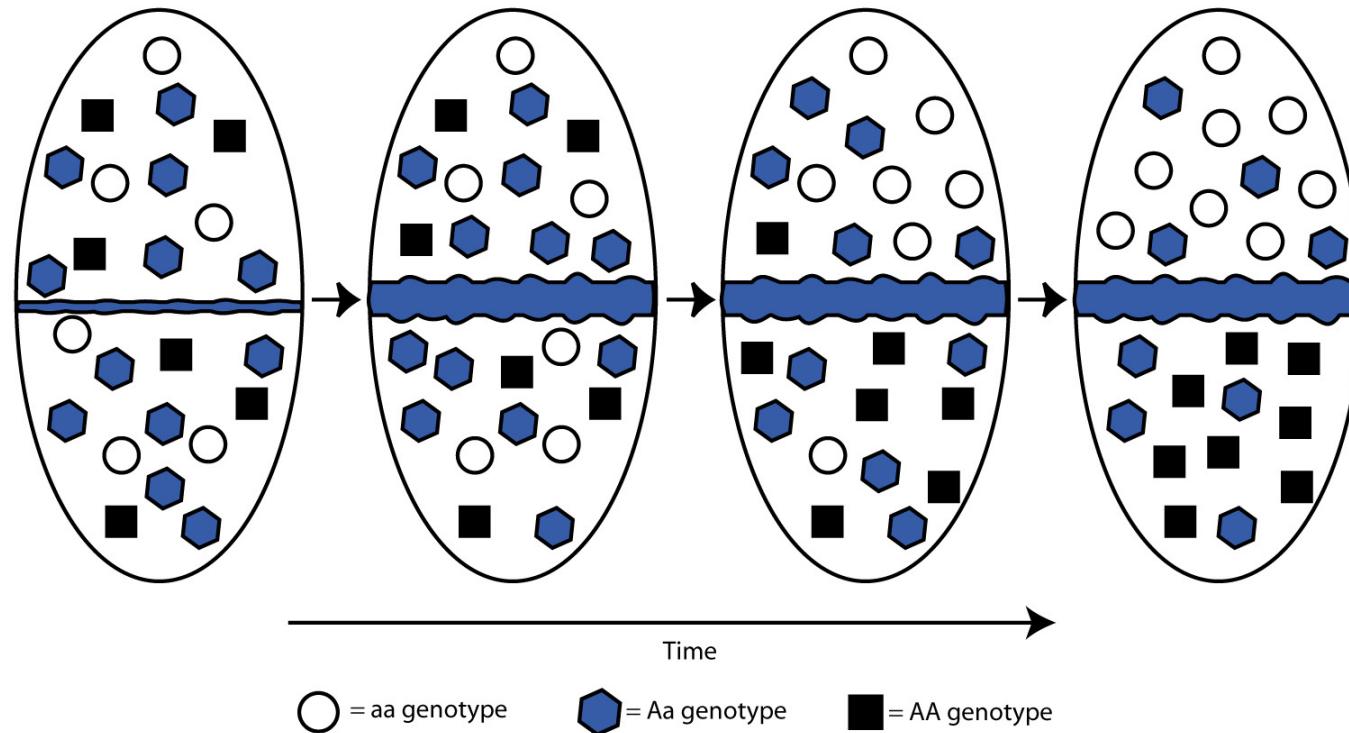


Population structure

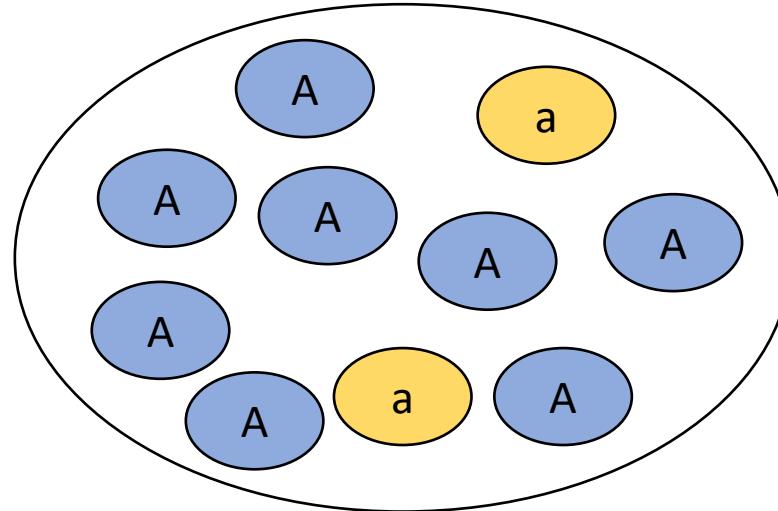
Week 3

Population structure

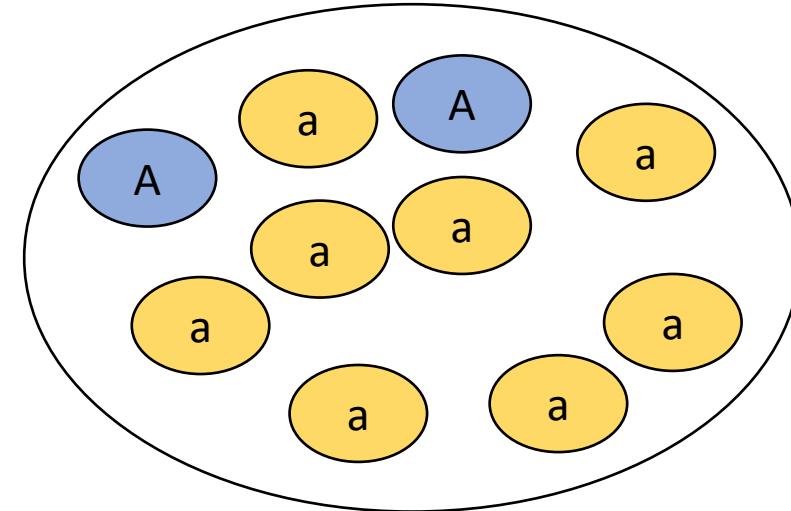
- When a population is not randomly mating because of geographic reasons
- Up to now: a single, random mating (panmictic) population
- Clearly not always the case!
- Example: separation by a geographic barrier (vicariance)



Wahlund effect



$$P = 0.8; q = 0.2$$



$$P = 0.2; q = 0.8$$

- Population structure generates non-random mating..
 - Violates HW
- But what if you don't know there is structure?
- Look at heterozygosity
 - Causes a deficit

F_{ST}

- Most common way to measure population subdivision:
 - Allele frequency differences between populations

$$F_{ST} = \frac{H_T - H_S}{H_T} = 1 - \frac{H_S}{H_T}$$

$H_S = \frac{1}{n} \sum_{i=1}^n 2p_i q_i$ Mean expected heterozygosity within subpopulations (assuming HW within subpops)

$H_T = 2\bar{p}\bar{q}$ Expected heterozygosity of the total population (assuming HW within the total pop)

the degree to which random mating (HW) expectations for the frequencies of heterozygotes are not met

