www.gephebase.org

Virginie Courtier-Orgogozo
Arnaud Martin (Washington DC)
Laurent Arnoult (former postdoc)
Stéphane Prigent (former postdoc)

www.gephebase.org

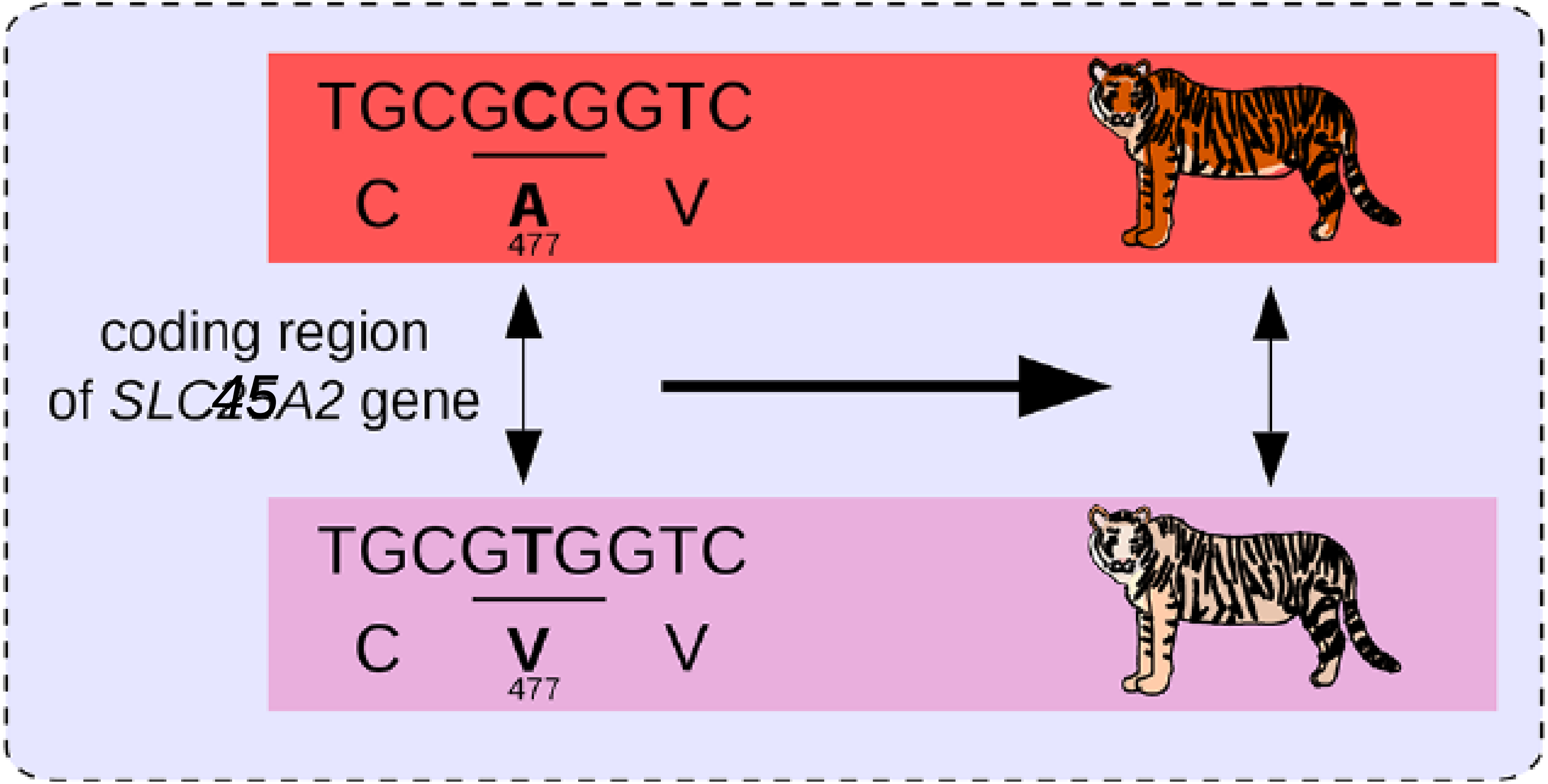
Includes Natural, Domesticated and Experimental Variation
but NO *LAB MUTANTS* and NO *CLINICAL TRAITS*

>1600 genes and mutations associated with natural phenotypic changes in animals and plants

Virginie Courtier-Orgogozo
Arnaud Martin (Washington DC)
Laurent Arnoult (former postdoc)
Stéphane Prigent (former postdoc)

Ge-phe

a Genetic VARIATION *causing* a Phenotypic VARIATION



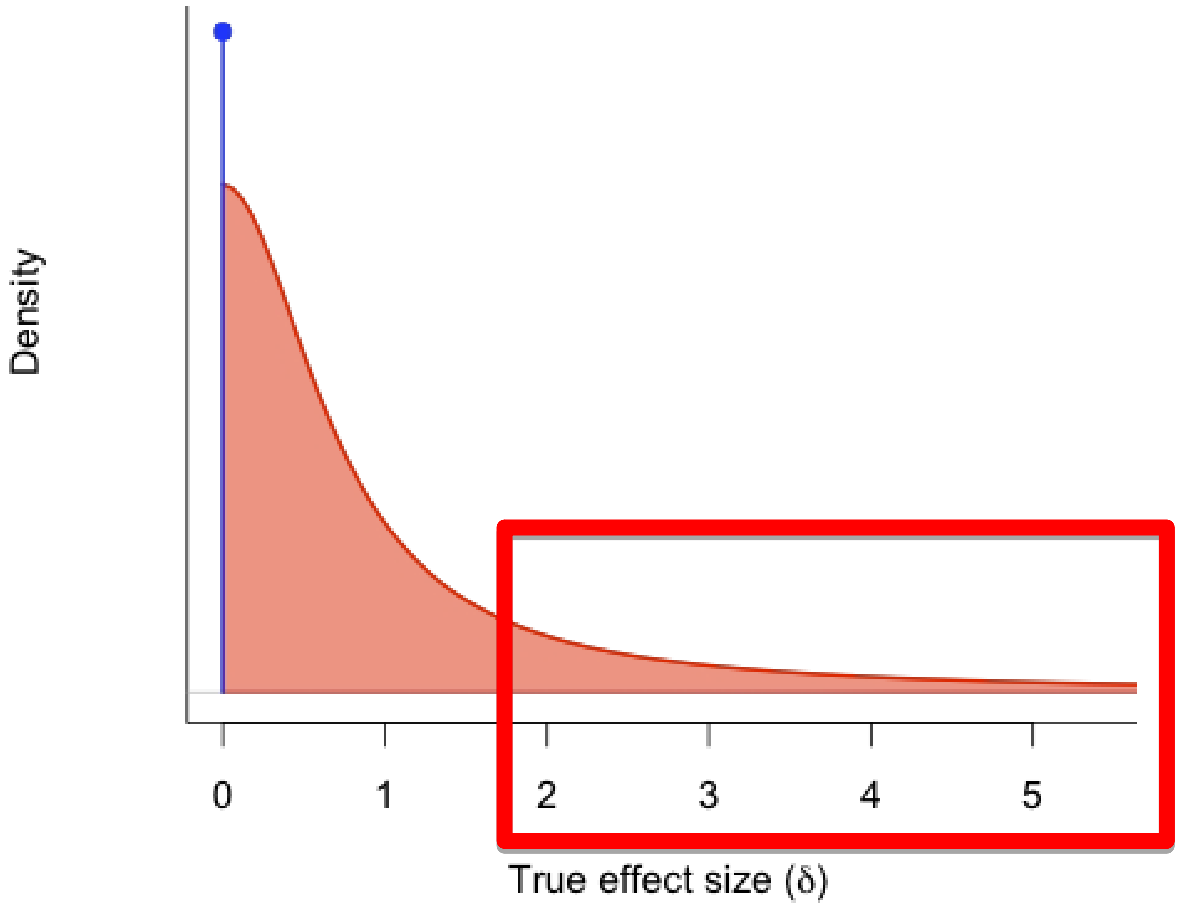
Orgogozo, Morizot & Martin *Frontiers Genet.* 2015

THE QTN PROGRAM AND THE ALLELES THAT MATTER FOR EVOLUTION: ALL THAT'S GOLD DOES NOT GLITTER

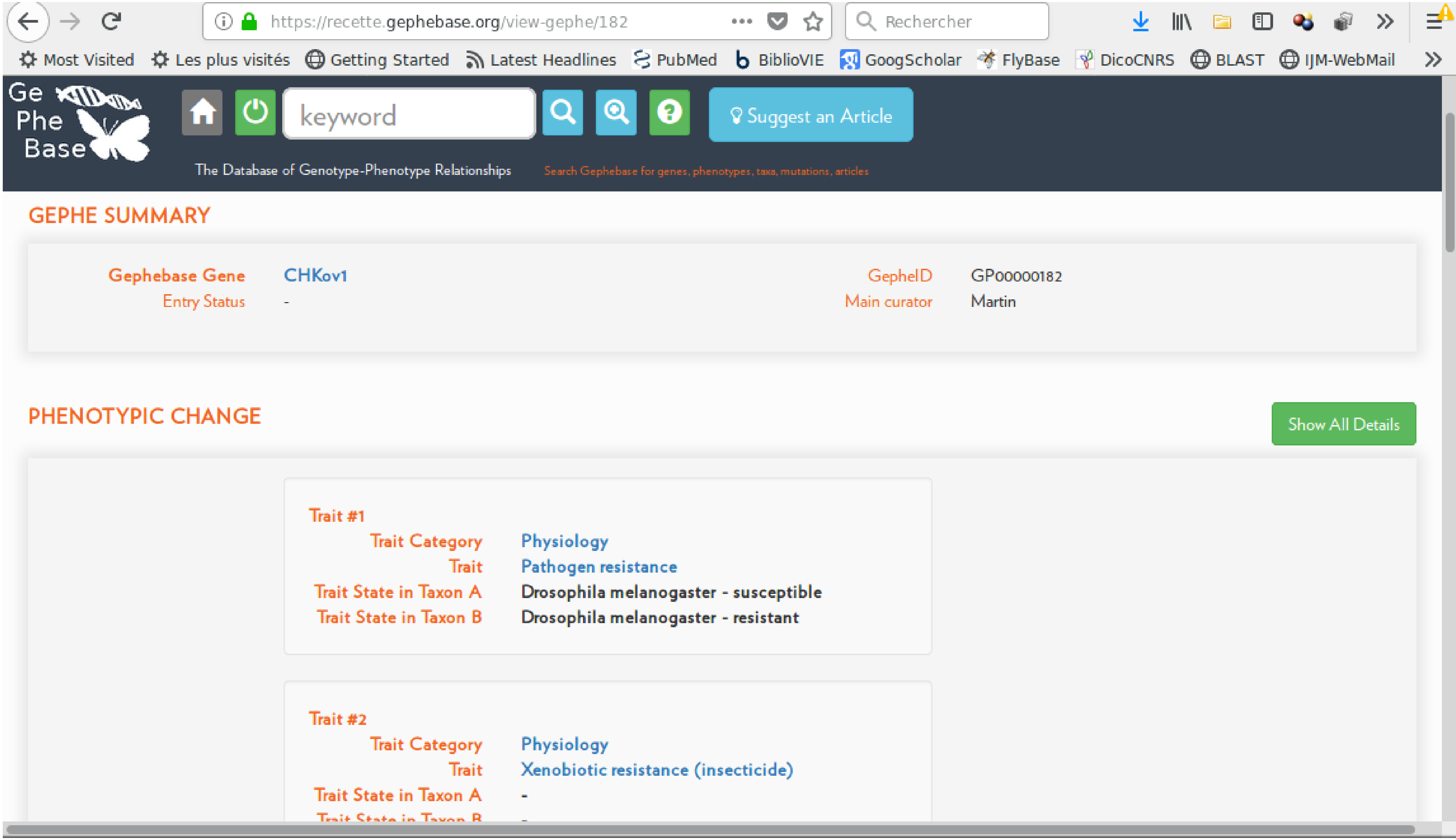
Matthew V. Rockman[1,2]

