

Insertions and deletions in the great tit genome

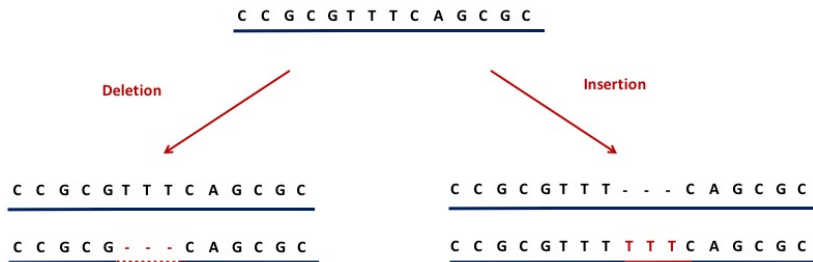
Henry Barton and Kai Zeng, University of Sheffield

PopGroup 50 - 06/01/17

Introduction

Insertions and deletions

- ▶ short INDELs: sections of DNA $< 50\text{bp}$ that are deleted or inserted in a genome



INDELs often overlooked

- ▶ Disproportionately occur in repetitive sequence
- ▶ Hard to align
- ▶ Often occur in hotspots
- ▶ 1/8 as frequent as SNPs in humans

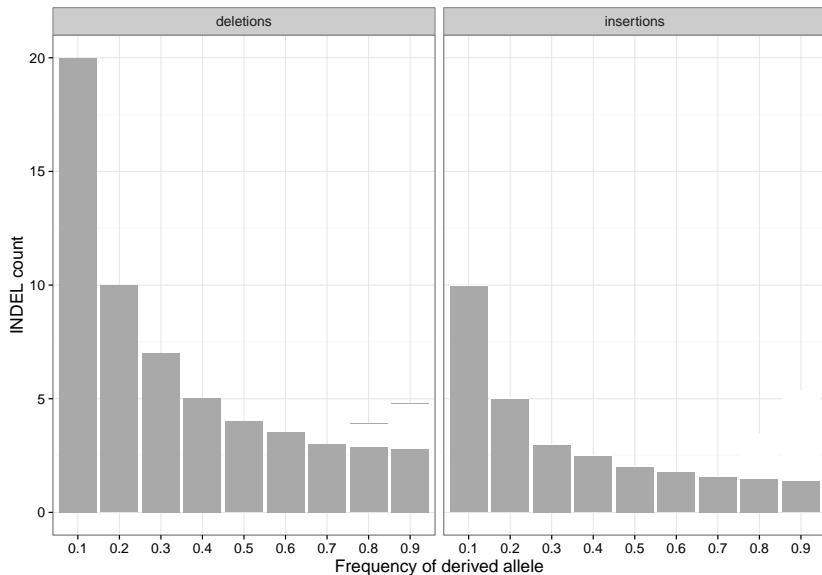
(Earl et al., 2014; Montgomery et al., 2013)

The importance of INDELs in genome evolution

- ▶ Influence genome size:
 - ▶ low deletion rate → large genomes?
 - ▶ high deletion rate → compact genomes?
- ▶ Insertions may be favoured:
 - ▶ biased gene conversion
 - ▶ minimum intron size
 - ▶ polarisation error

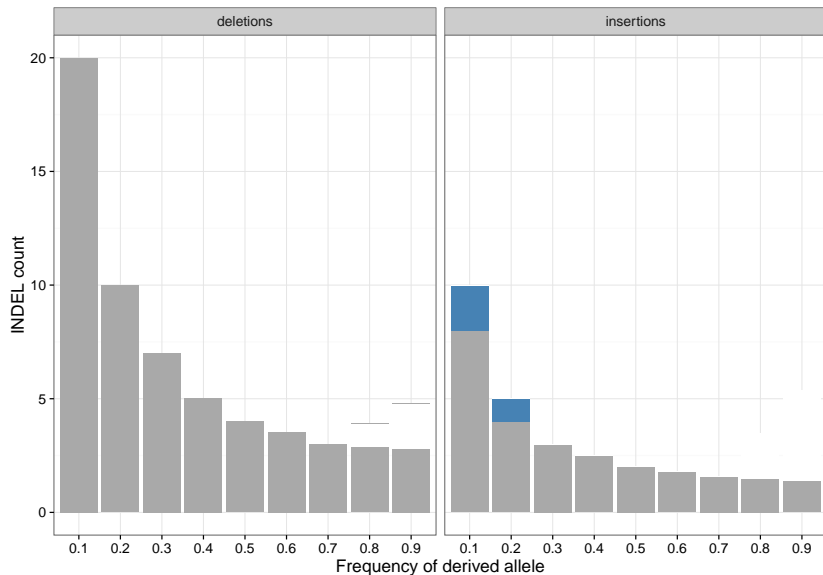
(Leushkin and Bazykin, 2013; Nam and Ellegren, 2012; Ometto et al., 2005; Sun et al., 2012)

