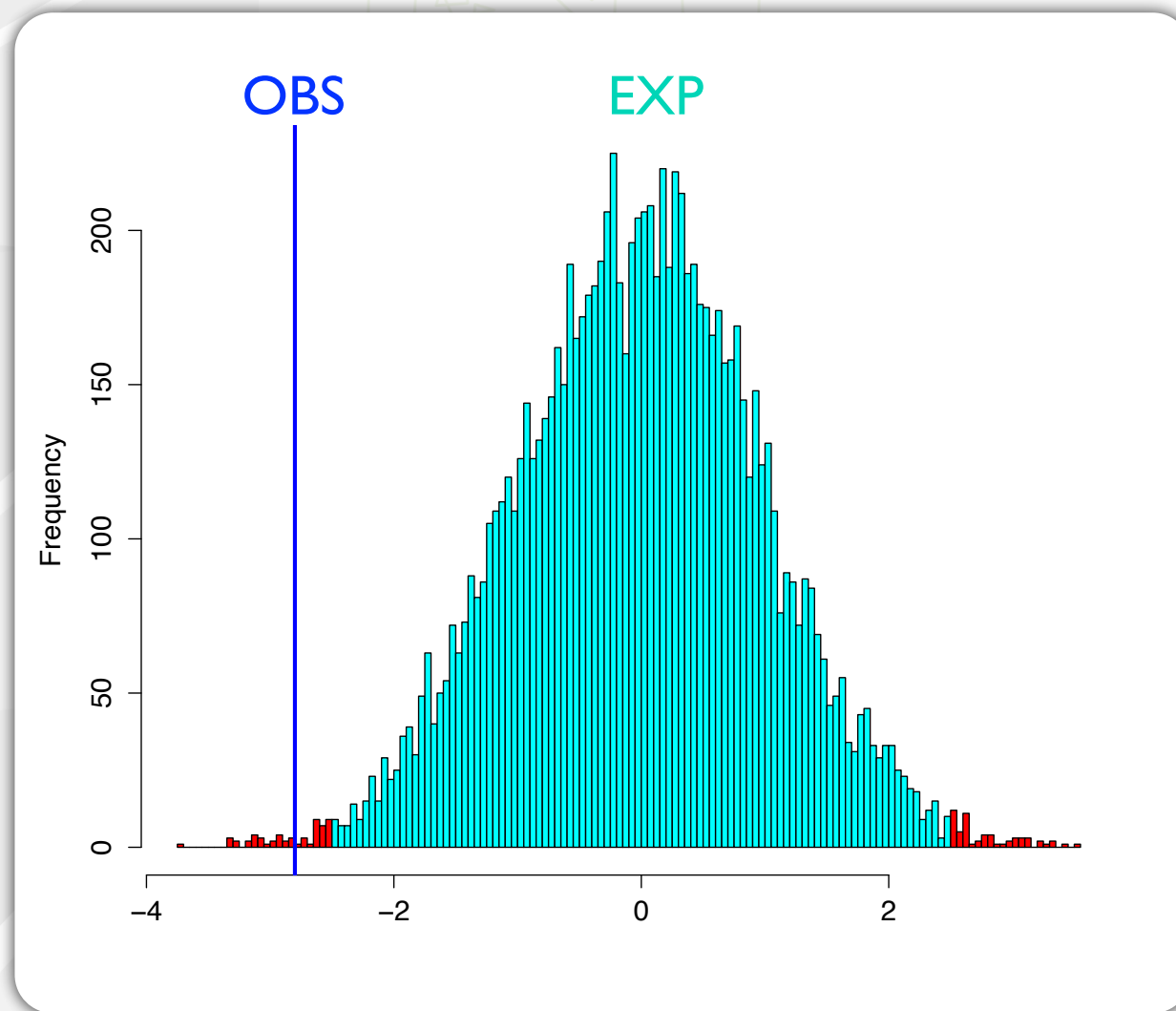


Detection of Natural Selection at given loci



Detection of Natural Selection at given loci

- ❖ Methodologies to detect the effect of natural selection using molecular data.
- ❖ Methodologies dependent on the time scale being studied:
 - ❖ Detection of incipient selection.
 - ❖ Detection of selective events in populations.
 - ❖ Detection of selective events between species.

Methodologies to detect the effect of natural selection using molecular data.

- Molecular data shows variants that have accumulated in populations and their frequencies.

Methodologies to detect the effect of natural selection using molecular data.

- Molecular data shows variants that have accumulated in populations and their frequencies.

```
sel=0      281
AY5Brek    AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
AYLW2      AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
AYP2       AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
AYSBlackJ  AACTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
DQ3Kwb     AACTCTAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
DQ0Kwb     AACTCTAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
ABlriuk    AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
AB2riuk    AATTCCAGTCAACATGCGTATCACTACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
AB5riuk    AATTCCAGTCAACATGCGTATCACTACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
EFWbChYun  AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
EFWbChFu   AACTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
EFWbChHai  AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
ChWbdb6    AACTCCAGTCAACATGCGTATCATCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
ChWb26     AATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
EFwb371    AACTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
ChWb48     AACTCCAGTCAACATGCGTATCACTACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
ChWb47     AACTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
ChWb65     AACTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
EUChiN     AACTCTAGTCAACATGCGTATCATCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
EFWbViet   AACTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
TypeIILanyu AACTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
TypeIILanyu AATTCCAGTCAACATGCGTATCACTGCCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
AYMeish1   AACTCCAGTCAACATGCGTATCACTACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
AYMeish2   AACTCCAGTCAACATGCGTATCACTACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
EFXiang    GATTCCAGTCAACATGCGTATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
AYJNP13    AACTCCAGTCAACATGCATATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAG
Phaco      AATTCCAGTCAACATGCATATCACCACCATTAGATCACGAGCTTAATTACCATGCCGCGTGAAACCAGCAACCCGCTTGGCAA
```

