Chorion Patterning: A Window into Gene Regulation and *Drosophila* Species' Relatedness

Matthew G. Niepielko,[†,1] Robert A. Marmion,[†,1] Kenneth Kim,[2] David Luor,[2] Chelsea Ray,[2] and Nir Yakoby*[,1,2]

[1]Center for Computational and Integrative Biology, Rutgers, The State University of NJ

[2]Biology Department, Rutgers, The State University of NJ

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: yakoby@camden.rutgers.edu.

Associate editor: Stuart Newfeld

Abstract

Changes in gene regulation are associated with the evolution of morphologies. However, the specific sequence information controlling gene expression is largely unknown and discovery is time and labor consuming. We use the intricate patterning of follicle cells to probe species' relatedness in the absence of sequence information. We focus on one of the major families of genes that pattern the *Drosophila* eggshell, the *Chorion protein* (*Cp*). Systematically screening for the spatiotemporal patterning of all nine *Cp* genes in three species (*Drosophila melanogaster*, *D. nebulosa*, and *D. willistoni*), we found that most genes are expressed dynamically during mid and late stages of oogenesis. Applying an annotation code, we transformed the data into binary matrices that capture the complexity of gene expression. Gene patterning is sufficient to predict species' relatedness, consistent with their phylogeny. Surprisingly, we found that expression domains of most genes are different among species, suggesting that *Cp* regulation is rapidly evolving. In addition, we found a morphological novelty along the dorsalmost side of the eggshell, the dorsal ridge. Our matrix analysis placed the dorsal ridge domain in a cluster of epidermal growth factor receptor associated domains, which was validated through genetic and chemical perturbations. Expression domains are regulated cooperatively or independently by signaling pathways, supporting that complex patterns are combinatorially assembled from simple domains.

Key words: **tissue patterning, oogenesis, eggshell, EGFR signaling.**

Article

Introduction

The amassing of genomic information for more than a decade suggests that species diversity is strongly associated with noncoding regions of the genome (Mann and Carroll 2002; Shubin et al. 2009). Numerous groups have begun analyzing individual regulatory sites to characterize how changes in cis-regulatory modules (CRMs) affect gene expression and consequently morphology. For example, the difference in the development of extra embryonic tissues, the amnion and serosa in *Anopheles gambiae* and a single amnioserosa in *Drosophila melanogaster*, is suggested to be associated with modifications in dorsal binding clusters in the short gastrulation locus (Goltsev et al. 2007). Another example is the dependency of trichome formation on the expression of *shavenbaby* (*svb*) in fly larvae (Frankel et al. 2011). In *D. melanogaster*, *svb* expression depends on the action of six CRMs; the lack of trichomes in *D. sechellia* and *D. ezoana* is associated with nucleotide changes within specific modules (Frankel et al. 2012). Thus, modifications in CRMs can lead to patterning and morphological differences.

The characterization of gene regulation requires a considerable amount of work, especially when numerous CRMs derive the full pattern of a gene (Stathopoulos and Levine 2005; Frankel et al. 2012). The challenge is even greater with unsequenced species. Because changes in gene patterning reflect modifications in gene regulation, we used *Drosophila* eggshell patterning to study gene regulation and species' relatedness. Eggshell morphogenesis is preceded by an extensive tissue patterning that mediates the formation of different functional chorionic structures (Dobens and Raftery 1998; Waring 2000; Berg 2005; Yakoby, Bristow, et al. 2008). The eggshell is an intricate 3D structure that is formed by a monolayer of follicular epithelium engulfing the developing oocyte (Hinton and Service 1969; Spradling 1993; Horne-Badovinac and Bilder 2005). The eggshell protects the developing embryo from the environment and at the same time allows gas exchange through specialized structures, including a posterior porous aeropyle and anterior tubelike dorsal appendages (Hinton and Service 1969; Margaritis et al. 1980; Dorman et al. 2004).

Here, we analyzed the dynamics and diversities of *Chorion protein* (*Cp*) expression, a family of nine genes patterning the *Drosophila* eggshell (reviewed in Waring [2000] and Cavaliere et al. [2008]). We developed an annotation system to capture the complexity of follicle cell patterning in three *Drosophila* species. This system transforms 2D images into binary matrices of simple expression domains. We show that expression patterns are sufficient to determine species evolutionary relationships. Surprisingly, we found that expression domains of most genes are highly diverse among species, suggesting that

Cp regulation is rapidly evolving. By combining image analysis and experimental validation, we linked the expression of genes in specific domains to their regulation by bone morphogenetic protein (BMP) and epidermal growth factor receptor (EGFR) signaling; two major pathways underlying eggshell patterning (Berg 2005). Furthermore, we experimentally validated our computational prediction that the dorsal ridge (DR), a lumen-like structure along the dorsal side of eggshells from *D. willistoni* and *D. nebulosa*, is regulated by EGFR signaling. Our annotation system provides an alternative way to examine gene regulation in the follicle cells. Furthermore, we show that analysis of tissue patterning is a powerful approach to determine relatedness among species, especially when DNA sequences are unavailable. We propose that this annotation system can be extended to other tissues that have recognizable pattering domains with clear boundaries.

Results

Chorion Patterning Is Dynamic and Reflects Eggshell Morphologies

The DR is a lumen-like structure along the dorsalmost side of eggshells from *D. willistoni* and *D. nebulosa*; two species in the subgenus *Sophophora*. This structure is absent in *D. melanogaster* eggshells (fig. 1*A*–*C*). The DR was reported and characterized structurally only in eggshells from Hawaiian *Drosophila* species (Margaritis et al. 1983). We assume that different structures reflect changes in follicle cell patterning among species. Thus, we selected one family of genes that participate in eggshell formation, the *Cp* genes (Waring 2000; Fakhouri et al. 2006). This family includes nine genes: *Cp7fa*, *Cp7fb*, *Cp7fc*, *Cp15*, *Cp16*, *Cp18*, *Cp19*, *Cp36*, and *Cp38* (Spradling 1981; Griffin-Shea et al. 1982; Parks et al. 1986). We focused on four developmental stages of oogenesis: S10A, S10B, S11, and S12 (Spradling 1993) across three *Drosophila* species. In *D. melanogaster*, the expression patterns of seven of the nine *Cp* genes (excluding *Cp7fa* and *Cp19*), were published with different levels of spatial resolution (Griffin-Shea et al. 1982; Parks et al. 1986; Yakoby, Bristow, et al. 2008). Our results are consistent with the known patterns in *D. melanogaster* and include a complete collection of all genes across three species (supplementary fig. S1*A*-I, Supplementary Material online).

We found that *Cp* genes are expressed dynamically and in different domains of the follicle cells of *Drosophila* species (fig. 1*D*–*L* and supplementary fig. S1*A*-I, Supplementary Material online). Interestingly, in *D. willistoni* and *D. nebulosa*, we observe gene patterning that highly correlates with the DR morphologies, reflecting the shorter DR in *D. willistoni* than *D. nebulosa* (fig. 1*B* and *C*). Specifically, *Cp* genes that are patterned in the future DR domain of *D. willistoni* including *Cp7fa*, *Cp16*, and *Cp19* are shorter in length when compared with the corresponding patterns in *D. nebulosa* (fig. 1*K* and *L* and supplementary fig. S1*A*, *E*, and *G*, Supplementary Material online). The same correlation was observed in the repression domain of *Cp7fc* in the future DR domain (fig. 1*F* and supplementary fig. S1*C*, Supplementary Material online).

