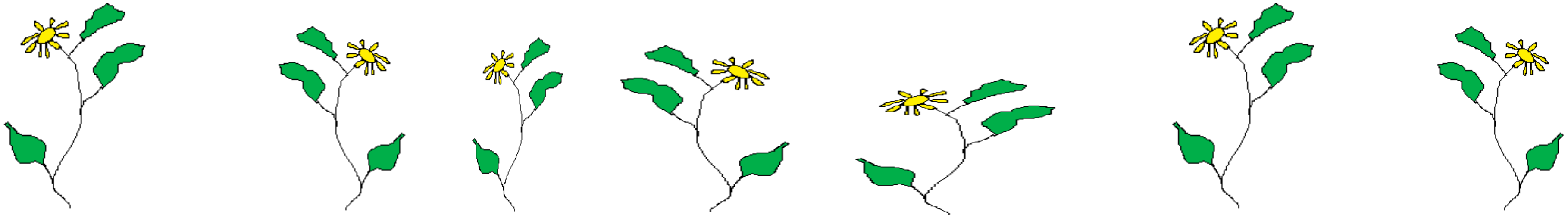
Genome-wide analysis of adaptation and speciation in *Senecio* & *Silene*

Dmitry A. Filatov

University of Oxford



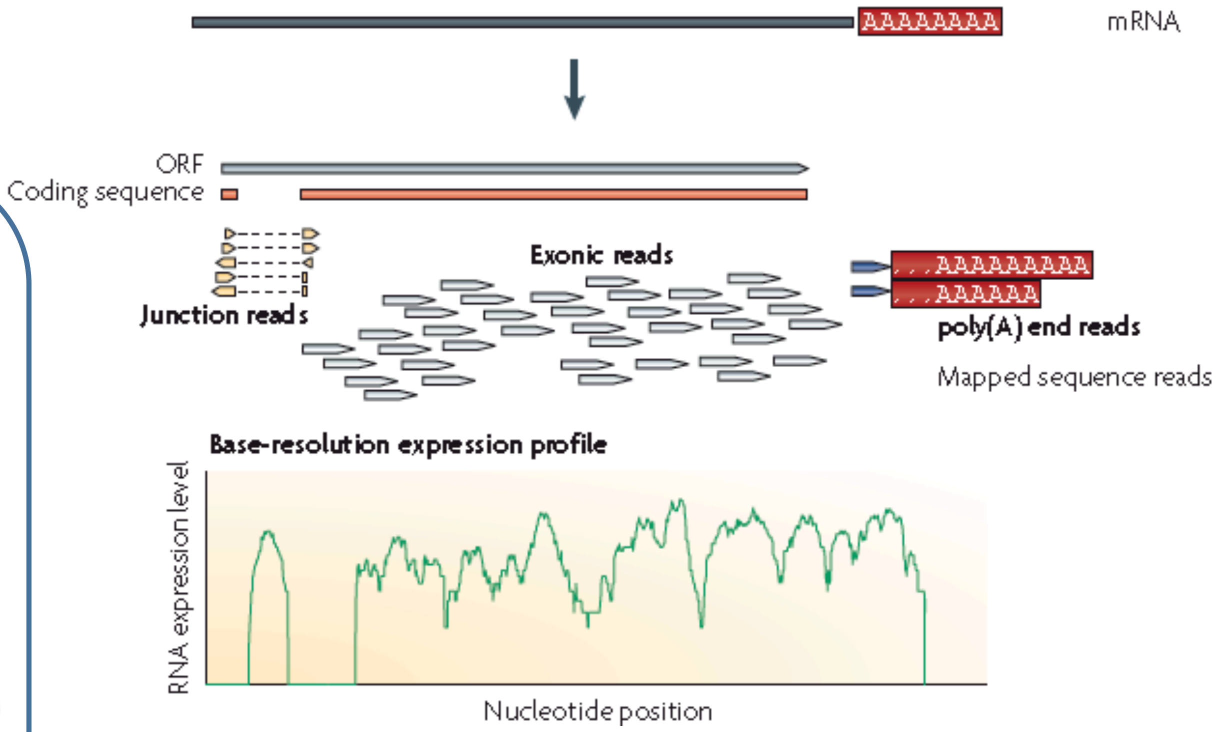
Project outline: the big questions

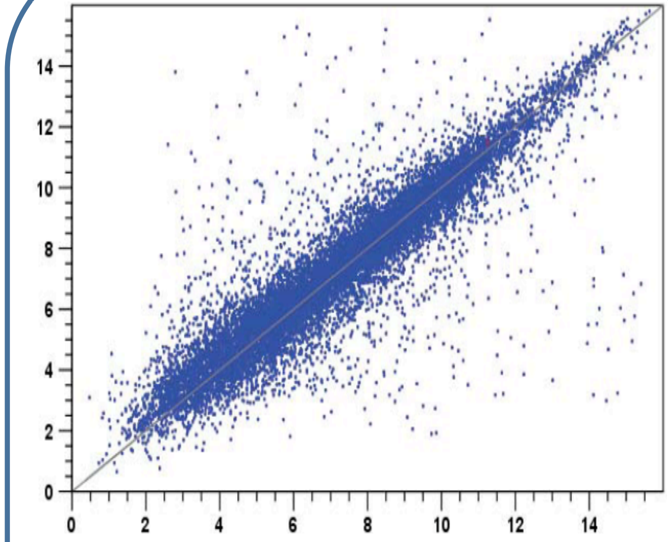
- How do species remain distinct despite gene flow?
- Are phenotypic differences between species (e.g. high/low altitude Senecio) maintained by contrasting selection pressures?
- Do genes under diversifying selection form “genomic speciation islands”?
- What is the proportion of genes that have been involved in adaptation and speciation processes in Senecio?
- Relative importance of evolution at non-coding regulatory sequences and in the protein-coding regions in adaptation and speciation
