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

Keisuke Atsumi[1*], Malgorzata Lagisz[2] and Shinichi Nakagawa[2]

[1] Graduate School of Environmental Science, Hokkaido University, N10W5 Sapporo, Hokkaido 060-0810, Japan

[2] Evolution & Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales, Australia

[*] Corresponding author: Keisuke Atsumi (k.atsumi115@gmail.com)

Summary

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 3208 effect sizes from 39 animal species pairs. Although additivity sets phenotypic distributions of F1s to be intermediate, F1s also showed dominance and maternal inheritance. 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.

(Introduction)

The long history of animal and plant breeding has taught us that hybridization is a powerful source of phenotypic novelty. While hybrid populations usually distribute their phenotypes throughout the range of parental lineages, they often exhibit novel phenotypes, as well as novel variability (i.e. phenotype mean and variance of hybrid populations beyond the range of parental species)[1–4]. Novel phenotypes enable hybrids to exploit novel niches and ultimately become new speces[5,6]. Naturally, the distribution of hybrids' phenotypic traits (hereafter, hybrid phenotypic distributions) 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[1,2,7].

Novel phenotypes are produced by various mechanisms, such as chromosomal recombination, dominance, epistasis, and maternal/paternal inheritance[1,2,8–12]. 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 an intermediate phenotype (i.e. the mid-parent value, averaged trait value of both parental species). However, F1 hybrids often resemble more one of the two parents (dominance), and have different phenotypes between reciprocal crosses[13]. Moreover, F1 hybrids sometimes exhibit novel phenotypes[1,2,14]. Although novel variability — especially smaller phenotypic variation in F1 hybrids — has been utilized in agriculture to enhance yield stability[3], the prevalence of novel variations has not been examined[4,15–17]. Fragmented evidence above suggests that non-additive genetic effects are important sources of novel phenotypic distributions of hybrid populations. 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.

Here we investigate phenotypic mean and variability of F1 hybrids of male mating traits (i.e. traits used during mating that females do not possess) across animal taxa (Fig. 1a). By focusing on male mating traits, we exclude the possible biases resulting from any sex differences in phenotypic distributions. Given a wide prevalence of sexual dimorphism[18–20], 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's rule)[21,22]. Male sexual traits also directly relate to the reproductive success of hybrids.

By employing formal phylogenetically controlled meta-analytic techniques for the first time, we test if phenotypic mean and variability of F1 hybrids are larger or smaller than mid-parent value (Fig. 1b), and are affected by crossing direction (father from species A with mother from species B, or vice versa; Fig. 1c). 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 that influence phenotypic novelty and variability of male mating traits in F1 hybrids (Fig. 3 and 4).

Results

Although trait inheritance pattern is the core 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 Information, section *Data description*). We extracted two sets of 1604 effect sizes comparing phenotypic mean (lnRR) and variability (lnCVR), 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. 1a) and only 13% of species pairs (5/39) were vertebrates, including bony fish, frog, bird and rodent species. Most species pairs were male heterogametic (92%, 34/37 species pairs, excluding two fishes of which sex-determination system is unknown). Both reciprocal crosses were viable in 59% of species pairs (23/39); and geographic ranges of parental lineages overlap in 36% of species pairs (14/39).

F1 hybrids show dominance and maternal inheritance

Phenotypic mean of F1 hybrids resembled that of parental species with smaller phenotypic mean ($spSS_M$, partial dominance) and also maternal species (maternal inheritance). These findings clearly indicate discriminating reciprocal cross directions is essential to understanding the inheritance mode of traits during hybridization. Phenotypic means of F1 hybrids from both cross directions ($hybLS_M$ and $hybSL_M$; see Fig. 1b) were significantly smaller than mid-parent value by 12.5% and 16.9%, respectively (Fig. 2a; $P < 0.001$ each), indicating partial dominance biasing toward smaller trait size. This trend was qualitatively the same when we excluded traits showing novel phenotype (i.e. values outside the range of both parental species; Fig. 2c).

Because we focused on male mating traits, we initially expected that hybrids resemble father species. Conversely, male-mating traits of F1 hybrids were similar to those of the males of maternal species

(*hybLS*$_M$ was significantly larger than *hybSL*$_M$ by 4.9%; Fig. 2a; $P < 0.001$). This pattern is consistent with non-male-mating traits across animals and plants[13]. Given that maternal inheritance is particularly profound for morphology, including body size[23], mothers seem to directly influence male-mating morphological traits, and even sound traits that are partially determined by morphology and body size[24,25]. Maternal effects on male-mating traits may mask effect of male sex chromosome on these traits. Indeed, previous comparative analyses have shown that sexual dimorphism is not associated with sex determination systems, suggesting little effect of sex chromosome on male-mating traits[26–28].

In contrast to phenotypic mean, average phenotypic variability of male F1 hybrids was similar to the average value of parents' variability (Fig. 1b), which aligns with previous comparative study on non-male-mating traits across animals and plants[13]. A caveat is that hybrid phenotypic variability is either larger or smaller than that of both parents in most trait observations (74.9%, see *Novel variability*), resulting in large heterogeneity in parent–hybrid difference in phenotypic variability (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. 2c, d). Due to similar mating traits with moderate phenotypic variability, F1 hybrid males may tend to backcross more often with their mother species or with parental species with smaller phenotypic mean, potentially biasing the direction of gene flow and, therefore, influencing the course of evolution. Nonetheless, the magnitude of phenotypic difference between hybrids and parental species is highly heterogenous (total I^2 = 99.7%, which are partitioned into phylogeny I^2 = 38.8%, study I^2 = 1.1%, crossed lineage I^2 = ~0%, residual I^2 = 59.8%; Fig 2a, b), suggesting that any inheritance patterns are plausible in F1 hybrids (e.g., dominance, maternal inheritance and transgressive segregation).

Novel phenotype

F1 hybrids exhibit novel phenotypes in 64.7% (22/34) of species pairs, and 42.9% (143/333) of trait observations (Fig. 3a), indicating that non-additive genetic interaction is a powerful source of phenotypic novelty in male mating traits. 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[13]. Discordance with our result 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. Other comparative studies, which included F1 and later generation hybrids across animals, found that 29–

31% of traits showed novel phenotypes[1,2]. While non-additive interactions can enhance phenotypic novelty in any generation hybrids, theory suggests recombination after F1 hybrids is the major source of phenotypic novelty[1]. Novel phenotype is thus expected to be more frequent in later generation hybrids than in F1 hybrids. The selection against novel phenotypes may explain the unexpected contrast of our findings with previous comparative studies. While our study exclusively focused on experimentally-derived F1 hybrids, previous studies included natural hybrid populations[4,16]. In natural hybrid populations, extrinsic natural selection may remove novel phenotypes and lead to underestimation of phenotypic novelty. Hence, our work shows that hybrid populations may express novel phenotypes much more frequently than previously thought.

We anticipated that both genetic and phenotypic divergences between parental species positively relate to phenotypic novelty. This is because species divergence should be positively associated with the number of heterozygotic loci in F1 hybrids that enable novel phenotype expression[14] (see Supplemental Table S1 for the summary of hypotheses). However, genetic divergence between parents did not significantly associate with phenotypic novelty (main effect: $\beta = 0.26$, 95% credible interval [CI] = -0.41 – 0.9, $P = 0.420$: Fig. 3b). Rather, increasing phenotypic divergence between parental species generally reduced phenotypic novelty (main effect of phenotypic divergence: $\beta =$ -0.8, CI = -1.48 – -0.23, $P < 0.001$). Our prediction was not supported, presumably because the effects of inter-allelic interactions can be diverse. We assumed that inter-allelic interactions among those loci act in the same direction (e.g., all interactions increase trait value)[14]. If the sign varies among the interactions, however, genetic differentiation no longer necessarily increases phenotypic novelty. Alternatively, similar parental species could be heterozygous at phenotype-determining loci, which can allow novel phenotype expression by homozygous F1 hybrid individuals[29].

F1 hybrids without viable reciprocal cross tended to exhibit smaller trait mean than both parents (interaction term of reciprocal hybrids' viability with the relative trait size compared to parents: $\beta =$ 2.51, CI = 1.01 – 4.01, $P = 0.001$: Fig. 3b, see also Fig. 3e) and their phenotypic novelty tended to increase (main effect of reciprocal hybrids' viability: $\beta =$ -1.36, CI = -2.97 – 0.04, $P = 0.064$: Fig. 3b). Since inviability of the reciprocal hybrid cross can indicate stronger genetic incompatibility, our results imply that genetic incompatibility tends to increase phenotypic novelty of hybrids by reducing trait value (i.e. outbreeding depression). This finding could explain why hybrids from genetically diverged parental species pairs – likely to develop genetic incompatibility – rarely had larger phenotypic mean than parents (interaction term of genetic divergence with the relative trait

size: β = -0.59, CI = -1.19 – 0.02, P = 0.041: Fig. 3b, see also Fig. 3c).

