# LANDSCAPE GENETICS

Prof. Dr. Akin Pala

Twitter: @akinpala

# Landscape Genetics

- ***Landscape genetics*** is an emerging discipline that combines the fields
  - Population ***genetics***
  - ***Landscape*** ecology

# Landscape Genetics

- The field looks at the genotype x environment interaction, especially in terms of land and its conditions being the environment.
- Subjects such as habitat loss, extinction, population size, genetic diversity are also among the subjects of landscape genetics.

# Cougar vs wolf

- Montana Idaho state line
- Wolves were introduced
- Wolf Pack vs. Cougars.

# Terminology, Jargon

- Landscape genetics
  - Informs on the interactions between landscape features and genotypic processes such as gene flow and selection.
- Landscape genomics
  - A subfield of landscape genetics investigating how environmental factors affect patterns of adaptive genetic variation.

# Terminology, Jargon

- Adaptation
  - The process by which a population or a species globally or locally adapts to its environment.
- Adaptive genetic variation
  - Variation at loci that are under selection.

# Terminology, Jargon

- Covariance structure
  - To account for the repeated data or autocorrelation of data (dependency among errors in time), a matrix of pairwise covariances is incorporated into the model.
- Gene flow
  - In its widest sense, the transfer of alleles or genes from one entity to another: between populations.

# Terminology, Jargon

- Cross-validation approaches
  - Partition of a sample of data into complementary subsets, performing analysis on one subset (training set) and validating the analysis on another subset (validation or testing set).
- Environmental association
  - Landscape-genomic approach to detect loci at which allele frequencies are significantly correlated with environmental factors.

# Terminology, Jargon

- Genome scan
  - Genotyping of many samples at a large number of loci across a genome to see whether any variant is associated with a phenotypic trait or environmental factor.
- Isolation by distance
  - Decrease in genetic similarity among samples with increasing geographical distance.

# Terminology, Jargon

- Genetic clustering
  - Grouping a set of individuals based on their genotype.
- Genome-wide association study (GWAS)
  - A genomic method searching for loci in a genome scan that are significantly associated with phenotypic traits.

# Terminology, Jargon

- Mixed linear models
  - A statistical model (linear regression) containing both fixed (i.e., known) and random (i.e., unknown) effects.
- Migration-drift equilibrium
  - State in which the loss of alleles through genetic drift is compensated by the gain of alleles via gene flow.

# Terminology, Jargon

- Model species
  - A species for which the full genome and significant amounts of genomic information are available.
- Neutral genetic variation
  - Variation at loci that are not under selection. Neutral loci can show the signals of adaptation if they are strongly linked to loci under selection.
- Outlier loci
  - Loci identified as showing signs of selection using population-genomic methods and genome scans.

# Terminology, Jargon

- Overlay
  - A purely visual inspection of spatial coincidences between the location of genetic group or genetic boundaries and any landscape element in geographical information systems (GIS).
- Population genomics
  - Captures signals of selection by searching for loci in a genome scan that show higher or lower differentiation among populations (i.e., outlier loci) than expected under neutral conditions.

# Terminology, Jargon

- Seascape genetics
  - Landscape genetics in marine ecosystems.
- Single nucleotide polymorphism (SNP)
  - Genetic variation at a single position in a genome (usually for only two alleles).

# Why diversity?

- Genetic diversity is required to keep up with the climate change or changing conditions such as nutrition, migrating predators/prey, changing microorganisms including viruses (New selective challenges)
  - Cat or monkey AIDS.
- We need to know how the environment and the landscape itself affects gene flow.

# Why diversity?

- Extinction is usually due to global change
  - You have confounding factors when climate change combines with other factors.
  - You don't know what caused extinction.
- **Global change** refers to planetary-scale **changes** in the Earth system.
  - The system consists of the land, oceans, atmosphere (ozone depletion), polar regions, life, the planet's natural cycles and deep Earth processes.
    - Now includes human society

# Global change and genetic variation

- The key drivers causing loss of the diversity of genes, species, and ecosystems (biodiversity)
  - Overexploitation
  - Invasive alien species
  - Pollution
  - Climate change
  - Degradation, fragmentation, and destruction of habitats.
- The complete extinction of an entire species is the most eye-catching result of this biodiversity loss.

