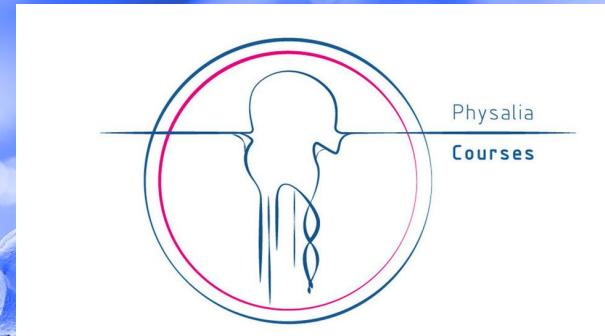


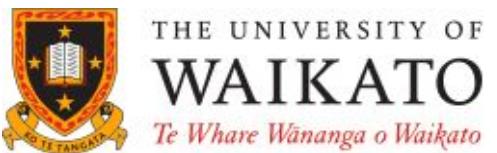
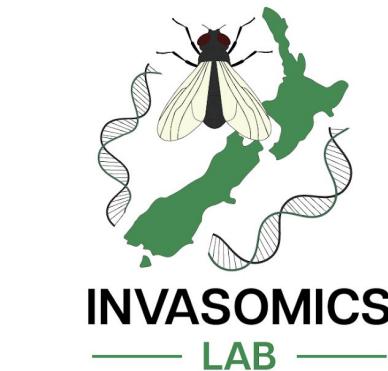
Invasion Genomics



Who I am

Dr Ang McGaughran
PI of the Invasomics Lab,
University of Waikato

Expertise:
Invasion genomics!
Invertebrate ecology



Who are you?

Pop into the slack channel:

Your name

Where you are zooming in from

Your career stage



Workflow

Day One: Movement dynamics

Day Two: Adaptation dynamics

Day Three: Environment dynamics



Workflow

Day One: Movement dynamics

Day Two: Adaptation dynamics

Day Three: Environment dynamics

Each day: Start with presentation, then move into R



Support

Dedicated slack channel – ask questions there!

My replies might come at odd times; thanks to
Carlo for help there also!



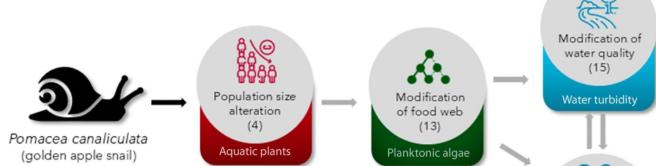
Invasive species

A species that has moved to an area outside of the area *in which it evolved*, and there causes harm to the environment, economy, or society

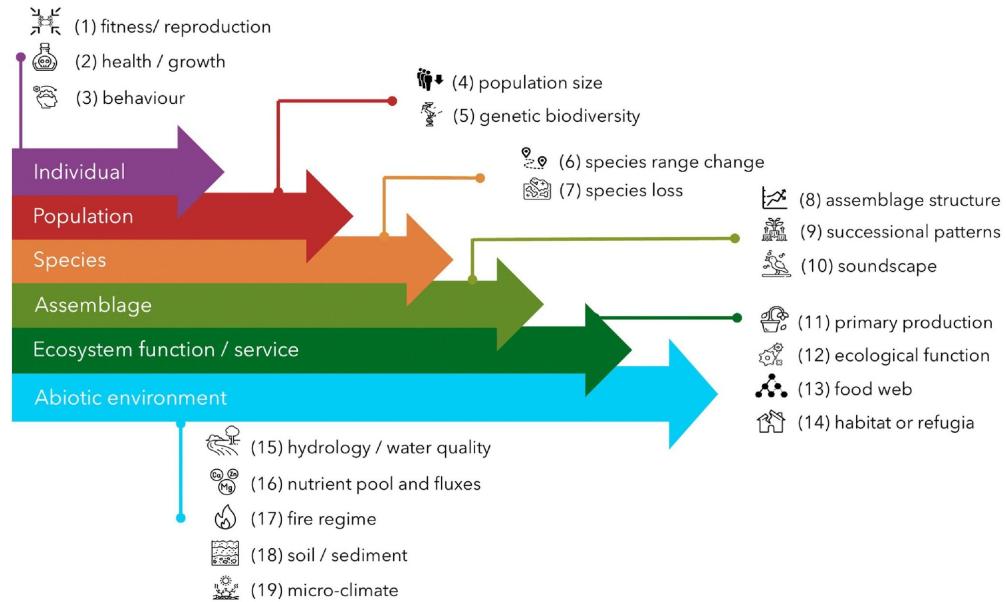
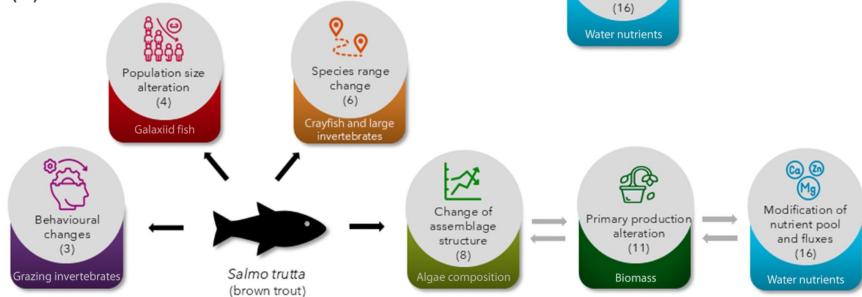


Invasion impact

(A)

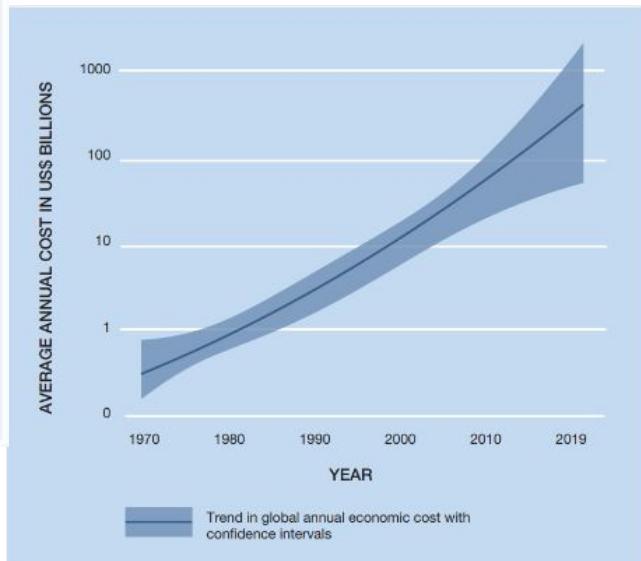
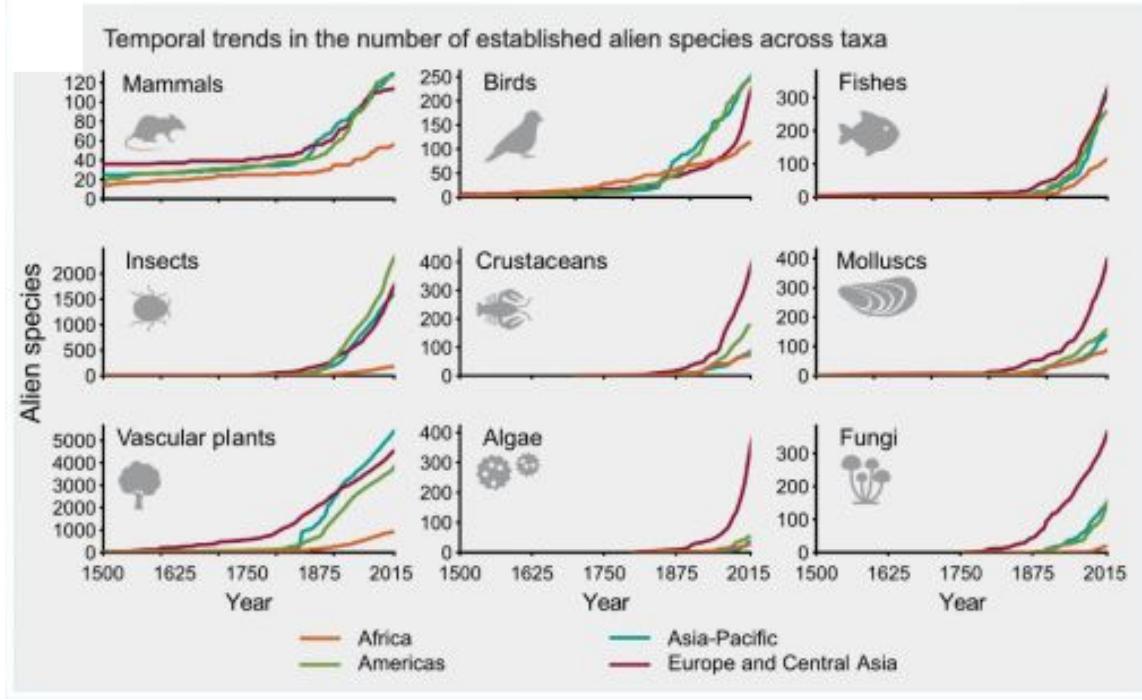


