

Selection efficacy differs between lifestages in the haploid-diploid *Marchantia polymorpha* subsp. *ruderalis*

Tianyu Shen¹, Philippe Gadiant², Justin Goodrich², and Hannes Becher^{3,*}

¹Morgan State University, Baltimore, USA

²Institute of Molecular Plant Sciences, Schools of Biological Sciences, University of Edinburgh, Max Born Crescent, Edinburgh EH9 3BF, UK

³The Roslin Institute, Royal (Dick) School of Veterinary Science, University of Edinburgh, Midlothian EH25 9RG, UK

*Correspondence to h.becher@ed.ac.uk

March 2024

Abstract

Ploidy has profound effects on evolution. Perhaps the most compelling effect is related to dominance: In cells with more than one genome copy, the effects of non-dominant deleterious variants are ‘masked’ to some extent, leading to a reduction in the efficacy of selection. This, in turn, leads to increased levels of nucleotide diversity, and it may also affect nucleotide substitution rates.

To test this predicted association between ploidy and the efficacy of selection, we studied genome scale per-gene patterns of genetic diversity and divergence in the genome of the haploid-diploid bryophyte *Marchantia*. We treated lifestage gene expression bias as a continuous covariate and accounted for genomic autocorrelation patterns using smoothing splines in a general additive regression model (GAM) framework.

Consistent with a lower efficacy of purifying selection, we found increased levels of sequence diversity, Watterson’s θ , and net divergence at non-degenerate sites in genes with diploid-biased gene expression. These genes also showed reduced levels of codon usage bias. In addition, we found chromosome 5 to be an outlier with overall decreased levels of diversity, the site-frequency spectrum skewed towards common alleles, and increased linkage disequilibrium.

In this study, we show the utility of generalized additive models in population genomics, and we present evidence for a ploidy associated difference in the efficacy of selection. We discuss parallels to the evolution of (diploid) sex chromosomes and why the patterns observed are unlikely to be mediated by masking.

1 Introduction

Ploidy has profound effects on the pace, course, and constraints on evolution in populations and species (Fox et al., 2020; Pandit et al., 2011; Schmickl and Yant, 2021). It varies at a fine scale across the tree of life, and many species comprise individuals of different ploidy levels (Kolář et al., 2017). 16 % of all flowering plants contain multiple ploidy levels (Rice et al., 2015). This includes crops like potato (De Jong and Tai, 1977) and sweet potato ([pereira_sweetpotato_2023](#))

as well as invasives/colonisers such as common reed (Pyšek et al., 2020), cordgrass (Fortune et al., 2008; Renny-Byfield et al., 2010), and monkeyflower (Salony et al., 2024). Ploidy can thus be seen as an individual-level property. In addition, ploidy commonly changes during a species' lifecycle as all sexual species, by definition, go through lifecycle stages of different ploidy levels. In plants, these lifestages are called 'gametophyte' and 'sporophyte', with the sporophyte having twice the ploidy level of the gametophyte. In the simplest case, the gametophyte is haploid and the sporophyte diploid.

When there are multiple genome copies, the concept of dominance becomes relevant. Any allele that is not fully dominant, i.e. with a dominance coefficient $h \leq 1$, should then be 'masked' to some extent if at least another 'wild-type' allele is present at the same locus. Increased ploidy should thus coincide with a reduced efficacy of selection and increased genetic load of deleterious mutations (Kondrashov and Crow, 1991). A reduced efficacy of selection should lead to higher levels of nucleotide diversity in any one population, due to an increased number of segregating slightly deleterious variants. A reduced selection efficacy can also affect non-synonymous divergence, d_A , for instance by reducing the number of advantageous substitutions relative to neutral ones. This is the theoretical explanation of faster-X effect (Charlesworth, Coyne, et al., 1987), where the X chromosome, whose loci are effectively haploid in males of many species, shows higher non-synonymous substitution rates than autosomes.

Carrying out within-genome comparisons is a powerful approach for studying the effects of ploidy on the efficacy of selection because it cuts out the potential for confounding with species or population idiosyncrasies. Much information has been gained from studies comparing genes on sex chromosomes to those located on autosomes (e.g. Hu et al., 2013; Langley et al., 2012; Thornton and Long, 2002; Zhou and Bachtrög, 2012) and by comparing X-chromosomal genes with differential expression sex bias (e.g. Campos, Johnston, et al., 2018; Rousselle et al., 2016). But sex chromosomal genes evolve under their own specific conditions and constraints and they only account for a limited fraction of the genome (Avery, 1984). A much larger proportion of the genome can be analysed in haploid-diploids. These are organisms that have separate haploid and diploid lifestages, both showing mitotic growth and developmental differentiation, accompanied by differential gene expression. The haploid-diploid setting allows one to test whether lifestage differences in gene expression, and thus ploidy, significantly affects nucleotide diversity and divergence. Accounting for a significant fraction of the tree of life, haploid-diploids cover a wide spectrum as to the relative importance and persistence of the two lifestages (Bourdareau et al., 2021; Coelho et al., 2007).

One possible issue with within-genome comparisons is that there tend to be autocorrelation patterns: Close-by genes tend to be similar in aspects such as their nucleotide diversity or the skew of their site-frequency spectrum. This observation has several potential causes which are not mutually exclusive, including selection at linked sites (Charlesworth, Morgan, et al., 1993; Maynard Smith and Haigh, 1974), which is mediated by the local rate of recombination and the strength of which may thus change over time (Booker et al., 2022). Further causes include genome-wide variation in mutation rates (Barroso and Dutheil, 2023; Castellano et al., 2020; Hodgkinson et al., 2009; López-Cortegano et al., 2021; Smith et al., 2018), and the mutagenic potential of recombination (Halldorsson et al., 2019; Hinch et al., 2023). Genomic autocorrelation must be taken into account if reliable conclusions are to be drawn from within-genome comparisons.

To investigate whether lower ploidy is associated with a higher efficacy of selection (in analogy the the faster-X effect), we chose to study the haploid-diploid liverwort *Marchantia polymorpha* subsp. *ruderalis*. This taxon has long served as a scientific model organism (Bowman, 2016). It is commonly known, and referred to below, as *Marchantia*. *Marchantia* is a member of the bryophytes, the non-vascular land plants. Typical for a bryophyte, *Marchantia*'s haploid gametophyte lifestage dominates the lifecycle while its diploid sporophyte stage is always dependent on the gametophyte, persisting for a few weeks and eventually releasing large quantities of meiotic spores. See Bowman et al. (2022) for details on *Marchantia*'s lifecycle. *Marchantia* is a cosmopolitan ruderal taxon, which disperses via airborne meiotic spores and vegetative gametophytic propagules called 'gemmae'. Both are haploid. Surprisingly, and despite the effective ploidy differences experienced by genes with differential lifestage expression differences, previous genome scale studies in bryophytes did not find significant effects of lifestage expression differences on the efficacy of selection (Szövényi et al., 2013) or the rate of molecular evolution (Linde et al., 2021).

The purpose of this study is twofold: (1) we follow up on the question of whether there are ploidy effects on the efficacy of selection in bryophytes using novel data and methodology, and (2) we demonstrate the utility of the generalized additive model (GAM) in population genomics. To achieve this, we analysed genome scale patterns of genetic diversity in the bryophyte *Marchantia*. We characterised nucleotide diversity, linkage disequilibrium, codon usage bias, and the net divergence from another subspecies called *Marchantia polymorpha* subsp. *polymorpha*. We then characterised the distribution of per-gene lifestage expression differences, and assessed how it was associated with diversity and divergence patterns. We found that levels of nucleotide diversity and divergence at non-synonymous sites were significantly increased in genes whose expression was biased towards the diploid sporophyte lifestage. We discuss possible reasons for this observation and parallels to the evolution in diploid sex chromosome systems.

2 Results

2.1 Data overview

We generated 38.3 Gbp of short-read whole-genome resequencing data (2.0-3.4 Gbp per individual) from 13 individuals of *Marchantia polymorpha* subsp. *ruderalis* (*Marchantia*). These individuals had been collected in southern Edinburgh, UK, with the sampling area spanning approximately 10 km, see Figure 1. We also reanalysed 115 Gbp of data generated by Sandler et al. (2023). These were from Canadian samples of both *Marchantia* (15 individuals) and the subspecies *M. p.* subsp. *polymorpha* (8 individuals). The Canadian range spanned a much larger area of several 100 km across. The proportion of successfully mapped reads was considerably higher for our data, as expected for DNA extracted from axenically cultivated samples. The mapping rates of our data ranged from 71 to 96 % with a mean of 93 %, resulting in comparable overall mapping depths for both datasets. After k-mer based masking of repeat regions in the reference assembly, 143 Mbp of autosomal regions (68.6 %) remained for us to carry out variant calling. Screening unmasked coding regions, we identified 3 442 987 fully degenerate (4-fold) and 12 471 802 non-degenerate (0-fold) nucleotide sites.