Low-hanging but juicy

The shiny
"**large effect**" loci we can document experimentally



CHKov1



GEPHE SUMMARY

Gephebase Gene	CHKov1	GepheID	GP00000182
Entry Status	-	Main curator	Martin

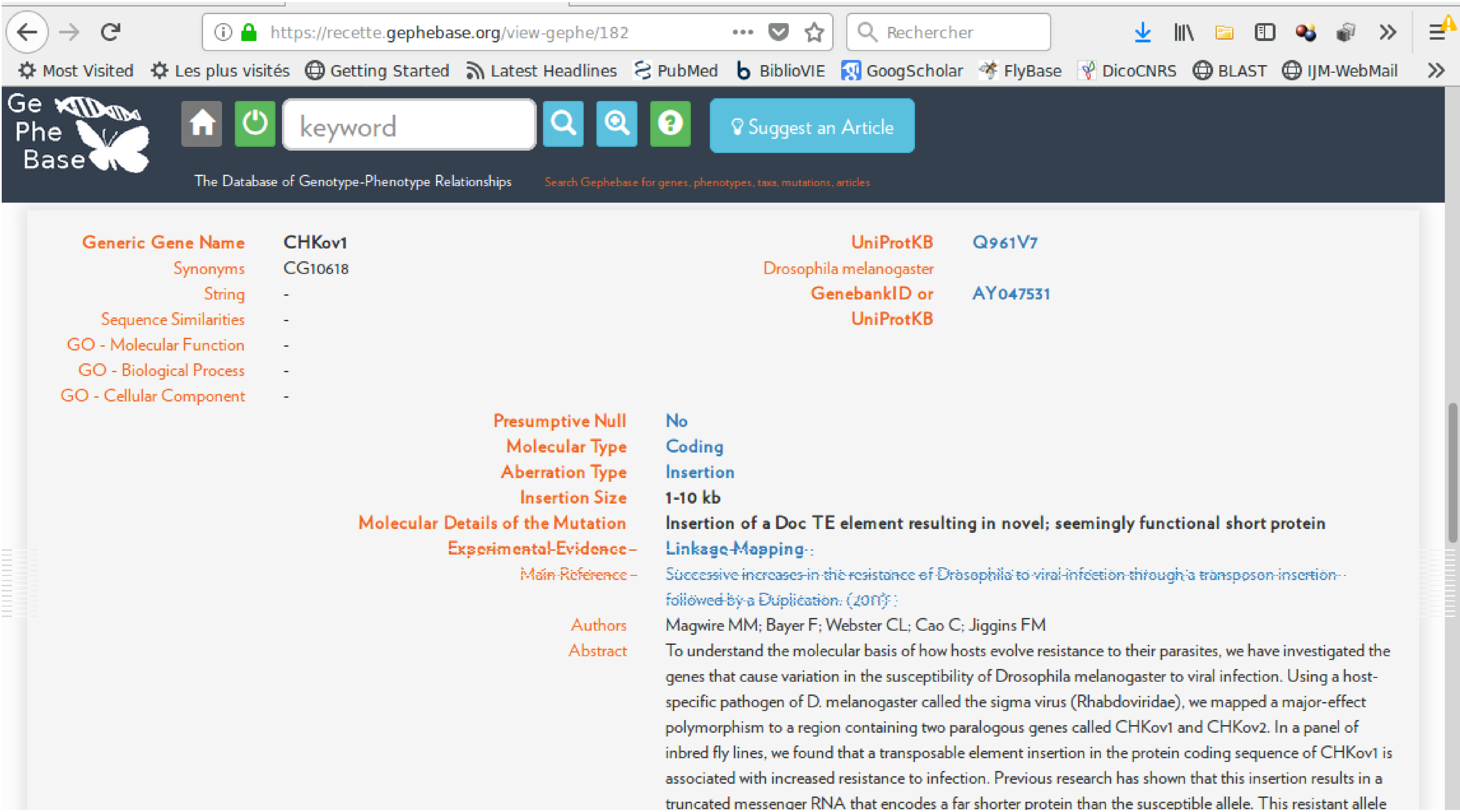
PHENOTYPIC CHANGE

Trait #1

Trait Category	Physiology
Trait	Pathogen resistance
Trait State in Taxon A	Drosophila melanogaster - susceptible
Trait State in Taxon B	Drosophila melanogaster - resistant

Trait #2

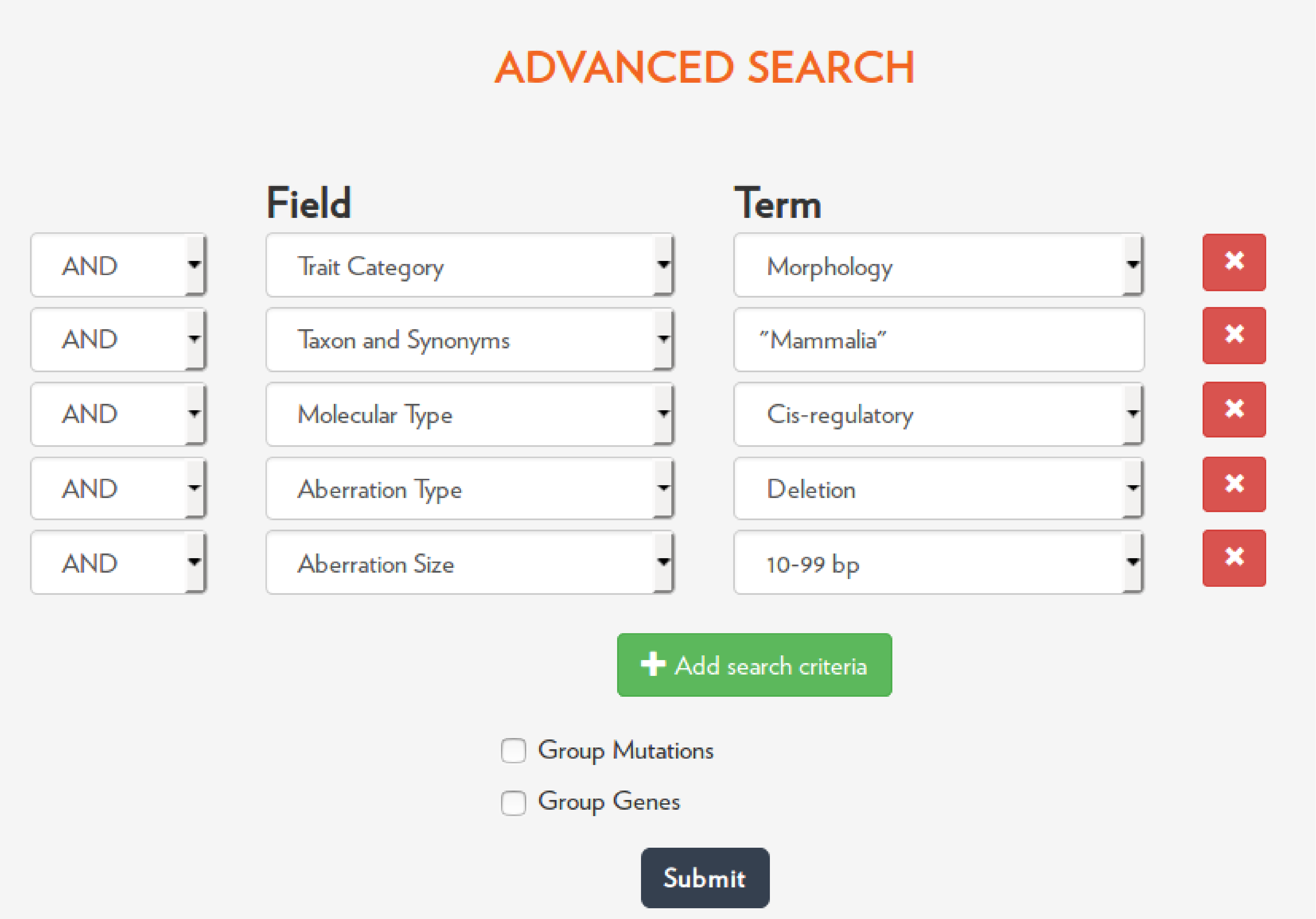
Trait Category	Physiology
Trait	Xenobiotic resistance (insecticide)
Trait State in Taxon A	-
Trait State in Taxon B	-



Generic Gene Name	**CHKov1**
Synonyms	CG10618
String	-
Sequence Similarities	-
GO - Molecular Function	-
GO - Biological Process	-
GO - Cellular Component	-

UniProtKB Drosophila melanogaster	**Q961V7**
GenebankID or UniProtKB	**AY047531**

Presumptive Null	**No**
Molecular Type	**Coding**
Aberration Type	**Insertion**
Insertion Size	**1-10 kb**
Molecular Details of the Mutation	**Insertion of a Doc TE element resulting in novel; seemingly functional short protein**
Experimental Evidence –	**Linkage Mapping**
Main Reference –	Successive increases in the resistance of Drosophila to viral infection through a transposon insertion followed by a Duplication. (2011)
Authors	Magwire MM; Bayer F; Webster CL; Cao C; Jiggins FM
Abstract	To understand the molecular basis of how hosts evolve resistance to their parasites, we have investigated the genes that cause variation in the susceptibility of Drosophila melanogaster to viral infection. Using a host-specific pathogen of D. melanogaster called the sigma virus (Rhabdoviridae), we mapped a major-effect polymorphism to a region containing two paralogous genes called CHKov1 and CHKov2. In a panel of inbred fly lines, we found that a transposable element insertion in the protein coding sequence of CHKov1 is associated with increased resistance to infection. Previous research has shown that this insertion results in a truncated messenger RNA that encodes a far shorter protein than the susceptible allele. This resistant allele



