Many ways to study the genomic basis of adaptation and adaptation with genomic tools…

Claire Mérot & Anna Tigano
Physalia Courses
September 2020

Analytical approaches

GWAS

Comparative genomics

Transcriptomics

Experimental evolution

QTL mapping

Epigenetics

Population genomics

Plenty of approachs

- Local adaptation / population genomics
  - Other statistics
- Comparative genomics
- Trait-focused genetics/genomics
  - GWAS
  - QTL
- Multi-omics
  - Transcriptomics
  - Epigenomics
  - Proteomics
- Functions and experiments
  - Experimental evolution / experimental selection
  - Candidate genes and pathways
  - Common garden
  - Gene-editing

Local adaptation / population genomics

Site frequency spectrum:

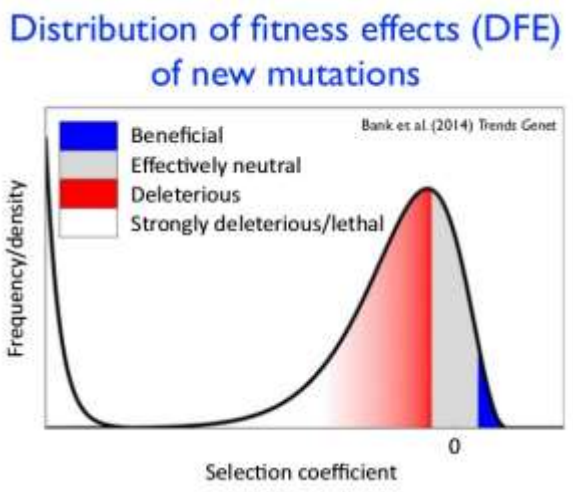
- Tajima's D: tests for excess of low or intermediate frequency alleles
- Distribution of fitness effects (DFE)

Signatures of selective sweeps:

- SweepFinder
- Extended haplotype homozygosity (iHS)

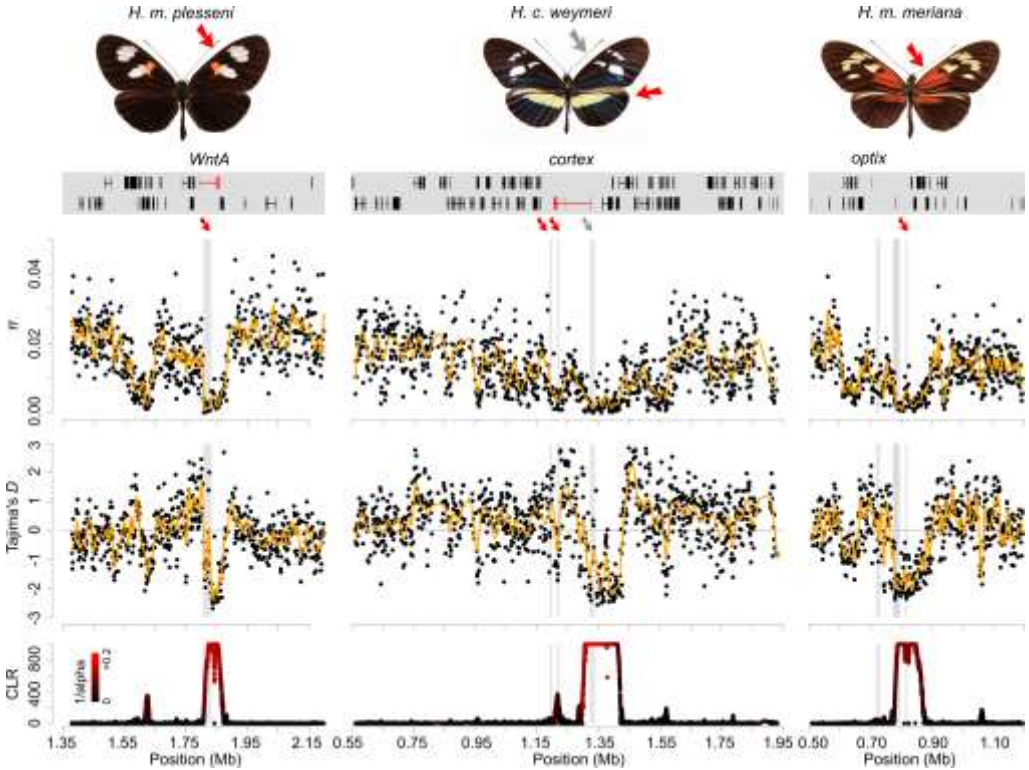
Inference of Distribution of Fitness Effects and Proportion of Adaptive Substitutions from Polymorphism Data
Tataru et al
GENETICS *November 1, 2017 vol. 207 no. 3 11031119;*
https://doi.org/10.1534/genetics.117.300323

polyDFEv2.0



Local adaptation / population genomics

Sweep finder, Tajima’s D



Targeted sequence capture towards colour loci

-> sweepFinder

-> low pi (diversity),

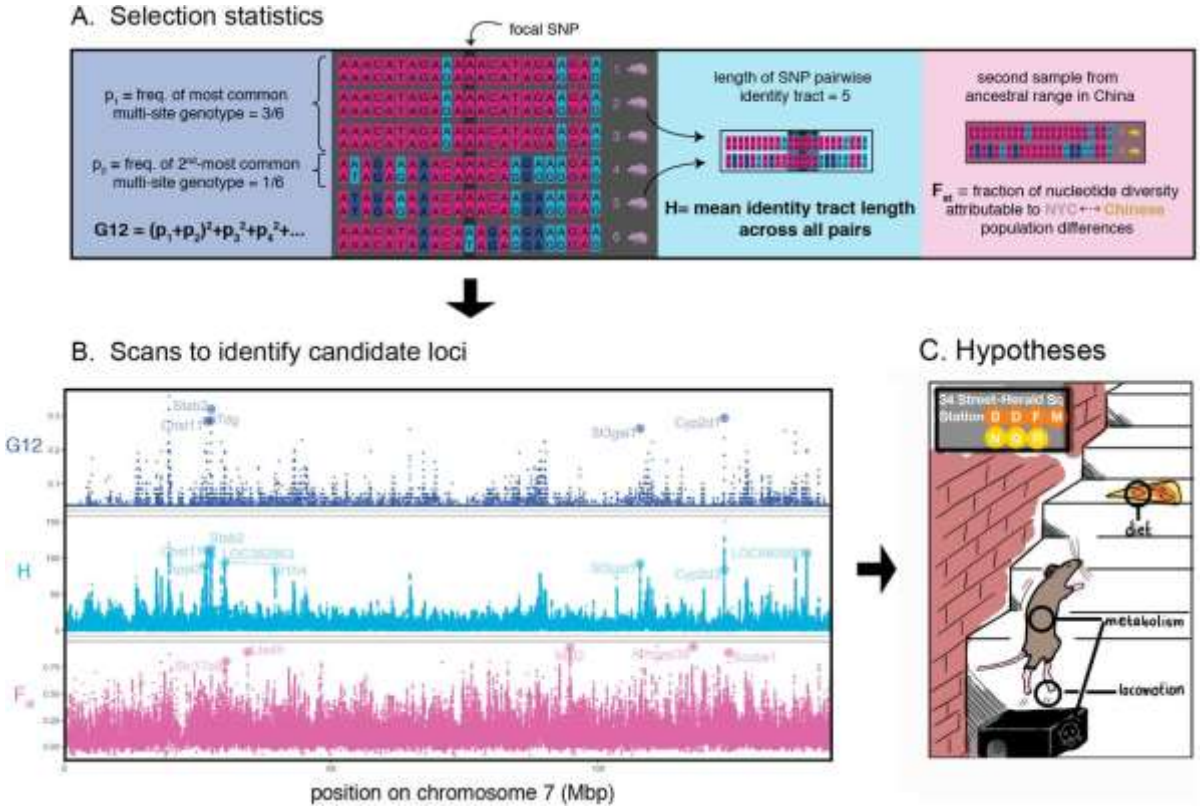
-> Low Tajima’s D

Moest M, Van Belleghem SM, James JE, Salazar C, Martin SH, Barker SL, et al. (2020) Selective sweeps on novel and introgressed variation shape mimicry loci in a butterfly adaptive radiation. PLoS Biol 18(2): e3000597. https://doi.org/10.1371/journal.pbio.3000597

Local adaptation / population genomics

Haplotypes

A signal of adaptation in New York City rats?

