Non-additive genetic effects induce novel phenotypic distributions in male mating traits of F1 hybrids

Abstract

Hybridization is a source of phenotypic novelty and variation because of increased additive genetic variation. Yet, the roles of non-additive allelic interactions in shaping phenotypic mean and variance of hybrids have been underappreciated. Here we examine the distributions of male-mating traits in F1 hybrids via a meta-analysis of 3,208 effect sizes from 39 animal species pairs. Although additivity sets phenotypic distributions of F1s to be intermediate, F1s also showed recessivity and resemblance to maternal species. F1s expressed novel phenotypes (beyond the range of both parents) in 65% of species pairs, often associated with increased phenotypic variability. Overall, however, F1s expressed smaller variation than parents in 51% of traits. While genetic divergence between parents did not impact phenotypic novelty, it increased phenotypic variability of F1s. By creating novel phenotypes with increased variability, non-additivity of heterozygotic genome may play key roles in determining mating success of F1s, and their subsequent extinction or speciation.

Key words

Meta-analysis, Transgressive segregation, Developmental stability,
Heterosis, Outbreeding depression, Phenotypic uniformity

Introduction

The long history of animal and plant breeding has taught us that interspecific hybridization is a powerful source of phenotypic novelty. While hybrid populations usually distribute their phenotypes throughout the range of parental species, they often exhibit novel phenotypes, as well as novel variability (i.e. more extreme phenotypic mean or variability than observed in either parent species) (Edmands 1999; Janick 1999; Stelkens and Seehausen 2009). Novel phenotypes enable hybrids to exploit novel niches and ultimately become new species (Rieseberg et al. 1999; Kagawa and Takimoto 2018). Naturally, the distribution of hybrids' phenotypic traits has attracted much attention of evolutionary biologists. Despite long interests of breeding and evolutionary biologists, however, hybrid phenotype has never been subjected to the formal meta-analysis that accounts for different types of statistical non-independence and mean-variance relationship (Rieseberg et al. 1999; Stelkens and Seehausen 2009; Nakagawa et al. 2015).

Novel phenotypes are produced by various mechanisms, such as mutation, chromosomal recombination, dominance, epistasis, and cross direction effects including epigenetics (DeVicente and Tanksley 1993; Cockett et al. 1996; Lamkey and Edwards 1999; Rieseberg et al. 1999). Chromosomal recombination and rearrangements increase additive (heritable) genetic variation in hybrid populations at second and later generations. In F1 hybrids characterized by heterozygous genome, additive effects are expected to result in the averaged trait size of both parental species. However, F1 hybrids often resemble more one of the two parents (dominance or recessivity), and have different phenotypes between reciprocal hybrids (cross direction effects) (Thompson et al. 2021). Moreover, F1 hybrids sometimes exhibit novel phenotypes (Lamkey and Edwards 1999; Rieseberg et al. 1999; Stelkens and Seehausen 2009). Although novel variability — especially smaller phenotypic variation in F1 hybrids — has been utilized in agriculture to enhance yield stability (Riggs 1988; Janick 1999), taxon-wide prevalence of novel variability has not been examined (Vetukhiv 1953; Wallace 1955; Edmands 1999). Fragmented evidence above suggests that non-additive genetic effects are important sources of novel phenotypic distributions of hybrid populations (Alibert and Auffray 2003; Chen 2013; Wei and Zhang 2018). Because phenotypic distribution determines survival and reproductive success of hybrids, it is particularly relevant to the gene flow and backcrossing of hybrid populations, and, ultimately, speciation.

Genetic divergence between parental species has been hypothesized as a factor shaping phenotypic distribution of F1 hybrids. Heterozygotic loci in F1 hybrids can increase phenotypic novelty by

inducing novel allelic interactions (e.g., dominance and epistasis) (Lamkey and Edwards 1999; Rieseberg et al. 1999). Genetic divergence between parents, which enhances heterozygosity of F1 hybrids, is thus expected to increase phenotypic novelty of F1 hybrids (Stelkens and Seehausen 2009). Further, accrued genetic difference between parents can harm hybrid fitness (i.e. outbreeding depression) due to increasing deleterious allelic interactions (i.e. genetic incompatibility) (Moyle and Nakazato 2010) or collapse of beneficial gene interactions that evolved within the parental species (Dagilis et al. 2019). Outbreeding depression of F1 hybrids could result in smaller trait size than both parents (Arnold and Hodges 1995; Wei and Zhang 2018), and increase phenotypic variability by inducing developmental instability (Alibert and Auffray 2003). Also, F1 hybrids between genetically close parents may perform better than both parents by being released from deleterious allelic interactions within species (i.e. heterosis or hybrid vigor) (Dagilis et al. 2019). Under heterosis, F1 hybrids can exhibit greater trait size and / or reduced trait variability than parents because of enhanced developmental stability (Arnold and Hodges 1995; Edmands 1999).

Here we investigate phenotypic mean and variability of F1 hybrids of male mating trait size (i.e. traits used during mating that females do not possess) across animal taxa (Fig. 2a). By focusing on male traits, we exclude the possible biases resulting from any sex differences in phenotypic distributions. Given a wide prevalence of sexual dimorphism (Poissant et al. 2010), sexes should be distinguished when comparing phenotypic distributions between F1 hybrids and parents. Phenotypic distribution can be sexually-biased in F1 hybrids, but not in parents, due to sex-biased mortality in F1 hybrids (Haldane 1922; Schilthuizen et al. 2011). Male mating traits also directly relate to the reproductive success and have been reported relatively well in the literature. In contrast, little information is available for female-specific sexual traits and mate preference of female hybrids.

By employing formal phylogenetically controlled meta-analytic techniques, we test if phenotypic mean and variability of F1 hybrids are larger or smaller than mid-species value (Fig. 2b), and are affected by cross direction (father from species A with mother from species B, or vice versa; Fig. 2c). Further, we provide the first quantification of how often F1 phenotype is more / less variable than parents. By using a phylogenetic comparative method, we investigate potential factors including genetic divergence between parents that influence novelty and variability of male mating trait size of F1 hybrids (Fig. 4 and 5: hypotheses for each factor are summarized in Supplementary Table S1).

Material and methods

Data collection

Literature search. We conducted systematic review of literature and followed Preferred Reporting Items for Meta-Analyses (PRISMA, Fig. 1) when reporting our meta-analytic workflow. We searched journal articles that contain 'hybrid*' AND 'male*' AND ('grow*' OR 'size.' OR 'length.' OR 'mass' OR 'weight.' OR 'behav*' OR 'trait.' OR 'phenotyp*') in the title, abstract or keywords, by using Web of Science and Scopus on October 20, 2017. We also searched journal articles through backward / forward citations of reviews of hybrid fitness and phenotype (Haldane 1922; Arnold and Hodges 1995; Burke and Arnold 2001; Reinhold 2002; Turelli and Moyle 2007; Stelkens and Seehausen 2009; Schilthuizen et al. 2011; Reinhold and Engqvist 2013).

Inclusion criteria. To be included in our analysis, the primary study had to report means and variance of male sexual phenotypes in hybrids and both non-hybrid crosses (i.e. parental species; note that we use the offspring from non-hybrid crosses as the proxy of the phenotypes of the parental species, the "parents"); separately report the phenotype of reciprocal hybrid crosses; use different species or subspecies for the crosses; experimentally obtain all crosses and raise them in same environment; provide all necessary statistics (means, standard deviations/errors and sample sizes) for effect size calculations. Relevant studies deemed to meet the above criteria, based on titles and abstracts, were screened as full-texts by two of the authors (KA and ML).