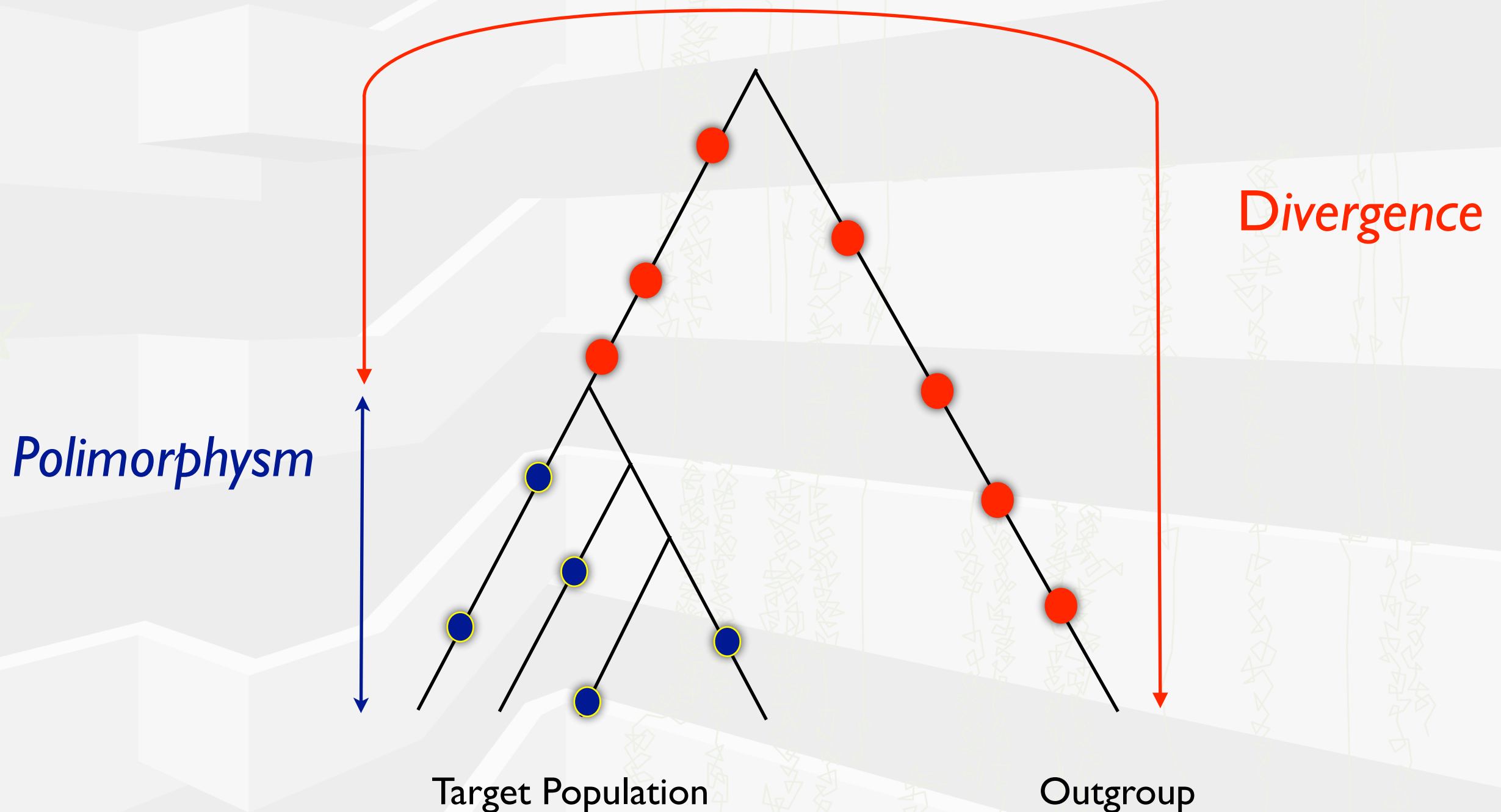

Methodologies to detect the effect of natural selection using molecular data.

- Molecular data shows variants that have accumulated in populations and their frequencies.
- Molecular data indirectly show the effect of natural selection.

Methodologies to detect the effect of natural selection using molecular data.

