Autopolyploid populations

Dataset

In this practical, you will be working with empirical datasets from the diploid-autotetraploid *Biscutella laevigata* species complex (Brassicaceae). Within each ploidy, we will focus on a pair of lowland vs alpine environments populations (i.e. four populations in total) genotyped with whole-genome sequencing.

Pop	Ploidy	Elevation_class	N samples
V2B	2	low	8
A2S	2	high	7
FUL	4	low	7
S3	4	high	8

General objectives of the practical

Autopolyploids properties require genetic data to be analysed with appropriate tools (Gallais 2003, Meirmans et al. 2018). Here, we will address three aspects of genetic analyses that are directly impacted by ploidy: inheritance, linkage disequilibrium and populations' genetic diversity and pairwise differentiation.

Tetrasomic inheritance

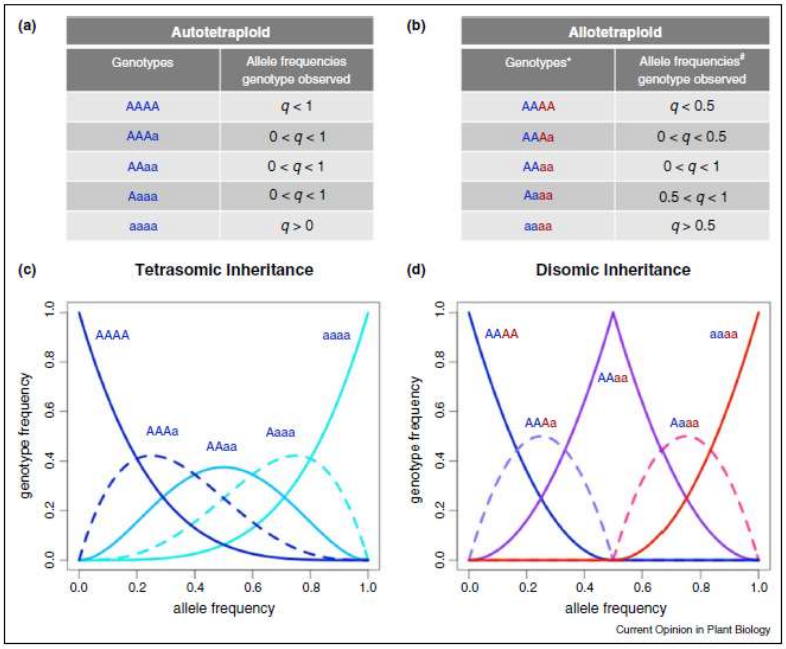
Polyploids have classically been classified based on their mode of origin (phylogenetic) and chromosome pairing behaviour (cytogenetic). Accordingly, autopolyploids are formed within a single population or species by whole-genome duplication and are expected to form (possibly deleterious) multivalent configurations during meiosis. Allopolyploids, on the other hand, are derived from interspecific hybrids and display bivalent pairing. These two categories are to be understood as endpoints of a continuum and polyploids formed along which species would be found (Stebbins 1971, Parisod et al. 2010).

Although chromosome pairing was long the only information available for assigning the "type" of polyploidy, it became clear that the key property distinguishing autopolyploids from allopolyploids is the inheritance mode (allelic segregation), which is largely determined by the structural differentiation of the different sets of chromosomes that pair at meiosis. In contrast to allotetraploids that mostly form bivalents according to the pairing of homeologous chromosomes and thus present disomic inheritance, autotetraploids can form either bivalents or multivalents through the random pairing of homologous chromosomes, supporting tetrasomic inheritance (e.g. bivalent pairing but tetrasomic inheritance in *Arabidopsis arenosa*; Hollister et al., 2012). Note that the mode of inheritance can vary along chromosomes according to their structure.

Assessing the mode of inheritance of tetraploid populations is also of particular interest if you want to carry out demographic inferences using coalescent models (one of the assumption being that chromosomes are exchangeable; Arnold et al., 2012). The Inheritance mode can be visually assessed through 1) the allele frequency spectrum, as the structure between duplicated chromosome sets due to pairing preference expectedly creates an enrichment of

alleles at 50% frequency relative to an allele frequency spectrum with unstructured chromosomes and 2) visualising whether observed genotypes frequencies according to allele frequencies follow expectations under tetrasomic vs disomic inheritance (Lloyd & Bomblies, 2016).

