Cross-kingdom transcriptomic trends in the evolution of hybrid gene expression

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Genetics

at Massey University, Manawatū, New Zealand

Anna Helena Martinson-Behling

2021

Abstract

The interbreeding of two genetically distinct parental lineages may lead to the formation of a new, reproductively isolated hybrid species. Hybridisation may occur with (allopolyploidisation) or without (homoploid hybridisation) a concomitant increase in chromosome number. Both allopolyploid and homoploid hybrids face a suite of near-instantaneous and longer-term biological repercussions with the potential to impede both their formation and establishment. Central to these challenges is the rewiring of gene-regulatory networks caused by the merger of the distinct genomes inherited from both parental species. Nonetheless, the success of hybrid species is well documented across the eukaryote tree of life. The applications and benefits of hybrid species drive tangible economic and cultural impacts, on both a local and global scale. Here in Aotearoa New Zealand, hybrid species permeate multiple facets of life including pastoral health and arable crop yield, long-established viticulture and microbrewery practices, and even the intrinsic nature of the land through the presence of diverse native allopolyploid flora.

Despite the commonalities in hybrid formation, prior research on the evolution of hybrid gene expression has almost entirely focused on single hybrid species or a few gene families. This thesis presents the first cross-kingdom, transcriptome-wide study to explore the fates of genes following hybridisation. Each allopolyploid system (plants, animals and fungi) is paired with a closely-related homoploid hybrid to decouple the influence of increased chromosome number from genome merger on post-hybridisation expression patterns. Genome merger, not changes in chromosome number, has the greatest effect across all study systems. Across kingdoms, genes that are differentially expressed in parental species preferentially have more similar expression in hybrid descendants, likely as a consequence of regulatory cross-talk within the hybrid nucleus. Further commonalities are highlighted in the prevalence of gene loss or silencing among extremely differentially expressed genes in hybrid species. These general patterns suggest that the evolutionary process of hybridisation leads to common high-level expression outcomes, regardless of the particular species or kingdom.

Acknowledgments

The completion of this thesis and my Masters degree would not be have been possible without the support of many people.

I would first like to extend my deepest gratitude to my primary supervisor, Murray Cox. Murray, it has been a true privilege to be your student. Thank you for believing in me, for the many opportunities you have given me, for the time you have taken to help me grow in academia and as a person, and for teaching me that even the most challenging of obstacles can be overcome. I would also like to extend my gratitude to my co-supervisor, Patrick Biggs, for having me as your student, and for your scientific advice and support. To David Winter, my officemate and bioinformatics guru over these past few years: thank you for all that you have taught me, and for your seemingly endless patience that matched my seemingly endless stream of questions. The base in bioinformatics I have developed to take me forward into my PhD and beyond is largely due to you, and for that I will always be immensely grateful. Throughout my degree I had the privilege of producing two manuscripts. My sincere thanks to my co-authors Murray Cox, Lara Shepherd (Te Papa), David Winter and Austen Ganley (University of Auckland) for your encouragement, insight, and constructive criticism. I would like to thank my lab mates in the Computational Biology Research Group for your willingness to teach me new concepts, your constructive feedback on my work, and your kindness and comradeship.

Thank you to the Massey Foundation donors for the financial support I have received throughout my degree from the Massey Sciences Scholarship, the Joan Dingley Memorial Scholarship in Mycology, and the Massey University Masterate Scholarship. My sincere gratitude must also go to Jade Robertson, Lizzy Kent and Nicky Rees: thank you for believing in me, and for your constant support, advocacy, and encouragement. I'm extremely grateful to Jan Schmid and Dragana Gagic for the conversation we had back in 2016 that led me onto the path of bioinformatic research, and for your support since. My thanks also go to Liz Evans, my high school biology teacher, for fostering my passion for biology at a crucial time. Thank you to Kathryn Stowell for your guidance, to Natisha Magan and Bronwen Comrie-Evans for your assistance, and to Jasna Rakonjac, Rosie Bradshaw and Alan Murray for the accommodations made during my undergraduate degree - without which I am not sure I would have made it to postgraduate-level study.

My heartfelt thanks to my family and friends, both here in New Zealand and around the

world, for your support through this time. My special thanks to Mike Lusby for your programming assistance, Renée Tonkin for your scientific advice, and Scott Miller for always being there. To my Davide, I thank you for being by my side with your unwavering encouragement, support and companionship. This degree is particularly special to me, being the time in which I met you. Finally, I wish to thank my immediate family: my Mum, Dad and brother Ben. Thank you for growing my passion for learning, and particularly science, from a young age. Thank you for teaching me to lead with a questioning mind, and for encouraging me to find my own path in life. You have been there for me, quite literally, from Day 1. Now, on Day 10,189, I wish to dedicate this thesis to the three of you.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Hybrid species and their applications

In many biological systems, reproductive isolation barriers maintain species integrity through the broad inhibition of interspecific (between species) genetic exchange [1–4]. Such barriers may be pre- or post-zygotic (acting before or after mating and fertilisation), and include phenological variation [5], hybrid seed inviability [6, 7], hybrid lethality [8], genetic incompatibilities [9], and selection [10]. In some instances, however, interspecific hybridisation may result in the production of fertile offspring [11], with potential implications for speciation and adaptive radiation [12, 13]. In those instances, reproductive barriers become indispensible in maintaining the integrity of the hybrid species, preventing them from re-integrating with their parents. The leaky reproductive barriers that exist between species have enabled widespread hybridisation; in at least 25% of plant species and 10% of animal species [14], and are hypothesised to have evolved as an adaptive optimum [15].

Through their widespread existence, plant, animal and fungal hybrid species have become central to the functioning of numerous industries, on both a local and global scale. Hybridism pervades arable crop development, in cultivars such as cotton [16], wheat [17], rice [18], maize [19], Brassicas [20], potato [21], tomato [22], soybean [23], sugarcane [24], sunflower [25], peanut [26], coffee [27], oats [28], and tobacco [29]. Hybrid species are also prominent among fungi with relevance to the brewing and viticultural industries [30–32], and among species in the aquaculture industry [33–37]. Here, in Aotearoa New Zealand, where agriculture forms the backbone of the present-day economy [38], the nitrogen-fixing capabilities of a hybrid white clover have enabled pasture establishment in regions with significant mineral deficits [39]. Certain hybrid endophytic symbionts of pasture grasses produce natural insecticides, protecting their host plant and providing a considerable return to the agronomy sector [40–42]. Moreover, New Zealand's diverse presence of native flora contains numerous species of hybrid mosses [43, 44], ferns [45, 46] and angiosperms [47, 48], playing a integral role in the *mauri*, or vital essence, of our land.

1.2 The genetic consequences of hybridisation

Hybrid speciation, where hybridisation has played a prominent role in the formation of a novel species, is most easily identified in allopolyploids. These are interspecific hybrids with duplicated genomes (the prefix 'allo-' meaning 'different', 'poly-' meaning 'many', and 'ploidy' referring to the number of chromosome sets in a cell); their elevated ploidy level acting as a strong facilitator of reproductive isolation from their parental species [49]. Contrary to the conventional bifurcation of Darwinian evolution where new species evolve from a single common ancestor [50], in allopolyploidy, multiple ancestral species hybridise to form a single new species (Figure 1). Polyploidy is also present in autopolyploids ('auto-' meaning 'same'). These are not hybrids, rather forming through whole genome duplication within species [51–53]. The widely accepted continuum of polyploidy, spanning these two primary classifications, also accounts for intermediates whose chromosomes exhibit both homologous (as in autopolyploidy) and homeologous (as in allopolyploidy) pairing during cell division [54, 55].

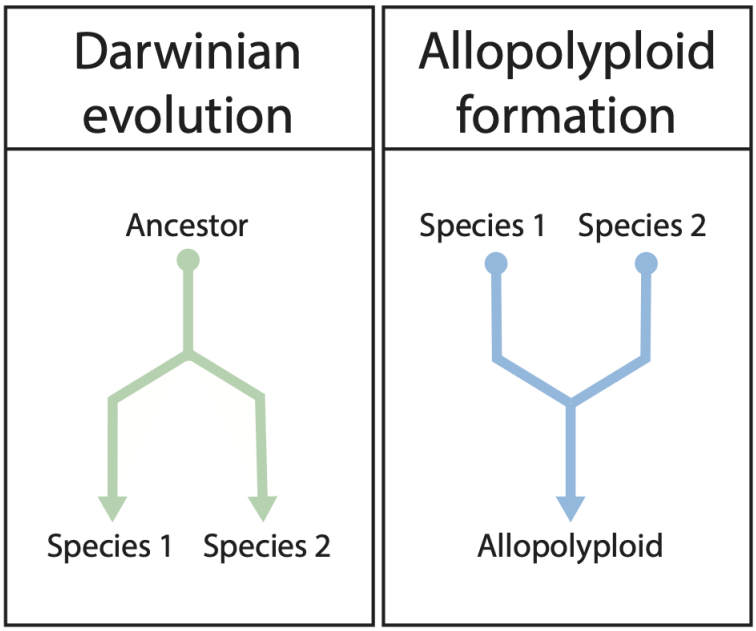

Figure 1: **Contrasting mechanisms of Darwinian evolution and allopolyploidisation.**

The hybridisation and concomitant ploidy elevation that occurs within the nascent allopolyploid nucleus 'shocks' the prior genomic state [56], inducing an extensive array of structural and biochemical changes. Rapid genomic changes have been observed in al-

lopolyploid rice [57], oilseed rape [58] and carp [59], as well as in *Carassius auratus red var.* (red crucian carp) x *Megalobrama amblycephala* (wuchang bream) allopolyploid hybrids [60]. More specifically, polyploidisation has the potential to result in gene loss, occasionally to an extent that sees the organism return to a diploid state through diploidisation. Evidence for diploidisation is seen in *Saccharomyces cerevisiae* (brewer's or baker's yeast), the descendent of an ancient allopolyploidisation event [61, 62]. The genomic restructuring following allopolyploidisation is not limited to within each parental contribution or subgenome: the recent development of three-dimensional genome mapping has supported the notion of 'intergenomic crosstalk', with evidence of inter-subgenomic A/B compartment switching and chromatin reorganisation in allopolyploid cotton [63]. Beyond the nuclear genome, allopolyploidisation has been demonstrated to influence the epigenomes [64–66], organellar genomes [67], small RNAs [68, 69] and the gene expression profiles of these species [70–72].

The definition of hybrid speciation applies less cleanly to homoploid hybrids (hybrids without a duplicated genome). Karyotypically distinct from allopolyploids, their absence of ploidy elevation may confer only weak reproductive isolation from their parental species, however, hybrid speciation can be inferred in some homoploid hybrids where marker alleles have different parental origins [49]. Further, the marked impact of hybridisation on genomic structure and gene expression can enable nascent homoploid hybrid species to colonise novel ecological niches, ameliorating the risk of back-crossing with their parental species through spatial and ecological isolation [73]. While homoploid hybrid speciation is able to eventuate within only a few generations [74, 75], it may require several hundred generations for complete genomic stabilisation of the nascent hybrid species [73].

The challenge of hybridisation lacking ploidy elevation lies in the resulting tendency towards sterility, exacerbated by higher degrees of parental divergence. Nevertheless, it appears that homoploid hybrid speciation may not be as rare as once thought [76], with data supporting the existence of a number of species in the plant kingdom [77–83] and further examples throughout animals [5, 84–87] and fungi [88–90]. The genetic consequences faced by homoploid hybrids could be considered a subset of those faced by allopolyploids; limited to those consequences arising from hybridisation alone. Homoploid hybrid chromosomal rearrangements, for instance, are purported to result from the increased transposon activity that appears to often accompany hybridisation [91–93].

1.3 Studying hybrid gene expression

The suite of hybridisation-induced biological repercussions faced by allopolyploid and homoploid hybrid species is vast, reaching different regions of the genome. Genome-wide gene expression studies are a useful approach to investigate the spread of some of these repercussions on a genome-wide scale. The core phases of a typical analysis of gene expression entail transcriptome profiling, differential expression, and interpretation [94]. Normally, transcriptome profiling will involve the alignment of reads onto an available reference genome or transcriptome. The complex inheritance of hybrid genomes presents an additional obstacle to this stage of analysis: how to define the parentage of each hybrid read. HyLiTE (hybrid lineage transcriptome explorer) [95] is a package that overcomes this obstacle by identifying diagnostic single nucleotide polymorphisms (SNPs) within RNA-seq reads, indicative of parental origin, to subsequently align hybrid reads to each parental subtranscriptome.

Changes in gene expression due to hybridisation can be inferred through the comparison of corresponding homeolog (parentally-derived hybrid gene copy) and ortholog (parental gene copy) expression levels. Within these relative expression levels, a number of general patterns may arise. Notably, some of these patterns might be non-additive, displaying a deviation from the parental mean [71]. The widespread biased expression of one parental subgenome, 'subgenome dominance', is a phenomenon associated with hybrid species [96]. Subgenome dominance may be a direct consequence of biased genetic loss [97], or may result from dissimilar transposable element density and methylation profiles between subgenomes [98]. Extreme, or transgressive, hybrid expression profiles that fall outside of the parental average have also been associated with hybridisation and may enable the hybrid to colonise novel niches that are inaccessible to their parental species [99, 100].

Research on the evolution of hybrid gene expression has largely focused on single hybrid species [86, 101–115]. While this knowledge inevitably contributes to the greater hybrid knowledgebase, it does so in a way that is restricted, usually through bespoke methodologies, and with system-specific conclusions that are not particularly applicable to other hybrid systems studied by other research groups. However, approaching hybrids from a general perspective, it is obvious that they all face a similar, and fundamental, challenge: how to coordinate biological processes following the (often, near instantaneous) merger

of multiple (often, considerably diverged) parental genomes within a single hybrid nucleus. Few research groups have performed comparative transcriptomic analyses of hybrid systems, including a study on the relative influence of hybridisation and shifting environmental conditions on hybrid yeast [116], the effects of temperature on hybrid yeast [117], the impact of light on allopolyploid soybean [118], and the comparative analysis of an allopolyploid fungal endophyte and cotton [70]. A recent study compared the 3D genomes of animal, plant and fungal eukaryote species [119], however, this study did not focus on hybrids. There is, therefore, a striking gap in the field for multiple-system, cross-kingdom comparative study of hybrid gene expression, and for a generalised framework to enable the comparison of transcriptomic profiles between species with markedly different genetic complements. It was also unclear, from the existing literature, just how influential hybrids are in New Zealand.

The framework of this thesis differs slightly from that of a 'typical' MSc thesis. During my Masters degree, I have produced two manuscripts. The first, a first-authored review article on the importance and prevalence of allopolyploidy in Aotearoa New Zealand, that was published in the Journal of the Royal Society of New Zealand in 2020. The second, a first-authored primary research article that investigates the impact of hybridisation (with and without ploidy elevation) on gene expression patterns, and the commonalities of those patterns across fungi, animals and plants. This has been submitted to Molecular Biology and Evolution in July 2021. These manuscripts are presented in chapters two and three, respectively. Additional validations, analyses and findings from this degree that were not included in either manuscript are presented as subsections to chapter three. The motivation for submitting this thesis partly by publication was to best present both the research performed, and the manuscripts produced, during this Masters degree.

2 A review of the importance and prevalence of allopolyploidy in Aotearoa New Zealand

The following publication is a review of the integral role of allopolyploid species to numerous industries in New Zealand [120]. While allopolyploids benefit humanity on a global scale, their local impact had not yet been fully evaluated in the literature. Since its online publication in the Journal of the Royal Society of New Zealand in October 2019, this paper has been cited once by a study on *Pseudomonas syringae* pv. *actinidiae* (Psa) resistance in a polyploid *Actinidia chinensis* (kiwifruit) population [121].

Subsequent research on some of the allopolyploid species that were described in the review has seen new data on abiotic stress tolerance in cotton [122–124] as well as insights into its genome evolution which could also be utilised in crop improvement strategies [125]. A review has been published on the implications of genomic hybridisation with and without genome duplication for the origin of new yeast lineages [126]. The transcriptomic response to elevated atmospheric carbon dioxide levels has been investigated in the highly traded allopolyploid *Coffee arabica* [127]. Moreover, there was a reassessment of fern and lycophyte diversity data on the Japanese archipelago, which found that putative allopolyploid species were more prevalent than putative autopolyploids [128].

My contributions to this manuscript were writing the paper and designing the figures.

JOURNAL OF THE ROYAL SOCIETY OF NEW ZEALAND
2020, VOL. 50, NO. 2, 189–210
https://doi.org/10.1080/03036758.2019.1676797

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

The importance and prevalence of allopolyploidy in Aotearoa New Zealand

Anna H. Behling[a], Lara D. Shepherd [b] and Murray P. Cox [a,c]

[a]School of Fundamental Sciences, Massey University, Palmerston North, New Zealand; [b]Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand; [c]Bio-Protection Research Centre, Massey University, Palmerston North, New Zealand

ABSTRACT

Allopolyploids arise when two or more species hybridise to form an entirely new species with a duplicated genome. Although initially met with an array of potentially catastrophic challenges triggered by the combination of two diverged parental subgenomes within a single cell, countless allopolyploids worldwide demonstrate exceptional biological resilience by not only living under these unique circumstances, but thriving. The archipelago of Aotearoa New Zealand is home to an unexpectedly large number of allopolyploid species, both indigenous and introduced. Here, we review the prevalence and importance of these species from a local perspective. The benefits of allopolyploid species permeate multiple facets of life in New Zealand, from pastoral health and arable crop yield, to long-established viticulture and brewery practices, to the intrinsic nature of the land through the presence of diverse native allopolyploid flora. Consequently, the motivation behind the pursuit of New Zealand's allopolyploid research extends beyond improvement of the global knowledgebase and also aims to drive tangible economic and cultural impacts on the country and the lives of its people.

ARTICLE HISTORY
Received 9 July 2019
Accepted 2 October 2019

KEYWORDS
New Zealand; polyploid; allopolyploid; agriculture; industry; conservation; speciation

Introduction

Polyploidy, the heritable phenomenon that generates species with increased chromosomal content, is a major force in evolution (Ren et al. 2018). It is widely accepted that polyploidy encompasses a continuum spanning its two primary classifications, autopolyploidy and allopolyploidy (Tayalé and Parisod 2013; Spoelhof et al. 2017). Autopolyploids, the prefix 'auto-' meaning 'same', are generated from whole genome duplication within species. Conversely, allopolyploids ('allo-', meaning 'different') originate from the hybridisation of two or more different species (Comai 2005). This continuum of polyploidy accounts for intermediates whose chromosomes exhibit both homologous (as in autopolyploids) and homeologous (as in allopolyploids) pairing during cell division. Intermediacy in polyploids was described in the seminal works by Stebbins (1947, 1950), who later

CONTACT Murray P. Cox m.p.cox@massey.ac.nz

asserted the inefficiency of selection in duplicated genomes, deeming polyploids to be evolutionary dead-ends (Stebbins 1971). Stebbins' view, the 'dead-end hypothesis', has seen increasing opposition through subsequent discoveries of recurrent polyploidy (Soltis and Soltis 1999), substantial dynamism within polyploid genomes (Scannell et al. 2006; Flagel and Wendel 2009; Parisod et al. 2009; Chester et al. 2012; Wang et al. 2015; Wang et al. 2018) and the presence of an ancestral polyploidy event at the base of the major land plant radiation (Jiao et al. 2011).

The genetic consequences of allopolyploidy

Allopolyploidy appears almost as an antithesis of conventional Darwinian evolution, where new species evolve from a single common ancestor, typically conceptualised as a bifurcating tree (Darwin 1859). In allopolyploidy, multiple ancestral species instead hybridise to form a single new species. Allopolyploid formation has been shown to be accompanied by extensive structural and biochemical changes to the genome, which have been described as 'genome shock' (McClintock 1984). Indeed, this phrase aptly captures the essence of this potentially catastrophic array of structural and biochemical changes that can occur, near-instantaneously, from ploidy elevation and hybridisation within the nascent allopolyploid genome. The suite of potential biological responses, both rapid and longer term, employed by the allopolyploid in response to genome shock have many potential manifestations, including genomic restructuring (Lim et al. 2008; Wu et al. 2015; Qin et al. 2016; Wang et al. 2018) and gene loss (Scannell et al. 2006; Gordon et al. 2009; Vallejo-Marín et al. 2015). Although the impact of allopolyploidy is most often viewed through the lens of nuclear genomic studies, organellar genomes (Sloan et al. 2018), small RNAs (Ha et al. 2009; Jiao et al. 2018) and the gene expression profiles of allopolyploid species (Yoo et al. 2014; Jung et al. 2015) are also affected. Extensive changes in allopolyploid gene expression ('transcriptome shock') are driven by the mixing of two dissimilar and poorly co-adapted sets of transcription factors and chromatin signatures (Cox et al. 2014).

The relative influences of hybridisation and ploidy elevation in driving post-allopolyploidisation genetic changes are highly debated, with these causal factors exhibiting lineage-specific differences. In allopolyploid *Fragaria* (strawberry) species, DNA methylation changes appear to be predominantly driven by genome doubling, whereas both genome doubling and hybridisation appear to impact gene expression changes (Wang et al. 2016). In contrast, comparative analysis of *Arabidopsis* auto- and allopolyploids has suggested that gene regulation is more influenced by hybridisation (Wang et al. 2006). A further study of synthetic allopolyploids obtained from *Chrysanthemum nankingense* (chrysanthemum) x *Tanacetum vulgare* (tansy) hybrids proposed that it is the interaction between the two processes that is important (Qi et al. 2018).

Nascent polyploids must adapt to an instantaneous increase in genomic material, with potential downstream consequences for cell regulatory functions. Such consequences are typically generated through non-proportional cellular expansion (Melaragno et al. 1993; Robinson et al. 2018), producing a stoichiometric imbalance between chromatin components and envelope proteins (Corredor et al. 2005). This scaling relationship may be further observed at the organismal level, particularly in polyploid plants (the gigas effect) (Sattler et al. 2016; Robinson et al. 2018), or there may be a compensatory reduction

in cell number that lessens the impact of increased cell size on organ size changes (Tsukaya 2008; Del Pozo and Ramirez-Parra 2014; Czesnick and Lenhard 2015). Diploidisation, the process by which a polyploid genome returns to a state of diploidy, commonly follows ploidy elevation, suggesting that the challenges associated with whole genome duplication are often too complex to allow the establishment of the nascent polyploid species (Wolfe 2001; Leitch and Bennett 2004). Alternatively, diploidisation may be triggered by the unavailability of any required ecological drivers, such as novel niche availability (Baduel et al. 2018). As a result, some presumed classical diploid lineages are in fact paleo-polyploids, having experienced single or multiple rounds of polyploidisation and diploidisation in their evolutionary history (Wolfe and Shields 1997; Wolfe 2001; Ozkan and Feldman 2009; Qiao et al. 2019).

As well as genomic changes, allopolyploidy can also yield novel phenotypes, largely through the diversity introduced by hybridisation (Hedrick 2013). Allopolyploids may be morphologically intermediate between both parents due to homeolog (parental gene copy) co-dominance at a given locus (Szymura and Farana 1978). Alternatively, hybrids may show a phenotypic similarity to one parent as a result of subgenome dominance or genomic imprinting (Heslop-Harrison 1990; Edger et al. 2017; Bird et al. 2018), or they may exhibit a phenotype beyond the range of either parental species, generated through transgressive expression (Rieseberg et al. 1999; López-Caamal and Tovar-Sánchez 2014). It is possible that some phenotypic changes observed in allopolyploids result from genome doubling, rather than hybridisation, with the multiple gene copies produced through duplication able to perform different functions. Support for this hypothesis is also provided by the phenotypic differences observed between autopolyploids, whose origin does not involve hybridisation, and their diploid ancestors (Segraves and Thompson 1999).

Advantages of allopolyploidy

The persistence of allopolyploidy across Eukarya (Yoo et al. 2013; Cox et al. 2014; Sehrish et al. 2014; Session et al. 2016; Matos et al. 2019) suggests that successful adaptation to all of these challenges may grant the allopolyploid advantages previously unavailable to its parent lineages. Their doubled and hybrid genomes can facilitate intergenomic heterosis (hybrid vigour); a phenotypic consequence where hybrid species demonstrate increased biological fitness when compared with either parental line (Baranwal et al. 2012; Fujimoto et al. 2018). At a molecular level, allopolyploid hybrid vigour may manifest as a buffering effect against deleterious recessive mutations (Gu et al. 2003), novel gene function innovation through neo- or subfunctionalisation of genes (Adams and Wendel 2005), and the evolution of complementary parental homeologs at a given genetic locus (Paterson 2005). Heterosis and transgressive phenotypes can enable some allopolyploids to outcompete their parents within the same ecological niche, or to colonise niches that are more extreme than those of either parental species. Thus, allopolyploidy is often associated with invasiveness (Ainouche et al. 2008; Kim et al. 2008; Pandit et al. 2011).

Allopolyploidy in New Zealand

New Zealand allopolyploids include introduced species; both those that underwent allopolyploidy prior to their arrival and those that were allopolyploids upon introduction, as well

Table 1. Prominent New Zealand allopolyploid taxa, with origin and commercial significance.

