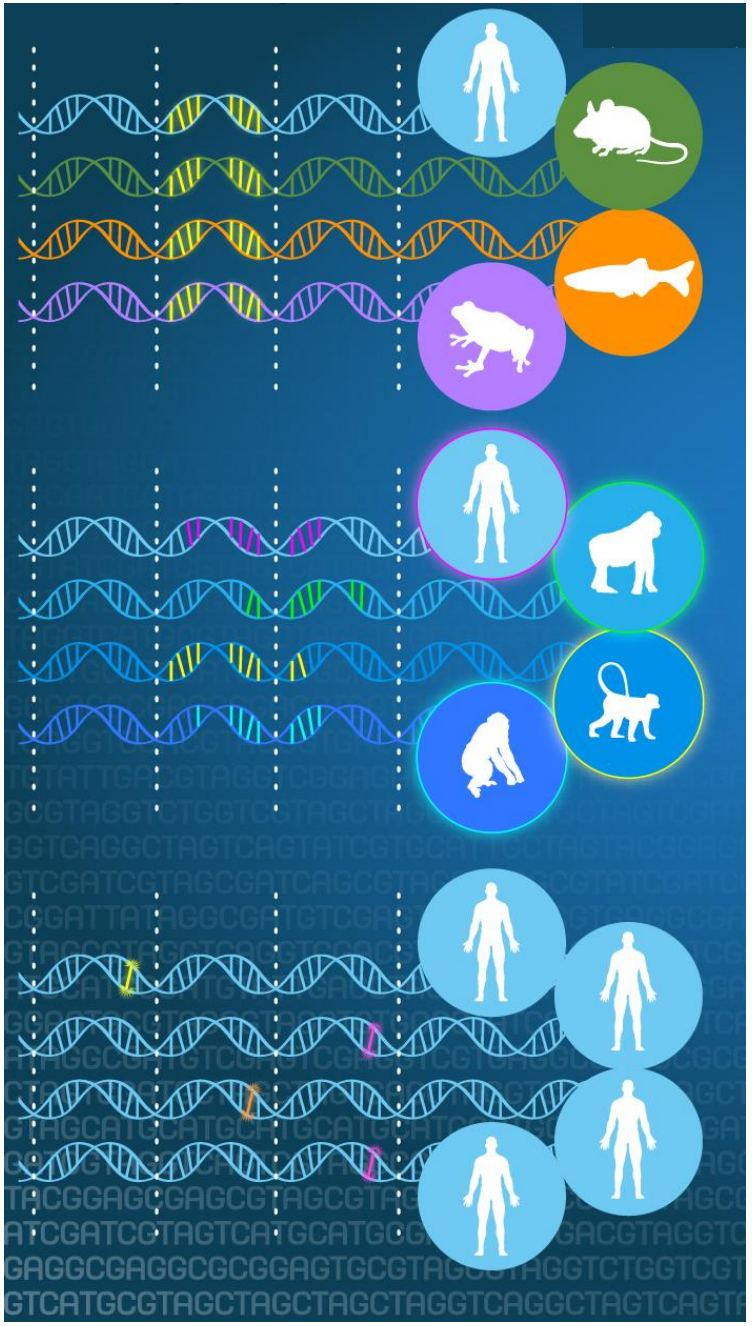
Comparative genomics

Ovidiu Paun
ovidiu.paun@univie.ac.at
http://plantgenomics.univie.ac.at

Slides available at
http://plantgenomics.univie.ac.at/MPGcourse



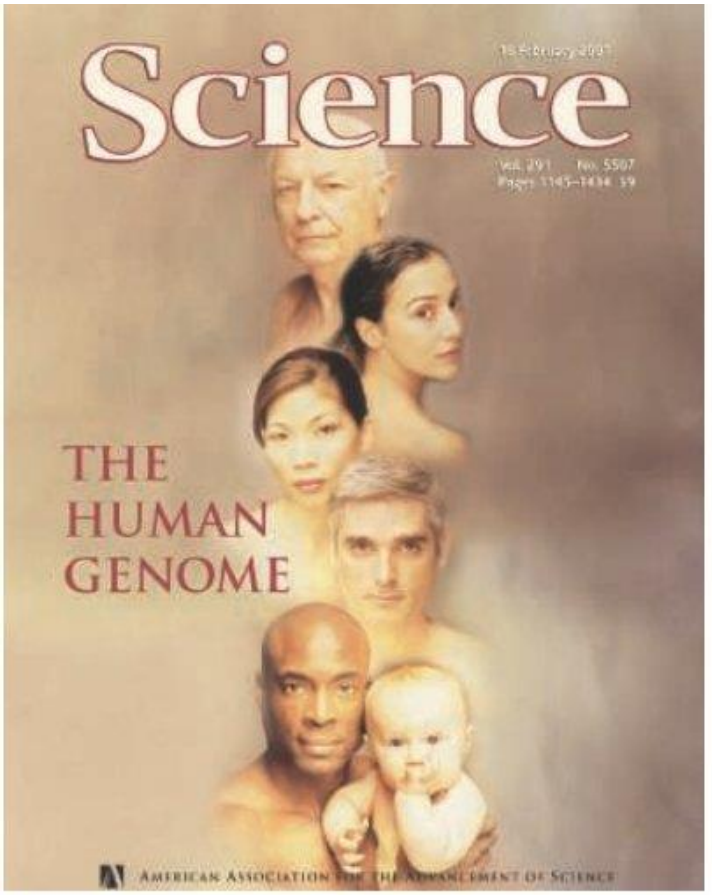
© www.genome.gov

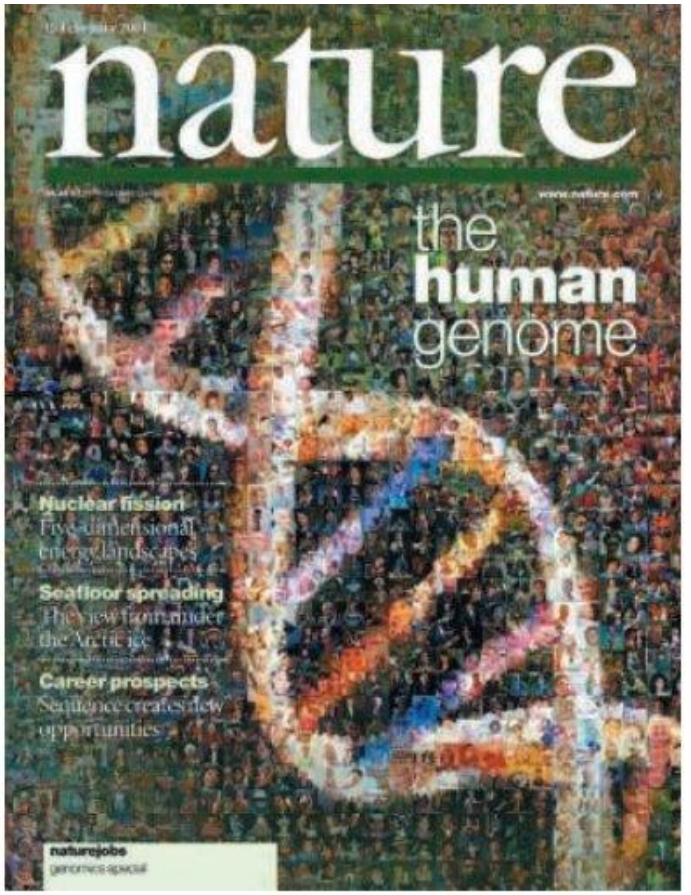
Why comparative genomics?

Evolution of genome features:

- Synteny
- Gene evolution
- Orthologs/paralogs

Human genome (Feb 2001)

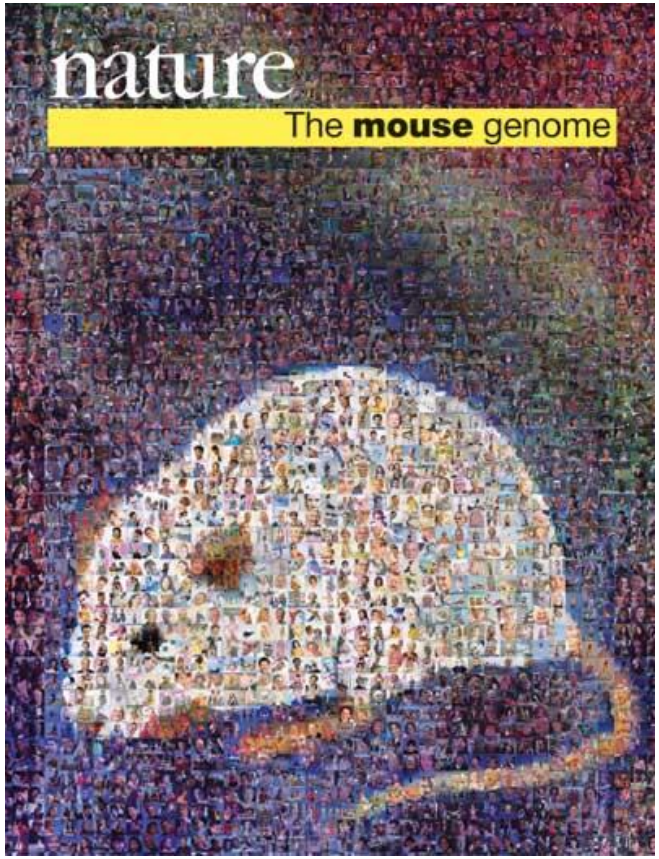


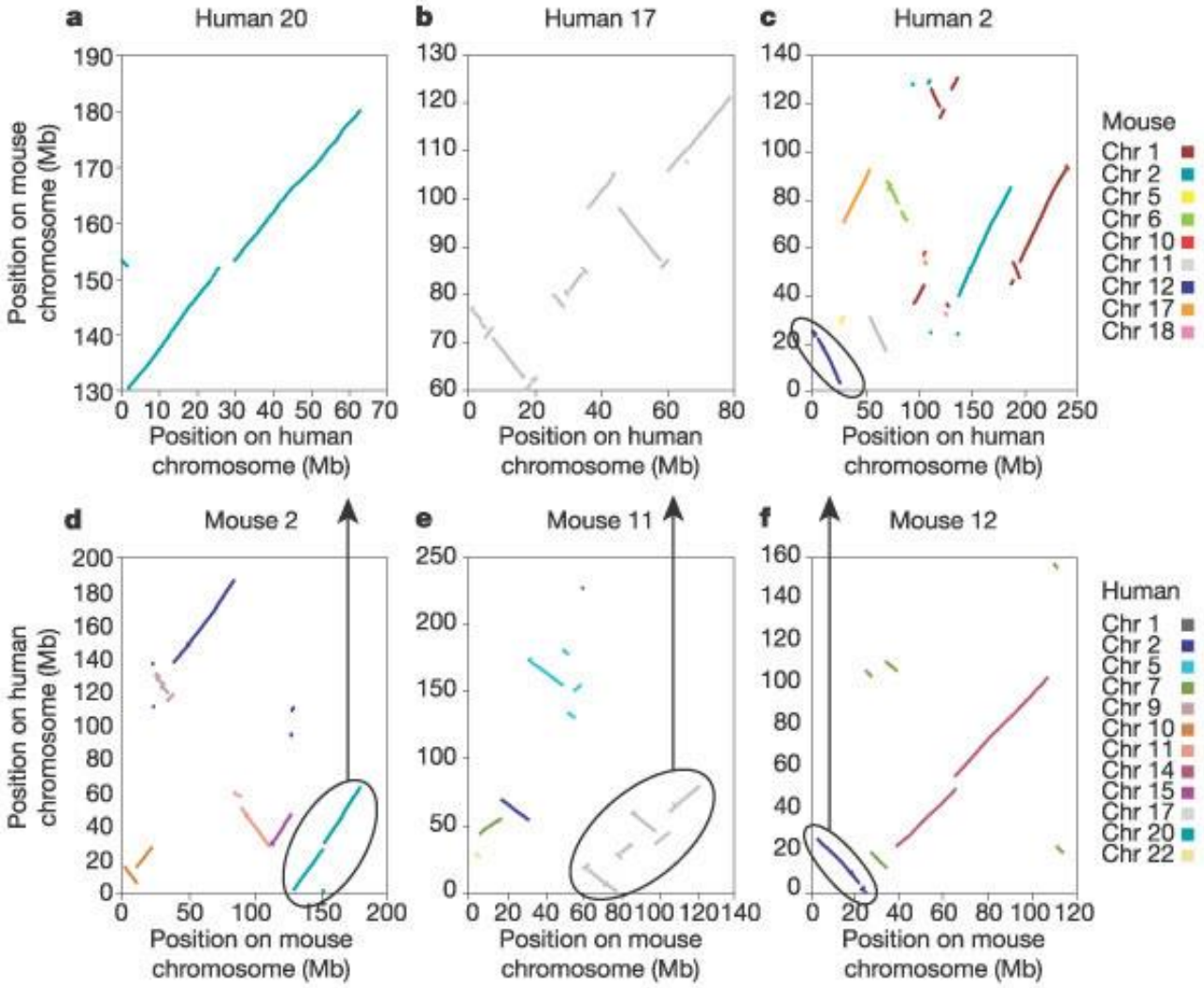


Ca 20,000 human protein-coding genes (ca 1.5% of the genome)
Biological functions of many genes still unclear