Lloyd and Bomblies, 2016



Tetrasomic and disomic inheritance patterns in polyploids. Autotetraploids (a) with tetrasomic inheritance retain random segregation of all four homologs. For biallelic loci this results in five possible genotypes. All genotypes can be observed whenever both alleles are segregating, regardless of allele frequency. In allotetraploids with disomic inheritance (b), the two sub-genomes (blue and red) segregate independently. While theoretically nine genotypes are possible, variants segregating in one sub-genome are generally fixed in the other subgenome, limiting the number of effective genotypes to five. The Hardy-Weinberg model extended for tetraploids with either tetrasomic (c) or disomic (d) inheritance patterns predicts the five genotype frequencies for any given allele frequency. With tetrasomic inheritance all five genotypes are observed when allele frequency = 0.5. With disomic inheritance an allele frequency of 0.5 represents fixed differences between the subgenomes and as a result all individuals have the duplex heterozygote genotype (AAaa). Thus, observed genotype frequencies of natural populations can be easily used to discriminate between the two modes of inheritance. In some cases, an allopolyploid can have segregation in both sub genomes, which is a more complicated situation not represented here, but which still generates patterns of genotype frequency distinct from those seen in autopolyploids

Linkage disequilibrium

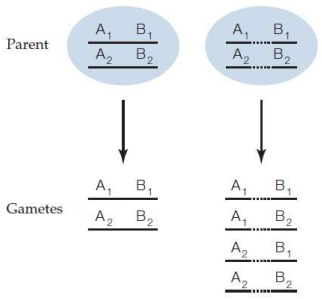


FIGURE 4.1 Linkage. The example on the left shows two closely linked loci such that only the two parental gametes are produced ($c = 0$), while that on the right shows two unlinked loci with all four gametes produced at equal frequency ($c = 0.5$).

Linkage disequilibrium (LD) is the non-random association of alleles in a population. While *linkage* describes the co-inheritance of sites in gametes because they are physically linked on chromosomes, loci in LD are not necessarily in linkage (e.g. selection on loci with epistasis) and inversely. Nevertheless, physical linkage is the major cause of LD, and many closely linked sites are in LD (Hahn 2019).

Genome wide LD reflects the population history, breeding system, and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates (Slatkin, 2008). LD is also commonly used to prune datasets for analyses requiring unlinked loci (e.g. in clustering analyses such as 'structure').

The estimation of LD in polyploids is limited. Gerard (2021) provides various methods to estimate LD in polyploids in the presence of genotype uncertainty by using genotype likelihoods. Haplotypic LD estimators assume autopolyploid individuals under Hardy-Weinberg equilibrium that exhibit polysomic inheritance. However, composite LD estimators are applicable to violations in both the random mating and the mode of inheritance assumptions.

F_{ST} scans

Genome scans can help teasing demographic and adaptive processes that shape patterns of genetic variation apart based on the rationale that selection acts locally, whereas demography acts globally across the genome (Hahn 2019). The most common approach consists of assessing genetic differentiation (i.e. F_{ST}-like metrics) across multiple loci between pairs of populations or among multiple populations and to identify loci presenting largely higher (directional selection) or lower ("balancing" selection) differentiation than the genome-wide loci (i.e. assumed as neutral).

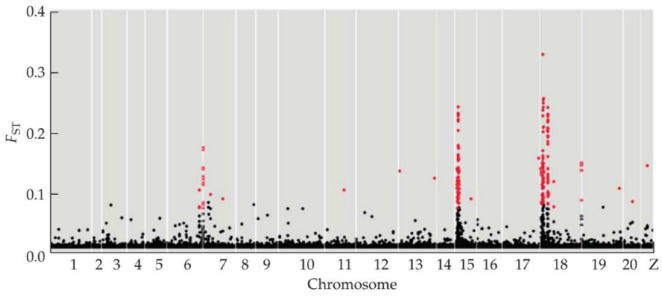


FIGURE 10.4 Genome-wide scan for selection using F_{ST}. Polymorphism data from two populations of the butterfly *Heliconius melpomene* were used to calculate F_{ST}. Significant outlier loci were identified using *BayeScan* (Foll and Gaggiotti 2008).