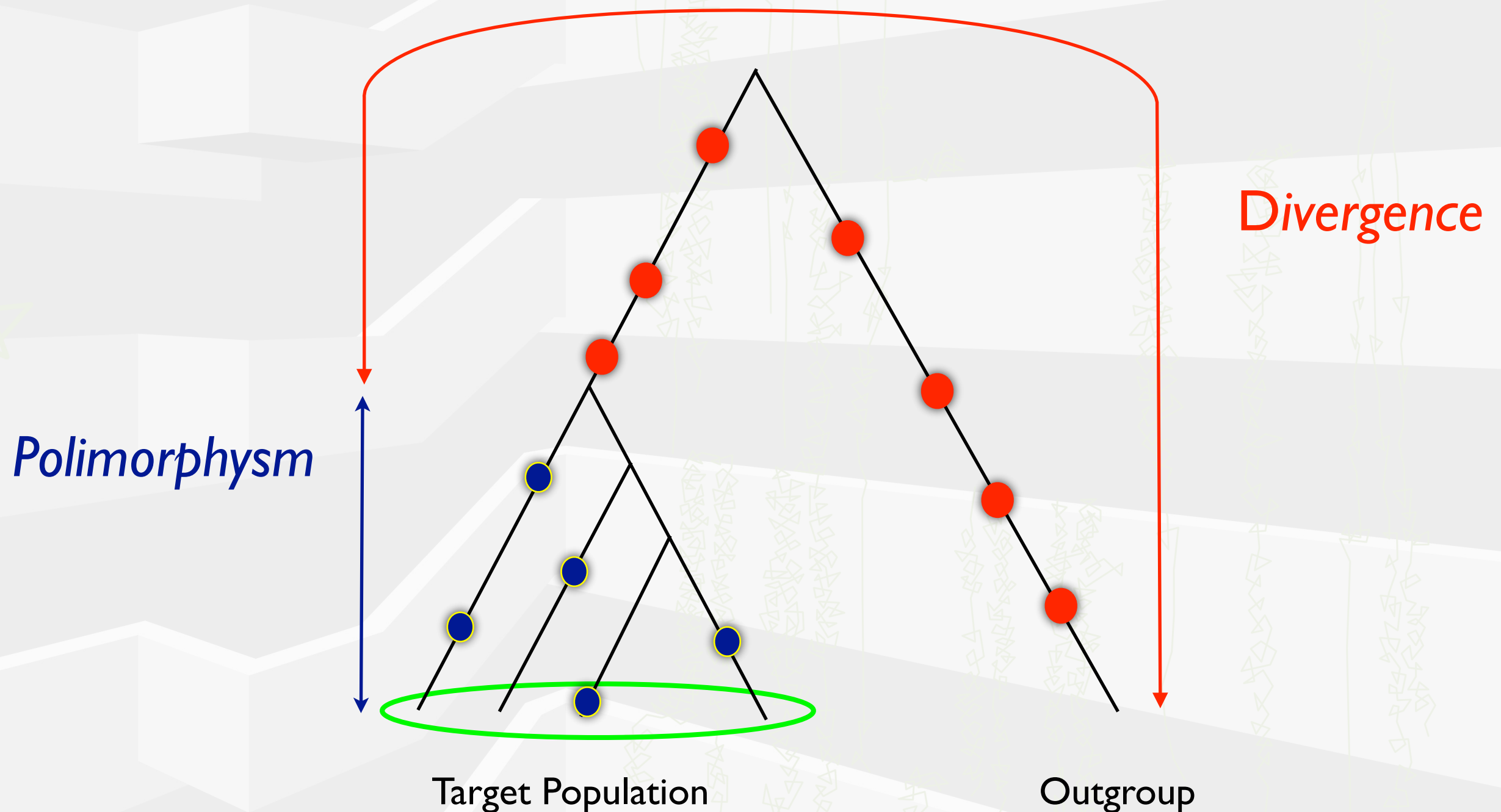
- Molecular data shows variants that have accumulated in populations and their frequencies.
- Molecular data indirectly show the effect of natural selection.
- The methodologies vary depending on the time scale to be studied.

The methodologies vary depending on the time scale to be studied.



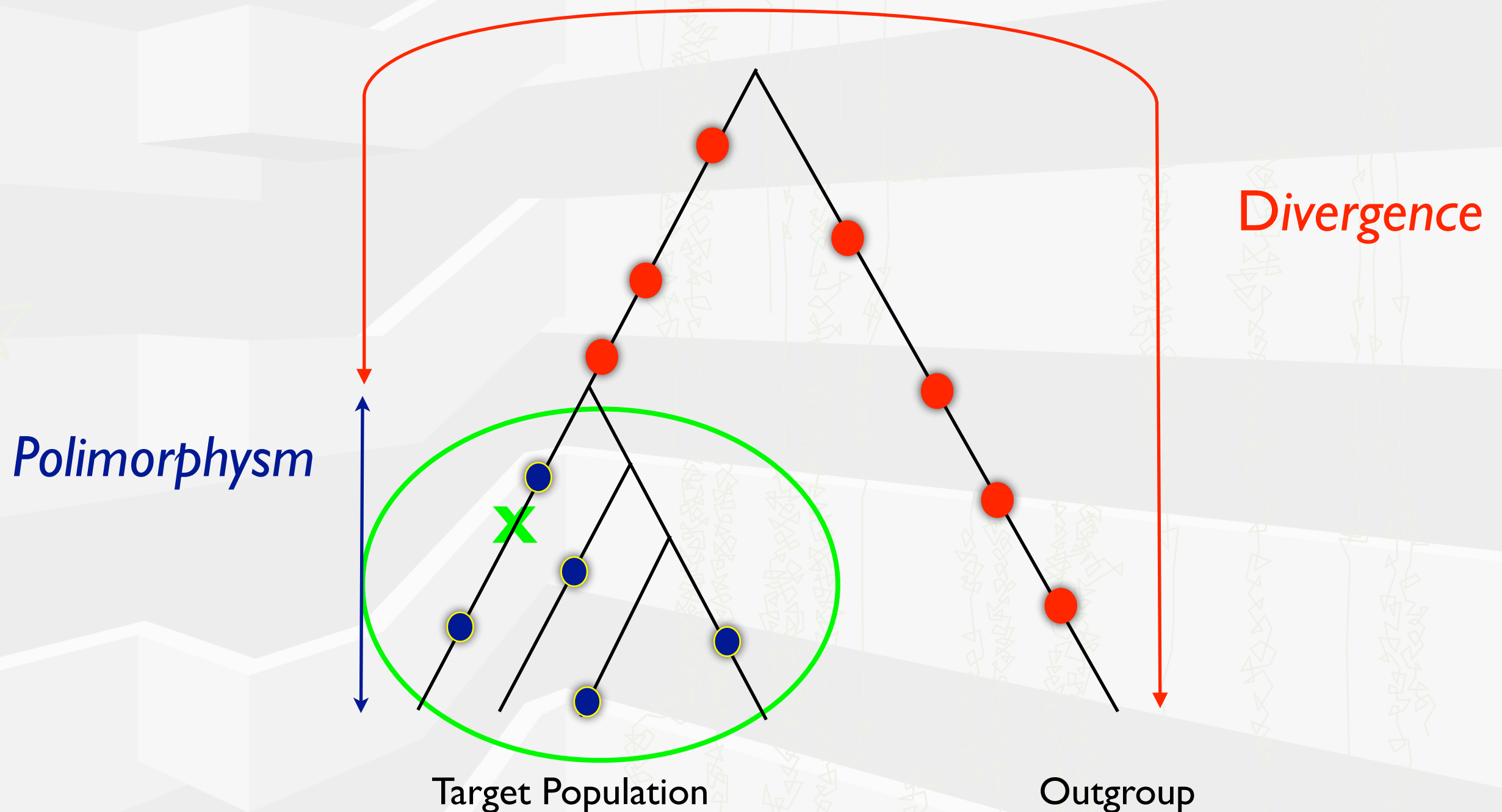
The methodologies vary depending on the time scale to be studied.