Importance of polarisation error



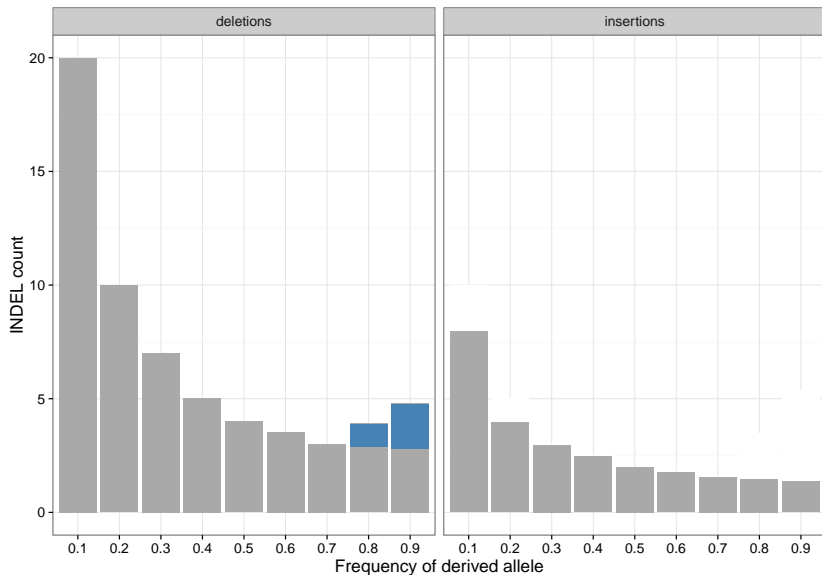
(see Hernandez et al., 2007)

Importance of polarisation error



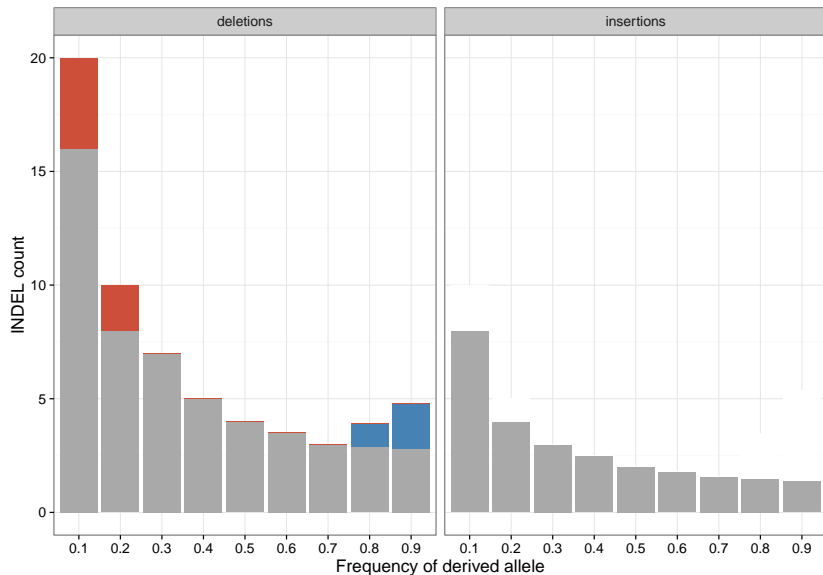
(see Hernandez et al., 2007)