- Do closely related species ‘share’ adaptive alleles? Such alleles sharing can significantly impact adaptation and speciation, especially in small (e.g. island) populations

Project outline: approaches

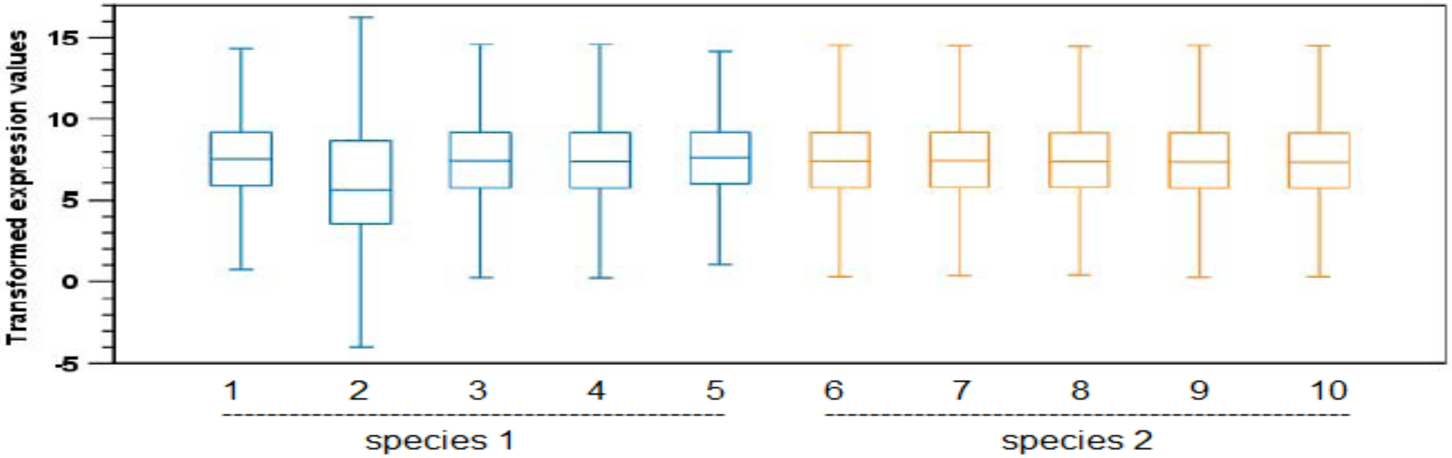
- Create a **reference genome** (*S. squalidus*)
- Sequence **transcriptomes** (incl mRNA and microRNA) for *S.aethnensis* and *S. chrysanthemifolius* and *S. squalidus*
- Identify **regulatory regions** (promoters, microRNA binding sites etc) that may have been involved in evolution of expression in Senecio species
- Sequence cDNA (polyA) for a sample of individuals from *S.aethnensis* and *S. chrysanthemifolius* and *S.squalidus* for **evolutionary genetic analyses**

