White pupae phenotype of tephritids is caused by parallel mutations of a MFS transporter

Mass releases of sterilized male insects, in the frame of sterile insect technique programs, have helped suppress insect pest populations since the 1950s. In the major horticultural pests Bactrocera dorsalis, Ceratitis capitata, and Zeugodacus cucurbitae, a key phenotype white pupae (wp) has been used for decades to selectively remove females before releases, yet the gene responsible remained unknown. Here, we use classical and modern genetic approaches to identify and functionally characterize causal wp− mutations in these distantly related fruit fly species. We find that the wp phenotype is produced by parallel mutations in a single, conserved gene. CRISPR/Cas9-mediated knockout of the wp gene leads to the rapid generation of white pupae strains in C. capitata and B. tryoni. The conserved phenotype and independent nature of wp− mutations suggest this technique can provide a generic approach to produce sexing strains in other major medical and agricultural insect pests.
The sterile insect technique (SIT) is a species-specific and environment-friendly approach to control their populations, which has been successfully applied as a component of area-wide integrated pest management programs. GSS requires two principal components: a selectable marker, which could be phenotypic or conditionally lethal, and the linkage of the male sex, ideally as close as possible to the male determining region. In a GSS, males are heterozygous and phenotypically wild type, whilst females are homozygous for the mutant allele thus facilitating sex separation. Puparium color was one of the first phenotypic traits exploited as a selectable marker for the construction of GSS. In all three species, brown is the typical puparium color. However, naturally occurring color mutants such as white pupae (wp) and dark pupae (dp) have occurred in the field or laboratory stocks. The wp locus was successfully used as a selectable marker to develop GSS for C. capitata, B. dorsalis, and Z. cucurbitae; however, its genetic basis has never been resolved.

Biochemical studies provided evidence that the white pupae phenotype in medfly is due to a defect in the mechanism responsible for the transfer of catecholamines from the hemolymph to the pupal cuticle. In addition, classical genetic studies showed that the wp phenotype is due to a recessive mutation in an autosomal gene located on chromosome 5 of the medfly genome. The development of translocation lines combined with deletion and transposition mapping and advanced cytogenetic studies allowed the localization of the gene responsible for the wp phenotype on the right arm of chromosome 5, at position 59B of the trichogen polytene chromosome map. In the same series of experiments, the wp locus was shown to be tightly linked to a temperature-sensitive lethal (tsl) gene (position 59B–61C), which is the second selectable marker of the VIENNA 7 and VIENNA 8 GSS currently used in all medfly SIT operational programs worldwide.

The genetic stability of a GSS is a major challenge, mainly due to recombination phenomena taking place between the selectable marker and the translocation breakpoint. To address this risk, a chromosomal inversion called D53 was induced and integrated into the medfly VIENNA 8 GSS (VIENNA 8D53). Cytogenetic analysis indicated that the D53 inversion spans a large region of chromosome 5 (50B–59C on trichogen polytene chromosome map) with the wp locus being inside the inversion, close to its right breakpoint.

Extensive genetic and cytogenetic studies facilitated the development of a physical map of the medfly genome. The annotated gene set provided opportunities for the identification of genes or loci associated with the wp phenotype, such as the wp and tsl genes. Salivary gland polytene chromosome maps developed for C. capitata, B. dorsalis, Z. cucurbitae, and B. tryoni show that their homologous chromosomes exhibit similar banding patterns. In addition, in situ hybridization analysis of several genes confirmed that there is extensive shared synteny, including the right arm of chromosome 5 where the C. capitata wp gene is localized. Interestingly, two recent studies identified SNPs associated with the wp phenotype in C. capitata and Z. cucurbitae that were also on chromosome 5.

In this work, we employ different strategies involving genetics, cytogenetics, genomics, transcriptomics, gene editing, and bioinformatics to identify independent natural mutations in a gene responsible for puparium coloration in three tephritid species of major agricultural importance, C. capitata, B. dorsalis, and Z. cucurbitae. We then functionally characterize causal mutations within this gene in C. capitata and B. tryoni resulting in development of new white pupae strains. Due to its conserved nature and widespread occurrence in many insect species of agricultural and medical importance, we also discuss the potential use of this gene as a generic selectable marker for the construction of GSS for SIT applications.

**Results**

**Resolving the B. dorsalis wp locus by introgression experiments.** The B. dorsalis white pupae phenotype was introgressed into B. tryoni to generate a strain referred to as the Bactrocera introgressed line (BIL, Supplementary Fig. 1). To determine the proportion of B. dorsalis genome introgressed into BIL, whole-genome sequence data from male and female B. dorsalis, B. tryoni, and BIL individuals were analyzed. Paired-end Illumina short read data from single B. oleae males (SRR826808) and females (SRR826807) were used as an outgroup. Single copy orthologs across the genome (n = 1,846) were used to reconstruct the species topology revealing a species-specific monophyly (Fig. 1a) consistent with published phylogenies. Reconstruction also showed monophyly between B. tryoni and BIL across 99.2% of gene trees suggesting the majority of loci originally introgressed from B. dorsalis have been removed during backcrosses.

Genomes were partitioned into 100 kb windows and pairwise absolute genetic distance (dXY) calculated between each species and BIL to estimate admixture. Bactrocera dorsalis was found to be highly similar to a small proportion of the BIL genome (Fig. 1b; purple), as indicated by dXY values approaching the median value of B. dorsalis vs B. tryoni (Fig. 1b; yellow).

Two formal tests for introgression were also carried out, the f estimator (Fig. 1c) and topology weighting (Fig. 1d). Three distinct local evolutionary histories (Fig. 1d) were tested using dXY and topology weighting across the B. dorsalis wp Quantitative Trait Locus (QTL) i) BIL is closest to B. tryoni (Fig. 1d; purple, expected across most of the genome), ii) BIL is closest to B. dorsalis (Fig. 1d; orange, expected at the wp locus), and iii) BIL is closest to B. oleae (Fig. 1d; green, a negative control). Across the nuclear genome the species topology was supported in 98.82% of windows. Both f and topology weighting confirmed a lack of widespread introgression from B. dorsalis into BIL with few (n = 42) discordant outlier windows. Genomic windows discordant across all three tests were considered candidate regions for the wp mutation. Four scaffolds accounting for 1.18% of the B. dorsalis genome met these criteria and only two, NW_011876372.1 and NW_011876398.1, showed homoyzgyous introgression consistent with a recessive white pupae phenotype (Supplementary Fig. 2).