(B)

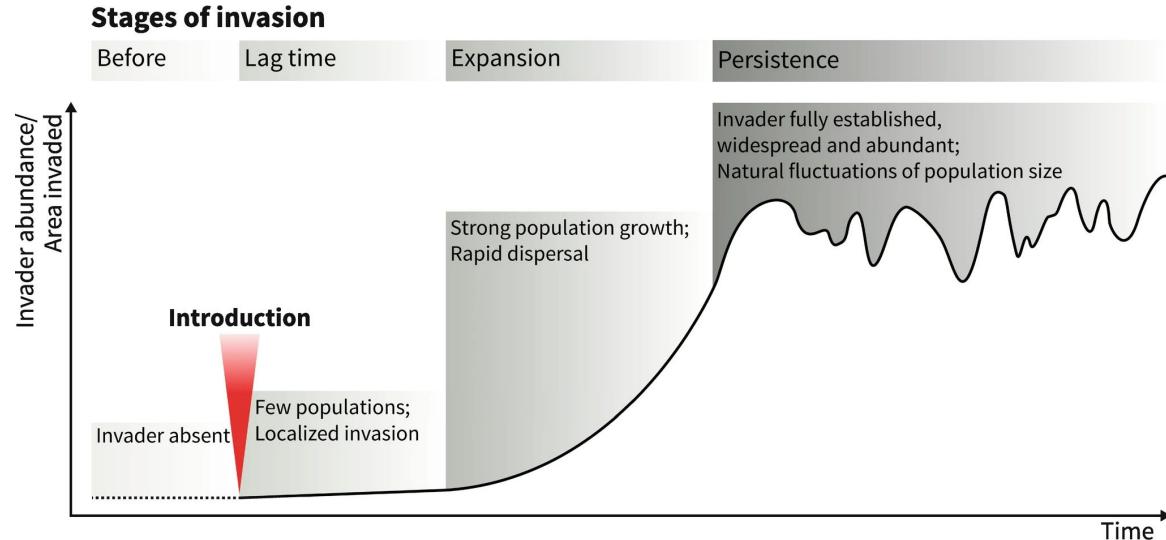


Trends in Ecology & Evolution

Invasion trends



Invasion curve



Stages of management

Prevention	Eradication	Containment	Long-term management
Control of vectors and pathways	Measures for eradication might be successful	Prevention of further spread; Public awareness typically begins	Minimise impact of invader; Protection of native species and resources

Movement Dynamics



Day One focus

Inferring:

Population structure

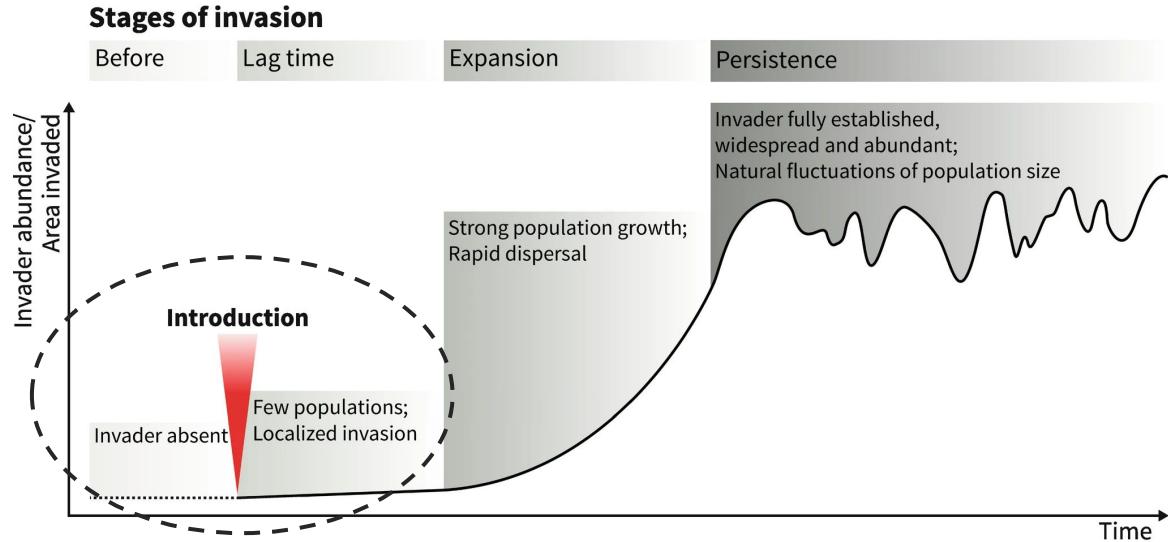
Invasion sources/pathways

Introduction to invasion genomics and the key kinds of research questions and methods we use

Overview of the tutorial dataset



Invasion curve



Stages of management

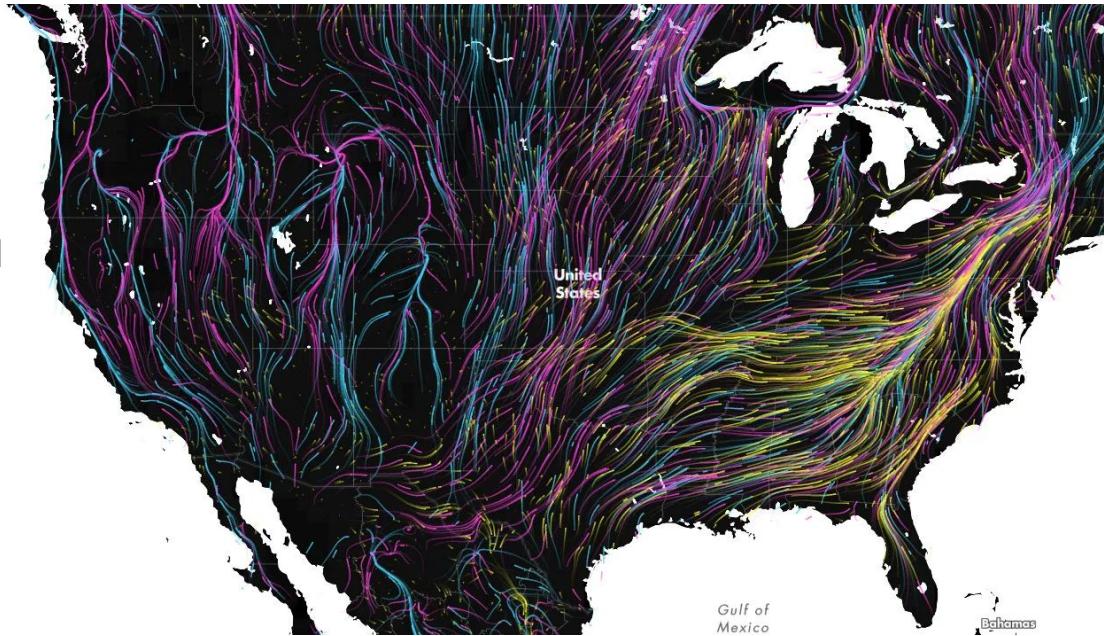
Prevention	Eradication	Containment	Long-term management
Control of vectors and pathways	Measures for eradication might be successful	Prevention of further spread; Public awareness typically begins	Minimise impact of invader; Protection of native species and resources

Movement dynamics

Why should we care about movement in invasion biology?

How can we use genomic data to tell us about movement?

MOVEMENT



Migration in motion: Visualising species movements due to climate change

<https://blog.nature.org/2016/08/19/migration-in-motion-visualizing-species-movements-due-to-climate-change/>

Genomic data

What kinds of questions can we ask?

Origin and source populations: Where did the invaders come from?