Importance of polarisation error



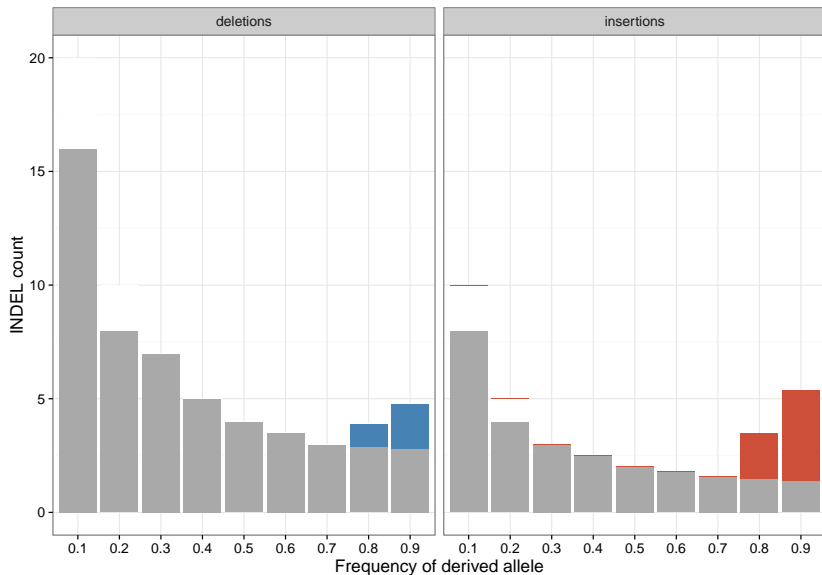
(see Hernandez et al., 2007)

Importance of polarisation error



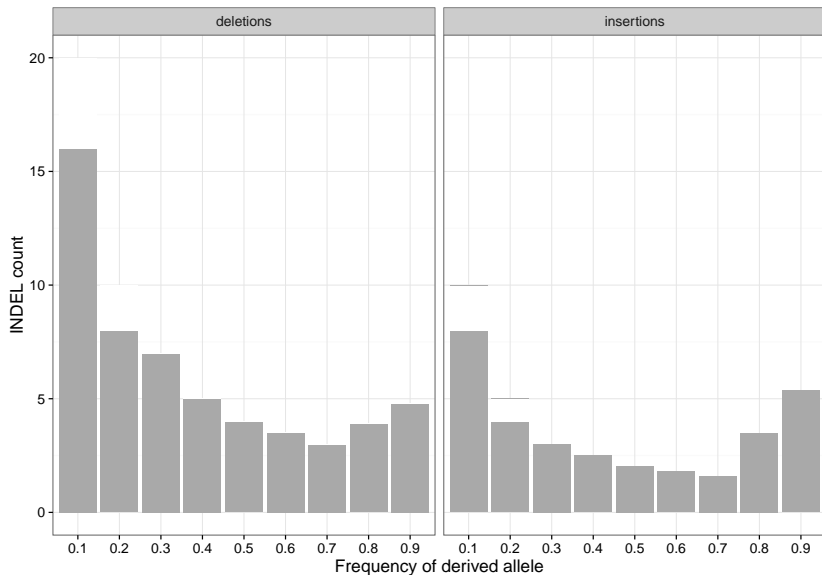
(see Hernandez et al., 2007)

Importance of polarisation error



(see Hernandez et al., 2007)

Importance of polarisation error



(see Hernandez et al., 2007)

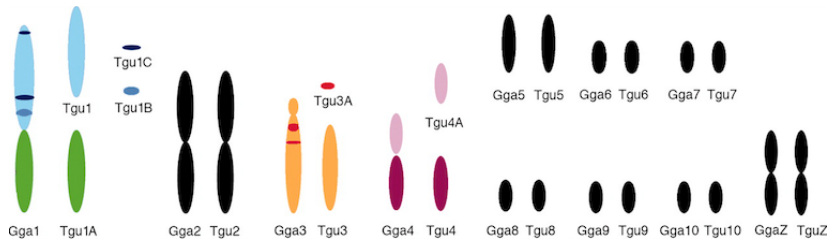
Aims

1. Overcome polarisation error using a novel modelling approach
2. Quantify the selective and mutational pressures acting on INDELs in great tits (*Parus major*)
3. Investigate how these pressures vary in different genomic contexts, ie chromosome length



Advantages of an avian system

- ▶ Highly dynamic recombination landscape
- ▶ Genomes consist of few large macrochromosomes and many small microchromosomes



(van Oers et al., 2014; Stapley et al., 2008)