To resolve breakpoints within the B. dorsalis wp QTL, a windowed analysis across NW_011876398.1 and NW_011876372.1 was performed using dXY (Fig. 1e), topology weighting (Fig. 1f), and f (Fig. 1g). The maximum range of the introgressed locus was 4.49 Mb (NW_011876398.1 was 2.9–5.94 Mb and NW_011876372.1 was 0–1.35 Mb) (Fig. 1e–g). The wp QTL was further reduced to a 2.71 Mb region containing 113 annotated protein coding genes through analyzing nucleotide diversity (π) among eight pooled BIL genomes (3.8 Mb on NW_011876398.1 to 0.73 Mb on scaffold NW_011876372.1, Supplementary Fig. 2).

**Resolving the C. capitata wp by genome sequencing and in situ hybridization.** Cytogenetic studies have determined the gene responsible for the white pupae phenotype to be localized on the right arm of chromosome 5, at position 59B of the trichogen polytene...
The equivalent of position 59B is position 76B of the salivary gland polytene chromosome map, inside but close to the right breakpoint of the D53 inversion (69C–76B on the salivary gland polytene chromosome map). Long read sequencing data were generated of the wild-type strain Egypt II (EgII, WT), the inversion line D53 and the genetic sexing strain VIENNA 8 (without the inversion; VIENNA 8D53 Deletion) (Supplementary Table 1) to enable a comparison of the genomes and locate the breakpoints of the D53 inversion, to subsequently narrow down the target region, and to identify wp candidate genes.

Chromosome 5-specific markers were used to identify the EgII_Ccap3.2.1 scaffold_5 as complete chromosome 5. Candidate D53 breakpoints in EgII scaffold_5 were identified using the alignment of three genome datasets EgII, VIENNA 8D53 Deletion, and D53 (see material and methods). The position of the D53 inversion breakpoints was located between 25,455,334 and 25,455,433 within a scaffold gap (left breakpoint), and at 61,880,224 bp in a scaffolded contig (right breakpoint) on EgII chromosome 5 (Ccap3.2.1; accession GCA_905071925) (Fig. 2a). The region containing the causal wp gene was known to be just next to the right breakpoint of the D53 inversion. Cytogenetic analysis and in situ hybridization using the WT EgII strain and the D53 inversion line confirmed the overall structure of the inversion, covering the area of 69C–76B on the salivary gland polytene chromosomes (Fig. 2), as well as the relative position of markers residing inside and outside the breakpoints (Fig. 2 and Supplementary Fig. 3). PCRs using two primer pairs flanking the predicted breakpoints (Supplementary Fig. 4) and subsequent sequencing confirmed the exact sequence of the breakpoints. Thereby, the wild-type status was confirmed for EgII flies and VIENNA 7D53 Deletion GSS males, which are heterozygous for the inversion. Correspondingly, these amplicons were not present in D53 males and females or in VIENNA 7D53 Deletion GSS females (all homozygous for the inversion) (Supplementary Fig. 4). Positive signals for the inversion were detected in D53 and VIENNA 7D53 Deletion GSS males and females, but not in WT flies using an inversion-specific primer pair (SupplementaryFig. 4).

Genome and transcriptome sequencing reveal a single candidate wp gene. Orthologs within the QTL of B. dorsalis, C. capitata, and scaffolds known to segregate with the wp phenotype in Z. cucurbitae (NW_011863770.1 and NW_011863674.1) were investigated for null mutations under the assumption that errors within a conserved gene result in white pupae. A single ortholog containing fixed indels absent from wild-type strains was identified in each species. White pupae B. dorsalis and BIL strains showed a 37 bp frame-shift deletion in the first coding exon of LOC105232189 introducing a premature stop codon 210 bp from the transcription start site (Fig. 3a). Presence of the deletion was confirmed in silico using whole genome resequencing from the wp and wildtype mapped to the reference, and by de novo assembly of Illumina RNAseq data transcripts (Fig. 3a).

In C. capitata, a D53 Nanopore read alignment on EgII showed an independent approximate 8150 bp insertion into the third exon of LOC101451947 disrupting proper gene transcription 822 bp from the transcription start site (Fig. 3b). The insertion sequence is flanked by identical repeats, suggesting that it may originate from a transposable element insertion. The C. capitata mutation was confirmed in silico, as in B. dorsalis, using whole genome sequencing and RNAseq data (Fig. 3b).

Transcriptome data from the white pupae-based genetic sexing strain of Z. cucurbitae revealed a 13 bp deletion in the third exon of LOC105232189 on scaffold NW_011863770.1 introducing a premature stop codon (Fig. 3c).

The candidate white pupae gene in all three species had a reciprocal best BLAST hit to the putative metabolite transport
hybridizations were done at least in duplicates and at least ten nuclei were analyzed per sample, scale bar one inside (the polytene chromosome map of chromosome 571 (left (L) and right (R) chromosome arm, linked at the centromeric region (C)) and the PacBio-Hi-C EgII lethal (Fig. 2c–f are provided as a Source Data file).

Fig. 2 Genomic positioning of the D53 inversion on chromosome 5 of C. capitata. a Chromosome scale assembly of C. capitata EgII chromosome 5. Shown are the positions of in situ mapped genes white (w), 6-phosphogluconate dehydrogenase (Pgd), glucose-6-phosphate 1-dehydrogenase (Zw), and sex lethal (Sxl), the position of the D53 inversion breakpoints (blue; LB = left breakpoint, RB = right breakpoint), and the relative position of white pupae (wp) on the polytene chromosome map of chromosome 5(1) (left (L) and right (R) chromosome arm, linked at the centromeric region (C)) and the PacBio-Hi-C EgII scaffold_5 (bp = base pairs), representing the complete chromosome 5 (Ccap3.2.1, accession GCA_905071925). The position of the yellow gene (y, LOC101455502) was confirmed on chromosome 5 70A by in situ hybridization, despite its sequence not been found in the scaffold assembly. b Schematic illustration of chromosome 5 without (EgII, WT) and with (D53) D53 inversion, with additional marker genes Curly (Cy), integrin-αPS2 (PS2a), white (w), chiron 536/38 (Ccs36/38), vitellogenin-1/2-like (Vg1 + 2). The inverted part of chromosome 5 is shown in light blue, the centromere in yellow. Two probes, one inside (y, 70A) and one outside (Pgd, 68B) of the left inversion breakpoint were used to verify the D53 inversion breakpoints by in situ hybridization. WT EgII is shown in c and e, D53 in d and f. Chromosomal segments are numbered, arrows in micrographs indicate in situ hybridization signal. In situ hybridizations were done at least in duplicates and at least ten nuclei were analyzed per sample, scale bar = 10 μm. All replicates led to similar results. The source data underlying Fig. 2c–f are provided as a Source Data file.