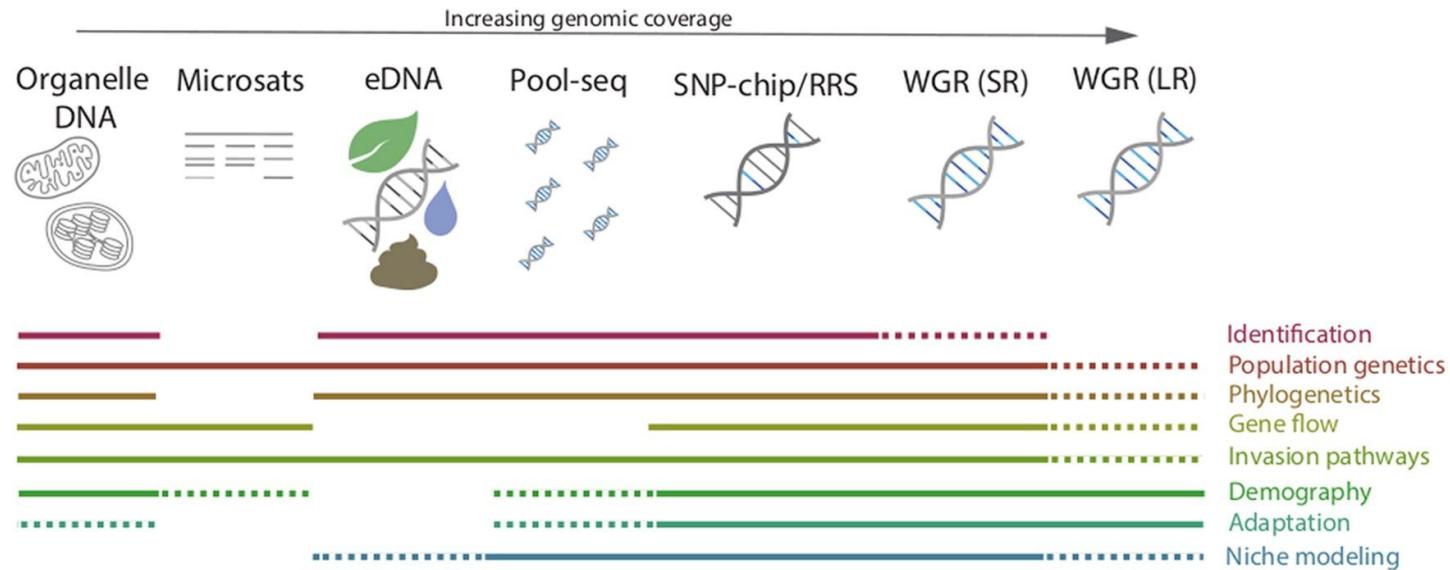
Number of introduction events: Single vs multiple introductions?

Genetic bottlenecks and founder effects: How did the introduction affect genetic diversity?

Cryptic invasions: Can we detect hidden or morphologically similar species?

Genomic tools

What kind of data can we use?



Common analyses

Population structure:

Populations exist across geographic space

Quantifying population structure tells us how different populations are

How much gene flow is occurring



F_{ST}

Ranges from 0 to 1; a measure of genetic distance across populations

Proportion of total observed allele frequency variance that is attributable to allele frequency differences between two groups, averaged across genome-wide SNPs (though note that other definitions of F_{ST} exist)

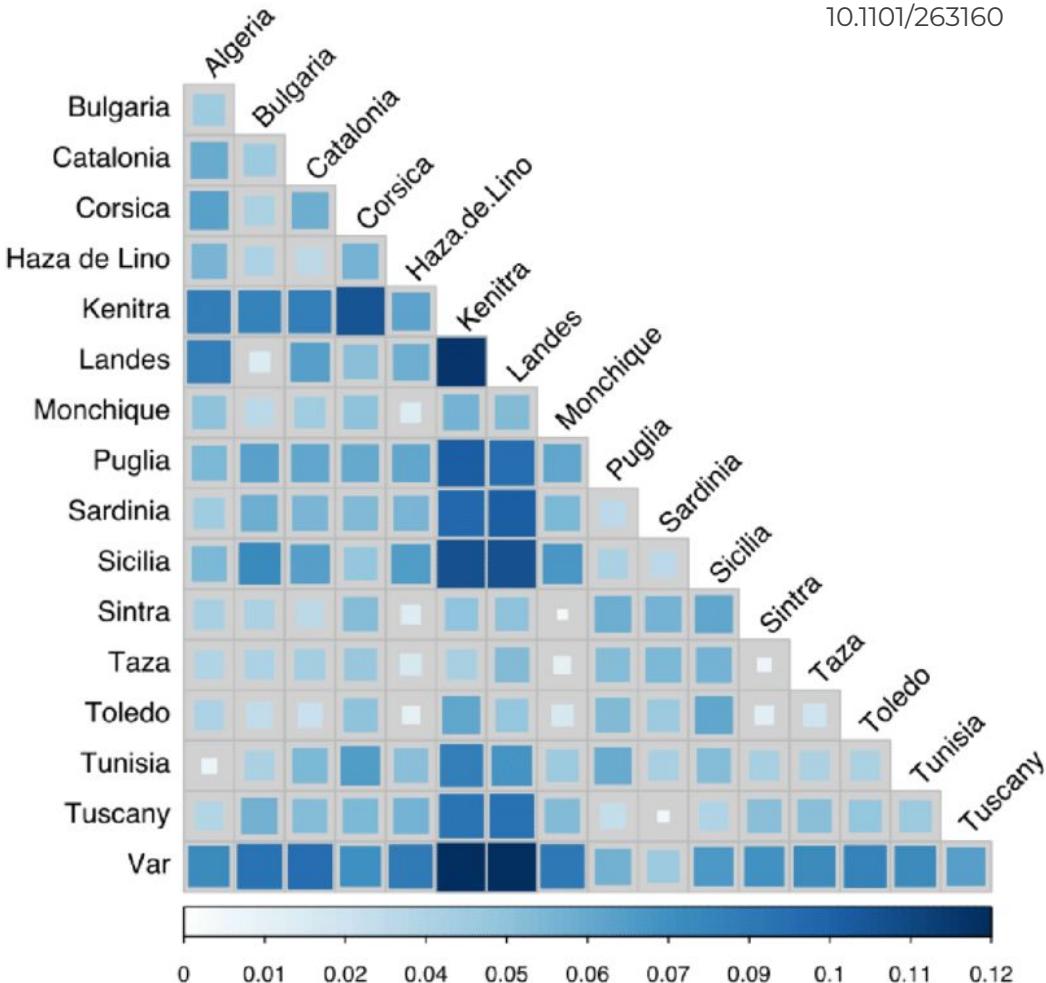
0 = no structure, freely exchanging genes, one large metapopulation

1 = full structure, populations fixed for alternate alleles, no gene flow

***Tractable with traditional population genetic tools, but genomic scale provides greater statistical power and precision from less individuals

Fst

Example shows ...



F_{ST}

Like PCA (next), F_{ST} is sensitive to sample size and number of SNPs, also potentially affected by SNP array ascertainment

Different demographic processes can lead to similar F_{ST} patterns

Common analyses

Visualising population structure

PCA

Bayesian clustering methods

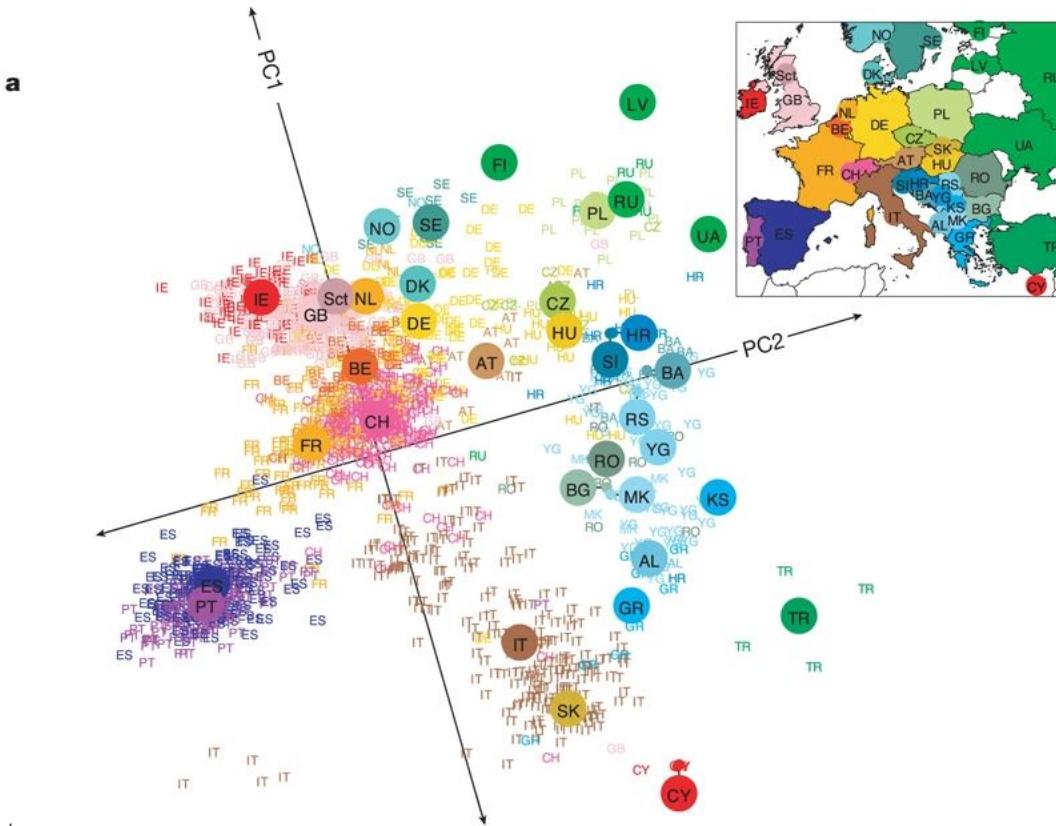
Illustrate the results of processes that affect the whole genome, e.g., population size, migration rates

