



Expression of Multiple *engrailed* Family Genes in Eyespots of *Bicyclus anynana* Butterflies Does Not Implicate the Duplication Events in the Evolution of This Morphological Novelty

Tirtha Das Banerjee^{1*}, Diane Ramos² and Antónia Monteiro^{1,3*}

¹ Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore, Singapore,

² Department of Natural Sciences, Daemen College, Amherst, NY, United States, ³ Science Division, Yale-NUS College, Singapore, Singapore

OPEN ACCESS

Edited by:

Marcus Kronforst,
University of Chicago, United States

Reviewed by:

Elizabeth Jenness Duncan,
University of Leeds, United Kingdom
Ralf Janssen,
Uppsala University, Sweden

*Correspondence:

Tirtha Das Banerjee
tirtha_banerjee@u.nus.edu
Antónia Monteiro
antonia.monterio@nus.edu.sg

Specialty section:

This article was submitted to
Evolutionary Developmental Biology,
a section of the journal
Frontiers in Ecology and Evolution

Received: 01 April 2020

Accepted: 22 June 2020

Published: 24 July 2020

Citation:

Banerjee TD, Ramos D and
Monteiro A (2020) Expression
of Multiple *engrailed* Family Genes
in Eyespots of *Bicyclus anynana*
Butterflies Does Not Implicate
the Duplication Events in the Evolution
of This Morphological Novelty.
Front. Ecol. Evol. 8:227.
doi: 10.3389/fevo.2020.00227

Gene duplication events often create genetic redundancy that can either lead to the appearance of pseudogenes or, instead, create opportunities for the evolution of novel proteins that can take on new functions. One of the genes which has been widely studied with respect to gene duplication is *engrailed* (*en*). *En*-family proteins are expressed in a morphological novelty, eyespots (in the center and in the outer gold ring), in the African squinting bush brown butterfly *Bicyclus anynana*, as well as in a more conserved pattern, the posterior compartment of a wing. In the present study, we used whole-genome sequencing and transcriptome data to show the presence of three *en*-family genes and their differential expression on the pupal wings of *B. anynana* using *in situ* hybridization. The results suggest two duplication events of *en*-family genes, the first evidence of a two-fold duplication in the Lepidoptera. We propose that all copies initially had posterior wing compartment expression and all copies subsequently gained a novel expression domain associated with eyespot centers. Two copies secondarily lost the posterior compartment expression, and one copy alone gained the outer ring expression domain. By dating the origin of both duplication events, however, we conclude that they predate the origin of eyespots by at least 60 mya, and hence our data does not support the retention of the multiple *en* gene duplicates in the genome via their involvement with the novel eyespot evolutionary innovation.

Keywords: gene duplication, *in situ* hybridization, eyespots, *engrailed*, *invected*, *Bicyclus anynana*

INTRODUCTION

Gene duplication plays an essential role in the evolution of novel proteins, and this process has also been proposed to lead to the evolution of major morphological innovations such as flowers; the chambers of the heart; and brains, bones, and cartilage in vertebrates (Theißen, 2001; Olson, 2006; Conant and Wolfe, 2008; Wagner and Lynch, 2010; Zhang et al., 2015). When genes duplicate, there is initially genetic redundancy which allows one of the copies to quickly accumulate mutations and become converted into a pseudogene (Lynch et al., 2001; Zhang, 2003). However, genetic redundancy can also lead to opportunity. When gene duplicates remain intact during the

evolutionary processes this usually means one of three things: (1) that either the double dose of RNA and/or protein increased fitness (Zhang, 2003); (2) that each copy divided the original functions and/or expression domains of the protein among themselves, a process known as sub-functionalization that reduces the total number of functions taken over by each copy and lessens pleiotropy (Force et al., 1999); (3) or that one of the copies gained a novel function, known as neo-functionalization (Force et al., 1999). Examples include sub-functionalization of the gene *engrailed* (*en*) in zebrafish, which resulted in the differential expression of the copies in the pectoral appendage, and in neurons (Force et al., 1999); and neo-functionalization of eosinophil-derived neurotoxin (EDN) and eosinophil cationic protein (ECP) in humans and old world monkeys from EDN, where after the duplication ECP developed a novel anti-pathogenic function (Zhang et al., 1998). Neo-functionalization has been of particular interest since the origin of novel gene duplicates has been loosely associated with the origin of evolutionary innovations (Conant and Wolfe, 2008; Wagner and Lynch, 2010), such as with beetle horns (Zattara et al., 2016), human vision (Yokoyama and Yokoyama, 1989), and betalain pigments in plants (Brockington et al., 2015).

The gene *engrailed* (*en*) has been heavily studied with respect to gene duplications (Force et al., 1999). Duplication of *en* in several independent metazoan lineages has generated a family of related genes, varying from one to four copies, where the paralogs display varying degrees of divergence in expression and function (Peterson et al., 1998; Abzhanov and Kaufman, 2000; Gibert et al., 2000; Damen, 2002; Marie and Blagburn, 2003; Peel et al., 2006).

In addition to its conserved function in segmentation during invertebrate embryogenesis (Manzanares et al., 1993; Brown et al., 1994; Fujioka et al., 2002), *en* is also associated with morphological novelties in several species. For instance, *en* is involved in neurogenesis (Patel et al., 1989), in determining the fate of glial and neuronal cells in grasshopper median neuroblast (Condrón et al., 1994), in axonal targeting in centipedes (Whittington et al., 1991), and in the patterning of insect wing veins (Guillén et al., 1995; Banerjee and Monteiro, 2020). In addition, while no function is yet known, *en* is expressed in precursor cells that build a mollusc's shell (Nederbragt et al., 2002), in the tentacles of the cephalopod *Sepia officinalis* (Baratte and Bonnaud, 2009), in bacteriocytes of the aphid, *Acyrtosiphon pisum* (Braendle et al., 2003), and in the eyespots of saturniid moths (Monteiro et al., 2006) and nymphalid butterflies (Keys et al., 1999; Brunetti et al., 2001).

Eyespots in butterfly wings are novel traits that appear to be specified by old transcription factors and signaling pathways, deployed in novel ways (Keys et al., 1999; Brunetti et al., 2001; Oliver et al., 2012; Monteiro, 2015). Using an antibody that detects a variety of Engrailed proteins (Patel et al., 1989), Keys et al. (1999) found that En-family proteins are expressed in both conserved and novel domains in butterfly wings. En-family proteins maintained their conserved expression domain, and likely their role, in establishing the posterior compartment of each embryonic body segment including that of the wings during the larval stage (Carroll et al., 1994; Keys et al., 1999; Banerjee and Monteiro, 2020). However, one or more of the genes encoding En or its paralog Invested (Inv) were also expressed in the

eyespot centers, in late larval and pupal wings, and in a pattern mapping to the ring of gold scales during the pupal stage of wing development in *Bicyclus anynana* butterflies (Brakefield et al., 1996; Brunetti et al., 2001). A detailed investigation of the number of *en* family gene copies present in the genome of *butterflies*, however, and of the expression domains of each paralog in eyespot patterns is still missing.

Here we provide additional insights on the association between the retention of duplicated genes and their involvement in the origin of novel traits by detailing the expression domain of multiple *en* copies in *B. anynana* wings, asking whether one or more copies are exclusively associated with the eyespot patterns, and testing whether the gene duplication events are closely associated with the origin of eyespots. We report the presence of three *en*-family genes in *B. anynana*, two of which contain all the five major domains of En-family proteins (Peel et al., 2006) and one contains four of these domains. We used sequencing data, to analyze the phylogenetic relationship between the copies, and whole-mount *in situ* hybridizations to examine the expression of all three copies in the pupal wings. Two of the copies are closely related to each other and have a single expression domain on the pupal wing, in the eyespot centers, while the other copy displays three wing expression domains, in the posterior compartment, in the eyespot centers, and in the golden ring. The timing between the last *en*-family gene duplication and the origin of eyespots, however, spans at least 60 mya. We discuss the implications of our data on the mechanisms of evolution of these morphological novelties.

MATERIALS AND METHODS

Rearing *Bicyclus anynana*

B. anynana butterflies were reared in the lab at 27°C, 60% relative humidity, and 12–12 h day–night cycle. Larvae were fed young corn leaves and adults were fed mashed bananas.

