UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Genetics of diversification: a hotspot locus for pigmentation evolution

Mestrado em Biologia Evolutiva e do Desenvolvimento

Carolina Martins Peralta Lopes da Silva

Dissertação orientada por:
Patrícia Beldade e Sara Magalhães

2015

Acknowledgments

Esta tese representa o término de uma etapa que foi, sem qualquer dúvida, importante para a minha aprendizagem tanto a nível profissional como pessoal. Quero agradecer a todas as pessoas que tive o prazer de conhecer, tanto aquelas com quem ainda mantenho contacto ou não, por toda a partilha de conhecimento e discussões de ideias que me ajudaram a chegar até este ponto. Um especial obrigado à Marta, João, Alexa, Ioana, Zé e Sardinha pela amizade nos últimos anos. Não sei o que faria sem vocês! À Tonecas, Bia e Ré por me ajudarem a pôr os pés na terra, obrigada pelos bons conselhos. À Diana por me conhecer melhor que ninguém, por compreender e ouvir os meus desabafos. À Susana, Sofia, Rita e Migueis por me obrigarem a abstrair do mundo da Biologia e pelo esforço em perceberem o meu mundo quando tal não era, de todo, possível. À Sofia, Diogo e Rúben pela amizade durante os últimos dois anos. Ao pessoal da HortaFCUL por, ao fim de um ano fora, ainda me fazerem sentir em casa.

Quero agradecer especialmente à Patrícia por me ter acompanhado durante este ano. Pela exigência, pela dedicação, pela paciência, por tudo aquilo que me ensinou e que será importante para o resto da minha carreira como investigadora. À Sara pela disponibilidade, interesse neste projecto e revisão.

A todo o pessoal do IGC, pela entreajuda que me surpreendeu e ainda hoje me surpreende. Um sincero obrigado à Elvira e à Marta pela amizade, apoio, por todo o acompanhamento, por tudo o que me ensinaram. À Tânia, Ana, Julien, Luís, Gonçalo, Andreia e a todos por fazerem deste ano tão agradável. Ao Pedro pela ajuda com as borboletas e disponibilidade sempre que necessário e ao Diogo por me ensinar e ajudar a injectar as pequenas pérolas.

A special thank you to my *in situ* mentor Kohtaro whose help and discussions were invaluable to this work. To Arnaud for is visit, amazing ideas and help injecting *Bicyclus* embryos. To Suzanne, whom I owe the pleasure of continuing such amazing work and for being available whenever I needed. To the barefoot man, Erik, that randomly appeared: thank you for all the good conversations that we had and for reviewing this thesis.

Ao Rui, por nunca me ter falhado. Por estar sempre do meu lado. Por tudo. Não teria conseguido sem ti.

Last but not least, agradeço a quem tenho o prazer de partilhar o mesmo património genético, aka, a minha família. Um especial e profundo obrigado aos meus pais (que nunca será suficiente) que sempre me apoiaram em tudo. Obrigada.

Resumo

Um dos grandes desafios da biologia evolutiva é caracterizar os *loci* e as mutações que contribuem para a variabilidade fenotípica de características adaptativas. Existem exemplos clássicos que demonstram a base genética de determinadas características em que o seu papel adaptativo é conhecido, nomeadamente as manchas nas asas da mosca da fruta, os espinhos das barbatanas dos peixes esgana-gata ou a pigmentação do pelo de ratinhos. Contudo, questões como "quais as classes e regiões génicas que contribuem para a variabilidade fenotípica?", "serão os *loci* e os alelos que contribuem para polimorfismos intraespecíficos os mesmos que estão por detrás da divergência entre espécies?" ou "serão os *loci* que possuem alelos de grande efeito estudados em laboratório os mesmos que contribuem para a variabilidade em populações naturais?" ainda estão por resolver de forma definitiva. Utilizar diferentes espécies e diferentes caracteres adaptativos é essencial para adquirir conhecimento suficiente para obter respostas gerais a estas questões.

A pigmentação dos insetos, por exemplo, é uma característica que tem uma diversificação surpreendente e que se sabe ser importante para comunicação visual, termorregulação e até mesmo resistência a agentes patogénicos. O facto de diferentes estudos já terem conseguido fazer a ligação entre variação fenotípica, variação genotípica e o processo de desenvolvimento desta característica, faz dela um óptimo alvo para estudar a base genética da evolução adaptativa. Os padrões de cor nas asas das borboletas, em particular, têm uma diversificação extraordinária, com cores e padrões tanto divergentes como convergentes entre diferentes espécies pelo mundo fora. Aliado ao facto das asas serem um tecido em que diferentes estudos do desenvolvimento são possíveis, fazem deles um dos casos mais interessantes da área de biologia evolutiva e do desenvolvimento.

Estudos distintos demonstraram que o melanismo nas traças *Biston betularia*, o mimetismo nas borboletas *Heliconius* e os padrões reminiscentes de olhos (chamados daqui em diante, *eyespots*) nas borboletas *Bicyclus anynana* são padrões com um grande valor adaptativo na ecologia destas espécies, nomeadamente por estarem envolvidos em evitar predação. Recentes estudos independentes, que estudavam variação intraespecífica nestes lepidópteros, implicaram a mesma região genómica em variação entre diferentes espécies. Como esta região está subjacente à variação adaptativa em espécies distintas, com padrões de asas também eles distintos, foi proposto chamar-lhe um locus-chave, um *hotspot,* para a diversificação. Perceber qual a organização genética deste locus e o papel das regiões regulatórias do mesmo, irá elucidar de que forma mudanças evolutivas em mecanismos genéticos do desenvolvimento levam à variação morfológica.

Neste trabalho, focámo-nos no *locus*-chave da borboleta *B. anynana*. Os padrões das asas destas borboletas, os *eyespots*, são uma característica que tem vindo a ser bastante estudada nos últimos vinte anos. Esta espécie é um bom organismo modelo para tentar responder às questões acima referidas uma vez que existe conhecimento acerca da sua ecologia, uma vasta disponibilidade de mutantes em laboratório e técnicas para estudar o seu desenvolvimento. Vários estudos mostraram quais os genes que estão por de trás do desenvolvimento destas estruturas, sendo a maioria deles conservados no desenvolvimento embrionário e tendo sido co-optados (desempenham novas funções) nomeadamente para a padronização das asas. Um dos modelos propostos para explicar a morfologia destas estruturas é a existência de um morfogéneo (por exemplo, Wingless), libertado do centro do *eyespot* (chamado *focus*) que vai ser "lido" pelas células que estão em volta, de uma maneira dose-dependente. Consequentemente, nessas células, genes como *engrailed* e *distaless* vão ser activados e terão

um papel preponderante na activação das vias biossintéticas de produção de pigmentos de diferentes cores.

O *locus*-chave da *B. anynana* em estudo não contém qualquer gene antes implicado no desenvolvimento dos padrões das asas desta espécie. Este locus foi denominado BFS uma vez que os seus alelos foram implicados em três mutantes que apareceram espontaneamente no laboratório: Bigeye, Frodo e Spread. Comparando com o "estado selvagem" (daqui em diante denominado *wild-type*, WT), os três mutantes dão jus ao seu nome: Bigeye possuí eyespots maiores, Frodo possui o anel dourado maior e finalmente Spread tem alterações tanto na composição da cor (as escamas amarelas cobrem quase completamente as escamas pretas) como também no tamanho dos *eyespots*. O facto destas duas características serem definidas por processos diferentes, nomeadamente a força do sinal dado pelo *focus* e a sensibilidade de resposta a esse sinal por parte das células, respetivamente, faz com que haja um grande interesse em perceber quais as características génicas existentes no locus que possam influenciar tal variedade de componentes morfológicos. Foi demonstrado também que alelos deste locus são recessivos letais e estão implicados na polaridade segmentar da embriogénese de *B. anynana*. Comparando os defeitos morfológicos destes embriões e os padrões de expressão de genes conservados no desenvolvimento embrionário como engrailed, foi proposto que estes efeitos pleiotrópicos no desenvolvimento embrionário e na morfologia do *eyespot* são causados por um regulador negativo da via de sinalização Wingless.

Após um mapeamento exaustivo na busca dos genes associados aos fenótipos mutantes, foi identificada uma região genómica de interesse de cerca de 40kb que inclui um único gene, *washout,* que está imediatamente a montante do gene *domeless*. *washout* é um regulador do citoesqueleto de actina enquanto *domeless* é o único receptor conhecido que activa a via de sinalização JAK-STAT. Dado que ambos estes genes estão associados ao desenvolvimento embrionário da mosca da fruta, e que alelos do locus BFS causam distúrbios durante a embriogénese, ambos são bons candidatos para a região genómica por de trás dos fenótipos do BFS. *domeless*, é particularmente um bom candidato pois faz parte de uma via de sinalização que interage com a via de sinalização de Wingless. A existência de uma região não-codificante que esteja a regular a expressão de genes relacionados com os efeitos pleiotrópicos observados é também uma hipótese que tem de ser tida em conta.

Neste estudo mostrámos que ambos os genes são expressos durante o desenvolvimento embrionário e das asas das larvas, de uma forma consistente com o seu envolvimento nestes processos. Identificámos também variação nucleotidíca nestes genes existente entre os diferentes alelos mutantes e discutimos de que forma pode estar a influenciar os fenótipos descritos. Finalmente, tentámos implementar uma técnica que irá permitir fazer uma análise funcional de genes ou alelos em organismos que não são modelos de laboratório clássicos, como é o caso da *B. anynana*. Apesar de apenas termos resultados preliminares no que toca à tentativa de causar uma disrupção nestes genes, quando optimizada, esta técnica irá permitir resolver a identidade do BFS locus. Num futuro próximo, esperamos que o estudo deste locus e dos seus ortólogos traga conhecimento acerca da base genética para a formação de padrões de asas de borboletas e que ajude a estabelecer princípios gerais acerca da base genética da variação e diversificação de características adaptativas.

Palavras-chave: Genética da diversificação, padrões de cores das borboletas, *domeless*, *washout,* CRISPR/CAS9

Abstract

Unravelling the nature and effects of the loci underlying adaptive evolution is a key challenge in evolutionary biology, now made easier by increasingly accessible tools and methods. Lepidopteran wing pigmentation patterns are highly diversified and include several models in studies of evolution and development. The fact that there is knowledge on both their adaptive value and underlying development makes them well suited to study the genetic basis of adaptive evolution. Recent independent studies aimed at identifying the loci responsible for intra-specific variation in different lepidopteran species converged into an orthologous genomic region. This locus was proposed to be a "hotspot" for pigmentation evolution. The *Bicyclus anynana* orthologue of the hotspot locus, called BFS locus, carries allelic variation implicated in wing pattern mutants, affecting eyespot morphology and also embryonic development. By means of linkage mapping, the genomic region to where the BFS locus locates was narrowed down to approximately 40 kb containing the gene *washout* and just outside the implicated region, the gene *domeless.* This project aimed to test both candidate genes by: 1) describing expression patterns during embryonic and wing development, 2) associate nucleotide differences to the segregating genotypic classes, 3) attempting a functional analysis to characterize the potential role of these genes in wing pattern formation. We found that both genes are expressed during embryonic and wing development in a manner consistent with their contribution to BFS phenotypes. In addition, we found nucleotide differences that may be causative for mutant phenotypes. Continued work on the functional analysis is needed to implicate the genes in eyespot development and phenotypic variants. This type of integrative analysis can shed needed light on the origin and diversification of adaptive novel traits.

Keywords: Genetics of diversification, butterfly colour patterns, *domeless*, *washout,* CRISPR/CAS9

Contents

I. Introduction

1. State of the art

One of the most challenging aims of evolutionary biology is to characterize the loci and individual mutations that contribute to phenotypic variation in adaptive traits[1]. Different study species and traits have emerged in the quest to unravel the genetic basis of adaptive evolution; some of them in classical lab models such as *Drosophila* including wing spots[2] and sex combs[3] and others in less well established models such as stickleback spines[4] and beach mice pigmentation[5].

Which gene classes (e.g. transcription factors or enzymes) and gene regions (coding or regulatory[6] contribute to phenotypic variation? Are the same loci and alleles that contribute to intra-specific polymorphism those underling species divergence[7]? Are the loci that carry mutant alleles of large effect analysed in the lab the same that contribute to variation in natural populations[8]? These are some of the not yet fully resolved questions that highlight the need for more studies in different species and different traits, especially those for which we have knowledge from laboratory analysis and ecological relevance.

One compelling example is insect pigmentation, a trait of striking phenotypic diversification that has well described adaptive role, not only in visual communication (for mating and predator avoidance) but also in thermoregulation and pathogen resistance[9,10]. Its contribution to linking variation in genotypes to variation in phenotypes and developmental processes makes this trait well suited to study the genetics basis of adaptive evolution[11]. Lepidopteran wing coloration, in particular, is one of the most interesting case studies of evolutionary developmental biology, with an astonishing intra- and inter-specific diversity allied to tractability for developmental studies. Well known studies of evolutionary adaptive role of lepidopteran wing patterns include melanism and its reversal in *Biston betularia* moths[12], mimicry in *Heliconius* butterflies[13] and evo-devo of eyespots in *Bicyclus anynana* butterflies[14].

Independent studies pursing the identification of the loci responsible for intra-specific variation on the species mentioned above, converged into a single genomic region implicated in the different species. Although lepidopteran species are rather under-represented in terms of genomic resources (except for the model silkworm, *Bombix mori,* that has a fully sequenced and annotated genome[15]), efforts to gene annotation and methods to assign genes to genetic maps have been growing[16]. In *B. betularia*, the construction of a linkage map to identify the chromosomal region containing the locus controlling the *carbonaria*-typical polymorphism (melanic form) mapped the morph to a linkage group (LG) that is orthologous with a high level of synteny with *B. mori* chromosome 17[17]. Curiously, the orthologous chromosome in the butterfly *Heliconius melpomene* contains HmYb, a locus responsible for colour pattern variation implicated in *Heliconius erato* and *Heliconius numata*[18]. Finally, three pleiotropic mutations with large effect on the *B. anynana* eyespot morphology, were mapped to a single locus - the BFS locus[19]. As in the latter examples, this locus was assigned to the orthologous region of LG 17. Together, these findings led to the suggestion that this genomic region and pigmentation loci therein behaved like genomic hotspots for diversification of lepidopteran pigmentation[20]. This supports the idea that developmental and genetic constrains frame evolutionary change in a way that only a small number of loci have a significant potential to underlie morphological diversity. Also, reports show that evolutionary relevant mutations tend to accumulate in a particular set of genes, often in cis-regulatory gene regions[7]. Studies of specific genes underlying parallel and convergent evolution are of particular interest since they can reveal potential adaptive hotspots. Such examples include the gene *yellow* for abdominal and wing pigmentation of *Drosophila* species [21], *bmp4* for feeding strategies in cichlid fishes[22] and Darwin's finches[23] and finally the *svb/ovo* regulatory region that explains multiple cases of convergence in Drosophila larval bristles[24] and evolution of excretory duct morphology in *Caenorhabditis elegans*[25].

Butterfly wing pattern genes are a powerful system for assessing potential adaptive hotspots, not only because they play a key role in adaptation but also because morphological variation is well known in this system[20]. Understanding the genetic organization of the recently found hotspot locus for wing patterning in lepidopterans will elucidate how changes in developmental mechanisms can lead to morphological variation.

2. *Bicyclus anynana* and eyespot evo-devo

The eyespots in *B. anynana* wings are a valuable system for study evolutionary and developmental processes that shape morphological variation. The presence of high phenotypic diversity, detailed description of its adaptive role in ecology and evolution, together with the availability of lab mutants, growing knowledge in genetics and development underlying eyespot morphology, makes *B. anynana* an excellent model to study evolution and development[14]. Of all pattern elements found in butterflies, eyespots are those whose underlying development is best understood (reviewed in[14]).

Butterfly wings are composed by two epidermal sheets that are supported by veins. Wing development starts during larval life, during which there is increase in size, venation system building, and finally, the onset of wing patterning including the establishment of eyespot foci (the eyespot centre that has organizing properties for the formation of this pattern). Throughout the pupal stage, the scale maturation and pigment deposition occurs and colour patterns on the wing surfaces are formed by arrangement of monochromatic scales on a single cell layer.

Different studies on gene expression patterns in larval and pupal wings implicated a number of genes in eyespot development. A candidate gene approach based on *Drosophila melanogaster* knowledge about wing development showed that conserved genes such as *engrailed* (*en)* and *Distaless* (Dll) perform similar functions in defining compartments in *B. anynana* and play a role in different stages of eyespot formation (*e.g.*[26]). The first stage of eyespot formation is the establishment of the eyespot in a particular location, which occurs in the final larval instar. Conserved members of Notch, Hedgehog, Wingless and TGF-B pathways have been found to be expressed in and around eyespot foci at this stage (reviewed in[27]). Exactly how all these genes interact and regulate each other and their downstream pathways is not known yet. One of the proposed models (reviewed in[27]) to explain how the eyespot ring establishment occurs suggests that these genes act upstream of Wingless (Wg) and Decapentaplegic (Dpp), potentially acting as the ring focus-derived inducing signals (focal signal) during pupal wing development (although functional tests are required to confirm this model). Depending on the strength of the focal signal, eyespot size can vary whereas colour composition is known to involve variability in the response threshold to the focal signal[14]. In the cells surrounding the centre of the eyespot (*focus*), genes that activate the pigment biosynthetic pathway, such as *en* (expressed in the outer, golden ring) and Dll and *sal* (inner, black ring) are potentially upregulated after cellular exposure to the focal signal.

This project aimed to resolve the identity of a locus implicated in variation in eyespot morphology (size and colour composition) which is an orthologous region of *B. betularia* and *Heliconius sp*. butterflies colour pattern loci.

3. The *B. anynana* BFS locus

The BFS locus was named after the three mutant phenotypes associated to it: Bigeye, Frodo and Spread[19]. Compared to wild-type (WT) butterflies, Bigeye (BE) mutants have dramatically enlarged eyespots while Frodo (Fro) mutants have eyespots with a broader outer golden ring. The eyespots of Spread (Spr) are very large and the golden scales almost completely substitute the black scales (Fig. 1a). Experimental crosses between individuals of the same mutant

phenotype (see Fig. 1b left and middle panels) resulted in one fourth of the offspring dying during embryogenesis and two thirds of adults having aberrant eyespots. This segregation is consistent with a recessive lethal allele with dominant effects on eyespot morphology. The embryos that segregate in crosses and died during embryogenesis (see Fig. 1b left and middle panels) showed the same morphological defects, which suggests that the three mutations were alleles at the same locus. This was confirmed by complementation tests: crosses between BE and either Fro or Spr still yielded embryonic lethality of one fourth of the progeny, consistent with these all being alleles at the same locus[19].