Data extraction. From primary studies, we extracted the phenotypic measures and the number of individuals in each group of animals. We extracted data from figures using R package *metaDigitse* 1.0 (Pick et al. 2019). We estimated standard deviation from range data, if necessary (Wan et al. 2014). The observations that had any negative trait sizes were removed from the dataset, but this procedure did not reduce included studies and species.

Dominance and cross direction effects

Effect sizes. For phenotypic means, we calculated the log ratio (ln*RR* (Hedges et al. 1999)) from the non-hybrid parent with the smaller trait size (small non-hybrid) to each reciprocal hybrid cross (arrows 2–3 in Fig. 2b), mid-species (arrow 4 in Fig. 2b) and the other parent (non-hybrids with larger trait size — large non-hybrid: arrow 1 in Fig. 2b). We evaluated phenotypic variability in the analogous way with phenotypic mean, but we used the log ratio between coefficient of variation

(ln*CVR*, Nakagawa et al. 2015) and aligned parents by phenotypic variability (small and large non-hybrids for parents with small and large coefficient of variation of trait size, respectively).

Meta-analyses. To estimate the overall mean of each effect size, we first estimated meta-analytic means of each effect size (effect sizes 1–3 in Fig. 2b). We then asked whether phenotypic mean or variability is greater or smaller than mid-species (comparing effect sizes 2–3 with 4 in Fig. 2b), by using meta-analytic model that included mid-species value as an intercept. To investigate cross direction effects, we tested whether the reciprocal hybrid crosses differ in their mean phenotype (compared effect size 2 with 3 in Fig. 2b).

All meta-analytic models were phylogenetically controlled and included following random effects: study ID, trait observation ID and strain ID that discriminated intraspecific genetic strains or populations. Phylogenetic effects were included as a random effect as a covariance matrix for large non-hybrid parental species. We did not include species as random effect because this effect should be accommodated by the strain random effect. A phylogenetic tree was created based on Open Tree of Life database (Hinchliff et al. 2015) (topology of the tree: Fig. S1–2). We identified species in the Open Tree Taxonomy using the R package *rotl* 3.0.10 (Michonneau et al. 2016) and computed branch lengths using the default settings of the *compute.brlen* function in the R package *ape* 5.3 (Paradis and Schliep 2019). All meta-analytic models were implemented using the R package *metafor* 2.1-0 (Viechtbauer 2010). All analyses were conducted in R 3.6.3 (R Core Team 2020).

Novelty in phenotypic means and variabilities

Effect sizes. For phenotypic mean, we calculated ln*RR* of large non-hybrid to hybrid crosses, and ln*RR* of hybrid crosses to small non-hybrid at each trait observation (Fig. 2c). As the sign of these effect sizes should be positive when hybrid phenotypic mean lies within the ranges of parents, negative effect sizes indicate novel phenotypic means. These effect sizes also indicate the relative trait size compared with parents – whether mean trait size is greater or smaller than both parents and whether hybrids exceed maternal or paternal species' trait size (Fig. 2c). Again, we evaluated phenotypic variability in the analogous way with phenotypic mean by using lnCVR and aligning parents by phenotypic variability size.

Predictor variables. We additionally collected genetic distance between parental species and heterogametic sex because we *a priori* expected them to influence phenotypic distribution of hybrids.

We took the mitochondria *cytochrome c oxidase I* (*COI*) DNA sequences (up to 20 sequences per parental species depending on their availability) from GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) and aligned them using ClustalW (Thompson et al. 1994). Alignments with 450 bp as the minimum length were manually optimized for calculating the genetic distance. We calculated maximum composite likelihood distance with uniform substitution rate and homogenous pattern (Tamura et al. 2004) in the software MEGA 7.0.26 (Kumar et al. 2016). If multiple sequences were available for a pair of species, we calculated the average of all possible pairs of sequences. We assigned heterogametic sex at family or genus level, if deemed evolutionary constant within the taxon (Ashman et al. 2014).

Bayesian models. To find the factors affecting novelty in phenotypic means and variances, we conducted logistic regression with the novelty status of phenotypic mean or variance (whether phenotypic mean or variance of hybrids is more extreme than of either parents) as a binary response variable (0 = not novel; 1 = novel). We ran the analogous logistic models for phenotypic means and variances. As predictors, the models included: (*i*) natural logarithm of genetic divergence between parents; (*ii*) phenotypic divergence between parents in the focal trait – absolute value of ln*RR*; (*iii*) range overlaps between parental species – allopatry or not; (*iv*) viability of reciprocal hybrids – completely inviable or not; (*v*) trait type – morphology or sound traits; and the relative trait size or variability compared to parents – (*vi*) greater or smaller and (*vii*) exceeding maternal or paternal species. The model also included interaction terms of species or trait level factors (*i–v*) with parental species compared (*vi* and *vii*). Range overlaps and viability of hybrids were assessed based on the primary studies.

The main effects indicate how species- and trait-level characteristics influence the probability of F1 hybrids to exhibit novel phenotypic means or variability in any directions (*i–v*), and whether phenotypic means or variability is biased toward certain direction (*vi* and *vii*). The interaction terms indicate how species and trait level characteristics affected the relative trait size or variability compared to parents (interactions with *vi*, greater trait size than both parents; those with *vii*, trait size exceed maternal species value). Two (all) fish and three orthoptera species-pairs were not used in regression because of the lack of data for genetic distance or sex-determination system. To visualize significant factors' influences on the relative trait size or variability (Fig. 4c–e and 5c–e), we ran simpler models for each significant factor. As predictors, we included the focal factors (either one of *i–v*), the compared parental species – large or small (*vi*), and the interaction between them. As

random effects, these models included trait study ID, strain ID (discriminated genetic strains or populations within a species), and standard error of the response variable. Phylogenetic effects were included as a random effect as a covariance matrix for large parental species (Fig. S1–2). Standard error of the binary response variable was estimated as $\frac{\pi}{\sqrt{3n}}$, where *n* is the summation of sample sizes of the hybrid and non-hybrid crosses compared.

Lastly, we conducted an ordinal regression mixed-effects model to examine the link between phenotypic novelty and variability. The model included relative trait variability of hybrids compared to parents (ordered: greater than both parents > within the range of parents > smaller than both parents) as a response, and novelty in phenotypic means (novel or non-novel) as a predictor variable. This model also included trait study ID, strain ID, trait observation ID, cross direction and phylogeny (tree topology, Fig. S3) as random effects.

All Bayesian analyses were conducted in R package *MCMCglmm v.*2.29 (Hadfield 2010), using phylogenetic comparative hierarchical models described in Hadfield and Nakagawa (2010). We used a weakly informative Gelman prior for the fixed effects, and inverse-Wishart priors for the variances of the random effects. The residual variance was fixed to 1, as this cannot be estimated with binary or trinary data. Parameter estimates were subsequently scaled under the assumption that the true residual variance is 0. We ran the analysis for 60,000 iterations with a burn-in of 5,000 and a thinning interval of 10.