Mining for *en* Paralogs and Splice Variants

The sequences of *en*-family genes and splice variants of *B. anynana* were obtained via whole genome sequencing data [*Bicyclus anynana* v1.2 (Nowell et al., 2017)], transcriptome [*B. anynana*; eyespot (Özsu and Monteiro, 2017) and pupal brain (Macias-Munoz et al., 2016)], and 5' and 3' RACE (Rapid Amplification of cDNA Ends). Protein sequences of En and Inv of *D. melanogaster* were obtained from FlyBase. En-family protein sequences of other Lepidopteran species were obtained from Lepbase (Challis et al., 2016), NCBI, EMBL and transcriptome data (Ferguson et al., 2010; Daniels et al., 2014; Connahs et al., 2016) and analyzed for conserved common and paralog-specific domains.

Sequence Alignment and Phylogenetic Tree Construction

Nucleotide sequence alignment was carried out using ClustalW (Thompson et al., 1994) with the default parameters in 'SLOW/ACCURATE' pairwise alignment option in GenomeNet

and visualized in Geneious v10.1.3 (Kearse et al., 2012). Nucleotides were converted into protein sequence via NCBI ORF finder. The protein sequences were aligned via ClustalW (Thompson et al., 1994) with gap open cost: 10, and gap extend cost 0.1; and afterward, manually edited and visualized in Geneious v10.1.3 (Kearse et al., 2012).

A maximum likelihood tree was inferred using RAxML v8.1.20 ran with model PROTGAMMAJTT and default parameters with 100 bootstraps (Stamatakis, 2014) using ETE v3.1.1 (Huerta-Cepas et al., 2016) implemented on GenomeNet. The same tree topology was also obtained using the following other methods and software:

- (a) PhyML v20160115 ran with model JTT and parameters: -f m -nclases 4 -alpha e -pinv e -o tlr -bootstrap 100 (Guindon et al., 2010), using ETE3 v3.1.1 (Huerta-Cepas et al., 2016);
- (b) fasttree with slow NNI and MLACC = 3 (to make the maximum-likelihood NNIs more exhaustive) (Price et al., 2009) using ETE v3.1.1 (Huerta-Cepas et al., 2016); and
- (c) geneious tree builder using Jukes-Cantor as the genetic distance model, UPGMA as tree build method, and 1000 bootstraps (Kearse et al., 2012).

Probe Design of the *en*-Family Genes

Sequence used for *en* probe: A 628 bp region in the exon 1 of *en* CDS was chosen for designing a probe. This region is absent in *inv* and *inv-like* and does not align well with these genes (~35% bp similarity) (Supplementary Figure S5).

Sequences used for *inv* and *inv-like* probes: The regions used for probe preparation are Exon 1 of *inv*, which is not present in *inv-like*, and the 5' UTR of *inv-like*, which is not present in *inv*. The sequence used for the *inv* probe has 24.9% identical bps to that of the *en* gene sequence, and 36.03% identical bps to that of the *inv-like* gene sequence in their overlapping regions. The sequence used for the *inv-like* probe has 30.2% identical bps to that of the *en* gene sequence, and 36.8% identical bps to that of the *inv* gene sequence in the overlapping region (Supplementary Figure S5). This level of sequence similarity is lower than that observed between highly unrelated genes such as *en* and *optix*, which share 41.2% identical sequences. The presence of a very low sequence identity and a small overlap as observed in Supplementary Figure S5 makes it unlikely for the two probes to produce results due to cross-reaction.

Control stainings: *inv-like* and *inv* are mostly identical in the CDS region. Due to this limitation, we used a 182 bps long region in the 5' UTR of *inv-like* (Figure 1A) which is small in size and can produce non-specific stainings. To test for presence of non-specific stainings we performed *inv-like* sense stainings using a probe prepared with the same sequence (Supplementary Figures S1p,q). Sense stainings were also performed for *en* and *inv* (Supplementary Figures S1d-f,j-l).

Prior to probe preparation, amplified fragments of each paralog were sequenced and aligned to ensure that the correct amplicons were used for *in situ* hybridization.

In situ Hybridization

One to two days old pupal wings, timed for moment of pupation using a time-lapse camera, were dissected in 1X PBS at room temperature based on a previously published protocol (Banerjee and Monteiro, 2020) and transferred into 1X PBST with 4% formaldehyde for 30–35 min. Afterward, the wings were washed three times in 1X PBST and treated with 20 mg/ml proteinase K (NEB, P8107S) in 1 ml 1X PBST and then with 2 mg/ml glycine in 1X PBST. Wings were then washed three times in 1X PBST and gradually transferred into a prehybridization buffer (composition in Supplementary Table S4) and incubated at 60°C for 1 h prior to transfer into hybridization buffer (composition in Supplementary Table S4) supplemented with 100 ng/μl of probe. The wings were incubated in hybridization buffer at 60°C for 16–20 h and washed five times with prehybridization buffer at 60°C to remove any unbound probe. To prevent cross-hybridization of the probes, experiments with each probe were performed on separate days. Afterward, wings were brought to room temperature and gradually transferred to 1X PBST and then incubated in block buffer (composition in Supplementary Table S4) for 1 h. 1:3000 dilution of anti-Digoxigenin was then added and incubated for 1 h followed by five washes with block buffer. The wings were then transferred to an alkaline-phosphatase buffer (composition in Supplementary Table S4) supplemented with NBT/BCIP. Imaging was done under a Leica DMS1000 microscope.

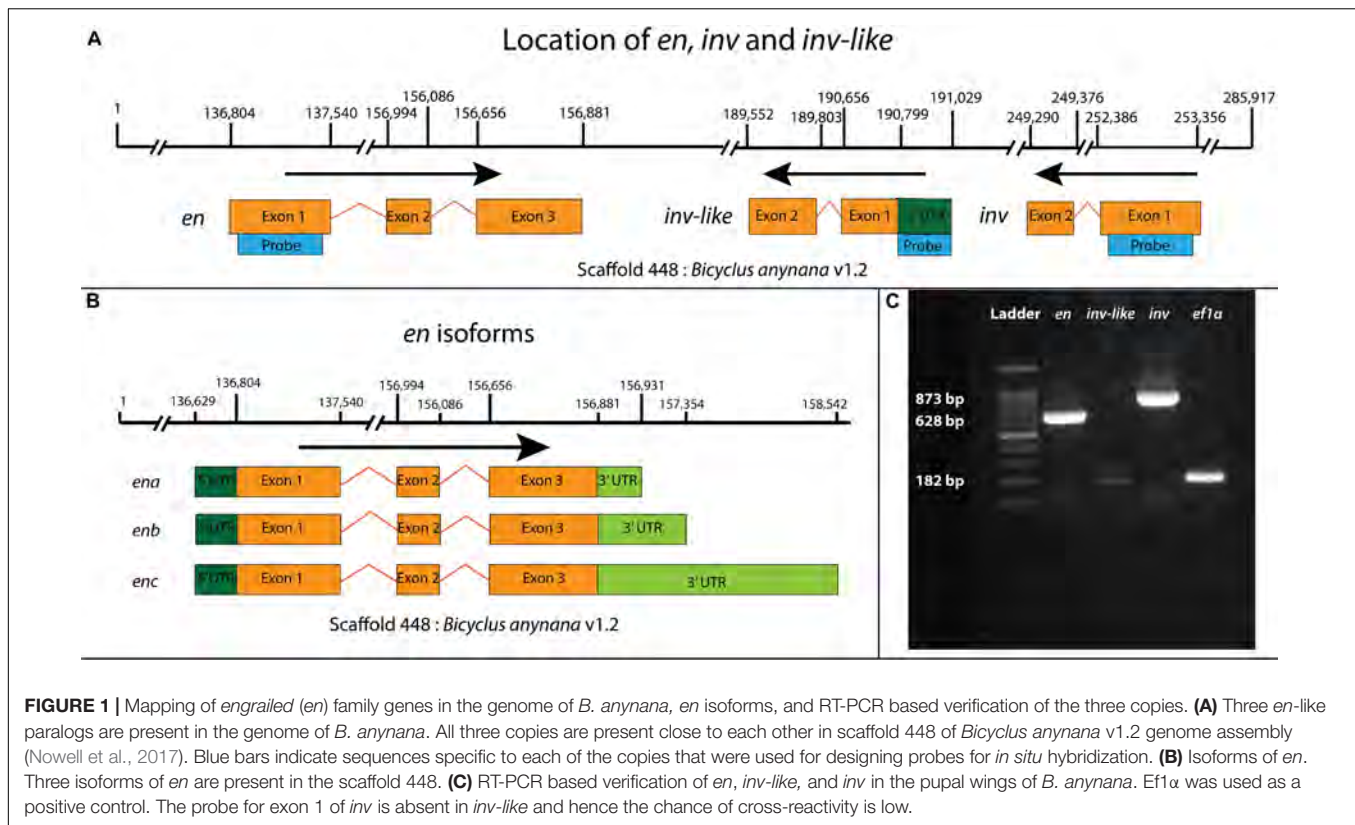
Antibody Staining

One to two days old pupal wings, timed using a time-lapse camera, were dissected in 1X PBS and immediately transferred into a fixation buffer supplemented with 4% formaldehyde at room temperature (composition in Supplementary Table S5) for 30–35 min. Wings were then moved to ice and washed four times with 1X PBS and transferred to block buffer for 1 day (composition in Supplementary Table S5). The next day wings were incubated in primary antibodies against En/Inv [1:15, mouse 4F11, a gift from Patel et al. (1989)], followed by anti-mouse AF488 (Invitrogen, #A28175) secondary antibody. Wings were then washed and mounted on an inhouse mounting media (composition in Supplementary Table S5) and imaged under an Olympus fv3000 confocal microscope.

