Emergence of a floral colour polymorphism by pollinator-mediated overdominance

Roman T. Kellenberger1,6, Kelsey J.R.P. Byers1,7, Rita M. De Brito Francisco2, Yannick M. Staedler3, Amy M. LaFountain4, Jürg Schönenberger3, Florian P. Schiestl1 & Philipp M. Schlüter1,5

Maintenance of polymorphism by overdominance (heterozygote advantage) is a fundamental concept in evolutionary biology. In most examples known in nature, overdominance is a result of homozygotes suffering from deleterious effects. Here we show that overdominance maintains a non-deleterious polymorphism with black, red and white floral morphs in the Alpine orchid Gymnadenia hellicani. Phenotypic, metabolomic and transcriptomic analyses reveal that the morphs differ solely in cyanidin pigments, which are linked to differential expression of an anthocyanidin synthase (ANS) gene. This expression difference is caused by a premature stop codon in an ANS-regulating R2R3-MYB transcription factor, which is heterozygous in the red colour morph. Furthermore, field observations show that bee and fly pollinators have opposite colour preferences; this results in higher fitness (seed set) of the heterozygous morph without deleterious effects in either homozygous morph. Together, these findings demonstrate that genuine overdominance exists in nature.

1 Department of Systematic and Evolutionary Botany, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland. 2 Department of Plant and Microbial Biology, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland. 3 Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, 1030 Vienna, Austria. 4 Department of Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Road, Storrs, CT 06269-3043, USA. 5 Institute of Botany, University of Hohenheim, Garbenstraße 30, 70599 Stuttgart, Germany. 6 Present address: Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK. 7 Present address: Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. Correspondence and requests for materials should be addressed to R.T.K. (email: rk615@cam.ac.uk) or to P.M.Süt. (email: philipp.schlueter@uni-hohenheim.de)
O

overdominance is defined as a fitness advantage of individuals based on heterozygosity at a single locus\(^3\). Historically, overdominance was considered an important mechanism maintaining polymorphism in natural populations, as it provides a simple explanation for the retention of multiple alleles in a population\(^2\). Although a range of putatively overdominant cases have been identified, very few are documented well enough to (1) establish a clear connection between genotype, phenotype and the selective force, and (2) exclude any influence of other genetic loci or other selection pressures on the examined phenotypic trait\(^4,5\). The well-understood examples are usually associated with intense, short-term selective pressures such as host–pathogen interactions\(^6\)–\(^8\) or artificial selection\(^9\)–\(^12\). In these cases, homozygosity of the derived allele generally has a detrimental effect on fitness, and cessation of selection would revert the fitness advantage of the heterozygote. Overdominance is thus often seen as a rather short-lived phenomenon, and its perceived significance in the maintenance of polymorphism has somewhat diminished in favour of other modes of selection\(^16,17\).

Polymorphism of visual traits such as colour has been studied for over a century, as it is easily accessible and can provide insights into fundamental evolutionary processes\(^18\). In plants, polymorphism of floral display traits is often highly fitness relevant due to their impact on the interactions of a plant with pollinators\(^19,20\). Most cases of floral polymorphism have been associated with antagonistic pleiotropy\(^21\)–\(^24\), negative frequency dependence\(^25,26\) or spatiotemporal heterogeneity\(^27\)–\(^29\). To our knowledge, the action of heterozygote advantage has been proposed in the colour-polyomorphic annuals *Cosmos bipinnatus*\(^30\), *Ipomoea purpurea*\(^31\) and *Sisyrinchium* sp.\(^32\). However, selection on the *Cosmos* colour morphs may not be strictly overdominant as the morphs also differ in number and size of flowers\(^30\) likewise, pollinator-mediated selection in the other two species seems to be influenced by other factors (pleiotropic effects of the colour locus in *Ipomoea*\(^33,34\) and reproductive interference in *Sisyrinchium*\(^35\)).

A yet unexplored floral colour polymorphism exists in a population of the Alpine orchid *Gymnadenia* (syn. *Nigritella* *rheicicani* located on the Alpine plateau Pflatsch (Seiser Alm) in Northern Italy (Supplementary Table 1). Unlike in other locations, where the single inflorescences of *G. rheicicani* bear uniformly ‘black’ flowers, this population is strikingly polymorphic with 62% (wild type) ‘black’, 28% ‘red’ and 10% ‘white’ individuals (census 2015, Fig. 1a). Within the population, there is no difference in the spatial distribution between morphs, which often grow in close proximity (5 cm) to each other (Supplementary Fig. 1), and morphs do not switch colour throughout their perennial life cycle. Variation in floral