Considered in a landscape context, can identify dispersal barriers, areas of high/low gene flow



PCA

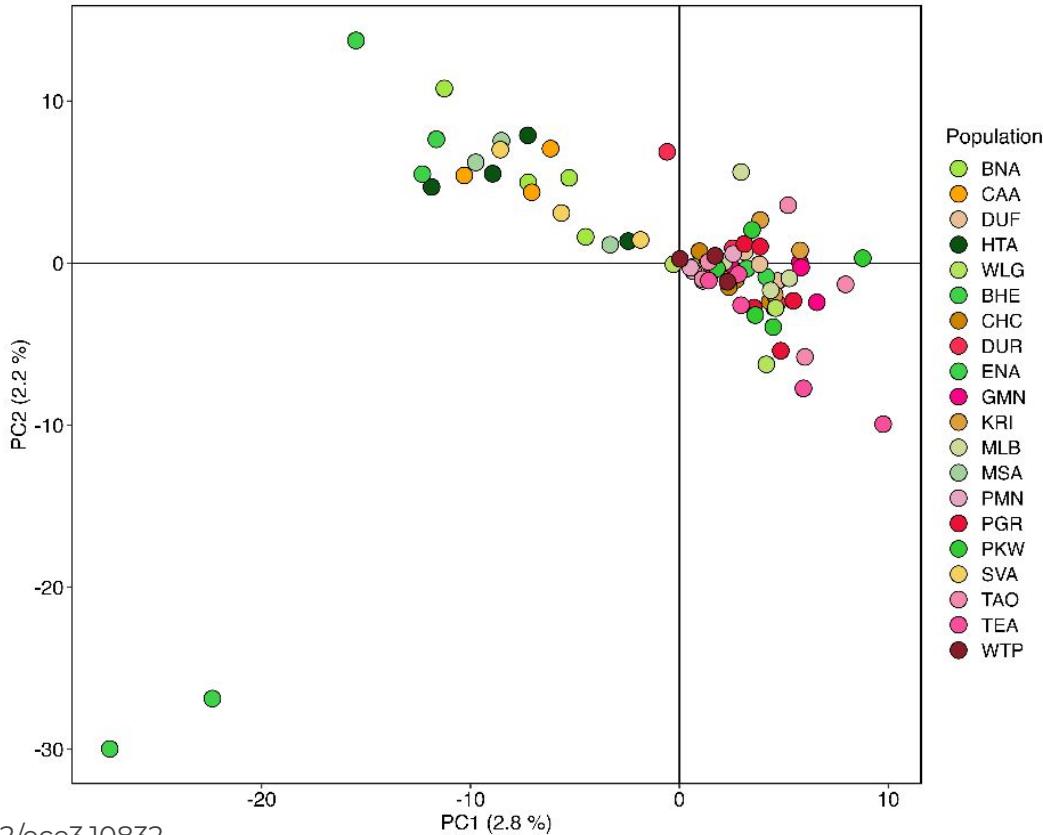
MOVEMENT



<https://doi.org/10.1038/nature07331>

PCA

MOVEMENT



PCA

PCA efficiently summarises key patterns in the data in a few dimensions

But, interpreting the underlying factors leading to these patterns can be challenging:

Recently isolated groups in a populations' history can have a big influence; populations with larger sample sizes can make up larger proportions of the overall variation; batch effects in genotyping (e.g., when combining data from multiple sequencing runs); ascertainment bias

PCA assumes that SNPs are independent (e.g., unlinked) → removing those sites can reduce power compared to methods that can retain them

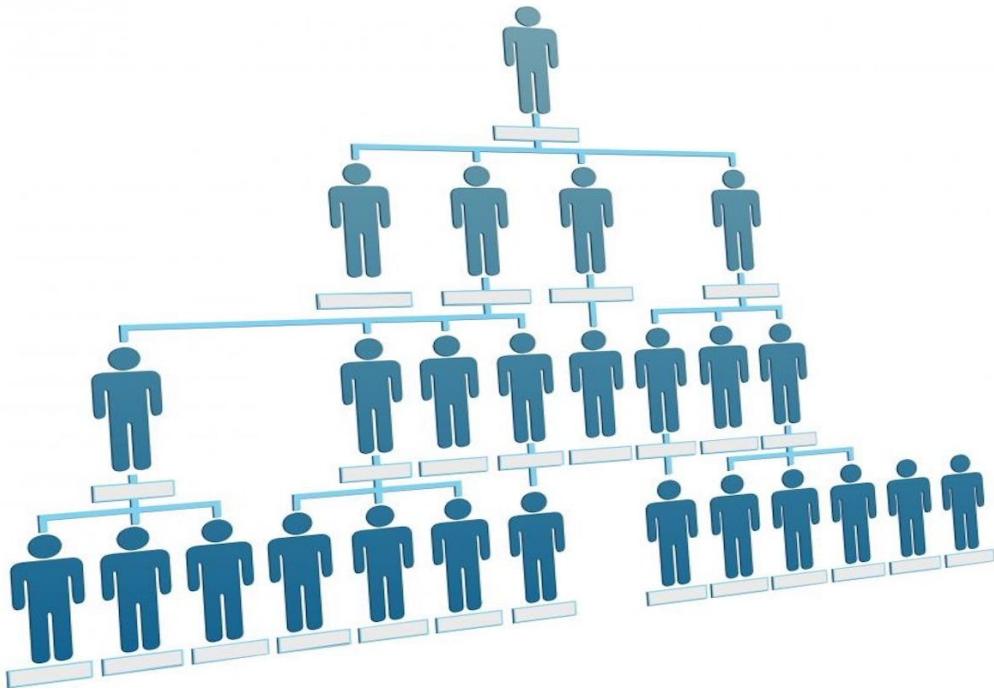
Bayesian clustering

Many methods:

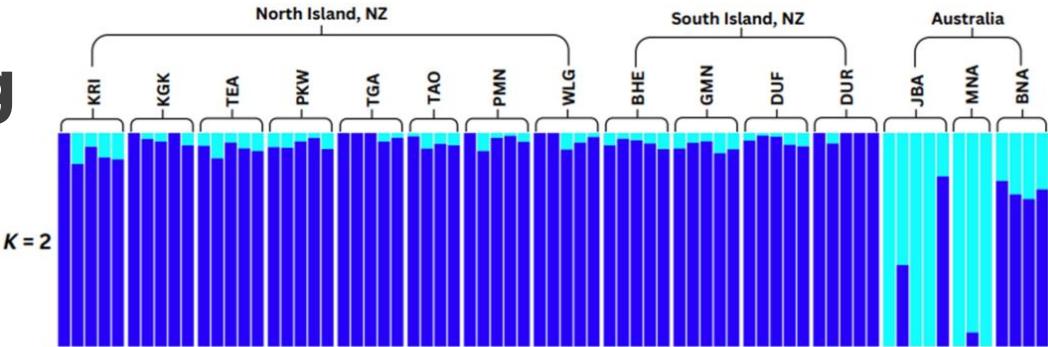
sNMF, admixture, STRUCTURE, ...

Based on allele or haplotype sharing + LD

Infers number of genetic clusters in the data *and* the shared ancestry between individuals