Despite *Marchantia*'s capacity for clonal reproduction and the modest size of our sam-

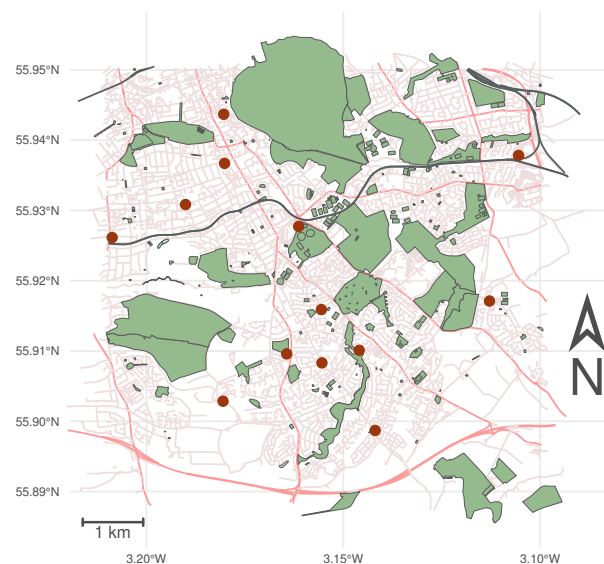


Figure 1: Sampling sites of *Marchantia* (maroon dots) in southern Edinburgh, UK.

plung area, we found that all individuals were genetically distinct. Pairwise per-nucleotide differences at 4-fold sites ranged from 0.00176 to 0.00255, suggesting an appreciable degree of sexual reproduction. This is in line with the frequent observation of *Marchantia*'s iconic reproductive structures, the gametangiophores.

2.2 Marked differences in chromosomal diversity patterns

Summary statistics of genetic diversity (Table 1) were in line with the numbers reported by Sandler et al. (2023). However, we computed these on a per-chromosome basis, which revealed considerable inter-chromosomal variation. Chromosome 5 turned out to be an outlier with reduced levels of both genetic diversity, measured as nucleotide diversity, π , and the standardised number of segregating sites, θ_w . Chromosome 5 also showed an increased skew of the site-frequency spectrum towards common alleles indicated by reduced values of $\Delta\theta_w = 1 - \pi/\theta_w$ (Becher et al., 2020), see Figure 2A-C. The same chromosome also showed increased levels of linkage disequilibrium, see Figure 2D.

There was no significant association between chromosome length and any measure of genetic diversity.

2.3 Lifestage expression differences

We chose to measure the per-gene level of lifestage expression bias as the proportion of haploid expression (PHE). This statistic indicates, per gene, what proportion of the overall expression came from haploid tissues. PHE ranges, per definition, from 0 (exclusively sporophytic/diploid) to 1 (exclusively gametophytic/haploid). We found the distribution of the PHE values was shifted towards expression in the gametophyte, which is the predominant lifestage, see Figure 3. A majority of 63% of genes showed haploid biased expression, i.e. with $PHE > 0.5$. The top 10% PHE genes were expressed at least 16.5 times more in haploid tissues, whereas the bottom 10% PHE genes were expressed only at least 3 times more in diploid tissues (arrows in Figure 3). There were 485 genes that were expressed only in haploid tissues, but not a single gene was expressed exclusively in diploid tissues.

Table 1: Per-chromosome summary statistics for *Marchantia* from Edinburgh. All statistics were computed per gene and then averaged per chromosome. Each is given for both fully degenerate (subscript 4) and non-degenerate sites (subscript 0), with three significant digits. π – average number of pairwise nucleotide differences; θ_w – standardised number of segregating sites; $\Delta\theta = 1 - \pi/\theta_w$ – a measure of the skew of the site-frequency spectrum; d_A – net divergence from subspecies to *M. p. polymorpha*, n – number of genes considered.

Chr	π_0	π_4	θ_{w0}	θ_{w4}	$\Delta\theta_{w0}$	$\Delta\theta_{w4}$	d_{A0}	d_{A4}	n
1	0.000711	0.00242	0.000731	0.00248	0.0834	0.0176	0.00518	0.0218	1858
2	0.00102	0.00269	0.00101	0.00279	0.0694	0.0618	0.00635	0.0236	1547
3	0.0011	0.00315	0.00109	0.00316	0.0742	0.0262	0.00546	0.0231	1515
4	0.000939	0.00254	0.000976	0.00261	0.109	0.0569	0.00651	0.0234	1427
5	0.000814	0.00215	0.000795	0.00201	-0.0125	-0.0507	0.00703	0.0258	1396
6	0.00143	0.00368	0.00139	0.00368	0.0543	0.0462	0.00547	0.0188	1290
7	0.000909	0.00234	0.000925	0.00239	0.0926	0.0452	0.00659	0.0247	1172
8	0.000932	0.00266	0.000954	0.00264	0.0768	0.0434	0.00558	0.0217	1075
all	0.000971	0.0027	0.000973	0.00272	0.0694	0.032	0.00599	0.0229	11280

The genomic distribution of PHE was non-random with significant differences between chromosomes (ANOVA, $F_{14067,7} = 8.87$, $p = 6.08 \times 10^{-11}$). While statistically significant, this only accounted for a minute proportion of the variance in PHE. The range of per-chromosome mean PHE values was 0.012, and the adjusted r^2 values were 0.004 when accounting for chromosome effects. It increased to 0.01 when additionally fitting splines to allow for variation in PHE along each chromosome.

2.4 Associations between measures of genetic diversity and lifestage expression bias

We then tested whether patterns of genetic diversity and divergence covaried with lifestage expression bias measured as PHE by using generalized additive regression models (GAMs; Wood, 2017). GAMs allow one to combine the usual linear combination of predictor variables used in other regression models with smoothing splines. This is extremely useful, for instance to take account of autocorrelation patterns, which would otherwise violate the regression assumption of independent residuals. In our models, we included covariates of genetic diversity identified in previous studies: overall expression level, tissue specificity, local gene density (Slotte et al., 2011; Wright et al., 2004), number of exons, CDS length, and GC content (Larracuenta et al., 2008; Yang and Gaut, 2011). In addition, we included smoothing splines (Wood, 2003) to take account of any autocorrelation of the respective regression outcome variable along the chromosomes.

Below, we report the implied differences between hypothetical haploid and diploid-specific genes which are identical for all other covariates. This is one way of interpreting the coefficients estimates obtained from the regression models. It is important to note that these values differ from the observed values of actual haploid or diploid-specific genes, of which there are few, and which differ not only in their degree of lifestage expression bias but also with respect to all other covariates. In the remainder of the results section, we will focus on the Edinburgh dataset and mention the results obtained from the Canadian dataset for comparison.

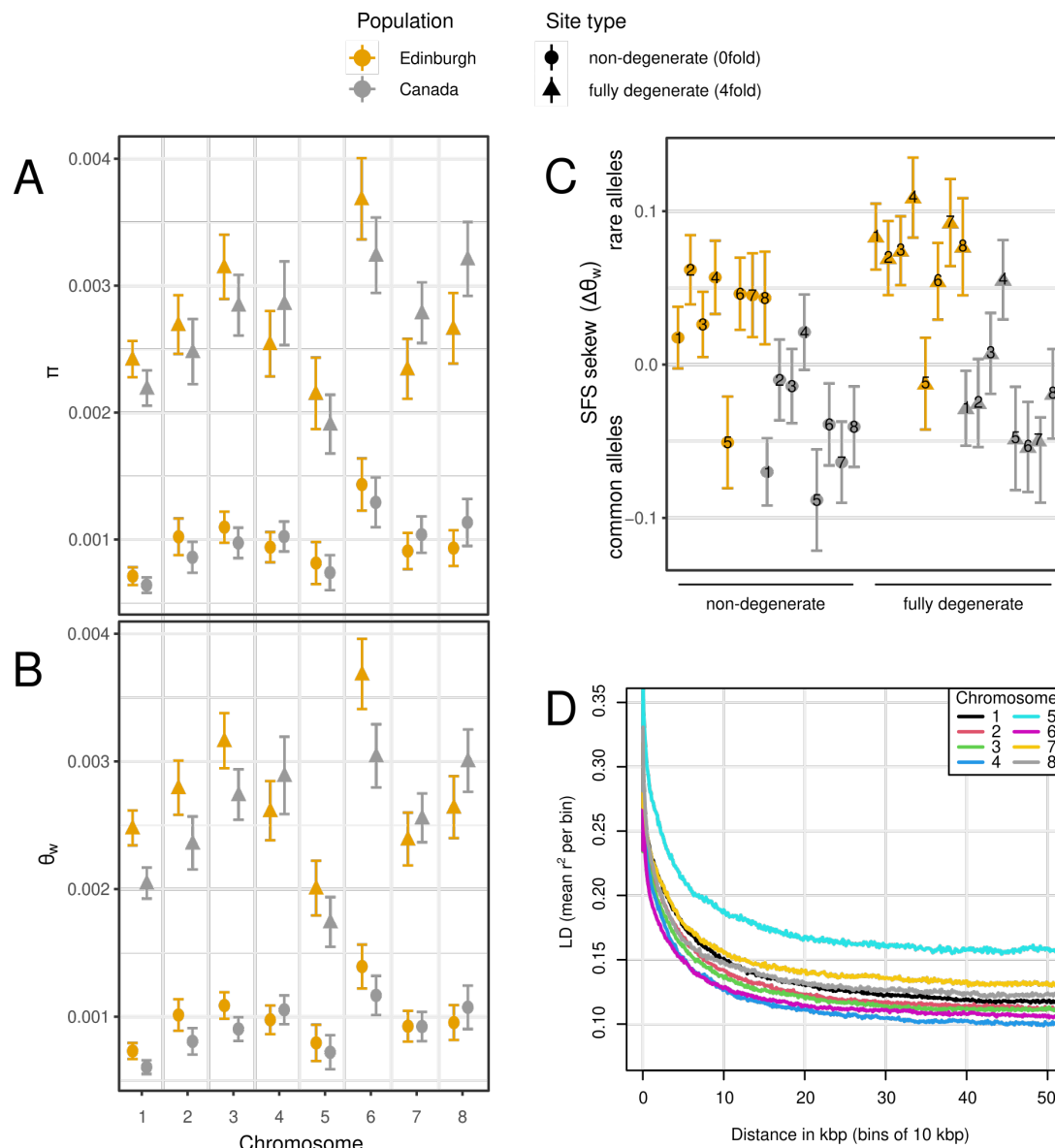


Figure 2: Summary statistics computed per gene, aggregated per chromosome of *Marchantia* from Edinburgh (mustard) and Canada (grey). Circles indicate the values for non-degenerate, and triangles the values for fully degenerate sites. The error bars indicate 95%-confidence intervals. Panel A shows nucleotide diversity, panel B shows Watterson's θ (the normalised number of segregating sites). A and B share the same x-axis, and the statistics are given per nucleotide. Panel C shows the skew of the site-frequency spectrum as $\Delta\theta_w$ with higher values indicating an excess of rare alleles. Panel D shows the decay of the linkage disequilibrium measured as the squared correlation coefficient between allele biallelic sites.