protein CG14393 in Drosophila melanogaster and contains a Major Facilitator-like superfamly domain (MFS_1, pfam07690), suggesting a general function as a metabolite transport protein. In situ hybridization on polytene chromosomes of C. capitata and Z. cucurbitae was used to confirm the presence of the wp locus in the same syntenic position on the right arm of chromosome 5 (Fig. 3d–f). Therefore, all three species show a mutation in the same positional orthologous gene likely to be responsible for the phenotype in all three genera.

Knockout of the MFS gene causes white pupae phenotypes. An analogous B. dorsalis wp mutation was developed in B. tryoni by functional knockouts of the putative Bt_wp using the CRISPR/Cas9 system. A total of 591 embryos from the Ourimbah laboratory strain were injected using two guides with recognition sites in the first coding exon of this gene (Fig. 4a). Injected embryos surviving to adulthood (n = 19, 3.2%) developed with either wild-type brown (n = 12) or somatically mosaic white-brown puparia (n = 7, Supplementary Fig. 5). Surviving G0 adults were individually backcrossed into the Ourimbah strain, resulting in potentially wp −/− (CRISPR) heterozygous brown pupae (Fig. 4c). Five independent G0 crosses were fertile (three mosaic white-brown and two brown pupae phenotypes). G1 offspring were sibling mated and visual inspection of G2 progeny revealed that three families contained white pupae individuals. Four distinct frameshift mutations were observed in screened G2 progeny (Fig. 4a) suggesting functional KO of putative Bt_wp is sufficient to produce the white pupae phenotype in B. tryoni. Capillary sequencing of cloned Bt_wp amplicons revealed deletions ranging from a total of 4–155 bp, summed across the two guide recognition sites, introducing premature stop codons.

In C. capitata, CRISPR/Cas9 gene editing was used to knockout the orthologous gene and putative Cc_wp, LOC101451947, to confirm that it causes a white puparium phenotype. A mix of recombinant Cas9 protein and the gRNA_MFS, targeting the third exon and thereby the MFS domain of the presumed Cc_wp CDS (Fig. 4b), was injected into 588 EgII WT embryos of which 96 developed to larvae and 67 pupated. All injected G0 pupae showed brown pupal color. In total, 29 G0 males and 34 females survived to adulthood (9.3%) and were backcrossed individually or in groups (see material and methods) to a strain carrying the naturally occurring white pupae mutation (wp −/− (nat); strain #1402_22m1B)23 (Fig. 4d). As white pupae is known to be monogenic and recessive in C. capitata, this complementation assay was used to reveal whether the targeted gene is responsible for the naturally occurring white pupae phenotype or if the mutation is located in a different gene. G1 offspring would only show white pupae phenotypes if Cc_wp was indeed the white pupae gene, knocked-out by the CRISPR approach, and complemented by the natural mutation through the backcross (wp −/− (nat)− (CRISPR)). In the case that the Cc_wp is not the gene...
Fig. 3 Identification of the wp mutation in the transcriptomes of *B. dorsalis*, *C. capitata*, and *Z. cucurbitae*. The gray graphs show expression profiles from the candidate wp loci in WT (wp⁺) and mutant (wp⁻) flies at the immobile pupae stages of *a* *B. dorsalis*, *b* *C. capitata*, and *c* *Z. cucurbitae*. The gene structure (not drawn to scale) is indicated below as exons (arrows labeled E1–E4) and introns (dashed lines), the Major Facilitator Superfamily (MFS) domain is shown in blue. The positions of independent wp mutations (*Bd*: 37 bp deletion, *Cc*: approximate 8150 bp insertion, *Zc*: 13 bp deletion) are marked with black dashed boxes in the expression profiles and are shown in detail below the gene models based on de novo assembly of RNAseq data from WT and white pupae phenotype individuals (nucleotide and amino acid sequences). Deletions are shown as dashes, alterations on protein level leading to premature stop codons are depicted as asterisks highlighted in black. In situ hybridization on polytene chromosomes for *d* *B. dorsalis*, *e* *C. capitata*, and *f* *Z. cucurbitae* confirmed the presence of the wp locus on the right arm of chromosome 5 in all three species (arrows in micrographs). In situ hybridizations were done at least in duplicates and at least ten nuclei were analyzed per sample, scale bar = 10 µm. The source data underlying Fig. 3d–f are provided as a Source Data file.
carrying the natural \(w_{p}^{-}\) mutation, a brown phenotype would be observed for all offspring. Here, five out of 13 crosses, namely M1, M3, F2, F3, and F4, produced white pupae phenotype offspring. The crosses generated 221, 159, 70, 40, and 52 G1 pupae, of which 10, 30, 16, 1, and 1 pupa respectively, were white. Fifty-seven flies emerged from white puparia were analyzed via non-lethal genotyping, and all of them showed mutation events within the target region. Overall, eight different mutation events were seen, including deletions ranging from 1 to 9 bp and a 46 bp deletion combined with a 2 bp insertion (Fig. 4b). Five mutation events
Fig. 4 CRISPR/Cas9-based generation of homozygous wp\(^{(CRISPR)}\) lines in B. tryonii and C. capitate. A schematic structure of the wp CDS exons (E1, E2, E3, E4) including the MFS domain in B. tryonii (a) and C. capitata (b) are shown. Positions of gRNAs targeting the first and third exon in B. tryonii and C. capitata, respectively, are indicated by green arrows. Nucleotide and amino acid sequences of mutant wp alleles identified in G1 individuals are compared to the WT reference sequence in B. tryonii (a) and C. capitata (b). Deletions are shown as dashes, alterations on protein level leading to premature stop codons are depicted as asterisks highlighted in black. Numbers on the right side represent InDel sizes (bp = base pairs). Crossing schemes to generate homozygous wp\(^{(CRISPR)}\) lines in B. tryonii (c) and C. capitata (d) show different strategies to generate wp strains. Bright-field images of empty puparia are depicted for both species. Genotype schematics and corresponding PCR analysis (for C. capitata) validating the presence of CRISPR-induced (orange) and natural (blue, for C. capitata) wp mutations are shown next to the images of the puparia. c Injected G2 B. tryonii were backcrossed to the Ourimbah laboratory strain resulting in uniformly brown G2 offspring (depicted as illustration because no images were acquired during G1). G1 inbreeding led to G2 individuals homozygous for the white pupae phenotype. d Injected WT G0 flies were crossed to flies homozygous for the naturally occurring wp\(^{−}\) allele (wp\(^{−}\)(nat)), wp\(^{−}\)(nat) (457 bp amplicon) and wp\(^{−}\)(CRISPR) or WT (724 bp amplicon) alleles were identified by multiplex PCR (left lane; L = NEB 2 log ladder). White pupae phenotypes in G1 indicated positive CRISPR events. G2 flies with a white pupae phenotype that were homozygous for the wp\(^{−}\)(CRISPR) allele were used to establish lines. PCR was done once for each individual, wp\(^{−}\)(CRISPR) alleles were verified and further analyzed via sequencing. The source data underlying Fig. 4d are provided as a Source Data file.