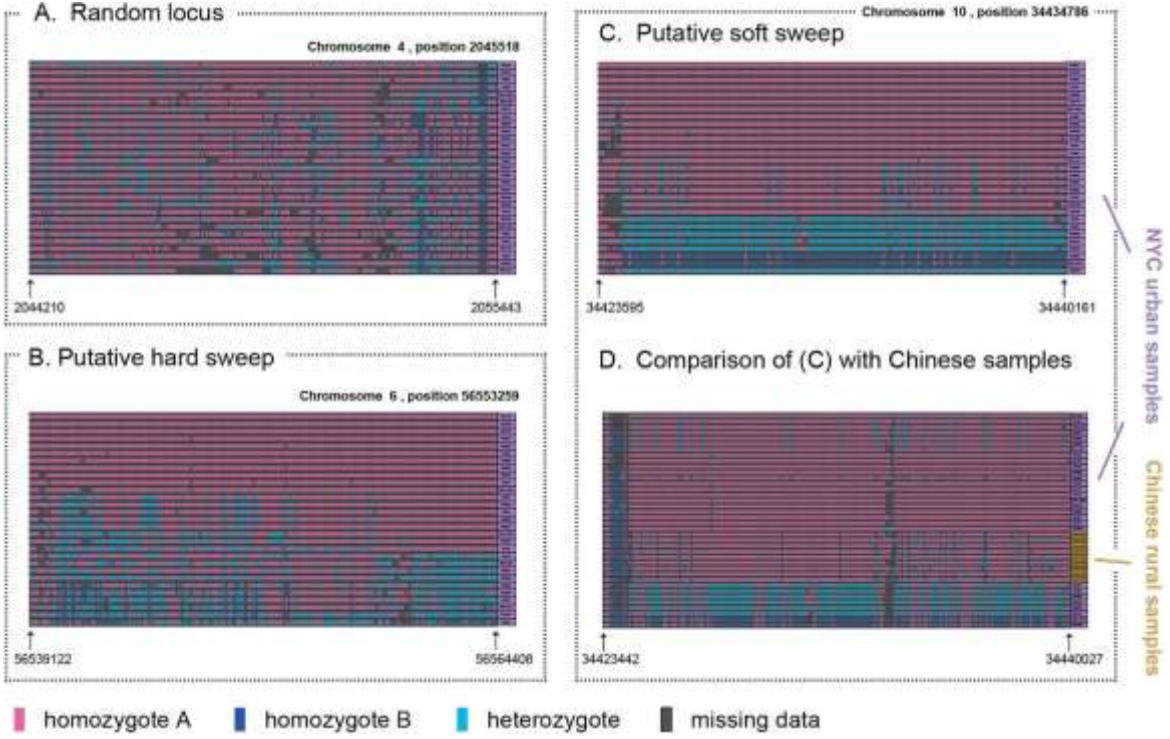


Harpak, A., Garud, N., Rosenberg, N. A., Petrov, D. A., Combs, M., Pennings, P. S., & Munshi-South, J. (2020). Genetic Adaptation in New York City Rats. *BioRxiv*. https://doi.org/10.1101/2020.02.07.938969

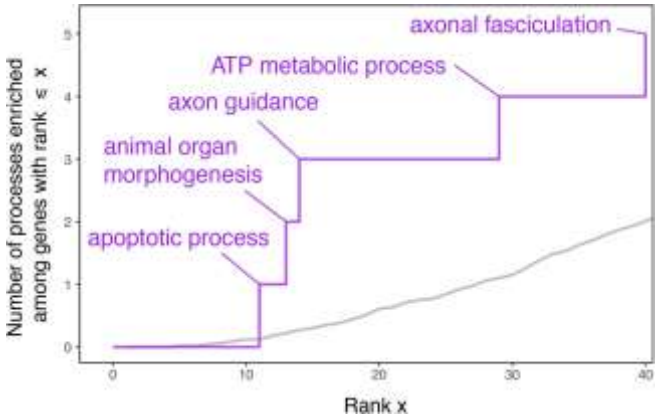
Methods to find hard sweep and soft sweeps based on haplotype + asymptotic MK test

Local adaptation / population genomics

Sweep signal in haplotypes



Biological processes in outlier loci



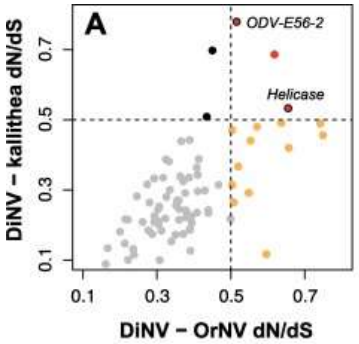
Harpak, A., Garud, N., Rosenberg, N. A., Petrov, D. A., Combs, M., Pennings, P. S., & Munshi-South, J. (2020). Genetic Adaptation in New York City Rats. *BioRxiv*. https://doi.org/10.1101/2020.02.07.938969

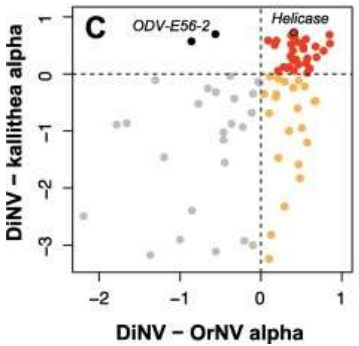
Population genomics/ comparative genomics

Comparison to outgroup:

- ω = dN/dS ratio of the number of non-synonymous to synonymous sites between target species and an outgroup
- McDonald Kreitman test: variation in rate of evolution across the genome and fixed vs. polymorphic sites
- Fay and Wu's H: Test of neutrality by comparing frequency of derived vs. ancestral alleles

asymptoticMK: A Web-Based Tool for the Asymptotic McDonald–Kreitman Test
Benjamin C. Haller and Philipp W. Messer
G3: GENES, GENOMES, GENETICS *May 1, 2017 vol. 7 no. 5 1569-1575; https://doi.org/10.1534/g3.117.039693*





A virus(DiNV) compared to two outgroups (OrNV and DiNV)

⇒ Genes with high dN/dS ratio

⇒ Genes with high alpha value following McD-K test

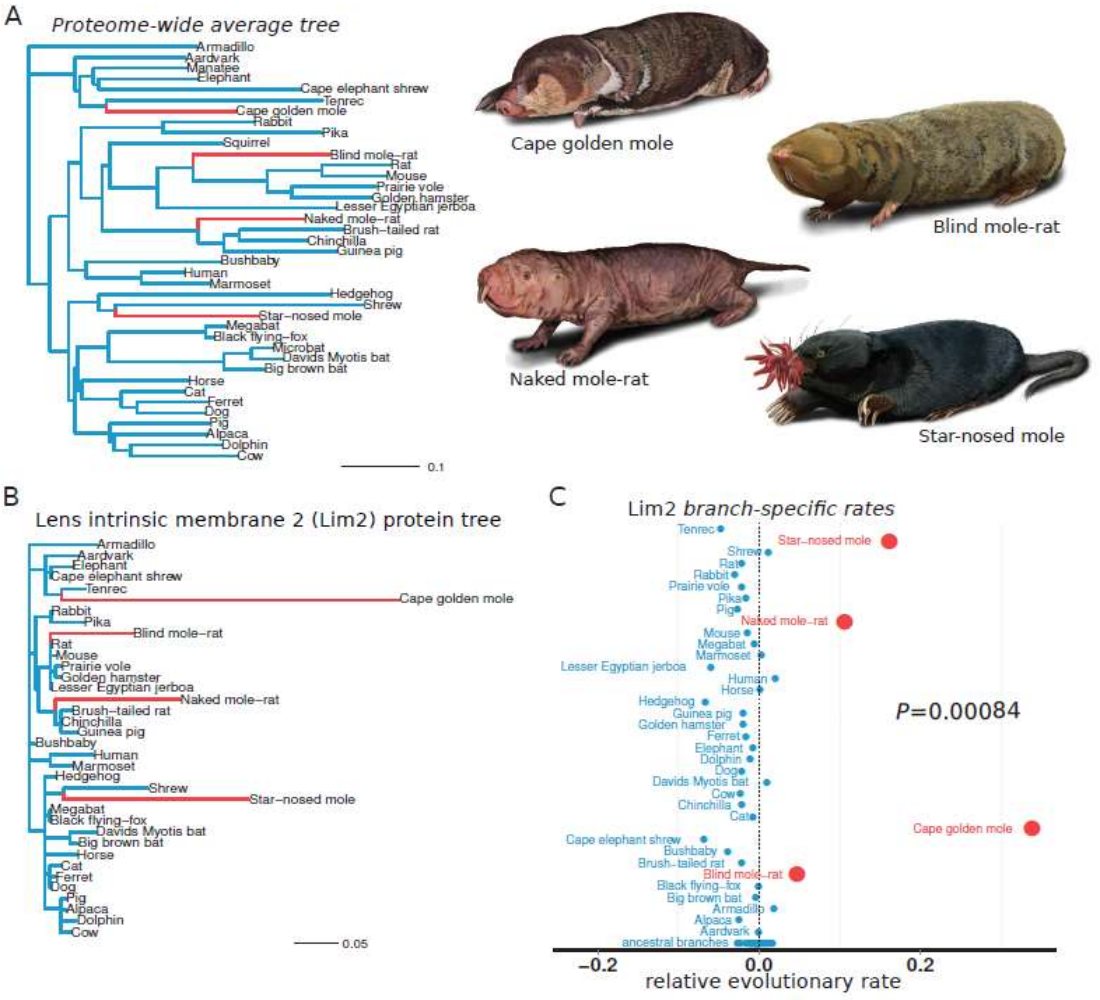
⇒ Indicate rapid evolution under selection

Hill & Unckless, 2018
https://doi.org/10.1016/j.meegid.2017.11.013

Comparative genomics

Study adaptive convergence across species

- What are the genes that evolved at a more rapid rate in organisms adapted to similar ecological conditions? That evolved comparable traits?



