

Evolutionary analyses of positively selected sites in the interferon-induced innate immunity factor *Mx1* of bats



seit 1558

Martin Hölzer

RNA Bioinformatics and
High Throughput Analysis

Friedrich-Schiller-University Jena

31st TBI Winterseminar in Bled

February 29, 2016

Run!

Now, here, you see, it takes all the running you can do, to keep in the same place.

– Red Queen to Alice

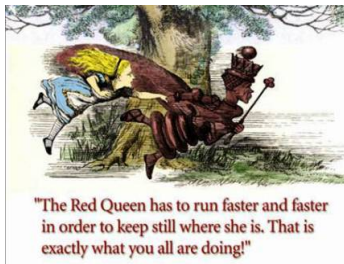


"The Red Queen has to run faster and faster in order to keep still where she is. That is exactly what you all are doing!"

Run!

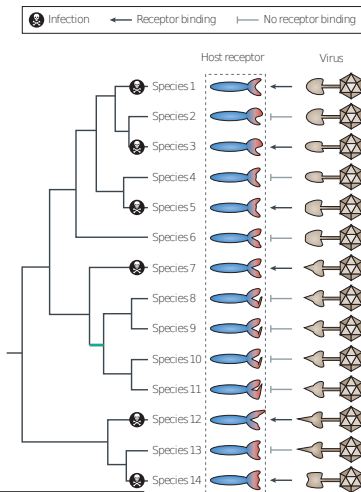
Now, here, you see, it takes all the running you can do, to keep in the same place.

– Red Queen to Alice



- “arms race” between host and virus

Co-Evolution and detecting natural selection



Sironi, Manuela, *et al.* "Evolutionary insights into host-pathogen interactions from mammalian sequence data." *Nature Reviews Genetics* 16.4 (2015): 224-236.

Co-Evolution and detecting natural selection

Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14							
1	...	AAA	GGA	TTG	ATT	AGG	AGT	GCA	AAC	CGT	ACT	CGC	AAG	ATC	AAT	TAC	CTT	AGG	...		
2	...	AAA	GGA	TTG	ATT	AGG	GGT	GGC	AAC	TAT	ACG	CAT	AAA	ATC	AAAT	TAT	CTT	CTC	AGG	...	
3	...	AA	CGA	TTG	ATT	AGA	GGT	GGC	AAC	TAT	ACT	CAT	AAA	ATC	AAAT	TAT	CTT	CTC	AGG	...	
4	...	AAA	GGA	TTG	ATT	AGA	AGT	ACC	AA	A	CAT	ACG	ACT	AAA	ATC	AAT	TAT	CTG	AGG	...	
5	...	AAA	GGA	TTG	ATT	AGA	AGT	ACC	AA	T	CA	C	ACT	AAA	ATC	AAT	TAT	CTT	AGG	...	
6	...	AAA	GGA	TTG	FTT	AGA	AG	C	GCC	AAC	CA	A	ACG	CCT	AAA	AT	T	TAT	CTG	AGG	...
7	...	AAA	AGA	TT	C	ATT	AGA	CGT	GCC	AAC	CAT	ACT	TCT	AAA	ATC	AAT	TAC	CTT	AG	...	
8	...	AAA	GGA	CT	G	ATT	AGA	A	G	T	TCC	AAC	C	T	T	ACT	A	G	A	...	
9	...	AAA	GGA	TTG	ATT	AGA	A	G	T	TCC	AAC	C	T	T	ACT	A	G	A	...		
10	...	AAA	GGA	TTG	ATT	AGA	A	G	T	TCC	AAC	C	T	T	ACT	A	G	A	...		
11	...	AA	F	GGA	TTG	ATT	AGA	A	G	T	TCC	AAC	C	T	T	ACT	A	G	A	...	
12	...	AAA	GG	G	TTG	ATT	AGA	A	G	F	GCC	AAC	CA	G	ACT	CCT	AAA	ATC	AG	...	
13	...	AAA	GGA	TTG	AT	C	AGA	A	G	T	C	C	AAC	CAT	ACT	CCT	AA	G	ATC	AG	...
14	...	AAA	GG	G	TT	A	CTT	AGA	G	G	T	G	C	C	C	C	CT	CT	AG	...	

Synonymous substitution
(no amino acid replacement)

Non-synonymous substitution
(amino acid replacement)

$dN/dS < 1$
(negative selection)

$dN/dS \sim 1$
(neutral evolution)

$dN/dS > 1$
(positive selection)

Lineage-specific selection
(episodic selection)

Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14					
1	...	K	G	L	I	R	S	A	N	R	T	R	K	I	N	Y	L	R	...
2	...	K	G	L	I	R	G	G	N	Y	T	H	K	I	N	Y	L	R	...
3	...	K	G	L	I	R	G	G	N	Y	T	H	K	I	N	Y	L	R	...
4	...	K	G	L	I	R	S	T	K	H	T	T	K	I	N	Y	L	R	...
5	...	K	G	L	I	R	S	A	N	H	T	T	K	I	N	Y	L	R	...
6	...	K	G	L	F	R	S	A	N	Q	T	P	K	I	N	Y	L	R	...
7	...	K	R	F	I	R	R	A	N	H	T	S	K	I	N	Y	L	R	...
8	...	K	G	L	I	R	T	S	N	L	T	T	R	M	N	Y	L	R	...
9	...	K	G	L	I	R	T	S	N	L	T	T	R	M	N	Y	L	R	...
10	...	K	G	L	I	G	T	S	N	L	T	T	R	M	N	Y	L	R	...
11	...	N	G	L	I	R	T	S	N	L	T	T	R	M	S	Y	L	R	...
12	...	K	G	L	I	R	R	A	N	Q	T	P	K	I	S	Y	L	R	...
13	...	K	G	L	I	R	N	P	N	H	T	P	K	I	S	Y	L	R	...
14	...	K	G	L	L	R	G	A	T	N	T	P	K	I	N	Y	L	R	...

Sironi, Manuela, *et al.* "Evolutionary insights into host-pathogen interactions from mammalian sequence data." *Nature Reviews Genetics* 16.4 (2015): 224-236.

