Insect Biochemistry and Molecular Biology 153 (2023) 103896

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

Differential regulation of the *Tor* gene homolog drives the red/green pigmentation phenotype in the aphid *Myzus persicae*

Nasser Trissi [a], Bartlomiej J. Troczka [a,1], Luke Ozsanlav-Harris [a,1], Kumar Saurabh Singh [a], Mark Mallott [a], Veenu Aishwarya [b], Andrias O'Reilly [c], Chris Bass [a,**], Craig S. Wilding [c,*]

[a] College of Life and Environmental Sciences, Biosciences, University of Exeter, Penryn Campus, Penryn, Cornwall, UK
[b] AUM LifeTech, Inc., Philadelphia, PA, 19104, USA
[c] School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK

ARTICLE INFO

Keywords:
Carotene
Carotenoid
Aphid
Colour polymorphism
Gene expression

ABSTRACT

In some aphid species, intraspecific variation in body colour is caused by differential carotenoid content: whilst green aphids contain only yellow carotenoids (β-, γ-, and β,γ-carotenes), red aphids additionally possess red carotenoids (torulene and 3,4-didehydrolycopene). Unusually, within animals who typically obtain carotenoids from their diet, ancestral horizontal gene transfer of carotenoid biosynthetic genes from fungi (followed by gene duplication), have imbued aphids with the intrinsic gene repertoire necessary to biosynthesise carotenoids. In the pea aphid, *Acyrthosiphon pisum* a lycopene (phytoene) desaturase gene (*Tor*) underpins the red/green phenotype, with this locus present in heterozygous form in red individuals but absent in green aphids, resulting in them being unable to convert lycopene into the red compounds 3,4-didehydrolycopene and torulene. The green peach aphid, *Myzus persicae*, separated from the pea aphid for ≈45MY also exists as distinct colour variable morphs, with both red and green individuals present. Here, we examined genomic data for both red and green morphs of *M. persicae* and identified an enlarged (compared to *A. pisum*) repertoire of 16 carotenoid biosynthetic genes (11 carotenoid desaturases and five carotenoid cyclase/synthase genes). From these, we identify the homolog of *A. pisum Tor* (here called *carotene desaturase 2* or *CDE-2*) and show through 3D modelling that this homolog can accommodate the torulene precursor lycopene and, through RNA knockdown feeding experiments, demonstrate that disabling *CDE-2* expression in red *M. persicae* clones results in green-coloured offspring. Unlike in *A. pisum*, we show that functional *CDE-2* is present in the genomes of both red and green aphids. However, expression differences between the two colour morphs (350–700 fold *CDE-2* overexpression in red clones), potentially driven by variants identified in upstream putative regulatory elements, underpin this phenotype. Thus, whilst aphids have a common origin of their carotenoid biosynthetic pathway, two aphid species separated for over 40MY have evolved very different drivers of intraspecific colour variation.

1. Introduction

Carotenoids are a diverse group of natural tetraterpenes produced by photosynthetic plants and algae, as well as some bacteria, archaea and fungi, and are the most common pigment compounds found in nature (Maoka, 2020). They are essential compounds in photosynthetic organisms due to their photoreceptive and antioxidant properties enabling collection and transfer of light energy to chlorophylls through singlet-singlet excitation transfer. Carotenoids include two distinct groups: the hydrocarbon carotenes and the more chemically complex xanthophylls, with the latter being the more numerous group with over 800 individual compounds identified to date (Maoka, 2020).

Most animals are unable to synthesize their own carotenoid compounds *de novo* and instead acquire them through dietary intake (Maoka, 2020). In the animal kingdom dietary carotenoids provide a variety of benefits including enhanced immune responses (Chew and Park, 2004), protection from oxidative stress (Fiedor and Burda, 2014; Miki, 1991) and synthesis of visual pigments (Cazzonelli, 2011; von

* Corresponding author.
** Corresponding author.
E-mail addresses: c.bass@exeter.ac.uk (C. Bass), c.s.wilding@ljmu.ac.uk (C.S. Wilding).
[1] Authors contributed equally.

https://doi.org/10.1016/j.ibmb.2022.103896
Received 28 October 2022; Received in revised form 22 December 2022; Accepted 22 December 2022
Available online 29 December 2022

Lintig, 2012). In some organisms carotenoids also play important roles in visual signalling, sexual signalling, warning colouration and crypsis (Maoka, 2020).

Aphids (superfamily: *Aphidoidea*) often exhibit distinct colour polymorphisms with green, red or pink individuals present. Although under the same conditions body colouration remains stable in the parthenogenetically reproducing female clones (Moran and Jarvik, 2010) it can be influenced by various biotic and abiotic factors including temperature (Caillaud and Losey, 2010), light intensity (Alkhedir et al., 2010), endosymbiotic bacteria (Tsuchida et al., 2010, 2014) and nutrient depravation (Wang et al., 2019). In the well-studied pea aphid *Acyrthosiphon pisum* body colouration is thought to provide protection against some natural enemies, although this effect is countered by increased susceptibility to others (Losey et al., 1997) and, more recently, the effect of selection pressure is thought to be minor (Ives et al., 2020). There is also evidence that in this aphid, host plant preference affects colour morph prevalence, with pink aphids predominately found on pea (*Pisum sativum*) and at lower frequencies on red clover (*Trifolium pratense*) (Simon et al., 2003). Red coloured *A. pisum* have been shown to be more likely to drop off plants and produce a higher proportion of winged offspring (Weisser and Braendle, 2001).

Aphid colouration in *A. pisum* is associated with the type of carotenoid compounds present; green *A. pisum* contain only yellow carotenoids (β,β-carotene (β,β-carotene), γ-carotene (β,ψ-carotene), and β,γ-carotene) whereas red aphids possess red carotenoids (torulene (3′,4′-didehydro-β,ψ-carotene) and 3,4-didehydrolycopene (3,4-didehydro-ψ,ψ-carotene) in addition to β-, γ-, and β,γ–carotenes (Moran and Jarvik, 2010).

The biosynthesis of carotenes involves a cascade reaction series (Takemura et al., 2021), the first step of which involves the condensation of geranylgeranyl pyrophosphate into phytoene by phytoene synthase following which phytoene is desaturated to produce lycopene. Subsequent enzymatic steps can then convert lycopene into the red-coloured 3,4-didehydrolycopene and torulene, or the yellow γ,γ-carotene, γ,ψ-carotene or β,γ-carotene.

Moran and Jarvik (2010) demonstrated that unlike most members of the animal kingdom, the pea aphid is capable of *de novo* synthesis of carotenoid compounds with aphid colouration linked to the activity of carotenoid biosynthetic genes gained through horizontal gene transfer from fungi, which have been integrated into the aphid genome and subsequently duplicated (Fukatsu, 2010; Moran and Jarvik, 2010). Through functional analysis, Takemura et al. (2021) has since elucidated the full biosynthetic pathway, demonstrating that just four enzymes are needed to complete this pathway in *A. pisum*.