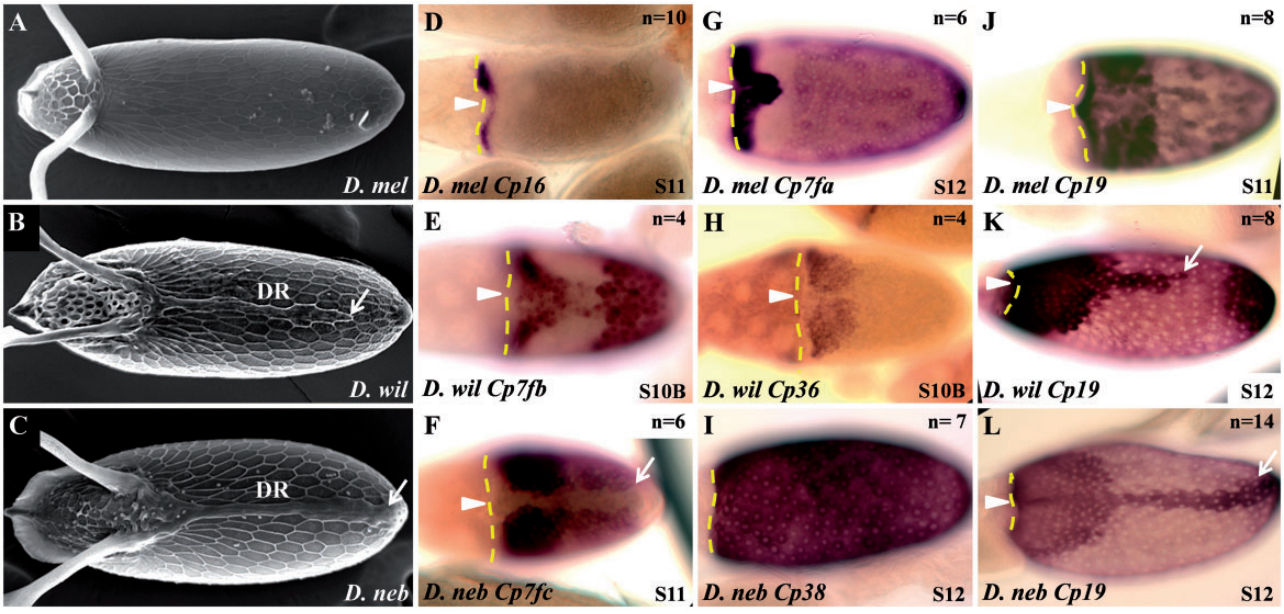


Fig. 1. *Drosophila* eggshell morphologies and chorion patterning are diverse. (*A*–*C*) Scanning electron images of the wild-type eggshells of *Drosophila melanogaster* (*D. mel*), *D. willistoni* (*D. wil*), and *D. nebulosa* (*D. neb*). The eggshells of *D. wil* and *D. neb* have an additional structure called the DR that begins from the bases of dorsal appendages and extends toward the posterior along the dorsalmost side of the eggshell (*B*, *C*). The DR varies in length between species; the DR of *D. wil* does not reach the posterior end of the eggshell (*B*, white arrow), while the DR in *D. neb* reaches the most posterior end of the eggshell (*C*, white arrow). (*D*–*L*) Examples of different patterns of *Cp* genes at different developmental times in the follicle cells of *D. mel*, *D. wil*, and *D. neb*. (*F*, *K*, *L*) Examples of *Cp* gene patterns that reflect DR morphology (white arrow points to the most posterior end of the future DR domain). In all images, broken yellow line denotes anterior region of the follicle cells, white arrowhead denotes dorsal midline, images are dorsal views, and anterior is to the left. The "n" represents the number of similar images to the one that is represented in this figure.

In *D. melanogaster*, no similar patterns were found (supplementary fig. S1A-I, Supplementary Material online, and Yakoby, Bristow, et al. 2008).

Transformation of 2D Patterning Images into Digital Information and Species' Relatedness

The entire patterning collection includes 108 images (nine *Cp* genes over four developmental stages across three species) (supplementary fig. S1A-I, Supplementary Material online). We were particularly interested to understand the complexity of gene patterning over developmental and evolutionary axes. Previously, a code that is based on six simple shapes of patterning domains that were combinatorially assembled into more complex patterns using Boolean operations have successfully described the entire 2D image collection of eggshell gene patterning in *D. melanogaster* (Yakoby, Bristow, et al. 2008). This code focuses on the dorsal anterior domain and thus excludes the posterior domain and the new DR domain. Also, this code uses a minimal selection of shapes that are not exclusive. Furthermore, identical patterns can be annotated in various correct ways, which would make the computational comparisons between annotations impractical.

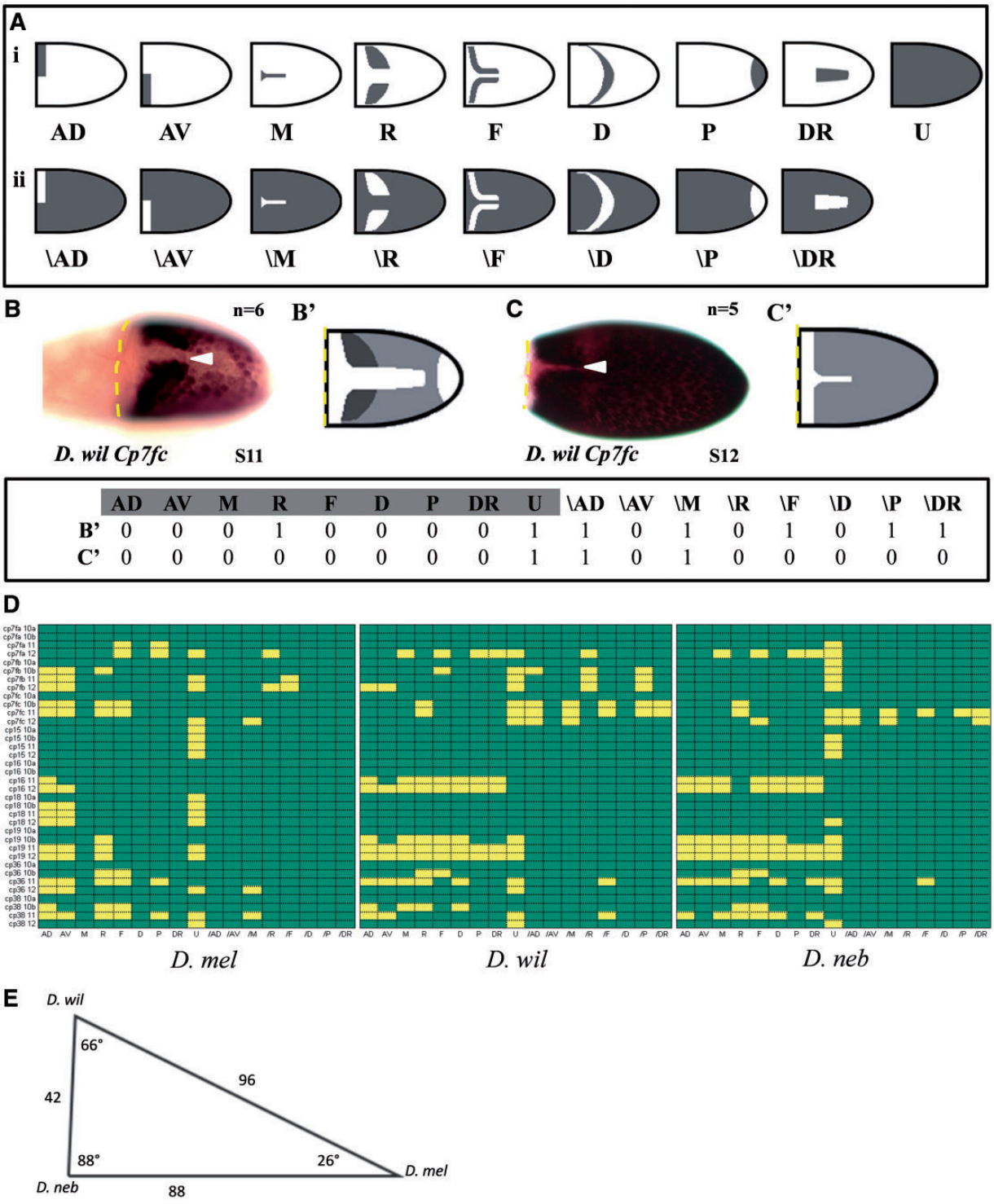
To analyze follicle cell patterning systematically, we used a similar concept but modified the code to generate exclusive domains that cover the entire follicle cells (fig. 2*Ai* and *ii*). Specifically, in the original code, the anterior (A) domain reflects genes that are regulated by BMP signaling, and it spans the cells overlaying the border between the oocyte and the nurse cells (Yakoby, Bristow, et al. 2008). Here, the A domain is split into the anterior–dorsal (AD) and the anterior–ventral (AV) domains. The midline (M) represents the high levels of EGFR activation, and in the original code it includes the AD domain (Yakoby, Bristow, et al. 2008). Now, the M is separate from the AD. The roof (R) and floor (F) represent two domains that build the top and bottom, respectively, of the future dorsal appendages (Ruohola-Baker et al. 1993; Deng and Bownes 1997; Ward and Berg 2005). In the original code, the dorsal (D) domain includes AD, M, F, and R. Now, the D domain represents an arch shaped at the intermediate/low levels of EGFR signaling. The posterior (P) and DR are new shapes that represent the future aeropyle and DR domains, respectively. Uniform (U) represents an expression throughout the follicle cells. In addition, we use U to describe the absence or presence of domains in patterns with uniform expression. In this way, individual domains can be added on top of, or subtracted from a uniform expression to capture, all patterning domains (fig. 2*Aii*).