RESULTS

Three *en*-Family Copies Are Present in the Genome of *Bicyclus anynana*

Whole-genome sequencing (Nowell et al., 2017) and transcriptomic data (Macias-Munoz et al., 2016; Özsü and Monteiro, 2017) revealed three *en*-family genes present in the genome of *B. anynana* within a single scaffold. We named the first copy *engrailed* (*en*) based on the presence of specific sequence domains described below. It has three exons and three splice variants. The 5' UTR of the splice variants are the same, but the 3' UTR region has different lengths across the three variants. The sequences are present between 136,804 and 158,542 bps



of Scaffold 448: *Bicyclus anynana* v1.2 (Figures 1A,B). The transcript is translated into a 352 aa sequence containing all the five conserved domains of En-family proteins and the two En specific domains (Hui et al., 1992) (Figure 2A). The second copy we named *invected* (*inv*). It has two exons present between 249,290 and 235,356 bps of the same scaffold (Figure 1A). The transcript is translated into a 479 aa sequence containing all the five conserved domains of En-family proteins and the Inv specific LVSG and RS domains (Hui et al., 1992; Peel et al., 2006) (Figure 2A). The third copy we named *invected-like* (*inv-like*). It has two exons with a 5' UTR as identified using a pupal brain (Macias-Munoz et al., 2016) and an eyespot-specific transcriptome (Özsu and Monteiro, 2017) (see **Supplementary File** for sequence). The sequence is present between 189,552 and 191,029 bps of the same scaffold (Figure 1A). The transcript is translated into a 125 aa sequence containing four domains of En-family proteins with high similarity to Inv (Figure 2A and results below for details).

Conserved Domains of En-Family Proteins in *B. anynana*

En and Inv proteins contain all five domains of high conservation previously identified for En-family proteins, while Inv-like contains three full domains and one partial domain (Figure 2A) (Logan et al., 1992). The function of most of these regions has been determined: EH1 and EH2 are binding sites for the transcriptional co-repressor Groucho (Tolkunova et al., 1998) and the co-activator Exd (Peltenburg and Murre, 1996)

respectively, EH4 is the homeodomain (Fjose et al., 1985) and EH5 represses En targets (Han and Manley, 1993). The function of EH3 is not known. The EH1 domain of En and Inv are 13 aa each and share 61.5% identity. The EH2 domain of En and Inv are 19 and 21 aa respectively and share 90.5% identity, while Inv-like has only three aa from this domain. The EH3 domain is highly divergent among species and paralogs. En has 11 aa, while Inv and Inv-like have 15 aa each for this domain, respectively. The EH4 domain is a 60 aa homeodomain region which is highly conserved in the three paralogs. Inv and Inv-like are 100% identical in this domain while En has 95% identity with the other two. The EH5 domain is an 18 aa region which is also highly conserved. Inv and Inv-like are 100% identical in this region while En shares 94.4% identity with the other two (Figure 2B).

Phylogenetic Analysis of En-Family Proteins in Lepidoptera

Phylogenetic analysis of the highly conserved sequences from the EH2 to EH5 domains of En-family proteins from nine lepidopteran species using RaxML (Randomly Accelerated Maximum likelihood) (Stamatakis, 2014) resulted in the clustering of all En sequences as a sister lineage to Inv and Inv-like sequences in Lepidoptera (Figure 2C). The same results were obtained using PhyML (Guindon et al., 2010) (Supplementary Figure S2), fasttree (Price et al., 2009) (Supplementary Figure S3), and Geneious Tree builder (Kearse et al., 2012) (Supplementary Figure S3). The protein sequences of Inv and Inv-like at the conserved EH2 to EH5 have 98.9% similarity;

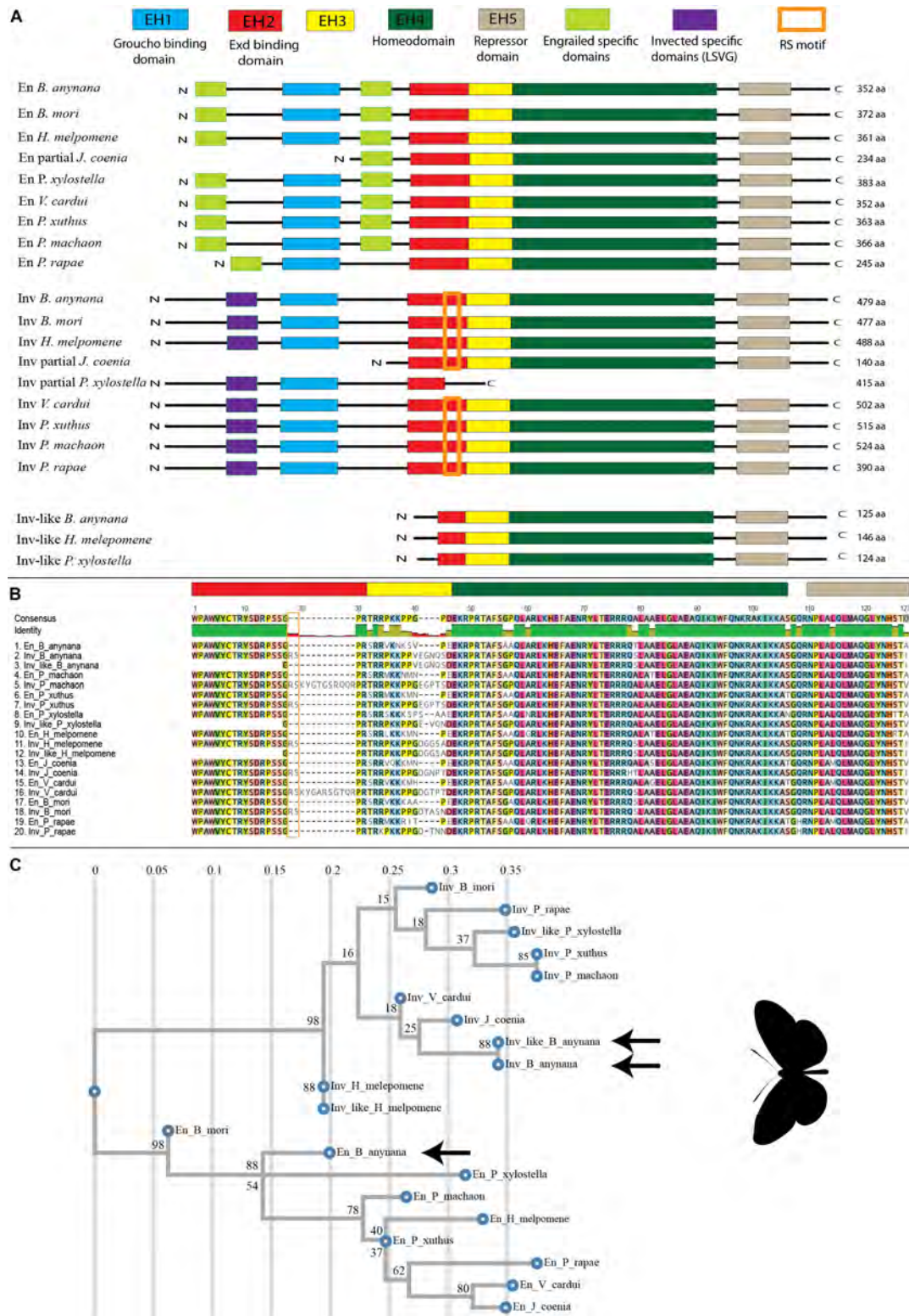


FIGURE 2 | Alignment and Phylogenetic tree of En-family proteins in Lepidoptera. **(A)** En-family proteins have five conserved domains named EH1 to EH5. En and Inv contain all the five En-family domains along with gene-specific domains, while Inv-like only contains three complete domains (EH3 to EH5) and one partial domain (EH2). **(B)** Alignment of En-family proteins around the conserved EH2-EH5. **(C)** Phylogenetic tree of En-family protein created using RAxML (Stamatakis, 2014). En proteins cluster together; and Inv, and Inv-like cluster together. Arrows point to the *B. anynana* paralogs. The values on the top axis represent mean number of amino acid substitutions per site and the node values represent branch support calculated based on 100 bootstraps.

while *En* has 90.1% similarity to *Inv*, and 91.2% similarity to *Inv-like*. Furthermore, the nucleotide sequences of *inv* and *inv-like* are 98.25% similar; while *en* is 80.07% similar to *inv*, and 81.81% similar to *inv-like* at the conserved EH4 and EH5 regions (Supplementary Figure S5 and Supplementary Table S1). The presence of highly conserved sequences between *Inv* and *Inv-like* compared to *En* indicate that *inv* and *inv-like* underwent a more recent duplication.