Approach - data

Sample and pipeline

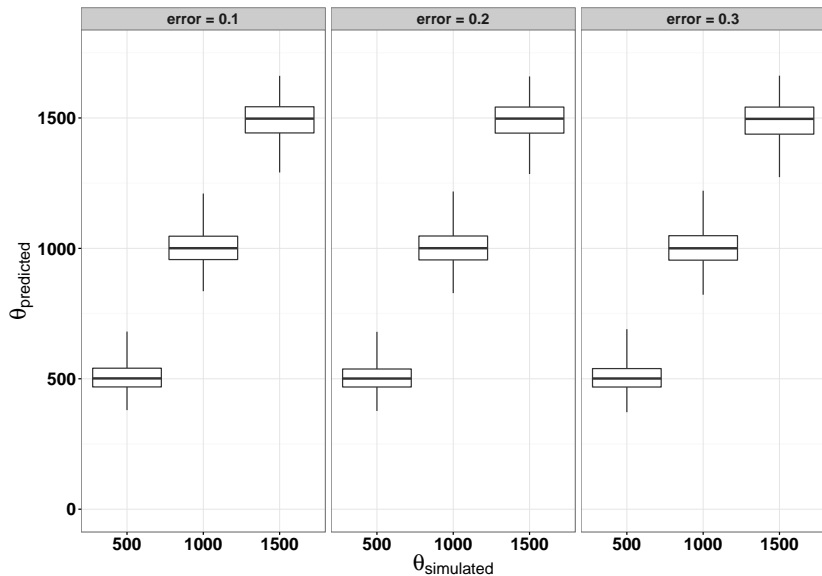
- ▶ 10 european great tit males
- ▶ high coverage (44x)
- ▶ variant calling with GATK
- ▶ multispecies alignment between zebra finch, flycatcher and great tit
- ▶ parsimony based polarisation

Approach - model

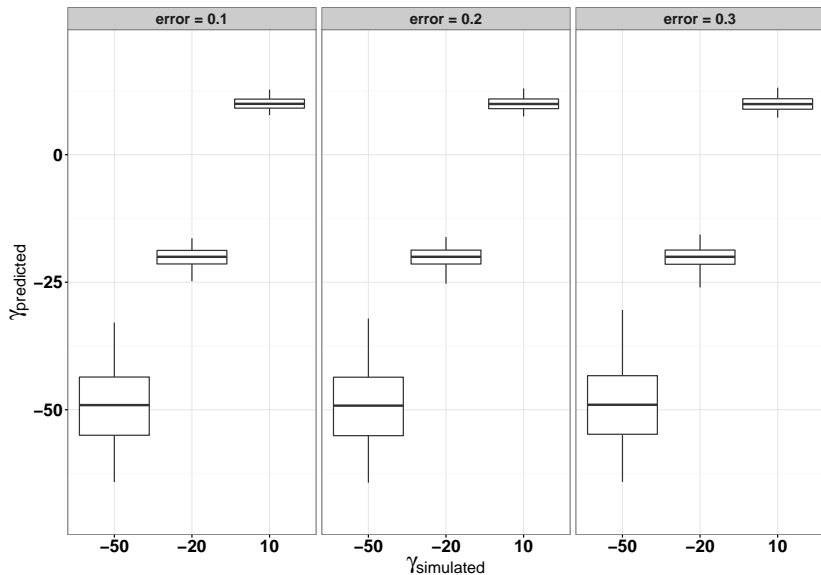
A novel maximum likelihood approach

- ▶ Adapting Glémin et al. (2015) model to INDELs
- ▶ estimates from the unfolded site frequency spectrum:
 - ▶ insertion rate ($\theta_i = 4N_e\mu_i$)
 - ▶ deletion rate ($\theta_d = 4N_e\mu_d$)
 - ▶ insertion selection coefficient ($\gamma_i = 4N_es_i$)
 - ▶ deletion selection coefficient ($\gamma_d = 4N_es_d$)
 - ▶ deletion bias ($\kappa = \frac{\mu_d}{\mu_i}$)
- ▶ Controls for demography using neutral SNPs (Eyre-Walker et al., 2006)
- ▶ Estimates and accounts for polarisation error

