Genomic offset studies for bird species conservation

Bay *et al.* (2018)
https://doi.org/10.1126/science.aan4380

Miller *et al.* (2023)
https://doi.org/10.1111/mec.17199

Chen *et al.* (2022)
https://doi.org/10.1038/s41467-022-32546-z

What is genomic offset?

- Local adaptation is widespread in wild populations

- Within a species range, different populations are adapted to different environmental conditions due to local adaptation
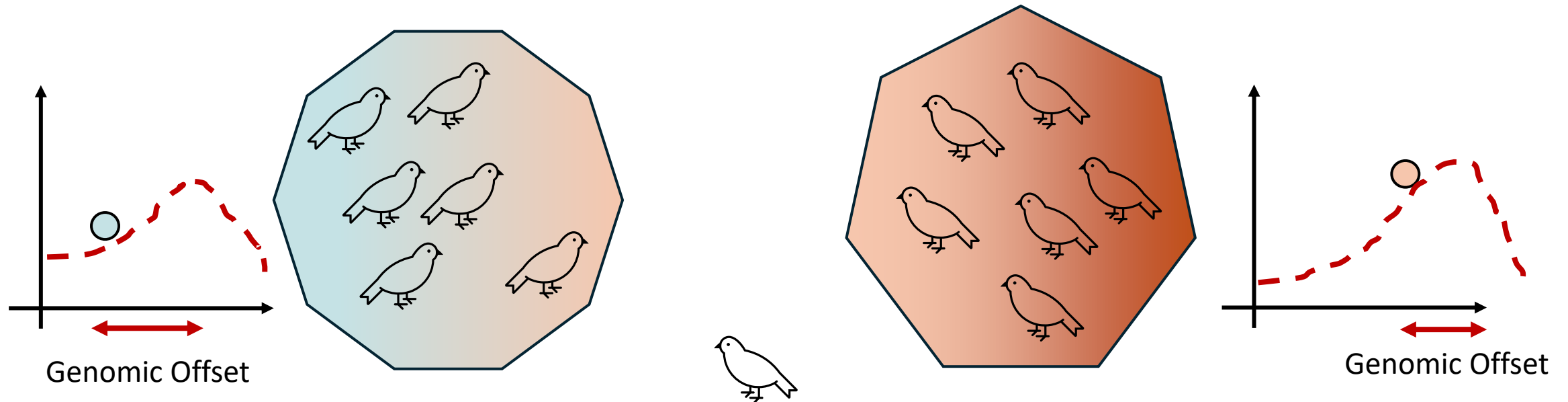
Manuel Bettencourt
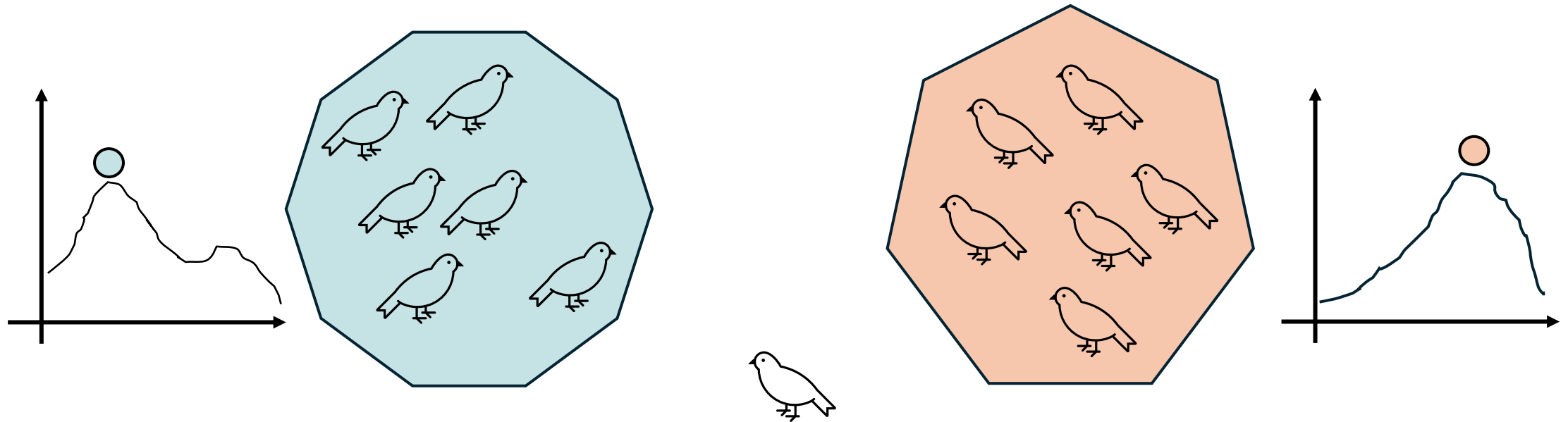
What is genomic offset?

- A changing climate disturbs the conditions to which populations are locally adapted, resulting in different levels of maladaptation
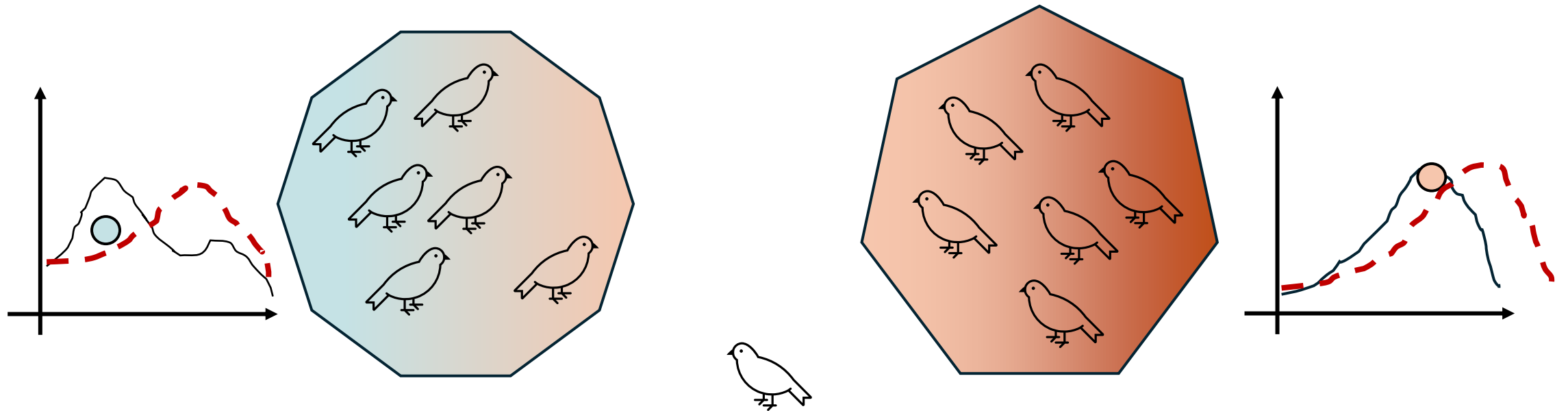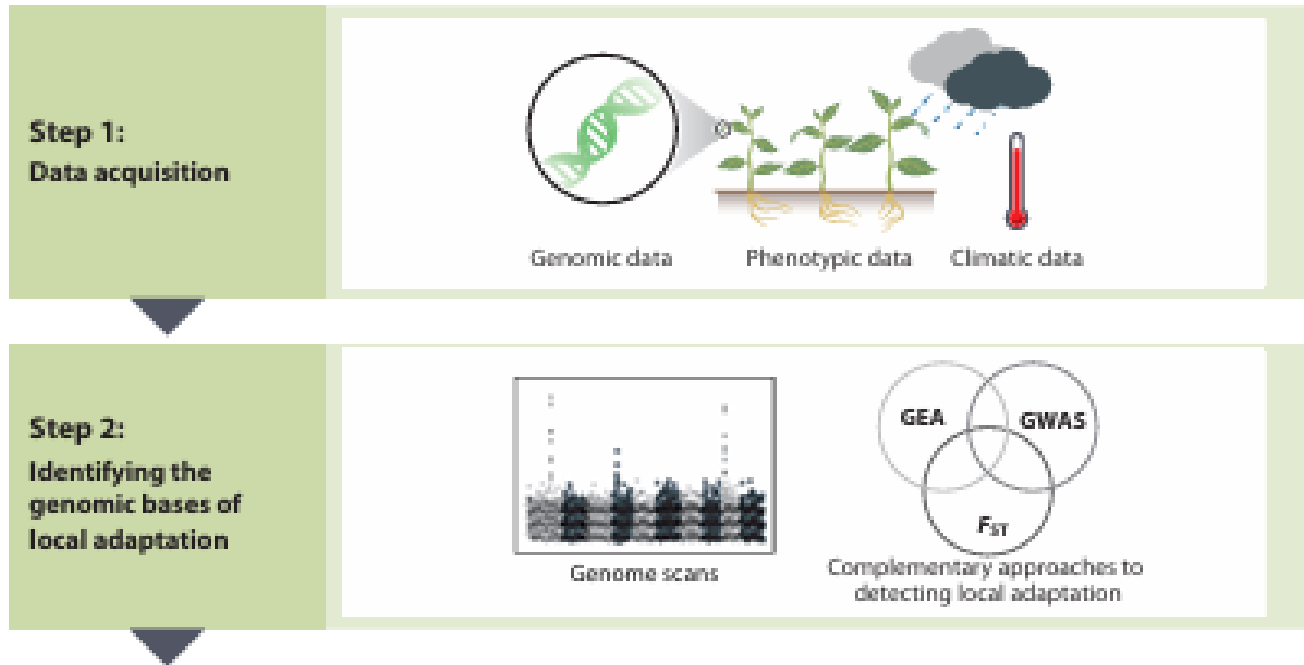
What is genomic offset?