Early Comparative Genomics

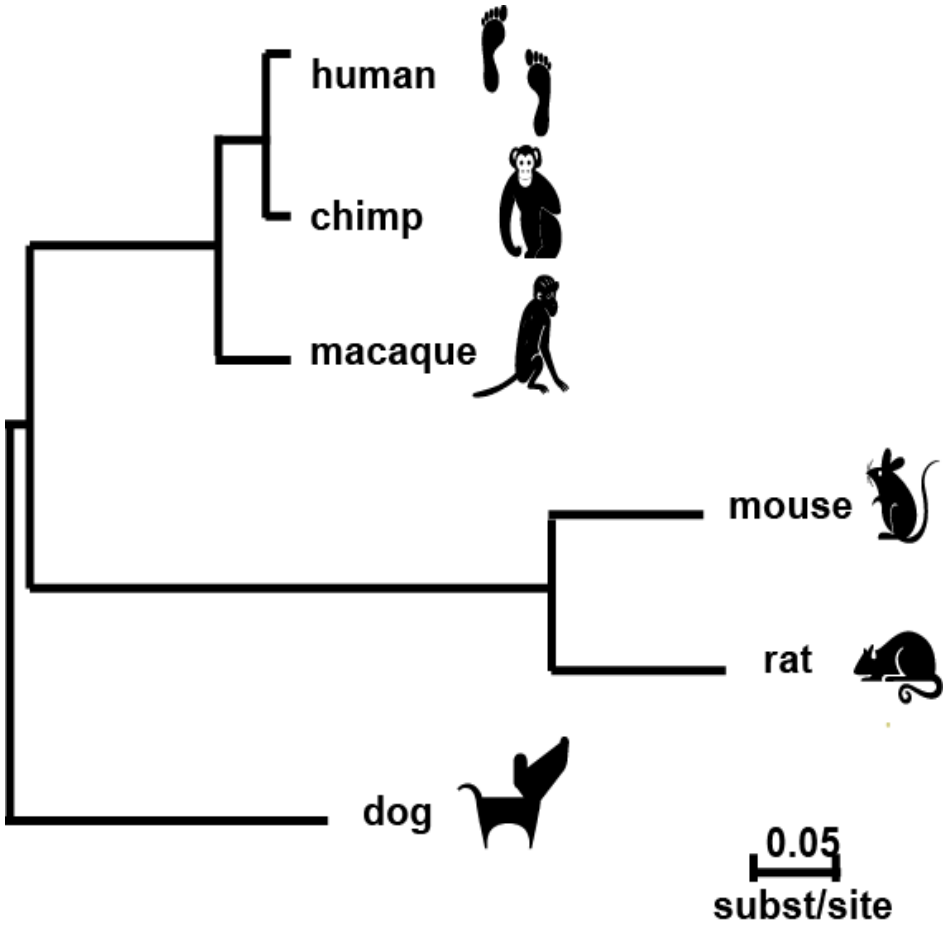
Chinwalla et al. 2002,
Nature 420





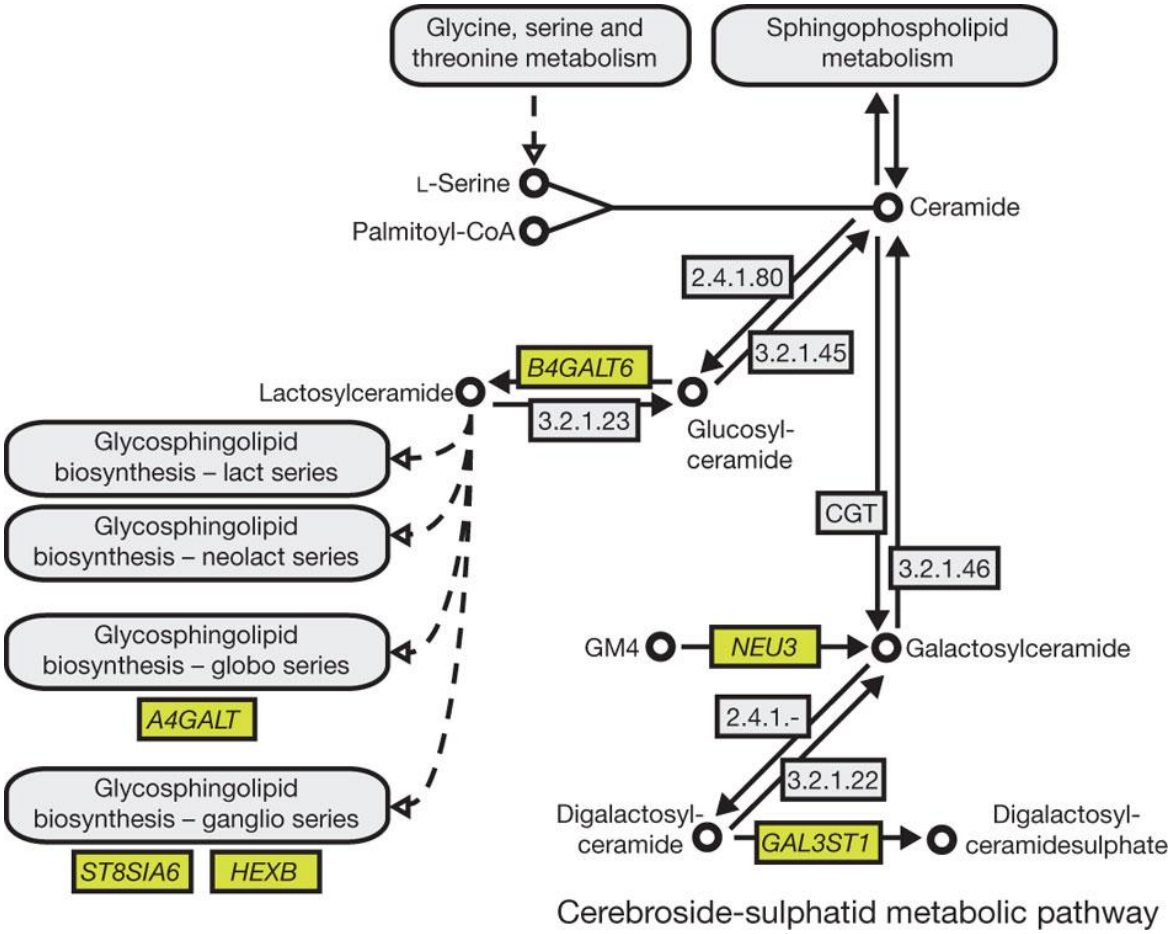
Dot plots comparing mouse and human chromosomes (80MYA)
Large scale synteny
85% (60-99%) identity of protein-coding regions

More mammalian genomes



Rhesus Macaque Consortium *et al*. 2007, *Science* 316

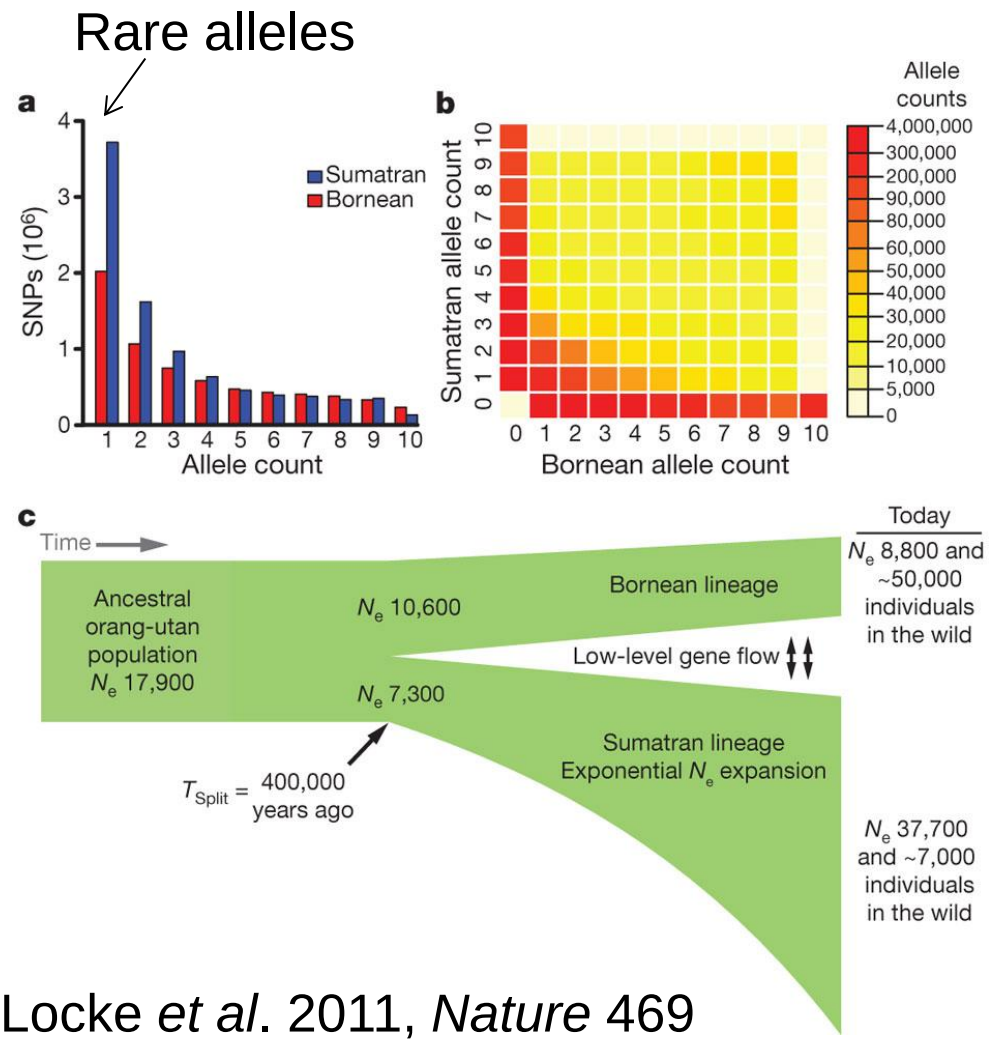
Orang-utan genome(s)



Locke *et al*. 2011, *Nature* 469

Evidence for positive selection in primates: ‘visual perception’ and ‘glycolipid metabolic processes’, important for the nervous system

Orang-utan genome(s)

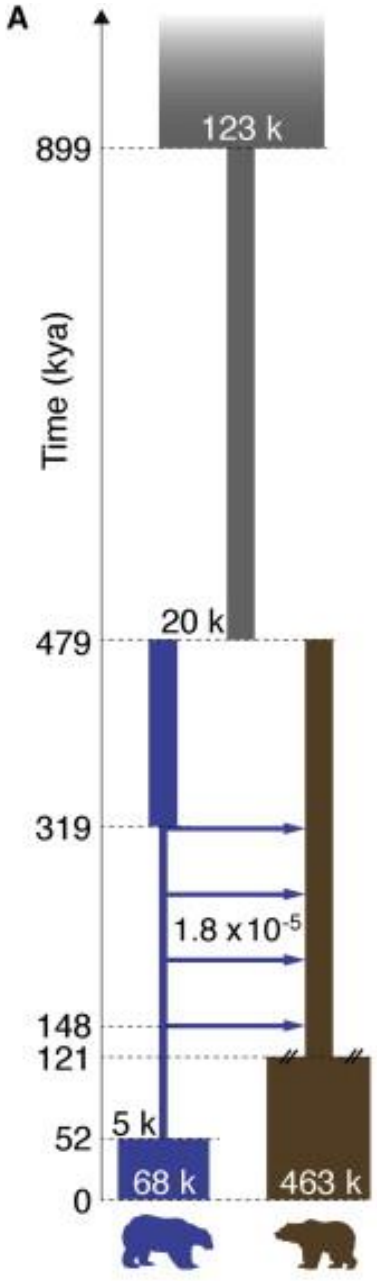


Locke *et al*. 2011, *Nature* 469

More recent split than thought

Census and N_e show opposing tendencies

Polar bear genome



Polar bear evolution:

- *De novo* assembly of a PB reference genome (101x)
- Re-sequenced 79 Greenlandic PB + 10 BrB (3.5x - 22x)
- Adaptation to a hyperlipid diet
- Significant signal for positive selection - 9 of 16 genes associated to reorganisation of cardiovascular system

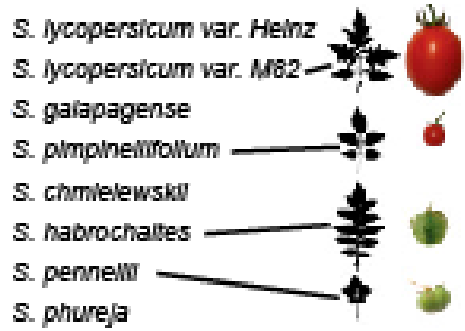
Liu *et al.* 2014 *Cell* 157

Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato

Daniel Koenig[a,b,1], José M. Jiménez-Gómez[a,c,1], Seisuke Kimura[a,d,2], Daniel Fulop[a,2], Daniel H. Chitwood[a], Lauren R. Headland[a], Ravi Kumar[a], Michael F. Covington[a], Upendra Kumar Devisetty[a], An V. Tat[a], Takayuki Tohge[e], Anthony Bolger[f], Korbinian Schneeberger[b,g], Stephan Ossowski[b,h], Christa Lanz[b], Guangyan Xiong[i], Mallorie Taylor-Teeples[a,j], Siobhan M. Brady[a,j], Markus Pauly[i], Detlef Weigel[b,3], Björn Usadel[f,k,l], Alisdair R. Fernie[e], Jie Peng[m], Neelima R. Sinha[a], and Julin N. Maloof[a,3]

www.pnas.org/cgi/doi/10.1073/pnas.1309606110 PNAS | Published online June 26, 2013 | E2655–E2662

- domesticated tomato
- S.gal. island colonization and adaptation
- S.chm. high altitude, drought tolerant
- S.hab. high altitude, chilling tolerant
- S.pen. desert adapted



PNAS

Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato

Daniel Koenig[a,b,1], José M. Jiménez-Gómez[a,c,1], Seisuke Kimura[a,d,2], Daniel Fulop[a,2], Daniel H. Chitwood[a], Lauren R. Headland[a], Ravi Kumar[a], Michael F. Covington[a], Upendra Kumar Devisetty[a], An V. Tat[a], Takayuki Tohge[e], Anthony Bolger[f], Korbinian Schneeberger[b,g], Stephan Ossowski[b,h], Christa Lanz[b], Guangyan Xiong[i], Mallorie Taylor-Teeples[a,j], Siobhan M. Brady[a,j], Markus Pauly[i], Detlef Weigel[b,3], Björn Usadel[f,k,l], Alisdair R. Fernie[e], Jie Peng[m], Neelima R. Sinha[a], and Julin N. Maloof[a,3]