$$\bar{p} = \bar{q} = 0.5$$

$$H_T = 0.5$$

$$H_S = 0.0$$

$$F_{ST} = 1.0$$

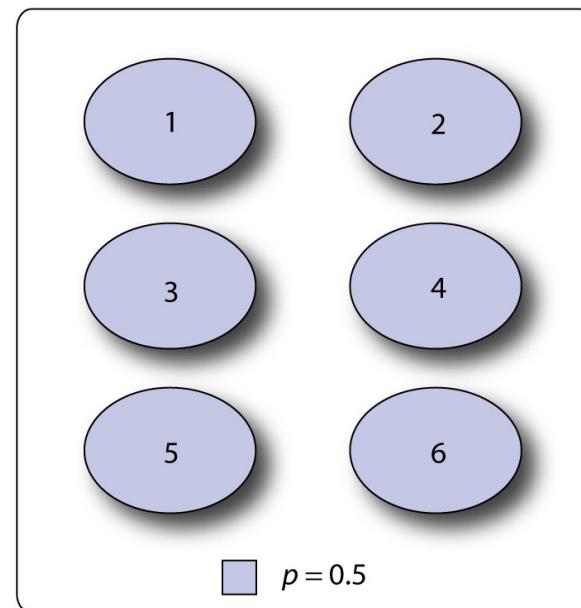
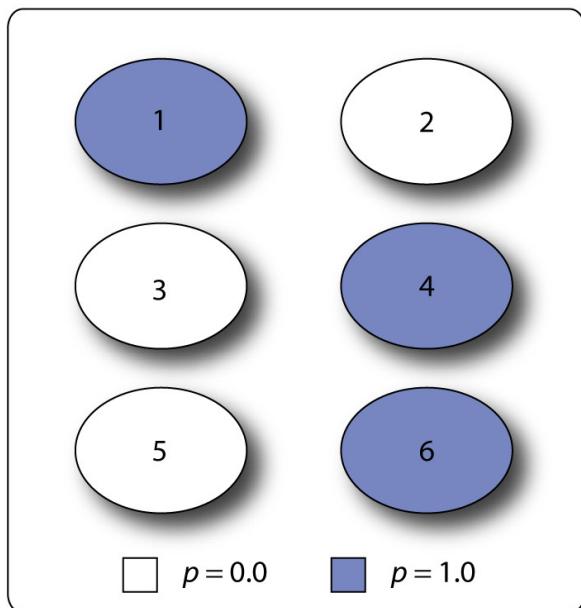
$$\bar{p} = \bar{q} = 0.5$$

$$H_T = 0.5$$

$$H_S = 0.5$$

$$F_{ST} = 0.0$$

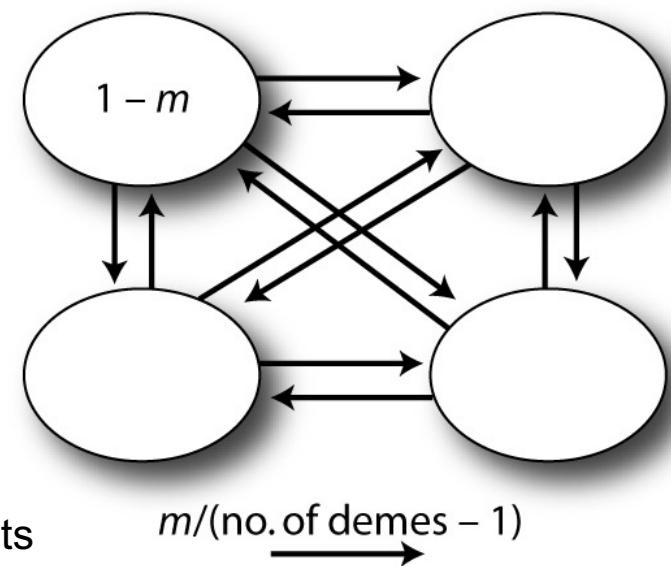
$$F_{ST} = \frac{H_T - H_S}{H_T}$$



Island Model

- Very large (infinite) number of subpopulations
- All subpopulations have the same size
- Gene flow (m) between all pairs of subpopulations
- Gene flow is equal between all pairs of subpopulations
- Gene flow is symmetrical between all subpopulation pairs
- Random mating within populations
- No selection

m : proportion of migrants



F_{ST} and migration

Assuming island model and that $m \ll N_e$ it can be shown that (Wright 1931, 1951)

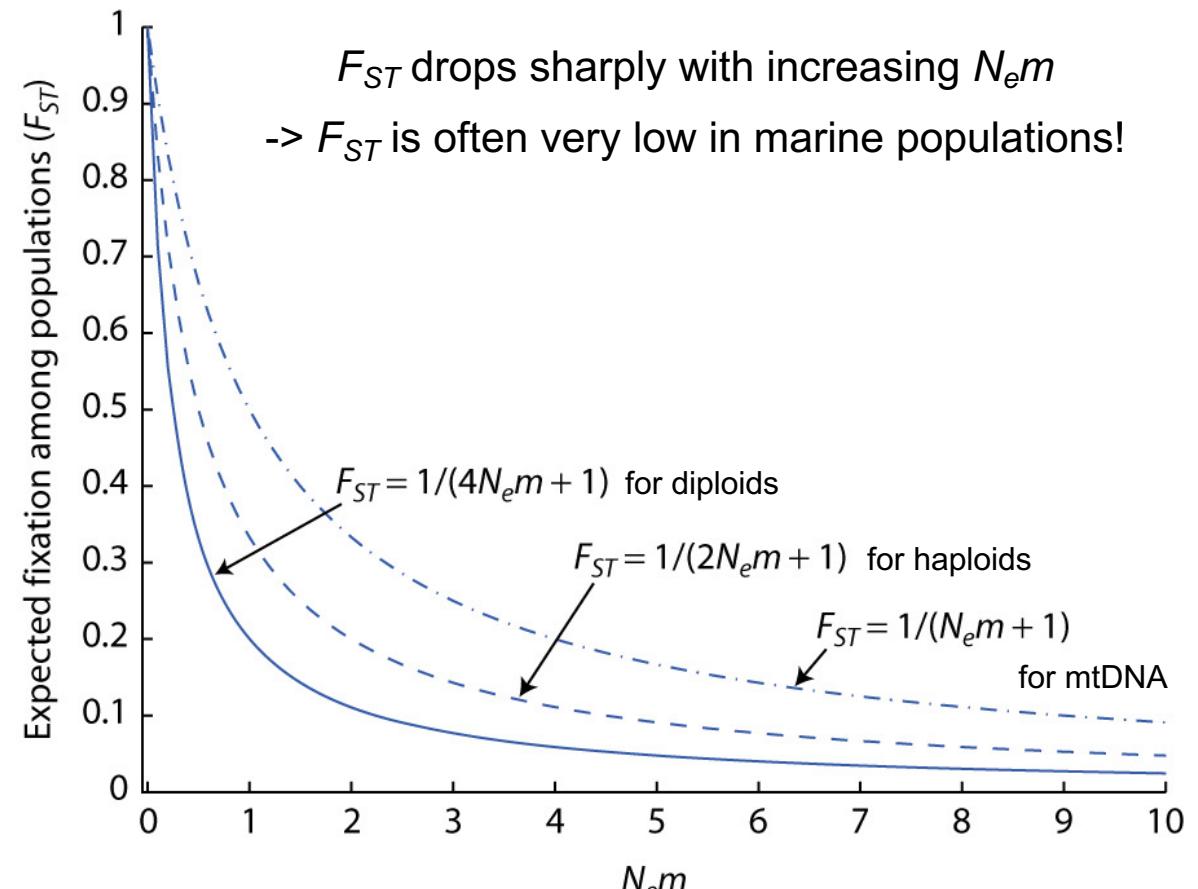
$$F_{ST} \approx \frac{1}{4N_e m + 1}$$

$N_e m$: effective number of migrants

~one migrant per generation ($N_e m = 1$) is sufficient to prevent genetic divergence

