Adaptive processes through time in a global agricultural pest moth

Dr Ang McGaughran | Te Aka Mātuatua School of Science | University of Waikato | New Zealand

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

ADALEP: Adaptation à l'environnement chez les Lépidoptères
October 24-25 2023 | Versaille, France

Talk outline

01

Introductions

Me & the system

02

Wet lab

Sample age effects on NGS

03

Dry lab I

Value of the bycatch

04

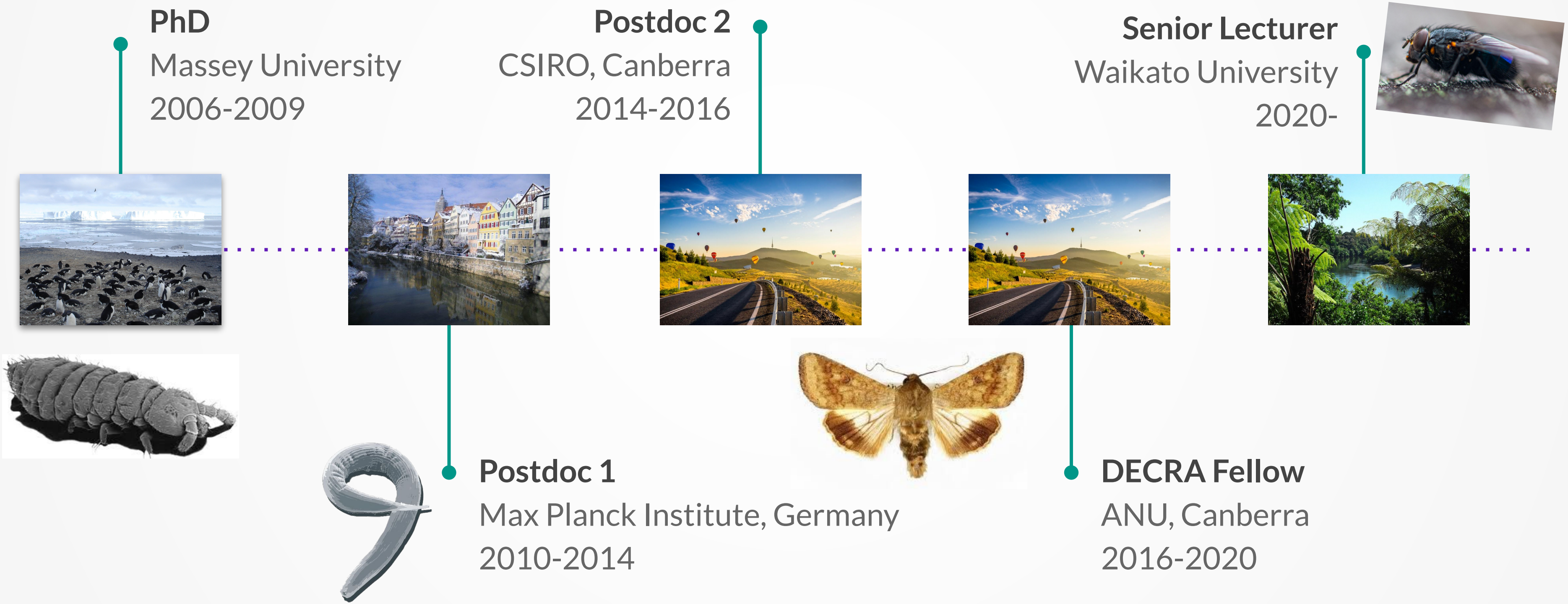
Dry lab II

Mechanistic insights

01

Introductions

Ang McGaughran



PhD
Massey University
2006-2009

Postdoc 1
Max Planck Institute, Germany
2010-2014

Postdoc 2
CSIRO, Canberra
2014-2016

DECRA Fellow
ANU, Canberra
2016-2020

Senior Lecturer
Waikato University
2020-

Invasomics Lab

How do populations **respond** to different environments?

Combining ecology with genomic approaches to examine evolutionary processes

Invasomics Lab

Helicoverpa armigera

A major pest causing **billions** of dollars in crop losses and management

Helicoverpa armigera

Highly polyphagous, diverse defences, highly motile

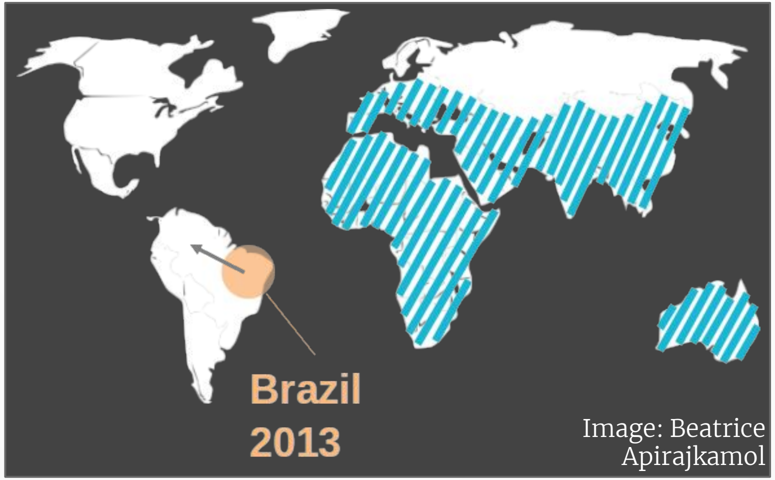
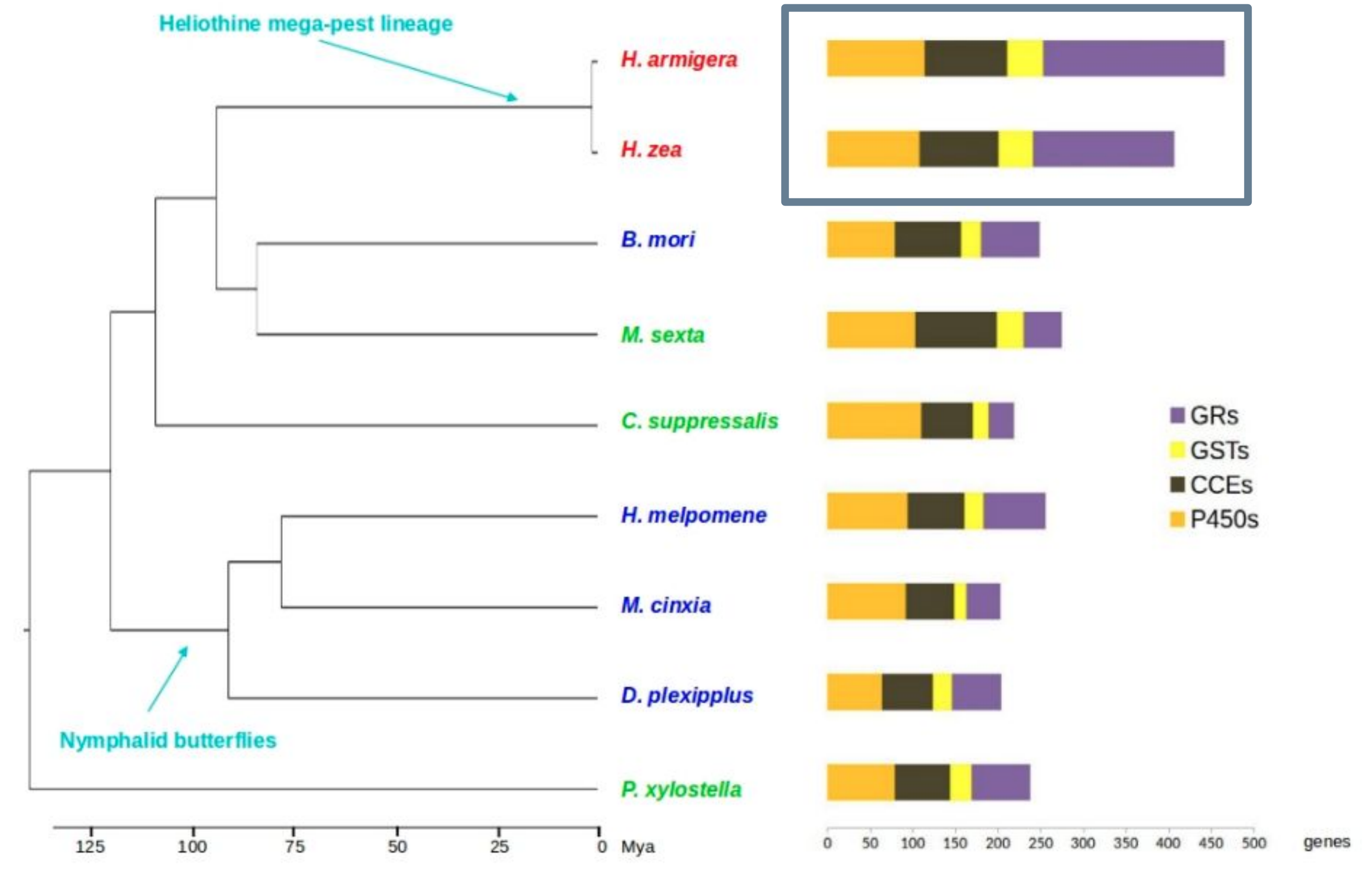


