

## Evo2: Molecular Evolution

Sexual dimorphism = physical and behavioral differences between males and females of the same species. This is common in many animal species.

In most species, it is the males that show extreme or extravagant secondary sexual characteristics. Examples include antlers and colorful ornaments (such as feathers). Charles Darwin explained the occurrence of such traits by his theory of sexual selection, which has two main components:

- 1) male-male competition = males must compete directly with each other for resources, territory, access to females, *etc.*
- 2) Female choice = females choose to mate with males that they find “attractive”.

Although males and females may differ greatly in phenotype, they are nearly identical in genotype. In species like human and *Drosophila*, the only genetic difference between males and females is the male-specific Y chromosome, which has very few genes. About 99.8% of all genes are present in both males and females.

Human:  $\approx 25,000$  genes total,  $\approx 60$  on the Y chromosome

*Drosophila*:  $\approx 15,000$  genes total,  $\approx 40$  on the Y chromosome

This suggests that most of the phenotypic differences between males and females are caused by differences in expression of genes present in both sexes.

Sex-biased genes have higher expression in one sex than in the other

Male-biased genes have higher expression in males than in females

Female-biased genes have higher expression in females than in males

Unbiased genes are expressed equally in males and females

It is now possible to use microarrays or RNA-seq to detect genes that are expressed differently between males and females. Furthermore, one can use comparative genomics to compare protein/DNA sequence divergence between species. With these data, one can address questions like:

How does sex-biased gene expression affect the rate of molecular evolution?

Do male-biased genes evolve faster than female-biased genes?

As an example, one can look at sex-biased gene evolution in the fruitfly *Drosophila melanogaster* and its close relatives, such as *Drosophila simulans*.

The rate of protein evolution can be measured by the divergence in the amino acid sequence of a protein between species, where:

$D = \# \text{ of differences} / \# \text{ of aligned amino acids}$

On average, proteins of male-biased genes evolved much faster (have higher  $D$ ) than those of female-biased or unbiased genes.

Is this due to a higher mutation rate for male-biased genes? To rule out this possibility, one can look at DNA sequence divergence.

The genetic code is degenerate:

64 possible codons

3 are stop codons

61 encode 20 different amino acids

This means that many amino acids are encoded by more than one codon.

Nonsynonymous changes = alter the encoded amino acid

Synonymous changes = do not alter the amino acid

$dN$  = # of nonsynonymous differences/# of nonsynonymous sites

$dS$  = # of synonymous differences/# of synonymous sites

The ratio  $dN/dS$  measures protein divergence standardized by the mutation rate. We assume that the mutation rate is the same for nonsynonymous and synonymous sites.

$dN/dS$  is higher for male-biased genes than for female-biased or unbiased genes. This suggests that the difference is due to selection, not mutation.

Possibilities:

- 1) there is more positive (or sexual) selection acting on male-biased genes
- 2) there is less negative selection acting on male-biased genes

Current evidence from *Drosophila* supports possibility number 1.

There is a positive correlation between  $dN/dS$  and recombination rate for male-biased genes. Selection (both positive and negative) is expected to be more efficient in regions of high recombination.

What about humans?

In a comparison of human vs. chimpanzee, genes expressed in testis (which are male-biased genes) were the fastest evolving. Genes expressed in brain were the slowest evolving.

Figures:

Example: Amino acid sequence of a protein

Methionine Proline Valine Serine Threonine Leucine Glycine Isoleucine Lysine Phenylalanine ...

Met Pro Val Ser Thr Leu Gly Ile Lys Phe ...

