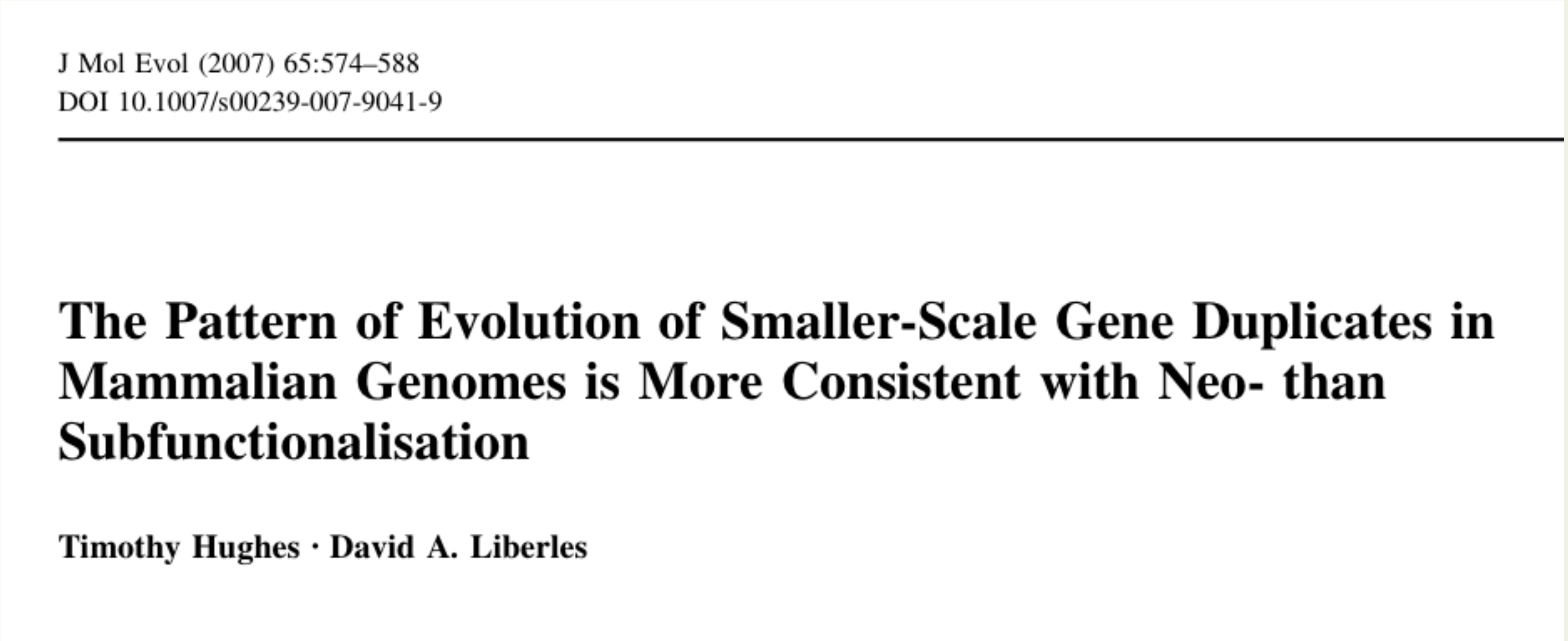


J Mol Evol (2007) 65:574–588
DOI 10.1007/s00239-007-9041-9

The Pattern of Evolution of Smaller-Scale Gene Duplicates in Mammalian Genomes is More Consistent with Neo- than Subfunctionalisation

Timothy Hughes · David A. Liberles

JC by Tõnu 4.04.2011

Intro

- Gene duplication and the accompanying release of negative selective pressure on the duplicate pair is thought to be the key process that makes functional change in the coding and regulatory regions of genomes possible.
- There are a number of models for the fate of gene duplicates, the two most prominent of which are ***neofunctionalisation*** and ***subfunctionalisation***, but it is still unclear which is the dominant fate.

