doi:10.1111/evo.14336

# Mapping the evolution of accurate Batesian mimicry of social wasps in hoverflies

**Alice Leavey,[1,2] Christopher H. Taylor,[1] Matthew R. E. Symonds,[3] Francis Gilbert,[1] and Tom Reader[1]**

[1]*School of Life Sciences, University of Nottingham, Nottingham NG7 2RD, United Kingdom*

[2]*E-mail: AliceLeavey@outlook.com*

[3]*Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Burwood, VIC 3125, Australia*

Received March 5, 2020
Accepted August 4, 2021

**Hoverflies (Diptera: Syrphidae) provide an excellent opportunity to study the evolution of Batesian mimicry, where defenseless prey avoid predation by evolving to resemble defended "model" species. Although some hoverflies beautifully resemble their hymenopteran models, others seem to be poor mimics or are apparently nonmimetic. The reasons for this variation are still enigmatic despite decades of research. Here, we address this issue by mapping social-wasp mimicry across the phylogeny of Holarctic hoverflies. Using the "distance transform" technique, we calculate an objective measure of the abdominal pattern similarity between 167 hoverfly species and a widespread putative model, the social wasp, *Vespula germanica*. We find that good wasp mimicry has evolved several times, and may have also been lost, leading to the presence of nonmimics deep within clades of good mimics. Body size was positively correlated with similarity to the model, supporting previous findings that smaller species are often poorer mimics. Additionally, univoltine species were less accurate wasp mimics than multivoltine and bivoltine species. Hence, variation in the accuracy of Batesian mimics may reflect variation in the opportunity for selection caused by differences in prey value or signal perception (influenced by body size) and phenology or generation time (influenced by voltinism).**

**KEY WORDS:** **Batesian mimicry, distance transform, evolution, image analysis, similarity, Syrphidae.**

Batesian mimicry, where palatable prey avoid predation by evolving features to resemble defended model species (Bates 1862), not only provides an iconic example of adaptation by natural selection, but also presents a paradox that has challenged evolutionary theory for the last 159 years (Gilbert 2005; Ruxton et al. 2018). Theory predicts that constant selection pressures imposed by predation should improve mimetic accuracy (Cuthill and Bennett 1993; Dittrigh et al. 1993; Edmunds 2000; Gilbert 2005; Rotheray and Gilbert 2011). However, mimicry is frequently far from perfect (Speed and Ruxton 2010; Edmund and Reader 2014; Taylor et al. 2016a). Attempts to comprehend the existence of imperfect mimicry have produced an extensive series of hypotheses (see McLean et al. 2019 for a review). Although some of these hypotheses are now regarded as implausible, great uncertainty remains over which factors are most important in the persistence of imperfect mimicry.

One of the best-known systems for the study of imperfect mimicry is provided by hoverfly mimics (Diptera: Syrphidae), which are probably defenseless, and their harmful hymenopteran models. Many hoverflies imitate Hymenoptera behaviorally (Golding et al. 2005; Penney et al. 2014), acoustically (Moore and Hassall 2016), and morphologically, in the form of color, pattern, shape, and size (Howarth et al. 2004; Penney et al. 2012; Taylor et al. 2017). However, many supposedly mimetic hoverflies do not accurately resemble their putative models, and others are apparently not mimetic at all. The hoverfly clade therefore provides an ideal opportunity to study how mimetic accuracy has evolved.

The study of Batesian mimicry is often hampered by difficulties in defining and quantifying mimicry. Hoverflies have typically been classified as Batesian mimics based on behavioral studies using putative or model predators under controlled

conditions, or entirely subjectively, and often inconsistently, by humans (Taylor et al. 2013; Edmunds and Reader 2014). Even attempts to quantify mimicry more objectively have relied on somewhat ad hoc selections of variables or landmarks, often using features that will be perceived very differently depending on the signal receiver (e.g., RGB color values) (Dittrigh et al. 1993; Azmeh et al. 1998; Holloway et al. 2002; Penney et al. 2012). Consequently, our understanding of variation in the accuracy of mimicry among hoverfly species may be at odds with the perception of real predators in the wild. Furthermore, the mimetic status of many hoverflies, especially those that are not conspicuous to the human eye, remains completely unknown.

Correlations between mimicry and life history traits can provide important insights into the factors that have driven the evolution of mimicry. For instance, we might expect mimicry to be related to body size because larger species are more conspicuous to predators, or more valuable prey, whereas smaller species may benefit more from other antipredation strategies such as crypsis (Holen and Johnstone 2004). Wilson et al. () found that body size does not correlate strongly with mimetic fidelity in hoverflies, but they did not account for phylogeny (and hence shared evolutionary history) in their analysis. By contrast, a phylogenetically controlled analysis suggested that large hoverfly species are indeed better mimics (Penney et al. 2012). However, neither of these studies explicitly considered hoverflies that are thought to be nonmimics. Studies examining mimicry in coral snakes have found that good Batesian mimicry could gradually evolve from nonmimetic ancestral species, and that maladaptive mimetic patterns can break down, resulting in poor mimics being deeply nested in a clade of good mimics (Kikuchi and Pfennig 2010; Hodson and Lehtinen 2017). However, life history traits that could be associated with the evolution of mimicry, such as diet or body size, were not considered in these analyses. Additionally, the relative abundance and phenology of mimics and models can impact the selection pressure for good mimicry, factors that are likely to be influenced in insects by voltinism, which can vary substantially among species (Howarth and Edmunds 2000; Finkbeiner et al. 2018; Hassall et al. 2019). Only by analyzing life history traits and phylogenetic history together can we make clear inferences about the evolvability of mimetic accuracy, but this has yet to be attempted for any large taxonomic group, such as the Syrphidae (Gilbert 2005; Rotheray and Gilbert 2011).

In this study, we build on previous attempts to quantify variability in visual mimetic accuracy among hoverfly species, and to identify the possible drivers of that variability, with a comprehensive phylogenetically controlled analysis of hoverfly abdominal patterns, features which are detectable by almost any visual system. The key questions we address are: (i) how has the accuracy of wasp mimicry evolved across the hoverfly phylogeny? and (ii) what predicts the evolution of high fidelity in wasp mimics? We use a "distance transform" method for image analysis (Taylor et al. 2013) to quantify the similarity of Holarctic hoverflies from 108 genera to the common and widespread social wasp model, *Vespula germanica*. The distance transform approach allows rapid semiautomated evaluation of mimetic accuracy across large numbers of taxa, which can easily be re-run with different subsets of data, model taxa, and so on. We focus on wasp mimicry because it is the most widespread form of mimicry in hoverflies, likely to be homologous across species, and most easily quantified using our objective image analysis. Having verified that our measure of similarity correlates well with existing measures and similarity scores for two additional social wasp models, we then plot pattern similarity onto the hoverfly phylogeny, and test for associations with key life history traits. For the first time in a study of this kind, we include hoverflies that are not considered to be mimics, so that we can identify the positions in the phylogeny where wasp mimicry first evolved.

# Methods

## HYMENOPTERAN MODEL SELECTION

We chose to study mimicry of the German wasp (*V. germanica*), a widespread and abundant noxious social wasp considered to be a model for many hoverfly mimics in the Holarctic region (Gilbert 2005). *Vespula germanica* is very similar in appearance to other *Vespula* species (Table S1; see SENSITIVITY TESTS section), which are also likely models for hoverfly mimicry, but *V. germanica* is the most widely distributed and the most common species in the genus (CABI 2019). Our specific objective was to study the evolution of social wasp mimicry alone, rather than all forms of Batesian mimicry in hoverflies. Where we find a hoverfly species is a poor wasp mimic, or a nonmimic relative to wasps, this could be because it is entirely nonmimetic, but it could also be because it is a conspicuous mimic of another defended model. Other relevant putative models for hoverfly mimics include honeybees (*Apis mellifera*) and bumblebees (*Bombus* spp.).