**M P V S T L G I K F ...**

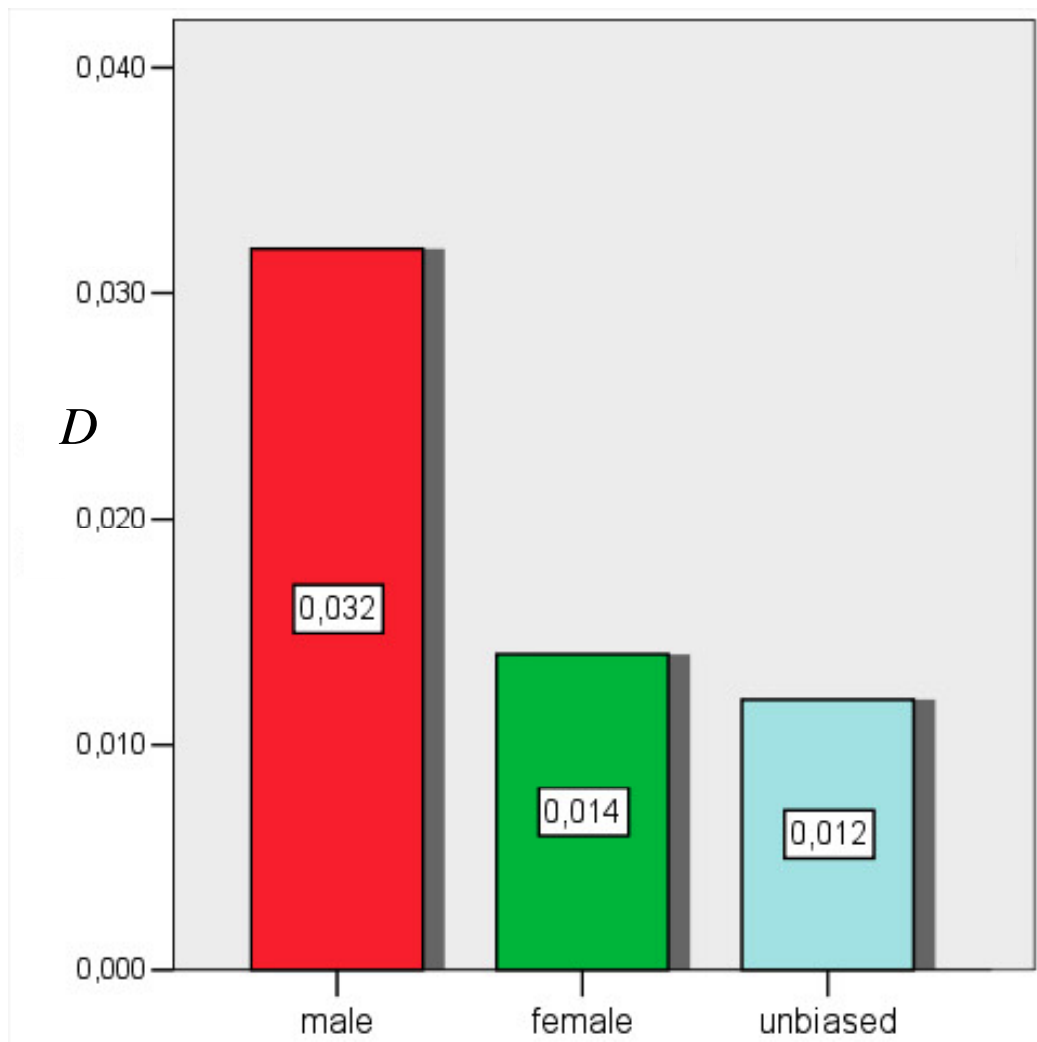
Alignment of the protein between two species

sp1: **M P V S T L G I K F**

sp2: **M P A S T L G L K F**

$$\text{Divergence} = D = \frac{\# \text{ of differences}}{\# \text{ of amino acids}}$$

$$D = \frac{2}{10} = 0.20 = 20\%$$



*D. melanogaster/D. simulans*

non-synonymous

TCT = Ser  
CCT = Pro  
ACT = Thr  
GCT = Ala

synonymous

GGT = Gly  
GGC = Gly  
GGA = Gly  
GGG = Gly

**Alignment of a protein-coding sequence between two species**

Pro Ala Ser Thr Gly Val Pro Gly Val Thr  
Sp. 1: CCC GCA TCG ACC GGA GTG CCT GGG GTC ACA