Bayesian clustering



Bayesian clustering

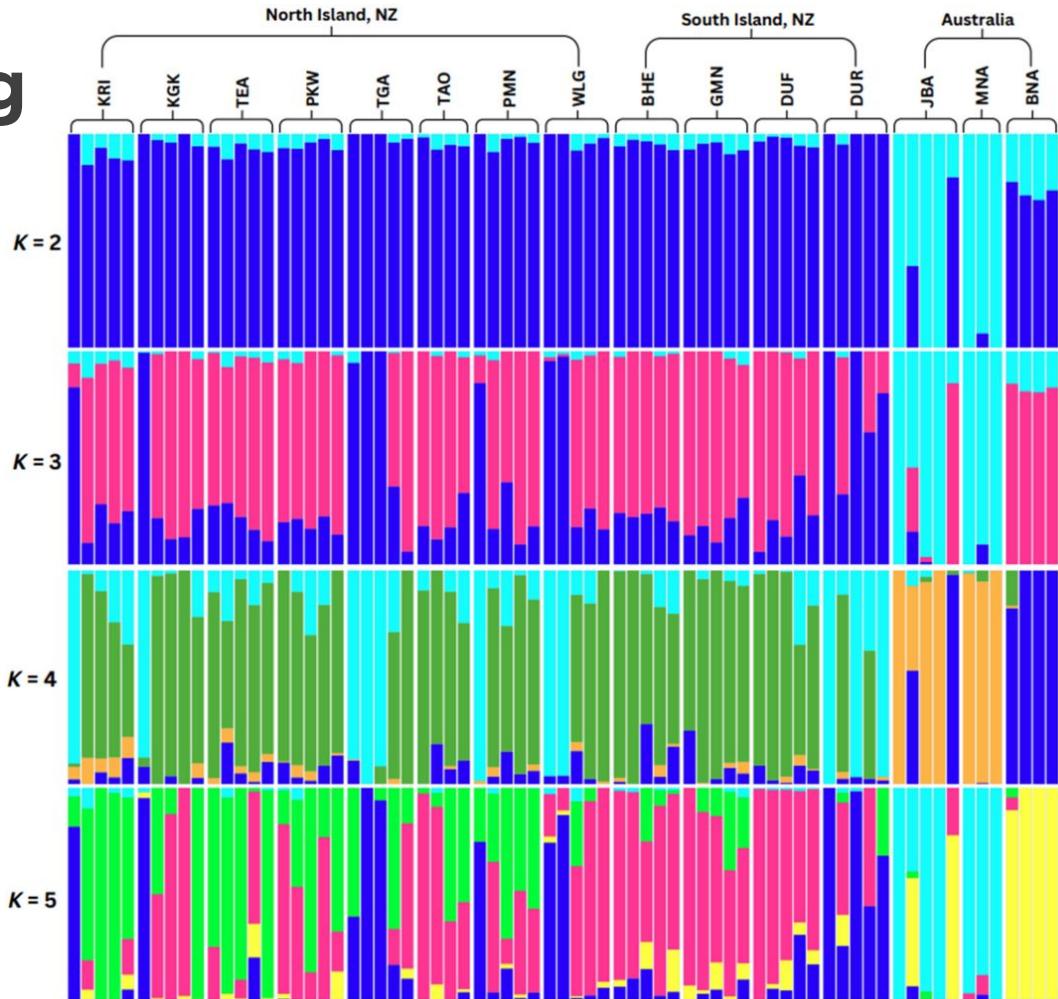
$K = 2$ optimal, but in practice we tend to look at a range of values

What does this plot show?

Clear similarity among invasive populations - no structure; high gene flow

Hint of BNA derivation?

<https://doi.org/10.1002/ece3.10832>



Bayesian clustering

Choosing appropriate K values can be challenging – though common to relate to geography, unlikely that any given K has a true biological interpretation

Results can change depending on the set of SNPs, composition of the individuals in the dataset, and K value

E.g., some individuals might be placed in clusters that reflect isolation relative to other samples versus true admixture patterns

Population assignment

As for other methods, can address diverse questions from measuring population connectivity to detecting dispersal and migration events

Model-based clustering methods like STRUCTURE designed to infer population structure, can also probabilistically assign individuals to inferred populations

Bayesian approaches assign individuals based on posterior probability they originate from each putative source population given their genotype – individual assigned to population with highest posterior probability or considered a migrant

Also ML approaches (e.g., R's assignPOP, extends DAPC)

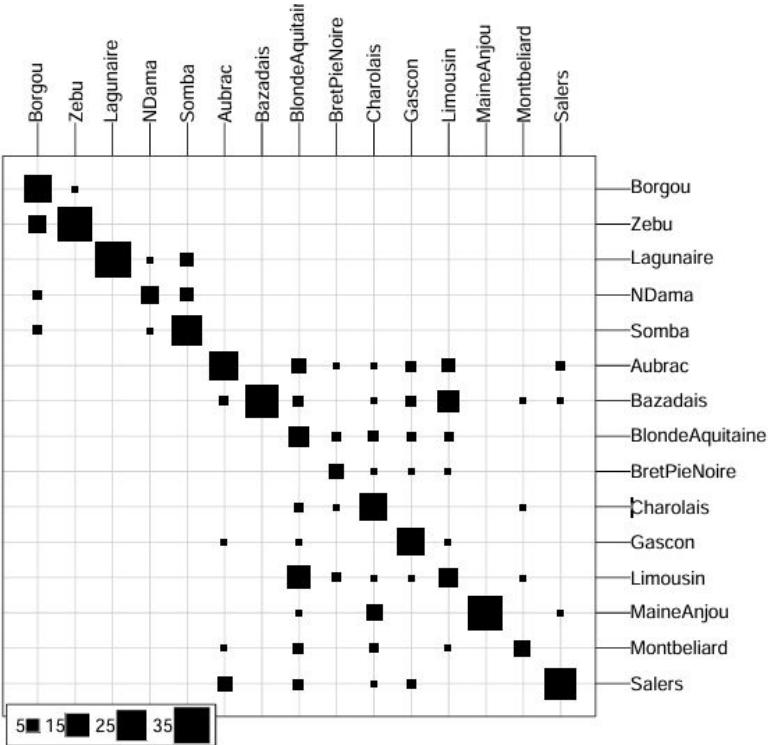
Population assignment

DAPC

K-means to infer number of clusters based on maximising variation between groups and build model; then predict population membership for samples not included in the training dataset

Columns = actual clusters of predicted individuals, rows = their inferred clusters

Groups fairly well retrieved, but some individuals not so good



Case study 1

Evolutionary Applications

Open Access



ORIGINAL ARTICLE

Open Access



Tracking the North American Asian Longhorned Beetle Invasion With Genomics

Mingming Cui, Amanda D. Roe , Brian Boyle, Melody Keena, Yunke Wu, W. Evan Braswell, Michael T. Smith, Ben Gasman, Juan Shi, Marion Javal, Geraldine Roux ... [See all authors](#)

First published: 19 November 2024 | <https://doi.org/10.1111/eva.70036> | Citations: 2

Case study 1

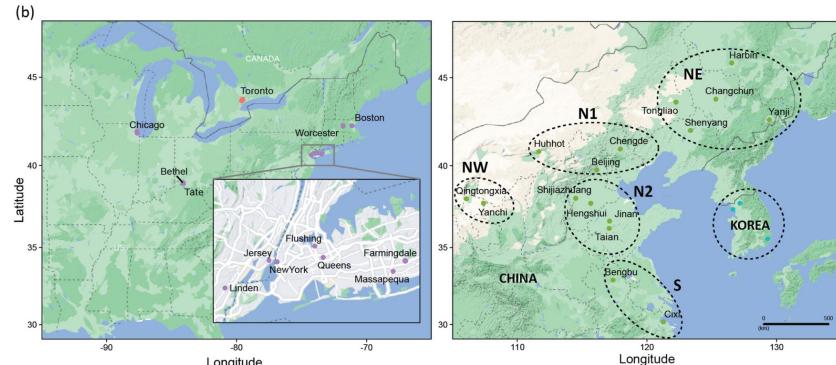
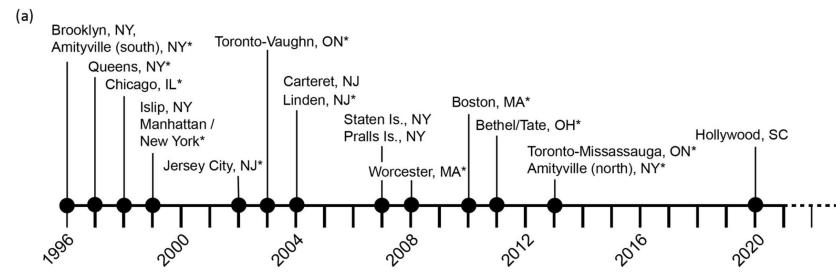
Asian longhorned beetle (*Anoplophora glabripennis*)