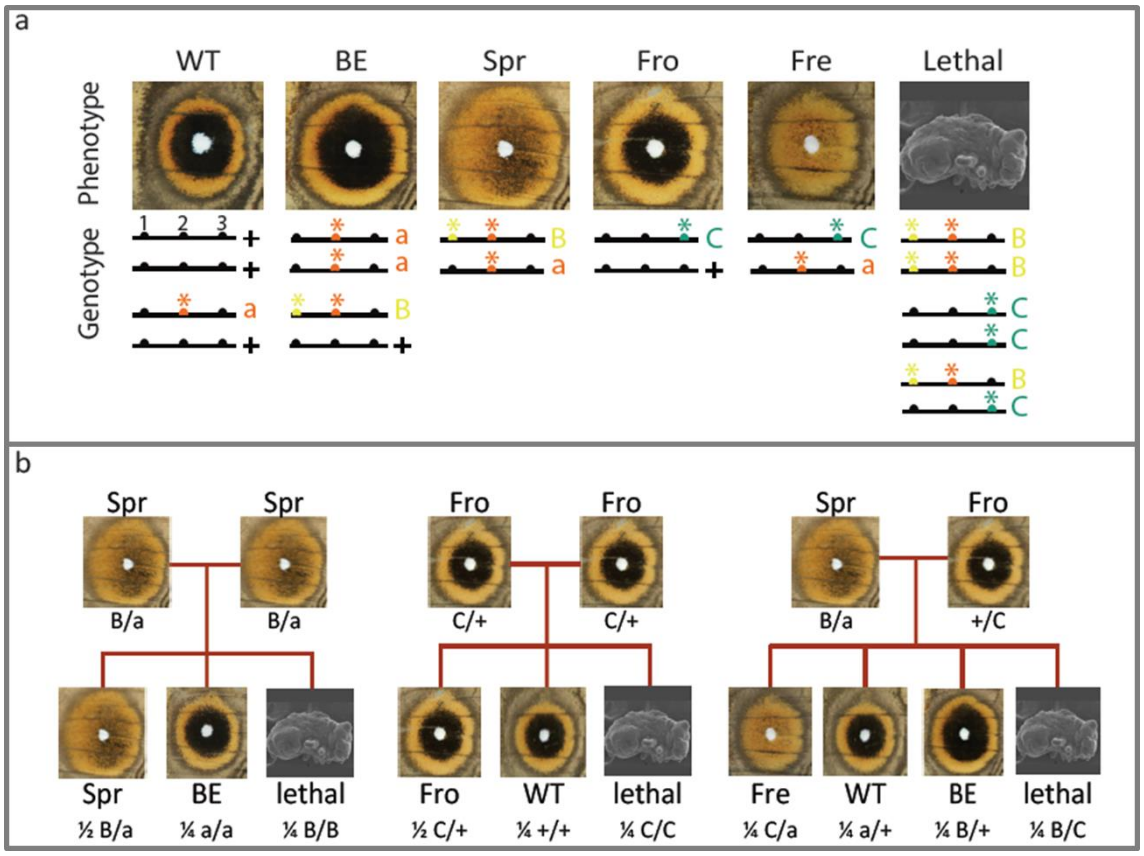


Figure 1. Model for the BFS locus effects and segregation of mutant phenotypes. a. Current model for allelic variation for BFS locus (*cf.*[19]). Eyespot phenotypes and corresponding genotypes are shown, wherein lines represent the locus and dots the proposed nucleotide positions (123, order and distance are arbitrary) with sequence variation (*cf.*[19]). Stars of different colours represent the different variants: black for WT allele (+), and yellow, orange or blue for mutations in first, second or third site, respectively. Based on segregation analysis, different alleles (a, B or C) were proposed and are represented with different letters and colours. The a allele (1*3) does not have an obvious effect on eyespot morphology (WT phenotype) when in heterozygosity with the + allele (123) but affects eyespot size (BE phenotype) when in homozygosity. A mutation in the first and second sites (**3) characterize the B allele and have a dominant effect on eyespot size: a single copy produces a BE phenotype. The a and B alleles together (1*3/**3) affect both eyespot size and colour composition (Spr phenotype). The C allele (12*) has a dominant effect on eyespot colour: together with the + allele, it produces the Fro phenotype, and together with the a allele, produces the Fre phenotype. Three genotypes result in embryonic lethality: homozygous and heterozygote of B and C alleles. Figure adapted from[19]. **b**. Maintenance of lab stocks (Spr in left panel, Fro in middle panel, and Fred in right panel). In all these crosses, one fourth of the progeny dies during embryogenesis, as expected for recessive lethal alleles, and the adult progeny falls into two (left and middle panels) or three (right panel) phenotypic classes.

It was shown that lethal alleles at the BFS locus do not affect specification of segment number nor establishment of segment polarity genes, *en* and *wg*, but rather disrupt their correct maintenance[19]. In *D. melanogaster*, loss-of-function or experimental knock-down of negative

regulators of Wg signalling (*e.g. axin* causes identical morphological defects in embryos[28]). Since Wg has been proposed as an eyespot-inducing morphogen[29], mutations in a negative regulator of the Wg pathway would affect the distribution of black and gold scales in the eyespot rings. In order to explain the segregation of eyespot and the embryonic lethal phenotypes, a model with different combination of alleles was proposed (Fig. 1a). In this model there are three alleles with three possible mutation sites which, in combination, result in the different phenotypes observed. The BE phenotype can be produced by a single copy of the recessive lethal B allele (in yellow) or two copies of the non-lethal a allele (in orange; each obtained by a different cross; see Fig. 1b) whereas the Spr phenotype is produced by a copy of a and B. The Fro phenotype is produced by a single copy of the recessive lethal C allele (in blue). In this model, a and C alleles carry a single mutation in different sites and the B allele carries two, one of which is common with the a allele (Fig. 1a). This model was put forward as the most parsimonious way of explaining BFS segregation. More specifically, it explains why Spr initially appeared in BE stock, why BE is segregating in a cross between Spr individuals (and not WT segregating instead, as it would be expected if it was a single causative mutation in the BFS) and also why Spr segregates in a cross between two BE (for instance one BE, with two copies of the a allele and other BE with only one copy of B allele).

The BFS locus, assigned to *B. anynana* linkage group (LG) 17, shows high levels of synteny with the orthologous chromosome of *B. mori*[16]. Because *axn* and *doubletime* (*dbt*) are genes linked to the Wg pathway in *D. melanogaster* (*e.g*. [30]) and are present in the orthologous chromosome of *B. mori*, they were prime candidates for the BFS locus. A mapping study was carried out to investigate size polymorphism alleles in these candidate genes[31]. Since there was a large number of recombinants between markers in the candidate genes and BFS alleles, consistent with them being distant loci, they were excluded from the BFS locus (Fig. 2a). Further mapping, using a recombinant panel and assessing recombinants in increasingly close markers and BFS, allowed to narrow down the interval that contains the BFS locus to approximately 40 kb (37 kb; Fig. 2b-d for details). *In silico* annotation of the BAC clone sequence was done with a web-available tool designed for annotation of genomic sequence in *B. mori* [32]. Of the 22 putative open reading frames identified, only four showed significant sequence similarity to known proteins: lethal (2) k05819, tyrosine phosphatase, WD repeat domain 13 protein, and serine/threonine protein kinase).

Additional work on BFS locus included manual annotation of the 40 kb region. Only one gene was found inside the mapped region – *washout* (*wash*) – which has five exons across the region. Approximately 300 bp from the location of the first *wash* exon, a putative tyrosine phosphatase encoding gene is found and has similarity to the *domeless (dome)* gene of *D. melanogaster*. Because of their known roles in embryonic development, and in particular, the role *dome* has in the JAK-STAT pathway, which was shown to negatively regulate Wnt/Wg signalling pathway[33], both these genes were considered as good candidates for the BFS locus. In this work, we tested these candidate genes by: 1) characterizing them in terms of sequence conservation by comparing *B. anynana* sequences with well described sequences in other model organisms and described role in development of other invertebrates found in the literature, 2) describing the spatial patterns of expression during embryonic and wing development of WT individuals of *B. anynana* by *in situ* hybridization (ISH), 3) attempting to characterize their putative role in wing pattern development with a genome editing tool, the CRISPR/CAS9 technique and 4) identifying nucleotide variants in coding sequence and intergenic region between both genes that distinguish proposed BFS alleles by sequence analysis. With these analysis, we found that both genes are expressed during embryonic development (as it was described in other insects) and during wing development (which was not yet been described before). Because we also found putative DNA polymorphisms consistent with being causative for BFS phenotypes in both candidate genes, we could not exclude one of them for being the gene behind the pleiotropic effects of the BFS locus.

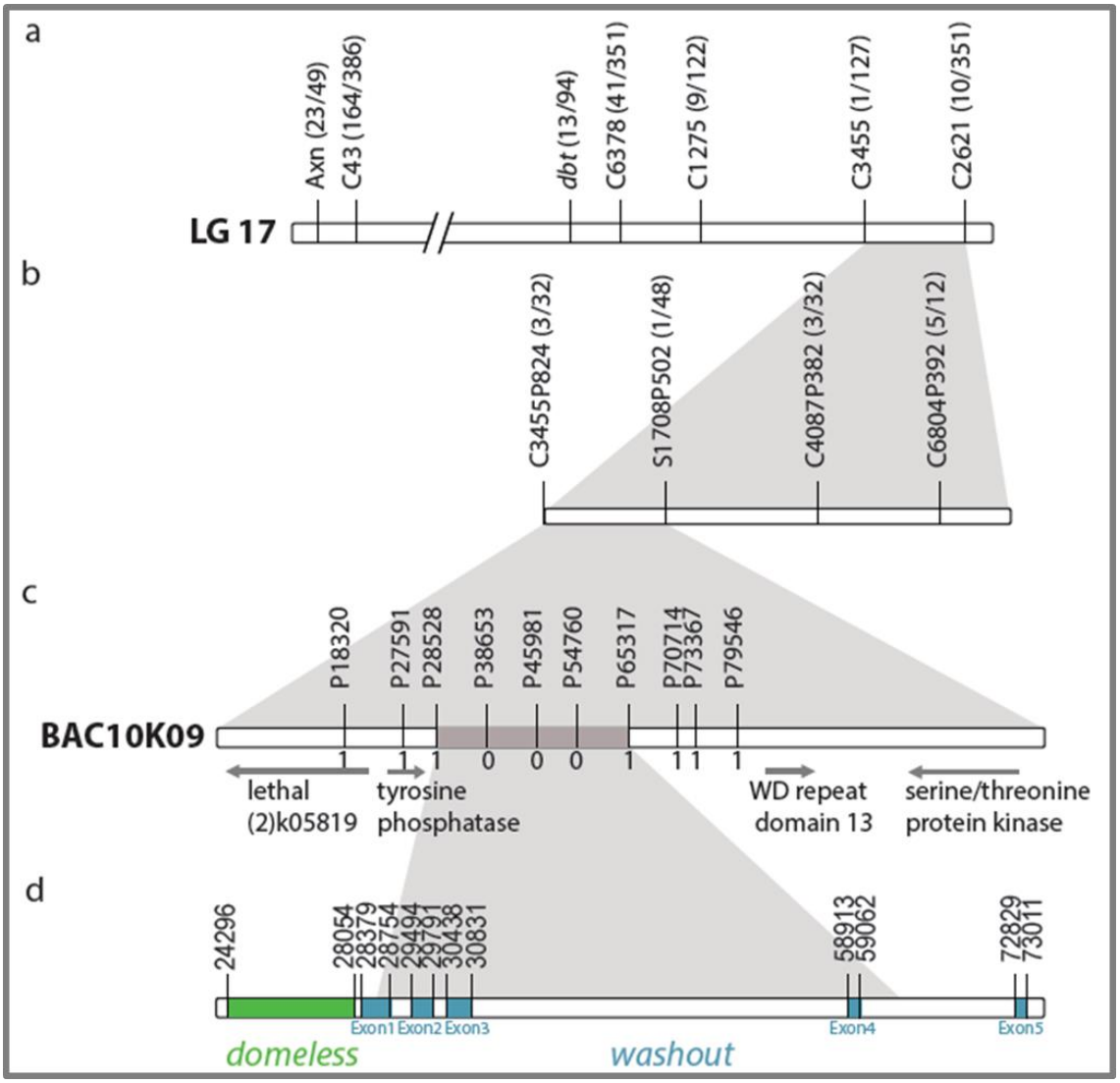


Figure 2. Mapping and annotation of the BFS locus. a) Recombination rates (# recombinant offspring / # total genotyped offspring) between the BFS locus and molecular markers (indel polymorphisms in *Axn*, *dbt*, and five ESTs) in *B. anynana* linkage group (LG) 17 mapped the BFS locus to a region between loci C3455 and C2621. **b**) Recombination rates between the BFS locus and SNPs in four genes known to localize between loci C3455 and C2621 mapped the BFS locus to a genomic region between loci C3455 and S1708. Because the orthologous region of *B. mori* is ca. 150 Kb long, the flanking loci were used to screen and select a corresponding BAC clone in *B. anynana*. **c**) *in silico* annotation of the selected BAC clone (10k09) sequence predicted four proteins with a known function and allowed development of further SNP markers (vertical bars) that allowed to narrow down the interval containing the BFS locus. The number after the P corresponds to the nucleotide position in BAC sequence, and the numbers under each bar correspond to the number of recombinant progeny between each of the markers and the BFS locus. This mapped the target locus to a ca. 40 Kb region between BAC nucleotides 26368 and 62361 (grey). **d**) Manually curated sequence analysis of this region identified exons for genes *dome* (green) and *wash* (blue). Number represent positions in BAC 10K09. All data in this figure are unpublished and part of the PhD thesis of Suzanne Saenko[31].

II. Materials and Methods

1. Experimental animals

All *B. anynana* stocks were reared in climate rooms at 27°C, 60-70% relative humidity and 12hr: 12hr light-dark cycle. Larvae were reared on maize plants and adults fed on sliced banana[34]. Both mutant stocks used in this study, Spr and Fro, carry recessive lethal alleles[19] which prevents the establishment of pure-breeding stocks. These two lines were maintained with active selection in favour of mutant phenotypes in each generation by removing the remaining segregating phenotypes (BE in Spr stock and WT from Fro stock). In the Fro stocks (genotype C/+) it is possible to recover, each generation, individuals with the following genotypic classes: C/C (die as embryos displaying a characteristic phenotype), C/+ (adults with altered eyespots) and ++ (adults with WT phenotype). Spr inheritance, however, is more complex (*cf.* Fig. 1b;[19]). Each generation, the Spr stock (genotype B/a; *cf.* Fig. 1b) segregates for Bigeye individuals (genotype a/a, *cf.* Fig. 1b) and defective embryos (B/B).

To obtain individuals that are homozygous for the mutant alleles (B/B or C/C), we dissected developing embryos from the Spr or Fro stocks, respectively. Recessive lethal mutant embryos can be distinguished from WT during early stages of development[19]. At 27°C, *B. anynana* embryogenesis lasts approximately 4 days (100 hr) – with one hour corresponding to 1% developmental time (DT)[35]. Time-controlled eggs from both mutant stocks were collected and transferred to Petri dishes. Spr and Fro embryos were let to develop until ~70% DT and ~90% DT, respectively, as those are the stages when differences between WT embryos and mutants with disturbed embryogenesis are clearer[19]. Eggs were dechorionated in 50% bleach solution for one minute, rinsed twice with water and dissected in 1x phosphate-buffered saline (PBS) under a light scope (Olympus SZX7).

2. Characterization of candidate genes

The BFS locus was mapped to a 40 kb interval (contained within sequenced BAC clone 10K09) on LG17 of *B. anynana*[31]. *In silico* annotation of the BAC with Kaikogaas, a tool customized for *B. mori* genomic data[32], allowed the annotation of one of the candidate genes in this study, *dome,* adjacent to the BFS locus (Fig. 2c). A manually curated sequence analysis allowed to identify a single gene inside the BFS interval, *wash* (Fig. 2c). *dome* is predicted to be a single exon gene with a length of 3759 bp (from position 24296 to 28054 in BAC10K09; *cf.* Fig. 2d) and *wash* to have five exons from position 28379 to 73011 in BAC10K09, having a total length of 44632 bp (*cf.* Fig. 2d). *wash* but not *dome* exonic boundaries were confirmed by amplification of WT complementary DNA (cDNA), sequencing and alignment of sequences with BAC 10k09 (previous work by Saenko, unpublished data).

Coding DNA sequences (CDS) of *dome* (3759 bp) and *wash* (1392 bp) were used for running tBLASTx analysis (default settings) against NCBI and EnsemblMetazoa databases to find orthologues in other species. All sequence hits with e-values below to 1e-45 were selected, translated (ExPASy translate tool) and used for phylogenetic reconstruction using MEGA version 5. Phylogenetic relationships were estimated using Neighbor-Joining with complete deletion and bootstrap (1000 replications). Although *D. melanogaster* is phylogenetically distant from *B. anynana,* this organism is to our knowledge the one in which both genes have more detailed descriptions of both protein structure and functions. Therefore, we have used amino acid sequences from *D. melanogaster* Dome (GenBank accession number NP_523412) and Wash

(GenBank accession number AGA18702) to assess sequence similarity, homology and predict protein domains.

Sequence similarity was assessed by aligning the amino acid sequences from *D. melanogaster* and *B. anynana* with Blastp tool on NCBI (blast.ncbi.nlm.nih.gov/Blast.cgi), with default settings.

Conserved protein domains of the candidate genes were searched using NCBI conserved-domain search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and SMART (simple modular architecture research tool, http://smart.embl-heidelberg.de/), both run with *B. anynana* amino acid sequences (default settings). Wash domains where only found in NCBI database whereas Dome domains were found in both databases. Accession numbers and e-values presented in the main text (*cf.* Characterization of candidate genes *dome* and *wash*) correspond to hits in SMART (Dome) and NCBI (Wash). Given the rather high e-values, we also confirmed predicted domains by running a ClustalX alignment in Bioedit (default settings, Fig. S1) with amino acid sequences of Dome and Wash from *B. anynana* and their corresponding orthologues from *D. melanogaster*, where data on described domains is available (*cf.* [36] and [37], respectively).

3. cDNA library of embryo and wing development

In order to build a cDNA library of both embryonic and wing developing specimens, RNA was extracted from a series of developmental stages of *B. anynana.* To cover all embryonic development, we let females from a large outbred WT population lay eggs for two hours and collected approximately 450 eggs that were split between time points: 0, 24h, 48h, 72h and 100h. A total of 40 embryos were pooled for each of three biological replicates of each time point, homogenised with a pestle and stored in 400ul of Trizol (Invitrogen) at -80°C until RNA extraction.

For wing development, we sampled during the last larval instar and during early pupal life from individuals of the outbred WT population. Both larvae and pupae were anesthetized in cold PBS and dissected under sterile, RNAse-free conditions (RNAse Zap,Sigma-Aldrich). Wing imaginal discs of 13 5th instar larvae were staged according to the tracheal extension into the vein lacunae as described elsewhere[38]. We had at least two biological replicates per stage, with pools of 4 wings of the same individual. Pupal wings were staged by timing the hours after pupation, by using time-lapsed photographs taken in ten minute intervals with a digital camera (Canon 1100D). All wings from each of the 16 individuals were collected in different time points, from 0 to 22h after pupation, in each two hours. Dissected tissues were homogenised with a pestle and stored in 400 µl of Trizol (Invitrogen) at -80°C until RNA extraction.

Total RNA was extracted with Direct-zol RNA extraction Kit (Zymo Research), following manufacturer's instructions. Samples were run on an 1% agarose gel to verify RNA integrity and measured with Nanodrop spectrophotometer (Thermo Scientific) to verify yield. Trace gDNA was removed with RQ1 DNAse (Promega), following manufactures' instructions. 1 µg of total RNA of each of 38 samples was synthetized in cDNA using M-MLV reverse transcriptase (Promega) following manufacturer's instructions except oligo (dT) primer incubation period (60 min instead of 15 min). cDNA from each time point stage (embryos, larval wings and pupal wings) was pooled and a standard PCR (semi-quantitative PCR) was done to verify in which stages candidate genes *dome* and *wash* were expressed. PCRs were performed in a total volume of 25 µl with 0,25µg of template, 10µM of forward and reverse primers, 2,2 units of GoTaq® DNA Polymerase (M300, Promega), 5x Green GoTaq Flexi Buffer (Promega), 2,5 mM of each DNTP (Takara), 25 mM $MgCl_2$ and MilliQ water to amount to 25 µl . PCR conditions were: 94°C for 4

min; 35 cycles of 94°C for 30 s, X°C for 30 s, 72°C for Ys; end with 10min at 72°C where X (anneling temperature) varied according to the sets of primers used and Y (extension) to the length of the fragments amplified (see Table S1).