Sp. 2: CCC GCG TCG ACC GGA GTG CCA GGG GTC CCA  
Pro Ala Ser Thr Gly Val Pro Gly Val Pro

■ = non-synonymous sites

■ = synonymous sites

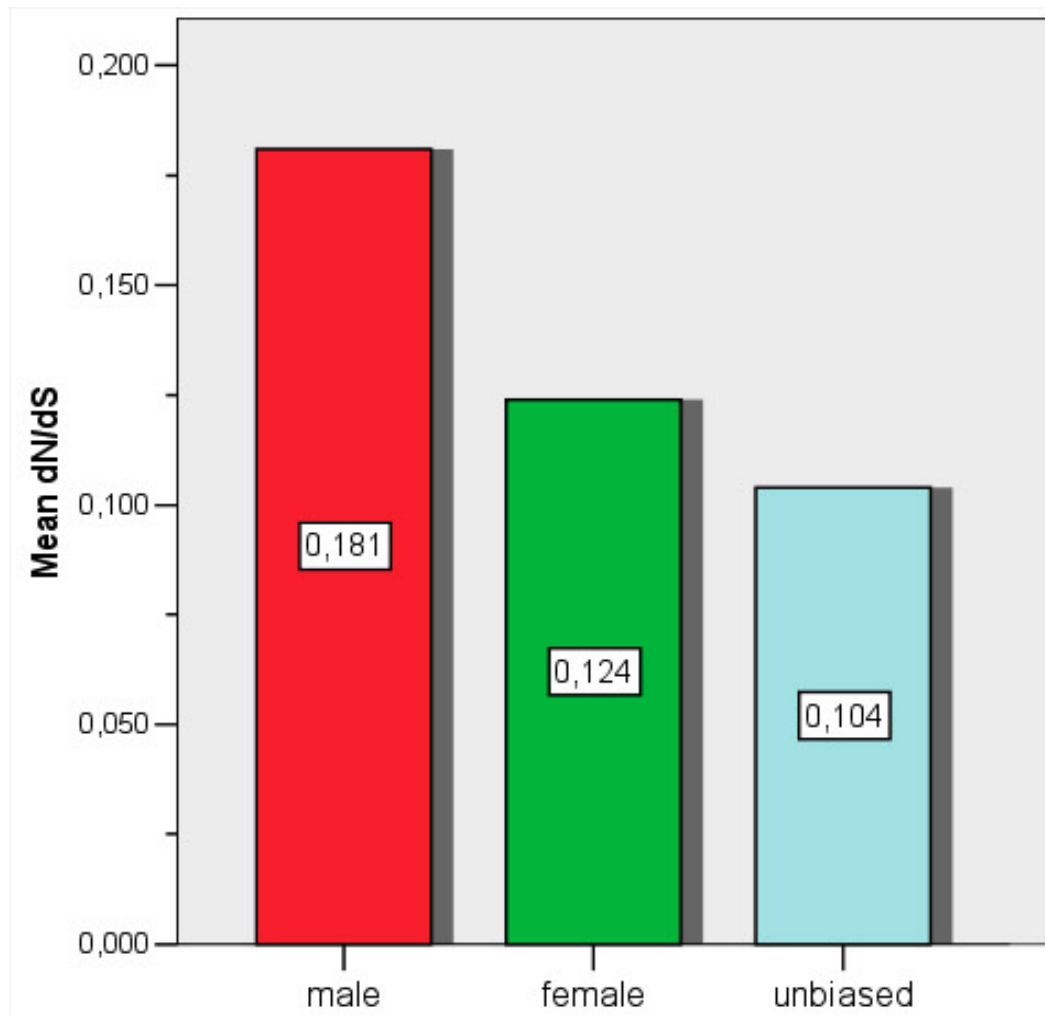
$$dN = \frac{\# \text{ non-syn. changes}}{\# \text{ non-syn. sites}}$$