## IMAGE SELECTION

We used images of hoverfly abdomens to characterize mimetic accuracy. Logistical constraints, including a shortage of high-quality images, meant we could not sample all Holarctic hoverfly species. As the species in most hoverfly genera/subgenera have similar color patterns, we chose a single representative species from each for analysis (see the Supporting Information). If many species looked similar to the human eye, the one with a distribution that most widely overlapped with that of *V. germanica* was included. Where species had similar distributions, the most abundant species (according to expert opinion, see below) was included. Some genera/subgenera (25 out of 108) contained

several widely distributed, abundant species with conspicuously different abdominal patterns. In these cases, we included multiple representative species, one for each obvious type of pattern, except where good-quality images were unavailable. Thus, the taxonomic units used here are color-pattern groups usually corresponding to genera or subgenera, but occasionally to species groups within them (Table S2): we use the term "operational taxonomic unit" (OTU) to denote these groups. For the full list of species used, see the dataset in the Supporting Information.

Hoverfly and wasp images were sourced primarily from reliable internet sites run by taxonomic experts where species identification was judged to be accurate by the research community (see the Supporting Information). Multiple images were sourced from Taylor et al. (2017) and Speight and de Courcy Williams (2018). Images were selected following a hierarchy of rules for quality, sexual dimorphism, and intraspecific variation. To meet the criteria for quality, the images were of alive or recently dead specimens to avoid color fading, except *Chrysosyrphus nasuta* that, due to a lack of good images, was from an artist's drawing. The images we used had variable backgrounds, depending on how the image was acquired, so we ran Wilcoxon test comparing mimetic accuracy between images from natural and artificial backgrounds to ensure our results were not impacted by the image sources.

The abdomen was used for analysis because the color pattern is, in general, much more distinctive and variable on the abdomen than on the thorax in dipterans and hymenopterans (Marchini et al. 2017), and the abdomen is typically conspicuous to potential predators. Studies have previously shown that abdominal color patterns of both hoverflies and wasps consist of clearly delineated contrasts in both achromatic and chromatic dimensions, and do not contain hidden ultraviolet signals (Taylor et al. 2016b), meaning that the spatial elements of the pattern are visible to all but the most primitive of visual systems.

Images were only used where they showed a clear dorsal view of the abdomen, without obvious distortion of the pattern. Images with glare, reflections, and obstructions from pollen or wings were rejected unless no alternative was available. Where the best image included minor examples of such imperfections, these were corrected by eye in the image preprocessing stage using ImageJ (Abràmoff et al. 2004), for example, by exploiting symmetry of the pattern to fill in obscured areas. It is important to note that, because we relied on photographs in the public domain, the selection of images we used was probably not entirely representative of natural inter- and intraspecific variation. Photographs of larger, more brightly colored species or individuals, and those with striking patterns, are probably more likely to be made available in the sources we used, because they are easier to photograph, or more interesting or detectable to photographers and entomologists.

Images of males were used by default, except where images of females were of significantly higher quality. Most of the selected species were not conspicuously sexually dimorphic. There were four instances where females had to be chosen despite the presence of conspicuous sexual dimorphism, defined as a distinct difference in pattern markings not simply due to differences in the shape or size of the abdomen: *Baccha elongata*, *Hiatomyia willistoni*, *Mixogaster breviventris*, and *Nausigaster punctulata*. Some multivoltine hoverflies, especially *Eristalis* spp., exhibit phenotypic variation in color pattern due to seasonal variation, so an image of the most commonly recorded pattern was selected for analysis (Holloway et al. 1997). *Merodon equestris*, a bumblebee mimic, was not included because it has widely variable and distinct color morphs (Mengual et al. 2006).

## PHYLOGENY RECONSTRUCTION

Recently, much progress has been made in our understanding of hoverfly phylogeny at the genus level (Mengual et al. 2018; Pauli et al. 2018; Moran and Skevington 2019; Moran et al. 2021), but its overall architecture remains little changed from the study of Rotheray and Gilbert (1999) as modified by Ståhls et al. (2003). We used a phylogeny based on morphological data from Katzourakis et al. (2001), excluding non-Holarctic genera and a few that lack good-quality images. This phylogeny is in turn based on Rotheray and Gilbert's (1999, 2008) cladistic study of larval characters in Palaearctic genera, and is very similar to recent skeleton trees based on transcriptomics (Pauli et al. 2018) and anchored enrichment genetic data (Young et al. 2016). A comprehensive phylogeny from anchored enrichment data is currently being constructed, but is still a long way from publication (JH Skevington, pers. comm.).

The Katzourakis et al. (2001) tree was updated using more recent molecular phylogenies of restricted subgroupings and 17 extra OTUs were added; if no data on their placement were available, the relationship was left as a polytomy (see Table S2). Our semi-resolved, literature-based tree was formed using Mesquite (Version 3.6, Maddison and Maddison 2018). In the absence of a comprehensive resolved phylogeny, combining published trees is often better than, for example, estimating the phylogeny using proxies from DNA sequences in GenBank (Beaulieu et al. 2012) and leaving parts unresolved where molecular data are not available. Phylogenies that covered most of the species used in this study took precedence over less densely sampled studies. Trees extrapolated from model-based approaches, such as Bayesian and maximum likelihood, took priority over those inferred from distance-based methods or parsimony (Beaulieu et al. 2012). These published data were used to resolve as much of the tree as possible to create a "master tree," which was then imported into R version 3.5.2 (R Core Team 2018) for analysis using the packages *ape* (Paradis and Schliep 2019) and *geiger*

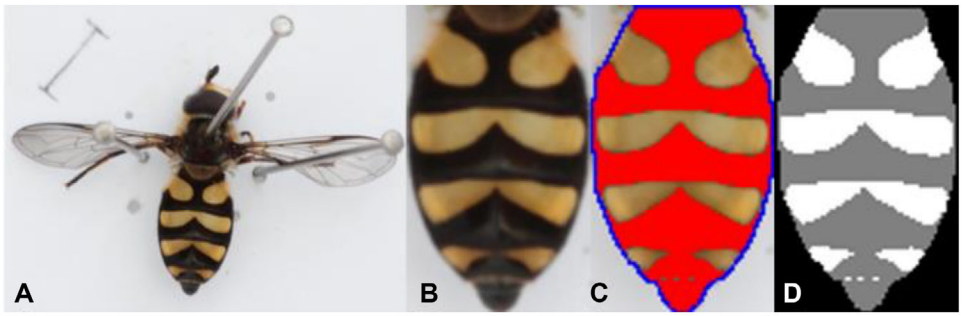


**Figure 1.** The stages of image preprocessing. (A) Original image of *Didea fasciata*. (B) After rotation, cropping, and scaling. (C) Abdomen outlined in blue and black areas masked with red using ImageJ. (D) Final binary image from MATLAB.

(Harmon et al. 2007). Branch lengths were calculated using the "Grafen" algorithm, where the depth of nodes is equal to the number of daughter species descend from that node (Grafen 1989), and polytomies were made dichotomous (with zero length) using the "compute.brlen" and "multi2di" functions in the *picante* package (Kembel et al. 2010). The final tree was constructed and visualized using *RColorBrewer* (Neuwirth and Neuwirth 2011) and the "contMap" function in *phytools* (Revell 2012).

## IMAGE PREPARATION

Following image selection, three wasp species and a total of 167 OTUs within 108 genera of Holarctic hoverflies were selected for processing and analysis (see the dataset in the Supporting Information). Image preprocessing was carried out in ImageJ. First, images were rotated so that the top of the scutellum was horizontal, with the tip of the abdomen facing downward. Images were cropped to the smallest area containing the abdomen, from the tip of the abdomen to where the scutellum meets the two sides (Taylor et al. 2013). Without changing the aspect ratio, each image was scaled to the height of 100 pixels to standardize abdomen size and the abdomen outlined in blue (Fig. 1). In all cases, we were able to identify two distinct colors in the abdominal pattern: a pale color (typically yellow, white, or orange) and a dark background color (typically black or dark brown). Images were "segmented" based on their light and dark components using color thresholding and paintbrush tools. Although in most cases, the color pattern was formed by pigmentation of the tergites, colored hairs sometimes played a role. The hairs outside the true outline of the abdomen were only included if they were dense enough to (1) obscure the true outline or (2) form a border just as strong as the true outline. Hairs within the outline of the abdomen were only included if they would be conspicuous regardless of the strength or direction of any light. All 167 images were preprocessed, saved as TIFF files, and converted into a binary format using MATLAB (Fig. 1; Taylor et al. 2013; Mathworks 2018).