Expression of *en* Paralogs in the Pupal Wings of *B. anynana*

In situ hybridization results indicate expression differences between the three *en* paralogs in the pupal wings of *B. anynana*. *en* is expressed in the posterior compartment of the wing, in the eyespot centers, and in the golden ring area surrounding each center (Figures 3d–f; Supplementary Figures S1a–c). *inv* and *inv-like*, however, are only present in the eyespot centers (Figures 3i–n and Supplementary Figures S1g–i,m–o).

DISCUSSION

Duplication of *en* Paralogs in Holometabolous Insects

Recent research proposed that a common ancestor to all Lepidoptera has undergone a large-scale genome duplication that is consistent with a whole-genome duplication, while hexapod ancestors to both Diptera and Lepidoptera have undergone two other large scale bursts of gene duplication (Li et al., 2018). These duplications are believed to play important roles in the evolution of hexapods by increasing genetic redundancy and freeing up redundant copies to evolve new functions (Li et al., 2018; Doyle and Coate, 2019). The duplication of *en* into *en* and *inv* happened to a common ancestor to all hexapods including the holometabolous orders Diptera (containing *Drosophila*) and Lepidoptera (butterflies and moths) (Peel et al., 2006). Two of the families within the Lepidoptera, including Pieridae (containing *Pieris rapae*), and Nymphalidae (*Junonia coenia*, *Vanessa cardui*, and *B. anynana*) have butterflies whose genomes and transcriptomes have at least two copies of *en*-family genes (Daniels et al., 2014; Challis et al., 2016; Connahs et al., 2016) with characteristic *En*-family conserved domains. *Inv* has the conserved RS and LSVG motifs present across holometabolous insects (Peel et al., 2006), and *En* proteins one or both *En*-specific motifs (Figure 2A, see Supplementary File for sequences and motifs).

The gene phylogeny suggested that the duplication of a common ancestor into *inv* and *inv-like* might have occurred independently in at least three different lineages: *B. anynana*, *Heliconius melpomene*, and *Plutella xylostella* (Figure 2C). This phylogenetic pattern, of *Inv-like* clustering with its paralog *Inv* in two of these lineages (*H. melpomene* and *B. anynana*) before clustering with its other orthologs (*Inv-like*), can also derive from the phenomenon of gene conversion (discussed below). Independent evidence using genomic copy position and orientation suggests that the duplication of these genes likely

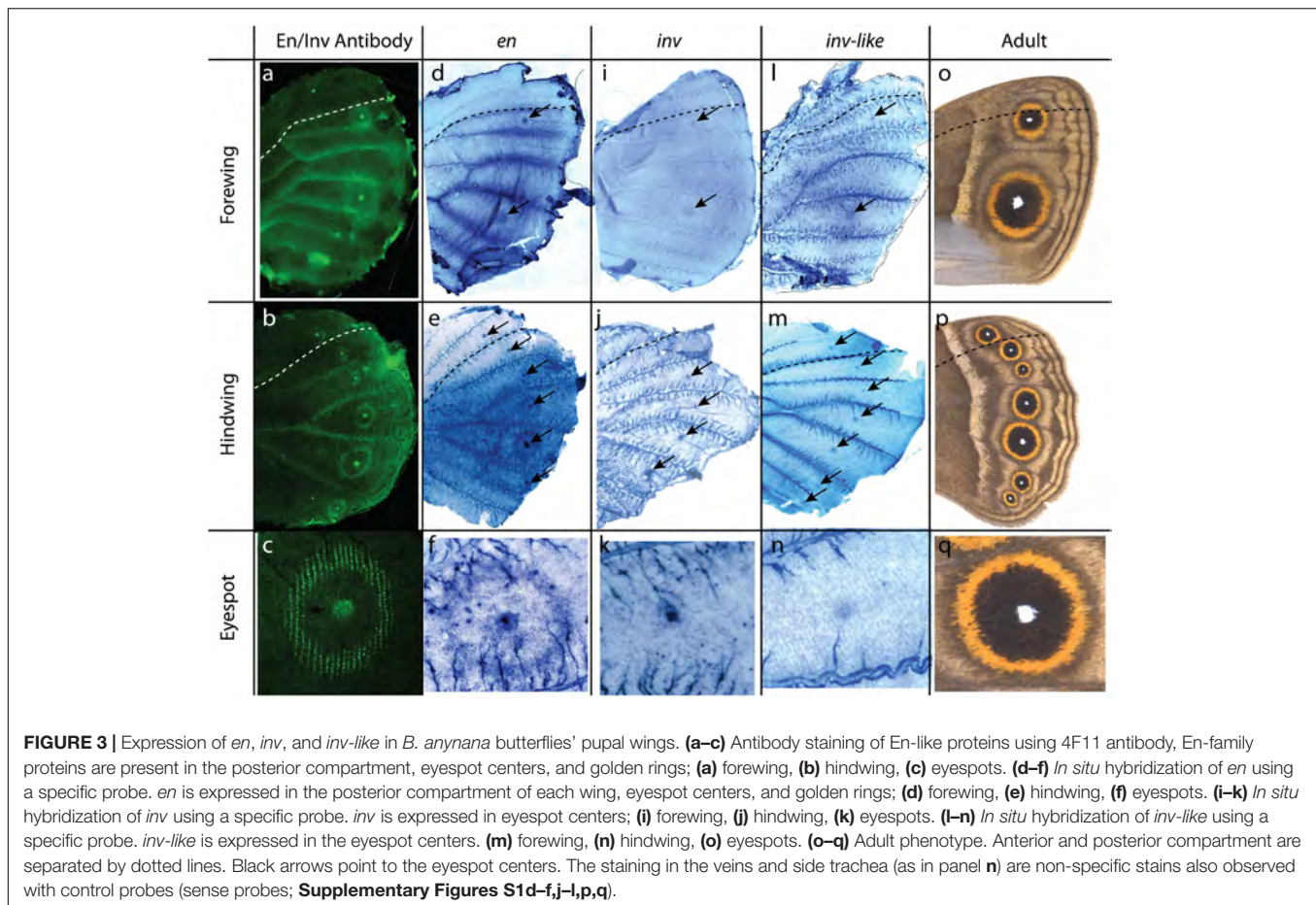
happened to a lineage ancestral to the divergence of all three species. The three *en*-like paralogs from *P. xylostella* map to contig unitig_1988 of *P. xylostella* pacbioV1 genome (Challis et al., 2016); and the same three paralogs from *H. melpomene* map to scaffold Hmel207002 of *H. melpomene* Hmel2 genome (Davey et al., 2015; Challis et al., 2016). All three paralogs map next to each other and in the same genomic orientation as in *B. anynana* (Supplementary Figures S6, S7, Supplementary Table S1, and Figure 1A; sequence in Supplementary File). The conserved genomic architecture suggests that this gene duplication event happened only once in a common ancestor to the three lineages, rather than in the three different lineages independently, and this must have happened at least 130 mya, which is an estimated date for the split of *P. xylostella* moths from the two butterflies obtained from a recent phylogenomic study that included fossil calibrations (Kawahara et al., 2019).