(B, D, E, G, H) caused frameshifts and premature stop codons. The remaining three (A, C, F), however, produced deletions of only one to three amino acids. Mutants were either inbred (mutation C) (Fig. 4d) or outcrossed to WT EgII (mutation A–H), both in groups according to their genotype. This demonstrated that C6 wp is the gene carrying the wp\(^{−}\)(nat), and that even the loss of a single amino acid without a frameshift at this position can cause the white pupae phenotype. Offspring from outcrosses of mutation A, D, and H, as well as offspring from the inbreeding (mutation C), were genotyped via PCR, and wp\(^{−}\)(CRISPR) and wp\(^{−}\)(−)(−)(CRISPR) positive flies were inbred to establish homozygous wp\(^{−}\)(CRISPR) lines.

Discussion

White pupae (wp) was first identified in C. capitata as a spontaneous mutation and was subsequently adopted as a phenotypic marker of fundamental importance for the construction of GSS for SIT\(^{6,9}\). Full penetrance expressivity and recessive inheritance rendered wp the marker of choice for GSS construction in two additional tephritid species, B. dorsalis and Z. cucurbitae\(^{1,12}\), allowing automated sex sorting based on pupal color. This was only possible because spontaneous wp mutations occur at relatively high rates either in the field or in mass rearing facilities and can easily be detected\(^{6,9}\). Despite the easy detection and establishment of wp mutants in these three species, similar mutations have not been detected in other closely or distantly related species, such as B. tryonii, B. oleae, or Anastrepha ludens, despite large screens being conducted. In addition to being a visible GSS marker used to separate males and females, the wp phenotype is also important for detecting and removing recombinants in cases where sex separation is based on a conditional lethal gene such as the ts1 gene in the medfly VIENNA 7 or VIENNA 8 GSS\(^{6,7}\). However, it took more than 20 years from the discovery and establishment of the wp mutants to the large-scale operational use of the medfly VIENNA 8 GSS for SIT applications\(^{6,9}\) and the genetic nature of the wp mutation remained unknown. The discovery of the underlying wp mutations and the availability of CRISPR/Cas genome editing would allow the fast recreation of such phenotypes and sexing strains in other insect pests. Isolation of the wp gene would also facilitate future efforts towards the identification of the closely linked ts1 gene.

Using an integrated approach consisting of genetics, cytogenetics, genomics, transcriptomics, and bioinformatics, we identified the white pupae genetic locus in three major tephritid agricultural pest species, B. dorsalis, C. capitata, and Z. cucurbitae. Our study clearly shows the power of employing different strategies for gene discovery, one of which was species hybridization. In Drosophila, hybridization of different species has played a catalytic role in the deep understanding of species boundaries and the speciation processes, including the evolution of mating behavior and gene regulation\(^{24–28}\). In our study, we took advantage of two closely related species, B. dorsalis and B. tryonii, which can produce fertile hybrids and be backcrossed for consecutive generations. This allowed the introgression of the wp mutant locus of B. dorsalis into B. tryonii, resulting in the identification of the introgressed region, including the causal wp mutation via whole-genome rescuing and advanced bioinformatic analysis.

In C. capitata, we exploited two essential pieces of evidence originating from previous genetic and cytogenetic studies: the localization of wp to region 59B and 76B on chromosome 5 in the trichogen cells and salivary gland polytene chromosome map, respectively\(^{15,29}\), and its position close to the right breakpoint of the large inversion D53\(^{8}\). This data prompted us to undertake a comparative genomic approach to identify the exact position of the right breakpoint of the D53 inversion, which would bring us in the vicinity of the wp gene. Coupled with comparative transcriptomic analysis, this strategy ensured that the analysis indeed tracked the specific wp locus on the right arm of chromosome 5, instead of any mutation in another, random locus which may participate in the pigmentation pathway and therefore result in the same phenotype. Functional characterization via CRISPR/Cas9-mediated knockout resulted in the establishment of new white pupae strains in C. capitata and B. tryonii and confirmed that this gene is responsible for the puparium’s coloration in these tephritid species. Interestingly, the wp phenotype is based on three independent and very different natural mutations of this gene, a rather large and transposon-like insertion in C. capitata, but only small deletions in the two other tephritids, B. dorsalis and Z. cucurbitae. In medfly, however, CRISPR-induced in-frame deletions of one or three amino acids in the MFS domain were sufficient to induce the wp phenotype, underlining the importance of this domain for correct coloration of the puparium.