Results

Dataset

Although trait inheritance pattern is a key question of evolutionary biology, a modest number of publications reported results of crossing species reciprocally while separating offspring trait data by sexes. We found 25 such published studies (Supplementary material, section *Data description*). We extracted two sets of 1,604 effect sizes comparing phenotypic mean and variability, respectively, between parents and hybrids, based on 401 male traits observations from 39 species pairs. More than half species crossings involved *Drosophila* (Diptera, 59%, 23/39 species pairs, Fig. 2a) and only 13% of species pairs (5/39) were vertebrates, including bony fish, frog, bird and rodent species. Most

219 species pairs were male heterogametic (92%, 34/37 species pairs, excluding two fishes of which
220 sex-determination system is unknown). Both reciprocal hybrids were viable in 59% of species pairs
221 (23/39); and geographic ranges of parental lineages overlap in 36% of species pairs (14/39).
222
223 ***Dominance in phenotypic means***
224 F1 hybrids of both cross directions showed partial recessivity. Hybrids whose maternal species had
225 larger trait size (large non-hybrid mother × small non-hybrid father) and the reciprocal hybrids (large
226 father × small mother) on average showed 12.5% and 16.9% smaller phenotypic mean than
227 mid-species value, respectively (Fig. 3a; $P < 0.001$ each, Table S3). This trend was qualitatively the
228 same when we excluded traits showing novel phenotypic means (i.e. values outside the range of both
229 parental species; Fig. 3c). Nonetheless, the magnitude of phenotypic difference between hybrids and
230 parental species is highly heterogenous, much of which are either due to phylogeny or unexplained
231 (total $I^2 = 99.7\%$, which are partitioned into phylogeny $I^2 = 38.8\%$, study $I^2 = 1.1\%$, crossed lineage
232 $I^2 = \sim 0\%$, residual $I^2 = 59.8\%$; Fig 3a–b). Such a high heterogeneity suggests that any inheritance
233 patterns are plausible in F1 hybrids (e.g., dominance, cross direction effect and exhibiting novel
234 phenotypic means). –
235
236 ***Crossing direction effects in phenotypic means***
237 Hybrids were more similar to the males of maternal species than of paternal species: F1 hybrids from
238 large mother species exhibited 4.9% larger trait size than the reciprocal hybrids (large father × small
239 mother) (Fig. 3a; $P < 0.001$, Table S4). Same tendency was detected when we excluded traits
240 showing novel phenotype (hybrids having larger maternal species had 8.7% larger phenotypic
241 means; $P < 0.001$, Table S4).
242
243 ***Dominance and cross direction effects in phenotypic variabilities***
244 F1 hybrids exhibited similar phenotypic variability (coefficient of variation, CV) to the mid-species
245 value (Fig. 3b, Table S5). Phenotypic variability did not significantly differ between F1 hybrids
246 whose maternal species had larger trait variability and the reciprocal hybrids (Table S6). Yet, these
247 do not necessarily mean that phenotypic variability inherits additively. In most trait observations,
248 hybrid phenotypic variability is either larger or smaller than that of both parents (74.9%, see *Novelty*
249 *in phenotypic variabilities*). We also note that detection power of the differences in CVs is lower than
250 in comparisons of means or variances alone, because errors of CVs contain sampling errors of both
251 means and variances (Nakagawa et al. 2015; Senior et al. 2020). The magnitude of phenotypic

variability was highly heterogenous, and was partially dependent upon phylogeny and study identities (total I^2 = 75.6%, partitioned into: phylogeny I^2 = 19.5%, study I^2 = 11.4%, crossed lineage I^2 = ~0%, residual I^2 = 44.7%; Fig. 3c, d).

Novelty in phenotypic means

F1 hybrids exhibit novel phenotypic means in 64.7% (22/34) of species pairs, and 42.9% (143/333) of trait observations (Fig. 4a). This indicates that non-additive genetic interaction is a powerful source of phenotypic novelty in male mating traits.

Genetic divergence between parents did not significantly associate with novelty in phenotypic means (main effect: β = 0.26, 95% credible interval [CI] = -0.41 – 0.9, P = 0.420: Fig. 4b). Rather, genetic divergence significantly reduced the probability of hybrids to have larger trait size than both parents (interaction term of genetic divergence with the relative trait size: β = -0.59, CI = -1.19 – 0.02, P = 0.041: Fig. 4b–c). Increasing phenotypic divergence between parental species generally reduced the novelty (main effect of phenotypic divergence: β = -0.8, CI = -1.48 – -0.23, P < 0.001). Compared to F1 hybrids with viable reciprocal hybrids, those without viable reciprocal hybrids were more likely to exhibit smaller trait mean than both parents (interaction term of reciprocal hybrids' viability with the relative trait size compared to parents: β = 2.51, CI = 1.01 – 4.01, P = 0.001: Fig. 4b and e) and, albeit non-significant, tended to show novel phenotypic means (main effect of reciprocal hybrids' viability: β = -1.36, CI = -2.97 – 0.04, P = 0.064: Fig. 4b). Novel phenotypic means were not biased toward any directions, as F1 hybrids exhibited either larger or smaller trait size than parents (main effect of relative trait size, P > 0.05, Table S7) and significantly exceeded phenotypic means of either maternal or paternal species (main effect of exceeding maternal or paternal species, P > 0.05, Table S7). This indicates that recessivity and crossing direction effect had little impact on novelty in phenotypic means.

Novelty in phenotypic variabilities

Phenotypic variability of F1 hybrids rarely lied within the range of parental species (6.1%, 31/33 of species pairs; 25.1%, 248/331 of trait observations, Fig. 5a). Compared to the parents, hybrids exhibited smaller phenotypic variability in 51.4% of trait observations and had greater variability less frequently (24.8%, Fig. 5a).

Novel phenotypes were more variable than non-novel phenotypes of which mean lie within the range

of parents (ordinal phylogenetic random regression: $\beta = 1.01$, CI = 0.45 – 1.62, $P < 0.001$). That is, novel phenotypes, regardless of relative trait size compared to parents, were more likely to exhibit greater variability than parents (30.0 % of trait observations) in comparison to non-novel phenotypes (18.5 %).

F1 hybrids between genetically distant parents were likely to be phenotypically more diverse than parents, whereas those between genetically close parents were typically more homogenous than parents (interaction term of genetic divergence with the relative trait variability compared to parents: $\beta = 1.25$, CI = 0.47 – 2.00, $P = 0.002$: Fig. 5b–c and Table S8). F1 hybrids were more phenotypically variable than parents in traits with large parental divergence (interaction term of phenotypic divergence with the relative trait variability: $\beta = 0.83$, CI = 0.01 – 1.61, $P = 0.036$: Fig. 5b). Note that novelty in variabilities did not associate with phenotypic divergence when species-level moderators were not included in the model (Fig. 5d). F1 hybrids tended to vary less than parents in sound traits, and to vary more than parents in morphology (interaction term of sound traits vs. morphological traits with the relative trait variability: $\beta = -3.29$, CI = -5.21 – -1.38, $P = 0.001$: Fig. 5b and e).

Discussion

By synthesizing 401 male traits observation data from 39 species pairs across the animal kingdom, we illuminated important roles of non-additivity in phenotypic distribution of F1 hybrids. We detected recessivity and cross direction effect on phenotypic means of the hybrids (Fig. 3a and c). Novel phenotypic means were found in 65% of species pairs (Fig. 4a), highlighting the importance of non-additive allelic interactions in creating novel phenotypes. Novelty in phenotypic means was not associated with parental genetic divergence but enhanced by genetic incompatibility (when reciprocal hybrids are inviable, Fig. 4b). Similar to observations made in plant agricultural breeding (Riggs 1988; Janick 1999), F1 hybrids exhibited smaller phenotypic variability than both parents in majority of trait observations (Fig. 5a). We showed that genetic differentiation between parents increases phenotypic variability of F1 hybrids (Fig. 5b).