4. Expression Analysis

For riboprobe synthesis, we first cloned one fragment of *wash* (340 bp fragment between primers washout.F1 and wash.exon1.R1; Table S1) and two of *dome* (a 852 bp fragment between primers domeless.F2 and domeless.R2, and a 809 bp fragment between primers domeless.F3 and domeless.R3; Table S1). Although ideally riboprobes should be against 3'UTR (Patrícia Beldade, personal communication), *dome* and *wash* probes were designed in either middle or 5' of the gene, respectively, since primers that amplified 3' UTR did not work (data not shown).

All fragments were amplified from pooled samples from all stages of embryonic cDNA (both genes were more expressed during embryonic development comparing with wing development) using the following PCR conditions: 94°C for 4 min; 40 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 1 min; end with 10 min at 72°C. For PCR components and final volume see section above. PCR products were analysed in 2% agarose gels and bands of expected size were extracted and cleaned-up with Nucleospin Gel and PCR Clean-Up (Macherey-Nagel), following manufacturer's instructions. PCR was performed as described above (see Table S1). 20 ng of cleaned PCR product was ligated into the pGem® T-easy Vector (Promega) and inserted in DH 52 competent cells, following manufacturer's protocol. A PCR with DNA of a single colony as template was performed with promotor specific primers for T7 and SP6 with the following conditions: 94°C for 3 min; 35 cycles of 94°C for 30 s, 58°C for 30 s, 72°C for 45 s; end with 10 min at 72°C. Colonies in which fragment insertion was successful were let to grow in lysogeny broth (LB) overnight and where plasmid DNA was isolated with Nucleospin® Plasmid (Macherey-Nagel). Plasmids were sequenced with T7 and SP6 primers (see DNA extractions and sequencing analysis section for sequencing protocol) and sequences aligned to BAC10K09 using ClustalW (Geneious v8.1) to confirm both identity and orientation of insert. 1000ng of plasmid DNA was then used for synthesis of *in situ* hybridization antisense and sense riboprobes with either T7 or SP6 RNA polymerase (Promega) and labelled with Fluorescein RNA labelling mix (Roche Applied Science), according to manufacturer's instructions. All probes were run on an 1% agarose gel and measured with Nanodrop spectrophotometer (Thermo Scientific) to assess quality and concentration.

To assess expression patterns of the candidate genes in embryonic development, approximately 500 eggs were collected for one hour periods and kept at 27°C until further processing. Eggs with embryos collected at different hours were dechorionated in 50% bleach solution for one minute, rinsed twice with MilliQ water and fixed in 10% formaldehyde in 1x PBS / 50mM EGTA solution for 1-2 hours. They were then washed twice in 1x PBS, dissected with fine forceps and staged according to their morphological and physiological maturation, as described in *Manduca sexta* embryonic development[39] (see Table S2 for stages and sample number) . Wing imaginal discs of 5th instar larvae from outbred WT population were dissected (Table S2), staged as described in cDNA library section and fixed in 10% formaldehyde in 1x PBS / 50mM EGTA for 30-60min. Fixed embryos and wings were gradually dehydrated in increasing concentrations of methanol (33%, 66% and 100%) and stored at -20°C until use. *In situ* hybridization was performed with antisense and sense probes (as negative control) for both genes and a positive control probe with *dll* (kindly provided by Kohtaro Tanaka), a gene in which expression patterns during embryonic and wing development are known[26,40]. All probes were used with a final concentration of 150 ng/µL. *in situ* hybridization was performed following the protocol in [41] with

the following modifications: embryos and larval wing discs were incubated for 3min and pupal wings for 5min in 4,5 µg/mL proteinase K at 37°C and hybridization was carried out at 60°C, overnight. For solutions used thorough *in situ* hybridization protocol see Annex III (List of solutions). All samples were mounted in 70% glycerol in PBS in slides, observed under Leica DMLB2 upright microscope, equipped with a IDS colour CCD camera and using uEye Cockpit software.

5. Functional analysis

In order to perform genetic engineering on *B. anynana*, CRISPR/Cas9 technique[42] was used to induce small double strand breaks (DSB) in the candidate genes. All steps in this experiment were adapted from Basset & Liu[43], with assistance from post-doctoral researchers Arnaud Martin (UC Berkeley) and Diogo Manoel (IGC).

Target sequences of both *dome* and *wash* were found using ZiFit Target tool – CRISPR/Cas9 nucleases (http://zifit.partners.org/ZiFiT/), using as template the coding sequences of the genes. This tool finds sequences that are followed by an NGG protospacer adjacent motif (PAM) sequence, necessary for the induction of endonucleolytic cleavage of the DNA by the Cas9 protein. Because there are no tools yet available to calculate off-targets for this species, potential off-targeting was not considered. To maximize mutagenic efficiency and a clear assay, target sequences were chosen based on 1) site in coding sequence (close to 5' region to induce a frameshift and not a complete failure to produce a functional protein); 2) GC content higher than 50% at 3' region of target sequence (*cf.*[44]); 3) absence of SNPs between genotypic classes studied in this work (*cf.* Fig. 8). A T7 polymerase binding site and sgRNA (single synthetic guide RNA) backbone (the region complementary to the common reverse primer that contains the remaining sgRNA sequence) were added to the target sequences to allow the production of the sgRNA template and *in vitro* transcription (*cf.* Fig. 6). All oligonucleotides (sequences in Table S3) were ordered from Sigma-Aldrich. Even though we noticed a mistake (one nucleotide missing, *cf.* Table S3) in the sgRNA backbone (sequence that primes to CRISPR R oligonucleotide) of all sgRNAs, we estimate this not to pose a problem because 1) PCR to generate the template for transcription worked even though there is a mismatch, which means that 5' to 3' strand as a nucleotide missing in sgRNA backbone but 3' to 5' strand has the correct sequence (from CRISPR R); 2) T7 polymerase uses the 3' to 5' sequence as template for transcription.

Production of sgRNA template and *in vitro* transcription were carried out as in [43] with the following exceptions: gel extraction of the PCR product with Nucleospin Gel and PCR Clean-Up (Macherey-Nagel); in vitro transcription reaction was purified with MEGAClear Kit (Ambion), following manufacturer's instructions. Cas9 protein (PNA Bio, CP01) was ressuspended at 1µg/uL in 0,015% Phenol-Red (Sigma-Aldrich)/RNAse free water.

Injections were performed in embryos with 2-4% DT because, if one or both of candidate genes studied lead to lethality when in homozygosity[19], early injections when cellularisation has not yet begun would lead to embryonic lethality. Hence, with this experiment, we aimed to obtain mosaic adult individuals where cells with induced mutation are phenotyped considering causative differences in pigmentation and colour patterns.

Embryonic chorion was soften and cleaned from fungi in 10% Benzalkonium Chloride Solution (Sigma-Aldrich) and rinsed twice in MilliQ water. The embryos were then dried in filter paper and mounted in strips of tape glued to a lid of a petri dish. Injections were done with glass capillary needles composed of borosilicate glass (Sutter Instrument, CORNING 7740) pulled in a Flaming/Brown Micropipette Puller (Sutter Instrument model P-97) with the following program:

P=500; Heat=543; Pull=20; Vel: 100 and Time=150 and broken with a fine forceps. CRISPR injection mixes were prepared with Cas9 at 250ng/ul and sgRNAs at 125ng/ul in nuclease-free water with a final volume of 10µl or Cas9 at 333 ng/ul and sgRNAs at 150 ng/ul in nuclease-free water with a final volume of 7,5 µl (*cf.* Table S4) . Each 2µl of injection mix was loaded in the needles with an Eppendorf Femotip while the rest of the mix was kept on ice to avoid RNA degradation. Because of the variability of each needle and injection pressure, it is not possible to quantify how much volume was injected per embryo. After injection, wet cotton was added to the petri dishes to keep humidity high and maize leafs to feed the freshly eclosed larvae. Injected eggs were left in the climate room (27°C) to develop and were transferred to a new cage with maize plants by either transferring the maize leafs or with a fine brush. Freshly eclosed butterflies were frozen at -20°C for further phenotyping/genotyping.

6. DNA extractions and sequence analysis

To obtain genomic DNA (gDNA) of the six genotypic classes as explained above (a/a; B/a; B/B; +/+; C/+ and C/C) plus outbred WT stock (referred as [+/+] to distinguishing from the WT segregating in Fro population) we used either stored thoraxes from adult female butterflies or dissected aberrant embryos. All gDNA extractions from thoraxes were performed individually (and then five individuals were pooled) opposed to embryos, where approximately 50 Fro embryos and 35 Spr embryos were dissected, scored and pooled immediately before extraction. gDNA extractions were performed with Dneasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. Extraction products were run on an 1% agarose gel to test for DNA integrity and measured with Nanodrop spectrophotometer (Thermo Scientific) to verify quality and yield.

In order to sequence the complete CDS of candidate genes (*i.e.* all exons) *wash* and *dome*, different sets of primers were used on gDNA from the different genotypic classes (see Table S1). gDNA from five females of each genotypic class were pooled in equimolar amounts except for the recessive lethal embryos. For PCR components and conditions see sections above. PCR products were analysed in 2% agarose gels and bands of expected size were extracted and cleaned-up with Nucleospin Gel and PCR Clean-Up (Macherey-Nagel), following manufacturer's instructions. Sanger sequencing was performed in 3130XL Genetic Analyser of Applied BioSystems using BigDye Terminator v1.1 Cycle Sequencing Kit, according to the manufacturer's instructions, with both forward and reverse primers used previously for amplification of each fragment.

Manual curation of the reads was performed to check for quality of peaks, misread nucleotides and heterozygosity (in which the proper IUPAC nucleotide code was given for each base). Phred score, the numerical estimate of error probability for a given base[45], was assessed and sequences with a percent of high quality base calls in a sequence read-out (HQ%) greater than 40 were used. CDS of *dome* was covered at least twice in the seven genotypic classes, in independent sequencing assays (*e.g.* using both forward and reverse primers for the same template and/or a different template sequenced with other sets of primers). *wash* coverage was not complete for all genotypic classes (*cf.* grey areas Fig. 8 and Table 1) as certain CDS regions had HQ % below the threshold defined. In order to investigate variants in other genotypes, BAC10K09 sequence was used for running BLASTN analysis against *B. anynana* genome database (Version 0.4; bicyclus.org). nBa.0.1.scaf01516 (Scaffold 01516) was the scaffold producing the most significant alignment opposed to other hits that correspond to short sequences and appear across different scaffolds (probably transposable elements).

SNP analysis was performed for the seven genotypic classes sequenced and Scaffold 01516, using as reference sequence BAC10K09. HQ % assessment, sequence assembly, alignments (ClustalW) and SNP analysis (default settings) were performed using Geneious v8.1.

III. Results and discussion

Starting from preliminary results of mapping and sequencing, we tested two candidate genes for BFS locus, a genomic region that affects embryonic development (notably, segment polarity) and eyespot development (specifically eyespot size and colour composition)[19]. We first characterized these genes in relation to what is known about the protein sequence and function described for other species. Secondly, we investigated the spatial patterns of expression of both candidate genes during embryonic and wing development. Thirdly, we described an experiment to attempt a functional analysis of the candidate genes on wing pattern development. Finally, we assessed the DNA sequence variation in distinct genotypic classes, including the embryonic and eyespot mutants.

1. Characterization of candidate genes *dome* and *wash*

To characterize the *B. anynana* candidate genes *dome* and *wash*, we compared them with a variety of invertebrate species to place them in a phylogenetic context. Also, we compared the amino acid sequences with those of *D. melanogaster* to highlight predicted functional domains. Ultimately, we compiled available literature describing the role of these genes in embryonic and wing development, with special attention to Wg signalling.

1.1 Protein sequence analysis

We compared the amino acid sequences of Dome and Wash with orthologous sequences from other insect species. We paid special attention to comparisons with *D. melanogaster,* for which these genes are better characterized in terms of protein structure and function [36] and [37], respectively).

To place Dome and Wash amino acid sequences of *B. anynana* in a phylogenetic context, we built neighbour-joining phylogenetic trees with other insect species orthologues. By building these unrooted trees, we aimed to assess sequence divergence among different insect orders. For both amino acid sequences of the candidate genes, all lepidopterans sequences included in the analysis are clustered. The same is observed for other insect orders such as Diptera (fruit fly and mosquitoes). This might indicate that sequence divergence within the lepidopterans is lower than that among insect orders.

Compared to the orthologous sequences of D. melanogaster, *B. anynana* Dome protein shares 25% identities (total length of 1252 amino acids, 44% positives, e-value = 4e-52) whereas Wash shares 29 % identities (total length of 467 amino acids, 42% positives, e-value = 4e-36). It has been demonstrated that sequence comparisons often provide good suggestions for gene functions[46]. In both protein comparisons, the sequence identities fall into the so-called "twilight zone"[47] in which the probability of sequence alignments unambiguously distinguishes between protein pairs is low. Therefore, it is inconclusive whether *B. anynana* candidate proteins share the same structure and it is even less certain whether they share the same function. However, conserved protein domains known to be important for both protein functions in D. melanogaster, were found using web-available tools (e.g. NCBI conserved-domain search or SMART tool). For Dome, those predicted domains are three fibronectin-type-III like domains, one transmembrane domain and one CBM (cd00063; e-values <4.06e-03) and for Wash, a domain that activates the Arp2/3 complex was found (pfam11945, accession cl13393; E-value 3.69 e-50). We also predicted the location of those domains by a ClustalX alignment of Dome and Wash from *B. anynana* and *D. melanogaster* (Fig. S1), to assess whether the variants found in the sequence variant analysis of the BFS genotypic classes occur in conserved domains.

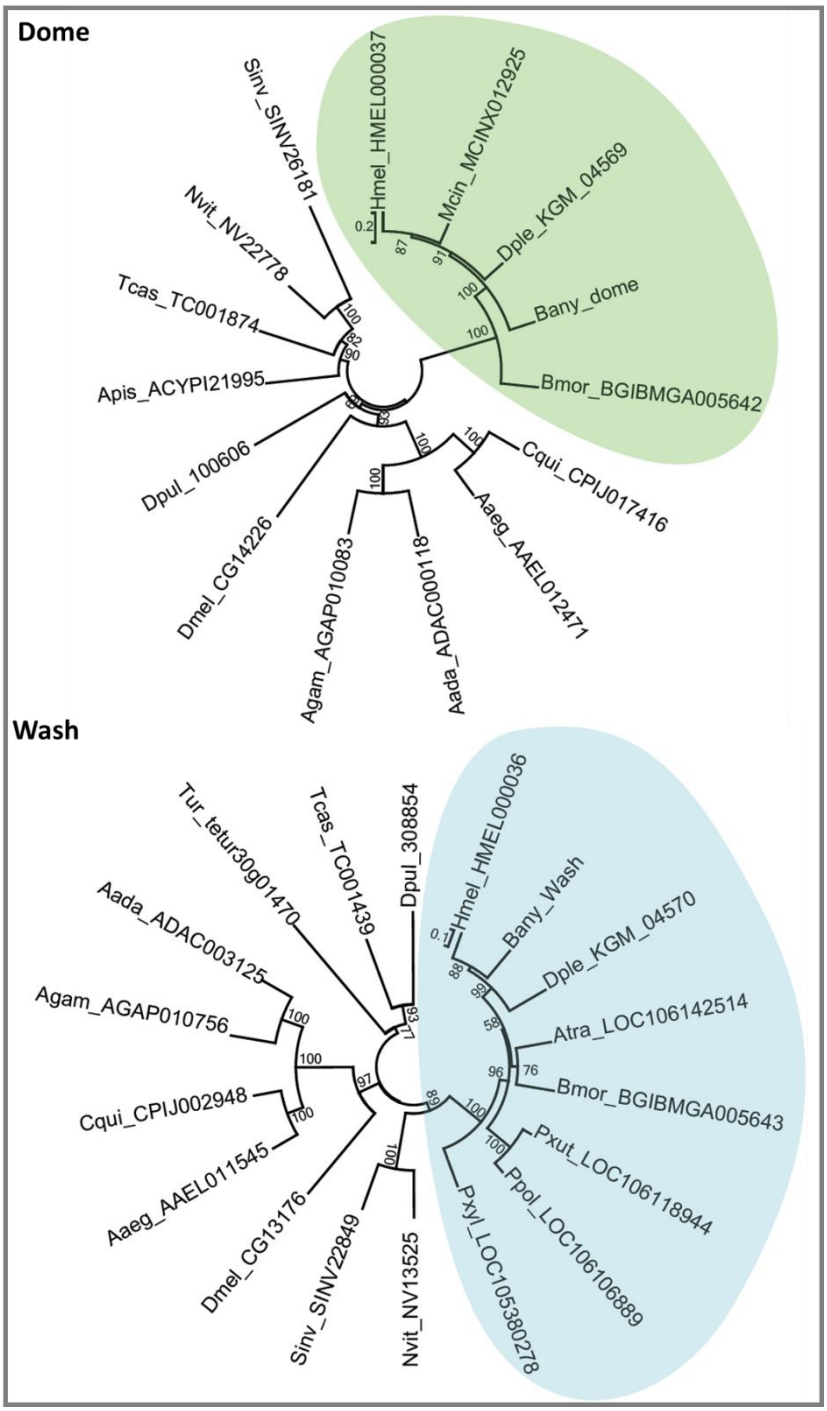


Figure 3. Phylogenetic trees of candidates Dome and Wash. Neighbour-joining unrooted trees generated with MEGA5 using aminoacid sequences of *B. anynana* (Bany) and the orthologous of different species: *Daphnia pulex* (Dpul, Cladocera), *Acyrthosiphon pisum* (Apis, Hemiptera), *Nasonia vitripennis* (Nvit, Hymenoptera), *Solenopsis invicta* (Sinv, Hymenoptera), *Tribolium castaneum* (Tcas, Coleoptera), *Drosophila melanogaster* (Dmel, Diptera); *Anopheles gambiae* (Agam, Diptera); Aaeg, *Aedes aegypti* (Diptera), *Anopheles adarlingi* (Aada, Diptera), *Culex quinquefasciatus* (Cqui, Diptera), *Heliconius melpomene* (Hmel, Lepidoptera), *Papilio xuthus* (Pxut, Lepidoptera), *Papilio polytes* (Ppol, Lepidoptera), *Bombyx mori* (Bmor, Lepidoptera), *Amyelois transitella* (Atra, Lepidoptera), *Plutella xylostella* (Pxyl, Lepidoptera), *Melitacea cinxia* (Mcin, Lepidoptera), *Danaus plexippus* (Dple, Lepidoptera). Lepidopteran clusters are highlighted in green (Dome) and blue (Wash). ENSEMBL or NCBI accession numbers are shown after underscore symbol that follows species code name. Numbers in nodes represent bootstrap values for 1000 replicates.