**Fig. 1** Ecological evidence for overdominance in *Gymnadenia rheicicani*. a Three *G. rheicicani* floral colour morphs are growing intermixedly at Pflatsch. b The fraction of red and white morphs is increasing at Pflatsch, suggesting a fitness advantage of the white, red or both morphs. c Red, intermediate plants have the highest reproductive fitness (ANOVA (\(n = 173\)), \(F(4, 168) = 8.935, P = 1.446 \times 10^{-6}\) and Tukey’s HSD post hoc test \(t_{\text{Black-Red}} = 3.334, t_{\text{Black-White}} = 0.003, t_{\text{Black-White}} = -1.168, t_{\text{Black-White}} = 0.467, t_{\text{Red-White}} = -3.250, P_{\text{Red-White}} = 0.004\); bars denote ± 1 standard deviation). d Shielding plants (ungrazed Ofenpass site) from pollinators yielded low seed set, indicating pollinator-dependent reproduction (two-sided t test (\(n = 42\)), \(t(34.818) = -4.939, P = 1.955 \times 10^{-5}\), the outlier value was caused by an accidentally trapped grasshopper leaving a hole in the net); centre lines denote medians, bounds of boxes denote first and third quartiles, whiskers denote 1.5× interquartile ranges. e Time-lapse recordings (2017), showing bee preference for dark and fly preference for bright plants, maximising seed set in the red morph (\(\chi^2\) test, Supplementary Fig. 1)
Coloration is thus likely genetically determined rather than a result of, for example, soil chemistry. To our knowledge, this population cannot be older than ca. 2000 years, when forests on Pfrightsch were cleared; with the first notion of the colour polymorphism is from 1906, with a photographic record starting in 1971. Moreover, frequency estimations conducted between 1997 and 2016 indicate an increase in relative abundance of the red and white colour morphs from total <5% to >40% (Fig. 1b, 1997–2014 data courtesy of Richard Lorenz). Although assessing evolutionary change from Hardy–Weinberg proportions in Gymnadenia populations is hampered by their pronounced age structure (cohorts of long-lived, periodically dormant individuals), the change in morph frequencies suggests the current action of strong...
selection pressure favouring the white, the intermediate or both alternate morphs in this system.

Here we use an eco-evo-devo approach to identify the genetic mechanism, the developmental basis and the evolutionary implications of the floral colour polymorphism in *G. rhellicani*. Taken together, our results suggest that this polymorphism has emerged by pollinator-mediated overdominance: the red colour morph is heterozygous for a single-nucleotide polymorphism (SNP), which introduces a premature stop codon in an *R2R3-MYB* transcription factor. This leads to reduced expression of an *anthocyanidin synthase* (ANS) gene, resulting in a reduction of red cyanidin pigment in flower tissue, apparently without affecting other phenotypic traits. The two main pollinator groups of this population, bees and flies, show opposite colour preferences and both visit the red colour morph in relatively high numbers. As a consequence, the red colour morph has the highest seed set and increases in abundance. Overdominance in *G. rhellicani* thus appears entirely based on colour, and unlike in most other overdominant systems in natural populations, the mutation does not seem to cause any deleterious effects in either of the homozygotes.

Results

Evolutionary implications of the colour polymorphism. To assess the selective forces maintaining the floral colour polymorphism in *G. rhellicani*, we first quantified both inflorescence survival (fraction of inflorescences eaten by insects and livestock as well as reappearance of inflorescences in the next season) and reproductive fitness (number of fruits × percentage of viable seeds) of 281 plants (Supplementary Table 2). While there was no difference in inflorescence survival between colour morphs (Supplementary Fig. 2a), red plants had a significantly higher seed set than either of the extreme colour morphs (Fig. 1c). This pattern could mainly be attributed to variation in the number of fruits (Supplementary Fig. 2b and c), which in orchids develop only after successful pollination\(^{40}\), supporting the hypothesis of pollinator-mediated heterozygote advantage. Flow cytometry of collected pollinia identified all colour morphs as diploid (Supplementary Fig. 2d), and shielding plants with glass fibre mesh showed that pollinators are indeed required for seed set (Fig. 1d). This rules out apomictic or autogamous reproduction as known from polyploid Gymnadenia species\(^{41}\).

*Gymnadenia rhellicani* is a nectar-producing plant visited by >60 different insect species from various orders\(^{42,43}\), and observations at Puflatsch have shown that pollination takes place during daytime with bees and flies as main carriers of pollinia (Supplementary Table 3). To test for differences in pollinator visitation, we (1) caught and identified insects on inflorescences in 2015, and placed ten cameras in front of (2) in situ pairs of black and red plants for four consecutive days in 2016, and (3) triplets of freshly cut inflorescences (one per colour morph) for 3 days in 2017, taking pictures every 60 s (2016: 25,007 frames; 2017: 14,419 frames). In 2016, we counted 160 landings of insects (18% bees, 81% flies and 1% other) and in 2017, we recorded 197 landings (24% bees, 64% flies and 12% other). While bees visit more dark than bright colour morphs, flies on the contrary visit more bright than dark colour morphs (Fig. 1e, Supplementary Fig. 2e, Supplementary Table 3). Although bees are less common, a limited survey in the population suggests that they are likely more efficient pollinators than flies (2 of 6 bees = 33% vs. 3 of 29 flies = 10% carried pollinia). Different preferences by bees and flies thus seem to exert opposite directional selection pressures on flower colour. The red, intermediate colour morph thereby attracts both pollinators in relatively high number, providing an explanation for the observed fitness advantage. Mapping of spectral reflectance data (λ = 300–700 nm) of flowers (Supplementary Fig. 2f) in the fly\(^{44,45}\) and bee\(^{46}\) visual spaces further suggested that both pollinator groups distinguish the colour morphs primarily based on luminance (Supplementary Fig. 2g and h).

Phenotypic basis for pollinator choice. Considering these findings, bees and flies could either distinguish colour morphs by flower coloration intensity only or could rely on other phenotypic traits that correlate with colour. We therefore quantified a series of other potentially pollinator-relevant traits including plant height, flower number and inflorescence temperature in the morning and at noon, and also performed micro-computed tomography (µCT) of flowers to assess three-dimensional (3D) floral morphology. Our analyses showed that the three colour morphs do not differ in any of these traits (Supplementary Fig. 3a–d). We also collected and analysed floral volatiles by headspace sorption and gas chromatography mass-selective detection (GC-MSD) and assessed the composition of the floral volatile bouquet\(^{47}\). Neither the whole volatile bouquet nor the emission of individual scent compounds were different between the colour morphs (with the exception of one minor compound, Supplementary Figs. 3e and 4, Supplementary Table 4). A comparison of temporal volatile emission between 10:00, 15:30 and 21:00 further showed that all colour morphs reduce volatile emission after sunset (Supplementary Fig. 3f). *Gymnadenia rhellicani* thus does not use floral volatiles to attract crepuscular or nocturnal pollinators. Altogether, these results suggest that the underlying mutation(s) only affect(s) flower coloration rather than acting in a pleiotropic fashion on multiple traits. This further strengthens the argument that flower colour may be under overdominant selection.