Novel variability

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. 4a), showing that phenotypic variability is usually inherited in non-additive ways during hybridization. Compared to the parents, hybrids exhibited smaller phenotypic variability in more than half of trait observations (51.4%, Fig. 4a) and greater variability less frequently (24.8%, Fig. 4a). Our results suggest that, counterintuitively, heterozygosity often tends to diminish phenotypic variability of hybrids, although sometimes it may also increase variability. Importantly, novel phenotypes were more variable than non-novel phenotypes of which mean lie within the range of parents (ordinal phylogenetic random regression: β = 1.01, CI = 0.45 – 1.62, $P < 0.001$). That is, novel phenotypes, regardless of relative trait value 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 %). This in turn indicates that phenotypic instability – characterized by the large phenotypic variability[4,30] – enhances phenotypic novelty of F1 hybrids. However, this is not a universal pattern, as smaller variability than parents was also frequently observed in both novel and non-novel phenotypes (40.4% and 49.2%). Even if F1 hybrids exhibit smaller phenotypic variability than parents, later-generation hybrids expand phenotypic variation through recombination[1,4,13]. Hence, the wide prevalence of small phenotypic variability in F1 does not indicate reduced evolvability of hybrid populations. Rather, our results emphasize the time lag between hybridizing event and expression of novel phenotypic variation in hybrid population[31].

We revealed that both genetic and phenotypic divergence between parental species enhanced phenotypic variability of F1 hybrids. F1 hybrids between genetically distant parents tended 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: β = 1.25, CI = 0.47 – 2.00, P = 0.002: Fig. 4b, see also Fig. 4c). Increasing variability with parental genetic divergence was also reported in later generation hybrids[2,29]. The common pattern in F1 and later generation hybrids implies that non-additive interactions contribute to phenotypic distribution of hybrid population. Additionally, genome-wide incompatibility may magnify phenotypic variability of F1 hybrids between well-diverged parents by

damaging developmental stability or phenotypic robustness[17] (see Supplemental Table S1 for the summary of hypotheses). On the other hand, genetically close parents yielded F1 hybrids with smaller phenotypic variability, possibly due to enhanced developmental stability (i.e. hybrid vigor)[4], rather than due to selective mortality arising from genetic incompatibility (i.e. individuals with anomalous phenotypes, resulting from epistatic interactions, are likely to die). Hybrid vigor is often observed in inbred strains due to the release from inbreeding depression[1,14] and expected in hybrids between genetically similar parents[32]. Notably, 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. 4b). Well-diverged traits may provide more diverse epistatic interactions among the relevant loci and increase phenotypic variability in F1 hybrids[4]. Note that novel variability did not associate with phenotypic divergence when species-level moderators were not included in the model (Fig. 4d).

We predicted that novel variability is more frequent in sound traits than in morphological traits, because of lower heritability of acoustic signals[33] (Supplementary Table S1). We found, however, that 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. 4b: see also Fig. 4e). 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.

Discussion

Novel mating phenotype of F1 hybrids could facilitate mating among F1s under trait-based assortative mating[5,34]. Meanwhile, large phenotypic variability of F1 hybrids would increase the overlap of phenotypic distribution with parents, thereby facilitating backcross and gene flow. Hence, under trait-based assortative mating, phenotypic similarity between parents could facilitate mating among F1 hybrids and limit backcrossing by increasing phenotypic novelty or reducing phenotypic variability. In addition to each trait value, trait integration can influence the attractiveness of F1 males, because multiple mating traits often interactively determine sexual attractiveness[35–37]. We found that trait-level factor (relative trait values of parents) influences F1 hybrid phenotypic mean

and variability (Fig. 3b and 4b). Our additional analysis further 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 Information, section *Trait mosaicism*), which was previously shown in non-male mating traits[13]. These facts indicate that phenotypic integration of parents could easily break in F1 hybrids, as reported in several studies[13,38–41] (but see[42]). Moreover, a variety of trait integration patterns can arise in F1 individuals because phenotypic variability varies among traits (according to relative trait values of parents: Fig. 4b). Despite the importance of determining fitness and mating pattern, mating traits integration of F1 hybrids has received little attention[13,31,43]. Mating pattern of F1 hybrids also depends on mate preference of parents and hybrids[44,45], of which inheritance pattern during hybridization varies across species pairs and preference components (Supplementary Table S2 for a summary of F1 female mate preferences)[46,47]. 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[32]. In contrast, we still know little how non-additive interactions influence in the phenotypic distribution of hybrid populations[48]. Since non-additive interactions are largely not heritable[49], increased phenotypic novelty or variability in F1 hybrids do not directly indicate enhanced evolvability of hybrid populations. For interactions among loci (epistasis) to become heritable, hybrids need to develop linkage disequilibrium[50]. Selfing is required to fix heterozygosity causing dominance[51], which is particularly difficult for animals. Nevertheless, such non-additive interactions can appear in any hybrid generations, including F1 hybrids. 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 insight into speciation.

Figures

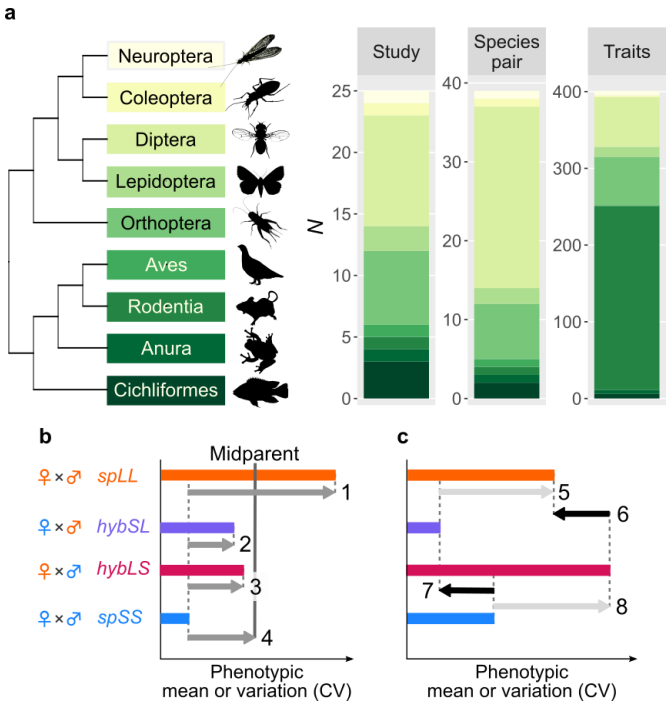


Fig. 1 | Dataset and effect size calculation. a Taxonomic diversity of the data set, at the phylogeny, study, species and observation levels; shades of green refer to the main taxonomic groups, as on the phylogenetic tree. Our systematic review identified 25 studies used 39 species pairs across animals that reported a total of 401 trait observations. **b–c** Schematic representation on the ways used to calculate effect sizes that compared phenotypic mean and variation among hybrids and parental species (arrows). **b** Calculation of effect size used in the formal meta-analytic models. To draw general patterns of phenotypic distribution of F1 hybrids, we quantified the difference in phenotypic mean and variability from one parental species (*spSS*) to each reciprocal hybrid cross (*hybLS* and *hybSL*, having *spSS* as father and mother, respectively), the other parental species (*spLL*) and midparent (the average of phenotypic mean or variability between two parental species). Comparisons are represented as grey arrows. *SpSS* was defined as the parental species with smaller phenotypic mean (denoted as $spSS_M$), in calculating differences in phenotypic mean. In calculating differences in phenotypic variability, *spSS* was defined as the parental species with smaller phenotypic variability (denoted as $spSS_V$). **c** Our approach to assess novel phenotype and variability

expression. By comparing phenotypic mean and variability between *spLL* and each hybrid crosses and between *spSS* and each hybrid crosses, we judged if F1 hybrids expressed novel phenotype and variability. These comparisons allowed us to determine relative trait size or variability comparing to parent species. That is, hybrids exceed upper or lower range of both parents' phenotypic mean or variability (i.e. greater or smaller trait size or variability than both parents) and; hybrids exceeded mother or father species' phenotypic mean or variability. Black arrows indicate novel phenotype or variability expression, but light grey arrows do not. In this hypothetical example, *hybLS* exceeded the upper range and mother species (arrow 6), but *hybSL* exceeded the lower range and father species (arrow 7).