Concerted Evolution of *En*-Family Proteins in *B. anynana*

En-family proteins seem to have evolved under concerted evolution in *B. anynana* (as well as in *H. melpomene*). Gene duplicates that undergo concerted evolution are often found in close physical proximity and maintain high sequence similarity due to unequal crossing over and frequent gene conversion (Zhang, 2003). Due to concerted evolution high sequence similarity is maintained between *En*-family members which has often led to misleading phylogenetic trees (Abzhanov and Kaufman, 2000; Marie and Bacon, 2000; Peel et al., 2006). In particular, higher sequence similarity is maintained between paralogs within a species when compared to similarity between orthologs across species, giving a false phylogenetic signal of multiple duplication events having happened independently in each lineage (Liao, 1999). Specific conserved LSVG and RS motifs in *Inv* paralogs, however, appear to be immune to gene conversion allowing researchers to conclude that the divergence of *en* and *inv* happened before the divergence of hexapods (Peel et al., 2006). We have identified both LSVG and RS motifs in *B. anynana* and other Lepidopteran species (Figures 2A,B) but also observed clear signs of concerted evolution in that *B. anynana* *Inv* has more sequence similarity to *B. anynana* *En* than to *Inv* from *D. melanogaster* (Diptera) (see Supplementary File for sequence). The same phenomenon of gene conversion is also likely happening to *Inv* and *Inv-like* within Lepidoptera, as discussed above.

Overlapping and Distinct Expression Patterns of *en* Paralogs in the Pupal Wings of *B. anynana*

Studies on *en*-family genes in butterflies have been primarily carried out using the 4F11 monoclonal antibody that binds to a conserved domain of the translated *En* and *Inv* proteins of *D. melanogaster* (Patel et al., 1989). These *En*-family proteins are expressed in the posterior compartment and eyespot centers of larval and pupal wings (Keys et al., 1999; Banerjee and Monteiro, 2020) and in the cells that will differentiate the golden ring



of the eyespots in pupal wings of *B. anynana* (Brunetti et al., 2001), but until now it was not clear how many copies of *en-like* genes and transcripts were contributing to the observed multiple protein expression domains in this species. We aimed to find if one or more copies of *en-like* genes were exclusively associated with the novel eyespot color patterns and if these novel expression domains might have contributed to the preservation of *en* paralogs in the *B. anynana* genome.

In situ hybridization against each *en* paralog in the present study indicates that they are expressed in different patterns in the pupal wings of *B. anynana*. The protein expression patterns are likely the result of the expression of all three *en* paralogs (**Figure 3**). The expression in the posterior wing compartment and in the golden ring areas is specific to *en*, while all three paralogs are expressed in the eyespot centers.

Previous work established that both *en* and *inv* are expressed in the posterior compartment of butterflies during larval wing development (Carroll et al., 1994; Banerjee and Monteiro, 2020), however, we show here that later in development only *en* is present in the posterior compartment of pupal wings. No data is currently available for the expression of *inv-like* during the larval stages. Previous studies using *in situ* hybridization showed that *en* in *B. anynana* is homogeneously expressed in the posterior compartment, while *inv* is present in the

upper posterior compartment in *B. anynana* (Banerjee and Monteiro, 2020) and *Junonia coenia* (Carroll et al., 1994). In the pupal stage, we only observed the expression of *en* in the posterior compartment (**Figures 3d–f** and **Supplementary Figures S1a–c**). The loss of *inv* in the posterior compartment during pupal wing development is most likely because this gene is no longer needed to perform its earlier larval function of setting up posterior veins (Banerjee and Monteiro, 2020). Venation patterning happens during the early fifth instar (Banerjee and Monteiro, 2020), where both *en* and *inv* are likely involved in activating other downstream diffusible signals such as Hedgehog and Bone Morphogenic Proteins that position the veins (Banerjee and Monteiro, 2020). Late pupal expression of *en* in the wings of *D. melanogaster*, however, is involved in the maintenance of the posterior compartment, as mutations of *en* at these late stage leads to an enlarged posterior compartment with venation defects (Garcia-Bellido and Santamaria, 1972; Lawrence and Morata, 1976; Blair, 1992). The maintenance of *en* expression in pupal wings of butterflies, suggests that this gene may be playing a similar role.

The expression of all three *B. anynana* *en-family* genes in the eyespot centers indicates that all three paralogs might have a redundant function in the differentiation or signaling

from the eyespot center which sets the rings surrounding it, but this requires functional evidence (Figure 3) (Beldade and Brakefield, 2002; Monteiro, 2015). Many lineages of nymphalid butterflies lost expression of En-family protein in the eyespot center (as evidenced by lack of protein expression detected by the 4F11 antibody (Oliver et al., 2012)) and yet they develop eyespots. This suggests that En-family proteins might not be necessary for eyespot center differentiation and/or signaling (Oliver et al., 2012).

The expression of *en* in the golden ring area indicates that this gene alone might play a role in its differentiation, but again this requires functional validation. *B. anynana goldeneye* spontaneous mutants, where the golden ring extends inward into the area of the central black scales in the eyespots, display an equally extended expression of En-family proteins into that area (Saenko et al., 2008) but it is still unclear whether *en*, or some other gene equally affected by the *goldeneye* mutation is actually responsible for the change in scale colors in the central disk.

Alternative 3' UTR Splicing of *en* in *B. anynana*

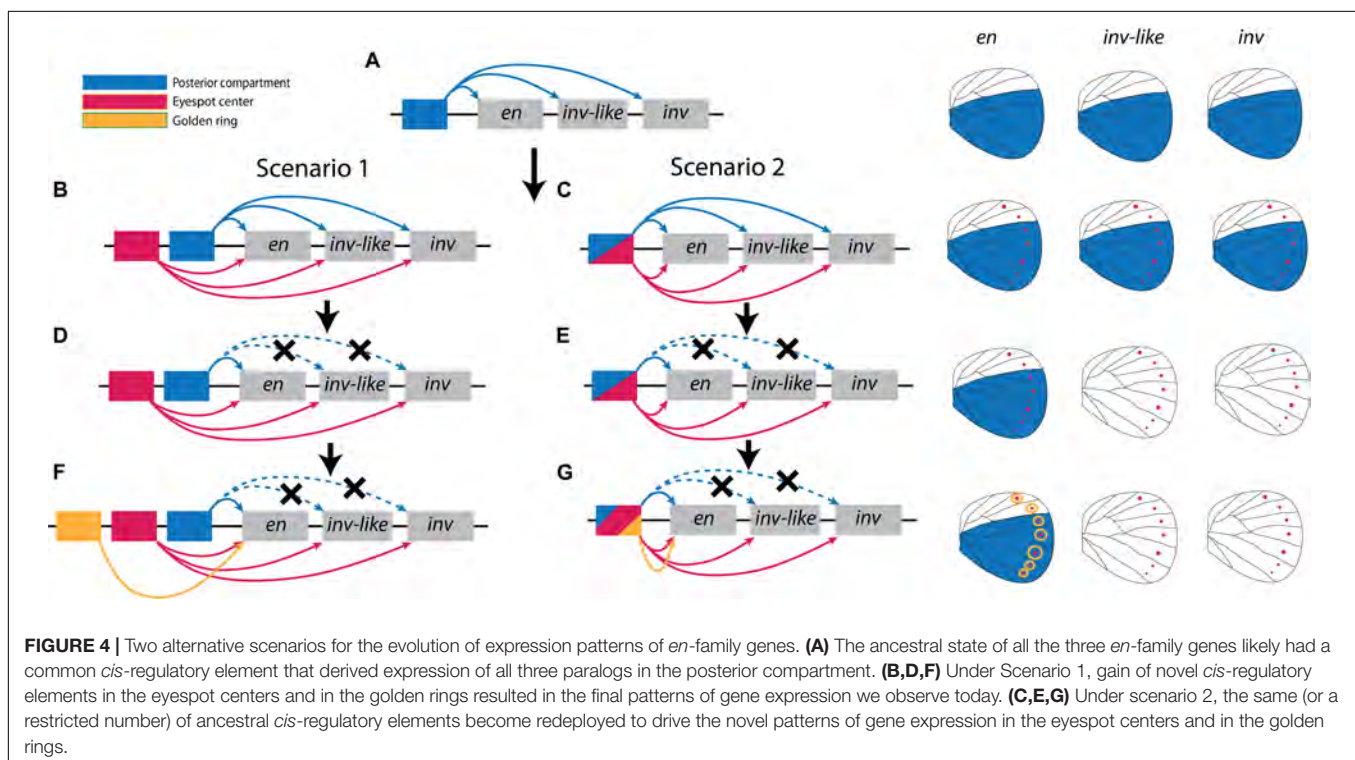
The different isoforms of *en* with varying 3' UTR length could be involved in controlling the level of proteins produced and their localization; and hence, they might also be playing a role in the differential expression of this paralog compared to *inv* and *inv-like*. Alternative 3' UTR lengths are achieved by the post-transcriptional process known as alternative cleavage and polyadenylation (APA) (Elkon et al., 2013). These 3' UTRs are known to control expression level and localization of proteins (Mayr and Bartel, 2009; Elkon et al., 2013; Mariella et al.,