Co-Evolution and detecting natural selection

Species

1	...	AAA	GGA	TTG	ATT	AGG	AGT	GCA	AAC	CGT	ACT	CGC	AAG	ATC	AAT	TAC	CTT	AGA	...
2	...	AAA	GGA	TTG	ATT	AGG	GGT	GGC	AAC	TAT	ACG	CAT	AAA	ATC	AAAT	TAT	CTT	AGG	...
3	...	AA	GGA	TTG	ATT	AGA	GGT	GGC	AAC	TAT	ACT	CAT	AAA	ATC	AAAT	TAT	CTT	AGG	...
4	...	AAA	GGA	TTG	ATT	AGA	AGT	ACC	AA	CAT	ACG	ACT	AAA	ATC	AAT	TAT	CTT	AGG	...
5	...	AAA	GGA	TTG	ATT	AGA	AGT	ACC	AA	CA	ACG	ACT	AAA	ATC	AAT	TAT	CTT	AGG	...
6	...	AAA	GGA	TTG	FTT	AGA	AG	GCC	AAC	CAA	ACG	CCT	AAA	AT	AAT	TAT	CTT	AGG	...
7	...	AAA	AGA	TT	ATT	AGA	CGT	GCC	AAC	CAT	ACT	TCT	AAA	ATC	AAT	TAC	CTT	AGA	...
8	...	AAA	GGA	CTG	ATT	AGA	AG	TCC	AAC	CT	ACT	ACT	AGA	ATG	AAT	TAT	CTT	AGG	...
9	...	AAA	GGA	TTG	ATT	AGA	AG	TCC	AAC	CT	ACT	ACT	AGA	ATG	AAT	TAT	CTT	AGG	...
10	...	AAA	GGA	TTG	ATT	GGA	AG	TCC	AA	CT	ACT	ACT	AGA	ATG	AAT	TAT	CTT	AGG	...
11	...	AA	GGA	TTG	ATT	AGA	AG	TCC	AAC	CT	ACT	ACT	GAA	ATG	AGT	TAT	CTT	AGG	...
12	...	AAA	GG	TTG	ATT	AGA	AG	GCC	AAC	CA	ACT	CCT	AAA	ATC	AGT	TAT	CTT	AGG	...
13	...	AAA	GGA	TTG	AT	AGA	AG	GCC	AAC	CAT	ACT	CCT	AA	ATC	AGT	TAT	CTT	AGG	...
14	...	AAA	GG	TT	ATT	AGA	G	GCC	A	A	ACT	CCT	AAA	ATC	AA	TAC	CTT	AGG	...

Synonymous substitution
(no amino acid replacement)

Non-synonymous substitution
(amino acid replacement)

$dN/dS < 1$
(negative selection)

$dN/dS \sim 1$
(neutral evolution)

$dN/dS > 1$
(positive selection)

Lineage-specific selection
(episodic selection)

Species

$$dS = \frac{\text{\# syn substitutions}}{\text{\# syn sites}}$$

$$dN = \frac{\text{\# non-syn substitutions}}{\text{\# non-syn sites}}$$

1	...	K	G	L	I	R	S	A	N	R	T	R	K	I	N	Y	L	R	...	
2	...	K	G	L	I	R	G	G	N	Y	T	H	K	I	N	Y	L	R	...	
3	...	K	G	L	I	R	G	G	N	Y	T	H	K	I	N	Y	L	R	...	
4	...	K	G	L	I	R	S	T	K	H	T	T	K	I	N	Y	L	R	...	
5	...	K	G	L	I	R	S	T	N	H	T	T	K	I	N	Y	L	R	...	
6	...	K	G	L	F	R	S	A	N	Q	T	P	K	I	N	Y	L	R	...	
7	...	K	R	F	I	R	R	A	N	H	T	S	K	I	N	Y	L	R	...	
8	...	K	G	L	I	R	T	S	N	L	T	T	R	M	N	Y	L	R	...	
9	...	K	G	L	I	R	T	S	N	L	T	T	R	M	N	Y	L	R	...	
10	...	K	G	L	I	G	T	S	N	L	T	T	R	M	N	Y	L	R	...	
11	...	N	G	L	I	R	T	S	N	L	T	T	R	M	S	Y	L	R	...	
12	...	K	G	L	I	R	R	A	N	Q	T	P	K	I	S	Y	L	R	...	
13	...	K	G	L	I	R	R	N	P	N	H	T	P	K	I	S	Y	L	R	...
14	...	K	G	L	L	R	G	A	T	N	T	P	K	I	N	Y	L	R	...	

Sironi, Manuela, *et al.* "Evolutionary insights into host-pathogen interactions from mammalian sequence data." *Nature Reviews Genetics* 16.4 (2015): 224-236.

Co-Evolution and detecting natural selection

Species

1	...	AAA	GGA	TTG	ATT	AGG	AGT	GCA	AAC	CGT	ACT	CGC	AAG	ATC	AAT	TAC	CTT	AGA	...		
2	...	AAA	GGA	TTG	ATT	AGG	GGT	GGC	AAC	TAT	ACG	CAT	AAA	ATC	AAAT	TAT	CTT	AGG	...		
3	...	AA	CGA	TTG	ATT	AGA	GGT	GGC	AAC	TAT	ACT	CAT	AAA	ATC	AAAT	TAT	CTT	AGG	...		
4	...	AAA	GGA	TTG	ATT	AGA	AGT	ACC	AA	A	CAT	ACG	ACT	AAA	ATC	AAT	TAT	CTT	AGG	...	
5	...	AAA	GGA	TTG	ATT	AGA	AGT	ACC	AA	T	CA	CG	ACT	AAA	ATC	AAT	TAT	CTT	AGG	...	
6	...	AAA	GGA	TTG	FTT	AGA	AG	C	GCC	AAC	CAA	ACG	CCT	AAA	AT	AAT	TAT	CTT	AGG	...	
7	...	AAA	AGA	TT	C	ATT	AGA	CGT	GCC	AAC	CAT	ACT	TCT	AAA	ATC	AAT	TAC	CTT	AGA	...	
8	...	AAA	GGA	CT	G	ATT	AGA	A	G	T	TCC	AAC	C	T	T	ACT	ACT	AGA	AT	G	...
9	...	AAA	GGA	TTG	ATT	AGA	A	G	T	TCC	AAC	C	T	T	ACT	ACT	AGA	AT	G	...	
10	...	AAA	GGA	TTG	ATT	G	GA	A	G	T	TCC	AA	T	C	T	ACT	ACT	AGA	AT	G	...
11	...	AA	F	GGA	TTG	ATT	AGA	A	G	T	TCC	AAC	C	T	T	ACT	ACT	GAA	AT	G	...
12	...	AAA	GG	G	TTG	ATT	AGA	A	G	A	GCC	AAC	CA	G	ACT	CCT	AAA	ATC	A	G	...
13	...	AAA	GGA	TTG	AT	C	AGA	A	A	T	C	C	AAC	CAT	ACT	CCT	AA	G	ATC	A	...
14	...	AAA	GG	G	TT	A	C	T	T	AGA	G	G	T	GCC	A	C	C	AAT	TAC	C	...