Image: Beatrice Apirajkamol

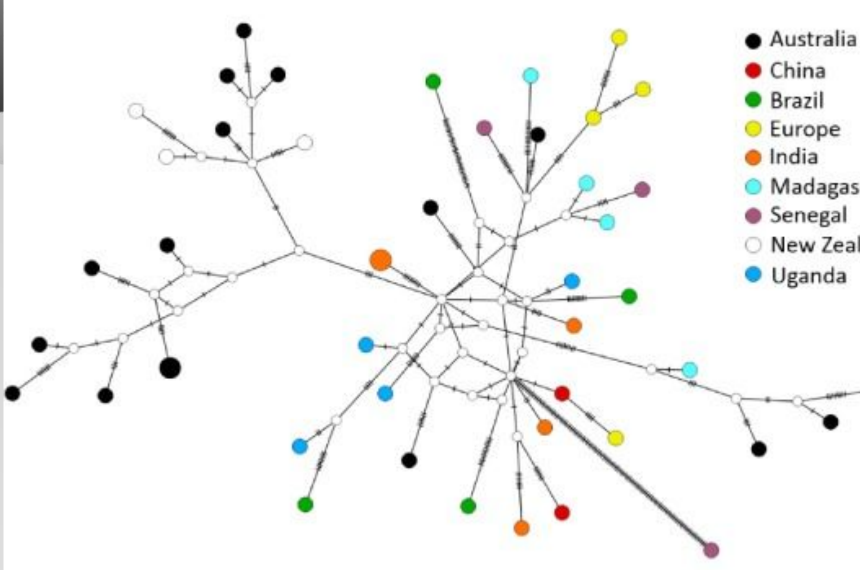
Novel genetic changes can quickly sweep through populations

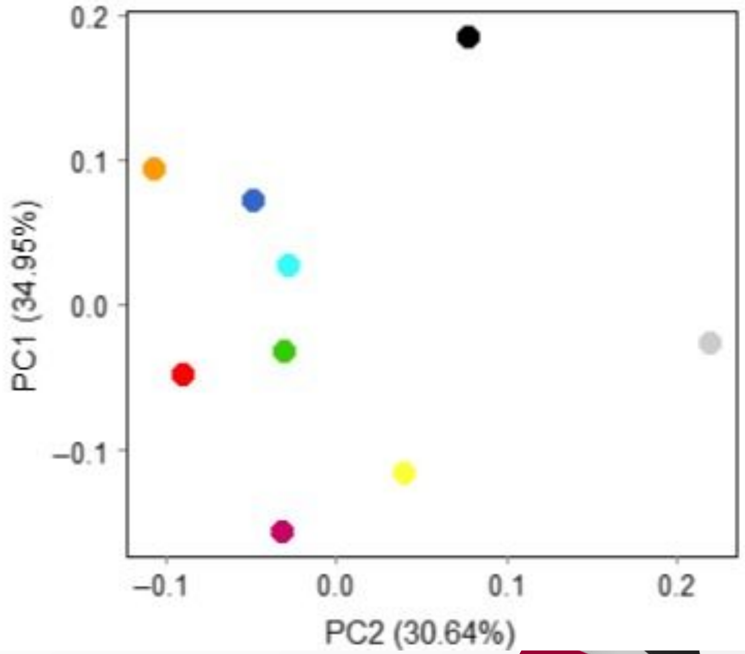
—**Rapid development of insecticide resistance**



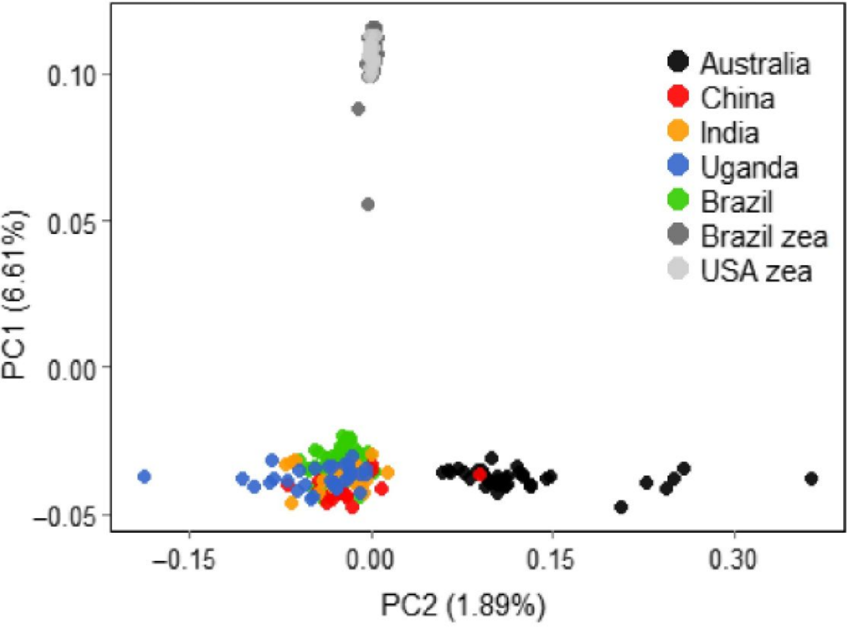
Hosts: one species or genus; one family; polyphagous

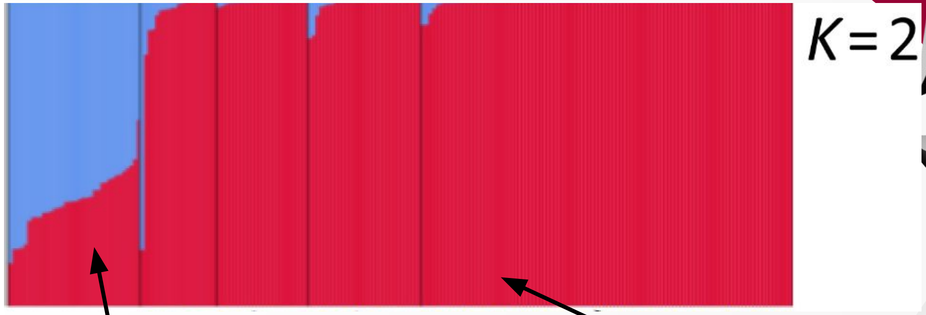
Population structure





Population structure





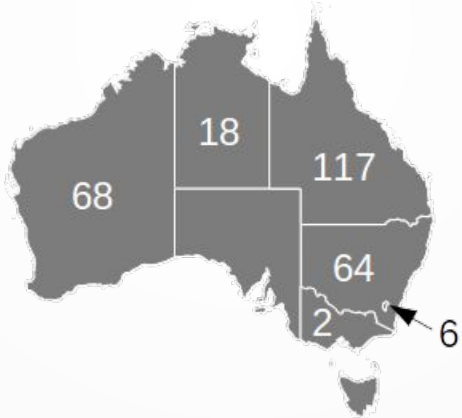
Australia: *H. armigera* ***conferta***

ROW: *H. armigera* ***armigera***

→ ???

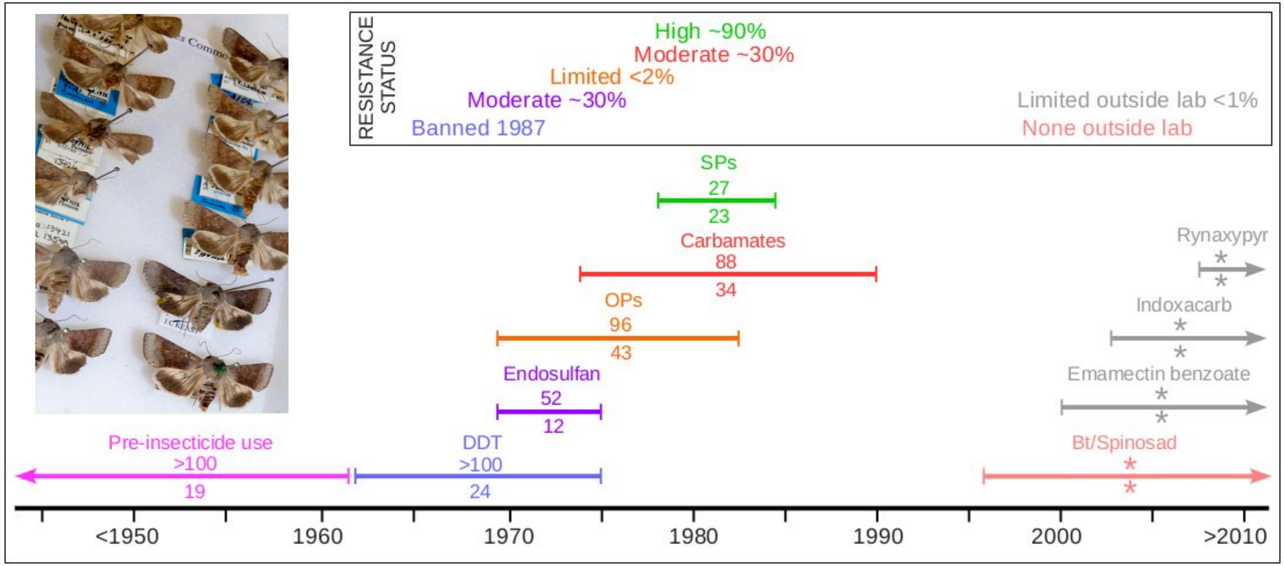
Project design

>260 samples from across Australia



Project **design**

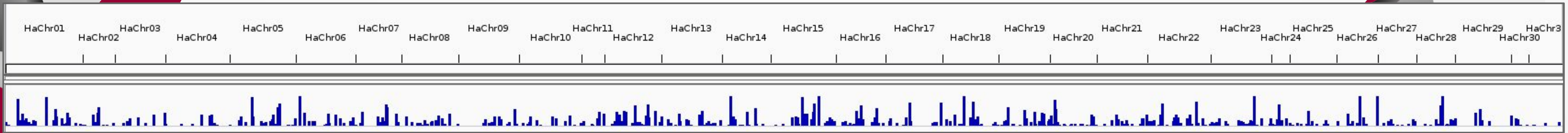
Museum specimens through time



Project design

Targeted capture

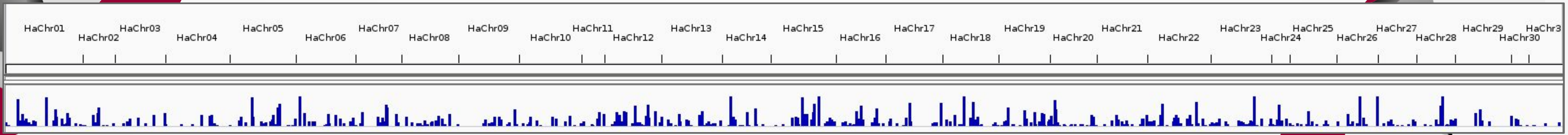
Genome 337 Mb



Project design

Gene families:

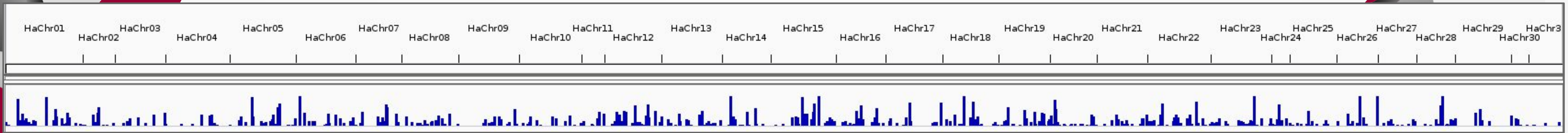
ABCs: ATP-binding cassette transporters; **CADs**: cadherin genes; **CCEs**: carboxyl/cholinesterases; **CRPs**: chemosensory receptor proteins; **GSTs**: glutathione S-transferases; **HSPs**: heat shock proteins; **ICHs**: ion channel-related genes; **LIPs**: lipases; **P450s**: cytochrome P450s; **SERs**: serine-proteases; **UGTs**: uridine diphosphate (UDP)-glucuronosyltransferases



Project design

Involved in:

LIPs/SERs (digestion, host use); CRPs/HRPs (environmental sensation/response); CADs/ABCs (Bt resistance); ICHs/CCEs/p450s/GSTs/UGTs (detoxification and insecticide resistance)



Problems to solve

Wet lab

How does sample age affect data quality?

Dry lab I

How connected are Australian populations?

Dry lab II

How does rapid evolution proceed mechanistically?

02

Wet lab

Study motivation

How does **sample age** affect data quality?

Key methods

‘Salting-out’ **DNA extraction** protocol using soaked moth abdomens

Modified Meyer Kircher **library prep** protocol:

No **shearing**

Use of **USER** (Uracil-Specific Excision Reagent) enzyme in the blunt end-repair step

On-beads clean-up protocol throughout (no elution)

Key methods

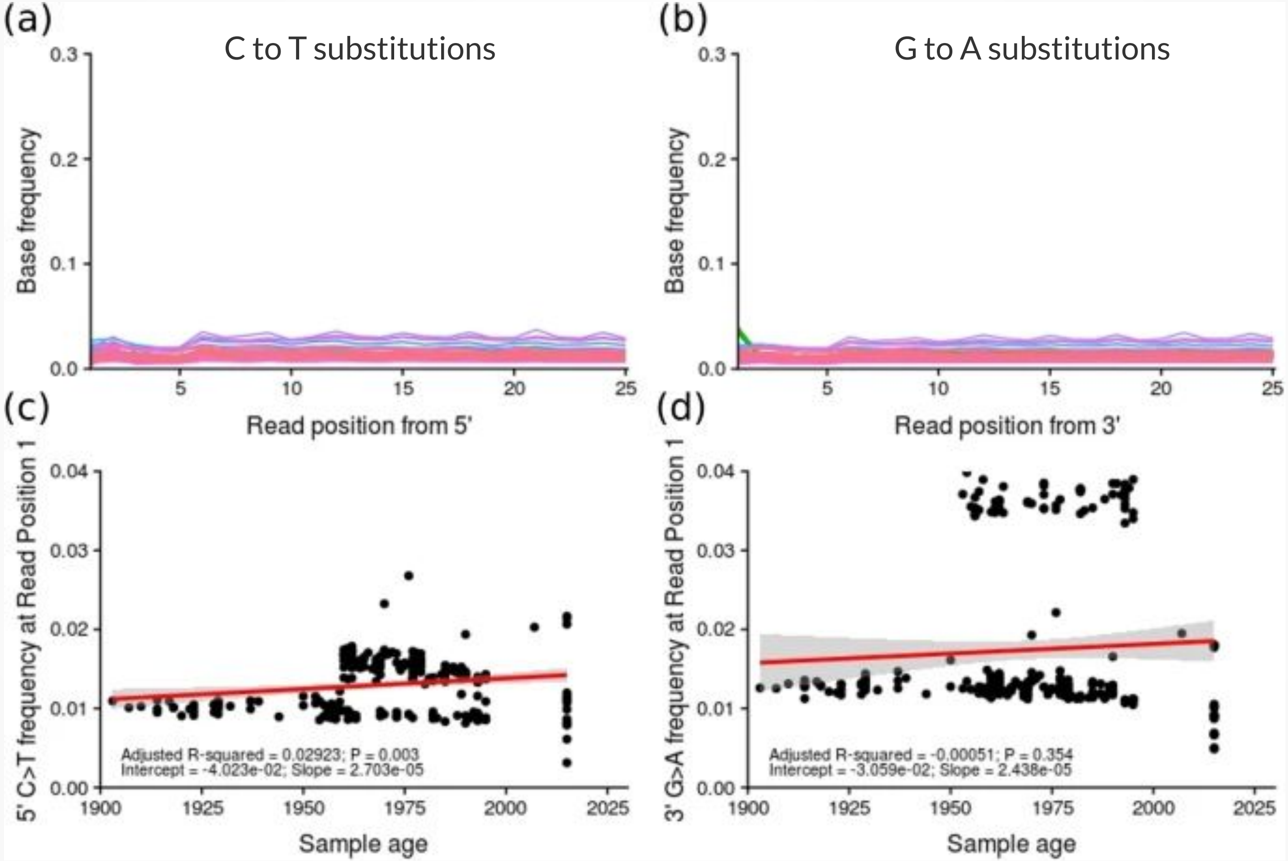
Mapdamage to examine frequency of C to T and G to A transitions at the 5′ and 3′ end of reads, respectively

Correlation analyses between sample age and various metrics of data quality:

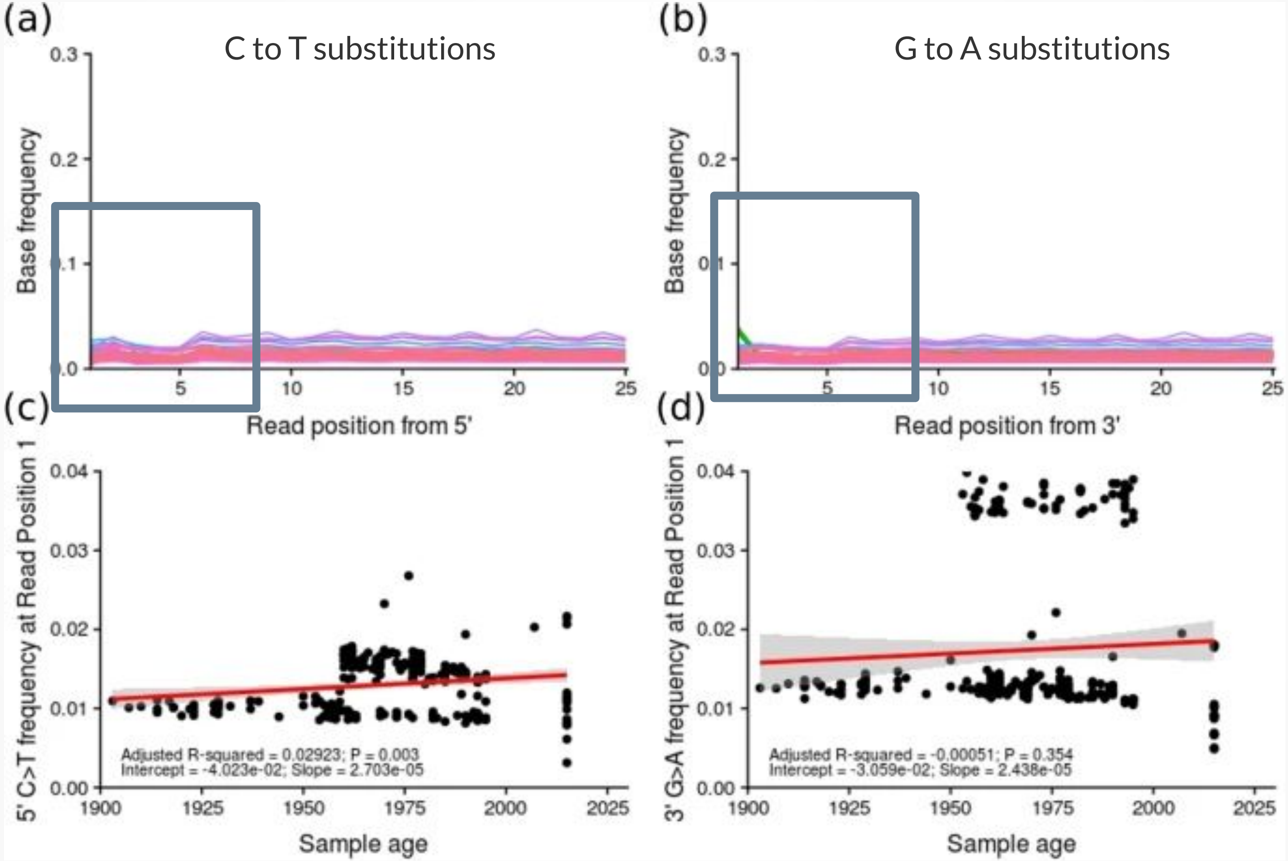
When reads aligned to *whole genome*

When reads aligned to ‘*baitome*’