Developmental basis of the colour polymorphism. To characterise the molecular basis of this polymorphism, we combined the phenotypic measurements with metabolomic and transcriptomic data from the same individuals by extracting anthocyanin and carotenoid pigments as well as RNA from the same pool of four open flowers per individual. Quantification of anthocyanins and their colourless precursors with ultra-high-performance liquid chromatography (UHPLC-MS/MS) identified cyanidin-3-glucoside and a derivative, (putatively) cyanidin-3-(6-malonylglucoside), as dominant pigments in the flowers. Another derivative, peonidin-3-glucoside, was also present to a lesser degree. The abundance of the two main cyanidin compounds is more than 7.5-fold lower in red, and more than 30-fold lower in white plants than in the black morph (Fig. 2b, Supplementary Fig. 5, Supplementary Tables 5 and 6). In contrast, HPLC quantification of carotenoids showed that the flowers mainly contain β-carotene and lutein with no difference in concentration between the three morphs (Supplementary Fig. 6a). Next, we performed messenger RNA-sequencing (mRNA-seq) of one plant per colour morph, three other wild-type black *G. rhellicani* plants from other populations, and four individuals of three closely related Gymnadenia species (Supplementary Table 2). We de novo assembled the Illumina HiSeq 2500 reads into a combined reference transcriptome, yielding a total of 836,101 contigs with a N\(_{50}\) length of 553 bp. With an additional 45 plants from the focal Puflatsch population, 32 plants from other populations and 16 Gymnadenia densiflora plants we prepared multiplexed libraries for RNA 3’ expression profiling by low-coverage mRNA-seq (Illumina HiSeq 2500), resulting in 166 million reads that were then mapped to the reference transcriptome. We next performed differential expression analysis between all three morphs and found 13 transcripts with significantly different expression levels.
Two of these, including the one with the highest statistical support, appeared to be derived from the same *anthocyanidin synthase* (*GrANS1*) gene. *ANS* encodes a key enzyme in the anthocyanin pathway (Fig. 2a, c, Supplementary Fig. 5, Supplementary Table 7).

Genetic mechanism underlying the colour polymorphism. A closer inspection of the white colour morph shows that the lateral lobes of the labellum still contain anthocyanin pigment (Fig. 2d), suggesting a mutation in an associated regulatory element rather than in the *GrANS1* gene itself. To find the underlying mutation,
we identified SNPs from the expression profiling data and performed a transcriptome-wide association study (TWAS). Here, spectral reflectance at 517 nm wavelength was used as the phenotype, since this is the mean of the two main cyanidins’ absorption maxima (516 and 518 nm). While we did not find any associations in GrANS1 itself (eight SNPs; \( P = 1 \) for all), three of the top ten SNP associations mapped to the 3’ non-coding region of the same GrMYB1 gene, a member of the R2R3-MYB transcription factor family involved in the regulation of anthocyanin production (Fig. 3a, c. Supplementary Table 8). Since the expression profiles cover only the 3’-end of transcripts, these SNPs may not be causative themselves, but indicate the presence of alleles segregating according to colour morph. Closer inspection of the full GrMYB1 transcript revealed a non-synonymous SNP with three allelic states: C (ancestral ‘wild-type’), G and A in the last exon of the coding region (Fig. 3a). Genotyping of all plants showed that black plants with high cyanidin content and GrANS1 expression are always homozygous for the wild-type allele (C/C), white plants with low cyanidin content and GrANS1 expression never contain the wild-type allele (G/G, G/A or A/A) and the red, intermediate colour morph is always heterozygous (Fig. 3b).

Evolutionary relationships and relative frequencies of GrMYB1 alleles (linked SNP positions 612 + 663) across the Alps; non-functional alleles at Puflatsch and Bondone. Map based on ALOS World 3D 30 m digital surface model from Japan Aerospace Exploration Agency, ©JAXA and administrative boundaries from GADM.

Discussion

Introduces into population genetics by T. Dobzhansky in 1951, the concept of overdominance initially raised considerable controversy within the scientific community. Since then, interest in overdominance has waned because obtaining clear and complete evidence of overdominant selection in natural populations has been surprisingly difficult, especially at the genetic level. In the few well-understood examples, heterozygotes usually have a fitness advantage because of deleterious effects in homozygotes, since (1) artificial breeding or pathogens select against wild-type homozygotes, and/or (2) the function of essential metabolic processes is impaired in derived homozygotes. In contrast, overdominance appears to have evolved under different circumstances in G. rhellicani. The investigated colour polymorphism is not connected to breeding or diseases, and the mutation does not appear to have any detrimental effect in the homozygous state. Similar to the well-known example of heterozygote advantage at the major histocompatibility complex in vertebrates, overdominance in G. rhellicani seems to be conditional on a single extrinsic factor (pollinator frequencies) instead of being maintained by superposition of an extrinsic with an opposing intrinsic factor.