Interrogation of genomic sequences showed the molecular basis of red-green colour polymorphism in the pea aphid is linked to presence of a heterozygous locus, *Tor*+, approximately 30 Kb in size and found only in red morphs (Moran and Jarvik, 2010). This allele contains a copy of one of the four genes identified by Takemura et al. (2021), a carotenoid desaturase gene, crucial for the production of torulene (although this carotenoid desaturase actually converts lycopene to 3,4-didehydrolycopene which is then converted to torulene by a separate carotenoid cyclase (Takemura et al., 2021)). It has been conclusively proven that green *A. pisum* completely lack *Tor*+ and so are not capable of producing torulene (Moran and Jarvik, 2010). A chromosome level genome assembly of *A. pisum* and cytogenetic experiments placed the location of all carotenoid metabolism genes into a single locus located on a sub-terminal portion of autosome one (Li et al., 2019; Mandrioli et al., 2016). With the expansion of genomic data from other aphids, carotenoid genes have also been found in other members of the aphid family (Nováková and Moran, 2011) as well as phylloxerids (Zhao and Nabity, 2017), gall midges (Cobbs et al., 2013) and spider mites (Altincicek et al., 2012).

In the current study we studied the green peach aphid *Myzus persicae*, a ubiquitous insect pest, capable of feeding on more than 400 host species over the world, with the ability to shift to novel host plants and with high insecticide resistance propensity (Singh et al., 2020). Like the pea aphid, *M. persicae* clones display a wide array of colour polymorphism ranging from green to brownish red. Previous studies have demonstrated that body colouration in *M. persicae* is under the control of a single locus with red dominant to green (Takada, 1981). Nováková and Moran (2011) showed that, in common with other aphids, carotene desaturases necessary for carotenoid biosynthesis are also found in *Myzus persicae*. However, *Myzus persicae* and *Acyrthosiphon pisum* last shared a common ancestor 45MYA (CI: 42.5–48.0 MYA) (Kumar et al., 2017), and thus whether in *M. persicae* it is the same genomic deletion of the *Tor* homolog that causes green colouration, or some other mechanism, is unknown. Although a number of horizontally transferred carotenoid genes have been identified *in M. persicae* (Nováková and Moran, 2011) the emergence of high quality chromosome level assemblies for this species (Mathers et al., 2017; Singh et al., 2021) now allow for a more detailed characterization of the genomic locus involved in carotenoid production.

In this paper we report a detailed characterization of the carotenoid gene locus in *M. persicae*, identify the *M. persicae* ortholog of the *Tor* carotenoid desaturase gene, and propose that unlike in *A. pisum* where it is absence of *Tor*+ in green morphs which underlies the colour phenotype, in *M. persicae*, both colour morphs carry the *Tor* + ortholog and it is differential regulation which mediates pigmentation in *M. persicae*. Upstream variants putatively driving this gene expression are identified.

2. Methods

2.1. Aphid clones

Red coloured *M. persicae* clones were collected from tobacco in Italy (124, 57), Greece (5191A), and Zimbabwe (5410R). Green *M. persicae* clones were collected from Italy (1X) from peach, from the UK (4106A, US1L) from potato and sugar beet, and from Germany (NS) from an unknown weed host plant. Clones were reared asexually on individual Chinese cabbage leaves (*Brassica napus* L var. chinensis cv. Tip-Top) in small plastic cups maintained at 18 °C under a 16:8-h light:dark regime for at least 1 year before molecular analyses.

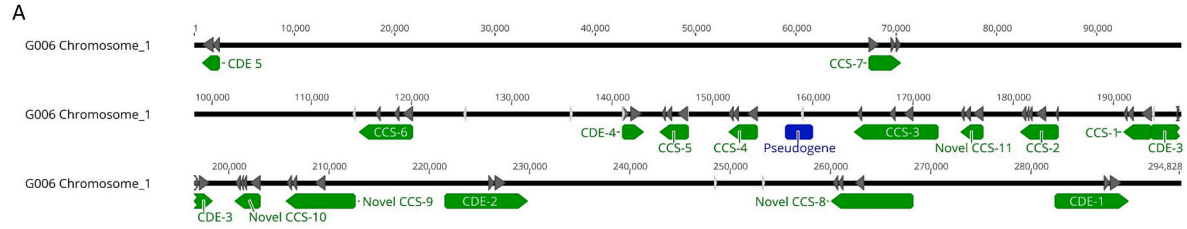
2.2. DNA and RNAseq and in silico data mining

We leveraged a recently published chromosome level assembly of the G006 clone (green in colour) of *M. persicae* (Singh et al., 2021) for data mining and expression analyses. BLAST within Geneious 10.2.3 (Biomatters, NZ) was used to identify carotenoid biosynthesis genes in *M. persicae* using sequences from *A. pisum* (Moran and Jarvik, 2010) as queries. All novel *M. persicae* carotenoid biosynthesis gene sequences identified were verified by reciprocal BLAST against the NCBI nucleotide (nt) database. DNA scaffolds containing genes of interest were extracted from AphidBase (Gauthier et al., 2007) (*M. persicae* clone G006) and manually curated using Geneious with publicly available DNA and RNA short read sequencing data (Singh et al., 2020). Identification of the ortholog of *Tor* (Gene ID:100169245) in *M. persicae* was undertaken through phylogenetic analysis of sequences using MEGAX (Kumar et al., 2018) with tree construction undertaken using the most appropriate model as determined through the MODEL function (LG + G for the CDE tree and JTT + G for the CCS tree). Gene order was also determined in a representative red aphid clone 5410R by leveraging long-single molecule sequencing (PacBio) reads that had been generated in a previous study (Singh et al., 2021).

Identification of colour-specific promoter and UTR modifications of the *Tor* ortholog was undertaken through alignment of upstream sequences from both the G006 and 5410R clones in Geneious.

Estimation of expression level was undertaken through alignment of publicly available RNAseq data from green (NS, 1X, 926B, 4106a, US1L, 23, 5410G) and red clones (5410R, 5191a, 124, JR, 124) to the appropriate genomic scaffold with calculation of transcripts per million (TPM)

A



B

	A. pisum Tor Gene	CDE-4	CDE-3	CDE-2	CDE-1	CDE-5
***A. pisum* Tor Gene**		74.053	86.7424	91.3208	83.5227	89.0566
CDE-4	74.053		78.4091	73.4848	73.4848	72.3485
CDE-3	86.7424	78.4091		87.6894	89.3939	86.7424
CDE-2	91.3208	73.4848	87.6894		84.2803	93.9623
CDE-1	83.5227	73.4848	89.3939	84.2803		83.5227
CDE-5	89.0566	72.3485	86.7424	93.9623	83.5227	

Fig. 1. (**A**) Pictorial representation of clone G006 chromosome 1 (black line) illustrating all 16 carotenoid biosynthesis genes identified in *M. persicae*. Genes are depicted in green below the chromosome. Exons are shown in dark grey on the chromosomes with pseudogenes in blue. CDE: carotenoid desaturase, CCS: carotenoid synthase/cyclase. (**B**) Matrix of amino acid identities between the five carotenoid desaturase genes identified in *M. persicae* and the *Tor* gene of *A. pisum*.

values for each gene conducted in Geneious.