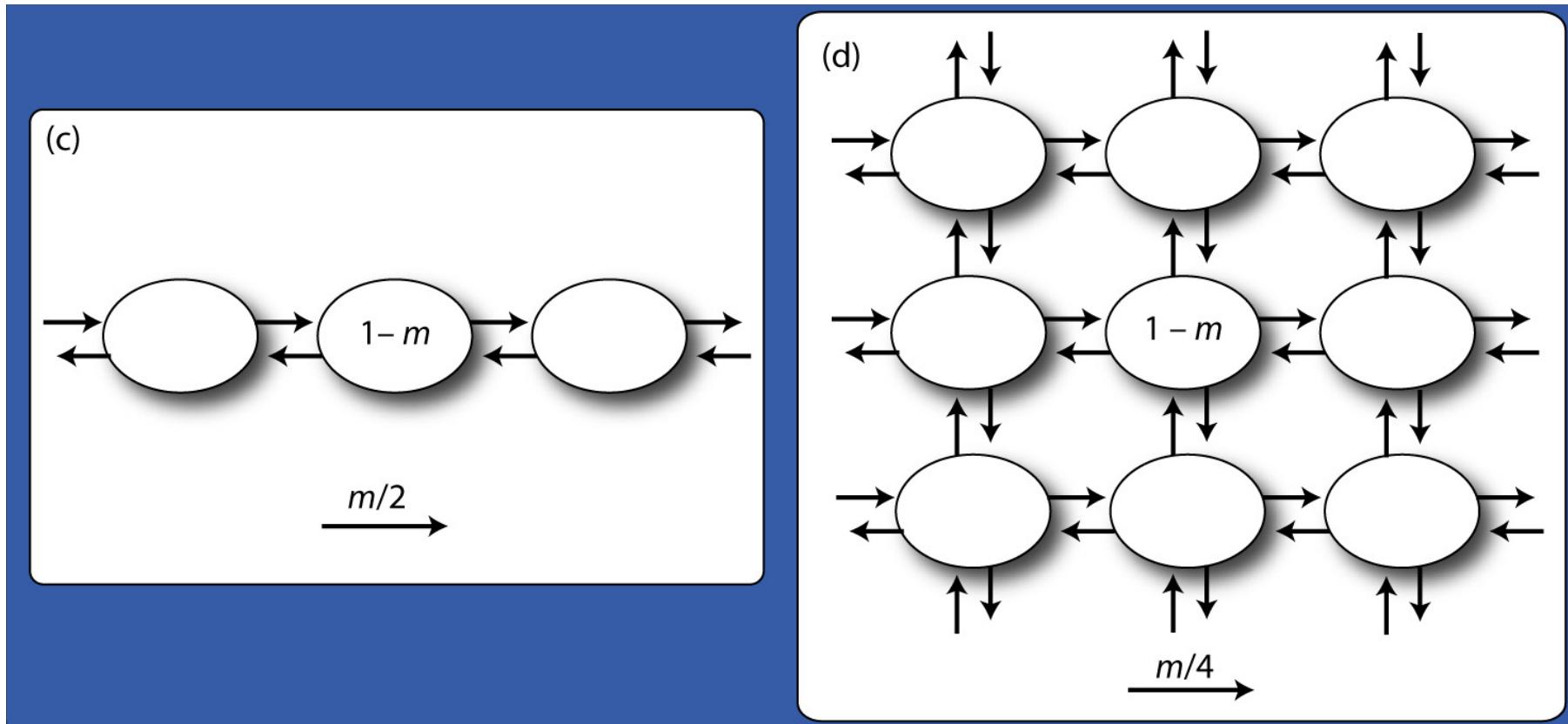
Suggests a way to estimate $N_e m$

$$N_e m \approx \frac{1 - F_{ST}}{4F_{ST}}$$



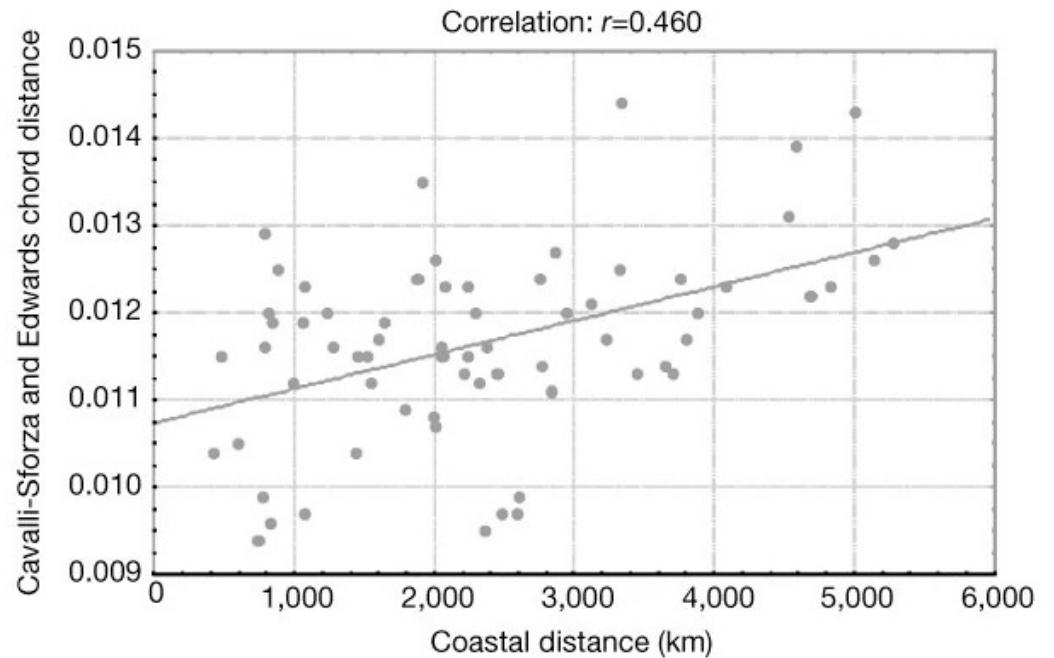
Stepping-stone model ((Kimura 1953; Kimura & Weiss 1964)

- Migration among neighbors only



Isolation by distance

Example: European eel (*Anguilla anguilla*) (Wirth & Bernatchez 2001)



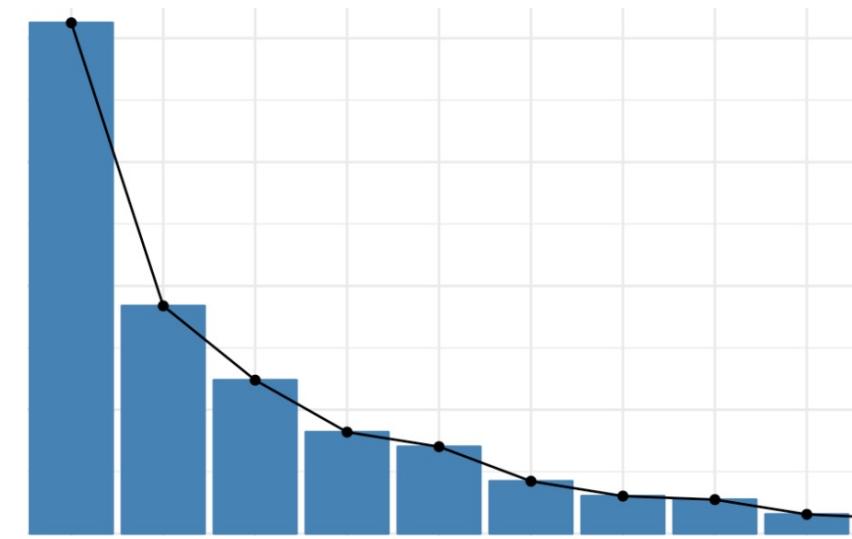
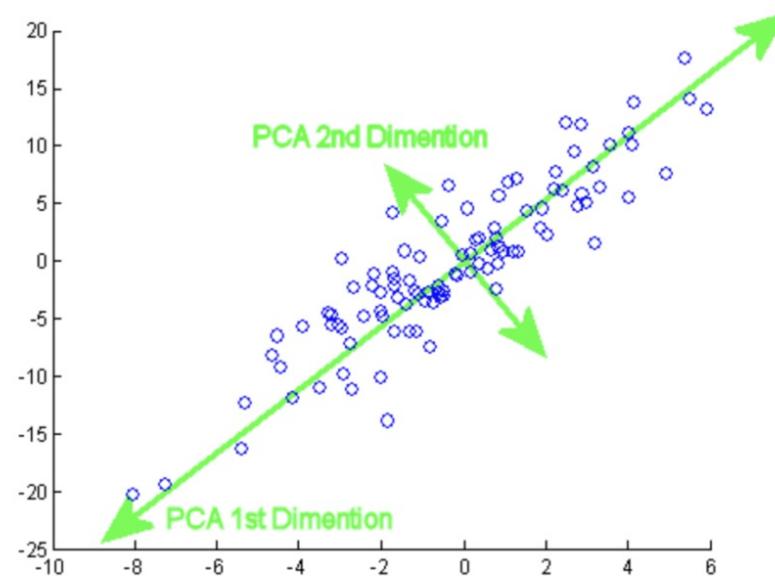
Typically measured as Cavalli-Sforza and Edwards chord distance