1.2 Described expression and function

In order to explore the mechanisms by which these candidate genes can affect embryogenesis and how each relates to changes in eyespot pattern and morphology, we conducted a literature survey for expression and functional analysis in other insect species.

dome was initially identified in *D. melanogaster* and it is, up till today, the only receptor known to be involved in the activation of the invertebrate Janus kinase/ signal transducers and activators of transcription (JAK-STAT) signalling pathway[36]. In this species, Dome has a cytokine binding homology module (CBM) that shares key amino acids required for signalling in the vertebrate cytokine class I receptors, suggesting common ancestry and was proposed to be conserved in invertebrates[36]. In *D. melanogaster*, this gene is expressed maternally and at later embryonic stages in tracheal pits, posterior spiracles, gut and Central Nervous System(CNS)[36]. Maternal and zygotic *dome* mutant embryos have segmentation defects which indicates that the gene plays a crucial role in early embryonic patterning and more specifically, for the expression of pair-rule genes. In *T. castaneum*, the JAK/STAT signalling pathway is not only required for the activation of transcription in early development but also at the level of segment polarity genes, for which it is required for maintenance[48]. In insect species such as beetles[48], bees[49], moths[50] and butterflies[51], orthologues of this gene have been found and predicted as JAK-STAT receptors. To our knowledge, *dome* functions have not yet been validated in lepidopterans with exception of a recent study suggesting that it might be implicated in eye development in *B. anynana*[51].

Wash (Wiskott-aldrich scar homolog) is a protein that belongs to the Wiskott-aldrich syndrome (Was) family[37]. Members of this protein family are involved in cytoskeleton reorganization and signal transduction by acting as effectors of Rho-GTPases and polymerizing actin via the Arp2/3 complex[52]. This protein is evolutionary conserved, especially in the C-terminal portion, which has an actin-binding domain (WH2) and other domains required for the activation of actin polymerization via Arp2/3 complex[37,53].

In *D. melanogaster*, *wash* is expressed maternally and is not detected at later stages of embryonic development[54]. Wash protein is expressed uniformly in the cytoplasm of all cells throughout embryogenesis and exhibits spatial and temporal enrichments above uniform levels[55]. During syncytial stages, Wash localizes to the cytoplasm surrounding migrating nuclei, during gastrulation Wash is enriched in cells undergoing shape changes and in later embryonic stages, Wash is expressed in posterior spiracles and enriched in the posterior midgut and Malpighian tubules[55]. Wash roles in biochemical activities have been shown to be essential for embryogenesis in *D. melanogaster*, since reduced *wash* expression leads to fused dorsal appendages as a result of improper dorsal-ventral patterning[55]. In addition, it has been shown that transheterozygous embryos of *wash* and *spire* (another gene involved in regulation of actin cytoskeleton) have segmentation defects due to effects on the maintenance of segment polarity genes[55].

Saenko *et a*l.[19] proposed that the BFS locus encodes a negative regulator of the Wnt/Wg pathway, since embryonic mutant phenotypes caused by different alleles of this locus, have segment polarity defects. Since both genes are expressed during embryonic development and mutant phenotypes in *T. castaneum* (*dome*) and *D. melanogaster* (*wash*) show that both play an important role in segment polarity, these genes are good candidates for the BFS locus. A gene of particular interest is *dome*, since it is known that the JAK/STAT pathway interacts with Wnt/Wg signalling [33]. However, there are no reports describing a role of these genes during wing development (with exception of the correlation of *dome* expressing in trachea tissue – where wing veins develop from – during *D. melanogaster* embryonic development).

Although functional tests are still missing, Wg is likely to be an eyespot-inducing morphogen[29]. Since Wg is known to upregulate *en* and Dll transcription[56], if we consider it as the focal signal then the cells that surround the foci will receive different concentrations of this morphogen and, depending on how close cells are to the focus, domains of Dll (black ring) and *en* (golden ring) are formed. The presence of a negative regulator of Wg and mutations affecting its function would cause a different response of the epidermal cells, leading to a different regulation of expression of *en* and Dll. Consequently, a different distribution of black and golden scales could occur as well as the size of the eyespot might be altered. The candidate genes for the BFS locus here studied might be or might regulate a negative regulator of the Wg pathway and being responsible for the BFS phenotypes.

2. Expression analysis of candidate genes in embryonic and wing development

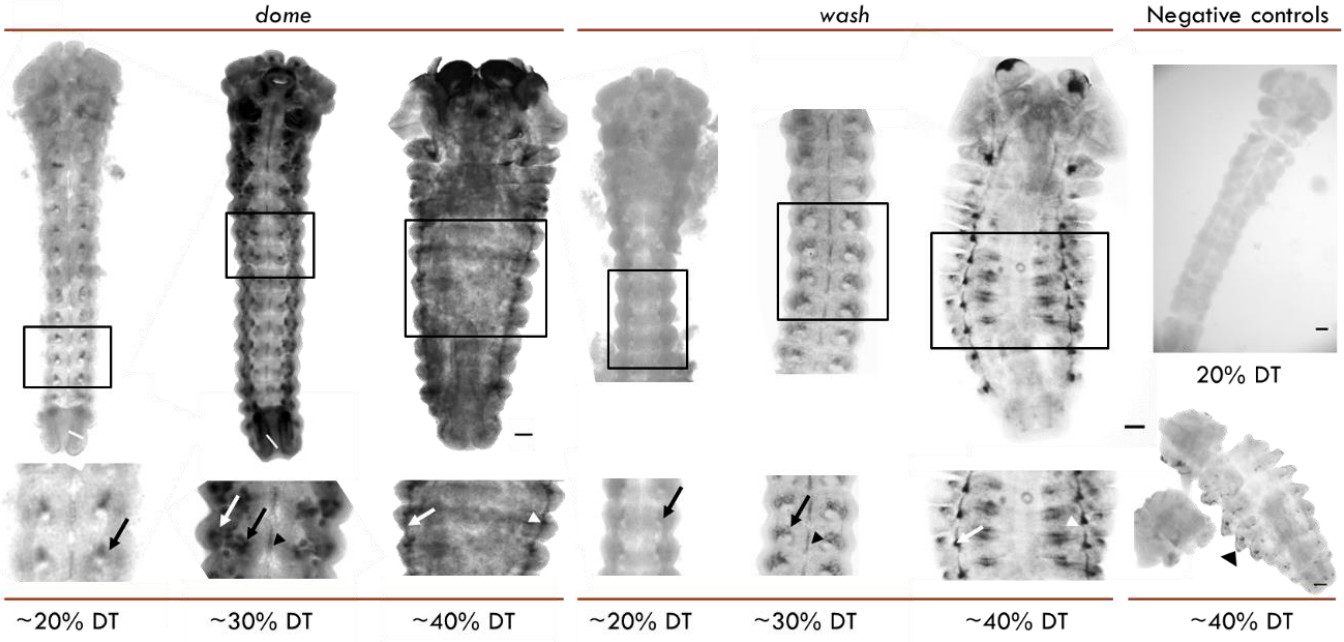
BFS alleles disturb embryonic development during the segmented germband stage (in homozygosity) and eyespot morphology (in heterozygosity). Therefore we expect our candidate genes to have patterns of expression consistent with a role in embryonic segment polarity and in defining eyespot size and colour composition. We have analysed expression patterns of the two candidate genes *dome* and *wash* during both embryonic and larval wing development by *in situ* hybridization.

2.1 Expression analysis during embryonic development

To assess whether the candidate genes are being expressed during embryonic development, we did *in situ* hybridization in WT embryos sampled from 15 to 65% DT (for sample number size see Table S2). Both *dome* and *wash* mRNA were detected throughout embryonic development, with similar spatial distributions (Fig. 4). At 20% DT, both genes are expressed in the anterior part of presumptive thoracic and abdominal appendages, although the detection of *wash* mRNA was lower in this stage of embryonic development. At later stages (30-40% DT), expression of both genes starts to appear in trachea, spiracles and in the CNS. These results are consistent with the expression patterns of *dome* in *D. melanogaster* embryos in spiracles[36]. In *D. melanogaster* embryos, only Wash protein was detected in spiracles. Expression of both genes in the anterior part of the presumptive thoracic and abdominal appendages has, to our knowledge, not been reported before. Likewise, in *T.castaneum dome* expression does not appear in distinct anterior segment domains but rather ubiquitously, although it was shown to be tightly linked to segment polarity establishment[48].

Lethal alleles at the BFS locus have been shown to affect the correct maintenance of segment polarity genes such as *engrailed* (*en*)[19]. The expression domains of En is typically confined to the posterior domains in WT embryos[57]. However, in BFS mutant embryos, expression of En is detected both in the posterior and in the anterior part of each segment. By comparing these morphological defects and gene expression of segment polarity genes between *B. anynana* and *D. melanogaster* embryos, it was proposed that the BFS locus encodes a negative regulator of the Wnt/Wg pathway[19]. Since the establishment of segments (or parasegments) is highly conserved in arthropods[58] regulators of these pathway are expected to be conserved as well. Conversely, because known negative regulators of this pathway (*e.g axin* and *doubletime*) were excluded from being in the BFS region by mapping, and the genomic region implicated for BFS mutants contains *wash* and adjacently *dome*, it is possible this region regulates Wnt/Wg pathway in butterflies. Here we show that both candidate genes for BFS locus are expressed during embryonic development. Interestingly, it seems that in early stages of development (20%

DT), when embryonic germ band is evident and thoracic appendages start to emerge, the expression is confined to the anterior part of the segments. It is conceivable that, in this location, they play a role in the regulation of segment polarity genes maintaining their expression in the posterior part of segments. Absence of this repressor in the anterior region of the segments could disrupt gene expression of segment polarity genes (e.g *en*) which would fail to be maintained only in the posterior part. Although at 20% DT, segment polarity is already established[39], it is possible that the expression of the regulator is still required for the maintenance of the segment polarity genes at this stage. Future studies should investigate *dome* and *wash* expression patterns earlier embryonic development (0-20% DT) and also raise the sample size number of the remaining developmental stages to support these results. It would also be interesting to assess whether expression patterns of *dome* and *wash* are disrupted or



absent during embryonic development of BFS mutants.

Figure 4. Patterns of expression of *dome* and *wash* in *B. anynana* WT embryos. Detection of *dome* (left), *wash* (middle), and negative control (sense probe for *dome*, right) mRNA in embryos at different stages of embryogenesis: *ca.* 20%, 30%, and 40% DT. For each gene and stage, we show a representative image of the whole embryo and, immediately underneath, a detail thereof (black inset). *dome* is expressed throughout embryonic development: in the anterior part of presumptive thoracic and abdominal appendages (black arrow) at *ca.* 20% DT (ca. 38 out of 47 embryos, see Table S2), in the presumptive trachea (lateral line, white arrow), presumptive thoracic and abdominal appendages (black arrow) and CNS (black arrowhead) at *ca.* 30% DT (ca. six embryos out of six , see Table S2) and in the trachea (white arrow) and spiracles (white arrowhead) at ~40% DT (ca. four embryos out of 10 , see Table S2). Note that at 40% DT there is unspecific binding (representative image with *dome* sense probe) in the cuticle (black arrowhead in negative control). *wash* is also expressed throughout embryonic development: in the anterior part of presumptive thoracic and abdominal appendages (black arrow) at ~20% (13 out of 25, see Table S2) embryos and 30% DT, in the presumptive CNS (black arrowhead) at ~30% (eight out of eight) and 40% DT (six out of nine embryos see Table S2), in trachea (white arrow) and spiracles (white arrow) at 40 % (four out of 11, in which unspecific binding was not seen). Anterior is to top; scale bar is 100 µm for whole embryo images except for details thereof (200 µm).

2.2 Expression analysis during wing disc development

To regulate eyespot morphology, the BFS locus candidate genes should affect properties of the eyespot organizer or the surrounding epidermis which respond to an organizer-derived signal

(reviewed in [27]). To determine the spatial distribution of *dome* and *wash* transcripts during establishment of this organizer, we performed *in situ* hybridization in WT larval wing discs.

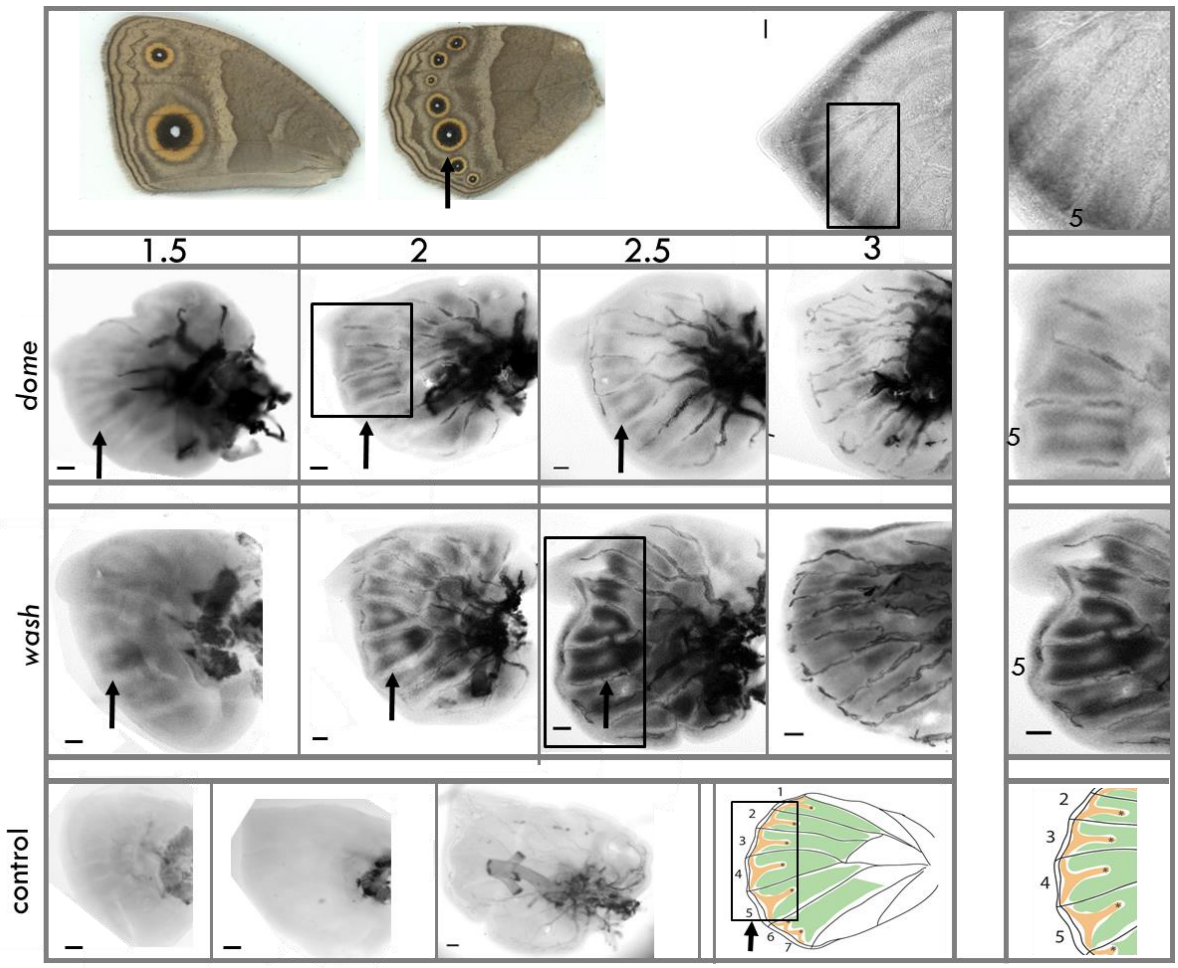


Figure 5. Patterns of expression of *dome* and *wash* in *B. anynana* WT developing larval wings. From top to bottom, each panel has wings assessed with different probes. An example of observed expression patterns are on the right with a detail thereof (black inset). **First panel** shows a representative image of *B. anynana* adult forewing and hindwing. On the right, positive control for the *in situ* hybridization with *dll* probe in a late 5th instar larval forewing in ventral view, stage 3 (all staging was done accordingly to [38]) and a detail thereof on the right (black inset). **Second panel -** *dome* expression starts to be clear from stage 1.5, where mRNA is detected in wing cells (between vein boundaries) except where the presumptive eyespot foci is developing. In later larval stages, expression is only detected proximately or not detected at all (stage 3). 1.0 – forewing, ventral view; 1.5, 2 & 2.5 – hindwing, ventral view; 3 & 3.75 forewing, ventral view. Black arrows point to the 5th eyespot indicated in the adult hindwing image. **Third panel** - *wash* expression pattern is similar to *dome* since it is detected between vein boundaries but not in the presumptive eyespot foci. 1.0 to 2.5 – hindwings, ventral view; 3 – forewing, ventral view. **Fourth panel** - Negative control with *dome* sense probe is shown for wings in 1.5, 2.0 and 3.0 stages (same absence of pattern for *wash* sense probes, *cf.* Table S2) for each stage. 1.0 – hindwing, ventral view; 1.5, 2 & 3.0 – forewings, ventral view. Last panel on the right with a scheme representing a hindwing with the pattern of expression of *dome* and *wash* (green) detected in this experiment and known expression patterns of *dll* in larval wing discs. Scale bars are 100 µm.

In early-fifth instar wings (0-0.75 stages; *cf.* [38]) neither *dome* nor *wash* mRNA was detected (in 17 and 16 wings, respectively; Table S2). In mid-fifth instar wings (1.0-2.75 stages, 16 and 11

wings for *dome* and *wash*, respectively), mRNA was detected in the epidermis of compartments between veins except where the presumptive eyespot foci is developing (Fig. 5). Similar to the expression patterns we observed in embryos, in developing larval wings the expression patterns of *dome* and *wash* are also similar. In late-fifth instar wings (3.0-4.0, 4 wings per gene), mRNA was also detected between vein boundaries although only proximately, or not detected at all (Fig. 5; Table S2).

Our results indicate that candidate genes *dome* and *wash* are expressed in developing wings, which, to our knowledge, has not been reported before. Interestingly, the expression is most profound in wing compartment regions bounded by veins, where the eyespots are normally found. mRNA locates in the surrounding region and seems not to overlap with the presumptive eyespot field and inter-vein midline (Fig. 5, bottom right image).

The expression of both candidate genes was detected during embryonic and larval wing development. Thus, none of the genes could be excluded or prioritized because they both have similar embryonic and wing expression patterns – as expected for a gene in the BFS locus. *dome* and *wash* are known to have different expression patterns and functions during embryonic development in *D. melanogaster* (see sections above). In addition, both genes are considered to be highly conserved throughout the evolutionary history of invertebrates. For that reason, the expression patterns of *dome* and *wash* during embryonic development is expected to be conserved in *B. anynana*. In contrast to the patterns of expression in *D. melanogaster,* we found strong correlations between expression patterns of both genes in *B. anynana.*

During the final larval instar, establishment of the location of eyespot foci takes place and one of the genes known to be expressed in the focal cells is *en.* However, how this and other genes described to be expressed in and around the focal cells regulate each other and the downstream pathway leading to focal signalling is not known. Considering the fact that the BFS locus affects segment polarity by altering the domains of expression of *en*, and that both candidate genes are expressed in all epidermis except where *en* expression locates to, it is conceivable that both candidate genes play a role in eyespot development by regulating, for instance, en expression domains. If the expression of the BFS candidate genes regulates the expressing *en*, then mutations in the BFS could affect both establishment of the organizer and the epidermal cells that surround it.