2.3. RNA knockdown validation

In order to functionally validate orthologs putatively involved in carotenoid biosynthesis, RNA knockdown feeding experiments were conducted. RNA knockdown assays were designed by AUM Lifetech, USA, using 2′-deoxy-2′-fluoro-β-D-arabinonucleic acid anti-sense oligonucleotides (FANAs), synthetic single-stranded nucleic acid analogs that can modulate gene expression by enzymatic degradation of a target RNA (Kalota et al., 2006). Sequences of oligos can be found in Supplementary Table S1. FANA stocks were dissolved to 1 mM in nuclease free water. Synchronized fourth instar nymphs were placed in plastic cups (25 mm in diameter) layered with a parafilm pouch containing 140 μl of complete artificial diet (adapted from Paliwal (2017)) and either 10 μl of 1 mM gene specific FANA oligos, 10 μl of 1 mM scrambled FANA (a FANA oligo designed with a scrambled (random) sequence not targeting any aphid gene) or just diet. Each treatment consisted of 4 replicates (N = 15 insects per replicate). Aphids were kept for seven days at 20 °C under a 16:8h light: dark regime after which they were visually inspected for colour change with surviving insects snap frozen in liquid nitrogen.

2.4. RNA extractions and quantitative real-time PCR

Aphid clones were age synchronized under identical environmental conditions and 4 biological replicates, each comprising 10 apterous females, were used for RNA extraction using the ISOLATE II RNA Mini Kit (Bioline, UK) following the manufacturer's instructions. Maxima H Minus First Strand cDNA Synthesis Kit (ThermoFisher Scientific, USA) with 1 μg total RNA was used for cDNA synthesis following the manufacturer's instructions. Quantitative PCR was used to examine the expression of candidate genes found to be differentially expressed by RNA-Seq (*CDE-2*, *CCS-3* and *CCS-9*) (primers shown in Supplementary Table S1) with SYBR® Green JumpStart™ Taq ReadyMix™ (SimgaAldrich, USA) as described previously (Bass et al., 2013). Data were analysed according to the $\Delta\Delta C_T$ method (Pfaffl, 2001) using the geometric mean of two housekeeping genes (*actin* and *para*) for normalization according to the strategy described previously (Vandesompele et al., 2002).

2.5. 3D modelling of the Tor homolog

Molecular models of the *A. pisum* Tor enzyme and the *M. persicae* CDE-2 homolog were generated using Alphafold2 (Jumper et al., 2021) as described previously (Duarte et al., 2022). The primary sequence of each protein was submitted to the ColabFold web interface (Mirdita et al., 2022) and AlphaFold2 was run using default options. For both *A. pisum* and *M. persicae* enzymes, the top-ranking model was selected and the thirty amino acids of the C-terminus were removed due to their low local Distance Difference Test (lDDT) scores. The 'iterative magic fit' function of SwissPDBViewer (Guex et al., 1999) was used to superimpose *A. pisum* and *M. persicae* model structures and to calculate the resulting root mean square deviation (RMSD).

Docking predictions for lycopene with either *A. pisum* Tor or *M. persicae* CDE-2 models were generated as described (Shi et al., 2020). A 3-dimensional structure of lycopene was generated using MarvinSketch (v19.22) of the ChemAxon suite (http://www.chemaxon.com). AutoDockTools (v1.5.7) (Molecular Graphics Laboratory, Scripps Research Institute, La Jolla, CA, USA) was used to merge the non-polar hydrogens in the models and to define rotatable bonds in lycopene. Automated ligand docking studies were performed using Auto-Dock Vina (version 1.1.2) (Trott and Olson, 2010) using a grid of 60 × 60 × 60 points (1 Å spacing), which encompassed the entire model. A figure was produced using PyMOL (DeLano Scientific, San Carlos, CA, USA).

2.6. Dual luciferase expression analysis of upstream regulatory variants

Quantification of the ability of haplotypes upstream of the *Tor* ortholog to drive differential expression was attempted using the dual luciferase assay system. In order to disentangle the putative regulatory effects of the GC indel from the direct-repeat region, multiple constructs were synthesised to represent the multiple combinations of the GC indel, and the presence of 1 or 2 direct repeats. However, unfortunately it did not prove possible to successfully produce a plasmid with 2 repeats together, likely due to the large total insert size when both repeats were present (over 11 Kb) and consequent difficulties transforming *E. coli*. Synthesised sequences were subcloned into pGL3-Basic (Promega) and transformed into Library Efficiency DH5α Competent Cells (Invitrogen). Plasmids were extracted with the GeneJet plasmid miniprep kit (Fermentas), sequenced and then adjusted to 400 ng/μl for use in dual luciferase assays using the Sf9 insect cell line. Approximately 1×10^6 cells per well were plated into 6-well plates 2 h prior to transfection and allowed to reach 60%–70% confluency. Insect GeneJuice Transfection Reagent (Novagen) was used for transfection of constructs and the Dual-Luciferase Reporter Assay (Promega) used for promoter activity measurements according to the manufacturer's protocols. 2 μg of reporter

Table 1
List of annotated carotenoid genes in *Myzus persicae*. Function based upon Interpro domain (Blum et al., 2020).

Name	NCBI accession no.	Function	Exons
CCS-7	XM_022317511	Squalene/phytoene synthase (IPR002060) Trans-isoprenyl diphosphate synthases, bacterial-type (IPR044843)	3
CCS-6	XM_022317500	Squalene/phytoene synthase (IPR002060) Trans-isoprenyl diphosphate synthases, bacterial-type (IPR044843)	3
CDE-4	XM_022317507	Carotenoid/retinoid oxidoreductase (IPR014105)	2
CSS-5	XM_022317503	Squalene/phytoene synthase (IPR002060)	3
CCS-4	XM_022317509	Squalene/phytoene synthase (IPR002060)	3
CCS-3	XM_022317504	Squalene/phytoene synthase (IPR002060)	3
CCS-2	XM_022317496	Squalene/phytoene synthase (IPR002060)	5
CCS-1	OP714165	Squalene/phytoene synthase (IPR002060)	4
CDE-3	XM_022317505	Carotenoid/retinoid oxidoreductase (IPR014105)	2
CDE-2	XM_022317517	Carotenoid/retinoid oxidoreductase (IPR014105)	2
CDE-1	XM_022317513	Carotenoid/retinoid oxidoreductase (IPR014105)	2
CDE-5	OP714166	Carotenoid/retinoid oxidoreductase (IPR014105)	2
CCS-8 (N)	XM_022317487	Squalene/phytoene synthase (IPR002060)	3
CCS-9 (N)	XM_022317518	Squalene/phytoene synthase (IPR002060) Trans-isoprenyl diphosphate synthases, bacterial-type (IPR044843)	3
CCS-10 (N)	OP714167	Squalene/phytoene synthase (IPR002060)	4
CCS-11 (N)	XM_022317501	Squalene/phytoene synthase (IPR002060)	3

constructs and pGL3 without insert (as a control) was co-transfected with 4 ng Renilla luciferase pRL-CMV using GeneJuice and incubated at 27 °C. 4 hr post-transfection, the transfection mixture was removed and replaced with supplemented Grace's Insect Medium (GIBCO). Following further incubation at 27 °C for 48 h and washing of cells with PBS, cells from each well were harvested in 500 μL passive lysis buffer (Promega) and luciferase activity measured on a GloMax 20/20 (Promega). Construct luciferase activity was normalized to Renilla luciferase activity as instructed in the manufacturer's protocol.