Mudelid –
neutraalne mudel e. null mudel

- Both copies are released from negative selective pressure and evolve neutrally
(*ratio of number of <u>replacement substitutions per replacement site</u> to number of <u>silent substitutions per silent site</u> equal to 1*).
- Both sequences are then effectively randomly exploring sequence space with the inevitable outcome that one of the duplicates eventually fixes a ***<u>null mutation</u>***.
- See mudel ei ole ilmselt tõene, kuna ta ennustab väga vähe duplikante, mis pole aga kooskõlas nende arvukusega genoomis.

Mudelid – The classical model (Ohno 1970)

- as long as one of the copies retains the ancestral function, the other copy is free to accumulate mutations that lead to either loss or gain/change in function (neofunctionalisation).
- Loss of function will remove a gene from selective pressure and it will eventually pseudogenise.
- the vast majority of potential mutations are classically thought to be of this kind, this is thought to be the most common fate of a duplicate under this model.

Mudelid: The duplication–degeneration–complementation (DDC) model

- takes into account the modularity of the regulatory regions of genes and demonstrates how *degenerative mutations* in complementary regions can *lead to retention* of both duplicate copies through the evolutionary requirement to retain all the regulatory regions of the original gene.
- As this model does not require beneficial mutations to explain the retention of both duplicates in a pair, it has been characterised as *near-neutral*.

Mudelid: teised near-neutral & mitte-neutraalsed

- “increased robustness” by maintaining conserved backup copy (Kuepfer et al
- “increased dosage” by increasing exp from a gene that is already highly expr little mutational capacity to further inc expression
- “dosage compensation” by maintainin expression levels for stoichiometric rea et al. 2006).

imply strong negative selective pressure on the coding regions of gene duplicates NO case SSD

presupposes that whole sets of interacting genes with strong stoichiometric constraints have been duplicated (assumes **WGD**)

Dataset - SSD

- Small-scale duplications SSD
- they exclude WGD because A WGD produces a context for the duplicated gene that is radically different from the context that results from a SSD.
- Species included:
  - Ciona intestinalis (sea squirt),
  - Canis familiaris (dog), *
  - Homo sapiens (human), *
  - Pan troglodytes (chimp),
  - Mus musculus (mouse), *
  - Rattus nor-vegicus (rat), *
  - Gallus gallus (chicken)
  - Xenopus tropicalis (frog).
- duplitseerunud geenide paarid: 1-1 suhe

Reaalselt andmetelt mõõdeti

- PAML paketi programmiga CodeML (Yang et al 1997) kõigi paaride kohta:
  - the cumulative number of silent substitutions per silent site (**S**)
  - the cumulative number of replacement substitutions per replacement site (**R**)
  - They are referring to dR/dS as to the *instantaneous rate* of accumulation of replacement substitutions
  - This explains our somewhat alternative notation: it is more common to refer to S as dS (where S stands for synonymous) and to R as dN (where N stands for nonsynonymous)

$dN/dS = \omega$

but this notation would result in confusing notation for the instantaneous rate of accumulation of replacement substitutions (dR/dS)

Pre-assumptions

- This study makes the key assumption that silent substitutions have no effect on fitness and, therefore, that silent substitutions per silent site accumulate at a rate proportional to time.

- Eventually, silent sites get saturated with substitutions, leading to inaccurate estimation of S, but this should not be the case before S < 3.