Origin	Taxon	Common name	Status	Study	Industry
Introduced	*Epichloë* spp.	Ryegrass endophyte	Allopolyploid	Campbell et al. (2017)	Agriculture
	Trifolium repens	White clover	Allopolyploid	Griffiths et al. (2019)	Agriculture
	Salmo salar	Atlantic salmon	Allotriploid	Harvey et al. (2017)	Aquaculture
	Saccharomyces pastorianus	Lager yeast	Allopolyploid	Casaregola et al. (2001)	Brewing
	Humulus lupulus	Hops	Intraspecific triploid	Beatson and Brewer (1994)	Brewing
	Saccharomyces cerevisiae	Brewer's yeast	Allopolyploid	Wolfe (2001); Pfliegler et al. (2012)	Brewing/ viticulture
	Dekkera bruxellensis	Wine spoilage yeast	Putative allopolyploid	Borneman et al. (2014)	Brewing/ viticulture
	Actinidia spp.	Kiwifruit	Possible allopolyploid origin	Atkinson et al. (1997)	Horticulture
	Avena sativa	Oats	Allopolyploid	Liu et al. (2017)	Horticulture
	Brassica napus	Oilseed rape	Allopolyploid	Chalhoub et al. (2014)	Horticulture
	Coffee arabica	Coffee	Allopolyploid	Lashermes et al. (2014)	Horticulture
	Gossypium spp.	Cotton	Allotetraploid	Wendel et al. (1995); Hu et al. (2015)	Horticulture
	Malus spp.	Apple	Possible allopolyploid origin	Chevreau and Laurens (1987); Tatum et al. (2005)	Horticulture
	Nicotiana tabacum	Tobacco	Allopolyploid	Bindler et al. (2011)	Horticulture
	Pyrus spp.	Pear	Possible allopolyploid origin	Evans et al. (2008)	Horticulture
	Triticum aestivum	Wheat	Allopolyploid	Ozkan et al. (2003)	Horticulture
	Zea mays	Maize	Putative allopolyploid	Gaut et al. (2000)	Horticulture
	Pilosella officinarum	Hawkweed	Allopentaploid	Morgan-Richards et al. (2004); Trewick et al. (2004)	Invasive weed
Native	*Acanthoxyla* spp.	Stick insect	Putative mosaic allotriploid	Buckley et al. (2008); Myers et al. (2013)	Native fauna
	Asplenium spp.	Spleenwort	Allopolyploid	Shepherd, Perrie, et al. (2008)	Native flora
	Lepidium spp.	Scurvy grass	Allopolyploid	Mummenhoff et al. (2004)	Native flora
	Leptinella spp.	Button daisy	Allopolyploid	Himmelreich et al. (2014)	Native flora
	Lobelia angulata	Pānakenake	Allopolyploid	Murray et al. (2004)	Native flora
	Pachycladon spp.	New Zealand rockcress	Allopolyploid	Joly et al. (2009); Mandáková et al. (2010)	Native flora
	Plantago spp.	Plantain	Allopolyploid	Ishikawa et al. (2009); Murray et al. (2010)	Native flora
	Polystichum neozelandicum	Shield fern	Allooctoploid	Perrie et al. (2003)	Native flora
	Sphagnum australe	Peat moss	Allotriploid	Karlin et al. (2009)	Native flora
	Sphagnum falcatulum	Peat moss	Allotriploid	Karlin et al. (2009)	Native flora

as an array of native and endemic allopolyploid taxa (Table 1). Allopolyploid species play an important role in the country's industries and economy, as well as in the *mauri*, or 'vital essence', of the land. Studies of New Zealand native allopolyploids offer complementary information and a unique perspective to global allopolyploid research, while improving our understanding of the evolution of a unique biota. Here, we discuss the importance and prevalence of allopolyploidy in Aotearoa New Zealand, in the context of the outcomes, timings and locations of their formation.

When Māori arrived in New Zealand from tropical Polynesia around AD 1250, they brought with them a number of tree and root crops (Leach and Stowe 2005). Only six of these species, whose cultivation was mostly marginal in New Zealand's temperate climate, are known to have survived into European times: *aute* (paper mulberry, *Broussonetia papyrifera*), *hue* (bottle gourd, *Lagenaria siceraria*), *kūmara* (*Ipomoea batatas*), *taro* (*Colocasia esulenta*), *tī pore* (*Corydyline fruticosa*) and *uwhi* (yam, *Dioscorea alata*). Polyploidy is inferred in the origins of three of these species – *kūmara* (Roullier et al. 2013), *tī pore* (Hinkle 2004) and *uwhi* (Nemorin et al. 2012) – but research to date indicates that these origins lie in autopolyploidy rather than allopolyploidy.

In contrast, post-colonial agriculture, which forms the backbone of the New Zealand economy today (Brooking 2006; Peden 2008), relies heavily on allopolyploids (Figure 1). Most introduced allopolyploids formed prior to their introduction to New Zealand, but there are also examples of allopolyploidy events occurring post-introduction.

Allopolyploidisation pre-introduction

Agricultural exports exceed NZD$25 billion per annum, most of which are products derived from pasture-grazing livestock (Ministry of Business, Innovation & Employment 2018). The 2017 census quantified national sheep, dairy cattle and beef cattle numbers at 27.5, 6.5 and 3.6 million, respectively (Stats NZ 2018). The human population of New Zealand is approaching 5 million. For the agricultural sector to maintain these consistently

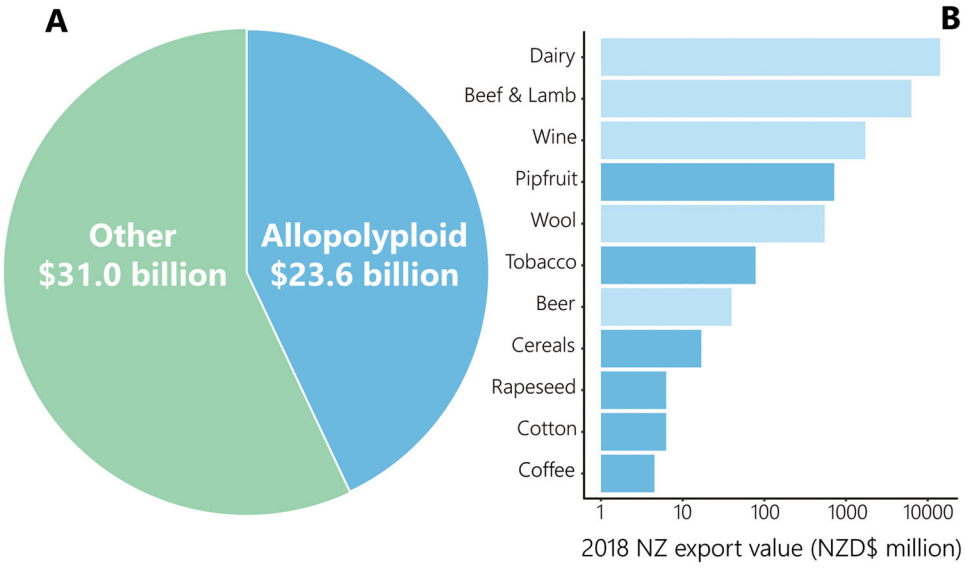

Figure 1. Economics of New Zealand allopolyploids. **(A)** Pie chart of New Zealand's total exports in 2018, separated into goods that derive directly or indirectly from systems that are reliant on allopolyploid species (blue) and other goods (green). Data were obtained using the New Zealand Trade Dashboard (Stats NZ 2019; accessed 24 June 2019) for the categories 'exports', 'goods' and '2018'. **(B)** Breakdown of allopolyploid export goods based on their individual contributions to the New Zealand economy in 2018 (NZ dollars, millions). Goods that are directly allopolyploid are represented by dark blue bars. Goods that derive indirectly from allopolyploid species are represented by pale blue bars. The *x*-axis is a log (base 10) scale.

high economic contributions, it is heavily reliant on pastoral quality and condition. Endophyte species of the genus *Epichloë* live as obligate symbionts of the cool-season grasses found in New Zealand pastures, particularly perennial ryegrass (*Lolium perenne*). An unexpectedly large number of *Epichloë* species have arisen via allopolyploidisation, with the genus subsequently becoming an emerging model system for studying fungal allopolyploidy (Schardl et al. 1994; Schardl et al. 2013; Cox et al. 2014; Campbell et al. 2017). Although central to New Zealand's agronomy, *Epichloë* species seldom form symbioses with native grasses (Rolston et al. 2002), although endemic *Epichloë* species are often allopolyploid (Leuchtmann et al. 2019). Agricultural strains were first introduced via seed brought by British immigrants during the 1800s (Stewart 2006). The coevolution between *Epichloë* endophytes and their hosts is thought to drive the specialisation and mutualistic cooperation of these symbioses (Leuchtmann et al. 1997; Saikkonen et al. 2004; Saikkonen et al. 2016). Able to persist asymptomatically within their host tissue, *Epichloë* fungi produce beneficial secondary metabolites that act as natural insecticides and enhance the survival of their grass host, in addition to improving the vegetative growth and drought tolerance of the pasture (Song and Nan 2015; Saikkonen et al. 2016).

Epichloë species are estimated to contribute NZD$200 million to the New Zealand economy per annum (Johnson et al. 2013). Given that one endophyte strain, *E. festucae* var. *lolii* AR37, introduced commercially at the 2006 Fieldays, is estimated to have taken NZD$12 million in development costs (Ministry of Business, Innovation & Employment 2018), this return has proved an immensely rewarding investment for New Zealand and its future. Different *Epichloë* strains make different compounds that protect their grass host from insect herbivory, with allopolyploid species often producing the most diverse array of compounds. While AR37 is not an allopolyploid, other allopolyploid strains such as *E. hybrida* Lp1 have been commercialised in other agricultural settings, and due to their diverse mechanisms of anti-insect protection, many contenders for third-generation commercial products are also allopolyploids. Beyond these economic incentives, *Epichloë* research also contributes fundamental information, such as the impacts of allopolyploidy on modulating fungal gene expression, to the growing global knowledge-base on allopolyploid species (Cox et al. 2014).

Epichloë endophytes are not the only allopolyploid species central to the functioning of the agricultural sector. White clover (*Trifolium repens*) is a key pastoral plant, owing to its quality forage material and nitrogen-fixing abilities (Charlton 2008). White clover emerged following an allopolyploidisation event between two European clover species confined to markedly different coastal and alpine habitats. Genetic analyses suggest that their unlikely co-habitation was driven by the reconfigured European landscape during the last glacial period (Griffiths et al. 2019). New Zealand's history of white clover research extends back some 80 years, incentivised by its advantages for pastoral agriculture. Nitrogen fixation by white clover has enabled pasture establishment in regions with significant mineral deficits, including the volcanic plateau and Northland gumlands (Brock et al. 1989). Moreover, a higher pastoral clover content has been positively correlated with dairy cow milk yields (Harris et al. 1998). White clover is a remarkable example of allopolyploidy-facilitated niche expansion that produced a ubiquitous temperate forage crop able to outcompete its two highly-specialised parental species, with surprising benefits for the modern pastoral economy. Continued research in New Zealand on the

interactions between the two parental subgenomes will enable the selection of superior white clover strains for use in pastoral breeding programmes (Griffiths et al. 2019).

Since the 1850s, pastoralism has dominated agricultural production in New Zealand, but crop production still constitutes an important source of economic gain in the agricultural sector (Peden 2008). New Zealand's major arable crops include a number of allopolyploid species formed prior to their introduction (Table 1). Cotton crops (*Gossypium* species) are a striking example of hybrid vigour through the evolution of complementary parental homeologs. Modern-day cotton cultivars are naturally-formed allotetraploids whose extinct diploid parental species, most closely related to extant *G. arboreum* and *G. raimondii* (Wendel et al. 1995), hybridised 1–2 million years ago (Hu et al. 2015). Notably, *G. raimondii* possesses a phenotype devoid of any spinnable fibre qualities (Hu et al. 2015). The natural union of complementary homeologs from these two divergent genomes in a common allopolyploid nucleus, in combination with intense human-mediated artificial selection, has seen the development of cotton species whose maximum yield, and fibre strength, length and fineness consistently surpass those of their parental species (Paterson 2005). New Zealand's cotton exports, a sector not widely recognised, exceeded NZD$6 million in 2018 (Figure 1B), mainly in the form of woven cotton fabrics, as opposed to combed or carded raw cotton fibres (Stats NZ 2019).

As yet another example of allopolyploidy in agriculture, many areas of New Zealand, including Nelson, Hawke's Bay and Central Otago, provide an ideal climate for the cultivation of apples (*Malus* species); a purportedly allopolyploid crop (Chevreau and Laurens 1987; Tatum et al. 2005) that has been grown on New Zealand soil since European settlement and exported since the 1880s. Several new apple varieties have been developed in New Zealand, including Royal Gala, Jazz and Lemonade. The sequencing of the domestic apple genome in 2010 gave unprecedented insight into the evolutionary history of the *Malus* genus, postulating a polyploidy event that occurred over 50 million years ago in the ancestral lineage from which another pipfruit, pears (*Pyrus* species), have also originated. This duplication may have enabled the expansion of key gene families, such as those implicated in carbohydrate metabolism, a gene family over-represented in apples (Velasco et al. 2010), and those responsible for the commonly observed red fruit-skin pigmentation (Chagné et al. 2013). However, contrary to earlier studies, more recent genomic analysis is favouring the view that apples may have originated through autopolyploidy (Velasco et al. 2010). The evolutionary history of pears is also unclear, with some studies supporting an allopolyploid origin (Evans et al. 2008), while others support the autopolyploidy hypothesis (Li et al. 2019). The latter study drew their conclusions from the observed unbiased subgenome evolution in the pear genome; a common trend among paleo-autopolyploids, but one that has also been observed in some paleo-allopolyploid lineages (Sun et al. 2017; Li et al. 2019).

Introduced allopolyploid taxa of importance to the New Zealand economy are also prevalent well beyond the agricultural and horticultural sectors. *Saccharomyces*, the fungal genus of the model organism *S. cerevisiae* (brewer's and baker's yeast), is instrumental in brewery practices worldwide. *Saccharomyces cerevisiae* is the descendent of an ancient allopolyploidisation event, although genetic analyses have revealed extensive diploidisation, resulting in only minor retention of duplicated gene copies, most commonly those associated with beneficial novel properties (Wolfe 2001; Pfliegler et al. 2012; Marcet-Houben and Gabaldón 2015; Wolfe 2015). Notably, *S. cerevisiae*, an ale

(top fermenting) yeast, is also a parent of the allopolyploid lager (bottom fermenting) yeast, *S. pastorianus* (Casaregola et al. 2001; Lodolo et al. 2008). However, it is the genomic contribution of its other parental species, *S. eubayanus*, that granted *S. pastorianus* the specific sugar and sulphite metabolism changes necessary for its lager-brewing domestication (Libkind et al. 2011).

The establishment of New Zealand vineyards also began in the early 1800s. Domesticated allopolyploid *S. cerevisiae* of course plays a central role in wine making too, dominating the initial (Mangado et al. 2018) and late (Bagheri et al. 2017) stages of fermentation due to its high sugar processing rates and ethanol tolerance. Recent evidence suggests that *S. cerevisiae* not only dominates this ecosystem, but also directly influences the other constituent species (Bagheri et al. 2017). Seemingly in parallel with the allopolyploidy observed in brewing yeasts, strains of *Dekkera bruxellensis*, a prominent industrial wine fermentation contaminant, also appear to be allopolyploids (Borneman et al. 2014). The separation of the *Saccharomyces* and *Dekkera* lineages occurred at least 200 million years ago, preceding the ancient allopolyploidisation event in the *S. cerevisiae* lineage (Schifferdecker et al. 2014). Perhaps this genotypic convergence has been driven by the biological pressure from an evolutionary arms race between these two fermentative yeasts. A purported link exists between the ploidy level of *Saccharomyces* species and culture productivity (Albertin and Marullo 2012), and is a plausible explanation for the strong presence of allopolyploids among brewing yeasts, particularly when considered in conjunction with hybrid vigour. Subsequently, it is in the interest of New Zealand brewers to explore the existence of novel local allopolyploid strains that may provide a unique new edge to New Zealand's longstanding brewing and viticultural practices.

Allopolyploidisation post-introduction

Allopolyploid taxa whose hybridisation occurred following their introduction to New Zealand are less common than the examples above where allopolyploidisation preceded their arrival in New Zealand. One case of hybridisation subsequent to arrival, which has benefited the brewing industry, is triploid hops (*Humulus lupulus*). The New Zealand brewing industry pioneered the development of triploid hops, first releasing a commercial variety in 1972 in response to the growing global demand for a seedless phenotype. Today, 24 New Zealand-developed triploid cultivars are available (New Zealand Hops Ltd. 2018), distinguishable beyond their seedlessness by their superior growth and lushness relative to diploid cultivars (Trojak-Goluch and Skomra 2018). Although strictly autopolyploids, triploid hops are formed by controlled crosses between genetically extremely distinct cultivars; most often between a tetraploid female and diploid male, and thus nonetheless contain diverse parental genomes (Beatson and Brewer 1994; Beatson et al. 2003).

New Zealand's large population of introduced flora provides ample opportunity for naturally-occurring allopolyploidy events among taxa following their arrival. Given the purported role of polyploidy (Pandit et al. 2011; te Beest et al. 2012; Baduel et al. 2018) and hybridisation (Ellstrand and Schierenbeck 2000; Hovick and Whitney 2014; Gaskin 2016) in plant invasions, it will be important to remain vigilant to such occurrences. Certainly, allopolyploidy appears to have been central to the invasiveness of the aggressive hawkweed (*Pilosella officinarum*), which was accidentally introduced over 100 years ago

and has since become the bane of New Zealand's high country (Morgan-Richards et al. 2004; Morgan-Richards et al. 2009). Data suggests that the hybridisation of *P. officinarum* with a related species (likely *Hieracium praealtum*) has occurred at least three times within New Zealand, generating plants that appear phenotypically similar to *P. officinarum*, but that may act as conduits for increased gene flow between the species and possess the potential to colonise from single seeds, an ability likely to facilitate the rapid spread of an invasive species (Morgan-Richards et al. 2004; Trewick et al. 2004). In contrast to the direct genetic manipulation of triploid hops, allopolyploid hawkweed appears to have arisen as a consequence of human-induced habitat disturbance (Morgan-Richards et al. 2009).

Native flora and fauna

Introduced allopolyploid species have become an integral part of the country: for its industries, its economy and its food-chain. Many of these species are relatively well studied, due to their analogous roles on a global scale and their economic importance (Figure 1). However, New Zealand native allopolyploid species are significantly less well understood than their non-native counterparts. This knowledge gap belies the importance of native allopolyploids to the distinctiveness of the New Zealand biota. Concordant with the global trend (Jiao et al. 2011; Barker et al. 2016), a wealth of native allopolyploid diversity is found in New Zealand's flora, with documented examples across mosses, ferns and angiosperms (flowering plants). Allopolyploidy is possibly largely absent from New Zealand native fauna, even for species with otherwise unusual reproductive strategies (Morgan-Richards et al. 2019). The sole case known to us is an ancient hybridisation inferred in the formation of the mosaic allotriploid, parthenogenetic *Acanthoxyla* stick insects (Buckley et al. 2008; Myers et al. 2013) (Table 1). Triploidy has been documented in the New Zealand freshwater snail *Potamopyrgus antipodarium* (Soper et al. 2013), although its origin appears to lie in autopolyploidy (Neiman et al. 2011).

Mosses

Mosses are considered to be one of the earliest forms of land plant that evolved in adaptation to the higher CO_2 levels present from the early Paleozoic Era (Shaw et al. 2011). New Zealand is home to more than 500 moss species, approximately 20% of which are considered endemic (Glenny et al. 2011). Although mosses are haploid-dominant organisms (Graham and Wilcox 2000), researchers have identified two species of the New Zealand native *Sphagnum* mosses (peat moss) that lack haploid gametophytes (Karlin et al. 2009; Karlin and Smouse 2017). *Sphagnum australe* and *S. falcatulum* are found throughout the South Island only as allodiploid or allotriploid cytotypes, and it further appears that the allotriploids are the predominant plant form, perhaps due to a competitive advantage over their allodiploid counterparts that allows them to inhabit a broader ecological niche. Hybridisation and polyploidy are relatively common within the *Sphagnum* genus (Ricca et al. 2008; Karlin et al. 2014), as is interploidal hybridisation (i.e. hybridisation between different ploidy levels) relative to angiosperms (Flatberg et al. 2006; Karlin et al. 2009; Meleshko et al. 2018). For *S. australe* and *S. falcatulum* allotriploids, one of the parental species is their respective allodiploid species (or ancestor

thereof). Both taxa also display intersectional allopolyploidy, with *S. australe* appearing to have parental species from the sections *Sphagnum* and *Rigida*, and *S. falcatulum* from the sections *Subsecunda* and *Cuspidata* (Karlin et al. 2009).

Ferns

Ferns are another prominent feature of the damp understory of forested areas throughout New Zealand; the number of distinct species is unusually high for a temperate country (approximately 200, with 44% endemicity) (Schönberger et al. 2018). Among vascular plants globally, ferns demonstrate the highest frequency of polyploid species, with polyploidy implicated in up to 31% of speciation events (Wood et al. 2009). Consequently, ferns have long been used as a model system for studying polyploidy and its genetic repercussions (DeMaggio et al. 1971). The genus *Asplenium* (spleenworts) contains more than 700 species worldwide. Its Austral group, a polyploid complex centred in New Zealand (Brownsey 1977; Shepherd, Holland, et al. 2008), is unique when compared with European and North American groups of the same genus due to its absence of diploid organisms (Dawson et al. 2000; Perrie and Brownsey 2005). Seven tetraploid and eight octoploid *Asplenium* species are found in New Zealand, with allopolyploidy implicated in the origin of seven of the octoploid species (Shepherd, Perrie, et al. 2008), mirroring the high frequency of allopolyploidy found in *Asplenium* worldwide (Lovis 1978; Reichstein 1981; Wagner et al. 1993; Schneider et al. 2017). Interestingly, one of the tetraploids is thought to be the parent of five of the octoploids, suggesting that this single species has played a central role in the evolution of this fern complex (Shepherd, Perrie, et al. 2008). Despite extensive sympatry and having the same parental species, the allopolyploids *A. cimmeriorum* and *A. gracillimum* have each evolved independently at least twice, with the four lineages reproductively isolated (Perrie et al. 2010), thus emphasising the importance of allopolyploidy in the generation of biological diversity. It is possible that the octoploids *A. shuttleworthianum* and *A. northlandicum* have extinct parental species; nuclear sequences of the *LFY* gene obtained from the allopolyploids could not be identified among the tetraploid taxa analysed by Shepherd, Perrie, et al. (2008). However, it is also possible that these sequences were inherited from unsampled Pacific species.

The genus *Polysticum* (shield ferns) contains an allopolyploid complex located in New Zealand, and provided the first evidence-based demonstration of allopolyploidy in the New Zealand fern flora (Perrie et al. 2003). Initially considered to be a single species with very high levels of morphological variability (Allan 1961; Brownsey and Smith-Dodsworth 1989), *P. richardii* is now recognised as an allopolyploid complex containing four unique evolutionary lineages: two allooctoploid subspecies and two purportedly-parental tetraploid species. Allooctoploid *P. neozelandicum* is widely distributed throughout New Zealand, including on several outer islands. However, the individual distributions of its two subspecies do not overlap. Notably, allooctoploid *P. neozelandicum* subsp. *zerophyllum* has a wider geographical distribution than either of its tetraploid parental species, despite possessing relatively lower intra-lineage genetic variability (Perrie et al. 2003). It is possible that the New Zealand *Polystichum* complex represents a further example of allopolyploidy-mediated niche expansion and hybrid vigour. However, a wider study of the geographic ranges of all New Zealand ferns and lycophytes has found no obvious correlation between range sizes and polyploidy (Mountier et al. 2018).

Angiosperms

Angiosperms are the largest group of vascular plants, numbering over 300,000 species worldwide. Phylogenomic evidence has implicated an ancient whole genome duplication event in the ancestry of all extant angiosperm lineages (Jiao et al. 2011). Allopolyploidy, specifically, has been instrumental in the evolution of many of these (Osabe et al. 2012; Lyu 2016), and importantly, the link between New Zealand and allopolyploid angiosperms extends well beyond those species that make major contributions to the economy via crop production. Among New Zealand native angiosperms, there is an excess of species with even haploid numbers (i.e. the number of chromosomes within a single set), and an excess of those with haploid numbers greater than 10–14. These two karyotypic features are strong indicators of polyploid ancestry and substantiate the key role of polyploidy in the evolution of the New Zealand flora (Murray and de Lange 2011).

Perhaps the best-studied example of native allopolyploidy is *Pachycladon* (New Zealand rockcress), a genus in the Brassicaceae that is related to the model organism *Arabidopsis thaliana*. The number of species in the *Pachycladon* genus increased from one to eleven in the last million years following an allopolyploidisation event during the Pleistocene (Joly et al. 2009; Mandáková et al. 2010). Today, ten *Pachycladon* species are endemic to the South Island, and exhibit considerable morphological and habitat diversity in mainly alpine environments. Phylogenetic modelling and molecular studies have provided strong evidence for an adaptive radiation in the New Zealand group, despite the absence of clear phenotypic adaptations (Joly et al. 2014), with allopolyploidy and subsequent genomic restructuring within the founding species being the purported driving force (Joly et al. 2009; Mandáková et al. 2010). An adaptive radiation in *Pachycladon* is an impressive example of hybridisation between significantly diverged parental species. The level of divergence between the parents prior to their hybridisation is thought to be greater than that of allopolyploids in *Gossypium*, a genus that has long boasted one of the highest known parental genetic divergences among allopolyploids (Joly et al. 2009).

Polyploidy is also frequent among the more than 200 species of the globally-widespread genus *Plantago* (Murray et al. 2010). Commonly known as plantains, *Plantago* primarily comprises small herbaceous plants and shrubs. The eleven species of *Plantago* in New Zealand are not monophyletic and form three distinct lineages that appear to derive from at least three relatively-recent long-distance dispersal events, likely from Australia (Tay et al. 2010). Cytological and molecular studies suggest that most, if not all, New Zealand polyploid *Plantago* are allopolyploids (Ishikawa et al. 2009; Murray et al. 2010). This feature appears to have been central to their evolutionary history and taxonomic complexity, generating six ploidy levels, from diploid to 16-ploid, and forming many sympatric species within the New Zealand group (Rahn 1957; Rattenbury 1957; Groves and Hair 1971; Murray et al. 2010; Meudt 2011).

Looking forward

Allopolyploid species, both native and introduced, are central to many key aspects of life in New Zealand, and thus the benefits they confer to the country and its people are vast, far-reaching and surprisingly underappreciated. Duplicated, hybrid genomes give allopolyploids access to a suite of potential benefits otherwise unavailable to their ancestral

species, assuming they survive the initial genome shock. The prevalence of allopolyploid species throughout New Zealand attests not only to their resilience in overcoming these challenges, but also to their ability to thrive and oftentimes surpass the biological fitness of their parental species. In some cases, such as the allopolyploid hawkweed, this ability is clearly demonstrated through invasiveness. In others, humans have been able to control the distribution of allopolyploid species, and even manipulate their ploidy level. New Zealand demonstrates a high level of polyploidy among its vascular plants: early studies characterised polyploidy in approximately 63% of angiosperms (Hair 1966). This value is concordant with the globally-observed positive correlation between polyploidy and latitude (Stebbins 1984; Brochmann et al. 2004; te Beest et al. 2012), but contrasts with the chromosomal stasis (i.e. the absence of chromosomal change) seen in other Pacific island groups (Murray and de Lange 2011; Stuessy et al. 2014). It has been postulated that the high number of introduced plant species (approximately 50%) among the native biota has generated both the opportunities, and the pressure, for increased hybridisation within New Zealand (Morgan-Richards et al. 2009). New Zealand allopolyploid research would benefit from a dedicated survey of the prevalence of allopolyploid taxa, both native and introduced, in place of the current rough estimates made from the independent prevalence of hybridisation and polyploidy, and the incomplete lists of known allopolyploid cases.