Recent synthesis showed that resemblance to maternal species was the norm in F1 hybrids across animals and plants (Thompson et al. 2021). Our analysis also revealed resemblance to maternal species of F1 hybrids, which is counterintuitive because we focused on male traits. Given that

maternal effects are particularly pervasive for morphology, including body size (Moore et al. 2019), mothers can directly influence male-mating morphological trait size, and even sound traits that are partially determined by morphology and body size (Ryan and Brenowitz 1985; Bennet-Clark 1998). Alternatively, maternally inherited mitochondrial genome may influence F1 male phenotypes.

Phenotypic novelty

We suggest that hybrid populations may express novel phenotypes much more frequently than previously thought. A recent comparative analysis, using non-male-mating traits, showed that F1 hybrids express novel phenotype only in 20% of species pairs across any plants and animal taxa (Thompson et al. 2021), which is one-third of the estimate from our study (65%). Discordance between these two results suggests that male-mating traits of animals more frequently express novel phenotype during hybridization, and/or using sex-aggregated data in the earlier study has led to underestimation of novel phenotype expression frequency. Alternatively, a large portion of novel phenotype observations in our dataset may arise from heterosis or outbreeding depression, though Thompson et al. (2021) attempted to exclude heterosis and outbreeding depression by choosing traits seemingly unrelated to fitness. Our dataset included any mating traits that are likely to reflect body growth (e.g., genitalia size) or behavioral deficit (e.g., compromised vocal traits). Hence, F1 hybrids may exhibit larger or smaller male trait sizes than both parents by growing faster or slower than parents (heterosis and outbreeding depression, respectively). Outbreeding depression, at least, is possibly reflected in our result — trait size declined in F1 hybrids without viable reciprocal hybrids and from genetically distant parents (Fig. 4b–c and e).

While non-additive interactions can enhance phenotypic novelty in any generation of hybrids, studies on quantitative trait loci (QTL) suggests that recombination after F1 hybrids is the major source of phenotypic novelty (DeVicente and Tanksley 1993; Rieseberg et al. 2003; Koide et al. 2019). Novel phenotype is thus expected to be more frequent in later generation hybrids than in F1 hybrids. Yet, other comparative studies, included F1 and later generation hybrids across animals, found that 29–31% of traits showed novel phenotypes (Rieseberg et al. 1999; Stelkens and Seehausen 2009), which is about half of our estimate. We exclusively focused on experimentally-derived F1 hybrids while previous studies included natural hybrid populations (Rieseberg et al. 1999; Stelkens and Seehausen 2009). In natural hybrid populations included in the previous syntheses, extrinsic natural selection may have removed novel phenotypes and led to underestimation of phenotypic novelty (Rieseberg et al. 1999; Stelkens and Seehausen 2009).

We anticipated that genetic divergence between parental species positively relates to phenotypic novelty. This is because parent divergence should increase the number of heterozygotic loci in F1 hybrids, and heterozygosity is expected to enhance phenotypic novelty (Lamkey and Edwards 1999). However, in our results, genetic divergence did not associate with the probability of F1 hybrids to express novel phenotypic means (Fig. 4b). This is in line with the previous finding across animals and plants that phenotypic means of F1 hybrids do not increase their degree of novelty along with genetic divergence between parents (Thompson et al. 2021). Our prediction was not supported, presumably because the effects of inter-allelic interactions can be diverse. We assumed that inter-allelic interactions among QTLs act in the same direction (e.g., all interactions increase trait size) (Lamkey and Edwards 1999). If the sign varies among the interactions, however, genetic differentiation no longer necessarily increases phenotypic novelty because interactions will cancel out effects of each other. Alternatively, F1 hybrid fitness can be maximized when genetic distance between the parents is small or moderate because of strong heterosis combined with weak genetic incompatibility (Moll et al. 1965; Wei and Zhang 2018; Dagilis et al. 2019). When heterosis increases trait size through improving fitness, novel phenotype expression can become frequent in hybrids between non-genetically distant parents. Genetic divergence between parents thus could be a poor predictor of phenotypic novelty of F1 hybrids.

Based on the similar reasoning with genetic divergence, we expected that F1 hybrids are likely to exhibit novel phenotypic means in traits with greater parental phenotypic divergence. However, our analysis detected the opposite pattern (Fig. 4b and d). When parental species are phenotypically similar, their QTL could be heterozygous. Here, some of F1 hybrids will be homozygous at QTL. Additive effects of QTL enable homozygous hybrids to express novel phenotype (Stelkens and Seehausen 2009). Hence, additivity and heterozygosity of parents may explain the novel phenotypic means of F1 hybrids from phenotypically close parents. Furthermore, we found that novel phenotypes (mean outside the range of parents) were more variable than non-novel phenotypes (mean within the range of parents). If phenotypic variability reflects developmental instability (Edmands 1999; Alibert and Auffray 2003), observed pattern suggests that novel phenotypes arise from developmental instability of hybrids. That is, hybrids with stable development may exhibit less extreme phenotype (within the range of parents), whereas hybrids suffering developmental instability may tend to exhibit novel phenotype. Moreover, F1 hybrids tend to exhibit novel phenotype when genetic incompatibility is strong (i.e. the reciprocal hybrids is inviable, Fig. 4b). We therefore

suggest that developmental instability arising from genetic incompatibility is an important source of phenotypic novelty in F1 hybrids.

Phenotypic variability

We revealed F1 hybrids tend to be less phenotypically variable than parents (Fig. 5a). Yet, this does not necessarily indicate reduced evolvability of hybrid populations — later-generation hybrids can expand phenotypic variation through recombination even if F1 hybrids exhibit smaller phenotypic variability than parents (Edmands 1999; Rieseberg et al. 1999). Rather, our results emphasize the time lag between hybridizing event and expression of novel phenotypic variation in hybrid population (Grant and Grant 2019).

As expected, genetic divergence between parental species enhanced phenotypic variability of F1 hybrids (Fig. 5b). Increasing phenotypic variability with parental genetic divergence was also reported in later generation hybrids (Stelkens and Seehausen 2009; Stelkens et al. 2009). The common pattern in F1 and later generation hybrids implies that non-additive interactions contribute to phenotypic distribution of hybrid population. Genetically close parents yielded F1 hybrids with smaller phenotypic variability, possibly due to heterosis enhancing developmental stability (Edmands 1999), rather than due to selective mortality caused by outbreeding depression (i.e. individuals with anomalous phenotypes, resulting from epistatic interactions, are likely to die). This is because heterosis is often observed in hybrids between genetically similar parental species (Dagilis et al. 2019), but outbreeding depression become more likely as parents genetically diverge (Moyle and Nakazato 2010; Dagilis et al. 2019).

We found that F1 hybrids tended to vary less than parents in sound traits, and to vary more than parents in morphology (Fig. 5b). Larger variability in mating-related traits may facilitate backcrossing due to greater overlap in phenotypic range with parental species. Hence, taxa relying predominantly on morphology-based mating traits (e.g., genitalia and coloration) might be more prone to gene flow resulting from backcrossing.