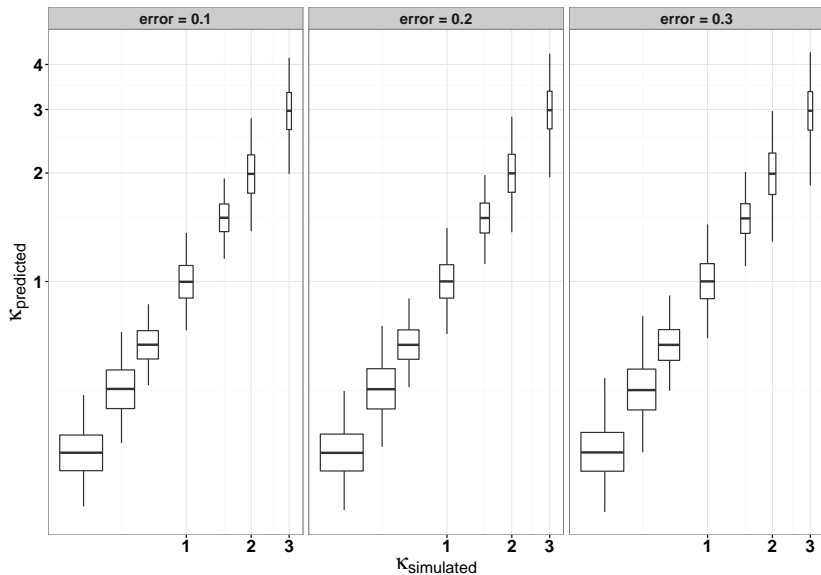
Predicts θ (mutation rate) well



Predicts γ (selection coefficient) well

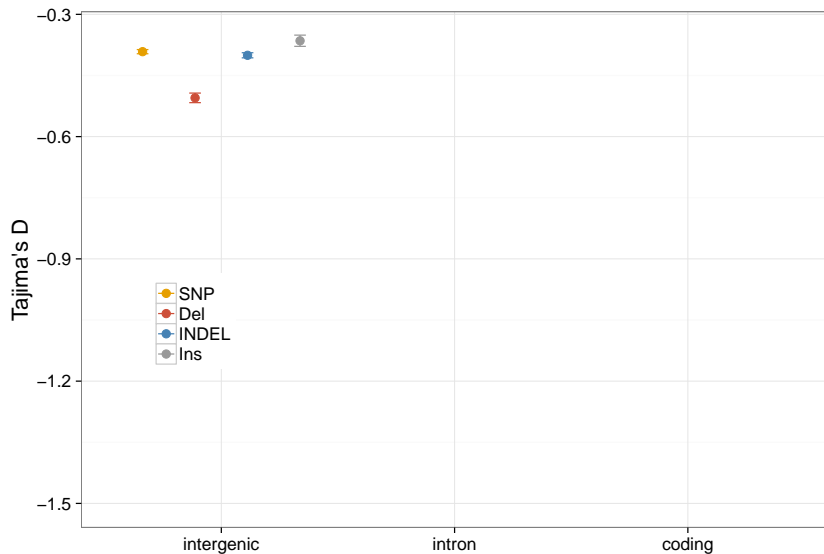


Predicts κ (deletion bias) well

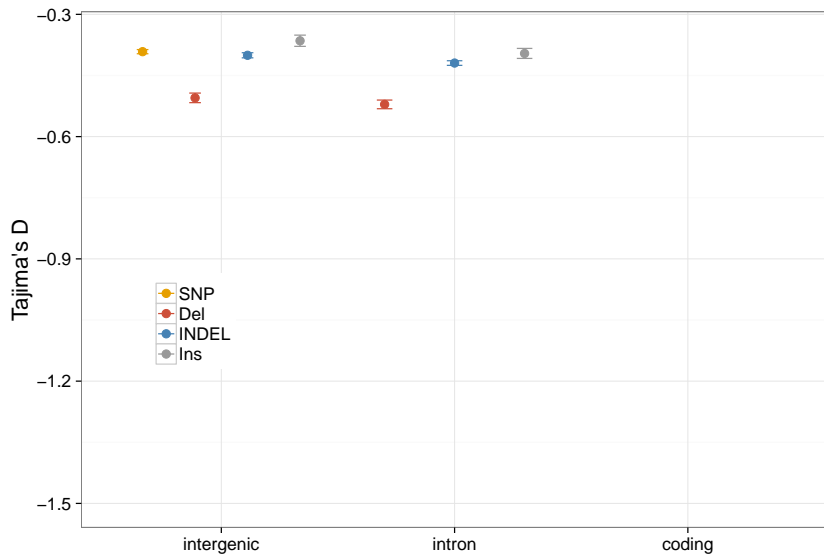


Regional Results