Incipient selection (start now)



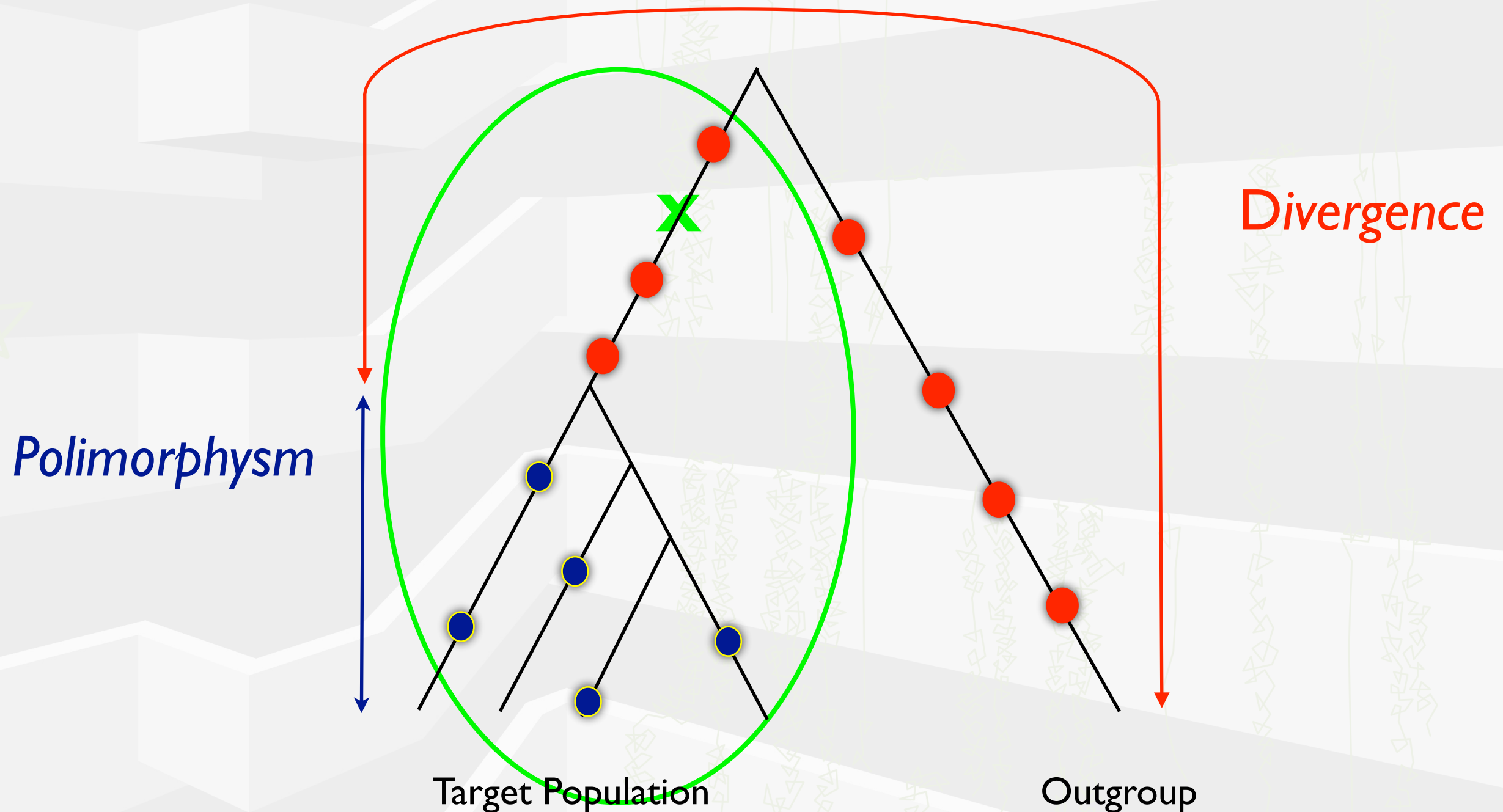
The methodologies vary depending on the time scale to be studied.

Selection within population (occurred and/or continues)



