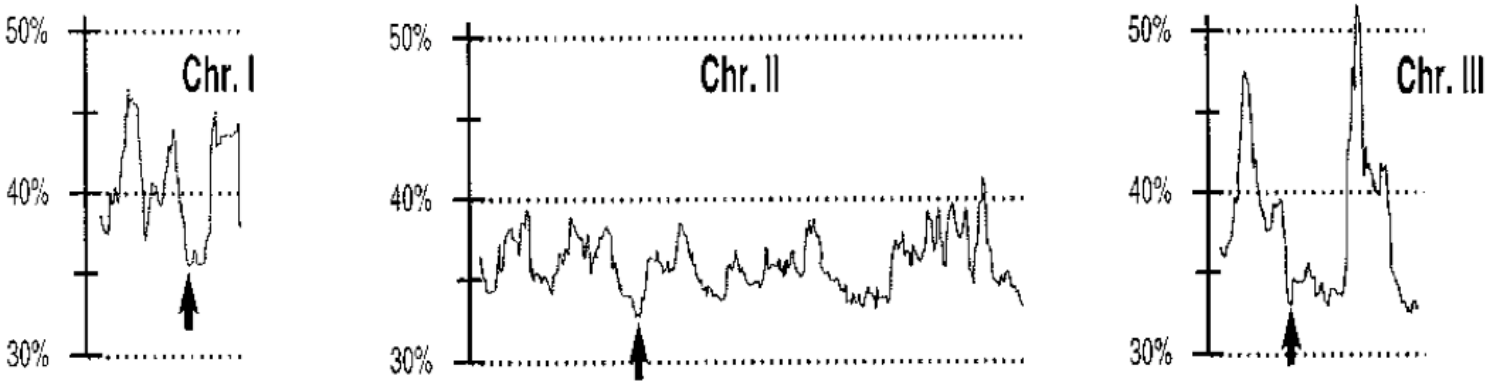
Models for the evolution of GC content in the asexual fungi *Candida albicans* and *Candida dubliniensis*

Marie-Claude Marsolier-Kergoat
(CEA/DSV/IBITECS/SBIGEM)

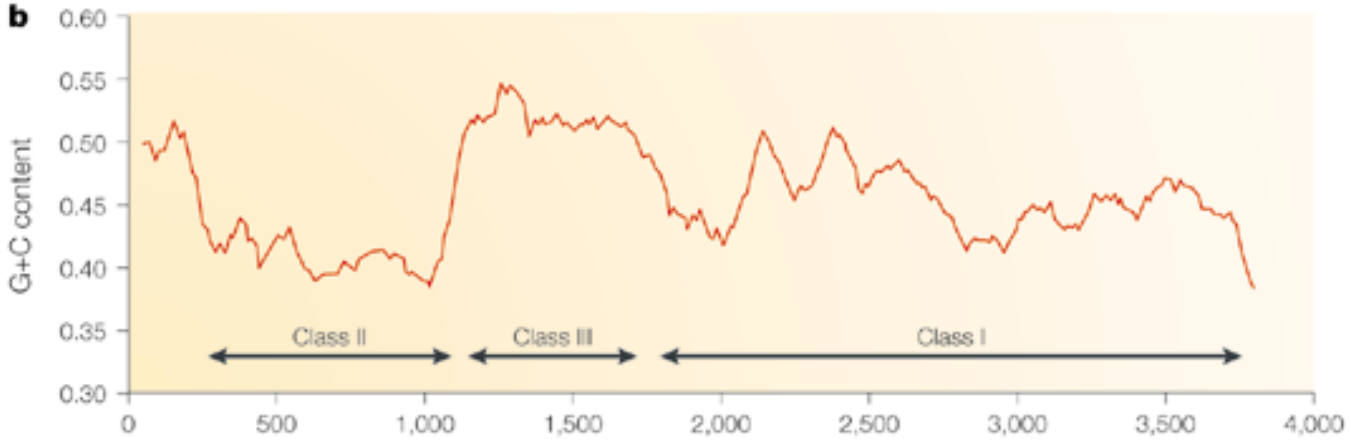
Marsolier-Kergoat MC (2013) Models for the evolution of GC content in the asexual fungi *Candida albicans* and *Candida dubliniensis*. Genome Biol. Evol. 5:2205-2216.

What determines genome-wide variations of GC content in eukaryotes ?



Bradnam et al., 1999

Variations in GC3 content along chromosomes I, II and III in *Saccharomyces cerevisiae*