Statistics New Zealand data (2019) suggest that allopolyploids, either directly or indirectly, collectively contribute around NZD$23 billion to the economy per annum (Figure 1A), with this value excluding the largest economic sector, tourism. The need to continue to invest in the development of superior crop strains that deliver higher yields with reduced energy, water and land area requirements is strengthened by the rapidly growing global population: by 2050, the world's population is projected to reach 9.8 billion people; 11.2 billion by the year 2100 (United Nations Department of Public Information 2017). The optimisation of arable crops is imperative to feed this rapidly growing global population, and the interests of New Zealand in pursuing this research rest not only on its own food sufficiency, but also on export opportunities to the rest of the world. Beyond these critical subsistence applications, allopolyploid species play an integral role in the long-established traditions of brewing and viticulture, which also make significant contributions to the New Zealand economy (Figure 1B). These practices could gain a distinctive and competitive edge in both sensory profile and culture productivity through the discovery and implementation of novel native allopolyploid yeast strains.

Questions remain about the role of allopolyploidy in the evolution of New Zealand's indigenous flora. The vast majority of New Zealand native plant species arrived via long distance dispersal (Winkworth et al. 2002; Wallis and Jorge 2018). Were the successful colonisers polyploid when they arrived, or did polyploidy occur subsequently in New Zealand? There is evidence that both situations occurred (Murray and de Lange 2011), but the relative contribution of allo- versus autopolyploidy has not yet been assessed. There are also questions about the timing of allopolyploidy events. Many Northern Hemisphere allopolyploids have Pleistocene origins (Abbott and Brochmann 2003), with the harsh environmental conditions of glacial periods thought to increase both hybridisation (Soltis et al. 2004) and the production of unreduced gametes (Mable 2004), and leading to an increased efficiency of allopolyploid species in colonising newly deglaciated areas

(Comai 2005). The age of many of New Zealand's allopolyploids are unknown but Pleistocene origins have been postulated for New Zealand *Asplenium* (Shepherd, Holland, et al. 2008), *Pachycladon* (Joly et al. 2009), *Lepidium* (Mummenhoff et al. 2004) and *Leptinella* (Himmelreich et al. 2014). The creation of time-calibrated phylogenies for additional taxa has the potential to provide insight into the geological drivers of allopolyploid formation within New Zealand's island context.

Finally, how will these allopolyploids, both of agricultural and environmental importance, adjust to a changing climate? Will increased plasticity from multiple genomes mean that allopolyploid species are more resistant to climate change? Or will allopolyploids tend to be more transient on the landscape as many early writers originally proposed? Ten years ago, Morgan-Richards et al. (2009) concluded their review of genetic analyses of hybridisation in New Zealand with an acknowledgement that hybridisation is a continuing focus of evolutionary biology. Today, it feels more important than ever to prioritise local allopolyploidy research to continue to improve the crops on which the economy depends, as well as to understand the processes that have generated the unique national biota of Aotearoa New Zealand.

Acknowledgements

We would like to thank Jennifer Tate (Massey University) for her constructive feedback, David Winter (Massey University) for his help during the initial stages of figure development, Thomas Buckley (Manaaki Whenua – Landcare Research) for his advice regarding allopolyploidy among New Zealand's native fauna, and the editor and two anonymous reviewers for their thoughtful and constructive comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Lara D. Shepherd http://orcid.org/0000-0002-7136-0017
Murray P. Cox http://orcid.org/0000-0003-1936-0236

References

Abbott RJ, Brochmann C. 2003. History and evolution of the Arctic flora: in the footsteps of Eric Hultén. Molecular Ecology. 12(2):299–313.

Adams KL, Wendel JF. 2005. Novel patterns of gene expression in polyploid plants. Trends in Genetics. 21(10):539–543.

Ainouche ML, Fortune PM, Salmon A, Parisod C, Grandbastien MA, Fukunaga K, Ricou M, Misset MT. 2008. Hybridization, polyploidy and invasion: lessons from *Spartina* (Poaceae). Biological Invasions. 11(5):1159–1173.

Albertin W, Marullo P. 2012. Polyploidy in fungi: evolution after whole-genome duplication. Proceedings of the Royal Society B: Biological Sciences. 279(1738):2497–2509.

Allan HH. 1961. Flora of New Zealand. Vol. I. Wellington: Government Printer.

Atkinson RG, Cipriani G, Whittaker DJ, Gardner RC. 1997. The allopolyploid origin of kiwifruit, *Actinidia deliciosa* (Actinidiaceae). Plant Systematics and Evolution. 205(1/2):111–124.

Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. 2018. The "polyploid hop": shifting challenges and opportunities over the evolutionary lifespan of genome duplications. Frontiers in Ecology and Evolution. 6(117):117.
Bagheri B, Bauer FF, Setati ME. 2017. The impact of *Saccharomyces cerevisiae* on a wine yeast consortium in natural and inoculated fermentations. Frontiers in Microbiology. 8:1988.
Baranwal VK, Kapoor S, Zehr UB, Mikkilineni V, Tyagi AK. 2012. Heterosis: emerging ideas about hybrid vigour. Journal of Experimental Botany. 63(18):6309–6314.
Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. 2016. On the relative abundance of autopolyploids and allopolyploids. New Phytologist. 210:391–398.
Beatson RA, Brewer VR. 1994. Regional trial evaluation and cultivar selection of triploid hop hybrids. New Zealand Journal of Crop and Horticultural Science. 22(1):1–6.
Beatson RA, Ferguson AR, Weir IE, Graham LT, Ansell KA, Ding H. 2003. Flow cytometric identification of sexually derived polyploids in hop (*Humulus lupulus* L.) and their use in hop breeding. Euphytica. 134(2):189–194.
Bindler G, Plieske J, Bakaher N, Gunduz I, Ivanov N, Van der Hoeven R, Ganal M, Donini P. 2011. A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development. Theoretical and Applied Genetics. 123(2):219–230.
Bird KA, VanBuren R, Puzey JR, Edger PP. 2018. The causes and consequences of subgenome dominance in hybrids and recent polyploids. New Phytologist. 220(1):87–93.
Borneman AR, Zeppel R, Chambers PJ, Curtin CD. 2014. Insights into the *Dekkera bruxellensis* genomic landscape: comparative genomics reveals variations in ploidy and nutrient utilisation potential amongst wine isolates. PLOS Genetics. 10(2):e1004161.
Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen A-C, Elven R. 2004. Polyploidy in Arctic plants. Biological Journal of the Linnean Society. 82(4):521–536.
Brock JL, Caradus JR, Hay MJM. 1989. Fifty years of white clover research in New Zealand. Proceedings of the New Zealand Grassland Association. 50:25–39.
Brooking T. 2006. Pasture, present and future – a brief history of pastoralism in New Zealand. Wellington: Ministry of Agriculture and Forestry.
Brownsey PJ. 1977. A taxonomic revision of the New Zealand species of *Asplenium*. New Zealand Journal of Botany. 15(1):39–86.
Brownsey PJ, Smith-Dodsworth JC. 1989. New Zealand ferns and allied plants. Auckland: David Bateman Ltd.
Buckley TR, Attanayake D, Park D, Ravindran S, Jewell TR, Normark BB. 2008. Investigating hybridization in the parthenogenetic New Zealand stick insect *Acanthoxyla* (Phasmatodea) using single-copy nuclear loci. Molecular Phylogenetics and Evolution. 48(1):335–349.
Campbell MA, Tapper BA, Simpson WR, Johnson RD, Mace W, Ram A, Lukito Y, Dupont P-Y, Johnson LJ, Scott DB, et al. 2017. *Epichloë hybrida*, sp. nov., an emerging model system for investigating fungal allopolyploidy. Mycologia. 109(5):715–729.
Casaregola S, Nguyen HV, Lapathitis G, Kotyk A, Gaillardin C. 2001. Analysis of the constitution of the beer yeast genome by PCR, sequencing and subtelomeric sequence hybridization. International Journal of Systematic and Evolutionary Microbiology. 51(4):1607–1618.
Chagné D, Lin-Wang K, Espley RV, Volz RK, How NM, Rouse S, Brendolise C, Carlisle CM, Kumar S, De Silva N, et al. 2013. An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. Plant Physiology. 161(1):225–239.
Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al. 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science. 345(6199):950.
Charlton D. 2008. Pastures – clovers. [accessed 2019 May 19]. http://www.TeAra.govt.nz/en/pastures/page-7.
Chester M, Gallagher JP, Symonds VV, da Silva AV C, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE. 2012. Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). Proceedings of the National Academy of Sciences of the United States of America. 109(4):1176–1181.

Chevreau E, Laurens F. 1987. The pattern of inheritance in apple (*Malus X domestica* Borkh.): further results from leaf isozyme analysis. Theoretical and Applied Genetics. 75(1):90–95.
Comai L. 2005. The advantages and disadvantages of being polyploid. Nature Reviews Genetics. 6 (11):836–846.
Corredor E, Díez M, Shepherd K, Naranjo T. 2005. The positioning of rye homologous chromosomes added to wheat through the cell cycle in somatic cells untreated and treated with colchicine. Cytogenetic and Genome Research. 109(1–3):112–119.
Cox MP, Dong T, Shen G, Dalvi Y, Scott DB, Ganley ARD. 2014. An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. PLOS Genetics. 10(3): e1004180.
Czesnick H, Lenhard M. 2015. Size control in plants–lessons from leaves and flowers. Cold Spring Harbor Perspectives in Biology. 7(8):a019190.
Darwin C. 1859. On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life. London: John Murray.
Dawson MI, Brownsey PJ, Lovis JD. 2000. Index of chromosome numbers of indigenous New Zealand pteridophytes. New Zealand Journal of Botany. 38(1):25–46.
Del Pozo JC, Ramirez-Parra E. 2014. Deciphering the molecular bases for drought tolerance in *Arabidopsis* autotetraploids. Plant, Cell & Environment. 37:2722–2737.
DeMaggio AE, Wetmore RH, Hannaford JE, Stetler DA, Raghavan V. 1971. Ferns as a model system for studying polyploidy and gene dosage effects. BioScience. 21(7):313–316.
Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, Yuan Y, Bewick AJ, Ji L, Platts AE, Bowman MJ, et al. 2017. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. The Plant Cell. 29(9):2150–2167.
Ellstrand NC, Schierenbeck KA. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? Proceedings of the National Academy of Sciences of the United States of America. 97 (13):7043–7050.
Evans KM, Govan CL, Fernández-Fernández F. 2008. A new gene for resistance to *Dysaphis pyri* in pear and identification of flanking microsatellite markers. Genome. 51(12):1026–1031.
Flagel LE, Wendel JF. 2009. Gene duplication and evolutionary novelty in plants. New Phytologist. 183(3):557–564.
Flatberg KI, Thingsgaard K, Såstad SM. 2006. Interploidal gene flow and introgression in bryophytes: *Sphagnum girgensohnii* × *S. russowii*, a case of spontaneous neotriploidy. Journal of Bryology. 28(1):27–37.
Fujimoto R, Uezono K, Ishikura S, Osabe K, Peacock WJ, Dennis ES. 2018. Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. Breeding Science. 68(2):145–158.
Gaskin JF. 2016. The role of hybridization in facilitating tree invasion. AoB Plants. 9(1):plw079.
Gaut BS, d'Ennequin MLT, Peek AS, Sawkins MC. 2000. Maize as a model for the evolution of plant nuclear genomes. Proceedings of the National Academy of Sciences USA. 97(13):7008–7015.
Glenny D, Fife AJ, Brownsey PJ, Renner MA, Braggins JE, Beever JE, Hitchmough R. 2011. Threatened and uncommon bryophytes of New Zealand (2010 revision). New Zealand Journal of Botany. 49(2):305–327.
Gordon JL, Byrne KP, Wolfe KH. 2009. Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. PLOS Genetics. 5(5):e1000485.
Graham LK, Wilcox LW. 2000. The origin of alternation of generations in land plants: a focus on matrotrophy and hexose transport. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 355(1398):757–767.
Griffiths AG, Moraga R, Tausen M, Gupta V, Bilton TP, Campbell MA, Ashby RL, Nagy I, Khan A, Larking A, et al. 2019. Breaking free: the genomics of allopolyploidy-facilitated niche expansion in white clover. The Plant Cell. 31(7):1466–1487.
Groves BE, Hair JB. 1971. Contributions to a chromosome atlas of the New Zealand flora: 15 miscellaneous families. New Zealand Journal of Botany. 9(4):569–575.

Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li W-H. 2003. Role of duplicate genes in genetic robustness against null mutations. Nature. 421(6918):63–66.
Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chem X, Wang X-J, Chen ZJ. 2009. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. Proceedings of the National Academy of Sciences of the United States of America. 106(42):17835–17840.
Hair JB. 1966. Biosystematics of the New Zealand flora, 1945–1964. New Zealand Journal of Botany. 4(4):559–595.
Harris SL, Auldist MJ, Clark DA, Jansen EB. 1998. Effects of white clover content in the diet on herbage intake, milk production and milk composition of New Zealand dairy cows housed indoors. The Journal of Dairy Research. 65(3):389–400.
Harvey AC, Fjelldal PG, Solberg MF, Hansen T, Glover KA. 2017. Ploidy elicits a whole-genome dosage effect: growth of triploid Atlantic salmon is linked to the genetic origin of the second maternal chromosome set. BMC Genetics. 18(1):34.
Hedrick PW. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology. 22(18):4606–4618.
Heslop-Harrison JS. 1990. Gene expression and parental dominance in hybrid plants. Development. 108:21–28.
Himmelreich S, Breitwieser I, Oberprieler C. 2014. Phylogenetic relationships in the extreme polyploid complex of the New Zealand genus *Leptinella* (Compositae: Anthemideae) based on AFLP data. Taxon. 63(4):883–898.
Hinkle AE. 2004. The distribution of a male sterile form of ti (*Cordyline fruticosa*) in Polynesia: a case of human selection? Journal of the Polynesian Society. 113(3):263–290.
Hovick SM, Whitney KD. 2014. Hybridisation is associated with increased fecundity and size in invasive taxa: meta-analytic support for the hybridisation-invasion hypothesis. Ecology Letters. 17(11):1464–1477.
Hu G, Wendel JF, Koh J, Yoo MJ, Chen S. 2015. Gene-expression novelty in allopolyploid cotton: a proteomic perspective. Genetics. 200(1):91–104.
Ishikawa N, Yokoyama J, Tsukaya H. 2009. Molecular evidence of reticulate evolution in the subgenus *Plantago* (Plantaginaceae). American Journal of Botany. 96(9):1627–1635.
Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature. 473 (7345):97–100.
Jiao W, Yuan J, Jiang S, Liu Y, Wang L, Liu M, Zheng D, Ye W, Wang X, Chen ZJ. 2018. Asymmetrical changes of gene expression, small RNAs and chromatin in two resynthesized wheat allotetraploids. Plant Journal. 93(5):828–842.
Johnson LJ, de Bonth ACM, Briggs LR, Caradus JR, Finch SC, Fleetwood DJ, Fletcher LR, Hume DE, Johnson RD, Popay AJ, et al. 2013. The exploitation of Epichloae endophytes for agricultural benefit. Fungal Diversity. 60:171–188.
Joly S, Heenan PB, Lockhart PJ. 2009. A Pleistocene inter-tribal allopolyploidization event precedes the species radiation of *Pachycladon* (Brassicaceae) in New Zealand. Molecular Phylogenetics and Evolution. 51(2):365–372.
Joly S, Heenan PB, Lockhart PJ. 2014. Species radiation by niche shifts in New Zealand's rockcresses (*Pachycladon*, Brassicaceae). Systematic Biology. 63(2):192–202.
Jung Y, Kawaura K, Kishii M, Sakuma S, Ogihara Y. 2015. Comparison of genome-wide gene expression patterns in the seedlings of nascent allohexaploid wheats produced by two combinations of hybrids. Genes and Genetic Systems. 90(2):79–88.
Karlin EF, Boles SB, Ricca M, Temsch EM, Greilhuber J, Shaw AJ. 2009. Three-genome mosses: complex double allopolyploid origins for triploid gametophytes in *Sphagnum*. Molecular Ecology. 18(7):1439–1454.
Karlin EF, Smouse PE. 2017. Allo-allo-triploid *Sphagnum* × *falcatulum*: single individuals contain most of the Holantarctic diversity for ancestrally indicative markers. Annals of Botany. 120 (2):221–231.

Karlin EF, Temsch EM, Bizuru E, Marino J, Boles SB, Devos N, Shaw AJ. 2014. Invisible in plain sight: recurrent double allopolyploidy in the African *Sphagnum×planifolium* (Sphagnaceae). The Bryologist. 117(2):187–201.
Kim S-T, Sultan SE, Donoghue MJ. 2008. Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. Proceedings of the National Academy of Sciences of the United States of America. 105(34):12370–12375.
Lashermes P, Combes M-C, Hueber Y, Severac D, Dereeper A. 2014. Genome rearrangements derived from homoeologous recombination following allopolyploidy speciation in coffee. The Plant Journal. 78(4):674–685.
Leach H, Stowe C. 2005. Oceanic arboriculture at the margins: the case of the Karaka (*Corynocarpus laevigatus*) in Aotearoa. The Journal of the Polynesian Society. 114(1):7–27.
Leitch IJ, Bennett MD. 2004. Genome downsizing in polyploid plants. Biological Journal of the Linnean Society. 82(4):651–663.
Leuchtmann A, Schardl CL, Penny D, Chung KR, Siegel MR. 1997. Coevolution by common descent of fungal symbionts (*Epichloë* spp.) and grass hosts. Molecular Biology and Evolution. 14(2):133–143.
Leuchtmann A, Young CA, Stewart AV, Simpson WR, Hume DE, Scott B. 2019. *Epichloe novae-zelandiae*, a new endophyte from the endemic New Zealand grass *Poa matthewsii*. New Zealand Journal of Botany. 1–18.
Li Q, Qiao X, Yin H, Zhou Y, Dong H, Qi K, Li L, Zhang S. 2019. Unbiased subgenome evolution following a recent whole-genome duplication in pear (*Pyrus bretschneideri* Rehd.). Horticulture Research. 6(1):34.
Libkind D, Hittinger CT, Valério E, Gonçalves C, Dover J, Johnston M, Gonçalves P, Sampaio JP. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proceedings of the National Academy of Sciences of the United States of America. 108 (35):14539–14544.
Lim KY, Soltis DE, Soltis PS, Tate J, Matyasek R, Srubarova H, Kovarik A, Pires JC, Xiong Z, Leitch AR. 2008. Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae). PLOS ONE. 3(10):e3353.
Liu Q, Lin L, Zhou X, Peterson PM, Wen J. 2017. Unraveling the evolutionary dynamics of ancient and recent polyploidization events in *Avena* (Poaceae). Scientific Reports. 7:41944.
Lodolo EJ, Kock JLF, Axcell BC, Brooks M. 2008. The yeast *Saccharomyces cerevisiae* – the main character in beer brewing. FEMS Yeast Research. 8(7):1018–1036.
López-Caamal A, Tovar-Sánchez E. 2014. Genetic, morphological, and chemical patterns of plant hybridization. Revista Chilena de Historia Natural. 87(1):16.
Lovis JD. 1978. Evolutionary patterns and processes in ferns. Advances in Botanical Research. 4:229–415.
Lyu J. 2016. Crop evolution: after allopolyploidization. Nature Plants. 2:16156.
Mable BK. 2004. 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. Biological Journal of the Linnean Society. 82(4):453–466.
Mandáková T, Heenan PB, Lysak MA. 2010. Island species radiation and karyotypic stasis in *Pachycladon* allopolyploids. BMC Evolutionary Biology. 10:367.
Mangado A, Morales P, Gonzalez R, Tronchoni J. 2018. Evolution of a yeast with industrial background under winemaking conditions leads to diploidization and chromosomal copy number variation. Frontiers in Microbiology. 9:1816.
Marcet-Houben M, Gabaldón T. 2015. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. PLOS Biology. 13: e1002220.
Matos I, Machado MP, Schartl M, Coelho MM. 2019. Allele-specific expression variation at different ploidy levels in *Squalius alburnoides*. Scientific Reports. 9(1):3688.
McClintock B. 1984. The significance of responses of the genome to challenge. Science. 226:792–801.
Melaragno JE, Mehrotra B, Coleman AW. 1993. Relationship between endopolyploidy and cell size in epidermal tissue of *Arabidopsis*. The Plant Cell. 5(11):1661–1668.

Meleshko O, Stenøien HK, Speed JDM, Flatberg KI, Kyrkjeeide MO, Hassel K. 2018. Is interspecific gene flow and speciation in peatmosses (*Sphagnum*) constrained by phylogenetic relationship and life-history traits? Lindbergia. 41(1):linbg.01107.
Meudt HM. 2011. Amplified fragment length polymorphism data reveal a history of auto- and allo-polyploidy in New Zealand endemic species of *Plantago* (Plantaginaceae): new perspectives on a taxonomically challenging group. International Journal of Plant Sciences. 172(2):220–237.
Ministry of Business, Innovation & Employment. 2018. Research, science and innovation system performance report: ryegrass endophytes. [accessed 2019 June 20]. https://www.mbie.govt.nz/assets/7693f53535/research-science-and-innovation-system-performance-report-2018.pdf.
Morgan-Richards M, Langton-Myers SS, Trewick SA. 2019. Loss and gain of sexual reproduction in the same stick insect. Molecular Ecology. 28(17):3929–3941.
Morgan-Richards M, Smissen RD, Shepherd LD, Wallis GP, Hayward JJ, Chan CH, Chambers GK, Chapman HM. 2009. A review of genetic analyses of hybridisation in New Zealand. Journal of the Royal Society of New Zealand. 39(1):15–34.
Morgan-Richards M, Trewick SA, Chapman HM, Krahulcova A. 2004. Interspecific hybridization among *Hieracium* species in New Zealand: evidence from flow cytometry. Heredity. 93(1):34–42.
Mountier CF, Case BS, Perrie L, Brownsey P, Paterson AM, Curran TJ, Buckley HL. 2018. Patterns of range size in New Zealand ferns and lycophytes. New Zealand Journal of Ecology. 42(2):248–261.
Mummenhoff K, Linder P, Friesen N, Bowman JL, Lee J-Y, Franzke A. 2004. Molecular evidence for bicontinental hybridogenous genomic constitution in *Lepidium* sensu stricto (Brassicaceae) species from Australia and New Zealand. American Journal of Botany. 91(2):254–261.
Murray BG, Datson PM, Lai ELY, Sheath KM, Cameron EK. 2004. Polyploidy, hybridization and evolution in *Pratia* (Campanulaceae). New Zealand Journal of Botany. 42(5):905–920.
Murray BG, de Lange PJ. 2011. Chromosomes and evolution in New Zealand endemic angiosperms and gymnosperms. In: Bramwell D, Caujapé-Castells J, editors. The Biology of Island Floras. Cambridge: Cambridge University Press; p. 265–283.
Murray BG, Meudt HM, Tay ML, Garnock-Jones PJ. 2010. New chromosome counts in New Zealand species of *Plantago* (Plantaginaceae). New Zealand Journal of Botany. 48(3–4):197–204.
Myers SS, Trewick SA, Morgan-Richards M. 2013. Multiple lines of evidence suggest mosaic polyploidy in the hybrid parthenogenetic stick insect lineage *Acanthoxyla*. Insect Conservation and Diversity. 6(4):537–548.
Neiman M, Paczesniak D, Soper DM, Baldwin AT, Hehman G. 2011. Wide variation in ploidy level and genome size in a New Zealand freshwater snail with coexisting sexual and asexual lineages. Evolution. 65(11):3202–3216.
Nemorin A, Abraham K, David J, Arnau G. 2012. Inheritance pattern of tetraploid *Dioscorea alata* and evidence of double reduction using microsatellite marker segregation analysis. Molecular Breeding. 30(4):1657–1667.
New Zealand Hops Ltd. 2018. Harvested to perfection. [accessed 2019 September 21]. https://nz.coop/harvested-to-perfection/.
Osabe K, Kawanabe T, Sasaki T, Ishikawa R, Okazaki K, Dennis ES, Kazama T, Fujimoto R. 2012. Multiple mechanisms and challenges for the application of allopolyploidy in plants. International Journal of Molecular Sciences. 13(7):8696–8721.
Ozkan H, Arumuganathan K, Tuna M. 2003. Nonadditive changes in genome size during allopolyploidization in the wheat (*Aegilops-Triticum*) group. Journal of Heredity. 94(3):260–264.
Ozkan H, Feldman M. 2009. Rapid cytological diploidization in newly formed allopolyploids of the wheat (*Aegilops-Triticum*) group. Genome. 52(11):926–934.
Pandit MK, Pocock MJO, Kunin WE. 2011. Ploidy influences rarity and invasiveness in plants. Journal of Ecology. 99(5):1108–1115.
Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien MA, Ainouche M. 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytologist. 184(4):1003–1015.

