące u większości eukariotów;
- 2 używalność kodonów bezpiecznych.

cel

Jak organizmy prokariotyczne zapobiegają PTC?

NMD

zjawisko zachodzące w komórkach eukariotycznych , polegające na rozpoznawaniu i niszczeniu mRNA zawierający PTC. Jest to proces kontroli jakości, zapobiega powstawaniu skróconych białek, które mogą być szkodliwe dla komórki.

- NMD-EJC (połączenie egzon-egzon zależne od introna);
- NMD-PABP (wymaga występowania PABP-polyA niezależnego od introna).

NMD

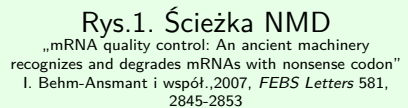
zjawisko zachodzące w komórkach eukariotycznych , polegające na rozpoznawaniu i niszczeniu mRNA zawierający PTC. Jest to proces kontroli jakości, zapobiega powstawaniu skróconych białek, które mogą być szkodliwe dla komórki.

- NMD-EJC (połączenie egzon-egzon zależne od introna);
- NMD-PABP (wymaga występowania PABP-polyA niezależnego od introna).

NMD

zjawisko zachodzące w komórkach eukariotycznych , polegające na rozpoznawaniu i niszczeniu mRNA zawierający PTC. Jest to proces kontroli jakości, zapobiega powstawaniu skróconych białek, które mogą być szkodliwe dla komórki.

- NMD-EJC (połączenie egzon-egzon zależne od introna);
- NMD-PABP (wymaga występowania PABP-polyA niezależnego od introna).



Kodony „kruche”, ryzykowne

Genetic code 1 : standard							
T T T	Phe	T C T	Ser	T A T	Tyr	T G T	Cys
T T C	Phe	T C C	Ser	T A C	Tyr	T G C	Cys
T T A	Leu	T C A	Ser	T A A	Stp	T G A	Stp
T T G	Leu	T C G	Ser	T A G	Stp	T G G	Trp
C T T	Leu	C C T	Pro	C A T	His	C G T	Arg
C T C	Leu	C C C	Pro	C A C	His	C G C	Arg
C T A	Leu	C C A	Pro	C A A	Gln	C G A	Arg
C T G	Leu	C C G	Pro	C A G	Gln	C G G	Arg
A T T	Ile	A C T	Thr	A A T	Asn	A G T	Ser
A T C	Ile	A C C	Thr	A A C	Asn	A G C	Ser
A T A	Ile	A C A	Thr	A A A	Lys	A G A	Arg
A T G	Met	A C G	Thr	A A G	Lys	A G G	Arg
G T T	Val	G C T	Ala	G A T	Asp	G G T	Gly
G T C	Val	G C C	Ala	G A C	Asp	G G C	Gly
G T A	Val	G C A	Ala	G A A	Glu	G G A	Gly
G T G	Val	G C G	Ala	G A G	Glu	G G G	Gly

Tabela.1 Tabela standardowego kodu genetycznego z pakietu *seqinr*

Kodony „kruche”, ryzykowne

Genetic code 1 : standard											
T T T	Phe	T C T	Ser	T A T	Tyr	T G T	Cys				
T T C	Phe	T C C	Ser	T A C	Tyr	T G C	Cys				
T T A	Leu	T C A	Ser	T A A	Stp	T G A	Stp				
T T G	Leu	T C G	Ser	T A G	Stp	T G G	Trp				
C T T	Leu	C C T	Pro	C A T	His	C G T	Arg				
C T C	Leu	C C C	Pro	C A C	His	C G C	Arg				
C T A	Leu	C C A	Pro	C A A	Gln	C G A	Arg				
C T G	Leu	C C G	Pro	C A G	Gln	C G G	Arg				
A T T	Ile	A C T	Thr	A A T	Asn	A G T	Ser				
A T C	Ile	A C C	Thr	A A C	Asn	A G C	Ser				
A T A	Ile	A C A	Thr	A A A	Lys	A G A	Arg				
A T G	Met	A C G	Thr	A A G	Lys	A G G	Arg				
G T T	Val	G C T	Ala	G A T	Asp	G G T	Gly				
G T C	Val	G C C	Ala	G A C	Asp	G G C	Gly				
G T A	Val	G C A	Ala	G A A	Glu	G G A	Gly				
G T G	Val	G C G	Ala	G A G	Glu	G G G	Gly				