Gene Duplicate Survival Analysis

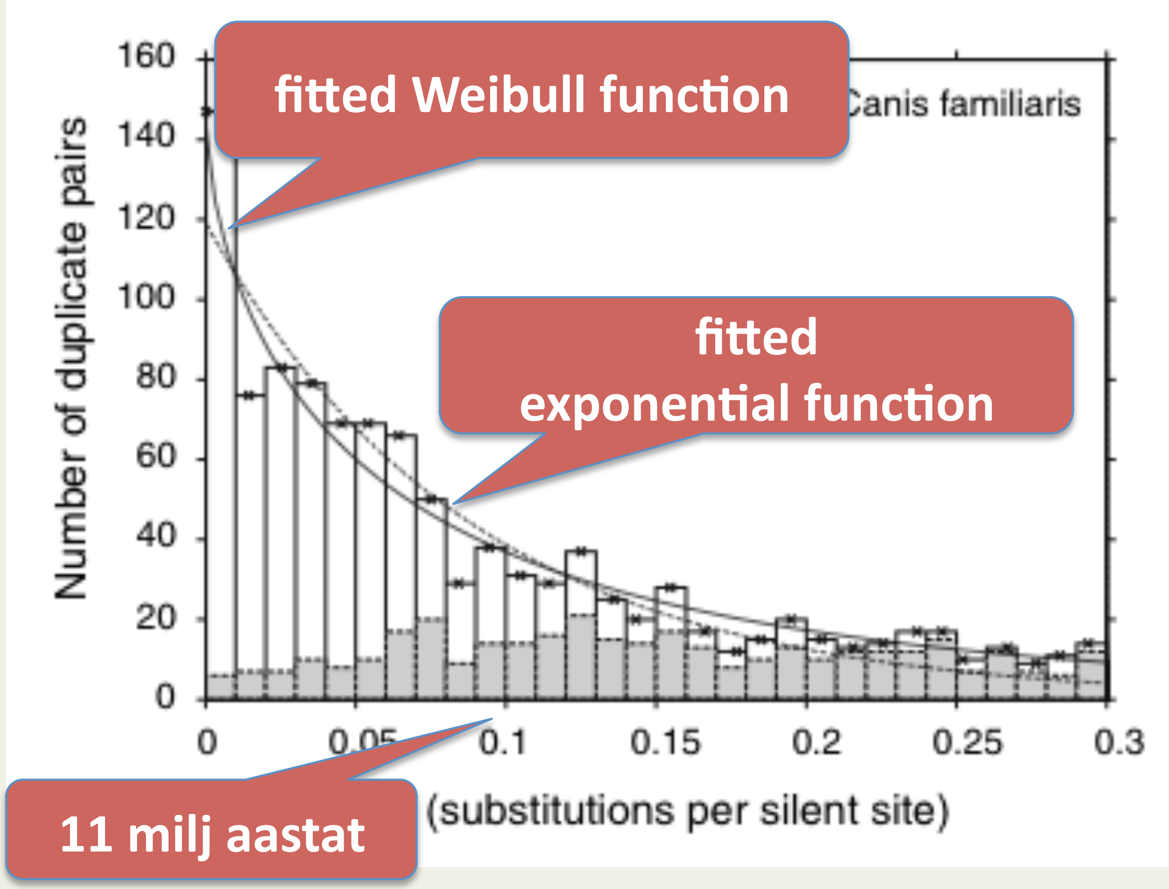
- To obtain estimates of the rates of gene duplication (birth) and pseudogenisation (death), they model the survival of duplicate pairs by assuming that gene ‘‘birth’’ and ‘‘death’’ are steady-state processes.

$$0 < S < 0.3$$

to ensure a high likelihood that the assumptions of S rate constancy and of non-saturation are valid

duplikaatide paarid jagati gruppidesse incremendiga S = 0.01 mis vastab ligikaudu 1.1 miljonile aastale.