In the following two examples, we demonstrate the use of the annotation system (fig. 2*B*, *C*, and *D*). Each pattern is transformed into a binary matrix that scores each domain with 0 or 1 for the absence or presence of expression, respectively. At stage 11, the pattern of *Cp7fc* in *D. willistoni* has an R domain and a U domain that lacks AD, M, F, P, and DR domains (fig. 2*B* and *B′*). At stage 12, *Cp7fc* is expressed in the U domain that lacks the AD and M domains (fig. 2*C* and *C′*). This annotation system enables computational analyses of the entire collection of patterns within and between species. Representing the gene per stage in the rows and the expression domains in the columns, we generated a *Cp* patterning profile for each species (fig. 2*D*). Our annotation system captures patterning dynamics in a matrix form, and it generates a spatiotemporal "fingerprint" profile for each species given a set of genes. These matrices can now be compared and analyzed.

To explore whether tissue patterning can be used as a tool to evaluate the relatedness among species, we compared the patterning matrices to predict the evolutionary distance between the three species. Distance between a pair of species was calculated as the absolute value of the difference between two matrices. Data were depicted as a triangle for display purposes (fig. 2*E*). Importantly, we found that the two DR species are most related, with *D. nebulosa* closer to *D. melanogaster* than *D. willistoni*. Species' relatedness remained the same even when the DR domains were set to zero in *D. willistoni* and *D. nebulosa* (not shown), suggesting that gene patterning is fundamentally different among species and not solely due to the presence of a DR domain. These results are in agreement with our sequence-based phylogenetics of these species (Niepielko et al. 2011). We suggest that gene patterning provides an additional source to evaluate distance among species. However, patterning experiments in additional species are required to further test this approach.

The Expression Landscape of DR Species Is Conserved

Except for the DR, the eggshells of the three species are similar, that is, two dorsal appendages, operculum, and main body cell imprints (fig. 1*A*–*C*). Therefore, we expected a high conservation of gene expression among species within specific domains. To test our expectation, individual domains were pairwise compared for shared expression of all genes at all developmental time points between species (fig. 3). The color represents the shared proportion of two species in a particular domain, and the numbers are the times each gene (up to four developmental stages of each of the nine genes) appeared in both species in the same domain, a theoretical maximum of 36. Using an arbitrary cutoff of 65%, we found that only genes that are expressed in the R domain are conserved among the three species (fig. 3*A*–*C*). The anterior domains (AV, AD) scored low when *D. melanogaster* was compared with the other two species (fig. 3*A* and *B*). However, these domains express similar genes when *D. willistoni* is compared with *D. nebulosa* (fig. 3*C*). In fact, we found that other domains, including the M, P, and DR domains, express similar genes in the two species with the DR (fig. 3*C*). To help elucidate the nature of domain sharing among genes in the species with a DR, we used domain cross comparisons this time, comparing the DR species together instead of comparing with each other (fig. 3*D*). We found that in *D. willistoni* and *D. nebulosa* many anterior and dorsal anterior domains, including the M, D, DR, and AD/AV, share many genes (fig. 3*D*), which is in contrast to the same domains in *D. melanogaster* with the exception of the AD/AV domain (fig. 3*E*). These differences may indicate



	AD	AV	M	R	F	D	P	DR	U	\AD	\AV	\M	\R	\F	\D	\P	\DR
B'	0	0	0	1	0	0	0	0	1	1	0	1	0	1	0	1	1
C'	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0

Fig. 2. Patterning domains, matrix transformation, and species' relatedness. (*A*) Cartoons depicting simple domains observed in follicle cell patterning. (*Ai*) Domains representing expression. Lateral views: anterior dorsal (AD) and anterior ventral (AV); Dorsal views: midline (M), roof (R), floor (F), dorsal (D), posterior (P), DR, and uniform (U). (*Aii*) Domains represent no expression in a uniform background and transformation of tissue patterning into binary matrices. Each domain in the pattern is given either 1 when present or a 0 for unrepresented. (*B*, *B*′) The pattern of *Cp7fc* in *D. wil* at S11 is constructed using R, U, \AD, \M, \F, \P, and \DR. (*C*, *C*′) At S12 *Cp7fc* is characterized as a combination of U, \AD, and \M. (*D*) Collection of binary matrices for all genes, at all developmental stages, for each species; the rows are the gene/stage, and each column is a specific patterning domain. (*E*) The sum of the absolute value of the difference between pairs of species is depicted as the line length of a triangle. *Drosophila neb* and *D. wil* are most related and *D. wil* is least related to *D. mel*. Arrowhead denotes the dorsal midline and anterior is to the left. Broken yellow line (*B*, *C*) denotes the anterior region of the follicle cells.