RNAseq

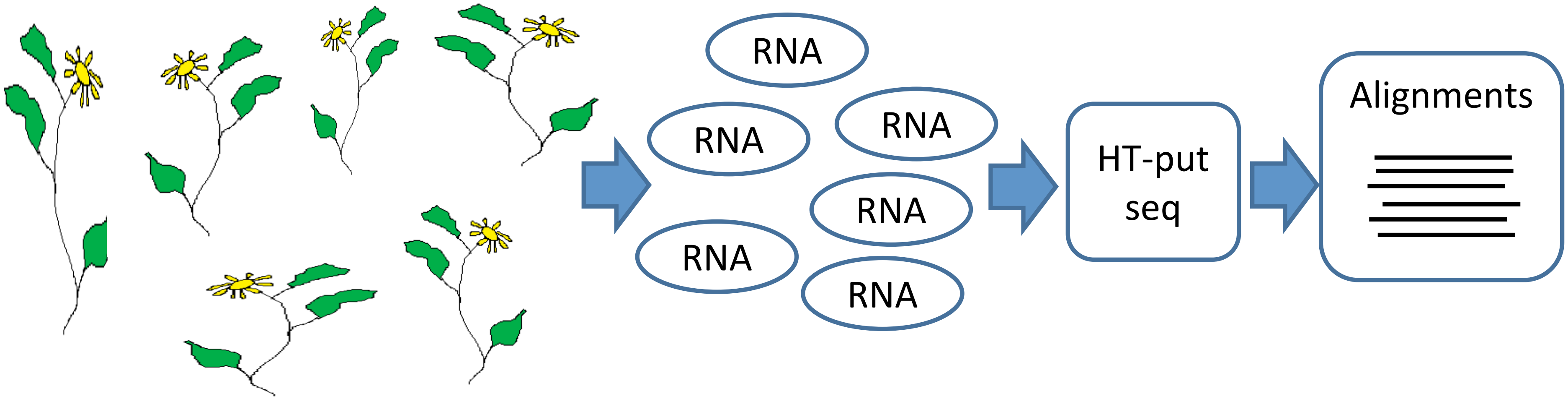




Is usually used for expression analysis



RNAseq for DNA polymorphism

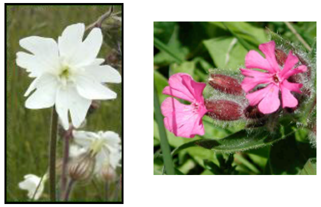


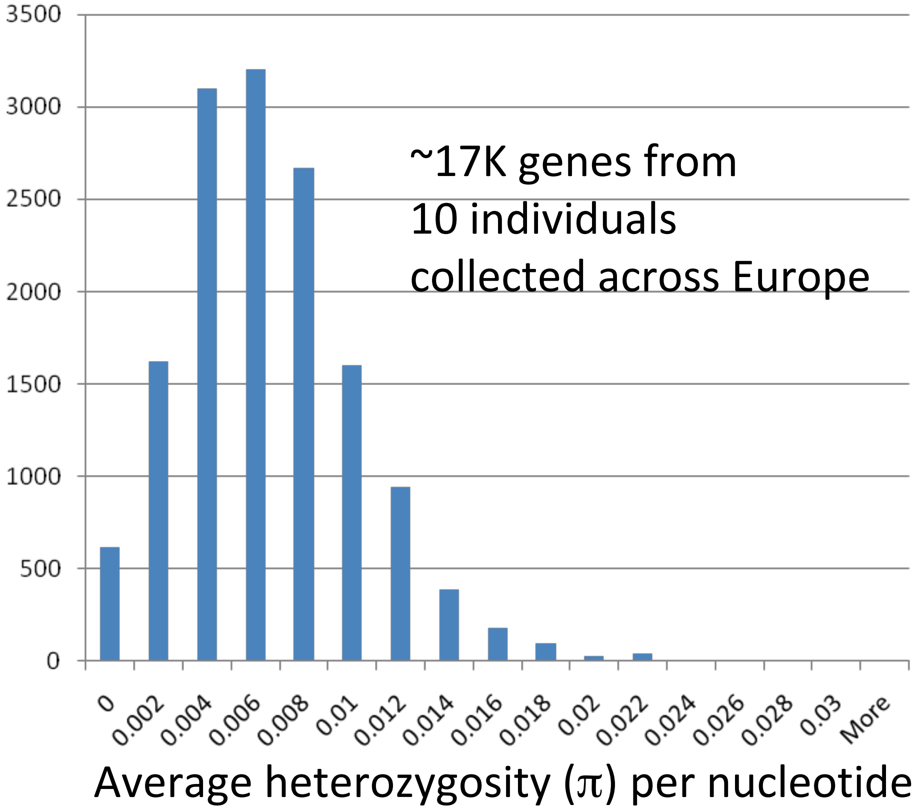
Coding regions only

➕
- •Polymorphism data for the majority of expressed genes
- •Relatively free from junk DNA
- •Comparison of silent/non-silent sites
- •Also gives expression data (by-product)