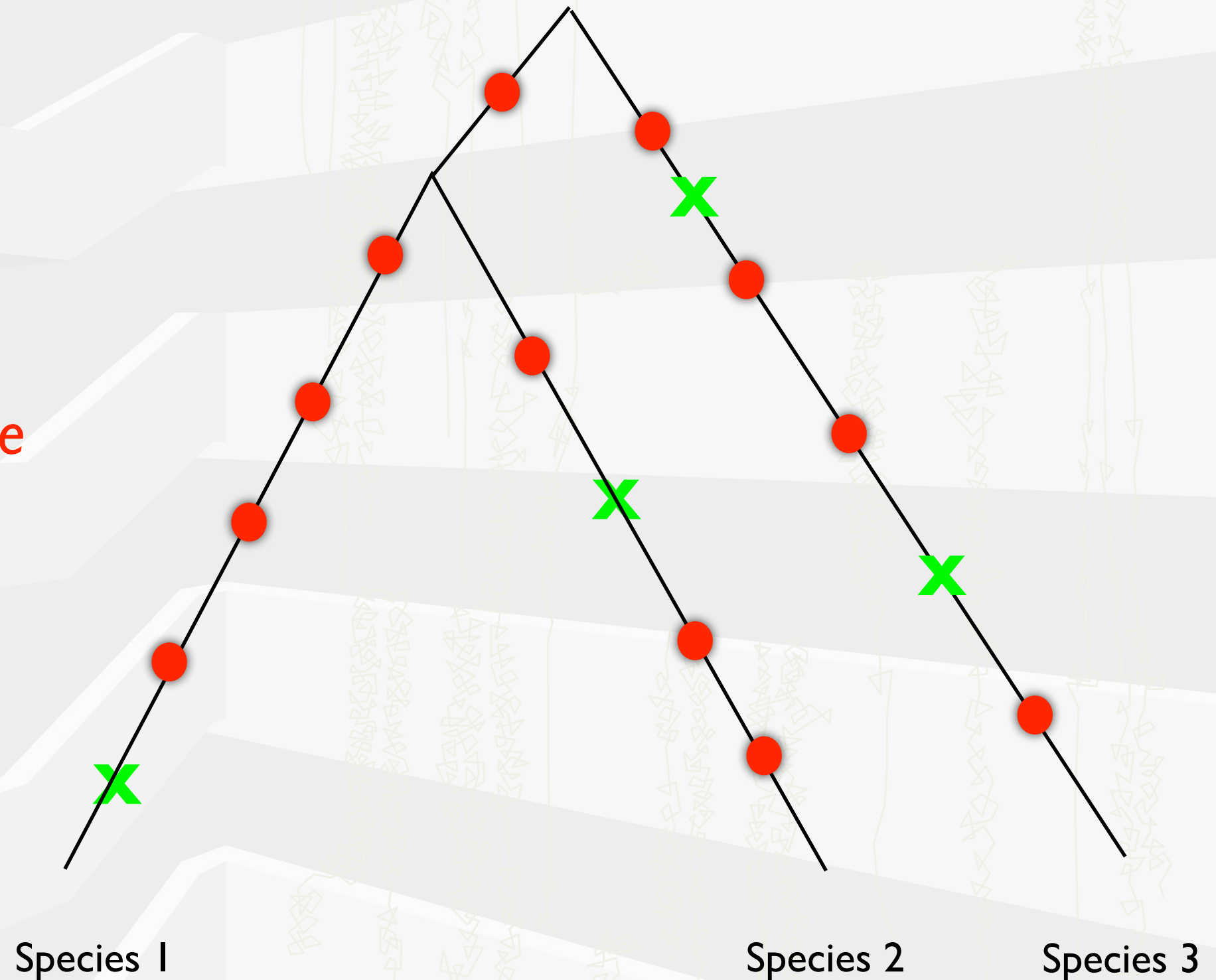
The methodologies vary depending on the time scale to be studied.

Selection in the species (adaptive mutations were fixed)



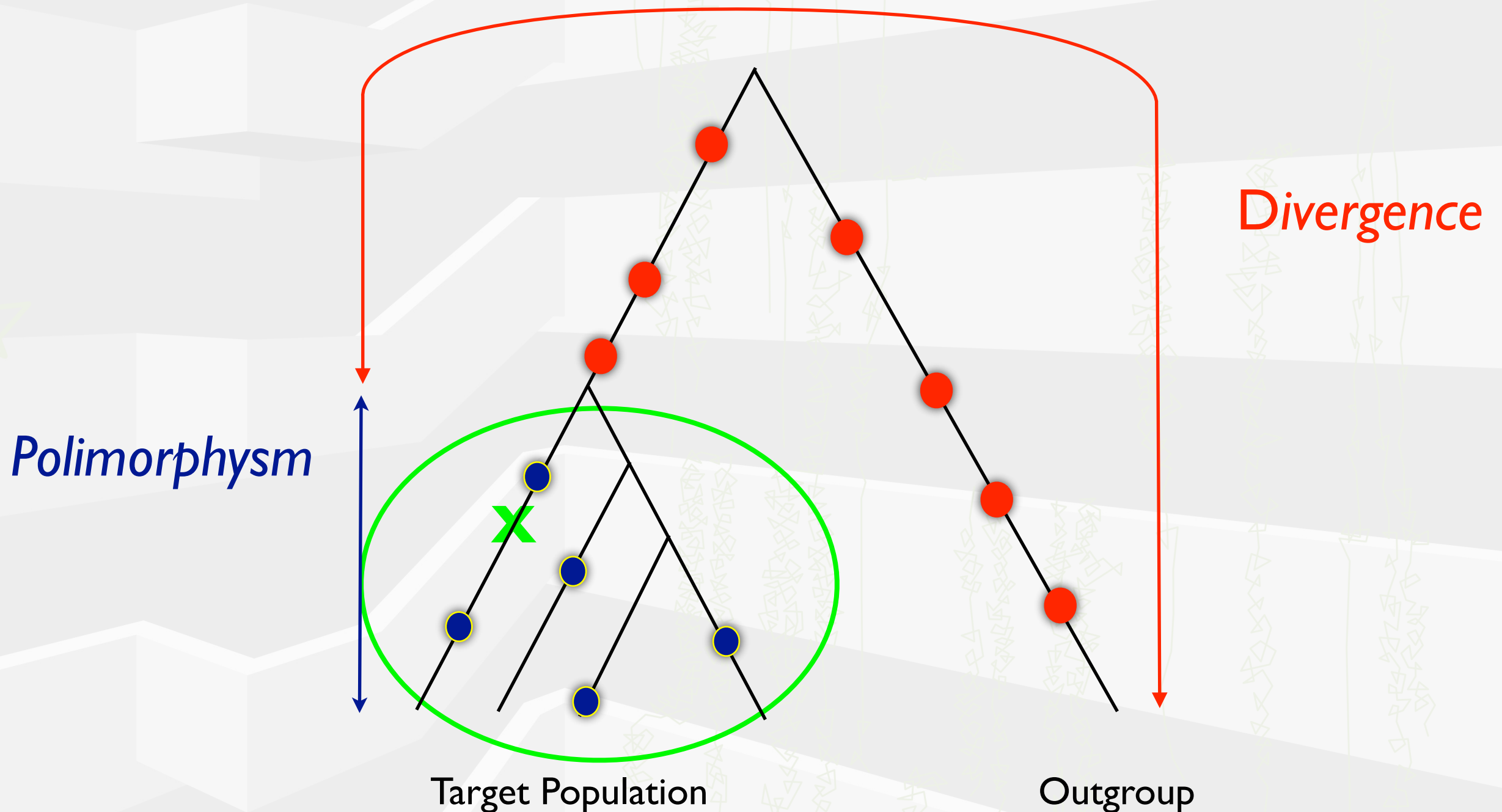
The methodologies vary depending on the time scale to be studied.