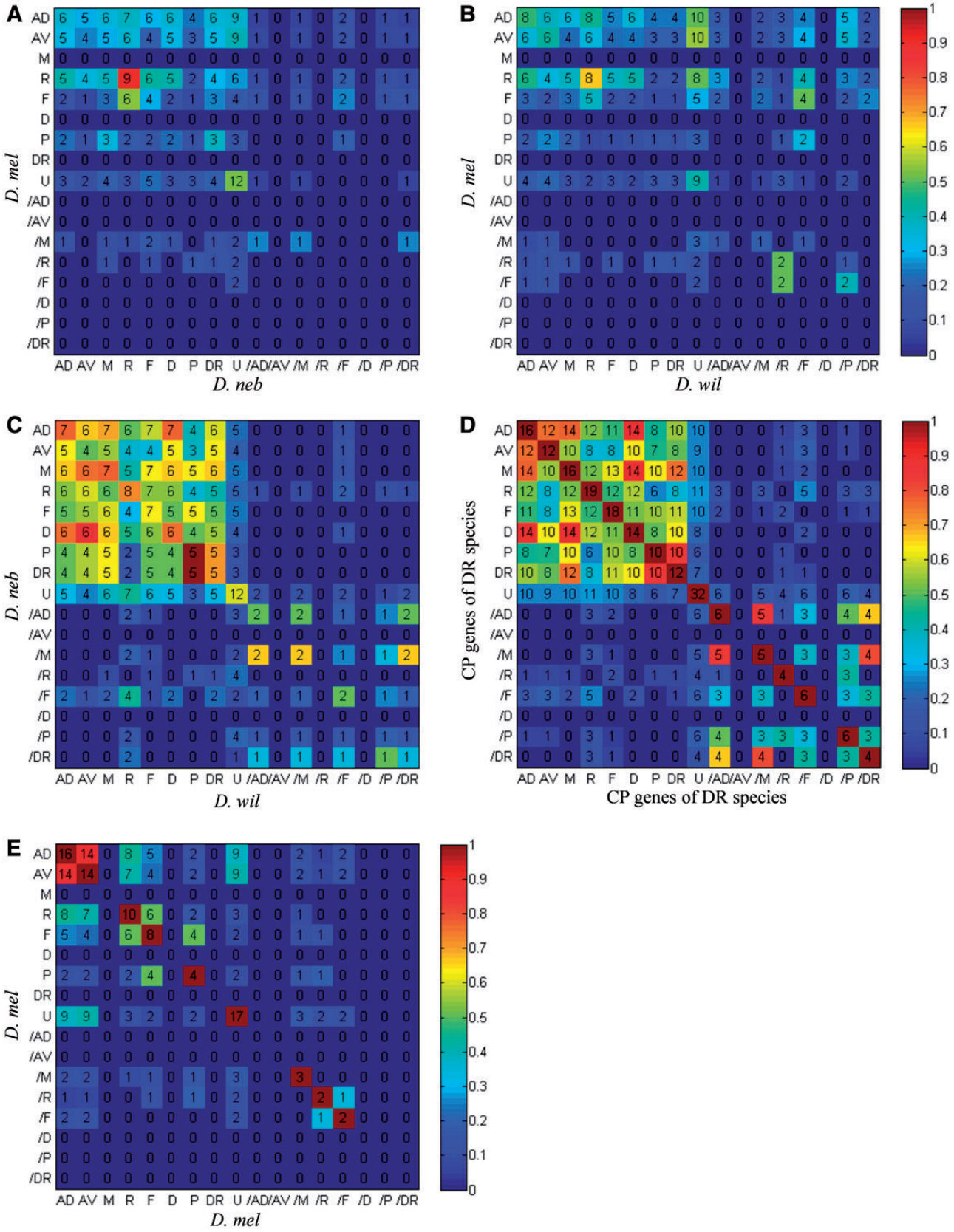


FIG. 3. The expression landscape of genes is diverse. (*A*–*E*) Pairwise comparisons between and within species. (*A*) Pair wise comparison between *D. melanogaster* and *D. nebulosa* shows high conservation of genes patterned in the roof (R) domain. (*B*) Pairwise comparison between *D. melanogaster* and *D. willistoni* shows a similar conservation of the R domain. (*C*) Pairwise comparison between *D. nebulosa* to *D. willistoni* shows high usage of the same genes in different dorsal and anterior domains at a given developmental stage. (*D*) *Cp* proteins in ridge species show high cross-domain correlations and are largely an accumulation of dorsal domains (upper left panels). Of particular interest, the *Cp* proteins show a correlation of DR, midline, dorsal, and posterior. (*E*) Within *D. melanogaster*, domains do not share genes with the exception of AD and AV domains. Numbers represent occurrences of the overlap. Percent calculated as the number over the highest occurrence of either domain being utilized.

on fundamental changes in the regulation of *Cp* genes during *Drosophila* evolution.

The DR Domain Is Associated with EGFR Signaling-Regulated Domains

Given the range of patterns and diverse use of genes across domains, we next investigated the relationship among expression domains. Initially, individual domains were summed by stage and plotted as simple bar graphs (fig. 4*A*–*C*). The difference between *D. melanogaster* and DR species was striking. Particularly, DR species are less enriched for anterior patterns (AV, AD). Also, in *D. melanogaster*, none of the patterns have unique expression in the midline (M) or dorsal (D) domains. Next, to determine domain relatedness, we used hierarchical clustering of the average expression between the three species (fig. 4*D*). We assume that domains that cluster together may be regulated by a similar input. As expected from domains that are regulated by the anteriorly located BMP signaling (Twombly et al. 1996), the AD and AV domains cluster with a bootstrap value of 98% (fig. 4*D*; the complete bootstrap analysis can be found in supplementary fig. S2*A*, Supplementary Material online). Of note, these domains also cluster when we looked at species individually (supplementary fig. S2*B*–*G*, Supplementary Material online).

The R domain appears as a separate domain (fig. 4*D*), which perhaps reflects the interplay between the EGFR and BMP signaling pathways controlling the R domain (Deng and Bownes 1997; Ward and Berg 2005; Yakoby, Lembong, et al. 2008). Interestingly, the R domain clusters with the F domain in *D. melanogaster* at stage 11 but still appears as a separate domain in all species during other developmental stages including stage 11 in *D. willistoni* and *D. nebulosa* (supplmentary fig. S2*B-G*, Supplementary Material online); we address this observation in the discussion. We found the U pattern to be least related to the other domains (fig. 4*D* and supplementary fig. S2, Supplementary Material online), suggesting that it is controlled by signals other than EGFR and BMP pathways.

The P, DR, M, and D domains cluster together with a bootstrap value of 85%. This is particularly interesting because the P, M, and D domains are regulated by EGFR signaling (Queenan et al. 1997; Yakoby, Bristow, et al. 2008), suggesting that the DR domain is regulated in a similar manner. To test the relatedness of DR to EGFR signaling, we looked for a gene that is expressed in M domain; an area with high levels of EGFR activation. Because none of the *Cp* genes is expressed in an M unique pattern in *D. melanogaster* (fig. 4*A*), we selected fasciclin III (FasIII), a gene that is expressed in the midline of *D. melanogaster* (Ward and Berg 2005; Shravage et al. 2007) and is a known target of EGFR signaling in the embryo (Dong et al. 1999). In *D. melanogaster*, FASIII is expressed in the AD, M, and F domains (fig. 4*E*). In *D. willistoni* and *D. nebulosa*, FASIII is also expressed in the future DR domain (fig. 4*F* and *G*). Interestingly, the patterns of FASIII in the future DR domains reflect the respective length of the DR in each species (fig. 1*B* and *C*). In the follicle cells, FASIII was shown to be regulated by BMP signaling (Shravage et al. 2007). Here, activation of EGFR in the posterior domain was sufficient to derive ectopic FASIII expression in this domain (fig. 4*H*, *H′*, *I*, and *I′*). These results support the clustering of the DR with other EGFR-regulated domains.

Complex Patterns Are Combinatorially Assembled from Simple Domains

Eggshell patterning is controlled by numerous signaling pathways including EGFR and BMP (Neuman-Silberberg and Schupbach 1994; Twombly et al. 1996; Deng and Bownes 1997; Dobens and Raftery 1998; Peri and Roth 2000; Marmion et al. 2013). To determine which domains are controlled by EGFR signaling, using the drug colchicine, we disrupted EGFR signaling by mislocalizing the oocyte nucleus (Peri and Roth 2000). In *D. melanogaster*, colchicine-affected eggshells have disrupted dorsal structures including the dorsal appendages and operculum (fig. 5*A*). In species with a DR, colchicine-affected eggshells lack the DR and have disrupted dorsal appendages (fig. 5*B* and *C*). Interestingly, in some eggshells of *D. willistoni* and *D. nebulosa*, the dorsal appendages could still be seen, suggesting that the DR is more sensitive to changes in the levels of EGFR than the dorsal appendages.

It was previously shown that different domains are regulated by EGFR and BMP pathways independently and cooperatively (Deng and Bownes 1997; Shravage et al. 2007; Yakoby, Bristow, et al. 2008; Yakoby, Lembong, et al. 2008; Lembong et al. 2009). In colchicine-treated egg chambers, all patterning domains except for the anterior, uniform, and posterior are disrupted (wild-type patterns in fig. 5*D*–*F* compared with colchicine treated flies fig. 5*G*–*I* and supplementary fig. S3*A-C*, Supplementary Material online). This is not surprising because the A and U domains are not regulated by EGFR signaling. The P domain is regulated by EGFR signalling; however, EGFR activation in this domain occurs before nucleus mislocalization. These results are consistent with the patterning changes of these genes when EGFR was activated or repressed uniformly throughout the follicle cells (supplementary fig. S3*D-F*, Supplementary Material online). The anterior domain is regulated by BMP signaling (Twombly et al. 1996; Yakoby, Bristow, et al. 2008), and thus, we were able to disrupt this domain by overexpressing the BMP ligand *decapentaplegic* (*dpp*) throughout the follicle cells (fig. 5*J*–*L* and Supplementary fig. S3*D-F*, Supplementary Material online). In these cases, the anterior domain now expands into a large dorsal dome-shaped domain that derives the formation of a large operculum (Twombly et al. 1996; Yakoby, Bristow, et al. 2008; Marmion et al. 2013). These results support the idea that complex patterns are combinatorially assembled and that DR patterning is regulated by EGFR activation.