ADVANCED SEARCH

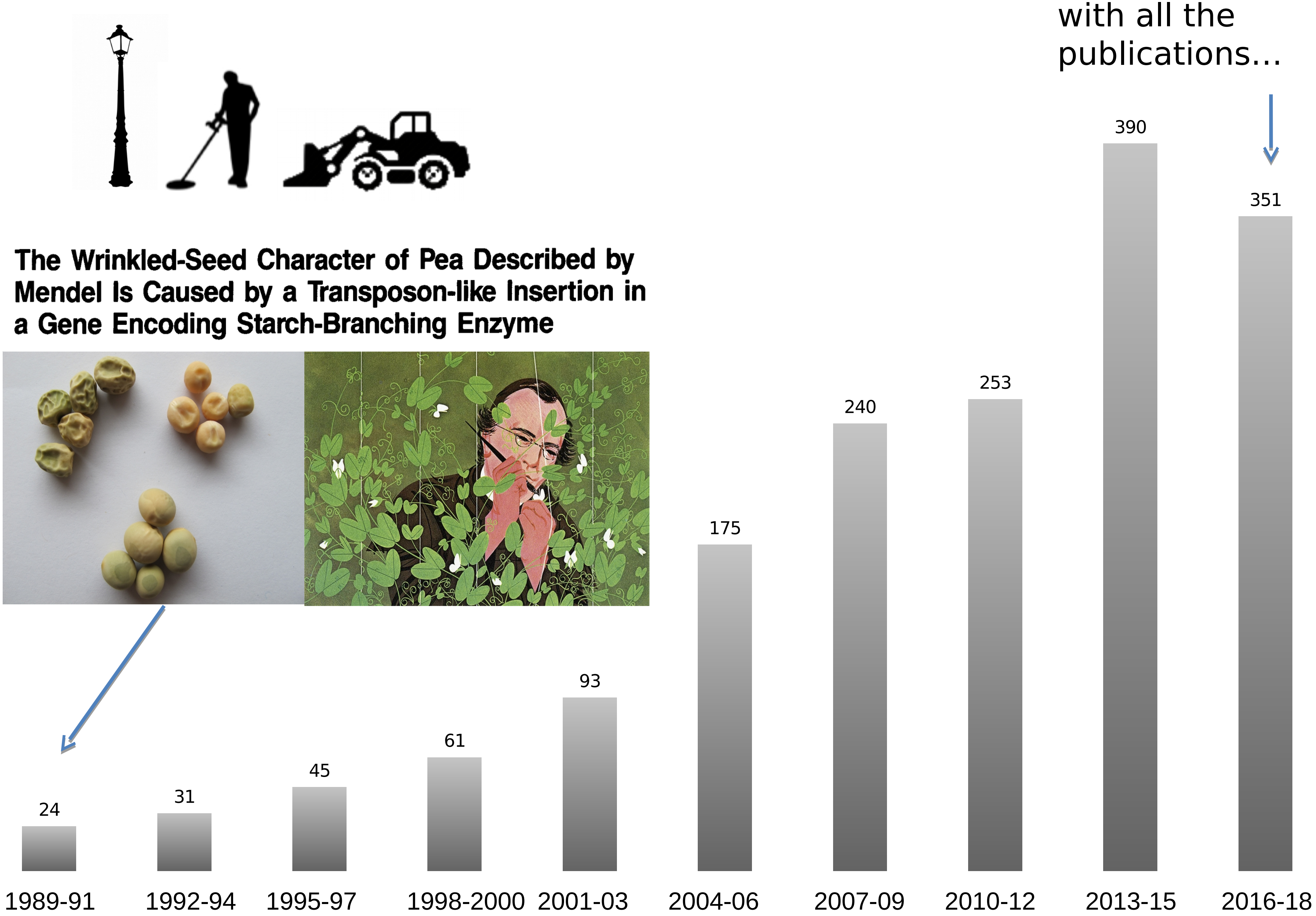
	Field	Term
AND	Trait Category	Morphology
AND	Taxon and Synonyms	"Mammalia"
AND	Molecular Type	Cis-regulatory
AND	Aberration Type	Deletion
AND	Aberration Size	10-99 bp

☐ Group Mutations

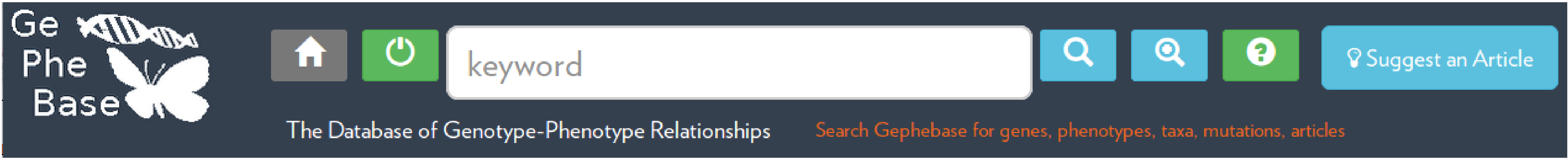
☐ Group Genes

Since 2015:
hard to keep up
with all the
publications…

The Wrinkled-Seed Character of Pea Described by Mendel Is Caused by a Transposon-like Insertion in a Gene Encoding Starch-Branching Enzyme



www.recette.gephebase.org



The Database of Genotype-Phenotype Relationships

Search Gephebase for genes, phenotypes, taxa, mutations, articles

@gephebase

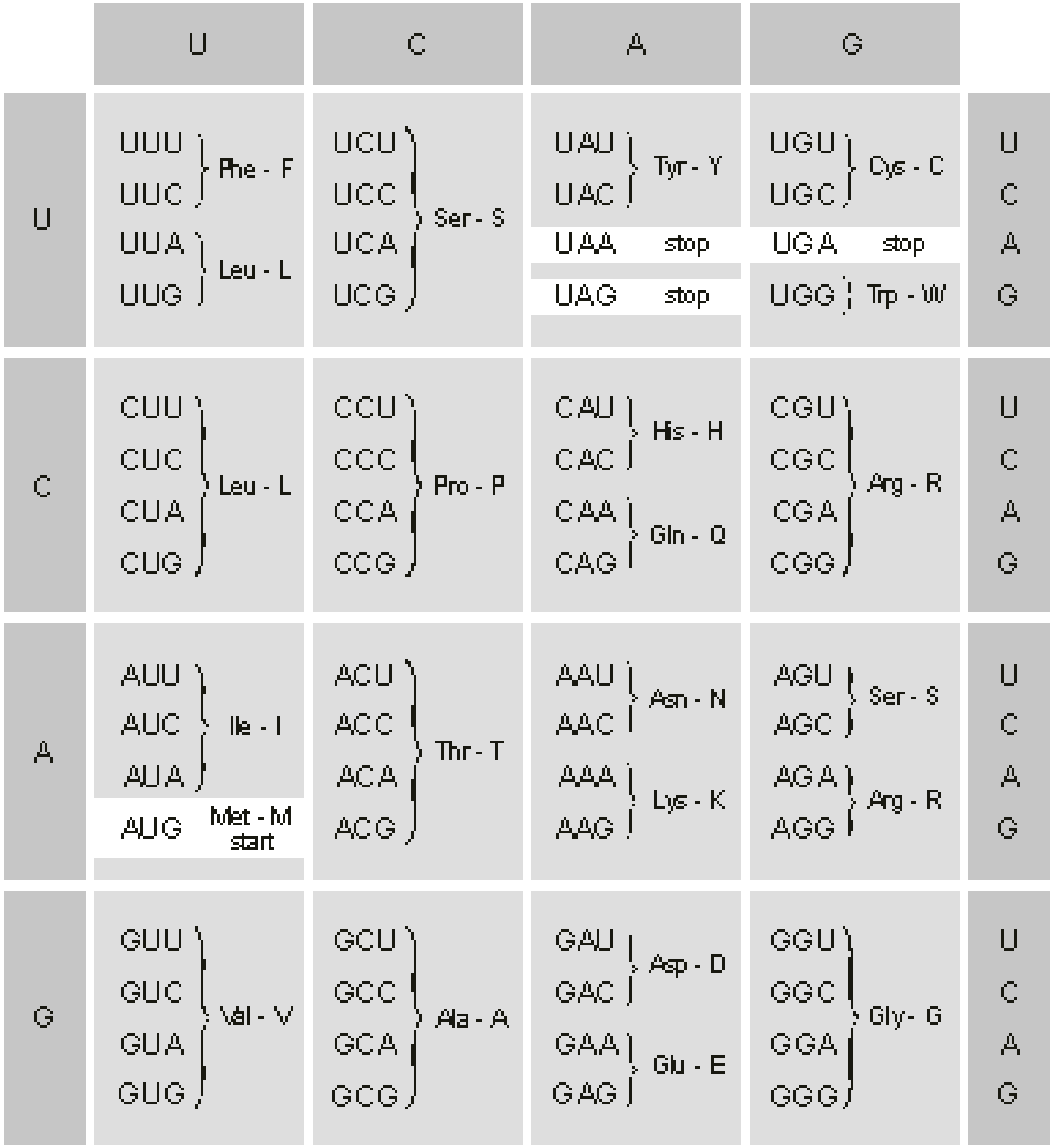
Virginie.courtier@ijm.fr

Meta-analysis of the coding SNP

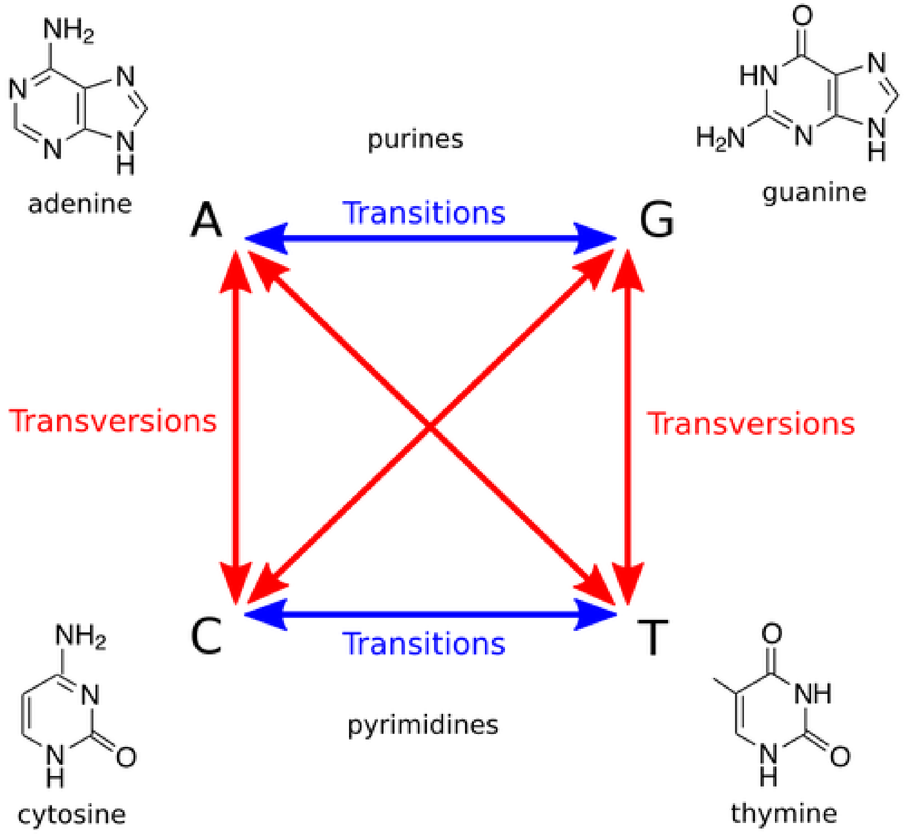
Transitions or transversions?

Which codon site?

Variation according to taxonomic group? Or to taxonomic status (experimental, domestication, intraspecific, interspecific, intergenera)?