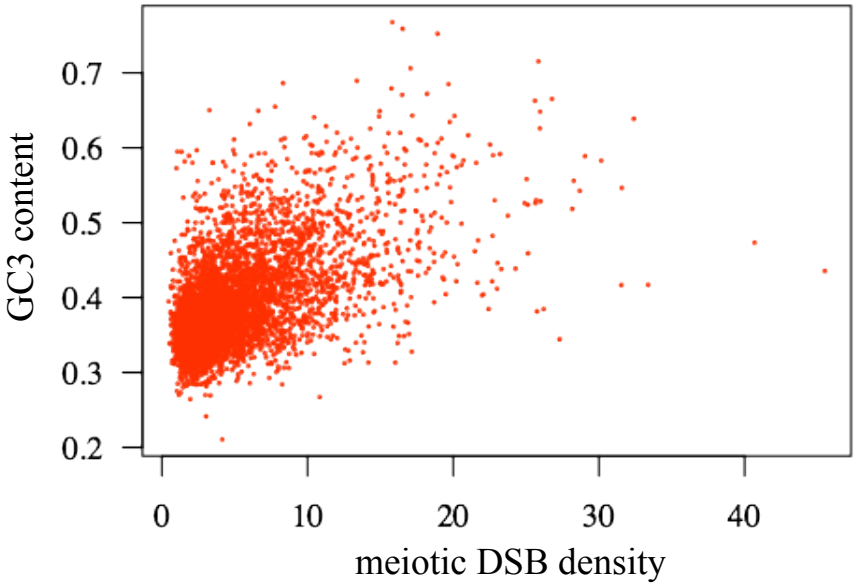


Eyre-Walker and Hurst., 2003

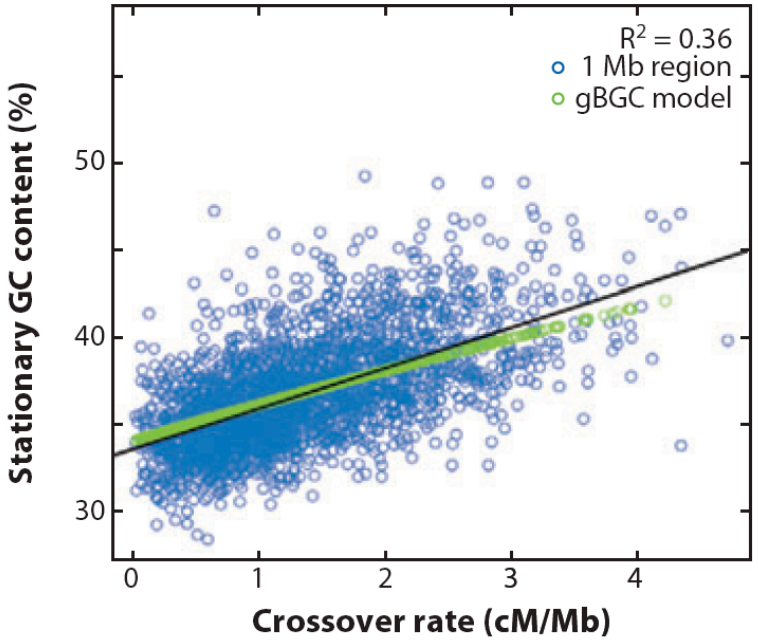
Variations in GC content along the MHC region of chromosome 6 in *H. sapiens*

What determines genome-wide variations of GC content in eukaryotes ?