Our study only lays the foundation for an understanding of the evolutionary and molecular processes in this system, and thus provides opportunities for further in-depth investigations: first, technical advances in ex situ propagation and genetic transformation of G. rhellicani may ultimately allow a direct functional verification of GrMYB1. Second, our study addresses plant–pollinator interactions from a plant’s perspective, but detailed pollinator behaviour and efficiency as well as the ultimate causes for colour preferences by the pollinators remain to be assessed. Third, our investigations were confined to flowering individuals over a few seasons, while long-term variation of...
selection as well as the effect of the mutation on other G. rhellicani life stages are currently unknown. However, pleiotropic effects are not expected to occur, as GrMYB1 seems to be restricted in its expression pattern. Indeed, mutations in R2R3-MYB transcription factors are frequently involved in evolutionary transitions of flower coloration, probably because of their minimal pleiotropic effects. This is also highlighted by the independent origin of the derived GrMYB1 alleles in the two polymorphic G. rhellicani populations.

In conclusion, our study provides clear evidence of ordinary polymorphism maintained by overdominant selection. Overdominance therefore should not be discounted as an explanation for the frequent occurrence of polymorphism in natural populations.

Methods

Study sites and sampling strategies. The main polymorphic G. rhellicani population contains around 4500 individuals distributed mainly over two ridges running east–west on the volcanic outcrop of the Alm Puflatsch, Italy (Supplementary Table 1, detailed sample sizes are listed in Supplementary Table 2). The whole area is subject to intense browsing by cattle and horses during the summer, and lower parts of both ridges are usually brown in early autumn. To enable cross-comparing analyses, the same set of 48 plants distributed over both ridges whenever possible. The second polymorphic G. rhellicani population contains around 650 individuals and is situated on Monte Bondone, Italy (Supplementary Table 1). For the locations of all other wild-type populations see Supplementary Table 1. To keep track of individual plants in all populations over multiple seasons, they were uniquely marked by burying passive integrated transponder (PTT) tags (HPT23, Biomark, USA) in 1.5 ml centrifuge tubes close to the shoots. Relocalisation of tagged plants was done with a handheld Global Positioning System (GPS) receiver and a Biomark HPR Plus reader with a BP Plus portable antenna. All experiments were carried out with the required collection and CITES permits issued by the relevant authorities.

Spatial distribution analysis. In July 2015, 14 plots were established in the Puflatsch population (Supplementary Fig. 1, Supplementary Table 1) with a dimension of 2 × 2 m, containing a total of 281 G. rhellicani plants (Supplementary Table 2). The plots were evenly distributed at 10 m intervals along the east–west axis of the entire Northern ridge, and a random azimuth of 0–5 m to the north or south. Within the plots, the position, height and colour of each flowering plant was recorded and all inflorescences were photographed. Spatial distribution of all G. rhellicani plants within all plots was modelled in R (R v.3.2.4, R Development Core Team 2016); randomness of the global spatial distribution pattern was assessed with a Kolmogorov–Smirnov Test, and differences in spatial distribution patterns between morphs were assessed with a Monte Carlo Permutation Test (1000 random permutations), both implemented in the R package spatstat v.1.56-056.

Morph frequency estimations. Census data from 1997, 1999 and 2011–2014 (courtesy of Richard Lorenz) are based on counts of inflorescences along east–west transects through the entire population, while data from 2015 to 2016 are based on counts of inflorescences within the 14 plots established in the population (see above). To ensure cross-comparability of estimations obtained with both methods, morph frequencies were additionally estimated along east–west transects by K.J.R. P.B. in 2015 when the plots were set up.

Fitness measurements. Number of flowers per inflorescence was inferred from the photographs of the plants within the 14 plots (see above) by hand-counting individual flowers with the aid of a custom marking script implemented in R. To correct for the fraction of flowers invisible on the photo, flower numbers were multiplied with a factor of 1.22, which was determined by comparing flower numbers between photos and actual inflorescences. In September 2015, the tagged plants were re-localised with a handheld GPS receiver (Garmin, Switzerland) and a Biomark HPR Plus reader with a BP Plus portable antenna. The number of eaten plants was recorded, differences in herbivory between morphs were computed in R with a χ² test and the remaining inflorescences were photographed. The number of capsules was inferred using the aforementioned custom R script. To assess seed quality, the content of five capsules per individual was photographed under a microscope with transmitted light and ×20 magnification (Carl Zeiss, Germany). Since orchid seeds have a transparent seed coat and do not contain an endosperm, seed viability can be estimated by assessing the presence of an embryo. Viability was estimated from 100 seeds per plant using the marking script. Number of capsules was multiplied by relative seed viability to produce reproductive fitness values. Resulting values >0 were BoxCox transformed using the R package caret v.6.0-76. To calculate fitness differences between morphs were then computed in R applying a Tukey’s honest significance (HSD) post hoc test on an Akaike information criterion optimised linear model with colour, standardised plant height and standardised flower number as fixed effects. In July 2016, the plots were revisited and differences in re-flowering between morphs were computed in R with a χ² test.

Pollinator exclusion. To avoid incidents with livestock at Puflatsch, this experiment was conducted in an ungrazed wild-type population on top of Ottenpass, Switzerland (Supplementary Table 1). The SNP data derived from low-coverage mRNA-seq (see below) suggests that this population shares its genetic ancestry with the colour-polymorphic population at Puflatsch (not shown). In June 2016, 40 plants in bud stage were randomly selected and PIT tagged, and 20 plants were covered with white glass fibre nets (1 mm mesh size) attached to cylindrical metal frames of 25 cm height × 10 cm diameter. In September 2016, capsules were counted and seed quality was determined as described above. Fitness differences were computed in R with a two-sided t test.