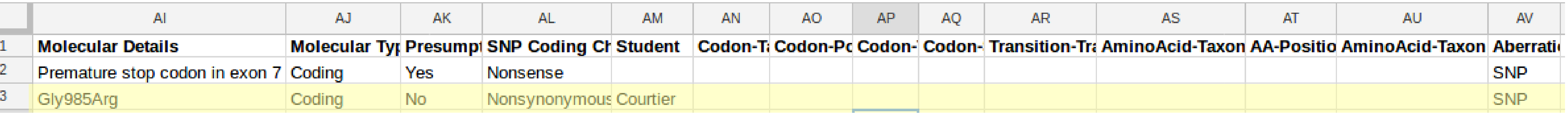


	U	C	A	G	
U	UUU UUC Phe - F UUA UUG Leu - L	UCU UCC UCA UCG Ser - S	UAU UAC Tyr - Y UAA stop UAG stop	UGU UGC Cys - C UGA stop UGG Trp - W	U C A G
C	CUU CUC CUA CUG Leu - L	CCU CCC CCA CCG Pro - P	CAU CAC His - H CAA CAG Gln - Q	CGU CGC CGA CGG Arg - R	U C A G
A	AUU AUC AUA Ile - I AUG Met - M start	ACU ACC ACA ACG Thr - T	AAU AAC Asn - N AAA AAG Lys - K	AGU AGC Ser - S AGA AGG Arg - R	U C A G
G	GUU GUC GUA GUG Val - V	GCU GCC GCA GCG Ala - A	GAU GAC Asp - D GAA GAG Glu - E	GGU GGC GGA GGG Gly - G	U C A G



CYtosine et TYmine are PYrimidines

Case 1: diapause in C. elegans desert strain *scd-2* gene



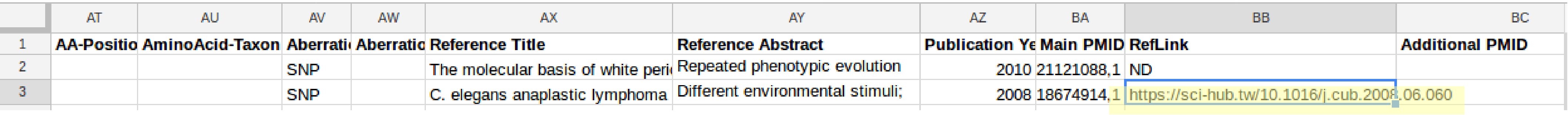
	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV
1	**Molecular Details**	**Molecular Typ**	**Presumpt**	**SNP Coding Ch**	**Student**	**Codon-T**	**Codon-P**	**Codon-**	**Codon-**	**Transition-Tra**	**AminoAcid-Taxon**	**AA-Positio**	**AminoAcid-Taxon**	**Aberrati**
2	Premature stop codon in exon 7	Coding	Yes	Nonsense										SNP
3	Gly985Arg	Coding	No	Nonsynonymous	Courtier									SNP

Gephebase-entries
https://tinyurl.com/y7pb3vth

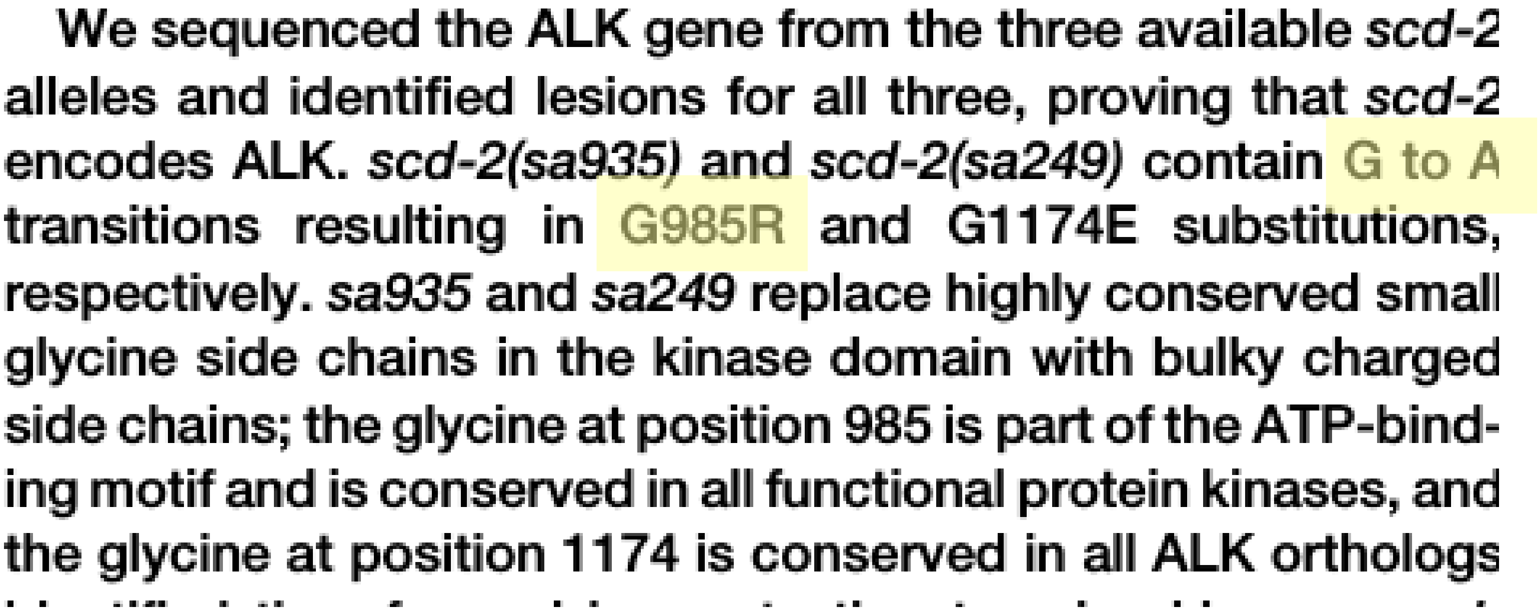
Protocol
https://tinyurl.com/y9eaxwk6

ExtractedData
https://tinyurl.com/ycuw25ax

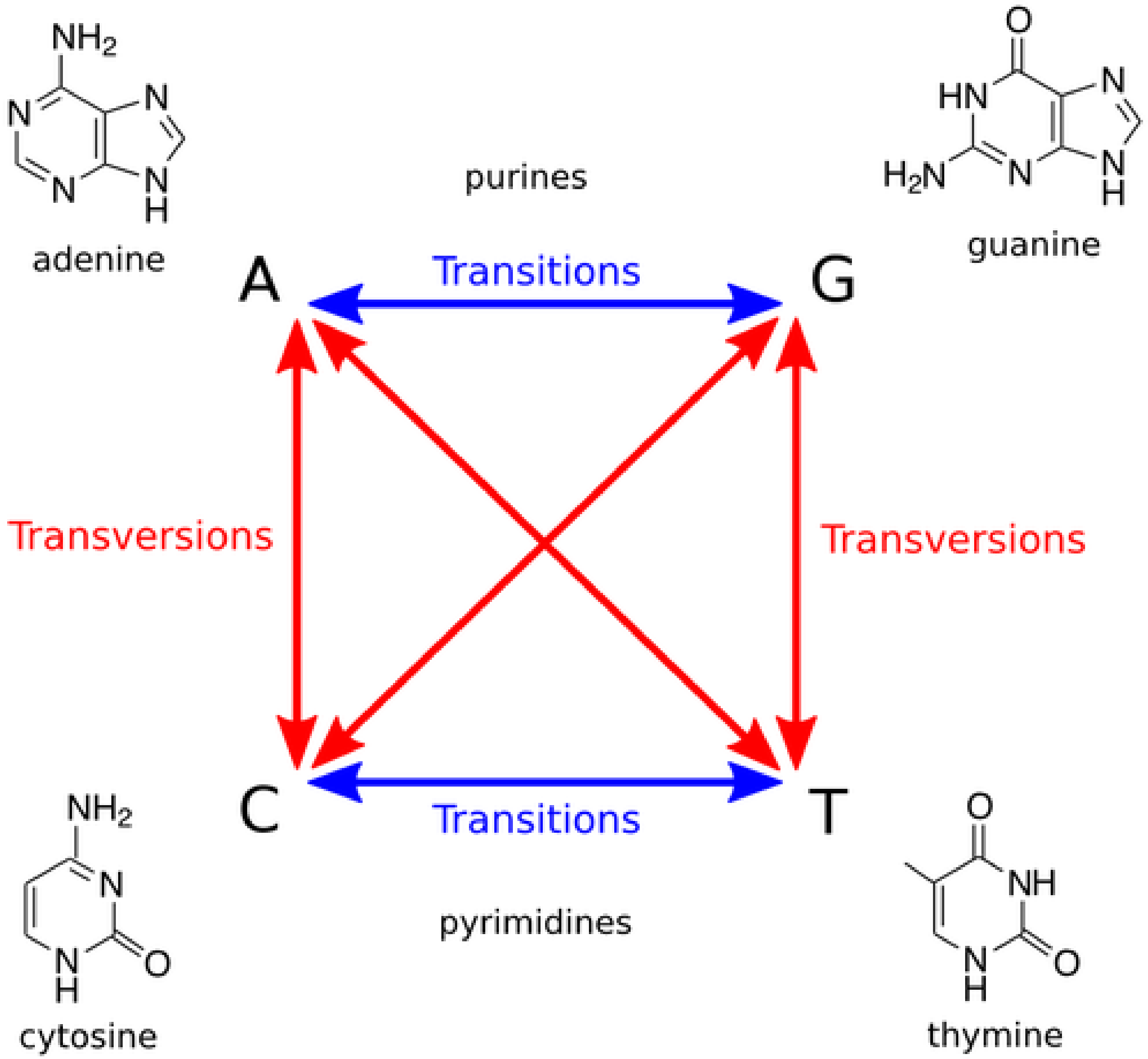
	U	C	A	G	
U	UUU Phe - F	UCU Ser - S	UAU Tyr - Y	UGU Cys - C	U
	UUC	UCC	UAC	UGC	C
	UUA Leu - L	UCA	UAA stop	UGA stop	A
	UUG	UCG	UAG stop	UGG Trp - W	G
C	CUU Leu - L	CCU Pro - P	CAU His - H	CGU Arg - R	U
	CUC	CCC	CAC	CGC	C
	CUA	CCA	CAA Gln - Q	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU Ile - I	ACU Thr - T	AAU Asn - N	AGU Ser - S	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA Lys - K	AGA Arg - R	A
	AUG Met - M start	ACG	AAG	AGG	G
G	GUU Val - V	GCU Ala - A	GAU Asp - D	GGU Gly - G	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glu - E	GGA	A
	GUG	GCG	GAG	GGG	G



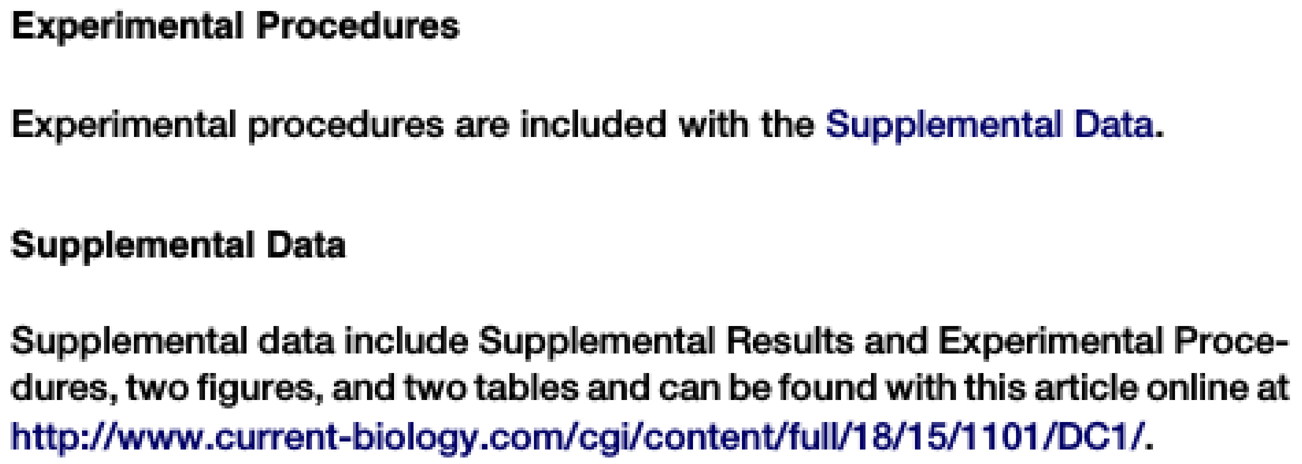
	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC
1	**AA-Positio**	**AminoAcid-Taxon**	**Aberrati**	**Aberratio**	**Reference Title**	**Reference Abstract**	**Publication Ye**	**Main PMID**	**RefLink**	**Additional PMID**
2			SNP		The molecular basis of white peri	Repeated phenotypic evolution	2010	21121088,1	ND	
3			SNP		C. elegans anaplastic lymphoma	Different environmental stimuli;	2008	18674914,1	https://sci-hub.tw/10.1016/j.cub.2008.06.060	