Synonymous substitution
(no amino acid replacement)

Non-synonymous substitution
(amino acid replacement)

$dN/dS < 1$
(negative selection)

$dN/dS \sim 1$
(neutral evolution)

$dN/dS > 1$
(positive selection)

Lineage-specific selection
(episodic selection)

Species

$$dS = \frac{\text{\# syn substitutions}}{\text{\# syn sites}}$$

$$dN = \frac{\text{\# non-syn substitutions}}{\text{\# non-syn sites}}$$

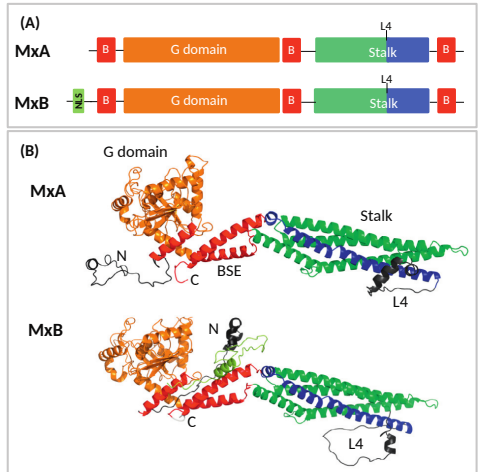
$$\omega = \frac{dN}{dS}$$

1	...	K	G	L	I	R	S	A	N	R	T	R	K	I	N	Y	L	R	...
2	...	K	G	L	I	R	G	G	N	Y	T	H	K	I	N	Y	L	R	...
3	...	K	G	L	I	R	G	G	N	Y	T	H	K	I	N	Y	L	R	...
4	...	K	G	L	I	R	S	T	K	H	T	T	K	I	N	Y	L	R	...
5	...	K	G	L	I	R	S	T	N	H	T	T	K	I	N	Y	L	R	...
6	...	K	G	L	F	R	S	A	N	Q	T	P	K	I	N	Y	L	R	...
7	...	K	R	F	I	R	R	A	N	H	T	S	K	I	N	Y	L	R	...
8	...	K	G	L	I	R	T	S	N	L	T	T	R	M	N	Y	L	R	...
9	...	K	G	L	I	R	T	S	N	L	T	T	R	M	N	Y	L	R	...
10	...	K	G	L	I	G	T	S	N	L	T	T	R	M	N	Y	L	R	...
11	...	N	G	L	I	R	T	S	N	L	T	T	R	M	N	Y	L	R	...
12	...	K	G	L	I	R	R	A	N	Q	T	P	K	I	S	Y	L	R	...
13	...	K	G	L	I	R	N	P	N	H	T	P	K	I	S	Y	L	R	...
14	...	K	G	L	L	R	G	A	T	N	T	P	K	I	N	Y	L	R	...

Sironi, Manuela, *et al.* "Evolutionary insights into host-pathogen interactions from mammalian sequence data." *Nature Reviews Genetics* 16.4 (2015): 224-236.

Mx GTPases

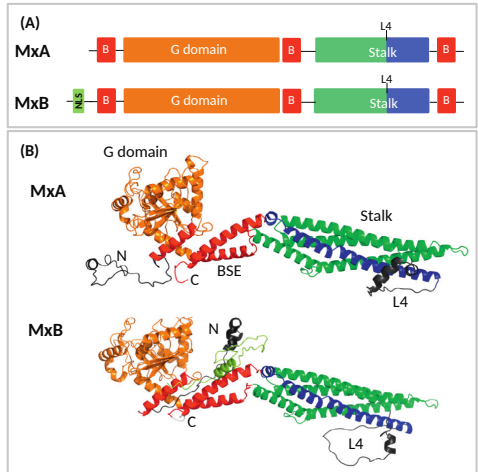
- Myxovirus resistance protein 1 (A) and 2 (B), two *Mx* copies in most mammals



Haller *et al.* "Mx GTPases: dynamin-like antiviral machines of innate immunity." *Trends in microbiology* (2015): 154-163.

Mx GTPases

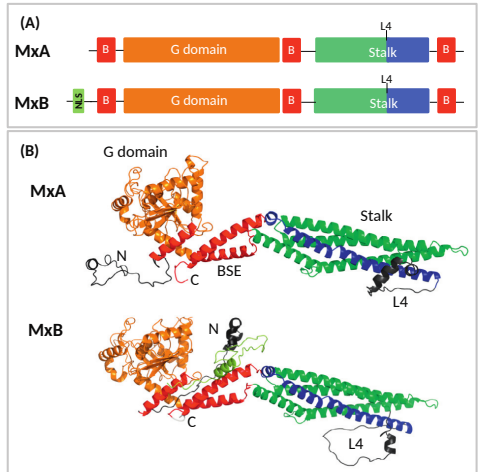
- Myxovirus resistance protein 1 (A) and 2 (B), two *Mx* copies in most mammals
- dynamin-like GTPases



Haller *et al.* "Mx GTPases: dynamin-like antiviral machines of innate immunity." *Trends in microbiology* (2015): 154-163.

Mx GTPases

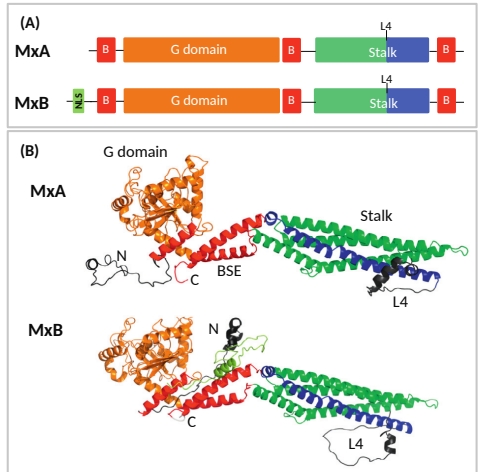
- Myxovirus resistance protein 1 (A) and 2 (B), two *Mx* copies in most mammals
- dynamin-like GTPases
- induced by IFN I/III



Haller *et al.* "Mx GTPases: dynamin-like antiviral machines of innate immunity." *Trends in microbiology* (2015): 154-163.

Mx GTPases

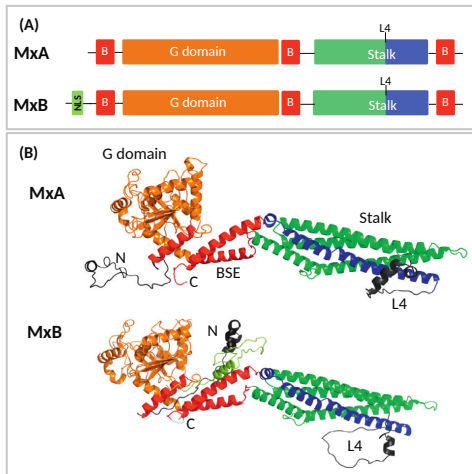
- Myxovirus resistance protein 1 (A) and 2 (B), two *Mx* copies in most mammals
- dynamin-like GTPases
- induced by IFN I/III
- key antiviral effector proteins in mammals



Haller *et al.* "Mx GTPases: dynamin-like antiviral machines of innate immunity." *Trends in microbiology* (2015): 154-163.