3. Results and discussion

3.1. Mapping and identification of carotenoid biosynthesis genes

Interrogation of publicly available genomic data from the chromosome level *M. persicae* clone G006 reference assembly (Singh et al., 2021) yielded sequences for five carotenoid desaturase genes (CDE) and seven carotenoid cyclase/synthase genes (CCS) (Fig. 1; Table 1). All identified sequences mapped to a single approx. 300 kb locus on autosome 1, further confirming recent cytogenetic reports on the carotenoid locus being linked with the sub-terminal portion of autosome 1 (Mandrioli et al., 2016). Further manual curation of the genomic data within the ~300 Kb section, where orthologs to all previously known carotenoid biosynthesis genes are located, yielded four additional carotenoid cyclase/synthase genes that were not automatically predicted (Fig. 1). These genes are referred to here as Novel CCS-8, Novel CCS-9, Novel CCS-10 and Novel CCS-11. This takes the total carotenoid biosynthesis genes found in *M. persicae* to 16 (11 carotenoid cyclase/synthase and 5 carotenoid desaturase genes), over double the number of genes annotated in *A. pisum* (Moran and Jarvik, 2010). However, we do note that this genomic analysis did not recover a desaturase gene equivalent to *Myzus persicae copy C* of Nováková and Moran (2011) and therefore there may be additional sequences in the genome. One region between CCS-3 and CCS-4 (annotated in blue in Fig. 1A) yielded BLAST hits (>50%) when carotenoid cyclase/synthase genes were used as queries in the initial search but did not contain any ORFs greater than 150bp. This section most likely represents the locus of a pseudogene that has accrued mutations and lost its function due to interruptions in the ORF.

3.2. CDE-2 is the ortholog of Tor and is found in both red and green M. persicae genome assemblies

Comparison of *M. persicae* CDE protein sequences showed that CDE-2 and CDE-5 form a clade with *A. pisum Tor* (Supplementary Figure 1). Whilst CDE-2 and CDE-5 show the highest identity (93.96%) and similarity (97.17%) of any gene pair, CDE-2 is closer (91.32% identity) to *Tor* than CDE-5 (89.06% identity) (Fig. 1B) indicating CDE-2 is most likely the ortholog of the *A. pisum Tor* gene. When aligned, gene architectures of *Tor* and CDE-2 were also very similar with a large intron in the 5′ UTR, two exons separated by a short intron and a large 3' UTR. Identical gene architectures were seen in both green (G006) and red (5410R) clones. However, despite their similarities at the amino acid level CDE-2 and CDE-5 are not neighbouring, tandemly repeated genes, instead being separated by over 220 kb (Fig. 1A).

3.3. CDE-2 undergoes differential gene expression between red and green morphs of M. persicae

Using publicly available RNAseq data for seven green and five red morphs we mapped raw reads to the G006 assembly and calculated average expression of carotenoid biosynthesis genes. Only 3 out of 16 genes were shown to be differentially expressed between red and green morphs: *CDE-2*, *CCS-3* and *CCS-2* (Fig. 2A), with the starkest change in expression for *CDE-2,* where red clones show approximately 100-fold higher expression than green morphs. We verified TPM calculations using qPCRs for these three differentially expressed genes (*CDE-2*, *CCS-3* and *CCS-9*) (Fig. 2B, C and D). In the case of *CDE-2* the high level of expression in red clones (5410R and 57) was clearly confirmed, showing approximately 350-700-fold overexpression when compared to 3 tested green clones (4106a, IX and US1L). *CCS*-3, whilst differentially expressed from RNASeq data, is actually up-regulated in the green morphs and differential expression in individual strains was not supported by qPCR (Fig. 2C). Whilst *CCS-9* is strongly differentially expressed in strain 57 and is upregulated in the other red strain studied (5410R) it is also upregulated in one of the green strains (1X) so the signal from this gene is not consistent. Since *CDE-2* is the likely homolog of *A. pisum Tor* its differential expression between red and green morphs provides strong evidence that this gene has a pivotal role in the *M. persicae* colour phenotype. Further investigation of the expression profile of *CDE-2* in red clones across different life stages and tissues showed that *CDE-2* is most highly expressed in bacteriocytes and during late-stage nymph development (Fig. 3), however, the change in expression level is relatively modest (approximately 3-fold higher) through the lifespan of the insect. It has been shown that the level of bacteriocytes reaches its peak in late-stage nymphs (Vogel and Moran, 2011) and that densities of the *Buchnera* endosymbiont are affected by host age, genotype and environmental factors with numbers increasing over the juvenile stages then beginning to drop after the aphid becomes an adult (Komaki and Ishikawa, 2000).