Source: S. Manel, R. Holderegger. 2013. Ten years of landscape genetics. Trends Ecol. Evol., 28, pp. 614–621 http://dx.doi.org/10.1016/j.tree.2013.05.012

# Global change and genetic variation

- Long before an entire species is finally lost, genetic diversity is significantly affected
- Distinct populations go extinct three times faster than entire species.

Source: S. Manel, R. Holderegger. 2013. Ten years of landscape genetics. Trends Ecol. Evol., 28, pp. 614–621 http://dx.doi.org/10.1016/j.tree.2013.05.012

# Global change and genetic variation

- The vanishing of populations is a process directly linked to the loss of intra-species genetic diversity.
- It is therefore of great importance that we understand how the loss of genetic diversity influences adaptive processes.

Source: S. Manel, R. Holderegger. 2013. Ten years of landscape genetics. Trends Ecol. Evol., 28, pp. 614–621 http://dx.doi.org/10.1016/j.tree.2013.05.012

# Global change and genetic variation

- In this respect, the key questions are
  - How has recent global change (land use and climate change) affected patterns of neutral and adaptive genetic variation
  - Are species likely to adapt to ongoing global change on an ecological time scale?
- The main objective of modern landscape genetics is to
  - Understand the effects of global change on genetic patterns to address these questions.

Source: S. Manel, R. Holderegger. 2013. Ten years of landscape genetics. Trends Ecol. Evol., 28, pp. 614–621 http://dx.doi.org/10.1016/j.tree.2013.05.012

# First ten years of landscape genetics

- Researchers try to combine ecology and population genetics.
- It is hard because of the disparities between the two scientific communities.
- The introduction of landscape genetics helped to stimulate research by facilitating the integration of researchers in the fields of population genetics, landscape ecology, and spatial statistics.

Source: S. Manel, R. Holderegger. 2013. Ten years of landscape genetics. Trends Ecol. Evol., 28, pp. 614–621 http://dx.doi.org/10.1016/j.tree.2013.05.012

# First ten years of landscape genetics

- Landscape genetics has increased our understanding of two main processes: gene flow and adaptation.

1. It has helped a move from the study of gene flow in a purely theoretical space characterized by geographical distances only
   - To the study of gene flow in heterogeneous and fragmented landscapes providing estimates of functional connectivity.

Source: S. Manel, R. Holderegger. 2013. Ten years of landscape genetics. Trends Ecol. Evol., 28, pp. 614–621 http://dx.doi.org/10.1016/j.tree.2013.05.012

# First ten years of landscape genetics

2. The new sequencing methods offer an empirical approach to investigate patterns of adaptive genetic variation
   - In wild and domestic species
   - In **real** landscapes
   - As opposed to theoretical modeling and simulations.
   - Making it possible to detect environmental factors driving the process of adaptation.

Source: S. Manel, R. Holderegger. 2013. Ten years of landscape genetics. Trends Ecol. Evol., 28, pp. 614–621 http://dx.doi.org/10.1016/j.tree.2013.05.012

# Genetic patterns depend on temporal and spatial scale

- Landscape genetics also highlights temporal-scale effects in addition to spatial effects.
- One prominent feature is lag times.
- Implementing a barrier rapidly results in a detectable genetic response
- The removal of a barrier will become measurable only with a time lag of many generations.

# Genetic patterns depend on temporal and spatial scale

- Until now, most landscape-genetic studies used genetic distances among populations to estimate gene flow.
  - However, genetic distances (mainly) reflect historical patterns of gene flow and might not be indicative of gene-flow patterns in contemporary landscapes.
- Empirical studies have found that genetic distances reflect contemporary landscapes better than historical landscapes (not gene flow).

# Genetic patterns depend on temporal and spatial scale

- The effects of spatial scale on genetic patterns have been tackled by restricting analyses to those populations within certain distance classes;
  - For example, maximal migration distance or neighboring populations.
- By doing so, researchers have found that unique effects of landscape are often detected only up to certain distances (changed after a certain distance).
  - For example, a damselfly stuck to streams during short-distance dispersal but used open agricultural land during long-distance dispersal.

# Genetic patterns depend on temporal and spatial scale

- A topic related to scale effects is population arrangement, which also influences landscape genetic patterns and has not yet been investigated in depth.
- Graph theory can help in understanding how spatially explicit patterns of gene flow are influenced.