Paterson AH. 2005. Polyploidy, evolutionary opportunity, and crop adaptation. Genetica. 123 (1):191–196.
Peden R. 2008. Farming in the economy. [accessed 2019 April 2]. http://www.TeAra.govt.nz/en/farming-in-the-economy/print.
Perrie LR, Brownsey PJ. 2005. Insights into the biogeography and polyploid evolution of New Zealand *Asplenium* from chloroplast DNA sequence data. American Fern Journal. 95(1):1–22.
Perrie LR, Brownsey PJ, Lockhart PJ, Large MF. 2003. Evidence for an allopolyploid complex in New Zealand *Polystichum* (Dryopteridaceae). New Zealand Journal of Botany. 41(2):189–215.
Perrie LR, Shepherd LD, De Lange PJ, Brownsey PJ. 2010. Parallel polyploid speciation: distinct sympatric gene-pools of recurrently derived allo-octoploid *Asplenium* ferns. Molecular Ecology. 19(14):2916–2932.
Pfliegler WP, Antunovics Z, Sipiczki M. 2012. Double sterility barrier between *Saccharomyces* species and its breakdown in allopolyploid hybrids by chromosome loss. FEMS Yeast Research. 12(6):703–718.
Qi X, Wang H, Song A, Jiang J, Chen S, Chen F. 2018. Genomic and transcriptomic alterations following intergeneric hybridization and polyploidization in the *Chrysanthemum nankingense* × *Tanacetum vulgare* hybrid and allopolyploid (Asteraceae). Horticulture Research. 5:5.
Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome Biology. 20(1):38.
Qin Q, Lai Z, Cao L, Xiao Q, Wang Y, Liu S. 2016. Rapid genomic changes in allopolyploids of *Carassius auratus red var.* (♀) × *Megalobrama amblycephala* (♂). Scientific Reports. 6:34417.
Rahn K. 1957. Chromosome numbers in *Plantago*. Botanisk Tidsskrift. 53:369–378.
Rattenbury JA. 1957. Chromosome numbers in New Zealand angiosperms. Transactions of the Royal Society of New Zealand. 84:936–938.
Reichstein T. 1981. Hybrids in European Aspleniaceae (Pteridophyta) in Botanica Helvetica. Berichte der Schweizerischen Botanischen Gesellschaft = Bulletin de la Societe botanique suisse. 91:89–139.
Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J. 2018. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. Molecular Plant. 11(3):414–428.
Ricca M, Beecher FW, Boles SB, Temsch E, Greilhuber J, Karlin EF, Shaw AJ. 2008. Cytotype variation and allopolyploidy in North American species of the *Sphagnum subsecundum* complex (Sphagnaceae). American Journal of Botany. 95(12):1606–1620.
Rieseberg LH, Archer MA, Wayne RK. 1999. Transgressive segregation, adaptation and speciation. Heredity. 83(4):363–372.
Robinson DO, Coate JE, Singh A, Hong L, Bush M, Doyle JJ, Roeder AHK. 2018. Ploidy and size at multiple scales in the *Arabidopsis* sepal. The Plant Cell. 30(10):2308–2329.
Rolston MP, Stewart AV, Latch GCM, Hume DE. 2002. Endophytes in New Zealand grass seeds: occurrence and implications for conservation of grass species. New Zealand Journal of Botany. 40(3):365–372.
Roullier C, Duputié A, Wennekes P, Benoit L, Fernández Bringas VM, Rossel G, Tay D, McKey D, Lebot V. 2013. Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). PLOS ONE. 8(5):e62707.
Saikkonen K, Wäli P, Helander M, Faeth SH. 2004. Evolution of endophyte-plant symbioses. Trends in Plant Science. 9(6):275–280.
Saikkonen K, Young CA, Helander M, Schardl CL. 2016. Endophytic *Epichloë* species and their grass hosts: from evolution to applications. Plant Molecular Biology. 90(6):665–675.
Sattler MC, Carvalho CR, Clarindo WR. 2016. The polyploidy and its key role in plant breeding. Planta. 243(2):281–296.
Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature. 440(7082):341–345.
Schardl CL, Florea S, Pan J, Nagabhyru P, Bec S, Calie PJ. 2013. The epichloae: alkaloid diversity and roles in symbiosis with grasses. Current Opinion in Plant Biology. 16(4):480–488.

Schardl CL, Leuchtmann A, Tsai HF, Collett MA, Watt DM, Scott DB. 1994. Origin of a fungal symbiont of perennial ryegrass by interspecific hybridization of a mutualist with the ryegrass choke pathogen, *Epichloë typhina*. Genetics. 136(4):1307–1317.
Schifferdecker AJ, Dashko S, Ishchuk OP, Piškur J. 2014. The wine and beer yeast *Dekkera bruxellensis*. Yeast. 31(9):323–332.
Schneider H, Liu H-M, Chang Y-F, Ohlsen D, Perrie LR, Shepherd L, Kessler M, Karger DN, Hennequin S, Marquardt J, et al. 2017. Neo- and Paleopolyploidy contribute to the species diversity of *Asplenium*—the most species-rich genus of ferns. Journal of Systematics and Evolution. 55 (4):353–364.
Schönberger I, Wilton AD, Brownsey PJ, Perrie L, Boardman KF, Breitwieser I, Cochrane M, de Pauw B, Fife AJ, Ford KA, et al. 2018. Checklist of the New Zealand flora – ferns and lycophytes [dataset]. Lincoln: Manaaki Whenua-Landcare Research.
Segraves KA, Thompson JN. 1999. Plant polyploidy and pollination: floral traits and insect visits to diploid and tetraploid *Heuchera grossulariifolia*. Evolution. 53(4):1114–1127.
Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA. 2014. Gene silencing via DNA methylation in naturally occurring *Tragopogon miscellus* (Asteraceae) allopolyploids. BMC Genomics. 15(1):701.
Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, Fukui A, Hikosaka A, Suzuki A, Kondo M, et al. 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. Nature. 538:336.
Shaw AJ, Szövényi P, Shaw B. 2011. Bryophyte diversity and evolution: windows into the early evolution of land plants. American Journal of Botany. 98(3):352–369.
Shepherd LD, Holland BR, Perrie LR. 2008. Conflict amongst chloroplast DNA sequences obscures the phylogeny of a group of *Asplenium* ferns. Molecular Phylogenetics and Evolution. 48(1):176–187.
Shepherd LD, Perrie LR, Brownsey PJ. 2008. Low-copy nuclear DNA sequences reveal a predominance of allopolyploids in a New Zealand *Asplenium* fern complex. Molecular Phylogenetics and Evolution. 49(1):240–248.
Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, Havird JC. 2018. Cytonuclear integration and co-evolution. Nature Reviews Genetics. 19(10):635–648.
Soltis DE, Soltis PS. 1999. Polyploidy: recurrent formation and genome evolution. Trends in Ecology & Evolution. 14(9):348–352.
Soltis DE, Soltis PS, Tate JA. 2004. Advances in the study of polyploidy since plant speciation. New Phytologist. 161(1):173–191.
Song H, Nan Z. 2015. Origin, divergence, and phylogeny of asexual *Epichloë* endophyte in *Elymus* species from western China. PLOS ONE. 10(5):e0127096.
Soper DM, Neiman M, Savytskyy OP, Zolan ME, Lively CM. 2013. Spermatozoa production by triploid males in the New Zealand freshwater snail *Potamopyrgus antipodarum*. Biological Journal of the Linnean Society. 110(1):227–234.
Spoelhof JP, Soltis PS, Soltis DE. 2017. Pure polyploidy: closing the gaps in autopolyploid research. Journal of Systematics and Evolution. 55(4):340–352.
Stats NZ. 2018. Agricultural production statistics: June 2017. New Zealand Government; [accessed 2019 June 20]. https://www.stats.govt.nz/information-releases/agricultural-production-statistics-june-2017-final.
Stats NZ. 2019. New Zealand Trade Dashboard. [accessed 2019 June 24]. https://statisticsnz.shinyapps.io/trade_dashboard/.
Stebbins GL. 1947. Demerec M, editor. Types of polyploids: their classification and significance. Vol. 1. Academic Press. Advances in genetics.
Stebbins GL. 1950. Variation and evolution in plants. Oxford: Oxford University Press.
Stebbins GL. 1971. Chromosomal evolution in higher plants. London: Edward Arnold.
Stebbins GL. 1984. Polyploidy and the distribution of the Arctic-alpine flora: new evidence and a new approach. Botanica Helvetica. 94:1–13.
Stewart AV. 2006. Mercer CF, editor. Genetic origins of New Zealand perennial ryegrass (*Lolium perenne*) cultivars. New Zealand Grassland Association. 'Breeding for success: diversity in action'

Proceedings of the 13th Australasian Plant Breeding Conference, Christchurch, New Zealand 18-21 April 2006.
Stuessy TF, Takayama K, López-Sepúlveda P, Crawford DJ. 2014. Interpretation of patterns of genetic variation in endemic plant species of oceanic islands. Botanical Journal of the Linnean Society. 174(3):276–288.
Sun H, Wu S, Zhang G, Jiao C, Guo S, Ren Y, Zhang J, Zhang H, Gong G, Jia Z, et al. 2017. Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. Molecular Plant. 10(10):1293–1306.
Szymura JM, Farana I. 1978. Inheritance and linkage analysis of five enzyme loci in interspecific hybrids of toadlets, genus *Bombina*. Biochemical Genetics. 16(3):307–319.
Tatum TC, Stepanovic S, Biradar DP, Rayburn AL, Korban SS. 2005. Variation in nuclear DNA content in *Malus* species and cultivated apples. Genome. 48(5):924–930.
Tay ML, Meudt HM, Garnock-Jones PJ, Ritchie PA. 2010. DNA sequences from three genomes reveal multiple long-distance dispersals and non-monophyly of sections in Australasian *Plantago* (Plantaginaceae). Australian Systematic Botany. 23(1):47–68.
Tayalé A, Parisod C. 2013. Natural pathways to polyploidy in plants and consequences for genome reorganization. Cytogenetic and Genome Research. 140(2–4):79–96.
te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesová M, Pysek P. 2012. The more the better? The role of polyploidy in facilitating plant invasions. Annals of Botany. 109(1):19–45.
Trewick SA, Morgan-Richards M, Chapman HM. 2004. Chloroplast DNA diversity of *Hieracium pilosella* (Asteraceae) introduced to New Zealand: reticulation, hybridization, and invasion. American Journal of Botany. 91(1):73–85.
Trojak-Goluch A, Skomra U. 2018. Breeding of triploid common hop cultivars (*Humulus lupulus* L. Polish Journal of Agronomy. 34:3–10.
Tsukaya H. 2008. Controlling size in multicellular organs: focus on the leaf. PLOS Biology. 6(7):e174.
United Nations Department of Public Information. 2017. World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100 – says UN. https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html.
Vallejo-Marín M, Buggs RJA, Cooley AM, Puzey JR. 2015. Speciation by genome duplication: repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*. Evolution. 69(6):1487–1500.
Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al. 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh. Nature Genetics. 42(10):833–839.
Wagner WH, Moran RC, Werth CR. 1993. Flora of North America Editorial Committee, editor. Aspleniaceae. Vol. 23. Oxford University Press on Demand. Flora of North America: volume 2: pteridophytes and gymnosperms.
Wallis GP, Jorge F. 2018. Going under down under? Lineage ages argue for extensive survival of the Oligocene marine transgression on Zealandia. Molecular Ecology. 27(22):4368–4396.
Wang T, Liu L, Ning C, Lü Z, Jia X, Gao Z, Qiao Y. 2016. Alterations of DNA methylation and gene expression during hybridization and polyploidization in *Fragaria* spp. Scientia Horticulturae. 201:218–224.
Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, et al. 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics. 172(1):507–517.
Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, Shen C, Li J, Yang Q, Zhang X. 2018. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. Nature Plants. 4(2):90–97.
Wang J, Ye LH, Liu QZ, Peng LY, Liu W, Yi XG, Wang YD, Xiao J, Xu K, Hu FZ, et al. 2015. Rapid genomic DNA changes in allotetraploid fish hybrids. Heredity. 114(6):601–609.
Wendel JF, Schnabel A, Seelanan T. 1995. Biderectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proceedings of the National Academy of Sciences of the United States of America. 92(1):280–284.
Winkworth R, Wagstaff S, Glenny DJ, Lockhart P. 2002. Plant dispersal N.E.W.S. from New Zealand. Trends in Ecology & Evolution. 17:514–520.

Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. Nature Reviews Genetics. 2(5):333–341.
Wolfe KH. 2015. Origin of the yeast whole-genome duplication. PLOS Biology. 13(8):e1002221.
Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. 387(6634):708–713.
Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences of the United States of America. 106(33):13875–13879.
Wu Y, Sun Y, Shen K, Sun S, Wang J, Jiang T, Cao S, Josiah SM, Pang J, Lin X, et al. 2015. Immediate genetic and epigenetic changes in F1 hybrids parented by species with divergent genomes in the rice genus (*Oryza*). PLOS ONE. 10(7):e0132911.
Yoo MJ, Liu X, Pires JC, Soltis PS, Soltis DE. 2014. Nonadditive gene expression in polyploids. Annual Review of Genetics. 48:485–517.
Yoo MJ, Szadkowski E, Wendel JF. 2013. Homeolog expression bias and expression level dominance in allopolyploid cotton. Heredity. 110:171–180.

3 A study of the shared transcriptomic responses to hybridisation, across kingdoms

The research presented in the following manuscript aimed to address the absence of cross-kingdom (plants, animals, fungi) transcriptomic studies among hybrid literature. The use of a generalised analytical framework enabled the comparison of global transcriptomic- and functional enrichment patterns between biological systems with markedly different genetic complements. However, research was limited to a small number of representative systems due to the lack of suitable available RNA-seq and genomic data. The manuscript was submitted to the journal of Molecular Biology and Evolution (MBE) in July 2021. A link to a GitHub repository containing the multiple supplementary files associated with the manuscript is given in the Data section at the end of this thesis.

This chapter also includes an addendum containing a series of data and code validations, in addition to other analyses that were ultimately not included in the paper. The purpose of the addendum is to present the entire body of work from this degree. Methods and results for each additional analysis or validation have been grouped together to improve readibility. Note: the additional analyses and validations occurred at multiple time points throughout this degree. Consequently, they may no longer be synchronised with the manuscript. An example where this has occurred is with the expression category labels. Four expression categories were used in the initial analyses:

- Parental expression inheritance (PEI)
- Homeolog expression blending (HEBl)
- Homeolog expression bias (HEBi)
- Homeolog expression reversal (HER)

However, a decision was subsequently made to differentiate PEI by differential or equal parental expression. The categories were also renamed for simplicity, becoming:

- Inherited equal
- Inherited differential expression
- Blending

- Bias
- Reversal

My contributions to this manuscript were designing and performing the experiments, analysing the data, writing the paper and designing the figures.

Cross-kingdom transcriptomic trends in the evolution of hybrid gene expression

Anna H. Behling[1], David J. Winter[1], Austen R. D. Ganley[2], Murray P. Cox[1*]

[1] School of Fundamental Sciences, Massey University, Palmerston North, New Zealand

[2] School of Biological Sciences, University of Auckland, Auckland, New Zealand

[*] Corresponding author: Murray Cox (m.p.cox@massey.ac.nz)

Abstract

Hybridisation is a route to speciation that occurs widely across the eukaryote tree of life. The success of allopolyploids (hybrid species with increased ploidy) and homoploid hybrids (with unchanged ploidy) is well documented. However, the formation and establishment of both types of hybrid is not straightforward, with a suite of near-instantaneous and longer-term biological repercussions faced by the new species. Central to these challenges is the rewiring of gene-regulatory networks caused by the merger of the distinct genomes inherited from both parental species. Research on the evolution of hybrid gene expression has largely focused on single hybrid species or a few gene families. Here, we present the first cross-kingdom, transcriptome-wide study to explore the fates of genes following hybridisation. We pair each allopolyploid system with a closely-related homoploid hybrid to decouple the influence of increased ploidy from genome merger on post-hybridisation expression patterns. Genome merger, not changes in ploidy, has the greatest effect across all study systems. We identify another common trend across kingdoms: genes that are differentially expressed in parent species preferentially have more similar expression in hybrid descendants, likely as a consequence of regulatory cross-talk within the hybrid nucleus. We also highlight the prevalence of gene loss or silencing among extremely differentially expressed genes in hybrid species. These general patterns suggest that the evolutionary process of hybridisation leads to common high-level expression outcomes, regardless of the particular species or kingdom.

Introduction

Hybrid species arise from successful interbreeding between two distinct parental species. Hybridisation is ubiquitous across the major eukaryotic kingdoms (Stukenbrock et al., 2012; Pereira et al., 2014; Schedina et al., 2014; Campbell et al., 2017; Edger et al., 2017; Bertioli et al., 2019), despite numerous pre- and post-zygotic reproductive isolating barriers that typically prevent the production of viable offspring from genetically divergent species. Hybridisation may occur with (allopolyploidisation) or without (homoploid hybridisation) a concomitant increase in ploidy level. Both outcomes involve the merger of genomes from each parental species, causing major disruptions at every level of an organism's cellular biology: genomic (McClintock, 1984; Qin et al., 2016), transcriptomic (Cox et al., 2014), proteomic (Holá et al., 2017; Ueno et al., 2019) and metabolomic (Zhang et al., 2019). However, successful adaptation to these challenges can confer new advantages on the hybrid, due to intergenomic heterosis (hybrid vigour) and enhanced genomic redundancy (Gu et al., 2003; Adams & Wendel, 2005; Baranwal et al., 2012; Fujimoto et al., 2018). Therefore, despite their initially improbable persistence, some hybrid species can even outcompete their parental species in the existing environment, or colonise transgressive niches unoccupied by either parent (Mallet, 2007; Kim et al., 2008; Li et al., 2014).

The effects of hybridisation on gene expression can be inferred by comparing the expression of orthologs (copies of a gene in the parents) with their corresponding homeologs (parentally-derived copies of a gene in the hybrid). Previous investigations into the evolution of hybrid gene expression have largely focused on single hybrid species (Combes et al., 2013; Yoo et al., 2013; Coate et al., 2014; Schedina et al., 2014; Matos et al., 2015; Zhang et al., 2015; Wang et al., 2016; Wu et al., 2016; Edger et al., 2017; Han et al., 2017; McElroy et al., 2017; Zhang et al., 2017; Qi et al., 2018; Zhang et al., 2018; Kryvokhyzha et al., 2019; Matos et al., 2019). While this research has furthered our knowledge of specific species, the different approaches and analytical frameworks employed by these studies preclude direct comparisons across different hybrid systems, thus preventing any generalisations about hybridism and gene expression more broadly. Most studies have also focused on a small number of gene families (Fulneček et al., 2009; Gong et al., 2014; Wen et al., 2019; Cui et al., 2020). A few studies have performed comparative transcriptomic analyses of multiple hybrid systems (Coate et al., 2012; Tronchoni et al., 2017; Hovhannisyan et al., 2020), with similarities noted in the gene

expression profiles of a fungal hybrid and plant hybrid (Cox et al., 2014). However, no systematic study of gene expression in multiple hybrid systems across kingdoms has been undertaken.

Here, we perform a comparative analysis of the transcriptome-wide impact of hybridisation on gene expression across the major eukaryotic kingdoms: plants, animals and fungi. We pair representative allopolyploid and homoploid hybrid systems from the same kingdom to investigate the relative effects of hybridisation versus an increase in ploidy level on gene expression. To perform transcriptomic comparisons across species with markedly different gene complements, we employ a structured and generalised 'fate of genes' expression framework, rather than bespoke approaches focusing on species-specific genes in each individual system. Our results bolster the growing view that hybridisation has a greater effect on gene expression than an increase in ploidy. We also show that most genes with expression differences between the parental species have more similar expression in the hybrid species, and we highlight the prevalence of gene loss or silencing among extremely differentially expressed genes in hybrids. These generalisations transcend the individual species upon which these analyses are performed, and therefore emphasise general expression outcomes following hybridisation, regardless of species or kingdom.

Results

Representative study systems

To compare gene expression in hybrid systems from different kingdoms, we sought datasets with RNA-seq data available for both the hybrids and their parent species. Following a systematic search of the literature, we identified only a limited number of systems from each kingdom that met all obligatory selection criteria (see Supplementary File 1), and we subsequently selected six representative systems for analysis in this study: one allopolyploid and one homoploid hybrid from each of animals, plants and fungi (Table 1). Most often, candidate systems did not meet the selection criteria due to an absence of suitable RNA-seq data, lacking either biological replicates, or with hybrid and parental data produced under different experimental conditions (Supplementary File 1). For animals and plants, we were able to acquire RNA-seq and genomic data for intra-genus pairings of allopolyploid and homoploid

hybrids, but obtained only intra-phylum allopolyploid and homoploid hybrids for the fungi (Supplementary File 2).

Classifying hybrid gene expression patterns

The animal, plant and fungal species used in this study vary greatly in their ploidy level, number of genes and lifestyles. Thus, we required a systematic and generalised framework to enable the inference of cross-kingdom patterns of gene expression. HyLiTE (Duchemin et al., 2015) was used to assign reads to homeologs using diagnostic SNP information to distinguish the different parental copies. As per the HyLiTE protocol, sequencing reads were mapped using one of the parental gene sets from each representative system as a reference (Supplementary File 2). We were able to align 0.6-15.6 million reads per replicate (Table 2), with the wide range of mapped reads reflecting the variation in size between the raw datasets. Next, to perform differential expression analyses, we adapted an approach used previously in Yoo et al. (2013) and Cox et al. (2014), based on the integration of parental (ortholog-ortholog) and hybrid (homeolog-homeolog) differential expression analyses. Expression between each ortholog or homeolog pair in the respective analysis is either differential (adjusted p value < 0.05, fold change > 2; towards one of the parental gene copies) or approximately equal, generating nine possible expression scenarios following the integration of the ortholog and homeolog results. These nine expression scenarios can then be grouped into the following five expression categories (Figure 1):

- **Inherited equal**: equal expression between parental orthologs remains equal in the hybrid homeologs
- **Inherited differential expression**: parental differential expression is inherited in the hybrid
- **Blending**: parental differential expression changes to equal expression in the hybrid
- **Bias**: equal parental expression changes to differential expression in the hybrid
- **Reversal**: differential expression occurs in both the parents and the hybrid, but the direction of expression bias is reversed in the hybrid

After excluding genes with very low coverage across all samples, we were able to analyse between 3632 and 9578 genes per system (Supplementary Table 1).

Parental expression differences tend to be lost in hybrids

We first tested our six systems for evidence of subgenome dominance; an observation sometimes seen among hybrid species where, following the merger of the two diverged parental subgenomes, there is unequal expression of parental contributions at the genome-wide level (Renny-Byfield et al., 2015; Edger et al., 2017; Ren et al., 2019; Bird et al., 2021). If subgenome dominance were present among our datasets, we would expect to see a substantial uni-directional parental bias in $\log_2$ expression fold changes. However, we did not identify any evidence of this: the median hybrid $\log_2$ fold change in expression ranged from -0.9 to 0.1 (Supplementary Table 2). Full data for the distribution of parental and hybrid $\log_2$ fold change in expression for each system can be found in Supplementary Figures 1-6.

Our six representative systems vary substantially in the proportion of genes with differential expression between the parents (ranging from 11.4% to 58.9%). This result has important ramifications for the proportion of genes in each of the expression categories described above, as the proportion of genes that are differentially expressed in the parental species, a statistic that is not related to hybridisation at all, strongly influences the number of genes that can fall in each of our five expression categories. For example, high levels of parental differential expression necessarily forces a smaller proportion of genes available for the inherited equal category. Indeed, as expected, we find a large and significant inverse correlation between the level of parental differential expression and proportion of genes in the inherited equal category (adj. $r^2 = 0.73$, $p = 0.019$) (Figure 2). Therefore, for each system we examine the expression outcomes ('fates') of genes in the hybrid compared to what they are in the parents.

We allocated all genes in the final dataset for each parent-hybrid system into the five expression categories, and then used this information to investigate what happens to the expression pattern (equal or differential) for each parental ortholog after it was inherited as a homeolog in the hybrid (Figure 3). We found that the majority of genes with expression differences between the parents lose this differential expression (blending) in the hybrid transcriptome. On average across the systems, 70% of differentially expressed parental genes (58-83%) were blended, while only an average of 21% of genes with equal expression (9-31%) between the parents gained an expression bias in the hybrid. This outcome was statistically significant in all systems.

The net result of this is a striking trend whereby genes typically show more equal expression between the two homeologs in the hybrid than between the two orthologs in the parents. This trend does not appear to be influenced by the degree of divergence in parental gene expression. Indeed, even for the plant allopolyploid system where the absolute number of genes becoming biased is greatest, the majority of differentially expressed genes are blended, so the proportion of differentially expressed genes that experience blending is still larger than the proportion of equally expressed genes gaining a bias (Figure 3).

One possible trivial explanation for the preponderance of blending is that it reflects a greater ability to detect a statistically-significant difference in expression in the parents versus the hybrids, as the read count per ortholog/homeolog is half in a hybrid species compared to the parents (assuming the same number of reads for each species). If this artefact explained our result, we would expect blended genes to predominantly be genes with borderline expression differences in the parental species. To test this hypothesis, we divided all genes with parental differential expression into deciles based on the level of differential expression and calculated the proportion of blended genes for each decile. We found no trend in the proportion of blended genes across deciles, except for a consistent drop in blending in the tenth decile; the most highly differentially expressed set of genes (Figure 4). Thus, we conclude that the propensity to blend expression in hybrids is likely a biological phenomenon rather than a statistical artefact.

A drop in the proportion of genes with blended expression was observed in the tenth decile of parental differential expression in all systems (Figure 4). We wondered if the lack of blending in this decile was the result of genes with very high levels of differential expression in the parent species being particularly recalcitrant to blending. To test this, we examined the fates of genes with an extreme (greater than 50-fold) difference in ortholog expression. However, while the numbers of these genes was too low to analyse in three of the hybrid systems, two of the remaining three hybrid systems showed limited reduction in blending, and the proportions of genes in the reversal category in all three systems were similar to those seen across all differentially expressed orthologs (Supplementary Figure 7). Thus, it remains unclear what is driving the reduction in the proportion of blending amongst genes in the tenth decile of parental differential expression.

Gene loss or silencing is prevalent among extremely differentially expressed homeologs

Extreme differential expression can include genes in the hybrid that are extremely differentially expressed, as well as genes in the parent species. We wondered whether parental extreme differential expression is associated with extreme differential expression in the hybrid, and so looked at this class of gene in the hybrids. Interestingly, while we saw no particular relationship between parental and hybrid extreme differential expression (Supplementary Table 3), we found that extreme differential expression is far more common in hybrids (between homeologs) than in the parents (Table 3). Hybridisation can cause the loss or silencing of hybrid gene copies (Nasrallah et al., 2007; Buggs et al., 2009; Buggs et al., 2010; Feldman et al., 2012; Cox et al., 2014; Lashermes et al., 2016), so we looked to see if this could explain the preponderance of extremely differentially expressed homeologs by determining how many of these homeologs have no reads mapped to one gene copy. In the reversal, inherited differential expression and bias categories, a high percentage of these homeologs have no reads mapped for one of the copies (57-100% of reversal genes; 33-98% of inherited differential expression genes; 67-94% of bias genes; Figure 5; Supplementary Files 3-5). Together, the genes with expression for only one homeolog comprise 0.2-7.4% of genes in the final gene sets across our representative systems, but are strongly overrepresented among the set of extremely differentially expressed genes.

Increased ploidy is not a major influence on altered hybrid gene expression patterns

Current literature suggests that hybridisation, rather than the increase in ploidy level, has a greater impact on post-allopolyploidisation gene expression patterns (Hegarty et al., 2005; Albertin et al., 2006; Hegarty et al., 2006; Wang et al., 2006; Chelaifa et al., 2010; Jung et al., 2015; Li et al., 2018). Our results are consistent with this, as there are no clear differences in gene expression patterns between the allopolyploid and homoploid hybrid systems (Figures 2-5). To more systematically test whether allopolyploid and homoploid hybrid expression patterns can be distinguished, we paired each kingdom's representative homoploid hybrid and allopolyploid systems to allow the comparison of hybrid gene expression patterns with and without an increase in parental ploidy level. We then performed hierarchical clustering of normalised expression category count data associated with genes that have changed gene expression category in the hybrid (the blending, bias and reversal categories). Count data were normalised by the total number of genes in these three categories of interest. We did not find any clustering of gene expression patterns as a consequence of increased ploidy (Figure 6),

suggesting that hybridisation has the greater influence on allopolyploid gene expression patterns.