Saccharomyces cerevisiae genes



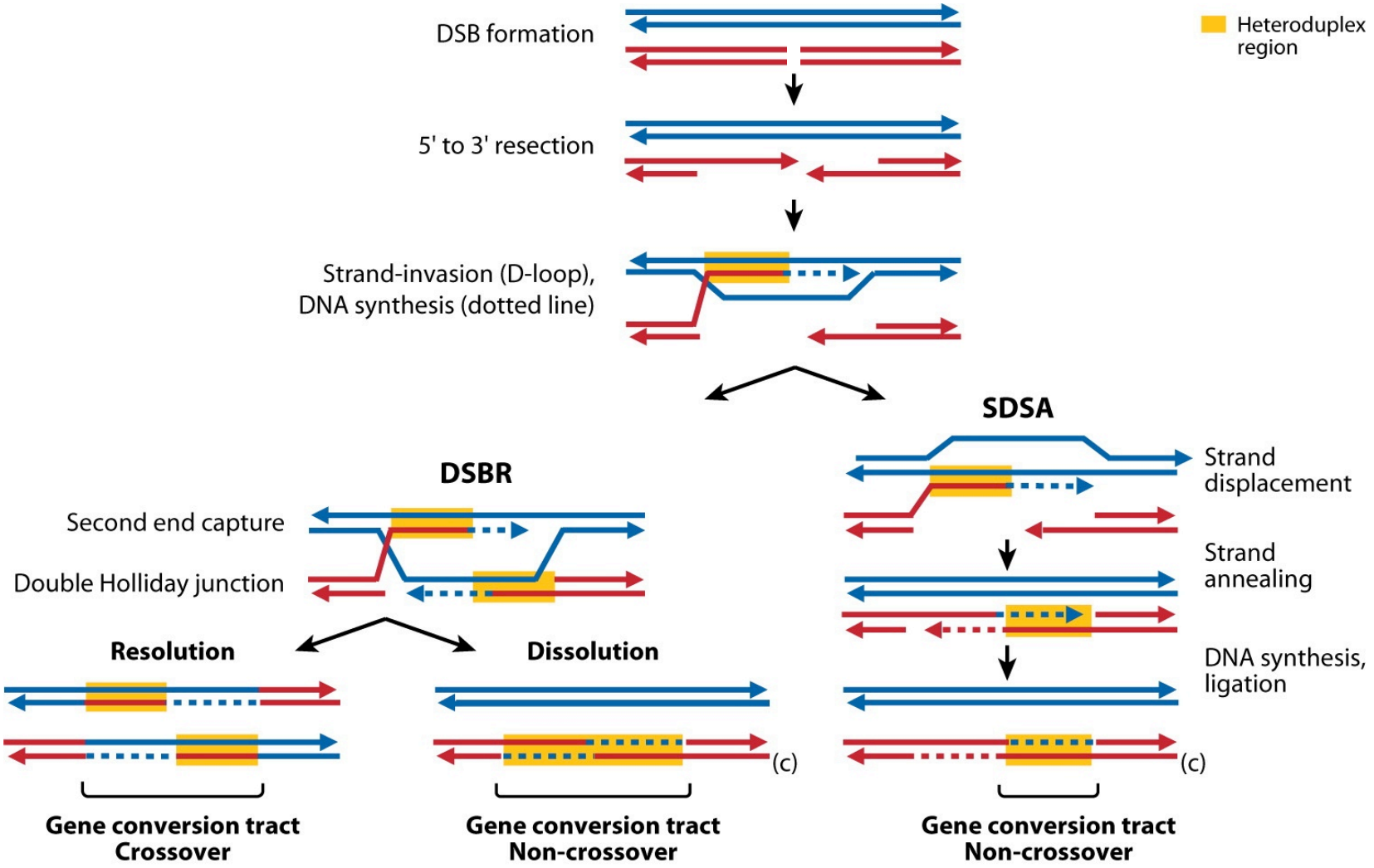
Human 1 Mb-long genomic regions



Duret and Galtier, 2009

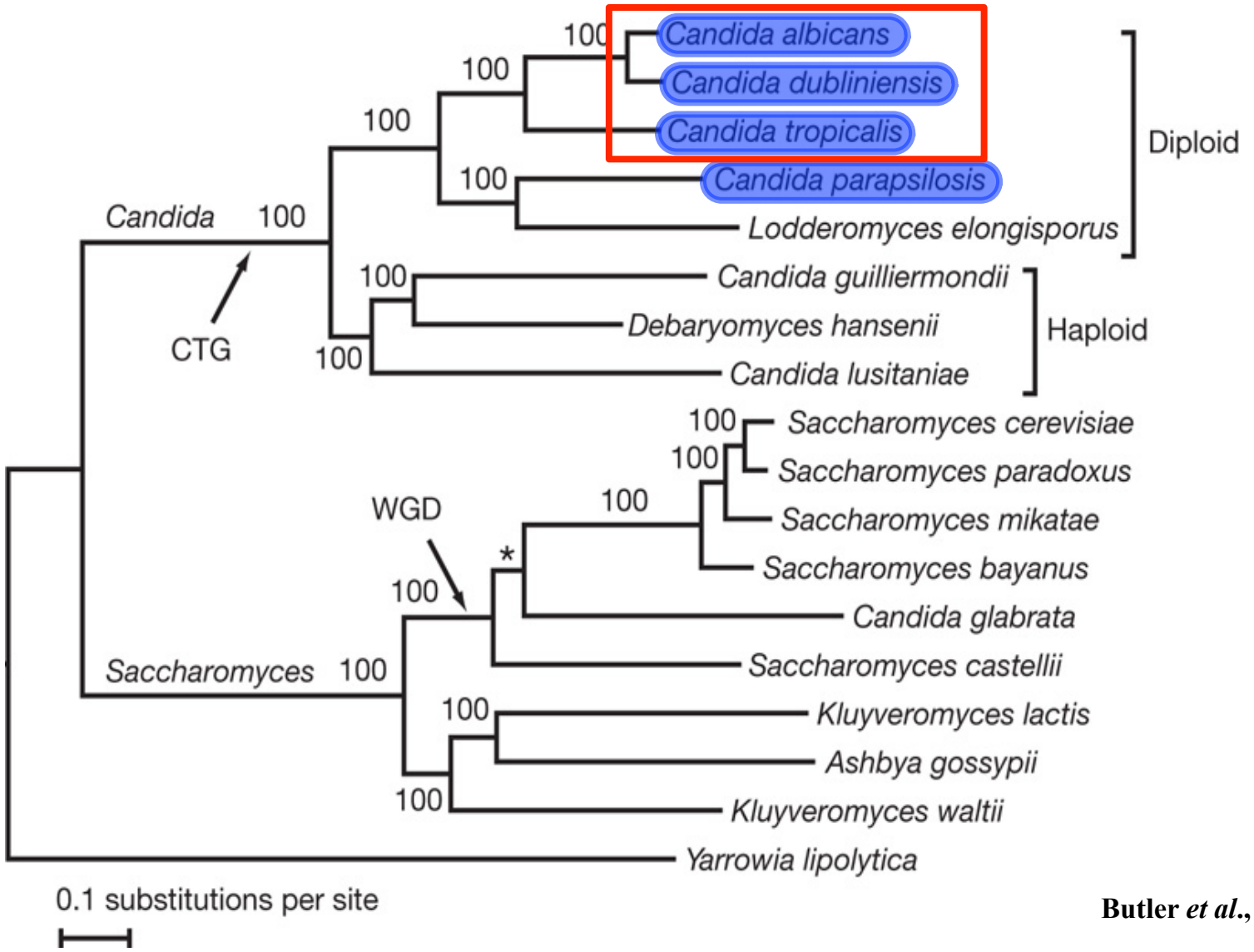
A strong correlation exists in many organisms between meiotic DSB density and GC content.