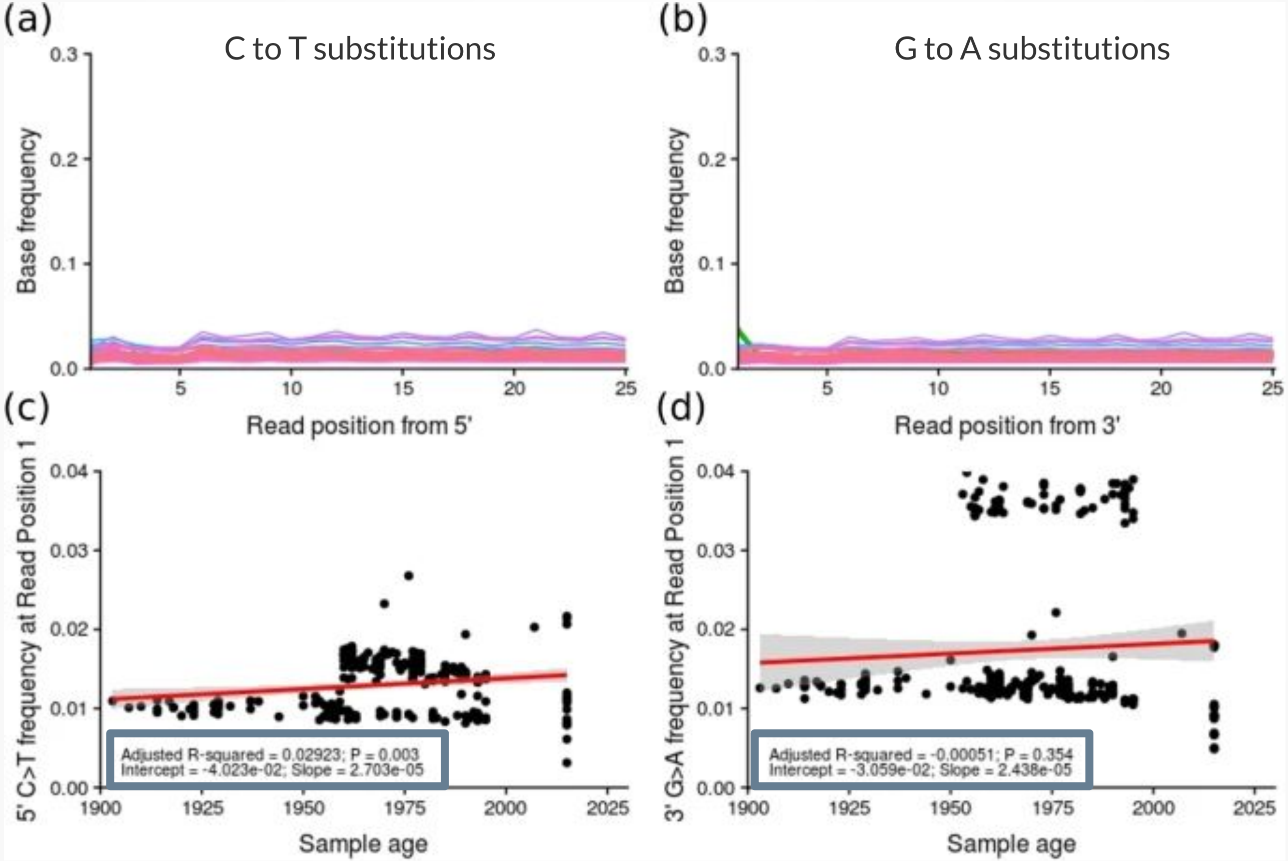
Effects of sample age



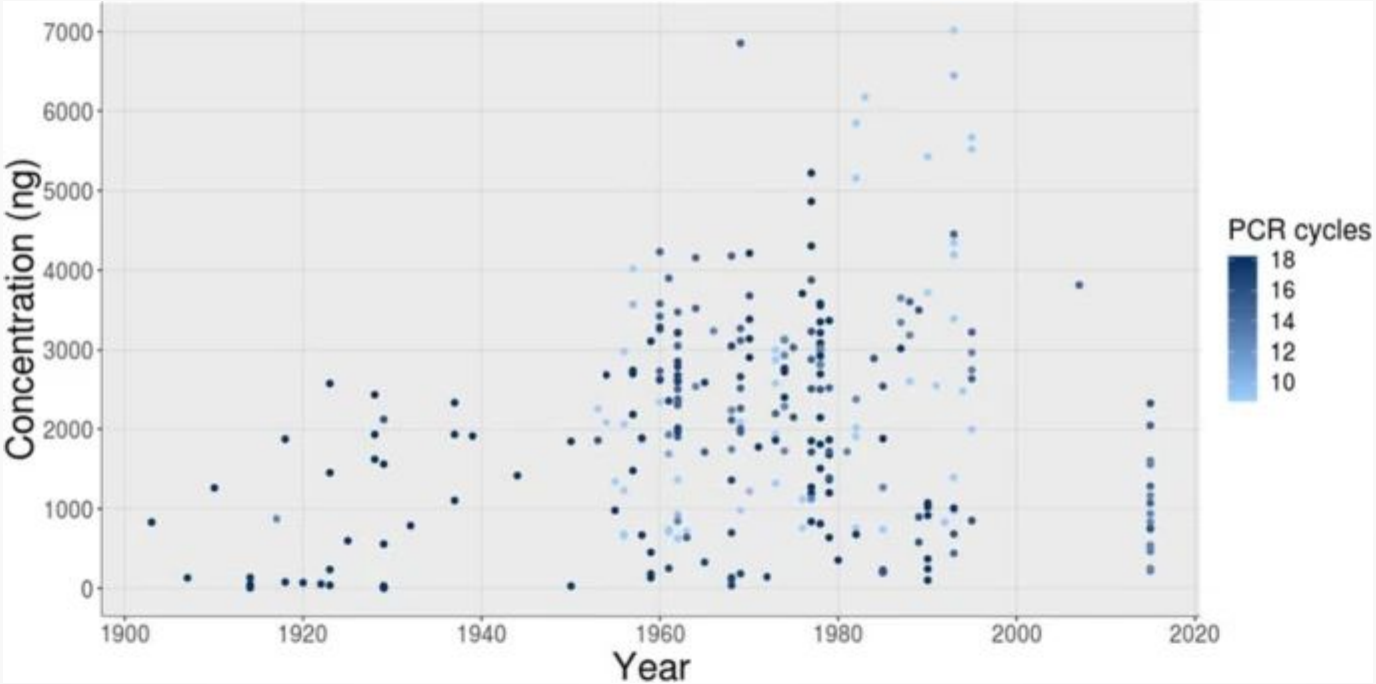
Effects of sample age



Effects of sample age



Effects of sample age

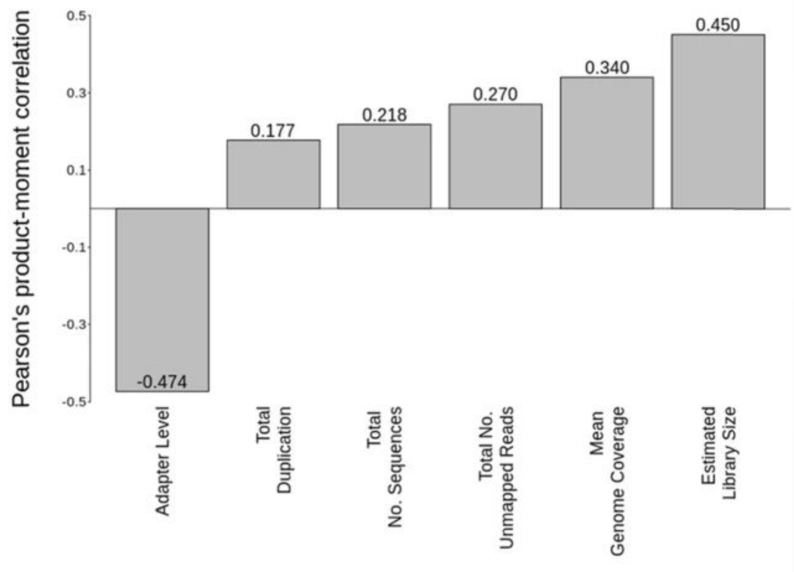


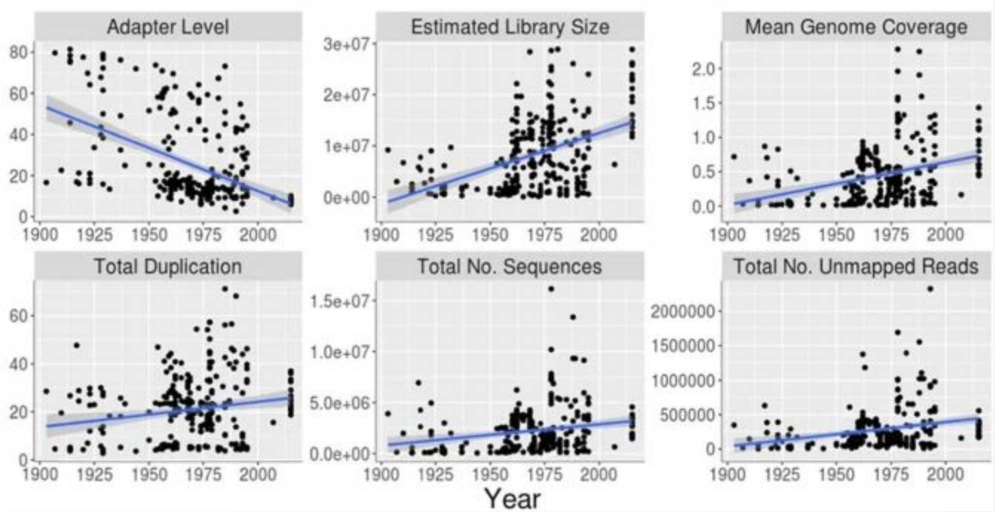
Older samples have **lower starting [C]:** T_{269}=3.83; *P*<0.01; *R*=0.23 (0.11:0.34)