Discussion

Here, we introduced a new approach to analyze the dynamics and diversities of 2D images that pattern the follicular epithelium. We focused on the family of *Cp* genes that have a highly conserved protein sequences and structures across fly species (Waring 2000). Eight of the nine proteins were isolated from *D. melanogaster* eggshells (Fakhouri et al. 2006). Analyzing *Cp*

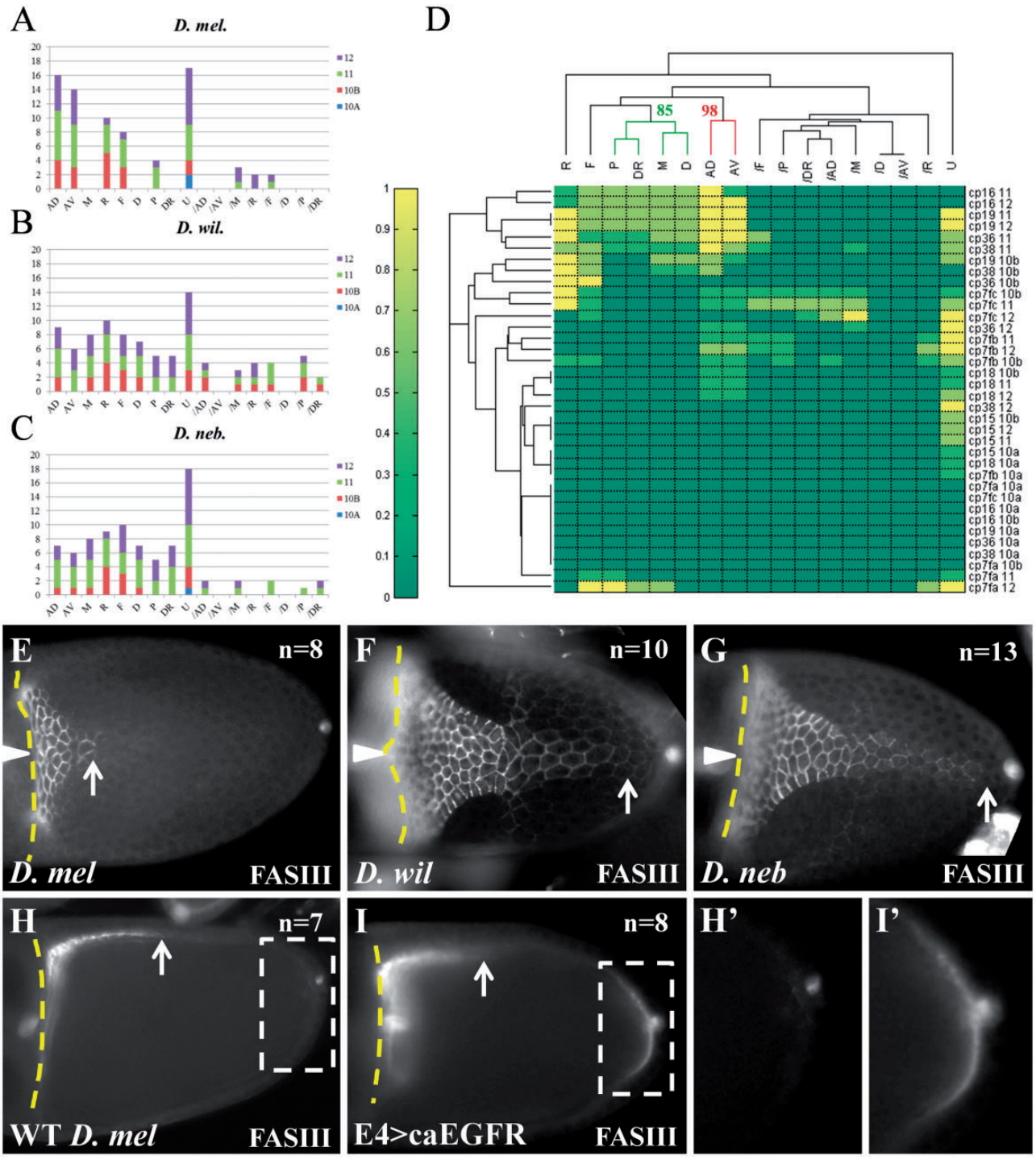


Fig. 4. Association of the DR domain with EGFR signaling. (*A*–*C*) Cumulative counts of domains used by *Cp* genes for (*A*) *D. melanogaster*, (*B*) *D. willistoni* and (*C*) *D. nebulosa*. (*D*) Clustergram of the average expression of all three species. Domains are clustered on the top and gene/stage on the left. (*E*–*G*) Wild-type expression of FASII) at stage 10B egg chambers in *D. melanogaster* (*D. mel*), *D. willistoni* (*D. wil*), and *D. nebulosa* (*D. neb*). (*E*) FASIII in *D. mel* is in the dorsal–anterior, midline, and floor domains. (*F*) In *D. wil*, FASIII is expressed in the dorsal–anterior, midline, floor, and DR domains. (*G*) In *D. neb*, FASIII is expressed in the dorsal–anterior, midline, floor, and DR domains. (*H*, *H*′) Sagittal section of FASIII in *D. mel*. (*I*, *I*′) Posterior activation of EGFR with E4 > caEGFR derives ectopic FASIII in the posterior domain. *H*′ and *I*′ are insets that are marked by a white broken line in *H* and *I*, respectively. (*E*–*G*) Dorsal views. Broken yellow line denotes the most anterior border of the FCs, white arrow points to the most posterior location of FASIII expression, and white arrowhead defines dorsal midline. Of note, posterior polar cells also express FASIII.

genes' patterning across fly species with different eggshell morphologies allowed us to address fundamental questions regarding the relationship between patterning domains along the developmental and evolutionary axes.

Two main signaling pathways pattern the *Drosophila* eggshell, EGFR and BMP (Berg 2005). These pathways act independently and cooperatively to pattern different domains of the eggshell (Yakoby, Lembong, et al. 2008). The early