$$dN = \frac{1}{20} = 0.05$$

$$dS = \frac{\# \text{ syn. changes}}{\# \text{ syn. sites}}$$

$$dS = \frac{2}{10} = 0.20$$

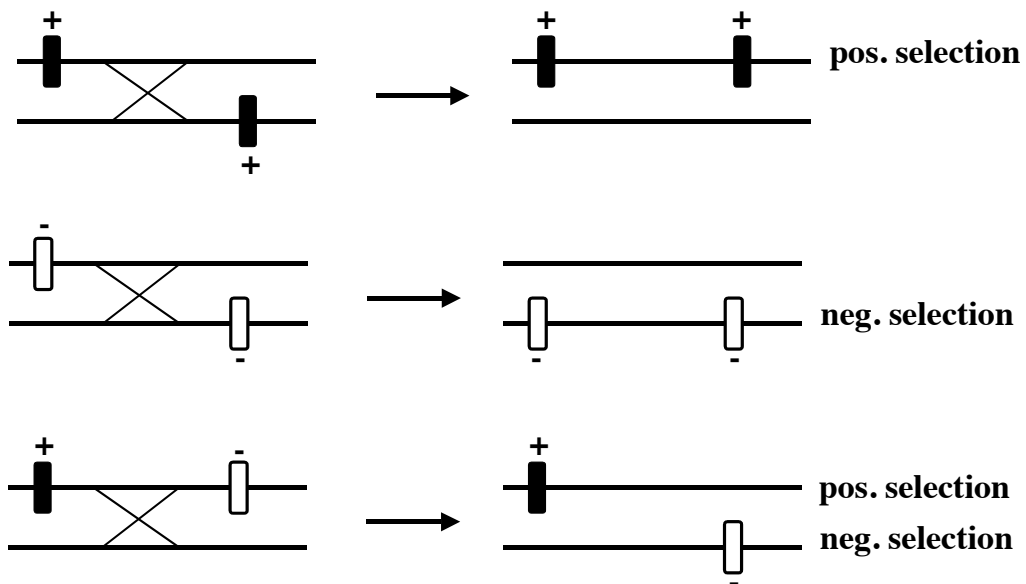
$$\frac{dN}{dS} = \frac{0.05}{0.20} = 0.25$$



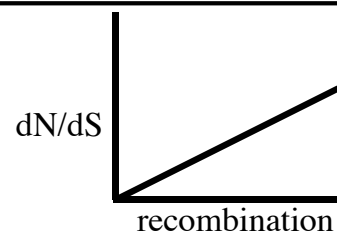
*D. melanogaster/D. simulans*

### dN/dS *versus* recombination: Hill-Robertson interference

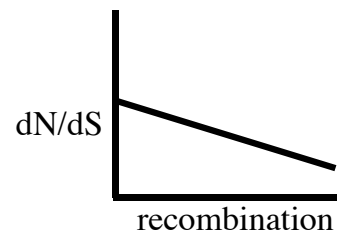
Selection (both positive and negative) is expected to be more efficient in regions of higher recombination. The local recombination rate varies along a chromosome.