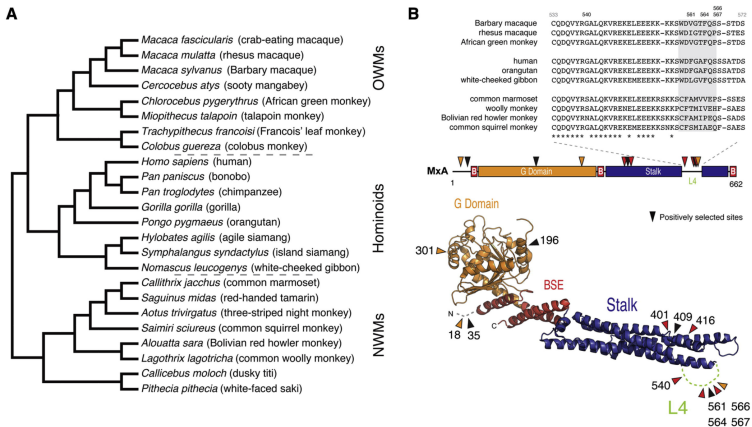
Mx GTPases

- Myxovirus resistance protein 1 (A) and 2 (B), two *Mx* copies in most mammals
- dynamin-like GTPases
- induced by IFN I/III
- key antiviral effector proteins in mammals
- blocking early steps of viral replication cycles (DNA, RNA)



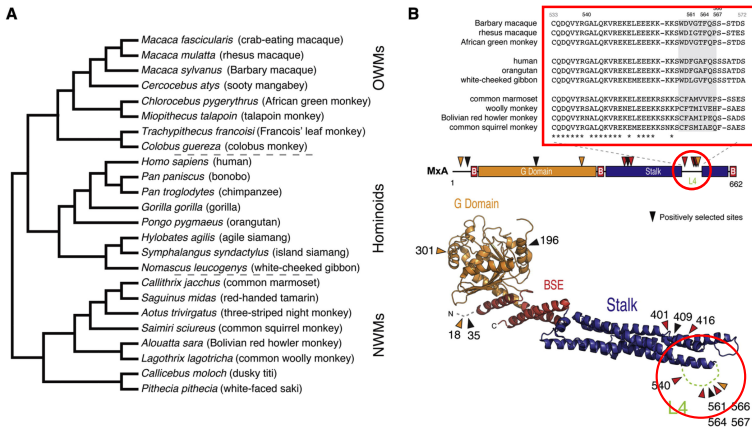
Haller *et al.* "Mx GTPases: dynamin-like antiviral machines of innate immunity." *Trends in microbiology* (2015): 154-163.

Mx GTPases



Mitchell, Patrick S., *et al.* "Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA." *Cell host & microbe* 12.4 (2012): 598-604.

Mx GTPases

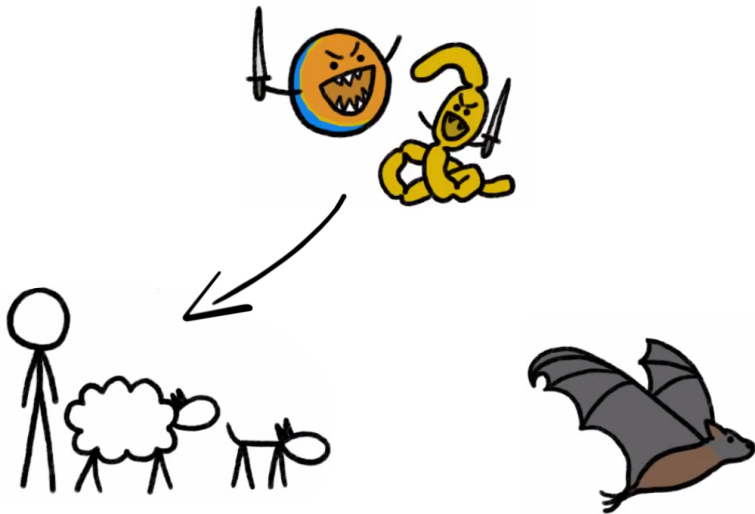


Mitchell, Patrick S., *et al.* "Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA." *Cell host & microbe* 12.4 (2012): 598-604.

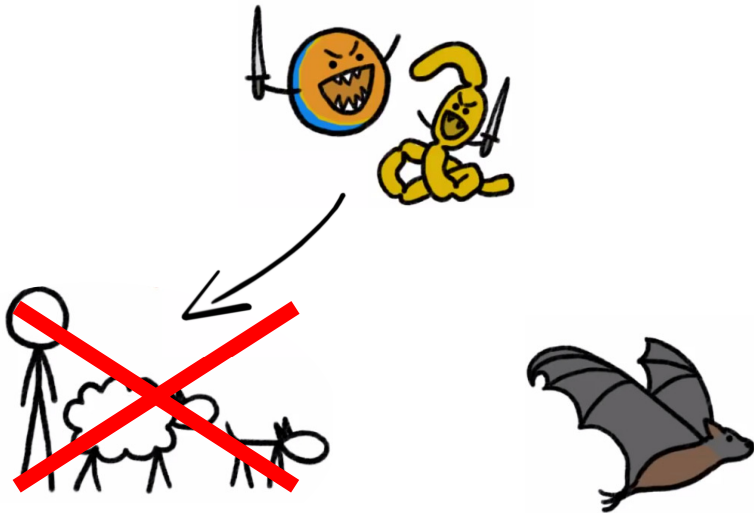
Mx1 in bats?



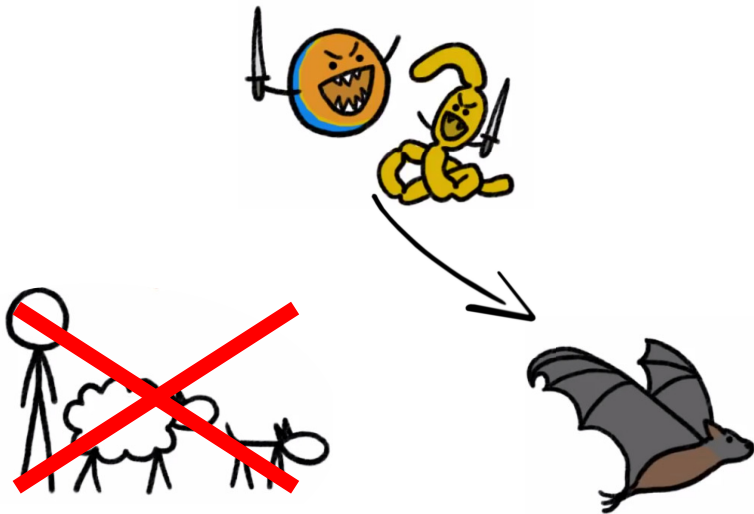
Mx1 in bats?