## SIMILARITY CALCULATION

A matrix of dissimilarity values was produced in MATLAB according to the methods in Taylor et al. (2013). To avoid misalignment and optimize the dissimilarity value, the "optim" parameter was set to "hy" and the "scal" parameter was set to "y." This shifted each image vertically to minimize mismatch between segments while keeping the height and aspect ratio the same (Taylor et al. 2013). To ease interpretation, results were scaled based on the highest number in the matrix, converted to similarity values, and squared. Henceforth, these values are referred to as "distance transform similarity scores." Images from nonmimetic species with entirely black abdomens were assigned the similarity value of zero. The ancestral estimates for similarity were calculated using the "fastAnc" function from *phytools*, which assumes a Brownian model of evolution (Revell 2012).

## OTHER MEASURES OF MIMETIC FIDELITY

We used classifications of mimicry from several sources to calibrate the measure of mimetic accuracy from our image analysis, and to establish a formal method for categorizing an OTU as a mimic. The calibration allowed us to determine whether our similarity measure actually predicts the behavior of representative vertebrates (humans and birds) when faced with a visual discrimination task similar to that required to identify models and mimics in real populations. First, we collected expert evaluations of mimetic accuracy from the literature (Gilbert, unpubl. data collated over the past 40 years from about 10,000 syrphid publications). Three categories were recognized: any OTU identified as a social wasp mimic was labeled either "good" or "poor," based on the expert descriptions given, whereas it was considered a "nonmimic" when there was no source to say otherwise.

Next, we gathered independent estimates of mimetic accuracy for a subset of overlapping OTUs from published studies of pigeon (Dittrigh et al. 1993) and human (Penney et al. 2012) evaluations of hoverfly images. To increase coverage to all 167 OTUs in our dataset, we also designed our own survey using human volunteers. In contrast to the published studies mentioned

above, which evaluated full-color images of the whole hoverfly, we surveyed perceptions of wasp mimicry in the binary images of abdomens created for the distance transform analysis. This permitted direct comparison of human perception of mimetic accuracy and distance transform similarity scores, based on the same characters. Nonexpert volunteers were recruited from a student population and were asked to compare the abdomen patterns of *V. germanica* and each of 30 hoverfly OTUs, randomly selected without replacement from the pool of 167 images. Volunteers rated the similarity of the pair of images from 1 (hoverfly is not mimetic) to 10 (perfect mimicry). Each pair of images was displayed via a website on the volunteer's computer screen until they decided on a rating and clicked the button. Overall, the survey was completed 98 times, and each image was assessed a minimum of eight and a maximum of 29 times.

## ANALYSES

All statistical analyses were conducted in R, version 3.5.2 (R Core Team 2018). Phylogenetic generalized least squares (PGLS) analyses were performed using the *caper* package to investigate the relationship between pattern similarity and ecological characteristics while correcting for phylogenetic effects (Orme et al. 2018). These traits included larval feeding ecology, voltinism, phenology (mostly from Speight 2018), and, as a proxy for body size, wing length (taken from Gilbert, unpubl. data [see above]; Stubbs and Falk 2002). The key flight periods were defined as "early" (March to May), "mid" (May to July), and "late" (July to September), based on quantitative data (primarily the Hoverfly Recording Scheme [www.hoverfly.org.uk], with gaps filled from Gilbert, unpubl. data [see above]). The PGLS approach considered the absence of phylogenetic independence between these traits by incorporating a covariance matrix between species into the model. Phylogenetic signal in the model was measured using a maximum likelihood estimation of the parameter lambda (Pagel 1999), which varies from 0 (phylogenetic independence of residuals) to 1 (strong association of residuals with phylogeny under the Brownian motion model of evolution). We estimated the degree of phylogenetic signal in the individual traits measuring mimicry (both the distance transform scores, and the human evaluation scores), by fitting intercept-only models predicting both traits.

PGLS analyses were performed using all ecological traits as explanatory variables, using similarity scores from the distance transform analysis (one for each wasp model) and our survey as separate response variables. Typically, it is not necessary to carry out nonphylogenetically controlled analyses in addition to PGLS (Freckleton 2009), but because there is some uncertainty over the phylogeny used, we also modeled the data using ordinary least squares (OLS) regression. Models with the best fit were identified using stepwise model selection based on Akaike's information criterion (AIC). This involved starting with the full model containing the complete set of predictors, then sequentially removing the least significant variable one at a time to find which model had the lowest AIC value.

To explore the impact of considering social wasp mimicry as a discrete as opposed to a continuous trait, we inspected the distribution of our *V. germanica* distance transform similarity scores for each category from the literature and identified a threshold score below which there are no recognized mimics (Fig. 2). We used this threshold to create a variable for mimicry as a binary trait (1/0). As a large number of hoverflies above this threshold were classified as nonmimics in the literature, we ran a second binary analysis where the threshold was defined by the point above which the number hoverflies classified as mimics by the literature exceeded the number of nonmimics. We also evaluated binary mimicry using the raw data from the literature evaluation, where "good" and "poor" mimics were grouped together under "mimics" and compared with OTUs for which no mimicry was reported. These three definitions of binary mimicry are subsequently referred to as "the mimicry threshold," "the majority threshold," and "the literature categories," respectively. For each definition of binary mimicry, a phylogenetic logistic regression was performed using the "phyloglm" function in *phyloglm*, which uses alpha ($\alpha$) to represent the strength of the phylogenetic signal (Ives and Garland 2009). A low alpha value denotes a strong association between phylogenetic structure and trait presence. Models in the phyloglm analysis were compared using AIC.

## SENSITIVITY TESTS

We ran a supplementary analysis using two additional social wasp models, *Vespula vulgaris* (the second most common member of the genus) and *Polistes dominula* (another widespread and common social wasp), to establish how sensitive our findings were to the choice of model taxon.

Our approach to image analysis is less effective where aposematic and mimetic patterns on the abdomen rely on colored hairs, as is the case with bees and some of their mimics, because the abdominal patterns of hairy species do not have uniform patches of color. In the distance transform algorithm, this leads to abnormally high similarity values when compared to a wide range of possible patterns, because the distances between matching pixels are small. Hence, we were unable to extend our analysis to include bee mimicry. For some hairy species, the distance transform measure of mimetic accuracy did not correspond well with evaluations of wasp mimicry made by volunteers or the literature (see OUR SURVEY section). We therefore explored the impact of the inclusion of hairy species in the dataset by classifying each species as hairy (with conspicuous hairs on the abdomen, $n = 32$) or not hairy ($n = 135$), and including this as a factor in the analysis of the relationship between the distance