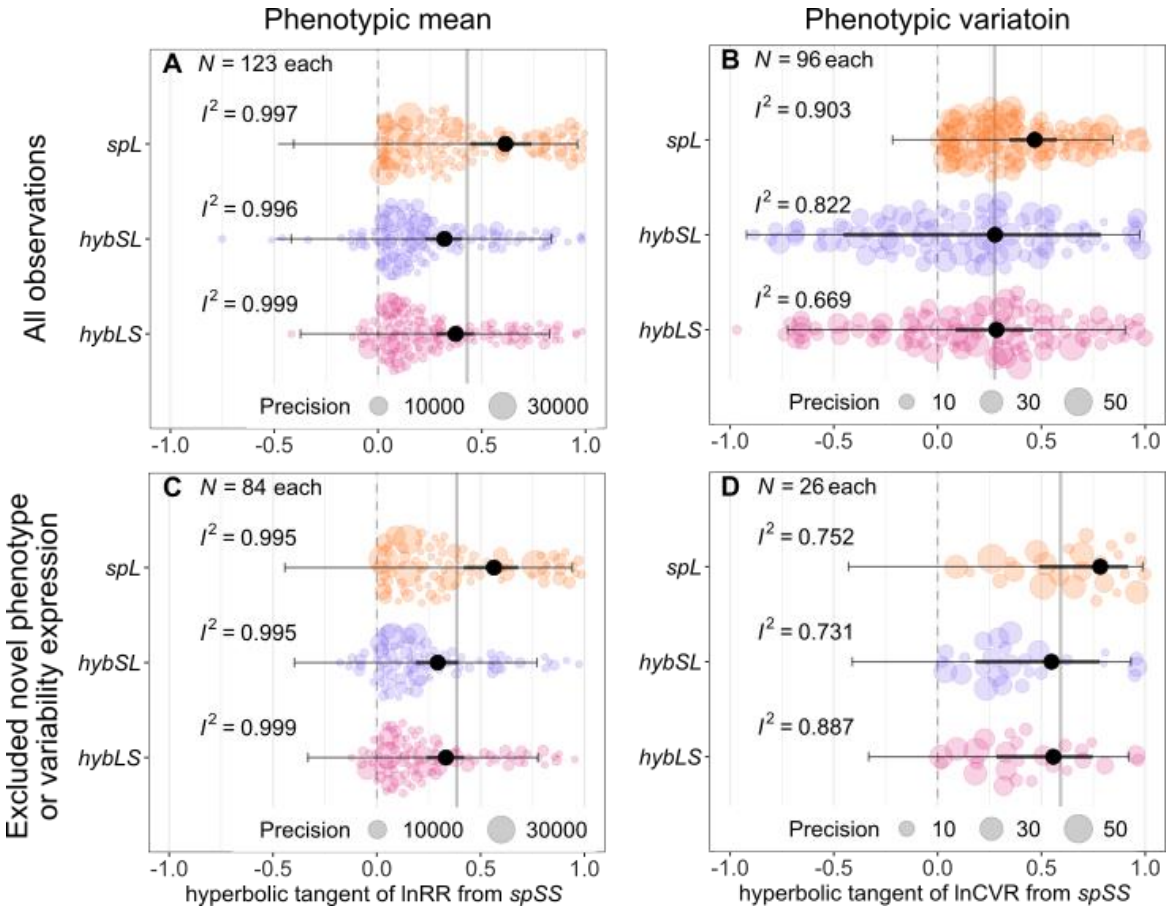


Fig. 2 | Relative phenotypic means and variations of hybrids *hybSL* and *hybLS* and parental species *spLL* compared to *spSS* (see Fig. 1). The meta-analytic mean (mean effect size) is shown with its 95% confidence interval (thick line) and 95% prediction interval (thin line). Individual effect sizes are represented as dots proportional to sample sizes. Dashed line indicates no difference from the *spSS* parental species (species with smaller phenotypic mean for a given trait in **a** and **c**; species with smaller phenotypic variability for a given trait in **b** and **d**), while grey vertical line indicates intermediate value between both parental species (midparent). **a** Differences in mean from parental species with smaller phenotypic mean ($spSS_M$) to each hybrid crosses ($hybLS_M$ and $hybSL_M$, having $spSS_M$ as father and mother, respectively) were highly heterogeneous across observations (represented by I^2). **b** Across observations, hybrids more varied in their phenotypic variability than parental species because of the frequent novel variability expression. Without the observations with novel variability expression (**d**), heterogeneity in phenotypic variation of hybrids is smaller than

phenotypic variation of parental species.

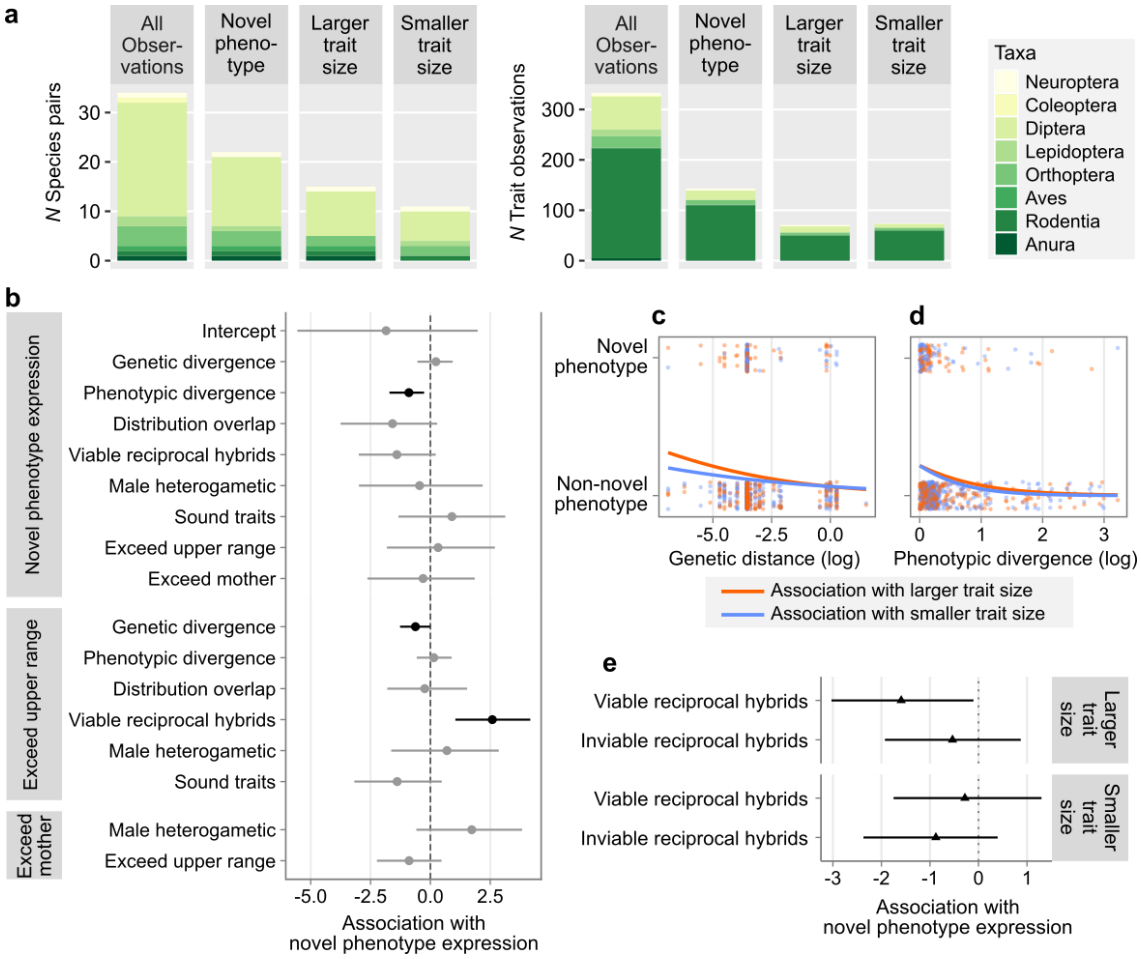


Fig. 3 | Novel phenotype expression. a Taxonomic distribution of novel phenotype expression (i.e. cases when hybrids' phenotypic mean lies outside the range of parental species, transgressive segregation). Novel phenotype expression was counted per species (count of species pairs whose hybrids exhibited novel phenotype in any traits in any direction; right) and per trait observation (when hybrids exhibited novel phenotype in any direction; left). Novel phenotypes with larger/smaller trait value compared to parents were also counted. **b** Result of the full model of Bayesian phylogenetic mixed logistic regression for novel phenotype expression probability (point estimates with 95% Confidence Intervals). Hypotheses for each factor are summarized in Supplementary Table S1. The main effects are categorized as the factors potentially increasing "novel phenotype expression", because those effects indicate overall impacts of each term on phenotypic novelty in any direction. To assess factors affecting direction of phenotypic novelty, the full model also included the interaction terms with compared parental species ($spLL_M$ vs. $spSS_M$, and mother species vs. father, see Fig. 1c), labeled

as “exceed upper range” and “exceed mother”, respectively. Compared parents used in judging the phenotypic novelty indicate the direction of novelty. For example, “novel phenotype exceeded $spLL_M$ and $spSS_M$” means that hybrids exceeded upper and lower phenotypic range of parents (greater or smaller trait size than parents), respectively. Therefore, these interaction terms show how the factors differently impacted on novel phenotype expression, depending on the direction of novelty. The interaction terms labeled as “exceed upper range” indicate how the factors biased novel phenotype toward larger trait size. The interaction terms labeled as “exceed mother” indicate how the factor biased novel phenotype toward exceeding the mother species’ phenotypic mean. 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 phenotype expression with larger/smaller trait size. In each panel, we conducted separate Bayesian phylogenetic mixed model that included the focal predictor variable, compared parental species ($spLL_M$ or $spSS_M$), and interaction between them, as predictors.