➖
- •Limited to transcribed regions
- •Coverage depends on expression

Example: Genome-wide DNA polymorphism in *Silene latifolia*



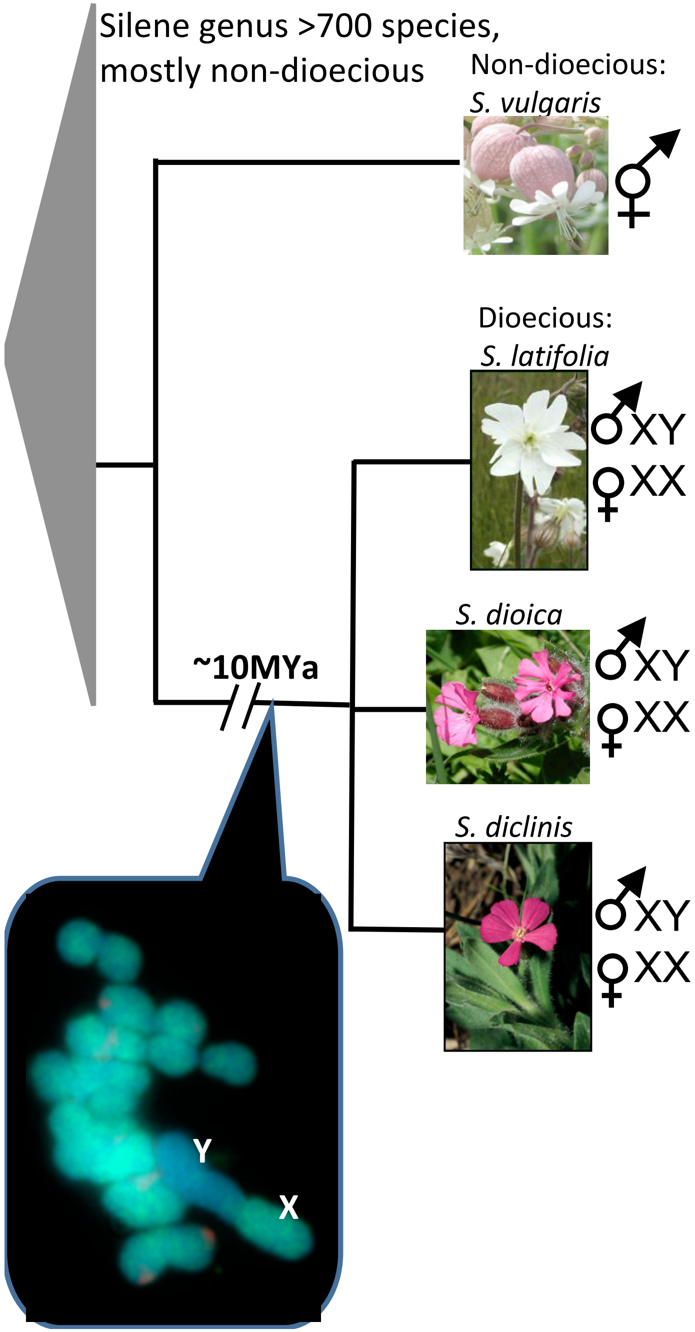


~17K genes from
10 individuals
collected across Europe

Can be used for:

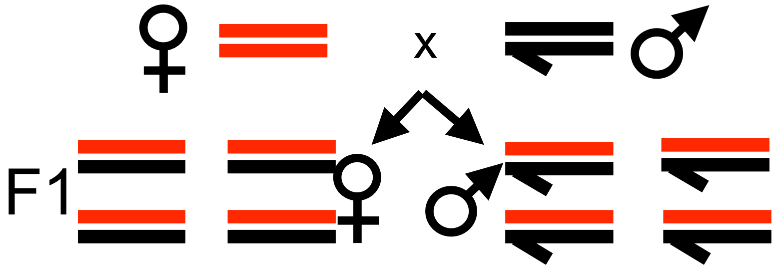
- Inference of past demography of the species
- Analysis of gene flow between the species
- Identification of genes that recently evolved under positive selection
- Etc…