Pollinator recordings. On the 11th of July 2015, insects observed on G. rhellicani inflorescences along east–west transects on both ridges were caught with nets in the morning (10:00) and after sunset (22:00), identified and checked for pollinia. During peak flowering in July 2016, ten Somikon (Pearl, Switzerland) and one Brinno TLC 200 (Brinno, Taiwan) time-lapse camera were placed in front of in situ pairs of black and red plants growing within ca. 10 m distance of each other at the Puflatsch site. The plants were recorded at 1 min intervals during four consecutive days, with cameras shutting off automatically at night. Two people (R.T.K. and K.J.R. P.B.) independently screened all images for number and type of pollinators. In July 2017, pollinator recordings were repeated for three consecutive days: three freshly cut inflorescences (one per morph, randomly chosen) were randomly placed in ca. 5 cm distance of each other in buried 15 ml tubes filled with water in front of the camera (9 Somikon and 3 Brinno). Differences in visitation of pollinator groups between morphs were calculated in R with Holm-adjusted p values.

Flow cytometry. Ploidy analysis was based on a protocol in Gross and Schiestl56 with minor modifications. Pollinia from three flowers per individual were ground in 200 μl Otto 1 buffer (0.1 M citric acid monohydrate in 0.5% Triton X-100) in a 1.5 ml reaction tube using a pestle. An internal standard solution was prepared by grinding 2 cm² leaf tissue of Phaseolus coccineus ‘Scarlett Emperor’ (Sativa Rhénau AG, Switzerland) in 1 ml Otto 1 solution, and filtering the suspension through a 30 μm CellTrics filter (Sysmex, Germany). After addition of 5 μl of this standard solution, each sample was filtered through a 30 μm CellTrics filter and centrifuged for 5 min at 380 × g. The supernatant was removed, and nuclei were resuspended in 400 μl Otto 1 solution. Samples were analysed on a Beckman Coulter Quanta flow cytometer (Beckman Coulter, USA) loaded with 160 μl Otto 2 solution (0.4 M disodium phosphate dehydrate with 4 μg ml⁻¹ 4, 6-diamidine-2’-phenyldiisothiocyanate (hydrochloride added before usage). Ploidy was determined as the ratio between the medians of the standard and sample peaks in the flow cytometric histogram. As G. rhellicani is generally diploid, the lowest sample peak detected was assumed to represent the haploid pollinaria of diploid individuals, and the two adjacent peaks correspond to the diploid and (if present) tetraploid state.

Spectral mapping. Spectral measurements were taken with an AvaSpec spectrometer (Avantes, Netherlands) from three flowers per individual. The three spectral curves per sample were averaged between 300 and 700 nm wavelength, smoothed and normalised. The curves were mapped into the R fourn space using a custom R script based on Trome44, and into the bee visual space using an R script published by Scedek et al.57 with minor modifications.

Temperature measurements. Inflorescence temperature was measured in the morning between 09:00 and 10:00 (60 plants) and in the afternoon between 12:00 and 13:00 (90 plants) on a cloudless day in July 2016. Temperatures were recorded with an infrared thermometer (Model JH-6606, Walter Werkzeuge, Austria) in two transects from east to west along the bottom and top part of the Northern ridge.

Micro-computed tomography. In July 2014, one flower from the lower part of the inflorescence was collected and stored in 1.5 ml FAA fixative (50% ethanol, 5% glacial acetic acid, 3.7% formaldehyde). Flowers were infiltrated twice with a contrast solution (1% phosphogluconic acid in FAA) and subsequently embedded in a polyester resin (model M02, DAKO® (DAKO, Invista, USA)) according to Stadelmann et al.58. 3D flower morphology was monitored by µCT, using a MicroXCT-200 imaging system (Xradia, USA) with a 90 kV Microfocus X-ray. All samples were scanned with the following parameters: acceleration voltage, 37 kV; source current, 200 μA; exposure time, 1.5 s; number of exposures, 728; camera binning, 4x, objective, large field of view (magnification: ×10.4; voxel size, 38,7475 μm). 3D data were exported in DICOM format. In Amira v.5.3.3 (Zuse Institute, Germany), a total of 21 geometric landmarks were positioned on each scanned flower: one at the base and tip of each petal, labellum and column, three along the left and right side of the labellum, one in the centre of the labellum and one at the tip of the spur (Supplementary Fig. 3d). Landmark positions were then compared among the three colour morphs with Procrustes ANOVA (analysis of variance) from the R package geomorph v.3.0.359.
Floral volatile analysis. Volatile organic compounds (VOC) were collected with non-destructive headspace sorption, a standard method also used in other Gymnadenia conopsea studies1. Parameters were determined by a GC/MSD system (Thermo Fisher Scientific, USA) equipped with a UVD340S diode array detector set at 450 nm for carotenoid identification. The two main carotenoids β-carotene and lutein were identified and quantified as described for the anthocyanins using calibration curves of authentic reference standards.