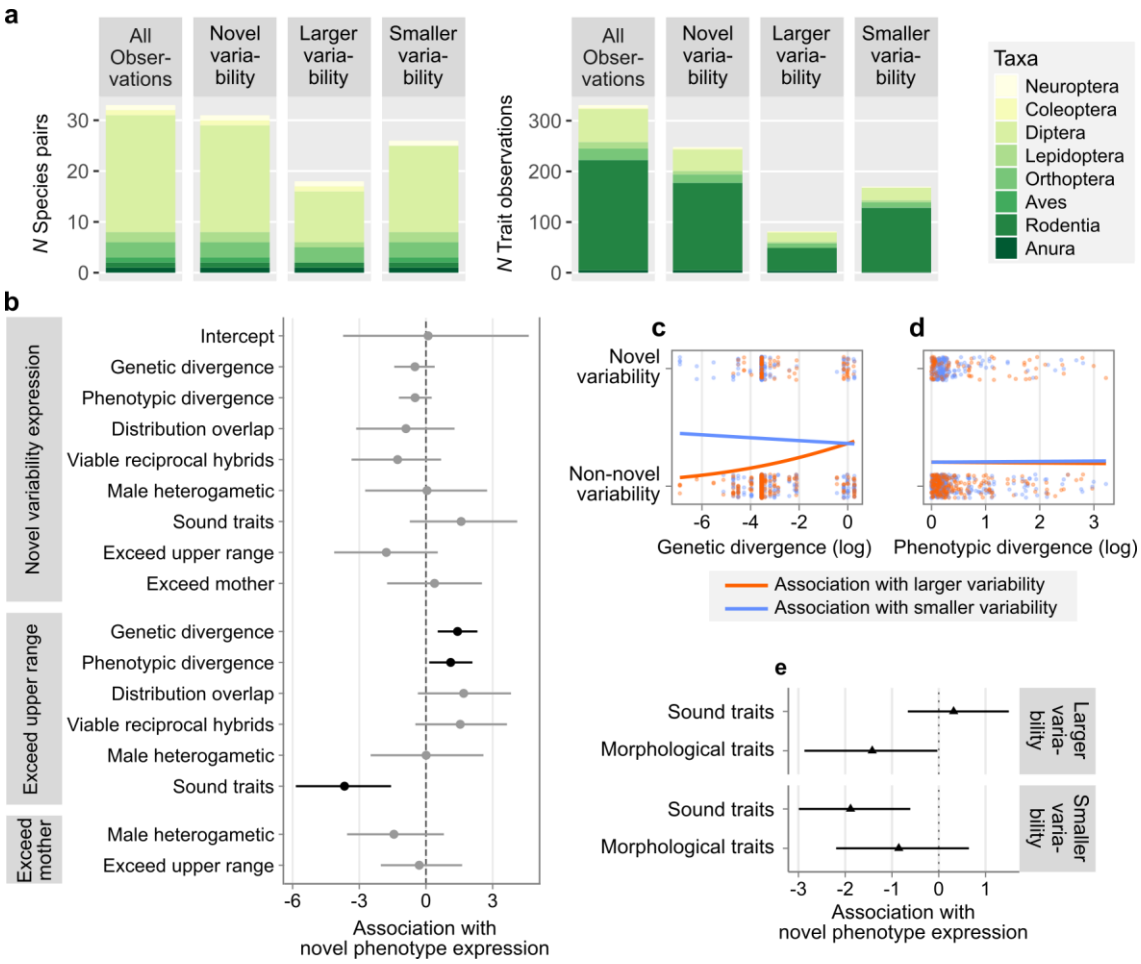


Fig. 4 | Novel variability expression. a Taxonomic distribution of novel variability expression (i.e. cases when hybrids' phenotypic variation lies outside the range of parental species). Counts of novel variability expression are analogous to Fig. 3a. **b** Result of the full model of Bayesian phylogenetic mixed logistic regression for novel variability expression (point estimates with 95% CI). Hypotheses for each factor are summarized in Supplementary Table S1. The categories of regression factors are analogous to Fig. 3b. The main effects are categorized as the factors potentially increasing "novel variability expression". The interaction terms labeled as "Exceed upper range" show how the factors biased novel variability toward larger trait variability (hybrids exceeded $spLL_V$ rather than $spSS_V$). The interaction terms labeled as "exceed mother" indicate how the factors biased novel variability toward exceeding the phenotypic variation of mother. 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 expression of novel larger/smaller variability. In each panel, we conducted separate Bayesian phylogenetic mixed models that included the focal predictor variable, compared parental species ($spLL_V$ or

$spSS_V$) and interaction between them as predictors.

Methods

Literature search and data collection We conducted systematic review of literature and followed Preferred Reporting Items for Meta-Analyses (PRISMA, shown in Extended Data Fig. S1) when reporting our meta-analytic workflow. We searched literature 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 (search date October 20, 2017). We also searched literature through backward / forward citations of reviews and meta-analysis of hybrid fitness and phenotype[2,21,52–58].

To be included in our meta-analysis, the primary study had to (*i*) report means and variance of male sexual phenotypes in hybrids and both pure crosses (i.e. parental species; note that we use the offspring from pure crosses as the proxy of the phenotypes of the parental species, the "parents"), and separately report the phenotype of reciprocal hybrid crosses, (*ii*) use different species or subspecies for the crosses, (*iii*) experimentally cross both hybrids and pure crosses and raise them in same environment, (*iv*) 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). This meta-analysis did not include books, theses, annuals or meeting reports in the search results. From original 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[59]. As the number of replicates, we used the number of studied individuals, but not the number of measurements. We estimated standard deviation from range data, if necessary[60]. The observations that had any negative trait values were removed from the dataset (83 observations from two studies that reported a total of 350 observations[61,62]), but this procedure did not reduce the total number of included studies and species. Male mating traits were classified into morphological (genital morphology and nuptial color pattern) or sound traits (stridulation and vocal song).

We collected the following predictor variables, which we a priori expected to explain variation in effect sizes: genetic and phenotypic divergence between parental species; heterogametic sex; distribution overlap between parental species, and inviability of the reciprocal hybrid cross. We calculated genetic distances for all parental species pairs using the *cytochrome c oxidase I* (*COI*) DNA sequence and then used natural logarithm of genetic distances in meta-regression. Up to 20

sequences per parental species (depending on their availability on NCBI GenBank) were taken from GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) and aligned in ClustalW[63]. Alignments with 450 bp as the minimum length were manually optimized for calculating the genetic distance. Then, we calculated maximum composite likelihood distance with uniform substitution rate and homogenous pattern[64] in the software MEGA 7.0.26[65]. If multiple sequences were available for a pair of species, we calculated the average of all possible pairs of sequences. We calculated phenotypic divergence as the absolute value of the natural logarithm of the ratio between the phenotypic means. We assigned heterogametic sex at family or genus level, if deemed evolutionary constant within the taxon[66–72]. For this reason, we did not assign value of the heterogametic sex in cichlid fishes that undergo frequent evolutionary changes in sex-determination system[73]. Range overlap of each species pair was classified into binary status (overlaps or not). We judged range overlap mostly using information from the included primary studies, but we also used other literature when the primary studies did not provide the information [74,75]. Range overlap included both sympatry and parapatry. Viability of the reciprocal hybrid cross was extracted from the included primary studies and coded as binary variable: at least partially viable or 100% inviable. Two fish and three orthoptera species-pairs were not used in meta-regression and Bayesian phylogenetic mixed models because of the lack of data for genetic distance or sex-determination system (fish, *Haplochromis burtoni* × *H. nubilus* and *Pundamilia nyererei* × *P. pundamilia*; orthoptera, *Chorthippus parallelus erythropus* × *C. parallelus parallelus*, *Gryllus armatus* × *G. rubens* and *Laupala kohalensis* × *L. paranigra*).