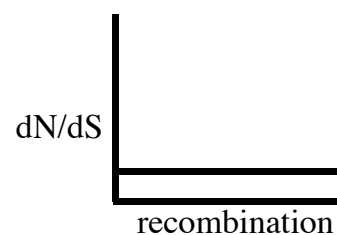
If there is frequent positive selection for amino acid replacements:  
dN/dS should be positively correlated with recombination rate

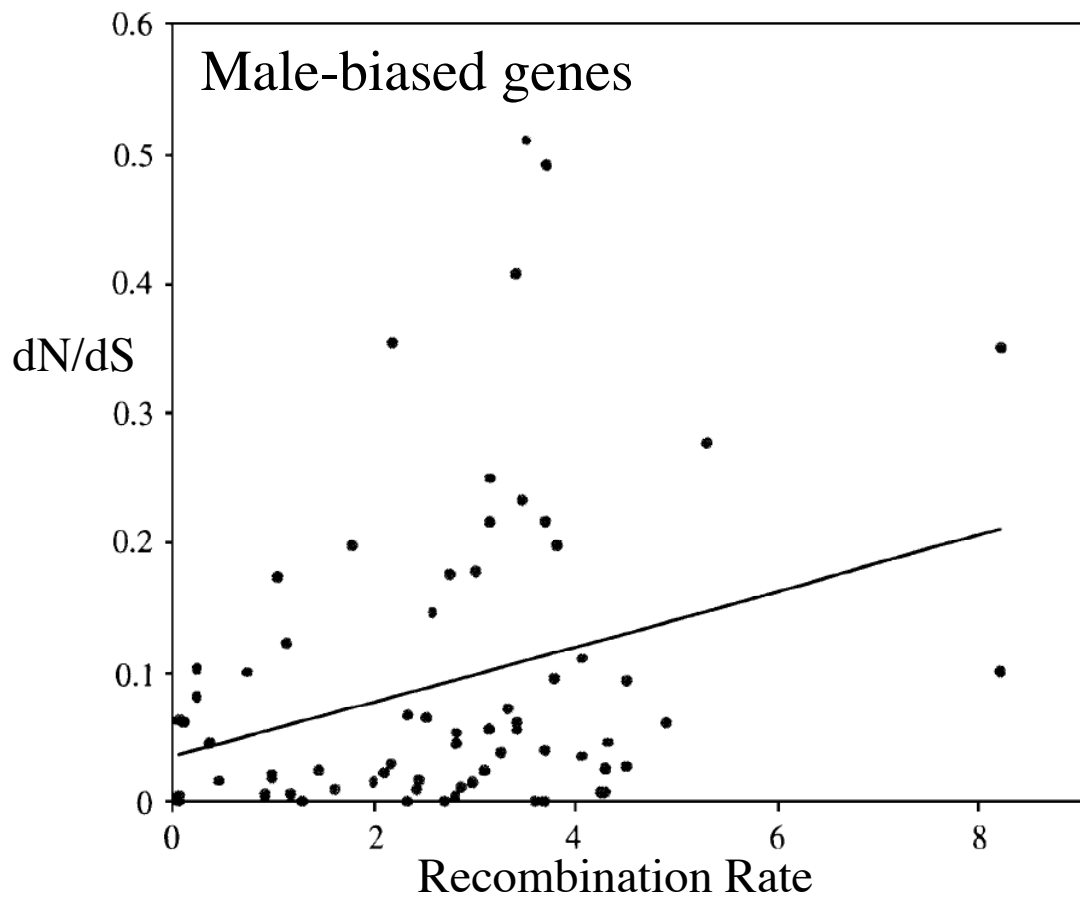


If there is weak negative selection against amino acid replacement:  
dN/dS should be negatively correlated with recombination rate