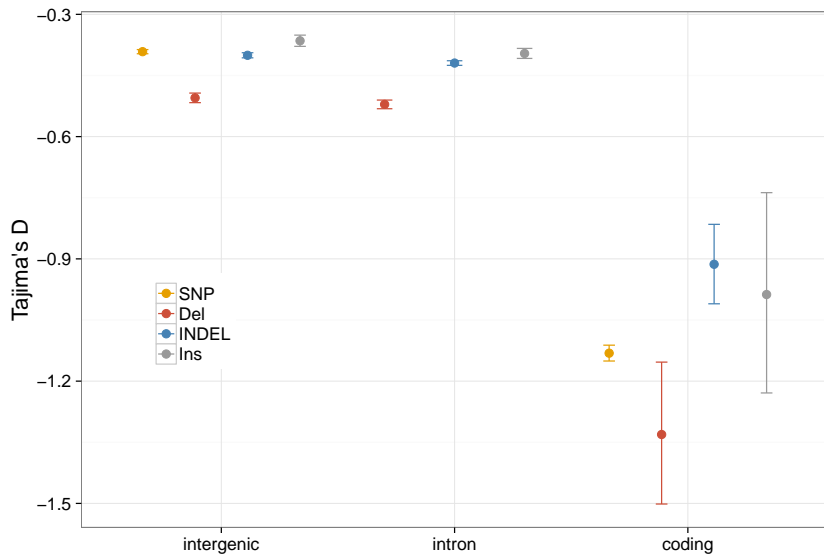
Regional variation in purifying selection



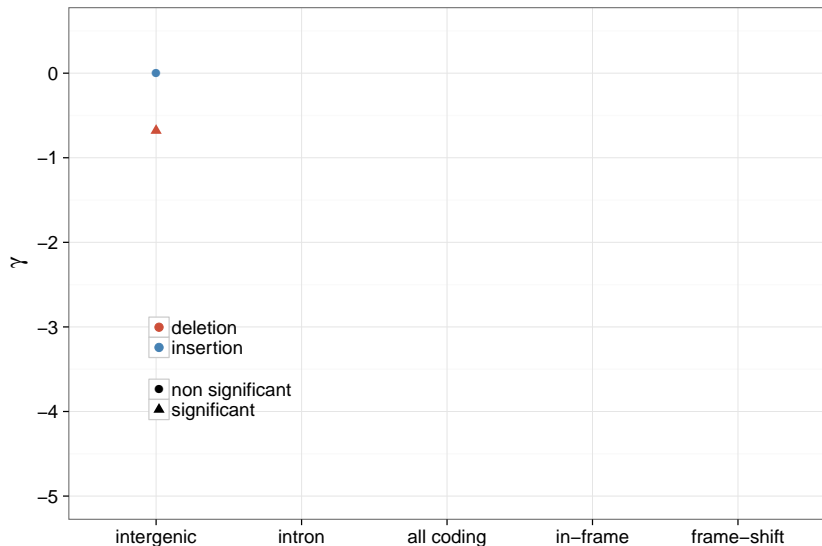
Regional variation in purifying selection



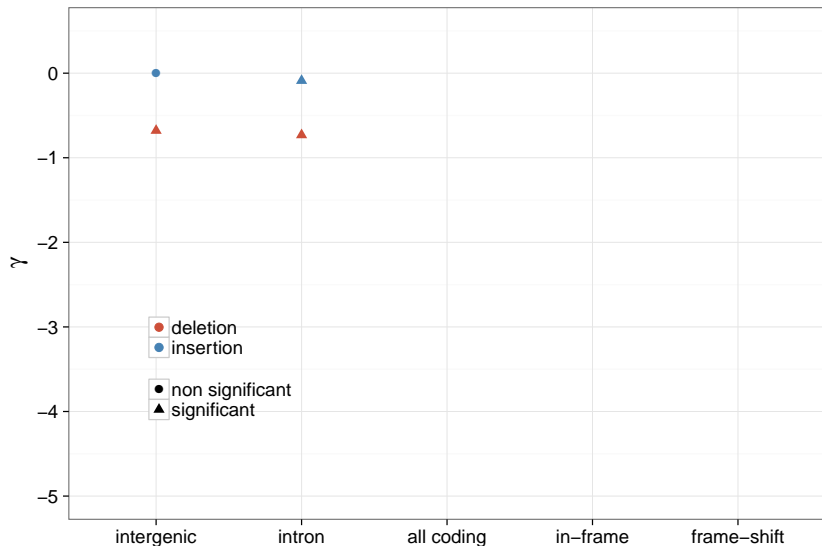
Regional variation in purifying selection