Partha, R.,... & Clark, N. L. (2017). Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *Elife*, *6*, e25884. https://doi.org/10.7554/eLife.25884.001

Trait-based approaches

Some local-adaptation/pop genomics approaches are also based on traits when contrasting populations with different phenotypes, morphotypes within a species or ecotypes…

⇒ **Need to characterize phenotypes and/or adaptative traits**

⇒ **Quantifying phenotypes can be complex (physiology, morphology, etc)**

Trait-based approaches

GWAS

Principle:

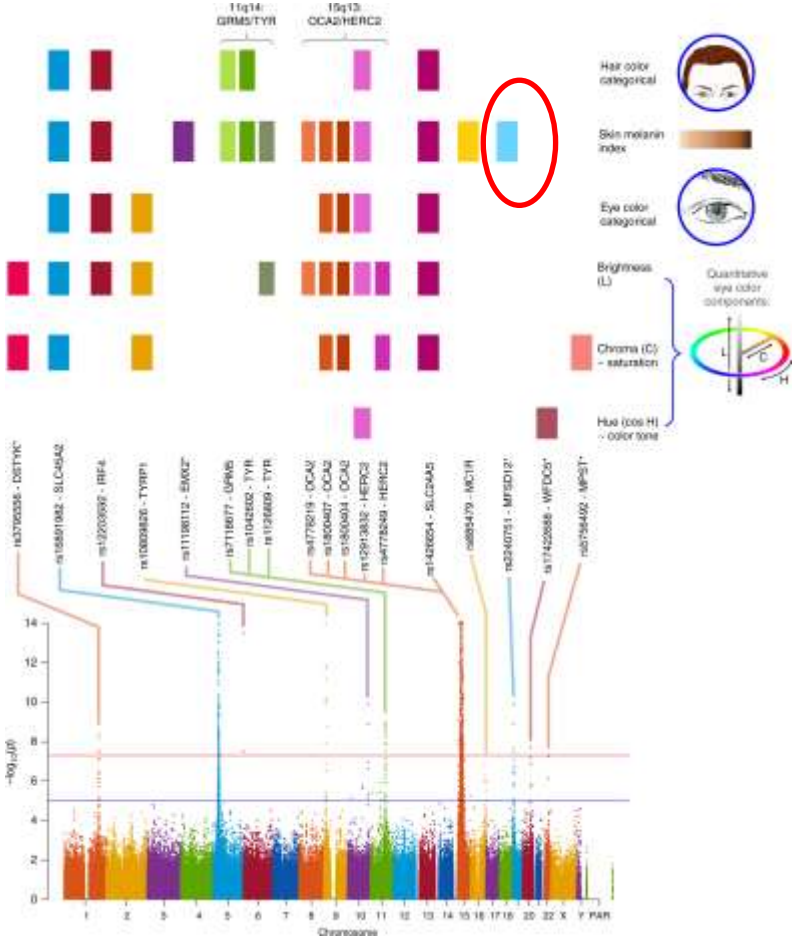
- **many individuals (hundreds to thousands)**
- **A trait is quantified for all individuals (case-control, quantitative)**
- **Many markers (usually SNP-array or Whole-genome)**

$\Rightarrow$ **Locus-by-locus association test**

$\Rightarrow$ **Polygenic scores**

Note many warning of the need to account for population structure

Trait-based approaches

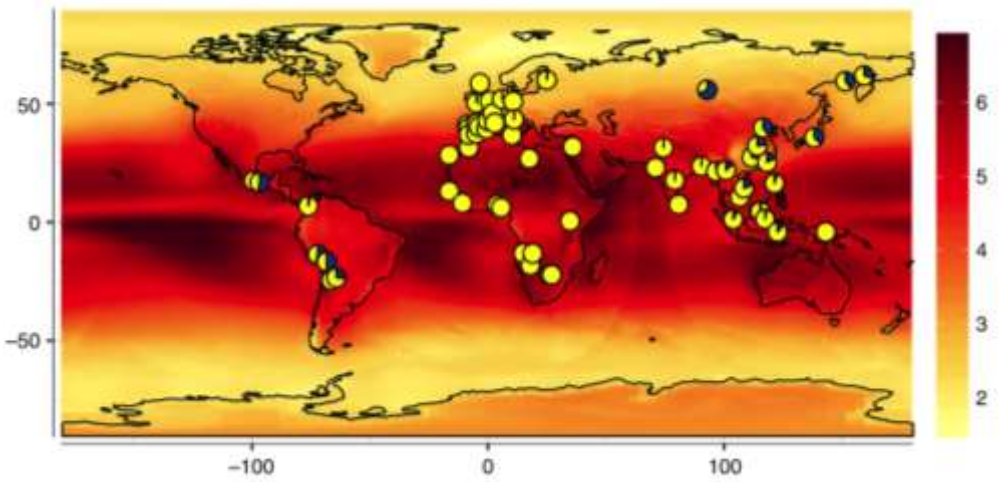


GWAS

6357 individuals

730,525 SNPs

pigmentation of skin and eyes



Adhikari, K., Mendoza-Revilla, J., Sohail, A. *et al.* A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. *Nat Commun* **10,** 358 (2019). https://doi.org/10.1038/s41467-018-08147-0

Trait-based approaches

QTL (quantitative trait loci)

QTL analyses are usually based on family-design with crosses (or recombinant lines or pedigree) between individuals with the phenotype of interest.

Linkage maps can be build with genetic markers - medium density is ok so RAD-seq is well-suited for that.

Then, the association between genotype and phenotype are tested in the progeny

Trait-based approaches

Huber, B., Whibley, A., Poul, Y. *et al.* Conservatism and novelty in the genetic architecture of adaptation in *Heliconius* butterflies. *Heredity* **114,** 515–524 (2015). https://doi.org/10.1038/hdy.2015.22

QTL (quantitative trait loci)