Or

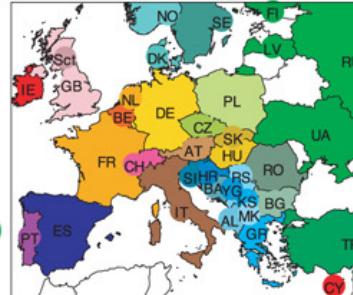
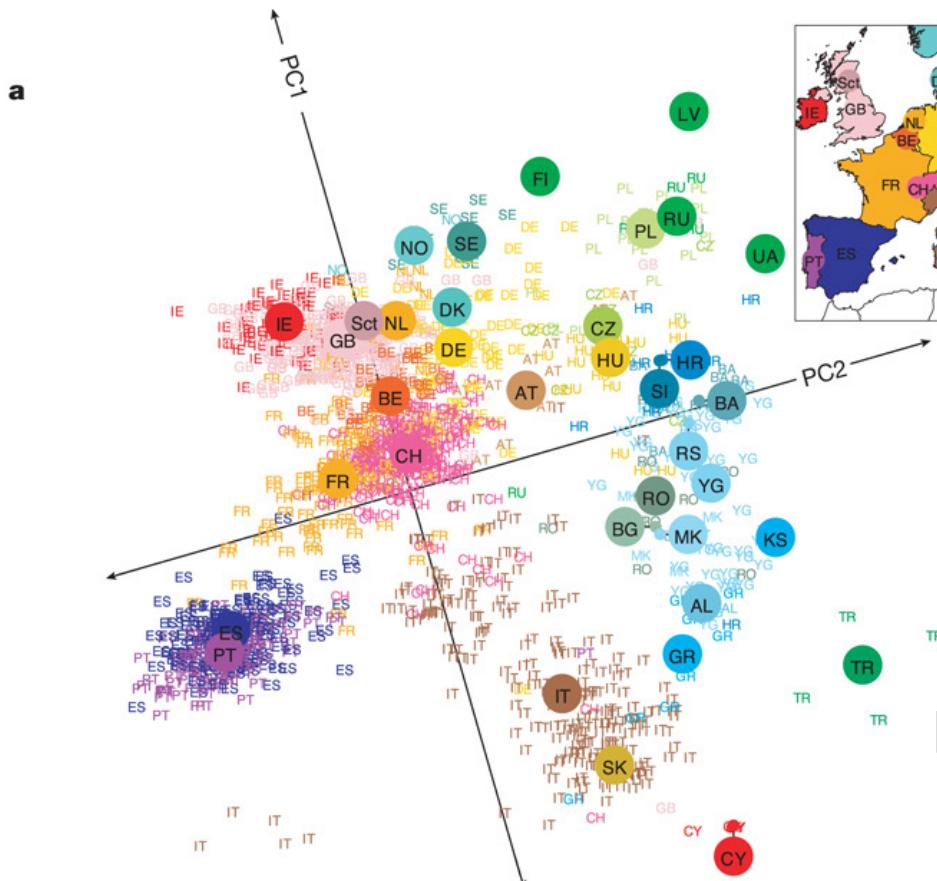
$$\frac{F_{ST}}{1 - F_{ST}}$$

Principal component analysis

- Goal: explain high dimension data in lower dimension while retaining information



Principal component analysis



Cannot assign individuals into populations

LETTERS

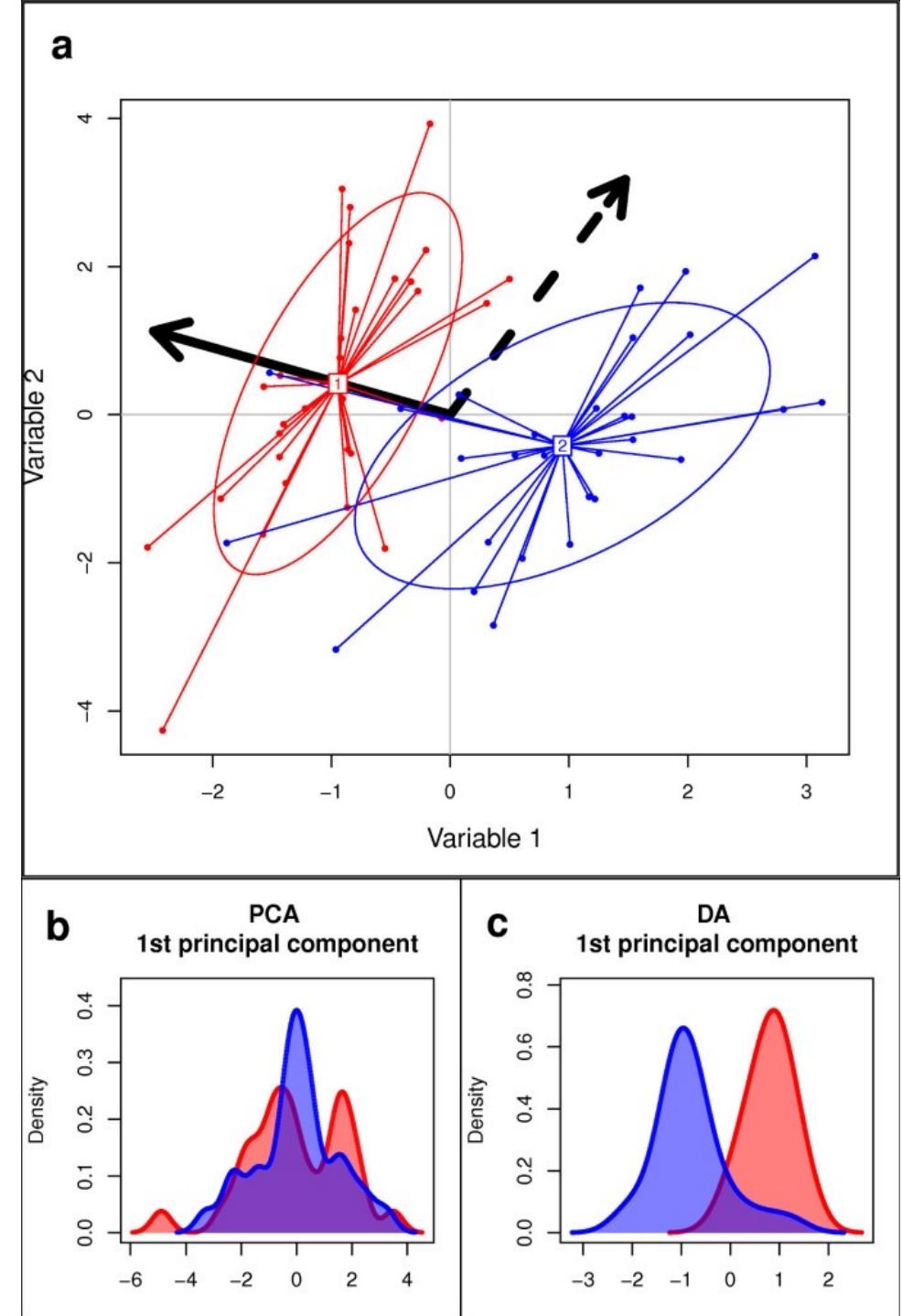
Genes mirror geography within Europe

John Novembre^{1,2}, Toby Johnson^{4,5,6}, Katarzyna Bryc⁷, Zoltán Kutalik^{4,6}, Adam R. Boyko⁷, Adam Auton⁷, Amit Indap⁷, Karen S. King⁸, Sven Bergmann^{4,6}, Matthew R. Nelson⁸, Matthew Stephens^{2,3} & Carlos D. Bustamante⁷

Vol 456 | 6 November 2008 | doi:10.1038/nature07331

Discriminant Analysis of Principal Components

- summarize the genetic differentiation between groups, while overlooking within-group variation.
- *discrimination* of individuals into pre-defined groups
- free of assumptions about Hardy-Weinberg equilibrium or linkage disequilibrium