2.4.1 Genetic diversity is reduced in gametophyte-biased genes

We found that the pairwise nucleotide diversity at non-degenerate sites, π_0 , significantly covaried with lifestage expression bias ($p = 3.61 \times 10^{-5}$), with higher levels of gametophyte expression bias associated with lower values of diversity. The coefficient estimate of -0.0907 on log10 scale implies that, all else being equal, the expected level of the nucleotide diversity at non-degenerate sites in a gametophyte specific gene is only 81.1 % of that seen in a sporophyte specific gene (95 % confidence interval: 73.5 %; 89.6 %).

All other covariates, except gene density, also had significant associations with π_0 , see Figure 4 (top row, top-left triangles). Tissue specificity of expression, τ , was associated with in-

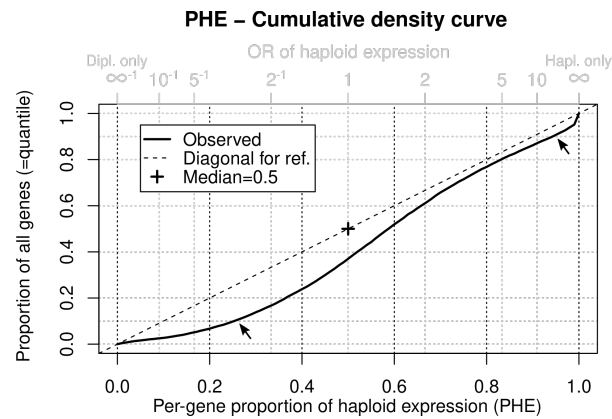


Figure 3: The distribution of per-gene lifestage expression bias, measured as the proportion of haploid expression (PHE). The fat black line shows the cumulative density curve of PHE, ranging from 0 to 1 on the black primary x-axis. The y-axis shows the PHE quantiles. A distribution with a median of PHE = 1/2 (odds ratio=1) would pass through the cross. The dashed diagonal line is given for reference. The arrows indicate the 10th and 90th percentiles, highlighting the distribution's asymmetry. The grey secondary x-axis applies to the same data, showing odds ratios instead of proportions.

creased diversity, whereas overall expression, number of exons, CDS length, and GC content all were associated with reduced diversity. In analogous models, nucleotide diversity at fully degenerate sites, π_4 , was not significantly associated with PHE, but all other covariates were. We observed qualitatively similar patterns when analysing the Canadian dataset, where π_0 was significantly associated with lifestage expression bias ($p = 0.00218$).

An alternative measure of genetic diversity, θ_w , showed similar behaviour to π . Increased gametophyte expression bias was significantly associated with lower values of θ_{w0} ($p = 2.06 \times 10^{-5}$). The coefficient estimate of -0.08 implies that haploid-specific genes have only 83.4 % of the diversity observed in diploid-specific ones (95 % CI: 76.8 %; 90.7 %). All other covariates except for gene density were significantly associated, with all coefficient estimates showing the same sign as for π . Expression bias was not significantly associated with the standardised number of segregating fully degenerate sites, θ_{w4} , but all other covariates were. The Canadian dataset showed analogous significance levels.

2.4.2 Codon usage is more optimal in haploid biased genes

The per-gene frequency of the optimal codon (Fop) was significantly increased in haploid biased genes ($p = 1.28 \times 10^{-20}$). The implied difference in Fop between haploid and diploid specific genes was 0.0241.

2.4.3 The site-frequency spectrum is not skewed by lifestage expression bias

We did not find significant associations between lifestage expression bias and the skew of the site frequency spectrum, measured as $\Delta\theta_w$, neither for fully degenerate nor non-degenerate sites. This was so for both the Edinburgh and Canadian sample. Overall, the site-frequency spectra of the Edinburgh samples tended to be skewed towards rare alleles, showing positive $\Delta\theta_w$, whereas the Canadian sample tended to show skew towards common alleles, negative $\Delta\theta_w$, see Figure 2C.

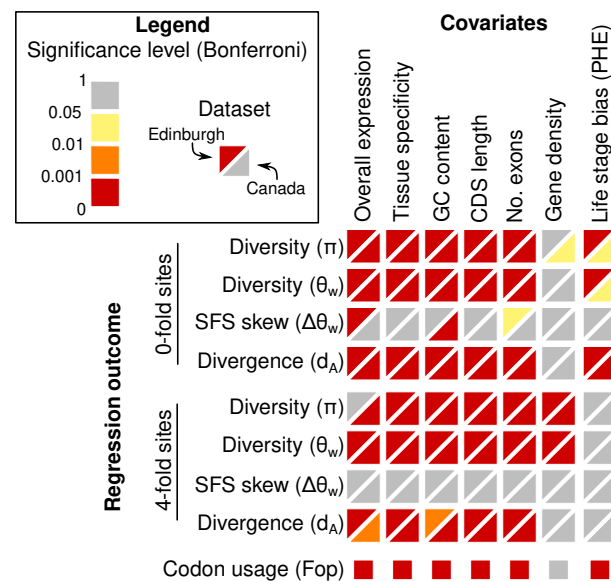


Figure 4: Significance levels of covariates regressed on summary statistics of genetic diversity and divergence for two Marchantia datasets using GAMs. Top-left triangles indicate the Edinburgh and bottom-right the Canadian dataset. Colours represent significance levels with thresholds indicated in the legend. The figure summarises the results of 17 regression models, and the p-values were corrected accordingly (Bonferroni). The significance level of the effect of life stage expression bias, the assessment of which was a key objective of this study, is shown on the very right. Fop (bottom row), the frequency of the optimal codon, which is a measure of codon usage bias, was estimated for the reference assembly. There are thus no separate results for Edinburgh and Canada.

2.4.4 Divergence at non-synonymous sites is reduced in gametophyte-biased genes

Using population data, we were able to estimate the net divergence, d_A , between *Marchantia* and *M. p. subsp. polymorpha*. The net divergence at non-degenerate sites was significantly reduced in genes with higher lifestage expression bias. The coefficient estimate of -0.130 on $\log_{10}$ scale implies a reduction of the net divergence at non-degenerate sites in haploid-specific genes to 74.2% of the level of diploid-specific genes (95% CI: 67.7%; 81.3%). All other covariates were significant, see Figure 4, with the same direction of effects as we found for π . Net divergence at fully degenerate sites did not covary with lifestage expression bias while all other associations remained significant. The picture for the Canadian samples was analogous.

3 Discussion

In this study, we characterised patterns of genetic diversity in a sample of *Marchantia* from Edinburgh, UK, to test whether there is an association between genes' lifestage expression bias and their patterns of genetic diversity and divergence. We also reanalysed a sample of *Marchantia* from Canada (Sandler et al., 2023). We found that genes with expression bias towards the haploid gametophyte lifestage showed, at non-degenerate sites, significantly reduced levels of genetic diversity and sequence divergence. Below, we first point out methodological differences from previous studies. We then discuss our findings in light of the expected efficacy of selection and ploidy.

3.1 Methodological advances

A higher efficacy of purifying selection at the haploid gametophyte lifestage has been reported in several systems (Arunkumar et al., 2013; Gossmann et al., 2014; Lipinska et al., 2019; Tanja Slotte et al., 2010). A higher efficacy of purifying selection on the X chromosome, which is haploid in males, is one explanation for the commonly observed faster X effect, which we mentioned in the introduction. In closer analogy with the haploid-diploid setting, the relative rate of adaptive substitutions, ω_{α} , on the X chromosome is highest in male-biased genes in *Drosophila*, which are effectively haploid when expressed (Campos, Johnston, et al., 2018). However, two previous studies on bryophytes paint a different picture. Linde et al. (2021) concluded that there was no evidence for lifestage differences in selection pressure as there was no increased d_N/d_S in the haploid lifestage in *Marchantia*. Studying the moss *Funaria hygrometrica*, Szövényi et al. (2013) found a complex relationship between d_N/d_S and lifestage expression bias – genes specific to either the sporophyte or gametophyte showed lower d_N/d_S values than unspecific genes. The authors explained this pattern by also taking into account tissue specificity, which tends to be lower in genes with lifestage specific expression.

Our study here differs in three key aspects. First, we used a continuous measure of lifestage expression bias, which we call the proportion of haploid expression. We chose this approach because the alternative – classifying genes into expression bins – always comes with a degree of arbitrariness and because classification does not allow for measurement error (noise). The more tissue samples are analysed in an experiment, the lower becomes the probability that a gene will be counted as lifestage specific. In addition, if there are bins for lifestage specific expression, the probability of a gene’s assignment to these bins will also depend on the relative number of tissues sampled for each lifestage. In the expression dataset that we used (Marpolbase Expression, <https://mbex.marchantia.info/>), there were far more gametophyte samples than sporophytic ones, making it more likely for low-expressed genes to appear gametophyte-specific by random chance. Indeed, we observed hundreds of gametophyte-specific genes but none that was sporophyte specific. This pattern is of course plausible given the biology of *Marchantia*, where the gametophyte lifestage is small and has a short duration. But the pattern is likely exacerbated by the tissue sampling available in the expression database. By treating expression bias as continuous, we treat it as data, which comes with uncertainty. Any gene with observed lifestage specific expression is not categorically different from any other genes. Lifestage specific genes just sit at the extremes of the continuous PHE scale, which ranges from 0 or 1.