Knowledge gap and future directions

In addition to each individual trait size, trait correlation (covariance) can influence the attractiveness of F1 males, because multiple mating traits often interactively determine sexual attractiveness (Rosenthal 2013). Our additional analysis revealed traits can vary in the strength and directions of

dominance within hybrids (i.e. several traits resemble one parent, but other traits resemble the other parent: Supplementary material, section *Trait mosaicism*), which was previously shown in non-male mating traits (Thompson et al. 2021). This indicates that trait correlation within parental species could easily break in F1 hybrids, as reported in several studies (Rieseberg and Ellstrand 1993; Matsubayashi et al. 2010; Selz et al. 2014; Thompson et al. 2021). Moreover, a variety of trait correlation patterns can arise in F1 individuals because phenotypic variability varies among traits (Fig. 3b and 5b). Despite the importance of determining fitness and mating pattern, correlation among mating traits of F1 hybrids has received little attention (Parsons et al. 2011). Mating pattern of F1 hybrids also depends on mate preference of parents and hybrids (Svedin et al. 2008; Chen and Pfennig 2020), of which inheritance pattern during hybridization varies across species pairs and preference components (Table S2 for a summary of F1 female mate preferences). Yet, we are still far from drawing general patterns of hybrid mate preference because it has rarely been studied. By filling knowledge gaps in trait integration and mate preference of F1 hybrids, we can better understand mechanisms of reproductive isolation and gene flow.

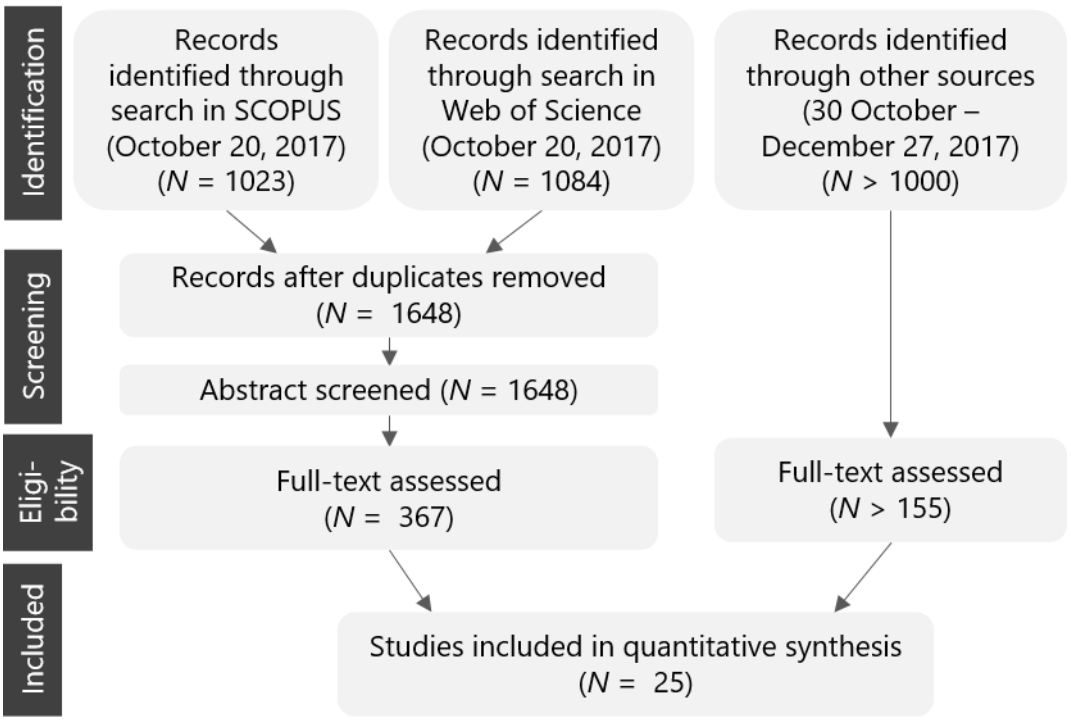
Non-additive interactions among species-specific alleles have received great attention, especially regarding genetic incompatibilities damaging hybrid fitness (Turelli and Moyle 2007; Dagilis et al. 2019). In contrast, we still know very little about how allelic interactions influence the phenotypic distribution of hybrid populations. Since allelic interactions are largely not heritable, increased phenotypic novelty or variability in F1 hybrids do not directly indicate enhanced evolvability of hybrid populations. Nevertheless, such non-additive interactions can appear in any hybrid generations, including F1 hybrids. Indeed, QTL studies in rice have shown that epistasis underlies novel phenotype expression by later generation hybrids (Mao et al. 2011; Koide et al. 2019).

Conclusions

By leveraging recent developments in meta-analysis, we have shown non-additive interactions are powerful sources of phenotypic novelty and stability in male reproductive traits. By providing phenotypic novelty and impacting phenotypic variation, non-additive allelic interactions may play key roles in determining early succession and dynamics of hybrid populations, and thus, the course of subsequent extinction or speciation. Finally, researchers can use the formal meta-analytic techniques we have developed in this study to synthesize growing empirical articles and to generate new insights into speciation.

449 ***Figures***

450



451

452

453 **Figure 1 | PRISMA diagram.** The flow of inclusion and exclusion of studies identified during the
454 literature search is shown. In brackets, we indicate the number of published literature (studies, *N*).

455

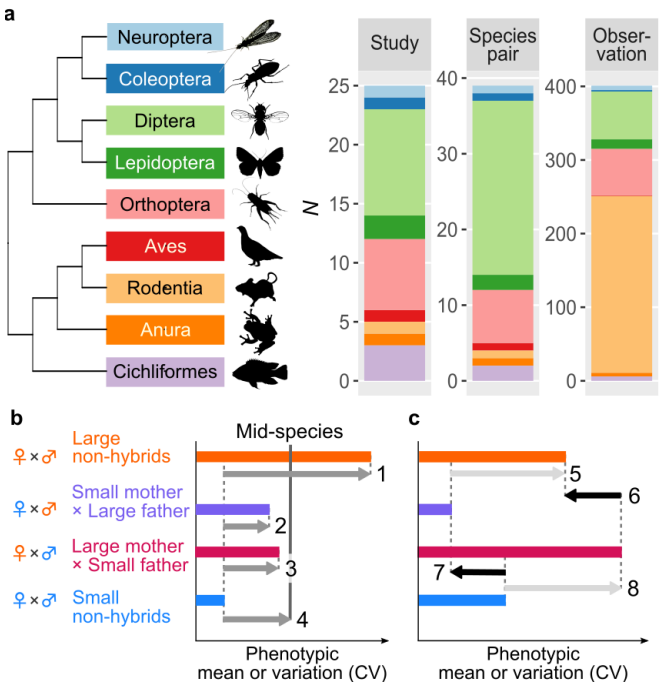


Figure 2[KA1] | **Dataset and effect size calculation. a** Taxonomic diversity of the data set, at the phylogeny, study, species and observation levels; shades of green refer to the main taxonomic groups, as on the phylogenetic tree. **b–c** Schematic representation on the ways used to calculate effect sizes that compared phenotypic mean and variation among males of hybrids and non-hybrid parental species (arrows). **b** Calculation of effect size used in the formal meta-analytic models (grey arrows). **c** Our approach to assess novel phenotype and variability expression. Black arrows indicate novel phenotypic means or variabilities, but light grey arrows do not. In this hypothetical example, hybrids from species with larger trait size or variability (large non-hybrids) females and species with smaller trait size or variability (small non-hybrids) males exceeded the large non-hybrids and maternal species (arrow 6), but the reciprocal hybrids (small female × large male) exceeded the small non-hybrids and father species (arrow 7).

