

Paleogenomes analysis project

Student: Peter Leontev

Scientific advisor: Tatiana Tatarinova, USC

Bees problem

- honey bees populations number may decline rapidly;
- Varroa mites represent one of the main bee's loss causes.

The main issue

Bees

Mites infestation

```
graph TD; A[Bees] --> C[Are there significant changes within the population gene pool?]; B[Mites infestation] --> C;
```

Are there significant changes
within the population gene pool?

What can be done via bioinformatics
given “old” and currently living bees
genomes?

Project goal

Try to recognize an error in ancient DNA using the information about methylation, SNP C->T and GC content

Data

- 32 ancient and 32 modern whole bee's genomes;
- Poor quality of ancient genomes.

Reads classification

- ~34% of unmapped reads for old genomes;
- ~70% of unmapped reads belong to bacteria;
- Reads classification is an ambiguous procedure.

Getting error rates distribution

- Find out the number of SNPs C->T for different genomic regions;
- Relate the number of SNP C->T with methylation;
- Is there any relationship between the number of SNP C->T, methylation ratio and GC content?

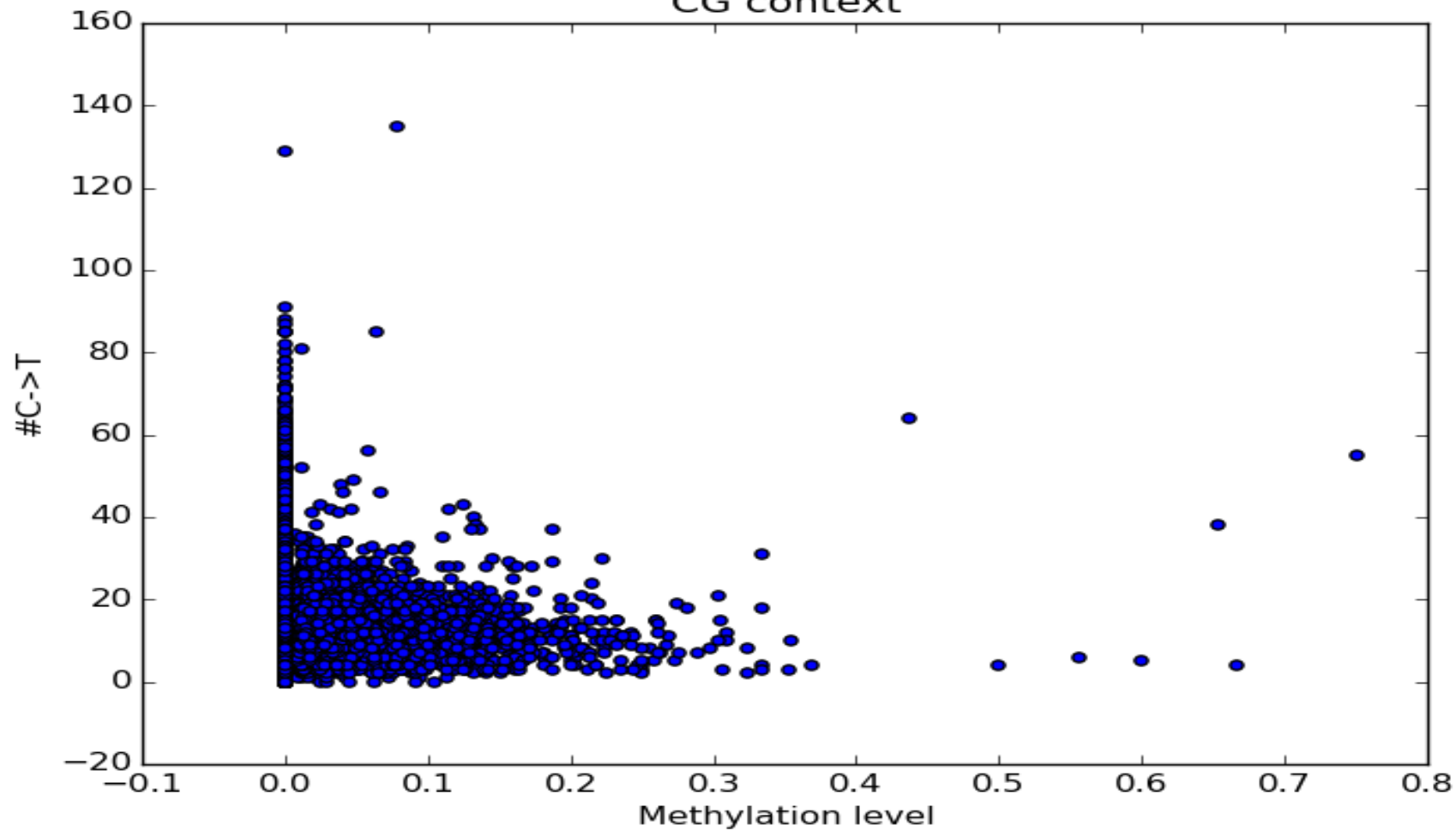
#C->T for intron-less genes

Promoter	exon	intron	UTR
9553	4806	-	44