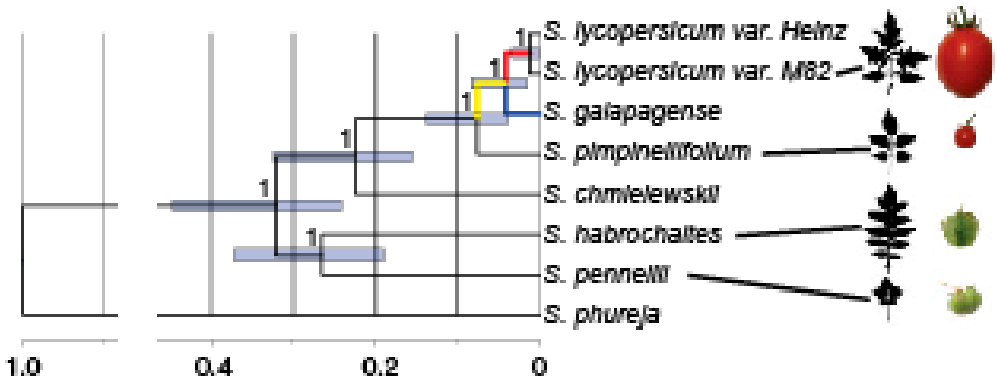
www.pnas.org/cgi/doi/10.1073/pnas.1309606110 PNAS | Published online June 26, 2013 | E2655–E2662

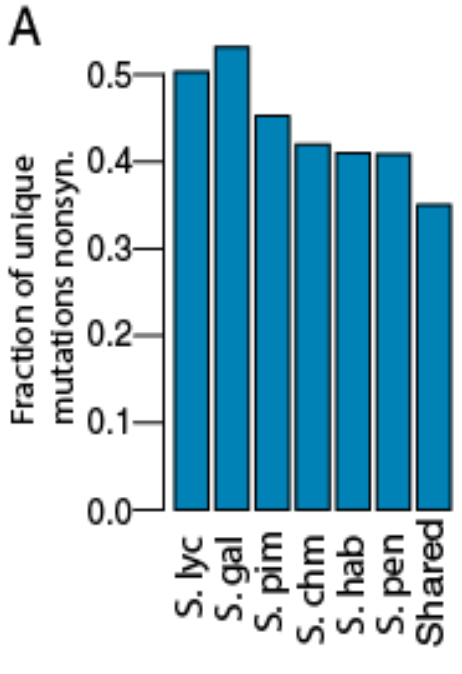
- Increased dN/dS after domestication in S. lyc., and during island colonization and adaptation in S. gal.

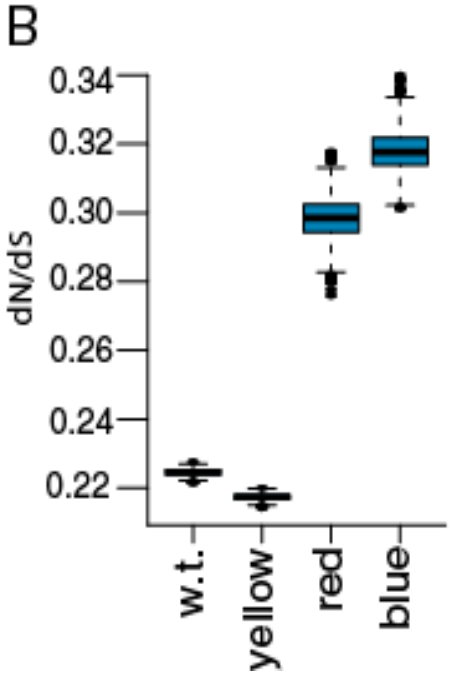
- Due to relaxed purifying selection and/or fixation of mutations during genetic bottleneck due to drift.

- Relaxed purifying selection elevates dN/dS by random substitution across the genome, but positive selection at specific loci.

- 51 genes under positive selection - pathogen response. Some response to abiotic factors, such as soil chemistry.







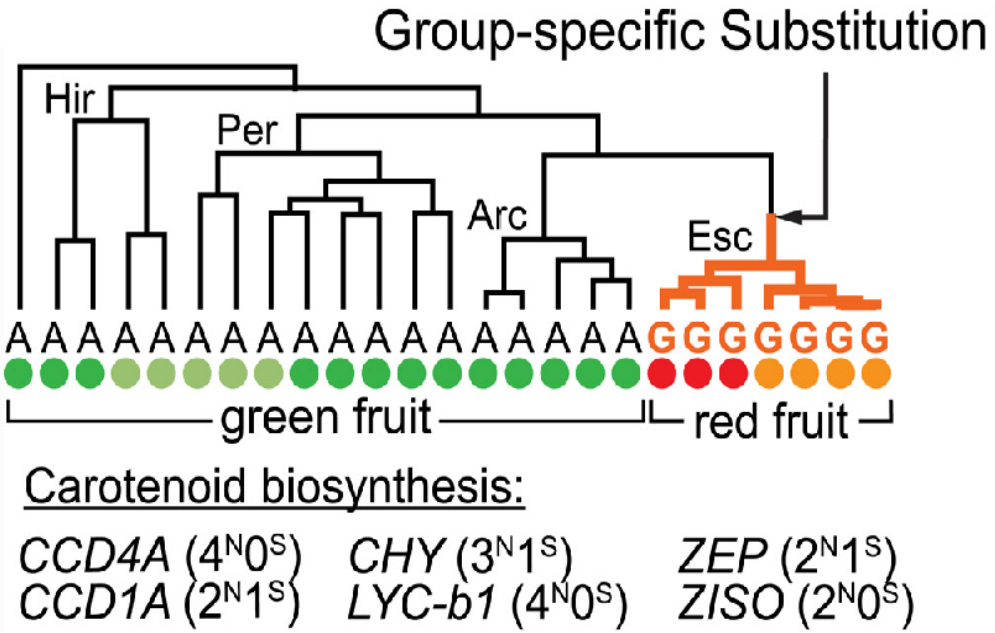
PLOS BIOLOGY

PLOS Biology | DOI:10.1371/journal.pbio.1002379 February 12, 2016

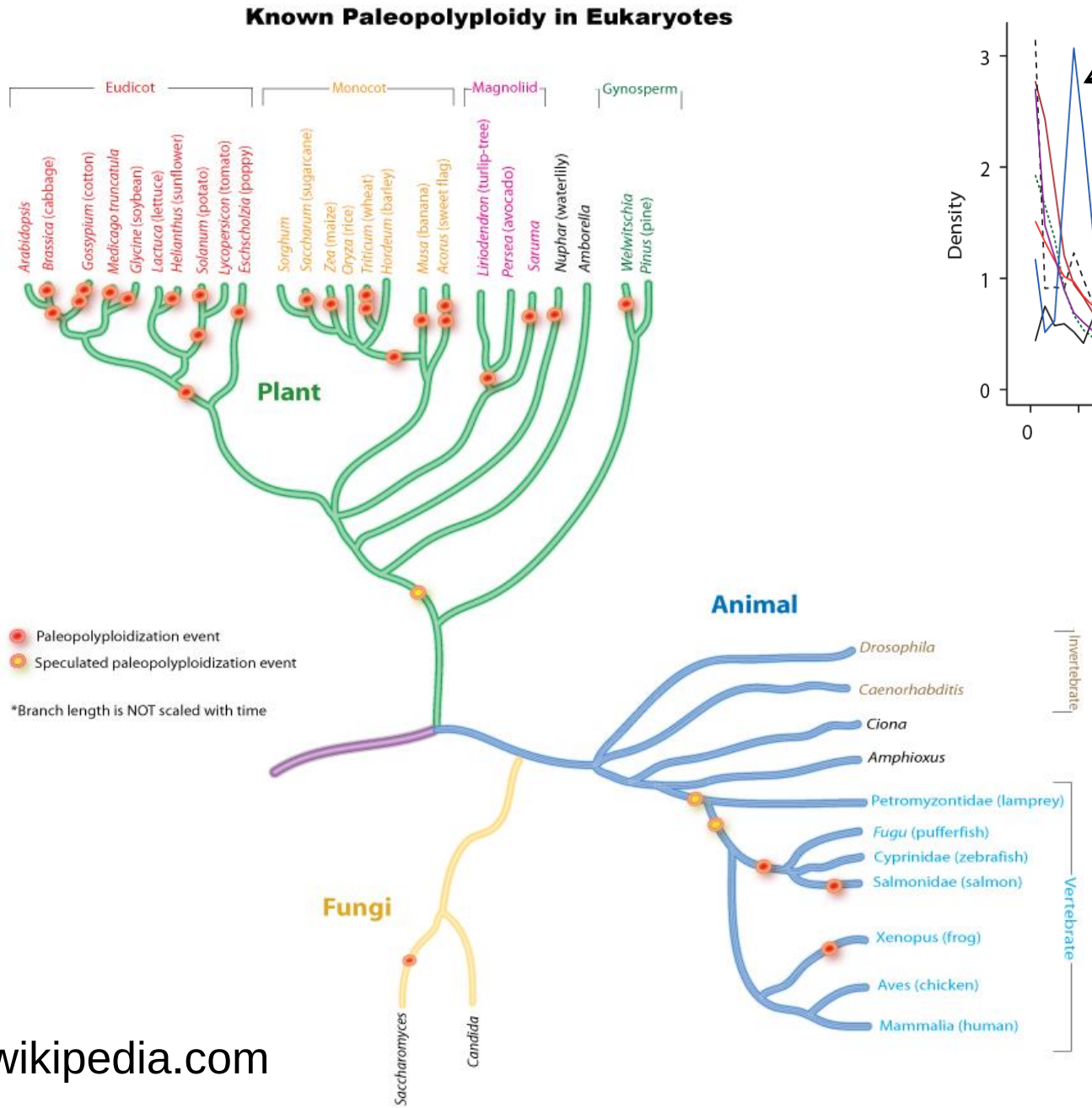
Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation

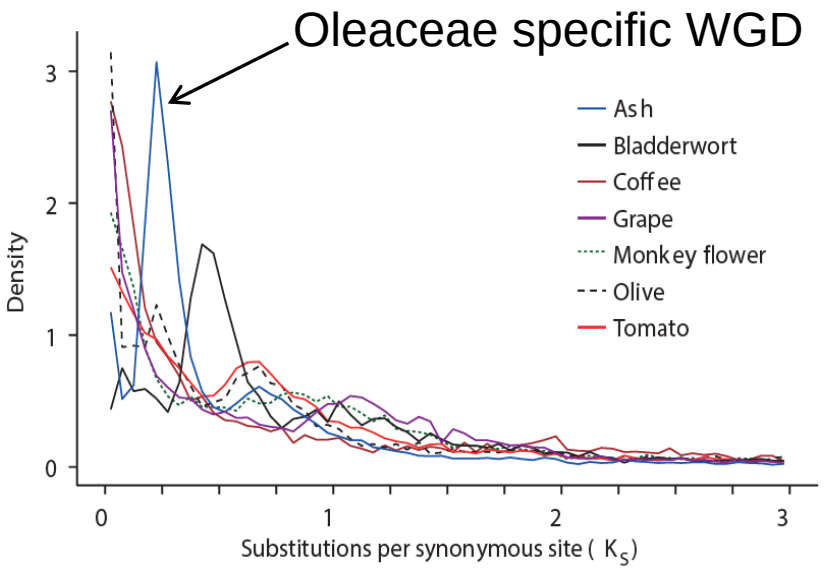
James B. Pease [1], David C. Haak [1,2], Matthew W. Hahn [1,3], Leonie C. Moyle [1*]

- Studied de novo evolution of lineage-specific traits
- 3.1% $d_N/d_S > 1$, $p < 0.01$ Esculentum; 4.7% in Arcanum and 4.0% in Hirsutum group
- Eg with functional consequences: 10 enzymes (33%) within the carotenoid biosynthesis are shared by red-fruited Esculentum
- Eg in *Ultraviolet Repair Defective 1* ortholog specific to Arcanum group, connected to adaptation to increased solar radiation at high altitude