Divergence



The methodologies vary depending on the time scale to be studied.

Selection within population (occurred and/or continues)



Selection within population (occurred and/or continues)

- Effect on the patterns of variability: Affected by selection: drops/increase in levels of variability in specific region of the genome, changes in the frequency spectrum, and unusual Linkage Disequilibrium.
- Methodologies for detection: Comparative analyses using neutral expectations. Differences in the polymorphism/divergence ratio between loci or across functional/neutral positions. Detection of excess of linkage disequilibrium.
- Detection power: Acceptable detection range.

Unusual drop in variability in the region of the genome adjacent to *Sdic* in *Drosophila*

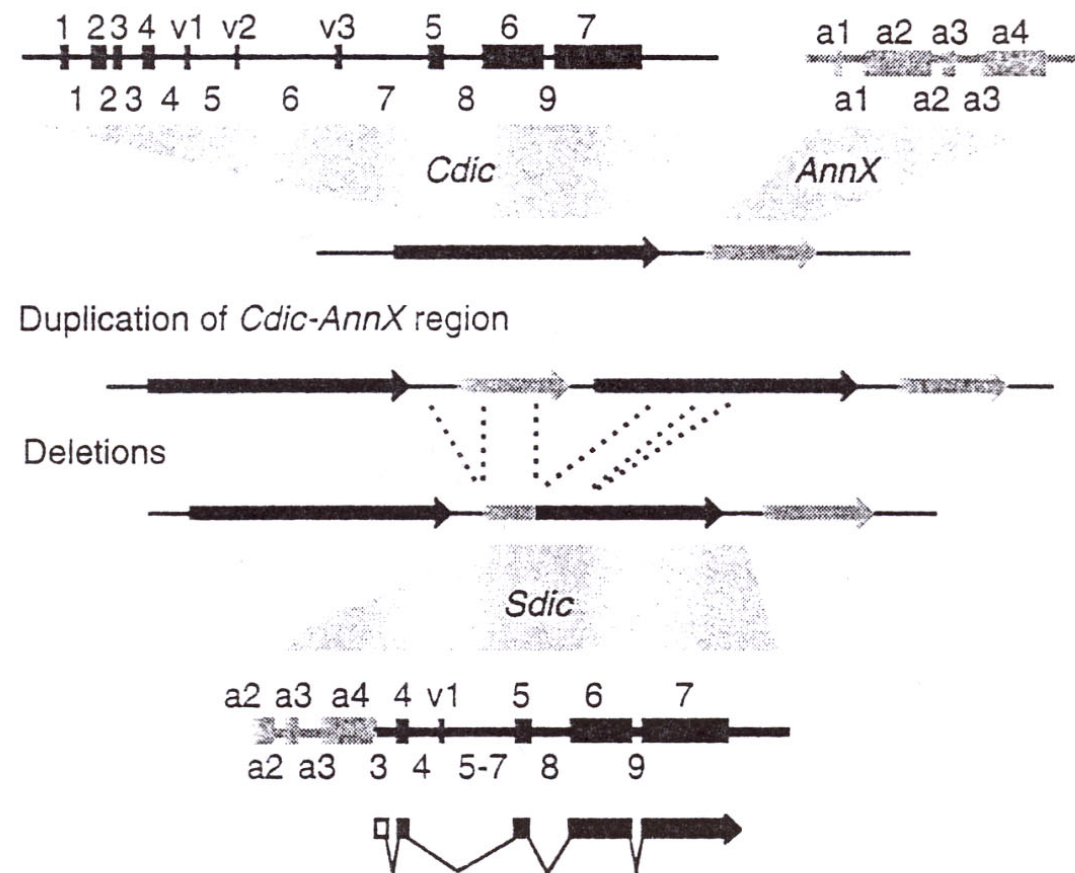


Figure 1 A model for the creation of the *Sdic* gene. Top, exon-intron structure of the *Cdic* and *AnnX* genes. Exons are numbered above the line and introns below the line; v denotes a variable (alternatively spliced) exon. The model below postulates a duplication of the *Cdic* (black arrow)-*AnnX* (grey arrow) region, followed by three deletions (dashed lines) to create *Sdic*. The *Sdic* gene has also been tandemly duplicated ~10-fold. At the bottom is the transcription and splicing pattern of the *Sdic* units.

