

day_1_afternoon_part_2

William Hemstrom

2023-07-13

Population structure detection with SNP data in R

We'll now work on how to explore and visualize population structure using the same tools (DAPC, PCA, and STRUCTURE) in R using some RADseq SNP data. First, let's discuss the data and background a bit:

Monarch Butterflies in the Pacific



Figure 1: Monarch Butterfly (*Danaus plexippus*), (c) William Hemstrom 2022

Monarch Butterflies (*Danaus plexippus*) are a charismatic, migratory insect species originally from North America. Over the last ~200 years, they have travelled to a series of islands across the Pacific Ocean. They reached Hawaii in the 1840s, Australia in the 1870s, and in roughly 1900, they reached the Mariana Islands, including Guam and Rota islands. They are non migratory almost everywhere in the western Pacific, although they secondarily re-acquired migratory behavior in southern Australia and New Zealand in the 1930s.

In North America, the migratory population can be roughly divided into two “populations” – one in western North America that spends the winter in California and their summer in the western United States and one in eastern North America that winters in central Mexico and summer in the eastern United States.

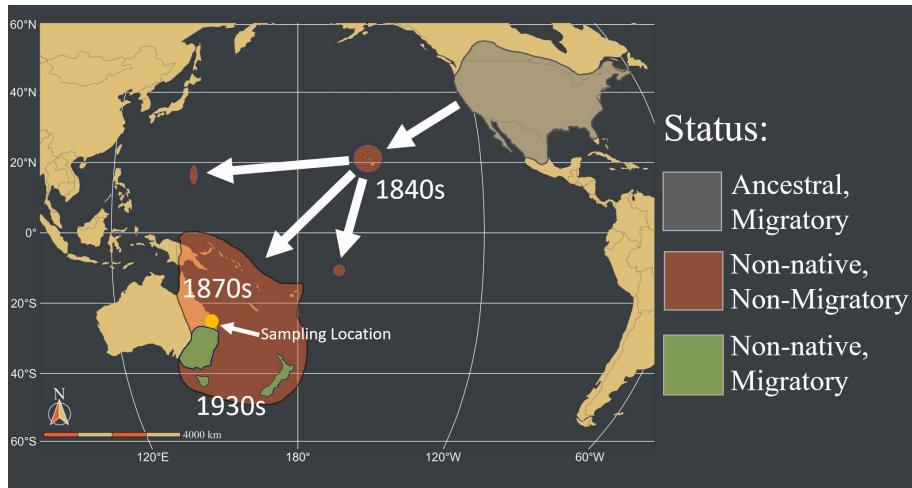
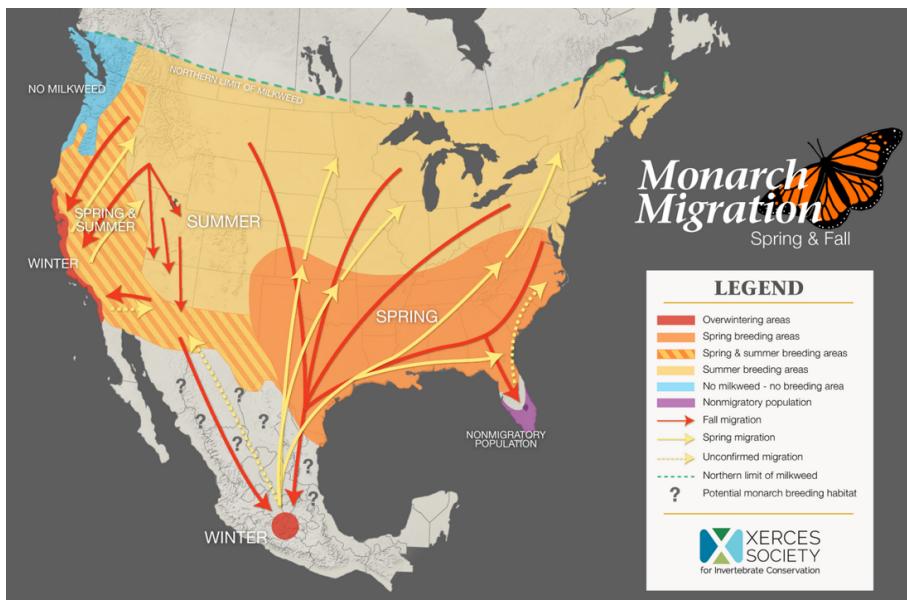


Figure 2: Expansion of monarch butterflies across the Pacific (Hemstrom et al in revision)



Note that they are also native, non-migrants in the Caribbean, Central, and South America, including Colombia where the first photo shown here was taken.

The Data

Located in the **Data** folder, there is a file called **monarchs.vcf.gz**. This is a compressed file in the Variant Call Format (**VCF**)

Questions to answer