Transcriptome assembly. mRNA-seq was performed for one individual per morph from Puflattsch, one individual from Ofenpass, Switzerland, two individuals from Chalandin, Switzerland, one G. densiflora plant from each of two populations near Tschierw, Switzerland and Davos Dorf, Switzerland, one G. odoratissima plant from a population in the Münstertal, Switzerland and one G. conopsea plant from a population near Cinnaus-chel-Brall, Switzerland (the latter two courtesy Karin Gross). RNA was extracted from the remaining 50% of the ground flowers used for UHPLC-MS/MS (see above). Total RNA was extracted with TRIzol reagent (Thermo Fisher Scientific, Massachusetts) according to the manufacturer’s protocol and subsequently purified using a Qiagen RNeasy MinElute Cleanup Kit (Qiagen, Netherlands). RNA was quantified on a 2100 Bioanalyzer (Agilent Technologies, USA) using the RNA 6000 nano assay (Agilent Technologies, USA). Libraries were prepared with a Lexogen QuantSeq 5 mRNA-Seq Kit (Lexogen, Austria) according to the manufacturer’s protocol. All libraries were pooled and sequenced single-end on one lane of an Illumina HiSeq 2500 sequencer.

Generation of mRNA expression profiles. Expression profiling was conducted for 45 plants from Puflattsch, for 26 plants from Chalandin and for 6 plants from Ofenpass (see Supplementary Table 1 for population locations). Total RNA was extracted from the other half of the ground flower tissue used for UHPLC-MS/MS (see above) using TRIzol according to the manufacturer’s protocol. Samples were quantified with a Qubit fluorometer and RNA HS Assay Kit (Thermo Fisher Scientific, USA) and libraries were prepared with a Lexogen QuantSeq 3 mRNA-Seq Kit (Lexogen, Austria) according to the manufacturer’s protocol. All libraries were pooled and sequenced single-end on one lane of an Illumina HiSeq 2500 sequencer.

Differential expression analysis. Raw expression profiles were trimmed to a minimal length of 25 pb with Trimmomatic v.0.3667. Trimmed reads were then aligned to the G. conopsea reference transcriptome using SENSE v.1.2.3170 and bowtie2 v.2.2.971 as aligner. Posterior mean counts per morph were subsequently extracted with a custom Perl script. Transcripts with fewer than 0.5 counts per million reads in at least 10 samples were removed, and negative binomial generalised log-linear models were fitted for transcripts with at least a two-fold expression difference using edgeR v.3.12.172. Transcripts with a significant expression difference were searched against the NCBI nucleotide database (nr) with blastx73.

Transcriptome-wide association study. Trimmed expression profile reads (see above) were mapped to the transcriptome reference transcriptome with bowtie2 v.2.2.971 and converted to indexed bam files. Variant calling of these files was performed using SAMtools v.1.3.173 and bcftools v.1.2.174 as described by Streisfeld and Rauscher75. In brief, the run was performed with default settings for priors until at least 50% of the output was indexed and converted to vcf format using bcftools v.1.3.174. Subsequently, conversion from vcf to hapmap format was performed in R v.3.2.3 using the biotools library function bdVc2Hapmap (https://sourceforge.net/projects/bi/). Variant calling of these files was performed with HaplotypeCaller and 1000 Genomes Project EUR panel data was used to call variants. Prior to association analysis, minor allele frequency was calculated. SNPs were selected with a minimum quality score of at least 15 and a minor allele frequency of at least 0.01. Missingness was allowed up to 0.05. Association analysis was performed for each transcript with a linear mixed model using the R package lme4 v.1.1-25. Genotyping of all flowering plants and a subset of non-flowering individuals was conducted for three sites in Italy: 88 plants from Puflattsch, 23 plants from Monte Bondone, 16 plants from Chalandin, Switzerland; five plants from each of two populations near Tschierw, Switzerland and Davos Dorf, Switzerland; and four plants in Austria: Rhaudersee, Wichtigsee and Heimberg. Genotypes were pheno-
**RNA interference.** Whole, expanding *G. rheiillici* inflorescences were cut at Oenopras (Supplementary Table 1) on 17 July 2017 and transferred to the lab in water-free conditions. The disk was placed on a tray of 10% sucrose, 1% propylene glycol, 1% MerckoPatch (Merck, Germany) in a Petri dish, and 22 inflorescences were inoculated by rubbing one side of the inflorescences in the solution, leaving the other side of the inflorescence untreated. Four additional inflorescences were mock inoculated the same way using 0.1 M phosphate buffer instead of dsRNA solution. Inflorescences were kept at 21°C/60% relative humidity/150 µmol m⁻² s⁻¹ light intensity at day (08:00 to 20:00) and 12°C/60% relative humidity/dark at night. After 7 days, five flowers each from two dsRNA-treated plants showing a phenotype and the four control plants were collected in liquid nitrogen. Total RNA was extracted from 50 mg of frozen tissues using a Qiagen RNaseasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer’s protocol, and 238 bp fragment in the last coding exon was amplified and sequenced as described above. Relative peak heights (fluorescence intensities) of each potential nucleotide (A, T, C, G) were normalized to the average peak height of the best-fitting nucleotide. The average and standard deviation of three replicate peak heights were calculated across the three reference genes. The peak heights were then converted from the average Cq of each target gene to obtain ΔCq. Ninety-nine per cent confidence intervals were calculated for the ΔCq values of the four control plants and compared to the ΔCq values of each treatment plant.

**Code availability.** All custom scripts are available from the corresponding authors upon request.

**Data availability.** The generated pCT data are deposited on Phaidra under accession number 912072, the UHP/LC-MS/MS data are deposited on figshare under accession number 7321880, the mRNA-seq and Lexogen QuantSeq data are deposited on NCBI SRA under accession number PRJNA504609, the *GrMYB1* allele sequences are deposited in the NCBI Genbank nucleotide database under accession numbers MK163677–MK163681, and the *GrANS1* allele sequences under accession numbers MK163682–MK163685. All other datasets generated and analysed in this study are deposited on figshare under accession number 7314731. A reporting summary for this article is available as a Supplementary Information file.