It is known that eyespot size is primarily dependent on properties of the focal signal whereas colour composition is entirely accounted for by properties of the tissue surrounding the foci (reviewed in [14]). It was shown that in Spr mutants, the ability of the entire wing to respond to eyespot-inducing signals is altered[19]. Consequently, the BFS candidate gene is predicted to be downstream of the eyespot-inducing signal (e.g Wg) but upstream of eyespot patterning genes (e.g *en* and Dll).

Because alleles of the BFS locus affect both size and colour composition, properties established during pupal wing development, it would be interesting to perform mRNA and protein detection of the BFS candidate genes in pupal wings to assess whether expression during signalling from the focus and response from epidermal cells occurs. Expression patterns during larval and pupal wing development in different Spr and Fro phenotypes would also new bring insight on the role of the candidate genes during different stages of eyespot development.

3. Functional validation of *dome* and *wash* during wing development

Expression pattern analysis and sequence polymorphisms consistent with BFS role in eyespot development and BFS allele segregation, respectively, are important insights but a functional validation is needed to implicate the candidate loci in these processes.

To test the potential role of *wash* and *dome* in wing and eyespot development, we tried to manipulate both candidate genes using genome editing with Clustered Regularly Interspaced Short Palindromic Repeats/associated sequence 9 (CRISPR/Cas9[42]). This technique holds promise of expanding gene functional assays outside the classical models.

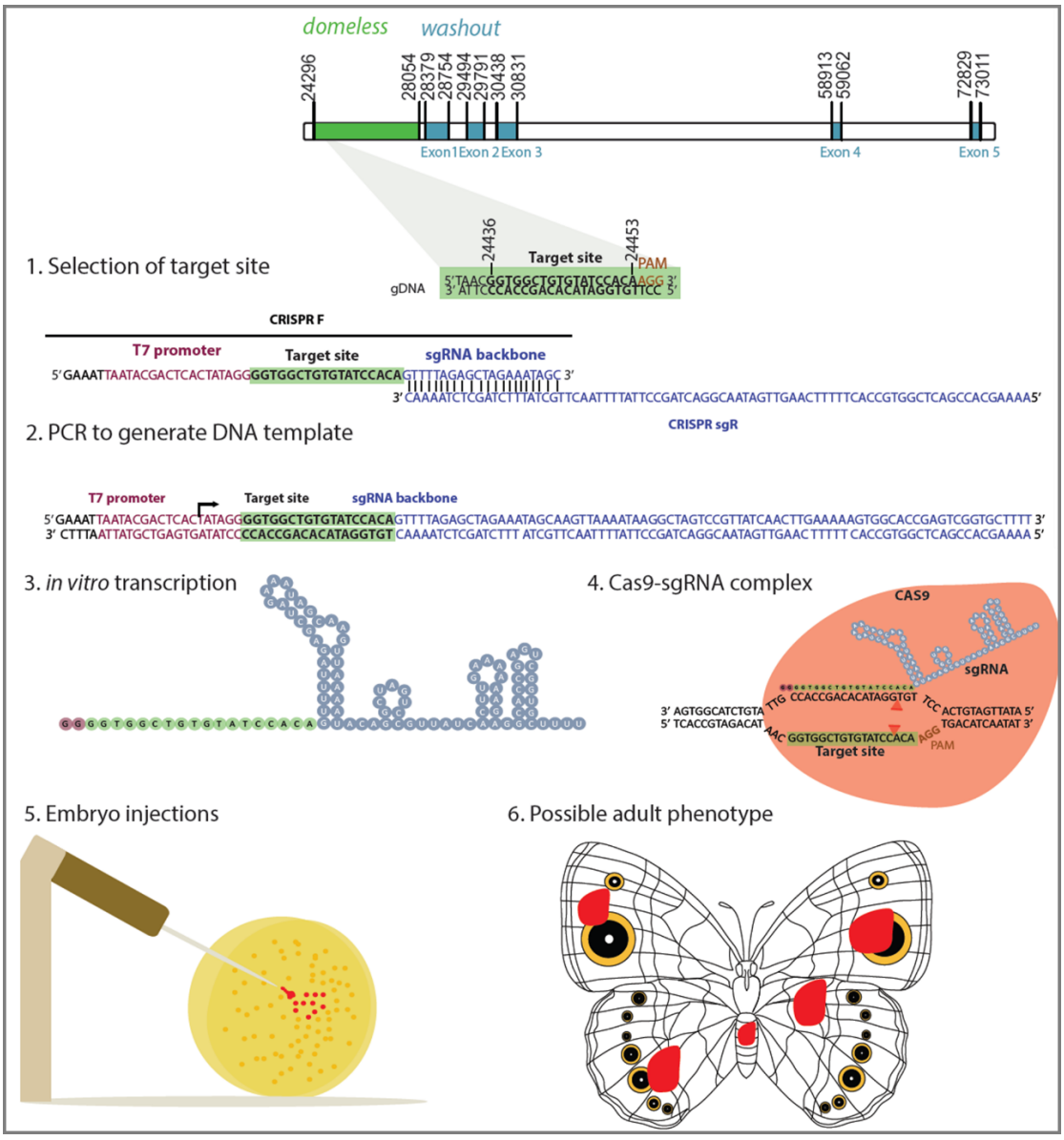


Figure 6. Experimental approach used for functional analysis with CRISPR-CAS9 system. 1. Selection of target site within the BFS locus (*dome* in green and *wash* in blue) with *dome* targeted as example. A target sequence (green) followed by a Protospacer Adjacent Motif (PAM; brown) was chosen (see Material and Methods) and CRISPR F was designed with both T7 promoter and sgRNA (single synthetic guide RNA) backbone, the region complementary to CRISPR sgR (purple). **2.** PCR was used to generate the template for *in vitro* transcription of the sgRNA. T7 transcription start site is indicated with an arrow. **3.** Predicted mature RNA sequence produced by *in vitro* transcription of the sgRNA. 4. Cas9-sgRNA complex showing the target site (green), sgRNA (green and blue), PAM (brown) and Cas9 protein indicated as the red oval. Cleavage site of the target is indicated with red arrows. 5. Injection of embryos in early embryonic

development stages (syncyncial stage) affects only some cells (in red). 6. Expected adult phenotypes are mosaics, with patches affected and non-affected cells by CRISPR/CAS9 (red). Image adapted from[43]).

Successful cases include different species of fish, mice, molluscs and protozoans (reviewed in [59]). Excitingly, this technique has also been used successively in different lepidopterans, with an efficient genome editing that yields expected phenotypes in different genes[60,61] and seems effective to study genes involved in wing patter development (Arnaud Martin, personal communication).

To test this technique in *B. anynana*, we injected embryos with the sgRNAs and purified Cas9 protein (Fig. 6, 4-6; see Materials and methods). Since one or both of our target genes cause embryonic lethality (see section above), we injected embryos with zero to four hours, in order to affect cells in the syncytial stage. Thus, we expected that not all cells may be edited and that adult phenotypes will have patches of cells wild-type and edited.

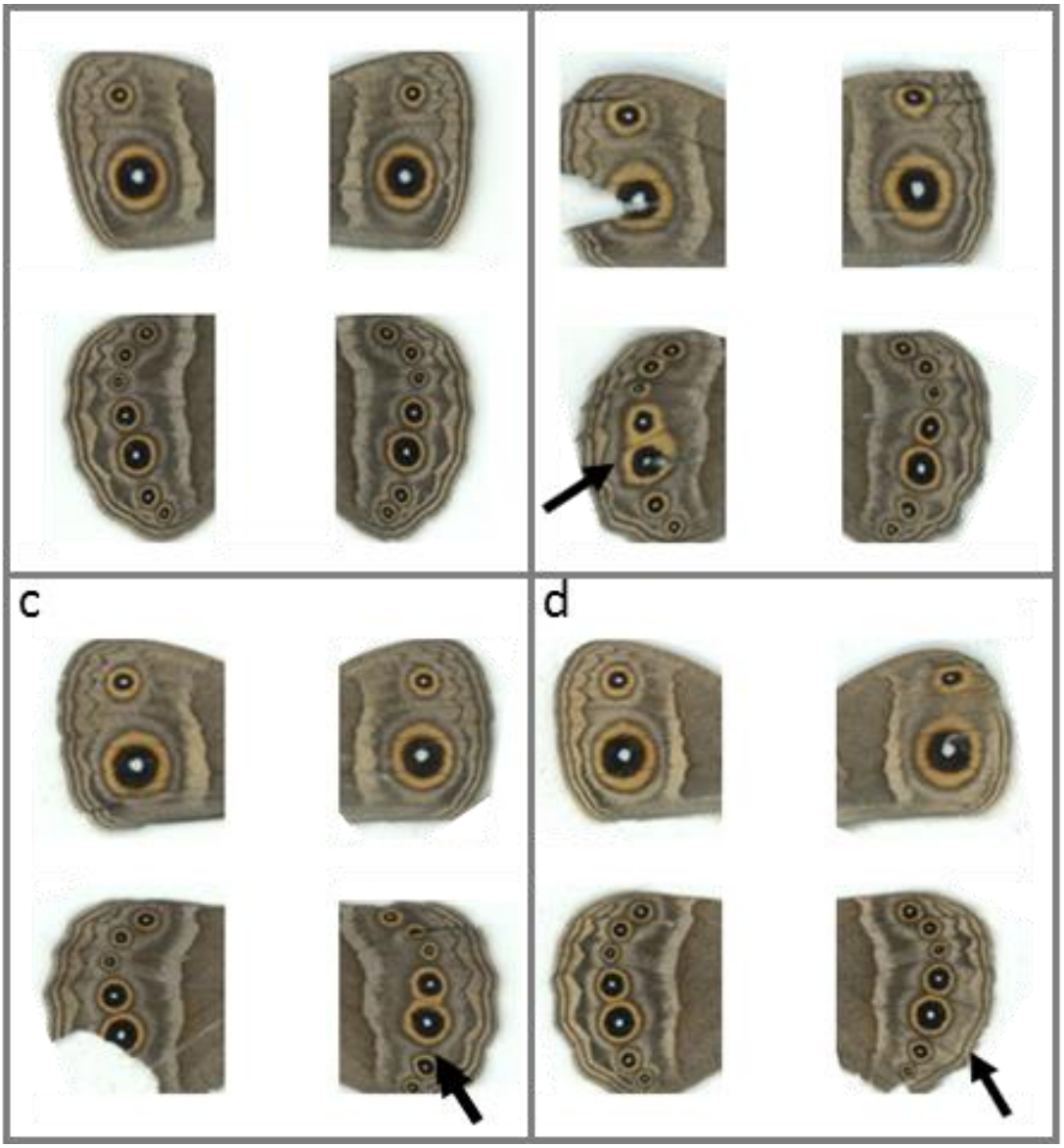


Figure 7. Phenotypes obtained after CRISPR/CAS9 manipulation. Ventral view of distal half of all four wings, forewings on top and hindwings bottom. **a.** WT phenotype; **b.** individual with disrupted 4th and 5th eyespots, with a broader golden ring that resembles Fro phenotype (black arrow, *cf.* Fig. 1); **c.** individual with wider 4th and 5th eyespots in the right hindwing, resembling a BE phenotype (black arrow, *cf.* Fig. 1) and **d.** individual with the dark band that surrounds the eyespots distally missing (from 3rd to 7th eyespot in the hindwing).

In the first injection trials (Table S4), in which we induced DSB in *dome*, the survival upon injection was low (10%), with only 18 individuals surviving until adulthood. From those, three individuals had differences in eyespot patterns (Fig. 7). These individuals appear to be mosaics, since only portions of the wing are affected and eyespot patterns resemble BFS mutant phenotypes (Fig. 1). Comparing to WT, one individual holds wider golden rings in the 4th and 5th eyespots (Fig. 7b) resembling Fro phenotype and one has broader eyespots than WT, resembling BE (Fig. 7c). Finally, we found one individual with the dark band that surrounds the eyespots distally, missing (Fig. 7d), or not as prominent as in WT. This phenotype is also present in Spr and Fro mutants. Although the survival upon injection in the second trial was higher (Table S4), a significant number of individuals died during the process of transference from the injection plate to the plant. From the individuals eclosed in this latter trial, so far no adult phenotype could be detected (15 out of 39 pupae for Dome sgRNA and 14 out of 65 pupae for Wash).

From this preliminary experiment, we obtained phenotypes that resemble BFS mutant but the sample size is too low for us to draw any conclusion. Moreover, in lab populations, variants with disturbed eyespots in only one wing appear occasionally. To show that those regions in the wing are mutants for the target genes, we will have to proceed to genotyping by extracting DNA from predicted edited cells and non-edited cells.

Future studies should aim to get: higher sample sizes, reduced handling and also trying other CRISPR/CAS9 protocols. Also, by inducing DSB specifically in our candidate DNA polymorphisms (see Sequence variants in different genotypic classes) for the different alleles of the BFS locus, we will be able to assess which one is the causative for the phenotypes observed. Unpublished studies with other butterflies targeting wing pigmentation genes look very promising. Most likely, this technique will help to resolve the identity of the hotspot locus and if that is the case, to define general aspects of the genetic basis of adaptive evolution.

4. Sequence variants in different BFS genotypic classes

In order to identify DNA sequence variants in *dome* and *wash* that could correspond to effector polymorphisms responsible for BFS phenotypes, we sequenced their CDS and intergenic region.

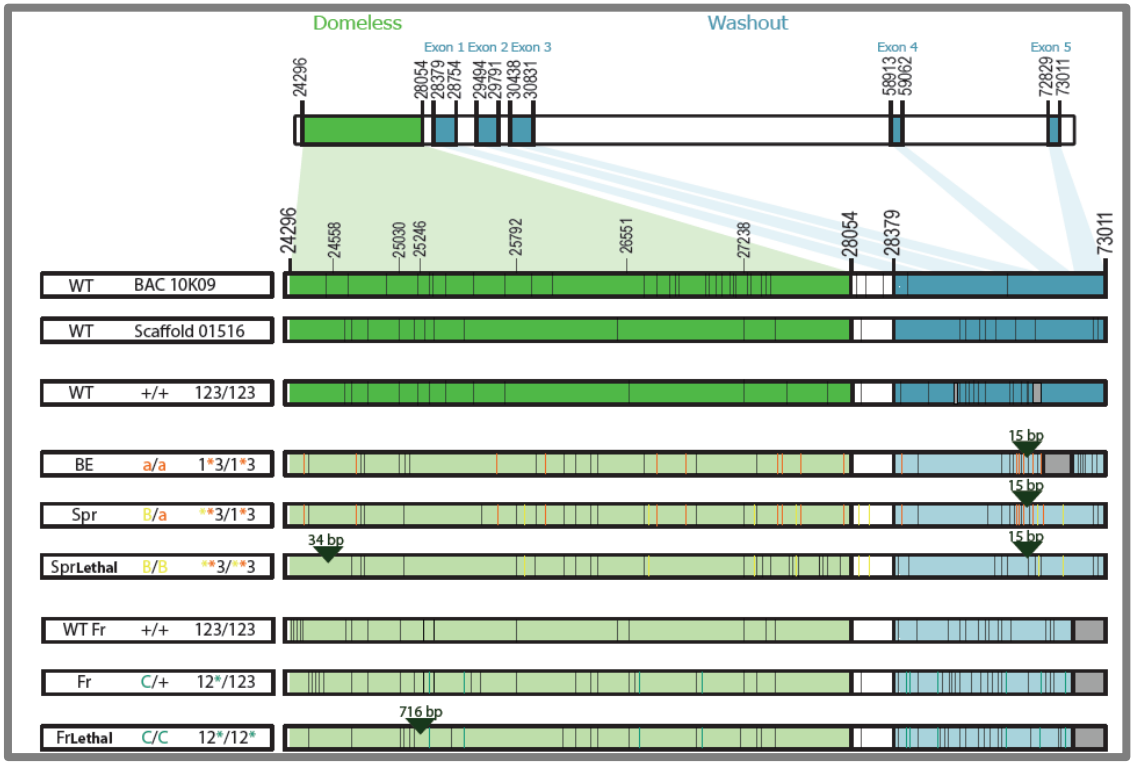


Figure 8. **Sequence variant analysis of *dome* and *wash*.** Each genotypic class and corresponding BFS alleles (*cf.* Segregation Model, Fig. 1) are represented (details on the left side boxes with phenotype, genotype and proposed site mutations *cf.* Fig. 1). Top bar represents *dome* (green) and *wash* (blue) annotation (*cf.* Fig. 2d). Sequence of *dome* and *wash* exons amplified from gDNA from seven genotypic classes (new data) were compared to corresponding sequences from BAC clone and from whole-genome sequence scaffold 01516. SNPs are represented with vertical lines, while insertion polymorphisms are represented by inverted triangles (length on top). Lines in BAC represent SNPs for which the allele in the BAC clone is not found in any other of the sequences. Lines in all other bars correspond to SNPs in relation to the BAC sequence. Line colour corresponds to allelic variants consistent with the proposed BFS alleles in Fig. 1a: a allele (found only in adult BE and Spr) in orange, B (found only and exclusively in Spr adults and dead embryos) in yellow, C (found only and exclusively in Fro adults and dead embryos) in blue. SNPs in grey are allelic variants that do not correspond to the BFS proposed model. Grey areas in the bars were not covered by sequencing. Details of all sequence variants in Table S5.

To generate the different genotypic classes corresponding to the BFS phenotypes, we used two types of crosses (Fig. 1b, left and middle panel) and extracted genomic DNA. We sequenced the following genotypic classes: BE, Spr, Spr lethal embryos (SprL), WT from Fro, Fro and Fro lethal embryos (FroL) plus individuals from the laboratory' outbred WT stock. Sequences were manually assembled and variant analysis was performed using BAC clone 10k09 as reference (See Materials and methods). All sequences were also compared with the available whole-genome sequence scaffold 01516 (see Materials and Methods). The regions analysed (Fig. 8 and Table 1) covered a total of approximately 5.5 kb: 48 bp of intergenic region 5' of *dome* (Intergenic1), 3759 bp of *dome* CDS, 324 bp of intergenic region between *dome* and *wash* (Intergenic2), and 1392 bp of *wash* CDS (coverage not complete for all genotypic classes in this gene).

Table 1. Sequencing coverage of candidate genes *dome* and *wash* and intergenic regions.

	Genomic Region							
	Intergenic 1	*dome*	Intergenic2	*wash*				
				Exon 1	Exon 2	Exon 3	Exon 4	Exon 5
Positions (BAC10k09)	24248-24295	24296-28054	28055-28378	28379-28754	29494-29791	30438-30831	58913-59062	72829-73011
Total length (bp)	48	3759	324	376	298	394	150	183
Sequencing Coverage								
WT	23	3759	381	357	292	301	150	183
BE	32	3759	381	376	298	394	0	183
Spr	48	3759	381	376	298	394	150	183
SprL	48	3759	381	376	298	394	150	183
WT Fro	48	3759	381	376	296	394	150	0
Fro	47	3759	381	376	272	394	150	0
FroL	48	3759	381	376	297	394	150	0
Variant analysis (BAC10K09 as reference sequence)								
Total variants	0	87	6	5	17	29	7	7
Synonymous	0	74	1	5	17	14	1	7
Non-synonymous	0	14	5	0	0	14	6	0
Conservative	-	2	0	-	-	3	0	-
Non-conservative	-	12	5	-	-	11	6	-
Indels	0	3	0	0	0	1	1	0

Note: summary data on total variants found in each sequenced region and corresponding number of amino acid changes

We found a total of 157 variants across the sequences accessed in all genotypic classes, comparing to BAC 10k09 (Fig. 8 and Table 1). This number includes variants only found in BAC 10k09 (*cf.* Table S5 and Materials and Methods). So that these variants can be putative causative DNA polymorphisms responsible for the mutant phenotypes, they have to be consistent with the proposed model of segregation of the BFS alleles (Fig. 1a) and match the remaining criteria proposed in Box 1. From the 33 variants found in homozygous and heterozygous mutants (21%) (*cf.* Table 2), five correspond to non-synonymous mutations (only one conservative) and two are

located in a non-coding region (only considered because one leads to the formation of a stop codon and the other to a starting codon).