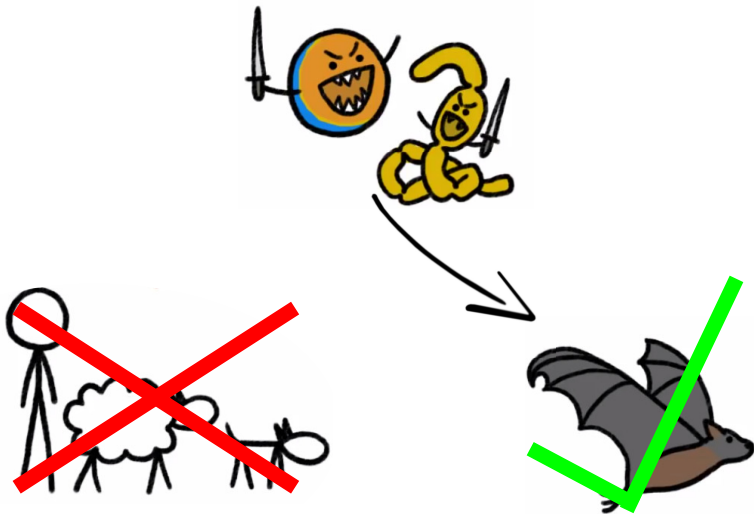
Mx1 in bats?



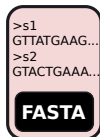
Mx1 in bats?



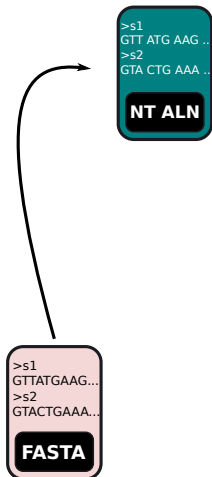
Mx1 in bats?



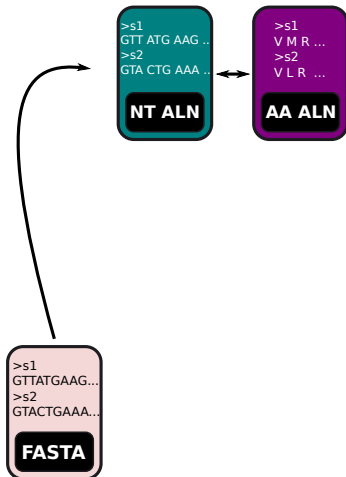
Workflow to detect positively selected sites



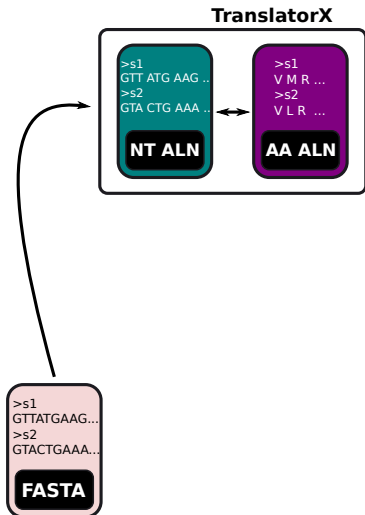
Workflow to detect positively selected sites



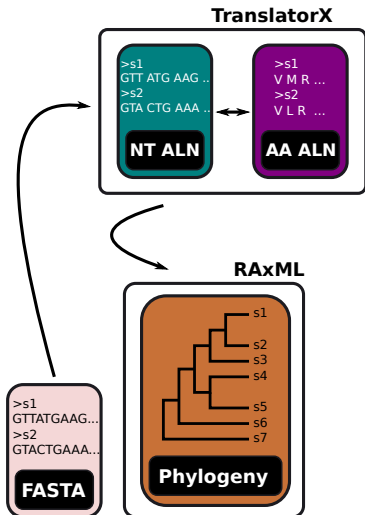
Workflow to detect positively selected sites



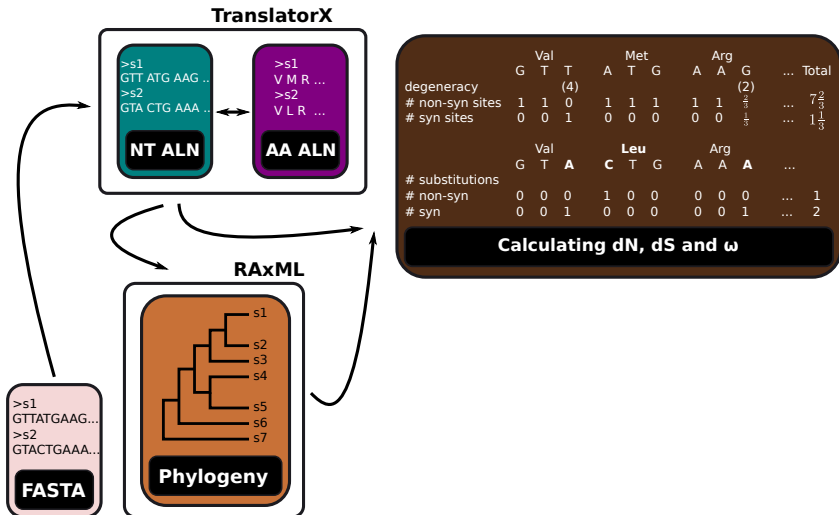
Workflow to detect positively selected sites



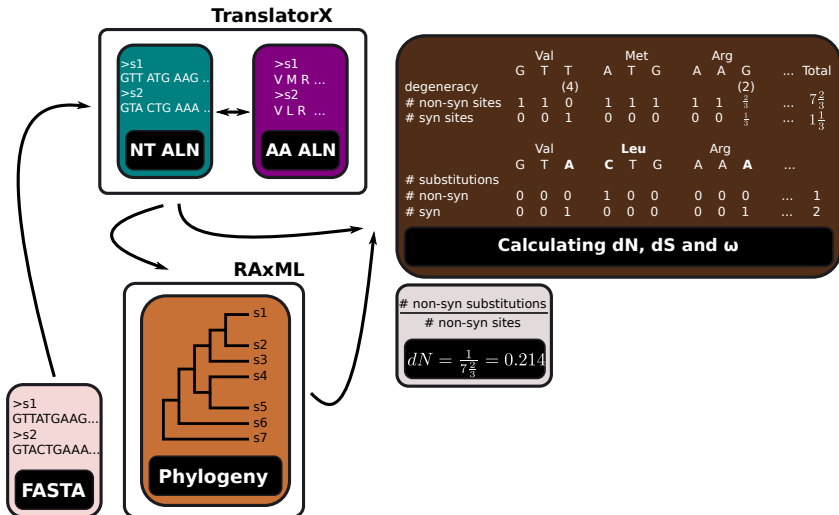
Workflow to detect positively selected sites