Effect size calculation We calculated four sets of effect sizes to investigate general patterns in hybrid male mating phenotypes and the factors underlying phenotypic novelty or novel variability expression. Effect sizes for phenotypic mean and variability (variation) were calculated in an analogous way. For each observation of each crossing experiment, phenotypic difference between parental species and hybrids in their mean and variability were calculated as the natural logarithm of the ratio between the two means (ln*RR*)[76] and between the coefficient of variation (ln*CVR*)[7], respectively. We aligned parental species according to their mean trait value prior to calculating ln*RR*, or their coefficient of variation (*CV*) before calculating ln*CVR* at each observation.

To draw general patterns of phenotypic mean in hybrids, we calculated ln*RR* of the parental species with the greater mean phenotype ($spLL_M$), each reciprocal hybrid cross ($hybLS_M$ and $hybSL_M$) and midparent value (i.e. average value between two parental species) to the other parental species ($spSS_M$) (Fig. 1b). Similarly, for phenotypic variability, we calculated ln*CVR* of the parental species with greater phenotypic variability ($spLL_V$), each reciprocal hybrid cross ($hybLS_V$ and $hybSL_V$) and

midparent value to the other parental species ($spSS_V$).

To assess if novel phenotype is expressed, we calculated ln*RR* of $spLL_M$ to both hybrids, and ln*RR* of both hybrids to $spSS_M$ at each trait observation (Fig. 1c). In the case of phenotypic variability, we calculated ln*CVR* of $spLL_V$ to each reciprocal hybrid cross, and ln*CVR* of each reciprocal hybrid cross to $spSS_V$. We calculated effect sizes for each reciprocal hybrid cross, because reciprocal hybrid crosses can differ whether they express novel phenotype or variability, and in the trait size or variability (polar overdominance)[12]. As the sign of these effect sizes should be positive when hybrid phenotypic mean or variability lies within the ranges of parental species, negative effect sizes indicate novel phenotype or variability expression. These effect sizes also indicate the relative trait size or variability compared with parents – whether mean trait size or trait variability is greater or smaller than both parents and; whether hybrids exceed mother or father species' trait size or variability. For example, if the effect size calculated from *spLL* had negative value, phenotypic mean or variability of hybrids was larger than parental species. If the effect size calculated from mother species was negative, hybrid exceeded phenotypic mean or variability of mother species.

General patterns of hybrid phenotypic mean and variability We conducted formal meta-analyses and meta-regression on the differences in phenotypic mean from hybrids to $spSS_M$ (calculated as ln*RR*) and on the differences in phenotypic variability from hybrids to $spSS_V$ (calculated as ln*CVR*). We first fitted meta-analytic model that contains only intercept to estimate the overall effect size mean (using effect sizes 1–3 in Fig. 1b). Second, we asked whether phenotypic mean and variability are greater or smaller than midparent value (the average of parental species). Using meta-regression, we compared the differences in phenotypic mean and variability from hybrids to *spSS* (effect sizes 2–3 in Fig. 1b) with those from midparent to *spSS* (effect size 4 in Fig. 1b). Midparent values in phenotypic mean and variability (*CV*) were calculated as $\frac{\overline{spLL_M}+\overline{spSS_M}}{2}$ and $\frac{CV_{spLL_V}+CV_{spSS_V}}{2}$, respectively. Hence, ln*RR* from midparent to $spSS_M$ was calculated as $\log_e \frac{\overline{spLL_M}+\overline{spSS_M}}{2} - \log_e \overline{spSS_M}$, and ln*CVR* from midparent to $spSS_V$ was calculated as $\log_e \frac{CV_{spLL_V}+CV_{spSS_V}}{2} - \log_e CV_{spSS_V} + \frac{1}{2\left(\frac{N_{spLL_V}+N_{spSS_V}}{2}-1\right)} - \frac{1}{2\left(N_{spSS_V}-1\right)}$. Third, we tested whether the reciprocal hybrid crosses differ in their mean phenotype and variability. We compared ln*RR* of $hybLS_M$ and $hybSL_M$ to $spSS_M$, and compared lnCVR of $hybLS_V$ and $hybSL_V$ to $spSS_V$ (compared effect size 2 with 3 in Fig. 1b). To compare reciprocal hybrid crosses, we restricted the dataset to the observations containing both reciprocal hybrids (sample sizes are shown in Fig. 2). 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 *spLL*. We did not include species as random effect because the effect should be accommodated by the strain random effect. A phylogenetic tree was created based on Open Tree of Life database[77] (topology of the tree: Extended Data Fig. S*2*). We searched for species names in the Open Tree Taxonomy, using the *tnrs_match_names* function in the R package *rotl* 3.0.10[78]. We computed branch lengths using the default settings of the *compute.brlen* function in the R package *ape* 5.3[79]. All meta-analytic models were implemented using the *rma.mv* function in the R package *metafor* 2.1-0[80]. All analyses were conducted in R 3.6.2[81].

Factors affecting phenotypic novelty and novel variability expression To find the factors affecting probability for F1 hybrids to express novel phenotype and variability, we constructed Bayesian phylogenetic mixed model, with the novelty status (whether phenotypic mean or variability is novel or not) as a binary response variable. We ran the analogous logistic regression model for the novel phenotype and variability expressions. As predictors, the models included (*i*) genetic divergence between parental species and (*ii*) phenotypic divergence between parental species in the focal trait, (*iii*) viability of reciprocal hybrids (binary: viable or completely inviable), (*iv*) trait type (binary: song or morphology), and (*v*) the parental species used as a reference in calculating effect sizes (binary) – *spLL* or *spSS* (the relative trait size or variability compared to parents – greater or smaller, respectively) and (*vi*) mother or father species (the relative trait size or variability compared to parents – exceeding mother or father species, respectively). As predictors, the models also included interaction terms of these species or trait level factors (*i–iv*), with parental species compared (*v* and *vi*). The main effects indicate how species- and trait-level characteristics influence the probability of novel phenotype or variability expression in any directions (*i–iv*), and whether novel phenotype or variability is biased toward certain direction (*v* and *vi*). The interaction terms indicate how species and trait level characteristics affected the relative trait size or variability compared to parents (interactions with *v*, greater trait size or variability than both parents; those with *vi*, trait size or variability exceed mother species value). As random effects, the models included trait study ID, strain ID (discriminated genetic strains or populations within a species), and standard error of each effect size. Phylogenetic effects were included as a random effect as a covariance matrix for *spLL* (Extended Data Fig. S2). In these logistic regressions, standard error of effect size was estimated as

$\frac{\pi}{\sqrt{3n}}$, where n is the summation of sample sizes of the hybrid cross and pure cross compared. From these full models, we identified the factors that significantly affected on novel phenotype and variability expression (shown in Fig. 2b and Fig. 3b, respectively).

We then employed simpler Bayesian phylogenetic mixed models to visualize significant factors' influences on the expression of novel phenotype with larger / smaller trait size, and of larger / smaller novel variability. The simpler models were separately conducted for each significant factor. As predictors, we included the focal factors (either one of *i–iv*), the compared parental species – *spLL* or *spSS* (*v*), and the interaction between them. Response variables and random effects of simper models were identical to full models. The results of the simpler models are shown in Fig. 2c–e and Fig. 3c–e for transgression in mean and variability, respectively.

To test if novel phenotypes of F1 hybrids tend to more variable than parents, we conducted ordinal Bayesian phylogenetic mixed regression. We included relative trait variability of F1 hybrids compared to parents (ordered: greater than both parents > within the range of parents > smaller than both parents) as a response, and phenotypic novelty of F1 hybrids (novel or non-novel) as a moderator variable. Random effects were identical to full models. The Bayesian analyses were conducted in R 3.6.2[81] using package *MCMCglmm* 2.29[82]. We used a weakly informative Gelman prior for the fixed effects[83], and inverse-Wishart priors for the variances of the random effects. The residual variance (overdispersion) was fixed to 1, as this cannot be estimated with binary 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.

Data and code availability

The analyses of the data were carried out with publicly available software, and all are cited in the Methods. Data, analysis code, and detailed results of analyses are available for download from **OSF address**.

Acknowledgements

K.A. thanks I. Koizumi for helpful comments, and T. Sota and Y. Takami for providing data. The project was supported by Human Resource Education & Development program of Hokkaido University and Grant-in-Aid for a JSPS Research Fellow (18J10096) for K.A., and by the Australian Research Council Discovery Project DP180100818, for S.N. and M.L..

Author contributions

K.A. and S.N. conceived the study and performed analyses. K.A. and M.L. collected data. All authors wrote the paper and approved the manuscript before submission.

Competing interests

The authors declare no competing interests.