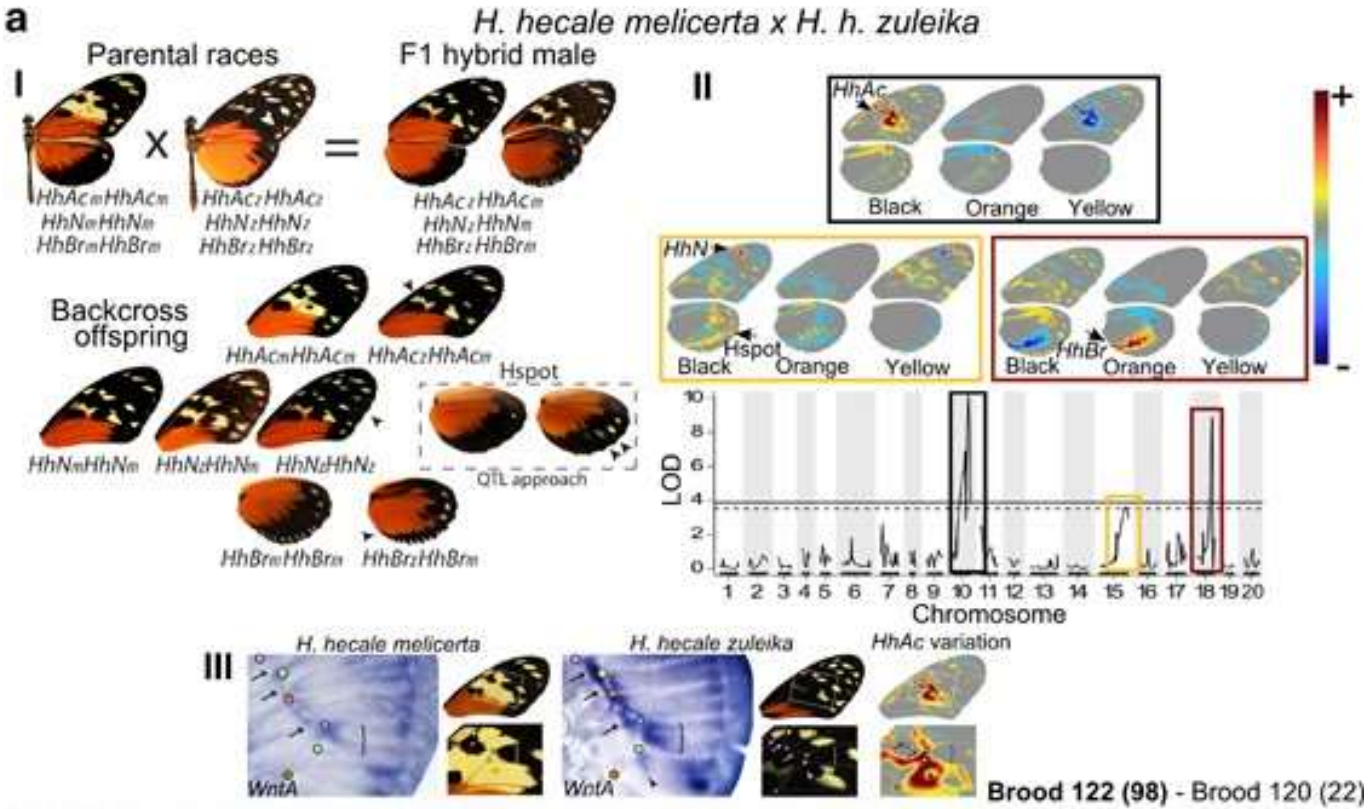
Targeted crosses and back-crosses

RAD-seq

Image analysis for colour pattern

Multivariate associations

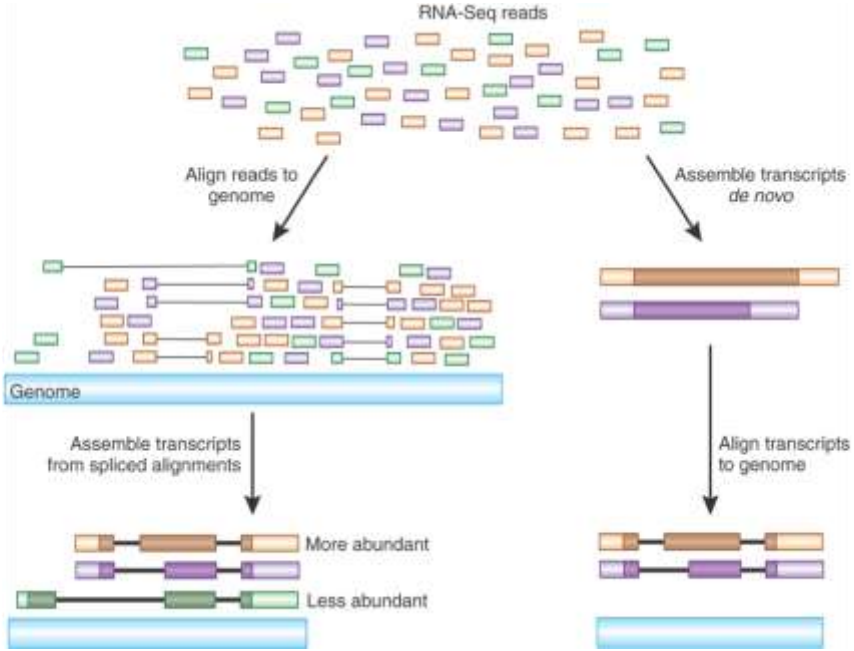
+ in situ hybridiation of WntA transcripts



Multi-omics

Transcriptomics

- Quantify RNA expression
- Can be done on the whole-organism or specific tissue
- qPCR on genes of interest or RNA-seq of all transcripts
- Need to build a transcriptome (collection of transcripts)



⇒Compare expression levels between two groups (differential expression)

⇒Find co-expressed genes

⇒eQTL if linked to phenotypes

⇒Allele-specific expression, isoform expression

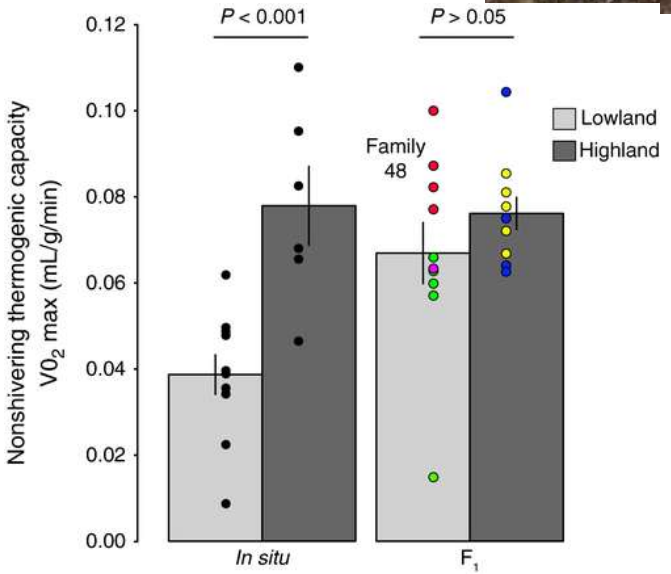
Multi-omics

RNA seq of brown adipose tissu
19 samples

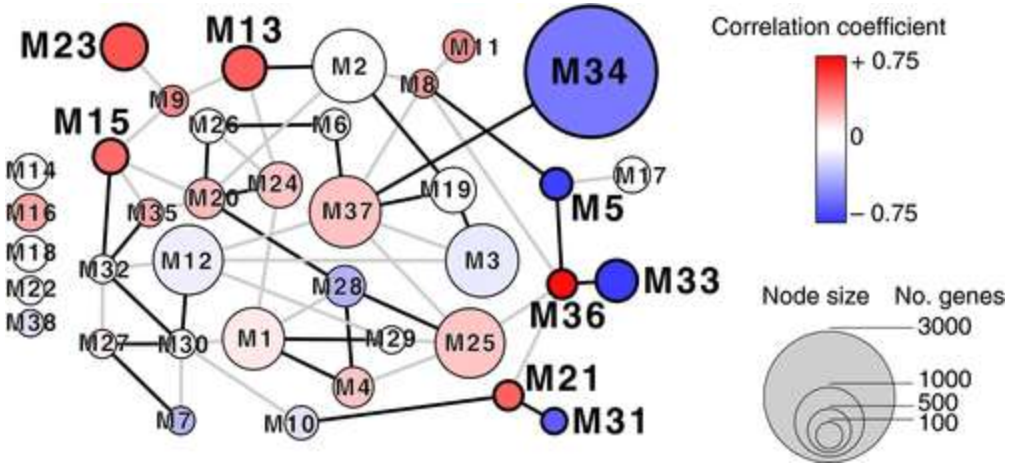
⇒ Genes co-expression network

⇒ Which modules are associated with thermogenic performance?

Transcriptomics



Field Lab



Velotta, J.P., Jones, J., Wolf, C.J. and Cheviron, Z.A. (2016), Transcriptomic plasticity in brown adipose tissue contributes to an enhanced capacity for nonshivering thermogenesis in deer mice. Mol Ecol, 25: 2870-2886. doi:10.1111/mec.13661