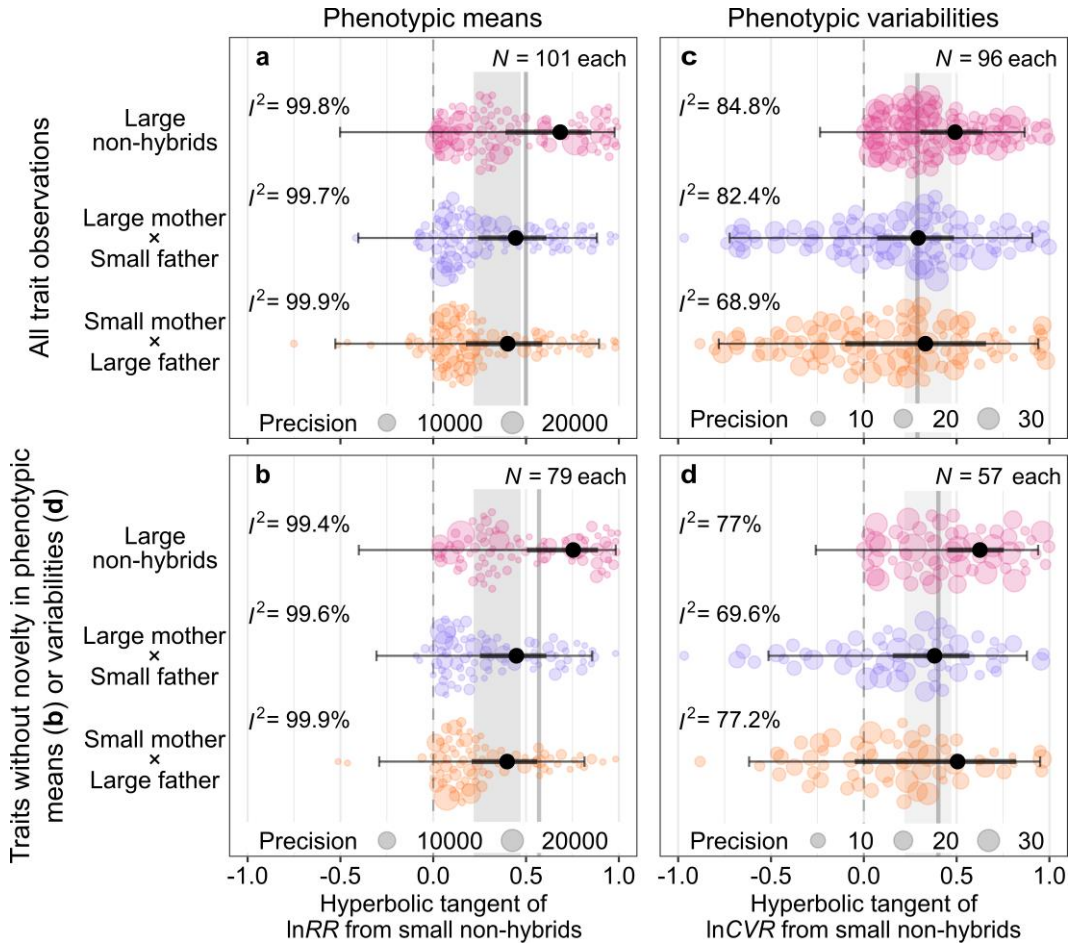


Figure 3 | Relative phenotypic means and variabilities of hybrids and large non-hybrids compared to small non-hybrids using orchard plots. The meta-analytic mean (mean effect size) is shown with its 95% confidence interval (thick line) and 95% prediction interval (thin line; see Nakagawa et al. 2021). Individual effect sizes are represented as dots proportional to sample sizes. Small and large non-hybrids indicate parental species with smaller and larger trait size for a given trait in **a** and **c**; species with smaller and larger trait size variability for a given trait in **b** and **d**. Dashed line indicates no difference from small non-hybrids, while grey vertical line indicates mid-species value (average of parental species in male trait size [**a** and **c**], and coefficient of variation of male trait size [**b** and **d**]).

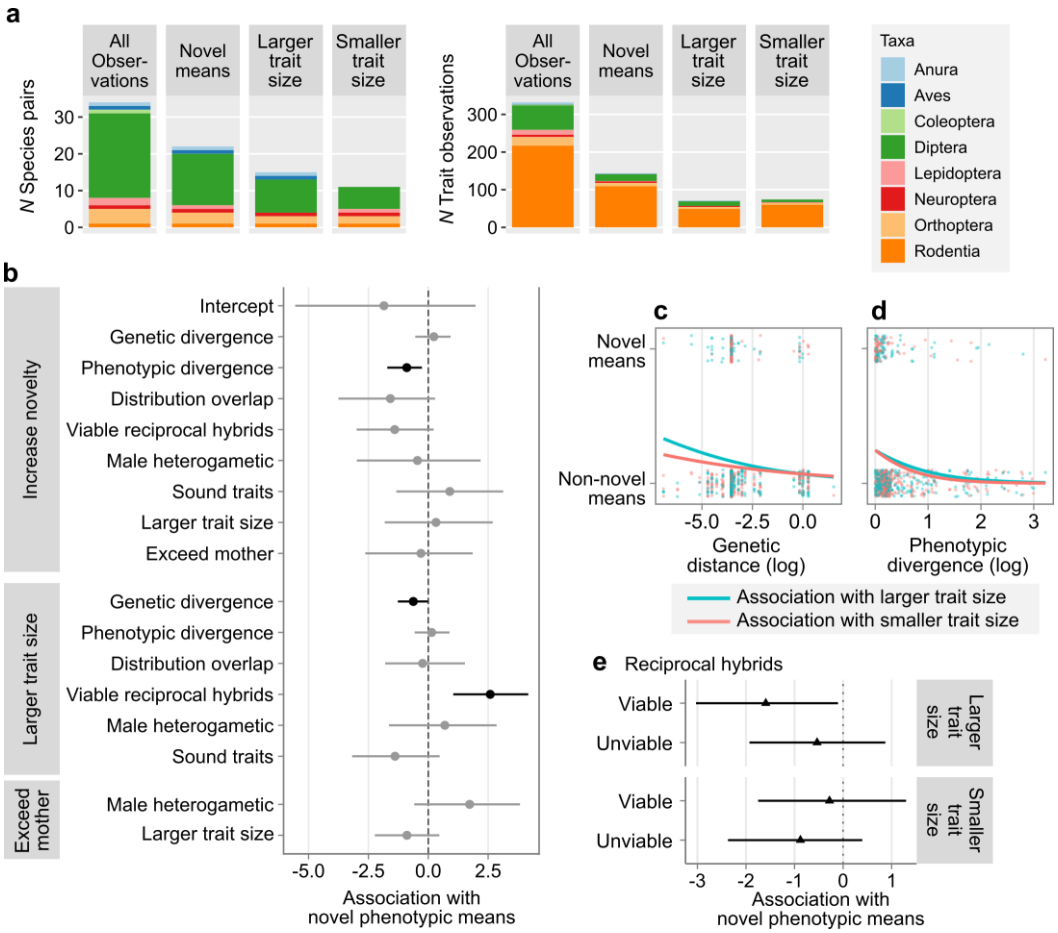


Figure 4 | Novelty in phenotypic means. a Taxonomic distribution of novel phenotypic means. Novel phenotypic means was counted per species (count of species pairs whose hybrids exhibited novel phenotype in any traits; right) and per trait observation (when hybrids exhibited novel phenotype in any direction; left). Novel phenotypes with larger/smaller trait size compared to parents were also counted. **b** Result of the full model for the probability of hybrids to exhibit novel phenotypic means (point estimates with 95% Confidence Intervals). The main effects are categorized as the factors potentially "Increase novelty" in phenotypic means. The interaction terms with compared parental species (large vs. small , and maternal vs. paternal, see Fig. 2c) were labeled as "Larger trait size" and "Exceed mother", respectively. Statistically significant and non-significant predictors are shown as black and grey, respectively. **c-e** Impacts of the three significant factors (identified in panel **b**) on novel phenotypic means with larger/smaller trait size.