Second, we generated population data, which allowed us to analyse patterns of nucleotide diversity within *Marchantia*. Population data are also required to estimate d_A , the net divergence (Nei, 1987; Nei and W.-H. Li, 1979). In the absence of population data, it is possible to estimate the pairwise nucleotide divergence, d_{xy} , from a pair of genome or transcriptome sequences. But the value of d_{xy} is affected by both the divergence between and diversity within the populations compared, even when comparing only two sequences. The distinction between d_{xy} and d_A is thus more important when comparing recently-diverged taxa where much of the expectedly low d_{xy} estimate is due to within-population (possibly ancestral) diversity.

Third and finally, we took account of ‘spatial’ autocorrelation along chromosomes by fitting smoothing splines using general additive models (GAMs). Ignoring autocorrelation leads to non-independent residuals in regression analyses, which can compromise significance tests

and lead to spurious conclusions. Early studies with small numbers of loci could safely ignore the possibility of autocorrelation (Begun and Aquadro, 1992; Payseur and Nachman, 2002) without detrimental effect. More recent genome-wide studies commonly use window-based approaches (Slotte et al., 2011), take the ratio of the same statistic at non-degenerate and fully degenerate sites such as π_0/π_4 , or explicitly take into account likely causes of autocorrelation (Becher et al., 2020; Campos, Halligan, et al., 2014). There are several advantages of fitting smoothing splines within the GAM framework. Being a non-parametric approach, it is not necessary to make assumptions as to the cause of autocorrelation. Also, because GAMs are ‘fully-fledged’ regression models, they are more flexible than rank regression. GAMs allow for more complex hypothesis testing and the inclusion of otherwise possibly confounding covariates. Using spline-based smoothing on chromosomal positions is arguably less arbitrary than selecting a window size for window-based analyses. Accounting for autocorrelation by smoothing, in turn, allows for the separate analysis of statistics computed on non-degenerate and fully degenerate sites rather than their ratios.

3.2 The efficacy of selection

It may appear plain that the efficacy of selection should differ between genes with different degrees of life cycle expression bias. The effect of masking in diploids or ‘unmasking’ in haploids (Beaudry et al., 2020) and the associated higher efficacy of selection in haploids is an extremely compelling argument in favour of such differences. There is however another mechanism potentially affecting the efficacy of selection in the opposite way. Because there is no additional genome where expression might happen at the same locus, haploids may show a higher degree of variation in gene expression levels (‘expression noise’) than organisms of higher ploidy (Cook et al., 1998). Wang and J. Zhang (2011) have compared the consequences of this phenomenon, which has been observed in yeast, to a depression of the effective population size in haploids relative to diploids. Increased expression noise in haploids would reduce the efficacy of selection. Expression noise might perhaps be mitigated by endopolyploidy, a condition where multiple genome copies are present in somatic cells. Endopolyploidy is common in plants and mosses in particular (Bainard, Newmaster, and Budke, 2020) but not known in the liverworts (Bainard and Newmaster, 2010) and is thus unlikely to matter for *Marchantia*. However, while important in yeast, is it not clear how relevant expression noise is in multicellular organisms Wang and J. Zhang (2011). For instance, gene expression from the functionally haploid mammal X chromosomes does not seem to be more noisy than from autosomes (Yin et al., 2009).

Because directional selection reduces the frequency of deleterious alleles, an increased efficacy of selection should lead to reduced nucleotide diversity. We observed the expected pattern at non-degenerate sites as π_0 was significantly reduced in genes with haploid-biased expression. Selection effects at fully degenerate sites are generally much weaker, leading to overall higher values of π_4 than π_0 , a pattern observed by Sandler et al. (2023), which we confirmed here. Fully degenerate sites may be affected by weak direct selection or by selection acting on linked nearby sites, the strength of which depends on the local, and possibly historic, recombination rate (Booker et al., 2022; Campos, Halligan, et al., 2014). Both effects should be affected by any difference in the efficacy of selection and there could in principle be a depression of π_4 at haploid-biased genes. We did not observe such a pattern. We did however find other

evidence for a difference in the efficacy of direct selection on degenerate sites. Codon usage differed, with the frequency of the optimal codon significantly higher at haploid-biased genes. This is in analogy to findings in *Drosophila*, where the effective codon usage bias is larger on the X than on the autosomes (Campos, Zeng, et al., 2013). Overall, these observations are evidence for a higher efficacy of selection at the gametophyte lifestage in *Marchantia*.

The efficacy of selection may also affect nucleotide substitution rates. But this effect is expected to be complex because nucleotide divergence may be due to adaptive, neutral, or even (slightly) deleterious substitutions. Depending on the distribution of fitness effects of mutations, a higher efficacy of selection might simultaneously reduce the number of slightly deleterious substitutions and increase the number of adaptive substitutions that overcome the effect of random genetic drift. For this reason d_N/d_S , sometimes called the ‘rate of evolution’, has been observed to be reduced in species with an increased efficacy of selection (Galtier, 2016).

We observed a significant decrease of non-synonymous substitutions between *Marchantia* and *M. polymorpha* subsp. *polymorpha* in genes with haploid expression bias. Considering that the proportion of adaptive mutations among all non-synonymous substitutions, α , was estimated to be small and not significantly different from zero (Sandler et al., 2023), it would seem plausible to attribute this pattern to a decreased number of slightly deleterious substitutions in genes with gametophyte-biased expression. This is analogous to an observation made in some butterflies. In butterflies, females are the heterogametic sex, carrying only one Z chromosome. Z-chromosomal genes with female-biased expression are thus functionally haploid just like autosomal genes with strong gametophyte expression bias in *Marchantia*. Rousselle et al. (2016) found that the rate of non-adaptive substitutions in Z-chromosomal genes with female expression bias was reduced compared to male biased ones, leading to an overall reduction of the rate of non-synonymous substitutions in haploid biased genes as we observed in *Marchantia*. While we found evidence for an increase in the efficacy of selection in haploid biased genes, this is not entirely analogous to the faster-X effect.

While the patterns we report suggest an increased efficacy of selection at the haploid gametophyte lifestage, it is important to note that our study cannot assess whether there is a causal relationship between ploidy and the efficacy of selection.

3.3 Within genome comparisons and the effects of ploidy

One rationale for studying the effect of lifestage expression bias is to test the effect of ploidy while excluding potentially confounding effects stemming from species or population idiosyncrasies, as we mentioned in the introduction. In the same organism, all autosomes must go through the same life cycle stages, experiencing the same changes in ploidy and the same levels of genetic drift. Any systematic differences between genes with expression bias towards one stage would thus necessarily be correlated with ploidy.

Besides ploidy, the basis for proposing effects of unmasking and expression bias, what else differs between the lifestages? One difference is the relative length of, and the number of cell divisions in, each stage. If some gene’s expression were biased towards the shorter life cycle stage, that gene would experience less transcription coupled repair. Transcription coupled repair is a repair mechanism that removes DNA adducts, which would otherwise cause mutations during DNA replication (Hanawalt and Spivak, 2008; Nicholson et al., 2024).

As a consequence, genes that are mainly active in the shorter life cycle stage may experience higher mutation rates. In bryophytes, the shorter lifestage is the sporophyte. The effect of reduced transcription coupled repair should thus act to increase levels of nucleotide diversity in the diploid lifestage, similar to masking. An important difference from masking is that lack of transcription coupled repair should also increase levels of synonymous diversity, which we did not observe.

A second aspect differing between the lifestages may be their population sizes. In bryophytes, one haploid gametophyte can, and usually does, carry numerous sporophytes. The realised lifestage specific population size will depend on numerous factors such as availability of suitable gametes, nutritional status of the carrying gametophyte, and (in dioicous species) the sex ratio. If there is a systematic difference in lifestage specific population size, that should affect the efficacy of selection of genes with biased expression (Bessho and Otto, 2017). Again, this should also affect levels of nucleotide diversity of fully degenerate sites, which we did not find to differ.

Third, even if genes are expressed in and under selection at both lifestages, this selection may be of antagonistic nature. Such ploidy antagonistic selection, may also interact with sexually antagonistic selection, can lead to increased diversity in some genes without lifestage expression bias (Immler et al., 2012). This could in principle cause a non-monotonic association between lifestage expression bias and nucleotide diversity.

Fourth, in many haploid-diploid species there is the possibility for life cycle short cuts, with the extreme being the emergence of asexuality (Hoshino et al., 2024). Such short cuts are common in bryophytes where there is vegetative propagation in the gametophyte stage. If the balance between vegetative and sexual reproduction is under genetic control, as modelled by Scott and Rescan (2017), then there are autosomal alleles whose average time spent in each lifestage differs from that of other genes. The respective loci would be expected to show deviant patterns of genetic diversity. However, this would presumably affect only a small number of loci, not skewing the overall picture.

Fifth, recent research has shown that evolutionary rates covary with gene age (Moutinho, Eyre-Walker, et al., 2022) and protein features such as ‘relative solvent accessibility’ and ‘intrinsic disorder’ (Moutinho, Trancoso, et al., 2019). While these are important determinants for the level of divergence between taxa, it seems less likely that these would covary with lifestage expression bias.