3.4. Functional validation of CDE-2

In order to functionally verify if high expression of *CDE*-2 has a

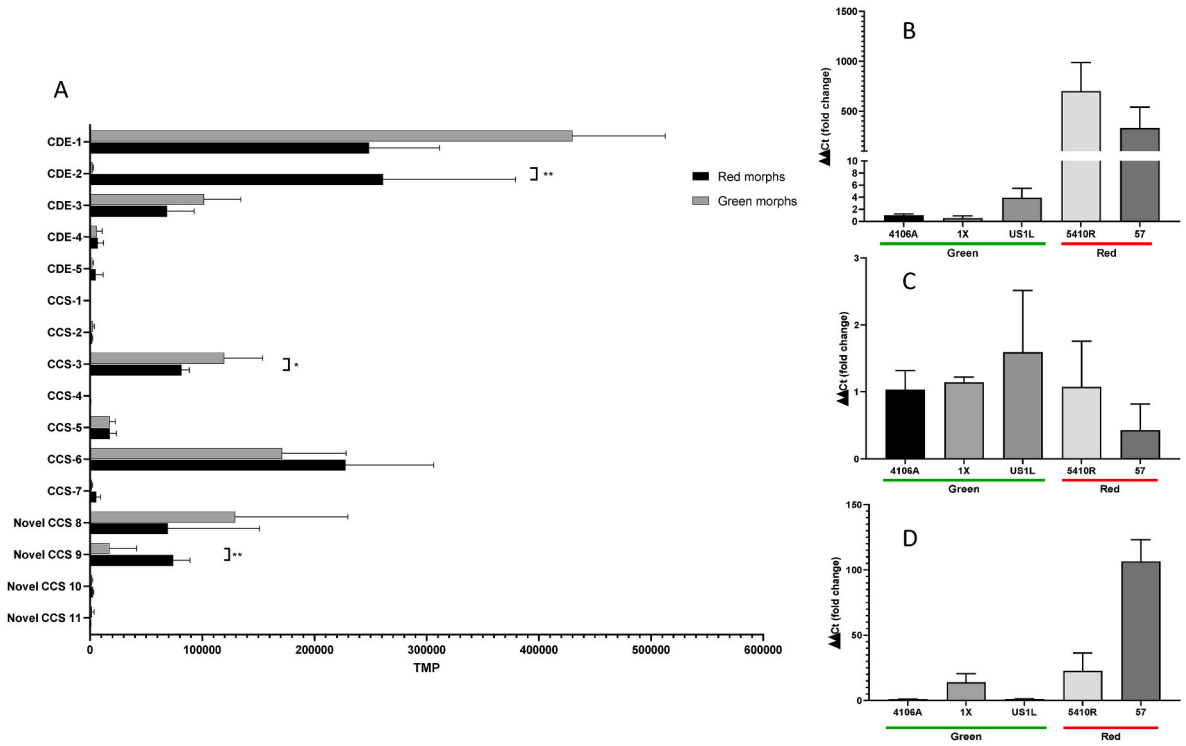


Fig. 2. (**A**) Average expression levels of all carotenoid desaturase (CDE) and carotenoid cyclase/synthase (CCS) in red (N = 5) and green morphs (N = 7) of *M. persicae*. Values shown are Transcripts Per Million (TPM) (error bars = standard deviation). Statistical significance was calculated using two-tailed t-tests for each gene pair. *P* values: * = $P < 0.01$ and ** = $P < 0.001$. (**B**) qPCR validation of differential expression of *CDE-2* (**B**), *CCS-3* (**C**) and *CCS-9* (**D**) using green clone 4106A as a reference. Error bars represent SEM.

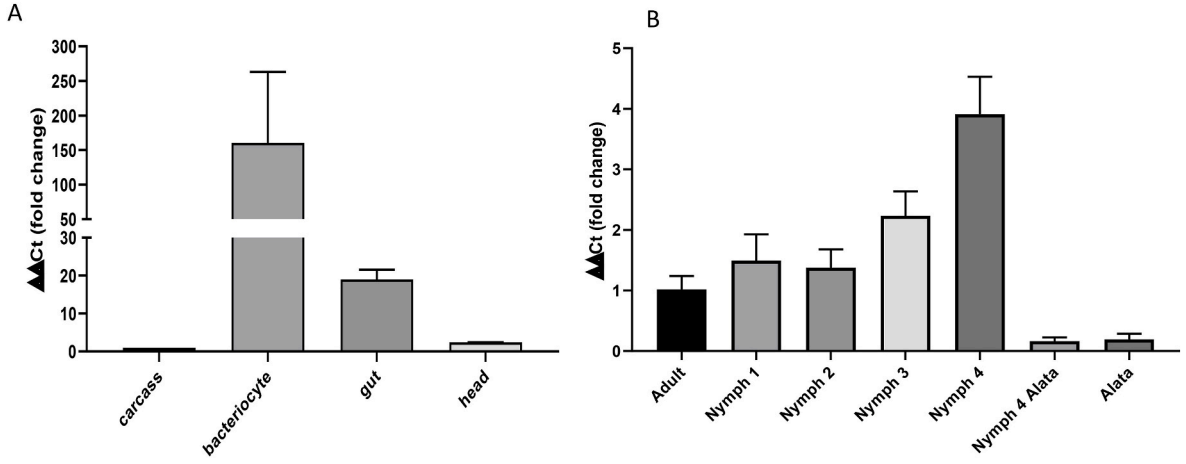


Fig. 3. (**A**) Expression of *CDE-2* in specific aphid tissues (shown relative to carcass) and (**B**) different life stages and adult forms relative to apterous adult in a representative red morph of *M. persicae*. Error bars represent SEM.

causal effect on *M. persicae* pigmentation, we conducted an RNA knockdown experiment feeding aphids on a diet containing one of six synthesised FANA oligos (Seq1-6, Table S1) that target various portions of the *CDE-2* transcript. For all six FANA aphid treatments, marked downregulation of *CDE-2* expression was observed compared to controls, with the greatest levels of knockdown observed for FANAs Seq3-6 (Fig. 4A). In a red aphid clone (clone 124) fed one of FANA Seq3-6, changes to pigmentation were observed after 7 days of continuous treatment, with around 50–60% of individuals in each treated replicate turning green (Fig. 4B). Molecular analysis of the red and green aphids from this experiment by qPCR showed a tight and consistent correlation between the level of *CDE2* expression and pigmentation, with *CDE2* expression levels 10-30-fold lower in aphids that turned green than those that remained red (Fig. 4C). Similar experiments in *A. pisum* using RNAi mediated knockdown of *Tor* resulted in a reduction of pigmentation in red and pink individuals (Ding et al., 2020). Our results indicate that FANA oligos are a viable tool for gene suppression experiments in *M. persicae* and can be used as an alternative to established RNAi techniques in aphids. Although classical RNAi approaches have been successfully used by others, it has been reported that endonucleases may limit the efficacy of orally delivered RNAi in *M. persicae* (Ghodke et al., 2019). The 2′-deoxy-2′-fluoro-beta-D-arabinonucleic acid sugar modification of FANA oligos increase resistance to degradation (Watts et al., 2009) and may provide improved resistance to endonucleases in