- Genomic offset is a genomics-based measure of the level of maladaptation a species is predicted to suffer in a specific new environment
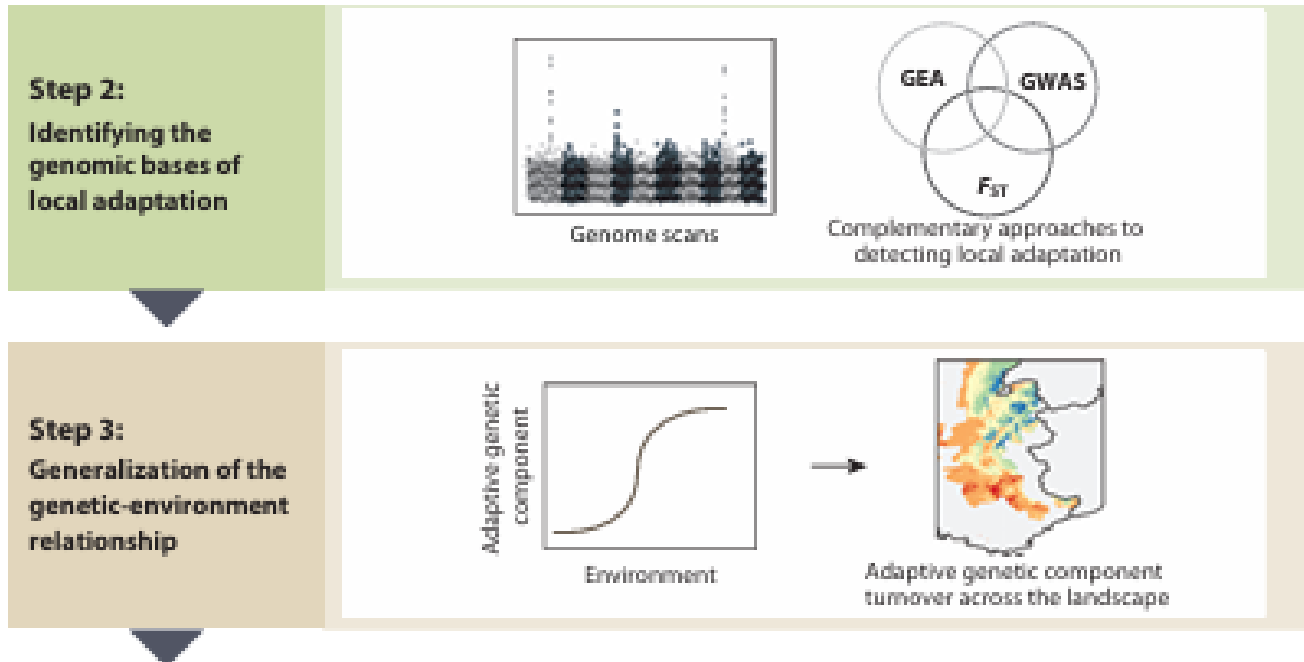

Genomic Offset

Genomic Offset

General workflow – discovery phase

Step 1:
Data acquisition

Genomic data   Phenotypic data   Climatic data

Step 2:
Identifying the genomic bases of local adaptation

Genome scans

GEA   GWAS
F_{ST}

Complementary approaches to detecting local adaptation

- Constructing a reference genome with whole genome sequencing methods

- Obtaining molecular polymorphism data

- Retrieving climatic data from public databases

Adapted from Capblancq et al. (2020). *Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes.*

General workflow – discovery phase

Step 2:
Identifying the genomic bases of local adaptation

Genome scans

Complementary approaches to detecting local adaptation

Step 3:
Generalization of the genetic-environment relationship

Adaptive genetic component

Environment

Adaptive genetic component turnover across the landscape

- Search for variants involved in local adaptation through **GEA – Genotype Environment Association** or F_{ST} based methods

- Now we have a collection of SNPs related to local adaptation

Adapted from Capblancq et al. (2020). *Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes.*

General workflow – prediction phase

Step 3:
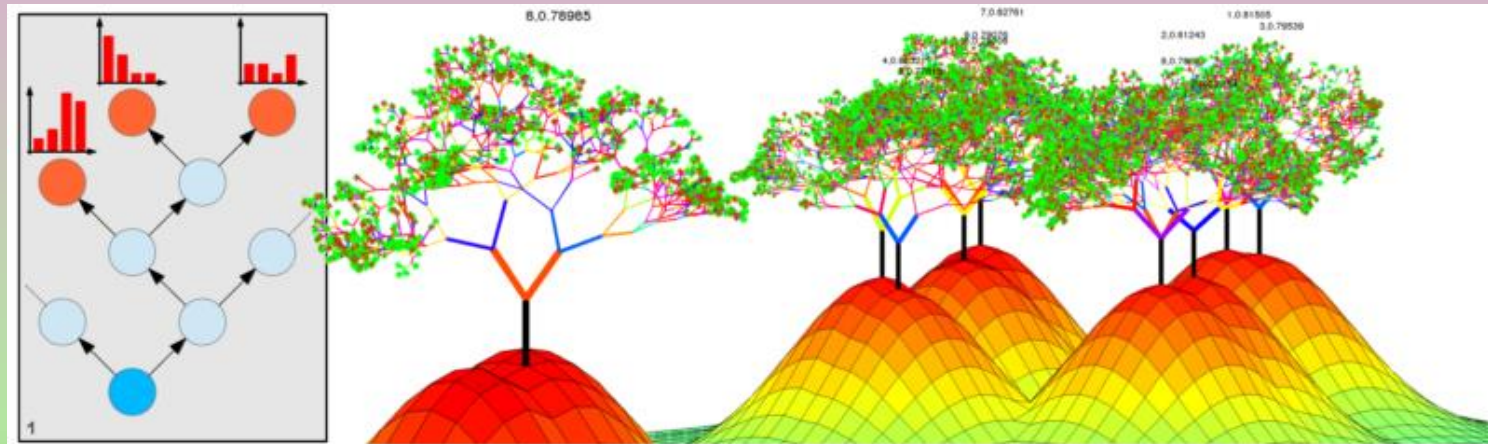Generalization of the genetic-environment relationship
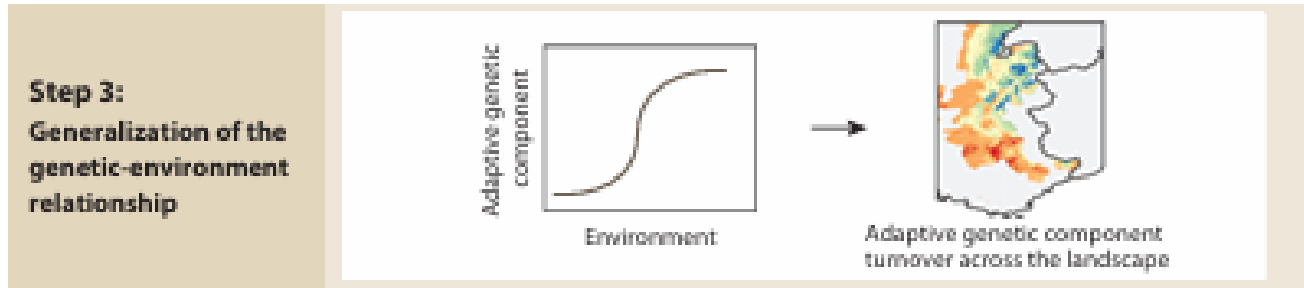
- Statistically model how the found SNPs vary with different environments

- This requires sophisticated statistical approaches based on machine learning classification methods and dimensionality reduction

Adapted from Capblancq et al. (2020). *Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes.*