Box 1. Criteria used to find variants consistent with the model proposed for BFS allele's segregation.

(a) Variants must be found in homozygosity and heterozygosity. For instance, if a variant is found in SprL (homozygous for B allele) it necessarily has to be present in Spr (heterozygous for B and a alleles). For each variant that could represent a, B or C alleles we used a colour code (orange, yellow and blue, respectively; see Fig. 1a and Fig. 8), **(b)** if present in both homozygous and heterozygous mutants, this event must have been confirmed in the sequencing chromatogram (in which that site has both alleles represented). **(c)** SNPs in CDS must be non-synonymous and non-conservative. In this scenario, the original amino acid is replaced by another with different physicochemical properties leading to a higher probability of altering protein structure.

Although two of the SNPs found in homozygous and heterozygous mutants, they do not fulfil all the criteria proposed in Box 1 (b). However, the fact that we do not see heterozygous peaks may be due to artefacts from sequencing (for instance if the background noise that did not allow the confirmation of a second peak). Hence, considering the current segregation model of the BFS mutants (Fig. 1a) and our criteria, there are six candidate DNA polymorphisms in coding regions: two possible representatives of the a allele, three of C allele and one of B allele (Table 3). The single non-conservative mutation found in *dome* (position 26823), for instance, is a good candidate for being the a allele. It occurs in the predicted extracellular domain and leads to the replacement of a leucine for an arginine (*cf.* Fig. S1, asterisk1), which are residues with different properties (both in polarity and hydrophobicity). This may cause an impact on the structure of the receptor and consequently affect its affinity to the JAK-STAT pathway ligands and/or its dimerization[62].

Table 2. Variants found in homozygous and heterozygous BFS mutants and corresponding amino acid changes. Each variant is a candidate for a BFS allele (a, B or C) and it is highlighted with the corresponding colour (orange, yellow or blue, respectively).

			dome								Interg2		*wash*									
source DNA	phenotype	position in BAC	25570	25918	26695	26823	26902	26965	27776	27916	28070	28167	28528	28639	28681	30675	30676	30685	30705	30755	30795	59045
Inbred lines	WT	BAC	C	C	T	T	C	T	C	G	C	C	C	C	C	G	C	G	T	C	A	G
	WT	Scaff01516	T	C	T	T	C	T	C	G	C	C	C	C	C	G	C	G	T	C	A	A
Outbred line	WT	+/+	C	C	T	T	C	T	C	G	C	C	C	C	C	G	C	G	T			G
Spr crosses	BE	a/a	C	T	T	G	T	T	C	T	C	C	T	C	C	A	G	C	C	C	G	
	Spr	B/a	C	Y	Y	K	Y	T	Y	K	Y	Y	T	C	C	M	S	S	Y	Y	R	G
	SPRLethal	B/B	C	C	C	T	C	T	T	G	T	T	C	C	C	G	C	G	T	T	A	G
Frodo Crosses	WT	+/+	T	C	T	T	C	T	C	G	C	C	C	C	C	G	C	G	T	C	A	G
	Fro	C/+	Y	C	T	T	C	Y	C	G	C	C	C	M	Y	G	C	G	T	C	A	R
	FroLethal	C/C	C	C	T	T	C	C	C	G	C	C	C	A	T	G	C	G	T	C	A	A
Amino acid			V	S	D	L>R	G	Y	L	P	Q>STOP	T>M	D	S	N	V	P>A	V>L	P	A>V	P	A>L

Each amino acid change is represented with > separating amino acid in WT and amino acid in mutant

In *wash*, the candidate SNP in position 59045 that leads to the replacement of an alanine for a threonine (*cf.* Fig. S1, asterisk2-6) occurs in the predicted V domain (WH2). This domain is part of a module required for binding and activation of actin polymerization via Arp 2/3 complex[37].

Although it does not occur in one of the four conserved residues that are essential for actin binding[37], changes in the properties of residues in this domain may lead to changes in the protein conformation.

Table 3. Non-synonymous mutations of candidate genes, corresponding aminoacid changes, predicted protein domains affected and predicted BFS allele.

Position	Gene	BAC	Amino acid	Properties	Mutant allele	Amino acid	Properties	Predicted domain
26823	***d***	T	Leu	Aliphatic, non-polar, hydrophobic	G	Arg	Positively charged, polar, hydrophilic	Extracellular
29503	***w2***	T	Leu	Aliphatic, non-polar, hydrophobic	A	His	Positively charged, polar, hydrophilic	WHD1-WHD2 boundary
30676	***w3***	C	Pro	Aliphatic, non-polar, neutral	G	Ala	Aliphatic, non-polar, hydrophobic	WHD2-P boundary
30755	***w3***	C	Ala	Aliphatic, non-polar, hydrophobic	G	Gly	Aliphatic, non-polar, neutral	P
30760	***w3***	C	Pro	Aliphatic, non-polar, neutral	T	Ser	Hydroxyl, uncharged, polar, neutral	P
59045	***w4***	G	Ala	Aliphatic, non-polar, hydrophobic	A	Thr	Hydroxyl, uncharged, polar, neutral	V

d - *dome*; *w* - *washout* followed by corresponding exon; P - Proline reach domain; V - portion of the conserved domain VCA (WH2)

In addition, we found insertions in different genotypic classes (*cf.* Table 1 and Table S5). In *dome,* we found a 34 bp insertion in SprL and a 787 bp in FroL and both lead to a frameshift. Interestingly, by blasting this latter sequence against *B. anynana* genome database, we found 20 hits in different scaffolds, which might indicate that this region could be a repetitive region, possibly a transposable element (TE). In *wash* the same 15 bp insertion that inserts five amino acids was found in all individuals that segregate in a Spr cross (without inducing frameshift; Fig. 1a).

The insertions in *dome* were only found in Spr and Fro lethal embryos. Although these do not follow the segregation model, (*cf.* Box 1), this may be due to a technical problem. When amplifying by PCR, it is possible that in the heterozygote there is preferential amplification of the smaller fragment where the insertion is absent. By amplifying with one primer inside the insertion and another outside, we will probably overcome this problem and see the insertion also in both heterozygous. The 34 bp insertion in Spr and SprL would be the B allele and the 787bp insertion in Fro and FroL, the C allele. Both insertions lead to a frameshift and occurrence of a premature stop codon in different sites (see Fig. S1). We predict that, in this truncated protein, the FnIII extracellular domains (most importantly, the CBM) are absent whereas FroL insertion occurs after the predicted domains. Since in *D. melanogaster* it has been shown that these domains are important for embryonic development, their absence in Spr lethal embryos could be related with the fact that they die earlier than Fro (Spr dies at 70% developmental time (DT) whereas Fro at about 90% DT [19].

The two SNPs found in non-coding regions are also possible candidates for the B allele (positions 28070 and 28167, Table 3). Both occur in the intergenic region between *dome* and *wash*, where the predicted 3' untranslated region (UTR) of *dome* and the 5' UTR of *wash* are. UTRs are known to play essential roles in post-transcriptional regulation of gene expression. 5' UTRs, for example, regulate translation initiation whereas 3' UTRs can spatially and temporally regulate mRNA expression and are essential for embryonic development[63]. Since mutations in these regions can strongly affect gene expression, it would be interesting to confirm whether these mutations occur in the UTRs. Moreover, because one of these SNPs (position 28167) leads to the formation of a start codon (ATG instead of ACG), this may have implications in the translation of *wash* gene. When a functional AUG precedes the main AUG of a gene, translation can occur in the upstream start codon and inhibit initiation of translation in the downstream one (first-AUG rule). Or, if an AUG is followed by an in-frame stop codon (which is the case for the ATG produced in the B allele) and both are upstream of the main AUG this will lead to the formation of an upstream open reading frame (uORF)[64]. Interestingly, it has been shown that uORFs play an important role in translational efficiency and mRNA stability, particularly in reduction of translation of the main ORF. Considering *wash* as a possible negative regulator of

the Wingless pathway, the appearance of a new ATG in the 5' UTR of the gene could lead to a reduction of Wash expression by 30-80%[64]. In homozygosity, this allele would probably lead to embryonic lethality and lead to changes in eyespot morphology by the absence of Wash inhibiting Wg, both phenotypic characteristics consistent with the B allele.

These results show that different DNA polymorphisms that follow the allele segregation proposed in the model (Fig. 1a), both in coding regions of *dome* and *wash* and in their intergenic region, could be responsible for the BFS phenotypes. Hence, we could not exclude the candidate genes from BFS locus. Because it was shown by experimental crosses and complementation tests that BFS mutants are different alleles within the same locus, one possible explanation is if the 40 kb region of the BFS locus contains a regulatory region. Even though we only analysed CDS of the region containing the BFS locus, mutations in regulatory regions may also be implicated in the BFS phenotypes, as it has been shown before in other studies of genetic basis of adaptive traits [65,66]. Such regulatory regions could be of the candidate genes *dome* and *wash* or other genes outside the mapped BFS locus and/or small non-coding RNAs[67]. Nonetheless, it is still necessary to confirm the variants found in this study (including those in which heterozygosity was not observed) not only with a higher sample number but also in other genotypic classes. For instance, segregating individuals from a Spr and Fro crosses, heterozygotes for B and C alleles or heterozygotes for B and C alleles (Fig. 1b, right panel).

IV. Concluding remarks

With this work we have characterized the BFS locus, a genomic region implicated in segment polarity in *B. anynana* embryos and in establishing eyespot size and colour composition in adult butterflies. We have found that candidate genes *dome* and *wash* have conserved domains in their amino acid sequences, known to be important for their functions in *D. melanogaster*. By considering each of these sequences in a phylogenetic context, with other invertebrate's orthologous, we predicted that they are conserved among lepidopterans. We also examined the expression patterns of the candidate genes throughout embryonic and wing development and found that both *dome* and *wash* are expressed in particular locations (e.g. anterior part of all segments in the embryo and in wing epidermis except where the presumptive eyespot organizer is). In addition, we have found nucleotide differences in both candidate genes *dome* and *wash* as well as in their intergenic region that can represent different alleles of the model proposed for the BFS effects. These results suggest that one or both candidate genes can harbour the allelic variation responsible for the BFS mutant phenotypes.

Further studies should aim to assess *dome* and *wash* expression patterns in pupal wings, when the establishment of the eyespot rings occurs. Also, it would be interesting to confirm whether the expression patterns of both candidate genes are disrupted in developing wings of BFS mutants or if they differ from other *Bicyclus* species whose phenotypes are similar to the BFS mutants (e.g. *B. taenias* with wider golden rings that resemble Fro phenotype[19]). Relative to nucleotide sequence variation found in this study, future experiments should confirm these results and also find the effector sites that correspond to the segregating alleles of the BFS locus. Subsequently, expanding this study to natural populations to assess both natural allelic variation in and across species and compare it with the hotspot-containing genomic region in different lepidopterans will shed light on the hotspot evolutionary history. Even though it seems unlikely that recessive lethal alleles of the BFS locus will contribute to natural variation, it is possible that non deleterious alleles at the same locus might.

Many conserved genes, such as the candidates explored in this work, have been shown to be expressed in suggestive patterns during eyespot formation. However, functional tests are still needed to confirm that they are indeed involved in pattern formation. With the fast pace of progress in methods for genetic engineering, it is now possible to perform functional analysis in non-classical models as *B. anynana*. We have tested the CRISPR/CAS9 technique, a tool that allows induction of site mutations with precision in the genome. Although the results shown here are very preliminary, it will be exciting to confirm not only the role of our candidate genes in eyespot development but also other genes that have been shown to be involved in pattern formation. This way, it will be possible to explore how gene networks for eyespot formation are established and how they have evolved.

Ultimately, the type of integrative analysis we present here will help to resolve the genetic basis of the hotspot locus for butterfly evolution. While this work was ongoing, studies in *Heliconius spp.*[68] and *B. betularia* moths (not published) identified the locus responsible for inter-specific variation in these species by mapping and population genetics. The implicated gene, *poikilomousa* (*poik*) is mapped outside the BFS region. Hence, we do not know if 1) our region contains a non-coding region involved in *poik* regulation, 2) our candidates' contribution to variation in *Heliconius spp*. butterflies and *B. betularia* moths. Taking all of this into account, more studies in these lepidopteran species are still needed to confirm whether the same locus is implicated in inter-specific variation. If the locus turns out to be the same, then we are more close to understand why and how this particular genomic region is contributing to adaptive variation and diversification of lepidopteran wing patterns. Obtain an accurate knowledge of the genetic basis of adaptive traits in different species is only possible with integrative studies that analyse all organisational levels, from genes to populations and including the developmental processes that underlie such traits. The study presented here represents a first step in this

direction and in the future, such approach on different organisms that have an amazing diversity and known adaptive traits could then lead to a better understanding and to derive general principals of the loci of morphological evolution.

V. References

1. Stern, D. Perspective: evolutionary developmental biology and the problem of variation. *Evolution* **54,** 1079–1091 (2000).

2. Gompel, N., Prud, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing : cis -regulatory evolution and the origin of pigment patterns in Drosophila. **433,** 481–487 (2005).

3. Tanaka, K., Barmina, O., Kopp, A. & Carroll, S. B. Distinct the mechanisms underlie developmental sex combs diversification of Drosophila evolutionary. **106,** 4764–4769 (2014).

4. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484,** 55–61 (2012).

5. Hoekstra, H. E., Hirschmann, R. J., Bundey, R. a, Insel, P. a & Crossland, J. P. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313,** 101–4 (2006).

6. Prud'homme, B., Gompel, N. & Carroll, S. B. Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* **104 Suppl ,** 8605–12 (2007).

7. Stern, D. & Orgogozo, V. Is genetic evolution predictable? *Science* **323,** 746–751 (2009).

8. Haag, E. & True, J. Perspective: From mutants to mechanisms? Assessing the candidate gene paradigm in evolutionary biology. *Evolution.* **55,** 1077–1084 (2001).

9. Wittkopp, P. J., Carroll, S. B. & Kopp, A. Evolution in black and white: genetic control of pigment patterns in Drosophila. *Trends Genet.* **19,** 495–504 (2003).

10. True, J. R. Insect melanism: the molecules matter. *Trends Ecol. Evol.* **18,** 640–647 (2003).

11. Wittkopp, P. J. & Beldade, P. Development and evolution of insect pigmentation: genetic mechanisms and the potential consequences of pleiotropy. *Semin. Cell Dev. Biol.* **20,** 65–71 (2009).

12. Cook, L. M. & Saccheri, I. J. The peppered moth and industrial melanism: evolution of a natural selection case study. *Heredity* **110,** 207–12 (2013).

13. Joron, M., Jiggins, C. D., Papanicolaou, a & McMillan, W. O. Heliconius wing patterns: an evo-devo model for understanding phenotypic diversity. *Heredity* **97,** 157–67 (2006).

14. Beldade, P. & Brakefield, P. M. The genetics and evo-devo of butterfly wing patterns. *Nat. Rev. Genet.* **3,** 442–52 (2002).

15. Suetsugu, Y., *et al.* Large Scale Full-Length cDNA Sequencing Reveals a Unique Genomic Landscape in a Lepidopteran Model Insect, *Bombyx mori*. G3 (Bethesda) **3,** 1481–1492 (2013).

16. Beldade, P., Saenko, S. V, Pul, N. & Long, A. D. A gene-based linkage map for Bicyclus anynana butterflies allows for a comprehensive analysis of synteny with the lepidopteran reference genome. *PLoS Genet.* **5,** e1000366 (2009).

17. van't Hof, A. E., Edmonds, N., Dalíková, M., Marec, F. & Saccheri, I. J. Industrial melanism in British peppered moths has a singular and recent mutational origin. *Science* **332,** 958–60 (2011).

18. Ferguson, L. *et al.* Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. *Mol. Ecol.* **19 Suppl 1,** 240–54 (2010).

19. Saenko, S. V, Brakefield, P. M. & Beldade, P. Single locus affects embryonic segment polarity and multiple aspects of an adult evolutionary novelty. *BMC Biol.* **8,** 111 (2010).

20. Papa, R., Martin, A. & Reed, R. D. Genomic hotspots of adaptation in butterfly wing pattern evolution. *Curr. Opin. Genet. Dev.* **18,** 559–64 (2008).

21. Prud'homme, B. *et al.* Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* **440,** 1050–3 (2006).

22. Roberts, R. B., Hu, Y., Albertson, R. C. & Kocher, T. D. Craniofacial divergence and ongoing adaptation via the hedgehog pathway. **108,** 13194-13199 (2011).

23. Abzhanov, A. *et al.* The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* **442,** 563–7 (2006).

24. Sucena, E., Delon, I. & Jones, I. Regulatory evolution of shavenbaby / ovo underlies multiple cases of morphological parallelism. **24,** 935–938 (2003).

25. Wang, X. & Chamberlin, H. M. Evolutionary innovation of the excretory system in Caenorhabditis elegans. *Nat. Genet.* **36,** 231–2 (2004).

26. Carroll, S., Gates, J. & Keys, D. Pattern formation and eyespot determination in butterfly wings. *Science* **265,** 109-114 (1994).

27. Monteiro, A. Origin, development, and evolution of butterfly eyespots. *Annu. Rev. Entomol.* **60,** 253–71 (2015).

28. Hamada, F. *et al.* Negative regulation of Wingless signaling by D-axin, a Drosophila homolog of axin. *Science* **283,** 1739–1742 (1999).

29. Monteiro, A., Glaser, G., Stockslager, S., Glansdorp, N. & Ramos, D. Comparative insights into questions of lepidopteran wing pattern homology. *BMC Dev. Biol.* **6,** 52 (2006).

30. Zhang, L. *et al.* Regulation of wingless signaling by the CKI family in Drosophila limb development. *Dev. Biol.* **299,** 221–37 (2006).

31. Shimomura, M. *et al.* KAIKObase: an integrated silkworm genome database and data mining tool. *BMC Genomics* **10,** 486 (2009).

32. Ekas, L. a, Baeg, G.-H., Flaherty, M. S., Ayala-Camargo, A. & Bach, E. a. JAK/STAT signaling promotes regional specification by negatively regulating wingless expression in Drosophila. *Development* **133,** 4721–9 (2006).

33. Saenko, S. V. Evo-devo of novel traits : the genetic basis of butterfly colour patterns. *Evol. Biol. Inst. Biol. Leiden (IBL), Fac. Sci. Leiden Univ.* (2010).

34. Brakefield, P. M., Beldade, P. & Zwaan, B. J. The African butterfly Bicyclus anynana: a model for evolutionary genetics and evolutionary developmental biology. *Cold Spring Harb. Protoc.* **5,** pdb.emo122 (2009).

35. Saenko, S. V, French, V., Brakefield, P. M. & Beldade, P. Conserved developmental processes and the formation of evolutionary novelties : examples from butterfly wings Conserved developmental processes and the formation of evolutionary novelties : examples from butterfly wings. *Phil. Trans. R. Soc. B* **383,** 1549-1555 (2008).