2019). Shorter 3' UTRs are usually translated at higher rates, leading to higher protein levels (Mayr and Bartel, 2009). Future studies could focus on understanding the spatial and temporal expression of the alternative 3' UTR of *en* to investigate if they have diverged from each other.

en Paralogs and Eyespot Evolution

En-family proteins are differentially expressed in Nymphalid butterflies (Brunetti et al., 2001; Oliver et al., 2012). Eyespots are believed to have originated in a lineage sister to the sub-family Danainae, and now revised to have originated around 70 mya (Kawahara et al., 2019), via co-option which resulted in the expression of a set of genes in the eyespot centers, one of which includes En/Inv (Oliver et al., 2012). However, after the duplication, many lineages lost expression of En-family proteins in the eyespot centers (Oliver et al., 2012). *V. cardui*, for example, has expression in the golden rings but not in the eyespot centers (Brunetti et al., 2001), while the closely related *J. coenia* has expression of En-family proteins in the eyespot centers and the surrounding black scales, but not in the golden ring (Brunetti et al., 2001).

The *in situ* hybridization data on the pupal wings of *B. anynana* (Figure 3) reveal that two of the *en* paralogs (*inv* and *inv-like*) either lost two ancestral expression domains (posterior compartment and gold ring) or, alternatively, that one paralog (*en*) gained novel expression domains at these two locations after the duplication event. Given that *inv* is expressed in the posterior compartment of larval wings of *B. anynana*, but is absent in pupal wings, it might be more parsimonious to propose that *inv* might have had both



larval and pupal wing expression but subsequently lost its pupal expression.

We propose two equally parsimonious scenarios for the evolution of expression domains of *en* paralogs (Figure 4). The two scenarios are not exclusive. In both scenarios, the ancestral copy of all the present paralogs in *B. anynana* already had expression in the posterior compartment of the wing because this is the pattern of expression for these genes in *D. melanogaster* in the pupal stage (Blair, 1992), which shared a common ancestor with Lepidoptera around 290 mya (Misof et al., 2014) (Figure 4A). This would be facilitated if all copies shared a common set of regulatory elements, as is observed for the regulation of *en* and *inv* in *D. melanogaster* (Gustavson et al., 1989; Cheng et al., 2014). After eyespots originated, around 70 mya (Oliver et al., 2012) (this is a revised estimate based on a new phylogenetic study (Kawahara et al., 2019)), all paralogs gained a novel expression domain in the eyespot centers. This might have happened in a few different ways. A novel *cis*-regulatory element might have evolved in a genomic region that affects the expression of all *en*-copies simultaneously in the eyespot center (Scenario 1, Figure 4B), or a pre-existent *cis*-regulatory element that was already enhancing all copies simultaneously might have been re-used in a novel location (eyespot centers) due to the co-option of a gene regulatory network to that novel location, containing *en/inv* as genes positioned within the co-opted network (Monteiro and Podlaha, 2009; Monteiro and Gupta, 2016) (Scenario 2, Figure 4C). Subsequently, in both scenarios, the expression of *inv* and *inv-like* in the posterior compartment is lost (Figures 4D,E). Finally, either a new *cis*-regulatory element driving *en* expression in the golden ring is gained (Scenario 1, Figure 4F), or a pre-existent *cis*-regulatory element is reused for the novel *en* expression in the golden ring, again via gene network co-option (Scenario 2, Figure 4G). It is, however, also possible that the expression of *en* in association with the gold ring is gained in all three copies at once and subsequently lost in the *inv/inv-like* copies (not shown).

An important point to make here is that the timing of the first duplication event of *en* into *en* and *inv* paralogs happened around 480 mya or earlier (Peel et al., 2006; Misof et al., 2014), and the duplication *inv* into *inv* and *inv-like* happened no more recently than around 130 mya (see “Discussion” section above). Both of these duplications happened much earlier than the origin of eyespots, estimated to be around 70 mya (Oliver et al., 2012, 2014; Kawahara et al., 2019). This suggests that the duplication and maintenance of *en*-like paralogs in the genome of *B. anynana* as well as in the genome of other Lepidoptera is not particularly well associated with the origin of eyespots.

Future studies focused on the expression pattern of the *en* paralogs in model species such as *J. coenia* and *V. cardui*, as well as on the sequencing of open chromatin region surrounding the genes in *B. anynana* using techniques such as FAIRE and ATAC, will help us identify how the expression of *en*-like genes and their *cis*-regulatory elements diversified in the context of eyespot evolution. Functional studies will also be important to discover whether *en*-family

members are actually involved in the regulation of eyespots in nymphalid butterflies.

CONCLUSION

The present study provides the first evidence of the expression of multiple gene duplicates associated with the development of eyespots, a trait novel to nymphalid butterflies (Monteiro, 2015). By studying gene transcripts rather than conserved domains of *en/inv* protein we found that *en* paralogs are expressed in both overlapping and distinct domains. We proposed two models for the evolution of the expression pattern in *en* paralogs, where either loss or gain of novel expression domains resulted in differential expression of the paralogs in the wings on *B. anynana*. Furthermore, we showed that the duplication of the three genes happened much earlier than the origin of eyespots. Hence, this study does not lend support to the idea that these particular gene duplicates were retained in Lepidopteran genomes due to their involvement in the development of this particular evolutionary novelty. Future studies focused on understanding the interaction of *en* paralogs with each other as well as other known genes that play roles in eyespot development will continue to illuminate the evo-devo of eyespot patterns.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

TB, DR, and AM wrote the manuscript. TB performed bioinformatic analysis, *in situ* hybridization, and phylogenetic analysis. TB and AM performed antibody stainings. DR found the splice variants of *en*. TB, DR, and AM analyzed and visualized the results. All authors read and agreed to the final version of the manuscript.

FUNDING

This work was supported by the Ministry of Education, Singapore grant MOE2015-T2-2-159, the National Research Foundation, Singapore Investigatorship award NRF-NRFI05-2019-0006, and the Department of Biological Sciences LHK fund GL 710221. TB was supported by a Yale-NUS scholarship.

ACKNOWLEDGMENTS

We would like to thank Nipam Patel for the anti-Engrailed/Invected 4F11 antibody and Heidi Connahs for help with the transcriptome analysis. We would like to thank DBS-CBIS confocal facility for access to confocal microscopes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fevo.2020.00227/full#supplementary-material>

FIGURE S1 | Replicates of *en*, *inv*, *inv-like* and control *in situ* hybridization staining in *B. anynana* pupal wings. **(a–c)** Pupal hindwings stained with probe against *en*. The transcript is present in the posterior compartment, in the eyespot centers and in the eyespot golden ring area. **(g–i)** Pupal hindwings stained with probe against *inv*. The transcript is present only in the eyespot centers **(m–o)** Pupal hindwings stained with probe against *inv-like*. The transcript is present only in the eyespot centers. **(d–f)** Pupal wings stained with *en* sense probe (control). **(j–l)** Pupal wings stained with *inv* sense probe. **(p,q)** Pupal wings stained with *inv-like* sense probe. Non-specific staining is observed in the veins and trachea. Anterior-posterior compartments are separated by dotted lines. Black arrows point to eyespot centers.

FIGURE S2 | Phylogenetic tree of En-family proteins in Lepidoptera created using PhyML (Maximum likelihood. See “Materials and Methods” in the main text for details). The tree clusters En together; and Inv and Inv-like together. Vertical lines indicate mean number of amino acid substitutions per site. Numbers above branches represent branch support calculated based on 100 bootstraps.

FIGURE S3 | Phylogenetic tree of En-family proteins in Lepidoptera created using fasttree (Maximum likelihood. See the “Materials and Methods” section in the main text for details). The tree clusters En together; and Inv and Inv-like together.