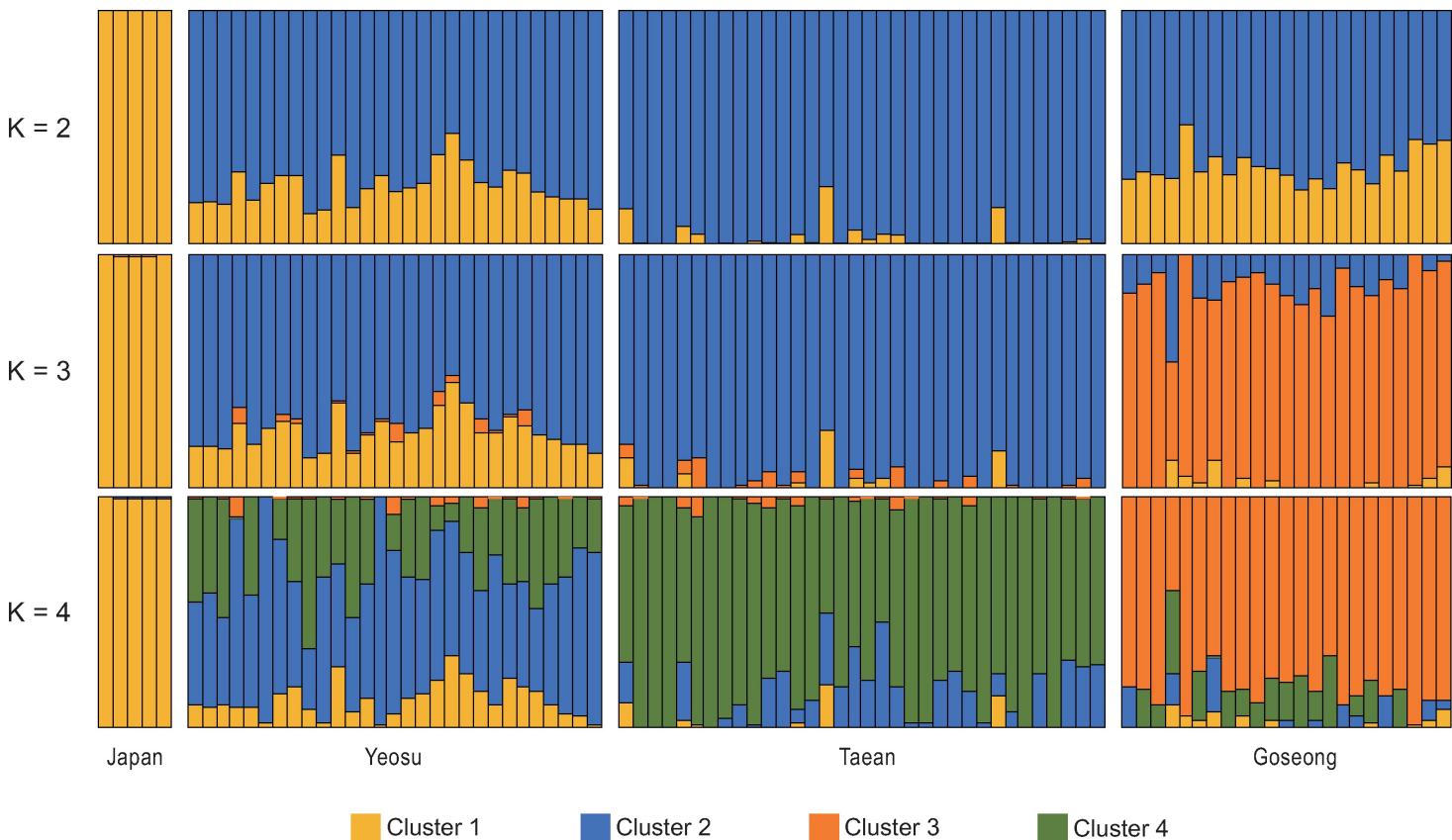


Structure

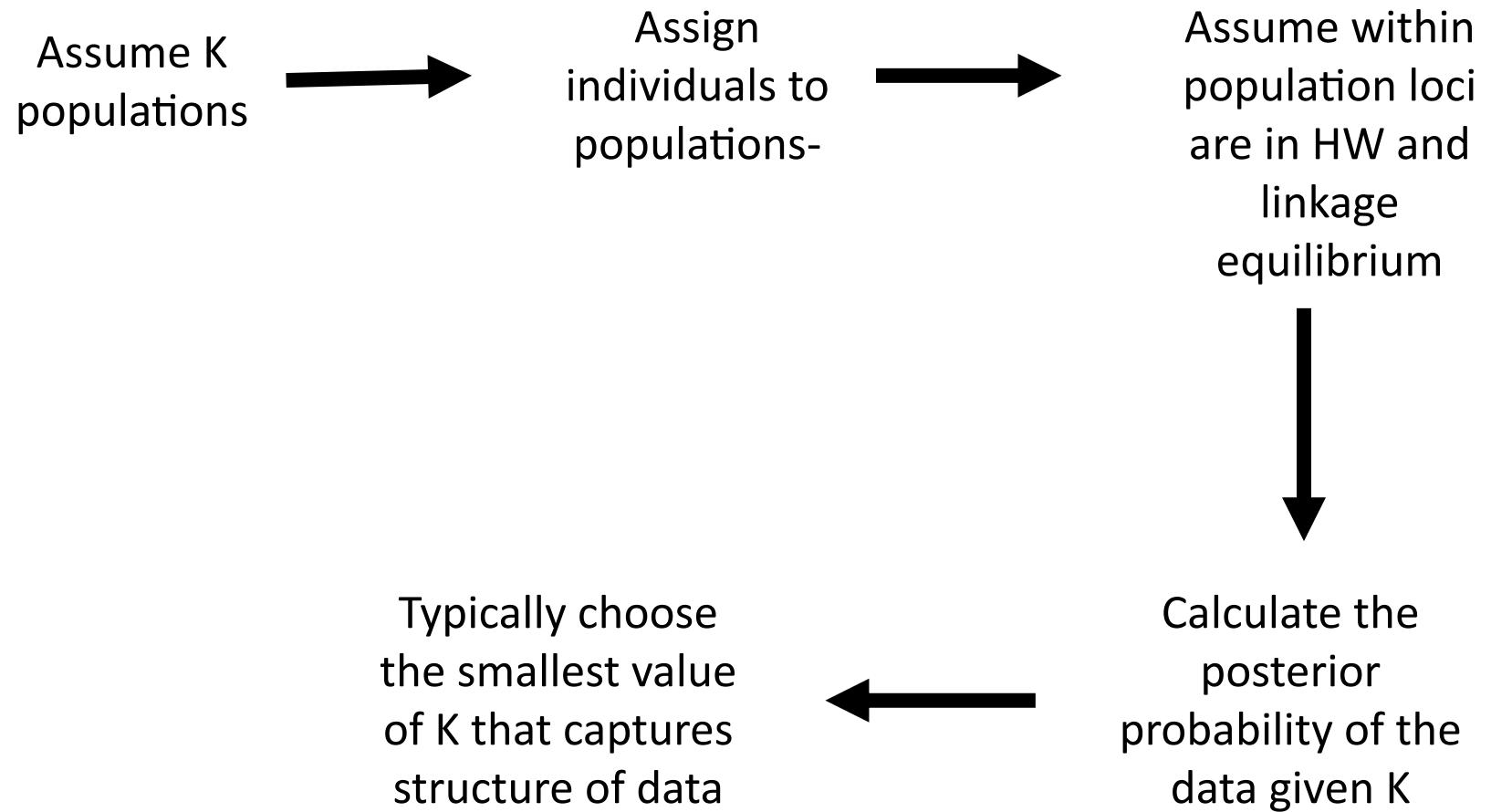
- Identifies the number of hypothetical ancestral populations.
- Introduce population structure
- Find groupings that are not in disequilibrium or out of HW
- Estimate the probability of the data given each model.
- Can assign individuals to populations

Genotyping-by-Sequencing of the regional Pacific abalone (*Haliotis discus*) genomes reveals population structures and patterns of gene flow