Tabela.2 Tabela standardowego kodu genetycznego z pakietu *seqinr* z ryzykownymi kodonami

Kodony „kruche”, ryzykowne

Genetic code 1 : standard							
T T T	Phe	T C T	Ser	T A T	Tyr	T G T	Cys
T T C	Phe	T C C	Ser	T A C	Tyr	T G C	Cys
T T A	Leu	T C A	Ser	T A A	Stp	T G A	Stp
T T G	Leu	T C G	Ser	T A G	Stp	T G G	Trp
C T T	Leu	C C T	Pro	C A T	His	C G T	Arg
C T C	Leu	C C C	Pro	C A C	His	C G C	Arg
C T A	Leu	C C A	Pro	C A A	Gln	C G A	Arg
C T G	Leu	C C G	Pro	C A G	Gln	C G G	Arg
A T T	Ile	A C T	Thr	A A T	Asn	A G T	Ser
A T C	Ile	A C C	Thr	A A C	Asn	A G C	Ser
A T A	Ile	A C A	Thr	A A A	Lys	A G A	Arg
A T G	Met	A C G	Thr	A A G	Lys	A G G	Arg
G T T	Val	G C T	Ala	G A T	Asp	G G T	Gly
G T C	Val	G C C	Ala	G A C	Asp	G G C	Gly
G T A	Val	G C A	Ala	G A A	Glu	G G A	Gly
G T G	Val	G C G	Ala	G A G	Glu	G G G	Gly

Tabela.3 Tabela standardowego kodu genetycznego z pakietu *seqinr* z ryzykownymi kodonami i aminokwasami

Metoda 1

Porównywanie FCU, NFCU, FAU, NFAU regionów nie- i podlegającym NMD z kontrolą. Wszystkie współczynniki obliczano dla każdego genu w grupach kodonów synonimicznych i o takiej samej zawartości GC i tylko dla tych grup gdzie istniały nie- i ryzykowne kodony. Sprawdzano organizmy posiadające obie ścieżki NMD: człowiek, mysz a kontrole wykonano na genomie *Drosophila Melanogaster* gdzie występuje tylko ścieżka NMD-PABP. Porównano także genomy drożdży *Schizosaccharomyces pombe* (obie ścieżki NMD) i *Saccharomyces cerevisiae* (NMD-PABP).

„Preventing dangerous nonsense: selection for robustness to transcriptional error in Human Genes ”

B.P. Cusack i współ., 2011, *PLos Genetics* 7, 10

Grupy dla kodonów:

1. TCA, TCT (seryna);
2. TCG, TCC (seryna);
3. CGA, CGT (arginina);
4. GGA, GGT (glicyna);
5. TTG, CTT, CTA (leucyna).

Grupy dla aminokwasów:

1. tyrozyna, lizyna, asparagina, fenyloalanina;
2. glutamina, kwas glutaminowy, cysteina, seryna, histydyna, treonina, kwas asparaginowy, walina.

Grupy dla kodonów:

1. TCA, TCT (seryna);
2. TCG, TCC (seryna);
3. CGA, CGT (arginina);
4. GGA, GGT (glicyna);
5. TTG, CTT, CTA (leucyna).

Grupy dla aminokwasów:

1. tyrozyna, lizyna, asparagina, fenyloalanina;
2. glutamina, kwas glutaminowy, cysteina, seryna, histydyna, treonina, kwas asparaginowy, walina.