Few commonalities in functional enrichment exist between study systems

Finally, we wondered whether the shared tendency of the hybrid transcriptome to blend the expression of differentially expressed parental genes might be driven by selective effects. To test this, we looked for shared overrepresentation of functional GO characteristics within our different expression categories. We chose DAVID (Huang et al., 2009a, 2009b) to test for the functional enrichment of GO terms because it allows the user to test enrichment at a defined GO level, enabling a direct comparison of GO terms across the different systems. Lists of RefSeq identifiers corresponding to genes (or orthologs from closely related species where genes were not present in the DAVID knowledgebase) in the blending, bias, reversal and inherited differential expression categories were submitted to DAVID against a custom background of each system's total gene set. The inherited equal expression category was not submitted because the corresponding lists of gene identifiers were too large, relative to the background, to produce an informative enrichment result. Across the four categories, 54-100% of the enriched GO terms were found in only one system, 0-45% in two systems, and 0-16% in three systems. Only one GO term (oxidoreductase activity in inherited DE genes) was shared between four systems, and when we compared across all six systems, we saw no clear pattern of shared functional enrichment across the five most general GO levels (1-5) in the molecular function, biological process or cellular component ontologies (Supplementary Files 6-7). These results indicate that while expression patterns have common outcomes following hybridisation across very different organisms, the individual genes that change are, as might be expected, species-specific.

Discussion

Hybrid studies to date have largely focused on single hybrid species. While these studies have improved knowledge about each particular system, their bespoke approaches impede the synthesis of data into broad generalisations about hybrid gene expression. Here, we have utilised a standardised framework to make comparisons between study systems across three eukaryote kingdoms: animals, plants and fungi. We performed separate differential expression

analyses of orthologs and homeologs that, when integrated, enabled classification of the expression of each gene into one of five categories. Through this approach, we found that the proportion of genes with inherited equal expression in hybrid species is inversely correlated with the level of parental differential expression. This is a consequence of different levels of parental differential expression, which in turn likely reflects variable genetic divergence between the parental species. This results in different distributions of genes in the expression categories, and makes the direct comparisons of category counts between systems an inaccurate means of comparing transcriptomic differences in hybrids. Thus, we separately investigated the fate of genes with equal and with differential parental gene expression in allopolyploid and homoploid hybrid species across our three study kingdoms.

Consistent patterns of hybrid gene expression are likely the outcomes of interactions between homeologs and their regulatory factors

A gene with approximately equal expression between parental species has two possible fates when the orthologs are combined in the hybrid nucleus: either the equal expression is inherited, or an expression bias may arise in the hybrid. Our results show that inheritance of approximately equal expression is most likely. In contrast, differentially expressed parental genes have three possible fates following hybridisation: either the parental differential expression is inherited; differential expression is inherited but its direction is reversed; or the differential expression is lost in the hybrid (blending). Unlike equal parental expression, we found that the predominant outcome for differentially expressed parental genes was not inheritance, but blending of their expression in the hybrid. This major trend is species-independent, in addition to being independent of the extent of differential expression between the parental species and any change in ploidy in the hybrid.

It is of great interest to understand why consistent patterns of hybrid gene expression are observed in systems separated by vast phylogenetic distances. One possible explanation is that these represent adaptive changes in expression; for example by certain types of genes being deleterious when expressed at different levels in a hybrid. However, we found no evidence for certain functional gene types displaying a consistent pattern of hybrid expression, as might be expected if the adaptative hypothesis were true. The absence of such a pattern is consistent with previous observations among allopolyploid and homoploid hybrid species (Adams, 2007; Bassene et al., 2010). Instead, as we have previously proposed (Cox et al., 2014), we suggest that consistent patterns of hybrid gene expression are primarily driven by interactions between

homeologs and their regulatory factors in the hybrid nucleus. For example, in the case of blending, if parental differential expression is determined by the presence of a *trans*-acting gene regulatory factor (such as a transcription factor) in one parental species and absence in the other, the interaction of this regulatory factor with both homeologs within the hybrid nucleus may result in equal expression for both homeologs (Figure 7). Our observation of the predominance of blending of hybrid gene expression is consistent with this mechanism. Also consistent with this mechanism is our result suggesting that hybridisation, rather than change in ploidy level, is the major driver of hybrid expression patterns, as this mechanism is dependent on genome merger, but independent of an increase in ploidy. Data from studies on other allopolyploid and homoploid hybrid species including *Senecio* (Hegarty et al., 2005; Hegarty et al., 2006), *Brassica* (Albertin et al., 2006), *Arabidopsis* (Wang et al., 2006), *Spartina* (Chelaifa et al., 2010), wheat (Jung et al., 2015) and Cyprinidae fish (Li et al., 2018) have also suggested that genome merger has a greater effect on gene expression patterns than an increase in ploidy.

A common phenotypic pattern among hybrids is intermediacy: a condition where hybrids develop a phenotype that is intermediate to their parental species (Hermansen et al., 2011; Schneider et al., 2011; Salamone et al., 2013; Rubini Pisano et al., 2019). Phenotypic intermediacy can be conceptualised as a blending of parental phenotypes, and impacts particular morphological traits. Our observation of preferential blending of parental differential expression in hybrid species is interesting in this regard, as this blending at the transcriptional level may, in part, underlie blending (or intermediacy) at the phenotypic level. However, the predominant trend towards blending of hybrid gene expression is not likely to solely explain phenotypic intermediacy, as a number of hybrid species display transgressive phenotypes (Rieseberg et al., 1999; Stelkens et al., 2009; Dittrich-Reed & Fitzpatrick, 2013; Koide et al., 2019). It would be interesting to see whether the degree of intermediate versus transgressive phenotypes correlate with the level of blending when looking across many different hybrid systems, but this will require more systems and better phenotyping than is currently available.

The impact of genome shock limits blending of differentially expressed genes

We found that there are many more extremely differentially expressed genes in the hybrids compared to the parents, and most of these genes in the hybrid are extremely differentially expressed because reads map to only one homeolog. By definition, none of the genes with reads mapping to only one homeolog can fall into the blended category. Thus, if these genes were

removed, the proportion of differentially expressed parental genes that are blended would be even greater. This further emphasises the dominance of the blending trend across these hybrid systems. Non-expression of one homeolog could be the result of gene loss or complete gene silencing. Both phenomena have been observed in hybrid species as a response to genome merger (Nasrallah et al., 2007; Buggs et al., 2009; Buggs et al., 2010; Feldman et al., 2012; Cox et al., 2014; Lashermes et al., 2016). It is, however, difficult to infer the relative proportions of each, and it would be interesting to determine the relative proportions of these two outcomes across different hybrid systems in a similar manner to what we have done here with expression, for example by using PCR or by obtaining whole genome sequences of the hybrids.

Building a more robust picture of generalised hybrid transcriptomic responses

In this study, we have identified transcriptomic trends that are consistent in allopolyploid and homoploid hybrid systems from across the tree of life. However, our ability to generalise these findings across eukaryote hybrid systems is limited by our analyses being restricted to a single representative allopolyploid and homoploid system from each of the plant, animal and fungi kingdoms. Despite the widespread uptake of RNA-seq analysis, the lack of suitable datasets to perform robust comparisons of expression patterns in hybrids and their parent species remains the main limitation to extending the kind of analysis undertaken here to a broader set of hybrid systems. Future work would benefit from a more standardised approach to collecting transcriptomic data across animal, plant and fungal hybrids to facilitate systematic cross-study comparisons, and thus further explore the expression generalities across kingdoms that we identify in this work. The general expression trends identified here in hybrid species across a wide phylogenetic range will also need to be taken into consideration when interpreting the expression of genes in individual hybrid species.

Methods

Representative species chosen for analysis

A broad survey of the literature was performed to identify allopolyploid and homoploid hybrid systems with available RNA-seq and corresponding genomic data (see Supplementary File 1 for details). The viability of systems was assessed using a number of pre-defined obligatory and preferential criteria:

- Non-normalised RNA-seq data was available for at least two biological replicates of each constituent species (obligatory)
- The RNA-seq data was extracted from the same tissue, or from cells grown in the same medium, for each parent and hybrid species in the system (obligatory)
- A genome sequence or gene models were available for at least one of the parental species (obligatory)
- The systems contained naturally-occurring allopolyploid or homoploid hybrid species (preferential)
- The systems had minimal phylogenetic distance between the allopolyploid and homoploid hybrid (preferential)
- Both parental species were extant, as opposed to being close relatives of extinct parents (preferential)

Hybrid systems were required to meet all obligatory criteria in order to be considered for analysis. Although it was important to minimise phylogenetic distance between the allopolyploid and homoploid hybrid to limit any potential taxon-specific differences observed in gene expression patterns, of the preferential criteria, more weight was given to systems with naturally-occurring hybrids due to the practical inevitability of allopolyploid and homoploid hybrid representatives being at least from different species.

Acquisition and curation of gene sequences

To determine the level of sequence duplication within gene sets and standardize between species, each gene set was submitted to CD-HIT v4.8.1 (Fu et al., 2012), with the user-defined similarity threshold set to 0.95. The resulting representative sequences from each gene cluster were re-submitted to CD-HIT to confirm a 1:1 gene sequence to cluster ratio.

RNA-seq data processing

HyLiTE (hybrid lineage transcriptome explorer) v2.0.2 (Duchemin et al., 2015) was used to generate read count matrices for differential expression analysis. Single nucleotide polymorphisms within the RNA-seq reads, indicative of parental origin, were identified by HyLiTE, and subsequently used to classify reads in the hybrid to the parental subtranscriptomes. If a read contained only hybrid-specific and/or masked SNPs, HyLiTE classified the read as 'unknown'. If a read could be assigned to either parent, it was classified as 'uninformative'. RNA-seq read sequences and the CD-HIT-reduced reference gene sequences were provided to HyLiTE as input data, in addition to a 'protocol file' that defined the different species, their parent-hybrid relationships, the biological replicates, and corresponding expression data files. HyLiTE used the protocol file to map the RNA-seq reads to the gene sequences with Bowtie 2 v2.3.4.1 (Langmead & Salzberg, 2012). To be consistent with single-end read sets for many species, if paired-end reads were available, only the forward reads were mapped.

Validation of read count data

Extensive automated and manual validation checks of read count data were performed. Code and example files are available online (https://github.com/annabehling/multi_bowtie).

Differential expression analysis

HyLiTE read count matrices were processed with hyliter (https://github.com/dwinter/hyliter) and provided as input to DESeq2 v1.28.1 (Love et al., 2014) for differential expression analysis. Parental read count totals included the HyLiTE '+N' columns (these were reads with clear parental assignment that also contained one or more hybrid-specific SNPs). Two separate differential expression analyses were performed on each dataset. First, expression was compared between orthologs in the parental species, then the expression of homeologs of these same genes was compared in the hybrid.

Gene classification

Differential expression was defined as a fold change > 2, with an adjusted p value < 0.05. The fate of gene expression due to allopolyploid and homoploid hybridisation was characterised by

combining the results from the two differential expression analyses into five gene expression categories (Figure 1).

Genes were subsequently defined as extremely differentially expressed if they had a fold change > 50 in either the parental or hybrid differential expression analysis.

Read counts for genes with extreme differential expression

To test for non-expression among the absolute hybrid read count data, we calculated the mean of the homeolog read counts for our gene sets using the two HyLiTE hybrid read summary files, and subsequently ordered for convenience by the mean values for parent 1. Extremely differentially expressed homeologs in the reversal and inherited equal categories were then ranked in the ordered lists.

Enrichment of functional categories

To achieve a generalised comparison of functional enrichment, sequence identifiers corresponding to genes in the blending, bias, reversal and inherited differential expression categories were submitted to the DAVID bioinformatics resource v6.8 (Huang et al., 2009a, 2009b) for GO term annotation and enrichment analysis. For consistency, RefSeq (mRNA or protein) was used as the sequence identifier type across all study systems. Functional annotation by DAVID was restricted to those systems with genes present in the DAVID knowledgebase. In the following instances where this requirement was not met by the representative systems, Proteinortho v6.0.8 was used to identify single copy orthologs from a closely related species that was available in the DAVID database.

Epichloë:

• *Epichloë* currently does not have genes present in the DAVID knowledgebase.

• Single copy *Metarhizium robertsii* orthologs were generated from *Epichloë* protein sequences. *Metarhizium robertsii*, a closely related member of the Clavicipitaceae, has the highest number of RefSeq mRNA IDs in the RefSeq database, which is used by the DAVID knowledgebase.

Squalius:

• *Squalius* currently does not have genes present in the DAVID knowledgebase.

• Single copy *Danio rerio* orthologs were generated from *Squalius* translated CDS sequences. *Squalius* and *Danio* are both Cyprinidae, unlike an alternative candidate orthologous system that was considered, *Salmo salar* (Salmonidae). The study species, *Sq. pyrenaicus* and *Sq. alburnoides*, exhibit a high level of conserved sequences with the *D. rerio* genome (Inácio et al., 2012).

Genes in the specific functional categories were run against a custom background of all genes in the final datasets. Annotations were limited to the query species. The resulting charts were filtered by count and EASE thresholds of 1; all other settings were the default. The charts were subsequently downloaded and imported into R where they were filtered using the same significance threshold that was applied to the classification of differentially expressed genes ($p < 0.05$).

Data availability

All raw data used in this study are publicly available, with accession numbers listed in Supplementary File 2.

Acknowledgements

This work was supported by the Massey University Masterate and Joan Dingley Memorial scholarships. We thank Monika Karmin and Kelli Smith for their constructive feedback on figures.

References

Adams, K. L. (2007). Evolution of duplicate gene expression in polyploid and hybrid plants. *Journal of Heredity*, *98*(2), 136-141. 10.1093/jhered/esl061

Adams, K. L., & Wendel, J. F. (2005). Novel patterns of gene expression in polyploid plants. *Trends in Genetics, 21*(10), 539-543. 10.1016/j.tig.2005.07.009

Albertin, W., Balliau, T., Brabant, P., Chèvre, A.-M., Eber, F., Malosse, C., & Thiellement, H. (2006). Numerous and rapid nonstochastic modifications of gene products in newly synthesized *Brassica napus* allotetraploids. *Genetics*, *173*(2), 1101-1113. 10.1534/genetics.106.057554

Baranwal, V. K., Kapoor, S., Zehr, U. B., Mikkilineni, V., & Tyagi, A. K. (2012). Heterosis: emerging ideas about hybrid vigour. *Journal of Experimental Botany*, *63*(18), 6309-6314. 10.1093/jxb/ers291

Bassene, J. B., Froelicher, Y., Dubois, C., Ferrer, R. M., Navarro, L., Ollitrault, P., & Ancillo, G. (2010). Non-additive gene regulation in a citrus allotetraploid somatic hybrid between *C. reticulata* Blanco and *C. limon* (L.) Burm. *Heredity*, *105*(3), 299-308. 10.1038/hdy.2009.162

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., . . . Schmutz, J. (2019). The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nature Genetics*, *51*(5), 877-884. 10.1038/s41588-019-0405-z

Bird, K. A., Niederhuth, C. E., Ou, S., Gehan, M., Pires, J. C., Xiong, Z., . . . Edger, P. P. (2021). Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. *New Phytologist*, *230*(1), 354-371. 10.1111/nph.17137

Buggs, R. J. A., Doust, A. N., Tate, J. A., Koh, J., Soltis, K., Feltus, F. A., . . . Soltis, D. E. (2009). Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. *Heredity*, *103*(1), 73-81. 10.1038/hdy.2009.24

Buggs, R. J. A., Elliott, N. M., Zhang, L., Koh, J., Viccini, L. F., Soltis, D. E., & Soltis, P. S. (2010). Tissue-specific silencing of homoeologs in natural populations of the recent

allopolyploid *Tragopogon mirus*. *New Phytologist, 186*(1), 175-183. 10.1111/j.1469-8137.2010.03205.x

Campbell, M. A., Tapper, B. A., Simpson, W. R., Johnson, R. D., Mace, W., Ram, A., . . . Cox, M. P. (2017). *Epichloë hybrida*, sp. nov., an emerging model system for investigating fungal allopolyploidy. *Mycologia, 109*(5), 715-729. 10.1080/00275514.2017.1406174

Chelaifa, H., Monnier, A., & Ainouche, M. (2010). Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina* × *townsendii* and *Spartina anglica* (Poaceae). *New Phytologist, 186*(1), 161-174. 10.1111/j.1469-8137.2010.03179.x

Coate, J. E., Bar, H., & Doyle, J. J. (2014). Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). *The Plant Cell, 26*(1), 136. 10.1105/tpc.113.119966

Coate, J. E., Powell, A. F., Owens, T. G., & Doyle, J. J. (2012). Transgressive physiological and transcriptomic responses to light stress in allopolyploid *Glycine dolichocarpa* (Leguminosae). *Heredity, 110*, 160. 10.1038/hdy.2012.77

Combes, M.-C., Dereeper, A., Severac, D., Bertrand, B., & Lashermes, P. (2013). Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytologist, 200*(1), 251-260. 10.1111/nph.12371

Cox, M. P., Dong, T., Shen, G., Dalvi, Y., Scott, D. B., & Ganley, A. R. D. (2014). An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. *PLOS Genetics, 10*(3), e1004180. 10.1371/journal.pgen.1004180

Cui, Y., Su, Y., Wang, J., Jia, B., Wu, M., Pei, W., . . . Yu, J. (2020). Genome-wide characterization and analysis of CIPK gene family in two cultivated allopolyploid cotton species: sequence variation, association with seed oil content, and the role of GhCIPK6. *International Journal of Molecular Sciences, 21*(3), 863. 10.3390/ijms21030863

Dittrich-Reed, D. R., & Fitzpatrick, B. M. (2013). Transgressive hybrids as hopeful monsters. *Evolutionary Biology, 40*(2), 310-315. 10.1007/s11692-012-9209-0

Duchemin, W., Dupont, P. Y., Campbell, M. A., Ganley, A. R., & Cox, M. P. (2015). HyLiTE: accurate and flexible analysis of gene expression in hybrid and allopolyploid species. *BMC Bioinformatics, 16*(8). 10.1186/s12859-014-0433-8

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., . . . Puzey, J. R. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *The Plant Cell, 29*(9), 2150. 10.1105/tpc.17.00010

Feldman, M., Levy, A. A., Fahima, T., & Korol, A. (2012). Genomic asymmetry in allopolyploid plants: wheat as a model. *Journal of Experimental Botany, 63*(14), 5045-5059. 10.1093/jxb/ers192

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics, 28*(23), 3150-3152. 10.1093/bioinformatics/bts565

Fujimoto, R., Uezono, K., Ishikura, S., Osabe, K., Peacock, W. J., & Dennis, E. S. (2018). Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. *Breeding Science, 68*(2), 145-158. 10.1270/jsbbs.17155

Fulneček, J., Matyášek, R., & Kovařík, A. (2009). Faithful inheritance of cytosine methylation patterns in repeated sequences of the allotetraploid tobacco correlates with the expression of DNA methyltransferase gene families from both parental genomes. *Molecular Genetics and Genomics, 281*(4), 407-420. 10.1007/s00438-008-0420-8

Gong, L., Olson, M., & Wendel, J. F. (2014). Cytonuclear evolution of rubisco in four allopolyploid lineages. *Molecular Biology and Evolution, 31*(10), 2624-2636. 10.1093/molbev/msu207

Gu, Z., Steinmetz, L. M., Gu, X., Scharfe, C., Davis, R. W., & Li, W.-H. (2003). Role of duplicate genes in genetic robustness against null mutations. *Nature*, *421*(6918), 63-66. 10.1038/nature01198

Han, S., Liu, H., Yan, M., Qi, F., Wang, Y., Sun, Z., . . . He, G. (2017). Differential gene expression in leaf tissues between mutant and wild-type genotypes response to late leaf spot in peanut (*Arachis hypogaea* L.). *PLOS One*, *12*(8), e0183428. 10.1371/journal.pone.0183428

Hegarty, M. J., Barker, G. L., Wilson, I. D., Abbott, R. J., Edwards, K. J., & Hiscock, S. J. (2006). Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Current Biology*, *16*(16), 1652-1659. 10.1016/j.cub.2006.06.071

Hegarty, M. J., Jones, J. M., Wilson, I. D., Barker, G. L., Coghill, J. A., Sanchez-Baracaldo, P., . . . Hiscock, S. J. (2005). Development of anonymous cDNA microarrays to study changes to the *Senecio* floral transcriptome during hybrid speciation. *Molecular Ecology*, *14*(8), 2493-2510. 10.1111/j.1365-294x.2005.02608.x

Hermansen, J. S., Saether, S. A., Elgvin, T. O., Borge, T., Hjelle, E., & Saetre, G.-P. (2011). Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular Ecology*, *20*(18), 3812-3822. 10.1111/j.1365-294X.2011.05183.x

Holá, D., Benešová, M., Fischer, L., Haisel, D., Hnilička, F., Hniličková, H., . . . Wilhelmová, N. (2017). The disadvantages of being a hybrid during drought: a combined analysis of plant morphology, physiology and leaf proteome in maize. *PLOS One*, *12*(4), e0176121. 10.1371/journal.pone.0176121

Hovhannisyan, H., Saus, E., Ksiezopolska, E., Hinks Roberts, A. J., Louis, E. J., & Gabaldón, T. (2020). Integrative omics analysis reveals a limited transcriptional shock after yeast interspecies hybridization. *Frontiers in Genetics*, *11*, 404. 10.3389/fgene.2020.00404

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, *37*(1), 1-13. 10.1093/nar/gkn923

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols, 4*(1), 44-57. 10.1038/nprot.2008.211

Inácio, A., Pinho, J., Pereira, P. M., Comai, L., & Coelho, M. M. (2012). Global analysis of the small RNA transcriptome in different ploidies and genomic combinations of a vertebrate complex – the *Squalius alburnoides*. *PLOS One, 7*(7), e41158. 10.1371/journal.pone.0041158

Jung, Y., Kawaura, K., Kishii, M., Sakuma, S., & Ogihara, Y. (2015). Comparison of genome-wide gene expression patterns in the seedlings of nascent allohexaploid wheats produced by two combinations of hybrids. *Genes & Genetic Systems, 90*(2), 79-88. 10.1266/ggs.90.79

Kim, S.-T., Sultan, S. E., & Donoghue, M. J. (2008). Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. *Proceedings of the National Academy of Sciences of the United States of America, 105*(34), 12370. 10.1073/pnas.0805141105

Koide, Y., Sakaguchi, S., Uchiyama, T., Ota, Y., Tezuka, A., Nagano, A. J., . . . Kishima, Y. (2019). Genetic properties responsible for the transgressive segregation of days to heading in rice. *G3 (Bethesda), 9*(5), 1655-1662. 10.1534/g3.119.201011

Kryvokhyzha, D., Milesi, P., Duan, T., Orsucci, M., Wright, S. I., Glémin, S., & Lascoux, M. (2019). Towards the new normal: transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (*Capsella bursa-pastoris*). *PLOS Genetics, 15*(5), e1008131. 10.1371/journal.pgen.1008131

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods, 9*(4), 357-359. 10.1038/nmeth.1923

Lashermes, P., Hueber, Y., Combes, M.-C., Severac, D., & Dereeper, A. (2016). Inter-genomic DNA exchanges and homeologous gene silencing shaped the nascent allopolyploid coffee genome (*Coffea arabica* L.). *G3 (Bethesda), 6*(9), 2937-2948. 10.1534/g3.116.030858

Li, A., Liu, D., Wu, J., Zhao, X., Hao, M., Geng, S., . . . Mao, L. (2014). mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid

heterosis in nascent hexaploid wheat. *The Plant Cell, 26*(5), 1878-1900. 10.1105/tpc.114.124388

Li, W., Liu, J., Tan, H., Luo, L., Cui, J., Hu, J., . . . Liu, S. (2018). Asymmetric expression patterns reveal a strong maternal effect and dosage compensation in polyploid hybrid fish. *BMC Genomics, 19*(1), 517. 10.1186/s12864-018-4883-7

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12), 550. 10.1186/s13059-014-0550-8

Mallet, J. (2007). Hybrid speciation. *Nature, 446*(7133), 279-283. 10.1038/nature05706

Matos, I., Machado, M. P., Schartl, M., & Coelho, M. M. (2015). Gene expression dosage regulation in an allopolyploid fish. *PLOS One, 10*(3), e0116309. 10.1371/journal.pone.0116309

Matos, I., Machado, M. P., Schartl, M., & Coelho, M. M. (2019). Allele-specific expression variation at different ploidy levels in *Squalius alburnoides*. *Scientific Reports, 9*(1), 3688. 10.1038/s41598-019-40210-8

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science, 226*(4676), 792. 10.1126/science.15739260

McElroy, K. E., Denton, R. D., Sharbrough, J., Bankers, L., Neiman, M., & Lisle Gibbs, H. (2017). Genome expression balance in a triploid trihybrid vertebrate. *Genome Biology and Evolution, 9*(4), 968-980. 10.1093/gbe/evx059

Nasrallah, J. B., Liu, P., Sherman-Broyles, S., Schmidt, R., & Nasrallah, M. E. (2007). Epigenetic mechanisms for breakdown of self-incompatibility in interspecific hybrids. *Genetics, 175*(4), 1965-1973. 10.1534/genetics.106.069393

Pereira, C. S. A., Aboim, M. A., Ráb, P., & Collares-Pereira, M. J. (2014). Introgressive hybridization as a promoter of genome reshuffling in natural homoploid fish hybrids (Cyprinidae, Leuciscinae). *Heredity, 112*(3), 343-350. 10.1038/hdy.2013.110

Qi, X., Wang, H., Song, A., Jiang, J., Chen, S., & Chen, F. (2018). Genomic and transcriptomic alterations following intergeneric hybridization and polyploidization in the *Chrysanthemum nankingense* × *Tanacetum vulgare* hybrid and allopolyploid (Asteraceae). *Horticulture Research, 5*, 5. 10.1038/s41438-017-0003-0

Qin, Q., Lai, Z., Cao, L., Xiao, Q., Wang, Y., & Liu, S. (2016). Rapid genomic changes in allopolyploids of *Carassius auratus red var.* (♀) × *Megalobrama amblycephala* (♂). *Scientific Reports, 6*, 34417. 10.1038/srep34417