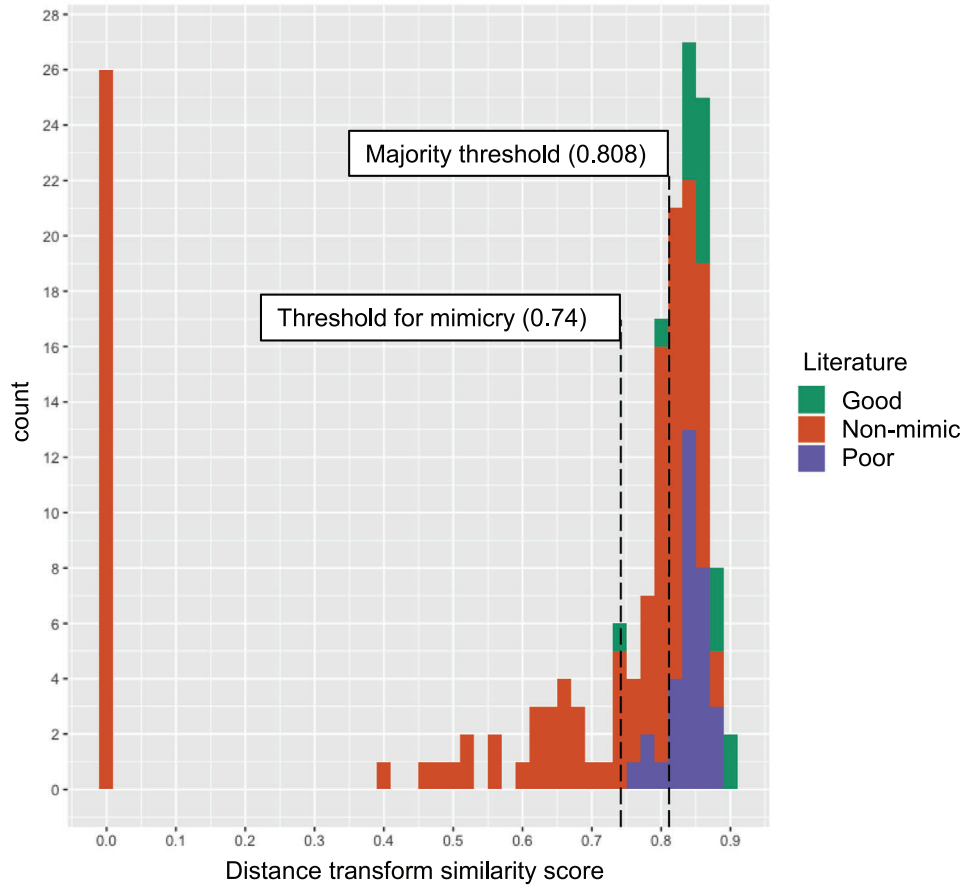


**Figure 2.** **Frequency distribution of similarity scores describing the accuracy of social wasp mimicry in 167 species of hoverfly, color coded according to categories identified from the literature. The threshold for mimicry divides possible mimics from species that have never been considered to be mimics by experts, whereas the majority threshold marks the point above which most species are considered mimics. Bin width = 0.02.**

transform score and similarity to *V. germanica* as perceived by our volunteers. We also ran a supplementary phylogenetic analysis for *V. germanica* distance transform similarity scores without the hairy species included.

We were concerned about the influence of sampling bias in the estimation of phylogenetic signal in our main analysis, caused by the repeated sampling of some genera in which phenotypes varied conspicuously among species (see above). We therefore conducted a second analysis with a reduced version of our *V. germanica* dataset. We repeated the PGLS and *phyloglm* binary analysis 1000 times with just one randomly selected species from each genus in which we had data for multiple species, and generated Higher Posterior Density (HPD) confidence intervals for the model coefficients averaged across all 1000 trees.

# Results

## QUANTIFYING MIMETIC SIMILARITY

Abdominal pattern similarity of hoverflies to *V. germanica* was widely distributed (Fig. S1). The distance transform analysis identified the three best *V. germanica* mimics as *Spilomyia interrupta*, *Caliprobola speciosa*, and *Helophilus pendulus* (Fig. S2). Aside from the all-black species, the three lowest similarity scores were obtained from *Hadromyia grandis*, *Pyrophaena rosarum*, and *Volucella pellucens* (Fig. S2). This result was the same in our analysis excluding species with hairy abdomens, but the choice of model taxon had some impact on the ranking of the mimics (Table 1; Fig. S2). Nevertheless, the similarity scores in relation to *V. vulgaris* (Spearman's rank: $r_s = 0.83$, $P < 0.001$) and *P. dominula* (Spearman's rank: $r_s = 0.78$, $P \leq 0.001$) were strongly and significantly correlated with those for *V. germanica* (Fig. S3). The similarity scores of every hoverfly species in relation to all three wasp models are provided in the dataset in the Supporting Information. The image background, and therefore the image source, did not impact the similarity score (Wilcoxon rank sum test: $W = 2729.5$, $P = 0.17$). Inspection of the distribution of distance transform similarity scores for species classified as mimics in the literature suggested a threshold of 0.74, below which hoverflies are never considered to be social wasp mimics

**Table 1.** **A summary of the major conclusions obtained from the main *Vespula germanica* analysis, and whether they were supported by our different supplementary analyses and sensitivity tests. Each box refers to evidence either in support of (in bold) or in contrast to each conclusion. "NA" means this conclusion was not tested in this analysis. *Not including species with all-black abdomens.**

| | Supplementary analyses | | | Sensitivity tests | | |
|---|---|---|---|---|---|---|
| Conclusion | Human survey | Binary analysis—mimicry threshold | Binary analysis—majority threshold | Wasp model type | Hairiness | PGLS with one species per genus |
| Identity of the top three and bottom three mimic taxa* | Dataset in the Supporting Information | NA | NA | Figure S2 | **Figure S2** | NA |
| Location of threshold to divide mimics and nonmimics | NA | NA—threshold used in analysis | Figure 2 | Figure S13 | **Figure S4** | NA |
| PGLS was a better fit than OLS | **Table S3** | NA | NA | **Table S3** | **Table S3** | NA |
| Mimicry has evolved many times | NA | **Figure 3** | **Figure S11** | **Figure S13** | **Figure S12** | NA |
| Mimicry is sometimes lost in clades of good mimics | NA | **Figure 3** | **Figure S11** | **Figure S13** | **Figure S12** | NA |
| Phylogenetic signal for wasp mimicry is significant but not strong | **THE EVOLUTION OF MIMICRY section** | **Table S6** | **Table S6** | **Figure S13** | **Figure S12** | Table S6—weak signal |
| The best predictors of mimetic accuracy were wing length and voltinism | **Table S3** | **Table S5** | Table S5—only wing length | **Table S3** | **Table S3** | **Table S5** |
| Smaller species are significantly less mimetic than larger species | **Table 2** | **Table S6** | **Table S6** | **Table S4** | **Table S4** | **Table S6** |
| Univoltine species are significantly less mimetic than multivoltine species | **Table 2** | Table S6 | NA | **Table S4** | **Table S4** | **Table S6** |

(Fig. 2). This threshold was the same when species with hairy abdomens were excluded (Fig. S4). The threshold above which the majority of species were considered mimics by the literature was 0.808 (number of mimics above threshold = 45; number of nonmimics above threshold = 42; Fig. 2). These two thresholds were used to divide mimics from nonmimics for subsequent analyses.

### DISTANCE TRANSFORM AND PREVIOUS STUDIES

Our distance transform similarity scores for hoverflies differed significantly across descriptive categories found in the literature (see OTHER MEASURES OF MIMETIC FIDELITY section above), with "nonmimics" having the lowest similarity to *V. germanica* (Kruskal-Wallis: Chi-squared = 52.83, df = 2, $P < 0.001$). Although the difference between "good" and "poor" mimics was not significant (Dunn's test: $z = 1.07$, $P = 0.14$), "good" mimics were marginally more similar on average (Fig. S5). The results when hairy species were excluded were qualitatively similar (Fig. S6). Distance transform similarity scores were significantly positively correlated with similarity analyses from published studies of pigeon (Spearman's rank: $r_s = 0.73$, $P = 0.02$; Fig. S7A; Dittrigh et al. 1993) and human perception ($r_s = 0.74$, $P = 0.0002$; Fig. S7B; Penney et al. 2012).

### OUR SURVEY

Volunteer perception of wasp mimicry in binary images of hoverfly abdomens in our survey varied significantly between mimics

and "nonmimics," as defined by the literature (Kruskal-Wallis: Chi-squared = 57.89, df = 2, $P < 0.001$), but not between "good" and "poor" mimics (Dunn's test: $z = 0.84$, $P = 0.20$; Fig. S8). The average perceived similarity in our survey was also positively correlated with survey ratings from Penney et al. (2012) (Pearson's correlation coefficient: $r = 0.86$, $P < 0.001$; Fig. S9). The ranking of distance transform similarity scores was also negatively correlated with the survey results ($r_s = -0.75$, $P < 0.001$; Fig. S10)—species with a higher similarity score in the distance transform analysis were typically perceived to be more similar to *V. germanica* in our survey. Many of the species with hairy abdomens appeared to be outliers, with a low survey score but relatively high distance transform similarity ranking (Fig. S10). A two-way ANOVA indicated that hairy species have significantly higher distance transform similarity ranks overall ($F_{(1,164)} = 281.63$; $P < 0.001$), and their relationship with survey score is weaker, although not significantly so ($F_{(1,163)} = 3.266$; $P = 0.073$). The results of subsequent sensitivity tests where species with hairy abdomens were excluded from the *V. germanica* dataset are summarized in Table 1.