Age distribution of duplicate pairs



fitted Weibull function

fitted exponential function

11 milj aastat

Fig. 1 Age distribution of duplicate pairs. Total column height: counts of duplicate pairs within each interval of size 0.01 S. Crosses: median S value for each interval. Shaded area: duplicate pairs with R/S significantly different from 1. Nonshaded area: duplicate pairs with R/S not significantly different from 1. Solid line: fitted Weibull function. Dotted line: fitted exponential function

nats valemeid

- Two common survival functions are the Weibull function

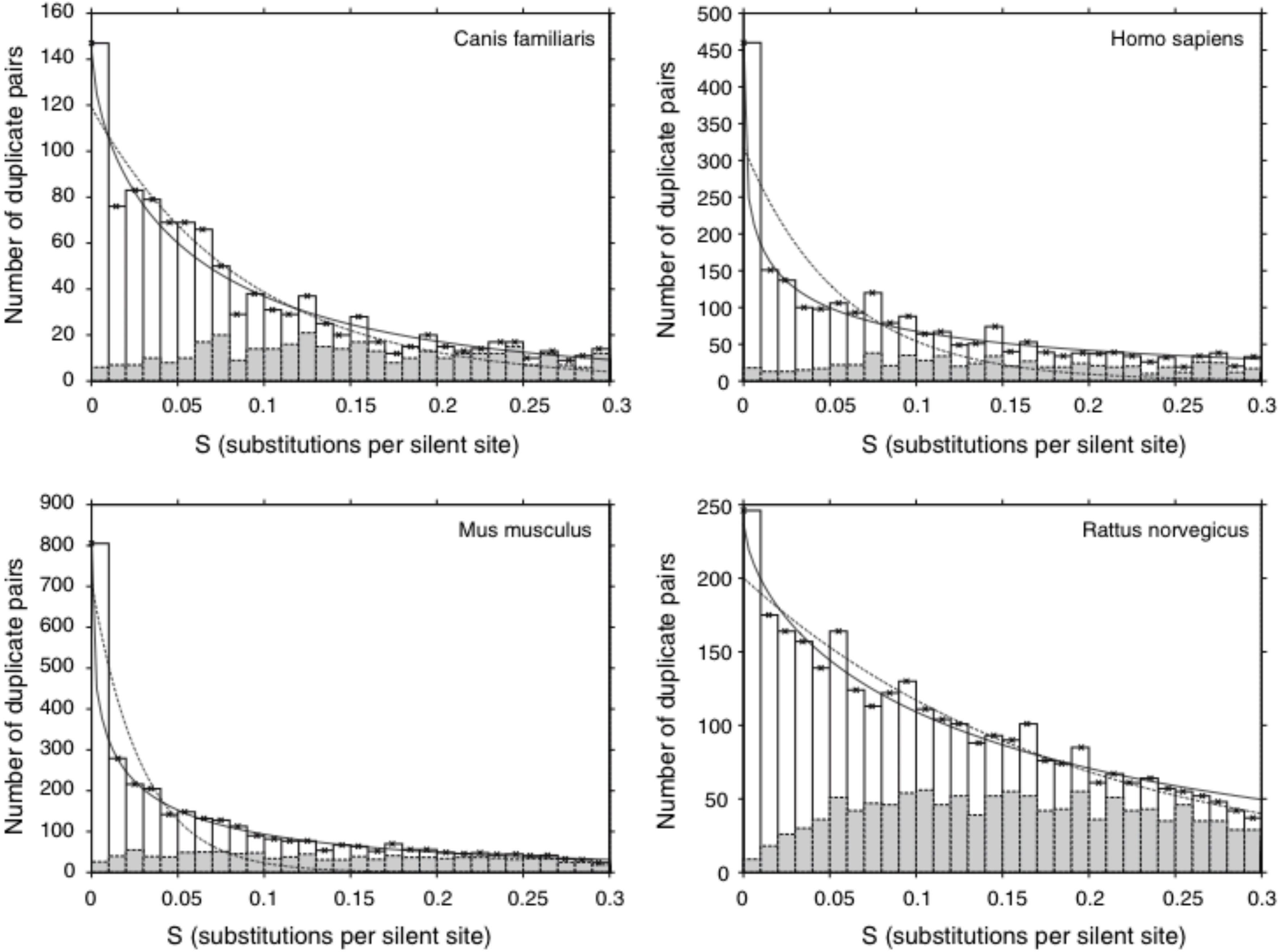
$$Q(t) = e^{\rho_1{}^{t^{\rho 2}}}$$

where T is the time of death and its special case, the exponential function where $\rho_2 = 1$.