Grupy dla kodonów:

1. TCA, TCT (seryna);
2. TCG, TCC (seryna);
3. CGA, CGT (arginina);
4. GGA, GGT (glicyna);
5. TTG, CTT, CTA (leucyna).

Grupy dla aminokwasów:

1. tyrozyna, lizyna, asparagina, fenyloalanina;
2. glutamina, kwas glutaminowy, cysteina, seryna, histydyna, treonina, kwas asparaginowy, walina.

Grupy dla kodonów:

1. TCA, TCT (seryna);
2. TCG, TCC (seryna);
3. CGA, CGT (arginina);
4. GGA, GGT (glicyna);
5. TTG, CTT, CTA (leucyna).

Grupy dla aminokwasów:

1. tyrozyna, lizyna, asparagina, fenyloalanina;
2. glutamina, kwas glutaminowy, cysteina, seryna, histydyna, treonina, kwas asparaginowy, walina.

Grupy dla kodonów:

1. TCA, TCT (seryna);
2. TCG, TCC (seryna);
3. CGA, CGT (arginina);
4. GGA, GGT (glicyna);
5. TTG, CTT, CTA (leucyna).

Grupy dla aminokwasów:

1. tyrozyna, lizyna, asparagina, fenyloalanina;
2. glutamina, kwas glutaminowy, cysteina, seryna, histydyna, treonina, kwas asparaginowy, walina.

Grupy dla kodonów:

1. TCA, TCT (seryna);
2. TCG, TCC (seryna);
3. CGA, CGT (arginina);
4. GGA, GGT (glicyna);
5. TTG, CTT, CTA (leucyna).

Grupy dla aminokwasów:

1. tyrozyna, lizyna, asparagina, fenyloalanina;
2. glutamina, kwas glutaminowy, cysteina, seryna, histydyna, treonina, kwas asparaginowy, walina.

Grupy dla kodonów:

1. TCA, TCT (seryna);
2. TCG, TCC (seryna);
3. CGA, CGT (arginina);
4. GGA, GGT (glicyna);
5. TTG, CTT, CTA (leucyna).

Grupy dla aminokwasów:

1. tyrozyna, lizyna, asparagina, fenyloalanina;
2. glutamina, kwas glutaminowy, cysteina, seryna, histydyna, treonina, kwas asparaginowy, walina.

Grupy dla kodonów:

1. TCA, TCT (seryna);
2. TCG, TCC (seryna);
3. CGA, CGT (arginina);
4. GGA, GGT (glicyna);
5. TTG, CTT, CTA (leucyna).

Grupy dla aminokwasów:

1. tyrozyna, lizyna, asparagina, fenyloalanina;
2. glutamina, kwas glutaminowy, cysteina, seryna, histydyna, treonina, kwas asparaginowy, walina.

Grupy dla kodonów:

1. TCA, TCT (seryna);
2. TCG, TCC (seryna);
3. CGA, CGT (arginina);
4. GGA, GGT (glicyna);
5. TTG, CTT, CTA (leucyna).

Grupy dla aminokwasów:

1. tyrozyna, lizyna, asparagina, fenyloalanina;
2. glutamina, kwas glutaminowy, cysteina, seryna, histydyna, treonina, kwas asparaginowy, walina.

Grupy dla kodonów:

1. TCA, TCT (seryna);
2. TCG, TCC (seryna);
3. CGA, CGT (arginina);
4. GGA, GGT (glicyna);
5. TTG, CTT, CTA (leucyna).

Grupy dla aminokwasów:

1. tyrozyna, lizyna, asparagina, fenyloalanina;
2. glutamina, kwas glutaminowy, cysteina, seryna, histydyna, treonina, kwas asparaginowy, walina.