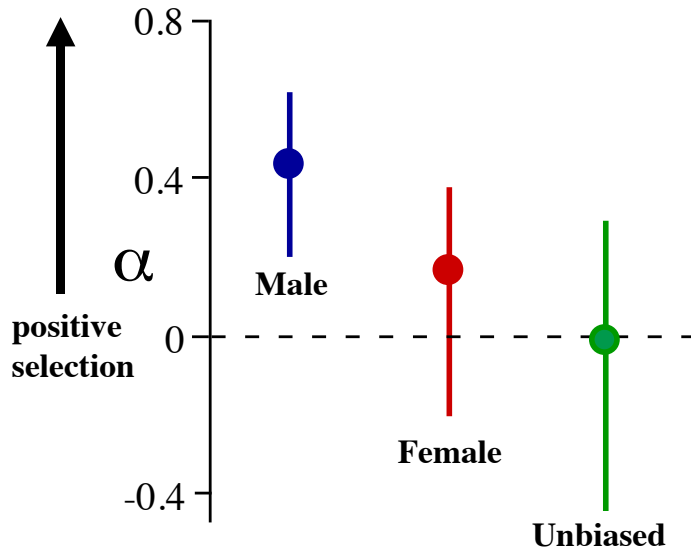


If most amino acid replacements are neutral:  
dN/dS should be not be correlated with recombination rate

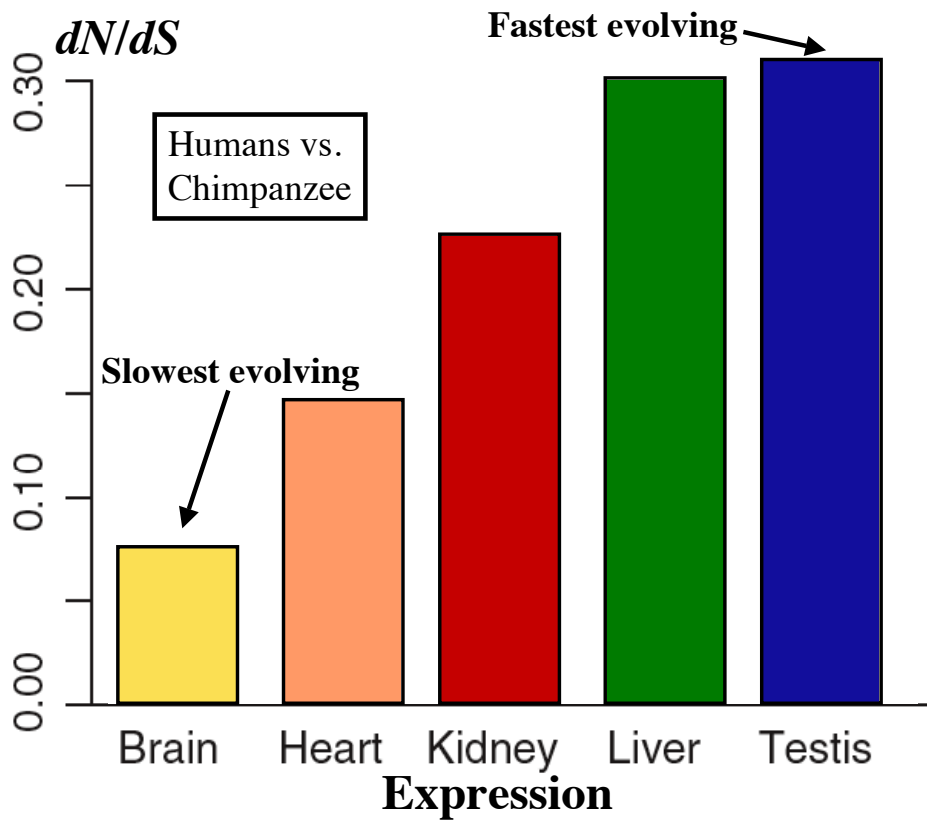








$\alpha$  = % of a.a. differences between species due to positive selection



## Speciation

The species problem: the long-standing failure of biologists to agree on how we should identify species and how we should define the word 'species'

A commonly-used species definition is the **Biological Species Concept (BSC)**, which defines species as "groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups." This definition is most associated with Ernst Mayr (1904-2005), a Harvard evolutionary biologist.

The BSC is useful for extant sexually reproducing species, but not for extinct species (e.g. fossils) or those that reproduce asexually (e.g. bacteria)

Under the BSC, the defining aspect of a species is reproductive isolation

The two major types of reproductive isolation are:

-- Pre-zygotic (or pre-mating) – the different organisms do not mate with each other under natural conditions. This may be due to behavioral or physical barriers to mating.