Require **more PCR cycles:** T_{269}=-5.56; *P<0.01*; *R*=-0.32 (−0.42:-0.21)

Effects of sample age

Correlation between sample age and various aspects of library quality for *full* genome



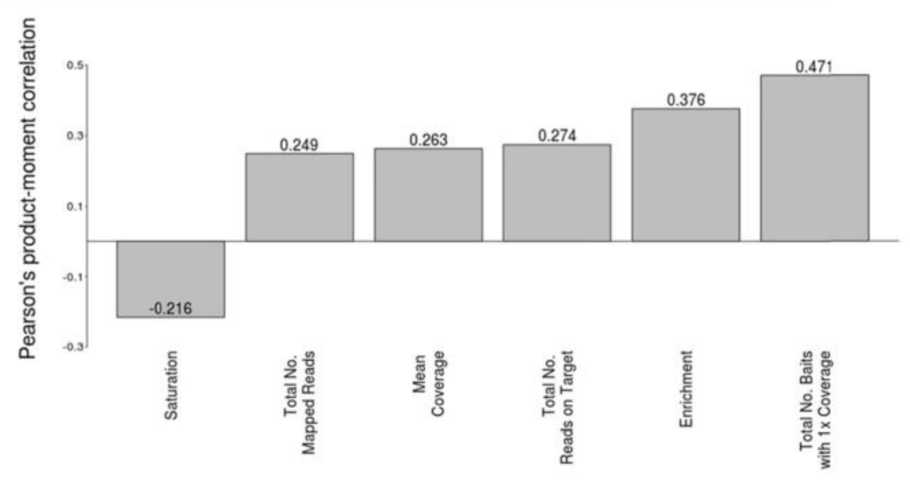


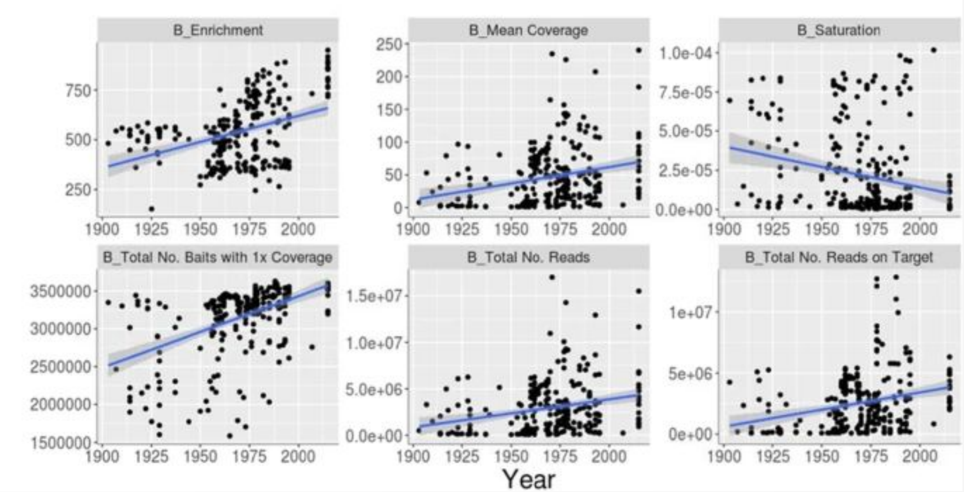
Younger samples = **better quality**

Older samples = **more adapters (up to 82%)**

Effects of sample age

Correlation between sample age and various aspects of library quality for *captured* reads





Younger samples = **better quality**

Saturation: indication of whether higher sequence depth = a higher percentage of covered positions

Key points

Age has a **big impact**, as expected: negative relationship between age and data quality

No signs of **deamination damage** (USER enzyme?)

High **adapter contamination** (adapter dimers similar in size to library, hard to remove) → dimer free methods

Percentage of unmapped reads not related to sample age

More information

Effects of sample age on data quality from targeted sequencing of museum specimens: what are we capturing in time?

Angela McGaughran ✉

BMC Genomics **21**, Article number: 188 (2020) | Cite this article

2224 Accesses | **9** Citations | **16** Altmetric | Metrics

03

Dry lab I

Study motivation

How **connected** are Australian populations?

Is there any value in the **bycatch?**

Key methods

Used data from McGaughran (2020) – 262 samples, combined with 53 samples from Anderson et al. (2016)

Substantially increased **geographic coverage:**

All Australian states except TAS, plus Brazil, China, France, India, Madagascar, New Zealand, Senegal, Spain, Uganda

Aligned reads to *H. armigera* reference mitogenome using **BWA mem**

Key methods

Called variants with **GATK**

(a) Subset data into at least 5% coverage of ref genome (n=250), 10%, …, 65% (n=56)

(b) Gathered 817 location-tagged **mtCOI** sequences from GenBank and trimmed to retain at least 65% coverage of first 653 bp (n=648; 518 from GenBank)

Key methods

Performed **DAPC** analysis in R (adegenet) to explicitly test for the presence of exclusive geographic distributions for distinct *H. a. armigera* and *H. a. conferta* clusters

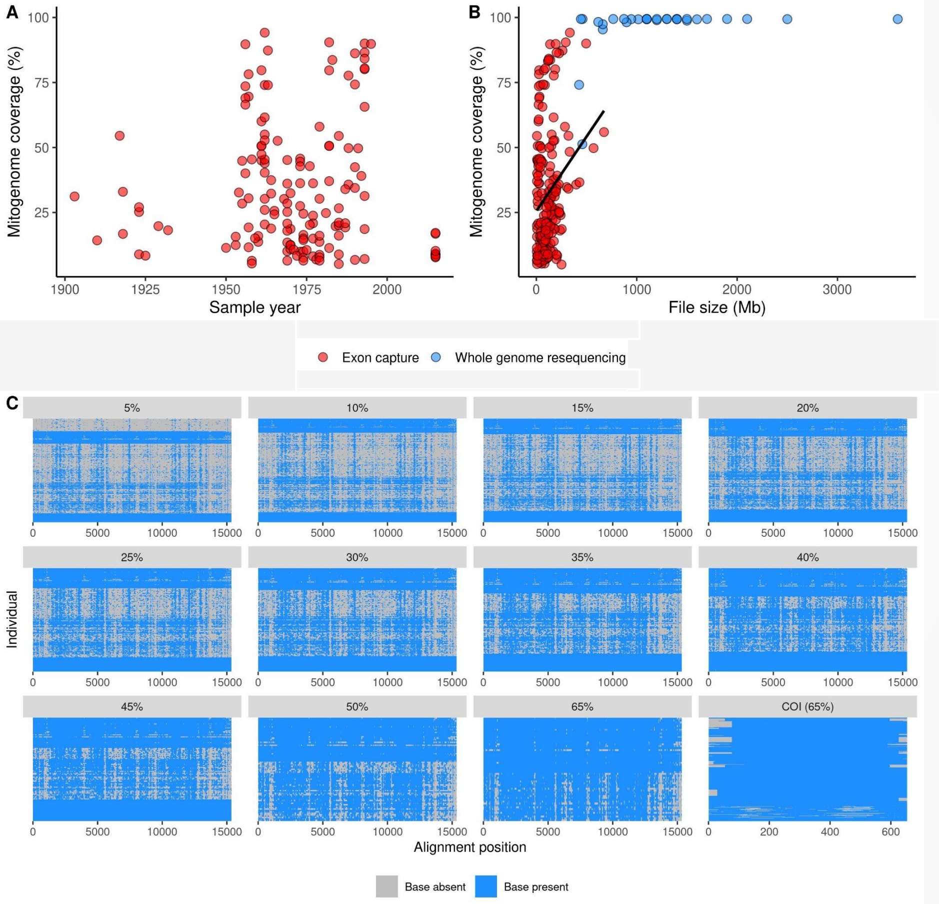
Phylodynamic analysis (Bayesian Coalescent Skyline in BEAST), to infer demographic history

Bycatch coverage

(a) Proportion of mitogenome coverage versus sampling age; and (b) file size

(c) Coverage heatmaps for: 5%-65% mitogenome datasets; COI (n=648) dataset

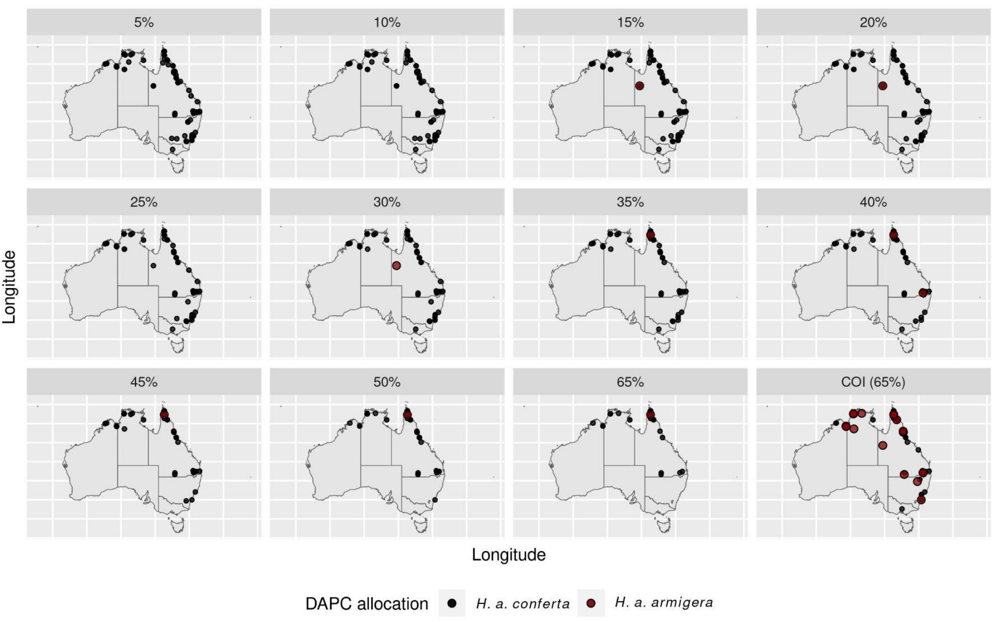
In panel (c), individuals are represented as rows and are plotted in a random order