Mechanisms of GC-biased gene conversion



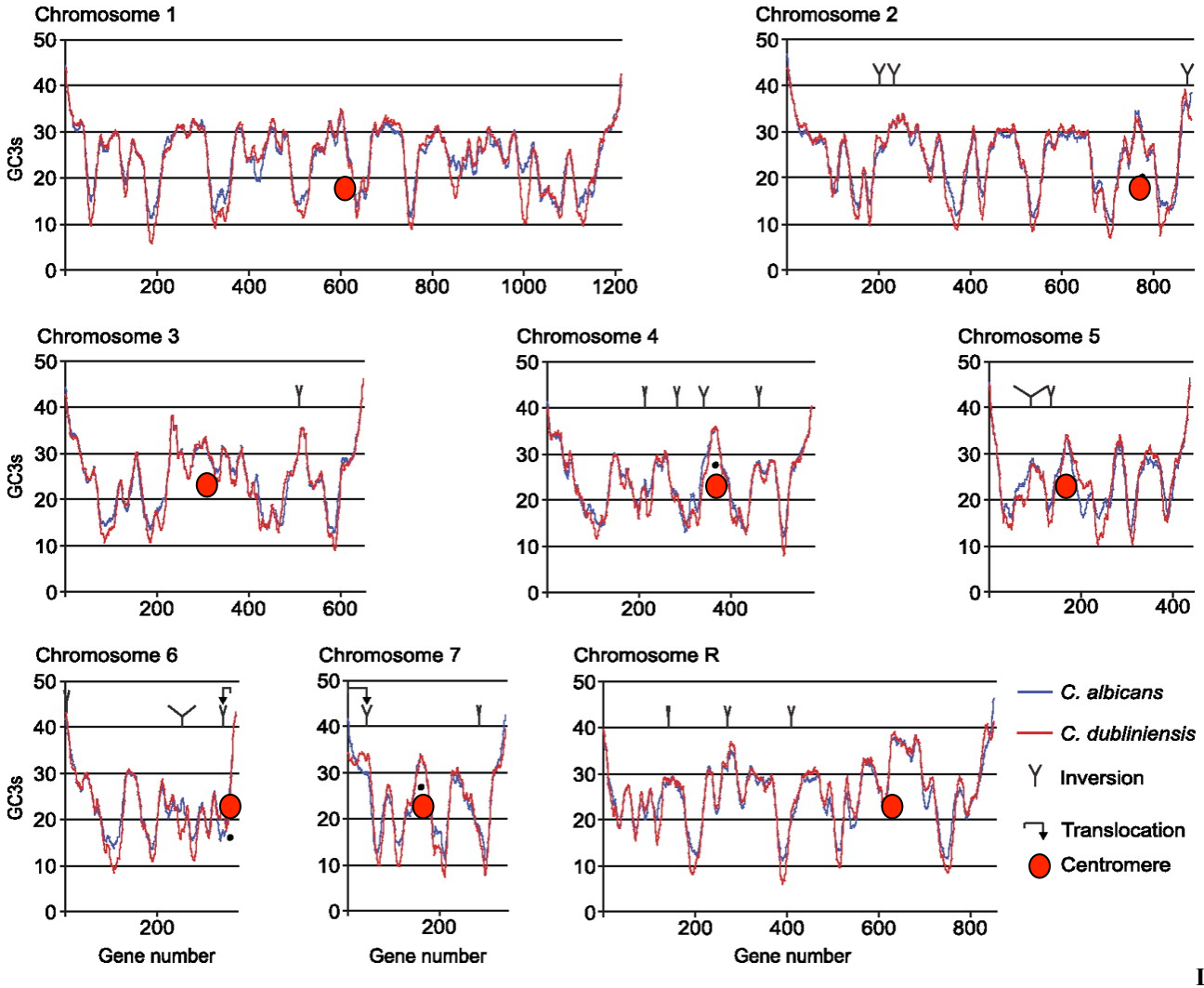
Duret and Galtier, 2009

Candida phylogeny: presence of asexual species



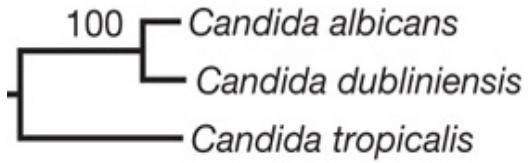
Butler *et al.*, 2009

Chromosomal GC3s profiles in *Candida albicans* and *C. dubliniensis*

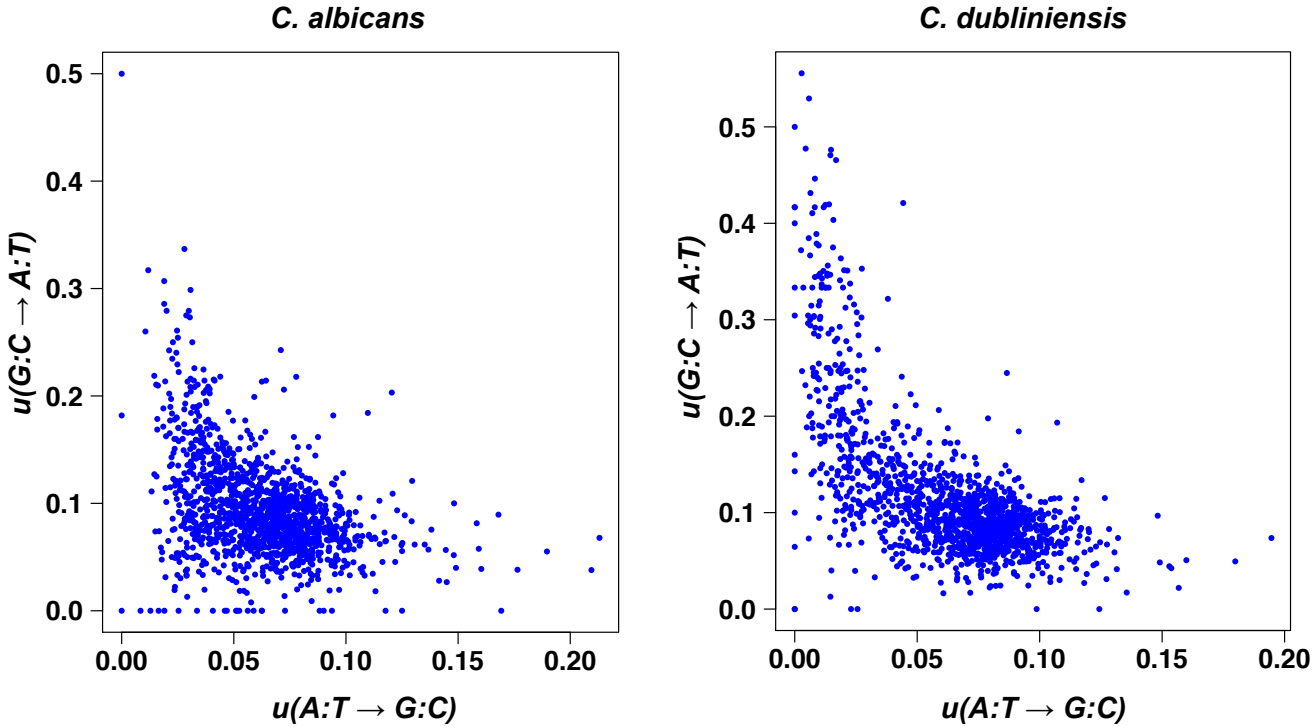


Lynch et al., 2010

Substitution patterns in *Candida*



Substitution patterns at third codon positions in the lineages of *Candida