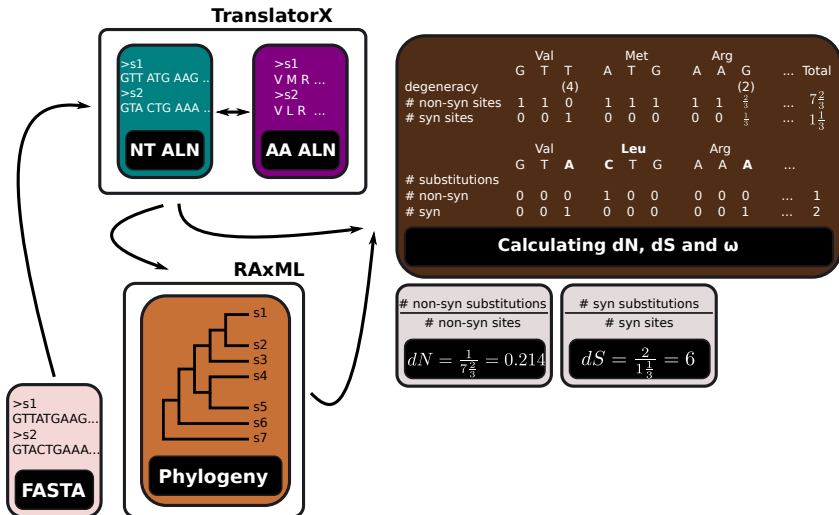
Workflow to detect positively selected sites



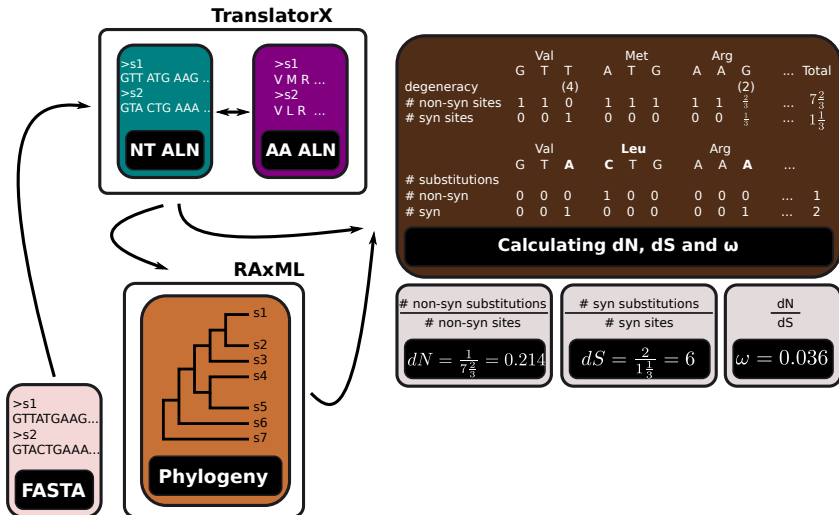
Workflow to detect positively selected sites



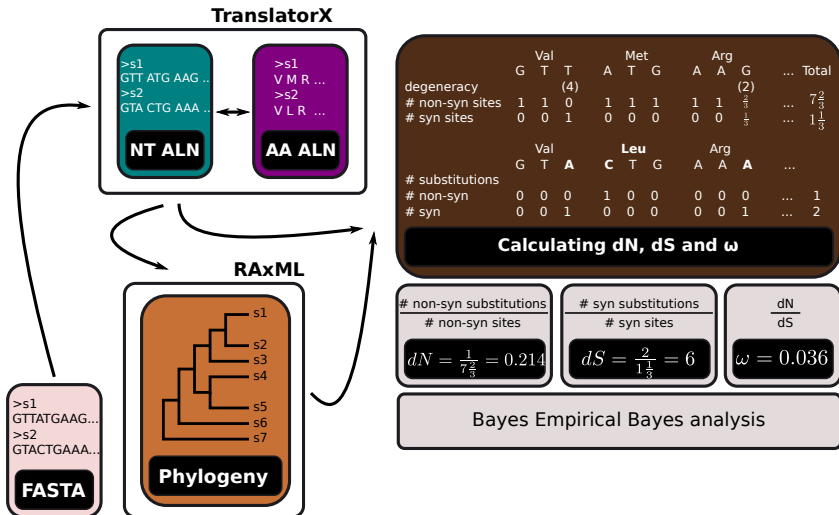
Workflow to detect positively selected sites



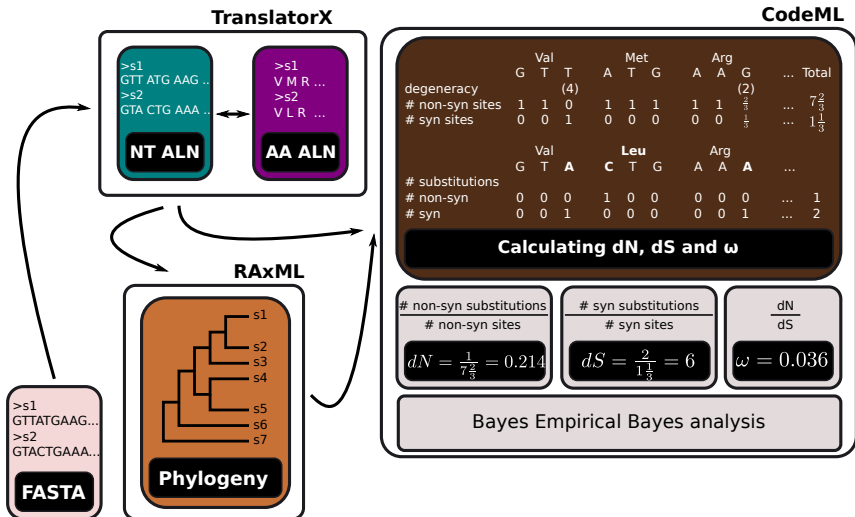
Workflow to detect positively selected sites



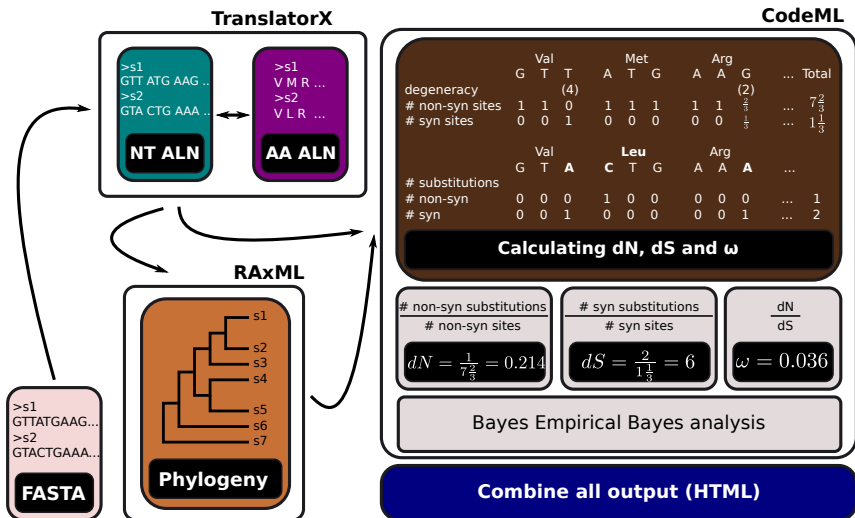
Workflow to detect positively selected sites