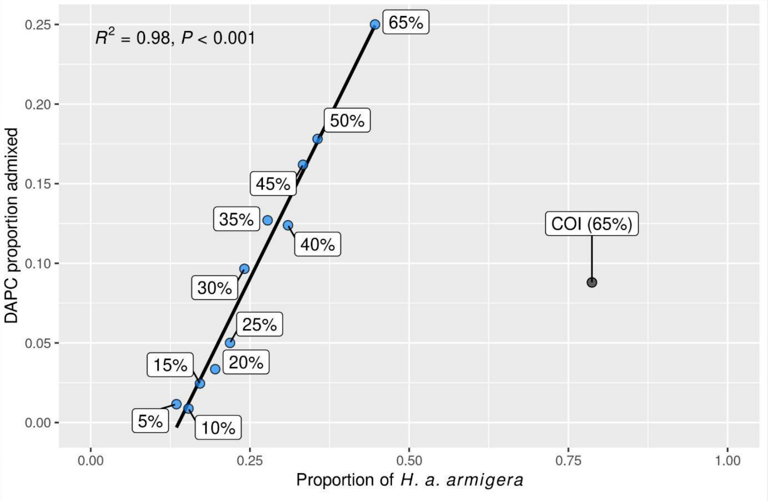


Leo Featherstone; University of Melbourne

Subspecies support

DAPC results for 5-65% mitogenome datasets



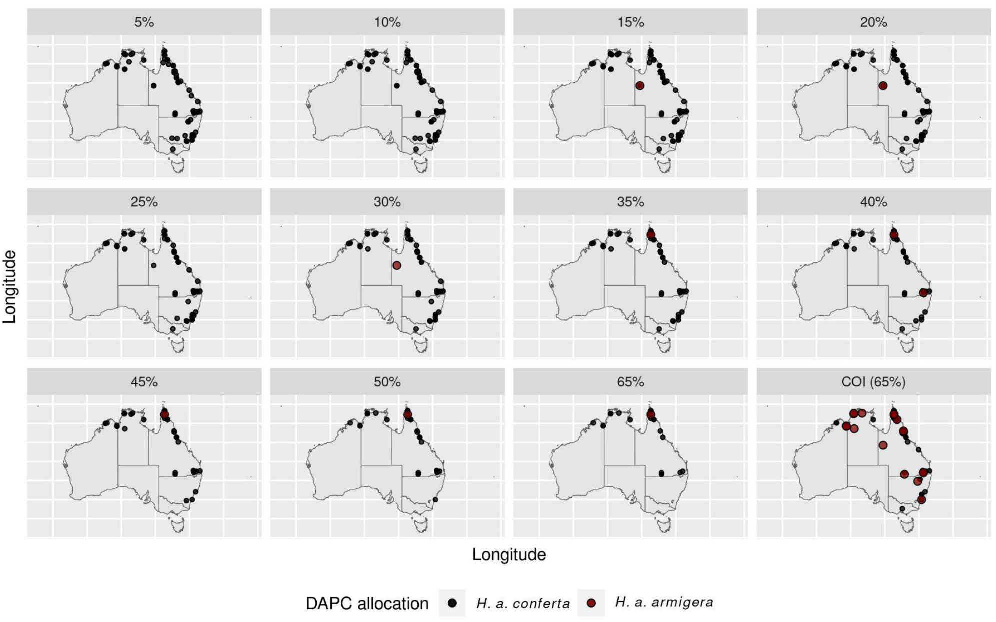


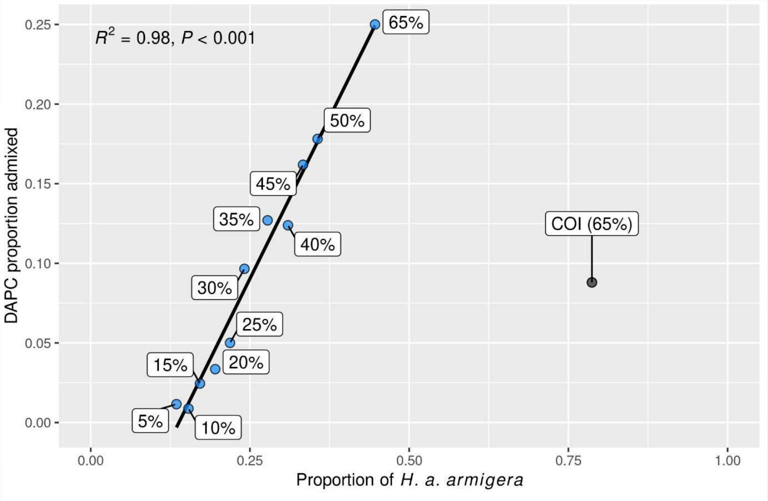
Supports distinct **Australasian subspecies** (left)

Significant effect of coverage and dataset composition (right) – admixture proportion increased with proportion of *H. a. armigera*

Subspecies support

DAPC results for 5-65% mitogenome datasets





COI dataset had higher proportion of *H. a. armigera* but lower admixture (~10%)

Sampling bias alone **does not explain** the increased admixture in the mitogenome data

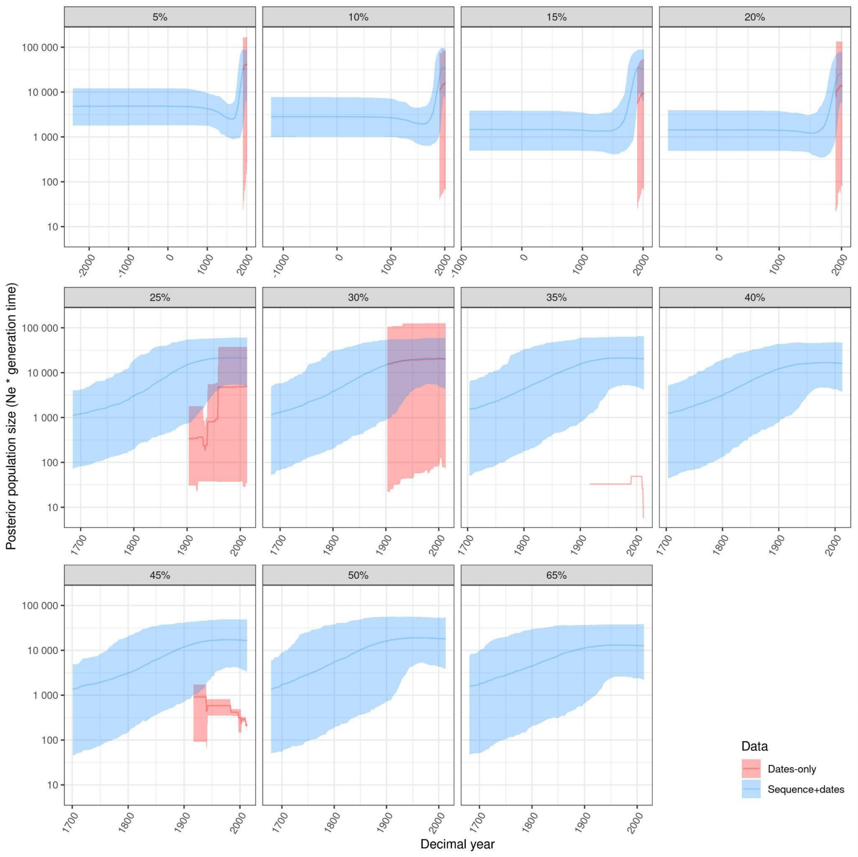
Higher coverage in COI data allowed **clearer separation** of *armigera* and *conferta*

Demographic analyses

Continual **increase in population size** for each dataset

Posterior population trajectory for 5-20% coverage datasets much **older**

Lower coverage → older population trajectory (> divergence among sequences)



Key points

Mitogenomes assembled from bycatch with up to **75% missing data** were able to return evolutionary inferences consistent with higher coverage datasets and the broader literature surrounding *H. armigera*

Key effects of dataset coverage and composition

Value of museum species as important records of historical change via the bycatch, but **caution** that missing data doesn't allow confounding factors to drive inference

04

Dry lab II

Study motivation

How does rapid evolution proceed **mechanistically**?