### THE EVOLUTION OF MIMICRY

Social wasp mimicry, as revealed by distance transform analysis of hoverfly abdominal patterns, was patchily distributed over the phylogeny (Fig. 3). When we defined species as mimics or nonmimics by calibrating similarity scores using the literature (see above), transitions between states of mimicry appear to have happened repeatedly, both from nonmimetic to mimetic and vice versa. According to ancestral state estimations using our mimicry threshold of 0.74, *V. germanica* mimicry has evolved 35 times, 13 of these being at ancestral nodes (47 and 16 times, respectively, using the majority threshold of 0.808) and there were seven instances (12 using the majority threshold, three of these being at ancestral nodes; Fig. S11) where nonmimics were found deep within a clade of mimics (Fig. 3). When binary mimicry was defined by the literature evaluation, mimicry evolved 28 times, nine of these being at shared ancestral nodes (Fig. 3). The Pipizinae were all nonmimics, whereas Eristalinae and Syrphinae contained species that were quite variable in their mimetic accuracy. Microdontinae, the earliest evolving subfamily, had high similarity results and therefore the two species we examined were considered to be accurate mimics of *V. germanica*. The pattern of repeated evolution of mimicry was broadly similar, regardless of the choice of wasp model (Fig. S13).

The phylogenetic signal associated with the distance transform similarities to *V. germanica* was significantly different from zero, but not strong, because the observed value was also significantly different from one ($\lambda = 0.63$, 95% CI = 0.59–0.81, $P$ ($\lambda = 0$) $< 0.001$, $P$ ($\lambda = 1$) $< 0.001$). The same was true for both *V. vulgaris* and *P. dominula* (Fig. S13), but the phylogenetic signal was slightly weaker in analyses of the *V. germanica* similarity survey ($\lambda = 0.43$, 95% CI = 0.15–0.71, $P$ ($\lambda = 0$) $< 0.001$, $P$ ($\lambda = 1$) $< 0.001$) and sensitivity tests (Table 1).

### LIFE HISTORY CORRELATES OF MIMICRY

The fit of the PGLS models was better than equivalent OLS models for all three wasp species, which establishes that the evolution of mimicry is constrained by phylogeny (see Table S3). The best statistical models for the distance transform scores for each wasp and the survey similarity scores for *V. germanica* all revealed that the most significant variables explaining mimetic similarity were wing length and voltinism (Tables 1 and 2). Smaller species were significantly less mimetic than larger species (Fig. S14) and univoltine species were significantly worse mimics than multivoltine species, with bivoltine somewhere between the two (Fig. S15). There were no noticeable relationships between mimicry and larval feeding ecology (Fig. S16). Species that emerge later in the year were typically slightly better mimics, but this effect of phenology was not significant (Fig. S16). Our analysis of *V. germanica* mimicry as a binary trait showed qualitatively similar results, with the results varying to some extent depending on which species were selected when reanalyzing the data excluding all but one species per genus (Tables S6 and S7).

## Discussion

Our study provides the first systematic and quantitative description of the repeated evolution of social wasp mimicry across the entire Holarctic hoverfly family. Distance transform analysis of abdominal patterns provides a measure of mimetic accuracy that can be applied to large numbers of taxa simultaneously and is not tied to a particular visual system. Our results show that this measure strongly corroborates other assessments of mimetic accuracy from expert and nonexpert humans and birds, and extends our understanding of variation in abdominal patterns to species for which wasp mimicry has not previously been evaluated, or has been considered to be absent. We found that accurate wasp mimicry has probably evolved repeatedly in hoverflies, and may also have been lost. We also found that mimetic accuracy is predicted by life history: it correlates positively with a proxy for body size, and is associated with voltinism. This implies that hoverfly ecology influences the tendency for species to evolve wasp mimicry (or indeed the reverse), giving us an insight into origins of the tremendous variation in morphology we see across the family.

Our results suggest social wasp mimicry has evolved repeatedly at scattered positions throughout the phylogeny, regardless of which threshold we use to distinguish between mimics and nonmimics. The phylogenetic signal for wasp similarity was significant but not strong, suggesting some relationship between

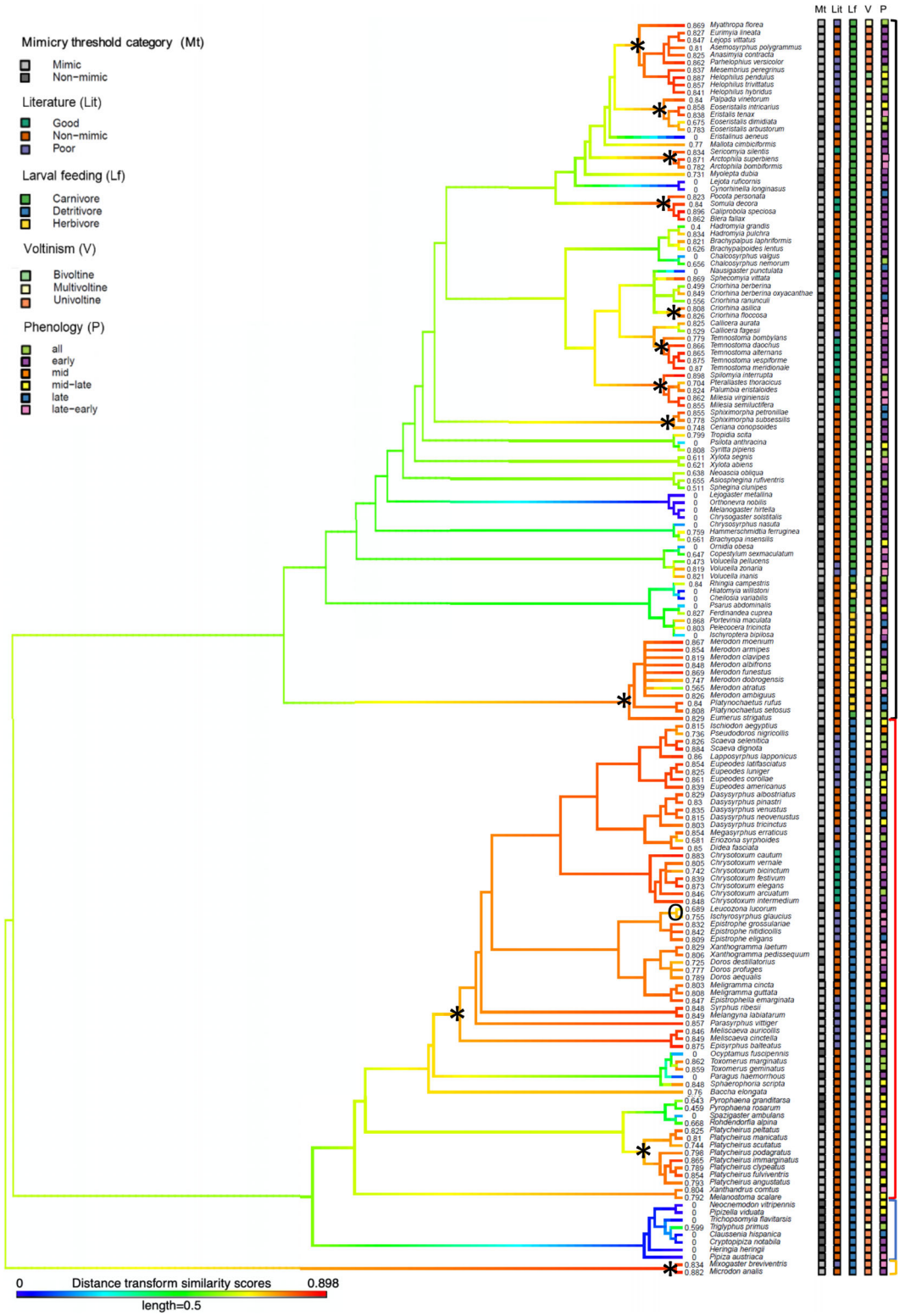


**Figure 3.** A literature-derived phylogeny of 167 hoverfly species. Warmer tip colors represent higher similarity to, and hence better mimicry of, the social wasp *V. germanica*. Tips are labeled with the distance transform similarity scores and a color-coded grid to represent the ecological traits investigated (for abbreviations, see key). Defining mimicry as a binary trait using the mimicry threshold (0.74) allowed us to identify ancestral nodes where social wasp mimicry evolved (*) and was lost (O) according to "fastAnc" ancestral state estimates under Brownian evolution. Blank nodes before an * are nonmimetic. Subfamilies (indicated by the brackets on the far right): black = Eristalinae; red = Syrphinae; blue = Pipizinae; yellow = Microdontinae (Chandler 1998; Stubbs and Falk 2002).