Ren, L., Li, W., Qin, Q., Dai, H., Han, F., Xiao, J., . . . Liu, S. (2019). The subgenomes show asymmetric expression of alleles in hybrid lineages of *Megalobrama amblycephala* × *Culter alburnus*. *Genome Research, 29*(11), 1805-1815. 10.1101/gr.249805.119

Renny-Byfield, S., Gong, L., Gallagher, J. P., & Wendel, J. F. (2015). Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Molecular Biology and Evolution, 32*(4), 1063-1071. 10.1093/molbev/msv001

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity, 83*(4), 363-372. 10.1038/sj.hdy.6886170

Rubini Pisano, A., Moré, M., Cisternas, M. A., Raguso, R. A., & Benitez-Vieyra, S. (2019). Breakdown of species boundaries in *Mandevilla*: floral morphological intermediacy, novel fragrances and asymmetric pollen flow. *Plant Biology, 21*(2), 206-215. 10.1111/plb.12924

Salamone, I., Govindarajulu, R., Falk, S., Parks, M., Liston, A., & Ashman, T.-L. (2013). Bioclimatic, ecological, and phenotypic intermediacy and high genetic admixture in a natural hybrid of octoploid strawberries. *American Journal of Botany, 100*(5), 939-950. 10.3732/ajb.1200624

Schedina, I. M., Hartmann, S., Groth, D., Schlupp, I., & Tiedemann, R. (2014). Comparative analysis of the gonadal transcriptomes of the all-female species *Poecilia formosa* and its maternal ancestor *Poecilia mexicana*. *BMC Research Notes, 7*, 249. 10.1186/1756-0500-7-249

Schneider, J. V., Schulte, K., Aguilar, J. F., & Huertas, M. L. (2011). Molecular evidence for hybridization and introgression in the neotropical coastal desert-endemic *Palaua* (Malveae,

Malvaceae). *Molecular Phylogenetics and Evolution, 60*(3), 373-384. 10.1016/j.ympev.2011.05.010

Stelkens, R. B., Schmid, C., Selz, O., & Seehausen, O. (2009). Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. *BMC Evolutionary Biology, 9*, 283. 10.1186/1471-2148-9-283

Stukenbrock, E. H., Christiansen, F. B., Hansen, T. T., Dutheil, J. Y., & Schierup, M. H. (2012). Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proceedings of the National Academy of Sciences of the United States of America, 109*(27), 10954-10959. 10.1073/pnas.1201403109

Tronchoni, J., García-Ríos, E., Guillamón, J. M., Querol, A., & Pérez-Torrado, R. (2017). Transcriptomic analysis of *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* hybrids during low temperature winemaking. *F1000Research, 6*, 679. 10.12688/f1000research.11550.3

Ueno, N., Kashiwagi, M., Kanekatsu, M., Marubashi, W., & Yamada, T. (2019). Accumulation of protein aggregates induces autolytic programmed cell death in hybrid tobacco cells expressing hybrid lethality. *Scientific Reports, 9*(1), 10223. 10.1038/s41598-019-46619-5

Wang, J., Tian, L., Lee, H.-S., Wei, N. E., Jiang, H., Watson, B., . . . Chen, Z. J. (2006). Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics, 172*(1), 507. 10.1534/genetics.105.047894

Wang, X., Zhang, H., Li, Y., Zhang, Z., Li, L., & Liu, B. (2016). Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA-sequencing. *New Phytologist, 209*(3), 1264-1277. 10.1111/nph.13678

Wen, J., Guo, P., Ke, Y., Liu, M., Li, P., Wu, Y., . . . Du, H. (2019). The auxin response factor gene family in allopolyploid *Brassica napus*. *PLOS One, 14*(4), e0214885. 10.1371/journal.pone.0214885

Wu, Y., Sun, Y., Wang, X., Lin, X., Sun, S., Shen, K., . . . Liu, B. (2016). Transcriptome shock in an interspecific F1 triploid hybrid of *Oryza* revealed by RNA sequencing. *Journal of Integrative Plant Biology*, *58*(2), 150-164. 10.1111/jipb.12357

Yoo, M. J., Szadkowski, E., & Wendel, J. F. (2013). Homeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity*, *110*, 171-180. 10.1038/hdy.2012.94

Zhang, C., Lin, C., Fu, F., Zhong, X., Peng, B., Yan, H., . . . Zhao, L. (2017). Comparative transcriptome analysis of flower heterosis in two soybean F1 hybrids by RNA-seq. *PLOS One*, *12*(7), e0181061. 10.1371/journal.pone.0181061

Zhang, J., Li, G., Li, H., Pu, X., Jiang, J., Chai, L., . . . Jiang, L. (2015). Transcriptome analysis of interspecific hybrid between *Brassica napus* and *B. rapa* reveals heterosis for oil rape improvement. *International Journal of Genomics, 2015*, 11. 10.1155/2015/230985

Zhang, L., Ma, C., Chao, H., Long, Y., Wu, J., Li, Z., . . . Li, M. (2019). Integration of metabolome and transcriptome reveals flavonoid accumulation in the intergeneric hybrid between *Brassica rapa* and *Raphanus sativus*. *Scientific Reports*, *9*(1), 18368. 10.1038/s41598-019-54889-2

Zhang, M., Liu, X.-K., Fan, W., Yan, D.-F., Zhong, N.-S., Gao, J.-Y., & Zhang, W.-J. (2018). Transcriptome analysis reveals hybridization-induced genome shock in an interspecific F1 hybrid from *Camellia*. *Genome, 61*(7), 477-485. 10.1139/gen-2017-0105

Figures

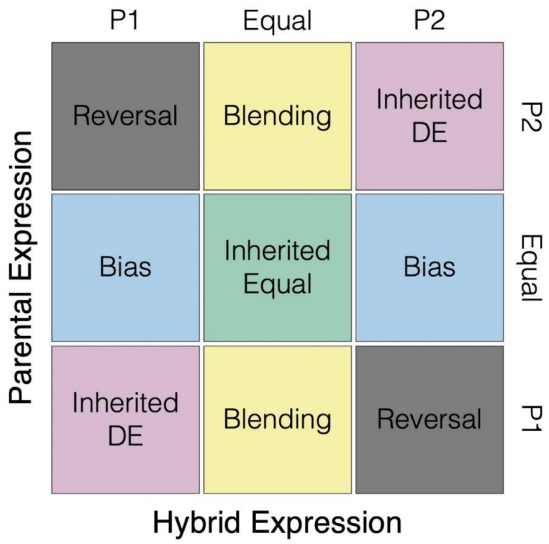

Figure 1: Fate of hybrid gene expression relative to parental expression. Following ortholog-ortholog (parental species) and homeolog-homeolog (hybrid species) differential expression analysis, each gene is defined as differentially expressed towards arbitrarily defined parent 1 (P1) or parent 2 (P2), or equally expressed. These nine hybrid expression outcomes can be grouped into five classes of hybrid gene expression, as indicated by the coloured boxes.

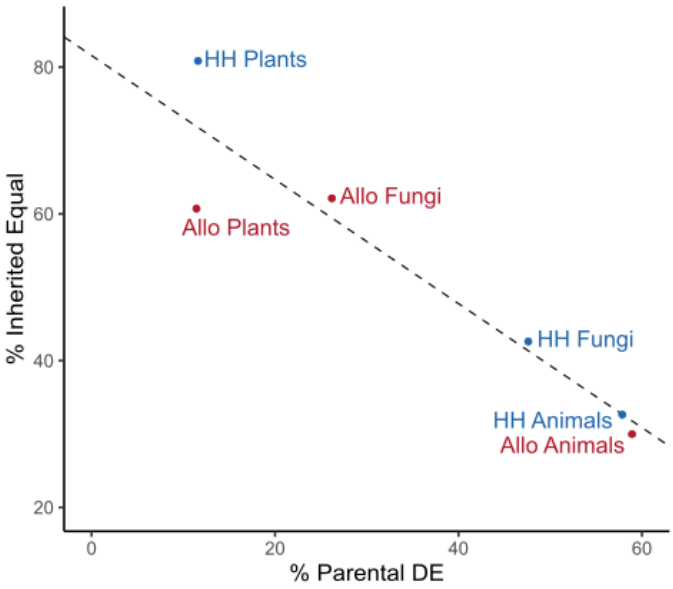

Figure 2: Inherited equal expression in hybrid species is dependent on the proportion of differential expression in the parent species. The percentage of differentially expressed parental orthologs plotted against the percentage of hybrid genes in the inherited equal category gives an inverse linear correlation across the study systems. The percentage of differentially expressed orthologs and homeologs was calculated relative to the total number of genes in the final datasets. Allo: allopolyploid; HH: homoploid hybrid.

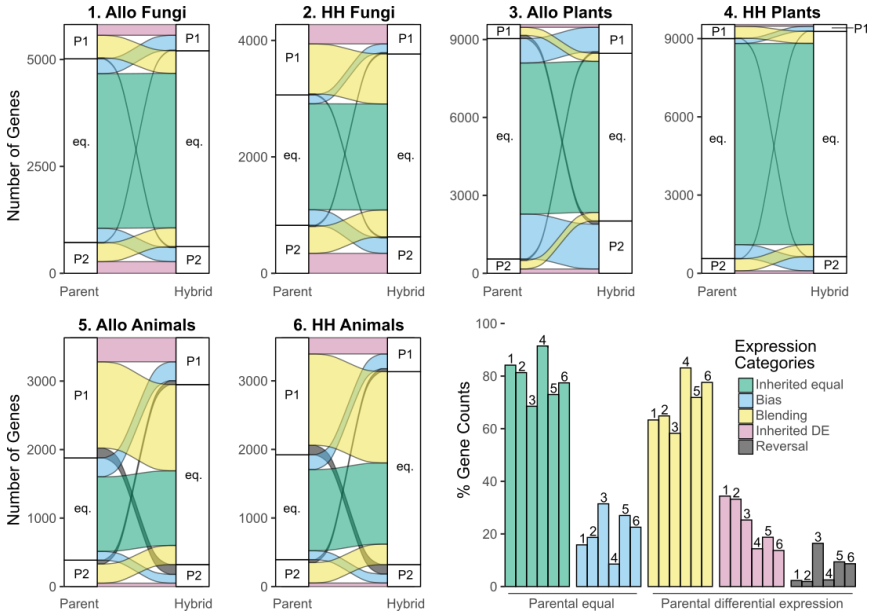

Figure 3: Genes with differential expression in parental species are more likely to have similar expression levels in hybrid species. Alluvial plots: the different fates of gene expression due to hybridisation are shown as alluvial plots 1-6. The heights of the white boxes represent the absolute numbers of genes with differential expression in either direction (P1 or P2; fold change > 2, adjusted $p < 0.05$) or equal expression (eq.). Note that the number of genes differs across the systems. Bar chart: the bar chart (bottom right) shows the relative proportions of equally expressed genes that are inherited or gain an expression bias, and the relative proportions of differentially expressed genes that are inherited, blended or reversed.

Differentially expressed parental genes are more likely to be blended in the hybrid (yellow) than inherit differential expression (pink), while equally expressed parental genes are more likely to be inherited as equal expression (green) than gain an expression bias in the hybrid (blue). Expression reversals (grey) are rare. Within each colour group, the bars are ordered to correspond with the alluvial plots, as indicated by the numbers above the bars. Allo: allopolyploid; HH: homoploid hybrid.

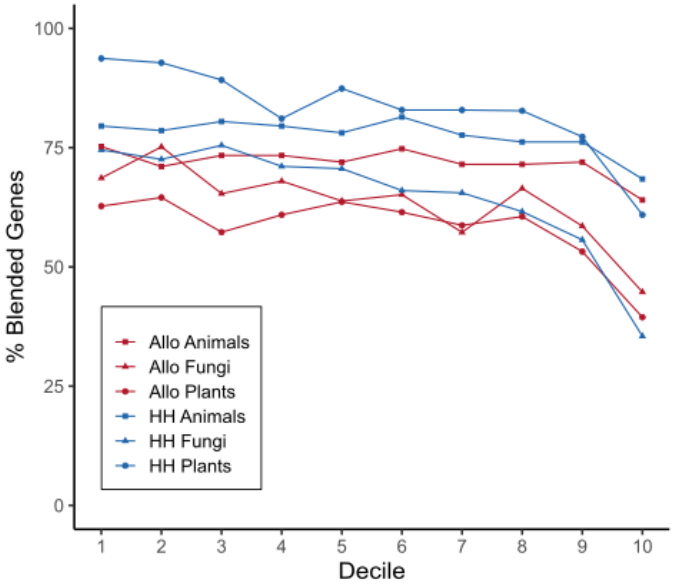

Figure 4: The proportion of blended genes is not dependent on the level of parental differential expression, except for the most highly differentially expressed genes. Genes were divided into deciles by parental expression difference, and the proportion of blended genes for each system are plotted for each decile. Allo: allopolyploid; HH: homoploid hybrid.

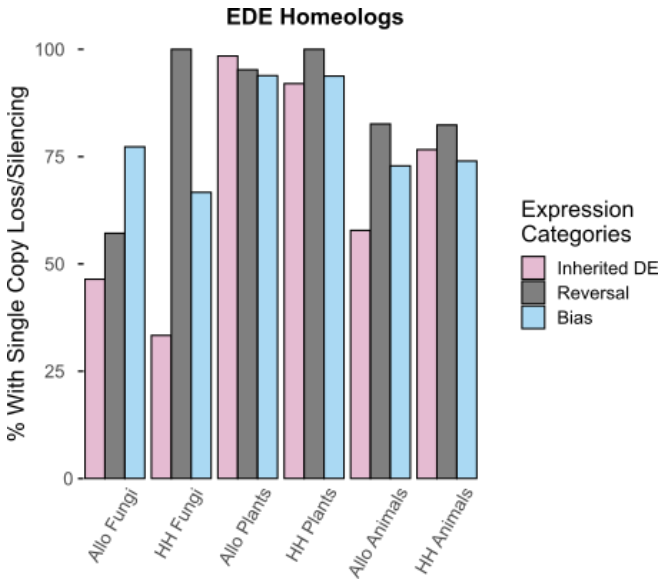

Figure 5: Most extremely differentially expressed homeologs are due to gene loss or silencing of one homeolog in the hybrid. Extreme differential expression was defined as > 50-fold difference in expression. Plots are shown separately for extremely differentially expressed homeologs in the inherited DE, bias and reversal categories. Allo: allopolyploid; HH: homoploid hybrid.

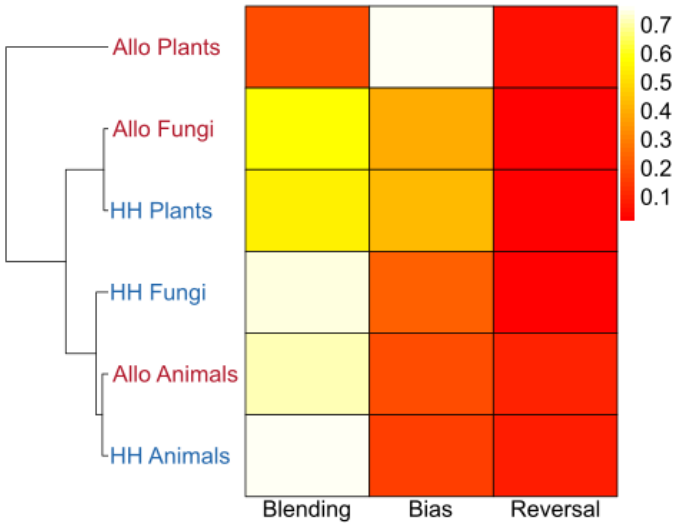

Figure 6: Change in genome ploidy does not strongly influence the distribution of hybrid gene expression patterns. Hierarchical clustering of non-inherited (blending, bias and reversal) expression category counts does not group systems based on whether they have undergone a change in genome ploidy or not. Category count data were normalised by the total number of genes in the three respective expression categories. The heatmap colour scale corresponds to the normalised count data. Allo: allopolyploid; HH: homoploid hybrid.

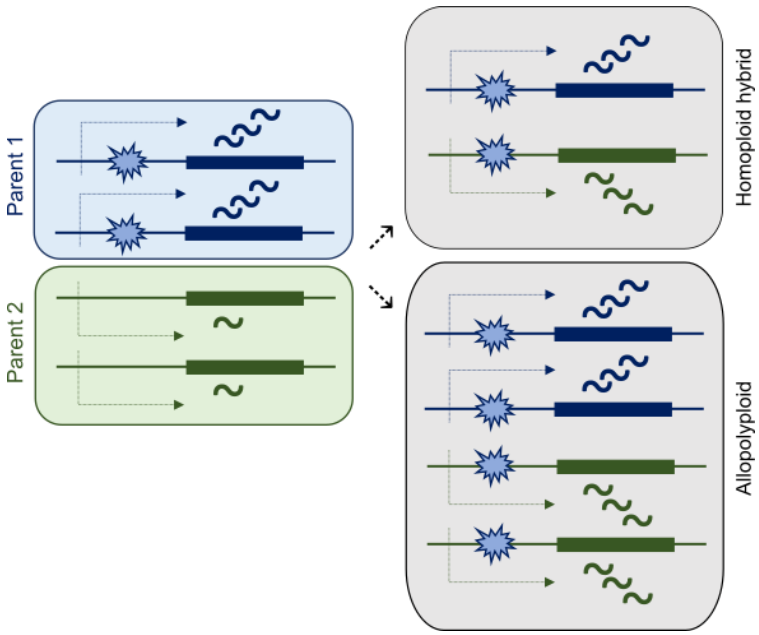

Figure 7: Blended hybrid gene expression is likely driven by interactions between homeologs and their regulatory factors. A representative gene in the nucleus of parent 1 (left) shows *trans*-activation via transcription factor binding (blue star bursts) and subsequent high levels of expression (wiggly lines) of the gene. In contrast, the lack of this transcription factor in parent 2 results in low levels of expression of the orthologous gene in parent 2. When both parental gene copies and the transcription factor are combined in a single homoploid hybrid nucleus (top right) or allopolyploid nucleus (bottom right), the ability of the transcription factor to bind to both parentally-derived homeologs results in a more similar, or 'blended', level of expression.

Tables

Table 1. Representative allopolyploid and homoploid hybrid study systems used for cross-kingdom analysis.

	Allopolyploid	Homoploid Hybrid
Fungi	*Epichloë canadensis*	*Saccharomyces cerevisiae* x *paradoxus*
Plants	*Gossypium hirsutum* TX2094	*G. arboreum* x *raimondii*
Animals	*Squalius alburnoides* (PAA)	*Sq. alburnoides* (PA)

Table 2. Read mapping by system. Number of reads (millions) mapped to each biological replicate in each representative system. The number of reads mapped against the *Epichloë* species is substantially lower than the total number of RNA-seq reads as the *Epichloë* endophytes were grown *in planta* to measure expression in their natural state, thus the majority of reads in the RNA-seq libraries were from their grass host.

	Replicate	n Reads (millions)	n Reads Mapped (millions)	% Reads Mapped
Epichloë elymi	1	313	1.4	0.4
E. elymi	2	309	0.6	0.2
E. amarillans	1	318	1.2	0.4
E. amarillans	2	315	1.2	0.4
E. canadensis	1	304	2.7	0.9
E. canadensis	2	311	8.1	2.6
Saccharomyces cerevisiae	1	35.3	13.4	38.0
Sa. cerevisiae	2	32.1	12.1	37.7
Sa. paradoxus	1	35.1	11.1	31.6
Sa. paradoxus	2	35.8	11.5	32.1
Sa. cerevisiae x *paradoxus*	1	31.9	11.4	35.7
Sa. cerevisiae x *paradoxus*	2	30.7	10.6	34.5
Gossypium arboreum	1	9.0	5.8	64.4
G. arboreum	2	5.4	3.8	70.3
G. raimondii	1	8.2	5.5	67.1
G. raimondii	2	5.8	3.8	65.5
G. hirsutum TX2094	1	7.5	5.1	68.0
G. hirsutum TX2094	2	8.4	5.3	63.1
G. arboreum x *raimondii*	1	7.1	4.8	67.6
G. arboreum x *raimondii*	2	6.2	4.0	64.5
Squalius pyrenaicus	1	22.7	10.1	44.5
S. pyrenaicus	2	22.7	9.9	43.6
Sq. alburnoides	1	27.8	14.2	51.1
Sq. alburnoides	2	27.8	13.8	49.6
Sq. alburnoides (PAA)	1	28.1	15.6	55.5
Sq. alburnoides (PAA)	2	28.1	15.2	54.1

	Sq. alburnoides (PA)	1	20.9	10.5	50.2
	Sq. alburnoides (PA)	2	20.9	10.3	49.3

1

2

Table 3. Extreme differential gene expression is more prevalent among homeologs (in the hybrid) than orthologs (in the parent species). Numbers of extremely differentially expressed (EDE) orthologs and homeologs in each representative system. Allo: allopolyploid; HH: homoploid hybrid.

	EDE Orthologs	EDE Homeologs
Allo Fungi	15	79
HH Fungi	1	10
Allo Plants	0	746
HH Plants	1	246
Allo Animals	33	168
HH Animals	30	207

3.1 Validations

3.1.1 Testing the validity of HyLiTE read count data

To validate the accuracy of the read count matrices generated by HyLiTE, a two-stage check of stringent read mapping followed by manual read assignment was performed. As this validation was not system-specific, the homoploid hybrid *Gossypium* dataset was chosen at random. First, to ensure that only high quality reads that exceeded a certain length were used in stringent mapping (poor quality reads would not give an accurate depiction of read count distributions, while short reads could map ambiguously to a number of genomic regions), fastq sequence quality control was performed with the SolexaQA v3.1.7.1 [129] 'dynamictrim' function, filtering bases with a score of Q30 or higher. Q30-scoring reads were subsequently trimmed to greater than or equal to 50 bp, using the SolexaQA 'lengthsort' function, before being mapped to the reference gene sequences. As the 'lengthsort' function can only take one single-end read file as an argument each time, the Python script auto_trim.py (`https://github.com/annabehling/multi_bowtie`) was written to automate this step for all files.

Mapping of the quality filtered and trimmed reads was performed with Bowtie 2. Two further Python scripts, also available in the above GitHub repository, were written to process fastq files for troubleshooting more efficiently. Following Bowtie 2 indexing of the reference gene models, the first script, multi_bowtie.py, was used to generate SAM files for a directory of fastq files. By specifying '–score-min C,0,0' in the code, mapping was performed with zero mismatches. The second script, index_sam.py, was used to circumvent the inability of samtools v1.7 [130] to give index statistics directly from SAM files, by automating the sequential use of the sort, index and idxstats functions. The script multi_bowtie_vis.R was written to process the resulting TSV files in R v4.0.2 [131].

The regression analysis of the HyLiTE read count data and Bowtie 2 stringent mapping read count data produced high correlation coefficients for both biological replicates of each constituent species of the hybrid system (Figure 2). The highest correlation coefficients (r= 0.99, 0.98) were for the reads belonging to *G. arboreum*. The *G. raimondii* reads gave the lowest correlation coefficients (r= 0.82, 0.83), while the homoploid hybrid correlation coefficients fell in the middle (r=0.95, 0.96). This was expected as the reference gene models were from *G. arboreum*. All regression analyses showed limited variation from a linear trend line: a promising indicator of confidence in the correlation coefficients.

The regression analyses are not centered on the linear trend lines because the high stringency mapping was performed with no mismatches, in addition to the preliminary quality filtering.

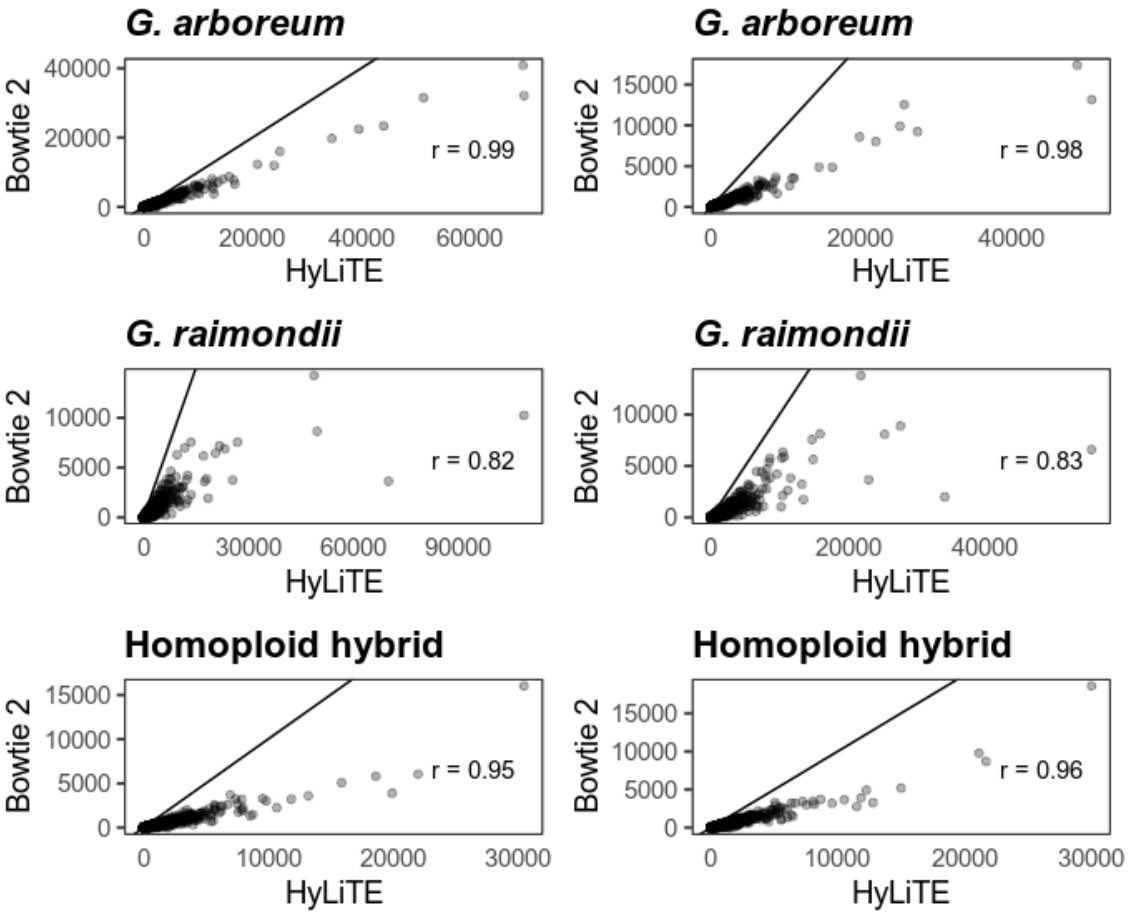

Figure 2: **Regression analysis of Bowtie 2 stringent mapping and HyLiTE read count data produces high correlation coefficients.** Data shown is for two replicates (L-R) of the parental species *Gossypium arboreum* and *G. raimondii*, and their homoploid hybrid. Points on the plots have 30% opacity, to emphasise overlap. Each plot shows a linear trend line (x = y).