		PABP-dependent NMD	EJC-dependent NMD	Genes	NFCU	P-value	Ratio
Human genes							
Multi-exon		+	+	20563	0.47 (0.41–0.54)		1.00
Single-exon:	All	+/-	-	2422	0.43 (0.34–0.51)	$<10^{-15}$	0.92
	Non-histone	+	-	2367	0.43 (0.34–0.51)	$<10^{-15}$	0.92
	Histone	-	-	55	0.32 (0.23–0.45)	$<10^{-9}$	0.68
Mouse genes							
Multi-exon		+	+	20263	0.47 (0.40–0.53)		1.00
Single-exon:	All	+/-	-	3582	0.42 (0.33–0.51)	$<10^{-15}$	0.89
	Non-histone	+	-	3533	0.42 (0.33–0.51)	$<10^{-15}$	0.89
	Histone	-	-	49	0.22 (0.14–0.33)	$<10^{-15}$	0.47
Fly genes							
Multi-exon		+	-	11643	0.54 (0.48–0.60)		1.00
Single-exon:	All	+/-	-	2498	0.53 (0.45–0.62)	0.0002	0.98
	Non-histone	+	-	2466	0.53 (0.45–0.62)	0.002	0.98
	Histone	-	-	32	0.38 (0.38–0.38)	$<10^{-9}$	0.69

	PABP-dependent NMD	EJC-dependent NMD	Genes	FAU	<i>P</i> -value	Ratio
Human genes						
Multi-exon	+	+	20573	0.23 (0.20–0.27)		1.00
Single-exon:	+/-	-	2424	0.19 (0.16–0.24)	$<10^{-15}$	0.83
Mouse genes						
Multi-exon	+	+	20284	0.24 (0.21–0.27)		1.00
Single-exon:	+/-	-	3589	0.19 (0.16–0.24)	$<10^{-15}$	0.79
Fly genes						
Multi-exon	+	-	11643	0.23 (0.20–0.26)		1.00
Single-exon:	+/-	-	2498	0.23 (0.20–0.26)	0.97	1.00

	PABP-dependent NMD	EJC-dependent NMD	Genes	NFAU	P-value	Ratio
Human genes						
Multi-exon	+	+	20573	0.44 (0.39 – 0.48)		1.00
Single-exon	+/-	-	2424	0.38 (0.32 – 0.47)	$<10^{-15}$	0.88
Mouse genes						
Multi-exon	+	+	20284	0.43 (0.39 – 0.48)		1.00
Single-exon	+/-	-	3588	0.36 (0.31 – 0.46)	$<10^{-15}$	0.84
Fly genes						
Multi-exon	+	-	11643	0.41 (0.37 – 0.46)		1.00
Single-exon	+/-	-	2498	0.42 (0.37 – 0.48)	$<10^{-4}$	1.02

Dodatkowe wnioski:

- dla genów o współczynniku $\frac{K_a}{K_s} \sim 1$ organizm obniża zawartość ryzykownych kodonów i aminokwasów;
- $\frac{K_a}{K_s} \sim 0$ różnice mogą się akumulować tylko na poziomie używalności kodonów (selekcja na poziomie aminokwasów);
- FCU i NFCU w pojedynczo egzonowych i wieloegzonowych genach jest niezależna od kodonów optymalnych;
- w ostatnim egzonie w wieloegzonowych genach jest o 8% mniejsze FCU niż w całości (7% u myszy, bez różnic u muchy)

Dodatkowe wnioski:

- dla genów o współczynniku $\frac{K_a}{K_s} \sim 1$ organizm obniża zawartość ryzykownych kodonów i aminokwasów;
- $\frac{K_a}{K_s} \sim 0$ różnice mogą się akumulować tylko na poziomie używalności kodonów (selekcja na poziomie aminokwasów);
- FCU i NFCU w pojedynczo egzonowych i wieloegzonowych genach jest niezależna od kodonów optymalnych;
- w ostatnim egzonie w wieloegzonowych genach jest o 8% mniejsze FCU niż w całości (7% u myszy, bez różnic u muchy)

Dodatkowe wnioski:

- dla genów o współczynniku $\frac{K_a}{K_s} \sim 1$ organizm obniża zawartość ryzykownych kodonów i aminokwasów;
- $\frac{K_a}{K_s} \sim 0$ różnice mogą się akumulować tylko na poziomie używalności kodonów (selekcja na poziomie aminokwasów);
- FCU i NFCU w pojedynczo egzonowych i wieloegzonowych genach jest niezależna od kodonów optymalnych;
- w ostatnim egzonie w wieloegzonowych genach jest o 8% mniejsze FCU niż w całości (7% u myszy, bez różnic u muchy)

Dodatkowe wnioski:

- dla genów o współczynniku $\frac{K_a}{K_s} \sim 1$ organizm obniża zawartość ryzykownych kodonów i aminokwasów;
- $\frac{K_a}{K_s} \sim 0$ różnice mogą się akumulować tylko na poziomie używalności kodonów (selekcja na poziomie aminokwasów);
- FCU i NFCU w pojedynczo egzonowych i wieloegzonowych genach jest niezależna od kodonów optymalnych;
- w ostatnim egzonie w wieloegzonowych genach jest o 8% mniejsze FCU niż w całości (7% u myszy, bez różnic u muchy)

Dodatkowe wnioski:

- dla genów o współczynniku $\frac{K_a}{K_s} \sim 1$ organizm obniża zawartość ryzykownych kodonów i aminokwasów;
- $\frac{K_a}{K_s} \sim 0$ różnice mogą się akumulować tylko na poziomie używalności kodonów (selekcja na poziomie aminokwasów);
- FCU i NFCU w pojedynczo egzonowych i wieloegzonowych genach jest niezależna od kodonów optymalnych;
- w ostatnim egzonie w wieloegzonowych genach jest o 8% mniejsze FCU niż w całości (7% u myszy, bez różnic u muchy)

Metoda 1

Badanie rozkładów kodonów tylko dla aminokwasów z kodonami ryzykownymi i poczwórnie zdegenerowanych (leucyna, seryna, arginina i glicyna).

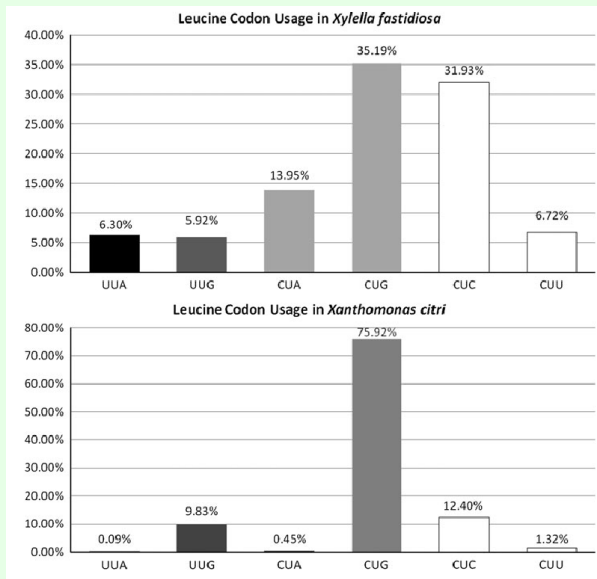
„The usage of codons which are similar to stop codons in the genomes of *Xylella fastidiosa* and ”

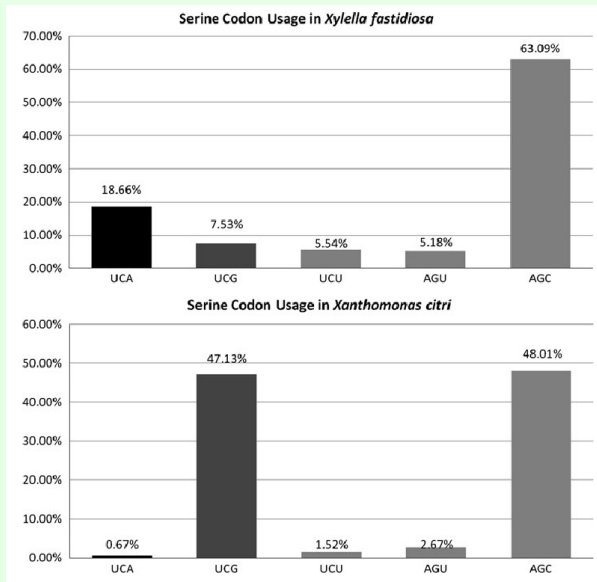
D. Galves-dos-Santos i współ., 2011, *Curr Microbiol* 62, 1090-1095