Multi-omics

Transcriptomics

⇒ Enrichment analysis for GO terms

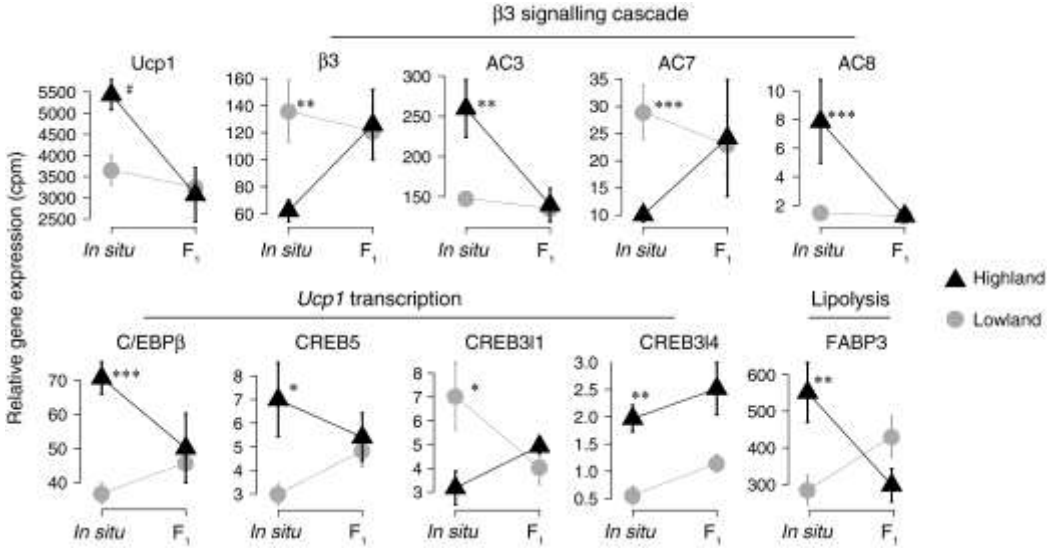
GO term	Description	*P*	*q*	Enrichment
GO:0001525	Angiogenesis	2.42E-09	4.37E-06	3.67
GO:0048514	Blood vessel morphogenesis	8.13E-05	1.03E-02	4.01
GO:0007219	*Notch* signalling pathway	8.61E-05	1.07E-02	3.7
GO:0043547	Positive regulation of GTPase activity	4.66E-07	2.19E-04	2.76
GO:1904018	Positive regulation of vasculature development	1.05E-04	1.27E-02	3.41
GO:0045765	Regulation of angiogenesis	1.25E-07	8.31E-05	3.68
GO:0070372	Regulation of ERK1 and ERK2 cascade	6.79E-05	9.15E-03	2.93
GO:0048729	Tissue morphogenesis	2.94E-07	1.69E-04	2.82
GO:0001570	Vasculogenesis	4.13E-04	3.74E-02	3.94

Velotta, J.P., Jones, J., Wolf, C.J. and Cheviron, Z.A. (2016), Transcriptomic plasticity in brown adipose tissue contributes to an enhanced capacity for nonshivering thermogenesis in deer mice. Mol Ecol, 25: 2870-2886. doi:10.1111/mec.13661

Multi-omics

Transcriptomics

⇒ Candidate genes



- β3 adrenergic receptor signalling cascade
- *Ucp1* transcription
- lipolysis

=> facilitate non-shivering thermogenesis within brown adipocytes

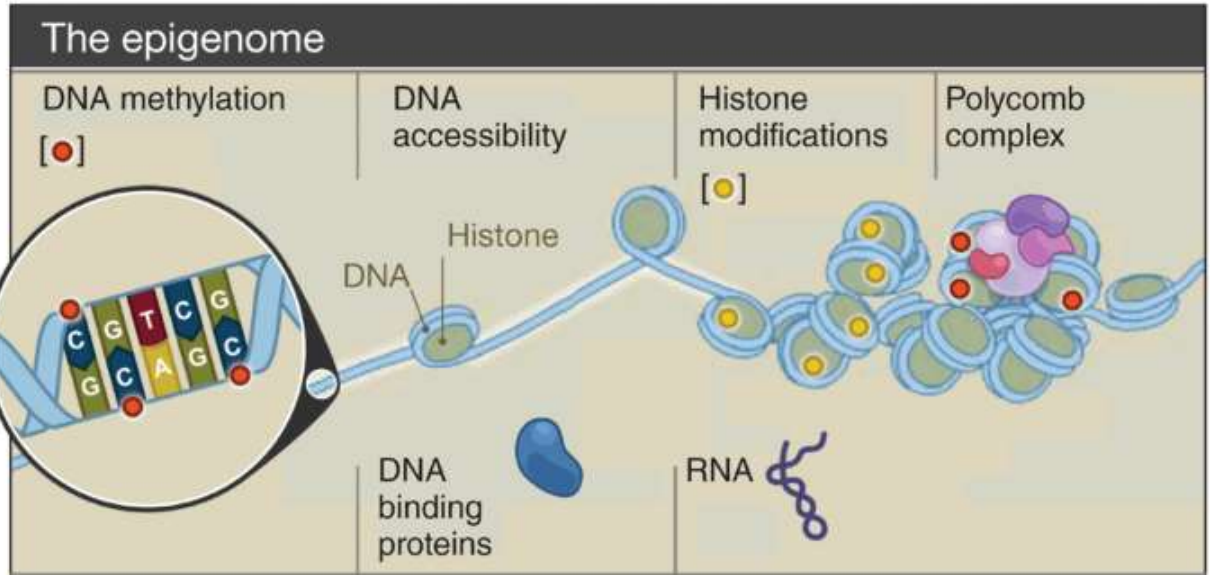
Velotta, J.P., Jones, J., Wolf, C.J. and Cheviron, Z.A. (2016), Transcriptomic plasticity in brown adipose tissue contributes to an enhanced capacity for nonshivering thermogenesis in deer mice. Mol Ecol, 25: 2870-2886. doi:10.1111/mec.13661

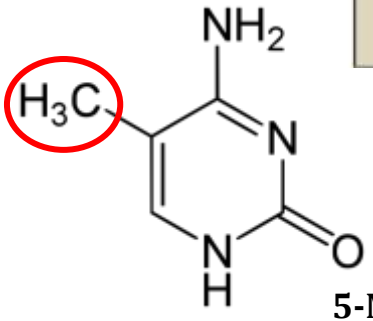
Multi-omics

Epi-genomics

Epigenome is change to the DNA or histones, can be plastic and/or heritable, vary between tissues

The most accessible aspect for studies in ecology and evolution will be DNA methylation

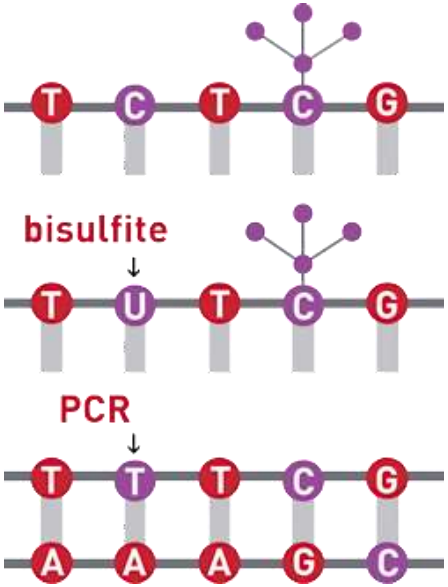




5-Méthylcytosine

Multi-omics

Epi-genomics : DNA methylation



use of bisulfite treatment of DNA before routine sequencing

⇒ Unmethylated C are converted into T

⇒ Remaining C are the ones that were metylated

100%	50%	✘
C	C	C
C	C	T
C	C	T
C	T	T
C	T	T
C	T	T
C/C	C/C	C/T

Studied variation is the % of methylated cytosine at a given position

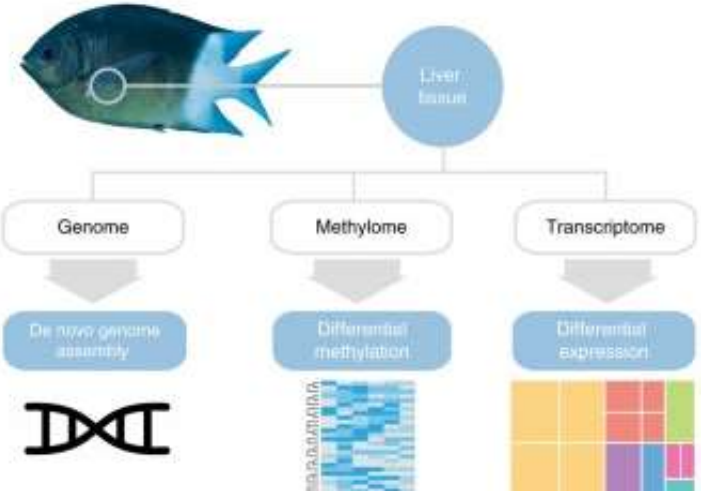
Cannot look at sites with C/T polymorphism

⇒ Variation in methylation level between tissues, between samples, between sexes, between populations, between ecotypes, etc…