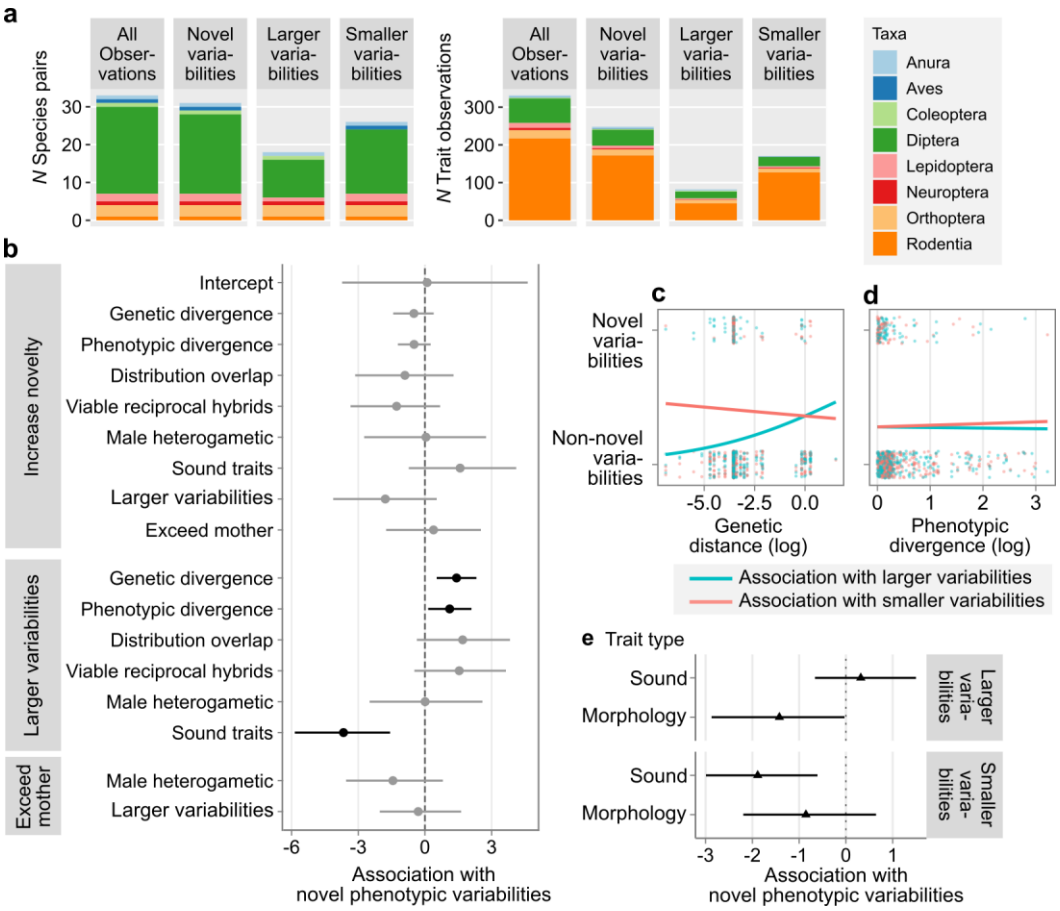


Figure 5 | Novelty in phenotypic variabilities. a Taxonomic distribution of novel phenotypic variability. Counts of novel phenotypic variability are analogous to Fig. 4a. **b** Result of the full model (point estimates with 95% CI). The categories of regression factors are analogous to Fig. 4b. Statistically significant and non-significant predictors are shown as black and grey, respectively. **c-e** Impacts of the three significant factors (identified in panel **b**) to the larger/smaller phenotypic variability.

Competing interests

The authors declare no competing interests.

References

Alibert, P., and J.-C. Auffray. 2003. Genomic coadaptation, outbreeding depression and developmental instability. Pp. 116–134 *in* M. Polak, ed. Developmental instability: causes and consequences. Oxford University Press, New York.

Arnold, M. L., and S. a Hodges. 1995. Are natural hybrids fit or unfit relative to their parents? Trends Ecol. Evol. 10:67–71.

Ashman, T. L., D. Bachtrog, H. Blackmon, E. E. Goldberg, M. W. Hahn, M. Kirkpatrick, J. Kitano, J. E. Mank, I. Mayrose, R. Ming, S. P. Otto, C. L. Peichel, M. W. Pennell, N. Perrin, L. Ross, N. Valenzuela, and J. C. Vamosi. 2014. Tree of Sex: A database of sexual systems. Sci. Data 1.

Bennet-Clark, H. C. 1998. Size and scale effects as constraints in insect sound communication. Philos. Trans. R. Soc. B 353:407–419.

Burke, J. M., and M. L. Arnold. 2001. Genetics and the fitness of hybrids. Annu. Rev. Genet. 35:31–52.

Chen, C., and K. S. Pfennig. 2020. Female toads engaging in adaptive hybridization prefer high-quality heterospecifics as mates. Science 367:1377–1379.

Chen, Z. J. 2013. Genomic and epigenetic insights into the molecular bases of heterosis. Nat. Rev. Genet. 14:471–482.

Cockett, N. E., S. P. Jackson, T. L. Shay, F. Farnir, S. Berghmans, G. D. Snowder, D. M. Nielsen, and M. Georges. 1996. Polar overdominance at the ovine callipyge locus. Science 273:236–238.

Dagilis, A. J., M. Kirkpatrick, and D. I. Bolnick. 2019. The evolution of hybrid fitness during speciation. PLoS Genet. 15:e1008125.

DeVicente, M. C., and S. D. Tanksley. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585–596.

Edmands, S. 1999. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. Evolution 53:1757.

Grant, P. R., and B. R. Grant. 2019. Hybridization increases population variation during adaptive radiation. Proc. Natl. Acad. Sci. 116:23216–23224.

Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R Package. J. Stat. Softw. 33:1–22.

Hadfield, J. D., and S. Nakagawa. 2010. General quantitative genetic methods for comparative biology: Phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J. Evol. Biol. 23:494–508.

Haldane, J. B. S. 1922. Sex ratio and unisexual sterility in hybrid animals. J. Genet. 12:101–109.

Hedges, L. V., J. Gurevitch, and P. S. Curtis. 1999. The meta-analysis of response ratios in experimental ecology.

Hinchliff, C. E., S. A. Smith, J. F. Allman, J. G. Burleigh, R. Chaudhary, L. M. Coghill, K. A. Crandall, J. Deng, B. T. Drew, R. Gazis, K. Gude, D. S. Hibbett, L. A. Katz, H. Dail Laughinghouse, E. J. McTavish, P. E. Midford, C. L. Owen, R. H. Ree, J. A. Rees, D. E. Soltisc, T. Williams, and K. A. Cranston. 2015. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc. Natl. Acad. Sci. 112:12764–12769.

Janick, J. 1999. Exploitation of Heterosis: Uniformity and Stability. Pp. 319–333 *in* J. G. Coors and S. Pandey, eds. The Genetics and Exploitation of Heterosis in Crops. CIMMYT, Madison, WI, USA.

Kagawa, K., and G. Takimoto. 2018. Hybridization can promote adaptive radiation by means of transgressive segregation. Ecol. Lett. 21:264–274.