We sequenced the ALK gene from the three available *scd-2* alleles and identified lesions for all three, proving that *scd-2* encodes ALK. *scd-2(sa935)* and *scd-2(sa249)* contain G to A transitions resulting in G985R and G1174E substitutions, respectively. *sa935* and *sa249* replace highly conserved small glycine side chains in the kinase domain with bulky charged side chains; the glycine at position 985 is part of the ATP-binding motif and is conserved in all functional protein kinases, and the glycine at position 1174 is conserved in all ALK orthologs



CYtosine et TYmine are PYrimidines



Experimental Procedures

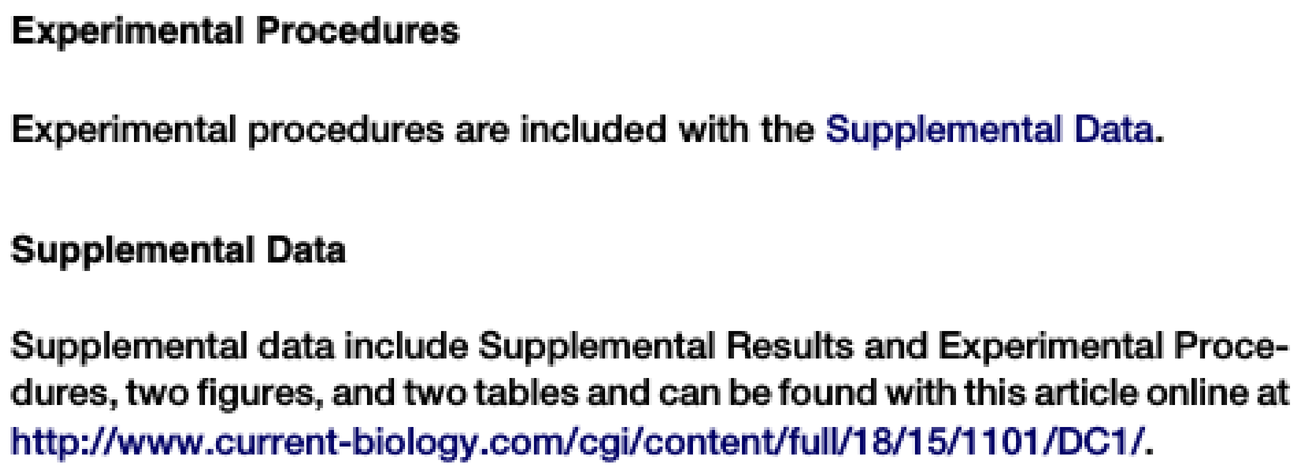
Experimental procedures are included with the Supplemental Data.

Supplemental Data

Supplemental data include Supplemental Results and Experimental Procedures, two figures, and two tables and can be found with this article online at http://www.current-biology.com/cgi/content/full/18/15/1101/DC1/.

No relevant data provided in the supplementary data
Accession numbers not provided

No sequence data except the wild-type sequence
when I searched “scd-2 C elegans” in NCBI Nucleotide



Experimental Procedures

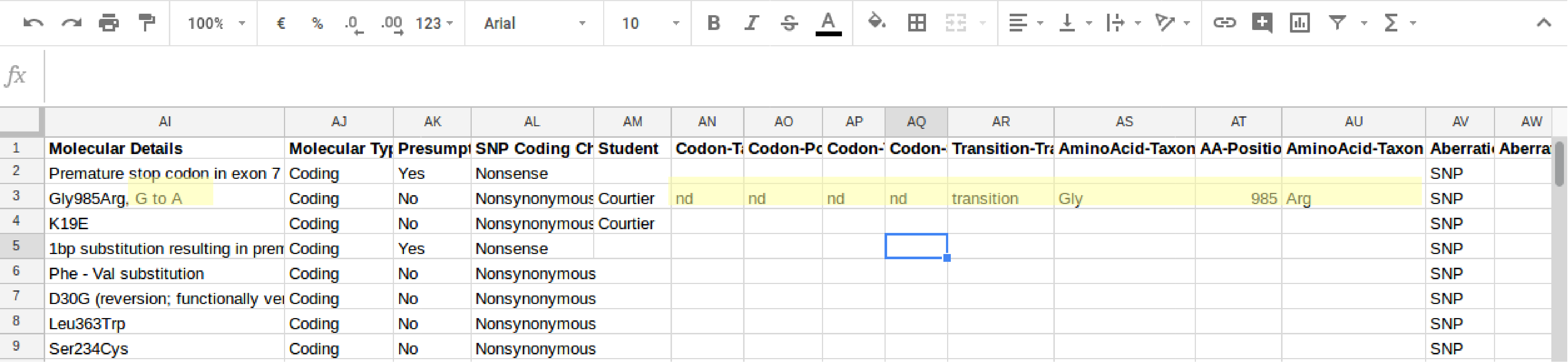
Experimental procedures are included with the Supplemental Data.

Supplemental Data

Supplemental data include Supplemental Results and Experimental Procedures, two figures, and two tables and can be found with this article online at http://www.current-biology.com/cgi/content/full/18/15/1101/DC1/.

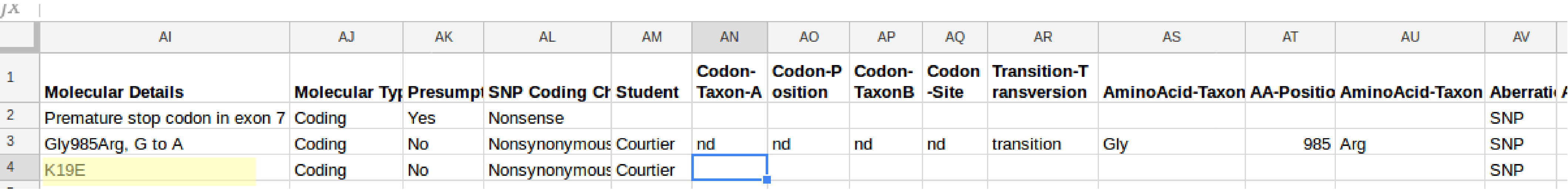
No accession numbers provided

No sequence data except the wild-type sequence
when I searched “scd-2 C elegans” in NCBI Nucleotide

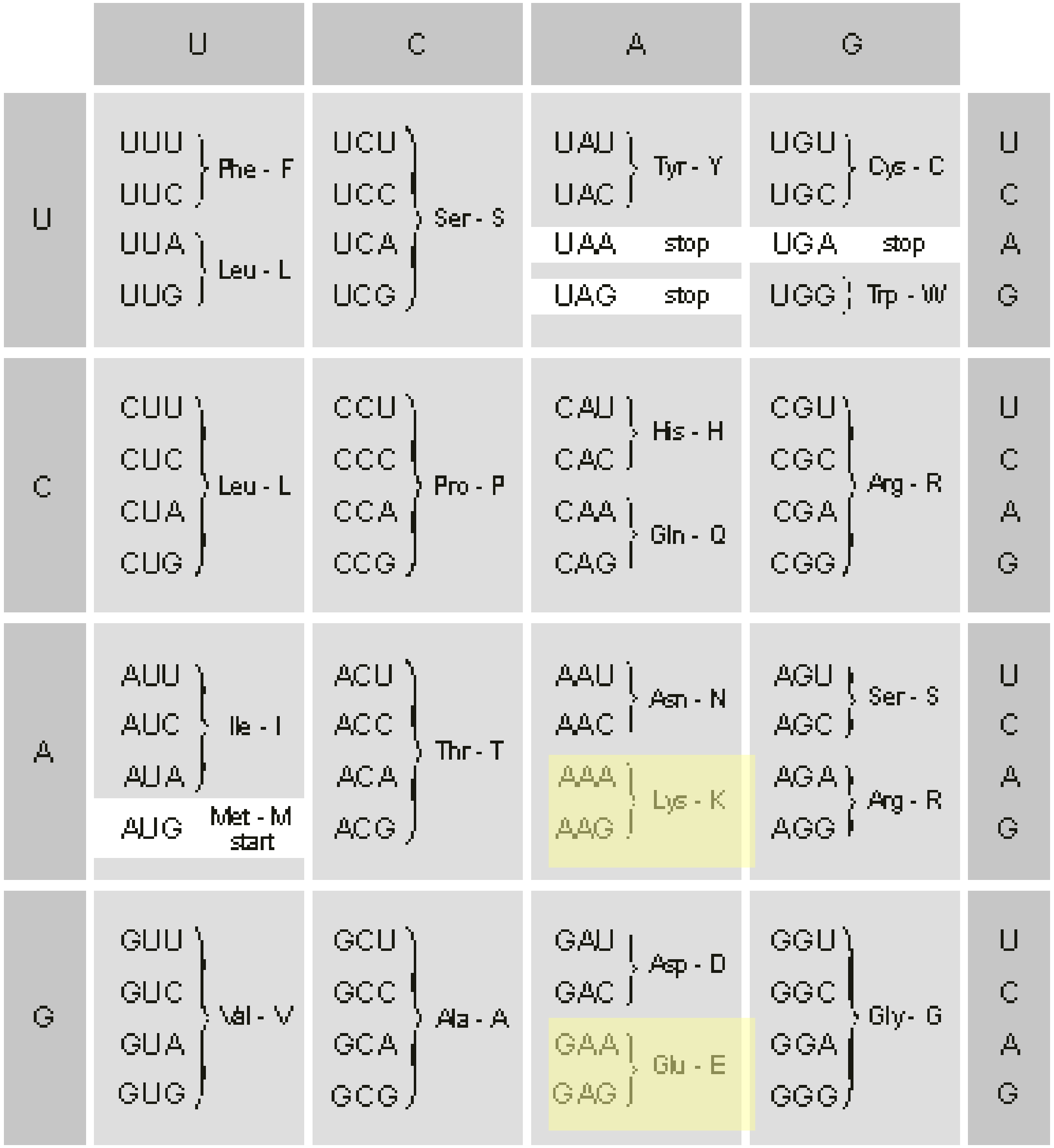


	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
1	**Molecular Details**	**Molecular Typ**	**Presumpt**	**SNP Coding Ch**	**Student**	**Codon-T**	**Codon-P**	**Codon-**	**Codon-**	**Transition-Tra**	**AminoAcid-Taxon**	**AA-Positio**	**AminoAcid-Taxon**	**Aberrati**	**Aberra**
2	Premature stop codon in exon 7	Coding	Yes	Nonsense										SNP	
3	Gly985Arg, G to A	Coding	No	Nonsynonymous	Courtier	nd	nd	nd	nd	transition	Gly	985	Arg	SNP	
4	K19E	Coding	No	Nonsynonymous	Courtier									SNP	
5	1bp substitution resulting in prem	Coding	Yes	Nonsense										SNP	
6	Phe - Val substitution	Coding	No	Nonsynonymous										SNP	
7	D30G (reversion; functionally ve	Coding	No	Nonsynonymous										SNP	
8	Leu363Trp	Coding	No	Nonsynonymous										SNP	
9	Ser234Cys	Coding	No	Nonsynonymous										SNP	