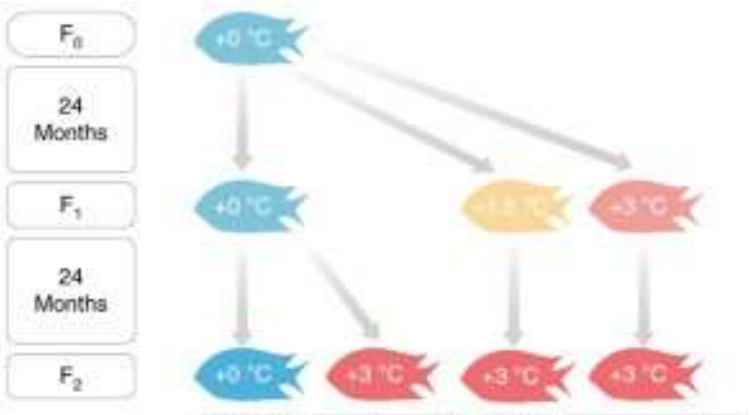
Multi-omics

Epi-genomics : DNA methylation

- coral reef fish, *Acanthochromis polyacanthus*
- liver genome, methylomes and transcriptomes



Exposed to current day (+0 °C) or future ocean temperatures (+3 °C) for one generation, two generations and incrementally across generations

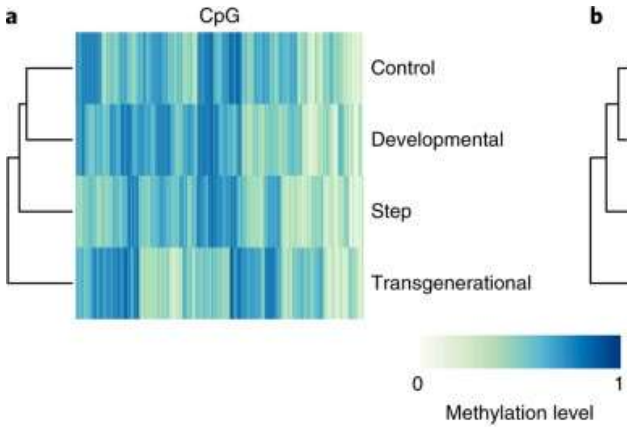


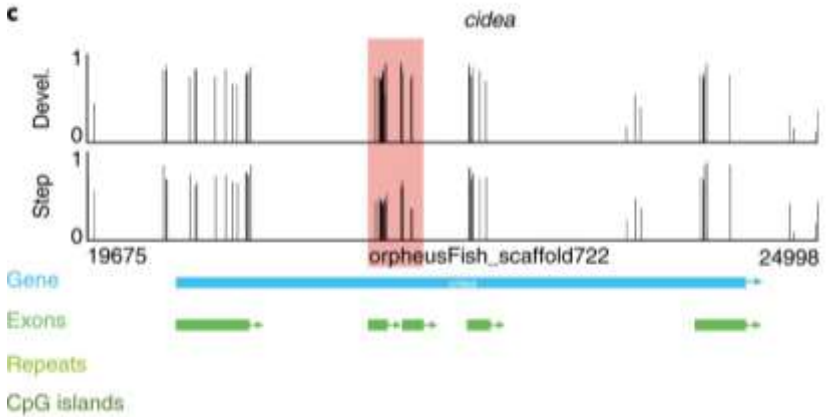
Genome	No. of samples	1			
	No. of bases (Gb)	249.8 (5 PE reads) and 217.7 (7 MP reads)			
Methylome	No. of samples	4	4	4	4
	No. of bases (Gb)	443.7	476.5	488.9	473
Transcriptome	No. of samples	5	5	4	5
	No. of bases (Gb)	29.2	26.3	14	23.4

Ryu, T., Veilleux, H.D., Donelson, J.M. *et al.* The epigenetic landscape of transgenerational acclimation to ocean warming. *Nature Clim Change* **8,** 504–509 (2018). https://doi.org/10.1038/s41558-018-0159-0

Multi-omics

Epi-genomics : DNA methylation





2,467 differentially methylated regions (DMRs) and 1,870 associated genes that respond to higher temperatures within and between generations

Some genes also show differential expression (but not many of them)

⇒ Association between DNA methylation and transgenerational acclimation to climate change

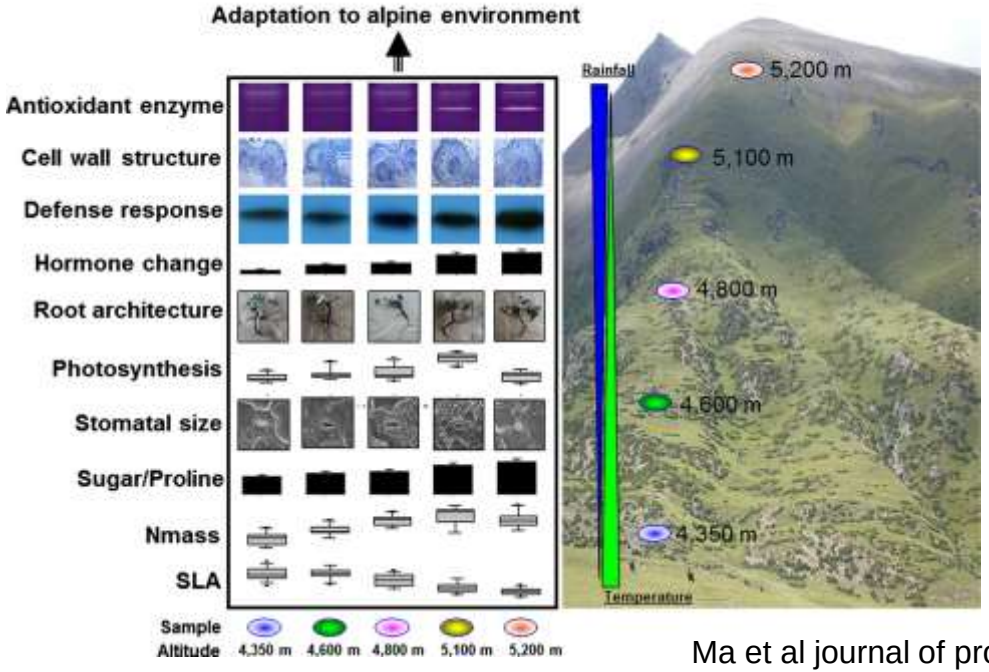
Ryu, T., Veilleux, H.D., Donelson, J.M. *et al.* The epigenetic landscape of transgenerational acclimation to ocean warming. *Nature Clim Change* **8,** 504–509 (2018). https://doi.org/10.1038/s41558-018-0159-0

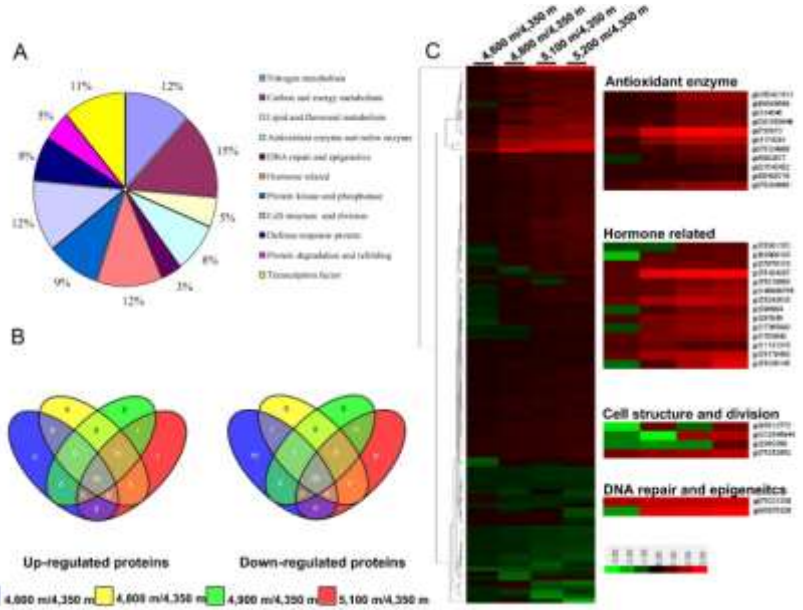
Multi-omics

Proteomics : characterization of all proteins with large-scale mass spectrometry

⇒ an analysis of morphology and proteome along an altitudinal gradient

⇒ Identify pathways involved in adaptation (hormones, anti-oxydant, epi-genetic regulations, etc.