**Table 2. Coefficients from the best PGLS models describing the relationship between life history traits and mimetic similarity scores for 167 species of hoverfly for *Vespula germanica*. Similarity scores were either calculated by pattern analysis ("distance transform") or from a survey of human volunteers ("survey"). SEM = standard error.**

| | | | Coefficients | SEM | *t*-value | *P*-value |
|---|---|---|---|---|---|---|
| Distance transform | Intercept (Univoltine) | | 0.284 | 0.159 | 1.783 | <0.001 |
| | Wing length | | 0.041 | 0.010 | 4.074 | <0.001 |
| | Voltinism | Bivoltine | 0.116 | 0.051 | 2.293 | 0.023 |
| | | Multivoltine | 0.163 | 0.072 | 2.263 | 0.025 |
| Survey | Intercept (Univoltine) | | 1.696 | 0.590 | 2.875 | 0.005 |
| | Wing length | | 0.122 | 0.049 | 2.467 | 0.015 |
| | Voltinism | Bivoltine | –0.116 | 0.269 | –0.430 | 0.668 |
| | | Multivoltine | 1.022 | 0.384 | 2.661 | 0.009 |

evolutionary history and mimetic fidelity, but with some lability. Similarity to *V. germanica* in the most basal of the taxa used, *Mixogaster* and *Microdon*, indicates that mimicry evolved early. However, this is a very provisional result because we could only sample two species of this very diverse predominantly Neotropical subfamily (552 species, Reemer and Stahls 2013a). Despite this, the deepest nodes had similarity estimates lower than our threshold, suggesting that the basal character state for the Syrphidae was nonmimicry of wasps, and that our Microdontinae may not appropriately represent the ancestral phenotype (although one of them, *Mixogaster*, is thought to be basal among the Microdontinae: Reemer and Stahls 2013b).

Our results suggest that wasp mimicry has occasionally been lost deep within a clade of good wasp mimics; thus, to assume that conspicuous wasp-mimetic hoverflies always evolve from nonmimetic ancestral phenotypes may be inappropriate (Fig. 3; see also Kikuchi and Pfennig 2010; Hodson and Lehtinen 2017). The loss of mimetic accuracy could result from an alteration in the selective environment, meaning that wasp mimicry was no longer an advantageous adaptation. For example, none of the ecological traits examined for *Leucozona lucorum* were noticeably different relative to its closely related taxa, so one possible explanation for the loss of mimetic resemblance to wasps could be a change in hymenopteran model. *Leucozona lucorum* has been described as "a little bumblebee-like," unlike closely related taxa that have been identified more with mimics of social and solitary wasps (Röder 1990). This supports the conclusion that additional research on the similarity between hoverflies and other models is needed to understand the evolution of this multifaceted trait fully (see below).

In all our analyses, wing length was a good predictor of wasp mimicry (Tables 1 and 2). Larger species were typically better wasp mimics, which corresponds with experimental results and theoretical hypotheses from previous papers (Penney et al. 2012; Taylor et al. 2016a). There may be greater selection pressure on larger hoverflies to deceive predator visual systems because they are more nutritionally profitable prey items (Penney et al. 2012). Smaller hoverflies also take longer to warm up to flight temperatures (Morgan and Heinrich, 1987), potentially increasing the thermoregulatory cost of pale colors (Taylor et al. 2016a), since darker colors allow hoverflies to warm up more rapidly (Holloway et al. 1997). Thus, thermoregulatory costs might act in opposition to selection for accurate mimicry, especially in smaller species. This is demonstrated by the 26 species with entirely black abdomens, which all had wing lengths below 10 mm (Fig. S14). Alternatively, small size may enable predators to discriminate prey from models, and hence there is no benefit for a small species evolving to be mimetic.

Voltinism was also an explanatory variable for pattern similarity. Multivoltine species had significantly more similar abdomen patterns to *V. germanica*, and were therefore better mimics, than univoltine species, with bivoltine species being intermediate (Table 2). More generations per year may lead to better mimicry because there are more chances for selection to act in a given time frame (Gillman and Wright 2014). Furthermore, univoltine species emerge at a particular time of year for a relatively short time, and if this does not coincide with a high abundance of models there may be less selection for good mimicry (Howarth and Edmunds 2000; Finkbeiner et al. 2018; Hassall et al. 2019). Multivoltine species are essentially present all year round, and so are bound to coincide with the peaks of wasp abundance in spring, when queens search for nests, and late summer when the nest is at maximum size (Tryjanowski et al. 2010). Although phenology was not a significant predictor of wasp mimicry (Table S3), results suggest that the earliest emerging species could in general be the weakest mimics, which is somewhat consistent with this hypothesis (Fig. S16B).

The selection and definition of traits for study by evolutionary biologists is always influenced by human perception, and is by necessity somewhat arbitrary. To the human eye, mimicry is clearly present in some hoverflies, and absent in others, but studying this variation scientifically requires us to define the trait more

precisely, answering questions about sensory modality (e.g., are we considering only visual mimicry?), specificity (e.g., are we considering mimicry of one model species or several?), and variability (e.g., is mimicry a quantitative or discrete trait)? By choosing to study similarity to the abdomen pattern of *V. germanica*, we were able to make considerable progress in quantifying variation in mimicry across the hoverflies. Interestingly, despite the variable approach to the characterization of mimicry in the literature, our tightly defined quantitative measure typically corresponded very well with more subjective evaluations from other published studies. The correspondence was not perfect, however, and the descriptive classification of hoverflies as "good" or "poor" mimics in particular was not a strong predictor of our similarity scores. The failure to differentiate between good and poor mimics may either be because humans perceive mimicry in a fairly binary manner, or because the classification into "good" and "poor" in the literature has not been made in a consistent or systematic way.

By comparing two different benchmarks for wasp mimicry to how it is categorized by the literature, we aimed to gain insight into the effects of different methods for defining mimicry as a discrete trait. Figure 2 and our binary analyses highlight how wasp mimicry is more of a continuous spectrum than a binary, or categorical trait, which has important implications for how future studies define mimicry. It is also important to note the majority threshold for mimicry was still passed by 52% of the hoverflies studied here, suggesting that wasp mimicry could be a much more prevalent feature of natural communities than previously estimated (22%: Gilbert 2005; Kikuchi et al. 2021). Even the vaguest resemblance to a noxious or abundant model can afford protection to a mimic, perhaps because the optimal predator behavior may be to avoid risks by not sampling even poor mimics whenever possible, resulting in relaxed selection on mimetic accuracy (Gilbert 2005; Pfennig and Kikuchi 2012; Sherratt and Peet-Paré 2017). Just as Nicholson (1927) claimed almost 100 years ago, our results suggest that the literature may have underestimated the amount of mimicry in nature, potentially by overestimating the gap in predation pressure among mimics (Dittrigh et al. 1993).

An alternative explanation for our apparent detection of previously undescribed mimics is that the taxa with intermediate accuracy (in Fig. 2) may actually have abdomens that are never perceived to be mimetic by predators. The subjective evaluations of wasp mimicry from the literature were typically made on the basis of the entire appearance, and possibly even the behavior, of the organism. Some species with nonmimetic abdomens may thus be regarded as mimics for other reasons, and this may mean that the thresholds we used (in Fig. 2) are poorly positioned to define abdominal pattern mimicry. Additionally, a taxon was defined as a "nonmimic" of *V. germanica* when there was no literature to say otherwise, but many of these taxa were reported to be good mimics of other models that themselves resemble wasps. For example, the "nonmimic" *Microdon analis* has been described as a good honeybee mimic (Röder 1990) but also received a high similarity score when compared to *V. germanica*. Essentially, the overshadowing by more obvious putative models has contributed to the inconclusive definition of Batesian mimicry (Gilbert 2005). Evidently, subjective literature assessments are not a reliable source for defining mimetic accuracy.