- They use these survival functions and **S** as a proxy for time to model the mean of the Poisson distribution:

$$E(N_{S_i}) = N_0 e^{\rho_1 S_i^{\rho 2}}$$

where N_{Si} the number of duplicate pairs observed at age S_i, and N_0, ρ_1 and ρ_2 are fitted parameters.


Canis familiaris
Homo sapiens
Mus musculus
Rattus norvegicus
Number of duplicate pairs
S (substitutions per silent site)
0
0.05
0.1
0.15
0.2
0.25
0.3
160
140
120
100
80
60
40
20
500
450
400
350
300
250
200
150
100
50
900
800
700
600
500
400
300
200
100
250
200
150
100
50

The hazard (väljasuremise) function λ(*t*)

- is defined as the event (death/ failure/ pseudogenisation) rate at time *t* conditional on survival to time *t* or later:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{Pr(t < T < t + \Delta t | T > t)}{\Delta t} = -Q'(t)/Q(t) \tag{2}$$

Using **S** as a proxy for time:

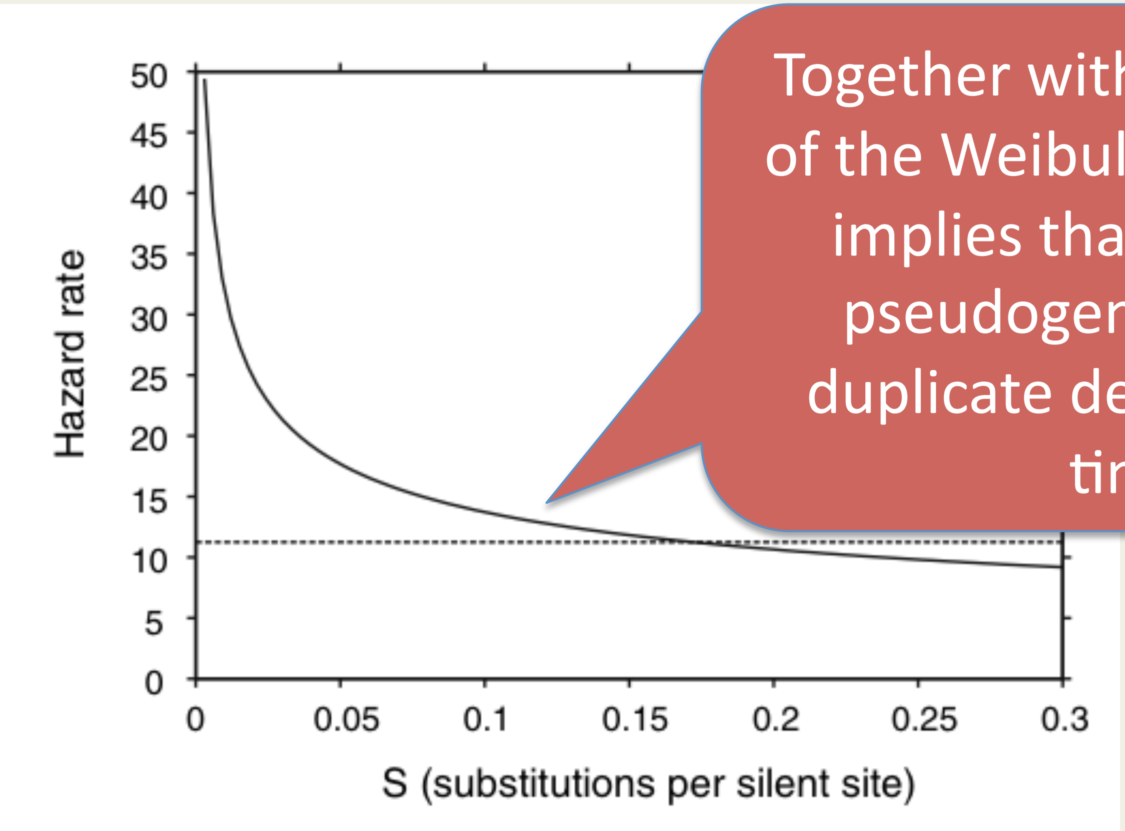
$$\lambda(S) = -\rho_1 \rho_2 S^{\rho_2 - 1}$$