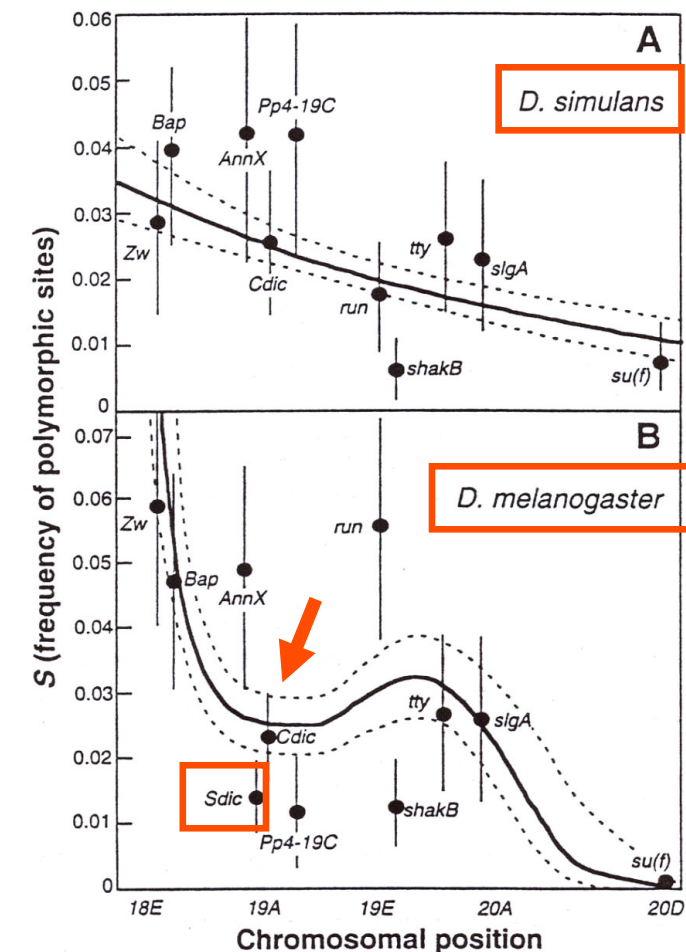


Fig. 1. (A) Results from *D. simulans* showing a monotonic decrease in the proportion of polymorphic sites (*S*) as a function of gene location at the base of the X chromosome. (B) Results from *D. melanogaster* showing a significant trough in the proportion of polymorphic sites (*S*) in the region around *Sdic*. The delimiter on each point is the approximate 50% confidence interval.

Balancing selection in the *Adh* gene in *Drosophila melanogaster*?