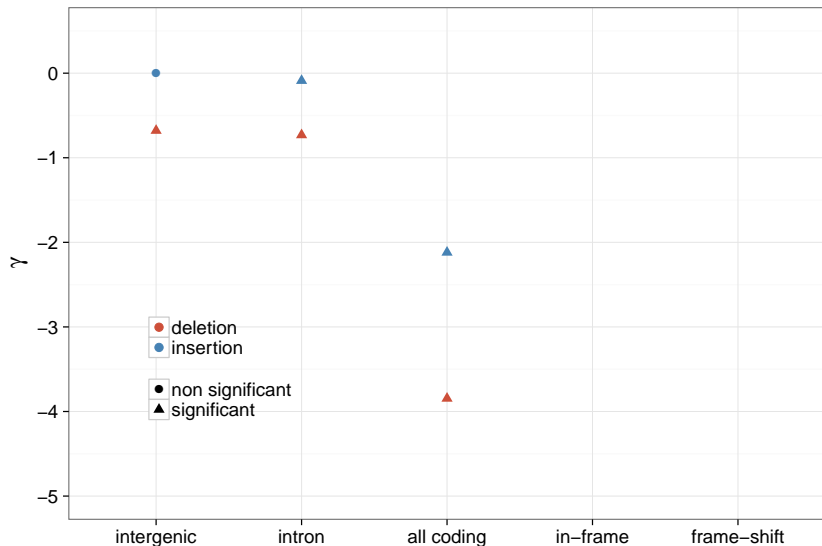
INDELs deleterious in coding regions



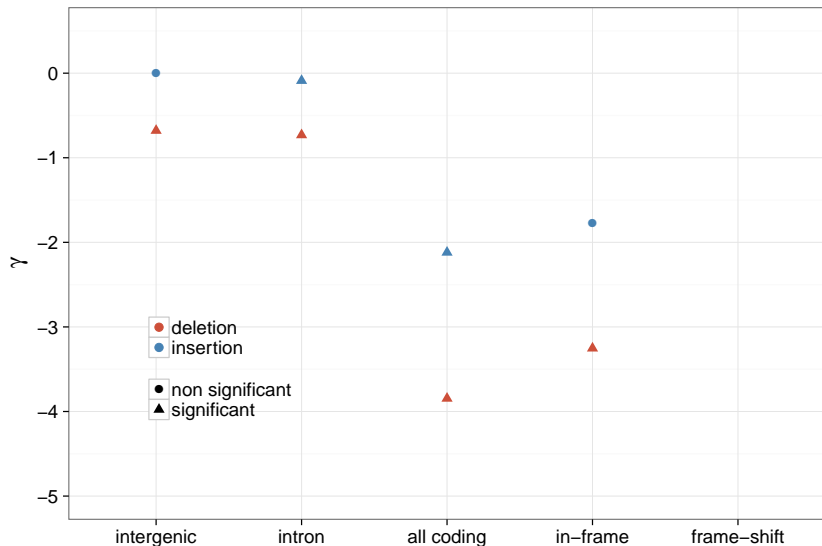
INDELs deleterious in coding regions



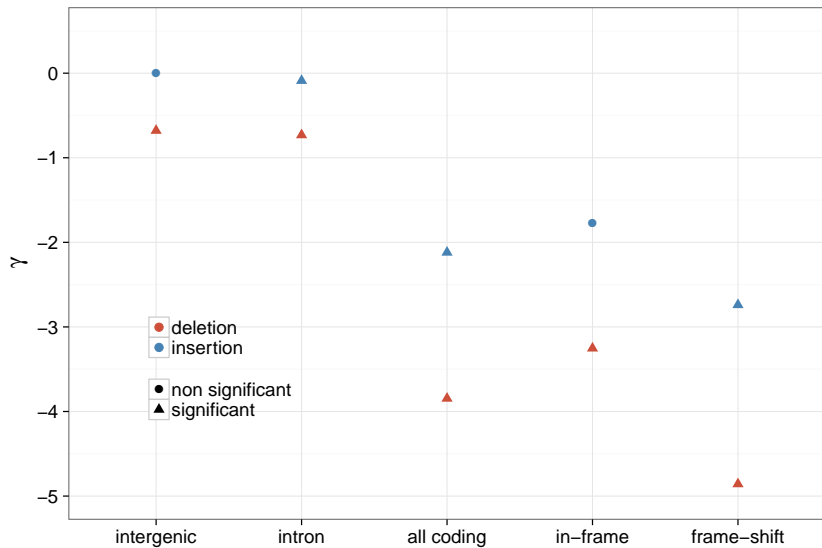
INDELs deleterious in coding regions



INDELs deleterious in coding regions

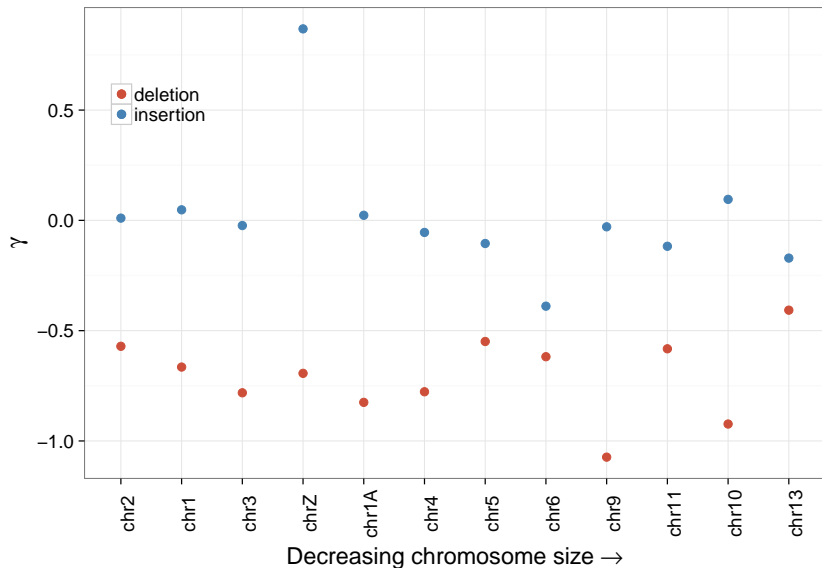


INDELs deleterious in coding regions

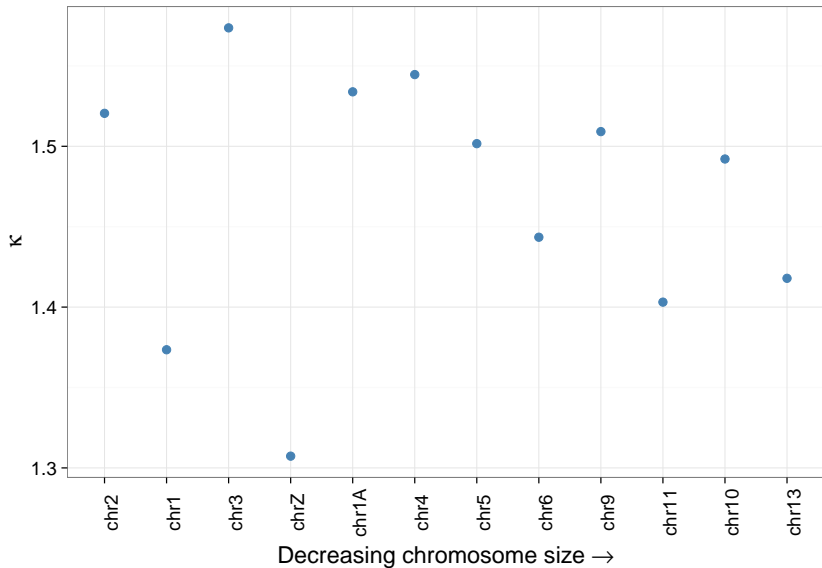


Recombination Results

No relationship between γ (selection coefficient) and chromosome length



No relationship between κ (deletion bias) and chromosome length



Concluding remarks

- ▶ INDELs in intergenic and intronic regions largely neutral
- ▶ Coding INDELs strongly deleterious
- ▶ No evidence for neutral role of deletion rates in chromosome size reduction
- ▶ Interesting in future to include:
 - ▶ divergence data
 - ▶ distribution of fitness effects
- ▶ model packaged in the form of 'anavar'

Acknowledgements

- ▶ Kai Zeng
- ▶ Pádraic Corcoran
- ▶ Toni Gossmann