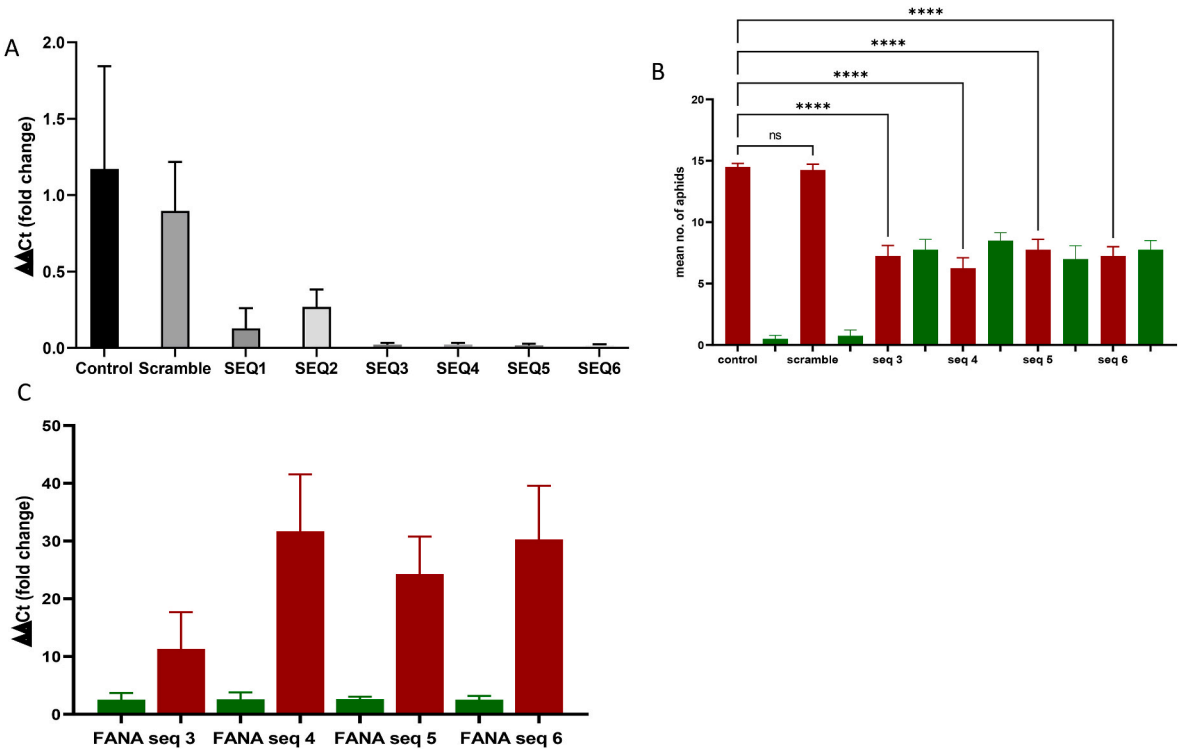


Fig. 4. Results of RNA knockdown of *CDE-2* using FANA oligonucleotides. (**A**) Expression level of *CDE-2* in aphids (clone 124) fed one of six different FANA oligos (Seq1-6) compared to controls comprising aphids fed diet without FANAs (labelled Control) or aphids fed on a FANA oligo designed with a scrambled (random) sequence not targeting any aphid gene (labelled Scramble) (**B**) Mean number of aphids (clone 124) from four cups (N = 15 per cup), which changed body colour from red (red bars) to green (green bars) following feeding for 7 days on a diet containing one of FANA oligos Seq3-6. Error bars represent SEM. Significantly different results between treatment and controls as determined by one-way ANOVA are denoted with **** corresponding to *P* value of <0.0001. **C)** Relative expression of *CDE-2* in aphids which turned green and those which remained red following feeding on a diet containing one of FANA oligos Seq3-6. Error bars represent 95% confidence limits with 4 biological reps each including 5 adults.

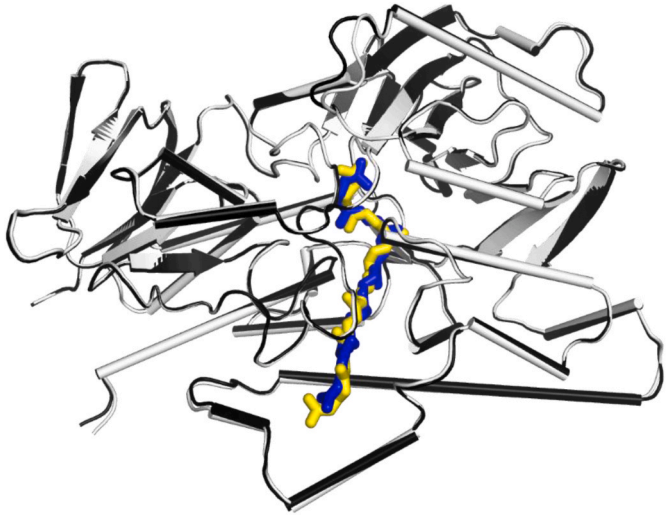


Fig. 5. Docking predictions for lycopene with enzyme models. Lycopene is shown as yellow sticks in the *A. pisum* Tor model (black) and as blue sticks in the *M. persicae* CDE-2 model (white).

M. persicae. However, it is important to acknowledge that despite the high levels of *CDE2* knockdown by the tested FANA oligos we did not observe a universal response in pigmentation (i.e. not all treated aphids turned green). Further optimisation of this technology in aphids may lead to more consistent results.

3.5. 3D modelling of CDE-2

Molecular models of *A. pisum* Tor and *M. persicae* CDE-2 were generated using Alphafold2 (Jumper et al., 2021) in order to assess their structural similarity. When the two enzyme models were superimposed, the root mean square deviation (RMSD) was 0.29 Å for the backbone atoms, indicating very close structural alignment (Fig. 5). As lycopene is a substrate for *A. pisum* Tor (Takemura et al., 2021), we next sought to compare docking predictions for this ligand with both enzyme models. Lycopene was found to adopt a similar docking pose in a narrow, hydrophobic centrally-located cavity present in both models (Fig. 5) and the estimated free energy of binding (ΔG_b) was an equivalent −10.95 kcal/mol for *A. pisum* Tor and −10.66 kcal/mol for *M. persicae* CDE-2. Furthermore, it was found that all 48 amino acids found within 6 Å of the docked ligand are conserved between *A. pisum* Tor and *M. persicae* CDE-2, indicating both enzymes share an identical set of binding contacts. Taken together, these results suggest that *M. persicae* CDE-2 is a close structural homolog of *A. pisum* Tor and that lycopene can be accommodated in the *M. persicae* CDE-2 active site as a putative substrate.

3.6. Promoter characterisation

In order to understand what is driving differential expression of *CDE-2* between colour morphs we examined the genomic sequence of *CDE-2* in both the green morph (G006) and red morph (5410R) *M. persicae* clones. Unlike in *A. pisum* where the *Tor* gene is completely absent in green morphs (Moran and Jarvik, 2010) alignment of this region in G006 and 5410R shows this locus is present in both. Mapping sequencing reads from other red and green clones did not identify any

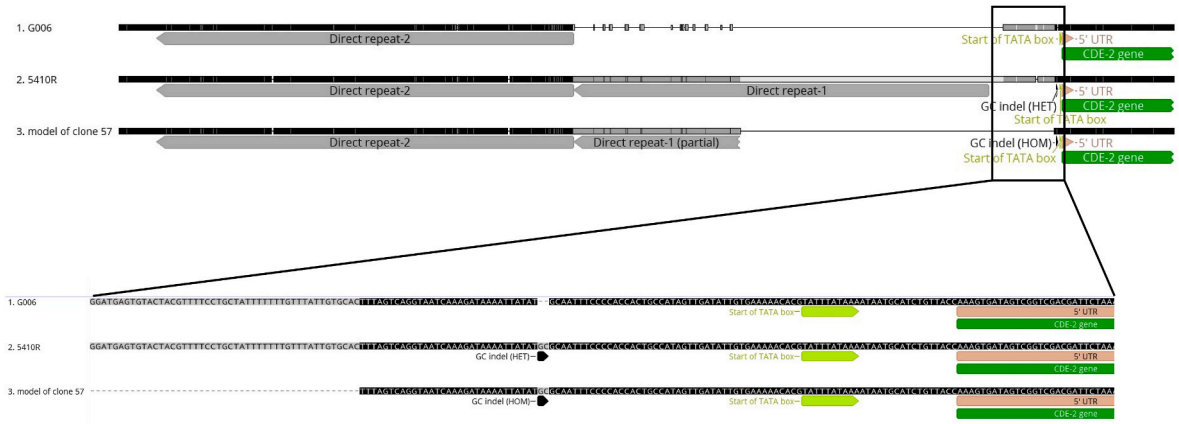


Fig. 6. Pictorial representation of the region upstream of *CDE-2*. The upper panel depicts the structure of the ≈14 kb region upstream of the gene in the green G006 morph in comparison to red morphs (5410R and clone 57). Large repeat regions are indicated. The lower panel shows a close up of the region immediately upstream of the 5′UTR with the GC indel position indicated.