Sixth and most importantly, there may be imprinting which causes parent-of-origin bias to allele expression in the sporophyte stage. The potential for imprinting has been pointed out by Haig (2013) and Haig and Wilczek (2006) in relation to life-stage conflict in bryophytes. Recent work has uncovered extreme imprinting in *Marchantia*, with expression of most sporophyte-expressed genes being strongly biased towards alleles inherited from the maternal gametophyte (Montgomery, Hisanaga, et al., 2022). Such a drastic expression reduction of paternal alleles is bound to reduce the amount of selection acting on these, in particular if the respective gene’s expression shows a strong sporophyte bias. Assuming that deleterious alleles tend to be rare and thus mainly to be present in haploid state, imprinting would approximately halve the selection acting on a sporophyte expressed deleterious allele. Masking would modify the selection by the dominance coefficient h . Depending on the value of h of mutant alleles in *Marchantia*, imprinting could have a stronger ($h \geq 1/2$), similar ($h \approx 1/2$), or weaker effects ($h \leq 1/2$) than masking on the reduction of the efficacy of selection. Nevertheless, both

imprinting and masking effects should act in the same direction, reducing the efficacy of selection in the sporophyte compared to the gametophyte and thus increasing diversity in the sporophyte compared to the gametophyte.

Because there are numerous phenomena that covary with ploidy, notably the massive imprinting at the sporophyte stage, it is not possible to attribute the observed lifestage differences in selection efficacy to the effects of masking.

3.4 Conclusion and outlook

Using novel methodology, we found significant associations between patterns of diversity and divergence on one hand and the degree of per-gene lifestage expression bias on the other, in the haploid-diploid liverwort *Marchantia*. We see this as evidence for a higher selection efficacy at the gametophyte lifestage of *Marchantia*. This effect is unlikely to be caused by the masking of deleterious mutations in the diploid lifestage because of the pervasive imprinting observed in *Marchantia*'s sporophytes (Montgomery and Berger, 2024).

Recent and ongoing theoretical research keeps adding to our understanding of evolution in haploid-diploids, where allelic fixation probabilities can differ considerably from species with other life cycles (Bessho and Otto, 2017, 2022) and lifestage expression bias may affect the architecture of local adaptation (Zwaenepoel et al., 2024) in analogy to sex chromosomal expression (Fraïsse and Sachdeva, 2021; Lasne et al., 2017). Promising new simulation tools have become available (Haller and Messer, 2023; Sorojsrisom et al., 2022), which may help isolating the effect of ploidy while varying other life cycle parameters. It is an exciting time to be working on haploid-diploids.

4 Methods

4.1 Plant collection, cultivation, DNA extraction, and sequencing

We collected 13 individuals of *Marchantia polymorpha* subsp. *ruderalis* in southern Edinburgh (UK). We then established axenic lines from single gemmae sterilised with NaDCC (Duckett et al., 2004), which we grew on plates of a modified Johnson's medium (Mulvey and Dolan, 2023) until the thalli reached approximately 2 cm.

We extracted genomic DNA using a phenol chloroform protocol (Lu et al., 2024). The DNA was sequenced at Novogene (Cambridge, UK) to yield approximately 5X coverage depth (discounting for PCR duplicates).

4.2 Genomic sequencing data and related bioinformatic analyses

We downloaded the newly generated sequencing data in FASTQ format from Novogene's web service. We also downloaded the data generated by Sandler et al. (2023). We trimmed all reads using Skewer (Jiang et al., 2014), discarding all 3' nucleotides with a PHRED quality score less than 30. We downloaded from <https://marchantia.info/> the the MpTak v6.1 reference assembly and annotation, and we aligned all sequencing reads using bwa-mem2 (Vasimuddin et al., 2019). We coordinate-sorted the resulting BAM files using SAMtools (Li and Durbin, 2009), marked duplicates using Picard tools (<https://broadinstitute.github.io/picard/>), and finally called variants with Freebayes (Garrison and Marth, 2012).

To restrict our downstream analyses to reliable single-copy regions, we generated a BED file of repetitive regions using the k-mer based masking program Red (Girgis, 2015). Starting from the reference assembly and annotation, we used Simon Martin’s Python script `codingSiteTypes.py` from the repository https://github.com/simonhmartin/genomics_general to identify fully degenerate and non-degenerate nucleotide sites. We developed Python code using the package scikit-allel (Miles et al., 2023) to compute population genetic statistics per gene.

To assess the decay of linkage disequilibrium, we used PopLDdecay (C. Zhang et al., 2019). Because this program runs only on diploid VCF files, we generated a phased ‘diploid’ VCF by joining pairs of individuals into diploid variant call columns.

4.3 Gene expression statistics

We downloaded the Marpolbase Expression data from <https://marchantia.info/mbex/> (accessed 11th April 2023). We then removed samples that might have contained RNA from both sporophyte and gametophyte tissues. We retained only genes with a summed TPM value > 10 . We then averaged transcription levels per gene and tissue. In a second step, we averaged per-gene tissue means per lifestage. We then divided the average gametophyte expression values by the sum of the average expression values from gametophyte and sporophyte, to obtain the proportion of haploid expression (PHE). PHE being a proportion, its values ranged from 0 to 1. We also computed the overall average expression value for each gene as the average of the per-tissue values. We finally computed the tissue specificity of gene expression, τ (Yanai et al., 2005), for each gene. To compute τ , we used haploid tissues only to avoid excessive correlation with PHE.

4.4 Other per-gene statistics

We computed each of the following statistics for each gene, separately for non-degenerate and fully degenerate sites. We computed nucleotide diversity using the scikit-allel package. We assessed the skew of the site-frequency spectrum both using the statistic $\Delta\theta$ (Becher et al., 2020), which is a simple function of the nucleotide diversity and standardised number of segregating sites $\Delta\theta = 1 - \pi/\theta_w$. We also assessed the net divergence, d_A (Nei, 1987; Nei and W.-H. Li, 1979), between *Marchantia* and *M. p.* subsp. *polymorpha*. We obtained this by computing d_{xy} and then subtracting half of each subspecies’ nucleotide diversity. To compute the frequency of the optimal codon, we used codonW (Peden, 1999).

4.5 Statistical analyses

We used R (R Core Team, 2023) for all statistical analyses. To analyse the per-gene data and its association with expression in different life stages accounting for background variation in levels of genetic diversity we used the package `mgcv` (<https://CRAN.R-project.org/package=mgcv>). In the analyses of nucleotide diversity, θ_w , and d_A , we used $\log_{10}$ transformation on the outcome variable.

5 Acknowledgements

HB is indebted to Christelle Fraïsse for discussions that lead to this inception of this study, which begun as TS's Bioinformatics MSc project at the University of Edinburgh's School of Biological Sciences. HB is grateful to Alex D Twyford who allowed him to take on two MSc students during his time at the Twyford Lab. We thank Matthew Hartfield, Konrad Lohse, and Brian and Deborah Charlesworth for their comments on an earlier version of this manuscript. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (<http://www.ecdf.ed.ac.uk/>).

6 Data availability

The sequencing reads generated are available from the sequence read archive, BioProject PRJNA1121287. The analysis code is available from Zenodo: [www/gggbbb](https://zenodo.org/record/1388888). [Both available upon acceptance of this manuscript.]

7 Author contributions

TS carried out bioinformatic analyses and computed per-gene summary statistics. PG and JG supplied access to wetlab infrastructure, helped with plant husbandry, and carried out DNA extractions. HB designed the study and wrote the first draft of the manuscript. All authors read, commented on, and approved the final manuscript.

8 Competing interests

The authors declare no competing interests.

References

- Arunkumar, Ramesh et al. (Nov. 1, 2013). "Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*". In: *Molecular Biology and Evolution* 30.11, pp. 2475–2486. ISSN: 0737-4038. DOI: [10.1093/molbev/mst149](https://doi.org/10.1093/molbev/mst149).
- Avery, P. J. (Dec. 1984). "The population genetics of haplo-diploids and X-linked genes". In: *Genetics Research* 44.3, pp. 321–341. ISSN: 1469-5073, 0016-6723. DOI: [10.1017/S0016672300026550](https://doi.org/10.1017/S0016672300026550).
- Bainard, Jillian D and Steven G Newmaster (June 29, 2010). "Endopolyploidy in bryophytes: Widespread in mosses and absent in liverworts". In: *Journal of Botany* 2010. Publisher: Hindawi, e316356. ISSN: 2090-0120. DOI: [10.1155/2010/316356](https://doi.org/10.1155/2010/316356).
- Bainard, Jillian D, Steven G Newmaster, and Jessica M Budke (Mar. 29, 2020). "Genome size and endopolyploidy evolution across the moss phylogeny". In: *Annals of Botany* 125.4, pp. 543–555. ISSN: 0305-7364. DOI: [10.1093/aob/mcz194](https://doi.org/10.1093/aob/mcz194).
- Barroso, Gustavo V. and Julien Y. Dutheil (2023). "The landscape of nucleotide diversity in *Drosophila melanogaster* is shaped by mutation rate variation". In: *Peer Community Journal* 3. ISSN: 2804-3871. DOI: [10.24072/pcjournal.267](https://doi.org/10.24072/pcjournal.267).