Paleopolyploidy





Sollars *et al*. 2017, *Nature*

wikipedia.com

LETTER

2012|VOL488|NATURE

doi:10.1038/nature11241

The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants

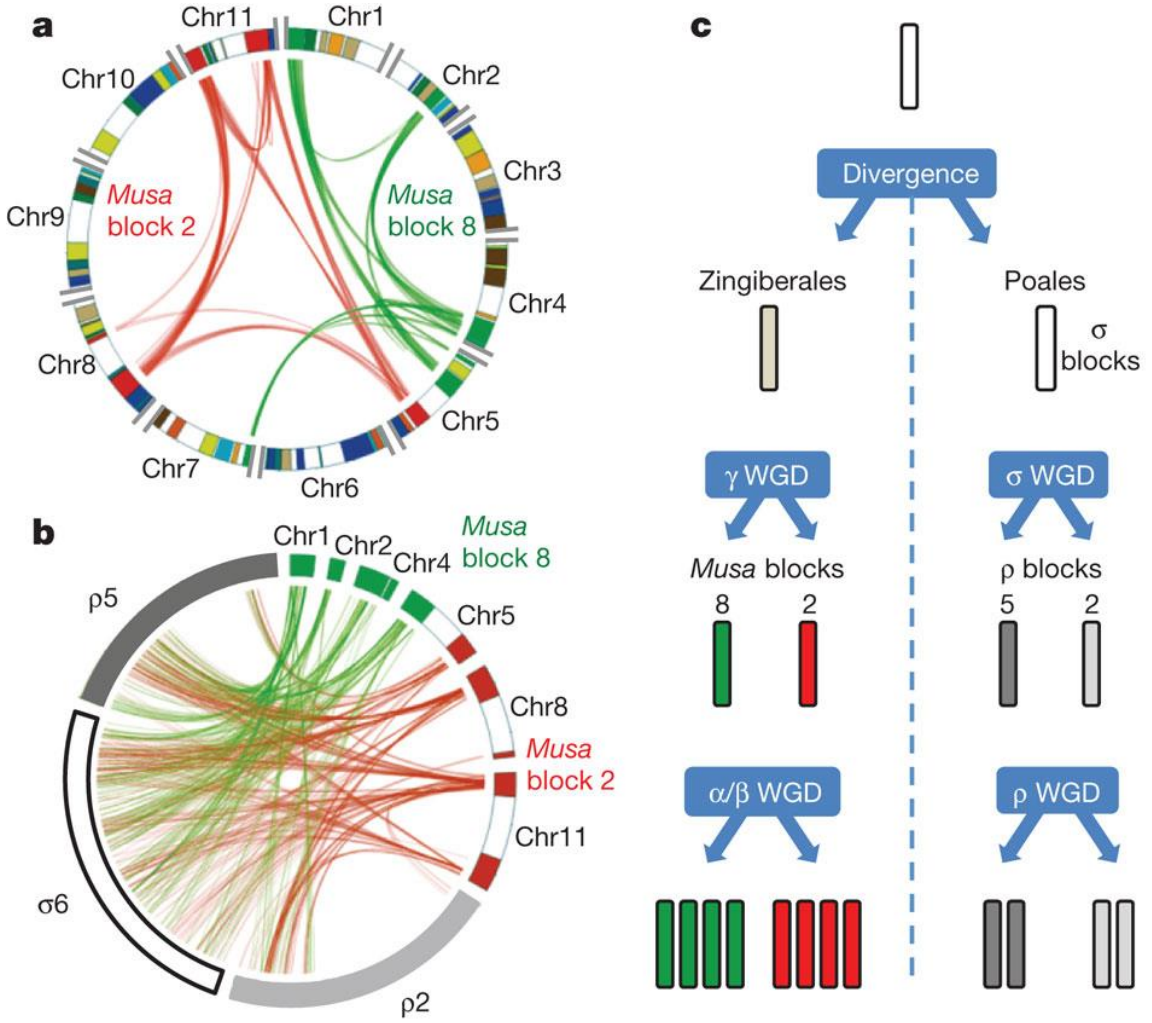
A D'Hont *et al.*

Signals of paleopolyploidy

a, Paralogs between chr. ancestral blocks 2 (red) and 8 (green).

b, Orthologs of *Musa* ancestral blocks 2 and 8 with rice ancestral blocks ρ2, ρ5 and σ6.

c, Representation of the deduced WGD event.



1000 genome projects

- 1000 Human (www.1000genomes.org)
- 1K *Drosophila* (www.dpgp.org/1K/
- 1001 *Arabidopsis* (www.1001genomes.org/)
- 10K Vertebrate Genome Porject (http://genome10k.soe.ucsc.edu/)

Two main reasons for genetic variation within a population or between species

Natural selection (survival of the fittest)

Mutation and drift (survival of the luckiest)

Gillespie, J.H. 1998. *Population genetics: a concise guide*. John Hopkins Univ. Press, Baltimore.

Hartl, D.L.& A.G. Clark. 1997. *Principles of population genetics*. Sinauer Associates, Sunderland, Massachusetts.

Mining polymorphism data

Disentangle the effect of evolutionary forces

- Mutational process
- Drift
- Population dynamics and structure
- Selection (which kind?)

Several possible approaches

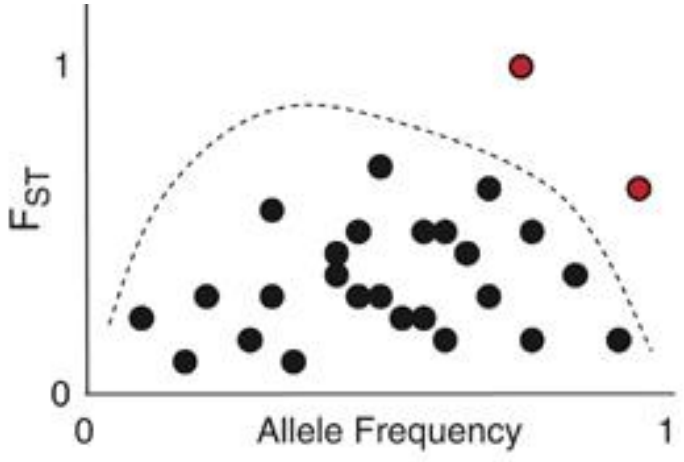
Candidate genes or loci

Blind approaches or genome scans

Unusual levels of polymorphism

Unusual patterns of polymorphism

Phylogenetic approach



Stinchcombe & Hoekstra 2008, *Heredity*

Positive and negative selection

Genotype	AA	Aa	aa
Frequency	p^2	$2p(1-p)$	$(1-p)^2$
Fitness	1	1+s	1+2s

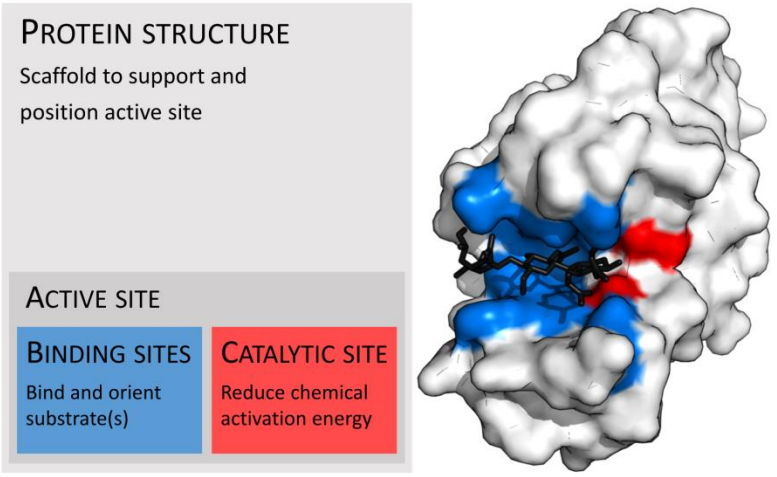
s is the selection coefficient

s ~ 0: **neutral evolution**

s < 0: **negative (purifying) selection**

s > 0: **positive selection (adaptive evolution)**

The rationale

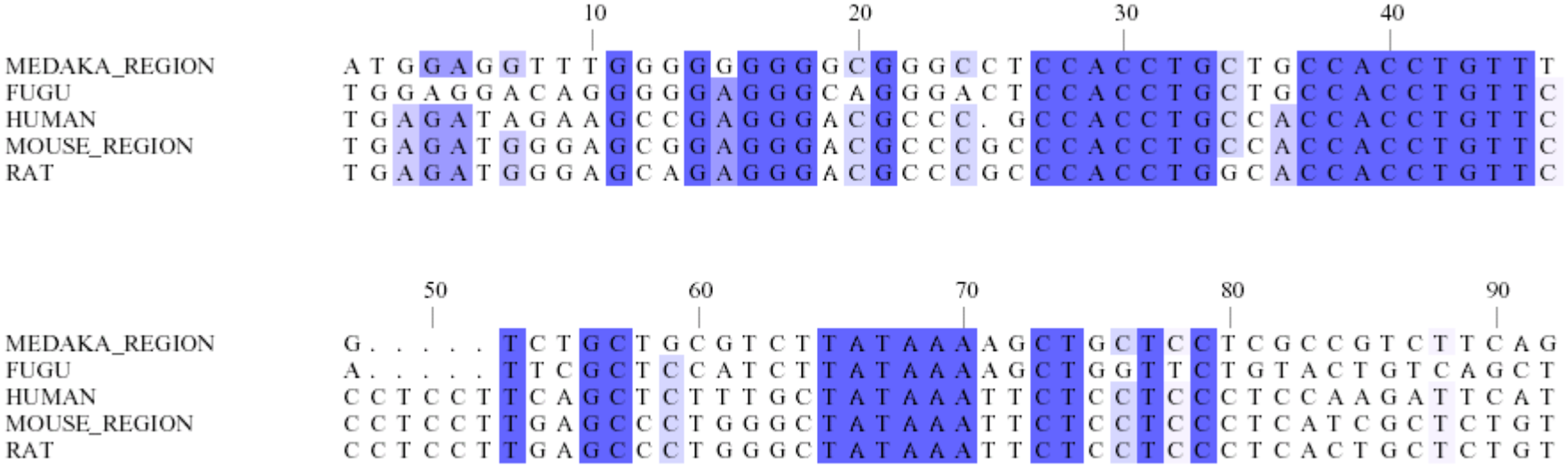


https://wikipedia.org

- Functionally important regions tend to be more conserved than non-functional regions
- Examples:
  - Exons are often more conserved than non-coding parts of the genome (genes)
  - The binding site composition is well conserved even between remote species
- Positive selection is more interesting than negative selection - **evolutionary innovations** and **species divergence**

The rationale

- Conserved regions are a good starting point to look for functionally important elements in the genome.



Neutral theory of molecular evolution (Kimura 1968)

the number of new mutations arising in a diploid population

$$2N\mu$$

the fixation probability of a new mutation by drift

$$1/2N$$

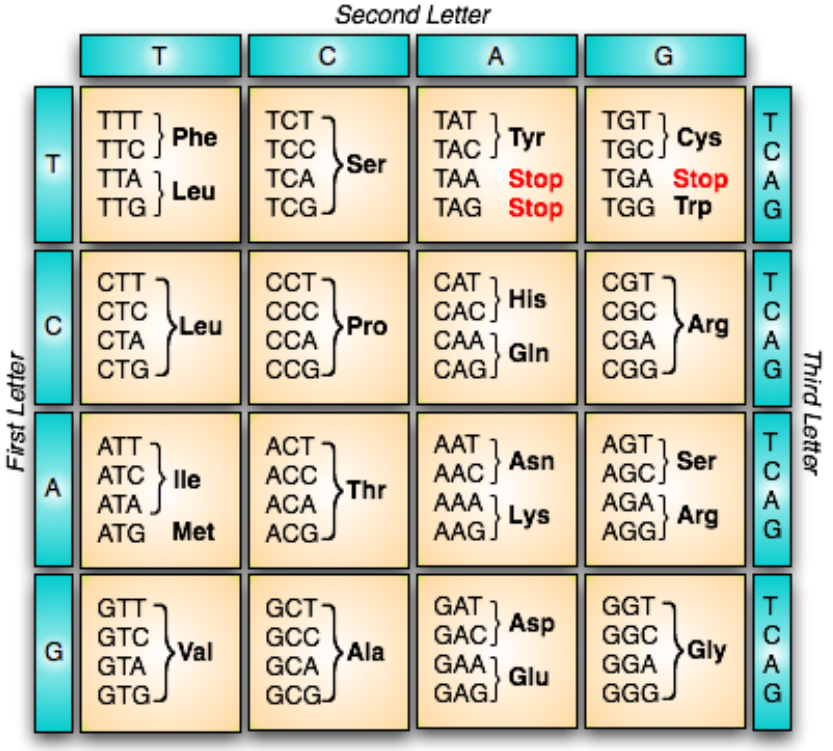
the substitution (fixation) rate

$$k = 2N\mu \times 1/2N$$

neutral theory:

$$k = \mu$$

The genetic code determines the impact of a mutation



Kimura (1968)

d_S: number of synonymous substitutions per synonymous site (K_S)

d_N: number of nonsynonymous substitutions per nonsynonymous site (K_A)

ω: the ratio d_N/d_S - measures selection at the protein level

An index of selection

rate ratio	mode	example
$d_N/d_S < 1$	purifying (negative) selection	house keeping genes
$d_N/d_S = 1$	neutral evolution	pseudogenes
$d_N/d_S > 1$	diversifying (positive) selection	genes of the immune system

Comparison between 2 protein-coding DNA sequences

Estimation of dN and dS between 2 sequences:

A. Counting methods

B. Codon substitution models

C. ML method

Counting method

Why use d_N and d_S? Why not use raw counts?

example:

gene of 300 codons from a pair of species
5 synonymous differences
5 nonsynonymous differences

5/5 = 1

Why **don't** we conclude that rates are equal (i.e., **neutral evolution**)?

Why do we use d_N and d_S?

Relative proportion of different types of mutations in hypothetical protein coding sequence.

Type	Expected number of changes			
	All 3 Positions	1st positions	2nd positions	3rd positions
Total mutations	300 (100%)	100	100	100
Synonymous	75 (25%)	4	0	69
Nonsyonymous	213 (71%)	91	96	27
nonsense	12 (4%)	5	4	4

Modified from Li and Graur (1991). Note that we assume a hypothetical model where all codons are used equally and that all types of point mutations are equally likely.