The evaluation of mimicry as a trait is complicated considerably by the choice of model taxon with which putative mimics are compared. If similarity scores were high for several different models, this could be evidence for the multimodel hypothesis, whereby some mimetic phenotypes are predicted to be an optimal intermediate between several aposematic models (Edmunds 2000; Sherratt 2002). However, mimicry of animals as different as bumblebees and social wasps can involve very different morphological (and other—e.g., behavioral, or perhaps even acoustic) characters, presumably encoded by different sets of genes. If we want to explore the pattern of selection on mimicry across the phylogeny, it seems sensible to start by focusing on a more narrowly defined trait, where it is likely the mimetic phenotypes exhibited by different species are mostly homologous. So, we chose to examine visual mimicry of the social wasp, *V. germanica*. *Vespula germanica* is the most common and widespread species of social wasp across the Holarctic, so it provides a reasonable best guess at the phenotypic target for selection on this form of mimicry. Our results were largely insensitive to this choice: hoverfly similarity to two other social wasps (*V. vulgaris* and *P. dominula*) showed similar patterns across the phylogeny, and similar associations with life history traits. A fascinating unanswered question is how social wasp mimicry in hoverflies is related to mimicry of other Hymenoptera. For example, to what extent were the genes and corresponding phenotypes involved in wasp mimicry co-opted in honeybee or even bumblebee mimicry (or vice versa) during diversification of the lineage? Are the different forms of mimicry seen in hoverflies, corresponding to different model taxa, driven by similar predators, and associated with similar life history traits? Only by addressing these questions with further research will we understand the extent to which it is reasonable to consider hoverfly mimicry of any hymenopteran to be a meaningful single trait.

This research has provided insights into the ecological and evolutionary factors that shape complex phenotypes by advancing our understanding of mimetic pattern evolution in a well-studied Batesian system (Penney et al. 2012; Kikuchi and Pfennig 2013; Marchini et al. 2017). Our results suggest that wasp mimicry is a relatively labile trait that has evolved repeatedly, and that this is at least partly predictable from life history.

Because these conclusions apply specifically to the hoverfly abdomen in its visual mimicry of social wasps, further work is needed to explore the extent to which different forms of mimicry (e.g., toward other model Hymenoptera, and in other sensory modalities) show similar patterns of evolution. It is clear to us, however, that objective phylogenetically controlled comparative studies of mimicry continue to illuminate the selective forces that shape the evolution of phenotypes in natural populations.

## AUTHOR CONTRIBUTIONS

TR and AL conceived and designed the study. TR and FG supervised the project. AL collected the data and performed each analysis. CHT wrote the MATLAB code and assisted image analysis. FG provided hoverfly expertise and MRES assisted with the phylogenetic comparative analysis. AL and TR took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

## ACKNOWLEDGMENTS

The authors would like to thank P. Reader for coding the website for testing our student volunteers.

## DATA ARCHIVING

Datasets are available on Dryad (https://doi.org/10.5061/dryad.15dv41nxx).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## LITERATURE CITED

Abràmoff, M. D., P. J. Magalhães, and S. J. Ram. 2004. Image processing with ImageJ. Biophotonics Int. 11:36–42.

Azmeh, S., J. Owen, K. Sørensen, D. Grewcock, and F. Gilbert. 1998. Mimicry profiles are affected by human-induced habitat changes. Proc. R. Soc. B Biol. Sci. 265:2285–2290.

Bates, H. W. 1862. XXXII. Contributions to an insect fauna of the Amazon valley. Lepidoptera: Heliconidæ. Trans. Linn. Soc. Lond. 23:495–566.

Beaulieu, J. M., R. H. Ree, J. Cavender-Bares, G. D. Weiblen, and M. J. Donoghue. 2012. Synthesizing phylogenetic knowledge for ecological research. Ecology 93:S4–S13.

CABI. 2019. Vespula germanica distribution map. Invasive Species Compendium. CAB International, Wallingford, U.K. Available via www.cabi.org/isc.

Chandler, P. J. 1998. *Checklists of insects of the British Isles*. Handbooks *for the* identification of British insects. Royal Entomological Society, Lond.

Cuthill, I. C., and A. T. Bennett. 1993. Mimicry and the eye of the beholder. Proc. R. Soc. Lond. B Biol. Sci. 253:203–204.

Dittrigh, W., F. Gilbert, P. Green, P. McGregor, and D. Grewcock. 1993. Imperfect mimicry: a pigeon's perspective. Proc. R. Soc. Lond. B Biol. Sci. 251:195–200.

Edmunds, M. 2000. Why are there good and poor mimics? Biol. J. Linn. Soc. 70:459–466.

Edmunds, M., and T. Reader. 2014. Evidence for Batesian mimicry in a polymorphic hoverfly. Evolution 68:827–839.

Finkbeiner, S. D., P. A. Salazar, S. Nogales, C. E. Rush, A. D. Briscoe, R. I. Hill, M. R. Kronforst, K. R. Willmott, and S. P. Mullen. 2018. Frequency dependence shapes the adaptive landscape of imperfect Batesian mimicry. Proc. R. Soc. Lond. B Biol. Sci. 285:20172786.

Freckleton, R. P. 2009. The seven deadly sins of comparative analysis. J. Evol. Biol. 22:1367–1375.

Gilbert, F. 2005. The evolution of imperfect mimicry. Pp. 231–288 *in* M. Fellowes, G. Holloway, and J. Rolff, eds. Insect evolutionary ecology. CABI Publishing, Wallingford, U.K.

Gillman, L. N., and S. D. Wright. 2014. Species richness and evolutionary speed: the influence of temperature, water and area. J. Biogeogr. 41: 39–51.

Golding, Y. C., M. Edmunds, and A. R. Ennos, 2005. Flight behaviour during foraging of the social wasp Vespula vulgaris (Hymenoptera: Vespidae) and four mimetic hoverflies (Diptera: Syrphidae) *Sericomyia silentis*, *Myathropa florea*, *Helophilus* sp. and *Syrphus* sp. J. Exp. Biol. 208:4523–4527.

Grafen, A. 1989. The phylogenetic regression. Philos. Trans. R. Soc. Lond. B Biol. Sci. 326:119–157.

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2007. GEIGER: investigating evolutionary radiations. Bioinformatics 24:129–131.

Hassall, C., J. Billington, and T. N. Sherratt. 2019. Climate-induced phenological shifts in a Batesian mimicry complex. Proc. Natl. Acad. Sci. USA 116:929–933.

Hodson, E. E., and R. M. Lehtinen. 2017. Diverse evidence for the decline of an adaptation in a coral snake mimic. Evol. Biol. 44:401–410.

Holen, Ø. H., and R. A. Johnstone. 2004. The evolution of mimicry under constraints. Am. Nat. 164:598–613.

Holloway, G. J., C. G. Marriott, and H. J. Crocker. 1997. Phenotypic plasticity in hoverflies: the relationship between colour pattern and season in *Episyrphus balteatus* and other Syrphidae. Ecol. Entomol. 22:425–432.

Holloway, G., F. Gilbert, and A. Brandt. 2002. The relationship between mimetic imperfection and phenotypic variation in insect colour patterns. Proc. R. Soc. Lond. B Biol. Sci. 269:411–416.

Howarth, B., and M. Edmunds. 2000. The phenology of Syrphidae (Diptera): are they Batesian mimics of Hymenoptera? Biol. J. Linn. Soc. 71:437–457.

Howarth, B., M. Edmunds, and F. Gilbert. 2004. Does the abundance of hoverfly (Syrphidae) mimics depend on the numbers of their hymenopteran models? Evolution 58:367–375.

Ives, A. R., and T. Garland Jr. 2009. Phylogenetic logistic regression for binary dependent variables. Syst. Biol. 59:9–26.

Katzourakis, A., A. Purvis, S. Azmeh, G. Rotheray, and F. Gilbert. 2001. Macroevolution of hoverflies (Diptera: Syrphidae): the effect of using higher-level taxa in studies of biodiversity, and correlates of species richness. J. Evol. Biol. 14:219–227.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Kikuchi, D. W., and D. W. Pfennig. 2010. High-model abundance may permit the gradual evolution of Batesian mimicry: an experimental test. Proc. R. Soc. Lond. B Biol. Sci. 277:1041–1048.

Kikuchi, D. W., and D. W. Pfennig. 2013. Imperfect mimicry and the limits of natural selection. Q. Rev. Biol. 88:297–315.