**References.**

1. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796 (2009).
2. Dobzhansky, T. A review of some fundamental concepts and problems of population genetics. *Cold Spring Harb. Symp. Quant. Biol.* **20**, 1–5 (1955).
3. Delph, L. F. & Kelly, I. D. On the importance of balancing selection in plants. *New Phytol.* **101**, 45–56 (2014).
4. Gemmell, N. J. & Slate, J. Heterozygote advantage for fecundity. *PLoS ONE* **1**, e125 (2006).
5. Futuyma, D. J. *Evolution* 2 edn (Sinauer Associates, Sunderland, 2009).
6. Aidoo, M. et al. Protective effects of the sickle cell gene against malaria morbidity and mortality. *Proteomics* **3**, 1311–1312 (2002).
7. Frelinger, J. A. The maintenance of transferrin polymorphism in pigeons. *Proc. Natl. Acad. Sci. USA* **109**, 6329–6332 (2012).
8. Carrington, M. et al. HLA and HIV-1: heterozygote advantage and B35-Cw∗04 disadvantage. *Science* **283**, 1748–1752 (1999).
9. Adalsteinsson, S. Establishment of equilibrium for the dominant lethal gene for Manx taillessness in cats. *Theor. Appl. Genet.* **57**, 49–53 (1980).
10. Karlsson, E. K. et al. Efficient mapping of Mendelian traits in dogs through genome-wide association. *Nat. Genet.* **39**, 1321–1328 (2007).
11. Hillbertz, N. H. S. et al. Duplication of *GrMYB1* causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat. Genet.* **39**, 1318–1320 (2007).
12. Mosher, D. S. et al. A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *Plas. Sci. Eng.* **3**, e79 (2007).

**Received:** 14 June 2018 **Accepted:** 5 December 2018 **Published online:** 08 January 2019
13. Fujii, J. & Otsu, K. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science 253, 448 (1991).

14. Greaves, J., Redfern, P. & Gill, J. Warfarin resistance: a balanced polymorphism in the Norwegian rat. Genet. Res. 30, 257–263 (1977).

15. Sells, D., Kvitek, D. J., Dunn, B., Sherlock, G. & Petrov, D. A. Heterozygote advantage is a common outcome of adaptation in Sazachromyces cerevisiae. Genetics 203, 1401–1413 (2016).

16. Favaro, S. M. & McKinnon, J. S. Linking color polymorphism maintenance and speciation. Trends Ecol. Evol. 22, 71–79 (2007).

17. Hedrick, P. W. What is the evidence for heterozygote advantage selection? Trends Ecol. Evol. 27, 698–704 (2012).

18. Svensson, E. I. Back to basics: using colour polymorphisms to study evolutionary processes. Mol. Ecol. 26, 2204–2211 (2017).

19. Kay, Q. The Role of Preferential and Assortative Pollination in the Maintenance of Flower Colour Polymorphisms Vol. 6 (Academic Press, London, 1978).

20. Russell, A. L., Newman, C. R. & Papaj, D. R. White flowers finish last; pollen-foraging bumble bees show biased learning in a color floral polymorphism. Evol. Ecol. Evol. 31, 173–191 (2017).

21. Carlson, J. E. & Helling. K. E. Natural selection on inflorescence color polymorphisms in wild Protea populations: the role of pollinators, seed predators, and intraspecific correlations. Ann. J. Bot. 97, 934–944 (2010).

22. Irwin, R. E., Strauss, S. Y., Storz, S., Emerson, A. & Guibert, G. The role of herbivores in the maintenance of a flower color polymorphism in wild radish. Ecology 84, 1733–1743 (2003).

23. Jersáková, J., Kindlmann, P. & Renner, S. S. Is the colour dimorphism in Ipomoea purpurea (C. L. Moench) L. dependent on pollinator foraging in polymorphic Clarkia xanthanana ssp. xantiana populations: implications for flower colour evolution and pollinator interactions. Oikos 112, 412–421 (2006).

24. Barrett, S.C.H. (ed) Evolution and Function of Heterostyly (Springer Verlag, Berlin and Heidelberg, 2013).

25. Schemske, D. W. & Bierzychudek, P. Spatial differentiation for flower color in the desert annual Linanthus parryi: was Wright right? Evolution 61, 2528–2543 (2007).

26. Ortiz, P., Berjano, R., Talavera, M., Rodriguez-Zayas, L. & Arista, M. Flower color polymorphism in Lysimachia arvensis: How is the red morph maintained in Mediterranean environments? Perspect Plant Ecol. Evol. System. 17, 142–150 (2015).

27. de Jager, M. L. & Ellis, A. G. Floral polymorphism and the fitness implications of attracting pollinating and florivorous insects. Ann. Bot. (Lond.) 113, 213–222 (2013).

28. Moar, B. & Nattero, J. Pollinator response to flower color polymorphism and floral display in a plant with a single-locus floral color polymorphism: consequences for plant reproduction. Ecol. Res. 27, 377–385 (2012).

29. Subramaniam, B. & Rausher, M. D. Balancing selection on a floral polymorphism. Evolution 54, 691–695 (2000).

30. Takahashi, Y., Takakura, K.-i. & Kawata, M. Flower color polymorphism maintained by overdominant selection in Symphytum sp. J. Plant Res. 128, 933–935 (1993).