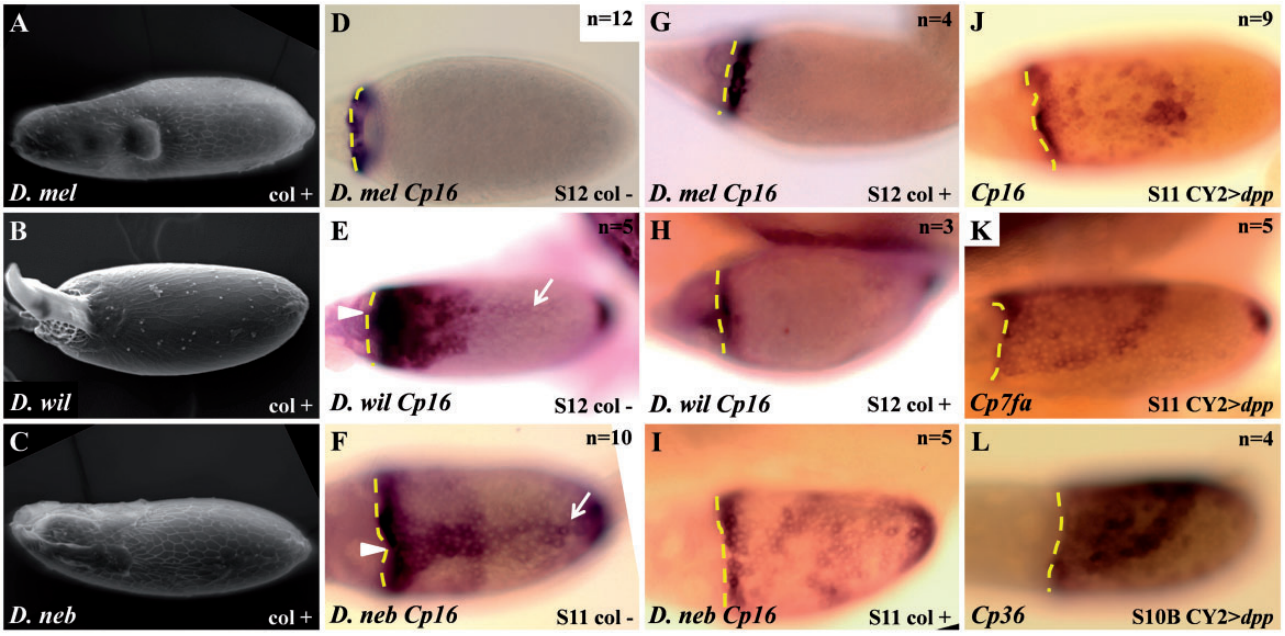


Fig. 5. Colchicine disrupts eggshell morphologies and numerous patterning domains. (*A*–*C*) Eggs laid by colchicine-treated flies have disrupted eggshell morphologies. (*A*) The eggshells of *D. melanogaster* lack dorsal appendages. (*B*) *D. willistoni* eggshells have a fused single dorsal appendage and no DR. (*C*) *D. nebulosa* eggshells have disrupted dorsal appendages and no DR. (*D-F*) The wild-type patterns of *Cp16* in *D. melanogaster*, *D. willistoni*, and *D. nebulosa*. (*D*) *Cp16* is restricted to the anterior domain in *D. melanogaster*. (*E*) In *D. willistoni*, *Cp16* is patterned in the anterior, midline, roof, floor, dorsal, DR, and posterior domains. (*F*) Patterning of *Cp16* in *D. nebulosa* includes the anterior, midline, floor, dorsal, DR, and posterior domains. (*G*–*I*) *Cp16* patterns in colchicine-treated flies. In all three species, the anterior and posterior patterning domains of *Cp16* are unaffected. In *D. willistoni* and *D. nebulosa*, cochicine treatments affected *Cp16* expression in all other domains including midline, roof, floor, dorsal, and DR (*H*, *I*). (*J*–*L*) Overexpression of *dpp* disrupts *Cp16* (*J*), *Cp7fa* (*K*), and *Cp36* (*L*) dorsal anterior patterns. Broken yellow line denotes the anterior border of the follicle cells, white arrowhead denotes dorsal midline, and white arrow points to most posterior domain of the future DR domain.

activation pattern of both pathways is highly conserved across multiple species (Kagesawa et al. 2008; Niepielko et al. 2011; Niepielko et al. 2012). By contrast, the late activation patterns are different. In particular, the late pattern of EGFR activation reflects the number of dorsal appendages (Kagesawa et al. 2008), whereas the late pattern of BMP signaling is highly associated with the species' phylogeny (Niepielko et al. 2012).

Patterning Domains Are Linked to the Signaling Inputs

To determine how expression domains associate in different species, we altered a previously developed code to annotate gene patterning of *D. melanogaster* eggshell (Yakoby, Bristow, et al. 2008). The new code has exclusive domains, which allows for the generation of binary matrices that can be analyzed for patterning differences among species in an unbiased manner. The R and F domains derive the top and bottom of the future dorsal appendages, respectively (Ruohola-Baker et al. 1993; Deng and Bownes 1997; Ward and Berg 2005; Osterfield et al. 2013). We expected similar *Cp* genes to associate with the two domains in the three species. Surprisingly, during stages 11 and 12, the two domains cluster together only at S11 in *D. melanogaster* (supplementary fig. S2, Supplementary Material online). We propose that differences in BMP signaling may dissociate the two domains in *D. willistoni* and *D. nebulosa*. Specifically, the dynamics of EGFR activation in *D. melanogaster* and *D. willistoni* are similar (Kagesawa et al. 2008); however, the patterns of BMP signaling are different (Niepielko et al. 2011). In *D. melanogaster*, the pattern of BMP signaling at stage 11 overlaps the R domain (Yakoby, Lembong, et al. 2008), whereas in *D. willistoni* and *D. nebulosa*, in addition to overlapping the R domain, BMP also signals in the F domain (Niepielko et al. 2011). Because stage 11 is relatively short (30 min [Spradling 1993]), it is possible that the dissociation of the R and F domains at S12 in *D. melanogaster* is due to the signaling of BMP in the R domain with a short temporal delay (Yakoby, Lembong, et al. 2008; Lembong et al. 2009).

Given that different domains reflect inputs from different pathways, it was interesting to find that the M, P, D domains cluster with the DR domain (fig. 4*D*). These domains are associated with different levels of EGFR activation (Yakoby, Bristow, et al. 2008), suggesting that the DR is regulated by EGFR signaling. Interestingly, unlike the dorsal midline restricted pattern of EGFR activation in *D. melanogaster* (Peri and Roth 2000) and *D. willistoni*, EGFR is also activated in the future DR domain (Kagesawa et al. 2008). Furthermore, mislocalization of the EGFR ligand Gurken (Grk) disrupts dorsal appendages in *D. melanogaster* (Peri and Roth 2000), and the same treatment eliminates the DR (fig. 5). In the future, it will be important to determine whether changes in the transition of the oocyte nucleus and its associated ligand Grk

(Neuman-Silberberg and Schupbach 1994; Sapir et al. 1998; Van Buskirk and Schupbach 1999) underlie the formation of a DR.

The clustering of EGFR-regulated domains is further supported by follicle cell patterning. The midline marker FASIII in *D. melanogaster* is also expressed in the future DR domains of *D. willistoni* and *D. nebulosa* (fig. 4*E*–*G*). Furthermore, the length of FASIII expression reflects the dorsal ride morphology, which is the case for *Cp* genes that pattern the DR domain (figs. 1 and 4*E*-*G*). The expression of FASIII is associated with BMP signaling in the follicle cells (Shravage et al. 2007); however BMP signaling is absent from the DR domain (Niepielko et al. 2011). Ectopic activation of EGFR in the posterior follicle cells induces FASIII expression in this domain (fig. 4*H*–*I*). We propose that *Fas*III is also regulated by EGFR signaling in the follicle cells, which is consistent with its regulation in the *Drosophila* embryo (Dong et al. 1999).

Patterning of *Cp* Genes Is Temporally Conserved but Spatially Diverse across Species

The *Cp* genes are clustered in two locations in the fly genome. The *Cp7fa*, *Cp7fb*, *Cp7fc*, *Cp 36*, and *Cp38* are on the X chromosome and *Cp15*, *Cp16*, *Cp18*, and *Cp19* are on the third. It was previously shown that genes on the X chromosome are expressed at earlier stages, whereas genes on the third chromosome are expressed later during chorion formation (Parks et al. 1986; Tolias et al. 1993). Looking at our complete data set, we found that in *D. melanogaster* most *Cp* genes are already expressed at stage 10B, and by stage 11 all of them are expressed (fig. 2*D*). A similar patterning progression was found in the other two species. Our analysis is further supported by a recent work out of the Spradling Lab, where they used microarrays to analyze the transcriptoms of stage-specific egg chambers. They showed that the expression of most *Cp* genes begins at stage 10B in *D. melanogaster* (Tootle et al. 2011). We conclude that the expression of the *Cp* gene family is temporally conserved. These results are consistent with the analysis of multiple gene families in *D. melanogaster* embryo patterning (Konikoff et al. 2012).

With only one exception (discussed later), we found that likelihood of a *Cp* gene expressed in the same domain and at the same developmental stage across the three species is low (fig. 3*A*–*C*). This suggests that the role of these proteins changed over eggshells' evolution. This is interesting because the sequences and structures of these proteins are conserved across fly species (Waring 2000). The finding that genes are not spatially conserved has one outstanding exception across the three species, the R domain. This domain shares more than 65% of the *Cp* genes across species (fig. 3*A*–*C*). The R domain is regulated by BMP and EGFR signaling (Deng and Bownes 1997; Peri and Roth 2000; Yakoby, Lembong, et al. 2008; Fuchs et al. 2012). All three species have similar patterns of late EGFR and BMP activation on the R domain (Kagesawa et al. 2008; Niepielko et al. 2011); thus, it is possible that the R domain maintained a similar gene regulation mechanism across species. In addition, the porous texture of the dorsal appendage is used for gas exchange; thus it is possible that the R domain is under an evolutionary pressure to maintain the same Cp proteins in this domain. The fly *Ceratitis capitata* lacks an R domain, and its eggshell lacks dorsal appendages (Vreede et al. 2013). Although the structure of Cp proteins between *D. melanogaster* and *C. capitata* is highly conserved (Waring 2000), it will be interesting to determine how the lack of an R domain affects *Cp* gene patterning.