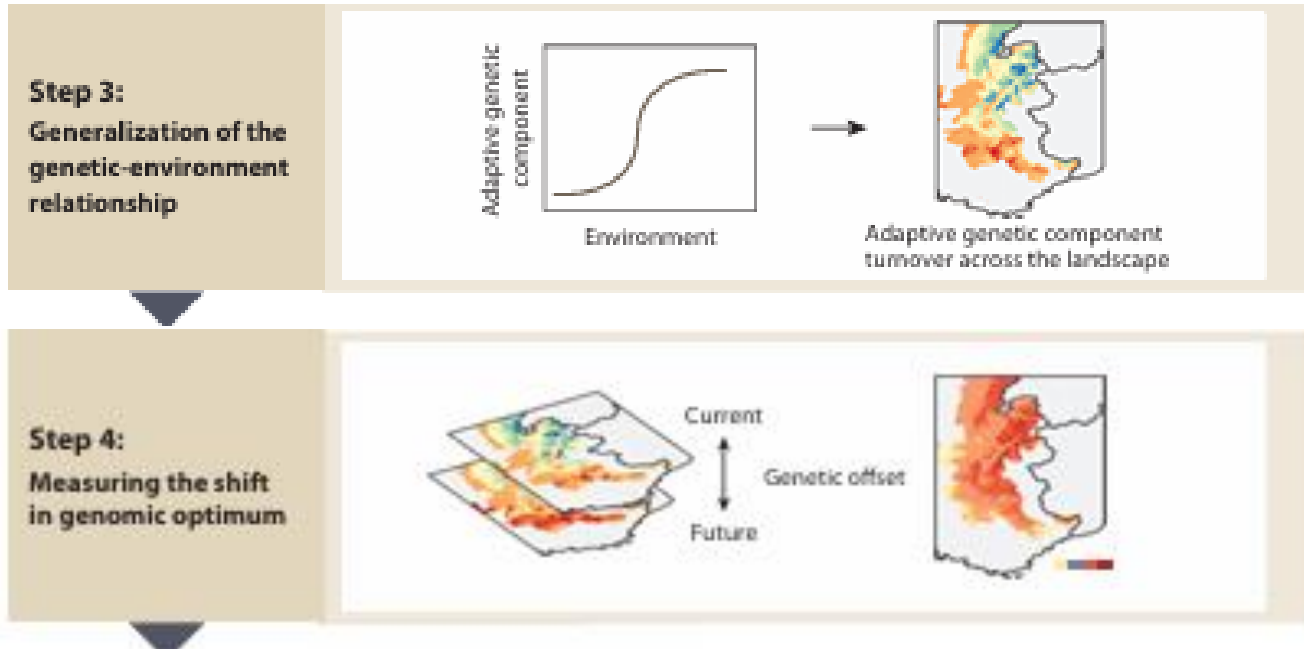# If you're curious – decision trees & random forests

- A supervised machine learning based method of classification/regression

- Gradient Forest is an adaptation of this method

- Check out these videos

  1. https://youtu.be/zs6yHVtxyv8?feature=shared (about decision trees)
  2. https://youtu.be/cIbj0WuK41w?feature=shared (about random forests)

Dasgupta, S. (2018, December 3). *De(coding) random forests*. Medium.
https://towardsdatascience.com/de-coding-random-forests-82d4dcbb91a1

General workflow – prediction phase

Step 3:
Generalization of the genetic-environment relationship

Adaptive genetic component

Environment

Adaptive genetic component turnover across the landscape

Step 4:
Measuring the shift in genomic optimum

Current

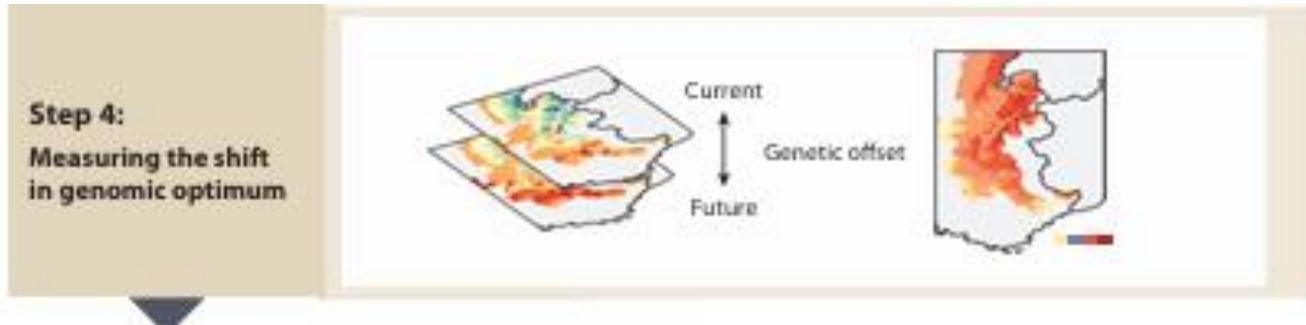Genetic offset

Future

- Extrapolate the inferred relation in space (interpolation)

- And in Time

Adapted from Capblancq et al. (2020). *Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes.*

General workflow – prediction phase

Step 4:
Measuring the shift in genomic optimum

- Predict future climate with existing climatic models

- Use the model to infer the 'genomic distance' between the current optimum and the future

Adapted from Capblancq et al. (2020). *Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes.*
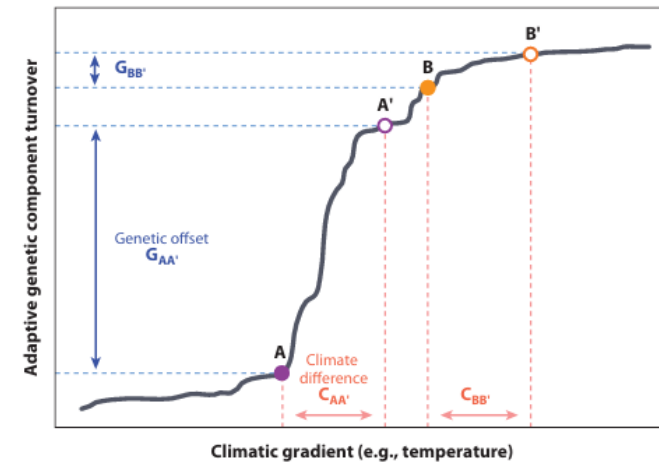
General workflow – prediction phase

- In the x-axis, the climatic gradient is represented (a multidimensional space of environmental variables)

- In y-axis, the adaptive alleles are represented (also a multidimensional space of the frequencies of each adaptive SNP)

- A predicted climate difference ($C_{AA'}$) leads to a predicted genomic offset ($G_{AA'}$)
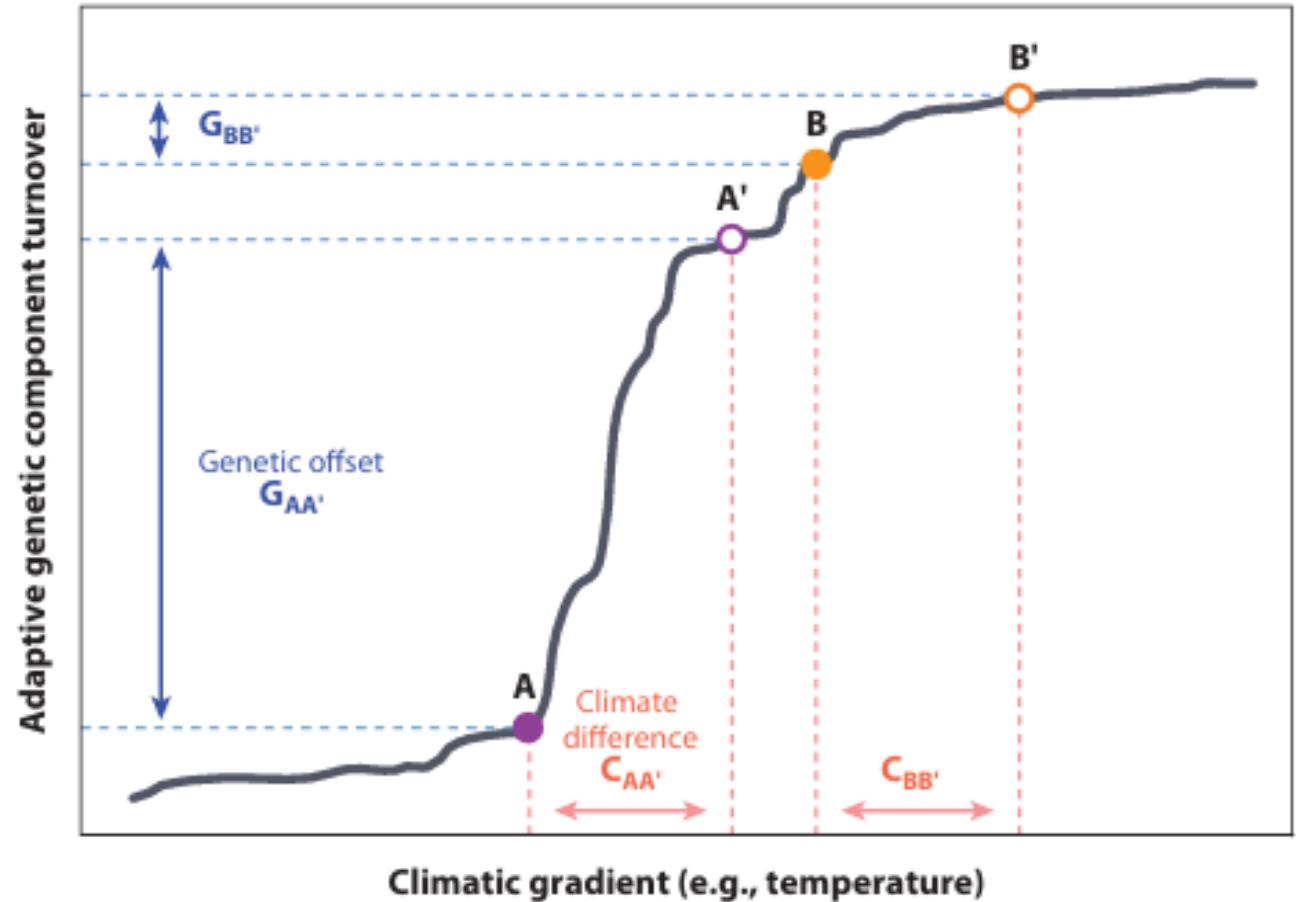
Bay et al. (2018)