coding sequence SNPs or insertions/deletions consistently different between red and green morphs. However, investigation of the putative promoter region directly upstream of *CDE-2* showed a two base pair (bp) GC indel 71bp upstream of the transcriptional start site (Fig. 6) with the GC present only in the red morphs analysed. Short read data from green morphs mapped to the 5410R locus also consistently showed the GC duplet to be absent, pointing to homozygous absence at this locus, whilst, on average, 46% of reads obtained from red morphs did contain the additional base pairs, indicating that the indel is heterozygous in most red clones. Using raw PacBio data for G006 and 5410R we were also able to compare structural differences between the promoter region of CDE-2 of red and green morphs. We identified a large tandem repeat, consisting of two 5.55 kb units 1.2 kb upstream of the transcriptional start site in red morphs, whilst only a single repeat unit is present in all green morphs (Fig. 6). When G006 reads (n = 8) were mapped to the 5410R locus all reads showed a large gap that spanned exactly the length of a single copy of the repeat, suggesting the deletion is homozygous. More in-depth analysis of the repeated region showed that two copies differ by only 3 SNPs. Short read data from other red morphs mapped to the locus showed equal numbers of reads for both versions of SNPs, indicating that the duplication is homozygous in red coloured aphids. NCBI and pfam BLAST searches yielded no significant hits for the single repeat sequence. We also did not identify any terminal inverted repeats (TIRs), which would indicate the origin of the sequence as a possible mobile element or that its duplication was mediated by a transposable element. It is therefore likely that this duplication arose through other means, possibly via homologous recombination (Reams and Roth, 2015). The carotenoid locus sits in the region of the autosome 1 frequently involved in recurrent chromosomal rearrangements (Mandrioli et al., 2016), which could explain the apparent diversity in promoter structure of *CDE-2* gene between red and green morphs.

Interestingly, we identified what appear to be two distinct upstream sequences in red morphs which allow us to speculate on the causative polymorphism driving differential expression. Clone 57 differs from other red clones in two respects. Firstly, it is homozygous for the upstream GC indel (which is heterozygous in other red morphs). Secondly, it differs from other red clones in having a 4.17 Kb deletion located 100bp upstream of the transcriptional start site (Fig. 6) that removes the majority of one of the tandem repeats. Lack of any short reads mapping to this region suggests this locus is homozygous. Expression of *CDE-2* in clone 57 is typical of other red morphs (TPM value of 263,212 compared to the average of 261,153 in red morphs) indicating that it is the GC indel, rather than two copies of the repeat region which is associated with altered expression. Additionally, the repeat itself is unlikely to be a transcriptional enhancer as there are no additional copies of it anywhere else in the genome. Even relatively modest changes to the promoter region can have significant effects on gene expression. In *M. persicae*, insertion of a CA motif within the promoter of the P450 gene *CYP6CY3* significantly upregulated its expression contributing to nicotine tolerance (Bass et al., 2013) whilst a single SNP in the core promoter of the *Culex quinquefasciatus CYP9M10* promoter is associated with significant expression changes (Itokawa et al., 2015). Thus, this GC indel is the potential causative driver of *CDE-2* over-expression in red *M. persicae*. To provide functional evidence of this, we created constructs with and without the GC indel or the repeat unit and used reporter gene assays to explore the effect of this on gene expression. Unfortunately, the promoter constructs generated were not active when expressed in a lepidopteran cell line (Supplementary Figure 2). Thus, demonstration of the causal role of the GC indel on expression will require further future investigation using alternative methodologies such as CRISPR/CAS genome editing. Nevertheless, the absolute association between the indel and colouration is suggestive of a regulatory role for this polymorphism.

4. Conclusions

Here we provide evidence that *M. persicae CDE-2* is the homolog of *A. pisum Tor* and that whilst *CDE-2* is found in both red and green forms it is highly differentially expressed between red and green colour morphs. Our RNAi data and 3D modelling strongly support the involvement of this gene in the red/green phenotype. Whilst our investigation of promoter polymorphisms implicate a GC-indel as the regulatory variant driving this phenotype, further confirmatory work, possibly involving CRISPR-modification is needed.

Acknowledgments

We thank Emma Randall for technical support with gene cloning.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ibmb.2022.103896.

References

Alkhedir, H., Karlovsky, P., Vidal, S., 2010. Effect of light intensity on colour morph formation and performance of the grain aphid *Sitobion avenae* F. (Homoptera: Aphididae). J. Insect Physiol. 56, 1999–2005.

Altincicek, B., Kovacs, J.L., Gerardo, N.M., 2012. Horizontally transferred fungal carotenoid genes in the two-spotted spider mite *Tetranychus urticae*. Biol. Lett. 8, 253–257.

Bass, C., Zimmer, C.T., Riveron, J.M., Wilding, C.S., Wondji, C.S., Kaussmann, M., Field, L.M., Williamson, M.S., Nauen, R., 2013. Gene amplification and