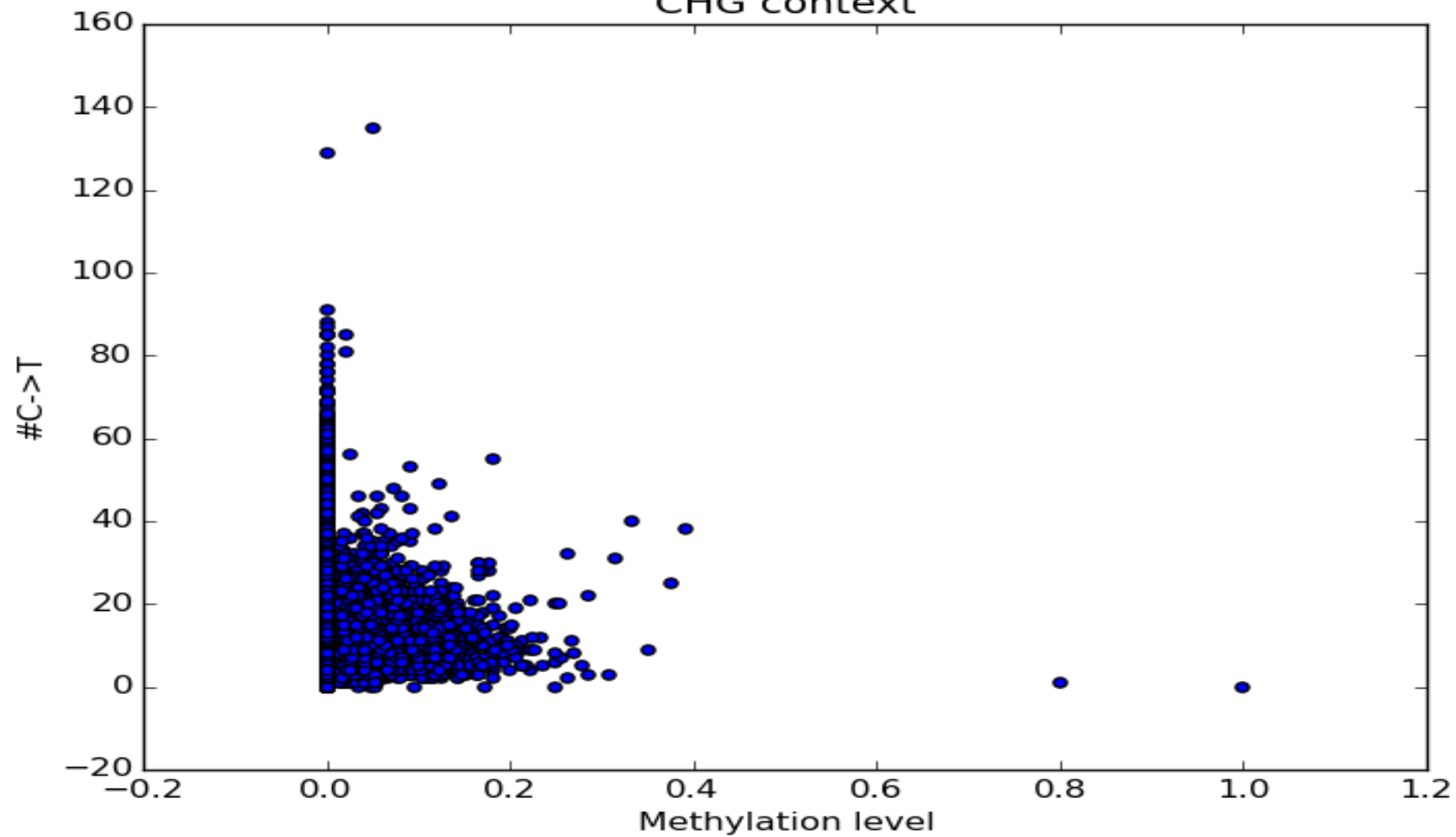
#C->T for intron-containing genes

Promoter	exon	intron	UTR
90152	163156	934751	53606

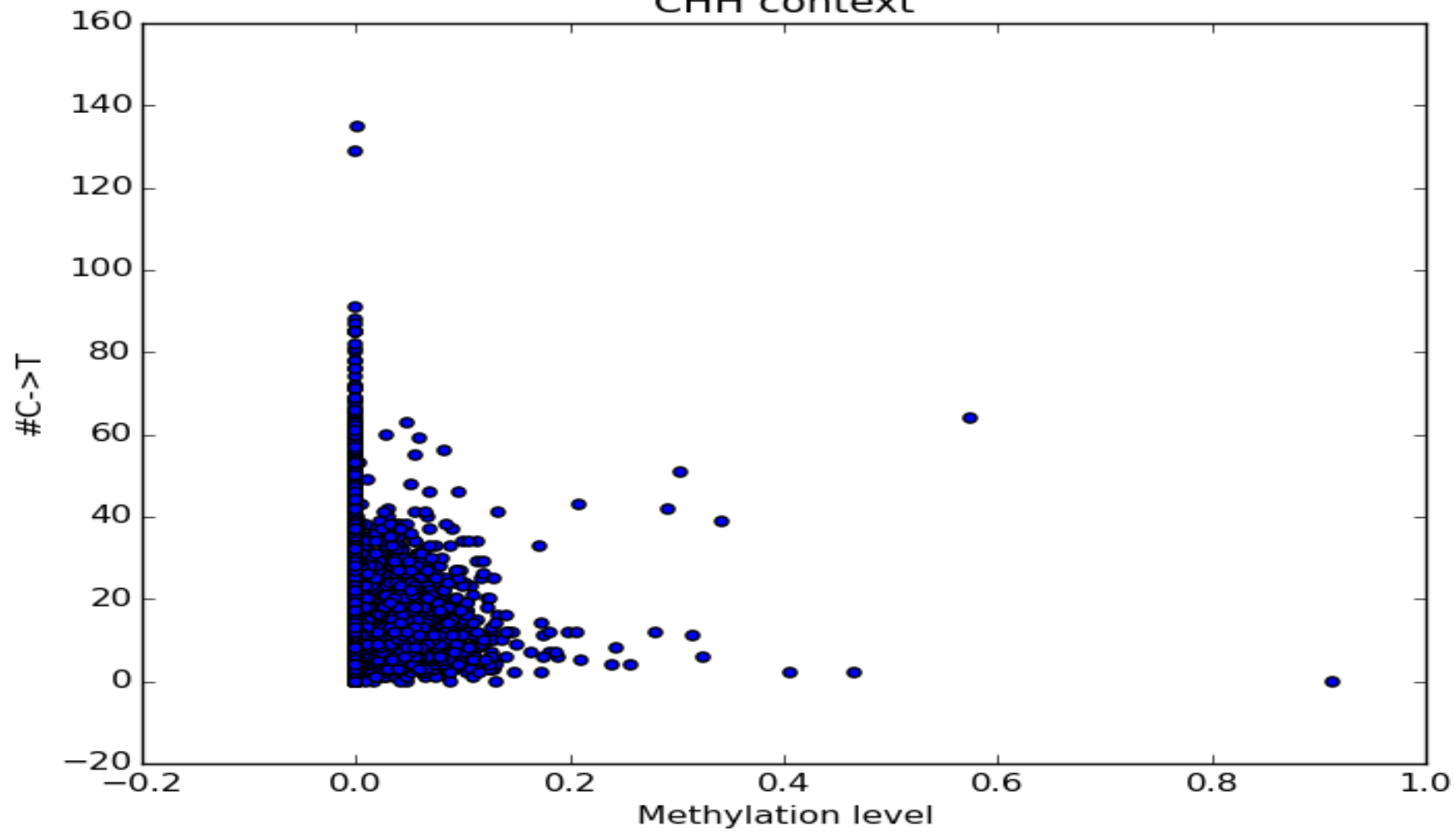
CG context



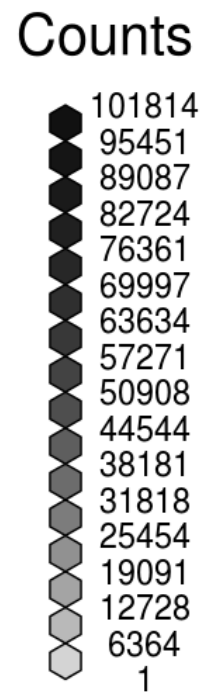
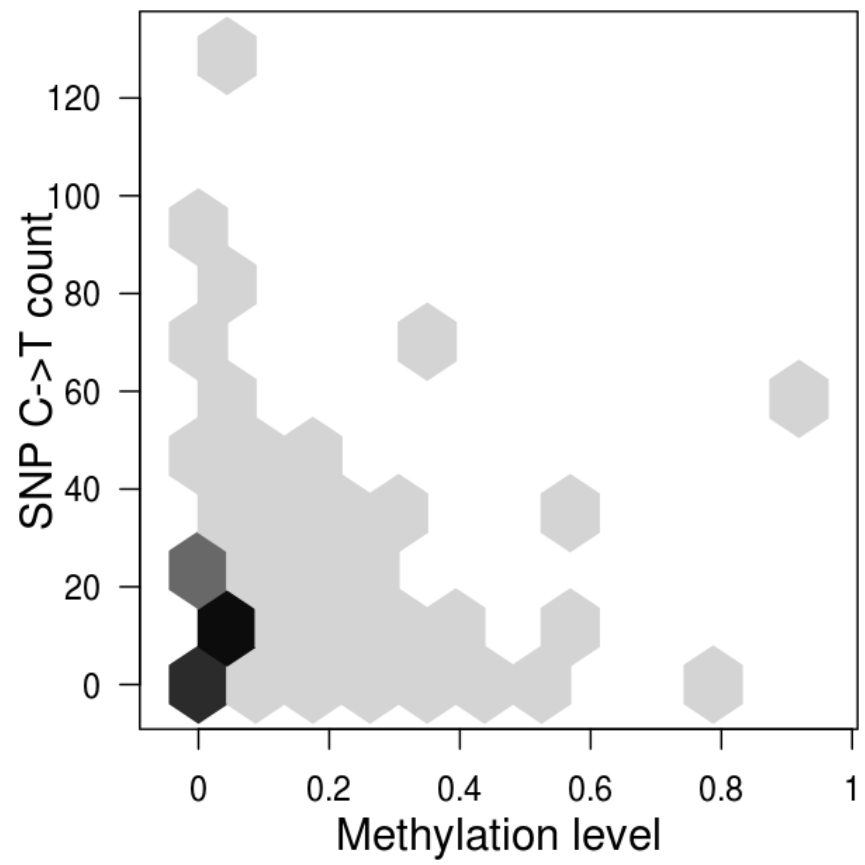
CHG context



CHH context



SNP C->T = F(Methylation)



Regression model

- No obvious correlations;
- Very rough error rate approximation.

Thank you! Questions?