ECOLOGICAL GENOMICS

Genomic signals of selection predict climate-driven population declines in a migratory bird

Rachael A. Bay,[1,2]* Ryan J. Harrigan,[1] Vinh Le Underwood,[1] H. Lisle Gibbs,[3] Thomas B. Smith,[1,4] Kristen Ruegg[1,5]

Reference genome assembly

- Illumina sequencing
- Paired-end 250bp reads
- Mate-pair libraries
- Genome annotation was based on protein coding sequences of the zebra-finch and chicken

RAD sequencing

- Illumina sequencing
- SbfI restriction enzyme
- Paired-end 100bp reads
- 229 individuals were sequenced
- Haplotype caller: 4 335 072 SNPs identified
- Quality control: discard low coverage individuals and low quality variants (minor alelle frequency<0.01, depth<8, genotype quality<30)
- Final results: 104 711 SNPs identified for 199 individuals

Modelling genomic variation due to climate

- Using **GradientForest,** a machine learning method
- Only alleles with >10% frequency were considered to avoid false positives
- PCA was done on the results (PC1,PC2,PC3 → RGB values)
- Spatial interpolation beyond sample collection sites
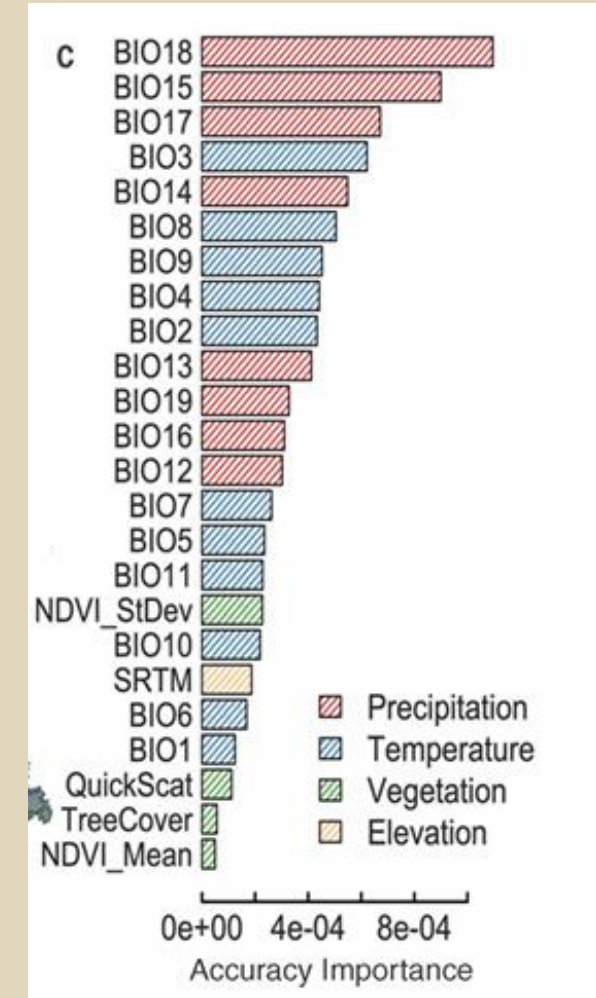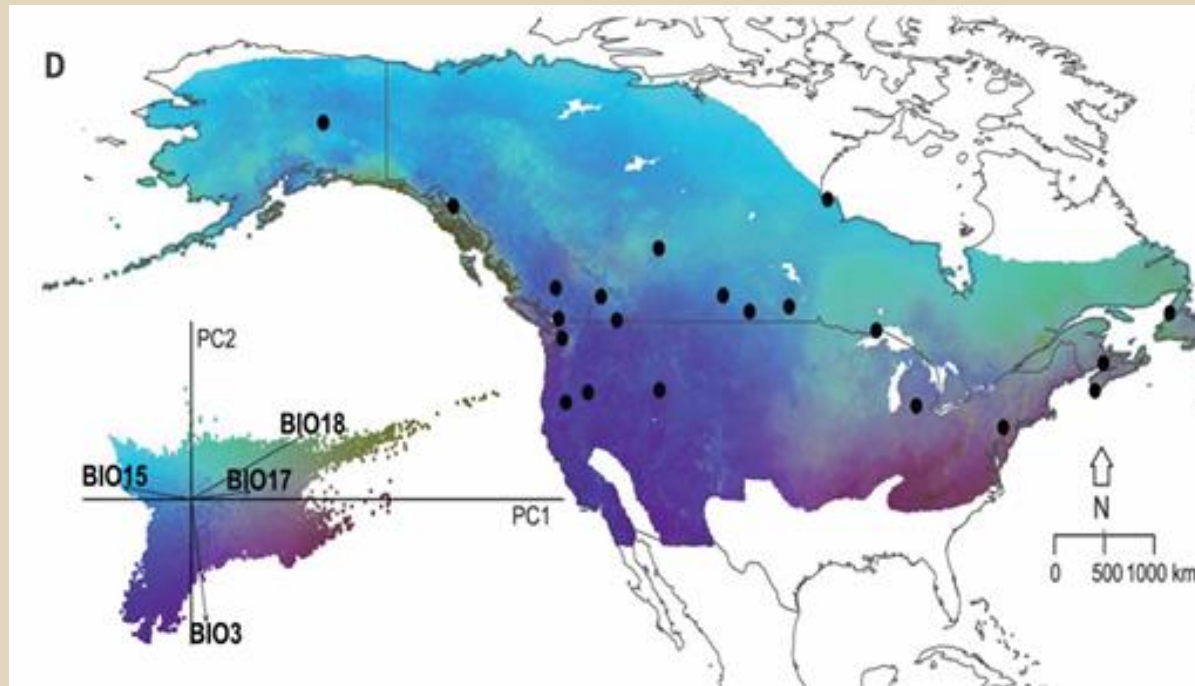
Predicting the future genomic 'vulnerability'

- The authors used climate change predictions from the IPCC under different RCP (representative concentration pathways)

- **GradientForest** was used again, considering the most relevant climatic variables

- The difference between their values in the future and the present is what the authors call genomic vulnerability
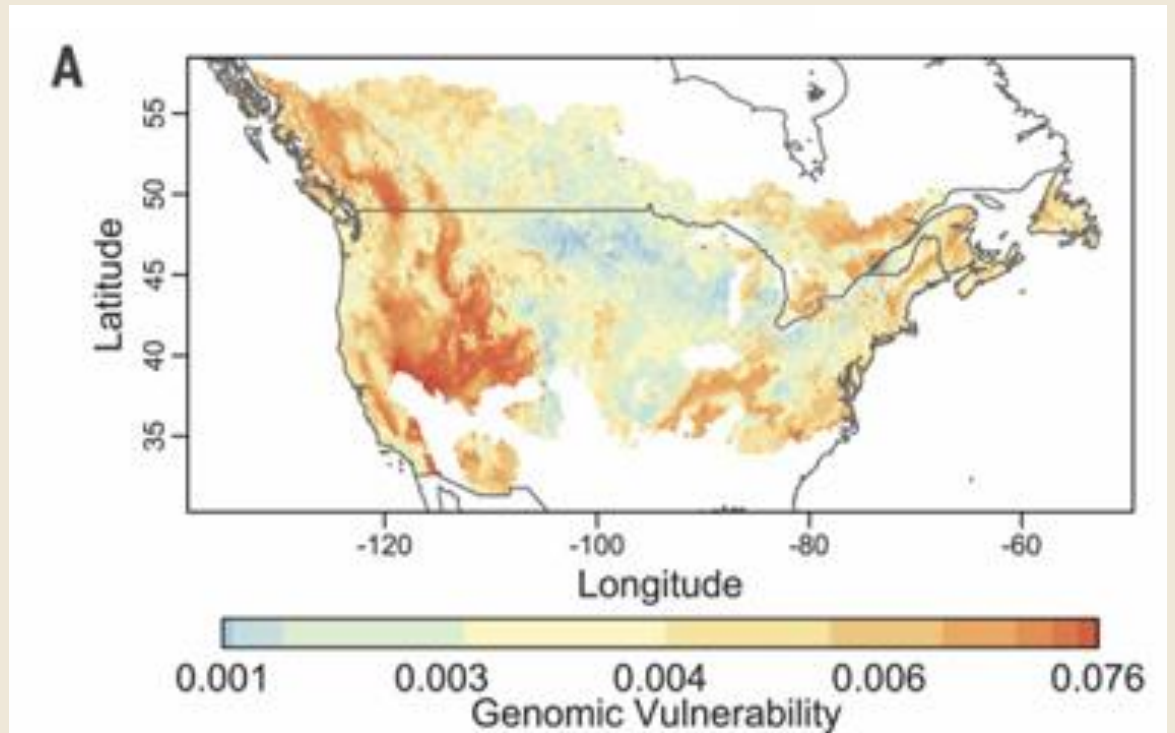
Validating the predictions