Dealing with biallelic SNPs under an island model, relative genetic differentiation is traditionally estimated as

$$F_{ST} = \frac{\pi_T - \pi_S}{\pi_T}$$

π_T being the nucleotide diversity, or heterozygosity, across all samples combined;

π_S being the same for each population.

In autotetraploids (with 4N chromosomes), the estimation of within population heterozygosity across loci is affected by their particular segregation under tetrasomic inheritance (e.g.

double reduction that increases the production of homozygotes as compared to expectations for diploids with 2N chromosomes). Accordingly, F_{ST} would appear lower in autoteraploid as compared to diploid populations with the exact same properties. Metrics independent of the ploidy level and double reduction (i.e. based on the average relatedness of individuals within and among populations; ρ) have been privileged (Ronfort et al. 1999).

The reliance of such metrics of genetic differentiation based on allele frequencies on variation within populations can lead to serious inflation in case of reduced variation such as expected at loci in low-recombining (e.g. pericentromeric) regions of chromosomes affected by background or linked selection. Accordingly, absolute measures of genetic differentiation or divergence (d_{XY}; i.e. excluding comparisons within populations) have been proposed. Following Cruickshank & Hahn (2014), the current practice consists of assessing and comparing F_{ST}-like and d_{XY} scans to identify candidate loci shaped by selection.

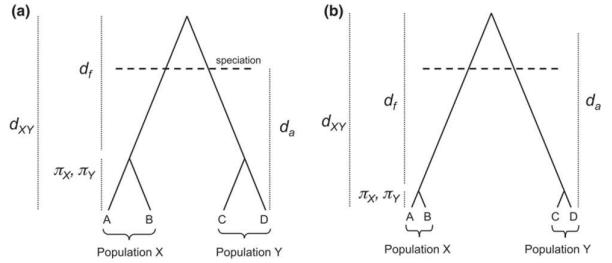


Fig. 1 Relative and absolute measures of divergence, with the effect of linked selection. Demonstrating the differences between π_X, π_Y, d_a, d_f and d_{XY}. Panels (a) and (b) both show example genealogies relating four sampled chromosomes (A, B, C and D) from two populations or species (X and Y). The statistics π_X and π_Y measure the average number of nucleotide differences between samples in population X and Y, respectively. d_{XY} measures the average number of nucleotide differences between each sample in population X and each sample in population Y, with no comparisons made within a population. d_a uses the average current levels of polymorphism as a measure of ancestral polymorphism, and subtracts this value from the total divergence (= $d_{XY} - [\pi_X + \pi_Y]/2$). d_f represents the total number of fixed differences between the two populations. For ease of comparison, the genealogies in the two panels have the same height, as do the genealogies for populations X and Y within each figure. The important distinction between panels a) and b) is that due to linked selection, there is a difference in π_X, π_Y, d_a and d_f between the two panels, but no difference in d_{XY}.

References

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157.

Hahn MW (2019) Molecular population genetics. Oxford University Press; Sinauer Associates, New York, Sunderland MA

Arnold, B., Bomblies, K., & Wakeley, J. (2012). Extending coalescent theory to autotetraploids. *Genetics*, *192*, 195–204.

Gallais A (2003) Quantitative genetics and breeding methods in autopolyploid plants. INRA éditions, Paris

Gerard, D. (2021). Pairwise linkage disequilibrium estimation for polyploids. *Molecular Ecology Resources*, *21*, 1230–1242.

Hollister, J. D., Arnold, B. J., Svedin, E., Xue, K. S., Dilkes, B. P., & Bomblies, K. (2012). Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLOS Genetics*, *8*, 1–10.

Lloyd, A., & Bomblies, K. (2016). Meiosis in autopolyploid and allopolyploid Arabidopsis. *Current Opinion in Plant Biology*, *30*, 116–122.

Meirmans PG, Liu S, & van Tienderen PH (2018) The analysis of polyploid genetic data. *Journal of Heredity*, *109*, 283–296.

Parisod, C., Holderegger, R., & Brochmann, C. (2010). Evolutionary consequences of autopolyploidy. *New Phytologist*, *186*, 5–17.

Ronfort J, Jenczewski E, Bataillon T, Rousset F (1998) Analysis of population structure in autotetraploid species. *Genetics*, *150*, 921–930.

Slatkin, M. (2008). Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, *9*, 477–485.

Stebbins GL. 1971. Chromosomal evolution in higher plants. London, UK: Edward Arnold.