It is worth noting that in the first stages of this study, we employed two additional approaches, which did not allow us to successfully narrow down the wp genomic region to the desired level. The first was based on Illumina sequencing of libraries produced from laser micro-dissected (Y,5) mitotic chromosomes that carry the wild-type allele of the wp gene through a translocation from the fifth chromosome to the Y. This dataset from the medfly VIENNA 7 GSS was comparatively analyzed to wild-type (Egypt II) Y and X chromosomes, and the complete genomes of Egypt II, VIENNA 7D53\(^{−}\) GSS, and a D53 inversion line in an attempt to identify the chromosomal breakpoints of the translocation and/or inversion, which are close to the wp locus (Supplementary Table 2). However, this effort was not successful due to the short Illumina reads and the lack of a high-quality reference genome. The second approach was based on individual scale whole-genome
resequencing/genotyping, and identifying fixed loci associated with pupal color phenotypes, which complemented the QTL analysis. Seven loci associated with SNPs and larger deletions linked to the white pupae phenotype were analyzed based on their respective mutations and literature searches for their potential involvement in pigmentation pathways. However, we could not identify a clear link to the pupal coloration as shown by in silico, molecular, and in situ hybridization analysis (Supplementary Figs. 6 and 7, Supplementary Table 3).

The wp gene is a member of a Major Facilitator Superfamily (MFS). Orthologs of white pupae are present in 146 of 148 insect species aggregated in OrthoDB and are single copy in 133 species. Furthermore, wp is included in the benchmarking universal single copy ortholog (BUSCO) gene set for Insecta and according to OrthoDB v10 has a below average evolutionary rate (0.87, OrthoDB group 42284aats5557) suggesting an important and evolutionarily conserved function (Supplementary Fig. 8). Its ortholog in Bombyx mori, muck, was shown to participate in the pigmentation at the larval stage, while 1D. melanogaster peak expression was during the prepupal stage after the larva had committed to pupation, which is the stage where pupal cuticle sclerotization and melanization occurs. It is known that the insect cuticle consists of chitin, proteins, lipids, and catecholamines, which act as cross-linking agents thus contributing to polymerization and the formation of the integument. Interestingly, the sclerotization and melanization pathways are connected and this explains the different mechanical properties observed in different medfly pupal color strains with the dark color cuticles being harder than the brown ones and the latter harder than the white color ones. The fact that the white pupae mutants are unable to transfer catecholamines from the hemolymph to the cuticle is perhaps an explanation for the lack of the brown pigmentation.

The discovery of the long-sought wp gene in this study and the recent discovery of the Maleness-on-the-Y (MoY) gene, which determines the male sex in several telegerrids, opens the way for the development of a generic approach for the construction of GSS for other species. Using CRISPR/Cas-based genome editing approaches, we can: (a) induce mutations in the wp orthologs of SIT target species and establish lines with wp phenotype and (b) link the rescue alleles as closely as possible to the MoY region. Given that the wp gene is present in diverse insect species including agricultural insect pests and mosquito disease vectors, this approach would allow more rapid development of GSS in SIT target species, members of diverse families, such as the agricultural pest species A. ludens, A. fraterculus, B. dorsalis, B. cor- recta, B. oleae, Drosophila suzukii, Cydia pomonella, Pectinophora gossypiella, Lobesia botrana; the livestock pests Glossina morsitans, G. pallidipes, G. palpalis gambiensis, G. austeni; and the mosquito disease vectors Aedes aegypti, Aedes albopictus, and Anopheles arabiensis. However, the biological quality of any new strain which is considered for SIT application should be first thoroughly tested in respect to their fitness and male mating competitiveness. In principle, these GSS will have higher fertility than the white color ones. The fact that the pathways are connected and this explains the different mechanistic opportunities for GSS based on visible markers.
Y chromosome. Thus, VIENNA 7 and VIENNA 8 males are heterozygous in the wp and ts loci but phenotypically wild type while VIENNA 7 and VIENNA 8 females are homozygous for both the mutant and wild type alleles. The males and females, and they die when exposed to elevated temperatures. The VIENNA 7 and VIENNA 8 GSS can be constructed with and without the D35 inversion (VIENNA 7/8D35+ or D35−). When the GSS have the inversion, females are homozygous (D35+/D35−) for D35 while males are heterozygous (D35+D35−).

To determine the position of the inversion breakpoint of C. capitata strains, high molecular weight (HMW) DNA was extracted from C. capitata lines (males and females of the WT EgII strain, the VIENNA 7D53− and 8D53−; GSS and the inversion line D53) and sequenced. Newly emerged, virgin and unfed males and females were collected from all strains. For DNA isolation and Nanopore sequencing, the HMW DNA was extracted using the phenol/chloroform Phase Lock Gel tubes (QuantBio). DNA for Illumina applications was extracted from individual flies (Supplementary table 1). Pigment de novo sequencing samples were prepared using AMPure beads (Beckman Coulter, UK) (6× volume QC check for concentration, size, integrity, and purity using Qubit (Qiagen, UK). Fragment Analyzer (Agilent Technologies) and Nanodrop (Thermo Fisher) machines. The samples were then processed without shearing using the PacBio Express kit 1 for library construction and an input of 4 µg DNA following the manufacturer’s protocol. Single- and double-stranded DNA were size-selected using the Sage Blunt-End Kit (Life Sciences). 0.75% cassette U1 marker in the range of 25–80 kb. The final library size and concentrations were obtained on the Fragment Analyzer before being sequenced using the Sequel 1.2 chemistry with V44 with a loading on plate concentration of 6 pM at 10 kb movie times. For Nanopore sequencing, the library was sequenced using the SQK-LSK109 10× Oxford Nanopore Technologies, Oxford, United Kingdom). Starting material for the ligation library preparation were 1–1.5 µg HMW gDNA for the ligation libraries and 400 ng for the rapid libraries. The prepared libraries were loaded onto FLO-PR0062 (Roche) 2× 4 flow cells. Data collection was carried out using a PromethION Beta with live high accuracy base calling for up to 72 h and with max scan intervals of 1.5 h. Each sample was sequenced at least twice. Data generated were 7.7 Gb for EgII male, 31.09 Gb for D53 male, 26.72 Gb for VIENNA 7/8D35− male, and 24.83 Gb for VIENNA 7D53− male. Run metrics are shown in Supplementary Table 4. The PacBio data were assembled using CANUv1.8 with two parameter settings: the first to avoid haplotype collapsing (genomeSize = 500 m corOutCoverage=−200 batOptions = –d 3 –b3 –d 3 r c 1 a 500 –cp 50) and the second to merge haplotypes together (genomeSize = 500 m corOutCoverage=200 correctedErrorRate=−0.15). The genome completeness was assessed with BUSCOv3.8−v3 using the dipteran gene set6. The two assemblies were found to be duplicated due to alternative haplotypes. To improve the contiguity and reduce duplication, haploMerger v20161203 was used6 and the assembly was assessed with BUSCO v3. Phase Genomics Hi-C libraries were made by Phase genomics from males (n = 2) of the same family used for PacBio sequencing. Initial scissoring was completed by Phase Genomics but edited using the Salsa57 v2.2 software (Macrogen Europe, Amsterdam). The expectation was to see a leftmost breakpoint in D53 read set alignments but not in VIENNA 7/8D53− and VIENNA 7D53−. This is the case here, since read alignments coming from both sides of the inversion are truncated at one position (Supplementary Fig. 9). Findings from genome version EgII_Ccap3.2 were extrapolated to the manually revised genome version EgII_Ccap3.2.1.