- There was a correlation between predicted genomic vulnerability and historical population trends
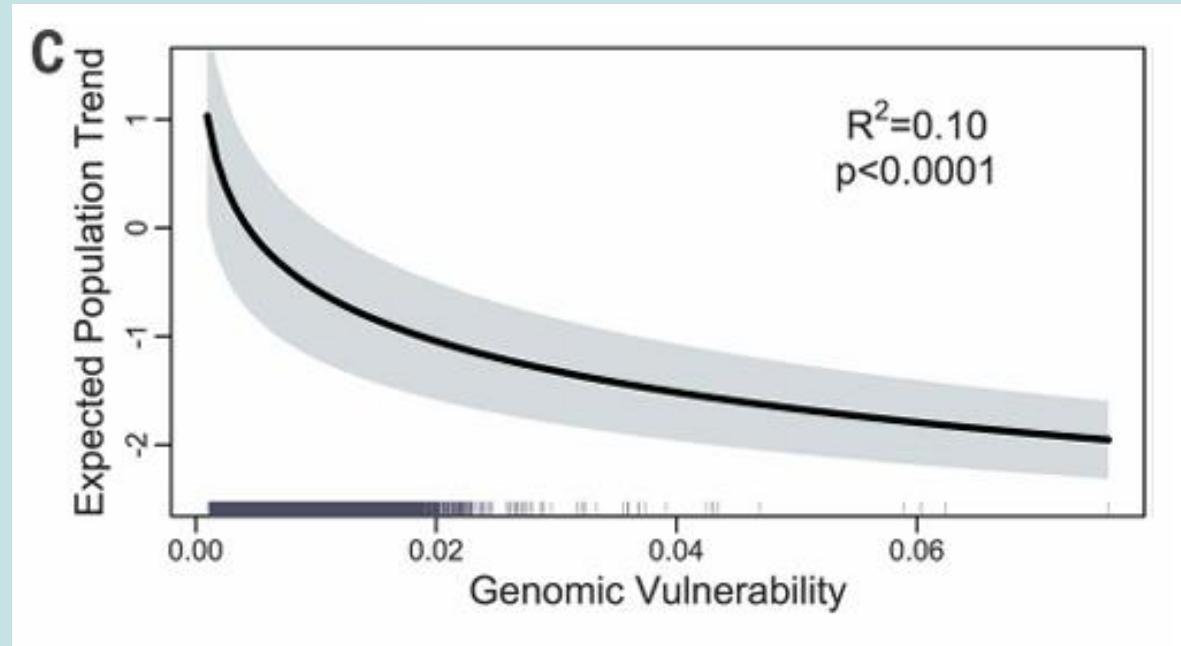
Miller et al. (2023)

ORIGINAL ARTICLE

Genomics-informed conservation units reveal spatial variation in climate vulnerability in a migratory bird

Caitlin V. Miller[1] | Christen M. Bossu[1] | James F. Saracco[2] | David P. L. Toews[3] | Clark S. Rushing[4] | Amélie Roberto-Charron[5] | Junior A. Tremblay[6] | Richard B. Chandler[4] | Matthew G. DeSaix[1] | Cameron J. Fiss[7] | Jeff L. Larkin[7] | Samuel Haché[8] | Silke Nebel[9] | Kristen C. Ruegg[1]
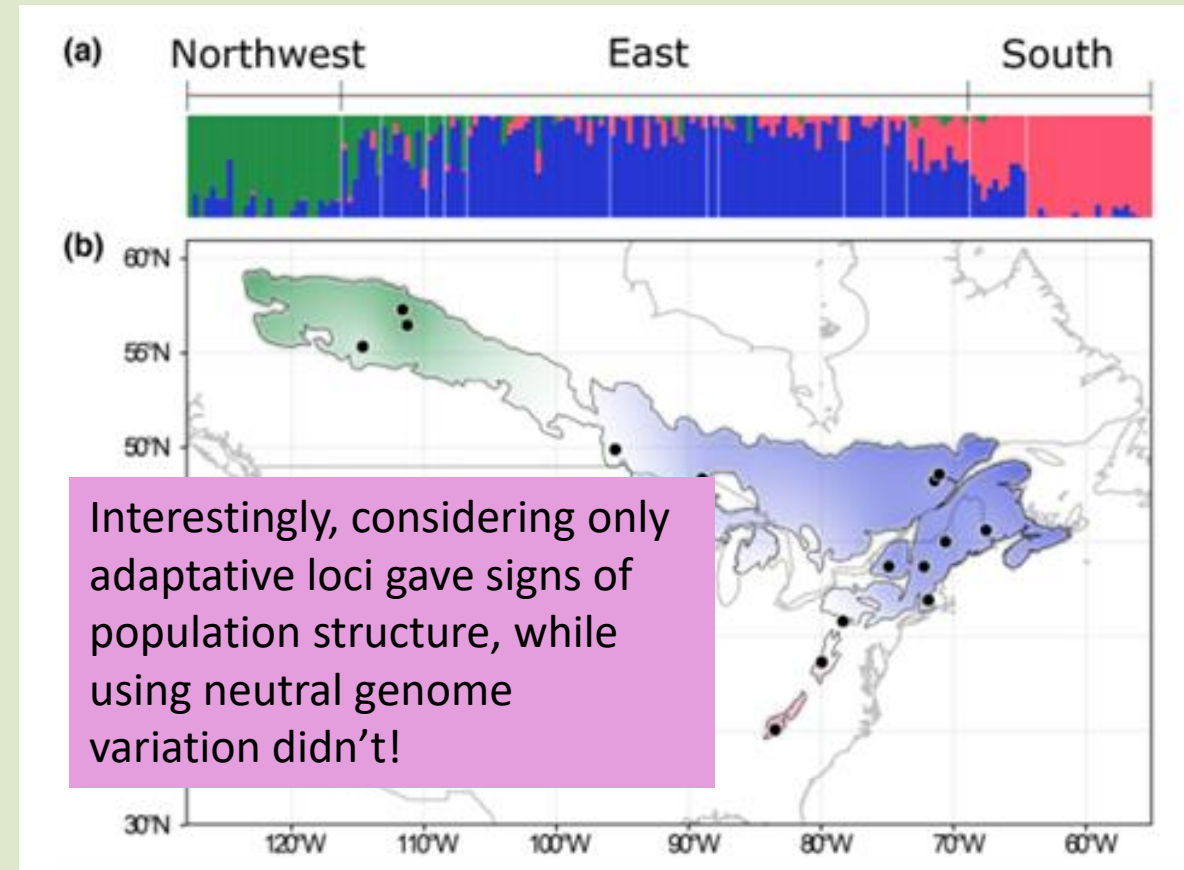
Reference genome assembly

- 10X sequencing (illumina) – a second generation long-read sequencing method
- 150bp paired-reads

lc-WGR

- Low coverage whole genome resequencing (2-3x coverage)
- Illumina sequencing
- Paired-end 150bp reads
- 181 individuals were sequenced
- Quality filtering using HaplotypeCaller and mpileup

Delineating relevant 'adaptive units'

- **GradientForest** was used to understand which climatic variables explained all genomic variation

- Adaptive loci were found using two complementary statistical methodologies: **LFMM** – Latent Factor Mixed Models and **RDA** – Redundancy Analysis

- To detect specific adaptive units, PCA was done on the resulting 11 441 putatively adaptive SNPs. ADMIXTURE was also used – both methods suggested 3 clusters

Interestingly, considering only adaptative loci gave signs of population structure, while using neutral genome variation didn't!