albicans* and *Candida dubliniensis*, using parsimony and *Candida tropicalis* as outgroup, reveal a strong anticorrelation between the transition rates *u(A:T > G:C)* and *u(G:C > A:T)* calculated on 1424 nonoverlapping 10-kb windows.

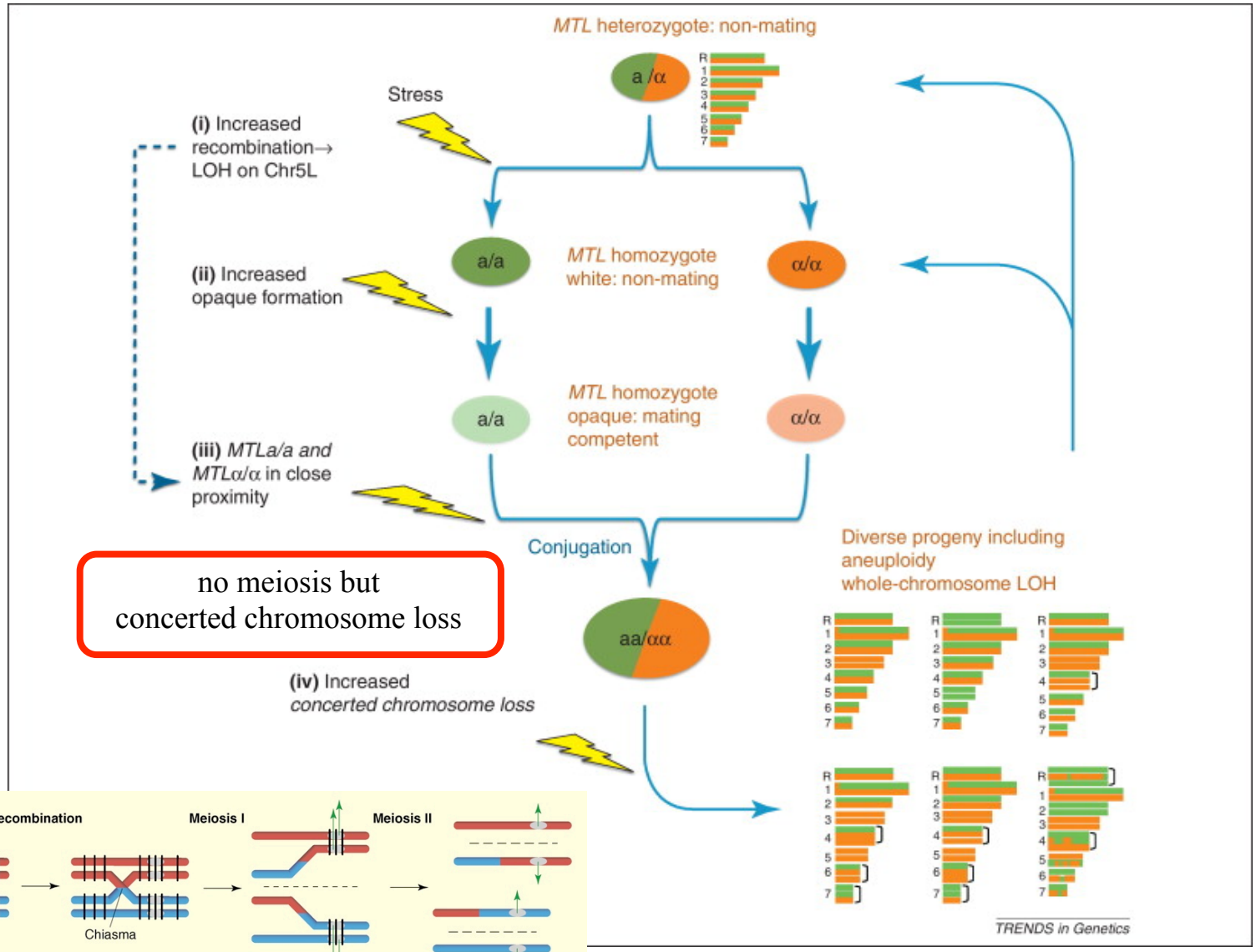


Two models

1. GC-biased gene conversion associated with mitotic recombination is responsible for variations in GC3 content: the higher the local recombination rate, the higher the GC content.