-- Post-zygotic (or post-mating) – the different organisms do mate with each other, but their hybrid offspring are inviable or infertile.

*Drosophila* species have complex courtship (mating) behavior that involves 'singing' by the males (vibrating one wing). *Drosophila* females tend to be choosy about the males that they mate with, typically rejecting males that are not of their own species. Thus, there is pre-zygotic reproductive isolation between most *Drosophila* species. In some cases, different *Drosophila* species will mate with each other, but the offspring from these matings either do not survive or are infertile. Thus, there is also post-zygotic reproductive isolation between *Drosophila* species.

## Ecological specialization

*Drosophila sechellia* is endemic to islands of the Seychelles archipelago. It feeds and lays its eggs exclusively on the fruit of *Morinda citrifolia* (it is a specialist).

*Drosophila simulans* is endemic to continental Africa, but is now found worldwide. It feeds and lays its eggs on many types of fruit (it is a generalist). However, *Morinda citrifolia* is toxic to *Drosophila simulans* and most other *Drosophila* species.

The divergence time between *D. sechellia* and *D. simulans* is less than 1 million years. This suggests that *D. sechellia* has rapidly evolved a new preference for *M. citrifolia*, as well as a resistance to its toxic effects, allowing it to occupy a new ecological niche.

Figure: *Drosophila* courtship behavior

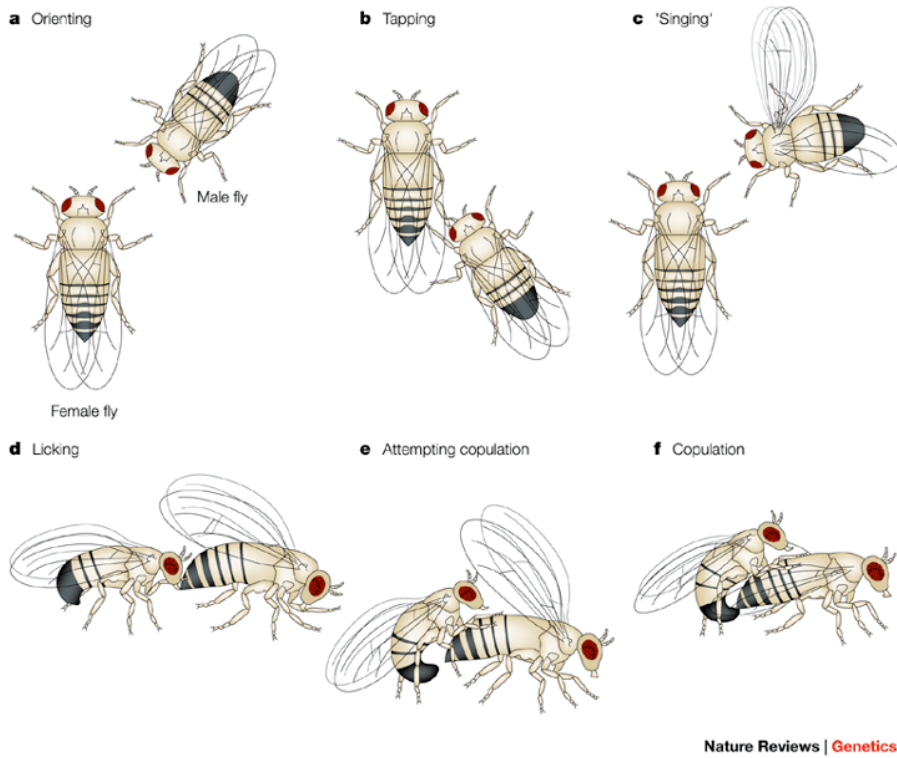


Figure: test of oviposition (egg-laying) preference and ecological specialization

## Experiment