Predicting the future genomic 'vulnerability'

- The authors used climate change predictions under different SSP (shared socioeconomic pathways)
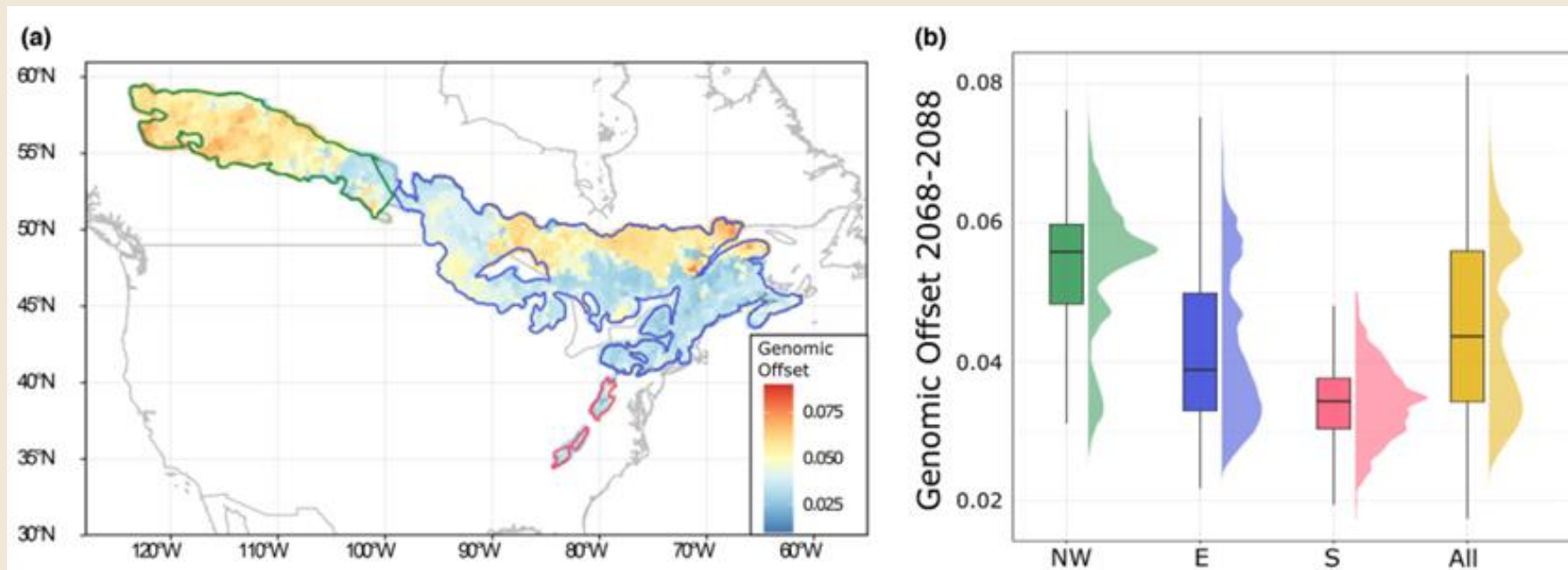

FIGURE 5   Predicted genomic offset across the Canada Warbler breeding range for 2061–2080 using shared socioeconomic pathway 585 (SSP585). Colours are used to represent AUs, green for the Northwestern AU, blue for the Eastern AU and pink for the Southern AU. (a) Map of predicted genomic offset at 100,000 random points across the breeding range. Coloured outlines represent the predicted AUs. (b) Box plots and density curves of genomic offset values for each AU and the entire range.

Genomic Offset – some key assumptions

- Current GEA reflect local adaptation and optima

- Space-for-time extrapolation

- Identified variants are in fact adaptive (population structure) / genomic offset translates into fitness differences
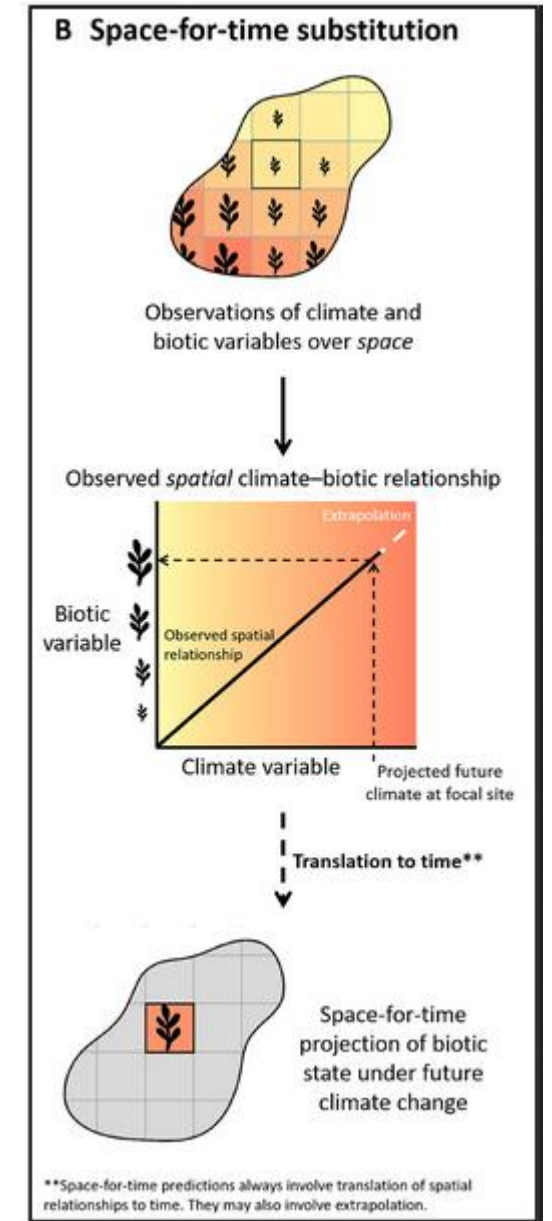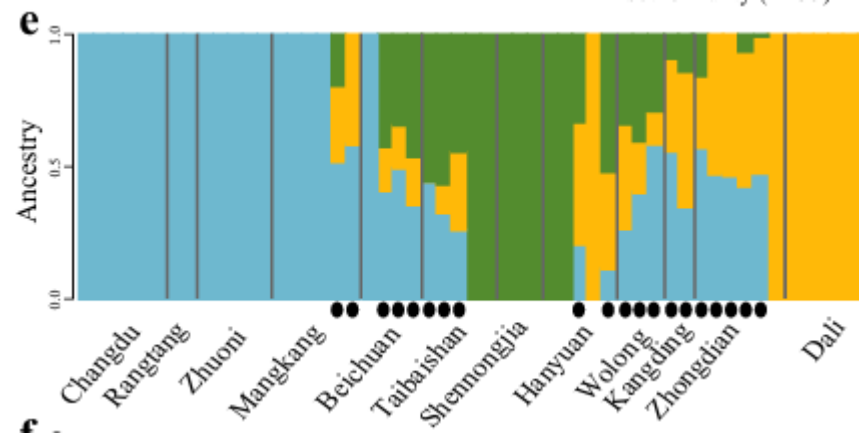
Chen et al. (2022)

The combination of genomic offset and niche modelling provides insights into climate change-driven vulnerability

Yilin Chen [1,2,6], Zhiyong Jiang [1,2,6], Ping Fan [1,2], Per G. P. Ericson [3], Gang Song[1], Xu Luo[4], Fumin Lei [1,2,5] ✉ & Yanhua Qu [1,2] ✉

Chen et al. (2022)

- They used three approaches to understand climate change driven vulnerability:

  ❑ **Genomic Offset**

  ❑ Species Distribution Modelling

  ❑ Landscape Genetics

Ecological genomics
(intraspecific variation in climatic adaptation)

Intraspecific variation in genomics ←→ Integrating intraspecific variation into ENM

Current climatic factors

Genomic offset ←→ Niche suitability change

Future climatic factors

Evolutionary rescue potential (Minor genome-niche interruption)

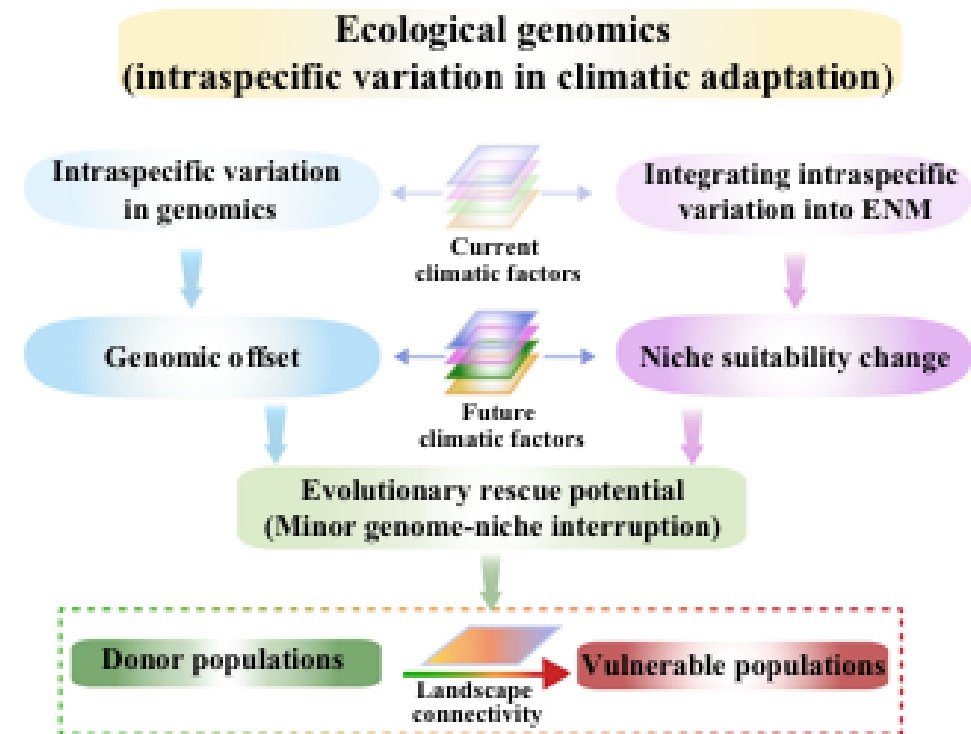Donor populations → Vulnerable populations
Landscape connectivity

Fig. 1 | A framework integrating ecological genomics, niche modelling and landscape genetic analysis to evaluate climate change-driven vulnerability. A schematic representation of climate change-driven vulnerability modelling. We combine ecological genomics and niche modelling to evaluate the population-specific responses of the two species to future climate change. We consider both the genomic offset and niche suitability change for potential evolutionary rescue. Only populations with minor genome-niche interruption can be regarded as a desired store of species' survival for climate change (i.e., being a donor). A landscape genetic approach is then used to evaluate potential rescue costs.