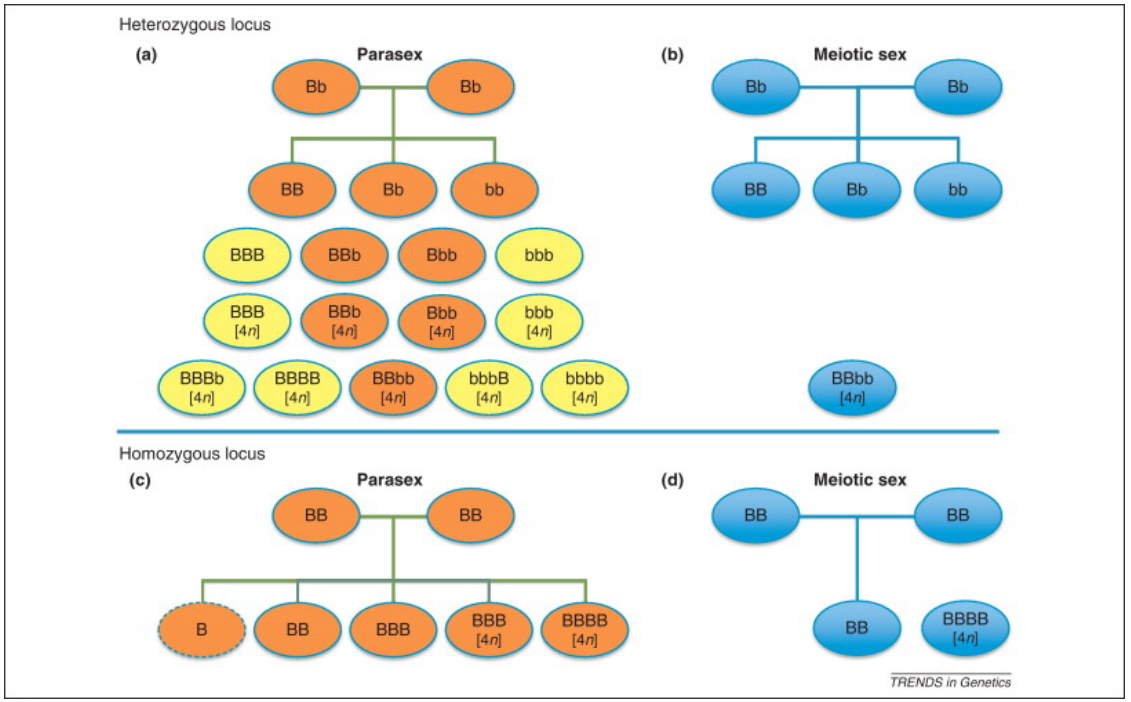
2. Replication errors are responsible for variations in GC3 content: the higher the concentration of dCTP and dGTP at the time when a sequence is replicated, the higher the GC content.

Candida albicans parasexual cycle



Berman and Hadany, 2009

Candida albicans parasexual cycle



- recombination events (LOH)
- aneuploidy

Berman and Hadany, 2009

The recombination model

Model developed by (Duret and Arndt, 2008):

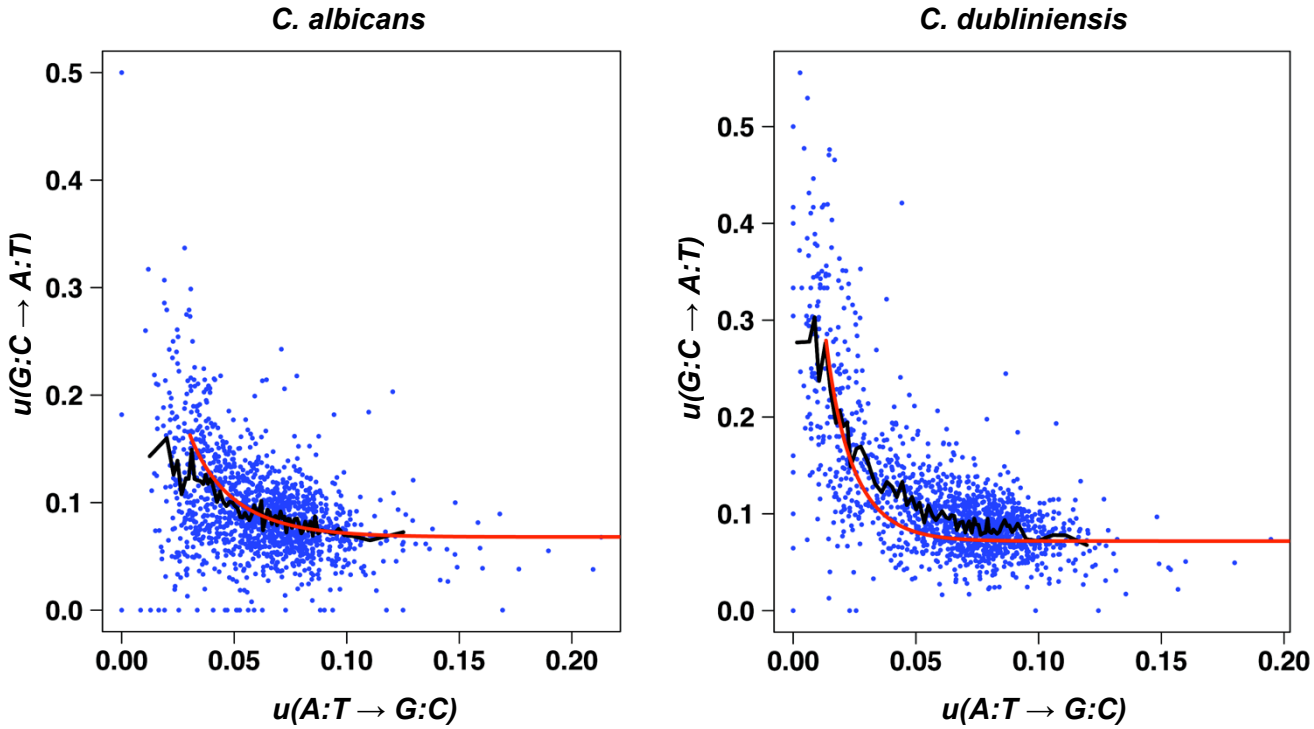
$$u(A{:}T \rightarrow G{:}C) = 2N(1-f)P(0)m(A{:}T \rightarrow G{:}C) + 2NfP(s)m(A{:}T \rightarrow G{:}C)$$
$$u(G{:}C \rightarrow A{:}T) = 2N(1-f)P(0)m(G{:}C \rightarrow A{:}T) + 2NfP(-s)m(G{:}C \rightarrow A{:}T)$$