References

1. Rieseberg, L. H., Archer, M. A. & Wayne, R. K. Transgressive segregation, adaptation and speciation. *Heredity* **83**, 363–372 (1999).
2. Stelkens, R. & Seehausen, O. Genetic distance between species predicts novel trait expression in their hybrids. *Evolution* **63**, 884–897 (2009).
3. Janick, J. Exploitation of Heterosis: Uniformity and Stability. in *The Genetics and Exploitation of Heterosis in Crops* (eds. Coors, J. G. & Pandey, S.) 319–333 (CIMMYT, 1999). doi:10.2134/1999.geneticsandexploitation.c30.
4. Edmands, S. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution* **53**, 1757 (1999).
5. Abbott, R. *et al.* Hybridization and speciation. *J. Evol. Biol.* **26**, 229–246 (2013).
6. Mallet, J. Hybrid speciation. *Nature* **446**, 279–283 (2007).
7. Nakagawa, S. *et al.* Meta-analysis of variation: Ecological and evolutionary applications and beyond. *Methods Ecol. Evol.* **6**, 143–152 (2015).
8. East, E. M. Inbreeding in corn. *Reports Connect. Agric. Exp. Stn. Years 1907–1908* 419–428 (1908).
9. Schnell, F. W. & Cockerham, C. C. Multiplicative vs. arbitrary gene action in heterosis. *Genetics* vol. 131 461–469 (1992).
10. Davenport, C. B. Degeneration, albinism and inbreeding. *Science* vol. 28 454–455 (1908).
11. Winge, Ö. & Roberts, C. Complementary action of melibiase and galactozymase on raffinose fermentation. *Nature* **177**, 383–384 (1956).
12. Cockett, N. E. *et al.* Polar overdominance at the ovine callipyge locus. *Science* **273**, 236–238 (1996).
13. Thompson, K. A., Urquhart-Cronish, M., Whitney, K. D., Rieseberg, L. H. & Schluter, D.

Patterns, predictors, and consequences of dominance in hybrids. *bioRxiv* (2019) doi:10.1101/818658.

14. Lamkey, K. R. & Edwards, J. W. Quantitative genetics of heterosis. in *The Genetics and Exploitation of Heterosis in Crops* (eds. Coors, J. G. & Pandey, S.) 31–48 (CIMMYT, 1999). doi:10.2134/1999.geneticsandexploitation.c4.
15. Wallace, B. Inter-population hybrids in *Drosophila melanogaster*. *Evolution* **9**, 302 (1955).
16. Vetukhiv, M. Viability of hybrids between local populations of *Drosophila pseudoobscura*. *Proc. Natl. Acad. Sci.* **39**, 30–34 (1953).
17. Alibert, P. & Auffray, J.-C. Genomic coadaptation, outbreeding depression and developmental instability. in *Developmental instability: causes and consequences* (ed. Polak, M.) 116–134 (Oxford University Press, 2003).
18. Karp, N. A. *et al.* Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat. Commun.* (2017) doi:10.1038/ncomms15475.
19. Poissant, J., Wilson, A. J. & Coltman, D. W. Sex-specific genetic variance and the evolution of sexual dimorphism: A systematic review of cross-sex genetic correlations. *Evolution* **64**, 97–107 (2010).
20. Fairbairn, D. J. Allometry for sexual size dimorphism: Pattern and process in the coevolution of body size in males and females. *Annual Review of Ecology and Systematics* vol. 28 659–687 (1997).
21. Haldane, J. B. S. Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* **12**, 101–109 (1922).
22. Delph, L. F. & Demuth, J. P. Haldane's rule: Genetic bases and their empirical support. *J. Hered.* **107**, 383–391 (2016).
23. Moore, M. P., Whiteman, H. H. & Martin, R. A. A mother's legacy: the strength of maternal effects in animal populations. *Ecol. Lett.* 1620–1628 (2019) doi:10.1111/ele.13351.

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

25. Ryan, M. J. & Brenowitz, E. A. The role of body size, phylogeny, and ambient noise in the evolution of bird song. *Am. Nat.* **126**, 87–100 (1985).

26. Dean, R. & Mank, J. E. The role of sex chromosomes in sexual dimorphism: Discordance between molecular and phenotypic data. *J. Evol. Biol.* **27**, 1443–1453 (2014).

27. Mank, J. E., Hall, D. W., Kirkpatrick, M. & Avise, J. C. Sex chromosomes and male ornaments: A comparative evaluation in ray-finned fishes. *Proc. R. Soc. B* **273**, 233–236 (2006).

28. Husby, A., Schielzeth, H., Forstmeier, W., Gustafsson, L. & Qvarnström, A. Sex chromosome linked genetic variance and the evolution of sexual dimorphism of quantitative traits. *Evolution* 1–11 (2012) doi:10.5061/dryad.3hq60.

29. Stelkens, R. B., Schmid, C., Selz, O. & Seehausen, O. Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. *BMC Evol. Biol.* **9**, (2009).

30. Lerner, I. M. *Genetic homeostasis*. (Oliver and Boyd, 1954).

31. Grant, P. R. & Grant, B. R. Hybridization increases population variation during adaptive radiation. *Proc. Natl. Acad. Sci.* **116**, 23216–23224 (2019).

32. Dagilis, A. J., Kirkpatrick, M. & Bolnick, D. I. The evolution of hybrid fitness during speciation. *PLoS Genet.* **15**, e1008125 (2019).

33. Hoffmann, A. A., Merilä, J. & Kristensen, T. N. Heritability and evolvability of fitness and nonfitness traits: Lessons from livestock. *Evolution* **70**, 1770–1779 (2016).

34. Kagawa, K. & Takimoto, G. Hybridization can promote adaptive radiation by means of transgressive segregation. *Ecol. Lett.* **21**, 264–274 (2018).

35. Höglund, J., Alatalo, R. V., Lundberg, A. & Rätti, O. Context-dependent effects of

tail-ornament damage on mating success in black grouse. *Behav. Ecol.* **5**, 182–187 (1994).

36. Rybak, F., Sureau, G. & Aubin, T. Functional coupling of acoustic and chemical signals in the courtship behaviour of the male *Drosophila melanogaster*. *Proc. R. Soc. B* **269**, 695–701 (2002).

37. Rosenthal, G. G. Individual mating decisions and hybridization. *J. Evol. Biol.* **26**, 252–255 (2013).

38. Ohshima, I. Genetic mechanisms preventing the fusion of ecotypes even in the face of gene flow. *Sci. Rep.* **2**, 1–6 (2012).

39. Selz, O. M., Lucek, K., Young, K. A. & Seehausen, O. Relaxed trait covariance in interspecific cichlid hybrids predicts morphological diversity in adaptive radiations. *J. Evol. Biol.* **27**, 11–24 (2014).

40. Rieseberg, L. H. & Ellstrand, N. C. What can molecular and morphological markers tell us about plant hybridization? *CRC. Crit. Rev. Plant Sci.* **12**, 213–241 (1993).

41. Matsubayashi, K. W., Ohshima, I. & Nosil, P. Ecological speciation in phytophagous insects. *Entomologia Experimentalis et Applicata* (2010) doi:10.1111/j.1570-7458.2009.00916.x.

42. Renaud, S., Alibert, P. & Auffray, J. C. Mandible shape in hybrid mice. *Naturwissenschaften* **96**, 1043–1050 (2009).

43. Parsons, K. J., Son, Y. H. & Craig Albertson, R. Hybridization promotes evolvability in African cichlids: Connections between transgressive segregation and phenotypic integration. *Evol. Biol.* **38**, 306–315 (2011).

44. Chen, C. & Pfennig, K. S. Female toads engaging in adaptive hybridization prefer high-quality heterospecifics as mates. *Science* **367**, 1377–1379 (2020).

45. Svedin, N., Wiley, C., Veen, T., Gustafsson, L. & Qvarnstrom, A. Natural and sexual selection against hybrid flycatchers. *Proc. R. Soc. B* **275**, 735–744 (2008).

46. Ritchie, M. G. The inheritance of female preference functions in a mate recognition system.

Proc. R. Soc. B **267**, 327–332 (2000).

47. Gottsberger, B. & Mayer, F. Dominance effects strengthen premating hybridization barriers between sympatric species of grasshoppers (Acrididae, Orthoptera). *J. Evol. Biol.* 921–930 (2019) doi:10.1111/jeb.13490.

48. Knief, U. *et al.* Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nat. Ecol. Evol.* **3**, 570–576 (2019).

49. Hartl, D. L. & Clark, A. G. *Principles of population genetics*. (Sinauer Associates, 2007). doi:10.2307/2531471.

50. Neher, R. A. & Shraiman, B. I. Competition between recombination and epistasis can cause a transition from allele to genotype selection. *Proc. Natl. Acad. Sci.* **106**, 6866–6871 (2009).

51. Busch, J. W. Heterosis in an isolated, effectively small, and self-fertilizing population of the flowering plant *Leavenworthia alabamica*. *Evolution* **60**, 184 (2006).