31. Moinnier, L. E. & Rausher, M. D. Selection on a floral color polymorphism in the common morning glory (Ipomoea purpurea): the effects of overdominance in seed size. Evolution 51, 608–614 (1997).

32. Coberly, L. C. & Rausher, M. D. Pleiotropic effects of an allele producing white flowers in Ipomoea purpurea. Evolution 62, 1076–1085 (2008).

33. Takahashi, Y., Takakura, K.-i. & Kawata, M. Spatial distribution of flower color induced by interspecific sexual interaction. PLoS ONE 11, e0163481 (2016).

34. Kral, F. Ein pollenanalytischer Beitrag zur Vegetationsgeschichte der Seiser Alm. Der Scherrn 57, 31–36 (1983).

35. Valiente, C. W. D., Gaus, P. W. & de Sarth, L. Flora der gefürsteten Grafschaft Tirol, des Landes Vorarlberg und des Fürstenthumes Liechtenstein: Bd. Die Farn- und Blütplänzenarten (Pteridophyta et Spermatophyta) 529–530 (Wagner’sche Universitäts-Buchhandlung, Innsbruck, 1906).

36. Reinhard, H. R. in Vol. Animals, Plants and Biophotes (ed Gold, P.) (ETH-Bibliothek, Zurich, 1971).

37. Moeen, A. & Ohm, D. B. Ecology and survival of Nigritella nigra, a threatened orchid species in Scandinavia. Nord. J. Bot. 22, 435–461 (2002).

38. Zhang, X. S. & O’Neill, S. D. Ovary and gametophyte development are coordinately regulated by auxin and ethylene following pollination. Plant Cell 5, 403–418 (1993).

39. Carrington, M., Klein, E. & Teppner, H. Evolution of polyploids in the European orchid genus Nigritella: evidence from allozyme data. Phytom 40, 239–275 (2000).
This study would not have been possible without the dedicated support of Mathias Kneubühler and his trusty Land Rover. We also thank Richard Lorenz and Giorgio Peraza for sharing their population census data, Andreas Tribsch for providing additional plant samples and Anitta Dirks for providing the E. coli HT115 strain. In addition, we thank Richard Lorenz, Benjamin Kellenberger, Johann Madl, Giorgio Peraza, Sebastiano Andreatta, Tobias Kellenberger, Karin Gross and Khalid Sedeek for fieldwork support, as well as Srignanakshi Iyer, Laura Piñeiro Fernández, Barbara Keller, Franz Huber, Beverley Glover, Wolf Blanckenhorn, Philip Jörg and Christoph Gasser for laboratory and technical assistance and advice. We further thank Marialuise Kiem, Amt für Landschaftsökologie Bozen, as well as the environmental authorities of Wallis, Uri, Graubünden, Carinthia, the Tyrol, Styria and Trento for issuing collection permits. This study was funded by a University of Zurich (UZH) Forschungskredit grant to R.T.K., a PLANT FELLOWS Postdoctoral Fellowship grant to K.J.R.P.B., and three travel grants from the Claraz Schenkung Zurich (two to R.T.K. and one to K.J.R.P.B.). P.M.S. acknowledges support by the Swiss National Science Foundation (SNF, project 31003A_155943).

Acknowledgements

This study would not have been possible without the dedicated support of Mathias Kneubühler and his trusty Land Rover. We also thank Richard Lorenz and Giorgio Peraza for sharing their population census data, Andreas Tribsch for providing additional plant samples and Anitta Dirks for providing the E. coli HT115 strain. In addition, we thank Richard Lorenz, Benjamin Kellenberger, Johann Madl, Giorgio Peraza, Sebastiano Andreatta, Tobias Kellenberger, Karin Gross and Khalid Sedeek for fieldwork support, as well as Srignanakshi Iyer, Laura Piñeiro Fernández, Barbara Keller, Franz Huber, Beverley Glover, Wolf Blanckenhorn, Philip Jörg and Christoph Gasser for laboratory and technical assistance and advice. We further thank Marialuise Kiem, Amt für Landschaftsökologie Bozen, as well as the environmental authorities of Wallis, Uri, Graubünden, Carinthia, the Tyrol, Styria and Trento for issuing collection permits. This study was funded by a University of Zurich (UZH) Forschungskredit grant to R.T.K., a PLANT FELLOWS Postdoctoral Fellowship grant to K.J.R.P.B., and three travel grants from the Claraz Schenkung Zurich (two to R.T.K. and one to K.J.R.P.B.). P.M.S. acknowledges support by the Swiss National Science Foundation (SNF, project 31003A_155943).

Author contributions

R.T.K., K.J.R.P.B. and P.M.S. conceived and designed the study. R.T.K., K.J.R.P.B. and P.M.S. conducted the field investigations. R.T.K., K.J.R.P.B. and F.P.S. performed all phenotypic and scent analyses. Y.M.S., K.J.R.P.B., R.T.K. and J.S. carried out the 3D morphometric analysis. R.M.D.B.F. and R.T.K. conducted the anthocyanin analysis. A.M.L.F., R.M.D.B.F. and R.T.K. carried out the carotenoid analysis. K.J.R.P.B., R.T.K. and P.M.S. performed the transcriptome assembly and analysis. R.T.K. and P.M.S. conducted the genetic analysis. R.T.K., P.M.S. and K.J.R.P.B. wrote the software used for the analysis. R.T.K. and P.M.S. wrote the original draft of the manuscript. All authors discussed and edited the manuscript. F.P.S., P.M.S. and J.S. provided resources for this study. P.M.S., R.T.K. and K.J.R.P.B. acquired funding for this project.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07936-x.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.