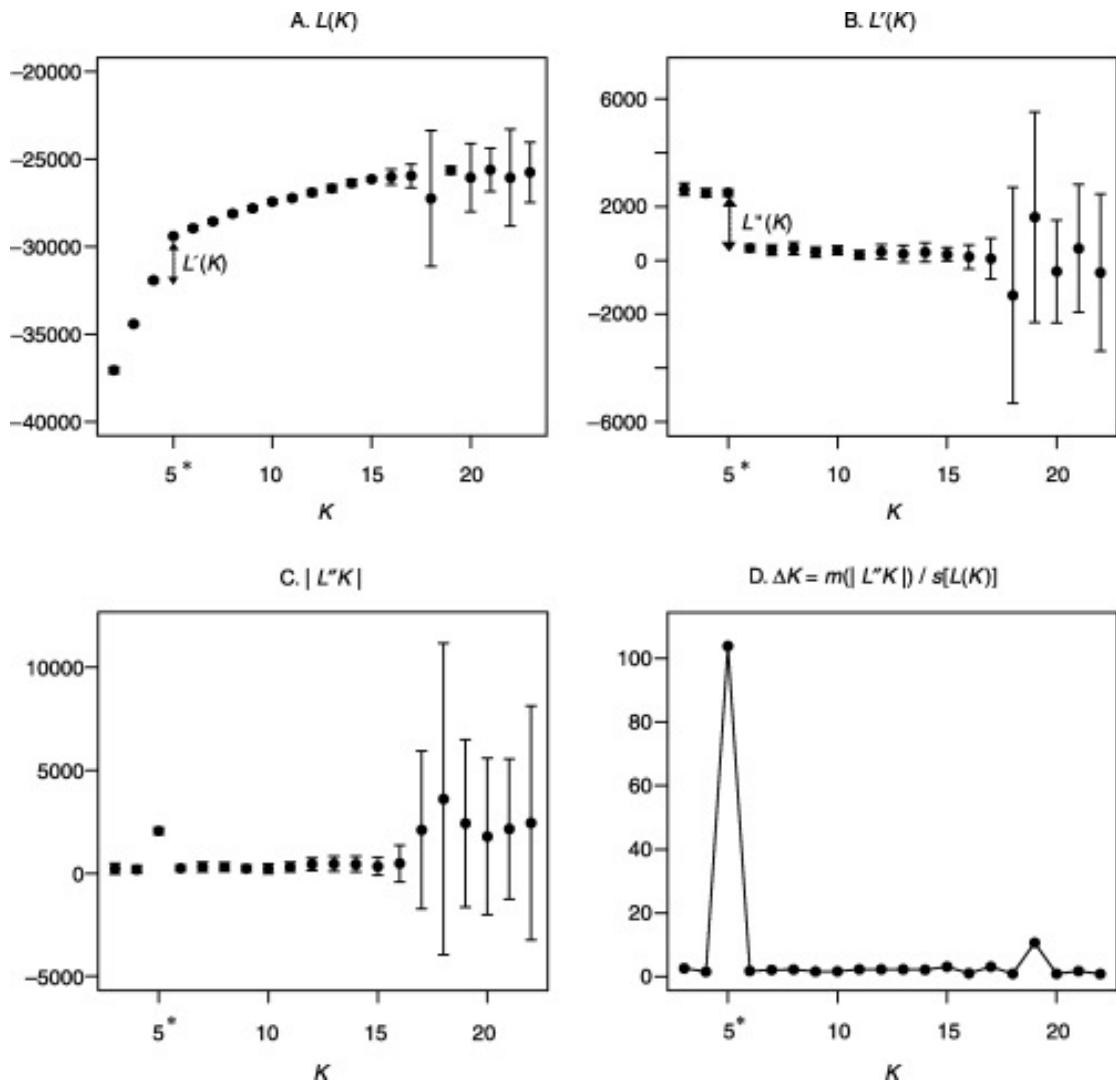
Bo-Hye Nam^{1‡}, Hyaekang Kim^{2‡}, Donghyeok Seol^{1,2,3}, Heebal Kim^{2,3}, Eun Soo Noh¹, Eun Mi Kim¹, Jae Koo Noh¹, Young-Ok Kim¹, Jung Youn Park¹, Woori Kwak^{1,3*}



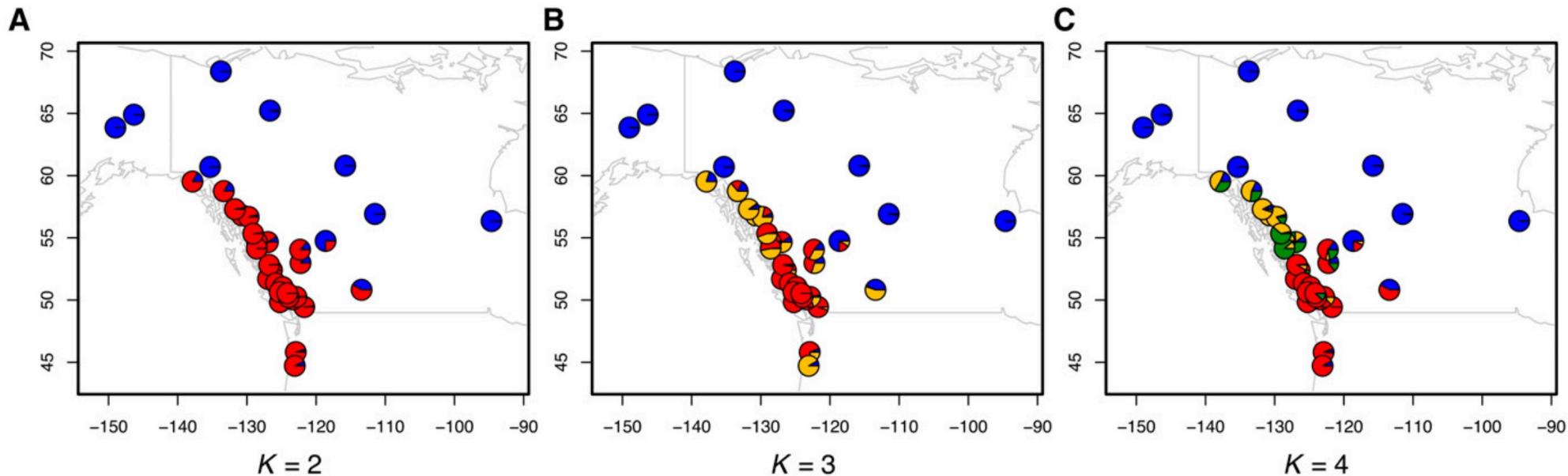
Structure



The Evanno method



- Problem with all of these is that populations aren't real
- Species are commonly continuously distributed



Infering Continuous and Discrete Population Genetic Structure Across Space

Gideon S. Bradburd,^{*1} Graham M. Coop,^{†,2} and Peter L. Ralph^{‡,2}

^{*}Ecology, Evolutionary Biology, and Behavior Graduate Group, Department of Integrative Biology, Michigan State University, East Lansing, Michigan 48824, [†]Center for Population Biology, Department of Evolution and Ecology, University of California, Davis, California 95616, and [‡]Institute of Ecology and Evolution, Departments of Mathematics and Biology, University of Oregon, Eugene, Oregon 97403

ORCID IDs: 0000-0001-8009-0154 (G.S.B.); 0000-0001-8431-0302 (G.M.C.); 0000-0002-9459-6866 (P.L.R.)