Patterns Are Combinatorially Assembled and Reflect Species' Relatedness

Gene patterning reflects different inputs that converge on the regulatory region of genes. Our genetic and chemical perturbations could differentially disrupt patterning domains. For example, perturbations in EGFR signaling disrupted most domains except for the anterior domain, which was disrupted by perturbations in BMP signaling (supplementary fig. S3, Supplementary Material online). Interestingly, a short fragment of regulatory DNA (84 bp) from the *Cp36* gene was able to recapitulate the full pattern of the gene (Tolias et al. 1993). By examining the two halves of this fragment, they successfully separated the anterior and posterior expression domains of the gene. These results further support our previous analysis of multiple gene patterns in *D. melanogaster*, which suggested that gene patterns are assembled combinatorially by inputs from different pathways (Yakoby, Bristow, et al. 2008).

Tissue patterning contains sufficient information to predict relatedness among species. This method has a similar sensitivity as the traditional utilization of sequence information. Not surprisingly, we were able to determine that *D. nebulosa* and *D. willistoni* are most related among the three species, as both belong to the *willistoni* subgenus. Furthermore, our analysis has the power to determine that *D. nebulosa* is more related to *D. melanogaster* than *D. willistoni* (fig. 2*E*). This result is in agreement with our previous finding that follicle cells' patterning by the type I BMP receptor Thickveins is sufficient to cluster species to their phylogenetic groups (Niepielko et al. 2011). Our method provides an opportunity to analyze other simple epithelial tissues including imaginal discs across species even when sequencing information is not currently available.

Materials and Methods

Flies: Genetic and Chemical Manipulations

The following *Drosophila* species were used in this study: *D. melanogaster* (wild-type OreR), *D. nebulosa* (a gift from D. Stern), *D. willistoni* (The San Diego Stock Center). Additional fly lines used: CY2-Gal4, E4-Gal4, and UAS-caEGFR (Queenan et al. 1997), USA-*dpp* and UAS-dnEGFR (Peri and Roth 2000). Overactivation of EGFR signaling in the posterior follicle cells was achieved by driving UAS-caEGFR with E4-Gal4. Uniform overexpression of BMP and EGFR signaling was completed using CY2-Gal4 to drive USA-*dpp* and UAS-caEGFR, respectively. Uniform reduction in EGFR signaling was done using CY2-Gal4 to drive UAS-dnEGFR. *Drosophila melanogaster* and *D. willistoni* flies were fed colchicine mixed with a yeast paste (25 μg/ml) for 24 h before

dissection and egg collection as previously described (Peri and Roth 2000). Colchicine treatment for *D. nebulosa* was carried out for 48 h at the same concentration. Colchicine treatment indirectly mislocalizes EGFR signaling by destabilizing microtubules involved in oocyte nucleus migration, and thus the EGFR ligand, Grk (Neuman-Silberberg and Schupbach 1994).

Gene Cloning, In Situ Hybridization, Immunoassays, Microscopy

RNA extractions from *D. melanogaster* and *D. willistoni* ovaries were carried out using RNeasy Mini Kit (Qiagen). cDNA was synthesized using Taqman Kit (Roch) and a partial region of all nine *Cp* genes was amplified using custom-designed primers (supplementary table S1, Supplementary Material online). Polymerase chain reaction (PCR) was done using the MJ Mini (BioRad) thermocycler, and products were cloned using StrataClone PCR Cloning kit (Stratagene). Plasmids were recovered using the QIAprep spin Miniprep Kit (Qiagen). Each gene was sequenced (GeneWiz) and compared with known sequences on FlyBase. RNA DIG-labeled probes were synthesized and in situ hybridization was performed (Yakoby, Bristow, et al. 2008). In situ hybridization for *D. nebulosa* was carried out using *D. willistoni* probes. Immunoassyas were done with the following antibodoes: Fasiclin III (FasIII–1:100, Developmental Studies Hybridoma Bank [DSHB]) staining was carried out as described (Yakoby, Lembong, et al. 2008). Secondary antibodies: 488 anti-mouse (Invitrogen) were used (1:1,000). Egg chambers were imaged using a Leica DM2500 compound microscope (Leica). SEM imaging was done as previously reported (Niepielko et al. 2011).

Matrices and Matrix Analysis

Gene patterns are represented as binary vectors consisting of mutually exclusive domains at four different developmental stages of *Drosophila* oogenesis (Spradling 1993; Yakoby, Bristow, et al. 2008). In the original combinatorial code (Yakoby, Bristow, et al. 2008), the anterior, dorsal, and midline domains overlap. Here, we modified them to be mutually exclusive. The anterior domain was split into AD and AV domains, and a domain for DR and posterior (P) was added along with repression domains (for the complete details see fig. 2). Representation and manipulation of matrices were conducted with MATLAB (The MathWorks, Natick, MA) and displayed using the *imagesc* command.

Accumulation of domain usage was summed in excel and displayed as a bar graph for each species and color coded by stage. Pairwise comparisons between domains were calculated in MATLAB as percent co-occurrence between all domain pairs of two species. Co-occurrences is depicted as a numeral and displayed using the *imagesc* command. The fraction is the co-occurrence value divided by the higher of the two domain utilizations. This fraction is represented by a color scale displayed underneath the co-occurrence value. Hierarchical clustering was conducted (Eisen et al. 1998) on an averaged expression matrix of all three species to determine expression domain relatedness. Bootstrap values were calculated by assembling a an unweighted pair group method (UPGMA) tree in Mega5 (Tamura et al. 2011) with 1,000 bootstrap trees, representing domain conservation with individual nucleotides. Distance was determined with the Euclidean distance metric and average linkage was used for tree generation. Clustergrams are generated such that genes cluster on one axis and domains cluster on the other.

Supplementary Material

Supplementary figures S1–S3 and table S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank T. Schupbach, D. Stern, M. O'Connor, and DSHB for fly species and reagents and also thank the critical comments made by J. Nam and V. Singh. The authors are also thankful for the fruitful discussions with members of the Yakoby Lab. M.G.N. and R.A.M. were partially supported by the Center for Computational and Integrative Biology, Rutgers-Camden. D.L. was supported by NSF-Q-STEP DUE-0856435 and C.R. by the NSF-REU-0754800 programs, Rutgers-Camden. The authors thank the Electron Microscopy Laboratory of the Biology Department at Rutgers-Camden for the use of the Leo 1450EP scanning electron microscope, grant number NSF DBI-0216233. R.A.M. was supported by NSF-CAREER Award (IOS-1149144) granted to N.Y. The research, publication, and M.G.N. were supported by the Rutgers Faculty Research grant number 281715 and by the National Institute of General Medical Sciences of the National Institutes of Health under award number 1R15GM101597-01A1, awarded to N.Y.

References

Berg CA. 2005. The *Drosophila* shell game: patterning genes and morphological change. *Trends Genet.* 21:346–355.

Cavaliere V, Bernardi F, Romani P, Duchi S, Gargiulo G. 2008. Building up the *Drosophila* eggshell: first of all the eggshell genes must be transcribed. *Dev Dyn.* 237:2061–2072.

Deng WM, Bownes M. 1997. Two signalling pathways specify localised expression of the broad-complex in *Drosophila* eggshell patterning and morphogenesis. *Development* 124:4639–4647.

Dobens LL, Raftery LA. 1998. *Drosophila* oogenesis: a model system to understand TGF-beta/Dpp directed cell morphogenesis. *Ann NY Acad Sci.* 857:245–247.