Koide, Y., S. Sakaguchi, T. Uchiyama, Y. Ota, A. Tezuka, A. J. Nagano, S. Ishiguro, I. Takamure,

and Y. Kishima. 2019. Genetic properties responsible for the transgressive segregation of days to heading in rice. G3 Genes, Genomes, Genet., doi: 10.1534/g3.119.201011.

Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33:1870–1874.

Lamkey, K. R., and J. W. Edwards. 1999. Quantitative genetics of heterosis. Pp. 31–48 *in* J. G. Coors and S. Pandey, eds. The Genetics and Exploitation of Heterosis in Crops. CIMMYT, Mexico City.

Mao, D., T. Liu, C. Xu, X. Li, and Y. Xing. 2011. Epistasis and complementary gene action adequately account for the genetic bases of transgressive segregation of kilo-grain weight in rice. Euphytica 180:261–271.

Matsubayashi, K. W., I. Ohshima, and P. Nosil. 2010. Ecological speciation in phytophagous insects. Entomol. Exp. Appl. 134:1–27.

Michonneau, F., J. W. Brown, and D. J. Winter. 2016. rotl: an R package to interact with the Open Tree of Life data. Methods Ecol. Evol. 7:1476–1481.

Moll, R. H., J. H. Lonnquist, J. V. Fortuno, and E. C. Johnson. 1965. The relationship of heterosis and genetic divergence in maize. Genetics 52:139–144.

Moore, M. P., H. H. Whiteman, and R. A. Martin. 2019. A mother's legacy: the strength of maternal effects in animal populations. Ecol. Lett. 1620–1628.

Moyle, L. C., and T. Nakazato. 2010. Hybrid incompatibility "snowballs" between Solanum species. Science 329:1521–1523.

Nakagawa, S., M. Lagisz, R. E. O'Dea, J. Rutkowska, Y. Yang, D. W. A. Noble, and A. M. Senior. 2021. The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond. Res. Synth. Methods 12:4–12.

Nakagawa, S., R. Poulin, K. Mengersen, K. Reinhold, L. Engqvist, M. Lagisz, and A. M. Senior. 2015. Meta-analysis of variation: Ecological and evolutionary applications and beyond.

582 Methods Ecol. Evol. 6:143–152.
583 Paradis, E., and K. Schliep. 2019. Ape 5.0: An environment for modern phylogenetics and
584 evolutionary analyses in R. Bioinformatics 35:526–528.
585 Parsons, K. J., Y. H. Son, and R. Craig Albertson. 2011. Hybridization promotes evolvability in
586 African cichlids: Connections between transgressive segregation and phenotypic integration.
587 Evol. Biol. 38:306–315.
588 Pick, J. L., S. Nakagawa, and D. W. A. Noble. 2019. Reproducible, flexible and high-throughput
589 data extraction from primary literature: The metaDigitise r package. Methods Ecol. Evol.
590 10:426–431.
591 Poissant, J., A. J. Wilson, and D. W. Coltman. 2010. Sex-specific genetic variance and the evolution
592 of sexual dimorphism: A systematic review of cross-sex genetic correlations. Evolution 64:97–
593 107.
594 R Core Team. 2020. R: A language and environment for statistical computing.
595 Reinhold, K. 2002. Maternal effects and the evolution of behavioral and morphological characters: A
596 literature review indicates the importance of extended maternal care. J. Hered. 93:400–405.
597 Reinhold, K., and L. Engqvist. 2013. The variability is in the sex chromosomes. Evolution 67:3662–
598 3668.
599 Rieseberg, L. H., M. A. Archer, and R. K. Wayne. 1999. Transgressive segregation, adaptation and
600 speciation. Heredity 83:363–372.
601 Rieseberg, L. H., and N. C. Ellstrand. 1993. What can molecular and morphological markers tell us
602 about plant hybridization? CRC. Crit. Rev. Plant Sci. 12:213–241.
603 Rieseberg, L. H., A. Widmer, A. M. Arntz, J. M. Burke, D. E. Carr, R. J. Abbott, and T. R. Meagher.
604 2003. The genetic architecture necessary for transgressive segregation is common in both
605 natural and domesticated populations. Philos. Trans. R. Soc. B 358:1141–1147.
606 Riggs, T. J. 1988. Breeding F1 hybrid varieties of vegetables. J. Hortic. Sci. 63:369–382.

607 Rosenthal, G. G. 2013. Individual mating decisions and hybridization. J. Evol. Biol. 26:252–255.
608 Ryan, M. J., and E. A. Brenowitz. 1985. The role of body size, phylogeny, and ambient noise in the
609 evolution of bird song. Am. Nat. 126:87–100.
610 Schilthuizen, M., M. C. W. G. Giesbers, and L. W. Beukeboom. 2011. Haldane's rule in the 21st
611 century. Heredity 107:95–102.
612 Selz, O. M., K. Lucek, K. A. Young, and O. Seehausen. 2014. Relaxed trait covariance in
613 interspecific cichlid hybrids predicts morphological diversity in adaptive radiations. J. Evol.
614 Biol. 27:11–24.
615 Senior, A. M., W. Viechtbauer, and S. Nakagawa. 2020. Revisiting and expanding the meta-analysis
616 of variation: The log coefficient of variation ratio. Res. Synth. Methods 11:553–567.
617 Stelkens, R. B., C. Schmid, O. Selz, and O. Seehausen. 2009. Phenotypic novelty in experimental
618 hybrids is predicted by the genetic distance between species of cichlid fish. BMC Evol. Biol. 9.
619 Stelkens, R., and O. Seehausen. 2009. Genetic distance between species predicts novel trait
620 expression in their hybrids. Evolution 63:884–897.
621 Svedin, N., C. Wiley, T. Veen, L. Gustafsson, and A. Qvarnstrom. 2008. Natural and sexual
622 selection against hybrid flycatchers. Proc. R. Soc. B 275:735–744.
623 Tamura, K., M. Nei, and S. Kumar. 2004. Prospects for inferring very large phylogenies by using the
624 neighbor-joining method. Proc. Natl. Acad. Sci. 101:11030–11035.
625 Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: Improving the sensitivity of
626 progressive multiple sequence alignment through sequence weighting, position-specific gap
627 penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
628 Thompson, K. A., M. Urquhart-Cronish, K. D. Whitney, L. H. Rieseberg, and D. Schluter. 2021.
629 Patterns, Predictors, and Consequences of Dominance in Hybrids. Am. Nat. 197:E000–E000.
630 Turelli, M., and L. C. Moyle. 2007. Asymmetric postmating isolation: Darwin's corollary to
631 Haldane's rule. Genetics 176:1059–1088.

632 Vetukhiv, M. 1953. Viability of hybrids between local populations of *Drosophila pseudoobscura*.
633 Proc. Natl. Acad. Sci. 39:30–34.
634 Viechtbauer, W. 2010. Conducting meta-analyses in R with the metafor. J. Stat. Softw. 36:1–48.
635 Wallace, B. 1955. Inter-population hybrids in *Drosophila melanogaster*. Evolution 9:302.
636 Wan, X., W. Wang, J. Liu, and T. Tong. 2014. Estimating the sample mean and standard deviation
637 from the sample size, median, range and/or interquartile range. BMC Med. Res. Methodol.
638 14:1–13.
639 Wei, X., and J. Zhang. 2018. The optimal mating distance resulting from heterosis and genetic
640 incompatibility. Sci. Adv. 4:1–8.
641