Weibull survival function

$$\lambda(S) = -\rho_1$$

Exponential survival function

Duplicate pair hazard functions



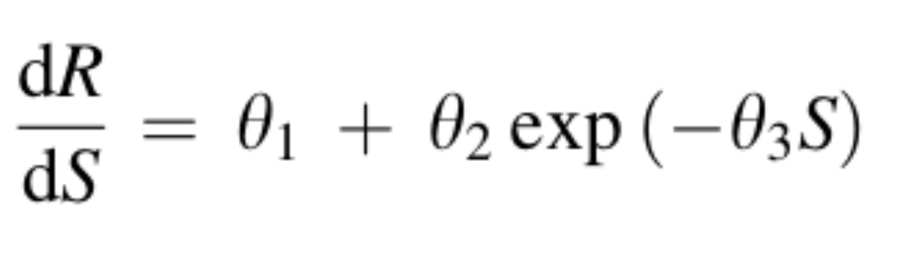
Together with the better fit of the Weibull function, this implies that the rate of pseudogenisation of a duplicate decreases with time.

Figure 2. Duplicate pair hazard functions. Solid line: hazard function for the Weibull survival function. Dotted line: hazard function for the exponential survival function

Accumulation of Replacement Substitutions Modeling

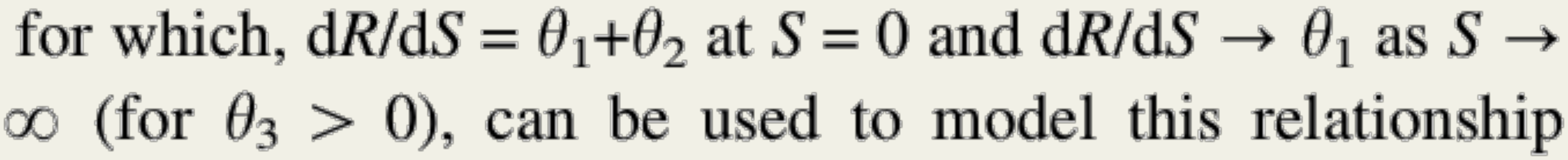
- In order to investigate the accumulation of replacement substitutions, they plot all duplicate pairs with S < 3 .

The equation:



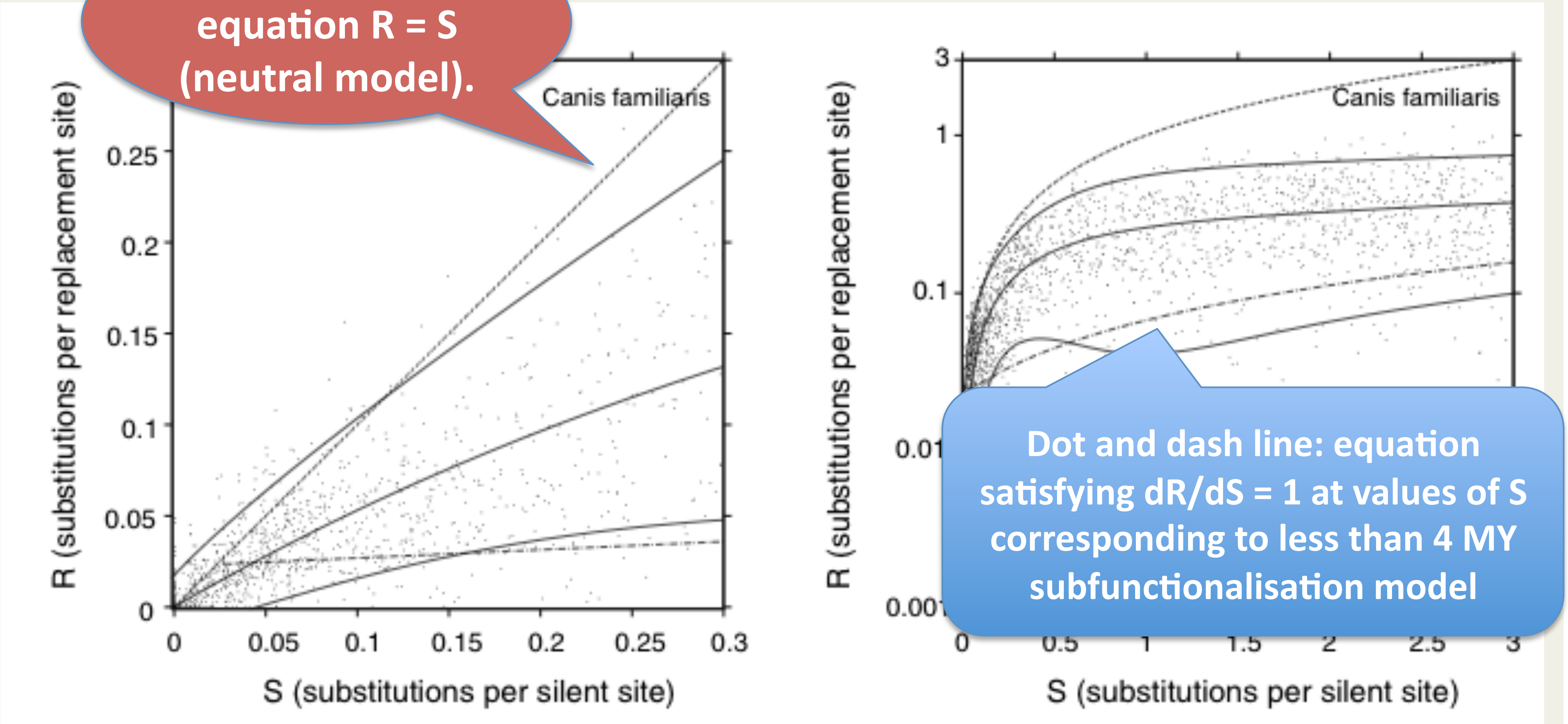
$$\frac{\mathrm{d}R}{\mathrm{d}S} = \theta_1 + \theta_2 \exp\left(-\theta_3 S\right)$$

Eq 3.



for which, $\mathrm{d}R/\mathrm{d}S = \theta_1+\theta_2$ at $S = 0$ and $\mathrm{d}R/\mathrm{d}S \to \theta_1$ as $S \to \infty$ (for $\theta_3 > 0$), can be used to model this relationship

Substitutions perreplacement site (R) as a f f substitutions per silent site (S)



Dashed line: equation R = S (neutral model).

Dot and dash line: equation satisfying dR/dS = 1 at values of S corresponding to less than 4 MY subfunctionalisation model

Solid lines: middle line is Eq. (4) fitted to the data, lowest and highest lines are the 5% and 95% quantiles of the distribution of R for a given value of S derived using Eqs. (4) and (5).

Conclusions

- In this paper, we have shown that the Weibull survival function provides the best fit to the age distribution data which, given the estimated values of the parameters, implies a decreasing convex hazard function which is most consistent with duplicate pairs following a neofunctionalisation model of evolution.
- These findings suggest that, for the smaller-scale duplicates that evade pseudogenisation, neofunctionalisation is a common fate.