Damage to forest ecosystems

Effective management requires knowledge of invasion history

2,768 genome-wide SNPs

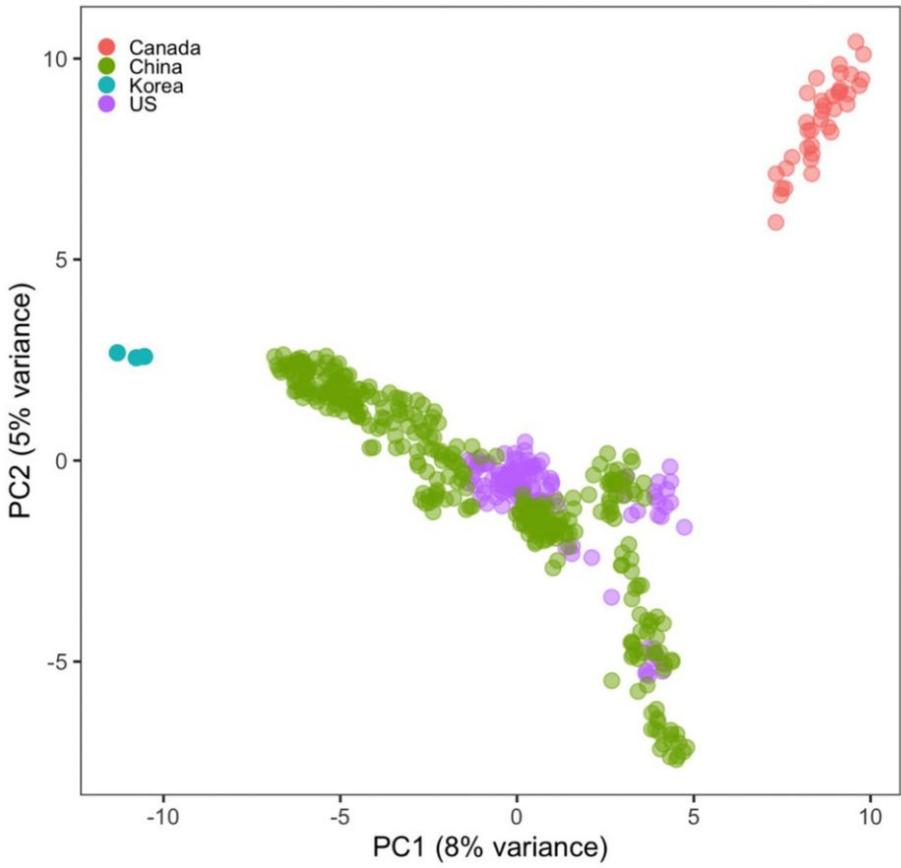
Focus on populations in N. America



Case study 1

Native populations show distinct regional clusters

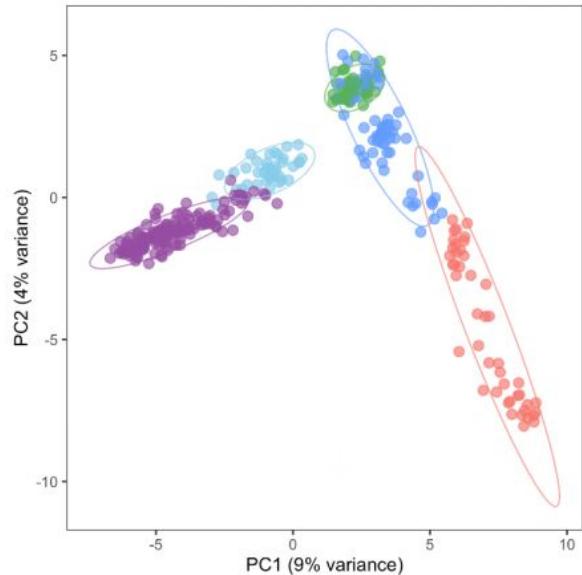
US samples grouped within Chinese regions



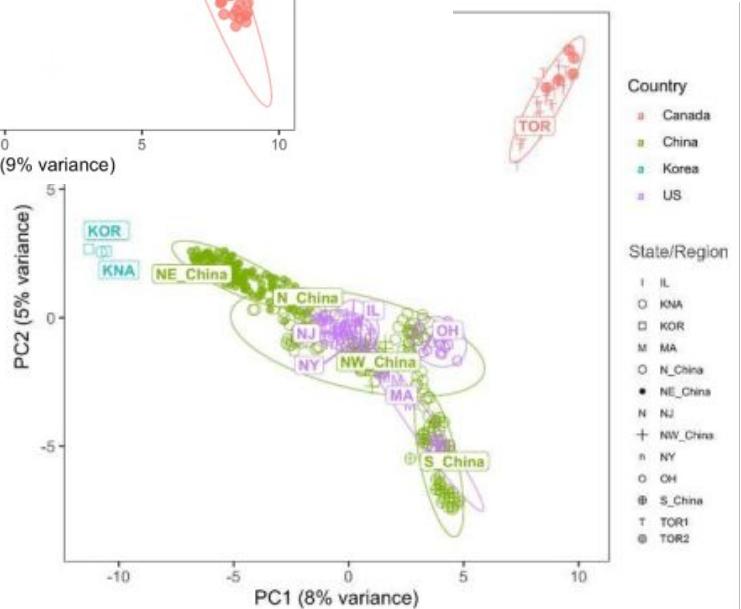
Case study 1

Native populations show distinct regional clusters

US samples grouped within Chinese regions

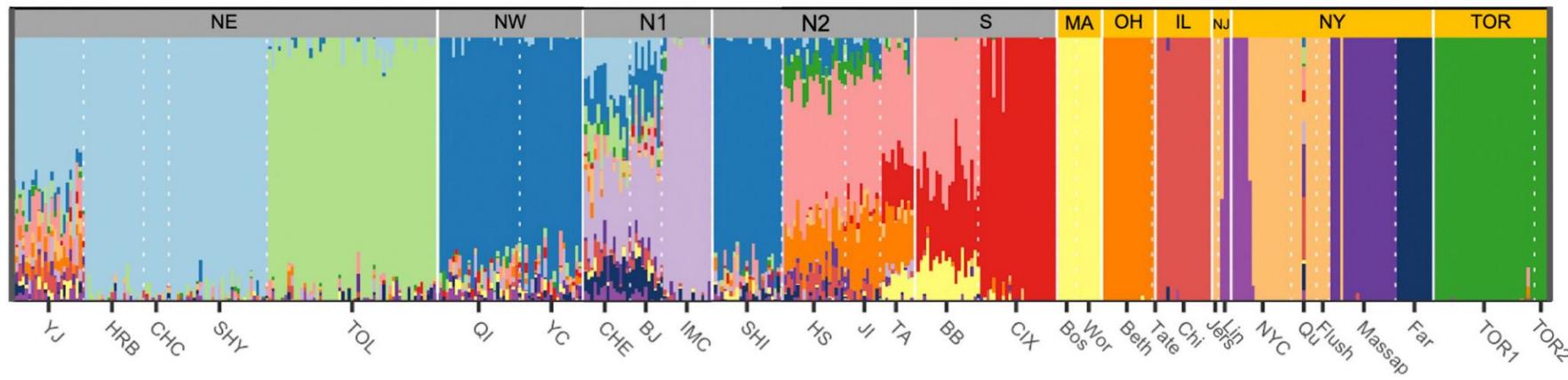


The different native pops



Case study 1

K = 14 optimal

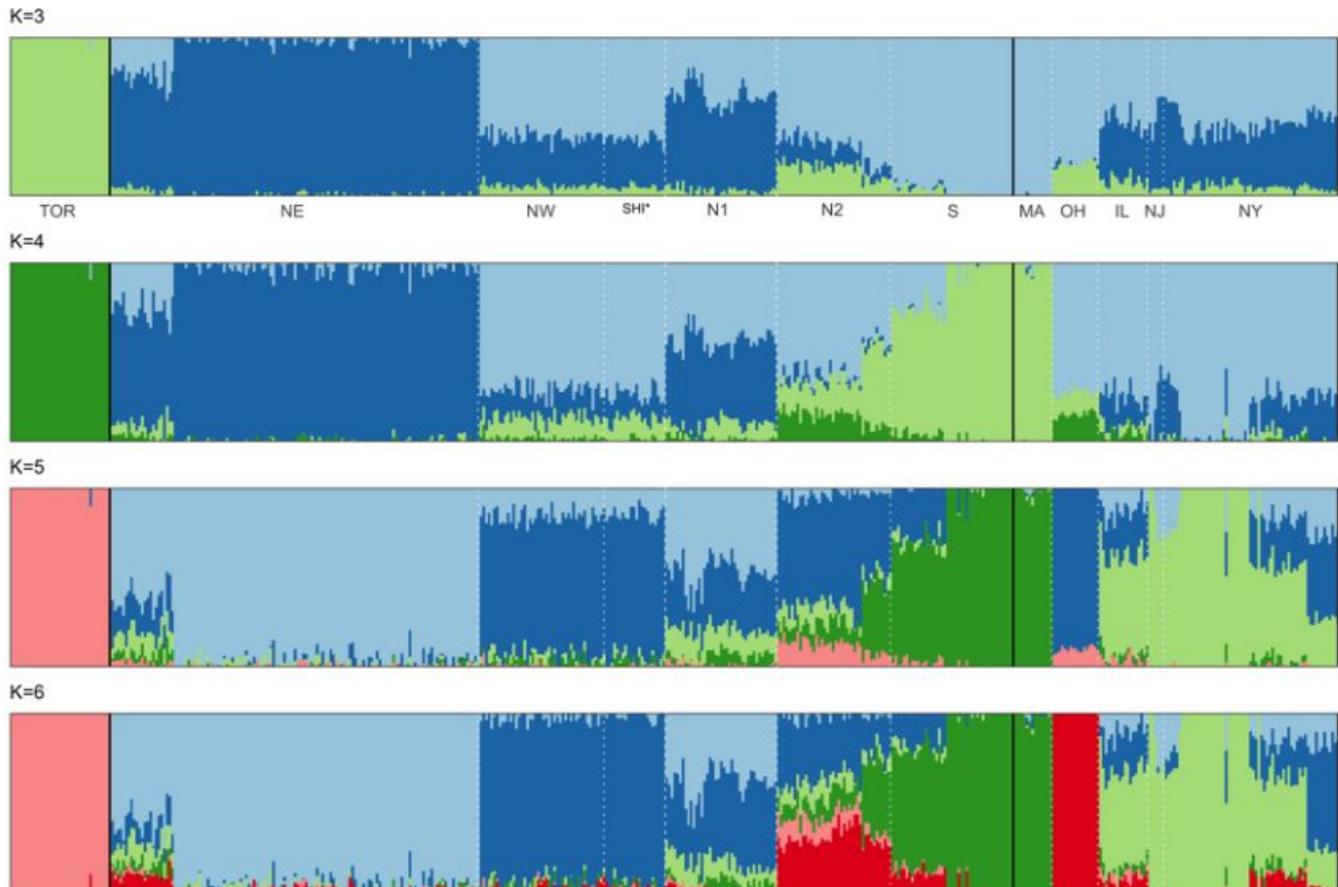


Case study 1

TOR (Canada)
very different

Also OH, IL

Greater mixing in
NY, NJ → NY
oldest (1999-2013),
more sources?