Predicted D35 inversion breakpoints were verified via PCRs on EgII, D35, and VIENNA 7/8D53− GSS male and female genomic DNA, using PhusionFlair. Polymerase in a 10 µL reaction volume [98 °C, 10 s; 35 cycles of (98 °C, 1 s; 56 °C, 90 s); 72 °C, 1 min]. Primers for the left breakpoint were designed based on IGV sequence information, primers for the right breakpoint were designed based on PacBio sequence information. The wild-type status of chromosome 5 (EgII male and female, VIENNA 7D53− male) was amplified using primer pairs P_1794 and P_1798 (150 bp) and P_1795 and P_1777 (690 bp). Chromosome 5 with the inversion (DS3 male and female, VIENNA 7D53− male and VIENNA 7D53− female) was verified using primer pairs P_1777 and P_1798 (1188 bp) and P_1795 and P_1795 (1152 bp) and amplicon sequencing (Macrogen Europe, Amsterdam).

Transcriptomic analysis of C. capitata, B. dorsalis, and Z. cucurbitae species were then conducted for RNA samples from 3rd instar larval and pre-pupal stages (Supplementary Table 1). Total RNA was extracted by homogenizing three larvae of the library insert and 1 µg was submitted for sequencing. Base calling with 100 ng of each sample was performed, and then using the NEBNext polyA selection and the Ultra II directional RNA library preparation protocols from NEB and sequenced on the Illumina NovaSeq 6000 using dual indexes as 150 bp paired end reads (library insert 500 bp). Individual libraries were sequenced to provide ≥1 million paired end reads per sample. Each replicate was then assembled separately using Trinity65 v2.8.5. The assembled transcripts from Trinity were mapped to the Capc2.3p genome using minimap266 (parameters -x splice:hq -uf). The Illumina reads were mapped with STAR60 v2.5.2.a. IGV61 v2.6 was used to view all data at a genomic and gene level. Given that the white pepper GusP1,26 was used to collect samples for RNA extraction from single larvae of Z. cucurbitae, larval sex was confirmed by a maleness-specific PCR on the MoV gene of Z. cucurbitae65 using cDNA synthesized with the OneStep RT-PCR Kit (Qiagen) and the primer pair ZcMoY1F and ZcMoY1R amplifying a 348 bp fragment. Complementary PCR reaction using the 1x Taq Master Mix kit (Qiagen) were [95 °C, 5 min; 30 cycles of (95 °C, 1 s; 51 °C, 1 s; 72 °C, 1 min); 72 °C, 10 min]. Presence of a PCR product indicated a male sample. Each, male and female sample was a pool of three individuals. Three replicates per sex and time point were collected.

Cytogenetic verification of D35 inversion and wp genes. Polytene chromosomes for in situ hybridization were prepared from third-instar larval salivary glands. In brief, the glands were dissected in 45% acetic acid and placed on a coverslip in a drop of 3:1:1 solution (3 parts glacial acetic acid: 2 parts water: 1 part lactic acid) until been transparent (approximately 5 min). The coverslip was picked up with a clean slide. After squashing, the quality of the preparation was checked by phase contrast microscope. Satisfaction preparations were left to flatten overnight at –20 °C and dipped into liquid nitrogen until the bubbling stopped. The coverslip was immediately removed with razor blade and the slides were dehydrated in absolute ethanol, air dried, and kept at room temperature. Coverslips were fastened with Permount to the salivary gland. Each sample was treated with the Extract me kit (Brlt SA), following the manufacturer’s protocol. NanoDrop spectrophotometer was used to assess the quality and quality of the extracted DNA which was then stored at –20°C until used. Primers (P_1790/P_1791, P_1821/P_1822, Fql Probe_F/R, vg1 Probe_F/R, Sd probe_F/R, y Probe_F/R, zv5 Probe_F/R, 16S probe_F/R, Zc probe_F/R, ZcF probe_F/R, ZcR probe_F/R, Zc2 probe_F/R, Zc3 probe_F/R, Zc4 probe_F/R, Zc5 probe_F/R, Zc6 probe_F/R) were designed for each targeted gene using the Geneious Prime software. PCR was performed in a 25 µL reaction volume using 12.5 µL PCR Master mix 2x Kit (Thermo Fisher Scientific), 60–80 ng DNA, and the following PCR settings [94 °C, 5 min; 35 cycles of (94 °C, 45 s; 56 °C, 30 s; 72 °C, 90 s); 72 °C, 1 min]. Primer labelling was carried out with the BigDye Terminator Ready Reaction Kit (Applied Biochemical) Roche). Prior to in situ hybridization, stored chromosome preparations were incubated with 2 µm for 2 min at 65°C. After denaturing in 0.1% NTA, 2% NaOH, dehydrating (2 min in 50%, 30%, 70%, and 95% ethanol), and air drying. Hybridization was performed on the same day by adding 15 µl of denaturated probe

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-020-20680-5 | www.nature.com/naturecommunications

9 Nature Communications [2021]12:491 | https://doi.org/10.1038/s41467-020-20680-5 | www.nature.com/naturecommunications
(boiled for 10 min and ice-chilled). Slides were covered with a siliconized coverslip, sealed with rubber cement, and incubated at 45°C overnight in a humid box. At the end of incubation, coverslips were floated off in 2× SSC and the slide washed in 2× SSC for 3× 20 min at 53°C.