u(A:T > G:C) = the substitution rate corresponding to the transitions from A:T to G:C

Constants:
m(A:T > G:C) = the mutation rate corresponding to the transitions from A:T to G:C
f = fraction of recombination hotspots in the genome
N = effective population size
P(s) = the probability that a mutation subject to GC-biased gene conversion of strength *s* will be fixed

$$u(A{:}T \rightarrow G{:}C) \approx (1-f)m(A{:}T \rightarrow G{:}C) + fm(A{:}T \rightarrow G{:}C)\frac{4Ns}{1-e^{-4Ns}}$$
$$u(G{:}C \rightarrow A{:}T) \approx (1-f)m(G{:}C \rightarrow A{:}T) - fm(G{:}C \rightarrow A{:}T)\frac{4Ns}{1-e^{4Ns}}$$

The recombination model



If the recombination model is correct, then the GC3 content depends on the mitotic recombination rate and reveals several features of mitotic recombination in *Candida* that differ from the usual characteristics of meiotic recombination:

1. The rate of mitotic recombination is not higher in smaller chromosomes
2. Mitotic recombination is not suppressed near centromeres
3. Mitotic recombination sites are not preferentially located in intergenes

The replication model

$$[E \bullet D_n] + dNTP \rightleftarrows [E \bullet D_n] \cdot dNTP \longrightarrow [E \bullet D_{n+1}] + PP_i$$

$$\Gamma(t_{rep}) = N_C + N_G = 1 - (N_A + N_T)$$

First model:

$$u(A{:}T \rightarrow G{:}C) \approx r_S \Gamma(t_{rep})$$
$$u(G{:}C \rightarrow A{:}T) \approx r_S [1 - \Gamma(t_{rep})]$$

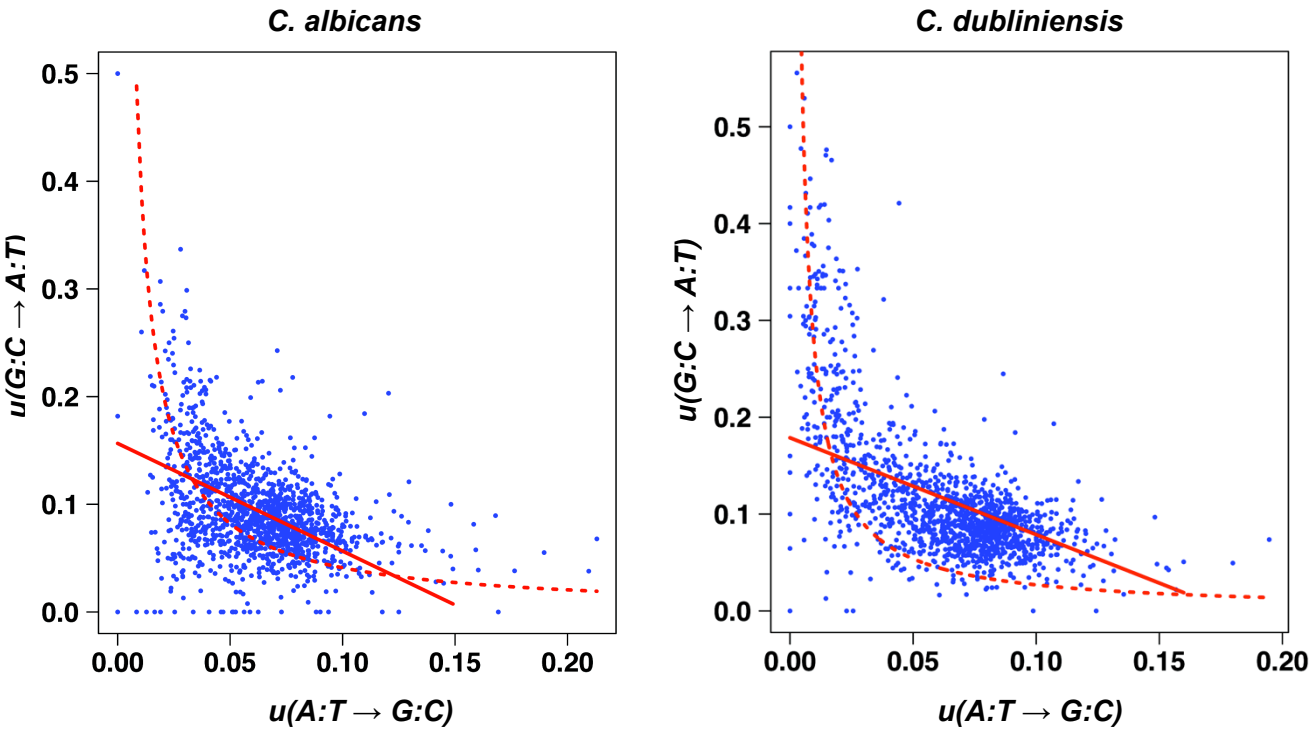
$$u(G{:}C \rightarrow A{:}T) \approx r_S - u(A{:}T \rightarrow G{:}C)$$