Why do we use d_N and d_S?

example, using d_N and d_S:

gene of 300 codons from a pair of species
5 synonymous differences
5 nonsynonymous differences

Synonymous sites = 25.5%
S = 300 x 3 x 25.5% = 229.5

Nonsynonymous sites = 74.5%
N = 300 x 3 x 74.5% = 670.5

d_S = 5/229.5 = 0.0218
d_N= 5/670.5 = 0.0075

d_N/d_S (ω) = 0.34 -> **purifying selection!!!**

Sequence evolution models

- Specify rates of replacement of each nucleotide
- These rates define the probabilities of all events that could happen at each instant, and the future depends on the present, but not on the past
- Hence the rates define the probabilities of all events over all times

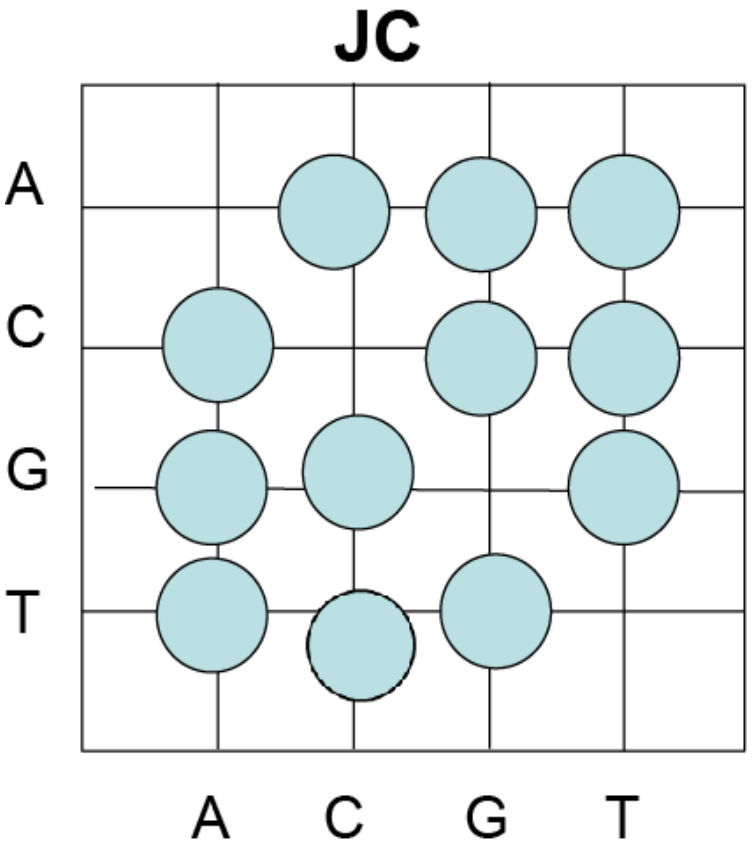
Can we model sequence evolution?

- Assuming each nucleotide evolves independent of other sites' evolution and of its past history (Jukes-Cantor 1969; Neyman 1971)

=> Model substitutions as Markov model

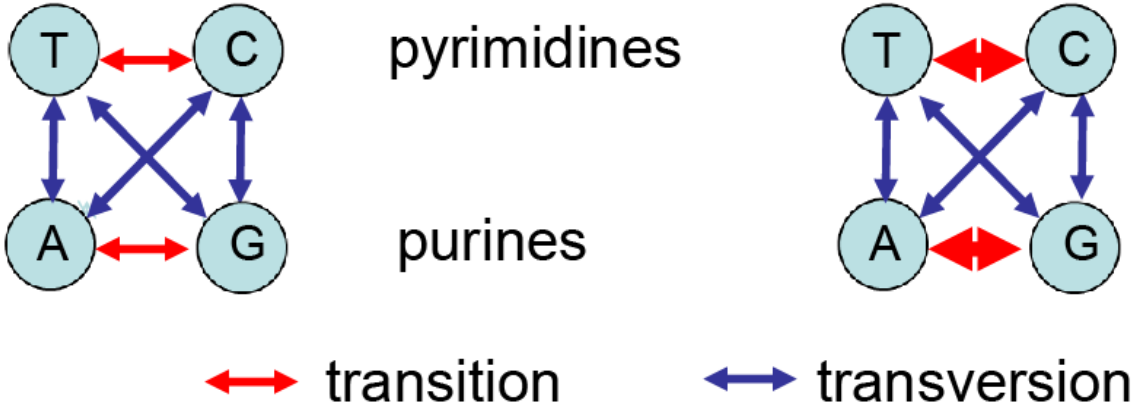
The Jukes-Cantor (JC) model

- ✓ All substitutions are equally likely.
- ✓ All nucleotides occur at the same frequency (25%).
- ✓ One parameter: the rate of substitution (α).



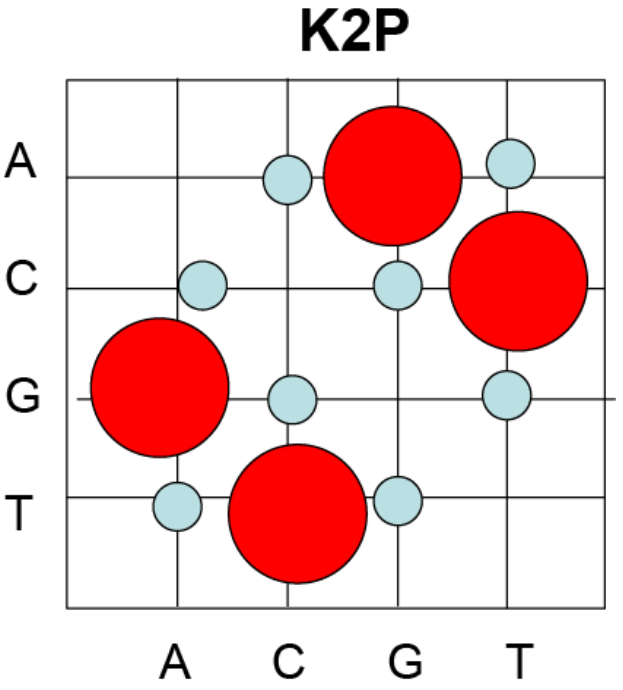
Real data have biases

e.g., Ts/Tv = 2.71 for *Drosophila* GstD1 gene

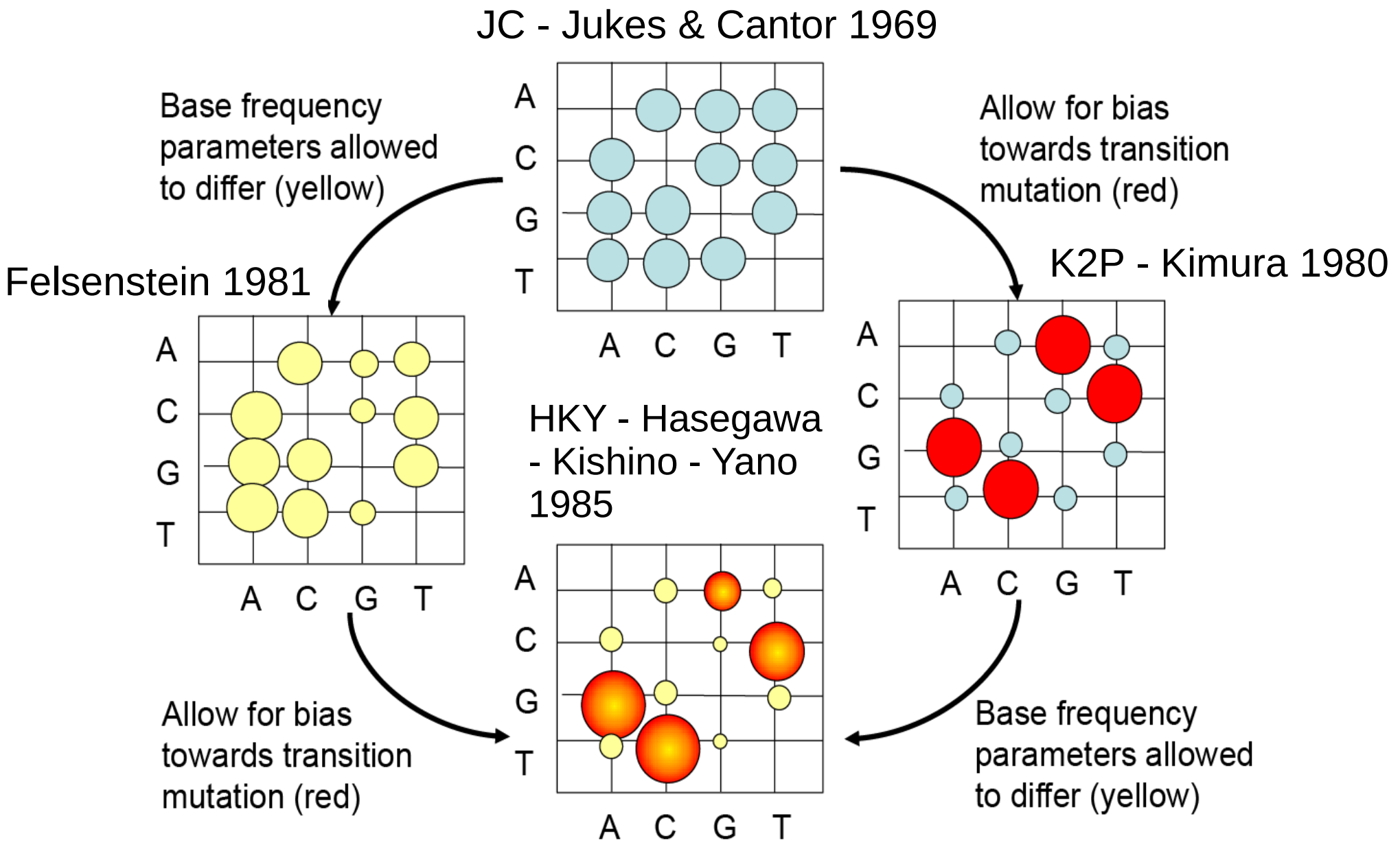


Kimura two parameter (K2P) model

- ✓ Transitions and transversions happen at different rates.
- ✓ All nucleotides occur at the same frequency.
- ✓ Two parameters: transition rate (α) and transversion rate (β).



Nested models



Whelan *et al*. 2001, *TREE* 17

Real data have biases

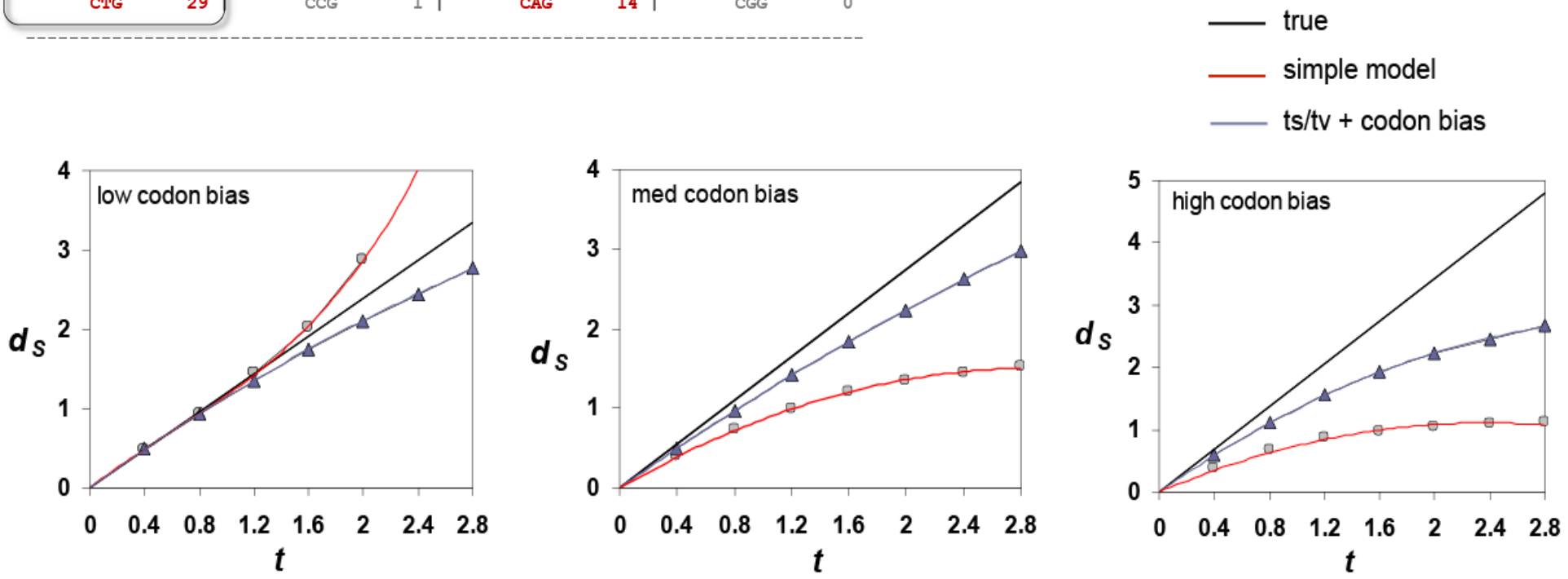
partial codon usage table for the *GstD* gene of *Drosophila*

Phe	F	TTT	0	Ser	S	TCT	0	Tyr	Y	TAT	1
		TTC	27			TCC	15			TAC	22
Leu	L	TTA	0			TCA	0	***	*	TAA	0
		TTG	1			TCG	1			TAG	0
Leu	L	CTT	2	Pro	P	CCT	1	His	H	CAT	0
		CTC	2			CCC	15			CAC	4
		CTA	0			CCA	3	Gln	Q	CAA	0
		CTG	29			CCG	1			CAG	14

Cys	C	TGT	0								
		TGC	6								
***	*	TGA	0								
Trp	W	TGG	8								
Arg	R	CGT	1								
		CGC	7								
		CGA	0								
		CGG	0								