Case 2: trichome density in *A. thaliana ETC2* gene

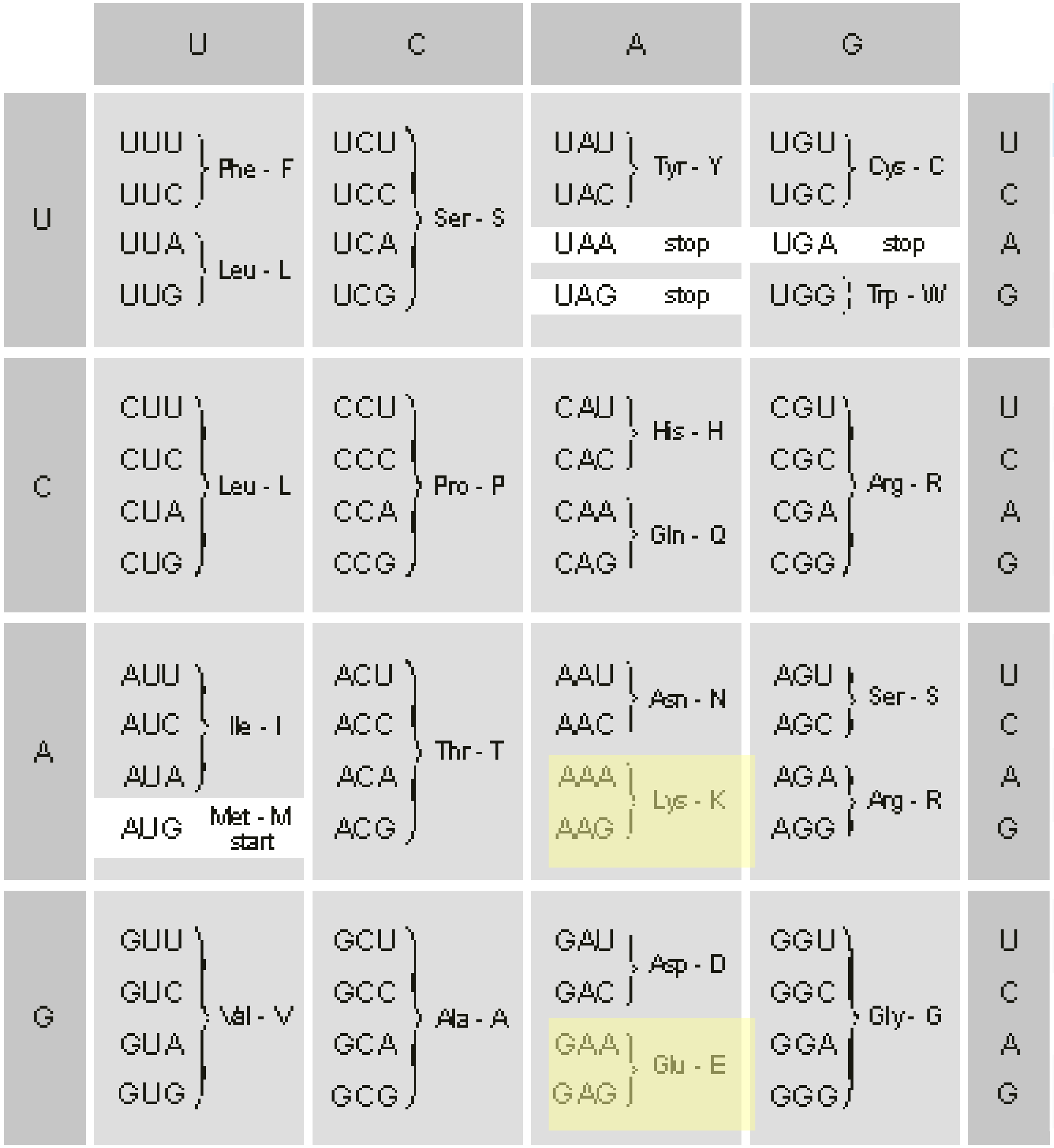


	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV
1	**Molecular Details**	**Molecular Typ**	**Presumpt**	**SNP Coding Ch**	**Student**	**Codon-Taxon-A**	**Codon-Position**	**Codon-TaxonB**	**Codon-Site**	**Transition-Transversion**	**AminoAcid-Taxon**	**AA-Positio**	**AminoAcid-Taxon**	**Aberrati**
2	Premature stop codon in exon 7	Coding	Yes	Nonsense										SNP
3	Gly985Arg, G to A	Coding	No	Nonsynonymous	Courtier	nd	nd	nd	nd	transition	Gly	985	Arg	SNP
4	K19E	Coding	No	Nonsynonymous	Courtier									SNP

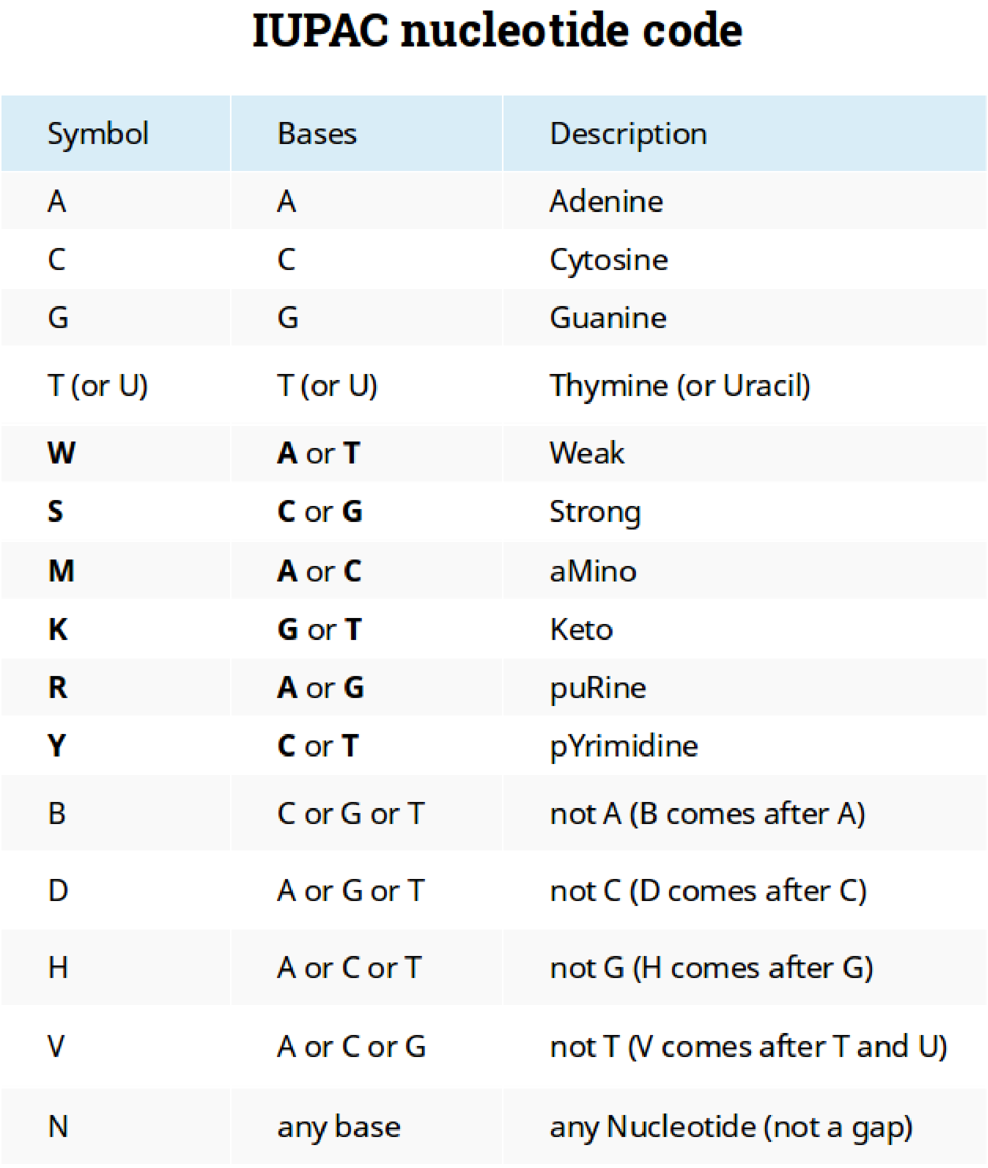


	U	C	A	G	
U	UUU Phe - F	UCU Ser - S	UAU Tyr - Y	UGU Cys - C	U
	UUC	UCC	UAC	UGC	C
	UUA Leu - L	UCA	UAA stop	UGA stop	A
	UUG	UCG	UAG stop	UGG Trp - W	G
C	CUU Leu - L	CCU Pro - P	CAU His - H	CGU Arg - R	U
	CUC	CCC	CAC	CGC	C
	CUA	CCA	CAA Gln - Q	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU Ile - I	ACU Thr - T	AAU Asn - N	AGU Ser - S	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA Lys - K	AGA Arg - R	A
	AUG Met - M start	ACG	AAG	AGG	G
G	GUU Val - V	GCU Ala - A	GAU Asp - D	GGU Gly - G	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glu - E	GGA	A
	GUG	GCG	GAG	GGG	G

GA AorG → AA AorG 1rst codon position



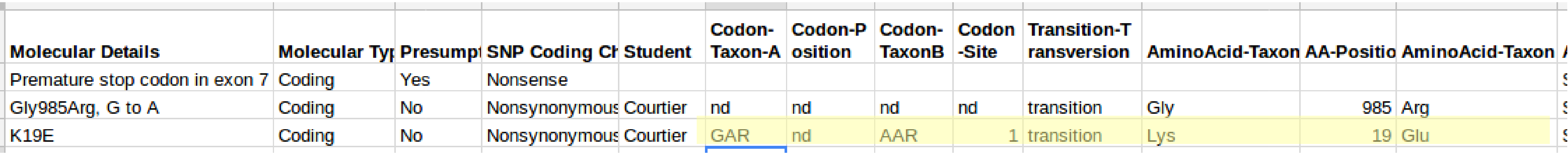
	U	C	A	G	
U	UUU UUC Phe - F UUA UUG Leu - L	UCU UCC UCA UCG Ser - S	UAU UAC Tyr - Y UAA stop UAG stop	UGU UGC Cys - C UGA stop UGG Trp - W	U C A G
C	CUU CUC CUA CUG Leu - L	CCU CCC CCA CCG Pro - P	CAU CAC His - H CAA CAG Gln - Q	CGU CGC CGA CGG Arg - R	U C A G
A	AUU AUC AUA Ile - I AUG Met - M start	ACU ACC ACA ACG Thr - T	AAU AAC Asn - N AAA AAG Lys - K	AGU AGC Ser - S AGA AGG Arg - R	U C A G
G	GUU GUC GUA GUG Val - V	GCU GCC GCA GCG Ala - A	GAU GAC Asp - D GAA GAG Glu - E	GGU GGC GGA GGG Gly - G	U C A G



IUPAC nucleotide code

Symbol	Bases	Description
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T (or U)	T (or U)	Thymine (or Uracil)
W	**A** or **T**	Weak
S	**C** or **G**	Strong
M	**A** or **C**	aMino
K	**G** or **T**	Keto
R	**A** or **G**	puRine
Y	**C** or **T**	pYrimidine
B	C or G or T	not A (B comes after A)
D	A or G or T	not C (D comes after C)
H	A or C or T	not G (H comes after G)
V	A or C or G	not T (V comes after T and U)
N	any base	any Nucleotide (not a gap)