Isolation of new sex-linked genes in *S. latifolia*



Principle:

Segregation of random genes (from cDNA)



The sons inherit
- Y-linked genes from the father
- X-linked genes from the mother

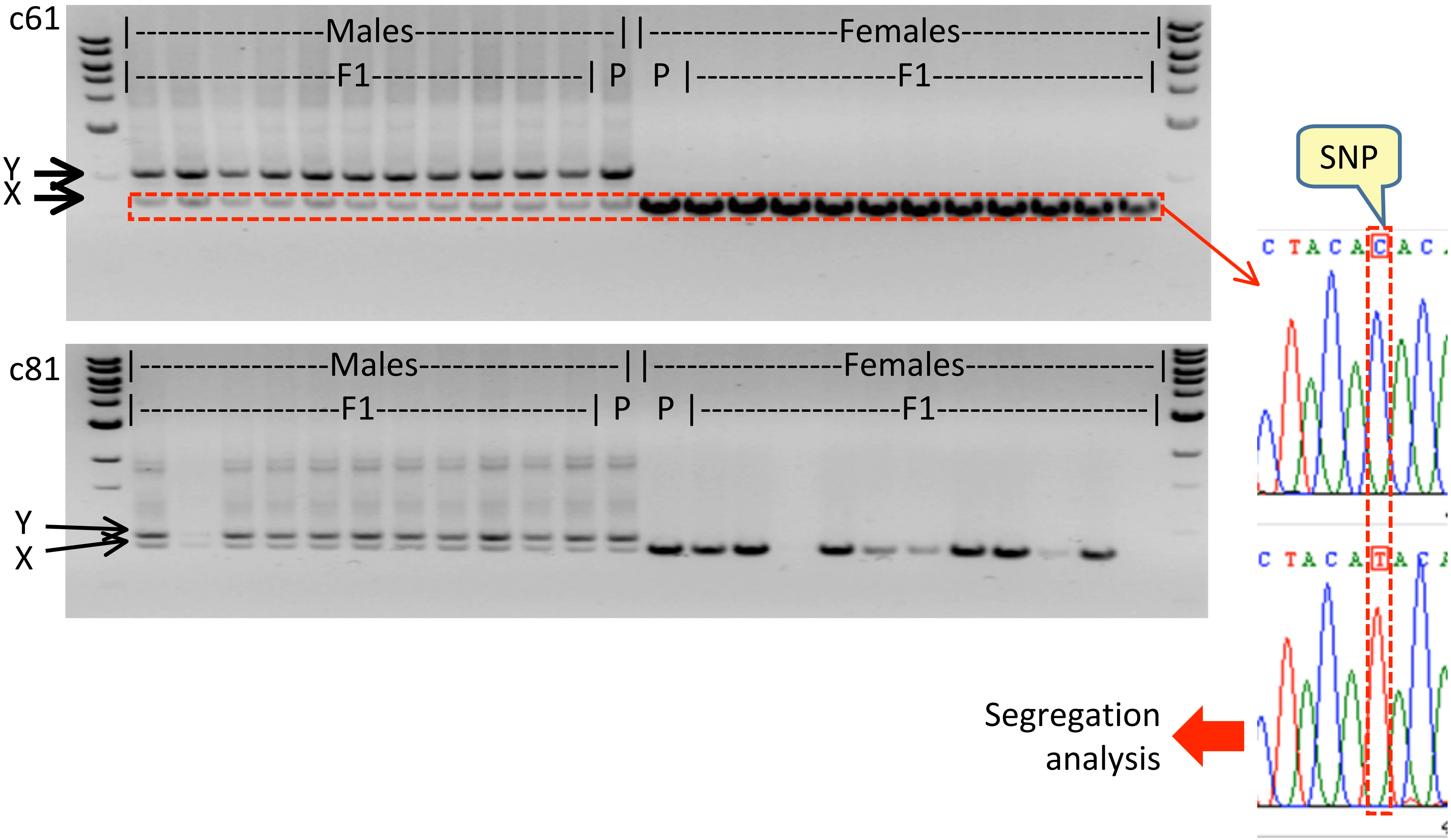
Method:

- Sequenced cDNA from parents and F_1 males & females of two families with Solexa/Illumina sequencing
- Identified SNPs and traced their segregation in *silico* and in the lab

Result:

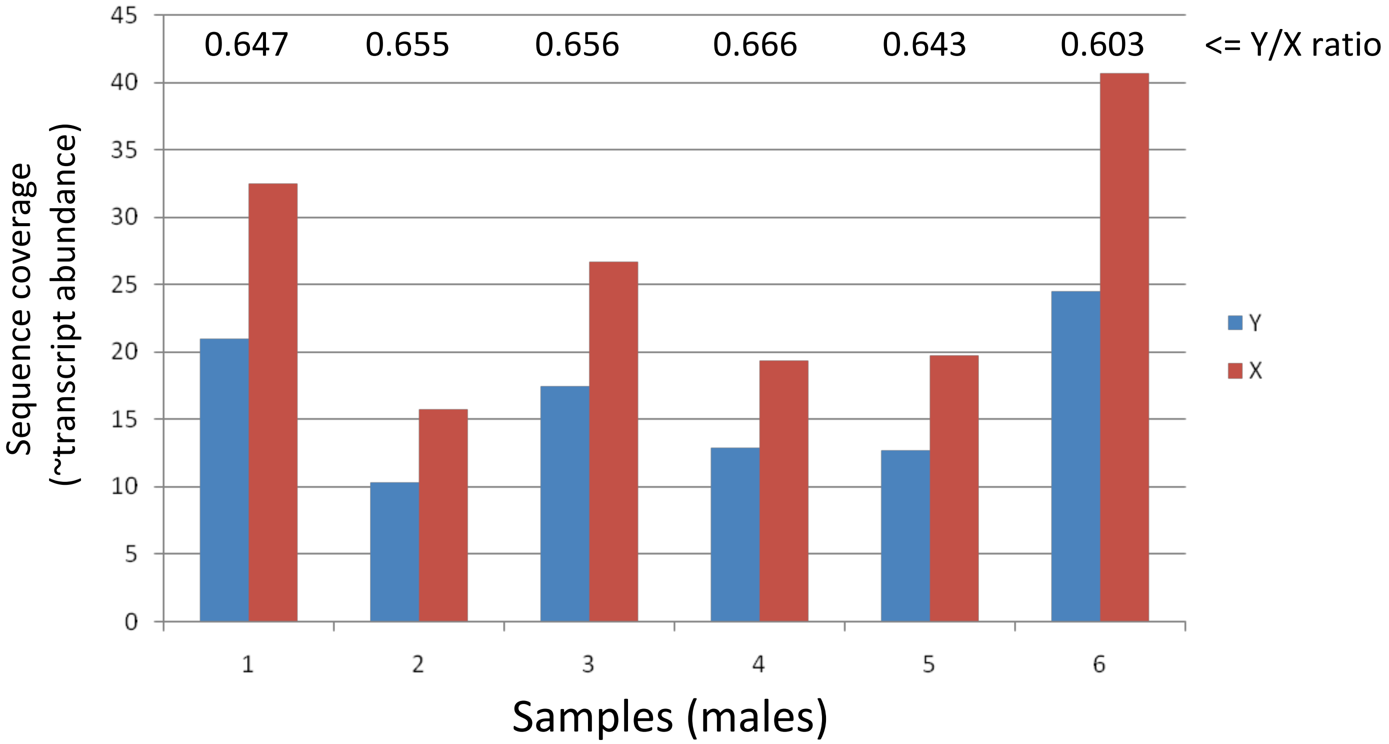
Identified 304 genes (including the known X/Y-linked genes) that consistently show X/Y segregation pattern in both families

Experimental verification of sex-linkage



Segregation analysis

Expression of Y versus X-linked genes



➡ Transcript abundance of Y-linked genes on average is only about 64% of that for X-linked homologues in males

Conclusions

RNAseq is a powerful tool for evolutionary genetics

Its use in this project will allow us to reveal the bases of adaptation and speciation in Senecio

Availability of the reference genome and identification of regulatory regions will allow us to extend evolutionary genetic inference beyond the coding/transcribed regions

Acknowledgements

Richard Abbott
Simon Hiscock
Matt Hegarty
Tom Batstone

Margarita Chibalina
Graham Muir
Jonas Sarasa-Marcuello

Funding:
NERC
John Fell fund (Oxford)

HT-Seq.Facilities:
Liverpool
Oxford