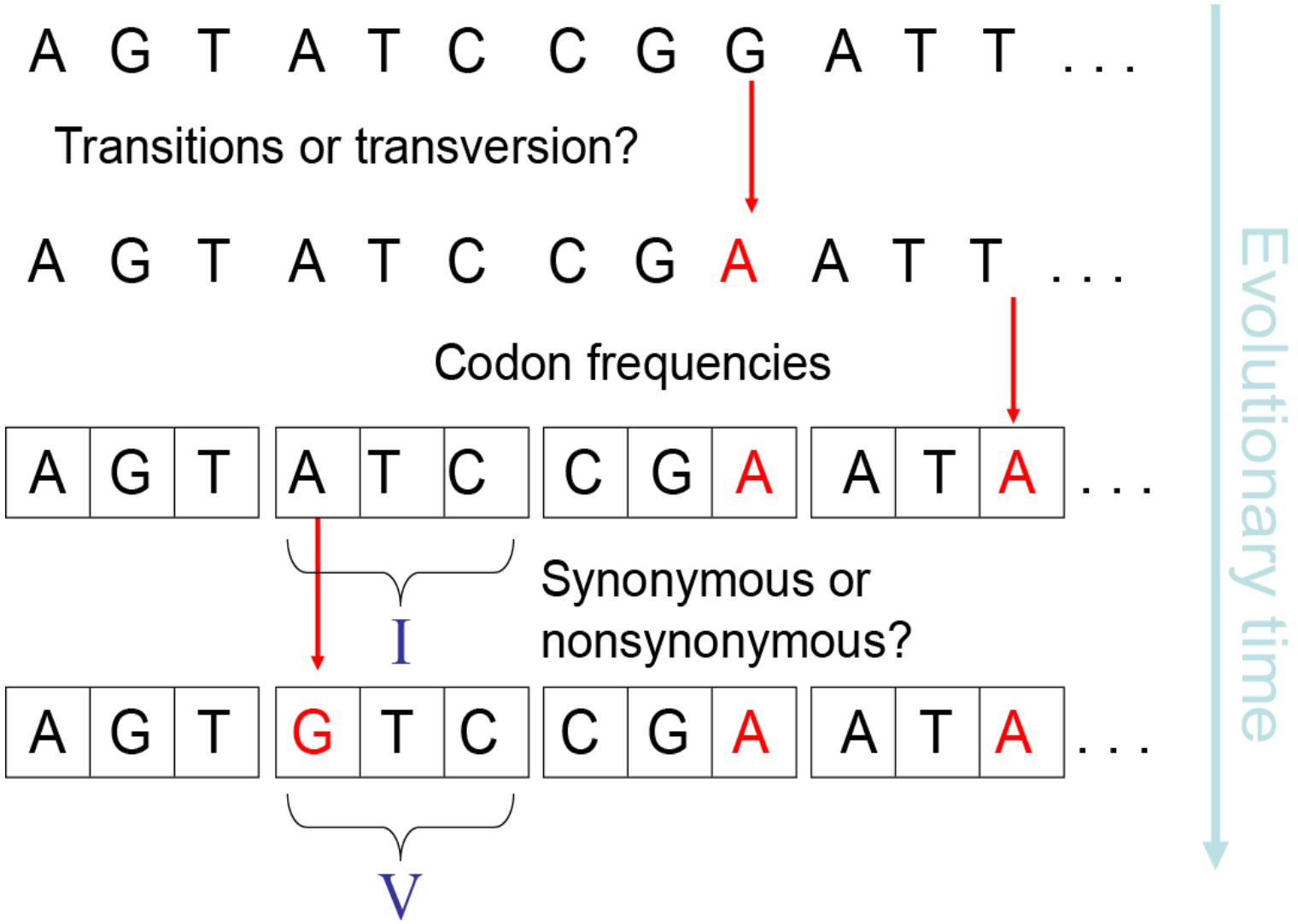
Preferred vs.
un-preferred codons

=> estimation bias



from Dunn *et al.* 2001, *Genetics* 157

Codon sequence evolution



Codon sequence evolution

dS and dN must be corrected for both the structure of genetic code and the underlying mutational process of the DNA

-this can differ among lineages and genes

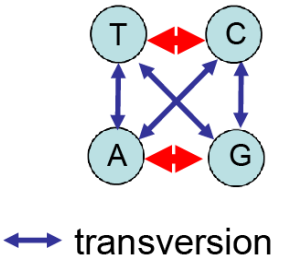
Correcting dS and dN for underlying mutational process of the DNA makes them sensitive to assumptions about the process of evolution

-evolution occurs at the population level (micro-evolution)

Markov chain model of codon substitution

Factors to consider:

- Transition/transversion rate ratio: κ
- Biased codon usage: πj for codon j
- Nonsynonymous/synonymous rate ratio: ω = dN/dS



- Synonymous

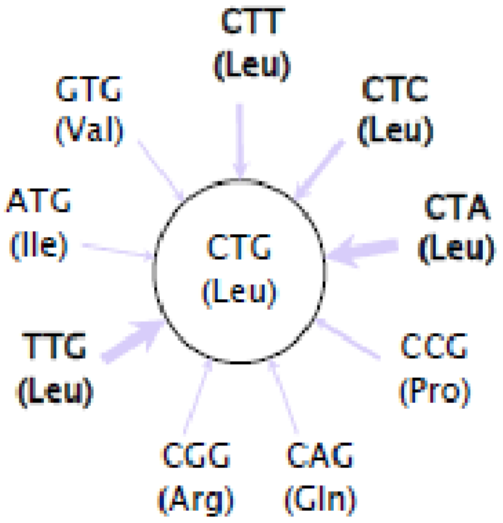
CTC (Leu) → CTG (Leu) πCTG

TTG (Leu) → CTG (Leu) κπCTG

- Nonsynonymous

GTG (Val) → CTG (Leu) ωπCTG

CCG (Pro) → CTG (Leu) κωπCTG

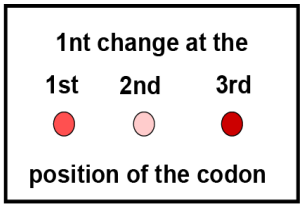


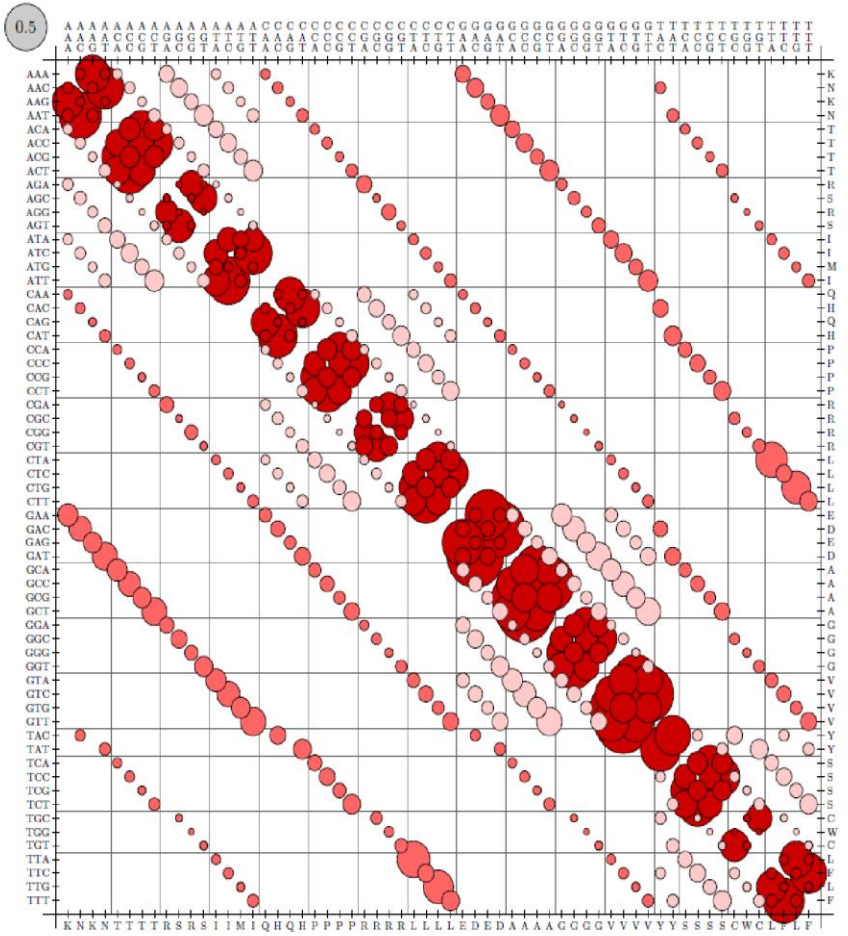
The (basic) codon model M0

$$Q_{ij} = \begin{cases} 0 & \text{if } i \to j \text{ is > 1 nucleotide substitution or j is a stop codon} \\ \pi_j & \text{if } i \to j \text{ synonymous transversion} \\ \pi_j K & \text{if } i \to j \text{ synonymous transition} \\ \pi_j \omega & \text{if } i \to j \text{ nonsynonymous transversion} \\ \pi_j K \omega & \text{if } i \to j \text{ nonsynonymous transition} \end{cases}$$

where
K : transition/transversion rate ratio
π_j : equilibrium frequency of codon *j*
ω : nonsynonymous/synonymous rate ratio

Parameter estimation via ML





(Goldman & Yang 1994 *Mol Biol Evol* 11
Muse & Gaut 1994 *Mol Biol Evol* 11)

The instantaneous rate matrix, Q is very big 61x61

Just a few parameters are needed to cover the 3721 transitions between codons!

Intentional simplification: all amino acid substitutions have the same ω

	to codon below:							
From codon below:	TTT (Phe)	TTC (Phe)	TTA (Leu)	TTG (Leu)	CTT (Leu)	CTC (Leu)	⋯→	GGG (Gly)
TTT (Phe)	—	$\kappa\pi_{TTC}$	$\omega\pi_{TTA}$	$\omega\pi_{TTG}$	$\omega\kappa\pi_{TTT}$	0	⋯→	0
TTC (Phe)	$\kappa\pi_{TTT}$	—	$\omega\pi_{TTA}$	$\omega\pi_{TTG}$	0	$\omega\kappa\pi_{CTC}$	⋯→	0
TTA (Leu)	$\omega\pi_{TTT}$	$\omega\pi_{TTC}$	—		0	0	⋯→	0
TTG (Leu)	$\omega\pi_{TTT}$	$\omega\pi_{TTC}$	$\kappa\pi_{TTA}$	—	0	0	⋯→	0
CTT (Leu)	$\omega\kappa\pi_{TTT}$	0	0	0	—	$\kappa\pi_{CTC}$	⋯→	0
CTC (Leu)	0	$\omega\kappa\pi_{TTC}$	0	0	$\kappa\pi_{TTT}$	—	⋯→	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	↘	
GGG (Gly)	0	0	0	0	0	0	0	—

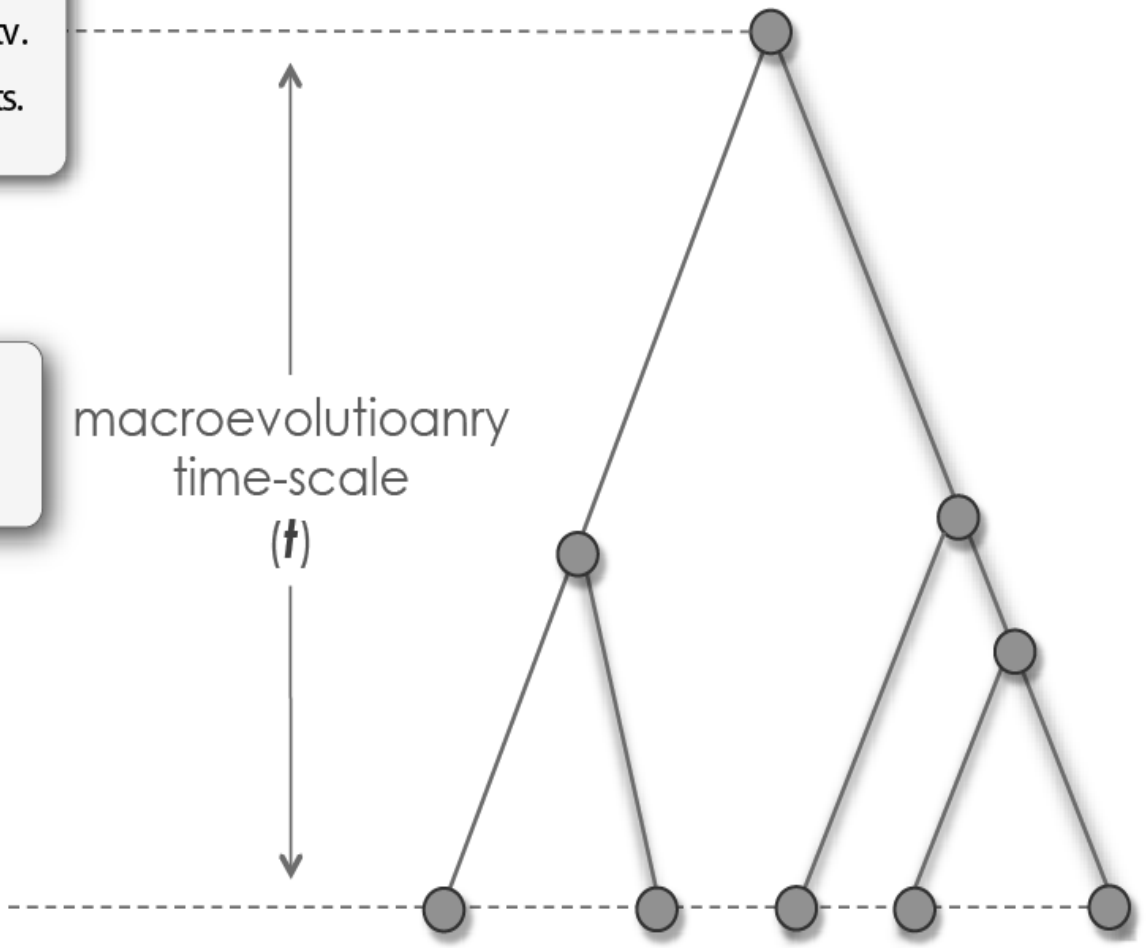
* This is equivalent to the codon model of Goldman and Yang (1994). Parameter ω is the ratio d_N/d_S, κ is the transition/transversion rate ratio, and π_i is the equilibrium frequency of the target codon (i).