- Beaudry, Felix E. G. et al. (Nov. 9, 2020). "Evolutionary genomics of plant gametophytic selection". In: *Plant Communications*. Special Issue on Plant Evolutionary Adaptation (Organizing Editors: Loren Rieseberg, Richard Abbott, Nolan Kane, Sophie Karrenberg, Daniel Ortiz-Barrientos, Suhua Shi) 1.6, p. 100115. ISSN: 2590-3462. DOI: [10.1016/j.xplc.2020.100115](https://doi.org/10.1016/j.xplc.2020.100115).
- Becher, Hannes, Benjamin C. Jackson, and Brian Charlesworth (2020). "Patterns of genetic variability in genomic regions with low rates of recombination". In: *Current Biology* 30.1, 94–100.e3. ISSN: 0960-9822. DOI: <https://doi.org/10.1016/j.cub.2019.10.047>.
- Begun, David J. and Charles F. Aquadro (Apr. 1992). "Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*". In: *Nature* 356.6369. Number: 6369 Publisher: Nature Publishing Group, pp. 519–520. ISSN: 1476-4687. DOI: [10.1038/356519a0](https://doi.org/10.1038/356519a0).
- Bessho, Kazuhiro and Sarah P. Otto (Jan. 2017). "Fixation probability in a haploid-diploid population". In: *Genetics* 205.1, 421 LP–440. DOI: [10.1534/genetics.116.192856](https://doi.org/10.1534/genetics.116.192856).
- (Feb. 1, 2022). "Fixation and effective size in a haploid-diploid population with asexual reproduction". In: *Theoretical Population Biology* 143, pp. 30–45. ISSN: 0040-5809. DOI: [10.1016/j.tpb.2021.11.002](https://doi.org/10.1016/j.tpb.2021.11.002).
- Booker, Tom R., Bret A. Payseur, and Anna Tigano (June 22, 2022). "Background selection under evolving recombination rates". In: *Proceedings of the Royal Society B: Biological Sciences* 289.1977. Publisher: Royal Society, p. 20220782. DOI: [10.1098/rspb.2022.0782](https://doi.org/10.1098/rspb.2022.0782).
- Bourdareau, Simon et al. (2021). "Alternation of generations in plants and algae". In: *Evolutionary Developmental Biology: A Reference Guide*. Ed. by Laura Nuño de la Rosa and Gerd B. Müller. Cham: Springer International Publishing, pp. 631–644. ISBN: 978-3-319-32979-6. DOI: [10.1007/978-3-319-32979-6_58](https://doi.org/10.1007/978-3-319-32979-6_58).
- Bowman, J L (Feb. 2016). "A brief history of Marchantia from Greece to genomics". In: *Plant and Cell Physiology* 57.2, pp. 210–229. ISSN: 0032-0781. DOI: [10.1093/pcp/pcv044](https://doi.org/10.1093/pcp/pcv044).
- Bowman, J L et al. (Oct. 1, 2022). "The renaissance and enlightenment of Marchantia as a model system". In: *The Plant Cell* 34.10, pp. 3512–3542. ISSN: 1040-4651. DOI: [10.1093/plcell/koac219](https://doi.org/10.1093/plcell/koac219).
- Campos, J L, Daniel L. Halligan, et al. (Apr. 2014). "The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*". In: *Molecular Biology and Evolution* 31.4. Publisher: Oxford University Press, pp. 1010–1028. DOI: [10.1093/molbev/msu056](https://doi.org/10.1093/molbev/msu056).
- Campos, J L, Keira J A Johnston, and Brian Charlesworth (Mar. 1, 2018). "The effects of sex-biased gene expression and X-linkage on rates of sequence evolution in *Drosophila*". In: *Molecular Biology and Evolution* 35.3, pp. 655–665. ISSN: 0737-4038. DOI: [10.1093/molbev/msx317](https://doi.org/10.1093/molbev/msx317).
- Campos, J L, Kai Zeng, et al. (Apr. 2013). "Codon usage bias and effective population sizes on the X chromosome versus the autosomes in *Drosophila melanogaster*". In: *Molecular Biology and Evolution* 30.4. Publisher: Oxford University Press, pp. 811–823. DOI: [10.1093/molbev/mss222](https://doi.org/10.1093/molbev/mss222).
- Castellano, David, Adam Eyre-Walker, and Kasper Munch (Jan. 1, 2020). "Impact of mutation rate and selection at linked sites on DNA variation across the genomes of humans and other homininae". In: *Genome Biology and Evolution* 12.1, pp. 3550–3561. ISSN: 1759-6653. DOI: [10.1093/gbe/evz215](https://doi.org/10.1093/gbe/evz215).
- Charlesworth, B., J. A. Coyne, and N. H. Barton (July 1987). "The relative rates of evolution of sex chromosomes and autosomes". In: *The American Naturalist* 130.1. Publisher: University of Chicago Press, pp. 113–146. ISSN: 0003-0147. DOI: [10.1086/284701](https://doi.org/10.1086/284701).
- Charlesworth, B., M. T. Morgan, and D. Charlesworth (Aug. 1993). "The effect of deleterious mutations on neutral molecular variation." In: *Genetics* 134.4, 1289 LP–1303.

- Coelho, Susana M. et al. (Dec. 30, 2007). "Complex life cycles of multicellular eukaryotes: New approaches based on the use of model organisms". In: *Gene. Marine Genomics* 406.1, pp. 152–170. ISSN: 0378-1119. DOI: [10.1016/j.gene.2007.07.025](https://doi.org/10.1016/j.gene.2007.07.025).
- Cook, Daniel L., Anthony N. Gerber, and Stephen J. Tapscott (Dec. 22, 1998). "Modeling stochastic gene expression: Implications for haploinsufficiency". In: *Proceedings of the National Academy of Sciences* 95.26. Publisher: Proceedings of the National Academy of Sciences, pp. 15641–15646. DOI: [10.1073/pnas.95.26.15641](https://doi.org/10.1073/pnas.95.26.15641).
- De Jong, H. and G. C. C. Tai (June 1, 1977). "Analysis of tetraploid-diploid hybrids in cultivated potatoes". In: *Potato Research* 20.2, pp. 111–121. ISSN: 1871-4528. DOI: [10.1007/BF02360270](https://doi.org/10.1007/BF02360270).
- Duckett, Jeffrey G. et al. (Mar. 2004). "In vitro cultivation of bryophytes: a review of practicalities, problems, progress and promise". In: *Journal of Bryology* 26.1. Publisher: Taylor & Francis, pp. 3–20. ISSN: 0373-6687. DOI: [10.1179/174328213X13789822578469](https://doi.org/10.1179/174328213X13789822578469).
- Fortune, P. M. et al. (2008). "The enigmatic invasive *Spartina densiflora*: A history of hybridizations in a polyploidy context". In: *Molecular Ecology* 17.19. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-294X.2008.03916.x>, pp. 4304–4316. ISSN: 1365-294X. DOI: [10.1111/j.1365-294X.2008.03916.x](https://doi.org/10.1111/j.1365-294X.2008.03916.x).
- Fox, Donald T. et al. (Sept. 1, 2020). "Polyploidy: A biological force from cells to ecosystems". In: *Trends in Cell Biology* 30.9. Publisher: Elsevier, pp. 688–694. ISSN: 0962-8924, 1879-3088. DOI: [10.1016/j.tcb.2020.06.006](https://doi.org/10.1016/j.tcb.2020.06.006).
- Fraïsse, C and Himani Sachdeva (Feb. 1, 2021). "The rates of introgression and barriers to genetic exchange between hybridizing species: sex chromosomes vs autosomes". In: *Genetics* 217.2, iyaa025. ISSN: 1943-2631. DOI: [10.1093/genetics/iyaa025](https://doi.org/10.1093/genetics/iyaa025).
- Galtier, Nicolas (2016). "Adaptive protein evolution in animals and the effective population size hypothesis". In: *PLoS Genetics* 12.1. ISBN: 9788578110796. ISSN: 15537404. DOI: [10.1371/journal.pgen.1005774](https://doi.org/10.1371/journal.pgen.1005774).
- Garrison, Erik and Gabor Marth (July 20, 2012). *Haplotype-based variant detection from short-read sequencing*. DOI: [10.48550/arXiv.1207.3907](https://doi.org/10.48550/arXiv.1207.3907). arXiv: [1207.3907\[q-bio\]](https://arxiv.org/abs/1207.3907).
- Girgis, Hani Z. (July 24, 2015). "Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale". In: *BMC Bioinformatics* 16.1, p. 227. ISSN: 1471-2105. DOI: [10.1186/s12859-015-0654-5](https://doi.org/10.1186/s12859-015-0654-5).
- Gossmann, Toni I. et al. (Mar. 2014). "Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*". In: *Molecular Biology and Evolution* 31.3. Publisher: Oxford University Press, pp. 574–583. DOI: [10.1093/molbev/mst226](https://doi.org/10.1093/molbev/mst226).
- Haig, David (Mar. 1, 2013). "Filial mistletoes: the functional morphology of moss sporophytes". In: *Annals of Botany* 111.3, pp. 337–345. ISSN: 0305-7364. DOI: [10.1093/aob/mcs295](https://doi.org/10.1093/aob/mcs295).
- Haig, David and Amity Wilczek (Jan. 4, 2006). "Sexual conflict and the alternation of haploid and diploid generations". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 361.1466. Publisher: Royal Society, pp. 335–343. DOI: [10.1098/rstb.2005.1794](https://doi.org/10.1098/rstb.2005.1794).
- Halldorsson, Bjarni V. et al. (Jan. 25, 2019). "Characterizing mutagenic effects of recombination through a sequence-level genetic map". In: *Science* 363.6425. Publisher: American Association for the Advancement of Science, eaau1043. DOI: [10.1126/science.aau1043](https://doi.org/10.1126/science.aau1043).
- Haller, Benjamin C. and Philipp W. Messer (May 2023). "SLiM 4: Multispecies eco-evolutionary modeling". In: *The American Naturalist* 201.5. Publisher: The University of Chicago Press, E127–E139. ISSN: 0003-0147. DOI: [10.1086/723601](https://doi.org/10.1086/723601).
- Hanawalt, Philip C. and Graciela Spivak (Dec. 2008). "Transcription-coupled DNA repair: two decades of progress and surprises". In: *Nature Reviews Molecular Cell Biology* 9.12. Publisher: Nature Publishing Group, pp. 958–970. ISSN: 1471-0080. DOI: [10.1038/nrm2549](https://doi.org/10.1038/nrm2549).
- Hinch, Robert, Peter Donnelly, and Anjali Gupta Hinch (Dec. 2023). "Meiotic DNA breaks drive multifaceted mutagenesis in the human germ line". In: *Science* 382.6674. Publisher:

- 614 American Association for the Advancement of Science, eadh2531. DOI: [10.1126/science.adh2531](https://doi.org/10.1126/science.adh2531).
- 615
- 616 Hodgkinson, Alan, Emmanuel Ladoukakis, and Adam Eyre-Walker (Feb. 3, 2009). "Cryptic
- 617 variation in the human mutation rate". In: *PLOS Biology* 7.2. Publisher: Public Library of
- 618 Science, e1000027. ISSN: 1545-7885. DOI: [10.1371/journal.pbio.1000027](https://doi.org/10.1371/journal.pbio.1000027).
- 619 Hoshino, Masakazu et al. (Aug. 16, 2024). "Parallel loss of sexual reproduction in field popu-
- 620 lations of a brown alga sheds light on the mechanisms underlying the emergence of asexu-
- 621 ality". In: *Nature Ecology & Evolution*. Publisher: Nature Publishing Group, pp. 1–17. ISSN:
- 622 2397-334X. DOI: [10.1038/s41559-024-02490-w](https://doi.org/10.1038/s41559-024-02490-w).
- 623 Hu, Tina T. et al. (Jan. 2013). "A second-generation assembly of the *Drosophila simulans* genome
- 624 provides new insights into patterns of lineage-specific divergence". In: *Genome research*
- 625 23.1. Publisher: Cold Spring Harbor Laboratory Press, pp. 89–98. ISSN: 1549-5469. DOI:
- 626 [10.1101/gr.141689.112](https://doi.org/10.1101/gr.141689.112).
- 627 Immler, Simone, Göran Arnqvist, and Sarah Perin Otto (Jan. 1, 2012). "Ploidally antagonis-
- 628 tic selection maintains stable genetic polymorphism". In: *Evolution* 66.1, pp. 55–65. ISSN:
- 629 0014-3820. DOI: [10.1111/j.1558-5646.2011.01399.x](https://doi.org/10.1111/j.1558-5646.2011.01399.x).
- 630 Jiang, Hongshan et al. (2014). "Skewer: a fast and accurate adapter trimmer for next-generation
- 631 sequencing paired-end reads". In: *BMC Bioinformatics* 15.1, p. 182. ISSN: 1471-2105. DOI:
- 632 [10.1186/1471-2105-15-182](https://doi.org/10.1186/1471-2105-15-182).
- 633 Kolář, Filip et al. (2017). "Mixed-ploidy species: progress and opportunities in polyploid re-
- 634 search". In: *Trends in Plant Science* 22.12, pp. 1041–1055. ISSN: 1360-1385. DOI: <https://doi.org/10.1016/j.tplants.2017.09.011>.
- 635
- 636 Kondrashov, Alexey S. and James F. Crow (May 1991). "Haploidy or diploidy: which is better?"
- 637 In: *Nature* 351.6324. Number: 6324 Publisher: Nature Publishing Group, pp. 314–315. ISSN:
- 638 1476-4687. DOI: [10.1038/351314a0](https://doi.org/10.1038/351314a0).
- 639 Langley, Charles H et al. (Oct. 1, 2012). "Genomic variation in natural populations of *Drosophila*
- 640 *melanogaster*". In: *Genetics* 192.2, pp. 533–598. ISSN: 1943-2631. DOI: [10.1534/genetics.112.142018](https://doi.org/10.1534/genetics.112.142018).
- 641
- 642 Larracuente, A M et al. (Mar. 1, 2008). "Evolution of protein-coding genes in *Drosophila*". In:
- 643 *Trends in Genetics* 24.3, pp. 114–123. ISSN: 0168-9525. DOI: [10.1016/j.tig.2007.12.001](https://doi.org/10.1016/j.tig.2007.12.001).
- 644
- 645 Lasne, Clémentine, Carla M. Sgrò, and Tim Connallon (Mar. 2017). "The relative contributions
- 646 of the X chromosome and autosomes to Local adaptation". In: *Genetics* 205.3. Publisher:
- 647 Genetics, pp. 1285–1304. ISSN: 1943-2631. DOI: [10.1534/genetics.116.194670](https://doi.org/10.1534/genetics.116.194670).
- 648 Li, Heng and R. Durbin (July 2009). "Fast and accurate short read alignment with Burrows-
- 649 Wheeler transform". In: *Bioinformatics* 25.14, pp. 1754–1760. ISSN: 1367-4803. DOI: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324).
- 650
- 651 Linde, Anna-Malin et al. (2021). "Rates and patterns of molecular evolution in bryophyte
- 652 genomes, with focus on complex thalloid liverworts, Marchantiopsida". In: *Molecular Phy-*
- 653 *logenetics and Evolution* 165, p. 107295. ISSN: 1055-7903. DOI: <https://doi.org/10.1016/j.ympev.2021.107295>.
- 654
- 655 Lipinska, Agnieszka P. et al. (Feb. 14, 2019). "Rapid turnover of life-cycle-related genes in the
- 656 brown algae". In: *Genome Biology* 20.1, p. 35. ISSN: 1474-760X. DOI: [10.1186/s13059-019-1630-6](https://doi.org/10.1186/s13059-019-1630-6).
- 657
- 658 López-Cortegano, Eugenio et al. (Sept. 1, 2021). "De novo mutation rate variation and its de-
- 659 terminants in *Chlamydomonas*". In: *Molecular Biology and Evolution* 38.9, pp. 3709–3723. ISSN:
- 660 1537-1719. DOI: [10.1093/molbev/msab140](https://doi.org/10.1093/molbev/msab140).
- 661 Lu, Yen-Ting et al. (Feb. 26, 2024). "Convergent evolution of water-conducting cells in *Marchan-*
- 662 *tia* recruited the ZHOUP1 gene promoting cell wall reinforcement and programmed cell
- 663 death". In: *Current Biology* 34.4. Publisher: Elsevier, 793–807.e7. ISSN: 0960-9822. DOI: [10.1016/j.cub.2024.01.014](https://doi.org/10.1016/j.cub.2024.01.014).
- 664

- Maynard Smith, John and John Haigh (Feb. 1974). "The hitch-hiking effect of a favourable gene". In: *Genetical Research* 23.1, pp. 23–35. ISSN: 0016-6723. DOI: [10.1017/S0016672300014634](https://doi.org/10.1017/S0016672300014634).
- Miles, Alistair et al. (May 17, 2023). *cggh/scikit-allele: v1.3.5*. DOI: [10.5281/zenodo.7945875](https://doi.org/10.5281/zenodo.7945875).
- Montgomery, Sean A. and Frédéric Berger (2024). "Paternal imprinting in *Marchantia polymorpha*". In: *New Phytologist* 241.3. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/nph.19377>, pp. 1000–1006. ISSN: 1469-8137. DOI: [10.1111/nph.19377](https://doi.org/10.1111/nph.19377).
- Montgomery, Sean A., Tetsuya Hisanaga, et al. (Aug. 23, 2022). "Polycomb-mediated repression of paternal chromosomes maintains haploid dosage in diploid embryos of *Marchantia*". In: *eLife* 11. Ed. by Mia T Levine and Jessica K Tyler. Publisher: eLife Sciences Publications, Ltd, e79258. ISSN: 2050-084X. DOI: [10.7554/eLife.79258](https://doi.org/10.7554/eLife.79258).
- Moutinho, Ana Filipa, Adam Eyre-Walker, and Julien Y. Dutheil (Sept. 13, 2022). "Strong evidence for the adaptive walk model of gene evolution in *Drosophila* and *Arabidopsis*". In: *PLOS Biology* 20.9. Publisher: Public Library of Science, e3001775. ISSN: 1545-7885. DOI: [10.1371/journal.pbio.3001775](https://doi.org/10.1371/journal.pbio.3001775).
- Moutinho, Ana Filipa, Fernanda Fontes Trancoso, and Julien Yann Dutheil (Sept. 2019). "The impact of protein architecture on adaptive evolution". In: *Molecular Biology and Evolution* 36.9, pp. 2013–2028. ISSN: 0737-4038. DOI: [10.1093/molbev/msz134](https://doi.org/10.1093/molbev/msz134).
- Mulvey, Hugh and Liam Dolan (July 24, 2023). "RHO GTPase of plants regulates polarized cell growth and cell division orientation during morphogenesis". In: *Current Biology* 33.14. Publisher: Elsevier, 2897–2911.e6. ISSN: 0960-9822. DOI: [10.1016/j.cub.2023.06.015](https://doi.org/10.1016/j.cub.2023.06.015).
- Nei, M (1987). *Molecular evolutionary genetics*. New York: Columbia University Press. 512 pp. ISBN: 0-231-06320-2. DOI: [10.7312/nei-92038](https://doi.org/10.7312/nei-92038).
- Nei, M and W-H Li (1979). "Mathematical model for studying genetic variation in terms of restriction endonucleases (molecular evolution/mitochondrial DNA/nucleotide diversity)". In: *Genetics* 76.10, pp. 5269–5273.
- Nicholson, Michael D. et al. (May 14, 2024). "DNA lesion bypass and the stochastic dynamics of transcription-coupled repair". In: *Proceedings of the National Academy of Sciences* 121.20. Publisher: Proceedings of the National Academy of Sciences, e2403871121. DOI: [10.1073/pnas.2403871121](https://doi.org/10.1073/pnas.2403871121).
- Pandit, Maharaj K., Michael J. O. Pocock, and William E. Kunin (2011). "Ploidy influences rarity and invasiveness in plants". In: *Journal of Ecology* 99.5. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2745.2011.01838.x>, pp. 1108–1115. ISSN: 1365-2745. DOI: [10.1111/j.1365-2745.2011.01838.x](https://doi.org/10.1111/j.1365-2745.2011.01838.x).
- Payseur, Bret A. and Michael W. Nachman (Mar. 1, 2002). "Gene density and human nucleotide polymorphism". In: *Molecular Biology and Evolution* 19.3, pp. 336–340. ISSN: 0737-4038. DOI: [10.1093/oxfordjournals.molbev.a004086](https://doi.org/10.1093/oxfordjournals.molbev.a004086).
- Peden, John F (1999). "Analysis of codon usage". PhD thesis. University of Nottingham. 215 pp.
- Pyšek, Petr et al. (2020). "Competition among native and invasive *Phragmites australis* populations: An experimental test of the effects of invasion status, genome size, and ploidy level". In: *Ecology and Evolution* 10.3. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.5907>, pp. 1106–1118. ISSN: 2045-7758. DOI: [10.1002/ece3.5907](https://doi.org/10.1002/ece3.5907).
- Renny-Byfield, Simon et al. (Apr. 2010). "Flow cytometry and GISH reveal mixed ploidy populations and *Spartina* nonaploids with genomes of *S. alterniflora* and *S. maritima* origin." In: *Annals of botany* 105.4, pp. 527–33. ISSN: 1095-8290. DOI: [10.1093/aob/mcq008](https://doi.org/10.1093/aob/mcq008).
- Rice, Anna et al. (Apr. 2015). "The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers". In: *New Phytologist* 206.1. Publisher: John Wiley & Sons, Ltd, pp. 19–26. ISSN: 0028-646X. DOI: <https://doi.org/10.1111/nph.13191>.
- Rousselle, Marjolaine et al. (Oct. 1, 2016). "Hemizygosity enhances purifying selection: Lack of fast-Z evolution in two satyrine butterflies". In: *Genome Biology and Evolution* 8.10, pp. 3108–3119. ISSN: 1759-6653. DOI: [10.1093/gbe/evw214](https://doi.org/10.1093/gbe/evw214).