Second model:

$$u(A{:}T \rightarrow G{:}C) \approx 2p_{nc}\eta \frac{\Gamma(t_{rep})}{1 - \Gamma(t_{rep})}$$
$$u(G{:}C \rightarrow A{:}T) \approx 2p_{nc}\eta \frac{1 - \Gamma(t_{rep})}{\Gamma(t_{rep})}$$

$$u(A{:}T \rightarrow G{:}C) \approx \frac{(2p_{nc}\eta)^2}{u(G{:}C \rightarrow A{:}T)}$$

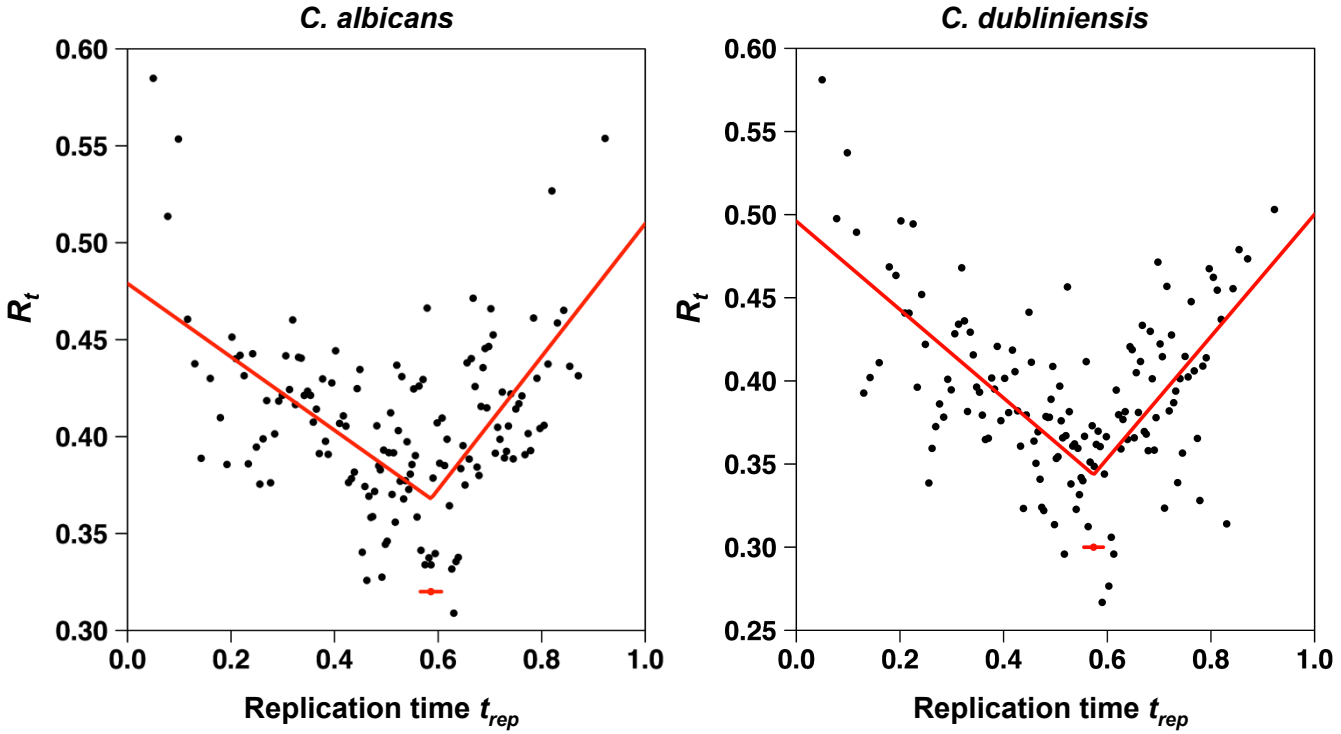
The replication model



The replication model

$$R_t = \frac{u(A{:}T \rightarrow G{:}C)}{u(A{:}T \rightarrow G{:}C) + u(G{:}C \rightarrow A{:}T)} \approx \frac{r_S \Gamma(t_{rep})}{r_S \Gamma(t_{rep}) + r_S[1 - \Gamma(t_{rep})]} = \Gamma(t_{rep})$$

Temporal variations of $\Gamma(t_{rep})$ translate into spatial variations of substitution rates, and, reciprocally, the variations of $\Gamma(t_{rep})$ can be inferred from the analysis of local substitution rates, knowing the current program of DNA replication.



Conclusions

- ✓ Two mutually non-exclusive models can be proposed to account for the variations of GC content in asexual organisms and in particular in *Candida* species.
- ✓ The validity of these models could be tested by the determination of either mitotic recombination rates or the proportions of dCTPs and dGTPs during the S phase.
- ✓ The models could also apply to sexual organisms along with GC-biased gene conversion linked to meiotic recombination.