microsatellite polymorphism underlie a recent insect host shift. Proc. Natl. Acad. Sci. USA 110, 19460–19465.
Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., Finn, R.D., 2020. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 49, D344–D354.
Caillaud, M.C., Losey, J.E., 2010. Genetics of color polymorphism in the pea aphid, *Acyrthosiphon pisum*. J. Insect Sci. 10, 95.
Cazzonelli, C.I., 2011. Carotenoids in nature: insights from plants and beyond. Funct. Plant Biol. 38, 833–847.
Chew, B.P., Park, J.S., 2004. Carotenoid action on the immune response. J. Nutr. 134, 257S–261S.
Cobbs, C., Heath, J., Stireman, J.O., Abbot, P., 2013. Carotenoids in unexpected places: gall midges, lateral gene transfer, and carotenoid biosynthesis in animals. Mol. Phylogenet. Evol. 68, 221–228.
Ding, B.-Y., Niu, J., Shang, F., Yang, L., Zhang, W., Smagghe, G., Wang, J.-J., 2020. Parental silencing of a horizontally transferred carotenoid desaturase gene causes a reduction of red pigment and fitness in the pea aphid. Pest Manag. Sci. 76, 2423–2433.
Duarte, A., Pym, A., Garrood, W.T., Troczka, B.J., Zimmer, C.T., Davies, T.G.E., Nauen, R., O'Reilly, A.O., Bass, C., 2022. P450 gene duplication and divergence led to the evolution of dual novel functions and insecticide cross-resistance in the brown planthopper *Nilaparvata lugens*. PLoS Genet. 18, e1010279.
Fiedor, J., Burda, K., 2014. Potential role of carotenoids as antioxidants in human health and disease. Nutrients 6, 466–488.
Fukatsu, T., 2010. A fungal past to insect color. Science 328, 574–575.
Gauthier, J.-P., Legeai, F., Zasadzinski, A., Rispe, C., Tagu, D., 2007. AphidBase: a database for aphid genomic resources. Bioinformatics 23, 783–784.
Ghodke, A.B., Good, R.T., Golz, J.F., Russell, D.A., Edwards, O., Robin, C., 2019. Extracellular endonucleases in the midgut of *Myzus persicae* may limit the efficacy of orally delivered RNAi. Sci. Rep. 9, 11898.
Guex, N., Diemand, A., Peitsch, M.C., 1999. Protein modelling for all. Trends Biochem. Sci. 24, 364–367.
Itokawa, K., Komagata, O., Kasai, S., Tomita, T., 2015. A single nucleotide change in a core promoter is involved in the progressive overexpression of the duplicated *CYP9M10* haplotype lineage in *Culex quinquefasciatus*. Insect Biochem. Mol. Biol. 66, 96–102.
Ives, A.R., Barton, B.T., Penczykowski, R.M., Harmon, J.P., Kim, K.L., Oliver, K., Radeloff, V.C., 2020. Self-perpetuating ecological–evolutionary dynamics in an agricultural host–parasite system. Nat. Ecol. Evol. 4, 702–711.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.
Kalota, A., Karabon, L., Swider, C.R., Viazovkina, E., Elzagheid, M., Damha, M.J., Gewirtz, A.M., 2006. 2′-Deoxy-2′-fluoro-β- d -arabinonucleic acid (2′F-ANA) modified oligonucleotides (ON) effect highly efficient, and persistent, gene silencing. Nucleic Acids Res. 34, 451–461.
Komaki, K., Ishikawa, H., 2000. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. Insect Biochem. Mol. Biol. 30, 253–258.
Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. Mega X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.
Kumar, S., Stecher, G., Suleski, M., Hedges, S.B., 2017. TimeTree: a resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. 34, 1812–1819.
Li, Y., Park, H., Smith, T.E., Moran, N.A., 2019. Gene family evolution in the pea aphid based on chromosome-level genome assembly. Mol. Biol. Evol. 36, 2143–2156.
Losey, J.E., Harmon, J., Ballantyne, F., Brown, C., 1997. A polymorphism maintained by opposite patterns of parasitism and predation. Nature 388, 269–272.
Mandrioli, M., Rivi, V., Nardelli, A., Manicardi, G.C., 2016. Genomic and cytogenetic localization of the carotenoid genes in the aphid genome. Cytogenet. Genome Res. 149, 207–217.
Maoka, T., 2020. Carotenoids as natural functional pigments. J. Nat. Med. 74, 1–16.
Mathers, T.C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S.T., Baa-Puyoulet, P., Bretaudeau, A., Clavijo, B., Colella, S., Collin, O., Dalmay, T., Derrien, T., Feng, H., Gabaldon, T., Jordan, A., Julca, I., Kettles, G.J., Kowitwanich, K., Lavenier, D., Lenzi, P., Lopez-Gomollon, S., Loska, D., Mapleson, D., Maumus, F., Moxon, S., Price, D.R.G., Sugio, A., van Munster, M., Uzest, M., Waite, D., Jander, G., Tagu, D., Wilson, A.C.C., van Oosterhout, C., Swarbreck, D., Hogenhout, S.A., 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. Genome Biol. 18, 27.
Miki, W., 1991. Biological functions and activities of animal carotenoids. Pure Appl. Chem. 63, 141–146.
Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. Nat. Methods 19, 679–682.
Moran, N.A., Jarvik, T., 2010. Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science 328, 624–627.
Nováková, E., Moran, N.A., 2011. Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. Mol. Biol. Evol. 29, 313–323.
Paliwal, D., 2017. Identification and Characterisation of New Aphid Killing Bacteria for Use as Biological Pest Control Agents. PhD. University of Reading.
Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res. 29, e45.
Reams, A.B., Roth, J.R., 2015. Mechanisms of gene duplication and amplification. Cold Spring Harbor Perspect. Biol. 7, a016592.
Shi, Y., O'Reilly, A.O., Sun, S., Qu, Q., Yang, Y., Wu, Y., 2020. Roles of the variable P450 substrate recognition sites SRS1 and SRS6 in esfenvalerate metabolism by CYP6AE subfamily enzymes in Helicoverpa armigera. Insect Biochem. Mol. Biol. 127, 103486.
Simon, J.-C., Carré, S., Boutin, M., Prunier–Leterme, N., Sabater–Muñoz, B., Latorre, A., Bournoville, R., 2003. Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. Proc. R. Soc. Lond. B Biol. Sci. 270, 1703–1712.
Singh, K.S., Cordeiro, E.M.G., Troczka, B.J., Pym, A., Mackisack, J., Mathers, T.C., Duarte, A., Legeai, F., Robin, S., Bielza, P., Burrack, H.J., Charaabi, K., Denholm, I., Figueroa, C.C., ffrench-Constant, R.H., Jander, G., Margaritopoulos, J.T., Mazzoni, E., Nauen, R., Ramírez, C.C., Ren, G., Stepanyan, I., Umina, P.A., Voronova, N.V., Vontas, J., Williamson, M.S., Wilson, A.C.C., Xi-Wu, G., Youn, Y.-N., Zimmer, C.T., Simon, J.-C., Hayward, A., Bass, C., 2021. Global patterns in genomic diversity underpinning the evolution of insecticide resistance in the aphid crop pest *Myzus persicae*. Commun. Biol. 4, 847.
Singh, K.S., Troczka, B.J., Duarte, A., Balabanidou, V., Trissi, N., Carabajal Paladino, L. Z., Nguyen, P., Zimmer, C.T., Papapostolou, K.M., Randall, E., Lueke, B., Marec, F., Mazzoni, E., Williamson, M.S., Hayward, A., Nauen, R., Vontas, J., Bass, C., 2020. The genetic architecture of a host shift: an adaptive walk protected an aphid and its endosymbiont from plant chemical defenses. Sci. Adv. 6, eaba1070.
Takada, H., 1981. Inheritance of body colors in *Myzus persicae* (Sulzer)(Homoptera : Aphididae). Appl. Entomol. Zool. 16, 242–246.
Takemura, M., Maoka, T., Koyanagi, T., Kawase, N., Nishida, R., Tsuchida, T., Hironaka, M., Ueda, T., Misawa, N., 2021. Elucidation of the whole carotenoid biosynthetic pathway of aphids at the gene level and arthropodal food chain involving aphids and the red dragonfly. BMC Zool 6, 19.
Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461.
Tsuchida, T., Koga, R., Fujiwara, A., Fukatsu, T., 2014. Phenotypic effect of "*Candidatus Rickettsiella viridis*," a facultative symbiont of the pea aphid (*Acyrthosiphon pisum*), and its interaction with a coexisting symbiont. Appl. Environ. Microbiol. 80, 525–533.
Tsuchida, T., Koga, R., Horikawa, M., Tsunoda, T., Maoka, T., Matsumoto, S., Simon, J.-C., Fukatsu, T., 2010. Symbiotic bacterium modifies aphid body color. Science 330, 1102–1104.
Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3 research0034.0031.
Vogel, K.J., Moran, N.A., 2011. Effect of host genotype on symbiont titer in the aphid—*Buchnera* symbiosis. Insects 2, 423–434.
von Lintig, J., 2012. Metabolism of carotenoids and retinoids related to vision. J. Biol. Chem. 287, 1627–1634.
Wang, X.-X., Chen, Z.-S., Feng, Z.-J., Zhu, J.-Y., Zhang, Y., Liu, T.-X., 2019. Starvation stress causes body color change and pigment degradation in *Acyrthosiphon pisum*. Front. Physiol. 10.
Watts, J.K., Katolik, A., Viladoms, J., Damha, M.J., 2009. Studies on the hydrolytic stability of 2′-fluoroarabinonucleic acid (2′F-ANA). Org. Biomol. Chem. 7, 1904–1910.
Weisser, W.W., Braendle, C., 2001. Body colour and genetic variation in winged morph production in the pea aphid. Entomol. Exp. Appl. 99, 217–223.
Zhao, C., Nabity, P.D., 2017. Phylloxerids share ancestral carotenoid biosynthesis genes of fungal origin with aphids and adelgids. PLoS One 12, e0185484.