- Salony, Susnata et al. (Aug. 5, 2024). "Establishment of polyploidy in natural populations of *Mimulus guttatus*". In: *Plant Systematics and Evolution* 310.5, p. 30. ISSN: 1615-6110. DOI: [10.1007/s00606-024-01914-1](https://doi.org/10.1007/s00606-024-01914-1).
- Sandler, George, Aneil F Agrawal, and Stephen I Wright (Nov. 1, 2023). "Population genomics of the facultatively sexual liverwort *Marchantia polymorpha*". In: *Genome Biology and Evolution* 15.11, evad196. ISSN: 1759-6653. DOI: [10.1093/gbe/evad196](https://doi.org/10.1093/gbe/evad196).
- Schmickl, Roswitha and Levi Yant (2021). "Adaptive introgression: how polyploidy reshapes gene flow landscapes". In: *New Phytologist* 230.2. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/nph.17204>. pp. 457–461. ISSN: 1469-8137. DOI: [10.1111/nph.17204](https://doi.org/10.1111/nph.17204).
- Scott, Michael F and Marie Rescan (Feb. 1, 2017). "Evolution of haploid–diploid life cycles when haploid and diploid fitnesses are not equal". In: *Evolution* 71.2, pp. 215–226. ISSN: 0014-3820. DOI: [10.1111/evo.13125](https://doi.org/10.1111/evo.13125).
- Slotte, T et al. (Jan. 1, 2011). "Genomic determinants of protein evolution and polymorphism in *Arabidopsis*". In: *Genome Biology and Evolution* 3, pp. 1210–1219. ISSN: 1759-6653. DOI: [10.1093/gbe/evr094](https://doi.org/10.1093/gbe/evr094).
- Slotte, Tanja et al. (Aug. 1, 2010). "Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size". In: *Molecular Biology and Evolution* 27.8, pp. 1813–1821. ISSN: 0737-4038. DOI: [10.1093/molbev/msq062](https://doi.org/10.1093/molbev/msq062).
- Smith, Thomas C. A., Peter F. Arndt, and Adam Eyre-Walker (Mar. 28, 2018). "Large scale variation in the rate of germ-line de novo mutation, base composition, divergence and diversity in humans". In: *PLOS Genetics* 14.3. Publisher: Public Library of Science, e1007254. ISSN: 1553-7404. DOI: [10.1371/journal.pgen.1007254](https://doi.org/10.1371/journal.pgen.1007254).
- Sorojsrisom, Elissa S. et al. (2022). "Selection on the gametophyte: Modeling alternation of generations in plants". In: *Applications in Plant Sciences* 10.2. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/aps3.11472>. e11472. ISSN: 2168-0450. DOI: [10.1002/aps3.11472](https://doi.org/10.1002/aps3.11472).
- Szövényi, Péter et al. (Aug. 2013). "Selection is no more efficient in haploid than in diploid life stages of an angiosperm and a moss". In: *Molecular Biology and Evolution* 30.8. Publisher: Oxford University Press, pp. 1929–1939. DOI: [10.1093/molbev/mst095](https://doi.org/10.1093/molbev/mst095).
- Thornton, Kevin and Manyuan Long (June 1, 2002). "Rapid divergence of gene duplicates on the *Drosophila melanogaster* X chromosome". In: *Molecular Biology and Evolution* 19.6, pp. 918–925. ISSN: 0737-4038. DOI: [10.1093/oxfordjournals.molbev.a004149](https://doi.org/10.1093/oxfordjournals.molbev.a004149).
- Vasimuddin, Md. et al. (May 2019). "Efficient architecture-aware acceleration of BWA-MEM for multicore systems". In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). ISSN: 1530-2075, pp. 314–324. DOI: [10.1109/IPDPS.2019.00041](https://doi.org/10.1109/IPDPS.2019.00041).
- Wang, Zhi and Jianzhi Zhang (Apr. 19, 2011). "Impact of gene expression noise on organismal fitness and the efficacy of natural selection". In: *Proceedings of the National Academy of Sciences* 108.16. Publisher: Proceedings of the National Academy of Sciences, E67–E76. DOI: [10.1073/pnas.1100059108](https://doi.org/10.1073/pnas.1100059108).
- Wood, Simon N. (2003). "Thin plate regression splines". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65.1. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9868.00374>. pp. 95–114. ISSN: 1467-9868. DOI: [10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- (May 3, 2017). *Generalized additive models: An introduction with R*. 2nd ed. New York: Chapman and Hall/CRC. 496 pp. ISBN: 978-1-315-37027-9. DOI: [10.1201/9781315370279](https://doi.org/10.1201/9781315370279).
- Wright, S I et al. (Sept. 1, 2004). "Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*". In: *Molecular Biology and Evolution* 21.9, pp. 1719–1726. ISSN: 0737-4038. DOI: [10.1093/molbev/msh191](https://doi.org/10.1093/molbev/msh191).
- Yanai, Itai et al. (Mar. 1, 2005). "Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification". In: *Bioinformatics* 21.5, pp. 650–659. ISSN: 1367-4803. DOI: [10.1093/bioinformatics/bti042](https://doi.org/10.1093/bioinformatics/bti042).

- 766 Yang, L and Brandon S. Gaut (Aug. 1, 2011). "Factors that contribute to variation in evolution-
767 ary rate among *Arabidopsis* genes". In: *Molecular Biology and Evolution* 28.8, pp. 2359–2369.
768 ISSN: 0737-4038. DOI: [10.1093/molbev/msr058](https://doi.org/10.1093/molbev/msr058).
- 769 Yin, Shanye et al. (July 13, 2009). "Dosage compensation on the active X chromosome mini-
770 mizes transcriptional noise of X-linked genes in mammals". In: *Genome Biology* 10.7, R74.
771 ISSN: 1474-760X. DOI: [10.1186/gb-2009-10-7-r74](https://doi.org/10.1186/gb-2009-10-7-r74).
- 772 Zhang, Chi et al. (May 15, 2019). "PopLDdecay: a fast and effective tool for linkage disequilib-
773 rium decay analysis based on variant call format files". In: *Bioinformatics* 35.10, pp. 1786–
774 1788. ISSN: 1367-4803. DOI: [10.1093/bioinformatics/bty875](https://doi.org/10.1093/bioinformatics/bty875).
- 775 Zhou, Qi and Doris Bachtrog (July 2012). "Sex-specific adaptation drives early sex chromosome
776 evolution in *Drosophila*." In: *Science (New York, N.Y.)* 337.6092, pp. 341–5. ISSN: 1095-9203.
777 DOI: [10.1126/science.1225385](https://doi.org/10.1126/science.1225385).
- 778 Zwaenepoel, Arthur, Himani Sachdeva, and Christelle Fraïsse (Aug. 22, 2024). "The genetic
779 architecture of polygenic local adaptation and its role in shaping barriers to gene flow". In:
780 *Genetics*, iyae140. ISSN: 1943-2631. DOI: [10.1093/genetics/iyae140](https://doi.org/10.1093/genetics/iyae140).