Paper for discussion

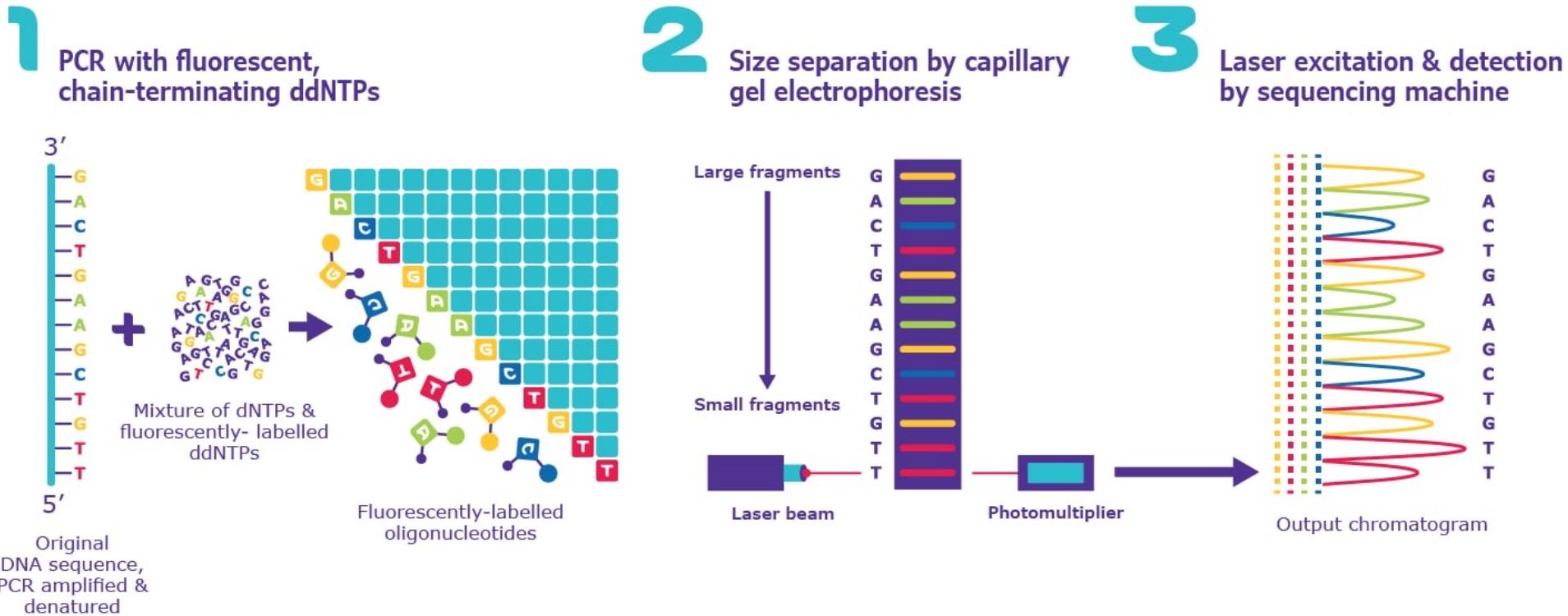
ORIGINAL ARTICLE

WILEY **MOLECULAR ECOLOGY**

Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*)

Amanda Xuereb¹  | Laura Benestan² | Éric Normandeau² | Rémi M. Daigle¹ |
Janelle M. R. Curtis³ | Louis Bernatchez² | Marie-Josée Fortin¹