Probability of substitution between codons over time, P(t)

$$Q_{ij} = \begin{cases} 0 & \text{if i and j differ by} > 1 \\ \pi_j & \text{for synonymous tv.} \\ \kappa\pi_j & \text{for synonymous ts.} \\ \omega\pi_j & \text{for non-synonymous tv.} \\ \omega\kappa\pi_j & \text{for non-synonymous ts.} \end{cases}$$

$$P(t) = \{p_{ij}(t)\} = e^{Qt}$$

exponentiate the **Q matrix** to obtain **substitution probabilities**



ML estimation of dS and dN

Numbers of substitutions are calculated from Qij and t

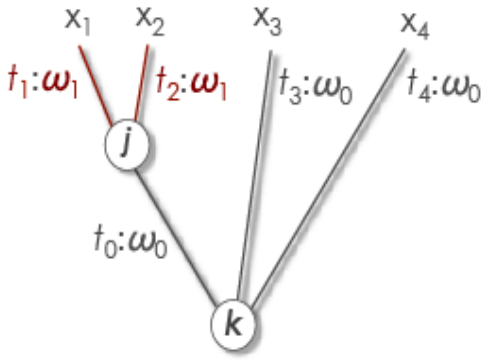
Number of sites (S and N, i.e., mutational opportunities) are calculated from Qij by fixing $\omega=1$

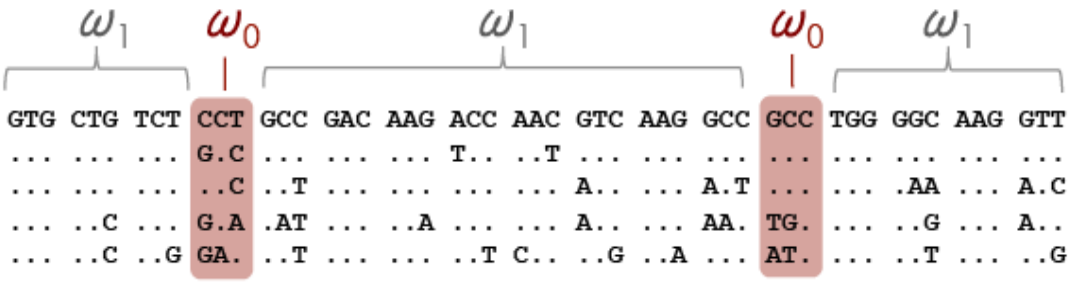
Potential problems:

- wrong sequence divergence, often too high (unreliable), sometimes too low (large sampling error)

-data quality control, alignment

Models for variation among branches and sites





branch models
(*ω* varies among branches)

site models
(*ω* varies among sites)

branch-site models
(combines the features of above models)

Model based inference

3 analytical tasks:

1) parameter estimation (e.g., ω)

2) hypothesis testing

3) make predictions (e.g., sites having ω > 1)

Parameter estimation

$$Q_{ij} = \begin{cases} 0 & \text{if } i \to j \text{ is > 1 nucleotide substitution or j is a stop codon} \\ \pi_j & \text{if } i \to j \text{ synonymous transversion} \\ \pi_j K & \text{if } i \to j \text{ synonymous transition} \\ \pi_j \omega & \text{if } i \to j \text{ nonsynonymous transversion} \\ \pi_j K \omega & \text{if } i \to j \text{ nonsynonymous transition} \end{cases}$$

where
K : transition/transversion rate ratio
π_j : equilibrium frequency of codon *j*
ω : nonsynonymous/synonymous rate ratio

t, K, ω - unknown constants estimated by ML

π - empirical (e.g., F3x4 codon frequency model)

Use a numerical hill-climbing algorithm to maximize the likelihood function

Likelihood ratio test for positive selection

The simpler (null) model **H0** has **q** parameters with log likelihood **L0**: variable selective pressure but no positive selection (**M1**)

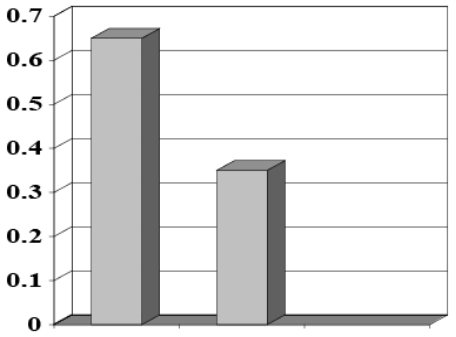
The more general (alternative) model **H1** has **p** parameters with log likelihood **L1**: variable selective pressure with positive selection (**M2**)

Compare twice the log likelihood difference

$2\Delta L = 2(L1 - L0)$

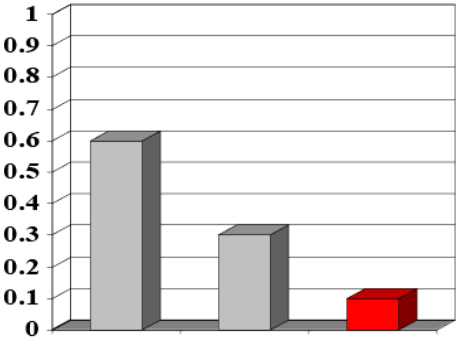
with a $\chi 2$ distribution with d.f. = p – q to test whether the simpler model is rejected

Model 1a



$\hat{\omega}$ = 0.5 (ω = 1)

Model 2a



$\hat{\omega}$ = 0.5 (ω = 1) $\hat{\omega}$ = 3.25

Weaknesses of codon-based methods

- Do not work for noncoding DNA
- Model assumptions may be unrealistic (but some assumptions matter more than others).
- The method detects positive selection only if it generates excessive nonsynonymous substitutions. It may lack power in detecting one-off directional selection or when the sequences are highly similar or highly divergent. It has little power with population data.
- Sensitive to sequence and alignment errors (Fletcher & Yang 2010 *Mol Biol Evol* 27; Privman et al. 2011 *Mol Biol Evol* 29; Jordan & Goldman 2012 *Mol Biol Evol* 29)

PAML (Phylogenetic Analysis by ML)

A program package by Ziheng Yang

Features include:

- estimating synonymous and nonsynonymous rates
- testing hypotheses concerning d_N/d_S rate ratios
- various amino acid-based likelihood analysis
- ancestral sequence reconstruction (DNA, codon, or AAs)
- various clock models
- simulating nucleotide, codon, or AA sequence data sets
- and more ……

PAML (Phylogenetic Analysis by ML)

Download PAML from:

http://abacus.gene.ucl.ac.uk/software/paml.html

For windows, Macs, and Unix/Linux

baseml	for bases
basemlg	continuous-gamma for bases
codeml	aaml (for amino acids) & codonml (for codons)
evolver	simulation, tree distances
yn00	dN and dS by YN00
chi2	chi square table
pamp	parsimony (Yang and Kumar 1996)
mcmctree	Bayes MCMC tree (Yang & Rannala 1997). Slow

Orchid diversity


Phalaenopsis ©Sagaflor


Caladenia ©N. Hoffman


Cattleya ©S. Wilson


Orchis italica ©T. Hughes


Ophrys apifera ©H. Bernd


Vanda ©D. Kulaga

D. fuchsii and *D. incarnata*

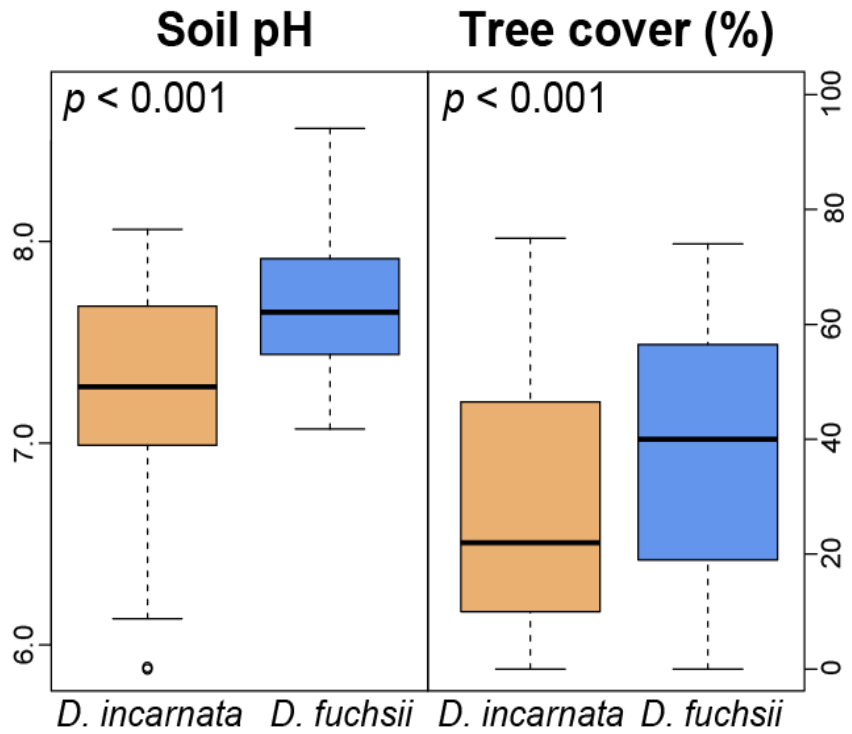
D. fuchsii and *D. incarnata*

Species	Soil pH				Soil moisture		Shade tolerance		Distribution
	5	6	7	8	Low	High	Low	Moderate	
D. incarnata s.l.	■	■	■	■	■	■	■		N + C Europe, W Asia
D. fuchsii		■	■	■	■	■	■	■	N + C Europe, W Asia

Based on field observations of RB in Britain, MH in Scandinavia and OP in the Alps and Pyrenees.

Paun *et al*. 2011 *BMC Evol Biol*

Dactylorhiza: microhabitat divergence



Francisco Balao

Landsat Tree Cover (~30 m resolution) at 691 localities

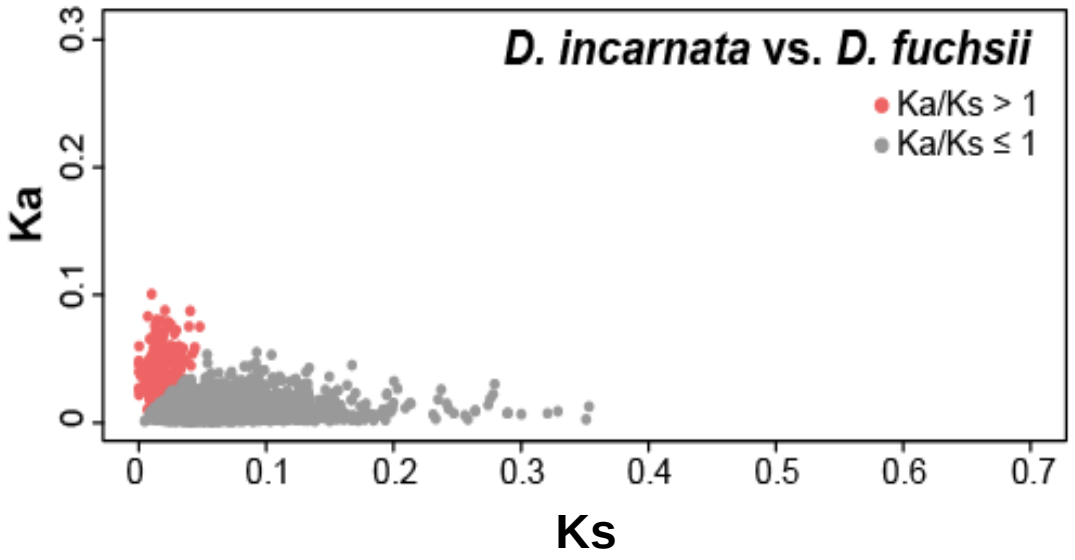
Soil pH measured at 14 European localities

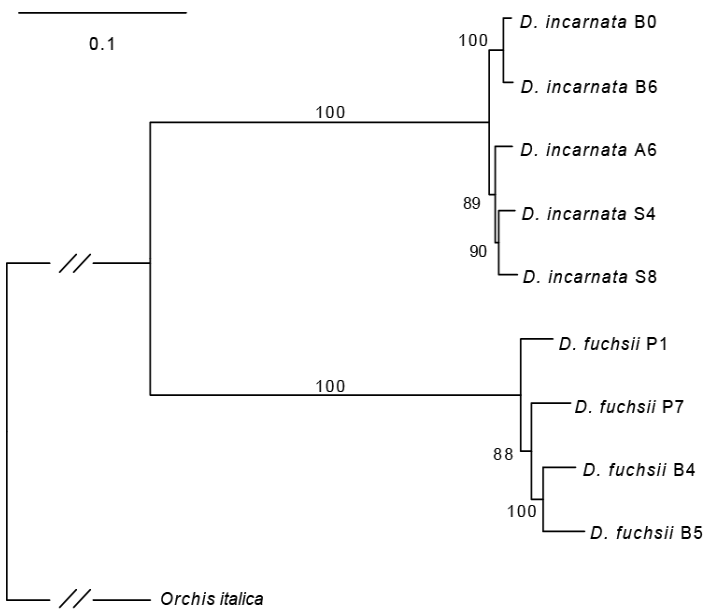
Parapatric species, with similar macroenvironmental preference, but distinct microhabitat optima

Balao *et al*. 2017 *Molec Ecol*

Coding sequence variation

- Mapping to *Orchis italica* transcriptome (De Paolo et al. 2014)
- Calling/filtering SNPs with GATK
- 23,185 indels and 727,350 SNPs
- 61 - 67% transcripts under purifying selection (i.e., Ka/Ks < 1, FDR < 0.1)