# Landscape genomics, a powerful tool for studying adaptive genetic variation

- Landscape genomics is a branch of landscape genetics.
- It aims to identify environmental or landscape factors that influence adaptive genetic diversity in wild species using genome scans with a large number of molecular markers genotyped per individual (SNP chips).
- The development of landscape genomics was greatly aided by the emergence of large genomic and environmental datasets in population genomics.

# Landscape genomics, a powerful tool for studying adaptive genetic variation

- So far, most landscape-genomic studies have used population genomic tools such as outlier locus detection.
- In these studies, the allelic distribution at identified outlier loci is a posteriori correlated with the distribution of environmental factors.

# Landscape genomics, a powerful tool for studying adaptive genetic variation

- Specific landscape genomic approaches have also been developed that directly correlate allele frequencies with environment factors.
- One important point in landscape-genomic studies is the need to account for genetic structure and/or demographic effects.
  - This can be done using linear mixed models.
  - G X E interaction

# Genotype Environment Interaction

- Data GXE;
- Title 'GXE Interaction';
- Input Rep Gene Y1 Y2 Y3 @@;
- Env=1; Y=Y1; Output;
- Env=2; Y=Y2; Output;
- Env=3; Y=Y3; Output;
- Drop Y1 Y2 Y3;
- Datalines;
- 1 1 64 17 8 1 2 76 20 59 1 3 38 51 66 1 4 90 110 30
- 4 1 47 52 55 4 2 117 64 68 4 3 81 45 58 4 4 58 73 33
- Proc Sort Data=GXE; By Gene Env Rep;
- Proc MIXED Data=GXE;
- Class Env Rep Gene;
- Model Y=Env Rep(Env) Gene Gene*Env; Test H=Env E=Rep(Env);
- lsmeans Gene*Env/pdiff;
- Run; Quit;

Source: http://analytics.ncsu.edu/sesug/2003/SD15-Bondari.pdf

# Landscape genomics, a powerful tool for studying adaptive genetic variation

- Genome scans with many genetic markers provide the opportunity
  - To investigate local adaptation in natural populations
  - Identify candidate genes under selection.
- Large single nucleotide polymorphism (SNP) datasets based on next-generation sequencing techniques has encouraged the development of new techniques for genome-wide association studies (GWAS).

# Landscape genomics, a powerful tool for studying adaptive genetic variation

- SNPs are dense throughout the genome of most organisms and are commonly observed in functional genes making them ideal markers to study adaptive molecular variation.
- GWAS search for associations of SNPs with particular phenotypic traits that are stronger than would be expected under neutral conditions.

# Landscape genomics, a powerful tool for studying adaptive genetic variation

- Stronger SNPs indicate genomic areas that are potentially under selection for a given trait.
- The GWAS approach is close to landscape genomic approaches that evaluate the association of SNPs with environmental factors instead of phenotypic traits.

# Selection signature

- In livestock, SNP studies showing the common patterns of genetic variation (HapMap) can provide insight into the differentiation of breeds and long-term selection for complex traits.
- When a favorable mutation occurs within a population under directional selection, the frequency of the favorable allele is likely to increase over time.
  - Directional selection, bi-directional selection, control

Source: Stella A., Ajmone-Marsan P., Lazzari B., Boettcher P., 2010. Identification of selection signatures in cattle breeds selected for dairy production. Genetics 185: 1451–1461.

# Selection signature

- DNA is composed of linear molecules
  - The probability of recombination is inversely proportional to the distance separating them
  - Nucleotides adjacent to the favorable mutation also tend to increase in frequency, in a sort of "hitch-hiking" process.
- This process leads to "signatures of selection"
  - Characterized by distributions of nucleotides around favorable mutations that differ statistically from that expected purely by chance.
- Detection of selection signatures can increase the understanding of the biology underlying a given phenotype and may provide tools to increase efficiency of selection.

Source: Stella A., Ajmone-Marsan P., Lazzari B., Boettcher P., 2010. Identification of selection signatures in cattle breeds selected for dairy production. Genetics 185: 1451–1461.

# Selection signature

- Various methods have been developed for detection of selection signature.
- In general, most of these methods are based on comparison of the distribution of allelic frequencies
  - by calculating population genetics statistics that are a function of allelic or genotypic frequencies
    - Assuming Hardy Weinberg equilibrium.
    - Partly covered in the genetics class, we even had chi square
  - $F_{ST}$ (change in allele diversity) and linkage disequilibrium
- In addition, specific tests for detecting signatures have been developed.