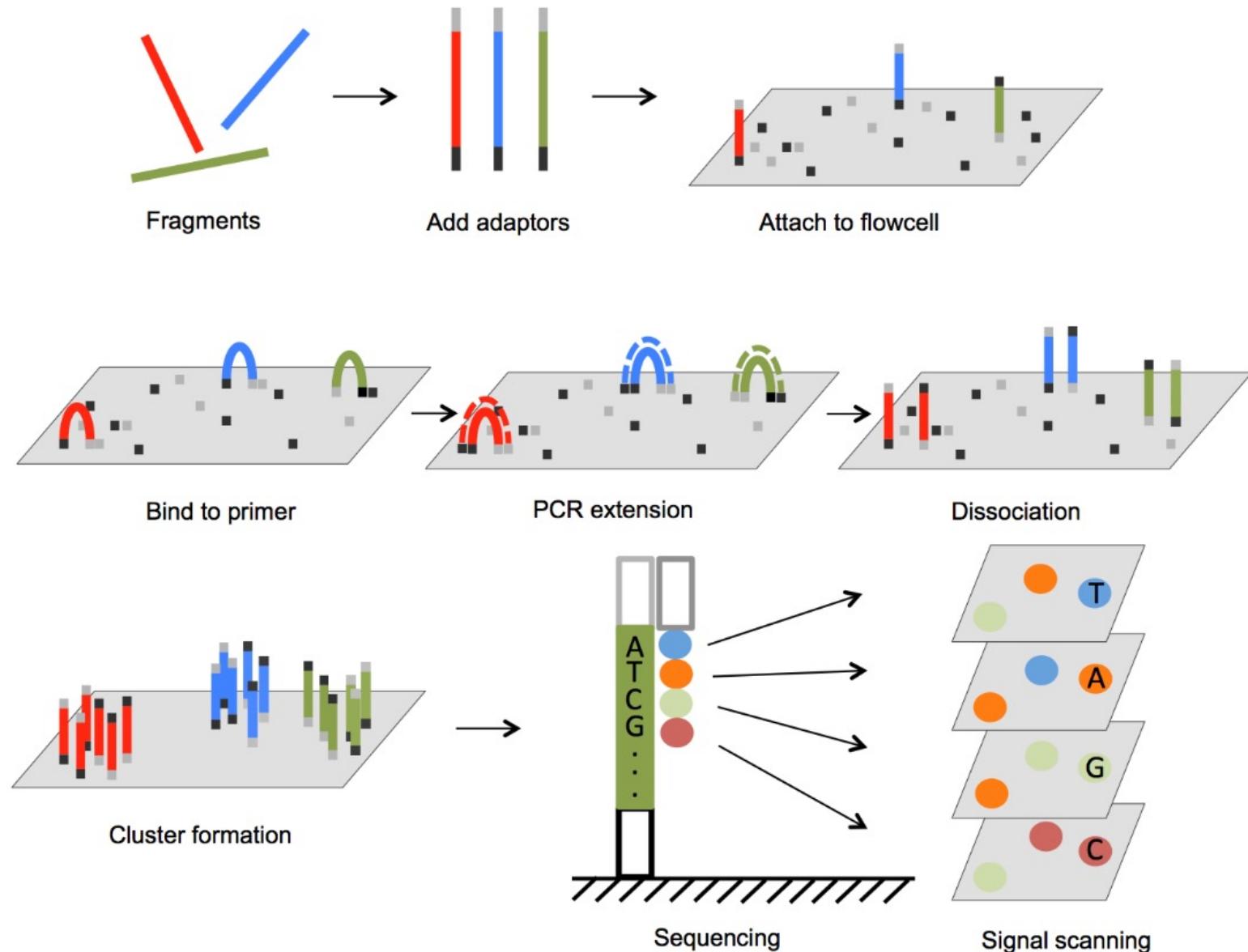
Sanger sequencing



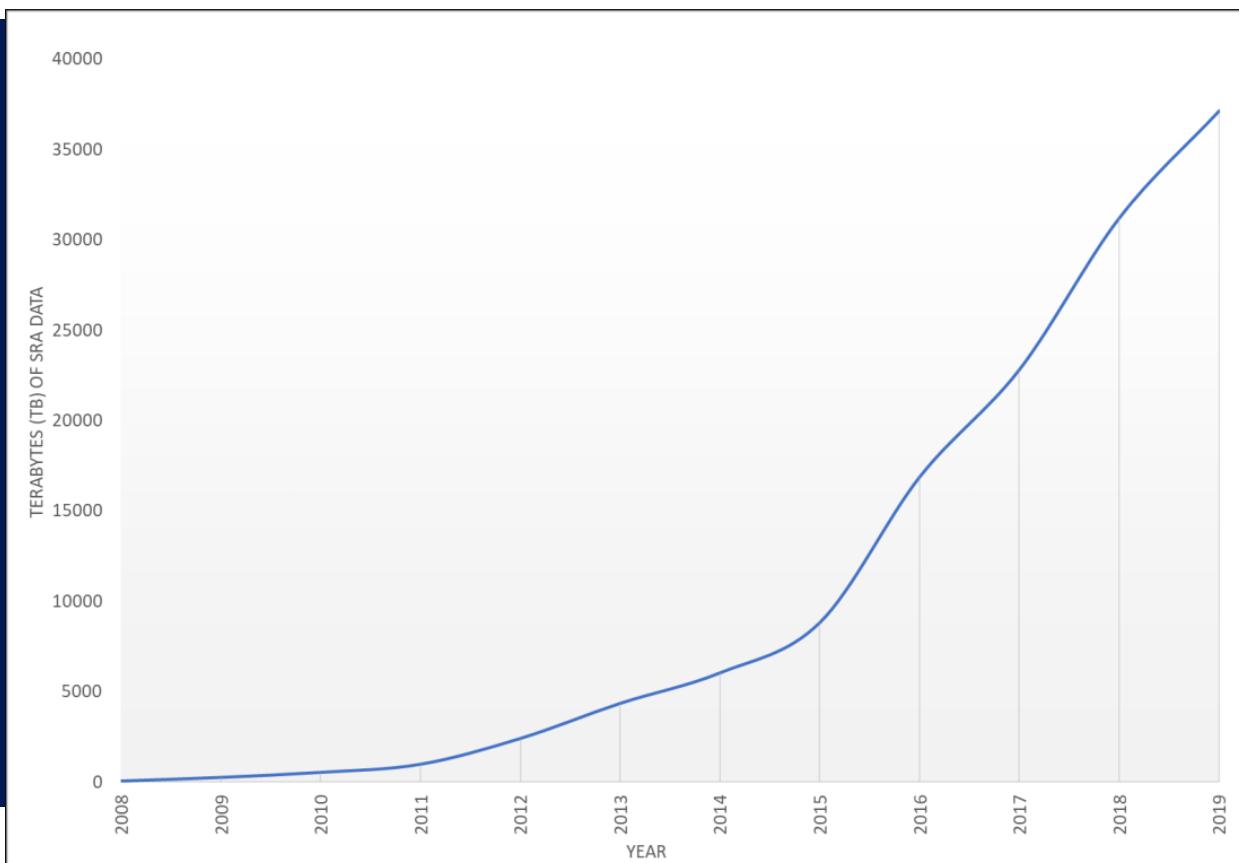
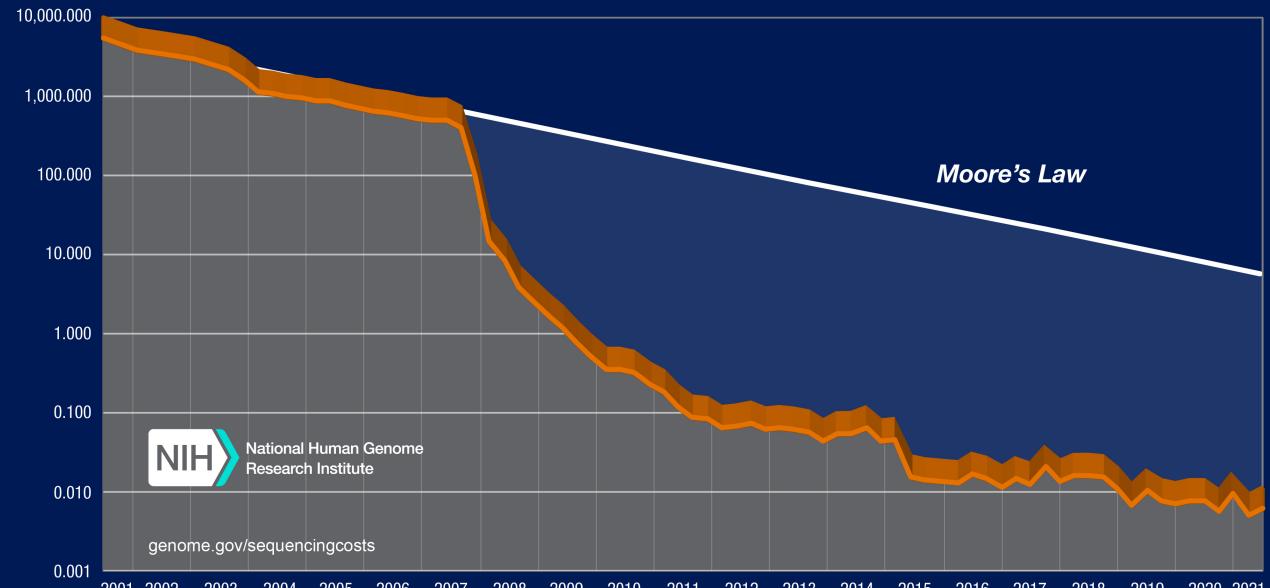
- A few hundred to ~1000 bp
- Low throughput
- Accurate

Illumina

- Short reads
 - 50-150 bp
- Very high throughput



Cost per Raw Megabase of DNA Sequence

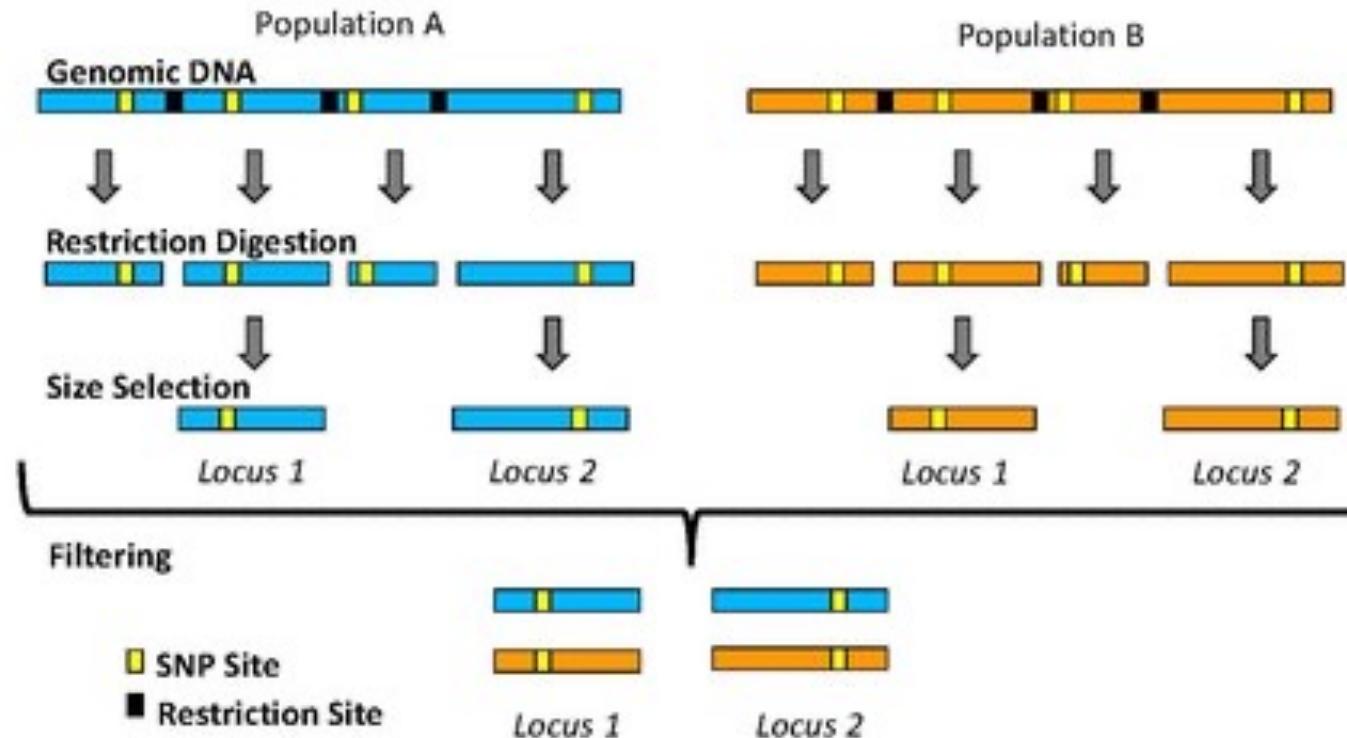


Data in NCBI

Reduced representation

- RADseq, GBS, sequence capture, ddRAD, 2b-RAD...
- Useful for population structure and similar
- But... misses lots of the genome

Restriction-site Associated DNA Sequencing (RADseq)



[https://en.wikipedia.org/
wiki/Restriction_site_ass
ociated_DNA_markers](https://en.wikipedia.org/wiki/Restriction_site_associated_DNA_markers)