Ma et al journal of proteomics 2015 https://doi.org/10.1016/j.jprot.2014.08.009

Experimental & genomics

Experimental evolution / Experimental selection

From a starting population which is usually clonal (in yeast/bacteria) or highly-diverse (diploid species, drosophila, insects, etc), several replicates are kept under controlled conditions (one or several treatments) for a given number of generations (5 to 1 000…)

Phenotypes can be regularly measured

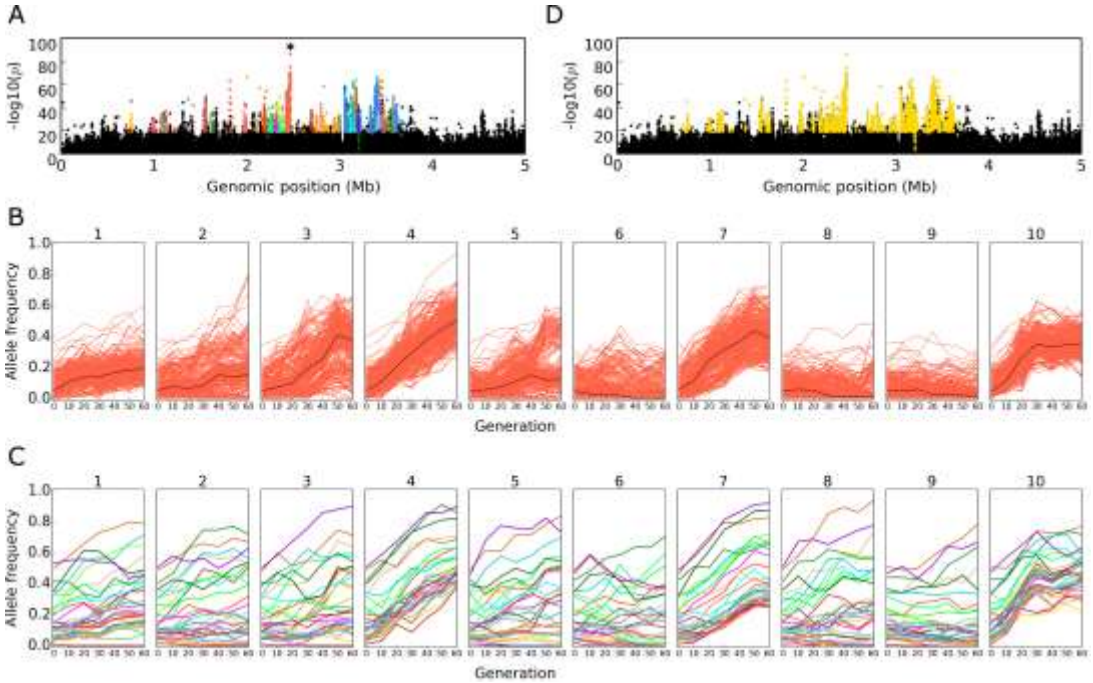
With genomics:

Usually pool-seq. The change of allelic frequencies is followed at regular intervals or contrasted between the beginning and the end of the experiment

$\Rightarrow$ Powerful way to follow evolution “under our eyes”!

Experimental & genomics

Experimental evolution / Experimental selection



Barghi N, Tobler R, Nolte V, Jakšić AM, Mallard F, Otte KA, et al. (2019) Genetic redundancy fuels polygenic adaptation in *Drosophila*. PLoS Biol 17(2): e3000128. https://doi.org/10.1371/journal.pbio.3000128

10 replicates of a *Drosophila simulans* population to a new temperature regime

Pool-seq

Convergent responses for several phenotypes

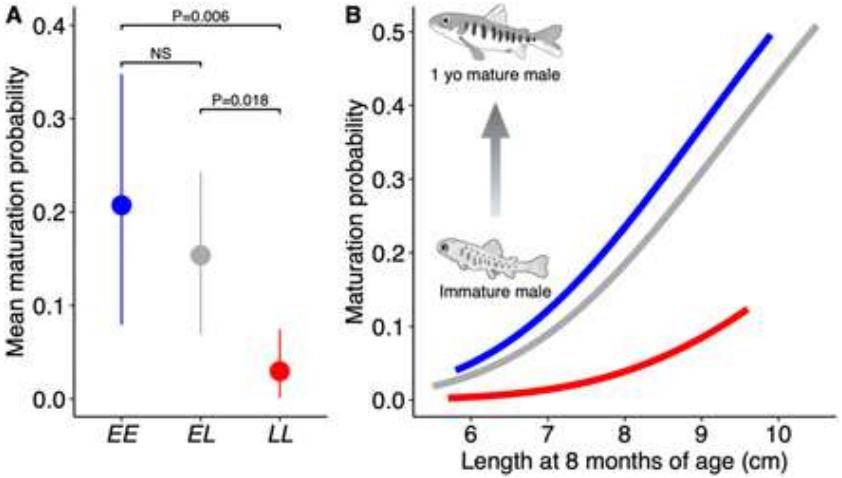
a strong polygenic response (99 selected alleles; mean s = 0.059)

⇒ redundancy : not the same loci contribute to the evolution of the same phenotypes between replicates

Experimental & genomics

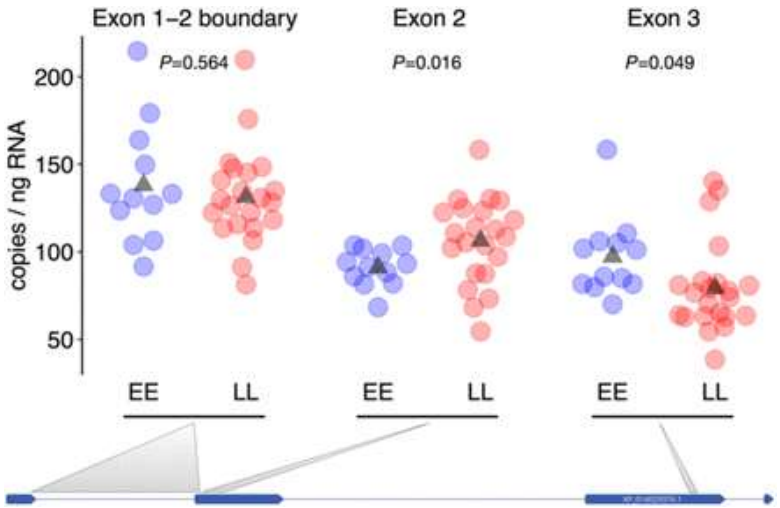
Cis-regulatory differences in isoform expression associate with life history strategy variation in Atlantic salmon
Jukka-Pekka Verta et al
bioRxiv 777300; doi: https://doi.org/10.1101/777300

Common-garden



A ***candidate gene*** from GWAS for age at maturity in Salmon : vgll3

⇒ Genotype and raise in controlled conditions 656 individuals



⇒ Trancriptomics : complex pattern of differential expression:
It depends on the exon…

⇒ A different isoform expressed in EE and LL : ***isoform-specific expression***

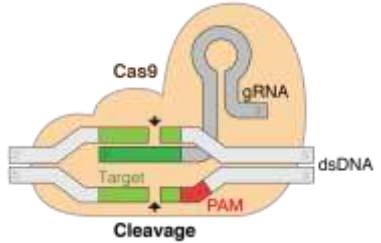
Experimental & genomics

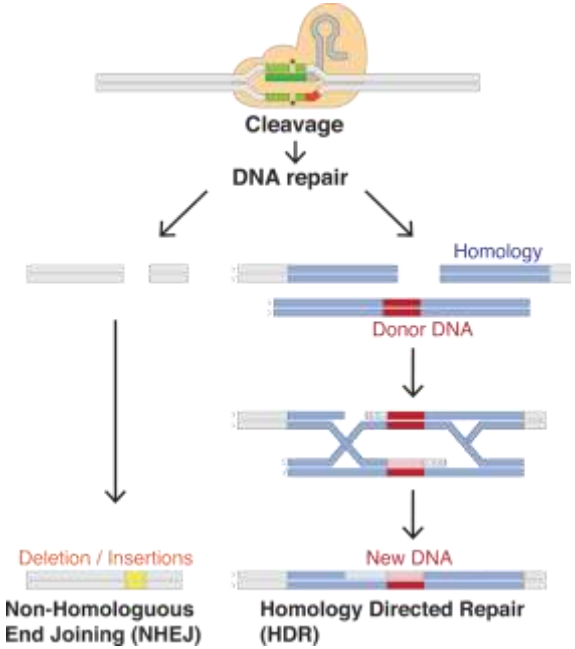
Gene-editing

Target a candidate gene to silence it or change its sequence:

-> used to be possible only in model species

-> CRISPR-Cas9 is a game-changer





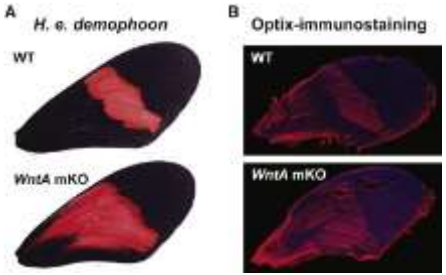
Experimental & genomics

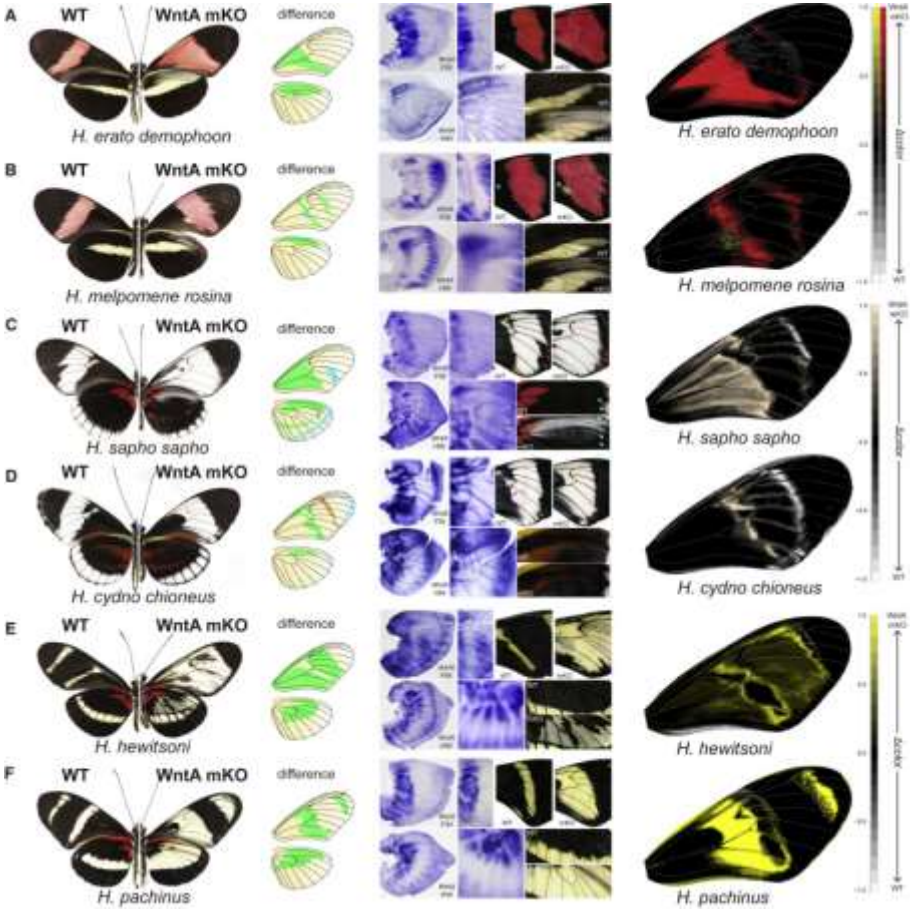
Gene-editing

Inject eggs and larva with CRISPR-CAs9 to make *WntA* KO mutant
-> generation 0: 30% with a mutant phenotype (indels in ~80% of PCR *WntA)*

⇒ Confirm the relation genotype/phenotype
⇒ Compare multiple species
⇒ Matches the area of influence during development

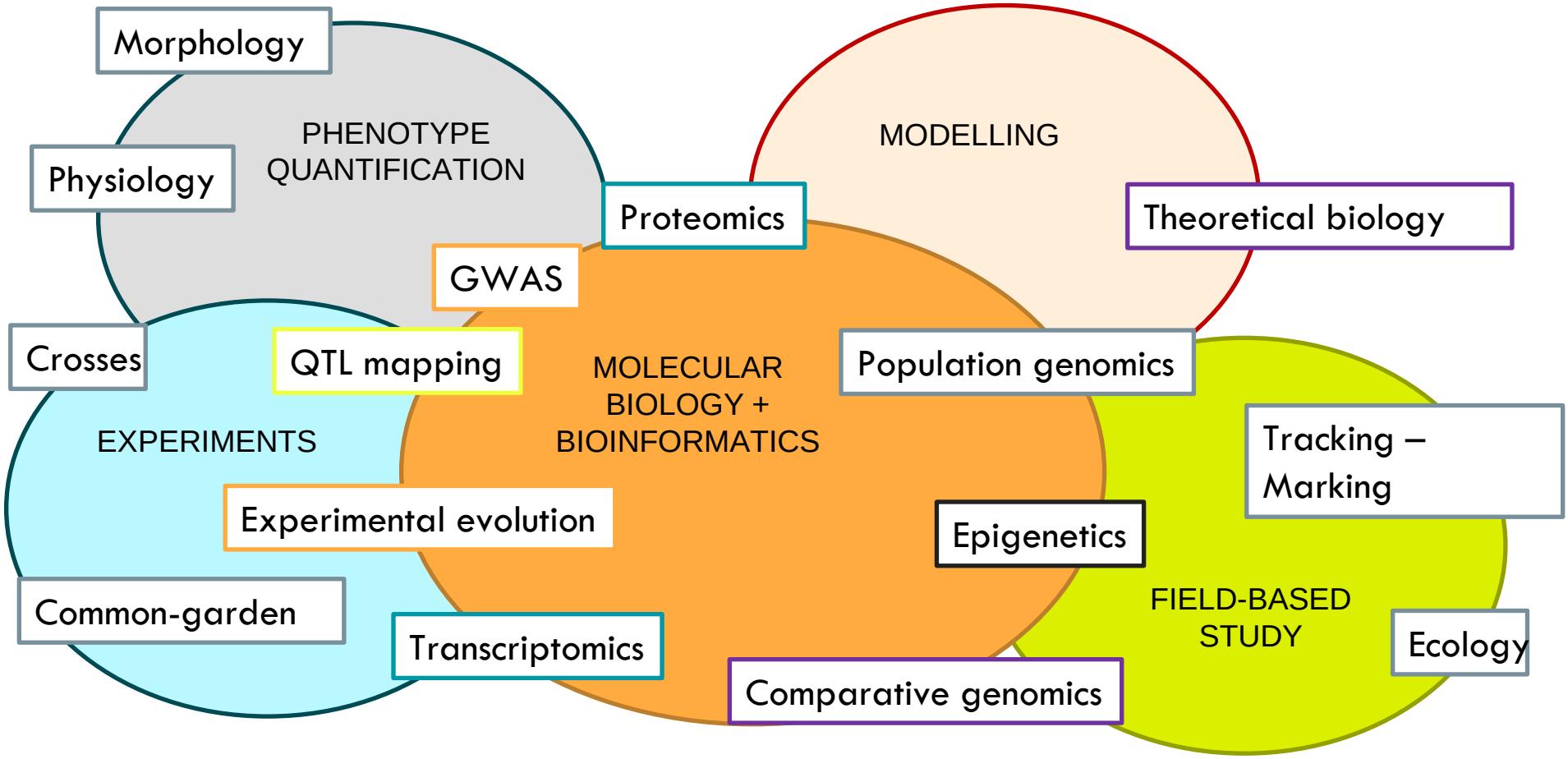
+ antibody immunostaining of transcripts





Concha, C., Wallbank, R. W., Hanly, J. J., Fenner, J., Livraghi, L., Rivera, E. S., ... & Morrison, C. (2019). Interplay between developmental flexibility and determinism in the evolution of mimetic Heliconius wing patterns. *Current Biology*, *29*(23), 3996-4009. https://doi.org/10.1016/j.cub.2019.10.010

Studying adaptation… Integrative biology!



Tutorial day 5

Most methods that we saw during the week will provide

⇒ **General knowledge about isolation-by-adaptation, the genetic architecture of adaptation, an idea of genomic variance related to possible ecological variation, etc …**

⇒ **Putatively-adapted SNPs, SVs or genomic regions**

- Can we point towards causal candidate genes or pathways ?

Local adaptation / population genomics

Gene annotation, gene ontology, gene enrichment

Genome + transcriptome + protein databases + transposable elements databases

⇒ By aligning the transcriptome on the genome we can know gene positions (and exon, intron, etc…)

⇒ The transcriptome can be annotated thanks to protein databases (protein sequences usually more conserved than DNA sequences)

⇒ Genes/Proteins are gather into functional categories called « gene ontology »
http://geneontology.org/docs/ontology-documentation/

⇒ Thanks to TE databases and repeat detection, the genome can be annotated for interspersed reapeats.

Tutorial day 5

We will:

- **Annotate the SNPs to know whether they belong to exon, intron, regulatory regions**
- **Look for genes at the proximity of our outlier SNPs**
- **Test for enrichment in the outliers for particular GO categories**
- **Investigate whether some of the CNV are transposable elements or repeated regions**

http://geneontology.org/docs/ontology-documentation/

- SFS

5 diploid individuals