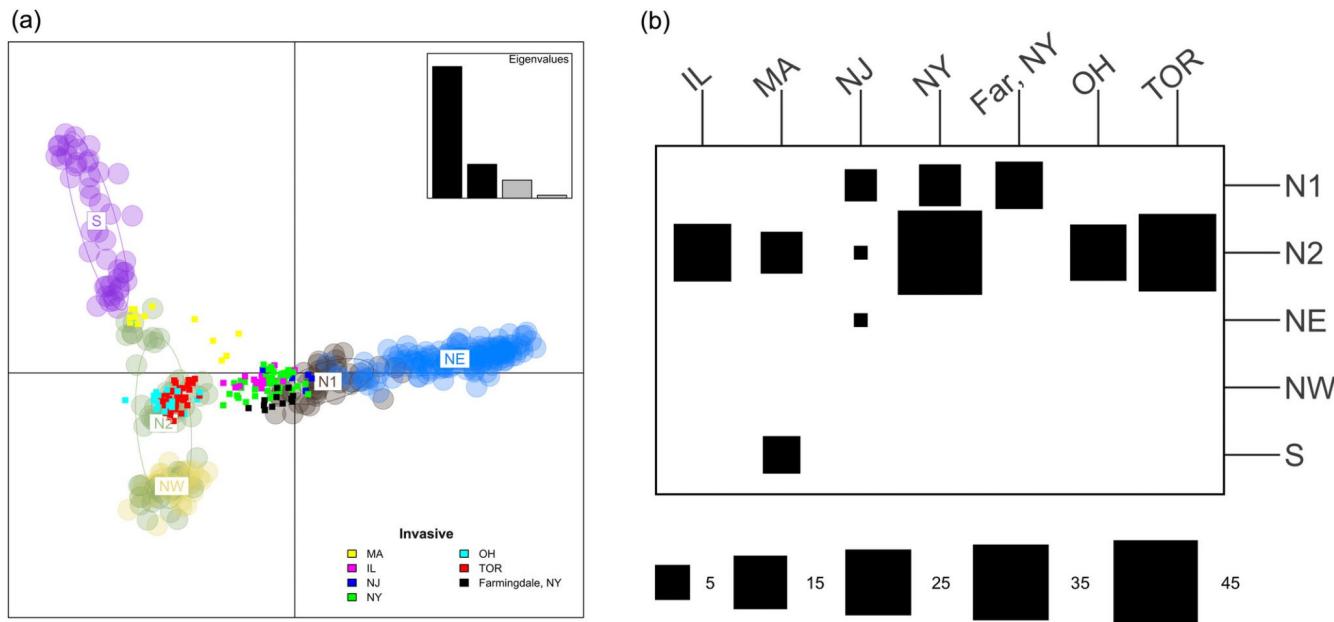


Case study 1

Population assignment

Used DAPC to assign invasive populations to native clusters

Most invasives assigned to N2 but also N1; NJ individuals from three sources



Case study 1

Multiple independent introductions from the native range

Traced invasion pathways and spread patterns

Used demographic modelling to infer post-introduction bottlenecks, followed by expansions

Now set up to quickly identify intercepted individuals, and examine other invasive populations → genomic biosurveillance



Case study 1

What kinds of questions can we ask?

Origin and source populations: Where did the invaders come from? 

Number of introduction events: Single vs multiple introductions? 

Genetic bottlenecks and founder effects: How did the introduction affect genetic diversity? 

Cryptic invasions: Can we detect hidden or morphologically similar species?

Case study 2

Cryptic invasions relatively common for plants and invertebrates, where taxonomic identification can be hard

Typically require molecular methods to identify

Case study 2

[Home](#) > [Biological Invasions](#) > Article

Genomics detects cryptic wallaby invasion

Invasion Note | [Open access](#) | Published: 24 August 2024

Volume 26, pages 3997–4003, (2024) [Cite this article](#)



[Download PDF](#) ↴

You have full access to this [open access](#) article

[Biological Invasions](#)

[Aims and scope](#) →

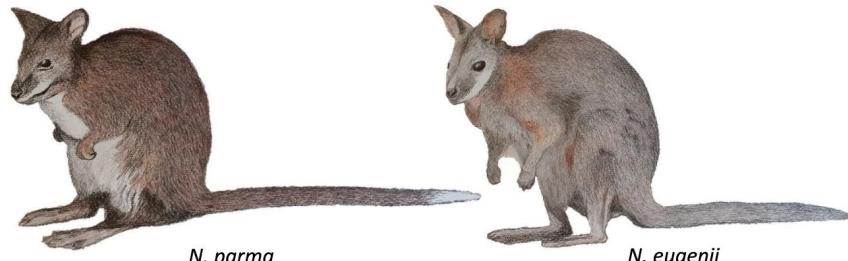
[Submit manuscript](#) →

[A. J. Veale](#) [J. C. McEwan](#), [R. Brauning](#), [T. Van Stijn](#) & [A. D. M. Latham](#)

1343 Accesses 2 Altmetric [Explore all metrics](#) →

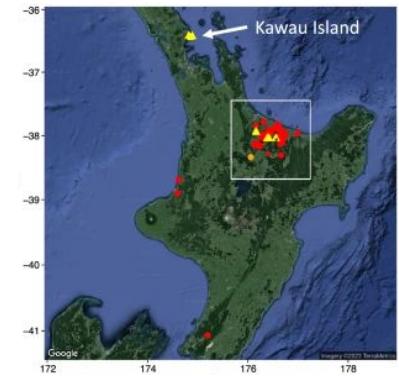
Case study 2

Invasive tammar wallaby



Around 1870, at least four species of wallaby introduced to Aotearoa New Zealand, on Kawau Island (2,070 ha) in the Hauraki Gulf

Two of the smaller species – parma wallaby (*Notamacropus parma*), tammar wallaby (*Notamacropus eugenii*) of similar size and colour



Records not well-kept, historians largely assumed all on the island were tammar

Spread of tammar wallabies to North Island ~1912, now large range (~4,000 km²) and damage to pasture, native plants, etc

Case study 2

GBS data (16,796 autosomal SNPs)



Mitochondrial DNA (~800 bp) for species ID

Case study 2

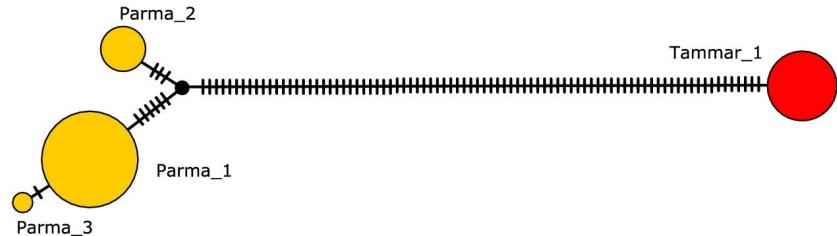
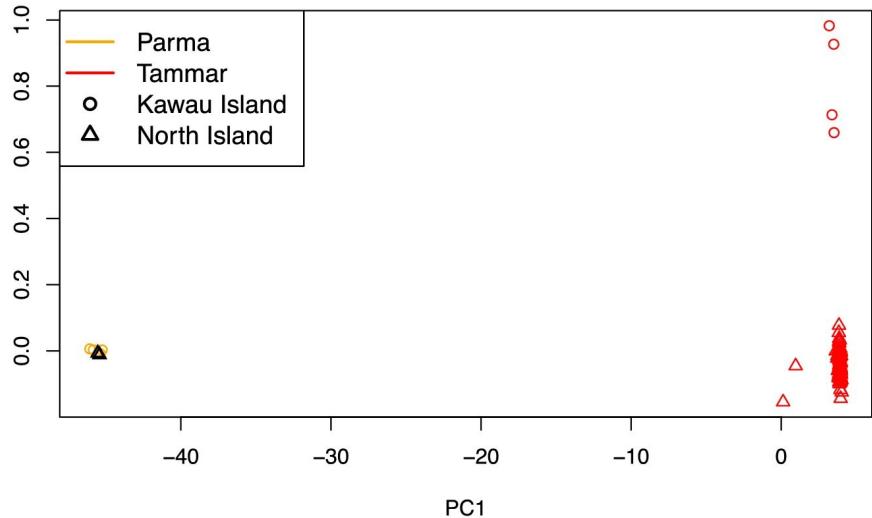
Two divergent species identified

58% of SNPs were fixed differences

No hybridisation

Also found three parma wallabies in the BoP (southeast) region (expected only tammar)

Further work on 183 other samples found another 17 parma species IDs



Case study 2

BoP parma population undetected for at least 30 years

Unclear origins



Long period without detection highlights need for greater monitoring efforts...