After 5 min wash in Buffer 1 (100 mM tris-HCl pH 7.5/1.5 M NaCl), the preparations were in Blocking solution (Blocking reagent 0.5% in Buffer 1) for 30 min, and then washed for 1 min in Buffer 1. The antibody mix was added to each slide and incubated overnight. The slides were incubated in 2× SSC for 3× 20 min at 53°C.

in 2× SSC for 3× 20 min at 53°C.

After 5 min wash in detection buffer (100 mM Tris-HCl pH 9.5/100 mM NaCl), the color was developed with 1 ml of NBT/BCIP solution during a 40 min incubation on the slides at room temperature. The removal of the NBT/BCIP solution was done by rinsing in water twice. Hybridization sites were identified using 40× or 100× oil objectives (phase or bright field) and a Leica DM 2000 LED microscope, with reference to the salivary gland chromosome maps. Well-spread nuclei or isolated chromosomes were photographed using a digital camera (Leica DMK 5400) and the LAS X software 3.7.8. All in situ hybridizations were performed at least in duplicates and at least ten nuclei were analyzed per sample.

Gene editing and generation of homozygous wp strains. For CRISPR/Cas9 gene editing in B. tryoni, purified Cas9 protein (Alcal-R Sp. Cas9 Nuclease V3, #1081058, 10 µg/mL) and guide RNAs (customized Alcal-R CRISPR/Cas9 crRNA, 2 nmol and Alcal-R CRISPR/Cas9 tracrRNA, #1072352, 5 nmol) were obtained from Integrated DNA Technologies (IDT). The guide RNAs were individually resuspended to a 100 µM stock solution with nuclease-free duplex buffer before use. The two customized 20 bp crRNA sequences (Bt_MFS-1 and Bt_MFS-2) were designed and tested using CRISPOR. Injection mixtures for microinjection of B. tryoni embryos comprised of 300 ng/µL Cas9 protein, 59 ng/µL of each individual crRNA, 222 ng/µL tracrRNA, and 1X injection buffer (0.1 mM sodium phosphate buffer pH 6.8, 5 mM KCl) in a final volume of 10 µL. The guide RNA complex containing the two crRNAs and tracrRNA was prepared by heating at 95°C for 5 min before cooling to room temperature. The Cas9 enzyme along with the other injection mix components were then added to the guide RNA complex and incubated at room temperature for 5 min to assemble the ribonucleoprotein (RNP) complexes.

Microinjections were performed in B. tryoni Ourimbah laboratory strain embryos that were anesthetized for a 1 h time period. Embryos were performed under an oil affin oil using borosilicate capillary needles (#30-0038, Harvard Apparatus) drawn out on a Sutter P-87 microinjection/brown micropipette puller and connected to an air-filled 20 ml syringe, a manual MM-3 micromanipulator ( Narishige) and a Keyence digital microscope VHX-5000. Image processing was conducted with a Keyence digital microscope. The injection mix contained 300 ng/µL Cas9 protein (1 µg/µL, dissolved in its formulation buffer (PNA Bio Inc, CP01), 200 ng/µL RNAi, and an end-concentration of 300 mM KCl. The mix was freshly prepared on ice followed by an incubation step for 10 min in 37°C to allow pre-assembly of the gRNA-Cas9 RNP complexes and stored on ice until use. Microinjections were conducted in WT EgII C. capitata embryos, collected over a 30–40 min period, chemically dechorionated (sodium hypochlorite, 3 min), fixed on double-sided sticky tape (Scotch 3M), and covered with halocarbon oil 700 (Sigma-Aldrich). Injections, siliconized glass needle tips (Q100-70-7.5, LOT171381; Science Products, Germany), drawn out on a laser-based micropipette puller (Sutter P-2000), were used with a manual micromanipulator (MN-151, Narishige). A dispense needle (tip diameter: 100 µm) was used. The microinjection mix was then added to the pipette tip and injected into the embryo. The embryos were incubated in an oxygen chamber (max. 2 psi), first instar larvae were transferred from the oil to larval food.

As complementation assay, reciprocal crosses between surviving G0 adults and virgin adults of the Ourimbah laboratory strain embryos, collected over a 30 min period, were performed for 10 min at 37°C to allow pre-assembly of gRNA-Cas9 RNP complexes and stored on ice until use. Microinjections were performed in the third CDS exon of CcMFS according to color of pupae (brown, mosaic, or white). C. capitata, non-lethal genotyping. Offspring of outcross cages showed brown pupae phenotype and either wp | CRISPR genotype. In order to make mutations A, D, and H homozygous, 40 flies (25 females, 15 males) were genotyped each, and wp CRISPR | CRISPR flies produced only white pupae offspring, based on either the wp wp | CRISPR or CRISPR wp | CRISPR genotype. 46 flies (46 females, 48 males) were genotyped, homozygous wp | CRISPR were inbred to establish a line (13 females, 8 males).

Molecular analyses of wp mutants and mosaics. In B. tryoni, genomic DNA was isolated for genotyping from G0 pupae using the DNeasy Blood and Tissue Kit (Qiagen). PCR amplicons spanning both BtMSF guide region sites were generated using Q5 polymerase (NEB) with primers BtMSF_Spinife and BtMSF_exon2R. Products were purified using MiniElute Purification Kit (Qiagen), ligated into pGEM-T-easy vector (Promega) and transformed into DH5α cells. Plasmids were purified with Wizard Plus SV Minipreps (Promega) and sequenced.

In C. capitata, non-lethal genotyping was performed to identify parental genotypes before setting up crosses. Therefore, genomic DNA was extracted from siblings of G0 flies and G2 flies of each crRNA/strand pair to establish an established protocol. Single legs of anesthetized flies were cut at the proximal femur, placed in vials containing ceramic beads and 50 µL buffer (10 mM Tris-Cl, pH 8.2, 1 mM EDTA, 25 mM NaCl), and homogenized for 15 s (6 min/s) using a FastPrep-24 S 5 homogenizer. Then, 28.3 µL buffer and 1.7 µL proteinase-K (2.5 U/mg) were added. The reaction mix was incubated for 1 h at 37°C, followed by 4 min at 98°C, and subsequently cooled down on ice and used for PCR. For G1 flies, PCR on wp was performed in a 25 µL reaction volume using the DreamTaq polymerase, primers P_1643 and P_1644, and 3.75 µL reaction mix, whereby different amplicon sizes were expected for different alleles (wp wp | CRISPR). 724 bp, wp wp | CRISPR, amplified via PCR settings [95°C, 3 min; 35 cycles of (95°C, 30 s; 56°C, 30 s; 72°C, 1 min); 72°C, 5 min]. The 724 bp PCR product was verified by gel electrophoresis and purified from the PCR reaction using the DNA Clean & Concentrator-5 kit. PCR products were sequenced (P_1644) and analyzed using Geneious Prime. In generation G2 flies, alleles were analyzed using multiplex PCR with primers P_1657, P_1643, and P_1644, to distinguish between the wp wp | CRISPR alleles (724 bp; P_1643/ P_1657) and wp wp | CRISPR alleles (724 bp; P_1657/ P_1644) using the above-described PCR protocol.