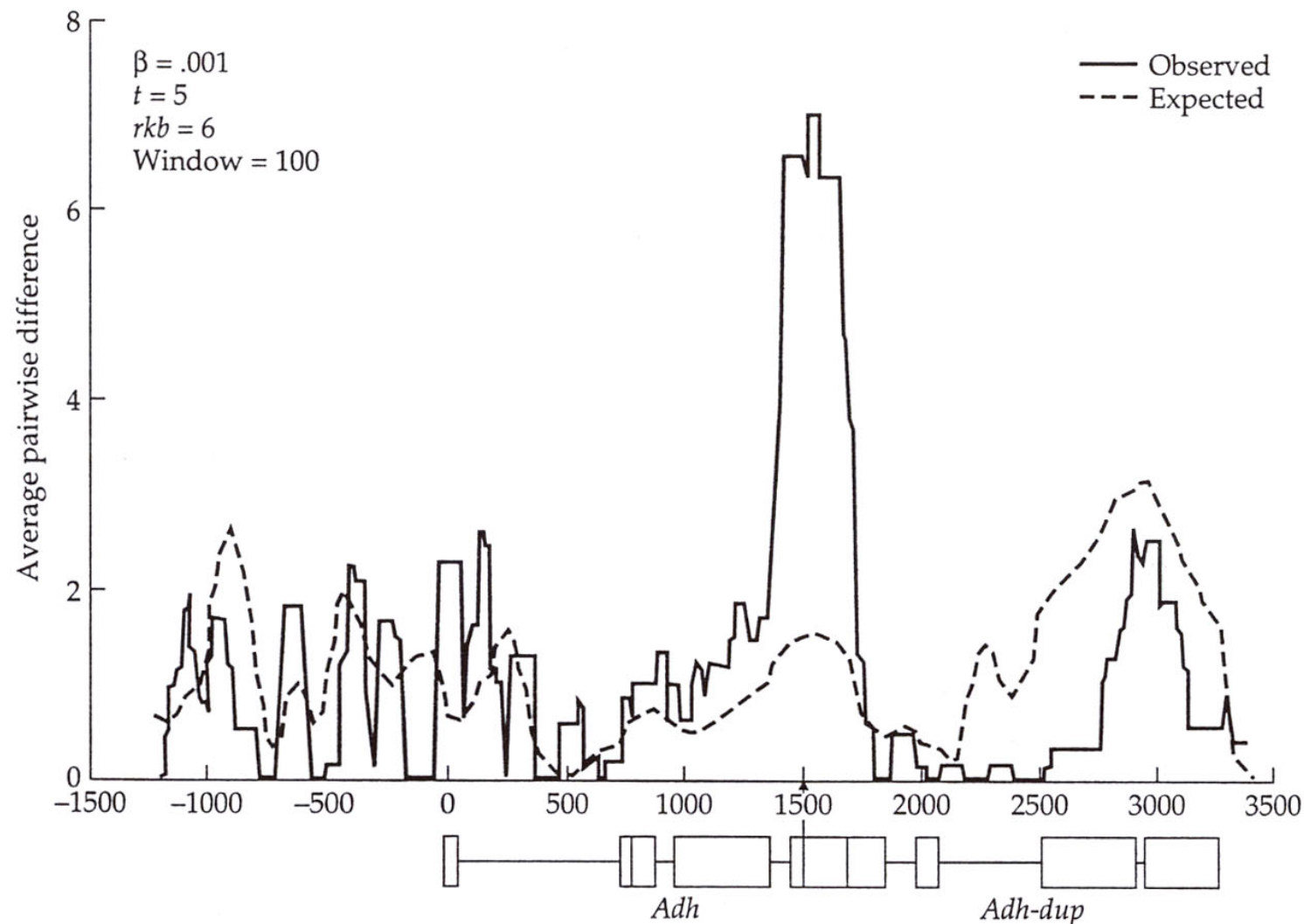


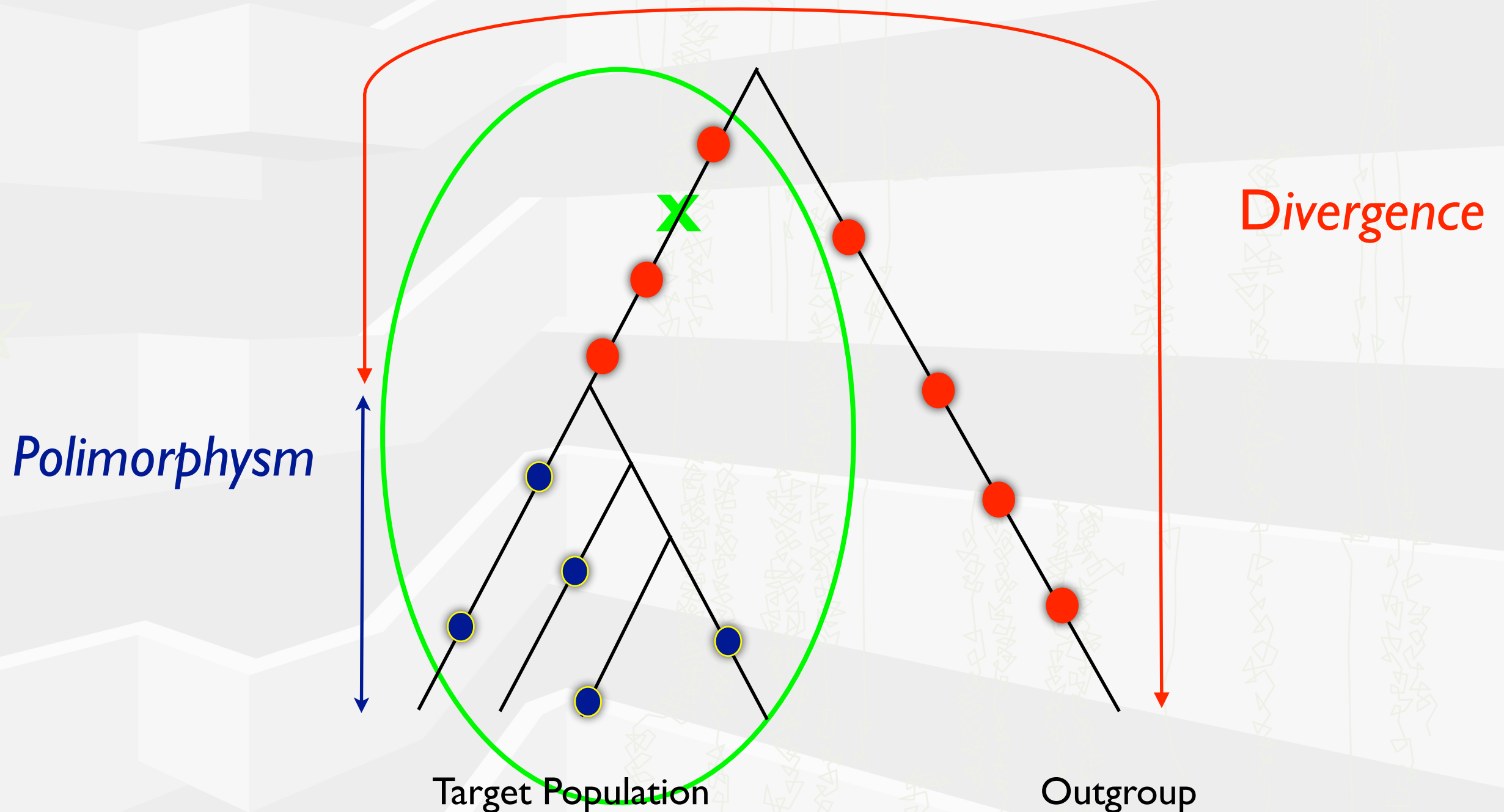
Figure 9.5 Sliding window of observed and expected polymorphism levels in a sample of 11 *Drosophila melanogaster* alleles. The average pairwise number of nucleotide differences in a window of 100 silent sites is plotted at each nucleotide position. The arrow at position 1490 marks the site of the *Adh^f* / *Adh^s* amino acid replacement polymorphism. See Kreitman and Hudson (1991) for more detailed information about the sliding-window method. From Kreitman and Hudson (1991).

Selection within population (occurred and/or continues)

- Some algorithms and software for selection detection at the genomic level:
 - ROH: Regions of homozygosity (e.g., with PLINK, Purcell et al. 2007)
 - EHH: Extended Haplotype Homozygosity. (Tang et al. 2007)
 - Genome scans: e.g., CLR Kim et al. 2002), Variscan (Vilella et al.)
 - Fst methods ex.PCADAPT (Luu et al 2017). Detects variants between populations with unexpected frequencies.
 - DnaSP (Rozas et al 2017): Multiple tests based on allele frequencies.
 - Other libraries based on multilocus analysis (R libraries, e.g., PopGenome, Pfeifer et al. 2014).
 - There are many other methods, for example based on population differentiation (Fst, Rsb...e.g., libraries in R like rehh, Gautier and Vitalis 2012)
- ...a review of methods in: Pavlidis and Alachiotis (J. of Biol. Res. 2017)

The methodologies vary depending on the time scale to be studied.

Selection in the species (adaptive mutations were fixed)



Selection in the species (adaptive mutations were fixed)

- Effect on patterns of variability: Differences in rates of functional versus neutral evolution. Differences between species.
- Methodologies for detection: Detection by comparison of the K_a/K_s ratio (non-synonymous vs. synonymous divergence) and the ratio of polymorphism (π_a/π_s , MKT). Detection of differential patterns on K/π among loci (HKA Test).
- Detection power: Quite powerful to detect adaptive events at genome level.