... Especially those that include genomic methods

Case study 2

What kinds of questions can we ask?

Origin and source populations: Where did the invaders come from?

Number of introduction events: Single vs multiple introductions?

Genetic bottlenecks and founder effects: How did the introduction affect genetic diversity?

Cryptic invasions: Can we detect hidden or morphologically similar species? 

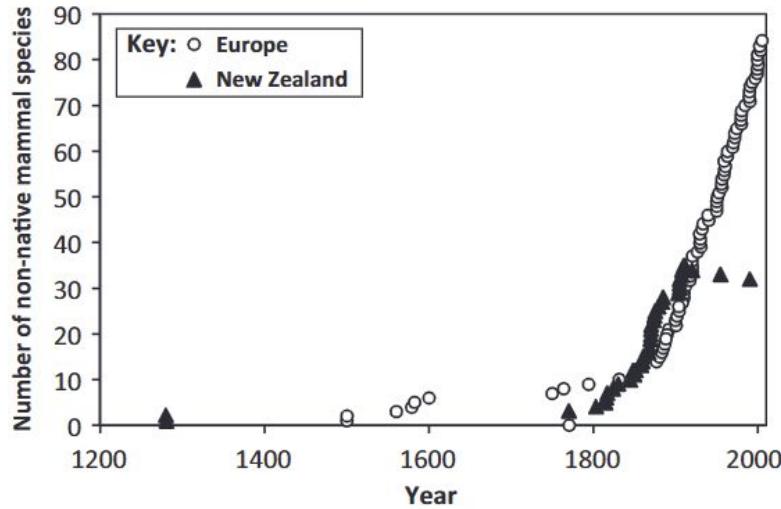
Genomic insights

Identifying invasion pathways

Detecting admixture: Mixing of different source populations can increase invasiveness

Predicting invasion success: Using genomic data to model potential spread and impact (e.g., based on GEA – Day Three)

...



Challenges and limitations

Sampling bias

Geographic coverage of the invasion is more important than the breadth of genomic coverage when only reconstructing invasion routes

Methodological biases

Some method assumptions (e.g., that populations are in migration-drift equilibrium → gene flow and F_{ST} estimates)

Demographic changes (e.g., founder/bridgehead effects) may weaken differentiation, impacting migration measures

Challenges and limitations

Distinguishing between neutral and adaptive variation is often an important first step

Ethical and logistical issues in global sampling

Everything is native (and possibly treasured) somewhere!

Data sovereignty

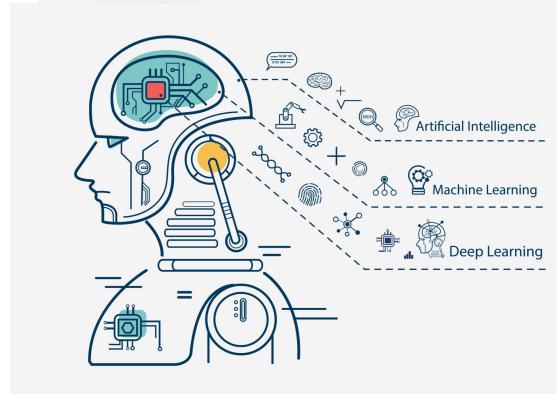
Future directions

Integration with ecological and environmental data

For example, iNaturalist and GBIF

Real-time genomic surveillance (e.g., portable sequencers)

Machine learning and AI in invasion genomics



Food for thought

Applied

Can genomic data help prevent invasions?

How might genomic tools be used in policy and management?

Fundamental

How does standing variation in different populations influence invasion success?

What role is admixture playing?



Conclusions

Introduction is a key stage of invasion

Genomic data can elucidate population structure, invasion history, and subsequent spread, which may aid management

Next steps

Course dataset

Heredity

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [heredity](#) > [articles](#) > article

Article | [Open access](#) | Published: 30 October 2023

Genomic signals of local adaptation across climatically heterogeneous habitats in an invasive tropical fruit fly (*Bactrocera tryoni*)

[Elahe Parvizi](#), [Amy L. Vaughan](#), [Manpreet K. Dhami](#) & [Angela McGaughan](#) 

[Heredity](#) 132, 18–29 (2024) | [Cite this article](#)

3998 Accesses | 6 Citations | [Metrics](#)

► Sci Rep. 2020 Jul 1;10:10788. doi: [10.1038/s41598-020-67397-5](https://doi.org/10.1038/s41598-020-67397-5) ↗

Genome-wide patterns of differentiation over space and time in the Queensland fruit fly

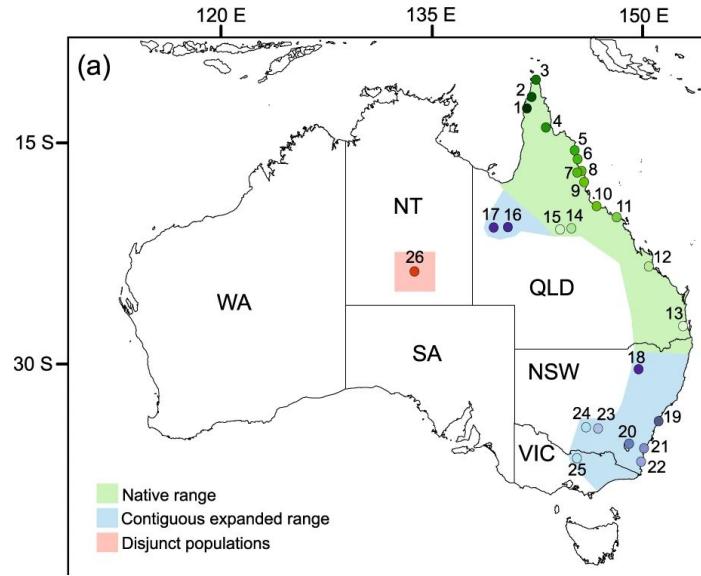
[Ángel-David Popa-Báez](#) ^{1,2,✉,#}, [Renee Catullo](#) ^{2,3,✉}, [Siu Fai Lee](#) ^{1,2}, [Heng Lin Yeap](#) ², [Roslyn G Mourant](#) ², [Marianne Frommer](#) ⁴, [John A Sved](#) ⁴, [Emily C Cameron](#) ⁵, [Owain R Edwards](#) ², [Phillip W Taylor](#) ¹, [John G Oakeshott](#) ^{1,2}



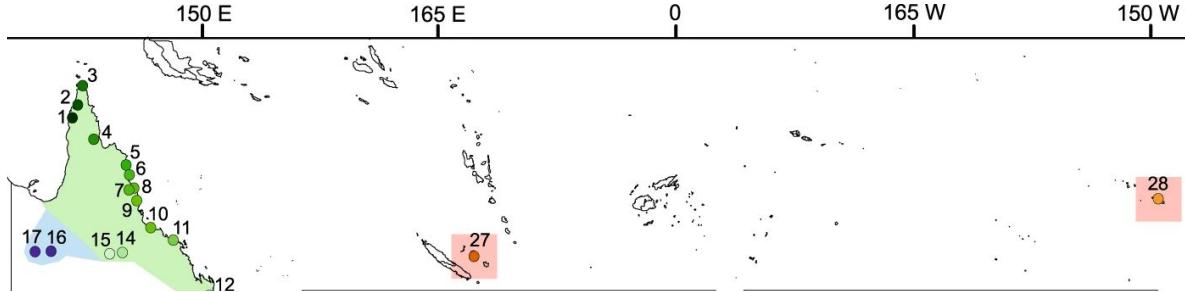
Next steps

Invasive tropical fruit fly (*Bactrocera tryoni*) -
Queensland fruit fly (Qfly)

Native to Australia; expanded range in
temperate and arid Australia; invasive in tropical
zones in the Pacific



28 populations, GBS



Next steps

We are not focusing today on how the genetic data file we'll be using was created (Qfly.vcf), but it is important to note that some filtering steps have been applied:

Biallelic SNPs only, MAF < 0.05, high LD and missing data removed, resulting in a final dataset of 6,707 SNPs; see full steps at:

https://github.com/Elahep/B.tryoni_PopGenomics/tree/main/1-variant-calling



***Depending on the analysis type, we may wish to remove or retain multi-allelic SNPs, set a lower MAF threshold (or higher), not remove high-LD sites, etc

Next steps

Please move to the Day One tutorial

Remember to ask questions via slack!

CREDIT: SlideEgg PowerPoint template

www.slideegg.com