36. Brown, S., Hu, N. & Hombría, J. Identification of the first invertebrate interleukin JAK/STAT receptor, the Drosophila gene domeless. *Curr. Biol.* **11,** 1700–1705 (2001).

37. Linardopoulou, E. V *et al.* Human subtelomeric WASH genes encode a new subclass of the WASP family. *PLoS Genet.* **3,** e237 (2007).

38. Reed, R. D., Chen, P. & Nijhout, H. F. Cryptic variation in butterfly eyespot development : the importance of sample size in gene expression studies. *Evol. Dev.* **9,** 2–9 (2007).

39. Broadie, K. S., Bate, M. & Tublitz, N. J. Quantitative staging of embryonic development of the tobacco hawkmoth, Manduca sexta. *Roux's Arch Dev Biol* **199,** 327–334 (1991).

40. Couso, J., Bate, M. & Martinez-Arias, A. A wingless-dependent polar coordinate system in Drosophila imaginal discs. *Science* **259,** 484-489 (1993).

41. Tanaka, K. *et al.* Multispecies Analysis of Expression Pattern Diversification in the Recently Expanded Insect Ly6 Gene Family. *Mol. Biol. Evol*. **32,** 1730–1747 (2015).

42. Harrison, M. M., Jenkins, B. V, Connor-giles, K. M. O. & Wildonger, J. A CRISPR view of development. *Genes Dev*. **28,** 1859–1872 (2014).

43. Bassett, A. & Liu, J. CRISPR / Cas9 mediated genome engineering in Drosophila. *Methods **69**, 128-136* (2014).

44. Ren, X. *et al.* Enhanced Specificity and Efficiency of the CRISPR/Cas9 System with Optimized sgRNA Parameters in Drosophila. *Cell Rep.* **9,** 1151–1162 (2014).

45. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8,** 186–194 (1998).

46. Joshi, T. & Xu, D. Quantitative assessment of relationship between sequence similarity and function similarity. *BMC Genomics* **8,** 222 (2007).

47. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **12,** 85–94 (1999).

48. Bäumer, D., Trauner, J., Hollfelder, D., Cerny, A. & Schoppmeier, M. JAK-STAT signalling is required throughout telotrophic oogenesis and short-germ embryogenesis of the beetle Tribolium. *Dev. Biol.* **350,** 169–82 (2011).

49. Evans, J., Aronstein, K. & Chen, Y. Immune pathways and defence mechanisms in honey bees Apis mellifera. *Insect Mol. Biol* **15,** 645–656 (2006).

50. Tanaka, H. *et al.* A genome-wide analysis of genes and gene families involved in innate immunity of Bombyx mori. *Insect Biochem. Mol. Biol.* **38,** 1087–110 (2008).

51. Macias-Munoz, A., Smith, G., Monteiro, A. & Briscoe, A. D. Transcriptome-wide differential gene expression in Bicyclus anynana butterflies. *Mol. Biol. Evol.* (2015).

52. Marchand, J., Kaiser, D., Pollard, T. & Higgs, H. Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. *Nat. Cell Biol.* **3,** 76-82 (2001).

53. Derivery, E. & Gautreau, A. Evolutionary conservation of the WASH complex: An actin polymerization machine involved in endosomal fission. *Commun. Integr. Biol.* **3,** 227–230 (2010).

54. Tomancak, P. *et al.* Systematic determination of patterns of gene expression during Drosophila embryogenesis. *Genome Biol.* **3,** 1–14 (2002).

55. Rodriguez-mesa, E., Abreu-blanco, M. T., Rosales-nieves, A. E. & Parkhurst, S. M. Developmental expression of Drosophila Wiskott-Aldrich Syndrome family proteins. *Dev. Dyn* **241,** 608–626 (2012).

56. Neumann, C. J. & Cohen, S. M. Long-range action of Wingless organizes the dorsal-ventral axis of the Drosophila wing. *Development* **880,** 871–880 (1997).

57. Fujioka, M., Jaynes, J. B. & Goto, T. Early even-skipped stripes act as morphogenetic gradients at the single cell level to establish engrailed expression. *Development* **4382,** 4371–4382 (1995).

58. Peel, A. D., Chipman, A. D. & Akam, M. Arthropod segmentation: beyond the Drosophila paradigm. *Nat. Rev. Genet.* **6,** 905–16 (2005).

59. Bono, J. M., Olesnicky, E. C. & Matzkin, L. M. Connecting genotypes, phenotypes, and fitness: harnessing the power of CRISPR/Cas9 genome editing. *Mol. Ecol.* **24,** 3810–3822 (2015).

60. Ma, S. *et al.* CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of BmKu70 in Bombyx mori. *Sci. Rep.* **4,** 4489 (2014).

61. Li, X. *et al.* Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nat. Commun.* **6,** 8212 (2015).

62. Brown, S. Novel level of signalling control in the JAK/STAT pathway revealed by in situ visualisation of protein-protein interaction during Drosophila development. *Development* **130,** 3077–3084 (2003).

63. Kuersten, S. & Goodwin, E. B. The power of the 3' UTR: translational control and development. *Nat. Rev. Genet.* **4,** 626–637 (2003).

64. Wethmar, K., Smink, J. J. & Leutz, A. Upstream open reading frames: Molecular switches in (patho)physiology. *BioEssays* **32,** 885–893 (2010).

65. Martin, A. & Orgogozo, V. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* ***67*** 1235–1250 (2013).

66. Hoekstra, H. E. & Coyne, J. a. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61,** 995–1016 (2007).

67. Shimoni, Y. *et al.* Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol. Syst. Biol.* **3,** 138 (2007).

68. Nadeau, N., Pardo-Diaz, C. & Whibley, A. The origins of a novel butterfly wing patterning gene from within a family of conserved cell cycle regulators. *BioRxiv* (2015).

VI. Annex 1

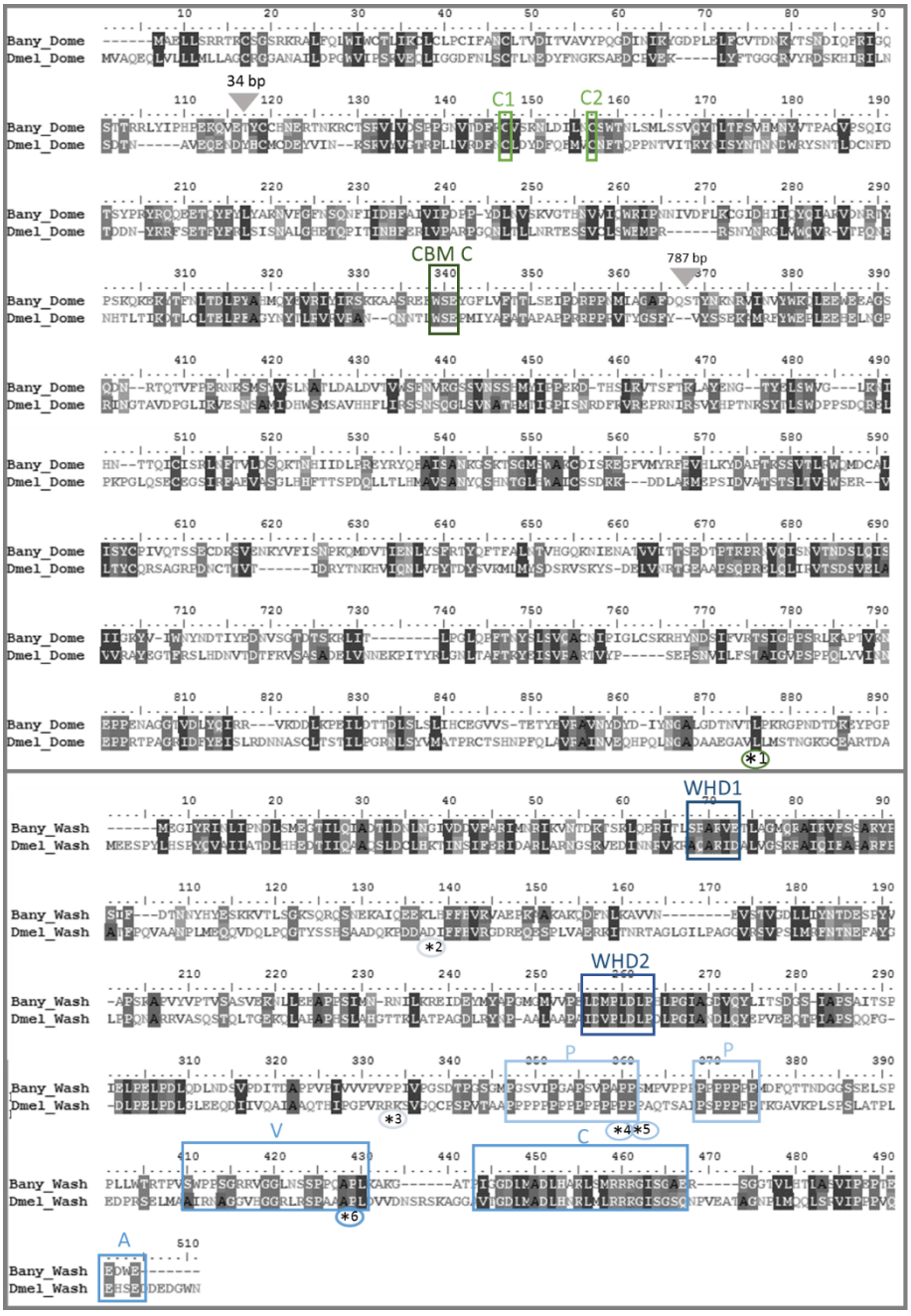


Figure S1. ClustalW alignments of amino acid sequences of *B. anynana* and *D. melanogaster* Dome and Wash. Top panel: Dome alignment. Conserved domains (prediction based on Stephen Brown, Hu, & Hombría, 2001) are shown in green boxes (different green shades correspond to different domains). C1 and C2 correspond to conserved cysteines in the N-terminal of the cytokine binding module (CBM N) and CBM C to C-terminal CBM. Triangles correspond to insertions found in SprL (34 bp insertion) and FroL (787 bp insertion). Bottom panel: Wash alignment. Predicted conserved domains (blue boxes) where inferred based on Linardopoulou et al., 2007. WHD1/2 – Wash homology domains 1/2); P – proline-rich domain; each component of the conserved VCA domain, V- WH2; C – central region and A. * and following number correspond to non-synonymous, non-conservative mutations found in the variant analysis (see Table 3). Both alignments were performed with ClustalW tool in Bioedit (default parameters).

VII. Annex 2

Table S1 –Primers used for sequencing and riboprobe synthesis.

Gene	Forward	Reverse	Amplicon Size	Position BAC (Start)	Position BAC (end)	Anneling Temperature (X)	Extension time (Y)	Used For: Sequencing (S)/ISH (I)_gene_exon
Domeless_0	GACAGTGCGCCATTGATCTT	ATTTGCGAAGGTACACAGGC	684	24163	24846	60°C	60s	S_*dome*
Domeless_1_1	GCCTGTGTACCTTCGCAAAT	CTTTGAGGGCACAGTCCATT	1195	24827	26041	60°C	90s	S_*dome*
Domeless_1_2	GTGGCAACCGAAAGGTGTAT	CTTTTGTTTGGACGGCAAAT	925	24265	25209	60°C	60s	S_*dome*
Domeless_2_1	ATTTGCCGTCCAAACAAAAG	CTTTGAGGGCACAGTCCATT	852	25170	26041	60°C	60s	S/I_*dome*
Domeless_2_2	AAGAAAGCTGCCTCTCGAGA	GTCCATGTACAGTATTGAGAGCA	943	25265	26207	60°C	60s	S_*dome*
Domeless_3_1	AATGGACTGTGCCCTCAAAG	TGTCTCCTAAAGCCCCATTG	809	26002	26830	64°C	60s	S/I_*dome*
Domeless_3_2	CACTTTGCCAGGACTTCAGC	TCTGCACACAAAGTTTGCTTC	1643	26431	28094	60°C	100s	S_*dome*
Domeless_4_1	CAATGGGGCTTTAGGAGACA	TCTGCACACAAAGTTTGCTTC	1283	26791	28094	60°C	90s	S_*dome*
Domeless_4_2	AATGGACTGTGCCCTCAAAG	ACATTTCATCTGAGGGTGGC	1121	26002	27122	60°C	90s	S_*dome*
Domeless_4_3	TTGCCACCACCAACAAAACA	TGTCTAGGGTGTCAGCGATC	733	27731	28463	59°C	60s	S_*dome*
Wash_e1_1	AATCAATCTGATACCCAACGA	CGATTGCCTCTGAGACTTCC	358	28396	28754	60°C	30s	I_wash_exon 1
Wash_e1_2	CTCGCAGTGACACAATAATCG	CGATTGCCTCTGAGACTTCC	632	28104	28754	60°C	60s	S_wash_exon 1
Wash_e1_3	TGTTCCTATAGTCACAGTGAGGT	CGTTGACACTTCGTTGACCA	1314	28281	29594	60°C	90s	S_wash_exon 1
Wash_e2_1	CTCGCAGTGACACAATAATCG	ACATAGGCCCCGAGAGATTT	1753	28104	29810	60°C	100s	S_wash_exon 2
Wash_e2_2	AATCTGAAGGCCGTGGTCAA	CGTCGCTTGTGATCAGGTAC	959	29562	30520	60°C	90s	S_wash_exon 2
Wash_e3_1	GGTACCGGAGTTGGACATGC	CCATCATTTGTTGTTTGGAAG	393	30438	30831	60°C	30s	S_wash_exon 3
Wash_e3_2	GGTACCGGAGTTGGACATGC	GCACATAAAACGAATCCTGCG	771	30438	31208	60°C	60s	S_wash_exon 3
Wash_e4_1	GCGCGGCGTCTACTAGTTT	TCGTACCCTCCTTACCCAAA	822	58529	59350	58°C	60s	S_wash_exon4
Wash_e5_1	CGGAACTCTAAATCGCTTGG	CATTTTGATTGAGCGACTGC	859	72562	73420	58°C	60s	S_wash_exon5

VIII. Annex 3

List of solutions used for *in situ* hybridization experiment

PBS 50mM EGTA For 50mL:
5mL PBS10X
5mL EGTA 0.5M
40mL dH2O

PBT = PBS 0.01% Tween 20 For 50 mL
50mL PBS
200uL Tween20 25%

Fix For each 1mL:
750 uL of PBS 50mM EGTA
250uL of formaldehyde 37%

Tween20 25% 12mL:
9mL sterile dH2O
3mL Tween 20 (very viscous, so just pour it) (Sigma#P-1379)
Mix gently until homogenous, Protect from light, Store @RT

EGTA 0.5M, 100mL:
19.02g in 90mL sterile dH2O
Should completely dissolve at RT once at pH=8 (adjust pH with NaOH)
Filter, Autoclave, Store @4C or RT
Post Fix (freshly prepared). For each 1mL:
150 uL formaldehyde 37%, 850 uL PBT.

Stop Solution = 2mg/mL glycine in PBT For 50mL :
10mL glycine 10g/L (this stock solution can be filtered, autoclaved and kept @4C)
5mL PBS 10x
200uL Tween20 25%
fill up with sterile dH2O until 50mL

PreHyb 50 mL:
10mL sterile dH2O
25mL formamide
12.5mL 20X SSC(check that pH=4.5 prior to addition)
200uL Tween20 25%
500uL 10mg/mL Salmon (or Herring) Sperm DNA (heat denature 5' @80C prior to addition)
fill up to 50mL with sterile dH2O. Store @-20C.

Hyb 50 mL:
5mL sterile dH2O
5mL glycine 10g/L
25mL formamide
12.5mL 20X SSC pH 4.5
200uL Tween20 25%
500uL 10mg/mL Salmon (or Herring) Sperm DNA (heat denature 5' @80C prior to addition)
fill up to 50mL with sterile dH2O. Store @-20C.

20x SSC, 1L:
900mL dH20
175.3g NaCl
88.2g Sodium Citrate, dihydrate
Once at room temperature, lower pH to 4.5. Fill-up to 1L. Filter, Autoclave, Store at RT

Alkaline Phosphatase Buffer, 50mL:
(freshly prepared)
250 uL of $MgCl_2$ 1M
5mL of NaCl 1M
5mL of Tris 1M (pH9.5)
200 uL Tween20 25%.

Staining Solution, each 1mL
(freshly prepared, protect from light)
1mL AP Buffer (check that pH=9.5)
3.5uL BCIP (protect from light)
4.5uL NBT (protect from light)

1M Tris pH=7.5, 1L:
121.1 g of Tris Base dissolved in 900mL H2O.
Once at room temperature lower pH to 7.5 (about 50mL of 12.1M HCl may be required)
Fill-up to 1L, filter, autoclave and store @RT

1M Tris pH=9.5, 1L:
121.1 g of Tris Base dissolved in 950mL H2O.
Once at room temperature adjust pH to 9.5
Fill-up to 1L, filter, autoclave and store @RT

Table S2 – Total number of embryos and larval wings assessed in expression analysis, with a specific staining (pattern).

	Embryo %DT									
Probe	0-15%	Pattern	15-25%	Pattern	30-40%	Pattern	40-50%	Pattern	55-65%	Pattern
Domeless_2 AS			25	25	4	3	4	2	16	0
Domeless_2 S	2	No staining	25	No staining	2	No staining		Staining in cuticle	2	Staining in cuticle
Domeless_3 AS			22	13	2	2	6	2	13	0
Domeless_3 S			1	No staining		No staining		Staining in cuticle		
Washout AS			25	13	8	8	9	6		
Washout S			19	No staining	3	No staining	5	Staining in cuticle		
	Larval wing disc stage									
	0-0.75	Pattern	1.0-1.75	Pattern	2.0-2.75	Pattern	3.0-3.75	Pattern	4.0	Pattern
Domeless_2 AS	13	2	15	11	5	4	4	3	1	1
Domeless_2 S	12	2	8	No staining	5	No staining	1	No staining		
Domeless_3 AS	4	2	8	3	3	3	0			
Domeless_3 S	5	No staining	3	No staining	2	No staining	0			
Washout AS	16	9	4	4	12	8	4	3		
Washout S	4	No staining	4	No staining	1	No staining	0			

IX. Annex 4

Table S3 – Oligonucleotides (sgRNAs) used for functional analysis. Nucleotides in grey represent target sequences

Name	Sequence
Dome_sgRNA_F1	GAAATTAATACGACTCACTATAGGGGTGGCTGTGTATCCACAGTTT*AGAGCTAGAAATAGC
Dome_sgRNA2_F2	GAAATTAATACGACTCACTATAGGGGTTGAACCAATTTGCGAGTTT*AGAGCTAGAAATAGC
WashE3_sgRNA	GAAATTAATACGACTCACTATAGGGGTGATGGCCGAGGGTGCGAGTTT*AGAGCTAGAAATAGC
sgRNA_R	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC

Note: a T is missing (*) in all sgRNA backbones (see Materials and methods).

Table S4 – Numbers of injection trials for CRISPR/Cas9 assay. sgRNA used, concentrations, number of eggs, survival and phenotypes obtained.

Injected with	Concetration (ng/uL)		# eggs	Survival upon injection	Survival upon injection (%)	Survied until adulthood	Phenotype different from WT
	sgRNA	Cas9					
Dome1_sgRNA	125	250	669	67	10,01	18	3
Dome1_sgRNA	150	333	555	387	69,73	39	0a
WashE3_sgRNA	150	333	475	421	88,63	65	0b

Note: Not all individuals eclosed by the time this dissertation was written a - 15 individuals; b- 14 individuals

X. Annex 5

Table S5 – Variant analysis of dome (d), intergenic region (I) and wash (w). All variants found in the genotypic classes of this study (CLS) werecompared with previous analysis (SVS) and to Scaffold 01516 using BAC10K09 as reference sequence (see Materials and methods).