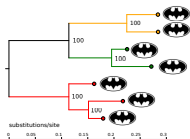
Workflow to detect positively selected sites



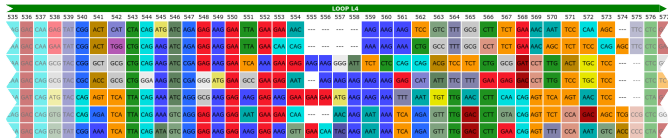
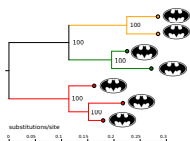
Workflow to detect positively selected sites



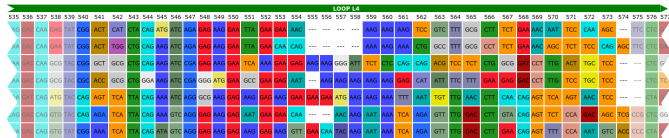
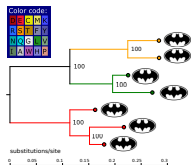
Positive selection in newly cloned bat *Mx1* sequences



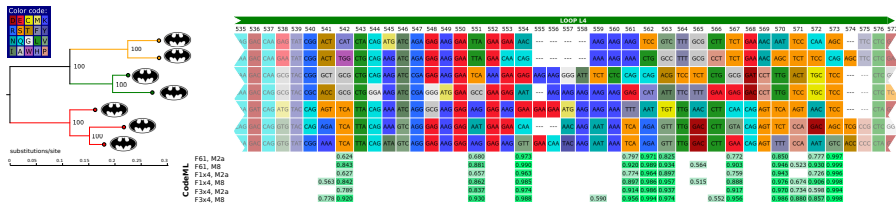
Positive selection in newly cloned bat *Mx1* sequences



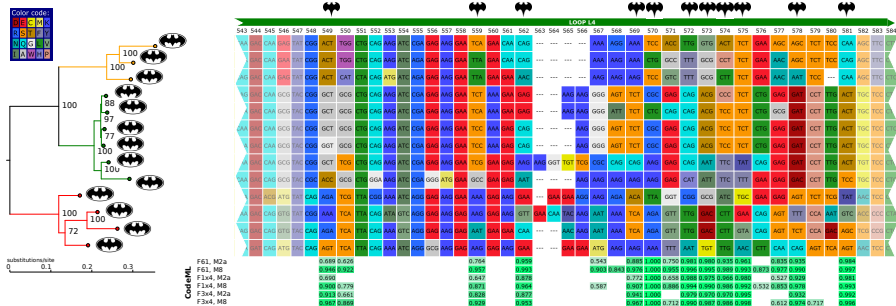
Positive selection in newly cloned bat *Mx1* sequences



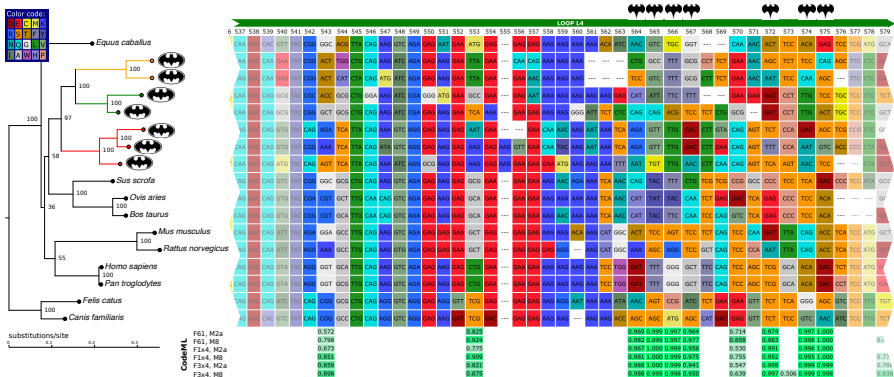
Positive selection in newly cloned bat *Mx1* sequences



Positive selection: more bat species



Positive selection: bats and other mammals



Thank you for your attention!

Thanks to all my group members!

Manja Marz
Konrad Sachse
Holger Bierhoff
Franziska Hufsky
Markus Fricke
Konstantin Riege
Emanuel Barth
Nelly F. M. Berrospi
Bertram Vogel
Michael Gaspar
Akash Srivastava



... and Georg and Jonas (*Mx*-guys from Universitätsklinikum Freiburg)

Calculating selective pressure, K_a/K_s

- K_a/K_s ratio is an indicator of selective pressure acting on protein-coding genes

Calculating selective pressure, K_a/K_s

- K_a/K_s ratio is an indicator of selective pressure acting on protein-coding genes
 - K_a : number of non-synonymous substitutions per non-synonymous site

Calculating selective pressure, K_a/K_s

- K_a/K_s ratio is an indicator of selective pressure acting on protein-coding genes
 - K_a : number of non-synonymous substitutions per non-synonymous site
 - K_s : number of synonymous substitutions per synonymous site

Calculating selective pressure, K_a/K_s

- K_a/K_s ratio is an indicator of selective pressure acting on protein-coding genes
 - K_a : number of non-synonymous substitutions per non-synonymous site
 - K_s : number of synonymous substitutions per synonymous site
- homologous genes with a K_a/K_s ratio above 1 are evolving under positive selection
 - at least some of the mutations concerned must be **advantageous**

Calculating selective pressure, K_a/K_s

- K_a/K_s ratio is an indicator of selective pressure acting on protein-coding genes
 - K_a : number of non-synonymous substitutions per non-synonymous site
 - K_s : number of synonymous substitutions per synonymous site
- homologous genes with a K_a/K_s ratio above 1 are evolving under positive selection
 - at least some of the mutations concerned must be **advantageous**
- if all the mutations are neutral or disadvantageous, the ratio will be in the range 0 to 1

Calculating selective pressure, K_a/K_s

- K_a/K_s ratio is an indicator of selective pressure acting on protein-coding genes
 - K_a : number of non-synonymous substitutions per non-synonymous site
 - K_s : number of synonymous substitutions per synonymous site
- homologous genes with a K_a/K_s ratio above 1 are evolving under positive selection
 - at least some of the mutations concerned must be **advantageous**
- if all the mutations are neutral or disadvantageous, the ratio will be in the range 0 to 1
- however, if some of the mutations are advantageous and some disadvantageous, the ratio could be less than 1