REFERENCES

- Abzhanov, A., and Kaufman, T. C. (2000). Evolution of distinct expression patterns for *engrailed* paralogues in higher crustaceans (Malacostraca). *Dev. Genes Evol.* 210, 493–506. doi: 10.1007/s004270000090
- Banerjee, T. D., and Monteiro, A. (2020). Dissection of larval and pupal wings of *Bicyclus anynana* butterflies. *Methods Protoc.* 3:5. doi: 10.3390/mps3010005
- Banerjee, T. D., and Monteiro, A. (2020). Molecular mechanism underlying venation patterning in butterflies. *bioRxiv*. [Preprint]. doi: 10.1101/2020.01.02.892760
- Baratte, S., and Bonnaud, L. (2009). Evidence of early nervous differentiation and early catecholaminergic sensory system during sepia officinalis embryogenesis. *J. Comp. Neurol.* 519, 539–549. doi: 10.1002/cne.22174
- Beldade, P., and Brakefield, P. M. (2002). The genetics and evo-devo of butterfly wing patterns. *Nature* 3, 442–452. doi: 10.1038/nrg818
- Blair, S. S. (1992). Engrailed expression in the anterior lineage compartment of the developing wing blade of *Drosophila*. *Development* 115, 21–33.
- Braendle, C., Miura, T., Bicke, R., Shingleton, A. W., Kambhampati, S., Stern, D. L., et al. (2003). Developmental origin and evolution of bacteriocytes in the aphid-Buchnera symbiosis. *PLoS Biol.* 1:21. doi: 10.1371/journal.pbio.0000021
- Brakefield, P. M., Gates, J., Keys, D., Kesbeke, F., Wijngaarden, P. J., Monteiro, A., et al. (1996). Development, plasticity and evolution of butterfly eyespot patterns. *Nature* 384, 236–242. doi: 10.1038/384236a0
- Brockington, S. F., Yang, Y., Fernando, G.-H., Covshoff, S., Hibberd, J. M., Sage, R. F., et al. (2015). Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytol.* 207, 1170–1180. doi: 10.1111/nph.13441
- Brown, S. J., Patel, N. H., and Denell, R. E. (1994). Embryonic expression of the single tribolium engrailed homology. *Dev. Genet.* 15, 7–18.
- Brunetti, C. R., Selegue, J. E., Monteiro, A., French, V., Brakefield, P. M., and Carroll, S. B. (2001). The generation and diversification of butterfly eyespot color patterns. *Curr. Biol.* 11, 1578–1585. doi: 10.1016/s0960-9822(01)00502-4
- Carroll, S. B., Gates, J., Keys, D. N., Paddock, S. W., Panganiban, G. E., Selegue, J. E., et al. (1994). Pattern formation and eyespot determination in butterfly wings. *Science* 265, 109–114. doi: 10.1126/science.7912449

Vertical lines indicate mean number of amino acid substitutions per site. Numbers above branches represents branch support.

FIGURE S4 | Phylogenetic tree of En-family proteins in Lepidoptera created using Geneious tree builder (Jukes-Cantor as the genetic distance model, and UPGMA as tree build method). The tree clusters En together; and Inv and Inv-like together. Numbers above branches represent branch support calculated based on 1000 bootstraps.

FIGURE S5 | DNA sequence alignment of *en*, *inv* and *inv-like* and the regions used for probe design. Sequence used for *inv* probe has 24.9% identical bps to that of the *en* sequence, and 36.03% identical bps to that of the *inv-like* sequence. Sequence used for *inv-like* probe has 30.2% identical bps to that of the *en* sequence, and 36.8% identical bps to that of the *inv* sequence.

FIGURE S6 | Mapping of *en*, *inv* and *inv-like* in the genome of *P. xylostella* (pacbio1). Location of the transcripts was obtained from Leptabase.

FIGURE S7 | Mapping of *en*, *inv* and *inv-like* in the genome of *H. melpomene melpomene* (Hmel2). Location of the transcripts was obtained from Leptabase.

TABLE S1 | En scaffold features of *B. anynana*, *H. melpomene melpomene*, and *P. xylostella*.

TABLE S2 | Similarity matrix of highly conserved regions (EH4 and EH5) of *en*, *inv* and *inv-like* nucleotide sequences in **Supplementary Figure S5**.

TABLE S3 | Primer table.

TABLE S4 | *In situ* hybridization buffers.

TABLE S5 | Immunohistochemistry buffers.

- Challis, R. J., Kumar, S., Dasmahapatra, K. K., Jiggins, C. D., and Blaxter, M. (2016). Lepbase: The Lepidopteran genome database. *bioRxiv*. [Preprint]. doi: 10.1101/056994
- Cheng, Y., Brunner, A., Kremer, S., Devido, S., Stefaniuk, C. M., and Kassis, J. A. (2014). Co-regulation of invected and engrailed by a complex array of regulatory sequences in *Drosophila*. *Dev. Biol.* 395, 131–143. doi: 10.1016/j.ydbio.2014.08.021
- Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482
- Condron, B. G., Patel, N. H., and Zinn, K. (1994). Engrailed controls glial/neuronal cell fate decisions at the midline of the central nervous system. *Neuron* 13, 541–554. doi: 10.1016/0896-6273(94)90024-8
- Connahs, H., Rhen, T., and Simmons, R. B. (2016). Transcriptome analysis of the painted lady butterfly, *Vanessa cardui* during wing color pattern development. *BMC Genomics* 17, 1–16.
- Damen, W. G. M. (2002). Parasegmental organization of the spider embryo implies that the parasegment is an evolutionary conserved entity in arthropod embryogenesis. *Development* 129, 1239–1250.
- Daniels, E. V., Murad, R., Mortazavi, A., and Reed, R. D. (2014). Extensive transcriptional response associated with seasonal plasticity of butterfly wing patterns. *Mol. Ecol.* 23, 6123–6134. doi: 10.1111/mec.12988
- Davey, J., Chouteau, M., Barker, S. L., Maroja, L. S., Baxter, S. W., Simpson, F., et al. (2015). Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. *G3*, 6 695–708. doi: 10.1534/g3.115.023655
- Doyle, J. J., and Coate, J. E. (2019). Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* 180, 1–52. doi: 10.1086/700636
- Elkon, R., Ugalde, A. P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* 14, 496–506. doi: 10.1038/nrg3482
- Ferguson, L., Lee, S. F., Chamberlain, N., Nadeau, N., Joron, M., Baxter, S., et al. (2010). Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. *Mol. Ecol.* 19, 240–254. doi: 10.1111/j.1365-294x.2009.04475.x