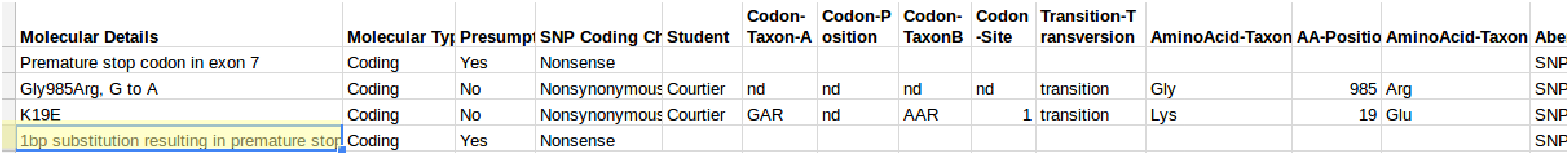
GA AorG → AA AorG 1rst codon position
GAR → AAR 1rst codon position



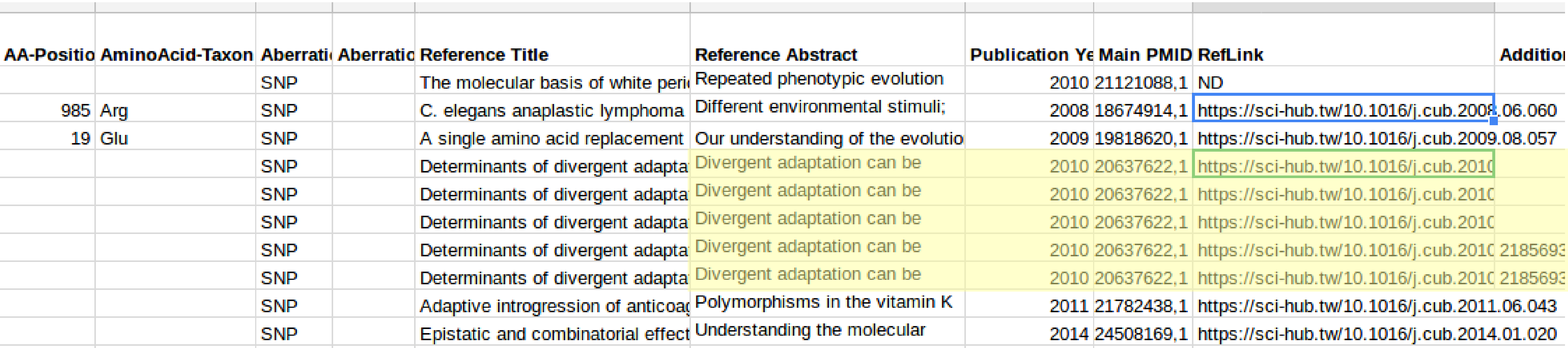
Molecular Details	Molecular Typ	Presumpt	SNP Coding Ch	Student	Codon-Taxon-A	Codon-Position	Codon-TaxonB	Codon-Site	Transition-Transversion	AminoAcid-Taxon	AA-Positio	AminoAcid-Taxon
Premature stop codon in exon 7	Coding	Yes	Nonsense									
Gly985Arg, G to A	Coding	No	Nonsynonymous	Courtier	nd	nd	nd	nd	transition	Gly	985	Arg
K19E	Coding	No	Nonsynonymous	Courtier	GAR	nd	AAR	1	transition	Lys	19	Glu

No need to check the reference in this case

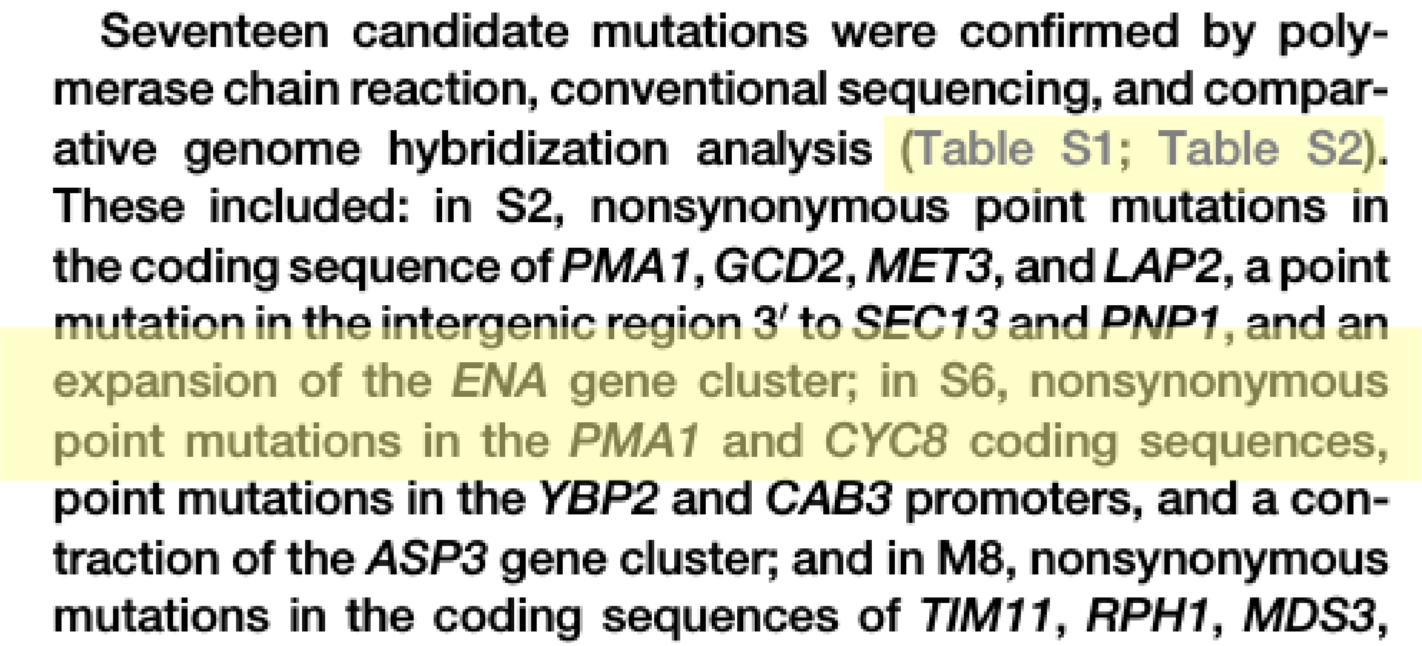
Case 3: salt tolerance (exerimental evol.) in S. *cerevisiae - CYC8* gene



Molecular Details	Molecular Typ	Presumpt	SNP Coding Ch	Student	Codon-Taxon-A	Codon-Position	Codon-TaxonB	Codon-Site	Transition-Transversion	AminoAcid-Taxon	AA-Positio	AminoAcid-Taxon	Abe
Premature stop codon in exon 7	Coding	Yes	Nonsense										SNP
Gly985Arg, G to A	Coding	No	Nonsynonymous	Courtier	nd	nd	nd	nd	transition	Gly	985	Arg	SNP
K19E	Coding	No	Nonsynonymous	Courtier	GAR	nd	AAR	1	transition	Lys	19	Glu	SNP
1bp substitution resulting in premature stop	Coding	Yes	Nonsense										SNP



AA-Positio	AminoAcid-Taxon	Aberrati	Aberratio	Reference Title	Reference Abstract	Publication Ye	Main PMID	RefLink	Additio
		SNP		The molecular basis of white peri	Repeated phenotypic evolution	2010	21121088,1	ND	
985	Arg	SNP		C. elegans anaplastic lymphoma	Different environmental stimuli;	2008	18674914,1	https://sci-hub.tw/10.1016/j.cub.2008.06.060	
19	Glu	SNP		A single amino acid replacement	Our understanding of the evolutio	2009	19818620,1	https://sci-hub.tw/10.1016/j.cub.2009.08.057	
		SNP		Determinants of divergent adapta	Divergent adaptation can be	2010	20637622,1	https://sci-hub.tw/10.1016/j.cub.2010	
		SNP		Determinants of divergent adapta	Divergent adaptation can be	2010	20637622,1	https://sci-hub.tw/10.1016/j.cub.2010	
		SNP		Determinants of divergent adapta	Divergent adaptation can be	2010	20637622,1	https://sci-hub.tw/10.1016/j.cub.2010	
		SNP		Determinants of divergent adapta	Divergent adaptation can be	2010	20637622,1	https://sci-hub.tw/10.1016/j.cub.2010	2185693
		SNP		Determinants of divergent adapta	Divergent adaptation can be	2010	20637622,1	https://sci-hub.tw/10.1016/j.cub.2010	2185693
		SNP		Adaptive introgression of anticoa	Polymorphisms in the vitamin K	2011	21782438,1	https://sci-hub.tw/10.1016/j.cub.2011.06.043	
		SNP		Epistatic and combinatorial effect	Understanding the molecular	2014	24508169,1	https://sci-hub.tw/10.1016/j.cub.2014.01.020	