Reference genome assembly

- 10X sequencing (BGI)
- Quality filtering: reads with >60% bases with Phred score <10 were removed, etc...
- Genome annotation was based on protein coding sequences of *Taeniopygia guttata* and chicken

Resequencing

- Illumina sequencing
- 55 individuals were sequenced
- **Coverage: 19.14x**
- Variants were classified with HaplotypeCaller and samtools (intersection)
- 10.3 million SNPs detected!
- Quality control: Only keep biallelic variants, several filtering criteria

Identifying adaptive loci

- The authors used both approaches we've seen so far:

- **GradientForest** was used to associate environmental variables with variation in SNP allele frequency. The 5 most relevant environmental variables were retained and SNPs with $R^2 > 0$ were considered → 5446 SNPs

- **LFMM**, **RDA** and **dbRDA** (distance-based RDA) were used to discover variants associated with variation along environmental gradients. The intersection variants of the three methods were considered. → 72 SNPs
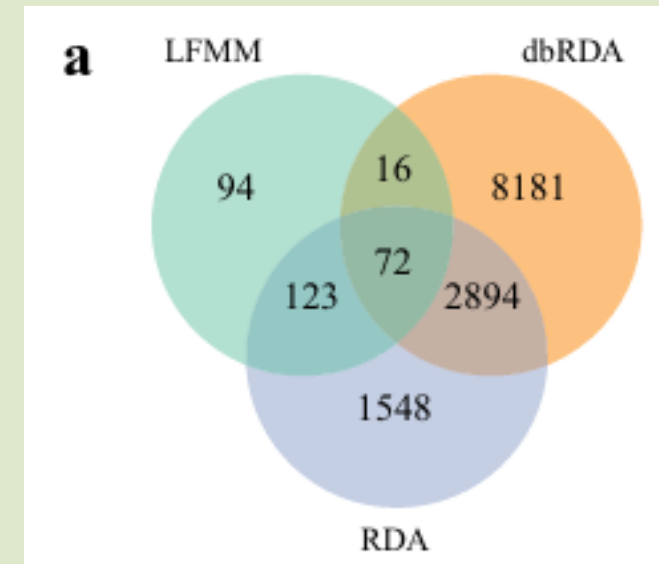
Modelling genomic variation due to climate

- **GradientForest** was used again, but another method was tried: Generalized Dissimilarity Modelling – **GDM**, which is based on generalized linear models
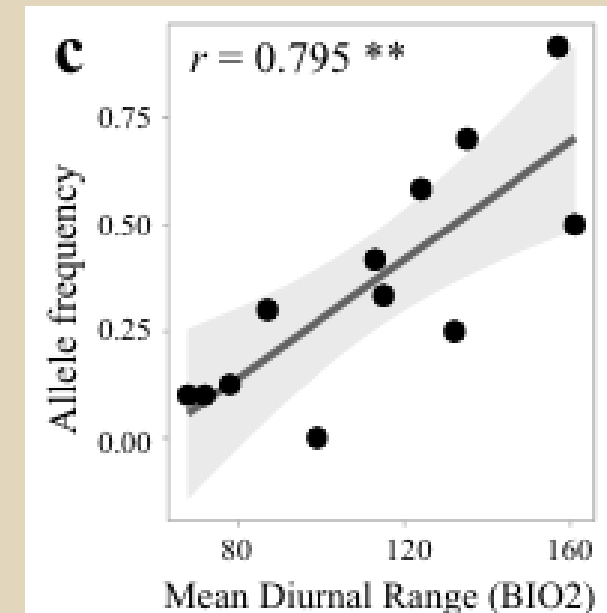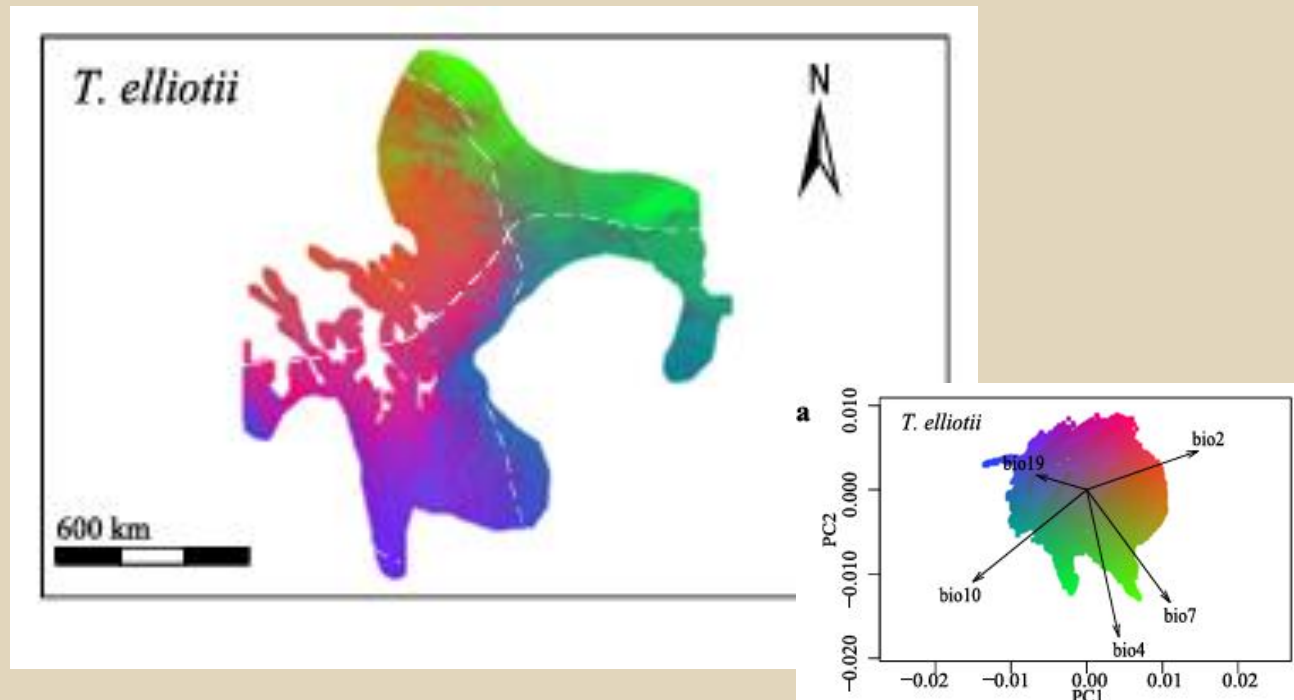
Predicting the future genomic 'vulnerability'

- The authors used climate change predictions from the IPCC under different RCP (representative concentration pathways)