Several genes from the stringent mapping files were further assessed in IGV v2.8.2 [132, 133], to check that the reads looked like they were mapping well, and with no mismappings over sites that have SNP differences between the parents. This process was manual and slow, and therefore limited to only a small number of genes. Three candidates were chosen from the approximately 35,000 cleaned *G. arboreum* gene models; their suitability assessed according to a number of selection criteria:

- The genes needed to have 150-200 reads mapped in the homoploid hybrid replicate 1 (obligatory)
- Of the 150-200 reads mapped, at least 100 in total needed to be assigned to parent 1 (*G. raimondii*) and parent 2 (*G. arboreum*) (obligatory)
- One gene was needed from each of three categories: high-level expression, mid-level expression and low-level expression in parent 1, relative to parent 2 (obligatory)
  - High-level expression: 75-100% of parental reads mapped to parent 1
  - Mid-level expression: 45-55% of parental reads mapped to parent 1
  - Low-level expression: 0-25% of parental reads mapped to parent 1
- Genes were longer than 1000 bp (preferential)

If too few reads were mapped to candidate genes, it would make the margin of error for parental counts too great; too many reads and their manual assignment became impractical. To further reduce the margin of error for determining parental origin, over half of the reads mapped to a gene had to be parentally-assigned; many genes exhibited a majority of reads mapped with unknown origin. The genes were preferentially of a length greater than 1000 bp to increase the chance of having a high number of diagnostic SNPs and enable accurate validation.

An R script, igv_picker.R, was written to generate short-lists of candidate genes in the high, mid and low expression categories that met all obligatory criteria. IGV was subsequently used to identify which candidate genes were greater than 1000 bp, or as close as possible in the instance where no candidates in an expression level category met the preferential criterion. One gene from each expression level was subsequently chosen for manual mapping (Table 1).

To confirm that the stringently mapped reads were mapping with no mismatches, IGV was first used to view the alignment of the corresponding BAM files for replicate 1 of *G. raimondii*, *G. arboreum* and the homoploid hybrid against each of the three genes, with the default setting 'show mismatched bases' selected. No SNPs (mismatches) were found in any reads aligned to the three genes chosen for manual mapping.

Next, parental assignment of the homoploid hybrid HyLiTE reads was performed through the identification of diagnostic SNPs in the parental HyLiTE reads. HyLiTE BAM files for replicate 1 of both parents and the hybrid were used to deduce parental origin for each read, as the SNPs were required for read assignment and these were not present in the stringent mapping BAM files. The parental assignment of reads was performed under the following conditions:

- If a hybrid read contained one or more SNPs present in only one parent, the read was assigned to that parent.
- If the above condition was fulfilled, but the hybrid read contained one or more additional SNPs not present in either parent, the above parental assignment was maintained.
  - Note: in HyLiTE this would be classified as 'parent+N', which for this manual read assignment was counted in with the relevant parental tally.
- If there was an absence of SNPs in a hybrid read, and one parent also had one or more reads without SNPs aligned to the same genic region, the read was assigned to that parent.

Conversely, parental assignment could not be made under the following conditions:

- If the absence of SNPs in a hybrid read was matched in the same genic region by neither or both parents.
- If a hybrid read had only hybrid-specific SNPs.
- If a hybrid read had parental SNPs that were found in the same quantity in both parents.

The proportions of parent 1 read assignment obtained from the HyLiTE and manual IGV approaches were identical for the high-level expression gene (100%, 100%), highly similar

for the low-level expression gene (24.3%, 27.9%) and similar for the mid-level expression gene (45%, 30%) (Table 2). The level of these similarities support the validity of the HyLiTE read count data. Differences between the respective proportions are likely to have arisen through differences between the assignment approaches. HyLiTE performs automatic masking of SNPs with low statistical support, based on a binomial distribution conditioned on the read coverage level. This feature improves the accuracy of read assignment to homeologs for genes with low to mid level expression [95].

Gene	Gene size (bp)	n reads	n P1 reads*	n P2 reads*	% P1 reads	P1 expression
lcl|NC_030674.1_cds_XP_017610228.1_35580	1012	170	114	0	100	high
lcl|NC_030664.1_cds_XP_017626417.1_3301	1074	180	45	55	45	mid
lcl|NC_030671.1_cds_XP_017643504.1_25816	827	172	34	106	24.3	low

Table 1: **Genes chosen for IGV manual read assignment.** *Parental read totals are inclusive of the parent+N reads.

Gene	Method	n P1 reads	n P2 reads	% P1 reads
lcl|NC_030674.1_cds_XP_017610228.1_35580	HyLiTE	114	0	100
	IGV	91	0	100
lcl|NC_030664.1_cds_XP_017626417.1_3301	HyLiTE	45	55	45
	IGV	29	69	30
lcl|NC_030671.1_cds_XP_017643504.1_25816	HyLiTE	34	106	24.3
	IGV	41	106	27.9

Table 2: **Comparison of parental read assignment in HyLiTE (automated) and IGV (manual).**

3.1.2 Testing the arbitrariness of the parental definition

Parental definitions ('parent 1', 'parent 2') were required at multiple stages during differential expression analyses and gene classification. By default, R takes the factor of the lowest value (i.e. the parent whose name comes first alphabetically) as the reference level when performing differential expression analyses with DESeq2 [134]. Consequently, levels were set in the relevant code.

To test that setting the levels made the parental definition arbitrary, the relevant code for the homoploid hybrid plant analyses were run through, firstly with *G. arboreum* as parent 1 in all instances and secondly with *G. raimondii* as parent 1 in all instances. The gene classification numbers derived from each analysis were compared. The parental definition was determined to be arbitrary (`https://github.com/annabehling/DEA_and_fit`). The following species of each genus were subsequently defined as parent 1. Note, parent 1 is not necessarily the source of the gene models.

- ***Epichloë*:** *E. elymi*
- ***Saccharomyces*:** *S. paradoxus*
- ***Gossypium*:** *G. raimondii*
- ***Squalius*:** *S. pyrenaicus*

3.1.3 Testing for extraneous effects of no independent filtering

As the gene classification code incorporated p value and expression fold change-based filtering, it was not necessary to use the default independent filtering parameter when viewing the DESeq2 differential expression analysis results. This also enabled user-defined calculation of the adjusted p value, which was achieved through the base-R function 'p.adjust()' and the application of the Benjamini–Hochberg procedure to control the false discovery rate.

To confirm that no other values, aside from the adjusted p value, were altered when removing independent filtering, the results of a differential expression analysis were viewed with and without independent filtering; the two sets of columns (aside from adjusted p value) compared and found to be equal (`https://github.com/annabehling/DEA_and_`

`fit`).

3.1.4 Testing the accuracy of expression category assignment

The classification of genes into each expression category was based on the HyLiTE read count data, specifically, on the read.summary.txt and expression.txt output files.

The **expression.txt** file contains gene-ordered information about the expression level of each parent and hybrid biological replicate.

The **read.summary.txt** file contains gene-ordered information about the parental origin (P1 or P2) of every read for every gene in the hybrid. If there is inadequate diagnostic information to make a read assignment, the read is classified as unknown or uninformative.

- **Unknown:** the read contains only masked or hybrid-specific SNPs, and could be assigned to neither parent
- **Uninformative:** the read contains only ambiguous SNPs, and could be assigned to either parent

To test that genes were being accurately classified into each of the expression categories, genes with mock HyLiTE read count data matching each of the four expression categories (Table 3, Table 4) were submitted to the downstream R pipeline. The correct expression category assignment occurred for all four genes.

GENE	P1%rep1	P1%rep2	HH%rep1	HH%rep2	P2%rep1	P2%rep2
pei_dummy	1000	1000	1100	1100	100	100
hebl_dummy	1000	1000	1100	1100	100	100
hebi_dummy	550	550	1100	1100	550	550
her_dummy	1000	1000	1100	1100	100	100

Table 3: **Mock HyLiTE expression.txt file** for genes predicted to be classified in the inheritance ('pei_dummy'), blending ('hebl_dummy'), bias ('hebi_dummy') and reversal ('her_dummy').

GENE	P1	P1+N	P2	P2+N	UNINFORMATIVE	UNINFORMATIVE+N	UNK	UNK+N
pei_dummy	1000	0	100	0	0	0	0	0
hebl_dummy	550	0	550	0	0	0	0	0
hebi_dummy	1000	0	100	0	0	0	0	0
her_dummy	100	0	1000	0	0	0	0	0

Table 4: **Mock HyLiTE read.summary.txt file** for genes predicted to be classified in the inheritance ('pei_dummy'), blending ('hebl_dummy'), bias ('hebi_dummy') and reversal ('her_dummy'). The same mock file was duplicated for replicate 1 and 2.

3.2 Additional analyses

3.2.1 No shared patterns of expression among transgressively expressed genes

Transgressive expression was defined as mean hybrid expression greater than two times the highest (parent 1 or parent 2) mean parental expression, or less than half of the lowest (parent 1 or parent 2) mean parental expression. Read count data was normalised by the total expression level for each biological replicate. Based on this definition, genes with transgressive expression could fall under any of the five expression categories. Therefore, to visualise shared patterns among genes with transgressive expression, relative expression-category counts for transgressively expressed genes were compared to the relative expression-category counts for all genes in the final gene sets. The general similarity between the two sets suggested that there were no novel expression trends shared by transgressively expressed genes (Figure 3).

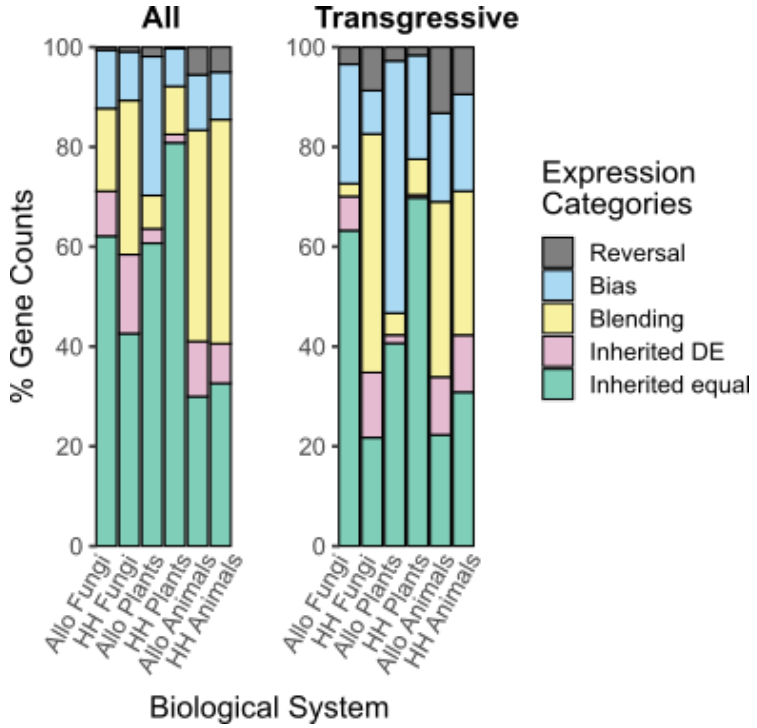

Figure 3: **Relative expression category counts for genes with transgressive expression are similar to those for all genes.** Transgressive expression was defined as mean hybrid expression $>$ 2x the highest mean parental expression, or $<$ 0.5x the lowest mean parental expression, following the normalisation of read count data.

3.2.2 topGO enrichment of GO terms

In addition to the functional enrichment analyses performed with the DAVID bioinformatics resource [135, 136], further functional enrichment analyses were performed using topGO v2.40.0 [137]. GO term enrichment analyses were performed for the blending, bias, reversal and inherited differential expression categories. The respective custom filtered annotations produced or obtained in the initial steps were provided as input to produce the topGO data objects. Only genes in the final, filtered datasets were used as the background, as these were the only genes that had any chance of having an expression category classification. Enrichment testing of topGO data objects implemented Fisher's exact test and the default 'weight01' algorithm. The default algorithm was used as the performance of each topGO algorithm on empirical data has been shown to be highly subjective [138]. The resulting tables were filtered by $p < 0.05$.

The representative systems used in this study ranged from model organisms (*Saccharomyces cerevisiae*) to those with minimal database representation (*Squalius* spp.). Subsequently, a range of precursory steps were taken in advance of functional enrichment analysis, specific to each study system.

***Epichloë*:**

- Amino acid sequences for the *Epichloë* gene models were submitted to Pannzer2 (protein annotation with Z-score) [139–143] to predict GO term annotations.
- GO annotations were produced using ARGOT (annotation retrieval of gene ontology terms) [144] and filtered with a positive predictive value (PPV) > 0.5. This value produces reliable functional annotations for well-characterised *Epichloë* genes without introducing false positives [145].

***Saccharomyces*:**

- To obtain GO annotations for the *S. cerevisiae* gene models, Ensembl Biomart [146, 147] could be used, as the Ensembl Genes 100 database had an *S. cerevisiae* (R64-1-1) dataset.
- A file containing all RefSeq NM accession numbers from the fasta gene model headers was submitted as input. These NM numbers correlated to RefSeq DNA IDs [148]

that were used to generate GO term accessions, and KEGG pathway and enzyme IDs.

- The output data was subsequently manipulated into a gene-to-GO format in R.

***Gossypium*:**

- Similar to *Epichloë*, the *G. arboreum* translated CDS sequences were submitted to Pannzer2 to predict GO term annotations using ARGOT and filtered with a PPV > 0.5.

***Squalius*:**

- The *Squalius* GO annotations were extracted from a prior annotation in the literature [86] and used as input for generating the topGO enrichment tables.

However,unlike DAVID, topGO does not allow the user to test functional enrichment whilst controlling the GO hierarchy level, a feature essential for comparative analyses of species with markedly different functional profiles. Therefore, functional enrichment analyses were performed using the DAVID bioinformatics platform, presented in the above manuscript, and the topGO analyses were not pursued further.

3.2.3 DAVID enrichment of KEGG pathways and COG orthologs

The DAVID bioinformatics platform was also used to test for the enrichment of KEGG pathways [149–151] and COG orthologs [152] among the expression category gene subsets. Enrichment of KEGG pathways and COG orthologs did not occur across all systems, so further comparisons could not be made.

3.2.4 Functional enrichment of GO terms among extremely differentially expressed genes

The full gene sets of the six representative systems yielded no clear pattern of shared functional enrichment in all GO ontologies, KEGG pathways on COG orthologs, in part due to low power where there were a limited number of genes in the gene subsets tested

for enrichment. As the extremely differentially expressed genes represented only a small proportion of the full gene sets, DAVID functional enrichment analysis among extremely differentially expressed genes was not performed.

4 Conclusion

4.1 The motivation behind this project

Hybridisation occurring with (allopolyploidisation) and without (homoploid hybridisation) concomitant ploidy increase pervades the eukaryote tree of life [89, 108, 153–155]. Due to the nature of their formation, all hybrids face a similar suite of challenges, triggered near-instantaneously and in the longer term, following the merger of two diverged parental subgenomes within a single hybrid nucleus. These challenges permeate the hybrid genome [56, 60], transcriptome [70], proteome [156, 157] and metabolome [158], with potentially catastrophic consequences. Nonetheless, hybrid species have developed a range of biological responses that enable them to not only survive these challenges, but to thrive. Indeed, hybrid species have often been associated with heterosis or 'hybrid vigour' that can facilitate invasiveness and the ability to dominate in their parental niche [49, 159].

Allopolyploid and homoploid hybrid species are integral to numerous industries, both globally and locally. They are commonly found amongst horticultural crop species [16–29], in the agricultural industry as vital components of pastoral health and arable crop yield [39, 41, 42], and in the aquaculture [35, 160] and brewing [32, 161] industries. Their widespread importance has driven countless studies of their biology, yet, the field has noticeably lacked the synthesis of individual hybrid studies to elucidate common trends in the response to hybridisation. The generalised framework used in this study enabled the comparison of the genome-wide gene expression profiles of hybrid species from animals, plants and fungi. One homoploid hybrid species from each kingdom was paired with a respective allopolyploid, enabling the dissociation of increased ploidy from genome merger on post-hybridisation expression patterns, and giving a total of six representative study systems.

4.2 Primary research findings

Following the integration of the results from two sets (parental and hybrid) of differential expression analyses, each gene was classified into one of five expression categories. This methodology circumvented the limitation of studying only the universal eukaryote genes in each gene set, for example using BUSCO [162], instead enabling the broad comparison of the global transcriptomic response to hybridisation. More specifically, this study

investigated the 'fate' of genes following hybridisation. This novel evolutionary framing in the existing literature, when paired with the broad comparative approach, identified a number of trends underlying the gene expression patterns of hybrids.

Genes that are differentially expressed in parental species were found to preferentially have more similar, 'blended', expression in the hybrid descendants. This observation was consistent with a previously proposed mechanism where patterns of hybrid gene expression are primarily driven by interactions between homeologs and their regulatory factors [70]. The preferential blending trend was observed in all but the most differentially expressed parental genes. Interestingly, extreme differential expression was found to be far more common between hybrid gene copies than parental genes. This appeared to be a consequence of frequent loss or silencing of one homeologous copy, both of which are important genomic consequences of hybridity. Moreover, when directly comparing the expression profiles of allopolyploid and homoploid hybrids, hybridisation was found to have a greater effect than increased ploidy on gene expression, across all three kingdoms.

4.3 Limitations and future research directions

According to the data, the above trends do appear general in nature, presenting strongly and statistically in allopolyploid and homoploid hybrids from animals, plants and fungi. The representative systems used in this study varied substantially in the proportion of genes with differential expression between the parents (ranging from 11.4% to 58.9%). Thus, one could reasonably apply such trends to the transcriptomic analysis of other eukaryote hybrids, irrespective of their kingdom, ploidy, or level of parental differential expression. Although these six representative systems enabled the elucidation of cross-kingdom commonalities, further testing with an increased number of biological systems is required to consolidate the degree of generality of such trends. It is also possible that further analyses will elucidate further species-specific features of hybrid gene expression. Here, the functional enrichment of expression category genes found no shared GO terms across all six systems. However, this was likely also a consequence of some expression categories having minimal representation among genes. Furthermore, for three of the representative systems, functional enrichment could only be measured among orthologs from closely related species, as genes from those particular species were not present in the DAVID knowledgebase. There was another inconsistency among the representative systems, as an intra-genus allopolyploid and homoploid hybrid pairing was not possible

within fungi (due to a lack of available RNA-seq data), although it is possible that this may have bolstered the generality of the trends observed, suggesting that such trends are not strictly limited to intra-genus pairings.

Additional future analyses could include polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) of genomic DNA to investigate whether the observed absence of read mapping among extremely differentially expressed homeologs is due to gene loss or complete silencing. One previous study of an allopolyploid *Epichloë* system suggested that it is gene loss, rather than regulatory changes, that explain the major differences in homeolog expression [70]. It would also be interesting to explore just how impactful the extent of parental divergence is on the hybrid's ability to blend parental differential expression. The proposed mechanism underlying blending is dependent on regulatory factors from one parent activating the expression of the other parental homeolog. This suggests that too much parental divergence could prevent the regulatory factor of one parental species recognising the binding site of the other. Taking the level of parental differential expression as a proxy for parental divergence, one might expect that orthologs with extreme differences in expression would exhibit a reduction in blending. While the data did show a drop in the proportion of genes with blended expression among the most highly differentially expressed genes, this did not appear to be the result of genes with extreme differential expression having particular difficulty blending. It remains unclear what was driving the reduction in the proportion of blending among the most differentially expressed parental genes. A comparative transcriptomic study of hybrid systems with a wide range of parental divergences may clarify this point.

4.4 Local motivation for continued hybrid research

This research also incorporated a review on the importance and prevalence of native and introduced allopolyploid species in Aotearoa New Zealand [120]. Many of these species drive benefits in the same industries as their global counterparts. While polyploidy has been inferred in the origins of three of the surviving crop species brought to New Zealand by Māori, research to-date suggests they are all autopolyploids [163–165]. Instead, it is among introduced species of agricultural and horticultural importance that numerous confirmed and possible allopolyploid species are found [28, 29, 154, 166–174]. The review also found that the longstanding brewing and viticultural practices of New Zealand could benefit from future discoveries of local allopolyploid strains, due to a purported link that

exists between the ploidy level of *Saccharomyces* species and culture productivity [175].

The 2018 collective export value of New Zealand goods that derived directly or indirectly from systems reliant on these and other allopolyploid species was calculated to be in excess of 23 billion New Zealand dollars [176]. It was this figure that, for me, really highlighted the tangible impact of allopolyploid species on our society, particularly when considering that the figure excluded the contribution of New Zealand's largest economic sector, tourism. The New Zealand tourism industry is undoubtedly bolstered by the native flora and fauna, which includes allopolyploid mosses [43, 44], ferns [45, 46] and angiosperms [47, 48], as well as a putative mosaic allotriploid stick insect [177, 178]. Of course, this review did not consider the impact of homoploid hybrid species, a possible and worthwhile avenue for a future review to explore. Nonetheless, when considering the present body of research, it is clear that further research on hybrid species will only continue to benefit humankind, both globally, and here in Aotearoa New Zealand, in the future.

5 Code

Scripts used to validate the HyLiTE read count data can be found at `https://github.com/annabehling/multi_bowtie`. This repository also contains code used to generate the short-lists of candidate genes in the high, mid and low expression categories for manual IGV mapping.

Scripts used to process the HyLiTE read count matrices, perform the differential expression analyses, fit the regression models, and classify each gene into the expression categories can be found at `https://github.com/annabehling/DEA_and_fit`. Code is provided for the usage of these scripts on all six representative system read count matrices. This repository also contains scripts used to identify extremely differentially expressed and transgressively expressed genes, in addition to further validation code.

Scripts used to process the raw DAVID functional enrichment analysis results, and to produce and process the topGO functional enrichment analysis results can be found at `https://github.com/annabehling/functional_enrichment`.

6 Data

All raw genomic and RNA-seq data used in these analyses are publicly available. Links to these repositories can be found in Supplementary File 2, at `https://github.com/annabehling/masters_supplementary`. All other supplementary files accompanying the manuscript in Chapter 3 are also available at this link.

Example TSV files produced in read count validation, and the HyLiTE expression.txt file and read.summary.txt files used in determination of gene candidates for manual read assignment can be found at `https://github.com/annabehling/multi_bowtie/tree/master/example_outfiles`.

All parental and hybrid HyLiTE read count matrices used in the analysis of all representative systems are available at `https://github.com/annabehling/DEA_and_fit/tree/master/all_count_matrices`.

All foreground and background gene lists used in the DAVID functional analyses, and subsequent output files, are available at `https://github.com/annabehling/functional_enrichment/tree/master/all_DAVID_files`.

References

1. Muñoz, A. G., Salazar, C., Castaño, J., Jiggins, C. D. & Linares, M. Multiple sources of reproductive isolation in a bimodal butterfly hybrid zone. *Journal of Evolutionary Biology* **23,** 1312–1320 (2010).

2. Dickinson, G. R., Lee, D. J. & Wallace, H. M. The influence of pre- and post-zygotic barriers on interspecific *Corymbia* hybridization. *Annals of Botany* **109,** 1215–1226 (2012).

3. Brix, K. V. & Grosell, M. Evaluation of pre- and post-zygotic mating barriers, hybrid fitness and phylogenetic relationship between *Cyprinodon variegatus variegatus* and *Cyprinodon variegatus hubbsi* (Cyprinodontiformes, Teleostei). *Journal of Evolutionary Biology* **26,** 854–866 (2013).

4. Sekine, D. *et al.* Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *The Plant Journal* **76,** 792–799 (2013).

5. Masello, J. F. *et al.* Additive traits lead to feeding advantage and reproductive isolation, promoting homoploid hybrid speciation. *Molecular Biology and Evolution* **36,** 1671–1685 (2019).

6. Coughlan, J. M., Wilson Brown, M. & Willis, J. H. Patterns of hybrid seed inviability in the *Mimulus guttatus* sp. complex reveal a potential role of parental conflict in reproductive isolation. *Current Biology* **30,** 83–93.e5 (2020).

7. Morgan, E. J. *et al.* Disentangling the components of triploid block and its fitness consequences in natural diploid–tetraploid contact zones of *Arabidopsis arenosa*. *New Phytologist* (2021).

8. Nakata, K. *et al.* Analysis of the possible cytogenetic mechanism for overcoming hybrid lethality in an interspecific cross between *Nicotiana suaveolens* and *Nicotiana tabacum*. *Scientific Reports* **11,** 7812 (2021).

9. Schumer, M., Cui, R., Rosenthal, G. G. & Andolfatto, P. Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLOS Genetics* **11,** e1005041 (2015).

10. Baiz, M. D., Tucker, P. K. & Cortés-Ortiz, L. Multiple forms of selection shape reproductive isolation in a primate hybrid zone. *Molecular Ecology* **28,** 1056–1069 (2019).

11. Barton, N. H. The role of hybridization in evolution. *Molecular Ecology* **10,** 551–568 (2001).

12. Fan, S. & Meyer, A. Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. *Frontiers in Genetics* **5,** 163 (2014).

13. Grant, B. R. & Grant, P. R. Watching speciation in action. *Science* **355,** 910 (2017).

14. Mallet, J. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20,** 229–237 (2005).

15. Servedio, M. R. & Hermisson, J. The evolution of partial reproductive isolation as an adaptive optimum. *Evolution* **74,** 4–14 (2020).

16. He, S. *et al.* Genomic divergence in cotton germplasm related to maturity and heterosis. *Journal of Integrative Plant Biology* **61,** 929–942 (2019).

17. Whitford, R. *et al.* Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *Journal of Experimental Botany* **64,** 5411–5428 (2013).

18. Cui, Y. *et al.* Hybrid breeding of rice via genomic selection. *Plant Biotechnology Journal* **18,** 57–67 (2020).

19. Duvick, D. N. Biotechnology in the 1930s: the development of hybrid maize. *Nature Reviews Genetics* **2,** 69–74 (2001).