	LEU	SER	ARG	GLY
Very risky	UUA	UCA		
Risky	UGG	UCG	CGA AGA	GGA
Safe	CUA CUG	UCU AGU ACG	CGU CGC CGG AGG	GGU GGC GGG
Very safe	CUC CUU			

Xanthomonas citri

<i>Xylella fastidiosa</i> Random codon usage	<i>Xanthomonas axonopodis</i> pv. <i>citri</i> Evident selection of optimal codons
<i>Leucine</i>	<i>Leucine</i>
UUA (Very risky): 298	UUA (Very risky): 8
UUG (Risky): 280	UUG (Risky): 871
CUA (Safe): 660	CUA (Safe): 40
CUG (Safe): 1665	CUG (Safe): 6730
CUC (Very safe): 1511	CUC (Very safe): 1099
CUU (Very safe): 318	CUU (Very safe): 117
<i>Serine</i>	<i>Serine</i>
UCA (Very risky): 310	UCA (Very risky): 26
UCG (Risky): 125	UCG (Risky): 1834
UCU (Safe): 92	UCU (Safe): 59
AGU (Safe): 86	AGU (Safe): 104
AGC (Safe): 1048	AGC (Safe): 1868
<i>Arginine</i>	<i>Arginine</i>
CGA (Risky): 154	CGA (Risky): 21
AGA (Risky): 75	AGA (Risky): 4
CGU (Safe): 492	CGU (Safe): 899
CGC (Safe): 1731	CGC (Safe): 4080
CGG (Safe): 168	CGG (Safe): 308
AGG (Safe): 25	AGG (Safe): 22
<i>Glycine</i>	<i>Glycine</i>
GGA (Risky): 398	GGA (Risky): 73
GGU (Safe): 322	GGU (Safe): 983
GGC (Safe): 2260	GGC (Safe): 6784
GGG (Safe): 115	GGG (Safe): 374





- założenie, że *Xylella fastidiosa* (wolno rosnąca) używa losowo kodonów wynika z częstszego używania kodonów bardzo ryzykownych niż *Xanthomonas citri* (szybko rosnąca).



- założenie, że *Xylella fastidiosa* (wolno rosnąca) używa losowo kodonów wynika z częstszego używania kodonów bardzo ryzykownych niż *Xanthomonas citri* (szybko rosnąca).
-

Używalność kodonów ryzykownych dla bazy prokaryotów

Materiały

Do analizy została wzięta baza genomów prokaryotów oprócz *Mycoplasmy* ze względu na różnice w kodzie genetycznym. Analiza używalności została wykonana na zbiorach sekwencji kodujących białko z i bez sekwencji kodujących białka rybosomalne.

Metoda

Dla każdego genomu został wyliczony średni współczynnik FCU, FAU, NFCU, NFAU, NRSCU oraz średnie GC.

Używalność kodonów ryzykownych dla bazy prokaryotów

Materiały

Do analizy została wzięta baza genomów prokaryotów oprócz *Mycoplasmy* ze względu na różnice w kodzie genetycznym. Analiza używalności została wykonana na zbiorach sekwencji kodujących białko z i bez sekwencji kodujących białka rybosomalne.

Metoda

Dla każdego genomu został wyliczony średni współczynnik FCU, FAU, NFCU, NFAU, NRSCU oraz średnie GC.