Dr Eli Parvizi;
University of Waikato

Andy Bachler;
CSIRO

Key methods

Data from McGaughran (2020) subset by decade to look at temporal patterns

Aligned to reference with **BWA mem**

Quality assessment with **ngsCAT**:

Mean coverage 18x (range: 1-234x; s.d.: 42x)

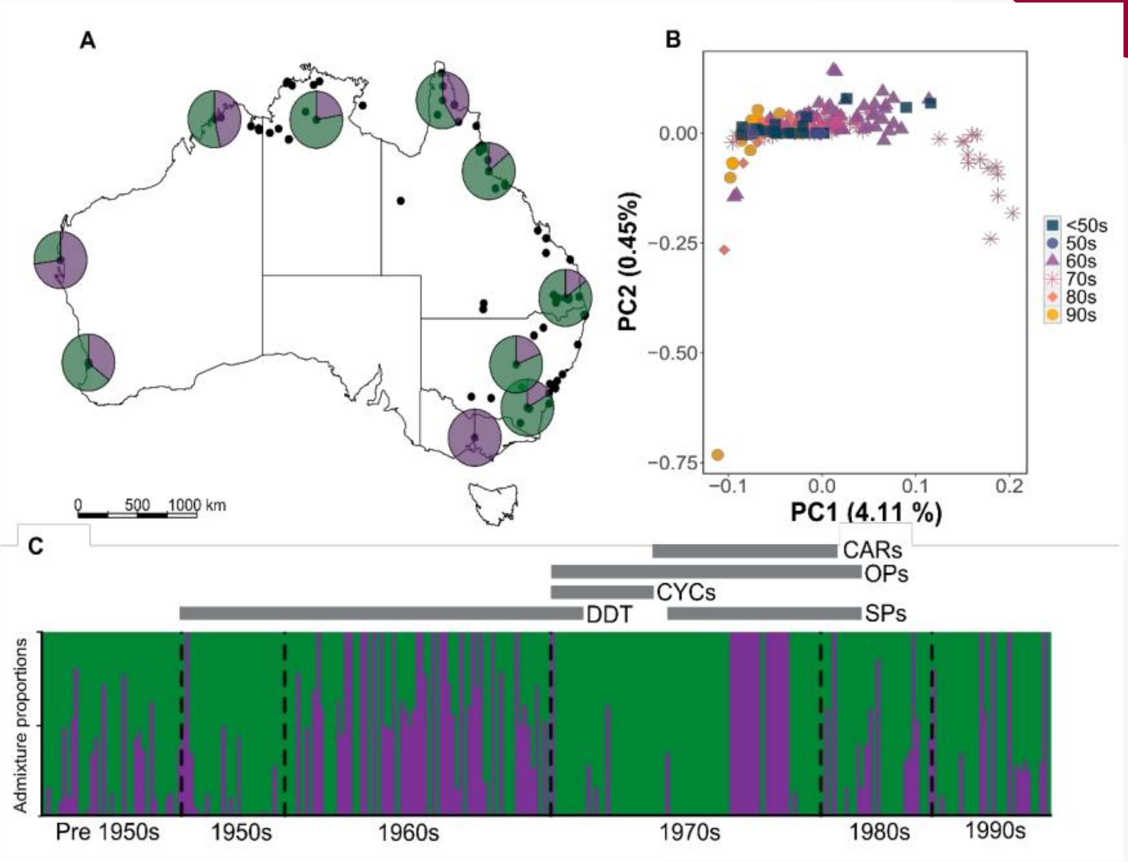
Mean reads on-target: 81% (range: 65-93%, s.d.: 4%)

Use of **ANGSD** to calculate genotype probabilities (59,798 variants)

Population structure

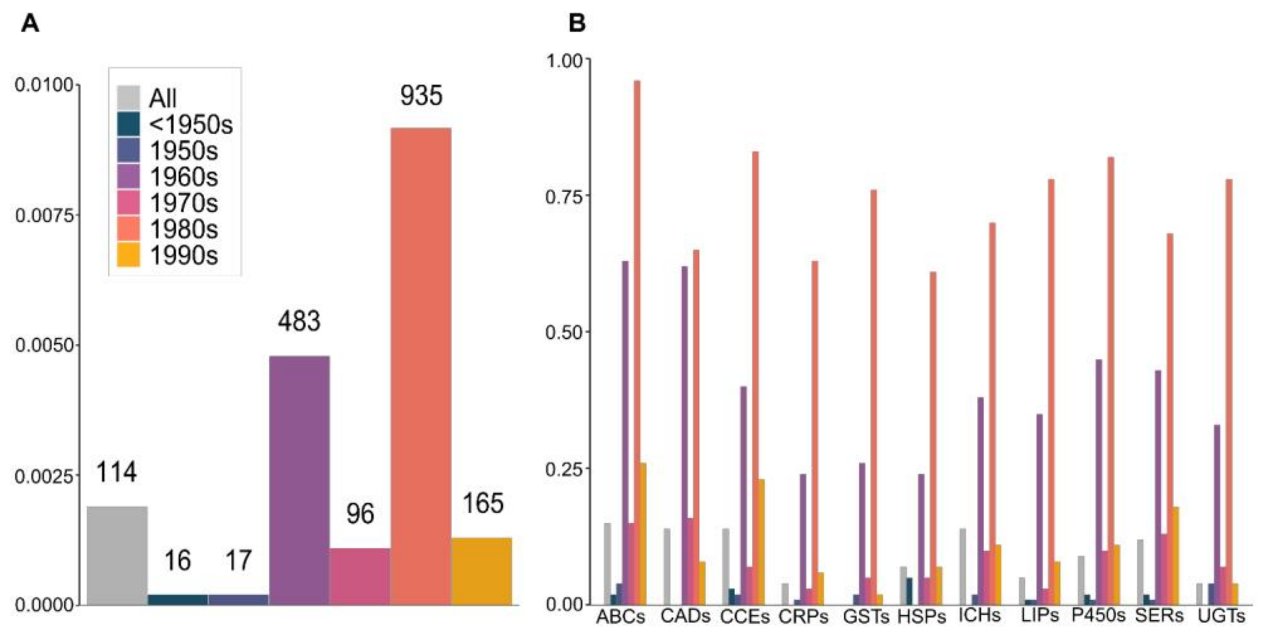
PCAngsd for PCA: **no overall clustering** related to decade or state, some local clustering for 70s/WA

NGSadmix for admixture plots: each cluster **well represented** through time and space, more mixing in 60s vs 70s?



Outliers

PCAdapt: Proportion of 1,247 exons containing SNP outliers, and their relationship to time and gene family

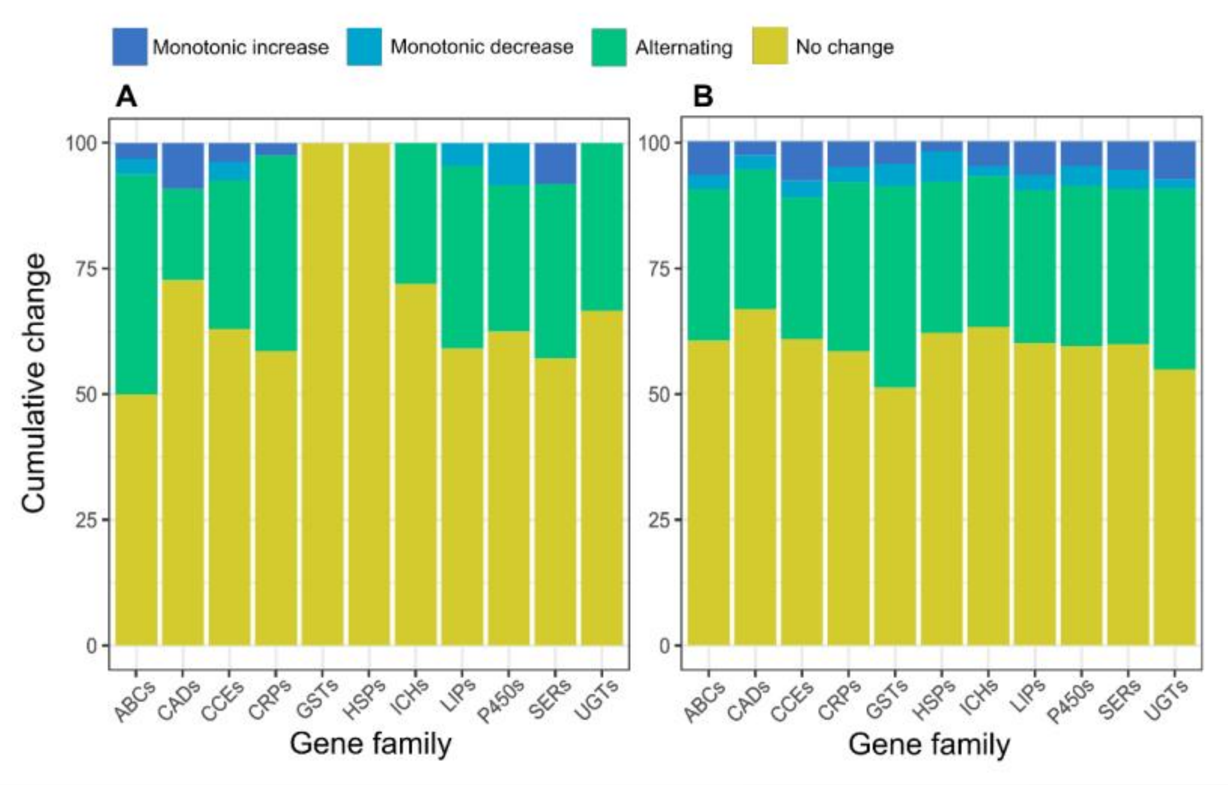


Increased outliers in 60s and 80s overall, and for all gene families → DDT/SPs?