20. Singh, S., Dey, S. S., Bhatia, R., Kumar, R. & Behera, T. K. Current understanding of male sterility systems in vegetable Brassicas and their exploitation in hybrid breeding. *Plant Reproduction* **32,** 231–256 (2019).

21. Rodríguez, F., Ghislain, M., Clausen, A. M., Jansky, S. H. & Spooner, D. M. Hybrid origins of cultivated potatoes. *Theoretical and Applied Genetics* **121,** 1187–1198 (2010).

22. De Maagd, R. A. *et al.* CRISPR/Cas inactivation of RECQ4 increases homeologous crossovers in an interspecific tomato hybrid. *Plant Biotechnology Journal* **18,** 805–813 (2020).

23. Taliercio, E., Eickholt, D., Rouf, R. & Carter, T. Changes in gene expression between a soybean F1 hybrid and its parents are associated with agronomically valuable traits. *PLOS One* **12,** e0177225 (2017).

24. Song, J. *et al.* Natural allelic variations in highly polyploidy *Saccharum* complex. *Frontiers in Plant Science* **7,** 804 (2016).

25. Radanović, A., Miladinović, D., Cvejić, S., Jocković, M. & Jocić, S. Sunflower genetics from ancestors to modern hybrids-a review. *Genes* **9,** 528 (2018).

26. Fávero, A. P. *et al.* Transference of multiple resistance to peanut through the development of cross-compatible complex hybrids of wild *Arachis*. *Genetics and Molecular Biology* **43,** e20190099 (2020).

27. Lashermes, P., Hueber, Y., Combes, M.-C., Severac, D. & Dereeper, A. Intergenomic DNA exchanges and homeologous gene silencing shaped the nascent allopolyploid coffee genome (*Coffea arabica* L.) *G3 (Bethesda)* **6,** 2937–2948 (2016).

28. Liu, Q., Lin, L., Zhou, X., Peterson, P. M. & Wen, J. Unraveling the evolutionary dynamics of ancient and recent polyploidization events in *Avena* (Poaceae). *Scientific Reports* **7,** 41944 (2017).

29. Bindler, G. *et al.* A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development. *Theoretical and Applied Genetics* **123,** 219–230 (2011).

30. Casaregola, S., Nguyen, H. V., Lapathitis, G., Kotyk, A. & Gaillardin, C. Analysis of the constitution of the beer yeast genome by PCR, sequencing and subtelomeric sequence hybridization. *International Journal of Systematic and Evolutionary Microbiology* **51,** 1607–1618 (2001).

31. Borneman, A. R., Zeppel, R., Chambers, P. J. & Curtin, C. D. Insights into the *Dekkera bruxellensis* genomic landscape: comparative genomics reveals variations in ploidy and nutrient utilisation potential amongst wine isolates. *PLOS Genetics* **10,** e1004161 (2014).

32. Marcet-Houben, M. & Gabaldón, T. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLOS Biology* **13,** e1002220 (2015).

33. Huang, X. *et al.* Fertilization and cytogenetic examination of interspecific reciprocal hybridization between the scallops, *Chlamys farreri* and *Mimachlamys nobilis*. *PLOS One* **6,** e27235 (2011).

34. Xiao, J. *et al.* Coexistence of diploid, triploid and tetraploid crucian carp (*Carassius auratus*) in natural waters. *BMC Genetics* **12,** 20 (2011).

35. Michalek, K., Ventura, A. & Sanders, T. Mytilus hybridisation and impact on aquaculture: a minireview. *Marine Genomics* **27,** 3–7 (2016).

36. Zhou, L. & Gui, J. Natural and artificial polyploids in aquaculture. *Aquaculture and Fisheries* **2,** 103–111 (2017).

37. Shao, G.-M. *et al.* Whole genome incorporation and epigenetic stability in a newly synthetic allopolyploid of gynogenetic gibel carp. *Genome Biology and Evolution* **10,** 2394–2407 (2018).

38. Ministry of Business, Innovation & Employment. *Research, science and innovation system performance report: ryegrass endophytes* Report (2018).

39. Brock, J. L., Caradus, J. R. & Hay, M. J. M. Fifty years of white clover research in New Zealand. *Proceedings of the New Zealand Grassland Association* **50,** 25–39 (1989).

40. Johnson, L. J. *et al.* The exploitation of Epichloae endophytes for agricultural benefit. *Fungal Diversity* **60,** 171–188 (2013).

41. Song, H. & Nan, Z. Origin, divergence, and phylogeny of asexual *Epichloë* endophyte in *Elymus* species from western China. *PLOS One* **10,** e0127096 (2015).

42. Saikkonen, K., Young, C. A., Helander, M. & Schardl, C. L. Endophytic *Epichloë* species and their grass hosts: from evolution to applications. *Plant Molecular Biology* **90,** 665–675 (2016).

43. Karlin, E. F. *et al.* Three-genome mosses: complex double allopolyploid origins for triploid gametophytes in *Sphagnum*. *Molecular Ecology* **18,** 1439–1454 (2009).

44. Karlin, E. F. & Smouse, P. E. Allo-allo-triploid *Sphagnum* × *falcatulum*: single individuals contain most of the Holantarctic diversity for ancestrally indicative markers. *Annals of Botany* **120,** 221–231 (2017).

45. Perrie, L. R., Brownsey, P. J., Lockhart, P. J. & Large, M. F. Evidence for an allopolyploid complex in New Zealand *Polystichum* (Dryopteridaceae). *New Zealand Journal of Botany* **41,** 189–215 (2003).

46. Shepherd, L. D., Perrie, L. R. & Brownsey, P. J. Low-copy nuclear DNA sequences reveal a predominance of allopolyploids in a New Zealand *Asplenium* fern complex. *Molecular Phylogenetics and Evolution* **49,** 240–248 (2008).

47. Ishikawa, N., Yokoyama, J. & Tsukaya, H. Molecular evidence of reticulate evolution in the subgenus *Plantago* (Plantaginaceae). *American Journal of Botany* **96,** 1627–1635 (2009).

48. Joly, S., Heenan, P. B. & Lockhart, P. J. A Pleistocene inter-tribal allopolyploidization event precedes the species radiation of *Pachycladon* (Brassicaceae) in New Zealand. *Molecular Phylogenetics and Evolution* **51,** 365–372 (2009).

49. Mallet, J. Hybrid speciation. *Nature* **446,** 279–283 (2007).

50. Darwin, C. *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life* (John Murray, London, 1859).

51. Comai, L. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6,** 836–846 (2005).

52. Tayalé, A. & Parisod, C. Natural pathways to polyploidy in plants and consequences for genome reorganization. *Cytogenetic and Genome Research* **140,** 79–96 (2013).

53. Spoelhof, J. P., Soltis, P. S. & Soltis, D. E. Pure polyploidy: closing the gaps in autopolyploid research. *Journal of Systematics and Evolution* **55,** 340–352 (2017).

54. Stebbins, G. L. *Variation and evolution in plants* (Oxford University Press, Oxford, 1950).

55. Stebbins, G. L. *Types of polyploids: their classification and significance* 403–429 (Academic Press, 1947).

56. McClintock, B. The significance of responses of the genome to challenge. *Science* **226,** 792 (1984).

57. Wu, Y. *et al.* Immediate genetic and epigenetic changes in F1 hybrids parented by species with divergent genomes in the rice genus (*Oryza*). *PLOS One* **10,** e0132911 (2015).

58. Gautam, M., Dang, Y., Ge, X., Shao, Y. & Li, Z. Genetic and epigenetic changes in oilseed rape (*Brassica napus* L.) extracted from intergeneric allopolyploid and additions with *Orychophragmus*. *Frontiers in Plant Science* **7,** 438 (2016).

59. Wang, J. *et al.* Rapid genomic DNA changes in allotetraploid fish hybrids. *Heredity* **114,** 601–609 (2015).

60. Qin, Q. *et al.* Rapid genomic changes in allopolyploids of *Carassius auratus red var.* (♀) × *Megalobrama amblycephala* (♂). *Scientific Reports* **6,** 34417 (2016).

61. Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440,** 341–345 (2006).

62. Gordon, J. L., Byrne, K. P. & Wolfe, K. H. Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLOS Genetics* **5,** e1000485 (2009).

63. Wang, M. *et al.* Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nature Plants* **4,** 90–97 (2018).

64. Parisod, C. *et al.* Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytologist* **184,** 1003–1015 (2009).

65. Qiu, T., Dong, Y. Z., Yu, X. M., Zhao, N. & Yang, Y. F. Analysis of allopolyploidy-induced rapid genetic and epigenetic changes and their relationship in wheat. *Genetics and Molecular Research* **16,** gmr16029303 (2017).

66. Marfil, C. F., Duarte, P. F. & Masuelli, R. W. Phenotypic and epigenetic variation induced in newly synthesized allopolyploids and autopolyploids of potato. *Scientia Horticulturae* **234,** 101–109 (2018).

67. Sloan, D. B. *et al.* Cytonuclear integration and co-evolution. *Nature Reviews Genetics* **19,** 635–648 (2018).

68. Jiao, W. *et al.* Asymmetrical changes of gene expression, small RNAs and chromatin in two resynthesized wheat allotetraploids. *Plant Journal* **93,** 828–842 (2018).

69. Ha, M. *et al.* Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 17835–17840 (2009).

70. Cox, M. P. *et al.* An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. *PLOS Genetics* **10,** e1004180 (2014).

71. Yoo, M. J., Liu, X., Pires, J. C., Soltis, P. S. & Soltis, D. E. Nonadditive gene expression in polyploids. *Annual Review of Genetics* **48,** 485–517 (2014).

72. Jung, Y., Kawaura, K., Kishii, M., Sakuma, S. & Ogihara, Y. Comparison of genome-wide gene expression patterns in the seedlings of nascent allohexaploid wheats produced by two combinations of hybrids. *Genes & Genetic Systems* **90,** 79–88 (2015).

73. Abbott, R. J. & Rieseberg, L. H. Hybrid speciation. *eLS* (2012).

74. McCarthy, E. M., Asmussen, M. A. & Anderson, W. W. A theoretical assessment of recombinational speciation. *Heredity* **74,** 502–509 (1995).

75. Buerkle, C., Morris, R., Asmussen, M. & Rieseberg, L. The likelihood of homoploid hybrid speciation. *Heredity* **84 (Pt 4),** 441–51 (2000).

76. Nieto Feliner, G. *et al.* Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity* **118,** 513–516 (2017).

77. Rieseberg, L. H. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *American Journal of Botany* **78,** 1218–1237 (1991).

78. Wolfe, A. D., Xiang, Q. Y. & Kephart, S. R. Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). *Proceedings of the National Academy of Sciences of the United States of America* **95,** 5112–5115 (1998).

79. Wang, X. R., Szmidt, A. E. & Savolainen, O. Genetic composition and diploid hybrid speciation of a high mountain pine, *Pinus densata*, native to the Tibetan plateau. *Genetics* **159,** 337–346 (2001).

80. James, J. K. & Abbott, R. J. Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. *Evolution* **59,** 2533–2547 (2005).

81. Pan, J., Zhang, D. & Sang, T. Molecular phylogenetic evidence for the origin of a diploid hybrid of *Paeonia* (Paeoniaceae). *American Journal of Botany* **94,** 400–408 (2007).

82. Brennan, A. C., Barker, D., Hiscock, S. J. & Abbott, R. J. Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: a study of *Senecio squalidus* (Asteraceae). *Heredity* **108,** 87–95 (2012).

83. Nevado, B., Harris, S. A., Beaumont, M. A. & Hiscock, S. J. Rapid homoploid hybrid speciation in British gardens: the origin of Oxford ragwort (*Senecio squalidus*). *Molecular Ecology* **29,** 4221–4233 (2020).

84. Mavárez, J. & Linares, M. Homoploid hybrid speciation in animals. *Molecular Ecology* **17,** 4181–4185 (2008).

85. Salazar, C. *et al.* Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *PLOS Genetics* **6,** e1000930 (2010).

86. Matos, I., Machado, M. P., Schartl, M. & Coelho, M. M. Gene expression dosage regulation in an allopolyploid fish. *PlOS One* **10,** e0116309 (2015).

87. Zhivotovsky, L. *et al.* Hybrids between chum *Oncorhynchus keta* and pink *Oncorhynchus gorbuscha* salmon: age, growth and morphology and effects on salmon production. *Journal of Fish Biology* **89** (2016).

88. Brasier, C. M. Rapid evolution of introduced plant pathogens via interspecific hybridization: hybridization is leading to rapid evolution of Dutch elm disease and other fungal plant pathogens. *BioScience* **51,** 123–133 (2001).

89. Stukenbrock, E. H., Christiansen, F. B., Hansen, T. T., Dutheil, J. Y. & Schierup, M. H. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 10954–10959 (2012).

90. Menardo, F. *et al.* Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. *Nature Genetics* **48,** 201–205 (2016).

91. Well, C. F. & Wessler, S. R. Molecular evidence that chromosome breakage by Ds elements is caused by aberrant transposition. *The Plant Cell* **5,** 515–522 (1993).

92. Madlung, A. *et al.* Genomic changes in synthetic *Arabidopsis* polyploids. *The Plant Journal* **41,** 221–230 (2005).

93. Lai, Z., Gross, B. L., Zou, Y. I., Andrews, J. & Rieseberg, L. H. Microarray analysis reveals differential gene expression in hybrid sunflower species. *Molecular Ecology* **15,** 1213–1227 (2006).

94. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biology* **17,** 13 (2016).

95. Duchemin, W., Dupont, P. Y., Campbell, M. A., Ganley, A. R. & Cox, M. P. HyLiTE: accurate and flexible analysis of gene expression in hybrid and allopolyploid species. *BMC Bioinformatics* **16** (2015).

96. Bird, K. A., VanBuren, R., Puzey, J. R. & Edger, P. P. The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist* **220,** 87–93 (2018).

97. Cheng, F. *et al.* Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLOS One* **7,** e36442–e36442 (2012).

98. Hollister, J. D. & Gaut, B. S. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research* **19,** 1419–1428 (2009).

99. Schwarzbach, A. E., Donovan, L. A. & Rieseberg, L. H. Transgressive character expression in a hybrid sunflower species. *American Journal of Botany* **88,** 270–277 (2001).

100. Hegarty, M. J. *et al.* Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in Senecio. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363,** 3055–3069 (2008).

101. Combes, M.-C., Dereeper, A., Severac, D., Bertrand, B. & Lashermes, P. Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New Phytologist* **200,** 251–260 (2013).

102. Yoo, M. J., Szadkowski, E. & Wendel, J. F. Homeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110,** 171–180 (2013).

103. Coate, J. E., Bar, H. & Doyle, J. J. Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). *The Plant Cell* **26,** 136 (2014).

104. Schedina, I. M., Hartmann, S., Groth, D., Schlupp, I. & Tiedemann, R. Comparative analysis of the gonadal transcriptomes of the all-female species *Poecilia formosa* and its maternal ancestor *Poecilia mexicana*. *BMC Research Notes* **7,** 249 (2014).

105. Zhang, J. *et al.* Transcriptome analysis of interspecific hybrid between *Brassica napus* and *B. rapa* reveals heterosis for oil rape improvement. *International Journal of Genomics* **2015,** 11 (2015).

106. Wang, X. *et al.* Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA-sequencing. *New Phytologist* **209,** 1264–1277 (2016).

107. Wu, Y. *et al.* Transcriptome shock in an interspecific F1 triploid hybrid of *Oryza* revealed by RNA sequencing. *Journal of Integrative Plant Biology* **58,** 150–164 (2016).

108. Edger, P. P. *et al.* Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *The Plant Cell* **29,** 2150 (2017).

109. Han, S. *et al.* Differential gene expression in leaf tissues between mutant and wild-type genotypes response to late leaf spot in peanut (*Arachis hypogaea* L.) *PLOS One* **12,** e0183428 (2017).

110. McElroy, K. E. *et al.* Genome expression balance in a triploid trihybrid vertebrate. *Genome Biology and Evolution* **9,** 968–980 (2017).

111. Zhang, C. *et al.* Comparative transcriptome analysis of flower heterosis in two soybean F1 hybrids by RNA-seq. *PLOS One* **12,** e0181061 (2017).

112. Qi, X. *et al.* Genomic and transcriptomic alterations following intergeneric hybridization and polyploidization in the *Chrysanthemum nankingense* × *Tanacetum vulgare* hybrid and allopolyploid (Asteraceae). *Horticulture Research* **5,** 5 (2018).

113. Zhang, M. *et al.* Transcriptome analysis reveals hybridization-induced genome shock in an interspecific F1 hybrid from *Camellia*. *Genome* **61,** 477–485 (2018).

114. Kryvokhyzha, D. *et al.* Towards the new normal: transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (*Capsella bursa-pastoris*). *PLOS Genetics* **15,** e1008131 (2019).

115. Matos, I., Machado, M. P., Schartl, M. & Coelho, M. M. Allele-specific expression variation at different ploidy levels in *Squalius alburnoides*. *Scientific Reports* **9,** 3688 (2019).

116. Hovhannisyan, H. *et al.* Integrative omics analysis reveals a limited transcriptional shock after yeast interspecies hybridization. *Frontiers in Genetics* **11,** 404 (2020).

117. Tronchoni, J., García-Ríos, E., Guillamón, J. M., Querol, A. & Pérez-Torrado, R. Transcriptomic analysis of *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* hybrids during low temperature winemaking. *F1000Research* **6,** 679 (2017).

118. Coate, J. E., Powell, A. F., Owens, T. G. & Doyle, J. J. Transgressive physiological and transcriptomic responses to light stress in allopolyploid *Glycine dolichocarpa* (Leguminosae). *Heredity* **110,** 160 (2012).

119. Hoencamp, C. *et al.* 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372,** 984–989 (2021).

120. Behling, A. H., Shepherd, L. D. & Cox, M. P. The importance and prevalence of allopolyploidy in Aotearoa New Zealand. *Journal of the Royal Society of New Zealand* **50,** 189–210 (2020).

121. Tahir, J. *et al.* QTL mapping for resistance to cankers induced by *Pseudomonas syringae* pv. *actinidiae* (Psa) in a tetraploid *Actinidia chinensis* kiwifruit population. *Pathogens* **9** (2020).

122. Cui, Y. *et al.* Genome-wide characterization and analysis of CIPK gene family in two cultivated allopolyploid cotton species: sequence variation, association with seed oil content, and the role of GhCIPK6. *International Journal of Molecular Sciences* **21,** 863 (2020).

123. Dong, Y. *et al.* Salt-tolerance diversity in diploid and polyploid cotton (*Gossypium*) species. *The Plant Journal* **101,** 1135–1151 (2020).

124. Rehman, A. *et al.* Genome wide identification, classification and functional characterization of heat shock transcription factors in cultivated and ancestral cottons (*Gossypium* spp.) *International Journal of Biological Macromolecules* **182,** 1507–1527 (2021).

125. Chen, Z. J. *et al.* Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nature Genetics* **52,** 525–533 (2020).

126. Gabaldón, T. Hybridization and the origin of new yeast lineages. *FEMS Yeast Research* **20,** foaa040 (2020).

127. Marques, I. *et al.* Transcriptomic leaf profiling reveals differential responses of the two most traded coffee species to elevated [CO_2]. *International Journal of Molecular Sciences* **21,** 9211 (2020).

128. Ebihara, A. & Nitta, J. H. An update and reassessment of fern and lycophyte diversity data in the Japanese Archipelago. *Journal of Plant Research* **132,** 723–738 (2019).

129. Cox, M. P., Peterson, D. A. & Biggs, P. J. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11,** 485–485 (2010).

130. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

131. R Development Core Team. *R: A Language and Environment for Statistical Computing* Computer Program. Version 4.0.2. 2020. `http://www.r-project.org`.

132. Robinson, J. T. *et al.* Integrative Genomics Viewer. *Nature Biotechnology* **29,** 24–26 (2011).

133. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14,** 178–192 (2012).

134. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15,** 550 (2014).

135. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37,** 1–13 (2009).

136. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).

137. Alexa, A. & Rahnenführer, J. *topGO: enrichment analysis for gene ontology* Computer Program. Version 2.36.0. 2019.

138. Alexa, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22,** 1600–1607 (2006).

139. Koskinen, P. & Holm, L. SANS: high-throughput retrieval of protein sequences allowing 50% mismatches. *Bioinformatics* **28,** i438–i443 (2012).

140. Koskinen, P., Törönen, P., Nokso-Koivisto, J. & Holm, L. PANNZER: high-throughput functional annotation of uncharacterized proteins in an error-prone environment. *Bioinformatics* **31,** 1544–1552 (2015).

141. Radivojac, P. *et al.* A large-scale evaluation of computational protein function prediction. *Nature Methods* **10,** 221–227 (2013).

142. Somervuo, P. & Holm, L. SANSparallel: interactive homology search against Uniprot. *Nucleic Acids Research* **43,** W24–W29 (2015).

143. Törönen, P., Medlar, A. & Holm, L. PANNZER2: a rapid functional annotation web server. *Nucleic Acids Research* **46,** W84–W88 (2018).

144. Fontana, P., Cestaro, A., Velasco, R., Formentin, E. & Toppo, S. Rapid annotation of anonymous sequences from genome projects using semantic similarities and a weighting scheme in gene ontology. *PLOS One* **4,** e4619 (2009).

145. Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G. & Thomas, R. Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology* **56,** 45–50 (2008).

146. Kinsella, R. J. *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database: The Journal of Biological Databases and Curation* **2011,** bar030 (2011).

147. Hunt, S. E. *et al.* Ensembl variation resources. *Database: The Journal of Biological Databases and Curation* **2018,** bay119 (2018).

148. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44,** D733–D745 (2016).

149. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28,** 27–30 (2000).

150. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* **44,** D457–D462 (2016).

151. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45,** D353–D361 (2017).

152. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* **28,** 33–36 (2000).

153. Pereira, C. S. A., Aboim, M. A., Ráb, P. & Collares-Pereira, M. J. Introgressive hybridization as a promoter of genome reshuffling in natural homoploid fish hybrids (Cyprinidae, Leuciscinae). *Heredity* **112,** 343–350 (2014).

154. Campbell, M. A. *et al. Epichloë hybrida*, sp. nov., an emerging model system for investigating fungal allopolyploidy. *Mycologia* **109,** 715–729 (2017).

155. Bertioli, D. J. *et al.* The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nature Genetics* **51,** 877–884 (2019).

156. Holá, D. *et al.* The disadvantages of being a hybrid during drought: a combined analysis of plant morphology, physiology and leaf proteome in maize. *PLOS One* **12,** e0176121 (2017).

157. Ueno, N., Kashiwagi, M., Kanekatsu, M., Marubashi, W. & Yamada, T. Accumulation of protein aggregates induces autolytic programmed cell death in hybrid tobacco cells expressing hybrid lethality. *Scientific Reports* **9,** 10223 (2019).

158. Zhang, L. *et al.* Integration of metabolome and transcriptome reveals flavonoid accumulation in the intergeneric hybrid between *Brassica rapa* and *Raphanus sativus*. *Scientific Reports* **9,** 18368 (2019).

159. Kim, S.-T., Sultan, S. E. & Donoghue, M. J. Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 12370 (2008).

160. Refstie, T. & Gjedrem, T. Hybrids between Salmonidae species. Hatchability and growth rate in the freshwater period. *Aquaculture* **6,** 333–342 (1975).

161. Bagheri, B., Bauer, F. F. & Setati, M. E. The impact of *Saccharomyces cerevisiae* on a wine yeast consortium in natural and inoculated fermentations. *Frontiers in Microbiology* **8,** 1988 (2017).

162. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31,** 3210–3212 (2015).

163. Hinkle, A. E. The distribution of a male sterile form of *ti* (*Cordyline fruticosa*) in Polynesia: a case of human selection. *The Journal of the Polynesian Society* **113,** 263–290 (2004).

164. Roullier, C. *et al.* Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.) *PLOS One* **8,** e62707 (2013).

165. Nemorin, A., Abraham, K., David, J. & Arnau, G. Inheritance pattern of tetraploid *Dioscorea alata* and evidence of double reduction using microsatellite marker segregation analysis. *Molecular Breeding* **30,** 1657–1667 (2012).

166. Griffiths, A. G. *et al.* Breaking free: the genomics of allopolyploidy-facilitated niche expansion in white clover. *The Plant Cell* **31,** 1466–1487 (2019).

167. Atkinson, R. G., Cipriani, G., Whittaker, D. J. & Gardner, R. C. The allopolyploid origin of kiwifruit, *Actinidia deliciosa* (Actinidiaceae). *Plant Systematics and Evolution* **205,** 111–124 (1997).

168. Chalhoub, B. *et al.* Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345,** 950 (2014).

169. Lashermes, P., Combes, M.-C., Hueber, Y., Severac, D. & Dereeper, A. Genome rearrangements derived from homoeologous recombination following allopolyploidy speciation in coffee. *The Plant Journal* **78,** 674–685 (2014).

170. Wendel, J. F., Schnabel, A. & Seelanan, T. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences of the United States of America* **92,** 280–284 (1995).

171. Tatum, T. C., Stepanovic, S., Biradar, D. P., Rayburn, A. L. & Korban, S. S. Variation in nuclear DNA content in *Malus* species and cultivated apples. *Genome* **48,** 924–930 (2005).

172. Evans, K. M., Govan, C. L. & Fernández-Fernández, F. A new gene for resistance to *Dysaphis pyri* in pear and identification of flanking microsatellite markers. *Genome* **51,** 1026–1031 (2008).

173. Ozkan, H., Arumuganathan, K. & Tuna, M. Nonadditive changes in genome size during allopolyploidization in the wheat (*Aegilops-Triticum*) group. *Journal of Heredity* **94,** 260–264 (2003).

174. Gaut, B. S., d'Ennequin, M. L. T., Peek, A. S. & Sawkins, M. C. Maize as a model for the evolution of plant nuclear genomes. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 7008–7015 (2000).

175. Albertin, W. & Marullo, P. Polyploidy in fungi: evolution after whole-genome duplication. *Proceedings of the Royal Society B: Biological Sciences* **279,** 2497–2509 (2012).

176. Web Page. 2019. `https://statisticsnz.shinyapps.io/trade_dashboard/`.

177. Buckley, T. R. *et al.* Investigating hybridization in the parthenogenetic New Zealand stick insect *Acanthoxyla* (Phasmatodea) using single-copy nuclear loci. *Molecular Phylogenetics and Evolution* **48,** 335–349 (2008).

178. Myers, S. S., Trewick, S. A. & Morgan-Richards, M. Multiple lines of evidence suggest mosaic polyploidy in the hybrid parthenogenetic stick insect lineage *Acanthoxyla*. *Insect Conservation and Diversity* **6,** 537–548 (2013).