Kikuchi, D. W., M. E. Herberstein, M. Barfield, R. D. Holt, and J. Mappes. 2021. Why aren't warning signals everywhere? On the prevalence of aposematism and mimicry in communities. Biol. Rev. Camb. Philos. Soc. https://doi.org/10.1111/brv.12760.

Maddison, W. P., and D. R. Maddison. 2018. Mesquite: a modular system for evolutionary analysis. Version 3.6. Available via http://www.mesquiteproject.org.

Marchini, M., D. Sommaggio, and A. Minelli. 2017. Playing with black and yellow: the evolvability of a Batesian mimicry. Evol. Biol. 44:100–112.

Mathworks. 2018. MATLAB. The Mathworks, Natick, MA.

McLean, D. J., G. Cassis, D. W. Kikuchi, G. Giribet, and M. E. Herberstein. 2019. Insincere flattery? Understanding the evolution of imperfect deceptive mimicry. Q. Rev. Biol. 94:395–415.

Mengual, X., G. Ståhls, A. Vujić, and M. A. Marcos-Garcia. 2006. Integrative taxonomy of Iberian *Merodon* species (Diptera, Syrphidae). Zootaxa 1377:1–26.

Mengual, X., G. Ståhls, P. Láska, L. Mazánek, and S. Rojo. 2018. Molecular phylogenetics of the predatory lineage of flower flies *Eupeodes-Scaeva* (Diptera: Syrphidae), with the description of the Neotropical genus *Austroscaeva* gen. nov. J. Zool. Syst. Evol. Res. 56:148–169.

Moore, C. D., and C. Hassall. 2016. A bee or not a bee: an experimental test of acoustic mimicry by hoverflies. Behav. Ecol. 27:1867–1774.

Moran, K. M., J. H. Skevington, S. Kelso, X. Mengual, K. Jordaens, A. D. Young, G. Ståhls, V. Mutin, S. Bot, M. van Zuijen, et al. 2021. A multigene phylogeny of the eristaline flower flies (Diptera: Syrphidae), with emphasis on the subtribe Criorhinina. Zool. J. Linn. Soc. https://doi.org/10.1093/zoolinnean/zlab006.

Moran, K. M., and J. H. Skevington, 2019. Revision of world *Sphecomyia* Latreille (Diptera, Syrphidae). ZooKeys 836:15–79.

Morgan, K. R., and B. Heinrich, 1987. Temperature regulation in bee-and wasp-mimicking syrphid flies. J. Exp. Biol. 133:59–71.

Nicholson, A. J. 1927. A new theory of mimicry in insects. Aust. Zool. 5:10–104.

Neuwirth, E., and M. E. Neuwirth. 2011. Package 'RColorBrewer'. CRAN 2011-06-17 08: 34: 00. Apache License 2.0.

Orme, D., R. Freckleton, G. Thomas, T. Petzoldt, S. Fritz, N. Isaac, and W. Pearse. 2018. caper: comparative analysis of phylogenetics and evolution in R. R Package Version 1.0.1. Available via https://CRAN.R-project.org/package=caper.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Paradis, E., and K. Schliep. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analysis in R. Bioinformatics 35: 526–528.

Pauli, T., T. O. Burt, K. Meusemann, K. Bayless, A. Donath, L. Podsiadlowski, C. Mayer, A. Kozlov, A. Vasilikopoulos, S. Liu, et al. 2018. New data, same story: phylogenomics does not support Syrphoidea (Diptera: Syrphidae, Pipunculidae). Syst. Entomol. 43:447–459.

Penney, H. D., C. Hassall, J. H. Skevington, K. R. Abbott, and T. N. Sherratt. 2012. A comparative analysis of the evolution of imperfect mimicry. Nature 483:461–464.

Penney, H. D., C. Hassall, J. H. Skevington, B. Lamborn, and T. N. Sherratt. 2014. The relationship between morphological and behavioral mimicry in hover flies (Diptera: Syrphidae). Am. Nat. 183:281–289.

Pfennig, D. W., and D. W. Kikuchi. 2012. Competition and the evolution of imperfect mimicry. Curr. Zool. 58:608–619.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available via https://www.R-project.org/.

Reemer, M., and G. Ståhls. 2013a. Generic revision and species classification of the Microdontinae (Diptera, Syrphidae). ZooKeys 288:1–123.

———. 2013b. Phylogenetic relationships of Microdontinae (Diptera: Syrphidae) based on molecular and morphological characters. Syst. Entomol. 38:661–688.

Revell, L. J. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Meth. Ecol. Evol. 3:217–223.

Röder, G. 1990. Biologie der Schwebfliegen Deutschlands (Diptera: Syrphidae). E. Bauer, Keltern-Weiler, Germany.

Rotheray, G., and F. Gilbert. 1999. Phylogeny of Palaearctic Syrphidae (Diptera): evidence from larval stages. Zool. J. Linn. Soc. 127:1–112.

Rotheray, G., and F. Gilbert, 2008. Phylogenetic relationships and the larval head of the lower Cyclorrhapha (Diptera). Zool. J. Linn. Soc. 153:287–323.

Rotheray, G. F., and F. Gilbert, 2011. The natural history of hoverflies. Forrest Text, Cardigan, U.K.

Ruxton, G. D., T. N. Sherratt, M. P. Speed, M. P. Speed, and M. Speed, 2018. Avoiding attack: the evolutionary ecology of crypsis, warning signals and mimicry. Oxford Univ. Press, Oxford, U.K.

Sherratt, T. N., 2002. The evolution of imperfect mimicry. Behav. Ecol. 13:821–826.

Sherratt, T. N., and C. A. Peet-Paré, 2017. The perfection of mimicry: an information approach. Philos. Trans. R. Soc. B Biol. Sci. 372:20160340.

Speed, M. P., and G. D. Ruxton, 2010. Imperfect Batesian mimicry and the conspicuousness costs of mimetic resemblance. Am. Nat. 176:1–14.

Speight, M. C. D. 2018. Species accounts of European Syrphidae. Pp. 1–305 *in* M. C. D. Speight, E. Castella, P. Obrdlik, and S. Ball, eds. Syrph the Net, the Database of European Syrphidae (Diptera). Vol. 103. Syrph the Net Publications, Dublin, Ireland.

Speight, M. C. D., and M. de Courcy Williams. 2018. European Syrphid Genera 2018: portraits of representative species. Syrph the Net, the Database of European Syrphidae. Vol. 102. Syrph the Net Publications, Dublin, Ireland.

Ståhls, G., H. Hippa, G. Rotheray, J. Muona, and F. Gilbert. 2003. Phylogeny of Syrphidae (Diptera) inferred from combined analysis of molecular and morphological characters. Syst. Entomol. 28:433–450.

Stubbs, A. E., and S. J. Falk. 2002. British hoverflies: an illustrated identification guide. British Entomological and Natural History Society, Wokingham, U.K.

Taylor, C. H., F. Gilbert, and T. Reader. 2013. Distance transform: a tool for the study of animal colour patterns. Methods Ecol. Evol. 4:771–781.

Taylor, C. H., T. Reader, and F. Gilbert. 2016a. Why many Batesian mimics are inaccurate: evidence from hoverfly colour patterns. Proc. R. Soc. Lond. B Biol. Sci. 283:20161585.

———. 2016b. Hoverflies are imperfect mimics of wasp colouration. Evol. Ecol. 30:567–581.

Taylor, C. H., J. Warrin, F. Gilbert, and T. Reader. 2017. Which traits do observers use to distinguish Batesian mimics from their models? Behav. Ecol. 28:460–470.

Tryjanowski, P., T. Pawlikowski, K. Pawlikowski, W. Banaszak-Cibicka, and T. H. Sparks. 2010. Does climate influence phenological trends in social wasps (Hymenoptera: Vespinae) in Poland? Eur. J. Entomol. 107:203–208.

Young, A. D., A. R. Lemmon, J. H. Skevington, X. Mengual, G. Ståhls, M. Reemer, K. Jordaens, S. Kelso, E. M. Lemmon, M. Hauser, et al. 2016. Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). BMC Evol. Biol. 16:143.

## *Supporting Information*

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary information
Supplementary information
Supplementary information