Image acquisition. Images of B. tryoni pupae were taken with an Olympus SZX6 microscope, Olympus DP74 camera, and Olympus LF-PS2 light source using the Olympus stream basic 2.3.3 software. Images of C. capitata pupae were taken with a Creative Standart digital microscope BX60. Image processing was conducted and analyzed with Adobe Photoshop CS5.1 software to apply moderate changes to image brightness and contrast. Changes were applied across the entire image.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this article is available as a Supplementary Information file. The datasets and insect strains generated and analyzed during the current study are available from the corresponding authors upon request. All sequences generated in this study for B. dorsalis, B. tryoni, Bactrocera introgression line (BIL), C. capitata and Z. cucurbitae samples are publicly available on NCBI within the EMB BioProject PRJEB36344 (for Ccap genome assembly EgII-3.2.1, WGS, PacBio, chromosome dissections, Illumina MiSeq, Illumina HiSeq 4000, RNAseq, Illumina NovaSeq 6000, Hi-C, and Nanopore data; see Supplementary Table 1 for detailed sample designation), BioProject PRJNA629430 (for WGS and Illumina DNAseq 2 × 250 PE data; see Supplementary Fig. 6 for detailed sample designation), and BioProject PRJNA682907 (for WGS and Illumina NovaSeq 6000 data; see Supplementary Table 1 for detailed sample designation). The source data underlying Figs. 1, 2e, 3a, 3d–f, 4b, and 4d, as well as Supplementary Figs. 3a–b, 4a–b, 4d, and 7 are provided as a Source Data file. Source data are available in this paper.

Received: 16 June 2020; Accepted: 14 December 2020; Published online: 21 January 2021
Acknowledgements

This study was financially supported by the Joint FAO/IAEA Insect Pest Control Subprogramme of the Joint FAO/IAEA Programme of Nuclear Techniques in Food and Agriculture, the Emmy Noether program of the German Research Foundation (SCHE 1833/1-1; to MFS), the LOEWE Center for Insect Biotechnology and Bioresources of the Hessen State Ministry for Higher Education, Research and the Arts (HMWK; to MFS), and the STPlus collaborative fruit fly program funded by the Hort Frontiers Fruit Fly Fund, part of the Hort Frontiers strategic partnership initiative developed by Hort Innovation, with co-investment from Macquarie University, USDA, and JLU Gießen and contributions from the Australian Government (FF17000 to MFS). The project was further supported by Hort Innovation, using the Apple & Pear, Strawberry, Citrus, Cherry, Summerfruit, Table Grape and Vegetable research and development levies (MTI1659 to PC, AC, EF), with co-investment from South Australian Research and Development Institute (SARDI) and Primary Industries and Regions South Australia (PIRSA) and contributions from the Australian Government. Hort Innovation is the grower-owned, not-for-profit research and development corporation for Australian horticulture. SWB was supported by the Australian Research Council (FT140101303) and the Hermon Slade Foundation grant HSF 18/6. Furthermore, this work was supported by the Canadian Foundation for Innovation (33408, to JR) and Genome Canada Genome Technology Platform awards (JR), as well as the International Atomic Energy Agency research contracts no. 23358 (JR) and no. 23379 (FM) as part of the Coordinated Research Project “Genetic approach for the development of genetic sexing strains for SIT applications”. The project benefitted from discussions at this CRP. Furthermore, resources were provided by the SCiNet project of the USDA-ARS, the Detection, Control, and Area-wide Management of Fruit Flies and Other Quarantine Pests of Tropical/Subtropical Crops, and a USDA-NIFA grant (0500-00093-001-00-D, 2040-22430-026-00-D, and 2017-67012-26087 to SBS and SMG). The USDA-ARS is an equal opportunity/affirmative action employer, and all agency services are available without discrimination. Mention of commercial products and organizations in this manuscript is solely to provide specific information. It does not constitute endorsement by USDA-ARS over other products and organizations not mentioned. The authors also wish to thank Tanja Rehling, Jakob Martin, and Johanna Rühl for technical assistance and Germano Solazzo for input on injections and primers design (Justus-Liebig University Gießen and Insect Pest Control Laboratory); Elena Isabel Cancio Martínez, Thilakasiri Dammalage, Sohel Ahmad, and Gültüar Pillwax for insect rearing (Insect Pest Control Laboratory), Shu-Huang Chen (McGill University) for technical assistance with Nanopore library preparations; and Arjen van’t Hof (University of Liverpool) for constructing libraries from micro-dissected chromosomes.

Author contributions

R.A.A., C.M.W., C.C., P.C., S.B.S., S.M.G., I.H., J.R., A.C.D., K.B., S.W.B., and M.F.S. designed the research; C.M.W., R.A.A., M.A.W., K.N., G.G., E.F., S.J.R., M.A.H., C.C., T.N.M.N., A.C., S.B.S., S.M.G., A.C.D., K.B., S.W.B., and M.F.S. performed the research; R.A.A., C.M.W., H.D., G.L., F.M., J.R., K.B., S.W.B., and M.F.S. contributed new reagents/analytic tools; C.M.W., R.A.A., M.A.W., K.N., G.L., G.G., H.D., S.W., T.N.M.N., A.C., S.B.S., S.M.G., I.H., J.R., A.C.D., K.B., S.W.B., and M.F.S. analyzed the data; R.A.A., C.M.W., K.N., G.L., G.G., S.J.R., S.W., A.C., S.B.S., S.M.G., I.H., J.R., A.C.D., K.B., S.W.B., and M.F.S. wrote the paper.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-20680-5.

Correspondence and requests for materials should be addressed to K.B., S.W.B. or M.F.S.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021