52. Reinhold, K. & Engqvist, L. The variability is in the sex chromosomes. *Evolution* **67**, 3662–3668 (2013).

53. Schilthuizen, M., Giesbers, M. C. W. G. & Beukeboom, L. W. Haldane's rule in the 21st century. *Heredity* **107**, 95–102 (2011).

54. Reinhold, K. Maternal effects and the evolution of behavioral and morphological characters: A literature review indicates the importance of extended maternal care. *J. Hered.* **93**, 400–405 (2002).

55. Arnold, M. L. & Hodges, S. a. Are natural hybrids fit or unfit relative to their parents? *Trends Ecol. Evol.* **10**, 67–71 (1995).

56. Burke, J. M. & Arnold, M. L. Genetics and the fitness of hybrids. *Annu. Rev. Genet.* **35**, 31–52 (2001).

57. Turelli, M. & Moyle, L. C. Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* **176**, 1059–1088 (2007).

58. Laurie, C. C. The weaker sex is heterogametic: 75 years of Haldane's rule. *Genetics* **147**, 937–951 (1997).

59. Pick, J. L., Nakagawa, S. & Noble, D. W. A. Reproducible, flexible and high-throughput data extraction from primary literature: The metaDigitise r package. *Methods Ecol. Evol.* **10**, 426–431 (2019).

60. Wan, X., Wang, W., Liu, J. & Tong, T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med. Res. Methodol.* **14**, 1–13 (2014).

61. Musolf, K., Meindl, S., Larsen, A. L., Kalcounis-Rueppell, M. C. & Penn, D. J. Ultrasonic vocalizations of male mice differ among species and females show assortative preferences for male calls. *PLoS One* **10**, (2015).

62. Derégnaucourt, S. Interspecific hybridization as a tool to understand vocal divergence: The example of crowing in quail (Genus *Coturnix*). *PLoS One* **5**, (2010).

63. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

64. Tamura, K., Nei, M. & Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci.* **101**, 11030–11035 (2004).

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

66. Vicoso, B. & Bachtrog, D. Numerous Transitions of Sex Chromosomes in Diptera. *PLOS Biol.* **13**, e1002078 (2015).

67. Traut, W., Sahara, K. & Marec, F. Sex Chromosomes and Sex Determination in Lepidoptera. *Sex. Dev.* **1**, 332–346 (2007).

68. Castillo, E. R., Marti, D. A. & Bidau, C. J. Sex and neo-sex chromosomes in orthoptera: A

review. *Journal of Orthoptera Research* (2010) doi:10.1665/034.019.0207.

69. Tauber, C. A. & Tauber, M. J. Inheritance of seasonal cycles in *Chrysoperla* (Insecta: Neuroptera). *Genet. Res.* **49**, 215–223 (1987).

70. Ezaz, T., Stiglec, R., Veyrunes, F. & Marshall Graves, J. A. Relationships between Vertebrate ZW and XY Sex Chromosome Systems. *Current Biology* vol. 16 (2006).

71. Miura, I. The late replication banding patterns of chromosomes are highly conserved in the genera *Rana*, *Hyla*, and *Bufo* (Amphibia: Anura). *Chromosoma* **103**, 567–574 (1995).

72. Serrano, J. & Yadav, J. S. Chromosome numbers and sex-determining mechanisms in Adephagan Coleoptera. *Coleopt. Bull.* **35**, 335-357. (1984).

73. Yoshida, K. *et al.* B chromosomes have a functional effect on female sex determination in lake victoria cichlid fishes. *PLoS Genet.* **7**, e1002203 (2011).

74. Spieth, H. T. *Drosophila heteroneura* and *Drosophila silvestris*: Head shapes, behavior and evolution. *Evolution* **35**, 921 (1981).

75. Otte, D. & Alexander, R. D. *Australian Crickets (Orthoptera: Gryllidae)*. (Academy of Natural Sciences of Philadelphia, 1983).

76. Hedges, L. V., Gurevitch, J. & Curtis, P. S. The meta-analysis of response ratios in experimental ecology. *Ecology* vol. 80 1150–1156 (1999).

77. Hinchliff, C. E. *et al.* Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci.* **112**, 12764–12769 (2015).

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

79. Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

80. Viechtbauer, W. Conducting meta-analyses in R with the metafor. *J. Stat. Softw.* **36**, 1–48 (2010).

81. R Development Core Team. R: A language and environment for statistical computing. (2019).

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

83. Gelman, A., Jakulin, A., Pittau, M. G. & Su, Y. S. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2**, 1360–1383 (2008).

84. Doherty, J. A. & Gerhardt, H. C. Acoustic communication in hybrid treefrogs: sound production by males and selective phonotaxis by females. *J. Comp. Physiol. A* **154**, 319–330 (1984).

85. Isoherranen, E., Aspi, J. & Hoikkala, A. Inheritance of species differences in female receptivity and song requirement between Drosophila virilis and D. montana. *Hereditas* **131**, 203–209 (1999).

86. Bridle, J. R., Saldamando, C. I., Koning, W. & Butlin, R. K. Assortative preferences and discrimination by females against hybrid male song in the grasshoppers Chorthippus brunneus and Chorthippus jacobsi (Orthoptera: Acrididae). *J. Evol. Biol.* **19**, 1248–1256 (2006).

87. Hochkirch, A. & Lemke, I. Asymmetric mate choice, hybridization, and hybrid fitness in two sympatric grasshopper species. *Behav. Ecol. Sociobiol.* **65**, 1637–1645 (2011).

88. Shaw, K. L. Interspecific genetics of mate recognition: Inheritance of female acoustic preference in Hawaiian crickets. *Evolution* **54**, 1303–1312 (2000).

Supplements

Supplementary Table S1 | Hypotheses for novel phenotype expression and phenotypic variability of F1 hybrids

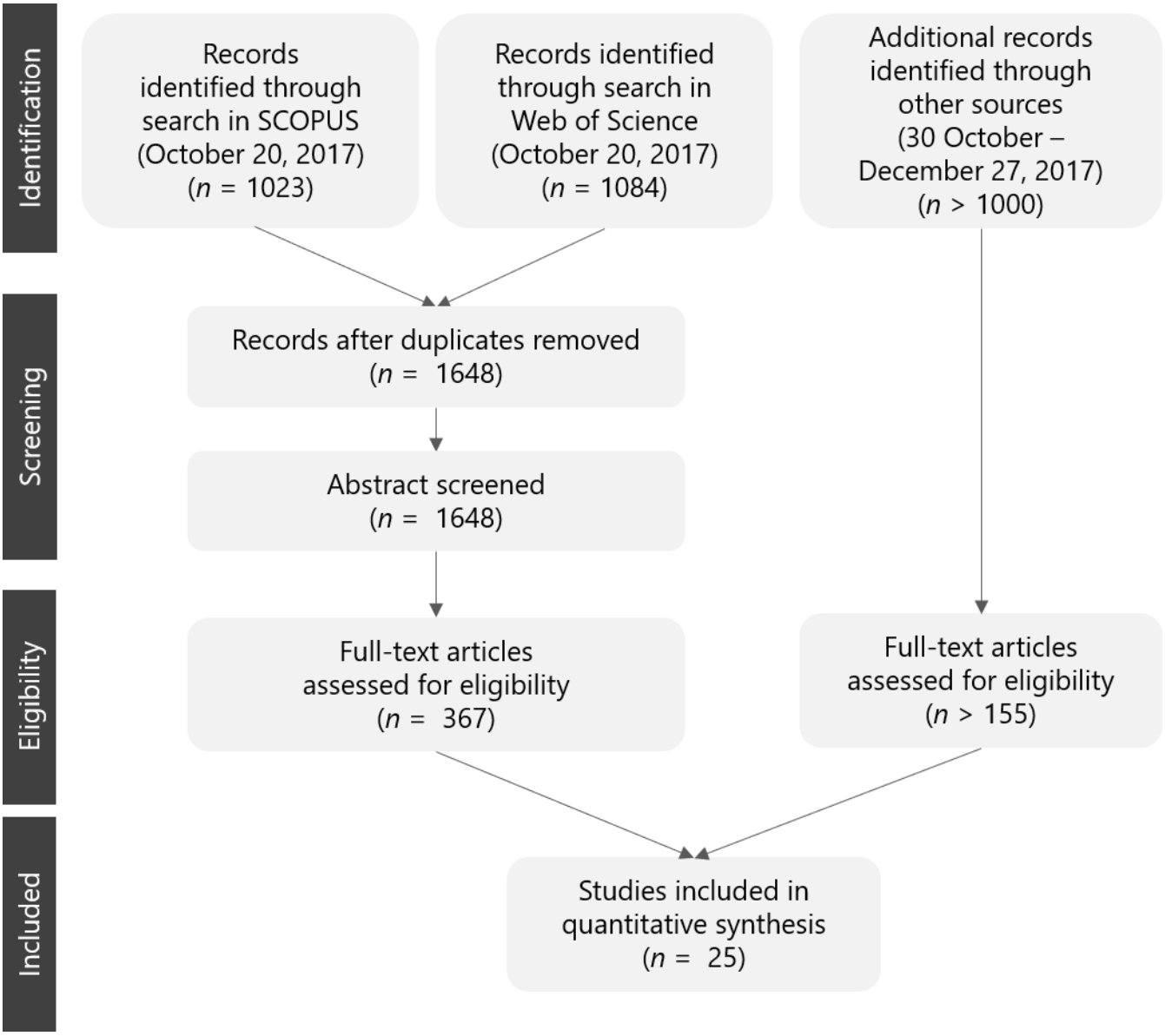
Factors	Hypothetical effect for novel phenotype expression frequency	Hypothetical effect for phenotypic variability
Genetic divergence	Increase. Positively correlate with the genetic divergence at phenotype-determining loci that facilitate novel phenotype expression through novel allelic interactions (i.e. over/underdominance, dominance, and epistasis)[1,14]	Increase. Increase variation because of increased diversity in novel allelic interactions at phenotype determining loci[4]; increase variation by damaging developmental stability[17,30]
Phenotypic divergence in focal trait	Increase. Indicates the genetic divergence at phenotype-determining loci that facilitates novel phenotype expression	Increase. Increases diversity in novel allelic interactions at phenotype determining loci
Inviability of reciprocal hybrids (i.e. genetic incompatibility)	Bias trait size. Reduces trait size by outbreeding depression	Reduce. Lethal non-additive interactions reduce genetic diversity of surviving hybrids
Distribution overlap between parental species	Increase. Reinforcement in range overlapping species pairs diverges phenotype determining loci to develop behavioral isolation	Reduce. Reinforcement reduces intraspecific variation at phenotype determining loci, and thus diminish genetic diversity of F1 hybrids at those loci
Crossing direction (parent-of-origin effect)	Genomic imprinting increases novel phenotype resembles either mother or father (i.e. polar overdominance)[12]	Not specified
Sex-determination system (male heterogamety vs. female	Increase in male heterogamety. In male heterogametic organisms, coadaptation between sex-chromosome and other chromosomes	Increase in male heterogamety. Reduced developmental stability increases phenotypic variability of heterogametic F1 males