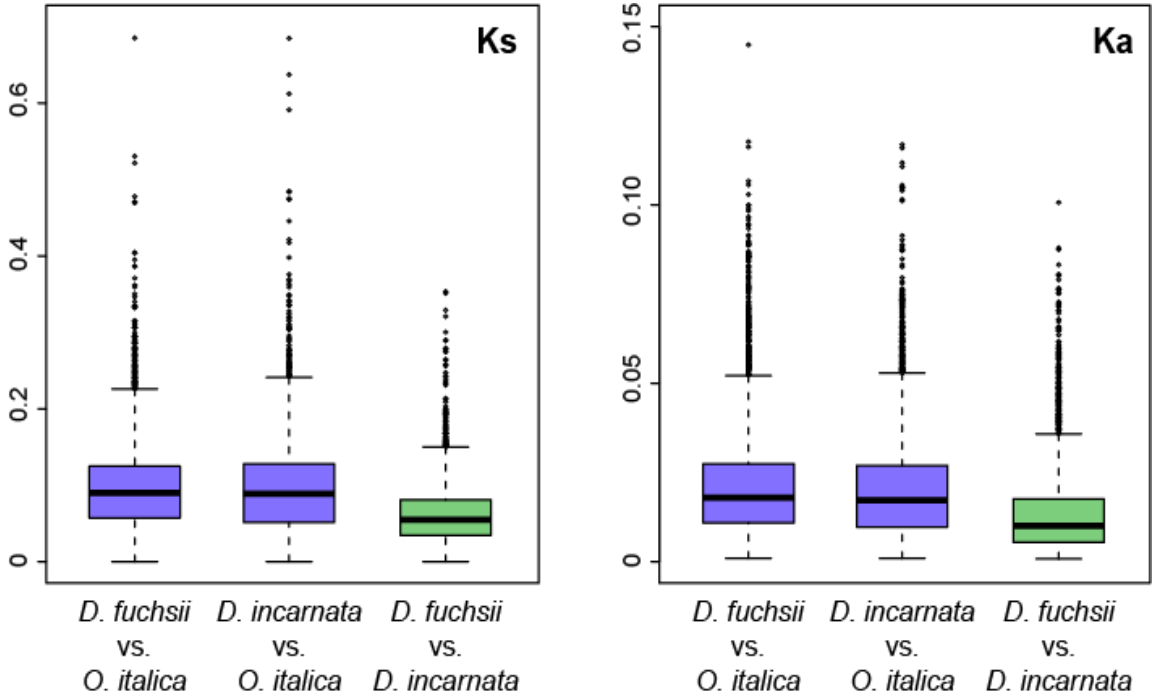




Balao *et al*. 2017 *Molec Ecol*

Adaptive coding sequence evolution

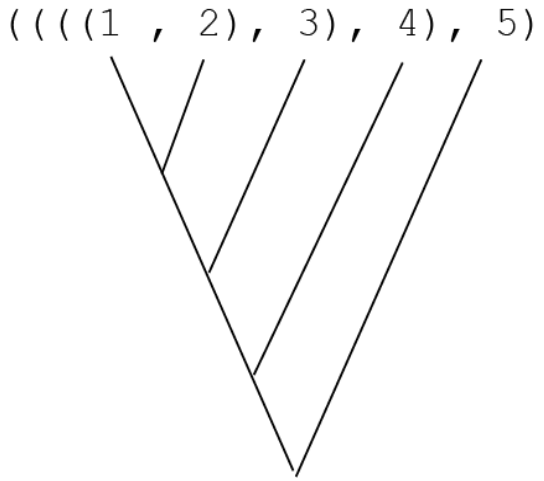
- *Df* and *Di* diverged 10.4 MYA (Ks = 0.06)
- Ka:Ks 1:5 - in line with
  A. thaliana and *A. lyrata* (5-10 MYA; Yang & Gaut 2011)
  Gossypium arboreum and *G. raimondii* (7-11 MYA; Senchina et al. 2003).



Balao *et al*. 2017 *Molec Ecol*

Running PAML

1. Sequence data file - plain text in PHYLIP format
2. Tree file as parenthetical notation
3. Control file (*.ctl)



Running PAML: the *.ctl file

```
  seqfile = seqfile.txt   * sequence data filename
 treefile = tree.txt      * tree structure file name
  outfile = results.txt   * main result file name

    noisy = 9      * 0,1,2,3,9: how much rubbish on the screen
  verbose = 1      * 1:detailed output
  runmode = 0      * 0:user defined tree

  seqtype = 1      * 1:codons
CodonFreq = 2      * 0:equal, 1:F1X4, 2:F3X4, 3:F61

    model = 0      * 0:one omega ratio for all branches

  NSsites = 0      * 0:one omega ratio (M0 in Tables 2 and 4)
                   * 1:neutral (M1 in Tables 2 and 4)
                   * 2:selection (M2 in Tables 2 and 4)
                   * 3:discrete (M3 in Tables 2 and 4)
                   * 7:beta (M7 in Tables 2 and 4)
                   * 8:beta&w; (M8 in Tables 2 and 4)

    icode = 0      * 0:universal code

fix_kappa = 0      * 1:kappa fixed, 0:kappa to be estimated
    kappa = 2      * initial or fixed kappa

fix_omega = 0      * 1:omega fixed, 0:omega to be estimated
    omega = 5      * initial omega
```

Excercises

1. ML estimation of the pairwise d_N/d_S (ω) ratio “by hand”. Use codeml to evaluate the likelihood for a variety of fixed pairwise ω values

2. Check your findings from exercise 1 by running codeml’s hill-climbining algorithm (still pairwise)

3. ML estimating of the branch-specific ω for *D. fuchsii*, for *D. incarnata* and for the branch of their most recent common recent ancestor

4. If time allows blastx the sequence, and use Ensembl to find which of the paralogues of this gene is due to the most recent duplication event (try using Oryza).

Adaptive coding sequence evolution

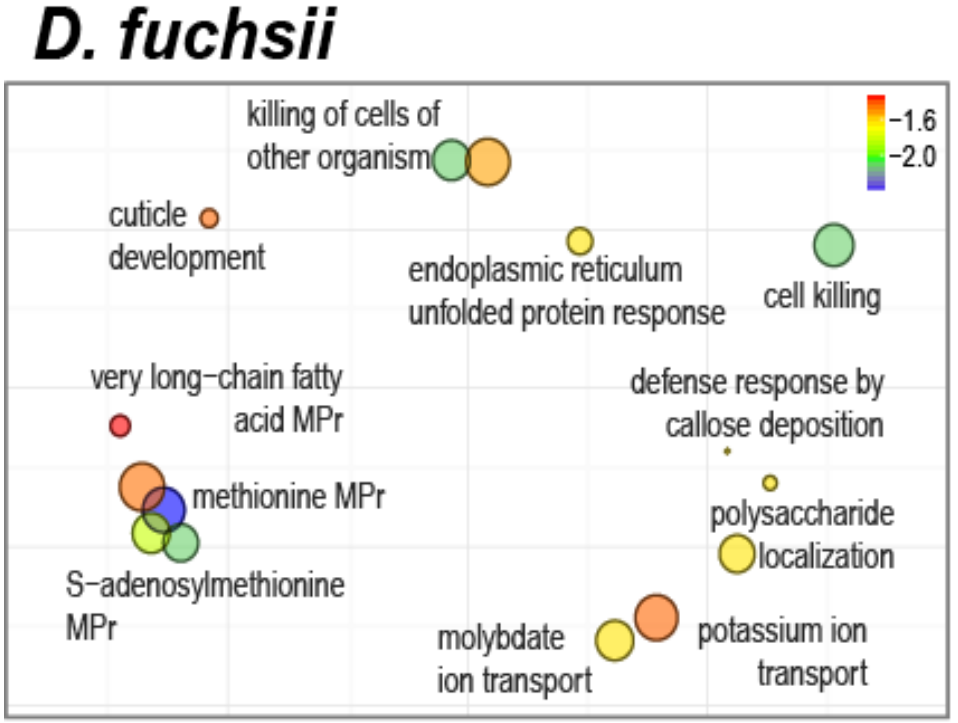
- ω >> 1 : 18 transcripts in *Df* and 14 transcripts in *Di*
- genes related to responses to biotic responses, including physical and chemical adaptations:

  Df: DEFENSIN J1-2 inhibits growth of pathogenic fungi

  HAI1 PHOSPHATASE - defence response by deposition of callose

  TETRAKITIDE α-PYRONE REDUCTASE 2 - flavonoid biosynthesis

  3-KETOACYL-SYNTHASE 10 - role in developing cuticular wax



Adaptive coding sequence evolution

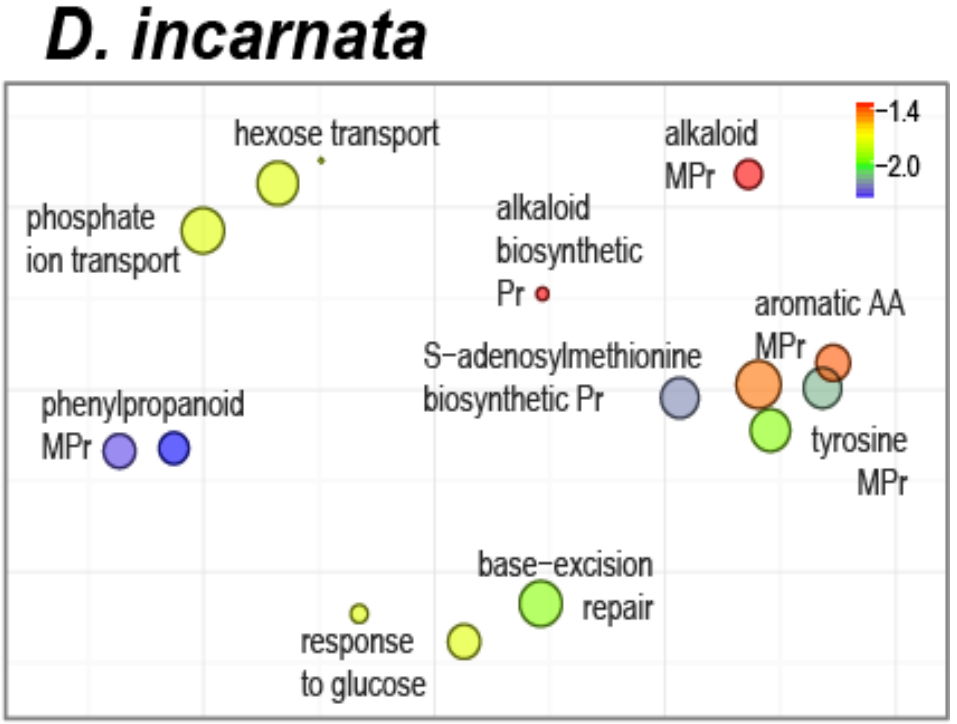
- ω >> 1 : 18 transcripts in *Df* and 14 transcripts in *Di*
- genes related to responses to biotic responses, including physical and chemical adaptations:

  Di: POLYPHENOL OXIDASE producing melanins protecting wounds

  PRIMARY AMINE OXIDASE wound-healing and cell-wall reinforcement

  YELLOW-LEAF SPECIFIC GENE 9 - viral defence response protein

  LACCASE - role in formation of lignin and alkaloid biosynthesis



Adaptive coding sequence evolution

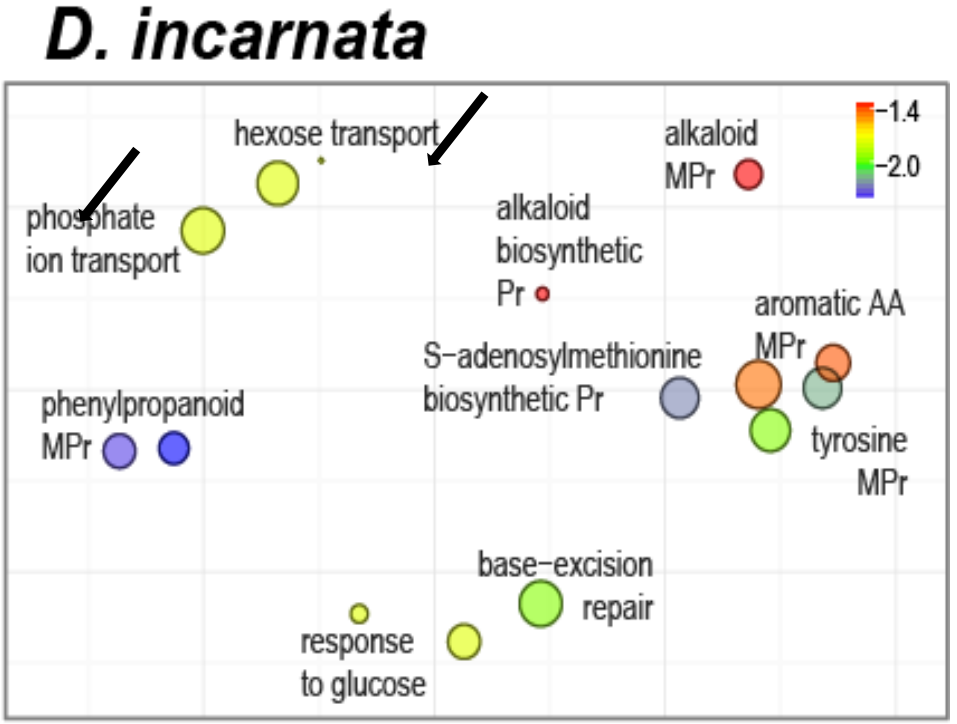
Di:

SUGAR CARRIER C - hexose transmembrane transport likely linked to hypoxia

PYRUVATE DECARBOXYLASE 1 - tolerance to root submergence

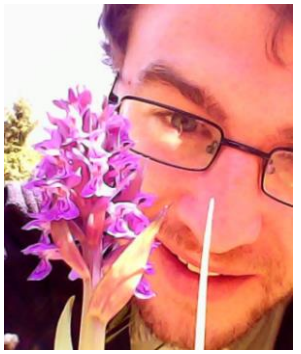
Df and *Di:*

Ion transporters



Highly divergent expression

- extensive DE (>30% of genes)
- DE related to abiotic adaptation or acclimation to common garden



Thomas Wolfe

Francisco Balao