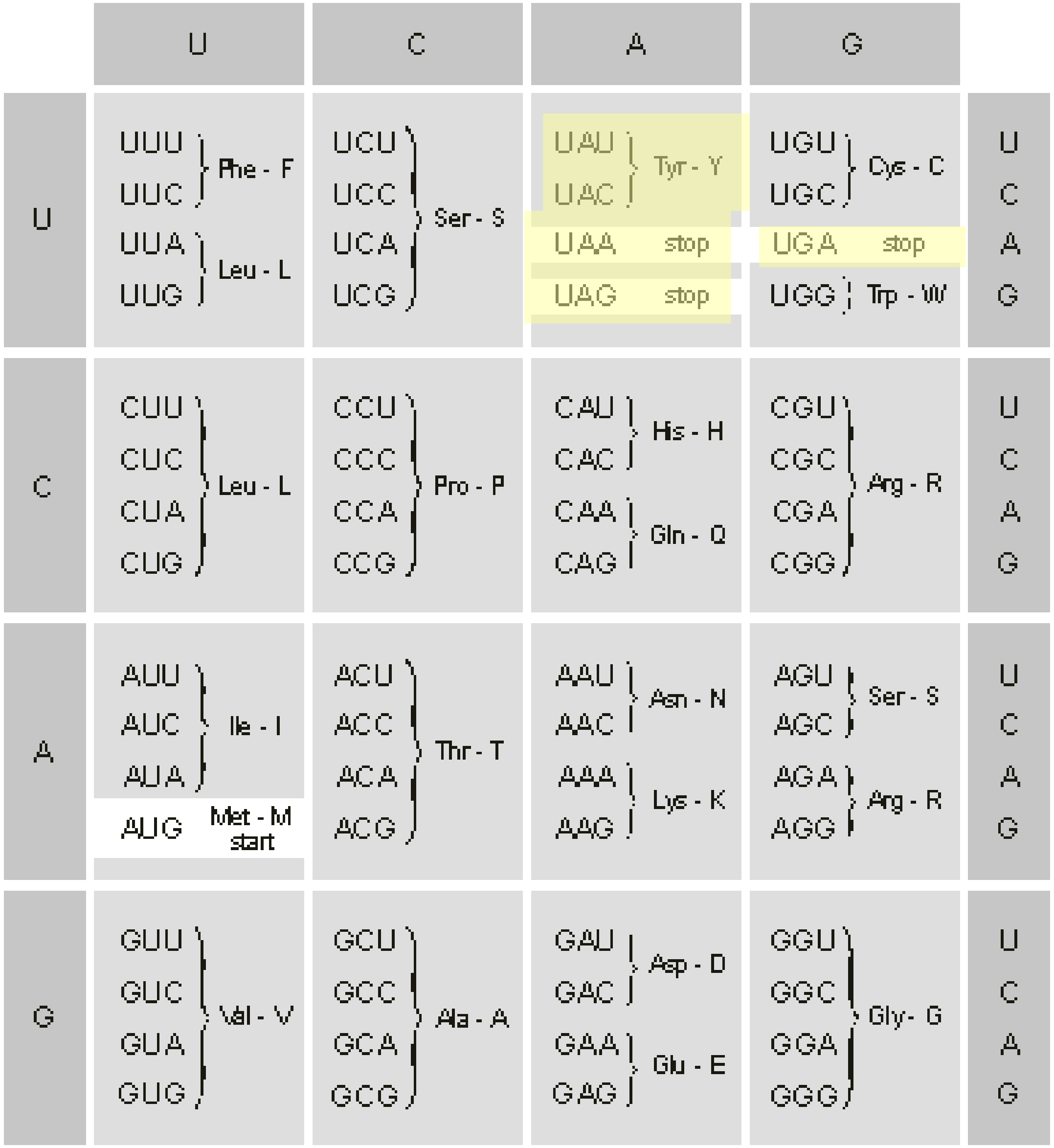


Seventeen candidate mutations were confirmed by polymerase chain reaction, conventional sequencing, and comparative genome hybridization analysis (Table S1; Table S2). These included: in S2, nonsynonymous point mutations in the coding sequence of *PMA1*, *GCD2*, *MET3*, and *LAP2*, a point mutation in the intergenic region 3′ to *SEC13* and *PNP1*, and an expansion of the *ENA* gene cluster; in S6, nonsynonymous point mutations in the *PMA1* and *CYC8* coding sequences, point mutations in the *YBP2* and *CAB3* promoters, and a contraction of the *ASP3* gene cluster; and in M8, nonsynonymous mutations in the coding sequences of *TIM11*, *RPH1*, *MDS3*,

Download supplementary tables

Table S1. SNPs and Gene-Cluster Size Changes in Haploid Representatives of S2, S6, and M8

Chromosome	Position, Base, and Amino-Acid Changes	Gene	Notes
S2 Haploid			
7	481971 G-C S-C	*PMA1*	Main adaptive determinant in high salt
7	646331 C-A M-I	*GCD2*	No effect on fitness detected
10	456758 C-A P-H	*MET3*	Confers a no-growth phenotype (auxotrophy) in low glucose
12	560742 C-G	Intergenic region 3' to YLR208W and YLR209C	No effect on fitness detected
14	530695-538465	Expansion, *ENA1*, *ENA2*, *ENA5*	Adaptive determinant in high salt. Synergistic with evolved allele of *PMA1*
14	543274 A-G D-G	*LAP2*	No effect on fitness detected
S6 Haploid			
2	464706, G -T, Y-*	*CYC8*	Adaptive determinant in high salt, synergistic with the **S6** evolved allele of *PMA1*
7	481584 A-C, L-W	*PMA1*	Adaptive determinant in high salt. Confers slow growth in YPD and low glucose
7	390007, G-A	Noncoding, promoter of *YBP2*	No effect on fitness detected
11	274875, T-A	Noncoding, promoter of *CAB3*	No effect on fitness detected
12	470406- 486202	Contraction, *ASP3-1*, *ASP3-2*, *ASP3-3*, *ASP3-4*	No effect on fitness detected
M8 Haploid			
4	1112209 C-A C-F	*TIM11*	No effect on fitness detected
5	525696 C-G T-I	*RPH1*	Fitness benefit after the diauxic shift in low glucose, not necessary to reconstitute the full **M8** phenotype
7	126872 T-G F-V	*MDS3*	Fitness benefit before the diauxic shift in low glucose. Evolved allele confers a sporulation deficiency: 10.4% ± 0.8 SE (n=12 measurements) vs. 30.1 ± 1.0 SE (n=50)
12	64832 C-T Q-Q	*UBI4*	No effect on fitness detected
14	467221 A-G D-G	*MKT1*	Fitness benefit after the diauxic shift in low glucose
15	432852 G-A R-K	*SGT1*	No effect on fitness detected



	U	C	A	G	
U	UUU UUC Phe - F UUA UUG Leu - L	UCU UCC UCA UCG Ser - S	UAU UAC Tyr - Y UAA stop UAG stop	UGU UGC Cys - C UGA stop UGG Trp - W	U C A G
C	CUU CUC CUA CUG Leu - L	CCU CCC CCA CCG Pro - P	CAU CAC His - H CAA CAG Gln - Q	CGU CGC CGA CGG Arg - R	U C A G
A	AUU AUC AUA Ile - I AUG Met - M start	ACU ACC ACA ACG Thr - T	AAU AAC Asn - N AAA AAG Lys - K	AGU AGC Ser - S AGA AGG Arg - R	U C A G
G	GUU GUC GUA GUG Val - V	GCU GCC GCA GCG Ala - A	GAU GAC Asp - D GAA GAG Glu - E	GGU GGC GGA GGG Gly - G	U C A G

G → T
Y → stop

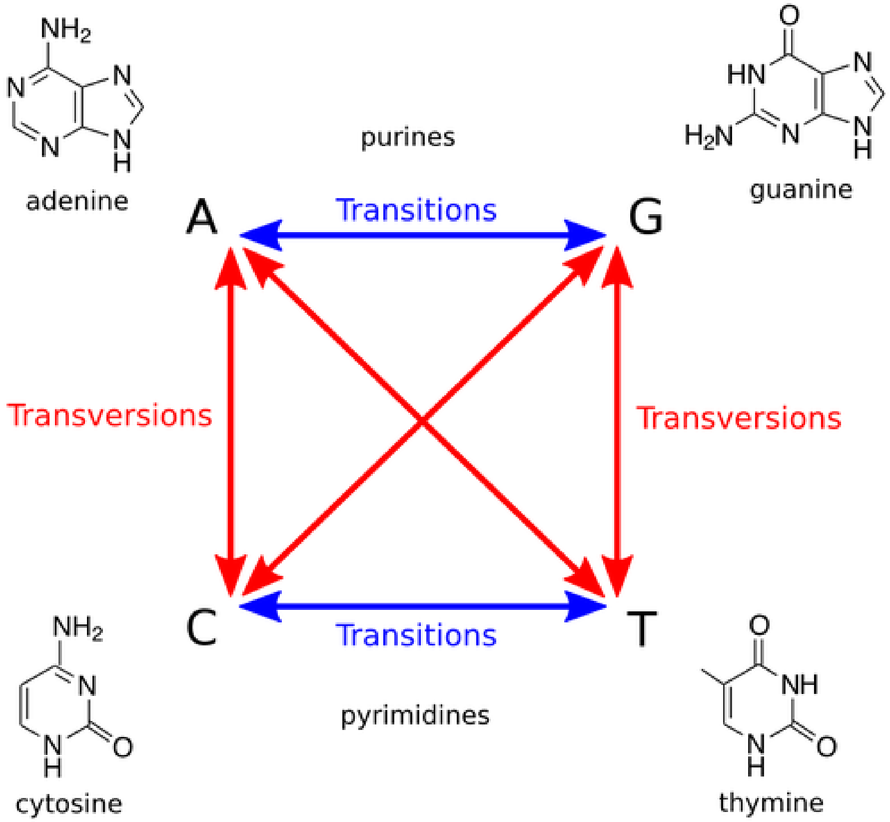
Impossible!
Error in Table S1

Let's consider that the error lies in the nucleotide report

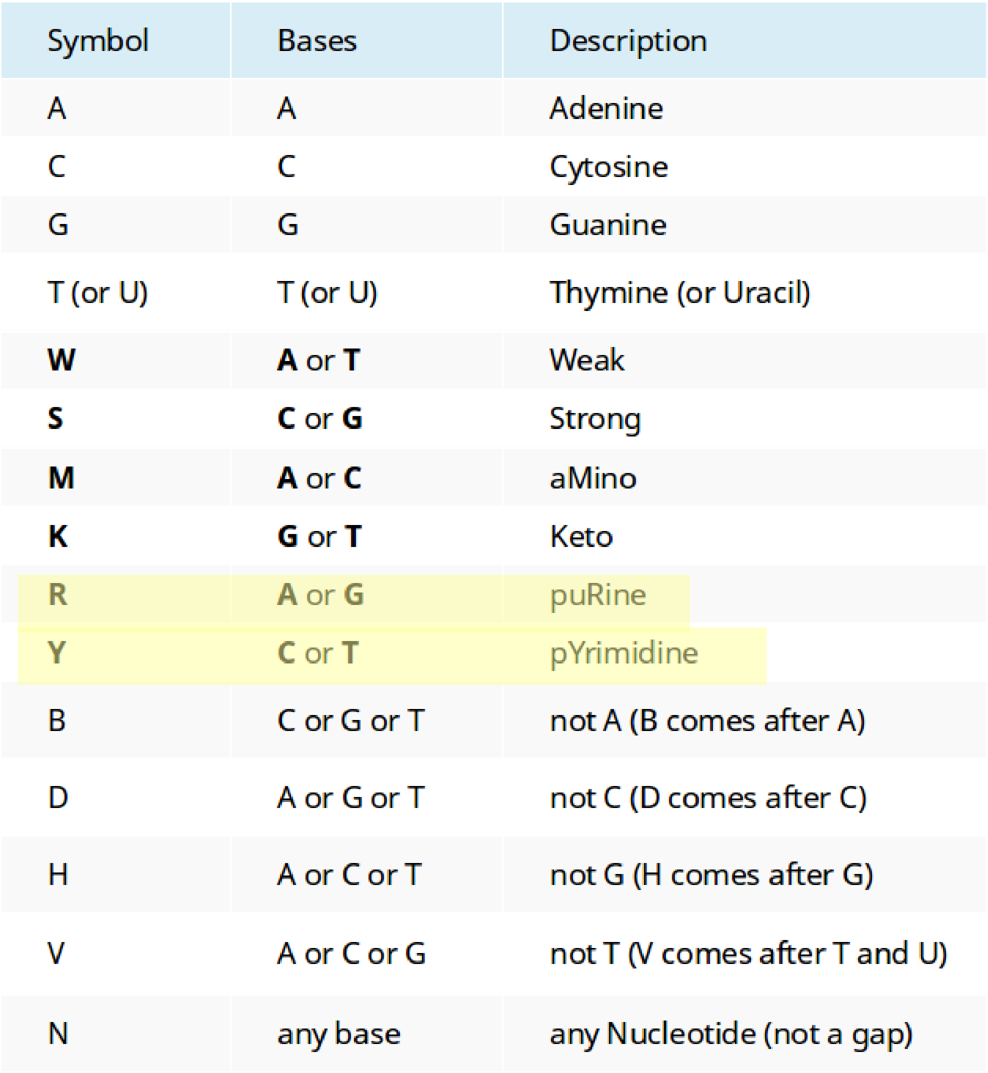
UAR → UAY

UAR → UAY

3rd position
Transversion



IUPAC nucleotide code



Symbol	Bases	Description
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T (or U)	T (or U)	Thymine (or Uracil)
W	**A** or **T**	Weak
S	**C** or **G**	Strong
M	**A** or **C**	aMino
K	**G** or **T**	Keto
R	**A** or **G**	puRine
Y	**C** or **T**	pYrimidine
B	C or G or T	not A (B comes after A)
D	A or G or T	not C (D comes after C)
H	A or C or T	not G (H comes after G)
V	A or C or G	not T (V comes after T and U)
N	any base	any Nucleotide (not a gap)

It's your turn!

Gephebase-entries
https://tinyurl.com/y7pb3vth

Protocol
https://tinyurl.com/y9eaxwk6

ExtractedData
https://tinyurl.com/ycuw25ax