				Genotype												
Position (BAC)	Gene	BAC	Scaff01516	CLS_BE	CLS_Spr	CLS_SprL	CLS_WT Fr	CLS_Fr	CLS_FrL	CLS_WT	PE75	SVS_BE	SVS_SprL	SVS_WTFr	SVS_FrL	Potencial Effects
24314	*d*	C	C	C	C	C	M	C	C	C						Synonymous
24319	*d*	A	A	A	A	A	R	A	A	A						Synonymous
24321	*d*	C	T	C	C	C	S	C	C	C						Heterozygotic site
24331	*d*	T	T	T	T	T	Y	T	T	T						Synonymous
24338	*d*	A	A	A	A	A	R	A	A	A						Heterozygotic site
24355	*d*	C	C	T	T	C	C	C	C	C						Synonymous
24403	*d*	T	T	C	C	T	T	C	C	T						Synonymous
24426	*d*	T	T	T	T	T	T	K	T	T						Heterozygotic site
24428	*d*	G	G	G	G	G	G	S	G	G						Heterozygotic site
24438	*d*	T	T	T	T	T	T	W	T	T						Heterozygotic site
24451	*d*	A	A	A	A	A	A	W	A	A						Synonymous
24556	*d*	G	A	A	A	A	A	A	A	A						Synonymous
24558	*d*	T	A	T	T	**insertion**[a]	T	T	T	T						Frameshift; premature Stop Codon
24667	*d*	C	T	C	C	C	T	C	C	T						Synonymous
24700	*d*	T	C	C	C	C	C	C	C	C						Synonymous
24706	*d*	A	T	A	A	T	T	T	T	T						Synonymous
24826	*d*	C	C	T	T	C	C	C	C	C						Synonymous
24877	*d*	A	A	G	G	G	A	A	A	A						Synonymous
24880	*d*	T	T	C	C	C	T	T	T	T						Synonymous
24912	*d*	C	T	C	C	C	T	T	C	T						Non-synonymous

Table S5 (continued)

		Genotype													
Position(BAC)	Gene	BAC	Scaff01516	CLS_BE	CLS_Spr	CLS_SprL	CLS_WT Fr	CLS_Fr	CLS_FrL	CLS_WT	PE75	SVS_BE	SVS_SprL	SVS_WTFr	SVS_FrL	Potencial Effects
25030	*d*	T	C	T	T	T	C	C	C	C						Synonymous
25126	*d*	G	G	A	A	A	G	G	A	G						Synonymous
25159	*d*	G	G	A	G	G	G	G	A	G						Synonymous
25189	*d*	G	A	G	G	G	G	A	A	G						Synonymous
25246	*d*	A	G	G	G	G	G	G	G	A						Synonymous
25349	*d*	A	A	A	A	A	A	A	**Insertion**[b]	A						Frameshift; premature Stop Codon
25372	*d*	C	T	C	C	C	T	T	C	C						Synonymous
25459	*d*	C	T	T	T	T	T	T	T	C						Synonymous
25478	*d*	G	A	A	A	A	A	A	A	A						Non-synonymous
25531	*d*	C	C	C	C	C	C	M	A	C						Synonymous
25570	*d*	C	T	C	C	C	T	Y	C	C						Synonymous
25645	*d*	C	T	T	T	T	T	T	T	C						Synonymous
25690	*d*	G	A	A	A	A	A	R	G	G						Synonymous
25699	*d*	A	G	G	R	A	G	R	A	A						Synonymous
25708	*d*	C	C	T	T	C	C	C	C	C						Synonymous
25771	*d*	T	C	C	C	C	C	C	C	T						Synonymous
25792	*d*	C	C	C	M	A	C	M	A	C						Synonymous
25831	*d*	A	A	A	G	G	A	A	A	A						Synonymous
25867	*d*	A	C	C	C	C	C	C	C	C						Synonymous
25906	*d*	G	G	A	A	A	G	G	G	G						Synonymous

Table S5 (continued)

		Genotype														
Position(BAC)	Gene	BAC	Scaff01516	CLS_BE	CLS_Spr	CLS_SprL	CLS_WT Fr	CLS_Fr	CLS_FrL	CLS_WT	PE75	SVS_BE	SVS_SprL	SVS_WTFr	SVS_FrL	Potencial Effects
25918	*d*	C	C	T	Y	C	C	C	C	C						Synonymous
26041	*d*	T	C	C	C	C	C	C	C	C						Synonymous
26062	*d*	T	T	T	T	T	T	T	C	T						Non-synonymous
26110	*d*	T	T	A	A	A	T	T	A	T						Synonymous
26233	*d*	C	C	T	T	T	C	Y	T	C						Synonymous
26248	*d*	A	A	C	C	C	A	C	C	A						Synonymous
26428	*d*	T	T	A	A	A	T	A	A	T						Synonymous
26551	*d*	T	T	T	T	T	C	T	T	C	T	T	T	C	T	Synonymous
26578	*d*	A	G	G	G	G	R	A	A	G	A	G	G	A	A	Non-synonymous
26638	*d*	A	A	A	A	A	A	R	A	A	A	A	A	A	A	Synonymous
26641	*d*	C	C	C	C	C	C	T	T	C	C	C	C	C	T	Synonymous
26662	*d*	G	A	G	G	G	A	A	A	A	G	G	G	A	A	Non-synonymous
26695	*d*	T	T	T	Y	C	T	T	T	T	T	T	T/C	T	T	Non-synonymous
26823	*d*	T	T	G	K	T	T	T	T	T	T	G	T/G	T	T	Non-synonymous
26842	*d*	T	C	T	Y	C	C	C	C	C	C	T	T/C	C	C	Synonymous
26889	*d*	G	T	T	T	T	T	T	T	T	G	G	G	G	G	Non-synonymous
26892	*d*	G	C	C	C	C	C	C	C	C	G	G	G	G	G	Non-synonymous
26896	*d*	G	C	C	C	C	C	C	C	C	G	G	G	G	G	Non-synonymous
26902	*d*	C	C	T	Y	C	C	C	C	C	C	C	C	C	C	Synonymous
26914	*d*	G	G	A	R	G	G	R	A	G	G	NA	G	G	A	Non-synonymous

Table S5 (continued)

		Genotype														
Position(BAC)	Gene	BAC	Scaff01516	CLS_BE	CLS_Spr	CLS_SprL	CLS_WT Fr	CLS_Fr	CLS_FrL	CLS_WT	PE75	SVS_BE	SVS_SprL	SVS_WTFr	SVS_FrL	Potencial Effects
26965	*d*	T	T	T	T	T	T	Y	C	T	T	T	T	T	C	Non-synonymous
27013	*d*	G	A	A	A	A	A	A	A	A	G	G	G	G	G	Synonymous
27037	*d*	A	G	G	G	G	G	G	G	G	A	A	A	A	A	Synonymous
27043	*d*	T	A	A	A	A	A	A	A	A	T	T	T	T	T	Synonymous
27067	*d*	G	A	A	A	A	A	A	A	A	G	G	G	G	G	Synonymous
27070	*d*	T	C	C	C	C	C	C	C	C	T	T	T	T	T	Synonymous
27073	*d*	C	T	T	T	T	T	T	T	T	C	C	C	C	C	Synonymous
27223	*d*	T	A	A	A	A	A	A	A	A	T	T	T	T	T	Synonymous
27226	*d*	A	A	A	A	T	A	A	A	A	A	A	T/A	A	A	Non-synonymous
27238	*d*	G	A	A	A	G	A	A	A	A	G	G	G	G	G	Non-synonymous
27241	*d*	C	T	T	T	T	T	T	T	T	C	C	C	C	C	Synonymous
27311	*d*	CAA	-CAA	-CAA	-CAA	-CAA	-CAA	-CAA	-CAA	-CAA	CAA	CAA	CAA	CAA	CAA	Insertion
27394	*d*	C	C	A	A	C	C	C	C	C	C	A	C	C	C	Synonymous
27463	*d*	A	A	T	T	T	A	T	T	A	A	T	T	A	T	Synonymous
27514	*d*	T	C	C	C	C	T	C	C	C	T	C	C	T	C	Synonymous
27520	*d*	C	C	C	C	G	C	C	C	C	C	C	C	C	C	Non-synonymous
27532	*d*	G	A	A	A	A	G	A	A	A	G	A	A	G	A/G	Synonymous
27536	*d*	T	A	A	A	T	A	A	A	A	T	T	T	T	T	Synonymous
27538	*d*	C	T	C	C	C	Y	T	T	C	C	C	C	C	T	Non-synonymous
27560	*d*	C	C	T	T	C	C	C	C	C	C	T	T	C	C	Synonymous

Table S5 (continued)

		Genotype													
Position(BAC)	Gene	BAC	Scaff01516	CLS_BE	CLS_Spr	CLS_SprL	CLS_WT Fr	CLS_Fr	CLS_FrL	CLS_WT	PE75	SVS_BE	SVS_SprL	SVS_WTFr	SVS_FrL	Potencial Effects
27589	*d*	G	G	T	T	G	G	G	G	G	G	T	G/T	G	G	Synonymous
27594	*d*	C	C	C	C	A	C	C	C	C	C	C	A	C	C	Non-synonymous
27652	*d*	G	G	G	G	A	G	G	G	G	G	G	G	G	G	Non-synonymous
27736	*d*	A	A	A	A	M	A	A	A	A	A	A	A	A	A	Heterozygotic site
27776	*d*	C	C	C	Y	T	C	C	C	C	C	C	C	C	C	Non-synonymous
27838	*d*	G	G	G	G	A	G	G	G	G	G	G	G/A	G	G	Non-synonymous
27916	*d*	G	G	T	K	G	G	G	G	G	G	NA	G/T	G	G	Synonymous
27997	*d*	A	A	A	A	A	A	A	A	A	A	A	A	T	A	Synonymous
28058	I	A	G	G	G	G	G	G	G	G	A	A	A	A	A	Non-synonymous
28070	I	C	C	C	Y	T	C	C	C	C	C	C	C	C	C	Non-synonymous
28072	I	G	T	G	G	G	T	T	G	T	G	G	G	G	G	Non-synonymous
28153	I	A	C	C	C	C	C	C	C	C	A	A	A	A	A	Synonymous
28167	I	C	C	C	Y	T	C	C	C	C	C	C	C	C	C	Heterozygotic site
28179	I	T	G	G	G	G	G	G	G	G	T	T	T	T	T	Non-synonymous
28439	*w*1	C	C	C	C	C	C	C	C	C	C	C/A	C	C	C	Non-synonymous
28452	*w*1	A	A	A	A	A	A	A	A	A	A	A	A	G	A	Non-synonymous
28468	*w*1	A	G	A	A	G	G	G	G	A	A	A	A/G	A	G	Synonymous
28528	*w*1	C	C	T	T	C	C	C	C	C	C	T	C/T	C	C	Synonymous
28536	*w*1	C	C	C	C	C	C	C	C	C	C	C	C	C	G/C	Synonymous
28639	*w*1	C	C	C	C	C	C	M	A	C	C	C	C	C	A	Synonymous

Table S5 (continued)

		Genotype													
Position(BAC)	Gene	BAC	Scaff01516	CLS_BE	CLS_Spr	CLS_SprL	CLS_WT Fr	CLS_Fr	CLS_FrL	CLS_WT	PE75	SVS_BE	SVS_SprL	SVS_WTFr	SVS_FrL	Potencial Effects
28657	*w*1	T	T	T	T	T	T	T	T	T	T	T	T	C	T	Synonymous
28681	*w*1	C	C	C	C	C	C	Y	T	C	C	C	C	C	T	Synonymous
28732	*w*1	A	G	G	G	A	G	G	G	G	A	A	A	A	A	Synonymous
29503	*w*2	T	T	T	T	T	T	W	W	T	T	T	T	T	T	Synonymous
29504	*w*2	T	T	T	T	T	T	T	W	T	T	T	T	T	T	Synonymous
29508	*w*2	T	T	T	T	T	T	Y	T	T	T	T	T	T	T	Synonymous
29511	*w*2	T	T	T	T	T	T	T	W	T	T	T	T	T	T	Synonymous
29588	*w*2	A	A	A	A	A	R	R	R	A	A	A	A	A	A	Synonymous
29602	*w*2	A	A	A	A	A	A	M	A	A	A	A	A	A	A	Synonymous
29607	*w*2	T	T	T	T	T	T	Y	T	T	T	T	T	T	T	Synonymous
29615	*w*2	C	T	C	C	C	T	Y	T	T	C	C	C	T	T	Synonymous
29618	*w*2	T	C	T	T	T	C	T	C	C	T	T	T	C	C	Synonymous
29621	*w*2	T	T	T	T	T	T	T	T	Y	A	A	A	A	A	Synonymous
29653	*w*2	A	A	A	A	A	A	M	A	A	A	A	A	A	A	Synonymous
29711	*w*2	A	A	A	A	A	A	A	A	G	A	A	A	A	A	Synonymous
29714	*w*2	G	A	G	G	G	A	G	A	A	G	G	G	A	NA	Synonymous
29725	*w*2	T	T	T	T	T	T	Y	T	T	T	T	T	T	T	Synonymous
29726	*w*2	A	A	A	A	A	A	A	C	A	A	A	A	A	A	Synonymous
29729	*w*2	G	A	G	G	G	A	G	G	A	G	G	G	G	G	Non-synonymous
29762	*w*2	A	A	A	A	A	A	R	A	A	A	A	A	A	A	Synonymous

Table S5 (continued)

		Genotype														
Position(BAC)	Gene	BAC	Scaff01516	CLS_BE	CLS_Spr	CLS_SprL	CLS_WT Fr	CLS_Fr	CLS_FrL	CLS_WT	PE75	SVS_BE	SVS_SprL	SVS_WTFr	SVS_FrL	Potencial Effects
30444	*w3*	G	G	G	G	C	G	G	G	G	G	G	G	G	G	Synonymous
30483	*w3*	C	C	C	C	A	C	C	C	C	C	C	C/A	C	C	Synonymous
30492	*w3*	T	T	T	T	T	C	C	C	T	T	T	T	C	C	Synonymous
30580	*w3*	A	T	T	T	T	T	T	T	A	A	A	A	A	A	Non-synonymous
30585	*w3*	A	A	A	A	A	A	G	G	A	G	A	A	G	NA	Synonymous
30594	*w3*	G	G	G	G	A	G	G	G	G	G	G	G	G	G	Synonymous
30637	*w3*	A	C	C	C	C	C	C	C	A	A	A	A	A	A	Non-synonymous
30662	*w3*	T	T	C	Y	T	T	T	T	Y	T	T	T	T	T	Non-synonymous
30663	*w3*	G	A	G	G	G	A	G	G	G	G	G	G	G	G	Synonymous
30666	*w3*	A	T	T	T	A	A	A	A	T	A	A	A	A	A	Synonymous
30675	*w3*	G	G	A	M	G	G	G	G	G	G	G	G	G	G	Synonymous
30676	*w3*	C	C	G	S	C	C	C	C	C	C	C	C	C	C	Non-synonymous
30683	*w3*	T	T	C	Y	T	T	T	T	Y	T	T	T	T	T	Non-synonymous
30685	*w3*	G	G	C	S	G	G	G	G	G	G	G	G	G	G	Non-synonymous
30705	*w3*	T	T	C	Y	T	T	T	T	T	T	T	T	T	T	Synonymous
30716	*w3*	T	T	A	W	T	T	T	T	W	T	T	T	T	T	Non-synonymous
30721	*w3*	T	T	**Insertion[c]**	**Insertion[c]**	**Insertion[c]**	T	T	T	W	T	T	T	T	T	Insertion
30728	*w3*	T	T	G	G	G	T	G	G	K	T	T	T	T	T	Non-synonymous
30744	*w3*	G	G	C	C	G	G	G	G	NA	G	G	G	G	G	Synonymous
30746	*w3*	G	G	G	G	A	G	G	G	NA	G	G	G	G	G	Non-synonymous

Table S5 (continued)

		Genotype													
Position(BAC)	Gene	BAC	Scaff01516	CLS_BE	CLS_Spr	CLS_SprL	CLS_WT Fr	CLS_Fr	CLS_FrL	CLS_WT	PE75	SVS_BE	SVS_SprL	SVS_WTFr	SVS_FrL	Potencial Effects
30753	*w3*	T	T	C	T	T	T	T	T	NA	T	T	T	T	T	Synonymous
30755	*w3*	C	C	C	Y	T	C	C	C	NA	C	C	C	C	C	Non-synonymous
30757	*w3*	C	C	C	C	G	C	C	C	NA	C	C	C	C	C	Non-synonymous
30758	*w3*	C	A	C	C	C	A	C	C	NA	C	C	C	C	C	Non-synonymous
30760	*w3*	C	C	C	C	C	C	T	T	NA	C	C	C	C	C	Non-synonymous
30795	*w3*	A	A	G	R	A	A	A	A	NA	A	A	A	A	A	Synonymous
30810	*w3*	C	T	C	C	C	C	T	C	NA	C	C	C	C	C	Synonymous
30829	*w3*	G	G	G	G	A	G	G	G	NA	G	G	G	G	G	Non-synonymous
30831	*w3*	G	G	G	G	G	A	G	A	NA	G	G	G	G	G	Synonymous
58941	*w4*	T	T	NA	T	T	G	G	T	T	T	T	T	T	T	Non-synonymous
58956	*w4*	C	C	NA	C	C	T	Y	C	C	C	C	C	C	C	Non-synonymous
58968	*w4*	T	T	NA	T	T	A	W	T	T	T	A	T	A	T	Non-synonymous
58986	*w4*	C	A	NA	C	C	C	C	A	C	C	C	C	C	A	Non-synonymous
59001	*w4*	C	A	NA	C	C	C	C	A	C	C	C	C	C	A	Non-synonymous
59023	*w4*	GC	GC	NA	TT	TT	GC	GC	GC	GC	GC	GC	GC	GC	GC	Synonymous
59045	*w4*	G	A	NA	G	G	G	R	A	G	G	G	G	G	G	Non-synonymous
72855	*w5*	C	C	A	C	C	NA	NA	NA	C	C	C	C	C	C	Synonymous
72882	*w5*	C	C	T	C	C	NA	NA	NA	C	C	C	C	C	C	Synonymous
72924	*w5*	C	C	A	C	C	NA	NA	NA	C	C	C	C	C	C	Synonymous
72945	*w5*	C	C	G	C	C	NA	NA	NA	C	C	C	C	C	C	Synonymous

Table S5 (continued)

		Genotype													
Position(BAC)	Gene	BAC	Scaff01516	CLS_BE	CLS_Spr	CLS_SprL	CLS_WT Fr	CLS_Fr	CLS_FrL	CLS_WT	PE75	SVS_BE	SVS_SprL	SVS_WTFr	SVS_FrL	Potencial Effects
72966	*w5*	C	G	C	C	C	NA	NA	NA	C	C	C	C	C	C	Synonymous
72978	*w5*	G	A	A	A	A	NA	NA	NA	G	G	G	G	G	G	Synonymous
72987	*w5*	G	G	A	A	A	NA	NA	NA	G	G	G	G	G	G	Synonymous

Note: **Insertion**[a] – 35bp; **Insertion**[b] – 716bp and **Insertion**[c] – 15bp