- Fjose, A., Mcginnis, W. J., and Gehring, W. J. (1985). Isolation of a homoeo box-containing gene from the engrailed region of *Drosophila* and the spatial distribution of its transcripts. *Nature* 1, 284–289. doi: 10.1038/313284a0
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.
- Fujioka, M., Yusibova, G. L., Patel, N. H., Brown, S. J., and Jaynes, J. B. (2002). The repressor activity of Even-skipped is highly conserved, and is sufficient to activate engrailed and to regulate both the spacing and stability of parasegment boundaries. *Development* 129, 4411–4421.
- Garcia-Bellido, A., and Santamaria, P. (1972). Developmental analysis of the wing disc in the mutant engrailed of *Drosophila melanogaster*. *Genetics* 72, 87–104.
- Gibert, J. M., Mouchel-Vielh, E., Quéinnec, E., and Deutsch, J. S. (2000). Barnacle duplicate *engrailed* genes: divergent expression patterns and evidence for a vestigial abdomen. *Evol. Dev.* 2, 194–202. doi: 10.1046/j.1525-142x.2000.00059.x
- Guillén, I., Mullor, J. L., Capdevila, J., Sánchez-Herrero, E., Morata, G., and Guerrero, I. (1995). The function of *engrailed* and the specification of *Drosophila* wing pattern. *Development* 121, 3447–3456.
- Guindon, S., Jean-François, D., Lefort, V., Anisimova, M., Hordijk, W., and Gascue, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Gustavson, E., Goldsborough, A. S., Ali, Z., and Kornberg, T. B. (1989). The *Drosophila* engrailed and invected genes: partners in regulation. *Expr. Funct.* 906, 893–906.
- Han, K., and Manley, J. L. (1993). Functional domains of the *Drosophila* engrailed protein. *EMBO J.* 12, 2723–2733. doi: 10.1002/j.1460-2075.1993.tb05934.x
- Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* 33, 1635–1638. doi: 10.1093/molbev/msw046
- Hui, C. C., Matsuno, K., Ueno, K., and Suzuki, Y. (1992). Molecular characterization and silk gland expression of Bombyx engrailed and invected genes. *Proc. Natl. Acad. Sci. U.S.A.* 89, 167–171. doi: 10.1073/pnas.89.1.167
- Kawahara, A. Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E. F. A., Donath, A., et al. (2019). Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc. Natl. Acad. Sci. U.S.A.* 116, 22657–22663. doi: 10.1073/pnas.1907847116
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Keys, D. N., Lewis, D. L., Selegue, J. E., Pearson, B. J., Goodrich, L. V., Johnson, R. L., et al. (1999). Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science* 283, 532–534. doi: 10.1126/science.283.5401.532
- Lawrence, P. A., and Morata, G. (1976). Compartments in the wing of *Drosophila*: a study of the engrailed gene. *Dev. Biol.* 50, 321–337. doi: 10.1016/0012-1606(76)90155-x
- Li, Z., Tiley, G. P., Galuska, S. R., Reardon, C. R., Kidder, T. I., Rundell, R. J., et al. (2018). Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci. U.S.A.* 115:253609.
- Liao, D. (1999). Concerted evolution: molecular mechanism and biological implications. *Am. J. Hum. Genet.* 64, 24–30. doi: 10.1086/302221
- Logan, C., Hanks, M. C., Noble-Topham, S., Nallainathan, D., Provart, N. J., and Joyner, A. L. (1992). Cloning and sequence comparison of the mouse, human, and chicken engrailed genes reveal potential functional domains and regulatory regions. *Dev. Genet.* 13, 345–358. doi: 10.1002/dvg.1020130505
- Lynch, M., O'Hely, M., Walsh, B., and Force, A. (2001). The probability of preservation of a newly arisen gene duplicate. *Genetics* 159, 1789–1804.
- Macias-Munoz, A., Smith, G., Monteiro, A., and Briscoe, A. D. (2016). Transcriptome-wide differential gene expression in bicyclus anynana butterflies: female vision-related genes are more plastic. *Mol. Biol. Evol.* 33, 79–92. doi: 10.1093/molbev/msv197
- Manzanares, M., Marco, R., and Garesse, R. (1993). Genomic organization and developmental pattern of expression of the engrailed gene from the brine shrimp *Artemia*. *Development* 118, 1209–1219.
- Marie, B., and Bacon, J. P. (2000). Two engrailed-related genes in the cockroach: cloning, phylogenetic analysis, expression and isolation of splice variants. *Dev. Genes Evol.* 210, 436–448. doi: 10.1007/s004270000082
- Marie, B., and Blagburn, J. M. (2003). Differential roles of engrailed paralogs in determining sensory axon guidance and synaptic target recognition. *J. Neurosci.* 23, 7854–7862. doi: 10.1523/JNEUROSCI.23-21-07854.2003
- Mariella, E., Marotta, F., Grassi, E., Gilotto, S., and Provero, P. (2019). The length of the expressed 3' UTR is an intermediate molecular phenotype linking genetic variants to complex diseases. *Front. Genet.* 10:714. doi: 10.3389/fgene.2019.00714
- Mayr, C., and Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138, 673–684. doi: 10.1016/j.cell.2009.06.016
- Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, K., Mayer, C. U., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767.
- Monteiro, A. (2015). Origin, development, and evolution of butterfly eyespots. *Annu. Rev. Entomol.* 60, 253–271. doi: 10.1146/annurev-ento-010814-020942
- Monteiro, A., Glaser, G., Stockslager, S., Glansdorp, N., and Ramos, D. (2006). Comparative insights into questions of lepidopteran wing pattern homology. *BMC Dev. Biol.* 6:52.
- Monteiro, A., and Gupta, M. D. (2016). *Identifying Coopted Networks and Causative Mutations in the Origin of Novel Complex Traits*. *Current Topics in Developmental Biology*. Berlin: Elsevier Inc, doi: 10.1016/bs.ctdb.2016.03.003
- Monteiro, A., and Podlaha, O. (2009). Wings, horns, and butterfly eyespots: how do complex traits evolve? *PLoS Biol.* 7:e37. doi: 10.1371/journal.pbio.1000037
- Nederbragt, A. J., Loon, E. V., and Dictus, W. J. A. G. (2002). Expression of patella vulgata orthologs of engrailed and dpp-BMP2 / 4 in adjacent domains during molluscan shell development suggests a conserved compartment boundary mechanism. *Dev. Biol.* 355, 341–355. doi: 10.1006/dbio.2002.0653
- Nowell, R. W., Elsworth, B., Oostra, V., Zwaan, B. J., Wheat, C. W., and Saastamoinen, M. (2017). A high-coverage draft genome of the mycalesine butterfly *Bicyclus anynana*. *Gigascience* 6, 1–7.
- Oliver, J. C., Beaulieu, J. M., Gall, L. F., Piel, W. H., and Monteiro, A. (2014). Nymphalid eyespot serial homologues originate as a few individualized modules. *Proc. R. Soc. B Biol. Sci.* 281:20133262. doi: 10.1098/rspb.2013.3262
- Oliver, J. C., Tong, X. L., Gall, L. F., Piel, W. H., and Monteiro, A. A. (2012). Single origin for nymphalid butterfly eyespots followed by widespread loss of associated gene expression. *PLoS Genet* 8:e1002893. doi: 10.1371/journal.pgen.1002893
- Olson, E. N. (2006). Gene regulatory networks in the evolution and development of the heart. *Science* 313, 1922–1927. doi: 10.1126/science.1132292
- Özsu, N., and Monteiro, A. (2017). Wound healing, calcium signaling, and other novel pathways are associated with the formation of butterfly eyespots. *BMC Genomics* 18, 1–14.
- Patel, N. H., Martin-Blanco, E., Coleman, K. G., Poole, S. J., Ellis, M. C., Kornberg, T. B., et al. (1989). Expression of engrailed proteins in arthropods, annelids, and chordates. *Cell* 58, 955–968. doi: 10.1016/0092-8674(89)90947-1
- Peel, A. D., Telford, M. J., and Akam, M. (2006). The evolution of hexapod engrailed-family genes: evidence for conservation and concerted evolution. *Proc. R. Soc. B Biol. Sci.* 273, 1733–1742. doi: 10.1098/rspb.2006.3497
- Peltenburg, L. T., and Murre, C. (1996). Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J.* 15, 3385–3393. doi: 10.1002/j.1460-2075.1996.tb00704.x
- Peterson, M. D., Popadić, A., and Kaufman, T. C. (1998). The expression of two *engrailed*-related genes in an apterygote insect and a phylogenetic analysis of insect engrailed-related genes. *Dev. Genes Evol.* 208, 547–557. doi: 10.1007/s004270050214
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077
- Saenko, S. V., French, V., Brakefield, P. M., and Beldade, P. (2008). Conserved developmental processes and the formation of evolutionary novelties: examples from butterfly wings. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363, 1549–1555.

- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Theißen, G. (2001). Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* 4, 75–85. doi: 10.1016/s1369-5266(00)00139-4
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Tolkunova, E. N., Fujioka, M., Kobayashi, M., Deka, D., and Jaynes, J. B. (1998). Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. *Mol. Cell. Biol.* 18, 2804–2814. doi: 10.1128/mcb.18.5.2804
- Wagner, G. P., and Lynch, V. J. (2010). Hox cluster duplications and the opportunity for evolutionary novelties. *Curr. Biol.* 20, 14603–14606. doi: 10.1073/pnas.2536656100
- Whittington, P. M., Meier, T., and King, P. (1991). Segmentation, neurogenesis and formation of early axonal pathways in the centipede, *Ethmostigmus rubripes* (Brandt). *Roux's Arch. Dev. Biol.* 199, 349–363. doi: 10.1007/bf01705928
- Yokoyama, S., and Yokoyama, R. (1989). Molecular evolution of human visual pigment genes. *Mol. Biol. Evol.* 6, 186–197.
- Zattara, E. E., Busey, H. A., Linz, D. M., Tomoyasu, Y., and Moczek, A. P. (2016). Neofunctionalization of embryonic head patterning genes facilitates the positioning of novel traits on the dorsal head of adult beetles. *Proc. R. Soc. B Biol. Sci.* 283:20160824. doi: 10.1098/rspb.2016.0824
- Zhang, J. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298. doi: 10.1016/s0169-5347(03)00033-8
- Zhang, J., Rosenberg, H. F., and Nei, M. (1998). Positive darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3708–3713. doi: 10.1073/pnas.95.7.3708
- Zhang, S., Zhang, J. S., Zhao, J., and He, C. (2015). Distinct subfunctionalization and neofunctionalization of the B-class MADS-box genes in *Physalis floridana*. *Planta* 241, 387–402. doi: 10.1007/s00425-014-2190-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Banerjee, Ramos and Monteiro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.