Calculating selective pressure, K_a/K_s

- K_a/K_s ratio is an indicator of selective pressure acting on protein-coding genes
 - K_a : number of non-synonymous substitutions per non-synonymous site
 - K_s : number of synonymous substitutions per synonymous site
- homologous genes with a K_a/K_s ratio above 1 are evolving under positive selection
 - at least some of the mutations concerned must be **advantageous**
- if all the mutations are neutral or disadvantageous, the ratio will be in the range 0 to 1
- however, if some of the mutations are advantageous and some disadvantageous, the ratio could be less than 1
- $K_a/K_s = dN/dS = \omega$

Calculating selective pressure, K_a/K_s

- K_a/K_s ratio is an indicator of selective pressure acting on protein-coding genes
 - K_a : number of non-synonymous substitutions per non-synonymous site
 - K_s : number of synonymous substitutions per synonymous site
- homologous genes with a K_a/K_s ratio above 1 are evolving under positive selection
 - at least some of the mutations concerned must be **advantageous**
- if all the mutations are neutral or disadvantageous, the ratio will be in the range 0 to 1
- however, if some of the mutations are advantageous and some disadvantageous, the ratio could be less than 1
- $K_a/K_s = dN/dS = \omega$ $\kappa = Ts/Tv$

Example: calculating K_a/K_s

	Val		Met		Arg		Thr					
G	T	T	A	T	G	A	A	G	A	C	C	Total

Example: calculating K_a/K_s

		Val		Met		Arg		Thr			
	G	T	T	A	T	G	A	A	C	C	Total
degeneracy			(4)				(2)			(4)	

Example: calculating K_a/K_s

	Val			Met			Arg			Thr			Total
	G	T	T	A	T	G	A	A	G	A	C	C	
degeneracy			(4)						(2)			(4)	
# non-syn sites	1	1	0	1	1	1	1	1	$\frac{2}{3}$	1	1	0	$9\frac{2}{3}$

Example: calculating K_a/K_s

	Val			Met			Arg			Thr			Total
	G	T	T	A	T	G	A	A	G	A	C	C	
degeneracy			(4)						(2)			(4)	
# non-syn sites	1	1	0	1	1	1	1	1	$\frac{2}{3}$	1	1	0	$9\frac{2}{3}$
# syn sites	0	0	1	0	0	0	0	0	$\frac{1}{3}$	0	0	1	$2\frac{1}{3}$

Example: calculating K_a/K_s

	Val			Met			Arg			Thr			Total
	G	T	T	A	T	G	A	A	G	A	C	C	
degeneracy			(4)						(2)			(4)	
# non-syn sites	1	1	0	1	1	1	1	1	$\frac{2}{3}$	1	1	0	$9\frac{2}{3}$
# syn sites	0	0	1	0	0	0	0	0	$\frac{1}{3}$	0	0	1	$2\frac{1}{3}$

- nucleotide sites give $9\frac{2}{3}$ non-synonymous sites and $2\frac{1}{3}$ synonymous sites in this peptide

Example: calculating K_a/K_s

	Val			Met			Arg			Thr			Total
	G	T	T	A	T	G	A	A	G	A	C	C	
degeneracy			(4)						(2)			(4)	
# non-syn sites	1	1	0	1	1	1	1	1	$2\frac{2}{3}$	1	1	0	$9\frac{2}{3}$
# syn sites	0	0	1	0	0	0	0	0	$1\frac{1}{3}$	0	0	1	$2\frac{1}{3}$

- nucleotide sites give $9\frac{2}{3}$ non-synonymous sites and $2\frac{1}{3}$ synonymous sites in this peptide

Val		Leu			Arg			Thr			Total
G	T	A	C	T	G	A	A	A	C	C	
		A	C				A				

Example: calculating K_a/K_s

	Val			Met			Arg			Thr			Total
	G	T	T	A	T	G	A	A	G	A	C	C	
degeneracy			(4)						(2)			(4)	
# non-syn sites	1	1	0	1	1	1	1	1	$\frac{2}{3}$	1	1	0	$9\frac{2}{3}$
# syn sites	0	0	1	0	0	0	0	0	$\frac{1}{3}$	0	0	1	$2\frac{1}{3}$

- nucleotide sites give $9\frac{2}{3}$ non-synonymous sites and $2\frac{1}{3}$ synonymous sites in this peptide

	Val			Leu			Arg			Thr			Total
	G	T	A	C	T	G	A	A	A	A	C	C	
# substitutions													
# non-syn	0	0	0	1	0	0	0	0	0	0	0	0	1

Example: calculating K_a/K_s

	Val			Met			Arg			Thr			Total
	G	T	T	A	T	G	A	A	G	A	C	C	
degeneracy			(4)						(2)			(4)	
# non-syn sites	1	1	0	1	1	1	1	1	$\frac{2}{3}$	1	1	0	$9\frac{2}{3}$
# syn sites	0	0	1	0	0	0	0	0	$\frac{1}{3}$	0	0	1	$2\frac{1}{3}$

- nucleotide sites give $9\frac{2}{3}$ non-synonymous sites and $2\frac{1}{3}$ synonymous sites in this peptide

	Val			Leu			Arg			Thr			Total
	G	T	A	C	T	G	A	A	A	A	C	C	
# substitutions													
# non-syn	0	0	0	1	0	0	0	0	0	0	0	0	1
# syn	0	0	1	0	0	0	0	0	1	0	0	0	2

Example: calculating K_a/K_s

Now,

$$K_a = \frac{\#nonsyn_substitutions}{\#nonsyn_sites} = \frac{1}{9\frac{2}{3}} = 0.103$$

and

$$K_s = \frac{\#syn_substitutions}{\#syn_sites} = \frac{2}{2\frac{1}{3}} = 0.857$$

Thus,

$$\frac{K_a}{K_s} = \frac{dN}{dS} = \omega = \frac{0.103}{0.857} = 0.12$$

CodeML: Output

- *CodonFreq=F3x4* (estimation of codon frequency distribution)
- *NSsites=M2a* (positive selection)

Model	NSsites	p	Parameters
M0 (one ratio)	0	1	ω
M1a (neutral)	1	2	p_0 ($p_1 = 1 - p_0$), $\omega_0 < 1, \omega_1 = 1$
M2a (selection)	2	4	p_0, p_1 ($p_2 = 1 - p_0 - p_1$), $\omega_0 < 1, \omega_1 = 1, \omega_2 > 1$
M3 (discrete)	3	5	p_0, p_1 ($p_2 = 1 - p_0 - p_1$) $\omega_0, \omega_1, \omega_2$
M7 (beta)	7	2	p, q
M8 (beta& ω)	8	4	p_0 ($p_1 = 1 - p_0$), $p, q, \omega_i > 1$

NOTE.—The site models are implemented using the control variable NSsites in CODEML, and p is the number of free parameters in the ω distribution.

⇒ [codeml.variable.mlc]