Dong X, Tsuda L, Zavitz KH, Lin M, Li S, Carthew RW, Zipursky SL. 1999. ebi regulates epidermal growth factor receptor signaling pathways in *Drosophila*. *Genes Dev.* 13:954–965.

Dorman JB, James KE, Fraser SE, Kiehart DP, Berg CA. 2004. Bullwinkle is required for epithelial morphogenesis during *Drosophila* oogenesis. *Dev Biol.* 267:320–341.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A.* 95:14863–14868.

Fakhouri M, Elalayli M, Sherling D, Hall JD, Miller E, Sun X, Wells L, LeMosy EK. 2006. Minor proteins and enzymes of the *Drosophila* eggshell matrix. *Dev Biol.* 293:127–141.

Frankel N, Erezyilmaz DF, McGregor AP, Wang S, Payre F, Stern DL. 2011. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* 474:598–603.

Frankel N, Wang S, Stern DL. 2012. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proc Natl Acad Sci U S A.* 109:20975–20979.

Fuchs A, Cheung LS, Charbinnier E, Shvartsman SY, Pyrowolakis G. 2012. Transcriptional interpertation of the EGF receptor signaling gradient. *Proc Natl Acad Sci U S A.* 109(5):1572–1577.

Goltsev Y, Fuse N, Frasch M, Zinzen RP, Lanzaro G, Levine M. 2007. Evolution of the dorsal-ventral patterning network in the mosquito, *Anopheles gambiae*. *Development* 134:2415–2424.

Griffin-Shea R, Thireos G, Kafatos FC. 1982. Organization of a cluster of four *chorion* genes in *Drosophila* and its relationship to developmental expression and amplification. *Dev Biol.* 91: 325–336.

Hinton HE, Service MW. 1969. The surface structure of aedine eggs as seen with the scanning electron microscope. *Ann Trop Med Parasitol.* 63:409–411.

Horne-Badovinac S, Bilder D. 2005. Mass transit: epithelial morphogenesis in the *Drosophila* egg chamber. *Dev Dyn.* 232:559–574.

Kagesawa T, Nakamura Y, Nishikawa M, Akiyama Y, Kajiwara M, Matsuno K. 2008. Distinct activation patterns of EGF receptor signaling in the homoplastic evolution of eggshell morphology in genus *Drosophila*. *Mech Dev.* 125:1020–1032.

Konikoff CE, Karr TL, McCutchan M, Newfeld SJ, Kumar S. 2012. Comparison of embryonic expression within multigene families using the FlyExpress discovery platform reveals more spatial than temporal divergence. *Dev Dyn.* 241:150–160.

Lembong J, Yakoby N, Shvartsman SY. 2009. Pattern formation by dynamically interacting network motifs. *Proc Natl Acad Sci U S A.* 106(9):3213–3218.

Mann RS, Carroll SB. 2002. Molecular mechanisms of selector gene function and evolution. *Curr Opin Genet Dev.* 12:592–600.

Margaritis L, Dellas K, Kalantzi M, Kambysellis M. 1983. The eggshell of Hawaiian *Drosophila*: structural and biochemical studies in *D. grimshawi* and comparison to *D. melanogaster*. *Wilhelm Roux's Arch Dev Biol.* 192:303–316.

Margaritis LH, Kafatos FC, Petri WH. 1980. The eggshell of *Drosophila melanogaster*. I. Fine structure of the layers and regions of the wild-type eggshell. *J Cell Sci.* 43:1–35.

Marmion RA, Jevtic M, Springhorn A, Pyrowolakis G, Yakoby N. 2013. The *Drosophila* BMPRII, wishful thinking, is required for eggshell patterning. *Dev Biol.* 375:45–53.

Neuman-Silberberg FS, Schupbach T. 1994. Dorsoventral axis formation in *Drosophila* depends on the correct dosage of the gene *gurken*. *Development* 120:2457–2463.

Niepielko MG, Hernaiz-Hernandez Y, Yakoby N. 2011. BMP signaling dynamics in the follicle cells of multiple *Drosophila* species. *Dev Biol.* 354:151–159.

Niepielko MG, Ip K, Kanodia JS, Lun DS, Yakoby N. 2012. The evolution of BMP signaling in *Drosophila* oogenesis: a receptor-based mechanism. *Biophys J.* 102:1722–1730.

Osterfield M, Du X, Schupbach T, Wieschaus E, Shvartsman SY. 2013. Three-dimensional epithelial morphogenesis in the developing *Drosophila* egg. *Dev Cell.* 24: 400–410.

Parks S, Wakimoto B, Spradling A. 1986. Replication and expression of an X-linked cluster of *Drosophila chorion* genes. *Dev Biol.* 117: 294–305.

Peri F, Roth S. 2000. Combined activities of Gurken and decapentaplegic specify dorsal chorion structures of the *Drosophila* egg. *Development* 127:841–850.

Queenan AM, Ghabrial A, Schupbach T. 1997. Ectopic activation of torpedo/Egfr, a *Drosophila* receptor tyrosine kinase, dorsalizes both the eggshell and the embryo. *Development* 124:3871–3880.

Ruohola-Baker H, Grell E, Chou TB, Baker D, Jan LY, Jan YN. 1993. Spatially localized rhomboid is required for establishment of the dorsal-ventral axis in *Drosophila* oogenesis. *Cell* 73:953–965.

Sapir A, Schweitzer R, Shilo BZ. 1998. Sequential activation of the EGF receptor pathway during *Drosophila* oogenesis establishes the dorsoventral axis. *Development* 125:191–200.

Shravage BV, Altmann G, Technau M, Roth S. 2007. The role of Dpp and its inhibitors during eggshell patterning in *Drosophila*. *Development* 134:2261–2271.

Shubin N, Tabin C, Carroll S. 2009. Deep homology and the origins of evolutionary novelty. *Nature* 457:818–823.

Spradling AC. 1981. The organization and amplification of two chromosomal domains containing *Drosophila chorion* genes. *Cell* 27: 193–201.

Spradling AC. 1993. Developmental genetics of oogenesis. In: Bate M, Martinez-Arias A, editors. The development of *Drosophila melanogaster*. Plainview: Cold Spring Harbor Laboratory Press. p. 1–70.

Stathopoulos A, Levine M. 2005. Genomic regulatory networks and animal development. *Dev Cell.* 9:449–462.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.

Tolias PP, Konsolaki M, Halfon MS, Stroumbakis ND, Kafatos FC. 1993. Elements controlling follicular expression of the s36 *chorion* gene during *Drosophila* oogenesis. *Mol Cell Biol.* 13:5898–5906.

Tootle TL, Williams D, Hubb A, Frederick R, Spradling A. 2011. *Drosophila* eggshell production: identification of new genes and coordination by Pxt. *PloS One* 6:e19943.

Twombly V, Blackman RK, Jin H, Graff JM, Padgett RW, Gelbart WM. 1996. The TGF-beta signaling pathway is essential for *Drosophila* oogenesis. *Development* 122:1555–1565.

Van Buskirk C, Schupbach T. 1999. Versatility in signalling: multiple responses to EGF receptor activation during *Drosophila* oogenesis. *Trends Cell Biol.* 9:1–4.

Vreede BM, Lynch JA, Roth S, Sucena E. 2013. Co-option of a coordinate system defined by the EGFr and Dpp pathways in the evolution of a morphological novelty. *Evo Devo* 4:7.

Ward EJ, Berg CA. 2005. Juxtaposition between two cell types is necessary for dorsal appendage tube formation. *Mech Dev.* 122:241–255.

Waring GL. 2000. Morphogenesis of the eggshell in *Drosophila*. *Int Rev Cytol.* 198:67–108.

Yakoby N, Bristow CA, Gong D, Schafer X, Lembong J, Zartman JJ, Halfon MS, Schupbach T, Shvartsman SY. 2008. A combinatorial code for pattern formation in *Drosophila* oogenesis. *Dev Cell.* 15: 725–737.

Yakoby N, Lembong J, Schupbach T, Shvartsman SY. 2008. *Drosophila* eggshell is patterned by sequential action of feedforward and feedback loops. *Development* 135:343–351.