Source: Stella A., Ajmone-Marsan P., Lazzari B., Boettcher P., 2010. Identification of selection signatures in cattle breeds selected for dairy production. Genetics 185: 1451–1461.

# Landscape genomics, a powerful tool for studying adaptive genetic variation

- Landscape genomics links patterns of adaptive genetic variation to environmental heterogeneity.
- Questions important for the management of biodiversity under global change:
  - What is the relationship between global and local patterns of adaptation?
  - How important are spatial scale and habitat heterogeneity in maintaining adaptive genetic variation?

# Prediction and conservation management

- Landscape genetics can inform conservation management by providing knowledge on:
  - Dispersal ecology
  - Fragmentation
  - Functional connectivity
  - The effectiveness of connectivity measures.
- Similarly, landscape genomics can determine which populations should be given priority when conserving adaptive genetic variation.

# Prediction, Multispecies Approach

- Landscape genetics is evolving from a descriptive science to a predictive science to have more impact on conservation management.
- This requires
  - An understanding of the behavior of multiple species in the same or multiple landscapes
  - Extrapolation from one species to other species and from one landscape to other landscapes.

# Multispecies Approach

- Multispecies landscape-genetic or -genomic studies are a promising approach to achieve successful conservation strategies
  - Because it is almost impossible to deduce general landscape effects on gene flow or local adaptation from single-species studies.
  - For example, Amos *et al.* used genetic data to rank various alternative models, from high-resistance models for sedentary species (i.e., species that barely move through the landscape) to a 'no isolation by distance' model for the most mobile species (i.e., the landscape has no effect on movement), in ten woodland birds.

# Multispecies Approach

- Landscape-genetics or -genomics studies could also compare species from diverse taxonomic groups or different habitat types in the same landscape.
  - Manel *et al.* showed that, for 13 wide-ranging alpine plants, the same environmental variables drive adaptive genetic variation at the whole biome scale of the European Alp mountains.
  - This example shows that carefully selecting a representative set of ecologically relevant species (e.g., common or dominant species) makes it possible to identify general patterns that go far beyond single-species studies.

# Multispecies Approach

- High-throughput next-generation sequencing and the reduction in processing costs per sample let us derive large genomic datasets for a large number of organisms.
- Comparing the genetic patterns of many species will make it possible to test hypotheses on what cause variation among species in the same or different ecosystems.

# Multispecies Approach

- Investigating several species also makes it possible to determine whether there is selection at the community level
  - Questions of importance not only for science but also for conservation biology.

# Resampling methods

- Regarding prediction, cross-validation of landscape genetic models is an important step in the elaboration of successful conservation strategies.
- Once you have a landscape genetic model for a species and for a particular landscape
  - This model's predictive power can be assessed using various cross-validation approaches
  - Leave-one-out: Pick one out, calculate heritability, put it back, pick another one out, calculate heritability etc.
  - K-fold: divide data into K roughly equal-sized parts

# Resampling methods

- Cross-validation, jackknife and the bootstrap are all resampling methods.
- The bootstrap is a flexible and powerful statistical tool that can be used to quantify the uncertainty associated with a given estimator or statistical learning method.
  - For example, it can provide an estimate of the standard error of a coefficient
- It is powerful because it can be used with various data types including discrete data.

# Prediction and conservation management

- If the model's predictive power is sufficient, landscapes could now be changed *in silico* (simulation) according to a specific change scenario.
  - For instance, does the construction of a residential area dramatically influence functional connectivity?
- Do ecological compensation areas significantly increase gene flow?
  - Few corresponding studies have been conducted.

# Prediction and conservation management

- Spatial graph theory and simulations can be helpful in addressing questions relevant to predictive conservation.
- They resolve gene-flow patterns at multiple spatial scales and can identify populations of particular relevance to the integrity of habitat networks, allowing the testing of various scenarios of landscape change.
  - So, simulations are useful

# Effects of Global Change on Genetic Variation

- The consequences of global change on genetic diversity have so far been understudied.
  - Essential for species persistence in the longer term
- Landscape genetics plays an important role in filling this gap.

# Effects of Global Change on Genetic Variation

- Understanding of the effects of global change on genetic diversity requires:
  - Knowledge of the molecular basis of
    - Adaptively relevant traits
    - Phenotypic variation (via adaptive diversity)
    - The distribution of genetic variation connected to population and species shifts (via neutral diversity).