heterogamety)	breaks down in F1 males. This reduces developmental stability of male-mating traits which increases novel phenotype expression	
Interaction between parent-of-origin effect and sex-determination system	Male heterogametic organisms tend to exhibit novel phenotype similar to father species, not mother species, due to effect of Y chromosome. Female heterogametic organisms do not show such trend	Not specified
Trait type (sound trait vs. morphology)	Increase in sound because of lower heritability compared to morphology[33]	Increase in sound because of lower heritability compared to morphology[33]

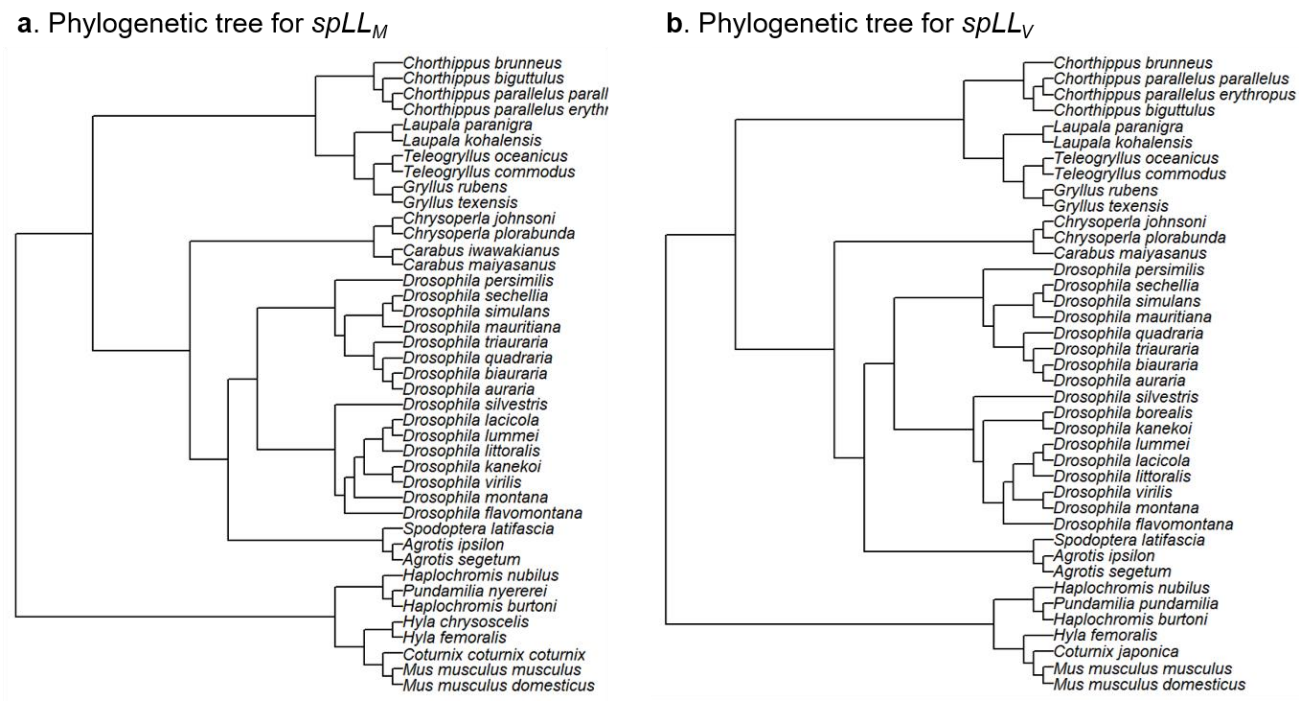
Supplementary Table S2 | Previously described female mate preference of F1 hybrids. Only 3 out of 7 studies detected additive inheritance in female mate preference during hybridization, showing that non-additivity pervades female mate preference of F1 hybrids. Dominance and parent-of-origin effect (maternal / paternal inheritance) were detected in 3 and 2 studies, respectively. Novel weak preference also appeared in 1 study. Importantly, components of female preference often varied in inheritance mode, indicating that integration of mate preference easily breaks down in F1 hybrids. This summary is based on a non-exhaustive and non-systematic review.

Taxon	Parental species	Reciprocal cross	Results	Reference
Anura: tree frog (*Hyla*)	*H. chrysoscelis* × *H. femoralis*	Viable	**Dominance**. Both reciprocal hybrids preferred hybrids over one parental species but not over the other parental species	Doherty 1984[84]
Diptera: fruit fly (*Drosophila*)	*D. virilis female* × *D. montana male*	Inviable	**Maternal / paternal inheritance**. Resembled mother in their receptivity, but resembled father in their song requirement.	Isoherranen 1999[85]
Orthoptera: bushcricket (*Ephippiger*)	*E. ephippiger* polysyllabic form × *E. ephippiger* monosyllabic form	Viable	**Additive**. Intermediate mate preferences without a large difference between reciprocal crosses	Ritchie 2000[46]
Orthoptera: grasshopper (*Chorthippus*)	*C. brunneus* × *C. jacobsi*	Viable	**Dominance and paternal inheritance**. Both reciprocal hybrids preferred one parental species over themselves. Preference function resembles that of the father.	Bridle 2006[86]

Orthoptera: grasshopper (*Chorthippus*)	*C. biguttulus* × *C. brunneus*	Viable	**Additive and dominance**. Several components of preference showed dominance, but other components showed additive inheritance. No maternal/paternal inheritance.	Gottsberger 2019[47]
Orthoptera: grasshopper (*Chorthippus*)	*C. parallelus* × *C. montanus*	Viable	**Novel preference**. Both reciprocal hybrids did not discriminate between males of two parental species	Hochkirch 2011[87]
Orthoptera: Hawaiian cricket (*Laupala*)	*L. kohalensis* × *L. paranigra*	Viable	**Additive**. Intermediate preference function, which was similar to reciprocal hybrids, resulting in preference for hybrids	Shaw 2000[88]



Extended Data Figure S1 | PRISMA diagram. The flow of inclusion and exclusion of studies identified during the literature search is shown. In brackets, we indicate the number of published literature (studies).



Extended Data Figure S2 | Phylogenetic tree used in meta-analysis. The analyses for phenotypic mean and variability used the phylogenetic tree of the parental species with larger phenotypic mean (**a**, $spLL_M$) and with larger phenotypic variability (**b**, $spLL_V$), respectively.