- Both gradientForest and GDM were used
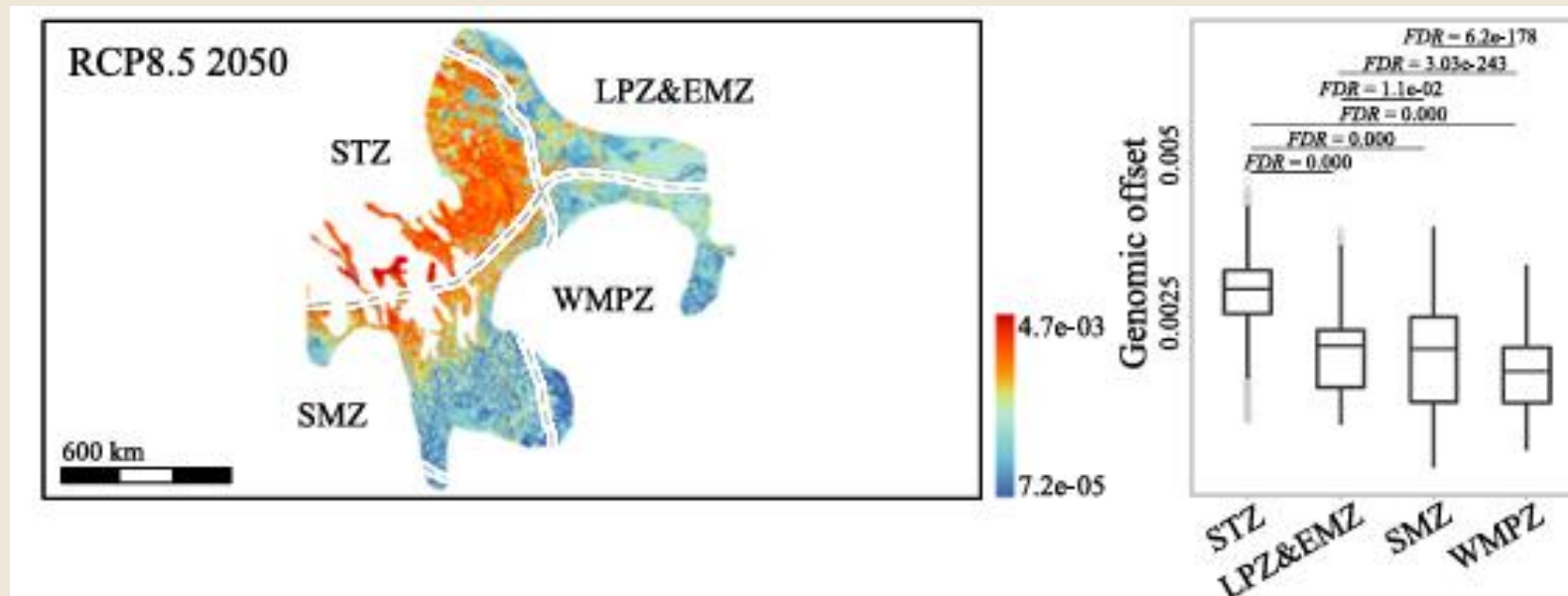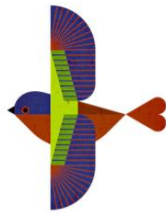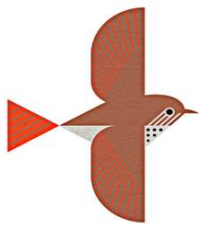
Comparison with SDM

- Warm adapted populations show the least genomic offset, yet they are predicted to experience the greatest decrease in niche suitability. The reverse is true for cold adapted populations

- Both measures were combined in a genome-niche index to discover which populations are climate 'tolerant' and thus allow for evolutionary rescue

- Furthermore, a landscape genetics approach was used to predict possible routes for rescuing-gene flow
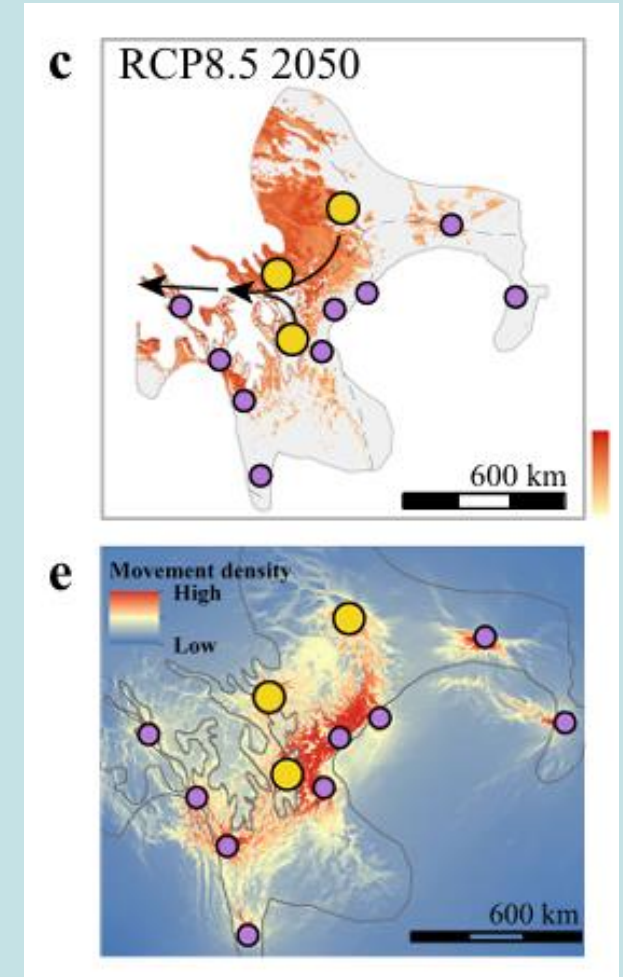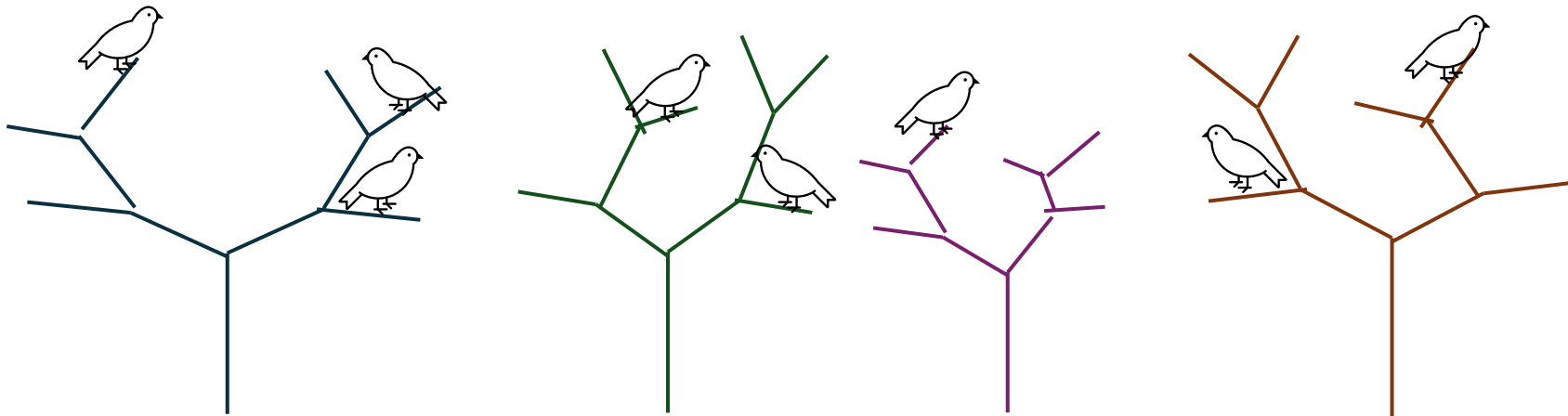
Conclusion

- The genomic offset approach has a lot of potential for studying local adaptation and maladaptation. In doing so, it can inform species conservation and management.

- It's a promising area of research, as statistical methods and algorithms are increasingly improved and studied. However, it still has some frailties.

- In conjunction with other genomic approaches, namely landscape genomics, and traditional methods, it's power for species conservation can only increase.

Genomic offset studies for bird species conservation

Bay *et al.* (2018)
https://doi.org/10.1126/science.aan4380

Miller *et al.* (2023)
https://doi.org/10.1111/mec.17199

Chen *et al.* (2022)
https://doi.org/10.1038/s41467-022-32546-z

References

Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science, 359*(6371), 83–86. https://doi.org/10.1126/science.aan4380

Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020). Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Annual Review of Ecology, Evolution, and Systematics, 51*(1), 245–269. https://doi.org/10.1146/annurev-ecolsys-020720-042553

Chen, Y., Jiang, Z., Fan, P. *et al.* The combination of genomic offset and niche modelling provides insights into climate change-driven vulnerability. *Nat Commun* **13**, 4821 (2022). https://doi.org/10.1038/s41467-022-32546-z

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology, 93*(1), 156–168. https://doi.org/10.1890/11-0252.1

Miller, C. V., Bossu, C. M., Saracco, J. F., Toews, D. P., Rushing, C. S., Roberto-Charron, A., Tremblay, J. A., Chandler, R. B., DeSaix, M. G., Fiss, C. J., Larkin, J. L., Haché, S., Nebel, S., & Ruegg, K. C. (2023). Genomics-informed conservation units reveal spatial variation in climate vulnerability in a migratory bird. *Molecular Ecology, 33*(1). https://doi.org/10.1111/mec.17199

Rellstab, C., Dauphin, B., & Exposito-Alonso, M. (2021). Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications, 14*(5), 1202–1212. https://doi.org/10.1111/eva.13205

Images

- https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSIESQniiIy1bkrcAaRT3pXcT5_dlPBR0uXXA&s

- https://www.birdgenoscape.org/species-list/#completedgenoscapes

- https://onlinelibrary.wiley.com/doi/10.1111/brv.13004

- https://inaturalist.nz/taxa/339863-Trochalopteron-elliotii

- https://en.wikipedia.org/wiki/Canada_warbler