# Prediction and conservation management

- The first step is
  - To conduct surveys of the current distribution of environmental factors within landscapes,
  - To determine how they influence the distribution of adaptive genetic variants.
- Annual temperature regimes influence the distribution of adaptive genetic variation in model and non-model organisms.

# Prediction and conservation management

- Modeling *Arabidopsis thaliana* genotypes sampled across the species' ecological niche based on the analysis of candidate SNPs for flowering time.
  - Later-flowering genotypes have a more restricted range potential and narrower niche breadths than earlier-flowering genotypes.
- This indicates that some geographical regions contain habitats suitable for some genotypes but not for others.
  - GXE interaction at its best

# Prediction and conservation management

- This study neatly illustrates how it is essential to consider the genetic diversity level when studying the impact of global change on species responses.
- Artificial selection requires study of genetic variance, heritability first to estimate accuracy and success.
- Lower heritability require harder selection

# Prediction and conservation management

- Studying contemporary gene flow using neutral genetic variation makes it possible to investigate the potential for adaptive genes to spread across landscapes.
- If this potential is high, gene flow will enable populations and species to adapt to changing environmental conditions not only due to standing genetic variation, but also owing to gene migration.

# Prediction and conservation management

- Shimizu *et al.* developed a mechanistic model of how genetic variation at the self-incompatibility locus in Arabidopsis halleri reacts and changes under specific temperature scenarios.
- Changes in allele frequencies at the self-incompatibility locus will have a substantial influence on gene flow patterns.
  - Similar hypotheses could be developed for many species, once loci that are relevant under global change have been identified.

# Prediction and conservation management

- We can combine spatial patterns of adaptive genetic variation with information on the pace of gene flow.
- Then we can forecast the future spatial distribution of genetic variation in a range of species under various global-change scenarios.
- Such applications are now feasible thanks to the ever-increasing sophistication of landscape-genetic and -genomic tools.

# Conclusion

- After many years of landscape-genetic research, it is fundamental to go beyond the mere description of patterns.
- Progress has been made toward a more theoretical interpretation of the processes
  - gene flow vs. selection vs. drift vs. recombination behind observed patterns of genetic variation.

# Conclusion

- Landscape genetics now offers promising tools to better understand genetic patterns–process relationships.
  - For example, landscape genomics already contributes to our understanding of the nature of genes involved in local adaptation to spatial environment differences.
- Landscape genomics identifies the loci or genomic regions under selection
  - Should be moving to dissecting and characterizing the underlying genes, genome architecture, molecular mechanisms, and ecological functions as well as their interactions.

# Conclusion

- Pattern–process analysis needs to go beyond the separation of neutral and adaptive genetic variation.
- It would be useful to use estimates of contemporary gene flow.
- Gene flow affects genetic variation, allele frequencies, inbreeding levels etc.

# Conclusion

- It would lead to the detailed analysis of
- How gene flow spreads alleles of adaptive relevance across landscapes,
- How strong gene flow counteracts local adaptation,
- How strong selection advantages of immigrating alleles have to be to become established in resident gene pools.
  - For instance, Andrew *et al.* studied adaptation under a regime of substantial gene flow in a sand dune species of sunflower.
  - They reported that isolation by adaptation increases divergence when gene flow decreases
    - Becaue natural selection against immigrants reduces gene flow between different dune habitats.

# Conclusion

- Similar approaches could be used to monitor gene escape from genetically modified organisms (GMOs) in situations where protective measures, such as isolating crop strips, have been implemented to reduce gene flow from GMO fields.
  - They could also be used to optimize reserve design by integrating the adaptive potential or adaptability of species.

# Conclusion

- Almost all studies make inferences from genetic markers such as microsatellites, SNPs, AFLPs etc.
- Using a whole genome approach and next generation sequencing techniques would benefit this field immensely
  - 20 birds from a mountain range, and 20 from a valley. Sequence their whole genomes and analyze the results
- This is the largest obstacle now, analyzing such large data and making sense out of it.

# Conclusion

- Landscape genetics increases awareness of the important effects that spatial patterns of landscape elements and environmental factors exert on neutral and adaptive genetic variation.
- In other words, landscape genetics has brought population genetics research closer to the empirical world as seen and experienced by organisms.
  - This endeavor should continue.