Temporal changes in AF

Up to 50% outlier allele frequencies **changed** from one period to the next

Changes mostly **alternating: antagonistic pleiotropy?**

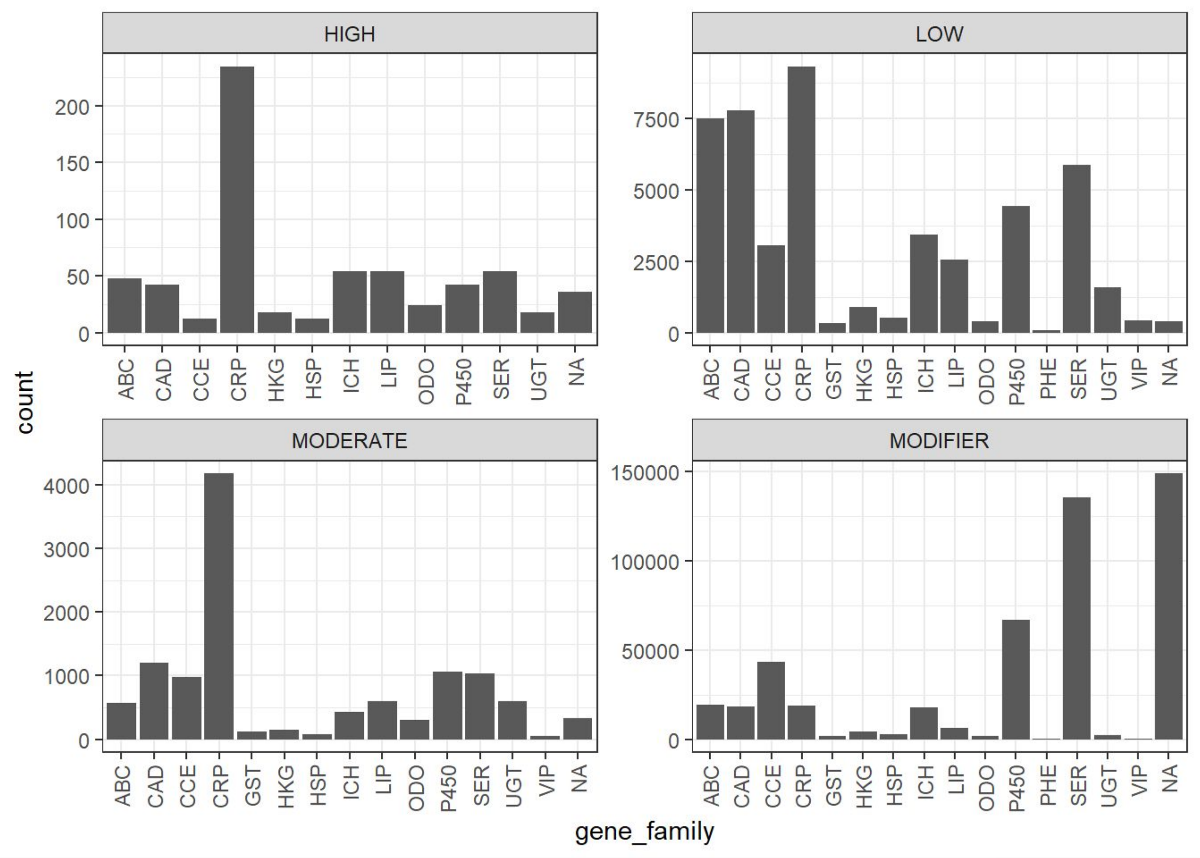


p1950s vs 1960s vs 1980s

p1950s vs 1980s vs 1990s

Variant impacts through time

SnpEff showed most SNPs had **low** deleterious gene effects (median across gene families = 2,046)



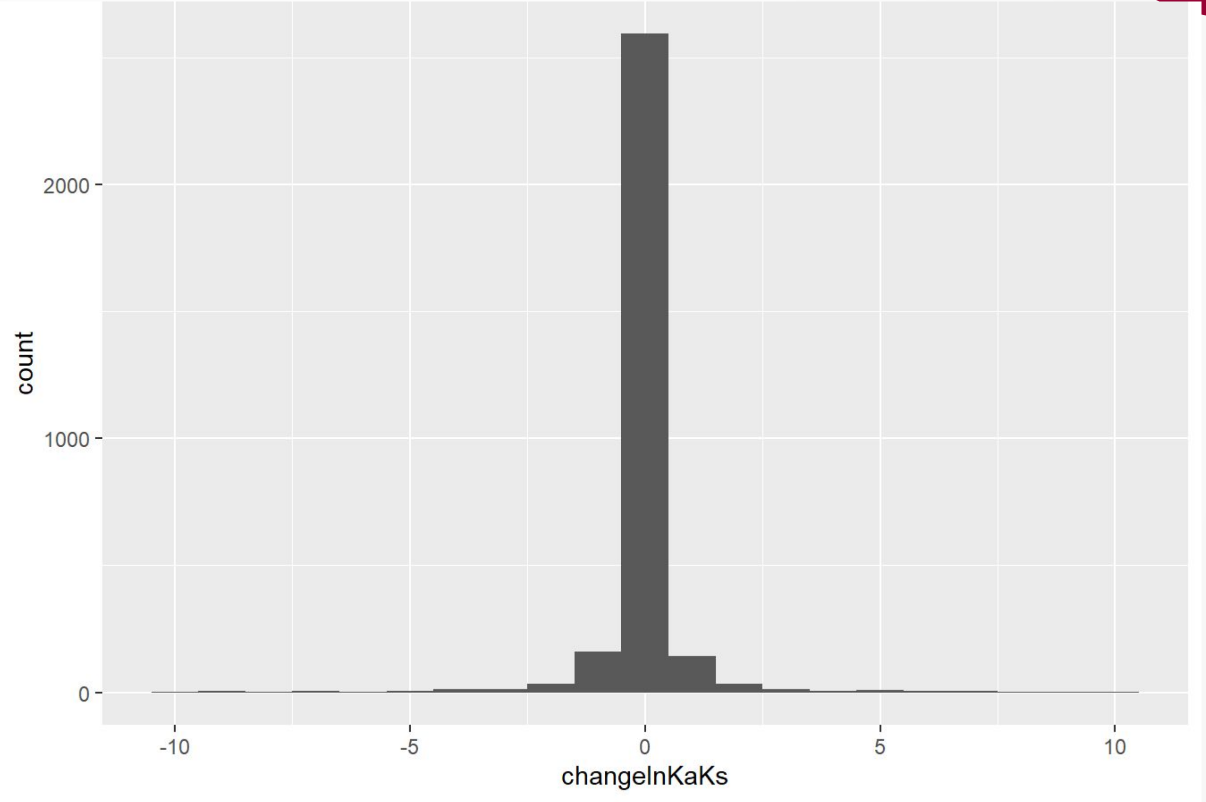
Variant impacts through time

Filtered out low-support variants

Calculated mean non-synonymous to synonymous variants per gene = base rate

Compared that to ratio of syn:non-syn variants over time

Most genes show **little change** in this ratio across decades

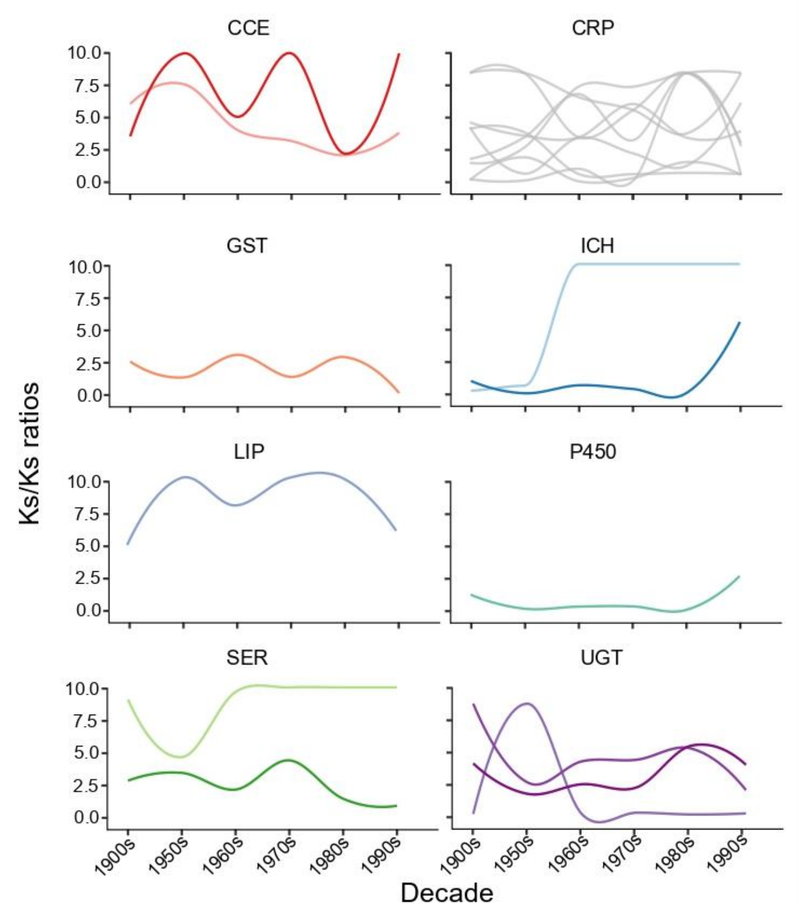


Temporal changes in AF: correlated

Variants with a **large amount** (±2-fold change) of change in this ratio for at least one sequential decade

21 genes, eight gene families

CRPs highest (n=9)



Key points

No clear population structure within Australia

Highest outliers in 60s and 80s (i.e., DDT/SPs)

Major outlier AF changes over time

Alternating selection – **antagonistic pleiotropy**?

21 interesting genes with major AF shifts (in missense:silent MAF ratio): function in resistance?

Acknowledgements

Museums

DAF (WA), QDPI, ASCT (NSW), MV, ANIC

Collaborators

Moritz lab @ ANU
Colleagues @ CSIRO
Leo Featherstone
Andy Bachler
Eli Parivzi

Funders

ARC (DECRA)

UoW

Thanks!

Do you have any questions?

amcgaugh@waikato.ac.nz
www.ang-mcgaughran.com
www.invasomics.com

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**